

A influência de fatores socioeconômicos na desistência dos estudantes da UNIFESP

Fabício Costa Souza ¹, Vanessa Ribeiro Neves ² Anderson da Silva Rosa ³,
Juliana Garcia Cespedes ⁴

Introdução

Segundo Mercuri e Polydoro (2004) o termo evasão pode indicar três situações na relação aluno e instituição de ensino, sendo eles: "a evasão de curso, definida como a saída definitiva do aluno de seu curso de origem, sem concluí-lo"; "a evasão de instituição, quando o aluno abandona a instituição de ensino em que está, podendo ou não mudar de curso" e "a evasão de sistema, quando o aluno desiste do ensino superior, abandonando por completo os estudos universitários".

A evasão estudantil pode ser vista como um fenômeno educacional complexo, que ocorre em todos os tipos de instituição de ensino e afeta o sistema educacional como um todo. Diante de sua importância, o tema tem sido objeto de investigações em todo o mundo, as quais buscam aprofundar os motivos, as causas e os impactos provenientes do abandono dos estudantes nos sistemas de ensino (MOROSINI et al., 2011).

Lobo, Filho e Lobo (2016) ressaltam que a evasão geram desperdícios tanto sociais quanto acadêmicos e econômicos, principalmente no setor público, no qual os recursos investidos oriundos de impostos não possui o seu devido retorno. Já no setor privado implica em perdas nas receitas. Além disso, destacam que a evasão ocasiona ociosidade de professores, funcionários, equipamentos e espaço físico.

Na literatura há trabalhos que visam investigar fatores influentes na decisão de desistência do curso de graduação em perspectivas qualitativas ou quantitativas. Entre esses trabalhos há aqueles que analisam o fenômeno entre os estudantes pertencentes a um curso em específico como Nicolau¹, Gagno e Laudares (2015), Rafael, Miranda e Carvalho (2015) e Lima e Machado (2014), numa visão mais geral, analisando um conjunto de cursos, como em Adachi (2009) e Fredenhagen (2014), ou no ensino superior como um todo, tais como, Pereira, Brasil e Samohyl (2004) e Cunha e Morosini (2009).

Nesse contexto, torna-se notória a importância de se identificar a existência de grupo de risco, ou seja, de possíveis fatores que caracterizam um estudante como provável desistente. Uma vez identificados pode-se definir políticas institucionais de permanência estudantil mais eficazes.

Para a realização deste trabalho será utilizado o modelo de regressão logística com abordagem Bayesiana visando a construção de um modelo estatístico para cada campus da Universidade Federal de São Paulo (UNIFESP), sendo esses: Baixada Santista, Diadema, Guarulhos, Osasco, São José dos Campos e São Paulo, para identificar os fatores socioeconômicos dos estudantes desistentes que ingressaram entre os anos de 2012 e 2016.

¹UNIFESP. e-mail: *fcsouza14@unifesp.br*

²UNIFESP. e-mail: *vanessaneves@gmail.com*

³UNIFESP. e-mail: *andersonsrosa83@gmail.com*

⁴UNIFESP. e-mail: *cespedes.juliana@gmail.com*

Material e métodos

Os bancos de dados utilizados nesse trabalho foram disponibilizados pela Pró-Reitoria de Graduação (PROGRAD) e pela Pró-Reitoria de Assuntos Estudantis (PRAE) da UNIFESP. O banco do perfil socioeconômico contém informações de 13402 estudantes ingressantes obtidas entre os anos de 2012 à 2016, já o da situação acadêmica contém informações de 30155 estudantes de 2007 à 2016.

Foram selecionadas 26 variáveis socioeconômicas para análise considerando apenas as variáveis com histórico nesse período. Após o pré-processamento dos dados, obteve-se 9911 estudantes, sendo 3331 desistentes, 5676 estudantes em curso, 741 formados e 163 estudantes com matrículas trancadas.

A variável resposta do modelo de regressão logística possui distribuição Binomial cuja natureza é binária (dicotômica), sendo atribuído o valor 1 ao evento de interesse (sucesso estatístico) e 0 ao complementar (insucesso).

Seja Y a variável resposta e x_i , $i = 1, \dots, k$ variáveis explicativas. A probabilidade de um estudante desistir do curso ao qual ingressou na universidade, dado o vetor de variáveis explicativas observado, é:

$$\Pi(x'_i\beta) = p[y = 1|x_i, \beta] = p_i$$

em que β é o vetor de dimensão $K + 1$ de coeficientes de regressão.

Considerando o modelo logístico:

$$\Pi(x'_i\beta) = \frac{e^{x'_i\beta}}{1 + e^{x'_i\beta}}$$

tem-se a função de ligação dada por:

$$g(x_i) = \ln \left[\frac{\Pi(x_i)}{1 - \Pi(x_i)} \right] = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

Dessa forma, a função de verossimilhança é dada por:

$$l(\beta|Y) = \prod_{i=1}^n (\Pi(x_i))^{y_i} (1 - \Pi(x_i))^{1-y_i} \quad (1)$$

Do ponto de vista Bayesiano a informação à posteriori é obtida por:

$$p(\beta|Y) = \frac{l(\beta|Y)p(\beta)}{\int l(\beta|Y)p(\beta)d\beta} \quad (2)$$

em que $p(\beta)$ é a distribuição à priori.

Considerando como priori a distribuição normal com hiperparâmetros μ e σ^2 , a distribuição à *posteriori* para a regressão logística pode ser definida como (MENDONÇA, 2008):

$$p(\beta|y) \propto \prod_{i=1}^n (\Pi(x_i))^{y_i} (1 - \Pi(x_i))^{1-y_i} \prod_{j=1}^k e^{-\frac{1}{2}(\frac{\beta_j - \mu_j}{\sigma_j^2})^2} \quad (3)$$

Segundo Ehlers (2011) quando a distribuição a *posteriori* for de difícil tratamento analítico, utiliza-se diversos métodos numéricos, dispostos na literatura, para a obtenção

de uma amostra da distribuição de interesse, sendo a mais comum a classe de métodos de Monte Carlo via Cadeias e Markov baseadas em simulações estocásticas.

Para analisar a colinearidade das variáveis explanatórias, foi implementando um programa, em linguagem R, que calcula a matriz de correlação utilizando o coeficiente de correlação de Pearson, teste χ^2 de Pearson e coeficiente de determinação R^2 para gerar todos os conjuntos de variáveis independentes possíveis.

Foram ajustados um modelo para cada conjunto de variáveis balancelando as classes pela técnica *undersampling* (PRATI; BATISTA; MONARD, 2003) e então, após selecionar o modelo com menor *DIC*, apresentou-se apenas as variáveis significativas.

Resultados e discussões

Adotando uma priori não informativa normal com média 0 e variância 1000, foi realizado um primeiro ajuste para cada conjunto. Para a obtenção das amostras a *posterioris* de interesse utilizou-se *burn-in* de 5000 com *thin* de 5 e então, na etapa de monitoramento adotou-se 30000 atualizações com *thin* de 10 e inicializando as cadeias com o valor zero.

Após avaliar a convergência da cadeia através da autocorrelação e histórico, foram analisadas as densidades, traços, quantis e as demais estatísticas.

Dentre as variáveis consideradas no ajuste do modelo aquelas que apresentaram significância em pelo menos um campus foram:

- | | |
|--|----------------------------|
| 1. Responsável pela manutenção na IES | 5. Perfil étnico-racial |
| 2. Sexo do aluno | 6. Idade de ingresso |
| 3. Já iniciou curso em IES Particular? | 7. Renda <i>per capita</i> |
| 4. Escolaridade da mãe | 8. Tempo de ensino médio |

Ajustou-se um modelo estatístico para cada Campus da UNIFESP. O Campus Baixada Santista contém 1763 estudantes ativos, sendo 571 desistentes e 1192 não desistentes. Para o campus Diadema tem-se 2886 estudantes ativos, sendo 832 desistentes e 2054 não desistentes. O número total de estudantes ativos no Campus de Guarulhos é de 2473, sendo 1056 desistentes e 1417 não desistentes. No Campus São José dos Campos tem-se 963 estudantes ativos, sendo 278 desiste e 685 não desistentes. O número de estudantes do campus Osasco é de 1264, sendo 385 desistentes e 879 não desistentes. O número de estudantes ativos no campus São Paulo é de 1231, sendo 209 desistentes e 1022 não desistentes. A tabela 1 apresenta o modelo final ajustado apenas com as variáveis significativas.

A curva ROC apresentou uma área sob a curva de 0,65 para os Campi Baixada Santista, Diadema e Osasco. O valor para Guarulhos foi de 0,59, para São José dos Campos foi 0,63 e para São Paulo 0,7, indicando um poder discriminatório pobre. A matriz de classificação do modelo final é apresentada na tabela (tabela 2) utilizando ponto de corte ótimo.

Para o campus Baixada Santista, observa-se que o estudante que é o próprio responsável pela manutenção na universidade tem probabilidade maior de desistir do curso do que aquele que possui ajuda de familiares (cônjuge, pais, irmãos ou outra pessoa). Esse modelo conseguiu prever corretamente 66% dos estudantes desistentes como pode ser observado na tabela 2.

Para o campus Diadema, observa-se que o estudante que possui menor renda per capita e maior idade de ingresso, tem maior chance de desistir do curso de graduação. Esse modelo conseguiu prever corretamente 62% dos estudantes desistentes como pode ser observado na tabela 2.

Tabela 1: Coeficientes do modelo final de cada campus.

Campus	Parâmetro	Média	Erro padrão	2,5%	Mediana	97,5%
Baixada Santista	Intercepto	0,981	0,170	0,652	0,979	1,322
	1	-1,293	0,188	-1,670	-1,292	-0,931
Diadema	Intercepto	-0,005	0,000	-0,025	-0,005	0,015
	6	0,006	0,000	0,002	0,006	0,010
	7	-0,003	0,000	-0,0211	-0,003	0,016
Guarulhos	Intercepto	-0,415	0,075	-0,565	-0,415	-0,269
	8	0,650	0,147	0,369	0,649	0,942
	7	0,223	0,038	0,150	0,223	0,300
Osasco	Intercepto	0,0393	0,150	-0,255	0,040	0,336
	1	-0,790	0,144	-1,074	-0,790	-0,505
	7	0,192	0,041	0,114	0,192	0,276
São José dos Campos	Intercepto	-0,356	0,135	-0,624	-0,357	-0,089
	5	0,278	0,094	0,096	0,276	0,466
	3	1,032	0,298	0,458	1,028	1,626
São Paulo	Intercepto	-0,003	0,264	-0,514	-0,002	0,520
	2	0,957	0,188	0,591	0,955	1,327
	4	-0,380	0,090	-0,557	-0,379	-0,204

Para o campus Guarulhos, o estudante que concluiu o ensino médio com mais de 3 anos e possui maior renda *per capita* tem maiores chances de desistir do curso. Esse modelo conseguiu prever corretamente 57% dos estudantes desistentes como pode ser observado na tabela 2.

Para o campus Osasco, o estudante que é o responsável pela manutenção na universidade tem maior probabilidade de desistir do curso comparado ao que possui ajuda de familiares (cônjuge, pais, irmãos, ou outra pessoa) e aquele que possui maior renda *per capita*. Esse modelo conseguiu prever corretamente 63% dos estudantes desistentes como pode ser observado na tabela 2.

Para o campus São José dos Campos, o estudante que é autodeclarado preto, pardo, amarelo ou indígena e aquele que já cursou em uma IES particular possuem maiores chances de desistir do curso pelo qual ingressou. Esse modelo teve uma acurácia de 64%.

E por fim, para o campus São Paulo, o estudante que é do sexo masculino e a mãe não é alfabetizada ou não concluiu o ensino fundamental possui maiores chances de desistir. Esse modelo conseguiu prever corretamente 66% dos estudantes desistentes como pode ser observado na tabela 2.

Conclusões

Todos os modelos tenderam a acertar mais estudantes não desistentes que desistentes, o que mostra que apenas as variáveis socioeconômicas não é o suficiente para concluir o fato de que o estudante pode desistir do curso. É preciso considerar que no ambiente acadêmico fatos como saúde mental e insatisfação são impactantes na continuidade do curso pelo qual ingressou, algo que as variáveis utilizadas não captaram.

Entrar em contato com os estudantes desistentes ainda parece uma abordagem atrativa no que diz respeito à investigar fatores influentes para enriquecer o questionário socioeconômico.

Tabela 2: Matriz de classificação dos modelos finais dos campi.

Campus		Situação	Não Desistente	Desistente	%
Baixada Santista	Observado	Não Desistente	239	32	88
		Desistente	151	120	44
		Total			66
Diadema	Observado	Não Desistente	603	229	72
		Desistente	409	423	51
		Total			62
Guarulhos	Observado	Não Desistente	774	282	73
		Desistente	630	426	40
		Total			57
Osasco	Observado	Não Desistente	271	114	70
		Desistente	173	212	55
		Total			63
São José dos Campos	Observado	Não Desistente	134	33	80
		Desistente	88	79	47
		Total			64
São Paulo	Observado	Não Desistente	136	24	85
		Desistente	84	76	48
		Total			66

Agradecimentos

CNPq, PROGRAD e PRAE.

Referências

- ADACHI, A. A. C. T. *Evasão e Evadidos nos Cursos de Graduação da Universidade Federal de Minas Gerais*. Belo Horizonte - MG: [s.n.], 2009. 49-71 p. Faculdade de Educação - Programa de Pós-Graduação em Educação.
- CUNHA, E. R.; MOROSINI, M. C. Evasão na educação superior: Uma temática em discussão. In: *Revista Cocar*. Belém - PA: [s.n.], 2009. p. 82-89.
- EHLERS, R. S. *Inferência Bayesiana*. [S.l.]: ICMC - USP, São Paulo, 2011.
- FREDENHAGEM, S. V. Evasão escolar no âmbito do instituto federal de Brasília. In: *Revista Eixo*. Brasília - DF: [s.n.], 2014.
- LIMA, E.; MACHADO, L. A evasão discente nos cursos de licenciatura da universidade federal de Minas Gerais. In: *Educação Unisinos*. São Leopoldo - RS: [s.n.], 2014. p. 121-129.
- LOBO, R. L.; FILHO, S.; LOBO, M. B. *Evasão no Ensino Superior: Causas e Remédios*. 2016. Rio Grande do Sul - RS. Disponível em: <http://robertolobo.com.br/index.php/2009/06/evasao-no-ensino-superior-causas-e-remedios/>.

MENDONÇA, T. S. *Modelos de Regressão Logística Clássica, Bayesiana e Redes Neurais para Credit Scoring*. [S.l.]: Universidade Federal de São Carlos, São Carlos, 2008. Dissertação - Mestrado.

MERCURI, E.; POLYDORO, S. A. J. *O compromisso com o curso no processo de permanência/evasão no Ensino Superior: algumas contribuições*. Taubaté - SP: [s.n.], 2004. Estudante universitário: características e experiências de formação.

MOROSINI, M. C. et al. A evasão na educação superior no brasil: uma análise da produção de conhecimento nos periódicos qualis entre 2000-2011. In: *Conferência LatinoAmericana Sobre El Abandono em La Educación Superior*. Managua - Nicaragua: [s.n.], 2011. p. 1-10.

NICOLAU¹, L. N.; GAGNO, K. G. dos A.; LAUDARES, F. A. L. Endendendo a evasão através do perfil socioeconômico dos alunos de licenciatura em física da ufrj. In: *XII Congresso Nacional de Educação (EDUCERE)*. Curitiba - PR: [s.n.], 2015.

PEREIRA, F. C. B.; BRASIL, G. H.; SAMOHYL, R. W. Análise fatorial e a evasão de alunos no ensino superior. In: *XXXVI - SBPO*. São João del Rei - MG: [s.n.], 2004.

PRATI, R. C.; BATISTA, G. E. A. P. A.; MONARD, M. C. *Uma experiência no Balanceamento Artificial de Conjuntos de Dados para Aprendizado com Classes Desbalanceadas utilizando Análise ROC*. 2003. São Carlos. Disponível em: http://conteudo.icmc.usp.br/pessoas/gbatista/files/atai2003_2.pdf.

RAFAEL, J. A. M.; MIRANDA, P. R. de; CARVALHO, M. P. de. Análise de evasão em um curso de licenciatura em matemática da rede federal de ensino nos seus primeiros cinco anos de implantação. In: *Revista Paranaense de Educação Matemática*. Campo Mourão - PR: [s.n.], 2015. p. 118-135.