

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE

CENTRO DE CIÊNCIAS EXATAS E DA TERRA

DEPARTAMENTO DE ESTATÍSTICA

CURSO DE ESTATÍSTICA

ALANE DANTAS DE AZEVÊDO LIMA

ANÁLISE DE DESEMPENHO NO ENEM: UM ESTUDO COM INGRESSANTES NA
UFRN EM 2016

NATAL/RN

2018

ALANE DANTAS DE AZEVÊDO LIMA

ANÁLISE DE DESEMPENHO NO ENEM: UM ESTUDO COM INGRESSANTES NA
UFRN EM 2016

Monografia de Graduação apresentada ao Departamento de Estatística da Universidade Federal do Rio Grande do Norte como requisito parcial para obtenção do título de bacharel em estatística.

Orientadora: Prof.^a Dr.^a Iloneide Carlos de Oliveira Ramos

Coorientadora: Prof.^a Dr.^a Ivone da Silva Salsa

NATAL/RN

2018

Universidade Federal do Rio Grande do Norte - UFRN
Sistema de Bibliotecas - SISBI
Catalogação de Publicação na Fonte. UFRN - Biblioteca Setorial Prof. Ronaldo Xavier de Arruda - CCET

Lima, Alane Dantas de Azevêdo.

Análise de desempenho no ENEM: um estudo com ingressantes na UFRN em 2016 / Alane Dantas de Azevêdo Lima. - 2018.
81f.: il.

Monografia (Bacharelado em Estatística) - Universidade Federal do Rio Grande do Norte, Centro de Ciências Exatas e da Terra, Departamento de Estatística. Natal, 2018.

Orientadora: Iloneide Carlos de Oliveira Ramos.

Coorientadora: Ivone da Silva Salsa.

1. Estatística - Monografia. 2. Desempenho no ENEM - Monografia. 3. Desempenho acadêmico - Monografia. 4. Análise exploratória de dados - Monografia. 5. Método de Ward - Monografia. 6. Meritocracia - Monografia. I. Ramos, Iloneide Carlos de Oliveira. II. Salsa, Ivone da Silva. III. Título.

RN/UF/CCET

CDU 519.2

ALANE DANTAS DE AZEVÊDO LIMA

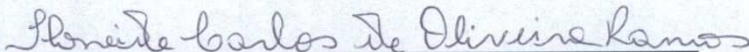
ANÁLISE DE DESEMPENHO NO ENEM: UM ESTUDO COM INGRESSANTES NA
UFRN EM 2016

Monografia de Graduação apresentada ao Departamento de Estatística da Universidade Federal do Rio Grande do Norte como requisito parcial para obtenção do título de bacharel em estatística.

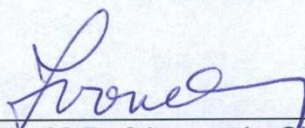
Orientadora: Prof.^a Dr.^a Iloneide Carlos de Oliveira Ramos

Coorientadora: Prof.^a Dr.^a Ivone da Silva Salsa

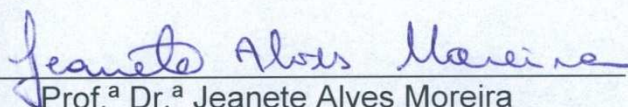
BANCA EXAMINADORA



Prof.^a Dr.^a Iloneide Carlos de Oliveira Ramos
Universidade Federal do Rio Grande do Norte
Orientadora



Prof.^a Dr.^a Ivone da Silva Salsa
Universidade Federal do Rio Grande do Norte
Coorientadora



Prof.^a Dr.^a Jeanete Alves Moreira
Universidade Federal do Rio Grande do Norte
Membro da banca examinadora

Dedico este trabalho à minha avó Maria Matilde de Almeida e seu irmão Pedro Irineu de Lima, *"In Memoriam"*, sem vocês eu não seria a pessoa que sou e talvez não estivesse aqui hoje. Espero que estejam orgulhosos. Estarão comigo a cada suspiro.

AGRADECIMENTOS

A Deus por guiar meus passos, por todas as oportunidades e todas as pessoas colocadas em minha vida.

A minha mãe Nice Dantas de Azevêdo Filha, por lutar por mim desde o dia que soube da gravidez, por todo o incentivo e apoio, principalmente o emocional, na academia e na vida; por todos os sacrifícios que eu sei que fez e pelos que a senhora me omitiu para que eu não me preocupasse. Pelo exemplo de mulher guerreira, detentora de uma força e coragem sem tamanho, saiba que a senhora é a pessoa que eu mais admiro na vida. Ao meu irmão Mateus Dantas de Azevêdo Lima, pelo apoio, por todo o carinho e cuidado comigo, por permanecer ao meu lado. Saibam que qualquer vitória minha é dedicada a vocês.

Ao meu pai Aluísio Cabral de Lima, por me incentivar e mesmo à distância conseguir sanar minhas dúvidas nas componentes comuns à estatística e matemática.

A professora Iloneide Carlos por aceitar a orientação, pela empolgação ao me receber e por todo o conhecimento compartilhado. A professora Ivone Salsa por ter aceitado me coorientar, pelo acolhimento fraterno, por toda a paciência e dedicação. Vocês formam uma dupla incrível, a ambas, o meu sincero e eterno agradecimento por cada minuto dedicado a mim neste trabalho e principalmente por acreditarem e confiarem em mim. À professora Jeanete Alves Moreira, por aceitar participar da banca e dar tantas contribuições para o enriquecimento deste trabalho.

A toda a minha família por todo o carinho e apoio, em particular a tio Nailson Dantas que me ofereceu sua casa para que eu pudesse cursar minha graduação em Natal.

Aos professores do departamento de estatística pelo compromisso em formar profissionais de qualidade, em especial aos professores Damião Nóbrega, Luz Milena, Carla Vivacqua e Dione Maria por terem ido além do quantitativo, por terem acreditado em mim e enxergado capacidade quando eu duvidei.

A todas as pessoas que trabalharam comigo na redação judiciária e nos gabinetes, enquanto eu estagiei no Tribunal de Justiça do Rio Grande do Norte, em especial à Ana Carolina de Souza Fernandes pelo acolhimento, pelos ensinamentos, por me ensinar tanto sobre olhar, cuidar e se preocupar com o outro e por ter se tornado

mais que uma colega de trabalho, mas uma grande amiga a quem eu desejo muito sucesso na vida.

A minha terapeuta Lílian Argolo, por me ajudar a me organizar, a me desfazer da bagagem emocional e a me encontrar em mim, assim como me mostrar que é possível conviver com a ansiedade.

Aos meus amigos que fiz nesta longa caminhada desde o início do curso até aqui, em particular à Rebeca Martins, Luana Romão, Caroline Beatriz e Ythalo Hugo por dividirem comigo suas histórias, pelos momentos de estudo, frustrações, euforia e felicidade. A Pedro Henrique, por me socorrer, a Hiranilson Andrade pelos momentos de discussão nesta reta final, a Diana Lopes pelos momentos de alegrias e descontrações. A Bruna Lôyza que conheceu o meu pior momento e ficou ao meu lado mesmo quando eu tentava evitar, mesmo quando eu não queria companhia, ela sabia que eu precisava.

A Welington Kaven por ter me acolhido e feito com que eu me sentisse em casa, por ter me deixado fazer parte da sua vida, por compartilhar instantes incríveis comigo e estar ao meu lado nos momentos de alegria, tristeza, angústia, felicidade, desespero; por ter me apoiado e respeitado cada decisão minha mesmo muitas vezes não concordando com ela, agradeço por ter se tornado um irmão para mim e por ter me presenteado com uma família que me acolheu como parte dela.

A minha noiva Amanda Bezerra de Sousa, por todo o apoio, por estar ao meu lado quando eu mais precisei, por planejar e sonhar comigo uma vida, por me mostrar o coração mais bonito e sincero que pude conhecer, por todo o amor, carinho, companheirismo, cuidado, sororidade e paciência principalmente nesta etapa final. Orgulho-me muito da pessoa que você é e me inspiro em você. Agradeço também à sua família Bruno, Rubinha, Aninha, Arlindo, Ciro, dona Francisca e seu Arlindo pelo acolhimento.

Por fim, agradeço ao Partido dos Trabalhadores – PT e ao presidente Luís Inácio Lula da Silva, pelo incentivo à educação, pelas políticas públicas que tiraram tantos da miséria e que permitiram qualidade de vida, além de proporcionarem melhores condições para que jovens do interior como eu, pudessem ter a chance de ter um diploma.

RESUMO

O acesso à educação impacta em todas as áreas da vida de uma pessoa. A crescente busca por mão de obra qualificada colocou o Ensino Superior numa vitrine e a globalização estimula a competição que é o princípio da meritocracia. Esta, no entanto, exclui a influência causada por agentes externos. O objetivo desta pesquisa é analisar, tanto o desempenho dos candidatos que fizeram o ENEM em 2015 e ingressaram na UFRN em 2016, quanto o respectivo desempenho acadêmico dos mesmos, ao fim do período letivo de 2016.2; além disto, nesse contexto, pretende-se averiguar a existência de alguma relação entre tais desempenhos e um certo conjunto de variáveis socioeconômicas. Para atender tais objetivos, primeiramente foi aplicada a técnica estatística Análise de Agrupamento, nomeadamente o método de *Ward*, por meio do qual os cursos de graduação ofertados pela UFRN, neste período, foram divididos em grupos de acordo com seus respectivos desempenhos no referido exame. Posteriormente, utilizou-se técnicas da Análise Exploratória, especificamente *Boxplots* e Gráficos em Barras, para relacionar os grupos de cursos com as variáveis socioeconômicas. Os resultados mostraram que o agrupamento possibilitou conhecer melhor certas características relacionadas aos grupos de menores desempenhos, de modo que, apesar de apresentarem medianas semelhantes no desempenho geral, um grupo apresentou bom desempenho em matemática e péssimo desempenho em redação, enquanto outro apresentou uma relação inversa. Além disso, a análise exploratória dos dados permitiu visualizar a influência das variáveis socioeconômicas no desempenho dos candidatos, tornando possível uma compreensão de que a disputa por uma vaga no Ensino Superior é, de certa forma, injusta, logo, não se pode falar em meritocracia e sim, em privilégios, daí, é inegável a necessidade de políticas públicas que permitam uma competição um pouco mais justa.

Palavras-Chave: Desempenho no ENEM. Desempenho Acadêmico. Análise Exploratória de Dados. Método de *Ward*. Meritocracia.

ABSTRACT

Access to education impacts on all areas of a person's life. The growing demand for skilled labor has placed Higher Education in a showcase, and globalization encourages competition, which is the principle of meritocracy. This, however, excludes the influence caused by external agents. The objective of this research is to analyze both the performance of the candidates who have taken the ENEM (National High School Exam) in 2015 and were accepted to UFRN (Federal University of Rio Grande do Norte) in 2016, and their respective academic performance, by the end of the second term of the 2016 school year. Moreover, in this context, we intend to investigate the existence of some relation between these performances and a certain set of socioeconomic variables. In order to meet these objectives, we first applied the Statistical Analysis technique, namely the Ward method, through the undergraduate courses offered by UFRN in this term, they were divided into subgroups according to their respective performance in the aforementioned exam. Subsequently, Exploratory Analysis techniques, specifically Boxplots and Bar Graphs, were used to relate the subgroups of academic courses with the socioeconomic variables. The results showed that the grouping allowed to know better certain characteristics related to the subgroups of lower performances, so that, although they presented similar medians in general performance, one subgroup presented good performance in mathematics and poor performance in writing, while another one presented an inverse relation. In addition, the exploratory analysis of the data allowed us to visualize the influence of the socioeconomic variables on the performance of the candidates, making it possible to understand that the dispute for a place in Higher Education is, in a way, unfair. Therefore, one cannot speak in meritocracy but in privileges, hence the need for public policies that allow a fairer competition is undeniable.

Key words: Performance in the ENEM. Academic achievement. Exploratory Analysis of Data. Ward's method. Meritocracy.

LISTA DE FIGURAS

Figura 1 – Dendrograma dos 110 cursos de graduação da UFRN em 2016, com relação aos desempenhos nas cinco áreas de conhecimento do ENEM.....	34
Figura 2 - Nota média no Exame Nacional do Ensino Médio - ENEM por curso de graduação, na UFRN, em 2016.....	37
Figura 3 - Desempenho no Exame Nacional do Ensino Médio - ENEM geral e por área de conhecimento por grupos de cursos formados pelo agrupamento.....	39
Figura 4 – Desempenho geral e por área de conhecimento no Exame Nacional do Ensino Médio – ENEM por Gênero	40
Figura 5 – Gráfico de dispersão do desempenho mediano em algumas áreas de conhecimento do ENEM considerando os grupos de cursos de Graduação, na UFRN, em 2016	41
Figura 6 – Distribuição percentual do gênero por grupos de cursos de graduação da UFRN, em 2016	42
Figura 7 – Desempenho geral no ENEM e na UFRN por grupo etário.....	45
Figura 8 – Distribuição percentual do grupo etário dos candidatos ingressantes considerando os grupos de cursos de graduação ofertados pela UFRN, em 2016 ..	46
Figura 9 – Desempenho geral no ENEM e na UFRN por sexo	47
Figura 10 – Desempenho geral no ENEM e na UFRN segundo a etnia dos candidatos que ingressaram nos cursos de graduação da UFRN, via SISU, em 2016	48
Figura 11 – Distribuição percentual da etnia dos candidatos ingressantes considerando os grupos de cursos de graduação ofertados pela UFRN, em 2016. .	49
Figura 12 – Desempenho geral no ENEM e na UFRN de acordo com o estado civil dos candidatos que ingressaram nos cursos de graduação da UFRN, via SISU, em 2016.	50
Figura 13 – Distribuição percentual do estado civil dos candidatos ingressantes conforme os grupos de cursos de graduação ofertados pela UFRN, em 2016.....	51
Figura 14 – Desempenho geral no ENEM e na UFRN considerando o meio de transporte mais utilizado pelos candidatos que ingressaram nos cursos de graduação da UFRN, via SISU, em 2016.....	52

Figura 15 – Distribuição percentual do meio de transporte mais utilizado pelos candidatos ingressantes por grupos de cursos de graduação ofertados pela UFRN, em 2016	53
Figura 16 – Desempenho geral no ENEM e na UFRN segundo tipo de escola que cursou o Ensino Médio	54
Figura 17 – Distribuição percentual do tipo de escola na qual o ingressante, via SISU 2016, cursou o Ensino Médio de acordo com os grupos de cursos de graduação ofertados pela UFRN, em 2016.....	55
Figura 18 – Desempenho geral no ENEM e na UFRN de acordo com modalidade de Ensino Médio cursado pelos candidatos que ingressaram nos cursos de graduação da UFRN, via SISU, em 2016.....	56
Figura 19 – Distribuição percentual da modalidade do Ensino Médio cursado pelos candidatos ingressantes segundo os grupos dos cursos de graduação ofertados pela UFRN, em 2016	57
Figura 20 – Desempenho geral no ENEM e na UFRN de acordo com o tipo de cursinho para prestar ENEM utilizado pelos candidatos que ingressaram nos cursos de graduação da UFRN, via SISU, em 2016.....	58
Figura 21 – Distribuição percentual do tipo de cursinho frequentado pelos candidatos ingressantes de acordo com os grupos de cursos de graduação ofertados pela UFRN, em 2016.....	59
Figura 22 – Desempenho geral no ENEM e na UFRN segundo a renda familiar do candidato que ingressou na UFRN em 2016	60
Figura 23 – Percentual da renda familiar segundo os grupos de cursos de graduação ofertados pela UFRN, em 2016.....	61
Figura 24 – Desempenho geral no ENEM de acordo com o grau de instrução dos pais do candidato que ingressou na UFRN em 2016.....	62
Figura 25 – Desempenho geral na UFRN dos candidatos de acordo com o grau de instrução dos pais do candidato que ingressou na UFRN em 2016.....	62
Figura 26 – Distribuição percentual do grau de instrução do pai segundo os grupos de cursos de graduação ofertados pela UFRN, em 2016	63
Figura 27 – Distribuição percentual do grau de instrução da mãe de acordo com os grupos de cursos de graduação ofertados pela UFRN, em 2016.....	64

LISTA DE TABELAS

Tabela 1 – Percentis 10, 25, 50,75 e 90 de cada área de conhecimento do ENEM por curso de graduação da UFRN.....	32
Tabela 2 – Distribuição geral dos valores de MC.....	44
Tabela 3 – Percentis calculados para a variável idade	44
Tabela 4 – Distribuição dos candidatos ingressantes na UFRN em 2016, por etnia	48
Tabela 5 – Distribuição percentual do status acadêmico dos candidatos ingressantes por grupos de cursos de graduação ofertados pela UFRN em 2016	65
Quadro 1 – Distribuição dos cursos de graduação ofertados pela UFRN, no ano de 2016, por grupo de cursos formados pela Análise de Agrupamentos, utilizando o método de Ward.....	35

LISTA DE ABREVIATURAS E SIGLAS

RN – Rio Grande do Norte

UFRN – Universidade Federal do Rio Grande do Norte

COMPERVE – Comissão Permanente de Concursos

SISU – Sistema de Seleções Unificadas

ENEM – Exame Nacional do Ensino Médio

INEP – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

CH – Ciências Humanas e suas Tecnologias

CN – Ciências da Natureza e suas Tecnologias

LC – Linguagens, Códigos e suas Tecnologias

MT – Matemática e suas Tecnologias

TRI – Teoria de Resposta ao Item

TCM ou TCT - Teoria Clássica da Medida ou Teoria Clássica do Teste

SIGAA – Sistema Integrado de Gestão de Atividades Acadêmicas

L – Licenciatura

B – Bacharelado

M – Matutino

V – Vespertino

N – Noturno

MV – Matutino e Vespertino

MN – Matutino e Noturno

VN – Vespertino e Noturno

MVN – Matutino, Vespertino e Noturno

NAT – Natal

MAC – Macaíba

SCR – Santa Cruz

CNO – Currais Novos

CAI – Caicó

MC – Média de Conclusão

MCN – Média de Conclusão Normalizada

SUMÁRIO

1	INTRODUÇÃO	17
2	OBJETIVOS	20
2.1	OBJETIVO GERAL.....	20
2.2	OBJETIVOS ESPECÍFICOS	20
3	METODOLOGIA.....	21
3.1	ENEM	22
3.2	ANÁLISE EXPLORATÓRIA.....	23
3.2.1	Separatrizes	23
3.2.2	<i>Boxplot</i>	25
3.2.3	Gráfico de Barras ou Colunas	27
3.3	ANÁLISE MULTIVARIADA	27
3.3.1	Análise de Agrupamento	27
3.3.2	Medida de Dissimilaridade.....	29
3.3.3	Método de <i>WARD</i>	30
4	ANÁLISE DOS RESULTADOS.....	31
4.1	ANÁLISE DE AGRUPAMENTO	33
4.2	ANÁLISE EXPLORATÓRIA.....	43
4.2.1	Média de Conclusão	43
4.2.2	Idade	44
4.2.3	Gênero	46
4.2.4	Etnia	47
4.2.5	Estado Civil.....	50
4.2.6	Meio de transporte que mais utiliza	51
4.2.7	Tipo de escola que cursou o Ensino Médio	53
4.2.8	Modalidade do Ensino Médio	55
4.2.9	Variável categórica “Frequentou cursinho para prestar ENEM”	57
4.2.10	Renda	59
4.2.11	Grau de Instrução dos Pais	61
4.2.12	<i>Status</i> acadêmico	65

5	CONSIDERAÇÕES FINAIS	66
	REFERÊNCIAS.....	68
	APÊNDICE.....	70

1 INTRODUÇÃO

A educação possui impacto em todas as áreas da vida dos cidadãos, e o acesso a ela permite que o indivíduo forme sua opinião, aprenda a proteger os interesses sociais e exerça seus direitos e deveres (SANDES, 2012). Em “Educação como Prática de Liberdade”, Paulo Freire diz que a educação é algo fundamental à vida do cidadão e que, desnudar-se da alienação é uma força de libertação.

Sendo a necessidade do desenvolvimento uma das prioridades mundiais, a relevância do Ensino Superior é inegável no âmbito das transformações na sociedade, assim como, promover o desenvolvimento econômico e social faz com que o mercado de trabalho exija cada vez mais a qualificação adequada de sua mão de obra. Portanto, a crescente busca pelo Ensino Superior pode ser percebida, principalmente pela necessidade de se adequar a este novo perfil (NEVES, 2007).

A competitividade, que é princípio da meritocracia, estimulada pela globalização, fez da educação sinônimo de produtividade, de modo que, na maioria das vezes, o aluno é cobrado a seguir sua vida escolar, mesmo sem a garantia da qualidade do ensino que ele recebe. A ideologia neoliberal que implanta a ideia de um mérito para seus esforços, exclui os agentes externos que tornam a competição desigual, tais como renda, qualidade de vida, acesso a saúde e educação, dentre outros. Desta forma, essa retribuição, pode ser alcançável apenas a uma parte da população, enquanto que, para os demais, torna-se inatingível (ESFORCE, 2013).

Avaliando esses aspectos, é perceptível que o ingresso no Ensino Superior, por exemplo, pode estar associado, não somente do quanto estudioso possa o aluno ser, mas sim, há um conjunto de fatores que exerce influência sobre o desempenho dos candidatos determinando seu ingresso ou não à universidade.

Segundo a COMPERVE, a Universidade Federal do Rio Grande do Norte – UFRN foi criada em 25 de junho de 1958, e, federalizada em 18 de dezembro de 1960. Essa criação se deu a partir de faculdades e escolas de Nível Superior, já existente, antes de 1958, na capital do Rio Grande do Norte – RN, as quais foram a origem da referida instituição. A UFRN, hoje, conta com cinco *campi* espalhados pelo Estado, sendo a principal instituição de Ensino Superior do RN. Esses *campi* estão localizados em: Natal (Campus Central), Macaíba (Escola Agrícola de

Jundiaí), Santa Cruz (Faculdade de Ciências da Saúde do Trairí), Currais Novos e Caicó (ambos vinculados ao Centro de Ensino Superior do Seridó – CERES).

Antigamente, o ingresso na UFRN se dava por meio de um vestibular realizado pela, então, Comissão Permanente de Vestibular – COMPERVE; posteriormente, a partir do ano de 2013, a UFRN aderiu, de forma parcial, ao Sistema de Seleção Unificada – SISU e passou a mesclar a entrada de novos alunos usando tal sistema e também o vestibular. Em 2014, o processo seletivo da UFRN passou a funcionar principalmente por meio do SISU. Desta forma, o ingresso permitiu uma Universidade mais ampla e diversa, partindo do pressuposto que, mesmo com as políticas sociais, nem todos os candidatos conseguiam obter isenção na inscrição do vestibular, e esta, diferentemente do ENEM, era muito cara. Quando o vestibular deixou de ser realizado, a COMPERVE passou então a ser chamada de Núcleo Permanente de Concursos, pois, o seu trabalho passou a ter outra natureza: realizar concursos públicos, no âmbito da UFRN ou não.

A motivação para esta pesquisa se deu a partir dos estudos desenvolvidos em 1994, pelo então concluinte de Estatística, Flávio Freire, hoje professor do Departamento de Demografia e Ciências Atuariais da UFRN. Esse estudo analisou três métodos distintos para realizar o agrupamento dos cursos de graduação ofertados pela UFRN durante os anos de 1992 a 1994, além de identificar o perfil dos candidatos que ingressaram na UFRN.

Esta pesquisa busca analisar o desempenho geral dos candidatos que fizeram o ENEM em 2015 e ingressaram, por meio do SISU, na UFRN em 2016, e ao mesmo tempo, analisar também o respectivo desempenho acadêmico desses candidatos, durante os períodos letivos de 2016.1 e 2016.2. Para isto, considerou-se a população como sendo todos os candidatos que fizeram o ENEM no ano de 2015 e ingressaram na UFRN no ano de 2016, via SISU, independente do Campus escolhido, e, no que se refere ao seu desempenho acadêmico, foi estabelecido como tempo limite, abril de 2017.

O conjunto de dados que compõe a parte empírica deste estudo foi disponibilizado pelo Núcleo Permanente de Concursos da UFRN – COMPERVE, e, para respaldar as análises estatísticas foi utilizado o *software* estatístico *R-Cran*.

O presente trabalho está estruturado em cinco capítulos; no primeiro deles, está a introdução, esta que está sendo desenvolvida; no segundo, são apresentados os objetivos (geral e os específicos) e as metas a serem alcançadas no decorrer do estudo. O terceiro discorre sobre a metodologia utilizada, trazendo esclarecimentos sobre o conjunto de dados a serem analisados, a população alvo, além das técnicas e análises estatísticas aplicadas a esse conjunto de dados. No quarto capítulo, está a análise dos resultados e as discussões tecidas a partir das constatações que foram reveladas por meio dessas análises. Por fim, no último capítulo, são apresentadas as considerações finais desta pesquisa; nele de maneira resumida, conclusões são expostas, com base nos resultados dos estudos desenvolvidos ao longo deste trabalho.

2 OBJETIVOS

2.1 OBJETIVO GERAL

Analisar o desempenho geral dos candidatos que fizeram o ENEM em 2015, e ingressaram, via SISU, nos cursos de graduação da UFRN em 2016, além do seu respectivo desempenho acadêmico durante os períodos letivos de 2016.1 e 2016.2.

2.2 OBJETIVOS ESPECÍFICOS

- Pesquisar possível relação entre o desempenho geral no ENEM dos alunos que ingressaram na UFRN em 2016 via SISU e certos grupos de cursos de graduação, nessa universidade.
- Averiguar a existência de alguma relação entre o desempenho geral no ENEM, do aluno que ingressou na UFRN em 2016, via SISU, e o seu respectivo desempenho acadêmico ao longo desse ano.
- Analisar a existência de uma possível relação entre o desempenho geral no ENEM e na trajetória acadêmica, dos ingressantes na UFRN em 2016 via SISU, com as seguintes variáveis socioeconômicas: Idade; Gênero; Estado Civil; Grau de instrução do pai e da mãe; Renda; Tipo de escola em que cursou o Ensino Médio; Modalidade do Ensino Médio; Etnia; Meio de transporte que mais utiliza; e, a variável categórica “Frequentou cursinho para prestar ENEM”.

3 METODOLOGIA

O presente estudo consiste em uma análise comparativa entre o desempenho dos ingressantes na UFRN em 2016 via SISU e seu respectivo desempenho acadêmico um ano após o seu ingresso. Em 2014, a UFRN adotou pela primeira vez o Sistema de Seleção Unificada, SISU, como a principal forma de entrada para ingressantes nos cursos de graduação por ela oferecidos. Em 2016, foram ofertadas 7.814 vagas, em 113 cursos de graduação espalhados pelos cinco *campi* distribuídos no Estado do Rio Grande do Norte.

Neste estudo a população considerada é composta por todos os alunos que fizeram o ENEM no ano de 2015 e ingressaram em algum curso de graduação na UFRN, via SISU, em 2016, independente do campus escolhido. Foram 6.664 ingressantes em 110 cursos de graduação, a diferença entre o número de cursos e o número de vagas ofertadas pela UFRN, se dá pelo fato de existirem outros meios de ingresso na instituição, como o ingresso via vagas remanescentes, ou o ingresso em cursos que dependem da conclusão do curso em Ciência e Tecnologia, como é o caso de algumas engenharias.

O banco é composto de tal maneira que cada linha representa um aluno ingressante, e nas colunas, estão às informações sobre seu desempenho, tanto no ENEM, quanto na UFRN, assim como, todas as variáveis referentes ao seu perfil socioeconômico.

Neste trabalho os dados da pesquisa foram submetidos às técnicas estatísticas denominadas Análise Exploratória, especificamente: Percentis, *Boxplots* e Gráfico de Barras, e a Análise Multivariada. Sendo que, dentre o leque de opções para tratamento de dados oferecidos na Análise Multivariada, foi utilizada, especificamente, a Análise de Agrupamentos, técnica que permite agrupar dados semelhantes entre si.

No decorrer da pesquisa fez-se necessário o uso de uma ferramenta gráfica, denominada Gráfico de Dispersão, cuja finalidade é identificar de maneira clara uma possível relação entre duas variáveis.

3.1 ENEM

Criado em 1998, com o objetivo de avaliar a qualidade da educação do País, o ENEM é hodiernamente um exame realizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – INEP que foi reestruturado em 2009, sob o governo do presidente Luiz Inácio Lula da Silva e da gestão de Fernando Haddad, como ministro da educação, de modo que, através do SISU, pudesse ser permitido o acesso unificado ao Ensino Superior, nas instituições Federais de Ensino.

A referida reestruturação implementou um novo modelo de prova para o ENEM, o qual contém 180 questões, divididas em quatro áreas de conhecimento, denominadas: Ciências Humanas e suas Tecnologias (CH), Ciências da Natureza e suas Tecnologias (CN), Linguagens, Códigos e suas Tecnologias (LC) e Matemática e suas Tecnologias (MT), além de uma redação, considerada uma quinta área.

No que concerne ao cálculo do desempenho dos candidatos, a partir da reforma do Enem, em 2009, foi implementada a Teoria de Resposta ao Item – TRI cuja metodologia sugere formas para representar a relação entre a probabilidade de um indivíduo escolher uma das alternativas de uma questão e seus traços latentes (características do indivíduo que não podem ser observadas diretamente). Esse método estatístico surge para suprir as limitações da Teoria Clássica da Medida – TCM ou a Teoria Clássica do Teste – TCT e uma de suas principais características é concentrar o foco nos itens da questão e não na prova, assim ele pode comparar populações distintas, ou mesmo, uma mesma população submetida a diferentes métodos. (ANDRADE, D. F. e TAVARES e VALLE, 2000).

Os dados de desempenho utilizados neste trabalho foram obtidos por meio do portal do INEP e os dados associados às variáveis socioeconômicas foram coletados em duas etapas, sendo a primeira no ato da inscrição no ENEM e a segunda por meio de um questionário disponibilizado pelo Sistema Integrado de Gestão de Atividades Acadêmicas – SIGAA, no ano de ingresso do aluno. Uma vez que, uma parte dos dados dependia exclusivamente das respostas dos ingressantes ao questionário disponibilizado, e que, alguns alunos cancelaram a matrícula antes de iniciar o curso, existem 385 dados faltantes para este conjunto de variáveis que compõe a segunda etapa da coleta de dados.

3.2 ANÁLISE EXPLORATÓRIA

A Estatística Descritiva é a etapa inicial da análise; ela se constitui em um ramo da Estatística que utiliza técnicas para organizar, descrever e resumir um conjunto de dados. Ela explora, dentre outros assuntos, as Medidas de Posição (média, moda e mediana), as Medidas de Dispersão (variância, desvio padrão, amplitude, etc.) e as Medidas associadas à forma da distribuição dos dados: a Assimetria e Curtose.

3.2.1 Separatrizes

Separatrizes são valores que dividem uma amostra (ou população) ordenada, de maneira crescente, em partes iguais quanto ao número de elementos em cada parte. Existem 3 tipos de separatrizes, são eles: os quartis, os decis e os percentis. Neste trabalho, apenas os quartis e os percentis serão utilizados.

Por serem valores calculados, e não necessariamente observações da amostra (ou população), algumas vezes é necessário se recorrer ao uso da interpolação para encontrar o valor exato da separatriz desejada.

3.2.1.1 Quartil

Os quartis são separatrizes que dividem uma amostra em quatro partes iguais. Para tal, dada uma série ordenada, são necessários 3 quartis para dividi-la em quatro partes, sendo que cada uma delas corresponde a um quarto de todos os dados, ou seja, 25%. Deste modo, o primeiro quartil (Q_1) divide os dados de tal maneira que 25% deles são menores ou iguais ao Q_1 , já o terceiro quartil (Q_3) separa os dados de modo que apenas 25% das observações são maiores ou iguais a ele, o segundo quartil (Q_2) é o valor que divide os dados, tal que 50% das observações são menores ou iguais a ele, ou seja, o segundo quartil é também a mediana. Assim, a análise destes valores possibilita uma singela visualização da tendência central e da dispersão de um conjunto de dados. O primeiro passo para a construção dos quartis é calcular o elemento quartílico, este indica a posição do quartil. O elemento quartílico pode ser calculado utilizando a Equação 1.

$$EQ_j = j * \frac{N}{4}$$

Sendo:

j o índice associado ao quartil

N o tamanho da população

Assim, os três quartis podem ser calculados por meio da Equação 2.

$$Q_j = L_i + \left(\frac{EQ_j - F_{antP}}{f_P} \right) * h \quad 2$$

Sendo:

L_i o limite inferior da classe que contém o quartil (valor real indicado pelo elemento quartílico)

F_{antP} a frequência acumulada da classe anterior à classe que contém o quartil

f_P a frequência simples da classe que contém o quartil

h a amplitude das classes

3.2.1.2 Percentil

Os percentis são separatrizes que dividem uma amostra (ou uma população) em cem partes iguais. Para tal, dada uma série ordenada, são necessários 99 percentis e cada uma das cem partes (P_i) corresponde a um por cento de todos os dados. Deste modo o décimo percentil (P_{10}) divide os dados de tal maneira que 10% dos dados são menores ou iguais ao P_{10} , do mesmo modo o P_{90} divide os dados de tal maneira que apenas 10% dos dados são maiores que ele. Assim, os percentis P_{25} , P_{50} e o P_{75} correspondem, respectivamente, ao primeiro, segundo e terceiro quartil. Os percentis podem ser calculados por meio da Equação 3.

$$P_i = i * \frac{N}{100} \quad 3$$

Sendo:

i o índice associado ao percentil

N o número total de observações

Os dados possuem distribuições diferentes e com a finalidade de avaliar melhor a sua variabilidade, assim como, obter uma estrutura mais consistente dos dados, para cada uma das cinco áreas de conhecimentos do ENEM (Ciências da Natureza e suas Tecnologias (CN), Ciências Humanas e suas Tecnologias (CH), Linguagens, Códigos e suas Tecnologias (LC), Matemática e suas Tecnologias (MT) e Redação), decidiu-se representar os valores de desempenhos mínimos por meio do percentil 10 (P_{10}) e os desempenhos máximos por meio do percentil 90 (P_{90}). Desta forma, o desempenho mínimo será representado pelo valor que separa os 10% dos candidatos que obtiveram menor desempenho.

No banco de dados original (6.664 observações para cada variável explorada) a Idade é apresentada como uma variável estatística discreta, assumindo 47 valores distintos entre si, sendo que, a frequência dessa variável está muito concentrada nos valores mais baixos da mesma. Esse particular comportamento da variável idade, no contexto estudado, acarretou dificuldades para analisá-la, por isso, decidiu-se subdividir os valores assumidos por essa variável, em classes, utilizando-se como limites, os percentis 10 (P_{10}); 25 (P_{25}); 50 (P_{50}) ou mediana; 75 (P_{75}) e 90 (P_{90}). Com tal procedimento, as frequências dessas classes ficaram mais diluídas, proporcionando maior visibilidade quanto ao comportamento a ser analisado.

3.2.2 *Boxplot*

O *Boxplot* ou Diagrama de Caixa é um gráfico que fornece importantes informações sobre a distribuição de um conjunto de dados, dentre as quais pode-se ressaltar a assimetria, dispersão e os *outliers* (pontos discrepantes). De maneira geral, os *boxplots* são especialmente úteis para realizar comparações entre dois ou mais grupos de dados, de um mesmo conjunto de dados, ou não; no último caso é importante que a escala utilizada seja a mesma, para que a comparação possa ser feita de maneira correta.

A construção da figura geométrica – um retângulo – que é exposto em um *boxplot* tem como base, os quartis: o primeiro, segundo e terceiro quartil, sendo que o 1º e o 3º quartis são os valores que estabelecem os limites do retângulo. Além disso, também fazem parte desse gráfico os limites inferior e superior dos dados

observados, cujos valores são usados no traçado das duas hastes: a inferior e a superior, respectivamente. Quanto ao segundo quartil, este representa a mediana, a qual é uma medida mais robusta quando existem *outliers* no conjunto de dados; no caso do *boxplot*, a mediana está representada por uma linha paralela à base do retângulo, traçada dentro do mesmo, de acordo com o valor assumido por essa medida estatística. A haste inferior do *boxplot* inicia no valor que determina: o valor mínimo; prosseguindo de forma perpendicular à base do retângulo (a linha associada ao primeiro quartil) até encontrar o ponto central dessa base. De forma semelhante, acontece com o desenho da haste superior: ela começa no meio da linha superior que “fecha” o retângulo (portanto paralela à sua base) a qual está associada ao terceiro quartil e termina no ponto associado ao limite superior. Os limites, inferior e superior, L_{Inf} e L_{Sup} , podem ser calculados por meio das Equações 4 e 5 respectivamente.

$$L_{Inf} = \max\{\min(dados); Q_1 - 1,5(Q_3 - Q_1)\} \quad 4$$

$$L_{Sup} = \min\{\max(dados); Q_3 + 1,5(Q_3 - Q_1)\} \quad 5$$

Pontos acima do limite superior ou abaixo do limite inferior são considerados *outliers*. A distância entre o primeiro e o terceiro quartil é chamada de amplitude interquartílica e representa uma medida de variabilidade. Quanto menor a distância entre esses quartis, ou seja, quanto mais achatado for o retângulo, isto significa que menor será a variabilidade dos dados em torno da mediana. Assim, quando dois ou mais diagramas de caixas são colocados lado a lado, é possível, não apenas comparar qual grupo possui maior ou menor mediana, como também, pode-se perceber a variabilidade de cada grupo de dados e a existência ou não de pontos discrepantes.

Os valores do desempenho geral do ENEM variam de 487,5 a 836,0 e os valores dos desempenhos na UFRN variam de 5 a 10. Como a intenção desta pesquisa é avaliar a diferença entre os grupos e não entre as áreas de conhecimento do exame ou entre os desempenhos no ENEM e na UFRN os limites dos gráficos para cada área pode não ser os mesmos.

3.2.3 Gráfico de Barras ou Colunas

O gráfico de barras é uma ferramenta gráfica estatística muito útil para resumir um grande volume de dados e aprender sobre a natureza de sua distribuição. Essa ferramenta é frequentemente utilizada para comparar categorias de uma variável, seja ela qualitativa ou quantitativa.

O gráfico consiste em barras retangulares, que podem ser visualizadas tanto horizontalmente quanto verticalmente, daí o fato de poder ser chamado gráfico de barras (horizontais) ou gráfico de colunas (verticais). Para cada categoria de uma variável, é desenhada uma barra de tamanho proporcional aos valores que ela representa.

Neste trabalho o gráfico será utilizado para comparar os níveis de cada variável socioeconômica dentro de certos grupos de cursos de graduação ofertados pela UFRN em 2016.

3.3 ANÁLISE MULTIVARIADA

A Análise Multivariada faz parte de uma área dedicada ao estudo de diferentes métodos estatísticos, muito utilizados quando o pesquisador precisa tratar mais de uma variável, as quais são medidas em cada elemento de uma, ou mais amostras.

Essa técnica de análise pode ser particionada em dois grupos, sendo, um deles de caráter exploratório e, o outro, inferencial. Dentre as técnicas exploratórias estão a Análise de Componentes Principais, Análise Fatorial, Análise de Correlações Canônicas, Análise de Agrupamentos, Análise de Discriminante e Análise de Correspondência; no tocante às inferenciais, estão as Estimações dos Parâmetros, os Testes de Hipóteses, Análise de Variância, Análise de Covariância e Regressões Multivariadas (MINGOTI, 2005).

3.3.1 Análise de Agrupamento

Análise de Agrupamento, *Cluster*, ou Conglomerados consiste em uma Técnica Estatística que utiliza uma medida de similaridade (variáveis qualitativas) ou de dissimilaridade (variáveis quantitativas) para formar grupos de tal forma que os

elementos dentro de cada grupo (*cluster*) sejam semelhantes entre si, e, em relação aos elementos de grupos diferentes sejam distintos.

A Análise de *Cluster* pode ser fragmentada em métodos Hierárquicos e não Hierárquicos. O Método de Agrupamento Hierárquico consiste em um procedimento que utiliza o mesmo passo até o final da análise, sendo que, em cada etapa, necessita do resultado anterior para realizar o procedimento seguinte, desta forma, um grupo mantém os mesmos objetos até o final do processo. Isto não ocorre com o Método de Agrupamento Não Hierárquico, pois, nesse caso, os grupos formados em cada etapa podem, ou não, ser desfeitos no passo seguinte, ou seja, os grupos podem, ou não, manter os mesmos elementos até que o critério estabelecido de parada seja atingido. Uma das principais diferenças entre eles é o fato de que o número de grupos formados pelo agrupamento hierárquico é definido apenas no final da análise, enquanto que, para os agrupamentos não hierárquicos, esse número é previamente fixado.

O Método de Agrupamento Hierárquico pode também ser dividido em dois tipos: o aglomerativo e o separativo. Em relação ao do tipo aglomerativo cada objeto do estudo inicia sozinho, e, na sequência, a cada passo dado, dois ou mais objetos vão sendo agrupados de acordo com a medida de similaridade ou dissimilaridade adotada. Este procedimento ocorre de maneira contínua, até que um único *cluster*, contendo todos os objetos, seja formado. Quanto ao tipo separativo, ocorre o contrário, objetos do estudo já começam formando um grande *cluster* e em seguida vão sendo separados, a cada passo dado de acordo com a medida de similaridade ou dissimilaridade adotada, até que cada objeto do estudo finalize o processo sozinho.

De maneira prática, a visualização dos Métodos de Agrupamento Hierárquicos se dá por meio de um gráfico conhecido por Dendrograma ou Diagrama de Árvore. A cada passo dado, no método de agrupamento hierárquico, o dendrograma liga os objetos, unindo-os de acordo com os níveis de similaridade, até que todos os objetos pertençam a um único grupo. Fixando um nível de similaridade, e, traçando-se uma linha transversal no dendrograma, é possível identificar a quantidade de grupos formados na altura de similaridade fixada. Por não ter hierarquia no momento em que os grupos estão sendo gerados, podendo cada

grupo se desfazer ao longo do processo, os métodos não hierárquicos não podem ser visualizados através do dendrograma, este gráfico apenas funciona quando o Método de Agrupamento Hierárquico é utilizado.

Dentre os métodos hierárquicos os mais conhecidos são: O método do vizinho mais próximo (ligação simples ou mínima distância); o do vizinho mais distante (ligação completa ou máxima distância); o método da ligação média (distância média); e o método de *Ward* (mínima variância). Dentre os métodos não hierárquicos os mais conhecidos são dois: o *K-Means* e o *Fuzzy C-Means*.

Este estudo utilizará o método de *Ward* para o tratamento e as análises de seus dados. Por meio desse método os cursos da UFRN serão divididos em grupos semelhantes com relação ao desempenho nas 5 áreas de conhecimento dos alunos que fizeram ENEM no ano de 2015 e ingressaram na UFRN no ano seguinte (2016). A escolha pelo método de *Ward* foi orientada pelos resultados obtidos por Silva (2013), para quem esse método se mostra melhor que os demais métodos hierárquicos, quando observado a similaridade entre o tamanho dos conglomerados.

3.3.2 Medida de Dissimilaridade

Existem várias medidas de dissimilaridade, no caso neste trabalho, foi utilizada aquela associada à distância euclidiana a qual pode ser visualizada como uma aplicação repetida do teorema de Pitágoras. Assim, a distância euclidiana entre dois objetos, Y_l e Y_k , sendo $l \neq k$, pode ser calculada por meio da Equação 6.

$$d(Y_l, Y_k) = \sqrt{\sum_{i=1}^n (Y_{i,l} - Y_{i,k})^2} \quad 6$$

Sendo:

$d(Y_l, Y_k)$ a distância euclidiana entre os pontos Y_l e Y_k .

$Y_{i,l}$ o valor da variável i para o objeto l

$Y_{i,k}$ o valor da variável i para o objeto k

Desta forma, quanto maior for a medida de dissimilaridade representada pela distância euclidiana $d(Y_l, Y_k)$, isto significa que menor será a semelhança entre os objetos do estudo.

3.3.3 Método de *WARD*

O método proposto por *Ward* (1963) é um Método Hierárquico Aglomerativo, também chamado de “Mínima Variância”; nele cada objeto do estudo inicia o primeiro passo sozinho, ou seja, o método começa considerando cada objeto como um único grupo, e, a cada passo dado, calcula-se a soma de quadrados dentro de cada grupo (MINGOTI, 2005).

O Método de *Ward* difere das demais técnicas de agrupamento no sentido de que a similaridade entre dois agrupamentos não é uma única medida de similaridade, mas a soma dos quadrados dentro dos agrupamentos feita sobre todas as variáveis. [...] No procedimento de *Ward*, a seleção de qual par de agrupamentos a combinar é baseada em qual combinação de agregados minimiza a soma interna de quadrados no conjunto completo de agrupamentos separados ou disjuntos. Em cada passo, os dois agrupamentos combinados são aqueles que minimizam o aumento na soma total de quadrados em todas as variáveis em todos os agrupamentos (HAIR JR. et al., 2006).

A Soma de Quadrados dos Desvios (SQD) pode ser obtida por meio da Equação 7.

$$SQD = \sum_{i=1}^N x_i^2 - \frac{1}{N} \left(\sum_{i=1}^N x_i \right)^2 \quad 7$$

Sendo:

N o tamanho da população

x_i o n -ésimo objeto do agrupamento

Entre o método de *Ward* e os demais métodos hierárquicos, as diferenças se dão pelo fato de que o método de agrupamento de *Ward* é apropriado apenas para variáveis quantitativas, tendo em vista que ele utiliza comparação entre médias, e ademais o método de *Ward* tem uma tendência a formar grupos de tamanhos aproximadamente iguais, diante da existência da atuação de um fator de ponderação, quando conglomerados de tamanhos distintos são comparados (MINGOTI, 2005).

4 ANÁLISE DOS RESULTADOS

A análise dos resultados desta pesquisa se inicia a partir do cálculo, por curso, dos percentis 10, 25, 50, 75 e 90 para cada uma das cinco áreas de conhecimento do ENEM (Ciências da Natureza e suas Tecnologias – CN; Ciências Humanas e suas Tecnologias – CH; Linguagens, Códigos e suas Tecnologias – LC; Matemática e suas Tecnologias – MT; e Redação). Assim, cada curso poderia ser representado por um ponto em um espaço de 25 dimensões.

A Tabela 1 é um recorte da tabela original, ela mostra uma parte dos valores desses percentis, por curso, apenas para se ter uma breve ideia de como está estruturada a tabela original. Nessa tabela, cada curso de graduação, exibe as separatrizes: P_{10} ; P_{25} ; P_{50} ; P_{75} e P_{90} . Essas medidas estatísticas, como já foi explicado anteriormente, foram tomadas como pontos de referência para estabelecer a classificação dos desempenhos: baixo, mediano e alto.

Para cada curso existem três siglas que indicam, respectivamente, o tipo de graduação (Licenciatura – L, Bacharelado - B), o turno do curso (Matutino – M, Vespertino – V, Noturno – N, Matutino e Vespertino – MV, Matutino e Noturno – MN, Vespertino e Noturno – VN, Matutino, Vespertino e Noturno - MVN) e a sigla da cidade onde está localizado o campus (Natal – NAT, Macaíba – MAC, Santa Cruz – SCR, Currais Novos – CNO, Caicó – CAI).

Tabela 1– Percentis 10, 25, 50,75 e 90 de cada área de conhecimento do ENEM por curso de graduação da UFRN

Cursos	CN					CH	LC	MT	Redação
	10%	25%	50%	75%	90%				
Administração / B / M / NAT	501,34	531,70	570,20	602,45	639,20
Arquitetura e Urbanismo / B / MVN / NAT	606,77	622,17	645,05	679,82	695,65
Biblioteconomia B / V / NAT	470,49	496,85	527,70	568,20	591,38
Biomedicina / B / MV / NAT	594,88	609,75	637,10	673,00	685,74
Ciências Biológicas B / MV / NAT	537,28	574,52	598,50	631,32	660,35
Ciências e Tecnologia B / MV / NAT	504,32	543,55	581,30	620,00	649,58
Direito B / M / NAT	570,90	602,50	638,10	664,70	694,40
Engenharia Civil B / MVN / NAT	605,50	626,85	647,50	677,70	705,55
Estatística B / M / NAT	493,99	516,15	537,20	570,90	605,29
Matemática L / MV / NAT	508,90	525,55	554,30	588,90	617,20
Pedagogia L / N / NAT	480,81	504,92	529,20	552,02	585,29
Serviço social B / V / NAT	485,06	501,50	529,45	575,60	599,15
...
Zootecnia B / MV / MAC	511,51	526,90	554,20	583,32	594,14

Fonte: COMPERVE

Os resultados obtidos por meio do cálculo dos percentis possibilitaram a criação de uma matriz com 110 linhas e 25 colunas, sendo, cada linha, um curso de graduação oferecido pela UFRN, e, nas colunas, cada uma das cinco áreas de conhecimento e os cinco percentis, 10, 25, 50, 75 e 90, para cada uma dessas áreas de conhecimento. Essa estrutura matricial (110 x 25) permite que cada curso dependa não apenas da mediana do desempenho geral, e sim de um conjunto de 25

percentis, ou seja, cada curso é representado por um ponto em um espaço de 25 dimensões, assim é possível preservar a consistência dos dados. Tendo em vista as características dessa estrutura matricial, a técnica de análise estatística escolhida foi a Análise de Agrupamentos, especificamente, o Método de *Ward*.

4.1 ANÁLISE DE AGRUPAMENTO

A utilização do método de *Ward* na estrutura matricial formada pelos referidos percentis possibilitou dividir os cursos de graduação ofertados pela UFRN em 2016, em quatro grupos. Esse agrupamento resultou em um dendrograma exposto na Figura 1, apresentada a seguir.

Fonte: COMPERVE

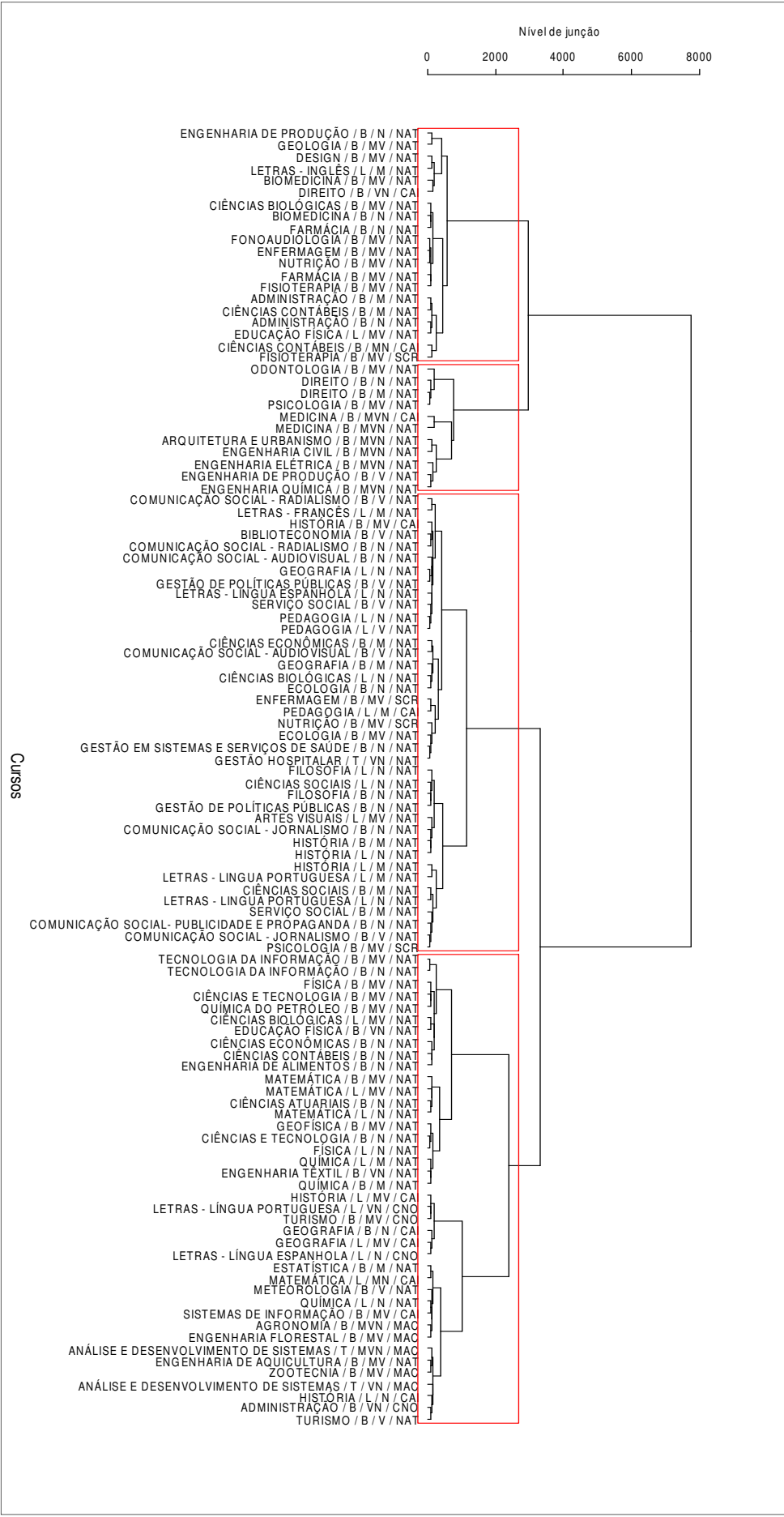


Figura 1 – Dendrograma dos 110 cursos de graduação da UFRN em 2016, com relação aos desempenhos nas cinco áreas de conhecimento do ENEM

O agrupamento visualizado no dendrograma, não necessariamente está em ordem crescente, em relação ao desempenho, assim, o grupo da extrema esquerda pode não ser o grupo com melhor desempenho. Dessa maneira, para melhor enxergar as demais análises, os grupos foram ordenados de acordo com o desempenho no ENEM 2015, dos candidatos ingressantes. Desta forma, os quatro grupos formados pela análise de agrupamentos utilizando o método de *Ward* foram ordenados de tal maneira que o grupo 1 representa os cursos cujos ingressantes alcançaram os maiores desempenhos; o grupo 2 está associado aos cursos cujos ingressantes tiveram um desempenho mediano e os grupos 3 e 4, em ambos, estão os cursos nos quais os ingressantes tiveram os menores desempenhos.

A distribuição dos 110 cursos de graduação ofertados pela UFRN no ano de 2016, nos quatro grupos ordenados pode ser visualizada por meio do Quadro 1.

Quadro 1 – Distribuição dos cursos de graduação ofertados pela UFRN, no ano de 2016, por grupo de cursos formados pela Análise de Agrupamentos, utilizando o método de *Ward*

Grupo	Curso
Grupo 1	ARQUITETURA E URBANISMO / B / MVN / NAT; DIREITO / B / M / NAT; DIREITO / B / N / NAT; ENGENHARIA CIVIL / B / MVN / NAT; ENGENHARIA DE PRODUÇÃO / B / V / NAT; ENGENHARIA ELÉTRICA / B / MVN / NAT; ENGENHARIA QUÍMICA / B / MVN / NAT; MEDICINA / B / MVN / CAI; MEDICINA / B / MVN / NAT; ODONTOLOGIA / B / MV / NAT; PSICOLOGIA / B / MV / NAT.
Grupo 2	ADMINISTRAÇÃO / B / M / NAT; ADMINISTRAÇÃO / B / N / NAT; BIOMEDICINA / B / MV / NAT; BIOMEDICINA / B / N / NAT; CIÊNCIAS BIOLÓGICAS / B / MV / NAT; CIÊNCIAS CONTÁBEIS / B / M / NAT; CIÊNCIAS CONTÁBEIS / B / MN / CAI; DESIGN / B / MV / NAT; DIREITO / B / VN / CAI; EDUCAÇÃO FÍSICA / L / MV / NAT; ENFERMAGEM / B / MV / NAT; ENGENHARIA DE PRODUÇÃO / B / N / NAT; FARMÁCIA / B / MV / NAT; FARMÁCIA / B / N / NAT; FISIOTERAPIA / B / MV / NAT; FISIOTERAPIA / B / MV / SCR; FONOAUDIOLOGIA / B / MV / NAT; GEOLOGIA / B / MV / NAT; LETRAS - INGLÊS / L / M / NAT; NUTRIÇÃO / B / MV / NAT.
Grupo 3	ARTES VISUAIS / L / MV / NAT; BIBLIOTECONOMIA / B / V / NAT; CIÊNCIAS BIOLÓGICAS / L / N / NAT; CIÊNCIAS ECONÔMICAS / B / M / NAT; CIÊNCIAS SOCIAIS / B / M / NAT; CIÊNCIAS SOCIAIS / L / N / NAT; COMUNICAÇÃO SOCIAL- PUBLICIDADE E PROPAGANDA / B / N / NAT; COMUNICAÇÃO SOCIAL - AUDIOVISUAL / B / N / NAT; COMUNICAÇÃO SOCIAL - AUDIOVISUAL / B / V / NAT; COMUNICAÇÃO SOCIAL - JORNALISMO / B / N / NAT; COMUNICAÇÃO SOCIAL - JORNALISMO / B / V / NAT; COMUNICAÇÃO SOCIAL - RADIALISMO / B / N / NAT; COMUNICAÇÃO SOCIAL - RADIALISMO / B / V / NAT; ECOLOGIA / B / MV / NAT; ECOLOGIA / B / N / NAT; ENFERMAGEM / B / MV / SCR; FILOSOFIA / B / N / NAT; FILOSOFIA / L / N / NAT; GEOGRAFIA / B / M / NAT; GEOGRAFIA / L / N / NAT; GESTÃO DE POLÍTICAS PÚBLICAS / B / N / NAT; GESTÃO DE POLÍTICAS PÚBLICAS / B / V / NAT; GESTÃO EM SISTEMAS E SERVIÇOS DE SAÚDE / B / N / NAT; GESTÃO HOSPITALAR / T /

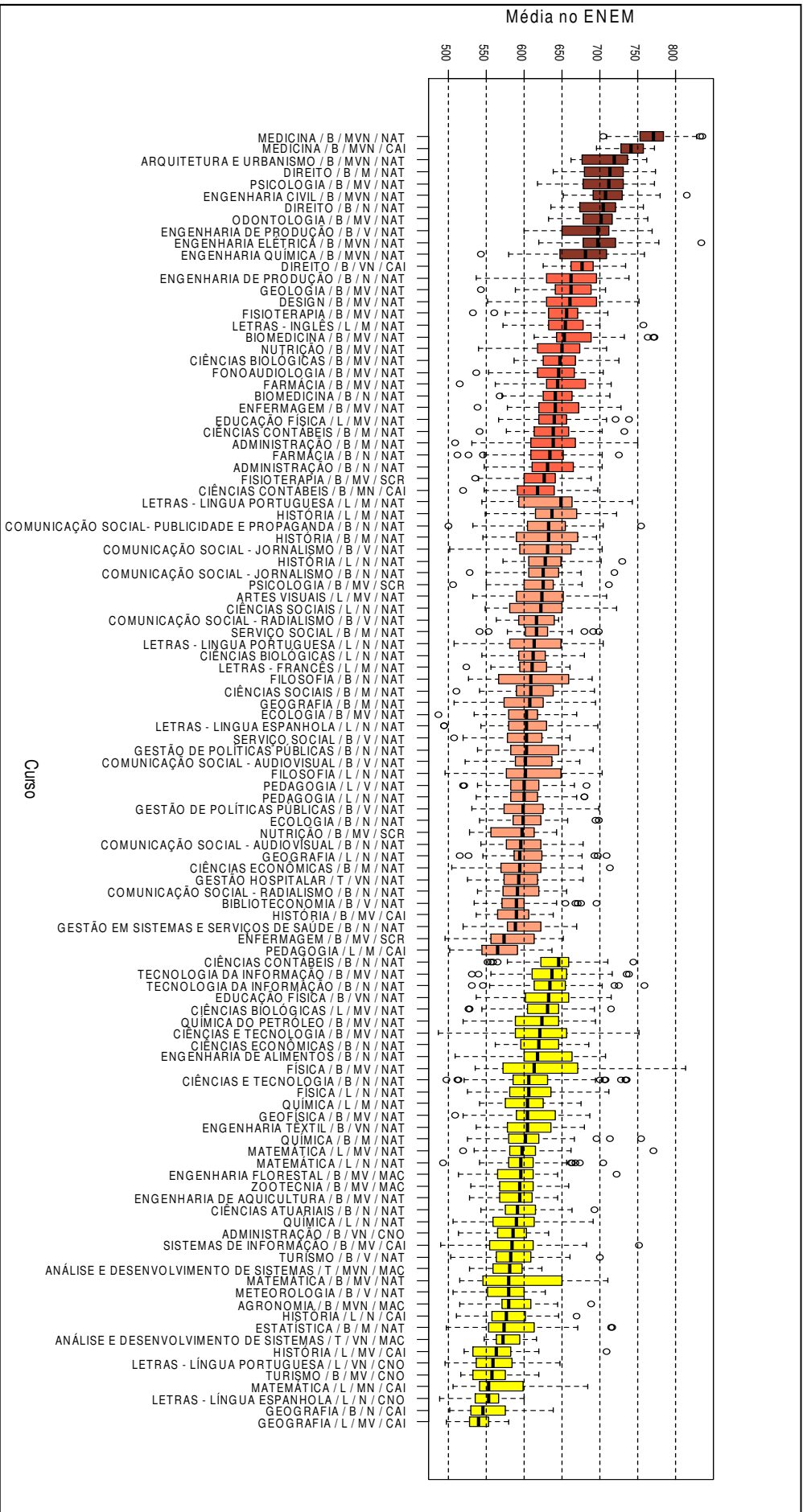
Grupo	Curso
	VN / NAT; HISTÓRIA / B / M / NAT; HISTÓRIA / B / MV / CAI; HISTÓRIA / L / M / NAT; HISTÓRIA / L / N / NAT; LETRAS - FRANCÊS / L / M / NAT; LETRAS - LINGUA ESPANHOLA / L / N / NAT; LETRAS - LINGUA PORTUGUESA / L / M / NAT; LETRAS - LINGUA PORTUGUESA / L / N / NAT; NUTRIÇÃO / B / MV / SCR; PEDAGOGIA / L / M / CAI; PEDAGOGIA / L / N / NAT; PEDAGOGIA / L / V / NAT; PSICOLOGIA / B / MV / SCR; SERVIÇO SOCIAL / B / M / NAT; SERVIÇO SOCIAL / B / V / NAT.
Grupo 4	ADMINISTRAÇÃO / B / VN / CNO; AGRONOMIA / B / MVN / MAC; ANÁLISE E DESENVOLVIMENTO DE SISTEMAS / T / MVN / MAC; ANÁLISE E DESENVOLVIMENTO DE SISTEMAS / T / VN / MAC; CIÊNCIAS ATUARIAIS / B / N / NAT; CIÊNCIAS BIOLÓGICAS / L / MV / NAT; CIÊNCIAS CONTÁBEIS / B / N / NAT; CIÊNCIAS E TECNOLOGIA / B / MV / NAT; CIÊNCIAS E TECNOLOGIA / B / N / NAT; CIÊNCIAS ECONÔMICAS / B / N / NAT; EDUCAÇÃO FÍSICA / B / VN / NAT; ENGENHARIA DE ALIMENTOS / B / N / NAT; ENGENHARIA DE AQUICULTURA / B / MV / NAT; ENGENHARIA FLORESTAL / B / MV / MAC; ENGENHARIA TÊXTIL / B / VN / NAT; ESTATÍSTICA / B / M / NAT; FÍSICA / B / MV / NAT; FÍSICA / L / N / NAT; GEOFÍSICA / B / MV / NAT; GEOGRAFIA / B / N / CAI; GEOGRAFIA / L / MV / CAI; HISTÓRIA / L / MV / CAI; HISTÓRIA / L / N / CAI; LETRAS - LÍNGUA ESPANHOLA / L / N / CNO; LETRAS - LÍNGUA PORTUGUESA / L / VN / CNO; MATEMÁTICA / B / MV / NAT; MATEMÁTICA / L / MN / CAI; MATEMÁTICA / L / MV / NAT; MATEMÁTICA / L / N / NAT; METEOROLOGIA / B / V / NAT; QUÍMICA / B / M / NAT; QUÍMICA / L / M / NAT; QUÍMICA / L / N / NAT; QUÍMICA DO PETRÓLEO / B / MV / NAT; SISTEMAS DE INFORMAÇÃO / B / MV / CAI; TECNOLOGIA DA INFORMAÇÃO / B / MV / NAT; TECNOLOGIA DA INFORMAÇÃO / B / N / NAT; TURISMO / B / MV / CNO; TURISMO / B / V / NAT; ZOOTECNIA / B / MV / MAC.

Fonte: COMPERVE

De maneira sucinta, é interessante destacar que o grupo 3 é composto, em sua maioria, por cursos da área das Ciências Humanas, enquanto que, o grupo 4 é composto majoritariamente por cursos da área tecnológica.

A Figura 2 mostra os *boxplots* dos cursos de graduação ofertados pela UFRN em 2016, ordenados de acordo com as medianas dos desempenhos de seus ingressantes no ENEM 2015, ou seja, de acordo com a ordem dos grupos; as cores variam para indicar o grupo ao qual aquele curso pertence.

Figura 2 - Nota média no Exame Nacional do Ensino Médio - ENEM por curso de graduação, na UFRN, em 2016



Fonte: COMPERVE

Observando as medianas dos *boxplots* dos cursos de graduação ofertados pela UFRN é importante perceber que à medida que o nível do grupo (1, 2, 3 ou 4) aumenta, as medianas decaem, exceto no grupo 4, que a semelhança com o grupo 3 é tão grande que dispendo um grupo sobre o outro não seria possível diferenciá-los.

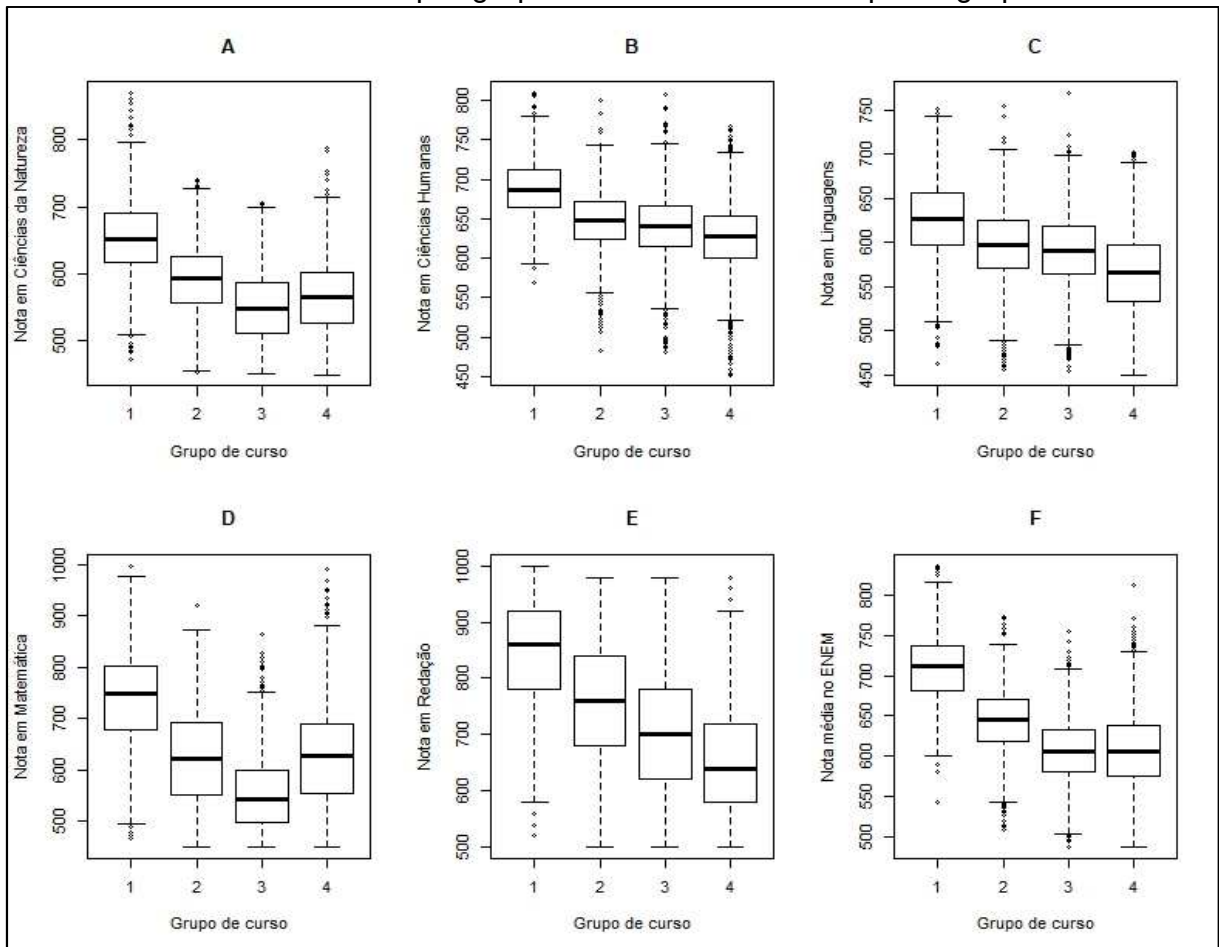
Neste caso, o decaimento seria o esperado, visto que, quanto maior o nível do grupo, menor é o desempenho dos candidatos ingressantes no ENEM.

A Figura 3 expõe seis grupos de *boxplots*, cada um deles representa uma das cinco áreas de conhecimento exploradas no ENEM, exceto o último que está associado à nota média geral nesse exame. Estes grupos foram codificados da letra A a F, de maneira que a distribuição das áreas de conhecimento pelos códigos se dá da seguinte maneira:

- A. Desempenho em Ciências da Natureza e suas Tecnologias - CN
- B. Desempenho em Ciências Humanas e suas Tecnologias – CH
- C. Desempenho em Linguagens, Códigos e suas Tecnologias – LC
- D. Desempenho em Matemática e suas Tecnologias – MT
- E. Desempenho em Redação
- F. Desempenho Geral no ENEM

Em cada uma dessas áreas de conhecimento, há quatro grupos correspondentes aos cursos de graduação ofertados pela UFRN em 2016, os quais foram gerados por meio da análise de agrupamento realizada, utilizando-se o método de *Ward*.

Figura 3 - Desempenho no Exame Nacional do Ensino Médio - ENEM geral e por área de conhecimento por grupos de cursos formados pelo agrupamento

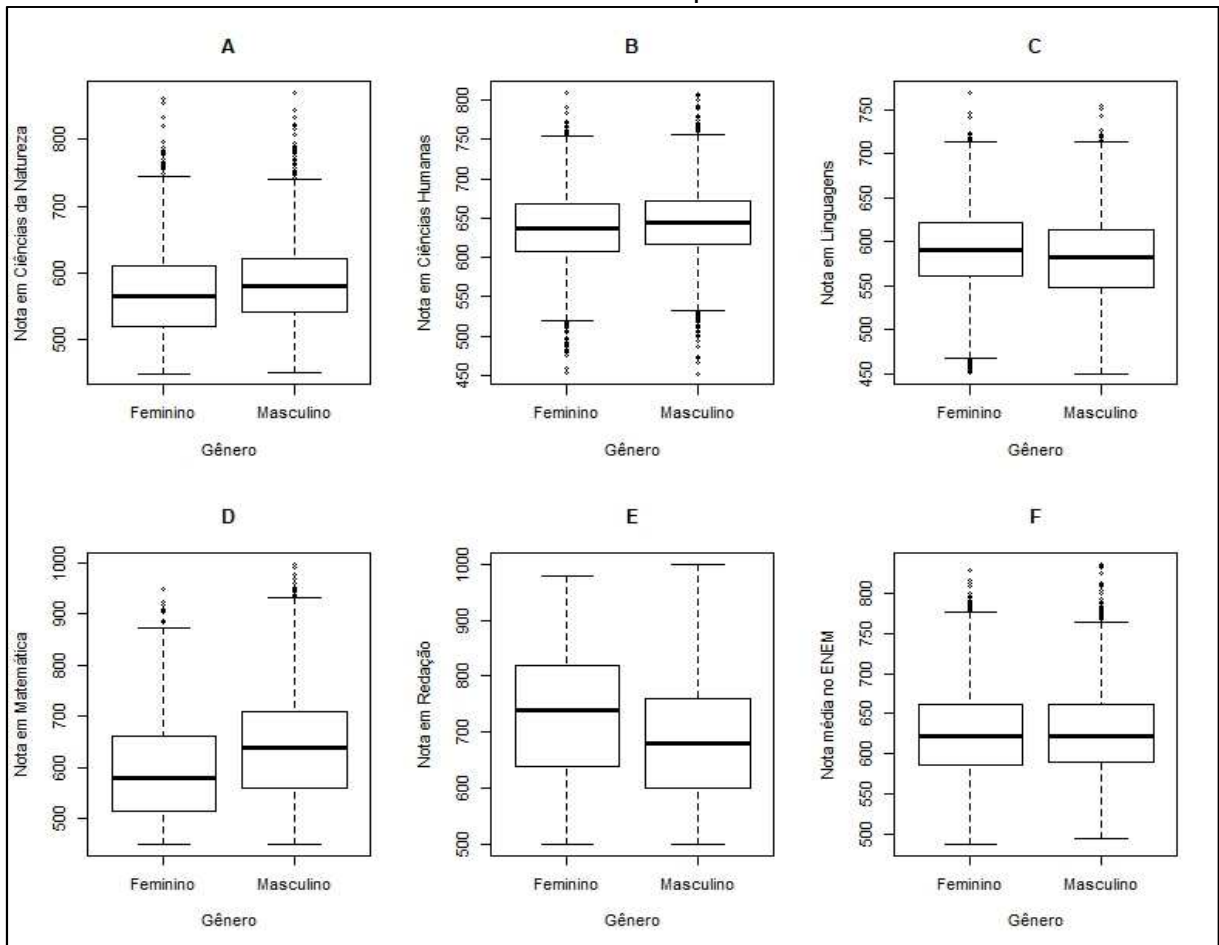


Fonte: COMPERVE

Analisando os gráficos em relação ao desempenho em cada uma das áreas de conhecimento do ENEM (A, B, C, D, E) dispostos na Figura 3 é possível perceber que não existe um padrão de decrescimento na mediana dos grupos em todas as áreas, em relação à Matemática e suas Tecnologias (*boxplot* D), por exemplo, a mediana do grupo 4 (baixo desempenho) se iguala a mediana do grupo 2 (desempenho mediano). Analisando o gráfico referente ao desempenho geral no ENEM (*boxplot* F), é importante chamar a atenção para os *boxplots* dos grupos 3 e 4; estes apresentam medianas aproximadamente iguais, no entanto, a variabilidade do grupo 4 é maior, assim como a quantidade de *outliers*.

A Figura 4 é similar à anterior, porém, cada grupo de *boxplots* compara o gênero dos alunos ingressantes por área de conhecimento do ENEM e pelo seu desempenho Geral.

Figura 4 – Desempenho geral e por área de conhecimento no Exame Nacional do Ensino Médio – ENEM por Gênero



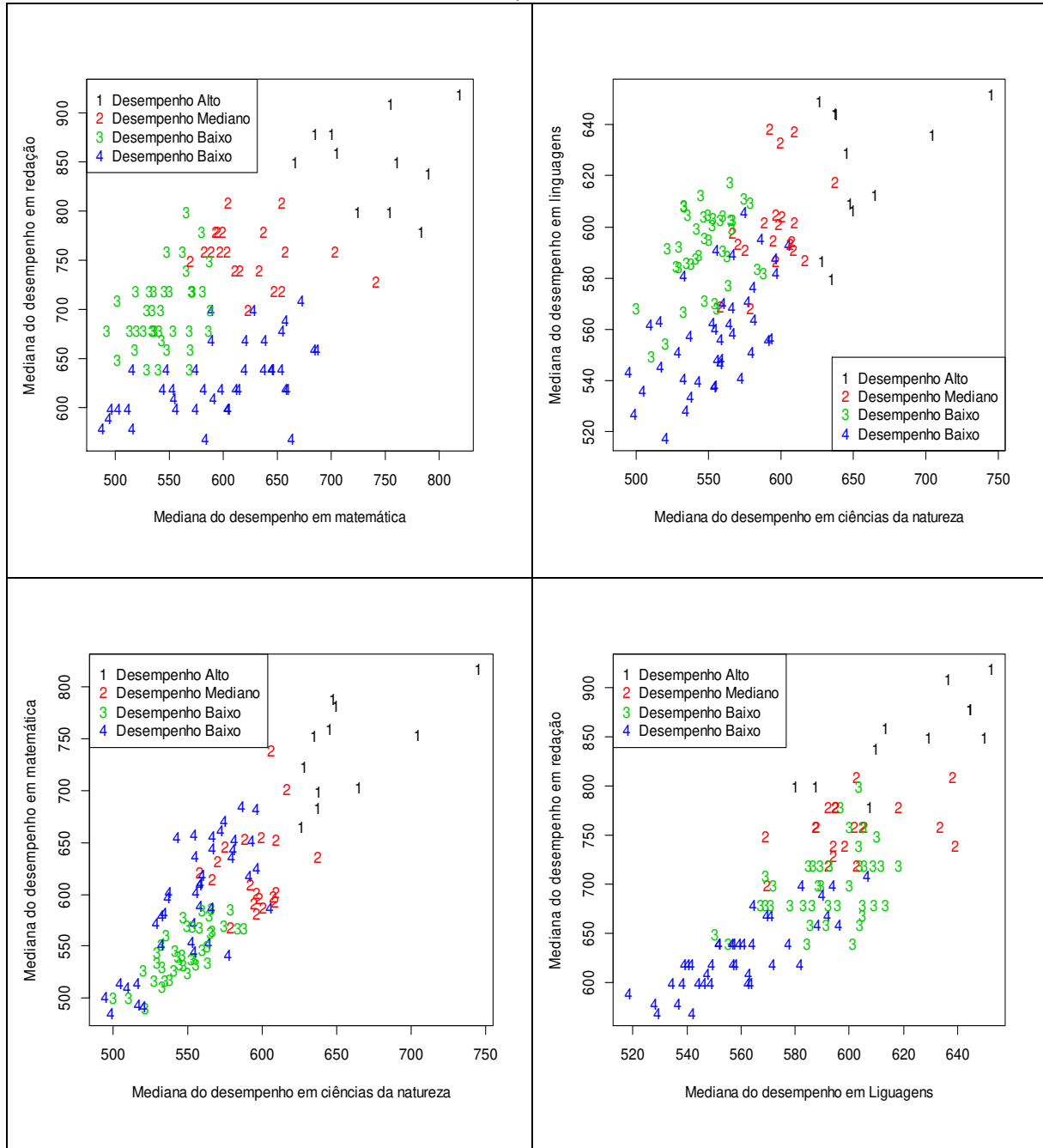
Fonte: COMPERVE

Analisando o desempenho geral dos candidatos ingressantes na UFRN em 2016, exibido no grupo de *boxplots* F, na Figura 4, nota-se que a mediana correspondente ao desempenho dos homens e a correspondente a das mulheres, são muito próximas. Todavia em relação ao desempenho em cada área de conhecimento do ENEM, ressalta-se que, em Matemática e suas Tecnologias (*boxplots* D) a mediana do desempenho dos homens é maior que a das mulheres, já em relação ao desempenho em Linguagem, Código e suas Tecnologias, e em Redação, (*boxplots* C e E respectivamente) essa relação se inverte.

Diante da semelhança apresentada na Figura 3, entre o grupo 3 e o 4, e, buscando entender melhor a estrutura dos dados apresentados no Quadro 1, decidiu-se analisar a dispersão dos dados dois a dois. A Figura 5, dispõe de um conjunto com 4 gráficos de dispersão entre as variáveis MT e Redação, superior

esquerdo; LC e CN, superior direito; CN e MT inferior esquerdo; LC e Redação, inferior direito.

Figura 5 – Gráfico de dispersão do desempenho mediano em algumas áreas de conhecimento do ENEM considerando os grupos de cursos de Graduação, na UFRN, em 2016



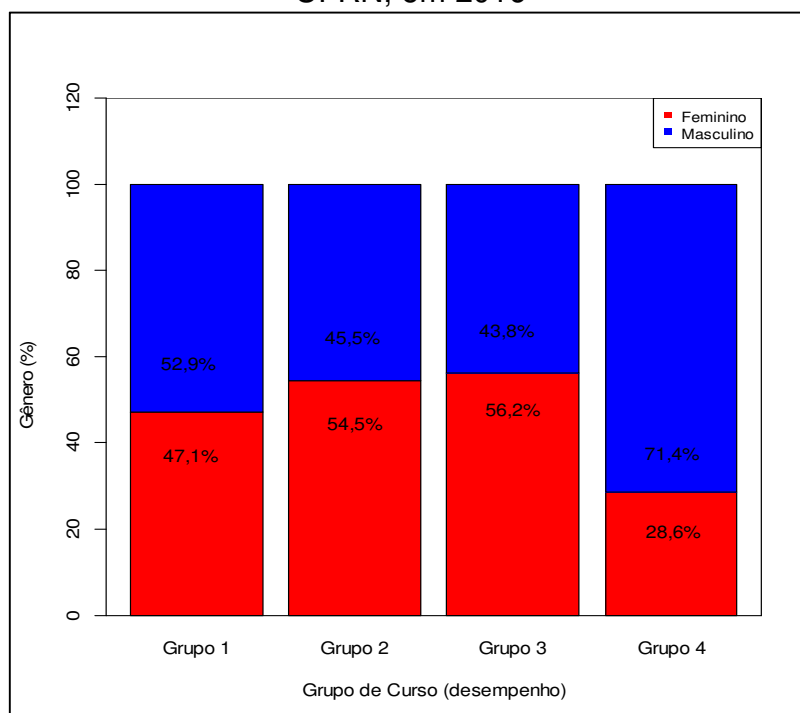
Fonte: COMPERVE

Pode-se observar, que algumas combinações de variáveis duas a duas apresentam correlação mais forte, essa relação pode ser visualizada por meio do crescimento quase linear, diferentemente de outras, que apresentam uma dispersão maior. De maneira clara, pode-se perceber a diferença existente entre os grupos de

cursos 3 e 4. Enquanto nos cursos do grupo 3 os ingressantes possuem bom desempenho em Redação e um desempenho péssimo em Matemática, para os cursos do grupo 4, essa relação ocorre de maneira contrária, ou seja, os ingressantes apresentam um alto desempenho em Matemática, no entanto, o desempenho em Redação é baixo. Assim, os grupos 3 e 4 são tão distintos que, refazendo a análise de agrupamentos e dividindo os cursos de graduação oferecidos pela UFRN em 2016 em 3 grupos, o método une os grupos 1 e 2, ao invés do 3 e 4, pois a diferença entre esses últimos é tão grande que a divisão em 3 grupos torna os grupos 1 e 2 semelhantes.

Sabendo da relação entre os desempenhos masculinos com matemática e femininos com redação, expostos na Figura 4, e da relação de desempenho dos grupos 3 e 4 com as referidas áreas de conhecimento, exibidos na Figura 5, optou-se por estudar a distribuição do gênero por grupo de cursos, a Figura 6 mostra um gráfico de barras com os percentuais de gênero, por grupo de cursos.

Figura 6 – Distribuição percentual do gênero por grupos de cursos de graduação da UFRN, em 2016



Fonte: COMPERVE

A distribuição do sexo por grupo de curso se dá de maneira aproximadamente simétrica, em torno dos 50%, exceto no grupo 4, pois neste 71,4% dos ingressantes são homens. Associando este resultado, ao disposto no Quadro 1,

é possível entender que esse dado é esperado, visto que o grupo 4 é composto majoritariamente por cursos da área das Ciências Exatas, tais como Estatística, Física, Matemática, dentre outros, e, historicamente esses são cursos cuja entrada tem grande predominância do sexo masculino.

4.2 ANÁLISE EXPLORATÓRIA

Partindo do pressuposto de que existe uma relação entre os desempenhos dos candidatos ingressantes na UFRN e os fatores socioeconômicos que o caracterizam, esta parte da pesquisa busca conhecer quais variáveis socioeconômicas possuem maior influência no desempenho no ENEM e, ainda, se essa influência alcança o desempenho acadêmico.

Para representar o desempenho acadêmico este trabalho utilizou a Média de Conclusão – MC em abril de 2017.

4.2.1 Média de Conclusão

A média ponderada do rendimento escolar final nas disciplinas cursadas em que o aluno conseguiu êxito ao longo do curso, é representada por MC. Não entram no cálculo dessas médias as disciplinas que foram trancadas, canceladas e reprovadas, assim como, as atividades complementares e os componentes curriculares cujo rendimento não seja expresso de forma numérica. A ponderação dessa média é feita pela carga horária do discente nos componentes curriculares que ele obteve sucesso.

A escolha por esse índice para retratar o desempenho acadêmico se deu porque a universidade, atualmente, utiliza-o para representar o rendimento de seus alunos no histórico escolar dos mesmos, conjuntamente com outro índice: a Média de Conclusão Normalizada – MCN, a qual é uma padronização da MC. A distribuição geral dos valores do MC associados aos dados estudados pode ser visualizada na Tabela 2 exposta a seguir.

Tabela 2 – Distribuição geral dos valores de MC

Estatísticas	Valores
Mínimo	0,00
Primeiro quartil	6,35
Mediana	7,29
Média	6,64
Terceiro quartil	8,15
Máximo	10,00

Fonte: COMPERVE

É interessante observar a diferença entre a média e a mediana, isso ocorre pelo fato da média ser sensível à valores discrepantes (*outliers*), por este motivo neste trabalho as comparações são feitas utilizando as medianas e não as médias.

4.2.2 Idade

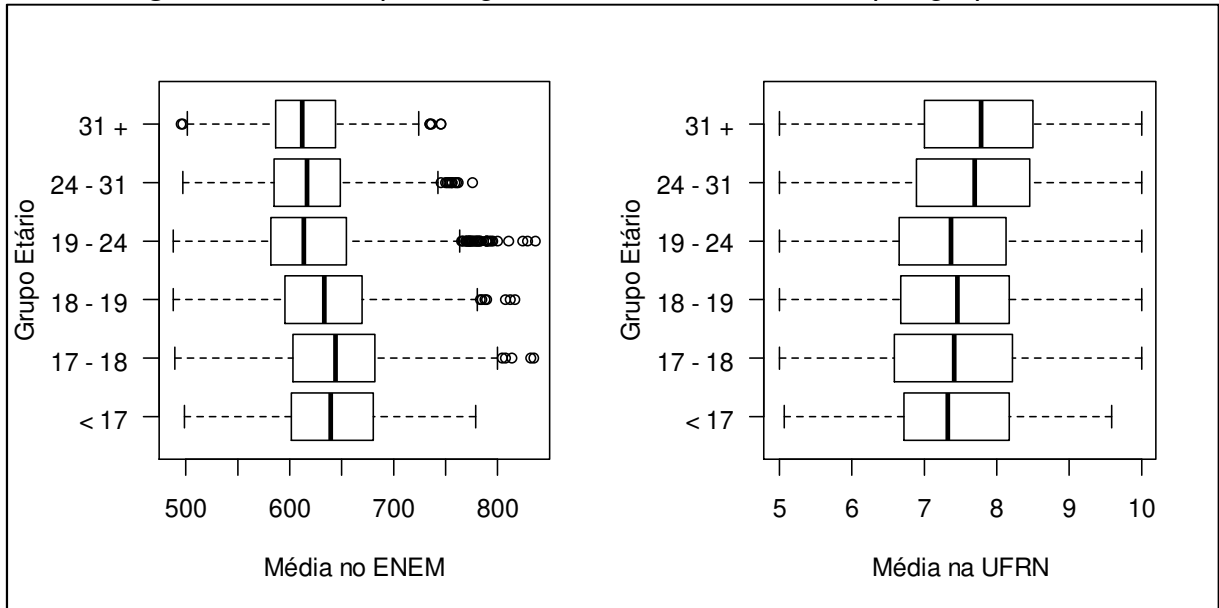
Com a finalidade de distribuir melhor as frequências das classes e assim conseguir maior compreensão no tocante à variável idade, esta, foi categorizada de acordo com os percentis 10, 25, 50, 75 e 90 conforme já foi explicado anteriormente. A Tabela 3 expõe os respectivos valores dos percentis calculados.

Tabela 3 – Percentis calculados para a variável idade

Percentil	Idade
10	17
25	18
50 ou mediana	19
75	24
90	31

Fonte: COMPERVE

A Figura 7 apresenta dois grupos de *boxplots*, um expondo o desempenho no ENEM, e outro, o desempenho na UFRN dos candidatos que fizeram ENEM em 2015 e ingressaram na UFRN em 2016. Ambos os grupos dispõem de seis gráficos cada um refletindo o desempenho dos candidatos por grupo etário.

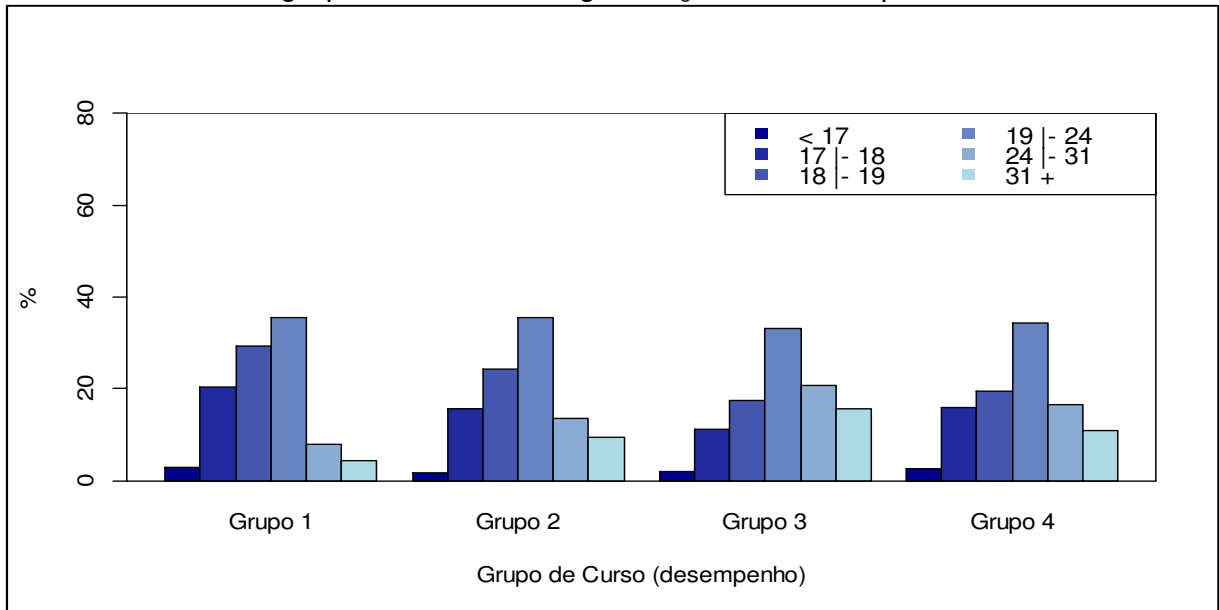
Figura 7 – Desempenho geral no ENEM e na UFRN por grupo etário

Fonte: COMPERVE

Analisando os desempenhos no ENEM por grupo etário, exposto na Figura 7, pode-se perceber que a idade influencia o desempenho dos alunos, de modo que, os candidatos ingressantes com idade entre 17 e 18 anos possuem desempenho médio melhor que os demais, e, à medida que a idade aumenta, o desempenho médio diminui. Esse resultado é esperado, visto que a faixa etária que apresentou melhor desempenho é justamente a dos alunos que estão concluindo o Ensino Médio. No tocante ao desempenho na UFRN, os candidatos ingressantes com idades mais altas possuem melhor desempenho mediano, este resultado pode indicar que pessoas com idades maiores possuem maior compromisso com os estudos.

A Figura 8 apresentada a seguir exibe a distribuição percentual dos grupos etários dos candidatos ingressantes dentro dos quatro grupos de cursos de graduação ofertados pela UFRN em 2016.

Figura 8 – Distribuição percentual do grupo etário dos candidatos ingressantes considerando os grupos de cursos de graduação ofertados pela UFRN, em 2016

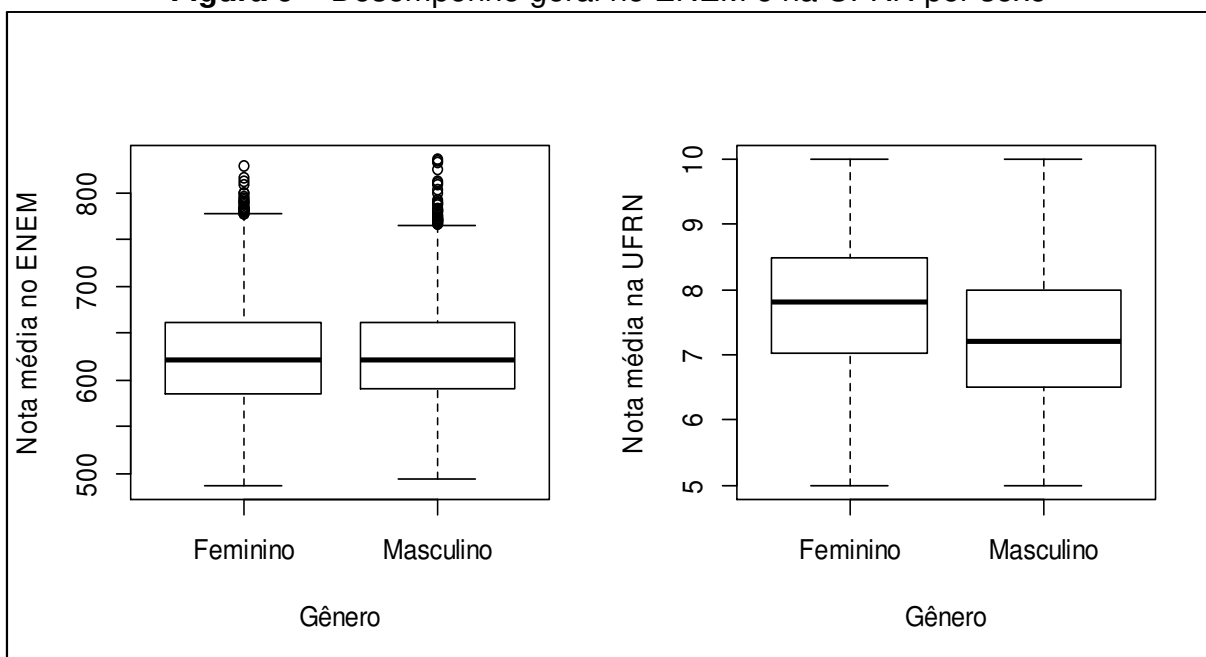


Fonte: COMPERVE

Analisando a Figura 8 é possível identificar que a proporção dos grupos etários 17-18 e 18-19, decaem quando o nível do grupo de cursos de graduação aumenta, enquanto que para os grupos etários de 24-31 e 31 + essa relação ocorre de maneira inversa, ou seja, a proporção deles aumenta à medida que o nível do grupo de cursos de graduação aumenta. É interessante frisar que a distribuição ocorre de maneira distinta, de modo que o grupo 1 apresenta uma leve assimetria, enquanto que o grupo 4 é o menos assimétrico.

4.2.3 Gênero

A Figura 9 dispõe de dois grupos de *boxplots*, um expondo o desempenho no ENEM (lado esquerdo), e outro o desempenho na UFRN (lado direito) dos candidatos que ingressaram na UFRN em 2016. Ambos os grupos dispõem de dois gráficos, cada um, refletindo o desempenho dos candidatos por sexo.

Figura 9 – Desempenho geral no ENEM e na UFRN por sexo

Fonte: COMPERVE

Analisando os desempenhos no ENEM, por sexo, exposto nos gráficos do lado esquerdo nota-se que não há diferença significativa entre as medianas dos desempenhos para os sexos, neste caso é interessante ressaltar que o percentil 90 (limite superior representado pela haste superior) do sexo feminino é maior que o masculino. Já para os desempenhos na UFRN, exibidos nos gráficos do lado direito, a mediana do sexo feminino é maior que a mediana do sexo masculino.

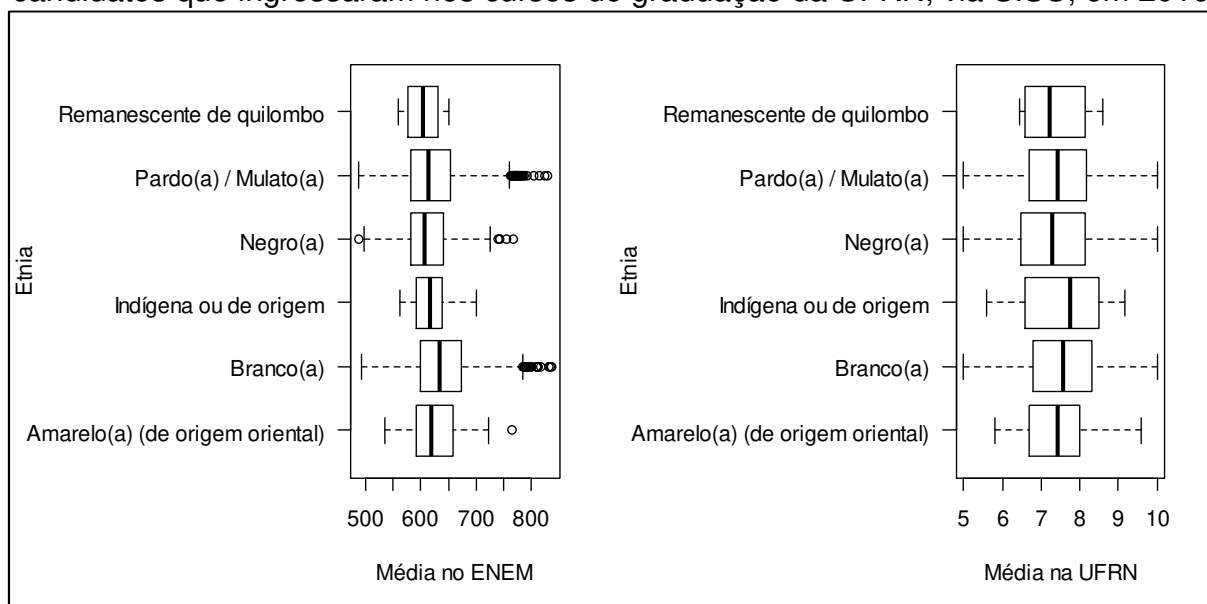
No trabalho realizado por Hiranilson Andrade Santos, com metodologia semelhante à utilizada neste, porém utilizando dados do vestibular da UFRN 2012, esse resultado para a variável sexo se repete. Considerando a pressão imposta pelas exigências do mercado de trabalho por mão de obra qualificada, e, sendo o ingresso no Ensino Superior a principal alternativa para obter tal qualificação, o nível de estresse e a ansiedade dos candidatos, é alto (GOMES; SPARTA, 2005). Associando esta informação à discussão realizada no parágrafo anterior, esse resultado pode ser um indicativo de que as mulheres são mais vulneráveis a fatores psicológicos que os homens.

4.2.4 Etnia

A Figura 10 dispõe de dois grupos de *boxplots*, um expondo o desempenho no ENEM e outro o desempenho na UFRN dos candidatos que ingressaram na

UFRN em 2016. Ambos os grupos dispõem de seis gráficos, cada um, refletindo o desempenho dos candidatos considerando a sua etnia.

Figura 10 – Desempenho geral no ENEM e na UFRN segundo a etnia dos candidatos que ingressaram nos cursos de graduação da UFRN, via SISU, em 2016



Fonte: COMPERVE

Analisando a Figura 10, tanto em relação ao desempenho no ENEM, quanto na UFRN, observa-se que existe diferença de desempenho entre as etnias. No que se referem ao desempenho no ENEM os grupos étnico branco e pardo / mulato são os grupos étnicos dos candidatos com maior desempenho mediano. Em relação à UFRN os grupos que se destacam são brancos e indígenas ou de origem indígena.

A Tabela 4 dispõe da distribuição dos candidatos ingressantes na UFRN em 2016, por etnia, o total não soma 6.664 alunos porque para esta variável houve 385 dados faltantes.

Tabela 4 – Distribuição dos candidatos ingressantes na UFRN em 2016, por etnia

Etnia	Total absoluto	Percentual
Pardo (a)/ Mulato (a)	2.968	47,27%
Branco(a)	2.751	43,81%
Negro(a)	477	7,60%
Amarelo (a) (de origem oriental)	43	0,68%
Indígena ou de origem	36	0,57%
Remanescente de quilombola	4	0,06%
TOTAL	6.279	100%

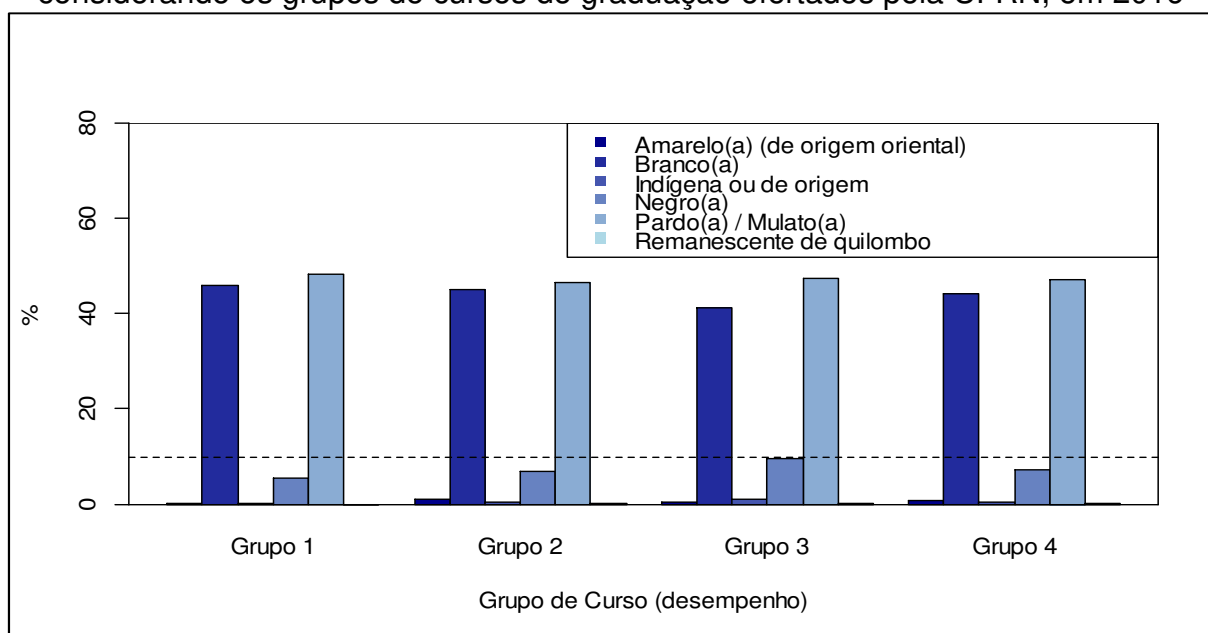
Fonte: COMPERVE

A análise da Tabela 4 possibilita chamar a atenção para o número absoluto e o percentual de negros, amarelos, indígenas e remanescente de quilombola.

Assim, apesar do resultado disposto na Figura 10, em que o branco e o indígena aparecem com maiores desempenhos médios na UFRN, esse resultado pode não refletir com fidelidade o resultado apresentado, visto que o percentual de pessoas indígenas ou de origem indígena que ingressou na UFRN é tão baixo que não garante a consistência desse resultado. Vale destacar que o percentual de brancos e pardos / mulatos representam 91,08% dos candidatos ingressantes.

A Figura 11 apresenta a distribuição percentual da etnia dos candidatos ingressantes dentro dos quatro grupos de cursos de graduação ofertados pela UFRN em 2016.

Figura 11 – Distribuição percentual da etnia dos candidatos ingressantes considerando os grupos de cursos de graduação ofertados pela UFRN, em 2016



Fonte: COMPERVE

A análise da Figura 11, deixa evidente que os grupos étnicos com maior percentual são os brancos e pardos / mulato, resultado que já era esperado dado o alto percentual da entrada destes dois grupos. Todavia é necessário chamar a atenção para o fato de que à medida que o nível do grupo de cursos aumenta, o percentual de candidatos declarados brancos e pardos / mulato apresentam uma leve queda, exceto no grupo 4.

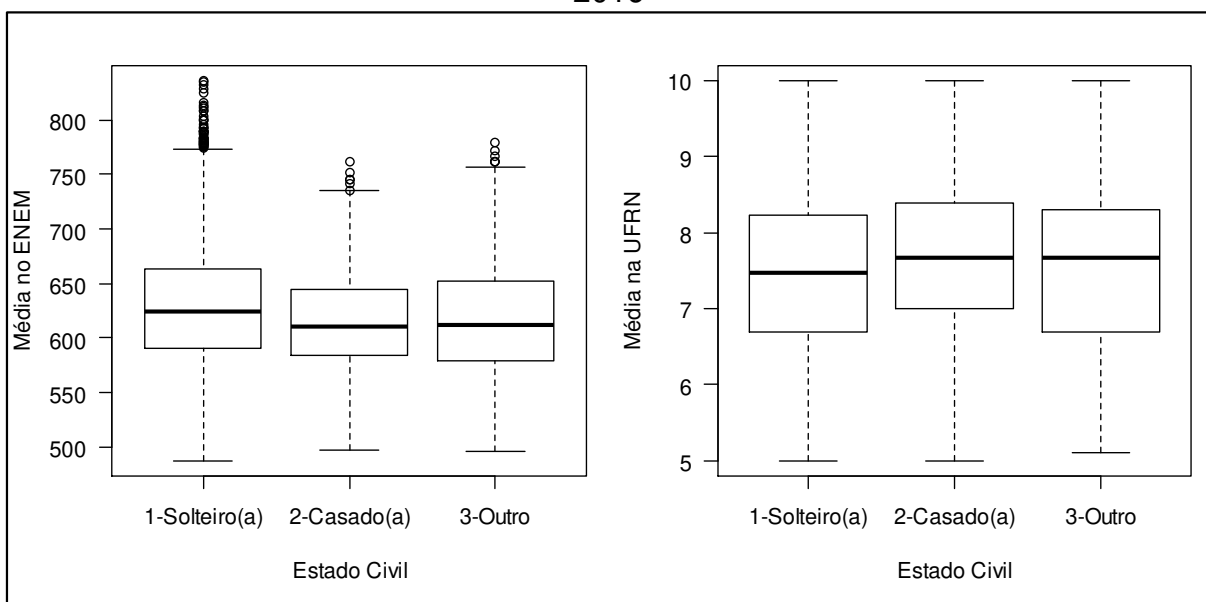
A linha tracejada, que marca em 10% o percentual da etnia, mostra que, em nenhum dos grupos de cursos, os candidatos autodeclarados negros ultrapassam este percentual. A distribuição da etnia acontece de maneira tão irregular que

existiram grupos étnicos que não puderam ser representados em todos os grupos, como é o caso dos remanescentes de quilombola, que não estão representados no grupo 1.

4.2.5 Estado Civil

A Figura 12 dispõe de dois grupos de *boxplots*, sendo um expondo o desempenho no ENEM e outro o desempenho na UFRN dos candidatos que ingressaram na UFRN em 2016. Ambos os grupos dispõem de três gráficos, cada um, estes refletem o desempenho do candidato em relação ao seu estado civil.

Figura 12 – Desempenho geral no ENEM e na UFRN de acordo com o estado civil dos candidatos que ingressaram nos cursos de graduação da UFRN, via SISU, em 2016

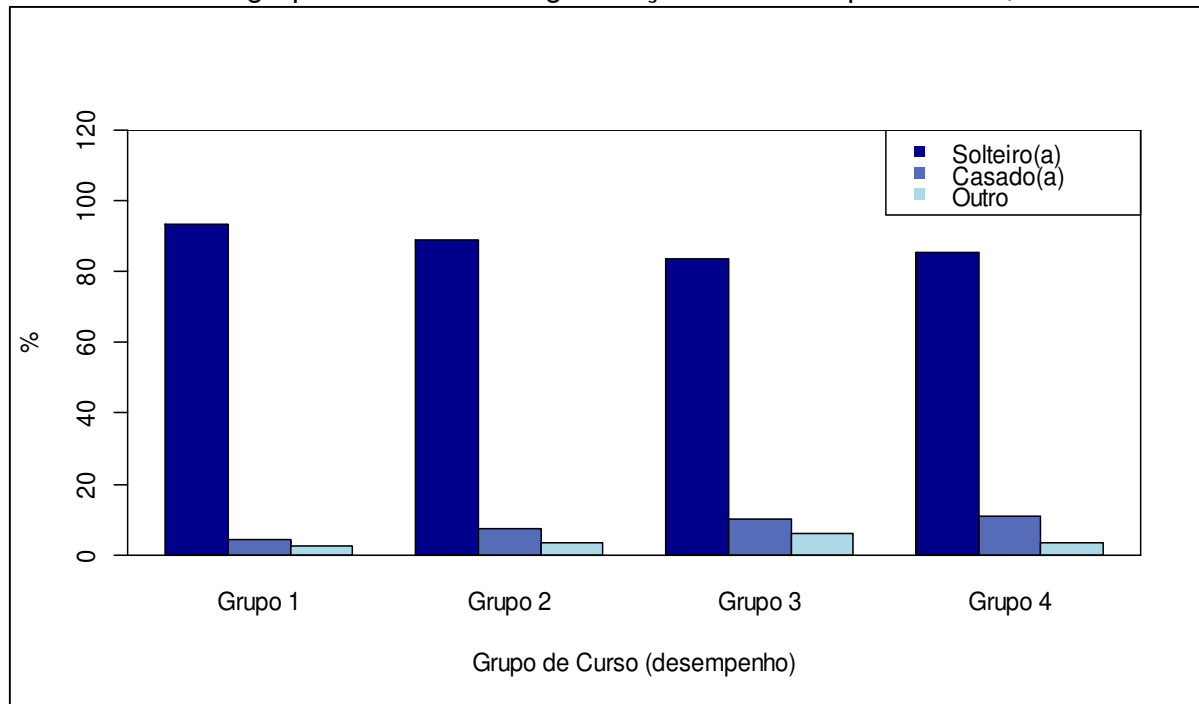


Fonte: COMPERVE

Analisando a Figura 12 em relação ao desempenho no ENEM pode-se mostrar que existe diferença no desempenho médio entre os solteiros e os casados: os solteiros possuem desempenho médio maior, o que de certa forma é esperado, supondo que estes possuem mais tempo para se dedicarem aos estudos. Esta relação, considerando os desempenhos na UFRN, surge de maneira inversa: a mediana do desempenho dos ingressantes casados aparece um pouco maior que as dos demais ingressantes, podendo indicar que os casados possuem maior compromisso.

A Figura 13 apresenta a distribuição percentual do estado civil dos candidatos ingressantes, dentro dos quatro grupos de cursos de graduação ofertados pela UFRN em 2016.

Figura 13 – Distribuição percentual do estado civil dos candidatos ingressantes conforme os grupos de cursos de graduação ofertados pela UFRN, em 2016



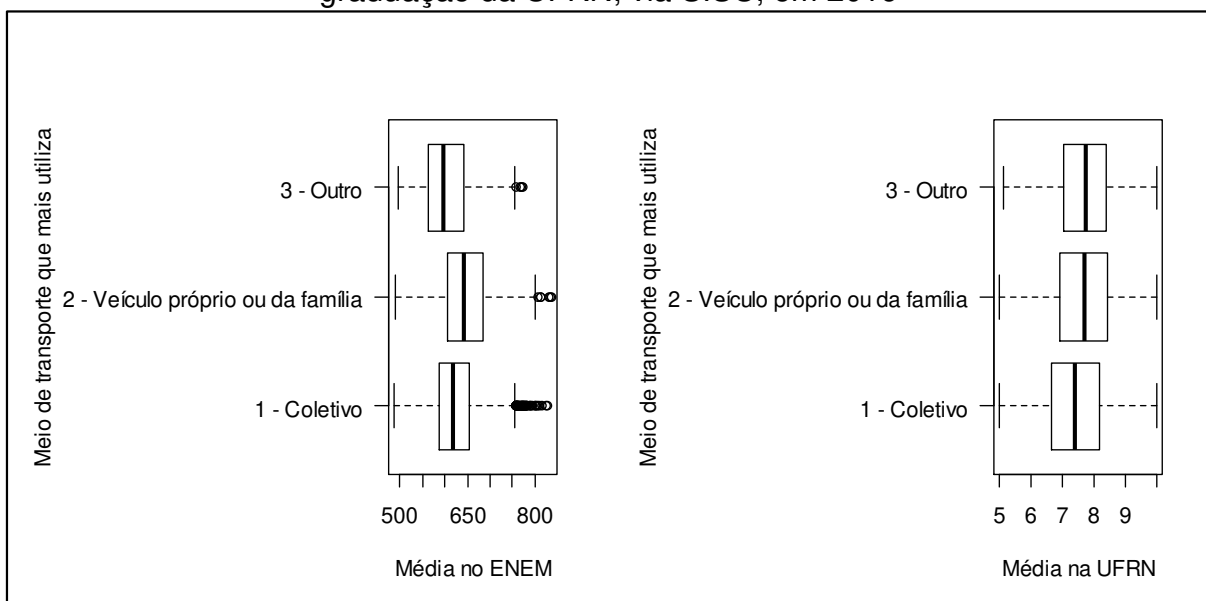
Fonte: COMPERVE

Essa Figura 13 mostra que os solteiros são maioria dentro dos grupos de cursos de graduação e que eles decrescem um pouco nos dois últimos grupos (3 e 4). Considerando que o aumento do nível do grupo implica em uma piora no desempenho dos candidatos ingressantes, este resultado é esperado visto que, pressupõe-se que os solteiros possuem mais tempo para se dedicar aos estudos.

4.2.6 Meio de transporte que mais utiliza

A Figura 14 é semelhante à Figura 12, no entanto, agora, para cada grupo de *boxplot* são analisados o meio de transporte que os candidatos ingressantes mais utilizam.

Figura 14 – Desempenho geral no ENEM e na UFRN considerando o meio de transporte mais utilizado pelos candidatos que ingressaram nos cursos de graduação da UFRN, via SISU, em 2016

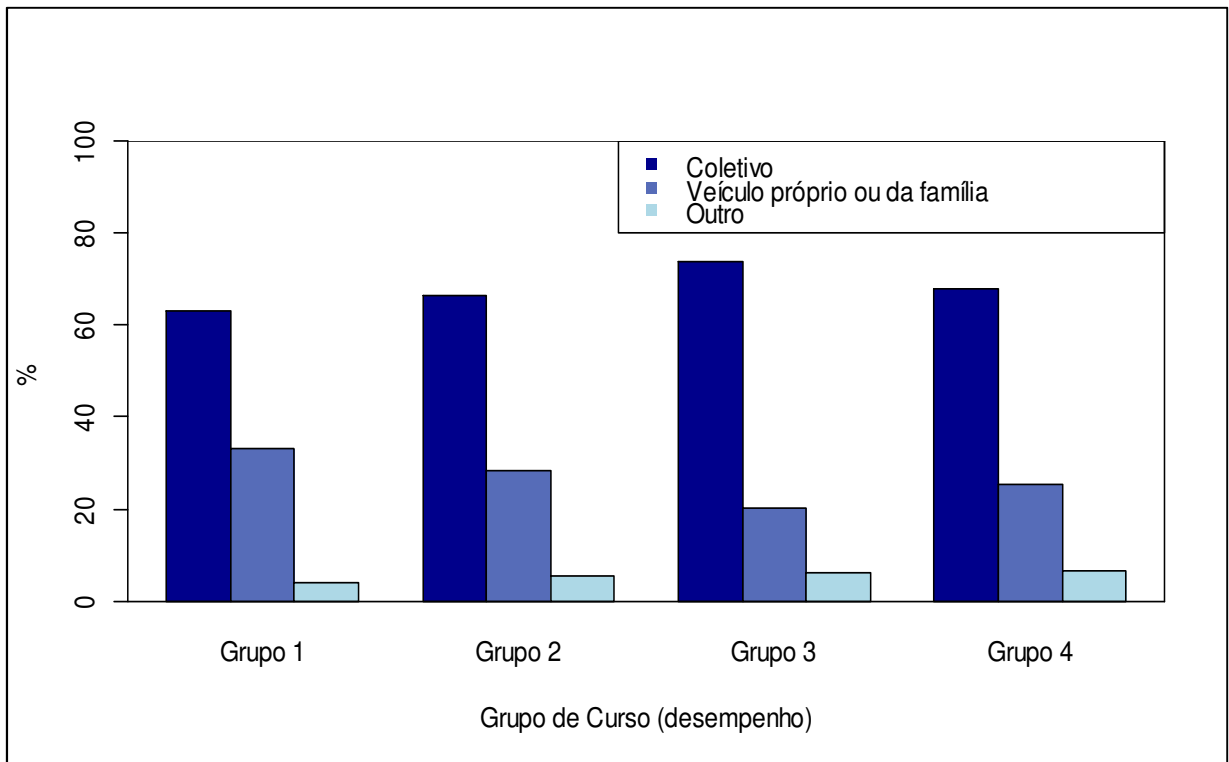


Fonte: COMPERVE

Analisando a Figura 14 em relação ao ENEM nota-se que existe diferença no desempenho médio dos candidatos de acordo com o meio de transporte que mais utilizam: os candidatos que utilizam veículo próprio ou da família, possuem desempenho médio maior que os demais. Já em relação à UFRN, os desempenhos médios dos candidatos que utilizam veículo próprio ou da família, e os dos que utilizam outro meio de transporte não apresentam tanta diferença, no entanto, ambos possuem desempenho médio maior do que os que utilizam transporte coletivo.

A Figura 15 apresenta a distribuição percentual do meio de transporte mais utilizado pelos candidatos ingressantes dentro dos quatro grupos de cursos de graduação ofertados pela UFRN em 2016.

Figura 15 – Distribuição percentual do meio de transporte mais utilizado pelos candidatos ingressantes por grupos de cursos de graduação ofertados pela UFRN, em 2016



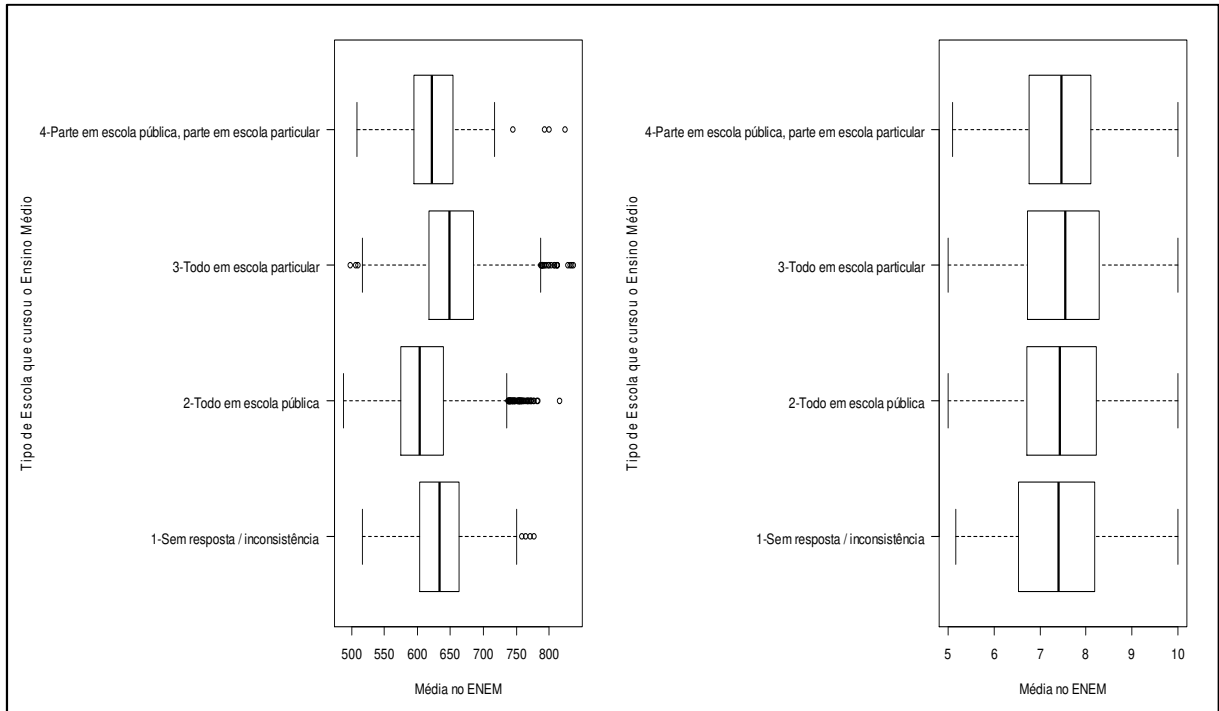
Fonte: COMPERVE

A Figura 15 mostra que o transporte coletivo é o transporte mais utilizado dentro de todos os grupos de cursos de graduação, tendo apenas ligeiro crescimento no grupo 3. Quanto ao transporte próprio ou da família o maior percentual está associado ao grupo 1 e o menor percentual ao grupo 3.

4.2.7 Tipo de escola que cursou o Ensino Médio

A Figura 16 dispõe de dois grupos de *boxplots*, sendo um, representando o desempenho geral no ENEM e o outro, o desempenho na UFRN. Cada um desses grupos é composto por quatro *boxplots* os quais apresentam os tipos de escola que os candidatos ingressantes cursaram o Ensino Médio.

Figura 16 – Desempenho geral no ENEM e na UFRN segundo tipo de escola que cursou o Ensino Médio

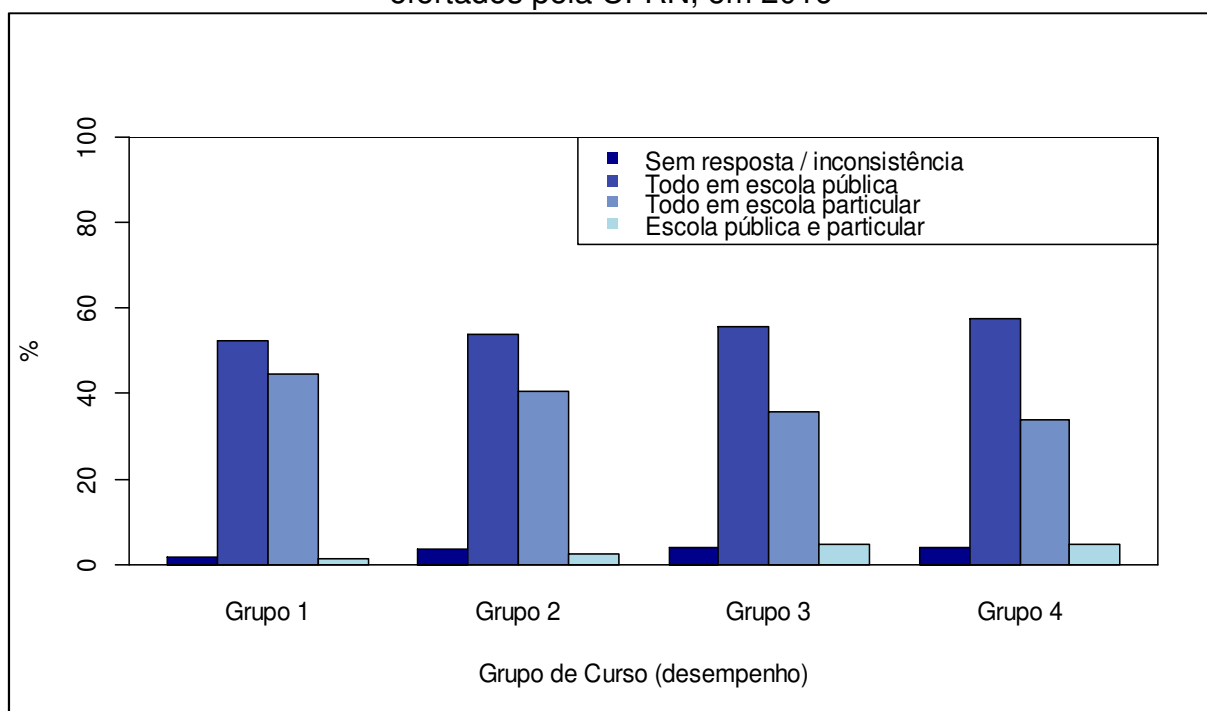


Fonte: COMPERVE

Em relação ao tipo de escola que cursou o Ensino Médio a Figura 16 mostra que existe distinção entre os desempenhos dos candidatos, de modo que os candidatos que cursaram o ensino superior todo em escola particular obtiveram desempenho mediano maior que os demais no ENEM.

A Figura 17 apresenta a distribuição percentual do tipo de escola que cursou o Ensino Médio pelos quatro grupos de cursos de graduação ofertados pela UFRN em 2016.

Figura 17– Distribuição percentual do tipo de escola na qual o ingressante, via SISU 2016, cursou o Ensino Médio de acordo com os grupos de cursos de graduação ofertados pela UFRN, em 2016



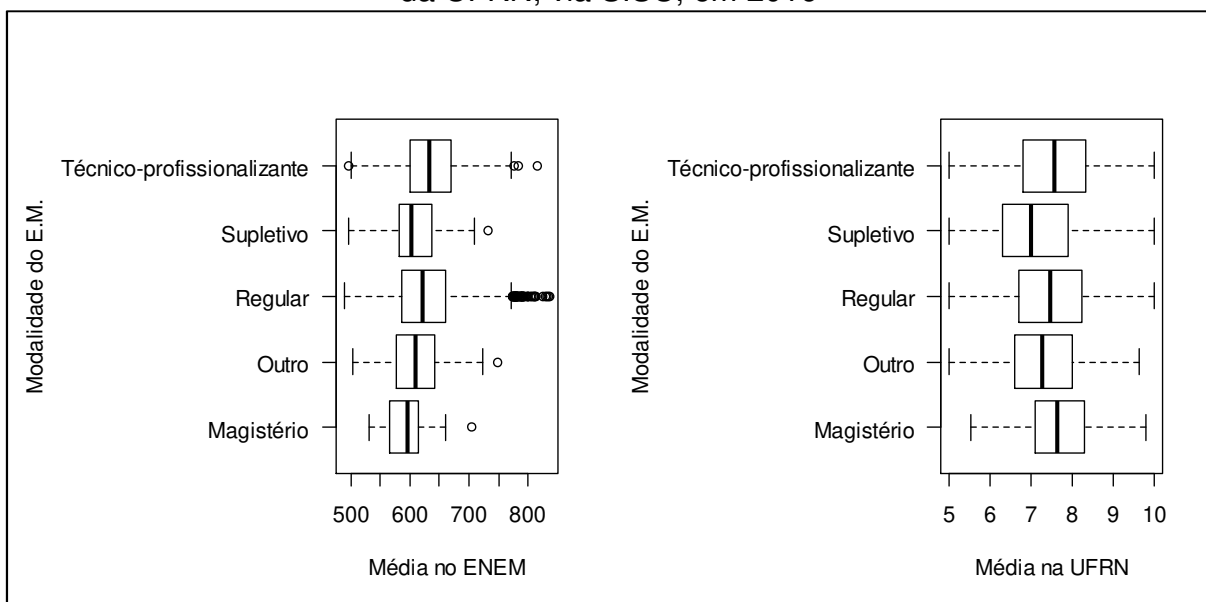
Fonte: COMPERVE

A análise da Figura 17 permite observar a relação existente entre o tipo de escola que os candidatos ingressantes cursaram o Ensino Médio e seu desempenho no ENEM dentro dos grupos de cursos de graduação. Assim, é necessário ressaltar que no grupo 1 apresenta percentuais muito parecidos quanto ao Ensino Médio cursado em escola todo em escola particular, além disso, os grupos de menor desempenho (3 e 4) possuem percentuais semelhantes quanto ao Ensino Médio cursado todo em escola particular e parte em escola pública, parte em escola particular.

4.2.8 Modalidade do Ensino Médio

A Figura 18 dispõe de dois grupos de *boxplots*: um expõe o desempenho no ENEM, e, o outro, o desempenho na UFRN, dos candidatos que ingressaram em 2016. Ambos os grupos dispõem de cinco gráficos cada um refletindo o desempenho do candidato em relação à modalidade do Ensino Médio cursado pelo mesmo.

Figura 18 – Desempenho geral no ENEM e na UFRN de acordo com modalidade de Ensino Médio cursado pelos candidatos que ingressaram nos cursos de graduação da UFRN, via SISU, em 2016

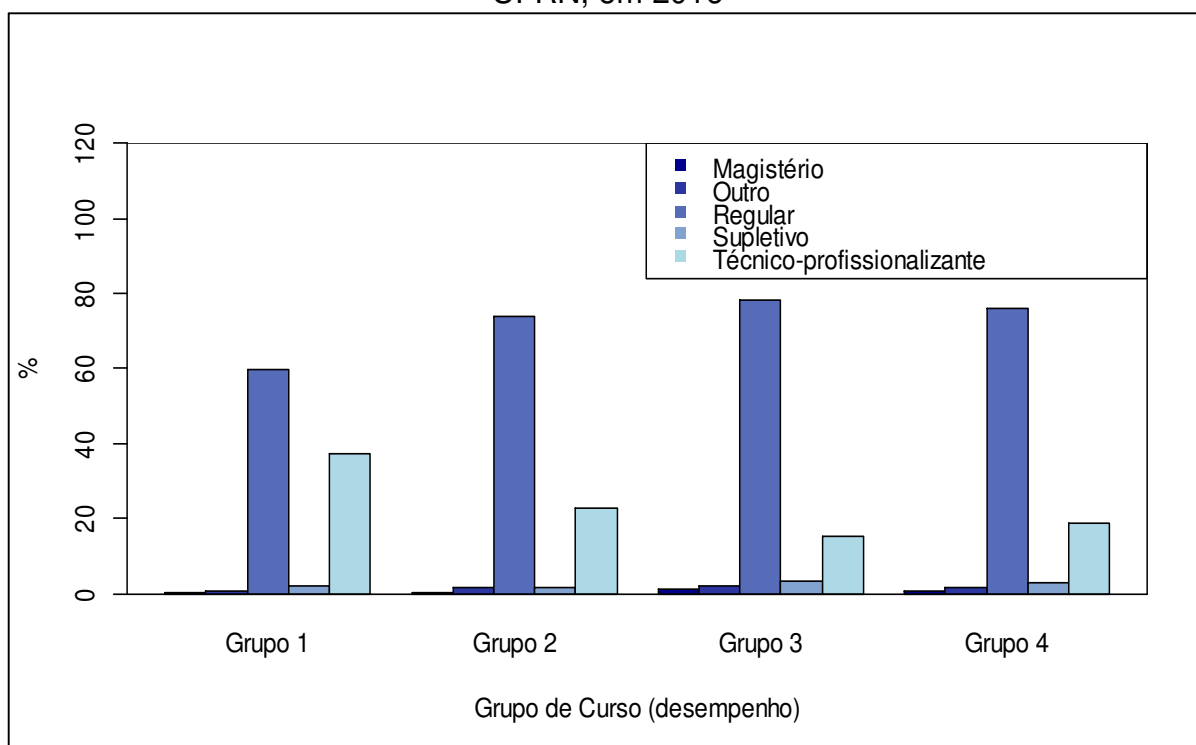


Fonte: COMPERVE

Analisando a Figura 18 é evidente a relação existente entre o desempenho médio dos candidatos no ENEM e na UFRN, e, a modalidade do Ensino Médio por eles cursado. É relevante evidenciar o fato de que, tanto no ENEM quanto na UFRN, os candidatos ingressantes que fizeram Ensino Médio técnico-profissionalizante possuem desempenho médio melhor que os demais, ao contrário dos candidatos ingressantes que cursaram o supletivo, estes apresentam desempenho médio abaixo dos demais.

A Figura 19 apresenta a distribuição percentual da modalidade do Ensino Médio cursado pelos candidatos ingressantes de acordo com os quatro grupos de cursos de graduação ofertados pela UFRN em 2016.

Figura 19 – Distribuição percentual da modalidade do Ensino Médio cursado pelos candidatos ingressantes segundo os grupos dos cursos de graduação ofertados pela UFRN, em 2016



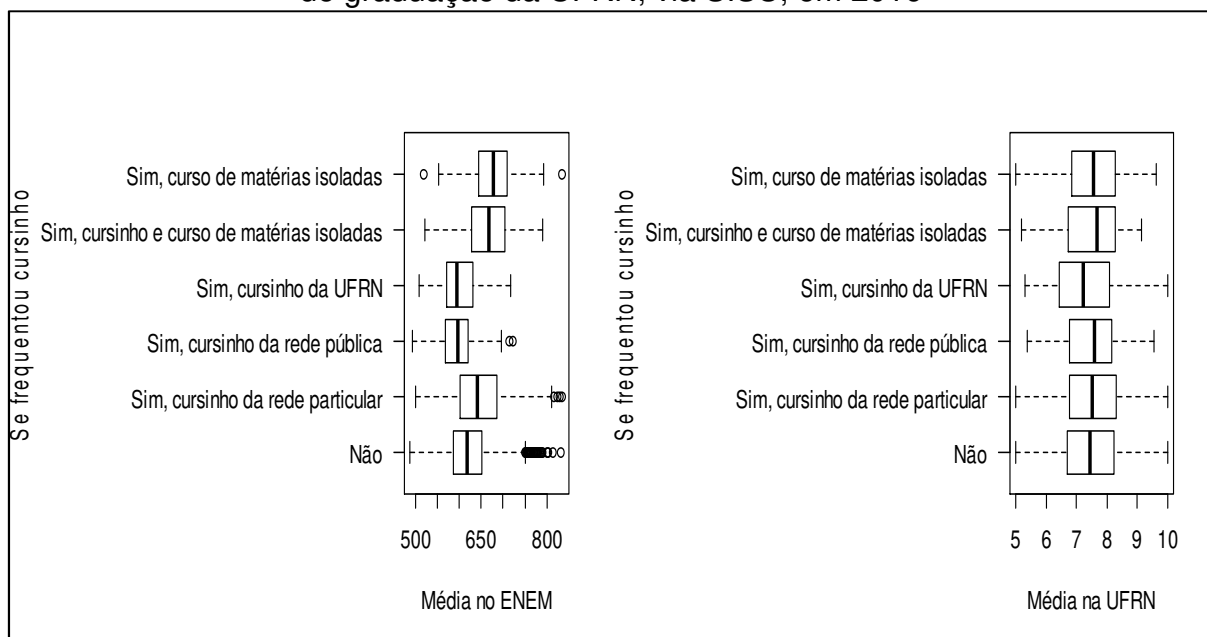
Fonte: COMPERVE

A Figura 19 mostra que o Ensino Médio regular é a modalidade que possui maior porcentagem em todos dos grupos de cursos de graduação. O grupo 1 apresenta o maior percentual quanto à modalidade de ensino técnico-profissionalizante, e o menor em relação ao ensino regular. Os grupos 3 e 4 (menores desempenhos), são muito parecidos quanto aos percentuais das modalidades de ensino regular e técnico-profissionalizante.

4.2.9 Variável categórica “Frequentou cursinho para prestar ENEM”

A Figura 20 dispõe de dois grupos de *boxplots*, sendo um expondo o desempenho no ENEM e outro o desempenho na UFRN dos candidatos ingressantes, via SISU, em 2016. Ambos os grupos dispõem de seis gráficos cada um mostra o desempenho do candidato em relação ao fato de ter frequentado cursinho para prestar ENEM e, em caso afirmativo, qual o tipo do cursinho.

Figura 20 – Desempenho geral no ENEM e na UFRN de acordo com o tipo de cursinho para prestar ENEM utilizado pelos candidatos que ingressaram nos cursos de graduação da UFRN, via SISU, em 2016



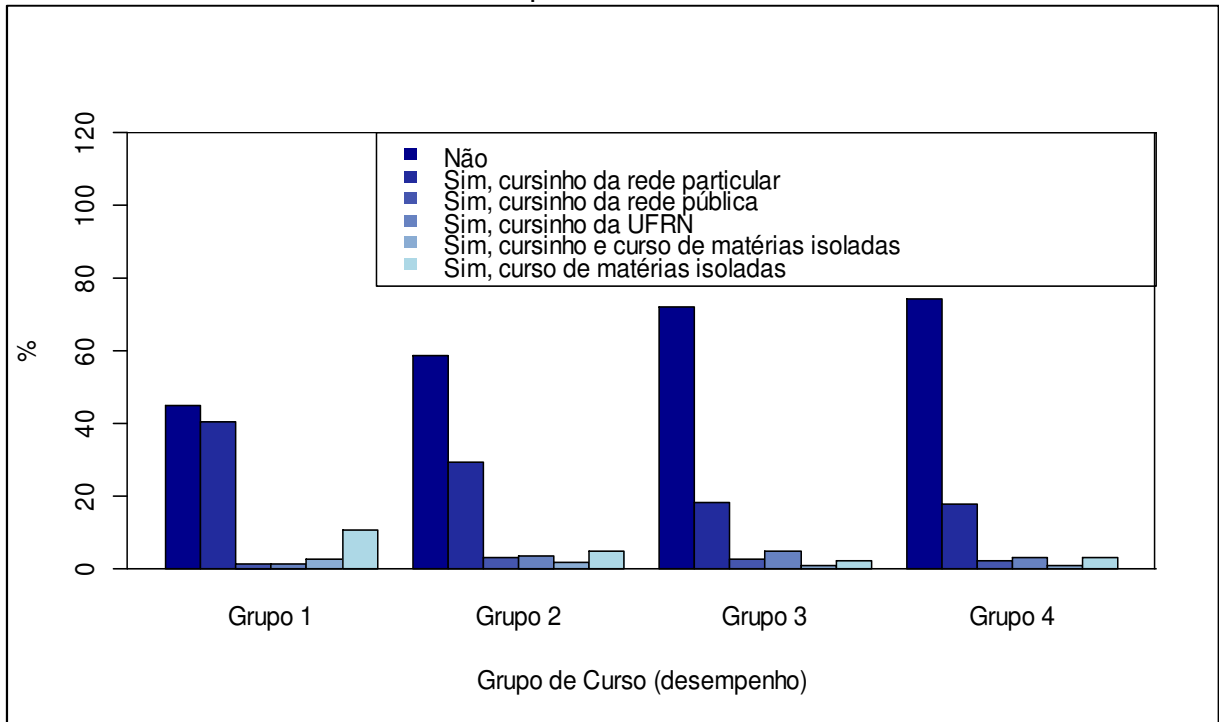
Fonte: COMPERVE

Analisando a Figura 20 em relação ao ENEM é perceptível que existe diferença no desempenho médio dos candidatos de acordo com a utilização do tipo de cursinho para prestar vestibular, de modo que candidatos que fizeram cursinhos isolados e os que fizeram cursinhos e matérias isoladas possuem maior desempenho, seguido dos que fizeram cursinho na rede privada. É importante ressaltar que os candidatos que não fizeram cursinho, obtiveram desempenho mediano melhor que os candidatos que tentaram o cursinho na rede pública ou os que tentaram o cursinho oferecido pela UFRN. Em relação ao desempenho na universidade, não há muita diferença entre as respostas, exceto, quando se trata dos candidatos que fizeram cursinho na própria instituição. Estes possuem desempenho médio menor que os demais.

Tanto para o desempenho no ENEM quanto na UFRN, o esperado seria que os candidatos que não fizeram cursinho tivessem desempenhos menores, no entanto isto não foi o que ocorreu.

A Figura 21 apresenta a distribuição percentual da variável “frequentou cursinho” dentro dos quatro grupos de cursos de graduação ofertados pela UFRN em 2016.

Figura 21 – Distribuição percentual do tipo de cursinho frequentado pelos candidatos ingressantes de acordo com os grupos de cursos de graduação ofertados pela UFRN, em 2016



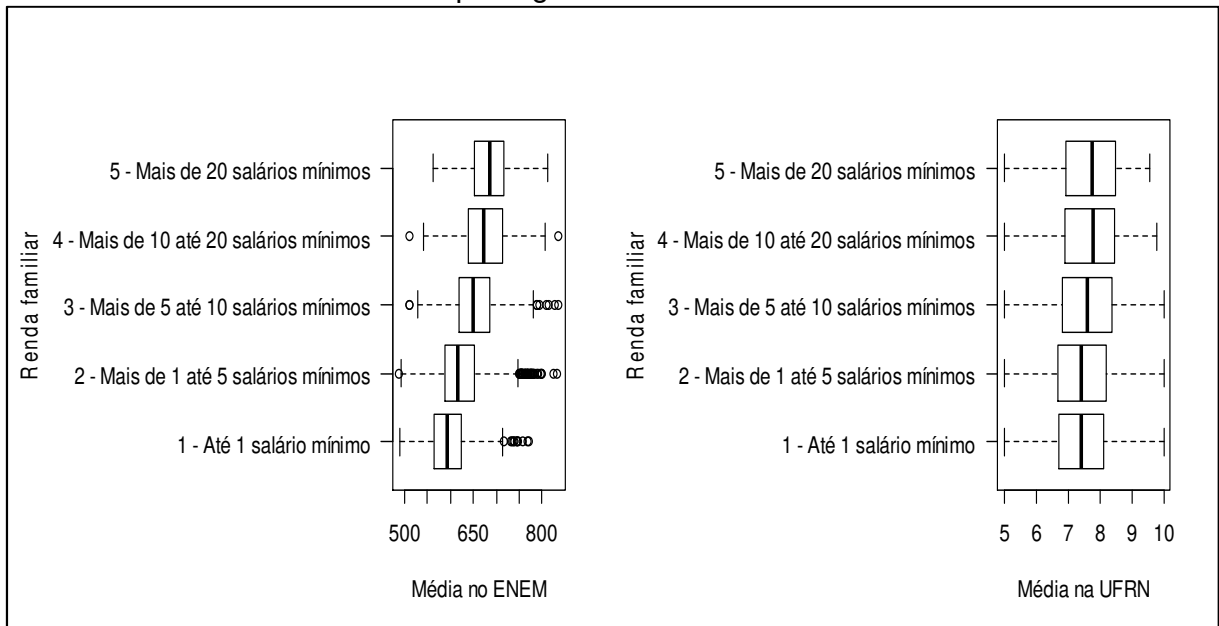
Fonte: COMPERVE

A análise da Figura 21 mostra que quanto ao cursinho de matérias isoladas o maior percentual está associado ao grupo 1, e este apresenta percentuais muito parecidos quanto as respostas “não fez cursinho” e “fez cursinho em rede particular”. As distribuições percentuais dos grupos 3 e do 4 (menores desempenhos) são muito parecidas.

4.2.10 Renda

A Figura 22 mostra grupos de *boxplot* associados ao desempenho geral no ENEM e na UFRN, de acordo com os dados da renda familiar dos ingressantes, via SISU, em 2016.

Figura 22 – Desempenho geral no ENEM e na UFRN segundo a renda familiar do candidato que ingressou na UFRN em 2016

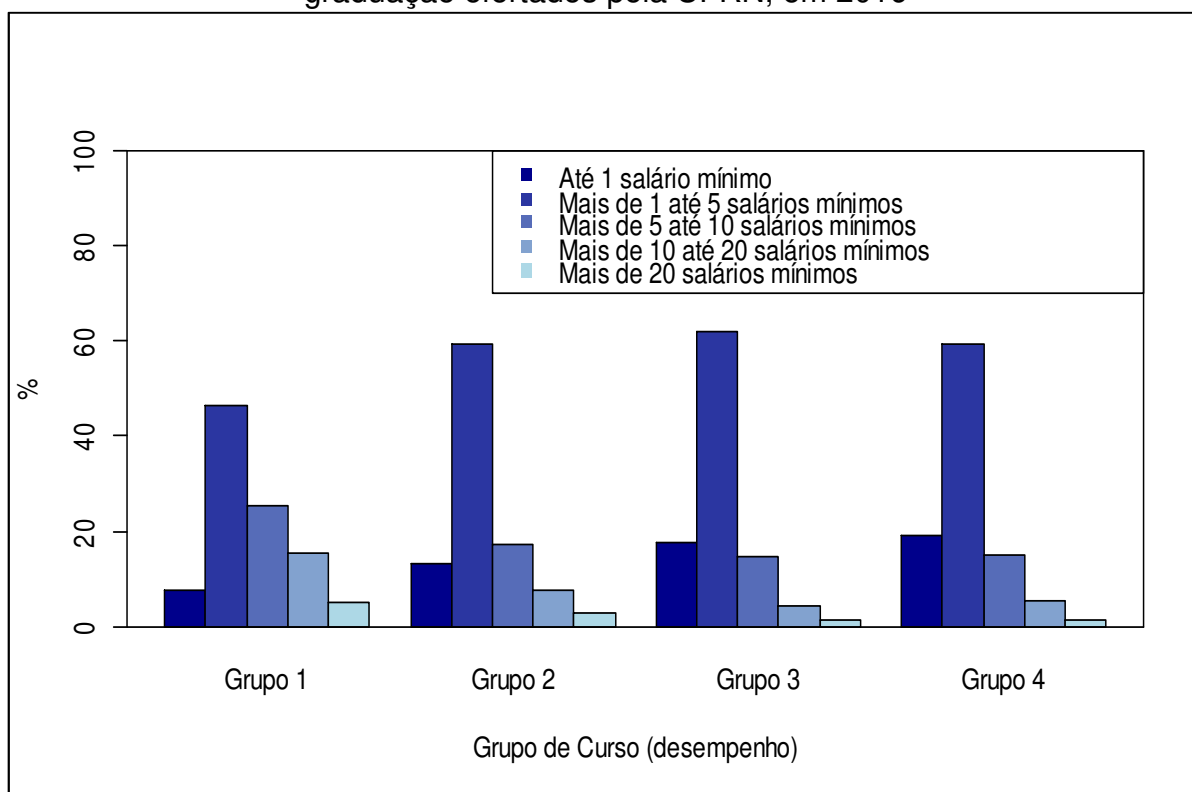


Fonte: COMPERVE

Analisando os desempenhos no ENEM no contexto da renda familiar do candidato, expostos nos gráficos do lado esquerdo da Figura 22, percebe-se nitidamente a influência da renda no desempenho dos candidatos, de maneira que, conforme a renda familiar aumenta, o desempenho do candidato também aumenta. Desse modo, a mediana do desempenho dos candidatos que possuem os maiores níveis de renda é maior, em torno de 100 pontos a mais, que a mediana dos desempenhos dos candidatos com menores níveis de renda. Porém quando se trata dos desempenhos na UFRN, essa relação não se mostra tão forte, ou seja, as medianas são de valores próximos entre si.

A Figura 23 apresenta a distribuição da renda familiar de acordo com os quatro grupos de cursos de graduação ofertados pela UFRN em 2016.

Figura 23 – Percentual da renda familiar segundo os grupos de cursos de graduação ofertados pela UFRN, em 2016



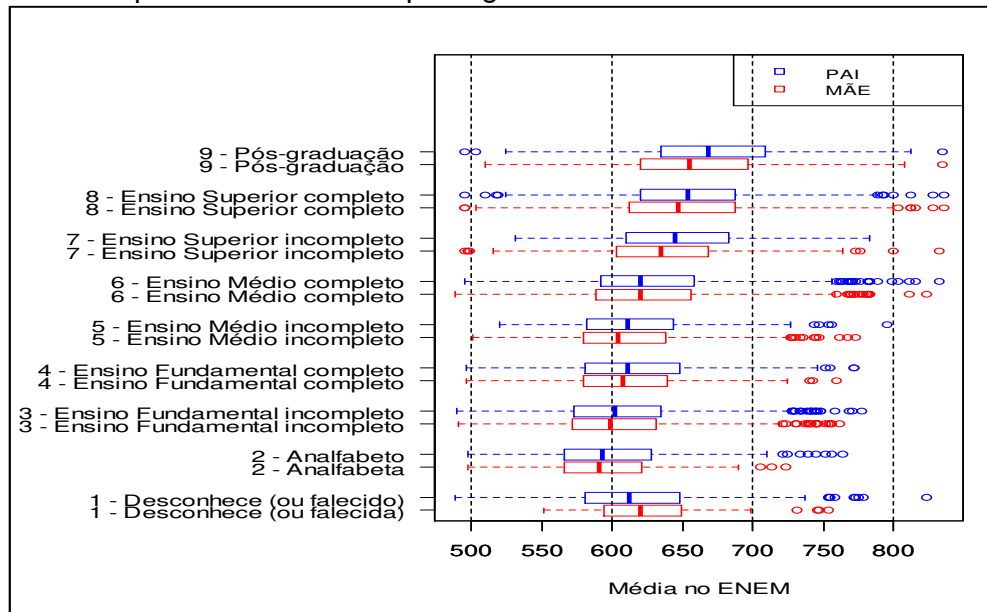
Fonte: COMPERVE

A análise da Figura 23 mostra que o percentual do nível de renda “mais de 1 até 5 salários mínimos” é o maior em todos os cursos, no entanto, quanto aos níveis de renda “mais de 5 até 10” e superiores os maiores percentuais estão associados ao grupo 1. Considerando a análise realizada da Figura 22 e o fato de que quanto maior o nível dos grupos de cursos de graduação pior o desempenho dos candidatos, este resultado é esperado.

4.2.11 Grau de Instrução dos Pais

A Figura 24 mostra os grupos de *boxplots* associados ao desempenho no ENEM, dos candidatos que ingressaram na UFRN em 2016. Este dispõe de dezoito gráficos cada um considerando desempenho do candidato segundo o grau de instrução dos pais.

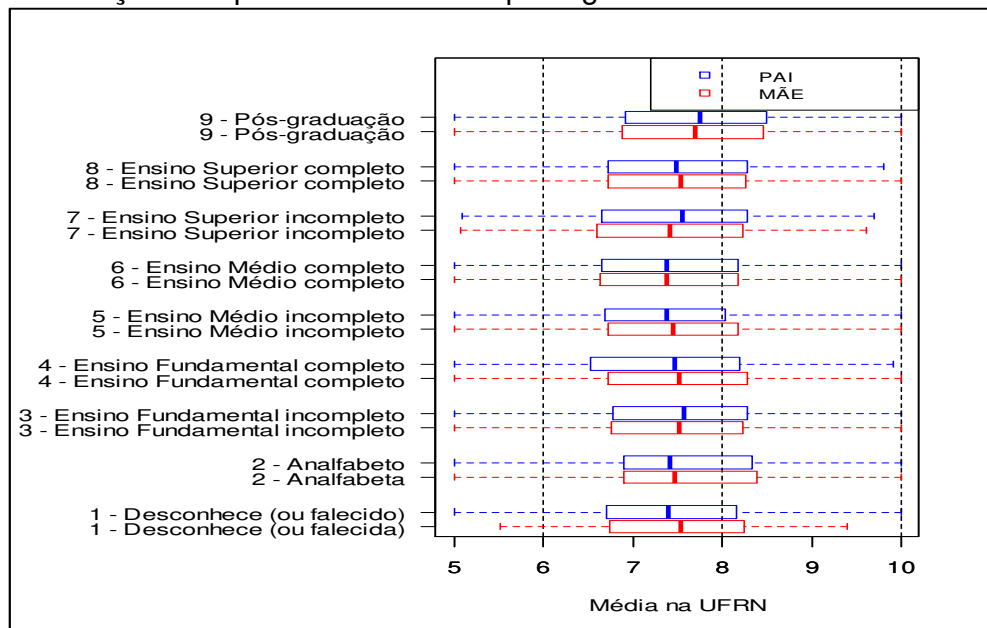
Figura 24 – Desempenho geral no ENEM de acordo com o grau de instrução dos pais do candidato que ingressou na UFRN em 2016



Fonte: COMPERVE

A Figura 25 mostra os grupos de boxplots associados ao desempenho na UFRN, dos candidatos que ingressaram, via SISU, em 2016. Este dispõe de dezoito gráficos cada um considerando desempenho do candidato segundo o grau de instrução dos pais.

Figura 25 – Desempenho geral na UFRN dos candidatos de acordo com o grau de instrução dos pais do candidato que ingressou na UFRN em 2016



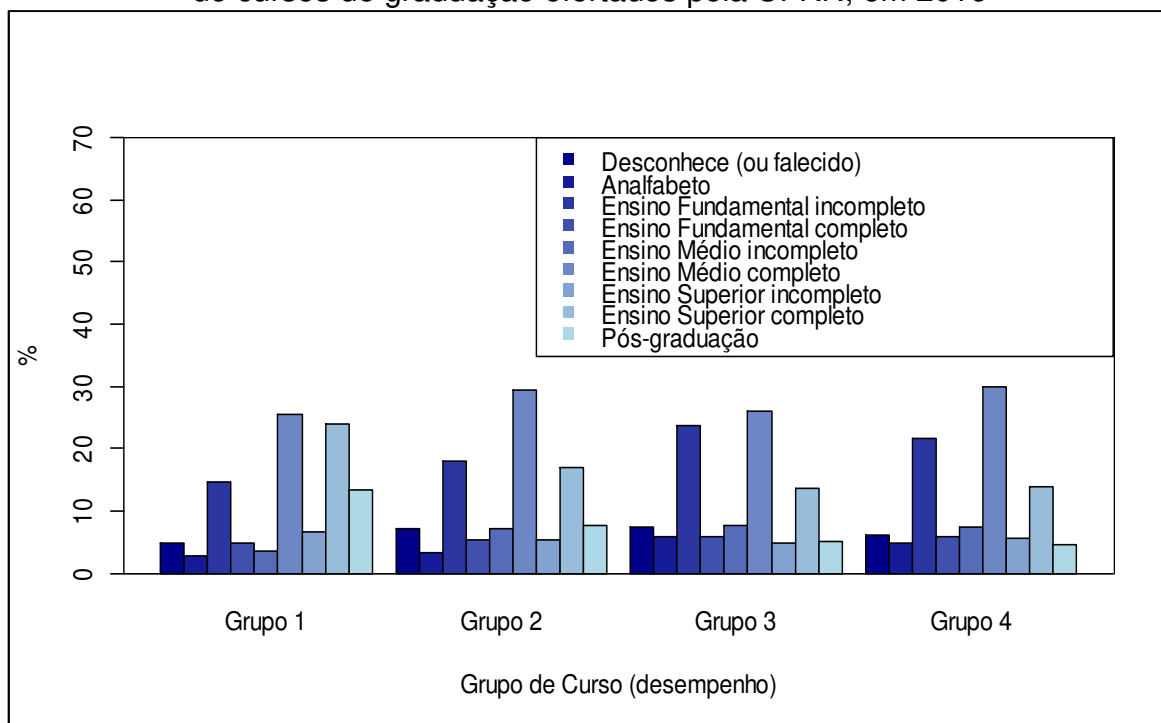
Fonte: COMPERVE

Analisando os desempenhos no ENEM em função da variável “grau de instrução dos pais do candidato”, expostos nos gráficos da Figura 24, é perceptível a influência dessa variável no desempenho dos candidatos, de tal maneira que, exceto nos casos associados às respostas “desconhece (ou falecido [a])”, à medida que a instrução dos pais aumenta, o desempenho do candidato também aumenta. É interessante comentar ainda que a mediana do desempenho dos candidatos ingressantes, em relação ao grau de instrução do pai é maior que as medianas destes desempenhos em relação ao grau de instrução da mãe.

No tocante ao desempenho na UFRN dos ingressantes, via SISU, em 2016, expostos na Figura 25, a variável “grau de instrução dos pais” parece não afetar de forma significativa as medianas dos desempenhos.

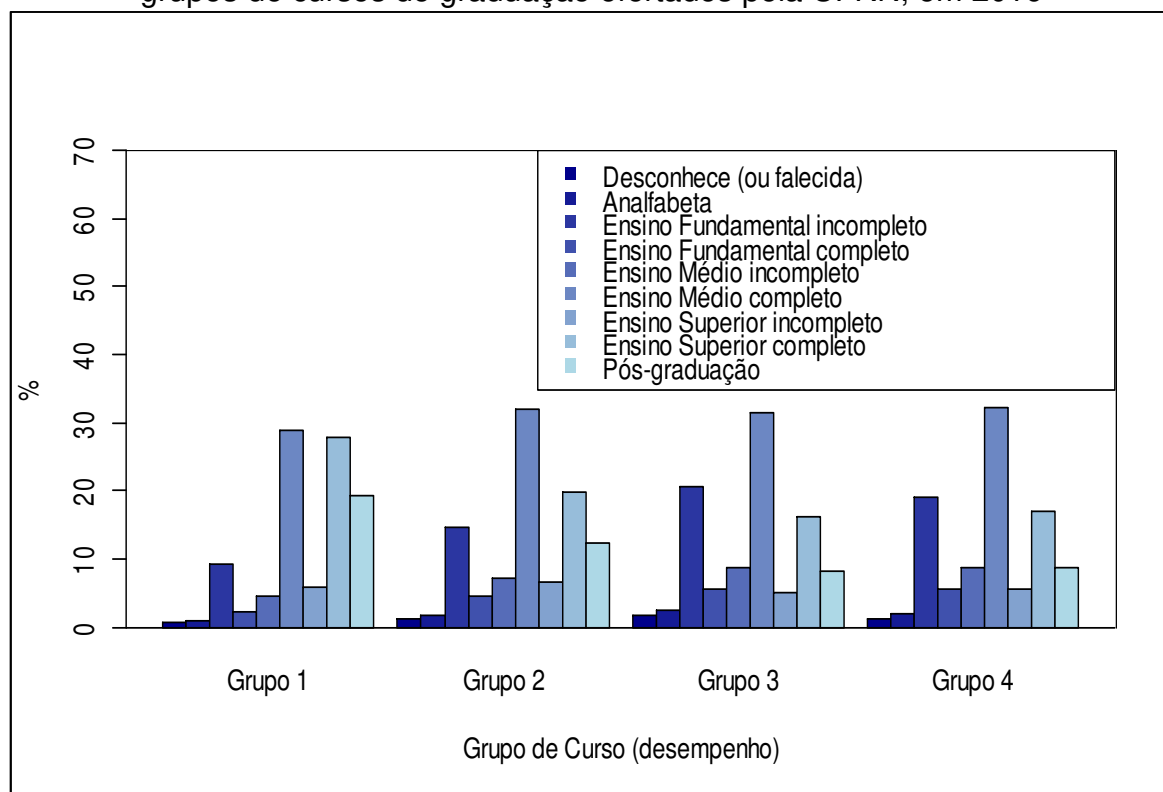
A Figura 26 apresenta a distribuição percentual do grau de instrução do pai considerando dos quatro grupos de cursos de graduação ofertados pela UFRN em 2016, e a Figura 27 apresentada na sequência, traz dados semelhantes, porém para o grau de instrução da mãe.

Figura 26 – Distribuição percentual do grau de instrução do pai segundo os grupos de cursos de graduação ofertados pela UFRN, em 2016



Fonte: COMPERVE

Figura 27– Distribuição percentual do grau de instrução da mãe de acordo com os grupos de cursos de graduação ofertados pela UFRN, em 2016



Fonte: COMPERVE

A análise da Figura 26 mostra que nos grupos 2 e 4 (médio e baixo desempenho respectivamente) o percentual dos pais dos ingressantes com Ensino Médio é semelhante e estes são os grupos onde este grau de instrução é maior. O percentual de pais dos ingressantes com Ensino Superior e o percentual com Pós-Graduação é mais alto quando associado ao grupo 1.

A Figura 27 mostra que a distribuição dos percentuais do grau de instrução das mães é parecida nos grupos 3 e 4 (menores desempenhos), e, o percentual de mães com Ensino Médio é maior no grupo 4. No tocante ao Ensino Superior e à Pós-Graduação estes apresentaram percentuais maiores quando associados ao grupo 1.

Analisando as Figura 26 e Figura 27 conjuntamente, é interessante comentar que o percentual do grau de instrução das mães é maior em relação ao grau de instrução dos pais, de modo que considerando os percentuais de mães desconhecidas ou falecidas e analfabetas, por exemplo, estes são baixos.

4.2.12 Status acadêmico

A Tabela 5 expõe a distribuição percentual do *status* acadêmico dos candidatos que fizeram ENEM 2015 e ingressaram na UFRN, via SISU, em 2016, por grupo de cursos de graduação.

Tabela 5 – Distribuição percentual do *status* acadêmico dos candidatos ingressantes por grupos de cursos de graduação ofertados pela UFRN em 2016

	Ativo	Cancelado	Concluído ¹	Trancado	Total
Grupo 1	88,53%	7,15%	0,00%	4,32%	100,00%
Grupo 2	80,44%	13,12%	0,08%	6,35%	100,00%
Grupo 3	74,99%	17,04%	0,06%	7,92%	100,00%
Grupo 4	70,24%	23,55%	0,03%	6,17%	100,00%

Fonte: COMPERVE

⁽¹⁾ Os percentuais de concluído não são zero por cento para todos os grupos de cursos de graduação porque os alunos podem tentar reingresso em outro curso ou no mesmo curso de origem e assim aproveitar as disciplinas já cursadas.

Considerando que os grupos de cursos de graduação estão ordenados de modo que quanto maior o nível do grupo pior o seu desempenho no ENEM 2015, a Tabela 5 possibilita entender a relação entre o *status* acadêmico dos alunos ingressantes e os grupos de cursos formados na análise de agrupamento. Assim, destaca-se que a porcentagem associada ao *status* de cancelamento de curso aumenta à medida que o nível do grupo aumenta. Este resultado pode ser interpretado como sendo um possível indicador de que os alunos ingressaram em cursos que não queriam, tendo em vista que os grupos 3 e 4 que são os que possuem maiores percentuais de cancelamento também são os que possuem pior desempenho no ENEM.

Como os cursos de graduação demoram em média 4 anos para serem concluídos, os ingressantes na UFRN, via SISU, em 2016 não tiveram o tempo necessário para tal, este fato justifica os resultados sobre o status de concluído.

5 CONSIDERAÇÕES FINAIS

A aplicação do método de *Ward* aos dados associados aos desempenhos dos candidatos ingressantes na UFRN, via SISU 2016, possibilitou dividir os cursos de graduação ofertados pela UFRN no referido ano, em quatro categorias de acordo com os níveis de desempenho no ENEM. Tais categorias são: o grupo 1, para os melhores desempenhos; o grupo 2 para desempenhos de nível mediano; o grupo 3 e o grupo 4 para desempenhos de nível baixo. A análise dos dados assim divididos possibilitou conhecer características que ocorreram nos grupos de menores desempenho, ou seja, nos grupos 3 e 4. Estes apresentaram medianas iguais no desempenho geral, porém quando analisados por área de conhecimento eles diferem, de tal forma que um grupo é oposto ao outro, quando as áreas são Matemática e Redação. Estas características, quando comparados com resultados de estudos semelhantes, utilizando os dados do vestibular, sugerem que o ENEM é um exame mais completo, capaz de captar melhor as diferenças.

É importante ainda falar que a utilização dos boxplots mostrou a relação existente entre às variáveis socioeconômicas e o desempenho médio dos candidatos ingressantes, de forma que, quando associados aos resultados obtidos por meio do gráfico de barras tornou possível caracterizar os grupos de cursos. Assim, o grupo 1 é composto pelos candidatos mais jovens e elitizados, visto que são os candidatos que mais utilizam carro próprio como meio de transporte, frequentaram cursinho particular ou cursinho isolado e estudaram em escolas particulares. Além disso, o grau de instrução dos pais é alto, o que sugere uma valorização da importância da educação e do ensino superior. O grupo 2 é semelhante ao grupo 1, porém em menores proporções. Já os grupos 3 e 4, estes são bem parecidos, possuindo os candidatos com idades mais elevadas, com as rendas mais baixas, tendo o coletivo como meio de transporte que mais utilizam, e para o grau de instrução dos pais, o percentual é alto nas categorias ensino fundamental incompleto e Ensino Médio completo. A distribuição da variável etnia se deu de forma desigual, de modo que 91% dos candidatos ingressantes são brancos ou pardos / mulatos. Outra variável que chama a atenção é a modalidade do ensino superior que mostra que os candidatos que cursaram o Ensino Médio em escolas técnicas possuem desempenho mediano maior, tanto no ENEM quanto na UFRN.

No que se referem aos desempenhos dos ingressantes na UFRN as variáveis socioeconômicas parecem não possuir tanta influência no desempenho mediano; talvez isto seja o resultado das várias políticas de permanência ofertadas pela universidade, tais como, auxílio moradia, creche, transporte, bolsas, projetos de apoio ao estudante, como o projeto de extensão Hábitos de Estudo, além dos atendimentos médicos e psicológicos gratuitos, dentre outros.

Como já citado anteriormente a ideia da meritocracia parte do princípio que existe igualdade na disputa pelo mérito, assim, se essa igualdade não existe, considerando os resultados obtidos neste trabalho que mostram a relação entre as variáveis socioeconômicas e os desempenhos dos candidatos ingressantes, essa disputa é injusta. Neste caso não há meritocracia, há uma parte da população que é privilegiada e outra que, para competir, precisa superar diversos fatores que não dependem diretamente dela, por isso a necessidade de políticas sociais, tais como a política de cotas raciais e sociais, bolsas de auxílio, dentre outras, que tentam reduzir a desigualdade e permitem uma disputa um pouco mais justa.

REFERÊNCIAS

Andrade DF, Tavares HR, Valle RC. **Teoria de Resposta ao Item: conceitos e aplicações**. São Paulo: Associação Brasileira de Estatística; 2000.

BRASIL. MEC – Ministério da Educação - ENEM. Disponível em: <<http://www.mec.gov.br/>>. Acesso em: 18 out. 2018.

COMPERVE – Núcleo Permanente de Concursos. Disponível em: <<http://www.comperve.ufrn.br/>> Acesso em: 10 set. 2018.

EVERITT, B.S.;DUNN, G. **Applied multivariate analysis**. London: Edward Arnold, 1991, 400p.

FREIRE, Flávio Henrique de Araújo. **O perfil do vestibulando da UFRN**. Natal: 1994. 77 p. Disponível em: <http://www.comperve.ufrn.br/conteudo/observatorio/uploads/publicacoes/monografias_08052013032537.pdf>. Acesso em: 23 out. 2018.

FREIRE, Paulo. **Educação como prática de liberdade**. Rio de Janeiro: Paz e Terra, 2010.

GOMES, William B.; SPARTA, Mônica. Importância Atribuída ao Ingresso na Educação Superior por Alunos do Ensino Médio. **Revista Brasileira de Orientação Profissional**, Universidade Federal do Rio Grande do Sul, Porto Alegre, p.45-53, 10 mar. 2005. Semestral. Disponível em: <<http://www.ufrgs.br/museupsi/lafec/16.pdf>>. Acesso em: 27 nov. 2018.

GUIA DO ESTUDANTE. **Tudo sobre o novo ENEM com 180 questões aplicadas em dois dias, a nova prova ganha mais conteúdo para substituir o vestibular em boa parte das universidades federais do Brasil**. Annual 2009, p.8. AcademicOneFile. Disponível em:<<http://link.galegroup.com/apps/doc/A206464727/AONE?u=capes&sid=AONE&xid=8b55b29d/>>. Acesso em: 18 out. 2018.

HAIR JUNIOR, Joseph F. et al. **Análise Multivariada de Dados**. 6. ed. Porto Alegre: Bookman, 2009. 688 p. Tradução de Adonai Schlup Sant' Anna.

INEP – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Disponível em: <<http://portal.inep.gov.br/web/guest/microdados>>. Acesso em: 10 set. 2018.

MINGOTI, Sueli Aparecida. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**, Ed 2013, 2 Reimpressão - UFMG.

NEVES, Clarissa Eckert Baeta. Desafios da educação superior. In: . **Sociologias**: Dossiê. 17. ed. Porto Alegre, 2007. p. 14-21. (Jan./jun.). Disponível em: <<http://www.scielo.br/pdf/soc/n17/a02n17.pdf>>. Acesso em: 19 nov. 2018.

R The R Project for statistical computing: Software. Version 3.5.1. Nova Zelândia:CRAN, 1993. Disponível em: <<https://www.r-project.org/>>.

SANDES, Hyran Ferreira. **O papel da educação na formação do cidadão brasileiro**. 2012. Disponível em: <<http://www.conteudojuridico.com.br/?artigos&ver=2.39463&seo=1>>. Acesso em: 19 nov. 2018.

TRIOLA, Mario F.. **Introdução à Estatística**. Tradução Vera Regina Lima de Farias e Flores, revisão técnica Ana Maria Lima de Farias, 10. ed. Rio de Janeiro: Ltc – Livros Técnicos e Científicos Editora S.A., 2008. 696 p.

UFRN – Universidade Federal do Rio Grande do Norte. Disponível em: <<https://www.ufrn.br/institucional/sobre-a-ufrn>>. Acesso em: 6 nov. 2018.

VIVACQUA, Carla Almeida. **Apostila análise agrupamentos**, UFRN-2018.

APÊNDICE

```

OVEU2016=read.table("D:/OVEU2016 - monografia Alane.csv", head=T, sep=";",
dec=",")
str(OVEU2016)
dim(OVEU2016)
attach(OVEU2016)
names(OVEU2016)
#### Análise de cluster (desempenho por curso) =====
# Prepara as variáveis para o agrupamento
CN.cur=split(NU_NOTA_CN,Nome_curso) # separa as notas por curso
CH.cur=split(NU_NOTA_CH,Nome_curso) # separa as notas por curso
LC.cur=split(NU_NOTA_LC,Nome_curso) # separa as notas por curso
MT.cur=split(NU_NOTA_MT,Nome_curso) # separa as notas por curso
Red.cur=split(NU_NOTA_REDACAO,Nome_curso) # separa as notas por curso
percCN=sapply(CN.cur,quantile,prob=c(.1,.25,.5,.75,.9))#pq percentil
percCH=sapply(CH.cur,quantile,prob=c(.1,.25,.5,.75,.9))#pq percentil
percLC=sapply(LC.cur,quantile,prob=c(.1,.25,.5,.75,.9))#pq percentil
percMT=sapply(MT.cur,quantile,prob=c(.1,.25,.5,.75,.9))#pq percentil
percRed=sapply(Red.cur,quantile,prob=c(.1,.25,.5,.75,.9))#pq percentil
perc=cbind(t(percCN),t(percCH),t(percLC),t(percMT),t(percRed))
      # Método de Ward #
wd<-hclust(dist(perc), method="ward.D")
# Dendrograma
dendwd <- as.dendrogram(wd)
x11(18,8)
par(mai= c(4, 0.82, 0.82, 0.42), cex=.7)
plot(dendwd, ylab="Nível de junção")
mtext("Cursos", side=1, line=27)
rect.hclust(wd, k=4, border="red")
# Corte dos grupos
g1=cutree(wd, k=4)
for (i in 1:length(g1)) OVEU2016$grupo[Nome_curso==names(g1)[i]]=g1[i]

```

```

table(OVEU2016$grupo)
# Calcula média entre as notas do ENEM
OVEU2016$mediaENEM=apply(cbind(NU_NOTA_CN,NU_NOTA_CH,NU_NOTA_L
C, NU_NOTA_MT, NU_NOTA_REDACAO),1,mean)
detach(OVEU2016)
attach(OVEU2016)
# Reordena grupos:
ord=order(tapply(mediaENEM,grupo,mean),decreasing=T)
g1.ord= vector()
for (i in 1:length(g1)) g1.ord[g1==ord[i]]=i
names(g1.ord)=names(g1)
## Cria a variável "grupo" (ordenada) no data frame:
for (i in 1:length(g1.ord))
OVEU2016$grupo[Nome_curso==names(g1.ord)[i]]=g1.ord[i]
table(OVEU2016$Nome_curso)
table(OVEU2016$grupo)
g1.ord[g1.ord==1] # Cursos do grupo 1 - desempenho alto
g1.ord[g1.ord==2] # Cursos do grupo 2 - desempenho médio
g1.ord[g1.ord==3] # Cursos do grupo 3 - desempenho baixo
g1.ord[g1.ord==4] # Cursos do grupo 4 - desempenho baixo
n1=length(g1.ord[g1.ord==1]) # Cursos do grupo 1 - desempenho alto
n2=length(g1.ord[g1.ord==2]) # Cursos do grupo 2 - desempenho médio
n3=length(g1.ord[g1.ord==3]) # Cursos do grupo 3 - desempenho baixo
n4=length(g1.ord[g1.ord==4]) # Cursos do grupo 4 - desempenho baixo
detach(OVEU2016)
attach(OVEU2016)
##### Fim Análise de cluster =====
#####Análise exploratória de desempenho via boxplots =====
summary(OVEU2016$mc)
# Diagrama de dispersão das medianas entre Matemática e Redação
x11()

```

```
plot(perc[,c(18,23)], pch=as.character(g1.ord), col= g1.ord, xlab="Mediana do
desempenho em matemática", ylab="Mediana do desempenho em redação")
```

```
legend("topleft", col=c(1,2,3,4), pch=c("1", "2","3","4"), cex=0.8, legend=c("
Desempenho Alto", "Desempenho Mediano", "Desempenho Baixo", "Desempenho
Baixo"))
```

```
# Diagrama de dispersão das medianas entre Ciências da Natureza e Linguagens
```

```
x11()
```

```
plot(perc[,c(3,13)], pch=as.character(g1.ord), col= g1.ord, xlab="Mediana do
desempenho em ciências da natureza", ylab="Mediana do desempenho em
linguagens")
```

```
legend("topleft", col=c(1,2,3,4), pch=c("1", "2","3","4"), cex=0.8, legend=c("
Desempenho Alto", "Desempenho Mediano", "Desempenho Baixo", "Desempenho
Baixo"))
```

```
# Diagrama de dispersão das medianas entre Ciências da Natureza e Matemática
```

```
x11()
```

```
plot(perc[,c(3,18)], pch=as.character(g1.ord), col= g1.ord, xlab="Mediana do
desempenho em ciências da natureza", ylab="Mediana do desempenho em
matemática")
```

```
legend("topleft", col=c(1,2,3,4), pch=c("1", "2","3","4"), cex=0.8, legend=c("
Desempenho Alto", "Desempenho Mediano", "Desempenho Baixo", "Desempenho
Baixo"))
```

```
# Diagrama de dispersão das medianas entre Linguagens e Redação
```

```
x11()
```

```
plot(perc[,c(13,23)], pch=as.character(g1.ord), col= g1.ord, xlab="Mediana do
desempenho em linguagens", ylab="Mediana do desempenho em redação")
```

```
legend("topleft", col=c(1,2,3,4), pch=c("1", "2","3","4"), cex=0.8, legend=c("
Desempenho Alto", "Desempenho Mediano", "Desempenho Baixo", "Desempenho
Baixo"))
```

```
# Desempenho por área de conhecimento e por grupo de curso
```

```
x11(18,15)
```

```
par(mfrow=c(2,3))
```

```
boxplot(OVEU2016$NU_NOTA_CN~OVEU2016$grupo, main="A",ylab="Nota em
Ciências da Natureza", xlab="Grupo de curso")
```

```
boxplot(OVEU2016$NU_NOTA_CH~OVEU2016$grupo,main="B", ylab="Nota em
Ciências Humanas", xlab="Grupo de curso")
```

```
boxplot(OVEU2016$NU_NOTA_LC~OVEU2016$grupo, main="C", ylab="Nota em
Linguagens", xlab="Grupo de curso")
```



```

boxplot(OVEU2016$NU_NOTA_MT~OVEU2016$grupo, main="D", ylab="Nota em
Matemática", xlab="Grupo de curso")

boxplot(OVEU2016$NU_NOTA_REDACAO~OVEU2016$grupo,          main="E",
ylab="Nota em Redação", xlab="Grupo de curso")

boxplot(OVEU2016$mediaENEM~OVEU2016$grupo, main="F", ylab="Nota média
no ENEM", xlab="Grupo de curso")

# Desempenho no ENEM segundo o perfil do ingressante:=====
# Desempenho por área de conhecimento e por gênero

x11(18,15)

par(mfrow=c(2,3))

boxplot(OVEU2016$NU_NOTA_CN~OVEU2016$Q1..Gênero,main="A", ylab="Nota
em Ciências da Natureza", xlab="Gênero")

boxplot(OVEU2016$NU_NOTA_CH~OVEU2016$Q1..Gênero, main="B", ylab="Nota
em Ciências Humanas", xlab="Gênero")

boxplot(OVEU2016$NU_NOTA_LC~OVEU2016$Q1..Gênero,main="C", ylab="Nota
em Linguagens", xlab="Gênero")

boxplot(OVEU2016$NU_NOTA_MT~OVEU2016$Q1..Gênero, main="D", ylab="Nota
em Matemática", xlab="Gênero")

boxplot(OVEU2016$NU_NOTA_REDACAO~OVEU2016$Q1..Gênero,          main="E",
ylab="Nota em Redação", xlab="Gênero")

boxplot(OVEU2016$mediaENEM~OVEU2016$Q1..Gênero,  main="F", ylab="Nota
média no ENEM", xlab="Gênero")

mc_sem_zero=OVEU2016$mc

mc_sem_zero[mc_sem_zero==0]=NA

# Desempenho geral no ENEM e na UFRN por sexo

x11(8,4)

par(mfrow=c(1,2))

boxplot(OVEU2016$mediaENEM~OVEU2016$Q1..Gênero, ylab="Nota média no
ENEM", xlab="Gênero")

boxplot(mc_sem_zero~OVEU2016$Q1..Gênero, ylab="Nota média na UFRN",
xlab="Gênero")

# Grau de instrução do pai e da mãe ENEM

nomes=levels(OVEU2016$Q27..Grau.de.instrução.do.pai)

n=length(nomes)

ord_boxplot = order(nomes, decreasing=F)

```

```

boxp1=boxplot(OVEU2016$mediaENEM~OVEU2016$Q27..Grau.de.instrução.do.pai
, plot=F)
boxp1$stats=boxp1$stats[,ord_boxplot]
boxp1$n=boxp1$n[ord_boxplot]
boxp1$conf=boxp1$conf[,ord_boxplot]
boxp1$names=boxp1$names[ord_boxplot]
group=numeric(length(boxp1$group))
o=0
for (i in ord_boxplot) {o=o+1; group[boxp1$group==i]=o}
boxp1$group=group
x11(15,20)
par(mai=c(1,2.8,1,.5))
bxp(boxp1, horizontal=T, las=1, xlim=c(1,11), xlab="Média no ENEM",
at=seq(1.15,n+.15,1), pars = list(boxwex = 0.25, staplewex = 0.5, outwex = 0.5),
border = 4)
abline(v=c(400,500,600,700,800,900), lty=2)
# da mãe
boxp2=boxplot(OVEU2016$mediaENEM~OVEU2016$Q29..Grau.de.instrução.da.mã
e, plot=F)
boxp2$stats=boxp2$stats[,ord_boxplot]
boxp2$n=boxp2$n[ord_boxplot]
boxp2$conf=boxp2$conf[,ord_boxplot]
boxp2$names=boxp2$names[ord_boxplot]
group=numeric(length(boxp2$group))
o=0
for (i in ord_boxplot) {o=o+1; group[boxp2$group==i]=o}
boxp2$group=group
bxp(boxp2, horizontal=T, las=1,xlim=c(1,11), add=T, at=seq(0.85,n-.15,1), pars =
list(boxwex = 0.25, staplewex = 0.5, outwex = 0.5), border = 2)
legend("topright", col=c(4,2), pch=0, cex=0.8, legend=c("PAI","MÃE"))
# Grau de instrução do pai e da mãe UFRN
nomes2=levels(OVEU2016$Q27..Grau.de.instrução.do.pai)
n=length(nomes2)
ord_boxplot = order(nomes2, decreasing=F)

```

```

boxp3=boxplot(mc_sem_zero~OVEU2016$Q27..Grau.de.instrução.do.pai, plot=F)
boxp3$stats=boxp3$stats[,ord_boxplot]
boxp3$n=boxp3$n[ord_boxplot]
boxp3$conf=boxp3$conf[,ord_boxplot]
boxp3$names2=boxp3$names2[ord_boxplot]
group=numeric(length(boxp3$group))
o=0
for (i in ord_boxplot) {o=o+1; group[boxp3$group==i]=o}
boxp3$group=group
x11(15,20)
par(mai=c(1,2.8,1,.5))
bxp(boxp3, horizontal=T, las=1, xlim=c(1,10), xlab="Média na UFRN",
at=seq(1.15,n+.15,1), pars = list(boxwex = 0.25, staplewex = 0.5, outwex = 0.5),
border = 4)
abline(v=c(2,4,6,8,10), lty=2)
# da mãe
boxp4=boxplot(mc_sem_zero~OVEU2016$Q29..Grau.de.instrução.da.mãe, plot=F)
boxp4$stats=boxp4$stats[,ord_boxplot]
boxp4$n=boxp4$n[ord_boxplot]
boxp4$conf=boxp4$conf[,ord_boxplot]
boxp4$names2=boxp4$names2[ord_boxplot]
group=numeric(length(boxp4$group))
o=0
for (i in ord_boxplot) {o=o+1; group[boxp4$group==i]=o}
boxp4$group=group
bxp(boxp4, horizontal=T, las=1, add=T, at=seq(0.85,n-.15,1), pars = list(boxwex =
0.25, staplewex = 0.5, outwex = 0.5), border = 2)
legend("topright", col=c(4,2), pch=0, cex=0.8, legend=c("PAI","MÃE"))
# Renda mensal da família no ENEM
x11(15,7)
par(mfrow=c(1,2))
par(mai= c(1.02, 3.4, 0.82, 0.42))

```

```

boxplot(OVEU2016$mediaENEM~OVEU2016$Renda, las=1,xlab="Média no
ENEM", horizontal=T, las=2)

  mtext("Renda familiar", side=2, line=15.5)

# Renda mensal da família na UFRN

par(mai= c(1.02, 3.4, 0.82, 0.42))

boxplot(mc_sem_zero~OVEU2016$Renda, las=1,xlab="Média na UFRN",
horizontal=T, las=2)

mtext("Renda familiar", side=2, line=15.5)

attach(OVEU2016)

  # ETNIA no ENEM

x11(15,7)

par(mfrow=c(1,2))

par(mai= c(1.02, 2.6, 0.4, 0.42))

boxplot(OVEU2016$mediaENEM~OVEU2016$Q4..Como.você.se.considera.,
las=1,xlab="Média no ENEM", horizontal=T, las=2)

mtext("Etnia", side=2, line=12)

  # ETNIA na UFRN #

par(mai= c(1.02, 2.6, 0.4, 0.42))

boxplot(mc_sem_zero~OVEU2016$Q4..Como.você.se.considera.,
las=1,xlab="Média na UFRN", horizontal=T, las=2)

mtext("Etnia", side=2, line=12)

  # Idade no ENEM

table(OVEU2016$Q3..Idade)

summary(OVEU2016$Q3..Idade) #Percentis 25,50 e 75

quantile(OVEU2016$Q3..Idade, c(.1,.9), type=4) #Percentis 10 e 90

OVEU2016$Idade[OVEU2016$Q3..Idade<17]<-" < 17"

OVEU2016$Idade[17<=OVEU2016$Q3..Idade&OVEU2016$Q3..Idade<18]<-"17 -18"

OVEU2016$Idade[18<=OVEU2016$Q3..Idade&OVEU2016$Q3..Idade<19]<-"18 -19"

OVEU2016$Idade[19<=OVEU2016$Q3..Idade&OVEU2016$Q3..Idade<24]<-"19 -24"

OVEU2016$Idade[24<=OVEU2016$Q3..Idade&OVEU2016$Q3..Idade<31]<-"24 -31"

OVEU2016$Idade[OVEU2016$Q3..Idade>=31]<-"31 +"

OVEU2016

OVEU2016$Idade=as.factor(OVEU2016$Idade)

```

```

# Idade no ENEM
x11(8,4)
par(mfrow=c(1,2))
par(mai= c(1.0, 1.0, 0.5, 0.42))
boxplot(OVEU2016$mediaENEM ~ OVEU2016$Idade, las=1, xlab=" Média no
ENEM", horizontal=T, las=2)
mtext("Grupo Etário ", side=2, line=4)

# Idade na UFRN
par(mai= c(1.0, 1.0, 0.5, 0.42))
boxplot(mc_sem_zero~OVEU2016$Idade, las=1,xlab=" Média na UFRN ",
horizontal=T, las=2)
mtext("Grupo Etário ", side=2, line=4)

# Estado Civil no ENEM
table(OVEU2016$Q2..Estado.civil)
x11(15,7)
par(mfrow=c(1,2))
par(mai= c(1.02, 0.8, 0.4, 0.2))
boxplot(OVEU2016$mediaENEM~OVEU2016$Q2..Estado.civil, las=1,xlab="Estado
Civil", horizontal=F, las=2)
mtext("Média no ENEM", side=2, line=3.0)

#Estado Civil na UFRN
par(mai= c(1.02, 0.8, 0.4, 0.2))
boxplot(mc_sem_zero~OVEU2016$Q2..Estado.civil, las=1,xlab="Estado Civil",
horizontal=F, las=2)
mtext("Média na UFRN", side=2, line=2.8)

# Tipo de Escola que cursou o E.M. - ENEM
x11(15,5)
par(mfrow=c(1,2))
par(mai= c(1.02, 4.5, 0.4, 0.42))
boxplot(OVEU2016$mediaENEM~OVEU2016$Q16..Tipo.de.escola.onde.cursou.o.E
nsino.Médio..ou.equivalente, las=1,xlab="Média no ENEM", horizontal=T, las=2)
mtext("Tipo de Escola que cursou o Ensino Médio", side=2, line=20.7)

# Tipo de Escola que cursou o E.M. - UFRN
par(mai= c(1.02, 4.5, 0.4, 0.42))

```

```

boxplot(mc_sem_zero~OVEU2016$Q16..Tipo.de.escola.onde.cursou.o.Ensino.Médi
o..ou.equivalente, las=1,xlab="Média no ENEM", horizontal=T, las=2)
mtext("Tipo de Escola que cursou o Ensino Médio", side=2, line=20.7)
#Modalidade do Ensino Médio
# ENEM
x11(15,7)
par(mfrow=c(1,2))
par(mai= c(1.02, 2.5, 0.82, 0.42))
boxplot(OVEU2016$mediaENEM~OVEU2016$Q20..Modalidade.deCurso.do.Ensino
.Médio, las=1,xlab="Média no ENEM", horizontal=T, las=2)
mtext("Modalidade do E.M.", side=2, line=11)
# UFRN
par(mai= c(1.02, 2.5, 0.82, 0.42))
boxplot(mc_sem_zero~OVEU2016$Q20..Modalidade.deCurso.do.Ensino.Médio,
las=1,xlab="Média na UFRN", horizontal=T, las=2)
mtext("Modalidade do E.M.", side=2, line=11)
#Frequentou cursinho
#ENEM
x11(15,7)
par(mfrow=c(1,2))
par(mai= c(1.02,3.6, 0.82, 0.42))
boxplot(OVEU2016$mediaENEM~OVEU2016$Q22..Frequentou.cursinho.para.prest
ar.ENEM, las=1,xlab="Média no ENEM", horizontal=T, las=2)
mtext("Se frequentou cursinho", side=2, line=17)
#UFRN
par(mai= c(1.02,3.6, 0.82, 0.42))
boxplot(mc_sem_zero~OVEU2016$Q22..Frequentou.cursinho.para.prestar.ENEM,
las=1,xlab="Média na UFRN", horizontal=T, las=2)
mtext("Se frequentou cursinho", side=2, line=17)
#Meio de transporte que mais utiliza
#ENEM
x11(15,7)
par(mfrow=c(1,2))
par(mai= c(1.02, 2.9, 0.82, 0.42))

```

```

boxplot(OVEU2016$mediaENEM~OVEU2016$Q13..Meio.de.transporte.que.mais.utiliza, las=1,xlab="Média no ENEM", horizontal=T, las=2)
mtext("Meio de transporte que mais utiliza", side=2, line=12.7)
#UFRN
par(mai= c(1.02, 2.9, 0.82, 0.42))
boxplot(mc_sem_zero~OVEU2016$Q13..Meio.de.transporte.que.mais.utiliza, las=1,xlab="Média na UFRN", horizontal=T, las=2)
mtext("Meio de transporte que mais utiliza", side=2, line=12.7)
# Boxplot para média no ENEM por curso (colorir por grupo de curso):
attach(OVEU2016)
x11(18,8)
par(mai= c(4, 0.82, 0.82, 0.42))
media.cur=split(OVEU2016$mediaENEM, OVEU2016$Nome_curso) # separa as médias por curso
mediana.media_ENEM=sapply(media.cur, median, na.rm=T)
ord.gru.mediana=order(5-g1.ord, mediana.media_ENEM, decreasing = T)
boxplot(media.cur[ord.gru.mediana], ylab="Média no ENEM", las=2, cex.axis=.7, col=c(rep(colors())[634],n1),rep(colors())[630],n2),rep(colors())[425],n3),rep("yellow",n4)))
mtext("Curso", side=1, line=18)
abline(h=seq(500,800,50), lty=2)
##### Análise exploratória de desempenho via gráfico de barras =====
# Gráfico de barras: Grupo de curso (desempenho) versus Gênero
x11()
barplot(100*prop.table(table(OVEU2016$Q1..Gênero,OVEU2016$grupo),2), col=c(2,4), ylim=c(0,120), names=c("Grupo 1","Grupo 2","Grupo 3","Grupo 4"), xlab="Grupo de Curso (desempenho)", ylab="Gênero (%)")
legend("topright", col=c(2,4), pch=15, cex=0.8, legend=c("Feminino","Masculino"))
abline(h=c(0,120), v=(4.98))
text(locator(n=8), c("47,1%", "52,9%", "54,5%", "45,5%", "56,2%", "43,8%", "28,6%", "71,4%"))
color=colorRampPalette(c("darkblue","lightblue"))
# Gráfico de barras: Grupo de curso (desempenho) versus Idade
x11()

```

```
barplot(100*prop.table(table(OVEU2016$Idade,OVEU2016$grupo),2),
col=color(6),beside=T, ylim=c(0,80), names=c("Grupo 1","Grupo 2","Grupo 3","Grupo
4"), xlab="Grupo de Curso (desempenho)", ylab="%")
```

```
legend("topright", col=color(6), ncol=2, pch=15, cex=1.1, legend=c("< 17","17 |-
18","18 |- 19", "19 |- 24", "24 |- 31", "31 +"))
```

```
abline(h=c(0,80), v=(29.08))
```

Gráfico de barras: Grupo de curso (desempenho) versus Etnia

```
x11()
```

```
barplot(100*prop.table(table(OVEU2016$Q4..Como.você.se.considera.,OVEU2016$
grupo),2), col=color(6),beside=T, ylim=c(0,70), names=c("Grupo 1","Grupo 2","Grupo
3","Grupo 4"), xlab="Grupo de Curso (desempenho)", ylab="%")
```

```
legend("topright", col=color(6), pch=15, cex=0.8, legend=c("Amarelo(a) (de origem
oriental)","Branco(a)","Indígena ou de origem", "Negro(a)", "Pardo(a) / Mulato(a)",
"Remanescente de quilombo"))
```

```
abline(h=c(0,70), v=(29.08))
```

```
abline(h=10, lty = 2)
```

```
table(OVEU2016$Q4..Como.você.se.considera.) # Total absoluto por etnia
```

```
round(100*table(OVEU2016$Q4..Como.você.se.considera.)/6279,2) #Percentual por
etnia
```

Gráfico de barras: Grupo de curso (desempenho) versus Estado Civil

```
x11()
```

```
barplot(100*prop.table(table(OVEU2016$Q2..Estado.civil,OVEU2016$grupo),2),
col=color(3),beside=T, ylim=c(0,120), names=c("Grupo 1","Grupo 2","Grupo
3","Grupo 4"), xlab="Grupo de Curso (desempenho)", ylab="%")
```

```
legend("topright", col=color(3), pch=15, cex=0.8, legend=c("Solteiro(a)", "Casado(a)",
"Outro"))
```

```
abline(h=c(0,120), v=(16.6))
```

Gráfico de barras: Grupo de curso (desempenho) versus Renda

```
x11()
```

```
barplot(100*prop.table(table(OVEU2016$Renda,OVEU2016$grupo),2), col=color(5),
beside=T, ylim=c(0,100), names=c("Grupo 1","Grupo 2","Grupo 3","Grupo 4"),
xlab="Grupo de Curso (desempenho)", ylab="%")
```

```
legend("topright", col=color(5), pch=15, cex=0.8, legend=c("Até 1 salário
mínimo","Mais de 1 até 5 salários mínimos","Mais de 5 até 10 salários
mínimos", "Mais de 10 até 20 salários mínimos", "Mais de 20 salários mínimos"))
```

```
abline(h=c(0,100), v=(24.9))
```

##Gráfico de barras: Grupo de curso (desempenho) versus Grau de Instrução do Pai

x11()

```
barplot(100*prop.table(table(OVEU2016$Q27..Grau.de.instrução.do.pai,OVEU2016$
grupo),2), col=color(9),beside=T, ylim=c(0,70), names=c("Grupo 1","Grupo 2","Grupo
3","Grupo 4"), xlab="Grupo de Curso (desempenho)", ylab="%")
```

```
legend("topright", col=color(9), pch=15, cex=0.8, legend=c("Desconhece (ou
falecido)","Analfabeto","Ensino Fundamental incompleto", "Ensino Fundamental
completo","Ensino Médio incompleto","Ensino Médio completo", "Ensino Superior
incompleto","Ensino Superior completo","Pós-graduação"))
```

```
abline(h=c(0,70), v=(41.58))
```

#Gráfico de barras: Grupo de curso (desempenho) versus Grau de Instrução da Mãe

x11()

```
barplot(100*prop.table(table(OVEU2016$Q29..Grau.de.instrução.da.mãe,OVEU2016
$grupo),2), col=color(9),beside=T, ylim=c(0,70), names=c("Grupo 1","Grupo
2","Grupo 3","Grupo 4"), xlab="Grupo de Curso (desempenho)", ylab="%")
```

```
legend("topright", col=color(9), ncol=2, pch=15, cex=0.8, legend=c("Desconhece (ou
falecida)","Analfabeta","Ensino Fundamental incompleto", "Ensino Fundamental
completo","Ensino Médio incompleto","Ensino Médio completo", "Ensino Superior
incompleto","Ensino Superior completo","Pós-graduação"))
```

```
abline(h=c(0,70), v=(41.58))
```

Gráfico de barras: Grupo de curso (desempenho) versus Tipo de Escola onde cursou o E.M.

x11()

```
barplot(100*prop.table(table(OVEU2016$Q16..Tipo.de.escola.onde.cursou.o.Ensino.
Médio..ou.equivalente,OVEU2016$grupo),2), col=color(4),beside=T, ylim=c(0,100),
names=c("Grupo 1","Grupo 2","Grupo 3","Grupo 4"), xlab="Grupo de Curso
(desempenho)", ylab="%")
```

```
legend("topright", col=color(4), ncol=2,pch=15, cex=0.8, legend=c("Sem resposta /
inconsistência ","Todo em escola pública", "Todo em escola particular","Escola
pública e particular"))
```

```
abline(h=c(0,100), v=(20.75))
```

Gráfico de barras: Grupo de curso (desempenho) versus Modalidade do E.M.

x11()

```
barplot(100*prop.table(table(OVEU2016$Q20..Modalidade.de.Curso.do.Ensino.Médi
o,OVEU2016$grupo),2), col=color(5),beside=T, ylim=c(0,100), names=c("Grupo
1","Grupo 2","Grupo 3","Grupo 4"), xlab="Grupo de Curso (desempenho)", ylab="%")
```

```
legend("topright", col=color(5), pch=15, cex=0.8, legend=c("Magistério", "Outro",
"Regular", "Supletivo", "Técnico-profissionalizante"))
```

```
abline(h=c(0,100), v=(24.9))
```

Gráfico de barras: Grupo de curso (desempenho) versus Tentou Cursinho

```
x11()
```

```
barplot(100*prop.table(table(OVEU2016$Q22..Frequentou.cursinho.para.prestar.EN
EM,OVEU2016$grupo),2), col=color(6),beside=T, ylim=c(0,100), names=c("Grupo
1","Grupo 2","Grupo 3","Grupo 4"), xlab="Grupo de Curso (desempenho)", ylab="% ")
```

```
legend("topright", col=color(6), pch=15, cex=0.8, legend=c("Não","Sim, cursinho da
rede particular","Sim, cursinho da rede pública","Sim, cursinho da UFRN","Sim,
cursinho e curso de matérias isoladas", "Sim, curso de matérias isoladas"))
```

```
abline(h=c(0,100), v=(29.05))
```

```
# Gráfico de barras: Grupo de curso (desempenho) versus Meio de transporte que
mais utiliza
```

```
x11()
```

```
barplot(100*prop.table(table(OVEU2016$Q13..Meio.de.transporte.que.mais.utiliza,O
VEU2016$grupo),2), col=color(3),beside=T, ylim=c(0,100),names=c("Grupo
1","Grupo 2","Grupo 3","Grupo 4"), xlab="Grupo de Curso (desempenho)", ylab="%")
```

```
legend("topright", col=color(3), pch=15, cex=0.8, legend=c("Coletivo","Veículo
próprio ou da família","Outro"))
```

```
abline(h=c(0,100), v=(16.6))
```

```
##### Análise exploratória de desempenho via tabela =====
```

```
# por grupo de cursos (criado segundo desempenho)
```

```
table(OVEU2016$grupo, status_abr_2017)
```

```
round(prop.table(table(OVEU2016$grupo, status_abr_2017),1)*100,2)
```