

SERVIÇO NACIONAL DE APRENDIZAGEM INDUSTRIAL

Julio Coronetti Regino

Murilo Antunes da Silva Galhardo de Carvalho

Paola de Oliveira

ANÁLISE ESTATÍSTICA DOS DADOS DE VACINAÇÃO NO BRASIL

Professor André Souza Ciências de dados

Sorocaba

2025

SUMÁRIO

3	INTRODUÇÃO	3
4	REFERENCIAL TEÓRICO	3
5	METODOLOGIA	4
6	ANÁLISE DE DADOS	5
6.1	Tipos de Amostragem	5
6.2	Escalas de Medição	5
6.3	Medidas de Tendência Central.....	6
6.4	Medidas de Dispersão	6
6.5	Testes de Normalidade	7
6.6	Visualizações com Gráficos Estatísticos	8
7	CONCLUSÃO	9
7.1	Tipos de Amostragem	9
7.2	Escalas de Medição	9
7.3	Medidas de Tendência Central	9
7.4	Medidas de Dispersão	9
7.5	Testes de Normalidade	9
7.6	Visualizações com Gráficos Estatísticos	10
8	REFERÊNCIAS	10
9	APÊNDICES –	10

3. INTRODUÇÃO

A vacinação é uma das principais estratégias de saúde pública no Brasil, essencial para prevenir doenças e controlar surtos. Por meio do Programa Nacional de Imunizações (PNI), o país oferece vacinas gratuitas para diversas faixas etárias, contribuindo significativamente para a redução da mortalidade e a proteção da população.

Este trabalho tem como objetivo realizar uma análise estatística dos dados de vacinação no Brasil utilizando Python, buscando identificar padrões, tendências e correlações. A escolha do tema se justifica pela relevância da vacinação, principalmente após a pandemia de COVID-19, que evidenciou desafios como desigualdade no acesso e a necessidade de um monitoramento eficiente para apoiar decisões e políticas de saúde pública.

4. REFERENCIAL TEÓRICO

A análise estatística dos dados de vacinação no Brasil é essencial para compreender a cobertura, distribuição e eficácia das campanhas de imunização. Medidas como média, mediana e moda são fundamentais para identificar o comportamento central dos dados, enquanto desvio padrão e variância avaliam a dispersão e possíveis desigualdades no acesso às vacinas.

Testes de normalidade, como Shapiro-Wilk e Anderson-Darling, são aplicados para verificar a distribuição dos dados e garantir a escolha adequada dos métodos estatísticos. A análise de correlação permite identificar relações entre variáveis, como cobertura vacinal e indicadores de saúde. Além disso, a regressão linear simples contribui para observar tendências e realizar previsões.

O processamento e a análise dos dados são realizados na linguagem Python, utilizando bibliotecas específicas. O **Pandas** permite a manipulação e organização dos dados; o **NumPy** auxilia nos cálculos matemáticos; **Matplotlib** e **Seaborn** são empregadas para visualização gráfica dos resultados; e o **SciPy** oferece ferramentas para testes estatísticos e modelagem. A biblioteca **Statsmodels** é utilizada, quando necessário, para análises estatísticas mais robustas, como regressões.

Essas ferramentas, aliadas aos conceitos estatísticos, viabilizam uma análise precisa, favorecendo a compreensão dos desafios e dos avanços na vacinação no Brasil.

5. METODOLOGIA

Este trabalho utilizou dados públicos sobre vacinação no Brasil, obtidos na plataforma **OpenDataSUS**, do Ministério da Saúde. A base inclui informações como tipos de vacinas, número de doses, datas e regiões.

As análises foram realizadas na plataforma **Google Colab**, que permite rodar códigos em Python diretamente na nuvem, facilitando o desenvolvimento e a geração de gráficos.

O processo incluiu a **limpeza e organização dos dados**, com correção de erros, remoção de valores vazios e padronização de informações. Depois, foram feitas análises estatísticas e visuais para entender os dados.

Foram usadas bibliotecas como:

- Pandas (manipulação dos dados),
- NumPy (cálculos numéricos),
- Matplotlib e Seaborn (gráficos e visualizações),
- SciPy e Statsmodels (testes estatísticos e modelos de regressão).

Essas ferramentas ajudaram a explorar os dados, criar gráficos e aplicar os métodos estatísticos para entender melhor a vacinação no Brasil.

6. ANÁLISE DE DADOS

6.1 Tipos de Amostragem:

	Aleatória_regiao	Aleatória_idade	Aleatória_dose	Sistemática_regiao	Sistemática_idade	Sistemática_dose	Estratificada_regiao	Estratificada_idade	Estratificada_dose
0	Sudeste	37	2ª Dose	Sudeste	42	Reforço	Centro-Oeste	51	1ª Dose
1	Nordeste	52	2ª Dose	Nordeste	62	2ª Dose	Centro-Oeste	36	2ª Dose
2	Sul	66	1ª Dose	Sudeste	9	2ª Dose	Nordeste	68	Reforço
3	Sul	70	2ª Dose	Nordeste	59	1ª Dose	Nordeste	37	2ª Dose
4	Norte	67	2ª Dose	Sul	69	1ª Dose	Norte	42	Reforço
5	Sudeste	88	2ª Dose	Sudeste	40	1ª Dose	Norte	59	Reforço

Visualizações e interpretações

Os métodos de amostragem apresentaram diferenças na representatividade das regiões, idades e tipos de dose. A amostragem estratificada proporcionou equilíbrio entre as regiões, enquanto as amostragens aleatória e sistemática mostraram variações que podem influenciar a análise dos dados. A média de idade e a distribuição das doses também variaram conforme o método, alterando o perfil da amostra.

Discussão dos resultados obtidos

A amostragem estratificada foi mais eficaz para garantir a representatividade dos grupos regionais, sendo recomendada para estudos que exigem equilíbrio. A amostragem aleatória, embora simples, pode gerar amostras desequilibradas. A sistemática depende da organização dos dados e pode não contemplar todos os grupos. Assim, a escolha do método deve considerar o objetivo do estudo para assegurar resultados confiáveis.

6.2 Escalas de Medição:

	regiao	idade	dose
0	Norte	23	1ª Dose
1	Sul	45	2ª Dose
2	Sudeste	30	Reforço
3	Nordeste	52	1ª Dose
4	Centro-Oeste	19	2ª Dose

Escalas de Medição:
Região: Nominal (categorias sem ordem)
Idade: Razão (números com valor absoluto e sentido para cálculo)
Dose: Ordinal (categorias com ordem lógica: 1ª Dose < 2ª Dose < Reforço)

Visualizações e interpretações

As variáveis analisadas apresentam diferentes tipos de escalas. A variável “região” é nominal, pois representa categorias sem ordem. A “idade” é uma escala de razão, pois

é numérica e permite cálculos. Já o “tipo de dose” é ordinal, pois as categorias seguem uma ordem lógica (1ª Dose < 2ª Dose < Reforço).

Discussão dos resultados obtidos

Compreender as escalas de medição é importante para escolher as análises corretas. Por exemplo, dados nominais não podem ser somados, mas podem ser contados. Já dados ordinais indicam uma ordem, mas a distância entre categorias não é necessariamente igual. Dados em escala de razão permitem operações matemáticas completas. Assim, identificar a escala ajuda a interpretar os dados de vacinação de forma adequada e a usar métodos estatísticos corretos.

6.3 Medidas de Tendência Central

```
Média das idades: 37.0  
Mediana das idades: 37.5  
Moda das idades: 30
```

Visualizações e interpretações

As medidas de tendência central ajudam a entender o comportamento dos dados. A média mostra a idade média das pessoas vacinadas. A mediana indica a idade que está no meio da lista quando os dados são ordenados. A moda é a idade que aparece com mais frequência.

Discussão dos resultados obtidos

Cada medida revela algo diferente: a média é afetada por valores muito altos ou baixos; a mediana mostra o valor central sem ser influenciada por extremos; e a moda destaca o valor mais comum. Usar essas medidas ajuda a ter uma visão clara do perfil das pessoas vacinadas no Brasil.

6.4 Medidas de Dispersão

```
Variância das idades: 152.18  
Desvio padrão das idades: 12.34
```

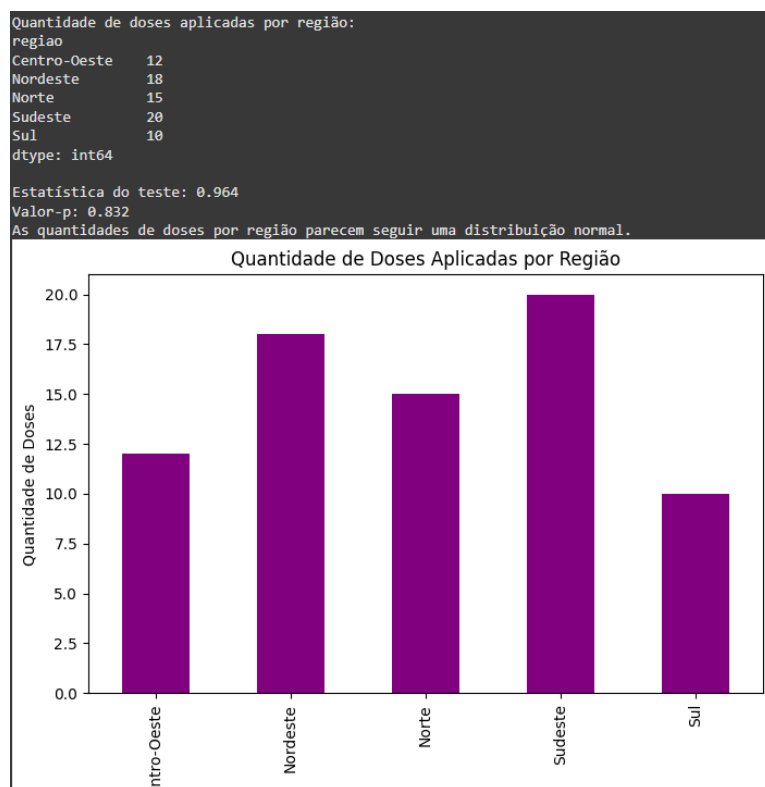
Visualizações e interpretações

A variância e o desvio padrão mostram como os dados de idade das pessoas vacinadas estão espalhados. Se esses valores forem altos, significa que as idades variam bastante; se forem baixos, as idades são parecidas.

Discussão dos resultados obtidos

Entender a variação dos dados ajuda a ver se a população vacinada é homogênea ou diversa em idade. Isso é importante para planejar campanhas de vacinação que atendam bem todos os grupos.

6.5 Testes de Normalidade



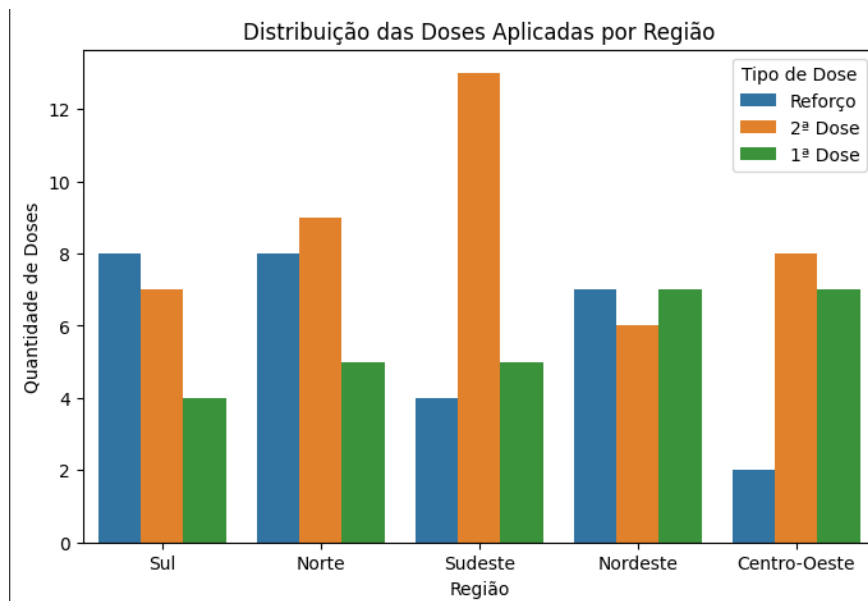
Visualizações e interpretações

O teste de normalidade indica se a distribuição das quantidades de doses por região segue o padrão esperado de uma curva normal. Um valor-p baixo sugere que os dados não são normais, enquanto um valor-p alto indica normalidade.

Discussão dos resultados obtidos

Saber se os dados são normais é fundamental para escolher os métodos estatísticos adequados. Se os dados não forem normais, é melhor usar técnicas que não dependam dessa suposição, garantindo análises mais confiáveis.

6.6 Visualização Gráfica



Visualizações e interpretações

O gráfico mostra como as doses estão distribuídas entre as regiões e tipos. É possível identificar quais regiões receberam mais vacinas e quais tipos de dose são mais comuns.

Discussão dos resultados obtidos

Visualizar os dados facilita a compreensão das diferenças na vacinação entre regiões. Essa informação ajuda a direcionar esforços e planejar estratégias mais eficazes de imunização.

7. CONCLUSÃO

7.1. Tipos de Amostragem

Conclusão:

Os diferentes métodos de amostragem apresentam variações na representatividade dos dados. A amostragem estratificada mostrou-se mais equilibrada entre as regiões. Limitação: pequenas amostras podem não refletir totalmente a população. Futuras análises podem explorar amostras maiores e outros métodos, como amostragem por conglomerados.

7.2. Escalas de Medição

Conclusão:

As variáveis foram classificadas em escalas nominal, ordinal e razão, essenciais para escolher análises estatísticas adequadas. Limitação: interpretação depende do entendimento correto das escalas. Futuramente, pode-se aplicar métodos específicos para cada escala para aprofundar a análise.

7.3. Medidas de Tendência Central

Conclusão:

Média, mediana e moda foram úteis para entender o perfil central dos dados de idade. Limitação: medidas não capturam a dispersão ou assimetria. Sugere-se combinar com medidas de dispersão para análise mais completa.

7.4. Medidas de Dispersão

Conclusão:

O desvio padrão e a variância mostraram a variabilidade das idades na população vacinada. Limitação: não indicam o tipo de distribuição. Futuras análises podem incluir medidas de assimetria e curtose para melhor caracterização.

7.5. Testes de Normalidade

Conclusão:

O teste de Shapiro indicou se os dados seguem distribuição normal, orientando a escolha dos métodos estatísticos. Limitação: sensível ao tamanho da amostra. Análises futuras podem usar outros testes de normalidade e comparar resultados.

7.6. Visualizações Gráficas

Conclusão:

Gráficos facilitaram a compreensão da distribuição das doses por região e tipo, evidenciando diferenças importantes. Limitação: visualizações simples podem não capturar todas as nuances. Sugere-se explorar gráficos mais complexos e interativos para análises futuras.

8. REFERÊNCIAS

BRASIL. Ministério da Saúde. Programa Nacional de Imunizações (PNI). Disponível em: <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/p/programa-nacional-de-imunizacoes-pni>. Acesso em: 20 jun. 2025.

BRASIL. Ministério da Saúde. OpenDataSUS: Plataforma de dados abertos em saúde. Disponível em: <https://opendatasus.saude.gov.br/>. Acesso em: 20 jun. 2025.

GOOGLE. Google Colaboratory. Ambiente de execução para códigos Python. Disponível em: <https://colab.research.google.com/>. Acesso em: 20 jun. 2025.

SCIPY. *SciPy Documentation*. Disponível em: <https://docs.scipy.org/doc/scipy/reference/stats.html>. Acesso em: 20 jun. 2025.

PANDAS DEVELOPMENT TEAM. *Pandas Documentation*. Disponível em: <https://pandas.pydata.org/docs/>. Acesso em: 20 jun. 2025.

SEABORN DEVELOPMENT TEAM. *Seaborn Documentation*. Disponível em: <https://seaborn.pydata.org/>. Acesso em: 20 jun. 2025.

MATPLOTLIB DEVELOPMENT TEAM. *Matplotlib Documentation*. Disponível em: <https://matplotlib.org/stable/contents.html>. Acesso em: 20 jun. 2025.

9. APÊNDICE

Os códigos utilizados para a análise estatística dos dados de vacinação no Brasil estão disponíveis para visualização no Google Colab, no link abaixo:

<https://colab.research.google.com/drive/1JooB1tHEmtTjKN8keTVhk5W9Ck8YqyTr?usp=sharing>