

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ - UTFPR

DAVI LIMA IACOMINI

**INTRODUÇÃO A MÉTODOS MATEMÁTICOS DE RANKING E
RATING**

CURITIBA

2025

DAVI LIMA IACOMINI

**INTRODUÇÃO A MÉTODOS MATEMÁTICOS DE RANKING E
RATING**

Introduction to mathematical ranking and rating methods

Dissertação apresentada como requisito para obtenção do título de Mestre em Matemática, área de concentração: Matemática na Educação Básica, no Programa Mestrado Profissional em Matemática em Rede Nacional - PROFMAT da Universidade Tecnológica Federal do Paraná - UTFPR.

Linha de pesquisa: Matemática na Educação Básica e suas Tecnologias.

Orientador: Prof. Dr. Andrés David Báez Sánchez.

CURITIBA

2025



[4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Esta licença permite remixe, adaptação e criação a partir do trabalho, para fins não comerciais, desde que sejam atribuídos créditos ao(s) autor(es) e que licenciem as novas criações sob termos idênticos. Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.



Ministério da Educação
Universidade Tecnológica Federal do Paraná
Campus Curitiba



DAVI LIMA IACOMINI

INTRODUÇÃO A MÉTODOS MATEMÁTICOS DE RANKING E RATING

Trabalho de pesquisa de mestrado apresentado como requisito para obtenção do título de Mestre Em Matemática da Universidade Tecnológica Federal do Paraná (UTFPR). Área de concentração: Matemática.

Data de aprovação: 02 de Outubro de 2025

Dr. Andres David Baez Sanchez, Doutorado - Universidade Tecnológica Federal do Paraná

Dr. Moiseis Dos Santos Ceconello, Doutorado - Universidade Federal de Mato Grosso (Ufmt)

Dr. Roy Wilhelm Probst, Doutorado - Universidade Tecnológica Federal do Paraná

Documento gerado pelo Sistema Acadêmico da UTFPR a partir dos dados da Ata de Defesa em 02/10/2025.

AGRADECIMENTOS

A Deus, que me capacitou e guiou no desenvolvimento deste trabalho, me dando muito mais do que mereço.

À minha esposa, Mariana, por me apoiar e incentivar nos dias em que tive que priorizar este curso. A meus pais, Franco e Marli, e meu irmão, Luca, por me mostrarem o caminho do estudo e da pesquisa.

Ao Prof. Dr. Andrés David Báez Sánchez, que aceitou me orientar semanalmente durante os últimos meses, e o fez de forma exemplar.

À Sociedade Brasileira de Matemática que, na busca da melhoria do ensino de matemática na Educação Básica, viabilizou a implementação do PROFMAT.

À CAPES, pela recomendação do PROFMAT por meio do parecer do Conselho Técnico Científico da Educação Superior.

RESUMO

A presente dissertação busca introduzir alguns aspectos matemáticos usados para a construção de *ratings* e *rankings*. Ranquear itens consiste em realizar uma ordenação, segundo determinados princípios. Esse processo de organização pode ser encontrado em muitos contextos, como em índices socioeconômicos, competições esportivas, rankings educacionais, etc. Neste trabalho, são considerados fundamentos matemáticos presentes no Índice de Gini e no Índice de Desenvolvimento Humano (IDH), e é explorada, por meio de regressões lineares, a correlação entre esses indicadores e o resultado brasileiro da avaliação do Programa Internacional de Avaliação de Alunos (PISA). Além disso, foram caracterizados os métodos de ranqueamento de Massey e Elo, muito presentes no contexto esportivo. Para tal fim, foi necessário um aprofundamento em tópicos de álgebra linear e funções, dentre os quais destacam-se a resolução de sistemas lineares sobredeterminados a partir da equação normal, da reparametrização e da pseudoinversa, e a função logística. Como produto educacional, foi produzida uma série de vídeos sobre alguns dos assuntos considerados, de forma alinhada com as orientações da BNCC.

Palavras-chave: ranqueamento; índices socioeconômicos; avaliação pisa; sistemas lineares sobredeterminados; método de elo.

ABSTRACT

This dissertation aims to introduce some mathematical aspects used to construct *ratings* and *rankings*. Ranking consists of ordering items according to certain principles. This process can be found in many contexts, such as socioeconomic index, sports competitions, educational rankings, etc. This work considers the mathematical foundations present in the Gini Index and in the Human Development Index (HDI), and explores, through linear regressions, the correlation between these indicators and the Brazilian results for the Program for International Student Assessment (PISA). Furthermore, the Massey and Elo ranking methods, often used in sports, were characterized. This required a more in-depth study of certain topics in linear algebra and functions, such as: overdetermined linear systems resolution using normal equations, reparameterization and pseudoinverse, and the logistic function. As an educational product, a series of videos was produced on some of the topics considered, in line with the BNCC guidelines.

Keywords: ranking; socioeconomic index; pisa test; overdetermined linear systems; elo's method.

LISTA DE ILUSTRAÇÕES

Figura 1 – Comunicação do estado do Paraná a respeito da nota do IDEB	10
Figura 2 – Representação visual do Exemplo 3.1	17
Figura 3 – A projeção do vetor b no espaço-coluna	24
Figura 4 – A curva de Lorenz em estratos discretos	37
Figura 5 – As regiões trapezoidais determinadas na poligonal de Lorenz	40
Figura 6 – Evolução do Índice de Gini brasileiro desde 2011	42
Figura 7 – Evolução do IDH brasileiro desde 1990.....	45
Figura 8 – Evolução da pontuação do Brasil no PISA.....	48
Figura 9 – (a) Relação entre IDH e pontuação PISA; (b) relação entre Índice de Gini e pontuação PISA.....	50
Figura 10 – (a) Relação entre IDH e pontuação PISA; (b) relação entre Índice de Gini e pontuação PISA.....	51
Figura 11 – O gráfico de uma função logística.....	78
Figura 12 – Função logística na população do Japão.....	79
Figura 13 – A superfície de pontuação esperada em função dos ratings atuais	82
Figura 14 – Confronto entre Carlsen e Nakamura; (a) até o recorde histórico de 2882 pontos; (b) até uma meta de 3000 pontos	90
Figura 15 – Confronto entre Carlsen e a média dos 10 melhores ratings: (a) até o recorde histórico de 2882 pontos; (b) até uma meta de 3000 pontos	91
Figura 16 – A evolução do rating de Magnus Carlsen.....	91
Figura 17 – O vídeo disponível no <i>youtube</i>	92
Figura 18 – O <i>site</i> do <i>software</i> Manim.....	95
Figura 19 – O <i>software</i> Microsoft Visual Studio.....	96
Figura 20 – Modificação de código com IA.....	98
Figura 21 – O <i>site</i> do <i>software</i> DaVinci Resolve	98
Figura 22 – A tela de edição do DaVinci Resolve	99
Figura 23 – Perfil dos espectadores dos vídeos	100
Figura 24 – Conhecimento prévio dos espectadores	101
Figura 25 – Avaliação da série de vídeos.....	101
Figura 26 – Retenção no vídeo a respeito do método de Elo	102

LISTA DE TABELAS

Tabela 1 – Distribuição das pessoas ocupadas conforme a renda obtida na atividade exercida, no Brasil, de acordo com a Pnad de 2003	36
Tabela 2 – Distribuição de renda ou consumo na Islândia	41
Tabela 3 – Ranking mundial do Índice de Gini	41
Tabela 4 – Parâmetros para normalização de índices do IDH	43
Tabela 5 – Exemplo do cálculo do IDH - exemplo do Brasil	43
Tabela 6 – Exemplo do cálculo do IDH - exemplo da Islândia	44
Tabela 7 – Ranking mundial do IDH	45
Tabela 8 – Ranking mundial do PISA	48
Tabela 9 – Olimpíadas de Verão Paris 2024: resultados do grupo B do handebol feminino.....	55
Tabela 10 – Rating encontrado para cada seleção a partir da inversa G_1	59
Tabela 11 – Rating encontrado para cada seleção a partir da inversa G_2	59
Tabela 12 – Rating encontrado para cada seleção a partir da pseudoinversa	60
Tabela 13 – Rating encontrado para cada seleção usando o método de Massey	61
Tabela 14 – Rating encontrado para todas as seleções das Olimpíadas	71
Tabela 15 – Ranking da ATP em 01/12/2024	72
Tabela 16 – Rating encontrado para os jogadores da ATP	73
Tabela 17 – Rating para os jogadores da ATP com jogos repetidos condensados.....	74
Tabela 18 – População do Japão.....	79
Tabela 19 – Valores de K no sistema de classificação de xadrez	85
Tabela 20 – Valores de K no ranking da FIFA	85
Tabela 21 – Duração média das visualizações	102

SUMÁRIO

1	INTRODUÇÃO	9
1.1	A matemática para ratings e rankings	9
1.2	Justificativa para a escolha do tema	11
1.3	Revisão da bibliografia	11
1.4	Objetivos	13
1.4.1	Objetivo geral	13
1.4.2	Objetivos específicos	13
1.5	Procedimentos metodológicos	13
1.6	Recurso educacional	14
1.7	Estrutura do trabalho	14
2	PRELIMINARES: SOBRE A SOLUÇÃO DE SISTEMAS LINEARES...	16
2.1	Inversas generalizadas	16
2.2	A inversa de Moore-Penrose	20
2.3	Sistemas lineares sobredeterminados: solução de mínimo erro quadrático	23
3	ÍNDICE DE GINI, ÍNDICE DE DESENVOLVIMENTO HUMANO E RANKING DO PISA	35
3.1	Coefficiente de Gini	35
3.2	Índice de Desenvolvimento Humano	42
3.3	Ranking do PISA	45
3.3.1	O formato da avaliação	46
3.3.2	Críticas acumuladas	48
3.4	Regressão linear entre alguns dos índices	49
3.4.1	Coefficiente de correlação de Pearson nas regressões	52
4	MÉTODO DE MASSEY PARA RANKEAMENTO	55
4.1	O método no caso todos contra todos	55
4.1.1	O ranking gerado a partir do método de Massey	66
4.2	Considerações sobre a matriz X com número de jogos diferentes	68
4.3	Comentários adicionais	74
5	O MÉTODO DE ELO	77
5.1	Função logística	77
5.2	A teoria do método de ranqueamento de Elo	79
5.3	Considerações sobre o coeficiente de pontuação esperada μ_{ij}	82
5.4	Considerações sobre as constantes K e ξ	84
5.5	Análise de confrontos e exemplos	87
6	PRODUTO EDUCACIONAL	92
6.1	Justificativa e Objetivos	92
6.2	Etapas de criação	93
6.3	Resultado	94
6.4	Softwares utilizados	95
6.4.1	Manim	95
6.4.2	Inteligências artificiais	96
6.4.3	Editores de vídeo	97
6.5	Divulgação e validação do produto	99
7	CONCLUSÃO	103
	REFERÊNCIAS	106

1 INTRODUÇÃO

O ranqueamento de itens tem como princípio organizá-los em ordem de importância de acordo com parâmetros pré-estabelecidos. Esse processo de organização é utilizado em muitos contextos, dos mais cotidianos aos mais históricos.

Uma importante aplicação de modelos de ranqueamento é vista em índices socioeconômicos, usados para medir qualidade de vida, desigualdade, qualidade de educação, entre outros. Como a matemática é usada para quantificar e ranquear esses aspectos da vida humana?

1.1 A matemática para ratings e rankings

Como exemplo de um índice amplamente usado no mundo todo, temos o Índice de Desenvolvimento Humano (IDH). Em 2022, o Brasil alcançou IDH igual a 0,760 figurando na 89ª posição entre 193 países (Sampaio, 2024). Como este número foi calculado? Que critérios são pensados para se estabelecer uma nota para o desenvolvimento de um país?

De acordo com o Relatório de Desenvolvimento Humano 2021-22, publicado pelo Programa das Nações Unidas para o Desenvolvimento (PNUD, 2023), o cálculo do IDH é realizado por meio da seguinte média geométrica:

$$IDH = (I_{SAUDE} \cdot I_{EDUCACAO} \cdot I_{RENDA})^{1/3}$$

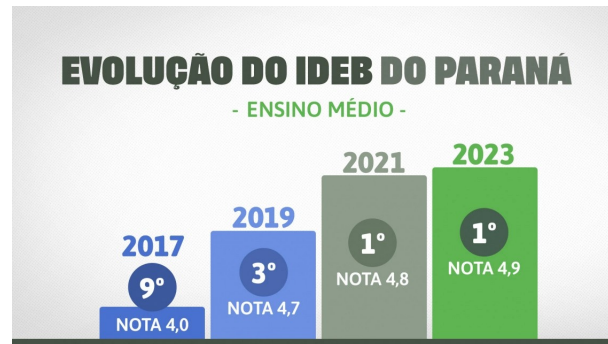
em que I_{SAUDE} é o índice de vida saudável, $I_{EDUCACAO}$ é o índice de acesso à educação e cultura e I_{RENDA} é o índice de padrão de renda adequada de vida. Cada um desses índices é construído a partir de pesquisas de órgãos estatísticos internacionais e, posteriormente, normalizados para que assumam valores de 0 a 1.

Outros índices podem ser encontrados em outros contextos. No ramo educacional, é comum serem definidos índices para um colégio ou para um sistema de ensino, de forma a melhor comparar tais organismos ou registrar o seu desenvolvimento ao longo do tempo. Mundialmente, o padrão para avaliação de sistemas educacionais no mundo é o *Programme for International Student Assessment* (Programa Internacional de Avaliação de Estudantes - PISA). Nos resultados de 2022, o Brasil esteve em 65º lugar, com 379 pontos. A média dos países da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) foi de 472 pontos (OCDE, 2023a). O cálculo da nota do PISA depende de recursos estatísticos avançados, como Teoria de Resposta ao Item, modelos logísticos e modelos de crédito parcial generalizado.

Já em um contexto nacional, o estado do Paraná alcançou nota 4,9 no Índice de Desenvolvimento da Educação Básica (IDEB) de 2023, nota que coloca o estado no primeiro lugar do Brasil (Paraná, 2024). Tal conquista foi amplamente divulgada em canais oficiais do governo e revela uma evolução do índice do estado nas últimas avaliações (Figura 1).

O cálculo do índice do IDEB também é resultado da união de outros índices, aqui ligados a uma prova chamada Sistema de Avaliação da Educação Básica (SAEB), que cobra conteúdos

Figura 1 – Comunicação do estado do Paraná a respeito da nota do IDEB.



Fonte: Paraná (2024).

de português e matemática. Em cada ano em que a prova é realizada, faz-se uma normalização da nota corrente com os mínimos e máximos da primeira prova, aplicada em 1997. A média das notas normalizadas é multiplicada por outro fator, a média harmônica das taxas de aprovação em cada série sendo contabilizada, resultando na nota final do IDEB (INEP, 2023).

Algumas instituições internacionais elaboram rankings educacionais de universidades, como é o caso da avaliação realizada pela QS Quacquarelli Symonds. Na edição de 2024 deste ranking, a Universidade de São Paulo (USP) foi eleita a melhor universidade da América Latina e Caribe (QS World University rankings, 2024).

No mesmo ranking, a Universidade Tecnológica Federal do Paraná (UTFPR) ficou na 104ª posição (QS World University rankings, 2024). Este ranking é construído levando em consideração a reputação acadêmica e publicações, empregabilidade, internacionalização, professores com doutorado e razão de professores para estudantes. Apesar de seu grande prestígio, o ranking não disponibiliza publicamente os métodos de cálculo que geram seu ranking, para preservar seu produto.

Atribuir uma nota para as universidades, e assim definir qual é a melhor, é uma boa ideia. Porém, esta premissa carrega vários desafios. Se o rating não for bem feito e matematicamente consistente, a divulgação destes resultados pode levar universidades medianas aos palcos, ou prejudicar a reputação de instituições que, na realidade, mereciam maior destaque.

Este dilema surge ainda diversas vezes no contexto digital atual. Frente a tantas informações disponíveis na internet, como determinar as que são mais relevantes para um usuário? A história mostra que a matemática possui as respostas para este questionamento. Com um bom ranqueamento, é possível afirmar com maior confiança qual resultado é mais interessante para um usuário que faz uma busca no mecanismo de pesquisa. Foi com uma boa fundamentação matemática, por exemplo, que foi criado o mecanismo de busca do *Google*.

Dois estudantes de pós-graduação pouco conhecidos utilizaram Cadeias de Markov para aprimorar ranqueamentos de mecanismos de busca e o método deles foi tão bem-sucedido que mais tarde se tornou o fundamento da recém-criada empresa, que mais tarde se tornaria o extremamente conhecido mecanismo de busca *Google* (Langville; Meyer, 2012, p. 1).

Portanto, a relevância do ranqueamento não se limita à educação e aos índices governamentais. Sugestões de músicas em aplicativos, rankings de times e seleções, até recomendações de filmes por plataformas com base em visualizações recentes são frutos do ranqueamento de itens. Por trás disso estão conceitos de matemática, como álgebra linear e teorias de funções, entre outros.

1.2 Justificativa para a escolha do tema

Pelo anterior, considera-se que o contexto de problemas de ranqueamento pode ser adaptado a um nível da matemática do ensino médio. A Base Nacional Comum Curricular (BNCC) afirma, na seção a respeito da área de matemática e suas tecnologias:

[...] no Ensino Médio o foco é a construção de uma visão integrada da Matemática, aplicada à realidade, em diferentes contextos. Consequentemente, quando a realidade é a referência, é preciso levar em conta as vivências cotidianas dos estudantes do Ensino Médio – impactados de diferentes maneiras pelos avanços tecnológicos, pelas exigências do mercado de trabalho, pelos projetos de bem viver dos seus povos, pela potencialidade das mídias sociais, entre outros. Nesse contexto, destaca-se ainda a importância do recurso a tecnologias digitais e aplicativos tanto para a investigação matemática como para dar continuidade ao desenvolvimento do pensamento computacional, iniciado na etapa anterior (Brasil, 2018, p. 528).

Considera-se que a classificação de itens faz parte da vivência cotidiana dos estudantes, principalmente em plataformas e sistemas digitais. O esporte também é elemento motivador e contexto próprio da realidade dos estudantes. A isto acrescenta-se a seguinte habilidade da Base:

(EM13MAT104) Interpretar taxas e índices de natureza socioeconômica (índice de desenvolvimento humano, taxas de inflação, entre outros), investigando os processos de cálculo desses números, para analisar criticamente a realidade e produzir argumentos (Brasil, 2018, p. 533).

Portanto, um estudo da matemática presente na formação de índices socioeconômicos nacionais e mundiais, acrescido de uma abordagem introdutória de métodos matemáticos de ranqueamento de itens, no contexto da álgebra linear, da solução de sistemas de equações e de funções, constitui um bom elemento motivador para os estudantes, embasado com a habilidade EM13MAT104 da BNCC.

1.3 Revisão da bibliografia

Neste trabalho, utilizam-se extensivamente dois livros principais para a caracterização e explicação dos principais sistemas de ranqueamento em esportes. Langville e Meyer foram a

principal referência para o reconhecimento de diferentes métodos de rating e ranking. Em seu livro, dissertam sobre sistemas como os de Massey, Colley, Keener, Elo e Markov, além dos avanços que tais métodos sofreram desde suas concepções iniciais (Langville, Meyer, 2012). Apesar do livro ser boa referência para boas introduções a cada sistema de ranqueamento, ele não explora com tanto rigor as justificativas para as escolhas feitas por seus criadores. Utiliza-se, então, os trabalhos próprios dos criadores dos métodos estudados, como o trabalho de monografia de Massey (1997).

É a partir de Langville e Meyer que se definem os termos que serão muito utilizados ao longo deste trabalho: rating e ranking. De acordo com os autores, “um ranking refere-se a uma lista ordenada de equipes, enquanto rating refere-se a uma lista de pontuações numéricas, uma para cada equipe” (Langville e Meyer, 2012, p. 5). Ou seja, uma lista de ratings, ao ser organizada de forma apropriada, dá origem a um ranking.

Para compreender decisões realizadas na construção dos métodos explorados neste trabalho, foi necessário um aprofundamento em conceitos de álgebra linear, em particular noções de resolução de sistemas sobredeterminados, inversas generalizadas e inversa de Moore-Penrose. Para isto, o trabalho baseia-se em Strang (2013). Seu livro traz a base necessária para compreender o que está por trás dos métodos explorados neste trabalho. Outras referências secundárias utilizadas para aprofundamento de Álgebra Linear foram Anton e Busby (2007) e Myers e Milton (1991).

Para complementar a pesquisa, busca-se também encontrar o que já foi abordado em dissertações do PROFMAT acerca deste mesmo assunto. No repositório presente no site oficial do programa, foram procuradas correspondências para as principais palavras-chave do presente trabalho. Em primeiro lugar, buscou-se o termo “rating”, que apresentou 0 registros. Em segundo lugar, buscou-se o termo “rank” e encontraram-se 3 registros, dos quais 2 estão de acordo com o escopo deste trabalho.

A dissertação de Batti (2015) conta rapidamente a história da criação do algoritmo Page-Rank, pilar fundamental da empresa Google, que efetua o ranqueamento de páginas de internet e, com isso, organiza os resultados de uma pesquisa no mecanismo de busca. O autor apresenta primeiro informalmente o funcionamento do algoritmo, construindo gradativamente a lógica formal nos ranqueamentos, com o diferencial da utilização de conceitos de grafos. Por fim, propõe-se um roteiro para o desenvolvimento destas ideias, no nível de dificuldade e abstração apropriado, em uma aula de Ensino Médio.

Já a dissertação de Pereira (2019) dá maior destaque para as matrizes de Markov, ou matrizes de transição de probabilidades, tema de estudo de diversas outras dissertações no programa. Além de Markov, também aborda matrizes irredutíveis, e o teorema de Perron-Frobenius. Ainda, em seu último capítulo, apresenta aplicações dos conceitos desenvolvidos em questões de olimpíadas de matemática, no algoritmo PageRank e no modelo de difusão de Ehrenfest, este último com consequência mais relevantes no campo da física, e assim apresentado de forma mais introdutória.

Outra pesquisa foi realizada no repositório com o termo “gini”, que apresentou um resultado. A dissertação de Costa (2022) explora reflexões econômicas no contexto da disciplina de Educação Financeira. O autor fundamenta a adição da disciplina no currículo escolar nos últimos anos e, posteriormente, analisa matematicamente índices utilizados para fundamentar políticas públicas, especificamente o IDH e o índice de Gini. Este conhecimento é então transposto para 10 atividades didáticas, com orientações para o professor, voltadas para estudantes do 6º ano do Ensino Fundamental.

Buscou-se então pelo termo “elo”, na tentativa de encontrar trabalhos que tratem do sistema de ranqueamento de jogadores de xadrez. O sistema retornou 172 resultados, mas todos estão em palavras como “modelo” ou “pelo”, e nenhum trata do tema deste trabalho.

1.4 Objetivos

1.4.1 Objetivo geral

Explorar fundamentos matemáticos presentes em índices socioeconômicos e contextos esportivos, através de um aprofundamento em tópicos de álgebra linear e funções.

1.4.2 Objetivos específicos

1. Conhecer diferentes índices socioeconômicos, seus fundamentos matemáticos, e a possível relação entre diferentes índices.
2. Caracterizar os métodos de ranqueamento de Massey e Elo, demonstrando suas escolhas metodológicas e as consequências dessas escolhas, a partir de conhecimentos de solução de sistemas lineares indeterminados e de funções logísticas.
3. Elaborar recurso educacional com aplicações da álgebra linear e funções no contexto de problemas de ranqueamento que estimulem um estudante na etapa do ensino médio.

1.5 Procedimentos metodológicos

O presente trabalho é resultado de pesquisa quantitativa e exploratória. Tais termos são definidos segundo as seguintes referências.

A pesquisa quantitativa se centra na objetividade. Influenciada pelo positivismo, considera que a realidade só pode ser compreendida com base na análise de dados brutos, recolhidos com o auxílio de instrumentos padronizados e neutros. A pesquisa quantitativa recorre à linguagem matemática para descrever as causas de um fenômeno, as relações entre variáveis, etc. (Fonseca, 2002, p. 20).

O objeto de pesquisa é reconhecido como fundamentado em critérios objetivos e com auxílio de métodos padronizados e baseados em linguagem matemática. Portanto, a pesquisa é qualitativa.

Estas pesquisas [exploratórias] têm como objetivo proporcionar maior familiaridade com o problema, com vistas a torná-lo mais explícito ou a construir hipóteses. Pode-se dizer que estas pesquisas têm como objetivo principal o aprimoramento de ideias ou a descoberta de intuições. Seu planejamento é, portanto, bastante flexível, de modo que possibilite a consideração dos mais variados aspectos relativos ao fato estudado (Gil, 2002, p. 41).

Parte do propósito desta pesquisa é analisar exemplos concretos e proporcionar maior familiaridade com problemas de ranqueamento, explicitando suas diferentes formas de resolução. Na busca de melhor cumprir este propósito exploratório, a pesquisa bibliográfica auxiliou na tentativa de captar as diversas relações, já analisadas e fundamentadas em trabalhos anteriores, presentes entre os conceitos trabalhados nos métodos de ranqueamento. Adotam-se os seguintes procedimentos metodológicos:

1. pesquisa bibliográfica em documentos oficiais de órgãos nacionais e internacionais e explicitação das metodologias de cálculo de índices socioeconômicos;
2. pesquisa bibliográfica sobre conceitos centrais da solução de sistemas lineares e exploração por meio de diversos exemplos;
3. descrição detalhada de diferentes métodos de ranqueamento;
4. desenvolvimento de produto educacional voltado para professores e estudantes de ensino médio.

1.6 Recurso educacional

Como produto deste trabalho de conclusão de curso, será produzida uma série de vídeos, publicados no site *youtube*. Serão quatro vídeos, cada um focado em um aspecto desta dissertação: um para o Coeficiente de Gini; um para o IDH; um para a nota do PISA e sua relação com os dois índices anteriores; e um para o método de ranqueamento de Elo. O público-alvo é professores de matemática, professores de educação financeira e estudantes que queiram se aprofundar no conteúdo. O objetivo é popularizar e democratizar o conhecimento explorado neste trabalho.

1.7 Estrutura do trabalho

O presente trabalho está estruturado da seguinte forma:

- No Capítulo 2 serão apresentados embasamentos teóricos a respeito das inversas generalizadas, e especificamente da pseudoinversa de uma matriz, na tentativa de encontrar uma solução ou uma solução aproximada de um sistema que envolve uma matriz singular. A partir destas bases, se buscarão métodos para a solução aproximada, com erro quadrático mínimo, de sistemas sobredeterminados.
- No Capítulo 3 serão explorados os métodos de ranqueamento de alguns índices socioeconômicos (Gini e IDH) e será apresentada uma breve introdução à forma de obtenção da nota do PISA. Tais resultados também serão relacionados na busca de uma correlação entre índices.
- No Capítulo 4 caracteriza-se o método de Massey para o ranqueamento de equipes num contexto esportivo, e serão explicadas suas decisões no método a partir da teoria dos capítulos anteriores.
- No Capítulo 5 caracteriza-se o método de Elo para ranqueamento de jogadores no contexto do xadrez, a partir da definição da função logística. São realizadas, então, diversas explorações por meio de análise de confrontos.
- No Capítulo 6 apresenta-se o recurso educacional, com suas justificativas e objetivos específicos.
- No Capítulo 7 apresentam-se as conclusões do presente trabalho.

2 PRELIMINARES: SOBRE A SOLUÇÃO DE SISTEMAS LINEARES

Antes de explorar os métodos matemáticos presentes em índices socioeconômicos e em ranqueamentos de times, será necessário lembrar alguns conceitos de álgebra linear. O método de Massey, por exemplo, depende de fundamentos de resolução de sistemas lineares, e exercícios de correlação entre índices utilizarão aproximações por mínimos quadrados. Portanto, o presente capítulo trata de resultados essenciais para uma boa compreensão dos capítulos seguintes.

Serão utilizadas, neste capítulo, definições segundo Strang (2013, p. viii), em particular os conceitos de espaço-coluna, definido como todas as combinações lineares das colunas de uma matriz, e posto, definido como o número de colunas ou linhas independentes. O número de colunas da matriz será denotado pela letra n , e o posto pela letra r .

Na solução de sistemas lineares $Ax = b$, há várias formas de se determinar x . Pode-se, por exemplo, efetuar operações elementares sobre as linhas pelo método do escalonamento. Pode-se também aplicar a Regra de Cramer, por meio dos determinantes das matrizes do sistema. Pode-se ainda, como será o foco desta seção, efetuar a multiplicação do sistema pela inversa de X . Se essa inversa existe, basta calcular $x = A^{-1}b$.

Mas nem todo sistema linear pode ser resolvido desta forma. Daremos destaque, neste trabalho, a sistemas lineares em que A não possui posto completo. Ou seja, se A é uma matriz quadrada $n \times n$ e seu posto é $r < n$, então o seu determinante é igual a 0 e, conseqüentemente, a matriz A não possui inversa.

2.1 Inversas generalizadas

A seguir é apresentado um exemplo que será explorado ao longo desta seção.

Exemplo 2.1. *Considere o sistema linear $Ax = b$:*

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 2 & 2 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

Pode-se procurar a inversa de A sabendo que $AA^{-1} = A^{-1}A = I$. Então, considerando o caso da multiplicação pela inversa à direita, temos o seguinte:

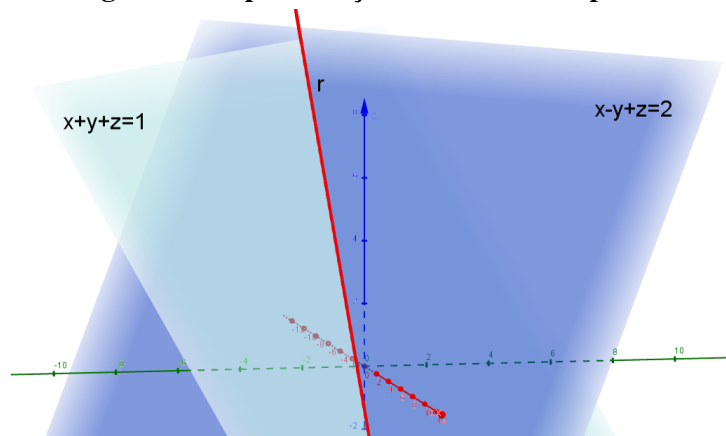
$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 2 & 2 & 2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Por meio deste sistema, deve ser verdade, por exemplo, que $a_{11} + a_{21} + a_{31} = 1$ e $2(a_{11} + a_{21} + a_{31}) = 0$, o que é impossível. Ou seja, a matriz A não possui inversa, podendo também ser chamada de singular.

De fato, a matriz A do Exemplo 2.1 não possui posto completo, uma vez que a primeira e a terceira colunas da matriz são iguais. Assim, em um sistema linear $Ax = b$ como o descrito no exemplo, não seria possível determinar um único valor do vetor x . Ele possui infinitas soluções.

É possível interpretar o sistema utilizando a geometria. Para isso, é importante destacar que a linha 3 pode ser encontrada pelo dobro da linha 1. Além disso, no vetor b , o elemento b_{31} é o dobro do elemento b_{11} . Ou seja, os planos formados pelas equações $x + y + z = 1$ e $2x + 2y + 2z = 2$ são coincidentes, de forma que a interseção destes planos coincidentes com o plano formado por $x - y + z = 2$ cria uma reta r . Assim, existem infinitos pontos desta reta r que pertencem a todos os três planos, e o sistema possui infinitas soluções. A Figura 2 traz uma representação visual deste caso.

Figura 2 – Representação visual do Exemplo 3.1



Fonte: O Autor.

Com isto, surge a pergunta: quando um sistema linear $Ax = b$ possui pelo menos uma solução? Esta pergunta pode ser respondida em termos de combinações lineares das colunas de A , como descrito no seguinte resultado.

Definição 2.1. (Strang, 2013, p. 71) *O sistema $Ax = b$ é solúvel se, e somente se, o vetor b pode ser expresso como uma combinação das colunas de A . Então, b está no espaço-coluna de A .*

Veja que no Exemplo 2.1, o vetor b pode ser escrito como combinação das colunas de A , de fato:

$$\frac{3}{2} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} + 0 \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

Outra possível combinação de colunas de A que gera o vetor b está abaixo.

$$\frac{3}{4} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} + \frac{3}{4} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

No Exemplo 2.1, a matriz A possui posto igual a 2. Assim, a imagem desta matriz é um espaço bidimensional em \mathbb{R}^3 , ou seja, um plano. Como b é um vetor de \mathbb{R}^3 , só existirá solução para o sistema se b pertencer ao plano gerado pelas colunas da matriz A . Nos dois exemplos acima, foi mostrado que o vetor b pertence a este plano, e pode ser gerado como combinação das colunas de A .

Neste trabalho, serão explorados sistemas lineares em que, frequentemente, não há uma única solução, mas sim infinitas. Uma possibilidade para determinar essas soluções está no conceito de inversas generalizadas, que será definido a seguir. Neste trabalho será utilizado como referência para este conceito o livro de Myers e Milton, com destaque para a Definição 5.3.1 e Teorema 5.3.4 (1991, p. 200-203).

Definição 2.2. *Seja A uma matriz $m \times n$. Uma matriz $G_{n \times m}$ tal que*

$$AGA = A$$

é chamada de inversa generalizada de A .

Desta Definição segue o seguinte Teorema.

Teorema 2.1. *Considere o sistema $Ax = b$. Se G é uma inversa generalizada de A , então $x = Gb$ é uma solução do sistema.*

Demonstração. Por definição,

$$AGA = A \Rightarrow AGAx = Ax$$

Considerando o sistema linear $Ax = b$, então:

$$AGb = b$$

Considerando $x_0 = Gb$, então $Ax_0 = b$, o que mostra que x_0 é uma solução do sistema. \square

Para o Exemplo 2.1, podemos encontrar a seguinte inversa generalizada.

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 2 & 2 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 2 & 2 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 2 & 2 & 2 \end{bmatrix}$$

Por definição, se G é tal que $AGA = A$, então $x_1 = Gb$ é uma solução do sistema. Neste caso:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{3}{2} \\ -\frac{1}{2} \\ 0 \end{bmatrix}$$

De fato, $x_1^T = [\frac{3}{2}, -\frac{1}{2}, 0]$ é solução do sistema original do exemplo. Ainda, outras matrizes podem ser usadas como inversas generalizadas, como mostrado abaixo.

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 2 & 2 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{20} & \frac{1}{4} & \frac{1}{10} \\ \frac{1}{10} & -\frac{1}{2} & \frac{1}{5} \\ \frac{1}{20} & \frac{1}{4} & \frac{1}{10} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 2 & 2 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 2 & 2 & 2 \end{bmatrix}$$

Assim, por definição, outra solução do sistema pode ser definida por meio da seguinte multiplicação.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \frac{1}{20} & \frac{1}{4} & \frac{1}{10} \\ \frac{1}{10} & -\frac{1}{2} & \frac{1}{5} \\ \frac{1}{20} & \frac{1}{4} & \frac{1}{10} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{3}{4} \\ -\frac{1}{2} \\ \frac{3}{4} \end{bmatrix}$$

Assim, foi encontrada outra solução, $x_2^T = [\frac{3}{4}, -\frac{1}{2}, \frac{3}{4}]$.

Nos dois cálculos acima, foram utilizadas inversas generalizadas sem a justificativa de como foram obtidas. Como é possível, portanto, determinar uma matriz G que atue como inversa generalizada? Myers e Milton (1991, p. 202) definem um algoritmo justamente para esse fim:

1. Na matriz A , tomar uma submatriz M que tenha inversa.
2. Encontrar a sua inversa, ou seja, M^{-1} .
3. Transpor o resultado encontrado, ou seja, $(M^{-1})^T$.
4. Substituir, na matriz A original, a submatriz M por $(M^{-1})^T$.
5. Substituir os demais elementos da matriz por 0.
6. Transpor a matriz resultante.

Este processo será demonstrado na matriz A do Exemplo 2.1. Tem-se que:

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 2 & 2 & 2 \end{bmatrix}$$

Portanto, uma inversa generalizada pode ser encontrada da seguinte forma.

$$M = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \longrightarrow M^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \longrightarrow (M^{-1})^T = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}$$

$$G^T = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \longrightarrow G = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Esta é exatamente a inversa generalizada utilizada para se determinar x_1 . Veja que na Definição 2.2 não há restrição sobre o tipo de matriz, sendo assim possível considerar uma inversa generalizada para matrizes que não são quadradas. Ainda, é possível encontrar diversas inversas generalizadas para uma mesma matriz. Se um sistema possui infinitas soluções, pode-se encontrar uma inversa generalizada para cada solução.

2.2 A inversa de Moore-Penrose

No Exemplo 2.1, encontramos duas inversas generalizadas para encontrar duas soluções para o sistema. Existem, porém, inversas generalizadas especiais, que são únicas e possuem propriedades diferentes sobre os seus resultados, como é o caso da inversa de Moore-Penrose, ou pseudoinversa, denotada por A^+ . Tal matriz pode ser definida da seguinte forma.

Definição 2.3. (Anton e Busby, 2007, p. 498) *Se A é uma matriz $m \times n$ com posto-coluna máximo, então*

$$A^+ = (A^T A)^{-1} A^T$$

Caso a matriz A não seja de posto completo, a pseudoinversa é definida por meio da decomposição de A em valores singulares, o que vai além do propósito deste trabalho. Alternativamente, é possível caracterizar A^+ por meio do cumprimento de algumas condições.

Definição 2.4. (Golub e Loan, 1996, p. 257) *Para a matriz A de dimensões $m \times n$, A^+ é definida como a única matriz que satisfaz as propriedades:*

i. $AA^+A = A$

$$ii. A^+AA^+ = A^+$$

$$iii. AA^+ = (AA^+)^T$$

$$iv. A^+A = (A^+A)^T$$

Sobre a Definição 2.4, destaca-se o fato de que a pseudoinversa é única, ou seja, há apenas uma matriz que satisfaz as 4 condições propostas. Ainda, há pseudoinversa para qualquer matriz A , e não há limitação para que ela seja quadrada: existe pseudoinversa, inclusive, para matrizes com $m > n$ (Anton e Busby, 2007, p. 497).

A matriz A^+ do Exemplo 2.1, calculada por meios computacionais no *software* Wolfram Alpha, resulta em

$$A^+ = \begin{bmatrix} \frac{1}{20} & \frac{1}{4} & \frac{1}{10} \\ \frac{1}{10} & -\frac{1}{2} & \frac{1}{5} \\ \frac{1}{20} & \frac{1}{4} & \frac{1}{10} \end{bmatrix}$$

que é exatamente a segunda inversa generalizada utilizada na seção anterior. Assim, a solução x_2 encontrada por meio desta inversa possui as propriedades enunciadas pelo teorema que segue.

Teorema 2.2. (Anton e Busby, 2007, p. 500) *Se A é uma matriz $m \times n$ e b é um vetor qualquer em \mathbb{R}^m , então*

$$x = A^+b$$

é a solução de mínimos quadrados de $Ax = b$ que tem norma mínima.

O termo “mínimos quadrados” será mais explorado na próxima seção, e está ligado à solução aproximada de sistemas. Se considerarmos $e = Ax - b$, ou seja, o erro da solução encontrada em relação ao sistema original, o uso da pseudoinversa possibilita minimizar o erro quadrático dado por $\|e\|$, e encontrar a solução x de norma $\|x\|$ mínima, onde $\|\cdot\|$ refere-se à norma euclidiana.

No exemplo anterior, as duas soluções encontradas não possuem erro com relação ao sistema. Porém, pode-se verificar que $\|x_1\| \approx 1,58$ e $\|x_2\| \approx 1,17$. O vetor encontrado com a pseudoinversa possui a menor norma euclidiana.

Nos cálculos baseados no Exemplo 2.1, em que b faz parte do espaço-coluna de A , apesar da matriz ser singular, há infinitas soluções. A seguir, apresenta-se um exemplo em que não há nenhuma solução exata para o sistema linear, e será necessário considerar o erro das soluções aproximadas obtidas através das inversas generalizadas.

Exemplo 2.2. *Considere o sistema linear $By = c$:*

$$\begin{bmatrix} 1 & 1 \\ 2 & 3 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix}$$

Perceba que o posto da matriz B é 2, uma vez que a terceira linha é a soma da primeira com a segunda. No entanto, no vetor c , temos que o terceiro elemento não é a soma do primeiro com o segundo. Sendo assim, não há possível solução para este sistema linear. Alternativamente, é possível observar que o vetor c não pode ser escrito como uma combinação linear das duas colunas de B .

Outra forma de perceber tal resultado é a partir da eliminação de Gauss. O escalonamento do sistema, em forma reduzida, se dá da forma:

$$\left[\begin{array}{cc|c} 1 & 1 & 2 \\ 2 & 3 & 5 \\ 3 & 4 & 8 \end{array} \right] \xrightarrow{L_2=L_2-2L_1} \left[\begin{array}{cc|c} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 3 & 4 & 8 \end{array} \right] \xrightarrow{L_3=L_3-3L_1} \left[\begin{array}{cc|c} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{array} \right] \xrightarrow{L_3=L_3-L_2} \left[\begin{array}{cc|c} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{array} \right]$$

A forma escalonada da matriz revela que $0y_1 + 0y_2 = 1$, ou seja, $0 = 1$. Torna-se claro que o sistema não possui solução. Se for necessário escolher um vetor \hat{y} para representar uma solução do sistema, seria necessário aceitar a existência de um erro. O menor erro quadrático pode ser obtido utilizando a pseudoinversa B^+ . Assim, por meios computacionais, pode-se calcular:

$$\hat{y} = B^+c = \begin{bmatrix} \frac{7}{3} & -\frac{5}{3} & \frac{2}{3} \\ -\frac{5}{3} & \frac{4}{3} & -\frac{1}{3} \end{bmatrix} \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix} = \begin{bmatrix} \frac{5}{3} \\ \frac{2}{3} \end{bmatrix}$$

O erro obtido nesta operação é dado por $e = B\hat{y} - c$. Podemos calcular este valor:

$$e = B\hat{y} - c = \begin{bmatrix} \frac{7}{3} \\ \frac{16}{3} \\ \frac{23}{3} \end{bmatrix} - \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ -\frac{1}{3} \end{bmatrix}$$

cuja norma é $\|e\| \approx 0,57$. Este erro, pelo Teorema 2.2, é o menor possível. Além disso, a norma euclidiana $\|\hat{y}\| \approx 1,79$ também é mínima.

2.3 Sistemas lineares sobredeterminados: solução de mínimo erro quadrático

Os problemas a serem explorados no próximo capítulo consistem em observações realizadas em esportes. Dispõe-se, atualmente, de grande variedade e quantidade de dados esportivos, abertos e livres, de forma que a obtenção de informação não é um problema. Porém, a sua organização em sistemas de ranqueamento pode ser desafiadora.

É justamente no conflito entre a variedade de informação e a quantidade de incógnitas que definem-se os sistemas sobredeterminados. Neles, há demasiadas informações para poucas incógnitas a serem definidas.

[...] um sistema sobredeterminado. O vetor b representa os dados de experimentos ou questionários, contendo muitos erros para se encontrar no subespaço S . Ao tentarmos escrever b como uma combinação dos vetores-base de S , percebemos que isto não é possível - as equações são inconsistentes, e não há solução para $Ax = b$. [...] Em geodesia, uma análise de mapeamento dos EUA tratou 2,5 milhões de equações com 400.000 incógnitas (Strang, 2013, p. 153).

Nos modelos que serão explorados, pode-se esperar que os dados não representem exatamente a realidade. Massey (1997) explicou bem tal fato:

em essência, estimativas de rating são escolhidas em uma tentativa de explicar os resultados de jogos que já ocorreram. Obviamente nem todo jogo pode ser justificado tanto por variações naturais na performance quanto por outras influências aleatórias. Por exemplo, se temos o caso em que A derrota B, B derrota C e C derrota A, então não existe sistema de ratings capaz de explicar os três resultados (Massey, 1997, p. 32).

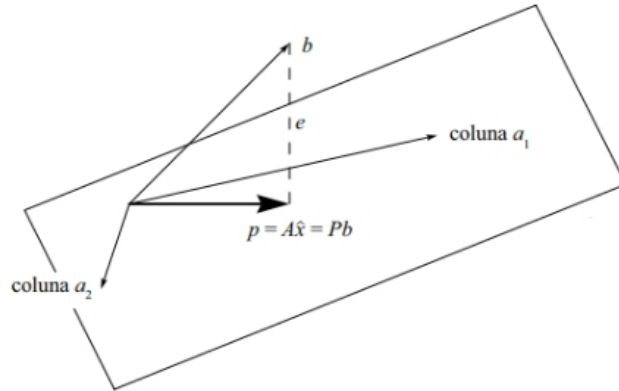
Assim, apesar da abundância de dados e recursos para manipulá-los, a solução de um sistema linear baseado em tais dados pode não existir. É preciso aceitar a presença de erros. Tal observação redireciona os esforços da solução de sistemas lineares para a minimização desses erros. Para melhor interpretar tal ideia, é interessante buscar uma visualização geométrica para solução de sistemas lineares, baseada em Strang (2013).

Pelo discutido na seção anterior, um sistema $Ax = b$ só terá solução se b faz parte do espaço-coluna de A . No caso de sistemas sobredeterminados, em que há muito mais observações do que o número de incógnitas, é provável que não exista vetor x tal que Ax resulte no vetor b . Ou ainda, o vetor b não pode ser escrito como uma combinação das colunas de A , de forma que ele não faz parte deste espaço-coluna.

Como o sistema $Ax = b$ não tem solução, uma alternativa é encontrar outro vetor, \hat{x} , tal que $A\hat{x}$ seja a melhor aproximação possível de b . O erro dessa aproximação é $e = A\hat{x} - b$. Como $A\hat{x}$ pertence ao espaço-coluna de A , o módulo de e representa a distância entre b e esse espaço-coluna. Para minimizar o erro, é preciso localizar $p = A\hat{x}$ que seja o ponto do espaço-coluna mais próximo de b .

Para determinar o ponto mais próximo, utiliza-se da projeção ortogonal, ou seja, o vetor p deve ser a projeção de b no espaço-coluna de A . E se $e = A\hat{x} - b$, então e deve ser perpendicular ao espaço-coluna de A . A Figura 3 ilustra esta situação.

Figura 3 – A projeção do vetor b no espaço-coluna



Fonte: Strang (2013, p. 162).

Nesta interpretação geométrica, os vetores a_1 e a_2 , que são colunas linearmente independentes de A , criam um espaço-coluna, e o vetor b não faz parte deste espaço. Sendo assim, o vetor $p = A\hat{x}$ é o vetor do espaço-coluna que está mais próximo possível de b (ou ainda, que torna e o menor possível). Este vetor e é a projeção ortogonal de b sobre o espaço-coluna.

O seguinte teorema mostra um caminho para determinar o vetor \hat{x} .

Teorema 2.3. Teorema fundamental da ortogonalidade (Strang, 2013, p. 144): O espaço-linha é ortogonal ao espaço nulo (em \mathbb{R}^n). O espaço-coluna é ortogonal ao espaço nulo à esquerda (em \mathbb{R}^m).

Isso mostra que o vetor de erro e deve estar no espaço nulo de A^T :

$$A^T(A\hat{x} - b) = 0$$

ou ainda

$$A^T A \hat{x} = A^T b \quad (1)$$

A Equação 1 é conhecida como Equação Normal. Uma forma mais rápida de alcançá-la é simplesmente multiplicando o sistema $Ax = b$ por A^T , à esquerda. Como vantagem do sistema novo encontrado, a matriz $A^T A$ é quadrada, simétrica e, quando as colunas de A são independentes, $A^T A$ também é inversível e, nesse caso, $\hat{x} = (A^T A)^{-1} A^T b$.

Uma solução da Equação Normal é uma solução de mínimos quadrados do sistema original (Strang, 2013, p. 162).

Os exemplos de sistemas normais abaixo são adaptados de Strang (2013) e Massey (1997).

Exemplo 2.3. Considere o seguinte sistema sobredeterminado de 4 equações e 3 incógnitas:

$$\left\{ \begin{array}{l} C + D + E = 3 \\ C + 2D + E = 5 \\ C + 3E = 6 \\ C = 0 \end{array} \right.$$

É possível reescrever este sistema de equações em termos matriciais da seguinte forma:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} C \\ D \\ E \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \\ 6 \\ 0 \end{bmatrix}$$

Este é um sistema sobredeterminado, por ter mais equações do que incógnitas. É simples perceber que a resolução deve considerar erro. Basta perceber que $C = 0$ leva à consequência de que $E = 2$, logo $D = 3/2$. Mas $C + D + E = 3$, de forma que a solução não pode ser exata.

Podemos determinar a solução de mínimos quadrados a partir da equação normal. Do lado esquerdo da equação, a matriz resultante será:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 0 & 0 \\ 1 & 1 & 3 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 4 & 3 & 5 \\ 3 & 5 & 3 \\ 5 & 3 & 11 \end{bmatrix}$$

Esta matriz possui determinante diferente de 0 e é, portanto, invertível. Do lado direito da equação tem-se que:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 0 & 0 \\ 1 & 1 & 3 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 14 \\ 13 \\ 26 \end{bmatrix}$$

Assim podemos calcular os resultados de C , D e E multiplicando a equação pela inversa da matriz obtida do lado esquerdo.

$$\begin{bmatrix} C \\ D \\ E \end{bmatrix} = \begin{bmatrix} 4 & 3 & 5 \\ 3 & 5 & 3 \\ 5 & 3 & 11 \end{bmatrix}^{-1} \begin{bmatrix} 14 \\ 13 \\ 26 \end{bmatrix}$$

$$\begin{bmatrix} C \\ D \\ E \end{bmatrix} = \begin{bmatrix} -0,12 \\ 1,46 \\ 2,02 \end{bmatrix}$$

Portanto, a solução com mínimo erro quadrático é $C = -0,12$, $D = 1,46$ e $E = 2,02$.

Exemplo 2.4. Constrói-se o seguinte sistema linear sobredeterminado $Xr = y$:

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & -1 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 7 \\ 2 \\ 1 \end{bmatrix}$$

Determinamos neste sistema a solução de mínimos quadrados por meio da Equação Normal. Do lado esquerdo da equação, multiplicamos pela transposta para encontrar a matriz 4×4 .

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ -1 & 0 & -1 & 0 & 1 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & -1 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 0 & -1 \\ -1 & 3 & -1 & -1 \\ 0 & -1 & 2 & -1 \\ -1 & -1 & -1 & 3 \end{bmatrix}$$

Já do lado direito, a multiplicação pela transposta traz como resultado o seguinte vetor:

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ -1 & 0 & -1 & 0 & 1 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 4 \\ 0 \\ 7 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ -10 \\ -1 \\ 5 \end{bmatrix}$$

O Sistema Normal obtido pode ser, então, escrito de forma reduzida como no modelo abaixo.

$$\left[\begin{array}{cccc|c} 2 & -1 & 0 & -1 & 6 \\ -1 & 3 & -1 & -1 & -10 \\ 0 & -1 & 2 & -1 & -1 \\ -1 & -1 & -1 & 3 & 5 \end{array} \right]$$

Diferentemente do que ocorreu no Exemplo 1, note que a matriz normal obtida aqui não é invertível! Isso torna-se claro a partir da eliminação de Gauss. Para isso chamaremos cada linha i da matriz de L_i , e efetuaremos as operações para tornar a matriz uma triangular superior:

$$\begin{array}{ccc} \left[\begin{array}{cccc|c} 2 & -1 & 0 & -1 & 6 \\ -1 & 3 & -1 & -1 & -10 \\ 0 & -1 & 2 & -1 & -1 \\ -1 & -1 & -1 & 3 & 5 \end{array} \right] & \xrightarrow{L_2=2L_2+L_1} & \left[\begin{array}{cccc|c} 2 & -1 & 0 & -1 & 6 \\ 0 & 5 & -2 & -3 & -14 \\ 0 & -1 & 2 & -1 & -1 \\ -1 & -1 & -1 & 3 & 5 \end{array} \right] \\ \\ \xrightarrow{L_4=2L_4+L_1} & \left[\begin{array}{cccc|c} 2 & -1 & 0 & -1 & 6 \\ 0 & 5 & -2 & -3 & -14 \\ 0 & -1 & 2 & -1 & -1 \\ 0 & -3 & -2 & 5 & 16 \end{array} \right] & \xrightarrow{L_3=5L_3+L_2} & \left[\begin{array}{cccc|c} 2 & -1 & 0 & -1 & 6 \\ 0 & 5 & -2 & -3 & -14 \\ 0 & 0 & 8 & -8 & -19 \\ 0 & -3 & -2 & 5 & 16 \end{array} \right] \\ \\ \xrightarrow{L_4=\frac{5}{3}L_4+L_2} & \left[\begin{array}{cccc|c} 2 & -1 & 0 & -1 & 6 \\ 0 & 5 & -2 & -3 & -14 \\ 0 & 0 & 8 & -8 & -19 \\ 0 & 0 & -\frac{16}{3} & \frac{16}{3} & \frac{38}{3} \end{array} \right] & \xrightarrow{L_4=\frac{3}{2}L_4+L_3} & \left[\begin{array}{cccc|c} 2 & -1 & 0 & -1 & 6 \\ 0 & 5 & -2 & -3 & -14 \\ 0 & 0 & 8 & -8 & -19 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right] \end{array}$$

Percebe-se, portanto, que a eliminação causou que toda a quarta linha assumisse valor 0. Disso decorre que a matriz $X^T X$ não é invertível, e não é possível determinar o vetor r

multiplicando a equação pela sua inversa, ou seja, por meio do cálculo de $r = (X^T X)^{-1} X^T y$. Ainda, o sistema possui infinitas soluções de quadrados mínimos. Como será possível escolher uma? Neste caso, serão utilizadas duas ferramentas para determinar a solução por mínimos quadrados.

Strang (2013) sugere que encontre-se a solução de mínimos quadrados por meio do uso da pseudoinversa. Comenta que no caso em que $Ax = b$ não tiver solução e, ainda, $A^T A$ não seja invertível, determine-se o vetor x^+ por meio de $x^+ = A^+ b$, em que A^+ é a pseudoinversa da matriz A . De acordo com o Teorema 2.2, a solução será a de mínimos quadros com norma euclidiana mínima.

Neste caso, calculamos a pseudoinversa da matriz por meio do software Wolfram Alpha, com a seguinte solução:

$$\begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & -1 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \end{bmatrix}^+ \begin{bmatrix} 4 \\ 0 \\ 7 \\ 2 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{bmatrix} = \begin{bmatrix} \frac{3}{8} & -\frac{1}{8} & 0 & \frac{3}{8} & \frac{1}{8} \\ -\frac{1}{4} & 0 & -\frac{1}{4} & 0 & \frac{1}{4} \\ -\frac{1}{8} & \frac{3}{8} & 0 & -\frac{1}{8} & -\frac{3}{8} \\ 0 & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & 0 \end{bmatrix} \begin{bmatrix} 4 \\ 0 \\ 7 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{19}{8} \\ -\frac{5}{2} \\ -\frac{9}{8} \\ \frac{5}{4} \end{bmatrix}$$

Esta multiplicação resulta na solução de mínimos quadrados do sistema, em que o vetor possui as entradas $r_1 = \frac{19}{8}$, $r_2 = -\frac{5}{2}$, $r_3 = -\frac{9}{8}$ e $r_4 = \frac{5}{4}$. Este vetor será chamado \hat{r} .

De acordo com Myers e Milton (1991), a matriz $X^T X$ só será invertível se X possuir posto completo. Note que, em qualquer linha de X encontram-se um elemento igual a 1, um elemento igual a -1 e o resto dos elementos iguais a 0 e, mais ainda, a soma das linhas de X anula-se. Dessa forma, as linhas de X são linearmente dependentes, e a equação normal possui infinitas soluções.

Desta forma, para determinar uma solução única para o sistema, pode-se adicionar uma restrição para a solução r . Por exemplo, pode-se definir um c arbitrário, tal que $\sum_{j=1, \dots, n} r_j = c$.

Ao adaptar este artifício ao seu método de ranqueamento de equipes esportivas, Massey decide utilizar $c = 0$, o que seria equivalente a considerar uma quinta linha adicional a $X^T X$, com elementos todos iguais a 1, e uma linha em $X^T y$, com o elemento 0.

Massey (1997) observa que o vetor $v = (1,1,1,1)$ está no espaço nulo da matriz X , uma vez que a soma das linhas de $X^T X$ resulta em 0. Do anterior, também segue que há uma restrição redundante, e ao invés de considerar a restrição adicional, $r_1 + r_2 + r_3 + r_4 = 0$, qualquer uma das restrições já existentes pode ser substituída por esta restrição.

No exemplo estudado, a equação $X^T X r = X^T y$ seria:

$$\begin{bmatrix} 2 & -1 & 0 & -1 \\ -1 & 3 & -1 & -1 \\ 0 & -1 & 2 & -1 \\ -1 & -1 & -1 & 3 \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{bmatrix} = \begin{bmatrix} 6 \\ -10 \\ -1 \\ 5 \end{bmatrix}$$

Seguindo a sugestão de Massey, é possível substituir todos os elementos da última linha de $X^T X$ por 1 e o último elemento de $X^T y$ por 0. Assim, teremos:

$$\begin{bmatrix} 2 & -1 & 0 & -1 \\ -1 & 3 & -1 & -1 \\ 0 & -1 & 2 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{bmatrix} = \begin{bmatrix} 6 \\ -10 \\ -1 \\ 0 \end{bmatrix}$$

Como a matriz obtida é, agora sim, invertível, pode-se resolver o sistema normalmente, de forma que encontram-se os resultados $r_1 = \frac{19}{8}$, $r_2 = -\frac{5}{2}$, $r_3 = -\frac{9}{8}$ e $r_4 = \frac{5}{4}$.

Curiosamente, tanto a solução com a utilização da pseudoinversa quanto a sugestão de Milton e Meyers resultam no mesmo vetor, que é a solução de mínimos quadrados para o sistema sobredeterminado. Este fato será mais explorado no capítulo seguinte.

Para a compreensão de que, realmente, os resultados encontrados são os que possuem erro quadrático mínimo, é válido comparar o resultado encontrado, $\hat{r} = [\frac{19}{8}; -\frac{5}{2}; -\frac{9}{8}; \frac{5}{4}]$ com outras soluções possíveis. Encontrando outras inversas generalizadas, diferentes da inversa de Moore-Penrose, é possível encontrar outros vetores que solucionam a Equação Normal. É o caso das inversas G_1 e G_2 , calculadas de acordo com as instruções de Myers e Milton (1991):

$$r' = G_1 X^T y = \begin{bmatrix} \frac{5}{8} & \frac{1}{4} & \frac{1}{8} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 \\ \frac{1}{8} & \frac{1}{4} & \frac{5}{8} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 6 \\ -10 \\ -1 \\ 5 \end{bmatrix} = \begin{bmatrix} \frac{9}{8} \\ -\frac{15}{4} \\ -\frac{19}{8} \\ 0 \end{bmatrix}$$

$$r'' = G_2 X^T y = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{5}{8} & \frac{1}{2} & \frac{3}{8} \\ 0 & \frac{1}{2} & 1 & \frac{1}{2} \\ 0 & \frac{3}{8} & \frac{1}{2} & \frac{5}{8} \end{bmatrix} \begin{bmatrix} 6 \\ -10 \\ -1 \\ 5 \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{39}{8} \\ -\frac{7}{2} \\ -\frac{9}{8} \end{bmatrix}$$

De fato, r' e r'' são soluções do sistema normal. Ao calcular o erro das soluções encontradas sobre o sistema original, $e = Xr - y$, pode-se encontrar a norma do vetor equivalente a este erro. Neste caso, são calculados os erros \hat{e} para \hat{r} , e' para r' e e'' para r'' :

$$\hat{e} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & -1 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} \frac{19}{8} \\ -\frac{5}{2} \\ -\frac{9}{8} \\ \frac{5}{4} \end{bmatrix} - \begin{bmatrix} 4 \\ 0 \\ 7 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{7}{8} \\ -\frac{19}{8} \\ -\frac{13}{4} \\ -\frac{7}{8} \\ -\frac{19}{8} \end{bmatrix} \Rightarrow \|\hat{e}\| \approx 4,83$$

$$e' = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & -1 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} \frac{9}{8} \\ -\frac{15}{4} \\ -\frac{19}{8} \\ 0 \end{bmatrix} - \begin{bmatrix} 4 \\ 0 \\ 7 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{7}{8} \\ -\frac{19}{8} \\ -\frac{13}{4} \\ -\frac{7}{8} \\ -\frac{19}{8} \end{bmatrix} \Rightarrow \|e'\| \approx 4,83$$

$$e'' = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & -1 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ -\frac{39}{8} \\ -\frac{7}{2} \\ -\frac{9}{8} \end{bmatrix} - \begin{bmatrix} 4 \\ 0 \\ 7 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{7}{8} \\ -\frac{19}{8} \\ -\frac{13}{4} \\ -\frac{7}{8} \\ -\frac{19}{8} \end{bmatrix} \Rightarrow \|e''\| \approx 4,83$$

Como o sistema normal tem como solução as soluções de norma mínima, então todas as três possíveis soluções do sistema apresentam o mesmo erro, que é mínimo, com relação ao sistema original. Apesar disso, $\|\hat{r}\| \approx 3,83$, enquanto $\|r'\| \approx 4,57$ e $\|r''\| \approx 6,10$; a solução encontrada com o uso da pseudoinversa é a que possui norma euclidiana mínima.

Exemplo 2.5. No seguinte sistema linear $Ax = b$, determine uma solução de mínimos quadrados.

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 4 \\ -4 \end{bmatrix}$$

Como as linhas da matriz A de novo somam 0, segue que A é singular e infinitas soluções podem existir para este sistema. Isto pode ser verificado por meio de três métodos. Em primeiro lugar, serão utilizadas inversas generalizadas para encontrar duas possíveis soluções. Em segundo lugar, será realizado o escalonamento da matriz para construir um conjunto solução generalizado, com base em parâmetros. Em terceiro lugar, a pseudoinversa poderá encontrar uma solução de norma mínima.

Como na matriz 2×2 do problema temos $L_1 = -L_2$, sabemos que o posto da matriz é 1. Ao lado direito da equação, também verifica-se que o primeiro elemento do vetor é igual ao inverso do segundo. Assim, o sistema linear possui infinitas soluções. Duas podem ser encontradas com as seguintes inversas generalizadas G_1 e G_2 , encontradas por meio do algoritmo descrito por Myers e Milton (1991).

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 4 \\ -4 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \end{bmatrix} \quad (2)$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ -4 \end{bmatrix} = \begin{bmatrix} 0 \\ -4 \end{bmatrix} \quad (3)$$

De fato, os pares encontrados em (2) e (3) são soluções do sistema. Para ir além e determinar um conjunto com todas as possíveis soluções a depender de um parâmetro, considere o escalonamento do sistema.

$$\left[\begin{array}{cc|c} 1 & -1 & 4 \\ -1 & 1 & -4 \end{array} \right] \xrightarrow{L_2=L_2+L_1} \left[\begin{array}{cc|c} 1 & -1 & 4 \\ 0 & 0 & 0 \end{array} \right]$$

Com isto, a equação $x - y = 4$ aponta as possíveis soluções do sistema. Esta equação pode ser interpretada como a equação de uma reta no plano, de forma que todos os infinitos pontos desta reta são soluções do sistema. Como esperado, os vetores em (2) e (3) fazem parte desta reta.

Dentre todos os infinitos pontos pertencentes à reta, como será possível encontrar o que descreve um vetor de norma euclidiana mínima? Como visto ao longo deste capítulo, essa solução é encontrada a partir da pseudoinversa. Por meios computacionais, obtém-se o seguinte.

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} 4 \\ -4 \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \end{bmatrix} \quad (4)$$

Verifica-se que o resultado também é solução do sistema linear. Porém, enquanto as soluções (2) e (3) apresentam normas euclidianas iguais a 4, o resultado em 4 apresenta norma $2\sqrt{2} \approx 2,82$, verificando que a pseudoinversa possui norma mínima.

Exemplo 2.6. *No seguinte sistema linear $Ax = y$, determine um vetor x que minimize o erro quadrático.*

$$\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ -2 \end{bmatrix}$$

O sistema será novamente resolvido de três formas: por inversas generalizadas para duas possíveis soluções, por parametrização para se obter um conjunto de todas as soluções, e pela pseudoinversa para se encontrar a de norma euclidiana mínima.

Percebe-se que, na matriz A , $-L_1 - L_3 = L_2$. Ou seja, o posto da matriz é 2. É possível encontrar inversas generalizadas G_1 e G_2 por meio do método descrito por Myers e Milton (1991) e assim encontrar duas possíveis soluções.

$$x_1 = G_1 y = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \\ -2 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} \\ \frac{4}{3} \\ 0 \end{bmatrix}$$

$$x_2 = G_2 y = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 0 \\ 2 \\ -2 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{2}{3} \\ -\frac{2}{3} \end{bmatrix}$$

Assim, duas possíveis soluções são $x_1 = (\frac{2}{3}, \frac{4}{3}, 0)$ e $x_2 = (0, \frac{2}{3}, -\frac{2}{3})$. Como segundo método para a solução do sistema, é possível pensar em uma interpretação geométrica, que permite encontrar o conjunto de todas as infinitas soluções. Considere a eliminação de Gauss do sistema em seu formato reduzido.

$$\left[\begin{array}{ccc|c} 2 & -1 & -1 & 0 \\ -1 & 2 & -1 & 2 \\ -1 & -1 & 2 & -2 \end{array} \right] \xrightarrow{L_2=2L_2+L_1} \left[\begin{array}{ccc|c} 2 & -1 & -1 & 0 \\ 0 & 3 & -3 & 4 \\ -1 & -1 & 2 & -2 \end{array} \right]$$

$$\xrightarrow{L_3=2L_3+L_1} \left[\begin{array}{ccc|c} 2 & -1 & -1 & 0 \\ 0 & 3 & -3 & 4 \\ 0 & -3 & 3 & -4 \end{array} \right] \xrightarrow{L_3=L_3+L_2} \left[\begin{array}{ccc|c} 2 & -1 & -1 & 0 \\ 0 & 3 & -3 & 4 \\ 0 & 0 & 0 & 0 \end{array} \right]$$

Assim, o sistema pode ser reescrito como o seguinte sistema de equações:

$$\begin{cases} 2a - b - c = 0 \\ 3b - 3c = 4 \end{cases}$$

Este sistema possui infinitas soluções. Manipulando as equações acima, pode-se chegar à conclusão de que:

$$b = c + \frac{4}{3}; a = c + \frac{2}{3}$$

Alternativamente, pode-se determinar o conjunto solução como:

$$S = \left\{ \left(c + \frac{2}{3}, c + \frac{4}{3}, c \right), c \in \mathbb{R} \right\}$$

Perceba que as soluções x_1 e x_2 encontradas se enquadram no conjunto, para os valores de $c = 0$ e $c = -\frac{2}{3}$. Outras soluções podem ser indicadas, como $x_3 = (\frac{5}{3}, \frac{7}{3}, 1)$, para $c = 1$, por exemplo.

O conjunto solução encontrado pode ser compreendido como a equação paramétrica de uma reta no espaço. Da mesma forma, cada equação do sistema original pode ser compreendida como um plano, dentre os quais a interseção é justamente a reta encontrada.

Para encontrar o resultado que possui norma euclidiana mínima, toma-se a solução encontrada por meio da pseudoinversa, calculada por meios computacionais.

$$A^+ = \begin{bmatrix} \frac{2}{9} & -\frac{1}{9} & -\frac{1}{9} \\ -\frac{1}{9} & \frac{2}{9} & -\frac{1}{9} \\ -\frac{1}{9} & -\frac{1}{9} & \frac{2}{9} \end{bmatrix}$$

Portanto, a solução de norma mínima será:

$$\hat{x} = A^+y = \begin{bmatrix} \frac{2}{9} & -\frac{1}{9} & -\frac{1}{9} \\ -\frac{1}{9} & \frac{2}{9} & -\frac{1}{9} \\ -\frac{1}{9} & -\frac{1}{9} & \frac{2}{9} \end{bmatrix} \begin{bmatrix} 0 \\ 2 \\ -2 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{2}{3} \\ -\frac{2}{3} \end{bmatrix}$$

Perceba que $\hat{x} = x_2$, ou seja, a solução de norma euclidiana mínima não é obtida exclusivamente pela pseudoinversa! Outras inversas generalizadas podem levar ao mesmo resultado. No próximo capítulo exploraremos mais a fundo que há formas diferentes de se obter a solução de norma euclidiana mínima.

A norma da solução encontrada pela pseudoinversa é $\|\hat{x}\| \approx 0,94$, mesmo valor encontrado para $\|x_2\|$. Já para as outras soluções encontradas, temos que $\|x_1\| \approx 1,49$ e $\|x_3\| \approx 3,03$.

3 ÍNDICE DE GINI, ÍNDICE DE DESENVOLVIMENTO HUMANO E RANKING DO PISA

Antes de explorar métodos matemáticos para ranqueamento de itens, que são mais elaborados e requerem conceitos de álgebra linear e resolução de sistemas, é útil verificar a necessidade de rankings em uma dimensão muito importante da formulação de políticas públicas: os indicadores sociais.

Os indicadores sociais, ou índices socioeconômicos, são largamente usados em debates públicos e meios de comunicação do Brasil e do mundo. É comum um cidadão ouvir a respeito de taxa de analfabetismo, nível de desemprego, taxa de inflação, Produto Interno Bruto (PIB) e outros. A propagação do uso de tais índices tornou muito difícil a defesa de qualquer medida pública ou iniciativa que não leve em conta o estado atual de seu objetivo utilizando um indicador social.

Jannuzzi define indicador social da seguinte forma:

Um Indicador Social é uma medida em geral quantitativa dotada de significado social substantivo, usado para substituir, quantificar ou operacionalizar um conceito social abstrato, de interesse teórico (para pesquisa acadêmica) ou programático (para formulação de políticas). É um recurso metodológico, empiricamente referido, que informa algo sobre um aspecto da realidade social ou sobre mudanças que estão se processando na mesma (Jannuzzi, 2001, p. 15).

Neste capítulo, o objetivo é desenvolver a base teórica e matemática utilizada na formação de dois índices: o Índice de Desenvolvimento Humano (IDH) e o Coeficiente de Gini. Depois, será considerado brevemente um importante ranqueamento internacional voltado à educação, a avaliação PISA.

3.1 Coeficiente de Gini

O coeficiente de Gini é um índice criado pelo estatístico italiano Corrado Gini, e é utilizado para medição da desigualdade de renda de uma região. O Banco Mundial define o índice, em seu glossário, da seguinte forma.

O índice de Gini mede a extensão em que a distribuição de renda (ou, em alguns casos, o consumo) entre indivíduos ou famílias dentro de uma economia se desvia de uma distribuição perfeitamente igual. Uma curva de Lorenz traça as porcentagens cumulativas da renda total recebida contra o número cumulativo de beneficiários, começando com o indivíduo ou família mais pobre. O índice de Gini mede a área entre a curva de Lorenz e uma linha hipotética de igualdade absoluta, expressa como uma porcentagem da área máxima sob a linha. Assim, um índice de Gini de 0 representa igualdade perfeita, enquanto um índice de 100 implica desigualdade perfeita (WORLD BANK GROUP, s.d.a).

O cálculo do índice de Gini, assim como outros índices de desigualdade, está apoiado na curva de Lorenz. Tal conceito é melhor entendido primeiro com um exemplo. A seguinte tabela tem valores retirados da Pesquisa Nacional por Amostra de Domicílios (Pnad) de 2003, e é baseada em Hoffmann (2006).

Tabela 1 – Distribuição das pessoas ocupadas conforme a renda obtida na atividade exercida, no Brasil, de acordo com a Pnad de 2003

Estrato	Porcentagem no estrato		Porcentagem acumulada	
	Da população	Da renda total	Da população	Da renda total
I	30	7	30	7
II	20	9	50	16
III	20	13	70	29
IV	10	10	80	39
V	10	16	90	55
VI	5	13	95	68
VII	4	19	99	87
VIII	1	13	100	100

Fonte: Hoffmann (2006).

A partir dos pontos encontrados nas duplas ordenadas formadas pelas colunas 4 e 5 da tabela 1, é possível determinar pontos em um plano cartesiano. Com a porcentagem acumulada da população no eixo horizontal e a porcentagem acumulada da renda total no eixo vertical, no plano ortogonal formam-se segmentos que lembram uma curva, como disposta na Figura 4. Como aqui está representado um caso discreto, pode-se utilizar a nomenclatura Poligonal de Lorenz. Se for considerado um estrato maior da população, os n valores podem ser generalizados para infinitos, gerando uma curva suave, e sempre crescente. O índice de Gini é definido exatamente nas áreas delimitadas pela curva descrita acima. Para o exemplo da tabela, o cálculo do índice será feito mais adiante neste capítulo.

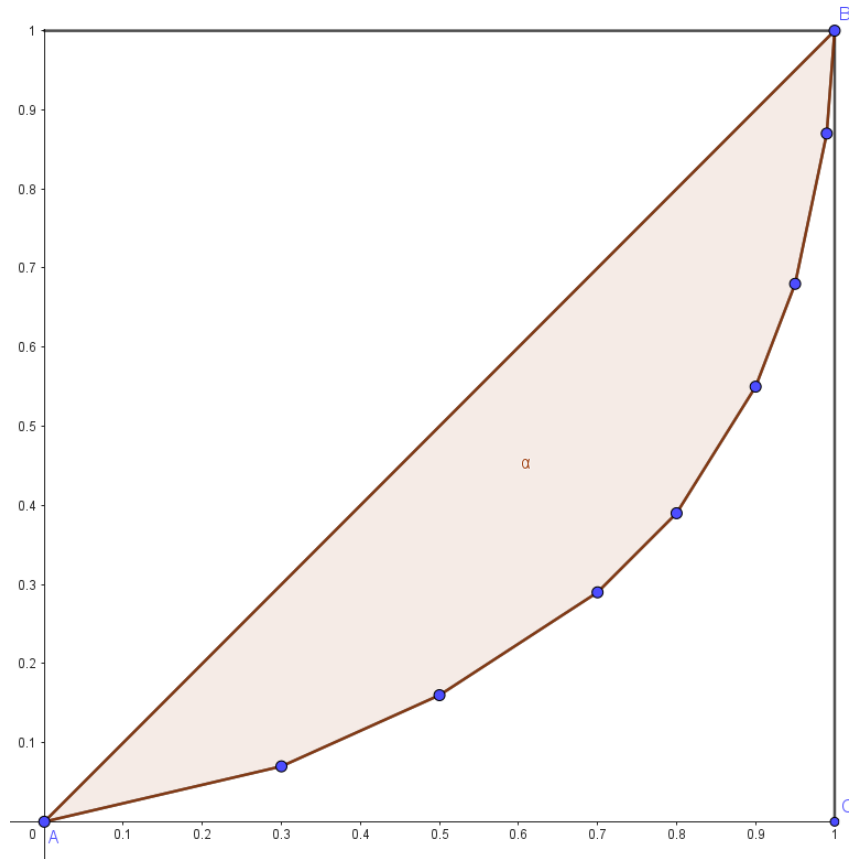
Com o exemplo anterior, pode-se definir formalmente a curva de Lorenz a partir de Gastwirth (1972).

Definição 3.1. *Seja a população ordenada de forma crescente com relação à renda mensal, e dividida em n estratos, de forma que Q_i é o número de pessoas representadas no estrato i . Ainda, sejam n números ordenados $x_1 \leq x_2 \leq \dots \leq x_n$ equivalentes à renda total acumulada mensalmente pelo estrato i da população. A curva de Lorenz $L(p)$ está no intervalo $(0,1)$ e é definida nos pontos (p_i, θ_i) , definidos da seguinte forma:*

- p_i é o valor da proporção acumulada da população até certo estrato i , ou seja,

$$p_i = \frac{\sum_{k=1}^i Q_k}{\sum_{t=1}^n Q_t} \quad (5)$$

Figura 4 – A curva de Lorenz em estratos discretos



Fonte: O autor.

- θ_i é o valor da proporção acumulada da renda até certo estrato i , ou seja,

$$\theta_i = \sum_{k=0}^i x_k / \sum_{t=0}^n x_t \quad (6)$$

Por definição, $L(0) = 0$ e $L(1) = 1$.

Utilizando-se p_i como parâmetros para o eixo horizontal, e θ_i para o eixo vertical, forma-se a curva de Lorenz. Perceba no exemplo de Hoffmann (2006), $p_i > \theta_i$, para qualquer i , de forma que a curva sempre se encontra abaixo da reta identidade, que representa total igualdade. Tal fato pode ser generalizado para toda curva de Lorenz, o que será provado a partir das seguintes proposições.

Proposição 3.1. *Em uma população organizada conforme a Definição 3.1, as rendas médias dos estratos são ordenadas de forma crescente.*

Considere que, caso a renda média do estrato $i + 1$ seja menor do que a renda média do estrato i , então pelo menos um indivíduo do estrato $i + 1$ possui renda mensal menor do que a média do estrato i . Neste caso, tal indivíduo deveria pertencer ao estrato i , e não ao $i + 1$. Dessa forma, como a população está ordenada de forma crescente com relação à renda mensal, sempre é o caso em que as rendas médias de cada estrato estão ordenadas de forma crescente.

Proposição 3.2. *Em uma poligonal de Lorenz, construída a partir de uma população dividida em n estratos, e formada por $n - 1$ segmentos de reta, o coeficiente angular de cada segmento de reta é sempre maior ou igual ao anterior. Ou seja, se a_i é o coeficiente angular do segmento de reta entre p_{i-1} e p_i , então*

$$a_i \leq a_{i+1}$$

Demonstração. Primeiro, para facilitar a notação, serão consideradas as seguintes constantes:

$$\sum_{k=1}^n Q_k = P \quad e \quad \sum_{k=1}^n x_k = R \quad (7)$$

Então, o coeficiente angular de um segmento de reta qualquer na poligonal de Lorenz é dado da seguinte forma.

$$a_i = \frac{\theta_i - \theta_{i-1}}{p_i - p_{i-1}} = \frac{x_1 + x_2 + \dots + x_i - (x_1 + x_2 + \dots + x_{i-1})}{\frac{R}{Q_1 + Q_2 + \dots + Q_i - (Q_1 + Q_2 + \dots + Q_{i-1})}}$$

$$a_i = \frac{\frac{x_i}{R}}{\frac{Q_i}{P}} \quad (8)$$

Assim, para provar que $a_i \leq a_{i+1}$, considere a subtração:

$$a_{i+1} - a_i = \frac{\frac{x_{i+1}}{R}}{\frac{Q_{i+1}}{P}} - \frac{\frac{x_i}{R}}{\frac{Q_i}{P}}$$

$$a_{i+1} - a_i = \frac{P}{R} \left(\frac{x_{i+1}}{Q_{i+1}} - \frac{x_i}{Q_i} \right)$$

Veja que $\frac{x_{i+1}}{Q_{i+1}}$ é a renda média da população no estrato Q_{i+1} . Como as rendas totais acumuladas por estrato estão organizadas de forma crescente, então as rendas médias dos estratos também estão organizadas de forma crescente. Dessa forma:

$$\frac{x_{i+1}}{Q_{i+1}} > \frac{x_i}{Q_i} \Rightarrow \frac{P}{R} \left(\frac{x_{i+1}}{Q_{i+1}} - \frac{x_i}{Q_i} \right) > 0$$

$$a_{i+1} > a_i \quad (9)$$

□

Portanto, a poligonal de Lorenz é formada por segmentos de reta sempre crescentes. Sendo assim, unindo este fato à definição de que $L(0) = 0$ e $L(1) = 1$, deve ser o caso em

que a curva deve estar sempre abaixo da reta identidade. Se a curva ultrapassasse a identidade, a inclinação cada vez maior dos segmentos de reta tornaria impossível que $L(1) = 1$.

Destacam-se assim, no triângulo ABC, duas situações na análise de desigualdade de uma região. Caso todas as famílias do estudo recebam a mesma renda, e perceba-se situação de perfeita igualdade, então a curva de Lorenz seria uma reta, exatamente a hipotenusa do triângulo (a reta identidade). Assim, cada porcentagem acumulada da população teria como correspondente o mesmo percentual acumulado da renda total. Opõe-se a esse caso a situação em que apenas uma pessoa detém 100% da renda acumulada da região, e todos os demais habitantes possuem 0%. Se este fosse o caso, a curva de Lorenz seria formada por retas coincidentes com os eixos X e Y.

Com a curva de Lorenz podemos construir um índice que determine o quão desigual é a distribuição de renda de uma região. Uma forma de fazer isto é determinar a “distância” entre a distribuição de renda verificada e a situação de perfeita igualdade; ou seja, o quão perto a curva de Lorenz está da reta da diagonal do quadrado. A seguinte definição é baseada em Hoffmann (2006, p. 337).

Definição 3.2. *Seja a região entre a reta de igualdade e a curva de Lorenz chamada de área de concentração, e denotada por α . O Índice de Gini é a razão entre a área de concentração e a área do triângulo ABC, isto é:*

$$G = \frac{\alpha}{0,5} = 2\alpha \quad (10)$$

Na Figura 4, a área de concentração está pintada de marrom, e denotada por α . Percebe-se, então, que quanto mais próximo o coeficiente de Gini estiver de 0, mais igualitária é a divisão de recursos na região, uma vez que a distância entre a curva de Lorenz e a reta de perfeita igualdade é mais estreita.

Cabe aqui ressaltar que neste trabalho não serão abordados os casos de famílias com rendas negativas; ou a consideração de que não há limite superior para os valores de x_i descritos na Definição 3.1; ou as diferentes demonstrações para casos discretos e contínuos de distribuições de renda. A elaboração sobre tais casos foge do foco deste trabalho, mas as conclusões são as mesmas para todos os casos, de acordo com Dorfman (1979).

Ainda utilizando a curva de Lorenz como referência, o índice de Gini poderia ser calculado de outras formas. Considerando a área abaixo da curva por meio do uso do cálculo integral, chegamos às seguintes conclusões.

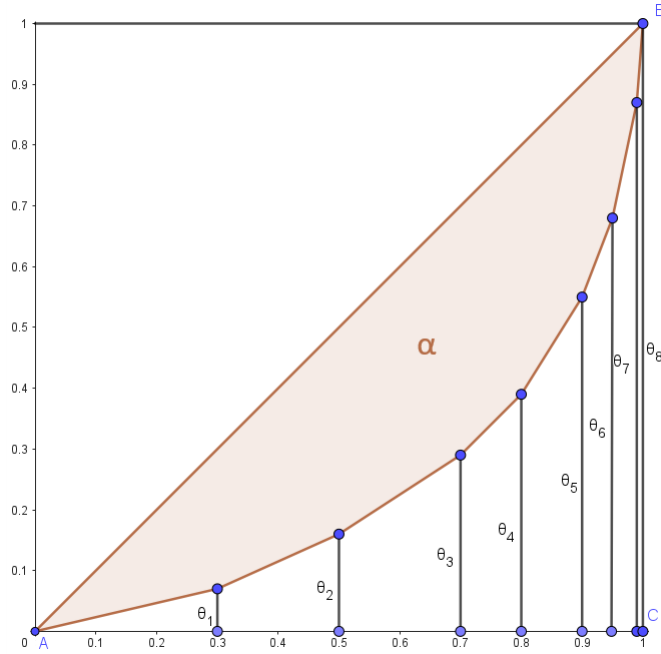
$$\alpha + \int_0^1 L(p)dp = 0,5 \quad (11)$$

$$\Rightarrow 2\alpha + 2 \int_0^1 L(p)dp = 1 \quad (12)$$

$$\Rightarrow G = 1 - 2 \int_0^1 L(p)dp \quad (13)$$

É claro que, na realidade das pesquisas estatísticas, não é possível entrevistar infinitas famílias e obter infinitos valores de renda para obter-se uma curva suave e integrável. Por isso, como verificado no exemplo da Figura 4, é possível calcular a área equivalente à integral usada acima usando as áreas das regiões trapezoidais delimitadas por θ_{i-1} e θ_i , em um processo semelhante à Soma de Riemann. A Figura 5 mostra como tais regiões seriam determinadas no caso exemplo.

Figura 5 – As regiões trapezoidais determinadas na poligonal de Lorenz



Fonte: O autor.

Fazendo assim, podemos elaborar a fórmula para o coeficiente de Gini, considerando $\theta_0 = 0$.

$$G = 1 - 2 \cdot \frac{1}{2} \sum_{i=1}^n (\theta_{i-1} + \theta_i)(p_i - p_{i-1}) \quad (14)$$

$$G = 1 - \sum_{i=1}^n (\theta_{i-1} + \theta_i)(p_i - p_{i-1}) \quad (15)$$

Tal soma pode ser simplificada de forma a obter-se a seguinte equação.

$$G = 1 - \left[1 + \sum_{i=1}^{n-1} \theta_i (p_{i+1} - p_{i-1}) - \theta_n p_{n-1} \right]$$

$$G = \theta_n p_{n-1} - \sum_{i=1}^{n-1} \theta_i (p_{i+1} - p_{i-1}) \quad (16)$$

A partir de (16), pode-se calcular o índice para o exemplo explorado por Hoffmann (2006). Assim,

$$G = 0,99 - 0,07(0,5 - 0) - 0,16(0,7 - 0,3) - \dots - 0,87(1 - 0,95) \quad (17)$$

$$G = 0,5388 \quad (18)$$

Com isto, fica definido o índice de Gini do país no período. Uma vez compreendida a forma de cálculo no exemplo brasileiro, podemos calcular o Índice de Gini para outros países, utilizando dados disponíveis abertamente online. Consideremos o caso da Islândia.

Tabela 2 – Distribuição de renda ou consumo na Islândia

Estrato	Porcentagem no estrato		Porcentagem acumulada	
	Da população	Da renda total	Da população	Da renda total
I	20	9,7	20	9,7
II	20	14,2	40	23,9
III	20	17,9	60	41,8
IV	20	22,3	80	64,1
V	20	35,9	100	100

Fonte: World Bank Group (s.d.b).

O índice pode ser calculado a partir da tabela de acordo com 16.

$$G = 0,8 - 0,097(0,4 - 0) - 0,239(0,6 - 0,2) - \dots - 0,641(1 - 0,6) \quad (19)$$

$$G = 0,242 \quad (20)$$

A partir de dados do Banco Mundial, pode-se obter o ranking de países com menor Índice de Gini, ou seja, menos desiguais. Na Tabela 3, os dados não são todos do mesmo ano, mas são os mais recentes de cada país.

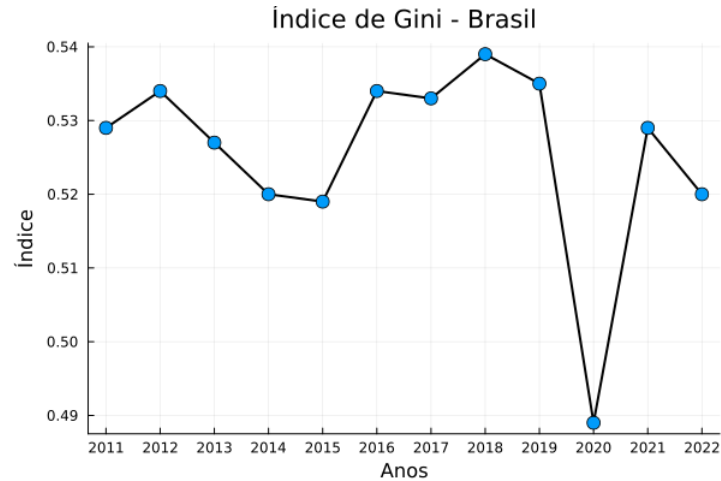
Tabela 3 – Ranking mundial do Índice de Gini

Posição	País	Índice de Gini
1	Eslováquia	0,241
2	Eslovênia	0,243
3	Bielorússia	0,244
4	Ucrânia	0,256
5	Moldávia	0,257
6	Holanda	0,257
7	Islândia	0,261
8	Tchéquia	0,262
9	Quirguistão	0,264
10	Emirados Árabes Unidos	0,263
163	Brasil	0,52

Fonte: World Bank Group (2025a).

É útil, ainda, analisar a variação do índice de Gini do Brasil ao longo dos anos. Tal análise está no gráfico da Figura 6.

Figura 6 – Evolução do Índice de Gini brasileiro desde 2011.



Fonte: World Bank Group (2025b).

3.2 Índice de Desenvolvimento Humano

Pela necessidade de se medir o desenvolvimento econômico de uma região, a literatura econômica se voltou para o PIB *per capita*, considerando que a divisão dos recursos produzidos por toda a nação entre os seus habitantes satisfaria as necessidades deste indicador social. Pelos motivos vistos na seção acima, a divisão e a captação dos recursos produzidos pela região não são iguais para toda a população, o que torna tal medida incompleta para seus fins. O Índice de Desenvolvimento Humano (IDH) surge a partir de estudos realizado pelo Instituto de Pesquisa e Desenvolvimento Social das Nações Unidas na década de 60, apesar de ser divulgado nesses termos apenas na década de 80 (Jannuzzi, 2001).

Jannuzzi descreve qual era o objetivo das Nações Unidas no desenvolvimento do IDH.

Na concepção do PNUD [Programa das Nações Unidas para o Desenvolvimento], o Desenvolvimento Humano deveria ser entendido como um processo dinâmico e permanente de ampliação das oportunidades dos indivíduos para a conquista de níveis crescentes de bem-estar. Para tanto, o processo de desenvolvimento deveria garantir, entre outros aspectos, oportunidades crescentes de acesso à educação e cultura, a condições de desfrutar uma vida saudável e longa e a condições de dispor de um padrão adequado de vida para a população (Jannuzzi, 2001, p. 120).

São exatamente estas três dimensões citadas que foram consideradas para a elaboração do IDH. Três indicadores são, portanto, aglutinados por meio de uma média geométrica para a elaboração do índice: um para nível educacional, um para esperança de vida, e outro para o PIB *per capita* ajustado segundo uma técnica que será elaborada a seguir.

Para a construção do índice, primeiramente tomam-se dados para expectativa de vida ao nascer, anos esperados de estudo, média de anos de estudo, e PIB *per capita* de fontes internacionais, quase em sua totalidade órgãos pertencentes às Nações Unidas. Esses valores são então normalizados para que assumam valores de 0 a 1, facilitando o manejo e leitura dos índices. Isso é feito por meio do seguinte cálculo.

$$\text{Índice normalizado} = \frac{\text{valor verificado} - \text{valor mínimo}}{\text{valor máximo} - \text{valor mínimo}} \quad (21)$$

Os valores usados como máximos e mínimos estão apresentados na Tabela 4. Para o índice sobre educação, os dois indicadores expostos são unidos por média aritmética.

Tabela 4 – Parâmetros para normalização de índices do IDH

Dimensão	Indicador	Mínimo	Máximo
Saúde	Esperança de vida à nascença (anos)	20	85
Educação	Anos de escolaridade previstos	0	18
	Média de anos de escolaridade	0	15
Padrão de vida	PIB <i>per capita</i> (2017 PPP\$)	100	75000

Fonte: PNUD (2023).

Na dimensão “Padrão de vida”, a sigla 2017 PPP\$ representa o estudo realizado em 2017 chamado *purchasing power parity*, ou paridade de poder de compra. O PIB *per capita* é, portanto, convertido nos termos deste PPP para considerar não apenas a “quantidade” de dinheiro, mas o que esse dinheiro pode comprar. Além disso, a normalização do índice nos padrões da Equação 21 é feita com a utilização do logaritmo natural.

Monteiro (2008) explica que o uso do logaritmo se justifica pelo interesse no aumento relativo, e não absoluto, da renda: “um real extra de renda, quando a renda é de 10 mil reais, não é insumo tão importante quanto um real extra quando a renda é de 100 reais” (Monteiro, 2008). O autor indica, ainda, que a exploração destes métodos de cálculo de índices pode ser uma boa atividade interdisciplinar entre matemática e geografia.

Para exemplificar o método de determinação do índice, calcularemos o caso do Brasil no relatório de 2021-22.

Tabela 5 – Exemplo do cálculo do IDH - exemplo do Brasil

Indicador	Valor
Esperança de vida à nascença (anos)	72,8
Anos de escolaridade previstos	15,6
Média de anos de escolaridade	8,1
PIB <i>per capita</i> (2017 PPP\$)	14370

Fonte: PNUD (2023).

Com tais dados e a metodologia descrita no capítulo é possível alcançar o índice divulgado pelo PNUD.

$$I_{SAUDE} = \frac{72,8 - 20}{85 - 20} \approx 0,812 \quad (22)$$

$$\text{Anos de escolaridade previstos} = \frac{15,6 - 0}{18 - 0} \approx 0,867 \quad (23)$$

$$\text{Média de anos de escolaridade} = \frac{8,1 - 0}{15 - 0} = 0,54 \quad (24)$$

$$I_{EDUCACAO} = \frac{0,867 + 0,54}{2} \approx 0,703 \quad (25)$$

$$I_{REND A} = \frac{\ln(14370) - \ln(100)}{\ln(75000) - \ln(100)} \approx 0,750 \quad (26)$$

$$IDH = (0,812 \cdot 0,703 \cdot 0,750)^{1/3} \approx 0,754 \quad (27)$$

Outro exemplo de cálculo de IDH pode ser feito sobre a Islândia, que figurou em altas posições no último ranking.

Tabela 6 – Exemplo do cálculo do IDH - exemplo da Islândia

Indicador	Valor
Esperança de vida à nascença (anos)	82,7
Anos de escolaridade previstos	19,2
Média de anos de escolaridade	13,8
PIB <i>per capita</i> (2017 PPP\$)	55782

Fonte: PNUD (2023).

No cálculo do IDH, os 19,2 anos de escolaridade previstos para a população islandesa serão desconsiderados, uma vez que o parâmetro máximo definido é de 18 anos. Assim, o cálculo se dá da seguinte forma.

$$I_{SAUDE} = \frac{82,7 - 20}{85 - 20} \approx 0,964 \quad (28)$$

$$\text{Anos de escolaridade previstos} = \frac{18 - 0}{18 - 0} = 1 \quad (29)$$

$$\text{Média de anos de escolaridade} = \frac{13,8 - 0}{15 - 0} = 0,92 \quad (30)$$

$$I_{EDUCACAO} = \frac{1 + 0,92}{2} = 0,96 \quad (31)$$

$$I_{REND A} = \frac{\ln(55782) - \ln(100)}{\ln(75000) - \ln(100)} \approx 0,955 \quad (32)$$

$$IDH = (0,964 \cdot 0,96 \cdot 0,955)^{1/3} \approx 0,959 \quad (33)$$

Desde o relatório composto pela PNUD de 2010, foram incluídos outros índices que tentam considerar outras dimensões do desenvolvimento humano na formação do seu valor final. Nos relatórios seguintes, o mesmo foi feito até o total de 6 índices no relatório aqui estudado. São eles: o próprio Índice de Desenvolvimento Humano; o Índice de Desenvolvimento Humano

Ajustado à Desigualdade (IDHAD); o Índice de Desenvolvimento Humano por Gênero (IDHG); o Índice de Desigualdade de Gênero (IDG); o Índice de Pobreza Multidimensional (IPM); e o Índice de Desenvolvimento Humano Ajustado às Pressões sobre o Planeta (IDHP).

Estes diversos índices revelam o quão distante estamos do consenso na medição deste objetivo tão subjetivo: o desenvolvimento. A escolha do que considerar ou ignorar no cálculo é crucial, e afeta o resultado drasticamente. No IDHAD, por exemplo, o Brasil alcançou valor 0,576, enquanto o topo da lista foi ocupado pela Islândia, com valor 0,915.

Com dados mais recentes do PNUD, já mais atualizados com relação ao relatório de 2021/2022, pode-se obter um ranking de 193 países e territórios a partir do IDH obtido. O início deste ranking, e a posição do Brasil, estão abaixo.

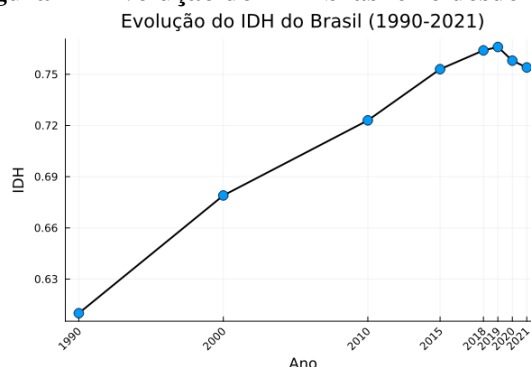
Tabela 7 – Ranking mundial do IDH

Posição	País	IDH
1	Suíça	0,967
2	Noruega	0,966
3	Islândia	0,959
4	Hong Kong, China	0,956
5	Dinamarca	0,952
5	Suécia	0,952
7	Alemanha	0,950
7	Irlanda	0,950
9	Singapura	0,949
10	Austrália	0,946
89	Brasil	0,760

Fonte: PNUD (2025).

Pode-se também obter um gráfico da variação dos resultados de IDH do Brasil ao longo dos anos, o que está na Figura 7.

Figura 7 – Evolução do IDH brasileiro desde 1990.



Fonte: PNUD (2023).

3.3 Ranking do PISA

O Programa Internacional de Avaliação de Estudantes (PISA) é uma iniciativa da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), cujo objetivo é avaliar

e classificar os países participantes de sua organização quanto à qualidade da educação. No relatório de resultados do PISA de 2022, a OCDE definiu as metas da avaliação da seguinte forma.

A ideia por trás do PISA reside em testar diretamente o conhecimento e as habilidades dos alunos, por meio de uma métrica acordada internacionalmente; vinculando isso com dados de alunos, professores, escolas e sistemas para entender as diferenças de desempenho; e, em seguida, aproveitar o poder da colaboração para agir com base nos dados, tanto criando pontos de referência compartilhados quanto aproveitando a pressão dos pares. [...] Em essência, o PISA mede o que importa e disponibiliza essas informações para educadores e formuladores de políticas, para que possam tomar decisões mais informadas (OCDE, 2023a, p. 5).

No final da seção anterior foi comentado como a medição do desenvolvimento humano é subjetiva e pode levar a resultados ilusórios. Nas notas do PISA, o mesmo acontece. Fica evidente na declaração do órgão a impressão de que há concordância internacional quanto à metodologia, os objetivos e o impacto que a avaliação tem nas políticas públicas do países participantes. Será visto mais à frente que isso não é o caso.

3.3.1 O formato da avaliação

Cada edição da avaliação é realizada a cada três anos, e participam dela uma amostra de estudantes com idade entre 15 anos e 3 meses e 16 anos e 2 meses. Tais estudantes devem ainda ter completado pelo menos 6 anos de educação formal, independentemente do formato ou tipo de escola em que cursaram esses anos. Em cada edição são avaliados três eixos do conhecimento, enquanto um foco especial é escolhido por vez: matemática, leitura e ciências. A última edição até a presente data, realizada em 2022, teve foco em matemática. A de 2018 priorizou leitura, e a de 2015 ciências.

Desde 2018 tem sido comum o uso de computadores para aplicação do teste na maioria dos países participantes. Em 2022, tal uso foi expandido. Além de maior facilidade para retenção e tratamento dos dados de respostas da prova, diversos outros parâmetros foram verificados, como o tempo gasto para cada questão e as ações realizadas antes de marcar a resposta escolhida.

O tempo de prova é de 2 horas para responder questões que misturam múltipla escolha e resposta escrita do estudante. As perguntas são construídas em grupos, narrando uma situação real em que se requer a contextualização de um conceito científico. As questões não são iguais para todos os estudantes. Além das perguntas relacionadas ao conhecimento, cada estudante respondeu também a um questionário com informações a respeito de suas casas, suas crenças, suas escolas e experiências.

A metodologia utilizada para a obtenção da nota do PISA é uma união de diversos recursos estatísticos. A descrição das ferramentas utilizadas é assim descrita no relatório técnico da edição de 2022.

O PISA 2022, assim como outros estudos internacionais de larga escala, utiliza uma abordagem da Teoria de Resposta ao Item (TRI) na análise e escalonamento dos dados e na medição de tendências entre os ciclos. O modelo TRI usado a partir do PISA 2015 passou por algumas modificações em comparação com os ciclos anteriores, que baseavam o escalonamento inteiramente no modelo Rasch. Para aumentar a capacidade do escalonamento em lidar com as complexidades dos dados de resposta do PISA, o PISA 2015 e ciclos posteriores implementaram um modelo híbrido que combinou uma abordagem Rasch com um modelo logístico de dois parâmetros e um modelo de crédito parcial generalizado (MPCG) usado quando apropriado (OCDE, 2023b, p. 16).

Todos estes são métodos psicométricos, ou seja, buscam “medir” o aprendizado do estudante e determinar o grau de seu desempenho a partir da relação entre a resposta dada e a dificuldade da questão. O mesmo tipo de método é utilizado no Exame Nacional do Ensino Médio (ENEM), especialmente quanto ao uso da Teoria de Resposta ao Item (INEP, 2025).

Uma vez que cada prova individual é processada pelo modelo híbrido citado acima, é realizada também uma modelagem da população. O objetivo é compreender a relação entre os resultados obtidos e as informações do questionário de situação social.

A abordagem de escalonamento empregada nas análises dos dados do PISA (modelagem populacional) combina a TRI e a modelagem de regressão latente para aumentar a precisão geral da medição e evitar possíveis vieses na estimação das relações entre proficiência e variáveis contextuais do questionário de contexto (QC). Uma vez que o modelo populacional é estimado, múltiplos valores plausíveis podem ser extraídos para cada aluno a partir de uma distribuição posterior de proficiência que leva em conta as fontes de incerteza nos dados (OCDE, 2023b, p. 206).

A Tabela 8 mostra o ranking da avaliação em Matemática, realizada em 2022.

Tabela 8 – Ranking mundial do PISA

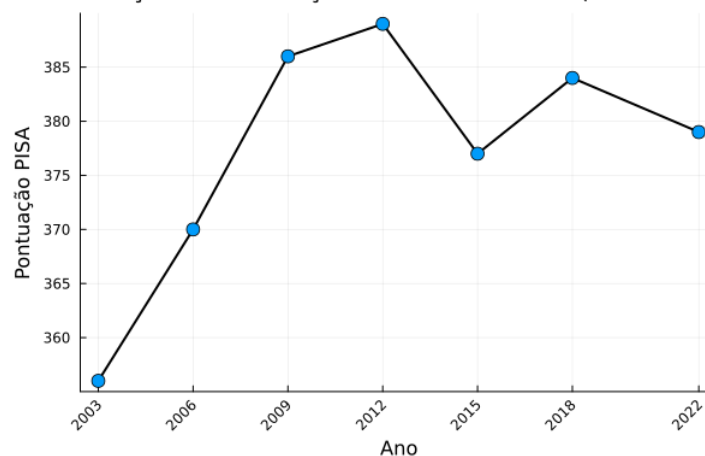
Posição	País	Pontuação
1	Singapura	575
2	Macau, China	552
3	Taiwan	547
4	Hong Kong, China	540
5	Japão	536
6	Coréia do Sul	527
7	Estônia	510
8	Suíça	508
9	Canadá	497
10	Países Baixos	493
65	Brasil	379

Fonte: OCDE (2023a).

Já a variação dos resultados do Brasil ao longo dos anos, na disciplina de Matemática, está disposta na Figura 8.

Figura 8 – Evolução da pontuação do Brasil no PISA.

Evolução da Pontuação do Brasil no PISA (2003-2022)



Fonte: OCDE (2023a).

3.3.2 Críticas acumuladas

O PISA se compromete a desenvolver uma prova metodologicamente robusta e transparente, para deixar em evidência o que é bom e o que falta na educação dos países, e assim permitir a melhor gestão de políticas públicas. Porém, ao longo de suas edições, acumularam-se pontos de crítica ao modelo de ranqueamento realizado, e aos seus efeitos nas secretarias de educação ao redor do mundo.

Um ponto de crítica se dá na formação da amostragem de estudantes para a análise de dados. É permitido que cada país exclua até 5% dos participantes, seja por meio da escola ou pelos estudantes individualmente. O relatório afirma que escolas podem ser excluídas por estarem localizadas em regiões de difícil acesso, por serem muito pequenas, ou devido a fatores operacionais que impedem a participação. Já estudantes podem ser excluídos por deficiências de

aprendizado ou por falta de conhecimento do idioma da avaliação (OCDE, 2023a). Tal permissão causa problemas no ranqueamento, uma vez que na edição de 2022 8 países ultrapassaram o limite de 5%, constando nas divulgações oficiais na posição conquistada, porém com um pequeno asterisco para indicar que as condições do programa não foram seguidas.

Em uma carta aberta aos organizadores da avaliação, publicada em 2014 por pesquisadores ao redor do mundo, a influência do PISA em políticas educacionais é vista com preocupação.

Administrado a cada três anos, os resultados do PISA são ansiosamente aguardados por governos, ministros da educação e conselhos editoriais de jornais, e são citados com autoridade em inúmeros relatórios de políticas públicas. Eles começaram a influenciar profundamente as práticas educacionais em muitos países. Como resultado do PISA, países estão reformulando seus sistemas educacionais na esperança de melhorar suas classificações. A falta de progresso no PISA tem levado a declarações de crise e “choque do PISA” em muitos países, seguidas por pedidos de renúncias e reformas de longo alcance de acordo com os preceitos do PISA. Nós estamos francamente preocupados com as consequências negativas dos rankings do PISA (Andrews et al., 2014).

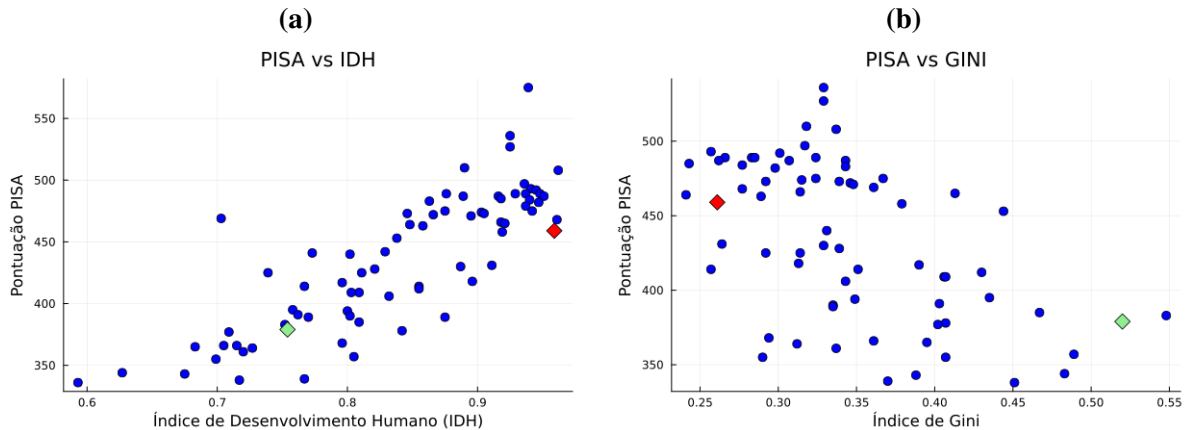
Na mesma carta aberta são compartilhadas preocupações como a confiança demasiada em resultados quantitativos de avaliações em larga escala; a busca por soluções rápidas e eficazes em curto prazo, ignorando que retornos em educação podem levar décadas; manutenção de uma visão de escola e educação que ignora suas dimensões físicas, cívicas, morais e artísticas; além de críticas à organização OCDE, que não possui a tradição e incumbência de influenciar decisões educacionais, como é o caso d/e órgãos como a UNESCO ou a UNICEF.

3.4 Regressão linear entre alguns dos índices

Os índices de Gini e IDH possuem bases matemáticas interessantes, e cumprem seu papel de análise quantitativa da qualidade de vida de uma região. Nesse sentido, espera-se que países com elevado desenvolvimento humano e baixa desigualdade de renda também obtenham bons desempenhos em outros indicadores socioeconômicos e em avaliações de larga escala. Seria interessante verificar como o resultado obtido em cada um desses índices influencia na pontuação obtida na avaliação PISA.

A relação entre tais pontuações pode ser melhor visualizada a partir de dois gráficos de dispersão. No primeiro, define-se o eixo horizontal com valores do IDH, e o eixo vertical com pontuações da avaliação PISA. No segundo gráfico, o eixo horizontal representa coeficientes de Gini, e o eixo vertical permanece com as pontuações PISA. Assim, cada ponto na Figura 9 representa um país. O ponto verde claro representa o Brasil, e o ponto vermelho representa a Islândia, os dois países sobre os quais foram feitas análises ao longo deste capítulo.

Figura 9 – (a) Relação entre IDH e pontuação PISA; (b) relação entre Índice de Gini e pontuação PISA



Fonte: OCDE (2023a); PNUD (2023); World Bank Group (2025a).

Destaca-se que nem todos os países incluídos em um ranking possuem pontuação em outros. Um exemplo disso são as regiões administrativas especiais da China (como Hong Kong e Macau), que figuram nas primeiras posições do ranking PISA, mas não constam no ranking de IDH do PNUD. Casos similares ocorrem no ranking para Índice de Gini do Banco Mundial. Dessa forma, foram considerados apenas os países e regiões presentes em ambos os rankings oficiais em cada gráfico.

Percebe-se nos gráficos uma provável generalização das relações entre os índices. Um país com IDH maior aparenta obter melhores resultados no PISA. Ainda, suspeita-se que quanto maior o índice de Gini, menor o aproveitamento na mesma avaliação. Para explicitar tal relação, pode-se buscar a reta que mais se ajusta ao conjunto de pontos, ou seja, que apresenta menor erro quadrático na diferença entre valor esperado (verificado na reta) e valor real (verificado na dispersão).

Considere o seguinte desenvolvimento matemático no caso da relação entre IDH e PISA, que ocorre de forma análoga na relação com o índice de Gini. Definindo-se o vetor \vec{I} com as pontuações no IDH de cada país, e o vetor \vec{P} com as pontuações PISA destes mesmos países, busca-se a seguinte relação.

$$\vec{I}\alpha + \beta = \vec{P}$$

Tal equação pode ser modificada construindo-se uma matriz A , formada por duas colunas: a primeira com o vetor \vec{I} , e a segunda com um vetor de mesma dimensão apenas com o número 1. Assim, obtém-se a seguinte equação.

$$A \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \vec{P}$$

Basta, então, solucionar o sistema linear e definir α e β . Sabemos, porém, que a matriz A não é invertível. Conforme visto no capítulo 2, uma possibilidade é recorrer à pseudoinversa de A . Assim, resolve-se o sistema linear com erro quadrático mínimo.

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = A^+ \vec{P}$$

Dessa forma são conhecidos os coeficientes α e β da reta que melhor se ajusta aos pontos em cada caso. Na relação entre PISA e IDH, a reta é a seguinte:

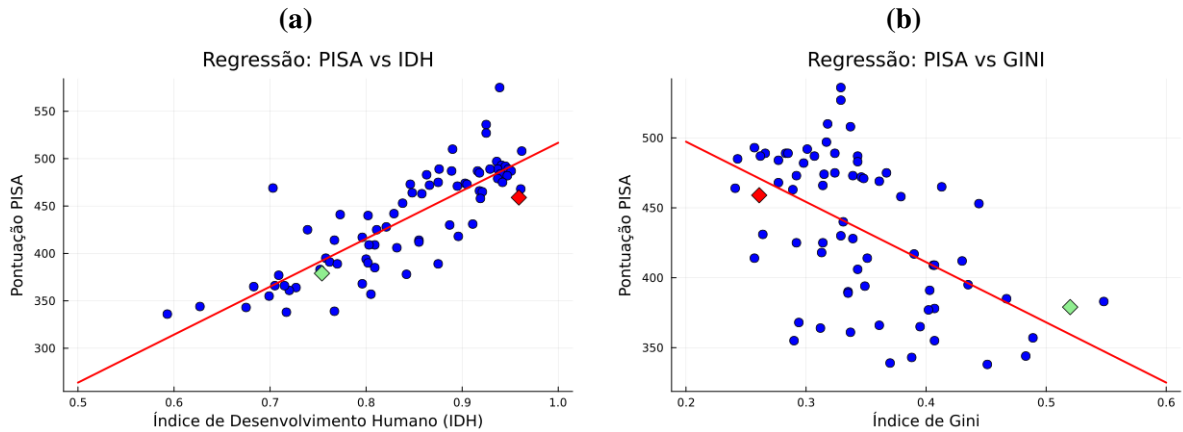
$$y = 506,073x + 10,6915 \quad (34)$$

Já na relação entre PISA e Gini, a reta é:

$$y = -430,716x + 583,44 \quad (35)$$

Pode-se agora adicionar tais retas ao gráfico de dispersão, obtendo o disposto na Figura 10. Novamente Brasil e Islândia estão destacados.

Figura 10 – (a) Relação entre IDH e pontuação PISA; (b) relação entre Índice de Gini e pontuação PISA



Fonte: OCDE (2023a); PNUD (2023); World Bank Group (2025a).

Vale lembrar que uma regressão não expressa relação causal entre duas variáveis. Não é possível afirmar com total certeza que um aumento no IDH causará aumento na pontuação PISA. Da mesma forma, não pode-se dar certeza de que, caso o Brasil diminua sua desigualdade, também obterá melhores índices educacionais. A reta encontrada relaciona os índices, e generaliza um padrão de comportamento deles. Seria necessário um estudo mais aprofundado, em conjunto com diversas outras áreas do conhecimento, para expressar, cientificamente, uma conexão entre desenvolvimento, desigualdade e desempenho educacional.

Tendo isso em mente, mantém-se o interesse em determinar um modelo para a melhoria do Brasil nas avaliações PISA. No resultado de 2022, a média da OCDE foi de 472 pontos, nota que pode ser considerada como meta para avaliações seguintes. Se considerarmos as retas de

regressão como relações válidas entre os índices estudados, então elas podem indicar caminhos para o Brasil melhorar sua nota no PISA.

No caso da relação entre IDH e PISA, considerando 34, tem-se:

$$\begin{aligned}472 &= 506,073x + 10,6915 \\461,3085 &= 506,073x \\x &\approx 0,911\end{aligned}$$

Isso significa que, se a reta de regressão representar a realidade, o Brasil poderá alcançar a média da OCDE se tiver IDH aproximado a 0,911. Isso é equivalente ao IDH atual da Espanha.

O mesmo cálculo pode ser feito para o Índice de Gini. Considerando 35, tem-se:

$$\begin{aligned}472 &= -430,716x + 583,44 \\-111,44 &= -430,716x \\x &\approx 0,258\end{aligned}$$

Assim, o Brasil poderá alcançar a média da OCDE se tiver o Índice de Gini igual a 0,258. Isso é aproximado com o valor obtido pelos Países Baixos.

3.4.1 Coeficiente de correlação de Pearson nas regressões

A partir da Figura 9, percebe-se a possibilidade de fazer generalizações quanto ao comportamento das variáveis envolvidas. Ainda assim, é notável a diferença entre o que vê-se ocorrer na relação entre PISA e IDH e na relação entre PISA e Gini. É clara a percepção de que os dados na figura (a) estão mais “unidos”, enquanto na figura (b) estão mais espalhados.

Na estatística, a noção do quão próximos estão os pontos de uma dispersão, e portanto o quão próximos estes mesmos pontos estarão da reta de regressão, é traduzida numericamente por meio do coeficiente de correlação, ou ainda coeficiente de Pearson. Neto (2006, p. 195) expõe a fórmula geral deste coeficiente da seguinte forma:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (36)$$

O coeficiente sempre está no intervalo entre -1 e 1 . Um resultado igual a 1 representa correlação perfeita positiva entre as grandezas; assim, se os valores de x_i aumentam, os valores de y_i aumentam sempre na mesma proporção, de forma que, no gráfico de dispersão, todos os

pontos pertenceriam à reta de regressão. O mesmo ocorre para um resultado igual a -1 , que representa correlação perfeita negativa. Um coeficiente igual a 0 mostra que não há correlação entre as grandezas.

Utilizando a fórmula 36, será feito o cálculo do coeficiente de Pearson, primeiro para a relação entre PISA e IDH. Em um primeiro momento é necessário determinar a média de cada índice.

$$\text{Média de pontuações PISA} = 434,986$$

$$\text{Média de índices de desenvolvimento humano} = 0,838$$

$$\text{Média de índices de Gini} = 0,346$$

Com estes valores, considerando x_i os i valores do vetor com resultados de IDH e y_i os i valores do vetor de pontuações PISA, encontra-se o coeficiente de Pearson para estes índices.

$$r = \frac{\sum_{i=1}^n (x_i - 0,838)(y_i - 434,986)}{\sqrt{\sum_{i=1}^n (x_i - 0,838)^2 \sum_{i=1}^n (y_i - 434,986)^2}} \approx \frac{306.942}{370.531} \approx 0,8283$$

Tal valor representa uma forte correlação positiva entre os índices. Ou seja, se o valor para um índice aumenta, é muito provável que o outro também aumente.

De forma análoga, considerando w_i como os i valores do vetor com índices de Gini e y_i definido como anteriormente, encontra-se o coeficiente de Pearson para estes índices.

$$r = \frac{\sum_{i=1}^n (w_i - 0,346)(y_i - 434,986)}{\sqrt{\sum_{i=1}^n (w_i - 0,346)^2 \sum_{i=1}^n (y_i - 434,986)^2}} \approx \frac{-132.302}{244.367} \approx -0,5414$$

Tal resultado confirma nossa hipótese. A correlação negativa para os valores de índices de Gini e pontuações PISA possui módulo menor do que o da correlação entre IDH e PISA. Assim representamos numericamente a percepção visual de “agrupamento” e “distanciamento” entre os pontos marcados no gráfico de dispersão.

Um coeficiente de Pearson de maior módulo também indica maior credibilidade para regressão linear criada, e a probabilidade de que novos dados sigam o padrão de comportamento descrito pela reta é maior. Dessa forma, o exercício imaginativo realizado no final da seção 2.4 pode ser visto com outros olhos. Como o coeficiente de correlação entre IDH e PISA é maior, então a conclusão de que o Brasil deve alcançar IDH 0,911 para atingir a meta da OCDE é mais

crível do que a determinação de que o Brasil deve alcançar índice de Gini igual a 0,258 para conquistar o mesmo feito.

Um possível fator que sustenta uma maior correlação entre IDH e PISA é a presença de critérios educacionais na formação do primeiro índice. Assim, o resultado obtido no IDH é fortemente influenciado no nível de educação do país, o que reflete também no PISA. Como o Índice de Gini considera apenas renda (que pode ou não ser influenciada pela educação), a correlação entre os índices é menor.

4 MÉTODO DE MASSEY PARA RANQUEAMENTO

Kenneth Massey criou, ao longo de sua vida, diversos métodos de ranqueamento. O que será aqui descrito foi inicialmente concebido por ele para times universitários de futebol americano, e publicado em sua tese (1997).

Massey baseia-se na concepção de que o saldo de pontos no confronto entre dois times deve refletir a diferença de ratings entre eles. Ou seja, se r_i e r_j são os ratings dos times i e j , a diferença y_k das suas pontuações após um jogo k deveria satisfazer, idealmente:

$$r_i - r_j = y_k \quad (37)$$

Esta lógica se mantém para todos os m jogos ocorridos entre n times em uma competição. Dessa forma, é possível estabelecer uma equação como (37) para cada partida da competição, e a resolução do sistema de equações formado poderá trazer os ratings de cada time e, consequentemente, suas classificações.

Na sequência deste capítulo, será utilizada como base teórica a obra de Langville e Meyer (2012), além da própria dissertação de Massey (1997).

4.1 O método no caso todos contra todos

O primeiro caso a ser considerado para o desenvolvimento do método de Massey se dá quando todos os times do campeonato enfrentam todos os demais uma única vez. Assim, é definida uma matriz X de m linhas e n colunas para refletir os resultados dos m jogos entre n times. Neste trabalho, utilizaremos como exemplo os jogos de handebol das Olimpíadas de Verão de Paris 2024, especificamente do grupo B da modalidade feminina, no qual estava presente o Brasil. A Tabela 9 traz os 15 confrontos entre os 6 times presentes no grupo.

Tabela 9 – Olimpíadas de Verão Paris 2024: resultados do grupo B do handebol feminino

Angola	19 x 30	Brasil
Angola	26 x 21	Espanha
Angola	24 x 38	França
Angola	31 x 31	Hungria
Angola	31 x 34	Países Baixos
Brasil	29 x 18	Espanha
Brasil	20 x 26	França
Brasil	24 x 25	Hungria
Brasil	24 x 31	Países Baixos
Espanha	24 x 32	França
Espanha	24 x 27	Hungria
Espanha	24 x 29	Países Baixos
França	31 x 28	Hungria
França	32 x 28	Países Baixos
Hungria	26 x 30	Países Baixos

Fonte: International Handball Federation (2024).

A partir desta tabela é criada a matriz X , com 15 linhas e 6 colunas. Cada coluna representa um país, em ordem alfabética (Angola, Brasil, Espanha, França, Hungria e Países Baixos), e cada linha representa uma partida entre duas equipes, atribuindo-se 1 para o vencedor e -1 para o perdedor.

$$X = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 \\ -1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

A esta matriz X associa-se o vetor y com 15 linhas e 1 coluna. Em cada linha deste vetor está exposta a diferença de pontuação entre o time vencedor e o time perdedor na partida representada pela linha de mesmo número na matriz X .

$$y^T = \left[11 \ 5 \ 14 \ 0 \ 3 \ 11 \ 6 \ 1 \ 7 \ 8 \ 3 \ 5 \ 3 \ 4 \ 4 \right] \quad (38)$$

Ou seja, a primeira linha da matriz X possui -1 na primeira coluna e 1 na segunda coluna, o que significa que na partida entre Brasil e Angola, o Brasil saiu como vencedor. Já o valor 11, da primeira linha do vetor y , representa a diferença entre os gols marcados pelo Brasil, 30, e pela Angola, 19.

Destaca-se que, na partida entre Angola e Hungria, ocorreu um empate. Nesse caso, foram utilizados na matriz X os valores 1 para Angola e -1 para a Hungria, enquanto no vetor y a partida foi representada com saldo 0. Em outras palavras, o empate está sendo representado

como uma vitória por 0 pontos de diferença. O mesmo poderia ter sido realizado atribuindo-se a vitória à Hungria e a derrota à Angola.

Com as matrizes X e y , monta-se um sistema linear do modo $Xr = y$. O propósito do método de Massey é determinar este vetor r , que declara os ratings de cada time, e assim definir suas classificações.

Porém, surge um problema: este sistema não possui solução. O sistema $Xr = y$ é sobre-determinado. Enquanto as matrizes X e y possuem 15 linhas, a matriz r possui apenas 6 linhas, que correspondem aos ratings das 6 seleções participantes. Nem sempre é possível encontrar um vetor solução da equação que não apresente erros na relação com a realidade.

Assim, seria mais justo reescrever (37) da seguinte forma:

$$r_i - r_j = y_k + e_k$$

Nesta equação, e_k ajusta o erro presente no resultado do jogo k . Assim, um desafio complementar do método é encontrar um vetor de ratings r que minimize o erro do sistema. Isso pode ser obtido pelo método de mínimos quadrados. Conforme estudado anteriormente neste trabalho, tal vetor pode ser encontrado pela solução da seguinte equação, também chamada de Equação Normal, ou Sistema Normal.

$$X^T X \hat{r} = X^T y$$

Chama-se a nova matriz obtida por $X^T X$ de M . Esta matriz poderia ter sido obtida sem cálculos, uma vez que apresenta na sua diagonal principal o número de jogos que cada seleção realizou, e os demais elementos são iguais a -1 . No exemplo explorado neste trabalho, M resulta em uma matriz 6×6 , com todos os elementos da diagonal principal iguais a 5. Assim:

$$X^T X = M = \begin{bmatrix} 5 & -1 & -1 & -1 & -1 & -1 \\ -1 & 5 & -1 & -1 & -1 & -1 \\ -1 & -1 & 5 & -1 & -1 & -1 \\ -1 & -1 & -1 & 5 & -1 & -1 \\ -1 & -1 & -1 & -1 & 5 & -1 \\ -1 & -1 & -1 & -1 & -1 & 5 \end{bmatrix}$$

No lado direito da equação, $X^T y$ é renomeado como p . Este vetor também pode ser obtido de forma direta, pois traduz, em cada linha, o saldo de gols no campeonato do time correspondente.

$$X^T y = p = \begin{bmatrix} -23 \\ 8 \\ -32 \\ 35 \\ -3 \\ 15 \end{bmatrix}$$

Assim, o Sistema Normal pode ser escrito como $M\hat{r} = p$, ou seja:

$$\begin{bmatrix} 5 & -1 & -1 & -1 & -1 & -1 \\ -1 & 5 & -1 & -1 & -1 & -1 \\ -1 & -1 & 5 & -1 & -1 & -1 \\ -1 & -1 & -1 & 5 & -1 & -1 \\ -1 & -1 & -1 & -1 & 5 & -1 \\ -1 & -1 & -1 & -1 & -1 & 5 \end{bmatrix} \hat{r} = \begin{bmatrix} -23 \\ 8 \\ -32 \\ 35 \\ -3 \\ 15 \end{bmatrix} \quad (39)$$

Note, no entanto, que M possui determinante igual a 0, de forma que suas colunas são linearmente dependentes. A matriz normal M não é invertível!

Como estudado no capítulo preliminar sobre inversas generalizadas, é possível encontrar soluções para o sistema $M\hat{r} = p$ por diversos métodos. Algumas soluções poderiam ser encontradas por meio das inversas generalizadas G_1 e G_2 , da matriz M , obtidas pelo algoritmo de Myers e Milton (1991):

$$r_1 = G_1 p = \begin{bmatrix} \frac{1}{3} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} & 0 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & 0 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -23 \\ 8 \\ -32 \\ 35 \\ -3 \\ 15 \end{bmatrix} = \begin{bmatrix} -\frac{19}{3} \\ -\frac{7}{6} \\ -\frac{47}{6} \\ \frac{10}{3} \\ -3 \\ 0 \end{bmatrix}$$

$$r_2 = G_2 p = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} \\ 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{6} \\ 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} -23 \\ 8 \\ -32 \\ 35 \\ -3 \\ 15 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{31}{6} \\ -\frac{3}{2} \\ \frac{29}{3} \\ \frac{10}{3} \\ \frac{19}{3} \end{bmatrix}$$

Estes vetores são soluções do sistema $M\hat{r} = p$ e, portanto, representam os ratings das equipes do exemplo. Dessa forma, obtêm-se as seguintes pontuações e seus respectivos rankings. É interessante destacar que as classificações são as mesmas para ambas soluções.

Tabela 10 – Rating encontrado para cada seleção a partir da inversa G_1

Seleção	Rating	Classificação
Angola	-6,33	5º lugar
Brasil	-2,33	3º lugar
Espanha	-7,833	6º lugar
França	3,33	1º lugar
Hungria	-3	4º lugar
Países Baixos	0	2º lugar

Fonte: O Autor.

Tabela 11 – Rating encontrado para cada seleção a partir da inversa G_2

Seleção	Rating	Classificação
Angola	0	5º lugar
Brasil	5,166	3º lugar
Espanha	-1,5	6º lugar
França	9,66	1º lugar
Hungria	3,33	4º lugar
Países Baixos	6,33	2º lugar

Fonte: O Autor.

Tanto r_1 quanto r_2 são soluções da Equação Normal, e por isso são soluções tais que a norma do erro para o sistema original é o menor possível. Porém, não são as soluções de norma euclidiana mínima. De fato, $\|\vec{r}_1\| \approx 11,08$ e $\|\vec{r}_2\| \approx 13,17$. Se buscamos uma solução com norma euclidiana mínima, é preciso utilizar uma inversa generalizada especial, a pseudoinversa.

Por meio da pseudoinversa M^+ , podemos encontrar o seguinte vetor \hat{r} :

$$\hat{r} = M^+p = \begin{bmatrix} \frac{5}{36} & -\frac{1}{36} & -\frac{1}{36} & -\frac{1}{36} & -\frac{1}{36} & -\frac{1}{36} \\ -\frac{1}{36} & \frac{5}{36} & -\frac{1}{36} & -\frac{1}{36} & -\frac{1}{36} & -\frac{1}{36} \\ -\frac{1}{36} & -\frac{1}{36} & \frac{5}{36} & -\frac{1}{36} & -\frac{1}{36} & -\frac{1}{36} \\ -\frac{1}{36} & -\frac{1}{36} & -\frac{1}{36} & \frac{5}{36} & -\frac{1}{36} & -\frac{1}{36} \\ -\frac{1}{36} & -\frac{1}{36} & -\frac{1}{36} & -\frac{1}{36} & \frac{5}{36} & -\frac{1}{36} \\ -\frac{1}{36} & -\frac{1}{36} & -\frac{1}{36} & -\frac{1}{36} & -\frac{1}{36} & \frac{5}{36} \end{bmatrix} \begin{bmatrix} -23 \\ 8 \\ -32 \\ 35 \\ -3 \\ 15 \end{bmatrix} = \begin{bmatrix} -\frac{23}{6} \\ \frac{4}{3} \\ -\frac{16}{3} \\ \frac{35}{6} \\ -\frac{1}{2} \\ \frac{5}{2} \end{bmatrix}$$

Uma vez que $\|r_1\| \approx 11,08$, $\|r_2\| \approx 13,17$ e $\|\vec{r}\| \approx 9,24$, o resultado encontrado pela pseudoinversa tem norma euclidiana menor do que as obtidas por outras inversas generalizadas, conforme esperado. As classificações dos países obtidas por tal solução do Sistema Normal estão na Tabela 12 que, apesar de constar com ratings diferentes, dispõe da mesma classificação obtida com as inversas generalizadas.

Tabela 12 – Rating encontrado para cada seleção a partir da pseudoinversa

Seleção	Rating	Classificação
Angola	-3,833	5º lugar
Brasil	1,33	3º lugar
Espanha	-5,33	6º lugar
França	5,833	1º lugar
Hungria	-0,5	4º lugar
Países Baixos	2,5	2º lugar

Fonte: O Autor.

Massey, em sua tese, porém, escolhe não utilizar a pseudoinversa para contornar a singularidade da matriz, mas soluciona este problema por meio de uma reparametrização, seguindo a sugestão de Myers e Milton (1991), conforme a argumentação do Exemplo 2.4. Substitui, então, os elementos da última linha de M por 1 e o elemento correspondente de p por 0. Assim, acrescenta a limitação de que a soma dos ratings das seleções seja igual a 0, e a dependência linear é eliminada.

Após tais adaptações, considera-se então para solução o sistema $\bar{M}\hat{r} = \bar{p}$, ou ainda:

$$\begin{bmatrix} 5 & -1 & -1 & -1 & -1 & -1 \\ -1 & 5 & -1 & -1 & -1 & -1 \\ -1 & -1 & 5 & -1 & -1 & -1 \\ -1 & -1 & -1 & 5 & -1 & -1 \\ -1 & -1 & -1 & -1 & 5 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} r_a \\ r_b \\ r_e \\ r_f \\ r_h \\ r_p \end{bmatrix} = \begin{bmatrix} -23 \\ 8 \\ -32 \\ 35 \\ -3 \\ 0 \end{bmatrix}$$

Como \bar{M} é invertível, vetor \hat{r} resultante é calculado a partir de $\hat{r} = (\bar{M})^{-1}\bar{p}$:

Tabela 13 – Rating encontrado para cada seleção usando o método de Massey

Seleção	Rating	Classificação
Angola	-3,833	5º lugar
Brasil	1,33	3º lugar
Espanha	-5,33	6º lugar
França	5,833	1º lugar
Hungria	-0,5	4º lugar
Países Baixos	2,5	2º lugar

Fonte: O Autor.

Este vetor \hat{r} representa, portanto, uma solução para o sistema $Xr = y$ com o menor erro quadrático. Ainda, este é o único vetor que apresenta a menor norma euclidiana.

Nota-se que os rankings encontrados pelos métodos matemáticos deste capítulo diferem do que realmente ocorreu nas olimpíadas de 2024. Enquanto o pódio deste trabalho é composto por França, Países Baixos e Brasil, a competição de Paris traz a Hungria em terceiro lugar, deixando o Brasil em quarto lugar, como a última equipe classificada para a próxima fase.

É interessante notar que existe uma relação entre as três soluções de mínimos quadrados encontradas, r_1 , r_2 e \hat{r} . Percebe-se, por exemplo, que os elementos de r_2 podem ser obtidos através de r_1 ; basta somar a cada um, uma quantidade de $\frac{19}{3}$. O mesmo acontece dos elementos de r_1 para \hat{r} , em que soma-se $\frac{5}{2}$, e de r_2 para \hat{r} , em que a cada elemento acrescenta-se $-\frac{23}{6}$. Podemos descrever uma operação como estas da seguinte forma:

$$\begin{bmatrix} -\frac{19}{3} \\ -\frac{7}{6} \\ -\frac{47}{6} \\ \frac{10}{3} \\ -3 \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{19}{3} \\ \frac{19}{3} \\ \frac{19}{3} \\ \frac{19}{3} \\ \frac{19}{3} \\ \frac{19}{3} \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{31}{6} \\ -\frac{3}{2} \\ \frac{29}{3} \\ \frac{10}{3} \\ \frac{19}{3} \end{bmatrix} \Rightarrow \begin{bmatrix} -\frac{19}{3} \\ -\frac{7}{6} \\ -\frac{47}{6} \\ \frac{10}{3} \\ -3 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ \frac{31}{6} \\ -\frac{3}{2} \\ \frac{29}{3} \\ \frac{10}{3} \\ \frac{19}{3} \end{bmatrix} = \frac{19}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Para todas as soluções de mínimos quadrados encontradas, verifica-se que se r_1 e r_2 são soluções do Sistema Normal (39), então $r_1 - r_2 = \lambda(1,1,1,1,\dots)$, para algum $\lambda \in \mathbb{R}$. De fato, esta propriedade será provada em geral na Proposição 4.1. É importante destacar que tal propriedade não foi encontrada em nenhuma das referências consultadas e sua demonstração é uma contribuição original deste trabalho.

Para demonstrar tal propriedade, primeiro é necessário considerar alguns lemas. Em primeiro lugar, o Lema 4.1 buscará uma generalização do formato das matrizes básicas do método de Massey, em que cada linha i representa um confronto entre dois times, representados nas j colunas. Considerando apenas um confronto por dupla de times, os valores 1 e -1 não podem estar presentes nas mesmas colunas em linhas diferentes.

Lema 4.1. *Seja X uma matriz de dimensões $\frac{j(j-1)}{2} \times j$, $j \in \mathbb{N}$, $j \geq 2$ tal que:*

- i. em cada linha i , todos os elementos são iguais a zero, a não ser em duas colunas, em que um elemento é igual a 1 e outro é igual a -1 (cada linha representa uma partida entre apenas duas equipes).*
- ii. cada linha i apresenta uma dupla diferentes de colunas que possuem os elementos 1 e -1 (há apenas um confronto entre duas equipes).*
- iii. cada coluna deve ter $j - 1$ elementos diferentes de 0 (cada equipe deve enfrentar todas as demais uma única vez).*

Nessas condições, o vetor $\vec{1} = (1, 1, \dots, 1) \in \mathbb{R}^j$, está no espaço nulo de X .

Pode-se compreender melhor tal resultado observando a matriz X do nosso caso concreto. Se é procurado o resultado do confronto entre Brasil e Angola, então há apenas uma linha em que os elementos da primeira e segunda colunas são diferentes de 0, que é a primeira linha. Em qualquer outra linha tem-se que ou o primeiro, ou o segundo, ou ambos os elementos das colunas são iguais a 0. Ainda, todas as colunas possuem 5 elementos diferentes de 0, uma vez que cada seleção participou de 5 partidas na fase de grupos da competição.

Se um vetor qualquer $\vec{v} = (v_1, v_2, v_3, \dots, v_n)$ está no espaço nulo de X , então $X\vec{v} = \vec{0}$. Com isto, pelo algoritmo da multiplicação de matrizes, multiplica-se cada linha i pelos elementos do vetor \vec{v} , e pela construção da matriz X , para qualquer dupla $k, p \in (1, \dots, n)$, $v_k - v_p = 0$. Assim, $v_k = v_p$, o que é equivalente a afirmar que $\vec{v} = \lambda \vec{1}$, em que $\vec{1} = (1, 1, \dots, 1)$ e $\lambda \in \mathbb{R}$.

Pode-se afirmar, portanto, que o espaço nulo de X é gerado pelo vetor $\vec{1}$ e, portanto, tem dimensão 1. O seguinte resultado permite afirmar que o espaço nulo da matriz $M = X^T X$ também tem essas características.

Lema 4.2. *$A^T A$ possui o mesmo espaço nulo que A . Assim, se $\vec{v} \neq 0$,*

$$A\vec{v} = \vec{0} \iff A^T A\vec{v} = \vec{0}$$

A implicação direta é evidente. Se $A\vec{v} = \vec{0}$, então pela associação da multiplicação, $A^T A\vec{v} = A^T \vec{0} = \vec{0}$.

Para provar a recíproca, assume-se que $A^T A\vec{v} = \vec{0}$ e a tese é de que $A\vec{v} = \vec{0}$. De fato:

$$\begin{aligned} A^T A\vec{v} = \vec{0} &\Rightarrow (A^T A\vec{v})^T = \vec{0}^T \\ (A\vec{v})^T A &= \vec{0}^T \Rightarrow (A\vec{v})^T A\vec{v} = \vec{0}^T \vec{v} \\ \|A\vec{v}\|^2 &= 0 \Rightarrow A\vec{v} = \vec{0} \end{aligned}$$

Com isto, verifica-se a equivalência. O Lema 4.2 é útil para determinar o posto de uma matriz A , pois por vezes torna-se mais produtivo observar o posto de $A^T A$.

Corolário 4.1. *O núcleo de $X^T X = M$ é $\lambda \vec{1}$.*

Pelo Lema 4.2, o espaço nulo de M e X são iguais. Por isso, o núcleo de M possui apenas vetores da forma $\lambda \vec{1}$, e a dimensão do espaço nulo de M é 1.

Lema 4.3. *Seja X uma matriz com a mesma descrição do Lema 4.1, então $X^T X = M$ é quadrada $n \times n$, simétrica e com posto $n - 1$.*

Pode-se provar o lema em três etapas. Em primeiro lugar, se X é $n \times m$, então por definição de multiplicação de matrizes, uma matriz $n \times m$ multiplicada por outra $m \times n$, como é o caso de X^T , terá dimensão $n \times n$. Em segundo lugar, sabe-se que uma matriz é simétrica se ela é igual à sua transposta. Como $(X^T X)^T = X^T X^{TT} = X^T X$, então a matriz M é simétrica. Por fim, como já foi mostrado que o espaço nulo de M tem dimensão 1, então pelo Teorema do Núcleo e da Imagem (Strang, 2013, p. 106), o seu posto deve ser $n - 1$.

Com a base deixada pelos lemas, segue a Proposição 4.1.

Proposição 4.1. *Sejam r_1 e r_2 duas soluções do sistema normal $Mr = p$. Então existe um $\lambda \in \mathbb{R}$ tal que*

$$r_1 = r_2 + \lambda \vec{1}$$

Demonstração. Veja que se multiplicarmos $r_1 - r_2$ por, M , obtemos o seguinte:

$$M(r_1 - r_2) = Mr_1 - Mr_2 = p - p = 0$$

Isso significa que o vetor $(r_1 - r_2)$ está no espaço-nulo de M . Mas foi provado que todos os vetores do espaço-nulo são da forma $\lambda \vec{1}$. Ou seja, $r_1 - r_2 = \lambda \vec{1}$ e $r_1 = r_2 + \lambda \vec{1}$. \square

Assim, todas as soluções do sistema são translações por meio do vetor descrito. Isso condiz com o que foi escrito por Strang: “Toda solução de $Ax = b$ é a soma de uma solução particular e uma solução de $Ax = 0$ ” (Strang, 2013, p. 82). Como neste contexto a dimensão do espaço-nulo é 1, a equação 40 descreve uma translação por meio do vetor $\vec{1}$.

$$x_{\text{completo}} = x_{\text{p (particular)}} + x_{\text{n (espaço nulo)}} \quad (40)$$

Com esta descoberta, foram descritas as ferramentas necessárias para explorar com mais detalhes determinados resultados dos capítulos anteriores. No Teorema 2.2, foi enunciado que o vetor solução de mínimos quadrados obtido pela pseudo inversa possui norma euclidiana mínima. No Teorema 4.1, será provado que, a solução obtida pelo método de Massey (reparametrização) tem a mesma propriedade. Primeiro, é interessante verificar que isto se confirma na matriz M do nosso problema concreto.

Tome \hat{r} o vetor obtido por meio do método de Massey. Sabe-se, por meio da Proposição 4.1, que todo vetor que também é solução do sistema é obtido por meio de uma translação pelo vetor $\vec{1}$. Então, qualquer outro vetor r solução do sistema será obtido por meio da equação $r = \hat{r} + \lambda\vec{1}$, para algum λ .

$$r = \hat{r} + \lambda\vec{1} = \begin{bmatrix} -\frac{23}{6} \\ \frac{4}{3} \\ -\frac{16}{3} \\ \frac{35}{6} \\ -\frac{1}{2} \\ \frac{5}{2} \end{bmatrix} + \lambda \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{23}{6} + \lambda \\ \frac{4}{3} + \lambda \\ -\frac{16}{3} + \lambda \\ \frac{35}{6} + \lambda \\ -\frac{1}{2} + \lambda \\ \frac{5}{2} + \lambda \end{bmatrix}$$

Assim, a norma euclidiana elevada ao quadrado $\|\vec{r}\|^2$ corresponde ao seguinte cálculo.

$$\|\vec{r}\|^2 = \left(-\frac{23}{6} + \lambda\right)^2 + \left(\frac{4}{3} + \lambda\right)^2 + \left(-\frac{16}{3} + \lambda\right)^2 + \left(\frac{35}{6} + \lambda\right)^2 + \left(-\frac{1}{2} + \lambda\right)^2 + \left(\frac{5}{2} + \lambda\right)^2$$

Para encontrar o mínimo da função $f(\lambda) = \|\vec{r}\|^2$, vamos igualar a derivada da função a zero e resolver para λ .

$$\begin{aligned} \frac{df}{d\lambda} &= 2 \left[\left(-\frac{23}{6} + \lambda\right) + \left(\frac{4}{3} + \lambda\right) + \left(-\frac{16}{3} + \lambda\right) + \left(\frac{35}{6} + \lambda\right) + \left(-\frac{1}{2} + \lambda\right) + \left(\frac{5}{2} + \lambda\right) \right] = 0 \\ &= 2 \left[\left(-\frac{23}{6} + \frac{4}{3} - \frac{16}{3} + \frac{35}{6} - \frac{1}{2} + \frac{5}{2}\right) + 6\lambda \right] = 0 \end{aligned}$$

Como no método de Massey, a reparametrização da matriz M inclui a restrição de que a soma de todos os ratings é igual a 0, então a derivação pode ser assim simplificada e tem-se que

$$\frac{df}{d\lambda} = 2[0 + 6\lambda] = 12\lambda = 0 \Rightarrow \lambda = 0$$

Determinemos também a sua segunda derivada.

$$\frac{d^2f}{d\lambda^2} = 12 > 0$$

Com isto entende-se que $\lambda = 0$ é, realmente, ponto de mínimo da função. Isso significa que, neste caso, a solução obtida pelo método de Massey realmente tem norma euclidiana mínima. O seguinte teorema apresenta o caso geral.

Teorema 4.1. *Sejam \hat{r} e \vec{r}_2 soluções do sistema $Mr = p$, tais que \hat{r} foi obtida a partir do método de Massey, e $\vec{r}_2 = \hat{r} + \lambda \vec{1}$. Se $\|\vec{r}_2\|^2 = f(\lambda)$, então $f(\lambda)$ atinge valor mínimo em $\lambda = 0$.*

Demonstração. Parte-se do fato de que $\vec{r}_2 = \hat{r} + \lambda \vec{1}$ e considere $\hat{r} = (r_1, r_2, \dots, r_n)$. Então a norma euclidiana ao quadrado de \vec{r}_2 deve ser:

$$\begin{aligned}\|\vec{r}_2\|^2 &= (r_1 + \lambda)^2 + (r_2 + \lambda)^2 + \dots + (r_n + \lambda)^2 \\ \|\vec{r}_2\|^2 &= r_1^2 + 2r_1\lambda + \lambda^2 + r_2^2 + 2r_2\lambda + \lambda^2 + \dots + r_n^2 + 2r_n\lambda + \lambda^2 \\ \|\vec{r}_2\|^2 &= n\lambda^2 + r_1^2 + r_2^2 + \dots + r_n^2 + 2\lambda(r_1 + r_2 + \dots + r_n)\end{aligned}$$

Novamente é usado o fato de que, no método de Massey, a reparametrização da matriz M acrescenta a restrição de que a soma de todos os ratings é igual a 0. Sendo assim:

$$\|\vec{r}_2\|^2 = n\lambda^2 + r_1^2 + r_2^2 + \dots + r_n^2 \quad (41)$$

De fato, todas as parcelas da soma são positivas, e a expressão atinge valor mínimo quando $\lambda = 0$. Com isso, a conclusão é que o resultado obtido pelo método de Massey possui norma euclidiana mínima. \square

Já foi visto, no Teorema 2.2, que a solução de um sistema linear que possui norma euclidiana mínima é a obtida pela pseudoinversa. Pode ser feito, então, o questionamento: a partir de dois vetores, soluções do sistema $Mr = p$ estudado, tais que os dois possuem mesma norma, o que pode-se afirmar a respeito destes vetores? Eles são iguais?

Apontar que dois vetores possuem a mesma norma não necessariamente implica em igualdade entre eles. Nas explorações deste capítulo, verificou-se que todas as soluções do sistema são relacionadas por meio de translações através do vetor $\vec{1}$. A Equação 41 mostra que a relação entre a norma de dois vetores é uma função quadrática na variável λ . Pode-se, portanto, interpretar que o gráfico das possíveis normas de r_2 é uma parábola, o que evidencia que é possível encontrar dois vetores de mesma norma mas com entradas diferentes: basta tomar valores de λ opostos (λ e $-\lambda$ possuem o mesmo quadrado). Porém, tanto o resultado da pseudoinversa quanto o obtido pela reparametrização são mínimos da função. O vértice da parábola obtida pela função quadrática é único, obtido quando $\lambda = 0$. Isso significa que só existe um vetor com norma euclidiana mínima. Como foram encontradas duas formas de obter vetores com tal propriedade, pode-se sim afirmar que os dois vetores são iguais.

Afirma-se, portanto, que o vetor de ratings será o mesmo escolhendo-se realizar a reparametrização de Massey ou calculando-se por meio da pseudoinversa. Justificam-se assim os resultados idênticos encontrados no nosso caso concreto.

4.1.1 O ranking gerado a partir do método de Massey

As conclusões obtidas até aqui ajudam na obtenção do melhor vetor possível para os ratings das seleções. Porém, uma análise mais cuidadosa dos resultados revela que o método de Massey, independentemente dos ratings obtidos, cria sempre o mesmo ranking. Mais do que isso, o ranking é o mesmo obtido apenas observando o saldo de gols do campeonato.

Tomemos o nosso caso concreto. Neste capítulo obtiveram-se três vetores de ratings: r_1 , r_2 e \hat{r} . O ranking obtido por meio destes três ranking revela sempre a mesma ordem: França em primeiro lugar, seguida por Países Baixos, Brasil, Hungria, Angola, e Espanha em último lugar.

$$\begin{aligned} \vec{r}_1 &= \left(-\frac{19}{3}, -\frac{7}{6}, -\frac{47}{6}, \frac{10}{3}, -3, 0 \right) \\ \vec{r}_2 &= \left(0, \frac{31}{6}, -\frac{3}{2}, \frac{29}{3}, \frac{10}{3}, \frac{19}{3} \right) \\ \vec{\hat{r}} &= \left(-\frac{23}{6}, \frac{4}{3}, -\frac{16}{3}, \frac{35}{6}, -\frac{1}{2}, \frac{5}{2} \right) \end{aligned}$$

A Proposição 4.1 justifica esta propriedade. Todos os vetores de ratings são translações por $\lambda \vec{1}$. Então se em \vec{r}_1 verifica-se que $v_k > v_p$ para quaisquer $k, p \in (1, \dots, n)$, então deve ser verdade que, em \vec{r}_2 verifica-se que $v_k + \lambda \vec{1} > v_p + \lambda \vec{1}$. De fato, todos os ratings produzem a mesma ordenação entre as seleções representadas.

A este resultado acrescenta-se que o ranking é o mesmo observado para o vetor com o saldo de gols de cada equipe no campeonato. Veja que o vetor com o saldo de gols já é representado a partir de $X^T y = p$. Isso pois cada linha de X^T representa as vitórias (com 1) e derrotas (com -1) de cada seleção, e tal linha é multiplicada por y , que traz a diferença de pontuação entre o time vencedor e o perdedor em cada jogo.

Perceba agora que $p = 6\vec{\hat{r}}$.

$$p = \begin{bmatrix} -23 \\ 8 \\ -32 \\ 35 \\ -3 \\ 15 \end{bmatrix} = \begin{bmatrix} -\frac{23}{6} \times 6 \\ \frac{4}{3} \times 6 \\ -\frac{16}{3} \times 6 \\ \frac{35}{6} \times 6 \\ -\frac{1}{2} \times 6 \\ \frac{5}{2} \times 6 \end{bmatrix} = 6 \begin{bmatrix} -\frac{23}{6} \\ \frac{4}{3} \\ -\frac{16}{3} \\ \frac{35}{6} \\ -\frac{1}{2} \\ \frac{5}{2} \end{bmatrix} = 6\vec{\hat{r}}$$

Assim, a ordem observada em qualquer rating é a mesma observada em $\vec{\hat{r}}$, que se repete em p . De fato, como $M\vec{\hat{r}} = p$, então $M\vec{\hat{r}} = 6\vec{\hat{r}}$, e $\vec{\hat{r}}$ é autovetor e 6 autovalor associado a matriz M .

Corolário 4.2. Dada a matriz M $n \times n$ e o vetor p obtidos no sistema normal, e \vec{r} a solução do sistema $Mr = p$ obtida por meio do método de Massey, tem-se que $p = n\vec{r}$.

Outra forma de perceber tal resultado é verificando que o rating de cada time, ao final do campeonato, pode ser obtido dividindo-se o saldo de gols pelo número de times. O trabalho de Franceschet, Bozzo e Vidoni (2017) é usado aqui como referência para provar tal veracidade. Perceba que é possível decompor a matriz M da seguinte forma, sendo D uma matriz diagonal com o número de jogos de cada time, e A uma matriz na qual cada termo i,j representa o número de jogos realizados entre os times i e j .

$$M = D - A$$

No caso do ranqueamento dos times do handebol feminino das olimpíadas de 2024, conforme visto no início deste capítulo, tal decomposição se daria da seguinte forma.

$$\begin{bmatrix} 5 & -1 & -1 & -1 & -1 & -1 \\ -1 & 5 & -1 & -1 & -1 & -1 \\ -1 & -1 & 5 & -1 & -1 & -1 \\ -1 & -1 & -1 & 5 & -1 & -1 \\ -1 & -1 & -1 & -1 & 5 & -1 \\ -1 & -1 & -1 & -1 & -1 & 5 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 5 \end{bmatrix} - \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Multiplicando a equação pelo vetor r , e lembrando que $Mr = p$, tem-se que:

$$Dr - Ar = p \Rightarrow r = D^{-1}Ar + D^{-1}p$$

Isso significa que, para cada time i :

$$r_i = \frac{1}{D_{ii}} \sum_j^n A_{ij}r_j + \frac{p_i}{D_{ii}}$$

Levando em conta que, no método de Massey, a soma de todos os ratings é igual a 0, obtemos o seguinte:

$$r_i = -\frac{r_i}{n-1} + \frac{p_i}{n-1} \Rightarrow r_i = \frac{p_i}{n}$$

Com tal resultado, chega-se à conclusão de que o método de Massey não é tão útil para determinar rankings esportivos: basta observar o saldo de gols de cada time no campeonato, e o mesmo ranking pode ser encontrado. Porém, ainda é um método interessante para atribuir uma nota, um rating, para cada time. Alguns esforços têm sido feitos no sentido de otimizar o método de Massey. Franceschet, Bozzo e Vidoni (2017) sugerem uma forma de temporalização

do método de forma a considerar o “momento” de cada time, exatamente no tempo em que a partida é jogada.

4.2 Considerações sobre a matriz X com número de jogos diferentes

Até o momento, o desenvolvimento do método de Massey foi aplicado na fase de grupos dos jogos de handebol feminino das Olimpíadas de Paris, de 2024. Foi analisado apenas o grupo B, que continha o Brasil. A observação apenas da fase de grupos traz a vantagem de considerar-se que todos os times enfrentam todos os restantes, e cada confronto ocorre uma única vez. Porém, o campeonato não terminou por aí, e mais jogos foram realizados. Além dos confrontos do grupo A, as quartas de final, semifinal e final trouxeram um desequilíbrio para a quantidade de jogos realizados por cada seleção. Enquanto os que ficaram na fase de grupos fizeram 5 jogos, os times que ficaram entre primeiro e quarto lugar fizeram 8 jogos. Isso traz problemas para o método de ranqueamento de Massey, como descrito no caso abaixo.

Da mesma forma, o arremessador Jarrod Parker teve uma média de corridas ganhas de 0.00 durante a temporada de 2011 pelo Arizona Diamondbacks. Embora isso seja impressionante, ele enfrentou apenas 22 rebatedores para alcançar essa média. Provavelmente há dezenas de arremessadores da liga principal que enfrentaram 22 rebatedores seguidos sem permitir uma corrida ganha. Parker teve uma média abaixo de qualquer arremessador vencedor do prêmio Cy Young. No entanto, essa estatística de ERA, que é frequentemente usada para comparar arremessadores, não é tão significativa neste caso (Charlier et al., 2014, p. 15).

No caso descrito, evidencia-se que é mais fácil manter estatísticas impressionantes em curtos períodos de tempo. No contexto abordado neste problema, também considera-se mais fácil manter um saldo de gols maior quando se participa de mais partidas. Assim, a diferença no número de partidas representa um grande desafio no método de Massey. Nesta seção, portanto, serão realizados experimentos visando a uma melhor compreensão do que ocorre no método caso equipes tenham número de jogos diferentes em um campeonato. Como o intuito é apenas exploratório, não serão realizadas demonstrações.

É útil exemplificar esta dificuldade a partir do exemplo concreto. Considere todos os jogos realizados pelas seleções, contando com quartas de final, semifinal, disputa de 3º lugar e final. Constrói-se a matriz X da mesma forma que anteriormente. Cada coluna representa uma seleção feminina de handebol, respectivamente: Angola, Brasil, Espanha, França, Hungria, Países Baixos, Alemanha, Coreia do Sul, Dinamarca, Eslovênia, Noruega e Suécia. Em cada linha está representado o confronto entre duas seleções, em que a vencedora recebe o valor 1 e o perdedor recebe o valor -1 .

$$X = \left[\begin{array}{cccccc|cccccccc}
 -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 \\
 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\
 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\
 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\
 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0
 \end{array} \right] \tag{42}$$

Define-se também o vetor y com os saldos de gols de cada partida, na ordem estabelecida pelas linhas da matriz X .

$$y^T = \begin{bmatrix} 11 & 5 & 14 & 0 & 3 & 11 & 6 & 1 & 7 & 8 & 3 & 5 & 3 & 4 & 4 & 1 & 3 & 19 & 1 \\ 12 & 7 & 6 & 6 & 8 & 8 & 9 & 2 & 7 & 4 & 4 & 4 & 3 & 4 & 17 & 3 & 4 & 5 & 8 \end{bmatrix}$$

Pode-se encontrar as soluções de mínimos quadrados a partir do Sistema Normal. Seguem a matriz $M = X^T X$ e o vetor $p = X^T y$:

$$X^T X = M = \begin{bmatrix} 5 & -1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 6 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & -1 & 0 \\ -1 & -1 & 5 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & -1 & 8 & -1 & -1 & -1 & 0 & 0 & 0 & -1 & -1 \\ -1 & -1 & -1 & -1 & 6 & -1 & 0 & 0 & 0 & 0 & 0 & -1 \\ -1 & -1 & -1 & -1 & -1 & 6 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 6 & -1 & -1 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 5 & -1 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & 8 & -1 & -2 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & 5 & -1 & -1 \\ 0 & -1 & 0 & -1 & 0 & 0 & -1 & -1 & -2 & -1 & 8 & -1 \\ 0 & 0 & 0 & -1 & -1 & 0 & -1 & -1 & -2 & -1 & -1 & 8 \end{bmatrix}$$

$$(X^T y)^T = p^T = \begin{bmatrix} -23 & -9 & -32 & 33 & -7 & 11 & -1 & -26 & 15 & -31 & 59 & 11 \end{bmatrix}$$

Perceba que na matriz M aqui definida os termos da diagonal não são todos iguais. Isso porque a diagonal representa o número de jogos que a seleção nesta linha representada jogou, em que o mínimo é 5 e o máximo é 8. Perceba também que, além das entradas 0 e -1 encontradas nos exemplos anteriores, aqui também verificamos a entrada -2 , que representa que houveram jogos repetidos no campeonato. De fato, Dinamarca e Noruega enfrentaram-se na fase de grupos e na semifinal. Também Dinamarca e Suécia enfrentaram-se na fase de grupos e na disputa da medalha de bronze.

Conforme o discutido no capítulo anterior, pode-se encontrar um vetor solução da equação $Mr = p$ a partir da pseudoinversa de M . Segue a ordem estabelecida a partir desta resolução.

Tabela 14 – Rating encontrado para todas as seleções das Olimpíadas

Seleção	Rating	Classificação
Angola	-6,48	11º lugar
Brasil	-2,27	8º lugar
Espanha	-7,98	12º lugar
França	3,59	4º lugar
Hungria	-2,75	10º lugar
Países Baixos	-0,01	6º lugar
Alemanha	2,64	5º lugar
Coreia	-1,68	7º lugar
Dinamarca	4,83	2º lugar
Eslovênia	-2,51	9º lugar
Noruega	9,00	1º lugar
Suécia	3,62	3º lugar

Fonte: O Autor.

Perceba que o pódio formado pelo método de Massey é diferente do observado nas Olimpíadas. No evento, os três primeiros lugares foram Noruega, França e Dinamarca, enquanto nos cálculos realizados obteve-se Noruega, Dinamarca e Suécia, com França apenas em quarto lugar.

É interessante, também, perceber que o ranking encontrado não condiz com a divisão do saldo de gols pelo número de seleções. Veja que:

$$\frac{p^T}{12} = \begin{bmatrix} -1,91 & -0,75 & -2,66 & 2,75 & -0,58 & 0,91 & -0,08 & -2,16 & 1,25 & -2,58 & 4,91 & 0,91 \end{bmatrix}$$

Tal operação produz ratings que não condizem com o obtido pela pseudoinversa. Além disso, o ranking aqui obtido também seria diferente do que obtivemos pelo método. Assim, chega-se à conclusão de que a divisão do saldo de gols pelo número de times só resulta no rating apropriado dos times pelo método de Massey quando todos os times realizam o mesmo número de jogos. Neste caso, como o número de jogos é diferente, não é possível tentar esta simplificação do método. Isso justifica-se com base em mudança das matrizes D e A descritas na prova de Franceschet, Bozzo e Vidoni (2017), e explorada na seção anterior.

Ainda com tal descoberta, surge uma pergunta: o que aconteceria ao condensar os jogos repetidos em um só, juntando também os saldos de gols de todos esses jogos? Juntando os jogos realizados e determinando o vencedor por meio do saldo de gols, o rating e ranking das seleções serão os mesmos? Será, então, possível realizar a divisão dos saldos pelo número de seleções? Será realizado mais um experimento neste sentido, em uma verificação empírica destes questionamentos. Porém, o caso que trabalhamos até aqui, de handebol feminino nas olimpíadas, não possui muitos jogos repetidos: apenas duas duplas de seleções se encontraram duas vezes no campeonato.

Foi eleito pelo autor, portanto, outro esporte em que acontecem mais confrontos entre os mesmos elementos: o tênis. Este esporte tem como associação organizadora o ATP (Association

of Tennis Professionals), que possui seu próprio ranking, baseado em um sistema de pontos atribuídos a posições conquistadas em diferentes campeonatos. No experimento que será realizado, serão considerados os primeiros 5 lugares no ranking da associação no dia 1º de dezembro de 2024, conforme a Tabela 15.

Tabela 15 – Ranking da ATP em 01/12/2024

RANK	JOGADOR	PONTUAÇÃO
1	Jannik Sinner	11830
2	Alexander Zverev	7915
3	Carlos Alcaraz	7010
4	Taylor Fritz	5100
5	Daniil Medvedev	5030

Fonte: Association of Tennis Professionals (2024).

Por meio do site da ATP, pode-se encontrar os resultados dos jogos que geraram tal ranking. Serão considerados tais jogos, realizados entre os 5 primeiros jogadores até o dia 1º de dezembro de 2024, para montar as matrizes X e y . Esclarece-se que o saldo de gols será considerado em relação à quantidade de sets ganhos pelos jogadores, não em relação aos pontos que marcaram dentro desses sets.

A matriz X formada possui 90 linhas e 5 colunas, representando 90 jogos entre os 5 jogadores. O vetor y também possuía 90 linhas e 1 coluna. Devido às suas dimensões, grandes demais para este trabalho, X e y não serão escritas explicitamente. A partir destes, monta-se a seguinte matriz M e o vetor p , obtidos a partir da multiplicação pela transposta de X à esquerda.

$$M = \begin{bmatrix} 36 & -6 & -10 & -5 & -15 \\ -6 & 48 & -11 & -12 & -19 \\ -10 & -11 & 31 & -2 & -8 \\ -5 & -12 & -2 & 21 & -2 \\ -15 & -19 & -8 & -2 & 44 \end{bmatrix}$$

$$p = \begin{bmatrix} 7 \\ 1 \\ 9 \\ -12 \\ -5 \end{bmatrix}$$

O cálculo do rating de cada jogador com o uso da pseudoinversa de M está descrito na Tabela 16

Tabela 16 – Rating encontrado para os jogadores da ATP

JOGADOR	RATING	CLASSIFICAÇÃO
Jannik Sinner	0,205	2º lugar
Alexander Zverev	-0,010	3º lugar
Carlos Alcaraz	0,317	1º lugar
Taylor Fritz	-0,499	5º lugar
Daniil Medvedev	-0,013	4º lugar

Fonte: O Autor.

Percebe-se que a ordem encontrada foi diferente da ordem estabelecida pela ATP. Isso se dá, provavelmente, pelo método de ranqueamento da associação considerar pontuações diferentes para cada campeonato. Em Massey, todos os jogos são considerados como equivalentes. Mas para a ATP, Grand Slams possuem pontuação diferente das ATP Finals, que por sua vez valem mais do que a United Cup, e assim por diante.

Resta ainda verificar se, condensando os jogos repetidos realizados entre os jogadores, o ranking mudará ou será o mesmo. Segue, portanto, a matriz X com a condensação de jogos repetidos.

$$X = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

$$y^T = [4 \ 4 \ 6 \ 3 \ 1 \ 2 \ 6 \ 4 \ 8 \ 0]$$

Por meio destas matrizes criamos M e p como segue:

$$M = X^T X = \begin{bmatrix} 4 & -1 & -1 & -1 & -1 \\ -1 & 4 & -1 & -1 & -1 \\ -1 & -1 & 4 & -1 & -1 \\ -1 & -1 & -1 & 4 & -1 \\ -1 & -1 & -1 & -1 & 4 \end{bmatrix}$$

$$p^T = (X^T y)^T = \begin{bmatrix} 1 & -3 & 15 & -8 & -5 \end{bmatrix}$$

Assim pode-se calcular o rating a partir dos jogos repetidos condensados a partir de $M^{-1}p$.

Tabela 17 – Rating para os jogadores da ATP com jogos repetidos condensados

JOGADOR	RATING	CLASSIFICAÇÃO
Jannik Sinner	0,2	2º lugar
Alexander Zverev	-0,6	3º lugar
Carlos Alcaraz	3	1º lugar
Taylor Fritz	-1,6	5º lugar
Daniil Medvedev	-1	4º lugar

Fonte: O Autor.

Os rankings verificados nas Tabelas 16 e 17 são iguais, apesar de os ratings serem diferentes. Porém, ambos diferem do ranking oficial da ATP, mostrando novamente que o método utilizado para ranqueamento impacta diretamente o resultado final.

4.3 Comentários adicionais

O método de Massey se baseia na lógica de que a diferença entre ratings de dois times é igual à diferença de pontuação no confronto entre os dois. Tal lógica leva a construção de um sistema linear sobredeterminado, no qual deve-se aceitar a presença de erros. Tal contratempo é solucionado com o uso do Sistema Normal, multiplicando-se ambos os lados do sistema original pela transposta da matriz X . O sistema resultante obtém respostas com erro quadrático mínimo. Com o desenvolvimento teórico dos capítulos anteriores, foi possível encontrar soluções com o uso de inversas generalizadas, em especial a pseudoinversa, que é especial por ter norma euclidiana mínima. A escolha de Massey em seu trabalho, porém, foi de realizar uma reparametrização que, como mostrado neste capítulo, resulta no mesmo vetor de norma mínima.

No decorrer do capítulo, também foram consideradas algumas propriedades das soluções encontradas. Verificando que os vetores soluções do sistema normal eram translações entre si, foi realizada uma prova original deste fato a partir de uma mais profunda caracterização das

matrizes X e M , seus espaços nulos e postos. A prova deste teorema permitiu explorar mais a fundo a razão pela qual encontramos a mesma resposta pelo algoritmo da pseudoinversa e pela reparametrização de Massey.

Foi mostrado, também, que a busca pelo melhor vetor de ratings neste ranqueamento não é tão interessante para a formulação de rankings; a ordenação dos times é a mesma observada no saldo de gols. Além disso, se em um campeonato todos os times se confrontam vezes iguais entre si, e o número de jogos é o mesmo, então para encontrar o melhor vetor de ratings basta dividir o vetor do saldo de gols pelo número de times.

Tendo em vista as limitações acima descritas, sugestões e modificações ainda hoje são feitas por autores para expandir as capacidades do método de Massey. Algumas dessas sugestões serão aqui listadas, sem desenvolvimento aprofundado, mas para apontar o que tem sido feito nesta teoria.

Uma sugestão, por exemplo, vem de Franceschet, Bozzo e Vidoni (2017), na ideia de adicionar considerar no ranqueamento os rankings do momento em que os times se encontraram.

Para um dado time i e tempo t , o método original de Massey classifica i de acordo com a diferença de pontos de i até o tempo t e as classificações dos times que i enfrentou até o tempo t . Note, no entanto, que a classificação de um time enfrentado j é calculada em relação ao tempo t , e não, como argumentamos que seria mais razoável, em relação ao tempo (possivelmente anterior) quando i e j se enfrentaram (Franceschet, Bozzo e Vidoni, 2017, p. 37).

A temporalização do ranqueamento seria interessante, por considerar o “momento” do time no momento do confronto. Outra ideia seria considerar o peso do momento do campeonato, e não do time. Por exemplo, a reta final de um campeonato, em que a posição em um pódio está sendo disputada, vale mais do que jogos intermediários em fase de grupos, em que a pressão por vezes é muito menos e mais administrável pelos jogadores. Jogos de muita pressão, clássicos e finais de campeonato devem ter um peso maior para o ranqueamento. Essa ideia é importante por ser utilizada em outro método de ranqueamento, o método Elo, que será discutido no próximo capítulo.

Na versão padrão dos métodos de Massey e Colley, um jogo no início da temporada contribui para a classificação de um time com o mesmo peso que um jogo no final da temporada. Isso pode ser um problema por várias razões. Um jogador estrela pode se machucar no meio da temporada ou um time pode amadurecer e melhorar ao longo do tempo. Além disso, parece especialmente valioso para jogos de torneio aumentar a classificação de um time que venceu seus últimos 10 jogos, mesmo que tenha tido um desempenho ruim no início da temporada. Isso é realizado atribuindo pesos aos jogos de acordo com a data em que foram jogados (Chartier et al., 2011, p. 5-6).

Por fim, um último contratempo descrito neste trabalho diz respeito à número diferente de jogos entre os times. Foi verificado que tal cenário cria condições desiguais para o ranque-

amento, uma vez que mais jogos podem criar mais oportunidades de bom saldo de gols, ou menos jogos podem criar melhores condições para manter boas estatísticas de aproveitamento. Uma sugestão para resolver este problema é proposta em Chartier et al. (2014).

Uma solução comum para esse problema de números díspares de jogos é simplesmente excluir atletas (ou equipes ou itens) com menos de um número mínimo de jogos. Embora os cortes mínimos sejam geralmente escolhidos com cuidado para a aplicação específica, essa abordagem é um pouco arbitrária. Além disso, uma classificação pode ser sensível ao corte escolhido. [...] Neste artigo, propomos uma alternativa à abordagem de corte mínimo para lidar com a questão do número desigual de jogos nos Métodos de Colley e Massey. Nossa abordagem utiliza um super-usuário, uma equipe fictícia (atleta ou item) que joga e vence todas as equipes reais que enfrenta. Forçamos as equipes a perder para o super-usuário em todos os casos, em vez de uma combinação de vitórias e derrotas, a fim de penalizar equipes que talvez estejam artificialmente infladas no ranking devido ao bom desempenho em apenas alguns jogos. As equipes abaixo do corte mínimo enfrentam o super-usuário várias vezes para que, com a adição desses jogos artificiais, haja muito mais paridade no número de jogos que as equipes jogam (Chartier et al., 2014, p. 17).

Pode-se perceber, portanto, que o método de Massey ainda é alvo de muitos esforços para o seu desenvolvimento, fortalecimento e potencialização de qualidades.

5 O MÉTODO DE ELO

O método de Elo foi criado por Arpad Élo, um físico húngaro e amante do xadrez. Seus esforços para a criação de um método de rating e ranqueamento de jogadores de xadrez deram origem a um método que até hoje é adaptado para diversos rankings esportivos.

Antes de analisar o método de ranqueamento, é importante compreender um aspecto relevante do seu desenvolvimento: a função logística, que será explorada na Seção 5.1. Já a teoria presente na Seção 5.2 é focada no método de Elo. A Seção 5.2 é baseada no livro de Langville e Meyer (2012).

5.1 Função logística

A função logística, que é base para o desenvolvimento do método de ranqueamento de Elo, foi proposta por Pierre Verhulst, matemático belga do século XIX, e é associada a um problema de dinâmica populacional. Ao longo de sua vida, Verhulst lecionou na Universidade de Bruxelas e na Academia Militar criada logo após a independência da Bélgica da Holanda (Bacaër, 2011).

O principal estudo sobre dinâmica populacional até esse momento era de Thomas Malthus, que argumentou que uma população tende a crescer em progressão geométrica. Assim, uma população dobra a cada período de tempo; ele mostrou, por exemplo, o caso dos Estados Unidos da América, que dobrava de população a cada 25 anos. A abordagem de Malthus foi considerada um tanto polêmica, o que trouxe muita atenção para o tema (Almeida, Porfiro, Pereira, 2024).

Verhulst, motivado pela pesquisa de Malthus, publicou em 1838 o artigo *Notice sur la loi que la population suit dans son accroissement*, em que argumenta que o crescimento em progressão geométrica só pode ocorrer temporariamente, em situações especiais (como o caso da colonização inicial da América do Norte) (Bacaër, 2011). De forma mais geral, diversas forças, como o tamanho do território e a dificuldade de encontrar comida, impedem um crescimento populacional exponencial. Em vez disso, ele propôs outro modelo baseado em equações diferenciais, que considerava a capacidade de suporte da região.

O modelo proposto por Verhulst foi chamado, por ele, de equação logística. Não é clara a razão da escolha deste nome. É provável que, enquanto trabalhava na Academia Militar, se tenha tornado familiar com a ideia de transporte e distribuição de alimentos e recursos para o exército, o que referencia a necessidade de suprimentos para uma população em um território. Verhulst também poderia estar adaptando a palavra *logis*, que em francês pode significar um local de moradia, novamente referenciando o suporte que uma região pode ter para a população que nela habita.

A função logística é definida da seguinte forma:

$$P(t) = \frac{M}{1 + Ae^{-kt}} \quad (43)$$

onde M , A e k são constantes positivas, e $P(t)$ corresponde à população no instante t .

Cada variável cumpre um papel no comportamento da função logística. A variável M é também chamada de “capacidade de suporte”, e determina o limite de crescimento da população. A variável k está relacionada com a velocidade de crescimento, ou ainda com a inclinação da curva de crescimento. Já a variável A pode ser definida a partir da noção de valor inicial. Perceba que:

$$P = \frac{M}{1 + Ae^{-kt}}$$

$$P + PAe^{-kt} = M$$

$$Ae^{-kt} = \frac{M - P}{P}$$

Assim, tomando $t = 0$, teremos o momento inicial de observação. Considerando P_0 a população inicial observada, tem-se que:

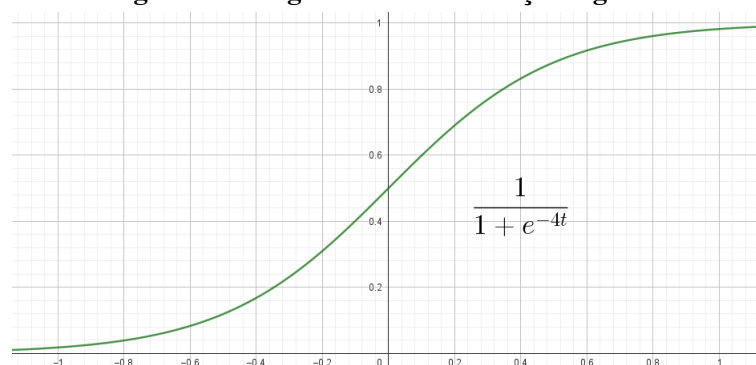
$$Ae^{-0k} = \frac{M - P_0}{P_0}$$

$$A = \frac{M - P_0}{P_0}$$

Assim, a variável A é definida como a diferença relativa entre a capacidade de suporte e a população inicial.

O formato padrão de uma função logística é retratado na Figura 11, em que foram utilizados $M = 1$, $A = 1$ e $k = 4$. Essa curva, em formato de S, também é chamada de sigmoide.

Figura 11 – O gráfico de uma função logística



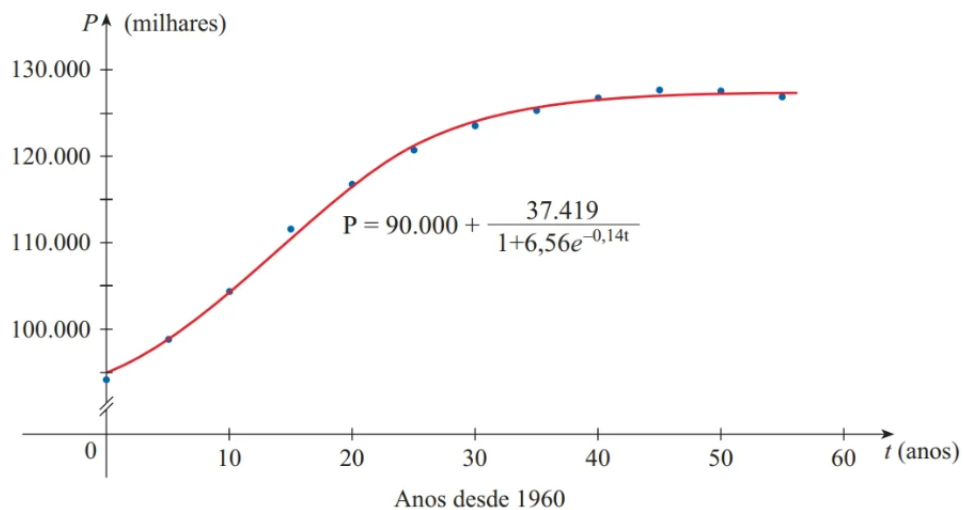
Fonte: O Autor.

Stewart, Clegg e Watson (2022a, p. 291) argumentam que essa função é usada com frequência para modelar crescimento de populações que apresentam, inicialmente, crescimento exponencial, mas que eventualmente alcançam um limite, ou suporte, devido aos recursos de sua localização. O formato da curva na Figura 11 ilustra a capacidade de modelar tais crescimentos populacionais. Os próprios autores modelam a equação para a população do Japão, com os dados da Tabela 18, produzindo o gráfico retratado na Figura 12 (Stewart, Clegg, Watson, 2022b, p. 587).

Tabela 18 – População do Japão

Ano	População (em milhares)
1960	94.092
1965	98.883
1970	104.345
1975	111.573
1980	116.807
1985	120.754
1990	123.537
1995	125.327
2000	126.776
2005	127.715
2010	127.579
2015	126.920

Fonte: Stewart, Clegg e Watson (2022b, p. 587).

Figura 12 – Função logística na população do Japão

Fonte: Stewart, Clegg e Watson (2022b, p. 587).

5.2 A teoria do método de ranqueamento de Elo

A ideia central do método de Elo é que o resultado de um jogo entre times de ratings diferentes deve ser, até certo ponto, previsível: espera-se que o time de maior rating vença, e quanto maior a diferença entre os ratings, maior é a expectativa de vencer. Ainda, uma atuação acima ou abaixo do esperado deve impactar adequadamente a atualização do rating futuro.

Inicialmente, a todos os times de um campeonato é atribuído um rating, ou uma pontuação. Isso pode ser feito considerando-se os resultados de partidas anteriores, ou simplesmente nivelando todos com pontuação 0. O rating inicial de um time i será chamado r_i . A cada vez que dois times se enfrentam, seus ratings são atualizados de acordo com o resultado, definindo-se $r_{i(\text{nov})}$. Para isso, é utilizada a variável S_{ij} , conforme a Equação (44), que representa a pontuação em disputa para o confronto entre os times i e j :

$$S_{ij} = \begin{cases} 1 & \text{se } i \text{ vence } j, \\ 0 & \text{se } i \text{ perde para } j, \\ 1/2 & \text{se } i \text{ e } j \text{ empatam} \end{cases} \quad (44)$$

O valor de S_{ij} relaciona-se com o que será chamado de *coeficiente de pontuação esperada* μ_{ij} , que representa o número de pontos que espera-se que o time i pontue contra o time j . Tal coeficiente dependerá da diferença entre os ratings, dada por $d_{ij} = r_{i(\text{antigo})} - r_{j(\text{antigo})}$, e está definido como:

$$\mu_{ij} = \frac{1}{1 + 10^{-d_{ij}/\xi}} \quad (45)$$

No contexto do xadrez, a constante ξ em (45) é geralmente considerada igual a 400. Nas próximas seções, será aprofundado o papel de tal constante e como ela pode ser adaptada para diferentes objetivos, mas nos cálculos e exemplos a seguir, será usado o valor de $\xi = 400$. Com isso, a nova pontuação de cada time é atualizada da seguinte forma:

$$r_{i(\text{novo})} = r_{i(\text{antigo})} + K(S_{ij} - \mu_{ij}) \quad \text{e} \quad r_{j(\text{novo})} = r_{j(\text{antigo})} + K(S_{ji} - \mu_{ji}) \quad (46)$$

Nesta fórmula, K é uma constante utilizada para determinar a variação da pontuação, baseada na diferença entre o resultado da partida e o resultado que era esperado. O valor de K pode ser escolhido de formas diferentes para cada jogo ou campeonato.

Para exemplificar o método e entender como as definições anteriores se dão na prática, será analisado um jogo do enxadrista Rafael Leitão, que figurava no número 1 do Brasil no ranking da Federação Internacional de Xadrez (FIDE) até o mês de julho de 2025. Os dados do jogo foram consultados no site oficial da FIDE (2024).

Exemplo 5.1. *Entre os dias 10 e 17 de outubro de 2024, foi realizado o III Manaus Chess Open, na cidade de Manaus. No momento do início do campeonato, Leitão possuía um rating de 2592 pontos e enfrentou 9 jogadores. Uma partida foi contra Jaime Sunye Neto, que detém também o título de Grande Mestre, com pontuação 2381. Determine o coeficiente de pontuação esperada dos dois jogadores e a atualização de suas pontuações para a vitória de Leitão, com $K = 10$.*

A partir da pontuação inicial dos jogadores, determina-se μ_{ij} , que representa o resultado esperado da partida entre Leitão (representado por i) e Sunye Neto (representado por j).

$$\mu_{ij} = \frac{1}{1 + 10^{-(2592-2381)/400}} = \frac{1}{1 + 10^{-211/400}} \approx 0,771$$

$$\mu_{ji} = \frac{1}{1 + 10^{-(2381-2592)/400}} = \frac{1}{1 + 10^{211/400}} \approx 0,228$$

Ou seja, a expectativa é de que Leitão pontue 0,771 sobre Sunye Neto. Como as pontuações possíveis, de acordo com (44), são 1, $1/2$ e 0, Leitão obterá pontuação positiva se ganhar, e negativa se empatar ou perder. Também Sunye Neto perderá pontos se perder a partida, mas ganhará pontos em caso de empate ou vitória.

Leitão venceu esta partida, definindo-se $S_{ij} = 1$ e $S_{ji} = 0$. Como neste campeonato foi definido $K = 10$, tem-se:

$$r_{(i,\text{novo})} = 2592 + 10(1 - 0,771) = 2594,29$$

$$r_{(j,\text{novo})} = 2381 + 10(0 - 0,228) = 2378,72$$

Ou seja, Leitão atualizou seu rating em aproximadamente $+2,3$ pontos neste jogo, e Sunye atualizou em aproximadamente $-2,3$ pontos.

É importante notar que o coeficiente μ_{ij} possui formato parecido com as funções logísticas estudadas no início deste capítulo. De fato, tomando-se $M = 1$, $A = 1$, $k = \ln(10)/400$ e $t = d_{ij}$, temos a seguinte igualdade.

$$\frac{M}{1 + Ae^{-kt}} = \frac{1}{1 + e^{(-\ln(10)/400) \cdot d_{ij}}} = \frac{1}{1 + 10^{-d_{ij}/400}}$$

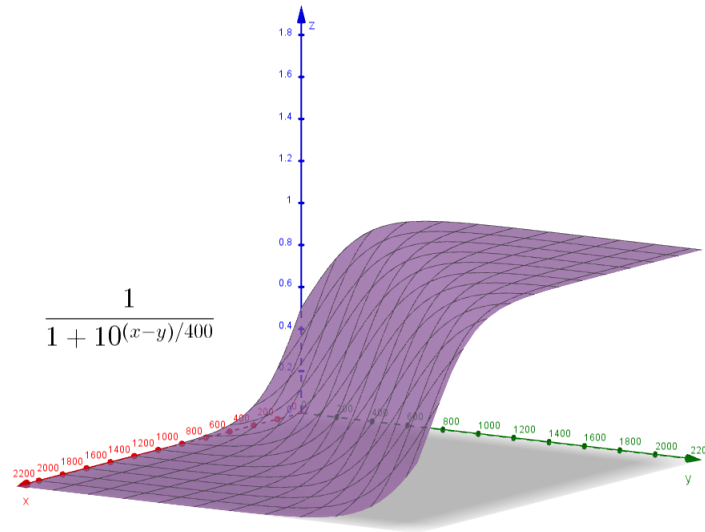
Portanto, pode-se compreender que a expectativa de pontuação de um time sobre o outro em uma partida cresce de acordo com uma função logística, exemplificada na Figura 11. Pela pontuação possível de S_{ij} dada por (44), o coeficiente de pontuação esperada deve estar limitado entre 0 e 1. O crescimento da curva é inicialmente exponencial, mas se estabiliza com o tempo. Caso $d_{ij} < 0$, a pontuação esperada é menor do que $1/2$, ou seja, uma vitória de i é pouco provável, e um empate já seria acima das expectativas. No caso $d_{ij} = 0$, ou seja, caso $r_{i(\text{antigo})} = r_{j(\text{antigo})}$, então a expectativa de pontuação é $1/2$; se os ratings são iguais, pode-se esperar um empate entre os competidores. Já se $d_{ij} > 0$, então um empate está abaixo das expectativas de vitória da equipe i .

Observando o confronto entre Leitão e Sunye Neto, percebe-se que, como o rating de Leitão era maior que o de Sunye Neto, então o μ_{ij} para Leitão era maior que $1/2$, enquanto μ_{ji} para Sunye Neto era menor que $1/2$.

Ainda observando a Figura 11, e considerando $M = 1$, percebe-se que a curva possui assíntotas horizontais em 0 e 1. Tomando esse fato para o contexto da expectativa de pontuação, pode-se afirmar que, independentemente da diferença de pontuação entre duas equipes, jamais será possível afirmar que uma vitória ou derrota seja certa. Pode-se apenas afirmar que a expectativa de pontuação é muito próxima de 1 ou de 0.

A Figura 13 mostra o comportamento do coeficiente de pontuação esperada μ_{ij} , considerando-o como função dependente das variáveis r_i e r_j (representadas em x e y). A superfície foi construída no ambiente GeoGebra.

Figura 13 – A superfície de pontuação esperada em função dos ratings atuais



Fonte: O Autor.

5.3 Considerações sobre o coeficiente de pontuação esperada μ_{ij}

O coeficiente μ_{ij} tem papel central no método Elo, e demanda uma observação mais atenta. Com base nas definições da seção anterior, apresentam-se três propriedades.

Propriedade 5.1. Considerando o coeficiente de pontuação esperada μ_{ij} conforme a equação (45), tem-se que $0 < \mu_{ij} < 1$

De fato, $0 < \mu_{ij} < 1$ é equivalente a afirmar que, na fração: (1) o numerador e o denominador são positivos; (2) o numerador é menor que o denominador. Ambas as afirmações são imediatas a partir da definição.

Propriedade 5.2. Considerando o coeficiente de pontuação esperada μ_{ij} conforme (45), tem-se que

$$\mu_{ij} + \mu_{ji} = 1$$

Demonstração. De fato, veja que:

$$\mu_{ij} + \mu_{ji} = \frac{1}{1 + 10^{-d_{ij}/400}} + \frac{1}{1 + 10^{-d_{ji}/400}} = \frac{1 + 10^{-d_{ji}/400} + 1 + 10^{-d_{ij}/400}}{(1 + 10^{-d_{ij}/400})(1 + 10^{-d_{ji}/400})}$$

$$\mu_{ij} + \mu_{ji} = \frac{2 + 10^{-d_{ij}/400} + 10^{-d_{ji}/400}}{1 + 10^{-d_{ij}/400} + 10^{-d_{ji}/400} + 10^{-d_{ij}/400} \cdot 10^{-d_{ji}/400}}$$

Uma vez que $-d_{ij} + (-d_{ji}) = -(r_i - r_j) - (r_j - r_i) = 0$, então:

$$\mu_{ij} + \mu_{ji} = \frac{2 + 10^{-d_{ij}/400} + 10^{-d_{ji}/400}}{1 + 10^{-d_{ij}/400} + 10^{-d_{ji}/400} + 10^{0/400}} = \frac{2 + 10^{-d_{ij}/400} + 10^{-d_{ji}/400}}{1 + 10^{-d_{ij}/400} + 10^{-d_{ji}/400} + 1}$$

$$\mu_{ij} + \mu_{ji} = \frac{2 + 10^{-d_{ij}/400} + 10^{-d_{ji}/400}}{2 + 10^{-d_{ij}/400} + 10^{-d_{ji}/400}} = 1$$

□

Propriedade 5.3. Após uma partida, a atualização dos ratings se dá de forma que os pontos que uma equipe conquista têm o mesmo módulo dos pontos que a outra equipe cede. Ou seja, sendo $r_{i(\text{nov})}$ o rating atualizado da equipe i após uma partida, e $r_{j(\text{nov})}$ o rating atualizado da equipe j após uma partida, ambos definidos de acordo com (46), tem-se que:

$$r_{i(\text{nov})} - r_{i(\text{antigo})} = -[r_{j(\text{nov})} - r_{j(\text{antigo})}]$$

Demonstração. Temos como hipótese que:

$$S_{ij} + S_{ji} = 1 \quad e \quad \mu_{ij} + \mu_{ji} = 1$$

Dessa forma, e considerando as equações em (46):

$$\begin{aligned} r_{i(\text{nov})} - r_{i(\text{antigo})} &= K(S_{ij} - \mu_{ij}) \\ &= K(1 - S_{ji} - (1 - \mu_{ji})) \\ &= K(1 - S_{ji} - 1 + \mu_{ji}) \\ &= K(-S_{ji} + \mu_{ji}) \\ &= -K(S_{ji} - \mu_{ji}) \\ &= -[r_{j(\text{nov})} - r_{j(\text{antigo})}] \end{aligned}$$

Assim, os pontos que uma equipe conquista têm o mesmo módulo dos pontos que a outra equipe cede ao fim de um confronto. □

As conclusões obtidas nas últimas três propriedades podem ainda ser generalizadas. Considerando os possíveis valores de S_{ij} , percebe-se que 1 ponto está sendo dividido entre os competidores. Pode-se considerar um sistema de ranqueamento que divide mais que 1 ponto, ou seja, M pontos, de forma que $S_{ij} + S_{ji} = M$. Se a fórmula do coeficiente de pontuação μ_{ij} também for adaptada, com M no numerador, será o caso que: (1) $0 < \mu_{ij} < M$; (2) $\mu_{ij} + \mu_{ji} = M$. Assim ainda será mantido um sistema que respeita a Propriedade 5.3, em que um competidor sempre “rouba” os pontos do outro.

5.4 Considerações sobre as constantes K e ξ

Outras observações podem ser realizadas sobre as constantes do método de Elo: a constante K , e a constante ξ . Nesta seção, serão consideradas suas interpretações dentro do sistema de rating.

Anteriormente, afirmou-se que K é utilizada para determinar a variação da pontuação nova em relação à antiga. Na prática, ela define o impacto que uma performance acima ou abaixo do esperado tem na atualização do rating. Para exemplificar tal afirmação, tome-se novamente o Exemplo 5.1, feito na Seção 5.2. Com $K = 10$, a variação dos ratings foi de 2,3 pontos. Caso fosse utilizado $K = 20$, teria-se, para Leitão, $r_{(\text{nov})} = 2596,57$, uma mudança de aproximadamente +4,56 pontos, e para Sunye Neto, $r_{(\text{nov})} = 2376,44$, uma mudança de aproximadamente -4,56 pontos. Como complementação da análise sobre o coeficiente K , considere o Exemplo 5.2

Exemplo 5.2. *Ainda no III Manaus Chess Open, Leitão enfrentou Roberto Junio Brito Molina, que detém título de Mestre, também com pontuação 2381. No caso de um empate entre os jogadores, determine a variação de suas pontuações para $K = 10$ e para $K = 20$.*

Será considerada aqui a mesma pontuação inicial para Leitão verificada na partida anterior, uma vez que a atualização de rating só ocorre ao fim do campeonato.

Como a pontuação de Leitão e Molina são as mesmas observadas para Leitão e Sunye Neto, pode-se considerar os mesmos $\mu_{ij} = 0,771$ e $\mu_{ji} = 0,228$. Como Leitão empatou esta partida, tornando $S_{ij} = S_{ji} = 1/2$, os ratings serão atualizados da seguinte forma.

$$r_{(i,\text{nov})} = 2592 + 10(1/2 - 0,771) = 2589,29$$

$$r_{(j,\text{nov})} = 2381 + 10(1/2 - 0,228) = 2383,72$$

Neste jogo, então, Leitão atualizou seu rating em aproximadamente -2,7 pontos, enquanto Molina teve uma mudança aproximada de +2,7 pontos. Novamente, pode-se escolher outro K para visualizar o seu papel dentro da definição dos ratings. Escolhendo-se $K = 20$, tem-se para Leitão $r_{(\text{nov})} = 2586,57$, uma mudança de -5,43 pontos, e para Molina $r_{(\text{nov})} = 2386,43$, uma mudança de +5,43 pontos.

Langville e Meyer destacam como a escolha desta constante tem grande impacto no ranqueamento.

Se K for muito grande, então dá-se muito peso à diferença entre as pontuações reais e esperadas, o que resulta em muita volatilidade nas classificações — por exemplo, um K grande significa que jogar apenas um pouco acima das expectativas pode gerar uma grande mudança nas classificações. Por outro lado, se K for muito pequeno, as fórmulas de Elo perdem sua capacidade de levar em conta a melhoria ou deterioração do desempenho, e as classificações se tornam muito

estáticas — por exemplo, um K pequeno significa que mesmo uma melhoria significativa no desempenho de alguém não pode gerar uma grande mudança nas classificações (Langville e Meyer, 2012, p. 55).

Originalmente, Elo determinou $K = 10$. Atualmente, cada federação, ao adaptar o método para o seu esporte, escolhe K de acordo com a competição na qual o jogo ocorre. Assim, o resultado obtido em uma partida mais importante tem mais peso em suas recompensas. A FIDE estabeleceu os valores de K da seguinte forma.

Tabela 19 – Valores de K no sistema de classificação de xadrez

Valor de K	Condição
40	Para um jogador novo na lista de classificação até que ele tenha completado eventos com pelo menos 30 jogos
20	Enquanto a classificação de um jogador permanecer abaixo de 2400
10	Uma vez que a classificação publicada de um jogador tenha atingido 2400 e permaneça nesse nível subsequentemente, mesmo se a classificação cair abaixo de 2400
40	Para todos os jogadores até seu 18º aniversário, desde que sua classificação permaneça abaixo de 2300

Fonte: FIDE (2025a).

Outra instituição esportiva que também utiliza o método de Elo para seus ranqueamentos é a *Fédération Internationale de Football Association* (FIFA). Seu ranking estabeleceu a constante K , de acordo com a competição, da seguinte maneira:

Tabela 20 – Valores de K no ranking da FIFA

Valor de K	Tipo de jogo
5	Partidas amistosas disputadas fora das janelas do Calendário de Partidas Internacionais
10	Partidas amistosas disputadas durante as janelas do Calendário de Partidas Internacionais
15	Partidas da fase de grupos das competições da Liga das Nações
25	Partidas de play-off e finais das competições da Liga das Nações
25	Partidas de qualificação para as competições finais das Confederações e para as competições finais da Copa do Mundo da FIFA
35	Partidas das competições finais das Confederações até a fase de quartas de final
40	Partidas das competições finais das Confederações a partir da fase de quartas de final; todas as partidas da Copa das Confederações da FIFA
50	Partidas das competições finais da Copa do Mundo da FIFA até a fase de quartas de final
60	Partidas das competições finais da Copa do Mundo da FIFA a partir da fase de quartas de final

Fonte: FIFA (2018).

Vale, porém, mencionar que a FIFA adaptou seu modelo de ranqueamento para que, em casos específicos, não sejam válidas as Propriedades 5.2 e 5.3. Em fases finais (mata-mata) de competições, e especialmente caso a eliminação se dê em disputa de pênaltis, o método sobre adaptações que o distanciam muito do apresentado por Arpad Elo.

Por fim, outra definição importante para o funcionamento do método de Elo reside na constante ξ , presente no coeficiente de pontuação esperada μ_{ij} . Sua função principal está em influenciar a extensão, ou propagação (em inglês, *spread*) dos ratings. Para entender melhor tal definição, é necessária uma manipulação algébrica da função μ_{ij} . Nesta manipulação, será fixado o valor $\xi = 400$ para verificar-se sua interpretação.

$$\mu_{ij} = \frac{1}{1 + 10^{-d_{ij}/400}}, \text{ com } d_{ij} = r_i - r_j$$

$$\mu_{ij} = \frac{1}{1 + 10^{r_j - r_i/400}} \Rightarrow \mu_{ij} = \frac{1}{1 + \frac{10^{r_j/400}}{10^{r_i/400}}}$$

$$\mu_{ij} = \frac{1}{\frac{10^{r_i/400} + 10^{r_j/400}}{10^{r_i/400}}} \Rightarrow \mu_{ij} = \frac{10^{r_i/400}}{10^{r_i/400} + 10^{r_j/400}}$$

Considerando este último formato do coeficiente, é interessante considerar a razão entre as expectativas de pontuação de cada time em uma partida, ou seja, $\frac{\mu_{ij}}{\mu_{ji}}$, em que obtém-se o seguinte.

$$\frac{\mu_{ij}}{\mu_{ji}} = \frac{\frac{10^{r_i/400}}{10^{r_i/400} + 10^{r_j/400}}}{\frac{10^{r_j/400}}{10^{r_i/400} + 10^{r_j/400}}} = \frac{10^{r_i/400}}{10^{r_j/400}}$$

$$\mu_{ij} = \mu_{ji} \cdot 10^{(r_i - r_j)/400} \quad (47)$$

Perceba que, caso a diferença de rating entre os times seja de 400 pontos, então $\mu_{ij} = 10\mu_{ji}$. Sendo μ_{ij} a expectativa de pontos que o time i pontuará no time j , então uma possível interpretação é de que o time i é 10 vezes melhor do que o time j .

Em um formato mais geral do método, a equação (47) pode ser escrita da seguinte forma:

$$\mu_{ij} = \mu_{ji} \cdot 10^{(r_i - r_j)/\xi} \quad (48)$$

Portanto, para cada ξ pontos de diferença no ranqueamento de dois times, a expectativa de vitória do time mais forte é dez vezes maior. A manipulação de ξ afeta o resultado do sistema de ranqueamento, traduzindo o quão “espalhados” estão os ranqueamentos das equipes, e até que ponto um rating mais alto se traduz em expectativa de vitória. A possibilidade de alteração de ξ e do fator K torna o método de Elo bastante personalizável, o que contribui para a sua popularidade em tantos esportes. A FIFA, por exemplo, utiliza $\xi = 600$ (FIFA, 2018).

A manipulação realizada acima também explicita a decisão, no método de Elo, de modificar o formato padrão da função logística, como mostrado em (43). As escolhas de M , A e k definem a pontuação atribuída à vitória, e dão origem ao número 10 no denominador, tornando possível a interpretação de uma expectativa de pontuação “10 vezes maior”, o que é mais didático e de simples interpretação do que a utilização do número de Euler.

5.5 Análise de confrontos e exemplos

Como um exercício, utilizando a teoria desenvolvida até aqui, pode-se realizar confrontos fictícios entre jogadores de xadrez para analisar qual a expectativa de pontuação de cada um. Tomando o número 1 mundial em julho de 2025, Magnus Carlsen (indicado por i), com 2837 pontos, e o número 1 brasileiro, Rafael Leitão (indicado por j), com 2596 pontos, têm-se as seguintes previsões.

$$\mu_{ij} = \frac{1}{1 + 10^{-d_{ij}/400}} = \frac{1}{1 + 10^{-(2837-2596)/400}} \approx 0,800$$

$$\mu_{ji} = \frac{1}{1 + 10^{-d_{ji}/400}} = \frac{1}{1 + 10^{-(2596-2837)/400}} \approx 0,199$$

Ou seja, em um confronto entre o melhor do mundo e o melhor do Brasil, espera-se que Carlsen pontue 0,8 pontos, o que significa que uma vitória aumentaria seu rating, mas um empate ou derrota causaria uma queda.

Atualmente, algoritmos e motores de xadrez, também chamados de *bots*, são comumente usados para treinamentos e outras análises. O motor mais famoso e poderoso no momento desta pesquisa chama-se Stockfish, que possui um rating estimado de 3640 pontos (Computer Chess Rating Lists, 2025). É interessante verificar o retorno do coeficiente de pontuação esperada em confrontos com esta ferramenta.

No caso de um confronto entre Stockfish (i) e Carlsen (j), têm-se os seguintes resultados.

$$\mu_{ij} = \frac{1}{1 + 10^{-d_{ij}/400}} = \frac{1}{1 + 10^{-(3640-2837)/400}} \approx 0,990$$

$$\mu_{ji} = \frac{1}{1 + 10^{-d_{ji}/400}} = \frac{1}{1 + 10^{-(2837-3640)/400}} \approx 0,009$$

Já em um confronto entre Stockfish (i) e Rafael Leitão (j), têm-se os seguintes resultados.

$$\mu_{ij} = \frac{1}{1 + 10^{-d_{ij}/400}} = \frac{1}{1 + 10^{-(3640-2596)/400}} \approx 0,997$$

$$\mu_{ji} = \frac{1}{1 + 10^{-d_{ji}/400}} = \frac{1}{1 + 10^{-(2837-3640)/400}} \approx 0,002$$

Assim, percebe-se um contexto atual em que ferramentas digitais ultrapassaram em muito a técnica humana para partidas de xadrez.

Outras explorações sobre valores de μ_{ij} também podem produzir resultados interessantes. Considerando a equação (47), propõem-se os seguintes exemplos.

Exemplo 5.3. *Qual deve ser a diferença de ratings de forma que a expectativa de pontos de uma equipe seja o dobro da expectativa de pontos da outra equipe? Ou seja, para que valor de d_{ij} tem-se que $\mu_{ij} = 2\mu_{ji}$?*

A partir de (47), encontra-se o seguinte:

$$\mu_{ij} = \mu_{ji} \cdot 10^{(r_i - r_j)/400} \Rightarrow \frac{\mu_{ij}}{\mu_{ji}} = 10^{d_{ij}/400}$$

Para que $\mu_{ij} = 2\mu_{ji}$, então deve ser o caso que $\frac{\mu_{ij}}{\mu_{ji}} = 2$. Então, considerando $\log(2) = 0,3$:

$$10^{d_{ij}/400} = 2$$

$$\log(10^{d_{ij}/400}) = \log(2)$$

$$d_{ij}/400 = 0,3$$

$$d_{ij} = 120$$

Portanto, se dois times possuem uma diferença inicial de ratings de 120 pontos, então, segundo o método de Elo, espera-se que um conquiste o dobro de pontos do que o outro.

Exemplo 5.4. *Um jogador de xadrez, já em fim de carreira, decide que não vai jogar todos os jogos de um campeonato. Já sem ambição de um crescimento acelerado no ranking, e com intuito de proteger seus pontos já conquistados, ele decide que só aceitará participar do confronto se a pontuação esperada para ele for igual ou maior que 0,8. Sendo assim, qual o menor d_{ij} que o jogador pode aceitar?*

Considerando a situação descrita e a definição (45), tem-se que:

$$\mu_{ij} \geq 0,8$$

$$\frac{1}{1 + 10^{-d_{ij}/400}} \geq 0,8$$

$$1 \geq 0,8(1 + 10^{-d_{ij}/400})$$

$$\frac{1}{0,8} - 1 \geq 10^{-d_{ij}/400}$$

$$\log\left(\frac{1}{4}\right) \geq \log(10^{-d_{ij}/400})$$

$$-0,6 \geq -d_{ij}/400$$

$$-240 \geq -d_{ij}$$

$$240 \leq d_{ij}$$

Assim, para que o jogador sempre tenha uma pontuação esperada de pelo menos 0,8 pontos, ele só deve jogar partidas em que a diferença para o rating do oponente seja de pelo menos 240 pontos.

Exemplo 5.5. *Sabe-se que a maior pontuação já conquistada na história do ranking da FIDE foi de 2882, em 2014, pelo próprio Magnus Carlsen, que ainda permanece como o número 1 do mundo. No momento da escrita deste trabalho, ele possui 2837 pontos. Pode-se, então, perguntar: quantas partidas consecutivas Carlsen deve ganhar para retomar sua pontuação máxima? E, para quebrar seu recorde e atingir uma pontuação igual a 3000, quantas partidas consecutivas deve ganhar?*

Para verificar quantas vezes Carlsen precisa ganhar consecutivamente para alcançar novamente sua pontuação máxima, é preciso primeiro escolher um oponente e verificar sua pontuação. Foi escolhido, então, o segundo lugar do mundo em julho de 2025, o americano Hikaru Nakamura, que possui 2804 pontos.

Perceba que é necessário atualizar, a cada partida, a pontuação de Carlsen, a pontuação de Nakamura, e o coeficiente de pontuação esperada. Para isso, foi construído o seguinte código em Julia.

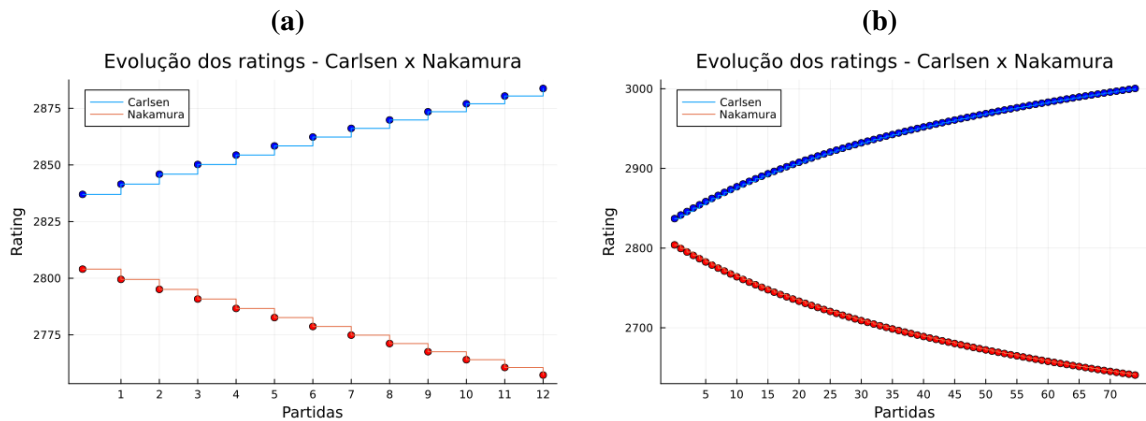
```

1  carlsen = 2837
2  meta = 2882
3  nakamura = 2804
4  count = 0
5
6  while carlsen <= meta
7      mu1 = 1/(1 + 10^((nakamura - carlsen)/400))
8      mu2 = 1/(1 + 10^((carlsen - nakamura)/400))
9      carlsen = carlsen + 10*(1 - mu1)
10     nakamura = nakamura + 10*(0 - mu2)
11     count = count + 1
12 end

```

A partir de tal código, é possível obter a Figura 14, que apresenta a evolução dos ratings de Carlsen e Nakamura nestes confrontos fictícios, supondo sempre a vitória de Carlsen.

Figura 14 – Confronto entre Carlsen e Nakamura: (a) até o recorde histórico de 2882 pontos; (b) até uma meta de 3000 pontos



Fonte: O Autor.

Percebe-se que Carlsen deve vencer Nakamura 12 vezes seguidas para alcançar novamente sua pontuação de 2882. Agora, se a variável “meta” for alterada para 3000, teremos que Carlsen deve vencer Nakamura 74 vezes consecutivas como ilustrado na parte (b) da Figura 14. Isso parece realmente inviável.

Vale a reflexão de que, como a cada partida a pontuação de Carlsen aumenta e a de Nakamura diminui, então o “prêmio” de cada vitória sempre diminui (pois é contra um jogador cada vez mais fraco). Isso explica a necessidade de tantas partidas. Portanto, pode-se considerar outra situação, em que Carlsen enfrenta um oponente fictício com pontuação fixa igual à média dos 10 primeiros lugares do ranking masculino, cujo valor hoje é de 2778,7. Neste contexto, o código ficaria da seguinte forma.

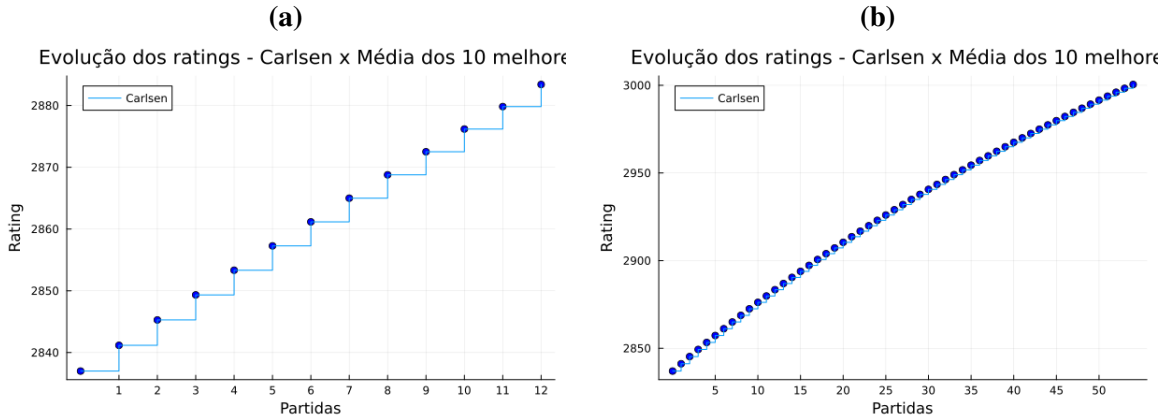
```

1 carlsen = 2837
2 meta = 2882
3 oponente = 2778,7
4 count = 0
5
6 while carlsen <= meta
7     mu = 1/(1+10^((oponente - carlsen)/400))
8     carlsen = carlsen + 10*(1-mu)
9     count = count + 1
10 end

```

Assim, Carlsen ainda deve ganhar 12 partidas para atingir seu máximo histórico. Já para alcançar a pontuação de 3000 pontos, a quantidade de partidas necessárias cai para 54. A Figura 15 mostra graficamente a evolução do rating de Carlsen no confronto fictício com a média dos 10 melhores ratings.

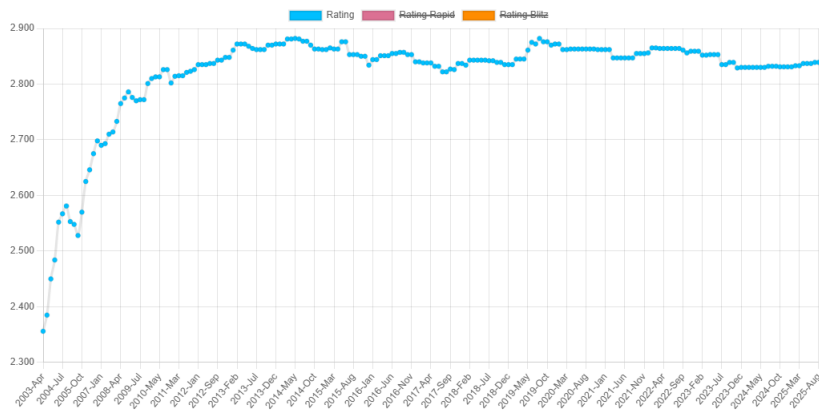
Figura 15 – Confronto entre Carlsen e a média dos 10 melhores ratings: (a) até o recorde histórico de 2882 pontos; (b) até uma meta de 3000 pontos



Fonte: O Autor.

Considerar uma sequência tão grande de vitórias é interessante, mas não representa fielmente a realidade. O histórico de partidas de um jogador frequentemente apresenta altos e baixos - mesmo o de Magnus Carlsen. A Figura 16 mostra a evolução do rating de Carlsen, de acordo com a FIDE (2025b).

Figura 16 – A evolução do rating de Magnus Carlsen



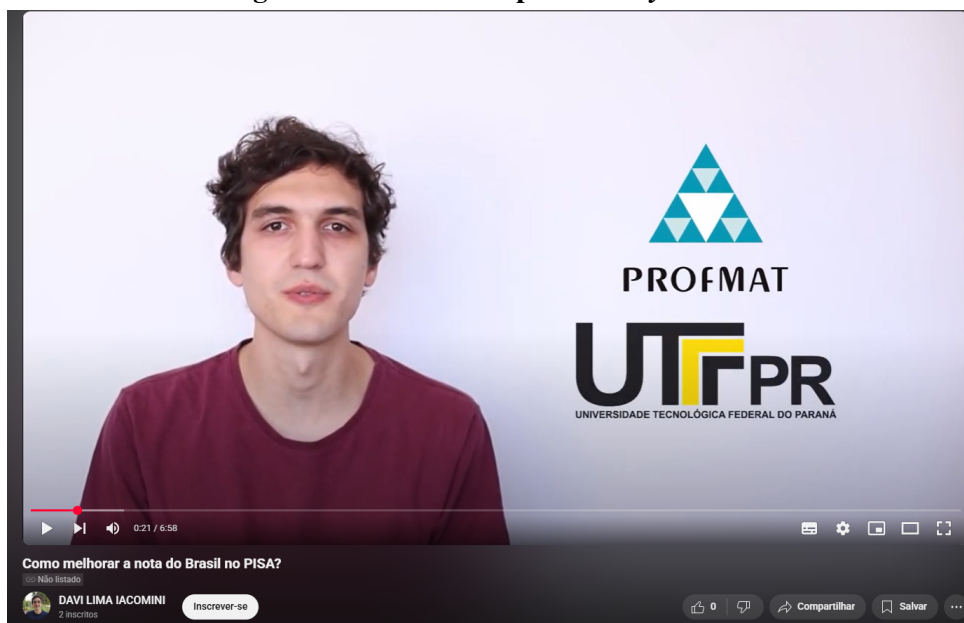
Fonte: FIDE, 2025b.

6 PRODUTO EDUCACIONAL

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) é a entidade responsável pela regulamentação dos Programas de Mestrado Profissional para Qualificação de Professores da Rede Pública de Educação Básica - ProEB, dentre os quais o PROFMAT faz parte. Na portaria nº 207, de 4 de Julho de 2024, enuncia que “para obtenção do título de mestre ou doutor no PROEB, os programas deverão contemplar a elaboração, pelos discentes, de uma dissertação ou tese, e de um recurso educacional” (CAPES, 2024).

Como produto educacional deste trabalho de mestrado, foi produzida uma série de vídeos a respeito de índices socioeconômicos e métodos matemáticos de ranqueamento. Os vídeos foram postados no site *youtube*. Foram explorados nestes vídeos o Coeficiente de Gini, o Índice de Desenvolvimento Humano (IDH) e a prova do Programa Internacional de Avaliação de Estudantes (PISA), seguindo o que foi desenvolvido no Capítulo 3 do presente trabalho. Também foi considerado o método de Elo para ranqueamento discutido anteriormente no Capítulo 5. A Figura 17 mostra uma captura do vídeo no *youtube*.

Figura 17 – O vídeo disponível no *youtube*.



Fonte: O Autor.

Este produto também foi cadastrado na plataforma eduCapes, e pode ser acessado através do link: <http://educapes.capes.gov.br/handle/capes/1131950>.

6.1 Justificativa e Objetivos

Considera-se que a produção científica deve ser conhecida pelo público geral para que seja devidamente valorizada. Menegotto afirma que cidadãos consomem resultados de pesquisa, mesmo sem reconhecer do que se trata:

Entretanto, em muitos países, como é o caso do Brasil, o financiamento da pesquisa é majoritariamente público, ou seja, são todos os cidadãos que pagam, via impostos, para terem essas pesquisas feitas. Ainda que esses cidadãos venham a utilizar as possíveis tecnologias advindas dessas pesquisas, as fazem – em grande parte do tempo – de maneira inconsciente. E em momentos de crises econômicas, pode ser que as ciências sejam diretamente afetadas, já que a população pode não enxergar com a devida importância o progresso científico, afinal de contas, apesar de se utilizarem dele, não sabem que o estão fazendo (Menegotto, 2021, p. 17-18).

Menegotto também afirma que, no Brasil, o consumo de conteúdo científico em formato textual ainda é de difícil acesso.

Com a fase inicial da Internet, havia muita informação na forma de textos. A resposta a essas perguntas já estava na Internet em formato escrito há bastante tempo, entretanto há de se considerar que há uma dificuldade inata na comunicação científica na forma de textos, uma vez que as capacidades interpretativas da linguagem escrita são um entrave na comunicação entre as pessoas, principalmente em países subdesenvolvidos. Por exemplo, os dados do PISA 2018 indicam que ainda hoje no Brasil, cerca de 50% dos jovens brasileiros não atingiram o mínimo de proficiência em leitura que todos os jovens devem adquirir até o final do ensino médio (Menegotto, 2021, p. 6).

Pelo anterior, o formato de vídeo foi escolhido pela sua popularidade e acessibilidade, sendo simples de se assistir e compartilhar. Assim, o propósito de democratização da pesquisa de mestrado é melhor cumprido, pois qualquer professor de matemática pode acessar o vídeo e se interessar pela pesquisa, podendo abordar este conhecimento em sua sala de aula. Também, estudantes podem assistir ao vídeo para se aprofundarem em seus estudos.

O objetivo do recurso educacional do presente trabalho é desenvolver uma série de vídeos acessíveis e práticos, que aproximem o conhecimento matemático desenvolvido nesta dissertação, de professores, estudantes e público geral, por meio de exemplos, animações e explorações de casos reais.

6.2 Etapas de criação

Com base nos temas discutidos na dissertação, foi pensado um roteiro para a disseminação do conhecimento. Considerando a habilidade EM13MAT104 da BNCC, que recomenda a exploração de métodos de cálculo de índices socioeconômicos, o conteúdo do Capítulo 3 deste trabalho pode ser considerado útil no planejamento de professores (Brasil, 2018). Além disso, para incentivar os estudantes para a pesquisa matemática, e devido a um diálogo com diversos conteúdos integrantes do Ensino Médio, como funções exponenciais e funções logarítmicas, mudança de bases, pensamento computacional, entre outros, o método de ranqueamento de Elo também foi escolhido.

Uma vez com o tema e roteiros definidos, foram consideradas referências de comunicadores de matemática em vídeo, na intenção de integrar melhores práticas e ferramentas para a criação do vídeo. Alguns canais utilizados como base foram os seguintes:

- 3Blue1Brown (<https://www.youtube.com/@3blue1brown>). Destaca-se principalmente o estilo de animação utilizado, a partir do software Manim, que foi adotado na visualização de fórmulas e gráficos.
- Dr. Trefor Bazett (<https://www.youtube.com/@DrTrefor>). O posicionamento de câmera apresentado, que mostra ao mesmo tempo o comunicador e o conteúdo, este sempre com movimentos e destaques, também foi adotado nos vídeos elaborados.
- Numberphile (<https://www.youtube.com/@numberphile>). É um canal modelo no quesito divulgação científica, inclusive para estudantes mais novos.
- BBC News Brasil (<https://www.youtube.com/@BBCNewsBrasil>). O veículo de comunicação possui vídeos muitos bons de divulgação matemática, nos quais, o ritmo de fala e imagens, serviram de referência.

Para o processo de gravação, foi realizada captação de imagem por câmera e de som por microfone, que posteriormente foram unidas na edição. O ambiente de gravação escolhido foi a própria residência do mestrando, com fundo em parede branca. O momento de captação foi misto, com preferências para imagens de manhã e à tarde, com iluminação natural benéfica para uma melhor qualidade de imagem.

Periodicamente nos vídeos criados, foram incluídas imagens, tabelas, gráficos e cálculos. Para acrescentar tais recursos, a gravação foi feita já com o orador deslocado do centro. Assim, qualquer mídia poderia ser colocada ao seu lado no momento da edição, a exemplo do que é feito pelo Dr. Trefor Bazett.

6.3 Resultado

O resultado final do produto educacional está nos links abaixo, todos disponíveis de forma pública no *youtube*.

- Como o Coeficiente de Gini quantifica a desigualdade? - <https://youtu.be/tIAbZCfYb7Y>
- Além da renda: como determinar o desenvolvimento humano - <https://youtu.be/bEQuV0uMiMM>
- Como melhorar a nota do Brasil no PISA? - <https://youtu.be/Zn1hdoh8XWM>
- O método de ranqueamento de Elo - <https://youtu.be/uQbHA7fjEOc>

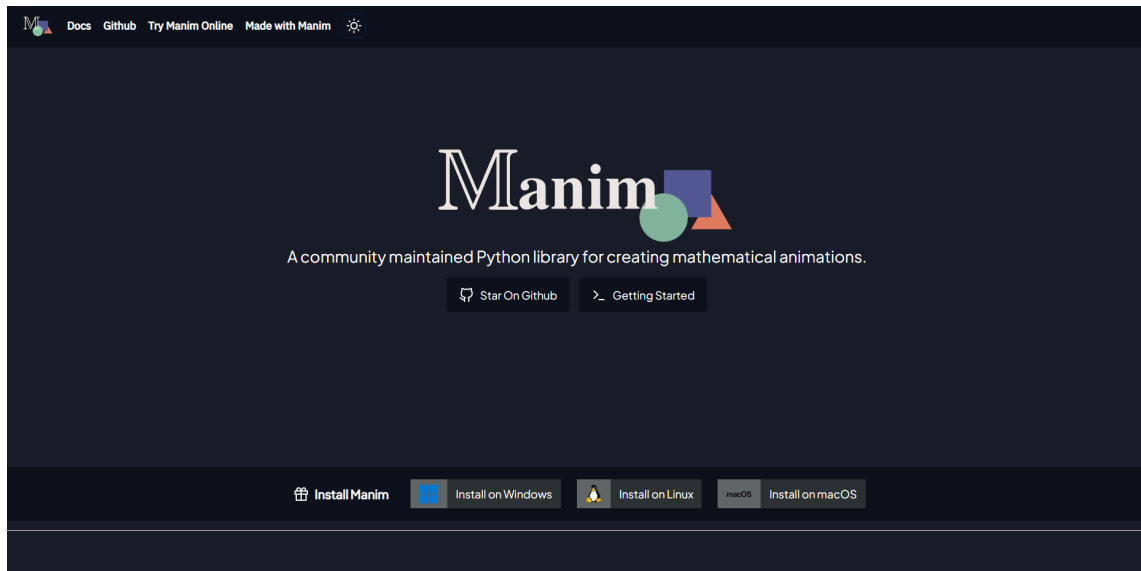
6.4 Softwares utilizados

Para a produção da série de vídeos, várias ferramentas foram utilizadas, na tentativa de criar o melhor produto final possível. Dividiremos os softwares utilizados em categorias.

6.4.1 Manim

O Manim (acrônimo para *Math Animation*) é uma biblioteca comunitária em Python para criação de animações de matemática. Seu *download* é gratuito, e tutoriais para seu uso estão disponíveis em diversos *sites* e canais. Para a instalação, recomenda-se o passo a passo da documentação oficial do projeto. Outra possibilidade é seguir as orientações presentes em tutoriais na plataforma *youtube*. A Figura 18 mostra a tela inicial do *site* oficial da biblioteca.

Figura 18 – O *site* do software Manim

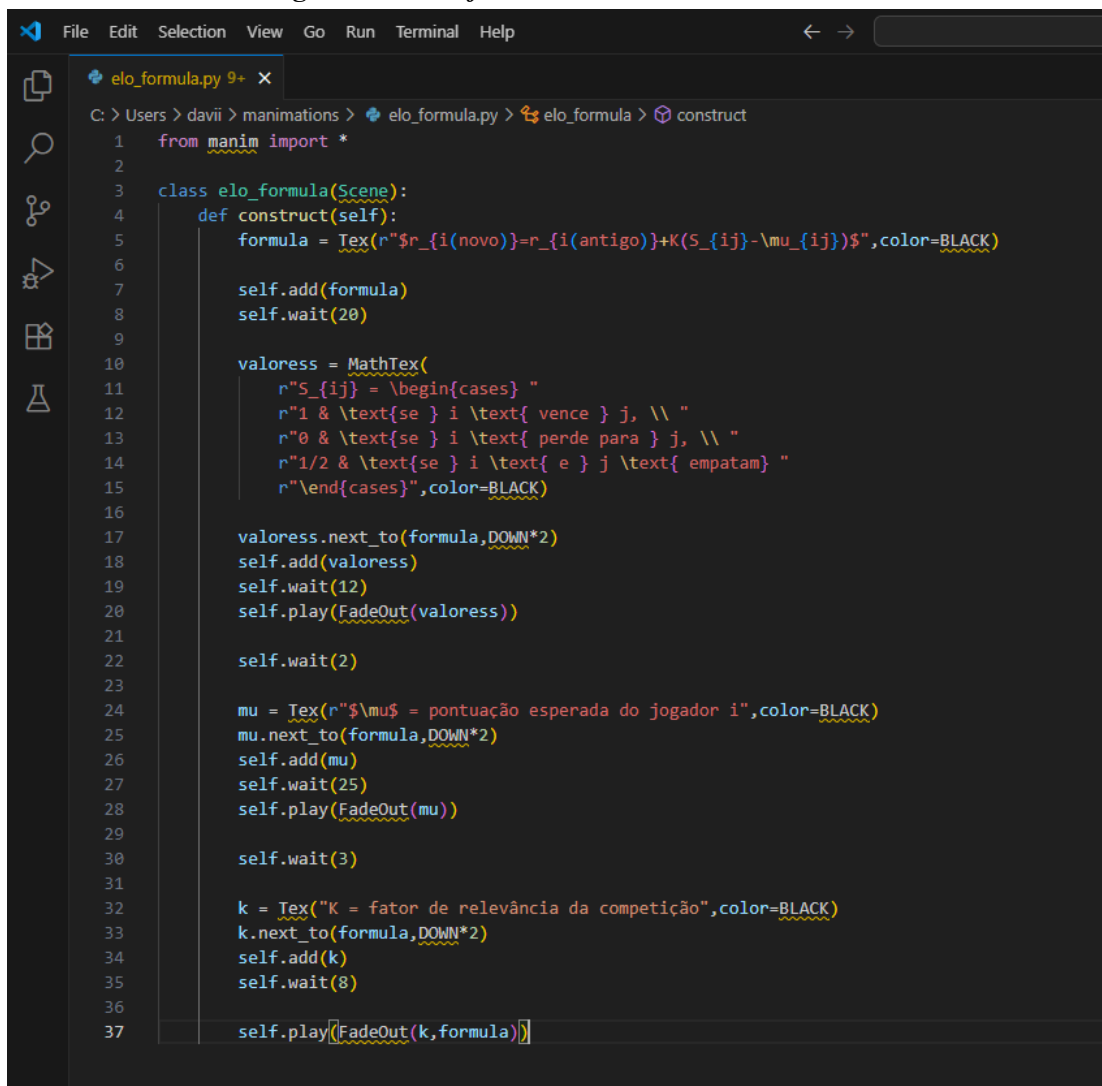


Fonte: Manim Community, 2025.

A criação de animações foi, em um primeiro momento, desafiadora. Depois de vários erros, a assimilação da sintaxe do Manim foi se tornando mais simples, e o tempo para a criação de cada animação diminuiu muito. O código foi desenvolvido no *software* Microsoft Visual Studio, como mostrado na Figura 19.

Através do Visual Studio, é possível gerar um vídeo, em formato .mp4, com o requerido no código, que é salvo automaticamente em uma pasta do computador. Porém, estes vídeos sempre são gerados com fundo preto. Para ser possível a sobreposição do vídeo gerado com a gravação realizada pelo autor, um código adicional foi utilizado para possibilitar um fundo transparente.

Figura 19 – O software Microsoft Visual Studio



```

C: > Users > davii > manimations > elo_formula.py > elo_formula > construct
1  from manim import *
2
3  class elo_formula(Scene):
4      def construct(self):
5          formula = Tex(r"$r_{i(novo)}=r_{i(antigo)}+K(S_{ij}-\mu_{ij})$",color=BLACK)
6
7          self.add(formula)
8          self.wait(20)
9
10         valoress = MathTex(
11             r"$S_{ij} = \begin{cases} "
12             r"1 & \text{se } i \text{ \text{vence } } j, \ \backslash \ "
13             r"0 & \text{se } i \text{ \text{perde para } } j, \ \backslash \ "
14             r"1/2 & \text{se } i \text{ \text{e } } j \text{ \text{empatam} } "
15             r"\end{cases}$",color=BLACK)
16
17         valoress.next_to(formula,DOWN*2)
18         self.add(valoress)
19         self.wait(12)
20         self.play(FadeOut(valoress))
21
22         self.wait(2)
23
24         mu = Tex(r"$\mu = \text{pontuação esperada do jogador } i$",color=BLACK)
25         mu.next_to(formula,DOWN*2)
26         self.add(mu)
27         self.wait(25)
28         self.play(FadeOut(mu))
29
30         self.wait(3)
31
32         k = Tex("K = fator de relevância da competição",color=BLACK)
33         k.next_to(formula,DOWN*2)
34         self.add(k)
35         self.wait(8)
36
37         self.play(FadeOut(k,formula))

```

Fonte: O Autor.

6.4.2 Inteligências artificiais

Apesar de tutoriais a respeito do Manim estarem disponíveis em diversas fontes, algumas animações podem ser realmente desafiadoras, principalmente para iniciantes. Para isso, é possível utilizar as inteligências artificiais como importante auxílio. Diversas ferramentas podem auxiliar no desenvolvimento de códigos para a criação das animações desejadas. Algumas possibilidades são o ChatGPT (<https://chatgpt.com/>), o Claude (<https://claude.ai/>) e o Gemini (<https://gemini.google.com/>). Neste trabalho, foi utilizado principalmente o Claude.

O fenômeno da utilização de inteligências artificiais para criação de código é recente, e é acompanhado do crescimento do uso de *chatbots* desde o início da década de 2020. Mesmo em ambientes profissionais, é comum que engenheiros e programadores utilizem IAs como suporte. Em um comunicado recente, por exemplo, o CEO do Alphabet (empresa dona do Google), Sundar Pichai, anunciou o seguinte.

Também estamos usando IA internamente para melhorar nossos processos de codificação, o que está aumentando a produtividade e a eficiência. Hoje, mais de um quarto de todo o novo código no Google é gerado por IA, depois revisado e aprovado pelos engenheiros. Isso ajuda nossos engenheiros a fazer mais e avançar mais rápido (Pichai, 2024).

Além do contexto profissional, o uso de IAs também é útil para aqueles sem muita experiência. O termo *vibecoding* surgiu para designar indivíduos sem conhecimento técnico de programação que, com o uso de ferramentas generativas, criam *apps* e *sites* para uso pessoal ou familiar (Roose, 2025). Em uma pesquisa com estudantes iniciantes em programação, conduzida por Kazemitabaar et al. (2023), foi obtida a seguinte conclusão:

Nossos resultados das tarefas de criação de código durante a fase de treinamento indicam que o uso de geradores de código de IA pode aumentar significativamente a conclusão de tarefas, melhorar a pontuação de correção, reduzir erros encontrados e diminuir o tempo de conclusão das tarefas. Além disso, nosso feedback qualitativo mostra que ter acesso ao gerador de código de IA reduziu o estresse e o desânimo, além de melhorar a disposição para continuar programando posteriormente (Kazemitabaar et al., 2023, p. 12-13).

O centro da utilização de um *chatbot* está na construção de um bom *prompt*, isto é, o direcionamento dado à inteligência artificial do que você quer que se faça. Uma sugestão de *prompt* para a geração de animações é a seguinte: “Sou professor de matemática e estou criando uma vídeo-aula a respeito de [TEMA]. Estou utilizando o software Manim para a criação de animações para a visualização do conceito [CONCEITO]. Me ajude na criação do código em Python para o desenvolvimento da seguinte animação: [DESCRIÇÃO DA ANIMAÇÃO]”.

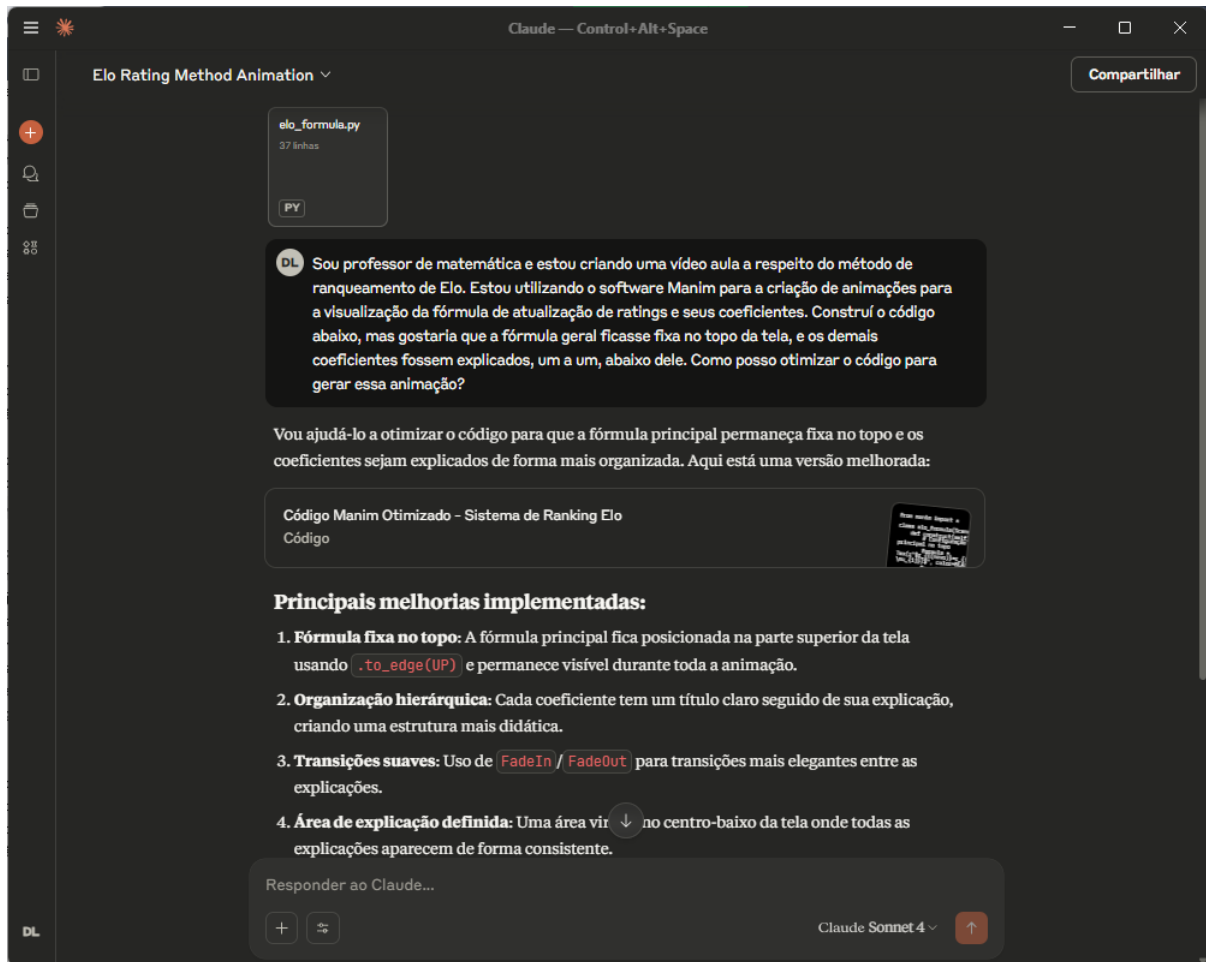
Outra possibilidade de *prompt*, no caso da criação de um código que está falhando em determinado ponto, é o seguinte: “Sou professor de matemática e estou criando uma vídeo aula a respeito de [TEMA]. Estou utilizando o software Manim para a criação de animações para a visualização do conceito [CONCEITO]. A descrição da animação desejada é [DESCRIÇÃO]. Construí o código abaixo, que apresentou os erros [ERROS]. Como posso otimizar o código para gerar a animação? [INSERIR CÓDIGO]”. Um exemplo da utilização de inteligências artificiais para criação de código Manim está na Figura 20.

Portanto, foram utilizadas inteligências artificiais neste trabalho como auxílio para o desenvolvimento de códigos de animação Manim.

6.4.3 Editores de vídeo

Um bom software de edição é essencial para dar qualidade ao vídeo, por meio do corte de partes desnecessárias, adição de trilhas de áudio, e posicionamento dos vídeos gerados no Manim. Há diversas opções gratuitas que são suficientes para o proposto neste trabalho. Utili-

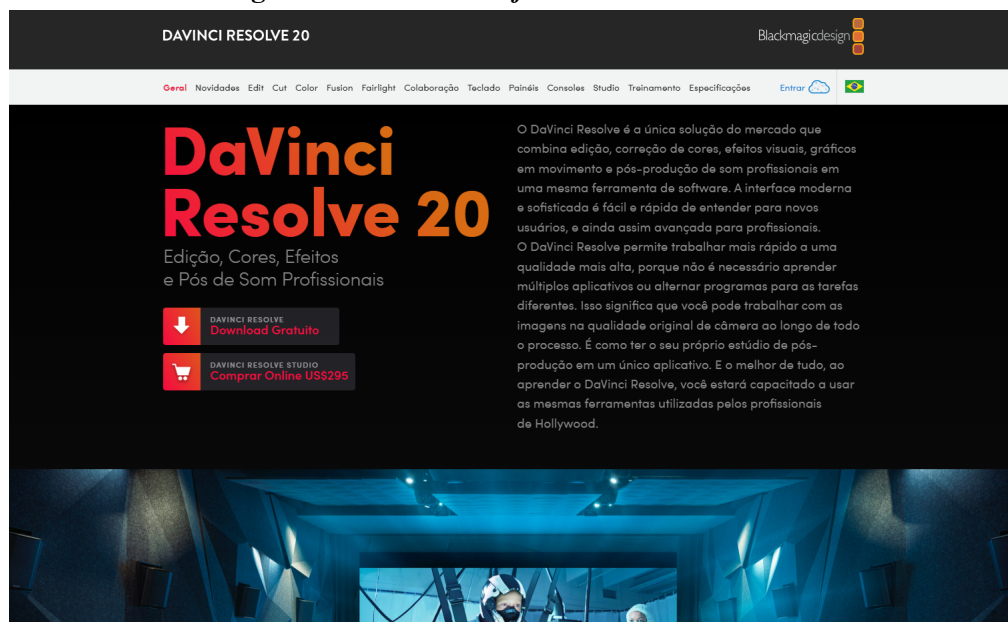
Figura 20 – Modificação de código com IA



Fonte: O Autor.

zamos neste trabalho o DaVinci Resolve, e verificamos dispor de todos os recursos essenciais, mesmo na opção gratuita. A Figura 21 mostra a página inicial do *site* de *download* do *software*.

Figura 21 – O *site* do *software* DaVinci Resolve

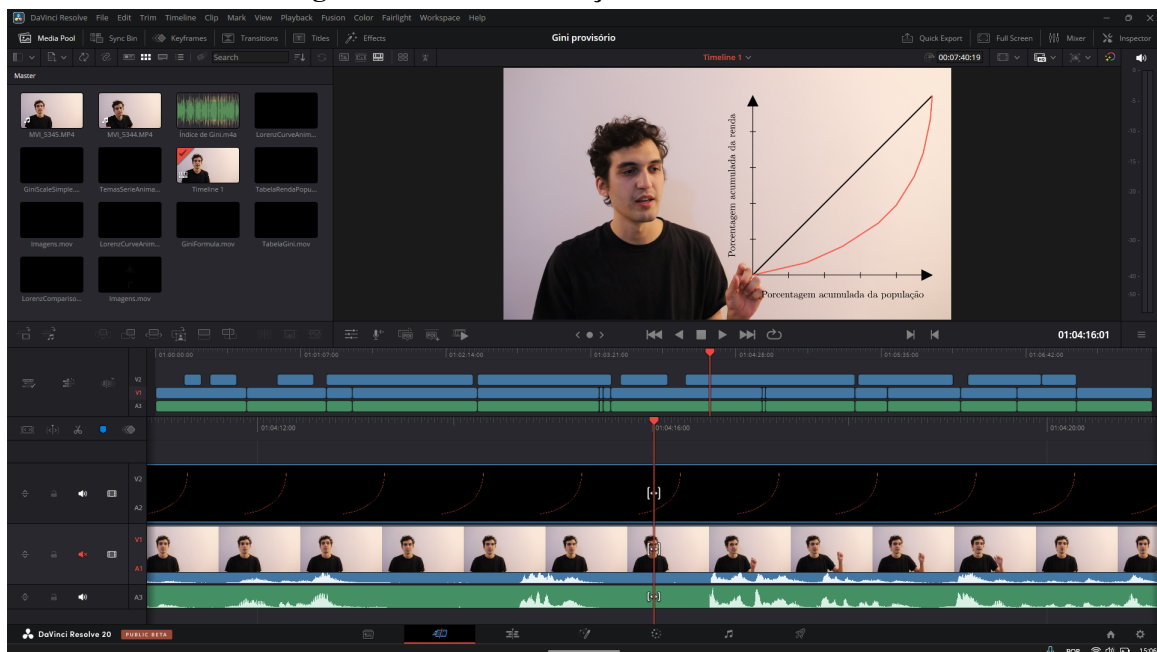


Fonte: Blackmagic Design, 2025.

O *software* é completo, e dispõe de ferramentas e recursos suficientes para produções de complexidade muito maior do que o proposto neste trabalho. Para o produto educacional produzido, foram utilizadas apenas a seção de escolha de mídias, a ferramenta de corte, e a aba *inspector*, para posicionamento de vídeos sobrepostos. O uso dessas funções foi bastante intuitivo, de forma que não foi necessário consultar guias ou tutoriais para sua utilização.

A Figura 22 mostra uma captura da tela de edição. Vale destacar a área intermediária, em que há faixas das cores azul e verde. Tais faixas representam os arquivos de vídeo da câmera, do vídeo gerado pelo Manim, e do áudio gravado pelo microfone, todos sobrepostos e alinhados para resultar no vídeo sendo mostrado ao centro. Este vídeo, por fim, foi exportado e publicado no *youtube*.

Figura 22 – A tela de edição do DaVinci Resolve



Fonte: O Autor.

6.5 Divulgação e validação do produto

Os 4 vídeos foram publicados no dia 11 de agosto de 2025. De acordo com estatísticas do *youtube*, até o fim do mesmo mês, o canal recebeu 417 visitas, possibilitando um tempo acumulado de exibição, neste período, de 15,3 horas. O vídeo mais assistido foi “O método de ranqueamento de Elo”, com 136 visualizações. Em seguida, “Como o Coeficiente de Gini quantifica a desigualdade?” teve 110 visualizações, “Como melhorar a nota do Brasil no PISA?” teve 94 visualizações e “Além da renda: como determinar o desenvolvimento humano” teve 77 visualizações.

Nos comentários e na descrição dos vídeos, foi fixado um *link* para um formulário, no qual o espectador poderia expor suas percepções acerca do vídeo, tanto em critérios técnicos quanto no seu aprendizado. Todas as respostas foram voluntárias e anônimas.

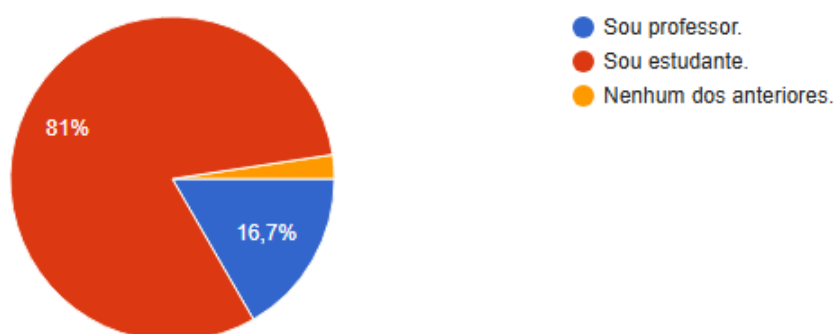
O formulário foi composto de 4 perguntas. A primeira trazia uma caracterização do espectador, entre professor, estudante ou nenhum dos dois. A segunda pergunta questionava o conhecimento prévio sobre os temas apresentados nos vídeos, classificados em nenhum, moderado, satisfatório, muito bom ou excelente. A terceira pergunta pedia uma avaliação entre discordo totalmente, discordo, não sei, concordo ou concordo totalmente, para as seguintes sentenças: a produção técnica foi satisfatória (som, imagem, iluminação, edição); o conteúdo foi bem apresentado; pude aprender algo com esta série; os vídeos despertaram um interesse meu pelo assunto; eu recomendaria este vídeo para um estudante ou colega. A quarta pergunta era opcional e aberta para feedbacks adicionais.

Ao todo, foram registradas 42 respostas. Quanto ao perfil dos espectadores, percebe-se que a grande maioria é de estudantes, como visto na Figura 23. O autor deste trabalho tem, no momento da escrita, 15 turmas de educação financeira do ensino médio. Assim, os estudantes atingiram facilmente a maioria das respostas.

Figura 23 – Perfil dos espectadores dos vídeos

Qual opção melhor te descreve?

42 respostas



Fonte: O Autor.

Como índices socioeconômicos são frequentemente apresentados em notícias e meios de comunicação, houve prevalência de declaração de conhecimento prévio moderado, satisfatório e excelente. Ainda assim, mais de um quarto dos espectadores alegou conhecimento nulo dos assuntos antes de assistir aos vídeos. A Figura 24 mostra esses resultados.

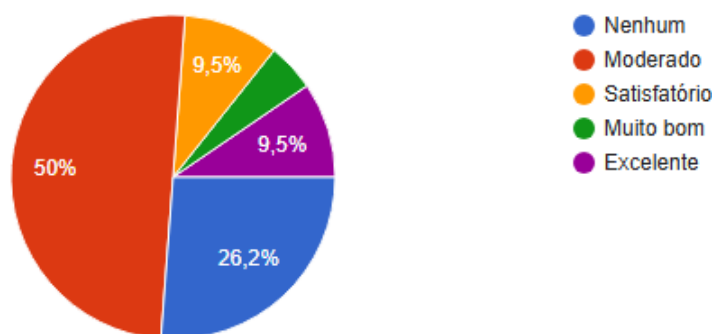
A questão de percepções sobre os vídeos trouxe respostas mistas. A maioria dos questionamentos foi respondida com “concordo totalmente”, enquanto “concordo” foi a segunda resposta mais escolhida. A opção “discordo totalmente” foi respondida apenas uma vez para cada pergunta. Os resultados são mostrados na Figura 25.

A última pergunta do formulário pedia feedbacks adicionais. Dentre as respostas obtidas, destacam-se as seguintes: “O conjunto de vídeos traz de um modo geral um conhecimento de fácil acesso e uma forma de explicação muito boa, é fácil perceber que o professor tem conhecimento sobre os assuntos aplicados”; “Achei ótimo, só podia falar um pouco mais rápido! Assisti tudo em 1.5x.”; “Vídeos muito bons que explicam o conteúdo facilmente, sendo per-

Figura 24 – Conhecimento prévio dos espectadores

Qual era o seu nível de conhecimento prévio a respeito dos temas apresentados nos vídeos?

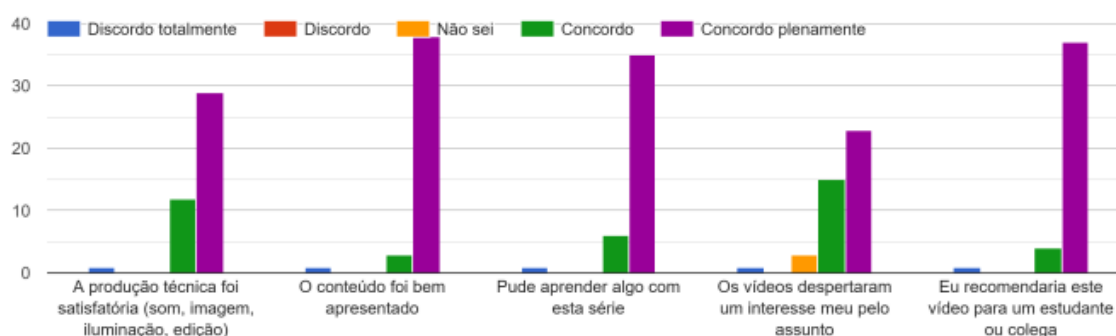
42 respostas



Fonte: O Autor.

Figura 25 – Avaliação da série de vídeos

Avalie a série de vídeos de acordo com os seguintes critérios.



Fonte: O Autor.

ceptível o conhecimento amplo do professor sobre a matéria aplicada”; “Excelente produção de mestrado. Quando podemos quantificar alguns índices para aproximar de um padrão, é ótimo. Ajuda muito o profissional qualquer área relacionada com esses índices.”; “Trabalho excepcional que mostra a importância e relação dos índices socio-economicos para pautar, por exemplo, as políticas públicas de educação. Seria importante que o mesmo fosse compartilhado com o maior número de pessoas ligadas a essas políticas e porque não gerar eventos como palestras, oficinas, etc sobre este tema com o objetivo de esclarecer e atualizar os conceitos envolvidos tanto na academia quanto para nossos estudantes da educação básica. Parabéns Professor Davi, aprendi muito e espero compartilhar com meus estudantes.”.

As estatísticas disponibilizadas pelo *youtube* permitem outras inferências, que o autor também reconheceu em sua prática. Ao apresentar o trabalho para estudantes (que são a maior parte dos espectadores), foi evidente um tempo de foco menor do que o esperado. O tempo de fixação em cada vídeo era curto demais, e alguns responderam ao formulário mesmo sem

visualizar todos os vídeos. A retenção dos espectadores está evidenciada na Tabela 21 e na Figura 26.

Tabela 21 – Duração média das visualizações

Vídeo	Duração média da visualização
Como o coeficiente de Gini quantifica a desigualdade?	1 : 44 (22,7%)
Além da renda: como determinar o desenvolvimento humano	1 : 41 (22%)
Como melhorar a nota do Brasil no PISA?	4 : 32 (65,2%)
O método de ranqueamento de Elo	1 : 14 (13,1%)

Fonte: O Autor.

Figura 26 – Retenção no vídeo a respeito do método de Elo



Fonte: O Autor.

É interessante notar que o vídeo com maior tempo médio de visualização é a respeito das notas do Brasil no PISA. É provável que este tema interesse mais a professores e público externo, e não seja tão atrativo para adolescentes. Esse público apresentou tempo de foco maior, retendo os espectadores no vídeo por mais tempo. Por outro lado, o método de Elo, mais ligado ao xadrez, pode chamar mais a atenção de um público jovem, que acessou o vídeo em maior quantidade, mas não apresentou retenção alta, apesar de dentro das médias da plataforma.

Sendo assim, vale considerar, para trabalhos futuros, a produção de menos vídeos, ou de vídeos de menor duração, de 1 a 5 minutos. Falas mais dinâmicas e diretas também podem ser consideradas. Apesar de limitar o tempo para desenvolvimento dos temas, tais modificações permitiriam reter a atenção de jovens e adolescentes por mais tempo.

7 CONCLUSÃO

É comum, no exercício da profissão de ensino de matemática, ouvir estudantes questionando o propósito do estudo de determinados tópicos. Afinal, qual é a necessidade de estudar matrizes ou funções no Ensino Médio? Para quê eles servem? Este trabalho buscou explorar princípios matemáticos utilizados no ranqueamento de itens e, assim, trouxe uma possível resposta para essas perguntas.

Em primeiro lugar, foi necessário aprofundar os conhecimentos de álgebra linear. Uma vez que matrizes singulares não possuem inversa, é possível determinar soluções de um sistema a partir das inversas generalizadas. Maior destaque foi dado à inversa de Moore-Penrose, a pseudoinversa, a qual apresenta a solução de norma euclidiana mínima, além de ser útil para a solução com erro mínimo de sistemas lineares sobredeterminados.

Outros resultados interessantes deste capítulo se deram na análise das Equações Normais, ou seja, aqueles sistemas em que $Ax = b$ são adaptados para $A^T Ax = A^T b$. Neste contexto, este trabalho se deparou com matrizes quadradas $A^T A$ que não possuíam inversa, o que não é usual. Assim, foi mostrado no Capítulo 4, que a sugestão de Myers e Milton (1991) e Massey (1997), da reparametrização do vetor b , leva a soluções idênticas à encontrada por meio da pseudoinversa. Mostrou-se que soluções de mínimos quadrados do sistema $Ax = b$, com a matriz A no formato explorado neste trabalho, são translações pelo vetor $\vec{1}$, do que decorre a prova de que a solução pela reparametrização tem norma euclidiana mínima. Destaca-se que tal propriedade não foi encontrada em nenhuma das referências consultadas, e sua demonstração é uma contribuição original deste trabalho.

Com a base desenvolvida no capítulo teórico, e de forma alinhada com orientações da BNCC, foram observados o IDH e o Índice de Gini, buscando revelar como se determina a pontuação atribuída a cada país, e o que ocorre no contexto brasileiro. Um destaque deste capítulo é o desenvolvimento sobre a área de concentração determinada entre a curva de Lorenz e a reta de perfeita igualdade. Considera-se relevante que estudantes e professores de matemática conheçam a relação de seu conteúdo com a realidade social da qual fazem parte, reconhecendo que diversas políticas públicas são decididas a partir de resultados matemáticos.

Destaca-se, ainda, o ranqueamento estabelecido a partir da avaliação do PISA. Inicialmente, esperava-se determinar com mais detalhes o processo de elaboração destes *ratings*, porém, a metodologia utilizada pelo programa mostrou-se bastante elaborada, com métodos estatísticos avançados. Seria necessário um foco maior apenas na avaliação para destrinchar melhor seus resultados. Ainda assim, as regressões lineares encontradas entre os resultados na avaliação do PISA e os demais índices estudados mostram como a educação é multidimensional, e como renda, padrão de vida, desigualdade e saúde também são importantes para se ter uma educação de qualidade.

O método de Massey foi dividido, principalmente, em duas seções: uma para ranqueamentos em campeonatos em que todos enfrentam-se entre si, em números iguais de jogos, e no caso em que times possuem números de partidas diferentes. Em ambos os casos, as noções

do Capítulo 2 foram essenciais. Uma conclusão não esperada, para o caso de números iguais de jogos, é que o ranqueamento obtido pelo método é o mesmo obtido pela análise do saldo de pontos da equipe no campeonato, de forma que a utilização de Massey é mais útil para a determinação de ratings do que de rankings. Algumas inovações recentes e possibilidades de adaptação foram descritas na seção final.

A respeito do método de ranqueamento de Elo, buscou-se conhecer bem a equação logística de Verhulst, base para a definição do coeficiente de pontuação esperada do método. Esta variável foi explorada e manipulada para se compreender suas propriedades, e diversos exemplos foram analisados. Considera-se este método mais acessível aos estudantes de ensino médio (por não depender de resultados de álgebra linear), apesar de poder ser mais elaborado a partir de conhecimentos de cálculo.

Para a produção da pesquisa, utilizaram-se o *software* de edição de texto *LaTeX*, e para a produção de gráficos o *GeoGebra* e a linguagem de programação Julia.

Como produto educacional, foram produzidos quatro vídeos explicativos dos seguintes temas: as bases matemáticas Índice de Gini; a forma de cálculo do IDH; as relações entre tais índices e o resultado da prova PISA; e o método de Elo. Para a produção dos vídeos, foram utilizados os *softwares* DaVinci Resolve, Manim, Microsoft Visual Studio e Claude. Os vídeos foram publicados na plataforma *youtube*, e foram coletados *feedbacks* de professores, colegas de mestrado, estudantes e comunidade geral.

Alguns conceitos desenvolvidos neste trabalho, relacionados ao Capítulo 3, foram apresentados no III Workshop Nacional Online do PROFMAT, realizado virtualmente no mês de agosto de 2025. Foi feita uma comunicação de 15 minutos apresentando os fundamentos matemáticos do Índice de Gini e do IDH. Depois, possíveis relações entre esses índices e os resultados na avaliação PISA foram expostas, por meio de regressões lineares. O trabalho está publicado nos anais do evento, sob o ISBN 978-65-01-66422-4 (Iacomini e Sánchez, 2025).

O PROFMAT contribuiu grandemente para o desenvolvimento profissional do autor, apresentando em suas disciplinas, obrigatórias e optativas, fundamentos úteis para a prática de sala de aula. Além disso, esta dissertação de mestrado foi relevante para suscitar questionamentos para o autor a respeito dos mais variados usos do ranqueamento, desde questões políticas e educacionais até esportes e entretenimento. Por fim, o programa confere crescimento pessoal, através do desafio constante, que garante desenvolvimento de habilidades e interesses.

É recomendada a continuidade de estudos a respeito de avaliação e ranqueamento, principalmente no contexto educacional. Provas como o SAEB e o PISA já se tornaram padrão para a avaliação de instituições de ensino e sistemas educacionais. Considera-se relevante não apenas determinar a forma de obtenção destas notas, mas também como os seus resultados impactam decisões institucionais e políticas educacionais em larga escala. Também pode ser questionada a pergunta simétrica: que impacto decisões governamentais têm em resultados de avaliações externas?

Espera-se que os conhecimentos desenvolvidos neste trabalho impactem a prática do autor e de outros professores que assistirem aos vídeos do produto, contribuindo para uma

educação contextualizada e significativa para o estudante, tanto na disciplina de Matemática quanto na disciplina de Educação Financeira.

REFERÊNCIAS

- ALMEIDA, P. H. A. de; PORFIRO, L. D.; PEREIRA, T. de L. B. A equação de verhulst: estudo e modelagem de uma equação diferencial ordinária para a dinâmica populacional de anápolis. **Revista Mirante**, v. 17, n. 1, p. 300–316, 2024. Disponível em: <https://www.revista.ueg.br/index.php/mirante/article/view/14536>. Acesso em: 1 maio 2025.
- ANDREWS, P. *et al.* **OECD and Pisa tests are damaging education worldwide - academics**. Londres, 2014. Disponível em: <https://www.theguardian.com/education/2014/may/06/oecd-pisa-tests-damaging-education-academics>. Acesso em: 20 fev. 2025.
- ANTON, H.; BUSBY, R. C. **Álgebra linear contemporânea**. Porto Alegre: Bookman, 2007. 602 p.
- ASSOCIATION OF TENNIS PROFESSIONALS. **PIF ATP rankings (singles)**. Londres, 2024. Disponível em: <https://www.atptour.com/en/rankings/singles>. Acesso em: 1 dez. 2024.
- BACAËR, N. **A short history of mathematical population dynamics**. Londres: Springer, 2011. 160 p.
- BATTI, J. ao C. B. **Um pouco da matemática usada no algoritmo PageRank do Google**. 2015. 63 f. Dissertação (Mestrado) — Universidade Federal de Santa Catarina Florianópolis, Santa Catarina 2015.
- Blackmagic Design. **DaVinci Resolve 20**. Melbourne, 2025. Disponível em: <https://www.blackmagicdesign.com/br/products/davinciresolve>. Acesso em: 17 jun. 2025.
- BRASIL. **Base Nacional Comum Curricular**. Brasília, 2018. Disponível em: https://www.gov.br/mec/pt-br/escola-em-tempo-integral/BNCC_EI_EF_110518_versaofinal.pdf. Acesso em: 10 out. 2024.
- CHARTIER, T. P. *et al.* Reducing the effects of unequal number of games on rankings. **IMAGE**, , n. 52, p. 15–23, 2014. Disponível em: <https://ilasic.org/wp-content/uploads/IMAGE/image52.pdf>. Acesso em: 24 nov. 2024.
- CHARTIER, T. P. *et al.* Sports ranking with nonuniform weighting. **Journal of quantitative analysis in sports**, v. 7, n. 3, p. 16, 2011.
- COMPUTER CHESS RATING LISTS. **CCRL 40/15**. s.l., 2025. Disponível em: <https://computerchess.org.uk/ccrl/4040/>. Acesso em: 23 abr. 2025.
- COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR. **Portaria nº 207, de 4 de Julho de 2024**. Brasília, 2024. Disponível em: <https://cad.capes.gov.br/ato-administrativo-detalhar?idAtoAdmElastic=15731#anchor>. Acesso em: 8 ago. 2025.
- COSTA, M. V. S. da. **Noções de economia para ambientes de educação financeira escolar: reflexões sobre desenvolvimento humano e distribuição de renda por meio dos índices de GINI e IDH**. 2022. 102 f. Dissertação (Mestrado) — Colégio Pedro II Rio de Janeiro, Rio de Janeiro 2022.
- DORFMAN, R. A formula for the gini coefficient. **The review of economics and statistics**, v. 61, n. 1, p. 146–149, 1979.

- FONSECA, J. J. S. da. **Metodologia da pesquisa científica**. Fortaleza, 2002. Disponível em: <http://www.ia.ufrj.br/ppgea/conteudo/conteudo-2012-1/1SF/Sandra/apostilaMetodologia.pdf>. Acesso em: 9 nov. 2024.
- FRANCESCHET, M.; BOZZO, E.; VIDONI, P. The temporalized massey's method. **Journal of quantitative analysis in sports**, v. 13, n. 2, p. 37–48, 2017.
- FÉDÉRATION INTERNATIONALE DE FOOTBALL ASSOCIATION. **FIFA/Coca-Cola men's world ranking procedures**. Zurique, 2018. Disponível em: <https://inside.fifa.com/fifa-world-ranking/procedure-men>. Acesso em: 10 abr. 2025.
- FÉDÉRATION INTERNATIONALE DES ÉCHECS. **Calculators for chess rating**. Lausanne, 2025. Disponível em: <https://ratings.fide.com/calc.phtml?page=change>. Acesso em: 17 abr. 2025.
- FÉDÉRATION INTERNATIONALE DES ÉCHECS. **Rating - Carlsen, Magnus**. Lausanne, 2025. Disponível em: <https://ratings.fide.com/profile/1503014/chart>. Acesso em: 5 ago. 2025.
- FÉDÉRATION INTERNATIONALE DES ÉCHECS. **Rating calculations - Leitao, Rafael BRA**. Lausanne, 2024. Disponível em: https://ratings.fide.com/calculations.phtml?id_number=2101246&period=2024-11-01&rating=0. Acesso em: 17 abr. 2025.
- GASTWIRTH, J. L. The estimation of the lorenz curve and gini index. **The review of economics and statistics**, v. 54, n. 3, p. 306–316, 1972.
- GIL, A. C. **Como elaborar projetos de pesquisa**. São Paulo: Editora atlas, 2002. 176 p.
- GOLUB, G. H.; LOAN, C. F. van. **Matrix computations**. Londres: The johns hopkins university press, 1996. 694 p.
- HOFFMANN, R. **Estatística para economistas**. São Paulo: Pioneira Thomson Learning, 2006. 432 p.
- IACOMINI, D. L.; BÁEZ-SÁNCHEZ, A. D. Alguns aspectos matemáticos Índice de gini, Índice de desenvolvimento humano e notas pisa. *In: Anais do III Workshop Nacional Online do Profmat*. Rio de Janeiro: Sociedade Brasileira de Matemática, 2025. Disponível em: <https://sites.google.com/view/iii-workshop-do-profmat/anais-e-certificados-2025>. Acesso em: 3 set. 2025.
- INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Exame Nacional do Ensino Médio (Enem)**. Brasília, 2025. Disponível em: <https://www.gov.br/inep/pt-br/aceso-a-informacao/perguntas-frequentes/exame-nacional-do-ensino-medio-enem>. Acesso em: 9 out. 2025.
- INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Nota informativa do IDEB 2023**. Brasília, 2023. Disponível em: https://download.inep.gov.br/ideb/nota_informativa_ideb_2023.pdf. Acesso em: 13 nov. 2024.
- INTERNATIONAL HANDBALL FEDERATION. **Teams & Groups**. Basileia, 2024. Disponível em: <https://www.ihf.info/competitions/women/307/2024-olympic-games-paris-womens-tournament/169543/groups-rankings>. Acesso em: 22 jul. 2025.
- JANNUZZI, P. de M. **Indicadores sociais no Brasil: conceitos, fontes de dados e aplicações**. Campinas: Alínea, 2001. 72 p.
- KAZEMITABAAR, M. *et al.* Studying the effect of ai code generators on supporting novice learners in introductory programming. **Proceedings of the 2023 CHI Conference**

on **Human Factors in Computing Systems**, , n. 455, p. 1–23, 2023. Disponível em: <https://dl.acm.org/doi/abs/10.1145/3544548.3580919>. Acesso em: 18 jun. 2025.

LANGVILLE, A. N.; MEYER, C. D. **Who's #1?: the science of rating and ranking**. New Jersey: Princeton University Press, 2012. 247 p.

Manim Community. **Manim**. s.l., 2025. Disponível em: <https://www.manim.community>. Acesso em: 17 jun. 2025.

MASSEY, K. **Statistical models applied to the rating of sports teams**. 1997. 84 f. Monografia (Honors Project) — Bluefield College Virginia 1997.

MENEGOTTO, F. N. **O YouTube como ferramenta de democratização da divulgação das ciências**. 2021. 118 f. Dissertação (Mestrado) — Universidade de Lisboa Lisboa, Portugal 2021.

MONTEIRO, M. A. A matemática do Índice de desenvolvimento humano – IDH. **Revista do Professor de Matemática**, , n. 67, p. 31–35, 2008. Disponível em: <https://rpm.org.br/cdrpm/67/8.html>. Acesso em: 18 nov. 2024.

MYERS, R. H.; MILTON, J. S. **A first course in the theory of linear statistical models**. Boston: PWS-KENT Publishing Company, 1991. 352 p.

NETO, P. L. de O. C. **Estatística**. São Paulo: Editora Blucher, 2006. v. 2 ed. 281 p.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT . **PISA 2022 results (Volume I): The state of learning and equity in education**. Paris: , 2023. Disponível em: https://www.oecd.org/en/publications/pisa-2022-results-volume-i_53f23881-en/full-report.html. Acesso em: 19 jan. 2025.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT . **PISA 2022 technical report**. Paris: , 2023. Disponível em: https://www.oecd.org/en/publications/pisa-2022-technical-report_01820d6d-en.html. Acesso em: 19 jan. 2025.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS. **Human Development Insights**. Nova York, 2025. Disponível em: <https://hdr.undp.org/data-center/country-insights#/ranks>. Acesso em: 19 mar. 2025.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS . **Relatório de desenvolvimento humano 2021-22: Tempos incertos, vidas instáveis: construir o futuro num mundo em transformação**. Nova York: , 2023. Disponível em: <https://www.undp.org/pt/brazil/desenvolvimento-humano/publications/relatorio-de-desenvolvimento-humano-2021-22>. Acesso em: 30 jan. 2025.

PARANÁ. **Paraná tem a melhor educação do Brasil no ranking geral do Ideb**. Curitiba, 2024. Disponível em: <https://www.aen.pr.gov.br/Noticia/Parana-tem-melhor-educacao-do-Brasil-no-ranking-geral-do-Ideb>. Acesso em: 4 out. 2024.

PEREIRA, L. S. **Matrizes de Markov: o Teorema de Perron-Frobenius; PageRank e outras aplicações**. 2019. 68 f. Dissertação (Mestrado) — Universidade de Brasília Brasília, Distrito Federal 2019.

PICHAU, S. **Q3 earnings call: CEO's remarks**. Califórnia, 2024. Disponível em: <https://blog.google/inside-google/message-ceo/alphabet-earnings-q3-2024/#full-stack-approach>. Acesso em: 18 jun. 2025.

RANKINGS, Q. W. U. **Latin America & The Caribbean 2025**. S.l., 2024. Disponível em: [hhttps://www.topuniversities.com/latin-america-caribbean-overall](https://www.topuniversities.com/latin-america-caribbean-overall). Acesso em: 11 out. 2024.

ROOSE, K. **Not a coder? With A.I., just having an idea can be enough.** Nova York, 2025. Disponível em: <https://www.nytimes.com/2025/02/27/technology/personaltech/vibecoding-ai-software-programming.html>. Acesso em: 18 jun. 2025.

SAMPAIO, A. **Brasil tem melhora no IDH, mas retrocede em ranking da ONU em 2022.** São Paulo, 2024. Disponível em: <https://www.cnnbrasil.com.br/economia/macroeconomia/brasil-tem-melhora-no-idh-mas-retrocede-em-ranking-da-onu-em-2022/>. Acesso em: 4 out. 2024.

STEWART, J.; CLEGG, D.; WATSON, S. **Cálculo, volume I.** São Paulo: Cengage Learning, 2022. 708 p.

STEWART, J.; CLEGG, D.; WATSON, S. **Cálculo, volume II.** São Paulo: Cengage Learning, 2022. 740 p.

STRANG, G. **Álgebra linear e suas aplicações.** São Paulo: Cengage Learning, 2013. 445 p.

WORLD BANK GROUP. **Gini index.** Washington, 2025. Disponível em: https://data.worldbank.org/indicator/SI.POV.GINI?most_recent_value_desc=false. Acesso em: 19 mar. 2025.

WORLD BANK GROUP. **Gini index - Brazil.** Washington, 2025. Disponível em: <https://data.worldbank.org/indicator/SI.POV.GINI?end=2022&locations=BR&start=1990&view=chart>. Acesso em: 19 mar. 2025.

WORLD BANK GROUP. **Metadata Glossary: Gini index.** Washington, s.d. Disponível em: <https://databank.worldbank.org/metadataglossary/gender-statistics/series/SI.POV.GINI>. Acesso em: 20 jan. 2025.

WORLD BANK GROUP. **World development indicators: distribution of income or consumption.** Washington, s.d. Disponível em: <https://wdi.worldbank.org/table/1.3>. Acesso em: 25 mar. 2025.