



12º Congresso UFSC de Controladoria e Finanças  
12º Congresso UFSC de Iniciação Científica em Contabilidade  
5º UFSC *International Accounting Congress*

Certificamos que o trabalho intitulado **Desempenho Escolar No Brasil: Análise Dos Resultados Do Saeb E Dos Gastos Públicos Com A Educação** de autoria de **Luana Elen de Souza, Leonardo Flach, Arthur Frederico Lerner**, foi apresentado por **Luana Elen de Souza** no **12o. Congresso UFSC de Controladoria e Finanças & 5o. UFSC International Accounting Congress & 12o. Congresso UFSC de Iniciação Científica em Contabilidade**, realizado na cidade de Florianópolis - SC, Brasil, de 19/09/2022 a 20/09/2022, contabilizando carga horária de 2hs.

Prof. Edilson Paulo  
Coordenador Geral



Consulte os anais do  
Congresso pelo QR Code



## Desempenho Escolar no Brasil: análise dos resultados do SAEB e dos gastos públicos com a Educação

**Luana Elen de Souza**  
Universidade Federal de Santa Catarina (UFSC)  
*e-mail: luana.elen@hotmail.com*

**Leonardo Flach**  
Universidade Federal de Santa Catarina (UFSC)  
*e-mail: leonardo.flach@gmail.com*

**Arthur Frederico Lerner**  
Universidade Federal de Santa Catarina (UFSC)  
*e-mail: arthurlerner\_@hotmail.com*

### Resumo

O presente estudo é uma replicação científica que descreve a relação entre os gastos na educação, produto interno bruto, contextos educacionais e o desempenho escolar mensurado por meio das médias das notas mais recentes da Prova SAEB 2019 das escolas de ensino médio nas regiões geográficas e unidades da federação do Brasil, por meio do método de regressão linear múltipla. O SAEB é o sistema de avaliação da educação básica, o qual permite o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) gerar informações sobre o desenvolvimento da educação básica brasileira e o que pode estar interferindo na educação dos alunos. A amostra foi composta por 27 estados do Brasil e foram analisadas 11 variáveis contextuais, nas quais 2 compuseram o modelo de regressão linear múltiplo: Índice de Regularidade Docente (IRD) e Taxa de Abandono (TAB). Essas variáveis obtiveram uma correlação múltipla de 78,4% com a variável dependente NPS, e um p-valor menor que 0,05 que rejeitou a hipótese nula, indicando que pelo menos uma das variáveis possui uma ligação com a constante. Assim, possuindo um coeficiente de determinação ajustado ( $R^2$  ajustado) de 0,58, resultando que 58% das variações na nota da prova SAEB são explicadas pela variação conjuntas de ambas variáveis do modelo 2. O estudo frisa a importância dos contextos escolares, bem como os estímulos do professor, o estado da economia do país e os gastos com a educação das unidades federativas brasileiras como forma de reduzir os efeitos desfavoráveis da atuação do desempenho escolar.

**Palavras-chave:** Desempenho escolar; Educação; Investimento em educação.

**Linha Temática:** Outros temas relevantes em contabilidade





## 1 Introdução

Devido o descuido que o Brasil teve ao longo de 200 anos, mesmo com o alto investimento com gastos na educação (baseado em 5,7% do PIB) em comparação aos países desenvolvidos, não tem tido tanto retorno na qualidade da educação como esperado. Segundo o Doutor em Educação Gabriel Grabowski (2015), as realidades dos países são bem diferentes, pois pelos 500 anos de educação no Brasil, é apresentado uma demanda de 46% da nossa população pela educação, sendo umas das maiores entre o mundo. Isso devido à grande taxa de analfabetos e baixa escolaridade.

Sabendo que a avaliação do SAEB é um sistema para avaliação do desempenho da educação no Brasil desde os anos iniciais até o ensino médio, foi avaliado as médias das notas das escolas do ensino médio (tanto pública como privada) no ano de 2019, nas unidades federativas do Brasil. Verificando o contexto dos indicadores educacionais disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), junto com os gastos da educação e o produto interno bruto gerados no ano mais recente em cada estado do Brasil.

Por meio de uma análise quantitativa, com dados em corte, foi adotado para análise o modelo de regressão linear múltipla e avaliado as influências das 11 variáveis independentes adotadas como: Gastos da Educação (GDE); Produto Interno Bruto (PIB); Alunos Por Turma (APT); Horas-Aulas Diária (HAD); Taxa de Distorção Idade-Série (DIS); Taxa de Aprovação (TAP), Taxa de Abandono (TAB); Indicador de Regularidade do Corpo Docente (IRD); Indicador de Esforço Docente (IED); Indicador de Complexidade de Gestão da Escola (ICG); Adequação de Formação Docente (AFD), com a variável dependente Notas da Prova Saeb (NPS). Dessa forma, verificar se podemos rejeitar a hipótese nula com 5% de nível de significância e assim averiguar se pelo menos uma variável explica a variação do NPS.

O objetivo do trabalho é evidenciar uma influência entre os GDE, o PIB de cada estado do Brasil e outros contextos educacionais disponibilizados pelo INEP com a variável NPS em 2019 em cada unidade federativa Brasileira através do método de análise regressão linear múltipla. Para dessa forma, viabilizar as possíveis melhoras no desempenho educacional para os alunos do Brasil.

## 2 Revisão da literatura científica

O SAEB é o sistema de avaliação da educação básica, o qual permite o INEP gerar informações sobre o desenvolvimento da educação básica brasileira e o que pode estar interferindo na educação dos alunos. Essa prova é feita a cada 2 anos, por meio de questionários, avaliando os níveis de aprendizado dos estudantes.

Todas as informações contextuais juntadas com o SAEB, permite monitorar o aprimoramento das políticas educacionais com evidências, além de que as médias das notas, juntamente com as taxas de aprovação, abandono e reprovação, as quais são mensuradas pelo Censo Escolar, formam o Índice de Desenvolvimento da Educação Básica, IDEB.

Essa prova é realizada desde 1990, porém apenas a partir de 2019 que a estruturação da prova se desenvolveu e começou a abranger também a educação infantil, ensino fundamental e o ensino médio.

O Produto Interno Bruto, PIB, é uma soma de todos os bens e serviços finais produzidos anualmente em um país, medido pelo preço que chega ao consumidor. E o seu crescimento é um medidor de fluxo de novos bens e serviços finais produzidos durante o período analisado. Indicando





a economia de uma localidade, como caso de estudo as unidades federativas do Brasil, sendo que o PIB do Brasil em 2019 foi de 7,3 trilhões. Um país com maior PIB, tende obter um maior IDH, o qual é o Índice de Desenvolvimento Humano.

O Brasil investe cerca de 5,7% do PIB na educação, ocasionando num valor superior do investimento na educação que países desenvolvidos geralmente disponibilizam. Porém mesmo o gasto na educação brasileira sendo alto, o investimento por aluno acaba sendo baixo, logo é insuficiente para a demanda necessária pela quantidade de população que o Brasil tem em relação aos demais países desenvolvidos.

Segundo o Professor José (2015), da universidade de São Paulo, o qual é ex-presidente da Associação de Pesquisa em Financiamento da Educação (FINEDUCA), argumenta que devido o descuido que o Brasil teve com a educação ao longo do tempo, apenas aumentando o investimento em gastos brasileiros na educação em cerca de 6% a 7% do PIB, teria melhorias na educação e assim alcançando os países desenvolvidos.

Foi escolhido o artigo científico “Contexto e desempenho escolar: análise das notas na Prova Brasil das escolas capixabas por meio de regressão linear múltipla”, como base do presente artigo, através do site de periódicos científicos Scielo para replicação da metodologia, mudando a amostra para o ano de 2019 e acrescentando as hipóteses de que a economia do Brasil e os gastos com a educação também afetam as notas das provas SAEB. Com um grupo de amostras diferentes, as quais seriam as médias das influências e notas do SAEB nas escolas públicas e privadas em cada estado brasileiro, e não apenas delimitado a uma escola.

O artigo que foi baseado o presente trabalho, analisou a Prova Brasil de 2013, em relação com os contextos educacionais disponibilizados pelo INEP, no qual o objetivo era, segundo Américo e Lacruz (2017), “contribuir com a rede estadual de educação do Espírito Santo”. As amostras utilizadas foram 244 escolas, com 10 variáveis contextuais, relacionadas pelo método de regressão linear múltipla de stepwise. Resultando na influência sob a constante as seguintes variáveis: Índice de Regularidade Docente, Indicador de Esforço Docente e Taxa de Abandono. Indicando um grau de associação de 47,9% com a constante e a explicação de 22% da sua variação.

A Prova SAEB de 2019, junto ao site do INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira) proporciona indicadores educacionais no contexto da escola e do professor em relação aos alunos, nas quais podemos destacar relevantes para o presente artigo:

- Alunos por turma;
- Horas-aula diária;
- Taxa de distorção idade-série;
- Taxa de aprovação;
- Taxa de abandono;
- Indicador de regularidade do corpo docente;
- Indicador de complexidade de gestão da escola;
- Indicador de esforço docente;
- Adequação de formação docente;
- Médias das notas da prova SAEB nas unidades federativas e estados brasileiros.

As demais variáveis independentes como o PIB (Produto Interno Bruto) que mede a riqueza econômica de uma localidade, que no caso são as unidades da federação, foram encontradas no site do IBGE (Instituto Brasileiro de Geografia e Estatística), dados dos anos mais recentes relativos a 2019. E no Portal da Transparência da Controladoria-Geral da União foram encontrados os dados





de distribuição das despesas com a educação por localidade das unidades da federação brasileira, o qual é a última variável independente GDE (Gastos de Educação) objetos de estudo.

De tal modo, que o objetivo geral desse artigo é medir a influência dessas variáveis em relação as notas da prova SAEB em cada unidade federativa do Brasil pela análise o uso do método de regressão linear múltipla e assim destacar como melhorar o desempenho escolar dos alunos do ensino médio tanto das escolas públicas como as privadas nos estados brasileiros.

### 3 Método de pesquisa

Para atingir o objetivo de pesquisa desse presente artigo, foi feito uma pesquisa descritiva, qualitativa e de dados em corte. As variáveis utilizadas e suas descrições se encontram na Tabela 1.

A atual pesquisa é uma réplica metodológica do artigo disponibilizado no site da Scielo, “Contexto e desempenho escolar: análise das notas na Prova Brasil das escolas capixabas por meio de regressão linear múltipla”, feito pelos alunos Bruno Luiz Américo e Adonai José Lacruz, alunos da pós-graduação em Administração da Universidade Federal do Espírito Santo.

Os dados iniciais de pesquisa foram retirados dos sítios eletrônicos do IBGE e INEP relativos ao ano de 2019 (pois ainda não possuía dados de 2020), buscando por Indicadores Educacionais, seus contextos e nos sítios eletrônicos de transparência do Governo para achar dados da economia do Brasil, filtrando com os dados mais recentes possíveis e convenientes com a pesquisa.

**Tabela 1**

Descrição das variáveis do modelo inicial de pesquisa.

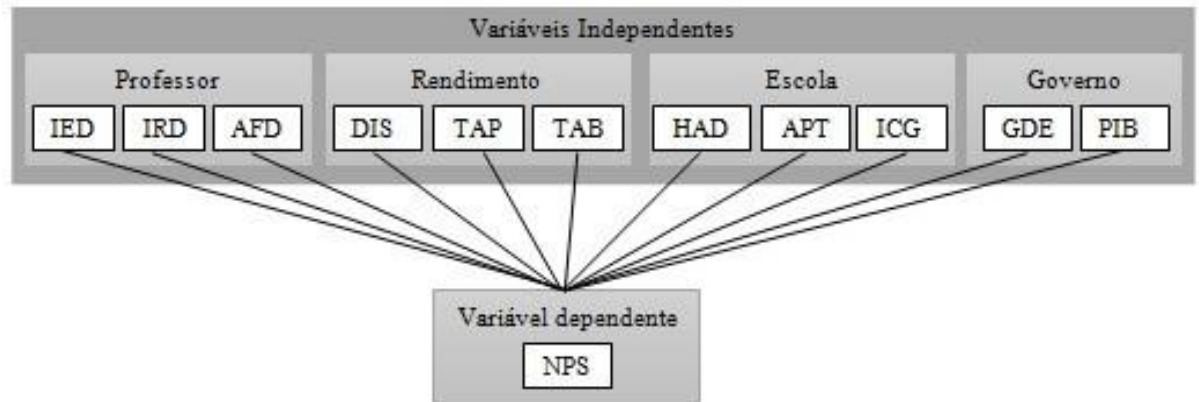
Variáveis				Descrição
<b>Independentes</b>	Notas da Prova SAEB de 2019	NPS	Intervalar (0 a 10)	Média das notas da prova no intervalo de 0 a 10, na qual, reflete o nível de aprendizado demonstrados pelo estudante avaliados, indicando a qualidade do ensino brasileiro.
	Alunos por turma	APT	Rácio	Média de matrículas pelas turmas nas escolas.
	Horas-aula diária	HAD	Rácio	Média de Horas-aula diária.
	Taxa de distorção idade-série	DIS	Rácio	Taxas de distorção idade-série.



Taxa de aprovação	TAP	Rácio	Taxa de aprovação de discentes.
Taxa de abandono	TAB	Rácio	Taxa de abandono dos discentes.
Indicador de regularidade do corpo docente	IRD	Intervalar (0 a 5)	Permanência dos professores nas escolas, o índice mede a irregularidade de vínculo do professor com a escola de 0 a 5. Foi utilizado os valores de 5.
Indicador de esforço docente	IED	Rácio	Percentual do nível de esforço dos docentes. Mensurado por número de escolas, turnos, alunos e etapas de lecionamento do professor. Quanto maior o indicador, maior é o esforço. Foi utilizado a soma dos níveis 4, 5 e 6 na variável explicativa.
Indicador de complexidade de gestão da escola	ICG	Intervalar (0 a 6)	A complexidade de gestão da escola é medida pelo número de turnos, etapas e modalidades disponíveis. Quanto maior a complexidade maior o nível. Foi utilizado o valor do nível 6 para variável explicativa.
Adequação de formação docente	AFD	Rácio	Vai do grupo 1 ao 5, na qual o 1 é o docente que possui formação superior com complementação na disciplina que leciona e o 5, é o qual o docente não possui o curso superior. Foi utilizado com variável explicativa o nível 1.
Gastos na Educação em 2019	GDE	Rácio	Valor gasto na educação pelo governo em cada estado do Brasil.
PIB dos estados do Brasil	PIB	Rácio	Último valor do Produto Interno Bruto disponível que indica a economia de cada estado do Brasil (2017).

Fonte: Elaborado pelos autores (2021).

Em seguida é mostrada na figura 1, a esquematização do problema de pesquisa, no qual separa a ideia de cada influencia em seu setor de atuação que pode contabilizar na nota média da prova SAEB em cada estado do país feita por todos os alunos do ensino médio das escolas do Brasil.



**Figura 1** - Modelo inicial de pesquisa  
 Fonte: Elaborado pelos autores (2021).

A pesquisa conta com 27 observações, que são as médias dos dados coletados de cada estado que compõe o Brasil. Relativos a uma média das notas da prova SAEB em cada estado, realizadas em todas as escolas que possuem ensino médio, sendo escola pública (municipal ou estadual) e escolas privadas.

Foi feito o processamento de dados através do software Stata (que foi utilizada para achar a correlação entre as variáveis utilizadas) e em grande maioria, foi utilizado também o software Gretl, onde foi feita toda a análise do modelo de regressão linear múltipla e suas validações dos pressupostos.

O método foi avaliado pela contribuição significativa do modelo de regressão, no qual foi identificado as correlações e hipóteses de independência entre a constante e as demais variáveis independentes, com nível de significância de 0,05 (5%).

## 4 Análise dos resultados

Para começar a analisar e fazer a extração de medidas foi feito a análise estatística descritiva das variáveis do modelo inicial de pesquisa na Tabela 2, para melhorar o entendimento dos resultados. Mostrando que possui 27 observações.

A variável dependente NPS, que seria as médias das notas da prova SAEB do Ensino Médio de cada estado brasileiro, apresenta dados que são bem pouco dispersas, apresentando um CV de apenas 7,65%. De acordo com cada média e Desvio padrão, as demais variáveis independentes apresentam uma dispersão considerada de baixa à moderada, com exceção de duas variáveis, as quais são a AFD (com CV=98,80%) e o PIB (com CV=169,26%), ambas estão com a dispersão bastante alta, ou seja, apresentam uma grande diferença em relação a cada estado.

**Tabela 2**  
 Estatística descritiva das variáveis

Variável	Média	Mediana	D.P.	CV (%)	Mín.	Máx.
IED	60,3	62,6	9,75	16,17	33,6	76,1
IRD	8,98	8,8	4,2	46,77	2,8	20

AFD	3,35	2,4	3,31	98,80	0,4	13,6
DIS	29,5	28,6	8,26	28,00	11,9	46,5
TAP	84,6	83,8	5,21	6,16	73,2	93,5
TAB	5,65	5,8	2,49	44,07	1,4	10,2
HAD	5,2	5,2	0,553	10,63	4,5	6,8
APT	29,2	29,2	3,19	10,92	21,9	35,2
ICG	2,79	2,7	1,49	53,40	0,5	7,3
GDE	1,87E+09	1,60E+09	1,71E+09	91,44	2,61E+08	6,97E+09
PIB	2,44E+05	1,13E+05	4,13E+05	169,26	1,21E+04	2,12E+06
NPS	4,69	4,65	0,359	7,65	4,1	5,27

Fonte: Elaborada a partir dos dados da pesquisa (2021).

A dispersão foi calculada pelo coeficiente de variação (CV), através da fórmula:

$$CV = (s/X) * 100$$

Onde,

s → é o desvio padrão;

X → é a média dos dados.

Assim, para analisar as correlações estatisticamente significativas entre as variáveis independentes com a variável dependente, foi feito na tabela 3 a análise de correlação de Pearson pelo software Stata.

**Tabela 3**

Matriz de Correlação de Pearson



Variáveis	NPS	IED	IRD	AFD	DIS	TAB	TAP	HAD	APT	ICG	GDE	PIB
NPS	1,0000											
IED	0,0433	1,0000										
IRD	0,3686**	0,1660**	1,0000									
AFD	0,1827**	0,2064**	0,1241*	1,0000								
DIS	0,7079**	0,1191*	0,5620**	-	0,0539	1,0000						
TAP	0,3589**	0,0905	0,0988	0,0115	-	0,5799**	1,0000					
TAB	0,7295**	0,0805	0,1157*	0,0859	0,7235	0,8225**	1,0000					
HAD	0,1432*	0,3854**	0,2861**	0,0331	-	0,1286*	0,4907**	0,4410**	1,0000			
APT	0,0218	0,1607**	0,2373**	0,4918**	0,1435**	0,3449**	0,2445**	0,3676**	1,0000			
ICG	0,2369**	0,0570	0,1301*	0,4325**	0,0142	0,2093**	0,0788	0,0443	0,1856**	1,0000		
GDE	0,4023**	0,1746**	0,1979**	0,2547**	0,1205*	0,0723	0,2188**	0,0359	0,0700	0,1710**	1,0000	
PIB	0,3461**	0,2600**	0,0318	0,2478**	0,4790**	0,2576**	0,3754**	0,0059	0,0691	0,0713	0,4453**	1,0000

\*\* e \* representam valores significativos ao nível 0,01 e 0,05, respectivamente.

Fonte: Elaborada a partir dos dados da pesquisa (2021).

Na análise de correlação, foi visto que 9 variáveis independentes demonstraram ter uma correlação significativa com a variável dependente e boas candidatas ao modelo de regressão. Segundo o critério de Miles e Shevlin (2001), podemos caracterizar as correlações altamente significativas quando os coeficientes de Pearson forem  $r > 0,5$ , moderadas quando forem  $0,3 < r \leq 0,5$  e baixa quando for  $0,1 < r \leq 0,3$ .

Sendo os valores mais próximos de -1 há uma correlação negativa e quando for mais próxima de +1 temos uma correção linear positiva. Intensificando sua força de correlação de acordo com os números sendo  $< -1$  e  $> +1$ .

Desse modo, podemos observar que as variáveis DIS, TAB tem uma alta correlação negativa com a variável dependente NPS, as variáveis GDE, PIB, TAP apresentam uma correlação moderada positivamente, a variável IRD teve uma correlação moderada negativamente e as

variáveis ICG, HAD (positivamente) e AFD (negativamente) obtiveram uma correlação baixa. As demais variáveis IED e APT não apresentaram correlação com a variável dependente.

Assim os fortes candidatos para o modelo de regressão são as variáveis DIS, TAB, GDE, PIB, TAP, ICG, IRD, HAD e AFD. Foram escolhidos 4 modelos para implementar a análise de regressão linear múltipla, nas quais não foram utilizados o DIS por apresentar um p-valor alto, e as variáveis HAD, ICG e AFD por apresentarem uma correlação baixa.

**Tabela 4**

Resumo dos Modelos de Regressão Linear Múltipla.

Modelo	Preditores	R	R <sup>2</sup>	R <sup>2</sup> ajustado	Erro-padrão	p-valor(F)
1	(constante), TAB	0,729475	0,532134	0,513419	0,250263	0,000016
2	(constante), TAB, IRD,	0,783567	0,613977	0,581809	0,232010	0,000011
3	(constante), TAB, IRD, TAP,	0,891747	0,795212	0,768500	0,172622	4,28e-08
4	(constante), TAB, IRD, TAP, PIB, GDE.	0,903479	0,816275	0,761158	0,175338	2,01e-06

Fonte: Elaborada a partir dos dados da pesquisa (2021).

Foi feita a verificação da validação dos pressupostos do modelo de regressão linear múltipla. A primeira verificação foi avaliada na Tabela 5 se houve problemas de colinearidade, mostrando se houve ausência de multicolinearidade a partir dos resultados da VIF (Variance inflation) e Condition index. Se o VIF for acima de 10, indica que a multicolinearidade pode estar influenciando as estimativas de mínimos quadrados do modelo.

**Tabela 5**

Validação dos Pressupostos – Regressores.

Modelo	Multicolinearidade			
	Variável	VIF	Dimensão	Condition Index
1	(constante)		1	1,000
	TAB	1,000	2	4,354
2	(constante)		1	1,000
	TAB	1,014	2	4,354
	IRD	1,014	3	6,494
3	(constante)		1	1,000
	TAB	3,102	2	5,041
	IRD	1,014	3	5,801
	TAP	3,091	4	84,855

4	(constante)		1	1,000
	TAB	4,274	2	2,614
	IRD	1,119	3	4,411
	TAP	4,000	4	6,523
	IED	1,149	5	7,543
	GDE	1,712	6	18,793
	PIB	1,532	7	119,841

Fonte: Elaborada a partir dos dados da pesquisa (2021).

De acordo com a análise feita na Tabela 5, foi visto que nenhuma das variáveis apresentou um VIP superior de 10, logo, há ausência de multicolinearidade nas variáveis independentes utilizadas. E pela a Condition Index, segundo Gujarati (2000), o qual diz que terá sérios problemas de multicolinearidade para valores maiores de 30, apresentou valores superiores a 30 nos modelos 3 e 4.

Desse modo, foi visto que o TAP e o PIB amplificam os efeitos da multicolinearidade nos modelos, e assim influenciando e reduzindo a correlação com as demais variáveis do modelo em relação à variável dependente.

A próxima análise de pressupostos feitas do modelo de regressão linear múltipla são os temas de perturbação estocástico, o qual verifica o teste de homocedasticidade de Breusch- Pagan e o teste de Normalidade de resíduos, ambas aplicadas pelo software Gretl.

**Tabela 6**

Validação dos Pressupostos – Termo de Perturbação Estocástico.

Modelo	Homocedasticidade Breusch-Pagan		Normalidade dos Resíduos	
	LM	p-valor	R <sup>2</sup>	p-valor
1	1,034179	0,309179	1,521	0,4674
2	0,554153	0,757997	3,123	0,20985
3	1,889358	0,595685	3,391	0,1835
4	6,200627	0,401095	2,358	0,30754

Fonte: Elaborada a partir dos dados da pesquisa (2021).

Pelo teste de Normalidade dos Resíduos, nenhum dos modelos podemos rejeitar a hipótese nula de normalidade da distribuição do termo estocástico, podendo assumir o pressuposto de normalidade.

No teste de Homocedasticidade de Breusch-Pagan também não permite rejeitar a hipótese nula de homocedasticidade do termo de perturbação estocásticos, devidos seus altos valores de p-valor.

Assim podemos verificar na tabela 4, que a variável TAB (que é a taxa de abandono) eo IRD (índice de regularidade docente) do modelo 2, tiveram uma correlação múltipla de 78,4% com a variável dependente NPS (nota da prova SAEB). Ocasionalmente num coeficiente R<sup>2</sup> ajustado de 0,58, demonstrando que 58% da divergência da nota da prova SAEB são explicadas pela divergência de ambas variáveis independentes escolhidas nesse modelo.

E por mais que os modelos 3 e 4, apresentaram uma porcentagem maior de correlação múltipla, que seria o grau de associação entre as variáveis, os dois modelos apresentaram problema de multicolinearidade, demonstrado na tabela 5, por isso foi adotado o modelo 2.

Com isso, foi feito a análise de variância ANOVA na tabela 7, o qual mostra um teste estatístico para ajuste do modelo em questão do teste F, mostrando se o modelo permite ou não rejeitar a hipótese nula (H0) e que pelo menos uma das variáveis independentes tenha influência sobre a variável dependente NPS.

**Tabela 7**  
ANOVA

Modelo		Soma dos Quadrados	gl	Quadrado da média	R <sup>2</sup>	F	p-valor	
1	(constante), TAB	Regressão	1,78088	1	1,78088	0,532134	28,4341	1,58E-05
		Resíduo	1,56579	25	0,0626318			
		Total	3,34667	26	0,128718			
2	(constante), TAB, IRD	Regressão	2,05478	2	1,02739	0,613977	19,0862	1,09E-05
		Resíduo	1,29189	24	0,0538289			
		Total	3,34667	26	0,128718			
3	(constante), TAB, IRD, TAP	Regressão	2,66131	3	0,887105	0,795212	29,7704	4,28E-08
		Resíduo	0,68536	23	0,0297982			
		Total	3,34667	26	0,128718			
4	(constante), TAB, IRD, TAP, IED, GDE, PIB	Regressão	2,73181	6	0,455301	0,816275	14,8098	2,01E-06
		Resíduo	0,614867	20	0,0307433			
		Total	3,34667	26	0,128718			

Fonte: Elaborada a partir dos dados da pesquisa (2021).

Analisando a Tabela 7, verificamos que com o uso das variáveis TAB e IRD, calculando nos valores da soma de quadrados, dividindo a regressão pelo total do modelo 2, temos 61% de redução do erro quadrado que ocorreria se fosse usada apenas a média da variável NPS para prever a variável dependente. O qual tem uma redução significativa se avaliar o valor do teste F de 19,0862, com o p-valor de 1,09E-05.

Assim, vemos na tabela 8 os coeficientes das variáveis apresentadas no modelo de regressão linear múltipla feita no *software* Gretl.

**Tabela 8**  
 Coeficientes

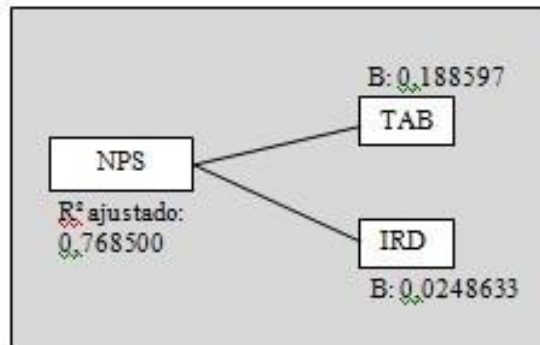
Modelo		Coeficientes da regressão		Significância estatística	
		B	Erro-padrão	T	p-valor
1	(constante)	5,27812	0,121179	43,56	<0,0001
	TAB	$\hat{\alpha}^{\wedge}0,104910$	0,0196743	$\hat{\alpha}^{\wedge}5,332$	<0,0001
2	(constante)	5,47221	0,141505	38,67	<0,0001
	TAB	$\hat{\alpha}^{\wedge}0,100117$	0,0183627	$\hat{\alpha}^{\wedge}5,452$	<0,0001
	IRD	$\hat{\alpha}^{\wedge}0,0246267$	0,0109173	$\hat{\alpha}^{\wedge}2,256$	0,0335
3	(constante)	10,3387	1,08378	9,539	<0,0001
	TAB	$\hat{\alpha}^{\wedge}0,188597$	0,0239014	$\hat{\alpha}^{\wedge}7,891$	<0,0001
	IRD	$\hat{\alpha}^{\wedge}0,0248633$	0,00812293	$\hat{\alpha}^{\wedge}3,061$	0,0055
	TAP	$\hat{\alpha}^{\wedge}0,0515872$	0,0114343	$\hat{\alpha}^{\wedge}4,512$	0,0002
4	(constante)	9,34661	1,28269	7,287	<0,0001
	TAB	$\hat{\alpha}^{\wedge}0,168422$	0,028496	$\hat{\alpha}^{\wedge}5,910$	<0,0001
	IRD	$\hat{\alpha}^{\wedge}0,0286035$	0,00867038	$\hat{\alpha}^{\wedge}3,299$	0,0036
	TAP	$\hat{\alpha}^{\wedge}0,0422967$	0,0132134	$\hat{\alpha}^{\wedge}3,201$	0,0045
	IED	0,00104518	0,00378058	0,2765	0,785
	GDE	3,72E-11	2,63E-11	1,414	0,1727
	PIB	$\hat{\alpha}^{\wedge}2,81514e-08$	1,03E-07	$\hat{\alpha}^{\wedge}0,2729$	0,7877

Fonte: Elaborada a partir dos dados da pesquisa (2021).

Na tabela acima foi verificado que a variável de gastos com a educação (GDE), obteve um p-valor 0,1727, demonstrando um valor maior que o nível de significância de 0,05, ou seja, a hipótese de que os coeficientes são iguais a zero não é válida, não tendo influência na variável dependente estatisticamente significativa. Da mesma forma acontece com a variável independente PIB (Produto Interno Bruto de cada estado do Brasil).

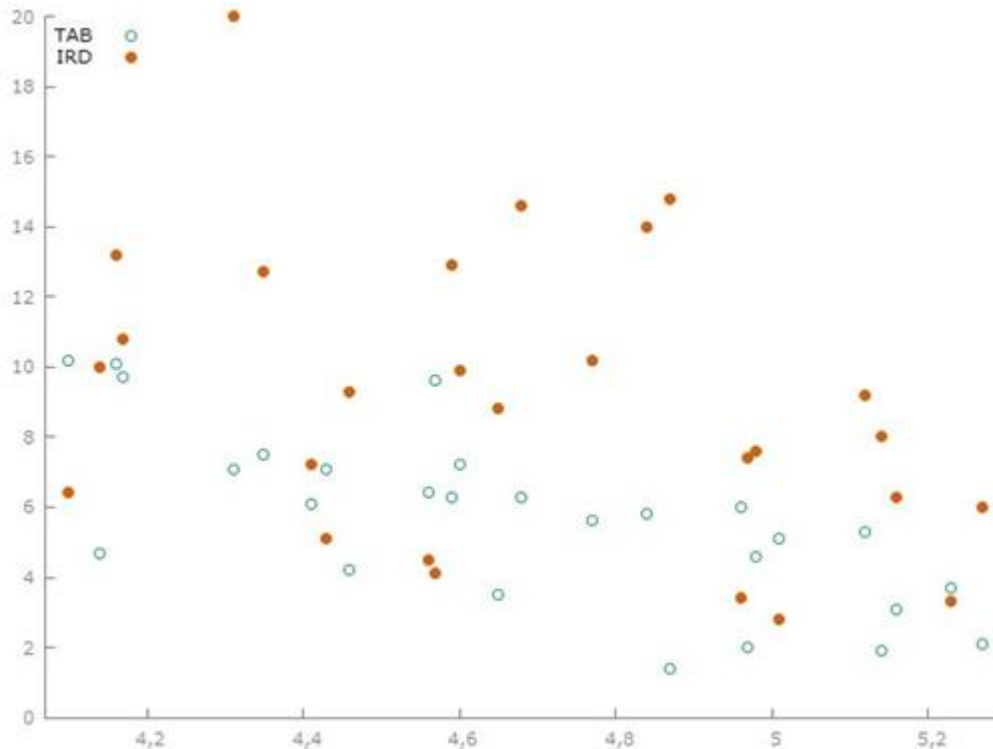
A constante B mede a importância da variável individual na previsão geral da nota da prova SAEB. E a variável mais relevante foi TAB, que é a taxa de abandono nas escolas de cada estado Brasil no Ensino Médio, em seguida da variável IRD (índice de regularidade docente), seguindo todos os critérios previstos em cada teste feito anteriormente.

Com tudo, podemos apresentar que o modelo 2 seria as melhores variáveis independentes que influenciam na variável dependente NPS, como podemos ver no modelo final na figura 2, abaixo:



**Figura 2 - Modelo Final**  
 Fonte: Elaboração própria (2021).

Na figura 3, é possível analisar que com os menores valores de IRD e TAB, maior é as médias das notas da prova SAEB nos estados do Brasil.



**Figura 3 - Gráfico de dispersão do Modelo 2.**  
 Fonte: Dados retidos do IBGE e INEP, gráfico elaborado pelo autor.

Quanto maior o IRD, maior é a irregularidade do corpo docente nas escolas, e quanto maior o TAB, maior a taxa de abandono dos alunos, os quais resultam em notas menores da prova SAEB.



## 5 Considerações finais

Observa-se que o estudo teve uma diferença de resultado em relação ao “Contexto e desempenho escolar: análise das notas na Prova Brasil das escolas capixabas por meio de regressão linear múltipla”, o qual resultou em (além das variáveis encontradas nesse artigo) mais uma variável independente no modelo, que seria o IED (indicador de esforço docente), porém nesse presente artigo essa variável não apresentou influência significativa na variável independente NPS (Notas das provas SAEB em cada estado do Brasil).

O objetivo do estudo era evidenciar se havia influência entre os Gastos da Educação (GDE), o Produto Interno Bruto (PIB) de cada estado do Brasil e outros contextos educacionais disponibilizados pelo INEP com a variável NPS (notas da prova SAEB), porém as análises e testes não demonstraram influência significativa com as variáveis PIB e GDE, pois ambos apresentaram elevados valores de p-valor, não rejeitando a hipótese nula e não tendo uma margem de segurança significativa.

Considerando um nível de significância de 5%, foi visto que as influências significativas nas notas da prova SAEB foram através das variáveis IRD e TAB do Modelo 2 de regressão linear múltipla, apresentando uma correlação múltipla de 78,4% com a variável dependente NPS, e um p-valor menor que 0,05 que rejeitou a hipótese nula, indicando que pelo menos uma das variáveis possui uma ligação com a constante. Assim, possuindo um coeficiente de determinação ajustado ( $R^2$  ajustado) de 0,58, resultando que 58% das variações na nota da prova SAEB são explicadas pela variação conjuntas de ambas variáveis do modelo 2 (conforme dados da tabela 4).

Outro fator que poderia ser avaliado em futuros trabalhos seria a taxa de desempregos, renda per capita e a média de remuneração de cada professor em cada estado, os quais poderiam ter uma ligação com o desenvolvimento dos alunos e nas notas da prova SAEB.

## Referências

- Américo, B. L., & Lacruz, A. J. (2017). Contexto e desempenho escolar: análise das notas na Prova Brasil das escolas capixabas por meio de regressão linear múltipla. *Revista de Administração Pública*, 51, 854-878.
- Controladoria-Geral da União (2022). Educação: Distribuição das despesas com educação por localidade. *Porta da Transparência*. <http://www.portaltransparencia.gov.br/funcoes/12-educacao?ano=2019>
- Flach, Leonardo (2020). Análise de Regressão Múltipla. UFSC, Departamento Ciências Contábeis, disciplina contabilometria, p. 01-55, 2020.
- Grabowski, Gabriel (2015). Quanto e como é o gasto do Brasil com educação: Por que aplicação não reverte em qualidade?. FUNDACRED. <https://www.fundacred.org.br/site/2019/04/16/quanto-e-como-e-o-gasto-do-brasil-com-educacao/>
- Gujarati, d. (2000). *Econometria básica*. São Paulo/SP: Ed.





Ministério da Educação (2022). Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). <https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/ideb/resultados>

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). <http://portal.inep.gov.br/indicadores-educacionais>

Miles, J., & Shevlin, M. (2001). Applying regression and correlation: A guide for students and researchers. Sage.

Fundacred (2019). Quanto e como é o gasto do Brasil com educação (2019). <https://www.fundacred.org.br/site/2019/04/16/quanto-e-como-e-o-gasto-do-brasil-com-educacao/>

Instituto Brasileiro de Geografia e Estatística – IBGE (2020). Produto Interno Bruto (PIB). <https://www.ibge.gov.br/explica/pib.php>

