



UNIVERSIDADE FEDERAL DO PIAUÍ
CENTRO DE CIÊNCIAS DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA
MESTRADO PROFISSIONAL EM MATEMÁTICA

Antonio de Alencar Costa Silva

Estudo de caso sobre análise de rendimento acadêmico:
uma abordagem multivariada

Teresina - 2024



Antonio de Alencar Costa Silva

Dissertação de Mestrado:

**Estudo de caso sobre análise de rendimento acadêmico: uma
abordagem multivariada**

Dissertação submetida à Coordenação do Programa de Mestrado Profissional em Matemática - Profmat, da Universidade Federal do Piauí, como requisito parcial para obtenção do grau de Mestre em Matemática na modalidade profissional.

Orientadora:

Prof. Dra. Jackelya Araujo da Silva

Teresina - 2024

Copyright © 2024 by Antonio de Alencar Costa Silva.

Direitos reservados, 2024 por Antonio de Alencar Costa Silva.

Universidade Federal do Piauí - UFPI, Centro de Ciência da Natureza - CCN, Programa de Pós-Graduação em Matemática, Mestrado Profissional em Matemática. Cep 64049-550 - Teresina, PI.

Nenhuma parte desta dissertação pode ser reproduzida sem a expressa autorização do autor.

FICHA CATALOGRÁFICA
Universidade Federal do Piauí
Sistema de Bibliotecas UFPI - SIBi/UFPI
Biblioteca Setorial do CCN

S586e Silva, Antonio de Alencar Costa.
Estudo de caso sobre análise de desempenho acadêmico:
uma abordagem multivariada / Antonio de Alencar Costa
Silva. -- 2024.
43 f. : il.

Dissertação (Mestrado Profissional) - Universidade
Federal do Piauí, Centro de Ciências da Natureza,
Programa de Pós-Graduação em Matemática, Teresina,
2024.
“Orientadora: Prof^ª. Dr^ª. Jackelya Araújo da Silva.”

1. Educação. 2. Matemática - estudo e ensino. 3. Índice
de desempenho. 4. Metodologia *K-means*. I. Silva, Jackelya
Araújo da. II. Título.

CDD 371.218

Bibliotecária: Caryne Maria da Silva Gomes - CRB3/1461



PROFMAT



UNIVERSIDADE FEDERAL DO PIAUÍ
CENTRO DE CIÊNCIAS DA NATUREZA
CENTRO DE EDUCAÇÃO ABERTA E À DISTÂNCIA
MESTRADO PROFISSIONAL EM MATEMÁTICA EM REDE NACIONAL



Dissertação de Mestrado submetida à coordenação Acadêmica Institucional, na Universidade Federal do Piauí, do Programa de Mestrado Profissional em Matemática em Rede Nacional para obtenção do grau de mestre em matemática intitulada: “**Estudo de caso sobre análise de rendimento acadêmico: Uma abordagem multivariada**”, defendida pelo mestrando Antonio de Alencar Costa Silva, em 22 de julho de 2024 e aprovada pela banca constituída pelas professoras:

Documento assinado digitalmente



JACKELYA ARAUJO DA SILVA
Data: 05/08/2024 12:07:16-0300
Verifique em <https://validar.iti.gov.br>

Profa Dra. Jackelya Araujo da Silva
Presidente da Banca examinadora

Documento assinado digitalmente



DEBORA BORGES FERREIRA
Data: 05/08/2024 18:11:31-0300
Verifique em <https://validar.iti.gov.br>

Profa Dra. Débora Borges Ferreira
Examinador Externo à Instituição

Documento assinado digitalmente



CLEIDE MAYRA MENEZES LIMA
Data: 05/08/2024 21:25:51-0300
Verifique em <https://validar.iti.gov.br>

Profa Dra. Cleide Mayra Menezes Lima
Examinador Externo ao PROFMAT

Documento assinado digitalmente



VALMARIA ROCHA DA SILVA FERRAZ
Data: 06/08/2024 10:24:32-0300
Verifique em <https://validar.iti.gov.br>

Profa Dra. Valmária Rocha da Silva Ferraz
Examinador Interno

Antonio de Alencar Costa Silva

**Estudo de caso sobre análise de rendimento acadêmico: uma
abordagem multivariada**

Dissertação submetida à banca examinadora
abaixo discriminada em defesa pública e aprovada
em 22 de julho de 2024.

BANCA EXAMINADORA

Prof. Dra. Jackelya Araujo da Silva (Orientadora)

Universidade Federal do Piauí (PROFMAT)

Profa. Dra. Débora Borges Ferreira

Universidade Federal do Rio Grande do Norte

Profa. Dra. Cleide Mayra Menezes Lima

Universidade Federal do Piauí (Coordenação do curso de Estatística)

Profa. Dra. Valmária Rocha da Silva Ferraz

Universidade Federal do Piauí (PROFMAT)

Teresina - 2024

Dedico esta dissertação ao Criador

Agradecimentos

Agradeço,

Ao Criador.

À minha esposa e aos meus filhos, pelo conforto e companheirismo que me proporcionam e, em especial, ao meu filho, Benedito Paulo, pelas contribuições no uso do LATEX.

Aos meus pais, por todas as lições que ainda reviso, àqueles que agiram como se fossem meus pais, muitas vezes sem saber de mim, àqueles que agem como se fossem meus genitores, às minhas irmãs e familiares, pelas contribuições diretas ou indiretas à minha vida. Ao professor Rocha (*in memoriam*), pela deferência e amizade, ao professor Paulo Jorge (*in memoriam*) e família, pela deferência e contribuições para a vida, ao professor João Ernesto, que ouviu e ajudou o garoto andarilho, ao professor João Benício (*in memoriam*), que me serviu em Teresina e insistiu para eu continuasse, ao professor Luiz Júnior que extrapolou o horário de trabalho para me atender.

Aos familiares que me acompanharam nos primeiros passos em Parnaíba, especialmente à família dos meus padrinhos, senhor José (*in memoriam*) e senhora Dolores.

Ao Paulo Tadeu e à Carmem, que em plena mocidade me serviram, à senhora Dolores, minha madrinha, pelos cuidados de mãe, ao senhor José Oliveira (*in memoriam*), meu padrinho, por ensinar-me a descobrir meu propósito de vida e a me tornar mais leve a cada dia. Gratidão!

Aos amigos Gervásio Barbosa, Camilo França, Jailson Ramos, Wilfred, Francisco José, Francisco das Chagas, João Pedro, dentre outros, pelo apoio, incentivo e motivação ao longo desses anos. Em particular, ao Edson Nascimento, que agiu incondicionalmente em um dia determinado de nossas vidas, o que me permite escrever esta dissertação com mais tranquilidade.

Ao Instituto Federal do Piauí, IFPI, através da Reitoria, Direção-Geral e de Ensino, Campus Pedro II, pelas condições oferecidas para o afastamento e acesso aos dados que foram base da geração deste trabalho.

Aos meus professores que sempre me ensinaram a prosseguir, em especial os do PROFMAT/UFPI, nas aulas e coordenações, e, particularmente, à Prof^a. Dra. Jackelya Araujo da Silva, que, de forma incansável, contribuiu para que eu concluísse o mestrado.

Aos colegas da turma PROFMAT/UFPI de 2022 que contribuíram para a aceleração do meu aprendizado, seja nos exercícios, seja no uso do LATEX.

Às professoras, membros convidados para a banca examinadora, pelas críticas e sugestões apresentadas a este trabalho.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES, à Universidade Federal do Piauí, UFPI, à Sociedade Brasileira de Matemática, SBM, pelo oferecimento deste curso, bem como pelo apoio financeiro da CAPES.

*“Não são tanto as coisas que não sabemos
que nos metem em confusões. São as coi-
sas que pensamos que sabemos”.*

Artemus Ward

Lista de Figuras

3.1	Boxplot dos coeficientes de rendimentos segundo os cursos técnicos A, B e C	16
3.2	Boxlot dos coeficientes de rendimentos segundo a escola de origem	17
3.3	Boxplot dos coeficientes de rendimentos segundo <i>status</i> de beneficiário . . .	18
3.4	Análise dos coeficientes de correlações entre as disciplinas: (A) - Turma 01, (B) - Turma 02 e (C) - Turma 03.	19
3.5	Apresentação do agrupamento para os dados referentes a turma 01.	21
3.6	Apresentação do agrupamento para os dados referentes a turma 02.	21
3.7	Apresentação do agrupamento para os dados referentes a turma 03.	22
3.8	Gráfico Biplot para a turma 01 - relações entre disciplinas e estudantes . .	25
3.9	Gráfico Biplot para a turma 02 - relações entre disciplinas e estudantes . .	26
3.10	Gráfico Biplot para a turma 03 - relações entre disciplinas e estudantes . .	26

Resumo

Este trabalho é uma proposta de metodologia para auxiliar o gestor na tomada de decisão para a construção de mecanismos de apoio ao estudante que vão além dos aspectos intelectuais. A metodologia *K-means* possibilitou explicar a variabilidade das notas nas turmas analisadas, identificou a formação de um número reduzido de estudantes que necessitam de acompanhamento educacional com o propósito de melhoria da vida acadêmica, permitiu observar a presença de pequenos grupos de notas que influenciaram positivamente na média geral da turma. Assim, com a utilização da metodologia e diante das avaliações institucionais, os profissionais envolvidos no processo de ensino e aprendizagem poderão tomar decisões acertadas com relação à melhoria dos índices da instituição da qual fazem parte e direcionar as melhores ações no sentido de maximizar os índices avaliativos em favor do bem-estar dos estudantes.

Palavras-chave: Educação, Índices avaliativos, Matemática, Gestão.

Abstract

This paper proposes a methodology to assist managers in making decisions to build student support mechanisms that go beyond intellectual aspects. The K-means methodology made it possible to explain the variability of grades in the classes analyzed, identified the formation of a small number of students who need educational support for the purpose of improving their academic life, and allowed us to observe the presence of small groups of grades that positively influenced the overall average of the class. Thus, by using the methodology and in light of institutional assessments, professionals involved in the teaching and learning process will be able to make informed decisions regarding improving the indices of the institution they are part of and direct the best actions to maximize the evaluative indices in favor of the well-being of students.

Keywords: Education, Evaluation indexes, Mathematics, Management.

Sumário

Resumo	vi
Abstract	vii
Sumário	viii
Introdução	1
1 Fundamentação Teórica	4
1.1 Breve histórico das avaliações educacionais	4
1.2 Rendimento acadêmico	7
2 Metodologia e Materiais do Estudo	11
2.1 Apresentação dos conjuntos de dados	11
2.2 Análise de agrupamento: K-means	12
2.3 Procedimentos para a pesquisa	14
3 Análise dos Dados e Discussão dos Resultados	15
3.1 Análise descritiva	15
Considerações Finais	27

Introdução

A busca por melhorias no povo brasileiro faz-se entender que é preciso investir em educação e com isso surgiu a necessidade de criar mecanismos para avaliar os progressos e retrocessos da educação. Na educação básica, a partir de 1990, eles vêm se traduzindo em uma aparelhagem estatal formada por provas e índices, como o Sistema de Avaliação da Educação Básica, SAEB, o Índice de Desenvolvimento da Educação Básica, IDEB, o Exame Nacional do Ensino Médio, ENEM, o Exame Nacional para Certificação de Competências de Jovens e Adultos, ENCCEJA, a Prova Brasil ou Avaliação Nacional de Rendimento Escolar, ANRESC, Programa Internacional de Avaliação de Estudantes, PISA, para avaliar a redução.

O IDEB é calculado a partir do fluxo escolar e da proficiência em Língua Portuguesa e Matemática. O fluxo é um valor entre 0 e 1 sendo obtido pelo quociente entre o número de alunos aprovados em uma turma e o número inicial de alunos da turma, enquanto a proficiência é um número entre zero e dez obtido pela média das notas dos alunos nas avaliações de Língua Portuguesa e de Matemática obtidas no SAEB. Assim, O IDEB é o resultado do produto da média obtida na proficiência pelo fluxo.

Note que algumas avaliações do sistema consideram somente duas disciplinas, embora os estudantes estejam cursando diferentes componentes curriculares.

O sistema educacional brasileiro é formado pelos ensinos da educação básica e superior. A educação básica vai desde a educação infantil ao ensino médio. O ensino superior fica a cargo das universidades, faculdades e institutos federais, estaduais e municipais de Educação.

Os Institutos Federais de Educação, Ciência e Tecnologia, um dos entes associados à educação básica através da educação profissional técnica de nível médio, estão em um contexto diferenciado com relação à oferta e formação de estudantes. A diferença é notada pela oferta de disciplinas que permitem aos alunos avançarem nos estudos, ao mesmo tempo que propõe uma formação profissional técnica de nível médio, além de ofertar cursos de graduação e pós-graduação.

Com relação às avaliações institucionais, os Institutos Federais de Educação são submetidos à avaliação do SAEB, bienalmente, e conseqüentemente, ao índice IDEB. Assim, há necessidade de que o gestor escolar tome decisões acertadas para as melhorias

do bem-estar estudantil, do fluxo escolar e do bom rendimento acadêmico.

No SAEB, são coletadas informações relevantes no que se refere à captação da descrição dos estudantes, por exemplo, aspectos econômicos, culturais e etnias. A atividade docente, ao tempo que permite um professor perceber diferentes rendimentos dos alunos em determinada disciplina, às vezes, a compreensão sobre as lacunas deixadas por essas diferenças são imperceptíveis para ele. Dessa forma, o gestor escolar, além da percepção sobre os números que refletem o bom rendimento acadêmico, deve ser sensível às questões que promovem o mau desempenho dos estudantes.

Neste contexto, uma análise individualizada dos rendimentos acadêmicos pode não refletir os déficits existentes em um grande grupo de estudantes do qual o gestor escolar é responsável. Segundo a matriz nacional comum, compete ao gestor escolar construir e adequar-se às normatizações em compasso com a educação e suas mudanças. O acompanhamento de um grande grupo de estudantes sobre os aspectos sociais, culturais, econômicos e acadêmicos envolve uma dinâmica muito complexa.

Diante da complexidade da análise de diferentes estudantes em itinerários educacionais, apresenta-se neste trabalho uma proposta de análise multivariada na qual os grupos de estudantes podem ser analisados para compreender o panorama do Instituto Federal de Educação, Campus Pedro II, Piauí.

Motivação

Diante das avaliações institucionais, os gestores escolares devem tomar decisões acertadas em relação à melhoria dos índices da instituição da qual fazem parte. Os professores, que estão envolvidos no processo de ensino-aprendizagem, por mais que acompanhem um número reduzido de estudantes, é o gestor escolar quem se torna o responsável por direcionar as melhores ações acadêmicas para maximizar os índices avaliativos da instituição.

Realizar o acompanhamento acadêmico de todos os estudantes, individualmente, é uma tarefa dificilmente executável. A proposta deste trabalho é apresentar uma metodologia que auxilie no acompanhamento educacional dos estudantes do Instituto Federal de Educação (IFPI), Campus Pedro II.

Objetivos

O objetivo geral deste estudo é apresentar uma metodologia para auxiliar o gestor acadêmico na descrição do panorama de rendimentos escolares do Instituto Federal de Educação (IFPI), Campus Pedro II. A metodologia considera uma análise estatística multivariada dos dados educacionais relativos ao primeiro semestre do ano de 2023.

Além disso, os objetivos específicos deste estudo se concentram em:

-
- (i) Realizar análise descritiva do conjunto de dados;
 - (ii) Analisar e descrever os aspectos nos quais a metodologia proposta neste trabalho poderá ser útil para auxiliar na descrição e acompanhamento de rendimentos acadêmicos;
 - (iii) Apresentar possibilidades do uso da metodologia para melhoria na tomada de decisão.

Estrutura do Trabalho

A partir do objetivo principal proposto neste estudo, de analisar dados educacionais de forma multivariada no que se refere aos rendimentos acadêmicos dos estudantes e aspectos sociais, este trabalho está dividido em quatro capítulos. A fundamentação teórica é o primeiro capítulo. Nele se apresentam aspectos legais sobre o IFPI Campus Pedro II, avaliações e rendimento escolar.

No segundo, apresenta-se a metodologia e os materiais dos estudos. No terceiro, faz-se a análise dos dados e discussão dos resultados e no quarto as considerações finais.

Capítulo 1

Fundamentação Teórica

Neste capítulo serão apresentados os aspectos teóricos que fundamentam este estudo. Serão abordados tópicos inerentes à metodologia, conteúdos relativos à sistemática de acompanhamento escolar.

1.1 Breve histórico das avaliações educacionais

O Instituto Federal do Piauí foi criado pelo artigo 5º, inciso XXIV, da lei n.º 11.892, de 29 de dezembro de 2008, pertence à Rede Federal de Educação Profissional, Científica e Tecnológica que, entre outras instituições, inclui, no inciso I da mesma lei, os Institutos Federais.

O Instituto Federal do Piauí, Campus Pedro II, atende público de jovens e adultos da Cidade de Pedro II, Piauí, oferecendo cursos técnicos de nível básico ou curso de nível superior conforme o artigo 21 da Lei de Diretrizes e Bases da Educação Nacional (LDB, 2024).

Assim, pelo caráter institucional de escola de ensino básico, o Campus Pedro II participa do conjunto de testes externos do Sistema de Avaliação da Educação Básica, SAEB, instrumentos de avaliação da Educação Básica vinculadas ao Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, INEP, que é a autarquia federal responsável pela realização de avaliações, pesquisas e levantamentos estatísticos educacionais do governo federal que produz dados anuais a respeito da educação básica e superior, respectivamente, através do Censo Escolar e do Censo da Educação Superior (INEP, 2024a).

Para atingir os objetivos educacionais, segundo o artigo 214 da Constituição Federal, o Plano Nacional de Educação, com duração de dez anos, será estabelecido por lei, e, através dele, o Sistema Nacional de Educação será articulado em regime de colaboração de maneira que estratégias, objetivos, metas e estratégias possam ser implementadas

assegurando manutenção e desenvolvimento do ensino em seus diversos níveis, etapas e modalidade por meio das ações integradas dos poderes públicos das diferentes esferas federativas.

No PNE (2014) apresentam-se as diretrizes para o sistema educacional brasileiro. A partir dessas considerações o IFPI Campus Pedro II em cumprimento às diretrizes, artigo 2º, aplica esforços que, entre outros, são a melhoria da qualidade do ensino, superação das desigualdades educacionais, com ênfase na promoção da cidadania e na erradicação de todas as formas de discriminação, formação para o trabalho, promoção humanística, científica e tecnológica, estabelecimento e aplicação de recursos públicos em educação, pode preocupar-se, cada vez mais, com o entendimento de parâmetros que melhorem o rendimento escolar, bem como a qualidade dele.

No que tange ao rendimento escolar, o PNE (2014), salienta através da meta 7 que a educação básica deve atingir melhores patamares observando o fluxo escolar e a aprendizagem de forma que a média nacional no IDEB seja de 5,2 no ensino médio e para que isso ocorra, de acordo como a estratégia 7.9, se deve orientar as políticas das redes e sistemas de ensino para diminuir a diferença entre as escolas com os menores índices e também a média nacional, e, também, segundo a estratégia 7.11, deve-se melhorar o desempenho dos alunos da educação básica nas avaliações da aprendizagem no Programa Internacional de Avaliação de estudantes, PISA, para se conseguir 473 na média em 2021, bem como, conforme a orientação 7.36, é necessário estabelecer políticas de estímulo às escolas que melhorarem o desempenho no IDEB para valorizar o mérito do corpo docente, da direção e da comunidade escolar.

O Instituto Nacional de Estudos e Pesquisas, INEP, a cada dois anos, segundo o parágrafo segundo do artigo 5º do PNE, fará a divulgação de estudos sobre a aferição da evolução do cumprimento das metas. Isso será realizado através dos indicadores da educação básica e da educação superior. Entre os indicadores da educação básica estão as taxas de rendimento (INEP, 2024a).

Essas taxas são informações produzidas anualmente pelo Inep através dos dados do Censo Escolar da Educação Básica que se incorporam ao cálculo do Índice de Desenvolvimento da Educação Básica, IDEB, sendo um indicador de qualidade educacional produzido e divulgado a cada dois anos. No IDEB se concentram as informações de desempenho dos estudantes nos testes padronizados do Sistema de Avaliação da Educação Básica, SAEB (INEP, 2024b).

As ações de caráter avaliativo, pertencentes ao SAEB, foram iniciadas em 1990, e, no decênio (1993-2003), o Brasil assumiu, juntamente com os outros nove signatários da Declaração Mundial sobre Educação para Todos, assegurar a universalização da educação básica (INEP, 2024a). Dessa forma, o SAEB vem contribuindo para a melhoria da qualidade e universalização do acesso à escola, por oferecer subsídios concretos para a formulação e o monitoramento das políticas desse nível de ensino.

Neste contexto, o Governo Federal, em 1995, lança a versão preliminar dos Parâmetros Curriculares Nacionais, PCN, que são materiais separados por disciplinas, não obrigatórios que devem subsidiar e orientar a elaboração e revisão curricular, formação continuada do professor, discussões pedagógicas nas escolas, produção de livros e outros materiais, e avaliação do sistema de educação. As ideias contidas neles norteiam o SAEB, pois elas nasceram na Declaração Mundial de Educação para Todos (MEC, 1997).

Segundo Alexandre (2015), dentre algumas avaliações e índices dos níveis da educação básica brasileira são o SAEB, o IDEB, o Exame Nacional do Ensino Médio, ENEM, o Exame Nacional de Certificação de Competências de Jovens e Adultos, ENCCEJA, a Prova Brasileira do Final do Ciclo de Alfabetização, Prova ABC, o Programa Internacional de Avaliação de Estudantes (PISA - Programme for International Student Assessment). No ano de 1997, a análise do desempenho dos alunos na avaliação foi realizada por meio dos níveis das escalas de proficiência. Isso foi conseguido pelo trabalho de especialistas das disciplinas avaliadas ao estabelecerem associações ou relações entre momentos dos ciclos escolares e os níveis de proficiência da escala.

Em 1999, os conhecimentos de ciências humanas foram avaliados. Os alunos das escolas públicas e particulares foram escolhidos por amostra e algumas séries foram contempladas. Além de História e Geografia, as provas versaram sobre Língua Portuguesa, Matemática e Ciências Naturais (Física, Química e Biologia). Já em 2001, os testes tratavam de Língua Portuguesa e Matemática e novas matrizes de referências foram adotadas e amostras de alunos das escolas públicas e particulares. Em 2003, ocorreu a consolidação do formato de 2001.

A partir de 2013, a Avaliação Nacional da Alfabetização (ANA) prevista no Pacto Nacional pela Alfabetização na Idade Certa (PNAIC) foi incorporada ao SAEB, anualmente. Avaliação censitária envolvendo alunos do último ano do ensino fundamental das escolas públicas versava sobre escrita, leitura e matemática. Na mesma prova, se incluiu testes de ciências humanas e ciências da natureza para estudantes do nono ano do ensino fundamental de escolas escolhidas por amostra, mantendo Língua Portuguesa e Matemática para estes estudantes, bem como, para os demais, conforme 2005. Em 2014, realizou-se a ANA, novamente.

Em 2015, a fim de tornar os dados coletados mais relevantes, é disponibilizada a Plataforma Devolutivas Pedagógicas, que traz itens da prova Brasil descritos e comentados por especialistas, possibilitando professores e gestores aprimorarem o aprendizado dos alunos. As escolas públicas foram escolhidas de forma censitária e as escolas particulares de forma amostral, sendo que as provas de Língua Portuguesa e Matemática foram aplicadas para 5^o e 9^o ano do ensino fundamental, enquanto as mesmas disciplinas foram aplicadas alunos do ensino médio, escolas públicas e particulares foram escolhidas de forma amostral. Em 2016, foi aplicada a 3^a prova da ANA.

Em 2017, a prova foi aplicada em caráter censitário para a 3^a série do ensino médio

e para esta mesma série, nas escolas particulares, foi aberta a possibilidade de adesão. Agora, não só as turmas do ensino fundamental poderiam apresentar resultados no SAEB, como no IDEB, de forma censitária. Em relação às disciplinas, ficaram mantidas as de 2015.

As matrizes de referência da Base Nacional Comum Curricular (2018), em Língua Portuguesa e Matemática, foram adotadas para a prova de 2019, nas turmas de 2º ano do ensino fundamental, as escolas foram escolhidas amostralmente a fim de que a transição para as novas referências fosse realizada e o IDEB não fosse tão impactado. Nas turmas de 9º ano, aplicou-se provas de Ciências da Natureza e Ciências Humanas, com referência na BNCC(2018) e mantidas as provas de Língua Portuguesa e Matemática, para escolas públicas, de forma censitária, para as privadas, por amostra, para o 5º e 9º do ensino fundamental, com as Matrizes de Referência de 2001. Para alunos dos 3º e 4º anos do ensino médio, foram mantidas a Língua Portuguesa e Matemática, com referência a 2001, escolha censitária para escolas públicas e amostral para particulares. Em 2021, repetiu-se o plano de 2019.

1.2 Rendimento acadêmico

Em relação aos resultados do SAEB 2021, o INEP afirma que um dos impactos importantes já identificados nas duas últimas edições do Censo Escolar foi o crescimento abrupto das taxas de aprovação da rede pública entre 2020 e 2021, quando comparado a 2019, antes da pandemia. Além disso, afirma que talvez esse aumento das taxas estejam relacionados aos ajustes e critérios recomendados e adotados pelas escolas e destaca que na formulação do indicador já se considera que o aumento da proficiência média nas avaliações pode não refletir uma efetiva melhora no desempenho do sistema educacional (INEP, 2024c).

O relatório do INEP em relação ao PISA 2022 (INEP, 2024d), afirma que, segundo Andreas Schleicher, diretor de Educação da Organização para a Cooperação e Desenvolvimento Econômico, OCDE, existem pontos a serem aprimorados no que diz respeito às relações entre alunos, escola e comunidade, apesar da resiliência brasileira nos impactos da pandemia. E acrescenta que o que acontece na escola influencia o bem-estar dos alunos, que é necessário despertar o protagonismo dos discentes no processo de ensino, fortalecer as relações da família com a escola, além de proporcionar a esses jovens, professores que os acompanhem nas respectivas trajetórias. Em Matemática, 73% dos alunos ficaram abaixo do nível 2, considerado o padrão mínimo; em leitura, 50% ficaram abaixo do mínimo e em Ciências, 55% ficaram abaixo do padrão mínimo.

Do ponto de vista da avaliação em busca de indicadores para a educação básica, tanto o SAEB quanto o PISA fazem referências a elementos subjetivos tais como ajustes dos critérios de aprovação e ambiente escolar, respectivamente.

A compreensão dos valores indicativos de taxas de rendimento escolar está no entendimento de que a qualidade da escola é um conceito dinâmico alcançável a partir das dimensões avaliadas pelos indicadores. Estes servem para a comunidade escolar avaliar e melhorar a qualidade da escola (Unicef et al., 2004).

O rendimento acadêmico dos estudantes deve ser analisado no contexto escolar de maneira geral. Os indicadores de adequação da formação docente, números de turnos de funcionamento das escolas, esforço escolar, número de alunos atendidos, indicadores financeiros educacionais, ajuda financeira para os estudantes, despesas, taxas de distorção de idades e número médio de horas-aulas podem fornecer informações à cerca do panorama da escola.

Muitos outros aspectos relacionados ao bem-estar dos estudantes podem estar associados ao seu bom desempenho na escola. Segundo Demir & Karabeyoglu (2015), existem muitos fatores que afetam direta ou indiretamente o desempenho do aluno. As taxas de frequência escolares foram citadas como prejudiciais ao desempenho acadêmico e sugerem que a melhoria do atendimento ao aluno seja um indicador direto como determinação do desempenho acadêmico. Nesse sentido, os resultados do estudo utilizando análise de agrupamento hierárquico mostraram que a avaliação dos alunos do ensino médio sobre o ambiente escolar explicou 83% da variação dos níveis do seu comprometimento com a escola.

As reformas curriculares pensadas em atender demandas educacionais em que se observam a prática do respeito, desenvolvimento pessoal e autonomia do estudante promovem um despertar para as potencialidades do aluno. De acordo com Bolivar (2015), quando abordou as reformas educacionais na Espanha, reformas de cima para baixo, principalmente curriculares, sem alterar como o professor é preparado para atender uma demanda diversificada ao longo do tempo, a sala de aula permanece inalterada.

Segundo Silva & Costa (2022), sobre o papel da tecnologia na escola, observou que o avanço das tecnologias são inevitáveis e os professores devem desenvolver competências digitais nos primeiros anos de sua formação e que a aquisição destas competências são essenciais na sociedade em que vivemos hoje. Dessa forma, uma reforma educacional que impacta sobre a melhoria no desempenho escolar deve ser fortalecida.

O sucesso escolar passa por diversos fatores. A tarefa da escola em fazer com que os estudantes desenvolvam habilidades teóricas e práticas, de conteúdos que promovam a melhoria humana não é uma tarefa fácil. Muitos aspectos das relações entre estudantes e ambiente escolar têm sido estudados. Alguns estudos estão voltados para a descoberta de novas metodologias que incentivam os alunos a permanecerem na escola Júnior, Santos & Maciel (2016), outros focados no bem-estar do estudante como ponto positivo para o sucesso acadêmico, segundo (Baker; Sigmon; Nugent, 2001). Outro aspecto estudado foi programas de auxílio financeiro na melhoria das taxas de repetência, por (Shirasu; Arraes, 2016).

Demir & Karabeyoglu (2015) concluíram que a principal variável que contribui de forma negativa e significativa para a evasão escolar está no comprometimento dos alunos com a escola e o controle dos pais. Diferentes autores estudaram sobre as relações entre alunos e desempenho acadêmico, e com base nesses resultados, sugeriram que o nível de frequência dos alunos é um determinante do sucesso acadêmico, bem como um indicador direto para o desempenho acadêmico, segundo (Phillips, 1997; Lehr; Sinclair; Christenson, 2004; Sheldon, 2007).

Diferentes perspectivas metodológicas sobre os rendimentos acadêmicos dos estudantes convergem para diagnósticos semelhantes que descrevem o panorama educacional. Os estudos apontam para uma seletividade e baixo rendimento sistêmico. Trabalhos que tratam da evasão escolar de crianças em idade escolar foram rastreados utilizando meta-análise e foi possível concluir que a realização educacional não se limita ao acesso à escola ou ao tempo geral de escolarização, segundo (Júnior; Santos; Maciel, 2016).

A evasão escolar e a falta de frequência estão associadas a uma série de resultados negativos para o bem-estar da criança, como baixo desempenho acadêmico, baixo vínculo escolar, marginalização social, uso de drogas, sexualidade e abandono escolar, segundo (Baker; Sigmon; Nugent, 2001; Garry, 1996; Loeber; Farrington, 2000; Sheldon, 2007; Sheldon; Epstein, 2004; Teasley, 2004).

De acordo com Bolivar (2015), a reforma educacional proposta na Espanha não obteve êxito pelo fato de estarem focadas no processo ao invés de estarem ligadas às exigências para uma melhor qualidade de ensino. As reformas não observaram o profissionalismo dos professores e suas competências, paixão que ‘move’ o agir pessoal e que não pode ser dissociada dos aspectos profissionais da arte de ensinar.

De forma similar ao pensamento exposto no trabalho de Bolivar (2015) pode-se afirmar que a falta de conhecimento sobre as motivações dos estudantes profissionais pode ser uma das causas de baixa aprendizagem e conseqüentemente desempenho acadêmico inferior ao esperado.

A tecnologia pode auxiliar na previsão e detecção de fatores que possam contribuir para um rendimento acadêmico inferior ao esperado. Alguns exemplos de diferentes metodologias que usam a tecnologia na educação podem ser citados, por exemplo (Júnior; Santos; Maciel, 2016; Souza, 2021; Soares; Silva, 2024; Lima; Barbosa, 2017).

Segundo Demir & Karabeyoglu (2015), existem muitos fatores que afetam o desempenho dos alunos direta e indiretamente no nível do ensino secundário. As baixas frequências escolares são citadas como prejudiciais ao desempenho acadêmico. As relações individuais, familiares e escolares são apontadas como fatores atuantes que podem determinar a qualidade da aprendizagem. O ambiente escolar e as relações professor-estudante e gestão foram observadas e concluiu-se que o bem-estar acadêmico é um dos fatores que podem levar o estudante a permanecer na escola, fortalecendo o comprometimento com

sua educação.

Capítulo 2

Metodologia e Materiais do Estudo

2.1 Apresentação dos conjuntos de dados

O IFPI-Campus Pedro II é uma das vinte unidades dos Institutos Federais em território piauiense. Oferece três cursos técnicos integrados ao ensino médio: Administração, Meio Ambiente e Informática.

A cidade Pedro II é conhecida como a “Suíça piauiense” por apresentar temperaturas mais amenas durante todo o ano. Se destaca em diversos atrativos turísticos, entre eles o Festival de Inverno de Pedro II, a extração da pedra opala e na produção de joias. Os setores econômicos que mais reuniram trabalhadores em Pedro II no ano de 2022 foram Administração Pública, Defesa e Seguridade Social, Comércio Varejista e Atividades de Atenção à Saúde Humana (SEBRAE, 2004).

O conjunto de dados utilizado neste trabalho é referente às notas avaliativas e informações sociais de estudantes do IFPI-Campus Pedro II. Os dados são secundários, obtidos através do próprio sistema educacional da referida instituição. Para manter o sigilo e a ética com relação às informações dos estudantes, realizou-se a codificação das informações sensíveis, mantendo somente os nomes dos cursos e disciplinas ofertadas.

As informações contidas no conjunto de dados apresentam a descrição do curso, se os estudantes moram no município de Pedro II, forma de ingresso, tipo de escola de origem, notas e frequências, e também, referência sobre o acesso a benefícios financeiros.

2.2 Análise de agrupamento: K-means

A análise de agrupamento permite a descoberta de relações existentes entre os exemplares, objetos ou sujeitos sob análise. De maneira geral, as técnicas de agrupamentos têm como base o uso de distâncias. Os exemplares que possuem menor “distância” são mais similares e aqueles que possuem as maiores “distâncias” são ditos dissimilares.

Neste estudo utilizou-se a análise de agrupamento *K-means*. É um agrupamento por partição, em que são criados grupos e, interativamente, os exemplares são realocados entre as partições de modo que o modelo de grupo mude e se ajuste melhor ao objetivo da maximização da dissimilaridade entre os grupos. Dessa forma, ao final da execução do algoritmo de agrupamento, os grupos são obtidos de maneira que exemplares similares são agrupados de forma que a similaridade intergrupos tenha sido minimizada e a relação intragrupos tenha sido maximizada.

O agrupamento *K-means* MacQueen (1967) é o algoritmo de aprendizado de máquina não supervisionado mais comumente usado para particionar um determinado conjunto de dados em um conjunto de k grupos. O número de grupos k é pré-especificado pelo analista utilizando um dos vários métodos de escolha para o valor de k : Método Elbow (Syakur et al., 2018), Método Silhouette (Rousseeuw, 1987) ou Método da Estatística Gap (Tibshirani; Walther; Hastie, 2001). Os métodos estão disponíveis no soft-ware utilizado neste trabalho e usou-se o método Elbow para identificação do número razoável de grupos.

O Método Elbow é bastante usual e consiste na interpretação de um gráfico de linhas com formato de “cotovelo” em que a curva apresenta a menor soma de quadrados totais no grupo, indicando menor variação interna. O eixo horizontal do gráfico corresponde ao número de grupos e o eixo vertical é a soma dos quadrados nos grupos para cada número de grupos (Syakur et al., 2018).

O algoritmo padrão *K-means* é o algoritmo de Hartigan & Wong (1979) que define a variação total no grupo como a soma das distâncias quadradas das distâncias euclidianas entre os itens e o centroide correspondente:

$$W(G_k) = \sum_{x \in G_k} (x_i - \mu_k)^2, \quad (2.1)$$

em que x_i , é o i -ésimo exemplar pertencente ao k -ésimo grupo, μ_k é o k -ésimo ponto médio do k -ésimo grupo seja minimizada. A variação total dentro do grupo como ocorre no *software* R - WSS é definida por

$$WSS = \sum_{k=1}^K W(G_k) = \sum_{k=1}^K \sum_{x_i \in G_k} (x_i - \mu_k)^2. \quad (2.2)$$

A soma de quadrados totais no grupo (WSS) mede a qualidade do agrupamento e

se deseja que seja a menor possível, de tal forma que haja menor variação no grupo.

Após a identificação do número razoável de grupos, o algoritmo distribui os objetos de forma aleatória nos, k , grupos. Com os centros iniciais para cada grupo, os objetos restantes são atribuídos ao centroide mais próximo, no qual o mais próximo é definido usando a distância euclidiana entre o objeto e a média do grupo, definida por:

$$d_{\text{euc}}(x, \mu_k) = \sqrt{\sum_{i=1}^n (x_i - \mu_k)^2} \quad (2.3)$$

No geral, as técnicas de agrupamentos se baseiam em similaridades ou diferenças entre os exemplares, a partir das características em análise. Os exemplares que possuem menor “distância” são mais similares e os que apresentam maiores “distâncias” são ditos dissimilares. Após a realocação dos objetos, os novos centros são obtidos e a cada observação é verificada se poderá estar mais próxima de um grupo diferente.

O algoritmo *K-means* pode ser resumido da seguinte forma:

1. Com as informações do conjunto de dados, especifica-se do número de grupos, k ;
2. Selecione aleatoriamente k objetos do conjunto de dados, como os centros dos grupos;
3. Atribui cada observação ao centroide mais próximo, com base na distância euclidiana, conforme equação (2.3) ;
4. Para cada um dos k grupos, atualize o centroide do grupo calculando os novos valores médios de todos os pontos de dados no grupo. O centroide de um k -ésimo agrupamento é um vetor de comprimento contendo as médias de todas as variáveis para as observações no k -ésimo agrupamento; p é o número de variáveis;
5. Minimizar, iterativamente, a soma de quadrados totais no grupo. Repetir as etapas 3 e 4 até que as atribuições do grupo parem de mudar ou o número máximo de iterações definido pelo usuário seja atingido.

A solução do agrupamento é uma apresentação visual da disposição dos objetos relativos às variáveis em um plano cartesiano. Com a apresentação gráfica, pode-se estudar: as relações de proximidade entre as variáveis; variabilidade das variáveis e relações existentes entre objetos de variáveis em estudo.

2.3 Procedimentos para a pesquisa

O conjunto de dados é do tipo secundário, não foi realizado nenhum filtro de ordenação alfabética para aplicação da metodologia. O conjunto de dados consta dos rendimentos escolares dos estudantes. Realizou-se a codificação das variáveis e consideraram-se somente os dados não faltantes.

Como exigência da metodologia, realizou-se a padronização dos dados e a Tabela 2.1 apresenta as variáveis envolvidas nesse trabalho.

Tabela 2.1: Apresentação geral das variáveis estudadas

Turma 01, 02 e 03
Disciplinas para os cursos A, B e C
Biologia, Matemática, História, Educação Física, Filosofia, Geografia, Física, Química, Sociologia
Escola de Origem:
Ensino público ou privado
Beneficário:
Se o estudante recebe algum auxílio financeiro institucional

Assim, de acordo com a Tabela 2.1, as codificações **A14**, **B14** e **C14**, por exemplo, referem-se ao décimo quarto estudante do cursos de A, B e C, respectivamente, do conjunto de dados das turmas 01, 02 ou 03.

Utilizou-se o *software R* (R Core Team, 2024) para as análises estatísticas e os pacotes *MASS* e *ggplot2* para as construções e apresentação gráfica.

Para a manipulação dos dados, busca ativa no conjunto de dados foi utilizado o pacote *tydiverse*. Para a metodologia de agrupamento, *k-means*, no *software R* existem pacotes que auxiliaram na aplicação e análise dos dados. Os pacotes *cluster*, *factoextra* foram utilizados.

Capítulo 3

Análise dos Dados e Discussão dos Resultados

3.1 Análise descritiva

O Instituto Federal do Piauí (IFPI)-Campus Pedro II oferece, integradamente ao Ensino Médio, os cursos técnicos em Administração, Informática e Meio Ambiente. Aquele que aspira a uma vaga de Ensino Médio no IFPI-Campus Pedro II deve submeter-se a uma das modalidades de ingresso, por exemplo, transferências ou exames classificatórios.

O conjunto de dados usado neste trabalho é referente às notas dos estudantes do IFPI-Campus Pedro II. As informações foram concedidas a pedido de quem realizou a pesquisa e trata-se de dados secundários.

A análise das notas médias referentes ao primeiro semestre de 2023 foi realizada para as turmas 01 e 02 de 117 e 119 estudantes dos três cursos citados acima. Além disso, dada a especificidade de registros de notas do IFPI - Campus Pedro II, consideraram-se os registros das avaliações do segundo bimestre de 89 estudantes da turma 03 dos cursos A, B e C.

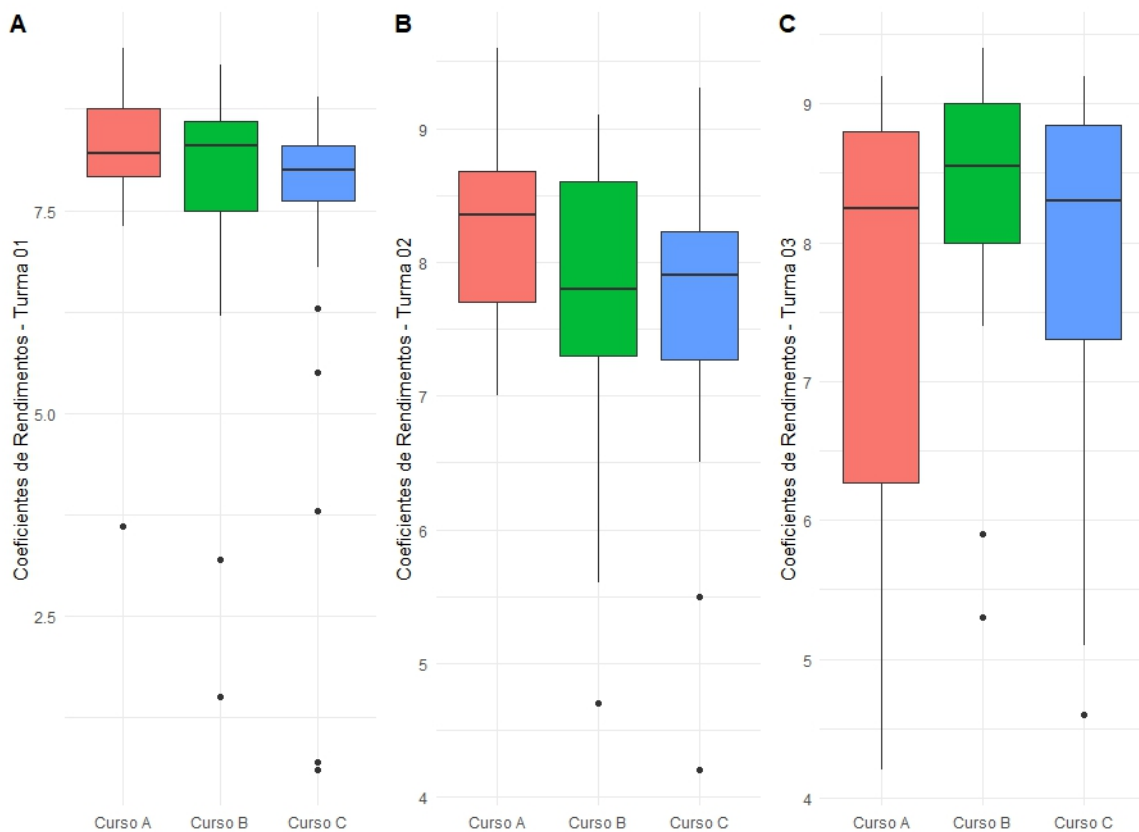
A representação gráfica *BoxPlot* mostrou a dispersão e a distribuição das notas de acordo com algumas variáveis. O *BoxPlot* dá uma ideia da posição, da dispersão, da assimetria, das caudas e dados discrepantes. A posição central é dada pela mediana e a dispersão pela diferença entre o terceiro e o primeiro quartil. As posições relativas ao primeiro, ao segundo e ao terceiro quartil dão uma noção da assimetria da distribuição. Os comprimentos das caudas são dados pelas linhas que se estendem do retângulo aos valores remotos e aos valores atípicos (Bussab; Morettin, 2010).

Na Figura 3.1, gráfico *BoxPlot*, nota-se que a mediana representativa dos coeficientes de rendimentos dos estudantes é maior que 7,0, que é o mínimo esperado para aprovação. De maneira geral, verifica-se que os coeficientes de rendimentos para a turma 01, em todos

os cursos, são menores em relação aos demais anos. Porém, a variabilidade é menor.

Para a turma 03, gráfico *BoxPlot*, (Figura 3.1 - C), o curso codificado com a letra A possui maior variabilidade dos coeficientes de rendimentos. Vale lembrar que a turma 03 possui uma sistemática de avaliação diferenciada e estamos diante de coeficientes de rendimentos bimestrais desses estudantes.

Figura 3.1: Boxplot dos coeficientes de rendimentos segundo os cursos técnicos A, B e C

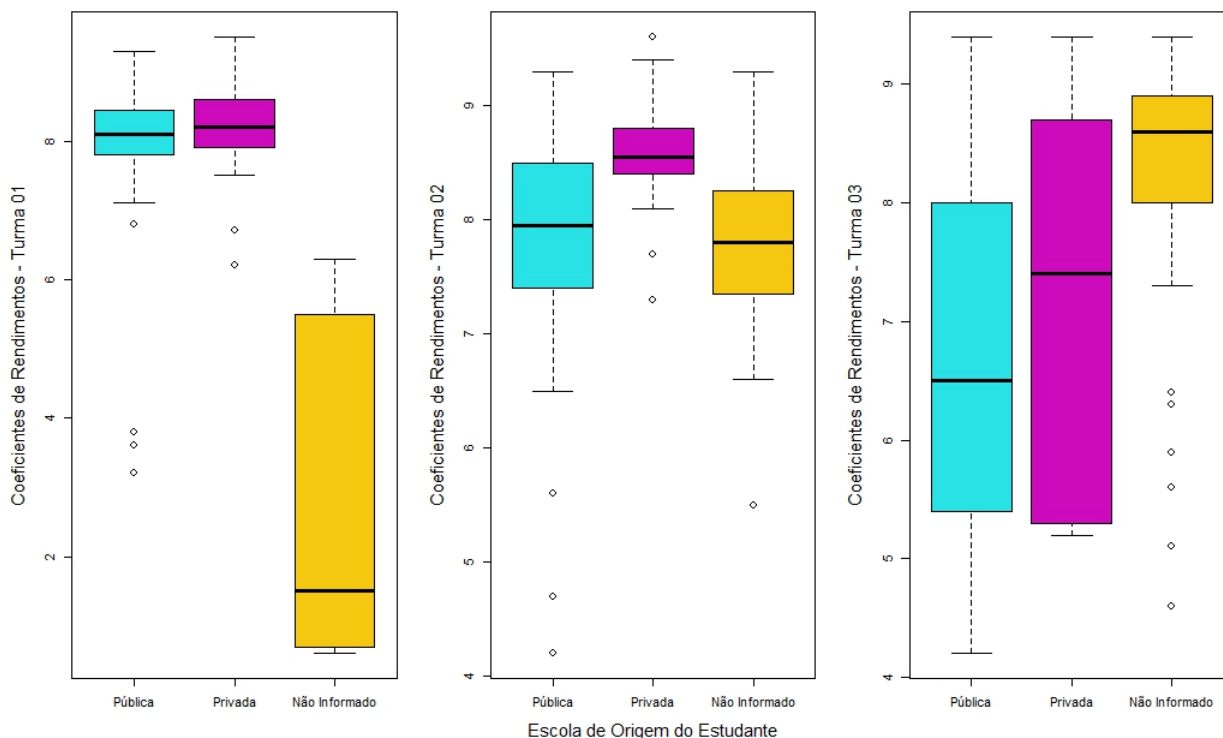


Ao se considerar as turmas 01, 02 e 03 desta sequência e seus respectivos cursos, é possível notar que a variabilidade do rendimento acadêmico aumenta, exceto para o curso codificado com a letra B, cuja variabilidade aumenta da turma 01 para a turma 02 e diminui da turma 02 para a 03.

Os alunos do curso codificado com a letra A possuem rendimentos acadêmicos com assimetria positiva na turma 01, ou seja, os seus rendimentos apresentam maior variabilidade acima da mediana. Enquanto nas turmas 02 e 03 a assimetria é negativa.

De acordo com o *Boxplot* da Figura 3.2 os alunos que concluíram o ensino fundamental na escola pública, bem como os que terminaram na escola privada têm a mediana dos rendimentos acadêmicos aproximadamente 8,0. Nota-se, ainda, menor variabilidade em relação às demais turmas e alguns coeficientes de rendimento inferiores à nota média exigida para aprovação.

Figura 3.2: Boxlot dos coeficientes de rendimentos segundo a escola de origem



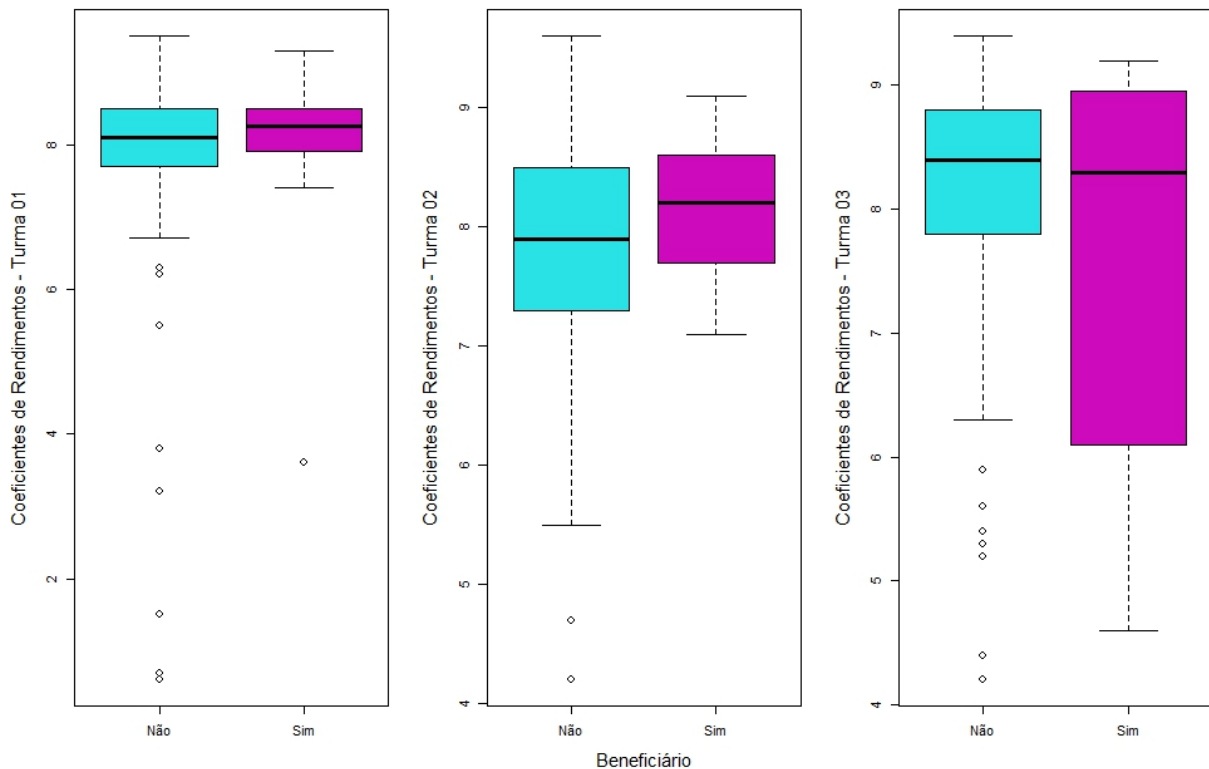
Ainda de acordo com o *BoxPlot* da Figura 3.2, é possível perceber a diferença mais acentuada na na turma 03. Há uma maior variação, sobretudo quando se analisa os coeficientes de rendimentos da rede privada de ensino. Ao contrário do que ocorre na turma 03, a turma 02 apresenta uma menor variabilidade e maiores notas associadas aos estudantes de escola privada.

Segundo a proposta de análise sobre os coeficientes de rendimentos, estudou-se a possível relação existente entre os estudantes beneficiários e o desempenho acadêmico medido através das notas nas disciplinas.

O beneficiário é o estudante que, por atender a critérios de vulnerabilidade social, recebe ajuda financeira na forma de bolsa estudantil fornecida pelo IFPI ou participa de programas de auxílio financeiro.

A Figura 3.3 apresenta um *Boxplot* dos coeficientes de rendimentos segundo o *status* de auxílio financeiro recebido pelos estudantes.

Figura 3.3: Boxplot dos coeficientes de rendimentos segundo *status* de beneficiário



No *BoxPlot* da Figura 3.3, nota-se que, em geral, os estudantes que participam de programas de auxílio financeiro apresentam notas medianas acima da média de aprovação do IFPI. Nas turmas 01 e 02, os coeficientes de rendimentos dos estudantes beneficiários são maiores que os coeficientes dos estudantes que não recebem auxílio financeiro. A maior variabilidade entre os beneficiários está na turma 03.

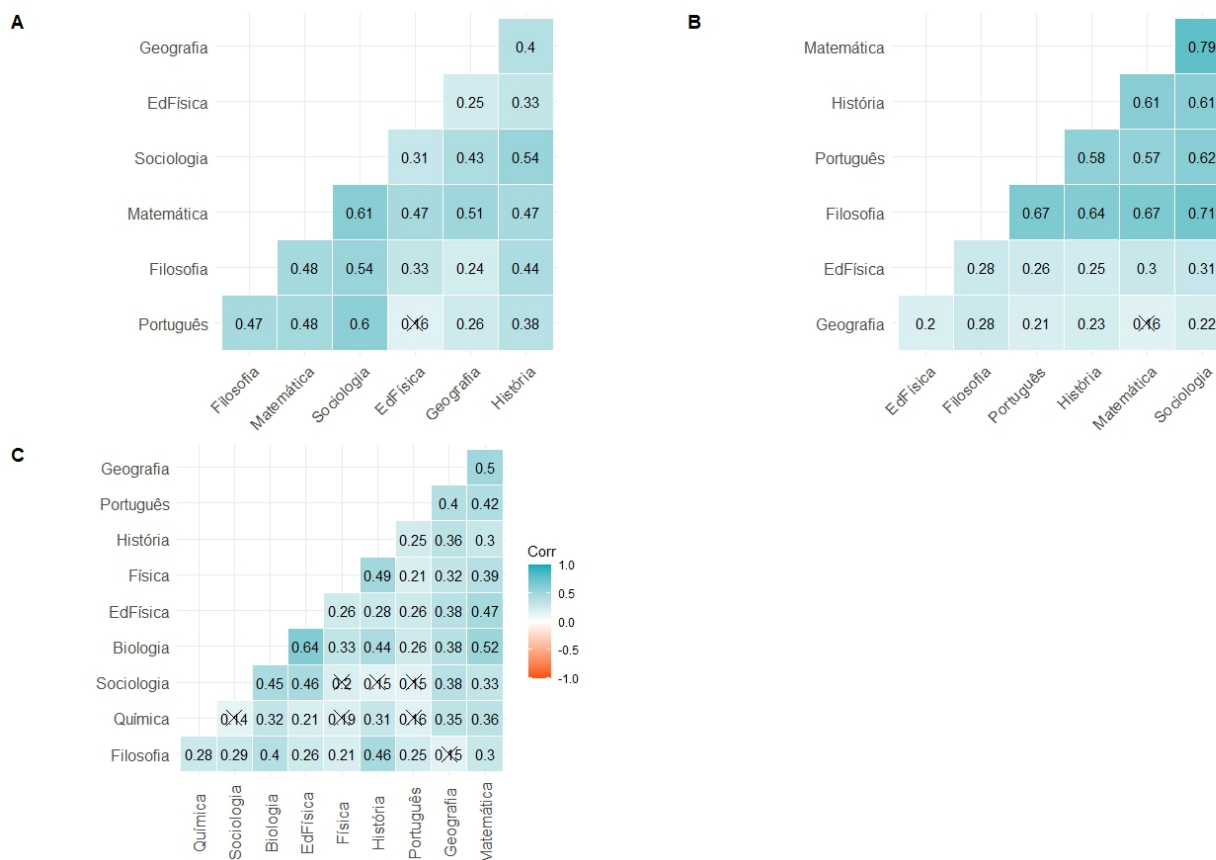
A Figura 3.4 apresenta as associações que possam haver entre as disciplinas cursadas pelos estudantes durante o primeiro bimestre de 2023. Para os cálculos das correlações das disciplinas duas a duas utilizou-se a equação 3.1:

$$corr(X, Y) = \frac{1}{n} \sum_{i=1}^n \frac{(x_i - \bar{x})}{dp(X)} \frac{(y_i - \bar{y})}{dp(Y)} \quad (3.1)$$

que representa a média dos produtos dos valores padronizados das notas das disciplinas X e Y e dp é o desvio padrão.

A Figura 3.4 apresenta as associações que possam haver entre as disciplinas cursadas pelos estudantes durante o primeiro bimestre de 2023. Nota-se não haver correlações negativas e a maior associação está entre as disciplinas de matemática e sociologia para a turma 02 (Figura 3.4 - B) e alguns coeficientes de correlações abaixo de 0.20, indicando uma fraca associação entre as disciplinas.

Figura 3.4: Análise dos coeficientes de correlações entre as disciplinas: (A) - Turma 01, (B) - Turma 02 e (C) - Turma 03.



Na Figura 3.4, pode-se notar que os coeficientes de correlações nas turmas 01 e 02 são maiores para as associações das disciplinas duas a duas, diferentemente do que ocorre para a turma 03 (3.4 - C), na qual os coeficientes são menores significando uma associação menor entre as disciplinas, segundo o coeficiente de rendimentos dos estudantes.

Dessa forma, percebe-se que existe associação entre duas disciplinas, mas os valores indicam que as correlações não são perfeitamente lineares. Elas são de outra natureza. Segundo os coeficientes de correlações, as disciplinas das turmas 01 e 02 (Figuras 3.4 - A e B) Português e Educação Física apresentam a menor correlação. Entre as disciplinas de Matemática e Sociologia, as associações são maiores.

Ao se utilizar os rendimentos dos estudantes para a análise das associações entre as disciplinas, percebeu-se haver uma desconexão dos itinerários das disciplinas em termos de coeficientes de correlação.

Naturalmente, espera-se que as disciplinas estejam de alguma forma interligadas e que essa ligação pode ser estudada sob diferentes aspectos, e isto se comprovou, visto que o coeficiente de correlação apontou associações fortes (valores próximos de 1) para algumas disciplinas em turmas distintas. Por exemplo, o caso das disciplinas de Matemática e Sociologia, ao se observar as turmas 01 e 02.

Por outro lado, as desconexões também são interligações que podem ser estudadas sob diferentes aspectos, entre eles, os baixos coeficientes de rendimentos, os quais podem ter causas variadas, não necessariamente ligados aos itinerários das disciplinas, como mostram os trabalhos de (Phillips, 1997; Lehr; Sinclair; Christenson, 2004; Sheldon, 2007) que estudaram a relação do estudantes com a escola e constataram que a frequência escolar e o ambiente favorável à presença do estudante na escola impactam o desempenho acadêmico.

O aspecto numérico existente no coeficiente de rendimento acadêmico não expressa o conjunto de variáveis que contribuem para o baixo rendimento escolar e não favorece ao professor e ao gestor na detecção de estudantes que necessitam de tratamento diferenciado em favor de sua melhoria acadêmica.

Dessa forma, as próximas análises tratam do agrupamento dos estudantes segundo as notas das avaliações nas disciplinas cursadas.

A formação de grupos, segundo um conjunto de variáveis, fornece um panorama dos rendimentos escolares. O agrupamento pode favorecer professores e gestores na implementação de dinâmicas de apoio aos estudantes que possuem necessidades de atendimento diferenciados.

Ao adotar o Método Elbow escolheu-se o número de grupos segundo a variabilidade para a estruturação do agrupamento. Porém, como a metodologia é uma proposta para o acompanhamento dos rendimentos escolares, o gestor poderá optar por um número maior de grupos. No entanto, a escolha por maiores números de grupos pode não favorecer na detecção de pequenos números de estudantes que necessitam de apoio acadêmico.

O vetor de médias de um estudante são variáveis que interagem na formação e definição do rendimento escolar. Essa interação entre as variáveis permite usar o k -means, que funciona de maneira que, se os alunos pertencem a um mesmo grupo, pode-se dizer que o algoritmo os considerou mais similares e, se os manteve separados, os classificou em menos similares.

O agrupamento, alocou estudantes em um mesmo grupo de forma que internamente apresentem as menores variações para média, e, em grupos distintos, de maneira que as variações das distâncias entre os centros dos grupos sejam as maiores.

As Figuras 3.5, 3.6, 3.7 apresentam os agrupamentos construídos segundo a metodologia descrita em 2.2.

Figura 3.5: Apresentação do agrupamento para os dados referentes a turma 01.

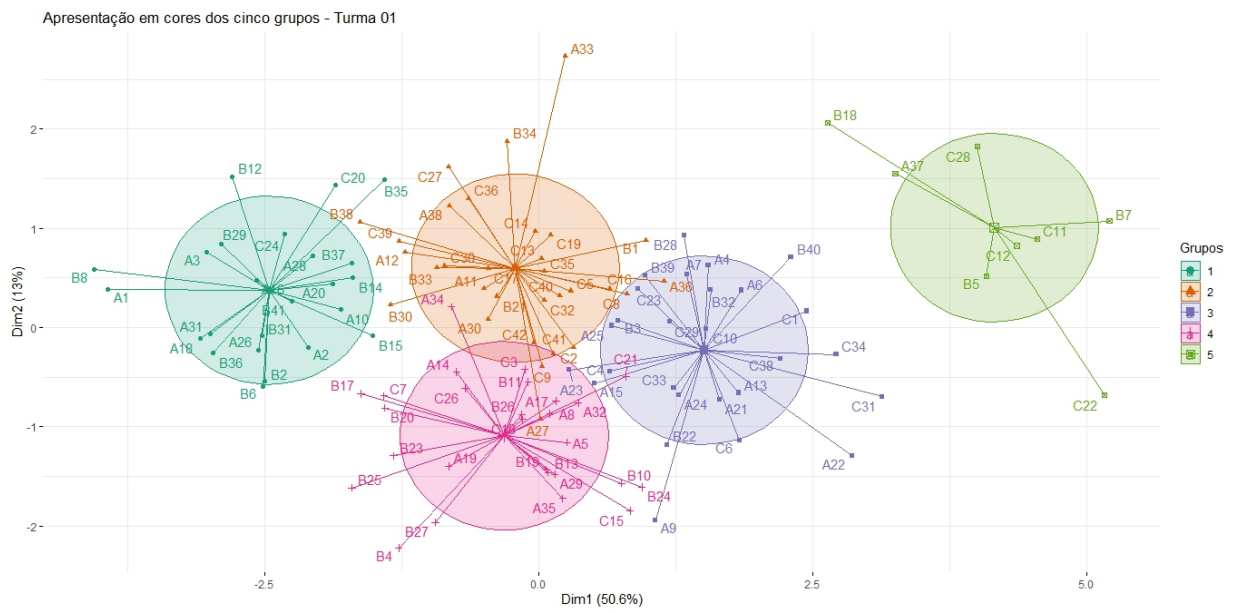


Figura 3.6: Apresentação do agrupamento para os dados referentes a turma 02.

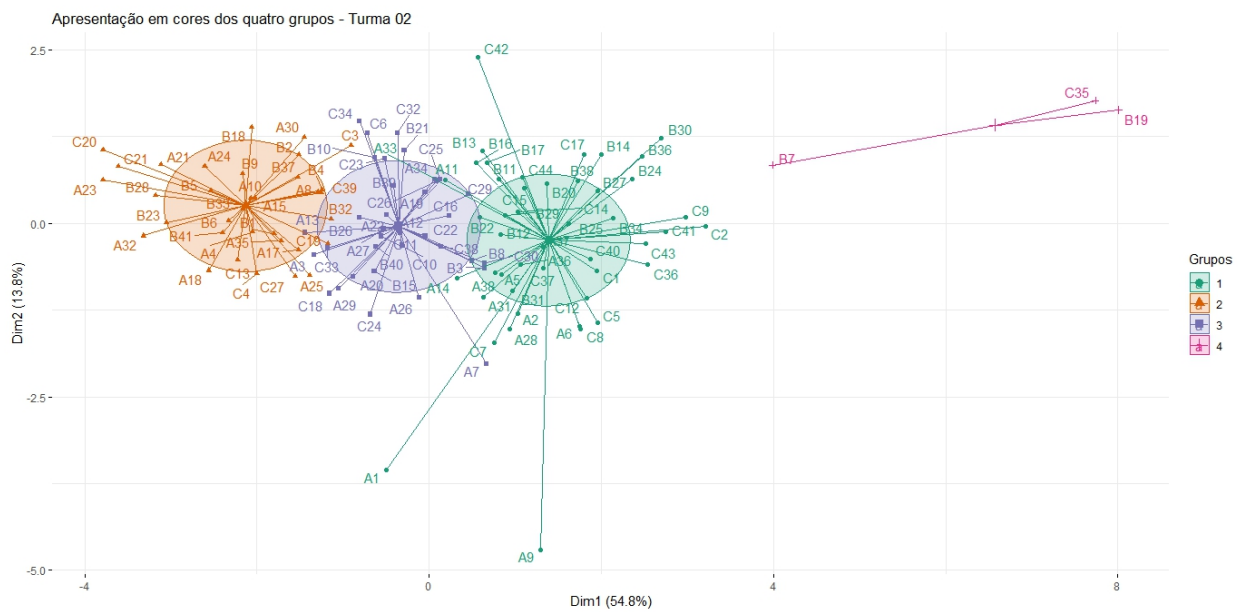
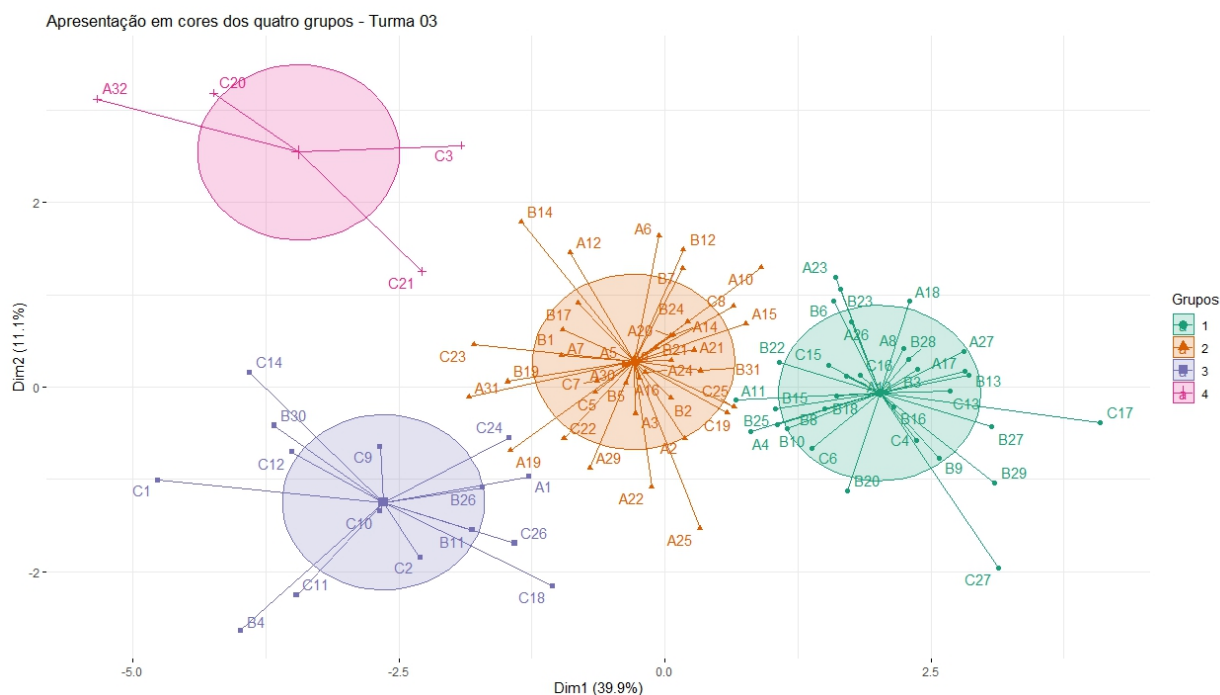


Figura 3.7: Apresentação do agrupamento para os dados referentes a turma 03.



Nas Figuras 3.5, 3.6, 3.7, observa-se a distribuição dos grupos de estudantes segundo os rendimentos acadêmicos nas disciplinas analisadas. Cada estudante possui um vetor de notas. O centro da elipse representa um vetor de médias para cada grupo de estudantes formado. Os segmentos de reta que partem do centro da elipse até a localização dos estudantes representam a distância entre os vetores de notas dos indivíduos e a média de notas do grupo e mostram o quanto as médias desses alunos variam em relação à média geral do grupo. Quanto maior a distância, maior é a variabilidade das notas.

Para a Figura 3.5 o agrupamento explica 63,3% da variabilidade do conjunto de dados. Possui um grupo de oito estudantes que se apresenta separadamente dos demais grupos, significando maior dissimilaridade com relação aos demais. De forma similar, a Figura 3.6 apresenta um agrupamento de três estudantes bastante dissimilares com relação aos demais grupos. Nota-se maior similaridade na turma de segundo ano, visto que as elipses formadas encontram-se agrupadas. A configuração gráfica da disposição das variáveis e indivíduos explica 68% da variabilidade das notas para o segundo ano. Para as notas do terceiro ano, na Figura 3.7 pode-se observar que a dissimilaridade entre os primeiro e o terceiro grupos não foi o suficiente para que a intersecção das respectivas regiões fosse vazia. Apresenta a formação de dois grandes grupos dissimilares com relação aos grupos 1 e 3. O percentual de explicação é de 51%.

A apresentação gráfica auxilia na detecção de indivíduos que possuem vetores de notas semelhantes e possibilita a identificação de padrões. Como proposta da metodologia, a apresentação dos valores de médias para cada grupo formado poderá auxiliar o gestor na tomada de decisão.

Tabela 3.1: Vetores de médias totais para a turma 01 e os agrupamentos

Vetores de médias para as disciplinas							
Educação física	Filosofia	Geografia	História	Português	Matemática	Sociologia	Número de estudantes
Médias totais para a turma 01							
7,965	8,091	8,053	7,978	8,256	7,365	7,856	117
Grupo 1							
8,250	8,996	8,629	8,883	8,971	8,662	8,821	24
Grupo 2							
7,639	7,932	8,184	7,874	8,771	7,510	8,148	31
Grupo 3							
7,552	7,730	8,167	7,611	7,541	6,115	7,348	27
Grupo 4							
8,793	8,056	8,093	8,048	7,967	7,852	7,689	27
Grupo 5							
6,975	7,338	5,300	6,662	7,500	5,487	6,112	8

Nota-se através da 3.1 que o vetor de médias para o primeiro ano apresenta valores maiores que a média exigida para aprovação. No entanto, a metodologia aplicada nesse estudo possibilitou a identificação de um grupo formado por oito estudantes em que somente as notas médias nas disciplinas de Português e Filosofia são aprovativas, possibilitou a redução do número de estudantes para que os profissionais envolvidos no processo de ensino e aprendizagem possam interferir no sentido de propor acompanhamento personalizado para melhoria acadêmica desses estudantes.

De acordo com a Tabela 3.1 o gestor poderá decidir pelo acompanhamento estudantil e investigar as causas dos rendimentos acadêmicos inferiores a média exigida. Por exemplo, para o grupo 2, o estudante B18 reside em Pedro II, proveniente de escola pública e não é beneficiário de programas de auxílio estudantil.

A Tabela 3.2, referente as notas dos estudantes da turma 02, mostra que a disciplina de Português apresenta média geral inferior a 7,0. No geral os melhores índices acadêmicos estão nas disciplinas de Geografia e Educação Física. Verifica-se, também, que o grupo 2, formado por 46 estudantes apresenta os melhores desempenhos acadêmicos em termos de notas avaliativas nas disciplinas. E que, três estudantes necessitam de acompanhamento acadêmico para a melhoria dos seus desempenhos.

Tabela 3.2: Vetores de médias totais para a turma 02 e os agrupamentos

Vetores de médias para as disciplinas							
Educação Física	Filosofia	Geografia	História	Português	Matemática	Sociologia	Número de estudantes
Médias totais para a turma 02							
7,778	7,566	8,609	8,010	6,771	8,176	7,692	119
Grupo 1							
7,311	7,530	8,232	7,738	6,981	8,254	7,716	37
Grupo 2							
8,215	8,596	9,089	8,900	7,611	9,263	8,454	46
Grupo 3							
7,745	6,509	8,379	7,258	5,573	7,127	7,052	33
Grupo 4							
7,200	3,833	8,433	6,000	4,500	2,100	2,733	3

Para a turma de estudantes da turma 03, de acordo com a Tabela 3.3, verifica-se que no geral as notas médias são próximas da nota exigida para aprovação. Nota-se que a disciplina de Matemática apresenta média geral igual a 5,746 e que para o grupo 1 composto de 33 estudantes a média é de 7,616. Percebe-se que os estudantes encontram alguma dificuldade na disciplina de matemática, além de grupos de estudantes apresentarem dificuldades em diferentes disciplinas.

Tabela 3.3: Vetores de médias totais para a turma 03 e os agrupamentos

Vetores de médias para as disciplinas										
Biologia	Educação Física	Filosofia	Física	Geografia	História	Português	Matemática	Química	Sociologia	Número de estudantes
Médias totais para a turma 03										
7,108	7,739	7,733	7,278	7,375	7,697	7,867	5,746	7,611	8,480	89
Grupo 1										
8,419	8,506	8,297	7,737	8,731	8,534	8,866	7,616	8,231	9,234	32
Grupo 2										
7,105	7,911	7,589	7,245	7,100	7,816	7,376	4,900	7,303	9,005	38
Grupo 3										
4,907	5,867	7,233	7,000	5,440	7,127	7,113	4,233	7,233	5,533	15
Grupo 4										
4,900	7,000	6,450	4,950	6,400	2,000	7,375	4,500	7,000	8,500	4

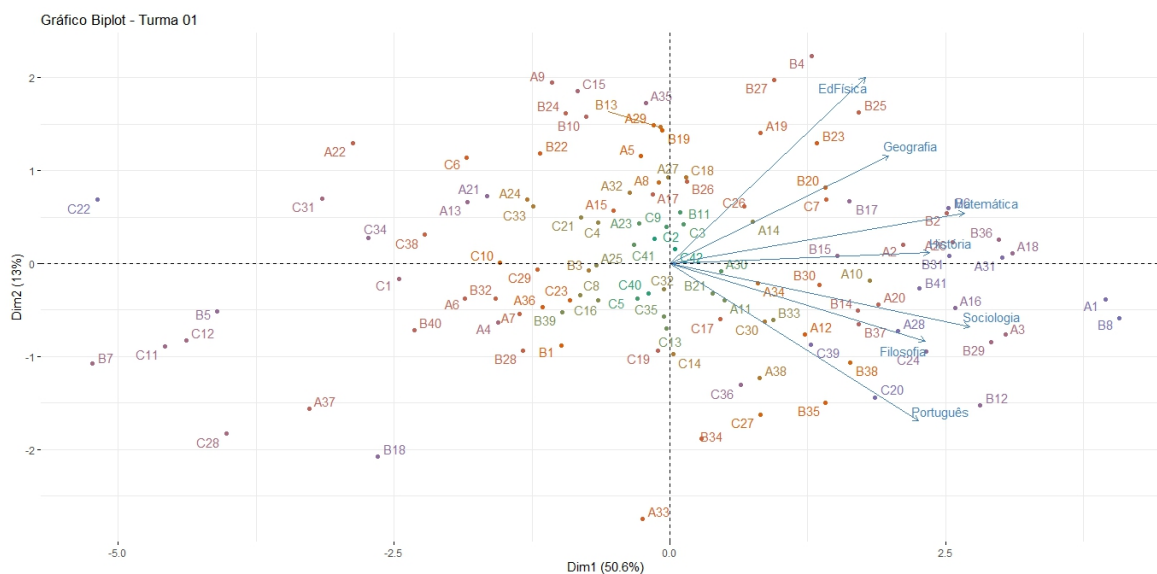
Os estudantes da turma 03 estão na fase de início da juventude e com um pouco mais de autonomia nos afazeres escolares. E muitos outros aspectos relacionados ao bem-estar e frequências escolares podem ser a causa do baixo rendimento acadêmico. Segundo Demir & Karabeyoglu (2015), existem outros fatores que podem afetar de forma direta ou indiretamente o desempenho do estudante: o ambiente escolar e comprometimento com a escola, por exemplo.

Dessa forma, investigar a causa do baixo rendimento acadêmico poderá auxiliar na identificação de metodologias e direcionamentos acadêmicos mais assertivos como propósito de melhoria de desempenho dos estudantes. De acordo com Bolivar (2015), as reformas educacionais sem que haja uma mudança na forma como o professor é preparado

para aquisição de outras competências, a sala de aula permanecerá inalterada. Para Phillips (1997), Lehr, Sinclair & Christenson (2004), Sheldon (2007) o acompanhamento da frequência dos estudantes é um indicador da relação da escola com o estudante. Mas, a realização educacional não se limita ao acesso à escola ou o tempo geral de escolarização (Júnior; Santos; Maciel, 2016).

Com o propósito de investigar a relação entre os rendimentos dos estudantes e as disciplinas cursadas, realizou-se a apresentação gráfica para expor as possíveis relações existentes.

Figura 3.8: Gráfico Biplot para a turma 01 - relações entre disciplinas e estudantes



A Figura 3.8, o comprimento dos vetores de disciplinas, significa a variabilidade. Quanto maior o comprimento, maior a variabilidade de notas para a disciplina. Os ângulos formados pelos vetores correspondem às associações entre as disciplinas. Ângulos menores que 90° significam que as disciplinas possuem alguma correspondência. Os pontos no gráfico representam os estudantes. Estudantes próximos ao vetor de disciplinas significam que a disciplina foi um fator importante para a disposição dos estudantes naquela região gráfica e o rendimento dos estudantes é influenciado pela disciplina.

Com base na Figura 3.8 verifica-se de maneira geral que as disciplinas, segundo as notas, estão associadas. A disciplina de Educação Física e Português apresentam um ângulo de aproximadamente 90° , significando uma associação muito fraca entre as disciplinas. A variabilidade das notas nas disciplinas são aproximadamente iguais. E que os estudantes se apresentam dispersos no gráfico.

Segundo as Figuras 3.9 e 3.10 os vetores de disciplinas são correlacionados positivamente, ou seja, apresentam correlação forte. Para turma 02, Figura 3.9, as disciplinas de Geografia e Matemática são as menos correlacionadas. Nota-se que os pontos representativos dos alunos estão mais concentrados próximo ao centro do plano determinado pelas dimensões 1 e 2, e, com isso, pode-se concluir que as notas nas disciplinas são homogêneas. Quando se considera a Figura 3.10, relativa a turma 03, verifica-se maior dispersão dos pontos e menor variabilidade das notas para a disciplina de Português.

Figura 3.9: Gráfico Biplot para a turma 02 - relações entre disciplinas e estudantes

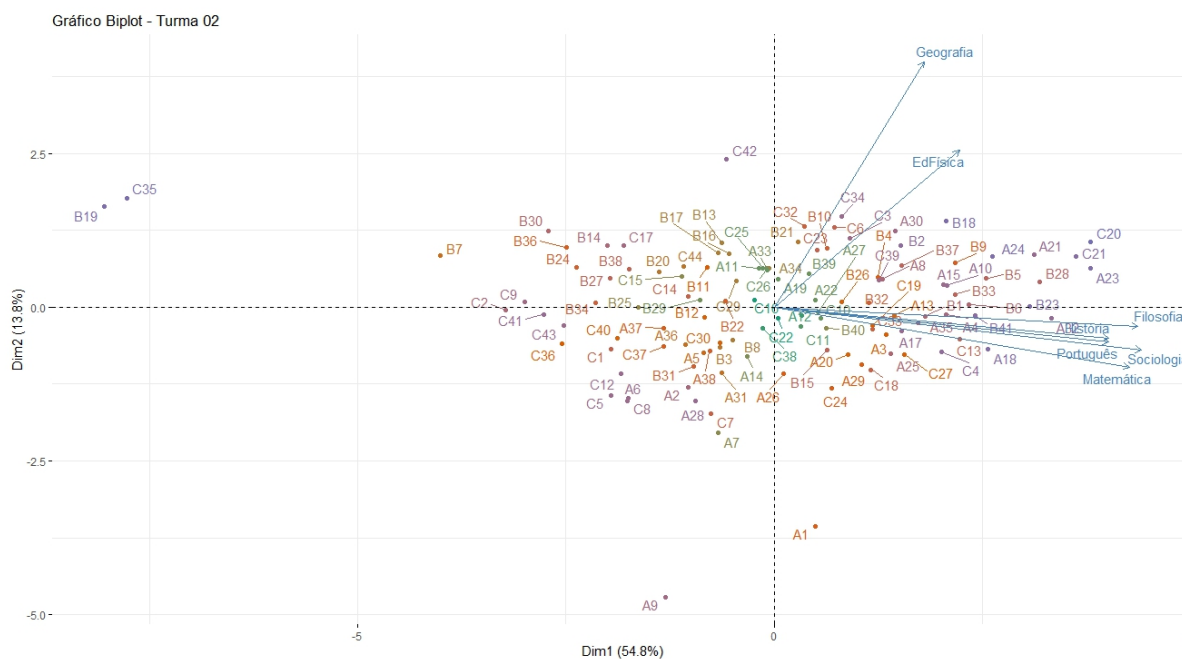
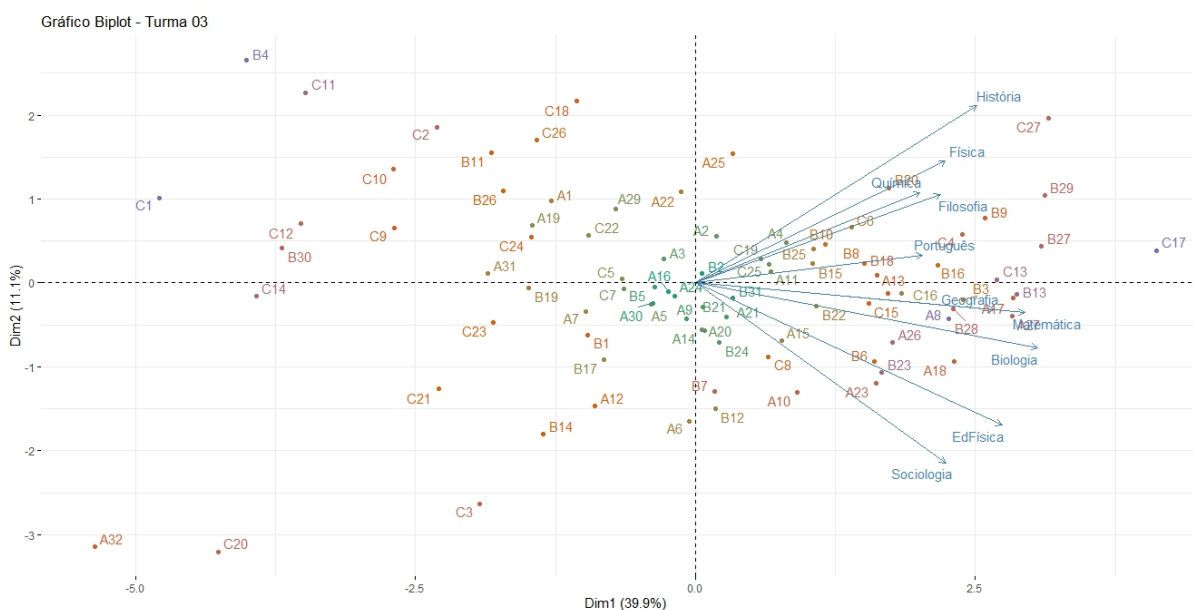


Figura 3.10: Gráfico Biplot para a turma 03 - relações entre disciplinas e estudantes



Considerações Finais

O acompanhamento educacional é uma tarefa que exige entre outras habilidades profissionais, o uso de uma metodologia que possa auxiliar na detecção de estudantes que apresentam necessidades de acompanhamentos diferenciados com propósito de melhoria da vida acadêmica.

A proposta da metodologia *k*-means utilizando as informações quantitativas das notas, contribuiu para explicar, 63,6%, 68,6% e 51% das variações dos rendimentos dos estudantes nas turmas 01, 02 e 03, respectivamente. Vale salientar que grande parte dos dados sociais, econômicos, entre outros, relativos aos estudantes não houve disponibilidade na sua totalidade. No entanto, entende-se que a captura dessas informações pode auxiliar o gestor no acompanhamento estudantil.

A metodologia aplicada aos dados nos dois primeiros bimestres de 2023 permitiu reduzir o número de informações a serem verificadas pelo gestor e a identificação dos alunos que possuem essas informações em comum. O *K*-means contribuiu para a formação de grupos de estudantes com notas mais similares, facilitando a tomada de decisão.

O descompasso observado entre as turmas 01, 02 e 03 de cada curso foi cada vez maior e também maior variação do rendimento acadêmico, exceto para o curso codificado com a letra B, cuja variação diminuiu das turmas 02 e 03. Podemos levantar algumas questões: o aumento da variabilidade das notas podem estar associadas às questões de identificação que o estudante tem com o curso? Estará associada ao bem-estar escolar? Será que a metodologia abordada em sala? Ou a falta de compromisso com a escola, a distância da escola, questões financeiras entre outros?

O trabalho aborda um panorama dos rendimentos acadêmicos sob a perspectiva de análise multivariada. A proposta da metodologia foi satisfatória, pois contribuiu para identificação de grupos menores de estudantes que facilita o acompanhamento escolar do gestor no sentido de tomar decisões assertiva à cerca da melhoria estudantil. Contudo, sugere-se como trabalhos futuros, realizar uma investigação sobre os aspectos externos que podem influenciar o rendimento acadêmico estudantil. Sugerimos, também, que os registros legais dos alunos, tais como, cidade em que reside, zona urbana ou rural, escola de egresso do estudante, status quanto à vulnerabilidade social e econômica, participações em olimpíadas, participações em atividades extra-classe relacionadas a esportes e condi-

ções especiais de aprendizagem sejam capturados pela escola para que se possa ter mais informações que auxiliarão na identificação e construção de tomadas de decisão.

Referências Bibliográficas

- ALEXANDRE. Sistemas de avaliação da educação básica no brasil. **Consultoria Legislativas, Brasília–DF**, 2015.
- BAKER, M. L.; SIGMON, J. N.; NUGENT, M. E. Truancy reduction: Keeping students in school. *juvenile justice bulletin*. ERIC, 2001.
- BOLIVAR, A. The comprehensive school in spain: A review of its development cycle and crises. **European Educational Research Journal**, SAGE Publications Sage UK: London, England, v. 14, n. 3-4, p. 347–363, 2015.
- BUSSAB, W. d. O.; MORETTIN, P. A. Estatística básica. In: **Estatística básica**. [S.l.: s.n.], 2010. p. xvi–540.
- DEMIR, K.; KARABEYOGLU, Y. A. Factors associated with absenteeism in high schools. **Eurasian Journal of Educational Research**, v. 16, n. 62, 2015.
- GARRY, E. M. **Truancy, first step to a lifetime of problems**. [S.l.]: US Department of Justice, Office of Justice Programs, Office of Juvenile, 1996.
- HARTIGAN, J. A.; WONG, M. A. A k-means clustering algorithm. *applied statistics*. **Applied Statistics**, v. 28, n. 100-108, p. 51, 1979.
- INEP. **Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira | Inep**. 2024. Disponível em: <<https://www.gov.br/inep/pt-br/assuntos/noticias/institucional/inep-completa-86-anos-de-fundacao::text=Exames>>. Acesso em: 12 de Março de 2024.
- INEP. **Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira | Inep**. 2024. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/saeb>>. Acesso em: 12 de Março de 2024.
- INEP, C. e. **Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira | Inep**. 2024. Disponível em: <<https://www.gov.br/inep/pt-br/assuntos/noticias/saeb/mec-e-inep-divulgam-resultados-do-saeb-e-do-ideb-2021::text=No>>. Acesso em: 12 de Março de 2024.
- INEP, P. **Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira | Inep**. 2024. Disponível em: <https://www.gov.br/inep/pt-br/assuntos/noticias/acoes-internacionais/divulgados-os-resultados-do-pisa-2022::text=Matem>>. Acesso em: 12 de Março de 2024.

JÚNIOR, F. T.; SANTOS, J. R. D.; MACIEL, M. de S. Análise da evasão no sistema educacional brasileiro. **Pesquisa e Debate em Educação**, v. 6, n. 1, p. 73–92, 2016.

LDB. **Lei de Diretrizes e Bases da Educação Nacional**. 2024. Disponível em: <https://www2.senado.leg.br/bdsf/bitstream/handle/id/642419/LDB_7ed.pdf>. Acesso em: 03 de Março de 2024.

LEHR, C. A.; SINCLAIR, M. F.; CHRISTENSON, S. L. Addressing student engagement and truancy prevention during the elementary school years: A replication study of the check & connect model. **Journal of education for students placed at risk**, Taylor & Francis, v. 9, n. 3, p. 279–301, 2004.

LIMA, I. dos S.; BARBOSA, C. Classificação do risco à evasão escolar por meio de protocolo eletrônico: um estudo realizado nas escolas públicas estaduais de ensino médio da cidade de feira de santana na bahia. **Pesquisa e Debate em Educação**, v. 7, n. 1, p. 203–218, 2017.

LOEBER, R.; FARRINGTON, D. P. Young children who commit crime: Epidemiology, developmental origins, risk factors, early interventions, and policy implications. **Development and psychopathology**, Cambridge University Press, v. 12, n. 4, p. 737–762, 2000.

MACQUEEN, J. Classification and analysis of multivariate observations. In: **5th Berkeley Symp. Math. Statist. Probability**. [S.l.: s.n.], 1967. p. 281–297.

MEC. **Parâmetros Curriculares Nacionais**. 1997. Disponível em: <<http://portal.mec.gov.br/seb/arquivos/pdf/livro01.pdf>>. Acesso em: 12 de Março de 2024.

PHILLIPS, R. W. **Educational facility age and the academic achievement and attendance of upper elementary school students**. [S.l.]: University of Georgia, 1997.

PNE. **Lei nº 13.005, de 25 de junho de 2014. Aprova o Plano Nacional de Educação-PNE e dá outras providências**. 2014. Disponível em: <<http://www.planalto.gov.br/ccivil03/ato2011-2014/2014/lei/113005.htm>>. Acesso em: 03 de Março de 2024.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2024. Disponível em: <<https://www.R-project.org/>>.

ROUSSEEUW, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. **Journal of computational and applied mathematics**, Elsevier, v. 20, p. 53–65, 1987.

SEBRAE. **DATA MPE Brasil**. 2004. Disponível em: <<https://datampe.sebrae.com.br/profile/geo/pedro-ii>>. Acesso em: 31 de Maio de 2024.

SHELDON, S. B. Improving student attendance with school, family, and community partnerships. **The Journal of Educational Research**, Taylor & Francis, v. 100, n. 5, p. 267–275, 2007.

SHELDON, S. B.; EPSTEIN, J. L. Getting students to school: using family and community involvement to reduce chronic absenteeism. **School Community Journal**, ERIC, v. 14, n. 2, p. 39–56, 2004.

SHIRASU, M. R.; ARRAES, R. d. A. Determinantes da evasão e repetência escolar. **ENCONTRO NACIONAL DE ECONOMIA**, v. 43, 2016.

SILVA, W. A. da; COSTA, F. A. Reflexões teóricas sobre o lugar e o papel das tecnologias digitais na formação inicial de professores em portugal. **Pesquisa e Debate em Educação**, v. 12, n. 1, p. 1–e35328, 2022.

SOARES, F. G. S.; SILVA, J. A. da. Análise multivariada aplicada para descrição do panorama educacional brasileiro. **Brazilian Journal of Development**, v. 10, n. 3, p. e68404–e68404, 2024.

SOUZA, V. F. de. Mineração de dados educacionais com aprendizagem de máquina. **Revista Educar Mais**, v. 5, n. 4, p. 766–787, 2021.

SYAKUR, M. et al. Integration k-means clustering method and elbow method for identification of the best customer profile cluster. In: IOP PUBLISHING. **IOP conference series: materials science and engineering**. [S.l.], 2018. v. 336, p. 012017.

TEASLEY, M. L. Absenteeism and truancy: Risk, protection, and best practice implications for school social workers. **Children & Schools**, Oxford University Press, v. 26, n. 2, p. 117–128, 2004.

TIBSHIRANI, R.; WALTHER, G.; HASTIE, T. Estimating the number of clusters in a data set via the gap statistic. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 63, n. 2, p. 411–423, 2001.

UNICEF et al. Indicadores da qualidade na educação, ação educativa. **Unicef, PNUD**, 2004.