UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

*modes*3.2.2.340K-*modes*subsubsection.3.2.2.3

Detecção de vulnerabilidade de estudantes do ensino fundamental público durante a pandemia de Covid-19 através de técnicas de agrupamento

Paula Ianishi

Dissertação de Mestrado do Programa de Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria (MECAI)

ICMC USP SÃO CARLOS

Paula Ianishi

Detecção de vulnerabilidade de estudantes do ensino fundamental público durante a pandemia de Covid-19 através de técnicas de agrupamento

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestra – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria. *VERSÃO REVISADA*

Área de Concentração: Matemática, Estatística e Computação

Orientador: Prof. Dr. Adriano Kamimura Suzuki

USP – São Carlos
Julho de 2021

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

I11d

Ianishi, Paula
Detecção de vulnerabilidade de estudantes do ensino fundamental público durante a pandemia de Covid-19 através de técnicas de agrupamento / Paula Ianishi; orientador Adriano Kamimura Suzuki. -- São Carlos, 2021.
64 p.

Dissertação (Mestrado - Programa de Pós-Graduação em Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria) -- Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2021.

1. Vulnerabilidade. 2. ensino público. 3. análise de agrupamento. 4. educação remota. 5. ensino na pandemia de Covid-19. I. Kamimura Suzuki, Adriano , orient. II. Título.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Saia Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

Paula Ianishi

Vulnerability detection of elementary school students during the Covid-19 pandemic using *clustering* techniques

Master dissertation submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP, in partial fulfillment of the requirements for the degree of the Master – Professional Masters in Mathematics, Statistics and Computing Applied to Industry. *FINAL VERSION*

Concentration Area: Mathematics, Statistics and Computing

Advisor: Prof. Dr. Adriano Kamimura Suzuki

USP – São Carlos
July 2021

Este trabalho é dedicado à todas as mulheres cientistas,
em especial da área de Exatas, que mesmo estando em menor número,
não deixaram de acreditar em si mesmas.

AGRADECIMENTOS

Agradeço, primeiramente, à Deus por me privilegiar com pais e um marido tão compreensíveis e pacientes. Em segundo lugar, gostaria de agradecer meus pais por sempre priorizarem meus sonhos, muitas vezes em detrimento de seus próprios. Ademais, gostaria de agradecer meu marido pelas conversas técnicas, mas principalmente por todo apoio emocional. Essa conquista é tão minha quanto de vocês.

Gostaria de agradecer, também, ao meu orientador prof. Dr. Adriano K. Suzuki por tornar essa defesa possível e pelas contribuições técnicas. Ademais, gostaria de agradecer ao Instituto de Ciências Matemáticas e Computação por ajudar a formar meu conhecimento.

Agradecimento, também, à diretora da escola que disponibilizou os dados para a realização da aplicação desse trabalho e por todas as discussões que sempre priorizaram ajudar os estudantes. Toda a minha admiração pelo trabalho que você e os demais profissionais da educação realizam.

Agradecimento especial às minhas amigas-irmãs Luciana, Marina e Gabriellen por sempre terem me acolhido em momentos de necessidade, que foram tantos! E ao o departamento de Estatística da Universidade Federal de São Carlos, principalmente, ao meu ex-orientador Rafael Izbicki por construirem tanto minha formação técnica na área de Estatística, mas principalmente minha formação pessoal.

“Quem nós somos não pode ser separado de onde viemos.”
(Malcolm Gladwell)

RESUMO

IANISHI, P. **Detecção de vulnerabilidade de estudantes do ensino fundamental público durante a pandemia de Covid-19 através de técnicas de agrupamento**. 2021. 63 p. Dissertação (Mestrado – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2021.

No contexto da pandemia de Covid-19, a educação, desde a básica até a superior, precisou introduzir aulas por meios virtuais. Parte dos estudantes, principalmente no setor público, não tinham acesso a equipamentos eletrônicos e/ou internet. Portanto, entender quais são os alunos que possuem esse tipo de vulnerabilidade foi fundamental para que as escolas pudessem emprestar equipamentos ou até direcionar recursos que empresas privadas ofereceram. Além disso, existe uma preocupação do setor educacional com o estado psicológico abalado que o isolamento social e mesmo o contágio de familiares e amigos provocou nos estudantes. Este trabalho se trata de um estudo de caso e tem como objetivo utilizar técnicas de agrupamento para identificar estudantes que apresentaram algum tipo de vulnerabilidade, durante o ínicio da pandemia, para que uma escola municipal da cidade de São Paulo pudesse atuar de acordo com a vulnerabilidade detectada. O método de agrupamento foi replicado várias vezes, de maneira a calcular probabilidades empíricas dos estudantes pertencerem a grupos vulneráveis para que a escola, em questão, pudesse priorizar o atendimento a esses estudantes e suas famílias.

Palavras-chave: Vulnerabilidade, ensino público, análise de agrupamento, educação remota, ensino na pandemia de Covid-19 .

ABSTRACT

IANISHI, P. **Vulnerability detection of elementary school students during the Covid-19 pandemic using *clustering* techniques**. 2021. 63 p. Dissertação (Mestrado – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2021.

In the context of the Covid-19 pandemic, education, from basic to higher, had to introduce classes through virtual means. Part of the students, mainly in the public sector, did not have access to electronic equipment and/or the internet. Therefore, understanding which students have this type of vulnerability was essential for schools to be able to lend equipment or even direct resources that private companies offered. In addition, there is a concern in the education sector with the psychological state that social isolation and even the contagion of family and friends provoked in students. This work is a case study and aims to use *clustering* techniques to identify students who presented some type of vulnerability during the beginning of the pandemic, so that a municipal school in the city of São Paulo could act in accordance with the vulnerability detected. The *clustering* method was replicated several times, in order to calculate empirical probabilities of students belonging to vulnerable groups so that the school in question could prioritize assistance to these students and their families

Keywords: Vulnerability, public education, *cluster* analysis, remote education, teaching in the Covid-19 pandemic.

LISTA DE ILUSTRAÇÕES

LISTA DE CÓDIGOS-FONTE

LISTA DE TABELAS

SUMÁRIO

CAPÍTULO

1

INTRODUÇÃO

A pandemia de Covid-19 trouxe um cenário inédito no mundo, no ano de 2020, nos mais diversos setores. No âmbito da educação, estima-se que cerca de 90% dos estudantes em todo o mundo precisaram se adaptar ao ensino remoto (ARRUDA, 2020; CRAWFORD *et al.*, 2020).

O isolamento social prolongado impacta as relações familiares, a rotina e a saúde mental de todas as comunidades, sobretudo aquelas de situação de vulnerabilidade socioeconômica (JOYE; MOREIRA; ROCHA, 2020). Além disso, as famílias estão sendo afetadas economicamente e devido à incerteza atrelada a situação, sentimentos como estresse e medo são comumente observados. Estudos têm apontado, também, que o fechamento de escolas aumenta o risco de violência contra crianças e adolescentes (SCHMIDT *et al.*, 2020).

Além do aspecto emocional, as comunidades mais vulneráveis possuem maior chance de não ter acesso à internet ou mesmo equipamentos eletrônicos que possibilitem os estudantes a frequentarem o ensino remoto (DIAS; PINTO, 2020). Nesse contexto, algumas instituições privadas fizeram doações de recursos financeiros e tecnológicos para o governo ou mesmo diretamente para as escolas.

É de extrema importância identificar estudantes em situação de vulnerabilidade para que sejam destinados recursos ou mesmo outros tipos de auxílio de acordo com as necessidades apontadas. Essa identificação pode ser feita por meio de questionários online ou por telefone, dado que, pelo contexto pandêmico, a interação presencial deve ser evitada.

Após a coleta de dados, existe a necessidade de identificar estudantes vulneráveis e priorizar determinados atendimentos. Esse processo pode ser feito de maneira manual, porém seria demasiadamente demorado e o resultado poderia ser tendencioso, no sentido de que o humano tomando as decisões poderia privilegiar estudantes conhecidos em detrimento à outros estudantes.

Quando o objetivo do estudo é identificar vulnerabilidade no sentido mais amplo da

palavra, ou seja, aquela definida por vários fatores como sociais, econômicos, familiares e psicológicos, vários autores propuseram a utilização de técnicas estatísticas não supervisionadas, entre eles Codeço *et al.* (2020) e Cohrs *et al.* (2013).

1.1 Objetivo Geral

Esse trabalho tem como objetivo estudar a multidimensionalidade da vulnerabilidade ao auxiliar uma escola municipal da cidade de São Paulo a identificar estudantes nessa situação durante o início da pandemia de Covid-19.

1.2 Organização do Trabalho

O Capítulo 2 apresenta uma breve descrição do conjunto de dados, bem como a análise descritiva. No Capítulo 3, apresentamos as técnicas estatísticas não supervisionadas utilizadas para lidar com o problema de identificação de vulnerabilidade. No Capítulo 4 apresentamos os resultados finais. E, finalmente, no Capítulo 5 fazemos uma conclusão sobre a metodologia e possíveis direções para trabalhos futuros.

CAPÍTULO

2

O CONJUNTO DE DADOS

O conjunto de dados é proveniente de um questionário enviado a totalidade de 800 estudantes de uma escola pública de ensino fundamental da cidade de São Paulo/SP. Os estudantes estão entre o primeiro e nono ano escolar.

O questionário foi elaborado pela própria direção da escola com o objetivo de mapear as necessidades dos estudantes durante a quarentena. O mesmo foi dividido em quatro sessões principais: bem-estar do estudante, acessibilidade a internet, equipamentos eletrônicos e materiais didáticos, vida familiar e exposição ao coronavírus.

O questionário foi enviado para os *e-mails* da totalidade de 800 estudantes e disponibilizado no site da escola. Caso o estudante tivesse dificuldade no preenchimento, ele era aconselhado a pedir ajuda de um adulto por ele responsável.

Para o caso dos estudantes que não responderam o questionário, os professores ficaram responsáveis por tentar contato com as famílias e para o caso da resposta não ter chegado por falta de internet ou equipamento, as respostas poderiam ser coletadas via telefone.

2.1 Descrição das sessões do questionário

2.1.1 Bem-estar do estudante

Essa sessão teve como objetivo avaliar a saúde mental dos estudantes. Uma psicóloga voluntária auxiliou a direção na construção das perguntas. Ela foi composta de três perguntas:

- Em uma escala de 1 a 5, como você se sente em casa nesse período de isolamento social? (sendo 1: péssimo, 2: mal, 3: regular, 4: bem, 5: excelente)
- Selecione as opções que você sentiu até o momento na quarentena: segurança, felicidade, alegria, ansiedade, tristeza e cansaço

- Quais dessas atividades você tem feito na quarentena: atividades do livro didático, leio, brinco, ver televisão, ver streaming, ouvir música, fazer chamada de vídeo com amigos e familiares

2.1.2 Acesso à internet, equipamentos eletrônicos e materiais

Essa sessão teve como objetivo avaliar a acessibilidade dos alunos a questões que os possibilitariam estudar remotamente. Ela foi composta de sete perguntas:

- Quais equipamentos eletrônicos você mais utiliza? (computador, notebook, tablet, celular, não uso nenhum equipamento eletrônico)
- Quais materiais didáticos você possui em casa? (livros didáticos, “trilhas da aprendizagem", “minha biblioteca")
- Com que frequência você pode utilizar os equipamentos? (tenho livre acesso, acesso restrito, não tenho acesso)
- Avalie sua conexão com a internet (boa, não muito boa, não tenho conexão)
- Qual é o melhor horário para você utilizar o equipamento eletrônico? (qualquer horário, manhã, tarde, em nenhum momento)
- Você consegue mexer no equipamento eletrônico e navegar na internet sozinho? (sim, as vezes peço ajuda, não)
- Pergunte a um adulto da sua família: com que frequência sua família pode te ajudar em atividades escolares na quarentena? (diariamente, a cada dois dias, uma vez por semana, raramente)

2.1.3 Vida familiar

Essa sessão teve como objetivo avaliar como a pandemia de Covid-19 afetou as relações familiares dos estudantes e como isso se traduz, ou não, no lado socioeconômico da vulnerabilidade. Ela foi composta de seis perguntas:

- Quantos adultos estão morando com você na quarentena? (1, 2, 3, 4, 5 ou mais)
- Quantas crianças estão morando com você na quarentena? (1, 2, 3, 4, 5 ou mais)
- Quem está cuidando das crianças e/ou adolescentes na sua casa? (mãe, pai, outro responsável da família, outro responsável que não é da família, nenhum adulto)
- Como é seu local de estudo? (silencioso, movimentado, muito movimentado, não tenho um lugar para estudar)

- Pergunte a um adulto da sua família: qual é a fonte de renda da sua família na quarentena? (bolsa família, auxílio emergencial, trabalho informal, trabalho formal, aposentadoria ou pensão, doações)
- Pergunte a um adulto da sua família: a renda mensal da sua família diminiu durante a quarentena (não, um pouco, muito, totalmente)

2.1.4 Exposição ao coronavírus

Essa sessão teve como objetivo avaliar quantos e quais estudantes estavam expostos ao vírus e quantos e quais estavam respeitando as medidas de prevenção indicadas pela OMS e como isso se traduz, ou não, no lado social da vulnerabilidade. Ela foi composta de três perguntas:

- Você faz parte do grupo de risco (tem diabetes, problemas cardíacos, respiratórios ou autoimunes)? (sim, não)
- Alguém que mora com você apresentou sintomas de Covid-19 ou foi diagnosticado por um médico? (sim mas sem diagnóstico, sim com diagnóstico, ninguém apresentou sintomas)
- Ao voltar da rua, quais medidas de higiene você e sua família adotaram? (lavamos as mão, retiramos sapatos, roupas, tomamos banho. etc)

2.2 Análise descritiva dos dados

Da totalidade de 800 estudantes, a escola obteve 578 (72%) respostas únicas distribuídas nos anos escolares de 1º a 9º ano do ensino fundamental (ver Tabela 1). É importante destacar que não pretendemos extrapolar os resultados obtidos para a população dos 800 estudantes da escola, pois não há garantias de que os estudantes não respondentes possuem as mesmas características dos respondentes. Na realidade, o fato de um aluno não ter respondido já é um indicativo de vulnerabilidade.

A Figura 1 mostra a distribuição dos respondentes por ano escolar. Os anos escolares que obtiveram mais respostas foram 3º ao 5º ano e 8º e 9º ano.

2.2.1 Bem-estar do estudante

Podemos observar que cerca de 89% dos estudantes estavam se sentindo de regulares até excelentes no período da pandemia (ver Tabela 2).

Em relação aos sentimentos presentes durante a pandemia, os estudantes poderiam selecionar múltiplas respostas. Os sentimentos mais selecionados foram segurança (71%) e felicidade por estar com a família (69%), sendo que 496 (85%) estudantes afirmaram estar sentindo segurança ou felicidade (ver Tabela 4). Em relação aos sentimentos negativos, os mais

Tabela 1 – Ano escolar dos respondentes

Ano escolar	Quantidade (%)
1º ano	45 (7,79%)
2º ano	51 (8,82%)
3º ano	77 (13,32%)
4º ano	75 (12,98%)
5º ano	80 (13,84%)
6º ano	56 (9,69%)
7º ano	56 (9,69%)
8º ano	71 (12,28%)
9º ano	67 (11,59%)

Tabela 2 – Questão 1: Em uma escala de 1 a 5, como você se sente em casa nesse período de isolamento social? (sendo 1: péssimo, 2: mal, 3: regular, 4: bem, 5: excelente)

Resposta	Quantidade (%)
1	22 (3,80%)
2	39 (6,75%)
3	193 (33,39%)
4	169 (29,24%)
5	155 (26,82%)

citados foram preocupação (46%) e ansiedade (43%) (ver Tabela 3), sendo que 387 (67%) estudantes afirmaram estar preocupados ou ansiosos (ver Tabela 5).

Tabela 3 – Questão 2. Selecione as opções que você sentiu até o momento na quarentena

Resposta	Quantidade (%)
Segurança	413 (71,45%)
Felicidade por estar com a família	401 (69,37%)
Alegria por brincar	297 (51,38%)
Tranquilidade	233 (40,31%)
Preocupação	267 (46,19%)
Medo	170 (29,41%)
Ansiedade	249 (43,08%)
Tristeza	177 (30,62%)
Angústia	142 (24,57%)
Cansaço	172 (29,76%)

Tabela 4 – Cruzamento dos sentimentos de segurança e felicidade

		Felicidade	
		Não	**Sim**
Segurança	**Não**	82 (14,19%)	83 (14,36%)
	Sim	95 (16,44%)	318 (55,01%)

Em relação às atividades escolares, 483 (83%) estudantes afirmaram estar fazendo atividades do livro didático ou o roteiro escolar (ver Tabela 7). Em relação a outras atividades

Tabela 5 – Cruzamento dos sentimentos de preocupação e ansiedade

		Ansiedade	
		Não	**Sim**
Preocupação	**Não**	191 (33,05%)	120 (20,76%)
	Sim	138 (23,87%)	129 (22,32%)

feitas na pandemia, as mais citadas foram ficar no celular, ficar na internet, brincar, ver streaming e ajudar em casa (ver Tabela 6).

Tabela 6 – Questão 3. Quais atividades você tem feito na quarentena?

Resposta	**Quantidade (%)**
Faço atividade do livro didático	296 (51,21%)
Faço roteiro	390 (67,47%)
Leio	386 (66,78%)
Brinco	423 (73,18%)
Fico no celular	433 (74,91%)
Fico na internet	422 (73,01%)
Faço chamada de vídeo	413 (71,45%)
Jogo	381 (65,91%)
Vejo TV	384 (66,43%)
Vejo streaming	420 (72,66%)
Ouço música	417 (72,14%)
Ajudo em casa	424 (73,36%)

Tabela 7 – Cruzamento das atividades do livro didático e roteiro

		Faço atividade do roteiro	
		Não	**Sim**
Faço atividade do livro didático	**Não**	95 (16,43%)	187 (32,35%)
	Sim	93 (16,09%)	203 (35,13%)

2.2.2 Acesso à internet, equipamentos eletrônicos e materiais

Afirmaram não utilizar computador, celular ou tablet, apenas 3% dos estudantes (ver Tabela 8). Os estudantes que afirmaram não ter equipamento eletrônico ou internet para assistir as aulas de maneira remota foram imediatamente auxiliados com o empréstimo de computadores da própria escola e/ou instalação de internet doada por uma empresa provedora.

Tabela 8 – Questão 4. Quais equipamentos eletrônicos você mais utiliza?

Resposta	**Quantidade (%)**
Computador	336 (58,13%)
Celular	466 (80,62%)
Tablet	141 (24,39%)
Não uso nenhum equipamento	16 (2,77%)

Apesar de apenas 51% dos estudantes estarem fazendo atividades do livro didático (ver Tabela 6), 85% dos estudantes possuem esse material em casa (ver Tabela 9) e quase metade dos alunos reponderam que tem livre acesso para utilizar os equipamentos eletrônicos (ver Tabela 10).

Tabela 9 – Questão 5. Quais materiais didáticos você possui em casa?

Resposta	**Quantidade (%)**
Livros didáticos	492 (85,12%)
“Trilhas de aprendizagem"	151 (26,12%)
“Minha biblioteca"	285 (49,31%)

Tabela 10 – Questão 6. Com que frequência você pode utilizar os equipamentos?

Resposta	**Quantidade (%)**
Não tenho acesso	5 (0,86%)
Tenho acesso restrito (só em alguns momentos do dia)	286 (49,48%)
Tenho livre acesso	287 (49,65%)

Cinco alunos (1%) responderam que não tinham acesso a internet e foram imediatamente auxiliados através de doações de empresas do ramo (ver Figura 9). A grande maioria (82%) dos respondentes afirmaram ter boa conexão de internet (ver Figura 11). Infelizmente, as empresas do ramo de internet auxiliaram apenas as famílias que não tinham conexão e não foi possível auxiliar as famílias que tinham conexão limitada (17%).

Tabela 11 – Questão 7. Avalie sua conexão com a internet (boa, não muito boa, não tenho conexão)

Resposta	**Quantidade (%)**
Não tenho conexão de internet	5 (0,86%)
Sim, mas minha conexão de internet não é muito boa	98 (16,95%)
Sim, tenho boa conexão de internet	475 (82,18%)

Por meio da questão 8, foram idenficados estudantes que apesar de terem acesso à internet e terem equipamentos eletrônicos tinham que dividir o equipamento com outros familiares e não conseguiam atender a 100% das atividades escolares. Esses casos representam 9% dos respondentes (ver Tabela 12). No caso de empréstimo de equipamentos, essas famílias foram auxiliadas, porém após o auxílio prestado para as famílias que não possuiam nenhum equipamento.

Tabela 12 – Questão 8. Qual é o melhor horário para você utilizar o equipamento eletrônico?

Resposta	**Quantidade (%)**
Manhã	101 (17,47%)
Tarde	274 (47,40%)
Em qualquer horário	257 (44,46%)
Em nenhum horário	53 (9,17%)

Afirmaram precisar algumas vezes de ajuda para manusear os equipamentos eletrônicos e acessar a internet 36% dos estudantes e, 9% dos alunos precisavam de ajuda em 100% do

tempo (Tabela 13). Paralelamente, 23% dos alunos afirmaram que não tinham ajuda diária de um adulto para realizar as atividades escolares (Tabela 14).

Dos 263 alunos que afirmaram necessitar de ajuda para acessar os equipamentos, frequentemente ou não, 95% afirmaram ter ajuda diária ou a cada dois dias. 6 dos estudantes desse público afirmaram que tem ajuda uma vez por semana e outros 6 estudantes têm ajuda raramente (tabela 15).

Tabela 13 – Questão 9. Você consegue mexer no equipamento eletrônico e navegar na internet sozinho?

Resposta	**Quantidade (%)**
Não, preciso de ajuda	53 (9,17%)
Sim, mas as vezes peço ajuda pra alguém	210 (36,33%)
Sim, tranquilamente	315 (54,50%)

Tabela 14 – Questão 10. Pergunte a um adulto da sua família: com que frequência sua família pode te ajudar em atividades escolares na quarentena?

Resposta	**Quantidade (%)**
A cada dois dias	97 (16,78%)
Diariamente	448 (77,51%)
Raramente	18 (3,11%)
Uma vez por semana	15 (2,60%)

Tabela 15 – Frequência de ajuda familiar para estudantes que necessitam de auxílio para navegar na internet e manusear o equipamento eletrônico

Frequência de ajuda	**Quantidade (%)**
A cada dois dias	48 (18,25%)
Diariamente	203 (77,19%)
Raramente	6 (2,28%)
Uma vez por semana	6 (2,28%)

2.2.3 Vida familiar

Afirmaram estar morando com 3 ou mais adultos 29% dos estudantes (Tabela 16) e, apenas 7% dos estudantes afirmaram estar morando com 3 ou mais crianças (Tabela 17).

Tabela 16 – Questão 11. Quantos adultos estão morando com você na quarentena?

Resposta	**Quantidade (%)**
1	96 (16,61%)
2	314 (54,33%)
3	95 (16,43%)
4	52 (8,99%)
5 ou mais	21 (3,64%)

Tabela 17 – Questão 12. Quantas crianças estão morando com você na quarentena?

Resposta	**Quantidade (%)**
0	199 (34,43%)
1	245 (42,39%)
2	90 (15,57%)
3	29 (5,02%)
4	8 (1,38%)
5 ou mais	7 (1,21%)

90% dos estudantes afirmaram estar morando com a mãe e apenas 56% afirmaram estar morando com o pai (ver Tabela 18). 7 crianças afirmaram não estar sendo cuidadas por nenhum responsável.

Tabela 18 – Questão 13. Quem está cuidando das crianças e/ou adolescentes na sua casa?

Resposta	**Quantidade (%)**
Mãe	523 (90,48%)
Pai	324 (56,06%)
Outro responsável da família	103 (17,82%)
Outro responsável fora da família	16 (2,77%)
Sem cuidado	7 (1,21%)

Apenas 67% dos estudantes afirmaram ter um local de estudo silencioso e 31% responderam que tem um local de estudo movimentado e que isso causa alguma ou muita dificuldade de concentração (ver Tabela 19).

Tabela 19 – Questão 14. Como é seu local de estudo?

Resposta	**Quantidade (%)**
Movimentado, tenho alguma dificuldade para me concentrar	161 (27,86%)
Muito movimentado, tenho bastante dificuldade para me concentrar	20 (3,46%)
Não tenho um lugar para estudar	11 (1,90%)
Silencioso, consigo me concentrar	386 (66,78%)

Apenas 26% das famílias não tiveram suas rendas afetadas pela pandemia (ver Tabela 20) e paralelamente, 40% das famílias citaram como fonte de renda o trabalho informal (ver Tabela 21).

Tabela 20 – Questão 15. Pergunte a um adulto da sua família: a renda mensal da sua família diminiu durante a quarentena

Resposta	**Quantidade (%)**
Não, continua a mesma	153 (26,47%)
Sim, muito	153 (26,47%)
Sim, totalmente	49 (8,48%)
Sim, um pouco	223 (38,58%)

Tabela 21 – Questão 16. Pergunte a um adulto da sua família: qual é a fonte de renda da sua família na quarentena?

Resposta	Quantidade (%)
Bolsa família	34 (5,89%)
Auxílio emergencial	107 (18,51%)
Trabalho informal	229 (39,61%)
Trabalho formal	308 (53,29%)
Aposentadoria	49 (8,48%)
Doações	35 (6,06%)

2.2.4 Exposição ao coronavírus

Afirmam fazer parte do grupo de risco para o novo coronavírus, 16% dos estudantes (ver Tabela 22) e, paralelamente, 12% dos estudantes apresentaram algum sintoma da doença ou moram com alguém que apresentou sintomas. Dos estudantes ou familiares que apresentaram sintomas (69 respondentes), quase 80% não foram diagnosticados por um médico (ver Tabela 23).

Tabela 22 – Questão 17. Você faz parte do grupo de risco?

Resposta	Quantidade (%)
Não	484 (83,74%)
Sim	94 (16,26%)

Tabela 23 – Questão 18. Alguém que mora com você, ou você mesmo, apresentou sintomas de Covid-19 ou foi diagnosticado por um médico?

Resposta	Quantidade (%)
Ninguém apresentou sintomas	509 (88,06%)
Sim, apresentou sintomas e foi diagnosticado	14 (2,42%)
Sim, apresentou sintomas, mas não foi diagnosticado	55 (9,52%)

Em relação às medidas que precisam ser adotadas ao voltar da rua, a medida mais adotada é lavar as mãos, porém 8% das famílias ainda não adotavam essa medida. Apenas 72% das famílias afirmaram também retirar as roupas ao chegar em casa (ver Tabela 24).

Tabela 24 – Questão 19. Ao voltar da rua, quais medidas de higiene você e sua família adotaram?

Resposta	Quantidade (%)
Retira o sapato	462 (79,93%)
Retira a roupa	416 (71,97%)
Lava as mãos	531 (91,87%)
Higieniza os produtos trazidos de fora	473 (81,83%)
Não adota nenhuma medida	4 (0,69%)

CAPÍTULO

3

TÉCNICAS DE AGRUPAMENTO

Nesse capítulo fazemos uma descrição do método utilizado nesse trabalho. Na Seção 3.1 descrevemos o problema de aprendizado não-supervisionado e fazemos uma breve revisão da literatura. Por fim, na Seção 3.2 descrevemos uma das técnicas mais utilizadas dentro do aprendizado não-superviosionado, a técnica de *clustering*.

3.1 Aprendizado não-supervisionado

O aprendizado não-superviosinado se caracteriza pela presença de variáveis $X_1, X_2, ..., X_p$ medidas em n observações, porém não temos uma variável resposta Y associada a essas p variáveis (JAMES *et al.*, 2013). Ao invés de realizar a tarefa de predição, o objetivo é encontrar características sobre as variáveis $X_1, X_2, ..., X_p$ sem a ajuda de um supervisor Y. Por exemplo, a análise de *cluster* busca encontrar subregiões do espaço X, dado que cada sub-região seja composta por observações similares entre si (FRIEDMAN; HASTIE; TIBSHIRANI, 2001).

Para o aprendizado supervisionado, temos medidas de qualidade de ajuste que são medidas diretas de sucesso. No sentido de avaliar a efetividade, o aprendizado não-supervisionado é muito mais desafiador, já que não existe nenhum tipo de "supervisor". Dada essa característica subjetiva da técnica, há uma crescente quantidade de métodos disponíveis na literatura.

Alguns métodos propõem melhorias para as tradicionais metodologias como, por exemplo, uma melhoria proposta para o método hierárquico de *clusters* que lida automaticamente com dados de característica *outlier* (ALMEIDA *et al.*, 2007). Outros autores estudam métodos em que os subgrupos de observações não são distintos, como o caso de algoritmos de *fuzzy clustering* (HUANG; CHUANG; CHEN, 2011) e outros, ainda estudam adaptações desse método para dados categóricos (HUANG; NG, 1999). Alguns autores ainda propõem variações de métricas de distância para agrupar observações semelhantes (TSAI; LIN, 2011).

Assim, como são inúmeras as variações dos métodos não-supersionados, sua importância

nas mais diversas aplicações têm crescido nas últimos décadas (FRIEDMAN; HASTIE; TIBSHIRANI, 2001). Na área da educação, que é o objeto de estudo dessa dissertação, há estudos referentes a motivação dos estudantes na educação física (ULLRICH-FRENCH; COX, 2009), outros autores estudaram o agrupamento de universidades e estudantes para finalidade de exploração de dados (HUBERTY; JORDAN; BRANDT, 2005), outros ainda estudaram o aprendizado dos estudantes através de *hiperlinks* da internet (ANTONENKO; TOY; NIEDERHAUSER, 2012). Ainda há aqueles que fizeram metanálise com *clusters* para a resolução de problemas na área da educação (VASCONCELOS *et al.*, 2007).

3.2 Análise de agrupamento

Análise de agrupamento ou análise de *cluster* são termos genéricos para denominar métodos que procuram subgrupos em um conjunto de dados. O objetivo geral desses métodos é encontrar grupos em que as observações dentro de cada grupo sejam semelhantes entre si e ao mesmo tempo, as observações pertencentes a diferentes grupos sejam dissimilares entre si (JAMES *et al.*, 2013).

Usualmente, a medida de dissimilaridade acima descrita, é computada por meio da distância euclidiana, porém existem inúmeras variações dessa métrica (TSAI; LIN, 2011). Para o caso p-dimensional e tomando como métrica de dissimilaridade a distância euclidiana, a dissimilaridade entre as observações i e i' é dada por (FRIEDMAN; HASTIE; TIBSHIRANI, 2001):

$$D(\mathbf{x_i}, \mathbf{x_{i'}}) = \sum_{j=1}^{p} d_j(x_{ij}, x_{i'j}) = \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2,$$

sendo $\mathbf{x_i} = (x_{i1}, ..., x_{ip})$ e $\mathbf{x_{i'}} = (x_{i'1}, ..., x_{i'p})$

Nas próximas seções, iremos focar nos dois principais métodos de análise de agrupamento: o método hierárquico e o método k-*means*, que é o principal método não-hierárquico. A principal diferença entre esses métodos, é que para utilizar o segundo método, precisamos de um número pré-definido de *clusters*. Em contrapartida, no método hieráraquico, obtemos uma figura com visual de árvore, chamada dendograma que nos possibilita visualizar cada possível número de *clusters* (JAMES *et al.*, 2013). Ademais, na Seção **??** apresentaremos uma variação do método k-*means* para dados de natureza caregórica chamada de método k-*modes*, já que o conjunto de dados que irá compor a aplicação dessa dissertação tem essa natureza.

3.2.1 Métodos hierárquicos de agrupamento

A vantagem de utilizar métodos hierárquicos de agrupamento é que eles não necessitam de um número de *clusters* pré-definido. A ideia é construir um gráfico com visual de árvore chamado de dendograma.

O dendograma é uma representação hierárquica em que os *clusters* de cada nível são obtidos pela junção de *clusters* do nível que está imediatamente abaixo. No nível mais baixo, cada *cluster* contém uma única observação. Em contrapartida, no nível mais alto, temos um único *cluster* com todas as observações. Além disso, a altura de cada nó é proporcional ao valor da dissimilaridade entre os dois grupos filhos (FRIEDMAN; HASTIE; TIBSHIRANI, 2001).

O algoritmo da construção do dendograma começa tratando cada observação como um *cluster* diferente. O próximo passo é, em um novo nível, unir os dois *clusters* que são mais similares resultando em $n-1$ *clusters*. O algoritmo procede dessa maneira até que haja apenas um *cluster* com todas as observações (JAMES *et al.*, 2013).

Como citado na Seção 3.2, a medida de dissimilaridade mais utilizada na literatura é a distância euclidiana. Porém, essa definição foi feita para encontrar a distância, ou dissimilaridade, entre duas observações. Os métodos hierárquicos necessitam da definição de dissimilaridade entre pares de *clusters*, compostos por várias observações.

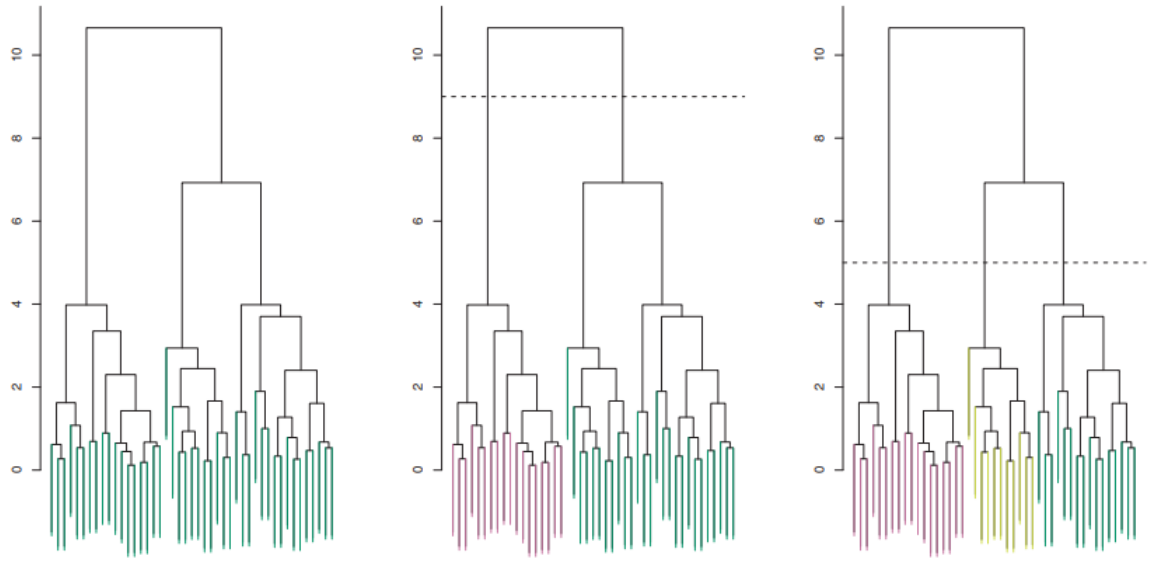
Existem algumas maneiras de definir a dissimilaridade entre grupos de observações, abaixo vamos descrever as principais delas (JAMES *et al.*, 2013):

- ***Complete linkage*** : a ideia consiste em computar todas as distâncias entre duas observações de *clusters* distintos. Esse método irá considerar a distância máxima como a distância entre os *clusters*.
- ***Single linkage*** : como no método anterior, todas as distâncias entre duas observações de *clusters* distintos serão computadas, porém esse método irá considerar a distância mínima como a distância entre os *clusters*.
- ***Average linkage*** : como nos métodos anteriores, todas as distâncias entre duas observações de *clusters* distintos serão computadas, porém esse método irá considerar a distância média como a distância entre os *clusters*.
- ***Centroid linkage*** : diferente dos métodos anteriores, esse método computa apenas a distância entre os centróides dos dois *clusters* de interesse.

Após definidos a medida de dissimilaridade e o *linkage*, obtemos o dendograma da esquerda na Figura 1, nesse caso foram usadas a distância euclidiana com complete *linkage*. O próximo passo é definir um corte horizontal, representado pela linha pontilhada. Por exemplo, o dendograma do meio apresenta um corte no número 9 do eixo vertical, o que produz apenas dois *clusters* distintos. Já o dendograma da direita apresenta um corte no número 5 do eixo vertical, o que nos leva a três *clusters* distintos (ver Figura 1).

Uma das desvantagens do método hierárquico é que a escolha do corte horizontal é subjetiva (FRIEDMAN; HASTIE; TIBSHIRANI, 2001). Todavia, a ideia principal é que observações dissimilares sejam atribuídas a *clusters* diferentes. A Figura 1 adaptada de James *et*

Figura 1 – Exemplo de dendogramas e cortes para definir o número de *clusters*



Fonte: Adaptado de James *et al.* (2013)

al. (2013), apresenta um dendograma a direita que separou as observações amarelas e verdes em *clusters* distintos, enquanto o dendograma do centro uniou essas mesmas observações em um único cluster. Sabemos que quanto mais alto é o nó que une dois grupos, mais dissimilares os mesmos são. Como as observações amarelas e verdes foram unidas em um ponto muito alto do eixo vertical, é compreensível a escolha de separar as observações em *clusters* distintos.

3.2.2 Métodos não-hierárquicos de agrupamento

3.2.2.1 Definição do número de clusters (k)

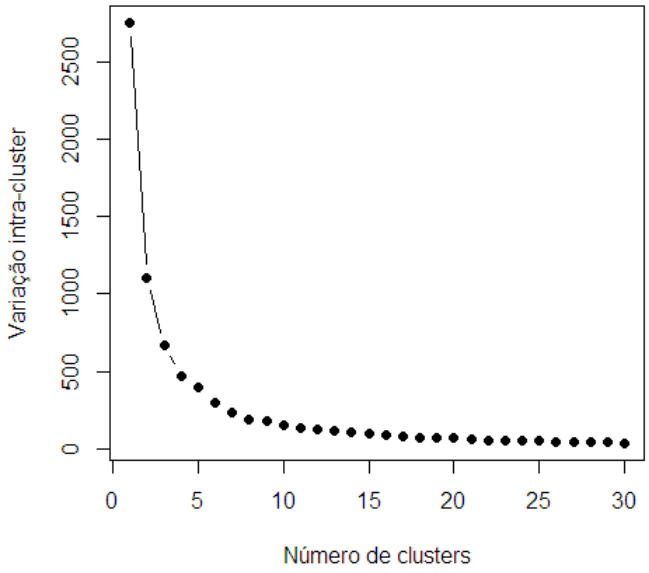
Uma das principais maneiras de definir o número de *clusters*, k, é dada através do próprio método hierárquico. A ideia é rodar o algoritmo definido na Seção 3.2.1, obter o dendograma e através dele definir o número k para posteriormente rodar algum algoritmo não-hierárquico.

Outra maneira para encontrar o número ótimo de *clusters* é através do gráfico de silhueta. A ideia é definir uma medida que possa capturar a variabilidade intra-*cluster* e computá-la para cada um dos possíveis valores de k (FRIEDMAN; HASTIE; TIBSHIRANI, 2001). Supondo que a medida de distância escolhida seja a distância euclidiana, a variação intra-*cluster* do q-ésimo *cluster* seria dada por (JAMES *et al.*, 2013):

$$W(C_q) = \frac{1}{|C_q|} \sum_{i,i' \in C_q} \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2,$$

sendo $|C_k|$ o número de observações pertencentes ao q-ésimo cluster. Ou seja, a variação intra-

Figura 2 – Exemplo de gráfico de silhueta para determinar o número ótimo de *clusters*



Fonte: Elaborada pela autora.

cluster é a soma das distâncias eucludianas de todos os pares de observações pertencentes a um mesmo *cluster* dividido pelo número de observações pertencentes ao q-ésimo cluster.

Dado que o objetivo da clusterização é encontrar grupos com observações semelhantes, podemos traduzir esse objetivo como minimizar a soma de todas as variações intra-cluster, ou matematicamente, minimizar $\sum_{k=1}^{K} W(C_k)$ (JAMES *et al.*, 2013). Todavia, sempre procuramos modelos parcimoniosos, para o caso da análise de agrupamento, a complexidade seria introduzida ao aumento o número de *clusters* k. A Figura 2 mostra exatamente esse *trade-off* e nesse caso, a queda da variação intra-*cluster* após $k = 4$ não é tão relevante a ponto de ser necessário introduzir complexidade ao modelo.

3.2.2.2 k-means

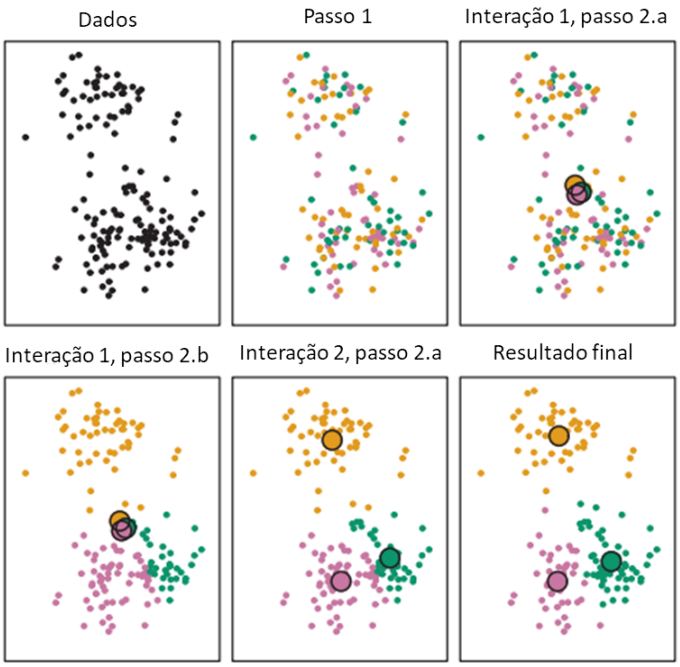
Dados $C_1, ..., C_k$ representando conjuntos de observações que contém cada cluster, o método k-*means* satisfaz duas propriedades matemáticas (JAMES *et al.*, 2013):

- $C_1 \bigcup C_2 \bigcup ... \bigcup C_k = \{1, ..., n\}$, ou seja, cada observação pertence a pelo menos um dos k *clusters*.
- $C_k \bigcap C'_k = \emptyset$, ou seja, nenhuma observação pertence a mais de um cluster.

O algoritmo k-*means* consiste de dois passos. Inicialmente, cada observação será atribuída

Figura 3 – Exemplo do progresso do algoritmo k-*means*

.



Fonte: Adaptado de James *et al.* (2013)

para um dos *k clusters* de maneira aleatória (Passo 1). O segundo passo se inicia com o cálculo dos centróides $(m_1, ..., m_k)$ dado pelas médias das variáveis para observações pertencentes a um mesmo *cluster* (Passo 2.a). Finalmente, cada observação x_i é atribuída ao centróide que se encontra mais próximo dela (Passo 2.b). Matematicamente, desejamos minimizar $(x_i - m_k)^2$ (FRIEDMAN; HASTIE; TIBSHIRANI, 2001). O segundo passo é repetido até que as atribuições das observações aos *clusters* não mudem.

A Figura 3 representa um exemplo do progresso do algoritmo k-*means*. Inicialmente, no passo 1, cada uma das observações foi atribuída a um *cluster* de maneira aleatória. Por esse motivo, após os centróides serem calculados, percebemos que estão praticamente sobrepostos (interação 1 no passo 2.a). Ainda na interação 1 no passo 2.b as observações são atribuídas ao *cluster* de centróide mais próximo. Finalmente, depois de 10 iterações, temos a representação gráfica do resultado final.

3.2.2.3 K-modes

Em 1997, Huang propôs uma variação ao método *k-means* cuja proposta era clusterizar dados categóricos. O autor propôs três grandes mudanças no algoritmo. A primeira era usar alguma medida de dissimilaridade que pudesse lidar com dados categóricos não ordinais, já que a distância euclidiana funciona apenas para dados numéricos. A segunda era trocar a média pela moda, isto é, o valor mais frequente, no cálculo dos centróides. E finalmente, a terceira, é utilizar um método baseado em frequência para atualizar as modas (HUANG, 1997).

A medida de dissimilaridade proposta é baseada no número de desigualdades dos p atributos entre dois objetos. Quanto maior o número de desigualdades, mais dissimilares os dois objetos são. Formalmente, a medida de dissimilaridade entre x_i e $x_{i'}$ é dada por (HUANG, 1997):

$$d(x_i, x_{i'}) = \sum_{j=1}^{p} \delta(x_{ij}, x_{i'j})$$

em que,

$$\delta(x_{ij}, x_{i'j}) = \begin{cases} 0, & \text{se } x_{ij} = x_{i'j} \\ 1, & \text{se } x_{ij} \neq x_{i'j} \end{cases}$$

Observe que $d(x_{ij}, x_{i'j})$ dá a mesma importância para cada categoria de um atributo. Se levarmos em conta a frequência das categorias, temos (HUANG, 1997):

$$d_{\chi^2}(x_i, x_{i'}) = \sum_{j=1}^{p} \frac{(n_{x_{ij}} + n_{x_{i'j}})}{n_{x_{ij}} n_{x_{i'j}}} \delta(x_{ij}, x_{i'j})$$

sendo n_{x_i} e $n_{x'_i}$ o número de observações no conjunto de dados que tem categorias x_{ij} e $x_{i'j}$ para o atributo j. A medida $d_{\chi^2}(x_i, x_{i'})$ é chamada de distância qui-quadrado e a mesma dá peso maior para categorias pouco frequentes.

O algoritmo *k-modes*, então, segue os mesmos passos do algoritmo *k-means* apenas trocando a distância euclidiana pela distância qui-quadrado e tendo o cálculo dos centróides feito através da moda em vez da média.

CAPÍTULO

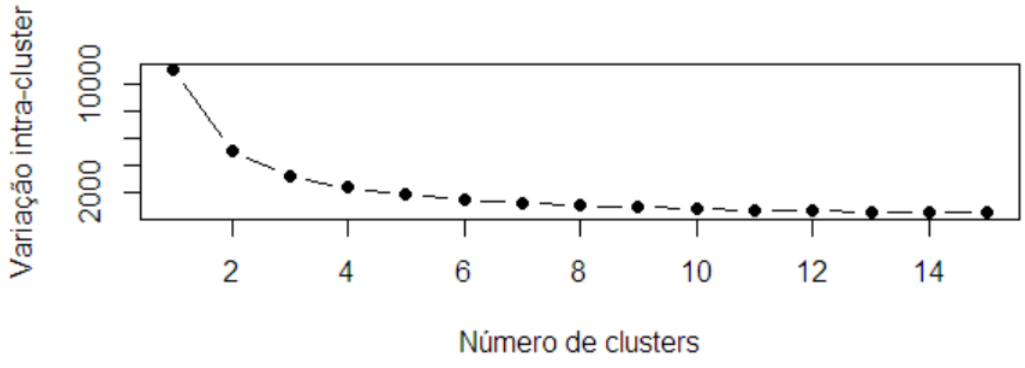
4

RESULTADOS

Para que o objetivo do estudo fosse concluído de maneira a priorizar atendimentos de alunos, fizemos uma análise de agrupamento chamada *k-modes* devido a característica categórica dos dados. A análise de agrupamento foi repetida 500 vezes de maneira a obtermos uma probabilidade empírica do respondente pertencer ao grupo menos frequente. A probabilidade era dada pelo número de vezes que o estudante ficou no *cluster* menos frequente (n_r) dividido pelo número de repetições ($B = 500$), ou seja, $P_r = \frac{n_r}{500}$

Para aplicar o método de *k-modes*, utilizamos a função *kmodes* do pacote *klaR* (WEIHS *et al.*, 2005) do software R (R Core Team, 2017). Já para definir o número ótimo de *clusters*, utilizamos o gráfico de silhueta para visualizar a variabilidade intra-*cluster* para cada um dos números de *clusters* entre 1 e 15. Ao utilizar esse método, foi escolhido número de *clusters* igual a 2, pois diminuir ainda mais a variabilidade intra-*cluster* não compensaria o aumento de complexidade da análise (ver Figura 4).

Figura 4 – Gráfico de silhueta para determinar o número ótimo de *clusters*



Fonte: Elaborada pela autora.

Para garantir que não houvesse mudança da nomenclatura entre os *clusters*, chamamos de *cluster* 1 o mais frequente e o *cluster* 2 foi o que possuia o menor número de observações. Ao final das 500 repetições, a média da proporção de observações que foram classificadas como pertencentes ao *cluster* 2 foi de 0.4. Portanto, para fazer a análise de variáveis que mais discriminaram a separação dos *clusters*, aplicamos a seguinte regra: caso o estudante tivesse probabilidade maior do que 0.4 era classificado como pertencente ao *cluster* menos frequente entitulado como *cluster* 2. Finalmente, a distribuição dos *clusters* pode ser visualizada na Tabela 25. Apenas ao final da comparação das variáveis nos *clusters* criados é que saberemos se algum dos *clusters* tem mais características de vulnerabilidade.

Tabela 25 – Distribuição dos *clusters*

Cluster	**Quantidade (%)**
1	276 (47,75%)
2	302 (52,25%)

Para identificar se as diferenças de proporções entre os cluster são estatisticamente significantes utilizamos a função *prop.test* do software R (R Core Team, 2017). A função foi configurada para testar a hipótese:

$$\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 \neq p_2 \end{cases}$$

sendo p_1 a proporção avaliada para o *cluster* 1 e p_2 a proporção avaliada para o *cluster* 2. Ou seja, quando p-valor $< 0,05$ encontramos evidências suficientes para rejeitar a hipótese de igualdade das proporções.

4.1 Variáveis de sentimento

A variável bem-estar do respondente na quarentena apresentou diferença de 14 pontos percentuais (p-valor $< 0,05$) quando comparada entre os *clusters* (ver Tabela 26). O *cluster* 2 é menos provável de respondido 4 ou 5, sendo que 1 representa o pior sentimento e 5 representa o melhor sentimento.

Tabela 26 – Distribuição da variável “Bem-estar na quarentena"pelos *clusters*

Cluster	**Bem-estar na quarentena**		**p-valor**
	1, 2 ou 3	**4 ou 5**	
1	37,50%	62,50%	-
2	51,50%	48,50%	-
Total	**43,94%**	**56,06%**	**0,001**

A variável "Senti felicidade"apresentou diferença de 11 pontos percentuais (p-valor $< 0,05$) entre os *clusters*, sendo que o *cluster* 2 é menos provável de respondido que sentiu felicidade durante a quarentena (ver Tabela 27).

Tabela 27 – Distribuição da variável “Senti felicidade"pelos *clusters*

Cluster	**Felicidade**		**p-valor**
	Não	**Sim**	
1	25,32%	74,68%	-
2	36,84%	63,16%	-
Total	**30,62%**	**69,38%**	**0,004**

É possível observar uma diferença de 22 pontos percentuais (p-valor $< 0,05$) entre a proporção de presença de alegria por brincar entre os *clusters* 1 e 2 (ver Tabela 28), sendo que os respondentes do *cluster* 2 possuem menor chance de terem vivenciado esse sentimento em sua quarentena.

Tabela 28 – Distribuição da variável “Senti alegria por brincar"pelos *clusters*

Cluster	**Alegria por brincar**		**p-valor**
	Não	**Sim**	
1	38,14%	61,86%	-
2	60,90%	39,10%	-
Total	**48,62%**	**51,38%**	$\approx$ **0**

Observando a variável de sentimento de preocupação, podemos observar uma diferença de 18 pontos percentuais (p-valor $< 0,05$) entre os grupos (ver Tabela 29), sendo que os respondentes do *cluster* 2 possuem maior chance de terem vivenciado esse sentimento em sua quarentena.

Tabela 29 – Distribuição da variável “Senti preocupação"pelos *clusters*

Cluster	**Preocupação**		**p-valor**
	Não	**Sim**	
1	62,18%	37,82%	-
2	43,98%	56,02%	-
Total	**53,81%**	**46,19%**	$\approx$ **0**

A variável de presença de medo apresentou uma diferença de 10 pontos percentuais (p-valor $< 0,05$) entre os grupos (ver Tabela 31), sendo que novamente o *cluster* 2 é mais provável de ter vivenciado esse sentimento em sua quarentena.

Tabela 30 – Distribuição da variável “Senti medo"pelos *clusters*

Cluster	**Medo**		**p-valor**
	Não	**Sim**	
1	75,32%	24,68%	-
2	65,04%	34,96%	-
Total	**70,59%**	**29,41%**	**0.009**

A variável de presença de ansiedade apresentou uma diferença de quase 25 pontos percentuais (p-valor $< 0,05$) entre os grupos (ver Tabela 31), sendo que novamente o grupo 2 é mais provável de ter presenciado esse sentimento em sua quarentena.

Tabela 31 – Distribuição da variável “Senti ansiedade"pelos *clusters*

Cluster	**Ansiedade**		**p-valor**
	Não	**Sim**	
1	68,27%	31,73%	-
2	43,61%	56,39%	-
Total	**56,92%**	**43,08%**	$\approx$ **0**

4.2 Variáveis de atividades realizadas durante a quarentena

Outro conjunto de variáveis importantes para explicar a segregação dos *clusters* foram as atividades realizadas durante a quarentena. Sobre a atividade de leitura, podemos observar diferença de 15 pontos percentuais (p-valor $< 0,05$) entre os grupos, sendo que respondentes pertencentes ao *cluster* 2 foram mais prováveis a não realizar tal atividade (ver Tabela 32).

Tabela 32 – Distribuição da variável “atividades: Ler"pelos *clusters*

Cluster	**Leio**		**p-valor**
	Não	**Sim**	
1	26,28%	73,72%	-
2	41,35%	58,65%	-
Total	**33,22%**	**66,78%**	$\approx$ **0**

Em relação a atividade brincar, a diferença foi de 15 pontos percentuais (p-valor $< 0,05$) entre os *clusters*, sendo que o *cluster* 2 também era menos provável de realizar a atividade (ver Tabela 33).

Tabela 33 – Distribuição da variável “atividades: brincar"pelos *clusters*

Cluster	**Brincar**		**p-valor**
	Não	**Sim**	
1	19,87%	80,13%	-
2	34,96%	65,04%	-
Total	**26,82%**	**73,18%**	$\approx$ **0**

Em relação a atividade de navegar pela internet, a diferença foi de quase 20 pontos percentuais (p-valor $< 0,05$) entre os *clusters*, sendo que o *cluster* 2 também era menos provável de realizar a atividade (ver Tabela 34).

Tabela 34 – Distribuição da variável “atividades: ficar na internet"pelos *clusters*

Cluster	**Ficar na internet**		**p-valor**
	Não	**Sim**	
1	17,95%	82,05%	-
2	37,59%	62,41%	-
Total	**26,99%**	**73,01%**	$\approx$ **0**

Em relação a atividade realizar chamada de vídeos com familiares e amigos, a diferença foi de 17 pontos percentuais (p-valor $< 0,05$) entre os *clusters*, sendo que o *cluster* 2 também era menos provável de realizar tal atividade (ver Figura 35).

Tabela 35 – Distribuição da variável “atividades: fazer chamada de vídeo"pelos *clusters*

Cluster	**Realizar chamada de vídeo**		**p-valor**
	Não	**Sim**	
1	20,51%	79,49%	-
2	37,97%	62,03%	-
Total	**28,55%**	**71,45%**	$\approx$ **0**

Em relação a atividade assistir filmes e séries via algum serviço de *streaming* a diferença entre os grupos foi de quase 26 pontos percentuais (p-valor $< 0,05$), sendo que o *cluster* 2 também era menos provável de realizar tal atividade (ver Tabela 36).

Tabela 36 – Distribuição da variável “atividades: assistir filmes/séries via *streaming*"pelos *clusters*

Cluster	**Ver streaming**		**p-valor**
	Não	**Sim**	
1	15,38%	84,62%	-
2	41,35%	58,65%	-
Total	**27,34%**	**72,66%**	$\approx$ **0**

Em relação a atividade ouvir música, a diferença foi de quase 14 pontos percentuais (p-valor $< 0,05$) entre os *clusters*, sendo que o *cluster* 2 também era menos provável de realizar tal atividade (ver Tabela 37).

Tabela 37 – Distribuição da variável “atividades: ouvir música"pelos *clusters*

Cluster	**Ouvir música**		**p-valor**
	Não	**Sim**	
1	21,47%	78,53%	-
2	35,34%	64,66%	-
Total	**27,85%**	**72,15%**	$\approx$ **0**

Em relação a atividade ajudar em casa, a diferença foi de 16 pontos percentuais (p-valor $< 0,05$), sendo que o *cluster* 2 era menos provável de realizar tal atividade (ver Tabela 38).

Tabela 38 – Distribuição da variável “atividades: ajudar em casa"pelos *clusters*

Cluster	Ajudar em casa		p-valor
	Não	Sim	
1	19,23%	80,77%	-
2	35,34%	64,66%	-
Total	**26,64%**	**73,36%**	$\approx$ **0**

4.3 Variáveis de acessibilidade tecnológica

Em relação a acessibilidade tecnológica que possibilitasse que os estudantes assistissem as aulas, a variável que mais discriminou os *clusters* foi a de possuir computador em casa. A diferença foi de quase 29 pontos percentuais (p-valor $< 0,05$), sendo que o *cluster* 2 era menos provável de possuir um computador (ver Tabela 39).

Tabela 39 – Distribuição da variável “Tenho computador"pelos *clusters*

Cluster	Tenho computador		p-valor
	Não	Sim	
1	28,53%	71,47%	-
2	57,52%	42,48%	-
Total	**41,87%**	**58,13%**	$\approx$ **0**

Em relação a qualidade da conexão de internet, o aluno avaliou subjetivamente a sua própria internet. O *cluster* 2 era menos provável de ter uma boa conexãode internet e a diferença entre os grupos foi de quase 14 pontos percentuais (p-valor $< 0,05$, ver tabela 41). Comparando com os resultados da Tabela 39, enfatizamos que os estudantes não necessariamente utilizam a internet apenas no computador, sendo possível a utilizar em celulares ou tablets. Inclusive, a própria internet a ser avaliada pode ser internet 3G, 4G ou banda larga.

Tabela 40 – Distribuição da variável “Tenho boa conexão de internet"pelos *clusters*

Cluster	Boa conexão de internet		p-valor
	Não	Sim	
1	10,90%	89,10%	-
2	25,94%	74,06%	-
Total	**17,82%**	**82,18%**	$\approx$ **0**

O *cluster* 2 era mais provável de necessitar de ajuda para mexer nos equipamentos eletrônicos e a diferença entre os grupos foi de quase 15 pontos percentuais (p-valor $< 0,05$, ver tabela 41).

4.4 Variáveis de vida familiar

Outro grupo de variáveis importantes para a segregação dos *clusters* foram variáveis de fonte de renda e diminuição de renda durante a quarentena e outras variáveis da vida familiar.

Tabela 41 – Distribuição da variável “Preciso de ajuda para mexer nos equipamentos eletrônicos"pelos *clusters*

Cluster	**Preciso de ajuda para mexer nos equipamentos eletrônicos**		**p-valor**
	Não	**Sim**	
1	61,22%	38,78%	-
2	46,62%	53,38%	-
Total	**54,50%**	**45,50%**	**0.001**

Os respondentes do *cluster* 2 eram menos prováveis de ter a figura paterna morando na mesma casa que os respondentes, sendo a diferença de 26 pontos percentuais entre os grupos (p-valor $< 0,05$, ver Tabela 42).

Tabela 42 – Distribuição da variável “morar com o pai"pelos *clusters*

Cluster	**Morar com o pai**		**p-valor**
	Não	**Sim**	
1	31,73%	68,27%	-
2	58,27%	41,73%	-
Total	**43,94%**	**56,06%**	$\approx$ **0**

Os respondentes do *cluster* 2 eram mais prováveis de ter como principal fonte de renda o auxílio emergencial, sendo a diferença de quase 16 pontos percentuais entre os grupos (p-valor $< 0,05$, ver Tabela 43).

Tabela 43 – Distribuição da variável “Fonte de renda: auxílio emergencial"pelos *clusters*

Cluster	**Auxílio emergencial**		**p-valor**
	Não	**Sim**	
1	88,78%	11,22%	-
2	72,93%	27,07%	-
Total	**81,49%**	**18,51%**	$\approx$ **0**

Os respondentes do *cluster* 2 também eram mais prováveis de ter como principal fonte de renda o trabalho informal, sendo a diferença de quase 27 pontos percentuais entre os grupos (p-valor $< 0,05$, ver Tabela 44).

Tabela 44 – Distribuição da variável Fonte de renda: trabalho informal pelos *clusters*

Cluster	**Trabalho informal**		**p-valor**
	Não	**Sim**	
1	72,76%	27,24%	-
2	45,86%	54,14%	-
Total	**60,38%**	**39,62%**	$\approx$ **0**

Os respondentes do *cluster* 1 eram mais prováveis de ter como principal fonte de renda o trabalho formal, sendo a diferença de 48 pontos percentuais entre os grupos (p-valor $< 0,05$, ver Tabela 44), sendo essa variável a que mais discriminou os *clusters* entre todas as outras.

Tabela 45 – Distribuição da variável “Fonte de renda: trabalho formal"pelos *clusters*

Cluster	**Trabalho formal**		**p-valor**
	Não	**Sim**	
1	24,36%	75,64%	
2	72,93%	27,07%	
Total	**46,71%**	**53,29%**	$\approx$ **0**

Os respondentes do *cluster* 1 eram mais prováveis de terem famílias cuja renda não diminuiu na quarentena ou diminuiu um pouco, enquanto respondentes do *cluster* 2 eram mais prováveis de terem suas rendas muito ou totalmente diminuídas (p-valor $< 0,05$, ver Tabela 46).

Tabela 46 – Distribuição da variável Diminuição da renda na quarentena pelos *clusters*

Cluster	**Renda diminuída durante a quarentena**				**p-valor**
	Não	**Sim, um pouco**	**Sim, muito**	**Sim, totalmente**	
1	32,37%	47,76%	14,42%	5,45%	-
2	19,55%	27,82%	40,60%	12,03%	-
Total	**26,47%**	**38,58%**	**26,47%**	**8,48%**	**0.001**

De acordo com as variáveis analisadas, o *cluster* 2 foi identificado como vulvenável tanto socioeconomicamente quando psicologicamente, pois foi mais provável de ter relatado ansiedade (diferença de 25 pontos percentuais entre os *clusters*), medo (diferença de 10 p.p.) e preocupação (diferença de 18 p.p.), e ainda foi menos provável de ter relatado felicidade (diferença de 11 p.p.), alegria (diferença de 22 p.p.) e bem-estar geral (diferença de 14 p.p.). O *cluster* 2 foi menos provável de assistir filmes e séries via *streaming* (diferença de 26 p.p.), ler (diferença de 15 p.p.), brincar (diferença de 20 p.p.), ficar na internet (diferença de 20 p.p.), fazer chamada de vídeo com familiares e amigos (diferença de 17 p.p.), ouvir música (diferença de 14 p.p.) e ajudar em casa (diferença de 15 p.p.). Foi, ainda, menos provável de possuir computador em casa (diferença de 29 p.p.), ter uma conexão de internet boa (diferença de 14 p.p.) e foi mais provável de precisar de ajuda para mexer em equipamentos eletrônicos e internet (diferença de 15 p.p.). Também, foi menos provável a ter trabalho formal como fonte de renda familiar (diferença de 48 p.p.), menos provável a ter o pai morando em casa na quarentena (diferença de 26 p.p.) e mais provável de ter o auxílio emergencial como fonte de renda familiar (diferença de 16 p.p.). Todas as variáveis apresentadas nessa Seção mostraram ter diferenças estatisticamente significantes entre *clusters*.

CAPÍTULO

5

CONCLUSÕES FINAIS

Com o objetivo de identificar vulnerabilidade em estudantes do ensino fundamental de uma escola pública na pandemia de Covid-19 utilizamos um método de agrupamento adaptado para dados categóricos chamado de *k-modes*. O objetivo foi concluído com sucesso, pois encontramos um grupo que possui características marcantes de vulnerabilidade psicológica, socioeconomica e mesmo tecnológica.

Em relação a variáveis que poderiam identificar características de vulnerabilidade psicológica, os estudantes que pertencem ao grupo mais vulnerável, eram mais prováveis de relatar sentimentos como medo, preocupação e ansiedade. Em contrapartida, eram menos prováveis de relatar sentimentos positivos commo bem-estar, felicidade e alegria por brincar.

Em relação a atividades realizadas durante a pandemia, o grupo mais vulnerável apresentou menor chance de ler, brincar, ficar na internet, fazer chamadas de vídeo com amigos e familiares, ver tv, ver streaming e até mesmo, ouvir música. Essas características podem representar vulnerabilidades sociais, devido a falta de equipamento e internet para fazer determinadas atividades.

Em relação a características de vulnerabilidade tecnológica, o grupo mais vulnerável se apresentou menos provavél de ter um computador em casa e menos prováveis a avaliar a qualidade da conexão de internet como boa. Os estudantes que não possuíam nenhum equipamento para assistir as aulas e/ou internet ou mesmo precisavam dividir equipamentos com outros familiares foram imediatamente ajudados pela escola e empresas voluntárias. Todavia, infelizmente, não houveram recursos suficientes para oferecer uma internet de maior qualidade para os estudantes que possuiam uma conexão ruim. O grupo mais vulnerável era, ainda, mais provável de precisar da ajuda de um adulto para mexer nos equipamentos eletrônicos.

Em relação a vida familiar, os estudantes que pertencem ao grupo mais vulnerável eram menos prováveis de estar morando com o pai durante a pandemia, porém ambos os grupos tinham chances parecidas de estar morando com a mãe. O grupo vulnerável também relatou estar

dependendo do auxílio emergencial e/ou trabalhos informais como principal fonte de renda da família. Esse grupo, ainda, apresentou maior chance de ter sua renda diminuída, em algum grau, durante a pandemia.

Como são mais de 300 estudantes identificados com características marcantes de vulnerabilidade, repetimos o processo de agrupamento várias vezes de maneira a encontrar uma probabilidade empírica de vulnerabilidade. Dessa maneira, os atendimentos psicológicos e demais programas sociais e voluntários da escola puderam contar com uma lista de priorização. A escola estima que 76 alunos receberam algum tipo de auxílio tecnológico e 18 alunos receberam atendimento psicológico.

Finalmente, como análises futuras, seria importante testar outras metodologias de agrupamento como métodos hierárquicos, bem como diferentes tipos de *linkage* (ver Seção 3.2.1), para que fosse possível comparar a sinergia entre métodos. Como os métodos não supervisionados não possuem medidas de qualidade de ajuste, como taxa de erro, poderíamos considerar o voto da maioria para determinar quem são os estudantes vulneráveis.

REFERÊNCIAS

ALMEIDA, J.; BARBOSA, L.; PAIS, A.; FORMOSINHO, S. Improving hierarchical cluster analysis: A new method with outlier detection and automatic clustering. **Chemometrics and Intelligent Laboratory Systems**, Elsevier, v. 87, n. 2, p. 208–217, 2007. Citado na página 35.

ANTONENKO, P. D.; TOY, S.; NIEDERHAUSER, D. S. Using cluster analysis for data mining in educational technology research. **Educational Technology Research and Development**, Springer, v. 60, n. 3, p. 383–398, 2012. Citado na página 36.

ARRUDA, E. P. Educação remota emergencial: elementos para políticas públicas na educação brasileira em tempos de covid-19. **EmRede-Revista de Educação a Distância**, v. 7, n. 1, p. 257–275, 2020. Citado na página 23.

CODEÇO, C. T.; VILLELA, D.; COELHO, F. C.; BASTOS, L. S.; CARVALHO, L. M.; GOMES, M. F.; CRUZ, O. G.; LANA, R. M.; VESPIGNANI, A.; PIONTTI, A. Pastore y *et al.* **Estimativa de risco de espalhamento da COVID-19 no Brasil e avaliação da vulnerabilidade socioeconômica nas microrregiões brasileiras**. [S.l.], 2020. Citado na página 24.

COHRS, F. M.; SOUSA, F. S.; TENÓRIO, J. M.; RAMOS, L. R.; PISA, I. T. Aplicação de análise de cluster em dados integrados de um estudo prospectivo: projeto epidoso como cenário. **Journal of health informatics**, v. 5, n. 1, 2013. Citado na página 24.

CRAWFORD, J.; BUTLER-HENDERSON, K.; RUDOLPH, J.; MALKAWI, B.; GLOWATZ, M.; BURTON, R.; MAGNI, P.; LAM, S. Covid-19: 20 countries' higher education intra-period digital pedagogy responses. **Journal of Applied Learning & Teaching**, Kaplan Singapore, v. 3, n. 1, p. 1–20, 2020. Citado na página 23.

DIAS, É.; PINTO, F. C. F. A educação e a covid-19. **Ensaio: Avaliação e Políticas Públicas em Educação**, SciELO Brasil, v. 28, n. 108, p. 545–554, 2020. Citado na página 23.

FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R. **The elements of statistical learning**. [S.l.]: Springer series in statistics New York, 2001. v. 1. Citado nas páginas 35, 36, 37, 38 e 40.

HUANG, H.-C.; CHUANG, Y.-Y.; CHEN, C.-S. Multiple kernel fuzzy clustering. **IEEE Transactions on Fuzzy Systems**, IEEE, v. 20, n. 1, p. 120–134, 2011. Citado na página 35.

HUANG, Z. A fast clustering algorithm to cluster very large categorical data sets in data mining. **DMKD**, Citeseer, v. 3, n. 8, p. 34–39, 1997. Citado nas páginas 40 e 41.

HUANG, Z.; NG, M. K. A fuzzy k-modes algorithm for clustering categorical data. **IEEE transactions on Fuzzy Systems**, IEEE, v. 7, n. 4, p. 446–452, 1999. Citado na página 35.

HUBERTY, C. J.; JORDAN, E. M.; BRANDT, W. C. Cluster analysis in higher education research. In: **Higher education: Handbook of theory and research**. [S.l.]: Springer, 2005. p. 437–457. Citado na página 36.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An introduction to statistical learning**. [S.l.]: Springer, 2013. v. 112. Citado nas páginas 35, 36, 37, 38, 39 e 40.

JOYE, C. R.; MOREIRA, M. M.; ROCHA, S. S. D. Educação a distância ou atividade educacional remota emergencial: em busca do elo perdido da educação escolar em tempos de covid-19. **Research, Society and Development**, v. 9, n. 7, p. e521974299–e521974299, 2020. Citado na página 23.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2017. Disponível em: <https://www.R-project.org/>. Citado nas páginas 43 e 44.

SCHMIDT, B.; CREPALDI, M. A.; BOLZE, S. D. A.; NEIVA-SILVA, L.; DEMENECH, L. M. Saúde mental e intervenções psicológicas diante da pandemia do novo coronavírus (covid-19). **Estudos de Psicologia (Campinas)**, SciELO Brasil, v. 37, 2020. Citado na página 23.

TSAI, D.-M.; LIN, C.-C. Fuzzy c-means based clustering for linearly and nonlinearly separable data. **Pattern recognition**, Elsevier, v. 44, n. 8, p. 1750–1760, 2011. Citado nas páginas 35 e 36.

ULLRICH-FRENCH, S.; COX, A. Using cluster analysis to examine the combinations of motivation regulations of physical education students. **Journal of Sport and Exercise Psychology**, Human Kinetics, Inc., v. 31, n. 3, p. 358–379, 2009. Citado na página 36.

VASCONCELOS, C.; LOPES, B.; COSTA, N.; MARQUES, L.; CARRASQUINHO, S. Estado da arte na resolução de problemas em educação em ciência. **Revista electrónica de enseñanza de las ciencias**, v. 6, n. 2, p. 235–245, 2007. Citado na página 36.

WEIHS, C.; LIGGES, U.; LUEBKE, K.; RAABE, N. klar analyzing german business cycles. In: BAIER, D.; DECKER, R.; SCHMIDT-THIEME, L. (Ed.). **Data Analysis and Decision Support**. Berlin: Springer-Verlag, 2005. p. 335–343. Citado na página 43.

APÊNDICE

A

CÓDIGOS-FONTE

Código-fonte 1 – Código-fonte da aplicação do algoritmo k-modes nos dados do formulário aplicado pela escola

```
dados = read.table("mapeamento.csv",head=T,sep=";")
#install.packages(c("FactoMineR", "factoextra"))
library("FactoMineR")
library("factoextra")

library("data.table")
#' Load useful packages
library(cluster)
library(dplyr)
library(ggplot2)
library(readr)
library(klaR)

head(dados)

#Tratamento da variável ano

dados$ano_escolar_trat = ifelse(dados$Ano.escolar %like% "1",1,
  ifelse(dados$Ano.escolar %like% "primeiro",1,
  ifelse(dados$Ano.escolar %like% "Primeiro",1,
  ifelse(dados$Ano.escolar %like% "2",2,
  ifelse(dados$Ano.escolar %like% "segundo",2,
  ifelse(dados$Ano.escolar %like% "Segundo",2,
  ifelse(dados$Ano.escolar %like% "3",3,
  ifelse(dados$Ano.escolar %like% "terceiro",3,
```

```
26:   ifelse(dados$Ano.escolar %like% "Terceiro",3,
27:   ifelse(dados$Ano.escolar %like% "4",4,
28:   ifelse(dados$Ano.escolar %like% "quarto",4,
29:   ifelse(dados$Ano.escolar %like% "Quarto",4,
30:   ifelse(dados$Ano.escolar %like% "5",5,
31:   ifelse(dados$Ano.escolar %like% "quinto",5,
32:   ifelse(dados$Ano.escolar %like% "Quinto",5,
33:   ifelse(dados$Ano.escolar %like% "6",6,
34:   ifelse(dados$Ano.escolar %like% "sexto",6,
35:   ifelse(dados$Ano.escolar %like% "Sexto",6,
36:   ifelse(dados$Ano.escolar %like% "7",7,
37:   ifelse(dados$Ano.escolar %like% "setimo",7,
38:   ifelse(dados$Ano.escolar %like% "Setimo",7,
39:   ifelse(dados$Ano.escolar %like% "8",8,
40:   ifelse(dados$Ano.escolar %like% "oitavo",8,
41:   ifelse(dados$Ano.escolar %like% "Oitavo",8,
42:   ifelse(dados$Ano.escolar %like% "9",9,
43:   ifelse(dados$Ano.escolar %like% "nono",9,
44:   ifelse(dados$Ano.escolar %like% "Nono",9,
45:   ifelse(dados$Ano.escolar %like% "Sétimo",7,
46:   ifelse(dados$Ano.escolar %like% "Primeira",1,
47:   ifelse(dados$Ano.escolar %like% "OItavo",8,
48:   ifelse(dados$Primeiro.Nome == "DAVI" & dados$TUTOR.MAIUSCULO
       =="RAFAEL",3,
49:   ifelse(dados$Primeiro.Nome == "MARIA" & dados$TUTOR.MAIUSCULO
       =="ELIETE",9,
50:   ifelse(dados$Primeiro.Nome == "DAPHNE" & dados$TUTOR.
       MAIUSCULO=="LÚCIO",2,
51:   ifelse(dados$Primeiro.Nome == "NICOLAS" & dados$TUTOR.
       MAIUSCULO=="CARINA",7,
52:   ifelse(dados$Primeiro.Nome == "LAYSLA" & dados$TUTOR.
       MAIUSCULO=="PAULO",5,NA))))))))))))))))))))))))))))))))))))))
53:
54:
55:
56: table(dados$ano_escolar_trat,useNA="always")
57:
58: dados$flag_seguranca = ifelse(dados$X2..MARQUE.O.QUE.VOCÊ.
       SENTIU.ATÉ.ESTE.MOMENTO.NA.QUARENTENA. %like% "SENTI SEGURAN
       ÇA EM CASA",1,0)
59: table(dados$flag_seguranca,useNA="always")
60:
```

```
61:
62: #Elbow Method for finding the optimal number of clusters
63: set.seed(123)
64: # Compute and plot wss for k = 2 to k = 15.
65: #The within-cluster simple-matching distance for each cluster.
66: k.max <- 15
67: wss <- sapply(1:k.max,function(k){mean(kmodes(db[,-c(1)], k)$
      withindiff)})
68:
69:
70:
71: plot(1:k.max, wss,type="b", pch = 19,
72:      xlab= "Número de clusters",ylab="Variação intra-cluster")
73:
74: B = 500
75:
76: matriz_cluster = matrix(rep(NA,B*nrow(db)),nrow(db),B)
77:
78: #db <- db %>% mutate_if(is.character,as.factor)
79: #db <- db %>% mutate_if(is.numeric,as.factor)
80: #res.mca = MCA(db[,-1],2)
81: #var <- get_mca_var(res.mca)
82: #ncol(var$coord)
83:
84: fit <- princomp(na.omit(db[,-1]))
85: summary(fit) # print variance accounted for
86: loadings(fit) # pc loadings
87: plot(fit,type="lines") # scree plot
88: fit$scores # the principal components
89: biplot(fit)
90:
91: for (i in 1:B){
92:   #boot = sample(1:nrow(db),nrow(db),replace = F)
93:   cluster.results <-kmodes(db[,-1], 2 ) #don't use the record
      ID as a clustering variable!
94:   if (table(cluster.results$cluster)[1]<table(cluster.results$
      cluster)[2])
95:     { cluster.results$cluster = ifelse(cluster.results$cluster
      == 1,2,1)
96:     }
97:   matriz_cluster[,i]=cluster.results$cluster
98: }
```

```
99:
100: for (j in 1:nrow(db)){
101:   db$cluster_prob[j] = table(matriz_cluster[j,])[2]/(table(
       matriz_cluster[j,])[1]+table(matriz_cluster[j,])[2])
102: }
103:
104:
105: corte = table(matriz_cluster)[2]/(table(matriz_cluster)[1]+
       table(matriz_cluster)[2])
106:
107: db$cluster = ifelse(db$cluster_prob>corte,2,1)
108:
109:
110: write.csv2(db, file = "resultados_v2.csv")
111:
112: #######TESTE DE HIPOTESE PARA DIFERENÇA DE PROPORCAO
113:
114: #bem-estar
115: p1 = .6250
116: p2= .4850
117: n1=276
118: n2=302
119:
120:
121: round(prop.test(x=c(p1*n1,p2*n2),n=c(n1,n2),alternative = "two.
       sided")[3]$p.value,3)
122:
123: #felicidade
124: p1 = .7468
125: p2= .6316
126: n1=276
127: n2=302
128:
129: round(prop.test(x=c(p1*n1,p2*n2),n=c(n1,n2),alternative = "two.
       sided")[3]$p.value,3)
130:
131:
132: #alegria
133: p1 = .6186
134: p2= .3910
135: n1=276
136: n2=302
```

```
137:
138: round(prop.test(x=c(p1*n1,p2*n2),n=c(n1,n2),alternative = "two.
       sided")[3]$p.value,3)
139:
140: #preocupacao
141:
142: p1 = .3782
143: p2= .5602
144: n1=276
145: n2=302
146:
147: round(prop.test(x=c(p1*n1,p2*n2),n=c(n1,n2),alternative = "two.
       sided")[3]$p.value,3)
148:
149: #medo
150:
151: p1 = .2468
152: p2= .3496
153: n1=276
154: n2=302
155:
156: round(prop.test(x=c(p1*n1,p2*n2),n=c(n1,n2),alternative = "two.
       sided")[3]$p.value,3)
157:
158: #ansiedade
159: p1 = .3173
160: p2= .5639
161: n1=276
162: n2=302
163:
164: round(prop.test(x=c(p1*n1,p2*n2),n=c(n1,n2),alternative = "two.
       sided")[3]$p.value,3)
165:
166: #ansiedade
167: p1 = .3173
168: p2= .5639
169: n1=276
170: n2=302
171:
172: round(prop.test(x=c(p1*n1,p2*n2),n=c(n1,n2),alternative = "two.
       sided")[3]$p.value,3)
173:
```

```
174: #leitura
175: p1 = .7372
176: p2= .5865
177: n1=276
178: n2=302
179:
180: round(prop.test(x=c(p1*n1,p2*n2),n=c(n1,n2),alternative = "two.
       sided")[3]$p.value,3)
181:
182:
183: #brincar
184: p1 = .8013
185: p2= .6504
186: n1=276
187: n2=302
188:
189: round(prop.test(x=c(p1*n1,p2*n2),n=c(n1,n2),alternative = "two.
       sided")[3]$p.value,3)
190:
191:
192:
193: #ficar na internet
194: p1 = .8205
195: p2= .6241
196: n1=276
197: n2=302
198:
199: round(prop.test(x=c(p1*n1,p2*n2),n=c(n1,n2),alternative = "two.
       sided")[3]$p.value,3)
200:
201:
202: #fazer chamada de video
203: p1 = .7949
204: p2= .6203
205: n1=276
206: n2=302
207:
208: round(prop.test(x=c(p1*n1,p2*n2),n=c(n1,n2),alternative = "two.
       sided")[3]$p.value,3)
209:
210:
211: #streaming
```

```
212: p1 = .8462
213: p2= .5865
214: n1=276
215: n2=302
216:
217: round(prop.test(x=c(p1*n1,p2*n2),n=c(n1,n2),alternative = "two.
       sided")[3]$p.value,3)
218:
219: #ouvir musica
220: p1=.7853
221: p2=0.6466
222: n1=276
223: n2=302
224:
225: round(prop.test(x=c(p1*n1,p2*n2),n=c(n1,n2),alternative = "two.
       sided")[3]$p.value,3)
226:
227: #ajudar em casa
228: p1=.8077
229: p2=0.6466
230: n1=276
231: n2=302
232:
233: round(prop.test(x=c(p1*n1,p2*n2),n=c(n1,n2),alternative = "two.
       sided")[3]$p.value,3)
234:
235: #tenho computador
236: p1=.7147
237: p2=0.4248
238: n1=276
239: n2=302
240:
241: round(prop.test(x=c(p1*n1,p2*n2),n=c(n1,n2),alternative = "two.
       sided")[3]$p.value,3)
242:
243:
244: #preciso de ajuda para mexer em equipamentos
245: p1=.3878
246: p2=.5338
247: n1=276
248: n2=302
249:
```

```
250: round(prop.test(x=c(p1*n1,p2*n2),n=c(n1,n2),alternative = "two.
       sided")[3]$p.value,3)
251:
252:
253: #morar com o pai
254: p1=.6827
255: p2=.4173
256: n1=276
257: n2=302
258:
259: round(prop.test(x=c(p1*n1,p2*n2),n=c(n1,n2),alternative = "two.
       sided")[3]$p.value,3)
260:
261:
262: #auxílio emergencial
263: p1=.1122
264: p2=.2707
265: n1=276
266: n2=302
267:
268: round(prop.test(x=c(p1*n1,p2*n2),n=c(n1,n2),alternative = "two.
       sided")[3]$p.value,3)
269:
270: #trabalho informal
271:
272: p1=.2724
273: p2=.5414
274: n1=276
275: n2=302
276:
277: round(prop.test(x=c(p1*n1,p2*n2),n=c(n1,n2),alternative = "two.
       sided")[3]$p.value,3)
278:
279: #trabalho formal
280:
281: p1=.7564
282: p2=.2707
283: n1=276
284: n2=302
285:
286: round(prop.test(x=c(p1*n1,p2*n2),n=c(n1,n2),alternative = "two.
       sided")[3]$p.value,3)
```

```
287:
288: #renda diminuida
289:
290: p1=.3237
291: p2=.1955
292: n1=276
293: n2=302
294:
295: round(prop.test(x=c(p1*n1,p2*n2),n=c(n1,n2),alternative = "two.
       sided")[3]$p.value,3)
```


ICMC
USP
SÃO CARLOS