



UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIENCIAS EXATAS E TECNOLOGIA
DEPARTAMENTO DE ESTATISTICA E CIENCIAS ATUARIAIS



Lorena França Andrade

**INFLUÊNCIA DAS CARACTERÍSTICAS SOCIOECÔNOMICAS SOBRE O
DESEMPENHO NO ENEM EM CIÊNCIAS EXATAS**

São Cristóvão – SE

2019

Lorena França Andrade

**INFLUÊNCIA DAS CARACTERÍSTICAS SOCIOECÔNOMICAS SOBRE O
DESEMPENHO NO ENEM EM CIÊNCIAS EXATAS**

**Trabalho de Conclusão de Curso apresentado ao
Departamento de Estatística e Ciências Atuariais
da Universidade Federal de Sergipe, como parte
dos requisitos para obtenção do grau de
Bacharel em Estatística.**

Orientador: José Rodrigo Santos Silva

São Cristóvão – SE

2019

Lorena França Andrade

**INFLUÊNCIA DAS CARACTERÍSTICAS SOCIOECÔNOMICAS SOBRE O
DESEMPENHO NO ENEM EM CIÊNCIAS EXATAS**

Trabalho de Conclusão de Curso apresentado ao Departamento de Estatística e Ciências Atuariais da Universidade Federal de Sergipe, como um dos pré-requisitos para obtenção do grau de Bacharel em Estatística.

Aprovado em ____/____/____, Nota Final_____.

Banca Examinadora:

Prof. Dr. José Rodrigo Santos Silva – Orientador
Professor da Universidade Federal de Sergipe

Hérica Santos da Silva
Dra. em Biometria e Estatística Aplicada

Evelyn Souza Chagas Oliveira
Dra. em Biometria e Estatística Aplicada

AGRADECIMENTOS

A Deus.

Aos meus pais, pelo incentivo e apoio incondicional.

Aos meus amigos da turma sobreviventes: Jaciele, Jesy, Laudiane e Richel, pelo companheirismo (ninguém solta a mão de ninguém), por tornarem o terror que disseram ser a Universidade em maravilhosos 4 anos, obrigada!

Meus sinceros agradecimentos ao meu orientador Dr. José Rodrigo que se fez muito competente em suas atribuições e que me aturou durante todo meu percurso acadêmico (com uma calma nunca vista), confiando a mim seus projetos de pesquisa, compartilhando suas ideias e reflexões: “Lorena, primeiro você arruma o problema depois a solução”, e assim fui me enfiando em tanta coisa, e hoje só tenho a agradecer, sou sua fã! Agradeço também aos demais professores do Decat, alguns mais íntimos que outros, mas todos essenciais para que eu pudesse chegar até aqui.

E tão importante quanto a todos já mencionados, à Pró-Reitora de Extensão Dra. Alaíde Hermínia de Aguiar, pela oportunidade, por confiar em meu trabalho e pela compreensão da importância que está sendo esse momento para mim, gratidão!

RESUMO

O presente trabalho tem por finalidade avaliar a relação do desempenho dos estudantes do estado de Sergipe no Exame Nacional do Ensino Médio (ENEM) com suas condições socioeconômicas. Por conseguinte, com o intuito de expor alguns pontos que podem ser melhorados a partir do conhecimento desses perfis nas instituições em Sergipe sob a ótica dos inscritos. Para a análise foram utilizados os microdados do ENEM 2016, disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Foram extraídas técnicas de Análise Fatorial (com Rotação Varimax) e Modelos Lineares Generalizados (Família Binomial com ligação logit) na análise dos dados. Ao longo do trabalho mostrou-se que no quesito matemática, é possível notar que os parâmetros apresentaram significância ($p < 0,001$), o que mostra a existência de associação significativa entre as variáveis (gênero, faixa etária, estado civil, raça, renda, computador, internet, abandono/reprovação, tipo de escola, turno, trabalha, local de residência, nível de escolaridade dos pais) e a nota de matemática no Enem 2016. Admitiu-se como desempenho satisfatório os estudantes que possuíram nota acima da mediana do Brasil (450 pontos). Foram selecionados 6 fatores, com a manutenção de 63,48% da variação total. Os resultados apontaram que o fator 1 (residentes em Aracaju, pai ou mãe com ensino superior, alta renda familiar e que nunca estudaram em escola pública), fator 3 (candidatos brancos ou amarelos) e fator 5 (estudantes oriundos do ensino regular, que estudaram em turno diurno e que nunca abandonaram ou reprovaram) funcionaram como fatores protetores ao mau desempenho na prova (Odds Ratio, respectivamente igual a 0,67, 0,95 e 0,87), enquanto que o fator 2 (candidatos com maior idade, casados ou em união estável, que já concluíram o ensino médio e que trabalham), fator 4 (candidatos cujos pais possuem ensino fundamental ou médio apenas) e fator 6 (pessoas do sexo feminino e pessoas sem acesso a computador ou internet) foram verificadas como fatores de risco ao desempenho satisfatório em Ciências Exatas e Tecnológicas (Odds Ratio, respectivamente, 1,05, 1,03 e 1,47).

Palavras Chaves: ENEM, Ciências Exatas e Suas Tecnologias, desempenho educacional.

ABSTRACT

This paper aims to evaluate the relationship between the performance of students from Sergipe state in ENEM and their socioeconomic conditions. Therefore, in order to expose some points that can be improved from the knowledge of these profiles in the institutions in Sergipe from the perspective of the subscribers. For the analysis we used the ENEM 2016 microdata, made available by INEP. Techniques of factor analysis (with Varimax rotation) and generalized linear models (binomial family with logit connection) were extracted in the data analysis. Throughout the work it was shown that in the mathematical question, it is possible to notice that the parameters presented significance ($p < 0,001$), which shows the existence of significant association between these variables and the mathematics grade in Enem 2016. It was admitted as satisfactory performance students who scored above the median (450 points). Six factors were selected, maintaining 63.48% of the total variation. The results showed that factor 1 (residents in Aracaju, father or mother with higher education, high family income and who never studied in a public school), factor 3 (white or yellow candidates) and factor 5 (students from regular education, who day shift and who never dropped out or failed) functioned as protective factors for poor test performance (Odds Ratio, respectively equal to 0.67, 0.95 and 0.87), while factor 2 (older candidates , married or in stable union, who have already completed high school and who work), factor 4 (candidates whose parents have only elementary or high school) and factor 6 (female and those without access to computer or internet) were verified as risk factors for satisfactory performance in Exact and Technological Sciences (Odds Ratio, respectively, 1.05, 1.03 and 1.47).

Keywords: ENEM, Exact Sciences and Their Technologies, educational performance

LISTA DE ILUSTRAÇÃO

Figura 1: Evolução do número de inscritos em milhões confirmados no ENEM de 1998 a 2016.	17
Figura 2: Modelo Conceitual dos Fatores Associados ao Desempenho Escolar.....	22
Figura 3: Modelos de distribuição Normal.....	28
Figura 4: Etapas da aplicação da análise de componentes principais.....	33
Figura 5: Ilustração das variâncias de três itens e suas relações com um fator hipotético.	37
Figura 6: Exemplo do efeito da rotação (ortogonal, oblíqua) de dois fatores.	39
Figura 7: Screeplot dos autovalores dos componentes principais.	48

LISTA DE QUADRO

Quadro 1: Proposta tridimensional de Matriz de Referência.	19
Quadro 2: Matriz de referência de Matemática e Suas Tecnologias.	19
Quadro 3: Adequação Amostral de acordo com a medida KMO.....	38

LISTA DE TABELAS

Tabela 1: Perfil dos avaliados no Enem, prova de matemática, Sergipe, 2016.....	455
Tabela 2: Candidatos distribuídos por pontuação, dos avaliados no Enem, prova de matemática, Sergipe, 2016.....	466
Tabela 3 – Análise de componentes principais.	477
Tabela 4 – Análise de fatores com rotação Varimax, comunalidade e variância específicas referentes às variáveis explicativas.....	500
Tabela 5 - Ajuste da regressão logística para o desempenho insatisfatório em matemática na prova do ENEM, Sergipe, 2016.....	511

Sumário

1. INTRODUÇÃO.....	10
2. OBJETIVOS	13
2.1 Geral.....	13
2.2 Específicos	13
3. REVISÃO LITERARIA.....	14
3.1 Ensino-Aprendizagem da Matemática no Ensino Médio	14
3.2 Exame Nacional do Ensino Médio (ENEM)	16
3.3 Fatores Associados ao Desempenho Escolar	21
4. METODOLOGIA.....	27
4.1 Teste De Hipótese	28
4.1.1 Shapiro-Wilk para Normalidade	28
4.1.2 U de Mann-Whitney	29
4.1.3 U de Kruskal-Wallis.....	31
4.2 Análise Multivariada.....	32
4.2.1 Análise de Componentes Principais (ACP)	32
4.2.2 Análise Fatorial (AF)	36
4.3 Modelos Lineares Generalizados	40
4.3.1 Família Exponencial.....	40
4.3.2 Modelo Binomial.....	41
4.3.3 Estimativa de β	42
5. RESULTADOS E DISCUSSÕES	43
6. CONCLUSÕES	54
REFERÊNCIAS	55

1. INTRODUÇÃO

O ensino médio é a última etapa da educação básica e tem por finalidade, segundo a Lei de Diretrizes e Bases da Educação (LDB) de 1996, consolidar e aprofundar os conhecimentos obtidos durante o ensino fundamental, garantindo um futuro profissional e o preparo para ser um cidadão ético, intelecto, com pensamento crítico e compreensão dos fundamentos científico tecnológicos dos processos produtivos, articulando a teoria com a prática no ensino de cada disciplina.

Continuamente escutamos por parte dos estudantes questionamentos sobre o porquê de estudar determinado assunto ou disciplina. Isso se dá devido toda a trajetória o estudante ser orientado para realizar um caminho formativo, e uma das vias deste caminho é feita pela escolha de pleitear uma vaga em instituições de ensino superior público ou privado.

O alcance deste propósito dar-se através do Exame Nacional do Ensino Médio (ENEM), cujo resultado pode ser utilizado como mecanismo único, alternativo ou complementar para acesso à educação superior, a depender da universidade.

Como base nas alterações do ensino médio, o ENEM foi criado em 1998, através da Portaria Ministerial nº 348, de 28 de maio de 1998, do Ministério da Educação e Desportos (BRASIL, 1998).

Em 2009, o ENEM é reformulado com a proposta de unificar os concursos vestibulares das universidades federais brasileiras. Em 2015, o ENEM fez parte do processo de ingresso nas 63 universidades federais e 48 delas substituíram completamente seus vestibulares pelo Exame (SANTOS, 2014). A importância do ENEM foi aumentando gradativamente por toda extensão longitudinal, passando de 157 mil inscritos, na primeira edição, para mais de 8,6 milhões em 2016 (INEP, 2016).

Aplicado anualmente o Exame “[...] tinha como objetivo fornecer informações sobre estratos específicos para ações do poder público e disponibilizar informações aos estudantes, para que eles mesmos avaliassem seu desempenho em comparação com os dados gerais, e não se voltar para a avaliação individual” (VIGGIANO e MATTOS, 2013, p.420).

Segundo Lopes e Péres (2010) o Enem além de servir como porta para entrada no ensino superior, serve como controle dos currículos da educação básica, e do desempenho da educação.

O ENEM examina através de redação e quatro áreas do conhecimento, incluindo a Matemática, o desempenho dos concluintes do ensino médio. Seguido por um instrumento de coleta socioeconômica de caráter obrigatório no ato da inscrição, possibilitando equiparar entre diferentes sistemas de ensino, regiões e estratos sociais, contextualizando o desempenho dos sujeitos, que participam desse certame, em situações pessoais, familiares, trajetória escolar, condição econômica, experiência de trabalho, como também nas percepções sobre a escola (GREMAUD et al., 2007; INEP, 2009).

Palermo et al (2011) definiu três níveis de análise para a caracterização do desfecho. O nível do aluno (características individuais: aspectos pessoais e familiares), nível da turma (características de sala de aula: composição dos pares, da turma, do professor, estilos e práticas pedagógicas) e nível da escola (características do estabelecimento: fatores sociais, composição de alunos e professores, características físicas, clima escolar, perfil do diretor, políticas e práticas escolares).

Por conseguinte, considerando a magnitude deste Exame, que se propõe a medir a qualidade da educação básica no seu término e ainda apresentar-se como porta de acesso ao ensino superior, manifesta-se o conteúdo central deste trabalho: **Qual é de fato a influência que as variáveis socioeconômicas têm sobre o desempenho no ENEM em Ciências Exatas de Sergipe?**

Sendo assim, visando à construção do real perfil socioeconômico dos candidatos à realização do ENEM e com o intuito de expor alguns pontos que podem ser melhorados a partir do conhecimento desses perfis nas instituições em Sergipe sob a ótica dos inscritos, realizou-se esta pesquisa para análise dos dados disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Anísio Teixeira (INEP) relativo ao ENEM do ano de 2016, com a finalidade de fundamentar a tomada de decisões, imprescindíveis para um processo de melhoramento contínuo.

O aspecto motivacional para a delimitação do tema foi a dificuldade em encontrar estudos que abordam a associação entre desempenho nessa prova de matemática e fatores socioeconômico associados ao estado de Sergipe, embora a literatura apresente diversos fatores ligado ao desempenho escolar de outros estados.

É evidente que estatísticas estão ficando cada vez mais indispensáveis para o sistema de informação e tem sido bastante utilizada em pesquisas científicas nas mais

variadas áreas do conhecimento, incluindo a área educacional. Alguns trabalhos na área da educação (ALMEIDA and PIMENTEL, 2010; BAKER et al. 2011) investigam base de dados provenientes do processo de ensino aprendizagem.

Espera-se, por fim, que esse trabalho contribua para subsidiar os gestores públicos na elaboração de programas e projetos, bem como incentivar novas pesquisas na área, traçar planos sociais e econômicos para o futuro de modo a que se possa cumprir uma função vital da avaliação.

A partir deste contexto, pensou-se em estruturar todo o trabalho em seis seções. A presente seção que explica de forma resumida o assunto qual será desenvolvido e as subsequentes que se dividem em: Objetivos, Revisão de Literatura, Metodologia, Resultados e Conclusões.

Na seção 2, os objetivos serão expostos de modo que o leitor compreenda a intenção deste trabalho, delimitando qual será o escopo do mesmo.

A seção 3 apresenta o referencial teórico, que é uma grande explanação sobre o tema com base em resultados encontrados por outros autores. Onde discutiremos sobre os aspectos institucional do Ensino Médio; a contextualização do Enem (principais características, competências, habilidades); a Matemática como forma de saber escolar no Brasil, a matemática no Enem, por fim as situações-problemas sobre a aprendizagem.

Na seção 4, serão apresentados os dados a serem estudados, os métodos aplicados para obtenção dos resultados que serão posteriormente apresentados e explicados na seção 5.

E por fim as conclusões na seção 6, que terá a função de arrematar todo o conteúdo que foi desenvolvido ao longo deste trabalho.

2. OBJETIVOS

2.1 Geral

A questão principal deste trabalho é explorar a relação que as variáveis socioeconômicas têm sobre o desempenho dos participantes do ENEM na área de Ciências Exatas e suas Tecnologias.

2.2 Específicos

- Caracterizar o perfil sociodemográfico dos estudantes que fizeram o ENEM;
- Relacionar o perfil sociodemográfico dos estudantes que fizeram o ENEM com o desempenho na prova de Ciências Exatas.

3. REVISÃO LITERARIA

3.1 Ensino-Aprendizagem da Matemática no Ensino Médio

Configurado na LDB (Lei nº 9394/96) como a última etapa da educação básica, o ensino médio corresponde apenas três anos de toda a trajetória (BRASIL, 1996a). Sendo assim, tem como finalidade consolidar e aprofundar os conhecimentos adquiridos no Ensino Fundamental, além de possibilitar o prosseguimento de estudos, em outras palavras, é a etapa transitória entre a educação fundamental e a educação superior.

O ensino médio está organizado em quatro áreas de conhecimento, cada qual com as suas competências específicas, que serão aplicadas no contexto formativo, são elas: Linguagens e Suas Tecnologias, Ciências da Natureza e Suas Tecnologias, Ciências Humanas e Suas Tecnologias e a Matemática e Suas Tecnologias. A cada uma das competências são relacionadas habilidades, que indicam as aprendizagens essenciais aos estudantes do Ensino Médio (BRASIL, 2017b). Essa organização curricular se deu a partir de um profundo estudo com equipes técnicas dos Sistemas Estaduais de Educação, professores e alunos da rede pública e representantes da comunidade acadêmica, com o intuito de oferecer uma educação básica de qualidade (BRASIL, 2018).

No caso particular da área da Matemática e suas Tecnologias elegeu três grandes competências como objetivos a serem perseguidos durante essa etapa da escolaridade básica e complementar do ensino fundamental para todos os brasileiros, são elas (BRASIL, 2006):

- representação e comunicação: que envolvem a leitura, a interpretação e a produção de textos nas diversas linguagens e formas textuais características dessa área do conhecimento;
- investigação e compreensão: competência marcada pela capacidade de enfrentamento e resolução de situações-problema, utilização dos conceitos e procedimentos peculiares do fazer e pensar das ciências;
- contextualização sociocultural: na forma de análise crítica das ideias e dos recursos da área e das questões do mundo que podem ser respondidas ou transformadas por meio do pensar e do conhecimento científico.

Com base nas competências descritas a proposta de Matemática dos Parâmetros Curriculares Nacionais para o Ensino Médio (PNCEM) é que cada escola e grupo de professores proponham um trabalho pedagógico que permita o desenvolvimento das competências almejadas (BRASIL, 2006).

Agregando o currículo do Ensino Médio deve-se garantir também espaço para que os alunos possam fazer “[...] interpretação da realidade e possa estar melhor preparado para sua inserção no mundo do conhecimento e do trabalho” (BRASIL, 2006).

Skovsmose (2001) discute a respeito do papel que a Matemática tem na sociedade, afirmando que a matemática interfere na realidade e tem o poder de “formatar nossa sociedade”. Defende ainda, que a escola deve capacitar os alunos para analisar as informações referentes à matemática com os quais se depara.

É de grande importância para o aluno perceber essa relação da matemática com o cotidiano em seu modo mais amplo, visto que a matemática desenvolve o raciocínio, garante uma forma de pensar, possibilita a criação e o aperfeiçoamento de ideias, algo que está diretamente relacionado com as atitudes que o aluno terá em sociedade (RODRIGUES, 2004). Esta ideia é assentida por Brasil (2006):

Espera-se que os alunos saibam usar a Matemática para resolver problemas práticos do cotidiano; para modelar fenômenos em outras áreas do conhecimento; compreendam que a Matemática é uma ciência com características próprias, que se organiza via teoremas e demonstrações; percebem a Matemática como um conhecimento social e historicamente construído; saibam apreciar a importância da Matemática no desenvolvimento científico e tecnológico (BRASIL, 2006, p.69).

Aprender e ensinar matemática requer mais que memorização e repetição, requisita-se que haja dinamismo e seja favorecedor do desenvolvimento do pensamento matemático como significado prático e coerente. E de acordo com D’ambrosio (2001) e Nacarato et al. (2009) a forma qual está sendo ensinada ainda é contestável.

Segundo Bessa (2007) e Toledo (2009) essa dificuldade no processo de ensino e aprendizagem da Matemática podem estar ligadas a diversos fatores, ao professor quando pensado em metodologias e práticas pedagógicas, ao aluno quanto ao desinteresse pela disciplina, à escola por não estimular o aprendizado e/ou por condições físicas, e ainda à família por falta de suporte e /ou falta de condições.

Silveira (2002) denota que existe um julgamento pré-constituído dos alunos sobre a matemática, que os leva a inferir o nível de dificuldade que a disciplina pode ter, fazendo com que o aluno tenha apatia tanto pelo professor, quanto pela disciplina, dificultando assim sua aprendizagem. Neste mesmo sentido Bicudo e Chamie (1994) em seu estudo identificou depoimentos de estudantes de ensino médio que dizem “O que eu acho ruim na Matemática são as fórmulas que temos que decorar (seno, cosseno, área, delta, etc.) muitas vezes sem entender como esta fórmula foi feita...”. Farago (2003) afirma que conhecer a história da matemática é fundamental para que se possa realizar então uma boa prática, facilitando assim o ensino-aprendizagem.

Para Sacramento (2008) os estudantes não compreendem a matemática que é ensinada na escola, e este motivo faz com que o cresça o número de reprovações e evasões. E ainda segundo Marques (2011) menos da metade dos alunos faz essa trajetória escolar (ensino médio-universidade) não sabem nem a matemática básica.

3.2 Exame Nacional do Ensino Médio (ENEM)

O ENEM, implantado em 1998, através da Portaria Ministerial nº 348, de 28 de maio de 1998 do Ministério da Educação e Desportos, e segundo Santos (2011) propõe mensurar o desempenho do aluno ao término da escolaridade básica através de modalidades estruturais da inteligência, denominadas competências, garantindo a certificação do mesmo, e auxiliando na elaboração de políticas educacionais e melhorias na educação brasileira (FERREIRA, 2014).

Segundo o Documento Básico do ENEM (DBE), de 2002, pretende-se atingir os seguintes objetivos:

- a. oferecer uma referência para que cada cidadão possa proceder a sua autoavaliação com vista às suas escolhas futuras, tanto em relação ao mercado de trabalho quanto em relação à continuidade de estudos;
- b. estruturar uma avaliação da educação básica que sirva como modalidade alternativa ou complementar aos processos de seleção nos diferentes setores do mundo do trabalho;
- c. estruturar uma avaliação da educação básica que sirva como modalidade alternativa ou complementar aos exames de acesso aos cursos profissionalizantes pós-médios e ao ensino superior (BRASIL, 2002, p.7-8).

Desde a sua implantação o ENEM era constituído por 63 questões de múltipla escolha e uma redação. As questões eram propostas como situações-problema ancoradas

na interdisciplinaridade e na contextualização, princípios curriculares dos PCN para o ensino médio. A matriz descrevia cinco competências gerais expressas no total de 21 habilidades (VIGGIANO e MATTOS, 2013).

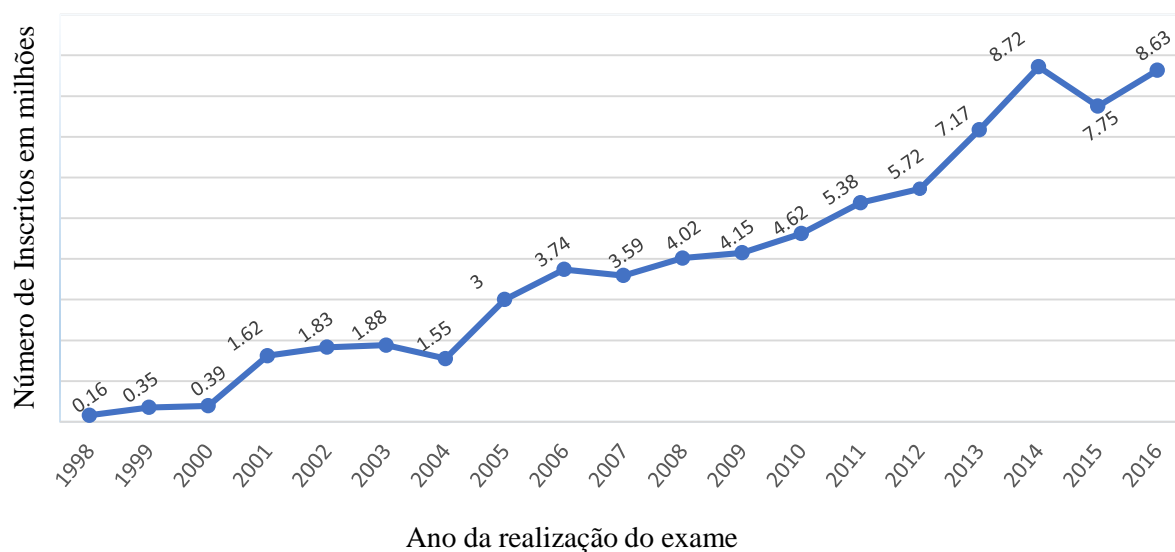
Durante a trajetória do ENEM referente ao número de inscritos dois grandes avanços ocorreram. O primeiro se deu entre 2000 e 2001 quando o governo adotou a isenção de pagamento da taxa de inscrição aos estudantes concluintes do ensino médio da rede pública (INEP, 2019). O segundo ocorreu entre 2004 e 2005 devido à criação do Programa Universidade para Todos (Prouni) através da medida provisória nº 213, de 10 de setembro, quando o Ministério da Educação vinculou a concessão de bolsas para os candidatos que havia cursado todo o ensino médio na rede pública com a nota do ENEM (SANTIAGO, 2012). Em 2005 esta medida virou Lei nº 11.096 e disponibilizou o artigo 3º que diz:

Art. 3º O estudante a ser beneficiado pelo PROUNI será pré-selecionado pelos resultados e pelo perfil socioeconômico do Exame Nacional do Ensino Médio – ENEM ou outros critérios a serem definidos pelo Ministério da Educação, e na etapa final, selecionado pela instituição de ensino superior, segundo seus próprios critérios, a qual competirá, também, aferir as informações prestadas pelo candidato (BRASIL, 2005, p. 1).

Devido essa nova demanda e supondo maiores avanços, em 2009 o ENEM anunciou mudanças, passando a ser a ferramenta de entrada parcial ou total nas universidades públicas e privadas, através do Sistema de Seleção Unificada (SISU) para universidades públicas, bem como pelo Programa de Universidade para Todos (PROUNI) e pelo Fundo de Financiamento ao Estudante do Ensino Superior (FIES) para faculdades privadas, e sua prova passa a ser constituída por 180 questões, dividida em quatro áreas, Ciências humanas e suas tecnologias, Ciências da Natureza e suas Tecnologias e redação, Linguagens e códigos e suas Tecnologias, Matemática e suas tecnologias e uma redação (VIGGIANO e MATTOS, 2013; SAPATINI, 2014).

A Figura 1 apresenta esse aumento no número de inscritos (em milhões), sob comparação do antigo ENEM até 2008 e o novo ENEM a partir do ano de alteração da matriz de referência 2009.

Figura 1: Evolução do número de inscritos em milhões confirmados no ENEM de 1998 a 2016.



Fonte: Reprodução/Inep-MEC

Segundo Ferreira (2014) essa mudança no exame em 2009 é composta por ações e operações mentais que todos os jovens e adultos deveriam desenvolver como recursos mínimos. A partir da mudança, cada área de conhecimento passa a possuir sua matriz de referências com suas próprias competências e habilidades com base em cinco eixos cognitivos que são (MIRANDA, 2010; FERREIRA 2014):

- I) **Dominar linguagens:** dominar a norma culta da Língua Portuguesa e fazer uso das linguagens matemática, artística e científica e das línguas espanhola e inglesa.
- II) **Compreender fenômenos:** construir e aplicar conceitos das várias áreas do conhecimento para a compreensão de fenômenos naturais, de processos históricos, geográficos, da produção tecnológica e das manifestações artísticas.
- III) **Enfrentar situações problemas:** selecionar, organizar, relacionar, interpretar dados e informações representadas de diferentes formas, para tomar decisões e enfrentar situações-problema.
- IV) **Construir argumentação:** relacionar informações, representadas em diferentes formas, e conhecimentos disponíveis em situações concretas, para construir argumentação consistente.

- V) **Elaborar propostas:** recorrer aos conhecimentos desenvolvidos na escola para elaboração de propostas de intervenção solidária na realidade, respeitando os valores humanos e considerando a diversidade sociocultural.

Com base nesses cinco eixos cognitivos apresentados, a matriz de referência do novo ENEM é composta por 30 competências e 120 habilidades, sendo 30 habilidades para cada área de conhecimento (FERREIRA, 2014), incluindo a Matemática e Suas Tecnologias, que pode ser distribuída na forma:

Quadro 1: Proposta tridimensional de Matriz de Referência.

Competências de Matemática e suas Tecnologias	Dominar linguagens (DL)	Compreender fenômenos (CF)	Enfrentar situação problema (SP)	Construir argumentação (CA)	Elaborar propostas (EP)
Competência de área 1	H1	H2	H3	H4	H5
Competência de área 2	H6	H7	H8	H9	-
Competência de área 3	H10	H11	H12	H13	H14
Competência de área 4	-	H15	H16	H17	H18
Competência de área 5	H19	H20	H21	H22	H23
Competência de área 6	-	-	H24	H25	H26
Competência de área 7	-	H27	H28	H29	H30

Fonte: Rabelo, 2013.

Uma vez que as habilidades são descritas por:

Quadro 2: Matriz de referência de Matemática e Suas Tecnologias.

Descritor	Descrição das Habilidades
Construir significados para os números naturais, inteiros, racionais e reais.	<p>H1 - Reconhecer, no contexto social, diferentes significados e representações dos números e operações - naturais, inteiros, racionais ou reais.</p> <p>H2 - Identificar padrões numéricos ou princípios de contagem.</p> <p>H3 - Resolver situação-problema envolvendo conhecimentos numéricos.</p> <p>H4 - Avaliar a razoabilidade de um resultado numérico na construção de argumentos sobre afirmações quantitativas.</p> <p>H5 - Avaliar propostas de intervenção na realidade utilizando conhecimentos numéricos.</p>

Utilizar o conhecimento geométrico para realizar a leitura e a representação da realidade e agir sobre ela.	<p>H6 - Interpretar a localização e a movimentação de pessoas/objetos no espaço tridimensional e sua representação no espaço bidimensional.</p> <p>H7 - Identificar características de figuras planas ou espaciais.</p> <p>H8 - Resolver situação-problema que envolva conhecimentos geométricos de espaço e forma.</p> <p>H9 - Utilizar conhecimentos geométricos de espaço e forma na seleção de argumentos propostos como solução de problemas do cotidiano.</p>
Construir noções de grandezas e medidas para a compreensão da realidade e a solução de problemas do cotidiano.	<p>H10 - Identificar relações entre grandezas e unidades de medida.</p> <p>H11 - Utilizar a noção de escalas na leitura de representação de situação do cotidiano.</p> <p>H12 - Resolver situação-problema que envolva medidas de grandezas.</p> <p>H13 - Avaliar o resultado de uma medição na construção de um argumento consistente.</p> <p>H14 - Avaliar proposta de intervenção na realidade utilizando conhecimentos geométricos relacionados a grandezas e medidas.</p>
Construir noções de variação de grandezas para a compreensão da realidade e a solução de problemas do cotidiano.	<p>H15 - Identificar a relação de dependência entre grandezas.</p> <p>H16 - Resolver situação-problema envolvendo a variação de grandezas, direta ou inversamente proporcionais.</p> <p>H17 - Analisar informações envolvendo a variação de grandezas como recurso para a construção de argumentação.</p> <p>H18 - Avaliar propostas de intervenção na realidade envolvendo variação de grandezas.</p>
Modelar e resolver problemas que envolvem variáveis socioeconômicas ou técnico-científicas, usando representações algébricas.	<p>H19 - Identificar representações algébricas que expressem a relação entre grandezas.</p> <p>H20 - Interpretar gráfico cartesiano que represente relações entre grandezas.</p> <p>H21 - Resolver situação-problema cuja modelagem envolva conhecimentos algébricos.</p> <p>H22 - Utilizar conhecimentos algébricos/geométricos como recurso para a construção de argumentação.</p> <p>H23 - Avaliar propostas de intervenção na realidade utilizando conhecimentos algébricos.</p>
Interpretar informações de natureza científica e social obtidas da leitura de gráficos e tabelas, realizando previsão de tendência, extrapolação, interpolação e interpretação.	<p>H24 - Utilizar informações expressas em gráficos ou tabelas para fazer inferências.</p> <p>H25 - Resolver problema com dados apresentados em tabelas ou gráficos.</p> <p>H26 - Analisar informações expressas em gráficos ou tabelas como recurso para a construção de argumentos.</p>
Compreender o caráter aleatório e não-determinístico	H27 - Calcular medidas de tendência central ou de dispersão de um conjunto de dados expressos em uma tabela de frequências de

dos fenômenos naturais e sociais e utilizar instrumentos adequados para medidas, determinação de amostras e cálculos de probabilidade para interpretar informações de variáveis apresentadas em uma distribuição estatística.	<p>dados agrupados (não em classes) ou em gráficos.</p> <p>H28 - Resolver situação-problema que envolva conhecimentos de estatística e probabilidade.</p> <p>H29 - Utilizar conhecimentos de estatística e probabilidade como recurso para a construção de argumentação.</p> <p>H30 - Avaliar propostas de intervenção na realidade utilizando conhecimentos de estatística e probabilidade</p>
---	--

Fonte: INEP, 2009.

Em 2017, o Exame passou ainda por uma nova mudança relativa aos dias de aplicação da prova. A partir de consulta pública, no ano citado, a prova que costumava ser aplicada em dois dias de um mesmo final de semana, passa a ser aplicada em dois domingos consecutivos, mantendo a estrutura de número de questões por área e a aplicação da redação (INEP, 2019). Essas alterações, na estrutura da prova e dias de aplicação, foram respaldadas pela Portaria MEC 109/09.

Embora o ENEM disponibilize o número de vagas de forma exponencialmente, os candidatos ainda se deparam com outras situações sempre existente na educação brasileira fazendo com que muitos deles não consigam adentrar nos cursos almejados.

Uma forma que o exame alcançou para a extração de informações sobre essas supostas situações foi através da inclusão do questionário socioeconômico de preenchimento obrigatório durante o ato da inscrição. Em 2014, esse questionário passa a ser utilizado para contextualizar as medidas de desempenho dos alunos da educação básica (INEP, 2014).

3.3 Fatores Associados ao Desempenho Escolar

Para Barbosa e Fernandes (2001), Scorzafave e Ferreira (2011) o desempenho escolar do aluno é resultado de uma diversidade de fatores que representam vários níveis da inserção social, e que merecem ser discutidos. Soares e Alves (2003) afirmam que “essas desigualdades são moduladas por filtros socioeconômicos, raciais, localização (urbana, rural) e por tipo de rede escolar (pública, particular)”. Nessa perspectiva, Soares (2004) aconselha ainda que analise esses fatores associados às

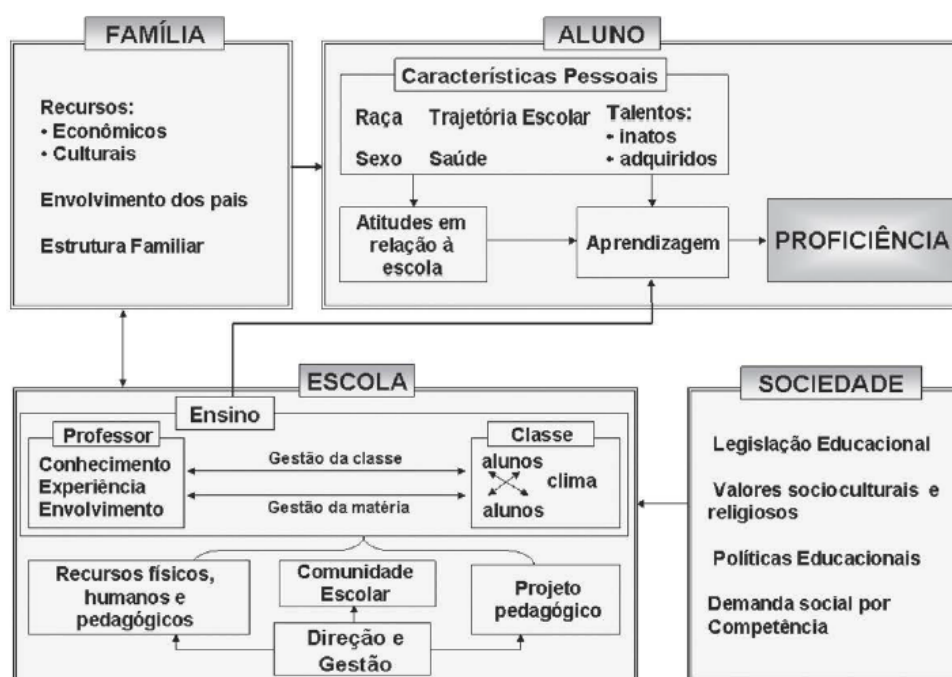
condições socioeconômicas levando em consideração aspectos familiares, do próprio aluno e do âmbito social no qual ele está inserido.

Melo e Arakawa (2012), definem pela existência de uma relação direta entre a escolaridade, a renda familiar e o desempenho no ensino, e observam que esta relação é mais clara nas regiões com baixo desenvolvimento. Sendo assim, afirmam que as diferenças regionais também são fatores que podem ser levados em consideração no que relaciona à persistência intergeracional educacional.

Partindo-se dessas constatações, diferentes estudos corroboram para a obtenção de fatores internos e externos à escola associados a esse desempenho. Dentre os estudos que investigaram esses fatores ao desempenho, a variável gênero foi uma das que mais se fez presente.

Segundo Menezes Filho (2007) e Luz (2006), os meninos tendem a ter um desempenho superior às meninas durante todo o seu percurso educacional. De acordo com o trabalho feito por Andrade (2007) utilizando a base de dados do Sistema de Avaliação da Educação Básica (Saeb) de 2001 para estudantes do 3º ano do ensino médio, foram encontrados resultados que confirmam que meninos atingem em média 5,92 pontos a mais que as meninas em matemática.

Figura 2: Modelo Conceitual dos Fatores Associados ao Desempenho Escolar



Fonte: Soares (2004, p.85)

Para Brito (1996) a concepção da Matemática como um domínio masculino é cultural:

“Culturalmente, são atribuídos “rótulos” aos indivíduos e afirmações que não são confirmadas através de pesquisas passam a ser consideradas como verdadeiras. Assim, cristalizou-se a idéia que a habilidade verbal é uma característica feminina e a habilidade matemática é uma característica masculina. Dentro desta concepção, os homens deveriam apresentar alta habilidade matemática e baixa habilidade verbal enquanto as mulheres apresentariam alta habilidade verbal e baixa habilidade matemática”. (Brito, 1996, p.75).

Outro fator achado em literaturas é referente ao grupo étnico. Essa variável é de grande impacto quando se refere a desempenho escolar, segundo estudiosos. Ferrão et al. (2001) e Menezes Filho (2007) destacam em seus estudos que os candidatos autodeclarados de cor preta apresentam resultados inferior que os demais, e ainda que os candidatos brancos têm um desempenho significativamente melhor que os negros, mas não que os pardos.

Quando referido à idade, existe uma afirmativa de que o desenvolvimento das habilidades intelectuais é relacionado com faixa etária (Anastasi, Urbina, 2000; Angelini, Alves, Custódio, Duarte, & Duarte, 1999). Segundo Seabra e Santos (2007) a idade é um influenciador do desempenho matemático, seu estudo mostrou que estudantes com idade abaixo de 20 anos tendem a ter um desempenho melhor que os com mais de 20 anos. Melo e Arakawa (2012) acredita que os concluintes do ensino médio podem estar em desvantagem quando comparados aos candidatos mais velhos. O mesmo supõe que após a saída do ensino médio os candidatos podem ter estendido seus estudos ou terem adquirido conhecimentos profissionais. Em contrapartida, acredita ser possível que o tempo que passou desde o término do ensino médio atue contra os participantes mais velhos, em caso de após a conclusão os mesmos não deram continuidade aos estudos possivelmente podem estar menos qualificados para prestar o exame.

Outro fator interessante para essa dependência é o estado civil. Para Masasi (2012), o estado civil do estudante não influencia no desempenho acadêmico. Por vivência, acredita-se que quando o estudante é solteiro tem menos responsabilidades,

consequentemente tem mais tempo livre para usar a favor do seu aprendizado. Andrade e Corrar (2008) afirmam que os solteiros têm melhor desempenho acadêmico. Nesse mesmo seguimento o Ibope (2011) em sua pesquisa afirma que os solteiros têm desempenho superior aos casados, e um dos motivos seria a preocupação que os mesmos tem com relação em construir uma futura carreira. É possível que o estudante quando casado tenha outros tipos de responsabilidades (filhos, casa, trabalho) e por isso tenham um desempenho inferior.

Em pesquisa, Melo e Arakawa (2012) asseguram que as responsabilidades fora de casa podem ser um fator influenciável negativamente ao desempenho. Luz (2006) fortalece a ideia quando confirma que o aluno que trabalha fora de casa tende a perda de proficiência. Sendo assim, sabe-se que uns dos motivos do início precoce no mercado de trabalho é recorrente da necessidade do capital financeiro de muitas famílias. Nessa perspectiva, Soares e Collares (2006) orientam que o fator relacionado ao capital financeiro seja “a primeira característica que deve ser considerada em estudos da influência da família no desempenho do aluno, principalmente em países como o Brasil, onde prevalecem altos graus de desigualdade de renda”. A literatura indica que a questão financeira está profundamente relacionada com a escolaridade dos pais (BARROS e MENDONÇA, 1997; BARROS et al., 2001).

Melo e Arakawa (2012) e Barbosa e Souza (2014) encontram evidências de que o nível de escolaridade dos pais está correlacionado positivamente ao de seus filhos. Em pesquisa sobre o assunto, Kleinke (2017, p. 45) concluiu:

Ao analisar candidatos cujos pais concluíram o ensino médio, verificamos diferenças de comportamentos entre pai e mãe: filhos de mães que possuem ensino médio completo têm um acréscimo na participação no decil superior; no caso dos pais há uma queda na transição. Esse resultado pode estar relacionado ao fato de o processo educacional familiar ainda ser predominantemente responsabilidade das mães, implicando maior interação destas — e de seu conhecimento — com os filhos. Esses dados reforçam a escolha da escolaridade da mãe como um indicador de capital cultural dos candidatos.

Inúmeros são os fatores associados à desvantagem social. Para Ferrão (2003), a desvantagem social é o responsável pela dificuldade no aprendizado, pois, aqueles cuja

condição social é baixa tendem a ter uma habitação, alimentação e saúde ruim, o que em conjunto pode tornar um estudante fisicamente fraco e abaladamente emocional.

Ainda assim, Andrade (2007) levanta a hipótese de que mesmo o aluno com renda baixa venha a estudar em uma escola de bom desempenho, este aluno tende a ter um desempenho superior comparado ao estudante de renda baixa de uma escola com poucos recursos.

Seguido o pensamento de Andrade (2007) informações a respeito das redes de ensino do caráter público ou privado são de grande importância também para o desempenho. Macedo (2004) em um estudo longitudinal confirmou que os estudantes de escolas particulares possuem rendimento superior que os de escolas públicas. Luz (2006) em seus estudos afirma ter encontrado diferença dentre os estudantes das escolas públicas, onde as escolas administradas pelo estado têm o desempenho melhor que as administradas pelo município. Segundo Soares (2004), seja a escola pública ou privada aquela que não possui bons recursos (humanos e financeiros) não tende a ser boas.

Uma forma que as instituições podem se fazer presente em relação a esses recursos financeiros é com a inclusão da tecnologia. Desde que se descobriu que computadores e internet também podem ser usados na educação, os pesquisadores começaram a realizar investigações para identificar qual é a sua relevância para o desenvolvimento do ensino e da aprendizagem. De acordo com os estudos da OECD (2005), o aluno que não tem acesso a este recurso está em franca desvantagem, tanto em termos de conhecimento de tecnologia quanto no seu desempenho em Matemática. A presença de computadores em casa e nas escolas públicas tem impacto positivo sobre o desempenho dos alunos. Mas se utilizado esses recursos para entretenimento não se obterá o mesmo resultado (GRACIOSO, 2006; DELPORTO & FERREIRA, 2007; MENEZES FILHO, 2007).

Outro fator que está associado ao desempenho escolar é a trajetória do aluno. As pesquisas educacionais, de maneira geral, apontam associação aos efeitos negativos sob esse fator (abandono e/ou repetência), sendo que as perdas tendem a se acentuar a cada ano perdido, levando a defasagem da idade-série. (BARBOSA & FERNANDES, 2001; ALBERNAZ, FERREIRA & FRANCO, 2002)

Simultaneamente Menezes Filho (2007) diz que o término tardio e o histórico de reprovação são fatores deprimentes para o desempenho escolar, mas que quando

encontram-se em série correta ou adiantada os estudantes tem um desempenho superior aos demais.

Por fim e sucedendo esse entendimento, estudos mostram que o fator trajetória é afetado pela localização espacial da escola (ALVES; LANGE; BONAMINO, 2010). Então por sua vez podemos concluir indiretamente que o local em que a escola está situada também é um fator que influencia negativamente o desempenho.

4. METODOLOGIA

O método a ser utilizado foi o transversal de caráter quantitativo, com o uso de dados de natureza secundária, uma vez que a coleta foi obtida no banco de microdados do MEC/Inep. De acordo com Richardson (1989), este método destaca-se pelo emprego da quantificação, tanto nas modalidades de coleta de informações, quanto no método através das técnicas estatísticas, desde as mais simples até as mais complexas.

O número de inscritos no Enem 2016 e que, portanto, constavam na base de dados inicial era de 8,6 milhões de participantes. Uma vez que o interesse desse estudo parte dos resultados obtidos na matemática do estado de Sergipe, a primeira iniciativa foi excluir os sujeitos que não faziam parte da UF-28, que não estiveram presentes na prova de Ciências Exatas e Suas Tecnologias, e os considerados treineiros (estudantes que ainda não concluíram o Ensino Médio), totalizando 67.821 participantes.

A segunda iniciativa se deu em definir as variáveis investigadas, salientando a escolha com base em literaturas sobre o tema, tais como: Desempenho na matemática, idade, gênero, etnia, renda familiar, escolaridade dos pais, tipo de escola, turno, abandono ou reprovação, ensino regular, possuir computador, possuir internet, trabalha, estado civil.

Posteriormente, os dados foram organizados no Excel 2016 e analisados por meio do software R (Project for Statistical Computing) versão 3.6.1.

As técnicas utilizadas foram a de Estatística descritiva para organizar, descrever e sumarizar os aspectos importantes do conjunto de dados, através do cálculo de média, desvio padrão, frequência e percentual. Seguidamente do teste de hipótese de Shapiro-Wilk para testar se a variável dependente provinha de uma população normal. Verificado essa não normalidade aplicou-se os testes de hipótese U de Mann-Whitney e Kruskal-Wallis que serviram para comparar se existia diferença significativa entre os grupos independentes.

A análise de componentes principais e a análise fatorial são técnicas multivariadas que serão aplicadas em terceiro instante, permitindo a análise e interpretação de conjuntos de dados com grande número de variáveis de forma simplificada, ou seja, as técnicas multivariadas serão utilizadas como instrumentos de

redução da quantidade de variáveis possibilitando a análise de um grande banco de dados através de fatores sem a ocorrência de perda significativa de informação.

Por fim, aplicaremos os Modelos Lineares Generalizados que irá nos permitir criar um modelo a partir da família exponencial de distribuição binomial, que nos permita identificar os perfis sociodemográfico que os candidatos tendem quando o desempenho é considerado satisfatório ou insatisfatório.

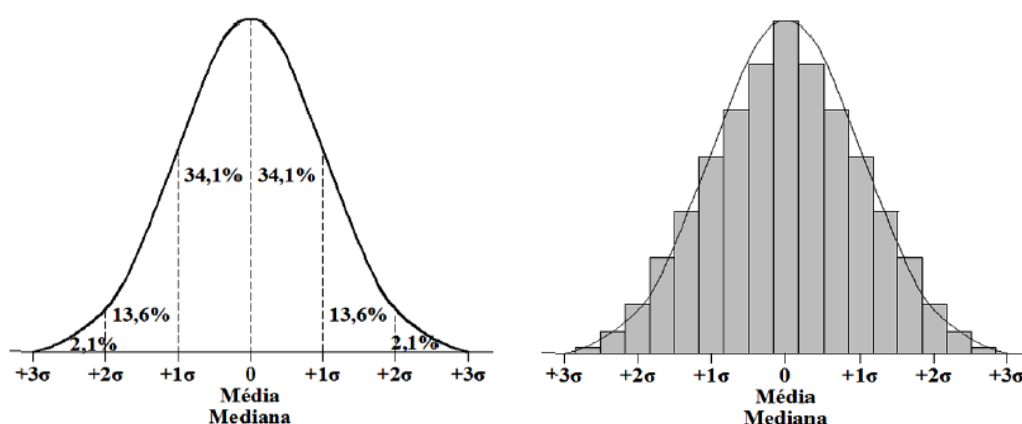
4.1 Teste De Hipótese

4.1.1 Shapiro-Wilk para Normalidade

A normalidade dos dados é uma das suposições frequentemente utilizada para determinar que tipo de teste estatístico será atribuído e muitos dos procedimentos estatísticos são testes paramétricos, os quais requerem que os dados sejam retirados de uma população normalmente distribuída.

A distribuição Normal é uma das mais importantes distribuições de probabilidades da estatística, conhecida também por Distribuição de Gauss ou Gaussiana. A mesma foi desenvolvida pelo matemático francês Abraham de Moivre em 1733 (Triola, 2008), e sua representatividade é feita através de um gráfico simétrico, em forma de sino (Figura 3).

Figura 3: Modelos de distribuição Normal



Fonte: Lopes et al. (2013)

Diversos métodos estatísticos indicam que os dados provem de uma distribuição normal, por exemplo: o histograma, boxplot e o gráfico de dispersão normal Q-Q plot, ainda assim, eles não informam se a distribuição está próxima suficiente da normalidade (FIELD, 2009; BARROS et al., 2012). Desse modo, para dar suporte às análises gráficas, técnicas de inferência estatística são necessários para determinar se uma distribuição é ou não normal.

Dentre as técnicas utilizadas, o teste de Shapiro-Wilk é, aparentemente, o melhor teste de aderência à Normalidade, uma vez que baseia-se nos valores amostrais ordenados elevados ao quadrado. Este fato também é confirmado pelos autores Cirillo e Ferreira (2003) e Öztuna *et al.* (2006).

Shapiro & Wilk (1965) desenvolveram o teste de *Shapiro-Wilk* mostrando que esse teste é eficiente para diferentes distribuições e tamanhos de amostras quando comparado aos resultados de outros testes, a estatística utilizada é:

$$W = \frac{b^2}{S^2} = \frac{(\sum_{i=1}^n a_i y_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (1)$$

Onde y_i é a variável aleatória observada e a_i são os coeficientes tabelados.

Os testes SW fornece o parâmetro valor de prova (p-valor ou significância), que pode ser interpretado como a medida do grau de concordância entre os dados e a hipótese nula H_0 , sendo H_0 correspondente à distribuição Normal. Quanto menor for o valor-p, menor é a consistência entre os dados e a hipótese nula. Então, a regra de decisão adotada para saber se a distribuição é Normal ou não é rejeitar H_0 :

- (i) Se $p\text{-valor} \leq \alpha$, rejeita-se H_0 , ou seja, não se pode admitir que o conjunto de dados em questão tenha distribuição Normal;
- (ii) Se $p\text{-valor} > \alpha$, não há evidências para rejeitar H_0 , ou seja, a distribuição Normal é uma distribuição possível para o conjunto de dados em questão.

4.1.2 U de Mann-Whitney

Desenvolvido primeiramente por Frank Wilcoxon em 1945, para comparar tendências centrais de duas amostras independentes de tamanhos iguais. Em 1947, H.B. Mann e D.R. Whitney generalizaram a técnica para amostras de tamanhos diferentes.

Apropriado para averiguar se dois grupos (X_1, X_2, \dots, X_n) e (Y_1, Y_2, \dots, Y_n) não relacionados (independentes) pertencem ou não à mesma população, o teste U de Mann-Whitney é uma das mais poderosas provas não paramétrica. Suas únicas exigências são de que a variável de mensuração seja de caráter quantitativo e não apresente sabidamente uma variação normal, ou ainda que não exista homogeneidade das variâncias (MATTAR, 1998; COOPER E SCHINDLER 2003).

Para Mattar (1998) a estatística U é a base para a decisão. E para a realização do teste é preciso considerar os seguintes procedimentos:

i) Formule as hipóteses:

$$H_0: \mu_x = \mu_y$$

$$H_1: \mu_x \neq \mu_y$$

ii) Defina o nível de significância (α);

iii) Coloque os dados dos dois grupos em uma única ordenação crescente (amostra combinada) e observam-se as suas posições (ranks) ocupadas pelos elementos X e Y. As observações empatadas atribuir a média dos postos correspondentes;

X	X	Y	X	Y	Y	Y	X
1	2	3	4	5	6	7	8

iv) Considere

n_1 = número de casos do grupo 1 (grupo menor)

n_2 = número de casos do grupo 2

$$n = n_1 + n_2$$

v) Calcule

R_1 = soma dos postos do grupo 1

R_2 = soma dos postos do grupo 2

vi) Calcule a estatística U:

$$\begin{aligned} U_1 &= n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 \\ U_2 &= n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2 \end{aligned} \quad (2)$$

vii) Escolha o menor valor entre U_1 e U_2 para ser o valor comparação (U), utilizado para testar H_0 .

viii) Se $n < 20$ utilizar os valores críticos de Mann-Whitney. Caso contrário, calcule-se a padronização (Z), dada por:

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \quad (3)$$

4.1.3 U de Kruskal-Wallis

Introduzido por Kruskal e Wallis em 1952, o teste é considerado uma extensão do teste de U de Mann-Whitney, permitindo averiguar três ou mais grupos independentes (CAMPOS, 1983; LEVIN e RUBIN, 2004). Esse teste também condiciona que a variável em análise seja medida em escala ordinal ou numérica e deva ser aplicado nos casos em que a amostra for pequena e/ou as suposições exigidas pela análise de variância (normalidade e igualdade das variâncias) forem violadas (FÁVERO, 2009).

Os procedimentos são equivalentes ao de Mann-Whitney, diferindo apenas a hipótese a ser testada de que todos os grupos são iguais contra a hipótese de que pelo menos dois grupos diferem entre si:

$$H_0: g_1 = g_2 = \dots = g_n$$

$$H_1: g_i \neq g_j, \text{ para pelo menos um par } i, j \in [1, \dots, k] (i \neq j)$$

E a estatística de teste definida por:

$$H = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(N+1) \quad (4)$$

Onde:

k = número de amostras

n_j = número de elementos da amostra j

N = total de observações do conjunto das k amostras

R_j^2/n_j = quadrado da soma dos postos/ranks da amostra j dividido pelo respectivo número de elementos

Como o valor de H é de certo modo influenciado por empates, se mais de 25% das observações estiverem empatadas uma correção deve ser feita. O efeito da correção para empates é aumentar o valor de H e assim tornar o resultado mais significativo do que seria (VIALI, 2008). Essa correção consiste em dividir H pelo fator C .

$$C = 1 - \frac{\sum T}{N(N+1)(N-1)} \quad (5)$$

Onde:

$T = t^3 - t$ (onde t é o número de valores empatados)

A estatística H então é comparada com a distribuição Qui-Quadrado com $k-1$ graus de liberdade, rejeitando H_0 se $H \geq \chi^2_{\alpha, k-1}$.

4.2 Análise Multivariada

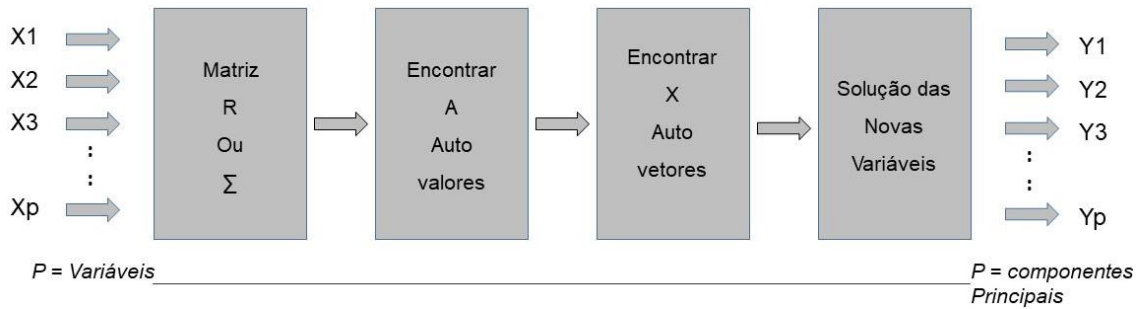
4.2.1 Análise de Componentes Principais (ACP)

Desenvolvida inicialmente com estudos de maximização de variâncias de Hotteling em 1933 (MORRISON, 1976), a ACP tem como principal objetivo retirar a multicolinearidade das p variáveis aleatórias reais, $X_1, X_2, X_3, \dots, X_p$, efetivamente correlacionadas entre si, transformando-a num outro conjunto de k variáveis $Y_1, Y_2, Y_3, \dots, Y_p$ não correlacionadas que expliquem o máximo possível a variância total das variáveis reais.

De acordo com Souza (2000), para a obtenção das componentes principais, é necessário calcular a matriz de variância-covariância (S) aplicada no caso de variáveis que apresentem distribuição normal, ou a matriz de correlação (R), que deve apresentar coeficientes de correlação não nulos. Quando utilizado matriz de covariância necessita-

se a verificação da unidade de medida das variáveis, quando muito diferentes entre si aconselha-se que seja efetuada uma padronização para retirar o efeito de escala. Essa padronização é feita de modo que o valor da média seja igual a zero e a variância igual a um, consequentemente calcula-se os autovalores, os autovetores e descreve as combinações lineares, que serão as novas variáveis.

Figura 4: Etapas da aplicação da análise de componentes principais



Fonte: Souza, A.M. (2000, p.25)

Segundo Hongyu et al. 2016, essas etapas podem ser expressas considerando uma matriz X com p variáveis dimensionais de médias $\mu = \mu_1, \mu_2, \dots, \mu_p$, variâncias $\sigma^2 = \sigma_1^2, \sigma_2^2, \dots, \sigma_p^2$ e n observações não independentes entre si.

$$X = \begin{bmatrix} x_{11} & \dots & x_{1p} \\ x_{21} & \dots & x_{2p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{np} \end{bmatrix}$$

A partir dessa matriz $X_{n \times p}$ é possível fazer uma estimativa da matriz de covariância simétrica (S) de ordem $p \times p$.

$$S = \begin{bmatrix} \hat{Var}(x_1) & \hat{Cov}(x_1, x_2) & \hat{Cov}(x_1, x_3) & \dots & \hat{Cov}(x_1, x_p) \\ \hat{Cov}(x_2, x_1) & \hat{Var}(x_2) & \hat{Cov}(x_2, x_3) & \dots & \hat{Cov}(x_2, x_p) \\ \hat{Cov}(x_3, x_1) & \hat{Cov}(x_3, x_2) & \hat{Var}(x_3) & \dots & \hat{Cov}(x_3, x_p) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{Cov}(x_p, x_1) & \hat{Cov}(x_p, x_2) & \hat{Cov}(x_p, x_3) & \dots & \hat{Var}(x_p) \end{bmatrix}$$

Comumente as variáveis em estudos apresentam-se em unidades de medidas diferentes entre si, e neste caso, o mais conveniente é que se padronize as mesmas fazendo a média igual a zero e a variância igual a um (REGAZZI, 2000).

$$Z_{ij} = \frac{X_{ij} - \bar{X}_j}{S(X_j)}, i = 1, 2, \dots, n \text{ e } j = 1, 2, \dots, p \quad (6)$$

Em que \bar{X}_j e $S(X_j)$ são respectivamente a média e o desvio padrão da característica j:

$$\bar{X}_j = \frac{\sum_{i=1}^n X_{ij}}{n} \quad (7)$$

$$S(X_j) = \frac{\sum_{i=1}^n (X_{ij} - \bar{X}_j)^2}{n-1} = \sqrt{\hat{V}ar(X_j)}, j = 1, 2, \dots, p \quad (8)$$

Após a padronização obtém-se a matriz de dados Z das variáveis padronizadas z_j que é equivalente a matriz de correlação R da matriz X.

$$Z = \begin{bmatrix} z_{11} & z_{12} & z_{13} & \dots & z_{1p} \\ z_{21} & z_{22} & z_{23} & \dots & z_{2p} \\ z_{31} & z_{32} & z_{33} & \dots & z_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & z_{n3} & \dots & z_{np} \end{bmatrix} \Rightarrow R = \begin{bmatrix} 1 & r(x_1x_2) & r(x_1x_3) & \dots & r(x_1x_p) \\ r(x_2x_1) & 1 & r(x_2x_3) & \dots & r(x_2x_p) \\ r(x_3x_1) & r(x_3x_2) & 1 & \dots & r(x_3x_p) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r(x_px_1) & r(x_px_2) & r(x_px_3) & \dots & 1 \end{bmatrix}$$

Com base na matriz de covariância ou correlação o próximo passo é calcular os autovalores e os seus respectivos autovetores. Com o auxílio da matriz identidade (I), monta-se a seguinte equação:

$$\det [R - \lambda I] = 0 \text{ ou } |R - \lambda I| = 0 \quad (9)$$

Sejam $\lambda_1, \lambda_2, \dots, \lambda_p$ as raízes da equação característica (autovalor) da matriz R ou S,

então $\lambda_1 > \lambda_2 > \dots > \lambda_p$. Para cada autovalor λ_i existe um autovetor a_i :

$$\begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{ip} \end{bmatrix}$$

tornando-se possível formar pares $(\lambda_1, a_1), (\lambda_2, a_2), \dots, (\lambda_p, a_p)$, definindo assim as combinações lineares que irão gerar as componentes principais (JOHNSON e WICHERN, 1998) :

$$\begin{aligned} Y_1 &= a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \\ Y_2 &= a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \\ &\vdots \\ Y_p &= a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p \end{aligned} \tag{10}$$

Cada autovalor corresponde à fração de variância representada pelo respectivo autovetor (NADAL e PANERAI, 1990). Sendo assim, a soma dos primeiros autovalores representa a proporção de informações retida na redução de p para k dimensões (MANLY, 2008).

O primeiro componente origina-se com base no autovetor de maior autovalor correspondente ao eixo para o qual os pontos têm maior variabilidade. O segundo maior autovetor corresponde ao segundo componente, e assim sucessivamente (MANLY, 2008).

A variabilidade inclusa nos componentes principais equivale à variação das variáveis originais, desta forma a contribuição (C_i) de cada componente (Y_i) para a variação é expressa em porcentagem pela variação total que pertence a cada componente (HONGYU, 2016; JOHNSON; WICHERN, 1998).

$$C_i = \frac{\widehat{var}(Y_i)}{\sum_{i=1}^p \widehat{var}(Y_i)} \cdot 100 = \frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \cdot 100 = \frac{\lambda_i}{traço(s)} \cdot 100 \tag{11}$$

A escolha do número de componentes principais é bastante difícil. É de grande importância que os componentes selecionados forneçam uma interpretação significativa tanto estatística quanto para a área específica a ser estudada (JOLLIFFE, 1972).

Não existe um modelo estatístico que ajude nessa tomada de decisão. Segundo Peres-Neto et al., (2005), a determinação dos componentes a serem mantidos pode ser feita obedecendo a diferentes critérios.

Em diversos casos, retêm-se modelos que expliquem pelos menos 80% da variação total (JOHNSON; WICHERN, 1998). Um outro método é sugerido por Jolliffe (1972), no qual os componentes com autovalor inferior a 0,7 são passíveis de descarte. E por fim o método de Kaiser que será utilizado neste trabalho (KAISER, 1960), que seleciona os componentes cuja variação (autovalor) é maior ou igual a 1 (um).

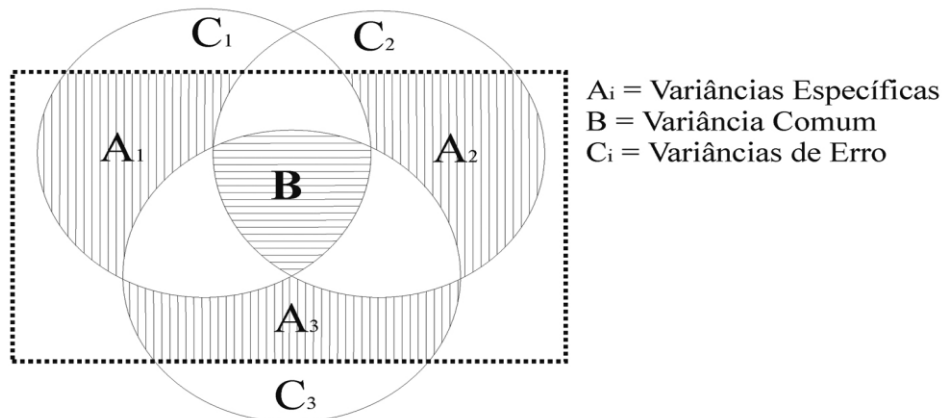
E ainda o teste de Cattell ou também conhecido como diagrama de scree plot, em que o número de componentes é determinado pela mudança brusca na curva do gráfico, determinando onde cessam os maiores e iniciam os menores valores referentes a variância (CATTELL, 1966), ou seja, identifica um ponto de inflexão na curva que liga os autovalores em um gráfico de escala normal ou logística, considerando o primeiro ponto após a sua primeira inflexão, como o número adequado de CP a serem retidos na análise (JOLIFFE, 2002; JACKSON, 2003).

4.2.2 Análise Fatorial (AF)

Assim como o método de Componentes Principais a Análise Fatorial também busca a redução de dados assumindo que a variância de uma variável é composta por três aspectos: a variância específica, que diz respeito à variação que não é compartilhada com as demais variáveis; a variância comum, referente a variância compartilhada entre todas as variáveis e a variância do erro, que trata da variação não explicada pelo componente ou fator (DAMÁSIO, 2012).

Uma diferença fundamental entre os dois métodos é que a ACP trabalha com a variância total observada (na Figura 5, $A_1 + B$), enquanto a AF trabalha somente com a variância partilhada dos itens (na Figura 5, apenas B) (DAMÁSIO, 2012).

Figura 5: Ilustração das variâncias de três itens e suas relações com um fator hipotético.



Fonte: Damásio, B.F. (2012)

Quando comparado o resultado da ACP com a AF nota-se que as cargas fatoriais e comunalidades da ACP são mais elevadas, e as taxas de variância explicadas elevadas em relação a AF, pois a ACP considera a variância específica dos itens e não somente a variância comum, isso gera inexatidão na sua interpretação, uma vez que o objetivo é entender as variáveis latentes que influenciam no grupo de variáveis (COSTELLO & OSBORNE, 2009).

A aplicação da análise fatorial requer que as variáveis $X_1, X_2, X_3, \dots, X_p$ sejam correlacionadas entre si, testando assim as hipótese de que a matriz de correlação é identidade $H_0: R = I$, contra a sua alternativa $H_1: R \neq I$ (FIELD, 2009).

Para testar essas hipóteses temos o teste de Bartlett sob as hipóteses H_0 , a distribuição da estatística deste teste converge para uma variável aleatória com distribuição qui-quadrado com $\frac{p(p-1)}{2}$ graus de liberdade, sob a hipótese alternativa (H_1) da normalidade do vetor aleatório $X = [X_1, X_2, X_3, \dots, X_p]$. A estatística é dada por:

$$\chi^2 = - \left(n - 2 - \frac{2p+5}{6} \right) \ln |R| \quad (12)$$

Valores do teste de esfericidade de Bartlett com níveis de significância (α) $p < 0,05$ rejeita a hipótese de que a matriz de dados é similar a uma matriz-identidade, indicando que a matriz é fatorável (TABACHNICK & FIDELL, 2007).

E ainda o critério de Kaiser-Meyer-Olkin (KMO), conhecido por teste de adequação da amostra, sugerindo a proporção da variância dos dados que pode estar

sendo explicada por uma variável latente (LORENZO-SEVA, TIMMERMAN & KIERS, 2011). O KMO indica, portanto, o quão adequado é a aplicação a análise fatorial (AF), quanto mais próximo de 1 (um) melhor o resultado, ou seja, mais adequada é a amostra a aplicação da análise fatorial (HAIR et al., 2005). O procedimento para a aplicação da adequação dos dados é dado por:

$$KMO = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} v_{ij}^2} \quad (13)$$

onde r_{ij} e v_{ij} são, respectivamente, os elementos na posição (i, j) da matriz de correlação amostral R, e da matriz $V = UR^{-1}U$, na qual $U = [(diag R^{-1})^{\frac{1}{2}}]^{-1}$.

Como regra de interpretação da estatística, os intervalos de análise dos valores de KMO são apresentados no Quadro 3 abaixo.

Quadro 3: Adequação Amostral de acordo com a medida KMO.

KMO	Adequação
> 0,9	Excelente
(0,8; 0,9]	Boa
(0,7; 0,8]	Média
(0,6; 0,7]	Aceitável
(0,5; 0,6]	Fraca
≤ 0,5	Inaceitável

Fonte: Kaiser, 1958.

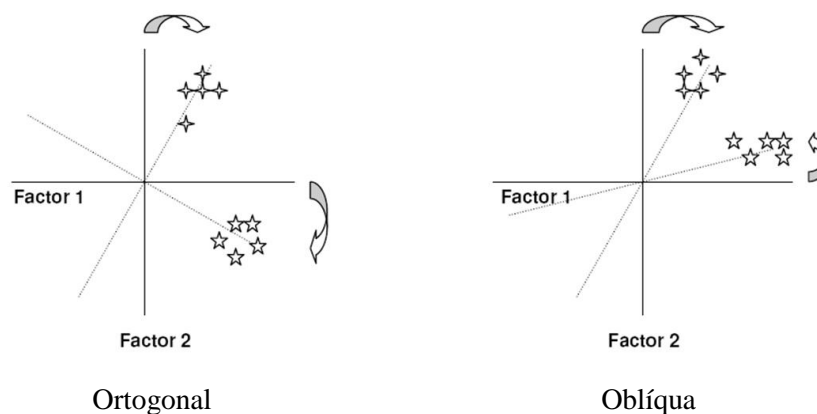
Após a identificação da adequação do método, o próximo passo e tão importante quanto é a retenção de fatores (ARTES, 1998; GLORFELD, 1995). Uma extração ineficaz dificulta a interpretação dos resultados de maneira apropriada (Hayton, Allen & Scarpello, 2004). A ineficácia na retenção dos fatores pode acarretar em dois tipos de erros: 1) superestimação de fatores, resultando em inadequação do poder explicativo; e 2) subestimação de fatores, resultando em perda significativa de explicação (PATIL et al., 2008; FRANKLIN, GIBSON, ROBERTSON, POHLMANN & FRALISH, 1995).

A retenção de fatores neste trabalho será realizada através da técnica de análise de componentes principais (ACP) definida com base nos métodos de Kaiser (eigenvalue > 1) e o Scree plot.

Com o objetivo de auxiliar na interpretação dos fatores obtidos pela ACP o método de rotação de fatores será utilizado para uma melhor interpretação sem modificar os resultados (COSTELLO & OSBORNE, 2009).

Existem dois tipos de rotações, a oblíqua e a ortogonal. A rotação oblíqua não mantém a ortogonalidade dos fatores, permitindo que os fatores sejam correlacionados entre si. E a rotação ortogonal que mantém a ortogonalidade dos fatores, assumindo que os fatores extraídos são independentes uns dos outros.

Figura 6: Exemplo do efeito da rotação (ortogonal, oblíqua) de dois fatores.



Fonte: Carvalho (2013).

Dentre a rotação ortogonal existe vários métodos, nomeadamente quartimax, equimax e varimax. Destes métodos o varimax é o sucedido e o mais utilizado, que busca minimizar o número de variáveis que tem altas cargas em um fator. Reis (2001) afirma que “é um método ortogonal e pretende que, para cada componente principal, existam apenas alguns pesos significativos e todos os outros sejam próximo de zero, isto é, o objetivo é maximizar a variação entre os pesos de cada componente principal, daí o nome Varimax”.

A última etapa da técnica refere-se, justamente, à interpretação e nomeação dos fatores por meio das cargas fatoriais. Considera-se que as cargas fatoriais maiores que 0,3 atingem o nível mínimo, cargas de 0,4 são consideradas mais importantes e quando

maiores que 0,5 são consideradas estatisticamente significativa (HAIR, ANDERSON, TATHAM e BLACK, 2005).

4.3 Modelos Lineares Generalizados

A seleção de modelos é uma parte importante de toda pesquisa em modelagem estatística e envolve a procura de um modelo parcimonioso que descreva bem o processo gerador dos valores observados que surgem em diversas áreas do conhecimento. Nelder e Wedderburn (1972) mostraram que um conjunto de técnicas estatísticas, comumente estudadas separadamente, podem ser formuladas, de uma maneira unificada, como uma classe de modelos de regressão. A essa teoria unificadora de modelagem estatística, uma extensão dos modelos clássicos de regressão, denominaram de modelos lineares generalizados.

A regressão logística é comumente usada como um modelo linear preditivo para diagnóstico e prognóstico de um conjunto de dados estimando a probabilidade de um evento binário (0 ou 1) relacionado com um conjunto de variáveis explicativas (LAI et al., 2009).

Essa técnica prediz a probabilidade de uma variável binária dependente pertencer ao evento de interesse em função de uma ou mais variáveis independentes, no caso dos componentes principais (ACP), sem que as mesmas apresentem uma distribuição Gaussiana e que possuam uma relação linear com as variáveis dependentes (HAIR et al., 2005).

4.3.1 Família Exponencial

Cordeiro e Demétrio (2010) definem a família exponencial como um conjunto de distribuições com características similares cuja função densidade pode ser escrita na seguinte forma

$$f(x; \theta) = h(x)\exp[\eta(\theta)t(x) - b(\theta)] \quad (14)$$

cujas funções $\eta(\theta)$, $b(\theta)$, $t(x)$ e $h(x)$ possuem em subconjuntos dos reais. As funções $\eta(\theta)$, $b(\theta)$ e $t(x)$ não são únicas. Como por exemplo, $\eta(\theta)$ pode ser multiplicada por uma constante k e $t(x)$ pode ser dividida pela mesma constante.

4.3.2 Modelo Binomial

Segundo Cordeiro e Demétrio (2010), a distribuição binomial foi deduzida por James Bernoulli em 1713, embora tenha sido encontrada anteriormente em trabalhos de Pascal.

Suponha que $Y = mP$ tenha distribuição binomial $B = (m, \pi)$, com função de probabilidade, sendo que P representa a proporção de sucessos em m ensaios independentes com probabilidade de sucesso π . A função geratriz de momento de Y é expressa por $M(t; \pi; m) = \{\pi[\exp(t) - 1] + 1\}^m$ e os seus momentos centrais, μ_{2r} e μ_{2r+1} , são de ordem $O(m^r)$, para $r = 1, 2, \dots$. O r -ésimo momento central de P é, simplesmente, $m^{-r}\mu_r$. Todos os cumulantes de Y são de ordem $O(m)$ e, portanto

$$\frac{Y - m\pi}{[m\pi(1 - \pi)]^{1/2}} \sim N(0,1) + O_p(m^{-1/2}) \quad (15)$$

sendo a taxa de convergência expressa pelo terceiro cumulante padronizado. A moda de Y pertence ao intervalo $\{(m + 1)\pi - 1, (m + 1)\pi\}$, e seus coeficientes de assimetria e curtose são respectivamente,

$$\frac{(1 - 2\pi)}{[m\pi(1 - \pi)]^{1/2}} \text{ e } 3 - \frac{6}{m} + \frac{1}{m\pi(1 - \pi)} \quad (16)$$

Se $y = mp$ é inteiro, um número de aproximações para as probabilidades binomiais são baseadas na equação

$$\begin{aligned} P(Y \geq y) &= \sum_{i=y}^m \binom{m}{i} \pi^i (1 - \pi)^{m-i} \\ &= B(y, m - y + 1)^{-1} \int_0^\pi t^{y-1} (1 - t)^{m-y} dt = I_\pi(y, m - y + 1), \end{aligned} \quad (17)$$

em que $I_\pi(y, m - y + 1)$ representa a função razão beta incompleta.

O modelo binomial é usado, principalmente, no estudo de dados na forma de proporções, como nos casos probito, logística (ou “logit”) e complemento log-log e na análise de dados binários, como regressão logística (CORDEIRO; DEMÉTRIO, 2010).

4.3.3 Estimativa de β

Para a obtenção da estimativa de máxima verossimilhança de β utilizamos o processo iterativo de Newton-Raphson, que pode ser reescrito como um processo iterativo de mínimos quadrados ponderados por

$$\beta^{m+1} = (X^T W^m X)^{-1} X^T W^m z^m, \quad (18)$$

$m=0,1,\dots$, onde $z = \eta + W^{-\frac{1}{2}} V^{-\frac{1}{2}} (y - \mu)$. Observe que a quantidade z faz o papel de uma variável dependente modificada, enquanto que W é uma matriz de pesos que muda a cada passo do procedimento iterativo. A convergência de equação acima ocorre em geral com um número infinito de passos, independentes dos valores iniciais utilizados. É usual iniciar com $\eta_i^{(0)} = g(y_i)$ para $i = 1, \dots, n$. Nesse caso, $\hat{\beta}$ assume a forma fechada $\hat{\beta} = (X^T X)^{-1} X^T y$.

5. RESULTADOS E DISCUSSÕES

Para a efetivação do estudo observa-se (Tabela 1) que no ano de 2016 o Exame Nacional do Ensino Médio (ENEM) foi predominante candidatos do gênero feminino (59,50%), com faixa etária superior a 18 anos (72,08%), solteiros (90,93%), autodeclarados cotistas (pardo e preto) (81,43%), que não possuíam computador (50,35%), mas que possuíam acesso à internet (54,69%), que nunca reprovou e/ou abandonou os estudos (75,62%), não residentes de Aracaju (64,67%) e com renda familiar de até dois salários mínimos (45,12%). Com relação às escolas, 80,13% dos candidatos veem de escola pública, e 64,02% frequentavam durante o turno diurno.

Ainda com o auxílio da Tabela 1 e utilizando os testes de hipóteses de U de Mann-Whitney e Kruskal-Wallis foi possível verificar a existência de diferença significativa entre os grupos referente à nota média. É possível notar que os candidatos masculinos apresentaram nota média superior as mulheres com $498,94 \pm 106,92$ contra $458,23 \pm 85,63$, respectivamente. Em conformidade com o estudo feito por Menezes Filho (2007) e Luz (2006), que afirmam que os homens tendem a ter o desempenho superior as mulheres em matemática durante toda a sua trajetória escolar.

Quando referido a faixa etária, os candidatos de menoridade (até 18 anos) se destacaram com $(491,53 \pm 108,23)$, estando em consonância com o pensamento de Seabra e Santos (2007) de que quanto mais novo for o candidato melhor será seu desempenho.

Para os candidatos que se autodeclararam solteiros obtiveram um desempenho mais favorável $(475,41 \pm 97,61)$ aos não solteiros $(467,79 \pm 89,32)$. Esse resultado vai contra a ideia de Masasi (2012) e a favor de Andrade e Corrar (2008) de que o estado civil é um influenciador do desempenho.

Assim como destacou Ferrão et. al (2001), o estudo identificou que os candidatos autodeclarados pretos ou pardos obtiveram um desempenho inferior $(470,41 \pm 92,17)$ aos candidatos brancos e amarelos $(494,01 \pm 113,42)$.

Quando analisado a renda observou-se que existe uma relação positiva com o desempenho em matemática, em sintonia com Ferrão (2003) que afirma quanto maior a pontuação do candidato, maior é a renda familiar do mesmo.

Em um século tão moderno e automatizado é esperado que estudantes que são inibidos da presença de um computador e internet tenham maiores desvantagens que aqueles que possuem essas tecnologias. Simetricamente ao estudo feito pela OECD (2005) foi possível notar diferença significativa entre os candidatos que tinham e os que não tinham acesso as tecnologias com $(499,04 \pm 109,02; 494,21 \pm 107,17)$ e $(450,73 \pm 75,98; 451,19 \pm 76,50)$, respectivamente.

Analisando o desempenho do candidato em relação a sua trajetória escolar, assim como Menezes Filho (2007), o trabalho resultou que os candidatos que se encontram com o histórico escolar regular tendem a um desempenho melhor $(480,80 \pm 101,40)$ que aqueles que estão em defasagem da idade-série $(455,86 \pm 78,48)$.

No quesito escola, ao avaliar a dependência administrativa e seguindo os resultados do estudo de Macedo (2004), os alunos oriundos de escola pública têm desempenho inferior $(460,39 \pm 82,60)$ aos que não são $(532,50 \pm 124,83)$.

Outro fator que tange a escola é o turno a qual o candidato estuda, para Carvalho (1998) os candidatos que não estudam no período diurno preenchem as vagas noturnas por questões inferiores de recursos. Em harmonia com essa ideia, o trabalho apresentou diferença entre as notas médias para quem estuda durante o dia com nota superior $(486,58 \pm 104,24)$ a quem não estuda $(453,62 \pm 77,92)$.

No quesito “trabalho” é notório um equilíbrio entre as notas médias dos grupos trabalha ou já trabalhou $(471,33 \pm 90,33)$ com quem nunca trabalhou $(477,72 \pm 102,29)$. Contradizendo a literatura de Luz (2006), o estudo identificou que o desempenho dos candidatos não é estatisticamente influenciado pelo fator trabalhar.

Outro fator influenciável é local onde o candidato reside, aqueles que cuja residência é localizada na capital do estado apresentaram um desempenho superior $(496,24 \pm 111,23)$ aos candidatos dos demais municípios $(462,96 \pm 85,85)$.

De acordo com o Inep (2018), o Brasil apresenta nota média em matemática equivalente a 489,5 pontos. Ciente dessa informação, salientamos a ineficiência do estado de Sergipe com nota média equivalente a $474,72 \pm 96,91$ pontos.

Tabela 1: Perfil dos avaliados no Enem, prova de matemática, Sergipe, 2016.

Variável	N	%	Nota Média	D.P.	P-valor
Gênero					
Feminino	40355	59,50%	458,23	85,63	< 0,001
Masculino	27466	40,50%	498,94	106,92	
Faixa etária					
Menor Idade	18938	27,92%	491,53	108,23	< 0,001
Maior Idade	48883	72,08%	468,20	91,32	
Estado civil					
Solteiro(a)	61671	90,93%	475,41	97,61	< 0,001
Não solteiro(a)	6150	9,07%	467,79	89,32	
Cotista					
Sim	55224	81,43%	470,32	92,17	< 0,001
Não	12597	18,57%	494,01	113,42	
Renda familiar					
Até 2 salários	30603	45,12%	449,51	75,41	< 0,001
3 a 4 salários	23912	35,26%	471,55	87,06	
Acima de 4 salários	13306	19,62%	538,39	125,04	
Computador					
Sim	33674	49,65%	499,04	109,02	< 0,001
Não	34147	50,35%	450,73	75,98	
Internet em casa					
Sim	37092	54,69%	494,21	107,17	< 0,001
Não	30729	45,31%	451,19	76,50	
Abandonou e/ou Reprovou					
Sim	16536	24,38%	455,86	78,48	< 0,001
Não	51285	75,62%	480,80	101,40	
Escola Pública					
Sim	54347	80,13%	460,39	82,60	< 0,001
Não	13474	19,87%	532,50	124,83	
Turno					
Diurno	43417	64,02%	486,58	104,24	< 0,001
Noturno	24404	35,98%	453,62	77,92	
Trabalha					
Sim ou já trabalhou	31869	46,99%	471,33	90,33	0,06
Nunca Trabalhou	35952	53,01%	477,72	102,29	
Local de Residência					
Aracaju	23964	35,33%	496,24	111,23	< 0,001
Outros	43857	64,67%	462,96	85,85	

D.P. = Desvio-Padrão; p significativo (p<0,001) pelo teste de Mann-Whitney e Kruskal-Wallis

Elaboração: Próprio Autor, 2019.

Com o auxílio da tabela 2, cujo o propósito é apresentar a frequência dos resultados obtidos em Sergipe no ano de 2016, observou-se que menos de 1% dos candidatos não conseguem responder corretamente aos itens recebendo assim a pontuação zero, que 45,35% dos candidatos atingiram pontuação entre 400 e 500 pontos, apenas 31,63% ultrapassaram 500 pontos e que nenhum candidato atingiu a nota máxima de 1000 pontos.

Tabela 2: Candidatos distribuídos por pontuação, dos avaliados no Enem, prova de matemática, Sergipe, 2016.

Pontuação	N	Frequência %
Nota 0	49	0,07%
Até 400	15563	22,95%
401-500	30758	45,35%
501-600	14143	20,85%
601-700	5065	7,47%
701-800	1769	2,61%
801-900	433	0,64%
901-999	41	0,06%
1000	0	0,00%
Total	67821	100,00%

Elaboração: Próprio autor

Em consequência das variáveis significativas e não significativas para o exame, e com o propósito de desenvolver um modelo que explique o desempenho dos candidatos na matemática, foi realizado o teste de esfericidade de Bartlett para verificar a adequação das variáveis selecionadas para a técnica multivariada.

Com o resultado $p\text{-valor} < 0,05$, nota-se pela estatística Qui-quadrado do teste de Bartlett que a hipótese nula foi rejeitada, ou seja, a matriz de correlações é estatisticamente diferente da matriz identidade ao nível de 5% de probabilidade.

Devido ao baixo valor obtido para o valor-p torna-se improvável que a matriz de correlações possua a forma da matriz identidade. Desta forma, pode-se pressupor que o uso da análise fatorial se torna pertinente, pois as correlações entre as variáveis foram significativas.

Dando continuidade à análise fatorial, o primeiro passo é realizar a retenção de fatores. O método utilizado para essa redução de massa foi o de análise de componentes

principais (Tabela 3) que nos retorna os autovalores, o percentual de variação referente a cada fator em relação a variabilidade total e o percentual de variância acumulada.

A quantidade de componentes retidos foi realizada tendo em vista o critério de Kaiser (1960) que considera adequado componentes cujo o autovalor seja maior ou igual a um. A Tabela 3 permite identificar que 6 componentes, atendem a esse critério conseguindo explicar 63,48% da variação total dos dados.

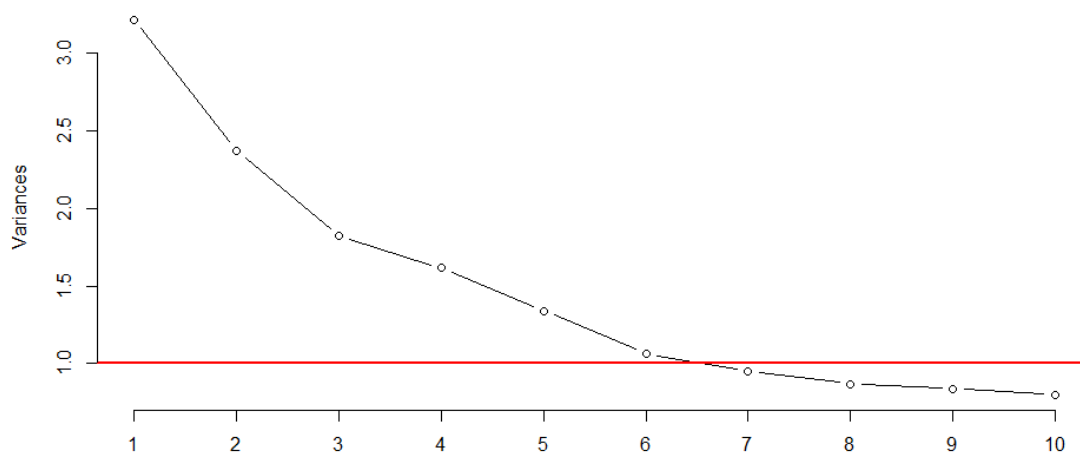
Tabela 3 – Análise de componentes principais.

Componente	Autovalor	Explicação (%)	Explicação Acumulada (%)
CP1	3,21	17,86	17,86
CP2	2,37	13,17	31,03
CP3	1,82	10,12	41,15
CP4	1,62	8,97	50,12
CP5	1,34	7,44	57,57
CP6	1,07	5,92	63,48
CP7	0,95	5,28	68,77
CP8	0,87	4,83	73,60
CP9	0,84	4,66	78,26
CP10	0,80	4,44	82,71
CP11	0,66	3,65	86,36
CP12	0,63	3,49	89,85
CP13	0,52	2,87	92,72
CP14	0,49	2,70	95,42
CP15	0,40	2,20	97,63
CP16	0,36	1,98	99,61
CP17	0,07	0,39	100,00
CP18	0,00	0,00	100,00

Elaboração: Próprio autor

Com o intuito de confirmar a fundamentação para a escolha dos seis componentes, o Teste Scree foi realizado resultando em um gráfico (Figura 7) onde a escolha do número de componentes deve estar acima do ponto de inflexão (Moita-Neto, 2009), ou seja, até o sexto componente é perceptível um decaimento brusco dos autovalores, e a partir do mesmo uma estabilidade, se aproximando cada vez mais de uma linha horizontal.

Figura 7: Screeplot dos autovalores dos componentes principais.



Elaboração: Próprio autor

A redução através do método de componentes principais foi satisfatória, reduzindo dezoito variáveis em apenas seis componentes, permitindo assim a análise dos dados. Não obstante, as cargas fatoriais da ACP tendem a ser bastante elevadas por considerar a variância específica de cada item (COSTELLO & OSBORNE, 2009).

Com o intuito de verificar os fatores latentes que explicam o desempenho em matemática, a técnica de Análise Fatorial foi realizada junto ao critério de rotação ortogonal Varimax que permite uma análise mais compreensível, fazendo com que cada variável apresente carga fatorial elevada em poucos fatores, ou em apenas um (ABDI, 2003).

Na tabela 4 estão apresentadas as comunalidade e variâncias específicas de todas as variáveis. A comunalidade é a proporção de variabilidade de cada variável que é explicada pelos fatores. O valor da comunalidade é o mesmo, independentemente se você usa cargas fatoriais rotacionadas ou não.

Quanto mais perto a comunalidade estiver de 1, mais a variável é explicada pelos fatores. Nesse seguimento, visto que apenas duas variáveis obtiveram comunalidade abaixo de 0,5 nos permite inferir a adequabilidade da análise fatorial perante aos dados analisados, retornado à formação dos seguintes fatores:

Fator 1: Residentes em Aracaju, pai ou mãe com ensino superior, alta renda familiar e que nunca estudaram em escola pública;

Fator 2: Candidatos com maior idade, casados ou em união estável, que já concluíram o ensino médio e que trabalham;

Fator 3: Candidatos brancos ou amarelos;

Fator 4: Candidatos cujos pais possuem ensino fundamental ou médio apenas;

Fator 5: estudantes oriundos do ensino regular, que estudaram em turno diurno e que nunca abandonaram ou reprovaram;

Fator 6: pessoas do sexo feminino e pessoas sem acesso a computador ou internet.

Tabela 4 – Análise de fatores com rotação Varimax, comunalidade e variância específicas referentes às variáveis explicativas.

Variáveis	Cargas Fatoriais						Comunalidade	Variância Específica
	Fator 1	Fator 2	Fator 3	Fator 4	Fator 5	Fator 6		
Residentes em Aracaju	0,67	0,21	0,02	0,14	-0,09	0,05	0,52	0,48
Maior idade	0,01	0,83	0,00	-0,04	-0,18	0,01	0,71	0,28
Sexo Feminino	0,07	0,06	0,00	0,02	0,28	0,46	0,30	0,70
Casados ou em união estável	0,00	0,56	0,01	-0,04	-0,05	-0,02	0,32	0,68
Branços ou amarelos	0,10	-0,02	0,97	-0,02	0,00	-0,04	0,96	0,04
Não Pardos ou negros	-0,10	0,02	-0,97	0,01	0,01	0,04	0,96	0,04
Já concluíram o ensino médio	0,00	0,69	-0,04	0,01	0,35	0,00	0,60	0,41
Pai ou mãe com ensino superior	0,61	-0,16	0,01	-0,57	0,06	-0,15	0,75	0,24
Pai e mãe estudou	-0,31	0,14	0,02	-0,73	-0,04	0,10	0,66	0,34
Pais com ensino fund. ou médio	-0,38	0,06	-0,02	0,90	-0,04	0,08	0,97	0,03
Alta renda familiar	0,66	0,01	0,09	-0,17	0,10	-0,35	0,60	0,40
Sem acesso a computador	0,24	0,10	0,04	0,00	0,13	-0,81	0,74	0,26
Sem acesso à Internet	0,20	0,04	0,04	0,00	0,11	-0,82	0,73	0,27
Trabalha	0,01	0,72	-0,03	0,08	-0,21	-0,06	0,57	0,43
Estudaram em escola pública	-0,68	0,13	-0,13	0,05	-0,04	0,15	0,53	0,47
Ensino Regular	-0,15	-0,17	-0,01	-0,04	0,65	-0,06	0,50	0,52
Turno Diurno	0,25	-0,38	-0,01	0,07	0,53	-0,07	0,50	0,50
Nunca abandonaram e/ou reprovaram	0,07	0,05	0,01	-0,01	0,72	0,04	0,53	0,47
Explicação (%)	0,12	0,13	0,11	0,10	0,09	0,10	-	-
Explicação Acumulada (%)	0,12	0,25	0,36	0,46	0,55	0,65	-	-

Elaboração: Próprio autor.

O Fator 1 que representa 12% da variação total dos dados, apresenta-se por assim dizer, os pontos positivos que um candidato poderia ter para que o desempenho seja satisfatório em relação a matemática, podendo ser considerado o fator benefício.

No fator 2, podendo ser chamado de fator dificuldade. Como o próprio nome já diz é o fator formado por variáveis que representam as principais ideias que impedem os candidatos de terem um desempenho satisfatório, sua representatividade na variação total dos dados é de 13%.

Não tão distante da ideia do fator 1, o fator 3 é um fator constituído por uma variável que segundo literaturas é favorável ao desempenho, sendo então chamado de fator vantagem.

Para o fator 4, constituído pelos candidatos cujos pais possuem apenas o ensino fundamental ou médio, representa 10% da variação total e chamaremos de fator modelo, uma vez que se acredita que o candidato tem seus pais ou responsáveis como algo a ser seguido.

Já o fator 5, representando 9% da variação total dos dados, é caracterizado como fator eficiência, chegando a ser um incentivo para que os futuros candidatos não reprovem e/ou abandonem os estudos. E por fim o fator 6, que pode ser considerado o fator exclusão digital, representando 10% da variação total dos dados.

Definido os seis fatores, o último passo é ajustar um modelo de regressão que estime qual desses perfis de candidatos tendem a um desempenho satisfatório (nota igual ou superior a 450 pontos) ou insatisfatório (inferior a 450 pontos) no exame em relação à matemática.

Tabela 5 - Ajuste da regressão logística para o desempenho insatisfatório em matemática na prova do ENEM, Sergipe, 2016.

Variáveis	Estimativa	Razão de Chance	Erro Padrão	Valor-P
Interseção	-0,074		0,008	< 0,001
Fator 1	-0,400	0,670	0,009	< 0,001
Fator 2	0,050	1,051	0,008	< 0,001
Fator 3	-0,051	0,950	0,008	< 0,001
Fator 4	0,033	1,034	0,008	< 0,001
Fator 5	-0,139	0,870	0,008	< 0,001
Fator 6	0,387	1,473	0,008	< 0,001

Elaboração: Próprio autor.

Neste modelo, é possível notar que os parâmetros apresentaram significância ($p < 0,001$), o que mostra a existência de associação significativa entre estas variáveis e a nota de matemática no Enem 2016.

Quanto as razões de chance, o fator 1 (Residentes em Aracaju, pai ou mãe com ensino superior, alta renda familiar e que nunca estudaram em escola pública) foi apontado pelo modelo de regressão como um fator protetor. Em média, o grupo oposto (residentes no interior, pais que não possuem ensino superior, baixa renda e que realizaram o ensino médio em escola pública) possui $1 / 0,67 = 1,49$ vezes mais chance de ter um desempenho insatisfatório em matemática em comparação com o outro grupo. Esse resultado está de acordo com o observado por Melo e Arakawa (2012) e Barbosa e Sousa (2014), que afirmaram que a escolaridade dos pais e sua renda familiar e os meios da comunidade onde residem também são fatores que influenciam no desempenho escolar.

O fator 2 (Candidatos com maior idade, casados ou em união estável, que já concluíram o ensino médio e que trabalham) caracterizou um grupo de risco em relação ao desempenho em matemática, onde as pessoas caracterizadas pelo fator 2 apresentaram 1,051 mais chance de ter um desempenho insatisfatório em comparação com o grupo oposto (estudantes, jovens e solteiros). Melo e Arakawa (2012) discutiram a influência da idade no desempenho, e levantaram a possibilidade de que o tempo que passou desde o término do ensino médio atue contra os participantes mais velhos.

O fator 3 (Candidatos brancos ou amarelos) apresentou uma razão de chance de 0,95, o que caracteriza um fator de proteção, implicando dizer que os candidatos que comportam esse grupo, tem 1,05 mais chances de ter um desempenho satisfatório em matemática do que os que não fazem parte desse grupo. Comparando com os testes realizados por Melo e Arakawa (2012) e Barbosa e Sousa (2014) sobre etnia, é possível ver que os resultados coincidem, confirmando que estudantes brancos alcançam melhores desempenho no ENEM, possível também através de um preparo feito por Menezes Filho (2007) que estudantes brancos têm um comportamento significativamente melhor que outras etnias na Avaliação da Educação Básica (SAEB).

No fator 4 (Candidatos cujos pais possuem ensino fundamental ou médio apenas) declarou através do modelo 1,034 mais chances de ter um desempenho insatisfatório em matemática, com aqueles que tem fatores antônimos a estes. Os

autores Melo e Arakawa (2012), afirmam que o desempenho educacional dos pais é um fator influenciável no desempenho dos filhos no ENEM. Além destes, os também autores Barbosa e Sousa (2014) encontraram evidências que o grau de escolaridades dos filhos é proporcional ao dos pais. Afirmam ainda que os níveis educacionais dos pais e das mães analisados conjuntamente se mostram com um maior poder explicativo da nota dos filhos no ENEM, do que quando analisados separadamente.

Contiguamente, o fator 5 (Candidatos estudantes oriundos do ensino regular, que estudaram em turno diurno e que nunca abandonaram ou reprovaram), obtiveram razão de chance equivalente à $1 / 0,850 = 1,18$ mais chances de desempenho satisfatório em matemática que aqueles que não tiveram ensino regular, que não estudaram no turno matutino e que já abandonaram ou reprovaram. De acordo com Menezes Filho (2007), o histórico de reprovações dos candidatos é uma variável diferencial que pode influenciar no desempenho escolar.

Por fim, o fator 6 (pessoas do sexo feminino e pessoas sem acesso a computador ou internet), demonstrou através do modelo ser um grupo de risco, onde os candidatos opostos ao fator 6, sendo do sexo masculino com acesso a computador e/ou internet apresentam 1,473 mais chances de ter um desempenho satisfatório em matemática.

6. CONCLUSÕES

Ao longo do trabalho mostrou-se a importância potencial de fatores socioeconômicos, nomeadamente a competência percebida nos resultados escolares em Matemática. Nesses termos, a perspectiva de desenvolvimento escolar deverá contemplar a melhoria da capacidade de auto avaliação dos alunos, principalmente entre os grupos de risco que foram identificados: Fator 2 (Candidatos com maior idade, casados ou em união estável, que já concluíram o ensino médio e que trabalham); Fator 4 (Candidatos cujos pais possuem ensino fundamental ou médio apenas) e Fator 6 (pessoas do sexo feminino e pessoas sem acesso a computador ou internet).

Pelo que foi exposto decorre inevitavelmente a relevância de que as técnicas multivariadas e o modelo obtido são praticáveis, permitindo determinar resultados estatísticos importantes que podem ser aplicados como providências para uma melhor qualidade de ensino. Destaca-se que a pesquisa cumpriu com seu propósito de identificar os fatores que influenciam a ineficiência no âmbito do programa pesquisado, obtendo resultados relevantes que podem ser utilizados gerencialmente pelas instituições privadas e públicas para traçar políticas ou estratégias que reduzam ou controlem esses fatores de riscos.

REFERÊNCIAS

ABDI, H. Factor rotations in factor analyses. Em: M. Lewis-Beck, A. Bryman & T. Futing (Orgs.), Encyclopedia of social sciences research methods (pp. 1-8). Thousand Oaks, CA: Sage, 2003.

ALBERNAZ, A.; FERREIRA, F. H. G.; FRANCO, C. Qualidade e equidade no ensino fundamental brasileiro. Pesquisa e planejamento econômico. Rio de Janeiro: Instituto de Pesquisa Econômica Aplicada, V. 32, n. 3, p.453-476, dez.2002.

ALMEIDA, A.; PIMENTEL, E. P. Mineração de dados no ensino a distância para fins de avaliação do uso das ferramentas de aprendizagem do ambiente tidia-ae. In Anais do III WAVAlia, 2010.

ALVES, F.; LANGE, W.; BONAMINO, A. A geografia objetiva de oportunidades educacionais na cidade do Rio de Janeiro. In: RIBEIRO, L. C de Q. et al. (Org.). Desigualdades urbanas e desigualdades escolares. Rio de Janeiro: Letra Capital, p. 67-90, 2010.

ANGELINI, A. L.; ALVES, I. C. B.; CUSTÓDIO, E. M.; DUARTE, W. F.; DUARTE, J. L. M. Matrizes progressivas coloridas de Raven - escala especial. São Paulo: Centro Editor de Testes e Pesquisa em Psicologia, 1999.

ANASTASI, A.; URBINA, S. Testagem psicológica. Porto Alegre: Artmed, 2000.

ANDRADE, J. X.; CORRAR, L. J. Condicionantes do desempenho dos estudantes de contabilidade: evidências empíricas de natureza acadêmica, demográfica e econômica. Revista de Contabilidade da UFBA, v.1, n.1, p. 62-74, 2008.

ANDRADE, J. M.; LAROS, J.A. Fatores associados ao desempenho escolar: estudo multinível com dados do Saeb/2001. Psicologia: Teoria e Pesquisa, Brasília, v.23, n.1, p.33-42, 2007.

ARTES, R. Aspectos estatísticos da análise fatorial de escalas de avaliação, Revista de Psiquiatria Clínica, 25(5), 223-228, 1998.

BAKER, R., ISOTANI, S.; DE CARVALHO, A. Mineração de dados educacionais: Oportunidades para o brasil. Revista Brasileira de Informática na Educação, 19, 2011.

BARBOSA, M. E. F.; FERNANDES, C. A escola brasileira faz diferença? Uma investigação dos efeitos da escola na proficiência em matemática dos alunos da 4ª série.

In: FRANCO, C. (org.) Avaliação, ciclos e promoção na Educação. Porto Alegre: ARTMED, 2001.

BARBOSA, Wescley de Freitas; SOUSA, Eliane Pinheiro de. Análise do Desempenho Educacional dos Estudantes Cearenses no Exame Nacional do Ensino Médio. In: XV Semana de Iniciação Científica URCA, 2012, Crato, Ceará, p. 1-24, 2014.

BARROS, R. P. et al. Determinantes do desempenho educacional no Brasil. Pesquisa Planejamento Econômico, v. 31, n.1, abr. 2001.

BARROS, R. P.; MENDONÇA, R. O impacto de gestão escolar sobre o desempenho educacional. Washington: BID, p. 39, 1997.

BARROS, M. V. G. et al. Análise de dados em saúde. Londrina: Midiograf, 2012.

BICUDO, M. A. V.; CHAMIE, L. M. S. Compreendendo e interpretando as dificuldades sentidas pelos alunos ao estarem com a Matemática. Revista Zetetiké, Campinas, ano 2, n. 2, p. 61–69, 1994.

BRASIL. Portaria MEC n. 438. Ministério da Educação e do Desporto, 28 MAI 1998. Disponível em <http://www.crmariocovas.sp.gov.br/pdf/diretrizes_p0178-0181_c.pdf>. Acesso em: 14/06/2019.

BRASIL. Base Nacional Comum Curricular – Ensino Médio. Documento homologado pela Portaria nº 1.570, publicada no D.O.U. de 21/12/2017, Seção 1, Pág. 146. Brasília, 21 de dezembro de 2017.

BRASIL. Ministério da Educação. ENEM Documento Básico. Brasília: INEP, 2002.

BRASIL. Ministério da Educação, Secretaria de Educação Básica, 2006. Ciências da natureza, matemática e suas tecnologias. Secretaria de Educação Básica. Orientações curriculares para o ensino médio. Brasília: 2006. (Volume 2).

BRASIL. Ministério da Educação. Políticas de Ensino Médio. Disponível em: <<http://portal.mec.gov.br/component/tags/tag/33672>>. Acesso em: 27/08/2019.

BRASIL. Lei nº 11.096, de 13 de janeiro de 2005. Institui o Programa universidade para todos – PROUNI. Diário Oficial da União, Brasília, 14 jan. 2005a. p.7. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2005/Lei/L11096.htm>. Acesso em: 14/06/2019.

BRASIL. Orientações Educacionais Complementares aos Parâmetros Curriculares Nacionais (PCN+). Ciências da Natureza e Matemática e Suas Tecnologias. Brasília: MEC, 2006.

_____. Lei nº 9.394, de 20 de dezembro de 1996. Estabelece as diretrizes e bases da educação nacional. Diário Oficial da União, Brasília, 23 dez. 1996a. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/19394.htm>. Acesso em: 20/08/2019.

BRITO, M. R. F. Um estudo sobre as atitudes em relação à Matemática em estudantes de 1º e 2º graus. Trabalho de Livre docência. Faculdade de educação - UNICAMP. Campinas: 1996.

CAMPOS, H. Estatística experimental não paramétrica. 2ª ed. Escola Superior de Agricultura “Luiz de Queiroz”, USP. Piracicaba, 1976.

CAMPOS, Luís Augusto; STORNI, Tiago; FERES JÚNIOR, João. A cor do ENEM-2012: comparações entre o desempenho de brancos, pardos e pretos. Textos para discussão do GEMAA (IESP-UERJ), n. 8, p. 1-16, 2014.

CIRILLO, M. A.; FERREIRA, D. F. Extensão do Teste para Normalidade Univariado Baseado no Coeficiente de Correlação Quantil-Quantil para o Caso Multivariado. Rev. Mat. Estat. Revista de Matemática e Estatística, São Paulo, v. 21, n. 3, p. 67–84, 2003.

COOPER, Donald R.; SCHINDLER, Pamela S. Métodos de Pesquisa em Administração. Porto Alegre: Bookman, 2003.

DEL PORTO, FABÍOLA BRIGANTE; FERREIRA, CLÉCIO DA SILVA. Os fatores socioeconômicos associados ao desempenho dos alunos no PISA 2003 (Brasil, México, Espanha e Portugal). In: XIII Congresso Brasileiro de Sociologia, 2007.

CORDEIRO, G. M.; DEMÉTRIO, C. G. B. Modelos Lineares Generalizados e Extensões. Piracicaba: [s.n.].

FAVÉRO, L. P. et al. Análise de dados: Modelagem multivariada para tomada de decisões. Rio de Janeiro: Elsevier, 2009.

FERRÃO, M. E. et al. O Saeb – Sistema Nacional de Avaliação da Educação Básica: objetivos, características e contribuições na investigação da escola eficaz. Revista Brasileira de Estudos de População, v. 18, n. 1/2, p.111-130, 2001.

FERRÃO, M. E. B.; FERNANDES, C. A escola brasileira faz diferença? Uma investigação dos efeitos da escola na proficiência em Matemática dos alunos da 4a série. In: FRANCO, C. (Org.). Promoção, ciclos e avaliação educacional. Curitiba: ArtMed, 2001.

FERRÃO, M. E. Introdução aos modelos de regressão multinível em educação. Campinas: Editora Komedi, 2003.

FERREIRA, E. M. Análise da Abrangência da Matriz de Referência do ENEM com Relação às Habilidades Avaliadas nos Itens de Matemática Aplicadas de 2009 a 2013. 2014. 64F. Dissertação (Pós-Graduação Profissional em Matemática em Rede Nacional). Universidade de Brasília, Brasília, 2014.

FIELD, A. Descobrindo a estatística usando o SPSS. São Paulo: Artmed, 2009.

FOLHA DE SÃO PAULO. Matrícula no 2º grau dobra em 10 anos. 16 de maio de 1998.

FRANKLIN, S. B.; GIBSON, D. J.; ROBERTSON, P. A.; POHLMANN, J. T. & FRALISH, J. S. Parallel Analysis: A method for determining significant principal components. *Journal of Vegetation Science*, 6(1), 99-106, 1995.

GLORFELD, L. W. An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement*, 55(3), 377-393, 1995.

GREMAUD, A. P.; FELÍCIO, F.; BIONDI, R. L. Indicador de efeito escola: uma metodologia para a identificação dos sucessos escolares a partir dos dados da prova Brasil. Brasília: INEP/MEC, 2007 (Texto para Discussão, n. 27). Disponível em: <<http://www.publicacoes.inep.gov.br>>. Acesso em: 14/06/2019.

HAYTON, J. C.; ALLEN, D. G.; SCARPELLO, V. Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, 7(2), pp.191-207, 2004. Disponível em: <http://journals.sagepub.com/doi/pdf/10.1177/1094428104263675>.

JACKSON, D. A. "Stopping Rules in Principal Components Analysis: A Comparison of Heuristical and Statistical Approaches", *The Ecological Society of America*, v.74, pp. 2204-2214, 2003.

JOLLIFFE, I. T. *Principal Components Analysis*, 2 ed, New York, Springer, 2002.

IBOPE. Desempenho em avaliações de alto nível, 2011 Disponível em: <<https://www.kantaribopemedia.com/kantar-ibo-pe-media-e-comscore-lancam-target-group-index-clickstream-no-brasil/>>. Acesso em: 16/08/2019.

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Matriz de referência para o Enem 2009. Brasília: Ministério da Educação, 2009.

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Ministério da Educação, 2009. Disponível em:< http://inep.gov.br/artigo/-/asset_publisher/B4AQV9zFY7Bv/content/definida-toda-a-logistica-do-enem-2009/21206>. Acesso em: 20/08/2019.

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Ministério da Educação, 2015. Disponível em:< <http://portal.inep.gov.br/web/guest/enem/historico>>. Acesso em: 16/07/2019.

INEP - Instituto Nacional De Estudos e Pesquisas Educacionais Anísio Teixeira. ENEM: resultado individual, 2016. Disponível em: <http://download.inep.gov.br/educacao_basica/enem/downloads/2016/apresentacao_final_resultados_2016.pdf>. Acesso em: 20/08/2019.

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Ministério da Educação, 2017. Disponível em:< http://inep.gov.br/artigo2/-/asset_publisher/GngVoM7TApe5/content/mec-e-inep-anunciam-mudancas-no-exame-em-funcao-de-consulta-publica/21206?inheritRedirect=false>. Acesso em: 20/08/2019.

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Ministério da Educação, 2018. Disponível em:<http://portal.inep.gov.br/artigo/-/asset_publisher/B4AQV9zFY7Bv/content/resultados-do-enem-2016-estaoliberados-para-consulta/21206>. Acesso em: 20/08/2019.

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Ministério da Educação, 2019. Disponível em:< <http://portal.inep.gov.br/web/guest/enem/historico>>. Acesso em: 16/07/2019.

KLEINKE, Maurício Urban. O Enem e o impacto de variáveis socioeconômicas e educacionais. In: Inep/MEC — Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. BOLETIM NA MEDIDA. Ano 5. N. 11. Brasília-DF, fev., 2017.

KRUSKAL, W. H. Historical notes on the Wilcoxon unpaired two-sample test, Journal of the American Statistical Association, 52, 356–360, 1957.

LAI, D. T.; LEVINGER, P.; BEGG, R. K., et al. "Automatic recognition of gait patterns exhibiting patellofemoral pain syndrome using support vector machine approach." IEEE Transactions on information Technology in Biomedicine. v. 13, n. 4, pp. 810-817, 2009.

LEVIN, Richard I.; RUBIN, David S. Estadística para administración y economía. Pearson Educación, 2004.

LOPES, A. C.; LÓPEZ, S. B. A performatividade nas políticas de currículo: caso do Enem. Educação em Revista, Belo Horizonte, v. 26, n. 1, p. 89-110, 2010.

LOPES, M. M.; et al. Utilização dos testes estatísticos de kolmogorov-Smirnov e Shapiro-Wilk para verificação da normalidade para materiais de pavimentação, v.21, n.1, p.59-66, 2013.

LORENZO-SEVA, U., TIMMERMAN, M. E. & KIERS, H. A. The hull method for selecting the number of common factors. Multivariate Behavioral Research, 46(2), 340-364, 2011.

LUZ, Luciana. S. Os determinantes do desempenho escolar: a estratificação educacional e o efeito valor adicionado. Encontro Nacional de Estudos Populacionais. Caxambu: Associação Brasileira de Estudos Populacionais. 2006.

MACEDO, G. A. Fatores Associados ao Rendimento Escolar dos alunos da 5ª série (2000) – uma abordagem do valor adicionado. Anais do XIV Encontro Nacional de Estudos Populacionais, ABEP, Caxambú-MG — Brasil, 2004.

MANN, H. B. and WHITNEY, D. R. On a Test of Whether One of Two Random Variables Is Stochastically Larger than the Other. Annals of Mathematical Statistics, 18, 50-60, 1947.

MATOS, Cristiano; VIGGIANO, Esdras. O desempenho de estudantes no ENEM 2010 em diferentes regiões brasileiras. Revista brasileira de estudos pedagógicos (online), Brasília, v.94, n. 237, p.417-438, 2013.

MATTAR, F. N. Pesquisa de marketing: execução, análise. 2. ed. São Paulo: Atlas, 224p, 1998.

MASASI, N. J. How personal attribute affect students' performance in Undergraduate Accounting Course: A Case of Adult Learner in Tanzania. International Journal of Academic Research in Accounting, Finance and Management Sciences, v.2, n.2, 2012.

MELO, Luzia Maria Cavalcante de; ARAKAWA, Victor Haselmann. Existe desigualdade regional na relação entre background familiar e desempenho escolar dos filhos? Evidências para as grandes regiões do Brasil. In: XVIII Encontro Nacional de Estudo Populacionais, 2012, Águas de Lindoia – São Paulo, p. 1-19, novembro 2012.

MENEZES FILHO, N. Os Determinantes do desempenho escolar do Brasil. Instituto Futuro Brasil/IBMEC. São Paulo, 2007.

MIRANDA, E. M.; ALVES, A. R.; MENTEN, M. L. M.; FREITAS, D.; ZUIN, V. G.; PIERSON, A. H. C. ENEM 2009: Articulação entre CTS, interdisciplinaridade e contextualização evidenciadas nas questões das Ciências da Natureza. In: II Seminário Ibero Americano Ciência- Tecnologia- Sociedade no Ensino das Ciências, 2., Distrito Federal. Anais... Brasília: UNB, 2010. p. 1-12, 2010.

MORRISON, D. F. Multivariate statistical methods. 2.ed. New York: McGraw-Hill, 1976.

NACARATO, Adair Mendes; MENGALI, Brenda Leme da Silva; PASSOS, Carmen Lúcia Brancaglioni. A Matemática nos anos iniciais do Ensino Fundamental: tecendo fios do ensinar e do aprender. Belo Horizonte: Autêntica, 2009.

NADAL, J.; PANERAI, R. B. “Estudo preliminar da classificação de arritmias cardíacas utilizando análise de componentes principais”, Revista Brasileira de Engenharia Biomédica, v. 7, n. 1, 1990.

NELDER, J. A.; Wedderburn, R. W. M. Generalized linear models. Journal of the Royal Statistical Society, A, 135, 370-384, 1972.

OECD. Are Students Ready for a Technology-Rich World? What PISA Studies Tell Us, 2005 Disponível em: <www.pisa.oecd.org/document/31/0,2340,en_32252351_32236173_35995743_1_1_1_1,00.html> Acesso em: 01/08/2019.

ÖZTUNA, D.; A. H. Elhan e E. Tüccar. Investigation of Four Different Normality Tests in Terms of Type 1 Error Rate and Power under Different Distributions. Journal of Medicine Cincinnati. v. 36, n. 3, p. 171–176, 2006.

PALERMO, G. A. Fatores associados ao desempenho escolar: uma análise da proficiência em matemática dos alunos do 5º ano do ensino fundamental da rede municipal do Rio de Janeiro. Dissertação (Mestrado em Estudos Populacionais e Pesquisas Sociais). Rio de Janeiro, 2011.

PATIL, V. H., SINGH, S. N., MISHRA, S. & DONAVAN, D. T. Efficient theory development and factor retention criteria: Abandon the ‘eigenvalue greater than one’ criterion. *Journal of Business Research*, 61(2), 162-170, 2008.

RABELO, M. L. Avaliação Educacional: fundamentos, metodologia e aplicações no contexto brasileiro. 1. ed. Rio de Janeiro: SBM, v. 1. 268p, 2013.

JOHNSON, R. A.; WICHERN, D. W. “Applied Multivariate Statistical Analysis”, Prentice Hall, New Jersey, 2002.

REIS, E. Estatística Multivariada Aplicada. Edições Silabo, Lisboa, 2ª edição, 2001.

RICHARDSON, R. J. Pesquisa social: métodos e técnicas. São Paulo: Atlas, 1989.

RODRIGUES, C. G.; RIOS-NETO, E. L. G.; PINTO, C. C. de X. Diferenças intertemporais na média e distribuição do desempenho escolar no Brasil: o papel do nível socioeconômico, 1997 a 2005. *Revista Brasileira de Estudos de População*, v. 28, n. 1, 2011.

SACRAMENTO, I. Dificuldades de aprendizagem em Matemática. I Simpósio Internacional do Ensino da Matemática - Salvador - Ba - setembro de 2008.

SANTIAGO, G. S. Habilidades e Competências de Leitura Segundo o Enem: Entre a teoria e a prática. 2012. 11 F. Dissertação (Pós-Graduação em Linguística Aplicada). Universidade de Taubaté, Taubaté, 2012.

SAPATINI, J. R. Categorização e Análise das Questões de Biologia do ENEM (1998-2012). 2014. 45 F. Monografia (Pós-Graduação em Ensino de Ciências). Universidade Tecnológica Federal do Paraná, Medianeira, 2014.

SEABRA, R. D.; SANTOS, E. T. Avaliando a aptidão espacial de estudantes em um curso de geometria gráfica. Em XVIII Simpósio Nacional de Geometria Descritiva e Desenho Técnico, Anais do GRAPHICA, Curitiba, 2007.

SANTOS, J. M. T. Exame Nacional do Ensino Médio: entre a regulação da qualidade do ensino médio e o vestibular. *Educar em revista*, Paraná, v.1, n.40, p.195-205, 2011.

SANTOS, Bárbara Ferreira. et al. ENEM já é o único vestibular para 48 das 63 federais. Estadão. 09 de novembro de 2014. Disponível em: <<http://brasil.estadao.com.br/noticias/geral,enem-ja-e-o-unico-vestibular-para-48-das-63-federais,1590097>>.

SCORZAFAVE, L.G. & FERREIRA, R. A. Desigualdade de Proficiência no Ensino Fundamental Público Brasileiro: uma Análise de Decomposição. *Revista Economia*, v.12, n. 2, p. 337-359, 2011

SEABRA, R. D. & SANTOS, E. T. Avaliando a aptidão espacial de estudantes em um curso de geometria gráfica. Em XVIII Simpósio Nacional de Geometria Descritiva e Desenho Técnico, Anais do GRAPHICA, Curitiba, 2007.

SHAPIRO, S. S. e WILK, M. B. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika Trust, London*, v. 52, p. 591–609. 3/4, 1965.

SOARES, J. F. O efeito da escola no desempenho cognitivo de seus alunos. In: REICE. *Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, Vol 2, n. 2, 2004.

SOARES, J. F.; ALVES, M.T. “Desigualdades raciais no sistema brasileiro de educação básica”. *Educação e Pesquisa*, vol.29, n.1, p.147-165, 2003.

SOARES, J. F.; COLLARES, A.C.M. Recursos familiares e o desempenho cognitivo dos alunos do ensino básico brasileiro. *DADOS: Revista de Ciências Sociais*, Rio de Janeiro, v.49, n.3, p.615-481,2006.

SOUZA, A. M. Monitoração e ajuste de realimentação em processos produtivos multivariados. 2000.

SOUZA LIMA, M. C. de. Resultados de Ingressantes no Ensino Superior via ENEM: Um Ensaio De Avaliação. Dissertação (Mestrado). Fundação Cesgranrio, Rio de Janeiro, 2011.

TABACHNICK, B.; FIDELL, L. Using multivariate analysis. Needham Heights: Allyn and Bacon, 2007.

TOLEDO, Marília Barros de Almeida; TOLEDO, Mauro de Almeida. *Teoria e Prática de Matemática: Como Dois e Dois*. 1. ed. São Paulo: FDT, 2009.

TRIOLA, M. F. *Introdução à Estatística*. 10ª Ed, Rio de Janeiro: Editora LTC.

VIALI, Lori: *Testes de Hipóteses Não Paramétricos – Departamento de matemática e estatística*. Apostila. (RFRGS) Porto Alegre, 2008. Disponível em: <http://www.mat.ufrgs.br/~viali/estatistica/mat2282/material/apostilas/Testes_Nao_Parametricos.pdf>. Acesso: 18/08/2019.

TRAVITZKI, R., CALERO, J. e BOTO, C. What does the national high school exam (enem) tell brazilian society? *Cepal Review*, 113, 157-174, 2014. Disponível em: <<https://doi.org/10.18356/5db107a1-en>>.

VIGGIANO, E., MATOS, C. O desempenho de estudantes no ENEM 2010 em diferentes regiões brasileiras. *Revista brasileira de estudos pedagógicos* (online), Brasília, v.94, n. 237, p.417-438, 2013.

WILCOXON, F. Individual comparisons by ranking methods, *Biometrics Bulletin*, 1, 80–83, 1945.