

## Efeitos das unidades federativas na renda disponível per capita por domicílio: uma análise multinível

Effects of federative units on per capita household disposable income: A multilevel analysis

Eduardo Ben-Hur de Queiroz Gomes<sup>1\*</sup>; Wilson Tarantin Junior<sup>2</sup>

Recebido: jun. 04, 2024

Aceito: jan. 13, 2025

<sup>1</sup>Especialista em Data Science e Analytics. Rua Conselheiro Ramalho, 636, Bela Vista, 01325-000, São Paulo, São Paulo, Brasil

<sup>2</sup>Doutor em Controladoria e Contabilidade. Rua Cezira Giovanoni Moretti, 580, Santa Rosa, 13414-157, Piracicaba, São Paulo, Brasil

\*Autor correspondente: [eduardo.benhur@gmail.com](mailto:eduardo.benhur@gmail.com)

**Resumo:** Este trabalho investigou os efeitos das unidades federativas brasileiras na renda per capita por domicílio, utilizando microdados da Pesquisa de Orçamentos Familiares (POF) 2017-2018, realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE). O objetivo foi analisar como a média de anos de estudo e a condição de segurança alimentar dos domicílios influenciam a renda disponível familiar per capita, considerando a forma como as unidades federativas podem impactar essa renda. Foram utilizados os microdados da POF aplicando regressão multinível, adequada para capturar os efeitos hierárquicos utilizando as variáveis explicativas: média de anos de estudos e condição de segurança alimentar. A escolha dessas variáveis foi sustentada pela literatura que indicou a educação como fator crucial na melhoria da renda familiar e na promoção do desenvolvimento econômico, e a segurança alimentar como indicador direto das condições de vida e saúde, influenciando a qualidade de vida e a capacidade de gerar renda. Os resultados mostraram diferenças significativas na renda entre domicílios de diferentes unidades federativas. A média de anos de estudo dos moradores e a condição de segurança alimentar impactaram significativamente a renda familiar per capita. Concluiu-se que as políticas públicas devem considerar essas variações regionais para promover distribuição de renda mais equitativa.

**Palavras-chave:** distribuição de renda; insegurança alimentar; "machine learning"; pesquisa de orçamentos familiares.



Este é um artigo publicado em acesso aberto (Open Access) sob a licença Creative Commons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

**Abstract:** This study investigated the effects of Brazilian federal units on per capita household income, using microdata from the 2017–2018 Pesquisa de Orçamentos Familiares (POF) conducted by the Instituto Brasileiro de Geografia e Estatística (IBGE). The objective was to analyze how the average years of schooling and the household food security status influence per capita disposable family income, considering the impact of federal units on this income. Microdata from the POF were analyzed using multilevel regression, an appropriate method for capturing hierarchical effects, using average years of schooling and household food security status as explanatory variables. The selection of these variables was supported by the literature, which highlights education as a crucial factor in improving family income and fostering economic development, while food security serves as a direct indicator of living and health conditions, influencing quality of life and income generation capacity. The results revealed significant differences in income among households across different federal units. Both the average years of schooling and household food security status had a significant impact on per capita family income. It was concluded that public policies should account for these regional variations to promote a more equitable income distribution.

**Keywords:** income distribution; food insecurity; machine learning; household budget survey.

## 1. Introdução

A desigualdade social é um desafio persistente no Brasil. Entender os fatores que influenciam a renda das famílias é crucial para a formulação de políticas públicas para combate da desigualdade<sup>[1]</sup>. Para isso são necessários estudos que analisem como variações contextuais das unidades federativas brasileiras impactam a renda dos domicílios e influenciam, de formas distintas, o seu comportamento em razão das condições de segurança alimentar e educação.

A Pesquisa de Orçamentos Familiares (POF) é amplamente reconhecida como uma importante fonte de dados sobre os orçamentos domésticos e as condições de vida da população brasileira. A POF utiliza uma amostra representativa da população brasileira e coleta dados ao longo de 12 meses, com o objetivo de entender o padrão de consumo das famílias e atualizar os índices de preços do consumidor, além de apresentar indicadores sobre qualidade de vida da população<sup>[2]</sup>.

O Instituto Brasileiro de Geografia e Estatística (IBGE) disponibiliza os microdados de pesquisa em seu portal, de modo a facilitar a exploração de sua base de dados sob um prisma diverso de possibilidades. Por exemplo, para entender as tendências de desigualdade no Brasil, Hoffmann e Vaz<sup>[3]</sup> analisaram a distribuição da renda e da despesa familiar per capita utilizando dados da POF. Eles descobriram que, entre 2008 e 2018, houve um aumento na renda familiar per capita média e uma redução na desigualdade, no entanto, essa redução foi desacelerada pelo aumento das disparidades inter-regionais, o que possibilitou uma avaliação em contextos.

A expressividade dos dados da POF permitiu o desenvolvimento de diversos trabalhos nos mais variados campos como, por exemplo, a busca por padrões de consumo alimentar baseados no orçamento das famílias na região metropolitana de São Paulo. Tal pesquisa foi realizada com o objetivo de avaliar como os níveis de renda das famílias podem estar relacionados aos padrões de consumo, seu comportamento entre regiões e ao longo do tempo<sup>[4]</sup>.

No Brasil, há uma relação significativa entre renda e níveis de escolaridade, com impactos em diversos desfechos de saúde. Pesquisas indicam que indivíduos com menor escolaridade e renda tendem a relatar maior perda dentária<sup>[5]</sup>, enquanto a expansão dos serviços voltados à saúde primária, como a Estratégia de Saúde da Família (ESF), demonstrou reduzir a mortalidade por todas as causas, beneficiando particularmente grupos em situação de vulnerabilidade socioeconômica com menor nível educacional<sup>[6]</sup>.

Russo e Dias<sup>[7]</sup> destacaram a importância da educação na melhoria do estado de saúde da população, observando retornos financeiros crescentes para indivíduos com melhor condição de saúde. As disparidades na prevalência de doenças crônicas também foram associadas a gradientes educacionais, reforçando o impacto da educação na saúde no Brasil<sup>[8]</sup>. Considerando que existem múltiplas possibilidades de análises com base na POF, e o fato de sua estrutura possibilitar abordagem multinível — isto é, indivíduos aninhados em domicílios, domicílios aninhados em estados e, estados aninhados em regiões —, o presente estudo analisou a renda familiar per capita sob essa perspectiva com o objetivo de verificar se a média de anos de estudo dos moradores do domicílio e sua condição de segurança alimentar exercem influência sobre a renda disponível familiar per capita. Considerou-se, ainda, a forma como as unidades federativas podem influenciar tal renda.

## 2. Material e Métodos

Este trabalho utilizou os microdados da Pesquisa de Orçamentos Familiares (POF) 2017-2018, lidos por meio de um script em linguagem R, fornecido pelo Instituto Brasileiro de Geografia e Estatística (IBGE). O script tem como principal função ler os arquivos em formato de texto e armazenar, no disco rígido, um arquivo serializado, compacto e de menor tempo de leitura, para consumo posterior.

Os dados selecionados para pesquisa foram de natureza quantitativa, obtidos a partir dos microdados dos questionários na POF. Esses dados foram submetidos a procedimentos de limpeza e pré-processamento. Em seguida aplicou-se a modelagem multinível para análise hierárquica, considerando domicílios aninhados em unidades federativas, com base em uma estratégia “step-up” de inclusão progressiva de variáveis. A transformação de Box-Cox foi aplicada à variável dependente com objetivo de estabilizar a variância e aproximar sua distribuição à normalidade. A significância estatística dos modelos foi avaliada por meio de testes de máxima verossimilhança, utilizados na interpretação dos resultados.

Segundo o IBGE<sup>[2]</sup>, a renda familiar per capita (RFPC) disponível foi obtida pela divisão da renda disponível total do domicílio pelo número de moradores. A renda disponível, por sua vez, corresponde à somatória dos rendimentos monetários e não monetários, subtraídos os impostos diretos e contribuições sociais, além de outras deduções.

### Carga da base de moradores e domicílios

Após a leitura inicial dos microdados e sua conversão para o formato “.rds” — o que facilitou a leitura e análise dos dados utilizando linguagem R por meio do software Rstudio® — foram carregadas as duas bases principais da POF. A base de moradores continha as características individuais dos componentes da família ou unidade de consumo, como idade, cor ou raça, sexo biológico, nível de escolaridade, entre outras variáveis, totalizando 57 variáveis e 178.431 observações. A base de domicílio, por sua vez, incluía as características dos domicílios dos entrevistados como, titularidade do imóvel, nível de segurança alimentar e materiais predominantes na construção, entre outras informações, totalizando 38 variáveis e 57.920 observações.

Para esse estudo, os dados foram tratados com a remoção de valores faltantes, discrepantes, e rendas negativas. Também foram aplicados filtros para excluir observações nas quais o indivíduo era empregado doméstico ou parente de empregado doméstico — seguindo a mesma premissa adotada pelo IBGE para cálculo de renda disponível familiar per capita, em que tais indivíduos não são considerados nos cálculos. Além disso, foram selecionados apenas os domicílios localizados em áreas urbanas.

### Disposição do conjunto de dados

O conjunto de dados utilizado para construção das análises — após as etapas de transformações e limpeza — foi composto por 44.712 domicílios e 2 variáveis explicativas, extraídas das bases dos microdados dos indivíduos e dos domicílios. Para os resultados preliminares, foram selecionadas as seguintes variáveis: (i) a média de anos de estudo dos moradores por domicílio, derivada da variável original anos de estudo de cada indivíduo; (ii) a condição de segurança alimentar do domicílio. A escolha dessas variáveis justificou-se por sua reconhecida influência sobre a renda familiar e por estarem diretamente relacionadas ao objetivo central da pesquisa. Essa delimitação contribuiu para um maior aprofundamento na análise da interação entre fatores socioeconômicos e as unidades federativas, conferindo objetividade ao estudo. Ambas as variáveis foram utilizadas como explicativas no modelo multinível. A variável de segurança alimentar foi obtida a partir da base de domicílios e foi incluída com o intuito de avaliar o desempenho da renda familiar per capita (RFPC) em função do nível segurança alimentar do domicílio.

As variáveis de peso amostral, que funcionam como fatores de expansão para estimativas populacionais, não foram utilizadas nos resultados deste estudo, uma vez que um de seus objetivos secundários foi a familiarização com as técnicas de modelagem multinível e sua aplicação em dados da POF. A escolha de um conjunto limitado de variáveis permitiu um aprendizado mais aprofundado da técnica e forneceu uma base sólida para futuras pesquisas com maior complexidade. Assim, as análises foram realizadas com base nas observações amostrais, sem ajustes por pesos.

A modelagem multinível, sobretudo com múltiplos níveis e variáveis, impõe desafios computacionais consideráveis<sup>[9]</sup>. A seleção de um número reduzido de variáveis possibilitou explorar as nuances metodológicas da técnica, e garantindo a robustez da análise, sem comprometer a capacidade de processamento e o tempo de execução do modelo.

Este estudo seguiu metodologias estatísticas validadas, alinhando-se a pesquisas anteriores que também utilizaram dados da POF. Um exemplo é o trabalho de aplicação de modelos de regressão de Poisson com variância robusta para investigar a insegurança alimentar, como realizado por Araújo et al.<sup>[10]</sup>. Esse tipo de abordagem reforça a relevância de análises estatísticas avançadas para compreensão de questões socioeconômicas, na formulação de políticas públicas direcionadas.

Souza et al.<sup>[11]</sup> evidenciaram o aumento da insegurança alimentar no Brasil entre 2013 e 2018, destacando desigualdades regionais significativas. Tal constatação embasou a escolha da condição de segurança alimentar como uma das variáveis explicativas deste estudo. De forma complementar, Palmeira et al.<sup>[12]</sup> analisaram os efeitos de programas governamentais na mitigação da insegurança alimentar no Nordeste do Brasil, identificando uma associação positiva entre políticas de proteção social e segurança alimentar. Esses achados reforçam a importância de abordar a interação entre variáveis contextuais e regionais, orientando tanto a escolha metodológica quanto a justificativa para o foco na variável de segurança alimentar e na média de anos de estudo dos moradores como centrais para a análise. Dessa forma, a seleção das variáveis explicativas foi guiada por sua relevância teórica e viabilidade prática de implementação, coadunando-se com o objetivo de explorar as dinâmicas socioeconômicas sob a ótica das desigualdades regionais.

## Descrição das variáveis

As variáveis selecionadas estavam alinhadas ao foco do estudo— educação e segurança alimentar — e esperava-se que outros fatores macroestruturais fossem capturados pela modelagem hierárquica das unidades federativas. A média de anos de estudo e a condição de segurança alimentar foram consideradas indicadores cruciais para análise da relação entre educação e bem-estar econômico. A literatura demonstrou a associação entre insegurança e níveis educacionais, conforme discutido por Hoffmann e Vaz<sup>[3]</sup>, Souza et al.<sup>[11]</sup> e Palmeira et al.<sup>[12]</sup>, evidenciando sua importância na compreensão das dinâmicas socioeconômicas e na formulação de políticas públicas eficazes.

A Tabela 1 apresenta as variáveis utilizadas no conjunto de dados tratado, “dataset”, incluindo a renda familiar per capita disponível, o nível de instrução da pessoa de referência no domicílio, e a média de anos de estudo dos moradores. Essas variáveis foram selecionadas por sua relevância na compreensão das disparidades de renda entre diferentes contextos domiciliares no Brasil.

**Tabela 1.** Variáveis utilizadas no “dataset”

Variável	Descrição
ID_domicílio	Variável resultante da concatenação das variáveis originais COD_UPA e NUM_DOM, responsáveis por caracterizar um único domicílio e o contexto de nível 1 no “dataset”
PC_RENDA_DISP	Renda disponível familiar per capita de cada domicílio
Instrução	Nível de instrução da respectiva pessoa de referência da unidade de consumo
media_anos_estudo_moradores	Média de anos de estudo dos moradores do domicílio calculada a partir da variável original ANOS_ESTUDO da base morador
Segurança_alimentar	Variável V61699 original da base de domicílio do tipo categórica, que representa a situação de segurança alimentar do domicílio

Fonte: Elaborado pelo autor com base nos dados da Pesquisa de Orçamentos Familiares 2017-2018 (IBGE)<sup>[2]</sup>

Evidenciou-se a concentração de domicílios por diferentes níveis de instrução da pessoa de referência (Tabela 2). A tabela demonstra que a maior quantidade de domicílios era composta por pessoas com Ensino Fundamental incompleto, seguido pelo Ensino Médio completo. Além disso, os dados revelam como a renda média variou significativamente de acordo com o nível de instrução. A pessoa de referência, por sua vez, foi aquela que era a responsável pelo pagamento das despesas dos domicílios de acordo com os seguintes contextos: pagamento do aluguel, em caso de domicílio alugado; pagamento das prestações do imóvel, no caso de domicílios não quitados; ou demais despesas da habitação.

**Tabela 2.** Quantidade de domicílios por nível de instrução da pessoa de referência

Nível de instrução	Quantidade de domicílios	Renda média familiar per capita (R\$)
Ensino Fundamental incompleto	15.424	1.244,41
Ensino Médio completo	11.617	1.684,48
Ensino Superior completo	6.422	4.011,30
Ensino Fundamental completo	3.807	1.450,39
Sem instrução	3.403	1.124,48
Ensino Médio incompleto	2.370	1.264,51
Ensino Superior incompleto	1.674	2.293,26
Total	44.717	

Fonte: Elaborado pelo autor com base em dados da Pesquisa de Orçamentos Familiares 2017-2018 (IBGE)<sup>[2]</sup>

A Tabela 3, apresenta a distribuição dos domicílios conforme os níveis de segurança alimentar, classificados em quatro categorias: segurança alimentar, insegurança leve, insegurança moderada e insegurança grave. Observou-se que a maioria dos domicílios apresentava segurança alimentar (27.129 domicílios), embora uma proporção significativa enfrentasse algum grau de insegurança. A renda média variou substancialmente entre os diferentes níveis de segurança alimentar, com os domicílios em situação de insegurança grave apresentando a menor renda média (R\$ 909,08), reforçando a correlação entre segurança alimentar e condições econômicas.

**Tabela 3.** Quantidade de domicílios por nível de segurança alimentar do domicílio

Nível de segurança alimentar	Quantidade de domicílios	Renda média familiar per capita (R\$)
Segurança	27.129	2.268,72
Insegurança leve	11.232	1.163,34
Insegurança moderada	4.012	987,41
Insegurança grave	2.344	909,08
Total	44.717	

Fonte: Elaborado pelo autor com base em dados da Pesquisa de Orçamentos Familiares 2017-2018 (IBGE)<sup>[2]</sup>

Em modelos multiníveis, os contextos dizem respeito às estruturas de aninhamento dos dados. Como os domicílios estavam contidos em uma única UF, definiu-se um constructo para avaliação dos efeitos aleatórios sobre a variável RFPC, considerando domicílios aninhados em unidades federativas — sendo domicílios como contextos de nível 1, i.e., unidade de análise mais granular e UF como contextos de nível 2, i.e., contexto de nível superior.

Para complementar a análise, realizou-se um estudo sobre o desbalanceamento dos dados no conjunto final. A Tabela 4 apresenta as quantidades de domicílios agrupados por UF, acompanhado da respectiva renda média familiar per capita em cada estado. Observou-se uma variação considerável entre os estados no que se refere à quantidade de domicílios e à renda média familiar per capita. Por exemplo, São Paulo apresentou o maior número de domicílios (3.403) e uma das maiores rendas per capita (R\$ 2.601,79), enquanto o Maranhão, com uma quantidade significativamente menor de domicílios (1.536), registrou a menor renda média per capita (R\$ 1.033,69). Esses dados foram fundamentais para o estudo, pois evidenciaram o desbalanceamento entre as UF, o que deve ser considerado nas análises multiníveis a fim de ajustar possíveis vieses regionais.

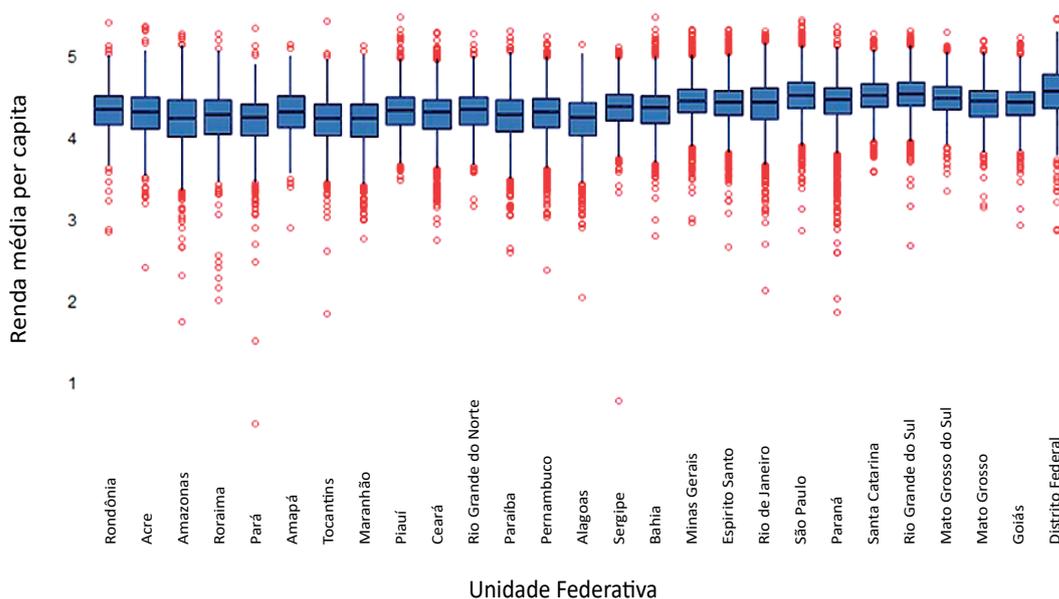
**Tabela 4.** Estudo sobre o desbalanceamento dos dados por Unidade Federativa

Unidade Federativa	Quantidade de domicílios	Renda média familiar per capita (R\$)
São Paulo	3.403	2.601,79
Minas Gerais	3.349	2.001,72
Rio de Janeiro	2.855	2.073,17
Espírito Santo	2.746	1.863,33
Pernambuco	2.664	1.378,33
Bahia	2.267	1.619,60
Ceará	2.193	1.300,17
Rio Grande do Sul	2.144	2.534,22
Paraná	2.123	2.032,13
Santa Catarina	1.932	2.271,23
Maranhão	1.536	1.033,69
Alagoas	1.526	1.092,51
Amazonas	1.488	1.302,82
Paraíba	1.470	1.299,54
Piauí	1.395	1.508,91
Mato Grosso do Sul	1.348	2.063,72
Rio Grande do Norte	1.250	1.475,26
Goiás	1.222	1.754,23
Sergipe	1.221	1.587,26
Distrito Federal	1.189	3.482,15
Mato Grosso	1.099	1.803,95
Pará	831	1.088,95
Acre	810	1.573,37

Unidade Federativa	Quantidade de domicílios	Renda média familiar per capita (R\$)
Tocantins	708	1.181,57
Roraima	682	1.328,67
Rondônia	650	1.483,42
Amapá	616	1.391,58
Total	44.717	

Fonte: Elaborado pelo autor com base em dados da Pesquisa de Orçamentos Familiares 2017-2018 (IBGE)<sup>[2]</sup>

A visualização da distribuição da variável RFPC por domicílio (Figura 1) evidenciou a possível necessidade de transformação da variável dependente, mais especificamente, a transformação de Box-Cox, com objetivo de melhorar o ajuste dos resíduos do modelo. Essa técnica estatística visa estabilizar a variância e aproximar a distribuição de uma variável à normalidade por meio de um parâmetro de transformação  $\lambda$ <sup>[13]</sup>. Embora o estudo não tenha como finalidade a predição da renda disponível per capita, a premissa de maximização do logaritmo da função de verossimilhança foi considerada relevante para a modelagem.



**Figura 1.** Renda média per capita disponível

Fonte: Dados originais da pesquisa

### Modelagem de dados e estratégia “step-up”

Na abordagem multinível, é possível definir quantos níveis forem necessários, desde que seja possível contextualizar a substancialidade das hierarquias. A estruturação dos contextos para estimação dos modelos ocorre com base nas características do conjunto de dados e nos efeitos aleatórios que se pretende capturar. Neste estudo, foram estimados modelos de regressão com e sem a inclusão de contextos aleatórios. Isso significa que foram analisadas as variações nos resultados, que se mostraram estatisticamente significativas quando os diferentes contextos foram levados em consideração.

Segundo Courgeau<sup>[14]</sup>, o uso de equações em níveis permite ao pesquisador transitar entre diferentes ciências sociais, evitando análises equivocadas que ignoram a estrutura hierárquica dos dados — como as relações entre regiões, famílias e indivíduos. A principal vantagem da modelagem multinível em relação aos modelos tradicionais é a possibilidade de incorporar o aninhamento natural dos dados<sup>[15]</sup>. Diferentemente dos modelos de regressão clássicos, os modelos multiníveis contemplam tanto as interações entre variáveis no componente de efeitos fixos, as interações entre os termos de erro e variáveis no componente de efeitos aleatórios<sup>[16]</sup>.

Com essa abordagem, foi estimado um modelo multinível para verificar a existência de diferenças estatisticamente significantes nas variâncias de intercepto e de inclinação, comparando com modelos tradicionais de regressão por meio do método “Restricted Estimation of Maximum Likelihood Restrict” (REML).

A estratégia adotada para modelagem foi chamada “step-up strategy” que consiste na inclusão progressiva de parâmetros a cada etapa, permitindo o refinamento do modelo com base nas significâncias estatísticas dos efeitos aleatórios, conforme o avanço na inserção de parâmetros<sup>[17]</sup>. Essa estratégia amplamente reconhecida, na literatura, proporciona maior controle sobre a inclusão de variáveis e melhor entendimento do comportamento dos efeitos aleatórios. Considerando que a modelagem multinível não se adequa ao uso da tradicional estratégia “stepwise”, foi necessário recorrer a algoritmos específicos para determinação dos erros padrão das variâncias dos componentes aleatórios, o que se mostrou para os modelos do tipo “Hierarchy Linear Modeling Level-2”, “Hierarchy Linear Modeling Level-2” (HLM2) e “Hierarchy Linear Modeling Level-3” (HLM3), desde que fossem estimados pelo pacote “nlme” do software R<sup>®1</sup>.

A aplicação da estratégia “step-up” envolveu a comparação entre quatro modelos. O primeiro modelo consistiu no modelo nulo estimado por Ordinary Least Squares (OLS). O segundo foi o modelo nulo do tipo HLM2. O terceiro considerou o modelo HLM2 com efeito aleatório de intercepto. Por fim, o quarto modelo envolveu o HLM2 com efeito aleatório de intercepto e de inclinação.

A comparação entre modelos foi realizada por meio do logaritmo da função de máxima verossimilhança, permitindo avaliar o ganho de explicação ao se considerar a estrutura hierárquica dos dados. Considerando a estratégia “step-up”, o modelo de regressão foi inicialmente estimado a partir do modelo nulo HLM2 cuja forma funcional é apresentada na Equação (1):

$$RFPC_{ij} = \beta_0_j + \varepsilon_{ij} \quad (1)$$

onde,  $RFPC_{ij}$ : renda familiar per capita disponível por domicílio  $i$ , pertencente ao grupo  $j$  sendo o fenômeno em estudo;  $\beta_0_j$ : intercepto do modelo; e,  $\varepsilon_{ij}$ : termos de erro.

A partir do modelo nulo, foi possível avaliar a existência e significância estatística dos efeitos aleatórios de intercepto no nível 2, onde  $\beta_0_j$ , para o modelo multinível, foi dado pela Equação (2):

$$\beta_0_j = \gamma_00 + u_0j \quad (2)$$

onde,  $\beta_0_j$ : intercepto médio para a variável dependente no nível 2, i.e., o valor médio esperado da variável dependente em todas as unidades quando as variáveis independentes no modelo são iguais a zero;  $\gamma_00$ : valor esperado da variável dependente para determinada observação  $i$  quando as demais variáveis explicativas forem zero; e,  $u_0j$ : termos de erro que indicam a existência de aleatoriedade nos interceptos.

Ao realizar os testes estatísticos, se  $u_0j$  for estatisticamente diferente de zero, foi possível afirmar que existem características das unidades federativas que impactam no desempenho médio da renda familiar per capita disponível.

Em seguida, foi avaliado um modelo com efeitos aleatórios de intercepto, onde a renda disponível foi estimada pela Equação (3):

$$RFPC_{ij} = \beta_0_j + \beta_1_j * X_{1ij} + \beta_2_j * X_{2ij} + \varepsilon_{ij} \quad (3)$$

onde,  $RFPC_{ij}$ : renda familiar per capita disponível por domicílio, sendo o fenômeno em estudo;  $\beta_0_j$ : intercepto do modelo;  $\beta_1_j$ : estimador do modelo que, nesse caso, multiplicará a variável média de anos de estudos dos indivíduos do domicílio;  $X_{1ij}$ : variável média de anos de estudo dos moradores do domicílio;  $\beta_2_j$ : estimador do modelo que, neste caso, multiplicará a condição de segurança alimentar do domicílio;  $X_{2ij}$ : variável referente ao nível de segurança alimentar do domicílio; e,  $\varepsilon_{ij}$ : termos de erro.

Destaca-se que a variável categórica referente ao nível de segurança alimentar do domicílio foi operacionalizada por meio da criação de variáveis “dummies”. Uma variável “dummy” é aquela que assume o valor de 1 ou 0 para indicar a ausência ou presença de um determinado atributo e, em casos de variáveis com mais de duas categorias, é necessário que seja operacionalizado o procedimento de  $n-1$  “dummies”, criando  $n-1$  variáveis para cada categoria preexistente.

A partir da Equação (3), tem-se para o nível 2, considerando efeitos aleatórios de intercepto, as Equações (4), (5) e (6):

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (4)$$

$$\beta_{1j} = \gamma_{10} \quad (5)$$

$$\beta_{2j} = \gamma_{20} \quad (6)$$

onde,  $\beta_{0j}$ : intercepto aleatório no nível 2 (unidade federativa j);  $\gamma_{00}$ : valor esperado da variável dependente para determinada observação quando as demais variáveis explicativas forem zero;  $u_{0j}$ : termos de erro que indicam a existência de aleatoriedade nos interceptos;  $\beta_{1j} = \gamma_{10}$ : valor esperado da variável dependente para uma determinada observação quando existir alteração unitária na variável média de anos de estudo dos moradores do domicílio (aumento unitário em anos de estudo eleva a RFPC em  $\gamma_{10}$ ); e  $\beta_{2j} = \gamma_{20}$ : valor esperado da variável dependente para uma determinada observação i quando existir alteração na variável nível de segurança alimentar do domicílio, que é uma variável qualitativa, ou seja, considerou-se um beta e um efeito aleatório para n-1 categoria após a transformação para variável "dummy".

Com isso, evoluiu-se do modelo nulo para o modelo com interceptos aleatórios e variáveis explicativas de nível 1, avaliando a significância estatística dos  $\beta_x$  de intercepto, chegando na Equação (7):

$$RFPC_{ij} = \gamma_{00} + \gamma_{10} * X_{ij} + \gamma_{20} * X_{ij} + u_{0j} + \varepsilon_{ij} \quad (7)$$

onde,  $RFPC_{ij}$ : é a renda familiar per capita do domicílio i no grupo j;  $\gamma_{00}$ ,  $\gamma_{10}$  e  $\gamma_{20}$ : são os coeficientes fixos (intercepto e efeitos das variáveis de nível 1);  $u_{0j}$ : efeito aleatório do grupo j no intercepto;  $X_{ij}$ : variáveis analisadas;  $\varepsilon_{ij}$ : erro aleatório no nível 1 (variação não explicada no domicílio i).

Em seguida, com a variância estatisticamente significativa nos efeitos aleatórios de intercepto, chegou-se ao modelo com efeitos aleatórios de intercepto e de inclinação, dado pela Equação (8):

$$RFPC_{ij} = \beta_{0j} + \beta_{1j} * X_{1ij} + \beta_{2j} * X_{2ij} + \varepsilon_{ij} \quad (8)$$

onde,  $RFPC_{ij}$ : renda familiar per capita do domicílio i no grupo j;  $\beta_{0j}$ ,  $\beta_{1j}$  e  $\beta_{2j}$ : coeficientes aleatórios (intercepto e inclinações variando por grupo j);  $X_{1ij}$ ,  $X_{2ij}$ : variáveis analisadas;  $\varepsilon_{ij}$ : erro aleatório no nível 1.

A partir da Equação (8), tem-se para o nível 2, considerando efeitos aleatórios de intercepto e de inclinação, as Equações (9), (10) e (11):

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (9)$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \quad (10)$$

$$\beta_{2j} = \gamma_{20} + u_{2j} \quad (11)$$

onde,  $\beta_{0j}$ : intercepto aleatório no nível 2 (unidade federativa j);  $\gamma_{00}$ : valor esperado da variável dependente para determinada observação quando as demais variáveis explicativas forem zero;  $u_{0j}$ : termos de erro que indicam a existência de aleatoriedade nos interceptos;  $\beta_{1j} = \gamma_{10}$ : valor esperado da variável dependente para uma determinada observação quando existir alteração unitária na variável média de anos de estudo dos moradores do domicílio (aumento unitário em anos de estudo eleva a RFPC em  $\gamma_{10}$ );  $u_{1j}$ : efeitos aleatórios de inclinação em relação a variável média de anos de estudo;  $\beta_{2j} = \gamma_{20}$ : valor esperado da variável dependente para uma determinada observação quando existir alteração nas categorias da variável referente as condições de segurança alimentar;  $u_{2j}$ : são os efeitos aleatórios de inclinação em relação a variável condição de segurança alimentar.

Com isso evoluiu-se do modelo com efeitos aleatórios de intercepto, para o modelo com efeitos aleatórios de intercepto e de inclinação, chegando na Equação (12):

$$RFPC_{ij} = \gamma_{00} + \gamma_{10} * X_{1ij} + \gamma_{20} * X_{2ij} + \gamma_{30} * X_{3ij} + \gamma_{40} * X_{4ij} + u_{0j} + u_{1j} * X_{1ij} + u_{2j} * X_{2ij} + \varepsilon_{ij} * X_{3ij} + u_{4j} * X_{4ij} + \varepsilon_{ij} \quad (12)$$

onde,  $RFPC_{ij}$ : renda familiar per capita do domicílio  $i$  no grupo  $j$ ;  $\gamma_{00}$ : intercepto fixo global (renda esperada quando todas as variáveis explicativas são zero);  $\gamma_{10}$ : efeito fixo da média de anos de estudos sobre a RFPC;  $\gamma_{20}$ ,  $\gamma_{30}$  e  $\gamma_{40}$ : efeitos fixos das categorias de segurança alimentar (operacionalizadas como dummies n-1);  $u_{0j}$ : efeito aleatório no intercepto de  $X_{1ij}$  (variação não explicada entre grupos);  $u_{1j}$ : efeito aleatório na inclinação de  $X_{1ij}$  (variação no impacto da educação entre grupos);  $X_{1ij}$ ,  $X_{2ij}$ ,  $X_{3ij}$ ,  $X_{4ij}$ : variáveis analisadas;  $u_{2j}$ ,  $u_{3j}$  e  $u_{4j}$ : efeitos aleatórios nas inclinações das dummies (variação no efeito das categorias de segurança alimentar entre os grupos)  $\epsilon_{ij}$ : erro aleatório no nível 1 (domicílio).

Destaca-se que, embora o modelo apresente apenas como variáveis explicativas, a categoria de referência (segurança alimentar) foi excluída para evitar multicolinearidade, seguindo a regra de n-1 dummies.

Por fim, segundo Tabachnick e Fidell<sup>[18]</sup>, se as variâncias dos termos aleatórios  $u_{0j}$  e  $u_{1j}$  forem estatisticamente diferentes de zero, não será adequado estimar os parâmetros por modelos tradicionais OLS. Sendo assim, foram comparados os resultados com modelos OLS, a fim de verificar o modelo que maximizou o logaritmo da função de máxima verossimilhança utilizando o software R®.

### Alteração na forma funcional do modelo proposto: transformação de Box-Cox

Box e Cox<sup>[13]</sup> propuseram uma transformação funcional com o objetivo de melhorar a normalidade dos resíduos e estabilizar a variância dos modelos de regressão. Segundo Fávero e Belfiore<sup>[15]</sup>, partindo de um modelo de regressão com uma variável  $Y$  dependente e aplicando a transformação proposta por Box e Cox, uma nova variável é gerada com um novo histograma, e os parâmetros da transformação de Box-Cox tornam a distribuição dos resíduos do modelo construído *a posteriori* mais aderente à normalidade.

A transformação proposta por Box-Cox segue a Equação (13):

$$Y_{\text{Box-Cox}}^* = \frac{Y^\lambda - 1}{\lambda} \quad (13)$$

onde:  $Y_{\text{Box-Cox}}^*$ : variável dependente transformada;  $Y$ : parâmetro estimado para normalizar resíduos, varia de  $-\infty$  até  $+\infty$  e, portanto, qualquer valor pode ser construído a partir dele; onde, o parâmetro  $\lambda$ : varia de  $-\infty$  até  $+\infty$  e, portanto, qualquer valor pode ser construído a partir dele.

### Clusterização e “scree plot”

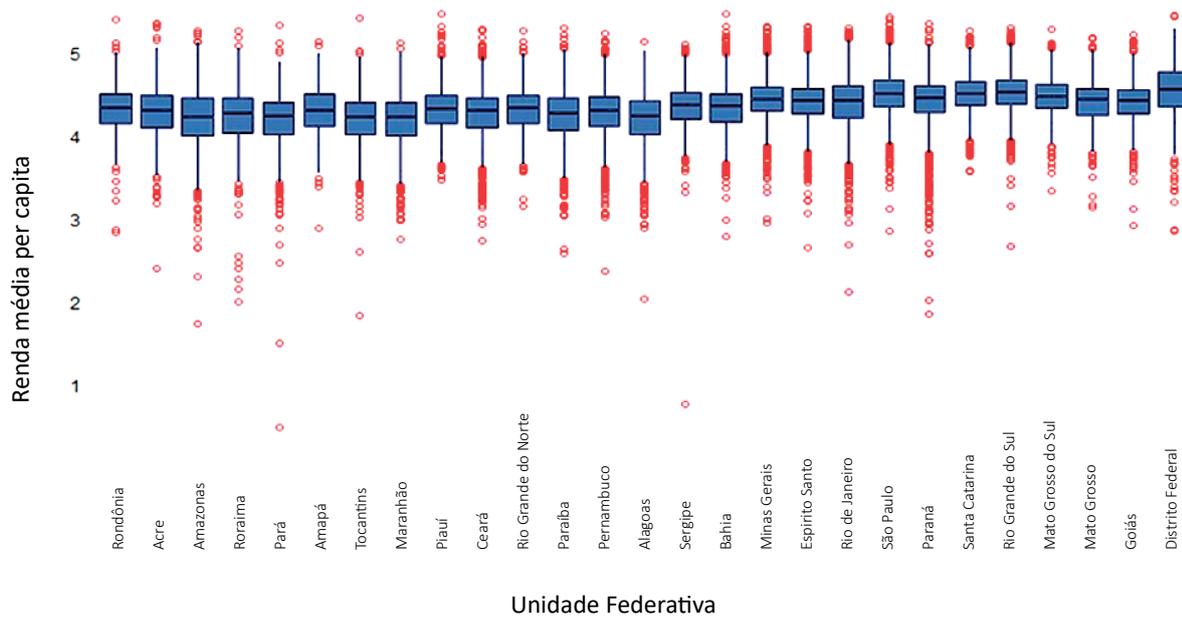
Para agrupar observações com características semelhantes, foram aplicadas técnicas de análise de agrupamento (clusterização). Utilizou-se a metodologia de clusterização “k-means” e o gráfico “scree plot” para identificar o número ideal de “clusters” — definidos em três para este estudo. A abordagem segue metodologia similar à utilizada por Damayanti et al.<sup>[19]</sup>, que empregaram o “scree plot” para otimização de “clusters” em sua análise de poluição ambiental na Indonésia.

## 3. Resultados e Discussão

Antes da estimação dos modelos HLM2, a variável RFPC foi submetida à transformação de Box-Cox com objetivo de avaliar sua distribuição e garantir maior robustez estatística nas análises hierárquicas.

Boas práticas metodológicas recomendam a aplicação do teste de Shapiro Francia nos resíduos do modelo para verificar a aderência dos resíduos à normalidade. Entretanto, considerando que a base de dados possuía 44.717 observações — número acima do limite de segurança para esse teste, no software R® (usualmente estabelecido em 5.000 observações) optou-se avaliar a normalidade dos resíduos via transformação de Box-Cox. Conforme destacado, a transformação de Box-Cox visa estabilizar a variância e aproximar a distribuição de uma variável dependente à normalidade, tornando a análise mais robusta e confiável<sup>[13]</sup>. Obtendo-se um  $\lambda$  igual ou muito próximo de 1, possivelmente os resíduos do futuro modelo seriam aderentes à normalidade. Fazendo uso da função “powerTransform” do pacote “car” na linguagem R, foi calculado o  $\lambda$  de Box-Cox e o valor obtido foi de  $\lambda = -0,1515865$ , isto é, numericamente diferente de 1. Assim, a modelagem foi conduzida com a variável RFPC transformada, gerando uma nova distribuição da variável e de seus resíduos.

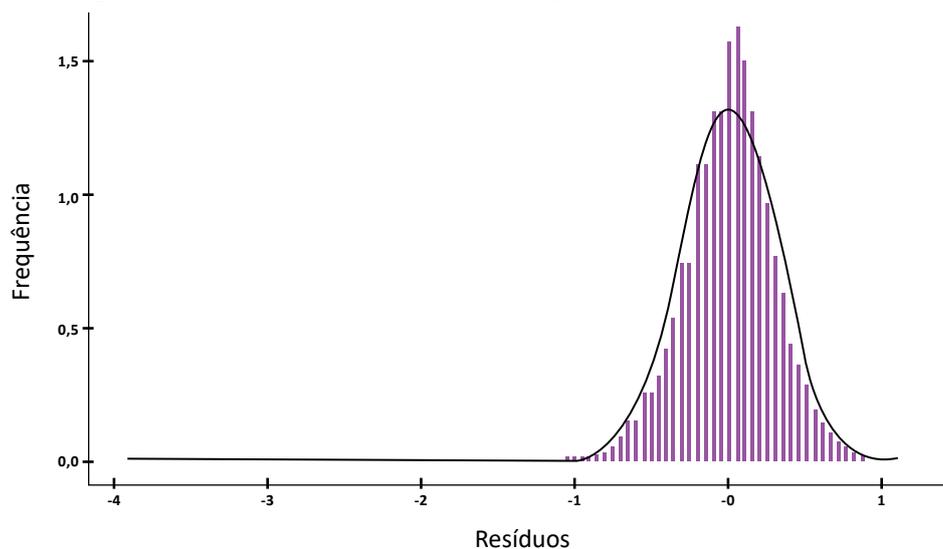
É possível avaliar a nova distribuição da variável transformada na Figura 2, quando comparada com a Figura 1.



**Figura 2.** Distribuição da nova variável dependente transformada por Box-Cox  
 Fonte: Resultados originais da pesquisa

### Modelo nulo OLS

Foram estimados dois modelos nulos: um modelo tradicional de regressão OLS e um Modelo HLM2. Para compará-los, utilizou-se a função *lrtest* do pacote *lmtest*, com base na maximização do logaritmo da função de verossimilhança (“LogLik”). O modelo com maior “LogLik” foi considerado mais adequado para avançar na estratégia “step-up”. Adicionalmente, foi realizada uma análise visual da distribuição dos resíduos da variável transformada. A Figura 3, ilustra a comparação entre a distribuição dos resíduos e uma curva normal teórica. Além disso, a transformação de Box-Cox foi aplicada à variável completa, sem distinção de contexto, como uma primeira abordagem metodológica. Esse procedimento resultou em um ajuste.



**Figura 3.** Distribuição dos resíduos do modelo nulo com transformação de Box-Cox e curva normal teórica para visualização  
 Fonte: Resultados originais da pesquisa

### Modelo nulo HLM2

Seguindo a estratégia de incremento de variáveis, e após entender que a variável transformada por Box-Cox aproxima a distribuição dos resíduos à normalidade, foi estimado um modelo nulo do tipo HLM2 multinível considerando apenas a variável transformada e o agrupamento por UF. Em comparação ao modelo OLS, observou-se um incremento no “LogLik”, indicando um ajuste superior. A Tabela 5, apresenta os resultados do modelo nulo HLM2 avaliando a variável dependente transformada por Box-Cox. O passo seguinte foi avaliar a significância estatística do componente aleatório de intercepto para seguir com a estratégia “step-up”, adicionando mais variáveis explicativas para o modelo com a intenção de maximizar o “LogLik” e capturar o comportamento do fenômeno com menor erro.

**Tabela 5.** Saída do modelo HLM2 Nulo transformado por Box-Cox

Nome	Modelo
	Modelo 1
(Intercepto)	4,35 ***
Erro padrão	(0,02)
Desvio padrão (Intercepto)	0,11
Número de observações	44.717,00
Log Likelihood ("LogLik")	-7.700,60
"Akaike Information Criterion" (AIC)	15.407,20
"Bayesian Information Criterion" (BIC)	15.433,33

Fonte: Resultados originais da pesquisa

Nota: Códigos de significância: \*\*\*p < 0,001; \*\*p < 0,01; \*p < 0,05

Com o auxílio do algoritmo de verificação das variações de intercepto e inclinação, foram estimadas as variâncias aleatórias de intercepto para o modelo HLM2, as quais se mostraram estatisticamente significantes. Esse foi o primeiro indicativo de que o agrupamento por UF capta efeitos estruturais relevantes na variação da RFPC. Na Tabela 6, são apresentados os resultados do algoritmo com a saída das variâncias dos termos aleatórios de intercepto e os p-valor de significância estatística.

**Tabela 6.** Output do algoritmo para determinação dos erros padrão das variâncias no componente de efeitos aleatórios

Componente de efeitos aleatórios	Estimativa de variância	Erro padrão	Z <sup>1</sup>	p-valor
(u <sub>0j</sub> ) <sup>2</sup>	0,01181730	0,003290947	3,590851	2,22e-16
(e) <sup>3</sup>	0,07865434	0,000526163	149,486644	2,22e-16

Fonte: Resultados originais da pesquisa

Nota: <sup>1</sup>Z: Estatística padronizada utilizada para avaliar a significância estatística das variâncias calculada com base na razão entre a estimativa da variância e seu erro padrão; <sup>2</sup>(u<sub>0j</sub>): Variância dos interceptos aleatórios no nível 2; <sup>3</sup>(e): Variância dos erros residuais no nível 1

Após a verificação da existência de variâncias no componente de efeitos aleatórios de intercepto para o modelo nulo, foi estimada a “intra-class correlation” (icc) para avaliar a proporção da variância do nível UF em relação à variância total dos termos de erro, isto é, calcular o percentual das diferenças na renda média que se devem à diferença entre UF. Com base no resultado obtido pelo algoritmo, obtemos a Equação (14):

$$icc = \frac{0,01181730}{0,01181730 + 0,07865434} = 13,06\% \quad (14)$$

Inicialmente, foi possível concluir que, para a amostra utilizada nesta análise após todas as suas transformações, identificou-se variância no componente de efeitos aleatórios de intercepto. Essa, por sua vez, se mostrou estatisticamente significativa, pois o p-valor do componente u<sub>0j</sub> foi menor do que 0,05. Além disso, com base no cálculo da correlação intraclass, avaliando a proporção da variância em relação à variância total dos termos de erro. Este resultado indica que aproximadamente 13,06% da variação na RFPC pode ser atribuída às diferenças entre os estados brasileiros. Ou seja, os contextos regionais exercem impacto significativo na renda familiar per capita disponível, validando a escolha pela modelagem multinível. Em outras palavras, ao considerar o conjunto de dados aninhados em contextos, foi possível afirmar que existem características das unidades federativas que impactaram no desempenho global da RFPC, entendendo-se que o ponto de partida da RFPC foi diferente entre estados brasileiros.

Logo em seguida, foi realizada uma comparação entre o “LogLik” do modelo OLS nulo contra o modelo HLM2 nulo (Figura 4).



**Figura 4.** Comparação dos “LogLik” OLS nulo Box-Cox e HLM2 nulo Box-Cox

Fonte: Resultados originais da pesquisa

Nota: Eixo X: Modelos comparados; Eixo Y: Magnitude do “Log Likelihood”. Quanto mais próximo de zero o valor do “LogLik”, melhor o ajuste do modelo aos dados

### Estimação do modelo com interceptos aleatórios HLM2 e variáveis explicativas de nível 1

Após as constatações iniciais sobre a existência de efeitos aleatórios no modelo nulo, foram adicionadas as variáveis explicativas média de anos de estudo dos moradores dos domicílios e nível de segurança alimentar ao modelo, com o objetivo de verificar os seus impactos sobre a variável dependente. Na estratégia “step-up”, as variáveis são incluídas uma a uma, permitindo a verificação de significância estatística de cada uma. Caso alguma variável não passe no teste de significância, ela é removida do modelo. Nesse presente caso, ambas as variáveis se mostraram estatisticamente relevantes.

A Tabela 7 apresenta os resultados dos modelos HLM2 nulo e HLM2 com interceptos aleatórios, comparando os efeitos de diferentes variáveis explicativas no desempenho da renda disponível familiar per capita. O Modelo 1 incluiu apenas os efeitos de intercepto, enquanto o Modelo 2 incorporou variáveis explicativas como a média de anos de estudos e o nível de segurança alimentar. A tabela também apresenta os critérios de avaliação do modelo, como os valores do logaritmo de máxima verossimilhança, “Akaike Information Criterion” (AIC) e “Bayesian Information Criterion” (BIC), para cada um dos modelos.

**Tabela 7.** Output do modelo HLM2 com interceptos aleatórios e variáveis explicativas

Nome	Modelo 1	Modelo 2
	HLM2 Nulo	HLM2 Interceptos
(Intercepto)	4,35 ***	4,17 ***
Erro padrão	(0,02)	(0,02)
Desvio padrão (Intercepto)	0,11	0,08
Desvio padrão (Observações)	0,29	0,25
media_anos_estudo_moradores		0,03 ***
SegurancaInseg_leve		-0,12 ***
SegurancaInseg_moderada		-0,16 ***
SegurancaInseg_grave		-0,19 ***
Número de observações	44.717,00	
Log Likelihood (“LogLik”)	-7.700,60	-1.545,86
“Akaike Information Criterion” (AIC)	15.407,21	3.105,71
“Bayesian Information Criterion” (BIC)	15.433,33	3.166,67

Fonte: Resultados originais da pesquisa

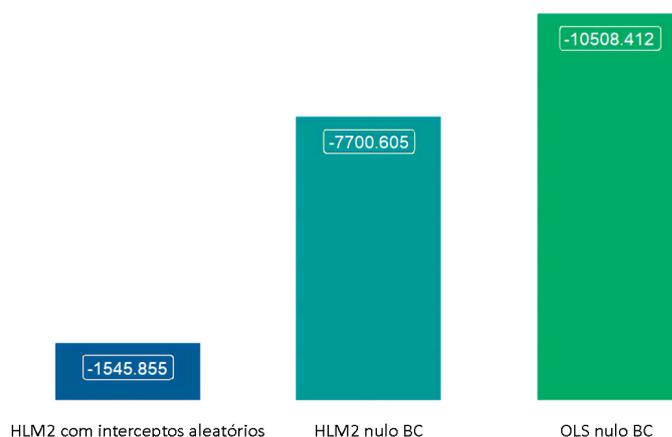
Nota: Códigos de significância: \*\*\*p < 0,001; \*\*p < 0,01; \*p < 0,05

No Modelo 2, a inclusão da média de anos de estudo e do nível de segurança alimentar proporcionou um incremento significativo na maximização do “LogLik”, evidenciando que essas variáveis capturam parte da variabilidade da renda disponível familiar per capita. Os critérios AIC e BIC também apontam para uma melhoria no ajuste do modelo com a adição das variáveis explicativas. Essa etapa confirma que tanto o nível educacional quanto a condição de segurança alimentar são fatores determinantes para explicar as disparidades na renda familiar per capita, reforçando a relevância teórica e empírica desses indicadores no contexto da análise multinível.

De acordo com a Pesquisa Nacional sobre Insegurança Alimentar no Contexto da Pandemia da covid-19 no Brasil (II VIGISAN), existe uma relação direta entre o nível de educação e a segurança alimentar. Famílias cujo chefe do domicílio possuía baixo nível educacional (4 anos de estudo ou menos) apresentaram maior prevalência de insegurança alimentar severa<sup>[20]</sup>.

Além disso, um estudo de Palmeira et al.<sup>[12]</sup> demonstrou que a renda familiar e os programas de transferência condicionada de renda, como o Bolsa Família, têm um impacto significativo na superação da insegurança alimentar. Isso mostrou que políticas voltadas para a proteção social, segurança alimentar e redução da pobreza foram essenciais para manter a segurança alimentar ao longo do tempo, especialmente em regiões mais vulneráveis, como o Nordeste do Brasil.

Na Figura 5 são apresentados os “LogLik” dos 3 modelos estimados anteriormente.

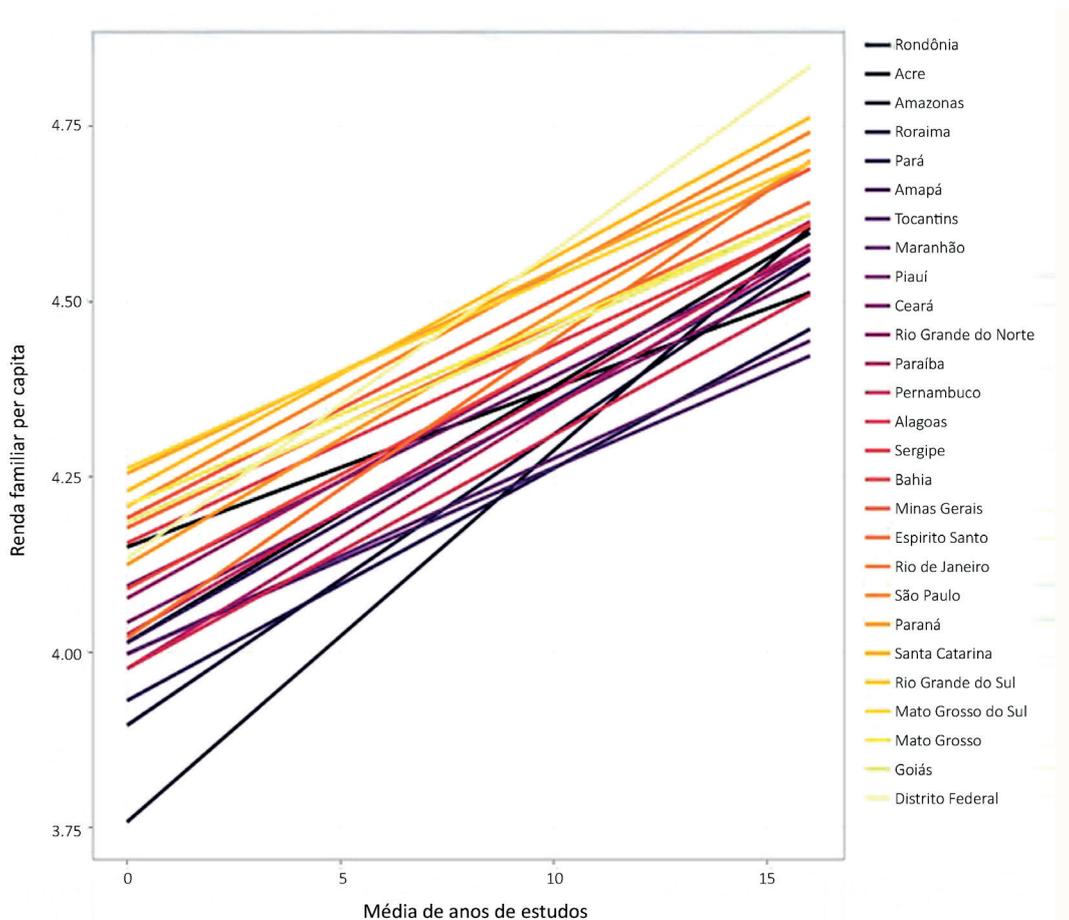


**Figura 5.** Comparação dos “LogLik” OLS nulo Box-Cox, HLM2 nulo Box-Cox e HLM2 com interceptos aleatórios

Fonte: Resultados originais da pesquisa

Nota: Eixo X: Modelos comparados; Eixo Y: Magnitude do "Log Likelihood". Quanto mais próximo de zero o valor do "LogLik", melhor o ajuste do modelo aos dados

A Figura 6 apresenta a visualização da variável RFPC em função da média de anos de estudo dos moradores, com ajuste linear por UF. Observam-se, de forma clara, as diferenças de intercepto capturadas pelos modelos anteriores, assim como indícios de que foram testadas estatisticamente no modelo com interceptos e inclinações aleatórias apresentado a seguir.



**Figura 6.** Visualização dos interceptos aleatórios da variável renda familiar per capita (RFPC) transformada por Box-Cox em função da média de anos de estudos dos moradores  
 Fonte: Resultados originais da pesquisa

**Estimação do modelo com interceptos e inclinações aleatórias HLM2**

Com o modelo HLM2, com interceptos e inclinações aleatórias, foi possível identificar diferentes inclinações nas trajetórias da RFPC quando considerados contextos hierárquicos em diferentes unidades federativas. Esses resultados sugerem que características de diferentes estados brasileiros podem influenciar de forma positiva ou negativa no comportamento da RFPC. Na Tabela 8 são apresentados os resultados dos efeitos aleatórios estimados para intercepto e inclinação, sendo que, para esse caso, os efeitos de inclinação consideraram apenas a variável média dos anos de estudo. As estimativas indicam que ambos os componentes aleatórios (intercepto e inclinação) são estatisticamente significativos, reforçando a robustez da modelagem hierárquica.

**Tabela 8.** Output do algoritmo para determinação dos erros padrão das variâncias no componente de efeitos aleatórios

Componente de efeitos aleatórios	Estimativa de variância	Erro padrão	Z <sup>1</sup>	p-valor
(u0j) <sup>2</sup>	0,0101689998	2,974808e-03	3,418372	0,001
(u1j) <sup>3</sup>	0,0000307276	9,792210e-06	3,137964	0,002
(e) <sup>4</sup>	0,0621416774	4,158559e-04	149,430800	0,000

Fonte: Resultados originais da pesquisa

Nota: <sup>1</sup>Z: Estatística padronizada utilizada para avaliar a significância estatística das variâncias calculada com base na razão entre a estimativa da variância e seu erro padrão; <sup>2</sup>(u0j): Variância dos interceptos aleatórios no nível 2; <sup>3</sup>(u1j): Variância das inclinações aleatórias no nível 2; <sup>4</sup>(e): Variância dos erros residuais no nível 1

A Tabela 9 apresenta os parâmetros do modelo nos elementos de efeitos fixos e a significância estatística das variáveis explicativas. Os coeficientes das variáveis explicativas foram fornecidos juntamente com seus respectivos

erros padrão, t-valor e p-valor, que indicam a significância de cada variável no modelo. Variáveis como a que representam a média de anos de estudo dos moradores e níveis de insegurança alimentar apresentam significância estatística, evidenciando seu impacto no desempenho da variável de estudo.

**Tabela 9.** Saída do modelo HLM2 com interceptos e inclinações aleatórias e a significância estatística das variáveis

Parâmetro	Valor	Erro padrão	t-valor	p-valor
(Intercepto)	4,175833	0,016969312	246,08143	0,00e+000
Media_anos_estudo_moradores	0,028896	0,001032089	27,99720	7,40e-149
Insegurança leve	-0,122494	0,004146906	-29,53877	6,42e-291
Insegurança moderada	-0,153910	0,005843170	-26,34017	2,82e-272
Insegurança grave	-0,185665	0,012566854	-14,77421	7,39e-260

Fonte: Resultados originais da pesquisa

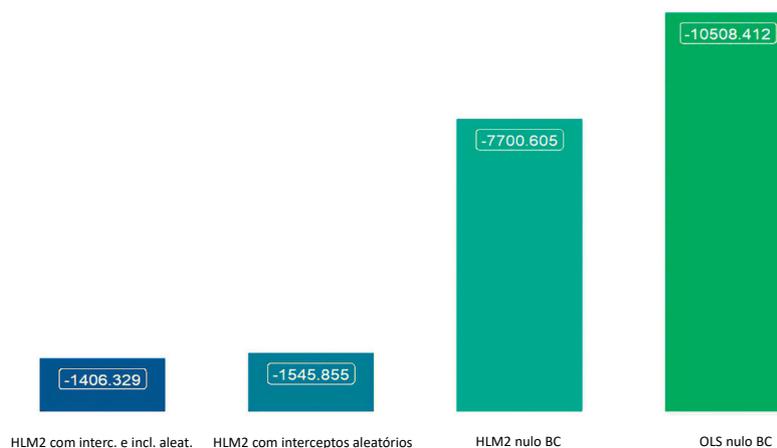
Os resultados indicam que o coeficiente da média de anos de estudo dos moradores apresentou efeito positivo estatisticamente significativo, evidenciando que o aumento da escolaridade está associado ao crescimento da RFPC. Os níveis de insegurança alimentar mostraram efeitos negativos crescentes conforme a gravidade da situação, com a insegurança grave exercendo o maior impacto negativo sobre a renda.

Esses achados estão em consonância com os resultados obtidos nos modelos anteriores, mas o modelo com interceptos e inclinações aleatórias, oferece uma compreensão mais refinada ao capturar a variação contextual entre as unidades federativas, tanto no ponto de partida (intercepto quanto no comportamento da variável anos de estudo (inclinação)).

A literatura corrobora a relação inversa entre renda e insegurança alimentar, evidenciando que menores níveis de renda estão associados a maiores graus de insegurança alimentar. Estudos como o de Palmeira et al.<sup>[12]</sup> que analisam dados do Nordeste brasileiro, e destacam a importância do acesso a programas governamentais para mitigar os efeitos da insegurança alimentar, especialmente em contextos de alta vulnerabilidade socioeconômica. Tais resultados reforçam a relevância de políticas públicas voltadas à segurança alimentar, particularmente em regiões mais desassistidas. Os achados de Palmeira et al.<sup>[12]</sup> sugerem que intervenções focadas no aumento e na estabilização da renda familiar podem ser eficazes na redução da insegurança alimentar.

Portanto, embora o presente estudo tenha se concentrado em modelar a relação entre educação, renda e segurança alimentar, estudos futuros podem aprofundar a análise considerando outras metodologias e/ou variáveis para explorar essas relações de maneira mais detalhada.

Na figura 7 são apresentados os "LogLik" dos quatro modelos estimados. O "LogLik" é uma medida estatística que avalia a qualidade do ajuste de um modelo aos dados. Quanto maior o valor do "LogLik", melhor o ajuste do modelo. É importante destacar que o "LogLik" é sempre um número negativo, e valores mais próximos de zero indicam um melhor ajuste. Observou-se um incremento no "LogLik" quando se considerou a variação aleatória de inclinação, o que indica que este modelo apresenta um ajuste superior aos demais.



**Figura 7.** Comparação dos "LogLik" OLS nulo Box-Cox, HLM2 nulo Box-Cox, HLM2 com interceptos aleatórios e HLM2 com interceptos e inclinações aleatórias

Fonte: Resultados originais da pesquisa

Nota: Eixo X: Modelos Comparados; Eixo Y: "Log Likelihood". Quanto mais próximo de zero o valor do "LogLik", melhor o ajuste do modelo aos dados

### Visualizações dos coeficientes aleatórios de intercepto e de inclinação

A Tabela 10 apresenta os componentes de efeitos aleatórios estimados para cada UF no modelo com interceptos e inclinações aleatórias. Os valores incluem tanto os efeitos aleatórios de interceptos quanto os efeitos aleatórios de inclinações associados às variáveis explicativas média de anos de estudo dos moradores e as categorias da variável segurança alimentar. As colunas indicam as variações nos efeitos de intercepto e inclinação para cada UF, refletindo a influência regional sobre os resultados do modelo.

**Tabela 10.** Componentes de efeitos aleatórios por Unidade Federativa

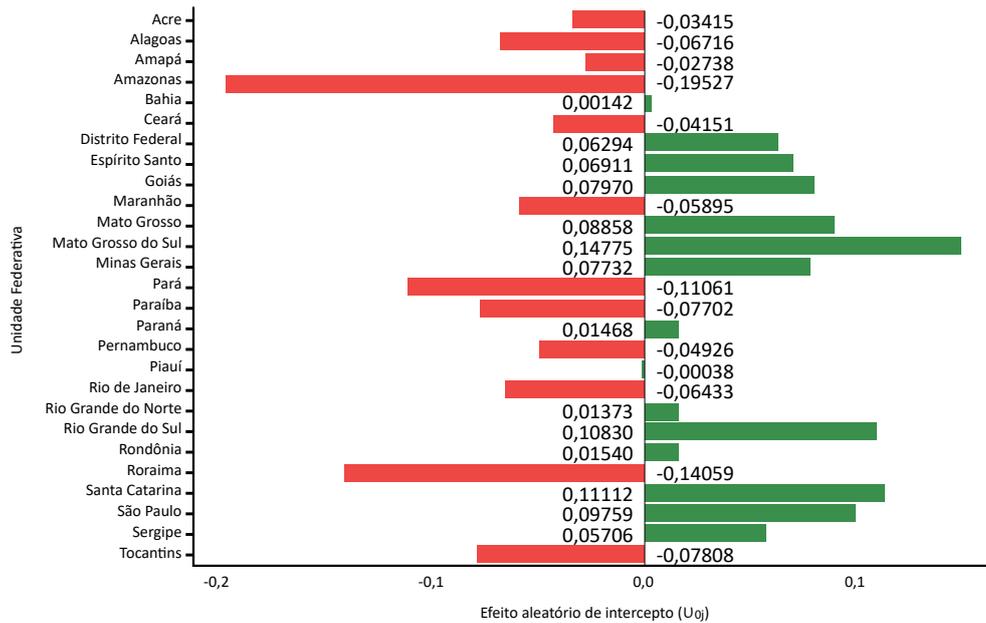
Unidade Federativa	u00 <sup>1</sup>	u10 <sup>2</sup>	u20 <sup>3</sup>	u30 <sup>4</sup>	u40 <sup>5</sup>
Rondônia	0,0154003	-0,0063865	0,0211899	0,0060529	0,0424659
Acre	-0,0341546	0,0018084	-0,0087935	-0,0086765	-0,0361680
Amazonas	-0,1952669	0,0132817	-0,0379009	-0,0487265	-0,1157089
Roraima	-0,1405920	0,0050410	-0,0229215	-0,0297751	-0,1027593
Pará	-0,1106133	-0,0012369	-0,0043164	-0,0198384	-0,0540591
Amapá	-0,0273824	-0,0003335	0,0021335	-0,0075157	-0,0113790
Tocantins	-0,0780810	-0,0054895	-0,0111281	-0,0088028	-0,0867533
Maranhão	-0,0589523	-0,0045038	0,0062224	-0,0140097	-0,0182118
Piauí	-0,0003792	-0,0027726	0,0100383	0,0021569	0,0242925
Ceará	-0,0415119	-0,0017373	0,0101687	-0,0133647	0,0016436
Rio Grande do Norte	0,0137307	-0,0007649	0,0003660	0,0025197	-0,0068146
Paraíba	-0,0770227	0,0032810	-0,0202713	-0,0207158	-0,0572381
Pernambuco	-0,0492627	0,0012613	0,0047447	-0,0101536	0,0149216
Alagoas	-0,0671558	-0,0021005	-0,0090820	-0,0113055	-0,0612922
Sergipe	0,0570587	-0,0035271	0,0182839	0,0104800	0,0816295
Bahia	0,0014203	-0,0000367	-0,0021640	0,0051760	0,0030620
Minas Gerais	0,0773159	-0,0008516	0,0080276	0,0249986	0,0420060
Espírito Santo	0,0691092	-0,0028891	-0,0005897	0,0151628	-0,0103696
Rio de Janeiro	-0,0643259	0,0089107	-0,0162268	-0,0187155	-0,0195911
São Paulo	0,0975896	0,0016863	0,0001756	0,0117213	0,0389022
Paraná	0,0146808	0,0026636	0,0055699	-0,0049551	0,0308435
Santa Catarina	0,1111214	-0,0017537	0,0043991	0,0274801	0,0307024
Rio Grande do Sul	0,1082989	0,0014078	0,0048356	0,0227220	0,0653283
Mato Grosso do Sul	0,1477542	-0,0044978	0,0192040	0,0387082	0,0838795
Mato Grosso	0,0885766	-0,0051442	0,0203743	0,0226796	0,0668672
Goiás	0,0797015	-0,0042571	0,0059124	0,0212470	0,0206919
Distrito Federal	0,0629428	0,0089412	-0,0082517	0,0054500	0,0331090

Fonte: Resultados originais da pesquisa

Nota: <sup>1</sup>u00: efeitos aleatórios de intercepto; <sup>2</sup>u10: efeito aleatório de inclinação em razão da média de anos de estudo; <sup>3</sup>u20: efeito aleatório de inclinação em razão da categoria insegurança alimentar leve; <sup>4</sup>u30: efeito aleatório de inclinação em razão da categoria insegurança alimentar moderada; <sup>5</sup>u40: efeito aleatório de inclinação em razão da categoria insegurança alimentar grave

A situação segurança da variável segurança alimentar foi capturada pelo intercepto após o procedimento de criação de variáveis n-1 “dummy” na modelagem.

Para observar graficamente o comportamento dos valores de u0j, ou seja, dos efeitos de interceptos aleatórios por UF, foi gerada a Figura 8, na qual foi possível observar os impactos positivos e negativos no intercepto de cada uma das unidades federativas. Essa foi a visualização de dados que permitiu demonstrar os efeitos aleatórios e constatar os resultados numéricos do modelo, bem como a significância estatística para existência de efeitos aleatórios de intercepto. Em outras palavras, existem características dos estados que fazem com que o ponto de partida da RFPC seja diferente e sofra um incremento positivo ou negativo no desempenho geral, a depender da UF em que o domicílio está localizado.

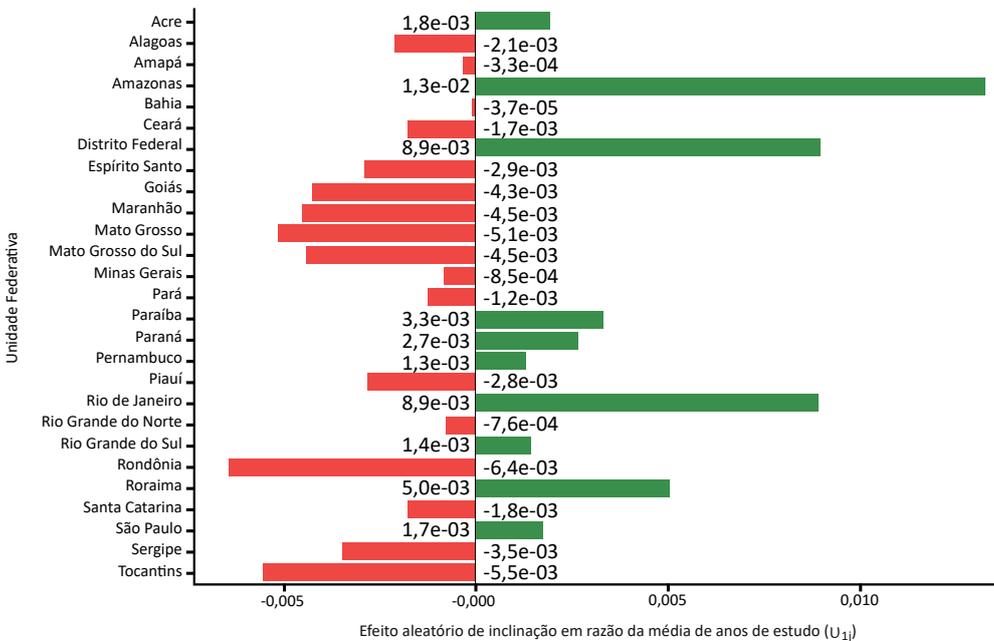


**Figura 8.** Efeitos aleatórios de intercepto por Unidade Federativa

Fonte: Resultados originais da pesquisa

Da mesma forma que foram avaliadas graficamente as variações aleatórias de intercepto para cada UF, foi possível gerar visualizações similares considerando os efeitos aleatórios de inclinação para cada uma das variâncias capturadas pelo modelo.

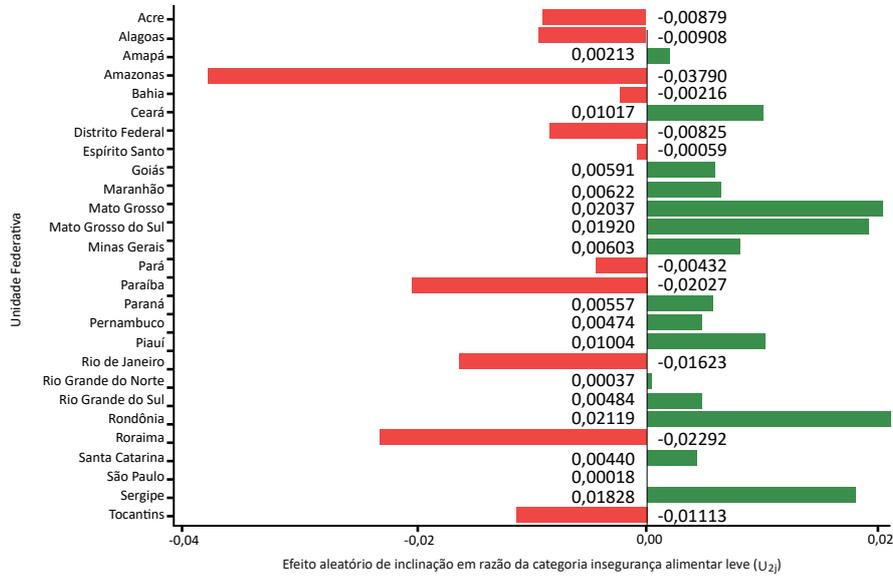
Na Figura 9 observa-se os efeitos aleatórios de inclinação em razão da variável média de anos de estudo dos moradores para cada uma das unidades federativas. Observou-se, portanto, que a quantidade média de anos de estudo dos moradores de um determinado município em uma determinada UF impactou de forma diferente a taxa de crescimento da RFPC, a depender da UF que o domicílio estava situado, isto é, houve coeficientes de inclinações aleatórias positivas para determinadas unidades federativas e negativas para outras.



**Figura 9.** Efeitos aleatórios de inclinação por Unidade Federativa em razão da variável explicativa média de anos de estudo dos moradores por domicílio

Fonte: Resultados originais da pesquisa

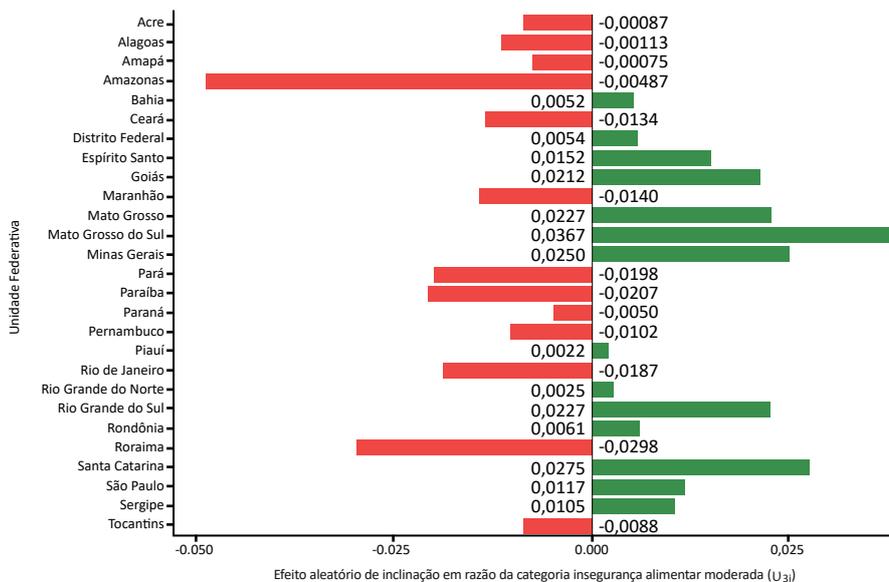
Na Figura 10, são apresentados os efeitos aleatórios de inclinação por UF em razão da variável explicativa nível de segurança alimentar, especificamente quando a categoria foi insegurança alimentar leve.



**Figura 10.** Efeitos aleatórios de inclinação por Unidade Federativa em razão da variável explicativa nível de segurança alimentar quando a categoria for insegurança leve  
 Fonte: Resultados originais da pesquisa

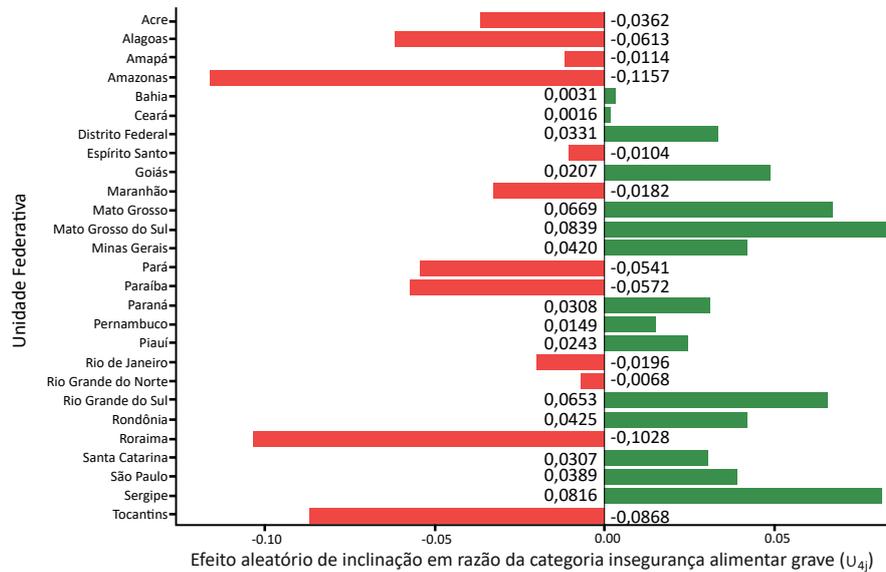
Assim como observado na Figura 9, onde se analisaram os efeitos aleatórios de inclinação da média de anos de estudo por UF, também se observou os coeficientes aleatórios de inclinação em razão da segurança alimentar do domicílio por UF (Figura 10). Em outras palavras, mesmo a categoria menos severa de insegurança alimentar impacta a renda de maneira diferenciada conforme a UF. À medida que se considera níveis mais severos de insegurança alimentar, observam-se diferenças ainda mais acentuadas nos coeficientes das inclinações aleatórias em cada UF.

Na Figura 11 apresenta-se os efeitos aleatórios de inclinação por UF em razão da categoria insegurança moderada da variável original categórica nível de segurança alimentar do domicílio.



**Figura 11.** Efeitos aleatórios de inclinação por Unidade Federativa em razão da variável explicativa nível de segurança alimentar quando a categoria for insegurança moderada  
 Fonte: Resultados originais da pesquisa

Na Figura 12 apresenta-se os efeitos aleatórios de inclinação por UF em razão da categoria insegurança grave da variável original categórica nível de segurança alimentar do domicílio.



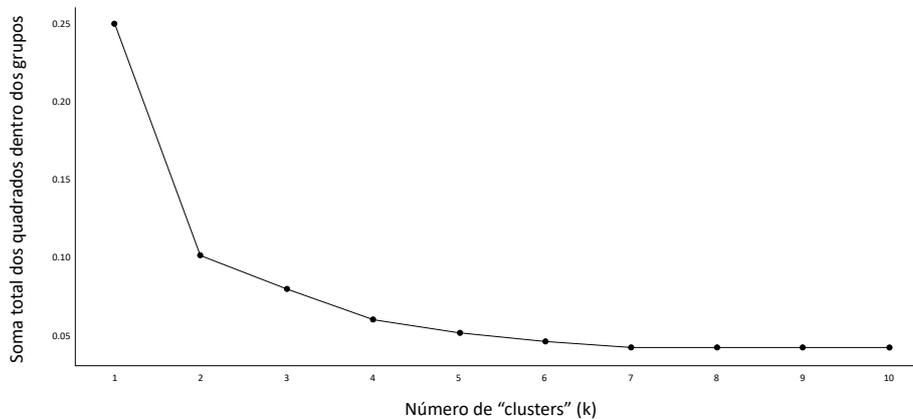
**Figura 12.** Efeitos aleatórios de inclinação por Unidade Federativa em razão da variável explicativa nível de segurança alimentar quando a categoria for insegurança grave

Fonte: Resultados originais da pesquisa

As figuras reforçam que, quanto mais grave a condição de segurança alimentar, maior foram a variação dos efeitos de inclinação entre os estados, sugerindo diferentes graus de resiliência social e econômica frente à vulnerabilidade alimentar.

Para resumir os resultados dos coeficientes aleatórios de interceptos e as inclinações estimados no modelo proposto, fez-se uso da técnica multivariada exploratória de agrupamento, ou “clustering”. Análises de agrupamento têm como um dos objetivos a alocação das observações em grupos homogêneos internamente e heterogêneos entre si. Para visualização dos resultados, fez-se uso do “clustering” por meio do qual foram obtidos os comportamentos semelhantes entre as UF em razão dos coeficientes aleatórios estimados.

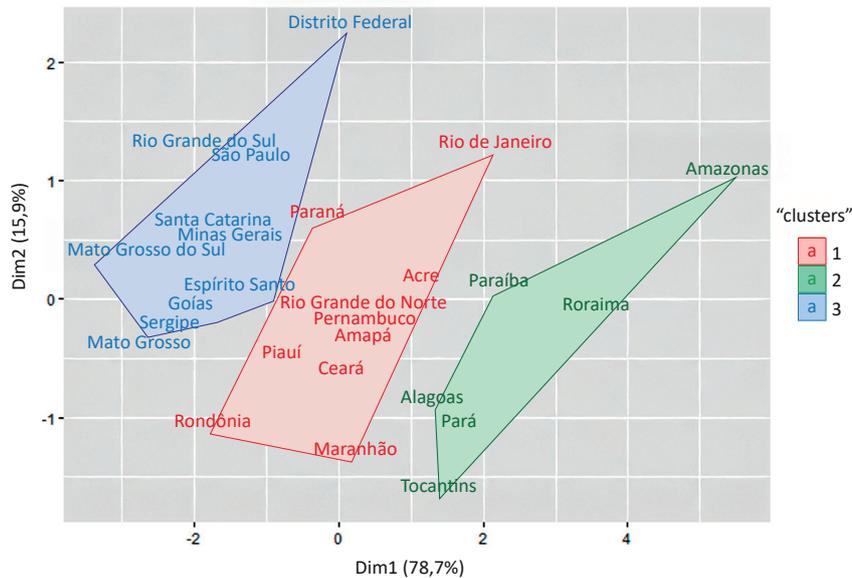
Neste estudo, utilizou-se a metodologia de clusterização “k-means” em conjunto com o “Scree plot” para identificação da quantidade ideal de “clusters”. Em outras palavras, procurou-se agrupar as unidades federativas que possuíam coeficientes aleatórios mais semelhantes entre si e, por meio do “Scree plot”, definiu-se 3 “clusters” agrupando as unidades federativas. Na Figura 13, é apresentado o gráfico com a queda da variância “intra-clusters” em função do número de “clusters” obtido pelo “Scree plot”.



**Figura 13.** “Scree plot” para definição do número de “clusters”

Fonte: Resultados originais da pesquisa

Na Figura 14 observa-se os “clusters” formados a partir da metodologia aplicada nos coeficientes agrupados para cada Unidade Federativa semelhante.



**Figura 14.** “Clusters” das Unidades Federativas

Fonte: Resultados originais da pesquisa

Nota: Dim1 (78,7%): representa a primeira dimensão da análise de componentes principais, que explica 78,7% da variância nos dados; Dim2 (15,9%): representa a segunda dimensão da análise, explicando 15,9% da variância nos dados

Na Tabela 11 observa-se as médias dos coeficientes aleatórios para cada "cluster" gerado, facilitando a visualização das características entre grupos.

**Tabela 11.** Output do algoritmo para determinação dos erros padrão das variâncias no componente de efeitos aleatórios

"Cluster"	Número	Média				
		$u00^1$	$u10^2$	$u20^3$	$u30^4$	$u40^5$
1	11	-0,0209	-0,0001	0,00250	-0,00478	0,00291
2	6	-0,1115	0,0021	-0,01692	-0,02433	-0,08045
3	10	0,0899	-0,0011	0,00740	0,01986	0,04506

Fonte: Resultados originais da pesquisa

Nota:  $u00^1$ : Média do efeito aleatório de intercepto;  $u10^2$ : Média do efeito aleatório de inclinação em razão da média de anos de estudo;  $u20^3$ : Média do efeito aleatório de inclinação em razão da categoria insegurança alimentar leve;  $u30^4$ : Média do efeito aleatório de inclinação em razão da categoria insegurança alimentar moderada;  $u40^5$ : Média do efeito aleatório de inclinação em razão da categoria insegurança alimentar grave

Os resultados médios dos “clusters” de unidades federativas demonstraram as características de similaridade das variações na renda disponível em razão da média de anos de estudo e da condição de segurança alimentar. Avaliando a renda em contextos de UF, foi possível notar os efeitos dos anos de estudo e da segurança alimentar de forma distinta em cada estado. Avaliando em termos de desigualdades sociais e implementação de políticas públicas, o presente estudo apresentou uma forma de analisar a RFPC em diferentes contextos, apresentando desde o grupo de estados com as menores rendas médias (“cluster” 2 na Figura 14) até o grupo de estados com as maiores rendas médias (“cluster” 3 na Figura 14), passando pelo “cluster” 1, que apresentou resultados intermediários no comportamento das variações e média da renda quando comparado aos “clusters” 2 e 3.

#### 4. Conclusão

Os resultados deste estudo indicaram que a média de anos de estudo dos moradores exerce um impacto positivo e estatisticamente significativo sobre a renda familiar per capita disponível. Por outro lado, a condição de segurança alimentar demonstrou ser um fator crucial na explicação das desigualdades de renda, uma vez que os domicílios em situação de insegurança — especialmente nos níveis moderado e grave — apresentaram reduções significativas na RFPC. Além disso, foram observadas diferenças expressivas entre as unidades federativas, com estados das regiões Norte e Nordeste que apresentaram menores pontos de partida (interceptos) da renda em comparação aos estados do Sul e Sudeste.

Concluiu-se que políticas públicas voltadas à redução das desigualdades sociais e econômicas devem levar em consideração essas variações regionais, promovendo ações mais específicas e sensíveis às particularidades — locais especialmente nos níveis moderado e grave — apresentaram reduções significativas na RFPC, especialmente em estados mais severamente afetados pela insegurança alimentar.

Ademais, este estudo apresentou limitações, como a ausência de uma análise detalhada sobre a interação entre variáveis contextuais e individuais, a aplicação da transformação de Box-Cox de maneira geral sem considerar contextos específicos, e a não utilização de pesos amostrais, o que pode ter impactado a representatividade das estimativas. Esses pontos sugerem a necessidade de replicar em estudos futuros a análise com ajustes por pesos, bem como considerar a inclusão de outras variáveis relevantes — como acesso a serviços públicos, infraestrutura e indicadores de saúde —, além da aplicação de metodologias mais recentes e a análise de dados longitudinais para uma compreensão mais aprofundada das dinâmicas socioeconômicas.

Em suma, os achados deste estudo contribuem para o debate sobre desigualdade regional no Brasil, evidenciando como variáveis educacionais e de segurança alimentar se manifestam de maneira distinta nos territórios, e reforçam a necessidade de estratégias públicas integradas e orientadas por evidências estatísticas.

**Contribuição dos autores:** Gomes, E.B.Q.: Conceitualização; Definição da Metodologia; Aquisição de Dados; Escrita e Edição. Tarantin Junior, W.: Conceitualização; Definição da Metodologia; Escrita e Edição.

**Agradecimentos:** Ao corpo docente do MBA em Data Science e Analytics da USP/Esalq pelo compartilhamento de conhecimento e suporte acadêmico essencial ao desenvolvimento desta pesquisa.

**Como citar:** Gomes, E.B.Q.; Tarantin Junior, W. 2025. Efeitos das unidades federativas na renda disponível per capita por domicílio: uma análise multinível. *Quaestum* 6: e2675824.

## Referências

- [1] Dedecca, C.S.; Trovão, C.J.B.M. Sobre desigualdades no Brasil: passado, presente e futuro. *RBEST Revista Brasileira de Economia Social e do Trabalho* 2(00): e020015. <https://doi.org/10.20396/rbest.v2i00.14825>.
- [2] Instituto Brasileiro de Geografia e Estatística [IBGE]. 2019. Pesquisa de orçamentos familiares 2017-2018. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv101670.pdf>. Acesso em: 14 out. 2021.
- [3] Hoffmann, R.; Vaz, D.V. 2021. Mensurando a desigualdade no Brasil: evidências a partir da renda e dos gastos das famílias. *RBEST Revista Brasileira de Economia Social e do Trabalho* 3(00): e021003. <https://doi.org/10.20396/rbest.v3i00.14841>.
- [4] Camargo, D.A.; Satolo, L.F. 2018. Padrões de consumo alimentar baseados no orçamento das famílias na região metropolitana de São Paulo. *Segurança Alimentar e Nutricional*. 25(3): 94-103. <https://doi.org/10.20396/san.v25i3.8652327>.
- [5] Ferreira, R.C.; Senna, M.I.B.; Rodrigues, L.G.; Campos, F.L.; Martins, A.E.B.L.; Kawachi, I. 2020. Education and income-based inequality in tooth loss among Brazilian adults: does the place you live make a difference? *BMC Oral Health* 20(246): 2-13. <https://doi.org/10.1186/S12903-020-01238-9>.
- [6] Hone, T.; Saraceni, V.; Coeli, C.M.; Trajman, A.; Rasella, D.; Millett, C.; Durovni, B. 2020. Primary healthcare expansion and mortality in Brazil's urban poor: a cohort analysis of 1.2 million adults. *PLOS Medicine* 17(10): e1003357. <https://doi.org/10.1371/JOURNAL.PMED.1003357>.
- [7] Russo, L.X.; Dias, J. 2016. The health influence on returns to education in Brazil: A nonlinear approach. *Economica* 17(2): 210-220. <https://doi.org/10.1016/J.ECON.2016.06.002>.
- [8] Beltrán-Sánchez, H.; Andrade, F.C.D. 2016. Time trends in adult chronic disease inequalities by education in Brazil: 1998–2013. *International Journal for Equity in Health* 15(139). <https://doi.org/10.1186/S12939-016-0426-5>.
- [9] Lazega, E.; Snijders, T.A.B. 2016. *Multilevel network analysis for the social sciences*. Springer, New York, NY, USA.
- [10] Araújo, M.L.; Nascimento, D.R.; Lopes, M.S.; Passos, C.M.; Lopes, A.C.S. 2020. Condições de vida de famílias brasileiras: estimativa da insegurança alimentar. *Revista Brasileira de Estudos da População* 37: 1-17. <https://doi.org/10.20947/S0102-3098a0110>.
- [11] Cherol, C.C.S.; Ferreira, A.A.; Lignani, J.B.; Salles-Costa, R. 2021. Regional and social inequalities in food insecurity in Brazil, 2013-2018. *Cadernos De Saúde Pública* 38(12): e00083822. <https://doi.org/10.1590/0102-311XEN083822>.
- [12] Palmeira, P.A.; Lignani, J.B.; Salles-Costa, R. 2022. Access to governmental programs/benefits and food insecurity in urban and rural areas of Northeast Brazil. *Ciência & Saúde Coletiva* 27(7): 2583-2595. <https://doi.org/10.1590/1413-8123202277.21592021EN>.
- [13] Box, G.E.P.; Cox, D.R. 1964. An analysis of transformations. *Journal of the Royal Statistical Society*. 26(2): 211-252. <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>.
- [14] Courgeau, D. 2003. *Methodology and epistemology of multilevel analysis*. Kluwer Academic Publishers, London, UK.
- [15] Fávero, L.P.; Belfiore, P. 2017. *Manual de análise de dados*. Elsevier, Rio de Janeiro, RJ, Brasil.
- [16] Goldstein, H. 2011. *Multilevel statistical models*. 4ed. John Wiley & Sons, West Sussex, UK.
- [17] Raudenbush, S.W.; Bryk, A.S. 2002. *Hierarchical linear models: applications and data analysis methods*. 2ed. Sage Publications, Thousand Oaks, CA, USA.
- [18] Damayanti, A.P.; Utami, W.D.; Novitasari, D.C.R.; Intan, P.K. 2023. Cluster Analysis of Environmental Pollution in Indonesia Using Complete Linkage Method with Elbow Optimization. *JTAM Jurnal Teori dan Aplikasi Matematika* 7(2): 399-408. <https://doi.org/10.31764/jtam.v7i2.12961>.
- [19] Tabachnick, B.G.; Fidell, L.S. 2013. *Using multivariate statistics*. 6ed. Pearson, Boston, Massachusetts, USA.
- [20] Rede Brasileira de Pesquisa em Soberania e Segurança Alimentar e Nutricional. 2022. *II VIGISAN: Inquérito Nacional sobre Insegurança Alimentar no Contexto da Pandemia da Covid-19 no Brasil – Suplemento I: Insegurança Alimentar nos estados*. Disponível em: <https://olheparaafome.com.br/wp-content/uploads/2022/06/Relatorio-II-VIGISAN-2022.pdf>. Acesso em: 20 dez. 2024.