Escola Secundária de Jácome Ratton

Dados Bidimensionais

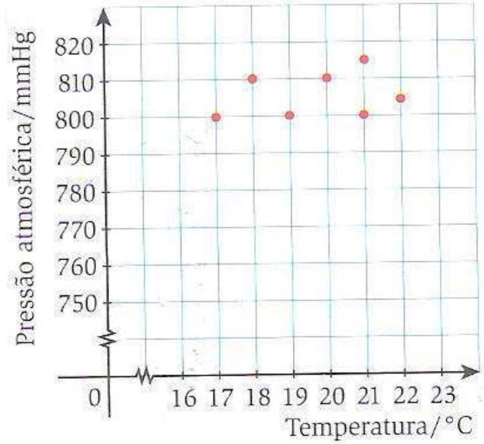
Ano Lectivo 2010/2011

Matemática Aplicada às Ciências Sociais

Dados bidimensionais ou bivariados – são dados obtidos de pares de variáveis.

A amostra de dados bivariados pode representar-se por $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.

Diagrama de dispersão ou nuvem de pontos representativa da distribuição estatística – é uma representação gráfica para os dados bivariados, em que cada par de dados (x_i, y_i) é representado por um ponto de coordenadas (x_i, y_i), num sistema de eixos coordenados.



Ponto médio ou Centro de gravidade do diagrama de dispersão – é o ponto cuja abcissa é a média dos valores da primeira variável e cuja ordenada é a média dos valores da segunda variável, ou seja, é o ponto $(\bar{x}, \bar{y})$.

Exemplo: Um professor tem uma turma com 18 alunos e pensa que os seus alunos que têm boas classificações em Matemática também são bons alunos em Física. No final do 1ºperíodo verificou que:

Aluno	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
Classificação em Matemática (x)	6	8	8	11	12	13	13	14	14	14	14	15	15	16	16	17	17	18
Classificação em Física (y)	5	7	10	10	12	12	10	15	15	13	12	13	16	17	12	15	18	18

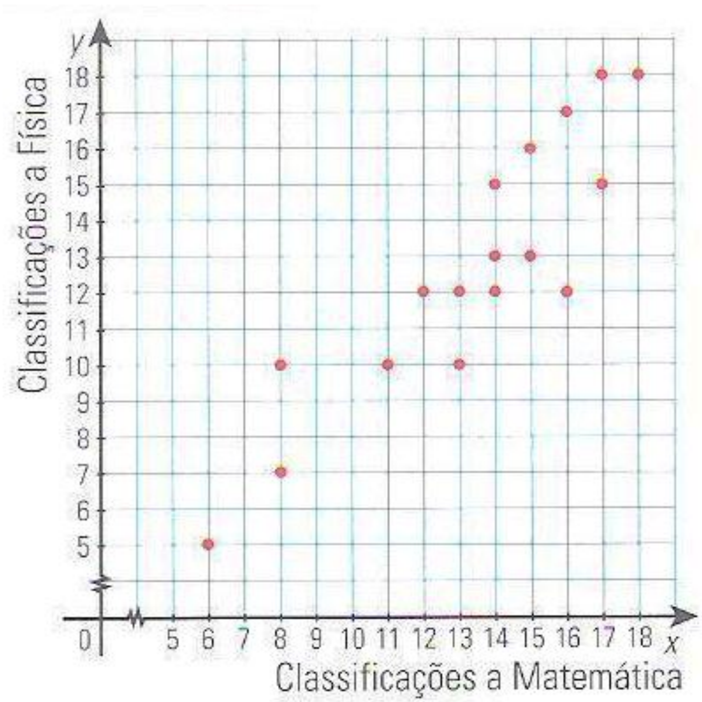
a) Represente os dados num diagrama de dispersão.

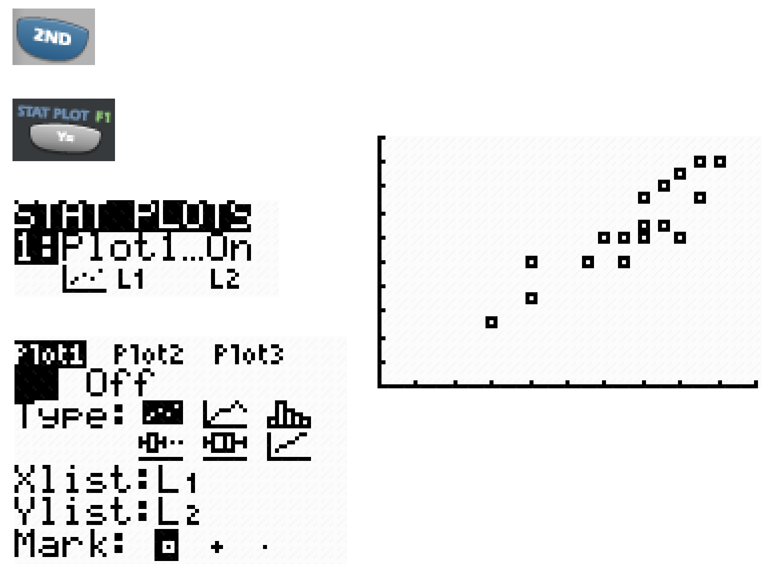
b) Explique se a ideia do professor faz ou não sentido.

c) Determine as coordenadas do centro de gravidade e assinale-o no diagrama de dispersão.

Aluno	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
Classificação em Matemática (x)	6	8	8	11	12	13	13	14	14	14	14	15	15	16	16	17	17	18
Classificação em Física (y)	5	7	10	10	12	12	10	15	15	13	12	13	16	17	12	15	18	18

a) Represente os dados num diagrama de dispersão.





b) Explique se a ideia do professor faz ou não sentido.

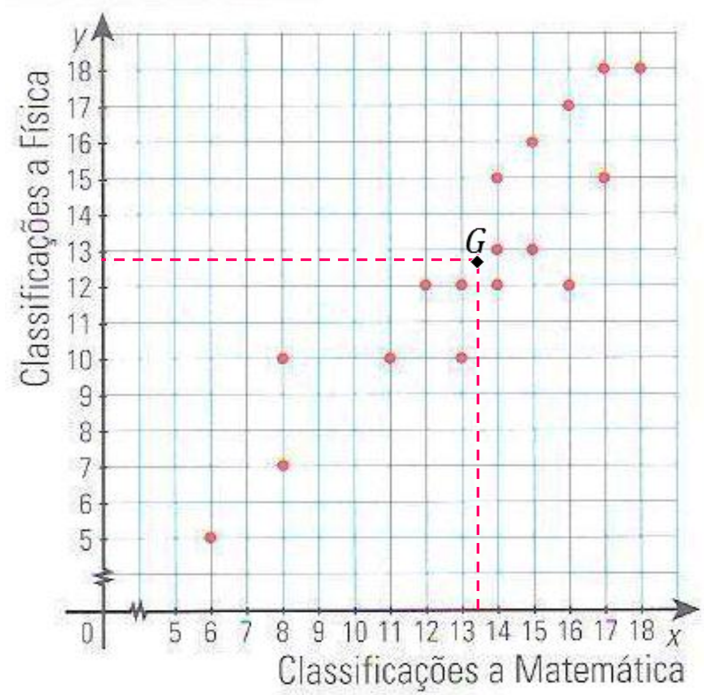
Sim, a ideia do professor faz sentido porque existe uma relação estatística acentuada entre as duas variáveis. .

Aluno	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
Classificação em Matemática (x)	6	8	8	11	12	13	13	14	14	14	14	15	15	16	16	17	17	18
Classificação em Física (y)	5	7	10	10	12	12	10	15	15	13	12	13	16	17	12	15	18	18

c) Determine as coordenadas do centro de gravidade e assinale-o no diagrama de dispersão.

CALC
2:2-Var Stats

$$\left.\begin{array}{l} \bar{x} \cong 13{,}4 \\ \bar{y} \cong 12{,}8 \end{array}\right\} \quad G(13{,}4; 12{,}8)$$



Análise gráfica de dados bidimensionais

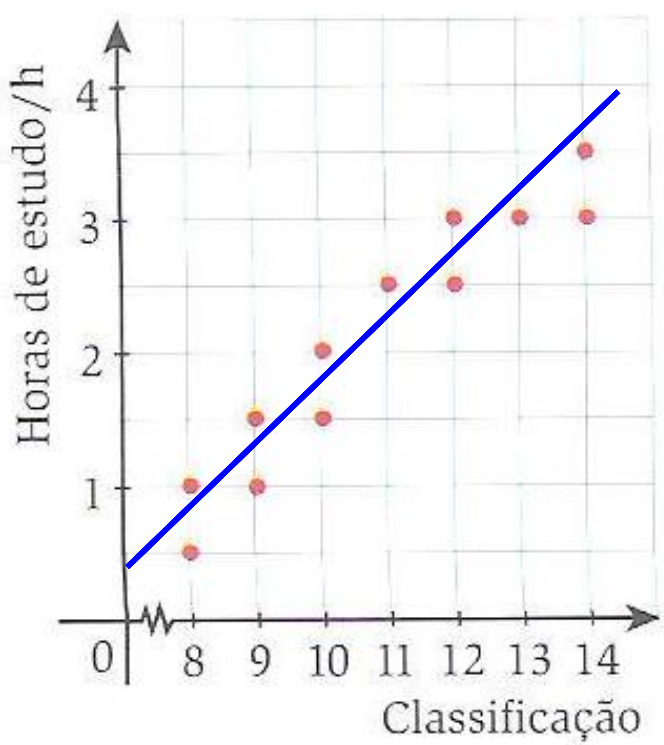
A tabela mostra a classificação num teste de estatística de 12 estudantes, as horas dedicadas à sua preparação, o número de horas gastas a ver televisão no fim-de-semana que precedeu o teste e a altura de cada estudante.

Classificação no teste	8	9	10	8	10	11	12	12	13	14	14	9
Horas de estudo para o teste	1	1,5	2	0,5	1,5	2,5	3	2,5	3	3,5	3	1
Horas a ver televisão	5	6	3	4	5	3	2	3	1	1	2	5
Altura	155	165	155	170	180	170	165	175	160	165	175	175

Faça um estudo gráfico para a classificação no teste e cada uma das outras variáveis. Comente cada um dos gráficos.

Classificação no teste	8	9	10	8	10	11	12	12	13	14	14	9
Horas de estudo para o teste	1	1,5	2	0,5	1,5	2,5	3	2,5	3	3,5	3	1
Horas a ver televisão	5	6	3	4	5	3	2	3	1	1	2	5
Altura	155	165	155	170	180	170	165	175	160	165	175	175

Classificação no teste e Horas de estudo para o teste



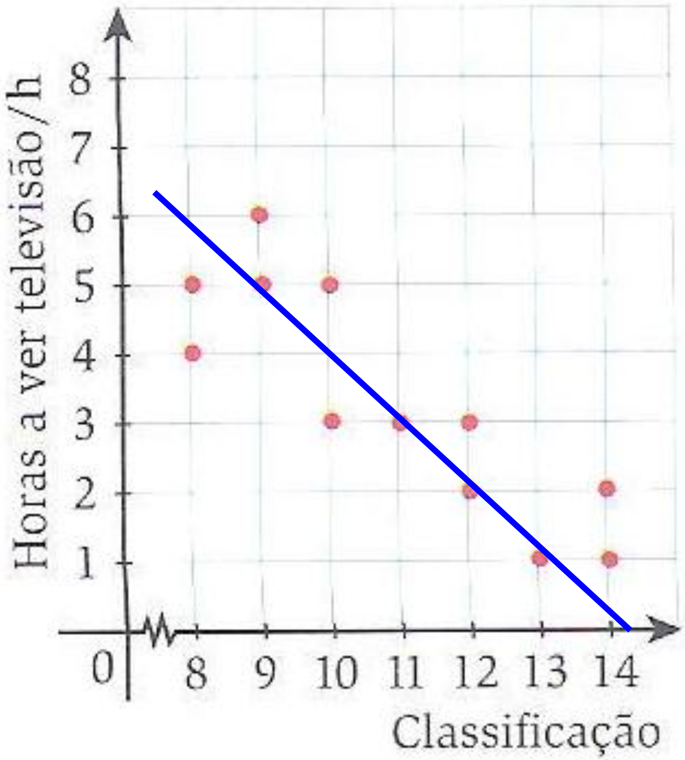
A recta que "melhor se aproxima" de todos os pontos do gráfico tem declive positivo.

Variáveis positivamente associadas ou associação positiva entre as variáveis – aos maiores valores de uma variável correspondem, geralmente, os maiores valores da outra.

Quem passa mais tempo a estudar obtém, geralmente, melhor classificação no teste.

Classificação no teste	8	9	10	8	10	11	12	12	13	14	14	9
Horas de estudo para o teste	1	1,5	2	0,5	1,5	2,5	3	2,5	3	3,5	3	1
Horas a ver televisão	5	6	3	4	5	3	2	3	1	1	2	5
Altura	155	165	155	170	180	170	165	175	160	165	175	175

Classificação no teste e Horas a ver televisão



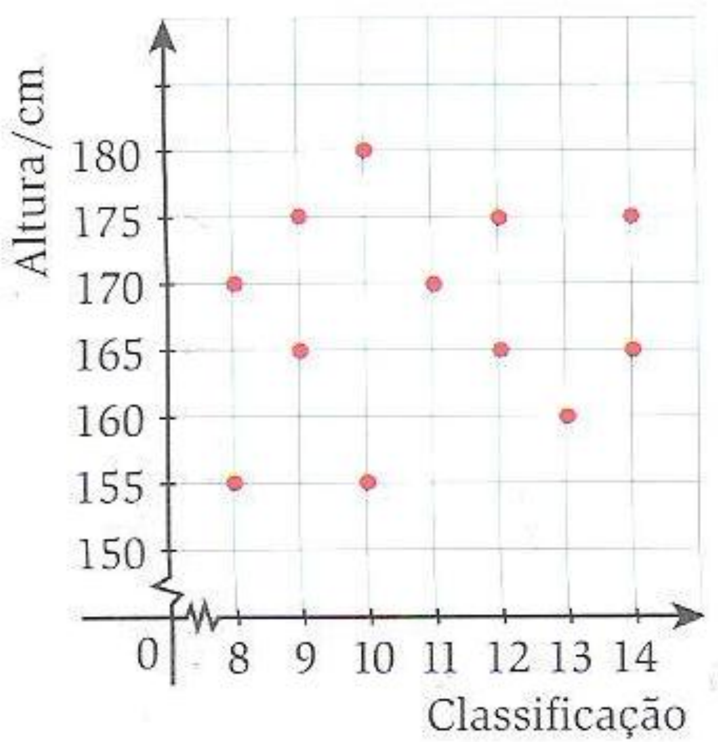
A recta que "melhor se aproxima" de todos os pontos do gráfico tem declive negativo.

Variáveis negativamente associadas ou associação negativa entre as variáveis – aos maiores valores de uma variável correspondem, geralmente, os menores valores da outra.

Quem passa mais tempo a ver televisão, geralmente, obtém pior classificação no teste.

Classificação no teste	8	9	10	8	10	11	12	12	13	14	14	9
Horas de estudo para o teste	1	1,5	2	0,5	1,5	2,5	3	2,5	3	3,5	3	1
Horas a ver televisão	5	6	3	4	5	3	2	3	1	1	2	5
Altura	155	165	155	170	180	170	165	175	160	165	175	175

Classificação no teste e Altura

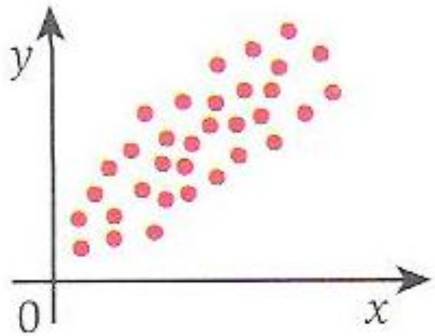


A nuvem de pontos encontra-se bastante dispersa assim **não existe uma associação clara entre as duas variáveis**.

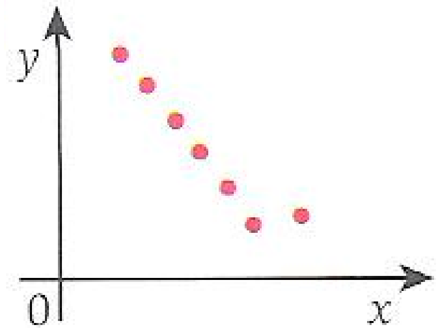
A altura do estudante não tem qualquer influência sobre a sua classificação no teste.

Exercicio: Observe os seguintes diagramas de dispersão:

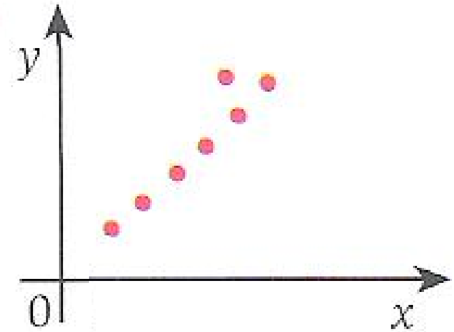
(A)



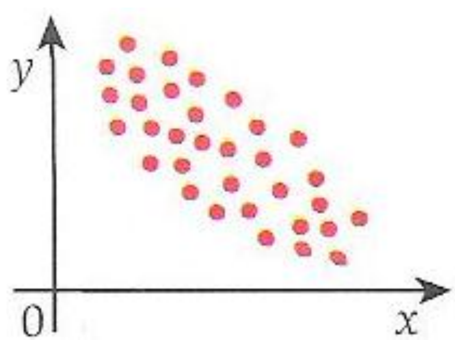
(B)



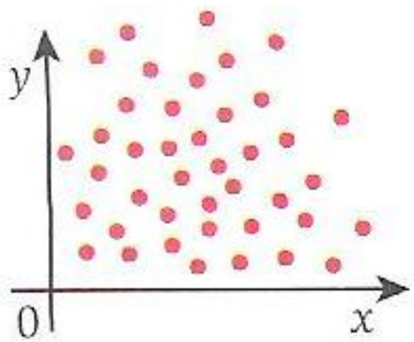
(C)



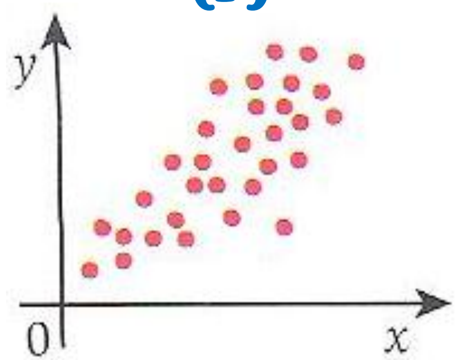
(D)



(E)



(F)



Indique, pela letra correspondente, aqueles em que se observa:

- uma associação positiva; **A, C e F**
- uma associação negativa; **B e D**
- não há uma associação clara entre as duas variáveis. **E**

Coeficiente de Correlação Linear

$$r = \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}\sqrt{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}}$$

x_i – Valores das observações de uma das variáveis;

y_i – Valores das observações correspondentes da outra variável;

$\bar{x}$ – média das observações de x_i;

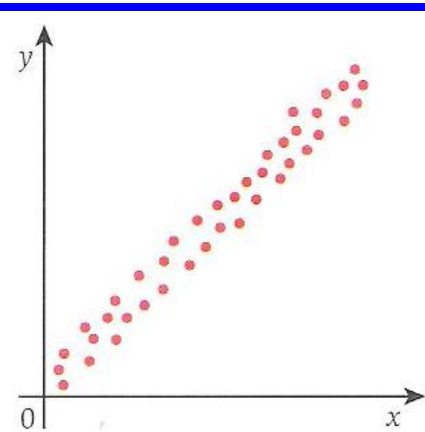
$\bar{y}$ – média das observações de y_i.

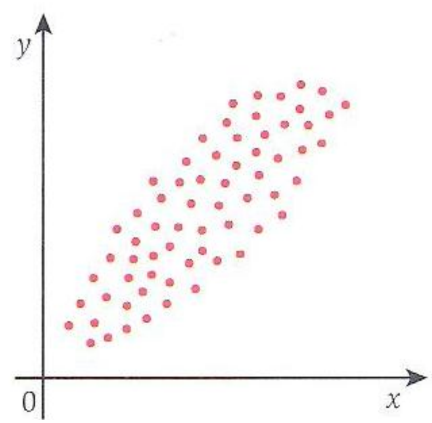
Exemplo:

$r > 0$

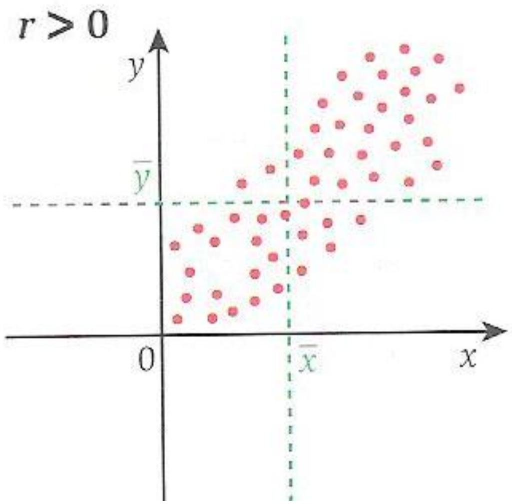
Maior grau de associação

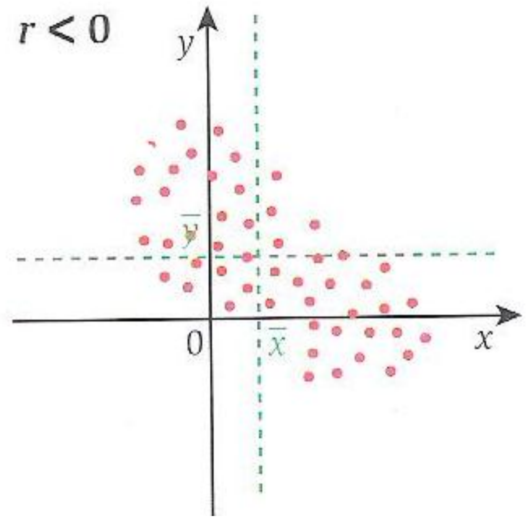
Maior coeficiente de correlação

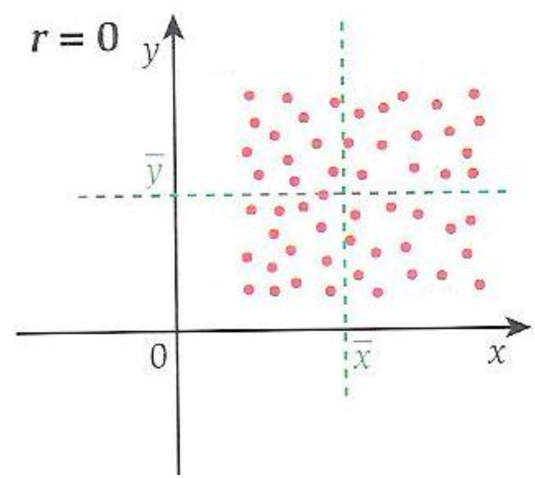




Interpretação geométrica do coeficiente de correlação







Propriedades do Coeficiente de Correlação Linear (r)

- O valor do coeficiente de correlação varia entre -1 e 1.

- Se **r** é positivo, a correlação é positiva (ou a associação entre as variáveis é positiva) e a dependência entre as variáveis é tanto mais forte quanto mais próximo estiver de 1.

- Se **r** é negativo, a correlação é negativa (ou a associação entre as variáveis é negativa) e a dependência entre as variáveis é tanto mais forte quanto mais próximo estiver de -1.

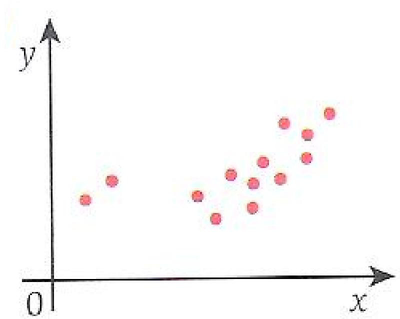
- Os valores -1 e 1 são atingidos quando à recta de regressão pertencem todos os pontos da nuvem.

- Quando r = 0 a correlação é nula e não existe associação entre as variáveis.

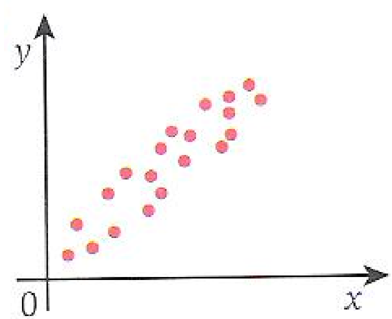
- A correlação não é afectada por uma mudança de unidades das variáveis.

Exercicio: Observe os seguintes diagramas de dispersão:

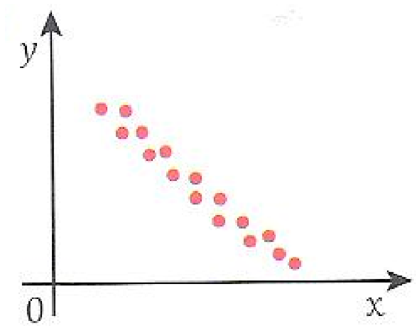
(A)



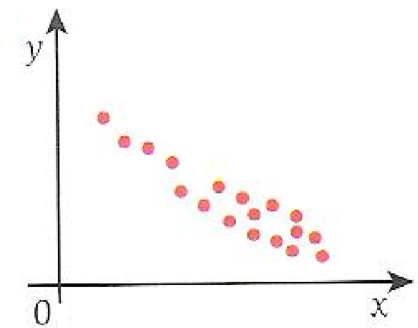
(B)



(C)



(D)

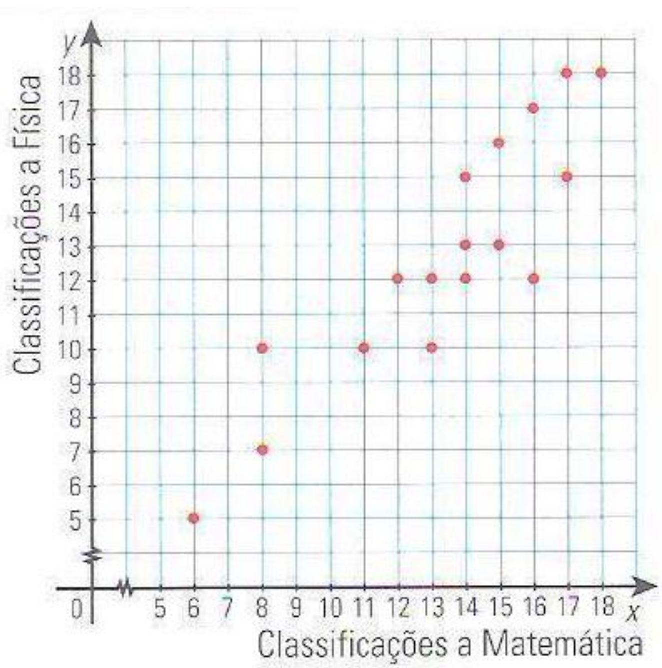


A cada um dos diagramas faça corresponder o valor de *r*.

r	**- 0,01**	**0,02**	**- 0,98**	**0,92**	**0,59**	**- 0,93**
Diagrama			**(C)**	**(B)**	**(A)**	**(D)**

Determinar o coeficiente de correlação com recurso à calculadora

Aluno	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
Classificação em Matemática (x)	6	8	8	11	12	13	13	14	14	14	14	15	15	16	16	17	17	18
Classificação em Física (y)	5	7	10	10	12	12	10	15	15	13	12	13	16	17	12	15	18	18



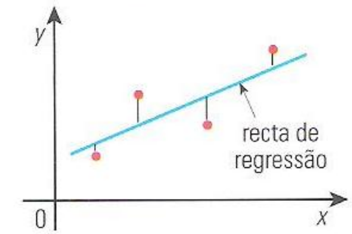
CALC

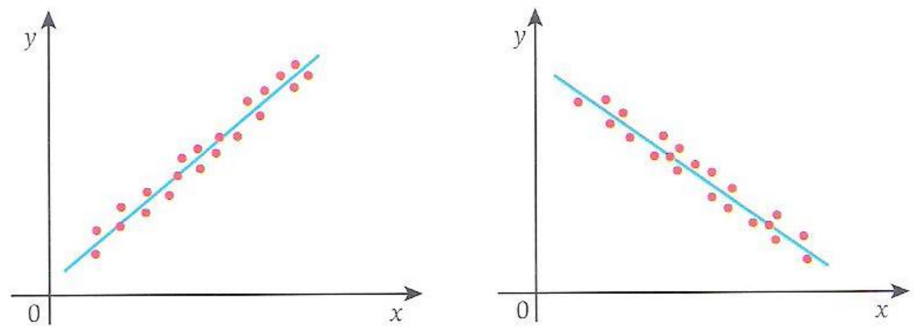
4:LinReg(ax+b)

r=.8930387401

Recta de Regressão

Quando duas variáveis estão fortemente correlacionadas os pontos do diagrama de dispersão colocam-se em torno de uma recta.

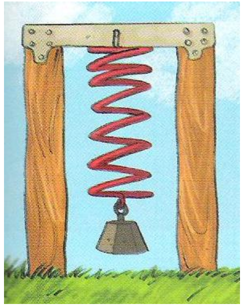




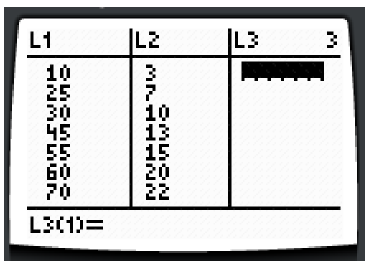
A recta que torna mínima a soma dos quadrados dos desvios dos pontos em relação à recta designa-se por **recta de regressão** e pode ser definida por uma equação do tipo: $y = ax+b$.

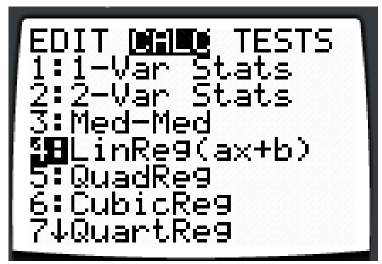
A recta de regressão passa no ponto $(\bar{x}, \bar{y})$ e o sinal do seu declive é o sinal do coeficiente de correlação.

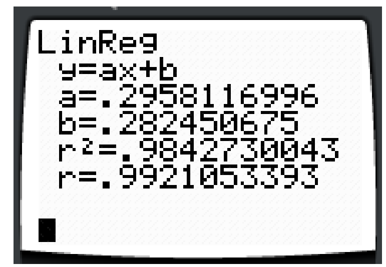
Exemplo: Suspenderam-se objectos de diferentes massas numa mola deformando-a e registaram-se os correspondentes alongamentos da mola, como se mostra na tabela:

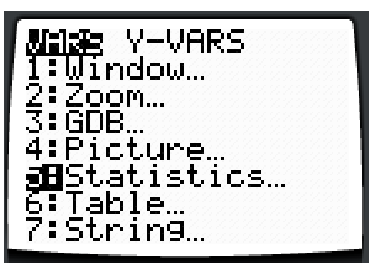


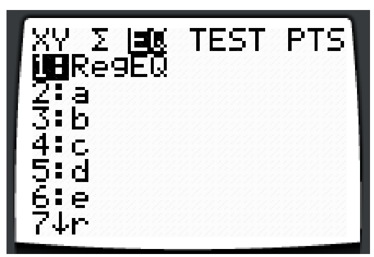
x_i (massa em g)	10	25	30	45	55	60	70	75	85	100
y_i (alongamento em mm)	3	7	10	13	15	20	22	23	25	29

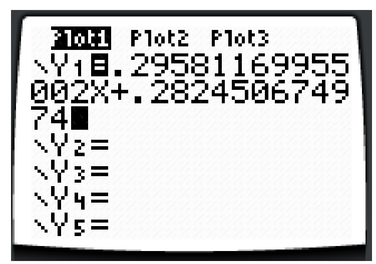


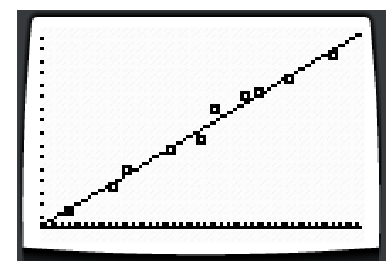






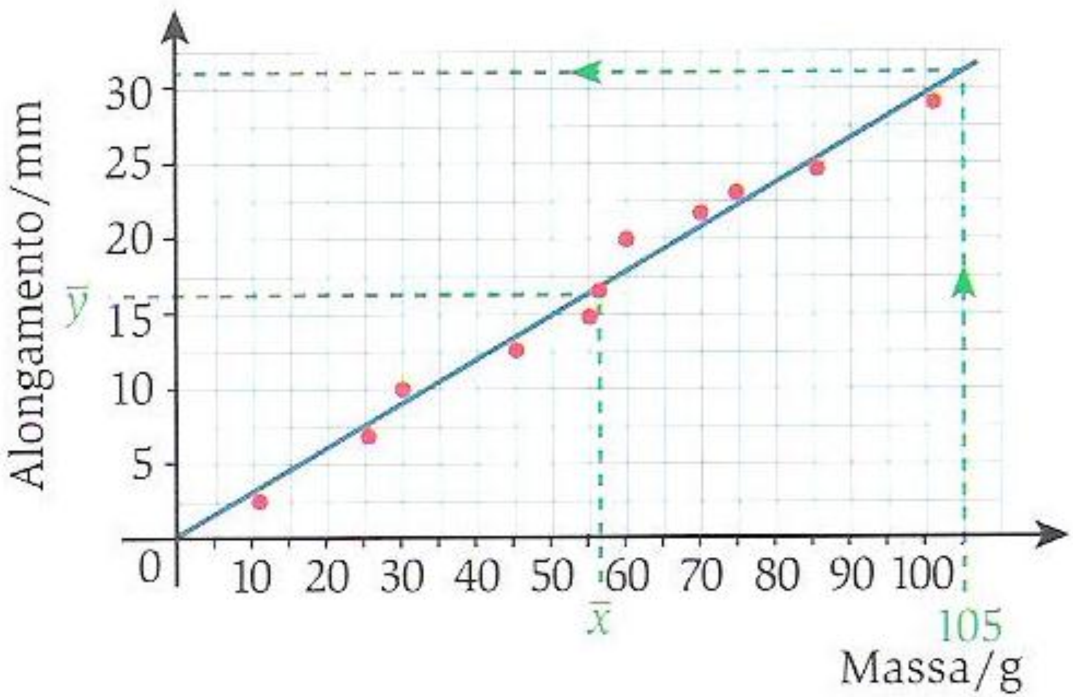






x_i (massa em g)	10	25	30	45	55	60	70	75	85	100
y_i (alongamento em mm)	3	7	10	13	15	20	22	23	25	29

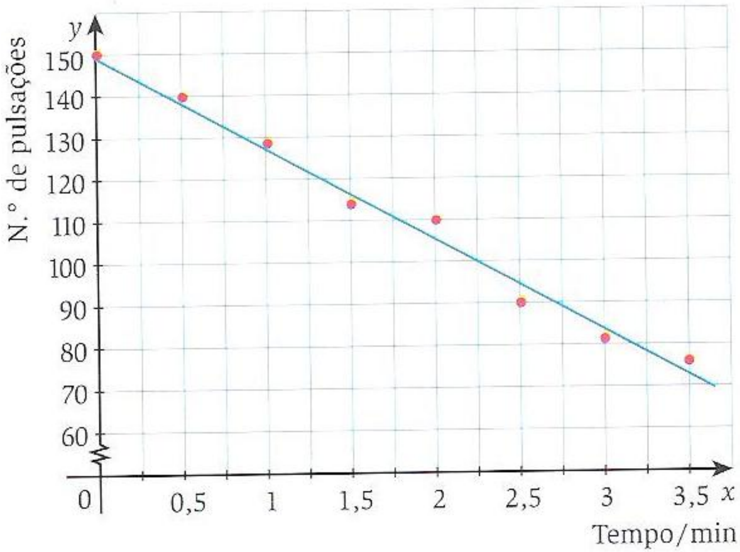
$$\bar{x} = \frac{10+25+30+45+55+60+70+75+85+100}{10} = 55{,}5\,; \quad \bar{y} = \frac{3+7+10+13+15+20+22+23+25+29}{10} = 16{,}7\,.$$

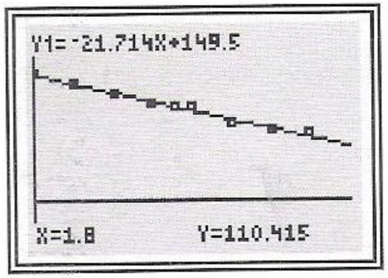


Recta de regressão para fazer estimativas

O Nuno pratica atletismo. Depois de terminada uma prova anotou as suas pulsações.

Tempo decorrido/min	0	0,5	1	1,5	2	2,5	3,0	3,5
N.° de pulsações	150	140	128	114	110	90	82	78





$$y = -21{,}714x + 149{,}5$$

Quantas pulsações teria o Nuno 1,8 min depois de terminada a prova?

$x = 1{,}8 \qquad y = -21{,}714 \times 1{,}8 + 149{,}5 \qquad \langle = \rangle\ y \cong 110{,}41$

1,8 min depois de terminar a prova o Nuno teria 110 pulsações.

Quantas pulsações teria o Nuno 30 min depois de terminada a prova?

Não faz sentido calcular y(30) porque 30 min depois de terminar a prova o Nuno terá a pulsação normal.

Exercicio: O comprimento y, em milímetros, de uma peça de metal foi medido a várias temperaturas x, em graus Celsius, obtendo-se os seguintes resultados:

Temperatura (°C)	60	65	70	75	80	85	90	95
Comprimento (mm)	50,2	50,5	50,5	51,0	52,1	53,0	54,2	55,1

a) Desenhe o diagrama de dispersão.

b) Obtenha, com a calculadora, a equação da recta de regressão e represente-a no gráfico.

c) Use a recta de regressão para estimar o comprimento da peça quando a temperatura é de 72ºC e quando a temperatura é de 130ºC.