

Submission: 20/Oct/2024; 1st round notif.: 20/Jan/2025; New version: 27/Feb/2025; 2nd round notif.: 25/Mar/2025;
Camera ready: 26/Apr/2025; Edition review: 26/Apr/2025; Available online: 19/Aug/2025; Published: 19/Aug/2025;

Autoformação e Literacia Estatística: projeto de Modelagem Matemática em Regressões Linear e Polinomial com Python

Title: *Self-Formation and Statistical Literacy: A Mathematical Modeling Project in Linear and Polynomial Regressions with Python*

Título: *Autoformación y Alfabetización Estadística: Proyecto de Modelado Matemático en Regresiones Lineal y Polinómica con Python*

Luciano Pontes da Silva
Universidade Federal de Sergipe
ORCID: [0000-0001-8923-3414](https://orcid.org/0000-0001-8923-3414)
pontesmatematicaufs@yahoo.com

Anne Alilma Silva Souza Ferrete
Universidade Federal de Sergipe
ORCID: [0000-0001-9637-6616](https://orcid.org/0000-0001-9637-6616)
aferrete21@gmail.com

Resumo

A sociedade informacional e sua produção de dados em larga escala pressupõe nos sujeitos pertencentes ferramentas e habilidades capazes de entender esses preceitos. O presente trabalho apresenta uma proposta de projeto de Modelagem Matemática, integrando as Tecnologias Móveis Digitais de Informação e Comunicação (TMDIC), mais precisamente no uso da linguagem de programação Python para construção de modelos de Regressão Linear e Polinomial. É apresentado, primeiramente, a teoria da Autoformação e os conceitos contidos nela, como Ecoformação e Heteroformação. Logo em seguida, é mostrado conceitos acerca da Educação Estatística, mais precisamente a Literacia Estatística e conceituando a Modelagem Matemática integrada as TMDIC com exemplos de projetos aplicados com essa abordagem. Por fim, a proposta de projeto, divida em três partes, onde serão discutidos o problema inicial, conceitos matemáticos essenciais e implementação de códigos para o modelo e a culminância do projeto. Essa proposta potencializa a percepção dos níveis de pluralidade da realidade, além da possibilidade de desenvolver posturas de questionamento desses níveis de pluralidade e intentando a sua modificação.

Palavras-chave: Autoformação; Literacia Estatística; Educação Estatística; Linguagem Python; Regressão Linear; Regressão Polinomial.

Abstract

The informational society and its large-scale data production presuppose that the individuals belonging to it possess tools and skills capable of understanding these precepts. This work presents a Mathematical Modeling project proposal, integrating Mobile Digital Information and Communication Technologies (MDICT), specifically using the Python programming language to build Linear and Polynomial Regression models. First, the theory of Self-Formation is presented, along with the concepts it contains, such as Ecoformation and Heteroformation. Next, concepts regarding Statistical Education are introduced, with a focus on Statistical Literacy, and the integration of Mathematical Modeling with MDICT is explained, providing examples of projects that apply this approach. Finally, the project proposal is divided into three parts, where the initial problem, essential mathematical concepts, and code implementation for the model, as well as the project's culmination, are discussed. This proposal enhances the perception of the levels of plurality in reality, along with the possibility of developing questioning attitudes towards these levels and aiming at their modification.

Keywords: Self-Formation; Statistical Literacy; Statistical Education; Python Language; Linear Regression; Polynomial Regression.

Resumen

La sociedad informacional y su producción de datos a gran escala presupone que los individuos que pertenecen a ella posean herramientas y habilidades capaces de comprender estos preceptos. Este trabajo presenta una propuesta de proyecto de Modelado Matemático, integrando las Tecnologías Móviles Digitales de Información y la Comunicación (TMDIC), más precisamente en el uso del lenguaje de programación Python para la construcción de modelos de Regresión Lineal y Polinómica. Primero, se presenta la teoría de la Autoformación y los conceptos que contiene, como la Ecoformación y la Heteroformación. Luego, se muestran conceptos relacionados con la Educación Estadística, más precisamente la Alfabetización Estadística, y se conceptualiza el Modelado Matemático integrado a las TMDIC, con ejemplos de proyectos aplicados con este enfoque. Finalmente, la propuesta del proyecto, dividida en tres partes, discute el problema inicial, los conceptos matemáticos esenciales, la implementación de códigos para el modelo y la culminación del proyecto. Esta propuesta potencia la percepción de los niveles de pluralidad de la realidad, además de la posibilidad de desarrollar actitudes de cuestionamiento hacia estos niveles de pluralidad, con la intención de su modificación.

Palabras clave: Autoformación; Alfabetización Estadística; Educación Estadística; Lenguaje Python; Regresión Lineal; Regresión Polinómica.

1 Introdução

A sociedade atual, denominada informacional, exige dos seus participantes habilidades e ferramentas para compreender e inferir sobre um montante de informação em constante crescimento, movimento denominado *Big Data* (Davenport, 2014). E os cidadãos, para poderem entender e modificar de forma benéfica a realidade que os cerca, precisam desenvolver adequadamente essas ferramentas e habilidades.

Dentro da educação, a Base Nacional Comum Curricular (BNCC) (Brasil. Ministério da Educação, 2017, p.272), delega à Educação Estatística, para os professores da matemática, parte desse desenvolvimento, quando propõe “a abordagem de conceitos, fatos e procedimentos presentes na vida cotidiana, das ciências e da tecnologia.” Esse documento ainda recomenda que os cidadãos precisam desenvolver habilidades para coletar, representar, interpretar e organizar dados, para fazerem julgamentos adequados da realidade.

Essa perspectiva é reforçada também nas diretrizes para a formação de professores (BNC-Formação), quando comenta que o professor deve:

Compreender, utilizar e criar tecnologias digitais de informação e comunicação de forma crítica, significativa, reflexiva e ética nas diversas práticas docentes, como recurso pedagógico e como ferramenta de formação, para comunicar, acessar e disseminar informações, produzir conhecimentos, resolver problemas e potencializar as aprendizagens.(Brasil. Conselho Nacional de Educação, 2020, p.13)

Essa Educação Estatística, voltada para uma compreensão geral da realidade, do questionamento de forma crítica desta, contrasta com os níveis de habilidades que os jovens brasileiros possuem. No último PISA (Programme for International Student Assessment), realizado pela OCDE (Organização para Cooperação e Desenvolvimento Econômico) 73% dos alunos brasileiros possuem conhecimento matemático abaixo do nível esperado (nível 2). A média de conhecimentos matemáticos é de 379, abaixo de países como Chile e Uruguai, com 412 e 409 respectivamente (Serviços e Informações do Brasil, 2024).

Esses dados ainda mostram que apenas 1% dos jovens brasileiros possuem nível elevado de conhecimentos matemáticos (nível 5). O texto sugere que a maioria destes não possuem conhecimento suficiente para exercer efetivamente sua cidadania, de forma crítica e contundente. O estudo apresenta também desafios à Educação para a Ciência e Literatura, com a proeficiência nesses campos abarcando problemas e ensejando metas.

Dentro desse contexto, no uso das tecnologias, o relatório da UNESCO (2023) mostra que os jovens reconhecem a potencialidade dessas ferramentas para a sua aprendizagem, porém para eles a tecnologia não supera a interação humana, reconhecendo assim o papel do professor e do ensino de forma efetiva. Ferrete (2007) comenta que essa estrutura social e o processo de ensino e aprendizagem integrados à tecnologia devem ser sempre discutidos para uma construção do conhecimento adequado.

Diante dessas assertivas, surge a seguinte questão: qual abordagem pode potencializar a resolução desses desafios dentro da formação docente, no que diz respeito à Educação Estatística e as Tecnologias Móveis Digitais de Informação e Comunicação (TMDIC)?

O objetivo deste trabalho é apresentar uma proposta de projeto dentro da Autoformação e da Literacia Estatística, utilizando a Modelagem Matemática integrada às TMDIC com os estudos de Regressão Linear e Polinomial, usando a linguagem de programação *Python*¹ no auxílio da construção do modelo de regressão.

Na primeira seção, será apresentada a Teoria da Autoformação, baseada nos estudos de Galvani (2002). Essa teoria trará uma abordagem dos entendimento do sujeito que se forma, entendendo as manifestações de si (eu), dos outros (hétero) e das coisas (Eco) como tríplice fundante da sua formação.

A segunda seção, dividida em duas subseções, trará primeiramente um panorama sobre a Educação Estatística dentro da abordagem *learning by doing* (aprender fazendo) que comunga com os objetivos desse trabalho, seguido do entendimento de Literacia Estatística na visão de autores fundamentais e apresentando a Modelagem Matemática integrada às TMDIC com exemplos de projetos relacionados à Educação Estatística.

Por fim, a proposta de projeto é discutida em três partes, sendo a primeira o problema condutor, trazendo a problemática do mundo real para a sala, emulando discussões acerca desse, seguido da modelagem do problema, introduzindo conceitos matemáticos necessários como o MSE e o R^2 , trazendo os códigos *Python* para a Modelagem e finalizando com a culminância.

Essa pesquisa é financiada com recursos da Coordenação de Aperfeiçoamento de Pessoal do Nível Superior (CAPES) incluído nos estudos do Núcleo de Pesquisa em Comunicação e Tecnologia (NUCA/UFS/CNPq). Faz parte de um projeto de pesquisa de Tese de Doutorado, da Universidade Federal de Sergipe (UFS), tendo como CAAE: 84415324.1.0000.5546², submetido ao Comitê de Ética em Pesquisa da UFS e já aprovado para execução, seguindo o que rege a Resolução CNS nº 466/2012 do Conselho Nacional de Saúde (Conselho Nacional de Saúde, 2012).

Os dados empíricos serão produzidos com alunos do curso de Licenciatura em Matemática, da Universidade Federal de Alagoas, Campus Arapiraca, o que será validado e discutido em trabalhos futuros. Cabe também o parêntese feito na seção da proposta, que é o critério da amostra baseado na problemática apresentada nessa introdução, mas podendo ser abarcada para outros nichos.

2 A Autoformação

Pascal Galvani é professor da Universidade de Quebec em Rimouski, na qual é coordenador da revista *Présences* e da coleção “*Écologie et formation*”. As suas pesquisas são pautadas dentro dos processos de Autoformação e Ecoformação. Sua tese foi centrada no significado de Autoformação com o uso do brasão para compreender sua dimensão semiótica.

Na perspectiva desse pensador, existe um processo tripolar, onde as Auto (o Eu, de si) concatenam-se com a Hétero (os outros e a influência desses) e a Eco (relativos ao meio físico, ambiente).

Galvani (2002) explica que a necessidade de pensar uma formação transdisciplinar perpassa

¹<https://www.python.org>

²O projeto pode ser consultado em <https://plataformabrasil.saude.gov.br/login.jsf>

duas questões vitais: o desequilíbrio ecológico, que é a busca intermitente do lucro e a crise antropológica, devido às confluências culturais na Era Moderna. Esse materialismo se resume em reduzir o mundo em realidade material, em divisão do conhecimento que recorta essas múltiplas realidades sem as coadunar.

O movimento proposto pelo autor dever inverter essa lógica, desenvolvendo uma abordagem interior da educação, o que o mesmo denomina autoformação. Essa nova abordagem deve conter bases pensadas na pluralidade dos níveis de realidade, sob a ótica transdisciplinar, comunicando dois conceitos fulcrais nesses níveis: auto (si) e formação. Também deve conter a questão antropológica, que desse modo requer uma abordagem transcultural. Segundo o mesmo:

Esse triplo movimento de tomada de consciência e de tomada de poder da pessoa sobre sua formação parece ser a base de uma definição conceitual da autoformação. A autoformação aparece aqui como o surgimento de uma consciência original na interação com o meio ambiente. A autoformação se caracteriza pelo imbricamento da reflexividade e da interação entre a pessoa e o meio ambiente (Galvani, 2002, p.4).

Pretende-se, dessa maneira, harmonizar os três eixos da perspectiva tripolar: Eu (Autoformação), os Outros (Heteroformação) e as Coisas (Ecoformação). Esses três não são dissociados uns dos outros, muito pelo contrário, formam uma engrenagem que se retroalimenta.

Esse esquema mostra que a abordagem tripolar se amálgama e não se pode dissociar os elementos, os quais compõem esse esquema, pois a autoformação é interna, da perspectiva do sujeito que se autoforma. Partindo desse esquema, Galvani (2002) explica o que são heteroformação e ecoformação:

O processo de formação conduzido pelo pólo hétero inclui a educação, as influências sociais herdadas da família, do meio social e da cultura, das ações de formação inicial e contínua, etc. Essa terraformação é definida e hierarquizada de maneira heterônima pelo meio ambiente cultural. A formação conduzida pelo pólo eco se compõe das influências físicas, climáticas, e das interações físicas corporais que dão forma à pessoa. Ela inclui também uma dimensão simbólica. O meio ambiente físico em todas as suas variedades (florestas, desertos, países temperados, metrópoles urbanas, etc.) produz uma forte influência sobre as culturas humanas, bem como sobre o imaginário pessoal, que organiza o sentido dado à experiência vivida (Galvani, 2002, p.3)

A autoformação então é o processo conduzido pelo sujeito. No esquema, S_1 e S_2 se caracterizam pela tomada de consciência dadas as retroações físicas e sociais. O S_3 por sua vez se caracteriza pela tomada do sujeito sobre si, sua subjetivação. Desse modo, a autoformação é a tomada de consciência de si (o Eu) da heteroformação (os Outros) e da ecoformação (as Coisas), num processo em moto contínuo. Essas interações com o físico e o social são embricadas e subvertem o esquema de educação transmissiva – tradicional (Galvani, 2002). A Figura 1 ilustra essas assertivas:

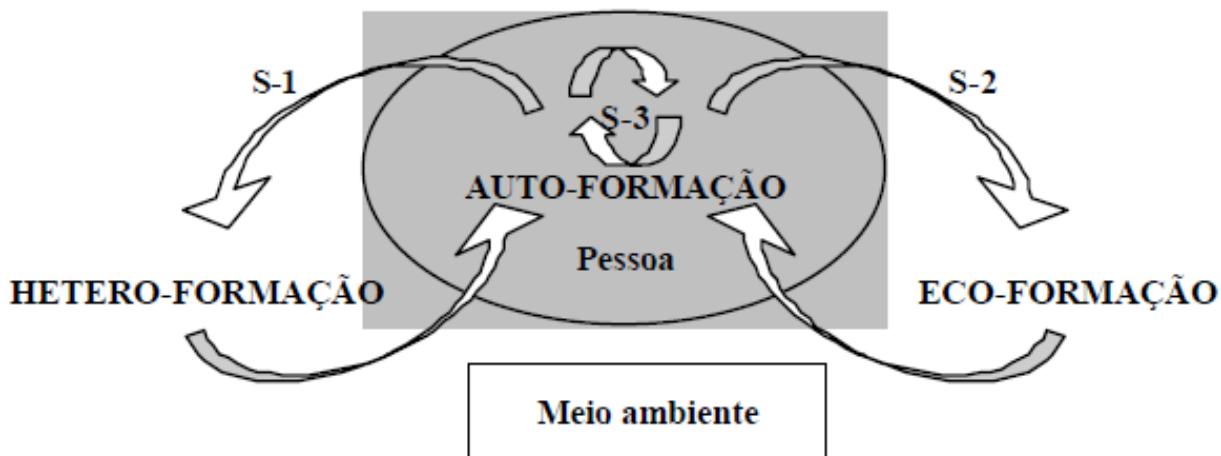


Figura 1: Autoformação.

A Autoformação, nesse ínterim, é caracterizada pela reflexividade e interação entre a pessoa e o meio ambiente, contendo as relações com as demais pessoas. Os três processos declinados da autoformação em questões de retroação são: a retroação de si (subjetivação), a retroação do meio ambiente social (socialização) e a retroação do meio ambiente físico (ecologização).

Essa base é oriunda da abordagem antropológica da formação, pois essas combinações entre reflexão das retroações e das interações tripolares demarcam essa concepção (Galvani, 2002). Em outras palavras: temos a dinâmica da morfogênese (manifestação dessas estruturas retroativas) e a metamorfose (transformação destas).

A abordagem transdisciplinar concatena o paradigma da complexidade teorizado por Morin (2000), dentro de um sistema complexo, associando a ideia da unidade e da multiplicidade, as quais se repelem e se unem num moto-contínuo e exigindo uma complementariedade desses dois polos. Portanto, a questão da formação nessa perspectiva também corrobora para uma união entre a subjetividade e as objetividades, dentro de uma praxeologia, sem uma hierarquização sobre as práticas ou teorias (Morin, 2000; Estrada, 2009).

Destarte, pressupõe-se que a teoria e a prática estão dentro de uma hierarquia na formação inicial matemática, onde então as disciplinas que ajudam o professor no seu local de trabalho, no caso específico da matemática, se sobrepõem às disciplinas pedagógicas. Um curso que visa formar professores não pode aderir à ideia de que o professor que sabe matemática sabe ensinar matemática (Ponte, 2002).

Para avançar no sentido dessa abordagem, o pensador conduz a sua teorização primeiramente na análise dos níveis de consciência reflexiva das auto para depois comentar sobre os níveis de interação que compõem a formação, aportando então essa perspectiva antropológica. Na perspectiva da consciência reflexiva o foco é na pessoa, no sujeito consciente e como se constitui como pessoa. Logo após, disserta sobre os níveis de pluralidade da realidade e como essas duas discussões formam os conceitos das *Auto*.

Essas discussões serão amalgamadas na próxima seção, na qual a Autoformação encontra os preceitos da Literacia Estatística, na tentativa de se produzir materiais de Modelagem Matemática

que possam subsidiar a formação do professor no que diz respeito à Educação Estatística.

3 A Educação Estatística: *learning by doing* (aprender fazendo)

A centralidade do professor será transduzida para um programa de estudo que pense o currículo na aprendizagem discente. A subjetividade é colocada como motor inicial do ensino. O aluno é convidado a participar ativamente da sua aprendizagem, pois aquele conhecimento tem relação com seu consuetudinário, com sua comunidade e seu locus de trabalho (Campos et al., 2021).

Esses alunos, nessa abordagem, devem então trazer para a aula problemas que condizem com os seus interesses, onde são instigados a formularem hipóteses, coletar dados e escolher quais os métodos estatísticos são mais adequados. Em suma, são levados a refletir sobre as soluções que formularam, levando em consideração os conceitos de incerteza e variabilidade, dentro da Estatística.

Garfield e Gal (1999) elaboraram algumas metas que perfazem essa capacitação dos alunos:

- Entender o propósito e a lógica das investigações estatísticas;
- Entender o processo de investigação estatística;
- Dominar as habilidades usadas no processo de investigação estatística;
- Entender a probabilidade, a chance, a incerteza, os modelos e a simulação;
- Entender as relações matemáticas presentes nos conceitos estatísticos;
- Desenvolver habilidades interpretativas para argumentar, criticar e refletir;
- Desenvolver habilidades para se comunicar estatisticamente, usando corretamente a sua terminologia.

Campos et al. (2021) sintetizam essas metas, quando colocam sobre desenvolver habilidades cooperativas e colaborativas, em trabalhos em equipe, desenvolver habilidades de transposição de saberes escolares para o cotidiano e de habilidades de questionamento de valores, grandezas, dados e informações. Para a obtenção no contexto escolar, esses autores colocam algumas estratégias, não engessando a maneira de desenvolvê-las, a saber:

1. O foco do ensino de Estatística deve ser desviado do produto para o processo. No trabalho com a inferência, por exemplo, é mais importante a compreensão dos processos de amostragem e da coleta de dados do que a obtenção do resultado final, conseguida através das fórmulas apropriadas e disponíveis em livros-textos ou apresentadas pelo professor.
2. Como consequência dessa valorização do produto, a análise e interpretação de dados estatísticos são mais importantes que as técnicas.

3. O uso de tecnologia deve ser incorporado ao ensino de Estatística, permitindo grandes possibilidades de simulações e mostrando que o cálculo pode ser feito pela máquina, mas a análise dos dados, interpretações e tomada de decisões, não.
4. A aprendizagem de Estatística fazendo estatística é a chave da motivação. Smith (1998) afirma que os trabalhos com projetos nos quais os alunos coletam dados, organizam esses dados, apresentam e interpretam resultados, produzem relatórios, gráficos, pareceres, etc. têm se mostrado extremamente frutífero para que as metas listadas acima sejam, ao menos parcialmente alcançadas. Para isso, é necessário produzir exemplos que tenham significação prática para os alunos [...].
5. Os alunos devem ser incitados a argumentar, interpretar e analisar, mais do que a calcular ou desenhar.
6. A implementação de estratégias de aprendizagem colaborativa e o encorajamento do trabalho em grupo têm suscitado casos de sucesso, como apontado por vários autores, como Garfield e Ben-Zvi (2008) e Smith (1998).
7. As avaliações devem estar voltadas para o cumprimento de metas e não para cálculos e aplicações de fórmulas (Campos et al., 2021, p.15)

Os pontos 1) e 2) são cruciais para o entendimento da abordagem que visa colocar o aluno de frente aos problemas que são familiares a ele. A análise, pautada num processo construído pelo próprio discente, corrobora para subverter o paradigma que foi posto por Galvani (2002). O ponto 3) é de também fundamental para esse trabalho, pois as TMDIC são o elo entre o processo do *learning by doing* e a autoformação. de Faria Gomes et al. (2017) explicitam que o uso das TMDIC na interação professor/aluno/objeto podem propiciar um salto qualitativo nas habilidades e capacidades tanto de alunos quanto de professores. E Alves et al. (2024) comentam sobre essas alterações de comportamento, atitude e emoções com relação à TMDIC e as pessoas, reforçando esse salto qualitativo.

Os pontos 4), 5) e 6) reforçam ainda mais a abordagem do aprender fazendo, pois no trabalho de Smith (1998) percebemos a modulação do foco nas aulas do professor para o aluno; as leituras não são mais passivas, sendo o mais importante ter atividades práticas, superando o que Skovsmose (2000) define como paradigma do exercício: situação didática na qual o professor apresenta a tarefa e o aluno trabalha na solução da mesma, o que seria a discussão sobre o ponto 7).

Na tentativa de desenvolver essas estratégias, pensando nas metas supracitadas, desenvolvemos as competências que são necessárias para a instrução dos estudantes de Estatísticas, baseadas nos trabalhos de Rumsey (2002), Garfield e Ben-Zvi (2008), Chance (1998), Delmas (2002) e Campos et al. (2021) que trazem três competências: A Literacia, o Raciocínio e o Pensamento Estatístico.

O conceito de competência aqui é percebido de acordo os estudos de Perrenoud (2000), sendo a mobilização de recursos cognitivos (capacidades, saberes, informações) para solucionar uma série de tarefas, de forma pertinente e eficaz. O sociólogo suíço acredita que a escola tem uma preocupação voltada mais para as fórmulas e ingredientes do que as competências. Para que haja

um efetivo desenvolvimento de destas é necessário trabalhar os alunos em situações complexas, não marginalizando as situações da vida deles.

Em questão de ordem destas, tomamos a postura de Delmas (2002), que prediz a relação embricada nas três. Em seu trabalho podemos encontrar duas interpretações sobre essas competências, onde a primeira é sobre uma interdependência e interseções parciais entre as mesmas. Se partirmos desse pressuposto, então podemos desenvolver essas competências de forma independente das demais e ao mesmo tempo atividades que possam desenvolver as três ou duas a duas, conforme a Figura 2 a seguir:

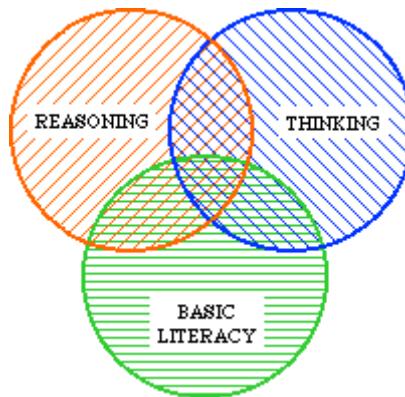


Figura 2: Domínios Independentes, com suas interseções *Fonte: Delmas, 2002.*

A segunda interpretação apresenta a Literacia Estatística como o conjunto que abarca as outras duas competências, o que significa que se o aluno possui literacia então ele tem o Raciocínio e o Pensamento Estatístico bem desenvolvido. Segundo Delmas (2002) essa interpretação requer mais vivência do aluno dentro e fora da sala de aula nos quesitos da Estatística, o que torna mais complexa essa abordagem, conforme a Figura 3:

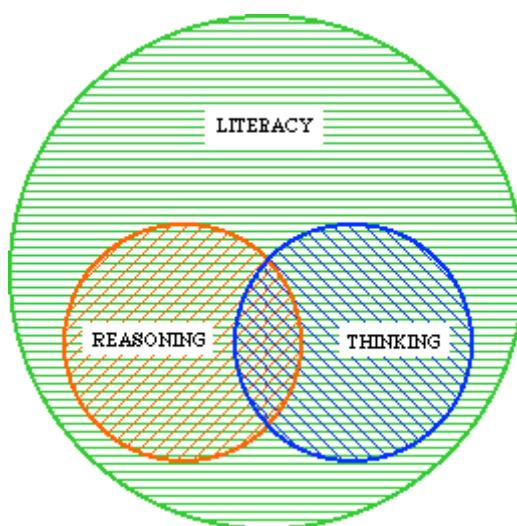


Figura 3: Literacia contendo as outras competências *Fonte: Delmas, 2002.*

Nesse trabalho será tomada a segunda postura, apesar da complexidade comentada pelo autor supracitado, pois o interesse é pautado aqui pelos problemas que surgem do cotidiano e os alunos de graduação possuem maior bagagem do que alunos do ensino básico, portanto a vivência da sala de aula e o público-alvo a qual o projeto está sendo proposto tem experiência em docência, incluindo estágios supervisionados ou programas de iniciação à docência (PIBID, Residência Pedagógica, dentre outros).

3.1 O que é a Literacia Estatística?

Na colaboração entre os estatísticos e os educadores matemáticos, percebeu-se alguns avanços na Educação Estatística, sobretudo na década de 1980, nos EUA, onde o projeto *Quantitative Literacy Project* (QLP) foi responsável por trazer ideias pedagógicas contudentes sobre a Literacia Estatística. Segundo Scheaffer (1990), alguns princípios devem ser seguidos na tentativa de trazer à tona essa literacia:

- análise de dados é a principal ação pedagógica;
- Estatística não é Probabilidade;
- medidas como mediana e quartis (amplitude quartílica) deve desempenhar um papel tão importante quanto a média e a variância;
- há mais de uma maneira de trabalhar um problema em estatística;
- devem ser usados dados reais e de interesse dos alunos;
- ênfase deve estar em bons exemplos e na valorização da intuição
- estudantes devem escrever mais e calcular menos;
- Estatística ensanada nas escolas deve ser importante e útil para os estudantes em seu cotidiano (Scheaffer, 1990, p.46).

Segundo Campos et al. (2021), o termo literacia³ define-se como a habilidade de ler, compreender, interpretar, analisar e avaliar textos escritos. Para a Literacia Estatística, essas habilidades estão relacionadas à argumentação estatística, ligada também às informações, organização de dados, símbolos e entendimento de conceitos-chave como incerteza e probabilidade.

Outros autores trazem versões diferentes dessa mesma ideia. Haack (1979), primeiro autor a pensar sobre esse termo, comenta em seu trabalho que essa competência está relacionada ao conhecimento técnico de estatística. Watson (1997) comprehende a Literacia em três estágios:

1. o entendimento básico da terminologia estatística;
2. o entendimento da linguagem estatística e dos conceitos inseridos num contexto de discussão social;

³Tradução do Inglês *literacy*. Alguns autores denominam "letramento", porém o termo nesse trabalho será utilizado na sua forma original, sem perda de sentido e significado.

3. o desenvolvimento de atitudes de questionamento nas quais se aplicam conceitos mais sofisticados para contradizer alegações que são feitas sem fundamentação estatística adequada (Watson, 1997, p.3)

Para Garfield e Gal (1999), a Literacia Estatística é o entendimento da linguagem estatística, isto é, suas terminologias, simbologia e trazem as habilidades necessárias para a interpretação de gráficos e tabelas, além de entender as informações advindas de veículos de informação. Sedlmeier (1999) aponta a Literacia como a arte de inferenciar racionalmente, baseando-se na quantidade extensa de informações e números promovidos pela mídia, configurando-se como indispensável para o exercício da cidadania.

Rumsey (2002) identifica a Literacia Estatística nessa mesma questão da cidadania. Para essa autora, os alunos precisam entender suficientemente as informações, para que se possa criticá-las de forma coerente e tomar boas decisões a partir desse entendimento. Essa autora ainda distingue essa competência da habilidade para atuar como pessoa educada estatisticamente, trazendo o termo cidadania estatística.

Desse modo, esses autores dissertam sobre a Literacia Estatística como o entendimento e o uso das ferramentas estatísticas adequadas para se tomar boas decisões em determinados problemas do cotidiano, tendo também a incumbência de comunicar às pessoas sobre as inferências desenvolvidas para a resolução.

3.2 Modelagem Matemática e as Tecnologias Móveis Digitais de Informação e Comunicação

Dentro da questão da Educação Matemática, onde está inserida a Educação Estatística, a abordagem que esse projeto toma é baseada na Modelagem Matemática, tendo a elaboração de projetos focada no desenvolvimento das competências, no caso a Literacia Estatística (Campos et al., 2021).

Bassanezi (2002) comenta que a Modelagem Matemática é um processo dinâmico, na tentativa de obter a validação de modelos matemáticos e consiste, fundamentalmente, em transformar situações do nosso cotidiano em problemas matemáticos, resolvendo-os assim com uma linguagem usual. Essa proximidade com os quesitos da Matemática Aplicada, área da Matemática que visa estudar fenômenos do mundo real, potencializa o papel pedagógico da Modelagem Matemática.

Dessa maneira, a Modelagem Matemática pode ser entendida como um método de ensino e aprendizagem, podendo ser aplicado nos diversos níveis escolares. Partindo do pressuposto de Barbosa (2007), este método se situa num ambiente de aprendizagem, pelo fato dos estudantes serem instigados pelo professor a investigarem situações retiradas do dia a dia desses.

Com isso, afirma-se que as Tecnologias são essenciais para se obter êxito nesses processos. Desse modo, salientam Campos et al. (2021, p.49): “vemos o papel da tecnologia informática como ator indispensável ao trabalho com Modelagem Matemática, quer como ferramenta de apoio operacional, quer como instrumento que venha contribuir para a superação de vários desafios.” Esses desafios podem ser entendidos como o desinteresse do aluno, a falta de habilidade para o mercado de trabalho, dentre outros.

Salientam ainda esses autores que o modo de pensar modelagem matemática juntamente com as Tecnologias devem pautar as investigações e reflexões acerca de determinado tema, para que assim possa ter significado para os participantes do projeto e não seja apenas para entender o conteúdo. A tecnologia, desse modo, deve estar integrada aos conhecimentos necessários para desenvolver esses projetos.

Algumas experiências com esse tipo de abordagem demonstram a capacidade de pensar a estatística para a literacia, superando o que Skovsmose (2000) chama de paradigma do exercício. Dentro do GPEE (Grupo de Pesquisas em Educação Estatística) da UNESP, duas delas, encontradas em Campos et al. (2021) podem ser explicitadas para a finalidade desse trabalho.

A primeira é referente a uma pesquisa com professores de Matemática, em um curso de extensão, de modo síncrono e assíncrono, com o objetivo de analisar e compreender o processo de interação, estabelecido dentro de um ambiente virtual de modelagem matemática, desenvolvido pelos professores com seus alunos, utilizando o Excel. Como resultados, os trabalhos possibilitaram, de modo colaborativo, o entendimento de conceitos sobre probabilidade e estatística nos três níveis de ensino (básico, médio e superior).

A segunda tem como problemática inicial as eleições presidenciais de 2006, utilizando também um ambiente virtual de aprendizagem, envolvendo professores, tanto das escolas públicas quanto das privadas. O projeto foi denominado "Modelagem Matemática e Eleições Presidenciais", no qual professores e alunos foram convidados a coletar hipertextos com intenções de voto nessa referida eleição. 13 professores participaram do projeto, oriundos de diferentes locais do país. Os dados produzidos foram disponibilizados para posteriores trabalhos com outros contextos similares (Campos et al., 2021).

Outros estudos que direcionam sobre essa questão são de Borba e Penteado (2007) que abordam a utilização da tecnologia informática na sala de aula e Borba et al. (2011) que discutem questões práticas e teóricas no Ensino a Distância (EaD), com um capítulo da obra reservado para Modelagem Matemática e EaD. Esses dois trabalhos mostram como um ambiente virtual, denominado Centro Virtual de Aprendizagem, facilita a interação dos professores e alunos sobre temas pertinentes à Educação Estatística.

O trabalho de Andrade (2008) que utilizou temas como o Alcoolismo na Adolescência como catalisador de ideias para projeto de modelagem matemática também pode ser destacado. A autora, baseando-se nos Parâmetros Curriculares Nacionais (PCN) e outros documentos relevantes, propôs a temática do projeto dentro da Ciência de Dados, por meio da Modelagem Matemática, com o auxílio do Excel. Os resultados dessa pesquisa foram enfatizados em categorias que se relacionam com a abordagem da modelagem.

Por fim, o trabalho de Cambraia et al. (2022) corrobora para o desenvolvimento do pensamento computacional em professores de informática, concatenando não somente os docentes, mas também o professor formador da disciplina que está diretamente ligada à essa temática. A análise desses autores explicitou os conhecimentos didáticos que os professores desenvolviam com seus alunos e a inter-relação entre o mesmo visto na graduação, tendo uma interface também com a autoformação.

Portanto, os projetos de intervenção desenvolvidos juntamente aos alunos, têm como suporte teórico a Modelagem Matemática com o uso das tecnologias, mais precisamente as TMDIC, pela

facilidade e popularização do uso dos smartphones, nos locais onde computadores e laboratórios possam ser insuficientes para o desenvolvimento desses projetos.

4 Proposta de projeto

Comentada as questões teóricas, emerge então elaborar um projeto de intervenção que conte empre os princípios da Autoformação de Galvani (2002) e as questões da Literacia Estatística, na perspectiva de Campos et al. (2021), Haack (1979), Scheaffer (1990), Garfield e Gal (1999), Watson (1997) e Rumsey (2002), utilizando a Modelagem Matemática e Tecnologias integradas como caminho para amalgamar essas duas assertivas.

A proposta de projeto se intitula "**Diabetes, Regressão Linear e Polinomial: ajustando modelos com o MSE (Mean Square Error ou Erro Quadrático Médio) e o R^2 (R quadrado)**" e tem como público alvo discentes de graduação, podendo ser Licenciatura ou Bacharelado em Matemática, Estatística e áreas afins, como Administração, Economia, etc. Como o pressuposto inicial é sobre Autoformação na perspectiva da Formação Docente, o projeto dará ênfase no primeiro público supracitado.

Para situar os preceitos da Autoformação e Literacia Estatística, pontuam-se como objetivos almejados:

1. Entender sobre os níveis de pluralidade da realidade (Eco e Heteroformação);
2. Debater sobre um problema do cotidiano estatisticamente, desenvolvendo a Literacia Estatística;
3. Trazer uma abordagem transdisciplinar para o problema, contemplando assim a Autoformação e a Literacia Estatística;
4. Potencializar a incorporação em práticas (futuras ou atuais) docentes dos discentes dentro dessas abordagens.

O projeto será apresentado em três partes, que serão explicitadas a seguir. Cada parte terá a sugestão do desenvolvimento do problema, não perdendo de vista os objetivos traçados, podendo também aparecer, conforme o contexto aplicado, outros objetivos. Os encontros (ou aulas) estão sugeridos nos títulos, porém também podem ser flexionados pelo mesmo motivo anterior.

4.1 Parte 1: Apresentando o Problema (2 a 3 encontros)

Nessa primeira parte, os alunos devem se familiarizar com o problema, trazendo estatísticas e questões sociais com o tema discutido, neste caso a Diabetes Mellitus. Para iniciar, temos o artigo do Ministério da Saúde (2024b), contendo informações técnicas e estatísticas. A outra sugestão é também um artigo do Ministério da Saúde (2024a) sobre o dia 26 de junho, Dia Nacional do Diabetes, que de maneira análoga traz alguns apontamentos.

Em ambos os artigos, podemos enfatizar algumas preposições, para que sejam debatidas entre os discentes, dentre as quais, sugere-se:

- "De acordo com a Sociedade Brasileira de Diabetes, existem atualmente, no Brasil, mais de 13 milhões de pessoas vivendo com a doença, o que representa 6,9% da população nacional"
- "O Brasil é o 5º país em incidência de diabetes no mundo, com 16,8 milhões de doentes adultos (20 a 79 anos), perdendo apenas para China, Índia, Estados Unidos e Paquistão. A estimativa da incidência da doença em 2030 chega a 21,5 milhões. Esses dados estão no Atlas do Diabetes da Federação Internacional de Diabetes (IDF)."
- "Mundialmente, o diabetes se tornou um sério problema de saúde pública, cujas previsões vêm sendo superadas a cada nova triagem. Por exemplo, em 2000, a estimativa global de adultos vivendo com diabetes era de 151 milhões. Em 2009, havia crescido 88%, para 285 milhões. Em 2020, calcula-se que 9,3% dos adultos, entre 20 e 79 anos (assombrosos 463 milhões de pessoas) vivem com diabetes. Além disso, 1,1 milhão de crianças e adolescentes com menos de 20 anos apresentam diabetes tipo 1."

Essas informações possuem dados que podem ser discutidos com os participantes, além de instigá-los a pensar além. Um exemplo disso é a informação sobre a taxa da população brasileira e os tratamentos oferecidos pelo Sistema Único de Saúde (SUS), em quais classes econômicas essas pessoas se encontram, etc.

A última informação pode ser trabalhada no sentido de pensar em estimativas, juntamente com as duas anteriores. Pode-se afunilar a pesquisa para os estados e municípios, já criando neles um sentimento "inferencial" sobre os dados, pois nesse sentido compete a eles entender a sua própria realidade e transmutar isso numa prática docente.

Essa sugestão de discussão pode ser emulada para uma pesquisa de campo. Os discentes podem ir aos postos de saúde de suas cidades ou bairros e conversar com profissionais como enfermeiros(as) ou médicos(as), podendo obter informações cruciais para uma culminância do projeto no final do percurso, o que é abarcado também pela Literacia Estatística (comunicar as informações).

Para discutir a diferença entre os tipos de Diabetes, a Figura 4 pode ser utilizada:



Figura 4: Tipos de Diabetes Fonte: Vale Mais Saúde, 2024.

Essa figura traz informações cruciais dentro da temática que está sendo trabalhada. Na Figura 4 os alunos podem além de discutirem sobre os tipos de Diabetes, colocarem experiências próprias, como parentes ou amigos que convivem com a doença.

A Figura 5 apresenta um acréscimo de informações, onde a sintomatologia é apresentada. Os alunos podem discutir sobre se já tiveram esses sintomas ou se já perceberam em algum fami-

liar ou amigo, ou se já coletoou na internet ou conhece alguém da área da saúde que relatou acerca desses.

A questão nessas duas figuras é trabalhar com os participantes de modo que, além de trazerm o problema dentro de uma perspectiva do cotidiano, possam também organizar os dados mais importantes. Pode ser sugerido a eles a criação de uma tabela com os sintomas e de acordo com os seus conhecimentos, classificá-los como mais visíveis ou mais silenciosos, ou até mesmo com uma pesquisa, verificando se esses sintomas estão dentro do conjunto de outras doenças (auto-imunes ou não).

Tipos:

- Tipo 1: causado pela destruição das células produtoras de insulina, em decorrência de defeito do sistema imunológico em que os anticorpos atacam as células que produzem a insulina. Ocorre em cerca de 5 a 10% dos diabéticos.
- Tipo 2: resulta da resistência à insulina e de deficiência na sua secreção. Ocorre em cerca de 90% dos diabéticos.
- Diabetes Gestacional: é a diminuição da tolerância à glicose, diagnosticada pela primeira vez na gestação, podendo ou não persistir após o parto. Sua causa exata ainda não é conhecida.
- Outros tipos: são decorrentes de defeitos genéticos associados com outras doenças ou com o uso de medicamentos.

Podem ser: defeitos genéticos da função da célula beta; defeitos genéticos na ação da insulina; doenças do pâncreas (pancreatite, neoplasia, hemocromatose, fibrose cística, etc.); induzidos por drogas ou produtos químicos (diuréticos, corticoides, betabloqueadores, contraceptivos, etc.).

Principais sintomas do DM tipo 1:

- vontade de urinar diversas vezes;
- fome frequente;
- sede constante;
- perda de peso;
- fraqueza;
- fadiga;
- nervosismo;
- mudanças de humor;
- náusea;
- vômito.

Principais sintomas do DM tipo 2:

- infecções frequentes;
- alteração visual (visão embaçada);
- dificuldade na cicatrização de feridas;
- formigamento nos pés;
- furúnculos.

Figura 5: Tipos e Sintomas de Diabetes *Fonte: Ministério da Saúde, 2024a.*

Após todas essas interações, para encerrar esse momento, os alunos podem produzir algum texto ou material midiático, como vídeos ou podcasts entre eles, na tentativa de explicar a doença e como as pessoas acometidas são impactadas por ela. Esse material não precisa, a priori, ser divulgado, porque pode ser utilizado na culminância do projeto.

Em termos de exemplificação, a Autoformação e a Literacia Estatística foram trabalhadas de forma conjunta, abarcando os objetivos 1) e 2), preparando os discentes ou participantes para entender, com as tecnologias, o problema na perspectiva da Modelagem Matemática.

4.2 Parte 2: Modelando o Problema (3 a 4 encontros)

Para essa seção, os discentes poderão utilizar tanto dispositivos móveis (celulares, notebooks, tablets) quanto computadores em um laboratório físico. A ideia aqui é não fechar as possibilidades para encontros síncronos, mas também deixar atividades que eles possam realizar em suas casas, ou em encontros virtuais em plataformas como Discord ou Meet.

Para a primeira atividade, serão utilizados dados reais de um paciente diagnosticado com Diabetes tipo 2⁴. O arquivo é do tipo ".csv", por sua formatação ser largamente empregada em análise de dados.

As variáveis são a taxa de glicose medida em jejum e os dias de medição (foram observados 50 dias). O objetivo é saber se é possível criar um modelo de Regressão que descreva o mais próximo possível as taxas de glicose ao longo do tempo.

Dividiremos essa parte em algumas subpartes, na tentativa de abranger os conhecimentos necessários para que os discentes possam entender efetivamente como o processo de modelagem do problema aconteceu.

4.2.1 Regressão Linear e Polinomial

Definição 1 (Regressão Linear) *Seja y uma variável dependente e o conjunto x_1, x_2, \dots, x_n as variáveis independentes, dentro de uma combinação linear. O modelo de Regressão Linear é dado por:*

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon \quad (1)$$

onde β_0 é o intercepto, β_i , $i = 1, 2, 3, \dots, n$ são os coeficientes de regressão e ε é o Erro Quadrático Médio.

Se for necessário para facilitar o entendimento dessa definição, o professor pode trazer alguns exemplos simples, encontrados em Almeida (2015), antes da modelagem principal. Utilizando a Regressão Linear Simples $y = \beta_0 + \beta_1 x_1 + \varepsilon$, podemos explicitar e discutir alguns conceitos:

1. O que β_0 significa na prática?
2. Como identificar as variáveis independentes e a variável dependente?
3. Como encontramos os melhores coeficientes de regressão?
4. Como saber se esse é um modelo bom ou não?

O terceiro e quarto item terão uma atenção especial. Antes de comentar sobre os mesmos, precisamos definir a Regressão Polinomial:

⁴o link para acesso a esses dados se encontra no apêndice 1

Definição 2 (Regressão Polinomial) Seja y uma variável dependente e x a variável independente. O modelo de Regressão Linear Polinomial (de grau k) é dado por:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_k x^k + \varepsilon \quad (2)$$

onde β_0 é o intercepto, $\beta_i, i = 1, 2, 3, \dots, n$ são os coeficientes de regressão e ε é o Erro Quadrático Médio.

Os trabalhos de Draper e Smith (1998) e Kutner et al. (2005) trazem exemplos das mais variadas áreas do conhecimento, podendo ser utilizados para facilitar a compreensão dos conceitos. Outros questionamentos podem aparecer com essa definição:

- Quais os termos que aparecem nas duas formas de Regressão?
- Por que temos variáveis em combinação linear e na Regressão Polinomial isso não acontece?
- Quando escolher entre Regressão Linear e Polinomial?
- Por que aparece o "Erro Quadrático Médio" (ε) nas duas formas?

Antes das tentativas de respostas dos alunos acontecerem na prática, precisamos definir outros conceitos que serão fundamentais para a modelagem.

4.2.2 Erro Quadrático Médio e R Quadrado

Definição 3 (Erro Quadrático Médio) Sejam $y_i, i = 1, 2, 3, \dots, n$, os valores observados no problema e sejam $\hat{y}_i, i = 1, 2, 3, \dots, n$, os valores preditos pela Regressão. O Erro Quadrático Médio (ou Mean Squared Error - MSE) é calculado como:

$$\varepsilon = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

Sem perda de generalização, podemos usar um exemplo simples para que os alunos possam perceber como calcular isso na prática:

Exemplo 1 Suponhamos três observações em um problema qualquer: $y_1 = 3$, $y_2 = 4$ e $y_3 = 5$ e calculamos na Regressão esses valores $\hat{y}_1 = 3.2$, $\hat{y}_2 = 3.9$ e $\hat{y}_3 = 4.1$. O Erro Quadrático Médio é calculado:

$$\begin{aligned} \varepsilon &= \frac{1}{3} \sum_{i=1}^3 (y_i - \hat{y}_i)^2 = \frac{1}{3} [(3 - 3.2)^2 + (4 - 3.9)^2 + (5 - 4.1)^2] \\ &= 0.286 \end{aligned}$$

A partir desse exemplo, podemos pensar em algumas questões:

- O erro é dividido pela quantidade de amostras, por esse motivo ele é "medio";

- A intenção é reduzir o erro a zero. Quanto menor a distância entre o valor observado e o valor predito, menor será a distância "ao quadrado" desses;
- O modelo é mais "forte" se o seu erro for o menor possível;
- A diferença $(y_i - \hat{y}_i)$ é chamada de **Resíduo**.

Para esse exemplo, podemos ver que o erro é bem próximo de zero, o que mostra que a Regressão utilizada descreve bem o fenômeno apresentado. O professor pode definir com os alunos qual o valor aceitável para um erro dentro de uma Regressão.

Definição 4 (R^2) Sejam y_i , $i = 1, 2, 3, \dots, n$, os valores observados no problema, sejam \hat{y}_i $i = 1, 2, 3, \dots, n$, os valores preditos pela Regressão e \bar{y} a média dos valores observados. o Coeficiente de Determinação (R^2) é calculado como:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \quad (4)$$

Exemplo 2 Usando o exemplo anterior, calculemos então a média dos valores observados:

$$\bar{y} = \frac{3 + 4 + 5}{3} = 4$$

e então calculamos o Coeficiente de Determinação:

$$\begin{aligned} R^2 &= 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} = 1 - \frac{(3 - 3.2)^2 + (4 - 3.9)^2 + (5 - 4.1)^2}{(3 - 4)^2 + (4 - 4)^2 + (5 - 4)^2} \\ &= 1 - \frac{0.86}{2} = 0.57 \end{aligned}$$

Levantamos então algumas preposições:

- O R^2 depende de ε ;
- $0 \leq R^2 \leq 1$, isto é, esse coeficiente mede a proporção de variabilidade do modelo;
- Se o valor for muito próximo de 0 (ou igual à 0) o modelo não explica o fenômeno, sendo considerado aleatório;
- Se o valor for próximo de 1 (ou igual a 1) então o modelo explica totalmente, ou próximo disso;
- A diferença $(y_i - \bar{y})^2$ é o **Resíduo Quadrático Médio**.

No exemplo, podemos ver que o valor do coeficiente é 0.57, ou seja, o modelo explica 57% dos dados, o que não seria considerado um bom modelo. Isso também demonstra a importância de não apenas pensar no Erro Quadrático Médio na construção de uma Regressão, seja Linear ou Polinomial.

Apresentados os conceitos-chave para o entendimento da modelagem do problema, utilizaremos agora a linguagem de programação *Python* e o compilador on-line e gratuito *Google Colab*. Lembrando também que esses códigos estão prontos, portanto, se os discentes envolvidos não tiverem prática nessa linguagem, poderão ser apresentados sem discussão das linhas, servindo como base de aprendizagem para o *Python*.

4.2.3 Códigos para a Análise dos Dados e Regressão

Com o compilador aberto, a primeira coisa é fazer *upload* do arquivo .csv que contém os dados, como mostra a Figura 6:

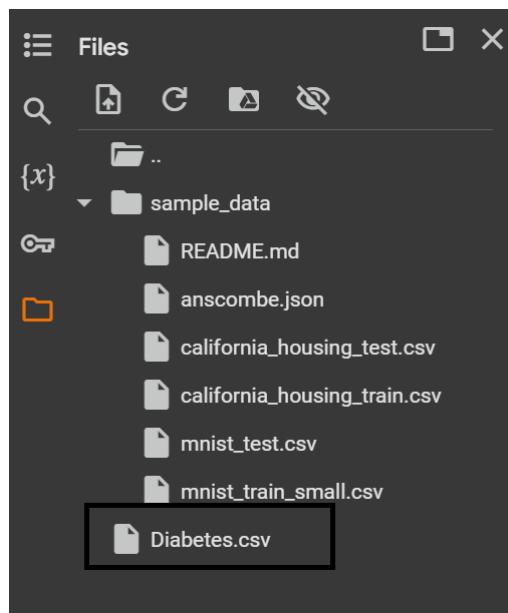


Figura 6: Upload do arquivo *Fonte: Os autores*.

Após o arquivo ter sido inserido na pasta "*sample_data*", as primeiras linhas de código se referem às bibliotecas que iremos utilizar na análise, como seguem na Figura 7:

```
import pandas as pd
from sklearn.metrics import mean_squared_error, r2_score
import numpy as np
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
import sklearn.model_selection as ms
import sklearn.linear_model as lm
import matplotlib.pyplot as plt
from sklearn.preprocessing import PolynomialFeatures
```

Figura 7: Importando as Bibliotecas no Python *Fonte: Os autores.*

Em seguida, devemos importar os dados aos quais fizemos o *upload*. Usaremos o formato de *Data Frame* importado da biblioteca "Pandas", para manipular os dados de acordo com as intenções do projeto, mostrando os 20 primeiros dias do experimento (Figura 9). Aproveitamos para deixar as colunas normalizadas para que o modelo de Regressão funcione (Figura 8). Também faremos um ".describe()" na coluna "Taxa/remedio" visualizando as informações do nosso conjunto de dados, como estatísticas descritivas (contagem, média, desvio padrão, mínimo, quartis (25%, 50%, 75% e 100%) e o máximo) na Figura 10:

```
df = pd.read_csv('/content/Diabetes.csv')
y = df['Taxa/remedio'].values.reshape(-1,1)
X = df['Dia'].values.reshape(-1,1)
df
df['Taxa/remedio'].describe()
```

Figura 8: Criando o Data Frame e Normalizando as colunas *Fonte: Os autores.*

	Taxa/remedio	Dia
0	118	1
1	144	2
2	114	3
3	120	4
4	128	5
5	118	6
6	145	7
7	107	8
8	116	9
9	103	10
10	127	11
11	130	12
12	135	13
13	117	14
14	134	15
15	119	16
16	143	17
17	118	18
18	140	19
19	128	20
20	125	21

Figura 9: *Data Frame* (20 primeiros dias) *Fonte: Os autores.*

Taxa/remedio	
count	51.000000
mean	123.274510
std	12.909033
min	99.000000
25%	115.500000
50%	121.000000
75%	131.500000
max	157.000000
dtype: float64	

Figura 10: *Data Frame* (.describe()) *Fonte: Os autores.*

Algumas informações podem ser discutidas, como o tipo de caracteres ser "float64", mais precisamente se são números reais, mas os números são inteiros, o que mostra uma característica

das estatísticas descritivas, além dos máximos e mínimos do conjunto, para se pensar no desvio padrão e homogeneidade da amostra.

Plotamos o gráfico de dispersão, que se refere aos pontos correspondentes (do dia e da taxa em função do mesmo) conforme o código da Figura 11, mostrado na Figura 12 :

```
plt.scatter(X,y, color = 'red')
plt.show
```

Figura 11: Código de Dispersão *Fonte: Os autores.*

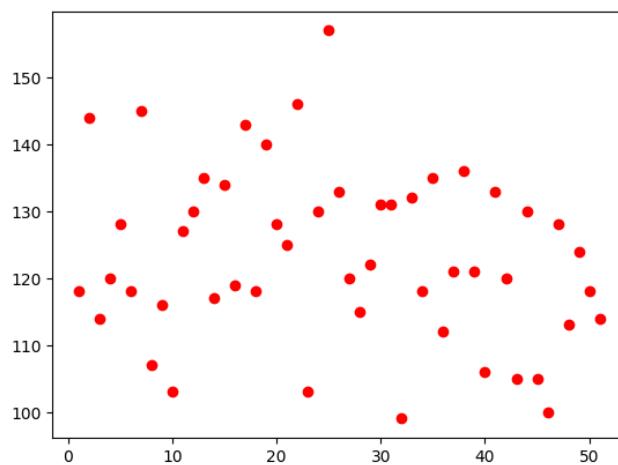


Figura 12: Gráfico de Dispersão *Fonte: Os autores.*

Agora utilizaremos um algoritmo para treino dos dados, tanto para o modelo linear, quanto para o modelo polinomial. Esse algoritmo pode ser entendido do que comumente chamamos de "Algoritmos de Aprendizagem de Máquina", pois podemos inserir dados futuros à nossa pesquisa, podendo assim ter resultados comparados com os reais e os do modelo de treinamento, conforme a Figura 13:

```
linear_regressor = lm.LinearRegression()
linear_regressor.fit(X, y)
poly_regressor = PolynomialFeatures(degree=2)
X_poly_2 = poly_regressor.fit_transform(X)
linear_regressor_2 = lm.LinearRegression()
linear_regressor_2.fit(X_poly_2, y)
```

Figura 13: Código de Treino de Modelo *Fonte: Os autores.*

Agora calculamos o MSE e o R^2 para o modelo linear:

```
rmse_1 = np.sqrt(mean_squared_error(y, linear_regressor.predict(X)))
r2_1 = r2_score(y, linear_regressor.predict(X))
print(rmse_1)
print(r2_1)
```

Figura 14: Código de MSE e R para Linear *Fonte: Os autores.*

o que nos retorna:

$$MSE = 12.494254368149049$$

e

$$R^2 = 0.044493984671139186$$

Seguimos calculando o MSE e o R^2 para o modelo polinomial:

```
rmse_2 = np.sqrt(mean_squared_error(y, linear_regressor_2.predict(X_poly_2)))
r2_2 = r2_score(y, linear_regressor_2.predict(X_poly_2))
print('rmse grau 2:', rmse_2)
print('r2 grau 2:', r2_2)
```

Figura 15: Código de MSE e R para Linear *Fonte: Os autores.*

obtendo:

$$MSE = 12.171660755893974$$

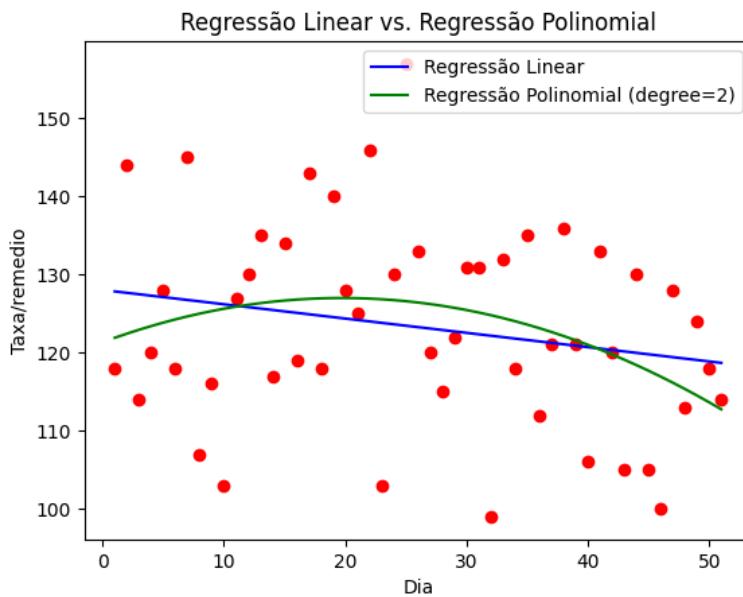
e

$$R^2 = 0.09319810850323074$$

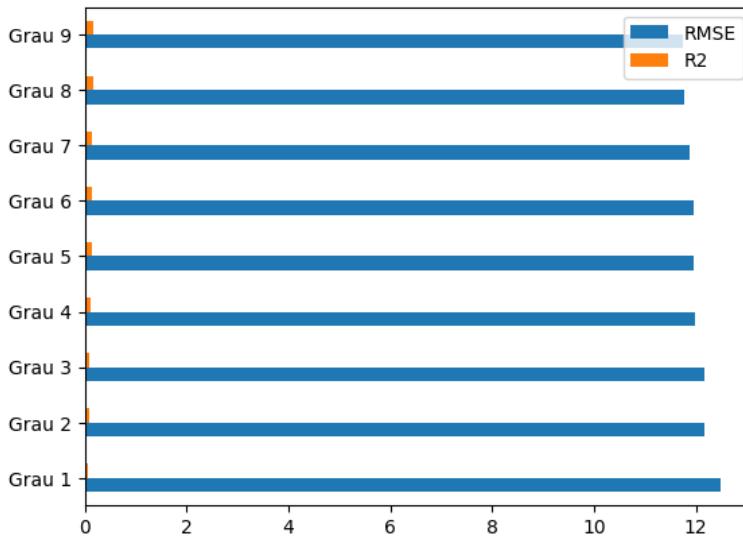
Plotamos agora o gráfico dos dois modelos. O objetivo de deixar em um único gráfico é ver quais as possibilidades de análise de qual regressão está correspondendo mais aos dados. Os gráficos podem ser plotados separadamente, para fins de estudos de cada um deles.

```
plt.scatter(X, y, color='red')
plt.plot(X, linear_regressor.predict(X), color='blue', label='Regressão Linear')
plt.plot(X, linear_regressor_2.predict(X_poly_2), color='green', label='Regressão Polinomial (degree=2)')
plt.title('Regressão Linear vs. Regressão Polinomial')
plt.xlabel('Dia')
plt.ylabel('Taxa/remédio')
plt.legend()
plt.show()
```

Figura 16: Código do gráfico das Regressões *Fonte: Os autores.*

Figura 17: Regressão Polinomial Quadrática e Linear *Fonte: Os autores.*

Os discentes agora podem fazer para a quantidade de graus que for necessário. Como sugestão, deixamos um gráfico de barra com os valores do MSE e do R^2 até o grau 9, como mostra a Figura 18:

Figura 18: Gráfico de Barras do MSE e R^2 até o grau 9 *Fonte: Os autores.*

Algumas questões podem ser levantadas:

1. O MSE está diminuindo de uma maneira significativa?
2. O mesmo se pergunta pro R^2 , se está aumentando dessa forma?

3. Quantos graus você testou?
4. O modelo de Regressão descreve bem esse problema?
5. Qual dos dois modelos (Linear ou Polinomial Quadrático) interceptou mais pontos no gráfico de dispersão? O que isso tem a ver com o MSE e o R^2 ?

Essas questões podem ser aplicadas no momento final do projeto, no qual os alunos deverão produzir e explicar as regressões com dados próprios. A divisão pode ser em grupos de três pessoas para que o trabalho de cada integrante seja evidenciado.

4.3 Parte 3: Culminância (2 a 3 encontros)

Nessa última parte do projeto de Modelagem Matemática, os discentes já devem ter os conhecimentos necessários para apresentarem seus próprios projetos. Essa parte final tem como objetivo suscitar alguns debates tanto acerca da importância de se conhecer a doença quanto para as questões da Regressão.

A sugestão de culminância que deixamos aqui está dividida em duas partes:

1. Assistir antes da culminância o vídeo "Diabetes"⁵ disponível no YouTube;
2. Trazer um profissional da área da Saúde (pode ser do curso de Medicina ou Enfermagem) para apresentar a doença, seus riscos e seus sintomas.

O debate com o profissional pode ainda abranger algumas dúvidas que os alunos têm sobre a doença e seus tratamentos, sobre hábitos alimentares que podem desenvolvê-la, tendo em vista também os próprios alunos e as suas condições como estudantes universitários. Essa problemática pode gerar outros projetos de intervenção dos alunos, auxiliados por estudantes e professores de outros cursos.

Nessa apresentação, os alunos podem explicitar como o MSE e o R^2 foram fundamentais para avaliar se o modelo é eficaz ou não, com os testes com graus polinomiais maiores, por exemplo, ou pesquisando sobre outro tipo de Regressão (Logaritmica, Poisson, Ridge, Lasso e com *Support Vector Machines*). Além disso, a discussão sobre os resíduos e a sua normalidade e homogeneidade pode ser fundamental para se entender quando escolher ou não um modelo de regressão.

Esses outros tipos de regressão podem ser apresentados em outros momentos e também discutidos com colegas da computação ou informática, na tentativa de perceber se os códigos estão bem desenvolvidos e assim fomentar ainda mais o projeto dentro de outras perspectivas. E, com o auxílio de professores de estatística e computação, realizar testes de normalidade e homogeneidade (Breusch-Pagan e Shapiro-Wilk, por exemplo).

Deve-se solicitar a eles, usando até mesmo uma inteligência artificial (*ChatGPT* ou a *Gemini* do *Google Colab*) para que criem previsões de valores nos modelos gerados. Os gráficos obtidos no final da lista de modelos podem ser não apenas de barra, mas de outros tipos. Também pode-se

⁵<https://www.youtube.com/watch?v=nyvu2euX8tM>

criar um gráfico sobre o crescimento ou decaimento dos dois parâmetros estudados no projeto, na regressão, para avaliar se é interessante usar outras regressões.

As apresentações também podem fornecer produtos, como manuais, cenários clicáveis ou cartilhas para os alunos de outros cursos e para as pessoas no geral, comentando sobre a importância de se conhecer a doença e explicando como analisar dados da vida real, tentando modelá-los, pode não ser uma tarefa fácil.

5 Conclusão

A sugestão de projeto sobre Modelagem Matemática apresentada nesse trabalho mostra uma perspectiva sobre essa abordagem da Educação Matemática dentro da Educação Estatística com a Literacia Estatística, mas com o acréscimo de pensar na Autoformação dos discentes.

Os meandros que esse projeto pode tomar, por ser uma proposta, ainda são desconhecidos, cabendo ao professor formador ter em mente todas as possibilidades, sejam elas ônus ou bônus e mitigar os problemas que surgirem com as etapas. Estas também podem ser subdivididas ou remanejadas, de acordo com cada necessidade e contexto.

A Autoformação, em consonância com os preceitos da Literacia Estatística, pode contribuir para sanar possíveis lacunas na formação, principalmente quando se trata da Educação Estatística, vista de forma técnica e sem correlação com a realidade e o cotidiano dos discentes. Se o graduando repetir práticas que são oriundas da sua formação, isso ficará ainda mais evidente.

Com as TMDIC aliadas e integradas à Modelagem Matemática, nesse caso utilizando uma linguagem de fácil manuseio como o *Python*, potencializa ainda mais essa abordagem, pois integra as ferramentas digitais à formação docente e dá subsídio ao futuro professor utilizar essas em suas práticas na sala de aula, síncronas ou não.

Desse modo, aliar as perspectivas que foram apresentadas nesse trabalho fornecem um olhar crítico sobre os níveis de pluralidade da realidade ao formando, percebendo a si mesmo no mundo que está inserido e capaz não somente de descrevê-lo, mas também modificá-lo. Desenvolver aprendizagem colaborativas, habilidades e competências se mostram como um passo fundamental para o intuito deste trabalho.

Por fim, o projeto servirá de base para produção de dados e assim, poderá ser analisado do ponto de vista científico sobre sua validação e adequações, para que se tenha o objetivo da proposta almejado.

Referências

- Almeida, R. N. d. (2015). *O método dos mínimos quadrados: estudo e aplicações para o ensino médio* [Dissertação (Mestrado Acadêmico em Matemática)]. Universidade Federal Fluminense. <https://uenf.br/posgraduacao/matematica/wp-content/uploads/sites/14/2017/09/28052015Renato-Neves-de-Almeida.pdf>

- Alves, M. M. S., Ferrete, A. A. S. S., Vasconcelos, A. D., & Santos, W. L. (2024). Humanização digital no ensino remoto: desafios vivenciados por professores da educação básica no panorama pandêmico [Acesso em: 07 jun. 2024]. *Revista Iberoamericana de Educación*, 94(1), 17–36. <https://rieoei.org/RIE/article/view/6072/4810>
- Andrade, M. M. (2008). *Ensino e aprendizagem de estatística por meio da modelagem matemática: uma investigação com o ensino médio* [Dissertação (Mestrado)]. Universidade Estadual Paulista, Instituto de Geociências e Ciências Exatas [193 f.].
- Barbosa, J. C. (2007). Mathematical modelling and parallel discussions. *Proceedings of the 5th Congress of the European Society for Research in Mathematics Education (CERME)*, 1, 1–10.
- Bassanezi, R. C. (2002). *Ensino-Aprendizagem com Modelagem Matemática: uma nova estratégia*. Contexto.
- Borba, M. C., Malheiros, A. P. S., & Amaral, R. B. (2011). *Educação a Distância on-line* (3^a ed.). Autêntica.
- Borba, M. C., & Penteado, M. G. (2007). *Informática e Educação Matemática* (3^a ed.). Autêntica.
- Brasil. Conselho Nacional de Educação. (2020). Resolução CNE/CP nº 2, de 20 de dezembro de 2019. Define as diretrizes curriculares nacionais para a formação inicial de professores para a educação básica e institui a Base Nacional Comum para a Formação Inicial de Professores da Educação Básica (BNC-Formação).
- Brasil. Ministério da Educação. (2017). Base Nacional Comum Curricular [Acesso em: 20 jun. 2024].
- Cambraia, A. C., Araújo, M. C. P. d., & Biondo, U. L. R. (2022). Conhecimento Didático do Conteúdo na Formação de Professores de Computação. *Revista Brasileira de Informática na Educação*, 30, 449–470. <https://doi.org/10.5753/rbie.2022.2525>
- Campos, C. R., Wodewotzki, M. L. L., & Jacobini, O. R. (2021). *Educação Estatística: teoria e prática em ambientes de modelagem matemática* (3^a ed.). Autêntica.
- Chance, B. L. (1998). Experiences with Authentic Assessment Techniques in a Introductory Statistics Course [Acesso em: 07 jun. 2024]. *Journal of Statistics Education*, 5(3). <https://jse.amstat.org/v5n3/chance.html>
- Conselho Nacional de Saúde. (2012, dezembro). Resolução CNS nº 466/2012: Diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos [Publicada no Diário Oficial da União em 13 de junho de 2013]. Recuperado janeiro 23, 2025, de <http://conselho.saude.gov.br/resolucoes/2012/Reso466.pdf>
- Davenport, T. H. (2014). *Big Data no trabalho: Derrubando mitos e descobrindo oportunidades*. Elsevier.
- de Faria Gomes, M. A., Viera, H. L., & Luna, R. A. (2017). Tecnologias Móveis Digitais da Comunicação e Informação: impactos nas práticas docentes e discentes [Acesso em: 07 jun. 2024]. *Informática na Educação: Teoria e Prática*, 20(2). <https://core.ac.uk/download/pdf/303959208.pdf>
- Delmas, R. C. (2002). Statistical Literacy, Reasoning, and Thinking: A Commentary [Acesso em: 07 jun. 2024]. *Journal of Statistics Education*, 10(3). https://jse.amstat.org/v10n3/delmas_discussion.html
- Draper, N. R., & Smith, H. (1998). *Applied Regression Analysis* (3rd). Wiley.
- Estrada, A. A. (2009). Os fundamentos da teoria da complexidade em Edgar Morin. *Akrópolis Umuarama*, 17(2), 85–90.

- Ferrete, A. A. S. S. (2007). *O computador no ensino de língua inglesa no CEFET/RN* [Tese (Doutorado em Educação)]. Universidade Federal do Rio Grande do Norte [159 f.].
- Galvani, P. (2002). A autoformação, uma perspectiva transpessoal, transdisciplinar e transcultural. Em C. et al. (Ed.), *Educação e Transdisciplinaridade II*. TRIOM.
- Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning*. Springer.
- Garfield, J., & Gal, I. (1999). Teaching and assessing statistical reasoning. Em L. Staff (Ed.), *Developing Mathematical Reasoning in Grades K-12* (pp. 207–219). National Council of Teachers of Mathematics.
- Haack, D. G. (1979). *Statistical Literacy: A Guide to Interpretation*. Duxbury Press.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models* (5th). McGraw-Hill.
- Ministério da Saúde. (2024a). 26/6 Dia Nacional do Diabetes [Acessado em: 09 out. 2024].
- Ministério da Saúde. (2024b). Diabetes [Acessado em: 09 out. 2024]. <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/d/diabetes>
- Morin, E. (2000). *Ciência com Consciência*. Bertrand.
- Perrenoud, P. (2000). Construindo Competências [Entrevista com Paola Gentle e Roberta Bencini]. *Nova Escola*, 19–31.
- Ponte, J. P. d. (2002). A vertente profissional da formação inicial de professores de Matemática. *Educação Matemática em Revista. Revista da Sociedade Brasileira de Educação Matemática*, (11a), 3–8.
- Rumsey, D. J. (2002). Statistical Literacy as a Goal for Introductory Statistic Courses [Acesso em: 07 jun. 2024]. *Journal of Statistics Education*, 10(3). <https://jse.amstat.org/v10n3/rumsey2.html#chance1997>
- Scheaffer, P. (1990). The ASA-NCTM Quantitative Literacy Project: an Overview. Em D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on the Teaching of Statistics (ICOTS-3)* (pp. 45–49, Vol. 1).
- Sedlmeier, P. (1999). *Improving Statistical Reasoning: Theoretical Models and Practical Implications*. Lawrence Erlbaum.
- Serviços e Informações do Brasil. (2024). Serviços e Informações do Brasil [Acessado em: 09 out. 2024].
- Skovsmose, O. (2000). Cenários para Investigação. *BOLEMA – Boletim de Educação Matemática*, 13(14), 66–91.
- Smith, G. (1998). Learning Statistics by doing Statistics [Acesso em: 07 jun. 2024]. *Journal of Statistics Education*, 6(3). https://jwilson.coe.uga.edu/EMAT6680Fa11/Shbaughnessy/LS_STAT6070/Portfolio%20Reflections/Reflection%206%20Learning%20Statistics%20by%20Doing%20Statistics.html
- UNESCO. (2023). A tecnologia na educação: os jovens clamam por equidade e sustentabilidade [Acesso em: 09 out. 2024].
- Vale Mais Saúde. (2024). Principais Tipos de Diabetes [Acessado em: 09 out. 2024]. <https://www.valemaissaude.com.br/materia/principais-tipos-de-diabetes>
- Watson, J. (1997). Assessing Statistical Thinking Using the Media. Em I. Gal & J. B. Garfield (Ed.), *The Assessment Challenge in Statistics Education* (pp. 107–121). IOS Press, on behalf of the ISI.

Apêndice 1

1. <https://drive.google.com/file/d/18Rt-a3OO6rg64C-sWZLtxusaIPWETpqv/view?usp=sharing>
Arquivo .csv com os dados utilizados