



08, 09, 10 e 11 de novembro de 2022
ISSN 2177-3866

PERFIL E DESEMPENHO DOS ALUNOS DE ENGENHARIA COM BASE NOS RESULTADOS DAS AVALIAÇÕES DO ENADE.

NATÁLIA DOS SANTOS ANDRADE
UNIVERSIDADE DE SÃO PAULO (USP)

FERNANDO FREIRE VASCONCELOS
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE DA UNIVERSIDADE DE SÃO PAULO - FEA

PERFIL E DESEMPENHO DOS ALUNOS DE ENGENHARIA COM BASE NOS RESULTADOS DAS AVALIAÇÕES DO ENADE

1 Introdução

No Brasil, os cursos de engenharia ainda fazem parte dos cursos com alta procura pelos estudantes no momento da escolha de um curso superior, sendo que alguns cursos ficam entre os 10 maiores índices de matrículas realizadas conforme resultados disponibilizados pelo censo da educação superior (Diretoria de Estatísticas Educacionais, 2020; INEP, 2020), mas quais fatores podem influenciar o desempenho dos estudantes.

Uma das formas de avaliação da efetividade do ensino superior é através do Exame Nacional de Desempenho dos Estudantes – ENADE, a prova aplicada aos estudantes serve para compor a nota do curso no Ministério da Educação – MEC, e como base para compor os indicadores de qualidade do ensino superior no Brasil (INEP, 2021a).

A avaliação do ENADE costuma ocorrer uma vez por ano, avaliando os alunos ingressos nos cursos e os alunos concluintes do mesmo, porém cada curso costuma ser avaliado em conjunto com as áreas de avaliação de cada edição, todo ano o governo divulga quais serão os cursos avaliados daquele ano (INEP, 2021a).

O Conceito ENADE é uma nota de 1 a 5 que cada curso de cada instituição recebe com base no desempenho dos estudantes no ENADE e é calculado com base na média ponderada das notas dos alunos conforme a fórmula abaixo, muitos cursos utilizam essa nota para divulgar a qualidade do curso e atrair mais alunos e a nota é calculada apenas se houver mais de dois alunos do curso realizando a prova, eq. (1).

$$NC_c = 0.25NP_{FG_c} + 0.75NP_{CE_c} \quad (1)$$

sendo, NC_c Nota dos concluintes do curso de graduação (conceito ENADE sem discretizar), NP_{FG_c} Nota padronizada em Formação Geral, NP_{CE_c} Nota padronizada em Componente Específico, (INEP, 2021b).

Quando há a convocação os alunos são obrigados a realizar a prova, podendo sofrer sanções como não obter o diploma, caso não realizem a prova.

Na literatura já é conhecido que o desempenho dos alunos nessas provas pode sofrer influência de diferentes variáveis como apresentado em Neto (2021), Meurer e Pereira (2020) e Medeiros et. al. (2019), que trazem aspectos socioeconômicos como faixa de renda, grau de escolaridade dos pais entre outros, localização da faculdade e modalidade do curso. Para analisar o desempenho dos alunos nessas provas, e entender quais fatores mais impactam existe diferentes ferramentas estatísticas que já foram e podem ser utilizadas.

Em Meurer e Pereira (2020), por exemplo, foi realizado uma regressão logística com base em Fávero et al. (2017), e para utilizar os dados referente a instituição foi feito uma análise fatorial com base em Hair Jr (2009), a fim de entender a relação entre o desempenho dos alunos de ciências contábeis com seu perfil e alguns aspectos dos cursos como infraestrutura, organização e oportunidades profissionais, utilizando os dados do ENADE 2018. Nesse trabalho, concluiu-se que variáveis como modalidade presencial de ensino, pai e mãe com formação superior, ser do gênero masculino, graduação em IES pública e maior grau de satisfação com a infraestrutura e organização da instituição de ensino influenciavam para um melhor desempenho os estudantes.

Testes qui-quadrado e Modelos de regressão foram utilizados em Mallmann et al. (2021) para estudar e relação entre a leitura de livros e o desempenho acadêmico dos alunos de ciências sociais

tendo como base os dados do ENADE 2018. Através da análise os autores concluíram que existe uma alta relação entre o desempenho acadêmico com a quantidade de livros lida.

E em Vieira e Schneiker (2021), através da aplicação de uma regressão linear múltipla encontrou-se relações entre o perfil do docente e o desempenho dos alunos, como por exemplo o grau de escolaridade da docência que se apresentou como um fator positivo no desempenho dos alunos.

Portanto, diante do que foi exposto, neste projeto foi desenvolvido uma análise do perfil e do desempenho dos alunos dos cursos de engenharia através dos resultados das avaliações que eles obtiveram no exame do ENADE e foi avaliado a evolução do desempenho nos últimos anos com base nos dados dos anos de 2019 e 2017 do conceito ENADE dos cursos utilizando modelo de regressão logística e modelo de regressão multinível.

O modelo de regressão logística ora apresentado consegue alcançar uma boa eficiência a probabilidade de um desempenho acima da média dos estudantes de acordo com o perfil do mesmo. No modelo conseguimos verificar que estudantes com renda de até 3 salários-mínimos tem chance de 54,9% inferior de atingir um bom desempenho nas avaliações do ENADE se comparado com estudantes quem tem uma renda familiar acima de 10 salários. Além disso, a cada ano acrescentado na idade também reduz em 4,9% as chances de obter um bom desempenho. Outra característica que chamou atenção no modelo proposto foi o aumento em 21% e 13% nas chances do bom desempenho quando o pai e/ou a mãe, respectivamente, tem nível superior em relação aos estudantes que seus pais não contam com este tipo de formação.

Além de tudo, ter estudado em uma instituição de ensino superior pública é a variável que mais afeta na probabilidade de um desempenho superior dos estudantes que chega a ter quase 3 vezes mais chances de obter um bom desempenho em comparação com estudante de universidade privada. No tocante às ocupações, o estudante trabalhar ao longo da graduação e/ou não ter sido bolsista por mérito reduz em 19,9% e 32,1% as chances de um bom desempenho, respectivamente. Já o estudante ser da região sul e do sexo masculino acresce em 29,5% e 20,7% às chances dele de obter um bom desempenho.

Na modelagem multinível verificou-se que existe uma relação entre os resultados dos cursos e a instituição de ensino que o estudante, sendo mais importante que área de avaliação e o curso. Assim, podem ser direcionadas políticas de melhoria das instituições de ensino públicas, com foco naquelas que reduzem a nota média do conceito do ENADE dos cursos de engenharia.

Destaca-se que como o trabalho tem um conteúdo mais aplicado, visando comparar os modelos mencionados na introdução e testados por outros autores com este ora proposto, que utiliza técnicas de *analytics* e *machine learning*, optou-se em partir diretamente para a metodologia utilizada.

2 Metodologia.

Para analisar o perfil e o desempenho dos alunos foi escolhido como fonte de dados a base de dados do ENADE 2019, utilizada na forma de microdados extraídos diretamente do site do INEP, (INEP, 2021c). A população final do estudo compreendeu 99.341 estudantes dos cursos de engenharia que tiveram respostas válidas no ENADE, para formar um banco de dados sem “*missing values*” nas variáveis de interesse do estudo. Os estudantes estavam divididos nos seguintes cursos:

Tabela 1. Cursos de Engenharia e alunos participantes do ENADE

Curso	Quantidade de Estudantes
ENGENHARIA DA COMPUTAÇÃO	2.676
ENGENHARIA CIVIL	37.636
ENGENHARIA ELÉTRICA	9.627

ENGENHARIA DE CONTROLE E AUTOMAÇÃO	2.796
ENGENHARIA MECÂNICA	14.444
ENGENHARIA DE ALIMENTOS	1.274
ENGENHARIA QUÍMICA	5.774
ENGENHARIA DE PRODUÇÃO	17.735
ENGENHARIA AMBIENTAL	5.759
ENGENHARIA FLORESTAL	1.620

Fonte: Resultados originais da pesquisa.

Os cursos de engenharia civil, engenharia mecânica e engenharia de produção foram os que mais tiveram estudantes participando da prova do ENADE em 2019, sendo que o curso de engenharia civil representa aproximadamente 38% do total de alunos participantes, Tabela 1.

O perfil dos estudantes foi identificado por uma análise exploratória na base, utilizando técnicas de estatística descritivas e de visualização dos dados, que auxiliaram no entendimento do perfil dos alunos que realizaram a prova.

Já para o desempenho dos estudantes, foi verificado em diferentes trabalhos as variáveis que se mostravam mais estatisticamente significantes. Em Meurer e Pereira (2020), por exemplo, variáveis como categoria do curso, gênero, modalidade, idade, trabalho e nível de instrução do pai e da mãe se mostraram estatisticamente significantes.

Por outro lado, em Nasu (2019), que faz uma análise sobre a relação entre o gênero e o desempenho na educação contábil, o estado civil e gênero aparecem como variáveis que influenciam o desempenho dos alunos.

E em Brandt et. al. (2020), que avaliou os fatores que influenciavam o desempenho dos estudantes em administração pública que obteve como resultado os fatores renda, escolaridade dos pais, gênero, raça, idade, financiamento dos estudos (se possui bolsa de estudos ou não), região e categoria da instituição que apareceram como variáveis que influenciaram positivamente o desempenho dos alunos.

Desta forma, a definição das variáveis a serem estudadas neste trabalho, foram selecionadas com base em algumas destas que já se mostraram significantes em trabalhos anteriores. As variáveis são modalidade, categoria, região, idade, sexo, estado civil, grau de escolaridade do pai, grau de escolaridade da mãe, renda, situação de trabalho e se foi bolsista de algum projeto ao longo da graduação (variável escolhida pelo autor).

Para descobrir os impactos dessas variáveis no desempenho dos estudantes, foi estimado um modelo de regressão logística baseado em Fávero et al. (2017). A técnica da regressão logística é utilizada para variáveis dependentes categóricas e é aplicada para estimar de ocorrência de um determinado evento em termos de probabilidade de ocorrência, identificando as características que cada grupo possui. O modelo testado foi, eq. (2):

$$\begin{aligned}
 \text{Desempenho} = & \alpha_0 + \beta_1 * \text{modalidade} + \beta_2 * \text{categoria} + \beta_3 * \text{região} + \beta_4 * \text{idade} \\
 & + \beta_5 * \text{sexo} + \beta_6 * \text{est. civil} + \beta_7 * \text{esc. Pai} + \beta_8 * \text{esc. Mãe} + \beta_9 * \text{renda} \\
 & + \beta_{10} * \text{sit. trabalho} + \beta_{11} * \text{bolsista}
 \end{aligned} \quad (2)$$

onde, α_0 é uma constante, *Desempenho* é a variável dependente e β_i ($i=1, 2, \dots, 11$) são os parâmetros obtidos para cada uma das variáveis escolhidas para explicar o desempenho dos alunos.

A variável categórica desempenho foi criada a partir da nota dos alunos dividindo os alunos em dois grupos em relação a média das notas, classificando como acima da média e abaixo da média.

Já para a análise da evolução do desempenho, existem alguns trabalhos que utilizam a metodologia hierárquica, multinível para analisar o efeito de algumas variáveis em relação ao desempenho dos alunos, instituições de ensino e evolução do desempenho ao longo do tempo. Como por exemplo, em Araujo (2021), é utilizado a modelagem multinível para identificar as variáveis que interferem no desempenho dos estudantes nos níveis de estudante, docente e instituição de ensino e estudar a influência da modalidade de ensino nesse quesito.

Para esta análise, foi utilizada a base de dados do Conceito ENADE dos anos de 2017 e 2019 que foram encontrados relacionando o curso, nome da instituição e município. Os dados foram extraídos do site do INEP, (INEP, 2021c), e para a análise foi criado um modelo hierárquico de regressão linear multinível com medidas repetidas, HLM2 com interceptos aleatórios para entender a influência da instituição no desempenho geral dos cursos de engenharia, conforme modelo abaixo, (Fávero et al., 2017), eq. (3):

$$\begin{array}{ll} \text{Nível 1} & \text{Conceito Enade}_{ij} = \beta_{0j} + \beta_{1j} * \text{Ano}_{ij} + \varepsilon_{ij} \\ \text{Nível 2} & \begin{cases} \beta_{0j} = \gamma_{00} + \nu_{0j} \\ \beta_{1j} = \gamma_{10} \end{cases} \end{array} \quad (3)$$

sendo, β_{ij} os interceptos, ε_{ij} os termos de erro, γ_{00} é o valor que se espera da variável dependente para cada observação, γ_{10} a mudança do valor esperado de acordo com a característica do grupo analisada e ν_{0j} os termos de erro que mostram que se há aleatoriedade nos interceptos.

O modelo multinível obtido foi comparado com o modelo HLM2 com interceptos aleatórios em relação a Área de Avaliação do curso, modelo OLS NULO e HLM2 Nulo para entender qual teve o melhor resultado.

O modelo multinível normalmente é utilizado quando se acredita em que nos dados há subgrupos, clusters, que influenciam a variável dependente de tal forma que se realizado a análise neles se obtém melhores resultados, ou seja, ao se utilizar o modelo multinível nesse trabalho espera-se compreender se há impactos na evolução da nota do conceito ENADE, de acordo com a área de avaliação do curso, instituição de ensino, ou se elas não interferem na evolução do desempenho.

Os dois modelos finais obtidos, modelo de regressão logístico e modelo multinível serão comparados a nível didático. O desenvolvimento desse trabalho foi realizado utilizando a linguagem R através da interface do RStudio® v. R 3.3.0+.

3 Resultados e Discussão

3.1 Perfil

Para melhor análise dos dados modelou-se as variáveis categóricas que na base estavam ajustadas em valores de A à E para os valores correspondentes ou aproximados do que foi disponibilizado no dicionário de variáveis, Tabela 2.

Tabela 2. Perfil dos estudantes de Engenharia

Variáveis	Categorias	Respostas	% De Participação
Categoria do Curso	Privado	64.753	67%
	Público	32.479	33%

Modalidade	Ead	1.963	2%
	Presencial	95.269	98%
Região	Centro-Oeste	5.982	6%
	Nordeste	16.179	17%
	Norte	4.619	5%
	Sudeste	53.292	55%
	Sul	17.160	18%
Idade	Numérica	Mínimo	20 anos
		Média	26,82 anos
		Máximo	72 anos
Sexo	Feminino	31.249	32%
	Masculino	65.983	68%
Estado Civil	Casado	13.244	14%
	Outro	83.988	86%
Raca	Amarela	2.495	4%
	Branca	58.399	60%
	Indígena	220	0%
	Preta	6.942	7%
Escolaridade_Pai	Sem Superior	70.554	73%
	Superior	26.678	27%
Escolaridade_Mãe	Sem Superior	62.836	65%
	Superior	34.396	35%
Renda	Acima De 10	13.133	14%
	Até 3 Salários	34.179	35%
	De 3 A 6	34.502	35%
	De 6 A 10	15.418	16%
Sittuação_Trabalhista	Não Trabalho	39.262	40%
	Trabalha	57.970	60%
Bolsista	Bolsista	24.726	25%
	Não Bolsista	72.506	75%
Desempenho	Alto	45.947	47%
	Baixo	51.285	53%

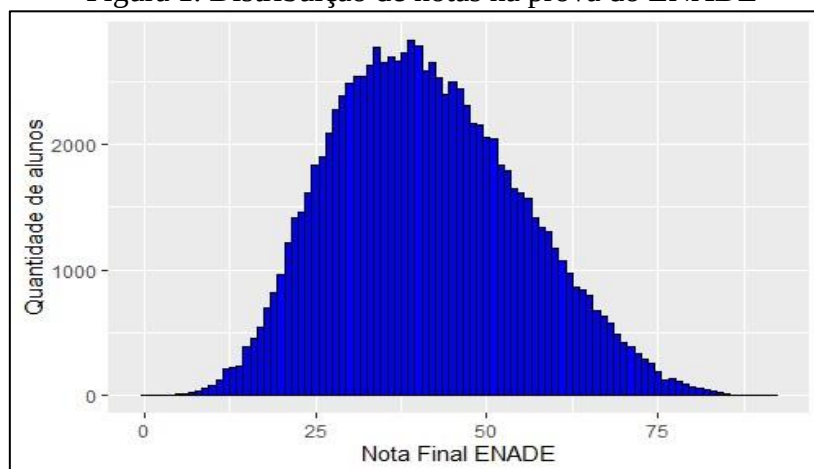
Fonte: Resultados originais da pesquisa.

Dos estudantes que realizaram a prova do ENADE em 2019, 67% estava inscrito em uma instituição de ensino privado e 98% na modalidade presencial. As regiões com maior número de estudantes são as regiões sudeste, sul e nordeste, respectivamente, sendo que o sudeste representa 55% do total de estudantes e já renda dos estudantes é na maioria dos casos de até 6 salários mínimos. Ao longo da graduação 60% dos alunos trabalharam e 75% não possuíam nenhum tipo de bolsa mérito. E na maioria dos casos os alunos eram brancos, do sexo masculino dos quais os pais não possuíam ensino superior. A faixa etária variou de 20 a 72 anos, porém em média os estudantes tinham 27 anos.

Esses resultados nos mostram que existe um perfil específico dos estudantes de engenharia e mostra que também que não há muita diversidade nos perfis quando falamos principalmente de gênero e raça.

A distribuição das notas dos estudantes na prova do ENADE se aproxima de uma curva normal, tendo como máxima a nota 91,6, mínimo 0,4 e média 41,26 e desvio padrão de 13,72, tal como demonstrado na Figura 1.

Figura 1. Distribuição de notas na prova do ENADE



Fonte: Resultados originais da pesquisa

3.2 Modelo de Regressão Logística

Utilizando o “software” RStudio® desenvolveu-se o modelo de regressão logística, primeiro considerando todas as variáveis e o segundo realizando o procedimento de “stepwise”, o qual é utilizado para retirar automaticamente do modelo as variáveis explicativas cujo parâmetro não se mostrar estatisticamente diferente de 0 de forma sucessiva até alcançar o melhor modelo (Fávero et al., 2017), tal como demonstrado na Figura 2.

Figura 2 - Teste da razão Verossimilhança para os modelos 1 e 2 com “stepwise”.

Likelihood ratio test					
Model 1: desempenho ~ (MODALIDADE + REGIAO + NU_IDADE + SEXO + EST_CIVIL + ESC_PAI + ESC_MAE + RENDA + SIT_TRAB + BOLSISTA + NT_GER + CAT_CURSO_Publico) - NT_GER					
Model 2: desempenho ~ REGIAO + NU_IDADE + SEXO + ESC_PAI + ESC_MAE + RENDA + SIT_TRAB + BOLSISTA + CAT_CURSO_Publico					
#Df	LogLik	Df	Chisq	Pr(>Chisq)	
1	17	-61048			
2	15	-61049	-2	1.7828	0.4101

Fonte: Resultados originais da pesquisa

Comparando os modelos, com o uso do procedimento “stepwise”, Figura 2, houve um incremento no logaritmo de máxima verossimilhança, que é um dos testes utilizado para comparação de modelos, além da saída de variáveis como modalidade e estado civil que já não se mostraram significantes no primeiro modelo, por possuírem um “p-value” de 0,546 e 0,220 respectivamente, ou seja maior do que 0,05.

A variável modalidade havia se mostrado significativo em trabalhos anteriores como em Meurer e Pereira (2020) para o curso de ciências contábeis, o que mostra que podem existir diferenças nos fatores que afetam o desempenho dos alunos para cursos e edições do ENADE diferentes. Desta forma o modelo final ficou o que está apresentado na Tabela 3 abaixo.

Tabela 3. Modelo final de regressão logística com “stepwise”

Variável Dependente: Desempenho	Est.	2,5%	97,5%	z-valor	p-valor
(Intercept)	1,468	1,359	1,5770	26,4210	0,000
Região Nordeste	0,133	0,069	0,1960	4,0860	0,000
Região Norte	-0,290	-0,373	-0,2070	-6,8290	0,000
Região Sudeste	0,114	0,056	0,1720	3,8650	0,000
Região Sul	0,258	0,195	0,3220	7,9940	0,000
Nu_idade	-0,050	-0,053	-0,0470	-32,9780	0,000
Sexo Masculino	0,188	0,159	0,2180	12,5050	0,000
Escolaridade_pai: superior	0,191	0,156	0,2250	10,8780	0,000
Escolaridade_mãe: superior	0,126	0,094	0,1570	7,7520	0,000
Renda até 3 salários-mínimos	-0,796	-0,844	-0,7470	-31,8420	0,000
Renda de 3 à 6 salários-mínimos	-0,517	-0,564	-0,4710	-21,7430	0,000
Renda de 6 à 10 salários-mínimos	-0,294	-0,345	-0,2420	-11,1970	0,000
Sit_trab trabalha	-0,222	-0,251	-0,1920	-14,7240	0,000
Bolsista não bolsista	-0,388	-0,421	-0,3540	-22,4400	0,000
CAT_cargo_Público	1,013	0,981	1,0450	61,5450	0,000

Fonte: Resultados originais da pesquisa

* Pseudo- R^2 (Cragg-Uhler) = 0.191

* Pseudo- R^2 (McFadden) = 0.111

* AIC = 122128.436

* BIC = 122271.030

Uma das formas de avaliar se o modelo consegue descrever de uma forma razoável o desempenho dos estudantes é calculando a acurácia do modelo. A acurácia pode ser obtida a partir da construção de uma matriz de confusão, que estipulando um valor percentual de corte para consideração de ocorrência de um determinado evento, conseguimos obter quais os erros e acertos do modelo se comparado com os valores reais.

No modelo obtido, Tabela 4, considerando um “cutoff” de 0.5, a acurácia obtida foi de 67,12%, ou seja, o modelo foi capaz de acertar 67% dos casos, o que pode ser considerado um bom resultado considerando que a sensibilidade e especificidade estão acima dos 50% e se compararmos com os resultados para os diferentes “cutoffs”.

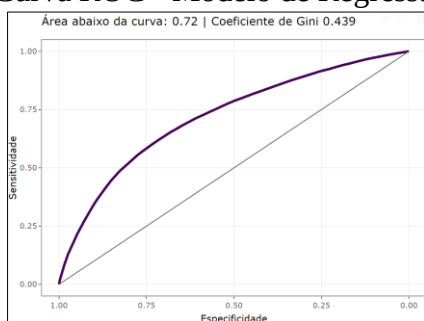
Tabela 4. Matriz de Confusão Modelo de Regressão Logística

Real	Classificação do modelo (“cutoff” = 0.5)		Comparativo - Percentual de Acerto				
	Nota > Média	Nota < Média	“Cutoff”	0.5	0.4	0.6	0.45
Nota > Média	27275	13054	Sensitividade	58,17%	71,73%	44,99%	64,38%
Nota < Média	19609	39403	Especificidade	75,11%	60,22%	84,87%	69,22%
			Acurácia	67,12%	65,65%	66,05%	66,94%

Fonte: Resultados originais da pesquisa.

Outro indicador para avaliar o modelo obtido é a análise da curva “Receiver Operator Characteristic” [ROC], Figura 3. Esta curva nos permite estudar a variação da especificidade e sensibilidade para diferentes valores de corte e a área da curva [ROC] dá a indicação do quão bom é um modelo, quanto mais próximo de 1 o valor, mais o modelo acerta, (Fávero et al., 2017). O que indica que o modelo desenvolvido tem uma boa taxa de acerto, obtendo uma área da curva [ROC] de 0,72.

Figura 3 - Curva ROC - Modelo de Regressão Logística



Fonte: Resultados originais da pesquisa.

Outra medida, utilizada para interpretar os resultados obtidos é a razão de chance. Ou seja, se alterarmos o valor de uma variável mantendo as demais constantes o quanto se altera a chance de ter um desempenho bom no ENADE, Fávero et al.(2017). Seu cálculo é dado pela equação abaixo, eq. (4):

$$Chance_{Y_i=1} = e^{Z_i} \quad (4)$$

sendo, Z_i os betas obtidos pelo modelo de regressão logística. Desta forma, calculou-se as razões de chance para cada uma das variáveis do modelo final e também o seu intervalo de confiança, Tabela 5.

Tabela 5. Razão de Chance das Variáveis Obtidas no Modelo de Regressão Logística

Variável Dependente: Desempenho	OR1	2,5%	97,5%
(Intercepto)	4,340	3,892	4,839
Nu Idade	0,951	0,949	0,954
Cat Curso Público	2,753	0,949	0,954
Região Nordeste	1,142	1,072	1,217
Região Norte	0,748	0,688	0,813
Região Sudeste	1,121	1,058	1,187
Região Sul	1,295	1,215	1,380
Sexo: Masculino	1,207	1,172	1,243
Escolaridade do pai: superior	1,210	1,169	1,252
Escolaridade da mãe: superior	1,134	1,098	1,171
Sit_Trab: Trabalha	0,801	0,778	0,825
Bolsista: não-bolsista	0,679	0,656	0,702
Renda até 3 Salários-mínimos	0,451	0,430	0,474
Renda de 3 à 6 Salários-mínimos	0,596	0,569	0,625
Renda de 6 à 10 Salários-mínimos	0,745	0,708	0,785

Fonte: Resultados originais da pesquisa.

Analisando as razões de chance temos que estudantes com renda de até 3 salários-mínimos tem chance de 54,9% menor de atingir um bom desempenho nas avaliações do ENADE se comparado com quem ganha acima de 10 salários. Conforme um ano é acrescentado na idade também reduz em 4,9% as chances de obter um bom desempenho, já o pai e/ou a mãe terem nível superior aumenta em 21% e 13% às chances do bom desempenho em relação aos pais que não tem, respectivamente.

Ter estudado em uma instituição de ensino superior pública é a variável que mais afeta o desempenho dos estudantes fazendo com que ele tenha quase 3 vezes mais chances de obter um bom desempenho. O fato de o estudante trabalhar ao longo da graduação e/ou não ter sido bolsista por mérito reduz em 19,9% e 32,1% as chances de um bom desempenho, respectivamente.

Já o estudante ser da região sul e/ou ser do sexo masculino aumentam e 29,5% e 20,7% às chances de obter um bom desempenho em relação aos alunos que são da região CO e/ou são do sexo feminino.

Comparando os resultados obtidos com outros estudos realizados, podemos perceber que muitas das variáveis que foram consideradas estatisticamente nesse estudo, tiveram significância em estudos anteriores considerando outras avaliações do ENADE para outros cursos e instituições Brandt et. al. (2020), Silva et. al. (2020) e Meurer e Pereira (2020). Entretanto os que mais se destacam para os alunos dos cursos de engenharia são a renda, se o aluno teve bolsa mérito e ter feito o ensino superior em instituição pública.

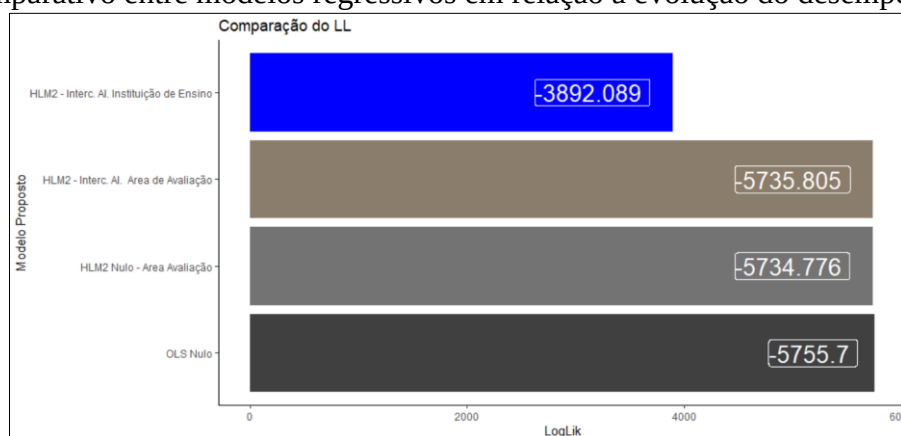
3.3 Modelo de Regressão Multinível

Na análise multinível foram utilizados uma amostra de 2111 cursos de instituições que estavam presentes nas avaliações de 2017 e 2019 e que tinham os mesmos dados de identificação e resultado válido de conceito ENADE, nota de 0 a 5, nos dois anos.

Com o uso da linguagem R na interface RStudio® desenvolveu-se também o modelo de regressão multinível, para essa análise foram feitos testes com diferentes modelos para compararmos os resultados com o modelo OLS Nulo, (Fávero et al., 2017).

Dentre os modelos avaliados, o que obteve o melhor logaritmo de máxima verossimilhança foi o modelo HLM2 com interceptos aleatórios considerando o nível 1 o ano do resultado ENADE e o nível 2 a Instituição de Ensino, Figura 4.

Figura 4 - Comparativo entre modelos regressivos em relação a evolução do desempenho dos cursos



Fonte: Resultados originais da pesquisa.

O resumo do modelo HLM2 com interceptos aleatórios considerando o nível 1 o ano do resultado ENADE e o nível 2 a instituição de ensino é dado pela Figura 5:

Figura 5 - Modelo Multinível HLM2 com interceptos aleatórios em relação a instituição de ensino

```
Linear mixed-effects model fit by REML
Data: BASE_CENADE
      AIC      BIC    logLik
7792.178 7817.568 -3892.089

Random effects:
Formula: ~1 | NOME_IES
(Intercept) Residual
StdDev:    0.7317447 0.5167088

Fixed effects: CONCC_ENADE ~ ANO
              Value Std.Error   DF   t-value p-value
(Intercept)  64.48863 16.047552 3665   4.018596 1e-04
ANO          -0.03090  0.007952 3665  -3.885123 1e-04
Correlation:
(Intr)
ANO -1

Standardized Within-Group Residuals:
      Min       Q1      Med       Q3      Max
-4.16249965 -0.56369943 -0.02406646  0.56349357  3.30636254

Number of Observations: 4222
Number of Groups: 556
```

Fonte: Resultados originais da pesquisa.

Logo, com base no modelo obtido na Figura 5, a equação da modelagem multinível é dada por eq. (5):

$$Desempenho_{ij} = 64,49 - 0,031 * Ano_{ij} + v_{0j} + \varepsilon_{ij} \quad (5)$$

Onde, v_{0j} um valor relativo a cada uma das instituições de ensino e ε_{ij} é o erro.

E analisando as variâncias estimativas dos componentes de efeitos aleatórios, conseguimos observar que existe significância estatística para o efeito aleatório nível instituição de ensino, pois o *p-value* é menor que 0,05, Tabela 6.

Tabela 6. Variâncias estimativas dos componentes de efeitos aleatórios do modelo multinível

	Componentes	Variância estimativa	Erro padrão	Z	p-valor
1	Var (v_{0j})	0,5354504	0,035	1.511.624,00	0,00
2	Var (e)	0,2669879	0,006	4.290.366,00	0,00

Fonte: Resultados originais da pesquisa.

Com os resultados obtidos é possível dizer que há uma estrutura hierárquica em relação aos resultados do conceito ENADE de cada curso das instituições, sendo os cursos agrupados pelas instituições de ensino a qual pertencem, o que possibilita interpretar de uma forma melhor em consideração à realidade e à aplicação de políticas públicas mais adequadas a cada realidade, Cruz (2010) e Finch et. al. (2019). Esses resultados obtidos são parecidos com os obtidos por Araújo (2021), em relação ao desempenho dos alunos de ciências contábeis.

4 Conclusões

Conforme visto no presente trabalho os estudantes dos cursos de engenharia são em sua maioria os que estão em instituições privadas, do sexo masculino, localizados principalmente na região sudeste, trabalham, não possuem bolsa por mérito e pai e mãe não possuem ensino superior.

Através do modelo de regressão logística obtido pode-se observar que das variáveis escolhidas o desempenho dos estudantes é influenciado, com uma acurácia alta, pelas variáveis: categoria, região, idade, sexo, estado civil, grau de escolaridade do pai, grau de escolaridade da mãe, renda, situação de trabalho e se foi bolsista por mérito.

Dentre as variáveis estudadas, ter feito ensino superior em uma instituição pública, ter a renda acima de 6 salários-mínimos, não trabalhar, ter sido bolsista por mérito, ser do sexo masculino, morar na região sul, ter os pais com ensino superior e ser jovem são as características que mais maximizam a probabilidade de o estudante possuir um alto desempenho na avaliação, revelando que as diferenças sociais influenciam diretamente no rendimento dos estudantes na prova do ENADE.

Desta forma, o modelo de regressão logística consegue estimar com uma boa acurácia a probabilidade de um desempenho acima da média dos estudantes de acordo com o perfil do mesmo e conseqüente pode ser utilizado para propor melhorias no ensino, tanto no ingresso de estudantes quanto em melhorar o desempenho dos perfis que, com base nos fatores que mais afetam, podem não gerar bom rendimento na prova do ENADE.

Já com a modelagem multinível realizada foi possível entender que existe uma relação entre os resultados dos cursos e a instituição de ensino a qual ele está atrelado, muito maior do que entre área de avaliação e o curso. Este resultado é importante, pois pode ser utilizado para a criação de medidas para a promoção e melhoria das instituições de ensino, principalmente daquelas que com o resultado do estudo estiverem impulsionando uma redução da nota do conceito ENADE dos cursos de engenharia que estão atrelados a ela.

Para futuros trabalhos com esse tema, sugiro aprofundamento da análise multinível acrescentando outros parâmetros das instituições e a realização do modelo regressivo em uma perspectiva multinível para captar a influência das características específicas no perfil e desempenho dos estudantes de engenharia.

5 Referências

Araujo, E. F. D. Relação entre desempenho de estudantes de Ciências Contábeis nas edições do ENADE e modalidade de ensino: uma análise multinível. Dissertação Pós-graduação em Contabilidade UFBA. Salvador. Disponível em: <https://repositorio.ufba.br/handle/ri/34104>. Acesso em: 01 maio 2022.

Brandt, J. Z., Tejedo-Romero, F., & Araujo, J. F. F. E. (2020). Fatores influenciadores do desempenho acadêmico na graduação em administração pública. *Educação e Pesquisa*, 46.

Cruz, C. C. M. S. D. (2010). *Modelos multi-nível: fundamentos e aplicações* (Doctoral dissertation).

Fávero, L. P., & Belfiore, P. (2017). *Manual de análise de dados: estatística e modelagem multivariada com Excel®, SPSS® e Stata®*. Elsevier Brasil.

Finch, W. H., Bolin, J. E., & Kelley, K. (2019). *Multilevel modeling using R*. Crc Press.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2009). *Análise multivariada de dados*. Bookman editora.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – INEP . 2020. Sinopse Estatística da Educação Superior 2019. Brasília: Inep.

Disponível em: <<http://portal.inep.gov.br/basica-censo-escolar-sinopse-sinopse>>. Acesso em: 05 jan. 2022.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – INEP. 2021a. Conceito ENADE. Disponível em: <<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/indicadores-educacionais/indicadores-de-qualidade-da-educacao-superior>>. Acesso em: 16 de out. 2021.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - INEP. 2021b. Exame Nacional de Desempenho dos Estudantes [ENADE]. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enade>>. Acesso em: 16 out. 2021.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - INEP. 2021c. Microdados ENADE: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enade>. Acesso em: 16 out. 2021.

Mallmann, C., Nasu, V., & Domingues, M. J. (2021). Relação entre a leitura de livros e o desempenho acadêmico: análise com discentes de ciências sociais aplicadas. *Revista de Educação e Pesquisa Em Contabilidade (REPeC)*, 15(2).

Medeiros Filho, A. E. C., Rodrigues, Y. S., Lopes, J. M., & Pontes Junior, J. A. F. (2019). Fatores associados ao desempenho discente no Exame Nacional de Desempenho dos Estudantes (ENADE): uma revisão integrativa. *Revista Expressão Católica*, 8(1), 88-96.

Meurer, A. M., & Pereira, V. H. (2020). Performance at Enade and the conditions of educational process from Accounting Sciences students. *Contextus–Revista Contemporânea de Economia e Gestão*, 18, 178-190.

Hideo Nasu, V. (2020). Investigação empírica acerca da relação entre sexo e desempenho acadêmico na educação contábil. *Advances in Scientific & Applied Accounting*, 13(1).

Neto, W. R. (2021). O uso de mineração de dados educacionais sob o ENADE como apoio ao processo de tomada de decisão de gestores do ensino superior. Universidade Federal do Ceará campus Quixadá. Disponível em: <http://www.repositorio.ufc.br/handle/riufc/58835>. Acesso em: 11 jan. 2022.

Silva, L. C. D., Campanelli, A., Silva, L. S. D., Silva, T. V. F. D., Silva, R. C. D., Garcia, R. D. R., & Magalhães, F. L. F. D. (2020). Graduação em Computação no Brasil: perspectiva a partir do Exame Nacional de Avaliação de Desempenho. *Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología*, (25), 57-67.

Vieira, A. M. D. P., & Schneiker, D. (2021). O perfil docente no ensino superior privado eo desempenho no Enade. *Revista Educação & Formação*, 6(2), 1-18.