

UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE
PRODUÇÃO

Mônica Cristina Bogoni Savian

**ESTUDO DOS FATORES DE RISCO ASSOCIADOS À EVASÃO DE
ALUNOS DE GRADUAÇÃO DA UNIVERSIDADE FEDERAL DE
SANTA MARIA**

Santa Maria, RS
2018

Mônica Cristina Bogoni Savian

**ESTUDO DOS FATORES DE RISCO ASSOCIADOS À EVASÃO DE ALUNOS DE
GRADUAÇÃO DA UNIVERSIDADE FEDERAL DE SANTA MARIA**

Dissertação apresentada ao Curso de Pós-Graduação em Engenharia de Produção, Área de Concentração em Gerência da Produção, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Mestre em Engenharia de Produção.**

Orientadora: Prof^a. Dr^a. Roselaine Ruviaro Zanini

Santa Maria, RS
2018

Savian, Mônica Cristina Bogoni
ESTUDO DOS FATORES DE RISCO ASSOCIADOS À EVASÃO DE
ALUNOS DE GRADUAÇÃO DA UNIVERSIDADE FEDERAL DE SANTA
MARIA / Mônica Cristina Bogoni Savian.- 2018.
90 p.; 30 cm

Orientadora: Roselaine Ruviaro Zanini
Dissertação (mestrado) - Universidade Federal de Santa
Maria, Centro de Tecnologia, Programa de Pós-Graduação em
Engenharia de Produção, RS, 2018

1. Evasão 2. Fatores de risco 3. Regressão Logística
I. Zanini, Roselaine Ruviaro II. Título.

Mônica Cristina Bogoni Savian

**ESTUDO DOS FATORES DE RISCO ASSOCIADOS À EVASÃO DE ALUNOS DE
GRADUAÇÃO DA UNIVERSIDADE FEDERAL DE SANTA MARIA**

Dissertação apresentada ao Curso de Pós-Graduação em Engenharia de Produção, Área de Concentração em Gerência da Produção, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Mestre em Engenharia de Produção**.

Aprovado em 12 de janeiro de 2018:

Roselaine Ruviaro Zanini, Dra. (UFSM)
(Presidente/Orientador)

Luciane Flores Jacobi, Dra. (UFSM)

Lorena Vicini, Dra. (UFFS)

Santa Maria, RS
2018

DEDICATÓRIA

Dedico este trabalho à minha família, principalmente aos meus pais Darci e Lidia que sempre me ensinaram o que é certo e ao meu esposo Evandro que sempre me incentivou a estudar.

AGRADECIMENTOS

Venho agradecer às muitas pessoas que contribuíram para que esse trabalho fosse concretizado:

- Aos meus pais, Darci e Lidia e ao meu esposo, Evandro por sempre me incentivarem a estudar, me apoiando nas minhas escolhas. Por entender a minha ausência em muitos momentos importantes da nossa vida. Muito obrigada, vocês são meu alicerce.

- À minha querida orientadora, Professora Dra. Roselaine Ruviano Zanini, que não mediu esforços durante o tempo percorrido até aqui, dedicando todo seu conhecimento, incentivo e paciência. Pela oportunidade de trabalhar ao seu lado e por ter entendido a distância em que me encontrava e mesmo assim ter confiado no meu trabalho. Pela excelente professora e pessoa que és.

- A todos os professores do Programa de Pós-Graduação em Engenharia da Produção - PPGEP da Universidade Federal de Santa Maria pela contribuição e conhecimento repassados, essenciais para minha formação.

- À Professora Luciane Flores Jacobi por estar comigo desde a graduação me apoiando.

- À banca, pelo aceite do convite e pelas considerações dadas, muito importantes para a conclusão deste trabalho.

- Ao Fernando Bordin da Rocha, Diretor do Centro de Processamento de Dados (CPD) da UFSM, pela disponibilidade dos dados necessários à realização dessa pesquisa.

- À Márcia e Fernando, funcionários da secretaria do PPGEP pela atenção, comprometimento e ajuda.

- À Universidade Federal de Santa Maria, pela oportunidade da realização do sonho de ser Mestre.

RESUMO

ESTUDO DOS FATORES DE RISCO ASSOCIADOS À EVASÃO DE ALUNOS DE GRADUAÇÃO DA UNIVERSIDADE FEDERAL DE SANTA MARIA

AUTORA: Mônica Cristina Bogoni Savian
ORIENTADORA: Roselaine Ruviano Zanini

Atualmente, as Instituições de Ensino Superior (IES), principalmente as públicas, manifestam grande preocupação com relação à adequada qualificação de seus estudantes e a garantia de resultados adequados em termos de números de diplomados que são liberados todo ano para o exercício da profissão. Desse modo, o presente estudo tem por objetivo identificar e estimar os fatores de risco associados à evasão de alunos de graduação da Universidade Federal de Santa Maria (UFSM) no período entre 2009 e 2015 por meio de modelos de regressão logística. O presente estudo foi de caráter quantitativo, descritivo, retrospectivo e aplicado. O CCNE (52,0%) e o CCS (11,6%) foram os centros que apresentaram o maior e o menor percentual de evasão, respectivamente. Sobre o perfil do discente, observou-se que, se trata de alunos com idade média entre 20 e 28 anos, o gênero varia de acordo com o centro de ensino analisado, de etnia branca, solteiros, com ingresso na universidade no primeiro semestre do ano, por ampla concorrência e em média 65,0% dos alunos não moram na cidade em que a universidade está localizada. Por meio dos modelos de regressão logística ajustados aos dados, observou-se, na maior parte dos casos que, quanto maior a idade do discente, maior o risco de evasão. Com relação às cotas, observou-se que, em geral, ingressar na universidade por cota de escola pública é fator de proteção em relação à ampla concorrência, enquanto que ingressar por outros tipos de cota, representa risco. Os primeiros semestres foram os que apresentaram maior risco de evasão em todos os cursos em que essa variável apresentou significância estatística no modelo, quando comparados com os últimos semestres da graduação. Quanto ao número de disciplinas aprovadas, o aluno ser bolsista ou participar de projetos observou-se que as variáveis se apresentaram como fator de proteção. A cada disciplina reprovada por frequência ou trancamento total realizado, o risco de evasão aumenta. Por meio dos modelos ajustados foi possível verificar quais os principais fatores associados à evasão e também as variáveis que mais contribuem e que apresentam maior risco para tal evento e assim, permitir aos gestores conhecimento para implementar ações que possam minimizar os índices de evasão na UFSM.

Palavras-chave: Evasão. Fatores de risco. Regressão Logística.

ABSTRACT

STUDY OF RISK FACTORS ASSOCIATED WITH EVASION OF UNDERGRADUATE STUDENTS OF THE FEDERAL UNIVERSITY OF SANTA MARIA

AUTHOR: Mônica Cristina Bogoni Savian

ADVISER: Roselaine Ruviaro Zanini

Currently, Higher Education Institutions (HEIs), especially public ones, express great concern regarding the adequate qualification of their students and assurance of adequate results in terms of numbers of graduates who are authorized every year to exercise their professions. Thus, the present study aims to identify and estimate the risk factors associated with the evasion of undergraduate students from the Federal University of Santa Maria (UFSM) in the period between 2009 and 2015 through logistic regression models. The present study was quantitative, descriptive, retrospective, and applied. The centers that presented the highest and the lowest percentage of evasion were the CCNE (52.0%) and the CCS (11.6%), respectively. About the profile of the student, it was observed that their mean age was between 20 and 28 years-old. The gender varies according to the center of education analyzed. They are from white ethnicity, unmarried, and admitted to the university in the first term by broad competition. On average, 65.0% of students do not live in the city where the university is located. Through logistic regression models adapted to the data, it was observed that in most cases the higher the student's age, the greater the risk of evasion. With regard to quotas, it was observed that admission to university under public school quota is generally a protection factor in relation to the wide competition, while admission under other types of quota represents a risk. The first semesters were the ones that presented the highest risk of evasion in all courses in which this variable presented statistical significance in the model when compared to the last semesters. It was observed that the variables appeared as a protection factor for the student as for the number of modules succeeded, being a scholarship holder or participating in projects. The risk of evasion increases at each module failed by excessive absence or enrollment cancellation. By means of the adjusted models, it was possible to verify the main factors associated with evasion as well as the variables that contribute the most and that present a greater risk for such event. Thus, it improves knowledge for the managers to implement actions that can minimize the evasion rates at UFSM.

Keywords: Evasion. Risk factors. Logistic Regression.

LISTA DE TABELAS

Tabela 1 – Número de IES, por Organização Acadêmica no ano de 2015	17
Tabela 2 – Número de cursos, matrículas, ingressos e concluintes por organização acadêmica no ano de 2015	19
Tabela 3 – Diplomados da UFSM, por Unidade Universitária, no período de 2014 e 2015	21
Tabela 4 – Estudantes matriculados por centro de ensino da graduação, no Campus de UFSM, no ano de 2016	21
Tabela 5 – Vagas e inscritos da Graduação, por Centro de ensino, no Campus da UFSM, no período entre 2014 e 2016	22
Tabela 6 – Associação entre fator de risco e doença em uma população.	32
Tabela 7 – Possíveis valores para algumas medidas de diagnóstico segundo as probabilidades ajustadas.....	48
Tabela 8 – Resultados da análise da evasão e não evasão de discentes de graduação da UFSM, para os centros de ensino CAL, CCSH e CE, no período de 2009 a 2015	56
Tabela 9 – Resultados da análise da evasão e não evasão de discentes de graduação da UFSM, para os centros de ensino CCNE, CT e CCR, no período de 2009 a 2015	58
Tabela 10 – Resultados da análise da evasão e não evasão de discentes de graduação da UFSM, para os centros de ensino CCS e CEFD, no período de 2009 a 2015	60
Tabela 11 – Percentual de evasão nos cursos, por Centro de ensino da UFSM, no período de 2009 a 2015	65
Tabela 12 – Resultado da análise de regressão logística de discentes dos cursos da UFSM, com melhor e pior desempenho em relação à evasão, no período de 2009 a 2015	69
Tabela 13 - Razões de chance (OR) que apresentaram significância no modelo ajustado de regressão logística de cada curso analisado, no período de 2009 a 2015	83

LISTA DE QUADROS

Quadro 1– Valores do modelo de regressão logística quando a variável dependente é dicotômica	38
Quadro 2 – Variáveis qualitativas (categóricas) e respectivas categorias.....	49

LISTA DE ABREVIATURAS E SIGLAS

CAL	Centro de Artes e Letras
CAFW	Colégio Agrícola de Frederico Westphalen
CCS	Centro de Ciências da Saúde
CCNE	Centro de Ciências Naturais e Exatas
CCR	Centro de Ciências Rurais
CCSH	Centro de Ciências Sociais e Humanas
CE	Centro de Educação
CEFD	Centro de Educação Física e Desportos
CEFET	Centro Federal de Educação Tecnológica
CEU	Casa do Estudante Universitário
CESNORS	Centro de Educação Superior Norte – RS/UFSM
CCSD	Clarck Country School District
CT	Centro de Tecnologia
CTISM	Colégio Técnico Industrial
COPERVES	Comissão Permanente do Vestibular
CPD	Centro de Processamento de Dados
EAD	Educação à Distância
HUSM	Hospital Universitário de Santa Maria
IES	Instituição de Ensino Superior
IF	Instituto Federal
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais
LUA	Latvia University of Agriculture
MEC	Ministério da Educação
MLG	Modelos Lineares Generalizados
MMQ	Método dos Mínimos Quadrados
OR	<i>Odds Ratio</i>
OSEP	Departamento Especial de Educação
PEG	Programa Especial de Graduação de Formação de Professores
PROGRAD	Pró-reitoria de Graduação
REUNI	Reestruturação e Expansão das Universidades Federais
RU	Restaurante Universitário
SIE	Sistema de Informação da Universidade Federal de Santa Maria
SISU	Sistema de Seleção Unificada
UFMG	Universidade Federal de Minas Gerais
UFMT	Universidade Federal de Mato Grosso
UFRGS	Universidade Federal do Rio Grande do Sul
UFSM	Universidade Federal de Santa Maria

SUMÁRIO

1 INTRODUÇÃO	12
1.1 DELIMITAÇÃO DO TEMA	13
1.2 PROBLEMÁTICA	14
1.3 OBJETIVOS	15
1.3.1 Objetivo geral	15
1.3.2 Objetivos específicos.....	15
1.4 JUSTIFICATIVA	15
2 EVASÃO NO ENSINO SUPERIOR	17
2.1 CENÁRIO DO ENSINO SUPERIOR NO BRASIL	17
2.2 CONTEXTUALIZAÇÃO DA UFSM.....	19
2.3 A EVASÃO	23
2.3.1 A evasão em universidades internacionais.....	23
2.3.1.1 Cálculo de taxas de evasão.....	26
2.3.2 A evasão no Brasil.....	27
2.3.2.1 Cálculo de taxas de evasão.....	29
3 REGRESSÃO LOGÍSTICA	31
3.1 REGRESSÃO LOGÍSTICA.....	31
3.2 RISCO RELATIVO	32
3.3 REGRESSÃO LOGÍSTICA SIMPLES	34
3.3.1 Estimação de parâmetros do modelo de regressão logística simples.....	36
3.3.2 Interpretação dos coeficientes do modelo de regressão logística	37
3.4 REGRESSÃO LOGÍSTICA MÚLTIPLA	39
3.4.1 Estimação de parâmetros do modelo de regressão logística múltipla.....	41
3.5 SELEÇÃO DE MODELOS.....	42
3.5.1 Método forward	43
3.5.2 Método backward	43
3.5.3 Método stepwise	44
3.5.4 Método de Akaike.....	45
3.6 ANÁLISE DE DIAGNÓSTICO E RESÍDUOS	46
3.6.1 Técnicas de diagnóstico	46
4 MATERIAIS E MÉTODOS	49
4.1 PROCEDIMENTOS METODOLÓGICOS.....	52
5 RESULTADOS E DISCUSSÃO	55
5.1 ANÁLISE DESCRITIVA.....	55
5.2 ANÁLISE DE REGRESSÃO LOGÍSTICA.....	65
5.2.1 Centro de Artes e Letras	66
5.2.2 Centro de Ciências Naturais e Exatas.....	67
5.2.3 Centro de Ciências Rurais.....	68
5.2.4 Centro de Ciências da Saúde.....	76
5.2.5 Centro de Ciências Sociais e Humanas	77
5.2.6 Centro de Educação.....	77
5.2.7 Centro de Educação Física e Desportos.....	78
5.2.8 Centro de Tecnologia	79
5.3 CONSIDERAÇÕES SOBRE OS RESULTADOS ANÁLISE DOS MODELOS DE REGRESSÃO LOGÍSTICA MÚLTIPLA.....	79
6 CONCLUSÃO	84
REFERÊNCIAS	86

1 INTRODUÇÃO

A evasão é considerada um problema universal que afeta o resultado dos sistemas educacionais. Estudantes que iniciam e não finalizam sua graduação são considerados perdas sociais, acadêmicas e econômicas. Tanto no setor público como no privado, a evasão é uma fonte de ociosidade de professores, funcionários, equipamentos e espaço físico, no setor público pelos recursos serem investidos sem o devido retorno e no privado por ser uma importante perda de receitas (SILVA FILHO et al., 2007).

No Brasil, as preocupações com a evasão no ensino superior iniciaram, efetivamente, em 1995 quando foi instituída a “Comissão Especial de Estudos sobre a Evasão nas Universidades Públicas Brasileiras”, por meio do Seminário sobre evasão nas universidades brasileiras”, organizado pela Secretaria de Educação Superior do Ministério da Educação (SESU/MEC), em fevereiro daquele ano, com o objetivo de identificar as causas e sugerir medidas para minimizar os índices de evasão nas instituições públicas de ensino superior do país.

De acordo com o Decreto nº 6.096, de 24 de abril de 2007 que institui o programa de apoio a planos de Reestruturação e Expansão das Universidades Federais (REUNI), atualmente, a evasão tem recebido maior destaque nas políticas federais brasileiras para a educação superior. A evasão que equivale à desistência do curso pelo aluno e a retenção que equivale à permanência prolongada no curso são os aspectos mais destacados dentro das instituições de ensino no conceito de fluxo escolar.

Diversos fatores relacionados ao fluxo escolar de graduação são enumerados como o desempenho do aluno no vestibular, sucesso acadêmico, identificação com a escolha do curso, adaptação em relação ao trabalho e estudo, entre outros. Tinto (1975), propôs o entendimento do processo da evasão em que se observa que estudantes pouco envolvidos com a instituição têm mais chances de abandonar o curso, considerando-se que os mesmos não se integram com os demais alunos da sua turma ou com a instituição como um todo.

A evasão deve ser entendida sob dois aspectos semelhantes, a evasão anual média que mede, em uma Instituição de Ensino Superior (IES), o percentual de alunos matriculados em um sistema de ensino, que não tendo se formado também não se matriculou no ano seguinte e a evasão total que mede o número de alunos

que entraram em uma determinada IES, mas não obtiveram o diploma ao final de certo período de tempo (SILVA FILHO et al., 2007). Os dois conceitos estão ligados, porém, os níveis de reprovação e as taxas de evasão por ano ao longo do curso devem ser considerados.

Segundo Silva Filho et al. (2007), tanto as instituições privadas quanto as públicas dão como principal razão da evasão a falta de recursos financeiros para o estudante prosseguir com seus estudos que também é o que o estudante declara quando perguntado sobre a principal razão da evasão. No entanto, verifica-se também que essa resposta é uma simplificação, uma vez que os principais fatores que desestimulam o estudante a priorizar o investimento de tempo ou financeiro para conclusão do curso são questões de ordem acadêmica e expectativas do aluno em relação à sua formação e à própria integração com a instituição. Ou seja, o estudante acredita que o custo benefício do “sacrifício” para obter um diploma superior na carreira escolhida não vale mais a pena.

Para Zago (2006), um fator importante a ser considerado é que somente 9% dos jovens brasileiros entre 18 e 24 anos frequenta o ensino superior, um dos índices mais baixos da América Latina. A autora ainda destaca que determinados estudos indicam que 25% dos potenciais alunos são carentes e não têm condições de ingressar no ensino superior, ainda que este seja gratuito. O aumento quantitativo do número de vagas foi considerável nos últimos anos, mas a sua concentração no ensino pago, ou seja, nas instituições particulares, não reduziu as desigualdades entre os grupos sociais nas últimas décadas.

Para Silva Filho e Hipólito (2009), a formação superior, além de preparar profissionais tecnicamente mais habilitados e competitivos no país ainda eleva o nível cultural e político de um povo. No Brasil, tem-se apenas 8% da população adulta com formação superior, enquanto que a Rússia, por exemplo, tem 55%, a Coreia 32%, Espanha 28% e o Chile 13%.

1.1 DELIMITAÇÃO DO TEMA

A pesquisa delimitou-se em analisar os alunos de graduação de cursos presenciais da UFSM, no Campus situado na cidade de Santa Maria/RS, que ingressaram na universidade no período entre 2009 e 2015.

De acordo com o MEC/SESu (1997) por meio da Comissão Especial de Estudos Sobre a Evasão nas Universidades Públicas Brasileiras, criada em março de 1995 com objetivo de estudar com profundidade o tema da evasão, a mesma pode ser classificada em três tipos:

- i. Evasão de curso: ocorre quando o estudante se desliga do curso superior em situações diversas, tais como: abandono (deixa de matricular-se), desistência (oficial), transferência ou reopção (mudança de curso), exclusão por norma institucional;
- ii. Evasão da instituição: configura-se pelo desligamento do estudante da instituição na qual está matriculado;
- iii. Evasão do sistema: caracteriza-se pelo abandono por parte do estudante de forma definitiva ou temporária do ensino superior.

Pode-se delimitar, para o presente estudo, a pesquisa sobre a evasão da instituição, em que o aluno abandona a universidade sem concluir nenhum curso, o que exclui a opção de mobilidade interna já que serão analisados os estudantes relacionados à área escolhida dentro da universidade em questão.

1.2 PROBLEMÁTICA

As maiores preocupações das IES atualmente, especialmente as públicas, devem ser a boa qualificação de seus estudantes e a garantia de resultados satisfatórios em termos de números de diplomados que são liberados todo ano para o exercício da profissão.

A evasão é um fenômeno complexo, comum às instituições universitárias no mundo contemporâneo. Sua complexidade e abrangência vêm sendo objeto de estudos e análises nos últimos anos, principalmente nos países de Primeiro Mundo (COMISSÃO ESPECIAL DE ESTUDOS SOBRE A EVASÃO NAS UNIVERSIDADES PÚBLICAS BRASILEIRAS, 1996).

Com base nas conjecturas realizadas na seção anterior e posteriormente na seção 1.4, pode-se definir que o problema de pesquisa abordará as questões relevantes sobre evasão, tentando identificar as principais causas que levam a este fato.

1.3 OBJETIVOS

1.3.1 Objetivo geral

Identificar e estimar os fatores de risco que estão associados à evasão de alunos de graduação da Universidade Federal de Santa Maria - UFSM, no período de 2009 a 2015.

1.3.2 Objetivos específicos

- Realizar estatística descritiva das variáveis utilizadas no estudo;
- Comparar, por Centro de ensino, as variáveis investigadas para a evasão;
- Identificar os cursos com maior e menor percentual de evasão;
- Analisar as características dos cursos de acordo com o percentual de evasão;
- Ajustar um modelo de regressão logística para identificar os fatores de risco associados à evasão;
- Estimar as probabilidades de casos que possuem situações especiais relativas à evasão;

1.4 JUSTIFICATIVA

O principal problema na educação superior atual é o desequilíbrio entre as entradas e saídas na universidade, em particular o desequilíbrio entre os estudantes matriculados e graduados. As taxas de abandono dos alunos universitários resultam no desperdício de dinheiro dos contribuintes, menor proporção de alunos graduados e, por conseguinte, menores oportunidades de emprego em posições altamente qualificadas (PAURA e ARHIPOVA, 2014).

Em 2006, um estudo realizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais (INEP) apontou que a evasão ocorre tanto entre as IES públicas como privadas. Enquanto as IES privadas apresentam índices aproximados de 53%, nas instituições públicas esse índice é de 33%.

Sendo assim, é de extrema importância para a instituição de ensino superior, a identificação do perfil de seus alunos regulares, evadidos ou concluintes, a fim de

que, políticas para controlar ou minimizar a evasão sejam implementadas na instituição.

2 EVASÃO NO ENSINO SUPERIOR

Esta revisão de literatura está desenvolvida em itens que envolvem o tema evasão no ensino superior. Primeiramente, realizou-se a pesquisa do atual cenário encontrado no ensino superior brasileiro, em seguida a visão de diferentes autores acerca da evasão em universidades internacionais e no Brasil, o cálculo de taxas e a contextualização da UFSM e por fim, o local onde será desenvolvida a pesquisa.

2.1 CENÁRIO DO ENSINO SUPERIOR NO BRASIL

No Brasil, o Instituto Nacional de Estudos Educacionais Anísio Teixeira (Inep), órgão do Ministério da Educação, tem publicado regularmente dados de matriculados, ingressantes e egressos no ensino superior bem como de instituições, número de cursos, matrículas, vagas oferecidas entre outras informações. Essas sinopses auxiliam a análise da evolução de indicadores ao longo de um período significativo de tempo o que colabora para uma formatação padronizada e a análise da evasão no ensino superior.

Tabela 1 – Número de IES, por Organização Acadêmica no ano de 2015

Tipo de IES	Instituições				Total
	Universidades	Centros Universitários	Faculdades	IF e CEFET	
Pública	107	9	139	40	295
Federal	63	-	4	40	107
Estadual	38	1	81	-	120
Municipal	6	8	54	-	68
Privada	88	140	1.841	-	2.069
Total	195	149	1.980	40	2.364

Fonte: Inep: Sinopse do Ensino Superior (2015).

Segundo dados do Inep (2015), apresentados na Tabela 1, o Brasil tinha em 2015 – data da última sinopse divulgada – um total de 2.364 Instituições de Ensino Superior (IES), 295 públicas e 2.069 privadas, sendo que dessas 195 universidades, 107 são públicas e 88 privadas, 149 Centros Universitários, 9 públicos e 140 privados, 1.980 faculdades, 139 públicas e 1.841 privadas e 40 Institutos Federais (IF) e Centro Federal de Educação Tecnológica (CEFET). Por meio desses dados

pode-se perceber que o número de IES privadas é 7 vezes maior que o de IES públicas.

Para melhor entendimento, pode-se discriminar universidade, faculdade e centro universitário por meio de algumas características. A universidade deve proporcionar obrigatoriamente atividades de ensino, pesquisa e extensão e possui total independência, podendo criar cursos sem requerer a autorização do Ministério da Educação (MEC). Os requisitos mínimos para a existência de uma universidade são pelo menos um terço do corpo docente possuir título de mestrado ou doutorado, um terço dos professores possuir contrato em tempo integral e desenvolver ao menos quatro programas de pós-graduação *stricto sensu* – mestrado e doutorado de boa categoria. Os centros universitários são frequentemente menores que as universidades e também necessitam de, no mínimo, um terço do corpo docente com mestrado ou doutorado e um quinto dos professores contratados em regime integral. Já as faculdades, por sua vez, são IES que apresentam um número menor de áreas de conhecimento em seus cursos e necessitam da autorização do MEC para introduzir um novo curso, neste caso, os docentes da instituição devem possuir, no mínimo, pós-graduação *latu sensu* (PORTAL EDUCAÇÃO, 2013).

O número de cursos oferecidos pelas IES, segundo dados do Inep (2015), apresentados na Tabela 2 foi de 33.501, desses 10.769 foram oferecidos por IES públicas, sendo que a maior parte por IES públicas federais, cerca de 6.313 e 22.732 por IES privadas. O número de matrículas foi de 8.027.297, dessas 1.952.145 na rede pública e 6.075.152 na rede privada. O total de ingressantes no ano de 2015 foi de 2.290.222, sendo 429.997 em IES públicas e 1.080.339 em IES privadas. O número de concluintes total foi de 1.150.067, sendo 239.896 na rede pública e 910.171 na rede privada.

Pode-se perceber, por meio do conjunto de dados disponíveis nas Tabelas 1 e 2, que existe, no ensino superior brasileiro um predomínio quantitativo de instituições privadas, o que permite inferir que as informações referentes aos alunos de IES privadas afetam os indicadores globais do ensino superior brasileiro, inclusive a evasão.

Tabela 2 – Número de cursos, matrículas, ingressos e concluintes por organização acadêmica no ano de 2015

Número de Cursos/Matrículas/Ingressos/Concluintes		Universidades	Centros Universitários	Faculdades	IF e CEFET	Total
Número de Cursos		16.057	4.483	11.703	1.258	33.501
Pública		8.761	127	623	1.258	10.769
	Federal	5.037	-	18	1.258	6.313
	Estadual	3.415	10	284	-	3.709
	Municipal	309	117	321	-	747
Privada		7.296	4.356	11.080	-	22.732
Matrículas		4.273.155	1.357.802	2.251.464	144.876	8.027.297
Pública		1.663.222	18.844	125.203	144.876	1.952.145
	Federal	1.068.101	-	1.658	144.876	1.214.635
	Estadual	545.485	1.570	71.578	-	618.633
	Municipal	49.636	17.274	51.967	-	118.877
Privada		2.609.933	1.338.958	2.126.261	-	6.075.152
Ingressos		1.510.336	584.323	773.890	51.673	2.920.222
Pública		429.997	6.008	46.683	51.673	534.361
	Federal	283.917	.	503	51.673	336.093
	Estadual	130.151	581	30.972	.	161.704
	Municipal	15.929	5.427	15.208	.	36.564
Privada		1.080.339	578.315	727.207	.	2.385.861
Concluintes		595.518	209.597	332.366	12.586	1.150.067
Pública		205.366	2.924	19.020	12.586	239.896
	Federal	121.604	-	257	12.586	134.447
	Estadual	76.328	118	10.324	-	86.770
	Municipal	7.434	2.806	8.439	-	18.679
Privada		390.152	206.673	313.346	-	910.171

Fonte: Inep - Sinopse do Ensino Superior (2015).

Evidentemente que as IES privadas mantêm a maioria dos alunos no ensino superior, seu peso é maior e a média nacional de evasão está mais próxima de seus índices.

2.2 CONTEXTUALIZAÇÃO DA UFSM

A Universidade Federal de Santa Maria (UFSM) é uma instituição de ensino superior público, fundada em 14 de dezembro de 1960 pelo professor José Mariano da Rocha Filho, no município de Santa Maria, que construiu credibilidade e tradição durante seu desenvolvimento. Atualmente, possui espalhadas pelo Rio Grande do Sul, dez unidades universitárias e quatro estabelecimentos de educação básica, técnica e tecnológica. Conta com 308 cursos de graduação e mais de 120 convênios internacionais, sendo a primeira universidade brasileira a priorizar a interiorização de ensino público. Segundo dados de dezembro de 2012 do Instituto Nacional de

Estudos e Pesquisas Educacionais (Inep), a UFSM, em pleno desenvolvimento com cursos, programas e projetos de pesquisa científica nas mais diversas áreas do conhecimento humano, ocupa a 18ª posição no ranking das melhores universidades do Brasil, além da assistência estudantil oferecida a fim de atender as necessidades de saúde, transporte e moradia (UFSM, 2016).

O campus da UFSM está localizado na região central do estado do Rio Grande do Sul (RS), na cidade de Santa Maria, no bairro Camobi, a aproximadamente 290 km de Porto Alegre, capital do estado, como pode ser observado na Figura 1. É no campus que se realiza a maior parte das atividades acadêmicas e administrativas. Existem, no centro da cidade, outras unidades acadêmicas e de atendimento à comunidade.

Figura 1 – Localização da UFSM



Fonte: Centro de Tecnologia; Website construído utilizando o Assistente Web – Centro de Processamento de Dados – CPD/2004; UFSM.

A extensão da cidade universitária é de 1.863,57 hectares, a universidade comporta, além do Hospital Universitário de Santa Maria (HUSM), com 335 leitos, duas agências bancárias, um posto dos correios, lancherias, dois restaurantes universitários, biblioteca central, com aproximadamente 92.535 volumes (livros e teses), bibliotecas setoriais, hospital veterinário, farmácia-escola comercial, planetário e museus, além dos centros de ensino, colégio técnico, anfiteatros e departamentos.

Tabela 3 – Diplomados da UFSM, por Unidade Universitária, no período de 2014 e 2015

Centro de Ensino	2014	2015
Centro de Artes e Letras – CAL	281	318
Centro de Ciências da Saúde – CCS	645	699
Centro de Ciências Naturais e Exatas – CCNE	378	451
Centro de Ciências Rurais – CCR	632	746
Centro de Ciências Sociais e Humanas – CCSH	771	1041
Centro de Educação – CE	556	641
Centro de Educação Física e Desportos – CEFD	118	228
Centro de Tecnologia – CT	507	657
Total	3.888	4.781

Fonte: Centro de Processamento de Dados (CPD) - Sistema de Informações Educacionais (SIE), Indicadores UFSM (2016).

Por meio da Tabela 3, verificam-se os alunos diplomados do Campus da UFSM no ano de 2014 e 2015 por Centro de ensino. Percebe-se que houve um crescimento do número de diplomados total em 2015 em relação a 2014 em todos os centros.

Tabela 4 – Estudantes matriculados por centro de ensino da graduação, no Campus de UFSM, no ano de 2016

Centro de Ensino – Graduação	Total
Centro de Artes e Letras – CAL	1.795
Centro de Ciências da Saúde – CCS	2.254
Centro de Ciências Naturais e Exatas – CCNE	1.491
Centro de Ciências Rurais – CCR	1.990
Centro de Ciências Sociais e Humanas – CCSH	3.353
Centro de Educação – CE	1.993
Centro de Educação Física e Desportos – CEFD	630
Centro de Tecnologia – CT	3.106
Total	16.812

Fonte: Centro de Processamento de Dados (CPD) - Sistema de Informações Educacionais (SIE), Indicadores UFSM (2016).

Outra informação importante para o desenvolvimento da presente pesquisa é o número de estudantes matriculados por Centro e nível de ensino na universidade. Procurou-se obter informações mais atualizadas e precisas acerca de números, neste caso utilizaram-se os dados apresentados no ano de 2016, apenas para análise do número de matriculados. Segundo o Centro de Processamento de Dados

(CPD), no ano de 2016 o número total de alunos de graduação matriculados no Campus da UFSM foi de 16.812. O Centro com maior número de estudantes matriculados em 2016 foi o Centro de Ciências Sociais e Humanas (CCSH), com 3.353 matrículas, seguido do Centro de Tecnologia (CT) com 3.106 estudantes matriculados. Esses números podem ser visualizados, por Centro de ensino na Tabela 4.

Na Tabela 5, pode ser verificada uma relação entre vagas e inscritos da graduação, separadamente por Centro de ensino, no período de 2014 a 2016. Percebe-se que o número de inscritos em 2016 foi o menor desde 2014 e que o número de vagas em 2016 diminuiu em relação ao ano de 2014, porém é maior que em 2015.

Os Centros que concentram o maior número de vagas e conseqüentemente de inscritos no período analisado são o CE que em 2016 tinha 3.800 inscritos para 906 vagas, em seguida pode-se destacar o CCSH com 3.405 inscritos para 847 vagas e o CCS com 2.398 inscritos para 492 vagas. O Centro com menor número de vagas e inscritos é o CEFD com 441 inscritos para 146 vagas. Analisando-se a concorrência, por Centro, em 2016, o CCS, CE e CCSH seriam os que apresentam o maior número de inscritos por vaga, 4,87, 4,19 e 4,14, respectivamente.

Tabela 5 – Vagas e inscritos da Graduação, por Centro de ensino, no Campus da UFSM, no período entre 2014 e 2016

Centro	Vagas 2014	Inscritos 2014	Vagas 2015	Inscritos 2015	Vagas 2016	Inscritos 2016
Centro de Artes e Letras – CAL	618	1.750	348	859	492	1.357
Centro de Ciências da Saúde – CCS	488	9.785	488	9.192	492	2.398
Centro de Ciências Naturais e Exatas - CCNE	674	1.443	429	877	469	1.166
Centro de Ciências Rurais – CCR	657	3.348	452	2.788	462	1.550
Centro de Ciências Sociais e Humanas - CCSH	815	5.497	815	5.401	847	3.504
Centro de Educação – CE	867	1.499	522	642	906	3.800
Centro de Educação Física e Desportos - CEFD	147	639	147	804	146	441
Centro de Tecnologia – CT	569	5.326	659	4.848	661	2.232
Total	4.835	29.287	3.860	25.411	4.475	16.448

Fonte: Centro de Processamento de Dados (CPD) - Sistema de Informações Educacionais (SIE), Indicadores UFSM (2016).

Em um levantamento prévio, realizado no Sistema de Informações Educacionais (SIE), no período compreendido entre 2009 e 2015, período o qual o estudo foi desenvolvido, pôde-se verificar que houve um aumento no número de

alunos evadidos até o ano de 2013 e que em 2014 e 2015 esse número diminuiu, porém ainda pode ser considerado expressivo.

2.3 A EVASÃO

De acordo com Sadler, Cohen e Kockesen (1997), dentre um significativo número de pesquisas da literatura sobre a questão de evasão de estudantes de graduação pode-se destacar as pesquisas de Spady (1970), Tinto (1975) e Cabrera, Nora e Casañeda (1992).

A primeira grande pesquisa sobre evasão estudantil foi realizada por Spady (1970), que forneceu o primeiro modelo teórico do processo de evasão no ensino superior. Este modelo propõe que a integração social (manifestada pelo compartilhamento de valores de grupo, desempenho acadêmico, congruência normativa e apoio dos amigos) e compromisso institucional, reduzem a probabilidade de um aluno evadir. O modelo sugere que as características familiares, pessoais e habilidades do aluno também se combinam para influenciar o processo de evasão.

Com base no trabalho de Spady (1970), Tinto (1975) forneceu um modelo teórico que descreveu o processo de integração dos alunos nos sistemas acadêmicos e sociais na instituição. Esse modelo abrange: atributos pré-colegiais; objetivos e compromissos dos alunos antes da entrada na universidade; experiências formais e informais da universidade; integração pessoal/normativa e, metas e compromissos após a entrada na universidade - tudo culminando na decisão de um aluno evadir. Cabrera, Nora e Casañeda (1992) apresentam a questão das finanças. Eles mostram que as mesmas têm um efeito direto sobre a persistência do aluno no curso, por meio de variáveis intervenientes, mais especificamente, com uma "integração acadêmica do estudante e processos de socialização".

2.3.1 A evasão em universidades internacionais

Como mencionado anteriormente, a evasão atinge IES, em âmbito internacional e nacional, inicia-se na educação básica e leva a consequências

negativas até a universidade. Na sequência serão apresentados alguns estudos internacionais em que o objetivo é análise da evasão em universidades.

Os estudos de Tinto (1975) são considerados a principal referência quando o assunto é evasão. O autor apresenta um modelo teórico para explicação da evasão e destaca a importância da integração acadêmica, estabelecida através de compromissos pessoais, sociais e acadêmicos, como elementos instauradores de um forte vínculo do estudante com a instituição. Tais elementos se transformariam em mecanismos capazes de evitar uma decepção com o curso ou com a instituição que acabasse por ocasionar o desligamento do mesmo.

Sadler, Cohen e Kockesen (1997) realizaram um estudo de evasão em uma universidade de Nova York com o objetivo de desenvolver um modelo que pudesse ser utilizado para prever quais estudantes estão em risco de evadir da universidade. Assim, este grupo de estudantes em risco seria então contatado pela instituição para identificação e abordagem das necessidades individuais de cada um.

O objetivo final do estudo dos autores era que a IES fosse capaz de identificar e interagir com estudantes em risco o mais cedo possível, reduzindo assim a probabilidade de abandono da instituição. Foi desenvolvido um modelo de regressão logística para prever a matrícula do segundo semestre, o qual é consistente com o método sugerido por Hosmer e Lemeshow (1989). A regressão logística foi escolhida porque permite a construção de modelos quando a variável dependente é dicotômica, neste caso, evadido ou não evadido. Os autores desenvolveram três modelos de regressão, que segundo eles, podem ser usados em momentos diferentes para identificar alunos em risco de evadir.

Para Paura e Arhipova (2014), a evasão universitária é um tema importante em muitos países, assim como na Letônia, uma vez que não é apenas o desperdício de dinheiro dos contribuintes que está em jogo, mas também é um critério utilizado para avaliar a própria instituição. Infelizmente, a política estratégica das instituições de ensino superior pode aumentar os números, mas não a qualidade dos estudantes de graduação.

A autora desenvolveu um estudo com o objetivo de analisar as causas de evasão de estudantes do primeiro ano do curso de Ciências da Engenharia na Latvia University of Agriculture (LUA) utilizando dados reais da própria instituição. A mesma utilizou o modelo de risco proporcional (modelo de Cox) para avaliar as causas de abandono dos estudantes. Alguns resultados mostraram que a taxa de evasão dos

alunos durante os primeiros 6 meses é de 23,2% e atingiram seu pico em 5 meses. Um percentual relativamente alto de estudantes decidiu não continuar seus estudos e deixou a universidade durante o primeiro ano letivo.

Araque, Roldán e Salguero (2009) desenvolveram modelos personalizados para obter o risco de o aluno evadir e também para analisar o perfil dos alunos que abandonaram os cursos. Neste estudo foram envolvidas três faculdades localizadas em Granada, Sul da Espanha. Em Engenharia de Software foram estudados 10.844 alunos, em Ciências Humanas, 39.241 e em Ciências Econômicas 25.745 alunos.

Os dados, correspondentes ao período de 1992 em diante, foram utilizados para obter um modelo de regressão logística para cada faculdade. Estes modelos e os dados da estrutura mostraram que certas variáveis aparecem repetidamente na explicação da evasão em todas as faculdades. Estas variáveis são, entre outras, a idade de início do curso, os estudos do pai e da mãe, o desempenho acadêmico, o sucesso, a nota média, o formulário de acesso e, em alguns casos, também o número de repetências necessárias para ser aprovado em uma disciplina.

Os alunos com fracas estratégias educativas e sem persistência para atingir seus objetivos na vida têm baixo desempenho acadêmico e baixas taxas de sucesso e isso implica um alto risco de abandono. Os resultados sugerem que cada centro universitário poderia considerar modelos semelhantes para elaborar um plano de ação específico para ajudar a reduzir a taxa de evasão, reduzindo custos e esforços.

A probabilidade de evasão de um estudante durante seu primeiro ano foi o estudo desenvolvido por ARULAMPALAM, NAYLOR e SMITH (2005). O estudo deu-se a partir de dados de nove coortes inteiras de estudantes de graduação em universidades do Reino Unido. Examinaram-se as coortes de estudantes de 1984-1985 a 1992-1993 que se matricularam em período integral em um curso de 3 ou 4 anos.

Os resultados mostraram não só que os alunos com maior dificuldade são mais propensos a evadir, mas também que a variação nas qualificações anteriores ao curso de graduação universitária do aluno também exerce uma influência sobre a probabilidade de evasão do mesmo. Também observaram que a evasão variou conforme as universidades e teve relação com uma medida da qualidade média dos estudantes universitários.

2.3.1.1 Cálculo de taxas de evasão

De acordo com Lehr et al. (2004) a estimação da taxa de evasão é uma tarefa complexa pois sua mensuração muda conforme o conceito aplicado na instituição. O cálculo da evasão escolar pode ser feito de várias maneiras, cada uma delas com seu próprio significado (SILVA FILHO e LOBO, 2012). O propósito da pesquisa precisa estar delimitado da maneira que o pesquisador possa optar pela fórmula mais condizente com seus objetivos de estudo.

Para o Clarck Country School District (CCSD), a inconsistência dos dados é considerada um desafio, pois não há um fornecimento correto dos mesmos e também não há um acompanhamento do aluno ao longo do tempo (NORA LUNA, 2003).

Lehr et al. (2004), destacou três tipos de taxas de evasão em que o Centro Nacional para Estatísticas da Educação e o Departamento Especial de Educação (OSEP), ambos do Departamento de Educação dos Estados Unidos, desenvolveram a fim de mensurar a evasão, a Taxa de Eventos ou de Incidência, Taxa de Estado ou de Prevalência e Taxa de Coorte ou Longitudinal, cada uma com uma definição e uma imagem ligeiramente diferente da magnitude do problema.

- A Taxa de Eventos, também referida como Taxa Anual ou de Incidência mede a proporção de alunos que abandonaram seus estudos em um único ano sem concluir. Segundo Nora Luna (2003), esse cálculo apresenta a proporção de alunos que evadiram de um ano para outro e assim, produz as menores taxas de evasão.
- A Taxa de Estado, também referida como Taxa de Prevalência, mede a proporção de alunos que não concluíram e não estão matriculados, independente de quando houve o abandono. Essa taxa tende a ser maior que a de eventos.
- A Taxa de Coorte, ou Taxa Longitudinal, mede o que ocorre com um determinado grupo, no decorrer de um intervalo de tempo, assim, tem-se estudantes desistentes quatro anos depois, e fornece a maior taxa de evasão.

2.3.2 A evasão no Brasil

De acordo com o mencionado, o universo da evasão no ensino superior é referenciado em estudos e pesquisas nacionais e internacionais. Têm-se, em âmbito nacional, alguns estudos que se sobressaem como, por exemplo, Silva Filho et al. (2007). Os autores estudaram a evasão no Brasil com base em dados oficiais do Instituto Nacional de Estudos e Pesquisas Educacionais (Inep), órgão do Ministério da Educação que tem divulgado publicamente, de forma regular, dados referentes aos matriculados, ingressantes e egressos do ensino superior.

O estudo de Neto, Cruz e Pfitscher (2008), com o objetivo de abordar os problemas e os riscos relacionados à utilização do índice de evasão escolar como meta para avaliação de desempenho institucional no contexto do ensino superior público, e propor formas para percepção da questão conforme a natureza das suas causas.

Polydoro (2000), com o objetivo de analisar o trancamento de matrícula no que se refere às condições envolvidas na saída e no retorno do estudante à instituição, com o propósito de compreender como se dá a inserção desta modalidade de evasão na trajetória acadêmica do universitário.

O estudo de Morosin et al. (2011), apresentando uma revisão de literatura sobre a evasão na educação superior no contexto brasileiro. A pesquisa revelou que, no período compreendido entre 2000 e 2005, no conjunto formado por todas as IES no Brasil, a evasão média foi de 22% e atingiu 12% nas instituições públicas e 26% nas particulares. O Estudo ainda inclui análises regionais dos índices de evasão anual média e da evasão por tipo de instituição. Pôde-se verificar correlação negativa entre os índices de evasão e a demanda por curso. Também foram apresentados dados internacionais a fim de possibilitar comparações e indicar que a evasão no Brasil não difere muito das médias internacionais.

O estudo desenvolvido por Baggi e Lopes (2011), analisa a produção teórica que aborda a evasão e a sua relação com a avaliação institucional a partir da Biblioteca de Teses e Dissertações entre 2008 e 2009.

Para as autoras, com base em outros estudos como o de Gaioso (2005) e de Silva Filho (2007), a evasão no ensino superior é um fenômeno complexo e que não pode ser analisado fora de um contexto histórico mais amplo, pois é reflexo da realidade de níveis anteriores de ensino, influenciando de diversas maneiras para o

abandono de um curso superior. As autoras ainda mostram que, dos 59 resultados encontrados para evasão, apenas 4 dentre os trabalhos estavam voltados para o ensino superior.

Pode-se considerar, analisando essas informações que, embora se observe um crescimento na produção acadêmica sobre a evasão no ensino superior, essa produção ainda é escassa.

Mercuri e Fior (2016), desenvolveram um estudo em uma universidade confessional da região sudoeste do Brasil com o objetivo de compreender o papel preditivo de características sobre o estudante, das interações que os mesmos estabelecem com os pares e do envolvimento acadêmico sobre a evasão. Os resultados obtidos pelas autoras, através de regressão logística univariada, mostraram que, de maneira isolada, idade, interações com os pares que envolvem conteúdos acadêmicos e o envolvimento acadêmico, tanto com as atividades obrigatórias como não obrigatórias tiveram papel preditivo na evasão. Os resultados da regressão múltipla indicam que idade e o envolvimento acadêmico com as atividades obrigatórias foram as variáveis que, em conjunto, melhor exercem papel preditivo para a evasão.

Considerando a técnica de regressão logística utilizada para a análise de dados pode-se citar ainda o trabalho de Abbad, Carvalho e Zerbini (2006). O objetivo das autoras foi identificar as variáveis explicativas da evasão em um curso gratuito, à distância, via internet, oferecido em nível nacional. Os resultados encontrados na pesquisa mostraram que alunos não concluintes são aqueles que tendem a não utilizar os recursos eletrônicos de interação (mural de notícias, chats, troca de mensagens eletrônicas).

Com o propósito de abordar o problema da evasão no ensino superior brasileiro por meio de um diagnóstico da sua ocorrência no ciclo básico da Universidade Federal de Minas Gerais (UFMG), Braga, Peixoto e Boguttchi (2003) desenvolveram um estudo em que o foco principal da análise é a compreensão da complexidade das relações que se desenvolvem no ambiente escolar e que podem vir a ser causadoras da evasão.

2.3.2.1 Cálculo de taxas de evasão

A evasão no Brasil pode ser calculada a partir dos três métodos mencionados anteriormente, destacados por Lehret al. (2004). A fórmula para calcular a taxa de evasão anual, elaborada pela Comissão Especial de Estudos sobre a Evasão nas Universidades Públicas Brasileiras (1995) é a seguinte:

$$\text{Evasão \%} = (N_i - N_d - N_r) / N_i * 100 \quad (1)$$

Onde N_i corresponde ao número de ingressantes no ano base, N_d número de diplomados e N_r o número de retidos.

O método mais utilizado no Brasil e no mundo é o cálculo da taxa anual de eventos por considerar, como mencionado acima, a proporção de alunos que evadiram de um ano para outro de acordo com a fórmula proposta pelo Instituto Lobo, desde 2006 (SILVA FILHO et al., 2007; SILVA FILHO e LOBO, 2012) que relaciona o número de matriculados, ingressantes e concluintes, como pode ser visualizada a seguir.

$$E_n = 1 - [M_n - I_n] / [M_{n-1} - C_{n-1}] \quad (2)$$

Sendo que, E corresponde à evasão, M ao número de matriculados, C ao número de concluintes, I número de ingressantes, n ao ano de estudo e $n-1$ o ano anterior. Segundo Silva Filho et al. (2012), independentemente do método utilizado, o importante é analisar a evasão, possibilitando meios de se criar estratégias e ações políticas para combater o fenômeno.

Uma ampla série de métodos vem sendo utilizados a procura das causas da evasão. Alguns, por meio de análise qualitativa, baseados em entrevistas aplicadas com os alunos como é o caso do estudo de Theóphilo e Moraes (2005). Outros, com dados específicos das instituições que permitem aplicações de testes estatísticos ou uma abordagem com base em censos como é o caso do estudo de Silva Filho et al. (2007). Contudo, segundo Furtado e Alves (2012), a análise de regressão logística vem sendo amplamente utilizada e tem apresentado resultados importantes. Podem-se citar como exemplos de pesquisas que utilizaram essa técnica, o estudo de Deke

(2003), onde a qualidade da educação foi associada à evasão, a pesquisa de Vitelli (2012), em que abordou a explicação de fatores que interferiam na probabilidade de evasão dos alunos, entre outros.

A técnica de análise de regressão logística será utilizada no presente estudo, a fim de identificar e estimar os fatores de risco associados à evasão de alunos de graduação.

3 REGRESSÃO LOGÍSTICA

Esta revisão de literatura está desenvolvida em itens que envolvem o tema modelos de regressão logística, ferramenta utilizada para a busca dos resultados propostos.

3.1 REGRESSÃO LOGÍSTICA

O objetivo dos métodos de regressão é descrever as relações entre a variável dependente ou resposta (Y) e a variável independente ou explicativa (X). Na regressão logística, a variável resposta é dicotômica ou categórica, neste caso, para o acontecimento de interesse “sucesso” atribui-se o valor 1 e para o evento complementar, “fracasso” o valor atribuído é 0.

A regressão logística tem se constituído em um dos principais métodos de análise estatística de dados. Alguns pesquisadores têm dicotomizado a variável resposta, mesmo quando a variável de interesse não é do tipo binário, de modo que a probabilidade de sucesso possa ser ajustada por meio da regressão logística. O principal motivo é a facilidade de interpretação dos parâmetros de um modelo logístico e também a possibilidade do uso desse tipo de metodologia na análise discriminante (PAULA, 2010).

Em modelos de regressão linear simples ou múltipla a variável dependente é de natureza contínua. Contudo, em algumas situações em que a variável dependente é qualitativa, expressa por duas ou mais categorias, ou seja, admite dois ou mais valores o método dos mínimos quadrados não oferece estimadores plausíveis. Nesse caso, a regressão logística permite o uso de um modelo de regressão para calcular ou prever a probabilidade de um evento específico.

As principais referências ao modelo de regressão logística podem ser encontradas em textos como de Cox (1972), Cox e Snell (1989), Breslowe Day (1980) com aplicação em pesquisa de câncer, estudo e análise de caso-controle, Agresti (1996) e Agresti (2002). Hosmer e Lemeshow (1989), em seu livro, dedicam exclusivamente o texto à análise de regressão logística. A aplicação do modelo à área de epidemiologia pode ser encontrada no texto de Kleinbaun, Kupper e Morgenstern (1982). Paula (2010) apresenta a análise introduzindo o modelo de

regressão logística para a resposta binária fazendo uma analogia com os métodos tradicionais de tabelas 2 X 2. Também discute a seleção de modelos logísticos, métodos de diagnóstico, alguns tipos de modelos de dose-resposta, sobredispersão e regressão logística condicional.

3.2 RISCO RELATIVO

Métodos clássicos de tabelas de contingência 2 X 2 são datados da década de 50 e foram motivados pelo interesse em inferência de alguns parâmetros com grande aplicabilidade na área biomédica, mais especificamente em Epidemiologia. Durante as décadas de 50 e 60 vários trabalhos foram publicados e as técnicas desenvolvidas são utilizadas até hoje especialmente na análise descritiva dos dados antes de um tratamento mais sofisticado por meio da análise de regressão (PAULA, 2010).

Para estimar a grandeza do efeito por meio de observações de uma única amostra podem-se considerar vários métodos disponíveis. Um exemplo é a *odds ratio* (OR) ou *razão de chances* que apresenta a informação sobre duas variáveis dicotômicas, medidas de amostras não emparelhadas (Pagano e Gauvreau, 2004).

Para definição de risco relativo, considera-se que os indivíduos de uma determinada população sejam classificados segundo um fator com dois níveis *A* e *B*, e por *D* e \bar{D} , respectivamente, a presença ou ausência de uma doença. Neste caso, pode-se perceber na Tabela 6, a descrição das proporções populacionais (PAULA, 2010):

Tabela 6 – Associação entre fator de risco e doença em uma população.

Doença	A	B	
D	P_1	P_3	$P_1 + P_3$
\bar{D}	P_2	P_4	$P_2 + P_4$
	$P_1 + P_2$	$P_3 + P_4$	$n = P_1 + P_2 + P_3 + P_4$

Fonte: Adaptado de Paula, 2010.

As quantidades podem ser definidas da seguinte forma:

$P_1/(P_1 + P_2)$ = proporção de indivíduos classificados como doentes no grupo

A;

$P_3/(P_3 + P_4)$ = proporção de indivíduos classificados como doentes no grupo B;

Foi denominado por Cornfield (1956) como sendo risco relativo entre os níveis A e B a razão entre as duas proporções expostas acima, neste caso tem-se:

$$RR = \frac{P_1/(P_1+P_2)}{P_3/(P_3+P_4)} = \frac{P_1(P_3+P_4)}{P_3(P_1+P_2)} \quad (3)$$

Se a doença for rara ($P_1 < P_2$ e $P_3 < P_4$), Cornfield (1956) observou que a proporção exposta acima assume a seguinte forma simplificada, a qual denominou de *Odds Ratio*:

$$OR = \frac{P_1P_4}{P_3P_2} \quad (4)$$

A interpretação da OR depende da sua magnitude e significância:

- Se $OR = 1$, o fator não é considerado de risco;
- Se $OR > 1$, o fator é considerado de risco;
- Se $OR < 1$, o fator é considerado de proteção.

Para o cálculo do intervalo de confiança para o estimador OR deve-se considerar a equação (5) que expressa um intervalo de confiança de 95% para uma média μ .

$$\left(\bar{X} - 1,96 \frac{\sigma}{\sqrt{n}}; \bar{X} + 1,96 \frac{\sigma}{\sqrt{n}} \right) \quad (5)$$

Em que \bar{X} é a média amostral, σ é o desvio padrão e n é o tamanho da amostra.

Porém, deve-se considerar também que a distribuição de probabilidade da razão de chances é simétrica à direita, assim a mesma pode assumir qualquer valor positivo entre 0 e infinito embora não possa tomar valores negativos. Neste caso, trabalha-se com uma escala logarítmica para o cálculo do intervalo de confiança de 95% para o logaritmo natural da razão de chances como pode ser observado na equação a seguir:

$$\left(\ln(\text{OR}) - Z_{\sigma/2} \widehat{\text{ep}}[\ln \text{OR}]; \ln(\text{OR}) + Z_{\sigma/2} \widehat{\text{ep}}[\ln \text{OR}] \right) \quad (6)$$

Sendo que a estimação do erro padrão do OR é dada por:

$$\widehat{\text{ep}} [\ln(\text{OR})] = \sqrt{\frac{1}{P_1} + \frac{1}{P_2} + \frac{1}{P_3} + \frac{1}{P_4}} \quad (7)$$

3.3 REGRESSÃO LOGÍSTICA SIMPLES

Para iniciar a discussão sobre regressão logística é interessante fazer uma breve explanação sobre os Modelos Lineares Generalizados (MLG).

Segundo McCullagh e Nelder (1989) a especificação de um MLG é dada por uma componente aleatória que consiste nas observações da variável aleatória Y , ou seja, com o vetor $y = (y_1, \dots, y_n)$, a qual identifica a distribuição de probabilidade da variável dependente. Uma componente sistemática, definida por meio de um vetor $\eta = (\eta_1, \dots, \eta_n)$, está associada ao conjunto de variáveis independentes em um modelo linear $\eta = \mathbf{X}\beta$, onde X é uma matriz que consiste nas variáveis independentes de n observações e β é um vetor de parâmetros do modelo, que especifica uma função linear entre as variáveis independentes. E uma função de ligação que descreve a relação matemática entre a componente sistemática e o valor esperado da componente aleatória, neste caso, seja $\mu_i = E(Y_i|x_i)$, com $i \in \{1, \dots, n\}$ então η_i é definida por $\eta_i = g(\mu_i)$ onde g é uma função monótona e diferenciável.

A função de ligação conecta os valores esperados das observações às variáveis explanatórias, para $i \in \{1, \dots, n\}$, segundo a fórmula a seguir:

$$g(\mu_i) = \sum_{j=1}^n \beta_j x_{ij} \quad (8)$$

Se a função g , dada acima, for a função identidade tem-se o modelo de regressão linear.

Podem ser consideradas, nessa situação, duas importantes classes de MLG, sendo uma delas constituída pelos modelos *logit*, em que a variável dependente

pode ser associada a uma variável aleatória Bernoulli, onde se enquadra a regressão logística e a outra classe composta pelos modelos *loglinear* na qual a variável dependente é associada a uma distribuição de Poisson.

A regressão logística binária é um caso particular dos MLG, mais especificamente dos modelos *logit*, neste caso, tem-se que o modelo de regressão logística simples envolve uma variável resposta binária, denotada por Y e uma variável independente denotada por x .

Considera-se $P(Y = 1|X = x_i) = \pi_i$ a probabilidade de “sucesso”, ou seja, o valor esperado de Y dado o valor de X de uma variável explicativa e $P(Y = 0|X = x_i) = 1 - \pi_i$ a probabilidade de “fracasso”. Para descrever a média condicional de Y dado X com a distribuição logística, utiliza-se a notação π_i (HOSMER e LEMESHOW, 1989).

Define-se a probabilidade de sucesso do modelo logístico simples como

$$\pi_i = \pi(x_i) = P(Y = 1|X = x_i) = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} \quad (9)$$

e a probabilidade de fracasso da forma

$$1 - \pi_i = 1 - \pi(x_i) = P(Y = 0|X = x_i) = \frac{1}{1 + \exp(\beta_0 + \beta_1 x_i)}, \quad (10)$$

onde β_0 e β_1 são parâmetros desconhecidos.

Observa-se uma diferença importante no que diz respeito à natureza da relação entre a variável resposta e as variáveis independentes entre o modelo de regressão linear e o modelo de regressão logística. Em qualquer problema de regressão a quantidade a ser modelada é o valor médio da variável resposta dado os valores das variáveis independentes. A quantidade denotada por $E(Y|X = x_i)$ é chamada de média condicional em que Y é a variável resposta e x_i os valores das variáveis independentes. Tem-se na regressão linear que $-\infty < E(Y|X = x_i) < +\infty$, porém, devido a natureza da variável resposta, na regressão logística, a variação é $0 < E(Y|X = x_i) < 1$.

Na regressão linear simples, a representação é dada por $E(Y|X = x_i) = \beta_0 + \beta_1 x_i$, na regressão logística, usando a definição de variáveis aleatórias, tem-se

$$E(Y|X = x_i) = 1P(Y_i = 1|X = x_i) + 0P(Y_i = 0|X = x_i) = \pi_i \quad (11)$$

Segundo Hosmer e Lemeshow (1989), no modelo de regressão linear, a variável resposta pode ser expressa por $Y_i = E(Y|X = x_i) + \varepsilon_i$, em que ε_i é o erro com distribuição normal, média zero e variância constante, porém, o mesmo não ocorre quando a variável resposta é dicotômica. O valor da variável resposta, dado x_i , é expresso por $Y_i = \pi(x_i) + \varepsilon_i$, como a quantidade ε_i , que pode assumir somente os valores $\varepsilon_i = 1 - \pi(x_i)$ para $Y_i = 1$ ou $\varepsilon_i = -\pi(x_i)$ para $Y_i = 0$ e, portanto ε_i tem distribuição com média zero e variância dada por $\pi(x_i)(1 - \pi(x_i))$.

Apresenta-se, por meio da expressão (12) a transformação logística de $\pi(x_i)$ sob a forma linear.

$$g(x_i) = \ln \left\{ \frac{\pi(x_i)}{1 - \pi(x_i)} \right\} \quad (12)$$

Em que $g(x) = \beta_0 + \beta_1 x$, ou seja, a expressão 13 é equivalente à expressão 12.

$$\pi(x_i) = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \quad (13)$$

As características do modelo de regressão logística são dadas por:

- i) A variável aleatória dependente se apresenta em escala nominal;
- ii) Para qualquer valor de x , $\pi(x_i)$ fica limitado ao intervalo (0,1);
- iii) Para $0,2 \leq \pi(x_i) \leq 0,8$, é essencialmente linear em x ;
- iv) Os coeficientes de regressão são facilmente interpretáveis, em termos da razão de chances e do tipo de amostragem dos dados.

3.3.1 Estimação de parâmetros do modelo de regressão logística simples

O método dos mínimos quadrados (MMQ) é o mais utilizado na regressão linear para a estimação de parâmetros β , conforme Hosmer e Lemeshow (1989). Quando o MMQ é usado em modelos com resultado dicotômico, os estimadores não apresentam propriedades estatísticas desejáveis para o ajuste do modelo, nesse

caso, utiliza-se o método de máxima verossimilhança. A função de verossimilhança expressa a probabilidade dos dados observados como uma função de parâmetros desconhecidos, sendo as estimativas resultantes aquelas que mais se aproximam do resultado.

Seja $\mathcal{B} = (\beta_0, \beta_1)$ o vetor de parâmetros relacionado com a probabilidade condicional $\mathbb{P} = (Y_i = 1 | x_i) = \pi(x_i)$. Nesse caso, pelo método da máxima verossimilhança, o estimador de \mathcal{B} , denotado por $\hat{\mathcal{B}}$ é a solução das equações de verossimilhança como segue

$$\sum_{i=1}^n [y_i - \pi(x_i)] = 0 \quad (14)$$

$$\sum_{i=1}^n x_i [y_i - \pi(x_i)] = 0 \quad (15)$$

Assume-se que as observações são independentes, nesse caso, a função de verossimilhança é dada por

$$L(\mathcal{B}) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (16)$$

Pelo princípio da máxima verossimilhança, o estimador $\hat{\mathcal{B}}$ é o valor que maximiza a expressão 16. Para encontrar o valor de $\hat{\mathcal{B}}$, diferencia-se a equação (16) relativamente à β_0 e β_1 e igualam-se as expressões resultantes a zero.

Aplicando-se o logaritmo em ambos os lados da equação (16) tem-se a expressão

$$\mathcal{L}(\mathcal{B}) = \sum_{i=1}^n [y_i \ln \pi(x_i) + (1 - y_i) \log(1 - \pi(x_i))] \quad (17)$$

3.3.2 Interpretação dos coeficientes do modelo de regressão logística

Para a interpretação dos coeficientes do modelo de regressão logística simples, consideram-se duas situações, a primeira em que a variável independente

também é do tipo categórica e a outra que a variável independente é do tipo numérica (quantitativa). Na primeira situação, a variável x será codificada como 0 ou 1. Quando $x = 1$, a chance de resposta é definida por $\pi(1)/[1 - \pi(1)]$ e, quando $x = 0$, tem-se que $\pi(0)/[1 - \pi(0)]$. O logaritmo da função, neste caso é dado por

$$g(1) = \log \pi(1)/[1 - \pi(1)] \quad \text{e} \quad g(0) = \log \pi(0)/[1 - \pi(0)] \quad (18)$$

Souza (2006) apresenta os valores do modelo de regressão logística quando a variável dependente é dicotômica, neste caso, esses valores, apresentados pela autora podem ser visualizados no Quadro 1:

Quadro 1– Valores do modelo de regressão logística quando a variável dependente é dicotômica

Variável Resposta Y	Variável Independente X	
	x = 1	x = 0
y = 1	$\pi(1) = \frac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)}$	$\pi(0) = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)}$
y = 0	$1 - \pi(1) = \frac{1}{1 + \exp(\beta_0 + \beta_1)}$	$1 - \pi(0) = \frac{1}{1 + \exp(\beta_0)}$

Fonte: Souza, 2006.

A razão de chances (*Odds Ratio*), denotada por ψ , é dada por

$$\psi = \frac{\pi(1)/[1 - \pi(1)]}{\pi(0)/[1 - \pi(0)]} \quad (19)$$

O logaritmo da razão de chances (*log-odds*) é

$$\log(\psi) = \ln \left[\frac{\pi(1)/[1 - \pi(1)]}{\pi(0)/[1 - \pi(0)]} \right] = g(1) - g(0) \quad (20)$$

Substituindo na expressão 20 as equações para o modelo de regressão logística apresentada no Quadro 1, a razão de chances é definida por:

$$\psi = \frac{\frac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)} / \frac{1}{1 + \exp(\beta_0 + \beta_1)}}{\left[\frac{\exp(\beta_0)}{1 + \exp(\beta_0)} \right] / \left[\frac{1}{1 + \exp(\beta_0)} \right]} = \frac{\exp(\beta_0 + \beta_1)}{\exp(\beta_0)} = \exp(\beta_1) \quad (21)$$

Desse modo, o logaritmo da razão de chances é dado por

$$\log(\psi) = \ln[\exp(\beta_1)] = \beta_1 \quad (22)$$

A razão de chances é definida como a chance de ocorrência de um evento entre indivíduos com o mesmo fator de risco, comparados com indivíduos não expostos, sujeitos ao evento. A mesma é considerada um parâmetro de grande interesse na regressão logística pelo motivo de fácil interpretação.

O risco relativo fornece o risco de desenvolvimento de uma determinada condição por um grupo quando comparado a outro em estudos prospectivos. Pode-se representar o risco relativo a partir da seguinte expressão

$$RR = \frac{\frac{\pi(1)}{\pi(1)[1 - \pi(1)]}}{\frac{\pi(0)}{\pi(0)[1 - \pi(0)]}} = \frac{\pi(1)}{\pi(0)} \quad (23)$$

Considera-se também a situação onde a covariável é contínua. Neste caso o coeficiente β_1 pode ser interpretado como a variação ocorrida no logaritmo da razão de chances quando há um acréscimo de “1” unidade em x , isto é, $\beta_1 = h(x + 1) - h(x)$, para todo valor de x , onde $h(x)$ é a função definida em 12.

3.4 REGRESSÃO LOGÍSTICA MÚLTIPLA

De acordo com Hosmer e Lemeshow (1989) o objetivo da técnica de regressão logística binária múltipla é a identificação de quais variáveis independentes influenciam no resultado e usá-las numa equação para prever a

probabilidade de as variáveis independentes explicarem o desfecho. As variáveis independentes podem ser categóricas e/ou contínuas, sendo que a variável dependente (Y) é normalmente binária e codificada como 0 (ausência) ou 1 (presença) do desfecho considerado.

Quando o interesse é controlar a associação entre dois fatores binários por meio de um terceiro fator, comumente chamado de fator de confundimento é necessária a estratificação com a eliminação ou redução da influência desses fatores na associação de interesse. Uma maneira eficiente de controle desses fatores de confundimento é por meio do ajuste de um modelo de regressão logística (PAULA, 2010).

Seja um conjunto de p variáveis independentes denotadas por $x_i^T = (x_{i0}, x_{i1}, \dots, x_{ip})$, o vetor da i -ésima linha da matriz (\mathbf{X}) das variáveis explicativas, em que cada elemento da matriz corresponde ao ij -ésimo componente (x_{ij}), em que $i = 1, 2, \dots, n$ e $j = 0, 1, \dots, p$, com $x_{i0} = 1$. O vetor de parâmetros desconhecidos denotado por $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$ e β_j é o j -ésimo parâmetro associado a variável explicativa x_j . A probabilidade de sucesso, no modelo de regressão logística múltipla, é dada por

$$\begin{aligned} \pi(x_i) = P(Y_i = 1 | \mathbf{X} = x_i) &= \frac{\exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})} = & (24) \\ &= \frac{\exp(x_i^T \beta)}{1 + \exp(x_i^T \beta)} \end{aligned}$$

e a probabilidade de fracasso como segue

$$\begin{aligned} 1 - \pi(x_i) = P(Y_i = 0 | \mathbf{X} = x_i) &= \frac{1}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})} = & (25) \\ &= \frac{1}{1 + \exp(x_i^T \beta)} \end{aligned}$$

A função de ligação usada para adequar a resposta média ao modelo linear, da mesma forma que no modelo logístico simples, é a função de ligação *logit*. A

equação (26) representa o modelo de regressão logística múltipla, segundo Paula (2010).

$$g(x_i) = \ln \left\{ \frac{\pi(x_i)}{1 - \pi(x_i)} \right\} = x_i^T \beta = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} \quad (26)$$

3.4.1 Estimação de parâmetros do modelo de regressão logística múltipla

A estimação dos parâmetros do modelo de regressão logística pode ser realizada por meio do método de mínimos quadrados ponderados (não iterativo), introduzido por Grizzle, Starmer e Koch (1969), e pode se visto em Paula (2010) e o método da análise discriminante devido a Cornfield (1962). Uma comparação entre o método da análise discriminante e o método da máxima verossimilhança é descrita por Efron (1975), Press and Wilson (1978) e Hosmer e Fisher (1983).

Assim como no modelo de regressão logística simples, a estimação dos parâmetros apresentada no presente estudo é pelo método da máxima verossimilhança.

Utiliza-se o processo iterativo de Newton-Raphson para encontrar o valor de β que maximiza o logaritmo da função de verossimilhança $l(\beta)$, neste caso implica a derivação de $l(\beta)$ em relação a cada parâmetro.

O logaritmo da função de verossimilhança $l(\beta)$ pode ser escrito de acordo à equação

$$l(\beta) = \sum_{i=1}^n [y_i x_i^T \beta - \ln(1 + \exp\{x_i^T \beta\})] \quad (27)$$

Derivando de $l(\beta)$ em relação a cada parâmetro tem-se

$$\frac{\partial l(\beta)}{\partial \beta_j} = \sum_{i=1}^n \left[y_i x_{ij} - \frac{\exp(x_i^T \beta)}{1 + \exp(x_i^T \beta)} x_{ij} \right] = \sum_{i=1}^n [y_i - \pi(x_i)] x_{ij} \quad (28)$$

Assim, o vetor escore $U(\beta)$ pode ser escrito como

$$U(\beta) = X^T y - X^T \pi = X^T (y - \pi) \quad (29)$$

A matriz de Informação de Fischer é dada por

$$I(\beta) = E \left[\frac{\partial^2 l(\beta)}{\partial \beta \partial \beta^T} \right] = X^T Q X \quad (30)$$

onde $Q = \text{diag}[\pi(x_i)(1 - \pi(x_i))]$ e X a matriz de dados, a inversa de $I(\beta)$ é a matriz de covariância assintótica dos estimadores de máxima verossimilhança dos parâmetros de regressão logística, cuja estimativa é dada por $[I(\beta)]^{-1}$.

O método de Newton Raphson é utilizado como método iterativo para solução das equações de máxima verossimilhança. Neste caso, têm-se as equações

$$\begin{aligned} \beta^{(t+1)} &= \beta^t + [I(\beta^{(t)})]^{-1} U(\beta^{(t)}); t = 0, 1, 2, \dots \quad (31) \\ &= \beta^{(t)} + [X^T Q^{(t)} X]^{-1} X^T (y - \pi^{(t)}) \end{aligned}$$

sendo que $\beta^{(t)}$ e $\beta^{(t+1)}$ são, respectivamente vetores de parâmetros estimados nos passos t e $t+1$.

3.5 SELEÇÃO DE MODELOS

Segundo Paula (2010), uma vez definido o conjunto de covariáveis a ser incluído num modelo de regressão logística, resta saber qual a melhor maneira de encontrar um modelo reduzido que inclua apenas as covariáveis e interações mais importantes para explicar a probabilidade de sucesso $\pi(x_i)$. Como a interpretação de parâmetros no modelo de regressão logística é crucial, ou seja, uma forma mecânica de seleção pode levar a um modelo sem sentido e de difícil interpretação. Particularmente, a inclusão de certas interações impõe a permanência no modelo de seus respectivos efeitos principais de ordem inferior, na ótica do princípio hierárquico. Muitas vezes, variáveis consideradas biologicamente importantes não devem ser deixadas de lado pela sua falta de significância estatística. Assim, a

seleção de um modelo logístico deve ser um processo conjugado de seleção estatística de modelos e bom senso.

Os métodos de seleção de modelos de regressão linear podem ser estendidos diretamente para os MLGs, porém, nos casos de regressão logística e Poisson, o teste da razão de verossimilhanças aparece como o mais indicado, pelo fato de ser obtido pela diferença de duas funções desvios (PAULA, 2010).

Apresentam-se em seguida alguns procedimentos mais usados para a seleção de modelos.

3.5.1 Método forward

Inicialmente, considera-se o modelo $\mu = \alpha$. Ajusta-se, para cada variável explicativa o seguinte modelo

$$\mu = \alpha + \beta_j x_j, \quad (j = 1, 2, \dots, q) \quad (32)$$

São testadas as hipóteses $H_0: \beta_j = 0$ contra $H_1: \beta_j \neq 0$. Seja P o menor nível descritivo dentre os q testes. Se $P < P_E$, a variável correspondente entra no modelo. Repete-se procedimento até que ocorra $P > P_E$.

3.5.2 Método backward

Considera-se o seguinte modelo para o início do procedimento

$$\mu = \alpha + \beta_1 x_1 + \dots + \beta_q x_q \quad (33)$$

São testadas as hipóteses $H_0: \beta_j = 0$ contra $H_1: \beta_j \neq 0$. Seja P o menor nível descritivo dentre os q testes. Se $P > P_S$, a variável correspondente sai do modelo. Repete-se procedimento até que ocorra $P \leq P_S$.

3.5.3 Método stepwise

Segundo Paula (2010), um dos métodos mais aplicados em regressão logística é o método *stepwise*. O mesmo baseia-se num algoritmo misto de inclusão e eliminação de variáveis explicativas segundo a importância das mesmas de acordo com algum critério estatístico. Esse grau de importância pode ser avaliado, por exemplo, pelo nível de significância do teste da razão de verossimilhanças entre os modelos que incluem ou excluem as variáveis em questão. Quanto menor for esse nível de significância tanto mais importante será considerada a variável explicativa. Como a variável mais importante por esse critério não é necessariamente significativa do ponto de vista estatístico, deve-se impor um limite superior P_E , para esses níveis descritivos, a fim de atrair candidatos importantes em princípio de entrada.

O presente método é uma mistura dos procedimentos anteriores. Inicia-se o processo com o modelo $\mu = \alpha$. Após incluir duas variáveis no modelo, verifica-se se a primeira não sai do modelo. O processo continua até que nenhuma variável seja incluída ou retirada do modelo. Geralmente adota-se $0,15 \leq P_E, P_S \leq 0,25$. Uma sugestão seria usar $P_E = P_S = 0,20$.

Descreve-se em seguida uma variante deste algoritmo usada por Hosmer e Lemeshow (1989). A etapa inicial começa com o ajustamento do modelo só com intercepto e é constituída pelos seguintes passos:

- i) Constroem-se testes da razão de verossimilhanças entre o modelo inicial e os modelos logísticos simples formados com cada uma das covariáveis do estudo. O mínimo dos níveis descritivos associados a cada teste será comparado com $P_E = 0,15$. Se P_E for maior inclui-se a covariável referente àquele nível mínimo e passa-se ao passo seguinte; caso contrário, a seleção é interrompida e seleciona-se o último modelo;
- ii) Partindo do modelo incluindo a covariável selecionada no passo anterior, introduzem-se individualmente as demais covariáveis. Cada um destes modelos com duas covariáveis é testado contra o modelo inicial deste passo. Novamente o mínimo dos níveis descritivos, se for menor do que P_E , implica a inclusão no modelo da sua respectiva covariável, e a passagem ao passo seguinte. Caso contrário, interrompe-se a seleção;

- iii) Compara-se o ajuste do modelo logístico contendo as covariáveis selecionadas nos passos anteriores com os modelos que dele resultam por exclusão individual de cada uma das covariáveis. Se o máximo dos níveis descritivos destes testes da razão de verossimilhanças for menor do que $P_E = 0,20$, a covariável associada a este nível permanece no modelo. Caso contrário, ela é removida. Em qualquer circunstância, o algoritmo segue para o passo seguinte.
- iv) O modelo resultante do passo anterior será ajustado, e antes de tornar-se o modelo inicial da etapa 2 (seleção de interações de primeira ordem das covariáveis incluídas), repetem-se os passos anteriores quantas vezes forem necessários até se ter a indicação de parada nestes passos ou todas as covariáveis inclusas no modelo.
- v) Uma vez selecionadas as covariáveis “importantes”, ou seja, os seus efeitos principais, na etapa 1, se dá entrada na etapa 2 por meio do passo (i) com o objetivo de selecionar as interações que envolvem aquelas covariáveis, e assim por diante.

Uma desvantagem deste procedimento é a de exigir as estimativas de máxima verossimilhança em cada passo, o que encarece o trabalho computacional, particularmente em grandes amostras.

3.5.4 Método de Akaike

O presente método, proposto por Akaike (1974) se diferencia dos procedimentos anteriores por ser um processo de minimização que não envolve testes estatísticos. O método consiste em selecionar um modelo que seja parcimonioso, ou seja, que esteja bem ajustado e tenha um número reduzido de parâmetros. Como o logaritmo da função de verossimilhança $L(\beta)$ cresce com o aumento do número de parâmetros do modelo, uma proposta razoável seria encontrar o modelo com menor valor para a função

$$AIC = -L(\hat{\beta}) + p, \quad (34)$$

em que p denota o número de parâmetros.

3.6 ANÁLISE DE DIAGNÓSTICO E RESÍDUOS

O modelo proposto, depois de ajustado, precisa apresentar uma boa descrição dos dados que foram observados para que não se conduzam inferências errôneas sobre o mesmo. Neste caso, é importante um estudo da robustez dos resultados obtidos, quanto aos aspectos de formulação do modelo e as estimativas dos seus parâmetros.

Para Souza (2006), a análise dos resíduos e diagnóstico é utilizada para detectar problemas, tais como

- Presença de observações discrepantes (pontos aberrantes);
- Inadequação das pressuposições para os erros aleatórios ou para as médias;
- Colinearidade entre as colunas da matriz do modelo;
- Forma funcional do modelo inadequada;
- Presença de observações influentes;

Algumas das medidas utilizadas para a análise de resíduos e diagnóstico são abordadas na sequência.

3.6.1 Técnicas de diagnóstico

Segundo Paula (2010), estudos de simulação apresentados, por exemplo, em Williams (1984) tem sugerido o resíduo componente do desvio t_{Di} para as análises de diagnóstico em MLG. Para os modelos binomiais o resíduo é expresso em 35, para $0 < y_i < n_i$, assumindo a forma descrita em 35.

$$t_{Di} = \pm \sqrt{\frac{2}{1 - \hat{h}_{ii}} \left\{ y_i \log \left(\frac{y_i}{n_i \hat{\pi}_i} \right) + (n_i - y_i) \log \left(\frac{n_i - y_i}{n_i - n_i \hat{\pi}_i} \right) \right\}^{1/2}} \quad (35)$$

Quando $y_i = 0$ ou $y_i = n_i$ o componente do desvio padronizado toma as formas respectivamente.

$$t_{Di} = -\frac{\{2n_i|\log(1 - \hat{\pi}_i)|\}^{1/2}}{\sqrt{1 - \hat{h}_{ii}}} \text{ e } t_{Di} = -\frac{\{2n_i|\log(\hat{\pi}_i)|\}^{1/2}}{\sqrt{1 - \hat{h}_{ii}}}, \quad (36)$$

O resíduo Studentizado t_{Si} que também é utilizado para avaliar a presença de observações aberrantes tem a forma que pode ser observada em 37.

$$t_{Si} = \frac{1}{\sqrt{1 - \hat{h}_{ii}}} \frac{(y_i - n_i \hat{\pi}_i)}{\{n_i \hat{\pi}_i (1 - \hat{\pi}_i)\}^{1/2}} \quad (37)$$

A distância de Cook aproximada é utilizada para medir a influência das observações nas estimativas dos coeficientes, neste caso tem-se

$$LD_i = \frac{\hat{h}_{ii}}{(1 - \hat{h}_{ii})^2} \frac{(y_i - n_i \hat{\pi}_i)^2}{n_i \hat{\pi}_i (1 - \hat{\pi}_i)} \quad (38)$$

Para Hosmer e Lemeshow (1989), \hat{h}_{ii} depende das probabilidades ajustadas $\hat{\pi}_i$, $i = 1, \dots, k$, e, conseqüentemente, os resíduos t_{Si} e t_{Di} e a medida de influência LD_i também dependem, neste caso, tem-se que

$$h_{ii} = n_i \pi_i (1 - \pi_i) \mathbf{X}_i^T (\mathbf{X}^T \mathbf{V} \mathbf{X})^{-1} \mathbf{X}_i, \quad (39)$$

com $V = \text{diag}\{n_1 \pi_1 (1 - \pi_1), \dots, n_k \pi_k (1 - \pi_k)\}$. Por meio de um estudo numérico, realizado por Hosmer e Lemeshow (1989), é mostrado que o comportamento de \hat{h}_{ii} em uma regressão logística pode ser muito diferente do comportamento de \hat{h}_{ii} em um modelo de regressão normal para uma matriz de modelo \mathbf{X} (PAULA, 2010).

Por meio da Tabela 7, podem-se visualizar os possíveis valores de algumas medidas de diagnóstico em função das probabilidades ajustadas.

Também são recomendados em regressão logística os gráficos da variável adicionada e de $|l_{max}|$ contra $\hat{\pi}_i$.

Tabela 7 – Possíveis valores para algumas medidas de diagnóstico segundo as probabilidades ajustadas

Medida	Probabilidade ajustada				
	0,0 – 0,1	0,1 – 0,3	0,3 – 0,7	0,7 – 0,9	0,9 – 1,0
$t_{S_i}^2$	grande ou pequeno	moderado	moderado ou pequeno	moderado	grande ou pequeno
LD_i	pequeno	grande	moderado	grande	pequeno
\hat{h}_{ii}	pequeno	grande	moderado ou pequeno	grande	pequeno

Fonte: PAULA, 2010.

Segundo Paula (2010), a Tabela 7 descreve, em função das probabilidades ajustadas, os possíveis valores de algumas medidas de diagnóstico. A medida \hat{h}_{ii} é interpretada de maneira semelhante à medida h_{ii} da regressão linear normal para $0,1 < \hat{\pi}_{ii} < 0,9$. \hat{h}_{ii} fica, em geral, pequeno quando $\hat{\pi}_{ii}$ é pequena ou alta, o que pode dificultar a detecção de pontos que estejam mais afastados no subespaço gerado pelas colunas da matrix X. Neste caso, sugerem-se os gráficos $t_{S_i}^2$, $t_{D_i}^2$ e LD_i contra as probabilidades ajustadas $\hat{\pi}_i$. Os gráficos podem informar sobre o posicionamento de pontos aberrantes e influentes com relação às probabilidades ajustadas.

4 MATERIAIS E MÉTODOS

Apresentam-se neste capítulo os procedimentos metodológicos da pesquisa com o intuito de atingir os objetivos geral e específicos propostos. São apresentadas a caracterização da pesquisa, local e tipo de estudo, bem como a forma de coleta de dados e os procedimentos metodológicos utilizados.

Trata-se de uma pesquisa quantitativa, com explanação em números das informações obtidas, e com isso a análise e classificação dos dados coletados por meio de técnicas estatísticas apropriadas, como testes de hipóteses e análise de regressão logística.

Classifica-se o presente estudo como descritivo, retrospectivo e aplicado, dado a análise das características das variáveis em estudo, baseado em períodos passados a fim de obter uma solução à problemática apresentada. O local da pesquisa é o Campus da UFSM, localizado na cidade de Santa Maria/RS.

Os dados utilizados a fim de obter os resultados propostos no presente estudo foram solicitados ao Centro de Processamento de Dados (CPD) da UFSM. O período considerado para a construção do banco de dados foi de acadêmicos ingressantes na universidade nos anos entre 2009 e 2015.

O estudo compreenderá os acadêmicos de graduação da UFSM, sendo excluídos da pesquisa os acadêmicos dos Centros de ensino dos Campi localizados nas cidades de Frederico Westphalen, Cachoeira do Sul, Palmeiras das Missões e Silveira Martins. Os acadêmicos do Colégio Técnico Industrial (CTISM) e do Colégio Politécnico da UFSM também não foram considerados na pesquisa.

As variáveis qualitativas (categóricas) estudadas com relação ao ingresso de acadêmicos para o estudo da evasão são listadas, no Quadro 2, como segue:

Quadro 2 – Variáveis qualitativas (categóricas) e respectivas categorias

(continua)

Variável Dependente	Categoria
Evasão	Evadido
	Não Evadido
Variáveis Independentes	Categoria
Gênero	Feminino
	Masculino

Quadro 2 – Variáveis qualitativas (categóricas) e respectivas categorias

(conclusão)

Etnia	Branca
	Não Branca
Estado Civil	Casado
	Solteiro
	Outros
Cota	Ampla Concorrência
	Escola Pública
	Outras
Turno do Curso	Diurno
	Integral
	Noturno
Período de Ingresso	1º Semestre
	2º Semestre
Período Atual (curso)	1º, 2º e 3º Semestre
	4º, 5º e 6º Semestre
	7º, 8º e 9º Semestre
	10º, 11º e 12º Semestre
Área de Conhecimento	Ciências Agrárias
	Ciências Biológicas
	Ciências da Saúde
	Ciências Exatas e da Terra
	Ciências Humanas
	Ciências Sociais
	Ciências Sociais Aplicadas
	Engenharias
	Linguística, Letras e Artes
	Música
	Odontologia
Mora no local onde estuda	Sim
	Não
Aluno Bolsista	Sim
	Não
Participação em Projetos	Sim
	Não
Frequenta RU	Sim
	Não
Mora na Casa do Estudante	Sim
	Não

Fonte: Elaborado pela autora com base nos dados enviados pelo CPD – UFSM.

As variáveis independentes de caráter quantitativo, utilizadas no estudo são apresentadas como segue:

- Idade do acadêmico no momento do ingresso à universidade;
- Número de empréstimos realizados na biblioteca;
- Número de disciplinas aprovadas;
- Número de disciplinas reprovadas;
- Número de reprovações por frequência;
- Número de disciplinas dispensadas por nota;
- Número de trancamentos parciais;
- Número de trancamentos totais;

Com relação à variável dependente, considerou-se como evasão o abandono, jubramento, transferência interna (dentro do mesmo centro) e externa (outro centro ou instituição) do curso, por parte do aluno. Os discentes classificados como alunos regulares, formados, falecidos e os com cancelamento de matrícula antes mesmo de iniciar a graduação foram considerados não evadidos.

Quanto à variável etnia, classificaram-se os discentes de cor de pele preta, parda, indígena, amarela e outras como não branca.

Para a variável estado civil foram classificados os alunos com união estável como casados e, como outros, os alunos desquitados, divorciados, separados judicialmente e viúvos.

Com referência à variável cota, classificou-se cota afro-brasileira e necessidades especiais como outras e a variável ampla concorrência abrange os alunos que não entraram na universidade pelo sistema de cotas. Quanto à cota Escola Pública, considerou-se nesta classificação alunos de Ensino Médio em Escola Pública – Baixa Renda, Ensino Médio em Escola Pública – independente de renda, Ensino Médio em Escola Pública – Cota Social e Ensino Médio em Escola Pública – autodeclarado preto, pardo ou indígena com cota social.

A variável período atual, na análise de regressão logística múltipla de alguns cursos, precisou de adaptação, sendo então classificada como primeiros (1º, 2º, 3º, 4º, 5º e 6º) e últimos semestres (a partir do 7º semestre), devido ao reduzido número de discentes em cada classe.

4.1 PROCEDIMENTOS METODOLÓGICOS

Realizou-se o ajuste do banco de dados, categorizando as variáveis e analisando a ocorrência de dados faltantes.

Após o ajuste dos dados foi investigado o comportamento das variáveis por meio de estatística descritiva. As variáveis qualitativas categóricas foram apresentadas em forma de frequência relativa e as quantitativas por meio de média e desvio padrão.

Para o ajuste da regressão logística, o primeiro passo foi verificar, individualmente, a existência de associação entre as variáveis independentes e a variável dependente. Neste caso, foi utilizado o teste qui-quadrado (χ^2) de associação para variáveis qualitativas.

Segundo Callegari-Jacques (2003) o teste qui-quadrado de associação é utilizado para testar a associação entre variáveis categóricas, assim como o coeficiente de correlação r é calculado e testado com o mesmo propósito em variáveis quantitativas. Para realizar um teste qui-quadrado de associação, os indivíduos de uma amostra são estudados quanto a duas variáveis qualitativas e os dados, organizados em uma tabela de contingência, na qual as linhas e as colunas representam as categorias das duas variáveis em análise. O único total fixo, neste teste é o total de indivíduos estudados.

As hipóteses a serem testadas são:

H_0 = As frequências observadas não são diferentes das frequências esperadas. Não existe diferença entre as frequências dos grupos. Portanto não há associação entre os mesmos.

H_1 = As frequências observadas são diferentes das frequências esperadas. Existe diferença entre as frequências dos grupos. Portanto há associação entre os mesmos.

Para verificar a diferença de médias entre duas amostras independentes, para as variáveis quantitativas, utilizou-se o teste t de Student. A pressuposição de normalidade das amostras foi testada com os testes Shapiro-Wilk e Kolmogorov-Smirnov e a pressuposição de homogeneidade de variâncias, verificada com o teste de Levene.

O próximo passo foi o ajuste do modelo de regressão logística, em que foram averiguados os dados a fim de responder à seguinte questão de investigação: Quais variáveis são preditoras/explicativas da evasão na UFSM?

Inicialmente foram identificados os cursos, dentro de cada Centro, que apresentaram o pior e o melhor desempenho em relação à evasão, com o objetivo de avaliar, os fatores de risco que levam o aluno a evadir.

Foram testadas, uma a uma, as variáveis independentes com a variável dependente, realizando-se análise de regressão univariada, e selecionadas as que apresentarem $p\text{-valor} \leq 0,25$ (HOSMER e LEMESHOW, 1989). Essas variáveis foram incluídas no modelo de regressão logística múltiplo.

Por fim, realizou-se a análise de diagnóstico e resíduos do modelo com o propósito de avaliar se o modelo escolhido foi ajustado com boa qualidade.

Primeiramente, calcularam-se as OR e os respectivos IC95% utilizando-se uma análise de regressão univariada, para escolher as variáveis candidatas a inclusão no modelo múltiplo. Posteriormente, realizou-se a análise múltipla verificando o efeito conjunto das variáveis independentes no desfecho, elaborando-se um modelo de regressão logística que permite estimar a chance de um aluno de graduação evadir da universidade, em função das variáveis que se mostraram significativas ao nível de 5% de significância e que estão presentes no modelo ajustado.

O método adotado para o ajuste das variáveis no modelo foi o *stepwise*, por se tratar de um estudo exploratório e ajuste de modelos aos dados analisados. Neste caso, inicia-se com um modelo que inclui apenas a constante e depois adiciona uma a uma as variáveis previsoras com base em critérios específicos. O critério é o valor da estatística *score*: a variável adicionada no modelo é a que apresentar a estatística *score* mais significativa (FIELD, 2009).

Para Vitelli, Rocha e Fritschi (2010) a análise de uma série de indicadores para avaliação da qualidade do ajuste do modelo de regressão logística contribui para decidir sobre o melhor modelo para explicar os dados analisados. O *log likelihood value* (-2LL) é umas das principais medidas de avaliação. Para comparar o desempenho dos modelos concorrentes ou o grau de aderência dos modelos, utilizou-se o R^2 de Cox & Snell (1989) que é baseado na verossimilhança-log do modelo novo, verossimilhança-log do modelo original e tamanho da amostra. Como

essa estatística nunca alcança o seu valor teórico máximo (1), Nagelkerke (1991) sugeriu uma correção. Quanto maior o seu valor, melhor a qualidade do ajuste.

A contribuição individual de cada um dos previsores no modelo ajustado de regressão logística, inclusive da constante, foi avaliada pela estatística de Wald, que informa se o coeficiente b de cada variável preditora é diferente de 0. Se isso acontecer, assume-se que a variável explicativa está contribuindo de forma significativa para a previsão da variável dependente (Y).

O teste de Hosmer e Lemeshow tem por finalidade verificar a existência de diferenças significativas de valores esperados pelo modelo e de valores observados nos dados, sendo útil para avaliar a qualidade do modelo, o qual não é capaz de produzir estimativas e classificações muito confiáveis caso exista diferença significativa entre os valores.

Segundo Field (2009), os dois principais objetivos em examinar os resíduos nos modelos de regressão logística são (1) isolar os pontos em que o modelo tem pouca aderência e (2) isolar pontos que exercem uma influência indevida no modelo. A fim de avaliar o primeiro, devem-se examinar principalmente os resíduos *studentizados* os *padronizados* e as estatísticas de desvio. Todas as estatísticas mencionadas tem uma propriedade em comum: em média em uma amostra normalmente distribuída, 95% dos casos devem apresentar valores entre $\pm 1,96$, e 99% dos casos devem conter valores entre $\pm 2,58$.

As estatísticas de influência, como a distância de Cook, foram utilizadas com o propósito de avaliar a influência dos casos individuais. A versão padronizada da distância de Cook, o *DFBeta*, também informa sobre a influência de alguns casos. Assim, os valores devem ser, preferencialmente, menores que 1. Um valor 0 indica que não há influência alguma, já um valor igual a 1 expressa influência completa sobre o modelo. A influência esperada é $\frac{(k+1)}{N}$, onde K é o número de previsores e N é o tamanho da amostra.

Para a realização das análises estatísticas foi utilizado o software PASW 18, considerando-se a significância de 5%.

5 RESULTADOS E DISCUSSÃO

Neste capítulo são apresentados os resultados e discussões da análise descritiva e dos modelos de regressão logística, para os cursos, dentro de cada Centro, que apresentaram o pior e o melhor desempenho em relação à evasão.

5.1 ANÁLISE DESCRITIVA

Foram estudados discentes distribuídos em Centros de ensino, considerando em cada um, o curso com maior e menor evasão.

Os resultados para os Centros de ensino foram distribuídos em três tabelas (Tabela 8, 9 e 10), de acordo com características de área comum de conhecimento.

Na Tabela 8 estão os resultados do CAL, CCSH e CE; na Tabela 9 pode-se observar os resultados do CCNE, CT e CCR e, na Tabela 10, estão os resultados do CCS e CEFD.

Consideram-se os resultados quanto ao percentual de discentes evadidos segmentado por Centro, verifica-se que o CCNE e o CAL foram os centros que apresentaram maior percentual de evasão, 52,0% e 49,1% respectivamente. Os centros com menor percentual de alunos evadidos são o CCS e o CCR com, respectivamente, 11,6% e 19,4%.

Observa-se que o CCNE apresentou mais da metade de alunos evadidos no período considerado. Neste mesmo sentido, em estudo realizado por Fritsch, Da Rocha e Vitelli (2015), no qual foram considerados alunos ingressantes no ano de 2003 em uma instituição de ensino superior privada foi observado que, entre outras, a área de Exatas apresentou taxa de evasão superior à institucional, o que vem de encontro aos resultados observados no presente estudo, já que a área de Ciências Exatas está inserida dentro do CCNE, o qual apresentou o maior percentual de evasão de todos os centros analisados.

Observa-se que há uma representatividade maior de mulheres no CAL, CCSH, CE, CCNE, CCR e CCS. Destaca-se que, de acordo com Frischenbruder, 1999 (apud Bardagi, 2007), diferenças entre os gêneros estão associadas ao nível de comprometimento, indicando que os estudos tendem a ser mais explorados pelas mulheres do que pelos homens, visto que estes acabam trocando o estudo pelo emprego, por terem uma maior percepção de oportunidade de trabalho.

Tabela 8 – Resultados da análise da evasão e não evasão de discentes de graduação da UFSM, para os centros de ensino CAL, CCSH e CE, no período de 2009 a 2015

(continua)

Variável		CAL (n=2.662)			CCSH (n=6.412)			CE (n=3.290)		
		\bar{E} (%)	E (%)	p-valor	\bar{E} (%)	E (%)	p-valor	\bar{E} (%)	E (%)	p-valor
Gênero	Feminino	32,0	31,0	0,880	39,7	14,8	<0,001*	54,9	24,1	0,034*
	Masculino	18,9	18,1		28,9	16,6		13,7	7,3	
Idade	$\bar{X} \pm s$	23,20 \pm 7,10	24,09 \pm 7,56	0,002*	22,20 \pm 6,67	24,78 \pm 8,19	<0,001*	27,12 \pm 7,82	27,65 \pm 7,68	0,069
Etnia	Branca	44,5	42,5	0,486	58,7	26,5	0,435	61,3	27,4	0,604
	Não Branca	6,4	6,6		10,0	4,8		7,7	3,6	
Estado civil	Casado	3,8	5,2	0,002*	4,5	3,5	<0,001*	12,4	6,8	0,005*
	Solteiro	44,3	41,2		63,6	26,9		53,5	22,6	
	Outros	0,9	1,2		0,7	0,8		2,2	1,4	
Cota	Ampla Concorrência	25,5	24,6	0,820	48,5	24,5	<0,001*	59,2	27,3	0,109
	Escola Pública	24,7	23,5		17,1	5,1		8,0	3,2	
	Outras	0,8	0,9		3,0	1,8		1,4	0,9	
Turno do curso	Diurno	41,0	38,2	0,002*	43,1	14,3	<0,001*	35,0	13,1	<0,001*
	Noturno	4,6	6,5		25,5	17,1		14,1	7,8	
	Integral	5,3	4,4		-	-		19,5	10,5	
Período de ingresso	1° (Semestre)	37,1	34,7	0,269	47,2	21,9	0,407	35,8	17,1	0,238
	2° (Semestre)	13,9	14,3		21,4	9,5		32,8	14,3	
Área de conhecimento	Ciências Agrárias	-	-	0,065	-	-	<0,001*	-	-	-
	Ciências Biológicas	-	-		-	-		-	-	
	Ciências da Saúde	-	-		-	-		-	-	
	Ciências Exatas e da Terra	-	-		-	-		-	-	
	Ciências Humanas	-	-		15,7	12,1		68,6	31,4	
	Ciências Sociais	-	-		1,7	0,6		-	-	
	Ciências Sociais Aplicadas	4,4	5,6		51,2	18,7		-	-	
	Engenharias	-	-		-	-		-	-	
	Linguística, Letras e Artes	46,5	43,4		-	-		-	-	
	Música	0,1	0,1		-	-		-	-	
Odontologia	-	-	-	-	-	-				
Período atual do curso (Semestre)	1°, 2° e 3° Semestre	4,2	26,2	<0,001*	3,5	16,1	<0,001*	31,5	23,6	<0,001*
	4°, 5° e 6° Semestre	13,1	17,4		15,5	10,3		12,2	5,5	
	7°, 8° e 9° Semestre	31,3	5,3		34,6	4,5		21,9	2,2	
	10°, 11° e 12° Semestre	2,3	0,2		15,0	0,5		3,0	0,2	

Tabela 8 – Resultados da análise da evasão e não evasão de discentes de graduação da UFSM, para os centros de ensino CAL, CESH e CE, no período de 2009 a 2015

(conclusão)										
Mora no local onde estuda	Sim	18,0	19,0	0,046*	25,7	12,0	0,321	26,6	13,2	0,042*
	Não	33,1	29,8		43,2	19,1		42,2	18,0	
Aluno bolsista	Sim	39,6	17,6	<0,001*	43,6	6,6	<0,001*	60,6	17,3	<0,001*
	Não	11,3	31,4		25,0	24,8		8,0	14,1	
Participação em projetos	Sim	32,9	13,3	<0,001*	34,0	3,5	<0,001*	25,5	2,7	<0,001*
	Não	18,0	35,7		34,6	27,9		43,1	28,7	
Mora na CEU	Sim	2,5	0,8	<0,001*	1,8	0,2	<0,001*	1,7	0,2	<0,001*
	Não	48,5	48,2		66,8	31,2		66,9	31,2	
Frequenta RU	Sim	46,0	24,8	<0,001*	56,2	13,4	<0,001*	33,2	8,2	<0,001*
	Não	5,0	24,3		12,4	18,0		35,4	23,2	
Empréstimos na biblioteca	$\bar{X} \pm s$	43,06 ± 65,45	27,22 ± 49,02	<0,001*	27,93 ± 44,27	17,55 ± 61,41	<0,001*	30,94 ± 44,91	24,71 ± 46,19	<0,001*
Disciplinas aprovadas	$\bar{X} \pm s$	28,09 ± 18,19	8,75 ± 10,68	<0,001*	32,24 ± 17,33	6,76 ± 8,35	<0,001*	29,91 ± 17,26	6,57 ± 9,33	<0,001*
Disciplinas reprovadas	$\bar{X} \pm s$	1,30 ± 2,11	1,01 ± 1,72	<0,001*	1,55 ± 2,59	1,64 ± 2,45	0,174	0,77 ± 2,01	1,40 ± 1,99	<0,001*
Reprovações por frequência	$\bar{X} \pm s$	4,08 ± 5,30	4,24 ± 4,88	0,493	2,41 ± 4,40	4,49 ± 4,98	<0,001*	1,42 ± 4,65	4,33 ± 5,39	<0,001*
Disciplinas dispensadas por nota	$\bar{X} \pm s$	1,52 ± 5,44	0,68 ± 3,93	<0,001*	0,65 ± 2,78	0,43 ± 2,14	<0,001*	0,20 ± 1,48	0,15 ± 0,99	0,354
Trancamentos totais	$\bar{X} \pm s$	0,26 ± 0,70	0,73 ± 1,05	<0,001*	0,24 ± 0,69	0,98 ± 1,21	<0,001*	0,15 ± 0,48	0,81 ± 0,94	<0,001*
Trancamentos parciais	$\bar{X} \pm s$	0,71 ± 1,20	0,31 ± 0,85	<0,001*	0,51 ± 0,99	0,23 ± 0,69	<0,001*	0,27 ± 0,78	0,28 ± 0,86	0,679

Fonte: Elaborado pela autora.

*Significativo ao nível de 5% de significância.

\bar{E} : Discente não evadido; E : Discente evadido.

$\bar{X} \pm s$: Média ± desvio padrão

Tabela 9 – Resultados da análise da evasão e não evasão de discentes de graduação da UFSM, para os centros de ensino CCNE, CT e CCR, no período de 2009 a 2015

(continua)

Variável	CCNE (n=3.785)			CT (n=4.219)			CCR (n=3.283)			
	\bar{E} (%)	E(%)	p-valor	\bar{E} (%)	E(%)	p-valor	\bar{E} (%)	E(%)	p-valor	
Gênero	Feminino	27,2	25,8	<0,001*	23,5	5,9	<0,001*	43,5	11,7	0,004*
	Masculino	20,8	26,2		52,7	18,0		37,1	7,7	
Idade	$\bar{X} \pm s$	22,28 ± 6,05	23,61 ± 7,27	<0,001*	19,73 ± 3,16	20,97 ± 4,83	<0,001*	20,58 ± 3,73	22,36 ± 6,53	<0,001*
Etnia	Branca	42,6	45,1	0,181	67,3	19,7	0,001*	71,5	16,3	0,003*
	Não Branca	5,5	6,7		9,2	3,8		9,2	3,0	
Estado civil	Casado	3,5	5,7	<0,001*	1,0	0,9	<0,001*	2,0	1,1	<0,001*
	Solteiro	43,8	45,2		75,5	22,4		78,7	17,7	
	Outros	0,5	1,2		0,1	0,1		0,2	0,2	
Cota	Ampla Concorrência	39,3	41,5	0,006*	52,8	16,7	<0,001*	56,3	14,2	0,001*
	Escola Pública	8,0	9,0		20,6	4,9		21,2	3,8	
	Outras	0,7	1,5		2,8	2,3		3,1	1,5	
Turno do curso	Diurno	28,3	27,5	<0,001*	76,1	23,9	-	57,2	16,0	<0,001*
	Noturno	8,0	12,4		-	-		-	-	
	Integral	11,8	12,1		-	-		23,4	3,4	
Período de ingresso	1º Semestre	36,2	38,0	0,128	27,9	7,2	<0,001*	40,0	9,0	0,158
	2º Semestre	11,8	13,9		48,2	16,7		40,5	10,4	
Área de conhecimento	Ciências Agrárias	-	-	<0,001*	-	-	<0,001*	80,6	19,4	-
	Ciências Biológicas	9,0	10,5		-	-		-	-	
	Ciências da Saúde	-	-		-	-		-	-	
	Ciências Exatas e da Terra	30,9	35,5		8,6	5,1		-	-	
	Ciências Humanas	8,1	6,1		-	-		-	-	
	Ciências Sociais	-	-		-	-		-	-	
	Ciências Sociais Aplicadas	-	-		4,1	0,7		-	-	
	Engenharias	-	-		63,5	18,0		-	-	
	Linguística, Letras e Artes	-	-		-	-		-	-	
	Música	-	-		-	-		-	-	
Odontologia	-	-	-	-	-	-				
Período atual do curso	1º, 2º e 3º Semestre	6,3	29,2	<0,001*	4,0	10,9	<0,001*	7,6	13,8	<0,001*
	4º, 5º e 6º Semestre	13,3	16,1		15,9	8,4		22,3	4,5	
	7º, 8º e 9º Semestre	25,7	6,4		24,6	3,4		18,2	1,0	
	10º, 11º e 12º Semestre	2,6	0,3		31,7	1,1		32,4	0,1	

Tabela 9 – Resultados da análise da evasão e não evasão de discentes de graduação da UFSM, para os centros de ensino CCNE, CT e CCR, no período de 2009 a 2015

(conclusão)										
Mora no local onde estuda	Sim	15,2	20,4	<0,001*	21,5	8,0	0,001*	22,3	7,6	<0,001*
	Não	33,0	31,4		54,9	15,7		58,7	11,4	
Aluno bolsista	Sim	35,3	16,7	<0,001*	54,3	4,9	<0,001*	73,5	7,3	<0,001*
	Não	12,7	35,3		21,8	18,9		7,1	12,1	
Participação em projetos	Sim	27,1	9,3	<0,001*	52,5	3,8	<0,001*	58,9	3,0	<0,001*
	Não	20,9	42,7		23,6	20,0		21,7	16,4	
Mora na CEU	Sim	2,4	1,0	<0,001*	1,7	0,2	0,002*	4,6	0,2	<0,001*
	Não	45,6	51,0		74,5	23,7		76,0	19,3	
Frequenta RU	Sim	40,7	21,5	<0,001*	72,2	17,2	<0,001*	77,8	12,9	<0,001*
	Não	7,3	30,5		4,0	6,7		2,8	6,6	
Empréstimos na biblioteca	$\bar{X} \pm s$	48,31 ± 50,17	31,92 ± 82,95	<0,001*	31,89 ± 38,90	22,40 ± 41,29	<0,001*	23,82 ± 30,55	12,57 ± 25,25	<0,001*
Disciplinas aprovadas	$\bar{X} \pm s$	22,39 ± 17,34	5,18 ± 7,22	<0,001*	36,23 ± 20,86	7,81 ± 9,21	<0,001*	46,27 ± 22,40	8,24 ± 10,19	<0,001*
Disciplinas reprovadas	$\bar{X} \pm s$	3,47 ± 4,60	2,16 ± 3,24	<0,001*	5,55 ± 6,71	4,56 ± 5,23	<0,001*	3,85 ± 5,38	4,38 ± 4,95	0,023*
Reprovações por frequência	$\bar{X} \pm s$	2,83 ± 4,95	3,72 ± 4,52	<0,001*	2,79 ± 5,31	5,34 ± 6,10	<0,001*	1,93 ± 4,93	4,09 ± 4,49	<0,001*
Disciplinas dispensadas por nota	$\bar{X} \pm s$	3,22 ± 8,56	1,48 ± 6,28	<0,001*	0,54 ± 2,35	0,29 ± 1,61	0,001*	1,58 ± 6,35	0,26 ± 1,79	<0,001*
Trancamentos totais	$\bar{X} \pm s$	0,30 ± 0,77	0,65 ± 1,00	<0,001*	0,16 ± 0,54	0,93 ± 1,17	<0,001*	0,17 ± 0,57	1,06 ± 1,17	<0,001*
Trancamentos parciais	$\bar{X} \pm s$	0,47 ± 0,93	0,24 ± 0,66	<0,001*	0,45 ± 0,92	0,26 ± 0,68	<0,001*	0,24 ± 0,69	0,17 ± 0,70	0,030*

Fonte: Elaborado pela autora.

*Significativo ao nível de 5% de significância.

\bar{E} : Discente não evadido; E : Discente evadido.

$\bar{X} \pm s$: Média \pm desvio padrão da amostra analisada

Tabela 10 – Resultados da análise da evasão e não evasão de discentes de graduação da UFSM, para os centros de ensino CCS e CEFD, no período de 2009 a 2015

(continua)

Variável	CCS (n=3.496)			CEFD (n=1.134)			
	\bar{E} (%)	E (%)	p-valor	\bar{E} (%)	E (%)	p-valor	
Gênero	Feminino	66,7	8,7	0,915	30,6	11,6	0,057
	Masculino	21,7	2,9		38,8	19,0	
Idade	$\bar{X} \pm s$	20,54 ± 3,82	21,79 ± 5,82	<0,001*	21,31 ± 4,57	22,06 ± 4,75	0,012*
Etnia	Branca	75,7	9,2	0,04*	57,4	24,9	0,639
	Não Branca	13,0	2,1		12,6	5,1	
Estado civil	Casado	2,8	0,9	<0,001*	2,3	2,4	0,003*
	Solteiro	85,6	10,0		67,0	27,5	
	Outros	0,5	0,1		0,5	0,4	
Cota	Ampla Concorrência	59,1	8,2	0,893	48,6	22,6	0,306
	Escola Pública	23,9	2,0		17,1	6,3	
	Outras	5,5	1,3		3,7	1,8	
Turno do curso	Diurno	59,3	4,9	<0,001*	69,4	30,6	-
	Noturno	-	-		-	-	
	Integral	29,2	6,6		-	-	
Período de ingresso	1° Semestre	46,0	6,4	0,178	35,3	15,5	0,974
	2° Semestre	42,4	5,1		34,1	15,1	
Área de conhecimento	Ciências Agrárias	-	-	<0,001*	-	-	0,070
	Ciências Biológicas	-	-		-	-	
	Ciências da Saúde	74,7	10,6		66,2	28,4	
	Ciências Exatas e da Terra	-	-		-	-	
	Ciências Humanas	-	-		-	-	
	Ciências Sociais	-	-		-	-	
	Ciências Sociais Aplicadas	-	-		-	-	
	Engenharias	-	-		-	-	
	Linguística, Letras e Artes	-	-		3,2	2,2	
	Música	-	-		-	-	
	Odontologia	13,8	1,0		-	-	
Período atual do curso	1°, 2° e 3° Semestre	3,2	7,2	<0,001*	3,1	18,3	<0,001*
	4°, 5° e 6° Semestre	17,7	3,2		17,3	8,1	
	7°, 8° e 9° Semestre	37,6	1,1		49,0	4,2	
	10°, 11° e 12° Semestre	30,0	0,0		-	-	

Tabela 10 – Resultados da análise da evasão e não evasão de discentes de graduação da UFSM, para os centros de ensino CCS e CEFD, no período de 2009 a 2015

								(conclusão)
Mora no local (cidade) onde estuda	Sim	25,4	4,0	0,004*	28,6	12,7	0,518	
	Não	63,3	7,3		41,7	17,0		
Aluno bolsista	Sim	79,4	2,0	<0,001*	62,5	9,0	<0,001*	
	Não	9,1	9,6		6,9	21,6		
Participação em projetos	Sim	68,5	0,8	<0,001*	47,0	5,8	<0,001*	
	Não	20,0	10,8		22,4	24,8		
Mora na CEU	Sim	1,8	0,0	0,012*	3,3	0,4	0,007*	
	Não	86,7	11,5		66,1	30,2		
Frequenta RU	Sim	86,0	7,7	<0,001*	64,5	18,7	<0,001*	
	Não	2,5	3,9		4,9	11,9		
Empréstimos na biblioteca	$\bar{X} \pm s$	37,55 ± 33,38	20,73 ± 33,12	<0,001*	25,68 ± 35,21	16,52 ± 41,23	<0,001*	
Disciplinas aprovadas	$\bar{X} \pm s$	43,36 ± 18,97	8,74 ± 12,09	<0,001*	37,56 ± 17,62	8,65 ± 10,73	<0,001*	
Disciplinas reprovadas	$\bar{X} \pm s$	1,35 ± 3,34	2,40 ± 3,64	<0,001*	2,53 ± 3,58	2,16 ± 3,41	0,101	
Reprovações por frequência	$\bar{X} \pm s$	0,93 ± 3,24	4,40 ± 6,10	<0,001*	2,35 ± 4,82	4,82 ± 6,52	<0,001*	
Disciplinas dispensadas por nota	$\bar{X} \pm s$	0,51 ± 2,78	0,31 ± 1,72	0,163	0,78 ± 3,01	0,91 ± 3,51	0,522	
Trancamentos totais	$\bar{X} \pm s$	0,12 ± 0,50	1,40 ± 1,40	<0,001*	0,25 ± 0,65	1,09 ± 1,14	<0,001*	
Trancamentos parciais	$\bar{X} \pm s$	0,25 ± 0,73	0,17 ± 0,64	0,043*	0,70 ± 1,10	0,33 ± 0,89	<0,001*	

Fonte: Elaborado pela autora.

*Significativo ao nível de 5% de significância.

\bar{E} : Discente não evadido; E : Discente evadido.

$\bar{X} \pm s$: Média ± desvio padrão da amostra analisada

Quanto à idade dos acadêmicos, observa-se que, em geral, a média de idade de alunos evadidos foi maior, variando de $20,97 \pm 4,83$ anos, no CT, a $27,65 \pm 7,68$ anos, no CE. Houve diferença significativa ($p < 0,05$) entre as médias de idade de discentes evadidos e não evadidos em todos os centros, com exceção do CE, que foi de $27,65 \pm 7,68$ anos e $27,12 \pm 7,82$ anos, respectivamente.

Pode-se observar, nas Tabelas 8, 9 e 10, maior representatividade de acadêmicos de cor branca em todos os centros considerados, apresentando associação significativa entre raça e evasão para os discentes do CT, CCR e CCS ($p\text{-valor} < 0,05$).

Os resultados indicaram que, em todos os Centros analisados, os estudantes solteiros tem maior representatividade na amostra, em média 90% é classificado como tal, o que vem de encontro com os achados do estudo de Assis (2013). Este autor analisou o perfil dos alunos ingressantes e evadidos com o objetivo de conhecer os principais fatores que influenciam a evasão nos cursos superiores de tecnologia em um IES privada.

Para os discentes do CAL, CCNE e CEFD observou-se maior percentual de evadidos entre os casados. Considerando que a evasão entre os discentes mais jovens é menor e a condição civil solteiro ser mais representativa, pode-se identificar uma melhor explicação relacionando a idade com o estado civil.

Quanto à forma de ingresso, em geral, ocorreu por ampla concorrência, ou seja, sem cotas. Uma parcela bem pequena dos ingressos, em média 3,7% foi por cotas raciais ou necessidades especiais, sendo que se observou associação significativa ($p < 0,05$) com a evasão no CCSH, CCNE, CT e CCR.

Em relação ao turno do curso, observa-se em geral, que a maioria estuda no período diurno, visto que a maioria dos cursos é ofertada durante o dia. O CCS e CCR não apresentam opções de cursos noturnos, no CEFD e CT, não há opção para o período noturno e integral, e o CCSH não dispõe da opção de turno integral. Observou-se associação significativa com a evasão para todos os Centros em que foi possível realizar a comparação, sendo que, em geral, a evasão foi maior em Centros e, conseqüentemente, em cursos noturnos.

O período de ingresso na universidade também foi investigado, em relação à evasão, verificando-se que, na maioria dos Centros, o ingresso foi no primeiro semestre do ano, como em geral, acontece na maioria das universidades, com

exceção do CT, que apresenta 64,9% de ingressantes no 2º semestre do ano e que se observou associação significativa ($p < 0,001$) com a evasão.

Sobre a variável área de conhecimento, utilizou-se a mesma a fim de caracterizar cada Centro, pode-se perceber que cada Centro apresentou no mínimo uma e no máximo três áreas de conhecimento.

Quando analisado o período no qual o discente está no curso, observa-se que os primeiros semestres (1º, 2º e 3º) são os responsáveis pelo maior percentual de evasão, em todos os Centros considerados, sendo que, à medida que o final do curso vai se aproximando, o percentual de evasão vai diminuindo, tendo assim, associação significativa com a evasão. Neste contexto, em seu estudo com o objetivo de identificar o perfil de alunos com potencial de evasão em uma Universidade Comunitária do Rio Grande do Sul, através da aplicação do processo de Descoberta de Conhecimento em Bases de Dados (DCBD), Paz e Casella (2017), verificaram que os alunos que estão em semestres iniciais tem maior tendência à evasão, o que vem de encontro aos resultados do presente estudo.

Também, pode-se observar que, em média, 65,0% dos discentes vieram de outros municípios para estudar em Santa Maria, sendo que, nestes casos, também se observou maior percentual de evasão.

Observou-se ainda que, alunos que não tinham bolsa ou não participaram de projetos durante a graduação, apresentaram maior percentual de evasão, apresentando associação significativa ($p < 0,001$) com a evasão em todos os Centros de ensino.

O percentual de discentes que moram na CEU foi baixo, em média 2,8%, mas observou-se que, em todos os Centros de ensino, a evasão é bem menor nestes casos. Por exemplo, no CCNE que é o Centro que apresentou maior percentual de evasão (52,0%), apenas 1,0% dos alunos evadidos moram na CEU.

Além disso, a evasão apresentou associação significativa ao fato de frequentar RU ($p < 0,001$), sendo maior entre os frequentadores.

Mensurou-se, como forma de avaliar a evasão, o número de empréstimos realizados na biblioteca, considerando todos os tipos de empréstimos do acervo.

Observou-se grande variação e que, na totalidade, o número de empréstimos foi superior entre os não evadidos, sendo, em média, 33,60 empréstimos por aluno não evadido e 21,70 por alunos evadidos. No CCNE, apesar de ser o Centro com maior percentual de evasão, o número médio de empréstimos (31,92) realizados por aluno evadido foi maior do que no CCS (20,73), Centro este que apresentou menor percentual de evasão.

O número de disciplinas aprovadas foi mensurado a fim de avaliar as características dos discentes quanto às variáveis acadêmicas. Verificou-se que, na sua totalidade, alunos não evadidos apresentaram número médio de disciplinas aprovadas superior em relação à alunos evadidos, o que também foi investigado no estudo realizado por Fritsch, Da Rocha e Vitelli (2015), em sua pesquisa realizada com alunos de graduação em uma instituição de ensino superior privada. Para os autores, quanto maior o percentual de aprovação que o aluno tem ao longo de sua trajetória no curso, menor tende a ser o seu percentual de evasão.

Mensurou-se também o número de disciplinas reprovadas e de reprovações por frequência, onde se observou que, em relação ao número de reprovações, alguns centros, como o CAL, CCNE, CT e CEFD apresentaram número médio maior para alunos não evadidos, o contrário do esperado. Pressupõe-se que seja pelo fato de que um aluno não evadido permanece mais tempo na universidade e, assim, cursa mais disciplinas do que um aluno evadido, obtendo mais reprovações. Verificou-se, também, que alunos evadidos apresentaram um número médio (4,43) maior de reprovações por frequência do que os outros (2,34).

Em relação a disciplinas dispensadas, observou-se que, em geral, o número médio é maior para discentes não evadidos.

Quanto a trancamentos parciais de disciplinas, verificou-se que, em geral, o número médio foi maior para alunos não evadidos, sendo que não houve diferença significativa com a evasão apenas no CE. Considerando os trancamentos totais, de semestre, verifica-se que o número médio foi superior entre os evadidos, com média de 0,96 e 0,23 entre os não evadidos, apresentando associação significativa com a evasão em todos os Centros de ensino.

5.2 ANÁLISE DE REGRESSÃO LOGÍSTICA

Foram identificados os cursos, dentro de cada Centro de ensino, que apresentaram o melhor e o pior desempenho em relação à evasão, com o propósito de avaliar os fatores de risco associados. Podem-se observar, por meio da Tabela 11, os cursos considerados, bem como o percentual de evasão de cada um.

Tabela 11 – Percentual de evasão nos cursos, por Centro de ensino da UFSM, no período de 2009 a 2015

Centro	Curso	% de Evasão
CAL	Letras - Licenciatura - Habilitação Português e Literatura Língua Portuguesa	32,6%
	Bacharelado em Letras-Português/Literaturas	62,0%
CCNE	Ciências Biológicas - Bacharelado	8,9%
	Matemática - Licenciatura e Bacharelado	83,1%
CCR	Medicina Veterinária	7,1%
	Curso Superior de Tecnologia em Alimentos	42,0%
CCS	Medicina	2,4%
	Terapia Ocupacional	21,0%
CCSH	Direito Diurno	8,0%
	Licenciatura em Sociologia	84,9%
CE	Programa Especial de Graduação de Formação de Professores (PEG)	27,2%
	Licenciatura em Educação Especial - Noturno	38,0%
CEFD	Educação Física - Bacharelado	29,2%
	Dança - Licenciatura	41,0%
CT	Engenharia Civil	11,5%
	Bacharelado em Sistemas de Informação	44,9%

Fonte: Elaborado pela autora.

Por meio da Tabela 12, verificam-se as variáveis que foram significativas para os modelos ajustados de regressão logística dos cursos com melhor e pior desempenho dentro de cada centro de ensino. São apresentadas as OR brutas e ajustadas dos modelos.

Foram incluídas, no modelo geral, todas as variáveis significativas na análise univariada, com exceção das variáveis disciplinas reprovadas e aprovadas, pois se entende que uma é complementar da outra, neste caso, avaliaram-se os modelos com as variáveis separadamente.

5.2.1 Centro de Artes e Letras

O curso de Bacharelado em Letras – Português/Literaturas da UFSM foi o que apresentou o pior desempenho em relação à evasão dentre os cursos do CAL, 62,0%. Verifica-se que, em relação à etnia, ser de raça não branca configura-se como fator de proteção para evasão, o mesmo pode ser verificado em disciplinas aprovadas, observa-se que, para cada disciplina aprovada, há 0,08 menos chance de o aluno evadir. Além disso, observa-se que a cada trancamento total aumenta 1,83 vezes o risco de evasão.

A qualidade do ajuste foi verificada pelo teste de Hosmer e Lemeshow, com p-valor = 0,33 e, assim, conclui-se que o modelo está bem ajustado aos dados. O valor de R^2 (Nagelkerke) indica que 43,0% da evasão é explicada pelas variáveis preditoras do modelo.

As variáveis significativas no modelo de regressão logística do curso de Letras/Licenciatura – Português e Língua Portuguesa, o qual foi avaliado com melhor desempenho em relação à evasão, no CAL, foram se o aluno frequenta RU, o número de disciplinas aprovadas, número de reprovações por frequência e o número de trancamentos totais. Verifica-se que o discente frequentar RU e o número de disciplinas aprovadas expressam um fator de proteção em relação à evasão, enquanto que o número de reprovações por frequência e trancamentos totais mostraram-se como fator de risco. O valor de R^2 (Nagelkerke) foi de 59,0% e a significância do teste de Hosmer e Lemeshow foi p-valor = 0,50, indicando um bom ajuste.

Para cada modelo ajustado de regressão logística no CAL, realizou-se a análise de resíduos e diagnóstico. Verificaram-se os resíduos padronizados, studentizados, as estatísticas de desvio, medidas de influência e a distância de Cook.

Para o modelo de Bacharelado em Letras, observou-se, em todos os casos, valores de DFBetas menores que 1 e a estatística de influência próxima ao valor esperado de 0,01. Os valores da distância de Cook estiveram todos abaixo de 1 e os resíduos padronizados apresentaram valores menores que 3, indicando bom ajuste do modelo.

Considerando a análise de resíduos do modelo ajustado do curso de Licenciatura em Letras, puderam-se observar valores não influentes, ou seja, menores que 1, através da distância de Cook. Com relação aos resíduos padronizados, verificaram-se em três casos valores maiores que 3, assim os mesmos foram investigados por se apresentarem

diferentes dos demais, porém eliminá-los do modelo não foi suficiente para melhorar o ajuste.

5.2.2 Centro de Ciências Naturais e Exatas

Os dois cursos avaliados no CCNE foram Ciências Biológicas – Bacharelado e Matemática – Licenciatura e Bacharelado, com 8,9% e 83,1% de alunos evadidos, respectivamente. Observa-se que as variáveis período atual e disciplinas reprovadas foram significativas no modelo ajustado de regressão logística em ambos os cursos analisados.

Verifica-se que no curso de Ciências Biológicas, estar cursando um dos primeiros semestres aumenta 28,28 vezes o risco de evasão quando comparado com os últimos semestres do curso, verificando-se também que a OR ajustada é maior que a OR bruta (18,12).

A participação em projetos é caracterizada como fator de proteção em relação à evasão no curso de Ciências Biológicas, ou seja, o aluno que se envolve em projetos durante a graduação tem 0,16 menos chance de evadir do que aquele que não participa.

A variável idade do discente foi testada no modelo de regressão logística ajustado do curso de Matemática e apresentou-se significativa, neste caso, o risco de evasão é de 13,0% maior a cada ano de idade do discente. Com relação à variável número de disciplinas aprovadas, a mesma foi categorizada e verifica-se que uma ou mais disciplinas aprovadas apresenta 4,32 vezes o risco de o aluno evadir quando comparada a nenhuma disciplina reprovada durante a graduação.

Para os dois modelos ajustados de regressão logística dos cursos do CCNE, o teste de Hosmer e Lemeshow apresentou p -valor $> 0,05$ (p -valor = 0,60 para o curso de Ciências Biológicas e p -valor = 0,41 para o curso de Matemática), o que indica que o modelo está bem ajustado aos dados. O valor de R^2 (Nagelkerke) foi de 55,0% para o curso de Ciências Biológicas e 48,0% para o curso de Matemática Licenciatura e Bacharelado.

Na análise de resíduos do modelo ajustado do curso de Ciências Biológicas, verificou-se, com a distância de Cook, que não houve pontos influentes no modelo e os resíduos padronizados não apresentaram valores maiores que 3.

A existência de pontos influentes no modelo ajustado para o curso de Matemática foi analisada através da distância de Cook, onde não se observou influência alguma, ou seja, em todos os casos os valores foram menores que 1.

5.2.3 Centro de Ciências Rurais

Ainda na Tabela 12, pode-se observar o modelo de regressão logística ajustado para o curso de Medicina Veterinária. As variáveis significativas presentes no modelo de regressão logística com o propósito de explicar a evasão são aluno bolsista, participação em projetos, número de empréstimos na biblioteca, número de disciplinas aprovadas e número de trancamentos totais.

Para o Curso Superior de Tecnologia em Alimentos, verificou-se que além das variáveis citadas como significativas no modelo de regressão logística do curso de Medicina Veterinária ainda pode-se observar a idade, cota (ampla concorrência, escola pública e outras) e frequentar RU. Neste caso, identifica-se semelhança entre a significância das variáveis para explicar a evasão dos cursos analisados dentro do mesmo centro. Por exemplo, o fato de o aluno ter bolsa ou participar de projetos é um fator de proteção para evasão.

O indicador de qualidade do modelo R^2 (Nagelkerk) foi 68,0% e 67,0% para o curso de Medicina Veterinária e Curso Superior de Tecnologia em Alimentos, respectivamente. O teste de Hosmer e Lemeshow apresentou p-valor=0,97 para o curso de Medicina Veterinária e p-valor=0,34 para o curso Superior de Tecnologia em Alimentos, o que indica que o modelo está bem ajustado aos dados.

Para ambos os modelos ajustados dos cursos analisados no CCR não se observou pontos influentes, através da Distância de Cook. Os resíduos padronizados apresentaram valores dentro do intervalo de ± 3 desvios-padrão.

Tabela 12 – Resultado da análise de regressão logística de discentes dos cursos da UFSM, com melhor e pior desempenho em relação à evasão, no período de 2009 a 2015

(continua)

Curso/Variáveis	OR bruto	IC 95%	p-valor	OR ajustado	IC 95%	p-valor
Letras/Licenciatura - Português e Literatura Língua Portuguesa (CAL)						
Frequência RU						
Não	1			1		
Sim	0,13	0,07 - 0,24	<0,001*	0,42	0,19 - 0,95	0,04*
Disciplinas aprovadas	0,91	0,88 - 0,93	<0,001*	0,92	0,92 - 0,90	<0,001*
Reprovações por frequência	1,05	1,01 - 1,09	0,02*	1,06	1,01 - 1,13	0,02*
Trancamentos totais	2,70	1,88 - 3,88	<0,001*	1,57	1,06 - 2,31	0,02*
	R ² = 0,59		Teste Hosmer e Lemeshow - p-valor = 0,50			
Bacharelado em Letras - Português/Literaturas (CAL)						
Etnia						
Branca	1			1		
Não Branca	0,52	0,27 - 1,02	0,06	0,42	0,19 - 0,92	0,03*
Disciplinas aprovadas	0,91	0,89 - 0,94	<0,001*	0,92	0,90 - 0,95	<0,001*
Trancamentos totais	2,73	1,76 - 4,21	<0,001*	1,83	1,15 - 2,91	0,01*
	R ² = 0,43		Teste Hosmer e Lemeshow - p-valor = 0,33			
Ciências Biológicas - Bacharelado (CCNE)						
Período atual						
Últimos Semestres	1			1		
Primeiros Semestres	18,12	4,99 - 35,72	<0,001*	28,28	4,74 - 48,11	<0,001*
Participação em projetos						
Não	1			1		
Sim	0,07	0,02 - 0,25	<0,001*	0,16	0,04 - 0,68	0,01*
Disciplinas reprovadas	1,17	1,04 - 1,30	0,01*	1,29	1,08 - 1,53	0,01*
Trancamentos totais	2,55	1,58 - 4,12	<0,001*	2,33	1,23 - 4,43	0,01*
	R ² = 0,55		Teste Hosmer e Lemeshow - p-valor = 0,60			

Tabela 12 – Resultado da análise de regressão logística de discentes dos cursos da UFSM com melhor e pior desempenho em relação à evasão, no período de 2009 a 2015

(continua)

Matemática - Licenciatura e Bacharelado (CCNE)						
Idade	1,12	1,01 - 1,23	0,03*	1,13	1,01 - 1,26	0,03*
Mora no local onde estuda						
Não	1			1		
Sim	3,22	1,34 - 7,72	0,01*	0,31	0,12 - 0,77	0,01*
Disciplinas reprovadas						
Nenhuma	1			1		
Uma ou mais	1,50	0,69 - 3,22	0,22	4,32	1,64 - 11,37	0,01*
R ² = 0,48 Teste Hosmer e Lemeshow - p-valor = 0,41						
Medicina Veterinária (CCR)						
Aluno Bolsista						
Não	1			1		
Sim	0,01	0,01 - 0,03	<0,001*	0,13	0,05 - 0,43	<0,001*
Participação em projetos						
Não	1			1		
Sim	0,04	0,01 - 0,09	<0,001*	0,24	0,08 - 0,73	0,01*
Empréstimos na biblioteca	0,94	0,91 - 0,96	<0,001*	0,98	0,96 - 0,99	0,04*
Disciplinas aprovadas	0,87	0,84 - 0,90	<0,001*	0,93	0,89 - 0,96	<0,001*
Trancamentos totais	2,79	2,08 - 3,74	<0,001*	1,58	1,12 - 2,14	0,01*
R ² = 0,68 Teste Hosmer e Lemeshow - p-valor = 0,97						

Tabela 12 – Resultado da análise de regressão logística de discentes dos cursos da UFSM com melhor e pior desempenho em relação à evasão, no período de 2009 a 2015

(continua)

Curso Superior de Tecnologia em Alimentos (CCR)						
Idade	1,05	1,02 - 1,08	0,01*	0,96	0,93 - 0,99	0,02*
Cota						
Ampla Concorrência	1			1		
Escola Pública	0,60	0,41 - 0,88	0,01*	0,47	0,27 - 0,84	0,01*
Outras	0,96	0,50 - 1,82	0,90	0,61	0,25 - 1,47	0,27
Aluno Bolsista						
Não	1			1		
Sim	0,10	0,07 - 0,14	<0,001*	0,64	0,37 - 0,93	0,03*
Participação em projetos						
Não	1			1		
Sim	0,09	0,06 - 0,13	<0,001*	0,46	0,26 - 0,81	0,01*
Frequência RU						
Não	1			1		
Sim	0,09	0,06 - 0,16	<0,001*	0,60	0,30 - 0,95	0,03*
Empréstimos na biblioteca	0,98	0,98 - 0,99	<0,001*	1,01	1,00 - 1,02	0,03*
Disciplinas Aprovadas	0,87	0,85 - 0,88	<0,001*	0,88	0,86 - 0,91	<0,001*
Trancamentos totais	3,39	2,63 - 4,37	<0,001*	1,93	1,45 - 2,57	<0,001*
$R^2 = 0,67$		Teste Hosmer e Lemeshow - p-valor = 0,34				
Medicina (CCS)						
Idade	1,12	1,03 - 1,20	0,01*	1,21	1,03 - 1,43	0,02*
Aluno bolsista						
Não	1			1		
Sim	0,01	0,01 - 0,14	<0,001*	0,01	0,01 - 0,02	<0,001*
Disciplinas reprovadas por frequência	2,48	1,84 - 3,33	<0,001*	3,16	1,59 - 6,26	0,01*
Trancamentos totais	5,59	3,16 - 9,88	<0,001*	2,39	1,02 - 5,87	0,05*
$R^2 = 0,82$		Teste Hosmer e Lemeshow - p-valor = 0,89				

Tabela 12 – Resultado da análise de regressão logística de discentes dos cursos da UFSM com melhor e pior desempenho em relação à evasão, no período de 2009 a 2015

(continua)

Terapia Ocupacional (CCS)						
Período atual						
Últimos Semestres	1			1		
Primeiros Semestres	17,46	9,10 – 61,93	<0,001*	29,63	39,95 – 89,35	<0,001*
Aluno bolsista						
Não	1			1		
Sim	0,02	0,01 - 0,04	<0,001*	0,18	0,08 - 0,38	0,01*
Participação em projetos						
Não	1			1		
Sim	0,01	0,01 - 0,04	<0,001*	0,16	0,04 - 0,67	0,03*
Disciplinas reprovadas por frequência	1,12	1,07 - 1,17	<0,001*	1,19	1,10 - 1,27	<0,001*
Trancamentos totais	4,43	3,15 - 6,22	<0,001*	2,38	1,55 - 3,64	<0,001*
		R ² = 0,77	Teste Hosmer e Lemeshow - p-valor = 0,82			
Direito Diurno (CCSH)						
Período atual						
Últimos Semestres	1			1		
Primeiros Semestres	14,05	5,08 – 35,27	<0,001*	33,53	19,33 – 52,27	0,02*
Disciplinas reprovadas	1,25	1,07 - 1,45	0,01*	1,43	1,02 - 2,18	0,04*
Disciplinas reprovadas por frequência	1,21	1,10 - 1,34	<0,001*	1,27	1,06 - 1,52	0,01*
Trancamentos totais	5,57	2,79 - 11,14	<0,001*	4,48	2,04 – 9,51	<0,001*
		R ² = 0,78	Teste Hosmer e Lemeshow - p-valor = 0,92			

Tabela 12 – Resultado da análise de regressão logística de discentes dos cursos da UFSM com melhor e pior desempenho em relação à evasão, no período de 2009 a 2015

(continua)

Licenciatura em Sociologia (CCSH)						
Cota						
Ampla Concorrência	1			1		
Escola Pública	0,85	0,34 - 2,11	0,72	1,74	0,49 - 6,18	0,40
Outras	0,20	0,05 - 0,78	0,02*	0,13	0,02 - 0,73	0,02*
Frequência no RU						
Não	1			1		
Sim	0,14	0,07 - 0,30	<0,001*	0,11	0,04 - 0,28	<0,001*
Disciplinas aprovadas	0,93	0,90 - 0,95	<0,001*	0,92	0,90 - 0,95	<0,001*
R ² = 0,42 Teste Hosmer e Lemeshow - p-valor = 0,35						
Programa Especial de Graduação de Formação de Professores (CE)						
Período atual						
4°, 5° e 6° Semestres	1			1		
1°, 2° e 3° Semestres	3,09	1,97 - 4,85	<0,001*	24,31	9,72 - 60,82	<0,001*
Disciplinas aprovadas	0,66	0,63 - 0,69	<0,001*	0,65	0,62 - 0,68	<0,001*
Trancamentos totais	11,05	8,47 - 14,40	<0,001*	3,51	1,95 - 6,32	<0,001*
R ² = 0,87 Teste Hosmer e Lemeshow - p-valor = 0,15						
Licenciatura em Educação Especial - Noturno (CE)						
Idade	0,97	0,96 - 0,99	0,23	0,98	0,92 - 0,99	0,05*
Período de ingresso						
1° Semestre	1			1		
2° Semestre	0,37	0,17 - 0,80	0,01*	0,31	0,11 - 0,92	0,04*
Participação em projetos						
Não	1			1		
Sim	0,09	0,04 - 0,16	<0,001*	0,28	0,12 - 0,65	0,01*
Disciplinas aprovadas	0,88	0,85 - 0,90	<0,001*	0,89	0,86 - 0,92	<0,001*
Trancamentos totais	2,11	1,59 - 2,78	<0,001*	1,56	1,16 - 2,10	0,01*
R ² = 0,60 Teste Hosmer e Lemeshow - p-valor = 0,69						

Tabela 12 – Resultado da análise de regressão logística de discentes dos cursos da UFSM com melhor e pior desempenho em relação à evasão, no período de 2009 a 2015

(continua)

Educação Física - Bacharelado (CEFD)						
Etnia						
Branca	1			1		
Não Branca	1,56	0,90 - 2,69	0,11	1,89	1,02 - 3,88	0,05*
Período atual						
Últimos Semestres	1			1		
Primeiros Semestres	16,85	10,25 – 27,71	<0,001*	25,65	13,97 – 47,09	<0,001*
Período de ingresso						
1º Semestre	1			1		
2º Semestre	1,49	1,03 - 2,17	0,04*	1,75	1,11 - 2,75	0,02*
Disciplinas reprovadas	0,89	0,88 - 0,91	<0,001*	1,10	1,03 - 1,17	0,01*
		R ² = 0,59	Teste Hosmer e Lemeshow - p-valor = 0,27			
Dança - Licenciatura (CEFD)						
Período atual						
Últimos Semestres	1			1		
Primeiros Semestres	14,09	7,25 - 23,78	0,03*	18,38	9,59 - 37,89	0,01*
Disciplinas reprovadas por frequência	1,13	0,99 - 1,30	0,07	1,21	1,04 - 1,41	0,02*
Trancamentos totais	3,36	1,43 - 7,89	0,01*	2,95	1,03 - 9,35	0,04*
		R ² = 0,55	Teste Hosmer e Lemeshow - p-valor = 0,46			

Tabela 12 – Resultado da análise de regressão logística de discentes dos cursos da UFSM com melhor e pior desempenho em relação à evasão, no período de 2009 a 2015

	(conclusão)					
Engenharia Civil (CT)						
Idade	1,16	1,08 - 1,25	<0,001*	1,10	1,03 - 1,22	0,04*
Cota						
Ampla Concorrência	1			1		
Escola Pública	0,48	0,23 - 1,01	0,05*	0,24	0,09 - 0,62	0,01*
Outras	2,93	1,37 - 6,26	0,01*	1,11	0,33 - 3,69	0,87
Período de ingresso						
1º Semestre	1			1		
2º Semestre	0,68	0,41 - 1,13	0,14	0,28	0,13 - 0,62	0,01*
Disciplinas aprovadas	0,88	0,86 - 0,91	<0,001*	0,88	0,85 - 0,91	<0,001*
Trancamentos totais	3,32	2,30 - 4,78	<0,001*	1,44	1,01 - 2,14	0,05*
R ² = 0,63 Teste Hosmer e Lemeshow - p-valor = 0,52						
Bacharelado em Sistemas de Informação (CT)						
Cota						
Ampla Concorrência	1			1		
Escola Pública	0,82	0,47 - 0,95	0,01*	0,75	0,34 - 0,93	0,01*
Outras	7,65	2,17 - 26,95	0,46	15,65	1,75 - 39,76	0,25
Período atual						
Últimos Semestres	1			1		
Primeiros Semestres	10,57	5,19 - 23,35	<0,001*	18,20	10,16 - 31,87	<0,001*
Disciplinas reprovadas por frequência	1,06	1,02 - 1,11	0,01*	1,24	1,13 - 1,35	<0,001*
Disciplinas dispensadas por nota	0,67	0,44 - 1,03	0,07	0,50	0,28 - 0,87	0,02*
Trancamentos totais	2,86	2,01 - 4,09	<0,001*	2,21	1,45 - 3,36	<0,001*
R ² = 0,69 Teste Hosmer e Lemeshow - p-valor = 0,76						

Fonte: Elaborado pela autora.

*Significativo ao nível de 5% de significância.

5.2.4 Centro de Ciências da Saúde

O CCS foi o Centro que apresentou o menor percentual de alunos evadidos dentre todos os centros avaliados no estudo, e o curso de Medicina, a menor taxa (2,4%), o mesmo foi observado no estudo realizado por Hoffmann (2016) que teve como objetivo conhecer os principais fatores que influenciam a evasão de alunos nos cursos superiores em uma IES pública. Verificou-se como variável significativa que explica a evasão neste curso, a idade do acadêmico, que apresentou 1,21 vezes o risco de evadir para cada ano a mais de idade. Observou-se, também como fator de risco de evasão, o número de disciplinas reprovadas por frequência e número de trancamentos totais realizados, 3,16 e 2,39 vezes o risco de evadir, respectivamente. 82,0% da evasão é explicada pelas variáveis preditoras do modelo e o p-valor referente ao teste de Hosmer e Lemeshow foi 0,89 indica que o modelo está bem ajustado aos dados.

Verificou-se, como variáveis explicativas para evasão, no modelo ajustado de regressão logística para o curso de Terapia Ocupacional, o período (semestre) atual, e pode-se observar que o risco de evasão nos primeiros semestres é 29,63 vezes o risco de evasão dos últimos semestres do curso. Verifica-se também que o risco diminuiu dos primeiros semestres quando ajustado no modelo de regressão múltipla com as variáveis aluno bolsista, participação em projetos, número de disciplinas reprovadas por frequência e número de trancamentos totais realizados. O indicador de qualidade R^2 (Nagelkerk) de explicação da evasão pelas variáveis preditoras do modelo foi de 77,0%.

A análise de resíduos do modelo ajustado para o curso de Medicina apresentou um ponto influente o qual foi verificado e observou-se que se tratava de um discente que apresentou um número alto de reprovações por frequência, o caso foi eliminado do modelo e testado novamente, mas não foi constatado melhora no ajuste do mesmo. Para os desvios padronizados foram observados dois casos com valores maiores que 3, as observações foram retiradas do modelo, porém não se observou melhora no ajuste.

O modelo ajustado do curso de Terapia Ocupacional também apresentou um ponto influente, verificado através da distância de Cook e 5 casos de valores maiores que 3, que inclusive um deles foi o ponto que apresentou influência no modelo, as observações foram retiradas do modelo, porém o mesmo não apresentou melhora nas estimativas.

5.2.5 Centro de Ciências Sociais e Humanas

Para o CCSH foram analisados os cursos de Direito diurno, o qual apresentou melhor desempenho em relação à evasão (8,0%) e de Licenciatura em Sociologia, que apresentou o maior percentual de evasão (84,9%) entre os cursos considerados.

Verificou-se, como variável explicativa para o modelo ajustado de regressão logística do curso de direito, o período atual, variável que apresentou 33,53 vezes o risco de evadir nos primeiros semestres, comparados com a categoria de referência que são os últimos semestres do curso. Observou-se, assim como no modelo de regressão logística ajustado para o curso de Ciências Biológicas, que a medida da OR ajustada a outras variáveis aumentou em relação à bruta. As variáveis número de disciplinas reprovadas, número de disciplinas reprovadas por frequência e número de trancamentos totais também apresentaram risco de evasão de 1,43, 1,27 e 4,43, respectivamente, para cada disciplina reprovada ou trancamento total realizado.

O modelo ajustado para o curso de Direito apresentou quatro pontos influentes, os quais foram retirados do modelo e o mesmo testado novamente, porém a eliminação desses pontos não apresentou melhora no ajuste. Com relação aos resíduos padronizados, houve dois casos em que o valor foi maior que 3.

Para o modelo do curso de Sociologia não se observou nenhum ponto influente, e em nenhum caso os resíduos padronizados apresentaram valores acima de ± 3 desvios.

5.2.6 Centro de Educação

O Programa Especial de Graduação de Formação de Professores apresentou no menor taxa de evasão dos cursos do CE, 27,2% enquanto que o curso de Licenciatura em Educação Especial apresentou percentual de 38,0%, o maior considerado dentre os cursos no Centro de ensino.

A variável período atual do curso apresentou-se significativa no modelo ajustado de regressão logística e assim como em outros cursos representa risco de evasão nos primeiros 3 semestres quando comparados com os últimos semestres, da categoria de base, neste caso o risco foi de 24,31 vezes a chance de evasão.

O número de trancamentos totais foi significativo nos dois cursos considerados, representando risco de evasão para cada trancamento realizado, 3,51 vezes no curso de Formação de Professores e 1,56 vezes no curso de Educação Especial.

Participar de projetos representou fator de proteção de 0,72 menos chance de evasão e, a cada disciplina aprovada, a chance de evasão é 0,11 vezes menor no curso de Educação Especial.

A análise de resíduos não apresentou nenhum ponto influente nos modelos ajustados dos cursos analisados e em relação aos resíduos padronizados, não houve valores fora do intervalo de ± 3 em ambos os modelos.

5.2.7 Centro de Educação Física e Desportos

Foram estudados os cursos de Educação Física – Bacharelado e Dança – Licenciatura no CEFD. Observou-se que a variável período atual foi significativa no modelo ajustado de regressão logística para ambos os cursos, apresentando 25,65 vezes o risco de evasão dos primeiros semestres do curso de Educação Física e 18,29 vezes o risco nos primeiros semestres do curso de Dança quando comparado com a categoria de referência definida como os últimos semestres do curso, o qual, na análise descritiva, verificou-se baixo percentual de evasão.

Ser de etnia não branca significou risco de evasão de 1,89 quando comparada a ser de etnia branca no curso de Educação Física, assim como ingressar no 2º semestre do ano significou um risco de 1,57 vezes maior que a entrada no 1º semestre. Número de disciplinas reprovadas e de trancamentos totais também apresentaram significância estatística, expressando risco de evasão.

Os dois modelos de regressão logística ajustados apresentaram R^2 (Nagelkerk) com boa explicação da evasão pelas variáveis preditoras, 59,0% e 57,0%, respectivamente para o curso de Educação Física e Dança.

Nenhum ponto influente foi observado no modelo ajustado para o curso de Dança e apenas alguns valores de resíduos padronizados, próximos de 3 mas em nenhum caso, ultrapassando esse valor. O mesmo foi observado no modelo ajustado para o curso de Educação Física, porém em 4 casos houve valores de resíduos padronizados fora do intervalo de ± 3 , os quais, assim como nos casos anteriores, foram retirados no modelo, porém não se observou melhora no ajuste.

5.2.8 Centro de Tecnologia

O curso de Engenharia Civil foi o que apresentou melhor desempenho (11,5%) com relação à evasão entre os cursos do CT e o curso de Bacharelado em Sistemas da Informação o que apresentou o pior desempenho (44,2%), como se pode observar na Tabela 11. Nos modelos ajustados de regressão logística para os dois cursos analisados, observou-se que a variável cota foi significativa para ambos os cursos, porém, para Engenharia Civil, verifica-se que ser de escola pública representa fator de proteção (0,66 menos chance de evasão) em relação à categoria de referência que é ampla concorrência enquanto que no curso de Sistemas de Informação ser de outras cotas apresentou fator de risco (15,65 vezes a chance de evasão) em relação à ampla concorrência.

Estar cursando os primeiros semestres representa, ao discente do curso de Sistemas de Informação, um risco de 18,20 vezes de evasão. Verifica-se também que a OR dessa variável, quando ajustada com as variáveis cota, número de disciplinas aprovadas, número de disciplinas dispensadas por nota e trancamentos totais apresenta um risco maior que o observado na OR bruta.

A variável número de trancamentos totais representou risco de evasão nos dois cursos analisados.

A explicação da evasão pelas variáveis preditoras do R^2 (Nagelkerk) foi de 63,0% e 69,0% e o p-valor do teste de Hosmer e Lemeshow foi $>0,05$, que significa que o modelo está bem ajustado aos dados.

Ao examinar os resíduos do modelo ajustado para o curso de Engenharia Civil, não se observou nenhum ponto influente no modelo que pudesse causar atenção e em todos os casos os valores dos resíduos padronizados estiveram próximos de ± 3 . O mesmo foi verificado no modelo ajustado para o curso de Sistemas de Informação.

5.3 CONSIDERAÇÕES SOBRE OS RESULTADOS ANÁLISE DOS MODELOS DE REGRESSÃO LOGÍSTICA MÚLTIPLA

Os modelos finais de cada curso incluíram variáveis derivadas de diversas áreas tanto em relação ao perfil do discente como de variáveis acadêmicas. Segundo Vitelli,

Rocha e Fritsch (2010), isto indica a existência de uma complexidade de fatores intervenientes no fenômeno da evasão e que os fatores diferem de um aluno para outro.

Conforme Gaioso (2005) e Schargel e Smink (2002), a idade do acadêmico é mencionada como um ponto importante quando se fala em evasão. Segundo os autores, quanto maior a idade, maior o nível de responsabilidade e compreensão sobre as necessidades sociais e profissionais. No presente estudo a variável idade se apresentou significativa no modelo de regressão ajustado para o curso de Matemática, Curso Superior de Tecnologia em Alimentos, Medicina, Educação Especial e Engenharia Civil, ou seja, apresentou-se como fator de risco para evasão em praticamente todos os cursos mencionados.

Com relação à variável gênero, não se observou significância estatística no modelo de regressão logística ajustado em nenhum dos cursos analisados o que pôde ser observado na maioria dos estudos encontrados. Essa variável difere bastante de acordo com a área de conhecimento, e conseqüentemente, o curso escolhido pelo discente. Por exemplo, no CT eram 71,0% de alunos do sexo masculino enquanto que no CCS 75,0% dos alunos eram do sexo feminino.

O estudo de Prim e Fávero (2013), com o objetivo de demonstrar e analisar a evasão de 42 estudantes de uma instituição de ensino superior em cursos tecnológicos no município de Blumenau/SC, no ano de 2012, identificou que 81,0% do total de alunos evadem nos 3 primeiros semestres do curso o que vem de encontro com o verificado no presente estudo, onde observou-se que a variável período atual foi a que apresentou maior risco, chegando à expressar, no curso de Direito, 33,53 vezes o risco de evasão nos primeiros semestres quando comparados aos últimos semestres do curso.

Hoffmann (2016) realizou um estudo de mapeamento de atributos de fatores que influenciam na evasão com relação a causas institucionais e observou que, dentre outras causas, pode-se assinalar a assistência estudantil. Os atributos que ajudam a estudar esta questão são o aluno ter bolsa, frequentar RU e morar na CEU. Com relação às causas individuais, na integração acadêmica, o autor destaca que os atributos que auxiliam no estudo desta questão são a etnia, estado civil, idade no ingresso, entre outras, as mesmas variáveis também foram consideradas no presente estudo, apresentando significância no modelo de regressão logística para explicar a evasão nos cursos considerados.

No estudo de Furtado e Alves (2012), em que o objetivo foi realizar uma análise dos fatores determinantes da evasão universitária na Universidade do Vale do Rio dos

Sinos (UNISINOS), os autores observaram uma relação crescente entre a variável disciplinas reprovadas e a possibilidade de evasão dos alunos da mesma forma que o constatado no presente estudo em que foi verificado risco maior de evasão a cada disciplina reprovada nos cursos de Ciências Biológicas, Matemática, Direito e Educação Física.

Também no estudo de Furtado e Alves (2012), os autores verificaram que a evasão, em média, aumenta com a distância da residência do estudante ao campus, os estudantes moradores de Dois Irmãos, localidade afastada da sede da universidade, apresentam probabilidade maior de evasão do que os alunos que residem em São Leopoldo, cidade em que está localizada a universidade. Observou-se no presente estudo que, no curso de Matemática a variável mora no local (cidade) onde estuda foi significativa no modelo de regressão ajustado apresentando fator de proteção em alunos que moram em Santa Maria.

Silva (2013), com o objetivo de estabelecer estratégias para conter a evasão, avaliou uma instituição privada de ensino, a partir de sua base de dados de acompanhamento discente no período de 2006 a 2009. O autor constatou através de um modelo *duration*, que as variáveis que influenciam na decisão do discente de evadir ou na redução do tempo que os alunos permanecem na instituição são, entre outras, número de reprovações, sexo e idade, assim como foi verificado nos modelos de regressão logística encontrados para explicar a evasão no presente estudo para os cursos avaliados.

Nagai e Cardoso (2017) buscaram comparar as principais razões que provocam a evasão nos cursos de Administração, Ciências Contábeis, Ciências Econômicas e Sistemas da Informação da Universidade Federal de Mato Grosso (UFMT) – Campus Rondonópolis, no período de 2010 a 2014. Os autores verificaram em seu estudo, através de modelos estatísticos de análise fatorial e comparativo entre médias, que entre as razões que provocam a evasão estão fatores relacionados ao curso e à instituição de ensino, e que, segundo Veloso e Almeida (2002) a evasão não é um processo que depende apenas do aluno, mas sim um fenômeno institucional, reflexo da ausência de uma política de permanência do aluno no curso de sua opção. Pode-se perceber, no presente estudo, que variáveis que consideraram a participação dos discentes em projetos e também se o aluno é bolsista foram significativas em alguns cursos, como por exemplo, no Curso Superior de Tecnologia em Alimentos, representando no modelo ajustado, fator de proteção em relação à evasão.

O estudo de Fritsch, Da Rocha e Vitelli (2015), indicou um perfil de discente com algumas características comuns. Quanto mais aprovação o aluno apresenta ao longo de sua trajetória no curso, menor tende a ser seu percentual de evasão. Neste estudo, verificou-se o número de disciplinas aprovadas como fator de proteção para explicar a evasão nos cursos de Letras – Português/Literaturas, Medicina Veterinária, Curso Superior em Tecnologia de Alimentos, Sociologia, Programa Especial de Graduação de Formação de Professores, Educação Especial e Engenharia Civil.

No estudo realizado por Rodrigues (2013), com o objetivo de identificar e apresentar informações e características relativas à evasão no curso de Ciência da Computação da Universidade Federal do Rio Grande do Sul (UFRGS), no período de 2000 a 2013, o autor verificou diferença entre ingressantes no primeiro e segundo semestres do ano, segundo ele, o fato se justifica quando considerado que a diferença entre os ingressantes dos dois períodos é relativa a seus desempenhos no vestibular para ingresso no curso, de acordo com o seguinte critério: os 75 primeiros colocados no concurso vestibular para o curso ingressam no 1º semestre letivo do ano e os 25 restantes no 2º. Assim, ao menos teoricamente, alunos que apresentaram desempenho superior tendem estar mais preparados para o curso, existindo menores chances de desistência por parte desses. Observou-se, no presente estudo, que a variável período de ingresso foi significativa para explicar a evasão nos cursos de Educação Especial, Educação Física e Engenharia Civil.

Na Tabela 13, pode-se observar um resumo das variáveis que apresentaram significância no modelo de regressão logística múltiplo de cada curso analisado no presente estudo.

A variável período atual foi a que apresentou maior risco de evasão nos modelos ajustados em que foi significativa. O número de trancamentos totais foi significativo, representando risco, para explicar a evasão em praticamente todos os cursos analisados. O número de disciplinas aprovadas apresentou-se como fator de proteção em todos os cursos em que a variável foi significativa para explicar a evasão, entre outros.

Tabela 13 - Razões de chance (OR) que apresentaram significância no modelo ajustado de regressão logística de cada curso analisado, no período de 2009 a 2015

Curso/Variáveis	Idade	Etnia	Cota	Período de Ingresso	Período Atual	Mora no local onde estuda	Aluno Bolsista	Participação em Projetos	Frequenta RU	Nº de empréstimos na biblioteca	Nº de disciplinas aprovadas	Nº de disciplinas reprovadas	Nº de reprovações por frequência	Nº de disciplinas dispensadas por nota	Nº de trancamentos totais
Letras – Licenciatura	-	-	-	-	-	-	-	-	0,42	-	0,92	-	1,06	-	1,57
Letras – Bacharelado	-	0,42	-	-	-	-	-	-	-	-	0,92	-	-	-	1,83
Ciências Biológicas	-	-	-	-	28,28	-	-	0,16	-	-	-	1,29	-	-	2,33
Matemática - Licenciatura e Bacharelado	1,13	-	-	-	-	0,31	-	-	-	-	-	4,32	-	-	-
Medicina Veterinária	-	-	-	-	-	-	0,13	0,24	-	0,98	0,93	-	-	-	1,58
Tecnologia em Alimentos	0,96	-	0,47	-	-	-	0,64	0,46	1,01	-	0,88	-	-	-	1,93
Medicina	1,21	-	-	-	-	-	0,01	-	-	-	-	-	3,16	-	2,39
Terapia Ocupacional	-	-	-	-	29,63	-	0,18	0,16	-	-	-	-	1,19	-	2,38
Direito Diurno	-	-	-	-	33,53	-	-	-	-	-	-	1,43	1,27	-	4,43
Licenciatura em Sociologia	-	-	0,13	-	-	-	-	-	0,11	-	0,92	-	-	-	-
PEG	-	-	-	-	24,31	-	-	-	-	-	0,65	-	-	-	3,51
Educação Especial – Licenciatura	0,98	-	-	0,31	-	-	-	0,28	-	-	0,89	-	-	-	1,56
Educação Física – Bacharelado	-	1,89	-	1,75	25,65	-	-	-	-	-	-	1,10	-	-	-
Dança – Licenciatura	-	-	-	-	18,29	-	-	-	-	-	-	-	1,21	-	2,95
Engenharia Civil	1,10	-	0,24	0,28	-	-	-	-	-	-	0,88	-	-	-	1,44
Sistemas de Informação - Bacharelado	-	-	0,75	-	18,20	-	-	-	-	-	-	-	1,24	0,5	2,21

Fonte: Elaborado pela autora.

6 CONCLUSÃO

O trabalho apresentou importantes evidências que qualificam o comportamento discente em uma instituição pública de ensino superior e assim, pôde atender aos objetivos propostos inicialmente, por meio da análise das variáveis selecionadas.

A partir da análise inicial verificou-se o percentual de evasão por Centro de ensino. O CCNE foi o Centro que apresentou o maior percentual (52,0%) entre todos os Centros analisados e o CCS foi o que apresentou menor evasão (11,6%).

Com relação à análise descritiva das variáveis utilizadas no estudo, pôde-se verificar o perfil do discente de cada Centro de ensino, bem como características dos mesmos a respeito de variáveis acadêmicas como número de disciplinas aprovadas, reprovadas, número de empréstimos realizados na biblioteca, participação em projetos, o aluno ter bolsa de estudo, entre outras.

Sobre o perfil do discente, observou-se que, se trata de alunos com idade média entre 20 e 28 anos, o gênero varia de acordo com o centro de ensino analisado, de etnia branca, solteiros, com ingresso na universidade no primeiro semestre do ano, por ampla concorrência e em média 65,0% dos alunos não moram na cidade em que a universidade está localizada.

Verificou-se, de um modo geral que, discentes que estão inseridos em projetos ao longo da graduação evadem menos que alunos não inseridos em tais programas, que não tem bolsa, não frequentam RU ou não moram na CEU.

Por meio dos modelos de regressão logística ajustados aos dados, identificou-se, na maior parte dos casos que, quanto maior a idade do discente, maior o risco de evasão, em alguns cursos, o fato de ser de etnia não branca também representa risco. Com relação às cotas, observou-se que, em geral, ingressar na universidade por cota de escola pública é fator de proteção em relação à ampla concorrência, entretanto, ingressar por outros tipos de cota, representa fator de risco. Os primeiros semestres são os que apresentam maior risco de evasão em todos os cursos em que essa variável apresentou significância estatística no modelo, quando comparados com os últimos semestres da graduação. Quanto ao número de disciplinas aprovadas e dispensadas por nota, percebeu-se que, as variáveis se apresentam como fator de proteção. Já, a cada disciplina reprovada, reprovada por frequência, trancamento total ou parcial realizado, o risco de evasão aumentou.

A qualidade dos modelos ajustados de regressão logística para cada curso com melhor e pior desempenho em relação à evasão dentro de cada Centro de ensino foi verificada pelo teste de Hosmer e Lemeshow e pelo valor de R^2 (Nagelkerke), onde se pôde observar que, a evasão foi explicada pelas variáveis preditoras utilizadas.

Por meio dos modelos ajustados foi possível verificar quais as principais causas de evasão e também as variáveis que mais contribuem e que apresentam maior risco para tal evento e assim, implementar ações que possam minimizar os índices de evasão na instituição estudada e em outras IES.

Espera-se que o estudo apresentado possa ser adaptado à realidade da instituição de ensino analisada e de outras instituições e utilizado por alunos, professores e coordenadores de cursos para um melhor entendimento da evasão e para a criação de políticas de combate ao fenômeno no ensino superior, de maneira que os altos índices do problema sejam combatidos.

Sugerem-se, como estudos futuros avaliar cursos de cada centro com particularidades em comum de forma a identificar novas relações, padrões, características e conhecimentos inerentes a cada área e assim, traçar estratégias específicas de combate à evasão.

REFERÊNCIAS

- ABBAD, G.; CARVALHO, R. S.; ZERBINI. A evasão em cursos via internet: explorando variáveis explicativas. **ERA eletrônica**, v. 5, n. 2, Art. 17, jul./dez. 2006.
- AGRESTI, A. **Categorical Data Analysis**, Second Edition. New York: Wiley. 2002.
- AGRESTI, A. **Introduction to the Analysis of Categorical Data**. New York: Wiley. 1996.
- AKAIKE. H. **A new look at statistical model identification**. IEEE Transactions on Automatic Control AU-19716-722. 1974.
- ARAQUE, F.; ROLDÁN, C.; SALGUERO, A. **Factors influencing university dropout rates**. Computers & Education. v. 53(3), 563–574, 2009.
- ARULAMPALAM, W.; NAYLOR, R. A.; SMITH, J. P. **Effects of in-class variation and student rank on the probability of withdrawal: cross-section and time-series analysis for UK university students**. Economics of Education Review, 24(3), 251–262, 2005.
- ASSIS, C. F. DE. **Estudo dos fatores que influenciam a evasão dos alunos nos Cursos Superiores de Tecnologia de uma Instituição de Ensino Superior Privada**. 2013. 105 p. Dissertação (Mestrado Profissional em Administração) - Fundação Cultural Dr. Pedro Leopoldo – FPL, Pedro Leopoldo, MG, 2013.
- BAGGI, C. A. S.; LOPES, D. A. **Evasão e Avaliação Institucional no Ensino Superior: uma discussão bibliográfica**. Avaliação, v. 16, n. 2, p. 355-374, São Paulo, 2011.
- BARDAGI, M. P. **Evasão e Comportamento Vocacional de Universitários: estudos sobre o desenvolvimento de carreira na graduação**. 2007. 242 f. Tese (Doutorado em Psicologia) - Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 2007.
- BRAGA, M. M.; PEIXOTO, M. C.; BOGUTCHI, T. F. **A Evasão no Ensino Superior Brasileiro: o caso da UFMG**. Avaliação: Revista da Avaliação da Educação Superior, [S.l.], v. 8, n. 3, 2003.
- BRESLOW, N. E.; DAY, N. E. **Statistical Methods in Cancer Research. v.1. The Analysis of Case-Control Studies**. Lyon: IARC, 1980.
- CABRERA, A.; NORA, A.; CASAÑEDA, M. **The role of finances in the persistence process: a structural model**. Research in Higher Education, 33, 571-594, 1992.
- CALLEGARI-JACQUES, S. M. **Bioestatística Princípios e Aplicações**. Porto Alegre: Artmed, 2003.
- COMISSÃO ESPECIAL DE ESTUDOS SOBRE A EVASÃO NAS UNIVERSIDADES PÚBLICAS BRASILEIRAS. **Diplomação, Retenção e Evasão nos Cursos de Graduação em Instituições de Ensino Superior Públicas**. ANDIFES/ABRUEM/SESu/MEC, 1996. Disponível em:

http://www.andifes.org.br/wpcontent/files_flutter/Diplomacao_Retencao_Evasao_Graduacao_em_IES_Publicas-1996.pdf. Acesso em 10 de janeiro de 2017.

CORNFIELD, J. **A statistical problem arising from retrospective studies**. In: Proceedings of the Third Berkeley Symposium, Berkeley, University of California Press, pgs.133-148, 1956.

CORNFIELD, J. **Joint dependence of the risk of coronary heart disease on serum cholesterol and systolic blood pressure: A discriminant function analysis**. Federation Proceedings, 21, 58-61, 1962.

COX, D.R. **Regression Models and Life Tables**. Journal Royal Statistical Society, Series B (Methodological), v. 34, n. 2, 187-220, 1972.

COX, D.R.; SNELL, E.J. **Analysis of Binary Data**. Second Edition. Chapman & Hall/CRC. 1989.

DEKE, J.A. **Study of the impact of public school spending on post secondary educational attainment using statewide school district refinancing in Kansas**. Economics of Education Review, v. 22, n. 3, p. 275-284, Jun. 2003.

EFRON, B. **The efficiency of logistic regression compared to normal discriminant function analysis**. Journal of the American Statistical Association, 70, 892-898, 1975.

FIELD, A. **Descobrimos a Estatística usando o SPSS**. 2. Ed. Porto Alegre: Artmed, 2009. 688 p.

FRISCHENBRUDER, S. L. **O desenvolvimento vocacional na adolescência: Autoconceito e comportamento exploratório**. 1999. Dissertação (não publicada) (Mestrado em Psicologia Social e da Personalidade) Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, RS, 1999.

FRITSCH, R.; DA ROCHA, C. S.; VITELLI, R.F. **A evasão nos cursos de graduação de uma instituição de ensino superior privada**. Revista Educação em Questão, Natal, v. 52, n. 38, p. 81-108, maio/ago, 2015.

FURTADO, V. V. A.; ALVES, T. W. **Fatores determinantes da evasão universitária: Uma análise com alunos da UNISINOS**. Contextus. Revista Contemporânea de Economia e Gestão. v. 10, n. 2, jul/dez 2012.

GAIOSO, N. P. L. **A evasão discente na educação superior no Brasil: na perspectiva de alunos e dirigentes**. 2005. 95p. Dissertação (Mestrado em Educação) - Universidade Católica de Brasília, Brasília, DF, 2005.

GRIZZLE, J. E.; STARMER, C. F.; KOCH, G. G. **Analysis of Categorical Data by linear models**. Biometrics, 25 (3), 489-504, 1969.

GÓMES, R. D.; FEIXAS, M. GAIRÍN, J. MUÑOZ, J.L. **Understanding Catalan University dropout from a comparative approach**. *Procedia - Social and Behavioral Sciences*. 46, 1424 – 1429, 2012.

HOFFMANN, I. L. **Metodologia para identificação de fatores estratégicos para acompanhamento sistemático da evasão em cursos de graduação**. 2016. Dissertação (Mestrado em Engenharia de Produção) Universidade Federal de Santa Maria, Santa Maria, RS, 2016.

HOSMER, D.; LEMESHOW, S. **Applied Logistic Regression**. New York: John Wiley and Sons, 1989.

HOSMER, T.; HOSMER, D. W.; FISHER, L. L. **A comparison of the maximum likelihood and discriminant function estimators of the coefficients of the logistic model for mixed continuous and discrete variables**. *Communications in Statistics*, B12, 577-593, 1983.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Sinopse Estatística da Educação Superior 2012**. Brasília: Inep, 2016. Disponível em: <http://portal.inep.gov.br/web/guest/sinopses-estatisticas-da-educacao-superior>. Acesso em: 24 de novembro de 2016.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Sinopse Estatística da Educação Superior 2015**. Brasília: Inep, 2016. Disponível em: <http://portal.inep.gov.br/superior-censosuperior-sinopse>. Acesso em: 31 de dezembro de 2016.

KLEINBAUM, D. G; KUPPER, L. L.; MORGENSTERN, H. **Epidemiologic Research**. Belmont, California: Lifetime Learning Publications, 1982.

LEHR, C. A.; JOHNSON, D. R.; BREMER, C. D.; COSIO, A.; THOMPSON, M. **Increasing Rates of School Completions: Moving From Policy and Research to Practice. A Manual for Policymakers, Administrators, and Educators**. University of Minnesota, 2004.

MCCULLAGH, P.; NELDER, J.A; **Generalized Linear Models**. London: Chapman & Hall, 2º Edição, 1989.

MEC/SESu. Comissão Especial de Estudos sobre a Evasão nas Universidades Públicas Brasileiras. Brasília: ANDIFES/ABRUEM/SESu/MEC. 1997.

MERCURI, E.; FIOR C. A. **Análise dos fatores preditivos da evasão em uma universidade confessional**. *Congressos CLABES*. 2016.

MOROSIN, M. C.; CASARTELLI, A. O.; SILVA, A. C. B.; SANTOS, B. S.; SCHMITT, R. E.; GESSINGER, R. M. **A evasão na Educação Superior no Brasil: uma análise da produção de conhecimento nos periódicos *Qualis* entre 2000-2011**. Porto Alegre/RS – Brasil. Faculdade de Educação – FACED. Pontifícia Universidade Católica do Rio Grande do Sul – PUCRS. P. 1-10, 2011.

NAGAI, N. P.; CARDOSO, A. L. J. **A evasão universitária: Uma análise além dos números.** Revista Estudo & Debate, Lajeado, v. 24, n. 1, p. 193-215, 2017.

NAGELKERKE, N. J. D. **A note on general definition of the coeficiente determination.** Biometrika. v. 78, p. 691 – 692. 1991.

NETO, O. A. P.; CRUZ, F.; PFITSCHER, E.D. **Utilização de metas de desempenho ligadas à taxa de evasão escolar nas universidades públicas.** Revista de Educação e Pesquisa em Contabilidade – REPEC. Brasília, n. 2, v. 2, art. 4, p. 54-74.

NORA LUNA, M. Ed. **Understanding graduation e Dropout Rate Calculations.** University of Nevada Cooperative Extension, 2013.

PAGANO, M.; GAUVREAU, K. **Princípios de Bioestatística.** São Paulo, Pioneira Thomson Learning, 2004.

PAULA, G. A. **Modelos de regressão com apoio computacional.** São Paulo: Instituto de Matemática e Estatística, Universidade de São Paulo, Dezembro, 2010.

PAURA, L.; ARHIPOVA, I. **Cause Analysis of students dropout rate in higher education study program.** Procedia - Social and Behavioral Sciences, 109, 1282-1286, 2014.

PAZ, F.J.; CAZELLA, S. C.; Identificando o perfil de evasão de alunos de graduação através da Mineração de dados Educacionais: um estudo de caso de uma Universidade Comunitária. In: VI CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 10. 2017, Recife. Anais dos Workshops do VI Congresso Brasileiro de Informática na Educação, Recife: WCBIE, 2017. Disponível em: <http://www.br-ie.org/pub/index.php/wcbie/article/view/7448/5244>. Acesso em 24 nov. 2017.

POLYDORO, S. A. J. **O Trancamento de Matrícula na Trajetória Acadêmica do Universitário: condições de saída e retorno à Instituição.** Tese (Doutorado em Educação). Faculdade de Educação – Universidade Estadual de Campinas. São Paulo, 2000.

PORTAL EDUCAÇÃO. **Educação e Pedagogia.** 2013. Disponível em: <https://www.portaleducacao.com.br/pedagogia/artigos/48642/entenda-a-diferenca-entre-universidade-faculdade-e-centro-universitario>. Acesso em: 02 de janeiro de 2017.

PRESS, S. J.; WILSON, S. **Choosing Between Logistic Regression and Discriminant Analysis.** Journal of the American Statistical Association, v. 73, n. 364, p. 699-705, 1978.

PRIM, L. A.; FÁVERO, J. D. **Motivos da evasão escolar nos cursos de ensino superior de uma faculdade na cidade de Blumenau.** Tecnologias para Competitividade Industrial, Florianópolis, n. Especial Educação, p. 53-72, 2013/2.

RODRIGUES, F. S. **Estudo sobre a Evasão no Curso de Ciência da Computação da UFRGS.** 2013. 87 p. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação) – Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 2013.

SADLER, W. E.; COHEN, F. L.; KOCKESEN, L. **Factors Affecting Retention Behavior: A model to predict at-risk students**. AIR 1997 Annual Forum Paper.1997-05-00. 22p.; Paper presented at the Annual Forum of the Association for Institutional Research (37th, Orlando, FL, May 18-21,1997).

SCHARGEL, F. P; SMINK, J. **Estratégias para Auxiliar o Problema de Evasão Escolar**. Rio de Janeiro: Dunya, 2002.

SILVA FILHO, R. L. L.; HIPÓLITO, O. **Financiamento e expansão do ensino superior. 2009**. Disponível em: <http://robertolobo.com.br/index.php/2009/06/financiamento-e-expansao-do-ensino-superior-brasileiro/>. Acesso em: 03 de janeiro de 2017.

SILVA FILHO, R. L. L.; LOBO, M. B. de C. M. **Esclarecimentos Metodológicos sobre os Cálculos de Evasão**. Instituto Lobo, 2012. Disponível em:http://www.institutolobo.org.br/imagens/pdf/artigos/art_078.pdf. Acesso em 11 de janeiro de 2017.

SILVA FILHO, R. L. L.; MOTEJUNAS, P. R., HIPOLITO, O.; LOBO, M. B. C. M. **A evasão no ensino superior brasileiro**. Cad. Pesqui., São Paulo, v.37, n.132, p.641-659, 2007.

SILVA, G. P. **Análise de evasão no ensino superior: uma proposta de diagnóstico de seus determinantes**. Revista Avaliação, Campinas; Sorocaba, SP, v. 18, n. 2, p. 311-333, jul. 2013.

SOUZA, E.C. **Análise da Influência Local no Modelo de Regressão Logística**. 2006. 102 p. Dissertação (Mestrado em Estatística e Experimentação Agrônômica) – Universidade de São Paulo, Escola Superior de Agricultura “Luiz de Queiroz”, São Paulo, 2006.

SPADY, W. **Dropouts from higher education: An interdisciplinary review and synthesis**. Interchange, 1, 64-85, 1970.

THEÓPHILO, C.R; MORAES, J.O. **Evasão no Ensino Superior: Estudo dos fatores causadores da evasão no Curso de Ciências Contábeis da Universidade Estadual de Montes Claros – UNIMONTES**. Montes Claros: UNIMONTES, 2005.

TINTO, V. **Dropout from higher education: a theoretical synthesis of recent research**. Review of Educational Research, 45: 89-125. 1975.

UNIVERSIDADE FEDERAL DE SANTA MARIA. 2016. **Centro de Processamento de Dados- Sistema de Informações Educacionais (SIE)**, Santa Maria, 2017. Disponível em <http://www.cpd.ufsm.br/>. Acesso em 30 nov. 2016.

UNIVERSIDADE FEDERAL DE SANTA MARIA. 2016. Disponível em: <http://w3.ufsm.br/frederico/index.php/institucional/historico>. Acesso em: 03 de janeiro de 2017.

VELOSO, T. C. M. A.; ALMEIDA, E. P. **Evasão nos cursos de graduação da Universidade Federal de Mato Grosso, campus universitário de Cuiabá: Um processo de exclusão.** Cuiabá: UFMT. 2002.

VITELLI, R.F; ROCHA, C. S.; FRITSCH, R. **Estudo sobre evasão nos cursos de graduação de uma instituição de ensino superior privada: aplicação de regressão logística.** Produção do edital 38/2010, Programa Observatório de Educação INEP/CAPES, Núcleo em Rede, Projeto nº 44, Indicadores de Qualidade e Gestão Democrática.

VIEGAS, V.A.F.; ALVES, T.W. **Fatores determinantes da evasão universitária: uma análise com alunos da UNISINOS.** Revista Contemporânea de Economia e Gestão, Vale do Rio dos Sinos, v. 10, n. 2, p 115-129, 2012.

ZAGO, N. **Do acesso a permanência no ensino superior: percursos de estudantes universitários de camadas populares.** Revista Brasileira de Educação, Rio de Janeiro, v. 11, n. 32, p. 226-237, 2006.

WILLIAMS, D. A. **Residuals in generalized linear models.** In: *Proceedings of the 12th. Internacional Biometrics Conference*, Tokyo, pp.58 – 64, 1984.