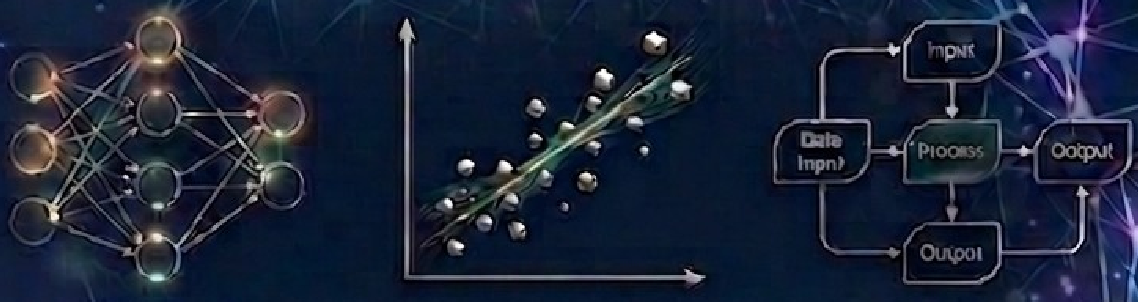
COLÉGIO PEDRO II

MANUAL PEDAGÓGICO PARA A INTRODUÇÃO AO APRENDIZADO DE MÁQUINA UTILIZANDO O MODELO DE REGRESSÃO LINEAR SIMPLES

~ UMA PROPOSTA PARA ESTUDANTES DO 3º ANO DO ENSINO MÉDIO ~



Ronaldo Cesar da Silva
Andreia Carvalho Maciel Barbosa

IMPERIAL
editora

Rio de Janeiro
2026

Introdução ao Aprendizado de Máquina Utilizando o Modelo de Regressão Linear Simples

Uma proposta para estudantes do Ensino Médio

Ronaldo Cesar da Silva

Andreia Carvalho Maciel Barbosa

Introdução ao Aprendizado de Máquina Utilizando o Modelo de Regressão Linear Simples

Uma proposta para estudantes do Ensino Médio

1ª Edição

IMPERIAL
editora

Rio de Janeiro, 2026

COLÉGIO PEDRO II

PRÓ-REITORIA DE PÓS-GRADUAÇÃO, PESQUISA, EXTENSÃO E CULTURA

BIBLIOTECA PROFESSORA SILVIA BECHER

CATALOGAÇÃO NA FONTE

S586 Silva, Ronaldo Cesar da

Manual pedagógico para a introdução ao aprendizado de máquina utilizando o modelo de regressão linear simples : uma proposta para estudantes do 3º ano do ensino médio / Ronaldo Cesar da Silva ; Andreia Carvalho Maciel Barbosa. – Rio de Janeiro : Imperial Editora, 2026.

62 p.

ISBN: 978-65-5930-230-7.

1. Matemática - Estudo e ensino. 2. Aprendizado por computador. 3. Inteligência artificial. 4. Modelagem semântica. 5. Modelos lineares (Estatística). I. Barbosa, Andreia Carvalho Maciel. II. Colégio Pedro II. III. Título.

CDD 510

Ficha catalográfica elaborada pela Bibliotecária Simone Alves – CRB-7: 5692.

RESUMO

DA SILVA, Ronaldo Cesar. **Introdução ao Aprendizado de Máquina utilizando o Modelo de Regressão Linear Simples:** Uma proposta para estudantes do 3° Ano do Ensino Médio. 2026. Produto educacional como parte integrante da Dissertação (Mestrado) – Colégio Pedro II, Pró-Reitoria de Pós-Graduação, Pesquisa, Extensão e Cultura, Programa de Mestrado Profissional em Matemática em Rede Nacional, Rio de Janeiro, 2026.

Este produto educacional pretende contribuir no campo da educação matemática com uma sequência de tarefas extracurriculares na qual serão apresentados conceitos sobre aprendizado de máquina, algoritmos supervisionados, regressão linear simples e ferramentas tecnológicas utilizadas para facilitar a análise de dados como planilhas eletrônicas (LibreOffice Calc), software matemático (GeoGebra) e linguagem de programação (Python). As atividades foram organizadas em cinco encontros de cem minutos (duas horas/aula), para serem realizadas no contraturno das aulas regulares. O modelo de Regressão Linear Simples foi apresentado gradualmente, desde o uso do celular como calculadora até a linguagem de programação Python e dessa forma, os estudantes podem participar ativamente do processo de ensino-aprendizagem, tendo a possibilidade de utilizar a tecnologia de forma ética e responsável, contribuindo para a formação de uma sociedade mais consciente quanto ao uso e desenvolvimento de aplicativos e ferramentas computacionais.

Palavras-chave: Aprendizado de Máquina, Regressão Linear Simples, Tarefas Extracurriculares.

Sumário

Apresentação das Atividades propostas

A sequência de atividades propostas para estudantes do 3° ano do ensino médio é composta por cinco encontros de 100 minutos (2 horas-aula) cada. O objetivo é introduzir o conceito de Inteligência Artificial e Aprendizado de Máquina utilizando o Modelo de Regressão Linear Simples, numa linguagem básica e de fácil entendimento para o estudante. São abordados temas como a origem do termo “inteligência artificial”, evolução da inteligência artificial ao longo do tempo, a inteligência artificial na educação, o uso ético da inteligência artificial, marco regulatório da inteligência artificial no Brasil, conceitos sobre aprendizado de máquina, tipos de aprendizado de máquina, algoritmos supervisionados, regressão linear simples e ferramentas tecnológicas utilizadas para facilitar a análise de dados como planilhas eletrônicas (LibreOffice Calc), software matemático (GeoGebra) e linguagem de programação (Python).

1° Encontro

Destina-se a abordar os temas inteligência artificial (IA), evolução da IA ao longo do tempo, aplicações da IA no cotidiano, uso da IA na educação, desafios éticos e sociais na utilização da IA, marco regulatório da inteligência artificial no Brasil, tipos de aprendizado de máquina (AM), com foco no aprendizado supervisionado e uma breve introdução ao modelo de regressão linear simples.

Aula com utilização do projetor como recurso multimídia. O objetivo dessa aula é compreender e discutir os conceitos básicos sobre inteligência artificial, identificar a influência da IA no mundo moderno e reconhecer situações do cotidiano nas quais a regressão linear simples pode ser utilizada como modelo preditivo. Com a finalidade de servir de apoio para os docentes no desenvolvimento das atividades do 1° Encontro, um modelo de slides está disponibilizado como apêndice do produto educacional.

Plano de Aula do 1° Encontro

- **Conteúdo a ser trabalhado:**

  Conceito de Inteligência Artificial;

  Evolução da Inteligência Artificial ao longo do tempo;

  Uso da Inteligência Artificial na educação;

  Desafios éticos e sociais na utilização da inteligência artificial;

  Marco regulatório da Inteligência Artificial no Brasil;

Tipos de Aprendizado de Máquina;
Introdução ao modelo de Regressão Linear Simples.

- **Objetivos:**

Compreender os conceitos básicos da Inteligência Artificial;
Observar os principais marcos da evolução da Inteligência Artificial;
Identificar os limites éticos do uso da Inteligência Artificial;
Identificar alguns casos de sucesso e erros no uso da Inteligência Artificial;
Reconhecer o Aprendizado de Máquina como uma subárea da Inteligência Artificial;
Compreender o modelo estatístico de Regressão Linear Simples;
Reconhecer situações do cotidiano nas quais o modelo de Regressão Linear Simples pode ser utilizado.

- **Habilidades relacionadas (BNCC):**

EM13CO05 - Identificar os limites da Computação para diferenciar o que pode ou não ser automatizado, buscando uma compreensão mais ampla dos limites dos processos mentais envolvidos na resolução de problemas.
EM13CO08 - Entender como mudanças na tecnologia afetam a segurança, incluindo novas maneiras de preservar sua privacidade e dados pessoais on-line, reportando suspeitas e buscando ajuda em situações de risco.

EM13CO09 - Identificar tecnologias digitais, sua presença e formas de uso, nas diferentes atividades no mundo do trabalho.

EM13CO10 - Conhecer os fundamentos da Inteligência Artificial, comparando-a com a inteligência humana, analisando suas potencialidades, riscos e limites.

- **Metodologia:**

Aula interativa visando o envolvimento ativo dos estudantes, com a utilização do projetor como recurso multimídia, tendo o professor a função de mediador do conhecimento produzido e facilitador do aprendizado, de acordo com o momento da atividade.

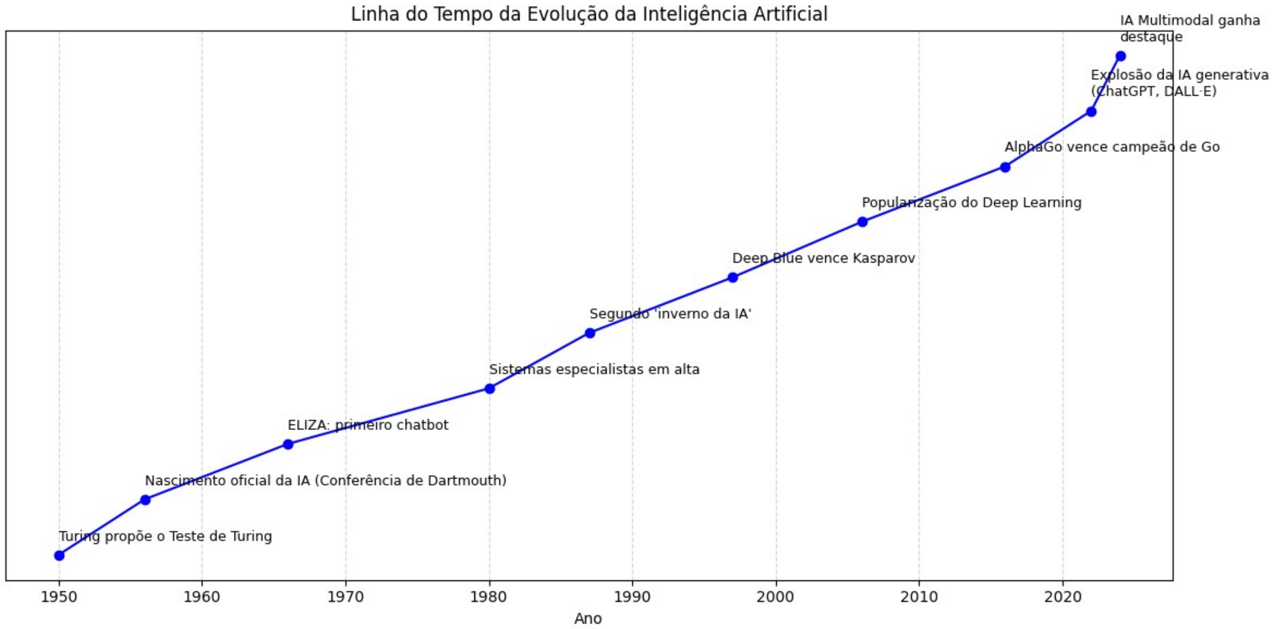
A atividade segue uma organização didática na qual o primeiro momento é de apresentação do tema, pois os conceitos necessários para um debate e exposição de ideias mais aprofundadas em torno dos conteúdos abordados estarão sendo desenvolvidos e além disso, como o conhecimento sobre o assunto varia de indivíduo para indivíduo, torna-se imprescindível que haja um nivelamento para posterior análise da produção de significado dos estudantes. Sendo assim, algumas definições sobre Inteligência Artificial devem ser discutidas, usando como base as definições constantes em Russell e Norvig (2013), que seguem basicamente quatro linhas gerais, fazendo referência ao pensamento e raciocínio e ao comportamento. A saber:

- **Modelos de sistemas/algoritmos que pensam como seres humanos:**
  "[Automatização de] atividades que associamos ao pensamento humano, atividades como a tomada de decisões, a resolução de problemas, o aprendizado..." (Bellman, 1978)
- **Modelos de sistemas/algoritmos que pensam racionalmente:**
  "O estudo das computações que tornam possível perceber, raciocinar e agir." (Winston, 1992)
- **Modelos de sistemas/algoritmos que agem como seres humanos:**
  "O estudo de como os computadores podem fazer tarefas que hoje são melhor desempenhadas pelas pessoas." (Rich and Knight, 1991)
- **Modelos de sistemas/algoritmos que agem de forma racional:**
  "IA ... está relacionada a um desempenho inteligente de artefatos." (Nilsson, 1998)

Numa visão mais abrangente, podemos definir Inteligência Artificial como sendo um campo da ciência da computação que se dedica a criar sistemas e máquinas capazes de simular a inteligência humana, realizando tarefas que normalmente exigem raciocínio, aprendizado e tomada de decisão. Em outras palavras, a IA busca desenvolver sistemas que possam aprender com dados, reconhecer padrões, resolver problemas e criar conteúdo.

Seguindo com a atividade, são explorados eventos históricos que foram relevantes no avanço da Inteligência Artificial ao longo do tempo e para tal, utilizamos a linha do tempo da figura 1 como apoio didático.

Figura 1: Marcos históricos da evolução da Inteligência Artificial



Fonte: Elaborado pelo autor

Com o intuito de despertar a curiosidade e o interesse dos estudantes, maior ênfase deve ser dada a eventos de grande impacto relacionados ao cotidiano, como carros autônomos, sistema de recomendação de filmes no streaming e compras na internet, assistentes virtuais, ChatGPT entre outros.

O uso da Inteligência Artificial na educação é visto de forma a causar reflexão, para que o estudante perceba que tais ferramentas podem ser usadas para personalizar o aprendizado, otimizar a gestão escolar e auxiliar professores. A IA pode adaptar o conteúdo, fornecer feedback imediato e analisar o progresso dos alunos, criando experiências de aprendizagem mais eficientes e inclusivas. Importante destacar que a IA tem limitações e não substituirá o esforço e dedicação nos estudos, mas pode ser uma grande aliada no processo de ensino-aprendizagem. Reportagem do jornal O Globo do dia 28/07/2025 (figura 2), mostra que o Conselho Nacional de Educação (CNE), estuda um meio de incluir a Inteligência Artificial de forma obrigatória, no currículo para a formação de professores.

Figura 2: Reportagem do Jornal O Globo do dia 28/07/2025


O GLOBO ines249 9

Brasil

Motéis se adaptam para a COP 30

NOVOS TEMPOS

Ensino de Inteligência Artificial deve se tornar obrigatório no currículo para formar professores

REDES JÁ ADOTAM TECNOLOGIA

São Paulo

Rio de Janeiro

Espírito Santo

PLANO POR ESCOLA

Fonte: Jornal O Globo

Na sequência da atividade, deve-se mostrar casos de uso da IA que foram prejudiciais a pessoas físicas e jurídicas além de estimularem discriminação. O uso ético e socialmente correto da IA será abordado, com ênfase nos esforços do governo brasileiro em criar leis e dispositivos de proteção dos direitos dos cidadãos, como a Lei Geral de Proteção de Dados Pessoais (LGPD) promulgada em 2018, que estabelece regras claras para a coleta, tratamento e armazenamento de dados pessoais, tanto em meios físicos quanto digitais, por pessoas físicas ou jurídicas, públicas ou privadas e o Marco Regulatório da Inteligência Artificial no Brasil que está em discussão e análise no Congresso Nacional, com o Projeto de Lei 2338/23. Este projeto visa definir normas para o desenvolvimento, implementação e uso responsável de sistemas de IA no país, buscando garantir segurança jurídica e ética, além de proteger direitos fundamentais.

Em prosseguimento, os modelos de Aprendizado de Máquina são apresentados como uma subárea da IA, com destaque ao modelo de Aprendizado Supervisionado e ao modelo estatístico de Regressão Linear Simples, que é o objeto central deste trabalho e será detalhadamente esmiuçado no 2° Encontro. Os momentos finais deste 1° Encontro são utilizados para a aplicação de uma folha de tarefas. Segue abaixo o modelo de perguntas que foram utilizadas:

Folha de tarefas 1

1) Inteligência Artificial e Aprendizado de Máquina são termos que representam o mesmo conceito? Justifique sua resposta.

2) Você considera que a Inteligência Artificial já influenciou ou estimulou em algum momento, sua tomada de decisão? Se sim, cite um exemplo.

3) Dentro de sua percepção sobre o que foi trabalhado em sala de aula, qual a diferença entre aprendizado supervisionado e aprendizado não supervisionado?

4) Dentro de sua percepção sobre o que foi trabalhado em sala de aula, qual a diferença entre correlação linear positiva e correlação linear negativa?

5) Cite um exemplo diferente dos vistos na aula, em que um modelo de regressão linear simples pode ser utilizado em um problema real.

2° Encontro

Tem a finalidade de aprofundar o conhecimento sobre o modelo de Regressão Linear Simples. Para tal, conceitos estatísticos como diagrama de dispersão, médias, desvio-padrão e variância devem ser revisitados, assim como a função polinomial do 1° grau, com o objetivo de nivelamento dos estudantes.

O segundo momento da aula tem o propósito de definir os conceitos de coeficiente de correlação linear, coeficiente de determinação e o método dos mínimos quadrados, que são a base para encontrarmos a reta que melhor se ajusta aos dados. Uma situação-problema com poucos dados amostrais é apresentada para servir de apoio para a introdução dos conceitos e realização dos cálculos. Os materiais de apoio são o quadro branco, caneta para quadro branco e projetor. Para agilizar, uma calculadora poderá ser utilizada.

A intenção é focar na teoria e cálculos matemáticos e deixar a utilização de recursos tecnológicos para os próximos encontros. Nos 30 minutos finais, uma folha de tarefas será entregue para reforçar o conteúdo e compor o material para análise de resultados.

Plano de Aula do 2° Encontro

- **Conteúdo a ser trabalhado:**

  Função Polinomial do 1° Grau;
  Gráfico de Dispersão;
  Média Aritmética, Variância e Desvio Padrão;
  Coeficiente de Correlação Linear;
  Coeficiente de Determinação;
  Modelo de Regressão Linear Simples.

- **Objetivos:**

  Compreender o conceito de Correlação Linear;
  Reconhecer os tipos de Correlação Linear;
  Calcular o Coeficiente de Correlação Linear entre duas variáveis;
  Calcular o Coeficiente de Determinação;
  Interpretar o Coeficiente de Determinação;
  Construir um modelo de Regressão Linear Simples;

Interpretar os resultados obtidos a partir do modelo de Regressão Linear Simples;

Aplicar a Regressão Linear Simples em situações reais.

- **Habilidades relacionadas (BNCC):**

  EM13MAT102 - Analisar gráficos e métodos de amostragem de pesquisas estatísticas apresentadas em relatórios divulgados por diferentes meios de comunicação, identificando, quando for o caso, inadequações que possam induzir a erros de interpretação, como escalas e amostras não apropriadas.

  EM13MAT302 - Resolver e elaborar problemas cujos modelos são as funções polinomiais de 1º e 2º graus, em contextos diversos, incluindo ou não tecnologias digitais.

  EM13MAT316 - Resolver e elaborar problemas, em diferentes contextos, que envolvem cálculo e interpretação das medidas de tendência central (média, moda, mediana) e das de dispersão (amplitude, variância e desvio padrão).

  EM13MAT409 - Interpretar e comparar conjuntos de dados estatísticos por meio de diferentes diagramas e gráficos, como o histograma, o de caixa (box-plot), o de ramos e folhas, reconhecendo os mais eficientes para sua análise.

  EM13MAT501 - Investigar relações entre números expressos em tabelas para representá-los no plano cartesiano, identificando padrões e criando conjecturas para generalizar e expressar algebricamente essa generalização, reconhecendo quando essa representação é de função polinomial de 1º grau.

- **Metodologia:**

Os pesquisadores em Educação Matemática, de forma geral, estimulam a participação ativa dos estudantes no processo de aprendizagem, porém, para que o processo seja efetivo no que propõe, os professores de Matemática da Educação Básica devem possuir conhecimento e habilidades suficientes para fazer a leitura correta do que está sendo produzido pelo aluno. Entendemos que o Modelo dos Campos Semânticos é uma fonte muito rica para a análise da produção de significados por parte dos estudantes.

De acordo com Loth e Silva (2013), a partir das ideias centrais do MCS, uma proposta de tarefa bem elaborada deve permitir ao professor, dentre outras situações, que perceba e deixe claro para os estudantes que os significados produzidos por eles e/ou os significados oficiais da matemática são alguns dentre os vários significados que podem ser produzidos a partir de uma tarefa.

Nessa linha de pensamento, uma situação-problema adaptada de Crespo (2004) foi apresentada para os estudantes com o propósito de aguçar a curiosidade e participação, além de gerar um ambiente propício para o desenvolvimento dos conteúdos trabalhados.

Situação-problema

Escolhemos de forma aleatória, 10 dentre 98 estudantes do 3° ano do Ensino Médio de uma escola estadual e verificamos suas notas em Matemática e Física no 1° bimestre. As notas estão na tabela abaixo:

Tabela 1 - Notas de Matemática e Física

Número do estudante	Matemática	Física
01	5,0	4,0
08	8,0	9,0
24	7,0	8,0
38	10,0	10,0
44	6,0	5,0
58	7,0	7,0
59	9,0	8,0
72	3,0	4,0
80	8,0	6,0
92	2,0	2,0

Fonte: Elaborado pelo autor

Vamos verificar a Média Aritmética Simples das notas de Matemática e Física desta amostra de 10 alunos, lembrando que a Média Aritmética Simples é calculada pela seguinte fórmula:

$$\overline{x}=\frac{\sum_{i=1}^{n} x_i}{n}$$

Sendo: $\overline{x}$ a média aritmética;

x_i os valores da variável;

n a quantidade de valores e

$\sum_{i=1}^{n} x_i$ o somatório dos números x_i, de modo que que i varia de $1 a n$.

Logo, temos:

- $Média\, em\, Matemática = \frac{5{,}0+8{,}0+7{,}0+10{,}0+6{,}0+7{,}0+9{,}0+3{,}0+8{,}0+2{,}0}{10} = 6{,}5$

- $Média\, em\, Física = \frac{4{,}0+9{,}0+8{,}0+10{,}0+5{,}0+7{,}0+8{,}0+4{,}0+6{,}0+2{,}0}{10} = 6{,}3$

Como a Média é uma medida de tendência central, tem-se um único valor como representante de um determinado grupo de dados e a presença de valores muito maiores ou muito menores em relação à Média, faz com que a Média Aritmética não consiga passar uma ideia fidedigna do perfil dos dados analisados. Quando a medida de tendência central não é suficiente para caracterizar um grupo de dados, são utilizadas medidas de dispersão, que medem o grau de variação desses dados e as mais utilizadas são a Variância e o Desvio Padrão.

A Variância tem por base os desvios em torno da Média Aritmética e o Desvio Padrão é a raiz quadrada da Variância. De acordo com Dante (2002), não é possível expressar a Variância na mesma unidade dos valores da variável, uma vez que os desvios são elevados ao quadrado, daí a importância da utilização do Desvio Padrão, que facilita a interpretação dos dados, uma vez que é expresso na mesma unidade dos valores observados. Suas fórmulas são:

Variância (s^2) = $\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}$ para uma população ou $\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$ para uma amostra.

Desvio Padrão (s) = $\sqrt{s^2}$

Vamos calcular a Variância e o Desvio Padrão das notas em Matemática e Física da amostra de 10 alunos.

- $Variância\, das\, notas\, em\, Matemática$

$$S^2 = \frac{(5{,}0-6{,}5)^2+(8{,}0-6{,}5)^2+(7{,}0-6{,}5)^2+(10{,}0-6{,}5)^2+(6{,}0-6{,}5)^2+(7{,}0-6{,}5)^2+(9{,}0-6{,}5)^2+(3{,}0-6{,}5)^2+(8{,}0-6{,}5)^2+(2{,}0-6{,}5)^2}{10-1}$$

$$S^2 = \frac{58{,}5}{9} = 6{,}50$$

- $Desvio\, Padrão\, das\, notas\, em\, Matemática$

$$S = \sqrt{6{,}5} = 2{,}55$$

- $Variância\, das\, notas\, em\, Física$

$$S^2 = \frac{(4{,}0-6{,}3)^2+(9{,}0-6{,}3)^2+(8{,}0-6{,}3)^2+(10{,}0-6{,}3)^2+(5{,}0-6{,}3)^2+(7{,}0-6{,}3)^2+(8{,}0-6{,}3)^2+(4{,}0-6{,}3)^2+(6{,}0-6{,}3)^2+(2{,}0-6{,}3)^2}{10-1}$$

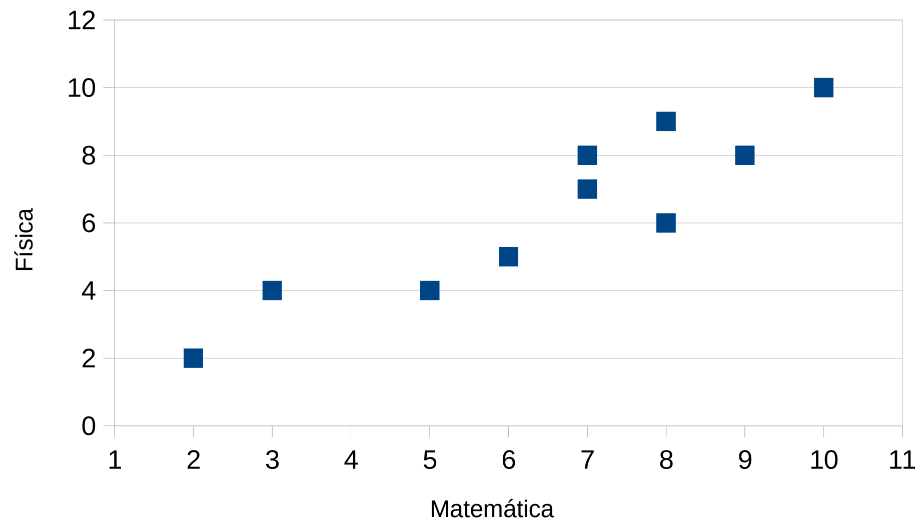
$S^2 = \frac{58,1}{9} = 6,45$

- *Desvio padrão das notas em Física*

$S = \sqrt{6,45} = 2,54$

Como os valores das médias e dos desvios padrão amostrais das notas em Matemática e Física ficaram muito próximos, uma pergunta torna-se pertinente: Será que existe correlação entre as notas de Matemática e Física, ou seja, será que o estudante que conseguiu ter um bom desempenho em Matemática obteve um bom desempenho em Física? Para entendermos melhor o que é correlação entre duas variáveis quantitativas (variáveis que podem ser expressas por números), um gráfico de dispersão[1] foi construído, considerando as notas de Matemática como sendo os valores x_i e as notas de Física sendo os valores y_i, formando pares ordenados (x_i, y_i). Gráfico de dispersão é a representação desses pares ordenados no plano cartesiano.

Figura 2: Gráfico de Dispersão das Notas de Matemática e Física



Fonte: O autor

Olhando para o gráfico de dispersão, parece haver algum tipo de associação entre as notas de Matemática e Física, pois quando uma aumenta, a outra também aumenta de

1 Alguns autores denominam diagrama de dispersão.

maneira aparentemente linear. Dizemos que uma correlação entre duas variáveis é linear, quando a relação entre as variáveis pode ser aproximada por uma reta, ou seja, pode-se aproximá-las por uma função polinomial do 1° grau.

Uma correlação linear é positiva quando uma variável está diretamente relacionada com a outra variável, ou seja, quando uma aumenta, a outra também aumenta. Quando uma variável está inversamente relacionada com a outra variável, dizemos que a correlação linear é negativa, pois uma aumenta e a outra diminui. Existem situações nas quais as variáveis não possuem correlação linear e isso acontece quando a correlação não é linear ou as variáveis não possuem qualquer relação entre si.

O instrumento estatístico utilizado para medir a correlação linear entre duas variáveis é o Coeficiente de correlação linear, que indica a intensidade e o sentido da correlação. Para encontrarmos o coeficiente de correlação linear entre as notas de Matemática e Física, foi utilizado o Coeficiente de correlação de Pearson (r), que é dado por:

$$r=\frac{n\sum_{i=1}^{n} x_i y_i-(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{\sqrt{[n\sum_{i=1}^{n} x_i^2-(\sum_{i=1}^{n} x_i)^2][n\sum_{i=1}^{n} y_i^2-(\sum_{i=1}^{n} y_i)^2]}}$$

Sendo x_i os valores assumidos pela variável x, y_i os valores assumidos pela variável y e n o número de pares de dados da amostra. Os valores limites para r estão no intervalo $[-1,+1]$.

- Valores de $r=+1$ indicam uma correlação linear perfeita e positiva;
- Valores de $r=-1$ indicam uma correlação linear perfeita e negativa;
- Valores de $r=0$ indicam que não existe correlação linear ou então, a relação entre as variáveis não é linear.

Segundo Crespo (2004), tem-se que:

- Se $0{,}3\leqslant|r|<0{,}6$, há uma correlação relativamente fraca entre as variáveis;
- Se $0\leqslant|r|<0{,}3$, a correlação é muito fraca e praticamente nada pode ser concluído sobre a relação entre as variáveis;
- Se $0{,}6\leqslant|r|<1$, há indícios de uma correlação significativa entre as variáveis.

Voltando à situação-problema, para encontrar o coeficiente de correlação linear de

Pearson entre as notas de Matemática e Física é indicado o uso de uma tabela para melhor visualização:

Tabela 2: Somatórios para o cálculo do coeficiente de correlação

Matemática (x_i)	Física (y_i)	$x_i y_i$	x_i^2	y_i^2
5,0	4,0	20,0	25,0	16,0
8,0	9,0	72,0	64,0	81,0
7,0	8,0	56,0	49,0	64,0
10,0	10,0	100,0	100,0	100,0
6,0	5,0	30,0	36,0	25,0
7,0	7,0	49,0	49,0	49,0
9,0	8,0	72,0	81,0	64,0
3,0	4,0	12,0	9,0	16,0
8,0	6,0	48,0	64,0	36,0
2,0	2,0	4,0	4,0	4,0
$\sum x_i = 65$	$\sum y_i = 63$	$\sum x_i y_i = 463$	$\sum x_i^2 = 481$	$\sum y_i^2 = 455$

Fonte: O autor

Como a amostra tem $n=10$, temos que:

$$r = \frac{10 \cdot 463 - 65 \cdot 63}{\sqrt{(10 \cdot 481 - 65^2)(10 \cdot 455 - 63^2)}}$$

$$r = \frac{4630 - 4095}{\sqrt{(4810 - 4225)(4550 - 3969)}}$$

$$r = \frac{535}{\sqrt{585 \cdot 581}}$$

$$r = 0{,}92$$

Esse valor indica uma forte correlação linear positiva entre as notas de Matemática e de Física, porém é importante destacar aos estudantes que correlação não implica causalidade. Apesar de uma correlação linear forte entre duas variáveis indicar que elas estão associadas de alguma forma, não podemos concluir que uma variável causa diretamente a mudança na outra. A correlação descreve a relação entre as variáveis, mas não indica necessariamente uma relação de causa e efeito.

Uma vez que $r = 0{,}92$ devemos definir um modelo de regressão linear amostral simples, com uma variável independente (x) (notas de Matemática) e uma variável dependente (y) (notas de Física), que terá o formato de uma função polinomial do 1° grau $(y=ax+b)$ e vai representar a reta que melhor se ajusta aos dados amostrais. Para isso é necessário determinar os parâmetros a (coeficiente angular) e b (coeficiente linear) da função e utilizaremos o Método dos Mínimos Quadrados (MMQ), que é uma técnica estatística usada para encontrar a melhor função de forma a minimizar a soma dos quadrados das diferenças (erros ou resíduos) entre os valores observados e os valores previstos pelo modelo. Essas diferenças são devidas aos fatores ou variáveis que não foram consideradas pelo modelo. Sendo assim, pelo MMQ, temos que o coeficiente angular (a) é dado por:

$$a=\frac{n\sum_{i=1}^{n} x_i y_i-(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n\sum_{i=1}^{n} x_i^2-(\sum_{i=1}^{n} x_i)^2}$$

e o coeficiente linear (b) é:

$$b=\bar{y}-a\bar{x}$$

onde n é o número de dados, $\bar{x}$ é a média dos valores x_i e $\bar{y}$ é a média dos valores y_i.

Na situação-problema apresentada, trabalhamos utilizando uma amostra para encontrarmos os parâmetros a e b e essa é uma estimativa da verdadeira equação de regressão, sendo assim, escrevemos a equação da seguinte forma:

$\hat{y}=ax+b$, onde $\hat{y}$ é o y estimado.

Com o auxílio de uma tabela, encontramos os valores utilizados para a definição dos parâmetros.

Tabela 3: Somatórios para o cálculo dos parâmetros do modelo de regressão

Matemática (x_i)	Física (y_i)	$x_i y_i$	x_i^2
5,0	4,0	20,0	25,0
8,0	9,0	72,0	64,0
7,0	8,0	56,0	49,0
10,0	10,0	100,0	100,0
6,0	5,0	30,0	36,0
7,0	7,0	49,0	49,0
9,0	8,0	72,0	81,0
3,0	4,0	12,0	9,0
8,0	6,0	48,0	64,0
2,0	2,0	4,0	4,0
$\sum x_i=65$	$\sum y_i=63$	$\sum x_i y_i=463$	$\sum x_i^2=481$

Fonte: O autor

Lembrando que o número de amostras é $n=10$, temos:

$$a=\frac{10\cdot 463-65\cdot 63}{10\cdot 481-(65)^2}=\frac{4630-4095}{4810-4225}=\frac{535}{585}$$

$$a=0{,}9145$$

Daí, como já foi calculado, temos $\bar{x}=6{,}5$ e $\bar{y}=6{,}3$ vem que:

$$b=6{,}3-0{,}9145\cdot 6{,}5$$

$$b=0{,}3557$$

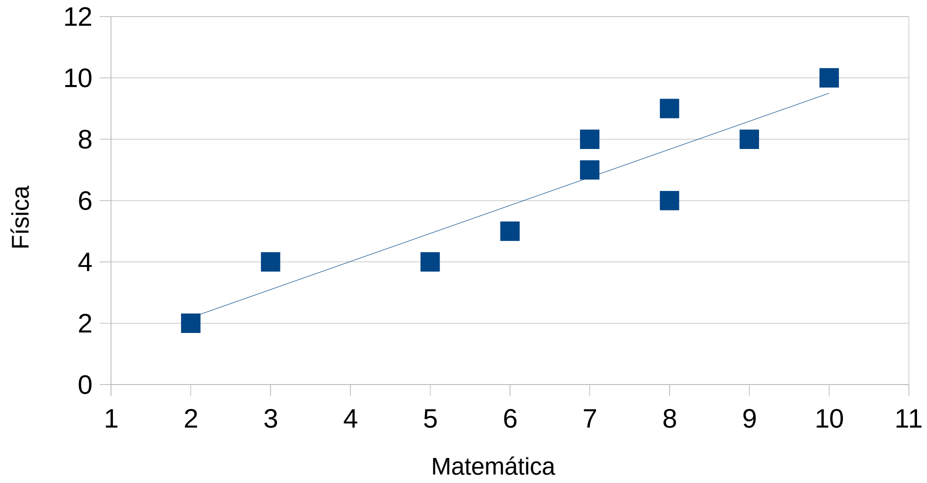
Arredondando para duas casas decimais tem-se $a=0{,}91$ e $b=0{,}35$.

Logo, $\hat{y}=0{,}91x+0{,}35$

Vale destacar que fazendo $x=6{,}5$, encontramos $\hat{y}=6{,}3$, o que mostra que as médias de x_i e y_i pertencem à reta de regressão linear encontrada. Para traçarmos a reta no plano cartesiano basta encontrar mais um ponto que pertença a reta e fazendo $x=0$ o valor encontrado é $\hat{y}=0{,}35$.

Segue a reta de regressão linear encontrada:

Figura 3: Reta de Regressão Linear



Fonte: O autor

Uma vez encontrada a reta de regressão linear, podemos estimar a nota em Física de um aluno que tirou $7{,}5$ em Matemática. Fazendo $x=7{,}5$ na função $\hat{y}=0{,}91x+0{,}35$, temos:

$\hat{y}=0{,}91\cdot 7{,}5+0{,}35\Rightarrow\hat{y}=7{,}2$

Vemos que os dados amostrais x_i variam no intervalo fechado $[2{,}10]$ e sendo assim, salvo que seja demonstrada através de considerações teóricas ou experimentais a possibilidade de extrapolação, é conveniente não trabalhar com dados fora do intervalo de observação, pois o modelo não foi ajustado levando em consideração valores fora do intervalo.

Com o modelo de regressão linear encontrado, podemos encontrar a qualidade do ajuste do modelo, ou seja, podemos inferir o quanto o modelo explica a variação dos dados. No caso do nosso exemplo, é possível verificar o quanto a nota em Matemática explica a nota em Física. Para tal será utilizada uma ferramenta estatística chamada de Coeficiente de Determinação (R^2) que no caso da regressão linear simples é o quadrado do Coeficiente de Correlação de Pearson (r).

Em nosso exemplo, encontramos o valor de $r=0,92$ logo, o valor do Coeficiente de Determinação é $R^2=0,85$. Isso significa que 85% da variação dos dados é explicada pelo modelo, ou seja, os 15% não explicados, referem-se a variáveis que não foram consideradas.

Na interpretação dos resultados obtidos, pode-se considerar que um estudante com notas elevadas em Matemática tem a tendência de ter boas notas também em Física, uma vez que Matemática e Física são disciplinas da área de exatas e que a Matemática é uma ferramenta importantíssima para a explicação de fenômenos físicos, porém não sabemos se as avaliações foram feitas no mesmo dia, se algum estudante não estava se sentindo bem de saúde para fazer uma das avaliações ou ainda se no dia de uma das verificações havia outra agendada. Sendo assim, vários fatores que poderiam interferir diretamente nas notas não foram considerados, e isso justifica o valor de R^2 não ser 100%. Outra verificação que pode ser feita através da reta de ajuste é que o aumento de uma unidade na nota em Matemática gera um aumento de $0,91$ na nota de Física. Segue a folha de tarefas do 2° Encontro.

Folha de tarefas 2

1) Uma amostra de cinco estudantes de um grupo de 50 foi retirada para investigar a relação entre a quantidade de horas de estudo por dia (x) e a nota final em Matemática (y) de um grupo de estudantes. Os dados coletados estão apresentados na tabela abaixo:

Horas de estudo por dia (x)	Nota final em Matemática (y)
2	63
4	69
6	84
8	92
10	99

a) Você considera que o estudante que investe uma quantidade maior de horas de estudo tem mais possibilidade de conseguir um bom desempenho na avaliação? Justifique sua resposta.

b) Represente os dados em um diagrama de dispersão.

c) Apenas observando o gráfico de dispersão, podemos dizer que existe correlação entre os dados? Justifique sua resposta.

d) Determine a reta de regressão linear e represente a reta no diagrama de dispersão.

e) Calcule o Coeficiente de Correlação Linear de Pearson e o Coeficiente de Determinação.

f) Use a reta de regressão encontrada para prever a nota final de um estudante que estuda 7 horas por dia.

3° Encontro

A BNCC propõe que os estudantes utilizem tecnologias, como calculadoras e planilhas eletrônicas, desde os anos iniciais do Ensino Fundamental (Brasil, 2018). Neste encontro, os estudantes serão estimulados a usar o LibreOffice Calc, por ser uma planilha inserida em um software livre e de código aberto, para criar retas de regressão linear a partir de uma quantidade maior de dados, pois quanto mais dados coletamos, melhor será o ajuste da reta e os dados serão processados de forma muito mais rápida do que se estivéssemos que fazer os cálculos à mão ou apenas com a calculadora. Os dados trabalhados podem ser inseridos manualmente ou importados de uma fonte externa. Com o uso do projetor mostrar-se-ão formas de encontrar a reta de regressão com comandos da própria planilha. Os alunos aprenderão a preparar um conjunto de dados, selecionar os dados, realizar a regressão, configurar a regressão e analisar os resultados. Com poucos toques no teclado, uma reta de regressão será apresentada, com os coeficientes de correlação linear e de determinação exibidos na tela. Com o uso da planilha, a análise dos dados pode ser feita de forma mais eficiente. Diferentes dados serão trabalhados pelos estudantes para que eles possam identificar e observar os dados que possuem boa correlação linear ou não e observar as retas de regressão obtidas. O ideal é que essa aula seja feita no laboratório de informática, mas em último caso, pode ser feita na sala de aula com a utilização de notebooks ou tablets da própria escola. Dependendo da quantidade de alunos inscritos na oficina, grupos podem ser formados para uso compartilhado dos computadores. A folha de tarefas será entregue no momento em que os alunos estiverem usando os computadores.

Plano de Aula do 3° Encontro

- **Conteúdo a ser trabalhado:**

  Apresentação da planilha eletrônica LibreOffice Calc;
  Inserção de dados na planilha;
  Construção de Gráfico de Dispersão;
  Função estatística Correl;
  Função estatística Rquad;
  Construção da Reta de Regressão Linear Simples.

- **Objetivos:**

  Inserir dados na planilha eletrônica LibreOffice Calc;

  Aplicar funções estatísticas para análise dos dados inseridos;

  Construir uma Reta de Regressão Linear Simples no LibreOffice Calc;

  Analisar os resultados obtidos com o auxílio da planilha eletrônica LibreOffice Calc.

- **Habilidades relacionadas (BNCC):**

  EM13MAT302 - Resolver e elaborar problemas cujos modelos são as funções polinomiais de 1º e 2º graus, em contextos diversos, incluindo ou não tecnologias digitais.

  EM13MAT501 - Investigar relações entre números expressos em tabelas para representá-los no plano cartesiano, identificando padrões e criando conjecturas para generalizar e expressar algebricamente essa generalização, reconhecendo quando essa representação é de função polinomial de 1º grau.

  EM13MAT409 - Interpretar e comparar conjuntos de dados estatísticos por meio de diferentes diagramas e gráficos, como o histograma, o de caixa (box-plot), o de ramos e folhas, reconhecendo os mais eficientes para sua análise.

  EM13CO09 - Identificar tecnologias digitais, sua presença e formas de uso, nas diferentes atividades no mundo do trabalho.

  EM13CO11 - Criar e explorar modelos computacionais simples para simular e fazer previsões, identificando sua importância no desenvolvimento científico.

  EM13CO12 - Produzir, analisar, gerir e compartilhar informações a partir de dados, utilizando princípios de ciência de dados.

- **Metodologia:**

O encontro inicia-se com a apresentação de uma situação-problema com o objetivo de ambientar os estudantes com a estrutura da planilha eletrônica LibreOffice Calc.

Situação-problema

Adaptado de https://www.ufrgs.br/probabilidade-estatistica/listas/listas_3/lista3.5-3.6.pdf

O gerente de uma indústria localizada em um país tropical suspeita que há uma correlação entre temperatura do dia e produtividade. Dados coletados aleatoriamente ao longo de um período de seis meses revelaram o seguinte:

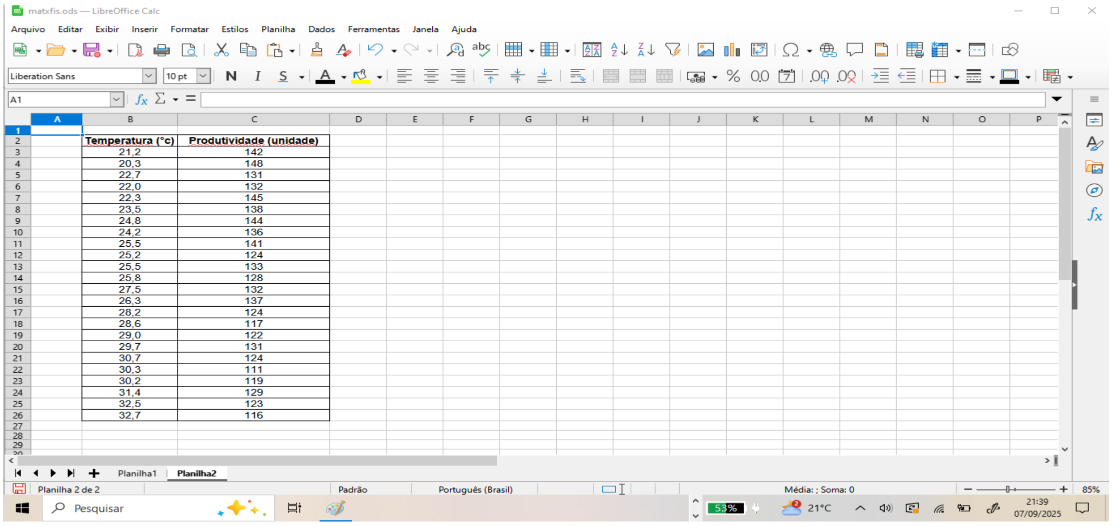
Tabela 4: Temperatura x Produtividade

Observações	Temperatura (°c)	Produtividade (unidade)
1	21,2	142
2	20,3	148
3	22,7	131
4	22	132
5	22,3	145
6	23,5	138
7	24,8	144
8	24,2	136
9	25,5	141
10	25,2	124
11	25,5	133
12	25,8	128
13	27,5	132
14	26,3	137
15	28,2	124
16	28,6	117
17	29	122
18	29,7	131
19	30,7	124
20	30,3	111
21	30,2	119
22	31,4	129
23	32,5	123
24	32,7	116

Fonte: O autor

Para começarmos a trabalhar com os dados, devemos inserí-los na planilha. Podemos inserir os dados manualmente, digitando um a um, ou importamos os dados de alguma fonte externa, como banco de dados ou outro arquivo que os contenha.

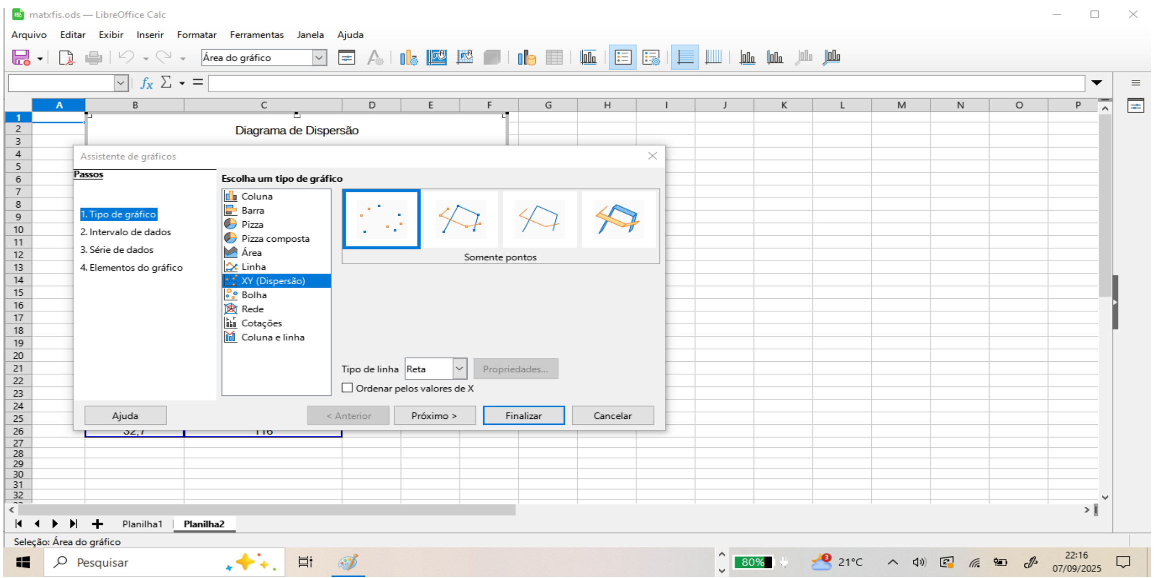
Figura 4: Inserção de dados na planilha



Fonte: O autor

Depois da inserção dos dados na planilha, o próximo passo é construir o gráfico de dispersão. Através da barra de menu, clique em Inserir – Gráfico – Tipo de Gráfico e em seguida escolha XY (Dispersão).

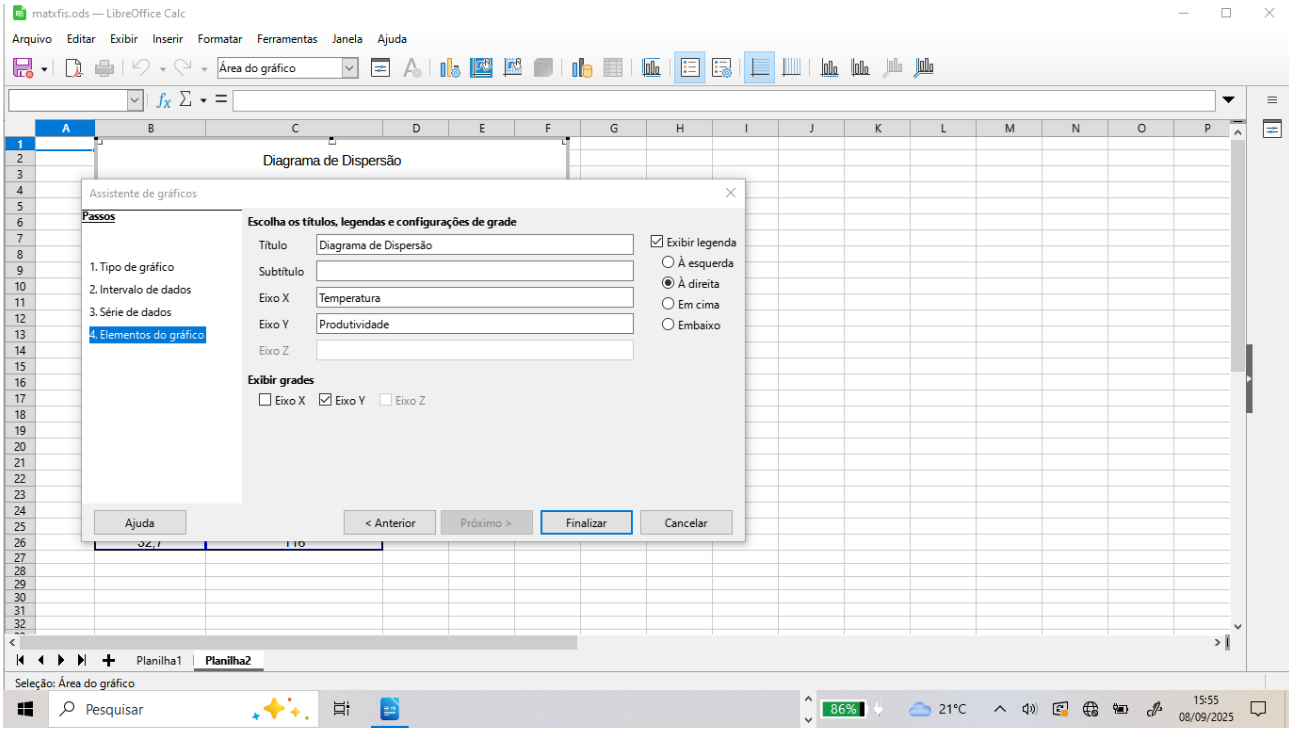
Figura 5: Inserção do gráfico de dispersão na planilha



Fonte: O autor

Ainda no Assistente de gráficos, clique no passo 4 - Elementos do gráfico para escolha do título, legenda e configuração de grade.

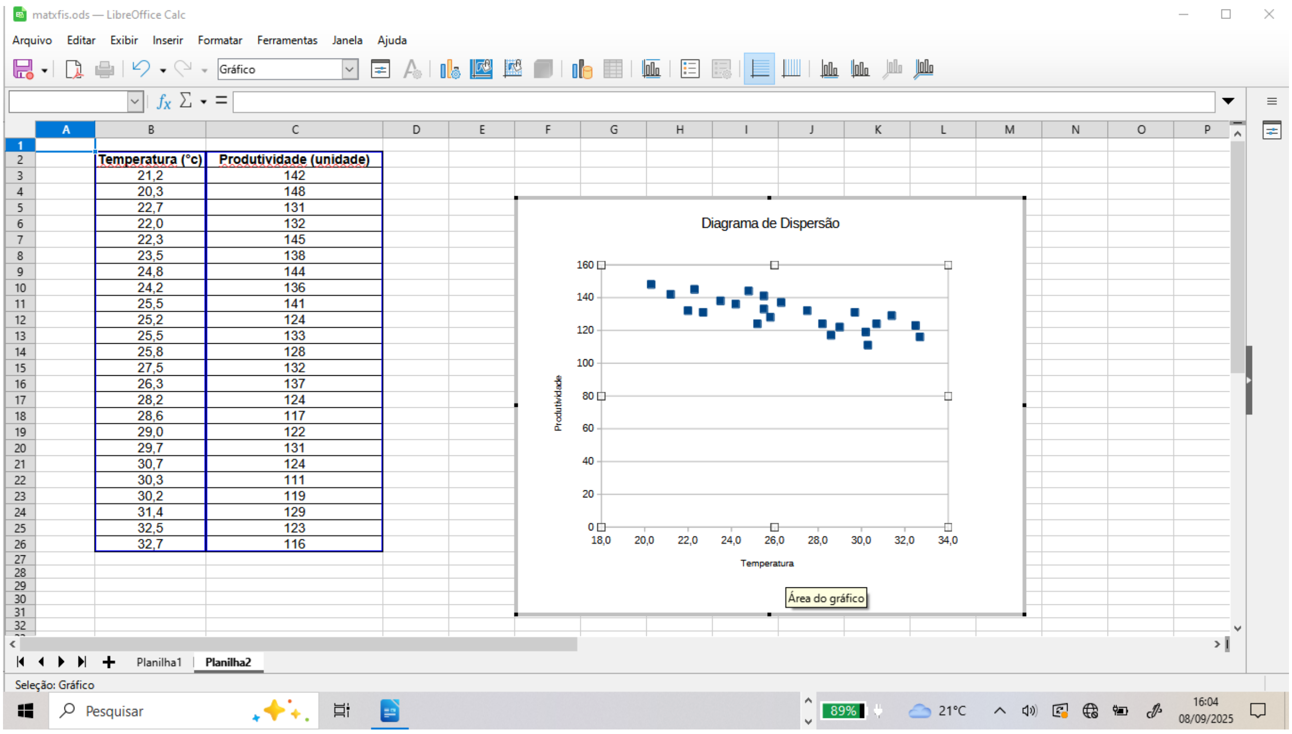
Figura 6: Configuração do gráfico de dispersão na planilha



Fonte: O autor

Na sequência, clicando em finalizar, o gráfico poderá ser visualizado.

Figura 7: Visualização do gráfico de dispersão na planilha

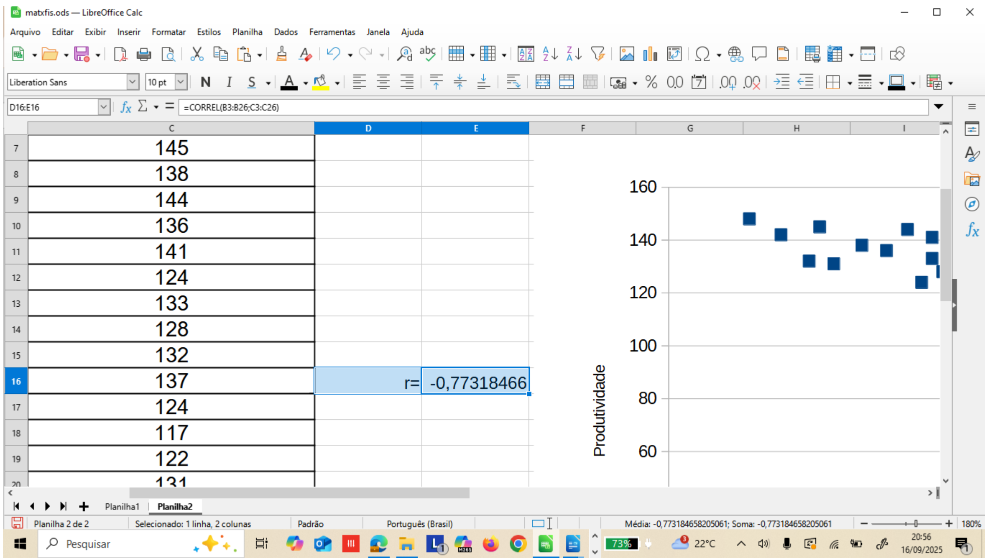


Fonte: O autor

Observando o gráfico, parece haver uma correlação linear negativa entre temperatura e produtividade e o Coeficiente de Correlação Linear de Pearson mostrará o grau dessa associação. Através da função CORREL, o LibreOffice fornece o valor de r.

Uma maneira de obter o coeficiente de correlação de dois intervalos de células é escolher uma célula da planilha e digitar: $=CORREL(intervalo\,1;intervalo\,2)$. No caso do problema em questão, o intervalo 1 está na coluna B (B3:B26) e o intervalo 2 está na coluna C (C3:C26). Sendo assim, a fórmula a ser digitada é $=CORREL(B3:B26;C3:C26)$. Teclando enter, O LibreOffice Calc calculará e exibirá o coeficiente de correlação entre os dois conjuntos de dados.

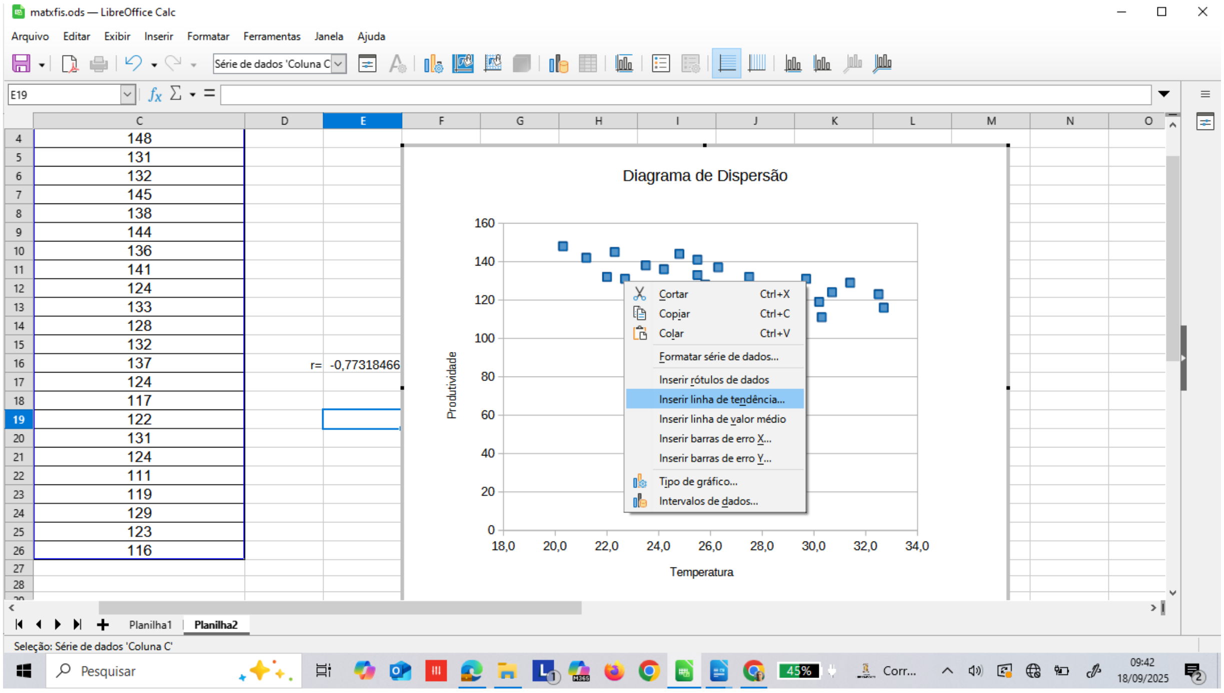
Figura 8: Visualização do coeficiente de correlação na planilha



Fonte: O autor

Uma vez encontrado o coeficiente de correlação entre temperatura e produtividade, vamos utilizar o LibreOffice Calc para determinar a reta de regressão que melhor se ajusta aos dados. Clicamos duas vezes sobre o gráfico para abrí-lo para edição e após, clicamos sobre um dos pontos de dados e no menu, escolhemos inserir linha de tendência.

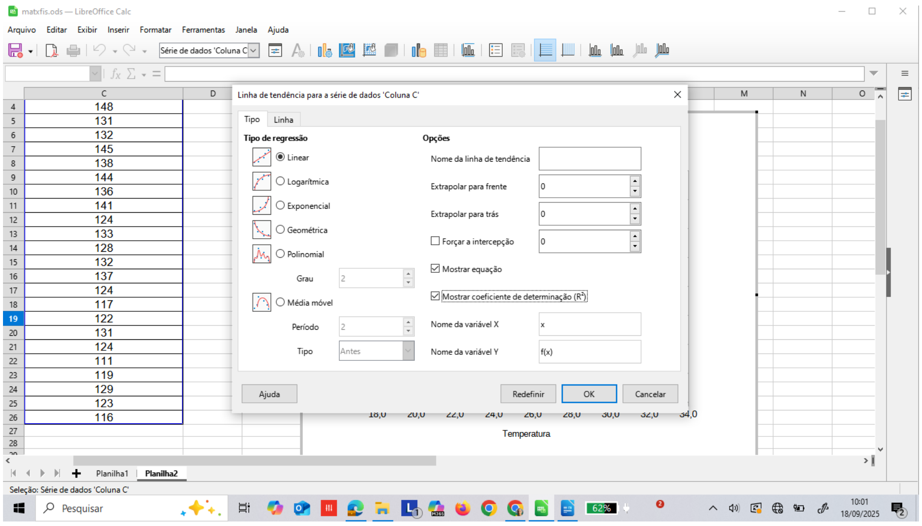
Figura 9: Inserção da linha de tendência na planilha



Fonte: O autor

Clicando em inserir linha de tendência, uma janela será exibida e escolhemos tipo de regressão linear e marcamos as opções de mostrar equação e mostrar coeficiente de determinação (R^2).

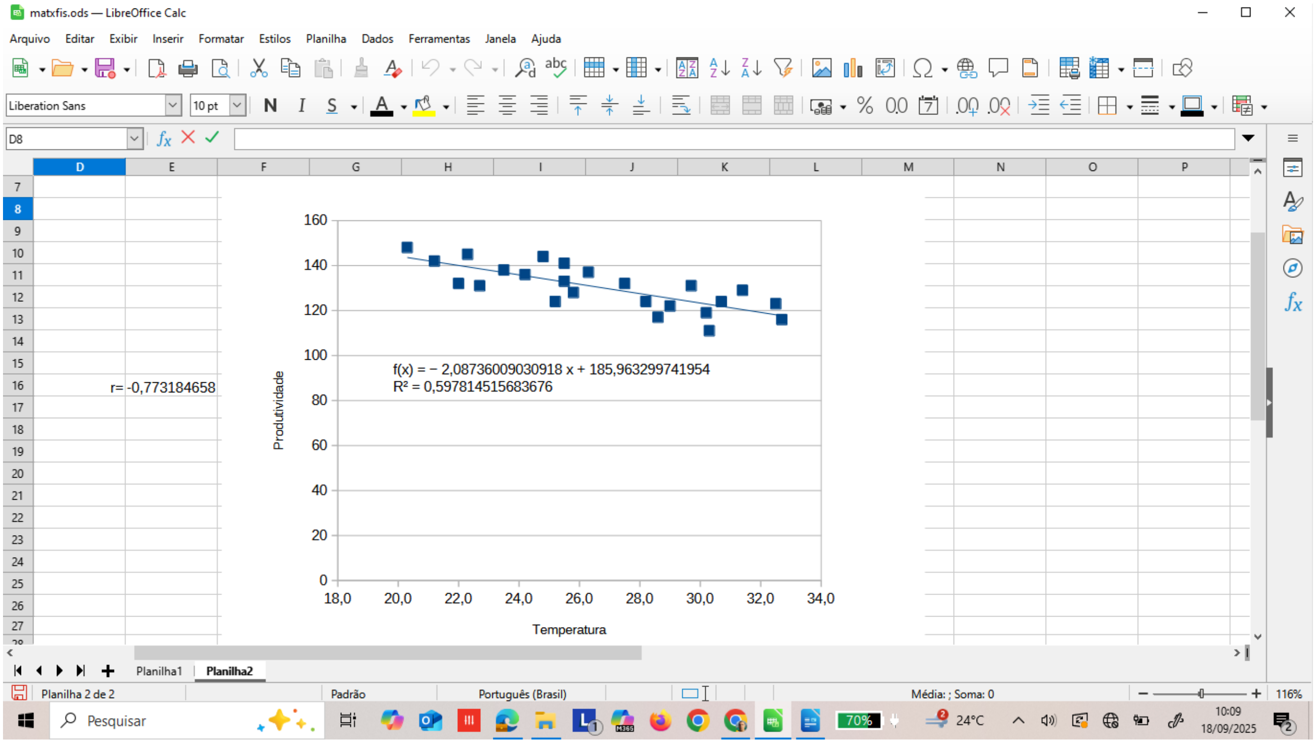
Figura 10: Inserção do coeficiente de determinação na planilha



Fonte: O autor

Ao clicar em OK, o gráfico surgirá com a reta de regressão linear e o coeficiente de determinação na tela.

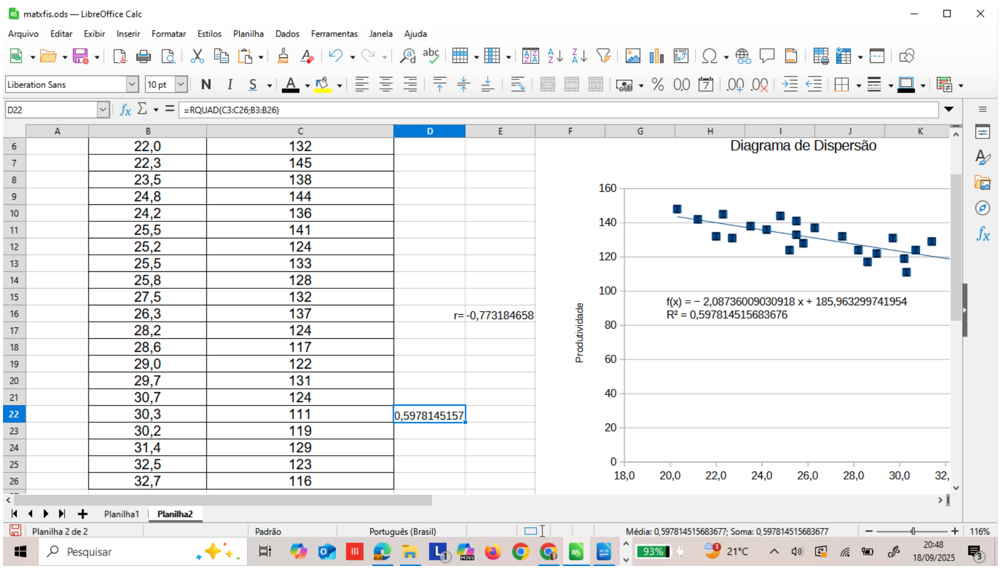
Figura 11: Exibição da reta de regressão e do coeficiente de determinação na planilha



Fonte: O autor

O coeficiente de determinação também pode ser encontrado através da função RQUAD. Na célula escolhida digitamos: $=RQUAD(Dados\ y\,;Dados\ x)$, onde os dados y estão no intervalo (C3:C26) e os dados x estão no intervalo (B3:B26). No caso, a função a ser digitada é $=RQUAD(C3:C26\,;B3:B26)$.

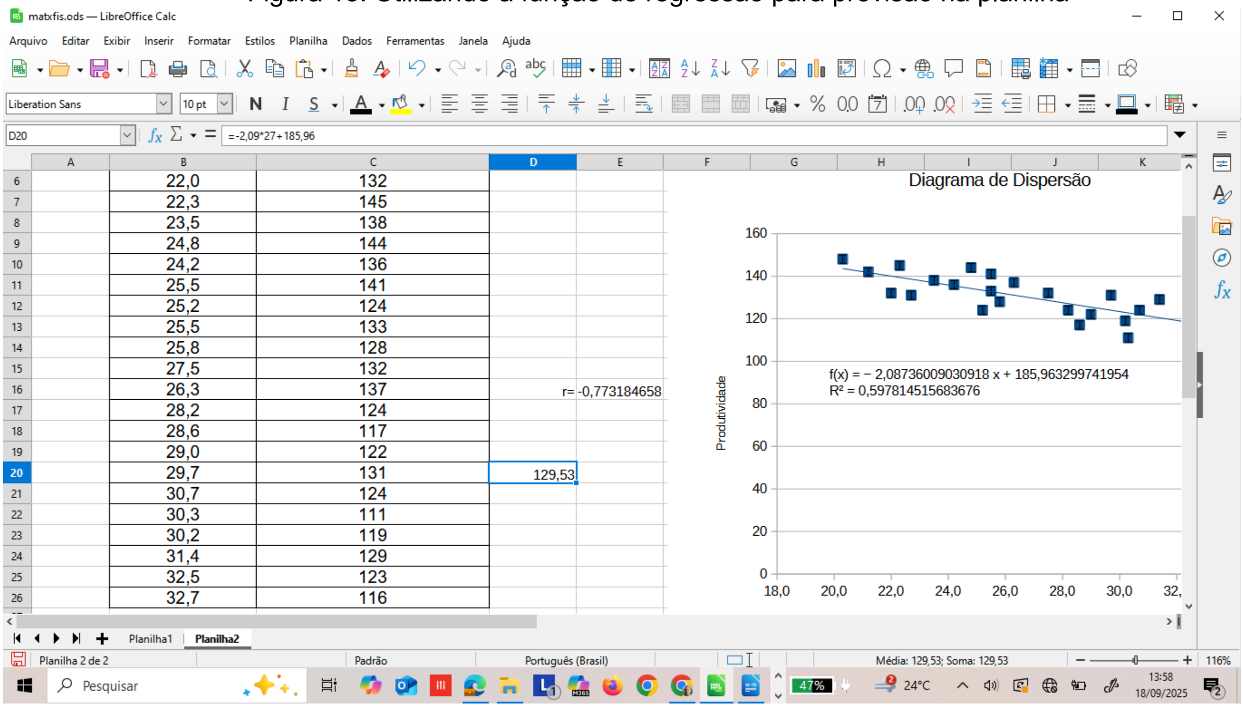
Figura 12: Outra forma de exibir o coeficiente de determinação na planilha



Fonte: O autor

Ajustando os coeficientes angular, linear e de determinação para duas casas decimais, temos a função de regressão igual a $y=-2{,}09\,x+185{,}96$ e $R^2=0{,}60$ o que indica que $60\,\%$ da variação dos dados é explicado pelo modelo. Os $40\,\%$ que não são explicados pelo modelo são devidos a variáveis que não foram consideradas. Observando a função de regressão, vemos que o aumento de $1°$ na temperatura causa uma diminuição de 2,09 unidades produzidas. Através do modelo de regressão encontrado, podemos estimar a produção para uma temperatura de $27°$. Utilizando a planilha, escolhemos uma célula onde queremos que apareça o resultado e digitamos = $-2{,}9\cdot 27+185{,}96$. Teclando enter, o resultado aparecerá na tela.

Figura 13: Utilizando a função de regressão para previsão na planilha



Fonte: O autor

De acordo com a reta de regressão, a produção estimada para a temperatura de $27°$ é $129{,}53$.

Folha de tarefas 3

1) Os dados a seguir correspondem à variável renda familiar e gasto com alimentação (em unidades monetárias) para uma amostra de 25 famílias.

Renda Familiar (x)	Gasto com Alimentação (y)
3	1,5
5	2,0
10	6,0
10	7,0
20	10,0
20	12,0
20	15,0
30	8,0
40	10,0
50	20,0
60	20,0
70	25,0
70	30,0
80	25,0
100	40,0
100	35,0
100	40,0
120	30,0
120	40,0
140	40,0
150	50,0
180	40,0
180	50,0
200	60,0
200	50,0

Utilizando a planilha LibreOffice Calc, faça os itens abaixo:

a) Organize os dados na planilha.

b) Represente os dados em um gráfico de dispersão.

c) Calcule o coeficiente de correlação.

d) Determine a reta de regressão linear e represente a reta no gráfico de dispersão.

e) Calcule o coeficiente de determinação e interprete o valor encontrado.

f) Use a reta de regressão encontrada para prever o gasto com alimentação de uma família com renda de 110 unidades monetárias.

g) De acordo com o modelo de regressão linear encontrado, o aumento de uma unidade monetária na renda familiar gera um aumento de quantas unidades monetárias no gasto com alimentação?

4° Encontro

Os estudantes são estimulados a utilizar o GeoGebra que é um software de matemática dinâmico e de código aberto, que reúne recursos geométricos, algébricos, estatísticos, e que possibilita a confecção de cálculos e o uso de planilhas através de várias janelas de exibição. Segundo Pazim e Fonseca (2022), o diferencial do GeoGebra é possuir uma interface totalmente dinâmica por meio de “botões” que permitem a omissão/exibição gráfica dos objetos explorados, facilitando a apresentação dos dados.

Por ser dinâmico, o estudante pode observar a mudança na função de regressão linear, no coeficiente de correlação e no coeficiente de determinação, logo após alguns dados serem alterados. O GeoGebra pode ser baixado no computador ou utilizado online e é uma poderosa ferramenta para exploração e visualização de conceitos matemáticos e o fato de muitos professores de matemática estarem familiarizados com este software, possibilita a aplicação da atividade de modelagem por regressão linear sem a necessidade de maiores capacitações. Os estudantes utilizaram o GeoGebra para criar um conjunto de dados, gerar um gráfico de dispersão, realizar a regressão linear e analisar os resultados. A dinâmica da aula é muito parecida com a do 3º Encontro.

Plano de Aula do 4° Encontro

- **Conteúdo a ser trabalhado:**
  Apresentação do software GeoGebra;
  Inserção de dados na planilha do Geogebra;
  Construção de Gráfico de Dispersão;
  Função estatística do Geogebra;
  Construção da Reta de Regressão Linear Simples usando a planilha e usando a janela de álgebra do GeoGebra.
- **Objetivos:**
  Inserir dados na planilha eletrônica do GeoGebra;
  Aplicar funções estatísticas para análise dos dados inseridos;
  Construir uma Reta de Regressão Linear Simples através da planilha eletrônica e através da inserção de pontos na janela de álgebra;
  Analisar os resultados obtidos com o auxílio do software dinâmico GeoGebra.
- **Habilidades relacionadas (BNCC):**
  EM13MAT302 - Resolver e elaborar problemas cujos modelos são as funções polinomiais de 1º e 2º graus, em contextos diversos, incluindo ou não tecnologias

digitais.

EM13MAT406 - Construir e interpretar tabelas e gráficos de frequências com base em dados obtidos em pesquisas por amostras estatísticas, incluindo ou não o uso de softwares que inter-relacionem estatística, geometria e álgebra.

EM13MAT501 - Investigar relações entre números expressos em tabelas para representá-los no plano cartesiano, identificando padrões e criando conjecturas para generalizar e expressar algebricamente essa generalização, reconhecendo quando essa representação é de função polinomial de 1º grau.

EM13CO09 - Identificar tecnologias digitais, sua presença e formas de uso, nas diferentes atividades no mundo do trabalho.

EM13CO11 - Criar e explorar modelos computacionais simples para simular e fazer previsões, identificando sua importância no desenvolvimento científico.

EM13CO12 - Produzir, analisar, gerir e compartilhar informações a partir de dados, utilizando princípios de ciência de dados.

- **Metodologia:**

Assim como no 3° encontro, uma situação-problema foi usada com o objetivo de ambientar os estudantes com a estrutura do software dinâmico GeoGebra. Como a instituição possuía acesso satisfatório à internet, foi utilizado o GeoGebra Classic online através do site https://www.geogebra.org/classic?lang=pt_PT.

Situação-problema

Consulta feita ao site da imobiliária QuintoAndar, no dia 05/10/2025, verificou o preço de venda de 08 apartamentos situados na rua Campos Sales, no bairro da Tijuca, Rio de Janeiro. A tabela abaixo mostra a área do imóvel e o valor de venda:

Tabela 5: Área x Valor de venda

Área (m^2)	Valor de venda (R$)
125	850000,00
125	980000,00
55	580000,00
67	595000,00
104	850000,00
149	1890000,00
150	1890000,00
35	300000,00

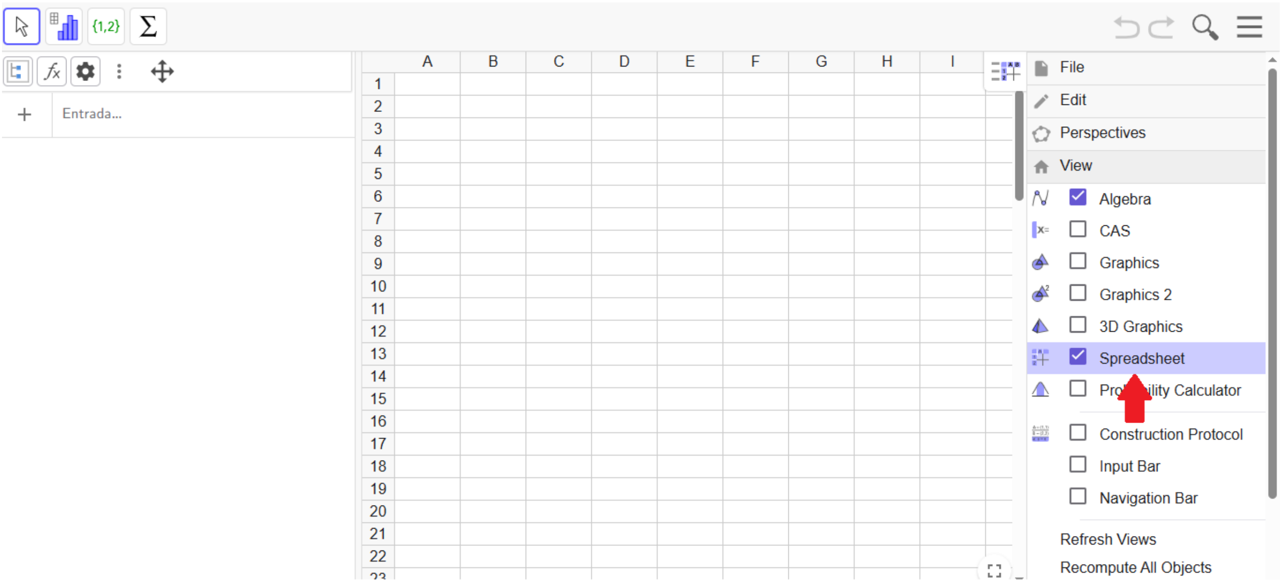
Após abrir a versão online do GeoGebra Classic, no menu superior, clique nas três barras horizontais que estão à direita.

Figura 14: Menu do GeoGebra

Fonte: O autor

No menu que se abre, escolha a opção de exibição da planilha (Spreadsheet) e a planilha aparecerá.

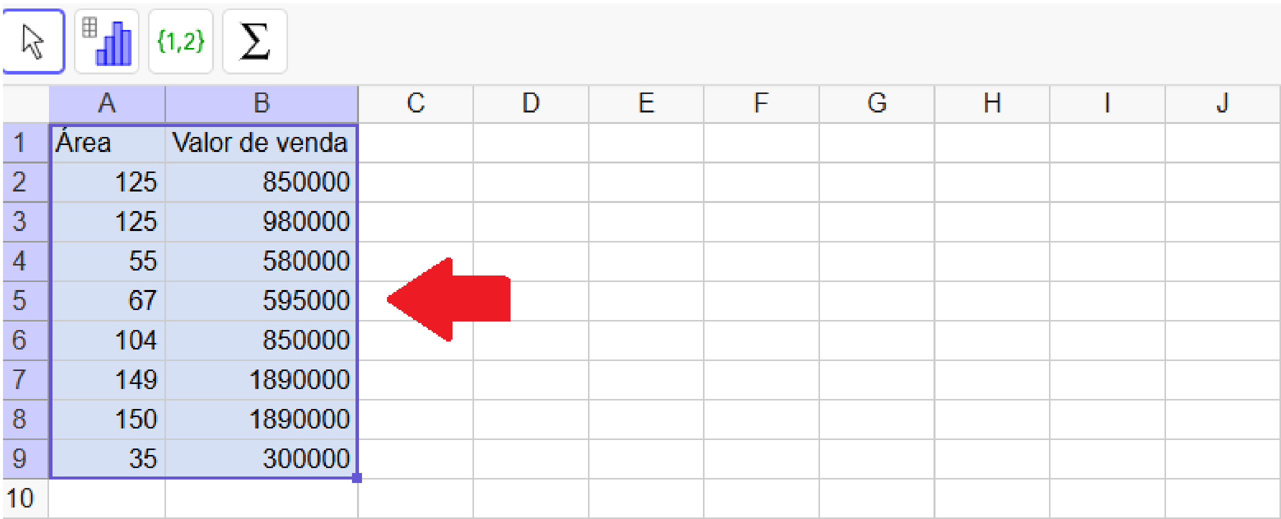
Figura 15: Exibição da planilha no GeoGebra



Fonte: O autor

O próximo passo é inserir os dados na planilha. A área do imóvel será a variável independente (x) e o valor de venda é a variável dependente (y). Depois, selecionamos todos os valores da planilha e para isso, deve-se apertar e manter pressionado o botão esquerdo do mouse, arrastando o cursor por todas as células que contêm os dados.

Figura 16: Inserção de dados na planilha do GeoGebra

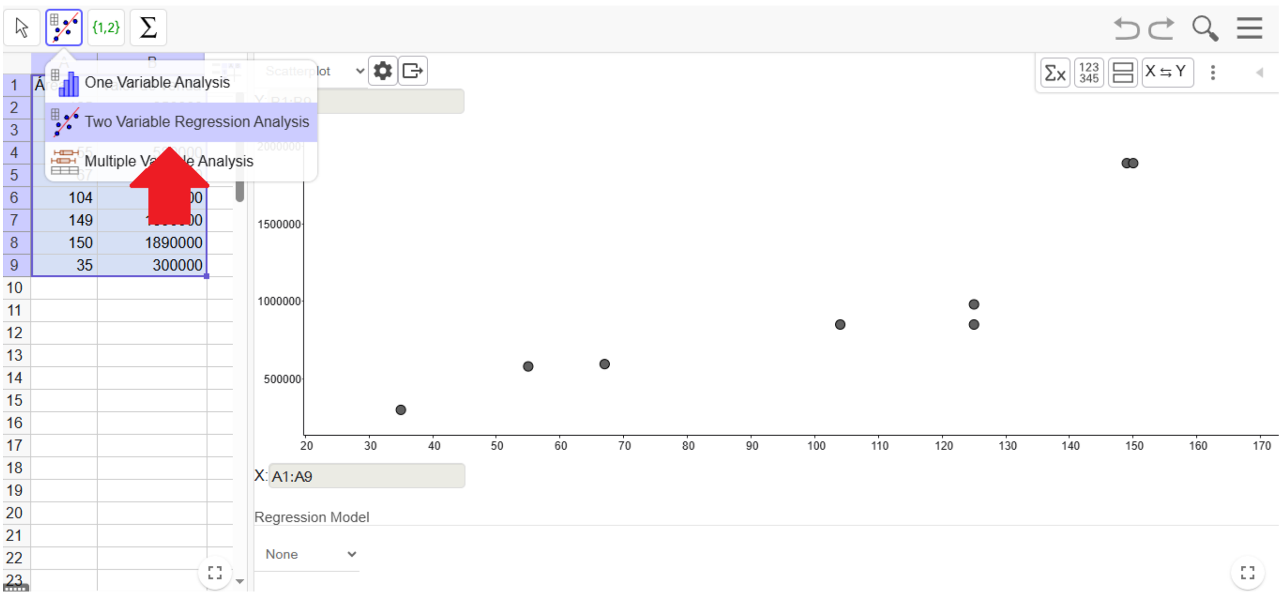


	A	B
1	Área	Valor de venda
2	125	850000
3	125	980000
4	55	580000
5	67	595000
6	104	850000
7	149	1890000
8	150	1890000
9	35	300000

Fonte: O autor

Na sequência, clicamos no segundo ícone do canto superior esquerdo da tela e escolhemos a análise bivariada (Two Variable Regression Analysis). Assim, aparecerá automaticamente na tela o gráfico de dispersão relativo aos dados analisados.

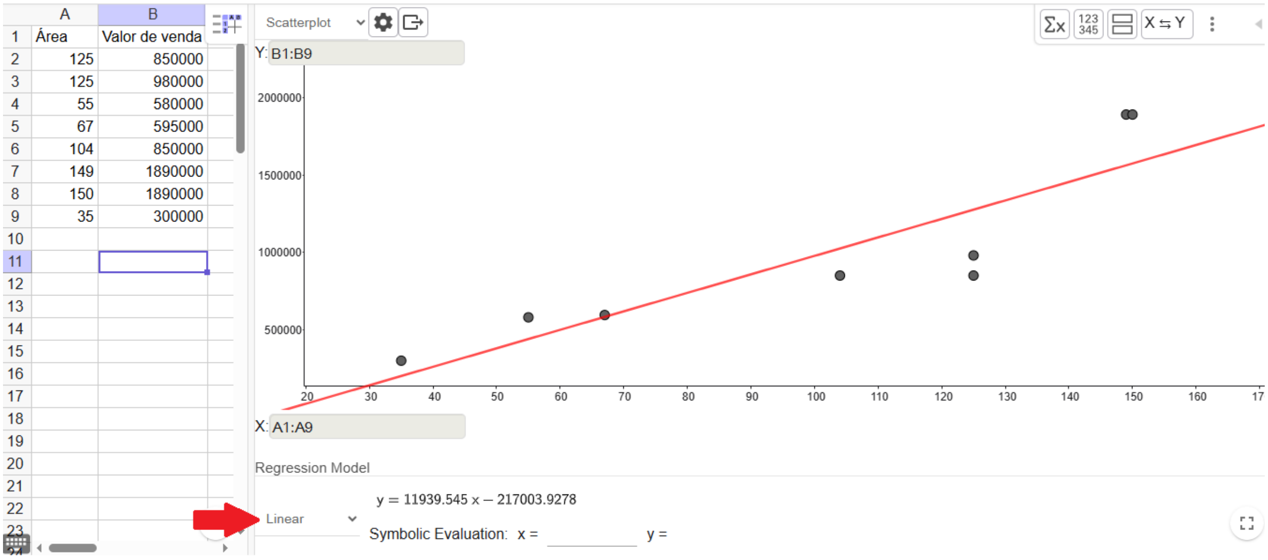
Figura 17: Exibição do gráfico de dispersão no GeoGebra



Fonte: O autor

Abaixo do gráfico de dispersão, temos a possibilidade de escolher o modelo de Regressão Linear (Regression Model) e com isso, a linha de tendência e a função de regressão aparecerão na tela.

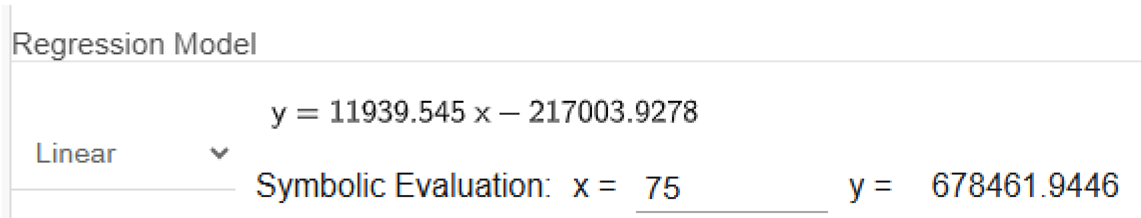
Figura 18: Inserção da linha de tendência no GeoGebra



Fonte: O autor

Abaixo da função de regressão, o GeoGebra nos possibilita estimar o valor de venda de acordo com a área do imóvel. Vamos ver, de acordo com o modelo, o valor de um apartamento de 75 m^2.

Figura 19: Fazendo previsão com o Geogebra

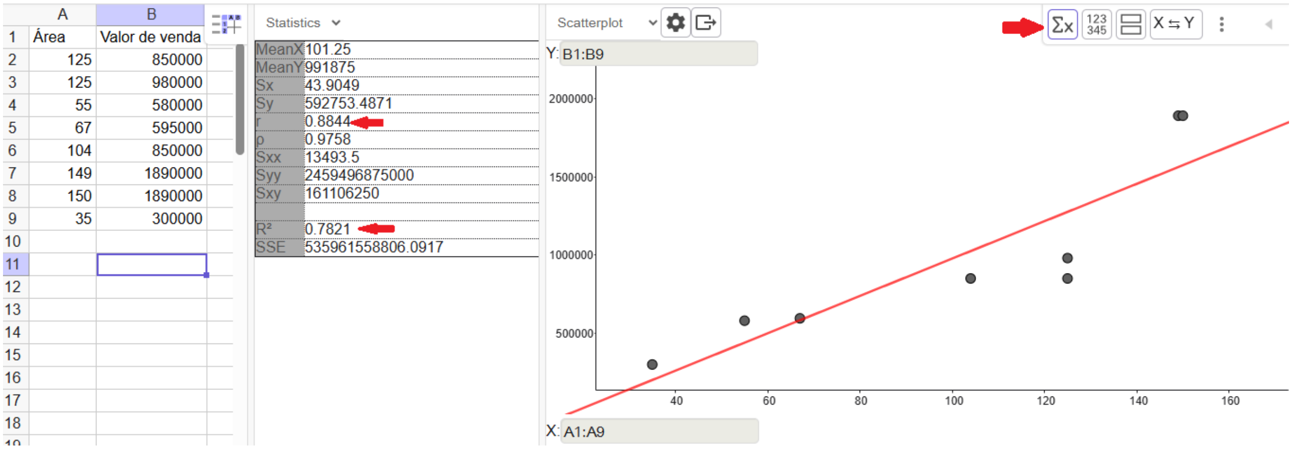


Fonte: O autor

De acordo com o modelo, um apartamento de 75 m^2 tem valor de venda de R$ 678.461,94.

Finalizando, podemos ver os valores estatísticos clicando no ícone $\sum x$, localizado acima do gráfico de dispersão, à direita.

Figura 20: Exibição dos dados estatísticos no GeoGebra

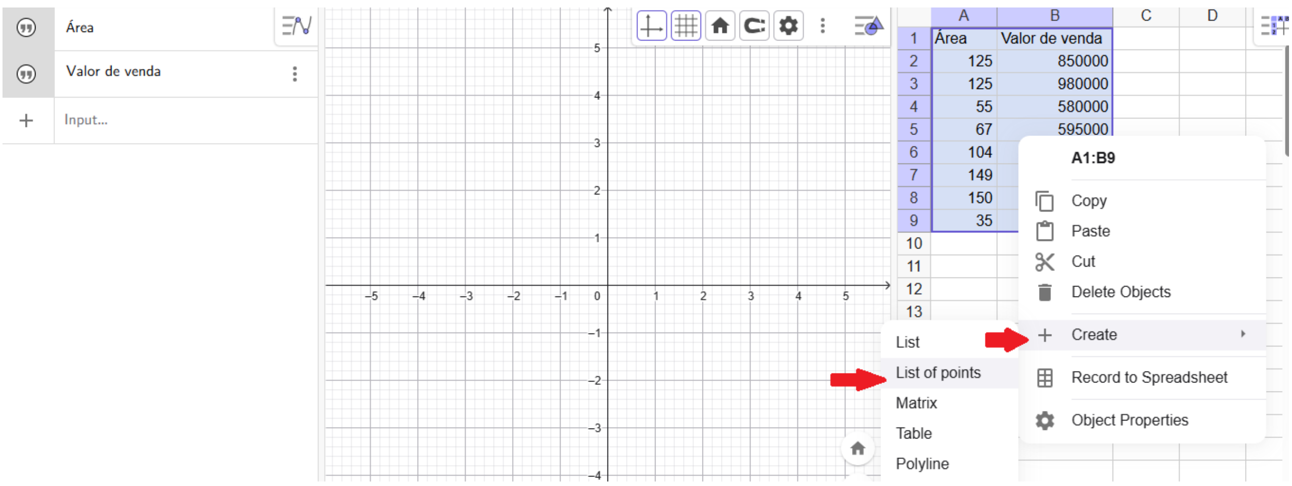


Fonte: O autor

Com a estatística do modelo exibida, verifica-se que o coeficiente de correlação (r) é igual a $0{,}8844$ e o coeficiente de determinação (R^2) é igual a $0{,}7821$.

Outra forma de criar um modelo de regressão linear com o uso do GeoGebra é fazendo uma lista de pontos. Essa lista será formada pelos pares ordenados (x, y) que representam os dados observados. Existem várias formas de criar uma lista de pontos, de acordo com a forma de inserção dos dados. Aproveitando que os dados do exemplo anterior já estão inseridos na planilha, vamos selecioná-los, apertando e mantendo pressionado o botão esquerdo do mouse, arrastando o cursor por todas as células que contêm os dados e na sequência, clicamos com o botão direito do mouse para escolher as opções criar (create) e lista de pontos (list of points).

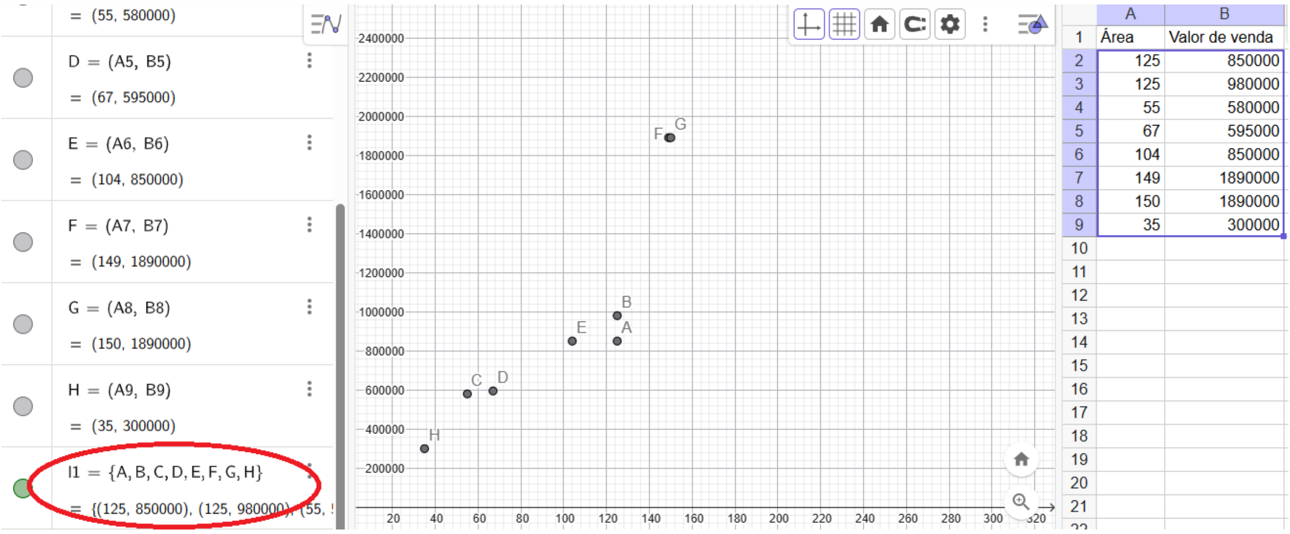
Figura 21: Seleção dos dados da planilha



Fonte: O autor

Assim, o GeoGebra irá gerar uma lista na janela de álgebra com todos os pontos selecionados e esses pontos também aparecerão na janela de visualização (gráfico 2D). Cabe destacar que um ajuste nos valores de máximo e mínimo dos eixos x e y foi feito para que as coordenadas aparecessem na tela, pois as ordenadas possuíam valores altos. Esse ajuste é feito nas configurações da janela de visualização, no ícone

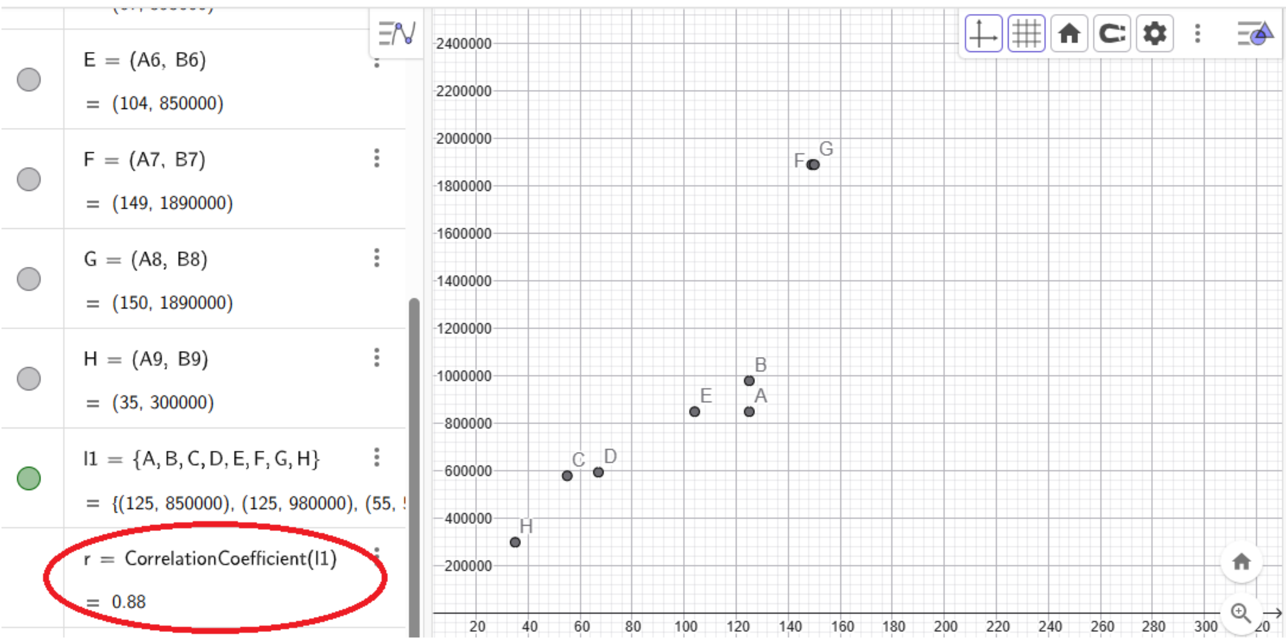
Figura 21: Lista de pontos na janela de álgebra e na janela de visualização



Fonte: O autor

Para encontrarmos o coeficiente de correlação linear (r) entre os valores, usamos na janela de álgebra, o comando $CoeficienteCorrelacao(l1)$ (CorrelationCoefficient).

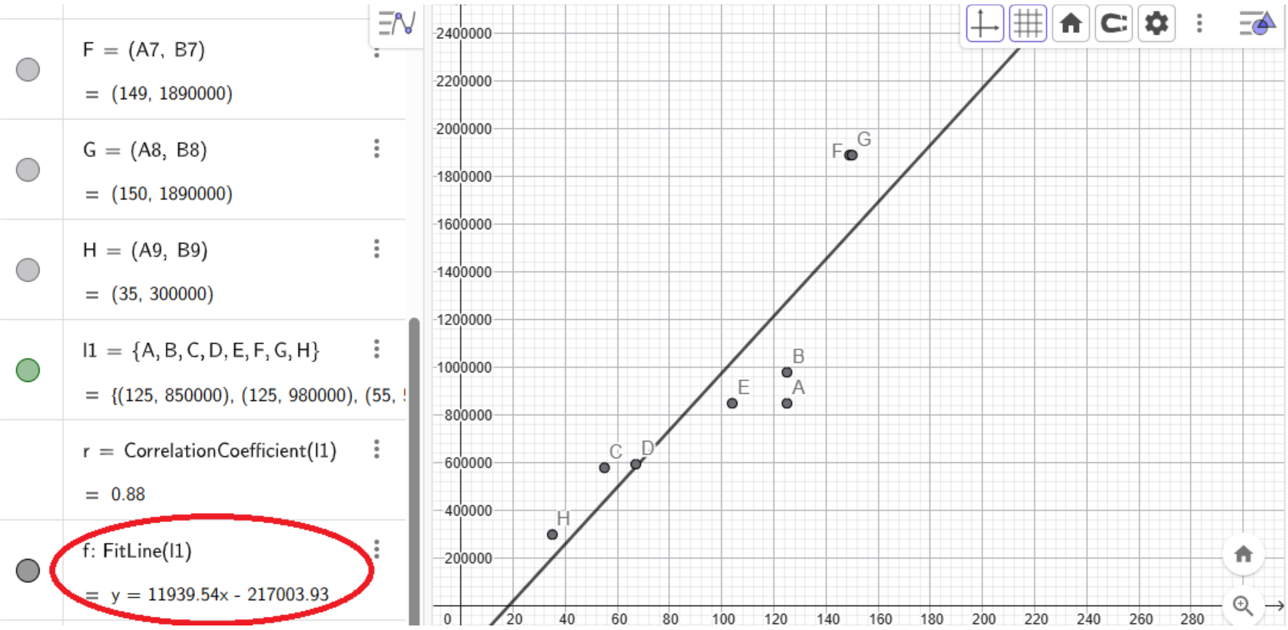
Figura 22: Exibição do coeficiente de correlação linear no GeoGebra



Fonte: O autor

O valor de $r = 0{,}88$ indica uma boa correlação linear entre os dados amostrais e sendo assim, para encontrarmos a reta de regressão linear que melhor se ajusta aos pontos, devemos usar o comando $Regressãolinear(l1)$ (FitLine), que retornará a reta de regressão na janela de visualização e a função de regressão na janela de álgebra.

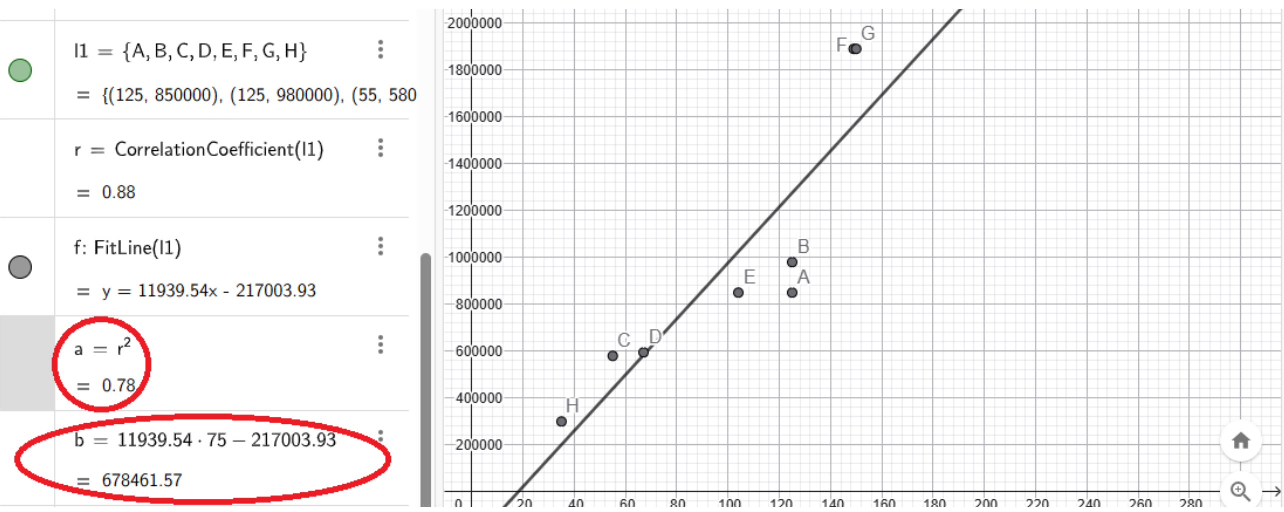
Figura 23: Exibição da função de regressão e da reta de regressão no GeoGebra



Fonte: O autor

O Coeficiente de Determinação (R^2) é encontrado elevando o valor de r ao quadrado e com a função de regressão encontrada, podemos prever o valor de venda de acordo com a área do imóvel. Temos que $R^2 = 0{,}78$ e de acordo com o modelo, um apartamento com área de $75\,m^2$ tem valor de venda de R\$ $678.461{,}57$.

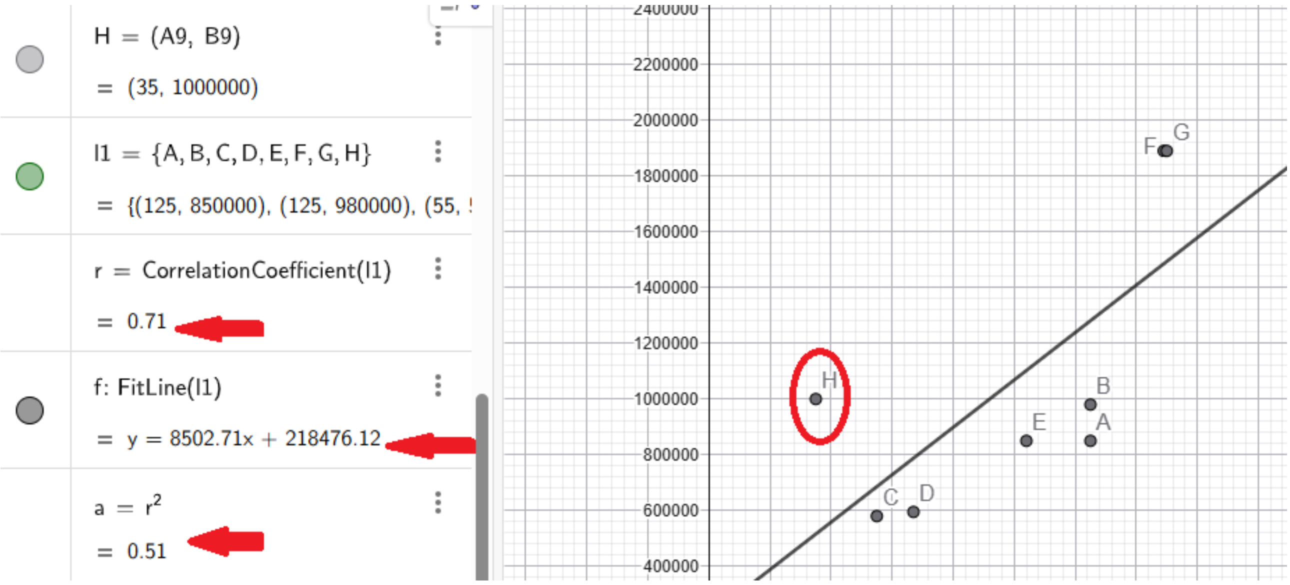
Figura 24: Exibição do coeficiente de determinação e da previsão de valor de venda no GeoGebra



Fonte: O autor

Essa segunda forma de apresentar a Regressão Linear com o uso do GeoGebra permite que o estudante verifique de forma dinâmica, as mudanças nos parâmetros do modelo e nos coeficientes de correlação linear e de determinação, quando dados atípicos (outliers) estão inseridos no conjunto de pontos da amostra, seja por erro de coleta ou eventos incomuns, causando distorções nos resultados e nas análises. Suponha que o imóvel de $35\,m^2$, esteja com valor de venda de R$ $1.000.000{,}00$. Clicando e arrastando o ponto H, vemos a mudança na reta de ajuste e nos coeficientes do modelo.

Figura 25: Mudança da reta de regressão e das estatísticas do modelo com a presença de outliers



Fonte: O autor

Fica claro que apenas um outlier causa mudanças significativas no modelo, como a diminuição da correlação entre as variáveis e a diminuição do Coeficiente de Determinação, o que causaria prejuízo na análise e nas previsões. Segue folha de tarefas do 4° Encontro.

Folha de tarefas 4

Exercício adaptado de Bussab e Morettin (2017)

1) Os dados abaixo referem-se a meses de experiência de dez digitadores e o número de erros cometidos na digitação de determinado texto.

Meses (x)	Erros (y)
1	30
2	28
3	24
4	20
5	18
6	14
7	13
8	10
9	7
10	6

Utilizando o GeoGebra, faça os itens abaixo:

a) Escolha uma forma de representar graficamente esse conjunto de dados.

b) Você considera que o coeficiente de correlação encontrado demonstra uma boa correlação entre as variáveis? Justifique sua resposta.

c) Determine a reta de regressão linear que melhor se ajusta aos dados da amostra.

d) Calcule o coeficiente de determinação e interprete o valor encontrado.

e) Use a função de regressão para prever a quantidade de erros de um digitador com 5 anos de experiência. Explique, com suas palavras, a diferença entre o valor encontrado pelo modelo e o valor coletado nos dados.

5° Encontro

Os estudantes entram em contato com o Python, que é uma linguagem de programação de alto nível[2], amplamente utilizada por ser considerada simples até para iniciantes. A linguagem foi criada por um programador de computador holandês chamado Guido Van Rossum, sendo a primeira versão do código apresentada em 1991 e nos dias de hoje, a comunidade ativa do Python possui vários desenvolvedores de suporte ao redor do mundo. Muitas IAs são programadas em Python devido a sua flexibilidade e por possuir várias bibliotecas[3] e frameworks[4] disponíveis.

As atividades de programação em Python propostas para este encontro devem ser realizadas online, através da plataforma Google Colab (Google Colaboratory) que é gratuita e permite criar e executar códigos em Python diretamente do navegador, sem a necessidade de baixar programas nem tampouco configurar os dispositivos utilizados. O uso do Google Colab traz diversas facilidades para o usuário, como integração com o Google Drive, bibliotecas de uso em Python pré-instaladas, possibilidade de combinar código executável, texto explicativo, equações, gráficos e visualizações em um documento único, entre outros. Cabe dizer que para utilizar a plataforma é necessário possuir uma conta Google.

Segundo Lins e Souza (2023), a BNCC através da competência denominada Cultura Digital, estimula o uso de atividades que envolvem linguagem de programação e uso de algoritmos nas aulas de matemática para auxiliar no processo de ensino/aprendizagem. Neste encontro, os estudantes aprendem a importar as bibliotecas utilizadas no modelo de regressão linear simples, preparar os dados, separá-los em dados de treino e dados de teste (os dados de treino servem para treinar o modelo e os dados de teste servem para avaliar a performance do modelo), fazer uma regressão linear no Python, analisar o modelo e visualizar os resultados. A folha de tarefas deve ser disponibilizada aos estudantes após a aula teórica, realizada com o apoio do projetor. São utilizados 30 minutos de apresentação da linguagem através de exemplos e o tempo restante é utilizado para que os estudantes possam executar as atividades de modelagem de regressão linear em Python. O objetivo principal desse encontro é apresentar a

2 Linguagem de programação projetada para ser fácil de entender e escrever, usando sintaxe próxima da linguagem humana.
3 Coleção de códigos utilizados com frequência pelos desenvolvedores/programadores.
4 Conjunto de códigos prontos que servem como estrutura para auxiliar no desenvolvimento da programação.

linguagem para os estudantes e mostrar que com poucos comandos, um modelo de regressão pode ser realizado.

Plano de Aula do 5° Encontro

- **Conteúdo a ser trabalhado:**
  Ambientação na plataforma Google Colab;
  Noções da linguagem de programação Python;
  Principais bibliotecas utilizadas em Python;
  Inserção dos dados e obtenção do coeficiente de correlação linear de Pearson;
  Divisão dos dados em treinamento e teste;
  Treinamento do modelo e previsão de valores;
  Avaliação do modelo e visualização dos dados;
  Análise dos resultados.

- **Objetivos:**
  Reconhecer a plataforma Google Colab como um ambiente de programação em Python;
  Identificar as principais bibliotecas utilizadas para a construção de um modelo de regressão Linear Simples em Python;
  Aplicar os códigos usados para a criação de um modelo simples de regressão;
  Preparar e programar o modelo;
  Mostrar e analisar os resultados obtidos.

- **Habilidades relacionadas (BNCC):**
  EM13MAT405 Utilizar conceitos iniciais de uma linguagem de programação na implementação de algoritmos escritos em linguagem corrente e/ou matemática.
  EM13CO04 Reconhecer o conceito de metaprogramação como uma forma de generalização na construção de programas, permitindo que algoritmos sejam entrada ou saída para outros algoritmos;

  EM13CO11 - Criar e explorar modelos computacionais simples para simular e fazer previsões, identificando sua importância no desenvolvimento científico;

EM13CO12 - Produzir, analisar, gerir e compartilhar informações a partir de dados, utilizando princípios de ciência de dados;

EM13CO15 - Analisar a interação entre usuários e artefatos computacionais, abordando aspectos da experiência do usuário e promovendo reflexão sobre a qualidade do uso dos artefatos nas esferas do trabalho, do lazer e do estudo.

- **Metodologia:**

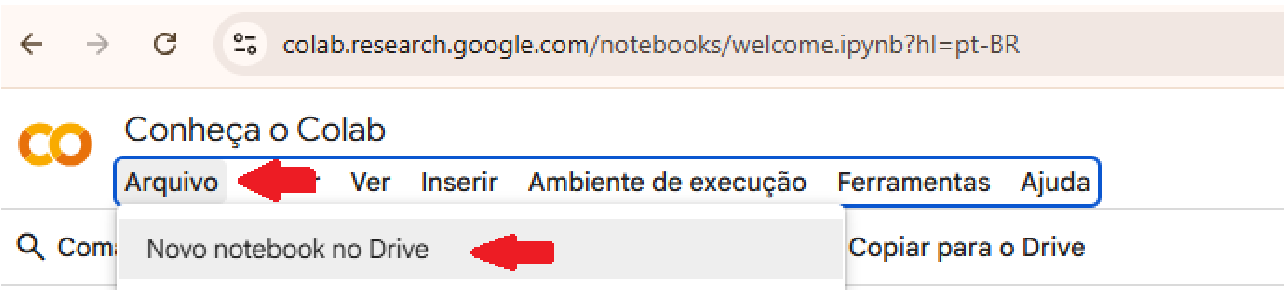
O encontro inicia-se com a apresentação da plataforma Google Colab que será o ambiente onde a programação em Python será realizada. O site utilizado foi https://colab.research.google.com/notebooks/welcome.ipynb?hl=pt-BR.

Figura 26: Tela inicial do Google Colab

Fonte: O autor

O Google Colab possui uma interface amigável e autoexplicativa, onde o estudante consegue navegar e conhecer todas as possibilidades que a plataforma pode oferecer. A inteligência artificial do Google (Gemini) está incorporada à plataforma e pode auxiliar nas dúvidas que o usuário tiver em relação aos códigos ou comandos a serem utilizados. Depois da breve apresentação, o estudante deve clicar em arquivo e novo notebook para criar uma janela para inserção dos códigos de programação.

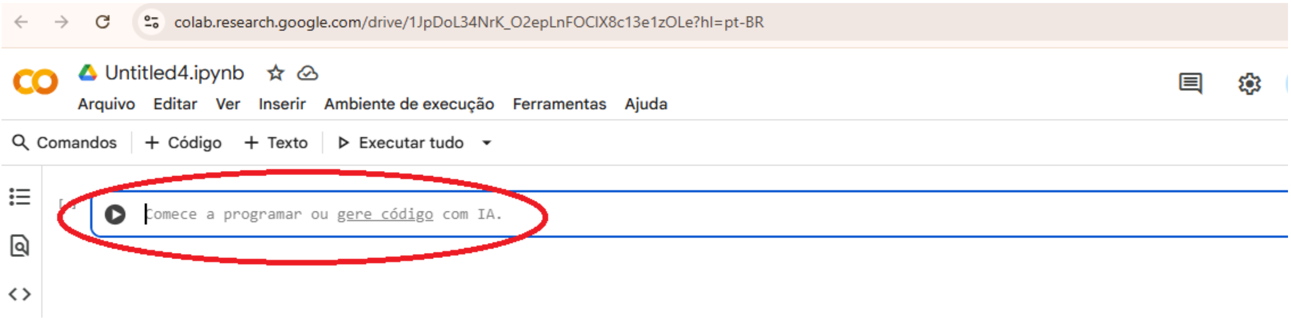
Figura 27: Abrindo a janela de inserção de códigos de programação



Fonte: O autor

Com a janela de programação aberta, o estudante pode começar a digitar os códigos e iniciar a programação.

Figura 28: Janela de inserção de códigos de programação



Fonte: O autor

Antes de iniciar com a inserção dos códigos para a realização do modelo de regressão linear, é importante apresentar algumas bibliotecas de ciência de dados e aprendizado de máquina que são utilizadas no Python. O uso dessas bibliotecas evita que o desenvolvedor tenha que escrever os códigos partindo do início. Segundo Piaz (2023), são elas:

- NumPy - Base para computação numérica em Python. Oferece arrays (vetores e matrizes) multidimensionais eficientes e funções matemáticas de alto nível. Para maiores informações acessar https://numpy.org/.

- Pandas - Principal biblioteca para manipulação e análise de dados. Fornece estruturas de dados flexíveis e eficientes como DataFrame (tabela) e Series (coluna). Para maiores informações acessar https://pandas.pydata.org/.

- Matplotlib - Biblioteca fundamental para a criação de gráficos e visualizações estáticas de alta qualidade em 2D e algumas em 3D. Para maiores informações acessar https://matplotlib.org/.

- Seaborn - Baseada no Matplotlib, fornece uma interface de alto nível para desenhar gráficos estatísticos atraentes e informativos. Para maiores informações acessar https://seaborn.pydata.org/.

Cabe ainda destacar as seguintes bibliotecas:

- SciPy - Extensão do NumPy, oferece módulos para otimização, álgebra linear, integração, interpolação, funções especiais, processamento de sinais e imagens. Para maiores informações acessar https://scipy.org/.
- Scikit-learn - Biblioteca essencial para Machine Learning "clássico", focada em simplicidade e eficiência. Utilizada para Classificação, regressão, clustering, pré-processamento de dados e validação cruzada. Para maiores informações acessar https://scikit-learn.org/.
- Statsmodels – Biblioteca que fornece classes e funções para a estimação de diversos modelos estatísticos, bem como para a realização de testes e exploração de dados estatísticos. É frequentemente vista como uma ferramenta que complementa o Scikit-learn ao oferecer uma abordagem mais focada em inferência estatística. Para maiores informações acessar https://statsmodels.org/.

Como nos últimos encontros, uma situação-problema foi o ponto de partida para o início da atividade.

Situação-problema

Um estudo foi realizado para mensurar o Produto Interno Bruto (PIB) de um país da América do Sul e o Consumo das Famílias. A tabela mostra os valores dos últimos 15 trimestres, em milhões de dólares americanos.

Tabela 6: Consumo das Famílias x PIB do país

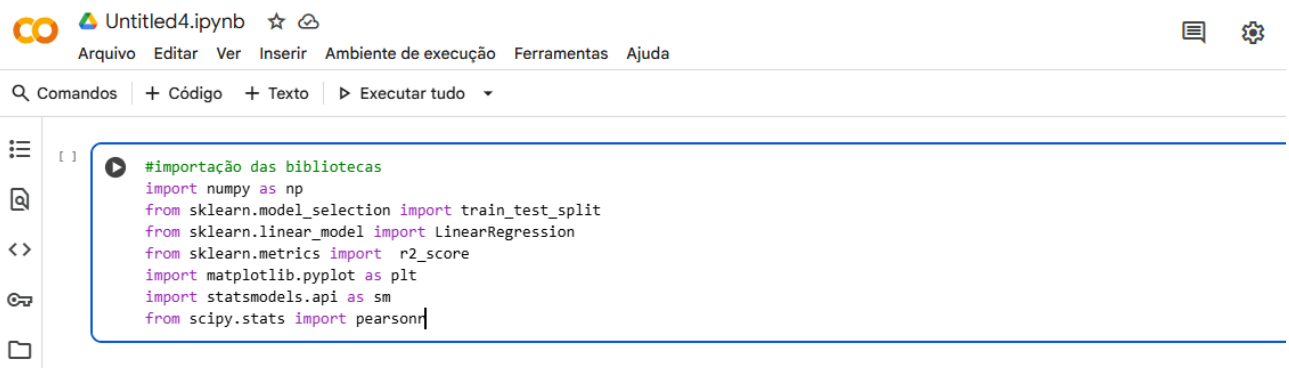
Consumo das Famílias (x)	PIB do País (y)
1275,60	1605,50
1312,00	1655,00
1282,90	1625,90
1295,40	1648,10
1320,80	1681,50
1356,10	1728,90
1325,00	1698,10
1337,90	1721,50
1364,00	1756,00
1401,50	1805,90
1370,10	1777,50
1383,20	1801,10
1409,80	1837,00
1448,50	1887,50
1417,10	1859,00

Fonte: O autor

Verifique se um modelo de Regressão Linear pode ser bem ajustado aos dados apresentados, utilizando o Python.

Para começar a trabalhar, vamos importar as bibliotecas necessárias, pois são a base para um projeto de Regressão Linear e Aprendizado de Máquina em Python.

Figura 29: Importação das bibliotecas do Python



Fonte: O autor

Essas bibliotecas preparam o ambiente para carregar, modelar e avaliar os dados. Quando colocamos o símbolo # no Python, significa que tudo que for escrito naquela linha não será considerado como código de programação. Isso é um comentário de linha que serve para adicionar notas e explicações, tornando o código mais legível. O significado de cada linha está a seguir:

- import numpy as np - Utilizada para realizar todas as operações matemáticas e estatísticas necessárias no modelo. O apelido np é universalmente usado.

- from sklearn.model_selection import train_test_split - Esta função é usada para dividir o seu conjunto de dados em dois subconjuntos: treino (train) e teste (test).

- from sklearn.linear_model import LinearRegression - Importa a classe principal para construir o modelo de Regressão Linear Mínimos Quadrados Ordinários (MQO). Essa função implementa o algoritmo para encontrar os coeficientes (angular e linear) da reta que minimiza a soma dos quadrados dos erros.

- from sklearn.metrics import r2_score – Importa a função para calcular o Coeficiente de Determinação (R^2).

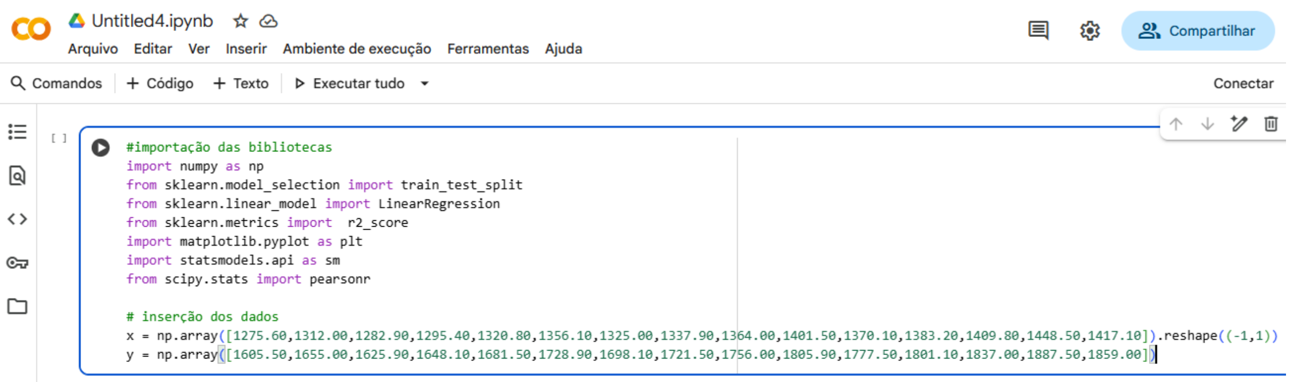
- import matplotlib.pyplot as plt – Como visto anteriormente, Matplotlib é a principal biblioteca para criar gráficos estáticos, animados e interativos em Python. O submódulo pyplot (plt) fornece uma interface semelhante ao MATLAB[5], essencial para plotar a nuvem de pontos (dispersão) e a reta de regressão ajustada para visualização.

- import statsmodels.api as sm - é usada para estimar modelos estatísticos, como a Regressão Linear, fornecendo um relatório estatístico completo.

- from scipy.stats import pearsonr - Importa a função específica para calcular o Coeficiente de Correlação de Pearson (r).

Depois de importar as bibliotecas necessárias, devemos inserir os dados. A inserção dos dados pode ser feita de diversas formas, através de uma lista de valores ou usando a biblioteca Pandas, que lê planilhas com muitos dados. Aqui foi usada uma lista de valores. Os valores x_i representam o consumo das famílias e os valores y_i representam o PIB.

Figura 30: Inserção dos dados na janela de programação



Fonte: O autor

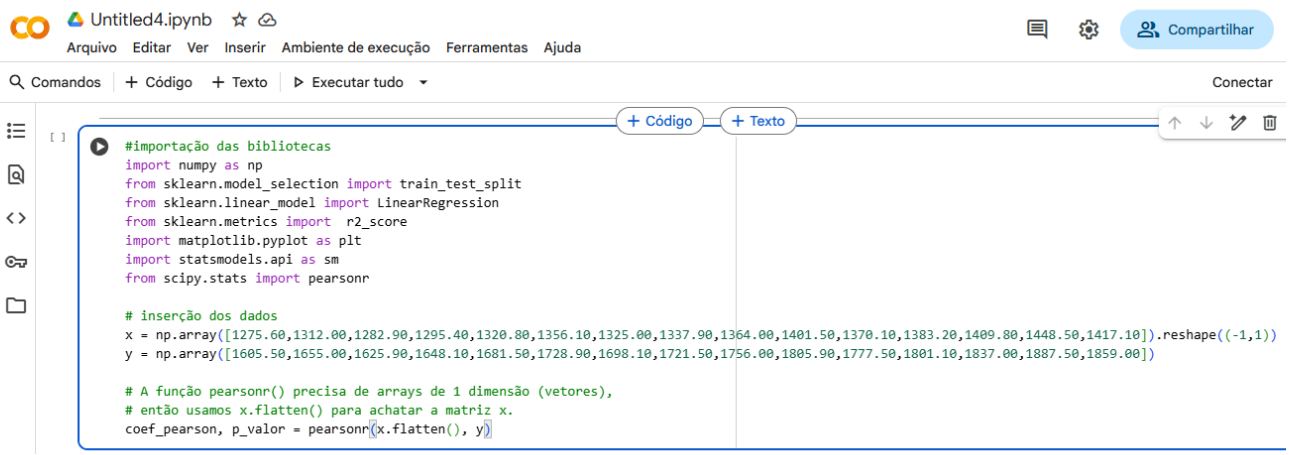
Em Python, a separação decimal se dá com a utilização do ponto (.) e os comandos $x=np.array$ e $y=np.array$ transformam a lista de valores em um Numpay Array, que é a estrutura de dados mais importante para computação numérica em Python.

5 Ambiente de programação e computação numérica de alto desempenho usado principalmente em engenharia e ciência para análise de dados, desenvolvimento de algoritmos e modelagem.

Na prática, é o primeiro passo para preparar os dados, convertendo-os em um formato otimizado para que os algoritmos de Aprendizado de Máquina possam processá-los de forma eficiente. O comando $reshape((-1,1))$ reorganiza a forma (as dimensões) do array (vetor/matriz) sem alterar os dados. O valor -1 informa ao NumPay para calcular o número de linhas automaticamente, garantindo que todos os elementos do array original sejam utilizados e o valor 1 obriga o array a ter exatamente uma coluna. Esse procedimento é necessário porque os modelos de Aprendizado de Máquina na biblioteca scikit-learn (onde a Regressão Linear está sendo implementada) requerem que a variável preditora (x) seja fornecida no formato de uma matriz bidimensional.

Na sequência, será colocado o comando para o cálculo do coeficiente de correlação de Pearson.

Figura 31: Comando para o cálculo do coeficiente de correlação

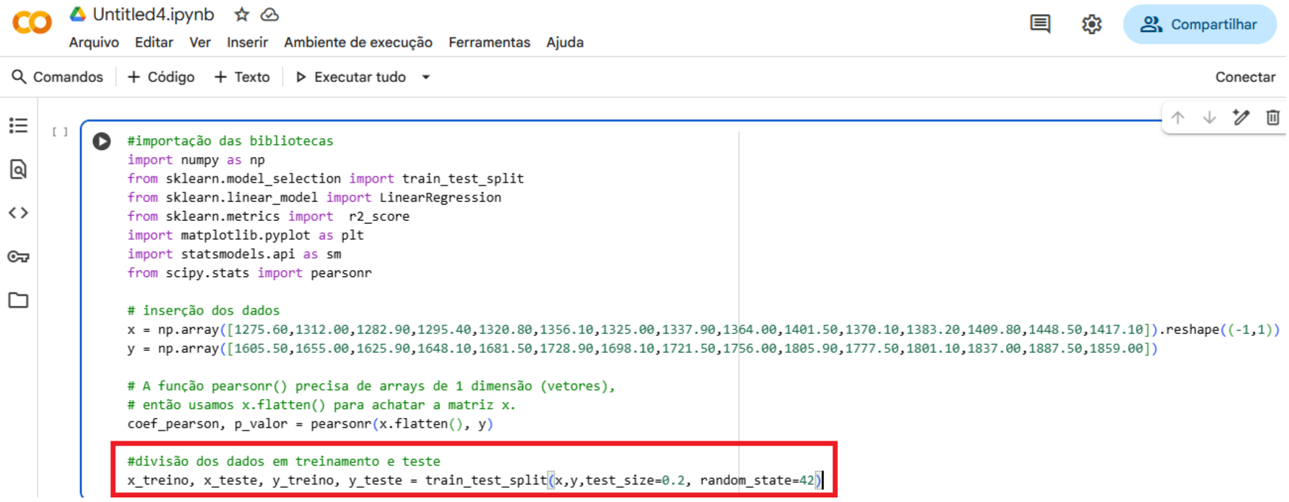


Fonte: O autor

Como a função pearsonr() precisa de arrays de 1 dimensão, o comando x.flatten() transforma a matriz x numa matriz de uma dimensão, para que o cálculo seja realizado.

Agora os dados devem ser separados em dados de treino e dados de teste. Os comandos estão na figura abaixo.

Figura 32: Separando os dados de treino e os dados de teste do modelo

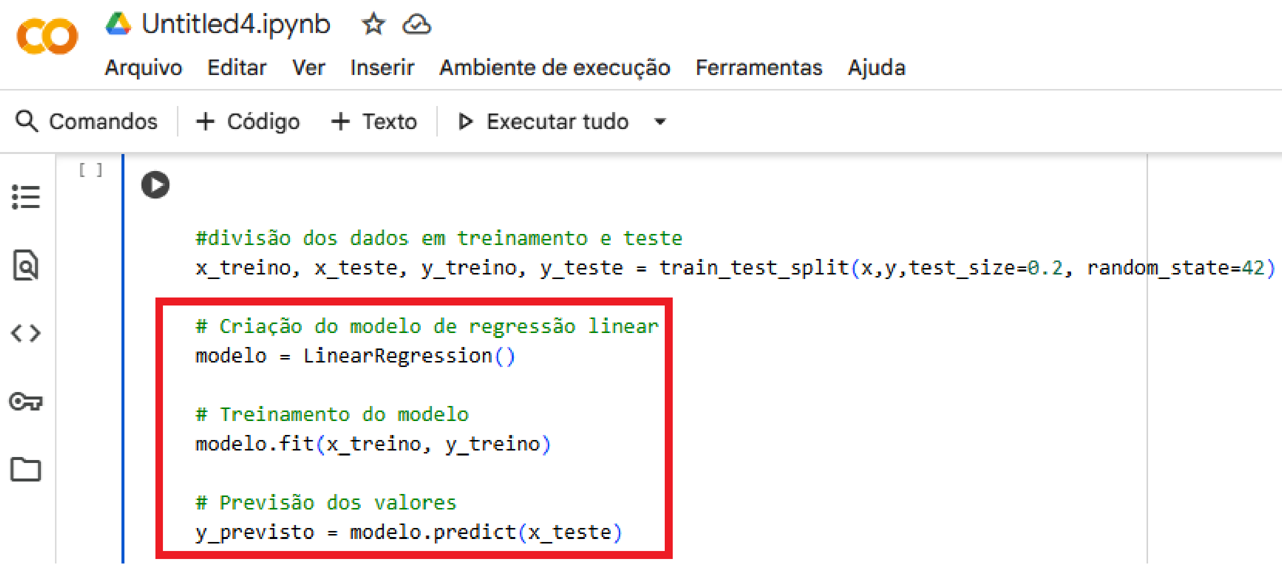


Fonte: O autor

A função train_test_split vem da biblioteca Scikit-learn e executa a divisão dos dados de forma aleatória e consistente. Test_size=0.2 indica que 20% dos dados serão separados para o conjunto de teste e o comando random_state=42 garante que a divisão aleatória dos dados seja reproduzível, ou seja, se o código for executado com random_state=42, o resultado da divisão dos dados será sempre o mesmo. O número 42 é aleatório, mas comum em ambientes de programação.

Os próximos passos são a criação do modelo de regressão linear, o treinamento do modelo e a previsão dos valores.

Figura 33: Criação e treinamento do modelo e previsão de valores



Fonte: O autor

O comando modelo = LinearRegression() é o algoritmo de Aprendizado de Máquina que cria um objeto de modelo de Regressão Linear com a função de treinar e fazer previsões. O comando modelo.fit(x_treino, y_treino) faz o modelo aprender as relações entre as variáveis, analisando os valores x_treino, ajustando os coeficientes angular e linear para criar uma reta que se aproxime ao máximo dos valores y_treino, com o objetivo de minimizar o erro entre o que o modelo prevê e os valores reais. Após a execução deste comando, o objeto modelo estará treinado e pronto para fazer previsões em novos dados que ele não conhece. O comando y_previsto = modelo.predict(x_teste) é a função que informa ao modelo já treinado, que deve aplicar o conhecimento adquirido para gerar resultados para novos dados.

Com o modelo já treinado, é o momento de visualizar os resultados. O comando print() é utilizado para exibir informações sobre os parâmetros, os comandos plt.scatter() e plt.plot() estão relacionados à criação de gráficos de dispersão, para a exposição da relação entre duas variáveis, o comando plt.legend coloca a legenda no gráfico e o comando plt.show() exibe o gráfico na tela.

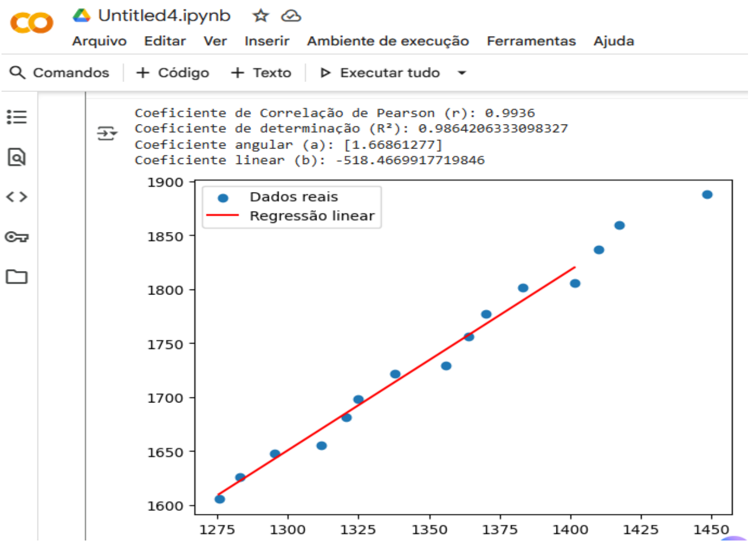
Figura 34: Comandos para a visualização dos resultados

```
# Mostrar o Coeficiente de Correlação de Pearson
print(f"Coeficiente de Correlação de Pearson (r): {coef_pearson:.4f}")

# Avaliação do modelo
print("Coeficiente de determinação (R²):", modelo.score(x_teste, y_teste))
print("Coeficiente angular (a):", modelo.coef_)
print("Coeficiente linear (b):", modelo.intercept_)

# Visualização dos dados
plt.scatter(x, y, label="Dados reais")
plt.plot(x_teste, y_previsto, color="red", label="Regressão linear")
plt.legend()
plt.show()
```

Fonte: O autor

Depois de todos os comandos e códigos inseridos, deve-se rodar o programa para a visualização do gráfico, dos coeficientes e dos parâmetros estatísticos.

Figura 35: Visualização dos resultados e da reta de regressão



Fonte: O autor

Arredondando os valores para duas casas decimais, observa-se que o Coeficiente de Correlação Linear de Pearson (r) é igual a $0{,}99$, o que mostra uma forte correlação linear positiva entre o consumo familiar e o PIB do país. O valor do Coeficiente Angular (a) é igual a $1{,}67$ e o valor do Coeficiente Linear (b) é $-518{,}47$, daí segue que a reta de Regressão Linear é da forma $y=1{,}67x-518{,}47$, com um ótimo Coeficiente de Determinação (R^2), demonstrando que 99% da variabilidade do consumo familiar é explicada pelo PIB do país.

Com os comandos e códigos explicados e usando como base a situação-problema apresentada, uma folha de tarefas foi disponibilizada para que os estudantes pudessem dar os primeiros passos em programação em Python, que é a linguagem mais utilizada no Aprendizado de Máquina.

Folha de tarefas 5

1) Os dados a seguir correspondem à variável renda familiar e gasto com alimentação (em unidades monetárias) para uma amostra de 25 famílias.

Renda Familiar (x)	Gasto com Alimentação (y)
3	1,5
5	2,0
10	6,0
10	7,0
20	10,0
20	12,0
20	15,0
30	8,0
40	10,0
50	20,0
60	20,0
70	25,0
70	30,0
80	25,0
100	40,0
100	35,0
100	40,0
120	30,0
120	40,0
140	40,0
150	50,0
180	40,0
180	50,0
200	60,0
200	50,0

Tendo como referência o que foi trabalhado no 5° Encontro, desenvolva os itens abaixo:

a) Em sua opinião, qual a vantagem da utilização do Google Colab na programação em Python?

b) Por que a linguagem de programação Python é considerada uma linguagem de alto nível?

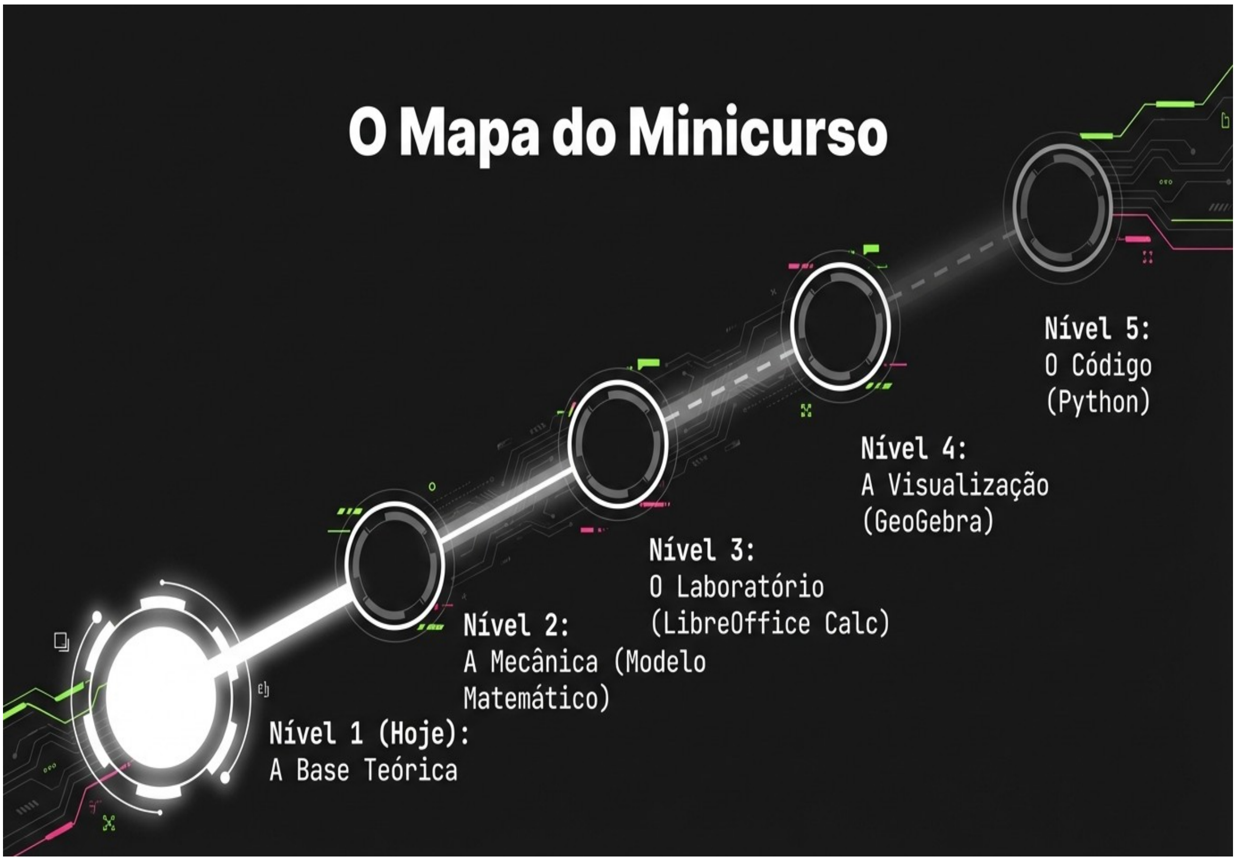
c) O uso das bibliotecas do Python facilita o trabalho do programador? Justifique a resposta.

d) Com os dados da tabela acima, construa um modelo de Regressão Linear Simples em Python.

REFERÊNCIA BIBLIOGRÁFICA

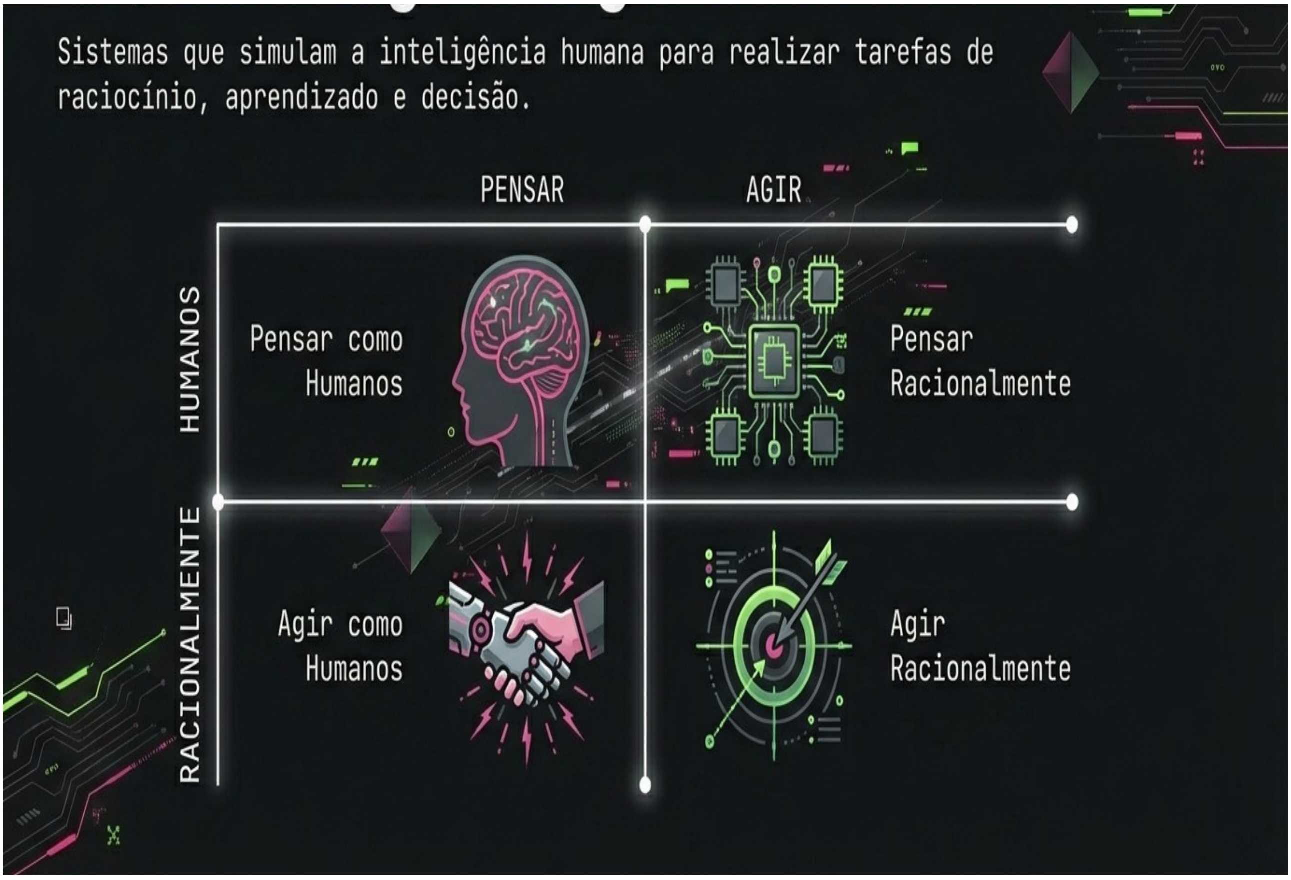
DA SILVA, Ronaldo Cesar. **Introdução ao Aprendizado de Máquina utilizando o Modelo de Regressão Linear Simples:** Uma proposta para estudantes do 3° Ano do Ensino Médio. Dissertação (Mestrado) – Colégio Pedro II, Programa de Mestrado Profissional em Matemática em Rede Nacional (PROFMAT), Rio de Janeiro, 2026.

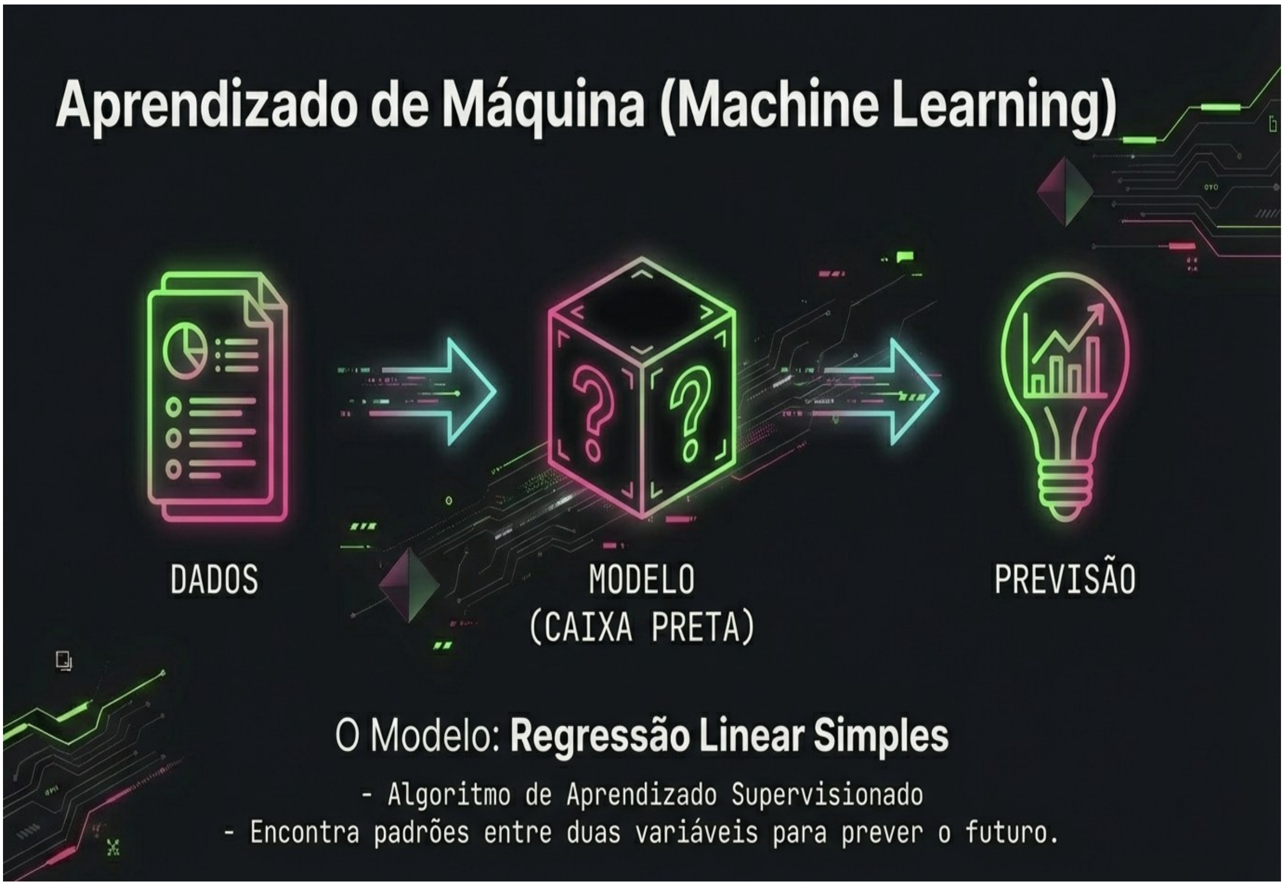
APÊNDICE – MODELO DE SLIDES PARA O 1° ENCONTRO

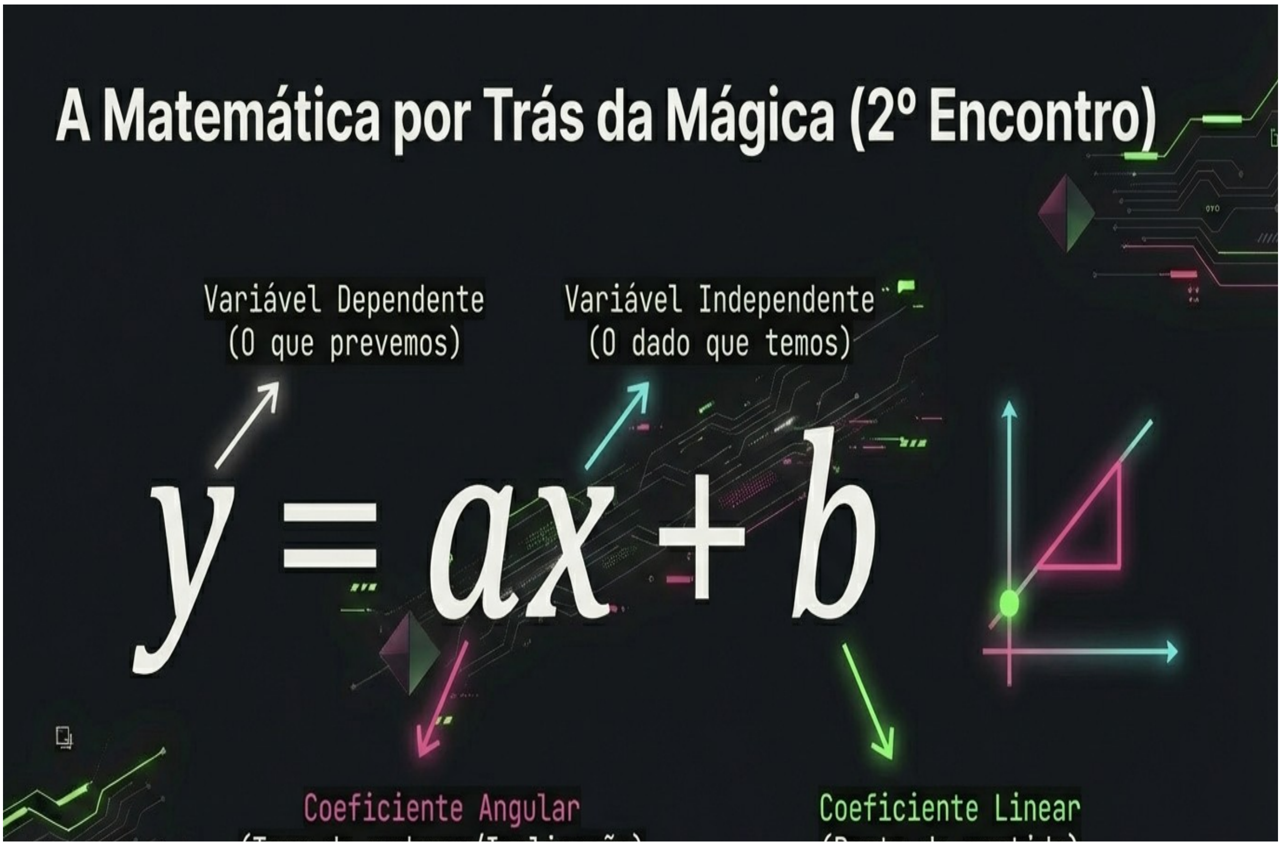

O Mapa do Minicurso
Nível 1 (Hoje):
A Base Teórica
Nível 2:
A Mecânica (Modelo
Matemático)
Nível 3:
O Laboratório
(LibreOffice Calc)
Nível 4:
A Visualização
(GeoGebra)
Nível 5:
O Código
(Python)

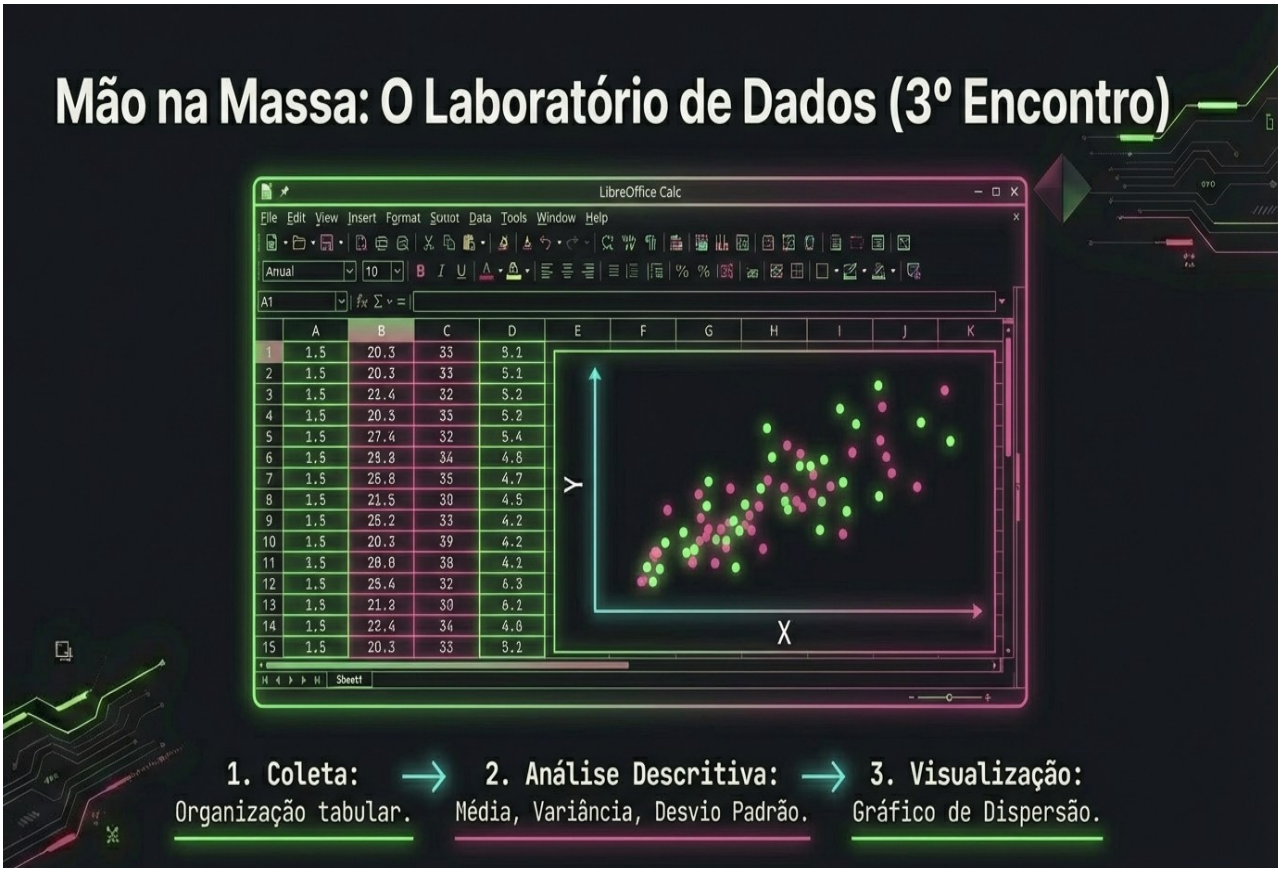

A Evolução da Inteligência (1950 - Hoje)
Do sonho de Turing à realidade Generativa.
1950
Teste de Turing
1956
Conferência de
Dartmouth
(Nascimento)
1997
Deep Blue vence
Kasparov
2010s
Popularização do
Deep Learning
2020s
IA Generativa
(ChatGPT, DALL-E)

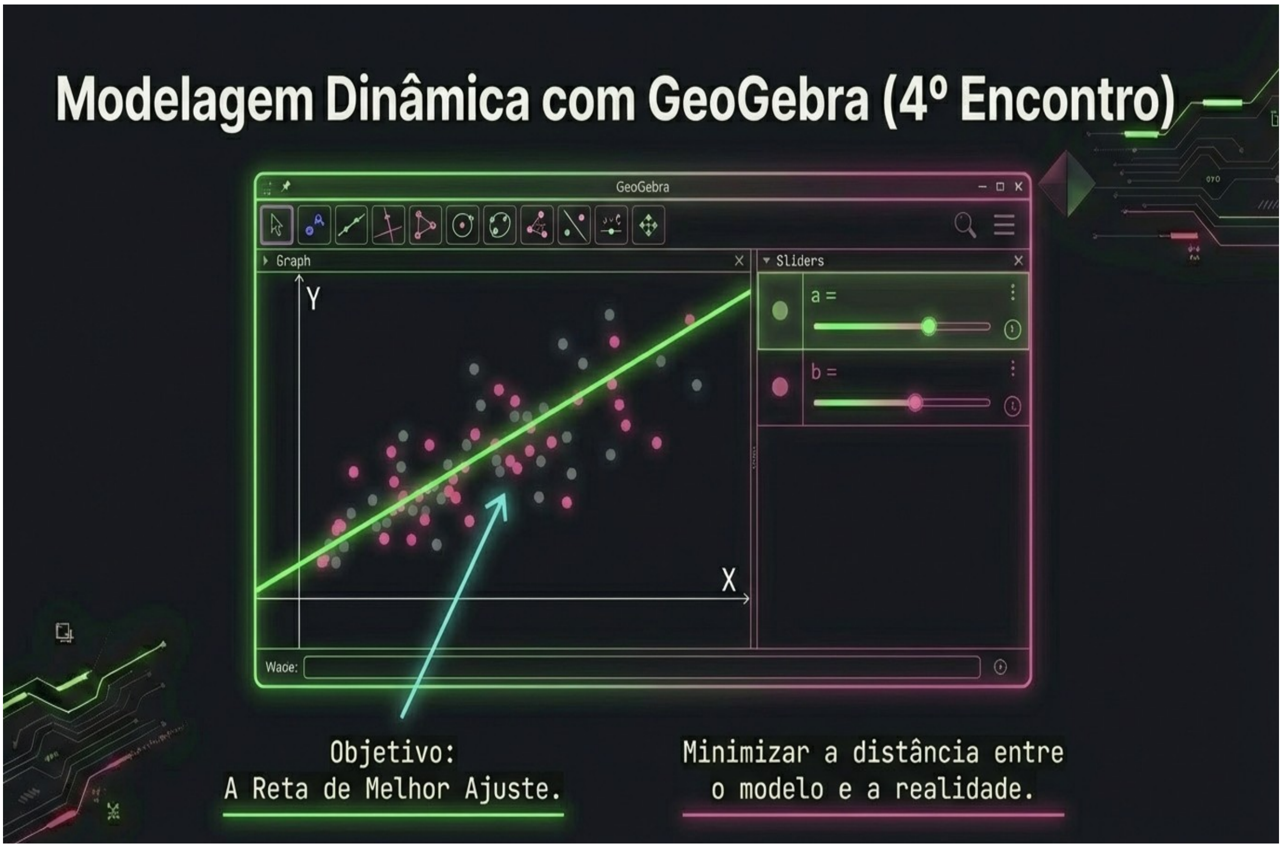
O que é Inteligência Artificial?

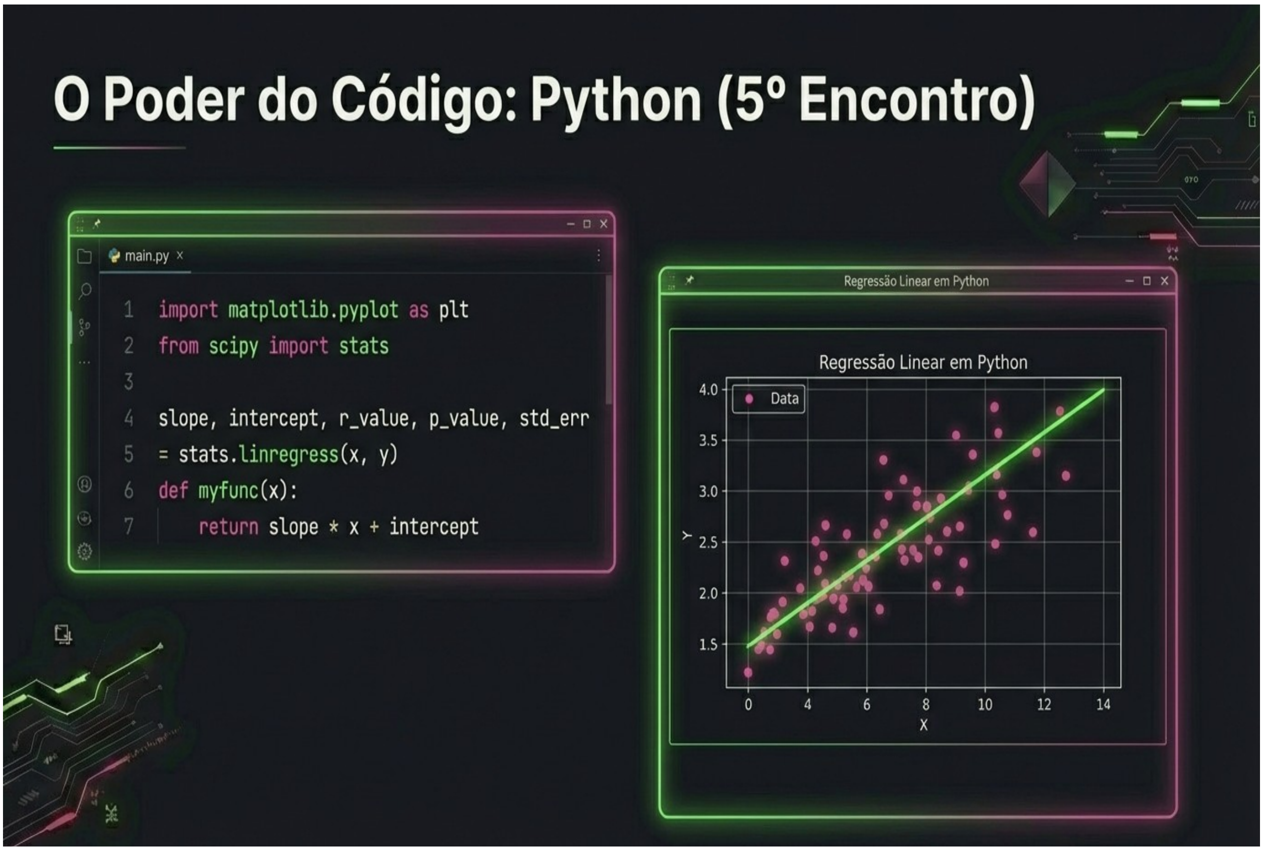



Aprendizado de Máquina (Machine Learning)
DADOS
MODELO
(CAIXA PRETA)
PREVISÃO
O Modelo: Regressão Linear Simples
- Algoritmo de Aprendizado Supervisionado
- Encontra padrões entre duas variáveis para prever o futuro.


A Matemática por Trás da Mágica (2º Encontro)
Variável Dependente
(O que prevemos)
Variável Independente
(O dado que temos)
y = ax + b
Coeficiente Angular
Coeficiente Linear


Mão na Massa: O Laboratório de Dados (3º Encontro)
LibreOffice Calc
Y
X
1. Coleta:
Organização tabular.
2. Análise Descritiva:
Média, Variância, Desvio Padrão.
3. Visualização:
Gráfico de Dispersão.


Modelagem Dinâmica com GeoGebra (4º Encontro)
GeoGebra
Graph
Sliders
a =
b =
Y
X
Objetivo:
A Reta de Melhor Ajuste.
Minimizar a distância entre
o modelo e a realidade.


O Poder do Código: Python (5º Encontro)
main.py
import matplotlib.pyplot as plt
from scipy import stats
slope, intercept, r_value, p_value, std_err
= stats.linregress(x, y)
def myfunc(x):
return slope * x + intercept
Regressão Linear em Python
Regressão Linear em Python
Data
4.0
3.5
3.0
2.5
2.0
1.5
Y
0
4
6
8
10
12
14
X