

Solución de los Problemas Propuestos

1. En la siguiente tabla se dan los datos correspondientes a las notas de Matemáticas de 60 alumnos de 1º Bachillerato.

Notas	IN: [1, 5)	SF: [5, 6)	BI: [6, 7)	NT: [7, 9)	SB: [9, 10]
N.º de alumnos	20	13	12	10	5

- a) Haz una tabla de frecuencias y porcentajes, simple y acumulada.
- b) Dibuja el correspondiente histograma.
- c) Representa los datos mediante un diagrama de sectores y mediante una poligonal acumulativa.

Solución:

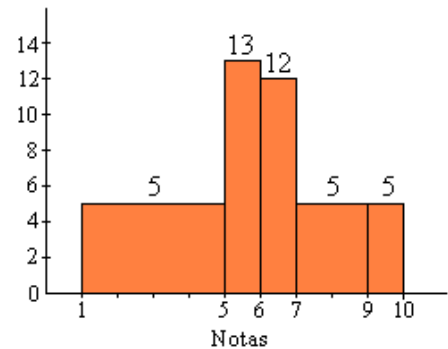
a)

Notas	M.c.	f_i	F_i	f_{ri}	%	%a
[1, 5)	3	20	20	0,333	33,3	33,3
[5, 6)	5,5	13	33	0,217	21,7	55
[6, 7)	6,5	12	45	0,200	20	75
[7, 9)	8	10	55	0,167	16,7	91,7
[9, 10]	9,5	5	60	0,083	8,3	100
TOTALES		60		1	100	

b) La altura de cada rectángulo se halla dividiendo la frecuencia que representa entre la longitud del intervalo:

$$20 : 4 = 5; \quad 13 : 1 = 13; \quad 12 : 1 = 12;$$

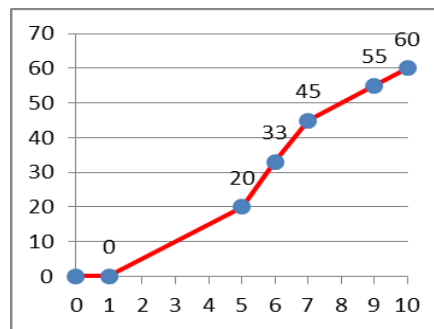
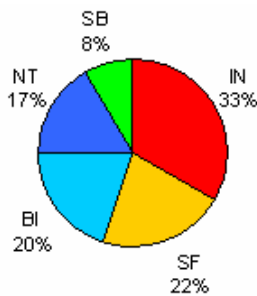
$$10 : 2 = 5; \quad 5 : 1 = 5.$$



c) Como a los 60 alumnos les corresponde 360° (el círculo completo) \Rightarrow a 1 alumno \rightarrow 6°.

Por tanto, la asignación de sectores para cada nota es:

IN: 120°; SF: 78°; BI: 72°; NT: 60°; SB: 30°. (En el gráfico se aproximan los porcentajes).



2. El número de turismos matriculados en España, para el período 2007/2018, se da en la siguiente tabla:

Año	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Miles de turismos	1634	1185	971	1000	818	711	742	890	1094	1230	1342	1425

- a) Tomando como base 100 el número de turismos matriculados en el año 2007 expresa en números índices la variación de la serie.
- b) Representa los datos mediante una poligonal simple usando Excel.

Solución:

a) Si asignamos a 2007 la base 100, multiplicando el número de vehículos matriculados en cada año por $\frac{100}{1634}$, se obtiene:

Año	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Miles de turismos	1634	1185	971	1000	818	711	742	890	1094	1230	1342	1425
Índice	100	72,5	59,4	61,2	50,0	43,5	45,4	54,5	67,0	75,3	82,1	87,2

(Puedes obtener pequeñas diferencias, de décimas, en los resultados, debidas al redondeo en los miles de turismos. Los datos proporcionados por el INE son los adjuntos).

b) Para representar con Excel:

- 1) Teclear los datos del problema.
- 2) Insertar gráfico → ir a gráficos recomendados (poligonal; barras; ...).
- 3) Títulos del eje → indicarlos.
- 4) Estilo de gráfico → buscar el preferido; también puede elegirse colores.



Nota: En el año 2008 se produjo en España, y en más países, una gran crisis económica, que se aprecia perfectamente en el gráfico.

3. La precipitación (P) y la temperatura media mensual (T) registradas en Soria a lo largo del año son:

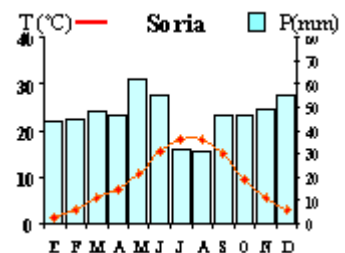
Mes	E	F	M	A	M	J	J	A	S	O	N	D
P (mm)	44	45	48	47	62	55	32	31	47	46	49	55
T (°C)	1,3	3,1	5,6	7,5	10,6	15,6	18,1	18,1	15	9,4	5,6	3,1

Representa gráficamente estos datos mediante un climograma.

Solución:

En estos gráficos la temperatura suele representarse mediante una línea poligonal; sus valores se indican en el eje vertical de la izquierda.

La precipitación se representa mediante un diagrama de barras (con escasa separación entre dos barras consecutivas); su magnitud suele referenciarse en el eje vertical de la derecha.



Nota: Puede verse otra forma de hacer este gráfico en el [Problema Resuelto n. 3 de Matemáticas I CCSS, Tema 12.0, página 111](#).

4. Siete estudiantes han leído este curso el siguiente número de libros: 3, 4, 5, 6, 5, 7, 5.

Para estos datos, determina:

- a) La media. b) La mediana. c) La moda. d) El rango.

Solución:

$$a) \bar{x} = \frac{3+4+5+6+5+7+5}{7} = \frac{35}{7} = 5.$$

b) Se ordenan los datos: 3, 4, 5, 5, 5, 6, 7. El cuarto dato es la mediana: $Me = 5$.

c) El 5 es el dato que más se repite: $Mo = 5$.

d) Rango: $7 - 3 = 4$.

5. En una empresa hay 3 directivos/as, 50 operarios/as y 8 vendedores/as. Los sueldos mensuales, en euros, de cada categoría son los siguientes: directivos/as, 4000 €; operarios/as, 1400 €; vendedores/as, 2000 €.

a) Halla la moda, la mediana y la media de los sueldos.

b) ¿Qué medida es más representativa del promedio?

Solución:

a) El sueldo que más se repite es la moda: 1400 €.

Mediana:

Ordenados los sueldos (en total hay 61 trabajadores/as) de menor a mayor, los 50 primeros datos son 1400. El valor del dato 31º, que sería el correspondiente a la mediana, vale 1400 €.

Media (ponderada):

$$\bar{x}_p = \frac{1400 \cdot 50 + 2000 \cdot 8 + 4000 \cdot 3}{50 + 8 + 3} = \frac{98000}{61} = 1606,56.$$

b) Ninguna es mala, aunque dado el peso de los operarios la más representativa sería la moda.

6. En primero de bachillerato de un centro escolar hay tres grupos, A, B y C, con 30, 35 y 25 alumnos/as, respectivamente. La nota media en Matemáticas fue, también respectivamente, de 5,3, 6,5 y 5,6. Halla la nota media de Matemáticas de todos los alumnos/as de primero.

Solución:

Se trata de una media ponderada: hay que tener en cuenta el número de alumnos de cada grupo.

$$\bar{x}_p = \frac{30 \cdot 5,3 + 35 \cdot 6,5 + 25 \cdot 5,6}{90} = 5,85.$$

7. El cociente intelectual de los 210 alumnos de un centro de bachillerato se da en la tabla:

Intervalo	[82, 90)	[90, 98)	[98, 106)	[106, 114)	[114, 122)	[122, 130)	[130, 138)	[138, 146)
Frecuencia	12	32	49	54	30	17	11	5

a) Calcula los cuartiles y el rango intercuartílico.

b) Halla la diferencia entre los deciles 3 y 6.

c) Calcula la puntuación necesaria para pertenecer al 15 % de alumnos con mayor cociente intelectual.

Solución:

Para contestar a estas preguntas es necesario hallar la tabla de frecuencias acumuladas.

Intervalo	[82, 90)	[90, 98)	[98, 106)	[106, 114)	[114, 122)	[122, 130)	[130, 138)	[138, 146)
Frecuencia	12	32	49	54	30	17	11	5
F. acum.	12	44	93	147	177	194	205	210

Hay 210 alumnos.

Si se dividen en cuatro cuartiles, cada uno de ellos contiene a 52,5 alumnos. Vendrían dados por el cociente intelectual de los alumnos 52,5 (para evitar sufrimientos elegimos al 53º), cuartil C1; alumno 105º, C2; alumno 158º, C3.

Si se dividen en deciles hay que hacer grupos de 21.

a) La posición de C1 es la del elemento 53º: $\frac{210}{4} = 52,5 \rightarrow 53^\circ$. Por tanto, $C1 = 98 + 9 \cdot \frac{8}{49} = 99,5$. ($53 = 44 + \underline{9} \rightarrow 98$ es el cociente intelectual del individuo 44º; los siguientes 49 individuos avanzan 8 puntos en total, luego $\underline{9}$ de ellos avanzarán $\underline{9} \cdot \frac{8}{49}$).

El valor de C2 es la puntuación del individuo situado en la posición de C2 es 105º (= 93º + 12º).

Por tanto, $C2 = 106 + 12 \cdot \frac{8}{54} = 107,8$. (106 es el cociente intelectual del individuo 93º).

La posición de C3 es 158. Por tanto, $C3 = 114 + 11 \cdot \frac{8}{30} = 116,9$.

Recorrido intercuartílico: $C3 - C1 = 17,4$.

b) La puntuación de D3 es la correspondiente al individuo 63ª $\Rightarrow D3 = 98 + 19 \cdot \frac{8}{49} = 101,1$.

La posición de D6 es la 126ª. Por tanto, $D6 = 106 + 33 \cdot \frac{8}{54} = 110,9$.

Por tanto, $D6 - D3 = 110,9 - 101,1 = 9,8$.

c) Hay que calcular el percentil 85, cuya posición es la 179ª.

Por tanto, $P_{85} = 122 + 2 \cdot \frac{8}{17} = 122,9$.

8. Se ha preguntado a 50 mujeres sobre su número de hijos, obteniéndose los resultados:

0 1 1 2 2 0 1 5 4 3 2 1 0 2 0 0 2 1 4 2 2 0 1 3 2
1 2 3 3 5 2 1 1 4 1 4 2 3 1 3 1 0 0 2 2 2 0 3 1 2

Construye la tabla de frecuencias y calcula la media, varianza y desviación típica.

Solución:

El número de hijos y la frecuencia correspondiente se indican en las dos primeras columnas de la siguiente tabla; las demás columnas servirán para el cálculo de los parámetros.

Hijos: x_i	f_i	$f_i \cdot x_i$	$f_i \cdot x_i^2$
0	9	0	0
1	13	13	13
2	15	30	120
3	7	21	63
4	4	16	64
5	2	10	125
Sumas	50	90	250

De donde:

$$\bar{x} = \frac{\sum f_i x_i}{n} = \frac{90}{50} = 1,8.$$

$$\sigma^2 = \frac{\sum f_i x_i^2}{n} - \bar{x}^2 = \frac{250}{50} - 1,8^2 = 1,76 \Rightarrow \sigma \approx 1,33.$$

Secuencia con la calculadora (con daos agrupados):

(1) Poner la calculadora en el modo SD y borrar los datos de memoria: **MODE** **2** **SHIFT** **AC**.

(2) Introducir los datos como sigue:

0 **SHIFT** **,** 9 **M+** 1 **SHIFT** **,** 13 **M+** 2 **SHIFT** **,** 15 **M+** 3 **SHIFT** **,** 7 **M+**
4 **SHIFT** **,** 4 **M+** 5 **SHIFT** **,** 2 **M+**.

(3) Los parámetros se obtienen pulsando:

$$\boxed{\text{SHIFT}} \boxed{2} \boxed{1} \boxed{=} \rightarrow \bar{x} = 1,8; \quad \boxed{\text{SHIFT}} \boxed{2} \boxed{2} \boxed{=} \rightarrow \sigma_x = 1,326649916.$$

Nota: Utilizando GeoGebra se obtiene:

$$\text{media}[\{0,1,2,3,4,5\}, \{9,13,15,7,4,2\}] \rightarrow \text{sale } 1,8.$$

$$\text{de}[\{0,1,2,3,4,5\}, \{9,13,15,7,4,2\}] \rightarrow \text{sale } 1,3266.$$

9. Se ha realizado una encuesta a los 40 empleados de una empresa para saber cuánto tiempo tardan en llegar desde su casa hasta su puesto de trabajo. Las respuestas, en minutos, son las siguientes:

30 42 37 50 15 35 90 65 38 45 30 12 78 20 35 41 25 32 85 25
41 28 50 30 20 60 14 36 48 32 27 30 76 30 51 28 25 22 17 10

a) Construye la tabla de frecuencias agrupando los datos en intervalos.

b) Calcula la mediana, la moda, la media y la desviación típica.

Solución:

a)

Tiempo	x_i	f_i	$f_i x_i$	$f_i x_i^2$	F_i
[10, 20)	15	5	75	1125	5
[20, 30)	25	9	225	5625	14
[30, 40)	35	12	420	14700	26
[40, 50)	45	5	225	10125	31
[50, 60)	55	3	165	9075	34
[60, 70)	65	2	130	8450	36
[70, 80)	75	2	150	11250	38
[80, 90]	85	2	170	14450	40
TOTALES		40	1560	74800	

b) La posición de la mediana es la 20ª, es decir, está en el intervalo [30, 40).

$$\text{Por tanto, } Me = 30 + 6 \cdot \frac{12}{10} = 37,2 \text{ minutos.}$$

El intervalo modal es [30, 40) y la clase modal es $Mo = 35$.

$$\text{Media: } \bar{x} = \frac{\sum f_i x_i}{n} = \frac{1560}{40} = 39 \text{ minutos.}$$

$$\text{Desviación típica: } \sigma = \sqrt{\frac{\sum f_i x_i^2}{n} - \bar{x}^2} = \sqrt{\frac{74800}{40} - 39^2} = 18,7 \text{ minutos.}$$

10. Los rendimientos medios (en kilogramos por hectárea) en España, para los cereales que se indican, fueron:

Año	2010	2011	2012	2013	2014
Trigo	2150	3100	2300	2830	2840
Maíz	9450	9220	9720	9510	9110

Halla los rendimientos medios para el quinquenio de cada cereal. ¿Qué cereal es más fiable?

Solución:

T representa al trigo y M al maíz.

$$\bar{x}(T) = 2644 \text{ kg/ha; } \sigma(T) = 358,7 \text{ kg/ha} \Rightarrow CV(T) = 13,57 \%$$

$$\bar{x}(M) = 9402 \text{ kg/ha; } \sigma(M) = 216 \text{ kg/ha} \Rightarrow CV(M) = 2,3 \%$$

Es más fiable el cereal que tenga asociado un coeficiente de variación menor: el maíz.

Con calculadora (datos imples):

Trigo → (1) Borrar los datos de memoria: Pulsar : **MODE** **1** **SHIFT** **AC**.

(2) 2150 **M+** 3100 **M+** 2300 **M+** 2830 **M+** 2840 **M+**.

(3) **SHIFT** **2** **1** → \bar{x} (media) → 2644; **SHIFT** **2** **2** → σ_n → 358,6976...

Maíz → (1) Borrar los datos de memoria: Pulsar **SHIFT** **AC**.

(2) 9450 **M+** 9220 **M+** 9720 **M+** 9510 **M+** 9110 **M+**.

(3) **SHIFT** **2** **1** → \bar{x} (media) → 9402; **SHIFT** **2** **2** → σ_n → 216,0925...

11. A un congreso asisten seis mujeres cuyas edades son:

27 34 38 42 33 36 (años)

a) Calcula la media y varianza de sus edades.

b) Cinco años después coinciden las mismas mujeres. A partir de los cálculos anteriores, halla la nueva media y varianza de sus edades.

Solución:

a) $\bar{x} = 35$ años. $\sigma^2 = 21,33$.

b) Pasados 5 años las edades son: 32, 39, 43, 47, 38, 41.

La nueva media sube 5 años: $\bar{x} [+5] = \bar{x} + 5 = 35 + 5 = 40$.

La varianza no cambia: $\sigma^2 [+5] = \sigma^2 = 21,33$.

12. El siguiente gráfico representa un total de 600 elementos. ¿Cuál es la frecuencia de cada categoría?

Solución:

Hay que resolver reglas de tres directas.

Si $360^\circ \rightarrow 600$; $1^\circ \rightarrow \frac{600}{360} = \frac{5}{3}$ elementos.

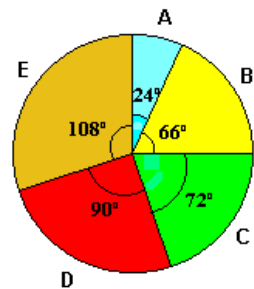
24° → 40 elementos;

66° → 110 elementos;

72° → 120 elementos;

90° → 150 elementos;

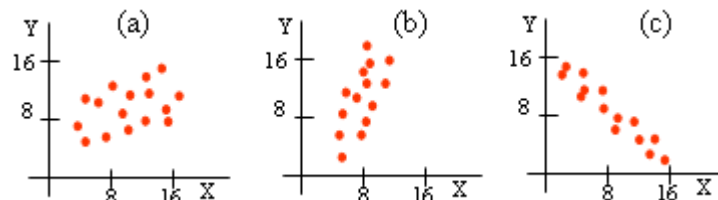
108° → 180 elementos.



Se obtiene:

Categoría	A	B	C	D	E
Frecuencia: f_i	40	110	120	150	180

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



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

Solución:

a) La recta de pendiente negativa debe corresponderse con la nube (c) → $y = -x + 16$.

La recta de ordenada -12 solo puede asociarse con la nube (b) → $y = 2x - 12$.

En consecuencia, (a) → $y = 0,5x + 5$.

b) La nube (a) es creciente y poco compacta: coeficiente de correlación positivo y no próximo a 1 $\rightarrow r = 0,4$. La nube (b) es directa y estrecha $\rightarrow r = 0,7$.
 La nube (c) es decreciente y estrecha: coeficiente de correlación negativo $\rightarrow r = -0,85$.

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

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

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

$$y = 350 - 2,1x; \quad y = 460 - 2,1x; \quad y = 406 + 2,1x.$$

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

Solución:

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

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

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

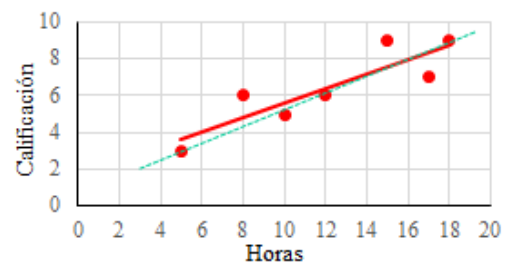
Estudio (h)	5	8	10	12	15	17	18
Calificación	3	6	5	6	9	7	9

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

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

Solución:

a) En la figura adjunta se representan los puntos de la tabla anterior. La línea de trazos puede ser una recta de regresión aceptable. (La buena es la de trazo continuo; se hallará en el problema que sigue).



b) Obviamente, la correlación es directa y fuerte.

16. Calcula, paso a paso (sin utilizar la calculadora en modo estadístico), el coeficiente de correlación y la recta de regresión asociada a los datos del problema anterior.

Solución:

El coeficiente de correlación vale $r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$. Para hallarlo hay que calcular las desviaciones típicas

marginales y la covarianza.

Utilizaremos las fórmulas:

$$\sigma_x = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2}; \quad \sigma_y = \sqrt{\frac{\sum y_i^2}{n} - \bar{y}^2}; \quad \sigma_{xy} = \frac{\sum x_i y_i}{n} - \bar{x} \cdot \bar{y}.$$

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

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

Con esto:

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

$$\sigma_x = \sqrt{\frac{1171}{7} - 12,14^2} = 4,46; \quad \sigma_y = \sqrt{\frac{317}{7} - 6,43^2} = 1,99; \quad \sigma_{xy} = \frac{601}{7} - 12,14 \cdot 6,43 = 7,80.$$

Por último, $r = \frac{7,80}{4,46 \cdot 1,99} = 0,88.$

Al ser r próximo a +1, la correlación entre las horas de estudio y las notas de examen es directa y fuerte: a más horas de estudio, mejor nota de examen.

La recta de regresión es

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

(Es la recta continua de la figura anterior).

Notas:

1. Los resultados anteriores están redondeados a las centésimas.

Con más precisión (con la calculadora): $r = 0,879689$; $y = 0,393004x + 1,656379$.

2. Los parámetros estadísticos σ_x , σ_y y σ_{xy} también pueden denotarse con las letras: s_x , s_y y s_{xy} , respectivamente.

3. Con **GeoGebra**, tecleando:

CoeficienteCorrelación({(5,3), (8,6),(10,5),(12,6),(15,9),(17,7),(18,9)}) se obtiene $\rightarrow 0,8797$.

AjusteLineal({(5,3), (8,6),(10,5),(12,6),(15,9),(17,7),(18,9)}) se obtiene $\rightarrow y = 0,393x + 1,6564$.

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

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

b) ¿Cuál es el valor que correspondería según dicha recta a $X = 7$?

Solución:

Se forma la tabla:

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

Se obtiene.

$$\bar{x} = 5; \sigma_x^2 = \frac{183}{5} - 5^2 = 11,6; \bar{y} = 6; \sigma_{xy} = \frac{102}{5} - 5 \cdot 6 = -9,6.$$

La ecuación de la recta de regresión es:

$$y - \bar{y} = \frac{\sigma_{xy}}{\sigma_x^2}(x - \bar{x}) \Rightarrow y = -0,8276x + 10,138.$$

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

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

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

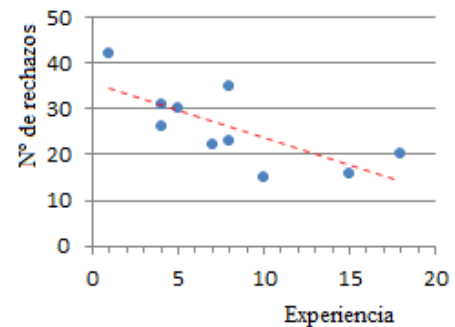
a) Representa el diagrama de dispersión correspondiente a esos datos. ¿Sugiere la gráfica alguna asociación lineal?

b) ¿Cómo calificarías la correlación?

Solución:

a) La nube de puntos es decreciente y puede aceptarse que, aunque no sea estrecha, sus puntos están cercanos a una recta como la que se ha trazado en la figura adjunta.

b) La correlación es claramente inversa y quizás pueda calificarse de fuerte, aunque algunos puntos [B(8, 35) o C(10, 15)] pueden sugerir alguna duda; por eso es necesario hacer un estudio cuantitativo.



19. Para los datos del problema anterior, halla con ayuda de la calculadora:

a) Las medias y desviaciones típicas marginales.

b) La covarianza.

c) El coeficiente de correlación lineal.

d) La recta de regresión de Y sobre X.

e) El número de rechazos que hay que esperar para una persona con 20 semanas de experiencia.

f) Detalla la solución con Excel.

Solución:

Sumas:

$$n = 10; \sum x_i = 80; \sum y_i = 260; \sum x_i^2 = 884; \sum y_i^2 = 7420; \sum x_i y_i = 1788.$$

a) $\bar{x} = 8; \sigma_x = 4,93963; \bar{y} = 26; \sigma_y = 8,12403$.

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

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

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

e) $y(20) = -1,19672 \cdot 20 + 35,5737 = 11,6391$, que se aproxima a 12, el entero más próximo.

f) Para hallar la solución con Excel:

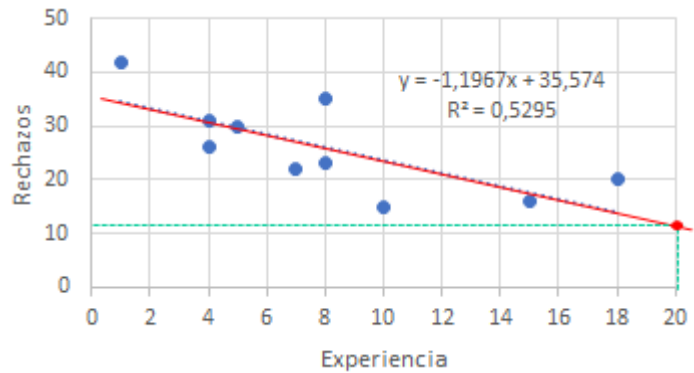
1) Teclar los datos.

2) Insertar gráfico → dispersión.



- 3) Títulos del eje.
- 4) Línea de tendencia. Opciones de línea de tendencia: presentar ecuación en el gráfico; presentar el valor R^2 en el gráfico.

Aparece el gráfico adjunto.
La línea roja y el punto (20, 11,6391) se han añadido a posteriori.



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

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

- a) Calcula el coeficiente de correlación lineal.
- b) Halla la expresión de la recta que permita estimar los valores fetales a partir de los maternos.

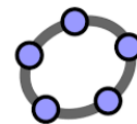
Solución:

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

Otros parámetros de interés:

$$\bar{x} = 7; \quad \sigma_x = 3,2659; \quad \bar{y} = 4,667; \quad \sigma_y = 2,1344.$$

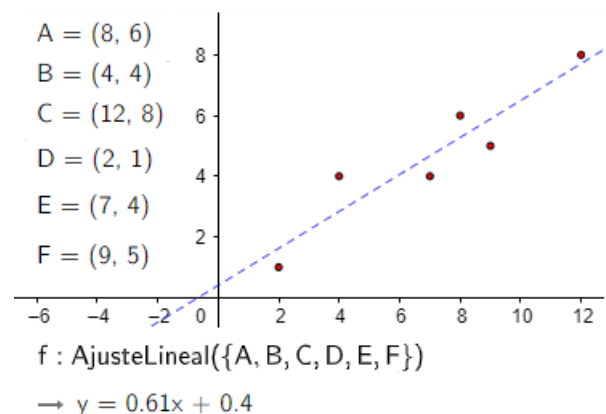
También se obtiene: $A = 0,40$; $B = 0,609$.



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

Nota: Este gráfico se ha hecho con GeoGebra. Para ello:

- 1) Se teclean las coordenadas de los puntos.
- 2) Con el botón derecho del ratón se abarcan todos los puntos: de manera automática se obtiene la línea de tendencia y su ecuación.
- 3) El tamaño y aspecto de los puntos y de la línea de tendencia se deciden en “propiedades”.



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

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

- a) Halla la recta de regresión.
- b) Sabiendo que una alumna obtuvo 130 puntos en el test, pero no realizó el examen de Matemáticas, predice, si es posible, la nota que hubiese obtenido.

Solución:

a) Utilizando la calculadora se obtiene que:

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

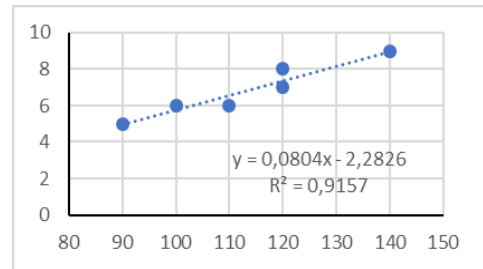
El coeficiente de correlación vale $r = 0,9569$.

b) Si una alumna obtuvo en el test 130 puntos, se puede estimar que su nota en Matemáticas sería:
 $y = 2,28 + 0,08 \cdot 130 = 8,12$.

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

→ El gráfico adjunto se ha confeccionado en Excel.

$R^2 = 0,9157$ es el coeficiente de determinación (el cuadrado del de correlación), que indica el porcentaje explicado de la variación de la nota de Matemáticas a partir de la puntuación obtenida en el test: el 91,57 %.



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

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

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

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

Solución:

a) Regresión de Y sobre X: $Y = 68,1853 + 0,621859 \cdot X$.

Los datos se introducen en el orden:

170 ['] 172 [M+] 173 ['] 177 [M+] ...
175 ['] 187 [M+]

Regresión de X sobre Y: $X = 77,4406 + 0,545082 \cdot Y$.

Los datos se introducen en el orden:

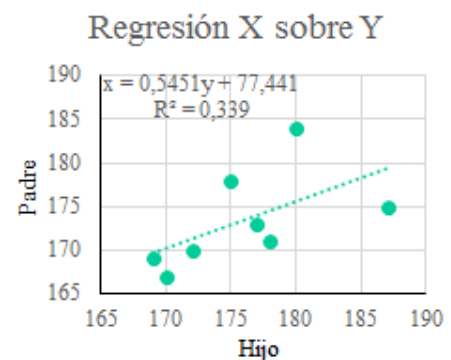
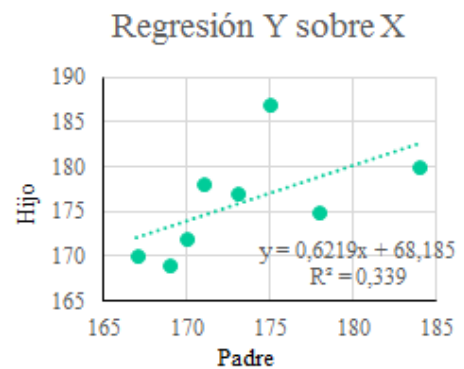
172 ['] 170 [M+] 177 ['] 173 [M+] ...
187 ['] 175 [M+]

(Observa que las rectas son distintas).

b) Si el padre (X) mide 174 cm (en la primera ecuación), entonces, para el hijo cabe esperar una estatura de $Y = 176,4$ cm.

Si el hijo (Y) mide 190 cm (en la segunda ecuación), entonces, para el padre puede suponerse una estatura de $X = 181$ cm.

En el gráfico de la regresión de X sobre Y, en el eje horizontal se representan los valores de altura del hijo (Y); en el vertical, los del padre (X).



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

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

a) Calcula, utilizando la recta de regresión, el diámetro que se puede predecir para árboles de 10 y 20 años.

b) Compara el resultado anterior con los valores observados en la tabla. Razona el porqué de las diferencias.

Solución:

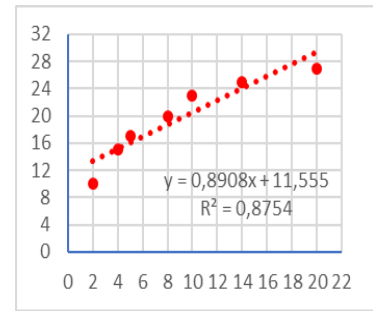
a) X = años; Y = diámetro.

$$\bar{x} = 9; \sigma_x = 5,83; \bar{y} = 19,57; \sigma_y = 5,55; r = 0,93563;$$

$$y = 11,55 + 0,89x.$$

b) $Y(10) = 20,45$; $Y(20) = 29,35$.

Las diferencias son debidas a que la recta de regresión da la media del valor esperado. Esto es, el diámetro medio de *muchos* árboles de 10 años será de 20,45 cm, pero cada árbol de 10 años tendrá un diámetro particular: el diámetro de los árboles de 10 años se distribuye normalmente con media 20,45 cm.



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

X: N.º de horas	0	1	2	3	4	5
Y: N.º de bacterias	12	19	23	34	56	62

Calcula:

- Las medias y desviaciones típicas de las variables, número de horas y número de bacterias.
- La covarianza de la variable bidimensional.
- El coeficiente de correlación, dando una interpretación del resultado.
- La recta de regresión de Y sobre X.

Solución:

Sumas:

$$\sum x_i = 15; \sum y_i = 206; \sum x_i^2 = 55; \sum y_i^2 = 9170; \sum x_i y_i = 701.$$

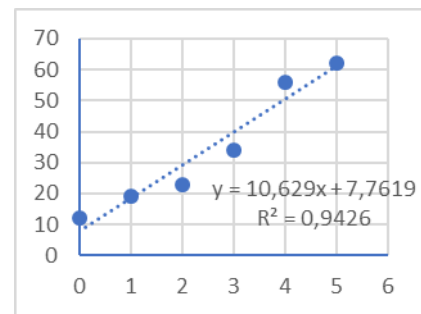
a) $\bar{x} = 2,5$; $\sigma_x = 1,70782$; $\bar{y} = 34,3333$; $\sigma_y = 18,6964$.

b) $\sigma_{xy} = 31$.

c) $r = 0,97086$. Su valor está próximo a 1, lo que indica que la correlación lineal es directa y muy fuerte.

Como el coeficiente de determinación, $r^2 = 0,9426$, se deduce que el 94,26 % de las variaciones de la variable Y se explican (están determinadas) por las variaciones de la variable X.

d) Con la calculadora se obtiene: $y = 10,6285x + 7,7619$.



Nota: El dibujo se ha hecho con Excel.

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

$$y = -2x + 8; \quad y = 0,8x + 2, \quad y = 1,5x + 1.$$

Solución:

Se sabe que el signo de la pendiente de la recta de regresión es el mismo que el de la correlación. Por tanto, como r es positivo, hay que descartar la primera recta.

Por otra parte, la recta de regresión pasa por el centro medio de la distribución, el punto $(\bar{x}, \bar{y}) = (2, 4)$

4) La única recta que lo contiene es la tercera.

En consecuencia, la recta de regresión pedida es $y = 1,5x + 1$.

26. Halla el centro medio de una distribución sabiendo que sus rectas de regresión valen:

De Y sobre X: $y = 2x + 2$.

De X sobre Y: $x = 0,45y - 0,2$.

Solución:

Como las dos rectas pasan por el centro medio, este punto será la solución del sistema

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

Cuya solución es: $x = 7, y = 16$.

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

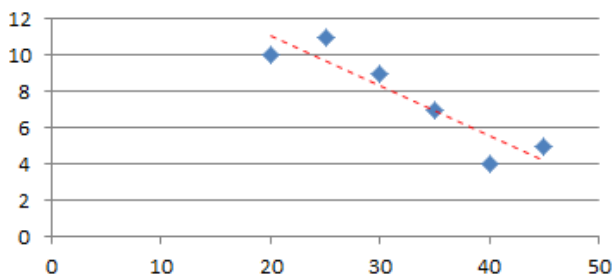
27. Una compañía de seguros de automóvil sospecha que el número de accidentes está en función de la edad del conductor. Para ello elige 100 personas de cada grupo de edad y contabiliza los accidentes totales del último año. Los datos fueron:

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

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

Solución:

a)



Correlación inversa, así lo indica la “línea de tendencia”.

b) Deben hacerse sumas, sumas de cuadrados y sumas de productos; para ello resulta eficaz la siguiente tabla:

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

Con esto:

$$\bar{x} = \frac{195}{6} = 32,5; \quad \bar{y} = \frac{46}{6} = 7,667.$$

$$\sigma_x = \sqrt{\frac{6675}{6} - 32,5^2} = 8,539; \quad \sigma_y = \sqrt{\frac{392}{6} - 7,667^2} = 2,56; \quad \sigma_{xy} = \frac{1375}{6} - 32,5 \cdot 7,669 = -20.$$

Por último, $r = \frac{-20}{8,539 \cdot 2,56} = -0,915.$

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

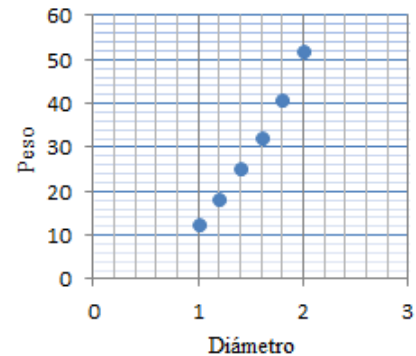
28. Se está experimentando la resistencia a la rotura de una determinada fibra textil. Para ello se ha medido el diámetro de la fibra y el peso que soporta hasta la rotura, obteniéndose los siguientes datos:

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

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

Solución:

- a) Dada la diferencia de rango entre las variables conviene cambiar la escala en los ejes, como se hace en la figura.



Claramente se adivina una correlación lineal.

Nota: Posiblemente un ajuste exponencial también sería bueno.

- b) Positiva y muy fuerte.

29. Con los datos del problema anterior, halla:

- a) La recta de regresión de Y sobre X, y determina la resistencia a la rotura de una fibra de 2,5 mm de diámetro.
- b) La recta de regresión de X sobre Y, y determina el diámetro mínimo de una fibra para que soporte más de 60 kg.

Solución:

Utilizando la calculadora:

a) $y = 39,0714x - 28,5238$.

Para $x = 2,5$ mm, $y = 69,1547$ kg.

b) $x = 0,02522y + 0,74112$.

Para $y = 60$ kg, $x = 2,25$ mm.

30. La siguiente tabla indica las horas de asistencia a un curso de informática y las notas obtenidas por seis alumnos:

Horas de asistencia (X)	2	1	7	5	4	3
Notas (Y)	2	1	10	6	5	3

- a) Representa la nube de puntos.
- b) Halla el coeficiente de correlación entre X e Y, e interprétalo.
- c) Halla la recta de regresión; y represéntala.
- d) Si una persona asistiera seis horas al curso, ¿qué nota cabe esperar para ella?

Solución:

- a) La nube de puntos es la representada en la figura.

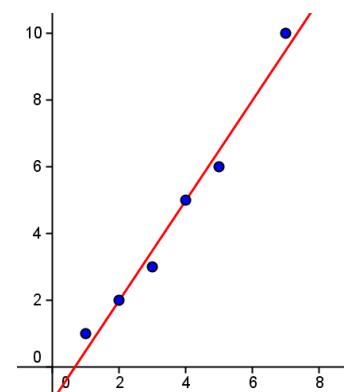
b) Utilizando la calculadora se tienen los siguientes resultados:

$$\bar{x} = 3,666... \quad \sum x_i = 22 \quad \sum x_i^2 = 104 \quad \sigma_x = 1,97203$$

$$\bar{y} = 4,5 \quad \sum y_i = 27 \quad \sum y_i^2 = 175 \quad \sigma_y = 2,98607$$

$r = 0,99061 \rightarrow$ Es una correlación directa y muy fuerte, lo que significa que las deducciones que puedan hacerse de las notas a partir de las horas de asistencia son bastante fiables.

- c) La recta de regresión es: $y = 1,5x - 1$. Es la adjunta.



→ Cálculo de los parámetros anteriores “a mano”.

X (x _i)	Y (y _i)	x _i ²	y _i ²	x _i ·y _i
2	2	4	4	4
1	1	1	1	1
7	10	49	100	70
5	6	25	36	30
4	5	16	25	20
3	3	9	9	9
$\sum x_i = 22$	$\sum y_i = 27$	$\sum x_i^2 = 104$	$\sum y_i^2 = 175$	$\sum x_i y_i = 134$
$\bar{x} = \frac{22}{6} = 3,666$	$\bar{y} = \frac{27}{6} = 4,5$	$\sigma_x = 1,972$	$\sigma_y = 2,986$	$\sigma_{xy} = 5,833$

$$\sigma_x = \sqrt{\frac{104}{6} - 3,666^2} = 1,972; \quad \sigma_y = \sqrt{\frac{175}{6} - 4,5^2} = 2,986; \quad \sigma_{xy} = \frac{134}{6} - \frac{22}{6} \cdot 4,5 = 5,833.$$

Recta de regresión: $y - \bar{y} = \frac{\sigma_{xy}}{\sigma_x^2} (x - \bar{x}) \Rightarrow y - 4,5 = \frac{5,833}{1,972^2} \left(x - \frac{22}{6}\right) \Rightarrow y = 1,5x - 1.$

Nota: En este caso, la manera más rápida de hallar el coeficiente de correlación y la recta de regresión es con **GeoGebra**, tecleando:

CoficienteCorrelación({(2,2),(1,1),(7,10),(5,6),(4,5),(3,3)}) → se obtiene r = 0,9906.
 AjusteLineal({(2,2),(1,1), (7,10),(5,6),(4,5),(3,3)}) → se obtiene y = 1.5x - 1.

d) Para una persona que asista 6 horas al curso puede estimarse una nota de:
 $y = 1,5 \cdot 6 - 1 = 8.$

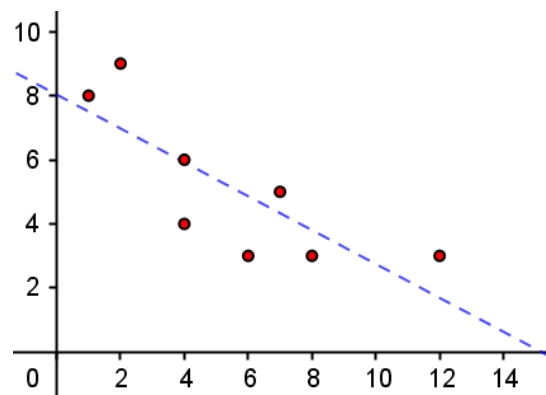
31. El número de faltas de asistencia a clase en los dos últimos meses, en la asignatura de Matemáticas, de ocho estudiantes de 1º de bachillerato, y su calificación final en dicha asignatura han sido:

Faltas (X)	1	2	4	4	6	7	8	12
Calificaciones (Y)	8	9	6	4	3	5	3	3

- a) Dibuja la nube de puntos asociada. ¿Qué tipo de correlación se observa entre las variables estudiadas?
- b) Calcula, indicando todos los pasos intermedios, el coeficiente de correlación y la recta de regresión de Y sobre X.

Solución:

a) A partir de la lectura de los valores de la tabla se observa una correlación negativa (la calificación de Matemáticas tiende a ser menor cuando aumentan las faltas de asistencia a clase)
 Representando los pares de valores: (1, 8), (2, 9), ... (12, 3), se observa una nube de puntos alargada y con tendencia decreciente; esto sugiere que la correlación es inversa y fuerte.



b) De acuerdo con las fórmulas de los parámetros hay que hacer sumas, sumas de cuadrados y sumas de productos; para ello resulta eficaz la siguiente tabla:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	8	1	64	8
2	9	4	81	18
4	6	16	36	24
4	4	16	16	16
6	3	36	9	18
7	5	49	25	35
8	3	64	9	24
12	3	144	9	36
$\sum x_i = 44$	$\sum y_i = 41$	$\sum x_i^2 = 330$	$\sum y_i^2 = 249$	$\sum x_i y_i = 179$

Con esto:

$$\bar{x} = \frac{44}{8} = 5,5; \quad \bar{y} = \frac{41}{8} = 5,125;$$

$$\sigma_x = \sqrt{\frac{330}{8} - 5,5^2} = 3,3166; \quad \sigma_y = \sqrt{\frac{249}{8} - 5,125^2} = 2,2044; \quad \sigma_{xy} = \frac{179}{8} - 5,5 \cdot 5,125 = -5,8125.$$

Por último, $r = \frac{-5,812}{3,3166 \cdot 2,2044} = -0,795$.

La correlación es inversa y fuerte: la nota de Matemáticas está condicionada por las faltas de asistencia a clase. (Como el coeficiente de determinación, $r^2 = 0,632$, puede asumirse que el 63,2 % de las variaciones en la nota se explican por la asistencia a clase).

La recta de regresión es

$$y - 5,125 = \frac{-5,8125}{3,3166^2} (x - 5,5) \Leftrightarrow y = 0,5284x + 8,0329.$$

(Es la que se ha trazado anteriormente).

c) Si $x = 5$, $y = -0,5284 \cdot 5 + 8,0329 = 5,3909$ puntos \rightarrow 5,4 puntos.

Nota: Comprueba los resultados utilizando GeoGebra.