

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
COORDENAÇÃO DO CURSO DE LICENCIATURA EM
MATEMÁTICA

JULIANO RODRIGUES DA SILVA

ANÁLISE MULTIVARIADA PARA DADOS QUANTITATIVOS
DISCRETOS: ESTUDO DE CASO DO DESEMPENHO DOS ALUNOS
NA PROVA PARANÁ

TRABALHO DE CONCLUSÃO DE CURSO

TOLEDO

2019

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
COORDENAÇÃO DO CURSO DE LICENCIATURA EM
MATEMÁTICA**

JULIANO RODRIGUES DA SILVA

**ANÁLISE MULTIVARIADA PARA DADOS
QUANTITATIVOS DISCRETOS: ESTUDO DE CASO DO
DESEMPENHO DOS ALUNOS NA PROVA PARANÁ**

Trabalho de Conclusão de Curso apresentado ao Curso de Licenciatura em Matemática da Universidade Tecnológica Federal do Paraná, Câmpus Toledo, como requisito parcial à obtenção do título de Licenciado em Matemática.

Orientadora: Prof. Dra. Regiane Slongo Fagundes

TOLEDO

2019

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
COORDENAÇÃO DO CURSO DE LICENCIATURA EM
MATEMÁTICA

TERMO DE APROVAÇÃO

O Trabalho de Conclusão de Curso intitulado **Análise multivariada para dados quantitativos discretos: estudo de caso do desempenho dos alunos na Prova Paraná** foi considerado **APROVADO** de acordo com a ata

n^o __ de __/__/----

Fizeram parte da banca examinadora os Professores:

Dra. Regiane Slongo Fagundes (Orientadora)

Ms. Naísa Camila Garcia Tosti

Dr. Gustavo Henrique Dalposso

TOLEDO

2019

Dedico este trabalho primeiramente a Deus por sempre me mostrar o caminho que devia seguir. A minha irmã, por sempre estar me apoiando e me encorajando nos momentos difíceis. Á minha família e amigos por compreenderem os momentos em que não pude estar presente.

AGRADECIMENTOS

Agradeço primeiramente a Deus, por me dar força e sabedoria nos momentos de desânimo.

Aos meus pais, Aparecida e José (*in memoriam*), que não pôde estar ao meu lado neste momento tão importante, mas que sempre torceu muito e nunca perdeu a fé nos meus sonhos. Saudade eterna.

Dedico esse trabalho aos meus professores da UTFPR. Essa conquista não seria possível se não fosse pela paciência e dedicação de cada um de vocês.

Aos meus colegas de trabalho, que sempre me incentivaram a continuar, que entenderam as inúmeras vezes que tive que me ausentar.

Dedico essa nova realização aos meus familiares, sobretudo à minha Irmã Edina, que ofereceu força, apoio e motivação em toda trajetória acadêmica.

A minha orientadora Regiane, por todo apoio e paciência ao longo da elaboração deste projeto.

Obrigada a minha banca, professoras Náisa, Regiane e professor Gustavo, por terem aceito o convite e por contribuírem com meu trabalho.

Por fim, mas não menos importante. Dedico aos meus amigos e ex professores, que sempre me incentivaram e me apoiaram nessa conquista.

Minha eterna gratidão a todos que de alguma ou de outra forma contribuíram.

Ensinar não é transferir conhecimento, mas criar possibilidades para a sua produção ou sua construção. Quem ensina aprende ao ensinar e quem aprende ensina ao aprender (Paulo Freire).

RESUMO

SILVA, Juliano Rodrigues. **Análise multivariada para dados quantitativos discretos: estudo de caso do desempenho dos alunos na Prova Paraná P.63.** Trabalho de Conclusão de Curso – Curso de Licenciatura em Matemática - Universidade Tecnológica Federal do Paraná. Toledo, 2019

O presente trabalho expõe uma investigação do desempenho dos alunos na primeira edição da Prova Paraná-2018, por meio da análise de uma escola do município de Toledo-PR, pertencente ao Núcleo Regional de Educação de Toledo. O objetivo foi aplicar método estatístico multivariados quantitativo discreto para estudar e comparar os desempenhos dos alunos. Os dados foram coletados junto a NRE-Toledo e escola. As séries avaliadas foram 6^o e 9^o anos do Ensino Fundamental e 1^o e 3^o anos do Ensino Médio. As informações coletadas foram digitalizadas e organizadas de acordo os descritores (conteúdos estruturantes) disponibilizados no site da SEED-Paraná. A partir da análise inicial, foi aplicado método estatístico multivariado de Análise de Agrupamento (conglomerado ou Análise de *clusters*), utilizando como método de agrupamento a distância Euclidiana e agrupamento hierárquico. Três métodos de ligação foram considerados: Método da ligação simples ou do vizinho mais próximo (*Single linkage*); Método da ligação completa ou do vizinho mais distante (*Complete linkage*); Método da média das distâncias (*Average Linkage*). O Coeficiente de Correlação Cofenético foi usado para avaliar a qualidade dos grupos formados. As análises foram implementadas utilizando o software gratuito e aberto, R Core Team (2019). O resultado geral obtido pelas turmas mostra que o índice de acerto dos alunos do Ensino Médio é menor que dos alunos do Ensino Fundamental. Das 20 questões avaliadas, em média, a porcentagem de respostas correta do Ensino Fundamental foi de 52,35%, enquanto o Ensino Médio apresentou apenas 32,30% . Isso indica que com passar do tempo os alunos mostraram ter maiores dificuldades com a matemática. Assim, a Prova Paraná precisa ser fortalecida, aprimorada e ampliada e se faz necessário que a SEED, o NRE e as escolas estabeleçam ações para a melhoria do ensino, visando sanar as dificuldades referentes as habilidades não desenvolvidas pelos alunos.

Palavras-chave: Prova Paraná-2018. Estatística multivariada. Análise de Agrupamento. Descritores de habilidades matemáticas.

ABSTRACT

SILVA, Juliano Rodrigues. **Multivariate analysis for discrete quantitative data: case study of student performance in Prova Paraná P.63.** Completion of course work – Licentiate degree in mathematics - Federal Technological University of Parana. Toledo, 2019.

The present work exposes an investigation of the performance of students in the first edition of the Paraná-2018 Test, through the analysis of a school in the municipality of Toledo-PR, belonging to the Regional Center for Education of Toledo. The objective was to apply discrete quantitative multivariate statistical method to study and compare student performance. The data were collected at the NRE-Toledo and school. The grades evaluated were 6th and 9th years of elementary school and 1st and 3rd years of high school. The information collected was digitalized and organized according to the descriptors, structuring contents, and the SEED-Paraná website was made available. From the initial analysis, a multivariate statistical method of Grouping Analysis (conglomerate or Cluster Analysis) was applied, using the Euclidean distance and hierarchical grouping as the grouping method. Three connection methods were considered: Simple connection method or the closest neighbor (*Single linkage*); Complete linkage method or the farthest neighbor (*Complete linkage*); Average Linkage method (*Average Linkage*). The Coefficient of Cofenetic Correlation was used to evaluate the quality of the groups formed. The analyses were implemented using free and open source software, Oksanen et al. (2016). The overall result obtained by the classes shows that the hit rate of high school students is lower than that of elementary school students. Of the 20 questions evaluated, on average, the percentage of correct answers in primary education was 52.35%, while secondary education presented only 32.30%. This indicates that with the passage of time students showed greater difficulties with mathematics. Thus, the Paraná Test needs to be strengthened, improved and expanded and it is necessary that SEED, the NRE and the schools establish actions for the improvement of teaching, in order to overcome the difficulties related to the skills not developed by the students.

Key-words: Prova Parana-2018. Multivariate statistics. Grouping Analysis. Descriptors of mathematical skills.

LISTA DE ILUSTRAÇÕES

2.1	Modelo do quadro de descritores da Avaliação: Prova Paraná	15
2.2	Esquema geral de procedimentos hierárquicos aglomerativos e divisivos. . .	19
4.1	Boxplot do número de acertos na Prova Paraná do 6 ^o ano do Ensino Fundamental.	26
4.2	Dendrograma formado a partir da Distância Euclidiana e Método da Média das Distâncias, com quatro <i>cluster</i> - sexto ano do Ensino Fundamental. . . .	27
4.3	Questão 24 retirada da Prova Paraná-2018, Ensino Fundamental - 6 ^o ano. .	29
4.4	Questão 36 retirada da Prova Paraná-2018, Ensino Fundamental - 6 ^o ano. .	29
4.5	Questão 33 retirada da Prova Paraná-2018, Ensino Fundamental - 6 ^o ano. .	30
4.6	Boxplot do número de acertos na Prova Paraná do 9 ^o ano do Ensino Fundamental.	32
4.7	Dendrograma formado a partir da Distância Euclidiana e Método da Média das Distâncias, com quatro <i>cluster</i> - nono ano do Ensino Fundamental. . . .	33
4.8	Questão 24 retirada da Prova Paraná-2018, Ensino Fundamental - 9 ^o ano. .	35
4.9	Questão 37 retirada da Prova Paraná-2018, Ensino Fundamental - 9 ^o ano. .	36
4.10	Questão 36 retirada da Prova Paraná-2018, Ensino Fundamental - 9 ^o ano. .	37
4.11	Boxplot do número de acertos na Prova Paraná do 1 ^o ano do Ensino Médio. .	39
4.12	Dendrograma formado a partir da Distância Euclidiana e Método da Média das Distâncias, com quatro <i>cluster</i> - primeira série do Ensino Médio.	40
4.13	Questão 23 retirada da Prova Paraná-2018, Ensino Médio - 1 ^o ano.	42
4.14	Questões 24, 22 e 35 retiradas da Prova Paraná-2018, Ensino Médio - 1 ^o ano. .	43
4.15	Questão retirada da Prova Paraná-2018	44
4.16	Boxplot dos resultados da 3 ^a série do Ensino Médio em relação a número de acertos na Prova Paraná.	46
4.17	Dendrograma formado a partir da Distância Euclidiana e Método da Média das Distâncias, com quatro <i>cluster</i> - terceira série do Ensino Médio.	47
4.18	Questão 26 retirada da Prova Paraná-2018, Ensino Médio - 3 ^o ano.	49
4.19	Questões 21 e 30 retiradas da Prova Paraná-2018, Ensino Médio - 3 ^o ano. .	50
B.1	Dendrogramas obtidos pelos métodos de agrupamento: Ligação simples (<i>Single Linkage</i>); Ligação Completa (<i>Complete Linkage</i>) e Média das Distâncias (<i>Average Linkage</i>) para o 6 ^o ano do Ensino Fundamental.	57

B.2	Dendrogramas obtidos pelos métodos de agrupamento: Ligação simples (<i>Single Linkage</i>); Ligação Completa (<i>Complete Linkage</i>) e Média das Distâncias (<i>Average Linkage</i> para o 9º ano do Ensino Fundamental.	57
B.3	Dendrogramas obtidos pelos métodos de agrupamento: Ligação simples (<i>Single Linkage</i>); Ligação Completa (<i>Complete Linkage</i>) e Média das Distâncias (<i>Average Linkage</i> para o 1º ano do Médio.	58
B.4	Dendrogramas obtidos pelos métodos de agrupamento: Ligação simples (<i>Single Linkage</i>); Ligação Completa (<i>Complete Linkage</i>) e Média das Distâncias (<i>Average Linkage</i> para o 3º ano do Ensino Médio.	58

LISTA DE TABELAS

4.1	Estatística descritiva dos resultados da Prova Paraná-2018, para o sexto ano do Ensino Fundamental.	25
4.2	<i>CCC</i> (Coeficiente de Correlação Cofenético) do 6º ano do Ensino Fundamental.	26
4.3	Grupos obtidos por meio da Análise de Agrupamento da Figura 4.2	28
4.4	Descritores com maior número de acertos- sexto ano do Ensino Fundamental.	31
4.5	Estatística descritiva dos resultados da Prova Paraná-2018, para o nono ano do Ensino Fundamental.	32
4.6	<i>CCC</i> (Coeficiente de Correlação Cofenético) para turmas do 9º ano do Ensino Fundamental	33
4.7	Grupos obtidos por meio da Análise de Agrupamento na Figura 4.7. . . .	34
4.8	Descritores com maior número de acertos- nono ano do Ensino Fundamental.	38
4.9	Estatística descritiva dos resultados do Prova Paraná-2018, para o primeiro ano do Ensino Médio	38
4.10	<i>CCC</i> (Coeficiente de Correlação Cofenético) para turmas do 1ª série do Ensino Médio.	39
4.11	Grupos obtidos por meio da Análise de Agrupamento da Figura 4.12. . . .	41
4.12	Estatística descritiva dos resultados da Prova Paraná-2018 para o terceiro ano do Ensino Médio	45
4.13	<i>CCC</i> (Coeficiente de Correlação Cofenético) para turmas do 3ª série do Ensino Médio.	46
4.14	Grupos obtidos por meio da Análise de Agrupamento da Figura 4.17 . . .	48

SUMÁRIO

LISTA DE ILUSTRAÇÕES	9
LISTA DE TABELAS	11
1 INTRODUÇÃO	13
2 REFERENCIAL TEÓRICO	15
2.1 A PROVA PARANÁ	15
2.2 A ANÁLISE DE AGRUPAMENTO	16
2.2.1 Técnicas Hierárquicas Aglomerativas	18
2.2.2 Método da ligação simples ou do vizinho mais próximo (<i>Single linkage</i>) . . .	19
2.2.3 Método da ligação completa ou do vizinho mais distante (<i>Complete linkage</i>)	20
2.2.4 Método da média das distâncias (<i>Average Linkage</i>)	20
2.3 Coeficiente de correlação Cofenético	21
2.4 Critérios para seleção do número de agrupamento	22
3 MATERIAL E MÉTODOS	23
3.1 COLETA DE DADOS	23
3.2 ANÁLISE DE DADOS	23
4 RESULTADOS E DISCUSSÕES	25
4.1 Descrição do 6 ^o ano do Ensino Fundamental	25
4.2 Descrição do 9 ^o ano do Ensino Fundamental	31
4.3 Descrição do 1 ^o ano do Ensino Médio	38
4.4 Descrição do 3 ^o ano do Ensino Médio	44
5 CONSIDERAÇÕES FINAIS	51
A Conjunto de dados avaliados	53
B Dendrogramas	57
C Rotina R	59
REFERÊNCIAS	61

1 INTRODUÇÃO

O objetivo da Prova Paraná é fornecer informações sobre o nível de apropriação dos conhecimentos, em relação aos conteúdos e habilidades considerados essenciais para a etapa de ensino avaliada. Além de ser uma ferramenta para os professores, equipes gestoras e pedagógicas na organização de ações e estratégias que contribuam para o ensino aprendizagem dos estudantes. (PARANÁ, 2019).

Participaram da primeira edição estudantes do 5^o, 6^o e 9^o ano do Ensino Fundamental, 1^o e 3^o ano do Ensino Médio. Os resultados obtidos foram analisados segundo os descritores de habilidades de cada questão, verificando assim os quais apresentaram os menores ou maiores desempenhos em diferentes área de conhecimento.

De acordo com os dados da SEED, cerca de 533 mil alunos de escolas municipais e estaduais participaram dessa edição, sendo cinco mil escolas em todo Paraná que aplicaram a prova com 40 questões, divididas em 20 questões de matemática e 20 de língua portuguesa.

De acordo com site G1 (CUNHA; BISCHOFF, 2019), o resultado da primeira edição da Prova Paraná mostra problemas no ensino de matemática. A matéria aponta que menos de um terço da prova com cálculos foi solucionada corretamente por alunos do Ensino Médio, algo que leva uma certa preocupação, pois grande parte desses alunos estarão cursando um curso do ensino superior no próximo ano.

Conforme o resultado disponibilizado pela SEED, os estudantes dos 6^o e 9^o ano do ensino fundamental e 1^o e 3^o ano do Ensino Médio, apresentaram resultado inferior a 55% da prova, ou seja, os alunos conseguiram desenvolver em torno de metade das competências e habilidades traçadas para cada série de ensino avaliada.

No ano de 2019 está sendo realizada a 2^a e 3^a Edição, sendo que nesta edição todas as series do Ensino Fundamental-nível 2 e Ensino Médio foram inclusas. Porém, pouco se foi discutido sobre os resultados da avaliação anterior, deixando de lado o caráter investigativo e os objetivos para o qual ela foi pensada. Apenas foi disponibilizado no site da Prova Paraná materiais de apoio e web conferência tendo como objetivo ajudar e dar explicações para os professores.

Diante do exposto, surgiu o interesse de fazer uma análise regionalizada do município de Toledo-PR, com objetivo de descrever o perfil dos estudantes do município e comparar com os dados gerais apresentados pela SEED. Dessa forma é possível contribuir para o redirecionamento das estratégias do Núcleo Regional de Educação- (NRE-Toledo) para os descritores com maiores dificuldades, visto que ainda não se tem uma análise específica por NRE ou Município.

Uma das formas de avaliar os diferentes descritores avaliados é a utilização

de técnicas estatísticas multivariadas. Segundo Mingoti (2005), a estatística multivariada, consiste em um conjunto de métodos estatísticos utilizados em situações nas quais várias variáveis são medidas simultaneamente, em cada elemento amostral. Em geral, as variáveis são correlacionadas entre si e quanto maior o número de variáveis, mais complexa torna-se a análise por métodos comuns de estatística univariada.

Técnicas analíticas multivariadas têm sido utilizadas em diferentes áreas como: Psicologia, Geologia, Química, Física, Engenharia, Agronomia, Ciências Sociais e Biológica, entre outras áreas. Ela pode ser aplicada com diversas finalidades, mesmo nos casos em que não se dispõe de antemão de um modelo teórico rigorosamente estruturado a respeito das relações entre as variáveis. A finalidade de sua aplicação pode ser de reduzir dados ou de simplificação estrutural, de classificar e agrupar, de investigar a dependência entre variáveis, de predição e de elaborar hipóteses e testá-las (JOHNSON; WICHERN, 2002).

Como a finalidade do trabalho é agrupar os descritores de habilidades de acordo com o índice de acertos das questões, o método mais indicado é a Análise de Agrupamento (Cluster Analysis), que classifica os objetos em categorias de similaridade. Para Prado (2015) essa técnica considera um conjunto inicial de objetos, aos quais são associadas medidas de várias grandezas, denominadas variáveis classificatórias, utilizadas para se obter grupos de objetos assemelhados em relação aos valores assumidos por essas variáveis. Devido a propriedade de hierarquia, é possível construir um gráfico chamado de Dendrograma ou Dendograma que representa a "árvore" ou a história do agrupamento (MINGOTI, 2005)

Assim, o objetivo do presente trabalho é aplicar a técnica de Agrupamento em dados quantitativos discretos para estudar e comparar o desempenho dos estudantes de uma escola do Núcleo Regional de Toledo (NRE-Toledo) na Prova Paraná 2018 ¹.

¹A escola aqui utilizada não terá seu nome divulgado. Apesar de termos a autorização para o uso dos dados analisados na pesquisa, como muitas outras escolas também fizeram a mesma prova, não queremos que essa escola sirva como parâmetros de comparação para outras. Assim, para não cometer injustiças, optamos por não apresentar o nome da escola.

2 REFERENCIAL TEÓRICO

2.1 A PROVA PARANÁ

A Prova Paraná tem como alguns dos seus objetivos: identificar as dificuldades apresentadas por cada um dos estudantes e apontar as habilidades já apropriadas no processo de ensino e aprendizagem, nas disciplinas de Língua Portuguesa e Matemática (PARANÁ, 2019).

Participam dessa avaliação os estudantes do 5^o, 6^o e 9^o ano do Ensino Fundamental, 1^a e 3^a séries do Ensino Médio. Esses resultados foram analisados segundo os descritores de cada questão, como apresentados na Figura 2.1. Por meio dos descritores verificou-se quais apresentaram os menores ou maiores desempenhos nas diferentes disciplinas.

**Relação de descritores da Prova Paraná – 2ª edição – Matemática
7º ano do Ensino Fundamental**

I. Espaço e Forma	
D1	Identificar a localização/movimentação de objeto em mapas, croquis e outras representações gráficas.
D2	Identificar propriedades comuns e diferenças entre poliedros e corpos redondos, relacionando figuras tridimensionais com suas planificações.
D3	Identificar propriedades comuns e diferenças entre figuras bidimensionais pelo número de lados, pelos tipos de ângulos.
D5	Reconhecer a conservação ou modificação de medidas dos lados, do perímetro, da área em ampliação e/ou redução de figuras poligonais usando malhas quadriculadas.
D11	Reconhecer círculo/circunferência, seus elementos e algumas de suas relações.
D44	Reconhecer figuras tridimensionais por meio de suas características.
II. Grandezas e Medidas	
D12	Resolver problema envolvendo o cálculo de perímetro de figuras planas.
D13	Resolver problema envolvendo o cálculo de área de figuras planas.
III. Números e Operações/Álgebra e Funções	
D13	Reconhecer e utilizar características do sistema de numeração decimal, tais como agrupamentos e trocas na base 10 e princípio do valor posicional.
D15	Reconhecer a decomposição de números naturais nas suas diversas ordens.
D21	Identificar diferentes representações de um mesmo número racional.

Figura 2.1: Modelo do quadro de descritores da Avaliação: Prova Paraná

Fonte: <http://www.provaparana.pr.gov.br/>

No Quadro 1 é apresentado os resultados gerais que foram disponibilizados pela Secretaria da Educação. Junto com esses dados foi informado que em alguns descritores os desempenhos dos estudantes foram muito baixo, como por exemplo: resolver problemas com frações (apenas 10% acertos); resolver problemas que envolvam porcentagem (10% de acertos) e reconhecer diferentes representações de números racional (11% de acertos).

Note pelo quadro 1 que o desempenho é decrescente em relação a série de

ensino, o que gera uma preocupação do embasamento teórico dos alunos para o ensino superior.

Quadro 1: Desempenho dos estudantes na Prova Paraná

Ano/ Série	REDE	Língua Portuguesa	Matemática
5º Ano do Ensino Fundamental	MUNICÍPIO	64,85%	69,25%
6º Ano do Ensino Fundamental	ESTADO	71,46%	49,17%
9º Ano do Ensino Fundamental	ESTADO	64,65%	39,77%
1ª Série do Ensino Médio	ESTADO	61,35%	28,55%
3ª/4ª Série do Ensino Médio	ESTADO	61,41%	29,62%

Fonte: <<http://www.provaparana.pr.gov.br/>>

Porém, é necessário um levantamento enquanto NRE. A avaliação só encontra sentido se analisada e compartilhada entre equipes pedagógicas e professores(as) e entre os(as) próprios(as) professores(as), tendo como direcionamento o que se estabelece nos PPP escolares. A partir dessas discussões é necessário reconstruir processo educativo, conjuntamente.

2.2 A ANÁLISE DE AGRUPAMENTO

Na literatura encontramos várias definições para a técnica de Análise de Agrupamento (conglomerado ou Análise de *clusters*), segundo Mingoti (2005), tem como objetivo dividir os elementos da amostra, ou população, em grupos de forma que os elementos pertencentes a um mesmo grupo sejam similares entre si com respeito às variáveis características que neles foram medidas. Os agrupamentos resultantes devem então exibir elevada homogeneidade interna (dentro dos agrupamentos) e elevada heterogeneidade externa (entre agrupamentos). Para Khattree e Naik (2000), a Análise de Agrupamento é uma técnica multivariada de grande aplicabilidade, principalmente o procedimento cluster, cujo objetivo da classificação é repartir os indivíduos em grupos homogêneos, de modo que cada seja bem diferenciado. Após a obtenção dos resultados, esses dados servirão para a definição do número de grupos distintos.

De acordo com Mardia et al. (1995), a Análise de Agrupamentos apresenta vantagem de reduzir o espaço multidimensional a uma medida de distância entre os objetos representados, esta em um espaço bidimensional, muito mais simplificado do que o espaço multidimensional. Portanto é considerada uma ferramenta de análise que visa á triagem de diferentes objetos, de modo que a associação entre eles seja a maior possível, se eles pertencem ao mesmo grupo e o mínimo em caso contrário.

Sendo assim, se a aglomeração for bem-sucedida, quando representados em um gráfico, os objetos dentro dos conglomerados estarão muito próximos enquanto os conglo-

merados distintos estarão afastados. Logo as similitudes entre os aglomerados mostraram os descritores com maiores dificuldades, além de mostrar os descritores com menores dificuldades.

Esse processo de agrupamento começa levando as medidas das p variáveis em cada um dos n objetos. Assim, a matriz $n \times p$ de dados é transformada em uma matriz $n \times n$ de semelhança ou, alternativamente, por medidas de distância onde são computadas as semelhanças ou distâncias entre pares de objetos pelas p variáveis. Logo, um algoritmo é selecionado com a finalidade de definir as regras que concernem ao agrupamento dos objetos em subgrupos com base nas semelhanças (VALLI, 2002).

A maioria dos métodos utilizados em análise de *cluster* requer uma medida de similaridade ou dissimilaridade entre os elementos quer a serem agrupados, a maioria das vezes é expressa como uma função distância para dados que têm propriedades métricas.

Quando trabalhamos com variáveis quantitativas é comum o uso de medidas de dissimilaridade e, logo, quanto menor os seus valores, mais similares serão os elementos que estão sendo comparados, podendo-se trabalhar com variáveis padronizadas quanto houver diferença na escala de medidas. Segundo Khattree e Naik (2000) e Cruz e Carneiro (2006), dentre as medidas de dissimilaridades conhecidas, a distância Euclidiana e a distância de Mahalanobis estão entre as medidas que mais se destacam, devido a sua maior utilização.

Neste estudo, para medir a distância entre os resultados de cada descritor. Utilizou-se a distância Euclidiana, que segundo Carmack (1971) citado por Prado (2015) a distância Euclidiana entre dois casos (i e j) é a raiz quadrada do somatório dos quadrados das diferenças entre os valores de i e j para todas as variáveis ($k=1,2,\dots,p$), isto é:

$$d_{ij} = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ik} - x_{jk})^2} \quad (2.1)$$

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ip} - x_{jp})^2} \quad (2.2)$$

Em que d_{ij} representa o valor observado da j -ésima variável no i -ésimo **indivíduo**.

Temos ainda as medidas de similaridade, que define o critério de quão similares são os indivíduos e se serão denominadas ou não a determinado padrão quando, quanto maior os seus valores mais similares serão os elementos que estão sendo comparados.

O coeficiente de correlação é uma medida de similaridade. Assim, inicialmente uma matriz $n \times n$ de similaridades entre todos os pares da observação é calculada (podendo ser a matriz com os coeficientes de correlação de Spearman ou de Person), e as colunas representando membros de um único agrupamento tenderão apresentar intercorrelações que serão agrupadas. Esse procedimento é repetidamente feito até que se agrupe todos os elementos.

Entre os procedimentos para agrupar os elementos podemos citar:

1) Matriz de similaridade, citada por Ferreira (2011), onde a montagem é realizada por meio de funções de correlação: $s_{ik} = |r_{ik}|$, $s_{ik} = (r_{ik})^2$, $s_{ik} = \frac{1+(r_{ik})^2}{2}$

2) Matriz de dissimilaridade, proposta por Rencher (2002), onde agrupamos os elementos através de $s_{ik} = 1 - (r_{ik})^2$.

Em muitas situações trabalhamos com dados qualitativos. Nestes casos, há duas formas de tratamento: ou transformamos estas variáveis para quantitativos e trabalhamos com medidas de similaridade descritas anteriormente, ou então trabalhamos com coeficientes de similaridades especialmente desenvolvidos para as variáveis qualitativas. Os dois mais utilizados são:

1) **Jaccard**: $\frac{a}{a+b+c}$, similaridade de indivíduos com dupla presença, excluindo a dupla ausência.

2) **Sorensen**: $\frac{2a}{2a+b+c}$, similaridade de indivíduos com dupla presença, excluindo a dupla ausência- peso duplo para $1 - 1$.

As técnicas de agrupamentos são frequentemente classificadas em dois tipos: técnicas hierárquicas, que na maioria das vezes são utilizadas em análises exploratórias dos dados com o intuito de identificar possíveis agrupamentos e o possível valor de g e as não hierárquicas, em que o número de grupos já esteja pré-estabelecido pelo pesquisador.

2.2.1 TÉCNICAS HIERÁRQUICAS AGLOMERATIVAS

As técnicas de conglomerados ou agrupamentos são frequentemente classificadas em dois tipos: técnicas hierárquicas e não hierárquicas, sendo que as hierárquicas são classificadas em aglomerativas e divisivas. As técnicas hierárquicas, na maioria das vezes, são utilizadas em análises exploratórias dos dados com o intuito de identificar possíveis agrupamentos e o valor provável do número de grupos. Já para o uso de técnicas não hierárquicas, é necessário que o valor do número de grupos já esteja pré-estabelecido pelo pesquisador (MINGOTI, 2005).

Segundo (MINGOTI, 2005), as técnicas hierárquicas aglomerativas partem do princípio de que no início do processo de agrupamento tem-se conglomerados, ou seja, cada elemento do conjunto de dados observado é considerado como sendo um conglomerado isolado. Em cada passo do algoritmo, os elementos amostrais vão sendo agrupados, formando novos conglomerados até o momento no qual todos os elementos considerados estão em um único grupo. Já nos métodos divisivos, todos os objetos pertencem inicialmente ao mesmo grupo, que vai sendo dividido, até que cada observação forme um grupo individualmente (JOHNSON; WICHERN, 2002). Veja na Figura 2.2.

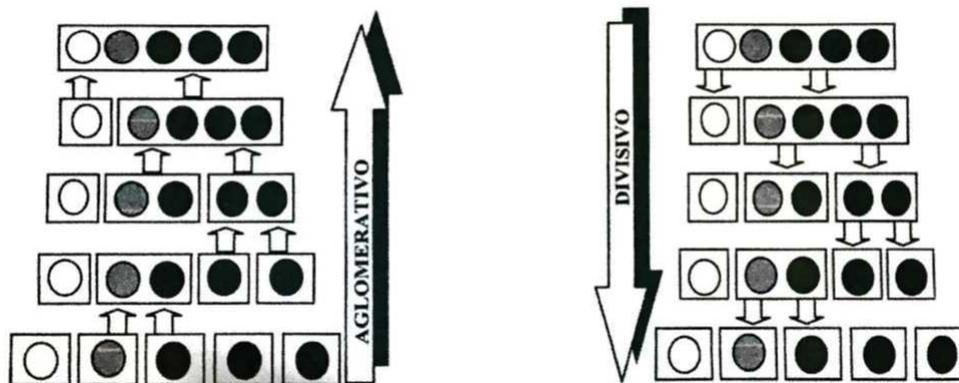


Figura 2.2: Esquema geral de procedimentos hierárquicos aglomerativos e divisivos.

Fonte: (MINGOTI, 2005)

O resultado de uma Análise de Agrupamento é representado em um dendrograma, onde apresenta o arranjo entre os objetos em escalas de distância. O dendrograma é um gráfico em forma de árvore no qual a escala vertical indica o nível de similaridade (ou dissimilaridade). No eixo horizontal, são marcados os elementos amostrais numa ordem conveniente relacionada à história de agrupamento. As linhas verticais, partindo dos elementos amostrais agrupados, têm altura correspondente ao nível em que os elementos foram considerados semelhantes, isto é, a distância do agrupamento ou o nível de similaridade (MINGOTI, 2005).

Existem vários métodos de agrupamento hierárquicos, a seguir falaremos sobre três deles. Método da Ligação Simples (conhecido também como Single Linkage ou Critério do Vizinho mais Próximo); Ligação Completa (conhecido também como Complete Linkage ou Critério do Vizinho mais Distante) e Método da média das distâncias (conhecido também como Average Likage).

2.2.2 MÉTODO DA LIGAÇÃO SIMPLES OU DO VIZINHO MAIS PRÓXIMO (*Single linkage*)

Segundo Sneath (1957) citado por Mingoti (2005, p.166), neste método, a similaridade entre dois conglomerados é definida pelos dois elementos mais parecidos entre si. Portanto a similaridade entre dois cluster é definida pela menor distância de qualquer ponto do 1º cluster para qualquer ponto do 2º cluster, assim a distância entre uma observação w e um grupo formado pelas observações u e v é dada por:

$$d_{(uv)w} = \min(d_{uw}, d_{vw}) \quad (2.3)$$

em que $D_{(uv)w}$ é a distância entre o grupo uv e a observação w ; e $\min(d_{uw}, d_{vw})$ é a menor distância entre os grupos de observações uw e vw (JUNIOR; THIEBAUT, 1999).

Segundo Anderberg (2014), algumas das características desse método são: Em geral grupos muito próximos pode não ser identificados; Permite detectar grupos de formas não-elípticas; Apresenta pouca tolerância a ruído, pois tem a tendência a incorporar os ruídos em um grupo já existente; Apresenta bons resultados tanto para distância euclidianas quanto para outras distâncias; e tem tendência a formar longas cadeias (enca-deamento).

2.2.3 MÉTODO DA LIGAÇÃO COMPLETA OU DO VIZINHO MAIS DISTANTE (*Complete linkage*)

Segundo Sneath (1957) citado por (MINGOTI, 2005, p.166). Neste método, a similaridade entre dois conglomerados é definida pelos elementos que são “menos semelhantes” entre si. Essa ligação é exatamente a oposta do Método da Ligação Simples. O algoritmo é iniciado encontrando a menor distância $D = \{d_{ik}\}$ e agrupando os objetos correspondentes (U e V) para formar o grupo (UV). No passo 3 do algoritmo, as distâncias entre (UV) e qualquer outro grupo são calculadas por (JOHNSON; WICHERN, 2002) :

$$d_{(uv)w} = \max\{d_{uw}, d_{vw}\} \quad (2.4)$$

Para Quintal (2006), este método tem tendência para encontrar clusters compactos compostos de objetos muito semelhantes entre si. Quando um objeto é acrescentado a um grupo, a distância do novo grupo aos restantes aumenta ou então fica inalterada. O método de ligação completa tende a formar grupos pequenos que depois serão aglutinados para formar grupos maiores.

2.2.4 MÉTODO DA MÉDIA DAS DISTÂNCIAS (*Average Linkage*)

Este método trata as distâncias entre dois aglomerados como a média das distâncias entre todos os pares de elementos que podem ser formados com os elementos dos dois conglomerados que estão sendo comparados (MINGOTI, 2005). Assim, se o conglomerado D_{uv} tem n_{uv} elementos e o conglomerado D_w tem n_w elementos, a distância entre eles será calculada por:

$$\bar{d}_{(uv)w} = \frac{\sum_{i \in C_{uv}} \sum_{k \in C_w} d_{ik}}{n_{uv}n_w} \quad (2.5)$$

Algumas características de método são: a menor sensibilidade a ruídos que os métodos citados anteriormente; apresenta bons resultados tanto para distâncias Eu-

clidianas quanto para outras distâncias; tem tendência a formar grupos com números de elementos similares (ANDERBERG, 2014; CRUZ; CARNEIRO, 2006).

Podemos representar graficamente os resultados de um procedimento hierárquico através do dendrograma no qual o objeto de estudo é colocado em um eixo e o outro eixo representa os passos do procedimento. Começando com cada objeto representado como um agrupamento separado, o dendrograma mostra graficamente como os agrupamentos são combinados em cada passo dos procedimentos até que todos estejam contidos em um único agrupamento.

Para validar os agrupamentos hierárquicos podemos utilizar uma medida chamada de Coeficiente de Correlação Cofenético para avaliar o grau de ajuste do agrupamento.

2.3 Coeficiente de correlação Cofenético

De acordo com Barroso e Artes (2003) e Cruz e Carneiro (2006) citado por (PRADO, 2015, p.17). O coeficiente de correlação linear de Pearson entre os elementos da matriz de dissimilaridade (matriz de distâncias entre os indivíduos, obtida a partir dos dados originais) e os elementos da matriz cofenética (matriz de distâncias entre os indivíduos, obtida a partir do dendrograma) é denominado Coeficiente de Correlação Cofenético (*CCC*). Esse coeficiente pode ser utilizado para avaliar a consistência do padrão de agrupamento de métodos de agrupamentos hierárquicos, sendo que valores próximos à unidade indicam melhor representação.

O (*CCC*) mede o grau de preservação das distâncias emparelhadas pelo dendrograma resultante do agrupamento em relação as distâncias originais (SNEATH; SOKAL, 1975) ou seja, tem como objetivo avaliar o grau de similaridade da matriz de dissimilaridade (distâncias) dos dados originais e a matriz de dissimilaridade preservadas pelo agrupamento. Segundo Rohlf (1970), se *CCC* igual ou acima de 0,70 o agrupamento é considerado adequado.

O (*CCC*) é calculado por (BUSSAB et al., 1990)

$$r_{cof} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (c_{ij} - \bar{c})(s_{ij} - \bar{s})}{\sqrt{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (c_{ij} - \bar{c})^2} \sqrt{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (s_{ij} - \bar{s})^2}} \quad (2.6)$$

Onde c_{ij} é valor de similaridade entre o indivíduos i e j , obtidos a partir da matriz cofenética; s_{ij} é valor de similaridade entre os indivíduos i e j , obtidos a partir da matriz de similaridade; $\bar{c} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n c_{ij}$; $\bar{s} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n s_{ij}$.

Já os métodos não hierárquicos mais utilizados são os das k -médias e o *Fuzzy*

c-Médias (*Fuzzy c-Means*), assim como redes neurais artificiais aplicadas à análise de clusters.

2.4 Critérios para seleção do número de agrupamento

Após a construção do dendrograma e realizado os agrupamentos, é necessário decidir o ponto de corte, ou seja, definir o número de grupos k que mais representa os dados da amostra, podendo interferir diretamente na qualidade do agrupamento. A seleção do número ótimo de grupos na etapa final dos estudos que utilizam os métodos hierárquicos de agrupamento é uma tarefa difícil para os pesquisadores (DIAS, 2014).

Segundo Mingoti (2005), uma questão de grande importância é de como se deve proceder na escolha do número final de grupos que define a partição do conjunto de dados analisados, ou em qual passo o algoritmo de agrupamento deve ser interrompido. Nos métodos hierárquicos o pesquisador pode escolher o número de grupos k , escolhendo assim um ponto de corte na escala de distância e cortando os ramos do dendrograma.

Existem vários trabalhos que fazem a comparação entre critérios de paradas, ou sugerem formas de trabalhar esta questão, entre eles o de Mejena(1977) apud. Ferreira (2011) propôs um critério para determinar o número de grupos k que otimiza a qualidade do ajuste do agrupamento aos dados. Este tem objetivo de buscar a maior amplitude nas distâncias de junção dos grupos formados.

3 MATERIAL E MÉTODOS

3.1 COLETA DE DADOS

Os dados da Prova Paraná-2018, utilizados nesse trabalho, foram obtidos junto ao NRE-Toledo. Este tem sede localizada no endereço Av. Tiradentes, 1001 - Centro, Toledo - PR, 85900-230

As informações foram fornecidas por meio de um relatório (PDF), neste foi possível verificar os descritores de maior e menor desempenho para cada série e área do conhecimento avaliada. Posteriormente os dados foram digitalizados, tabulados e processados.

3.2 ANÁLISE DE DADOS

De primeiro momento, foi realizada uma análise descritiva dos dados usando estatística básica, como: cálculo da média, mediana, desvio padrão e coeficiente de variação e, por fim, obtido o gráfico boxplot para cada série analisada. Depois da análise descritiva partiu-se para realização da Análise de Agrupamento, aplicada com objetivo de analisar a similaridade entre os resultados obtidos sobre cada descritor, ou seja, avaliar quais descritores possuem resultados semelhantes em relação a resultados da prova, além de descrever o perfil dos agrupamentos obtidos. Para isso as análises foram implementadas no Software estatístico R (OKSANEN et al., 2016).

A análise se iniciou a partir de uma matriz de dissimilaridade, que foi obtida por meio de uma medida de distância mais adequada, visto que os procedimentos em análise de agrupamento são influenciados pela natureza das variáveis ou dos atributos dos objetos. Sendo assim, foi considerado a natureza dos dados, isto é, o fato dos dados apresentarem grandezas (escalas) distintas, foi utilizada neste presente trabalho a distância Euclidiana para obtenção da matriz de distância.

Para a formação dos agrupamentos os métodos da Ligação Simples (Single Linkage), Ligação Completa (Complete Linkage) e Método da Média das Distância (Average Linkage) foram aplicados. Os agrupamentos usados por cada método de ligação foram representados graficamente pelo dendrograma, que auxiliou na identificação dos descritores com resultado similares entre si.

Para comparação e escolha de qual método de ligação seria mais adequado aos dados, foi feito o diagnóstico por meio do (CCC) entre as matrizes e os agrupamentos. Segundo (ROHLF, 1970), se CCC for igual ou acima de 0,70 o agrupamento é considerado

adequado, ou seja, quanto maior a concordância entre os agrupamentos, maior deverá ser o resultado do CCC calculado.

Após a análise estatística realizou-se a comparação entre os agrupamentos obtidos e as informações disponíveis no Site da SEED e em seguida disponibilizados ao NRE Toledo.

Todas as análises foram implementadas utilizando o software gratuito e aberto, R Core Team (2019).

4 RESULTADOS E DISCUSSÕES

A seguir apresenta-se a análise estatística dos dados. Inicialmente foi realizada a análise descritiva de cada série de ensino e em seguida foi aplicada a técnica multivariada de correspondência para agrupar os descritores de maior e menor desempenho. Por fim, apresenta-se uma breve discussão das questões que apresentaram maiores índices de erros.

4.1 Descrição do 6^o ano do Ensino Fundamental

Analisando a Tabela 4.1 verifica-se que a média de acertos das turmas variou de 16,10 a 17,25, sendo que em todas as turmas ocorreram questões em que somente um aluno acertou ou nenhum (em total de 90 alunos). Essa questão se refere ao descritor D-23: Resolver problemas com números reais envolvendo diferentes significados das operações. Já o maior índice de acertos nas diferentes turmas (Máximo) está relacionada a área de tratamento de informações. Ademais, os resultados apresentaram alto coeficiente de variação indicando altíssima dispersão amostral em relação ao valor médio. A turma C foi que apresentou a maior variação (52,93%), acompanhada da turma A e B respectivamente (52,23% e 46,45%).

Tabela 4.1: Estatística descritiva dos resultados da Prova Paraná-2018, para o sexto ano do Ensino Fundamental.

Turma	A	B	C
Mínimo	1,00	1,00	0,00
1 ^o quartil	10,00	13,25	9,25
Mediana	16,00	18,00	16,50
Média	17,15	17,25	16,10
3 ^o quartil	25,25	23,25	25,00
Máximo	31,00	29,00	29,00
Desvio Padrão	8,96	8,01	8,52
Variância	80,24	64,20	72,62
Coeficiente de Variação (%)	52,23	46,45	52,93

A Figura 4.1 apresenta o Boxplot do posicionamento dos acertos em relação aos quartis. Nota-se que nas três turmas 25% das questões apresentaram menos de 10 acertos e que as turmas A e C apresentaram comportamento levemente assimétrico positivo. No entanto, a turma B apresentou o maior valor para a mediana (Tabela 4.1).

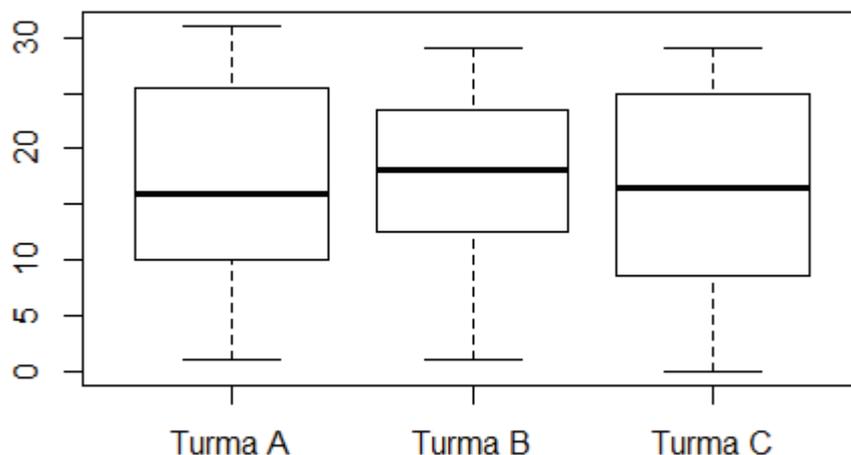


Figura 4.1: Boxplot do número de acertos na Prova Paraná do 6^o ano do Ensino Fundamental.

Para analisar as questões que apresentaram maiores índices de erros, a Análise de Agrupamento foi aplicada. O objetivo é agrupar os resultados com maior similaridade entre si. Considerando a natureza de cada variável agrupada, foi utilizada a medida de distância Euclidiana pois os dados apresentam a mesma escala mas com grandezas diferentes.

Para a formação dos agrupamentos foram aplicados os métodos de Ligação Simples (*Single Linkage*), Ligação Completa (*Complete Linkage*), Média das Distância (*Average Linkage*). Para verificar o ajuste da matriz de dissimilaridade a partir de cada método de agrupamento, calculou-se o (*CCC*). A Tabela 4.2 apresenta os valores obtidos. Verifica-se que para o sexto ano o método que melhor representou a matriz original foi o Método da Média das Distância (*Average Linkage*).

Tabela 4.2: *CCC* (Coeficiente de Correlação Cofenético) do 6^o ano do Ensino Fundamental.

Métodos	CCC
Método da Ligação Simples (<i>Single Linkage</i>)	0.6089
Método da Ligação Completa (<i>Complete Linkage</i>)	0.7388
Método da Média das Distâncias (<i>Average Linkage</i>)	0.7507

Definido o método de agrupamento, faz-se necessário decidir o ponto de corte, ou seja, definir o número de grupos k mais representativos dos dados da amostra. Este

passo interfere diretamente na qualidade do agrupamento. Aplicou-se o método proposto por Mojena (1977) apud. Ferreira (2011).

A Figura 4.2 apresenta o dendrograma que melhor representou o agrupamento dos descritores para as turmas do sexto ano. Note que quatro grupos foram formados. Os dendrogramas comparando cada método encontra-se no Anexo B.

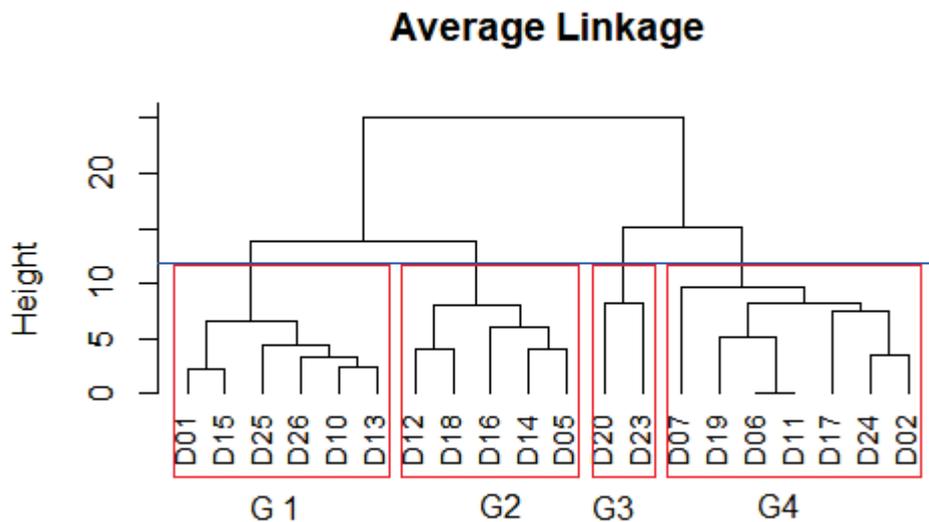


Figura 4.2: Dendrograma formado a partir da Distância Euclidiana e Método da Média das Distância, com quatro *cluster*- sexto ano do Ensino Fundamental.

Os quatro grupos formados a partir da análise do dendrograma estão separados na Tabela 4.3. O primeiro grupo é composto por 6 descritores, onde número de acerto total em cada questão varia entre 71 a 87 (78,88% a 96,67%). O grupo dois é composto por 5 descritores, tendo como número de acerto total variando entre 53 a 65 (58,88% a 72,22%). O terceiro grupo é composto por dois descritores, sendo esses os descritores que apresentaram um menor número de acertos, 15 e 2 acertos nessa respectiva ordem (16,67% e 2,22%). O último grupo formado, contém 7 descritores e seus resultados variam entre 25 e 41 acertos (27,77% a 45,55%).

Por meio das análise já feitas, tem-se que dois descritores apresentaram maior dificuldades sendo eles: D-20 Reconhecer/Identificar diferentes representações de um número racional e D-23 Resolver problemas com números reais envolvendo diferentes significados das operações, sendo esses descritores avaliados em questões específicas na prova.

Tabela 4.3: Grupos obtidos por meio da Análise de Agrupamento da Figura 4.2

Grupo 1	Grupo 2	Grupo 3	Grupo 4
D01 Identificar a localização/movimentação de objetos ou pessoas em mapas, croquis e outras representações gráficas;	D12 Resolver problemas envolvendo o cálculo de áreas de figuras planas;	D20 Reconhecer/Identificar diferentes representações de um número racional;	D07 Resolver problemas utilizando relações entre diferentes unidades de medida;
D15 Reconhecer a decomposição ou composição de números naturais nas suas diversas ordens;	D18 Resolver problemas com números reais envolvendo diferentes significados das operações;	D23 Resolver problemas com números reais envolvendo diferentes significados das operações;	D19 Resolver problemas com números reais envolvendo diferentes significados das operações;
D25 Identificar informações apresentadas em tabelas ou diferentes tipos de gráficos;	D16 Calcular o resultado de uma adição ou subtração de números naturais;		D06 Reconhecer figuras com simetria de reflexão e/ou identificar seus eixos de simetria;
D26 Identificar informações apresentadas em tabelas ou diferentes tipos de gráficos;	D14 Reconhecer números reais representados em diferentes contextos;		D11 Resolver problemas envolvendo o cálculo de perímetro de figuras planas;
D10 Resolver problemas envolvendo trocas entre cédulas e moedas do Sistema Monetário Brasileiro, em função de seus valores;	D05 Reconhecer a conservação ou modificação de medidas dos lados, do perímetro, da área em ampliação e/ou redução de figuras poligonais usando malhas quadriculadas;		D17 Calcular o resultado de uma multiplicação ou divisão de números naturais;
D13 Reconhecer e utilizar características do Sistema de Numeração Decimal;			D24 Resolver problemas que envolvam porcentagem;
			D02 Relacionar figuras tridimensionais à sua planificação ou vistas por meio de suas propriedades e vice-versa;

A primeira questão está relacionada ao descritor 20 Figura 4.3. Neste item a habilidade a ser avaliada é reconhecer diferentes representação de um número racional. Os alunos deveriam fazer a transformação da fração em número decimal e depois representá-lo em porcentagem. Sendo assim os alunos deveriam ter um grande domínio em divisão e saber conceitos envolvendo a porcentagem.

- 24) (M050083H6) Maurício gasta por mês $\frac{1}{5}$ do seu salário com alimentação. Esse gasto mensal corresponde a qual percentual de seu salário?
- A) 0,2%
 - B) 1,5%
 - C) 2%
 - D) 20% (alternativa correta)

Figura 4.3: Questão 24 retirada da Prova Paraná-2018, Ensino Fundamental - 6º ano.
Fonte: (PARANÁ, 2019)

A Figura 4.4 (questão 36) apresenta a questão que os alunos apresentaram maior dificuldade em que apenas dois (2) alunos, de um total de 90 alunos, conseguiram responder de maneira correta. A questão tinha como objetivo avaliar conhecimento sobre resolver problemas com números reais, envolvendo diferentes significados das operações. Para isso o aluno deveria saber o conceito de soma de frações com denominadores diferentes.

- 36) (M050034H6) Felipe gasta $\frac{1}{3}$ de seu salário no pagamento da mensalidade de sua faculdade e $\frac{2}{5}$ com as demais despesas fixas do mês. A fração do salário de Felipe usada no pagamento da mensalidade da faculdade e das demais despesas fixas do mês é
- A) $\frac{3}{15}$.
 - B) $\frac{3}{8}$.
 - C) $\frac{3}{5}$.
 - D) $\frac{11}{15}$. (alternativa correta)

Figura 4.4: Questão 36 retirada da Prova Paraná-2018, Ensino Fundamental - 6º ano.
Fonte: (PARANÁ, 2019)

As duas questões apresentada trabalham com conteúdo de frações, mostrando que os alunos apresentam grande dificuldades no conteúdo proposto. A questão 24 (Figura 4.3) mostra que além do conteúdo de frações os alunos apresentaram dificuldades no entendimento/aprendizagem do conceito de porcentagem.

Se tratando de um conteúdo de suma importância, é preocupante a dificuldade apresentada pelos alunos nesse conteúdo. Piaget classifica sete condições que coloca como

essenciais á existência de frações: a existência de uma totalidade divisível, a existência de um número determinado de partes, o esgotamento do todo, a relação entre número de partes e o número de cortes, a igualização das partes, a conceitualização de cada fração como parte de um todo em si, suscetível de novas direções e o atendimento ao princípio da invariância (POLESE et al., 2011)

De acordo com a estrutura disponível pela Diretrizes Curriculares da Educação Básica do Estado Paraná-Matemática (PARANÁ, 2008), no conteúdo de números fracionários, os alunos devem aprender a calcular MMC (Mínimo Múltiplo Comum) e MDC (Máximo Divisão Comum) além de trabalhar com relação de igualdade e transformação de fração e número decimal.

Andando na contramão dos dados apresentados acima, uma questão que chamou bastante a atenção por ter uma grande quantidade de acertos é a questão 33 da prova. Esta questão tinha como objetivo avaliar as habilidades dos estudantes em ler, analisar e interpretar informações e dados apresentados em tabelas. Em duas turmas todos os estudantes acertaram (100%) e apenas em uma cerca de 93,10% responderam de maneira correta a questão.

Questão 33 – D25 – Identificar informações apresentadas em tabelas ou diferentes tipos de gráficos.

33) (M052201E4) A tabela abaixo apresenta o número de atletas inscritos, por idade, para uma corrida de rua que acontece anualmente em uma cidade.

Idade	Número de inscritos
18 a 30 anos	879
31 a 40 anos	734
41 a 50 anos	410
Acima de 50 anos	651

De acordo com essa tabela, quantos atletas com mais de 50 anos se inscreveram para essa corrida de rua?

- A) 879
- B) 734
- C) 651 (alternativa correta)
- D) 410

Figura 4.5: Questão 33 retirada da Prova Paraná-2018, Ensino Fundamental - 6º ano.
Fonte: (PARANÁ, 2019)

Na Tabela 4.4 destaca outros descritores que apresentaram grande número de

acertos em cada turma.

Tabela 4.4: Descritores com maior número de acertos- sexto ano do Ensino Fundamental.

Descritor	Turma A	Turma B	Turma C
D10 Resolver problemas envolvendo trocas entre cédulas e moedas do Sistema Monetário Brasileiro, em função de seus valores	90,32%	90,00%	86,66%
D13 Reconhecer e utilizar características do Sistema de Numeração Decimal	93,55%	86,21%	93,10%
D26 Identificar informações apresentadas em tabelas ou diferentes tipos de gráficos	87,10%	90,00%	100%

4.2 Descrição do 9º ano do Ensino Fundamental

Observando a Tabela 4.5, que mostra os dados das turmas do nono ano, nota-se que a média de acertos das turmas variou de 13,75 a 15,10. Em todas a turmas ocorreram questões que somente dois alunos acertaram ou apenas um, em um total de 88 alunos. Essa questão se refere a descritor D-08: Resolver problemas utilizando as propriedades dos polígonos. Já o maior índice de acertos nas diferentes turma (Máximo) esta relacionado a área de tratamento de informações. Os resultados apresentaram um alto valor para o coeficiente de variação, onde indica uma dispersão amostral em relação ao valor médio. As turmas A e B, foram as que apresentaram maior índice de variação (54,94% e 54,55%), a turma C apresentou o menor (48,18%).

Tabela 4.5: Estatística descritiva dos resultados da Prova Paraná-2018, para o nono ano do Ensino Fundamental.

Turma	A	B	C
Mínimo	2,00	1,00	2,00
1º quartil	10,00	9,50	10,00
Mediana	14,00	15,00	13,00
Média	15,10	14,90	13,75
3º quartil	20,25	20,25	19,50
Máximo	32,00	27,00	26,00
Desvio Padrão	8,30	7,53	6,62
Variância	68,83	56,73	43,88
Coefficiente de Variação (%)	54,94	54,55	48,18

A Figura 4.6 mostra o Boxplot das três turmas do nono ano, indicando o posicionamento dos acertos em relação aos quartis. Observe que nas três turmas 25% das questões apresentaram menos de 10 acertos, como mostra o resultado do 1º quartil Tabela 4.5, resultado semelhante ao ocorrido no sexto ano. Nota-se que as turmas A e C apresentam comportamento assimétrico positivo a turma B foi que apresentou o maior valor de mediana (Tabela 4.5). Destaque para a Turma A que apresentou maior número de acertos nas questões.

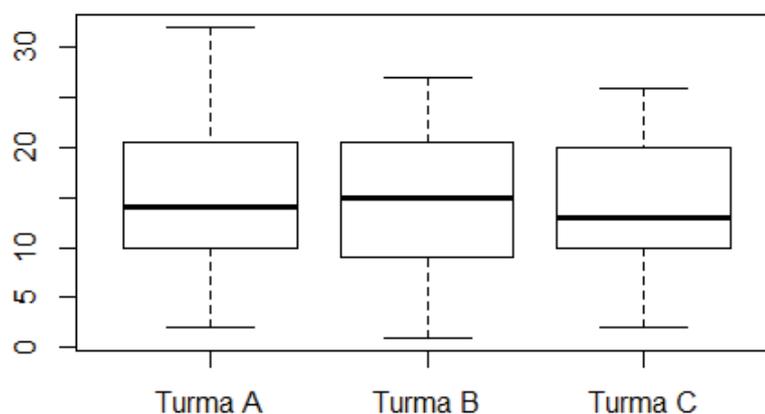


Figura 4.6: Boxplot do número de acertos na Prova Paraná do 9º ano do Ensino Fundamental.

Para analisar as questões que apresentaram maiores índices de erros, a Análise

de Agrupamento foi aplicada, utilizando novamente a medida de distância Euclidiana. A Tabela 4.6 apresenta os valores obtidos. Verifica-se que, como ocorreu para o sexto ano, o método que melhor representou a matriz original foi o Método da Média das Distância (*Average Linkage*).

Tabela 4.6: *CCC* (Coeficiente de Correlação Cofenético) para turmas do 9º ano do Ensino Fundamental

Métodos	CCC
Método da Ligação Simples (Single Linkage)	0.66
Método da Ligação Completa (Complete Linkage)	0.7468
Método da Média das Distâncias (<i>Average Linkage</i>)	0.7499

Na Figura 4.7 apresenta o dendrograma que melhor representou os dados para as turmas do nono ano. Foi usado o Método da Média das Distâncias (*Average Linkage*), separado por quatro grupos. Os dendrogramas comparando cada método encontra-se no Anexo B.

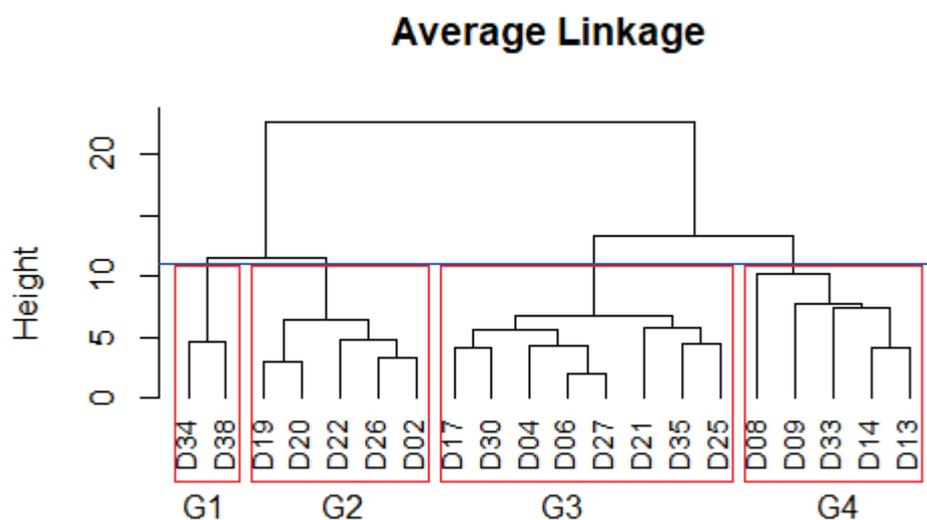


Figura 4.7: Dendrograma formado a partir da Distância Euclidiana e Método da Média das Distâncias, com quatro *cluster*- nono ano do Ensino Fundamental.

O dendrograma mostra como foi separados os quatro grupos sendo descritos na Tabela 4.7. De acordo a Figura 4.7 primeiro grupo (G1) é compostos por 2 descritores, em que o número de alunos que responderam corretamente variou de 78 a 85 (88,63% a 96,59%). O grupo dois (G2) é composto por 5 descritores, tendo 60 como menor número de resposta corretas e 72 como maior número de respostas corretas (68,18% a 81,81%). Terceiro grupo (G3) é formado pelos descritores cujo o número de acertos variou de 32 a

45 acertos (36,36% a 51,14%). O quarto grupo (G4) é formado por 5 descritores, sendo esses descritores que apresentaram o menor desempenho, tendo com número mínimo de respostas corretas 5 e o máximo 23 (5,68% a 26,14%).

Tabela 4.7: Grupos obtidos por meio da Análise de Agrupamento na Figura 4.7.

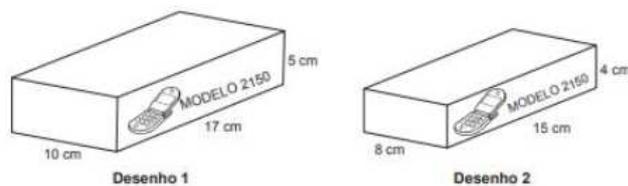
Grupo 1	Grupo 2	Grupo 3	Grupo 4
D34 Determinar a solução de um sistema de equações do 1º grau;	D19 Resolver problemas com números reais envolvendo diferentes significados das operações;	D17 Reconhecer/Identificar diferentes representações de um número racional;	D08 Resolver problemas utilizando as propriedades dos polígonos;
D38 Identificar informações apresentadas em tabelas ou diferentes tipos de gráficos;	D20 Resolver problemas com números reais envolvendo diferentes significados das operações;	D30 Resolver problemas envolvendo equações do 1º ou do 2º grau;	D09 Utilizar relações métricas do triângulo retângulo para resolver problemas;
	D22 Reconhecer números reais representados em diferentes contextos;	D04 Identificar figuras bidimensionais por meio de suas propriedades e vice-versa;	D33 Identificar a expressão algébrica que expressa uma regularidade observada em sequências de números ou figuras (padrões);
	D29 Identificar uma equação ou inequação do 1º grau que expressa um problema;	D06 Reconhecer ângulos como mudança de direção ou giros, identificando ângulos retos e não-retos;	D14 Resolver problemas envolvendo noção de volume;
	D02 Relacionar figuras tridimensionais à sua planificação ou vistas por meio de suas propriedades e vice-versa;	D27 Resolver problemas que envolvam porcentagem;	D13 Resolver problemas envolvendo o cálculo de áreas de figuras planas;
		D21 Reconhecer/Identificar diferentes representações de um número racional;	
		D35 Resolver problemas envolvendo noções de análise combinatória;	
		D25 Resolver problemas com números reais envolvendo diferentes significados das operações;	

Como apresentado na Tabela 4.7, nas questões do Grupo 4 o estudante deveria ter noções básica sobre como determinar o volume de um paralelepípedo retângulo, além

de ter uma boa interpretação do problema.

Outra questão com grande índice de erros foi do descritor 08, em que o aluno deveria ter habilidade de resolver problemas utilizando as propriedades dos polígonos. Para resolver a questão 37 o aluno precisaria saber conceitos de diagonal, para desenhá-las corretamente ou para deduzir uma maneira de calculá-la.

24) (M070054G5) Uma empresa reduziu as dimensões das caixas que são utilizadas para embalar os aparelhos celulares. Essas caixas possuem formato de paralelepípedo retângulo. Os desenhos 1 e 2 abaixo representam as dimensões da caixa antes e após a redução, respectivamente.



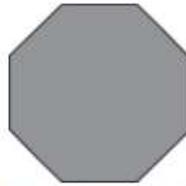
De acordo com esses desenhos, a caixa desse celular teve o seu volume reduzido em quantos centímetros cúbicos?

- A) 480 cm^3
- B) 370 cm^3 (alternativa correta)
- C) 51 cm^3
- D) 5 cm^3

Figura 4.8: Questão 24 retirada da Prova Paraná-2018, Ensino Fundamental - 9^o ano.

Fonte: (PARANÁ, 2019)

37) (SP101M) O pátio de uma escola, com o formato representado abaixo, foi decorado para a festa junina com fios de bandeirinhas em todas as suas diagonais. Para cada diagonal, foi usado apenas um fio.



Quantos fios de bandeirinhas foram usados para decorar o pátio dessa escola?

- A) 4
- B) 8
- C) 20 (alternativa correta)
- D) 40

Figura 4.9: Questão 37 retirada da Prova Paraná-2018, Ensino Fundamental - 9º ano.

Fonte:(PARANÁ, 2019)

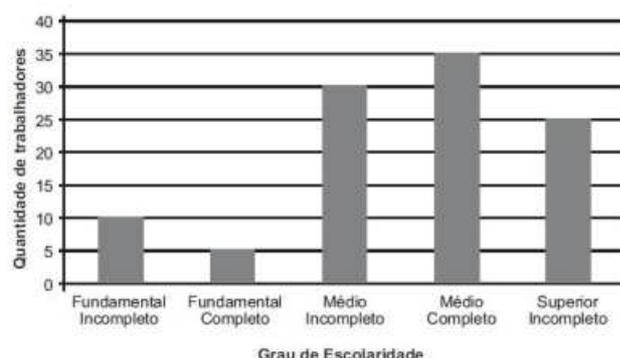
Note que as duas questões com maior dificuldade abordam conceitos de geometria e os alunos deveriam saber conceitos básicos para resolver as questões. No geral, esse conteúdo é importante, pois faz parte do cotidiano do aluno. Outros sistemas de avaliação já haviam identificado mesma dificuldade como o ENEM, SAEB e INAF. Isso nos mostra que o ensino da geometria, bem como todo o Sistema Educacional, deve ser analisado, aprofundando a investigação em busca de um diagnóstico para tomada de decisões mais eficazes (BARDIN, 2004).

De acordo com as Diretrizes Curriculares (PARANÁ (2008), p.37) “A geometria é rica em elementos que favorecem a percepção espacial e a visualização; constitui, portanto, conhecimentos relevantes, inclusive para outras disciplinas escolares”.

Como o trabalho não tem como objetivo, mostrar como se deve trabalhar esse conteúdo, mas sim demonstrar quais questões os alunos apresentaram maior dificuldades e os conteúdos associados, deixa-se esse caráter mais investigativo sobre o tema para outras discussões.

Uma questão que chamou muita atenção por ter índice alto de acerto é a questão 36 Figura 4.10. A mesma tinha como objetivo analisar a habilidade dos alunos em identificar informações apresentadas em tabelas ou diferentes tipos de gráficos. Portanto o estudante deveria observar o gráfico e verificar qual tabela o representava.

36) (M070028G5) O gráfico a seguir apresenta o grau de escolaridade dos trabalhadores de uma obra.



Qual é a tabela que apresenta os dados desse gráfico?

A)

Grau de Escolaridade	Quantidade de trabalhadores
Fundamental Incompleto	10
Fundamental	5
Médio Incompleto	30
Médio	35
Superior Incompleto	25

B)

Grau de Escolaridade	Quantidade de trabalhadores
Fundamental Incompleto	25
Fundamental	35
Médio Incompleto	30
Médio	5
Superior Incompleto	10

C)

Grau de Escolaridade	Quantidade de trabalhadores
Fundamental Incompleto	5
Fundamental	10
Médio Incompleto	25
Médio	30
Superior Incompleto	35

D)

Grau de Escolaridade	Quantidade de trabalhadores
Fundamental Incompleto	35
Fundamental	30
Médio Incompleto	25
Médio	10
Superior Incompleto	5

Figura 4.10: Questão 36 retirada da Prova Paraná-2018, Ensino Fundamental - 9º ano.

Fonte: (PARANÁ, 2019)

Note que a questão 36 não apresenta muita dificuldade, ou seja, os alunos teriam que associar os dados dispostos no gráfico com os dados mostrados na tabela. Talvez isso explique a porcentagem alta de acerto, aproximadamente 96,80%. Há outros descritores que apresentaram índice alto de alunos que resolveram de maneira correta, sendo mostrado na Tabela 4.8.

Tabela 4.8: Descritores com maior número de acertos- nono ano do Ensino Fundamental.

Descritores	Turma A	Turma B	Turma C
D38 Identificar in- formações apresentadas em tabelas ou diferentes tipos de gráficos;	94,12%	100%	96,30 %
D34 Determinar a solução de um sistema de equações do 1° grau.	88,24%	96,30%	81,48%
D19 Resolver problemas com números reais envol- vendo diferentes signifi- cados das operações	73,53%	88,89%	85,19%

4.3 Descrição do 1º ano do Ensino Médio

Analisando a Tabela 4.9 verificou-se que a média de acertos das turmas variou de 8 a 11,50 e em todas as turmas ocorreram questões em que somente dois alunos acertaram ou apenas um, sendo um total de 94 alunos. Essas questões são referentes a dois descritores: Resolver problemas envolvendo o cálculo de áreas de figuras planas (D-13) e Resolver problemas que envolvam porcentagem (D-27). Já o maior índice de acertos nas diferentes turmas (Máximo) está relacionado a área de Números e Álgebra. Todas as turmas apresentam um alto índice de coeficiente de variação que indica uma dispersão amostral em relação ao valor médio.

Tabela 4.9: Estatística descritiva dos resultados do Prova Paraná-2018, para o primeiro ano do Ensino Médio

Turma	A	B	C
Mínimo	2,00	1,00	1,00
1º quartil	5,75	6,00	4,50
Mediana	10,00	8,50	7,00
Média	11,50	10,75	8,00
3º quartil	16,25	16,50	12,50
Máximo	28,00	24,00	15,00
Desvio Padrão	6,87	6,59	4,84
Variância	49,21	43,46	23,47
Coeficiente de Variação (%)	59,75	61,33	60,56

Na Figura 4.11 é apresentado o gráfico Boxplot do posicionamento dos acertos

em relação aos quartis. Observa-se que nas três turmas, 25% das questões apresentaram menos de 5 acertos e as turmas apresentam comportamento assimétrico positivo. Nota-se também que a Turma C apresentou uma menor variância entre seus dados e menor média de acertos (Tabela 4.9).

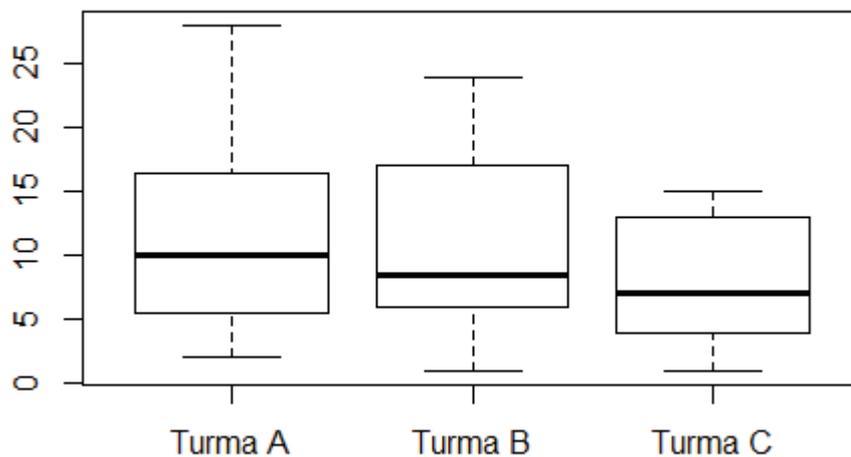


Figura 4.11: Boxplot do número de acertos na Prova Paraná do 1^o ano do Ensino Médio.

Para analisar as questões que apresentaram maiores índices de erros, a Análise de Agrupamento foi aplicada, utilizando novamente a medida de distância Euclidiana. A Tabela 4.10 apresenta os valores obtidos. Verifica-se que como ocorreu para o sexto e nono ano, o método que melhor representou a matriz original foi o Método da Média das Distâncias (*Average Linkage*).

Tabela 4.10: *CCC* (Coeficiente de Correlação Cofenético) para turmas do 1^a série do Ensino Médio.

Métodos	CCC
Método da Ligação Simples (Single Linkage)	0.8366
Método da Ligação Completa (Complete Linkage)	0.8387
Método da Média das Distâncias (<i>Average Linkage</i>)	0.8431

O dendrograma que melhor representou os dados para as turmas da primeira série foi do método da Média das Distância (*Average Linkage*) contendo quatro *cluster*, como mostra a Figura 4.12. Os dendrogramas comparando cada método encontra-se no Anexo B.

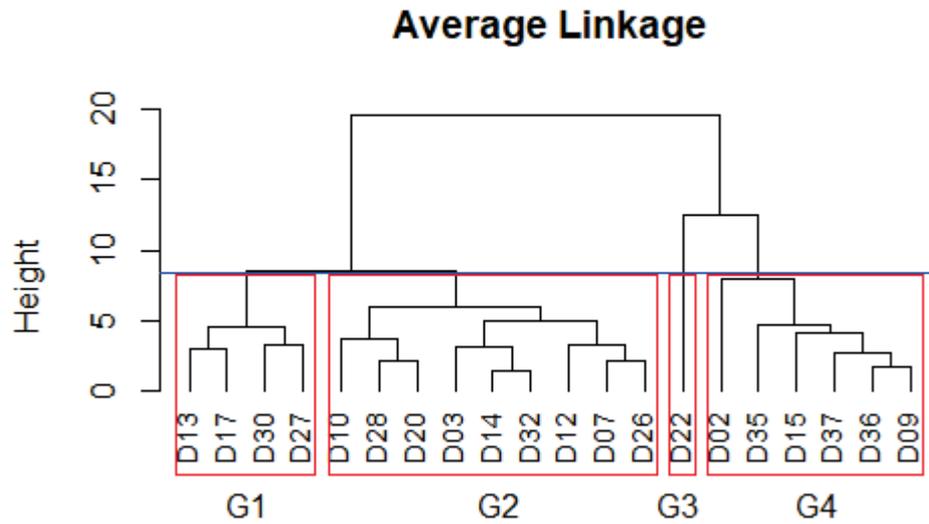


Figura 4.12: Dendrograma formado a partir da Distância Euclidiana e Método da Média das Distâncias, com quatro *cluster*- primeira série do Ensino Médio.

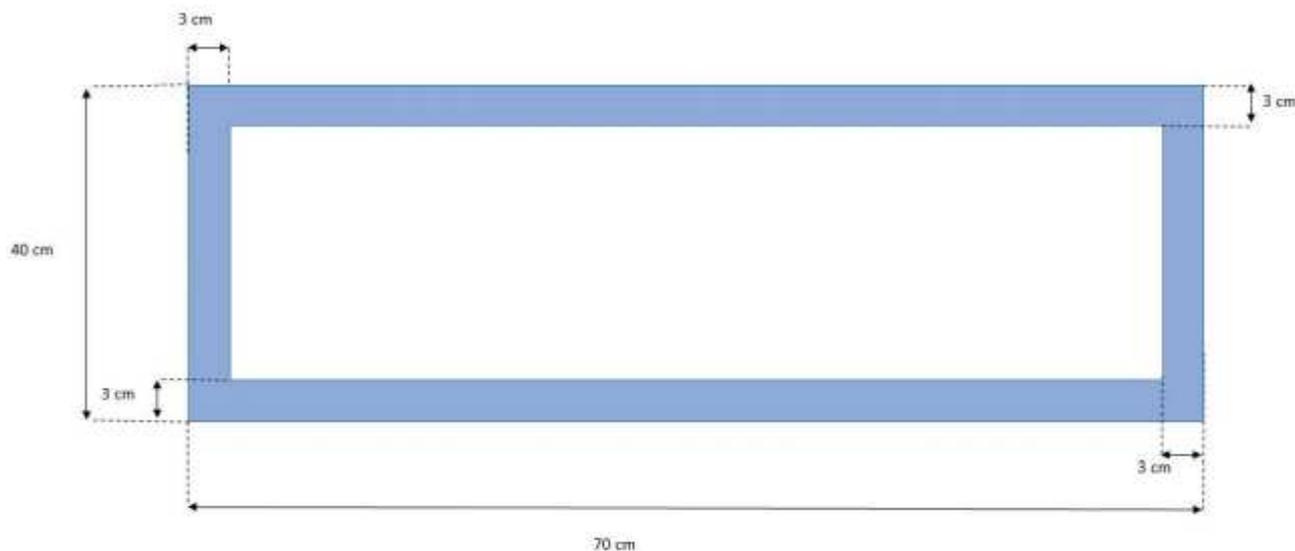
Em cada *cluster* (grupos) do dendrograma tem-se os descritores que os representam, sendo discriminados na Tabela 4.11. O primeiro grupo (G1) é formado por 4 descritores, tendo o número de respostas corretas entre 5 a 12 (5,32% a 12,77%). O segundo grupo (G2) formado por 9 descritores, onde apresentam a quantidade de respostas corretas 18 a 29 (19,15% a 30,85%). Terceiro grupo (G3), composto por apenas um descritor, sendo o de maior número de respostas corretas em toda a prova, cerca de 67 alunos responderam de maneira correta (71,28%). O último grupo (G4) é formado por 6 descritores, onde os resultado variam entre 43 a 53 (45,74% a 56,38%).

Tabela 4.11: Grupos obtidos por meio da Análise de Agrupamento da Figura 4.12.

Grupo 1	Grupo 2	Grupo 3	Grupo 4
D13 Resolver problemas envolvendo o cálculo de áreas de figuras planas;	D10 Reconhecer o seno, o cosseno e a tangente como razões entre os lados de um triângulo retângulo;	D22 Reconhecer números reais representados em diferentes contextos;	D02 Relacionar figuras tridimensionais à sua planificação ou vistas por meio de suas propriedades e vice-versa;
D17 Reconhecer/Identificar diferentes representações de um número racional;	D28 Resolver problemas que envolvam variação proporcional, direta ou inversa, entre grandezas;		D35 Resolver problemas envolvendo noções de análise combinatória;
D30 Resolver problema que envolva equação do 1 ^o grau;	D20 Resolver problema com números inteiros envolvendo as operações (adição, subtração, multiplicação, divisão, potenciação);		D15 Resolver problemas utilizando relações entre diferentes unidades de medida;
D27 Resolver problemas que envolvam porcentagem;	D03 Identificar propriedades de triângulos pela comparação de medidas de lados e ângulos;		D37 Resolver problemas envolvendo interpretação de informações apresentadas em tabelas ou diferentes tipos de gráficos;
	D14 Resolver problemas envolvendo noção de volume;		D36 Resolver problemas que envolvam noções de probabilidade;
	D32 Resolver problema que envolva equação do 2 ^o grau;		D09 Utilizar relações métricas do triângulo retângulo para resolver problemas;
	D12 Resolver problemas envolvendo o cálculo de perímetro de figuras planas;		
	D07 Reconhecer que as imagens de uma figura construída por uma transformação homotética são semelhantes, identificando propriedades e/ou medidas que se modificam ou não se alteram;		
	D22 Reconhecer números reais representados em diferentes contextos;		

De acordo com a Tabela 4.11 os descritores do grupo 1 foram os que os alunos apresentaram maior dificuldade. Dentro desse grupo se destacou o descritor 13, onde avaliaria a habilidade dos alunos em resolver problemas envolvendo o cálculo de área de figuras planas. Este conteúdo os alunos já saem com dificuldades do nono ano, como apresentado anteriormente. É possível verificar que essa dificuldade persiste na série seguinte.

23) (M100502H6) Em uma tela branca, retangular, de dimensões 70 cm e 40 cm, foi pregada uma moldura de metal cinza de 3 cm de largura para confeccionar um quadro retangular. A figura abaixo representa esse quadro já montado a partir dessa tela e dessa moldura.



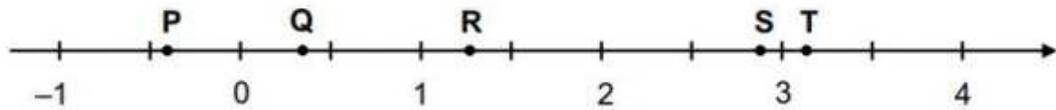
Qual é a medida, em centímetros quadrados, da região da tela branca coberta por essa moldura?

- A) 12
- B) 24
- C) 321
- D) 624 (alternativa correta)
- E) 696

Figura 4.13: Questão 23 retirada da Prova Paraná-2018, Ensino Médio - 1º ano.
fonte: (PARANÁ, 2019)

Outras três questões que chamaram atenção, são de três conteúdos diferentes apresentadas na Figura 4.14.

24) (M100060G5) Observe a reta numérica abaixo. Ela está dividida em segmentos de mesma medida.



Qual é o ponto que melhor representa a localização do número $\frac{1}{3}$ nessa reta?

- A) P.
- B) Q. (alternativa correta)
- C) R.
- D) S.
- E) T.

22) (M100054G5) Patrícia comprou um presente para sua mãe cujo valor seria dividido igualmente entre ela e seus 3 irmãos. No entanto, um dos irmãos não pôde mais contribuir e por isso Patrícia e os dois irmãos que contribuíram tiveram que pagar R\$ 40,00 a mais do que o valor combinado inicialmente.

Qual foi o valor desse presente comprado por Patrícia?

- A) R\$ 40,00
- B) R\$ 120,00
- C) R\$ 160,00
- D) R\$ 240,00
- E) R\$ 480,00 (alternativa correta)

35) (M100237H6) Na cantina de uma escola, o preço do sanduíche natural passou de R\$ 3,00 para R\$ 3,45.

A porcentagem que descreve o aumento no preço desse sanduíche é

- A) 0,15%.
- B) 0,45%.
- C) 13,00%.
- D) 15,00%. (alternativa correta)
- E) 45,00%.

Figura 4.14: Questões 24, 22 e 35 retiradas da Prova Paraná-2018, Ensino Médio - 1º ano.

Fonte: (PARANÁ, 2019)

A primeira questão trata-se do descritor 17, onde os alunos deveriam identificar a representação de um número racional na reta numérica. A segunda questão é referente ao descritor 30 e tinha como objetivo avaliar a habilidade dos alunos em resolver problemas que envolvam equação do primeiro grau. A interpretação dos dados era fundamental para resolução do problema proposto. Já a terceira questão tinha como objetivo avaliar a habilidade de resolver problemas que envolvem porcentagem e os alunos deveriam ter um pequeno domínio em regra de três simples. Todas questões citadas até aqui, exigiam a interpretação do problema proposto.

Já no grupo 3, temos os descritores onde os alunos tiveram menor dificuldade. Como mostra a Tabela 4.11, esse grupo conta apenas com o descritor D22, que tinha como objetivo avaliar a habilidade dos alunos em reconhecer números reais representados em diferentes contextos.

28) (M100418H6) Antônio tem em seu sítio 50 animais, sendo 8 cavalos, 12 vacas e 30 galinhas. A fração que representa a quantidade de cavalos em relação ao total de animais nesse sítio é

A) $\frac{1}{8}$.

B) $\frac{8}{50}$. (alternativa correta)

C) $\frac{8}{42}$.

D) $\frac{42}{50}$.

E) $\frac{50}{8}$.

Figura 4.15: Questão retirada da Prova Paraná-2018

Fonte: (PARANÁ, 2019)

Ao contrário do que aconteceu com as turmas do 6^o ano, onde os alunos apresentaram dificuldades com conteúdo e frações, o primeiro ano do Ensino Médio apresentou melhor desempenho, sendo um dos descritores com maior número de respostas corretas, cerca de 67 em um universo de 94 alunos (71,28%).

4.4 Descrição do 3^o ano do Ensino Médio

Analisando a Tabela 4.12 verifica-se que a média de acertos das turmas variou de 8,20% a 11,55%, sendo que em todas as turmas ocorreram questões em que apenas três, dois, um ou nenhum aluno conseguiu resolver (em um total de 119 alunos). Essas questões estão relacionadas com os descritores, D-23: Reconhecer a representação algébrica de uma função do 1^o grau dado o seu gráfico ou vice-versa e D30: Determinar a solução de um

sistema linear associando-o a uma matriz. O maior índice de acerto nas diferentes turmas (Máximo) está relacionada a área de Espaço e Forma. Os resultados apresentaram alto coeficiente de variação que indica altíssima dispersão em relação ao valor médio. A turma C apresentou o menor coeficiente cerca de 48,23%, acompanhado das turmas A e B com respectivamente 53,47% e 53,55% e a turma D que apresentou o maior coeficiente 54,77%.

Tabela 4.12: Estatística descritiva dos resultados da Prova Paraná-2018 para o terceiro ano do Ensino Médio

Turma	A	B	C	D
Mínimo	1,00	2,00	0,00	3,00
1 ^o quartil	4,75	6,75	6,75	6,00
Mediana	8,00	10,00	9,00	9,50
Média	8,20	11,55	9,30	9,55
3 ^o quartil	9,25	16,00	12,00	11,25
Máximo	19,00	26,00	19,00	21,00
Desvio Padrão	4,38	6,19	4,49	5,09
Variância	19,22	38,26	20,12	25,94
Coeficiente de Variação (%)	53,47	53,55	48,23	54,77

A Figura 4.16 apresenta o Boxplot do posicionamento dos acertos em relação aos quartis. Nota-se que em uma turma cerca de 25% das questões apresentaram menos de 5 acertos e nas outra turmas cerca de 25% das questões apresentaram menos de 7 acertos, esses resultados são apresentados pelo 1^o quartil. As turmas A e C apresentam comportamento assimétrico negativo e a turma B apresenta um comportamento levemente assimétrico positivo. No entanto, a turma A e B apresentaram valores discrepantes, que representa os descritores com maiores números de acertos. A turma B apresentou o maior valor para mediana (Tabela 4.12).

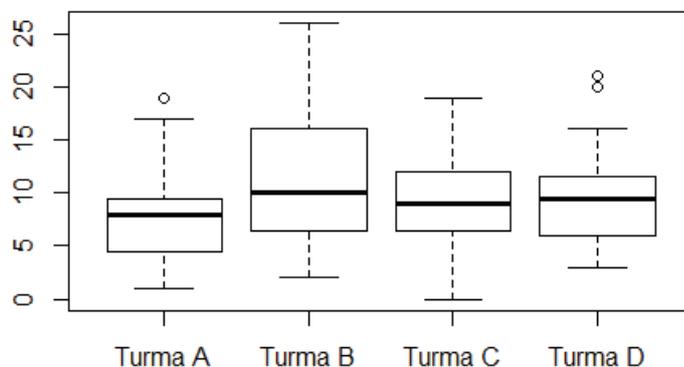


Figura 4.16: Boxplot dos resultados da 3ª série do Ensino Médio em relação a número de acertos na Prova Paraná.

A Análise de Agrupamento aplicada utilizando a medida de distância Euclidiana indicou que o melhor método que representou a matriz original foi o Método da Média das Distâncias (*Average Linkage*). A Tabela 4.13 apresenta os valores obtidos do CCC.

Tabela 4.13: *CCC*(Coeficiente de Correlação Cofenético) para turmas do 3ª série do Ensino Médio.

Métodos	CCC
Método da Ligação Simples (Single Linkage)	0.8387
Método da Ligação Completa (Complete Linkage)	0.5515
Método da Média das Distâncias (<i>Average Linkage</i>)	0.8647

O dendrograma que melhor representou os dados para as turmas da terceira série é apresentado na Figura 4.17 contendo quatro *cluster*. Os quatro grupos formados pelo dendrograma são compostos por descritores similares entre si, sendo descritos na Tabela 4.14. Os dendrogramas comparando cada método encontra-se no Anexo B.

No primeiro grupo (G1) mostra os dois descritores que obtiveram o maior índice de acerto, sendo 84 e 78. O segundo grupo (G2) é compostos pelo descritor que representou o menor índice, apenas 7 respostas corretas. Terceiro grupo (G3) composto por 4 descritores, com índice entre 43 a 47. O último grupo (G4) mostra os descritores com índice entre 20 a 39, tendo o maior número cerca de 13 descritores.

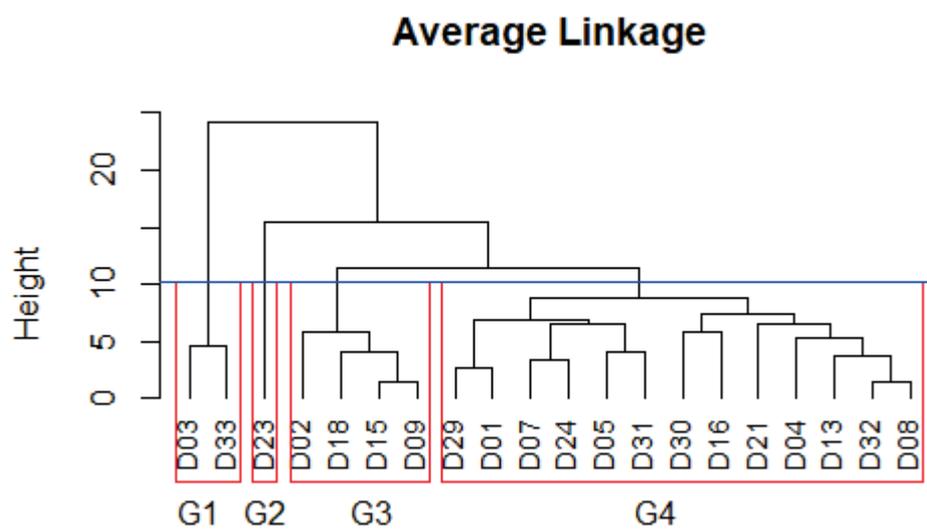


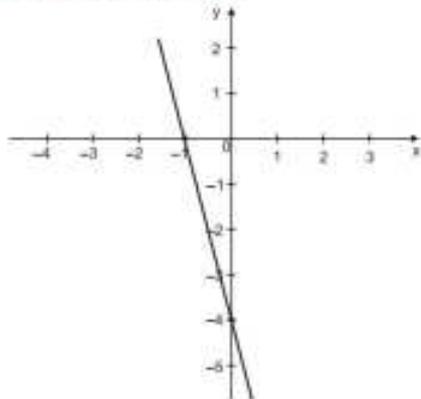
Figura 4.17: Dendrograma formado a partir da Distância Euclidiana e Método da Média das Distâncias, com quatro *cluster* - terceira série do Ensino Médio.

Tabela 4.14: Grupos obtidos por meio da Análise de Agrupamento da Figura 4.17

Grupo 1	Grupo 2	Grupo 3	Grupo 4
D03 Relacionar figuras tridimensionais à sua planificação ou vistas por meio de suas propriedades e vice-versa;	D23 Reconhecer a representação algébrica de uma função do 1º grau dado o seu gráfico ou vice-versa;	D02 Utilizar relações métricas do triângulo retângulo para resolver problemas;	D29 Identificar gráficos de funções trigonométricas (seno, cosseno, tangente) reconhecendo suas propriedades;
D33 Resolver problemas envolvendo interpretação de informações apresentadas em tabelas ou diferentes tipos de gráficos;		D18 Identificar a representação algébrica de uma função do 1º grau a partir dos dados de uma tabela	D01 Identificar figuras semelhantes mediante o reconhecimento de relações de proporcionalidade;
		D15 Resolver problemas que envolvam variação proporcional, direta ou inversa, entre grandezas;	D07 Interpretar geometricamente os coeficientes da equação de uma reta;
		D09 Relacionar a determinação do ponto de interseção de duas ou mais retas com a resolução de um sistema de equações com duas incógnitas;	D24 Reconhecer intervalos de crescimento/decrescimento, ponto(s) de máximo/mínimo e/ou zeros de funções reais representadas em um gráfico;
			D05 Resolver problemas que envolvam razões trigonométricas no triângulo retângulo;
			D31 Resolver problemas envolvendo noções de análise combinatória;
			D30 Determinar a solução de um sistema linear associando-o a uma matriz;
			D16 Resolver problemas que envolvam porcentagem;
			D21 Resolver problemas que envolvam progressões aritméticas ou geométricas;
			D04 Identificar a relação entre o número de vértices, faces e/ou arestas de poliedros expressa em um problema;
			D13 Resolver problema envolvendo a área total e/ou volume de um sólido (prisma, pirâmide, cilindro, cone, esfera);
			D32 Resolver problemas que envolvam noções de probabilidade;
			D36 Resolver problemas que envolvam noções de probabilidade;
			D08 Identificar a equação da reta apresentada a partir de dois pontos dados ou de um ponto e sua inclinação;

Como mostra Tabela 4.14, o grupo em que os alunos mostraram maior dificuldade esta representado por apenas um descritor. Sendo ele que avalia a habilidade dos estudantes em reconhecer representações algébricas de uma função do 1º grau a partir da representação gráfica. O aluno deveria lembrar que uma função do 1º grau é representada por $y = ax + b$, onde a e b são números reais.

Questão 26 – D23 – Reconhecer a representação algébrica de uma função do 1º grau dado o seu gráfico ou vice-versa.



Qual é a representação algébrica dessa função?

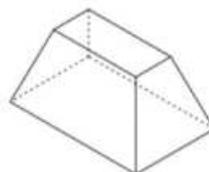
- A) $y = -4x - 4$ (alternativa correta)
- B) $y = -4x - 1$
- C) $y = -x - 4$
- D) $y = x - 4$
- E) $y = 4x - 4$

Figura 4.18: Questão 26 retirada da Prova Paraná-2018, Ensino Médio - 3º ano.
Fonte: (PARANÁ, 2019)

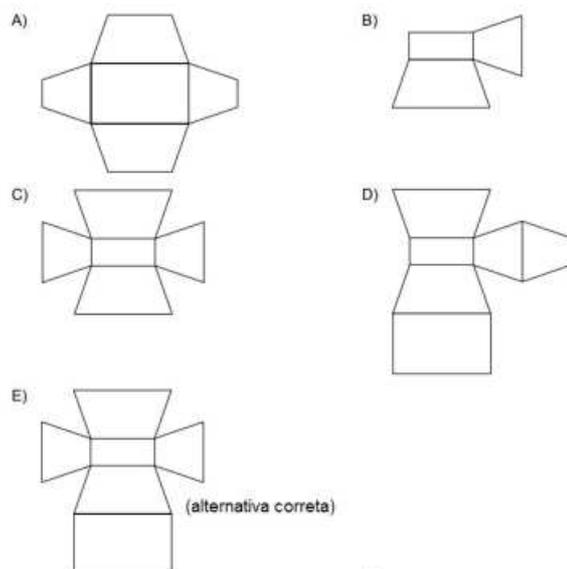
Na mesma Tabela 4.14, temos o grupo 1 que representa os descritores com maiores números de acerto, ou seja, os descritores que se mostraram com menor grau de dificuldade para os alunos (Figura 4.19).

Note que na primeira (questão 21) os alunos deveriam relacionar as figuras tridimensionais à sua planificação. Já na questão 30 os alunos deveriam resolver problemas envolvendo interpretação de informações apresentadas em tabelas ou diferentes tipos de gráficos. 84 alunos resolveram corretamente a primeira questão e 78 a segunda, em um total de 119 estudantes avaliados.

21) (M120430H6) Observe o sólido geométrico representado abaixo.



Uma das planificações desse sólido está representada em



30) (M100505H6) A tabela abaixo apresenta as três principais modalidades de atividades físicas praticadas pelos entrevistados na Pesquisa Nacional por Amostra de Domicílios (PNAD), em 2015.

Esportes e Atividades Físicas mais praticadas no Brasil – PNAD 2015			
Modalidade	Número de brasileiros (em milhões)	Homens (em %)	Mulheres (em %)
Caminhada	23,2	35,5	64,5
Futebol	16,6	94,5	5,5
“Fitness” (Ginástica)	8,2	34,4	65,6

Fonte: IBGE (Adaptado para fins didáticos)

De acordo com os dados dessa tabela, como é possível calcular a quantidade de mulheres brasileiras que praticavam a modalidade “Fitness” como uma atividade física em 2015?

- A) 34,4% de 8,2 milhões.
- B) 8,2% de 64,5 milhões.
- C) 65,6% de 8,2 milhões. (alternativa correta)
- D) 64,5% de 23,2 milhões.
- E) 65,6% de 48 milhões.

Figura 4.19: Questões 21 e 30 retiradas da Prova Paraná-2018, Ensino Médio - 3º ano.
Fonte: (PARANÁ, 2019)

5 CONSIDERAÇÕES FINAIS

Esta pesquisa é uma contribuição prática ao trabalho da gestão escolar de escolas públicas de ensino fundamental e médio e as demais escolas e gestão de sistemas. Servirá de subsídio para a reflexão crítica e a introdução de novas práticas voltadas para a tomada de decisão na escola, para a melhoria da qualidade da educação básica, além de contribuir em novos estudos na produção de conhecimentos relacionados à temática.

A análise possibilitou o estudo das dificuldades encontradas pelos alunos, mostrando os descritores com menores e maiores índices de acerto, sem perder de vista cada um dos objetivos específicos que é diagnosticar as habilidades com maiores defasagens e já apropriadas no ensino Básico do Paraná (PARANÁ, 2019) .

Foi possível verificar que os descritores com baixo desempenhos dos estudantes foram: resolver problemas com frações (apenas 10% acertos); resolver problemas que envolvam porcentagem (10% de acertos) e reconhecer diferentes representações de números racionais (11% de acertos).

A análise do sexto ano detectou algumas fragilidades. A maior dificuldade encontrada está no conteúdo de frações em que os acertos variaram de 2,22% a 16,67% . A segunda prova analisada foi do nono ano, onde as dificuldades encontradas foram nos conteúdos relacionados à Geometria, com apenas índices de 20,45% para a questão 24 que envolvia noções de volume e 5,68% para a questão 37 que fazia necessário conhecimento sobre polígonos.

A terceira série avaliada foi o primeira ano do Ensino Médio. Os alunos mostraram maior dificuldades em resolver problemas envolvendo cálculo de área de uma figura plana em que apenas 5,31% dos alunos acertaram a questão 23. Ademais, questões relacionadas a habilidade de Reconhecer/Identificar diferentes representações de um número racional, resolver problema que envolva equação do 1º grau e resolver problemas que envolvam porcentagem apresentaram 11.33%, 13.64% e 12.50 de acertos, respectivamente.

A ultima série analisada foi do terceiro ano. Porém os descritores presentes no Grupo 4, onde é composto por 13 habilidades, tiveram desempenho mais baixo. Nesse grupo os índices de acertos variaram de 16,81% a 35,29% e as habilidades requeridas eram diversas envolvendo noções de trigonometria, geometria plana e espacial, funções de 1º e 2º grau, probabilidade, porcentagem, entre outros,

As informações encontrados na análise colaboram com as informações disponibilizadas pela SEED. O resultado geral obtido pelas turmas mostra que o índice de acerto dos alunos do Ensino Médio é menor que dos alunos do Ensino Fundamental. Das 20 questões avaliadas, a média de respostas correta do Ensino Fundamental foi de 10,47, enquanto que o Ensino Médio apresentou apenas 6,46 acertos em média. Isso indica que

com passar do tempo os alunos mostraram ter maiores dificuldades com a matemática.

A análise dos dados estatísticos obtidos por meio da Prova Paraná, nos mostrou que os estudantes da escola, apresentaram as mesmas dificuldades comentadas pela SEED. Mostrando que o desempenho da maioria dos alunos na prova, deixou muito a desejar e por esse motivo a Prova Paraná precisa ser fortalecida e para esse aprimoramento é preciso que a NRE e escolas estabeleçam ações para a melhoria dos descritores.

Esta pesquisa investigou apenas uma escola pública, para comparar com os resultados disponibilizados pela SEED. No entanto o estudo não abrange toda complexidade do fenômeno estudado, por essa razão faz-se necessário novos estudos sistematizados que investiguem outras escolas e Núcleos Regionais de Educação do estado Paraná.

A Conjunto de dados avaliados

Quadro A1: Número de acertos, total de acertos e porcentagem por descritor do 6º ano do Ensino Fundamental.

Questão	Descritor	6A	6B	6C	Acertos	Porcentagem
21	D12	16	20	20	56	62,22
22	D14	21	16	17	54	60,00
23	D07	9	16	16	41	45,56
24	D20	5	3	7	15	16,67
25	D16	26	22	17	65	72,22
26	D06	10	15	6	31	34,44
27	D19	7	11	7	25	27,78
28	D10	28	27	26	81	90,00
29	D13	29	25	27	81	90,00
30	D18	16	21	16	53	58,89
31	D01	23	23	25	71	78,89
32	D24	10	8	10	28	31,11
33	D25	31	29	27	87	96,67
34	D15	25	24	25	74	82,22
35	D02	12	8	13	33	36,67
36	D23	1	1	0	2	2,22
37	D11	10	15	6	31	34,44
38	D26	27	27	29	83	92,22
39	D17	15	14	11	40	44,44
40	D05	22	20	17	59	65,56

Quadro A2: Número de acertos, total de acertos e porcentagem por descritor do 9º ano do Ensino Fundamental.

Questão	Descritor	9A	9B	9C	Acertos	Porcentagem
21	D04	15	12	14	41	46,59
22	D34	30	26	22	78	88,64
23	D21	14	18	8	40	45,45
24	D14	3	8	7	18	20,45
25	D35	10	16	13	39	44,32
26	D22	20	19	22	61	69,32
27	D33	10	3	10	23	26,14
28	D06	14	12	10	36	40,91
29	D25	14	18	13	45	51,14
30	D19	25	24	23	72	81,82
31	D13	5	6	10	21	23,86
32	D26	21	20	16	57	64,77
33	D02	20	21	19	60	68,18
34	D17	9	12	11	32	36,36
35	D27	14	14	10	38	43,18
36	D38	32	27	26	85	96,59
37	D08	2	1	2	5	5,68
38	D20	23	23	21	67	76,14
39	D09	10	8	4	22	25,00
40	D30	11	10	14	35	39,77

Quadro A3: Número de acertos, total de acertos e porcentagem por descritor do 1º ano do Ensino Médio.

Questão	Descritor	1A	1B	1C	Acertos	Porcentagem
21	D07	10	6	7	23	24,47
22	D30	5	5	2	12	12,77
23	D13	3	1	1	5	5,32
24	D17	5	2	3	10	10,64
25	D36	16	18	15	49	52,13
26	D26	10	7	5	22	23,40
27	D03	7	6	8	21	22,34
28	D22	28	24	15	67	71,28
29	D14	5	7	6	18	19,15
30	D10	9	13	7	29	30,85
31	D28	10	10	6	26	27,66
32	D12	11	6	3	20	21,28
33	D35	15	16	12	43	45,74
34	D02	22	13	15	50	53,19
35	D27	2	6	3	11	11,70
36	D37	19	19	15	53	56,38
37	D32	6	7	5	18	19,15
38	D15	18	20	11	49	52,13
39	D20	12	10	7	29	30,85
40	D09	17	19	14	50	53,19

Quadro A4: Número de acertos, total de acertos e porcentagem por descritor do 3º ano do Ensino Médio.

Descritor	3A	3B	3C	3D	Acertos	Porcentagem
D03	19	26	19	20	84	70,59
D18	8	16	10	9	43	36,13
D29	9	13	6	10	38	31,93
D05	12	10	7	6	35	29,41
D31	8	10	6	6	30	25,21
D23	1	2	0	4	7	5,88
D01	8	15	5	11	39	32,77
D15	9	16	12	13	50	42,02
D07	11	10	10	11	42	35,29
D33	17	22	18	21	78	65,55
D21	4	8	12	10	34	28,57
D04	8	4	9	3	24	20,17
D24	10	7	9	11	37	31,09
D02	9	18	14	16	57	47,90
D13	7	6	8	6	27	22,69
D30	5	8	4	3	20	16,81
D09	8	16	12	12	48	40,34
D32	3	6	9	6	24	20,17
D16	4	12	7	6	29	24,37
D08	4	6	9	7	26	21,85

B Dendrogramas

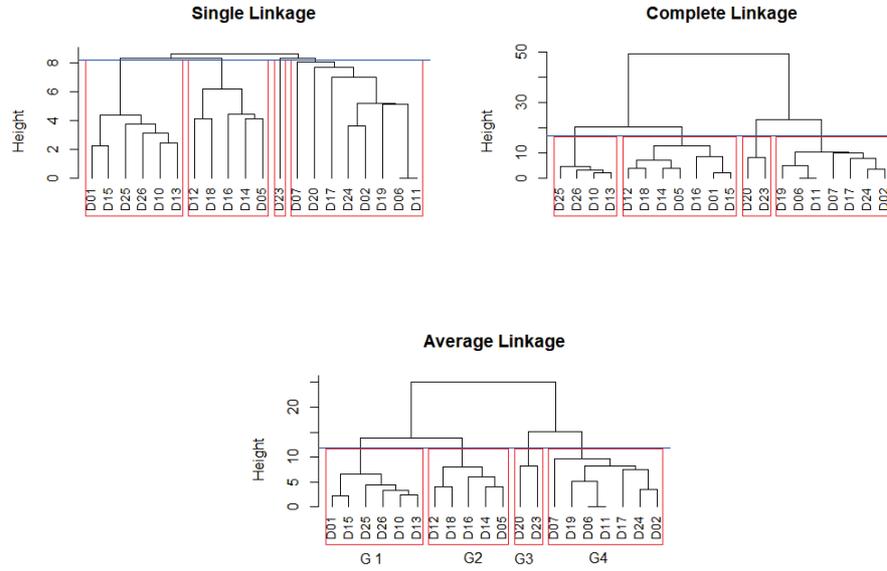


Figura B.1: Dendrogramas obtidos pelos métodos de agrupamento: Ligação simples (*Single Linkage*); Ligação Completa (*Complete Linkage*) e Média das Distâncias (*Average Linkage*) para o 6^o ano do Ensino Fundamental.

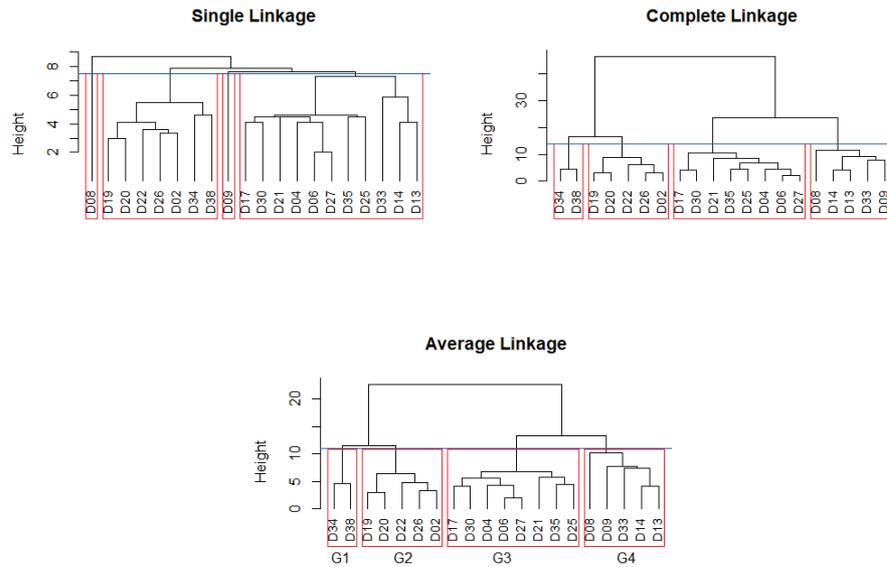


Figura B.2: Dendrogramas obtidos pelos métodos de agrupamento: Ligação simples (*Single Linkage*); Ligação Completa (*Complete Linkage*) e Média das Distâncias (*Average Linkage*) para o 9^o ano do Ensino Fundamental.

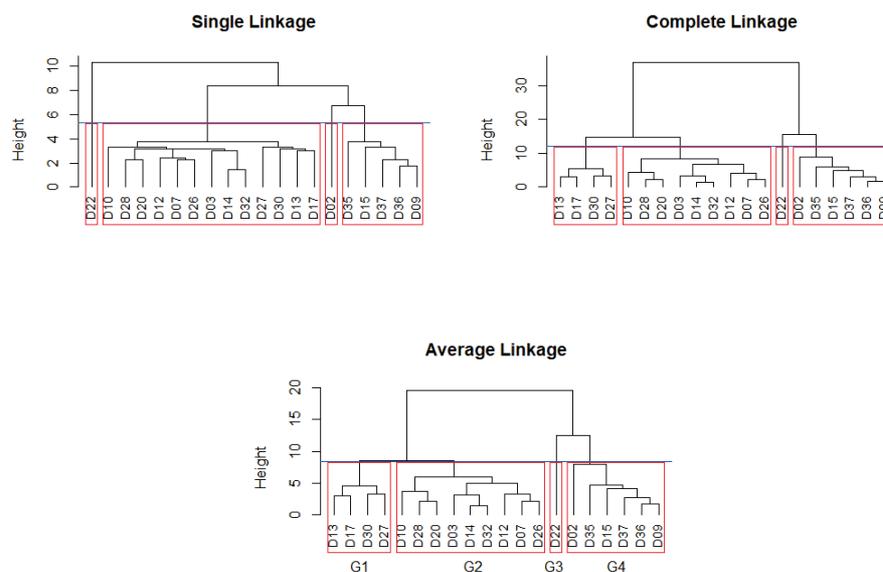


Figura B.3: Dendrogramas obtidos pelos métodos de agrupamento: Ligação simples (*Single Linkage*); Ligação Completa (*Complete Linkage*) e Média das Distâncias (*Average Linkage*) para o 1º ano do Médio.

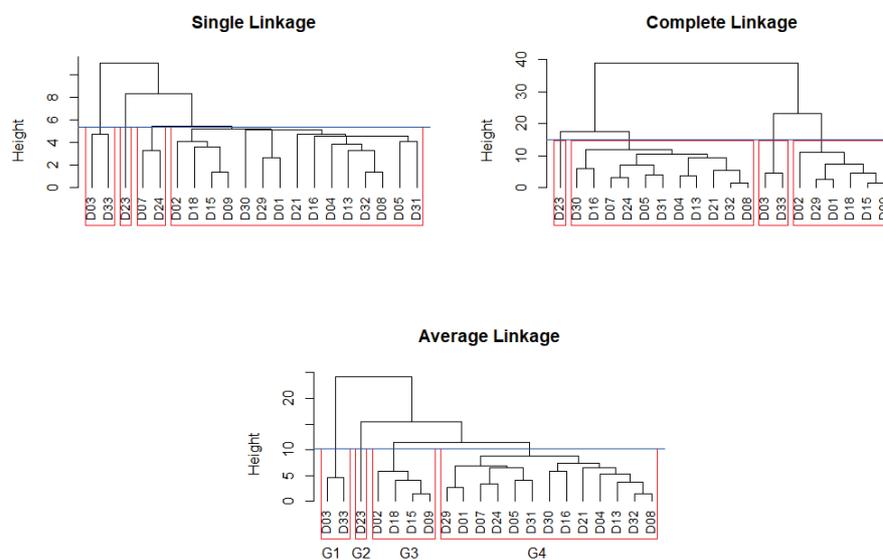


Figura B.4: Dendrogramas obtidos pelos métodos de agrupamento: Ligação simples (*Single Linkage*); Ligação Completa (*Complete Linkage*) e Média das Distâncias (*Average Linkage*) para o 3º ano do Ensino Médio.

C Rotina R

Análise de dados do prova Paraná
Roteiro de análise de agrupamento
Regiane Slongo Fagundes
Juliano Rodrigues da Silva

Criado em: 06/09/2019

Atualização: 10/11/2019

AGRUPAMENTO

** lendo um conjunto de dados em txt, header = T tem nome das colunas na 1ª linha

```
dados= read.table("sexto1.txt", header = T, row.names = 1)
```

```
summary(dados)
```

```
m1=mean(dados[,1])
```

```
m2=mean(dados[,2])
```

```
m3=mean(dados[,3])
```

```
v1=var(dados[,1])
```

```
v2=var(dados[,2])
```

```
v3=var(dados[,3])
```

```
sd1=sd(dados[,1])
```

```
sd2=sd(dados[,2])
```

```
sd3=sd(dados[,3])
```

```
cv1=sd1/m1*100
```

```
cv2=sd2/m2*100
```

```
cv3=sd3/m3*100
```

```
boxplot(dados)
```

** mostra os valores de junção no dendograma para o primeiro cálculo das distâncias Euclidianas

```
disteuc = dist(dados) distancia euclidiana disteuc
```

Método de ligação completa (Complete linkage) Medida de similaridade entre dois clusters é definida pela maior distância de qualquer ponto do 1º cluster para qualquer ponto do 2º cluster.

```

*****
    descritores.hc.complete =hclust(disteuc, method="complete")
    descritores.hc.complete $height
**Plotando o dendograma
    plot(descritores.hc.complete, main="Complete Linkage", xlab=, sub=, cex=.9,
hang=-1)
** definição do ponto de corte em k=??? Critério da amplitude das distâncias **
    rect.hclust(descritores.hc.complete, k = 4)
    abline(h=16.8, col="blue")
** coeficiente de correlação cofenética
    d1 = cophenetic(descritores.hc.complete)
    cor(disteuc,d1)
    cor.test(disteuc,d1)
*****

Método de ligação simples (Single linkage). Medida de similaridade entre dois clusters é
definida pela menor distância de qualquer ponto do 1º cluster para qualquer ponto do 2º
cluster.
*****

** Comandos que executam o algoritmo HC com distância Euclidiana e os tipos de linkage
    descritores.hc.single = hclust(disteuc, method="single")
    descritores.hc.single $height
**Plotando o dendograma plot(descritores.hc.single, main="Single Linkage", xlab=, sub=
, cex=.9, hang=-1)
** definição do ponto de corte em k=??? Critério da amplitude das distâncias **
    rect.hclust(descritores.hc.single, k = 4)
    abline(h=8.2, col="blue")
** coeficiente de correlação cofenética
    d3 = cophenetic(descritores.hc.single)
    cor(disteuc,d3)
    cor.test(disteuc,d3)
*****

Método
da média das distâncias (Average linkage). Medida de similaridade entre dois clusters é
definida pela média das distâncias de todos os pontos do 1º cluster em relação aos pontos
do 2º cluster.
*****

**Comandos que executam o algoritmo HC com distância Euclidiana e os tipos de linkage
    descritores.hc.average = hclust(disteuc, method="average")
    descritores.hc.average $height
** Plotando o dendograma plot(descritores.hc.average, main="Average Linkage", xlab=

```

```
, sub=, cex=.9, hang=-1)
** definição do ponto de corte em k=??? Critério da amplitude das distâncias **
    rect.hclust(descritores.hc.average, k = 4)
    abline(h=11.8, col="blue")
** coeficiente de correlação cofenética
    d2 = cophenetic(descritores.hc.average)
    cor(disteuc,d2)
    cor.test(disteuc,d2)
```

REFERÊNCIAS

- ANDERBERG, M. R. *Cluster analysis for applications: probability and mathematical statistics: a series of monographs and textbooks*. [S.l.]: Academic press, 2014. v. 19.
- BARDIN, L. Qualidade da educação: uma nova leitura do desempenho dos estudantes da 3ª série do ensino médio. *Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Brasília. Brasília: MEC/INEP, 2004.*
- BARROSO, L. P.; ARTES, R. Análise multivariada. *Lavras: Ufla*, p. 151, 2003.
- BUSSAB, W. d. O.; MIAZAKI, E. S.; ANDRADE, D. F. d. Introdução a análise de agrupamento. *Sao Paulo: USP-IME — IX SIMPÓSIO BRASILEIRO DE PROBABILIDADE E ESTATÍSTICA, 1990.*
- CRUZ, C. D.; CARNEIRO, P. C. S. Modelos biométricos aplicados ao melhoramento genético. *Viçosa: UFV*, p. 585, 2006.
- CUNHA, W.; BISCHOFF, W. Resultado da primeira edição da prova paraná mostra problemas no ensino de matemática. *G1- Paraná — Paraná RPC*, Curitiba, 2019. Disponível em: (<https://g1.globo.com/pr/parana/noticia/2019/04/17/resultado-da-primeira-edicao-da-prova-parana-mostra-problemas-no-ensino-de-matematica.ghml>). Acesso em: 10 nov. 2019.
- DIAS, A. *Seleção multivariada e identidade de modelos não lineares para o crescimento e acúmulo de nutrientes em frutos de mangueira. 2014*. Tese (Doutorado) — Tese (Doutorado em Estatística e Experimentação Agropecuária) – Coordenação do Programa de Pós-Graduação de Estatística e Experimentação Agropecuária. Universidade Federal de Lavras. Lavras, 2014.
- FERREIRA, D. F. Estatística multivariada. In: . [S.l.]: Lavras:Editora UFLA, 2011.
- JOHNSON, R. A.; WICHERN, D. W. *Applied Multivariate Statistical Analysis*. 5. ed. [S.l.]: ed. Englewood Cliffs: Prentice-Hall, 2002.
- JUNIOR, A. A.; THIEBAUT, J. Análise multivariada na avaliação da diversidade em recursos genéticos vegetais. *Campos dos Goytacazes-Universidade Estadual do Norte Fluminense-UENF, CCTA*, 1999.
- KHATTREE, R.; NAIK, D. N. Multivariate data reduction and discrimination with sas software. *Sas Institute Cary*, p. 574, 2000.
- MARDIA, K. V.; KENTK, J. T.; BIBBY, J. M. Multivariate analysis. London: Academic Press, p. 518, 1995.
- MINGOTI, S. A. *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. [S.l.]: Belo Horizonte: UFMG, 2005. 295 p.

- OKSANEN, J.; BLANCHET, F. G.; FRIENDLY, M.; KINDT, R.; MCGLINN, D.; MINCHIN, P. R.; O'HARA, R. B.; SOLYMOS, P.; STEVENS, M. H. H.; SZOECS, E.; WAGNER, H. *vegan: Community Ecology Package. R package version 2.4-0*. 2016. Disponível em: <https://CRAN.R-project.org/package=vegan>.
- PARANÁ, G. D. Diretrizes curriculares da educação básica matemática. *Secretaria de estado da educação do paraná*, 2008.
- PARANÁ, G. D. *Prova Paraná Avaliação Diagnóstica: Objetivos*. 2019. Disponível em: <http://www.provaparana.pr.gov.br/Pagina/Objetivos>.
- POLESE, F. O. et al. Análise de uma proposta construtivista de ensino de frações por meio da resolução de problemas. Pontifícia Universidade Católica do Rio Grande do Sul, 2011.
- PRADO, B. Q. de M. Análise de agrupamentos das taxas de incidência de dengue nos estados brasileiros. Universidade Federal de Uberlândia, 2015.
- QUINTAL, G. M. d. C. C. *Análise de clusters aplicada ao Sucesso/Insucesso em Matemática*. Tese (Doutorado) — Universidade da Madeira, 2006.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2019. Disponível em: <https://www.R-project.org/>.
- RENCHEER, A. C. *Methods of multivariate analysis*. 2. ed. [S.l.]: New York: John Wiley, 2002. 708 p.
- ROHLF, F. J. Adaptive hierarchical clustering schemes. In: _____. [S.l.]: Systematic Zoology, 1970. v. 18, cap. ., p. 58–82.
- SNEATH, P. H. A.; SOKAL, R. R. *Numerical taxonomy*. [S.l.]: New York: Hafner, 1975.
- VALLI, M. Análise de cluster. *Augusto Guzzo Revista Acadêmica*, n. 4, p. 77–87, 2002.