

## Estimativa do IDH e índices de escolaridade, renda e longevidade a partir da Pesquisa Nacional de Saúde (PNS)

*Estimation of HDI and education, income and longevity indices from the National Health Survey (PNS)*

Tcharlies Dejandir Schmitz<sup>1</sup>, Débora de Faria Ferreira Gomes<sup>1</sup>, Thiago Alam da Silva<sup>1</sup>, André Cristofolini Schneider<sup>1</sup>, Dalvan Antônio de Campos<sup>2\*</sup>

<sup>1</sup> PX Center, Joinville, Santa Catarina, Brasil.

<sup>2</sup> Programa de Pós-Graduação em Ambiente e Saúde, Universidade do Planalto Catarinense, Lages, Santa Catarina, Brasil.

\*Autor para correspondência: dalvandecampos@gmail.com

### RESUMO

O IDH, calculado tradicionalmente pelo IBGE com base em indicadores socioeconômicos e de saúde (como a longevidade obtida das tábuas de mortalidade regionais – PNUD, 2023), é essencial para políticas públicas. Contudo, esses métodos, ao usarem dados agregados, podem não captar as nuances individuais dos estilos de vida. A disponibilização de dados detalhados pela PNS (IBGE, 2020) permite repensar esses cálculos. Assim, o artigo propõe uma nova metodologia que utiliza perguntas específicas da PNS para calibrar um modelo estatístico que estima as componentes de educação, renda e longevidade, reconstruindo o IDH. Essa abordagem integra dados individuais aos tradicionais, oferecendo previsões mais sensíveis às variações regionais e às particularidades dos grupos estudados, enriquecendo o debate metodológico em estatísticas sociais e de saúde.

**Palavras-chave:** índice de desenvolvimento humano; pesquisa nacional de saúde; indicadores socioeconômicos; metodologia estatística.

### ABSTRACT

The HDI, traditionally calculated by the IBGE based on socioeconomic and health indicators (such as longevity obtained from regional mortality tables – UNDP, 2023), is essential for

public policies. However, these methods, by using aggregated data, may not capture the individual nuances of lifestyles. The provision of detailed data by the PNS (IBGE, 2020) allows us to rethink these calculations. Thus, the article proposes a new methodology that uses specific questions from the PNS to calibrate a statistical model that estimates the components of education, income, and longevity, reconstructing the HDI. This approach integrates individual data with traditional data, offering predictions that are more sensitive to regional variations and the particularities of the groups studied, enriching the methodological debate in social and health statistics.

**Keywords:** human development index; national health survey; socioeconomic indicators; statistical methodology.

## 1 INTRODUÇÃO

Utilizando dados individuais da PNS (IBGE, 2020), o artigo propõe uma nova metodologia para reconstruir o IDH, tradicionalmente calculado pelo IBGE a partir de indicadores socioeconômicos e de saúde (PNUD, 2023). Ao calibrar um modelo estatístico com perguntas específicas, a abordagem permite estimar as componentes de educação, renda e longevidade com maior sensibilidade às variações regionais e às nuances dos estilos de vida individuais, contribuindo para o debate metodológico em estatísticas sociais e de saúde.

## 2 MATERIAIS E MÉTODOS

Conduzida pelo IBGE em parceria com o Ministério da Saúde, a PNS coleta dados abrangentes sobre saúde, estilo de vida e condições socioeconômicas, tornando-a uma fonte confiável para calibrar modelos que estimem longevidade, renda e educação. De Brasil (2019), selecionamos perguntas específicas da PNS que capturam elementos desses indicadores para construir o modelo.

- PNS P0090: Quantos dias da semana consome verduras ou legumes?
- PNS P018: Quantos dias da semana consome frutas?
- PNS: P035: Quantos dias da semana realiza exercícios físicos por 30 a 60 minutos (caminhada, alongamento, musculação, outros)?

- PNS P03201: Nos últimos 30 dias, consumiu cinco ou mais doses de álcool em uma ocasião?
- PNS N00101: Considerando saúde como um estado de bem-estar físico e mental, e não somente a ausência de doenças, como é o seu estado de saúde?
- PNS D00901: Qual foi o curso mais elevado que frequentou?
- PNS D014: Concluiu este curso que frequentou
- PNS E01602: Qual era o rendimento bruto mensal ou retirada que fazia normalmente no seu trabalho ou estágio?
- PNS E01802: Qual era o rendimento bruto mensal ou retirada que fazia normalmente em outros trabalhos
- PNS F001021: Valor habitualmente recebido de aposentadoria ou pensão de instituto de previdência federal (INSS), estadual, municipal, ou do governo federal, estadual, municipal
- PNS F007021: Valor habitualmente recebido de pensão alimentícia, doação ou mesada em dinheiro de pessoa que não morava no domicílio
- PNS F008021: Valor habitualmente recebido de rendimento de aluguel ou arrendamento
- PNS VDF00102: Valor recebido em reais normalmente de algum juro de caderneta de poupança e de outras aplicações financeiras, dividendos, programas sociais, seguro-desemprego, seguro defeso ou outros rendimentos?

As questões foram escolhidas para identificar fatores do estilo de vida que impactam a saúde e a longevidade, além de abordar escolaridade e renda. O índice de educação é calculado a partir dos anos de estudo, conforme a metodologia do PNUD, convertendo cada nível de escolaridade no total de anos correspondentes (de 0 a 15 anos) e tirando a média dos anos estudados. Com o modelo desenvolvido, é possível calcular valores condicionais para indivíduos com características específicas, quantificar a incerteza das previsões e construir intervalos de confiança (Cumming, 2012). Para isso, as respostas da PNS são convertidas em pontuações (variável de entrada X), e a média de X é calculada para cada estado junto aos valores médios do IDH, longevidade, educação e renda, com os dados centralizados para facilitar os cálculos de covariância (Triola, 2018).

**Figura 1-** Cálculo da média das variáveis de entrada e variáveis de saída.

$$\mu_X = \frac{1}{n} \sum_{i=1}^n X_i$$

$$\mu_Y = \frac{1}{n} \sum_{i=1}^n y_i$$

A matriz de covariância (Figura 2) representa como as variáveis X e Y variam juntas. Ela é calculada considerando a diferença entre cada dado e sua média correspondente, multiplicada pelos pesos associados a cada ponto. O resultado é uma matriz que descreve a relação entre X e Y, dividida em blocos específicos: a covariância de X com ele mesmo, a variância de Y, e as covariâncias cruzadas entre X e Y (Johnson; Wichern, 2018).

**Figura 2-** Cálculo da matriz de covariância.

$$\Sigma = \text{Cov}([X, y]) = \frac{1}{n-1} \sum_{i=1}^n \begin{bmatrix} X_i - \mu_X \\ y_i - \mu_Y \end{bmatrix} \begin{bmatrix} X_i - \mu_X \\ y_i - \mu_Y \end{bmatrix}^T$$

Após calcular a matriz de covariância conjunta, ela é separada em blocos, como mostrado na Figura 3. A covariância de X com ele mesmo forma um bloco, enquanto a variância de Y compõe outro. As covariâncias cruzadas entre X e Y formam blocos que conectam as duas variáveis. Essa separação permite a utilização de fórmulas específicas para os cálculos condicionais.

**Figura 3-** Definição das covariâncias cruzadas

$$\Sigma_X = \Sigma_{1:m, 1:m}, \quad \Sigma_Y = \Sigma_{m+1, m+1}, \quad \Sigma_{YX} = \Sigma_{m+1, 1:m}, \quad \Sigma_{XY} = \Sigma_{YX}^T$$

Na Figura 4, temos a média condicional de Y dado X é calculada ajustando a média geral de Y com base nos valores observados de X. Esse ajuste considera como X influência Y, utilizando as covariâncias cruzadas e a inversa da covariância de X. O resultado é a média esperada de Y considerando as informações fornecidas por X.

**Figura 4-** Cálculo da média condicional das variáveis de entrada.

$$\mu_{Y|X} = \mu_Y + \Sigma_{YX} \Sigma_X^{-1} (X_{obs} - \mu_X)$$

A variância condicional visualizada na Figura 5, de Y dado X mede a incerteza remanescente em Y depois de considerar X. Esse cálculo subtrai a contribuição de X na explicação da variância total de Y. O valor resultante reflete o quanto ainda não foi explicado sobre Y, mesmo sabendo os valores de X.

**Figura 5-** Cálculo da variância condicional das variáveis de entrada.

$$\text{Var}(Y | X) = \Sigma_Y - \Sigma_{YX} \Sigma_X^{-1} \Sigma_{XY}$$

Com a variância condicional calculada, o erro padrão para um intervalo de confiança de noventa e cinco por cento é determinado. Esse erro fornece uma estimativa da amplitude em que o valor real de Y deve estar dado X, com 95% por cento de confiança, conforme mostrado na Figura 6. Isso é especialmente útil para realizar previsões mais confiáveis.

**Figura 6-** Cálculo a margem de erro com nível de confiança 95%.

$$\text{Erro}_{95\%} = 1.96 \cdot \sqrt{\text{Var}(Y | X)}$$

A metodologia proposta estima valores médios de IDH, longevidade, educação e renda diretamente das perguntas da PNS usando métodos estatísticos e outra derivada dos índices estimados, combinando as três dimensões fundamentais através da média geométrica (Figura 7).

**Figura 7-** Cálculo do IDH a partir dos índices dimensionais

$$\text{IDH} = \sqrt[3]{S \times E \times R}$$

O processo inclui o cálculo do erro propagado (Figura 8), permitindo modelar a incerteza das estimativas dimensionais.

**Figura 8-** Cálculo do erro propagado do IDH a partir dos pelos índices dimensionais

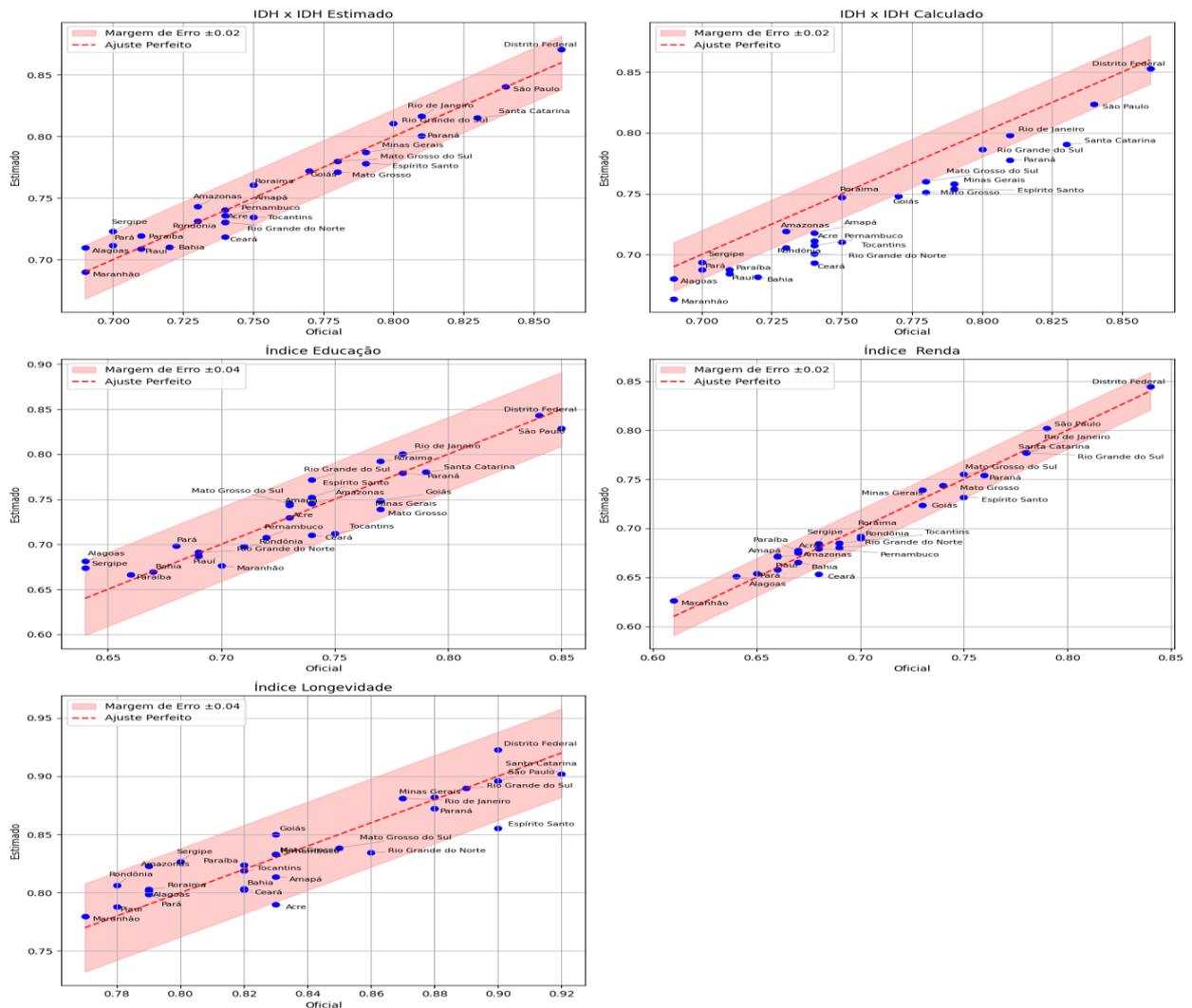
$$\frac{\Delta \text{IDH}}{\text{IDH}} = \frac{1}{3} \sqrt{\left(\frac{\Delta S}{S}\right)^2 + \left(\frac{\Delta E}{E}\right)^2 + \left(\frac{\Delta R}{R}\right)^2}$$

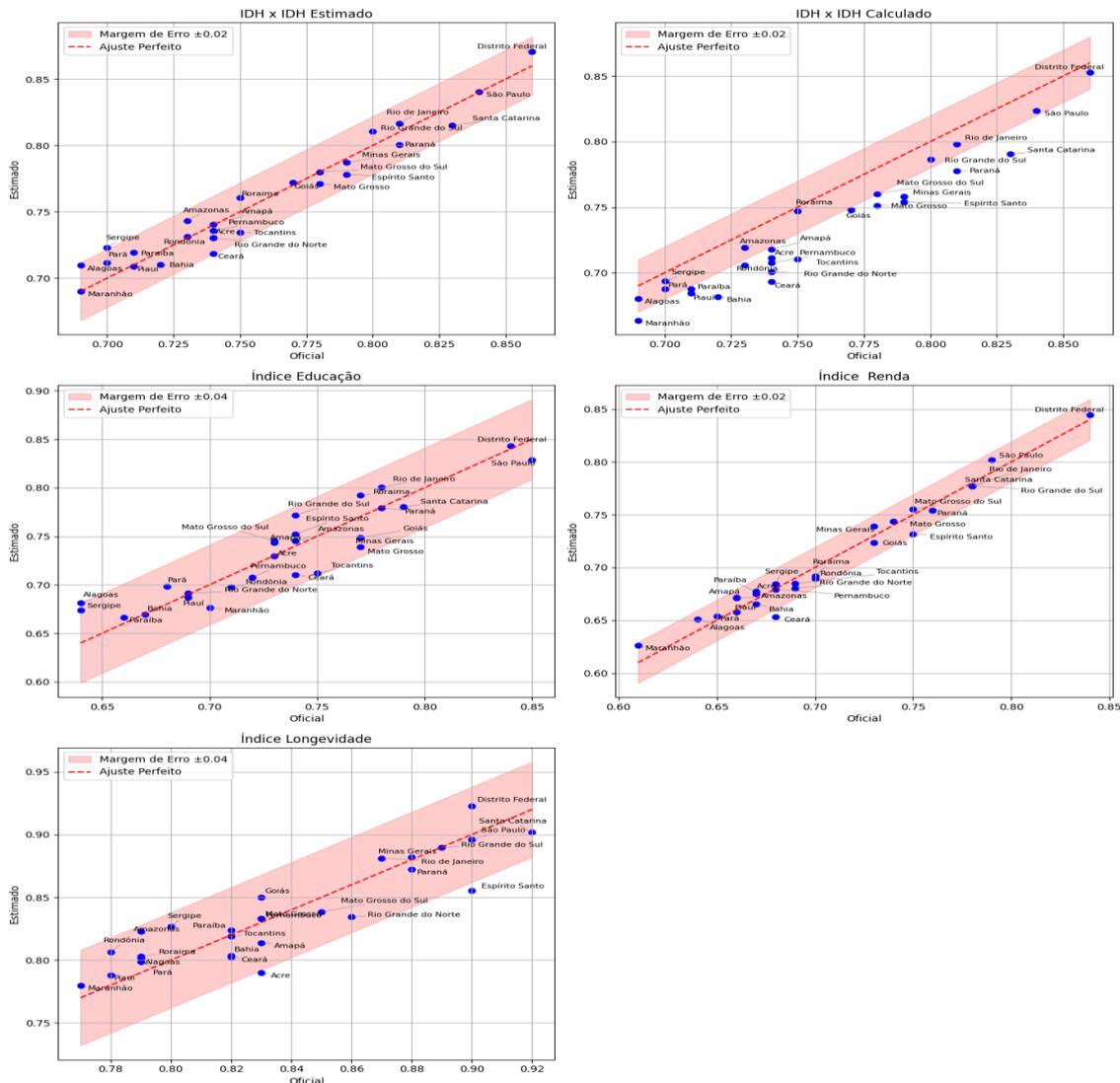
Neste ponto concluímos o processo matemático necessário para obter a os índices de longevidade, educação, renda e IDH a partir das respostas das perguntas selecionadas da PNS.

### 3 RESULTADOS

A análise dos dados evidencia claramente as nuances regionais dos indicadores de longevidade, renda, educação e IDH, refletindo as condições socioeconômicas locais. Essa comparação confirma a robustez da metodologia e demonstra a influência de fatores regionais, como serviços de saúde, infraestrutura e níveis educacionais, conforme ilustrado no gráfico da Figura 9.

**Figura 9-** Gráfico comparativo dos valores estimados com os valores oficiais.





Em síntese, a integração dos dados e a análise detalhada dos indicadores fortalecem a compreensão dos índices de educação, renda e longevidade e do IDH.

## 4 DISCUSSÃO

A nova metodologia integra indicadores individuais da Pesquisa Nacional de Saúde e dados tradicionais do IBGE para estimar o IDH, refletindo melhor as variações regionais e os estilos de vida específicos da população. Ao converter respostas de hábitos alimentares, atividade física e consumo de álcool em variáveis quantitativas, o modelo demonstra robustez

com um pequeno desvio entre os valores reais e os previstos. Essa abordagem permite uma análise detalhada, especialmente na estimativa da longevidade, ao complementar métodos convencionais com informações comportamentais. O uso de intervalos de confiança com um nível de 95% reforça a precisão das previsões e quantifica a incerteza dos resultados.

## 5 CONCLUSÃO

A integração dos dados individuais da Pesquisa Nacional de Saúde com os indicadores tradicionais do IBGE resultou em uma estimativa do IDH mais sensível e representativa, permitindo captar nuances regionais e comportamentais que os métodos convencionais não detectam. A calibração do modelo estatístico por meio de variáveis quantificáveis, associada à determinação de intervalos de confiança robustos, reforça a confiabilidade das previsões e fornece subsídios sólidos para a formulação de políticas públicas direcionadas. Essa abordagem não apenas aprimora a compreensão dos índices de educação, renda e longevidade, mas também abre caminho para pesquisas futuras que possam aprofundar a análise das disparidades regionais e a evolução dos indicadores ao longo do tempo.

## REFERÊNCIAS

BRASIL. Ministério da Saúde. Instituto Brasileiro de Geografia e Estatística. Pesquisa Nacional de Saúde 2019: Questionário. Rio de Janeiro: IBGE, 2019. Disponível em: <https://www.pns.icict.fiocruz.br/wp-content/uploads/2021/02/Questionario-PNS-2019.pdf>. Acesso em: 11 abr. 2025.

CUMMING, G. **Understanding the new statistics**: Effect sizes, confidence intervals, and meta-analysis. Routledge, 2012.

IBGE. Pesquisa Nacional de Saúde 2019: Ciclos de vida: Brasil e Grandes Regiões. Rio de Janeiro: IBGE, 2020.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. Pearson, 2018.

PNUD. Relatório de Desenvolvimento Humano 2023/2024: Rompendo o impasse: remodelando a cooperação em um mundo polarizado. Nova York: Programa das Nações Unidas para o Desenvolvimento, 2023.

TRIOLA, M. M. **Elementary statistics**. Pearson, 2018.