



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Mineração de Dados Aplicada ao Enade

Giovanni Torres Corso
Mateus Denucci Garcia Seabra Resende

Monografia apresentada como requisito parcial
para conclusão do Bacharelado em Ciência da Computação

Orientador
Prof. Dr. Jan Correa

Brasília
2018

Dedicatória

Dedico esse trabalho aos meus pais e todos que acompanharam a longa jornada que foi desenvolvê-lo.

-Mateus Denucci Garcia Seabra Resende

Dedico este trabalho aos meus pais e irmãos por acreditarem em mim e por me incentivarem com este trabalho até o final.

-Giovanni Torres Corso

Agradecimentos

Existe uma frase que identifico com uma certa frequência na minha vida: “Deus escreve certo com linhas tortas”. Por isso, primeiramente gostaria de agradecer a Deus por ter me guiado nesse caminho e tornado tudo possível. Gostaria de agradecer toda a minha família, em especial meus pais, que nunca mediram esforços para sempre me proporcionar o melhor. Outra frase famosa que com o passar dos anos se mostra cada vez mais verdadeira vida é: “Quem tem amigos tem tudo”. Tenho muitos amigos que me acompanharam nessa jornada e caso fosse citar todos este trabalho dobraria de tamanho. Por isso vou citar somente meu grande amigo Giovanni Torres que além de ser um grande companheiro na vida também foi minha dupla na realização desse trabalho. Gostaria também de agradecer os professores da UnB por todo conhecimento que me foi passado. Por último gostaria de agradecer meu professor orientador Jan Corrêa por me ter auxiliado trabalho.

-Mateus Denucci Garcia Seabra Resende

Agradeço primeiramente a Deus, por todas as oportunidades que me deu e por todo Seu auxílio e amor. Agradeço aos meus pais e irmãos por todo amor e suporte, por serem um dos pilares da minha vida me dando base e sustentação para realizar qualquer sonho e cumprir qualquer objetivo. Agradeço aos meus amigos, em especial os que estiveram mais próximos a mim nesse período: Maísa Menezes, Felipe Ken, Tarcísio Júnior, Victor Thuin, Luciano Santos por todo apoio que me deram e por estarem ao meu lado quando eu precisei. Em especial agradeço ao meu amigo-irmão Mateus Denucci por, além de sua amizade, ter realizado este trabalho comigo, por todo empenho e apoio tornando possível essa formatura. Agradeço também aos professores do Departamento de Ciência da Computação da UnB que contribuíram muito com meu desenvolvimento profissional, em especial ao meu orientador Jan Mendonça pela orientação e auxílio.

-Giovanni Torres Corso

Resumo

O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) disponibiliza os dados recolhidos por pesquisas, avaliações e exames realizados. Os dados referentes ao exame nacional de desenvolvimento dos estudantes (Enade) são um exemplo disso. Esses dados contêm diversas informações sobre os estudantes que realizaram o exame. Neste trabalho foram aplicadas técnicas de visualização, análise e mineração de dados por meio de recursos como tabelas, gráficos e a ferramenta Weka. O Enade do ano de 2014 foi estudado de forma geral por ser o ano mais atual contendo os cursos de computação, mas houve um foco diferenciado para o curso de Ciência da Computação (Bacharel) e para os alunos da Universidade de Brasília (UnB). Foram feitas diversas análises em cima dos dados como correlações entre colunas, classificação da nota com relação aos atributos e análise das respostas dos estudantes nos itens do exame. O objetivo dessas análises foi entender melhor o perfil dos estudantes do ensino superior no Brasil.

Palavras-chave: Mineração de dados, ENADE, graduação, Weka, UnB, Ciência da Computação

Abstract

The Anísio Teixeira National Institute for Educational Studies and Research (INEP) publishes data from various sources, including research, assessments, and past exams. One example is data on students from the National Assessment of Student Achievement (ENADE). This paper presents the results of applying visualization techniques, analyses, and data mining algorithms using resources like tables, charts, and the software Weka. The 2014 edition of ENADE was studied as a whole, but focused on comparing University of Brasilia (UnB) Computer Science bachelor's degree students with students pursuing the same degree at other universities. Several analyses were made with this data with the goal of better understanding the profile of Brazilian college students. They include correlation between attributes, classification of scores based on those attributes, and analysis of the students' answers on the exam.

Keywords: Data mining, ENADE, undergraduate, Weka, UnB, computer science

Sumário

1	Introdução	1
1.1	Problema	1
1.2	Motivação	2
1.3	Objetivo	2
1.3.1	Objetivos Específicos	2
1.4	Visão Geral do Trabalho	2
2	Referencial Teórico	4
2.1	Dados, Informação e Conhecimento	4
2.2	Mineração de Dados e o Processo de Extração do Conhecimento	5
2.2.1	CRISP-DM	5
2.2.2	Técnicas e Algoritmos de <i>DataMining</i>	7
2.3	WEKA	8
2.3.1	O formato “arff”	9
3	Estudo de Caso: Enade 2014	11
3.1	Revisões Bibliográficas	11
3.2	Coleta de Dados	12
3.3	Tratamento de Dados	12
3.4	Análise Geral dos Dados	14
3.4.1	Nota bruta na formação geral	14
3.4.2	Nota bruta na formação específica	15
3.4.3	Nota bruta da prova	15
3.5	Análise das tabelas dos atributos nas partes de Formação Geral, Compo- nente Específico e Nota Geral	16
3.5.1	Mediana e Média	17
3.5.2	Análise das tabelas	18
3.6	Boxplots	19

3.7	Análise do desempenho dos estudantes em cada questão - Universidade de Brasília - Ciência da Computação (Bacharel)	33
3.8	Correlações	44
3.8.1	Preparação prévia e adaptação dos dados	44
3.8.2	Transformações e análises	51
3.9	Mineração de Dados utilizando o Weka	59
3.9.1	Preparação dos Dados	59
3.9.2	Análise da mineração	61
4	Conclusões e Trabalhos Futuros	66
	Referências	69
	Apêndice	70
A		71
A.1	Regras Weka JRip	71
A.1.1	Ciência da Computação (bacharel)	71
A.1.2	Ciência da Computação (licenciatura)	73
A.1.3	Enade Geral nota CE	73
A.1.4	Enade Geral nota FG	78

Lista de Figuras

2.1	Ciclo de vida de um projeto de mineração de dados baseado em [1].	6
2.2	GUI do Weka no Windows.	10
2.3	Exemplo de arquivo .arff.	10
3.1	Questão 14 do Questionário do Estudante.	18
3.2	Questão 13 do Questionário do Estudante.	19
3.3	<i>Boxplots</i> sexo x nota_geral para todos os alunos do ENADE.	20
3.4	<i>Boxplots</i> questão i22 quantidade livros x nota geral para todos os alunos do ENADE.	21
3.5	<i>Boxplot</i> ação afirmativa x nota ger para todos os alunos do ENADE.	22
3.6	<i>Boxplot</i> renda família de etnia branca x nota geral para os alunos do ENADE (Parte 1).	25
3.7	<i>Boxplot</i> renda família de etnia branca x nota geral para os alunos do ENADE (Parte 2).	26
3.8	<i>Boxplot</i> renda família de etnia negra x nota geral para os alunos do ENADE (Parte 1).	27
3.9	<i>Boxplot</i> renda família de etnia negra x nota geral para os alunos do ENADE (Parte 2).	28
3.10	<i>Boxplot</i> específico do curso de CIC ação afirmativa x nota ger.	32
3.11	Questão 15 do ENADE 2014.	38
3.12	Questão 30 do ENADE 2014.	40
3.13	<i>Heatmap</i> das correlações do curso de CiC (bacharel).	56

Lista de Tabelas

3.1	Tabela comparando a nota bruta da formação geral.	14
3.2	Tabela comparando a nota bruta da formação específica.	15
3.3	Tabela comparando a nota bruta geral.	15
3.4	Tabela com as 20 maiores medianas da parte de Formação Geral apenas para alunos de Ciência da Computação (bacharel).	16
3.5	Tabela com as 20 maiores medianas da parte de Componente Específico apenas para alunos de Ciência da Computação (bacharel).	17
3.6	Tabela com as 20 maiores medianas da parte de Nota Geral apenas para alunos de Ciência da Computação (bacharel).	17
3.7	Tabela sexo x nota_geral para todos os alunos do ENADE.	23
3.8	Tabela da questão i22 quantidade livros x nota geral para todos os alunos do ENADE.	23
3.9	Tabela ação afirmativa x nota ger para todos os alunos do ENADE.	24
3.10	Tabela renda família de etnia branca x nota geral para os alunos do ENADE.	26
3.11	Tabela renda família de etnia negra x nota geral para os alunos do ENADE.	28
3.12	As 10 respostas com maior número de alunos com renda acima de 30 salários mínimos (com base em todos os alunos do ENADE).	29
3.13	As 10 respostas com maior número de alunos com renda de até 1,5 salário mínimo (com base em todos os alunos do ENADE).	29
3.14	Tabela específica do curso de CIC ação afirmativa x nota ger.	32
3.15	Tabela que mostra o numero de estudantes que realizaram a prova em cada uma das instituições utilizadas na análise.	35
3.16	Tabela comparando o percentual de acertos dos estudantes de Ciência da Computação (bacharel) da UnB nas questões de 1 a 6.	35
3.17	Tabela comparando o percentual de acertos dos estudantes de Ciência da Computação (bacharel) da UnB nas questões de 7 a 13.	36
3.18	Tabela comparando o percentual de acertos dos estudantes de Ciência da Computação (bacharel) da UnB nas questões de 14 a 20.	36

3.19	Tabela comparando o percentual de acertos dos estudantes de Ciência da Computação (bacharel) da UnB nas questões de 22 a 27.	37
3.20	Tabela comparando o percentual de acertos dos estudantes de Ciência da Computação (bacharel) da UnB nas questões de 28 a 35.	37
3.21	Tabela mostrando a media da nt_ce das 10 intuições com as maiores notas em Ciência da Computação (bacharel) no ENADE (com a adição da UnB).	38
3.22	Lista dos atributos usados e as respectivas medianas para o ENADE geral.	42
3.23	Lista dos atributos usados e as respectivas medianas para os alunos do curso de Ciência da Computação (bacharel).	43
3.24	Parte da tabela com atributos e suas médias para cada curso.	46
3.25	Cursos ordenados pela média do atributo "horas de estudo a parte por semana".	48
3.26	Cursos ordenados pela média do atributo "Considerando questões objetivas percebeu que aprendeu o conteúdo".	49
3.27	Cursos ordenados pela média do atributo "Renda familiar".	50
3.28	Cursos ordenados pela média do atributo "Situação trabalho"(em horas semanais).	51
3.29	Parte da tabela de correlações de instituições federais.	53
3.30	Parte da tabela de correlações de instituições privadas.	54
3.31	Parte da tabela com as diferenças entre correlações federal e privada.	55
3.32	Legenda para o <i>heatmap</i> 3.13.	57
3.33	Correlações das notas FG de instituições federais.	58
3.34	Correlações de atributos para o curso de Computação (Bacharelado) de instituições privadas.	59

Capítulo 1

Introdução

Atualmente uma grande quantidade de dados é gerada diariamente, contudo extrair informações destes dados nem sempre é uma tarefa fácil. Uma forma para extrair informações úteis dos dados é utilizando a mineração de dados. Existem ferramentas que auxiliam nesse processo de mineração de dados.

Segundo o site do INEP, o Exame Nacional de Desempenho dos Estudantes (Enade) avalia estudantes concluintes dos cursos de graduação em relação aos conteúdos programáticos, habilidades e competências adquiridas em sua formação. Esta prova avalia o desempenho destes estudantes. Os resultados obtidos dão base para os indicadores de qualidade da educação superior, o que permite conhecer e comparar a qualidade de ensino superior de diversos cursos de todo o Brasil.

Além disso, os microdados disponíveis no Enade contém informações sobre os alunos, tais como etnia, renda, tipo de ensino, nota, seu respectivo curso, entre outras. Estes dados contém informações que podem não ser percebidas tão facilmente por seres humanos, contudo, com ajuda da mineração de dados pode ser possível extrair relações, regras, ou padrões existentes nesses dados.

Realizando a mineração de dados nestes microdados, pode ser possível encontrar padrões relacionados aos atributos e perfis dos estudantes, ou alguma característica destes que influencie em suas notas, etc.

1.1 Problema

Os microdados do Enade contém diversos dados sobre os aproximadamente 481 mil alunos que realizam a prova, contudo a partir destes dados brutos é difícil fazer análises significativas. Dois dos objetivos da aplicação de algoritmos de mineração de dados em sobre este conjunto de dados são: gerar informações úteis para o entendimento humano e procurar padrões não elementares em cima de um grande conjunto de dados.

1.2 Motivação

Avaliar a educação superior no Brasil é um problema muito importante. Mas o formato que os dados são disponibilizados não permitem uma análise direta. Transformar esses dados e aplicar algoritmos de mineração de dados podem auxiliar essa tarefa de extrair informações.

1.3 Objetivo

O objetivo deste trabalho é buscar algum padrão ou regra que determine ou influencie diretamente na nota dos alunos. Para isso, tem-se como base informações sobre os alunos e seus respectivos cursos, o que permite fazer análises e relações entre as características dos alunos, seus cursos e suas notas correspondentes.

1.3.1 Objetivos Específicos

Afim de alcançar o objetivo geral, foram definidos os seguintes objetivos específicos:

- Extrair os dados do Enade;
- Entender os dados;
- Realizar o tratamento desses dados;
- Aplicar as técnicas e algoritmos de mineração de dados nos dados obtidos;
- Analisar os resultados obtidos após a aplicação das técnicas de mineração.

1.4 Visão Geral do Trabalho

Este trabalho está dividido nos seguintes capítulos:

- Capítulo 1: Introdução
Este capítulo introduz o assunto do trabalho juntamente com o problema a ser investigado, motivação e objetivos a serem alcançados.
- Capítulo 2: Referencial Teórico
Este capítulo apresenta um referencial sobre as teorias e ferramentas necessárias para a execução desse trabalho.
- Capítulo 3: Estudo de caso: Enade 2014
Esse capítulo contém majoritariamente o processo de mineração de dados. Desde a coleta de dados até a análise dos resultados encontrados na mineração.

- Capítulo 4: Conclusões e Trabalhos Futuros

Esse capítulo contém as conclusões do trabalho realizado e sugestões de trabalhos futuros.

Capítulo 2

Referencial Teórico

Este capítulo apresenta um referencial sobre as teorias e ferramentas necessárias para a execução desse trabalho.

2.1 Dados, Informação e Conhecimento

Neste trabalho são utilizados conceitos como dados, informação e conhecimento. Para que não haja uma má interpretação, estes conceitos estão explicados a seguir.

Segundo Valdemar W. Setzer [2] dados são definidos como uma sequência de símbolos quantificados ou quantificáveis. Sendo que quantificável significa algo que pode ser quantificado e depois reproduzido sem que se perceba a diferença para com o original. Dados então podem ser compreendidos como um conjunto de códigos que possuem ou não algum significado atribuído. Como exemplos de dados pode-se citar (mesmo que seu conteúdo não seja compreensível): textos, fotos, figuras, sons gravados.

De acordo com Shannon citado em [3] "Informação é o que é transferido de uma fonte a um destino por um canal de comunicação, medido pela incerteza (probabilidade) do que não é antecipável em relação ao que pode ser esperado e entendido pelo receptor (ou destinatário)".

Valdemar W. Setzer [2] caracteriza conhecimento como "uma abstração interior, pessoal, de algo que foi experimentado, vivenciado, por alguém (...). Nesse sentido, o conhecimento não pode ser descrito; o que se descreve é a informação (se entendida pelo receptor), ou o dado. Também não depende apenas de uma interpretação pessoal, como a informação, pois requer uma vivência do objeto do conhecimento. Assim, o conhecimento está no âmbito puramente subjetivo do homem ou do animal.". [4]Entende-se por descoberta de conhecimento o processo não trivial de identificar padrões nos dados que sejam válidos, potencialmente úteis e minimamente entendíveis.

2.2 Mineração de Dados e o Processo de Extração do Conhecimento

Diariamente, um número muito grande de dados de diversos tipos são coletados, armazenados e manipulados. Contudo, nem sempre esses dados são efetivamente utilizados.

Um ser humano é capaz de realizar essa tarefa sem muitas complicações se o conjunto de dados for de um tamanho razoável; entretanto, quando esse conjunto de dados se torna muito grande, um esforço enorme deve ser feito, ou, muitas das vezes é até impraticável realizar a tarefa.

Como alternativa, usam-se ferramentas e técnicas computacionais para auxiliar o ser humano nessa tarefa de Mineração de Dados [5], que é basicamente o processo de descoberta de padrões interessantes e úteis, assim como de conhecimento, de grandes volumes de dados. Esses dados podem incluir bases de dados, *data warehouses*, a *web*, entre outros repositórios de dados. Essas ferramentas podem ajudar o ser humano a identificar esses padrões, buscar correlações entre dados, estatísticas e de diversas outras formas que potencialize a utilização das informações obtidas.

2.2.1 CRISP-DM

O modelo CRISP-DM [6][7], cuja sigla significa *Cross-Industry Standard Process for Data Mining*, é um meio de guiar a mineração de dados. Pode ser considerada uma metodologia ou um modelo de processos. Como metodologia, ele descreve fases típicas de um projeto, as tarefas envolvidas em cada fase e uma explicação da relação entre essas tarefas. E como modelo de processos ele fornece uma visão geral do ciclo de vida de mineração de dados.

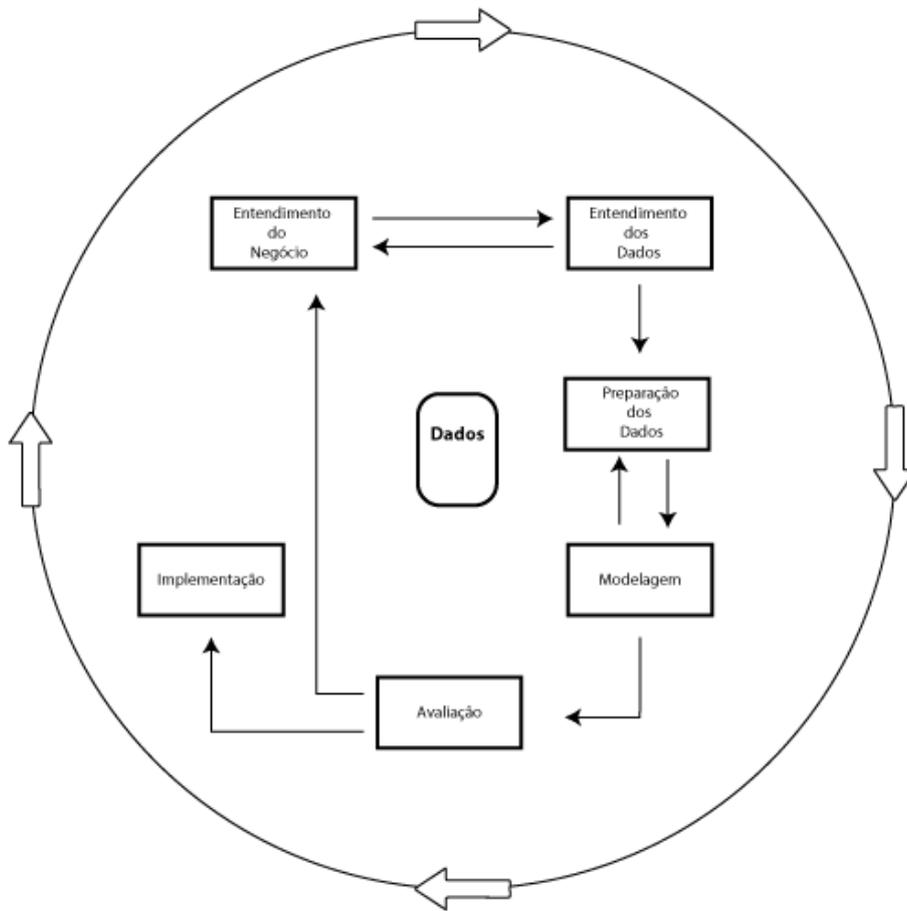


Figura 2.1: Ciclo de vida de um projeto de mineração de dados baseado em [1].

2.2.1.1 Entendimento do Negócio

Esta fase consiste em descobrir os objetivos, requerimentos e o que a empresa (ou organização) pretende obter com a mineração de dados, verificar também se a mineração pode ser utilizada para isso, e descobrir que tipo de dado pode ser coletado para montar um modelo.

2.2.1.2 Entendimento dos Dados

Esta parte do processo envolve analisar de forma mais aprofundada os dados disponíveis para uso, determina-se um conjunto de dados, sua qualidade, e avalia-se se ele é adequado para ser utilizado no futuro; essa fase é muito importante para evitar-se problemas durante a próxima fase - a preparação dos dados (que comumente é a mais longa do projeto).

2.2.1.3 Preparação dos Dados

Este passo é geralmente uma das partes mais importantes e mais trabalhosas do processo de mineração de dados. Envolve preprocessar os dados brutos para que os algoritmos de *machine learning* possam gerar um modelo. Algumas tarefas podem ser realizadas, como: agrupar conjuntos de dados ou registros; selecionar um subgrupo de dados específico; agregar registros, derivar novos atributos, separar dados para modelagem, remover ou realocar espaços brancos ou faltando, separar os dados em conjuntos de treino e teste.

2.2.1.4 Modelagem

Preparação e modelagem são fases que muitas vezes estão conectadas; resultados obtidos na modelagem geram informações que afetam a escolha de técnicas de processamento. Normalmente são feitas diversas iterações nessa fase; dificilmente uma questão estipulada no início será respondida com um único modelo e uma única execução, então costuma-se utilizar mais de um modelo e mais de uma execução de algoritmos até chegar a um resultado satisfatório.

2.2.1.5 Avaliação

É a fase de análise dos resultados. Precisa-se identificar se as descrições estruturais extraídas dos dados têm de fato algum valor para realizar previsões ou se são apenas falsos padrões.

2.2.1.6 Implantação

É a fase final do processo de CRISP-DM. Não importa o quão bom o modelo obtido é ou o quão bem esse modelo se encaixa com os dados, essas informações obtidas precisam ser aplicadas de alguma maneira. O processo CRISP-DM tem como objetivo final aumentar o conhecimento em cima dos dados obtidos e é justamente nessa fase, depois de passar por todas as outras cinco fases, que isso acontece. Essa fase pode ser simples como apenas a criação de um relatório.

2.2.2 Técnicas e Algoritmos de *DataMining*

A seguir são apresentadas técnicas de mineração de dados, sendo que para cada uma delas existem diversos algoritmos.

2.2.2.1 Clusterização

A Clusterização é considerada a classificação não supervisionada de padrões [8] de atributos, dados, vetores de atributos, entre outros elementos, formando então agrupamentos, também chamados *clusters*. Esses *clusters* são formados organizando-se os padrões baseados na similaridade existente entre os elementos que compõem cada agrupamento.

Nessa classificação não supervisionada, que é quando os dados não são previamente classificados ou rotulados, identifica-se uma dificuldade que é agrupar esses padrões de forma que exista alguma significância nesse padrão obtido em cada agrupamento, que é unido por padrões em comum, podendo-se então obter uma classificação, hipóteses novas ou alguma informação sobre as inter-relações dos dados.

Essa busca por padrões pode usar vários métodos possíveis, não existindo um método “genérico”, e, em consideráveis vezes, não chega-se a um resultado concreto ou expressivo.

2.2.2.2 Classificação

Classificação é uma técnica de mineração de dados que tem como função categorizar ou definir um conjunto de dados [9].

Na classificação cria-se uma espécie de modelo que determine a saída de uma nova instância de dados [10]. Pode-se, por exemplo, criar uma árvore de decisão de forma que se determine a saída dessa nova instância de dados, onde cada nó representa uma decisão a ser tomada, levando até as folhas da árvore, onde se mostra a saída prevista. Algoritmos de classificação também podem ser conhecidos como árvore de classificação ou de decisão, sendo assim um algoritmo de decisão ou classificação. Criam-se então regras que vão ter alguma finalidade, como por exemplo, e normalmente, tomar uma decisão ou realizar alguma previsão.

2.3 WEKA

A ferramenta que será utilizada no processo de mineração de dados é o software Weka ("*Waikato Environment for Knowledge Analysis*"). O Weka[11] é um pacote de algoritmos de aprendizado de máquina úteis em tarefas de mineração de dados. Ele contém ferramentas para: pré-processamento de dados, classificação, regressão, clusterização, regras de associação e visualização. Além de todos esses recursos, também pode ser utilizado para o desenvolvimento de novos algoritmos de aprendizado de máquina. O software começou a ser escrito em 1993 utilizando a linguagem de programação Java na Universidade de Waikato, Nova Zelândia. Weka é um software aberto e sua distribuição segue os termos da GNU (*General Public Licence*), o que torna seu uso livre, sem a necessidade

de compra de licença. É possível utilizar a plataforma por meio de uma GUI (*Graphical User Interface*) que permite a fácil manipulação de dados e visualização dos resultados ou ainda pode ser chamado diretamente de um código Java.

Porque o Weka foi escolhido:

- *Open source*: possui a licença GLP 3.0
- Multiplataforma: Disponível para Windows, Linux e Mac OS
- Uso de algoritmos complexos com uma interface simples

2.3.1 O formato “arff”

Será utilizada a ferramenta com a GUI, como mostra a Figura 2.2, e para isso os dados devem estar em um formato de arquivo especial chamado “arff” (sigla para *Attribute-Relation File Format*). Este arquivo está no formato ASCII e contém uma lista de instâncias que compartilha uma lista de atributos em comum. Arquivos ARFF [12] foram desenvolvidos no departamento de Ciência da Computação da Universidade de Waikato, Nova Zelândia para ser usado juntamente com o Weka. Esses arquivos podem ser gerados de maneira simples e são divididos em duas partes: cabeçalho e dados.

O cabeçalho, como pode ser visualizado na Figura 2.3, contém o nome da relação, a lista com os atributos contidos nos dados e seus respectivos tipos. Por exemplo:

```
@RELATION carro
@ATTRIBUTE velocidadeMaxima NUMERIC
@ATTRIBUTE ano NUMERIC
@ATTRIBUTE cor preto,branco,vermelho,prata
```

Os dados são as instâncias do seu banco de dados propriamente dito. Seguindo o exemplo acima uma possível área de dados teria o seguinte formato:

```
@DATA
150,2009,preto
250,2017,prata
170,2011,branco
198,2012,preto
```



Figura 2.2: GUI do Weka no Windows.

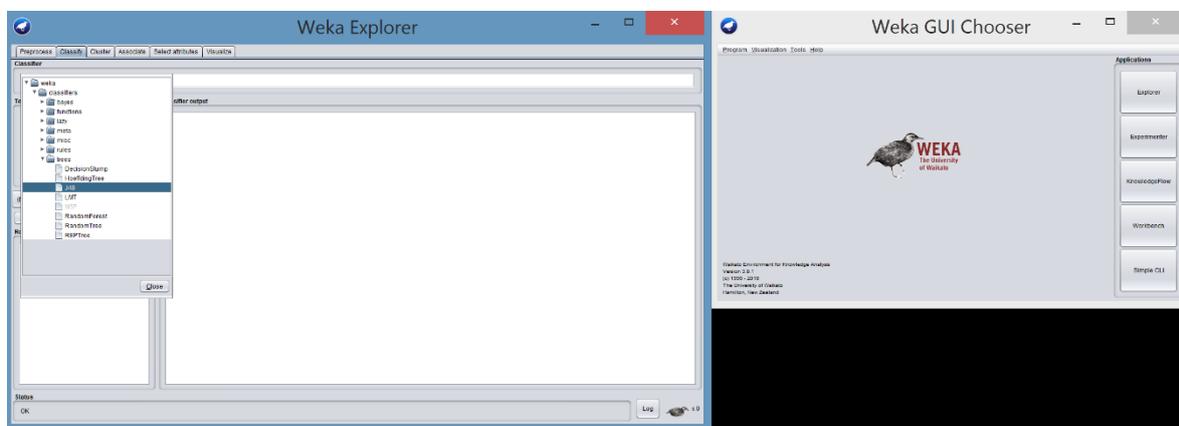


Figura 2.3: Exemplo de arquivo .arff.

Capítulo 3

Estudo de Caso: Enade 2014

O Exame Nacional de Desenvolvimento dos Estudantes (Enade) [13] é um dos procedimentos de avaliação do Sistema Nacional de Avaliação da Educação Superior (Sinaes), realizado pelo INEP. Ele tem como objetivo avaliar o rendimento dos alunos concluintes dos cursos de graduação, em relação aos conteúdos programáticos, habilidades e competências adquiridas em sua formação, assim avaliando a qualidade do curso de formação superior. A realização do exame é obrigatória e esta informação deve constar em seu histórico escolar.

O Enade 2014 foi escolhido como objeto do estudo pois, até a data presente, é o mais recente contendo dados sobre os cursos da área de Computação.

3.1 Revisões Bibliográficas

Alguns artigos pertinentes ao assunto abordado neste trabalho são citados a seguir.

No artigo "*Math and Science Students' Performance Analysis*." [14], por exemplo, os autores procuram descobrir fatores que influenciam o desempenho dos alunos na nota geral do Enade utilizando uma regressão linear.

No artigo "Técnicas de mineração de dados aplicadas aos microdados do ENADE para avaliar o desempenho dos acadêmicos do curso de Ciência da Computação no Rio Grande do Sul utilizando o software R." [15], os autores usaram clusterização hierárquica para achar agrupamentos de instituições baseado no desempenho dos alunos.

No artigo "Fatores relacionados aos concluintes do curso de Licenciatura em Matemática - uma análise dos resultados do Enade.", usaram o algoritmo de classificação *Naive Bayes* para tentar prever se a probabilidade da nota dos alunos foi "baixa", "mediana" ou "alta" usando os outros atributos. Esta precisão foi feita para os alunos do curso de licenciatura em matemática.

No artigo "Mineração de dados aplicada na base do enade com enfoque na criação de perfis dos estudantes que prestaram o exame utilizando o algoritmo J48"[16] o autor utiliza o Weka para achar árvores de decisão sobre os dados do Enade 2013 usando o desempenho dos alunos como classe.

3.2 Coleta de Dados

É possível baixar os micro dados do ENADE das edições entre 2004 (ano de início) e 2015 no próprio site do INEP e seu uso é livre. Os diretórios disponíveis após o *download* dos micro dados são:

- Dados (Contém o(s) arquivo(s) dos microdados do Enade 2014)
 - microdados_enade_2014 (csv);
- Inputs (Contém os seguintes inputs para leitura dos microdados)
 - input_enade (sas);
 - input_enade (R);
 - .Rhistory (Rhistory);
- Documentação
 - Manual do usuário (pdf);
 - Dicionário de variáveis (xls);
 - questionario_estudante_enade_2014 (pdf);
 - questionario_licenciaturas_enade 2014 (pdf);

O arquivo "microdados_enade_2014" no formato "csv" (*Comma-separated values*) pode ser visualizado de forma organizada no Microsoft Excel. A manipulação de arquivos csv pode ser feita de maneira eficiente com linguagens de programação como python. O uso de dados públicos é muito vantajoso em trabalhos de mineração de dados justamente por isso: fácil acesso, uso aberto e gratuito e formato já adequado.

3.3 Tratamento de Dados

Conforme a metodologia CRISP-DM [1] Seção 2.2.1.3, uma das etapas importantes para a mineração de dados é justamente o tratamento dos dados de forma que eles possam ser processados por algoritmos de aprendizado de máquina. Para utilizar a ferramenta

Weka é necessário criar o arquivo no formato “arff”. Para realizar essa tarefa foi utilizada a linguagem Python versão 3.5.2 devido a sua simplicidade para a manipulação de arquivos texto, por ter várias bibliotecas úteis que puderam ser utilizadas para a manipulação e tratamento dos dados, e por ser um *software* livre. Esse arquivo “.arff”, conforme mostrado na Seção 2.3.1, é composto por duas partes: cabeçalho e dados.

Primeiramente, utilizando os arquivos “Dicionário de Variáveis.xls”, que é um arquivo basicamente com os nomes das questões, “questionario_ estudante_enade_2014.pdf”, que são as perguntas direcionadas para cada aluno de forma mais pessoal, e “questionario_ licenciaturas_enade_2014.pdf”, que são questões apenas para os estudantes dos cursos de licenciatura, (que estão contidos na pasta DOCUMENTACAO) como referência, foi criado o cabeçalho do arquivo arff com as informações dos atributos. Então foram modificados os nomes de alguns atributos para que ficassem mnemônicos e sua análise fosse facilitada. Utilizando um *script* escrito em Python, os números correspondentes às opções de cada atributo foram substituídos pelo seu valor nominal para facilitar a análise. Esse tratamento dos dados pode ser verificado no exemplo:

- @attribute qe_i3 {a,b,c}
- @attribute i3_nacionalidade {"Brasileira.", "Brasileira naturalizada.", "Estrangeira."}

Existem dois campos denominados “vt_ace_ofg” e “vt_ace_oce” que são *strings* que armazenam o resultado das respostas do aluno na parte objetiva na formação geral e os acertos da parte objetiva do componente específico respectivamente. Essas *strings* são formadas por uma sequência de números onde: 0=Errado, 1=Certo, 8=Anulada pela comissão e 9=Anulada pelo índice de discriminação (correlação pontobisserial < 0,20). O campo “vt_ace_ofg” é composto por uma sequência de 8 números e “vt_ace_oce” por uma sequência de 27 números. Cada número dessa sequência se tornou um atributo nominal onde seus possíveis valores são: certo (caso o número seja 1), errado (caso o número seja 0) e NA (caso o número seja 8 ou 9). Dessa forma foram inseridos 35 novos atributos ao cabeçalho do arquivo arff. O objetivo disso é extrair as respostas de dentro desses vetores e usá-las como atributos de dados, de forma que um algoritmo de mineração de dados possa, posteriormente, analisar e possivelmente classificar esses dados. Um *script* desenvolvido em Python é responsável por passar linha a linha no arquivo de dados, extraído do microdados_enade_2014.csv, e adicionar os 35 novos campos em cada instância de acordo com os vetores correspondentes. Arquivo de cabeçalho e de dados são unidos e são adicionadas tags para completar a estrutura, criando o arquivo enade.arff, que é reconhecido pelo weka [17].

3.4 Análise Geral dos Dados

No total, o arquivo microdados possui 481720 instâncias onde cada instância representa um estudante inscrito no ENADE. Cada estudante possui três atributos relacionados as notas associadas a ele conforme mostrado no dicionario de variáveis Seção 3.2, são eles: "nt_fg", "nt_ce" e "nt_ger". A "nt_fg" é a nota bruta na parte de formação geral¹, "nt_ce" é a nota bruta na parte de componente específico² e "nt_ger" é a nota bruta geral da prova³. Dos 481720 inscritos no ENADE, 82,37% dos participantes de fato realizaram a prova e possuem uma nota atribuída.

No total, 9822 estudantes de Ciência da Computação (Bacharel) foram inscritos no ENADE. Destes, 84,26% de fato realizaram a prova e possuem uma nota atribuída.

Foram feitas algumas comparações analisando dados a respeito de todos os estudantes que compareceram no ENADE com dados referentes a somente alguns cursos. Essas comparações foram feitas utilizando os três atributos que dizem respeito as notas. Utiliza-se para essa comparação o termo "Quartil", que é uma separatriz que divide a área de uma distribuição em quatro domínios iguais. Exemplificando, o primeiro quartil ou 25 quartil é um valor onde 25% dos participantes obtiveram notas iguais ou inferiores a este valor; o segundo quartil, ou mediana, seria o valor onde 50% dos participantes obtiveram uma nota igual ou inferior a este valor, e assim por diante. A linha 'porcentagem' serve para indicar a porcentagem de alunos do total que tem essa nota associada e válida (os alunos que efetivamente participaram e possuem essa nota atribuída).

3.4.1 Nota bruta na formação geral

Tabela 3.1: Tabela comparando a nota bruta da formação geral.

	ENADE	Ciência da Computação (bacharel)
25 Quartil	42,1	48,1
Mediana	55,4	61,8
75 Quartil	67,8	73,1
Número de Instâncias	396776	8276
Percentual que fez a prova	82,37%	84,26%

A Tabela 3.1 compara os valores das notas brutas da formação geral de todos os estudantes que realizaram o ENADE com os estudantes de Ciência da Computação (Bacharel). Por meio dessa tabela é possível perceber que os estudantes de bacharelado em

¹Média ponderada da parte objetiva (60%) e discursiva (40%) na formação geral (0 a 100)

²Média ponderada da parte objetiva (85%) e discursiva (15%) no componente específico (0 a 100)

³Média ponderada da formação geral (25%) e componente específico (75%) (0 a 100)

Ciência da Computação obtiveram, na parte de formação geral, um desempenho de 7 a 15% superior a media dos estudantes do ENADE como um todo.

3.4.2 Nota bruta na formação específica

Tabela 3.2: Tabela comparando a nota bruta da formação específica.

	ENADE	Ciência da Computação (bacharel)
25 Quartil	29,1	28,3
Mediana	40,3	37,9
75 Quartil	52,5	49
Número de Instâncias	396776	8276
Percentual que fez a prova	82,37%	84,26%

A Tabela 3.2 compara os valores das notas brutas da formação específica de todos os estudantes que realizaram o ENADE com os estudantes de Ciência da Computação (Bacharel). Por meio dessa tabela é possível perceber que os estudantes de bacharelado em Ciência da Computação obtiveram uma nota mais baixa que a media dos estudantes do ENADE na parte de formação específica.

3.4.3 Nota bruta da prova

Tabela 3.3: Tabela comparando a nota bruta geral.

	ENADE	Ciência da Computação (bacharel)
25 Quartil	33,8	34,8
Mediana	44	43,75
75 Quartil	54,6	53,5
Número de Instâncias	396776	8276
Percentual que fez a prova	82,37%	84,26%

A Tabela 3.3 compara os valores das notas brutas da prova como um todo de todos os estudantes que realizaram o ENADE com os estudantes de Ciência da Computação (Bacharel). Embora seja difícil fazer uma comparação objetiva das provas na parte de componente específico devido às provas serem diferentes, é possível perceber que os estudantes de bacharelado em Ciência da Computação obtiveram um desempenho inferior quando analisada a mediana e o terceiro quartil. Já o primeiro quartil dos estudantes desse curso foi superior. Vale lembrar que a parte da prova referente ao componente específico é diferente para cada curso, logo fica difícil fazer uma comparação objetiva.

3.5 Análise das tabelas dos atributos nas partes de Formação Geral, Componente Específico e Nota Geral

Com base nos dados obtidos anteriormente referentes as notas de Formação Geral, Componente Específico e Nota Geral na Seção 3.4, decidiu-se fazer algumas análises em cima dos dados. Separaram-se os dados que foram considerados relevantes para esta análise em uma tabela, sendo eles: o atributo, os quartis (25 quartil, mediana e 75 quartil), o número de instâncias, que indica a quantidade bruta de alunos referente aquele atributo sendo analisado, e a porcentagem, que indica quanto esse número de instâncias representa do total de alunos de CiC que fizeram o ENADE. Os dados foram separados em três tabelas, uma para nota FG (de Formação Geral), uma para nota CE (de Componente Específico) e uma para nota GER (Geral).

Com isso, a ideia era analisar os dados e identificar algum padrão ou alguma relação significativa que pudesse existir entre os atributos e algum dos valores dos quartis. Essas tabelas e análises foram feitas tanto para alunos do Enade geral como também para alunos somente do curso de Ciência da Computação (bacharel). Por fim, os atributos de todas as tabelas (das notas FG, CE e GER) foram ordenadas pelas medianas, tanto de forma crescente como decrescente. Depois foram separados os 20 atributos com as maiores medianas (ordenados ainda pela mediana de forma decrescente), as Tabelas 3.4, 3.5 e 3.6 nos mostram esses valores referentes às tabelas geradas tendo como base os alunos de Ciência de Computação (bacharel) apenas.

Tabela 3.4: Tabela com as 20 maiores medianas da parte de Formação Geral apenas para alunos de Ciência da Computação (bacharel).

Atributo	25 Quaril	Mediana	75 Quartil	Numero Instancias	Porcentagem (%)
tp_def_fis."Mesa adaptada para cadeira de rodas"	90,3	90,3	90,3	1	0,01
i14_atividades_no_exterior."Sim; programa de intercambio financiado pelo Governo Federal (Marca; Brafitec; PLI; outro),"	65,6	72,3	78,9	39	0,4
i14_atividades_no_exterior."Sim; Programa Ciencia sem Fronteiras,"	62,5	71,6	80,7	270	2,75
qp_i9_tempo_gasto_p_concluir_a_prova.**	63,55	70,5	75,45	3	0,03
tp_def_vis."Prova em braile"	68,5	68,5	68,5	1	0,01
i11_tipo_bolsa_financiamento."ProUni Integral,"	57,45	67,8	76,85	543	5,53
i19_maior_incentivo_p_cursar."Lider ou representante religioso,"	55,15	67,8	73,85	7	0,07
qp_i1_grau_dificuldade_prova_parte_Formacao_Geral."Facil,"	55,6	67,7	77,2	1595	16,24
tp_def_fis."Carteira para canhoto"	57,775	67,35	73,025	8	0,08
i13_tipo_bolsa_academica."Bolsa de iniciacao cientifica,"	53,925	67,2	77,2	1106	11,26
i13_tipo_bolsa_academica."Bolsa PET,"	56,1	67	77,65	120	1,22
i13_tipo_bolsa_academica."Bolsa de monitoria/tutoria,"	52,5	66,8	76,25	364	3,71
qp_i9_tempo_gasto_p_concluir_a_prova."Entre tres e quatro horas,"	54,325	66,25	76	2334	23,76
i12_tipo_auxilio."Auxilio alimentacao,"	52,6	65,95	76	130	1,32
i19_maior_incentivo_p_cursar."Professores,"	52,5	65,8	74,5	153	1,56
i24_opportunidade_aprendizado_idioma_extrangeiro."Sim; parte na modalidade presencial e parte na modalidade semipresencial,"	52,975	65,8	76,3	440	4,48
i15_ingresso_por_politicas_acao_afirmativa."Sim; por sistema que combina dois ou mais criterios anteriores,"	52,6	65,7	75,1	191	1,95
i11_tipo_bolsa_financiamento."Nenhum; pois meu curso e gratuito,"	52,5	65,6	76,1	2864	29,16
i15_ingresso_por_politicas_acao_afirmativa."Sim; por ter estudado em escola publica ou particular com bolsa de estudos,"	52,5	65,4	75,4	609	6,2
i13_tipo_bolsa_academica."Bolsa de extensao,"	51,825	65,2	76	300	3,05

Tabela 3.5: Tabela com as 20 maiores medianas da parte de Componente Específico apenas para alunos de Ciência da Computação (bacharel).

Atributo	25 Quartil	Mediana	75 Quartil	Numero de Instancias	Porcentagem (%)
tp_def_fis "Mesa adaptada para cadeira de rodas"	79,9	79,9	79,9	1	0,01
i1_est_civil "Viuvo(a),"	58,2	58,2	58,2	1	0,01
i14_atividades_no_exterior "Sim; Programa Ciencia sem Fronteiras,"	48,5	57,55	66,075	270	2,75
i13_tipo_bolsa_academica "Bolsa PET,"	40,5	52,5	62,2	120	1,22
i14_atividades_no_exterior "Sim; programa de intercambio financiado pelo Governo Federal (Marca; Brafithec; PLI; outro),"	45,75	52,5	62,35	39	0,4
tp_sexo "N"	52	52	52	1	0,01
i13_tipo_bolsa_academica "Bolsa de iniciacao cientifica,"	40,5	51	62,1	1106	11,26
i12_tipo_auxilio "Auxilio alimentacao,"	36,4	48,5	58,1	130	1,32
in_vespertino "1"	36,9	48,5	58,7	2372	24,15
i13_tipo_bolsa_academica "Bolsa de monitoria/tutoria,"	36,4	47,75	59,15	364	3,71
i11_tipo_bolsa_financiamento "Nenhum; pois meu curso e gratuito,"	36,5	47,5	58,6	2864	29,16
i12_tipo_auxilio "Auxilio moradia e alimentacao,"	35,9	47,5	59,6	73	0,74
i41_coordenacao_do_curso_esteve_disponivel_p_orientar_alunos "Nao se aplica"	28,3	46,5	57,475	114	1,16
i51_atividades_durante_tcc_contribuiram_p_qualificar_sua_formacao_profissional "Nao sei responder"	35,4	46,5	58,1	593	6,04
in_matutino "1"	35,4	46,5	57,5	2794	28,45
qp_i2_grau_dificuldade_prova_parte_Componente_Especifico "Facil,"	32,9	46	60,6	277	2,82
qp_i4_enunciados_questoes_parte_Formacao_Geral_claros_e_objetivos "***"	45,5	45,5	45,5	1	0,01
in_noturno "0"	34,4	45	56,7	2899	29,52
i6_moradia_com_quem_onda "Em alojamento universitario da propria instituicao,"	32,8	44,75	59,875	24	0,24
i12_tipo_auxilio "Auxilio Permanencia,"	35,5	44,5	54,4	122	1,24

Tabela 3.6: Tabela com as 20 maiores medianas da parte de Nota Geral apenas para alunos de Ciência da Computação (bacharel).

Atributo	25 Quartil	Mediana	75 Quartil	Numero de Instancias	Porcentagem (%)
tp_def_fis "Mesa adaptada para cadeira de rodas"	82,5	82,5	82,5	1	0,01
i14_atividades_no_exterior "Sim; Programa Ciencia sem Fronteiras,"	52,175	61,8	68,25	270	2,75
i1_est_civil "Viuvo(a),"	59,5	59,5	59,5	1	0,01
i14_atividades_no_exterior "Sim; programa de intercambio financiado pelo Governo Federal (Marca; Brafithec; PLI; outro),"	50,2	58	64,2	39	0,4
i13_tipo_bolsa_academica "Bolsa PET,"	47,125	55,85	64,5	120	1,22
i13_tipo_bolsa_academica "Bolsa de iniciacao cientifica,"	45,2	54,7	64	1106	11,26
tp_sexo "N"	54,7	54,7	54,7	1	0,01
i12_tipo_auxilio "Auxilio alimentacao,"	41,9	53,75	60,825	130	1,32
in_vespertino "1"	42,3	52	61,6	2372	24,15
i11_tipo_bolsa_financiamento "Nenhum; pois meu curso e gratuito,"	42	51,8	61,4	2864	29,16
i13_tipo_bolsa_academica "Bolsa de monitoria/tutoria,"	41,85	51,25	62,125	364	3,71
i51_atividades_durante_tcc_contribuiram_p_qualificar_sua_formacao_profissional "Nao sei responder"	41	51,1	60,9	593	6,04
in_matutino "1"	40,9	50,5	60,9	2794	28,45
i12_tipo_auxilio "Auxilio moradia e alimentacao,"	43,4	50,3	61,6	73	0,74
qp_i4_enunciados_questoes_parte_Formacao_Geral_claros_e_objetivos "***"	50,2	50,2	50,2	1	0,01
tp_def_fis "Sala no terreo"	40,9	50	51,1	5	0,05
i6_moradia_com_quem_onda "Em alojamento universitario da propria instituicao,"	40,1	49,95	60,825	24	0,24
in_noturno "0"	40	49,8	59,9	2899	29,52
qp_i2_grau_dificuldade_prova_parte_Componente_Especifico "Facil,"	37,9	49,6	62,8	277	2,82
i8_renda_familia "Acima de 30 salarios minimos (mais de R\$ 21,720;01),"	37,6	49,5	59,35	143	1,46

3.5.1 Mediana e Média

A média aritmética simples de um conjunto de valores numéricos [18], também conhecida como média, é calculada somando-se todos os valores do conjunto e dividindo-se o resultado pelo número de elementos somados, que é igual ao número de elementos do conjunto, ou seja, a média de n números é sua soma dividida por n .

A mediana, é uma medida de localização do centro da distribuição dos dados. Ordenados os elementos da amostra, a mediana é o valor (pertencente ou não à amostra) que a divide ao meio, isto é, 50% dos elementos da amostra são menores ou iguais à mediana e os outros 50% são maiores ou iguais à mediana [19]. Para a sua determinação utiliza-se a seguinte regra, depois de ordenada a amostra de n elementos: Se n é ímpar, a mediana é o elemento médio; se n é par, a mediana é a média aritmética, isto é a metade da soma, dos dois elementos médios.

A vantagem da mediana em relação à média é que a mediana pode dar uma ideia melhor de um valor típico porque não é tão distorcida por valores extremamente altos ou baixos. No caso do nosso estudo, como um número considerável de alunos não se interessa em fazer a prova e acabam tirando muitas notas zero ou em extremos baixos, a média pode ser bastante prejudicada, enquanto a mediana mostra uma ideia melhor do desempenho de um aluno. Por exemplo, caso algum curso fizesse um boicote ao exame, a sua média seria muito prejudicada, contudo a mediana ainda daria uma ideia do desempenho geral dos alunos.

3.5.2 Análise das tabelas

Analisando as Tabelas 3.4, 3.5 e 3.6, conseguimos observar alguns dos atributos com valores significativos (não são considerados significativos os atributos com número muito pequeno de instâncias) dentro da amostra dos alunos de CIC que realizaram o ENADE, para todas as partes, por exemplo, o atributo “i14_atividades_no_exterior”, que representa o item 14 do Questionário do Estudante, como mostra a Figura 3.1, esteve presente entre as 20 melhores medianas em todas as partes e com mais de um tipo de resposta do atributo, sendo elas “Sim; programa de intercâmbio financiado pelo Governo Federal (Marca; Brafitec; PLI; outro),” com uma porcentagem de 0,40%, e "Sim; Programa Ciencia sem Fronteiras," com 2,75%, e apenas na parte de Formação Geral, mais um campo deste atributo apareceu entre as 20 melhores, sendo ele "Sim; outro intercambio nao institucional" com 1,11%.

14. Durante o curso de graduação você participou de programas e/ou atividades curriculares no exterior?
- A () Não participei.
 - B () Sim, Programa Ciência sem Fronteiras.
 - C () Sim, programa de intercâmbio financiado pelo Governo Federal (Marca; Brafitec; PLI; outro).
 - D () Sim, programa de intercâmbio financiado pelo Governo Estadual.
 - E () Sim, programa de intercâmbio da minha instituição.
 - F () Sim, outro intercâmbio não institucional.

Figura 3.1: Questão 14 do Questionário do Estudante.

Além deste, o atributo “i13_tipo_bolsa_academica”, que representa o item 13 do Questionário do Estudante, como mostra a Figura 3.2, também está presente em todas as partes, com mais de uma opção e com porcentagens consideráveis do total de alunos. São elas: “i13_tipo_bolsa_academica_”Bolsa de iniciacao cientifica,”com 11.26%, i13_tipo_bolsa_academica_”Bolsa PET,”com 1.22% e i13_tipo_bolsa_academica_”Bolsa de monitoria/tutoria,”com 3.71%. E apenas na parte de Componente Específico, ainda temos “i13_tipo_bolsa_academica_”Bolsa de extensao,”com uma porcentagem de 3.05%.

13. Ao longo da sua trajetória acadêmica, você recebeu algum tipo de bolsa acadêmica? No caso de haver mais de uma opção, marcar apenas a bolsa de maior duração.
- A () Nenhum.
 - B () Bolsa de iniciação científica.
 - C () Bolsa de extensão.
 - D () Bolsa de monitoria/tutoria.
 - E () Bolsa PET.
 - F () Outro tipo de bolsa acadêmica.

Figura 3.2: Questão 13 do Questionário do Estudante.

3.6 Boxplots

Em busca de novas ferramentas que possibilitassem uma análise visual dos valores dos dados, procurou-se gerar recursos que fossem uma alternativa para estas análises por tabela. Utilizando Python 3.5.2 [20] foram gerados diversos materiais com o objetivo de auxiliar na visualização desse grande volume de dados. Um dos materiais gerados foi o *boxplot*. Um *boxplot* [21] é uma ferramenta gráfica em forma de retângulo onde a base inferior e superior (topo) do retângulo representam o primeiro e o terceiro quartis respectivamente; uma linha que passa horizontalmente dentro do retângulo é a mediana (ou segundo quartil), e as linhas que se estendem para cima do terceiro quartil e para baixo do primeiro quartil até chegarem em duas hastes horizontais representam os limites superiores e inferiores respectivamente, o que indica que todos os valores estão dentro destes limites. Para cada um dos atributos (explicados nas seções 3.2, 3.3 e 3.4) existente foi criado um *boxplot* relacionando cada opção de resposta do atributo a *nota_fg*, *nota_ce*, *nota_ger*. Inicialmente, esses *boxplots* foram criados tendo como base todos os alunos do Enade.

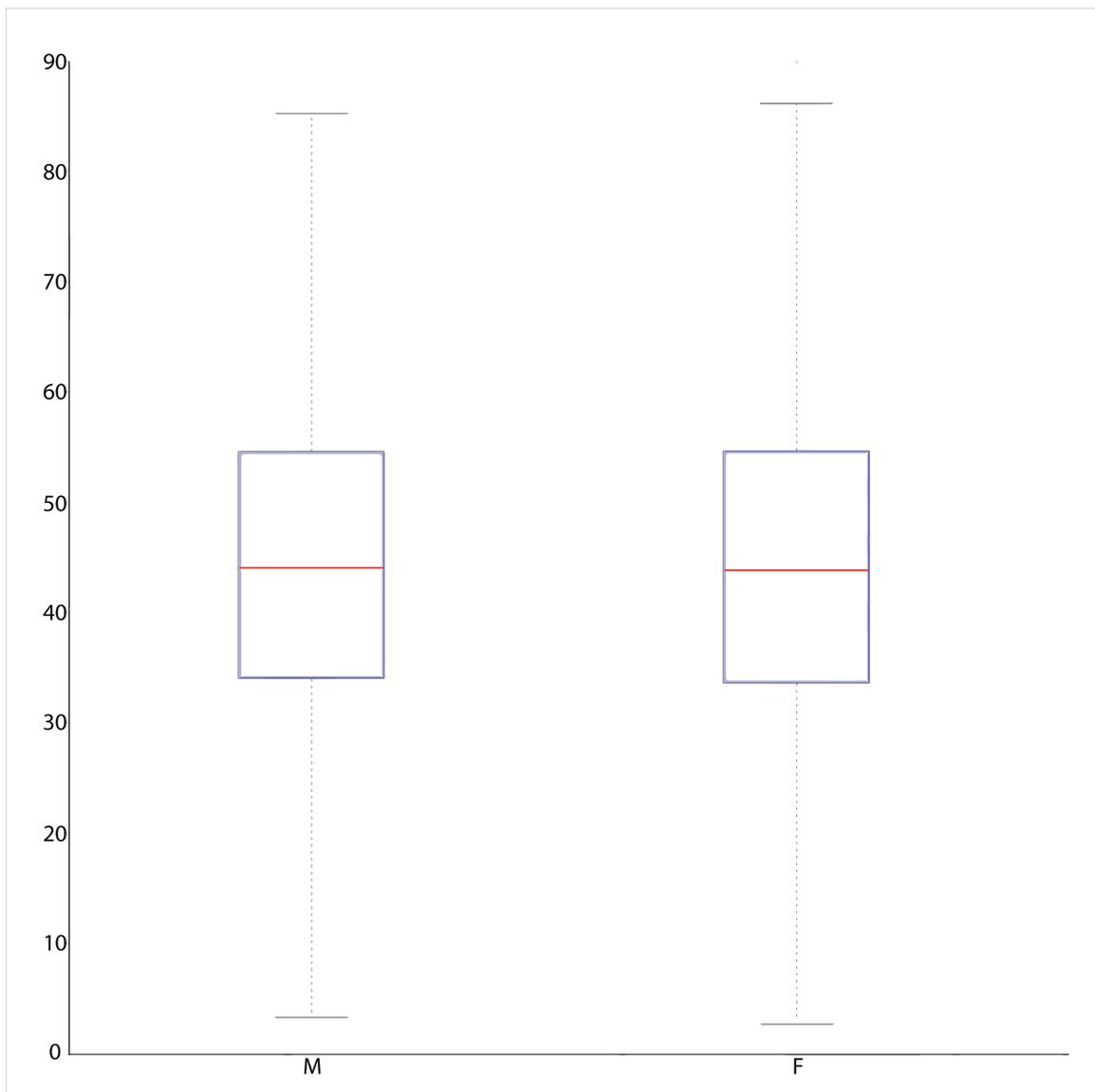


Figura 3.3: *Boxplots* sexo x nota_geral para todos os alunos do ENADE.

Como exemplo prático, pode ser visualizado o *boxplot* representado pela imagem 3.3. Nessa imagem pode-se ver a relação entre o atributo 'sexo', contendo os valores "M" e "F" (descritos na parte de baixo) e a 'nota_ger' (descrita na parte esquerda da imagem), é possível perceber que o limite máximo, o primeiro quartil, segundo quartil, terceiro quartil e o limite mínimo são bem próximos independente do sexo. Olhando para a escala à esquerda, é possível ter uma noção dos valores.

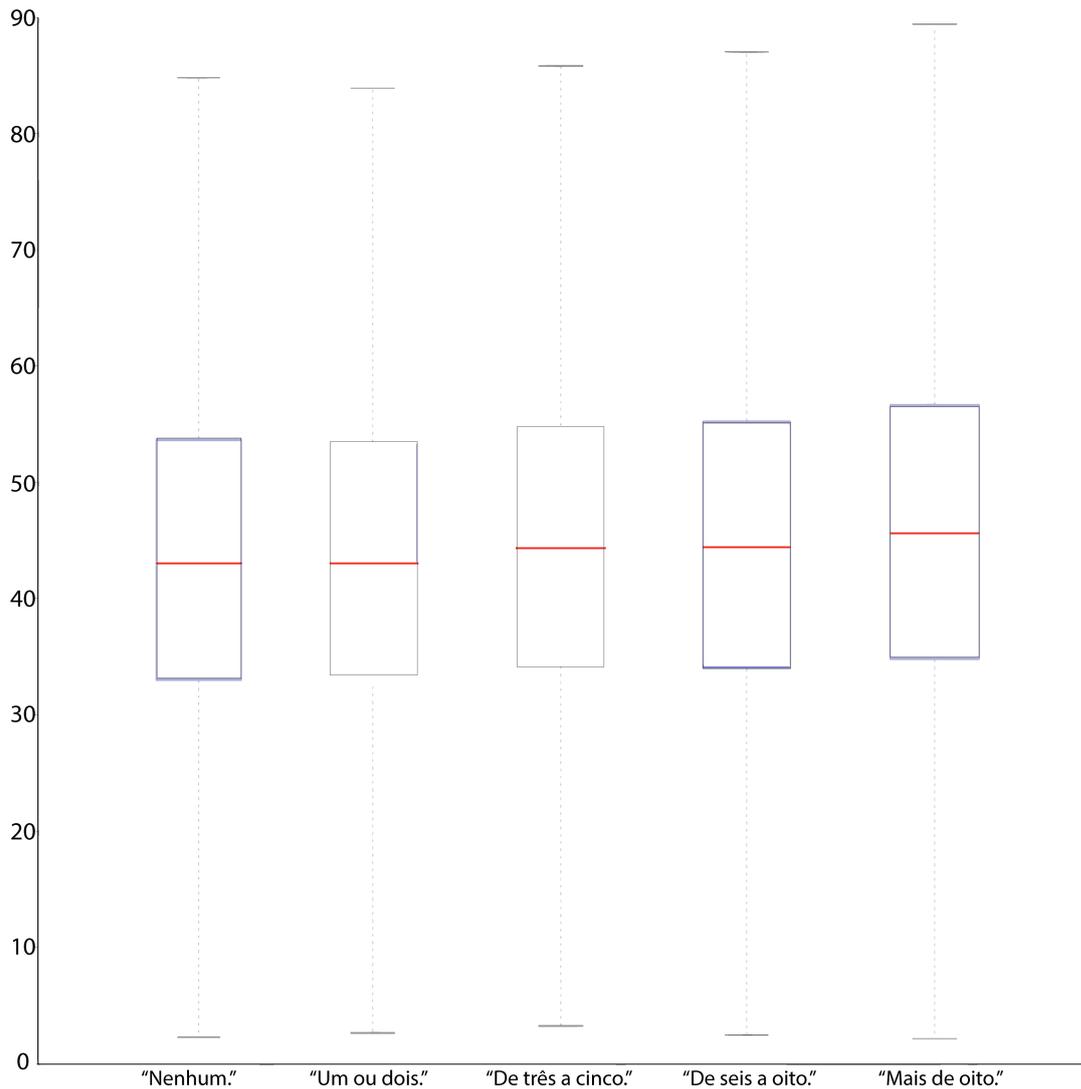


Figura 3.4: *Boxplots* questão i22 quantidade livros x nota geral para todos os alunos do ENADE.

Um outro *boxplot*, indicado pela imagem 3.4 acima mostra a relação entre a resposta da questão 22 do questionário do estudante (ver Seção 3.2) e a nota_ger. Cada um dos *boxplots* representa uma das opções que foi dada ao estudante com relação a pergunta “Quantos livros você leu no último ano”. As opções eram: “Nenhum.”, “Um ou dois.”, “De três a a cinco.”, “De seis a oito” e “Mais de oito.”. Com exceção do primeiro e segundo *boxplots* que são semelhantes, pode-se notar como o retângulo como um todo juntamente com a linha vermelha (equivalente a mediana) está um pouco mais alto a cada gráfico. Isso pode ser interpretado como: quanto maior o número de livros lidos no ano anterior, maior foi a nota geral bruta dos alunos.

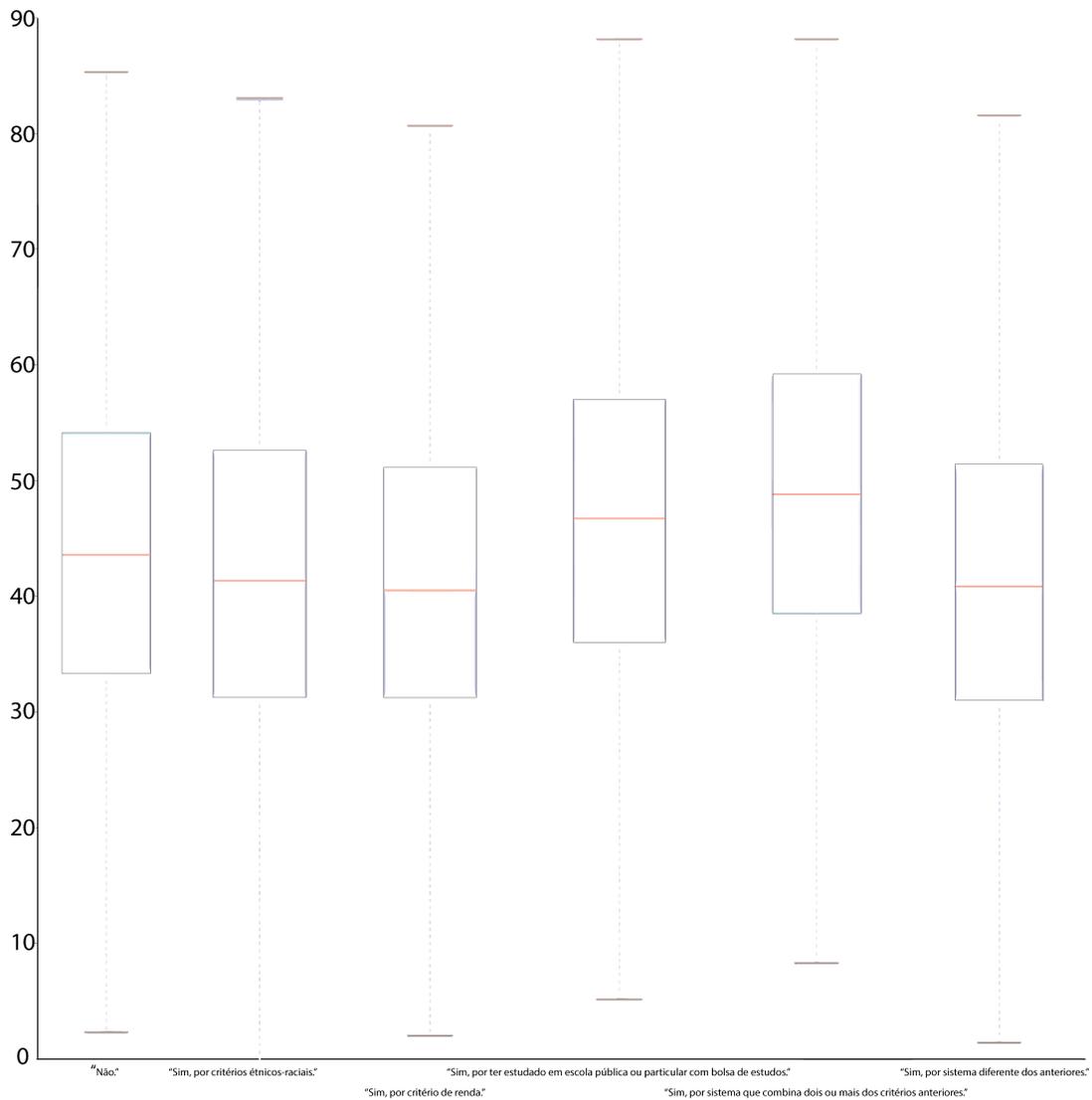


Figura 3.5: *Boxplot* ação afirmativa x nota ger para todos os alunos do ENADE.

A Figura 3.5 mostra a relação entre a questão 15 do questionário do estudante que diz respeito ao acesso a universidade por meio de ações afirmativas e a `nt_ger`, como é mostrado a seguir.

15. Seu ingresso no curso de graduação se deu por meio de políticas de ação afirmativa ou inclusão social?

- A () Não.
- B () Sim, por critério étnico-racial.
- C () Sim, por critério de renda.
- D () Sim, por ter estudado em escola pública ou particular com bolsa de estudos.
- E () Sim, por sistema que combina dois ou mais critérios anteriores.

F () Sim, por sistema diferente dos anteriores.

Pode-se notar que o *boxplot* da resposta “Sim, por sistema que combina dois ou mais das anteriores” está posicionado acima de todos os outros. Isso implica que todos os valores dos quartis deste *boxplot* (primeiro, segundo, terceiro) são maiores quando comparados com os respectivos valores dos quartis dos outros *boxplots* presentes na imagem.

Embora os *boxplots* sejam uma ferramenta gráfica útil como é possível reparar nas figuras colocadas como exemplo, eles não mostram de forma tão simples o valor exato dos quartis, sabendo disso, foram geradas tabelas para suprir essa falta.

Tabela 3.7: Tabela sexo x nota_geral para todos os alunos do ENADE.

nota_ger	"M"	"F"
Número de instâncias	165077	231680
Percentual de instâncias	34,27%	48,09%
Min	0	0
25 quartil	34,1	33,7
Mediana	44,1	43,9
75 quartil	54,6	54,7
Max	97,4	94,3

Agora com a Tabela 3.7 pode-se ver com mais exatidão que os números de fato são próximos independente do sexo confirmando o que foi observado previamente na Figura 3.3. Pode-se notar que a soma do percentual de instâncias não resulta em 100%, isso se deve ao fato de que nem todos os alunos inscritos de fato realizaram a prova ou possuem uma nota atribuída, como é explicado na Seção 3.4.

Tabela 3.8: Tabela da questão i22 quantidade livros x nota geral para todos os alunos do ENADE.

nota_ger	"Nenhum."	"Um ou dois."	"De três a cinco."	"De seis a oito."	"Mais de oito."
Número de instâncias	37571	129234	128330	43768	56574
Percentual de instâncias	7,80%	26,83%	26,64%	9,09%	11,74%
Min	0	0	0	0	0
25 quartil	33,1	33,2	34,3	34,1	34,9
Mediana	43,1	43,1	44,4	44,5	45,7
75 quartil	53,8	53,5	54,9	55,3	56,7
Max	94,7	93,6	97,4	91,8	94,3

Analisando a Tabela 3.8 novamente pode-se observar com detalhes o que foi analisado das imagens dos *boxplots* da Figura 3.4. As notas aumentam conforme o número de livros lidos também aumenta.

Tabela 3.9: Tabela ação afirmativa x nota ger para todos os alunos do ENADE.

nota_ger	"Não."	"Sim, por critério étnico-racial."	"Sim, por critério de renda."	Sim, por ter estudado em escola pública ou particular com bolsa de estudos."	"Sim, por sistema que combina dois ou mais critérios anteriores"	"Sim, por sistema diferente dos anteriores."
Número de instâncias	320758	6447	18006	25518	10218	14531
Percentual de instâncias	66,59%	1,34%	3,74%	5,30%	2,12%	3,02%
Min	0	0	0	0	0	0
25 quaril	33,8	31,7	31,7	36,5	39	31,4
Mediana	44	41,7	40,9	47,1	49,2	41,3
75 quaril	54,6	53	51,5	57,4	59,5	51,7
Max	97,4	92,2	88,9	91,5	91,9	89,7

A Tabela 3.9 confirma a análise visual da Figura 3.5 e podemos ver como os quartis da opção “Sim, por sistema que combina dois ou mais das anteriores” são, de fato, numericamente superiores aos outros.

Após analisar os dados em sua forma bruta, optou-se por fazer algumas relações das opções do atributo renda (sendo eles: "Ate 1,5 salario minimo (ate R\$ 1.086;00).", "De 1,5 a 3 salarios minimos (R\$ 1.086;01 a R\$ 2.172;00)", "De 3 a 4,5 salarios minimos (R\$ 2.172;01 a R\$ 3.258;00)", "De 4,5 a 6 salarios minimos (R\$ 3.258;01 a R\$ 4.344;00)", "De 6 a 10 salarios minimos (R\$ 4.344;01 a R\$ 7.240;00)", "De 10 a 30 salarios minimos (R\$ 7.240;01 a R\$ 21.720;00)", "Acima de 30 salarios minimos (mais de R\$ 21.720;01)) com as opções dos atributos: matutino, vespertino, ação afirmativa, raça e sexo. Essas relações foram comparados novamente com os três tipos de nota: nt_fg, nt_ce e nt_ger.

Um resultado interessante a ser observado foi: as Figuras 3.6 e 3.7 a seguir mostram o *boxplots* de cada opção do atributo renda relacionado com a opção “Branco” do atributo raça (usando nt_ger). A Tabela 3.10 mostra os valores correspondentes para os quartis.

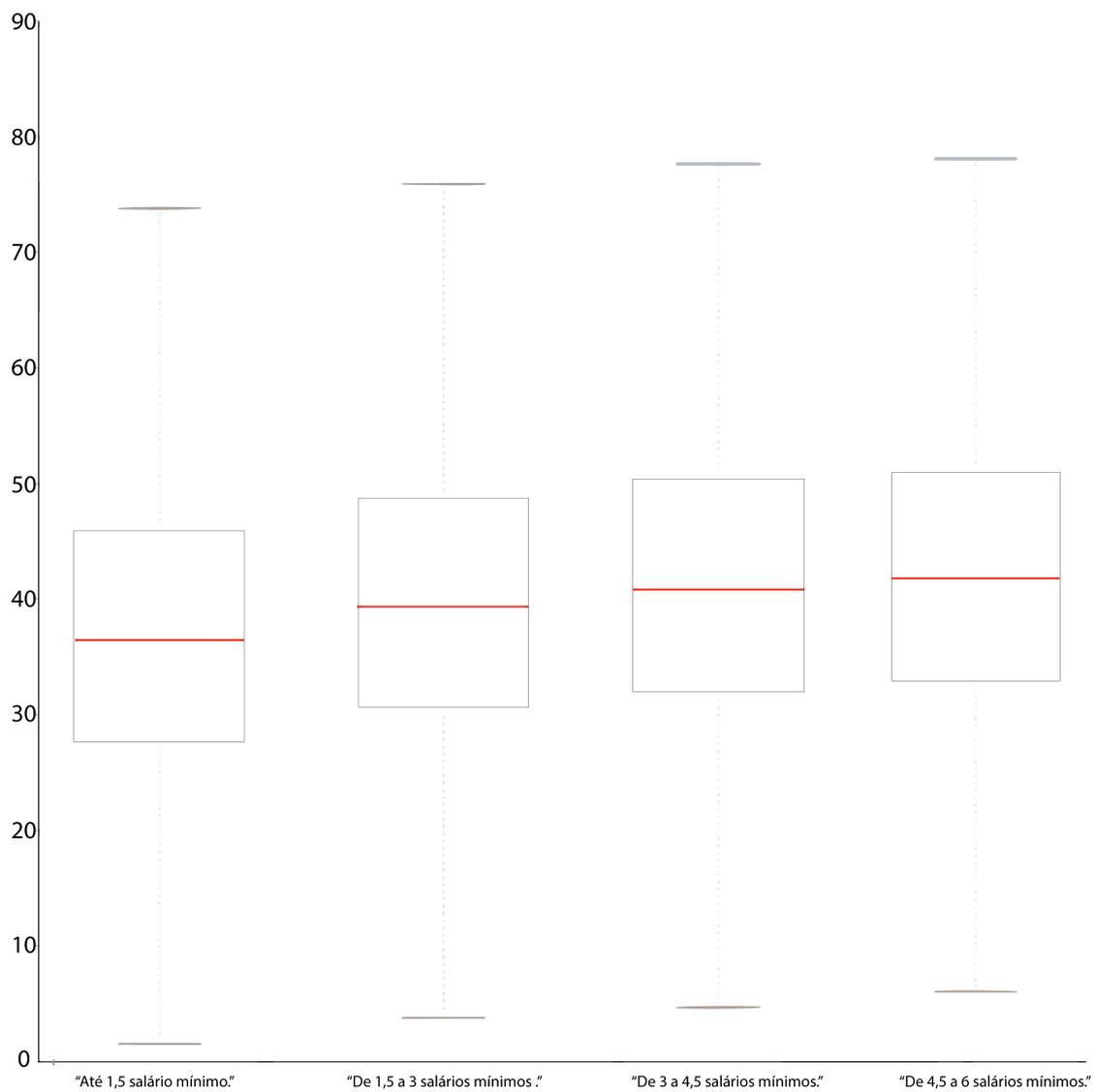


Figura 3.6: *Boxplot* renda família de etnia branca x nota geral para os alunos do ENADE (Parte 1).

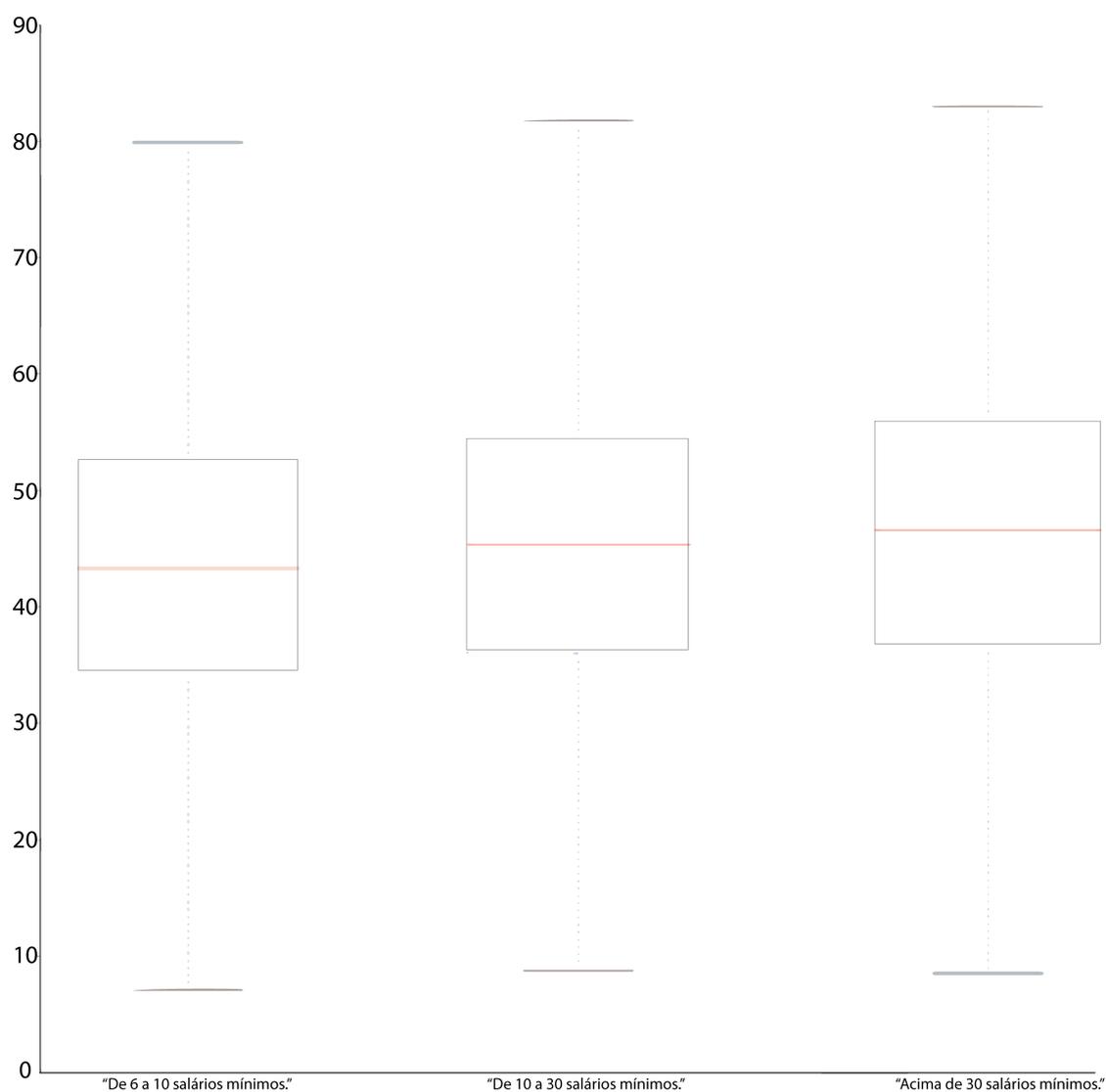


Figura 3.7: *Boxplot* renda família de etnia branca x nota geral para os alunos do ENADE (Parte 2).

Tabela 3.10: Tabela renda família de etnia branca x nota geral para os alunos do ENADE.

nota_ger	"Até 1,5 salário mínimo (até R\$1086,00)."	"De 1,5 a 3 salários mínimos (R\$1086,00 a R\$2172,00)."	"De 3 a 4,5 salários mínimos (R\$2172,01 a R\$3258,00)."	"De 4,5 a 6 salários mínimos (R\$3258,01 a R\$4344,00)."	"De 6 a 10 salários mínimos (R\$4344,01 a R\$7240,00)."	"De 10 a 30 salários mínimos (R\$7240,01 a R\$21720,00)."	"Acima de 30 salários mínimos (mais de R\$21720,01)."
Número de instâncias	26285	52236	42512	30388	33600	22295	3797
Percentual de instâncias com relação ao ENADE	5,46%	10,84%	8,83%	6,31%	6,98%	4,63%	0,79%
Min	0	0	0	0	0	0	0
25 quaril	30,8	34	35,6	36,7	38,1	40,1	40,9
Mediana	40,5	43,9	45,4	46,6	48,1	50,4	51,7
75 quaril	51,2	54,1	55,9	56,7	58,3	60,4	62,2
Max	87,9	93,6	90	91,6	96,1	97,4	95,3

As Figuras 3.8 e 3.9 e sua respectiva Tabela 3.11 mostram, de forma semelhante ao

parágrafo anterior, as opções do atributo renda relacionado com a opção “Negro” do atributo raça (usando `nt_ger`).

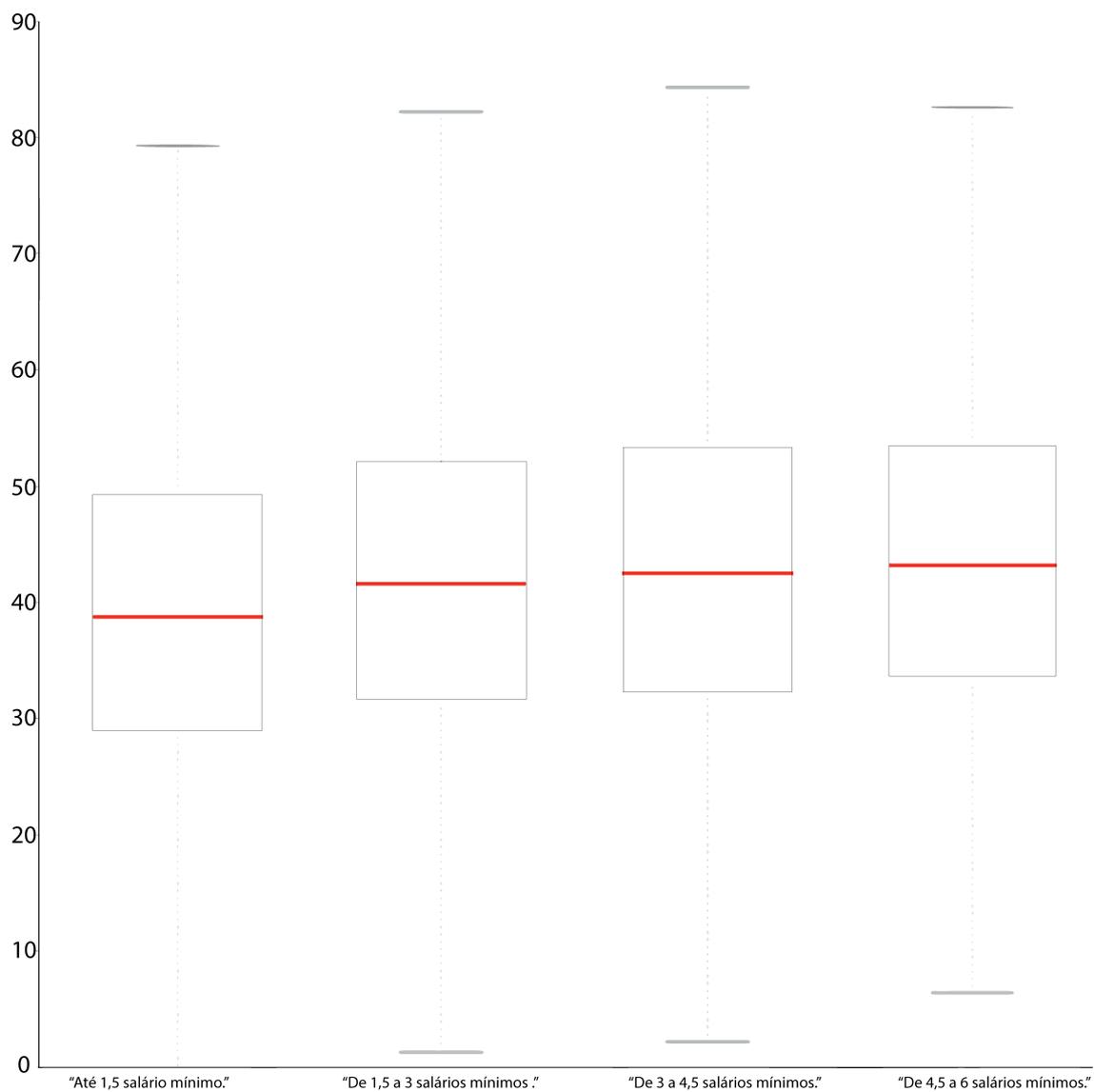


Figura 3.8: *Boxplot* renda família de etnia negra x nota geral para os alunos do ENADE (Parte 1).

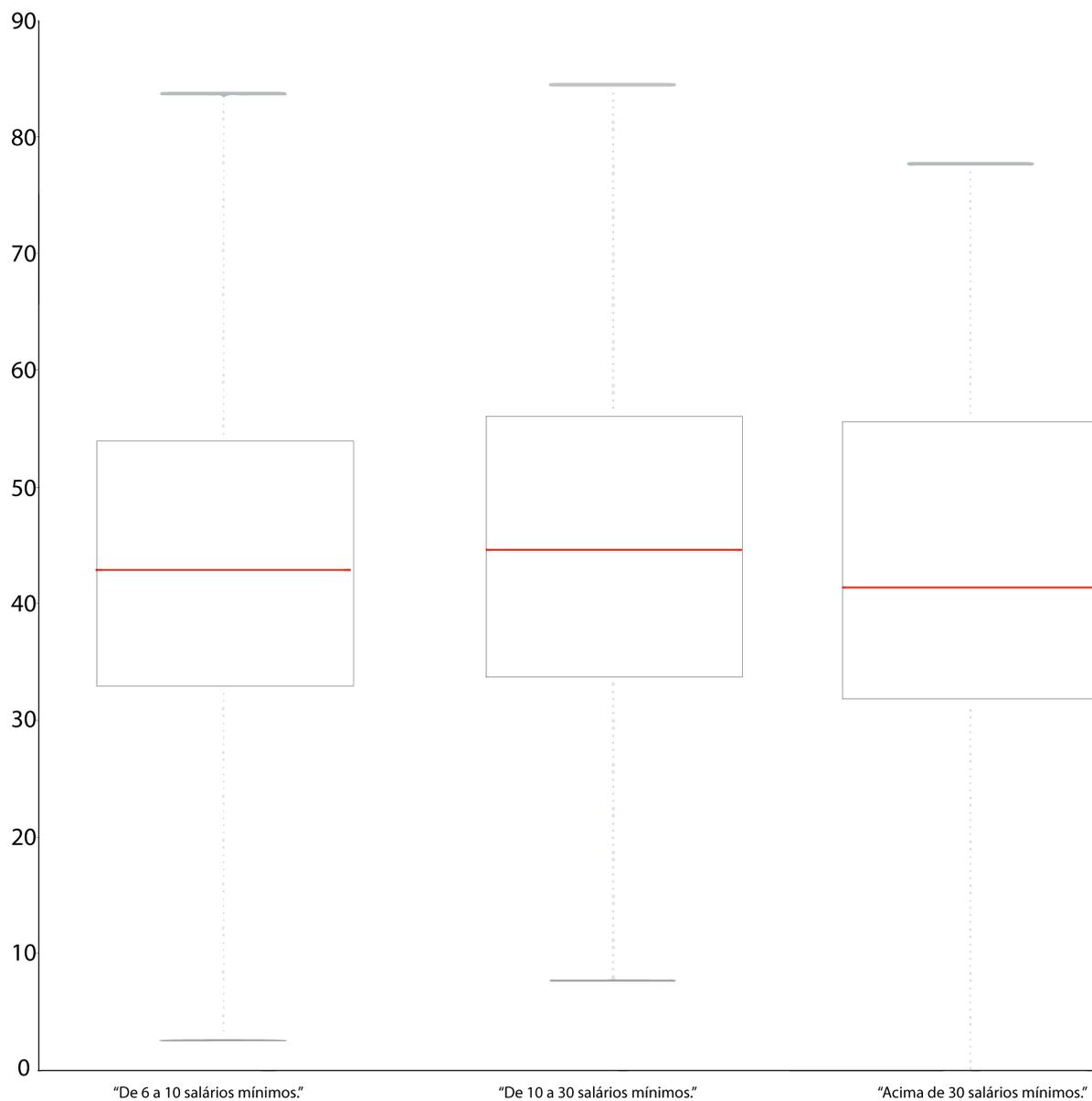


Figura 3.9: *Boxplot* renda família de etnia negra x nota geral para os alunos do ENADE (Parte 2).

Tabela 3.11: Tabela renda família de etnia negra x nota geral para os alunos do ENADE.

nota_ger	"Até 1,5 salário mínimo (até R\$1086,00)."	"De 1,5 a 3 salários mínimos (R\$1086,00 a R\$2172,00)."	"De 3 a 4,5 salários mínimos (R\$2172,01 a R\$3258,00)."	"De 4,5 a 6 salários mínimos (R\$3258,01 a R\$4344,00)."	"De 6 a 10 salários mínimos (R\$4344,01 a R\$7240,00)."	"De 10 a 30 salários mínimos (R\$7240,01 a R\$21720,00)."	"Acima de 30 salários mínimos (mais de R\$21720,01)."
Número de instâncias	11447	13417	6882	3584	2952	1125	89
Percentual de instâncias	2,38%	2,79%	1,43%	0,74%	0,61%	0,23%	0,02%
Min	0	0	0	0	0	0	0
25 quaril	29,7	32,2	33	34	33,6	34,4	32,2
Mediana	39,3	42,1	43	43,7	43,4	45,1	41,9
75 quaril	49,7	52,5	53,7	53,9	54,3	56,5	56
Max	92,2	87,5	91,1	87	91,1	84,9	78,1

Comparando as informações das Figuras 3.8 e 3.9 e das Figuras 3.6 e 3.7 é possível observar que para os alunos de etnia branca, conforme sua renda vai aumentando, a sua mediana tende a aumentar, enquanto para os alunos de etnia negra esse padrão de crescimento só é observado nos 4 primeiros *boxplots*. Cabe ressaltar que um dos *boxplots* que fugiu do padrão de crescimento, o *boxplot* referente aos alunos negros que tem renda familiar acima de 30 salários mínimos, representa um número de instâncias muito pequeno (0,02% de todos os alunos do ENADE).

Outra análise feita com relação a renda foi saber o perfil dos estudantes que estão na parcela com menor renda (ate 1.5 salários mínimos) e os estudantes que estão na parcela de maior renda (acima de 30 salários mínimos). Para isso foi criada uma tabela ordenando qual opção em cada pergunta foi marcada por cada aluno das parcelas citadas. As Tabelas 3.12 e 3.13 mostram os 10 atributos com as opções mais escolhidas dentre cada um dos grupos selecionados (maiores salários e menores salários).

Tabela 3.12: As 10 respostas com maior número de alunos com renda acima de 30 salários mínimos (com base em todos os alunos do ENADE).

Atributo	Numero de Instancias	Porcentagem sobre " acima de 30 salarios"	Porcentagem sobre o ENADE todo
i3_nacionalidade_"Brasileira,"	4897	98,20%	1,02%
i12_tipo_auxilio_"Nenhum,"	4876	97,77%	1,01%
i15_ingresso_por_politicas_acao_afirmativa_"Nao,"	4819	96,63%	1,00%
i21_familiares_tem_ensino_superior_"Sim,"	4791	96,07%	1,00%
i18_modalidade_ensino_medio_"Ensino medio tradicional,"	4530	90,84%	0,94%
i1_est_civil_"Solteiro(a),"	4454	89,31%	0,93%
i2_raca_etnia_"Branco(a),"	4029	80,79%	0,84%
i17_tipo_ensino_medio_"Todo em escola privada (particular),"	3985	79,91%	0,83%
i19_maior_incentivo_p_cursar_"Pais,"	3731	74,82%	0,78%
i14_atividades_no_exterior_"Nao participei,"	3660	73,39%	0,76%

Tabela 3.13: As 10 respostas com maior número de alunos com renda de até 1,5 salário mínimo (com base em todos os alunos do ENADE).

Atributo	Numero de Instancias	Porcentagem sobre "ate 1,5 salarios"	Porcentagem sobre o ENADE todo
i14_atividades_no_exterior_"Nao participei,"	76869	98,36%	15,96%
i3_nacionalidade_"Brasileira,"	76644	98,07%	15,91%
i17_tipo_ensino_medio_"Todo em escola publica,"	68420	87,55%	14,20%
i12_tipo_auxilio_"Nenhum,"	65436	83,73%	13,58%
in_vespertino_"0"	61771	79,04%	12,82%
i13_tipo_bolsa_academica_"Nenhum,"	59549	76,19%	12,36%
i24_oportunidade_aprendizado_idioma_extrangeiro_"Nao,"	58645	75,04%	12,17%
i15_ingresso_por_politicas_acao_afirmativa_"Nao,"	58415	74,74%	12,13%
tp_sexo_"F"	57622	73,73%	11,96%
i33_curso_possibilitou_aumentar_capacidade_reflexao_argumentacao_"Concordo Totalmente"	56707	72,56%	11,77%

Um total de 4987 alunos no Enade 2014 marcaram “renda superior a 30 salários mínimos” e 78154 marcaram “renda inferior a 1,5 salários mínimos”. A primeira coluna da tabela mostra o título da questão (atributo) juntamente com a opção mais escolhida. A segunda coluna diz quantos alunos (dentro desse perfil determinado) marcaram essa questão/opção. A terceira indica qual a porcentagem que esse número de instâncias

representa sobre o total de alunos desse perfil. A quarta indica a porcentagem sobre o total de alunos do ENADE que esse número representa.

Por exemplo, na Tabela 3.12, pode-se observar na primeira linha que 4897 alunos (que representa 98,2% do total de alunos que possuem renda familiar superior a 30 salários mínimos ou 1,02% do total de alunos do ENADE) que possuem renda familiar superior a 30 salários mínimos marcaram que sua nacionalidade é brasileira. Outro exemplo é na Tabela 3.13, onde pode-se observar na penúltima linha que 57622 alunos (que representa 73,73% do total de alunos que possuem renda familiar de até 1,5 salários mínimos ou 11,96% do total de alunos do ENADE) que possuem renda inferior a 1,5 salários mínimos marcaram que são do sexo feminino.

Essas 10 opções mais marcadas permitem observar algumas informações interessantes como o fato de que 89,31% das pessoas com renda superior a 30 salário mínima são solteiras e o fato de que 73,73% das pessoas com renda inferior a 1,5 salários mínimos são mulheres. O que chamou mais atenção foi o fato de que as questões i14 (sobre a participação de atividades no exterior), i12 (Sobre o tipo de auxílio), i15 (ingresso por políticas de ação afirmativa), e i3 (nacionalidade) aparecem nas 10 respostas com o maior número de alunos de ambas tabelas (renda máxima e mínima) e apresentam a mesma opção marcada. O que é relevante pois independente de as rendas serem muito divergentes, em ambos os grupos existe uma maioria que não utilizou nenhuma política de ação afirmativa ou teve algum tipo de auxílio, dentre todas as opções que haviam na questão, sendo elas:

13. Ao longo da sua trajetória acadêmica, você recebeu algum tipo de bolsa acadêmica? No caso de haver mais de uma opção, marcar apenas a bolsa de maior duração.

- A Nenhum.
- B Bolsa de iniciação científica.
- C Bolsa de extensão.
- D Bolsa de monitoria/tutoria.
- E Bolsa PET.
- F Outro tipo de bolsa acadêmica.

Já o atributo i17, que é referente à questão:

17. Em que tipo de escola você cursou o ensino médio?

- A Todo em escola pública.
- B Todo em escola privada (particular).
- C Todo no exterior.
- D A maior parte em escola pública.

E () A maior parte em escola privada (particular).

F () Parte no Brasil e parte no exterior.

também aparece nas 10 opções mais marcadas de ambos porém no grupo de renda superior a 30 salários mínimos a opção marcada foi “Toda em escola privada (particular)” e no grupo de “renda inferior a 1,5 salários mínimos” a opção “Todo em escola pública” foi a opção mais marcada.

Como explicitado anteriormente no Capítulo 3, o ENADE do ano de 2014 foi escolhido por ter cursos ligados a computação pois um dos objetivos desse trabalho é justamente examinar o perfil desses alunos. Todas as tabelas e *boxplots* que foram criados para o conjunto completo de dados do Enade também foram criados apenas para os alunos dos cursos: Engenharia da Computação, Ciência da Computação (Bacharelado) e Ciência da Computação (Licenciatura) separadamente. O fato de o curso de Ciência da Computação (bacharelado) ter mais representantes no Enade (quando comparado com Engenharia da Computação, Ciência da Computação (Licenciatura)) torna-o o curso de maior interesse para o estudo, por isso, uma parte maior das análises foram feitas direcionadas para este curso.

Análises dos *boxplots* e tabelas referentes à questão 15 para o conjunto de dados de todos os alunos do Enade já foram feitos como mostram as figuras, agora observando esses mesmos recursos porém gerados com base apenas nos dados dos alunos de ciência da Computação(bacharelado) encontramos um informação um pouco diferente.

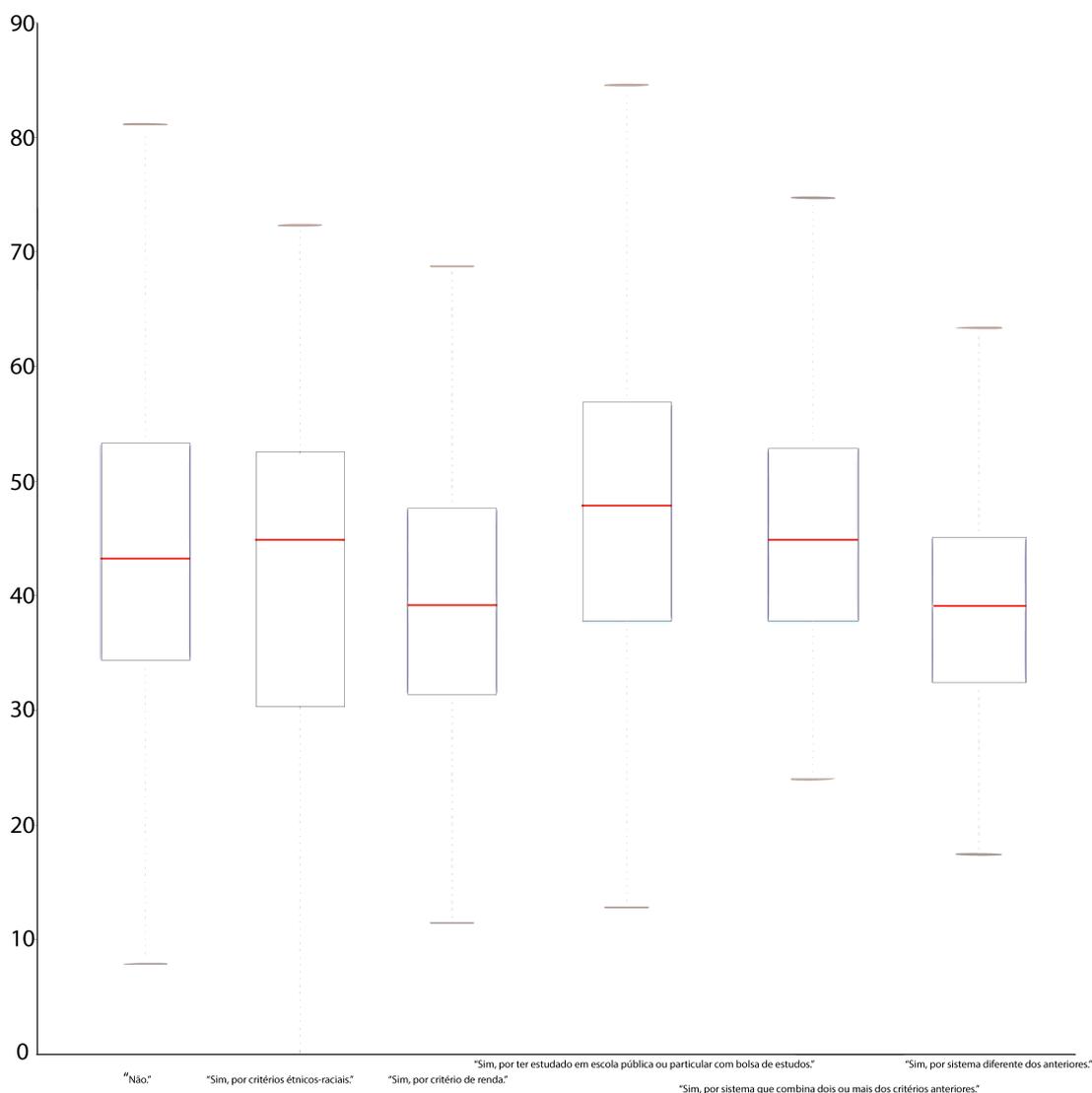


Figura 3.10: *Boxplot* específico do curso de CIC ação afirmativa x nota ger.

Tabela 3.14: Tabela específica do curso de CIC ação afirmativa x nota ger.

nota_ger	"Não."	"Sim, por critério étnico-racial."	"Sim, por critério de renda."	Sim, por ter estudado em escola pública ou particular com bolsa de estudos."	"Sim, por sistema que combina dois ou mais critérios anteriores"	"Sim, por sistema diferente dos anteriores."
Número de instâncias	6981	71	285	609	191	127
Percentual de instâncias	71,08%	0,72%	2,90%	6,20%	1,95%	1,29%
Min	0	0	11,6	0	13,3	6,8
25 quaril	34,7	30,6	31,7	38,1	38,1	32,65
Mediana	43,5	45,1	39,4	48,1	45,2	39,3
75 quaril	53,5	52,75	47,8	57,1	35,15	45,25
Max	96,1	72,7	74,8	85	75,2	75,3

Comparando os *boxplots* de ação afirmativa do enade geral 3.5 e o específico dos alunos de Ciência da Computação 3.10 e tabelas dos alunos do enade geral 3.9 e dos alunos de

CIC especificamente 3.14 podemos ver que os dados não estão dispostos nos mesmos padrões. A maior mediana não é mais de alunos que responderam “Sim, por sistema que combina dois ou mais das anteriores” e sim os que responderam “Sim, por ter estudado em escola pública ou particular com bolsa de estudos.” por exemplo. Pode-se observar que alunos cotistas tanto por critério étnico racial como por critério de ter estudado em escola pública ou particular com bolsa, possuem mediana maior do que alunos não cotistas.

3.7 Análise do desempenho dos estudantes em cada questão - Universidade de Brasília - Ciência da Computação (Bacharel)

A prova do ENADE é composta por 40 questões divididas em duas partes:

- Formação Geral (FG): composta de 10 questões, sendo 8 de múltipla escolha e 2 discursivas.
- Componentes específicos (CE): composta de 30 questões, sendo 27 de múltipla escolha e 3 discursivas.

O arquivo extraído na Seção 3.4 contem 4 campos que dizem respeito a essas 35 questões de múltipla escolha:

- "vt_esc_ofg": vetor de 8 caracteres onde cada carácter equivale a escolha do aluno na parte de formação geral. A ordem dos caracteres segue a ordem das questões da prova, por exemplo: o primeiro carácter do vetor equivale a opção marcada pelo aluno na questão 1 da parte de formação geral.
- "vt_ace_ofg": vetor de 8 caracteres onde cada carácter representa se o estudante acertou ou não o item na parte de formação geral. A ordem dos caracteres segue a ordem das questões da prova, por exemplo: o primeiro carácter do vetor mostra se o estudante acertou ou não a questão 1 da parte de formação geral.
- "vt_esc_oce": vetor de 27 caracteres onde cada carácter equivale a escolha do aluno na parte de componentes específicos. A ordem dos caracteres segue a ordem das questões da prova, por exemplo: o primeiro carácter do vetor equivale a opção marcada pelo aluno na questão 1 da parte de componentes específicos (equivalente a questão 9 da prova do ENADE).
- "vt_ace_oce": vetor de 27 caracteres onde cada carácter representa se o estudante acertou ou não o item na parte de componentes específicos. A ordem dos caracteres

segue a ordem das questões da prova, por exemplo: o primeiro carácter do vetor mostra se o estudante acertou ou não a questão 1 da parte de componente específico (equivalente a questão 9 da prova do ENADE).

Outro passo julgado interessante a ser tomado foi a análise do desempenho dos estudantes de Ciência da Computação (bacharel) da Universidade de Brasília (UnB) em cada uma das questões da prova na parte de Componente Específico (CE) e Formação Geral (FG) e a comparação destes desempenhos com o desempenho de estudantes das 10 universidades melhor classificadas (A nota final do ENADE não é calculada levando em consideração somente o desempenho dos estudantes.[22]) no ENADE 2014 com relação ao curso de Ciência da Computação [23]. A UnB não está entre essas 10 melhores, porem ela foi acrescentada na tabela também para comparação de resultados. Para isso, foram separadas em uma tabela essas instituições que participaram da análise, e foram recolhidas informações como o número de estudantes de CiC que realizaram a prova em cada uma delas, como mostra a Tabela 3.15.

Sabendo das instituições a serem avaliadas, foi criado um algoritmo que separava todos os alunos de Ciência da Computação que fossem dessas instituições. Com isso, o algoritmo coletava os dados referentes aos vetores de acertos de respostas "vt_ace_oce" e "vt_ace_ofg" (citados no início desta seção), e os organizava em suas respectivas questões. No final, foi calculado o percentual de acerto das questões dos estudantes de cada instituição, como pode-se observar nas Tabelas 3.16, 3.17, 3.18, 3.19, 3.20 . Essas tabelas estão ordenadas pela nota da instituição no ENADE para o curso de Ciência da Computação (bacharel).

Tabela 3.15: Tabela que mostra o numero de estudantes que realizaram a prova em cada uma das instituições utilizadas na análise.

Nome da Instituição	Número de Instâncias
UNIVERSIDADE DO OESTE PAULISTA	14
UNIVERSIDADE FEDERAL DO AMAZONAS	23
UNIVERSIDADE ESTADUAL DO CENTRO OESTE	15
UNIVERSIDADE FEDERAL DE SANTA CATARINA	46
UNIVERSIDADE FEDERAL DE ALAGOAS	113
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE	43
UNIVERSIDADE PAULISTA	302
UNIVERSIDADE FEDERAL DE SAO JOAO DEL REI	26
UNIVERSIDADE FEDERAL DA BAHIA	12
CENTRO UNIVERSITARIO DO INSTITUTO DE EDUCACAO SUPERIOR DE BRASILIA - IESB	10
UNIVERSIDADE FEDERAL DE MINAS GERAIS	66
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL	130
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO	92
UNIVERSIDADE ESTADUAL DE CAMPINAS	42
UNIVERSIDADE FEDERAL DE PERNAMBUCO	72
UNIVERSIDADE FEDERAL DE SANTA CATARINA	46
UNIVERSIDADE ESTADUAL PAULISTA JULIO DE MESQUITA FILHO	155
UNIVERSIDADE FEDERAL DA PARAIBA	39
UNIVERSIDADE FEDERAL DE SAO CARLOS	87
UNIVERSIDADE DE BRASILIA	24

Tabela 3.16: Tabela comparando o percentual de acertos dos estudantes de Ciência da Computação (bacharel) da UnB nas questões de 1 a 6.

Nome da Instituição	Q1_geral	Q2_geral	Q3_geral	Q4_geral	Q5_geral	Q6_geral
UNIVERSIDADE DO OESTE PAULISTA	71,43%	71,43%	100,00%	100,00%	50,00%	71,43%
UNIVERSIDADE FEDERAL DO AMAZONAS	69,57%	56,52%	91,30%	95,65%	43,48%	60,87%
UNIVERSIDADE ESTADUAL DO CENTRO OESTE	66,67%	40,00%	80,00%	100,00%	33,33%	53,33%
UNIVERSIDADE FEDERAL DE SANTA CATARINA	56,52%	50,00%	97,83%	95,65%	52,17%	69,57%
UNIVERSIDADE FEDERAL DE ALAGOAS	47,79%	38,94%	76,99%	88,50%	33,63%	63,72%
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE	46,51%	51,16%	86,05%	93,02%	39,54%	65,12%
UNIVERSIDADE PAULISTA	57,95%	48,01%	73,51%	87,42%	42,38%	60,27%
UNIVERSIDADE FEDERAL DE SAO JOAO DEL REI	53,85%	38,46%	80,77%	80,77%	34,62%	61,54%
UNIVERSIDADE FEDERAL DA BAHIA	58,33%	66,67%	91,67%	91,67%	33,33%	75,00%
CENTRO UNIVERSITARIO DO INSTITUTO DE EDUCACAO SUPERIOR DE BRASILIA - IESB	70,00%	50,00%	80,00%	100,00%	30,00%	60,00%
UNIVERSIDADE FEDERAL DE MINAS GERAIS	71,21%	42,42%	87,88%	95,46%	54,55%	75,76%
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL	65,39%	50,00%	92,31%	93,08%	45,39%	76,15%
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO	64,13%	48,91%	91,30%	90,22%	44,57%	68,48%
UNIVERSIDADE ESTADUAL DE CAMPINAS	59,52%	47,62%	80,95%	85,71%	42,86%	76,19%
UNIVERSIDADE FEDERAL DE PERNAMBUCO	55,56%	58,33%	91,67%	94,44%	56,94%	75,00%
UNIVERSIDADE FEDERAL DE SANTA CATARINA	56,52%	50,00%	97,83%	95,65%	52,17%	69,57%
UNIVERSIDADE ESTADUAL PAULISTA JULIO DE MESQUITA FILHO	60,00%	51,61%	84,52%	90,32%	43,23%	72,26%
UNIVERSIDADE FEDERAL DA PARAIBA	69,23%	43,59%	84,62%	97,44%	64,10%	74,36%
UNIVERSIDADE FEDERAL DE SAO CARLOS	62,07%	56,32%	88,51%	91,95%	49,43%	67,82%
UNIVERSIDADE DE BRASILIA	62,50%	45,83%	75,00%	79,17%	50,00%	75,00%

Tabela 3.17: Tabela comparando o percentual de acertos dos estudantes de Ciência da Computação (bacharel) da UnB nas questões de 7 a 13.

Nome da Instituição	Q7_geral	Q8_geral	Q9_especifico	Q10_especifico	Q12_especifico	Q13_especifico
UNIVERSIDADE DO OESTE PAULISTA	50,00%	100,00%	78,57%	78,57%	50,00%	57,14%
UNIVERSIDADE FEDERAL DO AMAZONAS	56,52%	86,96%	78,26%	78,26%	43,48%	60,87%
UNIVERSIDADE ESTADUAL DO CENTRO OESTE	60,00%	86,67%	53,33%	66,67%	60,00%	40,00%
UNIVERSIDADE FEDERAL DE SANTA CATARINA	73,91%	84,78%	82,61%	82,61%	50,00%	65,22%
UNIVERSIDADE FEDERAL DE ALAGOAS	55,75%	78,76%	67,26%	66,37%	28,32%	30,09%
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE	67,44%	90,70%	93,02%	76,74%	44,19%	55,81%
UNIVERSIDADE PAULISTA	60,60%	84,11%	70,53%	66,89%	26,49%	25,83%
UNIVERSIDADE FEDERAL DE SAO JOAO DEL REI	80,77%	92,31%	76,92%	76,92%	26,92%	57,69%
UNIVERSIDADE FEDERAL DA BAHIA	100,00%	100,00%	75,00%	91,67%	33,33%	50,00%
CENTRO UNIVERSITARIO DO INSTITUTO DE EDUCACAO SUPERIOR DE BRASILIA - IESB	80,00%	100,00%	90,00%	80,00%	10,00%	50,00%
UNIVERSIDADE FEDERAL DE MINAS GERAIS	74,24%	93,94%	86,36%	80,30%	43,94%	45,46%
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL	76,15%	92,31%	82,31%	73,85%	37,69%	45,39%
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO	76,09%	77,17%	83,70%	72,83%	34,78%	73,91%
UNIVERSIDADE ESTADUAL DE CAMPINAS	76,19%	83,33%	78,57%	66,67%	42,86%	54,76%
UNIVERSIDADE FEDERAL DE PERNAMBUCO	81,94%	95,83%	83,33%	88,89%	33,33%	55,56%
UNIVERSIDADE FEDERAL DE SANTA CATARINA	73,91%	84,78%	82,61%	82,61%	50,00%	65,22%
UNIVERSIDADE ESTADUAL PAULISTA JULIO DE MESQUITA FILHO	69,68%	82,58%	82,58%	72,90%	29,68%	35,48%
UNIVERSIDADE FEDERAL DA PARAIBA	74,36%	87,18%	87,18%	82,05%	48,72%	33,33%
UNIVERSIDADE FEDERAL DE SAO CARLOS	78,16%	86,21%	82,76%	78,16%	44,83%	52,87%
UNIVERSIDADE DE BRASILIA	83,33%	83,33%	70,83%	91,67%	33,33%	58,33%

Tabela 3.18: Tabela comparando o percentual de acertos dos estudantes de Ciência da Computação (bacharel) da UnB nas questões de 14 a 20.

Nome da Instituição	Q14_especifico	Q15_especifico	Q16_especifico	Q17_especifico	Q19_especifico	Q20_especifico
UNIVERSIDADE DO OESTE PAULISTA	35,71%	50,00%	21,43%	78,57%	35,71%	85,71%
UNIVERSIDADE FEDERAL DO AMAZONAS	47,83%	43,48%	56,52%	86,96%	4,35%	65,22%
UNIVERSIDADE ESTADUAL DO CENTRO OESTE	53,33%	46,67%	60,00%	86,67%	20,00%	80,00%
UNIVERSIDADE FEDERAL DE SANTA CATARINA	45,65%	78,26%	69,57%	84,78%	36,96%	93,48%
UNIVERSIDADE FEDERAL DE ALAGOAS	36,28%	35,40%	23,01%	66,37%	29,20%	59,29%
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE	62,79%	46,51%	53,49%	90,70%	32,56%	83,72%
UNIVERSIDADE PAULISTA	23,51%	31,46%	31,79%	65,23%	6,29%	68,87%
UNIVERSIDADE FEDERAL DE SAO JOAO DEL REI	46,15%	57,69%	57,69%	76,92%	34,62%	76,92%
UNIVERSIDADE FEDERAL DA BAHIA	75,00%	83,33%	50,00%	100,00%	50,00%	100,00%
CENTRO UNIVERSITARIO DO INSTITUTO DE EDUCACAO SUPERIOR DE BRASILIA - IESB	30,00%	50,00%	80,00%	100,00%	20,00%	100,00%
UNIVERSIDADE FEDERAL DE MINAS GERAIS	56,06%	66,67%	77,27%	96,97%	25,76%	86,36%
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL	63,08%	46,15%	66,15%	93,85%	27,69%	91,54%
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO	53,26%	43,48%	54,35%	89,13%	40,22%	90,22%
UNIVERSIDADE ESTADUAL DE CAMPINAS	61,91%	42,86%	73,81%	80,95%	16,67%	73,81%
UNIVERSIDADE FEDERAL DE PERNAMBUCO	62,50%	75,00%	62,50%	98,61%	54,17%	80,56%
UNIVERSIDADE FEDERAL DE SANTA CATARINA	45,65%	78,26%	69,57%	84,78%	36,96%	93,48%
UNIVERSIDADE ESTADUAL PAULISTA JULIO DE MESQUITA FILHO	42,58%	35,48%	40,65%	77,42%	23,23%	67,10%
UNIVERSIDADE FEDERAL DA PARAIBA	58,97%	51,28%	66,67%	94,87%	46,15%	87,18%
UNIVERSIDADE FEDERAL DE SAO CARLOS	49,43%	34,48%	63,22%	85,06%	16,09%	85,06%
UNIVERSIDADE DE BRASILIA	70,83%	95,83%	58,33%	95,83%	16,67%	70,83%

Tabela 3.19: Tabela comparando o percentual de acertos dos estudantes de Ciência da Computação (bacharel) da UnB nas questões de 22 a 27.

Nome da Instituição	Q22_especifico	Q23_especifico	Q24_especifico	Q25_especifico	Q26_especifico	Q27_especifico
UNIVERSIDADE DO OESTE PAULISTA	21,43%	57,14%	92,86%	92,86%	28,57%	78,57%
UNIVERSIDADE FEDERAL DO AMAZONAS	56,52%	43,48%	95,65%	91,30%	47,83%	82,61%
UNIVERSIDADE ESTADUAL DO CENTRO OESTE	40,00%	13,33%	73,33%	80,00%	33,33%	80,00%
UNIVERSIDADE FEDERAL DE SANTA CATARINA	21,74%	43,48%	91,30%	84,78%	41,30%	89,13%
UNIVERSIDADE FEDERAL DE ALAGOAS	33,63%	26,55%	68,14%	46,90%	35,40%	75,22%
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE	39,54%	41,86%	86,05%	67,44%	30,23%	90,70%
UNIVERSIDADE PAULISTA	23,84%	11,26%	75,17%	53,97%	29,47%	62,91%
UNIVERSIDADE FEDERAL DE SAO JOAO DEL REI	34,62%	34,62%	96,15%	57,69%	38,46%	92,31%
UNIVERSIDADE FEDERAL DA BAHIA	25,00%	58,33%	91,67%	75,00%	50,00%	75,00%
CENTRO UNIVERSITARIO DO INSTITUTO DE EDUCACAO SUPERIOR DE BRASILIA - IESB	10,00%	40,00%	90,00%	70,00%	10,00%	100,00%
UNIVERSIDADE FEDERAL DE MINAS GERAIS	56,06%	42,42%	90,91%	83,33%	56,06%	81,82%
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL	40,00%	49,23%	86,92%	81,54%	35,39%	81,54%
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO	25,00%	35,87%	90,22%	77,17%	35,87%	82,61%
UNIVERSIDADE ESTADUAL DE CAMPINAS	35,71%	38,10%	85,71%	71,43%	30,95%	76,19%
UNIVERSIDADE FEDERAL DE PERNAMBUCO	26,39%	56,94%	94,44%	84,72%	54,17%	90,28%
UNIVERSIDADE FEDERAL DE SANTA CATARINA	21,74%	43,48%	91,30%	84,78%	41,30%	89,13%
UNIVERSIDADE ESTADUAL PAULISTA JULIO DE MESQUITA FILHO	19,36%	27,74%	75,48%	60,65%	21,94%	71,61%
UNIVERSIDADE FEDERAL DA PARAIBA	20,51%	25,64%	84,62%	71,80%	30,77%	79,49%
UNIVERSIDADE FEDERAL DE SAO CARLOS	31,03%	36,78%	90,81%	75,86%	36,78%	73,56%
UNIVERSIDADE DE BRASILIA	33,33%	33,33%	91,67%	83,33%	33,33%	83,33%

Tabela 3.20: Tabela comparando o percentual de acertos dos estudantes de Ciência da Computação (bacharel) da UnB nas questões de 28 a 35.

Nome da Instituição	Q28_especifico	Q29_especifico	Q30_especifico	Q32_especifico	Q35_especifico
UNIVERSIDADE DO OESTE PAULISTA	28,57%	85,71%	28,57%	64,29%	28,57%
UNIVERSIDADE FEDERAL DO AMAZONAS	69,57%	56,52%	30,44%	65,22%	30,44%
UNIVERSIDADE ESTADUAL DO CENTRO OESTE	53,33%	66,67%	40,00%	46,67%	60,00%
UNIVERSIDADE FEDERAL DE SANTA CATARINA	63,04%	65,22%	41,30%	69,57%	30,44%
UNIVERSIDADE FEDERAL DE ALAGOAS	41,59%	47,79%	19,47%	53,98%	22,12%
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE	55,81%	69,77%	44,19%	51,16%	18,61%
UNIVERSIDADE PAULISTA	32,78%	42,72%	12,58%	46,69%	23,18%
UNIVERSIDADE FEDERAL DE SAO JOAO DEL REI	69,23%	69,23%	42,31%	69,23%	34,62%
UNIVERSIDADE FEDERAL DA BAHIA	58,33%	83,33%	8,33%	75,00%	41,67%
CENTRO UNIVERSITARIO DO INSTITUTO DE EDUCACAO SUPERIOR DE BRASILIA - IESB	40,00%	50,00%	10,00%	70,00%	20,00%
UNIVERSIDADE FEDERAL DE MINAS GERAIS	68,18%	66,67%	45,46%	50,00%	33,33%
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL	61,54%	65,39%	32,31%	66,15%	36,15%
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO	42,39%	64,13%	23,91%	59,78%	28,26%
UNIVERSIDADE ESTADUAL DE CAMPINAS	66,67%	66,67%	40,48%	64,29%	38,10%
UNIVERSIDADE FEDERAL DE PERNAMBUCO	66,67%	63,89%	22,22%	62,50%	30,56%
UNIVERSIDADE FEDERAL DE SANTA CATARINA	63,04%	65,22%	41,30%	69,57%	30,44%
UNIVERSIDADE ESTADUAL PAULISTA JULIO DE MESQUITA FILHO	20,65%	52,26%	14,19%	47,74%	30,32%
UNIVERSIDADE FEDERAL DA PARAIBA	64,10%	56,41%	28,21%	61,54%	15,39%
UNIVERSIDADE FEDERAL DE SAO CARLOS	56,32%	56,32%	20,69%	59,77%	22,99%
UNIVERSIDADE DE BRASILIA	45,83%	58,33%	4,17%	70,83%	25,00%

Conforme mencionado anteriormente, a nota do ENADE de um curso em uma determinada instituição não depende somente do desempenho de seus alunos. A Tabela 3.21 que mostra as 10 instituições com as maiores notas em Ciência da Computação (bacharel) no ENADE (com a adição da UnB), sua posição no ranking geral e está ordenada da maior média na nota do componente específico (nt_ce) para a menor. É possível ver que Universidade de Brasília que no ranking geral está decimo nono, quando comparada com relação à média da nt_ce, na verdade está na oitava posição.

Tabela 3.21: Tabela mostrando a media da nt_ce das 10 intuições com as maiores notas em Ciência da Computação (bacharel) no ENADE (com a adição da UnB).

Posição no ranking	Nome da instituição	nt_ce
9º	UNIVERSIDADE FEDERAL DA BAHIA	59,96
4º	UNIVERSIDADE FEDERAL DE SANTA CATARINA	58,22
6º	UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE	55,49
2º	UNIVERSIDADE FEDERAL DO AMAZONAS	55,47
8º	UNIVERSIDADE FEDERAL DE SAO JOAO DEL REI	54,99
3º	UNIVERSIDADE ESTADUAL DO CENTRO OESTE	53,91
1º	UNIVERSIDADE DO OESTE PAULISTA	53,76
19º	UNIVERSIDADE DE BRASILIA	53,55
10º	CENTRO UNIVERSITARIO DO INSTITUTO DE EDUCACAO SUPERIOR DE BRASILIA - IESB	48,39
5º	UNIVERSIDADE FEDERAL DE ALAGOAS	39,85
7º	UNIVERSIDADE PAULISTA	36,16

As questões 11, 18, 21, 31, 33 e 34 foram anuladas e por isso não aparecem nas tabelas. Como é possível ver na Tabela 3.18 a questão em que a UnB apresentou desempenho superior a todos as outras universidades da análise foi a questão 15 (sétima questão do componente específico).

QUESTÃO 15

Considere as seguintes expressões regulares cujo alfabeto é $\{a, b\}$.

$$R1 = a(a \cup b)^*$$

$$R2 = b(a \cup b)^*$$

Se $L(R)$ é a linguagem associada a uma expressão regular R , é correto afirmar que

- A** $L(R1) = L(R2)$.
- B** $L(R2) = \{w \mid w \text{ termina com } b\}$.
- C** existe um autômato finito determinístico cuja linguagem é igual a $L(R1) \cup L(R2)$.
- D** se $R3$ é uma expressão regular tal que $L(R3) = L(R1) \cap L(R2)$, então $L(R3)$ é uma linguagem infinita.
- E** um autômato finito não determinístico que reconheça $L(R1) \cup L(R2)$ tem, pelo menos, quatro estados.

Figura 3.11: Questão 15 do ENADE 2014.

Em uma análise mais detalhada foi constatado que 95,833% dos alunos da UnB marcou a opção 'C' e os 4,167% restantes deixaram a questão em branco.

Em contraste a esse cenário, a questão 30 (vigésima segunda questão do componente específico) foi a questão em que a UnB obteve o pior desempenho.

QUESTÃO 30

Uma fazenda possui um único poço artesiano que deve abastecer n bebedouros para o gado. Deseja-se determinar um projeto de ligação entre esses $n+1$ pontos através de encanamentos com a menor extensão total. Um algoritmo proposto para a solução do problema executa os seguintes passos:

1. Crie $n+1$ conjuntos unitários, cada um contendo um dos pontos a serem ligados entre si e insira esses conjuntos em um conjunto C .
2. Crie um conjunto D contendo um registro para cada combinação possível de dois pontos distintos a serem ligados. Cada registro deve conter os campos c_i , c_j e d , em que c_i e c_j são os dois pontos a serem ligados e d é a distância entre eles.
3. Enquanto D não estiver vazio faça:
 - 3.1. Remova o registro de D com o menor valor de distância d .
 - 3.2. Se os valores de c_i e c_j do registro removido pertencerem a conjuntos distintos de C , então:
 - 3.2.1. Substitua estes dois conjuntos pela união entre eles.
 - 3.2.2. Guarde o registro removido em um conjunto-solução.

Com base na descrição do problema e do algoritmo proposto, conclui-se que

- A** o problema exemplifica a obtenção de uma árvore geradora mínima, portanto está no conjunto P .
- B** o algoritmo é uma heurística para o Problema do Caixeiro Viajante, logo apresenta complexidade polinomial.
- C** o problema descrito é de otimização, logo pertence ao conjunto NP -difícil, mas não ao conjunto NP -completo.
- D** uma alternativa para a solução do problema é usar o algoritmo de Dijkstra para obtenção do caminho mínimo entre dois pontos.
- E** o passo de maior custo do algoritmo é a criação do conjunto D com as combinações de pontos, apresentando complexidade computacional $O(n!)$.

Figura 3.12: Questão 30 do ENADE 2014.

Em uma análise mais detalhada dessa questão é possível observar que somente 4,167%

dos alunos da UnB marcaram a opção "A", opção correta, 12,5% a opção B, 12,5% opção C, 25% opção D, 37,5% opção E e 8,333% deixaram a questão em branco.

Esse tipo de análise pode ser útil para analisar as áreas de conhecimento. Podendo assim estimular melhorias em algumas áreas individualmente em uma instituição e melhorá-la como um todo. As questões da prova podem ser visualizadas em http://download.inep.gov.br/educacao_superior/enade/provas/2014/03_computacao_bacharelado.pdf.

Foram também criadas tabelas que mostram as medianas das notas gerais, componente específico e formação geral para alguns cursos, como o de Ciência da Computação, Computação Licenciatura, Engenharia da Computação e a mediana geral do ENADE.

De acordo com uma matéria da “Folha de São Paulo” de 10 de Dezembro de 2017 [24] foi identificado que o desempenho dos alunos cotistas no final do curso é, em 33 de 64 cursos analisados, superior ou um pouco inferior aos alunos não cotistas (sendo pouco inferior considerada semelhante, cerca de até 5% inferior). A matéria informa que "A nota deles (alunos cotistas) é inferior, porém, especialmente nas exatas. Dos dez cursos em que os cotistas estão mais atrás dos demais, sete são dessa área."

Sabendo disso, pôde ser desenvolvida uma análise onde foram acrescentadas às tabelas anteriores que continham as medianas das notas gerais, de componente específico e formação geral para alguns cursos, filtros que foram considerados relevantes para estudo, tais como: alunos que concluíram o ensino médio em escolas públicas ou privadas, alunos negros, brancos, pardos, indígenas, assim como critérios de renda, entre outros. Além disso, as tabelas com os novos atributos foram geradas para todos os cursos do ENADE para enriquecer a análise, e não apenas para os cursos citados anteriormente. Como exemplo destas tabelas, podem ser visualizadas as Tabelas 3.22 e 3.23 que mostram como ficou a disposição dos novos atributos e suas respectivas medianas para cada um dos cursos e cada uma das notas, além disso, é mostrado também a porcentagem que o número de instâncias referentes a cada filtro representa do total de alunos do ENADE (onde 1,00 representa 100%). As tabelas representam as medianas dos novos atributos para o ENADE geral e para o curso de Ciência da Computação respectivamente.

Tabela 3.22: Lista dos atributos usados e as respectivas medianas para o ENADE geral.

Atributos	ENADE_FG	ENADE_CE	ENADE_GER	ENADE percentual do total de alunos do ENADE (%)
alunos brancos segundo grau escola publica	55	40,8	44,3	29,176
alunos negros segundo grau escola publica	52,1	38,1	41,5	7,085
alunos pardos segundo grau escola publica	52,3	37,8	41,4	22,875
alunos brancos segundo grau escola particular	63,1	45,2	49,5	12,363
alunos negros segundo grau escola particular	56,9	39,7	44	0,973
alunos pardos segundo grau escola particular	60,1	42,5	46,9	4,418
alunos brancos federal segundo grau escola publica	60,1	45	48,6	4,615
alunos negros federal segundo grau escola publica	54,7	40	43,8	1,518
alunos pardos federal segundo grau escola publica	55,3	39,8	43,6	5,117
alunos brancos faculdade particular segundo grau escola particular	60,7	42	46,5	5,863
alunos negros faculdade particular segundo grau escola particular	54,5	35,7	40,5	0,412
alunos pardos faculdade particular segundo grau escola particular	57,8	39	43,7	1,801
alunos brancos particular renda 1;5 a 3 salarios minimos	52,6	39,3	42,7	6,877
alunos negros particular renda 1;5 a 3 salarios minimos	51,1	38	41,2	1,474
alunos pardos particular renda 1;5 a 3 salarios minimos	50,8	37,8	41,2	4,912
alunos brancos particular renda 10 a 30 salarios minimos	62	42,5	47,4	2,595
alunos negros particular renda 10 a 30 salarios minimos	58,05	37,25	42,1	0,139
alunos pardos particular renda 10 a 30 salarios minimos	59,5	40,3	45	0,657
alunos brancos federal renda 1;5 a 3 salarios minimos	59,9	44,2	48	2,05
alunos negros federal renda 1;5 a 3 salarios minimos	55,3	41,4	44,8	0,665
alunos pardos federal renda 1;5 a 3 salarios minimos	55,4	40,2	43,9	2,279
alunos brancos federal renda 10 a 30 salarios minimos	69	53,2	56,9	1,33
alunos negros federal renda 10 a 30 salarios minimos	63,5	48,5	52,7	0,07
alunos pardos federal renda 10 a 30 salarios minimos	66,7	50,1	54,15	0,429
alunos particular brancos	55,4	40,2	43,9	27,173
alunos particular negros	51,4	37	40,6	4,334
alunos particular pardos	51,9	37,5	41,2	15,083
alunos particular brancos acao afirmativa racial	45	30,9	34,8	0,095
alunos particular negros acao afirmativa racial	56,3	42	45,4	0,128
alunos particular pardos acao afirmativa racial	52,3	38,4	41,9	0,191
alunos federal brancos	63,2	47,7	51,2	9,274
alunos federal negros	55,6	40,9	44,9	2,015
alunos federal pardos	57,5	41,9	45,7	7,401
alunos federal brancos sem acao afirmativa	63,5	48	51,5	7,774
alunos federal brancos acao afirmativa renda	56,5	41,7	44,6	0,094
alunos federal brancos acao afirmativa escola publica	63,6	48,1	51,5	1,074
alunos federal negros sem acao afirmativa	55,75	41	45	1,323
alunos federal negros acao afirmativa escola publica	58,1	42,9	46,6	0,195
alunos federal negros acao afirmativa racial	54,9	40,7	44,5	0,217
alunos federal negros acao afirmativa renda	51,5	37,25	43,4	0,032
alunos federal pardos sem acao afirmativa	58,2	42	46	5,482
alunos federal pardos acao afirmativa escola publica	59,1	44	47,8	0,921
alunos federal pardos acao afirmativa racial	58,1	41,9	45,7	0,159
alunos federal pardos acao afirmativa renda	53,3	40,4	43	0,125
alunos brancos federal segundo grau escola particular	66,8	50,8	54,5	3,83
alunos negros federal segundo grau escola particular	60,4	44,5	48,55	0,341
alunos pardos federal segundo grau escola particular	63,7	47,3	50,9	1,712
alunos brancos faculdade particular segundo grau escola publica	53,8	39,7	43,2	18,358
alunos negros faculdade particular segundo grau escola publica	51,3	37,5	40,8	3,534
alunos pardos faculdade particular segundo grau escola publica	51	37,4	40,9	11,923
alunos brancos federal segundo grau escola particular acao afirmativa	63,6	48,1	51,5	1,074
alunos negros federal segundo grau escola particular acao afirmativa	58,1	42,9	46,6	0,195
alunos pardos federal segundo grau escola particular acao afirmativa	59,1	44	47,8	0,921
alunos brancos federal segundo grau escola particular sem acao afirmativa	66,9	51	54,6	3,771
alunos negros federal segundo grau escola particular sem acao afirmativa	60,9	44,8	48,8	0,309
alunos pardos federal segundo grau escola particular sem acao afirmativa	64	47,5	51,1	1,647

Tabela 3.23: Lista dos atributos usados e as respectivas medianas para os alunos do curso de Ciência da Computação (bacharel).

Atributos	CiC (bacharel)_FG	CiC (bacharel)_CE	CiC (bacharel)_GER	CiC (bacharel) percentual do total de alunos do ENADE (%)
alunos brancos segundo grau escola publica	61,1	36,5	42,6	30,202
alunos negros segundo grau escola publica	59,5	32,4	39	3,346
alunos pardos segundo grau escola publica	60,7	35,4	41,3	13,25
alunos brancos segundo grau escola particular	64,4	42,5	48	23,184
alunos negros segundo grau escola particular	59,95	36,45	41,5	1,448
alunos pardos segundo grau escola particular	63	39,5	45,1	7,589
alunos brancos federal segundo grau escola publica	67,3	51,5	55,3	5,569
alunos negros federal segundo grau escola publica	64,7	46	50,6	0,683
alunos pardos federal segundo grau escola publica	67,4	49,5	53,8	3,264
alunos brancos faculdade particular segundo grau escola particular	61,8	35,9	42,3	9,833
alunos negros faculdade particular segundo grau escola particular	56,35	31,85	38,05	0,653
alunos pardos faculdade particular segundo grau escola particular	59,1	32,4	39,4	3,458
alunos brancos particular renda 1;5 a 3 salarios minimos	59,9	32,4	39,1	5,375
alunos negros particular renda 1;5 a 3 salarios minimos	63,25	30,3	38,85	0,877
alunos pardos particular renda 1;5 a 3 salarios minimos	57,1	29,8	37,1	2,785
alunos brancos particular renda 10 a 30 salarios minimos	62,55	37,7	43,65	4,498
alunos negros particular renda 10 a 30 salarios minimos	57,5	33,4	42,3	0,204
alunos pardos particular renda 10 a 30 salarios minimos	58,7	36,1	41,1	1,316
alunos brancos federal renda 1;5 a 3 salarios minimos	64,9	49,5	53	2,152
alunos negros federal renda 1;5 a 3 salarios minimos	59,4	48,25	50,1	0,275
alunos pardos federal renda 1;5 a 3 salarios minimos	68,8	48,75	54,2	1,448
alunos brancos federal renda 10 a 30 salarios minimos	68,7	54	56,85	2,999
alunos negros federal renda 10 a 30 salarios minimos	66,85	49	55,85	0,133
alunos pardos federal renda 10 a 30 salarios minimos	69,8	51,5	57,1	0,989
alunos particular brancos	60	33,9	40,4	30,885
alunos particular negros	57,65	31,3	36,6	3,019
alunos particular pardos	58,1	31,3	38	11,995
alunos particular brancos acao afirmativa racial	41,9	25	27,9	0,061
alunos particular negros acao afirmativa racial	68,9	40,5	46,6	0,082
alunos particular pardos acao afirmativa racial	58,05	32,65	39,35	0,102
alunos federal brancos	67,35	51,5	55,15	14,555
alunos federal negros	61,7	45,5	49,9	1,306
alunos federal pardos	67,5	49,5	54,1	6,375
alunos federal brancos sem acao afirmativa	67,1	51,5	54,8	12,76
alunos federal brancos acao afirmativa renda	53,15	50,25	49,05	0,102
alunos federal brancos acao afirmativa escola publica	70,3	54,25	57,3	1,54
alunos federal negros sem acao afirmativa	63,2	45	49	0,959
alunos federal negros acao afirmativa escola publica	65,2	49	55	0,133
alunos federal negros acao afirmativa racial	45,5	40	42	0,133
alunos federal negros acao afirmativa renda	59,35	32,85	39,45	0,02
alunos federal pardos sem acao afirmativa	67,9	50,5	54,3	5,263
alunos federal pardos acao afirmativa escola publica	66,5	48,5	53,8	0,734
alunos federal pardos acao afirmativa racial	67,5	48,25	55,35	0,143
alunos federal pardos acao afirmativa renda	58,2	43,75	50,25	0,082
alunos brancos federal segundo grau escola particular	67,4	51,5	55,25	7,956
alunos negros federal segundo grau escola particular	61,4	44,5	48	0,51
alunos pardos federal segundo grau escola particular	67,65	49,5	54	2,642
alunos brancos faculdade particular segundo grau escola publica	59,6	33,4	40	17,942
alunos negros faculdade particular segundo grau escola publica	57,8	30,3	36,1	2,06
alunos pardos faculdade particular segundo grau escola publica	58,35	30,3	37,5	7,405
alunos brancos federal segundo grau escola particular acao afirmativa	70,3	54,25	57,3	1,54
alunos negros federal segundo grau escola particular acao afirmativa	65,2	49	55	0,133
alunos pardos federal segundo grau escola particular acao afirmativa	66,5	48,5	53,8	0,734
alunos brancos federal segundo grau escola particular sem acao afirmativa	67,4	51,5	55,2	7,915
alunos negros federal segundo grau escola particular sem acao afirmativa	61,4	45,5	49,8	0,469
alunos pardos federal segundo grau escola particular sem acao afirmativa	67,9	49,5	54,1	2,581

3.8 Correlações

Tendo como base a tabela com os micro dados do ENADE gerados e tratados de acordo com a Seção 3.1, algumas alterações foram feitas, e que serão explicadas a seguir, para que os dados estivessem adequados para a análise que foi almejada e que será explicada mais à frente.

3.8.1 Preparação prévia e adaptação dos dados

Primeiramente, foram transformados os valores de alguns atributos com valores nominais para numéricos, o que possibilita fazer mais análises estatísticas. Os atributos que tiveram seus valores transformados foram:

- qp_i1_grau_dificuldade_prova_parte_Formacao_Geral
- qp_i2_grau_dificuldade_prova_parte_Componente_Especifico
- qp_i3_duracao_da_prova_considerando_extensao
- qp_i4_enunciados_questoes_parte_Formacao_Geral_claros_e_objetivos
- qp_i5_enunciados_questoes_parte_Componente_Especifico_claros_e_objetivos
- qp_i6_informacoes_e_instrucoes_p_resol_questoes_foram_suficientes_p_resolvelas
- qp_i7_alguma_dificuldade_ao_responder_a_prova
- qp_i8_considerando_questoes_objetivas_percebeu_que
- qp_i9_tempo_gasto_p_concluir_a_prova
- i7_qtde_familiares_em_casa
- i8_renda_familia
- i10_situacao_trabalho
- i22_qts_livros_leu_neste_ano
- i23_qts_horas_estudo_a_parte_p_semana

Para os atributos que tinham uma característica gradual, como alguns dos qps, que são as questões pessoais, (onde tem-se, por exemplo, desde “muito fácil” até “muito difícil”) foram colocados valores sequenciais como 0, 1, 2, 3, 4, 5, para viabilizar e facilitar o cálculo da média. Para os valores que ficavam entre um intervalo, como “de 1 a 3 salários mínimos”, foi utilizado o valor médio desse intervalo (por exemplo, se fosse entre 2000 e

3000 reais, o valor definido seria 2500), pois foi suposta uma distribuição equitativa de valores no intervalo. Para valores “acima de x”, foi utilizado o próprio valor x. Para valores “até y”, foi utilizada a média, ou seja, o valor $y/2$ (exemplo: “até 1 salário mínimo” ficaria a metade do valor do salário mínimo). No número de familiares apenas substituiu-se o nominal pelo valor numérico correspondente, por exemplo, “um familiar” substituiu-se por 1.

As exatas substituições que foram feitas podem ser visualizadas a seguir com detalhes, no formato

“<variável> = <valor nominal original>: <valor numérico modificado>, ...”

- renda = {"Ate 1;5 salario minimo (ate R\$ 1.086;00).":543,"De 1;5 a 3 salarios minimos (R\$ 1.086;01 a R\$ 2.172;00).":1629,"De 3 a 4;5 salarios minimos (R\$ 2.172;01 a R\$ 3.258;00).":2715,"De 4;5 a 6 salarios minimos (R\$ 3.258;01 a R\$ 4.344;00).":3801,"De 6 a 10 salarios minimos (R\$ 4.344;01 a R\$ 7.240;00).":5792,"De 10 a 30 salarios minimos (R\$ 7.240;01 a R\$ 21.720;00).":14480,"Acima de 30 salarios minimos (mais de R\$ 21.720;01).":21720,-1}
- livros = {"Nenhum.":0,"Um ou dois.":1.5,"De tres a cinco.":4,"De seis a oito.":7,"Mais de oito.":8,-1}
- horas_estudo = {"Nenhuma; apenas assisto as aulas.":0,"De uma a tres.":2,"De quatro a sete.":5.5,"De oito a doze.":10,"Mais de doze.":12,-1}
- horas_trabalho = {"Nao estou trabalhando.": 0,"Trabalho eventualmente.": 5,"Trabalho ate 20 horas semanais.":20,"Trabalho de 21 a 39 horas semanais.":30,"Trabalho 40 horas semanais ou mais.":40,-1}
- num_familiares = {"Nenhuma.":0,"Uma.":1,"Duas.":2,"Tres.":3,"Quatro.":4,"Cinco.": 5,"Seis.":6,"Sete ou mais.":7,-1}
- dificuldade_prova = {"Muito facil.": 1,"Facil.": 2,"Medio.": 3,"Dificil.": 4,"Muito dificil.": 5,"*": -1,"*": -1,-1}
- duracao_prova = {"Muito longa.": 5,"Longa.": 4,"Adequada.": 3,"Curta.": 2,"Muito curta.": 1,"*": -1,"*": -1,-1}
- clareza_prova = {"Sim; todos.": 5,"Sim; a maioria.": 4,"Apenas cerca da metade.": 3,"Poucos.": 2,"Poucos se apresentam.": 2,"Nao; nenhum.": 1,"*": -1,"*": -1,-1}
- informacao_suficiente = {"Sim; ate excessivas.": 5,"Sim; em todas elas.": 4,"Sim; na maioria delas.": 3,"Sim; somente em algumas.": 2,"Nao; em nenhuma delas.": 1,"*": -1,"*": -1,-1}

- dificuldade_resposta = {"Desconhecimento do conteudo.": 5, "Forma diferente de abordagem do conteudo.": 4, "Espaco insuficiente para responder as questoes.": 3, "Falta de motivacao para fazer a prova.": 2, "Nao tive qualquer tipo de dificuldade para responder a prova.": 1, "*": -1, ".": -1, :-1}
- aprendeu_conteudo = {"Nao estudou ainda a maioria desses conteudos.": 1, "Estudou alguns desses conteudos; mas nao os aprendeu.": 2, "Estudou a maioria desses conteudos; mas nao os aprendeu.": 3, "Estudou e aprendeu muitos desses conteudos.": 4, "Estudou e aprendeu todos esses conteudos.": 5, ".": -1, "*": -1, :-1}
- tempo_gasto = {"Menos de uma hora.": 0.5, "Entre uma e duas horas.": 1.5, "Entre duas e tres horas.": 2.5, "Entre tres e quatro horas.": 3.5, "Quatro horas e nao consegui terminar.": 4, ".": -1, "*": -1, :-1}

Todos os valores que não se encaixavam no padrão foram substituídos por -1 para facilitar a sua diferenciação dos demais valores, pois esse valor não existe na tabela normalmente. Os valores originais podem ser visualizados no Documento “Dicionário de variáveis” explicitado na Seção 3.1.

Em seguida, foram separados os cursos e foi feito um “vetor” para cada coluna com os alunos de cada curso. Com isso foi possível calcular a média aritmética de cada um desses atributos listados acima para cada curso, como mostra a Tabela 3.24. Os valores que não se adequavam ao padrão (-1) foram simplesmente ignorados para o cálculo da média, como se o aluno não existisse naquela coluna. Por exemplo: se em uma coluna com 10 valores existir um valor -1, o cálculo da média aritmética desta coluna teria o valor da soma dividido por 9, e não por 10. Além dessas novas colunas de atributos, foram incluídas também colunas com as médias das notas FG (formação geral), CE (componente específico) e GER (geral); além de ser feito isso para cada curso, foram calculadas também as médias de forma geral para todos os alunos que realizaram o ENADE.

Tabela 3.24: Parte da tabela com atributos e suas médias para cada curso.

Curso	qp_i1_grau_dificuldade_prova	qp_i2_grau_dificuldade	qp_i3_duracao_da_prova	qp_i4_enunciados_questoes
ARQUITETURA E URBANISMO	2,438	2,642	2,955	3,094
ARTES VISUAIS (LICENCIATURA)	2,135	2,123	2,371	2,301
CIENCIA DA COMPUTACAO (BACHARELADO)	2,251	2,744	2,528	3,026
CIENCIA DA COMPUTACAO (LICENCIATURA)	2,331	2,631	2,42	2,742
CIENCIAS BIOLOGICAS (BACHARELADO)	2,061	2,39	2,482	2,88
CIENCIAS BIOLOGICAS (LICENCIATURA)	2,226	2,408	2,468	2,68
CIENCIAS SOCIAIS (BACHARELADO)	1,05	1,149	1,465	1,516
CIENCIAS SOCIAIS (LICENCIATURA)	1,382	1,414	1,616	1,634
EDUCACAO FISICA (LICENCIATURA)	2,242	2,236	2,443	2,608
ENADE	2,303	2,443	2,565	2,77
ENGENHARIA	2,376	2,715	2,787	3,096
ENGENHARIA AMBIENTAL	2,526	2,783	2,846	3,256
ENGENHARIA CIVIL	2,539	2,801	2,856	3,255
ENGENHARIA DE ALIMENTOS	2,634	2,994	3,05	3,464
ENGENHARIA DE COMPUTACAO	2,189	2,572	2,53	2,956
ENGENHARIA DE CONTROLE E AUTOMACAO	2,446	2,841	2,692	3,157
ENGENHARIA DE PRODUCAO	2,4	2,645	2,716	3,051

Esta tabela possibilita analisar todos os cursos levando em conta os atributos citados, por exemplo, pode-se ordenar os cursos pelo atributo "horas de estudo a parte por semana" como nos mostra a Tabela 3.25, onde pode-se visualizar a média de horas de estudo a parte por semana de cada curso, ordenados de forma decrescente.

Tabela 3.25: Cursos ordenados pela média do atributo "horas de estudo a parte por semana".

Curso	i23_qts_horas_estudo_a_parte_p_semana
ARQUITETURA E URBANISMO	6,857
FISICA (BACHARELADO)	6,485
MATEMATICA (BACHARELADO)	5,859
ENGENHARIA QUIMICA	5,701
ENGENHARIA DE ALIMENTOS	5,658
ENGENHARIA	5,363
ENGENHARIA ELETRICA	5,117
QUIMICA (BACHARELADO)	5,074
ENGENHARIA DE COMPUTACAO	4,928
ENGENHARIA CIVIL	4,856
ENGENHARIA FLORESTAL	4,74
FISICA (LICENCIATURA)	4,717
ENGENHARIA AMBIENTAL	4,65
ENGENHARIA DE CONTROLE E AUTOMACAO	4,622
ENGENHARIA MECANICA	4,619
CIENCIAS BIOLOGICAS (BACHARELADO)	4,408
MUSICA (LICENCIATURA)	4,389
CIENCIA DA COMPUTACAO (BACHARELADO)	4,345
QUIMICA (LICENCIATURA)	4,322
FILOSOFIA (BACHARELADO)	4,29
LETRAS-PORTUGUES E ESPANHOL (LICENCIATURA)	4,255
CIENCIA DA COMPUTACAO (LICENCIATURA)	4,211
MATEMATICA (LICENCIATURA)	4,209
HISTORIA (LICENCIATURA)	4,183
FILOSOFIA (LICENCIATURA)	4,159
ENADE	4,148
CIENCIAS BIOLOGICAS (LICENCIATURA)	4,052
LETRAS-PORTUGUES E INGLES (LICENCIATURA)	4,026
ENGENHARIA DE PRODUCAO	4,013
ARTES VISUAIS (LICENCIATURA)	3,928
GEOGRAFIA (LICENCIATURA)	3,859
PEDAGOGIA (LICENCIATURA)	3,858
LETRAS-PORTUGUES (BACHARELADO)	3,845
LETRAS-PORTUGUES (LICENCIATURA)	3,831
SISTEMAS DE INFORMACAO	3,698
HISTORIA (BACHARELADO)	3,665
GEOGRAFIA (BACHARELADO)	3,658
CIENCIAS SOCIAIS (BACHARELADO)	3,527
TECNOLOGIA EM ANALISE E DESENVOLVIMENTO DE SISTEMAS	3,438
EDUCACAO FISICA (LICENCIATURA)	3,344
CIENCIAS SOCIAIS (LICENCIATURA)	3,238
TECNOLOGIA EM GESTAO DA PRODUCAO INDUSTRIAL	3,138
TECNOLOGIA EM AUTOMACAO INDUSTRIAL	3,041
TECNOLOGIA EM REDES DE COMPUTADORES	2,951

Também pode-se visualizar os cursos ordenados pela média atributo "qp_i8" considerando

_questoes_objetivas _percebeu_que_aprendeu_o_conteudo", como nos mostra a Tabela 3.26 . O atributo tinha valores que variavam entre 1 (não estudou ainda a maioria desses conteúdos) e 5 (estudou e aprendeu todos esses conteúdos), e baseado nisso foi calculada a média.

Tabela 3.26: Cursos ordenados pela média do atributo "Considerando questões objetivas percebeu que aprendeu o conteúdo".

Curso	qp_i8_considerando_questoes_objetivas_percebeu_que_aprendeu_o_conteudo
ENGENHARIA DE ALIMENTOS	3,214
ENGENHARIA CIVIL	3,102
ENGENHARIA QUIMICA	3,086
ENGENHARIA AMBIENTAL	3,065
ENGENHARIA FLORESTAL	3,049
ARQUITETURA E URBANISMO	2,976
ENGENHARIA DE PRODUCAO	2,931
ENGENHARIA DE CONTROLE E AUTOMACAO	2,916
ENGENHARIA MECANICA	2,893
ENGENHARIA ELETRICA	2,889
ENGENHARIA	2,769
TECNOLOGIA EM GESTAO DA PRODUCAO INDUSTRIAL	2,697
ENGENHARIA DE COMPUTACAO	2,686
CIENCIA DA COMPUTACAO (BACHARELADO)	2,683
CIENCIAS BIOLOGICAS (BACHARELADO)	2,673
QUIMICA (BACHARELADO)	2,65
PEDAGOGIA (LICENCIATURA)	2,625
FISICA (BACHARELADO)	2,543
ENADE	2,517
EDUCACAO FISICA (LICENCIATURA)	2,463
SISTEMAS DE INFORMACAO	2,462
LETRAS-PORTUGUES E INGLES (LICENCIATURA)	2,417
CIENCIAS BIOLOGICAS (LICENCIATURA)	2,344
QUIMICA (LICENCIATURA)	2,278
LETRAS-PORTUGUES E ESPANHOL (LICENCIATURA)	2,258
GEOGRAFIA (LICENCIATURA)	2,248
HISTORIA (LICENCIATURA)	2,226
FISICA (LICENCIATURA)	2,167
TECNOLOGIA EM AUTOMACAO INDUSTRIAL	2,161
MUSICA (LICENCIATURA)	2,152
MATEMATICA (BACHARELADO)	2,138
CIENCIA DA COMPUTACAO (LICENCIATURA)	2,125
GEOGRAFIA (BACHARELADO)	2,12
TECNOLOGIA EM REDES DE COMPUTADORES	2,12
MATEMATICA (LICENCIATURA)	2,114
TECNOLOGIA EM ANALISE E DESENVOLVIMENTO DE SISTEMAS	2,103
LETRAS-PORTUGUES (LICENCIATURA)	2,067
ARTES VISUAIS (LICENCIATURA)	2,03
LETRAS-PORTUGUES (BACHARELADO)	1,985
FILOSOFIA (LICENCIATURA)	1,928
FILOSOFIA (BACHARELADO)	1,571
CIENCIAS SOCIAIS (BACHARELADO)	1,505
CIENCIAS SOCIAIS (LICENCIATURA)	1,458
HISTORIA (BACHARELADO)	1,39

Outro exemplo pode ser visualizado na Tabela 3.27, onde os cursos estão ordenados pelas médias do atributo "Renda familiar".

Tabela 3.27: Cursos ordenados pela média do atributo "Renda familiar".

Curso	i8_renda_familia
ARQUITETURA E URBANISMO	6436,637
ENGENHARIA CIVIL	6147,522
ENGENHARIA QUIMICA	6120,906
ENGENHARIA MECANICA	5928,767
ENGENHARIA DE COMPUTACAO	5842,271
ENGENHARIA DE PRODUCAO	5797,38
ENGENHARIA DE CONTROLE E AUTOMACAO	5556,019
ENGENHARIA ELETRICA	5540,401
ENGENHARIA	5315,578
CIENCIA DA COMPUTACAO (BACHARELADO)	4693,088
ENGENHARIA AMBIENTAL	4673,8
ENGENHARIA DE ALIMENTOS	4552,871
FISICA (BACHARELADO)	4071,574
MATEMATICA (BACHARELADO)	4012,308
SISTEMAS DE INFORMACAO	3989,321
QUIMICA (BACHARELADO)	3855,053
CIENCIAS BIOLOGICAS (BACHARELADO)	3564,359
LETRAS-PORTUGUES (BACHARELADO)	3538,563
ENGENHARIA FLORESTAL	3522,411
TECNOLOGIA EM AUTOMACAO INDUSTRIAL	3302,082
TECNOLOGIA EM ANALISE E DESENVOLVIMENTO DE SISTEMAS	3280,657
CIENCIAS SOCIAIS (BACHARELADO)	3260,248
TECNOLOGIA EM GESTAO DA PRODUCAO INDUSTRIAL	3176,545
ENADE	3175,838
FILOSOFIA (BACHARELADO)	3174,553
HISTORIA (BACHARELADO)	3139,421
MUSICA (LICENCIATURA)	3035,953
GEOGRAFIA (BACHARELADO)	2982,247
TECNOLOGIA EM REDES DE COMPUTADORES	2879,899
ARTES VISUAIS (LICENCIATURA)	2507,903
EDUCACAO FISICA (LICENCIATURA)	2384,027
LETRAS-PORTUGUES E INGLES (LICENCIATURA)	2383,363
FISICA (LICENCIATURA)	2345,584
CIENCIA DA COMPUTACAO (LICENCIATURA)	2323,035
QUIMICA (LICENCIATURA)	2303,649
FILOSOFIA (LICENCIATURA)	2244,556
CIENCIAS SOCIAIS (LICENCIATURA)	2217,111
HISTORIA (LICENCIATURA)	2167,109
MATEMATICA (LICENCIATURA)	2139,208
CIENCIAS BIOLOGICAS (LICENCIATURA)	2125,394
LETRAS-PORTUGUES E ESPANHOL (LICENCIATURA)	2124,512
GEOGRAFIA (LICENCIATURA)	2006,921
PEDAGOGIA (LICENCIATURA)	1964,854
LETRAS-PORTUGUES (LICENCIATURA)	1832,46

E para finalizar, uma outra tabela como exemplo, a tabela contendo o atributo situação trabalho que representa a média do número de horas trabalhadas semanalmente, como mostra a Tabela 3.28.

Tabela 3.28: Cursos ordenados pela média do atributo "Situação trabalho"(em horas semanais).

Curso	i10_situacao_trabalho
TECNOLOGIA EM GESTAO DA PRODUCAO INDUSTRIAL	29,169
TECNOLOGIA EM AUTOMACAO INDUSTRIAL	27,085
SISTEMAS DE INFORMACAO	26,068
CIENCIA DA COMPUTACAO (LICENCIATURA)	24,681
TECNOLOGIA EM REDES DE COMPUTADORES	24,678
ENGENHARIA DE PRODUCAO	24,644
TECNOLOGIA EM ANALISE E DESENVOLVIMENTO DE SISTEMAS	23,672
ENGENHARIA DE CONTROLE E AUTOMACAO	22,531
CIENCIA DA COMPUTACAO (BACHARELADO)	21,86
ENGENHARIA ELETRICA	21,381
ENGENHARIA MECANICA	21,282
PEDAGOGIA (LICENCIATURA)	21,074
MATEMATICA (LICENCIATURA)	20,636
ENGENHARIA DE COMPUTACAO	20,111
ENGENHARIA CIVIL	19,672
LETRAS-PORTUGUES E INGLES (LICENCIATURA)	19,422
ENADE	19,318
ARTES VISUAIS (LICENCIATURA)	19,26
HISTORIA (LICENCIATURA)	18,791
LETRAS-PORTUGUES E ESPANHOL (LICENCIATURA)	18,377
GEOGRAFIA (LICENCIATURA)	18,295
MUSICA (LICENCIATURA)	17,816
EDUCACAO FISICA (LICENCIATURA)	17,232
ENGENHARIA AMBIENTAL	17,186
QUIMICA (LICENCIATURA)	16,806
LETRAS-PORTUGUES (LICENCIATURA)	16,402
FILOSOFIA (LICENCIATURA)	16,315
FISICA (LICENCIATURA)	16,27
CIENCIAS SOCIAIS (LICENCIATURA)	14,905
CIENCIAS BIOLOGICAS (LICENCIATURA)	14,732
QUIMICA (BACHARELADO)	14,589
ARQUITETURA E URBANISMO	14,02
ENGENHARIA	13,915
ENGENHARIA QUIMICA	13,692
GEOGRAFIA (BACHARELADO)	13,236
LETRAS-PORTUGUES (BACHARELADO)	12,149
FILOSOFIA (BACHARELADO)	11,183
CIENCIAS SOCIAIS (BACHARELADO)	10,469
ENGENHARIA DE ALIMENTOS	10,064
HISTORIA (BACHARELADO)	9,907
CIENCIAS BIOLOGICAS (BACHARELADO)	9,559
FISICA (BACHARELADO)	7,39
MATEMATICA (BACHARELADO)	6,847
ENGENHARIA FLORESTAL	6,235

3.8.2 Transformações e análises

Este próximo passo é importante pois foram feitas correlações com esses 14 atributos entre si, e cada um deles com as notas CE, FG e GER. Pode-se ter uma ideia da organização da tabela pela Tabela 3.31. Para ser feita a correlação, foi utilizada a biblioteca Pandas[20] do Python com a função corr(). Então obteve-se uma tabela onde as colunas são as correlações (por exemplo: “nota ENADE x renda”, ou mesmo “renda i8 x situação horas trabalho i10”), e as linhas são os cursos. As correlações [25] entre duas variáveis

são indicadas por valores entre -1 e 1. Quando o coeficiente de correlação se aproxima de 1, quando o valor de uma variável aumenta, o da outra variável também aumenta, o que caracteriza uma relação linear positiva. Quando o coeficiente se aproxima de -1, ainda há correlação mas ela é inversa, ou seja, quando o valor de uma variável aumenta, o da outra diminui. Quando o valor do coeficiente se aproxima de 0, não há relação entre as variáveis. É importante explicitar que esta correlação não indica causa e efeito, ou seja, a ocorrência de uma variável não é causa ou efeito da ocorrência de outra variável, por exemplo, uma correlação alta ou baixa não implica que uma variável interfere na outra.

Além disso, outro passo foi criar outras duas tabelas parecidas com a recém criada, contudo, uma delas contendo os valores (coeficiente de correlação) calculados apenas de alunos de instituições federais, a Tabela 3.29 mostra uma parte da tabela (pois a tabela é muito grande para ser visualizada por inteiro), e outra somente de alunos de instituições privadas, como pode ser visualizado na Tabela 3.30 uma pequena parte da tabela (para poder caber devido ao seu tamanho). Analisando essas tabelas, pode-se perceber a correlação entre alguns atributos, por exemplo na tabela de instituições privadas na coluna "qp_i1_grau_dificuldade_prova_parte_Formacao_Geral x qp_i2_grau_dificuldade_prova_parte_Componente_Especifico", nota-se um coeficiente de correlação muito alto para a maior parte dos cursos, chegando a valores de até 0,974, indicando que estes dois atributos estão muito correlacionados. Na tabela das instituições federais, por exemplo, o segundo menor valor de correlação é da relação "Nota_CE x i10_situacao_trabalho" com o valor de -0,217092655 para o curso de LETRAS- PORTUGUES E INGLES (LICENCIATURA), o que indica uma relação inversamente proporcional. Um exemplo de correlação muito próxima a zero (0) pode ser visualizada também na tabela de instituições federais, onde, para o curso de QUIMICA (LICENCIATURA), a correlação "Nota_GER x qp_i2_grau_dificuldade_prova_parte_Componente_Especifico" possui valor -0,000492745, o que indica que estes atributos possuem uma correlação muito fraca. As tabelas originais de correlação possuem 44 linhas e 134 colunas (44x134) cada uma das linhas contendo os cursos e cada uma das colunas contendo as correlações dos atributos.

Tabela 3.29: Parte da tabela de correlações de instituições federais.

Cursos	Nota_CE x i10_situacao_trabalho	Nota_CE x i22_qts_livros_leu_neste_ano	Nota_CE x i23_qts_horas_estudo_a_parte_p_semana
ARQUITETURA E URBANISMO	-0,052884255	-0,004083035	0,066812128
ARTES VISUAIS (LICENCIATURA)	-0,047537395	0,067851623	0,059891611
CIENCIA DA COMPUTACAO (BACHARELADO)	-0,070686319	0,055260326	0,095314654
CIENCIA DA COMPUTACAO (LICENCIATURA)	0,031388979	-0,040604092	0,023415905
CIENCIAS BIOLOGICAS (BACHARELADO)	-0,002803928	0,069424011	0,065310066
CIENCIAS BIOLOGICAS (LICENCIATURA)	-0,05551226	0,020175934	0,065115575
CIENCIAS SOCIAIS (BACHARELADO)	-0,042664709	0,029367385	0,077643776
CIENCIAS SOCIAIS (LICENCIATURA)	-0,012919447	0,023115255	0,050580473
EDUCACAO FISICA (LICENCIATURA)	-0,030885172	-0,022681863	0,061214182
ENGENHARIA AMBIENTAL	-0,053344398	0,066188498	0,075216165
ENGENHARIA CIVIL	-0,058039708	0,053843124	0,060279055
ENGENHARIA DE ALIMENTOS	-0,054851141	0,004533459	0,105292596
ENGENHARIA DE COMPUTACAO	-0,031049208	-0,004369208	-0,000117376
ENGENHARIA DE CONTROLE E AUTOMACAO	-0,08763521	0,053867808	0,045775394
ENGENHARIA DE PRODUCAO	-0,037244229	0,012140726	0,026926232
ENGENHARIA ELETRICA	-0,071224283	0,061505469	0,057128066
ENGENHARIA FLORESTAL	0,034634556	-0,004612828	0,081216927
ENGENHARIA MECANICA	-0,058754736	0,066866268	0,032100471
ENGENHARIA QUIMICA	-0,048411711	0,081484297	0,075976768
ENGENHARIA	-0,117574216	0,061650745	0,141737053
FILOSOFIA (BACHARELADO)	-0,019000682	0,079164836	0,118290955
FILOSOFIA (LICENCIATURA)	0,057096863	0,04138578	0,042225602
FISICA (BACHARELADO)	0,003498434	0,055300546	-0,00494399
FISICA (LICENCIATURA)	0,00926294	0,086731675	0,065725956
GEOGRAFIA (BACHARELADO)	-0,042553901	-0,032486518	0,04698688
GEOGRAFIA (LICENCIATURA)	-0,064327575	0,023308137	0,114547758
HISTORIA (BACHARELADO)	-0,02611351	0,037456673	0,066344116
HISTORIA (LICENCIATURA)	-0,058691995	0,044982286	0,081084384
LETRAS-PORTUGUES (BACHARELADO)	0,022389277	-0,001810592	0,02585371
LETRAS-PORTUGUES (LICENCIATURA)	-0,023676019	0,032544194	0,086923949
LETRAS-PORTUGUES E ESPANHOL (LICENCIATURA)	-0,046083353	0,093534452	0,098110034
LETRAS-PORTUGUES E INGLES (LICENCIATURA)	-0,148400665	0,06062279	0,079790961
MATEMATICA (BACHARELADO)	-0,083826781	0,036848412	0,098319111
MATEMATICA (LICENCIATURA)	0,044773883	0,066652023	0,093195622
MUSICA (LICENCIATURA)	0,031538057	0,015658319	0,055802477
PEDAGOGIA (LICENCIATURA)	-0,024315606	0,013439192	0,078612694
QUIMICA (BACHARELADO)	-0,058823908	0,077534265	0,044504726
QUIMICA (LICENCIATURA)	-0,005894402	0,055273146	0,086565286
SISTEMAS DE INFORMACAO	-0,00870654	0,098378997	0,065869736
TECNOLOGIA EM ANALISE E DESENVOLVIMENTO DE SISTEMAS	-0,016871865	0,047707064	-0,01531837
TECNOLOGIA EM AUTOMACAO INDUSTRIAL	-0,115118989	0,011848641	0,029249784
TECNOLOGIA EM GESTAO DA PRODUCAO INDUSTRIAL	0,03283932	-0,070233346	0,019233004
TECNOLOGIA EM REDES DE COMPUTADORES	0,018998177	0,113153875	0,13107917

Tabela 3.30: Parte da tabela de correlações de instituições privadas.

Cursos	Nota_CE x i10_situacao_trabalho	Nota_CE x i22_qts_livros_leu_neste_ano	Nota_CE x i23_qts_horas_estudo_a_parte_p_semana
ARQUITETURA E URBANISMO	-0,041340686	0,069693163	0,077797591
ARTES VISUAIS (LICENCIATURA)	0,027072478	0,052446569	0,076431284
CIENCIA DA COMPUTACAO (BACHARELADO)	0,001390805	0,056467398	0,060326777
CIENCIA DA COMPUTACAO (LICENCIATURA)	0,10926796	-0,008942976	0,028012656
CIENCIAS BIOLOGICAS (BACHARELADO)	-0,059083868	0,050105406	0,039581307
CIENCIAS BIOLOGICAS (LICENCIATURA)	0,007038176	0,0434132	0,111150568
CIENCIAS SOCIAIS (BACHARELADO)	-0,035010728	-0,090377501	0,064525229
CIENCIAS SOCIAIS (LICENCIATURA)	0,058673805	0,134016631	0,128664213
EDUCACAO FISICA (LICENCIATURA)	0,055692013	0,055921007	0,097344288
ENGENHARIA AMBIENTAL	-0,01351428	0,027211134	0,090416921
ENGENHARIA CIVIL	-0,011159251	0,058738897	0,042966534
ENGENHARIA DE ALIMENTOS	-0,095661699	-0,015759897	0,041980428
ENGENHARIA DE COMPUTACAO	-0,018447679	0,048822325	0,115838455
ENGENHARIA DE CONTROLE E AUTOMACAO	-0,058026001	0,049853387	0,069418306
ENGENHARIA DE PRODUCAO	-0,016225633	0,028602422	0,020007341
ENGENHARIA ELETRICA	-0,026447467	0,037474737	0,028398111
ENGENHARIA FLORESTAL	0,094629432	0,0648129	0,119717158
ENGENHARIA MECANICA	-0,009685031	0,016040173	0,048987986
ENGENHARIA QUIMICA	-0,028420241	0,084715143	0,081714266
ENGENHARIA	-0,061503803	0,030662195	0,038964347
FILOSOFIA (BACHARELADO)	0,104026901	0,097554778	0,162062852
FILOSOFIA (LICENCIATURA)	0,073423454	0,046831207	0,088086243
FISICA (BACHARELADO)	-0,231097775	-0,025811408	0,050832005
FISICA (LICENCIATURA)	0,033158819	0,135417797	0,140753448
GEOGRAFIA (BACHARELADO)	-0,077233702	0,018522598	0,052780249
GEOGRAFIA (LICENCIATURA)	0,077605058	0,046447123	0,0872901
HISTORIA (BACHARELADO)	-0,10109579	0,05838607	-0,008251705
HISTORIA (LICENCIATURA)	0,073925251	0,082865644	0,103233321
LETRAS-PORTUGUES (BACHARELADO)	0,079853903	-0,059954488	0,015343803
LETRAS-PORTUGUES (LICENCIATURA)	0,099925063	0,072545426	0,088951566
LETRAS-PORTUGUES E ESPANHOL (LICENCIATURA)	0,131846588	0,074398867	0,097935868
LETRAS-PORTUGUES E INGLES (LICENCIATURA)	0,048205654	0,063066848	0,087216958
MATEMATICA (BACHARELADO)	0,188811464	-0,054971105	0,700695108
MATEMATICA (LICENCIATURA)	0,104163037	0,031555163	0,081907005
MUSICA (LICENCIATURA)	0,139160566	0,038218999	0,070979896
PEDAGOGIA (LICENCIATURA)	0,071542171	0,038009649	0,072891452
QUIMICA (BACHARELADO)	-0,030016309	0,12202901	0,136889279
QUIMICA (LICENCIATURA)	0,008516055	0,077913121	0,036112607
SISTEMAS DE INFORMACAO	0,038496306	0,069417121	0,060719957
TECNOLOGIA EM ANALISE E DESENVOLVIMENTO DE SISTEMAS	0,057732109	0,071373245	0,071591431
TECNOLOGIA EM AUTOMACAO INDUSTRIAL	0,06057518	0,119755331	0,069954163
TECNOLOGIA EM GESTAO DA PRODUCAO INDUSTRIAL	0,03232345	0,065124792	0,068976418
TECNOLOGIA EM REDES DE COMPUTADORES	0,04433128	0,08315898	0,046886437

Com essas duas tabelas (federal e privada) de correlações, pôde ser gerada uma outra tabela semelhante a essas onde os valores das correlações seriam o valor da correlação da federal - (menos) valor da correlação da privada, utilizando os mesmos cursos e as mesmas colunas, como mostra a Tabela 3.31. Nesta tabela de diferença de correlações, pode ser analisado que, quanto mais próximo de zero (0) for o valor, mais semelhantes são as correlações. Se ambos atributos forem positivos e a subtração resultar em um número positivo, significa que a correlação da federal é mais forte que a da privada; se o resultado for negativo, significa que a correlação da privada é mais forte.

Tabela 3.31: Parte da tabela com as diferenças entre correlações federal e privada.

Curso	Nota_CE x i10_situacao_trabalho	Nota_CE x i22_qts_livros_leu_neste_ano	Nota_CE x i23_qts_horas_estudo_a_parte_p_semana
ARQUITETURA E URBANISMO	-0,00981179	-0,056046781	0,004418051
ARTES VISUAIS (LICENCIATURA)	-0,152606255	0,013475728	-0,042897229
CIENCIA DA COMPUTACAO (BACHARELADO)	-0,068026852	0,02784431	0,055254119
CIENCIA DA COMPUTACAO (LICENCIATURA)	-0,06810969	0,002734799	0,018946924
CIENCIAS BIOLOGICAS (BACHARELADO)	0,063359726	-0,004571474	0,031628809
CIENCIAS BIOLOGICAS (LICENCIATURA)	-0,135372377	-0,031860118	0,028055453
CIENCIAS SOCIAIS (BACHARELADO)	-0,026074653	-0,022803393	-0,041553038
CIENCIAS SOCIAIS (LICENCIATURA)	0,054197347	-0,040893816	0,001163625
EDUCACAO FISICA (LICENCIATURA)	-0,12734464	-0,072324757	0,006611216
ENGENHARIA AMBIENTAL	0,04945108	-0,015397932	0,032743111
ENGENHARIA CIVIL	-0,03798981	0,025309391	0,016082414
ENGENHARIA DE ALIMENTOS	0,012659007	0,112654333	-0,025699573
ENGENHARIA DE COMPUTACAO	0,06108636	0,010314884	-0,042899822
ENGENHARIA DE CONTROLE E AUTOMACAO	0,013717331	-0,04186856	0,033211952
ENGENHARIA DE PRODUCAO	-0,005731621	0,003306105	0,003540508
ENGENHARIA ELETRICA	-0,080875562	0,054193297	0,046509794
ENGENHARIA FLORESTAL	-0,036819124	-0,012608976	-0,146367139
ENGENHARIA MECANICA	-0,023888563	0,029445723	-0,009045075
ENGENHARIA QUIMICA	0,062020198	0,045397102	-0,020541362
ENGENHARIA	-0,056070413	0,030988551	-0,102772706
FILOSOFIA (BACHARELADO)	-0,123027584	-0,018389943	-0,043771898
FILOSOFIA (LICENCIATURA)	-0,016326591	-0,005445427	-0,045860641
FISICA (BACHARELADO)	0,23459621	0,081111954	-0,055775996
FISICA (LICENCIATURA)	-0,02389588	-0,048686123	-0,075027491
GEOGRAFIA (BACHARELADO)	0,034679801	-0,051009116	-0,00579337
GEOGRAFIA (LICENCIATURA)	-0,141932633	-0,023138986	0,027257658
HISTORIA (BACHARELADO)	0,07498228	-0,020929397	0,074595822
HISTORIA (LICENCIATURA)	-0,132617246	-0,037883358	-0,022148937
LETRAS-PORTUGUES (BACHARELADO)	-0,057464625	0,058143897	0,010509907
LETRAS-PORTUGUES (LICENCIATURA)	-0,123601082	-0,040001233	-0,002027617
LETRAS-PORTUGUES E ESPANHOL (LICENCIATURA)	-0,177929942	0,019135585	0,000174166
LETRAS-PORTUGUES E INGLES (LICENCIATURA)	-0,196606319	-0,002444058	-0,007425997
MATEMATICA (BACHARELADO)	-0,272638245	0,091819517	-0,602375997
MATEMATICA (LICENCIATURA)	-0,059389154	0,03509686	0,011288617
MUSICA (LICENCIATURA)	-0,107622509	-0,02256068	-0,015177419
PEDAGOGIA (LICENCIATURA)	-0,095857777	-0,024570457	0,005721243
QUIMICA (BACHARELADO)	-0,028807599	-0,044494745	-0,092384553
QUIMICA (LICENCIATURA)	-0,014410456	-0,022639975	0,050452679
SISTEMAS DE INFORMACAO	-0,047202846	0,028961875	0,005149779
TECNOLOGIA EM ANALISE E DESENVOLVIMENTO DE SISTEMAS	-0,074603974	-0,023666181	-0,086909802
TECNOLOGIA EM AUTOMACAO INDUSTRIAL	-0,17569417	-0,10790669	-0,040704379
TECNOLOGIA EM GESTAO DA PRODUCAO INDUSTRIAL	0,00051587	-0,135358138	-0,049743414
TECNOLOGIA EM REDES DE COMPUTADORES	-0,025333103	0,029994895	0,084192733

Foram criados também *heatmaps* para todas as correlações de cada um dos cursos para se ter uma melhor visualização dessas correlações em forma de imagem. Esses *heatmaps* foram criados utilizando a biblioteca Matplotlib do Python. Um exemplo de *heatmap* pode ser visualizado na Figura 3.13, onde, como exemplo de interpretação, pode-se observar que a correlação do item 12 (situação de trabalho) tem baixa correlação negativa com os itens 15 e 16 (que representam as notas nos componentes geral e específico respectivamente). O item 9 (tempo gasto para concluir a prova) tem uma correlação positiva com as notas, indicando que, em geral, os alunos que gastaram mais tempo na prova tiveram notas maiores. A Tabela 3.32 mostra a legenda para os números do *heatmap*.

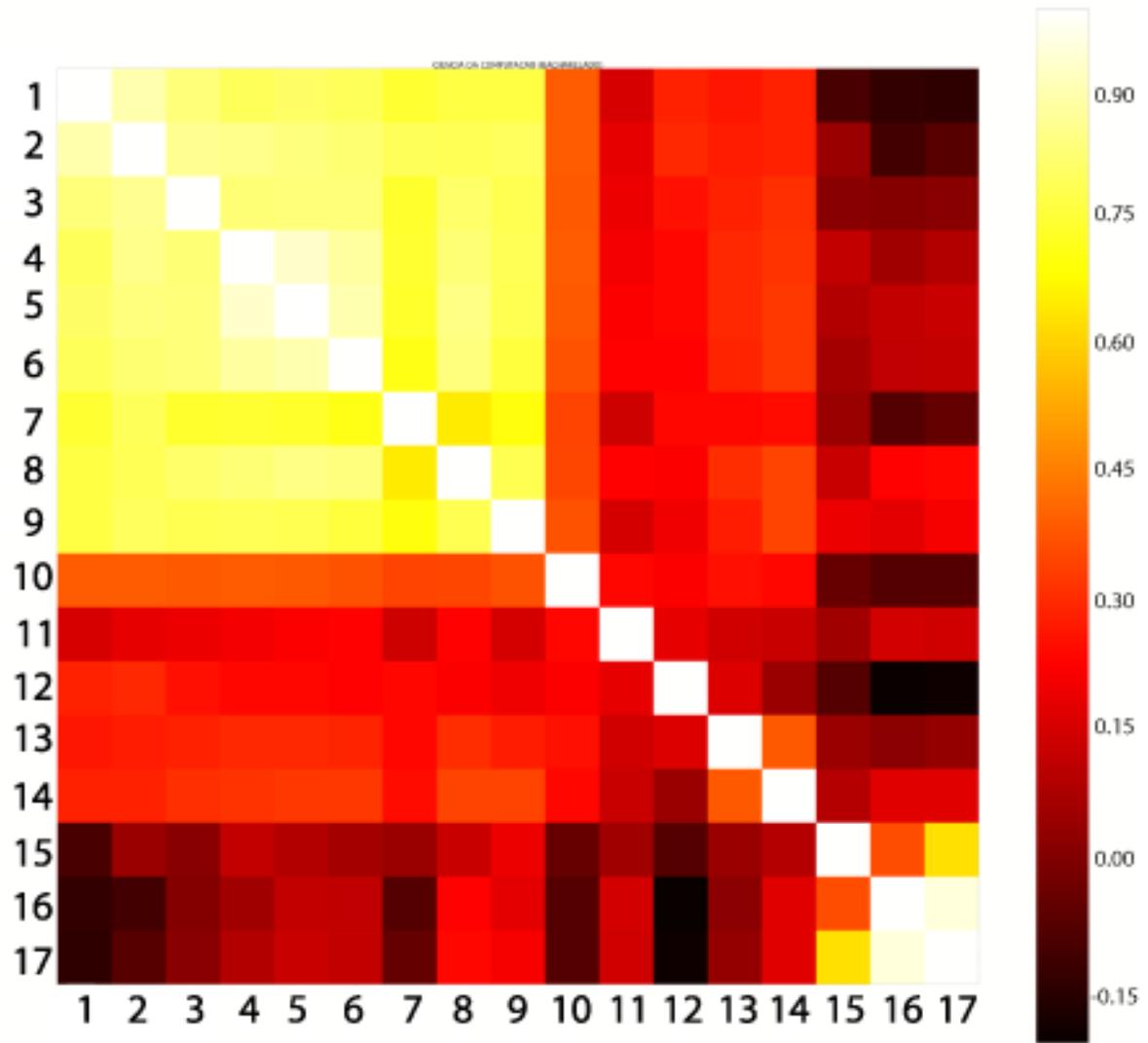


Figura 3.13: *Heatmap* das correlações do curso de CiC (bacharel).

Tabela 3.32: Legenda para o *heatmap* 3.13.

Número na imagem	Significado
1	qp_i1 Grau de dificuldade da prova de Formacao Geral
2	qp_i2 Grau de dificuldade da prova do componente específico
3	qp_i3 Duração da prova considerando extensão
4	qp_i4 Enunciados questões parte formação geral claros e objetivos
5	qp_i5 Enunciados questões parte componente específico claros e objetivos
6	qp_i6 Informações e instruções para resolver as questões foram suficientes
7	qp_i7 Dificuldade ao responder a prova
8	qp_i8 Considerando questões objetivas percebeu que já aprender o conteúdo
9	qp_i9 Tempo gasto para concluir a prova
10	i7 Quantidade de familiares em casa
11	i8 Renda família
12	i10 Situação trabalho
13	i22 Quantidade de livros que leu neste ano
14	i23 Quantidade de horas estudo a parte por semana
15	Nota da formação geral
16	Nota do componente específico
17	Nota Bruta

Considerando as notas de Componente Específico (Nota CE), Formação Geral (Nota FG) e Geral (Nota GER), foram selecionadas correlações de cada uma das notas com os seguintes atributos: renda familiar, situação trabalho, quantidade de livros lidos no ano, e quantidade de horas de estudo a parte por semana, gerando 4 colunas para cada tipo de nota (ex. “Nota_FG x i8_renda_familia”). Para cada combinação de nota com cada um desses atributos, foram criadas tabelas com base nos valores obtidos de instituições privadas, federais e da diferença entre as federais e privadas, logo, são 3 tabelas para cada tipo de correlação (de instituições federais, privadas e a diferença) totalizando 9 tabelas. A Figura 3.33 nos mostra, por exemplo, a tabela de nota FG das instituições federais, sendo que para este tipo de instituição (federal) ainda restam as tabelas de nota CE e GERAL. Com essas tabelas puderam ser analisados os valores das correlações, vendo o que de fato estava correlacionado com as notas.

Tabela 3.33: Correlações das notas FG de instituições federais.

Curso	Nota_FG x i8_renda_familia	Nota_FG x i10_situacao_trabalho	Nota_FG x i22_qts_livros	Nota_FG x i23_qts_horas_estudo
ARQUITETURA E URBANISMO	0,098769683	-0,052884255	-0,004083035	0,066812128
ARTES VISUAIS (LICENCIATURA)	0,160558752	-0,047537395	0,067851623	0,059891611
CIENCIA DA COMPUTACAO (BACHARELADO)	0,078065799	-0,070686319	0,055260326	0,095314654
CIENCIA DA COMPUTACAO (LICENCIATURA)	0,098731027	0,031388979	-0,040604092	0,023415905
CIENCIAS BIOLOGICAS (BACHARELADO)	0,109573726	-0,002803928	0,069424011	0,065310066
CIENCIAS BIOLOGICAS (LICENCIATURA)	0,100193947	-0,05551226	0,020175934	0,065115575
CIENCIAS SOCIAIS (BACHARELADO)	0,148388809	-0,042664709	0,029367385	0,077643776
CIENCIAS SOCIAIS (LICENCIATURA)	0,090743539	-0,012919447	0,023115255	0,050580473
EDUCACAO FISICA (LICENCIATURA)	0,071930265	-0,030885172	-0,022681863	0,061214182
ENGENHARIA	0,088733732	-0,117574216	0,061650745	0,141737053
ENGENHARIA AMBIENTAL	0,116389054	-0,053344398	0,066188498	0,075216165
ENGENHARIA CIVIL	0,090430303	-0,058039708	0,053843124	0,060279055
ENGENHARIA DE ALIMENTOS	0,131662237	-0,054851141	0,004533459	0,105292596
ENGENHARIA DE COMPUTACAO	0,063495234	-0,031049208	-0,004369208	-0,000117376
ENGENHARIA DE CONTROLE E AUTOMACAO	0,087871728	-0,08763521	0,053867808	0,045775394
ENGENHARIA DE PRODUCAO	0,109896389	-0,037244229	0,012140726	0,026926232
ENGENHARIA ELETRICA	0,108163404	-0,071224283	0,061505469	0,057128066
ENGENHARIA FLORESTAL	0,073359808	-0,117574216	-0,004612828	0,081216927
ENGENHARIA MECANICA	0,068425179	-0,058754736	0,066866268	0,032100471
ENGENHARIA QUIMICA	0,115207373	-0,048411711	0,081484297	0,075976768
FILOSOFIA (BACHARELADO)	0,171796677	-0,019000682	0,079164836	0,118290955
FILOSOFIA (LICENCIATURA)	0,133595747	0,057096863	0,04138578	0,042225602
FISICA (BACHARELADO)	0,222887519	0,003498434	0,055300546	-0,00494399
FISICA (LICENCIATURA)	0,11796221	0,00926294	0,086731675	0,065725956
GEOGRAFIA (BACHARELADO)	0,124233479	-0,042553901	-0,032486518	0,04698688

A maior correlação observada nestas tabelas foi a do curso de Matemática (bacharelado) em instituições privadas, na correlação “Nota_GER x i23_qts_horas_estudo_a_parte_p_semana” com um valor de 0,8861004927 (valor considerado muito alto, sabendo que a correlação vai até no máximo 1), e o segundo valor mais alto também foi observada no mesmo curso com a Nota CE (componente específico), na correlação “Nota_CE x i23_qts_horas_estudo_a_parte_p_semana” com o valor de 0,8339089482. Nas instituições federais, a maior correlação encontrada atribuída a alguma das notas foi no curso de Física (bacharelado) para a correlação “Nota_GER x i8_renda_familia” com o valor de 0,358095859, significativamente mais baixa que a das instituições privadas.

Em um passo seguinte, alguns cursos foram tomados como base, sendo eles: Ciência da Computação, bacharel e licenciatura, Engenharia da Computação, Educação Física licenciatura, Arquitetura, Filosofia licenciatura, Química Licenciatura, Física licenciatura, Matemática licenciatura e Ciências Biológicas bacharel. Para cada um dos cursos, foram selecionadas todas as correlações e ordenadas procurando assim alguma correlação significativa interessante que pudesse ser analisada, porém para correlações com valores mais altos, apenas relações consideradas triviais ou não interessantes, por exemplo, para o curso de Arquitetura a correlação com o valor mais alto, de 0,937, foi o do atributo "qp_i4_enunciados_questoes_parte_Formacao_Geral_claros_e_objetivos x qp_i5_enunciados_questoes_parte_Componente_Especifico_claros_e_objetivos", que é o atributo com uma das maiores correlação para todos os cursos analisados também, foi julgado como não interessante. Outro exemplo, para o curso de Ciência da Computação (bacharel), o atributo com terceiro maior valor de correlação foi "qp_i1_grau_dificul-

dade_prova_parte_Formacao_Geral x qp_i2_grau_dificuldade_prova_parte_Componente_Especifico" com valor igual a 0,899, também foi julgado como não interessante. Pode-se ter uma ideia de atributos não interessantes (triviais, ou intuitivos também) analisando os primeiros atributos da Tabela 3.34.

Também foram selecionadas as diferenças entre correlações de instituições federais e privadas (como foi citado anteriormente) para os cursos de Engenharia da computação, Ciência da Computação bacharel e licenciatura. A Tabela 3.34 mostra como ficou a tabela para o curso de Computação Bacharelado, assim pode-se ter uma ideia das tabelas dos outros cursos.

Tabela 3.34: Correlações de atributos para o curso de Computação (Bacharelado) de instituições privadas.

Atributos correlacionados	CIENCIA DA COMPUTACAO (BACH)
qp_i4_enunciados_questoes_parte_Formacao_Geral_claros_e_objetivos x qp_i5_enunciados_questoes_parte_Componente_Especifico_claros_e_objetivos	0,928086805
qp_i5_enunciados_questoes_parte_Componente_Especifico_claros_e_objetivos x qp_i6_informacoes_e_instrucoes_p_resol_questoes_foram_suficientes_p_resolvelas	0,905146235
qp_i1_grau_dificuldade_prova_parte_Formacao_Geral x qp_i2_grau_dificuldade_prova_parte_Componente_Especifico	0,899457776
qp_i5_enunciados_questoes_parte_Componente_Especifico_claros_e_objetivos x qp_i8_considerando_questoes_objetivas_percebeu_que	0,888757302
qp_i4_enunciados_questoes_parte_Formacao_Geral_claros_e_objetivos x qp_i6_informacoes_e_instrucoes_p_resol_questoes_foram_suficientes_p_resolvelas	0,876033922
qp_i6_informacoes_e_instrucoes_p_resol_questoes_foram_suficientes_p_resolvelas x qp_i8_considerando_questoes_objetivas_percebeu_que	0,871853849
qp_i2_grau_dificuldade_prova_parte_Componente_Especifico x qp_i5_enunciados_questoes_parte_Componente_Especifico_claros_e_objetivos	0,862008011
qp_i2_grau_dificuldade_prova_parte_Componente_Especifico x qp_i4_enunciados_questoes_parte_Formacao_Geral_claros_e_objetivos	0,855109978
qp_i4_enunciados_questoes_parte_Formacao_Geral_claros_e_objetivos x qp_i8_considerando_questoes_objetivas_percebeu_que	0,852913579
qp_i2_grau_dificuldade_prova_parte_Componente_Especifico x qp_i3_duracao_da_prova_considerando_extensao	0,849331846
qp_i3_duracao_da_prova_considerando_extensao x qp_i5_enunciados_questoes_parte_Componente_Especifico_claros_e_objetivos	0,843940283
qp_i3_duracao_da_prova_considerando_extensao x qp_i6_informacoes_e_instrucoes_p_resol_questoes_foram_suficientes_p_resolvelas	0,843892517
qp_i3_duracao_da_prova_considerando_extensao x qp_i8_considerando_questoes_objetivas_percebeu_que	0,83702133
qp_i1_grau_dificuldade_prova_parte_Formacao_Geral x qp_i5_enunciados_questoes_parte_Componente_Especifico_claros_e_objetivos	0,835407339
qp_i2_grau_dificuldade_prova_parte_Componente_Especifico x qp_i8_considerando_questoes_objetivas_percebeu_que	0,832376153
qp_i1_grau_dificuldade_prova_parte_Formacao_Geral x qp_i3_duracao_da_prova_considerando_extensao	0,829984382
qp_i2_grau_dificuldade_prova_parte_Componente_Especifico x qp_i6_informacoes_e_instrucoes_p_resol_questoes_foram_suficientes_p_resolvelas	0,828480322
qp_i3_duracao_da_prova_considerando_extensao x qp_i4_enunciados_questoes_parte_Formacao_Geral_claros_e_objetivos	0,825887512
qp_i1_grau_dificuldade_prova_parte_Formacao_Geral x qp_i8_considerando_questoes_objetivas_percebeu_que	0,814227348
qp_i1_grau_dificuldade_prova_parte_Formacao_Geral x qp_i6_informacoes_e_instrucoes_p_resol_questoes_foram_suficientes_p_resolvelas	0,811641617
qp_i1_grau_dificuldade_prova_parte_Formacao_Geral x qp_i4_enunciados_questoes_parte_Formacao_Geral_claros_e_objetivos	0,8025432
qp_i2_grau_dificuldade_prova_parte_Componente_Especifico x qp_i9_tempo_gasto_p_concluir_a_prova	0,794947721
qp_i5_enunciados_questoes_parte_Componente_Especifico_claros_e_objetivos x qp_i9_tempo_gasto_p_concluir_a_prova	0,793151879
qp_i8_considerando_questoes_objetivas_percebeu_que x qp_i9_tempo_gasto_p_concluir_a_prova	0,790264697
qp_i3_duracao_da_prova_considerando_extensao x qp_i9_tempo_gasto_p_concluir_a_prova	0,78485308

3.9 Mineração de Dados utilizando o Weka

Para realizar a mineração de dados foi utilizado o software Weka, conforme descrito na Seção 2.3. O algoritmo utilizado para a classificação dos dados foi o JRip [26].

3.9.1 Preparação dos Dados

Após entender como se utiliza o algoritmo e o formato necessário dos dados, iniciou-se a transformação destes dados.

Alguns atributos presentes no arquivo .arff inicial gerado, como descrito na Seção 2.3.1, foram removidos (do arquivo) para se encaixar no modelo desejado, por exemplo, os atributos "vt_gab_oce_orig", "vt_gab_oce_fin", "vt_esc_ofg", "vt_ace_ofg", "vt_esc_oce", "vt_ace_oce" foram removidos pois são do tipo *string* e o algoritmo JRip não trabalha com atributos do tipo *string*.

Os atributos "tp_pres", "tp_pr_ger", "tp_pr_ob_fg", "tp_pr_di_fg", "tp_pr_ob_ce", "tp_pr_di_ce", "tp_sfg_d1", "tp_sfg_d2", "tp_sce_d1", "tp_sce_d2", "tp_sce_d3" foram removidos pois são relacionados à presença do aluno em sala de aula ou se ele respondeu uma determinada questão discursiva ou não, o que estaria diretamente relacionado a nota. Para evitar essa relação trivial, estes atributos foram removidos.

Além destes, os atributos listados a seguir foram removidos pois se relacionam diretamente com "nt_ce":

```
form_gera_q1 a form_gera_q8
form_espc_q1 a form_espc_q27
nt_obj_fg
nt_fg_d1_pt
nt_fg_d1_ct
nt_fg_d1
nt_fg_d2_pt
nt_fg_d2_ct
nt_fg_d2
nt_dis_fg
nt_fg
nt_obj_ce
nt_ce_d1
nt_ce_d2
nt_ce_d3
nt_dis_ce
nt_ger"
```

Além das remoções, alguns dados foram discretizados. O JRip não trabalha com atributos numéricos, então foi necessária uma transformação desses atributos numéricos para atributos nominais. Para isso foi utilizado o filtro de pré processamento que o próprio weka fornece no pacote: `weka.filters.unsupervised.attribute Discretize`. O padrão colocado foi de 2 *bins* (divide os valores em duas classes) e `equal frequency = true` (distribui os atributos de forma que em cada classe exista uma quantidade igual de atributos, neste caso específico, como está dividido em duas classes, essa divisão seria equivalente à mediana 3.5.1). Ou seja, em um grupo ficam os valores inferiores a mediana e em outro os superiores ou iguais a mediana.

Depois da manipulação ser feita e o arquivo `arff` utilizado no Weka, foi colocado

“nt_ce” como a variável a ser descoberta. A nota do Componente Específico foi escolhida como a variável pois a nota de Formação Geral não representaria tão bem o curso especificamente, e a nota Geral seria afetada pela nota de Formação Geral, então foi julgada a nota CE como melhor opção. Após essa decisão, o classificador JRip pôde então ser executado.

3.9.2 Análise da mineração

O algoritmo foi executado para os cursos de Ciência da Computação Bacharel, Engenharia da Computação e Computação Licenciatura separadamente. Uma análise mais aprofundada foi feita no resultado dos alunos de CiC bacharel,

O número de instâncias foi de 9822;

O número de atributos analisados foi de 122;

O número de instâncias classificadas foi de 8276;

O número de instâncias classificadas corretamente foi de 5725 (69,17%);

O número de instâncias classificadas incorretamente foi de 2551 (30,82%);

O número de regras descobertas foi de 10, e foram selecionadas cinco para exposição, sendo elas:

- (i13_tipo_bolsa_academica = Bolsa de iniciação científica.) => nt_ce='(37.95-inf)' (57.0/25.0)
- (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (co_orgac = Universidade) and (i17_tipo_ensino_medio = Todo em escola privada (particular).) => nt_ce='(37.95-inf)' (679.0/262.0)
- (co_orgac = Universidade) and (in_vespertino = 1) and (i44_oferecidas_oportunidades_p_projetos_iniciacao_cientifica_atividades_estimulo_investigacao_cientifica = Concordo Totalmente) and (i25_motivo_escolha_curso = Vocação.) => nt_ce = '(37.95 - inf)' (178.0/34.0)
- (co_orgac = Universidade) and (i13_tipo_bolsa_academica = Bolsa de iniciação científica.) => nt_ce='(37.95-inf)' (291.0/96.0)
- (i25_motivo_escolha_curso = Vocação.) and (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (i18_modalidade_ensino_medio = Profissionalizante técnico (eletrônica, contabilidade, agrícola, outro).) => nt_ce = '(37.95-inf)' (79.0/27.0)

Com essas regras, algumas análises são possíveis, como por exemplo na regra "(i13_tipo_bolsa_academica = Bolsa de iniciação científica.) => nt_ce='(37.95 - inf)' (57.0

/ 25.0)"podemos interpretar que, na questão qe_i13 do questionário do estudante, que diz:

13. Ao longo da sua trajetória acadêmica, você recebeu algum tipo de bolsa acadêmica? No caso de haver mais de uma opção, marcar apenas a bolsa de maior duração.

- A () Nenhum.
- B () Bolsa de iniciação científica.
- C () Bolsa de extensão.
- D () Bolsa de monitoria/tutoria.
- E () Bolsa PET.
- F () Outro tipo de bolsa acadêmica.

os estudantes que marcaram a opção "Bolsa de iniciação científica" como resposta obtiveram uma nota na parte de Componente Específico superior ou igual a 37.95. Essa regra vale para 57 instâncias, já nas outras 25 que seguiram os padrões desta regra, os estudantes obtiveram nota inferior a 37.95 como é mostrado no final da regra "(57.0/25.0)".

Outro exemplo é a regra "(qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (co_orgac = Universidade) and (i17_tipo_ensino_medio = Todo em escola privada (particular).) => nt_ce='(37.95-inf)' (679.0/262.0)" que basicamente diz que, para estudantes que marcaram a opção "Estudou e aprendeu muitos desses conteúdos" como resposta na questão qp_i8, e também marcaram "Universidade" como Código da organização acadêmica da IES, e também cursaram o ensino médio todo em escola privada (particular), obtiveram uma nota superior ou igual a 37.95 na parte de Componente Específico em 679 casos, e inferior a 37.95 em 262 casos, como é informado no final da regra (679.0/262.0).

Além destes cursos, o algoritmo também foi executado utilizando os dados do ENADE de forma geral, contudo, o volume de dados é tão extenso que apenas uma parcela de 10% dos dados totais conseguiu ser processada pelo classificador pois a exigência de memória e o tempo de processamento não permitiram processar todos os dados do ENADE. Para isso foi utilizado o filtro de pré processamento que o próprio weka fornece no pacote weka.filters.unsupervised.instance.Resample. "No replacement" = *true* (esse filtro impede que a mesma instância seja utilizada duas vezes) e "sampleSizePercent" = 10.0 (esse filtro pega 10% da amostra total para fazer o processamento).

Para estes dados do ENADE geral, puderam ser observados os seguintes resultados:
O número de instâncias foi de 48172;
O número de atributos analisados foi de 122;

O número de instâncias classificadas foi de 39726;

O número de instâncias classificadas corretamente foi de 25632 (64,52%);

O número de instâncias classificadas incorretamente foi de 14094 (35,47%);

O número de regras descobertas foi de 31, sendo selecionadas as seguintes cinco para exposição:

- (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (i17_tipo_ensino_medio = Todo em escola privada (particular).) and (in_matutino = 1) and (qp_i9_tempo_gasto_p_concluir_a_prova = Entre três e quatro horas.) => nt_ce='(40.35-inf)' (226.0/69.0)
- (i2_raca_etnia = Branco(a).) and (i17_tipo_ensino_medio = Todo em escola privada (particular).) and (i11_tipo_bolsa_financiamento = Nenhum, pois meu curso é gratuito.) and (i53_oportunidades_p_estudantes_realizarem_intercambios_estagios_fora_do_pais = Concordo Totalmente) => nt_ce='(40.35-inf)' (312.0 / 108.0)
- (i17_tipo_ensino_medio = Todo em escola privada (particular).) and (i26_razao_escolha_inst_ensino_sup = Qualidade/reputação.) and (nu_idade = '(-inf-27.5]') and (qp_i9_tempo_gasto_p_concluir_a_prova = Entre três e quatro horas.) => nt_ce='(40.35-inf)' (108.0/39.0)
- (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (nu_idade = '(-inf-27.5]') and (co_orgac = Universidade) and (i13_tipo_bolsa_academica = Bolsa de iniciação científica.) => nt_ce='(40.35-inf)' (443.0/163.0)
- (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (qp_i9_tempo_gasto_p_concluir_a_prova = Entre três e quatro horas.) and (i17_tipo_ensino_medio = Todo em escola privada (particular).) => nt_ce='(40.35-inf)' (397.0/165.0)

De forma semelhante às análises feitas anteriormente para o curso de Ciência da Computação, algumas regras interessantes podem ser analisadas para o Enade geral.

Por exemplo, a regra:

"(i2_raca_etnia = Branco(a).) and (i17_tipo_ensino_medio = Todo em escola privada (particular).) and (i11_tipo_bolsa_financiamento = Nenhum, pois meu curso é gratuito.) and (i53_oportunidades_p_estudantes_realizarem_intercambios_estagios_fora_do_pais = Concordo Totalmente) => nt_ce='(40.35-inf)' (312.0/108.0)".

Nesta regra tem-se que, quando a raça ou etnia do participante foi marcada como "Branco(a)", o tipo de ensino médio foi "Todo em escola privada (particular)", o tipo de bolsa ou financiamento foi "Nenhum, pois meu curso é gratuito" e a opção para se houverem oportunidades para os estudantes realizarem intercâmbios ou estágios fora do país foi marcada como "Concordo totalmente", isso significa que 312 instâncias que estavam de acordo com esta regra ficaram com a nota superior a 40,35, e 108 instâncias com as mesmas características ficaram com a nota inferior a 40,35 como nos mostra "(312.0/108.0)" ao final da regra.

Além de fazer este tipo de análise para a nota de componente específico para todos os alunos do ENADE, foi julgado interessante realizar uma análise para as notas de formação geral para este mesmo grupo de alunos. Fizemos essa análise porque todos os alunos, independente do curso, tiveram as mesmas questões na parte de formação geral.

Nesta nova amostra, pode-se ser observado o seguinte:

O número de instâncias foi de 48172;

O número de atributos analisados foi de 122;

O número de instâncias classificadas foi de 39726;

O número de instâncias classificadas corretamente foi de 24511 (61,7001%);

O número de instâncias classificadas incorretamente foi de 15215 (38,2999%);

O número de regras descobertas foi de 25.

Dentre as regras encontradas, pode-se serem usadas cinco como exemplo, e são elas:

- (in_vespertino = 0) and (qp_i9_tempo_gasto_p_concluir_a_prova = Menos de uma hora.) => nt_fg='(-inf-55.55]' (257.0/50.0)
- (nu_idade = '(27.5-inf)') and (i8_renda_familia = De 1,5 a 3 salários mínimos (R\$ 1.086,01 a R\$ 2.172,00).) and (i11_tipo_bolsa_financiamento = Nenhum, embora meu curso não seja gratuito.) => nt_fg='(-inf-55.55]' (831.0/358.0)
- (i73_instituicao_estagio_obrg = Escola pública.) and (i8_renda_familia = Até 1,5 salário mínimo (até R\$ 1.086,00).) and (i11_tipo_bolsa_financiamento = Nenhum, embora meu curso não seja gratuito.) => nt_fg='(-inf-55.55]' (201.0/57.0)
- (i73_instituicao_estagio_obrg = Escola pública.) and (tp_sexo = F) and (i13_tipo_bolsa_academica = Nenhum.) and (i2_raca_etnia = Pardo(a)/mulato(a).) and (i8_renda_familia = De 1,5 a 3 salários mínimos (R\$ 1.086,01 a R\$ 2.172,00).) => nt_fg='(-inf-55.55]' (869.0/359.0)
- (i17_tipo_ensino_medio = Todo em escola pública.) and (i23_qts_horas_estudo_a_parte_p_semana = De uma a três.) and (co_catad = Privada com fins lucrativos) and (i81_acompanhamento_adequado_professores = Sim, durante todo o

tempo.) and (i78_fundamentacao_teorica_suficiente_p_docencia = Sim, completamente.) => nt_fg='(-inf-55.55]' (123.0/39.0)

Como exemplo de análise de regra, pode-se citar a regra "(in_vespertino = 0) and (qp_i9_tempo_gasto_p_concluir_a_prova = Menos de uma hora.) => nt_fg='(-inf-55.55]' (257.0/50.0)", por exemplo.

Nesta regra pode-se observar que, quando o turno vespertino for marcado como "0", o que significa que o turno não é o vespertino (e sim o matutino), e o tempo gasto para concluir a prova foi "Menos de uma hora", isso significa que 257 instâncias ficaram com a nota de até 55,55 e 50 instâncias ficaram com nota superior a 55,55.

Capítulo 4

Conclusões e Trabalhos Futuros

Neste trabalho foram realizadas análises e estudos acerca de dados referentes ao Exame nacional de desenvolvimento dos estudantes (ENADE), que tem como objetivo avaliar o rendimento dos alunos concluintes dos cursos de graduação, definindo assim a qualidade dos cursos. Mais especificamente, os dados utilizados foram os do Enade 2014.

Inicialmente foram coletados os dados do Enade 2014 (que podem ser baixados diretamente do site do Inep), contudo, os dados não estavam em um formato adequado para sua análise, então iniciou-se a fase de tratamento dos dados. Nesta fase os dados foram colocados em um padrão adequado para ser utilizado pelo software Weka, e também foram formatados de forma a ser mais simples e clara sua visualização e análise. Para realizar este trabalho, foi utilizada a linguagem Python.

Depois de todos os atributos e variáveis estarem devidamente tratados e colocados em um formato padrão, pôde-se começar o estudo e as análises. Vários tipos de análises foram feitas, principalmente relacionando as notas (geral, de componente específico e de formação geral) com os demais atributos.

Inicialmente, foram calculadas para o curso de Ciência da Computação e para o Enade como um todo, para cada uma das notas, valores tais como as medianas, 25 quartis e 75 quartis. Depois, foram feitas três tabelas separando estes dados de acordo com as notas (FG, CE e GER), e com os atributos indicados na tabela, então obteve-se informação de valores de quartis e medianas para os alunos que marcaram cada atributo separadamente. Com isso pôde-se procurar com mais detalhes algum padrão significativo entre os atributos e as notas. Destes atributos, foram selecionados os que têm as maiores medianas para cada uma das notas, então foi feita uma análise de quais atributos se repetiam para as maiores medianas de cada uma das notas.

Foram gerados *boxplots* para estes dados na intenção de auxiliar na visualização destes dados de uma forma alternativa. Esta ferramenta é bastante útil na análise visual dos dados, mostrando de forma clara também os quartis dos atributos. Foram feitas também

tabelas para cada um destes *boxplots* com a intenção de facilitar e visualizar com mais exatidão os valores dos quartis e da mediana.

Foram separados os alunos com a renda mais alta e com a renda mais baixa em duas tabelas diferentes. Essas tabelas contém quantas vezes cada opção de cada atributo foi escolhido. Por exemplo, o atributo "sexo" tem as opções: M (masculino) e F (feminino). Na tabela que diz respeito a renda mais alta pode ser visto que 3233 alunos marcaram a opção 'M'. Ambas as tabelas foram ordenadas pelas opções mais marcadas e esse resultado foi utilizado para analisar o perfil desses estudantes.

Foram geradas, para todos os cursos, tabelas contendo as medianas das notas (geral, componente específico e formação geral) e alguns atributos como renda, etnia, tipo de ensino médio, entre outros. As opções desses atributos foram combinados com os atributos em si, por exemplo, para o atributo renda tem-se a opção “renda de 1,5 a 3 salários mínimos” e para o atributo etnia tem-se a opção "negro", o que permite fazer as combinações: nota com a opção "negro", com a opção “renda de 1,5 a 3 salários mínimos” e assim por diante. Essas combinações trazem informações sobre as médias dos participantes com rendas mais altas, rendas mais baixas, ensino médio em escola pública ou privada, etc.

Foi feita também uma análise do desempenho dos alunos de Ciência da Computação da UnB em cada uma das questões de componente específico da prova, sendo assim possível avaliar o aprendizado dos alunos em cada questão. Esse desempenho foi comparado com estudantes das 10 universidades melhor classificadas.

Foi feita uma tabela com as medianas das notas (FG, GER e CE) para os cursos de Computação bacharel e licenciatura, engenharia da computação e o Enade geral. Tendo como referência a matéria “Cotista tem nota boa na universidade, mas recua em cursos com matemática” do jornal Folha de São Paulo de 10 de dezembro de 2017 que compara estudantes cotistas com não cotistas, foram acrescentados a esta tabela combinações entre essas notas e algumas das opções dos atributos: renda, etnia, tipo de ensino médio cursado e se entrou na universidade por algum tipo de ação afirmativa. O objetivo era comparar as medianas das notas de alunos com diferentes etnias, rendas, etc.

Outra transformação dos dados foi necessária para os passos seguintes. Os dados foram transformados para valores numéricos e colocados no padrão necessário.

Com isso, foram calculadas as médias aritméticas dos atributos envolvidos para cada um dos cursos. Logo, para cada curso, cada atributo teria uma média associada. Isso permitiu diversas análises como comparar os cursos com maior renda, comparar os cursos com mais horas de estudo a parte, entre outros.

Em seguida, esses atributos foram correlacionados entre si utilizando uma função de uma biblioteca do Python. Agora, para cada curso, a cada dois destes atributos haveria uma correlação entre eles. Foram separados os alunos de instituições públicas e privadas

em duas tabelas diferentes, e novamente feitas as correlações. Uma outra tabela foi gerada, onde agora o valor é o valor da correlação da tabela de instituições públicas menos o valor desta correlação da tabela de instituições privadas, ou seja a subtração do valor da correlação de instituições públicas - (menos) privadas.

Heatmaps evidenciam estas correlações, e foram criados para cada uma dessas correlações de cada um dos cursos, para melhorar a visualização destas correlações.

Para utilizar o algoritmo de mineração de dados do Weka, novamente os dados foram tratados e colocados no formato necessário para sua aplicação. O algoritmo de classificação JRIP foi executado para os cursos de Ciência da Computação Bacharel, Engenharia da Computação e Computação Licenciatura separadamente, gerando regras que relacionassem a nota da formação geral (FG) e a nota do componente específico (CE) com os demais atributos. O algoritmo encontrou diversas regras como por exemplo que alunos que não estudam no turno vespertino e que gastaram menos de uma hora para concluir a prova tiraram uma nota FG menor ou igual a mediana. Outros atributos, como renda, idade, etnia entre outros, também foram encontrados em regras semelhantes (que relacionam estes atributos com as notas).

Nesse trabalho foram descobertas muitas informações sobre os dados do ENADE. Foram geradas tabelas, gráficos, *heatmaps*, regras de classificação. Foram também analisados vários atributos que influenciam o desempenho dos alunos na prova do ENADE. Outros fatores foram analisados como correlações entre os atributos. Estas informações podem ser úteis para entender o perfil dos alunos que fazem a prova do ENADE e indicar alguns fatores que podem influenciar o desempenho destes alunos.

Todo o projeto pode ser encontrado em: https://github.com/MateusDenucci/Enade_TG

Como sugestão de trabalhos futuros, pode-se aprofundar estas análises fazendo mais comparações e relações entre os atributos. Além disso, pode-se utilizar os microdados do ENADE de anos diferentes e analisar a evolução do desempenho dos estudantes ao longo do tempo.

Referências

- [1] IAN H. WITTEN, EIBE FRANK, MARK A. HALL CHRISTOPHER J. PAL: *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2016. ix, 6, 12
- [2] SETZER, VALDEMAR W.: *Dado, informação, conhecimento e competência*. <https://www.ime.usp.br/~vwsetzer/dado-info.html>. 4
- [3] REZENDE, PEDRO A. D.: *Modelos de confiança para segurança em informática*. https://cic.unb.br/~rezende/trabs/modelos_de_confianca.pdf. 4
- [4] FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMITH P.: *From Data Mining to Knowledge Discovery: An Overview*. In *Advances in Knowledge Discovery and Data Mining*. AAAI Press, 1996. 4
- [5] JIAWEI HAN, MICHELINE KAMBER, JIAN PEI: *Data Mining, Concepts and Techniques*. 5
- [6] *Ibm spss modeler crisp-dm guide*. https://inseaddataanalytics.github.io/INSEADAnalytics/CRISP_DM.pdf. 5
- [7] PETE CHAPMAN, JULIAN CLINTON, RANDY KERBER ET AL.: *CRISP-DM 1.0 Step-by-step data mining guide*. Technical report, The Crisp-Dm consortium, 2000. 5
- [8] JAIN, A.K., M.N. MURTY e P.J. FLYNN: *Data clustering: A review*. <https://www.cs.rutgers.edu/~mlittman/courses/lightai03/jain99data.pdf>. 8
- [9] CHAPPLE, MIKE: *Classification in data mining*. <https://www.thoughtco.com/classification-1019653>, acesso em 2016-25-11. 8
- [10] ABERNETHY, MICHAEL: *Classificação e armazenamento em cluster*. <https://www.ibm.com/developerworks/br/opensource/library/os-weka2/index.html>, acesso em 2010-28-05. 8
- [11] EIBE FRANK, MARK A. HALL e IAN H. WITTEN: *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"*. Morgan Kaufmann, 2016. 8
- [12] *weka - arff*. <https://weka.wikispaces.com/ARFF>. 9
- [13] *Enade - inep*. <http://portal.inep.gov.br/enade>, acesso em 2015-20-10. 11

- [14] L. G. F. SILVA, M. E. P. S. ROCHA e R. A. A. FAGUNDES: *Enade: Math and science students' performance analysis*. http://www.revistaieeela.pea.usp.br/issues/vol15issue9Sept.2017/15TLA9_25Silva.pdf. 11
- [15] NICOLAS P. BOA VISTA¹, MICHELE F. FIQUEIRÓ², PATRICIA M. MOZZAQUATRO³: *Técnicas de mineração de dados aplicadas aos microdados do enade para avaliar o desempenho dos acadêmicos do curso de ciência da computação no rio grande do sul utilizando o software r*. <http://revistaeletronica.unicruz.edu.br/index.php/revistaeletronica/article/view/5401>. 11
- [16] CRETTON, NÍCOLLAS NOGUEIRA: *MineraÇÃO de dados aplicada na base do enade com enfoque na criação de perfis dos estudantes que prestaram o exame utilizando o algoritmo j48*. <https://mpoic.ucam-campos.br/wp-content/uploads/2017/07/N%C3%ADcollas-Nogueira-Cretton.pdf>. 12
- [17] DE OLIVEIRA, ALINE e LORENA PEREIRA: *Sistema de cotas*. <http://sistema-de-cotas.info/>. 13
- [18] *Médias - só matemática*. <http://www.somatematica.com.br/fundam/medias.php>. 17
- [19] *Moda e mediana*. <http://www.somatematica.com.br/estat/basica/pagina6.php>. 17
- [20] *Python data analysis library - pandas*. <https://pandas.pydata.org/>. 19, 51
- [21] *Boxplots*. <http://www.portalaction.com.br/estatistica-basica/31-boxplot>. 19
- [22] *Nota técnica daes/inep nº 57/2015*. http://www.ufrgs.br/sai/avaliacao-externa/nacional/arquivos_nacional/nota%20tecnica%20calcula%20cpc%20igc.pdf, acesso em 2015-27-10. 34
- [23] PATI, CAMILA: *Os melhores cursos superiores de computação do brasil*. <https://exame.abril.com.br/carreira/os-melhores-cursos-superiores-de-computacao-do-brasil/>, acesso em 2015-23-12. 34
- [24] *Cotista tem nota boa, mas recua em cursos com matemática*. <http://www1.folha.uol.com.br/educacao/2017/12/1942100-cotista-tem-nota-boa-na-universidade-mas-recua-em-cursos-com-matematica.shtml>. 41
- [25] *Coefficientes de correlação*. <http://www.abgconsultoria.com.br/blog/coeficientes-de-correlacao/>. 51
- [26] *Jrip algorithm*. https://www.eecs.yorku.ca/tdb/_doc.php/userg/sw/weka/doc/weka/classifiers/rules/JRip.html. 59

Apêndice A

Nesta seção se encontram os anexos referentes a materiais que não foram disponibilizados em seções anteriores.

A.1 Regras Weka JRip

A.1.1 Ciência da Computação (bacharel)

Todas as regras encontradas para o curso de Ciência da Computação (bacharel) com base na nota de componente específico foram:

- (i13_tipo_bolsa_academica = Bolsa de iniciação científica.) => nt_ce='(37.95-inf)' (57.0/25.0)
- (co_orgac = Universidade) and (in_vespertino = 1) and (i44_oferecidas_oportunidades_p_projetos_iniciacao_cientifica_atividades_estimulo_investigacao_cientifica = Concordo Totalmente) and (i25_motivo_escolha_curso = Vocaç o.) => nt_ce='(37.95-inf)' (178.0/34.0)
- (co_catad = Pessoa Jur dica de Direito P blico - Federal) and (in_matutino = 1) => nt_ce='(37.95-inf)' (1621.0/267.0)
- (co_orgac = Universidade) and (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conte dos.) and (i11_tipo_bolsa_financiamento = Nenhum, pois meu curso   gratuito.) and (ano_fim_2g = '(2007.5-inf)') and (co_regiao_curso = Sudeste) => nt_ce='(37.95-inf)' (129.0/13.0)
- (i11_tipo_bolsa_financiamento = Nenhum, pois meu curso   gratuito.) and (co_catad = Pessoa Jur dica de Direito P blico - Federal) and (qp_i9_tempo_gasto_p_concluir_a_prova = Entre tr s e quatro horas.) => nt_ce='(37.95-inf)' (96.0/16.0)

- (co_orgac = Universidade) and (in_vespertino = 1) and (i44_oferecidas_oportunidades_p_projetos_iniciacao_cientifica_atividades_estimulo_investigacao_cientifica = Concordo Totalmente) and (i25_motivo_escolha_curso = Vocação.) => nt_ce='(37.95-inf)' (178.0/34.0)
- (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (co_orgac = Universidade) and (i17_tipo_ensino_medio = Todo em escola privada (particular).) => nt_ce='(37.95-inf)' (679.0/262.0)
- (co_orgac = Universidade) and (i13_tipo_bolsa_academica = Bolsa de iniciação científica.) => nt_ce='(37.95-inf)' (291.0/96.0)
- (co_orgac = Universidade) and (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (i50_estagio_supervisionado_proporcionou_experiencias_diversificadas_p_formacao = Não sei responder) => nt_ce='(37.95-inf)' (240.0/96.0)
- (qp_i9_tempo_gasto_p_concluir_a_prova = Entre três e quatro horas.) and (co_curso = '(-inf-20035.5]') => nt_ce='(37.95-inf)' (521.0/255.0)
- (i25_motivo_escolha_curso = Vocação.) and (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (i18_modalidade_ensino_medio = Profissionalizante técnico (eletrônica, contabilidade, agrícola, outro).) => nt_ce='(37.95-inf)' (79.0/27.0)

Engenharia da Computação

Todas as regras encontradas para o curso de Engenharia da Computação com base na nota de componente específico foram:

- (i11_tipo_bolsa_financiamento = Nenhum, pois meu curso é gratuito.) and (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) => nt_ce='(40.9-inf)' (616.0/149.0)
- (in_matutino = 1) and (co_catad = Pessoa Jurídica de Direito Público - Federal) and (qp_i9_tempo_gasto_p_concluir_a_prova = Entre três e quatro horas.) => nt_ce='(40.9-inf)' (75.0/11.0)
- (in_matutino = 1) and (i26_razao_escolha_inst_ensino_sup = Qualidade / reputação.) => nt_ce='(40.9-inf)' (381.0/159.0)
- (qp_i9_tempo_gasto_p_concluir_a_prova = Entre três e quatro horas.) and (ano_fim_2g = '(2007.5-inf)') and (qp_i4_enunciados_questoes_parte_Formacao

_Geral_claros_e_objetivos = Sim, a maioria.) and (i58_professores_utilizaram_tecnologias_da_informacao_comunicacao_como_estrategia_de_ensino = Concorde Totalmente) => nt_ce='(40.9-inf)' (146.0/54.0)

- (i25_motivo_escolha_curso = Vocaç o.) and (qp_i1_grau_dificuldade_prova_parte_Formacao_Geral = F cil.) => nt_ce='(40.9-inf)' (79.0/29.0)
- (co_regiao_curso = Sul) and (i29_metodologias_no_curso_instigaram_a_aprofundar_conhecimentos = Concorde) => nt_ce='(40.9-inf)' (57.0/15.0)

A.1.2 Ci ncia da Computa o (licenciatura)

Todas as regras encontradas para o curso de Ci ncia da Computa o (licenciatura) com base na nota de componente espec fico foram:

- (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conte dos.) => nt_ce='(49.25-inf)' (765.0/278.0)
- (i28_disciplinas_favoreceram_atuacao_estagio_iniciacao_profissional = Concorde) and (qp_i5_enunciados_questoes_parte_Componente_Especifico_claros_e_objetivos = Sim, a maioria.) => nt_ce='(49.25-inf)' (126.0/44.0)

A.1.3 Enade Geral nota CE

Todas as regras encontradas para os alunos do Enade geral (levando em conta apenas 10% da amostragem total como explicado previamente) tendo como base apenas a nota de componente espec fico foram:

- (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conte dos.) and (i17_tipo_ensino_medio = Todo em escola privada (particular).) and (in_matutino = 1) and (qp_i9_tempo_gasto_p_concluir_a_prova = Entre tr s e quatro horas.) => nt_ce='(40.35-inf)' (226.0/69.0)
- (i2_raca_etnia = Branco(a).) and (i17_tipo_ensino_medio = Todo em escola privada (particular).) and (i11_tipo_bolsa_financiamento = Nenhum, pois meu curso   gratuito.) and (i53_oportunidades_p_estudantes_realizarem_intercambios_estagios_fora_do_pais = Concorde Totalmente) => nt_ce='(40.35-inf)' (312.0 / 108.0)
- (i17_tipo_ensino_medio = Todo em escola privada (particular).) and (i26_razao_escolha_inst_ensino_sup = Qualidade/reputa o.) and (nu_idade = '(-inf-27.5]') and (qp_i9_tempo_gasto_p_concluir_a_prova = Entre tr s e quatro horas.) => nt_ce='(40.35-inf)' (108.0/39.0)

- (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (nu_idade = '(-inf-27.5]') and (co_orgac = Universidade) and (i13_tipo_bolsa_academica = Bolsa de iniciação científica.) => nt_ce='(40.35-inf)' (443.0/163.0)
- (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (qp_i9_tempo_gasto_p_concluir_a_prova = Entre três e quatro horas.) and (i17_tipo_ensino_medio = Todo em escola privada (particular).) => nt_ce='(40.35-inf)' (397.0/165.0)
- (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (co_grupo = PEDAGOGIA (LICENCIATURA)) => nt_ce='(40.35-inf)' (6331.0/1805.0)
- (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (co_catad = Pessoa Jurídica de Direito Público - Federal) and (i17_tipo_ensino_medio = Todo em escola privada (particular).) => nt_ce='(40.35-inf)' (1621.0/406.0)
- (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (nu_idade = '(-inf-27.5]') and (qp_i9_tempo_gasto_p_concluir_a_prova = Entre três e quatro horas.) and (tp_sexo = M) => nt_ce='(40.35-inf)' (1800.0/617.0)
- (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (co_catad = Pessoa Jurídica de Direito Público - Federal) and (co_regiao_curso = Sudeste) and (ano_in_grad = '(2010.5-inf)') => nt_ce='(40.35-inf)' (256.0/63.0)
- (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (i17_tipo_ensino_medio = Todo em escola privada (particular).) and (in_matutino = 1) and (qp_i9_tempo_gasto_p_concluir_a_prova = Entre três e quatro horas.) => nt_ce='(40.35-inf)' (226.0/69.0)
- (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (nu_idade = '(-inf-27.5]') and (i25_motivo_escolha_curso = Vocação.) and (in_matutino = 1) and (i26_razao_escolha_inst_ensino_sup = Qualidade/reputação.) => nt_ce='(40.35-inf)' (458.0/131.0)
- (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (qp_i9_tempo_gasto_p_concluir_a_prova = En-

- tre três e quatro horas.) and (i17_tipo_ensino_medio = Todo em escola privada (particular).) => nt_ce='(40.35-inf)' (397.0/165.0)
- (qp_i5_enunciados_questoes_parte_Componente_Especifico_claros_e_objetivos = Sim, a maioria.) and (nu_item_oce_x = '(-inf-5.5]') and (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu todos esses conteúdos.) => nt_ce='(40.35-inf)' (960.0/317.0)
 - (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (qp_i9_tempo_gasto_p_concluir_a_prova = Entre três e quatro horas.) and (co_curso = '(-inf-88306]') and (qp_i2_grau_dificuldade_prova_parte_Componente_Especifico = Médio.) and (i1_est_civil = Solteiro(a).) => nt_ce='(40.35-inf)' (506.0/204.0)
 - (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (i2_raca_etnia = Branco(a).) and (i30_curso_proporcionou_experiencias_de_aprendizado_inovadoras = Concordo Levemente) and (in_noturno = 0) => nt_ce='(40.35-inf)' (273.0/88.0)
 - (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (qp_i5_enunciados_questoes_parte_Componente_Especifico_claros_e_objetivos = Sim, a maioria.) and (co_regiao_curso = Sul) and (ano_in_grad = '(-inf-2010.5]') => nt_ce='(40.35-inf)' (604.0/224.0)
 - (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (nu_idade = '(-inf-27.5]') and (co_orgac = Universidade) and (i13_tipo_bolsa_academica = Bolsa de iniciação científica.) => nt_ce='(40.35-inf)' (443.0/163.0)
 - (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (nu_idade = '(-inf-27.5]') and (tpsexo = M) and (qp_i1_grau_dificuldade_prova_parte_Formacao_Geral = Médio.) and (qp_i9_tempo_gasto_p_concluir_a_prova = Quatro horas e não consegui terminar.) => nt_ce='(40.35-inf)' (183.0/56.0)
 - (nu_item_oce_x = '(-inf-5.5]') and (qp_i4_enunciados_questoes_parte_Formacao_Geral_claros_e_objetivos = Sim, a maioria.) and (qp_i9_tempo_gasto_p_concluir_a_prova = Entre três e quatro horas.) and (qp_i1_grau_dificuldade_prova_parte_Formacao_Geral = Médio.) and (qp_i5_enunciados_questoes_parte_Componente_Especifico_claros_e_objetivos = Sim, a maioria.) and (co_grupo = PEDAGOGIA (LICENCIATURA)) => nt_ce='(40.35-inf)' (218.0/74.0)

- (nu_item_oce_x = '(-inf-5.5]') and (qp_i4_enunciados_questoes_parte_Formacao_Geral_claros_e_objetivos = Sim, a maioria.) and (i26_razao_escolha_inst_ensino_sup = Qualidade/reputação.) and (co_catad = Pessoa Jurídica de Direito Público - Federal) => nt_ce='(40.35-inf)' (380.0/142.0)
- (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (i11_tipo_bolsa_financiamento = ProUni integral.) => nt_ce='(40.35-inf)' (536.0/200.0)
- (i2_raca_etnia = Branco(a).) and (i17_tipo_ensino_medio = Todo em escola privada (particular).) and (i11_tipo_bolsa_financiamento = Nenhum, pois meu curso é gratuito.) and (i53_oportunidades_p_estudantes_realizarem_intercambios_estagios_fora_do_pais = Concordo Totalmente) => nt_ce='(40.35-inf)' (312.0/108.0)
- (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (nu_idade = '(-inf-27.5]') and (tp_sexo = M) and (i8_renda_familia = De 6 a 10 salários mínimos (R\$ 4.344,01 a R\$ 7.240,00).) => nt_ce='(40.35-inf)' (330.0/148.0)
- (nu_item_oce_x = '(-inf-5.5]') and (co_grupo = PEDAGOGIA (LICENCIATURA)) and (i78_fundamentacao_teorica_suficiente_p_docencia = Sim, em grande parte.) and (qp_i6_informacoes_e_instrucoes_p_resol_questoes_foram_suficientes_p_resovelas = Sim, na maioria delas.) and (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou a maioria desses conteúdos, mas não os aprendeu.) => nt_ce='(40.35-inf)' (253.0/84.0)
- (nu_item_oce_x = '(-inf-5.5]') and (co_grupo = PEDAGOGIA (LICENCIATURA)) and (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu todos esses conteúdos.) and (qp_i7_alguma_dificuldade_ao_responder_a_prova = Não tive qualquer tipo de dificuldade para responder À prova.) => nt_ce='(40.35-inf)' (256.0/75.0)
- (nu_item_oce_x = '(-inf-5.5]') and (i2_raca_etnia = Branco(a).) and (i75_etapa_modalidade_de_ensino_do_estagio_obrg = Ensino Fundamental - anos iniciais.) and (co_grupo = PEDAGOGIA (LICENCIATURA)) and (i10_situacao_trabalho = Trabalho 40 horas semanais ou mais.) => nt_ce='(40.35-inf)' (250.0/95.0)
- (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (co_regiao_curso = Sudeste) and (i11_tipo_bolsa_financiamento = Nenhum, pois meu curso é gratuito.) => nt_ce='(40.35-inf)' (562.0/234.0)

- (qp_i9_tempo_gasto_p_concluir_a_prova = Entre três e quatro horas.) and (i2_raca_etnia = Branco(a).) and (qp_i2_grau_dificuldade_prova_parte_Componente_Especifico = Médio.) and (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu todos esses conteúdos.) => nt_ce='(40.35-inf)' (181.0/69.0)
- (nu_item_oce_x = '(-inf-5.5]') and (i25_motivo_escolha_curso = Vocação.) and (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou e aprendeu muitos desses conteúdos.) and (qp_i4_enunciados_questoes_parte_Formacao_Geral_claros_e_objetivos = Sim, a maioria.) => nt_ce='(40.35-inf)' (717.0/339.0)
- (nu_item_oce_x = '(-inf-5.5]') and (qp_i9_tempo_gasto_p_concluir_a_prova = Entre três e quatro horas.) and (co_grupo = EDUCAÇÃO FÍSICA (LICENCIATURA)) and (qp_i6_informacoes_e_instrucoes_p_resol_questoes_foram_suficientes_p_resolveras = Sim, na maioria delas.) => nt_ce='(40.35-inf)' (149.0/55.0)
- (i2_raca_etnia = Branco(a).) and (nu_item_oce_x = '(-inf-5.5]') and (co_regiao_curso = Sul) and (i11_tipo_bolsa_financiamento = Nenhum, pois meu curso é gratuito.) => nt_ce='(40.35-inf)' (338.0/149.0)
- (nu_item_oce_x = '(-inf-5.5]') and (co_grupo = PEDAGOGIA (LICENCIATURA)) and (i75_etapa_modalidade_de_ensino_do_estagio_obrg = Ensino Fundamental - anos iniciais.) and (co_regiao_curso = Sudeste) and (i80_orientacao_supervisao_professores = Sim, em grande parte do tempo.) and (ano_fim_2g = '(-inf-2006.5]') => nt_ce='(40.35-inf)' (93.0/27.0)
- (nu_idade = '(-inf-27.5]') and (i17_tipo_ensino_medio = Todo em escola privada (particular).) and (i26_razao_escolha_inst_ensino_sup = Qualidade/reputação.) and (i57_professores_demonstram_dominio_dos_conteudos_abordados = Concordo) => nt_ce='(40.35-inf)' (281.0/116.0)
- (i17_tipo_ensino_medio = Todo em escola privada (particular).) and (i26_razao_escolha_inst_ensino_sup = Qualidade/reputação.) and (nu_idade = '(-inf-27.5]') and (qp_i9_tempo_gasto_p_concluir_a_prova = Entre três e quatro horas.) => nt_ce='(40.35-inf)' (108.0/39.0)
- (i2_raca_etnia = Branco(a).) and (qp_i1_grau_dificuldade_prova_parte_Formacao_Geral = Médio.) and (co_regiao_curso = Sul) and (i8_renda_familia = De 6 a 10 salários mínimos (R\$ 4.344,01 a R\$ 7.240,00).) => nt_ce='(40.35-inf)' (194.0/77.0)

- (nu_item_oce_x = '(-inf-5.5]') and (co_grupo = PEDAGOGIA (LICENCIATURA)) and (i32_curso_deu_oportunidade_p_trabalhar_em_equipe = Concorde) and (i66_atividades_academicas_possibilitaram_reflexao_convivencia_respeito_a_diversidade = Concorde Totalmente) => nt_ce='(40.35-inf)' (145.0/52.0)

A.1.4 Enade Geral nota FG

Todas as regras encontradas para os alunos do Enade geral (levando em conta apenas 10% da amostragem total como explicado previamente) tendo como base apenas a nota de formação geral foram:

- (co_grupo = PEDAGOGIA (LICENCIATURA)) and (i8_renda_familia = Até 1,5 salário mínimo (até R\$ 1.086,00).) => nt_fg='(-inf-55.55]' (3031.0/816.0)
- (i73_instituicao_estagio_obrg = Escola pública.) and (i79_experiencias_que_gostaria_de_proporcionar = Sim, durante todo o tempo.) and (i10_situacao_trabalho = Trabalho até 20 horas semanais.) and (nu_idade = '(27.5-inf)') => nt_fg='(-inf-55.55]' (532.0/115.0)
- (i73_instituicao_estagio_obrg = Escola pública.) and (i79_experiencias_que_gostaria_de_proporcionar = Sim, durante todo o tempo.) and (i72_etapa_modalidade_magisterio = Educação Infantil.) => nt_fg='(-inf-55.55]' (1206.0/379.0)
- (i73_instituicao_estagio_obrg = Escola pública.) and (i30_curso_proporcionou_experiencias_de_aprendizado_inovadoras = Concorde Totalmente) and (co_grupo = PEDAGOGIA (LICENCIATURA)) and (i23_qts_horas_estudo_a_part_e_p_semana = De uma a três.) => nt_fg='(-inf-55.55]' (1644.0/563.0)
- (i73_instituicao_estagio_obrg = Escola pública.) and (i79_experiencias_que_gostaria_de_proporcionar = Sim, durante todo o tempo.) and (i75_etapa_modalidade_de_ensino_do_estagio_obrg = Educação Infantil.) => nt_fg='(-inf-55.55]' (361.0/123.0)
- (i73_instituicao_estagio_obrg = Escola pública.) and (i13_tipo_bolsa_academica = Nenhum.) and (i8_renda_familia = Até 1,5 salário mínimo (até R\$ 1.086,00).) and (nu_idade = '(27.5-inf)') => nt_fg='(-inf-55.55]' (938.0/322.0)
- (i73_instituicao_estagio_obrg = Escola pública.) and (i30_curso_proporcionou_experiencias_de_aprendizado_inovadoras = Concorde Totalmente) and (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou alguns desses conteúdos, mas não os aprendeu.) and (co_grupo = PEDAGOGIA (LICENCIATURA)) => nt_fg='(-inf-55.55]' (157.0/40.0)

- (i73_instituicao_estagio_obrg = Escola pública.) and (i30_curso_proporcionou _experiencias_de_aprendizado_inovadoras = Concordo Totalmente) and (co _grupo = PEDAGOGIA (LICENCIATURA)) and (i79_experiencias _que _gostaria _de _proporcionar = Sim, durante todo o tempo.) and (i44_oferecidas _oportunidades _p _projetos _iniciacao _cientifica_atividades _estimulo_investigacao_cientifica = Concordo Totalmente) => nt_fg='(-inf-55.55]' (256.0/104.0)
- (i73_instituicao_estagio_obrg = Escola pública.) and (i8_renda_familia = Até 1,5 salário mínimo (até R\$ 1.086,00).) and (i11_tipo_bolsa_financiamento = Nenhum, embora meu curso não seja gratuito.) => nt_fg='(-inf-55.55]' (201.0/57.0)
- (i77_onde_atuar_em_cinco_anos = Em escola pública, como professor.) and (i13_tipo_bolsa_academica = Nenhum.) and (qp_i6_informacoes_e_instrucoes_p _resol_questoes_foram_suficientes_p_resolvelas = Sim, somente em algumas.) and (nu_item_oce_x = '(-inf-5.5]') => nt_fg='(-inf-55.55]' (854.0/302.0)
- (qp_i9_tempo_gasto_p_concluir_a_prova = Entre uma e duas horas.) => nt_fg = '(-inf-55.55]' (5524.0/2461.0)
- (i73_instituicao_estagio_obrg = Escola pública.) and (i23_qts_horas_estudo_a _parte_p_semana = De uma a três.) and (i30_curso_proporcionou _experiencias_de _aprendizado_inovadoras = Concordo Totalmente) and (qp_i8 _considerando _questoes_objetivas_percebeu _que = Estudou alguns desses conteúdos, mas não os aprendeu.) => nt_fg='(-inf-55.55]' (275.0/108.0)
- (i73_instituicao_estagio_obrg = Escola pública.) and (tp_sexo = F) and (i13_tipo _bolsa_academica = Nenhum.) and (i2_raca_etnia = Pardo(a)/mulato(a).) and (i8_renda_familia = De 1,5 a 3 salários mínimos (R\$ 1.086,01 a R\$ 2.172,00).) => nt_fg='(-inf-55.55]' (869.0/359.0)
- (i73_instituicao_estagio_obrg = Escola pública.) and (i30_curso_proporcionou _experiencias_de_aprendizado_inovadoras = Concordo Totalmente) and (i25 _motivo _escolha _curso = Valorização profissional.) => nt_fg='(-inf-55.55]' (595.0 / 243.0)
- (i77_onde_atuar_em_cinco_anos = Em escola pública, como professor.) and (i72_etapa_modalidade_magisterio = Educação Infantil.) and (i37_relacoes _professor _aluno _estimularam _estudo_aprendizado = Concordo Totalmente) => nt_fg='(-inf-55.55]' (382.0/149.0)

- (nu_idade = '(27.5-inf)') and (i8_renda_familia = De 1,5 a 3 salários mínimos (R\$ 1.086,01 a R\$ 2.172,00).) and (i11_tipo_bolsa_financiamento = Nenhum, embora meu curso não seja gratuito.) => nt_fg='(-inf-55.55]' (831.0/358.0)
- (i17_tipo_ensino_medio = Todo em escola pública.) and (i23_qts_horas_estudo_a_parte_p_semana = De uma a três.) and (co_catad = Privada com fins lucrativos) and (i81_acompanhamento_adequado_professores = Sim, durante todo o tempo.) and (i78_fundamentacao_teorica_suficiente_p_docencia = Sim, completamente.) => nt_fg='(-inf-55.55]' (123.0/39.0)
- (i73_instituicao_estagio_obrg = Escola pública.) and (qp_i1_grau_dificuldade_prova_parte_Formacao_Geral = Difícil.) and (tp_sexo = F) and (i10_situacao_trabalho = Trabalho até 20 horas semanais.) => nt_fg='(-inf-55.55]' (169.0/56.0)
- (nu_idade = '(27.5-inf)') and (qp_i9_tempo_gasto_p_concluir_a_prova = .) => nt_fg='(-inf-55.55]' (613.0/253.0)
- (i17_tipo_ensino_medio = Todo em escola pública.) and (qp_i9_tempo_gasto_p_concluir_a_prova = Entre duas e três horas.) and (i4_escolaridade_pai = Ensino Fundamental: 1º ao 5º ano (1ª a 4ª série).) and (qp_i4_enunciados_questoes_parte_Formacao_Geral_claros_e_objetivos = Poucos.) => nt_fg='(-inf-55.55]' (186.0/66.0)
- (i17_tipo_ensino_medio = Todo em escola pública.) and (i38_planos_de_ensino_contribuiram_desenvolvimento_atividades_academicas_estudos = Concordo Totalmente) and (qp_i6_informacoes_e_instrucoes_p_resol_questoes_foram_suficientes_p_resolveras = Sim, somente em algumas.) => nt_fg='(-inf-55.55]' (744.0 / 340.0)
- (in_vespertino = 0) and (qp_i9_tempo_gasto_p_concluir_a_prova = Menos de uma hora.) => nt_fg='(-inf-55.55]' (257.0/50.0)
- (qp_i1_grau_dificuldade_prova_parte_Formacao_Geral = Difícil.) and (qp_i2_grau_dificuldade_prova_parte_Componente_Especifico = Médio.) and (qp_i8_considerando_questoes_objetivas_percebeu_que = Estudou alguns desses conteúdos, mas não os aprendeu.) => nt_fg='(-inf-55.55]' (138.0/47.0)
- (i73_instituicao_estagio_obrg = Escola pública.) and (qp_i8_considerando_questoes_objetivas_percebeu_que = .) => nt_fg='(-inf-55.55]' (222.0/83.0)