

A predição da formatura e da evasão em uma universidade pública a partir de um modelo logístico

The prediction of graduation and dropout in a public university from a logistic model

Chaiane de Medeiros Rosa

Técnica em Assuntos Educacionais da Universidade Federal de Goiás. Goiânia, Goiás, Brasil.
chaianemr@hotmail.com - <http://orcid.org/0000-0001-8609-3487>

Fabiano Fortunato Teixeira dos Santos

Professor doutor da Universidade Federal de Goiás. Goiânia, Goiás, Brasil.
fortunato@ufg.br - <http://orcid.org/0000-0002-4455-8175>

Eder Angelo Milani

Professor doutor da Universidade Federal de Goiás. Goiânia, Goiás, Brasil.
edermilani@ufg.br - <http://orcid.org/0000-0001-5533-6693>

Recebido em 27 de março de 2019

Aprovado em 30 de julho de 2021

Publicado em 30 de setembro de 2021

RESUMO

O tema deste estudo é sobre os percursos terminativos da educação superior, sendo a evasão e a formatura. Tendo isso em vista, objetiva-se construir modelos logísticos preditivos desses dois desfechos em relação aos estudantes que ingressaram na Universidade Federal de Goiás em 2010, em cursos na modalidade presencial, os quais foram acompanhados por um período de oito anos. Para isso, foi feita consulta de microdados do Censo da Educação Superior Brasileira, de 2010 a 2017, publicados pelo Instituto Nacional de Estudos e Pesquisas Educacionais “Anísio Teixeira”. Para entender esses dois fenômenos - formatura e evasão - recorreu-se à técnica estatística da regressão logística e, além disso, foi feita pesquisa bibliográfica, em consulta a autores que fundamentam a técnica. Como resultado, verificou-se que a técnica estatística permite identificar variáveis que são fatores de risco ou protetivos para a formatura e a evasão; comparar categorias de uma mesma característica por meio da razão de chances; bem como avaliar a probabilidade de qualquer aluno se formar ou evadir. Trata-se, portanto, de uma ferramenta fundamental para a tomada de decisão dos gestores educacionais.

Palavras-chave: Educação superior; Evasão; Formatura.

ABSTRACT

The theme of this study is the terminative courses of higher education, which are dropout and graduation. With this in mind, the objective is to build predictive logistic models of these two outcomes in relation to students who entered the Federal University of Goiás in 2010, in courses in classroom mode, who were followed for a period of eight years. To this end, consultation was made with microdata from the Brazilian Higher Education Census, from 2010 to 2017, published by the National Institute of Educational Studies and Research "Anísio Teixeira". In order to understand these two phenomena - graduation and dropout - the statistical technique of logistic regression was used and, in addition, bibliographic research was carried out, in consultation with authors who support the technique. As a result, it was found that the statistical technique allows the identification of variables that are risk or protective factors for graduation and dropout; compare categories of the same characteristic using the odds ratio; as well as assessing the likelihood of any student graduating or dropping out. It is, therefore, a fundamental tool for decision-making by educational managers.

Keywords: High education; Evasion; Graduation.

Introdução

Neste estudo, parte-se do entendimento de que os processos educacionais formais se organizam em percursos formativos. No caso da educação superior, é possível acompanhar o itinerário formativo dos estudantes considerando a permanência, a desistência (evasão) e a conclusão (formatura). As duas últimas são condições terminativas em relação ao percurso, e representam, respectivamente, insucesso e sucesso (BRASIL, 2017), e por isso serão os desfechos aqui analisados.

No Brasil, a taxa de conclusão da educação superior em 2017, dentro da duração teórica, que consiste no período regular previsto para integralização, foi de 33%, número que coloca país em posição de inferioridade em relação à média dos países membros da Organização para a Cooperação e Desenvolvimento Econômico (OCDE)¹, que foi de 41%. E com prorrogação de três anos a mais para integralização do curso, chegou-se a uma taxa de conclusão de 50%, o que sinaliza que grande parte das trajetórias formativas estão sendo interrompidas (OCDE, 2019). No mais, dados do Censo da Educação Superior referentes aos ingressantes de 2010, nas

instituições de ensino do país, mostraram que a taxa de conclusão desses estudantes foi inferior à taxa de desistência durante todo o período observado, até 2016 (BRASIL, 2019).

Isso posto, este trabalho tem como objetivo construir modelos logísticos preditivos da evasão e da formatura dos estudantes que ingressaram na Universidade Federal de Goiás (UFG) em 2010, em cursos na modalidade presencial. Para isso, foi feita consulta de microdados do Censo da Educação Superior Brasileira, de 2010 a 2017, publicados pelo Instituto Nacional de Estudos e Pesquisas Educacionais “Anísio Teixeira” (Inep). Os dados objeto de análise são referentes aos 5791 alunos ingressantes na UFG em 2010, dos quais 40 foram excluídos do estudo por motivo de inconsistência, por não apresentarem informação em pelo menos uma das covariáveis, quando analisados de maneira longitudinal.

O período de acompanhamento dos mesmos foi até 2017, o que significa um tempo de observação de oito anos. Essa escolha foi feita tendo em vista que o documento intitulado “Metodologia de cálculo dos indicadores de fluxo da educação superior” aponta que o *prazo máximo* de acompanhamento da trajetória escolar do aluno no curso corresponde a duas vezes o tempo mínimo de integralização (BRASIL, 2017, grifo nosso). Logo, optou-se por fazer um estudo longitudinal de oito anos.

Os eventos de interesse deste estudo, também chamados de desfechos, são a formatura, que consiste na conclusão do curso, e a evasão, que significa a saída do curso anteriormente à sua conclusão, sendo esta subdividida em três categorias: evasão da UFG, evasão do curso e evasão total, que considera a união dos conjuntos formados pelos alunos que evadiram do curso e dos que evadiram da instituição. Portanto, dos 5751 alunos acompanhados até 2017, 3199 (55,6%) formaram-se, 245 (4,3%) evadiram-se do curso, 2055 (35,7%) evadiram-se da UFG, e o total de evadidos foi 2300 (40,0%).

Para entender esses dois fenômenos - formatura e evasão - recorreu-se à regressão logística, que se trata de um modelo muito eficaz para analisar dados, por meio do qual pesquisadores explicam uma variável dependente em função de múltiplas variáveis independentes (LEWIS-BECK, 2001). E, além da análise dos dados, esse estudo teve amparo em pesquisa bibliográfica, em consulta a autores que

ISSN: 1984-6444 | <http://dx.doi.org/10.5902/1984644443159>

fundamentam a técnica da regressão logística, sendo eles Hosmer Junior e Lemeshow (2000), Lewis-Beck (2001) e Hair Júnior et al. (2009).

É preciso ressaltar que a regressão logística é utilizada para construir um modelo baseado em um conjunto de informações, que pode ser univariado ou múltiplo, e que permite estimar a probabilidade de uma variável categórica binária assumir seus valores, a qual será denominada de variável resposta.

Essa técnica pode ser utilizada em várias áreas do conhecimento, como Economia e Finanças, Saúde, Geografia, Arqueologia e outros. No caso da Educação, a técnica vem sendo empregada em estudos sobre: desempenho acadêmico (CUNHA; CORNACHIONE JUNIOR; MARTINS, 2008; SILVA FILHO, 2017); proficiência em áreas do conhecimento (MAZULO, 2015); avaliação de programas e políticas públicas (MESQUITA, 2014); retenção acadêmica (PEREIRA et al., 2015); e outros, e tem se destacado nos estudos sobre predição da evasão escolar, na educação básica e superior (FRITSCH et al., 2015; GONZALES; NASCIMENTO; LEITE, 2016; SOUSA et al., 2018; DAVID; CHAYM, 2019).

Hosmer e Lemeshow (2000) chamam a atenção para o fato de que,

Antes de iniciar um estudo de regressão logística, é importante entender que o objetivo de uma análise usando esse método é o mesmo de qualquer técnica de construção de modelos usada em estatística: encontrar o modelo mais adequado e parcimonioso, mas biologicamente razoável para descrever a relação entre uma variável de resultado (dependente ou resposta) e um conjunto de variáveis independentes (preditoras ou explicativas) (HOSMER JUNIOR; LEMESHOW, 2000, p. 1, tradução nossa).

É necessário esclarecer que, quando uma variável dicotômica assume um de seus dois valores, significa uma representação do evento de interesse, que, nesse caso, são dois: a formatura e a evasão. Sendo assim, a regressão logística permite selecionar variáveis relevantes para os desfechos de interesse desse estudo, chamadas de covariáveis; estimar a razão de chances entre duas categorias de uma mesma covariável; bem como calcular a probabilidade de ocorrer a formatura ou a evasão de um aluno em particular.

Tendo isso em vista, esse estudo se encontra organizado em cinco tópicos, exibidos na sequência. O primeiro mostra a regressão logística como técnica para

predição da formatura e da evasão; os três tópicos que seguem propõem modelos de predição para os formados, da evasão da UFG e da evasão total, e o último apresenta a predição da evasão por área do conhecimento como uma possibilidade.

A regressão logística como técnica para predição da formatura e da evasão

Nesta pesquisa, o passo inicial para aplicação da técnica de regressão logística consistiu na seleção das covariáveis consideradas relevantes para este estudo, quais sejam: grau acadêmico (bacharelado ou licenciatura), sexo (masculino ou feminino), turno (matutino, vespertino, noturno ou integral), procedência (escola pública ou escola privada), bolsa de assistência estudantil (recebeu ou não), bolsa de atividade pedagógica (recebeu ou não)², idade em 2010 (até 24 anos, de 25 a 50 anos ou acima de 50 anos), participação em atividade complementar (participou ou não) e área do conhecimento³ (Ciências Exatas e da Terra, Ciências Biológicas, Engenharias, Ciências da Saúde, Ciências Agrárias, Ciências Sociais Aplicadas, Ciências Humanas ou Linguística, Letras e Artes).

É preciso demarcar que, do conjunto de covariáveis, três são relativas às características pessoais e fatores anteriores ao ingresso - sexo, idade em 2010 e procedência - e seis são relativas ao ambiente institucional, a fatores internos à instituição - grau acadêmico, turno, bolsa assistência estudantil, bolsa atividade pedagógica, atividade complementar e área do conhecimento.

No Quadro 1, estão apresentadas as covariáveis e suas categorias, acompanhadas do número que as identificam.

Quadro 1 – Covariáveis e suas categorias

COVARIÁVEL	CATEGORIAS
Grau acadêmico	Bacharelado (0) Licenciatura (1)
Sexo	Feminino (0) Masculino (1)
Turno	Matutino (0) Vespertino (1) Noturno (2) Integral (3)
Procedência	Não declarou (0) Escola pública (1) Escola privada (2)
Bolsa assistência estudantil	Não (0) Sim (1)
Bolsa atividade pedagógica	Não (0) Sim (1)
Idade em 2010	Até 24 anos (0) De 25 a 50 anos (1) Mais que 50 anos (2)
Participação em atividade complementar	Não (0) Sim (1)
Área do conhecimento	Ciências Exatas e da Terra (0) Ciências Biológicas (1) Engenharias (2) Ciências da Saúde (3) Ciências Agrárias (4) Ciências Sociais aplicadas (5) Ciências Humanas (6) Linguística, Letras e Artes (7)

Fonte: Brasil/MEC/Inep (2010-2017). Elaborado pelos autores (2020).

A análise descritiva dos dados mostrou o seguinte perfil dos ingressantes em 2010 na UFG:

- Em relação ao grau acadêmico, 69,1% eram do bacharelado e 30,9% da licenciatura.
- Quanto à área do conhecimento, 18,6% eram das Ciências Humanas, 16,7% das Ciências Sociais Aplicadas, 16,4% de Ciências Exatas e da Terra, 14,0% das Ciências da Saúde, 11,1% das Ciências Agrárias, 10,8% da Linguística, Letras e Artes, 7,9% das Engenharias e 4,5% das Ciências Biológicas.

ISSN: 1984-6444 | <http://dx.doi.org/10.5902/1984644443159>

- No quesito turno, 41,1% dos alunos estavam matriculados no período integral, 33,9% no noturno, 19,0% no matutino e 6,0% no vespertino.
- No tocante ao sexo, 53,8% dos estudantes eram do sexo feminino e 46,2% do masculino.
- Em termos de origem escolar, 69,2% eram procedentes de escola privada, 24,1% de escola pública e 6,7% não informou.
- Em referência à covariável idade, 81,6% dos estudantes tinham até 24 anos, 17,9% de 25 a 50 anos e 0,5% acima de 50 anos.
- Sobre os auxílios recebidos pelos alunos no período em estudo, constatou-se que 26,5% foram contemplados com algum tipo de assistência estudantil e 28,3% receberam bolsa relacionada à participação em atividade pedagógica. Ainda, 21,3% participaram de atividade complementar.

Na construção dos modelos de regressão logística para cada um dos quatro desfechos inicialmente analisados - formatura, evasão da UFG, evasão do curso e evasão total - optou-se pela utilização do método *Enter* do *software IBM SPSS Statistics 20* e nível de significância de 5,0%. Como etapa preliminar, foi verificada se havia multicolinearidade entre as covariáveis, que consiste no efeito combinado de duas ou mais covariáveis, e que pode impedir a obtenção de estimativas únicas dos coeficientes da regressão.

Como critério para concluir se é possível rejeitar a hipótese de multicolinearidade, seguiu-se a seguinte orientação de Hair Júnior et al. (2009): se a tolerância para cada covariável for maior do que 0,1, considera-se que a multicolinearidade não é relevante e pode ser desprezada. É preciso esclarecer que tolerância é a quantia de variabilidade da covariável selecionada não explicada pelas outras covariáveis, e indica se há uma relação linear forte entre elas (HAIR JÚNIOR et al., 2009).

As tolerâncias das covariáveis do estudo estão apresentadas no Quadro 2:

Quadro 2 - Tolerância das covariáveis

COVARIÁVEL	TOLERÂNCIA
Área do conhecimento	0,84
Grau acadêmico	0,82
Sexo	0,93
Idade	0,85
Turno	0,82
Procedência	0,93
Bolsa assistência estudantil	0,94
Atividade complementar	0,74
Bolsa atividade pedagógica	0,75

Fonte: Brasil/MEC/Inep (2010-2017). Elaborado pelos autores (2020).

Nenhuma medida de tolerância foi menor do que 0,1, nem mesmo menor que 0,82. Diante disso, pode-se descartar a multicolinearidade e não será excluída nenhuma covariável da próxima etapa.

O passo seguinte também é um crivo de seleção de covariáveis, sendo que a lógica é escolher aquelas que forneçam um bom modelo logístico univariado na presença da variável resposta em questão. O critério de seleção segue o sugerido por Hosmer Junior e Lemeshow (2000), que dizem que as covariáveis que apresentarem um valor p menor que 0,25 no modelo logístico univariado, são candidatas a fazer parte do modelo logístico múltiplo, juntamente com aquelas que têm importância conhecida no estudo em curso.

Levando em consideração cada desfecho, os modelos logísticos univariados construídos geraram os seguintes resultados. Para os desfechos formados, evadidos da UFG e total de evadidos, a única covariável que apresentou um valor p maior que 0,25 em alguma de suas categorias foi procedência, cujos valores foram, respectivamente, 0,817; 0,329 e 0,488, todos eles na categoria escola pública. Como apenas uma de suas categorias não foi significativa, essa covariável participará do modelo multivariado. No entanto, mesmo que todas as suas categorias tivessem se mostrado não significativas, essa covariável comporia o modelo, dada sua reconhecida importância.

Para o desfecho evadidos do curso, a covariável idade em 2010 apresentou valor p maior do que 0,25 em todas as suas categorias: 0,581 em até 24 anos, 0,308

em de 25 a 50 anos e 0,853 em mais que 50 anos. A covariável procedência também apresentou valor p maior do que 0,25 em todas as suas categorias: 0,552 em não declarou, 0,538 em escola pública e 0,422 em escola privada. Diante disso, essas duas covariáveis não participam do modelo multivariado. Já a covariável área do conhecimento apresentou valor p maior que 0,25 em apenas duas de suas categorias: 0,421 em Ciências Biológicas e 0,816 em Ciências Humanas, portanto, será mantida no modelo multivariado. Todas as demais covariáveis presentes no Quadro 1 apresentaram valor p menor do que 0,25 em todas as suas categorias e também participarão do modelo multivariado.

No mais, para cada um dos desfechos de interesse nesse estudo foi construído um modelo de regressão logística múltiplo utilizando as covariáveis selecionadas anteriormente. Eles permitem caracterizar, por meio do cálculo de probabilidades, se cada aluno é possível formado ou evadido, o que se dá quando o resultado dessa conta é maior que 50%. E a avaliação da qualidade do ajuste e precisão preditiva de um modelo de regressão logística é baseada no teste de Hosmer e Lemeshov, como se verá nas seções que seguem.

Um modelo de predição para os formados

No modelo final para a variável resposta formados, o teste de Hosmer e Lemeshov mostrou que não existem diferenças significativas entre os resultados preditos e os observados, com valor p igual a 0,489. Esse teste mede a acurácia do modelo, isto é, verifica se existem diferenças significativas entre as classificações realizadas pelo modelo e a realidade observada. Em termos de acurácia, o modelo foi capaz de classificar corretamente 76,4% dos formados e 61,3% dos não formados, resultando em uma precisão geral de 69,7%.

A exponenciação dos coeficientes logísticos significantes, chamada de *odds ratio* ou razão de chances, representa a chance de um evento ocorrer dada uma exposição particular, comparada com a chance do evento ocorrer na ausência da exposição. Um *odds ratio* maior que 1, indica que a exposição aumenta a chance da ocorrência do evento de interesse; um *odds ratio* menor do que 1, indica que a

exposição diminui a chance da ocorrência do evento de interesse; e um *odds ratio* igual a 1, indica que a exposição não afeta a chance do evento de interesse ocorrer.

O Quadro 3 apresenta as covariáveis presentes no modelo logístico, os coeficientes logísticos, seu nível de significância (valor p), os coeficientes logísticos exponenciados (Exp(B)) e o intervalo de confiança (I.C.) de 95% para os valores de Exp(B).

Quadro 3 - Covariáveis, coeficientes, odds ratio e IC para o modelo associado à formatura

COVARIÁVEL	CATEGORIAS	B	Valor p	Exp(B)	IC para Exp(B)
Grau acadêmico	Bacharelado				
	Licenciatura	-0,141	0,107	0,869	(0,732;1,031)
Sexo	Feminino				
	Masculino	-0,392	0,000	0,676	(0,597;0,765)
Turno	Matutino				
	Vespertino	0,087	0,433	1,091	(0,878;1,354)
	Noturno	0,511	0,001	1,667	(1,233;2,255)
	Integral	0,014	0,892	1,014	(0,827;1,244)
Procedência	Não declarou				
	Escola pública	0,467	0,000	1,595	(1,249;2,036)
	Escola privada	0,423	0,002	1,527	(1,173;1,987)
Idade em 2010	Até 24 anos				
	De 25 a 50 anos	0,959	0,046	2,610	(1,016;6,704)
	Mais que 50 anos	0,740	0,126	2,095	(0,812;5,405)
Área do conhecimento	Ciências Exatas e da Terra				
	Ciências Biológicas	-0,545	0,000	0,580	(0,450;0,746)
	Engenharias	0,310	0,078	1,363	(0,966;1,925)
	Ciências da Saúde	0,982	0,000	2,671	(1,918;3,719)
	Ciências Agrárias	1,259	0,000	3,522	(2,679;4,631)
	Ciências Sociais Aplicadas	0,596	0,000	1,815	(1,341;2,457)
	Ciências Humanas	1,096	0,000	2,991	(2,286;3,914)
Linguística, Letras e Artes	0,229	0,046	1,258	(1,004;1,576)	
Bolsa assistência estudantil	Não				
	Sim	0,475	0,000	1,608	(1,397;1,850)
Bolsa assistência pedagógica	Não				
	Sim	1,210	0,000	3,352	(2,856;3,934)
Atividade complementar	Não				
	Sim	0,638	0,000	1,363	(0,966;1,925)
Constante		-1,400			

Fonte: Brasil/MEC/Inep (2010-2017). Elaborado pelos autores (2020).

Para o desfecho formado, verificou-se que a chance de um aluno matriculado em um curso de licenciatura se formar é 0,869 vezes menor que a chance de um do

ISSN: 1984-6444 | <http://dx.doi.org/10.5902/1984644443159>

bacharelado. E ao comparar a chance de se formar entre homens e mulheres, constatou-se que primeiros têm uma chance menor do que as últimas, e o fator de multiplicação é 0,676.

No que diz respeito ao turno, e tendo como categoria de referência o matutino, um aluno do vespertino, noturno ou integral tem uma chance maior de se formar do que um do matutino. Vale destacar que alunos do noturno têm 1,667 vezes a chance de se formar do que os do matutino. Quanto à procedência, os provenientes de escola pública e de escola privada têm, respectivamente, 1,595 e 1,527 vezes a chance de um aluno que não declarou sua procedência de se formar.

No que tange à idade, alunos com mais de 24 anos têm mais que o dobro de chance de se formar quando comparados aos que possuem até 24 anos. Aqueles que estão na categoria de 25 a 50 anos têm 2,610 a chance de se formar, e os que estão na categoria mais que 50 anos têm 2,095 vezes a chance de se formar que os com idade até 24 anos.

A análise da covariável área do conhecimento, por sua vez, mostrou que apenas alunos de cursos da área de Ciências Biológicas têm chance de se formar menor que os de Ciências Exatas e da Terra, e o fator multiplicativo é 0,580. Vale ressaltar que, os maiores fatores de multiplicação foram os de cursos das Ciências Agrárias (3,522), Ciências Humanas (2,991) e Ciências da Saúde (2,671), o que significa que os alunos dos cursos dessas três áreas são os que mais possuem chance de se formar em comparação com os alunos dos cursos da área de Ciências Exatas e da Terra.

Quanto à assistência estudantil, o modelo mostrou que os estudantes que receberam algum benefício têm 1,608 vezes a chance de se formar do que os que não receberam. Os que receberam bolsa por participar de alguma atividade pedagógica têm 3,352 vezes a chance de se formar do que os que não desenvolveram. E os que desenvolveram alguma atividade complementar possuem 1,363 vezes a chance de se formar do que aqueles que não desenvolveram.

Como exemplo de aplicação do modelo, considere um aluno que ingressou em um curso de Ciências Exatas e da Terra, turno vespertino, grau bacharelado, sexo feminino, idade até 24 anos, proveniente de escola pública, beneficiado com bolsa de

ISSN: 1984-6444 | <http://dx.doi.org/10.5902/1984644443159>

assistência estudantil e bolsa de assistência pedagógica e tendo participado de alguma atividade complementar. O modelo de regressão logística dá a probabilidade de um aluno qualquer se formar, o que é dado por

$$1/[1+\exp(-y)],$$

onde $\exp(\cdot)$ é a função exponencial e

$$y=B_0+B_1X_1+B_2X_2+\dots+B_{19}X_{19},$$

em que B_0 é a constante do modelo; B_1, B_2, \dots, B_{19} são os coeficientes logísticos associados às categorias das covariáveis; e X_1, X_2, \dots, X_{19} são as categorias das covariáveis selecionadas.

Dessa forma, para o caso particular escolhido, $y=1,7393$ e $1/[1+\exp(-1,7393)]=0,8506$. Portanto, o aluno é caracterizado como possível formado, pois a probabilidade de se formar é 85,06%, logo, maior que 50%.

Um modelo de predição da evasão da UFG

É preciso dizer que o objetivo inicial do estudo era criar um modelo para a predição da evasão do curso, da UFG e total. Então, primeiramente, passou-se para a criação do modelo final para a variável resposta evadido do curso e se verificou que não existem diferenças significativas entre os resultados preditos e os observados (valor p igual a 0,385 no teste de Hosmer e Lemeshov). Em termos de acurácia, o modelo não foi capaz de classificar corretamente nenhum dos evadidos do curso, mas classificou 100,0% dos não evadidos, resultando em uma precisão geral de 95,7%. O motivo pelo qual o modelo foi aprovado no teste Hosmer e Lemeshov, mas reprovado no quesito relevância, é a pequena quantidade de alunos na situação evadido do curso, apenas 4,3% do total.

Portanto, embora o modelo logístico construído tenha sido considerado significativo, sua eficácia em prever a evasão do curso é nula, pois não conseguiu classificar nenhum aluno que evadiu de seu curso como evadido. Por essa razão, esse modelo foi desconsiderado. Todavia, esses não foram negligenciados no trabalho, e suas informações são parte do grupo total de evadidos, que é composto por discentes que evadiram apenas dos cursos ou da instituição UFG.

Desse modo, passou-se para a criação do modelo para a variável resposta evadido da UFG, e se concluiu que não existem diferenças significativas entre os resultados preditos e os observados (valor p igual a 0,158 no teste de Hosmer e Lemeshov). Em termos de acurácia, o modelo foi capaz de classificar corretamente 49,2% dos evadidos da UFG e 85,4% dos não evadidos, resultando em uma precisão geral de 72,5%.

O Quadro 4 apresenta as covariáveis presentes no modelo logístico, os coeficientes logísticos, seu nível de significância (valor p), os coeficientes logísticos exponenciados (Exp(B)) e o intervalo de confiança (I.C.) de 95% para os valores de Exp(B).

Quadro 4 - Covariáveis, coeficientes, odds ratio e IC para o modelo associado à evasão da UFG

COVARIÁVEL	CATEGORIAS	B	Valor p	Exp(B)	IC para Exp(B)
Grau acadêmico	Bacharelado	0,088	0,318	1,092	(0,919;1,299)
	Licenciatura				
Sexo	Feminino	0,402	0,000	1,495	(1,314;1,701)
	Masculino				
Turno	Matutino	0,007	0,953	1,007	(0,805;1,259)
	Vespertino	-0,362	0,026	0,696	(0,506;0,957)
	Noturno	0,274	0,010	1,316	(1,068;1,621)
	Integral				
Procedência	Não declarou	-0,665	0,000	0,514	(0,404;0,655)
	Escola pública	-0,621	0,000	0,537	(0,413;0,699)
Idade em 2010	Até 24 anos	-0,807	0,066	0,446	(0,189;1,054)
	De 25 a 50 anos	-0,624	0,157	0,536	(0,226;1,271)
	Mais que 50 anos				
Área do conhecimento	Ciências Exatas e da Terra	0,318	0,013	1,375	(1,068;1,770)
	Ciências Biológicas	-0,160	0,379	0,852	(0,596;1,218)
	Engenharias	-0,870	0,000	0,419	(0,295;0,595)
	Ciências da Saúde	-0,785	0,000	0,456	(0,344;0,605)
	Ciências Agrárias	-0,436	0,007	0,647	(0,472;0,886)
	Ciências Sociais Aplicadas	-0,932	0,000	0,394	(0,298;0,520)
	Ciências Humanas	-0,151	0,201	0,860	(0,683;1,084)
Bolsa assistência estudantil	Não	-0,784	0,000	0,457	(0,392;0,532)
	Sim				
Bolsa atividade pedagógica	Não	-1,456	0,000	0,233	(0,194;0,281)
	Sim				
Atividade complementar	Não	-0,835	0,000	0,434	(0,352;0,534)
	Sim				
Constante		0,959			

Fonte: Brasil/MEC/Inep (2010-2017). Elaborado pelos autores (2020).

Para o desfecho evadidos da UFG, verificou-se que a chance de um aluno matriculado em um curso de licenciatura evadir da UFG é 1,092 vezes a chance de um matriculado no bacharelado. No que diz respeito à covariável sexo, os homens têm uma chance maior de evadir da UFG que as mulheres, e o fator de multiplicação é 1,495.

No que diz respeito ao turno, e tendo como categoria de referência o matutino, a chance de um aluno do vespertino evadir da UFG é 1,007 vezes maior do que um do matutino; ou seja, alunos do matutino e do vespertino têm praticamente a mesma chance de evadir da universidade. Por outro lado, a chance de um aluno do noturno evadir é 0,696 vezes a chance de um aluno do matutino; ou seja, ser aluno do turno noturno é um fator protetivo da evasão da instituição. Para os alunos do turno integral, a chance de evasão é 1,316 vezes a chance dos alunos do turno matutino.

No que concerne à procedência, alunos precedentes de escola pública ou privada têm praticamente a mesma chance de evadir da UFG, do que alunos que não declararam a procedência. Aqueles provenientes de escola pública, têm chance de evadir 0,514 vezes menor, e os vindos de escola privada têm chance de evadir 0,537 vezes menor do que aqueles que não declararam sua procedência.

O modelo também mostrou que alunos com idade acima de 24 anos têm chance menor de evadir da UFG do que alunos com 24 anos ou menos. Aqueles com idade de 25 a 50 anos têm uma chance de evadir 0,446 vezes menor do que aqueles com 24 anos ou menos; e os que têm 50 anos ou mais têm chance de evadir 0,539 vezes menor do que os que têm no máximo 24 anos.

Tendo como categoria de referência a área do conhecimento Ciências Exatas e da Terra, apenas alunos dos cursos da área das Ciências Biológicas têm chance de evadir da UFG maior do que alunos de cursos da categoria de referência e o fator multiplicativo é 1,375. O destaque é dado para as Ciências Humanas e Ciências da Saúde, que correspondem às áreas cujos alunos têm as menores chances de evadir da UFG quando comparados a alunos da categoria de referência, de 0,394 e 0,419, respectivamente.

No quesito assistência estudantil, os discentes que receberam algum benefício têm 0,457 vezes a chance de se evadir da instituição do que os que não receberam. Aqueles que receberam bolsa decorrente de participação em alguma atividade pedagógica têm 0,233 vezes a chance de se evadir da UFG do que os que não desenvolveram. E os participantes de atividades complementares têm 0,434 vezes a chance de se evadir do que aqueles que não participaram.

Utilizando o mesmo exemplo apresentado anteriormente no modelo de predição para os formados, obtêm-se $y = -3,0977$ e $1/[1 + \exp(3,0977)] = 0,0432$. Logo, o aluno é caracterizado como possível não evadido da UFG, pois a probabilidade de se evadir da instituição é 4,32%, menor que 50%.

Um modelo de predição da evasão total

No modelo final para a variável resposta total de evadidos, não existem diferenças significativas entre os resultados preditos e os observados (valor p igual a 0,314 no teste de Hosmer e Lemeshov). Em termos de acurácia, o modelo foi capaz de classificar corretamente 55,7% dos evadidos e 81,1% dos não evadidos, resultando em uma precisão geral de 71,0%.

O Quadro 5 apresenta as covariáveis presentes no modelo logístico, os coeficientes logísticos, nível de significância (valor p), os coeficientes logísticos exponenciados ($\text{Exp}(B)$) e o intervalo de confiança (I.C.) de 95% para os valores de $\text{Exp}(B)$.

Quadro 5 - Covariáveis, coeficientes, *odds ratio* e IC para o modelo associado à evasão

COVARIÁVEL	CATEGORIAS	B	Valor p	Exp(B)	IC para Exp(B)
Grau acadêmico	Bacharelado Licenciatura	0,172	0,048	1,188	(1,001;1,409)
Sexo	Feminino Masculino	0,373	0,000	1,452	(1,279;1,648)
Turno	Matutino	-0,058	0,605	0,944	(0,759;1,174)
	Vespertino	-0,554	0,000	0,574	(0,422;0,782)
	Noturno	-0,093	0,373	0,911	(0,743;1,118)
	Integral				
Procedência	Não declarou	-0,595	0,000	0,552	(0,432;0,704)
	Escola pública	-0,565	0,000	0,568	(0,436;0,740)
	Escola privada				
Idade em 2010	Até 24 anos	-0,826	0,073	0,438	(0,178;1,079)
	De 25 a 50 anos	-0,640	0,167	0,527	(0,213;1,306)
	Mais que 50 anos				
Área do conhecimento	Ciências Exatas e da Terra	0,264	0,039	1,302	(1,014;1,673)
	Ciências Biológicas	-0,462	0,009	0,630	(0,445;0,893)
	Engenharias	-1,350	0,000	0,259	(0,184;0,365)
	Ciências da Saúde	-1,281	0,000	0,278	(0,210;0,367)
	Ciências Agrárias	-0,921	0,000	0,398	(0,292;0,542)
	Ciências Sociais Aplicadas	-1,159	0,000	0,314	(0,239;0,412)
	Ciências Humanas	-0,234	0,044	0,791	(0,630;0,993)
Bolsa assistência estudantil	Não	-0,673	0,000	0,510	(0,441;0,590)
	Sim				
Bolsa atividade pedagógica	Não	-1,262	0,000	0,283	(0,239;0,335)
	Sim				
Atividade complementar	Não	-0,674	0,000	0,510	(0,422;0,616)
	Sim				
Constante		1,375			

Fonte: Brasil/MEC/Inep (2010-2017). Elaborado pelos autores (2020).

Para o desfecho total de evadidos, as conclusões foram semelhantes às obtidas para o desfecho evadido da UFG, o que se deve à natureza dos dados. O conjunto total de evadidos é exatamente o conjunto evadidos da UFG, que contém 2055 alunos, acrescido dos 245 alunos pertencentes ao conjunto evadidos do curso. Logo, as diferenças esperadas entre os dois modelos são poucas.

De acordo com o Quadro 5, o que pode ser ressaltado é que 6 *odds ratio* aumentaram e 19 diminuíram. A maior diferença se deu no *odds ratio* da categoria integral da covariável turno, que diminuiu 30,8%, tendo passado de 1,316 para 0,911.

ISSN: 1984-6444 | <http://dx.doi.org/10.5902/1984644443159>

A menor diferença ocorreu na covariável atividade pedagógica, que teve seu *odds ratio* aumentado em 21,4%, mudando de 0,233 para 0,283.

Embora tenha ocorrido variação em todos os *odds ratio*, apenas duas delas mudaram o caráter protetivo ou não da covariável em questão. No modelo referente ao desfecho evasão da UFG, ser aluno do turno vespertino significava estar mais protegido da evasão do que um aluno do turno matutino (categoria de referência), pois o *odds ratio* dessa categoria é 1,007. Já no modelo relativo ao desfecho evasão total, esse mesmo *odds ratio* diminuiu para 0,944 e passou a ser um fator de risco para a evasão. Na modelagem da evasão da UFG, o *odds ratio* da categoria integral da covariável turno, valia 1,316 e, portanto, ser aluno do integral significa estar mais protegido da evasão do que um aluno do matutino. Entretanto, para a modelagem da evasão total, ser aluno do integral passou a ser um fator de risco, pois o *odds ratio* diminuiu para 0,911.

No modelo referente à evasão da UFG, o maior *odds ratio* associado à covariável área do conhecimento pertencia à categoria Ciências Biológicas, sendo os alunos desses cursos os que possuem maior chance de evadir, quando comparados aos alunos da categoria de referência (Ciências Exatas e da Terra). Isso se manteve no modelo proposto para a evasão total, porém, o menor *odds ratio* associado a essa covariável mudou. Para o desfecho evasão da UFG, a categoria que possuía o menor *odds ratio* era a das Ciências Humanas, mas, com a inclusão dos 245 alunos, o menor *odds ratio* passou a ser aquele associado à categoria das Ciências Sociais. Ou seja, no modelo proposto para a evasão da UFG, ser aluno de algum curso da área de Ciências Humanas significava ter uma chance de evadir da UFG 0,394 vezes menor do que a de um aluno de um curso da área de Ciências Exatas e da Terra. Com a inclusão dos 245 alunos, os estudantes das Ciências Sociais é que passaram a ter a menor chance de evadir quando comparados com a categoria de referência.

Valendo-se do mesmo exemplo apresentado no modelo de predição para os formados, obtém-se $y = -2,3698$ e $1/[1 + \exp(2,3698)] = 0,0855$. Sendo assim, o aluno é caracterizado como possível não evadido, pois a probabilidade de se evadir é 8,55%, menor que 50%.

A predição da evasão por área do conhecimento - uma possibilidade

Além dos quatro modelos apresentados anteriormente, outros 32 foram construídos. Para cada uma das oito áreas do conhecimento presentes no Quadro 1, foi criado um modelo de regressão logística, por se entender que o estudo individualizado dos desfechos já considerados pode ser relevante para o entendimento da formatura e da evasão na UFG. Devido à grande quantidade de informação, optou-se por apresentar na sequência apenas os aspectos que se destacaram.

Verificou-se que a área que apresentou o modelo com maior precisão geral, foi a Ciências da Saúde com o modelo envolvendo o desfecho evadido da UFG, cuja precisão é de 81,6%, posto que o evento de interesse foi acertado 13,8% das vezes e a não evasão 98,4% das vezes. Por outro lado, o modelo pertencente à área Linguística, Letras e Artes e que modela o desfecho formado é o que tem a menor precisão geral, de 63,0%, sendo que o desfecho de interesse foi acertado 46,3% das vezes e a não formatura 76,7%.

O modelo que mais acerta o desfecho categorizado como um (formatura ou evasão), é aquele associado à área Ciências da Saúde e vinculado ao desfecho formado. Para esse modelo, a porcentagem de acerto foi de 98,6%, enquanto a não formatura foi acertada 8,4% das vezes. Em contrapartida, o modelo que menos acerta o desfecho categorizado como um, está associado à área Ciências Agrárias, modelando o desfecho evasão da UFG e com uma precisão de 3,6%. A não evasão da UFG foi acertada 98,9% das vezes.

O desfecho formado é melhor modelado na área Ciências da Saúde, com 78,6% de precisão, e tem seu pior modelo na área Linguística, Letras e Artes, com 63,0% de precisão. Para o desfecho evadido da UFG, o modelo com maior precisão é aquele da área Ciências da Saúde, com 81,6% de acerto; e o de menor precisão é o da área Linguística, Letras e Artes, que acertou 68,5% das vezes. Quanto ao desfecho evadido, a maior precisão ocorreu na área Ciências da Saúde, cujo modelo associado teve 80,7% de acerto; e a menor precisão pertence ao modelo da área Linguística, Letras e Artes, com 64,2% de acerto.

Considerações finais

Com esse estudo, ficou evidente que a regressão logística permite identificar variáveis que são fatores de risco ou protetivos para a formatura e a evasão. Além disso, possibilita comparar categorias de uma mesma característica por meio da razão de chances, bem como avaliar a probabilidade de qualquer aluno se formar ou evadir.

Contudo, uma das restrições do modelo de regressão logística, é não permitir o cálculo de *odds ratio* entre quaisquer categorias de uma covariável, mas apenas entre a categoria de referência as demais, de uma mesma covariável. Ainda é preciso destacar que, dos três modelos propostos - formados, evadidos da UFG e total de evadidos - apenas os associados ao desfecho formados e total de evadidos, teve porcentagem de acertos superior a 50% no desfecho de interesse, de 76,4% e 55,7%, respectivamente. Já a menor precisão geral de acertos de todos os modelos foi superior dos formados, com 69,7% de acerto. O segundo maior foi o do total de evadidos, com 71,0% de acerto, e o melhor modelo foi o evadidos da UFG, com 72,5% de acerto, apesar de esse ter sido o pior modelo no quesito desfecho de interesse.

Nesse ponto, é preciso dizer que os modelos poderiam ter maiores percentuais de acerto caso o Censo da Educação Superior disponibilizasse dados que são imprescindíveis para o entendimento do perfil do aluno que se forma ou evade, como, por exemplo: renda familiar, estado civil, escolaridade dos pais, nota para ingresso na instituição, e outras.

Considerando o exposto, mesmo com suas limitações, a capacidade explicativa dos modelos pode ser considerada boa, e se acredita que a utilização dessa técnica estatística é de suma importância para subsidiar a tomada de decisões por parte dos gestores educacionais, para que possam agir de forma preventiva, evitando a evasão dos alunos que pertencem a grupos de risco.

Referências

BRASIL. Ministério da Educação. Diretoria de Estatísticas Educacionais. **Metodologia de Cálculo dos Indicadores de Fluxo da Educação Superior**. Brasília: MEC/DEED, 2017.

ISSN: 1984-6444 | <http://dx.doi.org/10.5902/1984644443159>

BRASIL. Ministério da Educação. Diretoria de Estatísticas Educacionais. **Censo da Educação Superior 2018** – divulgação dos resultados. Brasília: MEC/DEED, 2019. Disponível em: http://download.inep.gov.br/educacao_superior/censo_superior/documentos/2019/apresentacao_censo_superior2018.pdf. Acesso em: 25 fev. 2020.

BRASIL. Ministério da Educação. Instituto Nacional de Pesquisas e Estudos Educacionais Anísio Teixeira. **Censo da Educação Superior** – microdados. 2010 - 2017. Disponível em: <http://portal.inep.gov.br/web/guest/microdados>. Acesso em: 20 set. 2019.

CUNHA, Jacqueline Veneroso Alves; CORNACHIONE JUNIOR, Edgard B.; MARTINS, Gilberto de Andrade. Uma aplicação da regressão logística no inventário de estilos de aprendizagem de Canfield (LSI) sob a ótica das reprovações acadêmicas. **Revista de Contabilidade e Organizações**, v. 3, n. 2, p. 100-112, maio/ago. 2008.

DAVID, Larmartine Moreira Lima; CHAYM, Carlos Dias. Evasão universitária: um modelo para diagnóstico e gerenciamento de instituições de ensino superior. **Revista de Administração IMED**, Passo Fundo, v. 9, n. 1, p. 167-186, jan./jun., 2019

FRITSCH, Rosangela et al. 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.

GONZALES, Ricardo Alonso; NASCIMENTO, Janicleide Gonçalves; LEITE, Luciana Barone. Evasão em cursos a distância: um estudo aplicado na Universidade Corporativa da Secretaria da Fazenda do Estado da Bahia. **Rev. Serv. Público**, Brasília, v. 47, n. 4, p. 627-647, out./dez. 2016.

HAIR JÚNIOR, Joseph F. et al. **Análise multivariada de dados**. 6 ed. Porto Alegre: Bookman, 2009.

HOSMER JUNIOR, David W.; LEMESHOW, Stanley. **Applied logistic regression**. 2 ed. Nova York: John Wiley & Sons, 2000.

LEWIS-BECK, Michael S. Series Editor's Introduction. In: MENARD, Scott. **Applied Logistic Regression Analysis**. Thousand Oaks: Sage, 2001.

MAZULO, Emerson Silva. Análise de proficiência em Matemática por meio de regressão linear múltipla. **Revista Intersaberes**, v. 10, n. 21, p. 613-625, set./dez. 2015.

ISSN: 1984-6444 | <http://dx.doi.org/10.5902/1984644443159>

MESQUITA, Paulo Sérgio Belchior. **Um modelo de regressão logística para avaliação dos programas de pós-graduação no Brasil**. 90 f. 2014. Dissertação (mestrado em Engenharia de Produção) - Universidade Estadual do Norte Fluminense "Darcy Ribeiro", Centro de Ciência e Tecnologia, Campos dos Goytacazes, 2014.

OCDE. **Education at a glance 2019** – OCDE Indictors. Paris: OCDE Publishing, 2019. Disponível em: https://read.oecd-ilibrary.org/education/education-at-a-glance-2019_f8d7880d-en#page1. Acesso em: 01 mar. 2020.

PEREIRA, Alexandre Severino et al. Fatores relevantes no processo de permanência prolongada de discentes nos cursos de graduação presencial: um estudo na Universidade Federal do Espírito Santo. **Ensaio: aval. pol. públ. Educ.**, Rio de Janeiro, v. 23, n. 89, p. 1015-1039, out./dez. 2015.

SILVA FILHO, Rogério Luiz Cardoso. **Modelo de análise e predição do desempenho dos alunos dos Institutos Federais de Educação usando o ENEM como indicador de qualidade escola**. 93 f. 2017. Dissertação (Mestrado em Ciências da Computação) - Universidade Federal de Pernambuco, Centro de Informática, Recife, 2017.

SOUSA, Carolina Rodrigues de Oliveira et al. Fatores preditores da evasão escolar entre adolescentes com experiência de gravidez. **Cadernos de Saúde Coletiva**, Rio de Janeiro, v. 26, n. 2, p. 160-169, abr./jun. 2018.



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)

Notas

¹ Consultar países membros da OCDE em: <http://www.oecd.org/>

² Bolsas de pesquisa, extensão, monitoria ou estágio.

³ Áreas do conhecimento foram divididas considerando a classificação de árvore do conhecimento do CNPq, disponível em: [www.cnpq.br > documents > TabeladeAreasdoConhecimento](http://www.cnpq.br/documents/TabeladeAreasdoConhecimento).