

# Análise não paramétrica com python usando Kruskal-Wallis

Richard Matheus Avelino da Silva (CCT-UEPB)

Thaís Erika da Costa Azevedo (CCT-UEPB)

23 de maio de 2025

## 1 Introdução

A taxa de mortalidade infantil é um dos indicadores mais relevantes de saúde pública e progresso social, espelhando diretamente as condições de vida dos habitantes, o acesso a serviços de saúde, saneamento básico e políticas públicas direcionadas à infância. É crucial para a elaboração e avaliação de estratégias voltadas para a diminuição das desigualdades sociais e regionais no Brasil. Mesmo com os progressos observados nas últimas décadas, a mortalidade infantil ainda apresenta diferenças notáveis entre estados e regiões, afetadas por elementos como a pobreza, a urbanização, a infraestrutura e a disponibilidade de serviços de atenção básica. Na Paraíba, que se divide em quatro grandes regiões - Mata Paraibana, Agreste Paraibano, Borborema e Sertão Paraibano -, essas desigualdades podem ser intensificadas pelas diferenças socioeconômicas e de acesso aos serviços de saúde. Nesse contexto, torna-se crucial examinar se existe uma diferença estatisticamente relevante entre as regiões da Paraíba em relação aos índices de mortalidade infantil. Este trabalho utiliza uma metodologia estatística não paramétrica, mais apropriada considerando a natureza dos dados, com aplicação dos testes de Kruskal-Wallis [1] e Shapiro-Wilk [2], sendo os dados obtidos no site oficial do Instituto Brasileiro de Geografia e Estatística (IBGE) [3], com acesso realizado em 22 de maio de 2025.

## 2 Objetivo

Analisar os índices de mortalidade infantil nos municípios do estado da Paraíba, no ano de 2022, de acordo com sua distribuição regional (Mata Paraibana, Agreste Paraibano, Borborema e Sertão Paraibano), utilizando o teste não paramétrico de Kruskal-Wallis[1]. O objetivo é verificar se existem diferenças estatisticamente significativas nas medianas da mortalidade infantil entre as regiões, sem assumir a normalidade da distribuição dos dados.

## 3 Metodologia

### 3.1 Justificativa

A mortalidade infantil é um dos principais indicadores de qualidade de vida e desenvolvimento humano, refletindo diretamente as condições de saúde, saneamento, acesso a

serviços públicos e desigualdades sociais. No estado da Paraíba, observam-se diferenças socioeconômicas marcantes entre as regiões geográficas, o que pode impactar diretamente nos índices de mortalidade infantil. Diante disso, analisar a distribuição da mortalidade infantil entre as diferentes regiões do estado permite identificar desigualdades territoriais e orientar políticas públicas mais direcionadas. Contudo, como os dados podem não seguir uma distribuição normal e o número de municípios por região varia, torna-se mais apropriado o uso de técnicas estatísticas não paramétricas, que não exigem pressupostos rígidos quanto à distribuição dos dados. Assim, a escolha por métodos não paramétricos — em especial o teste de Kruskal-Wallis [1] — justifica-se por sua robustez diante de possíveis assimetrias ou heterogeneidade de variâncias, tornando a análise mais confiável e adequada à natureza dos dados.

### 3.2 Métodos

Este é um estudo de natureza quantitativa, descritiva e comparativa, com uma abordagem estatística não paramétrica. As informações utilizadas foram extraídas de um banco de dados contendo dados dos municípios do estado da Paraíba, referentes ao ano de 2022, incluindo a taxa de mortalidade infantil e a categorização regional de cada cidade (Mata Paraibana, Agreste Paraibano, Borborema e Sertão Paraibano), extraídos do site oficial do Instituto Brasileiro de Geografia e Estatística (IBGE) [3]. Os dados foram consolidados em uma única base, com a identificação da taxa de mortalidade infantil e da respectiva região de cada município. Inicialmente, foi realizada uma análise descritiva por meio do cálculo de medidas-resumo (mediana, mínimo, máximo e intervalo interquartil, além da construção de gráficos do tipo boxplot para caracterizar o comportamento da variável em cada região. Em seguida, a normalidade dos dados foi testada formalmente utilizando o teste de Shapiro-Wilk [2], cujos resultados indicaram ausência de distribuição normal. Diante disso, foi aplicado o teste de Kruskal-Wallis [1], adequado para comparar as medianas de uma variável numérica entre mais de dois grupos independentes, mesmo quando os pressupostos paramétricos não são atendidos. Todas as análises foram realizadas com o auxílio de software estatístico apropriado, assegurando a reprodutibilidade dos resultados e a consistência metodológica da abordagem utilizada.

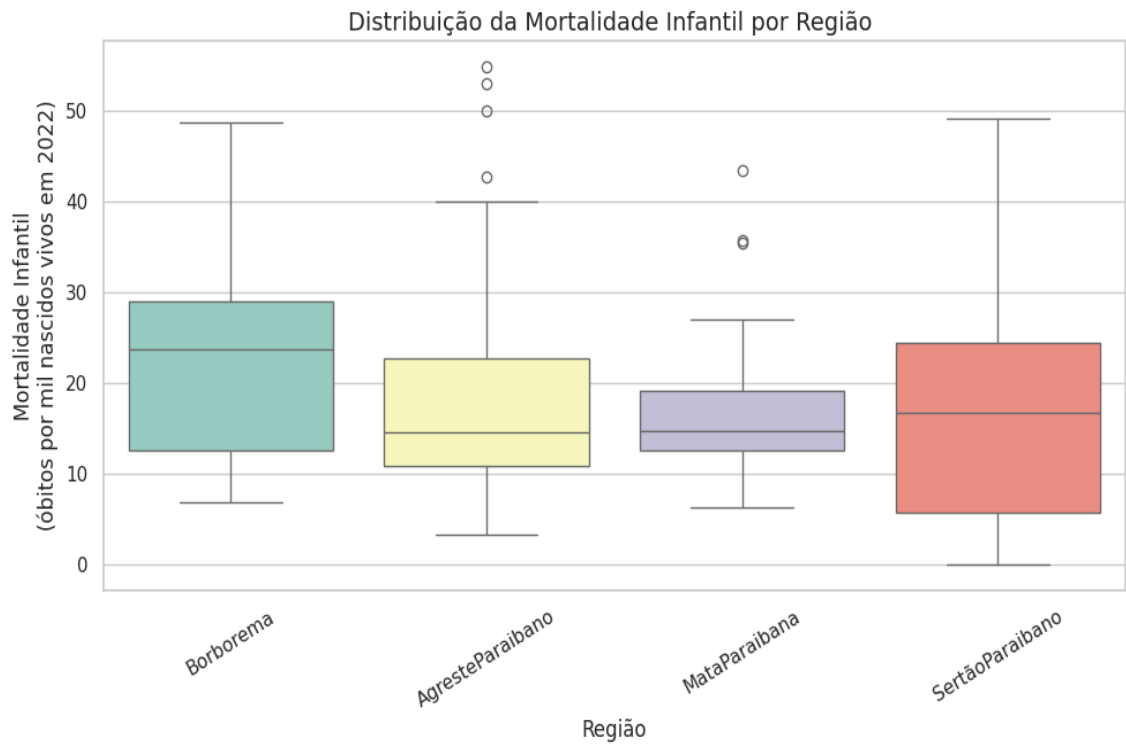
# 4 Resultados

Tabela 1: Medidas descritivas da mortalidade infantil por região na Paraíba (2022)

Medida	Borborema	Agreste Paraibano	Mata Paraibana	Sertão Paraibano
N	26	53	29	64
Média	22,53	18,78	16,90	17,10
Desvio padrão	10,47	12,34	9,11	13,36
Mínimo	6,90	3,37	6,33	0,00
1º Quartil	12,64	10,97	12,58	5,71
Mediana	23,74	14,60	14,78	16,74
3º Quartil	29,10	22,73	19,17	24,47
Máximo	48,78	54,95	43,48	49,18

A Tabela 1 resume an análise descritiva das informações recolhidas, abrangendo a quantidade de casos, médias, desvios padrão, valores mínimos, quartis e máximos para cada grupo analisado. Observa-se que as médias exibidas são bastante similares entre os grupos, e as medidas de dispersão apontam para uma variação comparável. Ademais, a similaridade na distribuição dos dados é reforçada pelos valores dos quartis e da mediana, indicando que não existem diferenças significativas nos perfis das amostras examinadas. Estes achados sugerem uma uniformidade nos dados, facilitando a interpretação dos resultados seguintes com base em grupos comparáveis,

Figura 1: Distribuição da Mortalidade Infantil por Região



Fonte: Fonte elaborado pelos autores (2025).

A Figura 1 apresenta a distribuição da taxa de mortalidade infantil por região do estado da Paraíba no ano de 2022. Observa-se que todas as regiões apresentam ampla dispersão nas taxas, com presença de outliers, especialmente nas regiões Agreste Paraibano e Sertão Paraibano. A região da Borborema apresenta uma mediana mais elevada em comparação com as demais, enquanto a região da Mata Paraibana possui uma mediana inferior e menor dispersão. Para avaliar a normalidade dos dados, foi utilizado o teste de Shapiro-Wilk, cujos resultados indicaram que a variável não segue distribuição normal em nenhuma das regiões analisadas ( $p < 0,05$ ). Diante disso, optou-se pela utilização do teste não paramétrico de Kruskal-Wallis para comparar as medianas das taxas de mortalidade infantil entre as quatro regiões. O resultado do teste indicou ausência de diferença estatisticamente significativa entre os grupos (Valor-p: 0.1801,  $p > 0,05$ ), sugerindo que, apesar das variações visuais observadas no boxplot, as diferenças entre as regiões não são suficientemente robustas do ponto de vista estatístico.

## 5 Conclusão

A partir dos achados do teste não paramétrico de Kruskal-Wallis [1], não se detectou uma diferença estatisticamente relevante entre as medianas das taxas de mortalidade infantil nas áreas da Paraíba (Mata Paraibana, Agreste Paraibano, Borborema e Sertão Paraibano) em 2022. Isso sugere que, mesmo considerando as possíveis disparidades socioeconômicas e estruturais entre as regiões, os índices de mortalidade infantil exibiram um comportamento estatisticamente parecido. Este resultado indica uma relativa uniformidade regional no indicador examinado, o que pode ser um reflexo de políticas públicas de âmbito estadual ou restrições dos dados na capacidade de identificar desigualdades locais. Para pesquisas futuras, sugere-se a avaliação de séries temporais ou a inclusão de variáveis extras que possibilitem a identificação de fatores específicos ligados à mortalidade infantil nos municípios da Paraíba.

## Referências

- [1] KRUSKAL, W. H.; WALLIS, W. A. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, v. 47, n. 260, p. 583–621, 1952.
- [2] SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). *Biometrika*, v. 52, n. 3/4, p. 591–611, 1965.
- [3] INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). *Paraíba: Cidades e Estados*. Disponível em: <https://www.ibge.gov.br/cidades-e-estados/pb.html>. Acesso em: 22 maio 2025.
- [4] REBACK, J. et al. pandas-dev/pandas: Pandas. Zenodo, 2020. Disponível em: <https://doi.org/10.5281/zenodo.3509134>