



Instituto de  
MATEMÁTICA  
E ESTATÍSTICA

UFRGS



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DEPARTAMENTO DE ESTATÍSTICA

**Mineração de Dados Educacionais:  
Análise de uma disciplina de Estatística em EaD**

**ALINE CAFRUNI GULARTE**

Porto Alegre  
2018

**ALINE CAFRUNI GULARTE**

**Mineração de Dados Educacionais:  
Análise de uma disciplina de Estatística em EaD**

Trabalho de Conclusão de Curso submetido  
como requisito parcial para a obtenção do  
grau de Bacharel em Estatística

Orientador  
Prof.<sup>a</sup> Dr.<sup>a</sup> Lisiane Priscila Roldão Selau

Coorientador  
Prof.<sup>a</sup> Dr.<sup>a</sup> Márcia Helena Barbian

Porto Alegre  
2018

Instituto de Matemática e Estatística  
Departamento

**Mineração de Dados Educacionais:  
Análise de uma disciplina de Estatística em EaD**  
Aline Cafruni Gularte

Banca examinadora:

Prof.<sup>a</sup> Dr.<sup>a</sup> Lisiane Priscila Roldão Selau  
Universidade Federal do Rio Grande do Sul

Prof.<sup>a</sup> Dr.<sup>a</sup> Márcia Helena Barbian  
Universidade Federal do Rio Grande do Sul

Prof.<sup>a</sup> Dr.<sup>a</sup> Luciana Nunes Neves  
Universidade Federal do Rio Grande do Sul

## AGRADECIMENTOS

Agradeço primeiramente a Deus por ter me dado saúde e força para superar todas as dificuldades.

A esta universidade que me proporcionou um ensino de qualidade com o excelente corpo docente que compõem o curso de Estatística.

Aos funcionários, em especial a Giovana que me atendia sempre com muito carinho e amor na COMGRAD e aos porteiros do Prédio A que sempre me cumprimentavam com muita alegria.

Um agradecimento em especial aos professores Hudson Torrent, Vanessa Leotti e Jandyra Fachel que marcaram a minha trajetória por serem tão dedicados e excelentes professores.

À professora Liane Werner que me convidou a ministrar aulas de Excel e que sempre me tratou com muito carinho.

Aos amigos/colegas que conquistei durante toda a minha trajetória, em especial aos amigos Lincon, Ângela, Rosana e a minha quase irmã Mariana Nolde que esteve em muitos momentos ao meu lado.

À professora Débora Feijó que me acolheu com todo o carinho e me ensinou tanta coisa durante o meu estágio no HCPA. Também agradeço a ela por ter me apresentado à Célia uma maravilhosa pessoa que se tornou uma grande amiga.

Agradeço também a professora Márcia Echeveste que me ensinou tantas coisas enquanto realizei minha bolsa de I.C. na Engenharia de Produção. Ela sempre acreditou no meu potencial e sempre me incentivou a não desistir dos meus sonhos de seguir a carreira acadêmica. Também devido a ela foi que conheci a Érika, Lindomar, Fernando, a amada professora Istefani, a querida Elaine e o Bernardo que se disponibilizaram tantas vezes em me ajudar.

Agradeço muito à professora Márcia Barbian por ter aceitado ser a minha coorientadora. Durante essa trajetória ela me ajudou muito com toda a sua dedicação, carinho, paciência e correções críticas no trabalho que foram essenciais para meu crescimento como profissional.

À professora Luciana Nunes que, com todo o seu carisma e alegria, aceitou participar da minha banca de TCC.

E agradeço a minha amada, querida, amiga, orientadora e eterna professora Lisiane Selau. Uma professora extremamente dedicada, que ama tanto a sua profissão e que se preocupa com o ensino de cada um dos seus alunos. É a professora mais inspiradora que eu já conheci. É nela que sempre me inspirei diariamente para atingir meus objetivos profissionais. Sempre me ensinou a nunca desistir do meu sonho, mesmo quando parecia ser impossível. Sem ela certamente minha trajetória até aqui seria muito mais difícil e triste. Muito obrigada professora pelas inúmeras correções no meu TCC e pela paciência de sempre.

Agradeço também aos meus amigos Isis, Cynara, Déborah e Amanda, que sempre compreenderam a minha ausência nesses últimos anos e que sempre torceram por mim.

A minha amiga/irmã/colega Lígia Latorre, que desde o momento em que nos conhecemos não deixamos mais de nos ajudar e apoiar. Foram incansáveis noites mal dormidas fazendo trabalhos e estudando para as provas. Momentos em que conversávamos e muitas vezes chorávamos sobre as dificuldades que enfrentávamos durante o curso devido à falta de coleguismo e cooperação dos colegas. Criamos um vínculo tão grande de amizade que será levado para a vida toda. Muito obrigada minha amiga por fazer parte da minha vida.

Agradeço a minha quase irmã Clarissa que me apoiou tantas vezes e que sempre esteve ao meu lado nessa minha trajetória.

Também não posso deixar de agradecer ao meu gatinho e companheiro Luan que ficava sempre ao meu lado e me dava muito carinho naquelas noites inacabáveis de estudo.

Aos meus familiares, em especial a minha vizinha Iria que foi essencial no meu crescimento devido a todo o seu amor que sempre teve por mim.

Aos meus pais, Luiz e Georgina, que me deram um ensino de qualidade. Que sempre estiveram ao meu lado. Principalmente à minha mãe, minha melhor amiga, que não tenho nem palavras para expressar o quão importante ela sempre foi pra mim. E ao meu pai que não dormia enquanto eu ficava de madrugada estudando, que acordava cedo para me acordar, pois eu não ouvia o despertador tocar e mesmo com todas as suas limitações sempre se preocupou comigo.

Ao meu irmão Álan e cunhada Caroline que sempre se preocuparam com os meus estudos, são a minha fonte de inspiração. Eles me ensinaram a importância de sempre desejar a melhorar profissionalmente. Obrigada por me amarem muito e sempre se preocuparem comigo.

Agradeço principalmente ao meu namorado Gardel Flores, não tenho nem palavras para escrever o quanto ele é importante para mim. Todos esses anos ele foi extremamente atencioso, dedicado e amoroso. Sempre me ajudou a superar todos os obstáculos desse percurso. Esteve ao meu lado nos meus momentos mais difíceis e felizes. O seu amor e carinho foram essenciais para o meu crescimento. Muito obrigada meu amor.

E por fim, agradeço a todos que direta ou indiretamente fizeram parte da minha formação. O meu muito obrigada!

“Fallo con orgoglio, E dove stai andando arriverai”  
Laura Pausini; Fai quello che sei, 2018

## RESUMO

O Ensino a Distância (EaD) tem assumido um papel relevante no processo de ensino-aprendizagem devido aos avanços das tecnologias digitais, dos sistemas de informação e comunicação. Um curso na modalidade a distância para ser bem desenvolvido, além de ter um planejamento didático cuidadoso e bons recursos tecnológicos, necessita do auxílio de um Ambiente Virtual de Aprendizagem (AVA), que por definição é um sistema elaborado para auxiliar no desenvolvimento de cursos EaD. Na UFRGS o Moodle é um AVA comumente utilizado por vários professores que trabalham com ensino a distância. A plataforma possibilita que o professor/tutor do curso acesse as informações dos alunos que são geradas toda vez que o estudante clica em um material (slides, vídeos, chats, fóruns, etc.). Essas informações, ao serem analisadas, podem ser transformadas em resultados que identifiquem as ações dos alunos dentro da plataforma Moodle e que tenham relação com o seu desempenho final na disciplina. Dentro deste contexto, neste trabalho, foram utilizadas técnicas de Minerações de Dados Educacionais para identificar as ações dos alunos, da disciplina de Probabilidade e Estatística, na plataforma Moodle que estejam associadas com seu desempenho final na disciplina. Foram analisados dados referentes as *logs* de interação de 452 alunos de oito turmas da disciplina de Probabilidade e Estatística da UFRGS na modalidade a distância nos períodos de 2016/01 a 2018/01. Com os resultados obtidos neste estudo é esperado que os professores/tutores possam recomendar aos futuros alunos quais as possíveis ações que eles devem executar na plataforma Moodle que poderão ajudá-los a melhorar seu desempenho final na disciplina.

**Palavras-chave:** Estatística. Ensino a distância. Ambiente Virtual de Aprendizagem. Moodle. Mineração de Dados Educacionais.

## ABSTRACT

Distance Learning (DL) has played a key role in the teaching-learning process due to advances in digital technologies and information and communication systems. For a remote mode course to be well developed, besides having a careful didactic planning and good technological resources, requires the assistance of a Virtual Learning Environments (VLE), which by definition is a system designed to assist in the development of DL courses. At UFRGS, Moodle is an VLE commonly used by several teachers who work with distance education, this platform allows the teacher/tutor of the course to access the students' information that are generated every time the student clicks on a material (slides, videos, chats, forums, etc.). This information can be analyzed and can be transformed into results that identify the actions of students within the platform. In this context, in this work, Educational Data Mining (EDM) techniques were used to identify the actions of students, of the probability and statistical course, in the Moodle platform that are associated with their final performance in the course. Data regarding the logs of interaction of 452 students from eight remote mode Probability and Statistics classes of UFRGS, in the period from 2016/01 to 2018/01. With the results obtained in this study it is expected that teachers/tutors can recommend to future students what actions they should take on the Moodle platform that can help them to improve their final performance in the course.

**Keywords:** Statistic. Teaching. Virtual Learning Environments. Moodle. Educational Data Mining.

## LISTA DE FIGURAS

Figura 1 - Mineração de Dados em Sistemas Educacionais. ....	18
Figura 2 - Síntese da Descoberta de Conhecimento de Dados. ....	19
Figura 3 - Informações das variáveis do banco de dados .....	25
Figura 4 - Descrição das variáveis de cada subárea.....	26
Figura 5 - Boxplot das notas das provas de cada área e nota final.....	28
Figura 6 - Percentual de alunos que acessaram aos slides e vídeos.....	29
Figura 7 - Percentual de alunos que acessaram aos arquivos do Moodle .....	30
Figura 8 - Comparação entre alunos que realizaram e não realizaram os simulados em relação à nota da prova separados pelo resultado de aprovação da área 1 .....	31
Figura 9 - Comparação entre alunos que realizaram e não realizaram os simulados em relação à nota da prova separados pelo resultado de aprovação da área 2 .....	32
Figura 10 - Comparação entre alunos que realizaram e não realizaram os simulados em relação à nota da prova separados pelo resultado de aprovação da área 3 .....	32
Figura 11 - Gráfico de dispersão: comparação entre as notas das provas e dos simulados separados por aprovação.....	33
Figura 12 - Gráficos de dispersão, densidade, Boxplots e histograma: comparação entre as notas das provas e notas dos simulados separados por aprovação .....	35
Figura 13 - Comparação entre slides e vídeos acessados durante o curso e o resultado de aprovação dos alunos da área 1 .....	36
Figura 14 - Comparação entre slides e vídeos acessados durante o curso e o resultado de aprovação dos alunos da área 2 .....	36
Figura 15 - Comparação entre slides e vídeos acessados durante o curso e o resultado de aprovação dos alunos da área 3 .....	37
Figura 16 - Fluxo do acesso aos materiais da área 1 para compreender o resultado médio das provas .....	38
Figura 17 - Fluxo do acesso aos materiais da área 2 para compreender o resultado médio das provas .....	39

Figura 18 - Fluxo do acesso aos materiais da área 3 para compreender o resultado médio das provas .....40

Figura 19 - Fluxo inverso: análise dos acessos aos materiais partindo da aprovação do aluno .....41

---

## SUMÁRIO

1.	INTRODUÇÃO .....	13
2.	METODOLOGIA .....	18
2.1	Disciplina MAT02219 - Probabilidade e Estatística em EaD da UFRGS.....	20
2.2	Coleta de dados .....	21
2.3	Pré-processamento dos dados .....	22
2.3.1	Extração, manipulação e agrupamento dos bancos de dados .....	23
2.3.2	Variáveis .....	24
2.4	Mineração de Dados Educacionais .....	26
3.	RESULTADOS.....	28
3.1	Análise dos Simulados .....	30
3.2	Análise dos Slides e Vídeos.....	36
3.3	Análise do Desempenho do Aluno .....	37
4.	CONCLUSÕES .....	42
5.	REFERÊNCIAS .....	44
	APÊNDICE A .....	46

## 1. INTRODUÇÃO

O Ensino a Distância (EaD) tem assumido, ao longo dos anos, um papel relevante, destacando o sentido de liberdade e autonomia, provocando uma nova visão para a educação. Além disso, tem-se a urgência de serem atribuídos novos valores, critérios e conceitos com o auxílio das tecnologias digitais, dos sistemas de informação e comunicação, possibilitando novas formas de ensino-aprendizagem (PETERS, 2006).

No contexto brasileiro, os cursos ofertados na modalidade a distância tiveram crescimento considerável nos últimos anos. De acordo com o Censo da Educação Superior de 2017 do Inep (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira), o Ensino Superior tem cerca de 8,3 milhões de estudantes em cursos de graduação, desses, 21,2% são do EaD. Em um comparativo do ano de 2016 com o ano de 2017, o número de matrículas na modalidade a distância cresceu 17,6%. Em contrapartida, o número de estudantes na modalidade presencial caiu 0,4% no mesmo período (INEP, 2018).

Mattar (2013), professor especialista em EaD, explica que o crescimento desta modalidade de ensino deve-se a rupturas de barreiras que impossibilitavam o aluno de estudar. Rodrigues (2010) acrescenta que este crescimento é devido, não somente à flexibilidade de horários, que possibilita uma maior autonomia para estudar, mas também pela economia, desde o valor da mensalidade até o gasto com o transporte, e principalmente devido ao crescente desenvolvimento dos mais diversos recursos tecnológicos. Entre os principais recursos estão: os computadores e o acesso fácil à internet que estão cada vez mais rápidos.

Sendo assim, para desenvolver um bom curso a distância, além de ter um planejamento didático cuidadoso e bons recursos tecnológicos, como videoaulas e o estímulo à interação professor-aluno, é indispensável o auxílio dos Ambientes Virtuais de Aprendizagem (AVA). Por definição, AVA é um sistema elaborado para auxiliar no desenvolvimento de cursos EaD. Este sistema é constituído por uma opção de mídia que possui chats, listas de discussões, fóruns, entre outras ferramentas que são utilizadas como

forma de facilitar o ensino em cursos presenciais e, mais intensamente, em cursos ofertados na modalidade a distância (NORONHA, 2005).

Com o propósito de criar uma plataforma de EaD que possibilitasse uma grande interação e englobasse um *layout* amigável, de fácil manuseio e acesso, Martin Dougiamas (1999) desenvolveu a plataforma Moodle. Segundo Senechal (2013), o Moodle segue os princípios do construcionismo social, ou seja, baseia-se na interação entre o ambiente (plataforma) e os envolvidos (alunos/professores), construindo um conhecimento a partir do aprendizado colaborativo.

Em outras palavras, o Moodle é um *software* livre de apoio à aprendizagem, executado em um AVA. Nele é possível que o professor elabore e disponibilize materiais didáticos de forma organizada, podendo dividi-los em tópicos ou seguir uma estrutura semanal. O aluno tem possibilidade de acessar materiais didáticos, postar atividades, realizar exercícios, debater sobre temas em fóruns, tirar dúvidas via mensagens e chats, e usufruir de outras ferramentas que o professor opte em usar (FILHO, 2005).

No mesmo instante em que o aluno interage com o Moodle, ou seja, quando exerce qualquer tipo de ação (clique) dentro do ambiente, ele acaba produzindo rastros. Machado (2018) conceitua estes rastros de cliques, ou *ClickStream*, como um registro da trajetória que um usuário de computador percorre ao clicar em página *web* ou aplicativo, armazenando informações, como por exemplo, os sites que visitou, quanto tempo gastou visualizando a página e quantas vezes acessou a página *web* e seus conteúdos.

Neste contexto, o *ClickStream* do aluno na Plataforma Moodle gera uma grande fonte de dados que, com o auxílio de análises estatísticas, podem detectar padrões de comportamento, os quais possibilitam criar mecanismos para aumentar a motivação do aluno pela disciplina ou indiquem discentes com maior risco de evasão. Quando o aluno acessa o sistema, clica no material didático, acessa fóruns, chats, ou outros recursos, ele gera fontes de informação que são armazenadas em *logs*. Esses dados podem ser acessados pelos professores e tutores de um curso. Tais comportamentos de acesso podem ser identificados no decorrer do curso para evitar que o não acesso aos materiais levem o aluno a evadir ou a ter um mau desempenho na disciplina.

Diante de tal cenário, Jozemberg (2016) explica que para facilitar o estudo e análises das fontes de dados extraídas nos *logs* do Moodle é fundamental utilizar técnicas e ferramentas da Mineração de Dados (MD), pois é através desta mineração que será possível coletar, organizar e analisar as informações de grandes bancos de dados, para que os dados gerados auxiliem na escolha de técnicas pedagógicas adequadas de forma ágil e confiável.

Além de tudo, a MD é uma área de pesquisa multidisciplinar, envolvendo basicamente Banco de Dados, Estatística e Aprendizado de Máquina. Fayyad *et al.* (1996) conceituam a MD como sendo a parte principal de um processo que tem como entrada uma Base de Dados e como saída um Conhecimento.

Conforme Anjewierden, Kollöffel e Hulshof (2007), a MD dá origem ao campo de pesquisa em Mineração de Dados Educacionais (MDE). Segundo Yadav e Pal (2012), a MDE auxilia no desenvolvimento de métodos que exploram os conjuntos de dados coletados pelos rastros de cliques em ambientes educacionais. Assim, pode-se utilizá-lo na busca de informações importantes a respeito de padrões de comportamento do aluno.

Romero e Ventura (2007) explicitam as particularidades da aplicação da mineração de dados. Segundo os autores, deve-se ter os dados disponíveis, os objetivos da mineração e compreender as técnicas a serem aplicadas, não apenas transformando os dados em conhecimento, mas também filtrando o conhecimento extraído para a tomada de decisões. Baker *et al.* (2012) acrescentam que, como em toda Mineração de Dados, há problemas com a extração das informações. Em MDE se faz necessária uma etapa dedicada à preparação de dados, de forma a adequá-los à análise que se pretende efetuar.

Em meados da década de 80, Mendonza e Swift (1981) abordaram sobre a importância do ensino da Estatística para todos os indivíduos. Segundo os autores, ter o conhecimento básico de Estatística permite que os indivíduos atuem na sociedade com mais confiança. Na percepção de Lopes (2008), uma vez que o estudante aprende a coletar, organizar, interpretar e comparar dados estatísticos para fundamentar conclusões, ele já está apto para realizar análises críticas sob diferentes aspectos científicos, tecnológicos e/ou sociais. De acordo com Pan (2003), devido a conscientização da im-

portância do ensino da estatística, as disciplinas de Probabilidade e Estatística já integram grande parte dos cursos de graduação das universidades e têm o objetivo de apresentar aos alunos as diferentes ferramentas utilizadas em pesquisas de todas as áreas de conhecimento.

Neste cenário, mais especificamente, a disciplina de Probabilidade e Estatística da UFRGS (Universidade Federal do Rio Grande do Sul), oferecida a 34 cursos na área de ciências exatas, no formato presencial e EaD, tem como objetivo ensinar conceitos e métodos estatísticos que possa ampliar a compreensão em relação aos fenômenos que sustentam as diferentes áreas de conhecimento. Esta disciplina possui os mesmos conteúdos tanto na modalidade presencial quanto a distância, sendo dividida em 3 áreas de Probabilidade e Estatística. A primeira área abrange os conteúdos relacionados à estatística descritiva, representações gráficas e probabilidade. A segunda aborda sobre variáveis aleatórias, modelos probabilísticos e estimação. A terceira e última área explana sobre testes de hipóteses, correlação e regressão linear simples.

Gularte e Selau (2017) realizaram um estudo comparativo das modalidades presencial e a distância da referida disciplina de Probabilidade e Estatística da UFRGS, objeto de estudo deste trabalho. A pesquisa apontou que não houve diferença significativa entre as médias finais dos alunos ao comparar as modalidades presencial e a distância ( $p > 0,05$ ). No final do curso foi enviada uma pesquisa de satisfação com o objetivo de avaliar a percepção dos alunos sobre a disciplina EaD. Os alunos responderam que ficaram bem satisfeitos com o formato da disciplina, fizeram muitas avaliações positivas quanto aos recursos utilizados e foram unânimes em dizer que recomendariam a disciplina na modalidade EaD para seus colegas.

Outro estudo sobre uma disciplina de Estatística Básica em EaD oferecida aos cursos de Ciências Sociais, Geografia e Biblioteconomia da UFRGS, realizado pela professora Luciana Neves Nunes (2010), apontou que oferecer estatística a distância para disciplinas de serviço é atraente para os alunos. Os alunos alegaram que o fato de ficarem dispensados de estarem sistematicamente presentes em um determinado local e horário, como acontece nas aulas presenciais, faz com que seu interesse pelas aulas aumente. Ou seja, os alunos ressaltaram a importância dessa inovação de ensino aten-

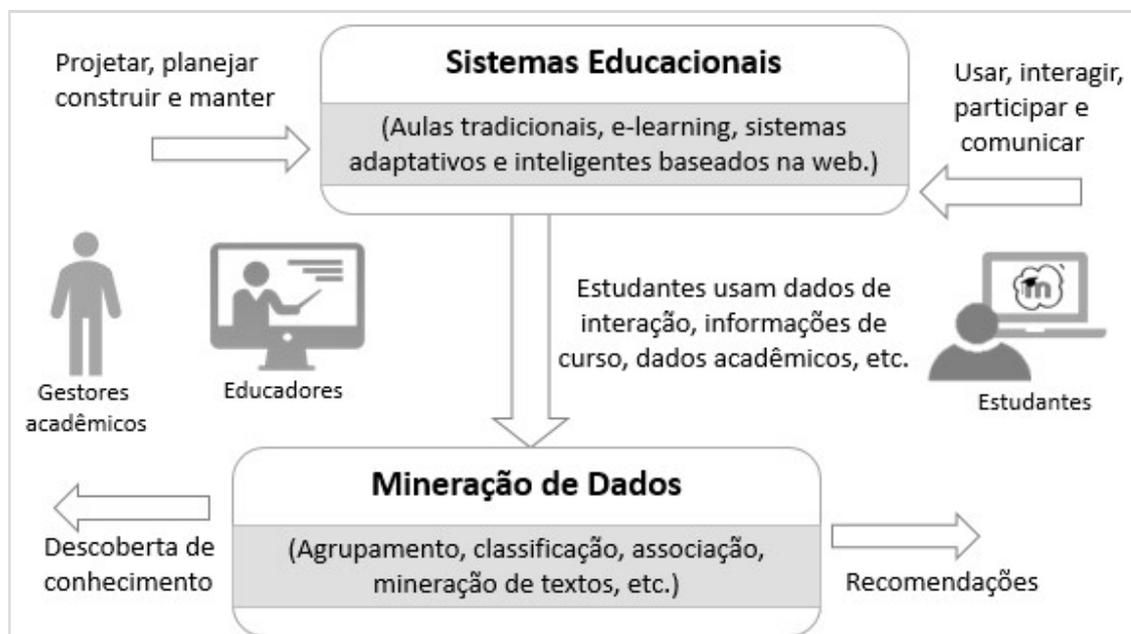
der a demanda do estilo de vida atual das pessoas e a importância de saber administrar o curto tempo disponível.

Dentro deste contexto, este trabalho tem como objetivo: identificar ações do aluno no uso da plataforma Moodle que tenham relação com o desempenho final na disciplina de Probabilidade e Estatística da UFRGS. Para isso os dados gerados pelo *ClickStream* do aluno em cada material ou ferramenta da plataforma serão analisados com base no processo de Mineração de Dados Educacionais, que constitui as seguintes etapas: coleta de dados; pré-processamento dos dados; mineração de dados; e por último, interpretar e avaliar os resultados. Na mineração de dados serão utilizadas as técnicas da Análise Exploratória de Dados (AED) que consistem em identificar quais são as ações do aluno no uso da plataforma Moodle que têm relação com o seu desempenho final na disciplina.

Este trabalho está estruturado em quatro seções. Após a introdução apresentada nesta seção inicial, a segunda seção traz os *softwares* e os métodos utilizados para a realização da Mineração de Dados Educacionais. Na terceira seção são apresentados os principais resultados obtidos na MDE. A última seção traz as considerações finais do estudo e as principais conclusões obtidas.

## 2. METODOLOGIA

As técnicas de Mineração de Dados Educacionais seguem um ciclo de formação de hipóteses, testes e refinamentos sobre as informações observadas durante o decorrer de um curso com apoio de um AVA. O ciclo de atividades envolvidas na aplicação de mineração de dados em sistemas educacionais, conforme representado na Figura 1 envolve os interesses de cada grupo de indivíduos (aluno, professores/tutores) com relação aos dados gerados pelo AVA, ou seja, os cliques que cada aluno gera ao acessar um material disponibilizado no Moodle. Também, gestores e educadores, cada qual com seu objetivo específico, possuem interesse nas descobertas de conhecimentos, pois acabam sendo favorecidos ao obterem um melhor planejamento e manutenção do curso em EaD. Além disso, professores/tutores podem fazer recomendações e orientações aos seus alunos que utilizam os AVA's para melhora do processo de aprendizagem (ROMERO; VENTURA, 2007).



**Figura 1** - Mineração de Dados em Sistemas Educacionais  
Fonte: Adaptado de Romero e Ventura, 2007

Em resumo, a aplicação da mineração de dados em sistemas educacionais pode ajudar a identificar quais são as necessidades específicas de cada um dos participantes do processo educacional. Com essa identificação é possível recomendar aos alunos quais as ações que eles devem tomar no AVA que favoreçam e melhorem sua aprendizagem. Os professores/tutores podem acessar, no decorrer do curso, todo o histórico do aluno, possibilitando classificar os alunos em grupos, com base em sua necessidade de orientação e monitoramento, descobrir os erros mais cometidos, encontrar as ações efetivas, etc. (ROMERO; VENTURA, 2007).

Desta forma, o presente estudo foi estruturado com base no processo de Mineração de Dados Educacionais, que será essencial na identificação das principais ações dos alunos no uso da plataforma Moodle que refletem em seu desempenho final. Segundo Romero, Ventura e Garcia (2008), o processo MDE em *e-learning* é constituído de quatro etapas que são representadas na Figura 2.



**Figura 2** - Síntese da Descoberta de Conhecimento de Dados  
Fonte: Adaptado de Romero e Ventura, 2007

**1º - Coleta de dados:** armazenar as informações das interações dos estudantes no ambiente em um banco de dados.

**2º - Pré-processamento dos dados:** é feita uma limpeza e transformação dos dados em formatos apropriados para a utilização da mineração de dados no próximo passo.

**3º - Aplicação da mineração de dados:** Análise Exploratória de Dados e algoritmos de mineração são utilizados para encontrar os padrões.

**4º - Interpretação e avaliação dos resultados:** os resultados são encontrados, interpretados e utilizados para tomada de decisões futuras que auxiliem os estudantes, professores, tutores e administradores na melhora do processo de ensino e aprendizagem.

## 2.1 Disciplina MAT02219 - Probabilidade e Estatística em EaD da UFRGS

A disciplina de Probabilidade e Estatística na modalidade a distância da UFRGS é dividida em três grandes áreas de ensino. Cada área é ministrada por materiais disponibilizados em slides, videoaulas, listas de exercícios, simulados, chats e fóruns. Além disso, outros materiais como manual de boas-vindas, cronograma, plano de ensino, outros materiais extras e apostila estão à disposição do aluno para complementar seus estudos.

A primeira área abrange os conteúdos relacionados à estatística descritiva, representações gráficas e probabilidade. A segunda aborda sobre variáveis aleatórias, modelos probabilísticos e estimação. A terceira e última área explana sobre testes de hipóteses, correlação e regressão linear simples.

Os slides contêm conceitos, exemplos de aplicações, exercícios resolvidos e sugestões de exercícios com gabarito; além de exercícios que devem ser realizados para auxiliar na fixação do estudo. Os vídeos apresentam a explicação detalhada de cada slide sendo mais completo assisti-lo do que apenas ler o material. Eles possibilitam uma maior compreensão daquilo que foi lido nos slides.

Os slides foram criados em .ppt, contendo conceitos, exemplos de aplicações, exercícios resolvidos e sugestões de exercícios com gabarito. Já os vídeos foram gravados pela professora que buscou desenvolvê-los com uma boa qualidade de áudio e vídeo. Para isso utilizou o *software* gratuito *Atube Catcher* que permite capturar a tela do computador, enquanto a matéria vai sendo explicada e gravada. Após as gravações, edições nos vídeos foram realizadas para juntá-los e acrescentar uma capa de entrada.

Os simulados são compostos por exercícios de múltipla escolha com dez questões, cada uma valendo 1 ponto. É por meio deles que os alunos se familiarizam com o estilo das provas, conhecem seus pontos fortes e fracos e podem definir estratégias para a resolução das questões. A disciplina possui um banco de dados no Moodle com mais de 1000 questões, separado nas três grandes áreas. Os simulados podem ser acessados uma única vez por cada aluno. O interessante é que, a cada acesso, as

questões aparecem de forma aleatória, possibilitando que diferentes questões apareçam para cada aluno.

O chat auxilia os alunos no esclarecimento de dúvidas. O professor/tutor estabelece um dia e horário para que os alunos possam entrar no chat, conversarem com o professor e monitor para debaterem sobre assuntos respectivos da área e sanar dúvidas. Ao todo são abertos três chats, os quais acontecem na semana que ocorrerá a prova para que os alunos possam tirar suas últimas dúvidas.

Se o aluno desejar entrar em contato para tirar dúvidas antes da abertura dos chats, ele poderá enviar mensagens no fórum. Cada uma das áreas possui seu respectivo fórum. Eles são abertos no início da disciplina e ficam disponíveis para acesso durante toda o semestre. O monitor da disciplina é quem controla os fóruns ajudando os alunos a sanarem suas dúvidas.

As provas de cada área são realizadas presencialmente. Elas são constituídas de dez questões de múltipla escolha, cada questão valendo 1 ponto. A nota final  $Y$  é obtida por média simples destas avaliações, produzindo o conceito final de cada estudante da seguinte forma: **A** ( $9 \leq Y \leq 10$ ), **B** ( $7,5 \leq Y < 9$ ), **C** ( $6 \leq Y < 7,5$ ). Caso  $Y$  seja menor que 6, o aluno tem direito a recuperação e caso persista esse score, seu conceito é **D**. Alunos que não atingem 75% de presença têm conceito **FF**.

## 2.2 Coleta de dados

As informações utilizadas são de uma base de dados provenientes de 8 turmas ministradas nos semestres de 2016/2 a 2018/1 da disciplina MAT02219 - Probabilidade e Estatística da UFRGS na modalidade EaD. Todas as turmas foram ministradas com o auxílio da Plataforma Moodle.

Tendo em vista a confidencialidade das informações dos alunos envolvidos, todos os dados de identificação foram omitidos, conforme descrito no TCLE (Termo de Consentimento Livre e Esclarecido), disponível no APÊNDICE A. De acordo com as

orientações do Comitê de Ética da UFRGS, o TCLE foi entregue a todos os 504 alunos que participaram da referida disciplina. Destes 504 alunos, 6 não concordaram em participar da pesquisa e 46 reprovaram com conceito FF (reprovação por falta), sendo excluídos da amostra. Posto isto, o número de observações analisadas no trabalho totalizou 452 alunos, perfazendo um montante de 112.051 interações armazenadas no *log* do Moodle, apresentadas na Tabela 1. Considera-se que os dados são provenientes de uma amostra de alunos do EaD do período de 2016/02 a 2018/01 e que os resultados obtidos neste trabalho serão utilizados para inferir para os alunos que farão a disciplina nos próximos semestres.

**Tabela 1-** Informações do total de alunos e dos logs separados por turmas

Turma	Quantidade		Média
	Alunos	Logs	
2015/02-L	45	12.209	271,31
2017/01-K	58	16.222	279,69
2017/01-L	50	15.014	300,28
2017/02-K	58	15.272	263,31
2017/02-L	62	17.094	275,71
2017/02-M	51	14.680	287,84
2018/01-K	68	20.109	295,72
2018/01-L	60	16.060	267,67
<b>Total</b>	<b>452</b>	<b>126.660</b>	<b>280,22</b>

### 2.3 Pré-processamento dos dados

Toda interação do aluno com o AVA, ou seja, toda informação gerada devido ao clique que o aluno executa sobre cada material disponibilizado na plataforma, é armazenada em um arquivo *log* que contém o id do usuário, o horário, a data do clique e o módulo que foi acessado. Cada um desses registros possui dados específicos sobre uma única interação do usuário com a plataforma.

Essas informações, quando extraídas, são exportadas em formato .csv. Porém as informações não são passíveis de serem analisadas diretamente. É necessário que

uma etapa de refinamento dos dados seja executada antes de realizar a mineração desses dados. Por isso, o pré-processamento de dados é essencial para que posteriormente seja realizada a mineração de dados.

### **2.3.1 Extração, manipulação e agrupamento dos bancos de dados**

Esta etapa constitui a extração, manipulação e agrupamento dos bancos de dados. Ela foi realizada no *Software Excel 2016* e no *Software R – versão 3.5.1*. Cada um dos passos executados é representado a seguir:

- I. Dados dos alunos de cada turma registrados nos *logs* foram exportados separadamente da plataforma Moodle em formato .csv e depois agrupados.
- II. Paralelamente, os dados dos simulados de cada área foram extraídos com suas respectivas notas (representadas de 0 a 10).
- III. Ambos os bancos de dados, dos passos I e II, foram combinados em relação as suas informações, fazendo com que cada linha pertencesse a um único aluno. No primeiro banco, as colunas ficaram contendo o nome das variáveis (ações) e a contagem de quantos cliques (interações) em cada material ocorreram no decorrer do curso. No segundo, as colunas contêm as notas em cada um dos simulados.
- IV. Um outro banco de dados foi criado apenas com as informações da nota final de cada aluno e das notas das provas de cada uma das três áreas.
- V. Os três bancos criados foram combinados. O banco final contém todas as informações respectivas de cada aluno. Ademais, uma variável denominada de “Aprovação” foi criada e acrescentada ao respectivo banco. A partir do conceito final do aluno, se sua nota final for maior ou igual a 6,0 ele está na categoria de ‘Aprovado’, caso contrário, ‘Reprovado’.

- VI. Por fim, registros de acessos dos professores, tutores e monitores foram excluídos do banco final (passo V), pois não pertenciam ao objeto de estudo (interações do aluno no AVA), e as informações dos 6 alunos que não autorizaram que seus dados fossem utilizados também foram retiradas, assim como dos 46 alunos com conceito final FF.

### 2.3.2 Variáveis

Após realizada a etapa de extração, manipulação e agrupamento dos bancos de dados, foi possível identificar todas as variáveis (materiais disponibilizados no curso) contidas no banco final (relatado na seção 2.3.1 – passo V).

Ao total foram contabilizadas **62** variáveis: 15 pertencentes ao conteúdo da primeira área (amostragem, estatística descritiva, representações gráficas e probabilidade), 9 variáveis da segunda área (variáveis aleatórias, modelos probabilísticos e estimação); e 13 da terceira área (testes de hipóteses, correlação e regressão). As demais são relacionadas às informações gerais da disciplina, aos simulados e notas das provas.

A Figura 3 apresenta as informações contidas em cada variável. Do total de 62 variáveis, 46 estão relacionadas aos cliques que os alunos realizaram nos materiais no decorrer do curso. Distribuídas nas três áreas, 7 variáveis correspondem as notas dos simulados, notas das provas e a nota final (média das três provas). As outras seis variáveis correspondem aos simulados, se o aluno realizou ou não, e quantos dias antes de cada prova ele o realizou. Para completar o total de variáveis, também consta a variável 'Aprovação', informando se o aluno foi aprovado ou reprovado na disciplina. A Figura 4 apresenta a descrição de cada subárea.

Informação		Conteúdo	Variável	Identificação da variável
MOODLE	GERAL	Identificação	ID	Identificação da linha
		Turma	TURMA	Turma
		Boas Vindas	boas_vindas	Boas Vindas
	LOGS	Cronograma	cronog	Cronograma da disciplina
		Plano de Ensino	plano_ensino	Plano de Ensino da disciplina
		Vídeo Extra	video_extra	Vídeos Extras
		Arquivos Extras	arq_extra	Arquivos Extras
		Orientações	orien_inicial	Orientações iniciais
		Acesso	acesso	Acesso a página do curso
		Chat	chat	Chats com professores e monitores
		Fórum	forum_areaX	Fóruns de dúvidas de cada área X (X = 1, 2 e 3)
		Slides	slide_areaX	Cliques nos slides da área X (X = 1, 2 e 3)
			slide_area1.K	Cliques no slide da área 1.K (K = 1 a 7)
			slide_area2.W	Cliques no slide da área 2.W (W = 1 a 4)
		Vídeos	slide_area3.Z	Cliques no slide da área 3.Z (Z = 1 a 6)
url_areaX	Cliques nos vídeos da área X (X = 1, 2 e 3)			
url_area1.K	Cliques no vídeo da área 1.K (K = 1 a 7)			
Lista de Exercícios	url_area2.W	Cliques no vídeo da área 2.W (W = 1 a 4)		
	url_area3.Z	Cliques no vídeo da área 3.Z (Z = 1 a 6)		
	lista_areaX	Listas de Exercícios de cada área X (X = 1, 2 e 3)		
SIMULADO	nota_simul_areaX	Nota final do simulado da área X (X = 1, 2 e 3)		
	fez_simul_areaX	Fez ou não o simulado da área X (X = 1, 2 e 3)		
	simulado_areaX	Acesso aos simulados		
	tempo_antes_areaX	Tempo que realizou o simulado área X antes da prova área X correspondente (X = 1, 2 e 3)		
PROVA PRESENCIAL	Nota das Provas	nota_areaX	Nota da prova X (X = 1, 2 e 3)	
		rec	Nota da recuperação	
	Notal Final <sup>2</sup>	nota_final	Nota final do curso	
	Aprovação	aprovacao	Aprovado ou Reprovado	

**Simulado<sup>1</sup>:** Cada uma das 3 áreas possuem as variáveis respectivas ao simulado, totalizando 9 variáveis.

**Nota Final<sup>2</sup>:** inclui nota das 3 provas mais a recuperação.

**Figura 3** - Informações das variáveis do banco de dados

<b>Nome</b>	<b>Descrição</b>
<b>Área 1: Estatística descritiva, representações gráficas e probabilidade</b>	
Material 1-1	Introdução
Material 1-2	Descritiva distribuições de frequência
Material 1-3	Representações gráficas
Material 1-4	Descritiva medidas descritivas
Material 1-5	Análise exploratória
Material 1-6	Probabilidade introdução e teoremas
Material 1-7	Probabilidade condicional e Bayes
<b>Área 2: variáveis aleatórias, modelos probabilísticos e estimação</b>	
Material 2-1	Variáveis aleatórias discretas e distribuições
Material 2-2	Variáveis aleatórias contínuas e distribuições
Material 2-3	Distribuições amostrais
Material 2-4	Estimação
<b>Área 3: testes de hipóteses, correlação e regressão linear simples</b>	
Material 3-1	Testes de hipóteses introdução
Material 3-2	Testes de hipóteses médias
Material 3-3	Testes de hipóteses variâncias
Material 3-4	Testes de hipóteses proporções
Material 3-5	Coefficiente de correlação
Material 3-6	Regressão linear simples

**Figura 4** - Descrição das variáveis de cada subárea

## 2.4 Mineração de Dados Educacionais

A Mineração de Dados Educacionais é constituída de Análise Exploratória de Dados (AED) e algoritmos de mineração, que são utilizados para encontrar os padrões desejados. Para este trabalho foram realizadas Análises Exploratórias de Dados para identificar os comportamentos dos alunos na plataforma Moodle em relação ao seu desempenho final na disciplina.

A AED é realizada para conquistar uma compreensão básica dos dados e das relações entre as múltiplas variáveis em estudo. Para explorar os dados e identificar alguns tipos de comportamento das possíveis variáveis de acesso em relação às variáveis de desempenho na disciplina – nota (prova 1, 2, 3 e nota final) e resultado (apro-

vado ou reprovado) – gráficos foram gerados, utilizando as funções do ggplot2 do software R versão 3.5.1, com auxílio do Excel versão 2016.

A análise gráfica foi bastante explorada neste estudo, buscando identificar as variáveis de acesso dos alunos no Moodle que estão relacionadas com seu desempenho em cada uma das provas. Por exemplo, foram cruzados os acessos e notas obtidas nos simulados com a nota obtida na prova correspondente. Entre os gráficos utilizados estão: barras, histograma, dispersão e boxplot.

Quando visualizado resultado importante nos gráficos, foram aplicados testes de hipóteses para comparação de médias, como o teste t de Student e calculados coeficientes de correlação de Spearman para identificar associações entre duas variáveis. Além disso, foram elaborados diagramas em formato de fluxogramas com informações de acesso aos diversos materiais da disciplina, relacionando-os com as médias obtidas nas provas de cada área.

O teste t de Student é utilizado para comparar médias de dois grupos de indivíduos para a mesma variável medida quando os dados seguem uma distribuição normal. A hipótese nula, na qual considera que as médias são iguais, não é rejeitada quando o valor de  $|t_{\text{observado}}| < |t_{\text{tabelado}}|$ , ou seja, a média dos dois grupos não diferem entre si. Para este estudo foi considerado o nível de significância de 5% ( $p < 0,05$ ). Este teste foi utilizado, pois partiu-se do pressuposto de que as médias das notas seguem distribuição normal assintota pelo Teorema Central do Limite (TCL) (WRIGHT, 2002).

O coeficiente  $r$  de Spearman mede a intensidade da relação entre duas variáveis. Ao invés do valor observado é usado apenas a ordem das observações. Deste modo, este coeficiente não é sensível a assimetrias na distribuição, nem à presença de outliers. O coeficiente varia entre -1 e 1. Quanto mais próximo estiver destes extremos, maior será a associação entre as variáveis. O sinal negativo da correlação significa que as variáveis variam em sentido contrário, isto é, as categorias mais elevadas de uma variável estão associadas a categorias mais baixas da outra variável (WRIGHT, 2002).

### 3. RESULTADOS

A disciplina de Probabilidade e Estatística da UFRGS na modalidade EaD, objeto deste estudo, foi ministrada nos períodos de 2016/02 a 2018/01, totalizando 8 turmas. Em um total de 452 alunos considerados neste estudo, 377 (83,4%) obtiveram nota final maior ou igual a 6,0, obtendo aprovação na disciplina, em contrapartida 75 (16,6%) foram reprovados. Desses aprovados, 16,71% obtiveram conceito final A ( $9,0 \leq \text{nota final} \leq 10,0$ ); 38,46%, conceito B ( $7,5 \leq \text{nota final} < 9,0$ ); e 44,83%, conceito C ( $6,0 \leq \text{nota final} < 7,5$ ).

A Figura 5 mostra a distribuição das notas de cada área e da nota final e seus respectivos valores discrepantes (*outliers*) representadas pelo boxplot. Observa-se que não há muita dispersão entre as notas finais dos alunos, observando-se poucos alunos com notas discrepantes. As notas das provas ficaram próximas da mediana 7, exceto a nota da área 2, na qual houve uma queda no valor da mediana e uma maior dispersão entre as notas.

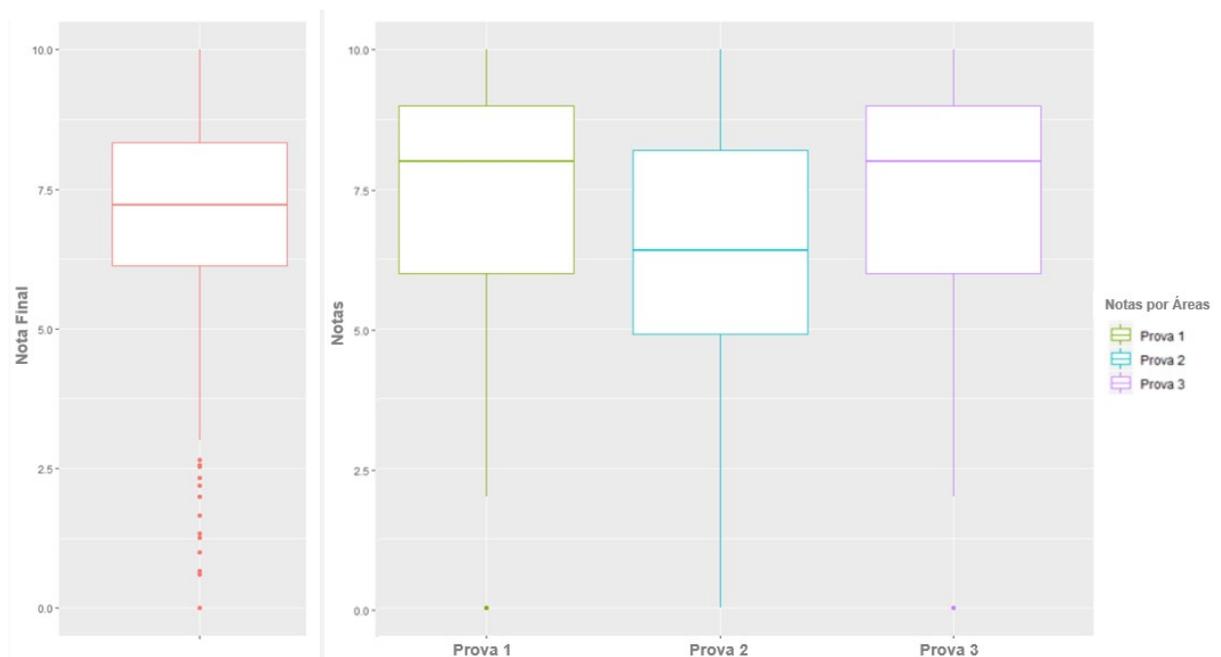
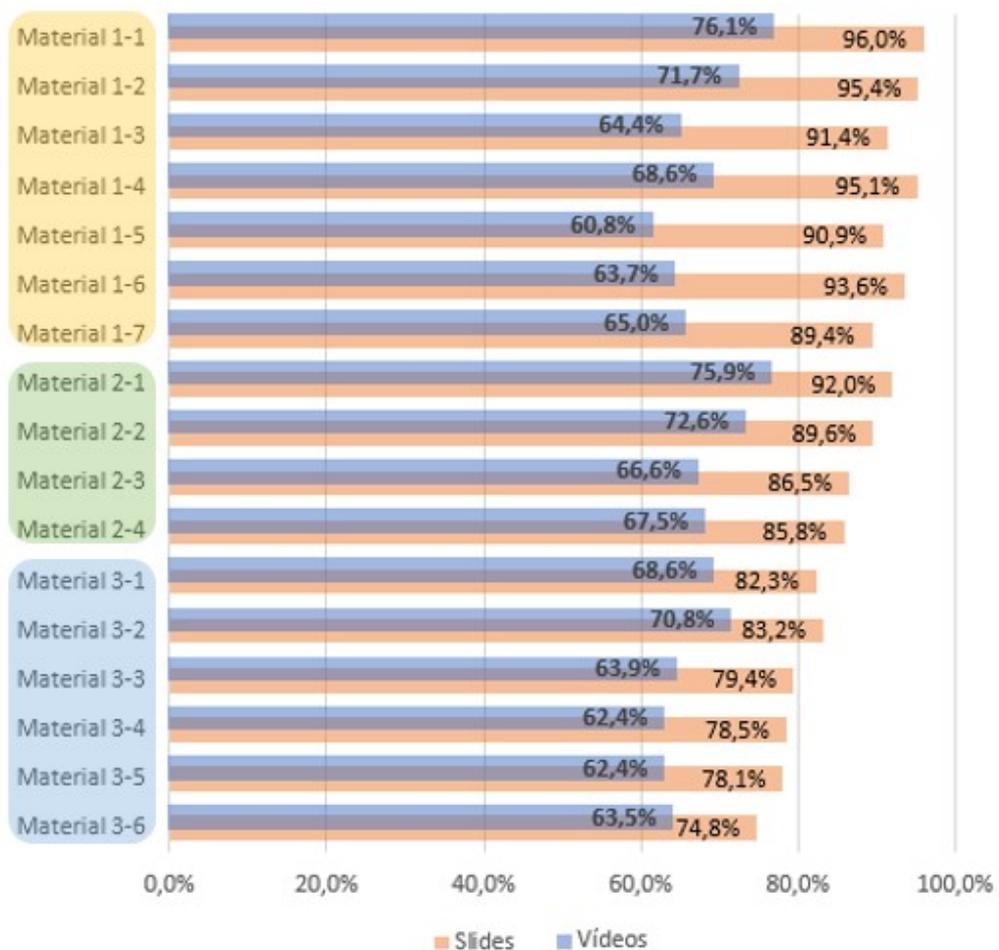


Figura 5 - Boxplot das notas das provas de cada área e nota final

Esta relação apresentada possivelmente ocorre, pois os alunos se deparam na primeira área com a estatística descritiva, que, por ser uma área da estatística com um

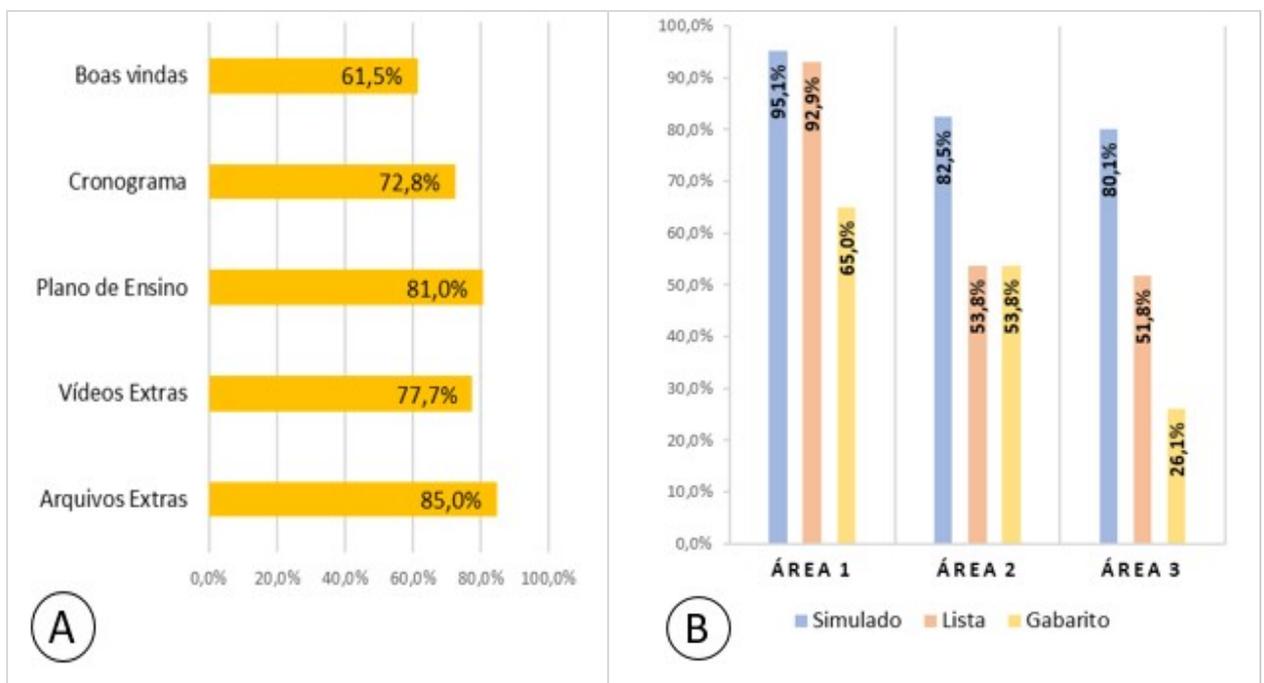
grau de facilidade de compreensão maior, os alunos tendem a ir melhor na prova. Já na segunda área, os alunos podem ter baixado seu desempenho pelo fato de se depararem com o assunto de distribuições de probabilidade que requer uma dedicação maior para entender seus conceitos e aplicações. A terceira área que apresenta assuntos da área inferencial da estatística possibilita que o aluno volte a melhorar seu desempenho.

A Figura 6 apresenta o percentual de alunos que acessaram os materiais de slides e vídeos de cada uma das 17 subáreas das três áreas do conteúdo. Ao analisar o gráfico, observa-se que os vídeos são menos acessados que os slides. Além disso, é possível observar que ao longo da disciplina os alunos tendem a acessar menos os materiais, independentemente se são slides ou vídeos com o conteúdo da disciplina.



**Figura 6** - Percentual de alunos que acessaram aos slides e vídeos

Pode-se também observar no gráfico A da Figura 7 que muitos alunos acabam não acessando materiais essenciais para a compreensão da disciplina, como arquivos de Boas Vindas, Cronograma e Plano de Ensino. No gráfico B pode-se identificar uma decrescente no número de alunos que acessaram os simulados, listas e gabaritos no decorrer do curso. Esse comportamento pode ter sido ocasionado devido à evasão dos alunos no decorrer do curso e/ou mau desempenho, entre outros fatores externos que não são possíveis identificar apenas analisando os cliques.



**Figura 7** - Percentual de alunos que acessaram aos arquivos do Moodle

### 3.1 Análise dos Simulados

Nesta seção serão apresentados os resultados das análises realizadas com os cliques e notas dos simulados.

Uma análise para entender a relação dos simulados com a nota das provas foi realizada. Conforme apresentado na Tabela 2, o aluno que realiza os simulados antes de cada prova, tende a ter um desempenho melhor do que os que não o realizam. Po-

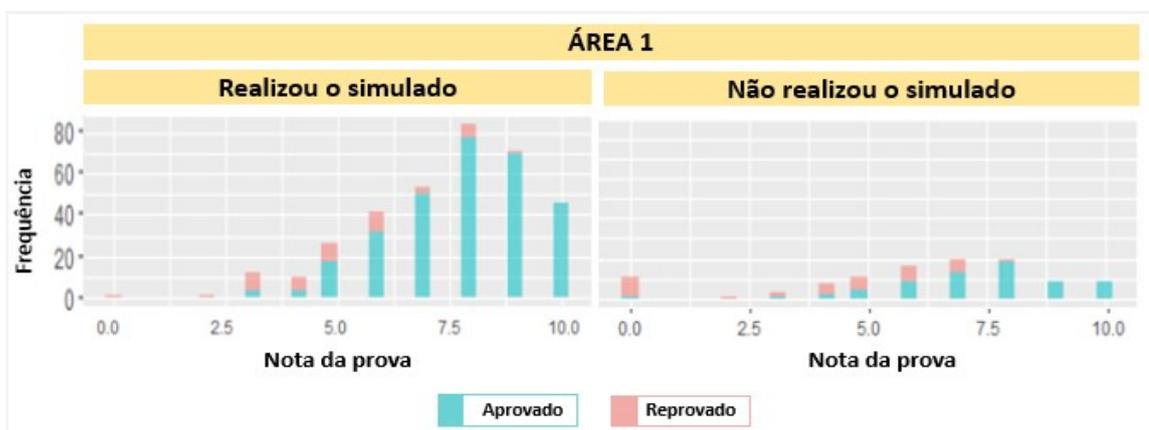
de-se observar que a nota média da Prova 1 foi de 7,52 entre os que fizeram o simulado, já os que não realizaram o simulado apresentaram uma média inferior, de 6,15, e isso ocorre com todas as demais provas.

**Tabela 2** - Comparação entre realização do simulado e a média das notas das provas

Notas	Simulado realizado											
	Sim				Não				Total			
	n	(%)	Média	(DP)	n	(%)	Média	(DP)	n	(%)	Média	(DP)
Prova 1	347	( 76,77)	7,52	( 1,86)	105	( 23,23)	6,15	( 2,77)	452	( 100,00)	7,20	( 2,18)
Prova 2	322	( 71,24)	6,84	( 2,29)	130	( 28,76)	4,74	( 2,99)	452	( 100,00)	6,23	( 2,68)
Prova 3	323	( 71,46)	7,71	( 1,82)	129	( 28,54)	5,60	( 3,42)	452	( 100,00)	7,11	( 2,57)

Partindo da premissa de que a média das notas das provas segue distribuição normal assíntota pelo TCL, a partir do teste t de Student verificou-se diferença significativa entre as médias das provas dos alunos que realizaram e não realizaram o simulado, para todas as 3 áreas. Com base nestas informações, pode-se afirmar que as médias das notas dos alunos que realizam os simulados antes das provas é significativamente maior que a dos alunos que não o realizam ( $p < 0,001$ ).

Este resultado também pode ser visto na Figura 8, Figura 9 e Figura 10 que apresentam dois gráficos, no qual comparam os alunos que realizaram e não realizaram o simulado em relação a sua nota da prova, discriminando-os pelo resultado de aprovação. Pode-se observar que os alunos que realizam o simulado tendem a tirar notas mais altas nas provas quando comparado aos que não o realizam.



**Figura 8** - Comparação entre alunos que realizaram e não realizaram os simulados em relação à nota da prova separados pelo resultado de aprovação da área 1

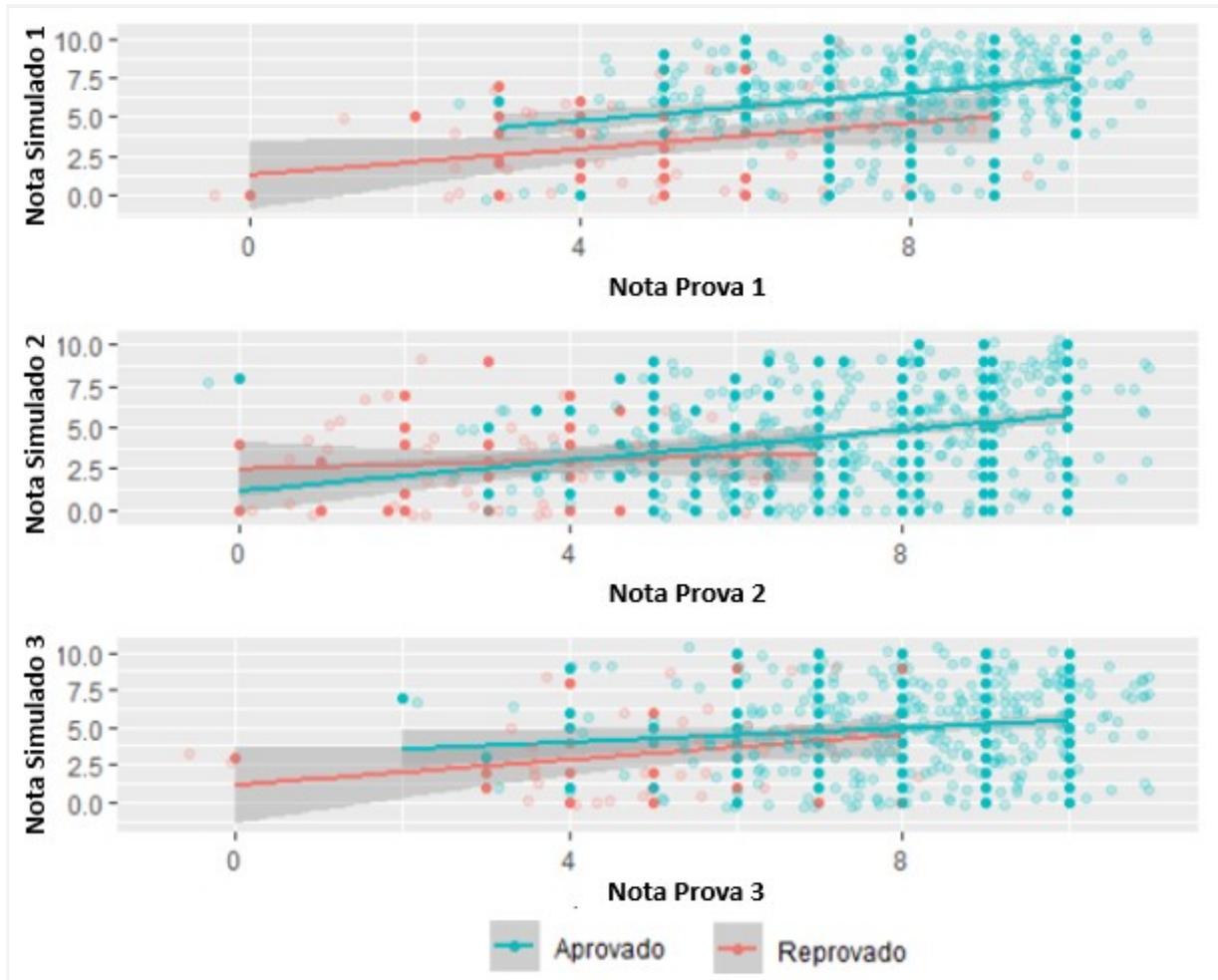


**Figura 9** - Comparação entre alunos que realizaram e não realizaram os simulados em relação à nota da prova separados pelo resultado de aprovação da área 2



**Figura 10** - Comparação entre alunos que realizaram e não realizaram os simulados em relação à nota da prova separados pelo resultado de aprovação da área 3

Considerando os alunos que fizeram os simulados, foi realizada uma análise de correlação para identificar se a nota do simulado tem associação linear positiva com a nota da prova. Os alunos que tiram notas altas no simulado tendem também a tirar notas altas na prova correspondente (coeficientes de correlação de Spearman ( $r$ ): área 1 ( $r = 0.46$ ), área 2 ( $r = 0.36$ ) e área 3 ( $r = 0.25$ )). Esta relação pode ser vista na Figura 11 que apresenta os gráficos de dispersão comparando as notas dos simulados em relação às notas das provas das três áreas, separando os alunos pelo resultado de aprovação.



**Figura 11** - Gráfico de dispersão: comparação entre as notas das provas e dos simulados separados por aprovação

Foram realizadas matrizes de gráficos para comparar todas as variáveis do estudo em relação ao desempenho do aluno, tanto por área como por desempenho final, foram criadas, conforme apresentada na Figura 12. Essa figura mostra uma matriz de gráficos que relaciona as notas das provas de cada área com as notas dos simulados, separando os alunos em aprovados e reprovados. A forma de olhar essa matriz é comparando as informações das linhas com sua respectiva coluna.

Ao olhar na Figura 12 para a coluna “Prova 1” e linha “Simulado 1” consegue-se ver o gráfico de dispersão (igualmente apresentado na Figura 11) que apresenta a rela-

ção linear entre as duas variáveis sendo possível identificar que alunos que tiram notas altas nos simulados tendem a tirar notas altas nas provas.

O comportamento das notas dos simulados em relação à classificação de aprovação do aluno pode ser identificada nos gráficos de densidade representados na diagonal principal dessa matriz. Ao analisar a nota do “Simulado 1” consegue-se observar que os alunos aprovados são aqueles que tendem a tirar notas acima da média nos simulados (curva azul à direita do gráfico). Também boxplots e histogramas são apresentados na matriz.

Os boxplots apresentados na coluna “Aprovação” são relacionados com cada uma das notas das provas, podendo perceber que alunos aprovados são os que têm melhor desempenho nos simulados, comparado aos que foram reprovados. Já os histogramas apresentados na última linha da matriz apresentam que há uma maior frequência de alunos que acessam aos simulados e são aprovados do que os reprovados.

Em suma, ao analisar essa matriz de gráficos, pode-se perceber que há indícios de que realizar os simulados está associado positivamente na melhora da nota das provas.

Também uma análise de correlação foi realizada para identificar se há relação entre quantos dias antes de cada prova era realizado o simulado comparado à nota da prova. Acreditava-se que o aluno que realizava os simulados com mais antecedência teria um desempenho melhor na prova, porém não houve correlação forte entre essas variáveis em nenhuma das 3 áreas (coeficiente de correlação de Spearman ( $r$ ): área 1 ( $r = -0.02$ ), área 2 ( $r = -0.06$ ) e área 3 ( $r = -0.09$ )).

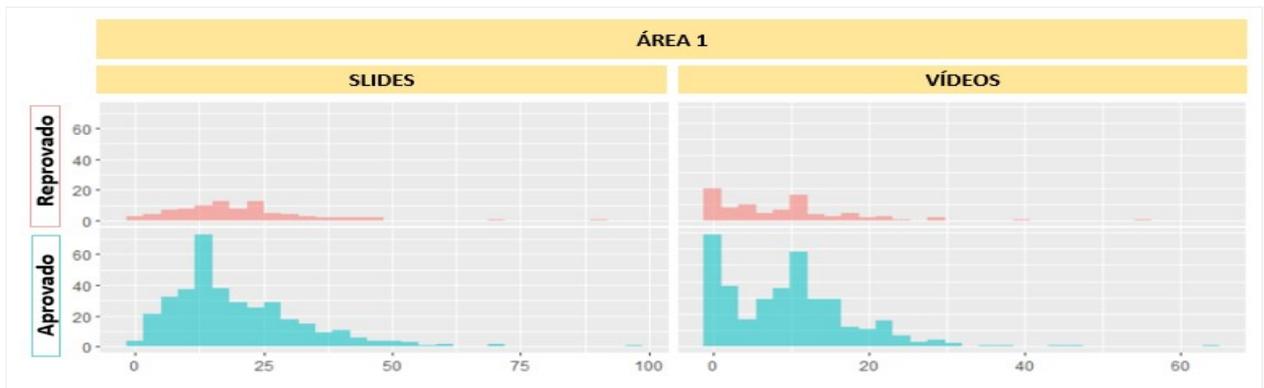


**Figura 12** - Gráficos de dispersão, densidade, boxplots e histograma: comparação entre as notas das provas e notas dos simulados separados por aprovação

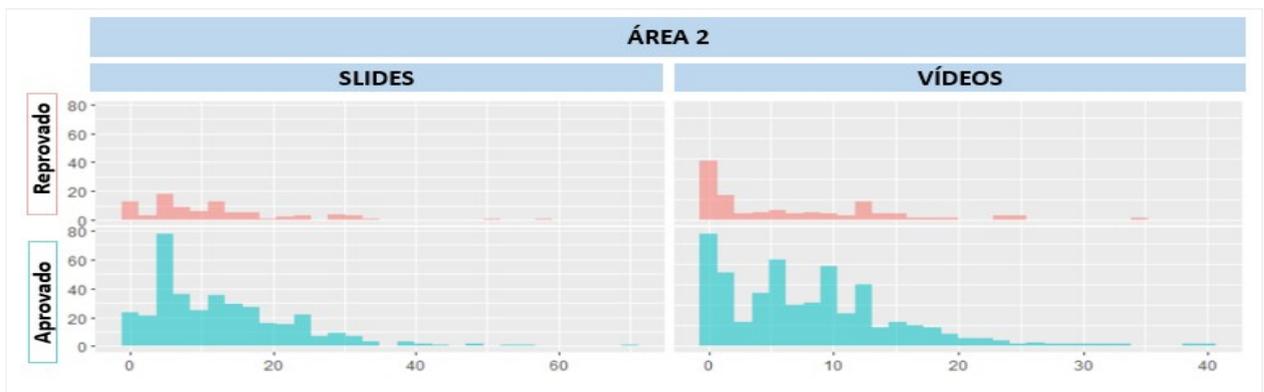
### 3.2 Análise dos Slides e Vídeos

Na seção 3.1 foi mostrada uma análise geral de cada uma das 17 subáreas. Nesta subseção serão apresentados os resultados das análises realizadas dos slides e vídeos por área. As 17 subáreas foram agrupadas dentro de suas respectivas áreas para melhor visualizar e compreender seu comportamento em relação às notas das provas, separados pelo resultado de aprovação.

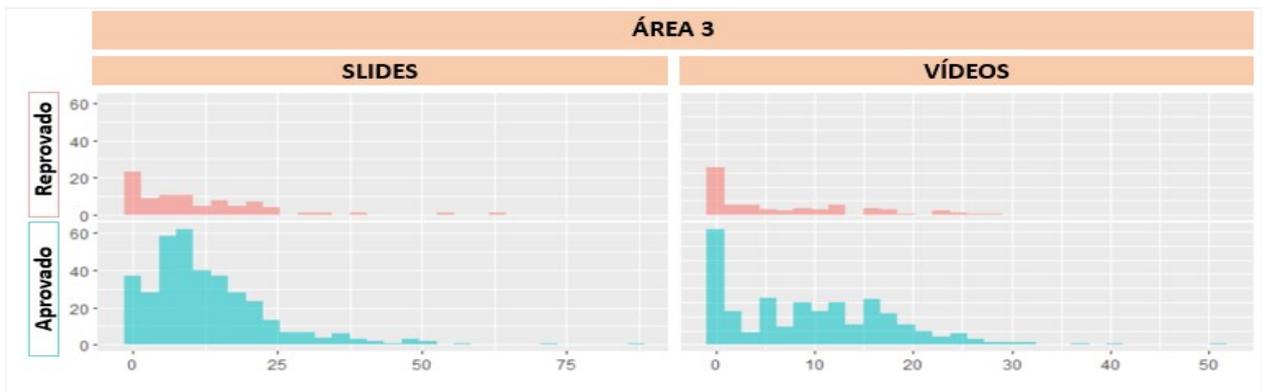
Uma comparação entre a aprovação final do aluno na disciplina e a quantidade de acessos (cliques) nos materiais de slides e vídeos foi realizada. Essa comparação pode ser vista na Figura 13, Figura 14 e Figura 15. Alunos aprovados tendem a acessar mais vezes os materiais, tanto de slides quanto de vídeos, quando comparados aos reprovados. Isso ocorre em todas as 3 áreas.



**Figura 13** - Comparação da quantidade de acessos aos slides e vídeos acessados durante o curso e o resultado de aprovação dos alunos da área 1



**Figura 14** - Comparação da quantidade de acessos aos slides e vídeos acessados durante o curso e o resultado de aprovação dos alunos da área 2

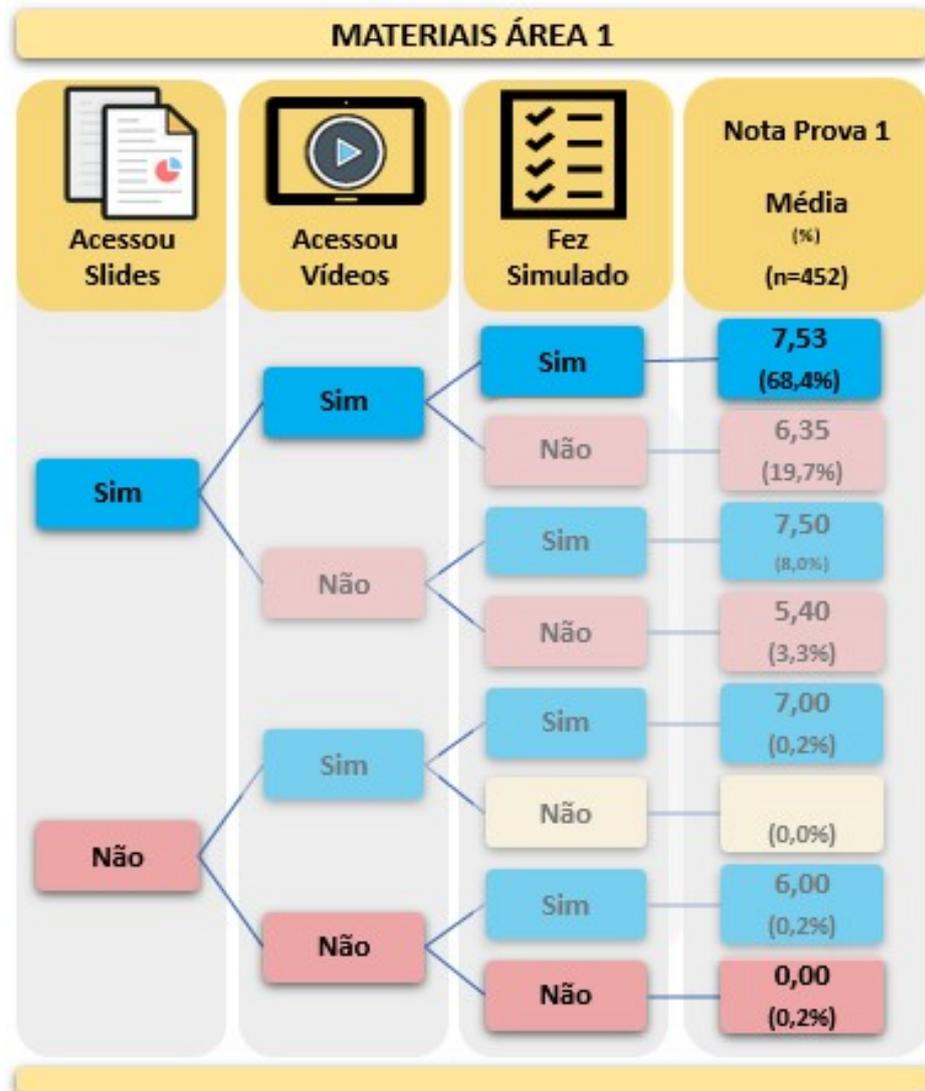


**Figura 15** - Comparação da quantidade de acessos aos slides e vídeos acessados durante o curso e o resultado de aprovação dos alunos da área 3

### 3.3 Análise do Desempenho do Aluno

Para compreender sobre o que ocorre com a média de cada uma das provas, passou-se de uma visão ampla para uma análise mais afunilada. Fluxos, apresentados na Figura 16, Figura 17 e Figura 18, foram criados para mostrar o que ocorre com a média das notas das provas quando o aluno acessa aos slides, vídeos e realiza os simulados.

Observa-se que alunos que tiram nota maior nas provas são aqueles que acessam aos slides, vídeos e realizam os simulados. As maiores médias das provas são dos alunos que pelo menos acessaram aos slides e vídeos e realizaram os simulados, tanto na área 1 quanto nas áreas 2 e 3, tendo como médias 7,53, 6,93 e 7,85, respectivamente.



**Figura 16** - Fluxo do acesso aos materiais da área 1 para compreender o resultado médio das provas

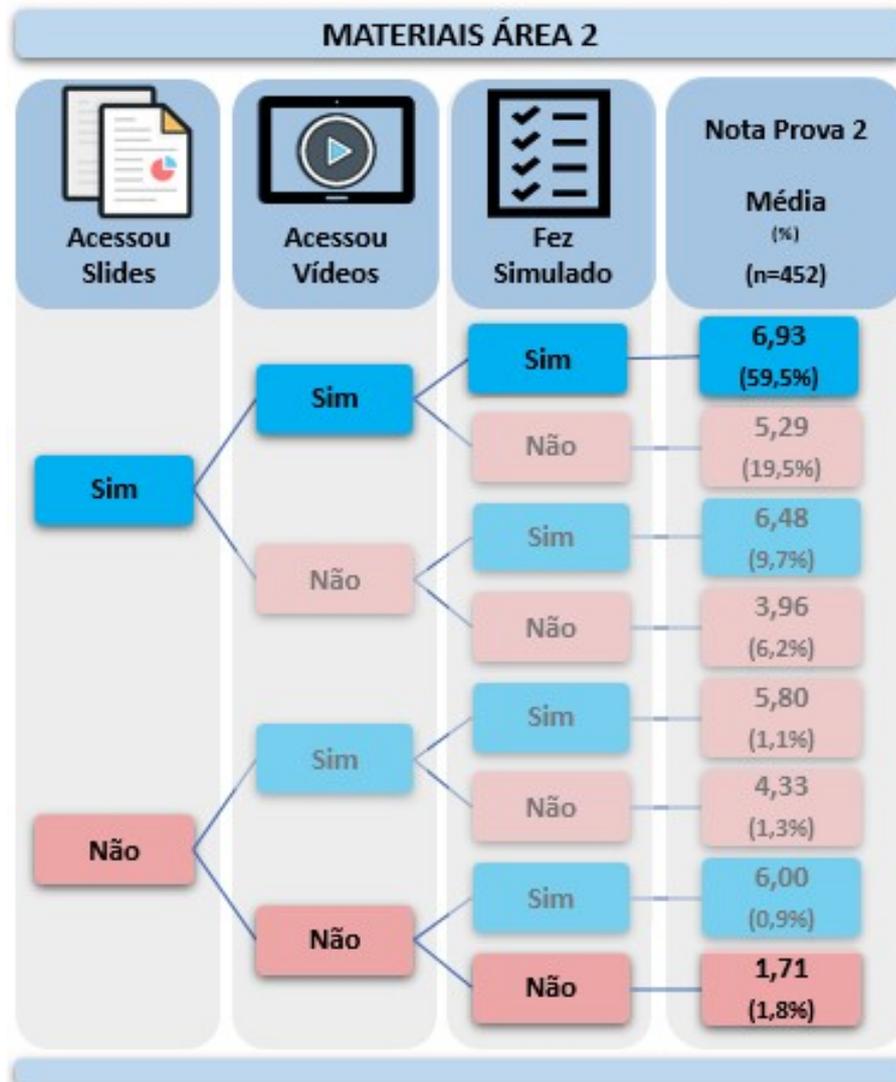
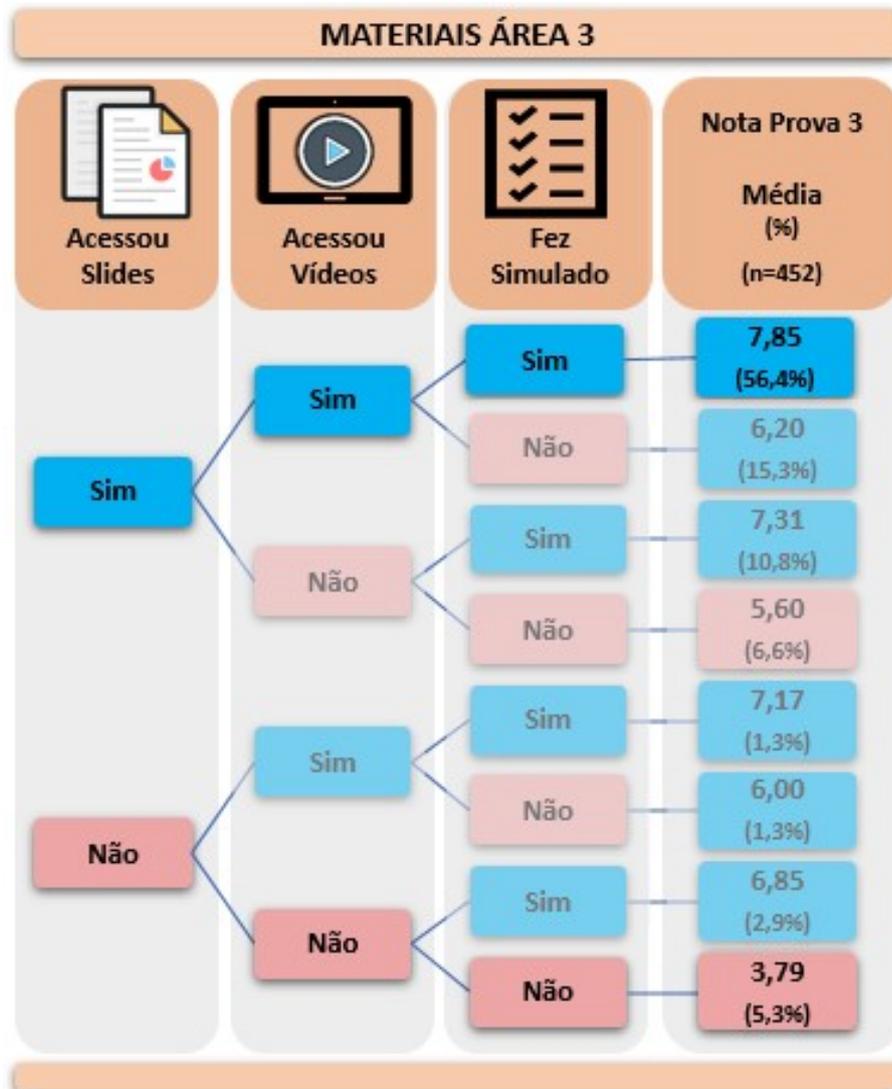


Figura 17 - Fluxo do acesso aos materiais da área 2 para compreender o resultado médio das provas



**Figura 18** - Fluxo do acesso aos materiais da área 3 para compreender o resultado médio das provas

A Figura 19 apresenta o fluxo inverso da análise dos acessos aos materiais, partindo da aprovação do aluno na disciplina. Ou seja, dado que o aluno foi ou não aprovado na disciplina, qual o percurso que ele realizou dentro da plataforma Moodle durante o período eletivo.

Pelo fluxo inverso é possível ter uma visão geral do que fora apresentada na Figura 16, Figura 17 e Figura 18 sobre o desempenho final do aluno na disciplina, se o aluno foi aprovado ou reprovado e o que talvez tenha o levado a esse resultado. Partindo do resultado final de aprovação, observa-se que 78,5% dos alunos que foram apro-

vados acessaram os slides ou os vídeos de cada assunto e o simulado dentro de cada área. Não houve nenhum caso de alunos que foram aprovados sem terem acessados os materiais da primeira área.

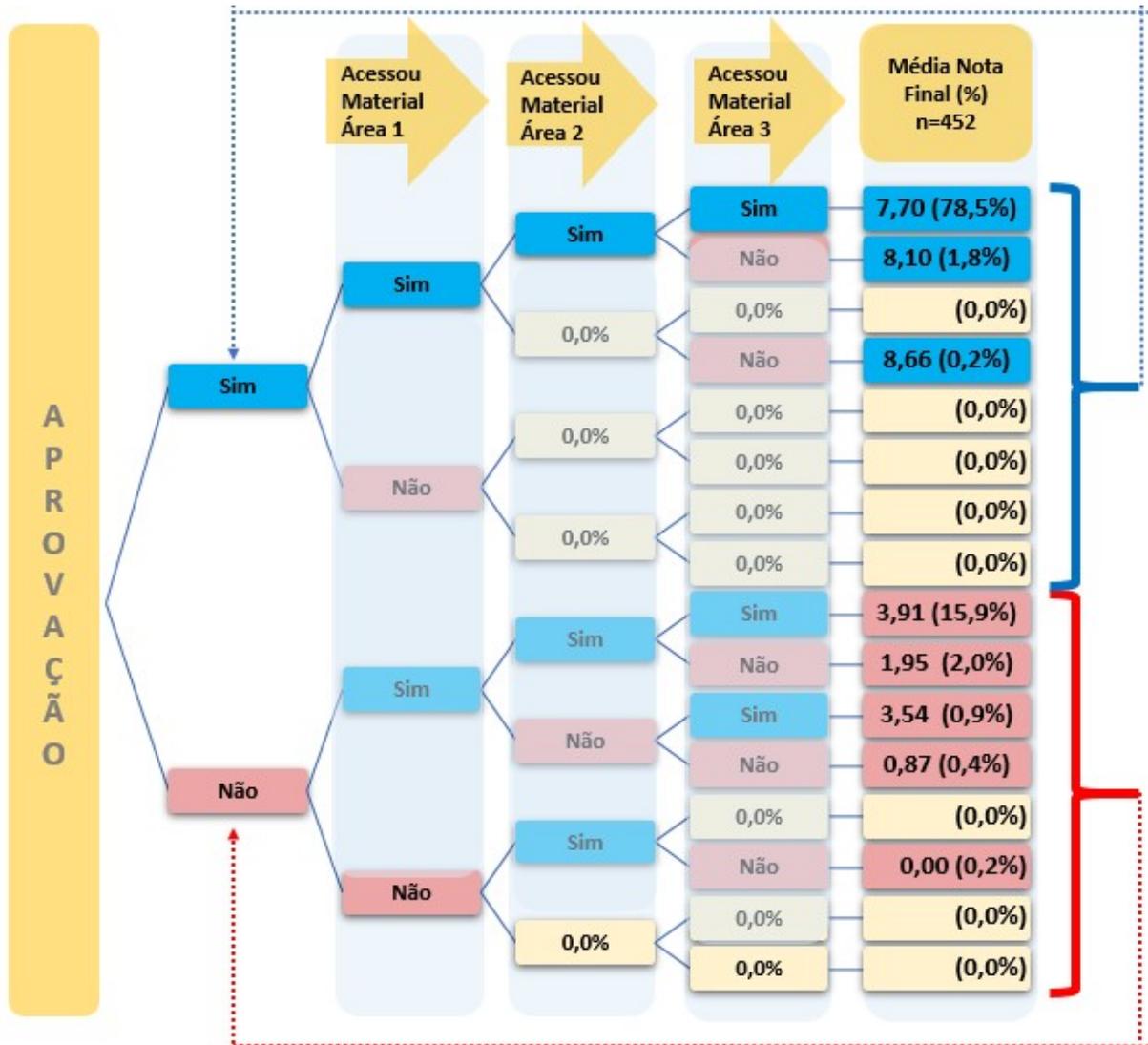


Figura 19 - Fluxo inverso: análise dos acessos aos materiais partindo da aprovação do aluno

#### 4. CONCLUSÕES

Esta pesquisa buscou identificar ações do aluno no uso da plataforma Moodle que tenham relação com o seu desempenho final na disciplina de Probabilidade e Estatística da UFRGS. Para isso os dados gerados pelo *ClickStream* do aluno em cada material ou ferramenta da plataforma foram analisados com base no processo de Mineração de Dados Educacionais que constitui técnicas de Análise Exploratória de Dados.

A partir dos resultados deste estudo foi possível observar que os alunos tendem a apresentar melhor aproveitamento ao participar ativamente de seu processo de aprendizagem. As tecnologias educacionais surgem para constituir um recurso promissor no ensino de Estatística. Isto ocorre porque a plataforma Moodle, que é um formato de Aprendizado Virtual, oferece um ambiente interativo e individualizado, no qual o aluno é constantemente convidado a interagir com os conteúdos, podendo escolher o tipo de mídia que se adapta melhor ao seu estilo de aprendizagem.

Durante o estudo foi possível observar que os alunos que acessam mais os materiais possuem um desempenho superior aos que acessam menos ou não acessam. Vale ressaltar que cada curso EaD possui diferentes materiais e métodos de aprendizagem. Para essa referida disciplina os professores tiveram a preocupação e um grande empenho ao criarem cada slide, vídeo e lista de exercícios para que o aluno pudesse estudar e compreender a matéria da melhor forma possível.

Com a Análise Exploratória de Dados foi possível observar que os alunos estão acessando mais os slides do que os vídeos. Um trabalho futuro do grupo de professores que trabalham no EaD é buscar aprimorar os vídeos.

Também, neste estudo, foi possível identificar que alunos que realizam o simulado tendem a ter um desempenho maior nas provas, comparado aos que não o realizam. Este resultado é de suma importância, pois para que a disciplina pudesse ter esses simulados, cerca de 1000 questões foram armazenadas em um banco de dados, algumas retiradas de concursos públicos e outras criadas pela própria professora que ao longo do planejamento desta disciplina foi alimentando esse banco.

Finalmente, a aplicação da mineração de dados em sistemas educacionais com o auxílio da Análise Exploratória de Dados pôde ajudar a identificar quais são as possíveis ações dos alunos no Moodle que refletem em seu desempenho final da disciplina de Probabilidade e Estatística da UFRGS. Com esse estudo será possível que os professores recomendem aos futuros alunos quais as ações que eles devem tomar na plataforma do Moodle que poderão ajudá-los no seu desempenho final. Não só os estudantes, mas os professores/tutores também são beneficiados com este estudo, pois ao acessar o histórico dos alunos no decorrer do curso poderão classificá-los em grupos com base em sua necessidade de orientação, monitoramento e alertá-los.

## 5. REFERÊNCIAS

- ANJEWIERDEN, A.; KOLLOFEL, B.; HULSHOF, C. Towards educational data mining: Using data mining methods for automated chat analysis to understand and support inquiry learning processes. International Workshop on Applying Data Mining in e-Learning (ADML 2007), Creta, Grécia, 2007.
- BAKER, R. S. J. D.; COSTA, E.; AMORIM, L.; MAGALHÃES, J.; MARINHO, T. Mineração de Dados Educacionais: Conceitos, Técnicas, Ferramentas e Aplicações. Jornada de Atualização em Informática na Educação, 2012.
- FILHO, A. R. P. Moodle: um sistema de gerenciamento de cursos. v. 1.5.2. Brasília: UnB, 2005.
- FILHOS, JOZEMBERG G. S. Análise de logs da plataforma Moodle utilizando técnicas de mineração de dados: um estudo de caso. Trabalho de Conclusão de Curso (Licenciatura em Ciência da Computação). Universidade Federal do Paraíba, 2016.
- GULARTE, A. C.; SELAU, L. P. R. Ensino de Estatística: um estudo comparativo entre modalidades presencial e a distância. In: VII Congresso Internacional de Ensino de Matemática, 2017, Canoas-RS. Anais do VII Congresso Internacional de Ensino de Matemática, 2017.
- INEP. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Sinopse Estatística da Educação Superior 2017. Brasília: Inep, 2018. Acessado em 19 de outubro de 2018. Disponível em: <<http://portal.inep.gov.br/basica-censo-escolar-sinopse-sinopse>>.
- LOPES, C. E. O ensino da estatística e da probabilidade na educação básica e a formação dos professores. Cadernos Cedes, 2008.
- MACHADO, F. N. R. Big data: o futuro dos dados e aplicações. São Paulo: Érica/Editora Saraiva, 224 p.: il, 2018.
- MATTAR, J. Tutoria e interação em educação a distância. São Paulo: Cengage Learning, 2012.
- MENDOZA, L.P.; SWIFT, J. Why teach statistics and probability: a rationale. In: SHULTE, A.P.; SMART, J.R. Teaching statistics and probability. Reston: Yearbook National Council of Teachers of Mathematics, 1981.
- NORONHA, A. B.; VIEIRA, A. R. A utilização da plataforma WebCT para desenvolvimento e implementação de disciplinas utilizando a internet. In: BARBOSA, Rommel Melgaço (org). Ambientes virtuais de aprendizagem. Porto Alegre: Artmed, 2005.

NUNES, L. N. Estatística à Distância: uma experiência. In: SINAPE - Simpósio Nacional de Probabilidade e Estatística, 19º, 2010, São Pedro-SP. Resumo Estendido do 19º Simpósio Nacional de Probabilidade e Estatística.

PAN, W. S. The challenges of teaching statistics in the current technology environment. *Journal of American Academy of Business*, Cambridge, v. 3, Sep. 2003.

PEREIRA, M. F. R.; MORAES, R. A.; TERUYA, T. K. Educação a distância (EaD): reflexões críticas e práticas. Uberlândia: Navegando Publicações, 2017.

PETERS, O. Didática do ensino a distância: experiências e estágio da discussão numa visão internacional. Tradução de Ilson Kayser. São Leopoldo: Unisinos, 2006.

R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2018.

ROMERO, C.; VENTURA, S.; GARCÍA, E. "Data Mining in Course Management Systems: Moodle Case Study and Tutorial". *Computers & Education*, no. 51, 2008.

ROMERO, C.; VENTURA, S. Educational data mining: A survey from 1995 to 2005. Elsevier, 2007.

RStudio Team. RStudio: Integrated Development Environment for R. RStudio, Inc., Boston, MA, 2016.

SENECHAL, A. C. L. Análise e pré-processamento de dados utilizando técnicas de mineração de dados educacionais para o Moodle. Monografia (Ciência da Computação), Universidade Federal de Lavras, 2013.

YADAV, S. K.; PAL. Data Mining Application in Enrollment Management: A Case Study. *International Journal of Computer Applications*. New York, v 41., n° 5, Mar. 2012.

WRIGHT, D. B. First steps in statistics. London: Sage Publication, 2002.

## APÊNDICE A

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

  
INSTITUTO DE MATEMÁTICA  
E ESTATÍSTICA  
UFRGS

**TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO**

Eu, **Lisiane Priscila Roldão Selau**, Professora do Instituto de Matemática e Estatística da Universidade Federal do Rio Grande do Sul, sou orientadora do Trabalho de Conclusão de Curso (TCC) da aluna **Aline Cafruni Gularte**, graduanda em Estatística pela mesma instituição.

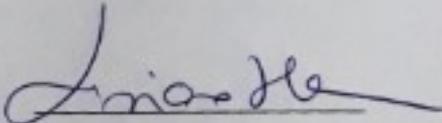
O referido TCC tem como objetivo analisar os dados dos alunos disponibilizados pela plataforma Moodle das disciplinas em EAD de Probabilidade e Estatística, visando compreender o perfil do aluno, suas principais necessidades e dificuldades durante a realização da disciplina, com o objetivo de desenvolver métodos de ensino mais eficazes.

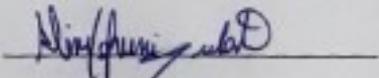
Sua participação neste estudo é muito importante e voluntária. Para que possamos atingir o objetivo proposto, solicito o seu consentimento para utilizar as informações disponibilizadas pela plataforma Moodle da disciplina EAD de Probabilidade e Estatística. Enfatizo que todas as informações serão tratadas de modo confidencial e anônima. **Os dados serão divulgados apenas para fins científicos, mantendo-se o cuidado de assegurar seu anonimato.**

Esta pesquisa é coordenada/orientada pela pesquisadora Lisiane Priscila Roldão Selau (professora do Departamento de Estatística – UFRGS), a quem você poderá contatar a qualquer momento que julgar necessário, através do e-mail [lisianeselau@gmail.com](mailto:lisianeselau@gmail.com) e será desenvolvida pela aluna Aline Cafruni Gularte, orientanda de graduação.

Por meio deste Termo de Consentimento Livre e Esclarecido, o aluno irá declarar que recebeu explicações sobre os objetivos do presente estudo, afirmando que foi esclarecido sobre a garantia de privacidade e do anonimato das suas informações, e que os dados utilizados servirão apenas para estudo e divulgação com fins científicos.

Porto Alegre, 22 de agosto de 2018.

  
Lisiane Priscila Roldão Selau

  
Aline Cafruni Gularte