

III – CORRELAÇÃO E REGRESSÃO

Profa. Dra. Nanci de Oliveira
nanci.oliveira@fatec.sp.gov.br

CORRELAÇÃO E REGRESSÃO

- ▶ São duas *técnicas* que envolvem uma forma de *estimação*.
- ▶ Vamos estudar situações que envolvem apenas duas variáveis.

CORRELAÇÃO LINEAR

- ▶ O termo *correlação* significa, literalmente, *correlacionamento*, pois indica até que ponto os valores de uma variável estão relacionados com os da outra.

Coeficiente r de Pearson

- ▶ O grau de relacionamento entre duas variáveis contínuas é sintetizado pelo **coeficiente de correlação** conhecido como “**r**” de Pearson (1857–1936), que desenvolveu a técnica:

$$r = \frac{[n(\sum xy) - (\sum x \sum y)]}{\sqrt{\{[n(\sum x^2) - (\sum x)^2][n \sum y^2 - (\sum y)^2]\}}}$$

onde \underline{n} = número de observações, \underline{x} e \underline{y} são as variáveis

O valor de r varia de -1 a +1, ou seja, $-1 \leq r \leq 1$

Classificação da correlação ou Grau de correlação

A correlação pode ser classificada em:

Forte: $0,6 \leq |r| \leq 1$

Fraca: $0,3 \leq |r| < 0,6$

Muito fraca: $0 \leq |r| < 0,3$

- Se $r = 1$ a correlação é *positiva perfeita*.
- Se $r = 0$ a correlação é *nula* (não tem correlação entre as variáveis).
- Se $r = -1$ a correlação é *negativa perfeita*.

REGRESSÃO LINEAR

- ▶ A regressão linear estabelece uma equação linear (equação de regressão) que descreva o relacionamento entre duas variáveis.
- ▶ A finalidade de uma equação de regressão é estimar valores de uma variável com base em valores conhecidos da outra.

Equação linear

- ▶ Tem a forma $y = ax + b$, onde a e b são valores que se determinam com base nos dados amostrais.

a = coeficiente angular da reta

b = coeficiente linear da reta

O método dos mínimos quadrados

- ▶ É método mais usado para ajustar um conjunto de pontos a uma reta.
- ▶ Nesse método, a soma dos desvios verticais dos pontos em relação à reta é igual a zero e a soma dos quadrados desses desvios é mínima.

O método dos mínimos quadrados

Este método consiste em resolver o sistema abaixo, em relação a a e b :

onde n é o número de observações.

$$\begin{cases} a \sum X^2 + b \sum X = \sum XY \\ a \sum X + nb = \sum Y \end{cases}$$

As equações acima podem ser removidas algebricamente em relação a a e b , resultando em duas fórmulas:

$$a = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2}$$

$$b = \frac{\sum Y - a \sum X}{n}$$

Erro padrão ou desvio padrão da estimativa

- ▶ A **precisão** das estimativas de regressão depende da dispersão (desvio padrão):
quanto maior a dispersão, menor a precisão das estimativas.
- ▶ Essa dispersão é dada pela seguinte fórmula:

$$S_e = \sqrt{\frac{\sum Y^2 - b \sum Y - a \sum XY}{n - 2}}$$

EXEMPLO

A tabela abaixo apresenta as notas de Matemática e de Estatística de 10 alunos de uma turma de um colégio.

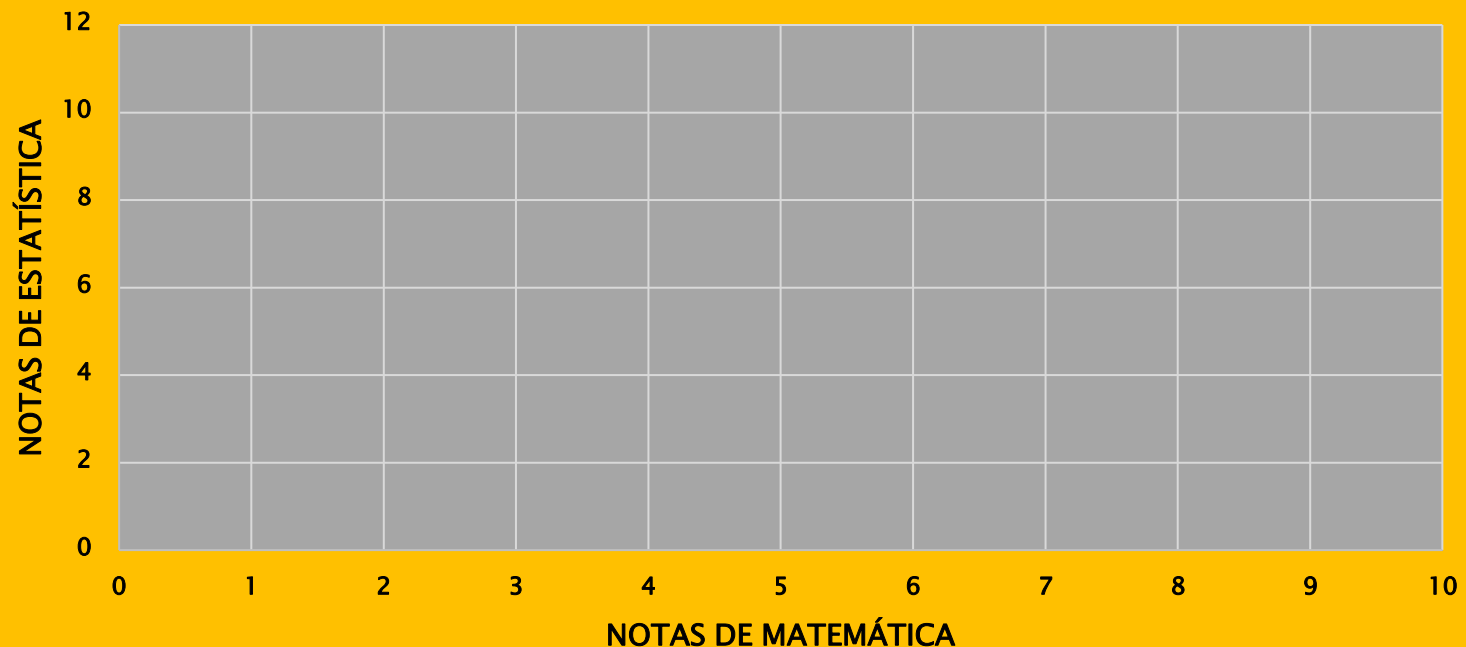
- 1) Faça o Diagrama de Dispersão.
- 2) Calcule o coeficiente de correlação linear (r de Pearson) entre as duas variáveis.
- 3) Classifique ou dê o grau de correlação linear.
- 4) Se a correlação for forte, encontre a equação da reta de regressão linear.
- 5) Se um aluno dessa turma tivesse nota 1 em Matemática, qual seria sua nota em Estatística?
- 6) Se um aluno dessa classe tivesse nota 4 em Matemática, qual seria sua nota de Estatística?

Alunos	Notas Mat. (X)	Notas Estat. (Y)
A	5	6
B	8	9
C	7	8
D	10	10
E	6	5
F	7	7
G	9	8
H	3	4
I	8	6
J	2	2
	$\Sigma X = 65$	$\Sigma Y = 65$

1) Solução

DIAGRAMA DE DISPERSÃO

Notas de Matemática e Estatística de 10 alunos de uma turma de um Colégio



2) Solução

Alunos	Notas Mat. (X)	Notas Estat. (Y)	XY	X ²	Y ²
A	5	6	30	25	36
B	8	9	72	64	81
C	7	8	56	49	64
D	10	10	100	100	100
E	6	5	30	36	25
F	7	7	49	49	49
G	9	8	72	81	64
H	3	4	12	9	16
I	8	6	48	64	36
J	2	2	4	4	4
	$\Sigma x = 65$	$\Sigma y = 65$	$\Sigma xy = 473$	$\Sigma x^2 = 481$	$\Sigma y^2 = 475$

$$r = \frac{[n(\sum xy) - (\sum x \sum y)]}{\sqrt{\{[n(\sum x^2) - (\sum x)^2][n \sum y^2 - (\sum y)^2]\}}}$$

n = 10 alunos

$\sum x = 65$	$\sum y = 65$	$\sum xy = 473$	$\sum x^2 = 481$	$\sum y^2 = 475$
---------------	---------------	-----------------	------------------	------------------

$$r = \frac{(10)(473) - (65)(65)}{\sqrt{[(10)(481) - (65)^2][(10)(475) - (65)^2]}} = \frac{4730 - 4225}{\sqrt{[4810 - 4225][4750 - 4225]}}$$

$$r = \frac{505}{\sqrt{585 \cdot 525}} = \frac{505}{\sqrt{307125}} = 0,9112$$

Logo, $r=0,9$.

3) Solução

- ▶ CLASSIFICAÇÃO: A correlação é positiva, forte.

4) Solução Algébrica

$\Sigma x = 65$	$\Sigma y = 65$	$\Sigma xy = 473$	$\Sigma x^2 = 481$	$\Sigma y^2 = 475$
-----------------	-----------------	-------------------	--------------------	--------------------

$$a = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2}$$

$$\Rightarrow a = \frac{(10) \cdot (473) - (65) \cdot (65)}{(10) \cdot (481) - (65)^2} \Rightarrow a = \frac{4730 - 4225}{4810 - 4225} = \frac{505}{585} = 0,8632$$

$$b = \frac{\sum Y - a \sum X}{n}$$

$$b = \frac{(65) - (0,8632) \cdot (65)}{10} \Rightarrow a = \frac{65 - 56,108}{10} = \frac{8,892}{10} = 0,8892$$

Logo, $y = ax + b \Rightarrow$

$$y = 0,8632x + 0,8892$$

5) Solução

- ▶ Para $x=1$, temos:

$$\begin{aligned}y &= 0,8632x + 0,8892 \\ \Rightarrow y &= 0,8632 \cdot (1) + 0,8892 \\ \Rightarrow y &= 0,8632 + 0,8892 \\ \Rightarrow y &= 1,7524\end{aligned}$$

- ▶ Portanto, se a nota de Matemática for 1, a nota de Estatística será 1,7524 ou 2 (arredondando para inteiro).

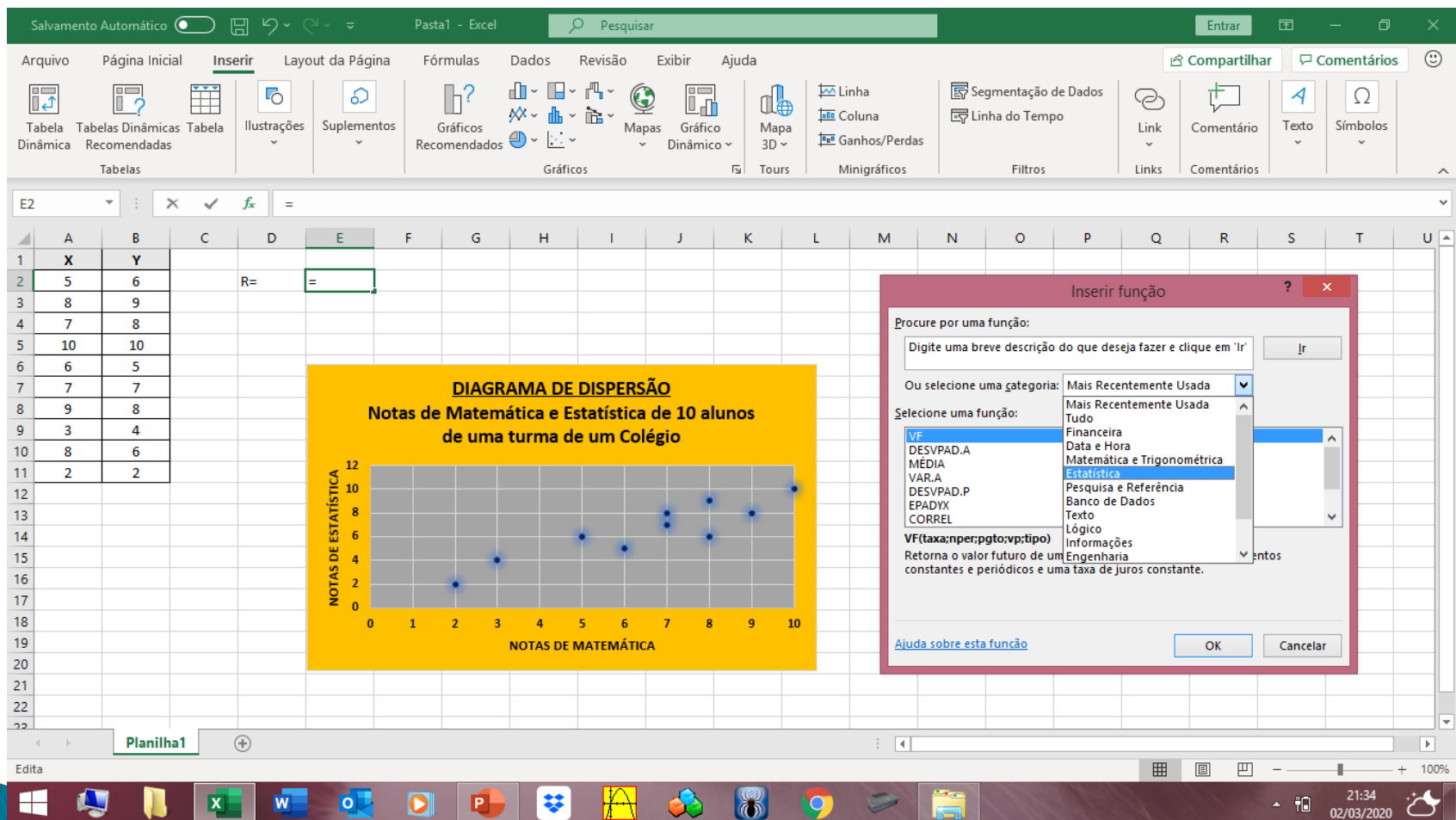
6) Solução

- ▶ Para $x=4$, temos:

$$\begin{aligned}y &= 0,8632x + 0,8892 \\ \Rightarrow y &= 0,8632 \cdot (4) + 0,8892 \\ \Rightarrow y &= 3,4528 + 0,8892 \\ \Rightarrow y &= 4,342\end{aligned}$$

- ▶ Portanto, se a nota de Matemática for 4, a nota de Estatística será 4,342 ou 4 (arredondando para inteiro).

RESOLVA O EXEMPLO ANTERIOR UTILIZANDO EXCEL



Salvamento Automático Pasta1 - Excel Pesquisar Entrar

Arquivo Página Inicial **Inserir** Layout da Página Fórmulas Dados Revisão Exibir Ajuda

Tabela Tabela Dinâmica Tabela Ilustrações Suplementos Gráficos Recomendados Mapas Gráfico Dinâmico Mapa 3D Tours Minigráficos Segmentação de Dados Linha do Tempo Link Comentário Texto Símbolos

E2 X Y R= =

	A	B
1	X	Y
2	5	6
3	8	9
4	7	8
5	10	10
6	6	5
7	7	7
8	9	8
9	3	4
10	8	6
11	2	2

DIAGRAMA DE DISPERSÃO
Notas de Matemática e Estatística de 10 alunos de uma turma de um Colégio

NOTAS DE ESTATÍSTICA

NOTAS DE MATEMÁTICA

Inserir função

Procure por uma função:
Digite uma breve descrição do que deseja fazer e clique em 'Ir'

Ou selecione uma categoria: Estatística

Selecione uma função:

- CONTAR.VAZIO
- CORREL**
- COVARIACÃO.P
- COVARIACÃO.S
- CRESCIMENTO
- CURT
- DESV.MÉDIO

CORREL(matriz1;matriz2)
Retorna o coeficiente de correlação entre dois conjuntos de dados.

[Ajuda sobre esta função](#)

OK Cancelar

Planilha1

21:37 02/03/2020

Argumentos da função

CORREL

Matriz1 = matriz

Matriz2 = matriz

=

Retorna o coeficiente de correlação entre dois conjuntos de dados.

Matriz1 é um intervalo de células de valores. Os valores devem ser números, nomes, matrizes ou referências que contenham números.

Resultado da fórmula =

[Ajuda sobre esta função](#)

OK Cancelar

Argumentos da função

CORREL

Matriz1 A2:A11 = {5;8;7;10;6;7;9;3;8;2}

Matriz2 B2:B11 = {6;9;8;10;5;7;8;4;6;2}

= 0,911242136

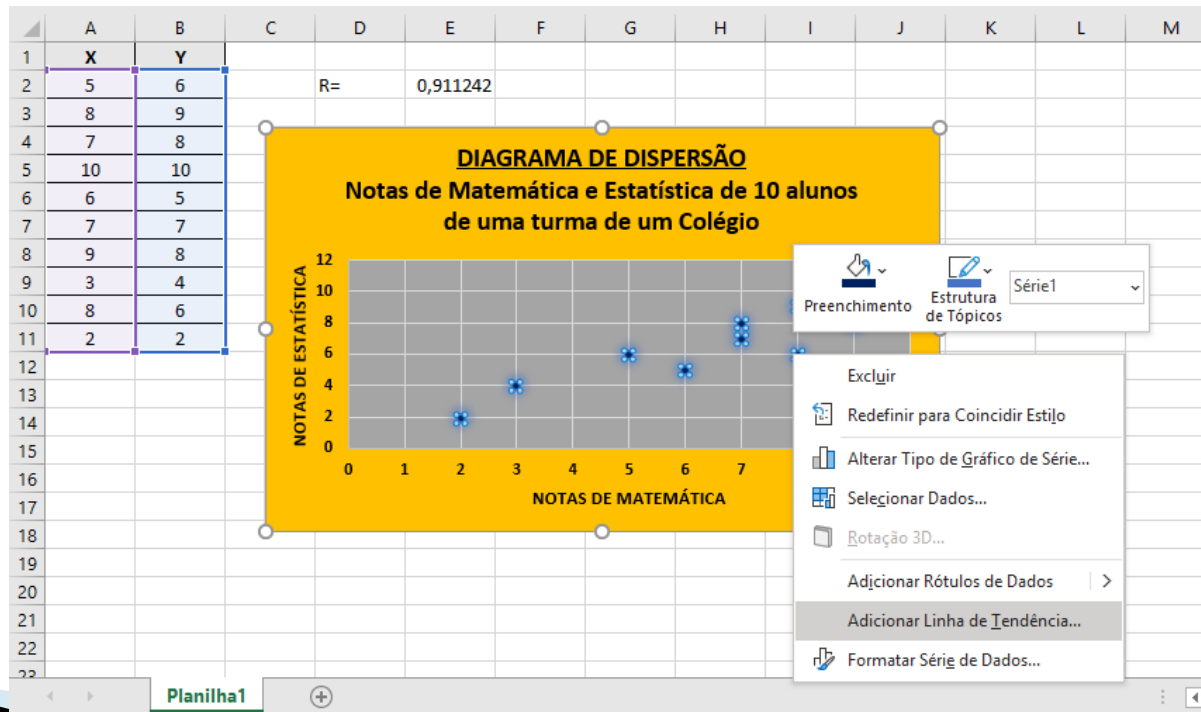
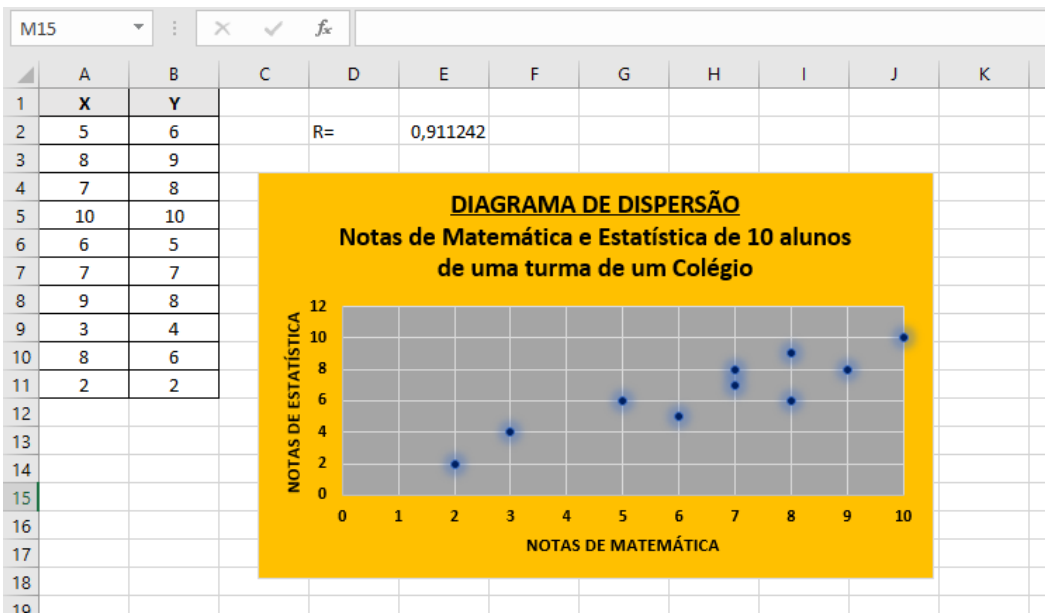
Retorna o coeficiente de correlação entre dois conjuntos de dados.

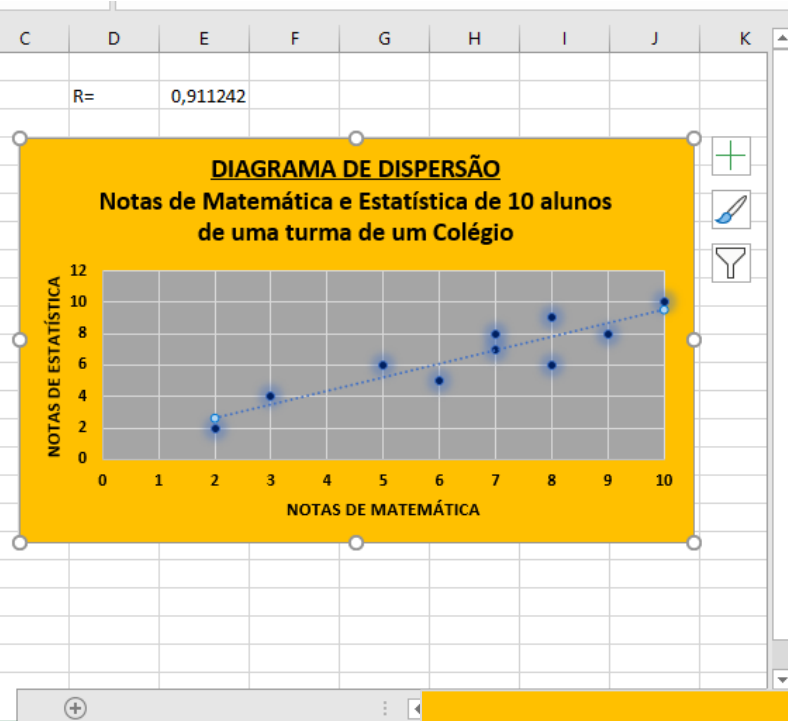
Matriz1 é um intervalo de células de valores. Os valores devem ser números, nomes, matrizes ou referências que contenham números.

Resultado da fórmula = 0,911242136

[Ajuda sobre esta função](#)

OK Cancelar





Formatar Linha de Tendência

Opções de Linha de Tendência



Opções de Linha de Tendência

- ☐ Exponencial
- ☒ Linear
- ☐ Logarítmica
- ☐ Polinomial
- ☐ Potência
- ☐ Média Móvel

Nome da Linha de Tendência

- ☒ Automático
- ☐ Personalizado

Previsão

Formatar Linha de Tendência

Opções de Linha de Tendência



Opções de Linha de Tendência

- ☐ Polinomial
- ☐ Potência
- ☐ Média Móvel

Nome da Linha de Tendência

- ☒ Automático
- ☐ Personalizado

Previsão

Avançar

Recuar

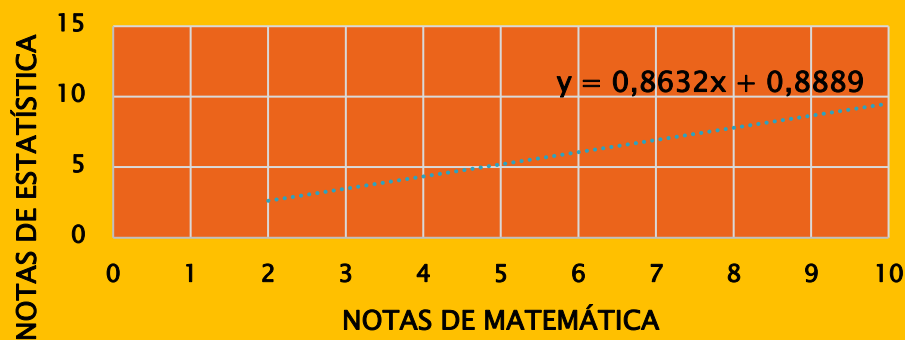
☐ Definir Interseção

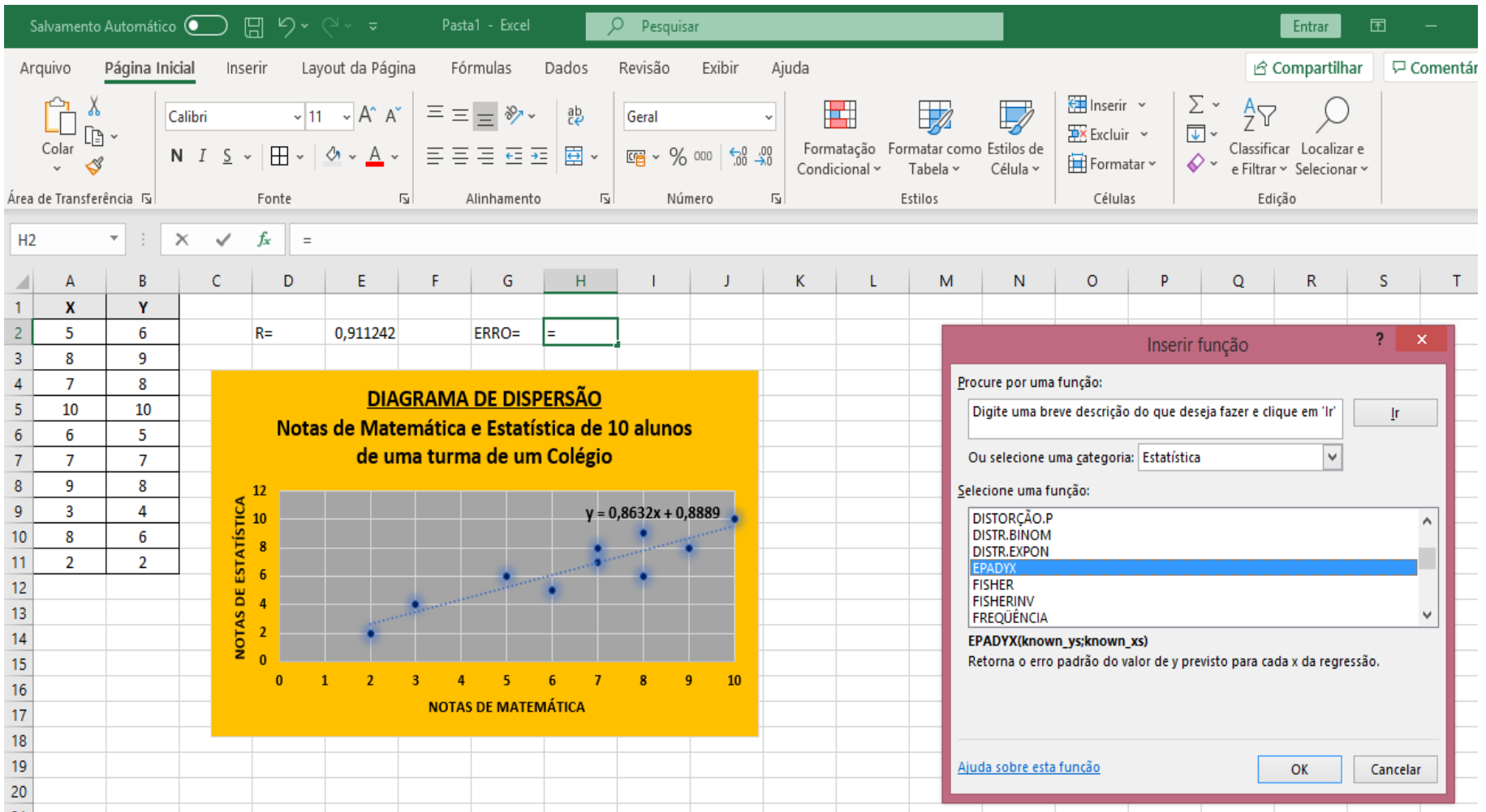
☒ Exibir Equação no gráfico

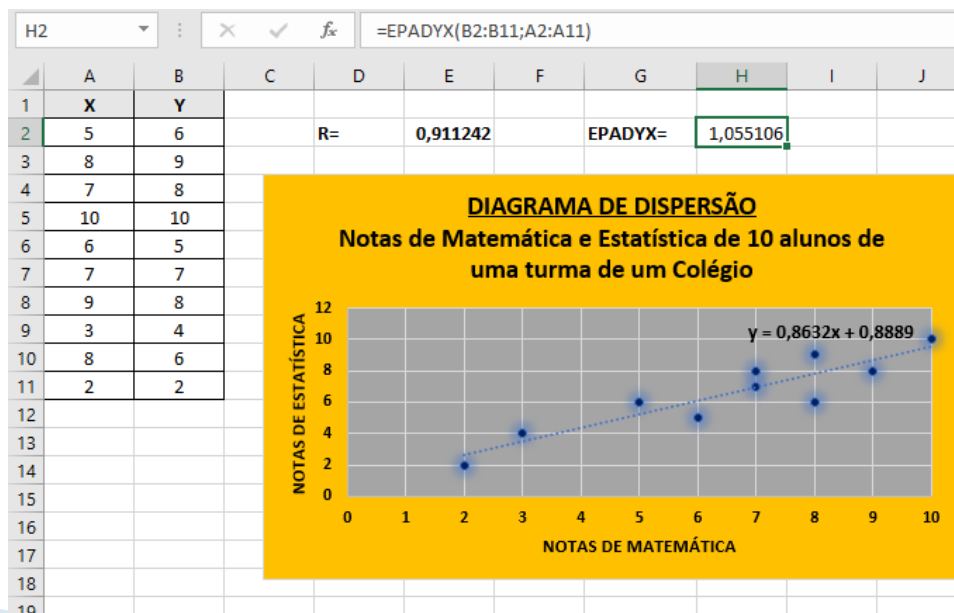
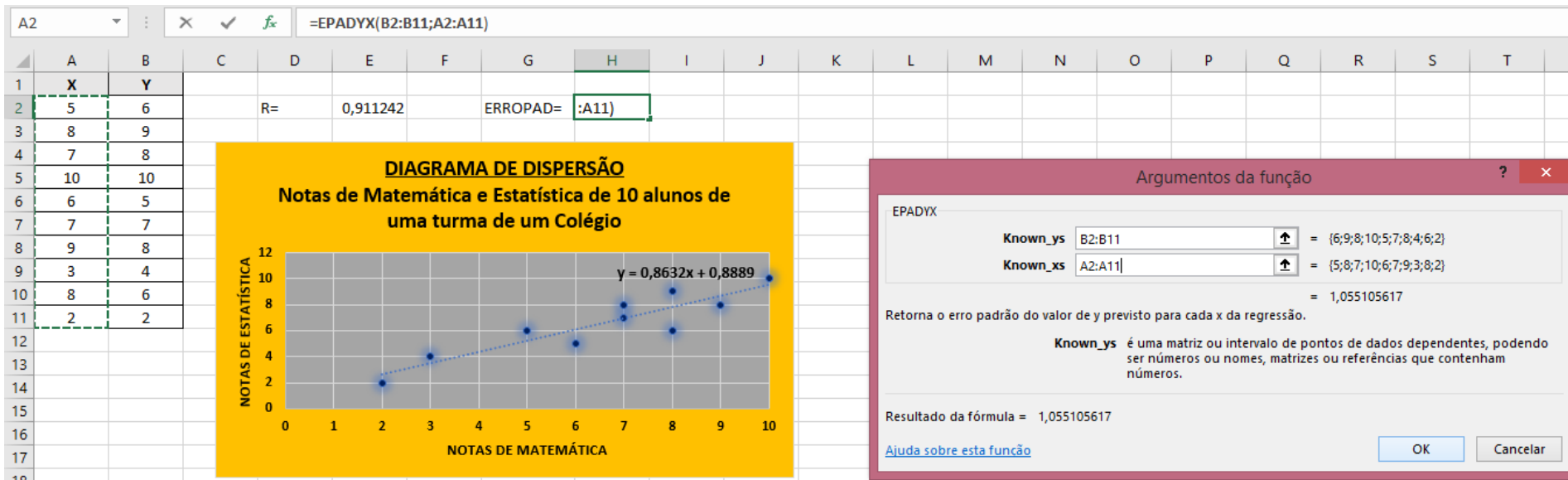
☐ Exibir valor de R-quadrado no gráfico

DIAGRAMA DE DISPERSÃO

Notas de Matemática e Estatística de 10 alunos de uma turma de um Colégio







REFERÊNCIAS BIBLIOGRÁFICAS

- ▶ BUSSAB, Wilson O. & MORETTIN, Pedro A. *Estatística Básica (Métodos Quantitativos)*, 4^a Edição, São Paulo: Atual Editora, 1987.
- ▶ SPINELLI, Walter & SOUSA, Maria Helena S. de. *Introdução à Estatística*. São Paulo: Editora Ática, 1990, cap. 9, p. 122–139.