

NOTA TÉCNICA FUNDAJ-NEES 02.2023

## Proposição de índice de classe de renda de alunos do ensino público de Pernambuco.

Data: 21/09/2023, Recife

**Autor:**

**Luís Henrique Romani de Campos**

A presente nota técnica tem por objetivo propor um indicador de renda familiar de alunos de escolas públicas de Pernambuco a partir da posse de bens em seus domicílios. Há uma larga bibliografia mostrando a correlação positiva entre a renda e o desempenho acadêmico. Assim, qualquer estudo empírico que trabalhe com resultados de avaliações de larga escala precisam capturar a renda de alguma forma para que esta seja uma variável de controle dos modelos estatísticos. A principal avaliação de larga escala do país é a Prova Brasil, que faz parte do Sistema Nacional de Avaliação da Educação Básica (SAEB). Nesta prova as crianças respondem perguntas sobre a posse de bens de sua residência. Esta abordagem é melhor do que a pergunta direta sobre a renda porque em geral as crianças não possuem informação sobre a renda monetária de seus pais. Contudo, gera dificuldades para a proposição de modelos estatísticos, pois são várias perguntas qualitativas para serem incorporadas.

A possibilidade de estabelecer a renda familiar pela educação dos pais é descartada porque a literatura mostra que há forte correlação entre desempenho escolar e a escolaridade dos pais. Então, construir um indicador de renda a partir da escolaridade pode gerar multicolinearidade nos modelos estatísticos e, no limite, impedir a identificação se um determinado efeito veio da renda ou da escolaridade dos pais.

A saída é a construção de um único indicador de renda seguindo a posse de bens. Uma possibilidade seria utilizar indicadores existentes em outras áreas do saber, notadamente o marketing, que também utiliza a posse de bens para caracterizar as famílias e direcionar suas ações. Ocorre que estes indicadores são estipulados observando a população como um todo e não um recorte desta. Em estudos cuja população seja composta por alunos da rede pública de educação há um viés para as rendas da base da estrutura social, dado que culturalmente alunos das classes mais ricas são matriculados na rede privada<sup>1</sup>.

Nesta nota técnica foi utilizada uma abordagem de análise multivariada, em particular a análise de clusters, para explorar o bando de dados dos alunos que realizaram a Prova Brasil nos anos de 2011 e 2019, no estado de Pernambuco, com o intuito de verificar quais os grupos de posse de bens emergem destes dados. Além disto, verificar se estes grupos apresentam desempenhos escolares distintos. Uma vez estabelecidos os grupos cujo desempenho tenha sentido empírico, é

---

<sup>1</sup> Existem exceções como as escolas de aplicação de universidades federais e estaduais ou, no ensino médio, os Institutos Federais, que dada sua qualidade diferenciada acaba atraindo alunos com rendas maiores. Mas estes casos não compõem a maioria dos alunos.

possível propor um indicador simples de ser replicado.

Na Tabela 1 estão explicitadas as posses de bens dos alunos do quinto ano que responderam ao questionário em 2011. Alguns pontos de destaque são: i) todas as famílias possuem ao menos uma televisão, o que já não ocorre com geladeira e banheiro; ii) o número de famílias com computador com acesso à internet é baixo em 2011; iii) possuir carro e contratar empregada todos os dias úteis são objetos de luxo.

**Tabela 1 – Percentuais de posse de bens alunos do 5º ano - 2011**

Bem	Uma*	Duas**	Três***	Mais de 3	Não tem
TV	52,09	47,91			
Geladeira	86,69	7,11			6,20
Lava roupa	45,52				54,47
Carro	19,92	3,24	1,45		75,38
Computador	23,27	5,95			70,77
Banheiro	76,23	13,51	1,97	1,01	7,27
Empregada	5,07	4,27	1,63		89,02

Fonte: Elaboração própria a partir de dados do INEP/SAEB.

\* em computador esta coluna indica computador com ligação com a internet. Para empregada indica uma diarista até duas vezes por semana.

\*\* em computador esta coluna indica computador sem ligação na internet. Para empregada indica uma, todos os dias úteis.

\*\*\* Para empregada esta coluna indica duas ou mais empregadas todos os dias úteis.

Na Tabela 2 a análise está focada nos alunos do nono ano de 2011. A maioria dos indicadores possui estrutura semelhante à do quinto ano, mas destacando-se que as diferenças são mais no sentido de maior posse de bens. Por exemplo, a proporção de famílias com duas televisões é maior e as proporções de famílias sem geladeira ou banheiro são menores. Os únicos indicadores menores são lava roupa e empregada.

**Tabela 2 – Percentuais de posse de bens alunos do 9º ano - 2011**

Bem	Uma*	Duas**	Três***	Mais de 3	Não tem
TV	42,98	57,02			
Geladeira	90,43	5,51			4,05
Lava roupa	36,96				63,03
Carro	20,36	2,63	0,82		76,18
Computador	29,26	6,98			63,75
Banheiro	75,49	15,84	1,98	0,66	6,02
Empregada	3,19	3,90	0,63		92,27

Fonte: Elaboração própria a partir de dados do INEP/SAEB. (as observações seguem as da Tabela 1)

Seguindo o raciocínio, a Tabela 3 apresenta a posse de bens dos alunos do terceiro ano do ensino médio. Manteve-se o padrão de posse, inclusive do maior nível de posse de bens, o que pode sugerir que a renda esteja ligada à manutenção do aluno até o final do ensino médio. Ou em outros termos, a falta de renda pode ser fator para o abandono escolar.

**Tabela 3 – Percentuais de posse de bens alunos do 3º ano do ensino médio - 2011**

Bem	Uma*	Duas**	Três***	Mais de 3	Não tem
TV	40,47	59,53			
Geladeira	92,65	4,10			3,24
Lava roupa	33,52				66,47
Carro	21,59	2,59	0,79		75,02
Computador	34,34	8,85			56,80
Banheiro	75,97	15,78	2,44	0,94	5,47
Empregada	3,39	3,32	0,21		93,07

Fonte: Elaboração própria a partir de dados do INEP/SAEB. (as observações seguem as da Tabela 1)

Em 2019, o questionário sobre a posse de bens apresentou ligeira alteração. Deixou-se de perguntar sobre a existência de empregada doméstica e abriu-se a possibilidade para a resposta “não tem” em todos os bens, o que não ocorreu em 2011 com televisão, por exemplo. Embora no questionário de 2011 existisse a categoria de resposta “não tem” para posse de bens, todos os estudantes entrevistados nesse ano mencionaram a existência de no mínimo um aparelho de televisão. Pela Tabela 4 pode-se ver que possuir lava roupa, carro e computador são marcadores de renda, além da posse de mais de uma televisão. De certa forma a estrutura de 2019 repete a de 2011, com a diferença de que o percentual de residências sem banheiro e sem geladeira reduziu de forma significativa.

**Tabela 4 – Percentuais de posse de bens alunos do 5º ano - 2019**

Bem	Um	Dois	Três ou mais	Não tem
TV	57,17	30,08	9,58	3,16
Geladeira	84,62	9,87	2,69	2,82
Lava roupa	30,52			69,47
Carro	26,37	4,27	2,20	67,15
Computador	26,07	7,22	3,45	63,26
Banheiro	74,11	19,09	4,10	2,68

Fonte: Elaboração própria a partir de dados do INEP/SAEB.

Procedeu-se a análise da posse de bens para o nono ano e para o terceiro ano do ensino médio, seguindo o procedimento de 2011. Em 2019 repete-se o padrão de quanto mais elevado o ano de estudo do aluno maior a posse de bens. Estas informações são úteis para a elaboração de um indicador síntese porque demonstram que há alguma manutenção no padrão de posse de bens ao longo do tempo, o que permite utilizar a mesma regra de cálculo, pelo menos desde 2011 até os dias atuais.

**Tabela 5 – Percentuais de posse de bens alunos do 9º ano - 2019**

Bem	Um	Dois	Três ou mais	Não tem
TV	57,52	30,21	9,10	3,17
Geladeira	88,17	8,83	1,74	1,26
Lava roupa	31,33			68,67
Carro	28,26	3,64	1,32	66,77
Computador	30,82	8,05	3,14	57,99
Banheiro	72,66	22,16	3,72	1,45

Fonte: Elaboração própria a partir de dados do INEP/SAEB.

**Tabela 6 – Proporções de posse de bens alunos do 3º ano do ensino médio - 2019**

Bem	Um	Dois	Três ou mais	Não tem
TV	64,49	24,28	6,95	4,27
Geladeira	92,76	5,51	0,74	0,97
Lava roupa	40,77			59,23
Carro	28,66	2,92	0,68	67,74
Computador	37,05	7,92	2,50	52,52
Banheiro	72,69	22,23	3,63	1,42

Fonte: Elaboração própria a partir de dados do INEP/SAEB.

Para verificar se existem grupos distintos de alunos, considerando o quesito posse de bens foram realizadas análises de cluster utilizando o procedimento k-means<sup>2</sup>. Como os dados são qualitativos (posse) a medida de distância utilizada foi a do tipo gowe3r. Considerando que em 2019 não existe a pergunta sobre empregada, não foi considerada esta variável para a análise. Foram realizadas 10 amostras aleatórias de 1.000 alunos e omitidos aqueles que não responderam as questões de posse de bens para o conjunto de alunos do quinto ano de 2011. Para cada uma destas amostras foram realizados testes gráficos para verificar qual o melhor número de clusters para aquela amostra. Em 60% das amostras o melhor número de clusters foi 5.

A Tabela 7 foi elaborada a partir de uma destas amostras e resume as características de cada um destes grupos. A coluna Cluster indica qual a numeração do grupo foi estabelecida no algoritmo. A coluna N apresenta o número de alunos que foram classificados naquele grupo. A coluna POR apresenta a média da proficiência em Português, enquanto que a MAT apresenta a proficiência em matemática. Não foram realizados teste de diferença de média. O cômputo foi para verificar se haveria alguma diferença para que fizesse sentido manter número tão grande de categorias de renda.

<sup>2</sup> Trata-se de algoritmos de mineração de dados não supervisionado. Ele classifica cada elemento em um grupo, considerando suas similaridades. Os grupos são formados para serem o mais heterogêneo possível. K representa o número de agrupamentos.

<sup>3</sup> A distância de Gower pode ser usada para medir o quanto diferente dois registros são. Estes registros podem conter dados lógicos, numéricos, categóricos e até textuais. A distância calculada é um número entre 0 (dois registros idênticos) e 1 (dois registros totalmente diferentes).

**Tabela 7 – Características por grupos a partir de amostra aleatória de 1.000 alunos**

Cluster	N	POR	MAT	TV	GEL	LRO	CAR	COM	BAN
1	220	168,2	181,9	100% possuem duas TVs	14,1% não possuem	100% não possuem	83,4% não possuem	83,6% não possuem	12,7% não possuem
2	124	178,5	197,2	30,6% possuem duas TVs	20,2% possuem duas	78,2% possuem	64,5% possuem	73,4% possuem com internet	57,3% possuem dois ou mais
3	125	175,1	184,3	100% possuem duas TVs	93,6% possuem uma	100% possuem	80,8% não possuem	75,2% não possuem	6,7% não possuem
4	221	162,9	182,2	100% possuem uma TV	9,95% não possuem	100% não possuem	91,8% não possuem	87,3% não possuem	15,4% não possuem
5	151	159,2	186,0	100% possuem uma TV	3,31% não possuem	100% possuem	82,8% não possuem	74,8% não possuem	2,4% não possuem
<b>Total</b>	<b>841</b>	<b>167,7</b>	<b>185,3</b>	<b>45,5%</b> <b>possuem</b> <b>duas</b> <b>TVs</b>	<b>7,6%</b> <b>não</b> <b>posuem</b>	<b>44,4%</b> <b>possuem</b>	<b>76,6%</b> <b>não</b> <b>possuem</b>	<b>71,6%</b> <b>não</b> <b>possuem</b>	<b>9,3%</b> <b>não</b> <b>possuem</b>

Fonte: Elaboração própria a partir de dados do INEP/SAEB.

Nitidamente o grupo 2 é o que apresenta o maior nível de posse de bens e, conseqüentemente, maior renda (riqueza). Isto porque os alunos deste grupo possuem duas geladeiras e dois banheiros ou mais, maiores percentuais de posse de lava-roupa, carro, computador com acesso à internet. Também é o grupo que apresenta os maiores níveis em proficiência.

De forma análoga é evidente que o grupo 4 é o com menores níveis de posse de bens e renda. 100% dos alunos deste grupo possuem apenas uma TV. Nenhum possui lava-roupa. Os níveis de ausência de posse de geladeira, carro, computador e banheiro são acima da média da amostra. Este também é o grupo com o menor nível de proficiência em português.

A diferenciação entre os grupos 1, 3 e 5 ocorre pela posse de TVs (100% dos alunos do grupo 5 possuem apenas uma TV, enquanto ocorre o oposto nos grupos 1 e 3); na posse de geladeira (o grupo 1 tem maior percentual de alunos sem este bem em sua residência); lava-roupa (100% dos alunos do grupo 3 possuem lava-roupa ao contrário dos demais); e banheiro (o grupo 1 possuem a pior situação, seguido pelo 3, sendo o 5 com menor percentual de alunos sem banheiro). Considerando estes fatos a Tabela 8 apresenta a proposta de um conversor de faixa de renda.

**Tabela 8 – Proposta de conversão de faixa de renda para a amostra**

Cluster	Faixa de renda
1	F2
2	F5
3	F4
4	F1
5	F3

Fonte: elaboração própria

O passo seguinte foi testar algoritmos simples e que permitissem a replicação para diversos anos. O algoritmo ad hoc inicial foi pontuar a quantidade de posse de bens. Com base nesta pontuação inicial comparou-se se sua classificação mimetizaria a do procedimento estatístico. A proposta que se mostrou mais viável está relatada na Tabela 9. Trata-se de estabelecer pontos a depender da posse de bens, obter a soma destes pontos e classificar os alunos pelo número total de pontos que possui.

**Tabela 9 – Pontuação por tipo de posse de bem**

Pontos	TV	GEL	LRO	CAR	COM	BAN
0	Não possuir	Não possuir	Não possuir	Não possuir	Não possuir	Não possuir
1	Uma	Um			Possuir, mas sem acesso à internet	Um
2	Duas ou mais	Duas ou mais		Possuir um ou mais	Possuir com acesso	Dois ou mais
3			Possuir uma ou mais			

Fonte: Elaboração própria.

A partir da amostra realizada para calibrar o algoritmo foi obtida estatística descritiva dos pontos por faixa de renda. Isto está relatado na Tabela 10. Além da média de pontos por classe foi calculado o intervalo de confiança. Com base nisto estabeleceram-se pontos de corte. Assim, um aluno cuja soma de pontos seja inferior ou igual a 4,5 será classificado na faixa de renda 1, e assim por diante. As últimas duas colunas referem-se à comparação entre a análise de cluster e a classificação da faixa de renda utilizando-se o ponto de corte. A taxa de acerto global é de 63,7%, onde se considerou acerto a coincidência entre a classificação pelo procedimento de cluster e a classificação pelo procedimento de pontos por categoria. A maior dificuldade de classificação está nos alunos da faixa 4 (cluster 3) que possuem características muito semelhantes às da faixa 3 (cluster 5), diferindo mais no número de televisões.

**Tabela 10 – Pontos por cluster para amostra em análise**

<b>FAIXA RENDA</b>	<b>CLUSTER</b>	<b>MÉDIA</b>	<b>IC – MIN</b>	<b>IC – MAX</b>	<b>CORTE</b>	<b>ERRO</b>	<b>ACERTO</b>
F1	4	4,10	3,8	4,4	4,5	22,2%	77,8%
F2	1	5,23	4,9	5,5	6	34,5%	65,5%
F3	5	7,60	7,2	8,0	8	25,2%	74,8%
F4	3	8,64	8,3	9,0	9	76,8%	23,2%
F5	2	9,79	9,4	10,2		37,1%	62,9%

Fonte: elaboração própria.

A sugestão é que seja utilizado o sistema de pontuação por ser simples e possível de ser replicável tanto em diversos anos da Prova Brasil, quanto em outros instrumentos de ensino público que possuam estes mesmos indicadores em seu instrumento. Resta comentar que não é viável utilizar a análise de cluster para cada ano (ou grupo de alunos) em separado. Isto porque para cada ano haverá um conjunto de pesos distintos, dada a metodologia da análise de clusters. Com pesos distintos os recortes de renda podem diferir ao longo dos anos ou do grupo de alunos. Já se utilizando o procedimento de pontuação a discriminação de renda será feita, todos os anos com o mesmo critério. Assim, o procedimento parte de uma análise de cluster específica e replica esta análise para demais bases com os mesmos parâmetros.

Esta proposta de metodologia de geração de faixas de renda a partir da posse de bens é válida para o período de análise desta nota técnica. Isto porque ao longo do tempo os hábitos de consumo se alteram, a renda média da população se altera, podendo interferir na estrutura da posse de bens. Assim, em trabalhos que envolvam outras amplitudes temporais sugere-se repetir a metodologia para validar se este sistema de pontos ainda é válido ou se é necessário ajustes nos pesos ou nos pontos de corte.