

**UNIVERSIDADE FEDERAL DO VALE DO SÃO FRANCISCO -
UNIVASF
COORDENAÇÃO DE PÓS-GRADUAÇÃO E EXTENSÃO
DISSERTAÇÃO - PROFMAT
EDILSON FRANCO DA SILVA JÚNIOR**

**UMA ANÁLISE MULTIVARIADA DO SUCESSO OU FRACASSO EM
MATEMÁTICA DOS ALUNOS DO 8º ANO DO ENSINO FUNDAMENTAL**

Juazeiro-Ba
2014

EDILSON FRANCO DA SILVA JÚNIOR

**UMA ANÁLISE MULTIVARIADA DO SUCESSO OU FRACASSO EM
MATEMÁTICA DOS ALUNOS DO 8º ANO DO ENSINO FUNDAMENTAL**

Dissertação apresentada ao Programa de Pós-Graduação PROFMAT (Mestrado Profissional em matemática em Rede Nacional) na Universidade Federal do Vale do São Francisco, como requisito parcial para a obtenção do grau de Mestre na área de Matemática.

Orientador: Prof Dennis Marinho Oliveira Ramalho de Souza

Co-orientadora: Prof^a Michele Rodrigues de Albuquerque

Juazeiro-Ba
2014

	Silva Júnior, E. F. da
S586a	Uma Análise Multivariada do sucesso ou fracasso em Matemática dos alunos do 8º ano do Ensino Fundamental / Edilson Franco da Silva Júnior . -- Juazeiro, 2014. ix ; 53f. 29 cm.
	Trabalho de Conclusão de Curso (Pós Graduação em Matemática) - Universidade Federal do Vale do São Francisco, Campus Juazeiro, Juazeiro, 2014. Orientador: Prof. Me. Dennis Marinho Oliveira Ramalho de Souza. Co-orientadora : Profª Me. Michele Rodrigues de Albuquerque
	Referências.
	1. Rendimento escolar. 2. Correlação cofenética. 3. Distância de Jaccard. I. Título. II. Souza. Dennis Marinho Oliveira Ramalho de III. Universidade Federal do Vale do São Francisco.
	CDD 510.7

Ficha catalográfica elaborada pelo Sistema Integrado de Biblioteca SIBI/UNIVASF



Universidade Federal do Vale do São Francisco
Mestrado profissional em Matemática em Rede Nacional
PROFMAT/UNIVASF



**UMA ANÁLISE MULTIVARIADA DO SUCESSO OU FRACASSO
EM MATEMÁTICA DOS ALUNOS DO 9º ANO DO ENSINO
FUNDAMENTAL**

Por:

EDILSON FRANCO DA SILVA JÚNIOR

Dissertação aprovada em 1º de outubro de 2014.

Orientador: Me. Dennis Marinho Oliveira Ramalho de Souza
(PROFMAT – UNIVASF)

Prof. Dr. Marcelo Silva de Souza Ribeiro
(Colegiado de Psicologia - UNIVASF)

Prof. Dr. Paulo José Pereira
(Colegiado de Engenharia de Produção - UNIVASF)

Juazeiro
2014

AGRADECIMENTOS

Aos meus pais (in memorian) pelo estímulo e apoio ao longo de toda minha vida acadêmica.

Aos professores do PROFMAT pela amizade e dedicação.

Aos colegas e amigos do curso pelo companheirismo ao longo da jornada.

À minha esposa Cristiane e aos meus filhos Edilson Neto, Eduardo, Ethiane e Ellen, que tanto me estimularam nas noites de sono perdidas.

Ao meu orientador Dennis, pela enorme contribuição na confecção deste trabalho.

Aos professores das escolas João Batista dos Santos e Conselheiro Luiz Viana, que desde o início apoiaram este projeto.

Por fim, à todos que direta ou indiretamente contribuíram para a minha vida acadêmica e profissional, os meus sinceros agradecimentos.

Muito Obrigado!!!

A Matemática não é algo mágico e ameaçadoramente estranho, mas sim um corpo do conhecimento naturalmente desenvolvido por pessoas durante um período de 5000 anos.

(Artur Frank Swetz)

RESUMO

O rendimento escolar em Matemática do aluno do 8º ano do Ensino Fundamental depende de inúmeros fatores pertinentes a vida que o envolve. São fatores familiares, sociais, psicológicos, afetivos ou outros que, diretamente ou indiretamente, contribuem para o seu sucesso ou insucesso. Desse modo, este trabalho apresenta resultados de um estudo estatístico de agrupamento, considerando 32 variáveis e uma amostra de 440 alunos catalogados em quatro escolas da rede pública de ensino, com o objetivo de verificar a influência dessas variáveis no sucesso ou fracasso nas notas finais da disciplina de matemática. Uma correlação cofenética de 97% e 91%, respectivamente, gerou o grau de proximidade das variáveis pela distância de Jaccard, tanto para sucesso como para o fracasso em Matemática. Concluiu-se que trabalhos envolvendo alunos da Educação Básica exigem uma amostra piloto para correções dos possíveis vieses. Além disso, o acesso as tecnologias, a necessidade de repensar o uso do celular como instrumento pedagógico, o incentivo do retorno dos pais ao estudo e a presença de assistente social na escola possibilitam um maior rendimento dos alunos na disciplina de matemática.

Palavras-chaves: Rendimento Escolar; Correlação Cofenética; Distância de Jaccard.

ABSTRACT

The academic performance of the student in Mathematics on 8th grade of elementary school depends on numerous relevant factors life that surrounds you. Are familial, social, psychological, emotional or other factors that directly or indirectly contribute to its success or failure. Thus, this paper presents results of a statistical study of grouping, considering 32 variables and a sample of 440 students in four schools cataloged the state education system, in order to check the influence of these variables on the success or failure of the final grades discipline of mathematics. A cophenetic correlation of 97% e 91%, respectively, generated the degree of proximity of the variables by Jaccard distance, both for success or failure in mathematics. It was concluded that studies involving basic education students require a pilot sample for correlations of possible biases. In addition, access technologies, the need to rethink the use of the use of cellphones as a pedagogical tool, the encouragement of parents return to study and the presence of a social worker at the school allow greater student achievement in the discipline of mathematics.

Palavras-chaves: Academic Performance; Cophenetic Correlation; Jaccard Distance.

LISTA DE ILUSTRAÇÕES

Gráfico 1 – relação entre número de homens e mulheres pesquisados.

Gráfico 2 – distribuição da faixa etária identificada nas escolas.

Gráfico 3 – autoafirmação e catalogação das raças declaradas.

Gráfico 4 – religiões declaradas pelos alunos pesquisados.

Gráfico 5 – quantidade dos sucessos em matemática no 8º ano do E. F.

Gráfico 6 – desempenho das escolas verificando as notas obtidas.

Gráfico 7 – agrupamento das variáveis correlacionadas por sucesso ou fracasso.

LISTA DE ABREVIATURAS E SIGLAS

FT-IR – microespectroscopia infravermelha.

HCA – análise de agrupamento hierárquico.

IBGE- instituto brasileiro de geografia e estatística

IDEP – índice de desenvolvimento da educação básica.

INEP – instituto nacional de estudos e pesquisas educacionais.

MDS – ministério da saúde.

MEC – ministério da educação.

OCDE – organização de cooperação para o desenvolvimento econômico.

PCA- análise por componentes principais.

SUMÁRIO

INTRODUÇÃO.....	10
2 REVISÃO DA LITERATURA.....	13
2.1 Análise de agrupamentos.....	16
2.2 Técnicas de análise de agrupamentos.....	18
2.2.1 TÉCNICAS DE HIERARQUIZAÇÃO.....	19
2.2.2 Definição do número de grupos.....	20
2.2.3 Análise de agrupamento hierárquico.....	21
2.3 Medidas de distância.....	22
2.4 Resultados de trabalhos científicos.....	27
3 METODOLOGIA.....	30
4 DEMONSTRAÇÃO E ANÁLISE DA PESQUISA.....	32
5 CONSIDERAÇÕES FINAIS.....	44
REFERÊNCIAS.....	45
APÊNDICE.....	51

INTRODUÇÃO

O 9º ano do Ensino Fundamental é um ano de grande importância para a formação intelectual do aluno, por esta razão é dada muita ênfase e trabalha-se mais em sala de aula com um objetivo de preparar o aluno para o ensino médio.

Este trabalho foi feito destinado ao aluno do 9º ano do ensino fundamental, mas tomando como base o desempenho do 8º ano, pois todas as verificações de aprendizagem referem-se ao ano de 2013.

Indicadores referentes à educação, no Brasil, como taxa de analfabetismo, taxa de atendimento escolar, taxa de escolarização líquida e, em decorrência, o nível de escolaridade média da população, melhoraram muito nos últimos 40 anos.

Segundo Afonso e Aubyn (2012), um sistema educacional é eficiente, se existe baixa distorção idade-série, alta taxa de aprovação e baixa de reprovação e evasão, e proficiência em disciplinas-chave do currículo.

O Brasil melhorou a qualidade do ensino de matemática, em torno de 4% nos últimos dois anos, mas ainda ocupa as últimas colocações na principal avaliação internacional sobre o desempenho dos alunos, feita pela Organização de Cooperação para o Desenvolvimento Econômico em 2013 (OCDE, 2013).

Preocupado com os dados educacionais apresentados pelas escolas brasileiras, o Instituto Nacional de Estudos e Pesquisas Educacionais (INEP), criou o Índice de Desenvolvimento da Educação Básica (IDEB), que é um indicador de qualidade educacional. Este índice combina informações do desempenho dos estudantes ao final das três etapas de estudo, como o 5º e 9º ano do ensino fundamental e o 3º ano do ensino médio, através de exames padronizados, como a Prova Brasil e o Saeb, somando informações do rendimento escolar (aprovação) dos alunos (MEC, 2007)

Estudos e análises sobre qualidade educacional raramente combinam as informações produzidas por esses dois tipos de indicadores, ainda que a complementaridade entre elas seja evidente. Um sistema educacional que reprova sistematicamente seus estudantes, fazendo com que grande parte deles abandone a escola antes de completar a educação básica, não é desejável, mesmo que aqueles que concluem essa etapa de ensino atinjam elevadas pontuações nos exames padronizados (LIMA, 2003).

Dessa forma, o IDEB foi desenvolvido para ser um indicador que sintetize informações de desempenho em exames padronizados com informações sobre rendimento escolar (taxa média de aprovação dos estudantes na etapa de ensino) e também sobre indicadores sociais que identificam o indivíduo no seu meio, sua família e as suas relações com a escola.

Indicadores educacionais como o IDEB são desejáveis por permitirem o monitoramento do sistema de ensino do país, como detectar escolas com baixo rendimento, e monitorar a evolução temporal do desempenho dos alunos.

A fórmula do IDEB leva em consideração a média da proficiência nas disciplinas de português e matemática através do Saeb e Prova Brasil, e também outro indicador do rendimento baseado na taxa de aprovação da etapa de ensino dos alunos.

Por esta razão, o IDEB não pode ser levado consideração como único componente para estudo do êxito ou não dos alunos, uma vez que avalia o aluno apenas nas disciplinas de português e matemática, deixando de lado os demais componentes curriculares.

Ao analisar o baixo rendimento, contribuindo para o sucesso ou fracasso do aluno, Bernadete Gatti (2012) analisou o comportamento dos professores e afirmou que os cursos universitários não preparam o professor para a realidade das escolas, sobretudo às públicas e não há uma preocupação do governo federal em reestruturar as disciplinas pedagógicas voltadas para o professor em sala de aula.

A revista Nova Escola em março de 2012 publicou que a partir de um estudo estatístico recente era necessário definir fatores determinantes para definir o sucesso ou não do aluno de disciplinas essenciais, como no caso, matemática. Daí, criar uma técnica que permita a análise de dados era primordial.

O objetivo era organizar dados observados em estruturas que façam algum sentido, ou desenvolver taxonomias capazes de classificar dados observados em diferentes classes, uma ideia utilizada primeiramente por Tyron (1939) com intenção de agrupar uma variedade de algoritmos de classificação diferentes, todos voltados para várias áreas de pesquisa.

O nome dado para esse tipo de estrutura foi de análise de agrupamentos ou análise de conglomerados, que consiste basicamente em agrupar dados de acordo com as similaridades entre eles, utilizando a estatística para formar grupos com

homogeneidade dentro do agrupamento e heterogeneidade entre eles (JOHNSON & WICHERN, 1992; CRUZ & REGAZZI, 1994).

O objetivo desse estudo foi aplicar uma técnica multivariada de análise de agrupamento, do tipo hierárquico, para avaliar as variáveis que influenciam no desempenho em matemática de alunos do 9º ano do ensino fundamental II, em quatro escolas da rede pública de ensino, sendo três na cidade de Petrolina-PE e uma na cidade de Casa Nova-BA, escolhidas devido a conveniência do pesquisador.

2. REVISÃO DA LITERATURA

Existem muitos trabalhos de análise multivariadas em diversas áreas de atuação, como a indústria química, a geografia que estuda o leito dos rios, a biologia com os seus micro-organismos, a administração com os recursos humanos de uma empresa e a matemática, que acumula dados da estatística para explicar os diversos casos de probabilidade, análise combinatória, na álgebra com o desenvolvimento das equações, como também na geometria das formas geométricas.

Qualquer que seja a decisão que se tome, sempre leva-se em consideração um grande número de fatores. Obviamente, nem todos pesam da mesma maneira na hora de uma escolha. Às vezes, a decisão é feita por intuição, sem identificar uma maneira sistemática para os fatores. Ou seja, não se identifica quais as variáveis que afetaram a decisão.

Segundo Moita (2014), estabelecer relações, encontrar ou propor leis explicativas é o papel próprio da ciência. Então, é necessário controlar, manipular, medir as variáveis que são consideradas relevantes ao entendimento do fenômeno analisado. Em todas as áreas de atuação, isto acontece, independente do trabalho ser na área de biologia, química, geografia, administração, sendo que a maior delas é de natureza epistemológica, em que a ciência não conhece a realidade, apenas a representa através de modelos e teorias dos diversos campos do conhecimento. Outra dificuldade é a universalização das explicações científicas. Isto implica e condiciona a pesquisa a uma padronização metodológica. Para isto, existe um aspecto essencial desta padronização, que é a avaliação estatística das informações.

A ciência analisa as variáveis isoladamente e a partir daí faz inferências sobre a realidade. Uma simplificação que pode trazer vantagens e desvantagens. De acordo com Moura (2006), um fenômeno depende de inúmeras variáveis, a análise falha, o mesmo não é suficiente conhecer informações estatísticas isoladas, mas também a totalidade destas informações obtidas pelo conjunto de variáveis.

Quando as relações existentes entre as variáveis não são percebidas, geram efeitos não esperados que complicam a interpretação do fenômeno. Agora no caso restrito de variáveis independentes entre si é possível, com razoável segurança,

interpretar um fenômeno complexo usando as informações estatísticas de poucas variáveis. As informações estatísticas de maior relevância neste tipo de análise são as medidas de tendência central e de dispersão dos dados.

Segundo Ayres (2012), em seu artigo sobre os resultados da pesquisa multivariada, o desenvolvimento tecnológico vindo das descobertas científicas tem aumentado o próprio desenvolvimento científico, ampliando em várias ordens de grandeza a capacidade de obter informações de acontecimentos e fenômenos que estão sendo analisados. Uma grande massa de informação deve ser processada antes de ser transformada em conhecimento. Dessa forma, cada vez mais necessita-se de ferramentas estatísticas que forneçam uma visão específica de determinado fenômeno utilizando múltiplas variáveis. Assim, análise multivariada corresponde a um grande número de métodos e técnicas que utilizam simultaneamente múltiplas variáveis na interpretação teórica do conjunto de dados obtidos.

Para que não haja qualquer mistificação dos métodos de análise multivariada, a estatística é pouca explorada nos artigos científicos, quando se compara uma análise significativa de um universo calculável de variáveis e condições de amostragem. Os diversos trabalhos sempre são empregados com subjetividade.

Moita (2014) fala do paradoxo interessante, que é a complexidade matemática que pode ser substituída por uma simplicidade didática. Através do uso do software estatístico, é possível realizar diversas pesquisas utilizando os métodos e técnicas estatísticos sem exigir do pesquisador um conhecimento específico na área. Porém, a facilidade no uso desses softwares para a realização de pesquisas esbarra em dois problemas. O primeiro refere-se à dificuldade dos pesquisadores em escolher qual o melhor método estatístico a ser aplicado em sua pesquisa. E o segundo consiste na dificuldade do usuário em interpretar a representatividade de cada símbolo, que variam de software para software, daí a necessidade de se trabalhar sempre com o mesmo software e o que apresenta as bibliotecas mais completas.

Ainda Moita (2014) enfatiza que ninguém faz ciência sem expectativa, e que esta surge do conhecimento teórico e do senso comum. A pesquisa científica consiste em traduzir uma expectativa em problema, e a partir daí chega-se a uma proposta de trabalho para, em seguida, escolher o procedimento metodológico

adequado. A estatística é parte constitutiva deste procedimento metodológico, que consiste, no início, na escolha da amostragem e seleção das variáveis, e, no final, com o tratamento, análise e inferência sobre os dados.

Os métodos multivariados possibilitam diversas finalidades e o primeiro passo é saber que conhecimento se pretende gerar, ou seja, o que se pretende afirmar a respeito dos dados. Exemplificando, esta diversidade é quando o interesse é quando as amostras se relacionam, ou seja, são semelhantes segundo as variáveis utilizadas no trabalho existem dois métodos principais: o HCA que é análise por agrupamento hierárquico e PCA que é análise por componentes principais.

Quando a finalidade principal é fazer previsão, segundo Moita (2014) temos muitas variáveis independentes e se pode encontrar uma variável dependente, a regressão linear múltipla e redes neurais são métodos indicados para esta situação. Com uma finalidade bem diversa, existem métodos de análise multivariada que podem ser usados na etapa primeira de uma pesquisa, na própria escolha das variáveis que vai descrever o sistema. É muito comum nos casos em que um processo necessita ser otimizado, como exemplos o simplex e o planejamento fatorial.

Moita (2014) continua afirmando que os métodos estatísticos são escolhidos de acordo com os objetivos da pesquisa, por isto, mostrar, predizer ou otimizar são obtidos por diferentes métodos. Portanto, a estatística multivariada, com os seus diferentes métodos difere de uma prateleira de supermercado abarrotada de produtos com a mesma função, pois cada método tem sua fundamentação teórica e sua faixa de aplicabilidade.

Pode-se dizer que em quase todas as áreas de pesquisa são mensuradas inúmeras variáveis e, em geral, essas devem ser analisadas conjuntamente. A análise multivariada é a área da estatística que trata desse tipo de estudo e existem várias técnicas que podem ser aplicadas, sendo que, a utilização dessas depende do tipo de dado que se deseja analisar e dos objetivos do estudo.

Segundo Anderson (1984), existem, basicamente, duas formas de classificar as análises multivariadas: as que permitem extrair informações a respeito da independência entre as variáveis que caracterizam cada elemento, tais como análise fatorial, análise de agrupamento, análise canônica, análise de ordenamento multidimensional e análise de componentes principais; e as que permitem extrair

informações a respeito da dependência entre uma ou mais variáveis ou uma com relação à outra, tais como a análise de regressão multivariada, análise de contingência múltipla, análise discriminante e análise de variância multivariada.

2.1 Análise de Agrupamento

A análise de agrupamento é uma técnica multivariada que tem por objetivo proporcionar uma ou várias partições na massa de dados, em grupos, por algum critério de classificação, de tal forma que exista homogeneidade dentro e heterogeneidade entre grupos (SNEATH & SOKAL, 1973; MARDIA et al. 1997).

Esta técnica sumariza dados para interpretação e utiliza métodos que procuram grupos excludentes, ascendentes, reduzindo as informações de um conjunto de n indivíduos para informações de um novo conjunto de g grupos, onde g é significativamente menor que n , resultando em um dendrograma de exclusão (MARDIA et al. 1997).

De acordo com Reis (1997), de modo sintético, a análise de agrupamento pode ser descrita da seguinte forma: dado um conjunto de n indivíduos para os quais existem informações sobre a forma p variáveis, o método de análise de agrupamento procede ao agrupamento dos indivíduos em função da informação existente, de tal modo que os indivíduos pertencentes a um mesmo grupo sejam tão semelhantes quanto possível e sempre mais semelhantes aos elementos do mesmo grupo do que aos elementos dos grupos restantes. Essa técnica é também chamada de técnica de partição, classificação ou taxonomia, embora o termo partição seja mais utilizado para uma das técnicas específicas da análise, aquela em que os indivíduos são divididos por um número preestabelecido de grupos.

Segundo Aaker et al. (2001), a premissa mais importante da análise de agrupamento é a de que a medida de similaridade ou dissimilaridade na qual o processo de agrupamento se baseia é uma medida válida de similaridade ou dissimilaridade entre os indivíduos. A segunda premissa mais importante é a de que existe uma justificativa teórica para estruturar os indivíduos em grupos. Como em outras técnicas multivariadas, também há teoria e lógica guiando e dando base à análise de agrupamento.

De um modo geral, é difícil avaliar a qualidade do processo de agrupamento. Não existem testes estatísticos padronizados para garantir que o resultado seja puramente aleatório. O valor do critério mensurado, legitimidade do resultado, aparência de uma hierarquia natural e confiabilidade de testes de divisão de amostra, oferecem informações úteis (BUSSAB et al., 1990). Entretanto, é difícil saber, exatamente, quais os grupos são semelhantes e quais objetos menos semelhantes por assim dizer. Geralmente, não é tarefa fácil selecionar um critério separação ou união entre os objetos. Muitos estatísticos recorrem aos métodos não tão triviais, como Bootstrap ou Jackknife para amostragem ou reamostragem aleatórias.

Na análise de agrupamento, é fundamental ter particular cuidado na seleção das variáveis de partida que vão caracterizar cada indivíduo, e determinar, em última instância, qual o grupo em que deve ser inscrito. Nesta análise, não existe qualquer tipo de dependência entre as variáveis, isto é, os grupos se configuram por si mesmo sem necessidade de ser definida uma relação causal entre as variáveis utilizadas. Essa análise não faz uso de modelos aleatórios, mas é útil por fornecer um sumário bem justificado de um conjunto de dados. As técnicas são exploratórias e a ideia é, sobretudo gerar hipóteses, mais do que testá-las, sendo necessária a validação posterior dos resultados encontrados através da aplicação de outros métodos estatísticos (REIS, 1997).

De um modo genérico, a análise de agrupamento compreende cinco etapas (AAKER et al., 2001):

1. A seleção de indivíduos ou de uma amostra de indivíduos a serem agrupados;
2. A definição de um conjunto de variáveis a partir das quais serão obtidas informações necessárias ao agrupamento;
3. A definição de uma medida de semelhança ou distância entre os indivíduos;
4. A escolha de um algoritmo de partição ou classificação dos grupos;
5. Por último, a validação dos resultados encontrados.

2.2 Técnicas de Análise de Agrupamentos

A análise de agrupamentos envolve algumas decisões subjetivas, como qual a técnica é a mais conveniente, de acordo com as circunstâncias. Inúmeras são os tipos de técnicas de agrupamento encontradas na literatura (JOHNSON & WICHERN, 1992; CRUZ & REGAZZI, 1994; MARDIA et al., 1997 ; AAKER et al., 2001; BARROSO & ARTES, 2003), cabendo ao pesquisador escolher aquela que mais se adequa ao seu propósito, uma vez que, as diferentes técnicas podem levar a diferentes soluções.

De um modo geral, ao empregar quaisquer procedimentos de análise de agrupamento, o pesquisador deve tomar cuidado com os seguintes aspectos (ALDENDEFER & BRASHIELD, 1984):

- 1) A maioria das técnicas de análise de agrupamento é procedimento relativamente simples que, geralmente não tem um embasamento teórico estatístico abrangente;
- 2) As técnicas de análise de agrupamento foram desenvolvidos com base em diversas disciplinas, e os vieses herdados de cada uma delas podem diferir muito entre si.
- 3) Métodos de agrupamentos diferentes podem gerar soluções diferentes para o mesmo conjunto de dados.
- 4) A estratégia da análise de agrupamentos busca uma estrutura, enquanto sua operação necessita de uma estrutura preestabelecida.

As próprias técnicas de agrupar podem ser classificadas em grupos, e diferentes autores produzem diferentes classificações. Cormack (1971) propõe a seguinte:

- a) A técnica hierárquica de agrupamento consiste em uma série de sucessivos agrupamentos ou sucessivas divisões de elementos, em que os elementos são agregados ou desagregados por aproximação ou por afastamento. As técnicas hierárquicas são subdivididas em aglomerativas e divisivas. Os grupos, na técnica hierárquica, são geralmente representados por um diagrama bidimensional chamado de dendrograma ou diagrama de árvore.

Nesse diagrama, cada ramo representa um elemento, enquanto a raiz representa o agrupamento de todos os elementos.

- b) As técnicas não-hierárquicas foram desenvolvidas para agrupar elementos em G grupos, em que G é a quantidade de grupos definida previamente. Nem todos os valores de G representam grupos satisfatórios, sendo que aplica-se o método várias vezes para agrupar diferentes valores de G , escolhendo os resultados que apresentem melhor interpretação dos grupos ou uma melhor representação gráfica (BUSSAB et al., 1990). A quantidade de grupos em G é inicialmente um valor aleatório o qual vai se modelando quando os primeiros resultados aparecerem, a exemplo k-médias.

A ideia central da maioria dos métodos por particionamento é escolher uma partição inicial dos elementos e, em seguida, alterar os membros dos grupos para obter-se a melhor partição (ANDERBERG, 1973).

Quando comparada com a técnica hierárquica, a técnica não-hierárquica é mais lenta, demorada e exige mais cálculos computacionais. De um modo geral, os métodos por particionamento diferem entre si pela maneira que constituem a melhor partição. Como qualquer classificação, existirão tipos que serão difíceis de classificar ou que poderão caber em mais de um grupo.

2.2.1 Técnicas de Hierarquização

De acordo com Reis (1997), as técnicas de hierarquização conduzem a uma hierarquia de partições $P_1, P_2, P_3, \dots, P_n$ do conjunto total dos n objetos em 1, 2, 3, n grupos. A denominação de hierárquicos advém do fato de que para cada par de partições P_i e P_{i+1} , cada grupo da partição P_{i+1} está sempre incluído num grupo da partição P_i . Esse tipo de técnica se baseia na construção de uma matriz de dissimilaridade ou distâncias em que cada elemento da matriz descreve o grau de diferença entre cada dois casos com base nas variáveis escolhidas.

Segundo Souza et al. (1997), os métodos hierárquicos se dividem em aglomerativos e divisivos. Entre os métodos aglomerativos, citam-se o do vizinho mais próximo, vizinho mais distante, da mediana, da média das distâncias e o proposto por Ward (1963). Entre os métodos divisivos, o mais conhecido é o de Edwards e Cavalli-Sforza (1965). Nos primeiros, parte-se de n grupos de apenas um

indivíduo, que vão ser agrupados sucessivamente até que se encontre apenas um grupo que incluirá a totalidade dos n indivíduos. O processo inverso é utilizado pelos métodos divisivos: parte-se de um grupo que inclui todos os indivíduos em estudo e por um processo sistemático de divisões sucessivas. Os métodos de análise de agrupamentos mais utilizados pelos pesquisadores são os hierárquicos aglomerativos.

Inicialmente o ponto comum a todos os métodos hierárquicos é a construção de uma matriz de similaridade ou de distância, sendo que este é o terceiro problema a se resolver em qualquer análise de agrupamento.

2.2.2 Definição do Número de Grupos

Determinar o número de grupos para uma base de dados é uma das tarefas mais difíceis no processamento de agrupamento. De acordo com Barroso & Artes (2003), o número de grupos pode ser definido a priori, através de algum conhecimento que se tenha sobre os dados, pela conveniência do pesquisador, por simplicidade, ou ainda pode ser definido a posteriori com base nos resultados da análise.

Segundo Aaker et al., (2001), para determinar o número apropriado de grupos, existem diversas abordagens possíveis: a primeira é que o pesquisador pode especificar antecipadamente o número de agrupamentos. Talvez, por motivos teóricos e lógicos, esse número seja conhecido. O pesquisador pode ter razões práticas para estabelecer o número de agrupamentos, com base no uso que pretende fazer dele. Em segundo lugar, o pesquisador pode especificar o nível de agrupamento de acordo com um critério. Se o critério do agrupamento for de fácil interpretação, tal como a média de similaridade interna do agrupamento, é possível estabelecer certo nível que ditaria o número de agrupamentos. Uma terceira abordagem é determinar o número de agrupamentos com base no padrão gerado pelo programa. As distâncias entre os agrupamentos em etapas sucessivas podem servir de guia, e o pesquisador pode escolher interromper o processo quando as distâncias excederem um valor estabelecido. Uma quarta abordagem é representar graficamente a razão entre a variância total interna dos grupos e a variância entre os grupos em relação ao número de agrupamentos.

O ponto em que surgir uma curva acentuada, um ponto de inflexão, seria a indicação do número adequado de agrupamentos. Aumentar esse número além desse ponto seria inútil e diminuí-lo seria correr o risco de misturar objetos diferentes.

Qualquer que seja a abordagem empregada, de um modo geral é aconselhável observar o padrão total de agrupamentos. Isto pode proporcionar uma medida da qualidade do processo de agrupamento e do número de agrupamentos que emergem nos vários níveis de critério de agrupamento. Geralmente mais de nível de agrupamento é relevante (AAKER et al. , 2001).

Em caso de não existir o conhecimento do número de grupos em que a população em estudo deverá ser dividida, um dos métodos usados quando se usa técnicas hierárquicas consiste na comparação gráfica do número de agrupamento com o respectivo coeficiente de fusão, isto é, o valor numérico de distância ou semelhança para o qual vários casos se unem para formar um grupo (REIS, 1997).

Desse modo, quando a divisão de um novo grupo não introduz alterações significativas no coeficiente de fusão, poderá se tomar essa partição como sendo ótima. Outro procedimento utilizado é o da comparação dos resultados obtidos por vários métodos diferentes de agrupamento. Neste caso, a aferição do grau de convergência entre os vários métodos de agrupamento será feito através de uma tabela de contingência, indicando o número das observações que se agrupam no mesmo agrupamento para o mesmo número de agrupamento. Então é possível verificar a maior ou menor estabilidade das soluções encontradas de maneira a concluir sobre a qualidade do agrupamento efetuado.

2.2.3 Análise de Agrupamento Hierárquico (HCA)

A análise de agrupamento hierárquico consiste no tratamento matemático de cada amostra como um ponto no espaço multidimensional descrito pelas variáveis escolhidas (MOITA NETO, 1998). Nesta técnica é possível tratar cada variável como um ponto no espaço multidimensional descrito pelas amostras, ou seja, se pode ter agrupamentos de amostras ou de variáveis de acordo com o interesse de cada situação. Então, quando uma determinada amostra é vista como um ponto no espaço de variáveis, é possível calcular a distância deste ponto a todos os outros

pontos, constituindo-se assim uma matriz que descreve a proximidade entre todas as amostras estudadas.

Para se calcular a distância entre dois pontos existem vários métodos, porém a mais comum e conhecida é usar a distância euclidiana para duas variáveis, aplicação direta do Teorema de Pitágoras ($a^2 = b^2 + c^2$), onde **a** é a distância e **b** e **c** são variáveis.

Baseado nessa matriz de proximidade entre as amostras se constrói um diagrama de similaridade denominado “dendrograma”, que é uma espécie de árvore. Existem alguns métodos de aglomerar matematicamente estes pontos no espaço multidimensional para formar os agrupamentos hierárquicos, sendo que cada um corresponde a um algoritmo específico que, de um modo geral, é uma particularidade de cálculos feitos no computador que usa as informações da matriz de proximidade com o objetivo de criar um dendrograma de similaridade.

As interpretações de um dendrograma de similaridade entre amostras, ainda segundo Moita (1998), fundamenta-se na intuição: duas amostras próximas, que devem ter também valores semelhantes para as variáveis medidas, bem próximas matematicamente no espaço multidimensional.

Pode-se interpretar que quanto maior a proximidade das medidas relativas à pesquisa, maior a similaridade entre elas. O próprio dendrograma hierarquiza esta similaridade de modo que pode-se ter uma visão bidimensional da similaridade ou dissimilaridade de todo o conjunto de amostras utilizados no estudo. Quando o dendrograma construído é das variáveis, a similaridade entre duas variáveis aponta uma relação muito forte entre estas duas variáveis do conjunto de dados estudados.

Em relação à aplicação da análise de agrupamento hierárquico, deve-se ter variáveis de escalas diferentes e precedida por um tratamento prévio de dados, pois se não for feito, as variáveis com valores mais altos serão mais importantes no cálculo que as de valores numéricos mais baixos.

2.3 Medidas de Distância

De acordo com Barroso & Artes (2003), há dois tipos de medidas de parença: a medida de similaridade, que diz que quanto maior o valor, maior a

semelhança entre os objetos e medidas de dissimilaridade afirmando que quanto maior o valor, menor será a semelhança entre os objetos. Geralmente as medidas de similaridade e de dissimilaridade são inter-relacionadas e facilmente transformáveis entre si (BUSSAB et al., 1990). Existe um grande número de coeficientes de similaridade e/ou dissimilaridade para caracteres binários disponíveis na literatura. Segundo Clifford & Stephenson (1975), tais coeficientes de similaridade podem ser facilmente convertidos em coeficientes de dissimilaridade. Se a similaridade for denominada s , a medida de dissimilaridade será o seu complementar $(1-s)$.

A maioria dos métodos de análise de agrupamento requer uma medida de similaridade ou dissimilaridade entre os elementos a serem agrupados, normalmente expressam como uma função distância ou métrica (DONI, 2004).

Seja C um conjunto, uma métrica em C é função $d: C \times C \rightarrow \mathbb{R}$, tal que para quaisquer $i, j, z \in C$, tenhamos:

1. $d(i, j) = d(j, i)$ simétrica;
2. $d(i, j) > 0$, se $i \neq j$;
3. $d(i, j) = 0$, se e somente se, $i=j$ e
4. $d(i, j) \leq d(i, z) + d(z, j)$ desigualdade triangular.

Além disso, é esperado que $d(i, j)$ aumente quando a dissimilaridade entre i e j aumentar.

Há várias medidas que podem ser utilizadas para distâncias ou dissimilaridade entre elementos de uma matriz de dados. Cormack (1971) descreve uma série de medidas possíveis: distâncias euclidiana, euclidiana quadrada e euclidiana padronizada, distância corda, distância de Nei, distância absoluta ou City – Block Metric, distância de Minkowski, distância Mahalanobis, distância de Chebychev.

Segundo Cormack (1971), as distâncias mais utilizadas em análise de agrupamento são:

1. Distância Euclidiana: a distância entre dois casos (i e j) é a raiz quadrada do somatório dos quadrados das diferenças entre valores de i e j para todas as variáveis ($v = 1, 2, 3, \dots, p$),

$$d_{ij} = \sqrt{\sum_{v=1}^p (X_{iv} - X_{jv})^2}$$

em que X_{iv} representa a característica do indivíduo i
 X_{jv} representa a característica do indivíduo j
 p indica o número de parcelas da amostra
 v número indivíduo da amostra

2. Distância Euclidiana quadrada: a distância entre dois casos (i e j) é definida como o somatório dos quadrados das diferenças entre os valores de i e j para todas as variáveis ($v= 1, 2, 3, \dots, p$)

$$d_{ij}^2 = \sum_{v=1}^p (X_{iv} - X_{jv})^2$$

em que X_{iv} representa a característica do indivíduo i
 X_{jv} representa a característica do indivíduo j
 p número de parcelas da amostra
 v número indivíduo na amostra.

3. Distância euclidiana ponderada.

Para Bussab et al., (1990) na realidade este coeficiente de similaridade e dissimilaridade está associado a uma questão frequente em análise de agrupamento, que é o da ponderação das variáveis, ou seja, o de dar mais peso para variáveis que o pesquisador julgar mais importante para definir semelhança.

$$d = \sqrt{(xi - xj)' S (xi - xj)}$$

S é uma matriz diagonal, tendo i -ésimo componente a variância s^2_i , isto é, $S = \text{diag. } (s^2_1, s^2_2, \dots, s^2_p)$ então:

$(xi - xj)'$ é uma matriz transposta;

$(x_i - x_j)$ é uma matriz;

X_i é o vetor de médias do i -ésimo indivíduo;

X_j é o vetor de médias do j -ésimo indivíduo.

Os casos particulares mais importantes sendo a distância ponderada por S , são:

- 1) $S=1$, a ponderação é a matriz identidade, tem-se então a distância euclidiana usual.
- 2) $S = [\text{diag. } (s^2_1, s^2_2, \dots, s^2_p)]^{-1}$, e tem-se a distância das variáveis padronizadas.
- 3) $S=V^{-1}$, onde V é a matriz de covariâncias, tem-se então a distância de Mahalanobis.

Esta última distância, além de ponderar pela variabilidade de cada uma das componentes, leva em conta também o grau de correlação entre elas. Este fato torna muito difícil a interpretação de resultados baseados neste coeficiente de similaridade dissimilaridade.

4. Distância de Mahalanobis também chamada de distância generalizada. Esta medida, ao contrário das apresentadas anteriormente, considera a matriz de covariância Σ para o cálculo das distâncias. Esta é a distância escolhida para o nosso trabalho, pois ela leva em consideração a estrutura de correlação existente nos dados.

Para Reis (1997), a distância generalizada D^2 de Mahalanobis também pode ser usada como técnica de comparação quando na separação entre diversos grupos, permitindo avaliar a extensão e a direção dos afastamentos entre os valores médios das variáveis usadas na discriminação. As diferenças entre cada par de grupos que estão sendo comparados são assim examinadas simultaneamente por meio de diversas variáveis, que podem ser correlacionadas de modo que a

informação fornecida por uma delas pode não ser independente da fornecida pelas demais.

O valor numérico da maior separação possível entre dois grupos quaisquer é chamado de distância generalizada entre os grupos e mede em escala independente da original, quando necessário transformar a variável, para tornar mais uniforme as suas medidas, variáveis do tipo dicotômicas ou transformações normalizadas ou transformações de caráter exponencial.

Desse modo, o valor da distância generalizada D^2 , ligando dois grupos é um número puro, com propriedades da distância comum, e mede a extensão com que diferem entre si em tamanho e forma.

A distância Generalizada de Mahalanobis entre os grupos i e j é usualmente estimada segundo Rao (1952) por:

$$D^2_{ij} = (X_i - X_j)' \cdot \Sigma^{-1} (X_i - X_j),$$

em que

X_i é o vetor de médias do i -ésimo grupo;

X_j é o vetor de médias do j -ésimo grupo;

Σ estimativa combinada da matriz da covariância/variância dentro dos grupos.

Para o cálculo de D^2 , supõe-se a existência de distribuição multinormal p -dimensional e a homogeneidade da matriz da covariância residual das unidades amostrais, restringindo-se, portanto o seu uso. Entretanto, considerável robustez para violação dessas hipóteses já foi demonstrada, que faz da distância de Mahalanobis uma opção de grande utilidade, principalmente pelo fato de D^2 ter analogia com outras técnicas multivariadas (CRUZ e REGAZZI, 1994). Além disso, a distância de Mahalanobis considera a variabilidade dentro de cada unidade amostral, e não somente a medida da tendência central, sendo, portanto uma medida mais aceitável, quando as unidades amostrais constituem um conjunto de

indivíduos e principalmente, quando as variáveis são correlacionadas (Riboldi, 1986).

Esse método de representação de diferenças entre grupos leva em consideração qualquer correlação que exista entre as variáveis utilizadas, e é também independente das unidades de medida com que as variáveis estão expressas.

2.4 Resultados de Trabalhos Científicos Envolvendo PCA e HCA

Dentre os trabalhos interessantes na área de química inorgânica que usa os métodos PCA e HCA podemos citar o de título “Análise exploratória dos teores de constituintes inorgânicos em sucos e refrigerantes de uva”.(FERREIRA, 2002)

Os teores de Ca, Cu, K, Mg, Na, P e S foram determinados simultaneamente em sucos e refrigerantes de uva utilizando espectrometria de emissão ótica com plasma de argônio, induzido como técnica analítica e os dados obtidos foram submetidos tanto ao PCA quanto ao HCA e os resultados mostraram que o Na foi o principal elemento para discriminação entre as amostras, diferenciando principalmente sucos e refrigerantes. A adição de conservantes e o teor de sais minerais também foram aspectos relevantes para a discriminação dos grupos. A análise exploratória mostrou-se útil para o tratamento dos dados em questão, produzindo informações que dificilmente seriam visualizadas diretamente na matriz de dados.

A determinação dos teores minerais em produtos industrializados é importante para garantir a qualidade e autenticidade dos mesmos. Para a determinação proposta, a técnica analítica utilizada mostrou-se viável, por ser multielementar e possibilitar a introdução direta, sem tratamento prévio das amostras, o que reduz consideravelmente o tempo dispensado às análises químicas.

Por outro aspecto, a interpretação sem tratamento específico do número de dados gerados é consideravelmente complicada, pois o trabalho afirma a importância da utilização da quimiometria (aplicação de método estatístico e matemático, assim como as questões baseadas em lógica matemática em dados de origem química) para o tratamento de dados químicos.

Quando se emprega a análise por componentes principais (PCA), evidencia-se as características comuns e discrepantes entre as diferentes amostras, importantes para o controle de qualidade dos produtos, porém, dificilmente visualizados diretamente da matriz de dados. A análise hierárquica de agrupamentos (HCA) complementou a análise de componentes principais, sendo outra forma de visualizar as semelhanças e diferenças na composição mineral dos diferentes refrescos.

Outro trabalho aplicativo de título “Aplicação de análise exploratória de dados na discriminação geográfica do quiabo do RN/PE”(SILVA, 2005) mostra a forte tendência de evidenciar uma planta hortícola de origem africana através de dados conforme a composição de macro e micronutrientes. Neste trabalho, as técnicas de reconhecimento de padrões não supervisionados (HCA e PCA) foram aplicadas ao conjunto de dados formados por variáveis, medidas por amostras de minerais feitas no quiabo, visando avaliar a constituição nutricional do quiabo, e a discriminação geográfica dos quiabos provenientes dos estados de Pernambuco (PE) e Rio Grande do Norte (RN).

As amostras foram agrupadas de acordo com o esperado, mostrando que a análise exploratória de dados permite a obtenção de informações rápidas e eficientes sobre a similaridade entre as amostras pela visualização gráfica. A PCA identificou as amostras provenientes de PE e RN, e a HCA confirmou os resultados produzidos pela PCA, discriminação geográfica de alimentos é hoje, relativamente simples e torna possível que analistas de alimentos utilizem estes métodos multivariados corretamente, para a construção de modelos de previsão ou classificação, com o objetivo de prever a origem de amostras alimentares.

O artigo científico de Engenharia Biomédica intitulado “Utilização da microespectroscopia infravermelha (FT-IR) para teste de algoritmos estatísticos na diferenciação dos micro-organismos *Candida albicans*, *Candida dubliniensis* e *Candida parapsilosis*”(UEHARA, 2012) mostra as espécies cândida e a devida identificação de infecções fúngicas e os diagnósticos microbiológicos utilizando as técnicas HCA e PCA na ajuda de captura de resultados eficientes.

Os três espectros médios obtidos na experiência foram analisados pelo método de ajuste de curva (não estatístico), e os resultados obtidos mostraram estruturas conformacionais diferentes nos três micro-organismos. Estes resultados confirmam a discriminação obtida através da análise estatística multivariada e visual. Conclui-se

que as análises estatísticas multivariadas baseadas em análise por componentes principais e análise de agrupamentos com uso de algoritmo. HCA é potencialmente útil para discriminar micro-organismos através de seus espectros infravermelhos de *Candida albicans*, *Candida dubliniensis* e *Candida parapsilosis* fornecem um conjunto de dados cuja estrutura de agrupamento é conhecida e que pode ser útil para testar e validar algoritmos estatísticos de análise de agrupamento.

Em relação a trabalhos na área específica de Matemática destaco dois: "Concepções de Matemática e aprendizagem matemática de alunos de engenharia" e "Atitudes em relação à matemática: Um estudo transversal com alunos da educação básica de escolas públicas do estado do Rio Grande do Sul".

As concepções de Matemática revela que o professor de matemática reflete pouco sobre a sua prática escolar e quase nada sobre as concepções de matemática e aprendizagem matemática. O trabalho utiliza técnicas multivariadas hierárquicas com uso de algoritmos e conclui que os professores devem estar atentos para trabalhar com seus alunos desenvolvendo estratégias de ensino que conduzam a uma aprendizagem relacional, a uma forma de pensar investigativa, de tal forma que as concepções de matemática e de conhecimento matemático deixem de constituir uma visão fragmentada da ciência.

No outro trabalho sobre os alunos da educação básica do Rio Grande do Sul, foram selecionados 345 alunos , 53% do sexo masculino e 47% do sexo feminino, com idades variando de 9 a 19 anos. O objetivo era identificar alguns fatores associados às atitudes positivas ou negativas em relação à matemática. Na avaliação aplicou-se o coeficiente alfa de Cronbach, depois uma análise descritiva pelo teste de Kolmogorov-Smirnov, teste de Qui-Quadrado, análise de variância e regressão logística por Forward/Wald.

Foi concluído que existe uma relação entre autodesempenho e atitude em relação à matemática, quanto melhor o desempenho mais favorável mostrou-se as atitudes. Os alunos que manifestaram terem uma boa interação com a matemática, principalmente com a resolução de problemas evidenciaram de forma predominante atitudes positivas. Na medida em que a idade do aluno avançou, percebeu-se uma redução na probabilidade de manifestar atitude favorável à matemática. Os professores devem buscar estratégias para desenvolver um sentimento afetivo e permanente do aluno com a matemática.

3. METODOLOGIA

Tendo como propósito o êxito ou não na aprovação do aluno e de acordo com os dados do INEP em relação ao desempenho dos alunos no 9º ano do ensino fundamental das escolas públicas de Petrolina-PE e Casa Nova-BA, são revelados estimadores muito baixos em relação à média nacional. Embora não se tenha uma nota específica para matemática, é possível obter uma ideia geral do rendimento do aluno por meio da aplicação de técnicas estatística. (INEP, 2011)

A partir da aplicação de questionário, com perguntas relacionadas à vida do aluno foi construído um banco de dados em planilha Excel, com base em uma pesquisa realizada em quatro escolas da rede pública de ensino, nos municípios de Petrolina-PE e Casa Nova-BA, selecionadas aleatoriamente, e que oferecem obrigatoriamente o 9º ano do ensino fundamental II. A partir deste banco de dados, foram selecionados alunos matriculados e com frequência regular nas aulas de matemática.

Foram contabilizados um número expressivo de 440 alunos, de forma aleatória, sem qualquer restrição ou condição imposta, pesquisados em três semanas ininterruptas. Para a coleta das informações e composição do banco de dados foi utilizado questionário individual. No entanto, observou-se alguns problemas com relação ao uso deste tipo de recurso para a coleta dos dados, tais como: o grupo estudado não está habituado com preenchimento de questionários; equívocos na grafia; falta de conhecimento sobre determinadas variáveis, estado civil, número de irmãos, entre outros. Considerando estes problemas, esta pesquisa considera um erro de 1% na perda de informação e ainda foi constatado uma amostra fortemente homogênea pelo cálculo de variância de 3,55, possibilitando uma amostra ideal de 83,5 pessoas.

As variáveis catalogadas dividiram-se em qualitativas e quantitativas, questionando: sexo, cor/raça, notas em matemática do 8º ano (igual ou superior a 5 pontos), repetência, turno (diurno ou noturno), boa relação com a escola, prática de esportes, se fuma e bebe, se possui auxílio nos exercícios em casa, religião, trabalho remunerado, benefício social, estado civil, número de filhos registrados, número de irmãos, estado civil dos pais, renda familiar, tipo de moradia, escolaridade do pai e da mãe, a quantidade de pessoas que habitam a residência,

acesso a microcomputador, celular, tablete, internet e se continuará seus estudos no ensino médio.

A partir das informações coletadas, foi realizada uma análise descritiva dos dados, e em seguida uma análise multivariada por agrupamento de índice de similaridade de Jaccard, a qual tornou-se possível comparar as variáveis que apresentam maior e menor correlação. Esta técnica possibilitou comparar o comportamento das variáveis dentro do mesmo grupo, bem como comparar os grupos a partir dessas mesmas variáveis.

Como existem variáveis de diferentes medidas é necessário transformar as variáveis qualitativas em variáveis quantitativas e a criação de um ponto de corte. Assim, os resultados que se apresentam acima de determinado valor substituímos a informação por 1 (presença) e abaixo desse valor substituímos por 0 (ausência), definindo de imediato um padrão para as variáveis em estudo.

A partir daí, a definição do padrão de medidas da variável e a obtenção da matriz dos coeficientes de similaridade de Jaccard, exigida pela dicotomização das variáveis e para tal função de Jaccard do tipo $\rho = a/(a+b+c)$, onde $0 \leq \rho \leq 1$ grau de similaridade entre variáveis ij ; a - ocorrência em ambas as variáveis; b e c - ocorrência de 1 em uma entre duas variáveis analisadas.

Para o cálculo do grau de ligação e posterior confecção do dendrograma foi aplicado o método do vizinho mais próximo, ao qual liga as variáveis com maior semelhança e estabelece grupo ou grupos dissociados. Ao final, para verificação da qualidade no cálculo das distâncias e a qualidade dos métodos de ligação entre variáveis aplicou-se o coeficiente de correlação cofenética.

4. DEMONSTRAÇÃO E ANÁLISE DA PESQUISA

A primeira variável analisada no estudo foi o sexo dos alunos. O gráfico 1 abaixo apresentou um comportamento semelhante aos dados do (IBGE,2011) o qual indica nos seus estudos um equilíbrio entre a quantidade de homens e mulheres. Nesse caso, a maioria do sexo feminino (52,3 %), contrapondo do masculino (47,7%), embora corresponda a uma faixa escolar não registrada pelo instituto.

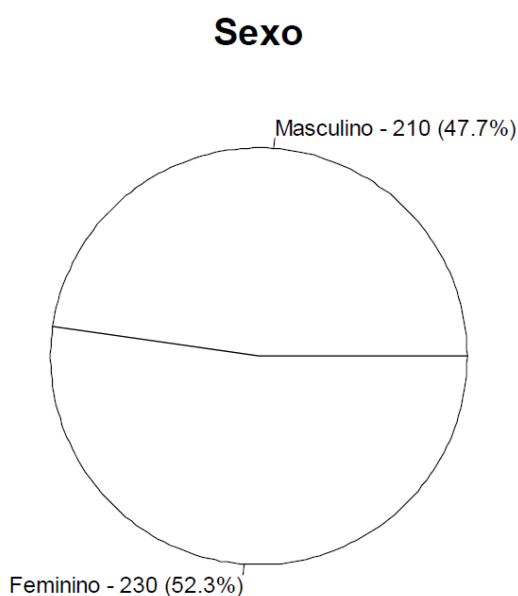


Gráfico 1: relação entre número de homens e mulheres pesquisados.

De acordo com o Ministério da Educação e a sua legislação regulatória para idade escolar do ensino fundamental de 06/06/2006, os indicadores nacionais apontam que, atualmente, 3,6% das crianças em idade escolar ainda não estão inseridas na escola. Do total de matriculadas registradas, 21,7% são repetentes da mesma série e apenas 51% concluirão o ensino fundamental, sendo 10,2 anos em média o tempo total na escola (MEC, 2007).

Porém, pode-se observar que esses dados fogem da referência idade/série definida pelo Ministério da Educação para a conclusão do ensino fundamental, que deve ser em nove anos, tendo os alunos idade média variando entre 11 e 14 anos, sendo que a última série deve ser cursada por alunos com idade de até 15 anos.

Assim, conclui-se que todo aluno acima de 15 anos, segundo o MEC, está fora da sua faixa etária (MEC, 2007),

O gráfico 2 apresenta a distribuição por idade dos alunos catalogados, onde a situação apontada pelo MEC é refletida nas quatro escolas estudadas. Mais de 20% desses alunos estão acima da idade mínima para cursar o ensino fundamental, estando com 15 anos ainda cursando o 9º ano. Pouco mais de 50% estão na idade série do curso, entre 12-14 anos. Apesar de o Brasil ter avançado no acesso e permanência dos alunos no Ensino Fundamental, esses resultados mostram a necessidade de corrigir essa distorção em termos de idade escolar.

Segundo Sergei(2208), o censo escolar de 1998 a 2005 revelou que insumos escolares tem pouca importância no desempenho educacional, principalmente no caso de escolas com maior precariedade das condições de infra-estrutura.

Distriubuição Aluno/Idade

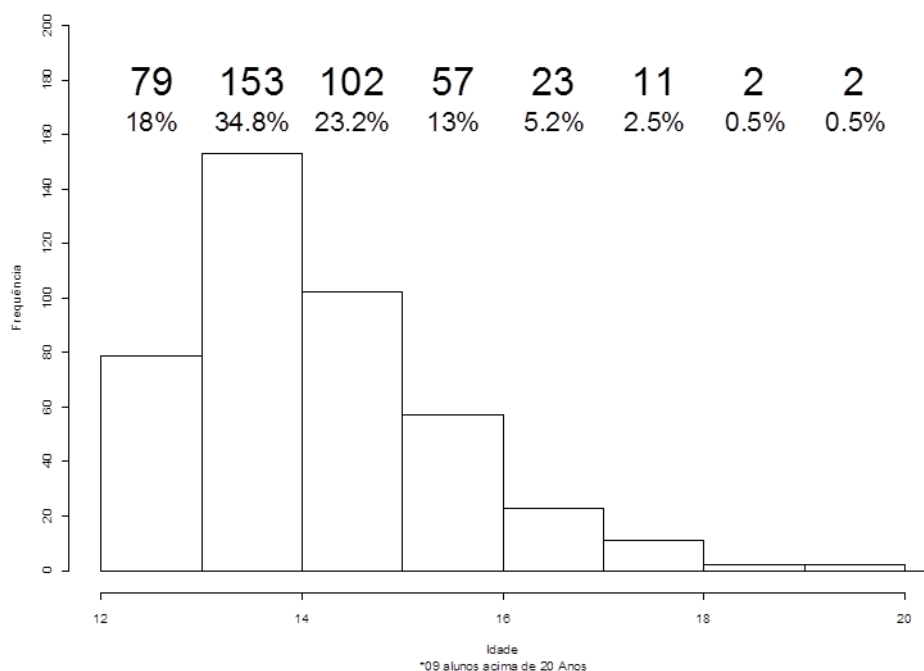


Gráfico 2: distribuição da faixa etária identificada nas escolas.

O trabalho de pesquisa revelou ainda que 34,8% dos alunos possuem idade de 14 anos, o que coloca a educação dos jovens com um percentual maior do que o da idade ideal. E se somarmos 1,4% dos que possuem 12 anos de idade com os 16,6% dos de 13 anos, e com os que estão com a idade limite de 15 anos (23,25%) teremos exatamente 86% dos alunos em idade escolar no 9º. ano do Ensino Fundamental, uma taxa muito elevada em relação a outras realidades encontradas pelo Brasil.

Se analisarmos esses resultados considerando as características da região pesquisada (Petrolina-PE e Casa Nova-BA), pode-se dizer que os dados apresentam, de certa forma, uma surpresa, uma vez que possuem alunos que abandonam a escola por diversos motivos, tais como: o trabalho infantil, gravidez na adolescência, problemas familiares, uso de entorpecentes e drogas ilícitas, e até mesmo a repetência, que faz com que muitos discentes fiquem fora da faixa etária ideal de atividades escolares.

Porém, um dos fatores que contribuem decisivamente para a manutenção da frequência do aluno na escola é o programa Bolsa Família, concedida pelo governo federal como incentivo ao aluno que vai a escola, complementando a renda familiar

e exigindo dos pais maior fiscalização sobre os filhos perante a frequência escolar.(PROGRAMA BOLSA FAMÍLIA, 2010)

Em relação aos aspectos de cor/raça, foram disponibilizadas para o grupo pesquisado quatro categorias: branca, preta, parda, e amarela. Os dados apresentados no gráfico 3 mostram que 67,5% declaram-se pardos, como a faixa de maior frequência, seguido por brancos com 23%, pretos com 8% e amarelo com 1,6%. Esse percentual tão abaixo da taxa nacional deve-se a dificuldade da autoafirmação da cor, onde os estudantes não reconhecem a sua raça, ficando longe da sua verdadeira identidade, fato verificado quando mais de 67% dos estudantes escolheram a cor parda. Isso reflete um problema jurídico, pois a Constituição Federal permite que o cidadão decida a sua raça, não podendo ser confrontado.

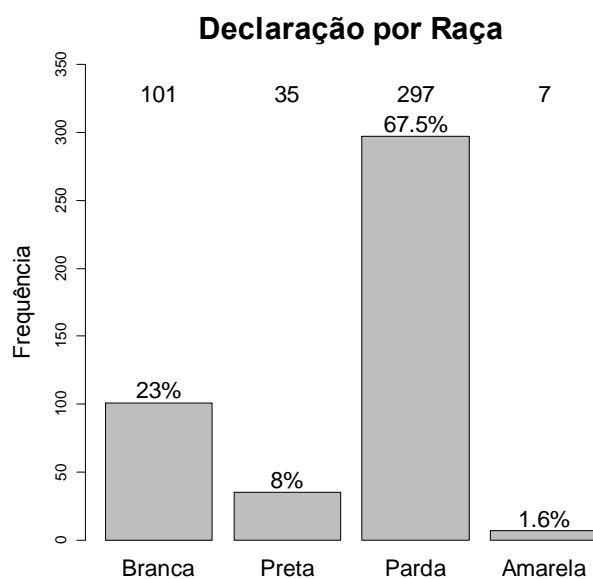


Gráfico 3: autoafirmação e catalogação das raças declaradas.

Em relação à variável religião, demonstrado no gráfico 4, a pesquisa identificou certo viés entre a realidade e a resposta dos questionários. Os dados mostram que grande parte dos alunos segue a religião católica, que no decorrer dos anos vem perdendo espaço para outras religiões. Outros fatores poderiam influenciar na resposta, tal como a discriminação.

Religião

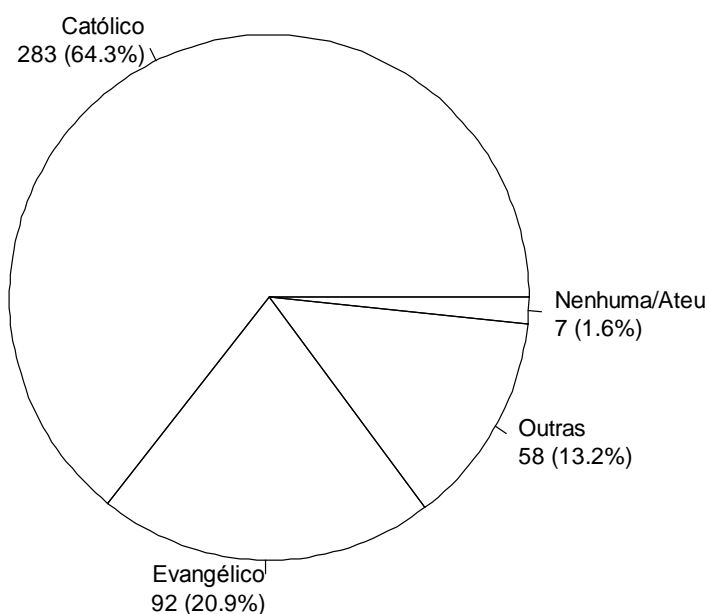


Gráfico 4: religiões declaradas pelos alunos pesquisados.

Segundo dados do IBGE (2010) no estado do Pernambuco, 66,3% são católicos, 20,3% evangélicos, 1,4% de outras religiões e 12% declaram não possuir religião. No estado da Bahia, a realidade é um pouco diferente: 51,2% são católicos, 18,1% evangélicos, 10,1% de religiões africanas, como o candomblé, cabula, umbanda e 20,6 não possuem religião. Porém, verificou-se que esses dados divergem dos levantados na pesquisa com os alunos do 8º ano do ensino fundamental, que, de certa forma, podem ter sido influenciados com a resposta de outros alunos ou até mesmo, por vergonha de declarar a sua verdadeira religião. Os percentuais não estão de acordo com os dados, embora tenha citado alguns aspectos e outros como a região, a cidade do interior, a classe econômica e os costumes.

Ainda em relação a esses dados, podemos destacar que, independentemente da escola, o grupo pesquisado, em quase sua totalidade, seguem alguma religião. Porém, mesmo apresentando significativa contribuição ética no desenvolvimento e formação do caráter e personalidade do jovem, ela não representa fator determinante, mas possui seu papel social. Ocorre que na adolescência, os ideais estão nascendo, mesmo que em áreas muito carentes. Dessa forma, os postulados religiosos norteiam os passos do iniciante nas experiências humanas, apresentando muitos dissabores e insucessos nas diferentes áreas de comportamento.

Observou-se ainda que, os estudantes do 9º ano do ensino fundamental, universo da pesquisa, em sua maioria adolescentes, o comportamento muda totalmente os seus valores, que podem alterar com frequência a sua vida social, empregatícia e até mesmo a sua vida sexual. Uma pesquisa do Data Popular (2012) aponta diferença considerável da percepção entre os mais jovens e mais velhos quanto à religião. O apego às crenças é proporcionalmente maior entre os mais velhos, sendo que 88,3% dos entrevistados com 50 anos ou mais concordam que é importante ter uma religião, ao passo que entre os jovens de 12 a 16 anos, faixa etária desse estudo, esse percentual cai para 76%.

Outro dado importante se refere ao sucesso e fracasso do aluno na disciplina de matemática, apresentado no gráfico 5. Foi adotada a média limite de 5,0 pontos para considerar sucesso, tanto para o estado da Bahia como para Pernambuco, como uma forma de igualar as médias.

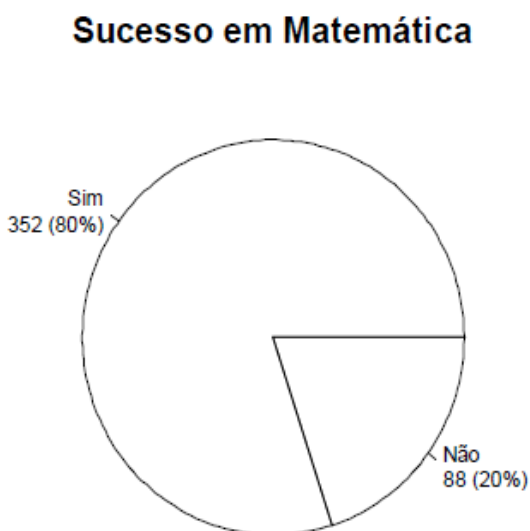


Gráfico 5: quantidade dos sucessos em matemática no 8º ano E.F.

Podemos dizer que o ideal seria que a média anual dos alunos, antes da recuperação, fosse originada de um processo avaliativo governamental, neutralizando a participação da escola na definição da nota, refletindo de maneira mais concreta a realidade do aluno. Isso porque são exames aplicados de forma rigorosa, uma vez que não é permitido ao professor de matemática ter acesso a avaliação. Porém, o que ocorre é que o aluno não é preparado para responder as avaliações externas. Isto é um fator decisivo para a possibilidade de padronização das notas e definição de um valor nacional ou regional adequado.

Outra consideração importante a fazer é o desinteresse dos alunos em responder as avaliações externas. Isso é resultado do pensamento dos estudantes de que a resolução ou não dessas provas pouco influenciará na sua vida escolar, pois são realizadas apenas para obter dados para verificação do conceito da escola na aplicação dos índices de aprendizagem. Assim, observa-se que, na maioria das vezes, a falta de interesse dos alunos na resolução das provas externas, leva-os a resolver as questões sem uso de regras ou conceitos técnicos, fato esse replicado em sala de aula com o professor, entregando as questões sem solucionar ou nem tentar solucionar, escolhendo aleatoriamente as alternativas.

Dessa forma, com os motivos apresentados anteriormente, somado a diversidade de professores e atividades escolares, fica difícil relacionar um conceito ideal para o aluno, embora os melhores fiquem um pouco acima da maioria esperada.

Nesta pesquisa, obtiveram sucesso com aprovação por média em matemática no 8º ano do ensino fundamental, exatamente 80% e fracasso de 20%, quantidade alterada quando colocamos a aprovação anual para a série seguinte. Um valor considerado alto quando comparamos com qualquer avaliação externa, o que indica uma contradição aos padrões governamentais, com clara e evidente aprovação em massa.

Em relação às notas obtidas, com base nas escolas pesquisadas, Jacob Ferreira, João Batista, Luiz Viana e Moysés Barbosa, apresentadas no gráfico 6, a que apresentou melhor desempenho foi o Colégio Estadual Conselheiro Luiz Viana com 71,4%, seguida da Escola Municipal Jacob Ferreira, com um percentual de aprovação de 69%, e o Colégio Estadual Moysés Barbosa ficou apenas um décimo deste percentual, com 68,9%. A Escola Estadual João Batista dos Santos

apresentou um índice de aprovação de 58,8%, embora tenha ficado na primeira colocação em relação à média mínima de aprovação de 28,6%.

De modo geral, a Escola Moysés Barbosa atingiu níveis de notas acima da média escolar e muito próximas, podendo ser considerada a que teve um desempenho melhor do que as demais escolas em estudo, quando se tem como objetivo uma média do desempenho escolar.

Deve-se levar em consideração que o Colégio Estadual Conselheiro Luiz Viana alcançou um percentual de aprovação alto devido sua média de aprovação ser no mínimo 5,0, valor adotado pelo estado da Bahia. Esta realidade é bem focada aqui no polo Petrolina/Juazeiro por causa da estreita distância entre as cidades, não reparada muito em outros municípios um pouco distantes e com pequenas populações dos dois estados. A estranheza em todo o processo de avaliação ficou por conta de não se ter nenhuma média anual 10, em nenhuma escola verificada. Por mais que se seja rigoroso, deve ter algum aluno com um potencial elevado para atingir a nota máxima conquistada na sala de aula.

A variabilidade dos êxitos escolares nas escolas pesquisadas deve-se ao fato de diferentes critérios estabelecidos nas escolas e forma de avaliação dos professores, embora tenha professores que atuam em ambos os estados.

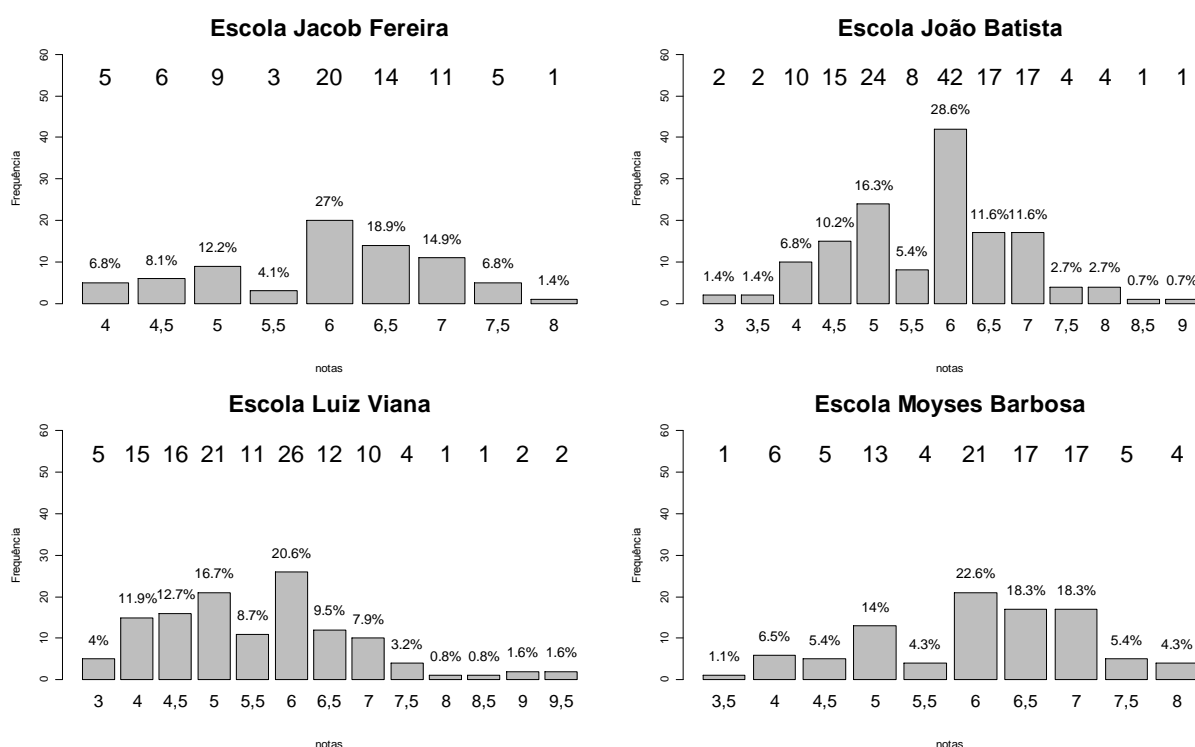


Gráfico 6: desempenho das escolas verificando as notas obtidas.

O gráfico 7 apresenta a análise de agrupamento de variáveis calculada por distância de Jaccard, onde dois grupos distintos foram criados, Grupo Sucesso e Grupo Fracasso. Estes geraram um ordenamento de variáveis, por ligação simples.

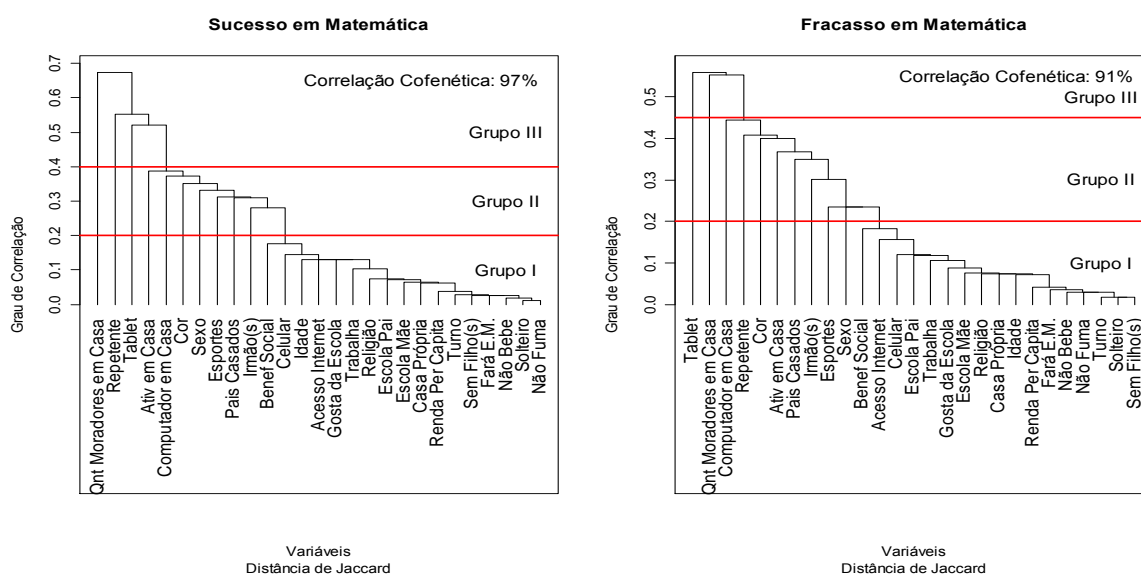


Gráfico7: agrupamento das variáveis correlacionadas por sucesso e fracasso.

Na variável fumo, o hábito de não fumar, foi constatado que este é um fator que possui maior correlação no grupo sucesso, do que no grupo fracasso. Ambos os grupos apresentam que o ato de não fumar é comum. Um estudo publicado em Israel em 2012 apontou que pessoas que fumam têm um Q.I. mais baixo do que as não fumantes. A amostra da pesquisa publicada foi composta por 20.000 pessoas, entre adolescentes de 16 a 21 anos no exército de Israel. Isto é conflitante, pois a maioria dos intelectuais é fumante e utilizam o mesmo para as suas produções artísticas

Outro estudo sobre o uso do tabaco, Simão (2013) apontou que o maior índice de fumantes está entre adolescentes a partir de 17 anos das classes sociais D ou E e que os fatores influentes são mães e amigos fumantes, separação dos pais, emoções físicas, abusos emocionais, físicos ou sexuais, uso de drogas ilícitas relacionadas e o álcool. Note-se que a maioria dos alunos pesquisados não faz parte desta classe acima dos dezessete anos.

Em relação ao turno, ambos os grupos são semelhantes, pois nos alunos pesquisados a maioria estuda no turno diurno e influenciou significativamente no

resultado. É possível que um estudo mais profundo, com equilíbrio nos horários dos estudantes para apresentar outra conjectura de variáveis. Pois, o aluno do curso noturno geralmente dispõe de menor tempo de estudo quando comparado com o aluno diurno, bem como qualidade nos índices de aprendizagem, cansaço físico do aluno e professor, tempo reduzido das aulas em relação ao diurno, naturalmente confeccionado pela própria disposição do horário.

A maioria dos alunos pesquisados apresenta intenção de continuar os estudos e cursar o Ensino Médio, mostrando um fator positivo encontrado no grupo sucesso. De fato, os alunos que fazem uma projeção para o seu futuro e desejam continuar seus estudos, as suas notas tendem a melhorar se comparada àqueles que desejam parar temporariamente ou definitivamente sua vida escolar. Essa variável indica um desejo ou objetivo de desenvolvimento pessoal, um fator social, fazendo com que se esforcem para novas conquistas.

Os alunos que não possuem filhos apresentaram um viés, onde é uma indicação que o aluno não tenha interpretado corretamente a pergunta, influenciando a formação do dendrograma, pois o estudante que tem filho apresenta melhor rendimento em relação ao estudante de menor rendimento que não tem filho. O entrevistado não teve culpa neste caso e a falha foi instrumento da pesquisa.

A renda familiar, de acordo com os dados apurados, não é fator primordial no sucesso ou fracasso dos alunos, embora haja alguns valores significativos. A posição está muito semelhante em ambos os grupos, situação explicada de acordo com o perfil da renda uniformemente distribuída entre os alunos das escolas públicas.

Os dados da pesquisa mostram, ainda, que a escolaridade dos pais é um fator decisivo no desempenho dos alunos, contribuindo para o seu rendimento em sala de aula. Aqueles em que os pais estudaram um pouco mais são influenciados para obter um melhor resultado escolar, no entanto, os alunos pertencentes ao grupo fracasso apresentaram pais com baixo nível de escolaridade. O IBGE (2010) confirma essas informações de escolaridade.

A idade-série do aluno não garante a obtenção de boas notas, pois o gráfico mostrou ser uma variável menos importante em nível de correlação no grupo sucesso.

O acesso a internet e ao computador em casa são variáveis que influenciam positivamente no rendimento escolar. Os alunos que apresentaram um menor contato com a internet e/ou não possuem computador nas suas residências apresentaram uma influência negativa no rendimento escolar de matemática. Barreto (2014) descreve que as novas tecnologias interativas e o uso da internet influenciam positivamente no processo de ensino-aprendizagem da matemática, para todas as idades e séries.

O uso de celulares, segundo a amostra, influencia negativamente o rendimento escolar em matemática, pois o grupo fracasso apresenta o celular como uma variável mais forte do que o grupo sucesso, o que é comprovado nos estudos de Teixeira (2012), que demonstra que muitas vezes o aluno usa o celular para filmar os exercícios ao invés de copiar a matéria e para colar nas provas. Entretanto, o celular pode ser utilizado para fins positivos como gravar as aulas, enviar mensagens sobre determinados assuntos vistos na sala de aula, e interagir com outros alunos.

Receber benefício social, ou seja, bolsa família ou similar, não influenciou igualmente para ambos os grupos, sucesso e fracasso, destacando que essa variável não foi importante nos grupos estudados. Fato contraditório, pois é possível que o aluno não tenha consciência que seus pais ou responsáveis recebam em seu nome, bem como as pesquisas de Ministério da Saúde (MDS, 2008) que afirmam que o benefício social aumentou a frequência e melhorou o desempenho escolar dos alunos.

A quantidade de irmãos é um dado de forte impacto no rendimento escolar dos alunos estudados, pois, tendo uma quantidade pequena de irmãos, este aluno terá uma maior atenção do seu responsável. Já os alunos que apresentam mais de dois irmãos foram inseridos no grupo fracasso.

A variável prática esportiva esteve presente no grupo fracasso, indicando uma influência negativa no rendimento de matemática dos alunos pesquisados. Isso revela uma contradição no primeiro momento, mas é totalmente aceitável, pois o aluno está faltando as aulas para praticar esportes, como jogar bola na rua, por exemplo, e por esta razão respondeu que faz atividades esportivas.

Em relação a variável cor/raça, os resultados apontaram que os que se consideraram da cor branca, parda e amarela possuem uma correlação mais forte

no grupo sucesso do que no grupo fracasso. Desses, os indivíduos que possuem a cor branca apresentam melhor desempenho em matemática do que os de cor preta, o que reforça ainda mais as políticas de igualdade entre as raças. Por essa razão, as últimas políticas governamentais enfatizam certa preocupação na inserção de cotas nas universidades. Segundo Lesme (2012), que utiliza a lei, as cotas não valem somente para negros, como também para indígenas e seus descendentes para compensar as desigualdades sociais, econômicas e educacionais.

A repetência escolar apresentou um comportamento inversamente proporcional, medida que está no grupo sucesso apresenta menor influência, no entanto no grupo fracasso apresentou-se como uma variável mais importante que no grupo anterior. De acordo com Aquino (2007), quanto mais o aluno repete, menor é o seu sucesso, gerando uma baixa autoestima nesse estudante. O autor declara ainda que o incentivo ao aluno de todas as formas pode despertar sua criatividade, intuição, sensibilidade e gosto pela matemática.

Observou-se ainda que o fato dos pais serem casados gera maior segurança nos estudantes e por esta razão suas notas são melhores, como mostram os dados apresentados no gráfico . A estrutura familiar contribui com o desempenho do aluno, podendo ser explicado pela fiscalização e controle da frequência do estudante pelos pais. Em contrapartida, a menor incidência de pais casados apresenta relação com o baixo rendimento em matemática, podendo, provavelmente estender para outras disciplinas, como português, história e demais. Souza (2009) destaca as inúmeras incumbências da família na educação do aluno e a ausência dos pais, casados ou não, que não comparecem às escolas, transferindo toda e qualquer responsabilidade das tarefas escolares para a instituição.

A pouca presença de tablets nesse estudo e a baixa contribuição em ambos os grupos pouco ou quase nada poderia concluir nesse gráfico 7, pois os tablets ainda apresentam um custo elevado diante da renda familiar do grupo estudado.

Esse estudo apresentou tanto para o grupo sucesso, quanto para o grupo fracasso, uma correlação cofenética, ou seja, correlação entre o método da Distância e Método da Similaridade aplicado, de 97% e 91%, respectivamente, sendo os dois dendrogramas satisfatórios para devidas análises.

5. CONSIDERAÇÕES FINAIS

A pesquisa de campo ao coletar informações dos alunos da educação básica antes de partir para a coleta definitiva, deverá obrigatoriamente ser testada por grupo controle para correções posteriores a coleta definitiva.

O acesso dos alunos às tecnologias mostrou estreita relação com desempenho positivo na disciplina de matemática. Isso evidencia a necessidade de aproximação do ensino de matemática com as tecnologias educacionais, devendo o professor explorar melhor esses recursos no processo de ensino e aprendizagem dos estudantes do ensino básico.

O celular em sala de aula, como é utilizado atualmente, é ferramenta nociva ao rendimento escolar em matemática. Daí a necessidade de transformá-lo em um instrumento pedagógico, redirecionando o seu uso para fins educacionais. Isso poderá ser feito através do uso de aplicativos que possibilitem interação entre os alunos, e deste com o conteúdo.

A parceria família e escola deverá ser reforçada no intuito de tornar o ambiente escolar similar a um ambiente familiar, bem como oportunizar aos pais que possuem formação incompleta retornarem à escola como estudantes. A presença dos pais na escola e a aproximação deste com o ambiente educacional, contribui para o estímulo dos alunos aos estudos

A presença do assistente social na escola também representou fator importante para auxiliar e orientar os alunos nas suas dificuldades de aprendizagem, reflexos de situações vivenciadas pelo educando no ambiente familiar ou dentro da própria escola.

Pode-se concluir que o trabalho foi bastante significativo, pois produziu dados úteis referentes ao desempenho escolar de alunos do 8º ano do ensino fundamental, aplicados a diversos fatores relevantes.

REFERÊNCIAS

AFONSO E AUBYN; **O equilíbrio e o desempenho escolar**. UPE (Universidade de Pernambuco)- Recife, 2012.

ALDENDERFER, M.S.; BLASHFELD, R.K. **Cluster analysis**. Beverly Hills; Sage, 1984, 547p.

ANDERSON, T.W. **An introduction to multivariate statistical analysis**. New York: John Wiley & Sons, 1984, 675p.

AQUINO, Julio G. **Erro e Fracasso na escola: alternativas teóricas e práticas**. Ed Summus, São Paulo, 2007.

AYRES, José Costa. **Técnicas Multivariadas**, Editora Modelo- Porto Alegre/Rs, 2012

BARBOSA, E. F. **Implantação da qualidade total na educação**. Belo Horizonte: Fundação Cristiano Ottoni, 1995.

BARRETO, Evanice Ramos Lima. **A influência dos computadores e Internet no processo ensino-aprendizagem de Matemática**. Revista Espaço Acadêmico-159- São Paulo, SP , agosto de 2014, v.9 nº.106.

BARROSO, L.P., ARTES, R. **Análise de Multivariada**. Lavras: UFLA, 2003. 157p.

BASE CURRICULAR COMUM PARA AS REDES PÚBLICAS DE ENSINO DA BAHIA. **MATEMÁTICA ENSINO FUNDAMENTAL E MÉDIO**. Salvador, Ba, 2009.

BOUROCHE, J.M. SAPORTA, G. **Análise de dados**. Rio de Janeiro: Zahar, 1972. 116p.

CARDOSO, M. A. G.; SANTOS P. M.; KHOURI, S.; PAULA JUNIOR, A. R.; UEHARA, M.; SAKANE, K. K., **Utilização da microespectroscopia infravermelha**

(FT-IR) para teste de algoritmos estatísticos na diferenciação dos micro-organismos Candida albicans, Candida dubliniensis e Candida parapsilosis.

Universidade do Vale do Paraíba, São Paulo, artigo original, 2012.

CLIFFORD, H. T.; STEPHENSON, W. **An introduction to numerical taxonomy.** London: Academic Press, 1975. 229p.

CORMACK, R. **A review of classification.** Journal of the Royal Statistical Society (Series A), v. 134, p. 321-367, 1971.

DATA POPULAR PESQUISAS NACIONAIS- **Pesquisas estatísticas e de opinião. Religião e Juventude.** São Paulo-SP, 2012.

DONI, M. V. **Análise de Cluster: métodos hierárquicos e de partição.** São Paulo: Mackenzie: 2004. 93f. Monografia de Pós Graduação – Universidade Prebisteriana Mackenzie, 2004.

EVERITY, B. **Cluster analysis.** London: Heinemann Educational Books, 1974. 136p.

FERREIRA, Edilene Cristina-**Análise exploratória dos teores de constituintes inorgânicos em sucos e refrigerantes de uva.** Eclético Químico , vol 27 São Paulo, 2002.

FRANCO, R.F., **Métodos Quimiométricos: Uma visão geral.** Conceitos Básicos de Quimiometria. Universidade Federal de Viçosa. Viçosa, Mg, vol. 1, 2013.

GAMA, M. de P. **Bases da análise de agrupamentos (Cluster Analysis).** Brasília: UNB, 1980. 229f. Dissertação de Mestrado em Estatística e Métodos Quantitativos- Universidade de Brasília. 1980.

GATTI, Bernadete, **Desempenho escolar do aluno,** editora livre, São Paulo/SP, 2010.

HAIR, Jr., J.F. et al. **Análise Multivariada de Dados**. 5. Ed. Porto Alegre: Bookman, 2005.

IBGE-INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA-CENSO 2010. **Educação e Religião**- Zona Industrial de Brasília, DF, 2011.

INEP – INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Ensino Básico**. Dados do Ensino Básico. Ministério da Educação e Desporto – Zona Industrial de Brasília, Brasília, DF, 2011.

JAIRO, S. F.; GILBERTO A. M. **Curso de Estatística**. São Paulo: Atlas. 1996, 320p.

JOHNSON R.A.; WICHERN D.W. **Applied Multivariate Statistical Analysis**. 3.ed. New Jersey: Prantice Hall, 1992. 642 p.

LESME, Adriano. **Vantagens e desvantagens das cotas raciais na educação**.Ed. Venâncio, São Paulo, 2012.

LOPES, L. F. D. **Identificação de fatores que influenciam na qualidade do ensino da matemática, através de análise fatorial**. 2007 52f. Dissertação (Mestrado em Estatística) – Universidade Federal de Santa Maria, RS, Brasil, 2007.

LOPES, M.P.D. **Gerenciamento da qualidade no ensino da matemática**. 2004. 104f. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Santa Maria, Santa Maria, 2004.

MARDIA, A. K. V.; KENT, J. T.; BIBBY, J. M. **Multivariate Analysis**. London: Academic Press, 1997, 518p.

MDS- Ministério do Desenvolvimento e Combate a Fome- **Bolsa Família**- Brasília,Df, 2008.

MINISTÉRIO DA EDUCAÇÃO- MEC- **Dados do censo escolar**- Gráfica do Senado- Brasília/DF , 2007

MEDEIROS, C. P. **Análise de Agrupamentos – Técnicas Estatísticas para Reconhecimento de Padrões**. 1. ed.- Universidade Federal de Santa Cruz. Santa Cruz do Sul, Rio Grande do Sul, 2012.

MOITA, J M, **Filosofia da Ciência, Estatística multivariada**. Jornal da Tarde, Salvador/Ba, 2014.

MOITA NETO J M, **Uma visão didática-metodológica de estatística multivariada**, Universidade de Brasília/UNB, Brasília/DF, 1998

MOURA, José Egídio, **As Técnicas Multivariadas na Educação** , Editora Científica, São Paulo, 2006.

Novaescola.abril.com.br / sucessos e fracassos do aluno, edição março 2012, Editora Abril, 2012.

ORGANIZAÇÃO DE COOPERAÇÃO PARA O DESENVOLVIMENTO ECONÔMICO, **Dados Estatísticos**, Editora Nobre, São Paulo, 2007.

PARÂMETROS CURRICULARES NACIONAIS (PCN). Terceiro e quarto ciclos do ensino fundamental. **Matemática**. Ministério de Educação e Desporto. Brasília,1998.

PARÂMETROS PARA EDUCAÇÃO BÁSICA DO ESTADO DE PERNAMBUCO. **MATEMÁTICA NÍVEL FUNDAMENTAL E MÉDIO**. Recife, Pe – 2012.

PROGRAMA BOLSA FAMÍLIA- **Considerações sobre o programa**. Ministério de Desenvolvimento e Combate a Fome, Brasília/DF , 2010.

RAO, C. R.; **Advanced Statistical methods in biometric research**. New York: John Wiley & Sons. 1952. 390p.

REIS, E.; **Estatística multivariada aplicada**. Lisboa, Portugal: Edições Silabo, 1997. 342p.

RIBOLDI, J. **Análise de agrupamento “Cluster Analysis”**. Piracicaba, Sp: ESALQ/USP, 1986. 49p. Monografia.

ROSÁRIO C.R. **Modelo de Aplicação de Análise Multivariada, através de técnicas estatísticas PCA e HCA sobre questionário de Avaliação de Satisfação de Clientes: Estudo de uma caso de uma empresa metalúrgica de embalagens metálicas** Santa Cruz do Sul, Rs, 2012.

SANTANA, E. R. S. **Adição e Subtração: o suporte didático influencia a aprendizagem do estudante**. Universidade Estadual de Santa Cruz, Ilheus, Ba 2012, 235p.

SERGEI e SATYRO, **O impacto de infra-estrutura escolar na taxa de distorção idade-série das escolas brasileiras de ensino fundamental de 1998 a 2005**. Ipea, Rio de Janeiro, maio de 2008.

SILVA, G B, **Aplicação de análise exploratória de dados na discriminação geográfica do quiabo do RN/PE**, Editora Segmento, São Paulo/Sp, 2005

SIMÃO, A.F., **I Diretriz Brasileira de Prevenção Cardiovascular**. Arq Brasileira de Cardiologia, 2013: 101 (6 SUP 2) – 1-63.

SOKAL, A. L.; MICHENER, C. D. **A STATISTICAL METHOD FOR EVALUATING SYSTEMATIC RELATIONSHIPS**: Bulletin of the Society University os Kansas. N.38, p. 109-1438, 1958.

SOUZA, Maria Ester do Prado, **Família/Escola: A importância dessa relação no desempenho escolar**. Artigo PDE . Santo Antônio de Platina, Pr, 2009.

TEIXEIRA, Laila. **A influência do uso do celular nas aulas de Matemática**. Ed. Paulistana, São Paulo, SP, 2012.

APÊNDICE 1

FORMULÁRIO DE PESQUISA PARA ANÁLISE MULTIVARIADA

- 1) Cidade: _____
- 2) Bairro: _____
- 3) Rua: _____
- 4) Ponto de referência: _____
- 5) Escola: _____
- 6) Sexo: _____
- 7) Cor: _____
- 8) Turno: _____
- 9) Recebe ajuda nos exercícios de casa: sim () não ()
- 10) Idade: _____
- 11) Religião: _____
- 12) Trabalha: sim () não ()
- 13) Recebe bolsa família: sim () não ()
- 14) Estado civil: _____
- 15) Número de filhos: _____
- 16) Número de irmãos: _____
- 17) Pais são casados: sim () não ()
- 18) Renda familiar: _____
- 19) Casa: própria () alugada () outros ()
- 20) Escolaridade do pai: _____
- 21) Escolaridade da mãe: _____
- 22) Quantas pessoas moram em casa: _____
- 23) Possui microcomputador: sim () não ()
- 24) Possui celular: sim () não ()
- 25) Possui tablete: sim () não ()
- 26) Tem acesso a Internet: sim () não ()
- 27) Pretende fazer ensino médio: sim () não ()
- 28) Já foi reprovado: sim () não ()
- 29) Gosta da escola: sim () não ()
- 30) Pratica algum esporte: sim () não ()
- 31) Fuma: sim () não ()
- 32) Bebe: sim () não ()

Nome : _____