



## **ANÁLISE DOS EXAMES DO ENADE PARA OS CURSOS DE COMPUTAÇÃO UTILIZANDO O SOFTWARE R**

<sup>1</sup>Flávio Barreiro Landes

<sup>2</sup>Laci Mary Barbosa Manhães

### **Resumo**

A base de dados pública disponível sobre os exames que avaliam a formação dos alunos de graduação das instituições públicas e privadas brasileira constitui um imenso repositório de informações. Neste artigo, foram realizadas diversas análises utilizando o ambiente R a partir da seleção de todos os alunos dos cursos de graduação em computação que realizam os exames do Enade de 2005, 2008, 2011 e 2014. A facilidade de acesso ao R e o conjunto de recursos disponíveis no software permitiram a seleção, remoção de dados indesejados, análise estatística e visualização dos resultados na forma de gráficos. Particularmente, foram analisados o quantitativo de pessoas concluintes por sexo, idade e notas nos componentes gerais, específicos e nota bruta geral nas quatro realizações dos exames do Enade.

**Palavras-chave:** Enade, Computação, Estatística, Software R.

### **Abstract**

The public databases concentrate all available information about test performed by undergraduate students of public and private educational institutions in Brazil, those databases are large repository of information. In this article, several analyses were performed by using R statistical software applied to select all undergraduate students in computing science undergraduate degree courses, the computer science students were tested in the Enade of 2005, 2008, 2011 and 2014. R is available as free software and the set of extensive statistical facilities allowed the use in this work, selecting the data, unwanted data removal, statistical analysis and graphs visualization of the results. In particular, the analysis were about the quantitative of the individuals by gender, age and score in the common knowledge, specific computer knowledge and final score in the four Enade's test.

**Keywords:** Enade, Computer Science, Statistic, Software R.

### **Introdução**

O setor da Educação gera, coleta e armazena uma grande quantidade de dados. O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) é uma autarquia Federal vinculada ao Ministério da Educação (MEC). O Inep coleta e mantém

---

<sup>1</sup> Universidade Federal Fluminense (UFF), [landesflavio@gmail.com](mailto:landesflavio@gmail.com)

<sup>2</sup> Universidade Federal Fluminense (UFF), [mary\\_manhaes@id.uff.br](mailto:mary_manhaes@id.uff.br)

dados e informações oriundas de todas as instituições de ensino públicas e particulares, abrangendo desde a educação básica até o ensino superior (INEP, 2014).

O Exame Nacional de Desempenho de Estudantes (Enade) avalia o rendimento dos alunos concluintes dos cursos de graduação em relação aos conteúdos programáticos, habilidades e competências adquiridas em sua formação (ENADE, 2018). O exame é obrigatório e deve constar no histórico escolar de todo aluno concluinte do curso de graduação. Os dados provenientes dos exames do Enade estão disponíveis, mas ainda são poucas as instituições que analisam estes dados a fim de transformá-los em informações úteis. A análise dos dados gerados por cada exame Enade pode trazer benefícios para as instituições, direcionando o trabalho de professores e gestores acadêmicos, como: aprimorar o currículo do curso, reduzir os índices de evasão, identificar estudantes que não conseguem cumprir com os procedimentos relativos à integralização curricular dentro do tempo médio para conclusão do curso de graduação etc.

A Ciência de Dados (*Data Science*) vem congregar o aporte de diversas áreas de estudo (Matemática, Estatística, Aprendizado de Máquinas, Inteligência Artificial, Banco de Dados, Recuperação da Informação, Visualização de Informação entre outras) para explorar adequadamente os dados, visando obter informações úteis (MANHÃES, 2015).

Para realizar as diversas fases dos processos de análise de dados é necessário o uso de várias ferramentas. Entretanto, o R se tornou indispensável neste processo porque congrega diversos fatores, como a facilidade de aquisição do software, utilização de vários conjuntos de recursos para efetuar diversas análises e variadas formas de visualização dos resultados, entre outras facilidades.

Segundo Martins (2010), o software R se tornou conhecido em 1996 quando os seus criadores, Ross Ihaka e Robert Gentleman, da Universidade de Auckland, Nova Zelândia, distribuíram seu código fonte de forma aberta (do inglês, *open source*). O R se destaca entre os softwares estatísticos, sendo muito utilizado pelas universidades e tendo crescente aceitação nas empresas (MORAES, SALES e COSTA, 2011). Destaca-se que, recentemente, o R ganhou notoriedade e tem sido utilizado na mineração de dados, como pode ser visto em Silva (2011) e Gomes (2017).

Este trabalho está organizado da seguinte forma: A próxima seção apresenta o objetivo do trabalho. Na seção Material e Métodos descrevem-se os procedimentos utilizados para acessar as bases de dados do Inep, a utilização do R para seleção dos

dados, análises e geração dos gráficos. Na seção Resultados e Discussão discutem-se as análises e os gráficos. Por fim, apresentam-se as conclusões.

## **Objetivo**

O presente artigo tem como objetivo efetuar investigação e análises estatísticas, através do software R, sobre os dados de todos os alunos que concluíram o curso de computação nos respectivos anos de 2005, 2008, 2011 e 2014 e fizeram a prova do Enade. Especificamente, o software R foi utilizado para auxiliar no trabalho de seleção dos atributos, análise estatística e apresentação dos resultados em forma de gráficos dos dados dos exames do Enade. Ressalta-se que a base de dados pública do Enade está disponível no site do Inep (INEP, 2018).

## **Material e Método**

O Inep torna público as suas bases de dados na forma de arquivos, os microdados, cada arquivo corresponde a um ano específico do exame do Enade (INEP, 2018). Os microdados são o menor nível de desagregação de dados recolhidos por pesquisas, avaliações e exames realizados sobre a educação brasileira (INEP, 2018). Para identificar e entender o conteúdo é necessário utilizar o dicionário de variáveis que acompanha cada microdado. O microdado pode ser convertido em formato de tabela. As linhas contêm os dados de cada aluno que realizou a prova e as colunas formam o conjunto de dados coletados ou gerados referente a um exame do Enade, em um ano específico.

O conjunto de informação disponível em um microdado de um exame Enade é grande e permite realizar diversas análises. Em linhas gerais, as bases do Enade contêm vários atributos dos alunos, como: sexo, idade, notas, questões assinaladas, o curso de graduação ao qual ele pertence, período do curso (matutino, vespertino ou noturno), entre outras informações (INEP, 2018).

Neste artigo, selecionamos todos os alunos dos cursos de graduação em computação. Os microdados do Enade, atualmente disponíveis no site do Inep, referentes aos alunos concluintes dos cursos de computação correspondem aos anos de 2005, 2008, 2011 e 2014. Especificamente, foram utilizados os seguintes atributos: código de identificação do aluno na base, descrição do grupo do curso de graduação (coluna co\_grupo), idade dos alunos (coluna nu\_idade), sexo (coluna tp\_sexo), indicador de concluinte/ingressante

(coluna in\_grad ou tp\_inscricao), a nota bruta na formação geral (coluna nt\_fg), a nota bruta no componente específico (coluna nt\_ce) e a nota bruta, nota geral, no exame (coluna nt\_ger). A nota bruta geral do exame (nt\_ger) é calculada em função das notas: formação geral (nt\_fg) e a nota bruta no componente específico (nt\_ce).

Ressalta-se que os exames do Enade aplicados nos anos de 2005 e 2008 avaliaram alunos ingressantes e alunos concluintes no curso de graduação em computação. Em 2011 e 2014 o exame do Enade avaliou apenas os alunos concluintes. Neste trabalho, foram selecionados todos os alunos concluintes nos respectivos exames do Enade de 2005, 2008, 2011 e 2014.

O processo passou pelas seguintes fases: (i) obtenção da base pública (micrdados) do Enade dos anos de 2005, 2008, 2011 e 2014; (ii) seleção dos atributos (colunas); (iii) verificação dos dados, a fim de identificar dados nulos ou inválidos; (iv) remoção dos dados nulos ou considerados inválidos; (v) análise dos dados, incluindo as estatísticas; e (vi) geração dos gráficos. Neste trabalho, foram utilizados a linguagem de programação R e o Ambiente de Desenvolvimento Integrado (IDE) RStudio para realizar as fases (ii) a (vi) descritas acima.

Antes da análise estatística, foi realizada a fase de seleção cujo objetivo era filtrar todos os alunos matriculados nos cursos de computação presentes nos micrdados. O dicionário de variáveis, que acompanha cada micr dado, traz na descrição do atributo (coluna co\_grupo) o valor correspondente ao grupo do curso de graduação oferecido pelas Instituições de Ensino Superior (IES) públicas e privadas do Brasil. A descrição do valor do código utilizado para identificar o curso de graduação está no dicionário de variáveis dos respectivos exames do Enade. A seleção dos alunos dos cursos de computação foi com base nos seguintes códigos de identificação dos cursos de graduação em computação no Brasil: para os exames do Enade realizados nos anos de 2005 e 2008, os cursos de computação eram identificados apenas pelo código 40; para o Enade de 2011, os códigos foram mais específicos, como 4004 para a identificação do bacharelado em computação, 4005 para licenciatura em computação, 4006 para o bacharelado em sistemas de informação e 4007 para engenharia de computação; para o Enade de 2014, os códigos 4004 bacharelado em computação, 4005 para licenciatura em computação, 4006 para o bacharelado em sistemas de informação. Para cada ano do Enade foram utilizados comandos do R específicos, de acordo com os códigos de curso descritos acima. Em anexo, encontra-se o exemplo de comando do R utilizado para filtrar os alunos dos cursos de computação que realizaram o exame do Enade de 2011.

Historicamente, os primeiros exames do Enade foram boicotados por muitos graduandos, talvez isso justifique a existência de vários dados faltosos e nulos, principalmente nos exames de 2005 e 2008. Mais especificamente, notas gerais (coluna nt\_geral) sem valor (em branco) ou com valores menores do que 1 (um). Destaca-se que foram excluídos todos os dados (linhas) dos alunos que tinham essas notas gerais inválidas, ou seja, os dados desses alunos não foram usados para as análises feitas neste trabalho.

A Tabela 1 mostra os códigos das áreas de enquadramento dos cursos avaliados neste estudo, bem como a quantidade de alunos com notas válidas (nota  $\geq 1$ ) depois de ter sido feita a seleção citada e a exclusão dos dados inválidos. Ressalta-se que todas as análises feitas e apresentadas na seção de Resultados e Discussão são sobre todos os alunos dos cursos de graduação em computação. Consideramos que a Tabela 1 mostra o quantitativo de alunos com notas válidas para cada Enade, isto é, cada quantitativo representa o número total de alunos concluintes dos cursos de computação, considerando que o exame do Enade é uma prova obrigatória. Foram analisadas 4 (quatro) populações distintamente, como segue:

**Tabela 1** - Ano do exame, código do curso e quantidade de alunos que fizeram o exame.

Ano	Cursos de graduação em Computação	
	Código do curso	Número de Alunos
2005	40	13866
2008	40	16326
2011	4004, 4005, 4006 e 4007	22603
2014	4004, 4005 e 4006	23302

Fonte: Os autores, 2018

Os códigos do R utilizados para fazer a seleção dos dados, bem como para a remoção, estão na seção Anexo deste artigo.

## Resultados e Discussão

O conteúdo público disponível no Inep permite que diversos segmentos possam realizar várias análises, com o auxílio do ambiente R, para demonstrar resultados específicos e mais gerais. O ambiente R e o RStudio também se mostraram eficazes em outros trabalhos relacionados encontrados na literatura, como segue:

Figueiró e Mozzaquattro (2017) analisaram a base do Enade de 2014 em que foram feitas análises estatísticas do desempenho dos estudantes do curso de Ciência da

Computação do Estado do Rio Grande do Sul, bem como foram feitas comparações para agrupar as instituições públicas e privadas do mesmo estado de acordo com seus desempenhos.

Em Farias-Santos e Noro (2017), o trabalho teve o objetivo de comparar o desempenho dos estudantes no Enade de 2010 que participaram do programa PET-Saúde em 2009 e 2010 com os que não participaram.

Moimaz, Amaral e Garbin (2017) estudaram as bases dos exames do Enade dos anos de 2004, 2007, 2010 e 2013 e analisaram os componentes específicos dos cursos de Odontologia, de modo que, entre outras análises, foram abordados os conteúdos dos enunciados das questões da área da Saúde Coletiva onde foram utilizadas as técnicas de lexicográficas de classificação hierárquica descendente e nuvem de palavras.

No estudo de Crepalde e Silveira (2016), foram analisados os dados do Enade de 2014 para investigar o desempenho dos alunos de acordo com diferenças de sexo, cor e renda financeira para instituições publica e privadas, bem como de acordo com os cursos. Além disso, também foram investigados aspectos relacionados às influências das instituições sobre o desempenho dos alunos.

Neste artigo, três itens foram analisados: sexo, idade e notas relacionados aos alunos concluintes dos cursos de computação nos anos do exame Enade de 2005, 2008, 2011 e 2014. Os detalhes dos resultados obtidos e os gráficos serão mostrados a seguir.

A primeira análise destina-se a identificar o porcentual de pessoas dos sexos masculino e feminino que estão concluindo os cursos de graduação em computação e verificar se o número de mulheres está aumentando em uma área onde a predominância é masculina. O R foi utilizado para verificar o porcentual de pessoas do sexo masculino e do sexo feminino, conforme é mostrado na Tabela 2. Pelos resultados obtidos, pode-se considerar que o sexo masculino ainda é predominante entre os graduandos dos cursos de computação.

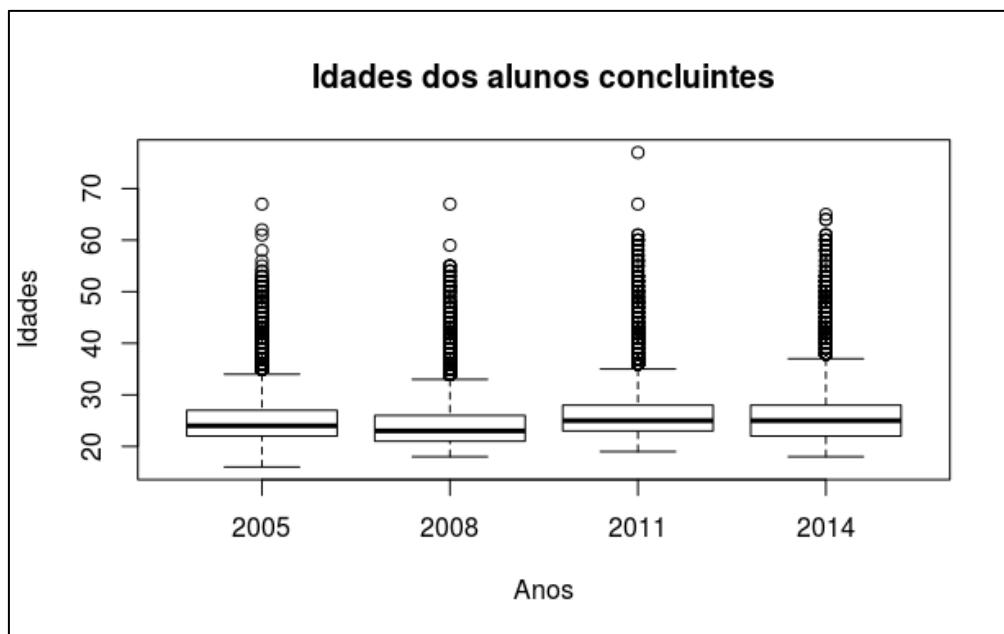
**Tabela 2** - Porcentual (%) de alunos do sexo masculino e do sexo feminino por ano do exame.

Sexo	Porcentual (%) de alunos do sexo masculino e do sexo feminino			
	2005	2008	2011	2014
Masculino	78,21	82,63	84,41	82,89
Feminino	21,79	17,37	15,59	17,11

Fonte: Os autores, 2018



A segunda análise destina-se a identificar a idade dos alunos, o objetivo é averiguar a tendência da faixa etária dos formandos nos cursos de computação. Para tanto, um gráfico do tipo *box plot* elaborado no R está representado pela Figura 2. Pelo gráfico é possível verificar que a média de idade dos alunos que estão se formando na computação é em torno dos 25 anos de idade. Além disso, é possível verificar que existe uma quantidade de alunos acima de 30 anos também concluindo os cursos de computação, estes são os *outliers*.

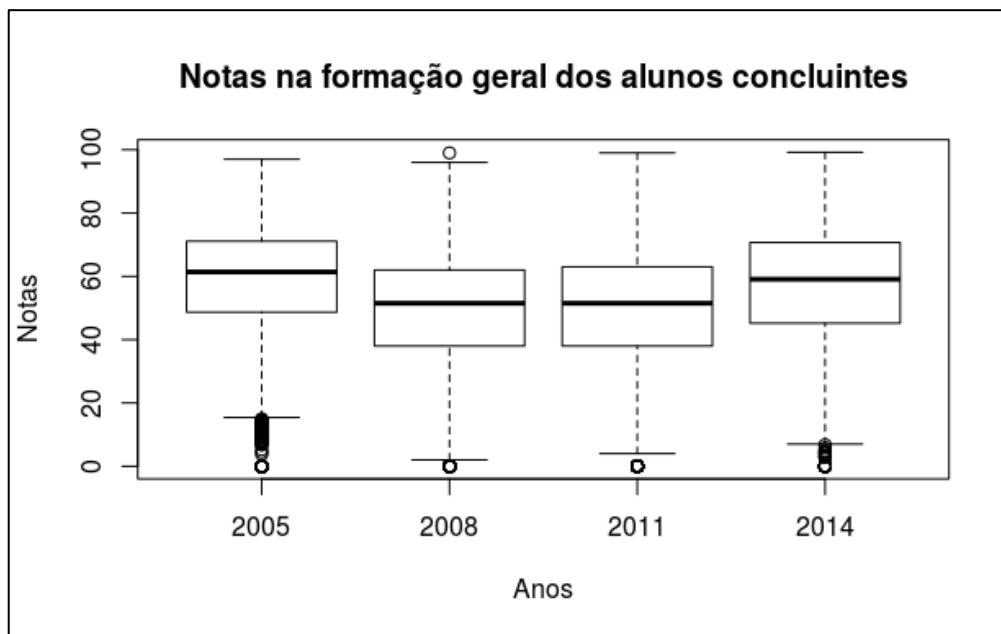


**Figura 1** – Gráfico das idades dos alunos concluintes nos cursos de computação.

Fonte: Os autores, 2018

A terceira análise destina-se a identificar as tendências das notas dos alunos dos cursos de computação, a fim de verificar, de maneira geral, o nível da formação oferecido pelos cursos de computação nas diversas IES.

A primeira nota investigada foi sobre a formação geral. As notas dos exames do Enade de 2005, 2008, 2011 e 2014 estão representadas pela Figura 3. Analisando o gráfico *box plot*, de maneira geral observa-se que as médias das notas ficaram abaixo ou próximas de 60. Além disso, também é notado que os *outliers* dos anos de 2005 e 2014, principalmente, demonstram que muitos alunos tiveram notas abaixo de 20 pontos. Todavia, em tais anos, é possível notar que grande parte dos alunos representados pelos intervalos interquartis nos *box plots* alcançou resultados melhores do que os demais anos observados.

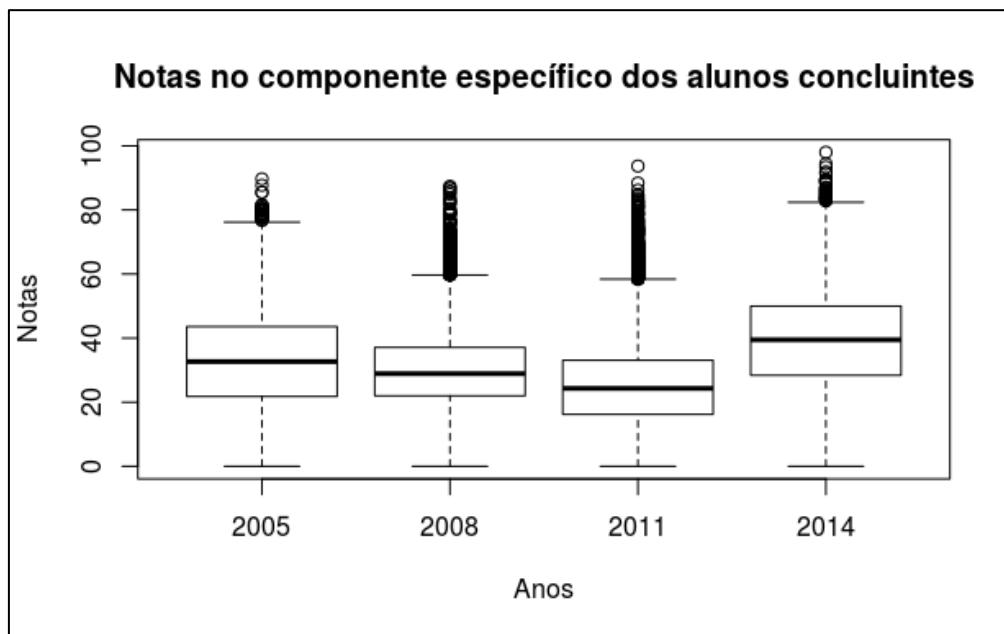


**Figura 2** - Notas na Formação Geral dos exames Enade.

Fonte: Os autores, 2018

A segunda nota investigada é o componente específico; as notas do componente específico representam os conhecimentos específicos da área de computação aprendidos nos cursos de computação.

A Figura 4 representa tais notas, bem como, explicitamente, demonstra que as notas estão baixas. As médias das notas estão menores ou em torno de 40. Também é possível verificar que alguns grupos de alunos se destacam por estarem bem acima da média, observando os *outliers*. Os intervalos interquartis que se destacam são os de 2005 e 2014, representando as maiores notas em comparação com os outros anos.



**Figura 3 - Notas no Componente Específico.**

Fonte: Os autores, 2018

Por fim, a terceira nota investigada é a nota geral obtida no exame do Enade para os alunos dos cursos de computação de todas IES para os exames de 2005, 2008, 2011 e 2014. Destaca-se que a nota bruta geral obtida no exame do Enade vai para o registro do aluno na IES.

Pelo gráfico, mostrado na Figura 5, é possível verificar que alguns alunos estão acima da média e outros alunos bem abaixo, considerando os *outliers*. Também nota-se que os intervalos interquartis com as maiores notas ocorreram nos anos de 2005 e 2014, sendo que neste último ocorreu o maior destaque.

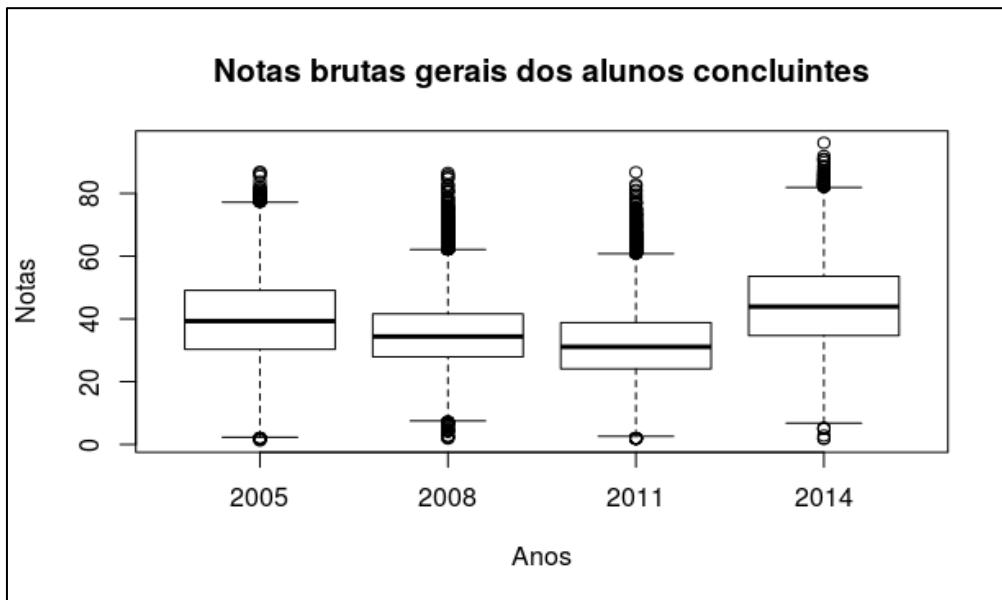


Figura 4 - Notas Brutas Gerais.

Fonte: Os autores, 2018

### Conclusão

O Exame Nacional de Desempenho de Estudantes (Enade) avalia o grau de desempenho dos alunos dos cursos de graduação tanto das instituições privadas quanto nas instituições públicas. Regulamente, o Inep disponibiliza as bases tornando-as dados públicos e passíveis de análise. Atualmente, estes dados constituem uma fonte livre para realizar diversas análises a fim de obter informações úteis para as IES. A comunidade acadêmica e demais interessados podem efetuar análises para tomar decisões, como: identificar o desempenho dos alunos concluintes a cada exame, identificar características específicas dos alunos, comparar o desempenho da IES em relação a outras IES, entre outras análises.

O R e o RStudio são recursos poderosos para realizar as análises estatísticas e a geração de gráficos para a investigação de dados relevantes nas bases de dados públicas. Particularmente, os exames do Enade disponíveis no site do Inep (INEP, 2018). Neste trabalho, investigamos os dados dos cursos de graduação em computação avaliados nos exames de 2005, 2008, 2011 e 2014. Mais especificamente, a investigação se deu sobre a distribuição por sexo dos alunos concluintes na computação, a idade dos concluintes e o desempenho das notas dos alunos concluintes com relação à formação geral, o componente específico e a nota bruta geral obtida pelos alunos no exame. Optamos por enfatizar as

análises gerais de todas as IES públicas e privadas. No entanto, como trabalho futuro pretendemos tratar os dados específicos de cada curso para cada IES.

As análises sobre a quantidade de alunos por sexo demonstraram que o sexo masculino ainda, majoritariamente, predomina nos cursos de computação. Em todos os anos analisados, o sexo masculino representa, aproximadamente, 80% dos alunos concluintes que fizeram a prova do Enade.

Sequencialmente, foram analisadas as idades; percebe-se que houve aumento gradativo das idades nos anos de 2011 e 2014 em comparação com os de 2008 e 2005, porém o aumento não foi de alta discrepância. No entanto, trabalhos futuros podem identificar e comparar as idades dos alunos ingressantes com a idade dos alunos concluintes. A fim de auxiliar os gestores das IES a identificar os motivos que levam a uma conclusão de curso de graduação tão tardia nos cursos de computação. Esta característica pode estar relacionada aos alunos que não conseguem a integralização curricular dentro do tempo médio previsto.

As notas dos alunos nos exames do Enade mostram a necessidade de um acompanhamento bem mais criterioso com relação à formação nos cursos de graduação em computação. As notas nas questões de formação geral demonstram que as médias estão abaixo ou em torno de 60, os melhores resultados foram nos anos de 2005 e 2014, e os piores resultados nos anos de 2008 e 2011, em torno da média 50. O componente específico, que representa o corpo do conhecimento na área de computação, obteve médias ainda menores; em 2014 observa-se a melhor média em torno de 40, seguido por 2005 com a segunda melhor média, sendo que os anos de 2008 e 2011 representaram a pior média nas notas do componente específico. As notas brutas gerais naturalmente também demonstram o mesmo padrão com as melhores médias em 2005 e 2014, bem como com as piores em 2008 e 2011.

Observa-se que a partir de 2005 as notas caíram, sendo que 2011 foi o pior resultado no exame do Enade para os alunos de computação. No entanto, 2014 mostrou um resultado melhor do que todos os exames anteriores. Outra constatação foi verificada comparando as notas da formação geral com as notas do componente específico, notoriamente as médias da formação geral são maiores do que as médias obtidas no componente específico. Esses fatos geram inquietações e implica na necessidade de futuras investigações. Não é possível a partir desta análise identificar os motivos destas discrepâncias nos resultados obtidos, mas é possível fazer algumas indagações como: de maneira geral, os cursos de computação melhoraram na formação de seus graduandos? Ou o exame do Enade sofreu

modificações na sua forma de avaliação? A resposta para tais interrogações remete para outras análises que podem ser feitas nas bases públicas disponíveis no Inep e, em particular, em dados de cada IES.

No mais, neste trabalho foi utilizado o software R com o objetivo de efetuar estatísticas sobre informações dos alunos dos cursos de computação que fizeram os exames do Enade aplicados nos anos de 2005, 2008, 2011 e 2014. Os recursos da linguagem permitiram efetuar todas as etapas necessárias para a geração das estatísticas, sendo tais etapas: seleção dos dados, remoção dos dados inválidos, cálculos estatísticos e geração de gráficos.

Por fim, a linguagem R é um ambiente de fácil utilização e com recursos promissores, permitindo a seleção de dados específicos disponíveis em bases de dados públicas, a facilidade de manipular volumes médios de dados, a obtenção da análise estatística dos dados e a representação dos resultados em várias formas de gráficos. Particularmente, o ambiente R ofereceu recursos gráficos para a análise de vários anos de exame do Enade. A manipulação dos dados, análise e visualização dos resultados utilizando o ambiente R permite que diversos profissionais da comunidade acadêmica possam conhecer e tomar decisões sobre a formação profissional nas IES, beneficiando a sociedade como um todo.

## Referências

CREPALDE, N. J. B. F.; SILVEIRA, L. S. Desempenho Universitário No Brasil: Estudo Sobre desigualdade educacional com dados do Enade 2014. **Revista Brasileira de Sociologia - Rbs**, [s.l.], v. 4, n. 7, p.211-238, jul. 2016.

ENADE 2018 - Exame Nacional de Desempenho de Estudantes. **INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Ministério da Educação.** Disponível em: <<http://portal.inep.gov.br/enade>>. Acessado em março de 2018.

FARIAS-SANTOS, B. C. S.; NORO, L. R. A. PET-Saúde como indutor da formação profissional para o Sistema Único de Saúde. **Ciência & Saúde Coletiva**, [s.l.] v. 22, n. 3, p. 997–1004, mar. 2017.

GOMES, G. L.; SEGUNDO, W. L. R. C. Estudo para integração entre a Plataforma Lattes a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) e o Banco de Teses e Dissertações da Capes. **Reciis – Rev Eletron Comun Inf Inov Saúde**, Rio de Janeiro, v. 11, p.1-5, nov. 2017.

INEP 2018 - Microdados. **INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Ministério da Educação.** Disponível em: <<http://inep.gov.br/web/guest/microdados>>. Acessado em março de 2018.

INEP. **Resumo Técnico do Censo da Educação Superior 2014.** Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2014.

MANHÃES, L. M. B., CRUZ, S.M.S., ZIMBRÃO, G. Predição do Desempenho Acadêmico de Graduandos Utilizando Mineração de Dados Educacionais. Tese de doutorado, 2015.

MARTINS, M. C. **Análise Conjunta de Regressões com desenvolvimentos computacionais em Linguagem R.** 2010. 84 f. Dissertação (Mestrado) - Curso de Estatística, Matemática e Computação, Especialidade Estatística Computacional, Ciências e Tecnologia, Universidade Aberta, Lisboa, 2010.



MOIMAZ, S. A. S.; AMARAL, M. A.; GARBIN, C. A. S. Enade: uma análise quanti-qualitativa dos exames nacionais de Odontologia. **ABENO**, [s.l.], v. 17, p. 97-108, 2017.

MORAES, R. F.; SALES, P. R. H.; COSTA, J. F. S. Utilização do software estatístico R na previsão de séries temporais: série prêmio de resseguro. **Cadernos do Ime – Série Estatística**, Rio de Janeiro, v. 30, p. 01-17, 2011.

SILVA, M. L. F. **Mineração de dados em redes sociais: um estudo de caso na ferramenta twitter sobre a organização little, brown book company**. 2011. 50 f. TCC (Graduação) - Curso de Gestão da Informação, Ciência e Gestão da Informação, Universidade Federal do Paraná, Curitiba, 2011.

VISTA, N. P. B.; FIGUEIRÓ, M. F.; MOZZAQUATRO, P. M. 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. **I Seminário de Pesquisa Científica e Tecnológica**, s. l., v. 1, p. 1-11, 2017.

## Anexo

```
# Códigos utilizados para efetuar a seleção dos dados. Como exemplo, os comandos utilizados para os microdados do exame Enade de 2011.
```

```
*****
```

```
# Leitura da tabela.
```

```
enade2011 = read.csv("microdados_enade_2011.csv", header = T, sep = ",");
```

```
# Filtra somente as linhas (alunos) referentes ao curso de Computação.
```

```
enade2011 = subset(enade2011, enade2011$co_grupo == 4004 | enade2011$co_grupo == 4005 | enade2011$co_grupo == 4006 | enade2011$co_grupo == 4007);
```

```
# Filtra somente as linhas que têm valores válidos na coluna nt_ger (nota geral).
```

```
enade2011 = subset(enade2011, enade2011$nt_ger != "");
```

```
# Filtra somente as linhas (alunos) que estão concluindo o curso.
```

```
enade2011 = subset(enade2011, enade2011$in_grad == 0);
```

```
# Filtra somente as linhas (alunos) que têm notas gerais maiores ou iguais a 1.
```

```
enade2011 = subset(enade2011, enade2011$nt_ger >= 1);
```

```
# Filtra somente as colunas desejadas.
```

```
enade2011 = subset(enade2011, select = c("co_grupo", "nu_idade", "tp_sexo", "nt_fg", "nt_ce", "nt_ger"));
```

```
# Grava a nova tabela limpa.
```

```
write.csv(enade2011, file = "microdados_enade_2011LIMPOS.csv");
```



```
#*****
# Comandos utilizados para efetuar a criação dos boxplots. Abaixo está um exemplo para
# a criação de um boxplot das Notas Brutas Gerais.
#*****

boxplot(enade2005$nt_ger, enade2008$nt_ger, enade2011$nt_ger, enade2014$nt_ger,
names = c("2005", "2008", "2011", "2014"), main = "Notas brutas gerais dos alunos
concluintes", xlab = "Anos", ylab = "Notas");
```

```
#*****
# Comandos para calcular a quantidade (%) de cada sexo (Masculino e Feminino) por
# ano. Abaixo se encontra o algoritmo usado no Enade de 2014, bem como no de 2011
# (onde foi trocado o ano). Em 2005 e 2008 o sexo masculino é representando por 1 e o
# feminino por 2, então esses valores foram substituídos no lugar do "M" e "F" no algoritmo.
#*****
```

```
m = 0;
f = 0;
for (sexo in enade2014$tp_sexo) {
  if (toupper(sexo) == "M") {
    m = m + 1;
  } else {
    if (toupper(sexo) == "F") {
      f = f + 1;
    }
  }
}
print(m / (m + f) * 100);
print(f / (m + f) * 100);
```