

COLÉGIO PEDRO II

Pró-Reitoria de Pós-Graduação, Pesquisa, Extensão
e Cultura

Mestrado Profissional em Matemática em Rede
Nacional

Ronaldo Cesar da Silva

**INTRODUÇÃO AO APRENDIZADO DE MÁQUINA
UTILIZANDO O MODELO DE REGRESSÃO LINEAR
SIMPLES: UMA PROPOSTA PARA ESTUDANTES DO 3º
ANO DO ENSINO MÉDIO**

Rio de Janeiro
2026



Ronaldo Cesar da Silva

**INTRODUÇÃO AO APRENDIZADO DE MÁQUINA UTILIZANDO O MODELO DE
REGRESSÃO LINEAR SIMPLES:**

Uma proposta para estudantes do 3º Ano do Ensino Médio

Dissertação de Mestrado apresentada ao Programa de Mestrado Profissional em Matemática em Rede Nacional, vinculado à Pró-Reitoria de Pós-Graduação, Pesquisa, Extensão e Cultura do Colégio Pedro II, como requisito parcial para obtenção do título de Mestre em Matemática.

Orientador(a): Professora Dra. Andreia Carvalho Maciel Barbosa.

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

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. – Rio de Janeiro, 2026.

120 f.

Dissertação (Mestrado Profissional em Matemática em Rede Nacional) – Colégio Pedro II, Pró-Reitoria de Pós-Graduação, Pesquisa, Extensão e Cultura.

Orientador: Andreia Carvalho Maciel Barbosa.

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

Ronaldo Cesar da Silva

**INTRODUÇÃO AO APRENDIZADO DE MÁQUINA UTILIZANDO O MODELO DE
REGRESSÃO LINEAR SIMPLES:**

Uma proposta para estudantes do 3º Ano do Ensino Médio

Dissertação de Mestrado apresentada ao Programa de Mestrado Profissional em Matemática em Rede Nacional, vinculado à Pró-Reitoria de Pós-Graduação, Pesquisa, Extensão e Cultura do Colégio Pedro II, como requisito parcial para obtenção do título de Mestre em Matemática.

Aprovado em: 27/02/2026.

Banca Examinadora:

Dr^a Andreia Carvalho Maciel Barbosa (Orientadora)
PROFMAT/CP2

Dr. Diego de Souza Nicodemos
PROFMAT/CP2

Dr^a Aline de Lima Guedes Machado
UERJ

Rio de Janeiro
2026

Dedico este trabalho à minha querida esposa Adriana e às minhas filhas, Ana Carolina e Ana Beatriz.

AGRADECIMENTOS

Agradeço imensamente a Deus, por me sustentar nos momentos de dúvida e hesitação.

Aos meus pais, Alcides e Valdinete, por terem me dado a vida, pelo amor e investimento na criação dos filhos e por me mostrarem a importância dos estudos.

À minha esposa Adriana, pelo companheirismo, incentivo, paciência e carinho durante o processo.

Aos professores do PROFMAT/CP II, por me proporcionarem muito mais do que conhecimento; vocês são exemplos de dedicação e amor à educação de qualidade!

Aos colegas da turma de 2023 do PROFMAT/CP II, por tantos momentos de aprendizagem e troca de experiências que tivemos juntos.

Aos meus superiores no CBMERJ, Ten-Cel Coelho, por ter me proporcionado todo apoio para que eu me dedicasse aos estudos e Ten-Cel Eduardo Ferreira, pelas inúmeras aulas de Python e trocas de ideias durante todas as etapas do processo.

Em especial, agradeço à minha orientadora e professora, Dr^a. Andreia Carvalho Maciel Barbosa, pelo conhecimento, paciência, afeto e humanidade que me ofereceu durante todo o processo de construção desse trabalho.

“O Brasil que precisamos construir, com oportunidade para todos, depende do êxito dos nossos esforços no campo da educação.”

(Gonzaga da Gama Filho)

RESUMO

DA SILVA, Ronaldo Cesar. **Introdução ao Aprendizado de Máquina utilizando o Modelo de Regressão Linear Simples**: Uma proposta para estudantes do 3º Ano do Ensino Médio. 2026. 106 f. 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.

O aprendizado de máquina é um campo da Inteligência Artificial que usa algoritmos com base em modelos estatísticos que fazem o computador reconhecer padrões através de dados, melhorando e aprimorando a resposta na medida em que são fornecidos mais dados. Este trabalho busca produzir significado contribuindo para uma introdução ao Aprendizado de Máquina (AM) para estudantes do 3º ano do ensino médio, usando o modelo preditivo de Regressão Linear Simples (ajuste de dados para uma função polinomial do 1º grau). Esta proposta 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). Sob o prisma dos Modelos dos Campos Semânticos (MCS) do professor Rômulo Campos Lins (1955-2017), a produção de significado acontece numa enunciação e quem produz uma enunciação é o autor. O leitor só se institui como autor na medida em que faz suas reflexões e constrói sua própria versão do enunciado e sendo assim, cada autor é único. Dessa forma, cada estudante participa ativamente do seu processo de ensino-aprendizagem, sendo capaz 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, Modelos dos Campos Semânticos, Regressão Linear Simples, Produção de Significados.

ABSTRACT

DA SILVA, Ronaldo Cesar. **Introdução ao Aprendizado de Máquina utilizando o Modelo de Regressão Linear Simples**: Uma proposta para estudantes do 3º Ano do Ensino Médio. 2026. 106 f. 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.

Machine learning is a field of Artificial Intelligence that uses algorithms based on statistical models that enable computers to recognize patterns in data, improving and refining their responses as more data is provided. This work aims to provide a meaningful introduction to Machine Learning (ML) for 3rd-year high school students, using the predictive model of Simple Linear Regression (data fitting to a first-degree polynomial function). This proposal intends to contribute to the field of mathematics education with a sequence of extracurricular tasks in which concepts about machine learning, supervised algorithms, simple linear regression, and technological tools used to facilitate data analysis, such as spreadsheets (LibreOffice Calc), mathematical software (GeoGebra), and programming languages (Python), will be presented. From the perspective of Professor Rômulo Campos Lins' (1955-2017) Semantic Field Models (SFM), the production of meaning occurs in an utterance, and the author is the one who produces an utterance. The reader only becomes an author to the extent that they reflect and construct their own version of the utterance; thus, each author is unique. In this way, each student actively participates in their teaching-learning process, being able to use technology ethically and responsibly, contributing to the formation of a society more aware of the use and development of computer applications and tools.

Keywords: Machine Learning, Semantic Field Models, Simple Linear Regression, Meaning Production.

LISTA DE FIGURAS

Figura 1 – Diagrama de Venn para a IA.....	23
Figura 2 – Tipos de Aprendizado de Máquina.....	29
Figura 3 – Gráficos de dispersão que sugerem uma tendência linear.....	32
Figura 4 – Gráficos de dispersão que sugerem uma tendência não linear.....	32
Figura 5 – Correlação linear perfeita positiva ($r = 1$) e correlação linear perfeita negativa ($r = -1$)	34
Figura 6 – Correlação linear positiva e correlação linear negativa ($-1 < r < 1$).....	34
Figura 7 – Modelo de Regressão Linear Simples.....	35
Figura 8 – Resíduos entre os valores observados e previstos.....	35
Figura 9 – Marcos Históricos da evolução da Inteligência Artificial.....	49
Figura 10 – Reportagem do jornal O Globo do dia 28/07/2025.....	50
Figura 11 - Gráfico de Dispersão das Notas de Matemática e Física.....	56
Figura 12 - Reta de Regressão Linear.....	59
Figura 13 - Inserção de dados na planilha.....	64
Figura 14 - Inserção do gráfico de dispersão na planilha.....	64
Figura 15 - Configuração do gráfico de dispersão na planilha.....	65
Figura 16 - Visualização do gráfico de dispersão na planilha.....	65
Figura 17 - Visualização do coeficiente de correlação na planilha.....	66
Figura 18 - Inserção da linha de tendência na planilha.....	66
Figura 19 - Inserção do coeficiente de determinação na planilha.....	67
Figura 20 - Exibição da reta de regressão e do R^2 na planilha.....	67
Figura 21 - Outra forma de exibir o R^2 na planilha.....	68
Figura 22 - Utilizando a função de regressão para previsão na planilha.....	68
Figura 23 - Menu do GeoGebra.....	71
Figura 24 - Exibição da planilha no GeoGebra.....	72
Figura 25 - Inserção de dados na planilha do GeoGebra	72
Figura 26 - Exibição do gráfico de dispersão no GeoGebra.....	73
Figura 27 - Inserção da linha de tendência no GeoGebra.....	73
Figura 28 - Fazendo previsão com o Geogebra.....	74
Figura 29 - Exibição dos dados estatísticos no GeoGebra.....	74
Figura 30 - Seleção dos dados da planilha.....	75
Figura 31 - Lista de pontos na janela de álgebra e na janela de visualização.....	75
Figura 32 - Exibição do coeficiente de correlação linear no GeoGebra.....	76
Figura 33 - Exibição da função de regressão e da reta de regressão no GeoGebra ..	76

Figura 34 - Exibição do coeficiente de determinação e da previsão de valor de venda no GeoGebra.....	77
Figura 35 - Mudança da reta de regressão e das estatísticas do modelo com a presença de outliers.....	77
Figura 36 - Tela inicial do Google Colab.....	81
Figura 37 - Abrindo a janela de inserção de códigos de programação.....	81
Figura 38 - Janela de inserção de códigos de programação.....	81
Figura 39 - Importação das bibliotecas do Python.....	83
Figura 40 - Inserção dos dados na janela de programação.....	85
Figura 41 - Comando para o cálculo do coeficiente de correlação.....	85
Figura 42 - Separando os dados de treino e os dados de teste do modelo.....	86
Figura 43 - Criação e treinamento do modelo e previsão de valores.....	86
Figura 44 - Comandos para a visualização dos resultados.....	87
Figura 45 - Visualização dos resultados e da reta de regressão.....	87

LISTA DE QUADROS

Quadro 1- IFDM/2023.....	46
--------------------------	----

LISTA DE TABELAS

Tabela 1 – Notas de Matemática e Física.....	54
Tabela 2 – Somatórios para cálculo do coeficiente de correlação.....	57
Tabela 3 – Somatórios para o cálculo dos parâmetros do modelo de regressão.....	58
Tabela 4 – Temperatura x Produtividade.....	63
Tabela 5 – Área x Valor de Venda.....	71
Tabela 6 – Consumo das Famílias x PIB.....	83

LISTA DE SIGLAS

AM – Aprendizado de Máquina
ANPD – Autoridade Nacional de Proteção de Dados
BNCC – Base Nacional Comum Curricular
CNE – Conselho Nacional de Educação
FIRJAN – Federação das Indústrias do Estado do Rio de Janeiro
GPT – Generative Pre-Trained Transformer
GPUs – Graphics Processing Unit
IA – Inteligência Artificial
IDEB – Índice de Desenvolvimento da Educação Básica
IFDM – Índice Firjan de Desenvolvimento Municipal
LGPD – Lei Geral de Proteção de Dados
MCS – Modelo dos Campos Semânticos
MEC – Ministério da Educação
MLGs – Modelos Lineares Generalizados
MMQ – Método dos Mínimos Quadrados
PIB – Produto Interno Bruto
SQE – Soma dos Quadrados Explicada
SQR – Soma dos Quadrados dos Resíduos
SQT – Soma dos Quadrados Totais

Sumário

1 INTRODUÇÃO.....	15
2 INTELIGÊNCIA ARTIFICIAL E APRENDIZADO DE MÁQUINA.....	19
2.1 Breve Histórico sobre a Evolução da Inteligência Artificial.....	19
2.2 Tipos de Inteligência Artificial.....	22
2.3 Aprendizado de Máquina (AM).....	24
2.4 Tipos de Aprendizado de Máquina.....	25
2.4.1 Aprendizado Supervisionado.....	26
2.4.1.1 Regressão.....	26
2.4.1.2 Classificação.....	26
2.4.2 Aprendizado Não Supervisionado.....	27
2.4.2.1 Agrupamento (Clustering).....	27
2.4.2.2 Redução de Dimensionalidade.....	27
2.4.3 Aprendizado Semissupervisionado.....	27
2.4.4 Aprendizado Autossupervisionado.....	28
2.4.5 Aprendizado por Reforço.....	28
2.5 O Marco Regulatório da Inteligência Artificial no Brasil e seu Diálogo com a LGPD.....	29
2.6 O Referencial do MEC para o Uso Responsável da Inteligência Artificial na Educação Brasileira.....	30
3 O MODELO DE REGRESSÃO LINEAR SIMPLES.....	31
3.1 Gráfico de Dispersão.....	31
3.2 Covariância.....	32
3.3 Coeficiente de Correlação Linear de Pearson (r).....	33
3.4 Regressão Linear Simples (RLS).....	34
3.5 O Método dos Mínimos Quadrados (MMQ).....	35
3.6 O Coeficiente de Determinação (R^2).....	37
3.7 Pressupostos do Modelo e Diagnóstico de Resíduos.....	38
3.8 Teste de hipóteses e Intervalo de Confiança.....	38
4 O MODELO DOS CAMPOS SEMÂNTICOS.....	39
4.1 A Regressão Linear como Campo Semântico.....	42
5 METODOLOGIA E ATIVIDADES PROPOSTAS.....	45
6 ANÁLISE DE RESULTADOS.....	93
7 CONSIDERAÇÕES FINAIS.....	99
REFERÊNCIAS.....	102
APÊNDICE A – Folha de Tarefas do 1º Encontro.....	105
APÊNDICE B– Folha de Tarefas do 2º Encontro.....	106
APÊNDICE C– Folha de Tarefas do 3º Encontro.....	107
APÊNDICE D– Folha de Tarefas do 4º Encontro.....	108
APÊNDICE E– Folha de Tarefas do 5º Encontro.....	109
ANEXO A – Termo de Assentimento Livre e Esclarecido.....	110
ANEXO B – Termo de Consentimento Livre e Esclarecido – Responsável Legal.....	111
ANEXO C – Termo de Consentimento Livre e Esclarecido – Maiores de Idade.....	113
ANEXO D – Parecer do Conselho de Ética.....	115

1 INTRODUÇÃO

O cenário contemporâneo do mercado de trabalho reflete uma valorização acentuada de profissionais capacitados no tratamento e análise de informações. De acordo com o Fórum Econômico Mundial (2023), funções como cientista de dados e especialistas em inteligência artificial figuram entre as ocupações com maior potencial de expansão global para os próximos anos. No contexto brasileiro, relatórios da Brasscom (2023) corroboram essa tendência, apontando um déficit estrutural de talentos qualificados frente à demanda crescente. Essa diferença entre oferta e demanda resulta em patamares salariais elevados e trajetórias de ascensão acelerada, conforme documentado nos guias salariais da consultoria Robert Half (2024), consolidando essas carreiras como pilares estratégicos da nova economia digital.

A Base Nacional Comum Curricular (BNCC) na área de Matemática e suas Tecnologias para o Ensino Médio possui Competências e Habilidades que estimulam o uso de tecnologias para tratamento e análise de dados. Como exemplo, pode-se citar a Competência Específica 5

Investigar e estabelecer conjecturas a respeito de diferentes conceitos e propriedades matemáticas, empregando recursos e estratégias como observação de padrões, experimentações e tecnologias digitais, identificando a necessidade, ou não, de uma demonstração cada vez mais formal na validação das referidas conjecturas. (BRASIL, 2018, p. 532)

Ainda na Competência Específica 5, dentro da área de Probabilidade e Estatística, a Habilidade (EM13MAT510) é: “Investigar conjuntos de dados relativos ao comportamento de duas variáveis numéricas, usando ou não tecnologias da informação e, quando apropriado, levar em conta a variação e utilizar uma reta para descrever a relação observada.”

A BNCC também possui Competências e Premissas Específicas da Computação, como a Competência Específica 2: “Analisar criticamente artefatos computacionais, sendo capaz de identificar as vulnerabilidades dos ambientes e das soluções computacionais buscando garantir a integridade, privacidade, sigilo e segurança das informações”.

Dentro da supracitada Competência, existem habilidades relevantes para o embasamento deste trabalho, como a Habilidade (EM13CO10): “Conhecer os fundamentos da Inteligência Artificial, comparando-a com a inteligência humana, analisando suas potencialidades, riscos e limites”, a Habilidade (EM13CO11): “Criar e explorar modelos computacionais simples para simular e fazer previsões,

identificando sua importância no desenvolvimento científico” e a Habilidade (EM13CO12): “Produzir, analisar, gerir e compartilhar informações a partir de dados, utilizando princípios da ciência de dados”.

A proposta desta pesquisa é aplicar uma sequência de tarefas, utilizando o conhecimento prévio dos estudantes sobre funções polinomiais do 1º grau e seus gráficos (reta), sobre o tema aprendizado de máquina através do modelo de regressão linear simples, incentivando o uso das tecnologias digitais.

O objetivo geral é analisar a produção de significados dos estudantes no âmbito do aprendizado de máquina, através do modelo dos campos semânticos (MCS), com foco restrito no modelo de regressão linear simples, com o intuito de que os estudantes consigam identificar situações do mundo atual que podem ser preditas utilizando o modelo de regressão estudado.

Especificamente, a partir dos discursos produzidos pelos estudantes, pretende-se:

- Construir significado sobre Aprendizado de Máquina o Modelo de Regressão Linear Simples.
- Investigar relações entre números expressos em tabelas para representá-los no plano cartesiano, identificando padrões e criando uma generalização que pode ser representada por uma função polinomial do 1º grau.
- Observar um conjunto de dados relativos a duas variáveis numéricas, com ou sem a utilização de computadores e identificar se essa relação entre as variáveis pode ser ajustada por uma reta.
- Identificar se um conjunto de pontos (x, y) no gráfico de dispersão pode ser ajustado através de um modelo de regressão linear simples.
- Utilizar planilhas eletrônicas, softwares matemáticos e linguagem de programação para a construção de um modelo de regressão linear simples.

Ainda usando como base de fundamentação a BNCC, vemos que a contextualização é tratada como um fator determinante em todas as áreas, pois é dessa forma que se prepara estudantes com iniciativa, bons argumentos e julgamentos. A BNCC salienta também

[...] que o uso de tecnologias possibilita aos estudantes aprofundar sua participação ativa nesse processo de resolução de problemas. São alternativas de experiências variadas e facilitadoras de aprendizagens que reforçam a capacidade de raciocinar logicamente, formular e testar conjecturas, avaliar a validade de raciocínios e construir argumentações. (BRASIL, 2018, p. 528)

A introdução de conceitos como aprendizado de máquina, inteligência artificial e técnicas de regressão linear simples para estudantes do 3º ano do ensino médio é uma forma de contextualizar conceitos matemáticos vistos anteriormente, tais como funções

polinomiais do 1º grau e seus gráficos, proporcionalidade, gráficos de dispersão, média e desvio-padrão, que quando vistos de forma isolada, num primeiro momento, parece não ter conexão entre si.

Ademais, a regressão linear simples é uma técnica estatística que não está presente nos livros didáticos de matemática do ensino médio, porém as competências e habilidades necessárias para que os estudantes do 3º ano possam acompanhar esse conteúdo, já foram adquiridas por eles ao longo da educação básica.

Os modelos de regressão linear são modelos mais simples do que outros utilizados em aprendizado de máquina, todavia o domínio dessa ferramenta estatística permite adquirir uma base sólida para outras modelagens mais sofisticadas.

Esta sequência de tarefas extracurriculares foi aplicada para estudantes do 3º Ano do Ensino Médio da Escola Estadual Bom Pastor, situada na periferia de Belford Roxo – RJ, com a finalidade de desenvolver a capacidade de identificar situações-problemas envolvendo duas variáveis que tenham correlação linear e encontrar uma reta de regressão linear simples que melhor se ajusta ao que foi proposto. Durante as atividades, os estudantes tiveram a oportunidade de explorar ferramentas, linguagens e programas de computador que facilitam e agilizam a análise e o tratamento de uma quantidade maior de dados, tornando a tomada de decisão mais eficiente, além de desenvolver a utilização da tecnologia de forma mais ativa e não apenas como usuários de redes sociais ou plataformas de conteúdos digitais.

As atividades foram realizadas em cinco encontros de cem minutos (dois tempos de aula por encontro) e durante o processo os estudantes foram convidados a pensar e produzir conhecimento sobre aprendizado de máquina, inteligência artificial e sua utilização no cotidiano, além de construírem modelos preditivos de regressão linear simples, a partir de conceitos estatísticos, com e sem a utilização de computadores.

Este trabalho possui sete capítulos, sendo o primeiro a introdução, que apresenta o tema abordado e a fundamentação de acordo com a BNCC, além dos objetivos e justificativas para a confecção da dissertação.

O segundo capítulo mostra uma visão geral sobre o conceito de Inteligência Artificial e Aprendizado de Máquina, passando pelo histórico de evolução da IA, comentando também sobre legislações brasileiras que visam a proteção de dados e o uso ético e responsável da IA.

O terceiro capítulo traz uma base teórica sobre os conceitos estatísticos de gráfico de dispersão, covariância, coeficiente de correlação e o modelo de regressão linear simples, com seus pressupostos e testes de validação do modelo.

O quarto capítulo faz um breve resumo sobre o modelo dos campos semânticos do professor Rômulo Lins, que serviu de base para a análise dos significados produzidos pelos estudantes durante as atividades e sua aplicabilidade no campo

semântico da regressão linear simples.

O quinto capítulo discorre sobre a metodologia da sequência de atividades aplicadas durante os cinco encontros, trazendo de forma detalhada como cada encontro aconteceu e as competências e habilidades envolvidas, segundo a BNCC. No sexto capítulo é feita uma análise de produção de significados de alguns dos estudantes envolvidos nas atividades, sob a luz do MCS.

Por derradeiro, o sétimo capítulo traz as considerações finais, mostrando os objetivos alcançados e algumas observações realizadas no decorrer do trabalho.

2 INTELIGÊNCIA ARTIFICIAL E APRENDIZADO DE MÁQUINA

A Inteligência Artificial (IA) é um dos campos mais promissores da ciência moderna, permeando desde a ficção até as mais variadas aplicações do nosso cotidiano.

Para compreender a Inteligência Artificial (IA), é interessante analisar primeiramente o conceito de inteligência humana. De acordo com Gardner (1995), a inteligência humana é um conjunto de capacidades cognitivas que permitem ao indivíduo resolver problemas ou elaborar produtos valorizados em um contexto cultural. Sob essa ótica, a inteligência envolve percepção, aprendizado, memória e, fundamentalmente, a capacidade de adaptação a novos cenários. Quando transportamos esse conceito para o campo computacional, a IA surge como uma tentativa de simular tais processos.

Segundo Rezende (2003), a IA é um campo da ciência da computação dedicado à criação de sistemas capazes de realizar tarefas que, se executadas por seres humanos, exigiriam inteligência, como o raciocínio lógico e a compreensão de linguagem.

Ao longo das décadas, diversas definições foram propostas, com abordagens que focam no "pensar como humanos", buscando replicar a arquitetura neural e os processos psicológicos. A vantagem desta visão é o aprofundamento na ciência cognitiva, mas sua dificuldade está na complexidade extrema e não totalmente compreendida do cérebro humano. Outras definições priorizam o "agir racionalmente", onde o sistema é avaliado pelo sucesso na execução de uma tarefa, independentemente de como processa a informação. Esta última é a definição mais aceita atualmente no mercado, pois foca em resultados práticos e eficiência, embora possa carecer da "sensibilidade" e do contexto ético inerentes ao julgamento humano.

2.1 Breve Histórico sobre a Evolução da Inteligência Artificial

O campo da IA é resultado de um sonho filosófico de longa data sobre a criação de máquinas pensantes, mas sua fundação científica moderna remonta a meados do século XX. O ponto de partida para essa evolução ocorreu em 1943, com o trabalho de Warren McCulloch e Walter Pitts¹, que apresentaram um modelo de neurônios artificiais capazes de calcular funções lógicas. Conforme detalhado por Russell e Norvig (2013, p. 16), esse período inicial foi caracterizado pelo desejo de entender se o cérebro poderia ser

1 Warren McCulloch (1898–1969), neuropsiquiatra, e Walter Pitts (1923–1969), matemático, foram os autores do artigo *A Logical Calculus of the Ideas Immanent in Nervous Activity* (1943). Neste trabalho, propuseram o primeiro modelo matemático de um neurônio biológico (o neurônio de McCulloch-Pitts), demonstrando que redes de neurônios artificiais simplificados poderiam calcular qualquer função lógica, o que lançou os fundamentos para o campo do conexionismo e das redes neurais profundas modernas.

mapeado como um dispositivo computacional.

O marco conceitual mais influente foi estabelecido por Alan Turing² em 1950, com a introdução do Teste de Turing, que estabeleceu as bases da visão moderna sobre o tema, ao publicar seu artigo sobre a capacidade das máquinas de pensar, introduzindo o teste que levaria seu nome e que ainda hoje serve como referência para avaliar o comportamento inteligente em sistemas artificiais (Russell & Norvig, 2013, p. 2).

Contudo, o nascimento oficial da IA ocorreu no workshop de Dartmouth em 1956, onde o termo "Inteligência Artificial" foi criado por John McCarthy³ (Russell & Norvig, 2013, p. 17). Esta fase foi dominada pela abordagem simbólica (ou clássica), que visava replicar o raciocínio humano por meio da manipulação de símbolos e regras lógicas explícitas. Em 1952, Arthur Samuel⁴ criou um jogo de damas no IBM 701 que conseguia aprender e melhorar seu desempenho e em 1957, Frank Rosenblatt⁵ desenvolveu o Perceptron, um dos primeiros algoritmos de aprendizado supervisionado baseado em redes neurais.

Esse período inicial foi marcado por um otimismo generalizado, levando à promessa de que "máquinas seriam capazes de realizar qualquer trabalho que um homem pudesse fazer, num prazo de vinte anos" (Russell & Norvig, 2013, p. 19).

Nos anos seguintes, o entusiasmo inicial gerou programas como o *General Problem Solver*, que tentava imitar as etapas de resolução de problemas humanos. Entretanto, a década de 1970 trouxe os primeiros grandes obstáculos, conhecidos como os "invernos da IA", quando a falta de poder computacional e a complexidade intratável de certos problemas levaram a um ceticismo generalizado e ao corte de subsídios governamentais para a pesquisa na área.

A recuperação do setor começou na década de 1980 com a ascensão dos sistemas especialistas, que utilizavam bases de conhecimento humano para resolver problemas específicos em nichos industriais. O desenvolvimento do algoritmo Backpropagation (Retropropagação) impulsionou o ressurgimento das Redes Neurais Artificiais, permitindo que modelos aprendessem a partir de dados de forma mais

2 Alan Mathison Turing (1912–1954) foi um matemático, cientista da computação e criptoanalista britânico, autor do artigo seminal *Computing Machinery and Intelligence* (1950), no qual propôs o "Jogo da Imitação", hoje conhecido como Teste de Turing.

3 John McCarthy (1927–2011) foi um influente cientista da computação e matemático norte-americano, a quem se atribui a cunhagem do termo "Inteligência Artificial" na proposta para a Conferência de Dartmouth em 1955.

4 Arthur Samuel (1901–1990) foi um pioneiro da computação na IBM e em Stanford. Em 1959, ele criou o termo "Aprendizado de Máquina" (*Machine Learning*), definindo-o como o campo de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados para cada tarefa específica.

5 Frank Rosenblatt (1928–1971), psicólogo e pioneiro da inteligência artificial, desenvolveu o "Perceptron" no Cornell Aeronautical Laboratory. Este modelo representou a primeira implementação prática de uma rede neural artificial capaz de aprender a classificar dados (como formas geométricas) através de um algoritmo de ajuste de pesos. O Perceptron é considerado o ancestral direto das redes neurais profundas (*Deep Learning*) utilizadas atualmente.

eficiente. De acordo com Russell e Norvig (2013, p. 22), esse período foi crucial para demonstrar que a IA poderia ter aplicações comerciais lucrativas, movendo o campo de laboratórios puramente acadêmicos para o interior das corporações. No final desta década, o renascimento das redes neurais e a introdução de técnicas estatísticas mais robustas permitiram que os sistemas passassem a "aprender" a partir de dados, em vez de dependerem apenas de regras lógicas rígidas inseridas manualmente por programadores.

Nos anos 90, houve um grande avanço na internet comercial e no aumento do poder de processamento, levando a aplicação da IA em sistemas de indexação e navegação, como os protótipos do Google. Em 1997, a vitória do Deep Blue⁶ sobre Garry Kasparov marcou o fim de uma era de busca simbólica, realizando uma transição da teoria para a aplicação prática de alto rendimento.

A partir do século XXI, a convergência entre o aumento exponencial de dados (o chamado Big Data), a maior capacidade de processamento computacional (impulsionada por GPUs) e o aprimoramento de algoritmos de aprendizado de máquina, especialmente o Aprendizado Profundo (Deep Learning), gerou grandes avanços.

A década de 2010 consolidou o Aprendizado Profundo como sendo a abordagem dominante para resolver muitos problemas de inteligência artificial. A popularização dos assistentes virtuais (como Siri em 2011) e o sucesso de IAs em jogos complexos, como o AlphaGo da DeepMind (que venceu o campeão mundial de Go⁷ em 2016), demonstraram o potencial da IA em escala global. Russell e Norvig (2013, p. 25) ressaltam que a disponibilidade de bases de dados massivas e algoritmos de aprendizado profundo permitiram avanços sem precedentes em visão computacional e fala.

O desenvolvimento de modelos como o Transformer⁸ (2017) e as subsequentes variantes do GPT (Generative Pre-trained Transformer) marcaram o surgimento da IA Generativa, que é uma subcategoria da IA, que se concentra na criação de novos conteúdos originais, em vez de apenas classificar ou analisar dados existentes. Assim sendo, ela é capaz de criar textos, imagens e sons novos com excelente qualidade. Essa tecnologia, está automatizando tarefas de comunicação e revolucionando a interação humano-máquina.

6 O Deep Blue foi um supercomputador desenvolvido pela IBM especificamente para jogar xadrez. Em maio de 1997, tornou-se o primeiro sistema computacional a vencer um campeão mundial de xadrez, Garry Kasparov, em um torneio de seis partidas sob condições de tempo regulamentares. Este evento é amplamente citado como um divisor de águas na história da inteligência artificial.

7 O Go é um jogo de tabuleiro estratégico de origem chinesa com mais de 2.500 anos de história.

8 Diferente de arquiteturas anteriores que processavam informações de forma linear (palavra por palavra), o *Transformer* analisa a sentença como um todo, atribuindo pesos de importância diferentes a cada parte da entrada. Essa abordagem resolveu o problema da "perda de memória" em frases longas, permitindo que a IA mantenha a coerência contextual em textos extensos, transformando o campo do Processamento de Linguagem Natural.

Atualmente, a IA está se expandindo para diversas áreas, desde diagnósticos médicos e agricultura de precisão até a automação de processos de trabalho, ao mesmo tempo em que levanta discussões cruciais sobre ética, explicabilidade e regulamentação.

2.2 Tipos de Inteligência Artificial

Para classificar os tipos de Inteligência Artificial, utiliza-se geralmente duas perspectivas, sendo uma baseada na capacidade e evolução (quão "inteligente" a IA é) e outra baseada na funcionalidade (como a IA processa informações). Dependendo do autor ou da referência utilizada, o entendimento sobre os tipos IA de pode variar.

Russell e Norvig (2013) organizam a IA em quatro categorias principais, dispostas em uma tabela que separa o foco entre racionalidade, pensamento e comportamento.

- Sistemas que pensam como humanos: Focam na modelagem cognitiva e em como o cérebro humano processa informações (ex: Redes Neurais que tentam replicar o raciocínio biológico).
- Sistemas que atuam como humanos: O objetivo é realizar tarefas que, se feitas por humanos, exigiriam inteligência. O Teste de Turing se encaixa aqui.
- Sistemas que pensam racionalmente: Baseiam-se na lógica e no "pensamento correto". Utilizam o silogismo e leis da lógica para chegar a conclusões irrefutáveis.
- Sistemas que atuam racionalmente (Agentes Inteligentes): Um agente racional age para alcançar o melhor resultado possível (ou o melhor resultado esperado em caso de incerteza).

Luger (2013) segue o eixo da IA como uma disciplina de resolução de problemas complexos. Ele categoriza os tipos de IA através de suas estratégias:

- IA Simbólica (Top-down): Processamento de símbolos e lógica formal para representar o conhecimento. É a IA baseada em regras claras e dedução.
- IA Conexionista (Bottom-up): Inspirada na biologia, foca no aprendizado a partir de dados (Redes Neurais), onde a inteligência emerge da conexão de pequenas unidades processadoras.
- IA Evolucionária: Algoritmos baseados na seleção natural e genética para encontrar soluções ótimas em grandes espaços de busca. Ela mantém uma população de soluções candidatas que 'evoluem' ao longo do tempo através de operadores de mutação e recombinação.

Rezende (2003) classifica a IA focando na estrutura dos Sistemas Inteligentes e suas aplicações práticas:

- Sistemas Especialistas: Programas que utilizam uma base de conhecimento e um motor de inferência para emular a capacidade de decisão de um especialista

humano em um domínio restrito.

- Aprendizado de Máquina (AM): Sistemas que melhoram seu desempenho através da experiência.
- Sistemas Híbridos: A combinação de diferentes técnicas (ex: Lógica Nebulosa/Fuzzy com Redes Neurais) para resolver problemas que uma única técnica não daria conta.

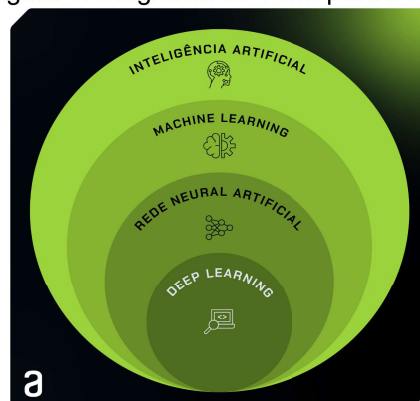
Já Teixeira (1990) traz uma visão mais filosófica e teórica, distinguindo dois conceitos fundamentais:

- IA Forte: A ideia de que uma máquina bem programada poderia realmente ter uma consciência e estados mentais reais, equiparando-se à mente humana.
- IA Fraca: A visão de que a IA é apenas uma ferramenta poderosa para simular processos mentais humanos para resolver problemas específicos, sem que a máquina de fato "pense" ou tenha consciência.

A diferença entre IA Forte (IA Geral) e IA Fraca (IA Estrita) está na capacidade de compreensão, versatilidade e autonomia do sistema. Enquanto a IA Fraca é uma ferramenta de especialização, baseada em algoritmos de reconhecimento de padrões, processamento de linguagem ou lógica matemática para resolver problemas dentro de parâmetros definidos, a outra é um conceito teórico de inteligência plena, um sistema que possuiria a capacidade intelectual de um ser humano. Ele poderia aprender, raciocinar, resolver problemas e se adaptar a qualquer tarefa e em qualquer contexto. Atualmente, a IA que está disponível para utilização é do tipo IA Fraca ou Estrita.

A IA Superinteligente seria o terceiro estágio evolutivo da IA, no qual a máquina superaria a inteligência humana em todos os campos. De acordo com Santaella (2023) a evolução da inteligência artificial não deve ser vista apenas como um incremento técnico, mas como um processo de coevolução que aponta para estágios onde a autonomia e a capacidade de processamento das máquinas podem transcender a inteligência biológica, alcançando o que a literatura classifica como superinteligência. A figura 1 mostra os subcampos da IA.

Figura 1: Diagrama de Venn para IA



2.3 Aprendizado de Máquina (AM)

Uma maneira didática de introduzir o conceito de aprendizado de máquina é fazer uma analogia à forma como os seres humanos aprendem através de suas experiências de vida. Quanto maior a quantidade de informações recebidas, interações com o ambiente, com outras pessoas, animais e objetos diversos, mais cultura e conhecimento conseguimos obter.

Arthur Samuel (1959), pioneiro no campo da inteligência artificial (IA), definiu Aprendizado de Máquina (AM) como um campo de estudo que oferece aos computadores a capacidade de aprender sem serem especificamente programados.

Segundo (Bhavsar et al. 2017; Alpaydin 2020) o AM é uma das principais subáreas da IA, e é composto por uma coleção de métodos criados a partir de modelos matemáticos baseados na teoria estatística que permitem aos computadores automatizar tarefas com base na descoberta sistemática de padrões nos conjuntos de dados disponíveis ou em experiências passadas. Outros pesquisadores em aprendizado de máquina também fizeram suas próprias definições sobre AM, mas todas convergem para as definições aqui citadas.

Os algoritmos e modelos de AM permitem que a máquina aprenda a partir de dados fornecidos, reconhecendo padrões e aprimorando sua acurácia e eficácia nas respostas de acordo com a quantidade e qualidade dos dados que alimentam o sistema.

Esses algoritmos de AM não são programados com regras específicas para executar uma tarefa diretamente. Esses sistemas são treinados usando dados e algoritmos que permitem melhorar seu desempenho para realizar a tarefa, ou seja, esses modelos de aprendizado de máquina permitem que computadores aprendam com os dados, sem que seja preciso programar essa tarefa, contribuindo para antever ou prever situações diversas.

É possível aplicar o aprendizado de máquina em muitas áreas do mercado, como previsão de vendas, detecção de fraudes, recomendação de produtos, diagnóstico médico entre outros e isso acontece devido ao enorme volume de dados que transitam pela internet e que podem ser armazenados e processados, como números, textos, imagens, vídeos e áudios. Todo tipo de informação que pode ser armazenada e transformada em dados, pode ser usada para treinar um algoritmo de aprendizado de máquina e o avanço nos processadores dos computadores faz com que o processamento dos dados se realize com uma velocidade cada vez maior.

Como exemplo, podemos citar o artigo de Dallagassa e de Oliveira (2024) sobre o uso de AM no diagnóstico de câncer de mama através de ultrassonografia, apresentado no XX Congresso Brasileiro de Informática em Saúde, realizado em Belo Horizonte – MG, no período de 08/10 a 11/10/2024.

De acordo com o artigo citado, foram processadas imagens de um banco de dados contendo ultrassonografias de mamas que apresentavam nódulos benignos e malignos e a análise das imagens foi realizada por meio de técnicas de AM, que demonstraram altas taxas de precisão na classificação de nódulos benignos e malignos com resultados acima de 90% de assertividade.

Sempre que os algoritmos ao longo de sua execução, no processamento dos dados, reconhecem padrões e adquirem “conhecimento” sobre o que está sendo trabalhado, estamos diante de um modelo de AM.

Segundo Dias (2024), deve-se ter atenção especial aos dados que alimentam o modelo para que a resposta seja adequada ao problema proposto. As inadequações que podem surgir são:

- Quantidade insuficiente de dados de treinamento: os modelos de aprendizado de máquina necessitam de uma grande quantidade de dados;
- Dados de treinamento não representativos: Espera-se que os dados utilizados para o treinamento consigam generalizar o modelo;
- Dados de baixa qualidade: Dados que se afastam muito dos dados da série (outliers) e dados que não são explicados pelo modelo (ruídos/erros);
- Características irrelevantes: São atributos que não têm impacto significativo na resposta do modelo, prejudicando sua precisão;
- Sobreajustamento dos dados do modelo: Acontece quando o modelo funciona bem com os dados de treino (dados utilizados para treinar o modelo) mas não funciona bem com os dados de teste (dados utilizados para testar o sistema) ou novos dados.
- Subajustamento: Ocorre quando o modelo não funciona bem com os dados de treino, dados de teste e nem com novos dados, mostrando que o modelo escolhido é muito simples para caracterizar o problema.

2.4 Tipos de Aprendizado de Máquina

Russell e Norvig (2013) definem os três tipos principais de aprendizado de máquina como sendo o aprendizado supervisionado, o aprendizado não supervisionado e o aprendizado por reforço. Há ainda o aprendizado semissupervisionado e atualmente, devido a evolução tecnológica, temos o aprendizado autossupervisionado. Tanto o

aprendizado semissupervisionado quanto o aprendizado autossupervisionado são uma combinação dos aprendizados supervisionado e não supervisionado.

2.4.1 Aprendizado Supervisionado

No aprendizado supervisionado é necessária a intervenção humana para rotular os dados, ou seja, nesses dados cada entrada possui a saída desejada conhecida para que o modelo seja treinado através desses dados rotulados. O objetivo é que o algoritmo aprenda a relação entre as entradas e saídas para, posteriormente, prever a saída correta para novas entradas não vistas, ajustando os parâmetros para minimizar os erros entre as previsões e as saídas desejadas. Esses dados rotulados devem ser separados de forma aleatória entre dados de teste (entre 20% a 30%) e dados de treino (entre 70% a 80%).

Como visto, os dados de treino servem para a criação do modelo e os dados de teste servem para verificar o desempenho do modelo. Dentro do aprendizado supervisionado existem as técnicas de regressão e classificação e de acordo com o problema a ser resolvido, deve-se utilizar uma dessas técnicas.

2.4.1.1 Regressão

Na técnica de regressão, o modelo é treinado com um conjunto de dados rotulados para prever um valor numérico e dessa forma, o modelo relaciona os dados de entrada com os dados de saída. Com o modelo ajustado, ele é usado para prever o valor de saída de novos dados que não foram rotulados. Como exemplo, uma técnica de regressão pode ser utilizada para encontrar o valor de venda de um imóvel a partir de dados de entrada como a área, o ano de construção, a quantidade de cômodos, vaga de garagem entre outros. O foco desse trabalho é a regressão linear simples, que descreve a relação entre uma variável de entrada (preditora) e uma variável de resposta (saída) numérica.

2.4.1.2 Classificação

Na técnica de classificação, o modelo é treinado com um conjunto de dados rotulados no intuito de classificar os dados em grupos ou classes. O modelo é preparado para encontrar características entre os dados de entrada e seus grupos específicos. Após o treinamento do modelo, ele será utilizado para prever o grupo de novos dados não rotulados. Pode ser utilizado na identificação de imagens e classificação de e-mails como spam ou não.

2.4.2 Aprendizado Não Supervisionado

O aprendizado não supervisionado caracteriza-se por não haver a intervenção humana para rotular/classificar os dados, sendo assim, o modelo recebe dados não rotulados em busca de padrões ou estruturas. Dentre as técnicas utilizadas, podemos citar o agrupamento (clustering) e a redução de dimensionalidade.

2.4.2.1 Agrupamento (Clustering)

Essa técnica consiste em separar um conjunto de dados em grupos (clusters) com características semelhantes entre si e diferentes dos demais grupos. O algoritmo analisa e identifica os dados de entrada e com base em padrões, similaridades e especificidades e forma grupos na saída desses dados, separando-os conforme suas semelhanças. Essa técnica permite criar um grupo de clientes com base em seus comportamentos de compra, ou agrupar dados semelhantes na busca de padrões em imagens e textos, por exemplo.

2.4.2.2 Redução de Dimensionalidade

A redução de dimensionalidade consiste em reduzir o número de variáveis a serem tratadas na intenção de encontrar um menor número de variáveis que possam representar de forma mais adequada as características do conjunto de dados tratados. Tem grande utilidade na detecção de bordas, no processamento digital de imagens e na procura de operações bancárias fraudulentas, por serem anomalias difíceis de detectar em uma grande quantidade de dados (Almeida, A.; Carvalho, F.; Menino, F. 2017).

2.4.3 Aprendizado Semissupervisionado

É caracterizado pela utilização de dados rotulados e dados não rotulados para o treinamento do modelo. Sua utilização justifica-se quando os dados rotulados são escassos ou a rotulagem dos dados demanda muito tempo ou dinheiro, porém os dados não rotulados estão à disposição em grande quantidade. O algoritmo pode utilizar os dados rotulados para identificar padrões e depois aplicá-los aos dados não rotulados, para que as previsões sejam mais precisas. Um exemplo seria a classificação de documentos de texto no qual poucos são classificados em categorias como “política”,

“esportes” e “saúde” e muitos outros não são rotulados. O algoritmo “aprende” com os dados rotulados e aplica esses padrões para classificar os documentos não rotulados.

2.4.4 Aprendizado Autossupervisionado

Essa técnica utiliza o aprendizado não supervisionado para situações nas quais, usualmente, é empregado o aprendizado supervisionado. Esses modelos geram rótulos implícitos a partir de dados não rotulados. As técnicas autossupervisionadas podem ser mais econômicas e menos demoradas, pois substituem parte ou toda a necessidade de rotular manualmente os dados de treinamento. São utilizadas para rotular grande quantidade de dados que são utilizados para treinamento de modelos de inteligência artificial de ponta.

2.4.5 Aprendizado por Reforço

No aprendizado por reforço, que baseia-se no processo de decisão de Markov⁹, o agente (algoritmo) “aprende” a partir de uma série de recompensas ou punições (Russel e Norvig. 2013). Cada ação bem-sucedida resulta em feedback positivo, enquanto ações inadequadas recebem feedback negativo ou penalidades. O agente utiliza a tentativa e erro para maximizar as recompensas. Os modelos são treinados para tomarem uma sequência de decisões em um ambiente incerto e complexo. Nessa abordagem, os agentes possuem um estado que é alterado após realizar uma ação que é executada de forma aleatória, com base nessa ação, os agentes podem ser penalizados ou recompensados. Caso a ação do agente gere recompensas, então ela será reforçada para o seu próximo estado (Goodfellow et al. 2016). Um exemplo desse tipo de aprendizado é o AlphaGo, programa de IA da que derrotou o campeão mundial de Go, um jogo de tabuleiro mais complexo do que o xadrez, em 2016. O programa utilizou deep learning e aprendizado por reforço profundo. A figura 2 apresenta os três principais tipos de aprendizado de máquina.

9 Andrei Andreevich Markov (1856–1922) foi um matemático russo famoso por desenvolver os processos de Markov, uma das bases fundamentais da probabilidade moderna e inteligência artificial.

Figura 2: Tipos de Aprendizado de Máquina



Fonte: <https://dataat.github.io/introducao-ao-machine-learning/introdu%C3%A7%C3%A3o.html#machine-learning>

2.5 O Marco Regulatório da Inteligência Artificial no Brasil e seu Diálogo com a LGPD

O debate sobre a regulamentação da Inteligência Artificial (IA) no Brasil ganhou significativa atenção nos últimos anos, culminando em diversas propostas legislativas que visam estabelecer um Marco Legal da IA. Embora o país ainda não possua uma lei final e consolidada dedicada exclusivamente à IA, o andamento do Projeto de Lei nº 2.338/2023 no Senado Federal sinaliza a intenção de criar um arcabouço normativo que equilibre inovação, direitos fundamentais e segurança jurídica. Este projeto, influenciado por modelos internacionais como o AI Act da União Europeia, adota uma abordagem baseada em riscos, classificando os sistemas de IA em diferentes níveis de periculosidade, de risco inaceitável a risco baixo, para aplicar exigências regulatórias proporcionais a cada categoria.

O objetivo principal do marco é promover o desenvolvimento e uso responsável da IA, garantindo a proteção dos direitos dos cidadãos e o respeito à dignidade humana, especialmente em contextos de alto impacto social, como nas áreas de crédito, saúde e justiça.

A ausência de um marco legal específico para a IA, contudo, não implica um vácuo regulatório. A Lei Geral de Proteção de Dados (LGPD - Lei nº 13.709/2018) é a principal referência para fornecer as diretrizes imediatas para a aplicação ética e legal dos sistemas de IA, atuando como um regulador transversal entre a proteção de dados e a utilização da IA. O elo entre os dois temas é imediato e inseparável, pois a maior parte dos sistemas de IA, em especial os modelos de aprendizado de máquina, dependem massivamente da coleta, processamento e análise de grandes volumes de dados pessoais. A LGPD, ao exigir o consentimento para o tratamento de dados de terceiros, ao estabelecer princípios como a finalidade, a necessidade e a transparência e determinar regras para o tratamento de dados sensíveis, impõe limites claros, obrigações e responsabilidade aos desenvolvedores e operadores de sistemas de IA.

O ponto de maior convergência reside na questão da transparência e da explicabilidade algorítmica. A LGPD, em seu Artigo 20, confere aos titulares o direito de solicitar a revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses, como decisões de crédito, definição de perfis de consumo ou avaliação de risco. Este dispositivo obriga as empresas a fornecerem informações claras sobre os critérios e procedimentos utilizados na decisão automatizada, exigindo que os sistemas de IA não sejam "caixas-pretas" impenetráveis. O Marco Legal da IA em tramitação busca robustecer essa exigência, impondo, para os sistemas de alto risco, obrigações de documentação técnica, avaliação de conformidade e adoção de medidas para garantir a auditabilidade dos algoritmos, de modo a permitir que a fiscalização da Autoridade Nacional de Proteção de Dados (ANPD), que tem um papel crucial na aplicação da LGPD, seja efetiva e detalhada.

O futuro Marco Legal da IA no Brasil pretende complementar a LGPD, criando regras para mitigar os riscos específicos gerados pelas tecnologias inteligentes. Enquanto a LGPD foca na proteção da privacidade e dos dados pessoais como direito fundamental, o Marco Regulatório da IA se dedica a disciplinar o impacto sistêmico da tecnologia em si, como a discriminação algorítmica, a manipulação comportamental e a segurança dos sistemas. O arcabouço legal brasileiro está, portanto, sendo construído sobre uma dupla camada de proteção: a LGPD como fundação ética e de privacidade, e o Marco da IA como o instrumento específico para a governança e fiscalização dos riscos tecnológicos avançados.

2.6 O Referencial do MEC para o Uso Responsável da Inteligência Artificial na Educação Brasileira

O Ministério da Educação (MEC) do Brasil, reconhecendo o impacto transformador da Inteligência Artificial (IA) no setor educacional, tem liderado o esforço para estabelecer um referencial para o desenvolvimento e uso responsável da IA na Educação. Este referencial, ainda em fase de construção democrática, notadamente por meio de consultas e audiências públicas coordenadas pelo Conselho Nacional de Educação (CNE), visa fornecer diretrizes e salvaguardas éticas, pedagógicas e regulatórias. O objetivo central é assegurar que a IA atue como uma aliada estratégica para a melhoria da qualidade do ensino e da aprendizagem, sem comprometer os direitos dos estudantes e professores nem exacerbar as desigualdades educacionais pré-existentes.

A proposta de referencial do MEC se estrutura em eixos temáticos que abordam tanto aspectos técnicos quanto questões éticas e de inclusão. Entre os focos principais, destacam-se a necessidade de rigorosa proteção de dados educacionais, garantindo a conformidade com a Lei Geral de Proteção de Dados (LGPD) e o uso responsável de

informações sensíveis de alunos e docentes. Outro eixo crucial é o combate a vieses algorítmicos, exigindo que os sistemas de IA sejam desenvolvidos e testados para evitar a discriminação e o reforço de estereótipos que possam prejudicar determinados grupos de estudantes, promovendo ativamente a equidade e a acessibilidade universal.

O aspecto da governança e transparência é considerado fundamental, uma vez que o referencial preliminar sinaliza a adoção de diretrizes como a supervisão humana significativa em todas as etapas da implementação da IA, assegurando que o julgamento pedagógico do professor prevaleça sobre as decisões automatizadas. Paralelamente, exige-se a transparência e explicabilidade dos sistemas, para que professores, estudantes e responsáveis compreendam os critérios e procedimentos pelos quais as ferramentas de IA chegam a conclusões, como avaliações e recomendações personalizadas. Essa responsabilização é reforçada pela necessidade de realizar uma avaliação de impacto algorítmico antes da adoção de sistemas de alto risco em escolas e universidades.

Além das garantias éticas, o referencial também se concentra em aspectos operacionais e de desenvolvimento profissional. É dada alta prioridade à formação continuada de professores e gestores, equipando-os com as competências digitais e o letramento em IA necessários para integrar a tecnologia de forma crítica e alinhada às finalidades pedagógicas do currículo. A questão dos direitos autorais e integridade acadêmica também é abordada, buscando estabelecer protocolos que orientem o uso ético da IA generativa por estudantes e professores, ao mesmo tempo em que se definem protocolos de uso por faixa etária, reconhecendo as diferentes necessidades de desenvolvimento de crianças e adolescentes.

3 O MODELO DE REGRESSÃO LINEAR SIMPLES

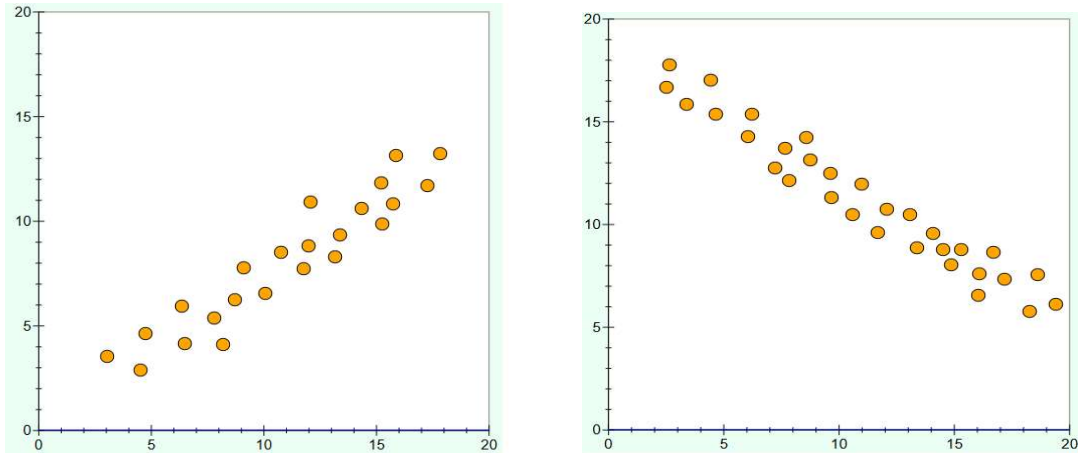
A fundamentação teórica deste capítulo foi baseada em Bussab e Morettin (2017), Triola (2010) e Gujarati e Porter (2011). Serão abordados os conceitos de Gráfico de Dispersão, Covariância e Correlação Linear como assuntos preliminares à Regressão Linear Simples.

3.1 Gráfico de Dispersão

Antes de qualquer análise numérica, a inspeção visual por meio do Gráfico de Dispersão é o primeiro passo para identificar a presença de tendências lineares, a dispersão dos dados e a existência de outliers (valores atípicos) que possam comprometer o modelo. Nele, cada par de observações (x_i, y_i) é um ponto no plano cartesiano. O gráfico de dispersão permite verificar se os pontos tendem a se organizar ao longo de uma linha reta. Se a nuvem de pontos sugerir uma tendência não linear, o

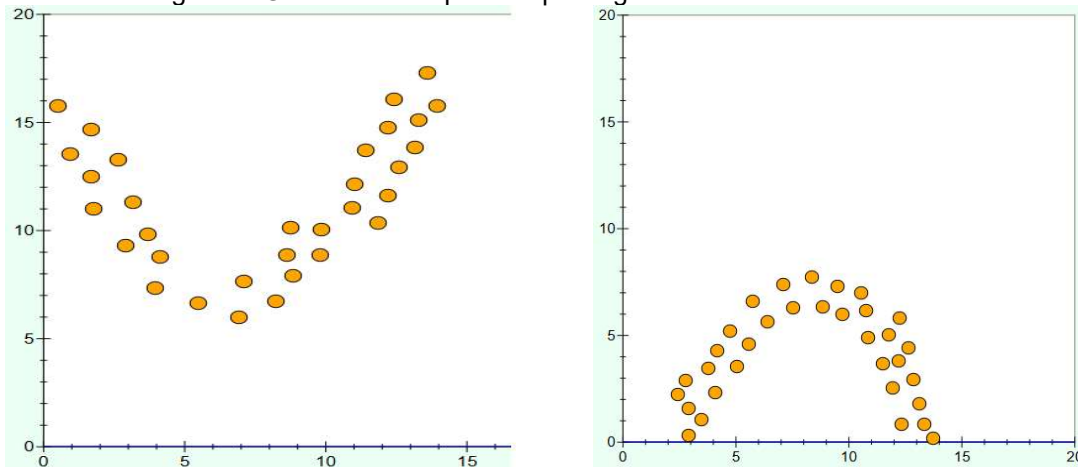
modelo de regressão linear simples será inadequado. As figuras 3 e 4 apresentam alguns gráficos de dispersão.

Figura 3: Gráficos de Dispersão que sugerem uma tendência linear



Fonte: O autor

Figura 4: Gráficos de Dispersão que sugerem uma tendência não linear



Fonte: O autor

3.2 Covariância

A covariância é uma medida estatística de associação linear que descreve a tendência de variação conjunta entre duas variáveis aleatórias. Seu propósito fundamental é medir a interdependência entre os dados. Quando o desvio de uma variável em relação à sua média é acompanhado por um desvio similar da outra variável, observa-se uma covariância positiva. Inversamente, quando os desvios ocorrem em sentidos opostos, a covariância é negativa.

A covariância de uma população de n observações é dada por:

$$Cov(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n}$$

onde x_i e y_i são os valores individuais em cada ponto e \bar{x} e \bar{y} são as médias aritméticas

de cada variável.

No caso de uma amostra, utiliza-se $n-1$ no denominador, para evitar o viés de estimativa¹⁰.

A covariância é uma medida difícil de ser interpretada pois é expressa pelo produto das unidades das variáveis, porém, se $Cov(X,Y) > 0$ pode-se dizer que existe uma tendência de associação positiva entre as variáveis, ou seja, quando uma variável aumenta, a outra tende a aumentar também. Do mesmo modo que, se $Cov(X,Y) < 0$, existe uma tendência de associação negativa entre as variáveis, pois quando uma variável aumenta, a outra tende a diminuir. Quando $Cov(X,Y) = 0$, dizemos que as variáveis são não correlacionadas.

3.3 Coeficiente de Correlação Linear de Pearson (r)

O coeficiente de Correlação Linear de Pearson mede a direção e a intensidade da relação linear entre duas variáveis. Ele é uma medida adimensional, o que permite comparar a força da associação entre diferentes tipos de dados. Sua fórmula tem origem na covariância das variáveis X e Y , normalizada pelo produto dos desvios padrão de cada variável:

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

Para facilitar os cálculos, a seguinte fórmula também é utilizada:

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

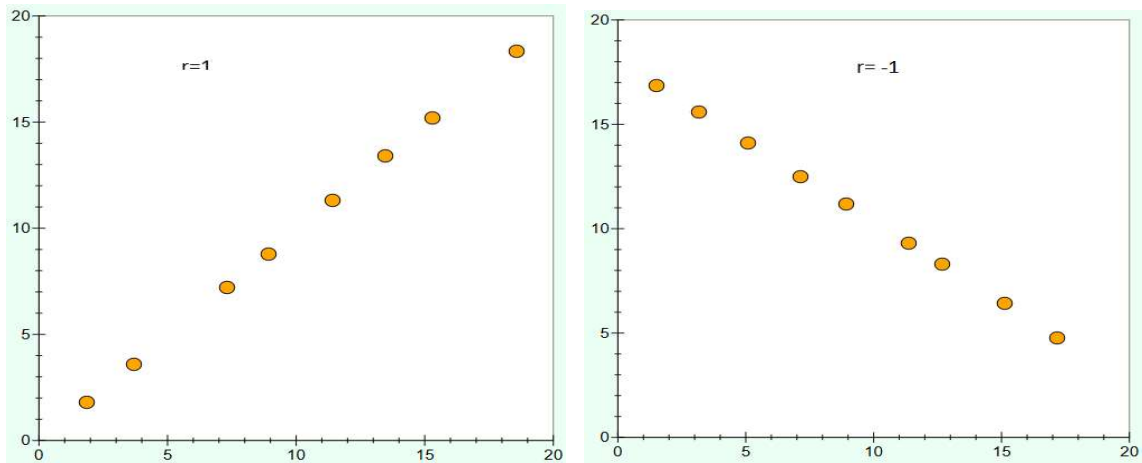
Sendo x_i os valores assumidos pela variável x , y_i os valores assumidos pela variável y e n o número de pares de dados da amostra. Os valores limites para r estão no intervalo $[-1, +1]$.

- 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 seja, a relação entre as variáveis não é linear.

¹⁰ Viés de estimativa é um conceito da inferência estatística que descreve a tendência de um estimador de superestimar ou subestimar sistematicamente o parâmetro real de uma população.

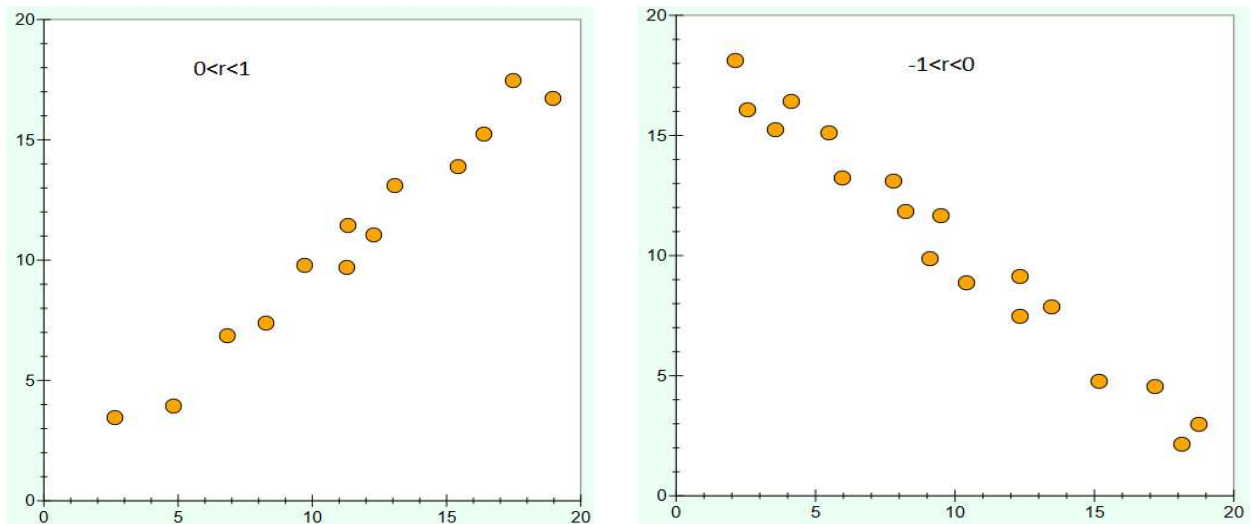
As figuras 5 e 6 mostram algumas correlações lineares.

Figura 5: Correlação linear perfeita positiva ($r=1$) e correlação linear perfeita negativa ($r=-1$)



Fonte: O autor

Figura 6: Correlação linear positiva e correlação linear negativa ($-1 < r < 1$)



Fonte: O autor

3.4 Regressão Linear Simples (RLS)

A Regressão Linear Simples (RLS) é uma técnica estatística que visa modelar a relação entre duas variáveis quantitativas, sendo uma variável dependente (ou resposta), denotada por Y e uma variável independente (ou preditora), denotada por X . O objetivo é encontrar uma função linear que descreva como Y varia em função de X , permitindo a predição de valores e a compreensão da intensidade dessa associação. Entretanto, esses pares ordenados (x_i, y_i) formados pelo conjunto de dados, dificilmente estarão em sua totalidade, sobre a reta de regressão, existindo uma diferença entre os valores observados e os valores previstos na função linear encontrada. Essa diferença é chamada de erro ou resíduo (ε) e pode representar variáveis não consideradas no modelo ou erro na coleta de dados. Adicionando o erro à função linear, temos:

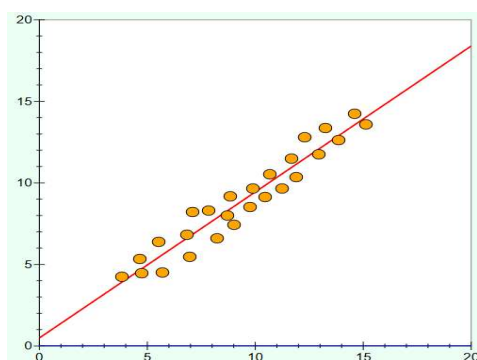
$$Y = a \cdot X + b + \varepsilon$$

que é o modelo de regressão linear simples, sendo a e b os parâmetros do modelo.

Segundo Bussab e Morettin (2017, pág. 451), o modelo é linear porque é representado por uma reta, mas numa visão mais abrangente, o termo linear refere-se ao fato dos parâmetros entrarem no modelo de forma linear.

De acordo com Crespo (2004, pág. 157), uma norma no uso de equações de regressão é a de nunca extrapolar, salvo quando considerações teóricas ou experimentais justifiquem tal possibilidade. A extrapolação é a previsão de valores fora do intervalo de dados observado. A figura 7 apresenta o gráfico de um modelo de Regressão Linear Simples.

Figura 7: Modelo de Regressão Linear Simples



Fonte: O autor

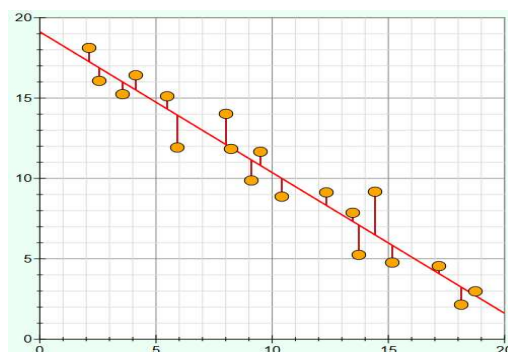
3.5 O Método dos Mínimos Quadrados (MMQ)

Para estimar os parâmetros a e b a partir de uma amostra, utiliza-se o Método dos Mínimos Quadrados (MMQ), que encontra a melhor função linear que minimiza a soma (SQR) dos quadrados das diferenças (erros ou resíduos) entre os valores observados (y_i) e os valores previstos pelo modelo (\hat{y}):

$$SQR = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Esse método garante que a reta resultante seja a que melhor se ajusta ao conjunto de dados, passando "pelo centro" da nuvem de pontos no gráfico de dispersão, de modo a equilibrar as distâncias verticais dos pontos acima e abaixo dela. A figura 8 mostra essas distâncias verticais.

Figura 8: Resíduos (Diferença entre os valores observados e previstos pelo modelo)



Fonte: O autor

Seja y_i o valor real observado e $\hat{y}=ax_i+b$, o valor predito pela reta, temos que o erro (resíduo) para cada ponto é $\varepsilon_i=y_i-(ax_i+b)$. Como alguns erros são positivos e outros negativos, para evitar que se cancelem, minimizamos a soma dos Quadrados dos Resíduos (SQR).

$$S(b, a) = \sum_{i=1}^n (y_i - b - ax_i)^2$$

Para minimizar a função S , calculamos as derivadas parciais em relação a b e a e igualamos cada derivada parcial a zero. Derivando em relação a b , temos:

$$\frac{\partial S}{\partial b} = -2 \sum_{i=1}^n (y_i - b - ax_i) = 0$$

Dividindo por -2 e distribuindo o somatório, temos:

$$\sum_{i=1}^n y_i - nb - a \sum_{i=1}^n x_i = 0$$

Isolando b e dividindo ambos os lados por n , temos:

$$b = \bar{y} - a\bar{x}, \text{ onde } \bar{y} \text{ e } \bar{x} \text{ são as médias das variáveis.}$$

Derivando em relação a a , temos:

$$\frac{\partial S}{\partial a} = -2 \sum_{i=1}^n x_i (y_i - b - ax_i) = 0$$

Substituindo o valor de b encontrado anteriormente e resolvendo a derivada parcial para a , temos:

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

Assim, os parâmetros a e b da função linear são:

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

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

Para facilitar os cálculos, o parâmetro a também pode ser escrito da forma:

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

O MMQ garante que a reta de regressão passe pelo ponto médio dos dados (\bar{x}, \bar{y}) , pois, como a reta de regressão estimada é $\hat{y} = b + ax$, se substituirmos x pela média \bar{x} , temos:

$$\hat{y} = b + a\bar{x}$$

Substituindo o valor de $b = \bar{y} - a\bar{x}$, temos:

$$\hat{y} = \bar{y} - a\bar{x} + a\bar{x} \Leftrightarrow \hat{y} = \bar{y}$$

Isso prova que, quando x é igual a \bar{x} , o valor previsto \hat{y} será igual a \bar{y} .

3.6 O Coeficiente de Determinação (R^2)

O coeficiente de determinação (R^2) indica o quanto da variação total da variável dependente (Y) é explicada pela variável independente (X). Essa indicação é feita em termos percentuais e para uma melhor compreensão, vamos observar a decomposição da soma dos quadrados totais (SQT). A SQT é a variação total de Y em relação a \bar{y} .

$$SQT = \sum_{i=1}^n (y_i - \bar{y})^2$$

Temos que $SQT = SQE + SQR$, onde SQE é a soma dos quadrados explicada, ou seja, é a parte da variação total que o modelo conseguiu explicar, e é dada por:

$$SQE = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

Como visto anteriormente, SQR é a soma dos quadrados dos resíduos e representa a variação que o modelo não conseguiu explicar.

$$SQR = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Sendo assim, o R^2 é definido por:

$$R^2 = \frac{SQE}{SQT} \Leftrightarrow R^2 = 1 - \frac{SQR}{SQT}$$

Na regressão linear simples, o coeficiente de determinação possui o mesmo valor do quadrado do coeficiente de correlação de Pearson: $R^2 = (r)^2$.

3.7 Pressupostos do Modelo e Diagnóstico de Resíduos

Para que os parâmetros a e b encontrados pelo MMQ sejam estatisticamente válidos, o modelo deve satisfazer alguns pressupostos básicos, que são:

- Linearidade – O modelo de regressão define uma relação linear entre a variável dependente (X) e a variável independente (Y). Essa linearidade se dá nos parâmetros.
- Independência das observações – Os resíduos não devem ter correlação entre si.
- Homocedasticidade — A variância dos erros deve ser constante para todos os valores de X .
- Normalidade da distribuição dos resíduos – Os erros devem seguir uma distribuição normal, com média zero e variância constante.

A observância destes pressupostos é conduzida através da análise do gráfico de resíduos, que permite identificar visualmente anomalias como padrões não lineares ou heterocedasticidade¹¹. A percepção de resíduos distribuídos aleatoriamente em torno de zero confirma que o modelo capturou adequadamente a estrutura dos dados, enquanto desvios sistemáticos sinalizam a necessidade de revisões na especificação do modelo.

3.8 Teste de hipóteses e Intervalo de Confiança

Uma vez validado o comportamento dos erros, procede-se aos testes de hipóteses, especialmente o teste t de Student sobre o coeficiente angular (a), visando determinar se a influência da variável independente sobre a dependente é estatisticamente significativa. A rejeição da hipótese nula indica que a inclinação da reta é significativamente diferente de zero, comprovando que a relação observada não advém do acaso amostral. Além disso, a construção de intervalos de confiança fornece uma faixa de valores prováveis para os parâmetros populacionais com um determinado nível de probabilidade. Se o intervalo de confiança para a inclinação não incluir o valor zero, reforça-se a evidência de uma relação linear real, conferindo maior robustez e precisão ao modelo.

11 A heterocedasticidade ocorre quando a variabilidade (dispersão) dos erros de um modelo de regressão não é constante ao longo das observações.

4 O MODELO DOS CAMPOS SEMÂNTICOS

O Modelo dos Campos Semânticos, criado pelo professor Rômulo Campos Lins (1955-2017), é um modelo epistemológico e uma ferramenta teórica de grande relevância para o campo da Educação Matemática, especialmente no que diz respeito à análise do discurso e da produção de significados em sala de aula. Esta teoria surgiu dos questionamentos e indagações do próprio autor, na qualidade de professor regente. Rômulo Lins relata que em particular, queria dar conta de caracterizar o que os alunos estavam pensando quando “erravam”, mas sem recorrer a esta ideia do erro (LINS, 2012, pág. 11). Ainda de acordo com o autor,

O que eu procurei fazer aqui foi oferecer, ao mesmo tempo, um modesto glossário das noções mais centrais ou mais interessantes do Modelo dos Campos Semânticos (MCS), e uma conversa minha comigo mesmo sobre coisas que me interessam em relação ao MCS. (LINS, 2012, p. 11)

Lins também ressalta que o MCS só existe em ação. Ele não é uma teoria para ser estudada, é uma teorização para ser usada (LINS, 2012, pág. 11). Sendo assim, o significado não está fixado em palavras ou conceitos de forma estática, mas emerge dinamicamente das relações estabelecidas entre enunciados, sujeitos e contextos.

O MCS não busca definir significados certos ou verdadeiros, mas compreender como esses significados são construídos, transformados e negociados em situações discursivas específicas. O termo “Campo Semântico” refere-se a um conjunto de enunciados que podem ser mobilizados por um sujeito para dar sentido a uma determinada situação.

Esses enunciados não formam um sistema fechado, mas um espaço de possibilidades semânticas em que cada novo enunciado pode modificar o campo e redefinir as relações entre seus elementos. Sendo assim, o MCS propõe uma leitura dinâmica e relacional do significado, em oposição a abordagens que o tratam de forma rígida e objetiva. O foco da análise pedagógica, mais do que avaliar se o estudante sabe algo no sentido tradicional, é compreender como é a sua produção de significado, ou seja, quais campos semânticos ele mobiliza e como se articulam. Para Lins, ensinar é oferecer modos legítimos de produção de significados e aprender é internalizar modos legítimos de produção de significados (LINS, 2008, p. 543)

Cabe enfatizar que para o MCS, o significado é sempre relativo ao contexto da enunciação. Um mesmo enunciado pode assumir sentidos distintos dependendo do lugar social do sujeito, do interlocutor e da situação discursiva.

Outra noção importante do MCS é a de espaço comunicativo que, diferente da conhecida e tradicional configuração de comunicação, em que duas pessoas falam uma em direção à outra, tem-se dois seres cognitivos falando na direção de um mesmo interlocutor (LINS, 2012, pág. 24).

Essa visão causa uma disruptura com a ideia de linguagem como um sistema

neutro de representação da realidade e introduz a noção de subjetividade e historicidade na produção de significados. O papel do professor não é mais o de transmissor de um saber fixo, e sim o de alguém que participa da construção de campos semânticos compartilhados, favorecendo a movimentação dos estudantes entre os diferentes modos de significar. Segue um resumo das noções centrais do MCS.

Campo Semântico - É um processo de produção de significado, em relação a um núcleo, no interior de uma atividade (LINS, 2012, pág. 17). É um modo legítimo de produção de significado e é legítimo porque está acontecendo. É no interior de campos semânticos que se produz conhecimento e significado e que objetos são constituídos (LINS, 2012, pág. 18). É um processo que envolve regras, que podem variar a todo momento, com estipulações e práticas que em um dado contexto, confere validade e coerência a produções de conhecimento.

Núcleo - O núcleo de um campo semântico é constituído por estipulações locais, que são, localmente, verdades absolutas, e que não requerem, localmente, justificação (LINS, 2012, pág. 26). Em se tratando de núcleo, SILVA e LINS (2013) salientam que na observação dos núcleos, numa dada atividade, podemos identificar a maneira de operar dos sujeitos bem como a lógica das operações ligadas ao processo de produção de significados para um texto (SILVA e LINS, 2013, pág. 6). O núcleo é a referência principal que traz sentido a todo o conhecimento que está sendo produzido num dado campo semântico.

Conhecimento - Um conhecimento consiste em uma crença-afirmação (o sujeito enuncia algo em que acredita) junto com uma justificação (aquilo que o sujeito entende como lhe autorizando a dizer o que diz) (LINS, 2012, pág. 12). No processo de produção de significados, existem algumas afirmações que a pessoa faz e que, tomando-as como absolutamente válidas, não sente necessidade de justificá-las. A essas crenças-afirmações, chamaremos de estipulações locais (SILVA e LINS, 2013, pág. 5).

Justificação - Não é justificativa. Não é explicação para o que digo. Não é algum tipo de conexão lógica com coisas sabidas. É apenas o que o sujeito do conhecimento (aquele que o produz, o enuncia) acredita que o autoriza a dizer o que diz (LINS, 2012, pág. 12). A justificação no MCS permite reconhecer que diferentes sujeitos podem produzir significados distintos e ter justificativas diferentes para o mesmo objeto.

Legitimidade/Verdade - Para o MCS, “verdadeiro” não é um atributo daquilo que se afirma (quando há produção de conhecimento), mas sim um atributo do conhecimento produzido. Já legitimidade aplica-se (ou não) a modos de produção de significado (LINS, 2012, pág. 21). Ainda de acordo com Lins, o que se internaliza não é conteúdo, não são conceitos, e sim legitimidades. Nesse sentido, pode-se entender que a luta pelo poder em uma cultura ou sociedade se manifesta no controle de quais são os modos de produção de significados legítimos (quais Campos Semânticos são aceitos).

Auto-Texto-Leitor - Quem produz uma enunciação é o autor. O autor fala sempre na direção de um leitor, que é constituído (produzido, instaurado, instalado, introduzido) pelo o autor. Quem produz significado para um resíduo de enunciação é o leitor. O leitor sempre fala na direção de um autor, que é constituído (produzido, instaurado, instalado, introduzido) pelo o leitor (LINS, 2012, pág. 14). Segundo Lins, uma vez que a produção de significados acontece numa enunciação, o leitor só se institui como tal na medida em que é autor, o autor. O texto é entendido como qualquer resíduo de enunciação para o qual o leitor produza algum significado (SILVA e LINS, 2013, pág. 7).

Resíduo de Enunciação - Algo com que me deparo e que acredito ter sido dito por alguém (LINS, 2012, pág. 27). No MCS, o resíduo tem importante papel na relação Autor-Texto-Leitor, pois a enunciação produzida pelo autor, gera um resíduo que pela interação com o leitor, inicia um processo de produção de significado por parte do leitor.

Interlocutor - O interlocutor é uma direção na qual se fala. Quando falo na direção de um interlocutor é porque acredito que este interlocutor diria o que estou dizendo e aceitaria/adotaria a justificação que me autoriza a dizer o que estou dizendo (LINS, 2012, pág. 27). De acordo com o autor, interlocutores são legitimidades, são seres cognitivos, não são seres biológicos. O autor diz que se falamos na direção do interlocutor, isso faz com que nunca estejamos cognitivamente sós (LINS, 2008, pág. 541).

Leitura Plausível/Leitura Positiva – A leitura plausível se aplica de modo geral aos processos de produção de conhecimento e significado; ela indica um processo no qual o todo do que eu acredito que foi dito faz sentido. A leitura positiva trata de saber de que forma uma coerência se compõe, seja na fala de uma pessoa ou num livro, mas

sem indicar falta de informação ou de reflexão por parte do autor. A leitura positiva é o oposto de uma leitura pela falta (LINS, 2012, pág. 23).

Significado/Objeto – Significado de um objeto é aquilo que efetivamente se diz a respeito de um objeto, no interior de uma atividade. Objeto é aquilo para que se produz, significado. Para o MCS não existe o significado de um objeto sem referência ao contexto em que se fala de um objeto (LINS, 2012, pág. 28).

A sala de aula deve ser um espaço para reflexão e produção de conhecimento e significado e o MCS é uma teoria que possibilita ao professor, uma leitura positiva da produção de significados sobre determinado objeto, por parte dos estudantes. Segundo Lins,

uma teoria do conhecimento tem, necessariamente, que nos permitir ler o que está acontecendo para que, eventualmente, possamos plausivelmente dizer do que é que se está falando aqui e quais são as legitimidades envolvidas. O que ela não pode fazer é o julgamento de valor, dizer o que é bom ou ruim, melhor ou pior, sob pena de se declarar responsável por um certo estado de coisas. (LINS, 2008, p. 537)

Essas noções centrais do MCS, permitem a criação de um espaço comunicativo e capacitam o docente a entender melhor a forma com que cada discente participa do processo de produção de significado.

Assim, o Modelo dos Campos Semânticos foi o referencial teórico utilizado para o desenvolvimento das atividades realizadas para introduzir o aprendizado de máquina através do modelo de regressão linear simples para estudantes do 3º ano do Ensino Médio, justamente por propiciar um ambiente de interação e estimular uma leitura das legitimidades já produzidas e as que serão produzidas pelos estudantes durante o processo.

Essa abordagem possibilita a apresentação de novos campos semânticos, como o da modelagem preditiva, onde o erro (resíduo) não é visto como algo a ser corrigido, mas como algo que faz parte e é legítimo ao modelo, sendo minimizado pelo método dos mínimos quadrados (MMQ).

4.1 A Regressão Linear como Campo Semântico

A Regressão Linear, que num primeiro momento pode parecer apenas uma ferramenta de cálculo estatístico, sob o olhar da Educação Matemática e do MCS desdobra-se em uma grande oportunidade de análise de campos semânticos, atuando

como um modo de produzir significado sobre conjuntos de dados e a correlação existente entre esses dados. Além das fórmulas utilizadas para encontrarmos os parâmetros da Reta de Regressão, o modelo de Regressão conecta a dispersão dos pontos em um gráfico a uma reta previsível, transformando a aleatoriedade aparente em uma estrutura que pode ser prevista. Sua essência está na busca por modelar a relação entre uma variável dependente (y) e uma ou mais variáveis independentes (x), assumindo que essa relação é, em sua forma mais simples, linear.

A profundidade de enxergar a Regressão Linear como campo semântico é percebida pela capacidade de mobilizar justificações em múltiplas dimensões. No plano das justificações estatísticas, a Regressão Linear Simples se apoia em fundamentos rigorosos, como o Método dos Mínimos Quadrados (MMQ), que não é meramente um algoritmo, mas sim um princípio de otimização que busca a reta que minimiza o erro quadrático total entre os valores observados e os valores previstos. Essa otimização matemática confere credibilidade e validade inferencial ao modelo, permitindo que se testem hipóteses sobre a inclinação da linha (coeficiente angular) e sua significância, o que é crucial para ir além da descrição e realizar previsões ou inferências causais.

Ao mesmo tempo, o modelo se sustenta em justificações práticas, destacando-se o ajuste visual. A Reta de Regressão, traçada no gráfico de dispersão, oferece uma representação visual imediata e intuitiva da tendência (ajuste pela reta). Um bom modelo de regressão é aquele que “parece certo” aos olhos de quem vê o gráfico de dispersão, onde a reta passa convincentemente pelo “meio” da nuvem de pontos, equilibrando a distância dos pontos acima e abaixo dela. Essa dimensão prática, de modo geral, é o primeiro passo para a validação do modelo, auxiliando na identificação de outliers (pontos de dados que diferem significativamente da maioria dos outros pontos do conjunto) ou na detecção de relações não-lineares que poderiam invalidar a suposição de linearidade.

A dimensão das justificações tecnológicas está ligada à prática moderna da Regressão Linear. O campo semântico se expande com o uso de planilhas eletrônicas, softwares estatísticos e linguagens de programação, como o Python, que tornaram a regressão acessível e aplicável a grandes conjuntos de dados. A tecnologia não apenas automatiza o cálculo complexo dos coeficientes, mas também facilita a avaliação da qualidade do ajuste, através de métricas como o R^2 (Coeficiente de Determinação) e a análise dos resíduos. Esses recursos tecnológicos permitem que o estudante se concentre mais na interpretação e na modelagem do problema e menos nos cálculos manuais, aumentando o alcance e a velocidade com que o significado é extraído dos dados.

O campo semântico da Regressão Linear encontra-se no dia-a-dia por meio das justificações cotidianas (interpretações espontâneas). O conceito subjacente de que a variação em uma coisa se relaciona com a variação em outra é uma forma de raciocínio profundamente humana. Quando o senso comum legitima afirmativas como "quanto mais se estuda, maior a nota", ou que "o preço do imóvel aumenta em função da sua área", está implicitamente empregando uma forma de raciocínio de Regressão Linear. A reta de regressão, nesse contexto, torna-se uma justificativa sobre a realidade, uma forma de explicar a correlação e covariação entre as variáveis. Essa interpretação espontânea humaniza o modelo, permitindo que resultados estatisticamente complexos sejam traduzidos em linguagem comum, facilitando a tomada de decisões em diversos campos profissionais.

A Regressão Linear Simples torna possível a construção de conhecimento e produção de significado sobre um conjunto de dados de múltiplas formas, baseada na estatística, na evidência empírica (prática), na eficiência computacional e tecnológica e no senso comum do cotidiano, devido à capacidade de transformar uma coleção de observações em uma relação funcional clara e preditiva, permitindo a análise dos dados trabalhados, de forma a orientar a tomada de decisão.

Com as premissas do MCS aplicadas ao Modelo de Regressão Linear, observa-se uma conexão entre a linguagem Matemática, que determina a reta de ajuste ideal, a linguagem gráfica, que torna visível a nuvem de pontos através do gráfico de dispersão, a linguagem estatística que testa sua validade e a linguagem da comunicação verbal, que após a análise dos dados, executa a tomada de decisão. É um modo de produção de significados, onde os cálculos são apenas uma ferramenta para uma leitura plausível da estrutura por trás dos dados.

Visando a garantia da coerência e a validade desses significados, o Campo Semântico da Regressão Linear exige uma consideração contínua de seus pressupostos fundamentais, como a normalidade dos resíduos, a homocedasticidade (variância constante dos resíduos) e a independência das observações. A não-violação desses pressupostos é a proteção estatística que assegura que a reta traçada não seja apenas um bom ajuste visual ou prático, mas uma representação válida para a população da qual a amostra foi retirada. Assim, além dos cálculos, também existe uma estruturação do pensamento sobre causalidade e correlação, oferecendo um modelo que é simultaneamente robusto e flexível o suficiente para ser aplicado a uma gama de fenômenos.

O Campo Semântico da Regressão Linear une a modelagem matemática com a interpretação humana, gerando um conhecimento que é objetivamente quantificável e

subjetivamente significativo e o papel dos resíduos, que são a diferença entre os valores observados e os valores previstos pelo modelo, é muito rico para a produção de significados sobre os modelos preditivos de Regressão. Os resíduos não são meramente "erros" a serem minimizados e sim uma fonte crucial de informação. Eles representam a porção da variação na variável dependente (y) que o modelo linear não conseguiu explicar. Portanto, o que o modelo "deixa de explicar" se torna tão significativo quanto o que ele explica.

Além disso, a Regressão Linear serve como base didática para a compreensão de modelos estatísticos mais avançados. Modelos de regressão mais complexos (como Regressão Logística ou Modelos Lineares Generalizados - MLGs) são extensões conceituais do modelo linear. Produzir significado sobre o Modelo de Regressão Linear Simples é uma justificativa pedagógica que abre as portas para a modelagem estatística e o Aprendizado de Máquina.

A Regressão Linear como campo semântico estabelece um espaço que favorece a movimentação entre diferentes modos de significar, onde a linguagem matemática/estatística, como a precisão do R^2 , encontra a intuição do bom ajuste visual, e onde a linguagem tecnológica das planilhas eletrônicas ou linguagens de programação interagem com as interpretações cotidianas sobre como as variáveis se relacionam e se ajustam por uma reta, tornando o significado acessível e aplicável.

5 METODOLOGIA E ATIVIDADES PROPOSTAS

Para a elaboração das atividades propostas, o referencial teórico utilizado foi o Modelo dos Campos Semânticos (MCS), do professor Rômulo Campos Lins. Segundo LINS (2012, pág. 17), Campo Semântico é “um processo de produção de significado, em relação a um núcleo, no interior de uma atividade”. Ainda de acordo com LINS (2012, pág. 18), “um campo semântico indica um modo legítimo de produção de significado”. É legítimo justamente por ser um processo que está acontecendo e sendo assim, está em constante transformação. Pela visão do MCS, a produção de significados é o ponto central de toda aprendizagem e no processo de produção de significados existem algumas afirmações feitas pelo sujeito que são tidas como válidas, sem necessidade de justificativa. Essas afirmações são chamadas de estipulações locais e o conjunto de estipulações locais contidas no interior de uma atividade é chamado de núcleo.

A partir das premissas do MCS aplicadas na produção de significados envolvendo o aprendizado de máquina e o modelo de regressão linear simples, uma seqüência de tarefas extracurriculares foi desenvolvida visando gerar reflexões por parte dos estudantes em relação ao tema abordado e suas aplicações no mundo contemporâneo.

A sequência de atividades propostas para estudantes do 3º ano do ensino médio é composta por cinco encontros de 100 minutos (2 horas-aula) cada. O objetivo foi 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. Foram abordados temas como a origem do termo “inteligência artificial”, desenvolvimento 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).

Para o bom desenvolvimento das atividades, foram revistos conceitos matemáticos e estatísticos como função polinomial do 1º grau, gráficos de dispersão, médias aritméticas, desvio padrão e variância, para nivelamento dos estudantes participantes.

A instituição escolhida para a realização das tarefas foi a escola estadual Bom Pastor, localizada no município de Belford Roxo, no estado do Rio de Janeiro. Tal escolha prendeu-se ao fato de ser um município que é frequentemente apontado como uma das cidades com os piores Índices de Desenvolvimento da Educação Básica (Ideb) no Brasil. Um estudo da Federação das Indústrias do Estado do Rio de Janeiro (Firjan), realizado em 2023 (IFDM/23) também coloca Belford Roxo na última posição em desenvolvimento socioeconômico no estado do Rio de Janeiro, considerando indicadores de saúde, educação e geração de emprego e renda.

Quadro 1: IFDM/2023

Melhores colocados	Piores colocados
1º Rio de Janeiro	92º Belford Roxo
2º Volta Redonda	91º Japeri
3º Resende	90º São Gonçalo
4º Macaé	89º Nova Iguaçu
5º Piraí	88º Queimados
6º Itaperuna	87º São João do Meriti
7º Petrópolis	86º São Francisco de Itabapoana
8º Porto Real	85º Duas Barras
9º Teresópolis	84º Cardoso Moreira
10º Areal	83º Varre-Sai

As atividades foram aplicadas após a liberação do Comitê de Ética em Pesquisa do Colégio Pedro II, disponível na Plataforma Brasil, através do Parecer nº 7.631.322, e as inscrições para essas atividades extracurriculares foram realizadas através de um formulário que foi preenchido e assinado pelos estudantes e seus responsáveis, onde tornou-se explícito todo o procedimento e atividades propostas. As atividades aconteceram no contraturno das aulas regulares, com 27 inscrições confirmadas. O cuidado com o número de inscritos justificou-se pela atenção dispensada à produção de significados de cada estudante e por termos apenas a presença do professor/pesquisador desta dissertação atuando como mediador. Ao final de cada encontro, uma folha de tarefas foi preenchida para que a produção dos discursos de cada estudante fosse documentada para posterior análise de resultados.

1º Encontro

Destinou-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 foi 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. Os 30 minutos finais serviram para a aplicação de uma folha de tarefas visando documentar a produção de significados e observar eventuais resíduos de enunciação. Segundo LINS (2012, pág. 27), resíduo de enunciação é “algo com que me deparo e que acredito ter sido dito por alguém”.

Plano de Aula do 1º Encontro

- **Conteúdo 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:**

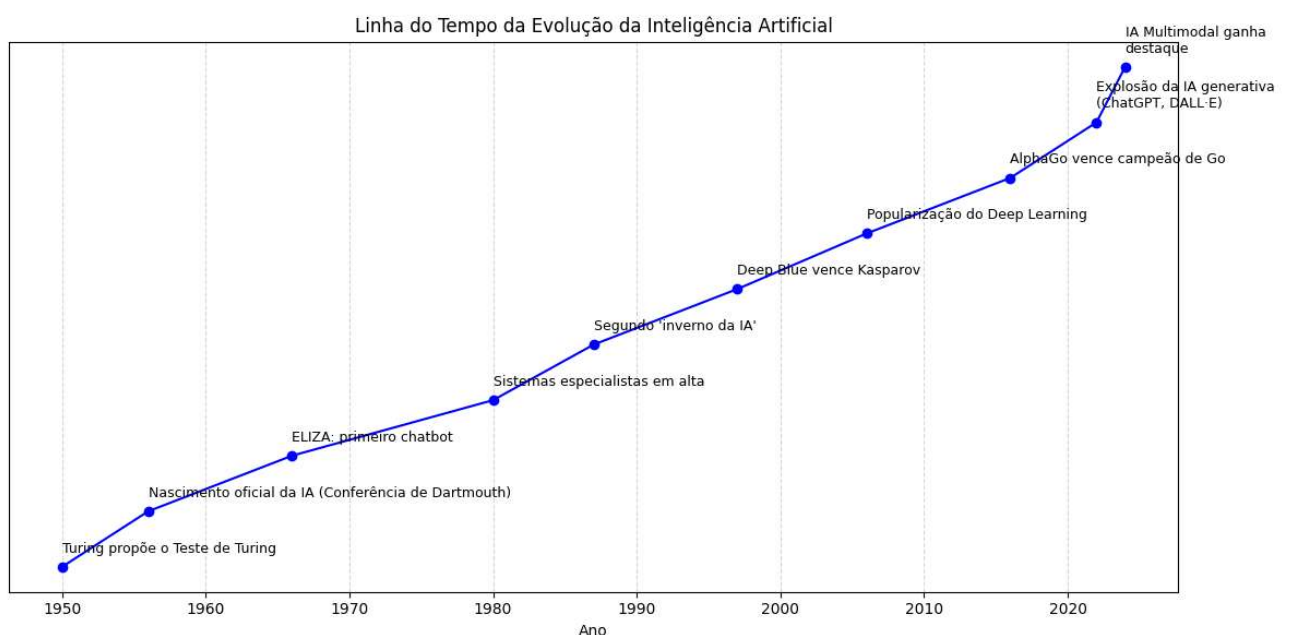
Aula interativa visando o envolvimento ativo dos estudantes, com a utilização do projetor como recurso multimídia, tendo o professor a função de mediador do conhecimento produzido e facilitador do aprendizado, de acordo com o momento da atividade.

A atividade seguiu uma organização didática na qual o primeiro momento foi de apresentação do tema, pois os conceitos necessários para um debate e exposição de ideias mais aprofundadas em torno dos conteúdos abordados foram sendo desenvolvidos e além disso, como o conhecimento sobre o assunto varia de indivíduo para indivíduo, tornou-se imprescindível um nivelamento para posterior análise da produção de significado dos estudantes. Sendo assim, algumas definições sobre Inteligência Artificial foram 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, foram explorados eventos históricos que foram relevantes no avanço da Inteligência Artificial ao longo do tempo e para tal, utilizamos a linha do tempo constante na figura 9, como apoio didático.

Figura 9: Marcos históricos da evolução da Inteligência Artificial



Fonte: O autor

Com o intuito de despertar a curiosidade e o interesse dos estudantes, maior ênfase foi 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 foi 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 10), 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 10: Reportagem do Jornal O Globo do dia 28/07/2025



Fonte: Jornal O Globo

Na sequência da atividade, foram mostrados casos de uso da IA de forma a prejudicar pessoas físicas e jurídicas além de estimular discriminação. O uso ético e socialmente correto da IA foi discutido, 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 foram 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 2º Encontro. Os momentos finais (30 minutos) deste 1º Encontro foram utilizados para a aplicação de uma folha de tarefas que servirá de base para a posterior análise de resultados. Segue a folha de tarefas do 1º Encontro.

Folha de tarefas do 1° Encontro

- 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

Teve 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 foram revisitados assim como a função polinomial do 1° grau, com o objetivo de nivelamento dos estudantes.

O segundo momento da aula teve 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 foi apresentada e serviu de apoio para a introdução dos conceitos e realização dos cálculos. Os materiais de apoio utilizados foram o quadro branco, caneta para quadro branco e projetor. Para agilizar, uma calculadora também foi utilizada.

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

Plano de Aula do 2° Encontro

- **Conteúdo 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 se 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 3º ano do Ensino Médio de uma escola estadual e verificamos suas notas em Matemática e Física no 1º 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: O 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:

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

Sendo: \bar{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 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:

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

$$\text{Desvio Padrão } (s) = \sqrt{s^2}$$

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*

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

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

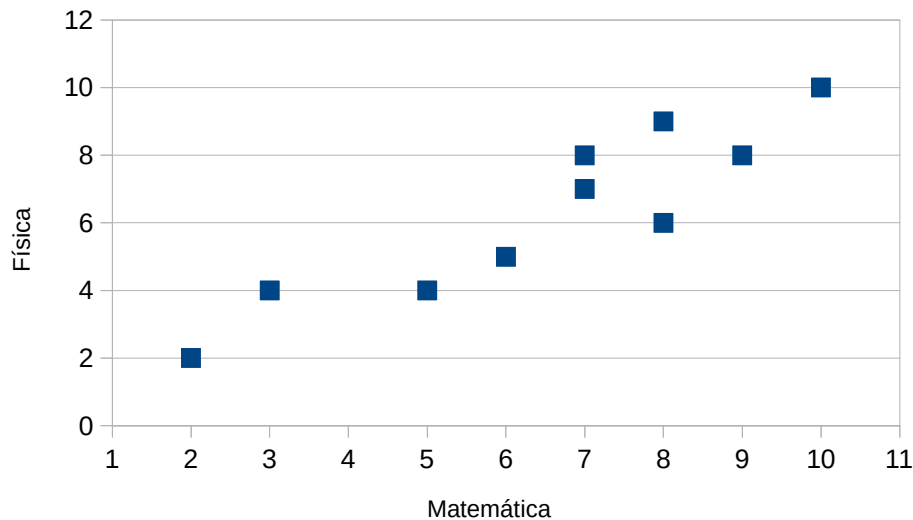
- *Desvio padrão das notas em Física*

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

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 dispersão¹² foi construído (Figura 12), 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 11: Gráfico de Dispersão das Notas de Matemática e Física



Elaborado pelo 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 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:

12 Alguns autores denominam diagrama de dispersão.

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

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 $0,3 \leq |r| < 0,6$, há uma correlação relativamente fraca entre as variáveis;
- Se $0 \leq |r| < 0,3$, a correlação é muito fraca e praticamente nada pode ser concluído sobre a relação entre as variáveis;
- Se $0,6 \leq |r| < 1$, 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 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 1º 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:

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

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:

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

$$a = 0,9145$$

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

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

$$b = 0,3557$$

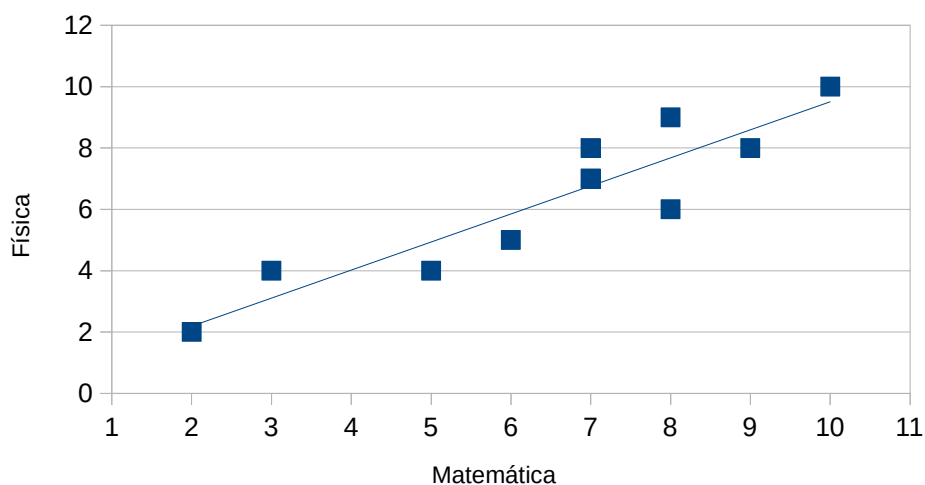
Arredondando para duas casas decimais tem-se $a=0,91$ e $b=0,35$.

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 $\hat{y}=0,35$.

A figura 12 representa a reta de regressão linear encontrada:

Figura 12: 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 do 2º Encontro

- 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

- 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.
- Represente os dados em um diagrama de dispersão.
- Apenas observando o gráfico de dispersão, podemos dizer que existe correlação entre os dados? Justifique sua resposta.
- Determine a reta de regressão linear e represente a reta no diagrama de dispersão.
- Calcule o Coeficiente de Correlação Linear de Pearson e o Coeficiente de Determinação.
- Use a reta de regressão encontrada para prever a nota final de um estudante que estuda 7 horas por dia.

3º 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 foram 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 mostramos formas de encontrar a reta de regressão com comandos da própria planilha. Os alunos aprenderam 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 foi apresentada, com os coeficientes de correlação linear e de determinação exibidos na tela. Com o uso da planilha, a análise dos dados pôde ser feita de forma mais eficiente. Diferentes dados foram trabalhados pelos estudantes para que eles pudessem 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 da escola, mas em último caso, pode ser feita na sala de aula com a utilização de notebooks ou tablets da própria escola. Devido à quantidade de alunos inscritos na atividade, grupos foram formados para uso compartilhado dos computadores. A folha de tarefas foi entregue no momento em que os alunos estavam usando os computadores.

Plano de Aula do 3º Encontro

- **Conteúdo 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

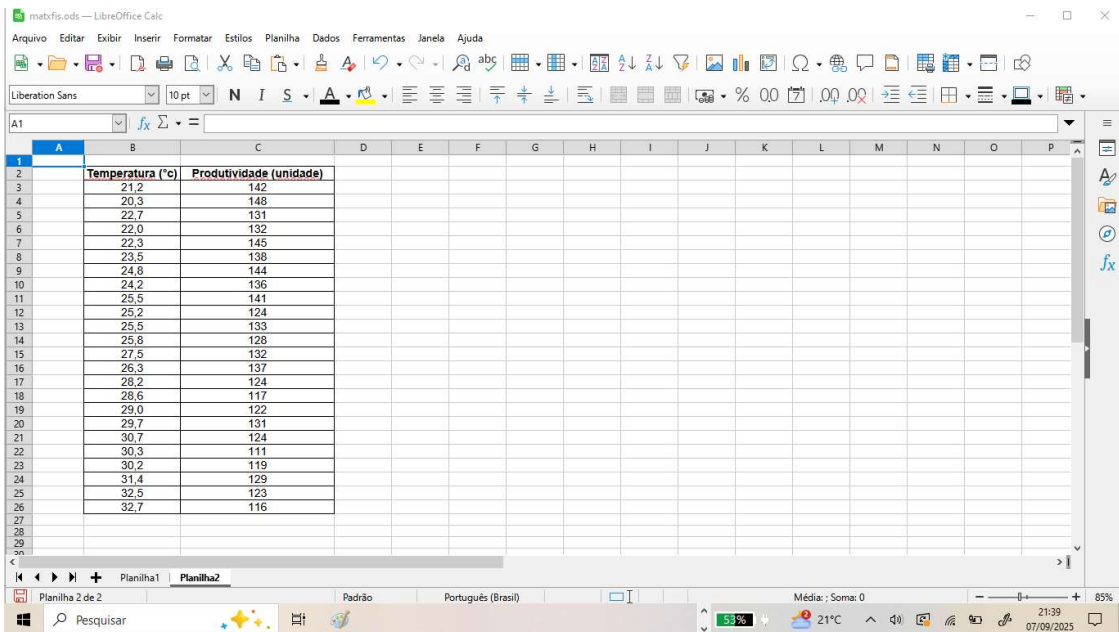
Observações	Temperatura (°c)	Produtividade (unidade)
1	21,2	142
2	20,3	148
3	22,7	131
4	22	132
5	22,3	145
6	23,5	138
7	24,8	144
8	24,2	136
9	25,5	141
10	25,2	124
11	25,5	133
12	25,8	128
13	27,5	132
14	26,3	137

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, conforme mostra a figura 13. 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.

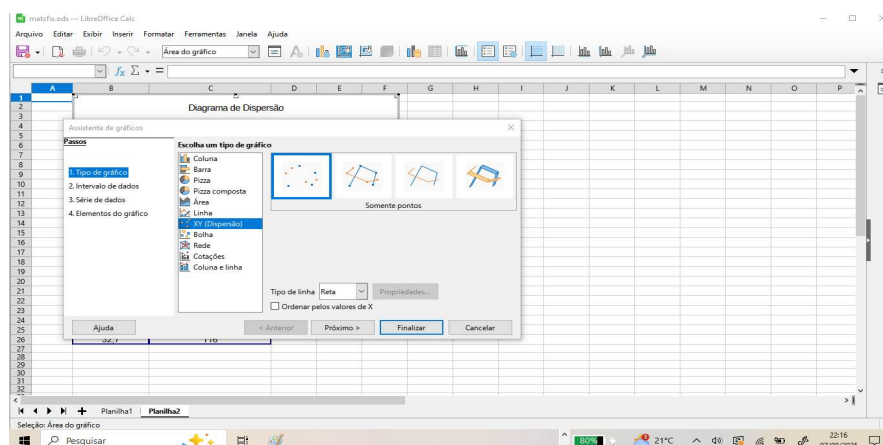
Figura 13: Inserção de dados na planilha



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).

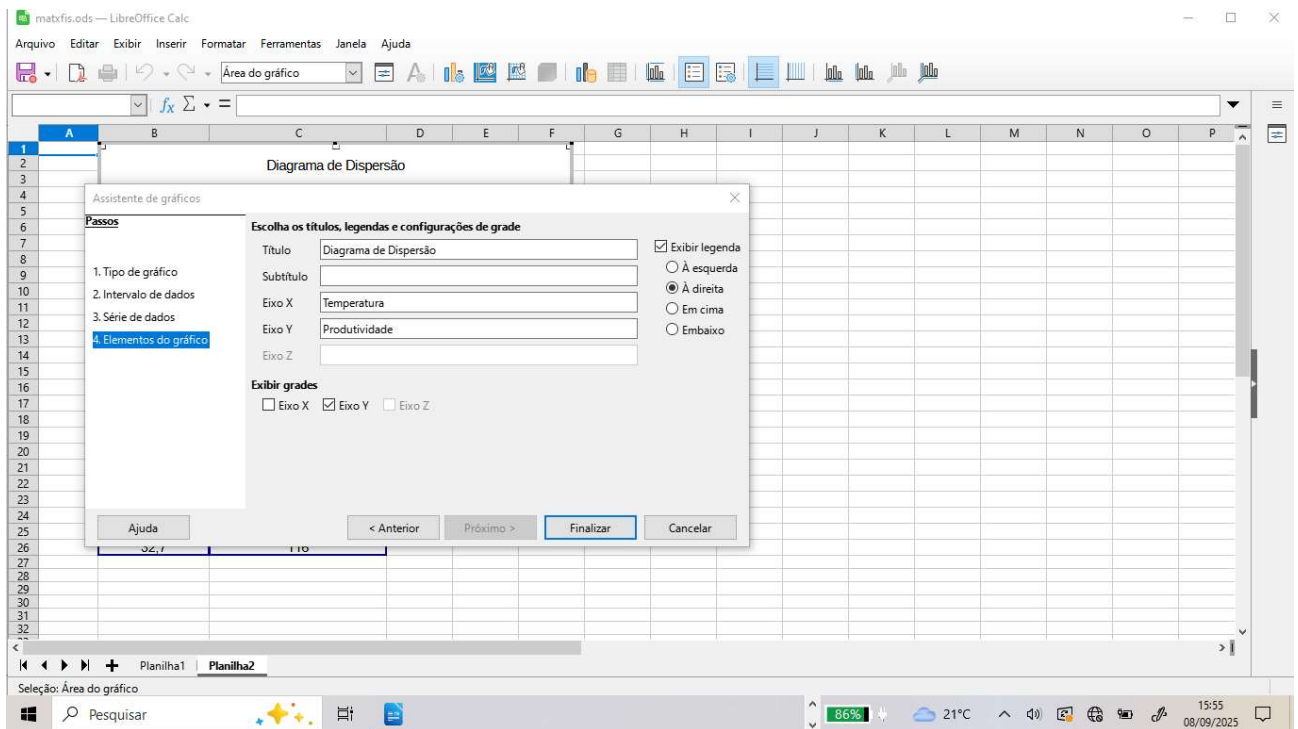
Figura 14: Inserção do gráfico de dispersão na planilha



Fonte: O autor

Ainda no Assistente de gráficos, clique no passo 4 - Elementos do gráfico para escolha do título, legenda e configuração de grade, conforme a figura 15.

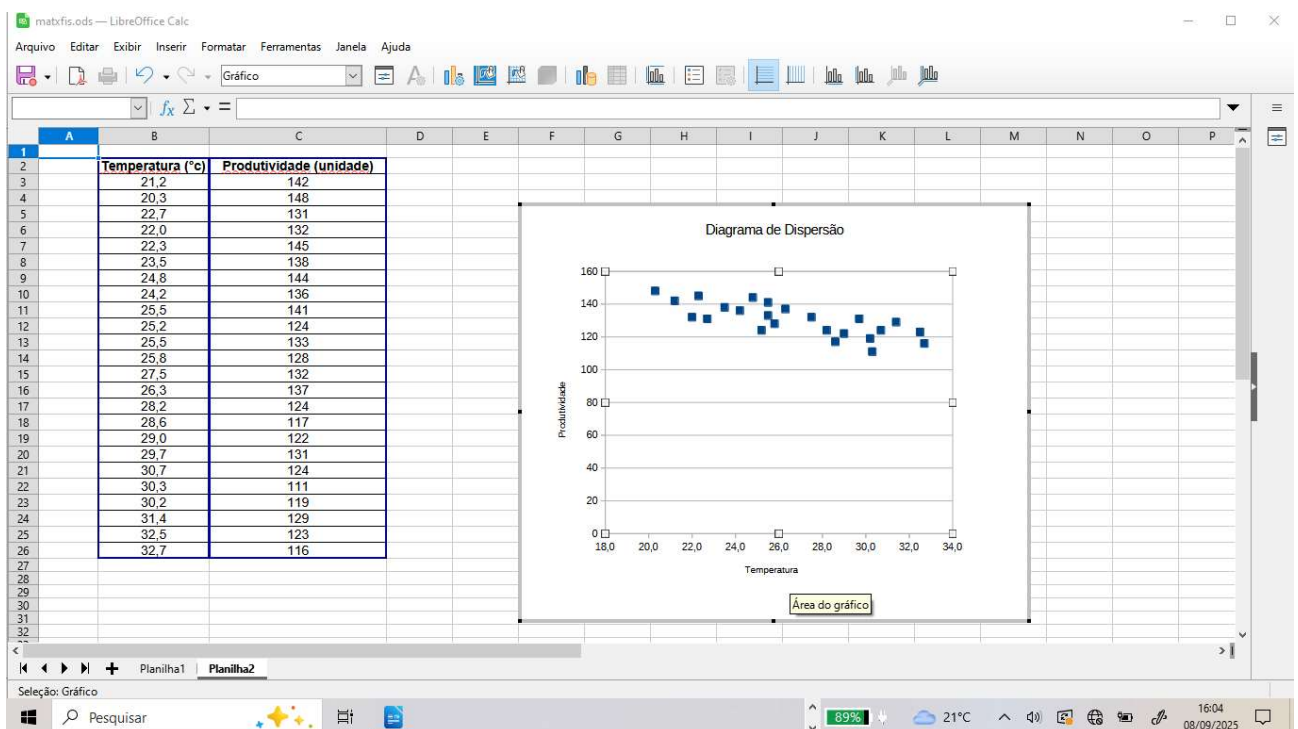
Figura 15: Configuração do gráfico de dispersão na planilha



Fonte: O autor

Na sequência, clicando em finalizar, o gráfico poderá ser visualizado, como mostra a figura 16.

Figura 16: Visualização do gráfico de dispersão na planilha

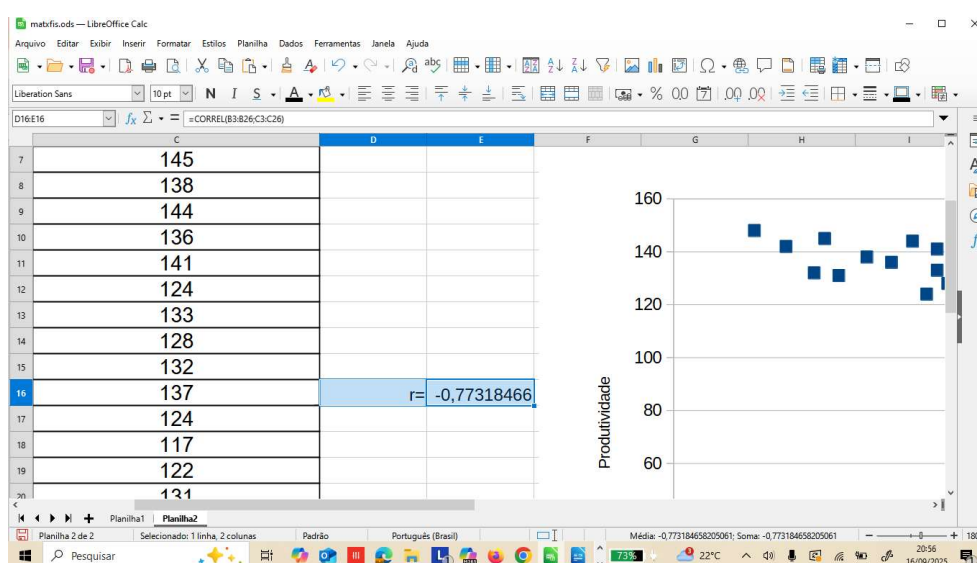


Fonte: O autor

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 17).

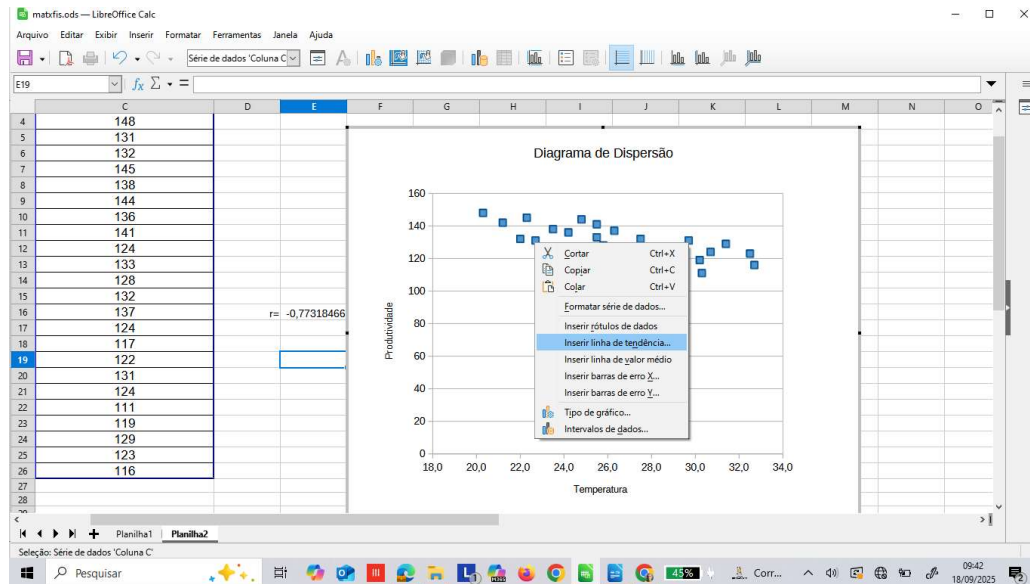
Figura 17: Visualização do coeficiente de correlação na planilha



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 abri-lo para edição e após, clicamos sobre um dos pontos de dados e no menu, escolhemos inserir linha de tendência (figura 18).

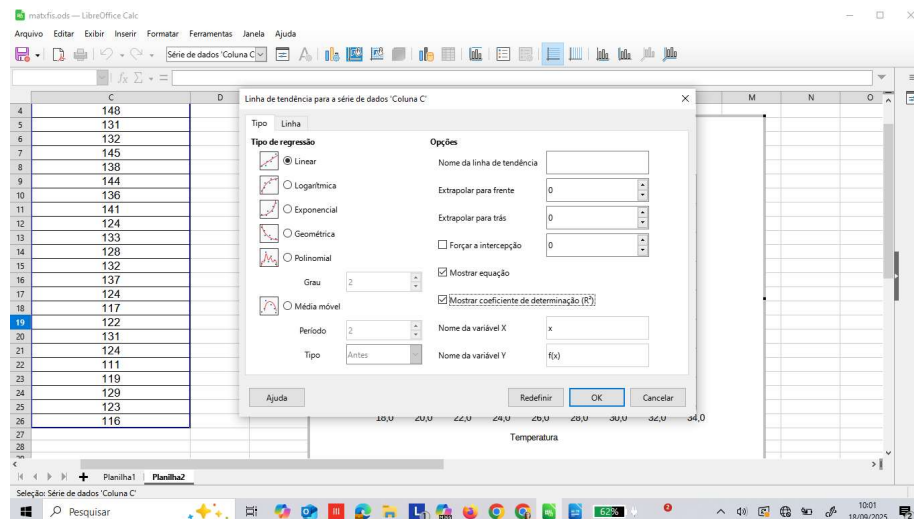
Figura 18: Inserção da linha de tendência na planilha



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 19).

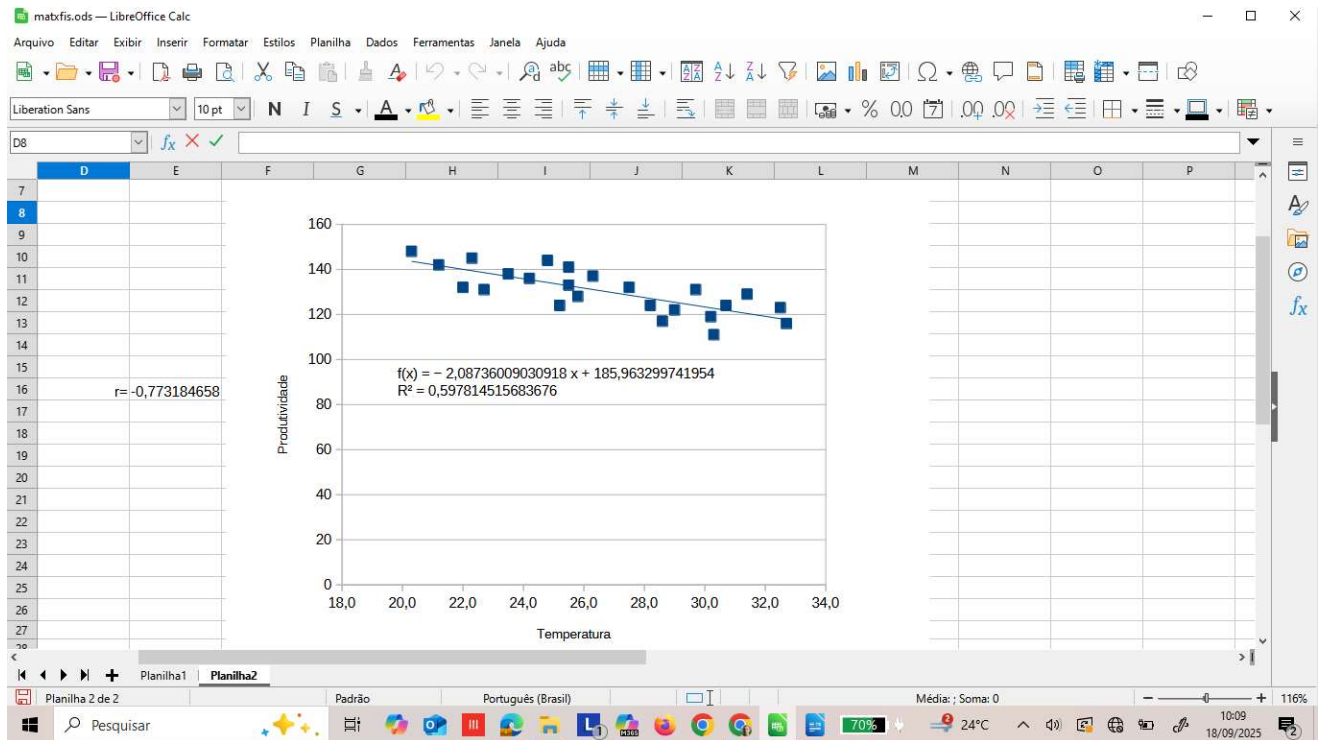
Figura 19: Inserção do coeficiente de determinação na planilha



Fonte: O autor

Ao clicar em OK, o gráfico surgirá com a reta de regressão linear e o coeficiente de determinação na tela, conforme exibido na figura 20.

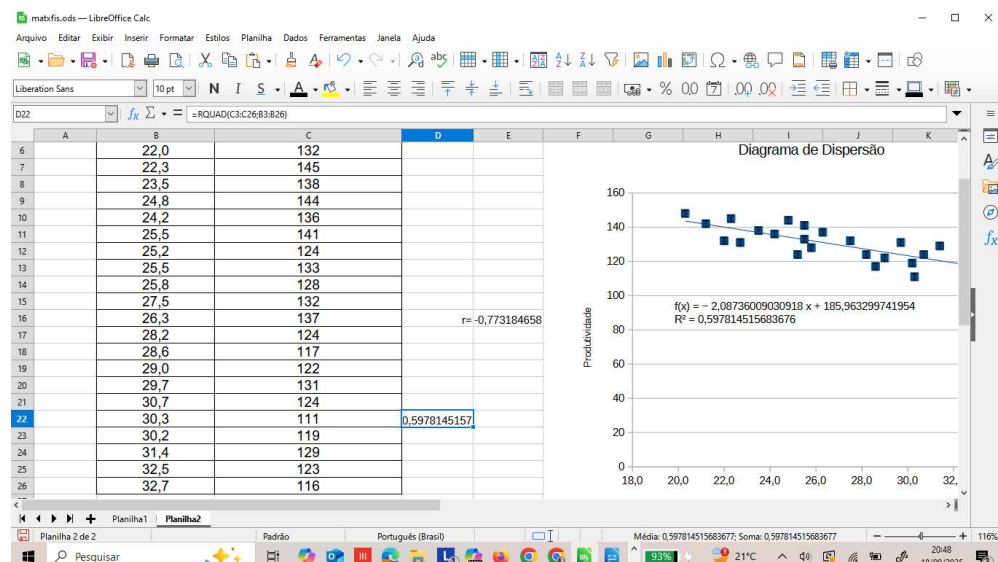
Figura 20: Exibição da reta de regressão e do coeficiente de determinação na planilha



Fonte: O autor

O coeficiente de determinação também pode ser encontrado através da função RQUAD. Na célula escolhida digitamos: $=RQUAD(\text{Dados } y; \text{Dados } x)$, 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 21).

Figura 21: Outra forma de exibir o coeficiente de determinação na planilha

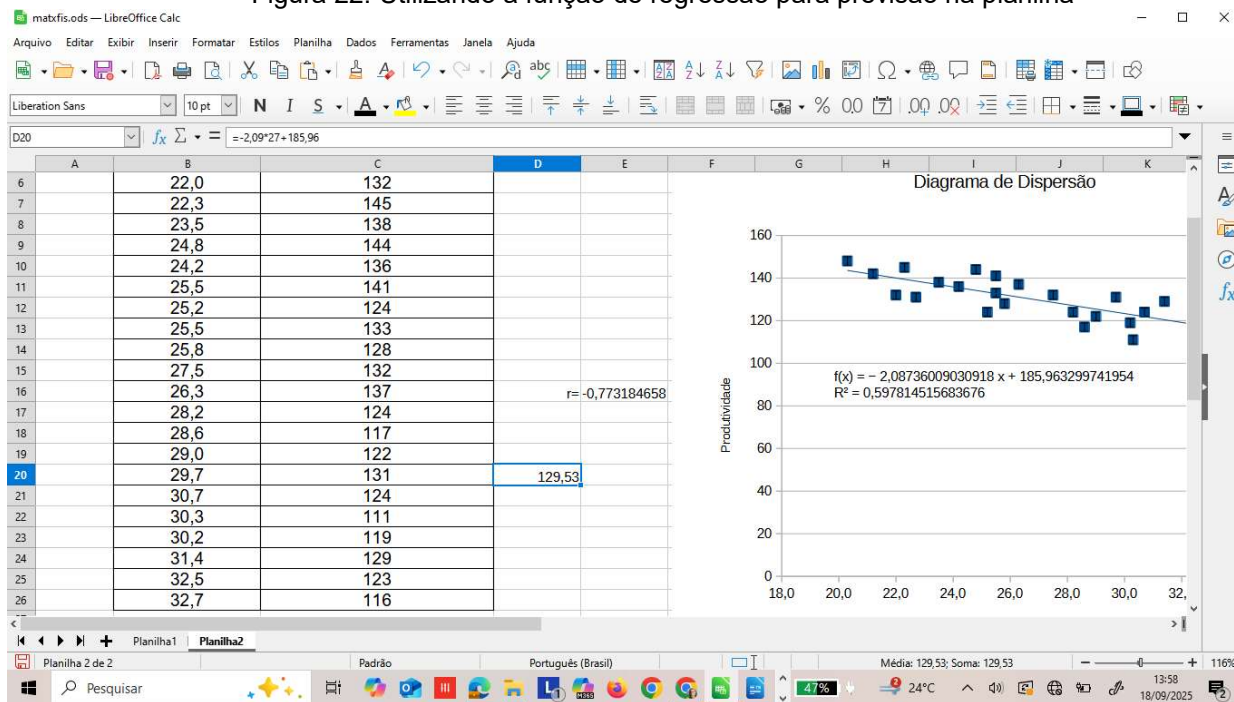


Fonte: O autor

Ajustando os coeficientes angular, linear e de determinação para duas casas decimais, temos a função de regressão igual a $y = -2,09x + 185,96$ 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 22).

Figura 22: 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 do 3º Encontro

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:

- Organize os dados na planilha.
- Represente os dados em um gráfico de dispersão.
- Calcule o coeficiente de correlação.
- Determine a reta de regressão linear e represente a reta no gráfico de dispersão.
- Calcule o coeficiente de determinação e interprete o valor encontrado.
- Use a reta de regressão encontrada para prever o gasto com alimentação de uma família com renda de 110 unidades monetárias.
- 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?

4º Encontro

Os estudantes foram 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 pôde 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 foi muito parecida com a do 3º Encontro.

Plano de Aula do 4º Encontro

- **Conteúdo 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 3º 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

Fonte: O Autor

Após abrir a versão online do GeoGebra Classic, no menu superior, clique nas três barras horizontais que estão à direita, como mostra a figura 23.

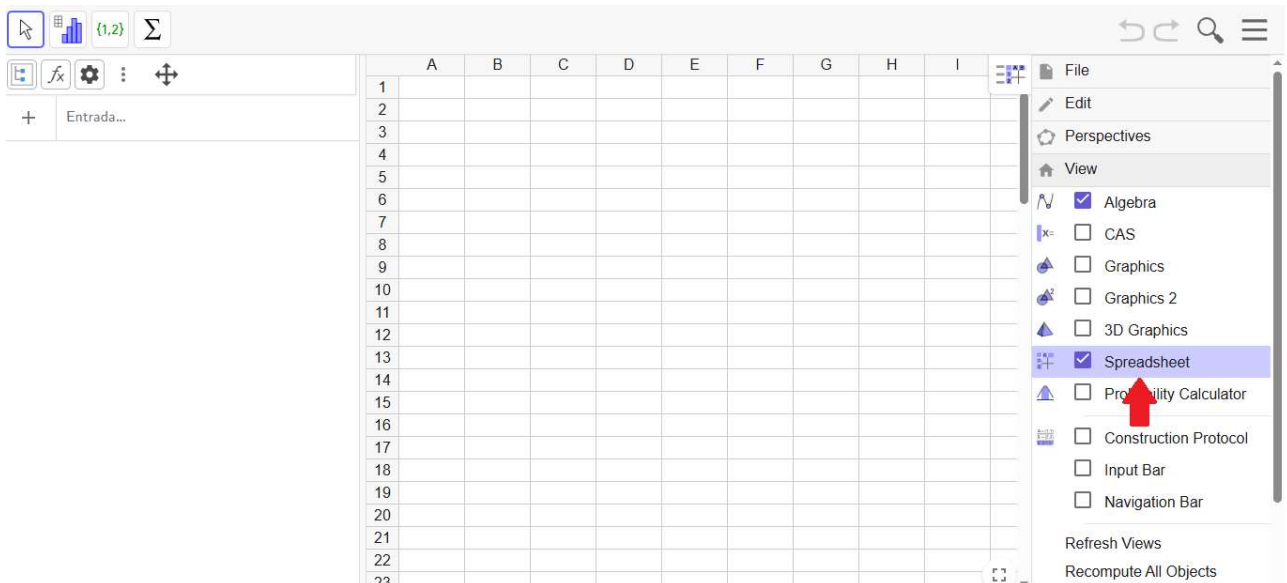
Figura 23: Menu do GeoGebra



Fonte: O autor

No menu que se abre, escolha a opção de exibição da planilha (Spreadsheet) e a planilha aparecerá (figura 24).

Figura 24: 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 25).

Figura 25: Inserção de dados na planilha do GeoGebra

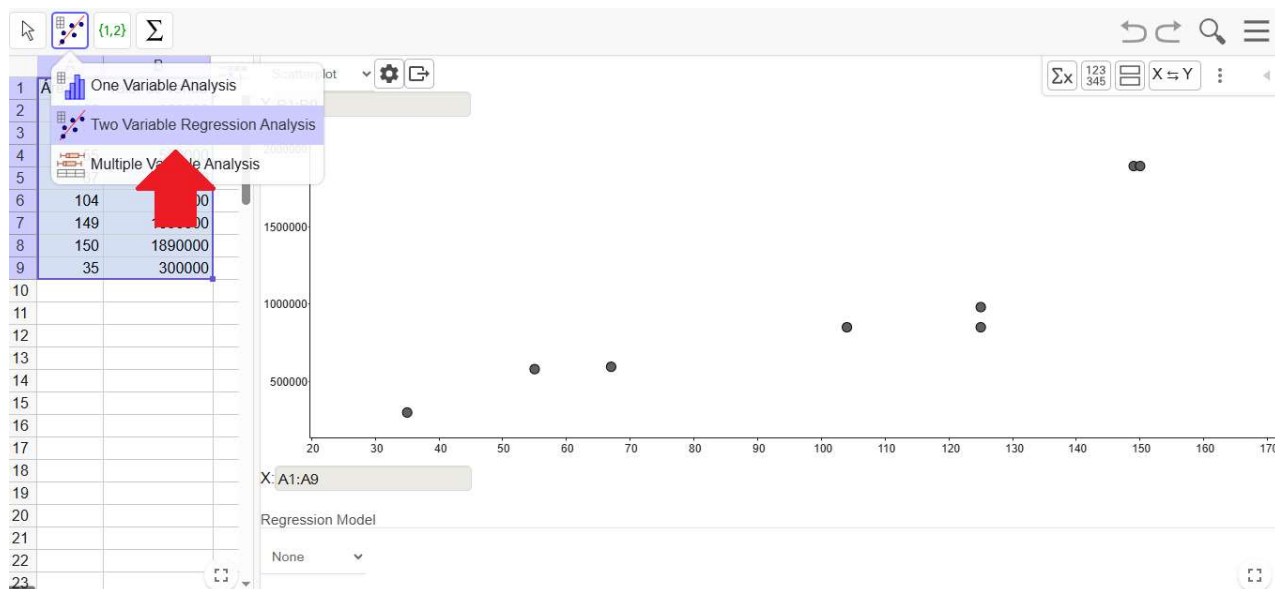
	A	B	C	D	E	F	G	H	I	J
1	Área	Valor de venda								
2	125	850000								
3	125	980000								
4	55	580000								
5	67	595000								
6	104	850000								
7	149	1890000								
8	150	1890000								
9	35	300000								
10										

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, como exibido na figura 26.

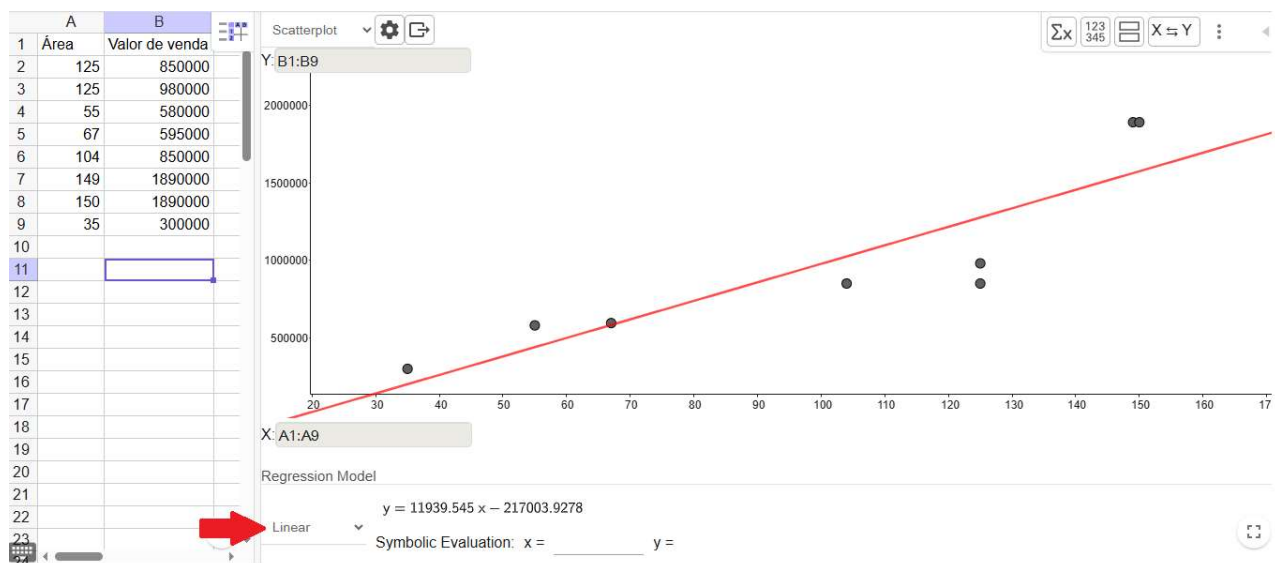
Figura 26: Exibição do gráfico de dispersão no GeoGebra



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 27).

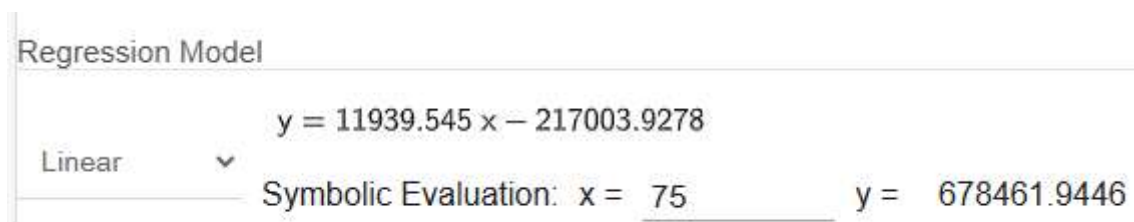
Figura 27: Inserção da linha de tendência no GeoGebra



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 , conforme apresentado na figura 28.

Figura 28: Fazendo previsão com o Geogebra

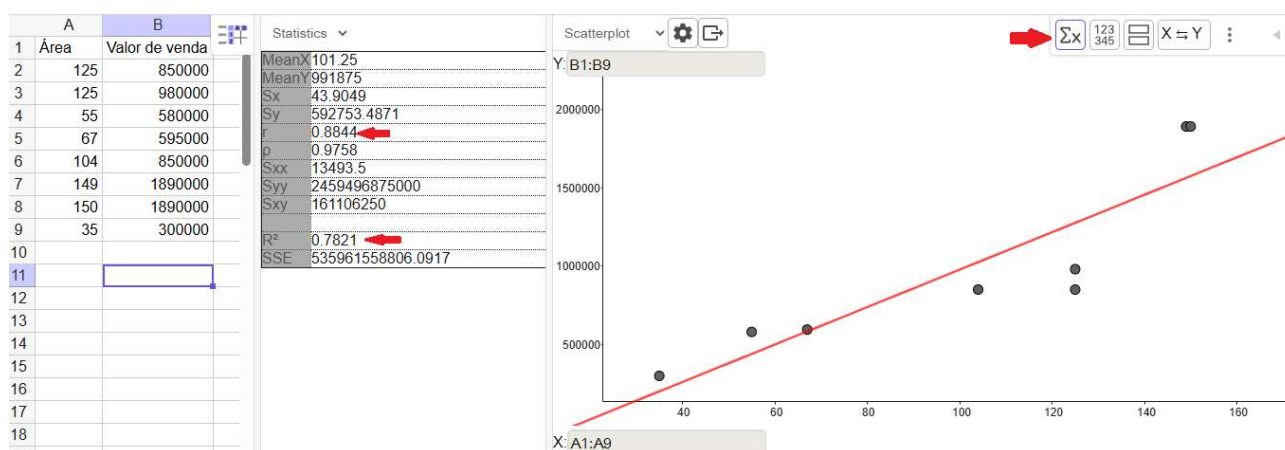


Fonte: O autor

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 29).

Figura 29: Exibição dos dados estatísticos no GeoGebra

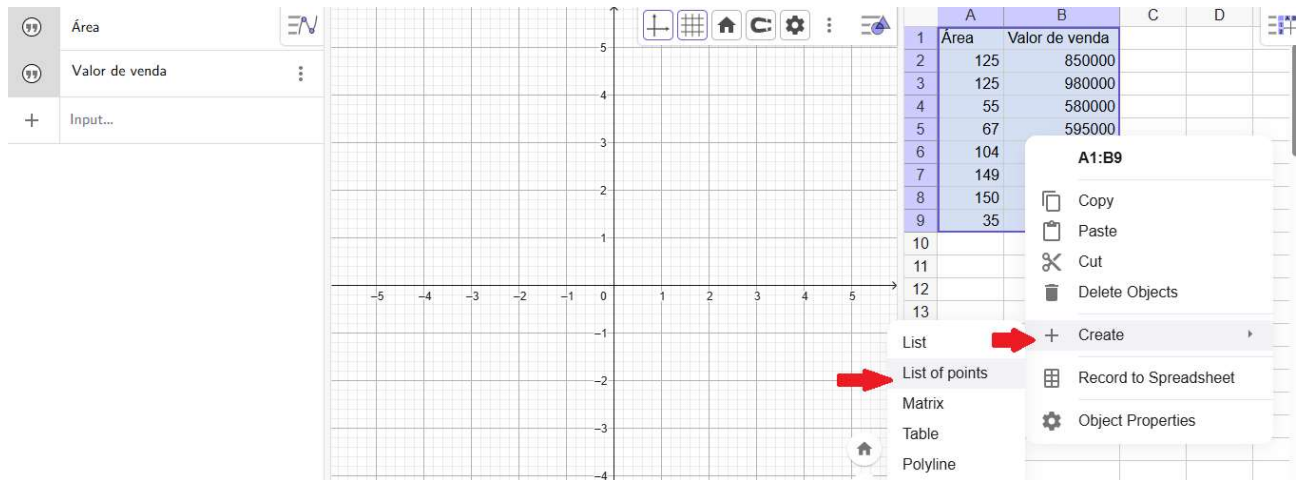


Fonte: O autor

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), como mostra a figura 30.

Figura 30: Seleção dos dados da planilha



Fonte: O autor


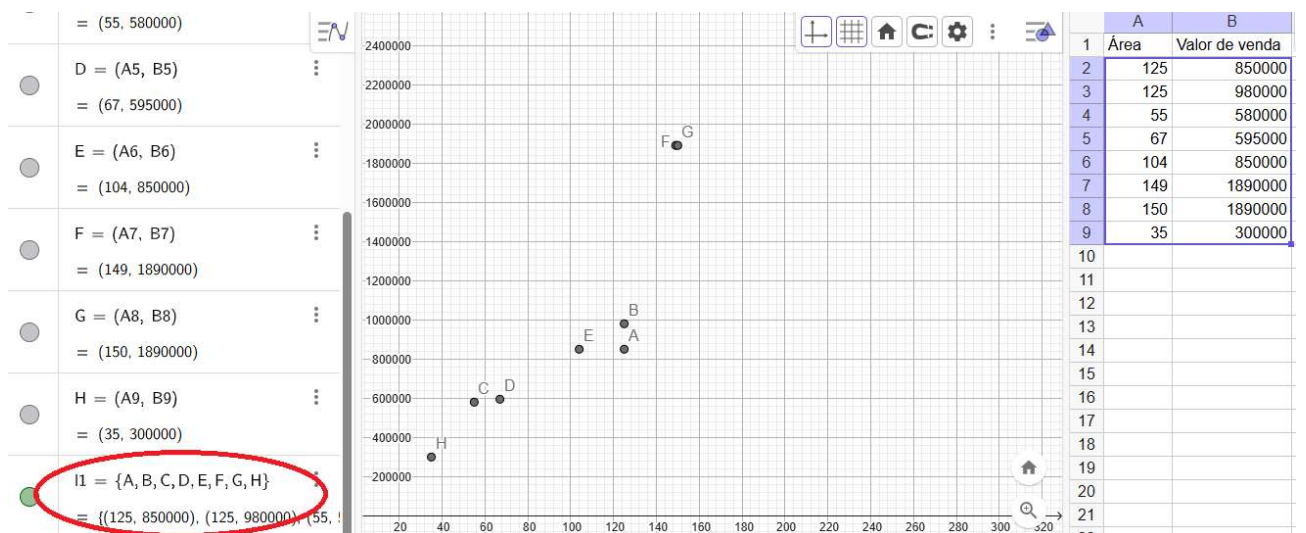
Assim, o GeoGebra irá gerar uma lista na janela de álgebra com todos os pontos selecionados e esses pontos também aparecerão na janela de visualização (gráfico 2D) (figura 31). Cabe destacar que um ajuste nos valores de máximo e mínimo dos eixos x e y foi feito para que as coordenadas aparecessem na tela, pois as ordenadas possuíam valores altos. Esse ajuste é feito nas configurações da janela de visualização,  no ícone

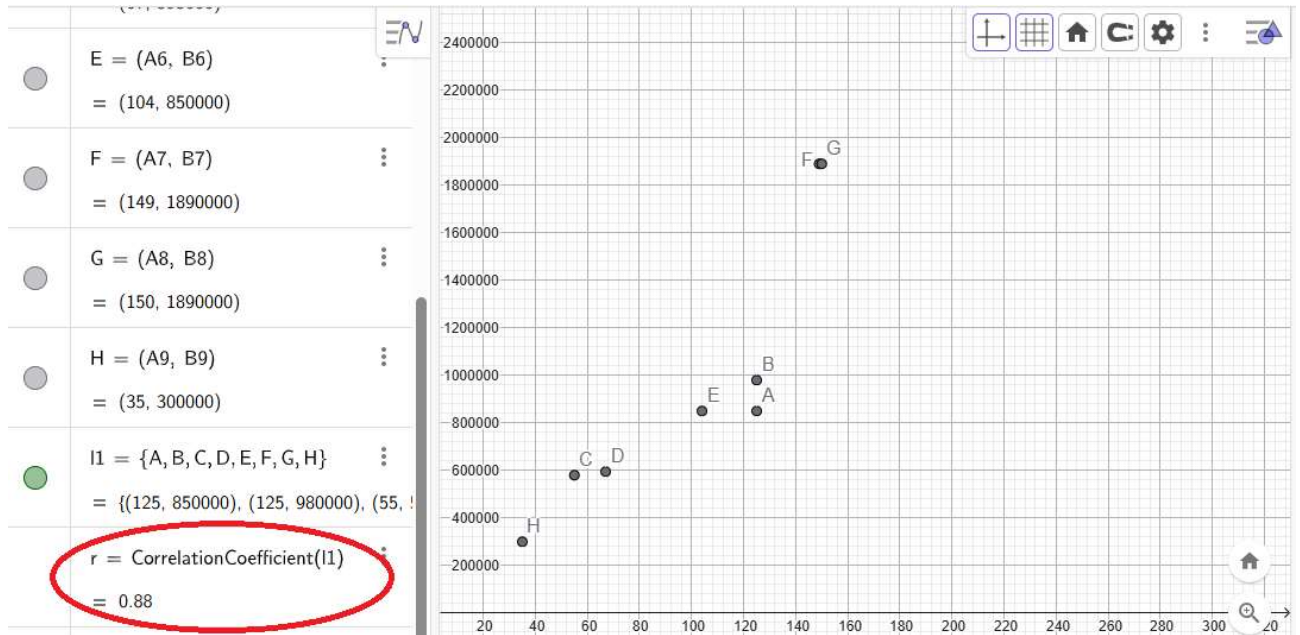
Figura 31: Lista de pontos na janela de álgebra e na janela de visualização



Fonte: O autor

Para encontrarmos o coeficiente de correlação linear (r) entre os valores, usamos na janela de álgebra, o comando *CoeficienteCorrelacao(I1)* (CorrelationCoefficient), como mostra a figura 32.

Figura 32: 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(I1)* (FitLine), que retornará a reta de regressão na janela de visualização e a função de regressão na janela de álgebra (figura 33).

Figura 33: Exibição da função de regressão e da reta de regressão no GeoGebra

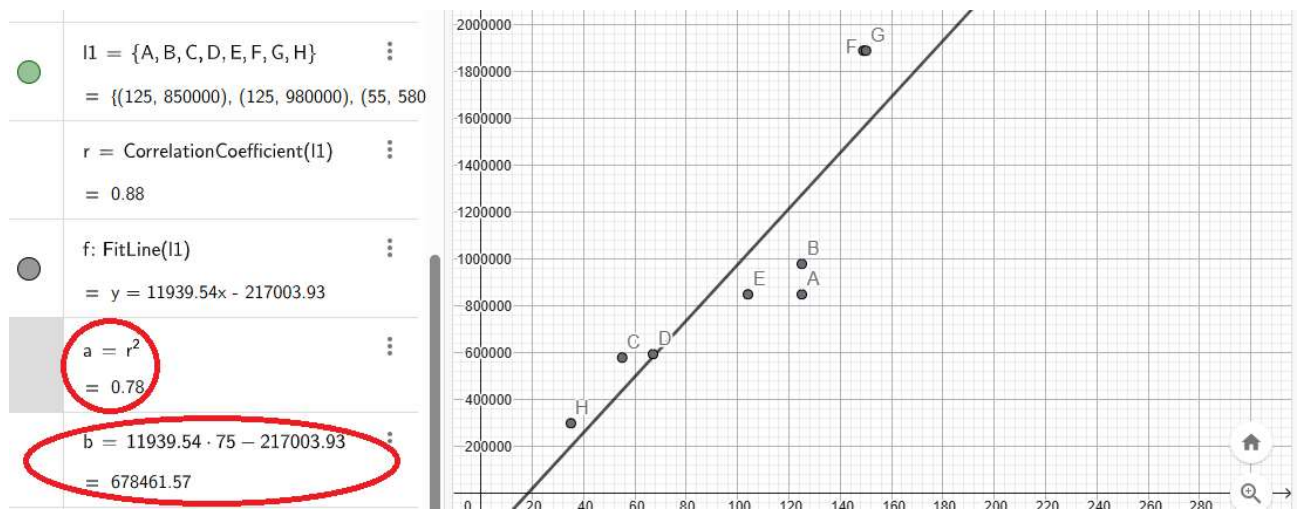


Fonte: O autor

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 $75m^2$ tem valor de venda de R\$ 678.461,57, como mostra a

figura 34.

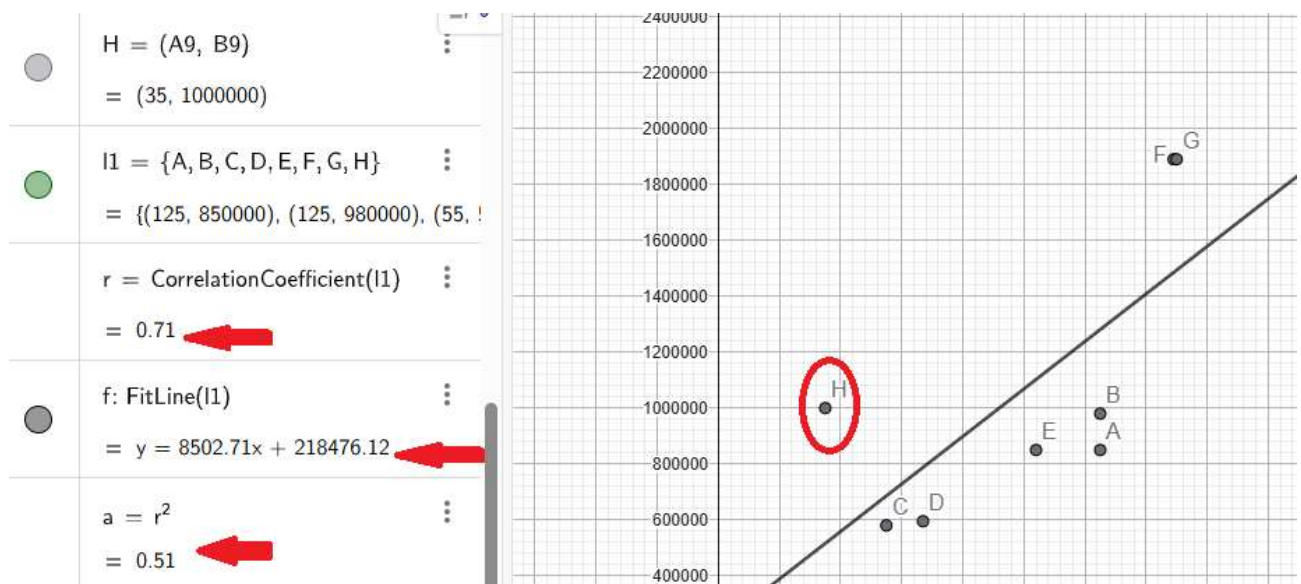
Figura 34: Exibição do coeficiente de determinação e da previsão de valor de venda no GeoGebra



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 $35m^2$, 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 35).

Figura 35: 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, trazendo prejuízo na análise e nas previsões. Segue folha de tarefas do 4º Encontro.

Folha de tarefas do 4° Encontro
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:

- Escolha uma forma de representar graficamente esse conjunto de dados.
- Você considera que o coeficiente de correlação encontrado demonstra uma boa correlação entre as variáveis? Justifique sua resposta.
- Determine a reta de regressão linear que melhor se ajusta aos dados da amostra.
- Calcule o coeficiente de determinação e interprete o valor encontrado.
- 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 tiveram contato com o Python, que é uma linguagem de programação de alto nível¹³, 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 bibliotecas¹⁴ e frameworks¹⁵ disponíveis.

As atividades de programação em Python propostas para este encontro foram 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, ferramenta de IA integrada, 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 aprenderam 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 foi disponibilizada aos estudantes após a aula teórica, realizada com o apoio do projetor. Foram 30 minutos de apresentação da linguagem através de exemplos e o tempo restante foi utilizado para que os estudantes pudessem executar as atividades de modelagem de regressão linear em Python. O objetivo principal desse encontro foi apresentar a linguagem para os estudantes e mostrar que com poucos comandos, um modelo de regressão pode ser realizado.

13 Linguagem de programação projetada para ser fácil de entender e escrever, usando sintaxe próxima da linguagem humana.

14 Coleção de códigos utilizados com frequência pelos desenvolvedores/programadores.

15 Conjunto de códigos prontos que servem como estrutura para auxiliar no desenvolvimento da programação.

Plano de Aula do 5º Encontro

- **Conteúdo 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 iniciou-se com a apresentação da plataforma Google Colab que foi o ambiente onde a programação em Python realizou-se. O site utilizado foi <https://colab.research.google.com/notebooks/welcome.ipynb?hl=pt-BR>. A figura 36 mostra a tela inicial do Google Colab.

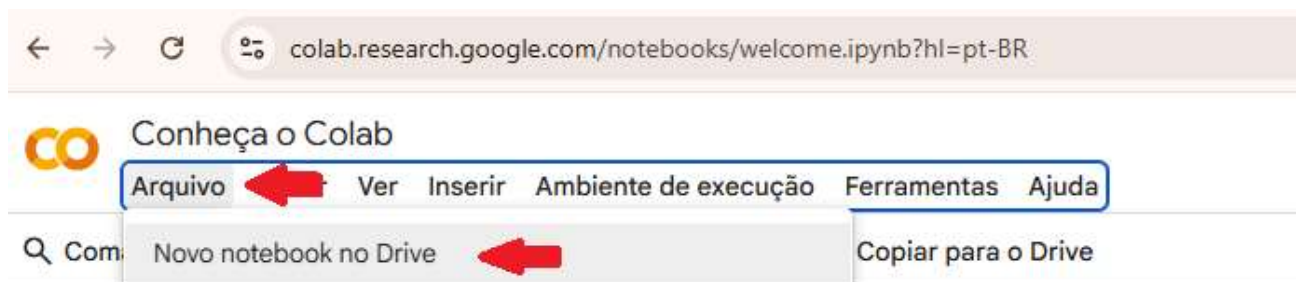
Figura 36: Tela inicial do Google Colab



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 37).

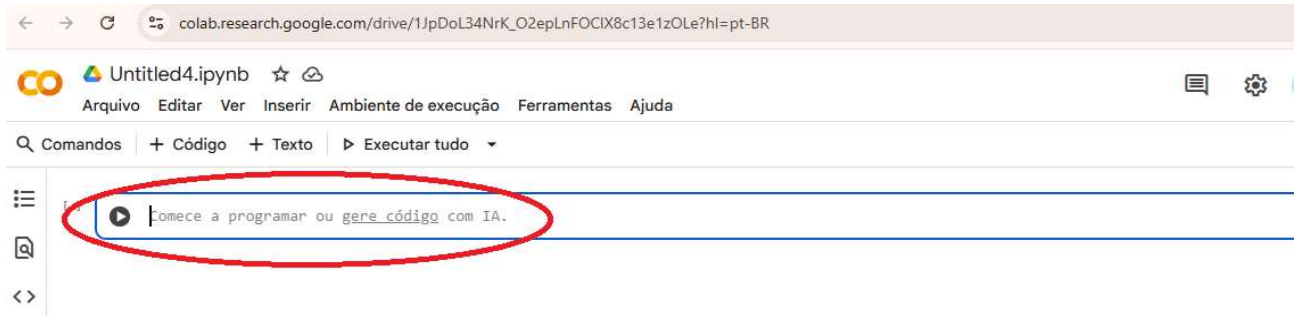
Figura 37: Abrindo a janela de inserção de códigos de programação



Fonte: O autor

Com a janela de programação aberta, o estudante pode começar a digitar os códigos e iniciar a programação, como mostra a figura 38.

Figura 38: 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 Piaç (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. A figura 39 mostra as bibliotecas utilizadas.

Figura 39: Importação das bibliotecas do Python



The screenshot shows a Jupyter Notebook window titled 'Untitled4.ipynb'. The interface includes a top menu bar with options like 'Arquivo', 'Editar', 'Ver', 'Inserir', 'Ambiente de execução', 'Ferramentas', and 'Ajuda'. Below the menu is a toolbar with icons for search, adding code cells, text cells, and running all cells. The main area displays a code cell with the following Python code:

```

#importação das bibliotecas
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score
import matplotlib.pyplot as plt
import statsmodels.api as sm
from scipy.stats import pearsonr

```

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 prefixo `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 resíduos.
- `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 MATLAB¹⁶, essencial

¹⁶ Ambiente de programação e computação numérica de alto desempenho usado principalmente em engenharia e ciência para análise de dados, desenvolvimento de algoritmos e modelagem.

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 40).

Figura 40: Inserção dos dados na janela de programação



```

#importação das bibliotecas
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score
import matplotlib.pyplot as plt
import statsmodels.api as sm
from scipy.stats import pearsonr

# inserção dos dados
x = np.array([1275.60,1312.00,1282.90,1295.40,1320.80,1356.10,1325.00,1337.90,1364.00,1401.50,1370.10,1383.20,1409.80,1448.50,1417.10]).reshape((-1,1))
y = np.array([1605.50,1655.00,1625.90,1648.10,1681.50,1728.90,1698.10,1721.50,1756.00,1805.90,1777.50,1801.10,1837.00,1887.50,1859.00])

```

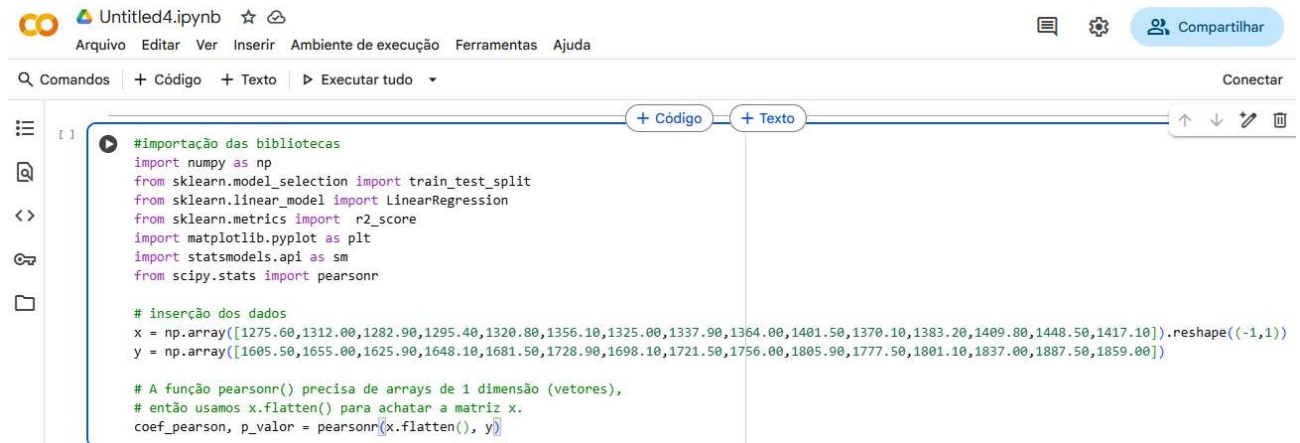
Fonte: O autor

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 Numpy Array, que é a estrutura de dados mais importante para computação numérica em Python. 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, como mostra a figura 41.

Figura 41: Comando para o cálculo do coeficiente de correlação



```

#importação das bibliotecas
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score
import matplotlib.pyplot as plt
import statsmodels.api as sm
from scipy.stats import pearsonr

# inserção dos dados
x = np.array([1275.00,1312.00,1282.90,1295.40,1320.80,1356.10,1325.00,1337.90,1364.00,1401.50,1370.10,1383.20,1409.80,1448.50,1417.10]).reshape((-1,1))
y = np.array([1605.50,1655.00,1625.90,1648.10,1681.50,1728.90,1698.10,1721.50,1756.00,1805.90,1777.50,1801.10,1837.00,1887.50,1859.00])

# A função pearsonr() precisa de arrays de 1 dimensão (vetores),
# então usamos x.flatten() para achatar a matriz x.
coef_pearson, p_valor = pearsonr(x.flatten(), y)

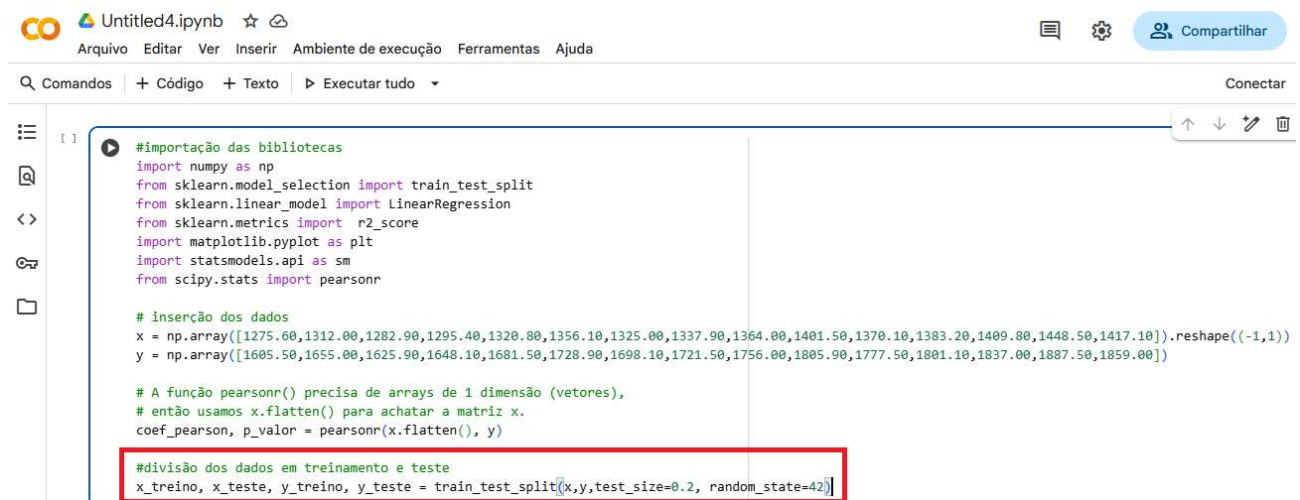
```

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 42.

Figura 42: Separando os dados de treino e os dados de teste do modelo



```

#importação das bibliotecas
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score
import matplotlib.pyplot as plt
import statsmodels.api as sm
from scipy.stats import pearsonr

# inserção dos dados
x = np.array([1275.00,1312.00,1282.90,1295.40,1320.80,1356.10,1325.00,1337.90,1364.00,1401.50,1370.10,1383.20,1409.80,1448.50,1417.10]).reshape((-1,1))
y = np.array([1605.50,1655.00,1625.90,1648.10,1681.50,1728.90,1698.10,1721.50,1756.00,1805.90,1777.50,1801.10,1837.00,1887.50,1859.00])

# A função pearsonr() precisa de arrays de 1 dimensão (vetores),
# então usamos x.flatten() para achatar a matriz x.
coef_pearson, p_valor = pearsonr(x.flatten(), y)

#divisão dos dados em treinamento e teste
x_treino, x_teste, y_treino, y_teste = train_test_split(x,y,test_size=0.2, random_state=42)

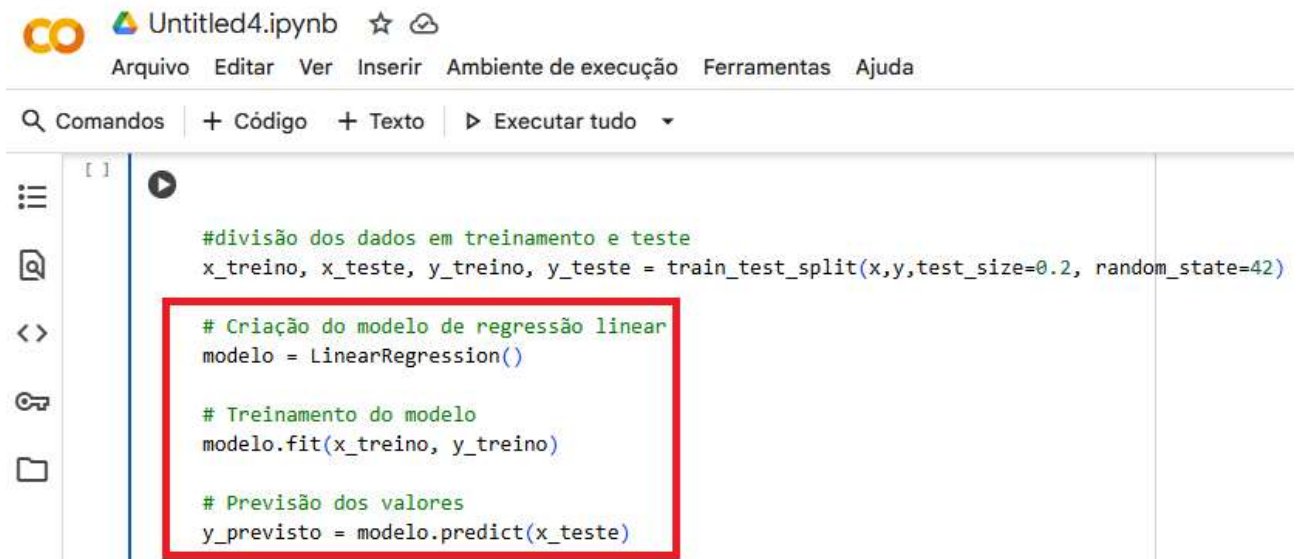
```

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, como mostra a figura 43.

Figura 43: Criação e treinamento do modelo e previsão de valores



```

#divisão dos dados em treinamento e teste
x_treino, x_teste, y_treino, y_teste = train_test_split(x,y,test_size=0.2, random_state=42)

# Criação do modelo de regressão linear
modelo = LinearRegression()

# Treinamento do modelo
modelo.fit(x_treino, y_treino)

# Previsão dos valores
y_previsto = modelo.predict(x_teste)

```

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 44).

Figura 44: 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_)

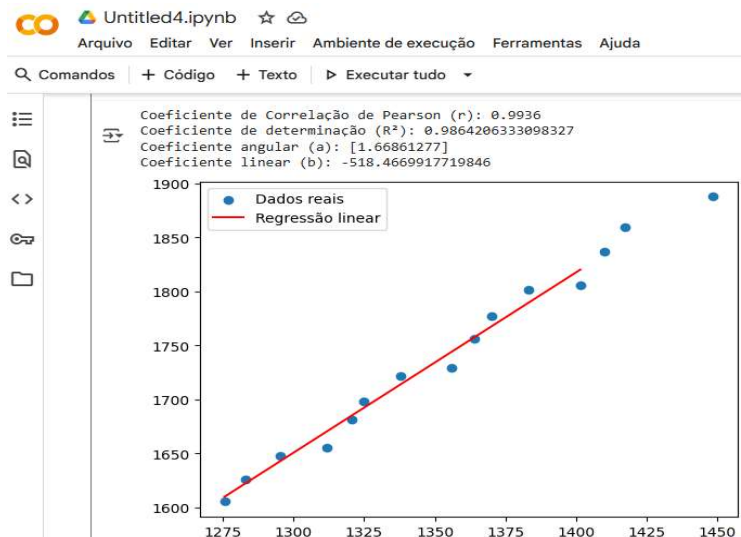
# Visualização dos dados
plt.scatter(x, y, label="Dados reais")
plt.plot(x_teste, y_previsto, color="red", label="Regressão linear")
plt.legend()
plt.show()

```

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, como mostra a figura 45.

Figura 45: 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 do 5° Encontro

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:

- Em sua opinião, qual a vantagem da utilização do Google Colab na programação em Python?
- Por que a linguagem de programação Python é considerada uma linguagem de alto nível?
- O uso das bibliotecas do Python facilita o trabalho do programador? Justifique a resposta.
- Com os dados da tabela acima, construa um modelo de Regressão Linear Simples em Python.

6 ANÁLISE DE RESULTADOS

A ênfase desta análise, sob a ótica do MCS, está no processo e não no resultado. Sendo assim, realizou-se a análise das legitimidades produzidas pelos estudantes durante os cinco encontros, levando-se em consideração as folhas de tarefas e os discursos que surgiram durante as atividades, através de apontamentos que auxiliaram nas reflexões sobre a produção de significados dos estudantes em relação aos temas abordados. Sobre conhecimento produzido, Lins afirma que

Como consequência de ser enunciado na direção de um interlocutor, e de ter mesmo sido produzido, todo conhecimento é verdadeiro. Isso não quer dizer que aquilo que é afirmado seja “verdade”. (LINS, 2012, pág. 21)

O desafio para o professor que se propõe a pesquisar e tentar compreender a “verdade” como atributo do conhecimento produzido pelo estudante quando se depara com um resíduo de enunciação, é saber colocar-se no lugar do aluno, despindo-se do papel de detentor do único conhecimento legítimo, pois somente com um olhar de empatia, sem juízo de valores como “isto está certo” e “aquilo está errado”, é que se pode fazer uma leitura positiva do que está sendo produzido.

Lins também sinaliza que “o que internalizamos, nos processos de humanização e do que se costuma chamar de desenvolvimento intelectual, são interlocutores, são legitimidades.” (LINS, 2012, pág. 20)

O MCS como base teórica para a análise das tarefas envolvendo conceitos como inteligência artificial, aprendizado de máquina e o modelo estatístico de regressão linear simples, possibilita uma movimentação entre os vários campos semânticos envolvidos, tornando ampla a produção de significados sobre esses conceitos. De acordo com Silva (2003), existem três grandes categorias no processo de produção de significados: o novo, a justificação e o dado.

O dado é aquilo que não precisa ser dito, pois está pressuposto no campo de significados do sujeito (estudante), e representa o conhecimento que o estudante já possui ou as crenças que assume como verdadeiras e que servem como referência para uma enunciação. Em nossas atividades, dado são os conceitos internalizados que cada estudante possui sobre, por exemplo, marcação de pontos no plano cartesiano, função polinomial do 1º grau e seus gráficos, média aritmética simples, desvio-padrão e inteligência artificial.

O novo é tudo o que é dito, expresso ou afirmado pelo estudante durante a atividade. É a produção de significado que acontece no momento da enunciação e representa o desenvolvimento cognitivo na compreensão do objeto, no interior das atividades. A produção de significados sobre os conceitos de correlação linear, modelo

preditivo de regressão linear simples, aprendizado supervisionado e a interpretação do modelo para a tomada de decisão, são exemplos do novo, no contexto de nossas atividades.

A justificação é o que autoriza o sujeito (estudante) a dizer o que diz, atuando como elemento de ligação entre o dado e o novo e através da justificação, o novo se estabelece, passando à condição de dado em situações posteriores. Sobre a justificação, Silva fala que

[...] a justificação tem o importante papel de ser o elo de ligação entre o novo e o dado. É a partir dela que ocorre o processo aonde o novo vai se transformando em dado frente a novas situações. (SILVA, 2003, pág. 69)

Segundo Lins e Giménez (1997), as justificações formam um vínculo entre as crenças-afirmações e os núcleos, e garantem, para o sujeito do conhecimento, que ele pode enunciar aquelas crenças-afirmações.

Esta pesquisa de caráter qualitativo, buscou realizar uma leitura positiva sobre a produção de significados de vinte e sete estudantes que participaram de todos os encontros. Sobre o perfil desses estudantes, verificou-se que são moradores de Belford Roxo, possuem faixa etária entre 15 e 17 anos, são do 3º Ano do Ensino Médio, possuem acesso à internet, seja por plano de dados móveis ou residencial, através do uso de celulares ou notebooks e desktops e costumam passar, em média, entre 5 a 6 horas por dia utilizando a internet, e na maioria das vezes, assistindo séries, jogando ou interagindo pelas redes sociais.

Durante os cinco encontros realizados, muitos conceitos foram trabalhados e discutidos, gerando uma quantidade significativa de material a ser analisado, e como exemplo, podemos citar a coerência interna entre os campos semânticos mobilizados e seus núcleos, as estipulações locais e a aderência aos objetos de estudo, os tipos de justificativas quanto ao uso da tecnologia, os interlocutores envolvidos na compreensão da correlação entre duas variáveis e na compreensão da Regressão Linear, os resíduos de enunciação dos estudantes nas concepções sobre IA e AM, na interpretação do modelo e na leitura de outliers, entre outros.

Devido às inúmeras combinações de análises possíveis e a limitação de tempo na confecção deste trabalho, separamos alguns resíduos de enunciação dos estudantes para serem analisados. Visando garantir a preservação da identidade dos estudantes, os mesmos serão retratados por números, como estudante 1, estudante 2 e assim por diante.

No 1º Encontro, a primeira pergunta da folha de tarefas foi: “Inteligência Artificial e Aprendizado de Máquina são termos que representam o mesmo conceito? Justifique sua

resposta.”

Vejamos alguns resíduos:

- **Estudante 23** – *Não, IA é um conceito amplo, e inclui qualquer sistema que imita a inteligência humana. E o aprendizado de Máquina é apenas uma parte da IA, ou seja, ela permite que a máquina aprenda com os dados.*
- **Estudante 14** – *Sim, pois uma precisa da outra para que tenha um bom funcionamento.*
- **Estudante 09** – *Não. IA é um conceito maior (máquinas simulando inteligência). Machine Learning é uma parte da IA onde as máquinas aprendem com dados.*
- **Estudante 19** – *Sim, pois o Aprendizado de Máquina é um subgrupo englobado na Inteligência Artificial.*

Observando a produção de significados desses estudantes sobre o conceito de Inteligência Artificial e Aprendizado de Máquina, verificamos que os estudantes 23 e 09 operaram em um campo semântico muito próximo do considerado correto, no sentido acadêmico, pois perceberam o Aprendizado de Máquina como um subconjunto de um conjunto maior, que é a Inteligência Artificial e isso fica exposto quando afirmam que “não” (não são o mesmo conceito) e nas palavras “o Aprendizado de Máquina é uma parte da IA ...” e “IA é um conceito maior ... Machine Learning é uma parte da IA”.

O resíduo do estudante 14 mostra um deslocamento de sentido, uma vez que está inserido em um campo semântico de operacionalidade e funcionamento da IA e não no campo semântico da classificação dos conceitos de IA e AM. Esse estudante faz uma observação prática de que sistemas de IA dependem de algoritmos de AM para serem eficientes. O “sim” (sim, são o mesmo conceito), traz como justificção a interdependência entre IA e AM, e como uma depende da outra para funcionar, então representam o mesmo conceito. Pelo olhar do MCS, o estudante não errou dentro do seu campo semântico, uma vez que a lógica de pensamento possui certa coerência interna. Esse estudante deve transitar do campo semântico funcional para o campo semântico conceitual.

A resposta do estudante 19 é muito interessante, uma vez que parte de uma premissa (dado) correta para chegar em uma conclusão (novo) oposta. O estudante percebe que o AM é um subgrupo da IA, mas a justificção do “sim” (sim, são o mesmo conceito) acontece por generalização. Para esse estudante, dizer que são o mesmo conceito significa dizer que se trata de um mesmo objeto. Nesse caso, o professor pode intervir, apresentando situações que possibilitem o “novo” a se tornar “dado”, com indagações do tipo “os números pares são naturais?”, “todos os números naturais são pares?”, “números naturais e números pares possuem o mesmo conceito?”.

Essas perguntas levam o estudante a incorporar justificações e criar novos interlocutores.

No 2º Encontro, o item a da questão única tem o seguinte texto: “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.”

- **Estudante 12** – *Sim, porque quem se dedica mais, pode conseguir uma nota melhor.*
- **Estudante 27** – *Depende, porque se o aluno já sabe a matéria, não precisa estudar muito.*
- **Estudante 01** – *Não, eu já estudei muito para uma prova e tirei nota baixa.*

Percebe-se que o estudante 12 já incorporou como “dado”, o conceito de correlação linear positiva, trabalhado no 1º Encontro, visto que considera o esforço nos estudos como uma possibilidade de obter uma nota melhor, ou seja, quanto mais estudo, maior a chance de ter boa nota. O fato do estudante ter utilizado a palavra “pode” em sua resposta, mostra que existe uma probabilidade maior de acontecer o fato, mas não garante que a boa nota será tirada. Essa resposta está englobada no campo semântico da regressão linear simples, onde existe uma aleatoriedade sustentada pelas variáveis que não foram consideradas no modelo.

O resíduo de enunciação do estudante 27 traz a informação sobre o conhecimento prévio através da palavra “depende”. O conhecimento prévio não é a variável independente do modelo. A variável independente do modelo em questão são as horas de estudo. Ele relativiza as horas de estudo, priorizando o conhecimento prévio. É como se o modelo de regressão linear $y=ax+b$, tivesse o conhecimento prévio como coeficiente linear (intercepto). Assim, mesmo com poucas horas de estudo (x), a nota ainda seria alta, pois o valor de b é alto. O campo semântico desse estudante não considera as horas de estudo como fator principal para um bom desempenho na prova. Ele traz o conhecimento prévio, que é um dos componentes do ruído (resíduo) do modelo, como fator principal para o sucesso na prova. Na visão do MCS, não há erro na análise do estudante, pois em sua produção de significados, quem possuir maior conhecimento prévio sobre o assunto irá garantir uma melhor nota. Cabe ao professor, gerar situações que facilitem a movimentação entre os campos semânticos do modelo de regressão linear simples, mostrando que vários modelos podem surgir, de acordo com a escolha da variável independente.

A resposta do estudante 01 traz uma experiência pessoal como justificção para a negação do modelo. Ele nega a correlação linear positiva existente entre as variáveis devido ao fato de focar no ponto que está muito distante da nuvem de pontos, o que

chamamos de outliers. A legitimidade desse estudante vem da própria vivência e no campo semântico dele, que considera que o modelo tem que funcionar sempre. Se o modelo falhou com ele, o modelo está invalidado. Ele está operando em um campo semântico determinístico e o professor tem que atuar para que o estudante se movimente para um campo semântico estocástico/probabilístico.

No 3º Encontro, o item f da questão única traz a pergunta: “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?”

- **Estudante 18** – *Gera um aumento de 5,65 unidades monetárias com alimentação.*
- **Estudante 10** – *O aumento com alimentação é de 5,40 unidades monetárias.*

De acordo com os dados do exercício, o modelo de Regressão Linear encontrado foi $y=0,25x+5,40$, e a resposta esperada era de um aumento de 0,25 unidades monetárias de gastos com alimentação.

O estudante 18, quando responde que o aumento é de 5,65, interpreta a pergunta como “qual o valor de y quando $x=1$?” e isso mostra que a reta de regressão encontrada por ele estava de acordo com o modelo, porém, o estudante operou em um campo semântico de valor da função no ponto. Ele usou como dado a função $y=0,25x+5,40$ e o valor da função quando $x=1$, como sendo o aumento de uma unidade. A justificativa foi a lógica da operação $0,25 \cdot 1 + 5,40 = 5,65$. O estudante entendeu o aumento de uma unidade como sendo a renda igual a um. Sendo o interlocutor uma direção na qual se fala, essas direções não convergiram devido ao fato do discente não estar operando no campo semântico da taxa de variação, no qual o olhar deve ser direcionado para o coeficiente angular, mas de qualquer forma, sua produção de significado é legítima dentro do campo semântico do valor da função no ponto $x=1$.

Uma forma de tentar movimentar esse estudante para o campo semântico da taxa de variação, é fazer uma comparação entre uma família A que possui renda de seis unidades monetárias e uma família B com renda de sete unidades monetárias, para ficarmos dentro do intervalo de dados da amostra. Utilizando a reta de regressão linear encontrada, verifica-se que o gasto com alimentação da família A é de 6,90 e o gasto com alimentação da família B é de 7,15. Assim, quando há o aumento de uma unidade monetária de renda ($7-6=1$), existe o aumento de $7,15-6,90=0,25$ no gasto com alimentação.

O estudante 10, em sua produção de significados, deslocou o foco para o coeficiente linear e usou como dado a função $y=0,25x+5,40$ e a palavra “aumento” (sinal

de adição). Ele encontrou justificação numa associação visual, uma vez que o número que está sendo somado é 5,40 e portanto, representa o aumento. Esse estudante está trabalhando num campo semântico que engloba o visual e a operação de adição, que vincula o intercepto ao ganho ou base para o aumento. A legitimidade está no fato do intercepto (+5,40) ser o acréscimo à parte variável, sendo identificado como o termo que representa a soma.

Esse estudante deve ser estimulado a ressignificar os papéis do coeficiente linear e do coeficiente angular, observando que o valor do intercepto é fixo e não se altera com as mudanças na variável x . A invariância do termo 5,40 é o novo dado para a transição entre os campos semânticos, pois não vale mais como justificção para o aumento, uma vez que o valor é constante.

No 4º Encontro, vale destacar o resíduo de enunciação do estudante 03, em resposta ao item e da questão única que tem o seguinte enunciado: “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.”

- **Estudante 03** – *Na tabela, quem tem 5 anos de experiência erra 18 vezes e fazendo as contas pela função, o erro foi de 18,45. O modelo errou.*

Fazendo uma leitura positiva, na tentativa de entender o que o estudante assumiu como legítimo para a produção de significado sobre os modelos de Regressão Linear Simples, vemos que a diferença entre o valor previsto pelo modelo e o valor observado ($18,45 - 18,00 = 0,45$) foi considerado um erro do modelo. Esse estudante está operando no campo semântico da exatidão matemática, ou seja, se o modelo é uma função, essa função deveria encontrar o valor exato. A diferença de 0,45 é considerada uma falha na validação do modelo e não como uma característica dos modelos de Regressão. A justificção baseia-se na imperfeição do modelo matemático quando confrontado com a realidade dos dados. Cabe ao professor, estimular esse estudante a trabalhar no campo semântico da modelagem estatística, onde a diferença entre o valor observado e o valor previsto é chamado de erro ou resíduo, que é minimizado pelo método dos mínimos quadrados. A função encontrada é a que melhor minimiza a soma dos quadrados das diferenças entre os valores observados e os valores previstos pelo modelo.

Esse estudante pode ser levado a pensar em outro digitador com 5 anos de experiência. Será que esse outro digitador também cometeria 18 erros, ou poderia cometer 19 ou 17 erros? Essa provocação tem o objetivo de criar um novo conhecimento, que é o fato do modelo ser usado para prever tendências.

Em relação ao 5º Encontro, como a linguagem Python não era de conhecimento de nenhum dos estudantes, um tempo maior do que o previsto foi utilizado na apresentação

da linguagem, e por isso realizamos uma leitura positiva global do item c da questão única, que pergunta se o uso das bibliotecas do Python facilita o trabalho do programador.

Pudemos observar que todos, em suas respostas consideraram a linguagem como sendo uma facilitadora no trabalho do desenvolvedor, pois possui bibliotecas que agilizam a programação, deixando mais tempo para a interpretação do modelo e as tomadas de decisão. A justificação para essas respostas foi a autoridade do professor, que “emprestou a legitimidade” para os estudantes, transformando o novo em dado.

A participação dos estudantes nas atividades evidenciou a estrutura das legitimidades construídas por eles ao longo da vida escolar. A dificuldade encontrada em interpretar o coeficiente angular como uma taxa de variação constante, demonstra que o conhecimento prévio sobre funções polinomiais do 1º grau ainda estava fortemente vinculado a exercícios de substituição numérica, e não à interpretação dos parâmetros. No MCS, isso significa que os interlocutores internos dos sujeitos operavam em campos semânticos onde a "verdade" de um enunciado depende da exatidão do resultado, dificultando a aceitação do "erro" como uma métrica de ajuste (resíduo) e não como uma falha do modelo.

Após as atividades, a comparação entre os modelos conceituais de IA e AM e as aplicações práticas de regressão linear simples mostraram uma evidente movimentação entre os campos semânticos da exatidão matemática e dos modelos probabilísticos, culminando em produção de significados legítimos por parte dos estudantes. Este professor/pesquisador observou que o uso da linguagem Python atuou como uma legitimidade emprestada. O desempenho nas tarefas de programação refletiu uma aceitação dessa ferramenta tecnológica sob a autoridade do professor, onde o "novo" se tornou "dado" sem que os estudantes produzissem uma justificação profunda que os levasse a legitimar, de forma própria, os resultados obtidos.

7 CONSIDERAÇÕES FINAIS

A ideia de realizar uma dissertação que pudesse contribuir no campo da Educação Matemática com uma sequência de tarefas extracurriculares na qual seriam apresentados conceitos sobre aprendizado de máquina, algoritmos supervisionados, regressão linear simples e ferramentas tecnológicas utilizadas para facilitar a análise de dados, surgiu desde o primeiro semestre letivo do mestrado. Na minha experiência como professor, ao longo de 21 anos de carreira, sentia falta de trabalhar os conceitos matemáticos de forma a aplicar a fundamentação teórica em situações de modelagem matemática. Essa dissertação proporcionou a este professor/pesquisador essa oportunidade, porém tive que vencer o desafio de conhecer a linguagem Python, estudar com mais profundidade a regressão linear simples e analisar a produção dos alunos com base no modelo dos

campos semânticos. A confecção das atividades propostas foi pensada e criada de forma a habilitar o estudante a construir e interpretar o modelo de regressão linear simples, utilizando desde o papel e a caneta, até uma linguagem de programação de alto nível, passando pela calculadora e planilhas eletrônicas.

A estratégia utilizada na proposição das atividades, possibilitou que os alunos construíssem suas próprias justificações antes de delegarem a tarefa à tecnologia. Essa gradatividade foi essencial para que o "novo" se tornasse "dado", mostrando que a inteligência artificial baseia-se em conceitos matemáticos e estatísticos.

Após a aplicação das atividades em uma escola estadual situada na periferia de uma cidade fluminense com um dos piores índices de desenvolvimento sócioeconômicos do estado do Rio de Janeiro, pude perceber que o objetivo do trabalho foi alcançado, pois houve uma produção de significados muito rica por parte dos estudantes, com uma troca de conhecimentos mútuo.

A visão do erro no modelo dos campos semânticos, constitui uma oportunidade para o professor interpretar e tentar entender como o aluno está pensando, melhorando significativamente o processo de ensino-aprendizagem.

Como a ênfase, durante as atividades propostas, foi dada ao processo e não ao resultado final, tivemos a oportunidade de observar a evolução de muitos estudantes. O fato de utilizar a tecnologia para algo diferente de jogos eletrônicos e redes sociais, possibilitou que esses estudantes ampliassem a visão sobre a aplicabilidade dos conceitos matemáticos e o uso das tecnologias.

Alguns estudantes, inclusive, disseram que estavam pesquisando mais sobre inteligência artificial e aprendizado de máquina, além de perceberem de forma mais profunda, como a inteligência artificial atua para influenciar suas tomadas de decisão, seja na escolha de filmes para assistir ou roupas para comprar, demonstrando que o conhecimento produzido extrapolou os limites da sala de aula e atingiu a esfera da cidadania crítica.

Quanto ao modelo de regressão linear simples, os estudantes gostaram da ideia de prever um resultado utilizando a reta de regressão ajustada. Os grupos de estudantes formados para a realização das atividades utilizando a planilha eletrônica e o software matemático, chegaram nos resultados esperados, alguns com a ajuda do professor, mas todos foram muito participativos no processo.

No quinto encontro, no qual foi apresentada a linguagem Python, os estudantes apresentaram mais dúvidas, e esse pesquisador observou que dois tempos de cinquenta minutos não foram suficientes para o melhor desenvolvimento da atividade. Seriam necessários programar mais dois tempos para reforçar o que foi trabalhado.

As atividades propostas no quinto capítulo desta dissertação resultaram na elaboração de um produto educacional, que pode ser adaptado de acordo com a realidade de cada docente, e está a disposição para servir como material de apoio aos colegas professores de Matemática, de forma a contribuir com as suas práticas em sala de aula, auxiliando os estudantes a terem uma melhor compreensão e entendimento sobre os conceitos abordados.

Como perspectivas futuras, esta dissertação converge para a necessidade de expandir o tempo de contato dos estudantes com linguagens de programação de alto nível, sugerindo que este produto educacional atue de forma a contribuir com currículos interdisciplinares mais robustos. Existe a possibilidade de investigar como a continuidade dessas práticas impacta na produção de significados dos estudantes em outras áreas da matemática, como o estudo de matrizes, pois o modelo de regressão linear simples pode ser escrito na forma matricial. Este produto educacional produzido, como dito anteriormente, não deve ser visto como um material pronto e acabado, mas como um convite à adaptação e à pesquisa contínua, consolidando a ideia de que a educação matemática deve ser dinâmica, inclusiva e profundamente conectada com as inovações tecnológicas do século XXI.

REFERÊNCIAS

ALMEIDA, A.; CARVALHO, F.; MENINO, F. Introdução ao machine learning. [S.l.]: Grupo DataAt, 2017. Disponível em <https://dataat.github.io/introducao-ao-machine-learning/>. Acesso em: 28 jun 2025.

ALPAYDIN, Ethem. **Introduction to machine learning**. 4. ed. Cambridge, MA: MIT Press, 2020.

BHAVSAR, P. et al. **Machine learning in transportation data analytics**. In: CHOWDHURY, M.; APON, A.; DEY, K. Data Analytics for Intelligent Transportation Systems. [S.l.]: Elsevier, 2017. p. 283-307.

BRASIL. Ministério da Educação. **Base Nacional Comum Curricular**. Brasília, 2018. Disponível em: <http://basenacionalcomum.mec.gov.br/>. Acesso em: 25 set. 2025.

BRASIL. Ministério da Educação. **Computação – Complemento à BNCC**. Brasília, 2022. Disponível em: <https://www.gov.br/mec/pt-br/escolas-conectadas/BNCCComputaoCompletoDiagramado.pdf>. Acesso em: 25 set. 2025.

BRASIL. Ministério da Educação. **Referencial para o Desenvolvimento e Uso Responsáveis de Inteligência Artificial na Educação**. Brasília, 2024. Disponível em: <https://www.gov.br/mec/pt-br/assuntos/noticias/2025/outubro/mec-abrira-consulta-publica-sobre-ia-na-educacao>. Acesso em: 27 nov. 2025.

BRASSCOM. **Relatório Setorial de TIC 2023**. São Paulo: Brasscom, 2023. Disponível em: <https://brasscom.org.br/relatorios/>. Acesso em: 31 jan. 2026.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, [2018]. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 10 out. 2025.

BRASIL. Senado Federal. **Projeto de Lei nº 2338, de 2023**. Dispõe sobre o uso da Inteligência Artificial no Brasil, e dá outras providências. Apresentado pelo Senador Rodrigo Pacheco. Brasília: Senado Federal, 2023. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>. Acesso em: 05 dez. 2025.

BUSSAB, Wilton de Oliveira; MORETTIN, Pedro Alberto. **Estatística Básica**. 9. ed. São Paulo: Saraiva, 2017.

CRESPO, Antônio Arnot. **Estatística fácil**. 18. ed. São Paulo: Saraiva, 2004.

DALLAGASSA, M. R., & de OLIVEIRA, A. J. C. (2024). **Uso de machine learning no diagnóstico de câncer de mama através de ultrassonografia**. *Journal of Health Informatics*, 16(Especial). Disponível em: <https://doi.org/10.59681/2175-411.v16.iEspecial.2024.1289>. Acesso em 03 out. 2025.

DANTE, Luiz Roberto. **Matemática**: contexto e aplicações: volume único. São Paulo: Ática, 2002.

DIAS, Jonathan Marques. **Aprendizado de máquina: previsão de evasão com python e recurso didático e tecnológico com Geogebra**. Dissertação (Mestrado Profissional em Matemática em Rede Nacional) - Universidade Federal do Triângulo Mineiro, Uberaba, MG, 2024

GARDNER, Howard. **Inteligências múltiplas**: a teoria na prática. Porto Alegre: Artmed, 1995.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep learning**. Cambridge, MA: MIT Press, 2016.

GUJARATI, Damodar N.; PORTER, Dawn C. **Econometria Básica**. 5. ed. Porto Alegre: AMGH, 2011.

LINS, Andréa; SOUZA, Lins. **Python para Matemáticos** – V colóquio de matemática da região nordeste. João Pessoa – PB: Disponível em https://sbm.org.br/wp-content/uploads/2023/09/Minicurso_Python_final2.pdf. Acesso em: 01 ago 2025.

LINS, Rômulo C; GIMENEZ, Joaquim. Perspectivas em aritmética e álgebra para o século XXI. 3. ed. Campinas: Editora Papyrus, 1997. v. 1. 250p. Disponível em: <http://sigma-t.org/permanente/1997a.pdf>. Acesso em 18 set. 2022.

LINS, Rômulo C. A diferença como oportunidade para aprender. *In*: XIV ENDIPE, 2008, Porto Alegre. **Trajetórias e processos de ensinar e aprender**: sujeitos, currículos e culturas. Porto Alegre: PUCRS, v.3. p. 530- 550, 2008.

LINS, Rômulo C. O modelo dos campos semânticos: estabelecimentos e notas de teorizações. *In*: ANGELO, Claudia L. et al. (Orgs). **Modelo dos Campos Semânticos e Educação Matemática - 20 anos de história**. São Paulo: Midiograf, 2012, p. 11- 30.

LOTH, Maria H. M; SILVA, Amarildo M. Tarefas aritméticas para o 6º Ano do Ensino Fundamental. *In*: **BOLEMA**: Boletim de Educação Matemática, v. 27, n. 46, p. 451- 465, Rio Claro, 2013. Disponível em: <https://www.scielo.br/j/bolema/a/XvvhMgXrswYDq9LX8rQ468j/?lang=pt&format=pdf>. Acesso em 19 out. 2025.

LUGER, George F. **Inteligência Artificial**. 6. ed. São Paulo: Pearson Education, 2013.

Pazim, R., & Fonseca, O. J. T. (2022). **Modelos matemáticos: análise de regressão com o GeoGebra**. (1a ed.). Fundação Uniselva. (MT Ciência – Série livros). Disponível em: <https://www.mtciencia.com.br/editora/livros>. Acesso em 04/08/2025

PIAZ, Marcelo Bragatto Dal. **Uso da linguagem python no auxílio à aprendizagem de estatística descritiva na educação básica**. Dissertação (Mestrado Profissional em Matemática em Rede Nacional) - Universidade Federal do Espírito Santo, Vitória, ES, 2023.

REZENDE, Solange Oliveira. **Sistemas Inteligentes**: fundamentos e aplicações. Barueri: Manole, 2003.

ROBERT HALF. **Guia Salarial 2024**. São Paulo: Robert Half, 2024. Disponível em: <https://www.roberthalf.com.br/guia-salarial>. Acesso em: 31 jan. 2026.

RUSSELL, Stuart; NORVIG, Peter. **Inteligência Artificial**. 3. ed. Rio de Janeiro: Elsevier, 2013.

SAMUEL, Arthur L. Some Studies in Machine Learning Using the Game of Checkers. **IBM Journal of Research and Development**, [S. l.], v. 3, n. 3, p. 210-229, 1959.

SANTAELLA, Lucia. **A inteligência artificial é inteligente?**. São Paulo: Paulus, 2023.

SILVA, Amarildo Melchiades da. **Sobre a dinâmica da produção de significados para a matemática**. 2003. Tese (Doutorado em Educação Matemática) – Instituto de Geociências e Ciências Exatas, Universidade Estadual Paulista, Rio Claro, 2003.

SILVA, Amarildo M; LINS, Rômulo C. **Sobre a dinâmica da produção de significados para a matemática**. *Jornal Internacional de Estudos em Educação Matemática*. v.6(2), 2013. Disponível em: <https://revista.pgsskroton.com/index.php/jieem/article/view/91>. Acesso em 20 ago. 2025.

TEIXEIRA, João de Fernandes. **O que é Inteligência Artificial**. São Paulo: Brasiliense, 1990.

TRIOLA, Mario F. **Introdução à Estatística**. 10. ed. Rio de Janeiro: LTC, 2010.

WORLD ECONOMIC FORUM. **The Future of Jobs Report 2023**. Geneva: WEF, 2023. Disponível em: <https://www.weforum.org/publications/the-future-of-jobs-report-2023/>. Acesso em: 2 fev. 2026.

APÊNDICE A – Folha de Tarefas do 1º Encontro**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.

APÊNDICE B– Folha de Tarefas do 2º Encontro**Folha de tarefas 2**

- 2) 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

- g) 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.
- h) Represente os dados em um diagrama de dispersão.
- i) Apenas observando o gráfico de dispersão, podemos dizer que existe correlação entre os dados? Justifique sua resposta.
- j) Determine a reta de regressão linear e represente a reta no diagrama de dispersão.
- k) Calcule o Coeficiente de Correlação Linear de Pearson e o Coeficiente de Determinação.
- l) Use a reta de regressão encontrada para prever a nota final de um estudante que estuda 7 horas por dia.

APÊNDICE C– Folha de Tarefas do 3º Encontro

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:

- Organize os dados na planilha.
- Represente os dados em um gráfico de dispersão.
- Calcule o coeficiente de correlação.
- Determine a reta de regressão linear e represente a reta no gráfico de dispersão.
- Calcule o coeficiente de determinação e interprete o valor encontrado.
- Use a reta de regressão encontrada para prever o gasto com alimentação de uma família com renda de 110 unidades monetárias.
- 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?

APÊNDICE D– Folha de Tarefas do 4º 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:

- Escolha uma forma de representar graficamente esse conjunto de dados.
- Você considera que o coeficiente de correlação encontrado demonstra uma boa correlação entre as variáveis? Justifique sua resposta.
- Determine a reta de regressão linear que melhor se ajusta aos dados da amostra.
- Calcule o coeficiente de determinação e interprete o valor encontrado.
- 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.

APÊNDICE E– Folha de Tarefas do 5º Encontro

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 5º Encontro, desenvolva os itens abaixo:

- Em sua opinião, qual a vantagem da utilização do Google Colab na programação em Python?
- Por que a linguagem de programação Python é considerada uma linguagem de alto nível?
- O uso das bibliotecas do Python facilita o trabalho do programador? Justifique a resposta.
- Com os dados da tabela acima, construa um modelo de Regressão Linear Simples em Python.

ANEXO A – Termo de Assentimento Livre e Esclarecido

TERMO DE ASSENTIMENTO LIVRE E ESCLARECIDO

Você está sendo convidado para participar da pesquisa: Introdução ao Aprendizado de Máquina utilizando o Modelo de Regressão Linear Simples: uma proposta para estudantes do 3º ano do Ensino Médio. Queremos analisar através de cinco folhas de tarefas, como se dará a produção de significados sobre aprendizado de máquina e o modelo de Regressão Linear Simples.

As pessoas que irão participar desta pesquisa possuem de 15 a 19 anos de idade. A pesquisa será feita no Colégio Estadual Bom Pastor – Belford Roxo - RJ. Durante as atividades, você irá responder a cinco folhas de tarefas. Depois das aulas expositivas, serão usadas somente as folhas de tarefas, a caneta e o recurso tecnológico disponível (notebook ou tablet). O uso de folha de tarefas, caneta e planilhas eletrônicas é considerado seguro, mas caso você se sinta desconfortável durante as atividades, pode falar diretamente com o pesquisador Ronaldo Cesar da Silva ou fazer contato pelo telefone (21) 98721-0223. Porém, muitas coisas boas que podem acontecer, como um maior conhecimento sobre aprendizado de máquina, inteligência artificial (IA), contato com planilhas eletrônicas, softwares matemáticos e linguagens de programação, além do fato desta pesquisa poder contribuir formalmente para o progresso da Educação Matemática nas escolas públicas.

Você não precisa participar desta pesquisa se não quiser. Ninguém ficará irritado ou chateado com você se disser “não”: a escolha é sua. Você pode pensar nisto e falar depois se você quiser. Você pode dizer “sim” agora e mudar de ideia depois e tudo continuará bem. É importante que você converse com seus responsáveis sobre a sua decisão. Saiba o que eles acham, fale a eles o que pretende fazer, se quer ou não participar. Você tem o tempo que precisar para isso. Também pode discutir com o pesquisador, quando quiser. Ele responderá todas as suas dúvidas, em qualquer momento.

Você não receberá nenhum dinheiro nem terá que pagar nada para participar da pesquisa. Ninguém saberá que você está participando da pesquisa, não falaremos a outras pessoas, nem daremos a estranhos as informações que você nos der. Os resultados da pesquisa poderão ser publicados, mas sem identificar as pessoas que participaram da pesquisa.

ASSENTIMENTO

Eu _____ li este termo e aceito participar da pesquisa.

Assinatura participante	do(a)	Data: ___/___/___
----------------------------	-------	----------------------

Eu, Ronaldo Cesar da Silva, obtive de forma apropriada e voluntária o Assentimento Livre e Esclarecido do participante da pesquisa.

Assinatura pesquisador(a).	do(a)	Data: ___/___/___
-------------------------------	-------	----------------------

ANEXO B – Termo de Consentimento Livre e Esclarecido – Responsável Legal

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO – RESPONSÁVEL LEGAL

Prezado(a) responsável/representante legal:

Gostaríamos de solicitar o seu consentimento para o(a) menor _____ participar como voluntário(a) da pesquisa denominada “Introdução ao Aprendizado de Máquina utilizando o Modelo de Regressão Linear Simples: uma proposta para estudantes do 3º ano do Ensino Médio”, realizada no âmbito da Pró-Reitoria de Pós-Graduação, Pesquisa, Extensão e Cultura – PROFMAT - CPIL e que diz respeito a uma pesquisa que subsidiará a Dissertação de conclusão do curso de Mestrado Profissional – PROFMAT. A pesquisa será realizada no Colégio Estadual Bom Pastor – Belford Roxo – RJ.

1. OBJETIVO: O objetivo do estudo é analisar a produção de significados dos estudantes no âmbito do aprendizado de máquina, com foco restrito no modelo de regressão linear simples, com o intuito de analisar e identificar situações do mundo atual que podem ser preditas utilizando o modelo de regressão estudado.

2. PROCEDIMENTOS: a forma de participação do (a) menor consistirá em assistir aos cinco encontros do minicurso e realizar as atividades constantes nas folhas de tarefas que serão entregues ao final de cada encontro. Essas folhas de tarefas servirão de base para que o pesquisador possa analisar a produção de significados realizada pelos estudantes sobre o “tema aprendizado de máquina e modelo de regressão linear simples”. Nos dois primeiros encontros, os participantes usarão apenas a folha de tarefas e caneta e nos três últimos, utilizaremos notebooks ou tablets para as atividades que requeiram o uso de planilhas eletrônicas, softwares matemáticos e linguagem de programação. Os registros serão feitos através das folhas de tarefas e cada aluno será identificado apenas por um número, ou seja, não haverá identificação nominal. Vale salientar também que fotografias podem ser realizadas durante a aplicação das tarefas, mas sem focar nos rostos ou partes do corpo que possam identificar algum estudante.

3. POTENCIAIS RISCOS E BENEFÍCIOS: Toda pesquisa oferece algum tipo de risco. Nesta pesquisa, o risco pode ser avaliado como mínimo, isto é, o participante pode apresentar timidez e se sentir envergonhado ou não se sentir confortável em participar de alguma atividade, ou pode apresentar algum desconforto em exteriorizar as suas respostas nas tarefas propostas. Objetivando minimizar esses riscos, o participante tem a possibilidade de decidir pela participação ou não em quaisquer atividades e a qualquer momento, sem problema algum. Por outro lado, são esperados os seguintes benefícios da participação na pesquisa: contribuição para o conhecimento sobre aprendizado de máquina, inteligência artificial e modelo de regressão linear simples, introdução à programação em Python e consequente aumento de sua autoestima, além de contribuir para a melhoria da Educação Matemática nas Escolas Públicas.

4. GARANTIA DE SIGILO: Os dados da pesquisa poderão ser publicados/divulgados em livros e revistas científicas. Asseguramos que a privacidade do(a) menor participante será respeitada e o seu nome ou qualquer informação que possa, de alguma forma, o (a) identificar, será mantida em sigilo. O pesquisador responsável se compromete a manter os dados da pesquisa em arquivo, sob sua guarda e responsabilidade, por um período mínimo de 5 (cinco) anos após o término da pesquisa.

5. LIBERDADE DE RECUSA: a participação do (a) menor neste estudo é voluntária e não é obrigatória. Você poderá se recusar a permitir que ele (a) participe do estudo, ou retirar seu consentimento a qualquer momento, sem precisar justificar. Se desejar que o (a) menor saia da pesquisa ele (a) não sofrerá qualquer prejuízo.

6. CUSTOS, REMUNERAÇÃO E INDENIZAÇÃO: a participação neste estudo não terá custos adicionais para você. Também não haverá qualquer tipo de pagamento devido à participação do (a) menor no estudo. Fica garantida indenização em casos de danos, comprovadamente decorrentes da participação na pesquisa, nos termos da Lei.

7. ESCLARECIMENTOS ADICIONAIS, CRÍTICAS, SUGESTÕES E RECLAMAÇÕES: você receberá uma via deste Termo de Consentimento Livre e Esclarecido (TCLE) e a outra ficará com o(a) pesquisador(a). Caso você concorde em participar, as páginas serão rubricadas e a última página será assinada por você e pelo pesquisador. O pesquisador garante a você livre acesso a todas as informações e esclarecimentos adicionais sobre o estudo e suas consequências. Você poderá ter acesso ao pesquisador Ronaldo Cesar da Silva pelo telefone (21)98721-0223 ou pelo e-mail: ronaldocesar73@gmail.com. Se você tiver alguma consideração ou dúvida sobre a ética da pesquisa, poderá entrar em contato com o Comitê de Ética em Pesquisa do Colégio Pedro II (CEP/CP II), situado no Endereço: Campo de São Cristóvão nº 177, prédio da Pró-Reitoria de Pós-Graduação, Pesquisa, Extensão e Cultura (PROPGPEC), sala 202-B – São Cristóvão – Rio de Janeiro, CEP 29921-903, pelo telefone: 21 3891-0020 ou pelo e-mail: cep@cp2.g12.br

CONSENTIMENTO

Eu, _____ li e concordo com a participação do menor _____ na pesquisa.

Assinatura do(a) responsável /representante legal	Data: ___/___/___
---	-------------------

Eu, Ronaldo Cesar da Silva, obtive de forma apropriada e voluntária o Consentimento Livre e Esclarecido do (a) responsável /representante legal pelo (a) menor participante da pesquisa.

Assinatura do pesquisador	Data: ___/___/___
---------------------------	-------------------

ANEXO C – Termo de Consentimento Livre e Esclarecido – Maiores de Idade

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO – MAIORES DE IDADE

Você está sendo convidado (a) a participar como voluntário (a) da pesquisa denominada “Introdução ao Aprendizado de Máquina utilizando o Modelo de Regressão Linear Simples: uma proposta para estudantes do 3º ano do Ensino Médio”, realizada no âmbito da Pró-Reitoria de Pós-Graduação, Pesquisa, Extensão e Cultura – PROFMAT - CPIL e que diz respeito a uma pesquisa que subsidiará a Dissertação de conclusão do curso de Mestrado Profissional – PROFMAT.

1. OBJETIVO: O objetivo do estudo é analisar a produção de significados dos estudantes no âmbito do aprendizado de máquina, com foco restrito no modelo de regressão linear simples, com o intuito de analisar e identificar situações do mundo atual que podem ser preditas utilizando o modelo de regressão estudado.

2. PROCEDIMENTOS: a sua participação consistirá em assistir aos cinco encontros do minicurso e realizar as atividades constantes nas folhas de tarefas que serão entregues ao final de cada encontro. Essas folhas de tarefas servirão de base para que o pesquisador possa analisar a produção de significados realizada pelos estudantes sobre o “tema aprendizado de máquina e modelo de regressão linear simples”. Nos dois primeiros encontros, os participantes usarão apenas a folha de tarefas e caneta e nos três últimos, utilizaremos notebooks ou tablets para as atividades que requeiram o uso de planilhas eletrônicas, softwares matemáticos e linguagem de programação. Os registros serão feitos através das folhas de tarefas e cada aluno será identificado apenas por um número, ou seja, não haverá identificação nominal. Vale salientar também que fotografias podem ser realizadas durante a aplicação das tarefas, mas sem focar nos rostos ou partes do corpo que possam identificar algum estudante.

3. POTENCIAIS RISCOS E BENEFÍCIOS: Toda pesquisa oferece algum tipo de risco. Nesta pesquisa, o risco pode ser avaliado como mínimo, isto é, o participante pode apresentar timidez e se sentir envergonhado ou não se sentir confortável em participar de alguma atividade, ou pode apresentar algum desconforto em exteriorizar as suas respostas nas tarefas propostas. Objetivando minimizar esses riscos, o participante tem a possibilidade de decidir pela participação ou não em quaisquer atividades e a qualquer momento, sem problema algum. Por outro lado, são esperados os seguintes benefícios da participação na pesquisa: contribuição para o conhecimento sobre aprendizado de máquina, inteligência artificial e modelo de regressão linear simples, introdução à programação em Python e consequente aumento de sua autoestima, além de contribuir para a melhoria da Educação Matemática nas Escolas Públicas.

4. GARANTIA DE SIGILO: os dados da pesquisa poderão ser publicados/divulgados em livros e revistas científicas. Asseguramos que a sua privacidade será respeitada e o seu nome ou qualquer informação que possa, de alguma forma, o (a) identificar, será mantida em sigilo. O pesquisador responsável se compromete a manter os dados da pesquisa em arquivo, sob sua guarda e responsabilidade, por um período mínimo de 5 (cinco) anos após o término da pesquisa.

5. LIBERDADE DE RECUSA: a sua participação neste estudo é voluntária e não é obrigatória. Você poderá se recusar a participar do estudo ou retirar seu consentimento a qualquer momento, sem precisar justificar. Se desejar sair da pesquisa você não sofrerá qualquer prejuízo.

6. CUSTOS, REMUNERAÇÃO E INDENIZAÇÃO: a participação neste estudo não terá custos adicionais para você. Também não haverá qualquer tipo de pagamento devido a sua participação no estudo. Fica garantida indenização em casos de danos, comprovadamente decorrentes da participação na pesquisa, nos termos da Lei.

7. ESCLARECIMENTOS ADICIONAIS, CRÍTICAS, SUGESTÕES E RECLAMAÇÕES: você receberá uma via deste Termo de Consentimento Livre e Esclarecido (TCLE) e a outra ficará com o pesquisador. Caso você concorde em participar, as páginas serão rubricadas e a última página será assinada por você e pelo pesquisador. O pesquisador garante a você livre acesso a todas as informações e esclarecimentos adicionais sobre o estudo e suas consequências. Você poderá ter acesso ao pesquisador Ronaldo Cesar da Silva pelo telefone (21) 98721-0223, ou pelo e-mail: ronaldocesar73@gmail.com. Se você tiver alguma consideração ou dúvida sobre a ética da pesquisa, poderá entrar em contato com o Comitê de Ética em Pesquisa do Colégio Pedro II (CEP/CP II), situado no Endereço: Campo de São Cristóvão nº 177, prédio da Pró-Reitoria de Pós-Graduação, Pesquisa, Extensão e Cultura (PROPGPEC), sala 202-B – São Cristóvão – Rio de Janeiro, CEP 29921-903, pelo telefone: 21 3891-0020 ou pelo e-mail: cep@cp2.g12.br

CONSENTIMENTO

Eu, _____ li e concordo em participar da pesquisa.

Assinatura do(a) participante	Data: ___/___/___
-------------------------------	-------------------

Eu, Ronaldo Cesar da Silva, obtive de forma apropriada e voluntária o Consentimento Livre e Esclarecido do(a) participante da pesquisa.

Assinatura do pesquisador	Data: ___/___/___
---------------------------	-------------------

ANEXO D – Parecer do Conselho de Ética

COLÉGIO PEDRO II



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Introdução ao Aprendizado de Máquina utilizando o Modelo de Regressão Linear Simples: uma proposta para estudantes do 3º ano do Ensino Médio.

Pesquisador: RONALDO CESAR DA SILVA

Área Temática:

Versão: 1

CAAE: 89054825.1.0000.9047

Instituição Proponente: Colégio Pedro II

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 7.631.322

Apresentação do Projeto:

As informações colocadas nos campos denominados "Apresentação do Projeto", "Objetivo da Pesquisa" e "Avaliação dos Riscos e Benefícios" foram retiradas do documento intitulado "PB_INFORMATIONES_BÁSICAS_DO_PROJETO_2542175.pdf" (submetido na Plataforma Brasil em 19/05/2025).

Objetivo da Pesquisa:

Segundo o pesquisador:

Objetivo Primário:

Este projeto pretende apresentar uma introdução aos conceitos básicos de aprendizado de máquina utilizando o modelo de regressão linear simples com base teórica na produção de significados sob a ótica dos modelos dos campos semânticos. O objetivo geral é analisar a produção de significados dos estudantes no âmbito do aprendizado de máquina, com foco restrito no modelo de regressão linear simples, com o intuito de analisar e identificar situações do mundo atual que podem ser preditas utilizando o modelo de regressão estudado.

Endereço: Campo de São Cristóvão, 177, São Cristóvão, prédio da PROPGPEC, Sala 202-B

Bairro: São Cristóvão **CEP:** 20.921-903

UF: RJ **Município:** RIO DE JANEIRO

Telefone: (21)2163-5730

E-mail: cep@cp2.g12.br

COLÉGIO PEDRO II



Continuação do Parecer: 7.631.322

Objetivo Secundário:

- 1- Observar como o estudante constrói significado sobre Aprendizagem de Máquina o Modelo de Regressão Linear Simples;
- 2- Investigar relações entre números expressos em tabelas para representá-los no plano cartesiano, identificando padrões e criando uma generalização que pode ser representada por uma função polinomial do 1º grau;
- 3- Observar um conjunto de dados relativos a duas variáveis numéricas, com ou sem a utilização de computadores e identificar se essa relação entre as variáveis pode ser ajustada por uma reta;
- 4- Identificar se um conjunto de pontos (x, y) no gráfico de dispersão pode ser ajustado através de um modelo de regressão linear simples.

Avaliação dos Riscos e Benefícios:

Segundo o pesquisador:

Riscos:

Toda pesquisa oferece algum tipo de risco. Neste trabalho, o risco pode ser avaliado como mínimo, isto é, o participante pode apresentar timidez e se sentir envergonhado ou não se sentir confortável em participar de alguma atividade, ou pode apresentar algum desconforto em exteriorizar as suas respostas nas tarefas propostas. Objetivando minimizar esses riscos, o participante tem a possibilidade de decidir pela participação ou não em quaisquer atividades e a qualquer momento.

Benefícios:

Contribuição para o conhecimento sobre aprendizado de máquina, inteligência artificial e modelo de regressão linear simples, introdução à programação em Python e consequente

Endereço: Campo de São Cristóvão, 177, São Cristóvão, prédio da PROPGPEC, Sala 202-B
Bairro: São Cristóvão **CEP:** 20.921-903
UF: RJ **Município:** RIO DE JANEIRO
Telefone: (21)2163-5730 **E-mail:** cep@cp2.g12.br

COLÉGIO PEDRO II



Continuação do Parecer: 7.631.322

aumento de sua autoestima.

Comentários e Considerações sobre a Pesquisa:

(1) Trata-se de uma pesquisa de caráter quantitativo/qualitativo com elementos da pesquisa-ação. Este projeto busca produzir significado contribuindo para uma introdução ao aprendizado de máquina (machine learning) para estudantes do 3º ano do ensino médio, usando o modelo preditivo de Regressão Linear Simples (ajuste de dados para uma função polinomial do 1º grau). O aprendizado de máquina é um campo da Inteligência Artificial que usa algoritmos com base em modelos estatísticos que fazem o computador reconhecer padrões baseando-se em dados, melhorando e aprimorando a resposta na medida em que são fornecidos mais dados. Esta proposta 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). Sob o prisma dos Modelos dos Campos Semânticos (MCS) do professor Rômulo Campos Lins (1955-2017), a produção de significado acontece numa enunciação e quem produz uma enunciação é o autor. O leitor só se institui como autor na medida em que faz suas reflexões e constrói sua própria versão do enunciado e sendo assim, cada autor é único. Dessa forma, cada estudante participa ativamente do seu processo de ensino-aprendizagem, sendo capazes 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.

(2) São esperados 50 participantes de pesquisa no Brasil como consta no arquivo intitulado "PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_2542175.pdf", postado em 19/05/2025.

(3) A duração do estudo será de aproximadamente 8 meses no Brasil, como consta no arquivo intitulado "PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_2542175.pdf", postado em 19/05/2025.

Considerações sobre os Termos de apresentação obrigatória:

Conferir item ¿Conclusões ou Pendências e Listas de Inadequações¿.

Endereço: Campo de São Cristóvão, 177, São Cristóvão, prédio da PROPGPEC, Sala 202-B
Bairro: São Cristóvão **CEP:** 20.921-903
UF: RJ **Município:** RIO DE JANEIRO
Telefone: (21)2163-5730 **E-mail:** cep@cp2.g12.br

COLÉGIO PEDRO II



Continuação do Parecer: 7.631.322

Recomendações:

Após a devida apresentação do projeto de pesquisa, recomenda-se, de acordo com o item 3.4.1 da Norma Operacional 001/2013, que o item "Critérios de Inclusão e Exclusão" esteja escrito de forma explícita no campo destinado da Plataforma Brasil. Solicita-se que essas recomendações sejam submetidas sob a forma de emenda.

Conclusões ou Pendências e Lista de Inadequações:

--- LER ATENTAMENTE O CAMPO "RECOMENDAÇÕES" ---

Considerações Finais a critério do CEP:

1. De acordo com o item X.1.3.b, da Resolução CNS n. 466/12, o pesquisador deverá apresentar relatórios semestrais - a contar da data de aprovação do protocolo - que permitam ao Cep acompanhar o desenvolvimento dos projetos.

Esses relatórios devem ser assinados pelo pesquisador responsável e conter as informações detalhadas - naqueles itens aplicáveis - nos moldes do relatório final contido no Ofício Circular n. 062/2011: <http://conselho.saude.gov.br/web_comissoes/conep/aquivos/conep/relatorio_final_encerramento.pdf>, bem como deve haver menção ao período a que se referem. As informações contidas no relatório devem ater-se ao período correspondente e não a todo o período da pesquisa até aquele momento. Para cada relatório, deve haver uma notificação separada. A submissão deve ser como Notificação (consultar pág. 69 no arquivo intitulado 1 - Manual Pesquisador- Versão 3.2, disponível no endereço <http://plataformabrasil.saude.gov.br/login.jsf>. Anexar em arquivo com recurso <copiar e colar>.

2. Eventuais emendas (modificações) ao protocolo devem ser apresentadas de forma clara e sucinta, identificando-se, por cor, negrito ou sublinhado, a parte do documento a ser modificada, isto é, além de apresentar o resumo das alterações, juntamente com a justificativa, é necessário destacá-las no decorrer do texto (item 2.2.1.H.1, da Norma Operacional CNS nº 001 de 2013)

3. O Cep lembra que o pesquisador deve ainda (1) encaminhar os resultados da pesquisa para publicação, com os devidos créditos aos pesquisadores associados e ao pessoal técnico integrante do projeto (Res. CNS 466/12 item XI.g); (2) divulgar os resultados para os participantes da pesquisa e para as instituições onde os dados foram obtidos (Norma Operacional nº 001/2013 item 3.4.14); (3) anexar os resultados da pesquisa na Plataforma

Endereço: Campo de São Cristóvão, 177, São Cristóvão, prédio da PROPGPEC, Sala 202-B
Bairro: São Cristóvão **CEP:** 20.921-903
UF: RJ **Município:** RIO DE JANEIRO
Telefone: (21)2163-5730 **E-mail:** cep@cp2.g12.br

COLÉGIO PEDRO II



Continuação do Parecer: 7.631.322

Brasil, garantindo o sigilo relativo às propriedades intelectuais e patentes industriais (Norma Operacional nº 001/2013 item 3.3.c) e (4) comunicar às autoridades competentes, bem como aos órgãos legitimados pelo Controle Social, dos resultados e/ou dos achados da pesquisa, sempre que esses puderem contribuir para a melhoria das condições de vida da coletividade, preservando, porém, a imagem e assegurando que os participantes da pesquisa não sejam estigmatizados (Res. CNS 466/2012 item III.1.m). Essas providências devem ser tomadas no prazo máximo de seis meses, contados a partir da data da emissão deste parecer.

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB INFORMAÇÕES BÁSICAS_DO_PROJETO_2542175.pdf	19/05/2025 14:01:11		Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_RESPONSAVEL_LEGAL.pdf	19/05/2025 13:58:51	RONALDO CESAR DA SILVA	Aceito
Outros	Documento_de_resposta_pendencias_assinado.pdf	19/05/2025 13:57:25	RONALDO CESAR DA SILVA	Aceito
Projeto Detalhado / Brochura Investigador	Projeto_de_etica_Ronaldo.pdf	05/05/2025 20:32:49	RONALDO CESAR DA SILVA	Aceito
Declaração de Instituição e Infraestrutura	Declaracao_Instituicao.pdf	05/05/2025 20:29:43	RONALDO CESAR DA SILVA	Aceito
Outros	Anexo_B.pdf	05/05/2025 20:26:55	RONALDO CESAR DA SILVA	Aceito
Outros	Anexo_A.pdf	05/05/2025 20:26:33	RONALDO CESAR DA SILVA	Aceito
Cronograma	CRONOGRAMA_E_CURRICULO_LATTES_assinado_assinado.pdf	04/05/2025 21:03:27	RONALDO CESAR DA SILVA	Aceito
Outros	Carta_de_Apresentacao_ao_Comite_de_Etica_assinado.pdf	03/05/2025 22:47:18	RONALDO CESAR DA SILVA	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_MAIORES_de_idade.pdf	03/05/2025 22:28:17	RONALDO CESAR DA SILVA	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	Termo_de_Assentimento_Livre_e_Esclarecido.pdf	03/05/2025 22:27:34	RONALDO CESAR DA SILVA	Aceito
Outros	Curriculo_Lattes.pdf	03/05/2025	RONALDO CESAR	Aceito

Endereço: Campo de São Cristóvão, 177, São Cristóvão, prédio da PROPGPEC, Sala 202-B
Bairro: São Cristóvão **CEP:** 20.921-903
UF: RJ **Município:** RIO DE JANEIRO
Telefone: (21)2163-5730 **E-mail:** cep@cp2.g12.br

COLÉGIO PEDRO II



Continuação do Parecer: 7.631.322

Outros	Curriculo_Lattes.pdf	22:01:24	DA SILVA	Aceito
Outros	TERMO_DE_CONFIDENCIALIDADE_a assinado.pdf	03/05/2025 21:55:57	RONALDO CESAR DA SILVA	Aceito
Outros	DECLARACAO_ISENCAO_DE_CUSTO S_ assinado.pdf	03/05/2025 21:55:03	RONALDO CESAR DA SILVA	Aceito
Orçamento	Orçamento_ assinado.pdf	03/05/2025 21:51:42	RONALDO CESAR DA SILVA	Aceito
Folha de Rosto	Ronaldo_Profmat_CPII_ assinado.pdf	01/05/2025 17:18:24	RONALDO CESAR DA SILVA	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

RIO DE JANEIRO, 11 de Junho de 2025

Assinado por:
ROGERIO MENDES DE LIMA
(Coordenador(a))

Endereço: Campo de São Cristóvão, 177, São Cristóvão, prédio da PROGPPEC, Sala 202-B
Bairro: São Cristóvão **CEP:** 20.921-903
UF: RJ **Município:** RIO DE JANEIRO
Telefone: (21)2163-5730 **E-mail:** cep@cp2.g12.br