

# Análise Estatística Descritiva de Causas de Morte por Doenças dos Países de 1990 até 2019

Resumo: Este relatório analisa dados sobre causas de morte causadas por doenças entre os anos de 1990 e 2019, abordando tuberculose, violência interpessoal e autoagressão. Os resultados da análise exploratória dos dados fornecem “insights” sobre a prevalência dessas causas de morte e suas tendências ao longo do tempo.

Aluno: Guilherme Rocha Duarte

Curso: Ciência de Dados e Inteligência Artificial

Matéria: Modelagem e Inferência Estatística

Professora: Natália Ribeiro de Souza Evangelista

<b>Introdução</b>	<b>3</b>
<b>População e amostra:</b>	<b>3</b>
<b>Tratamento dos dados:</b>	<b>3</b>
<b>Análise Exploratória dos Dados</b>	<b>4</b>
Óbitos mundiais e Tuberculose	4
Violência Interpessoal e Autoagressão	6
<b>Modelagem e Inferência dos Dados</b>	<b>7</b>
Intervalos de confiança	7
Teste de hipótese	8
Regressão Linear Simples/Múltipla	8
<b>Conclusão</b>	<b>9</b>
<b>Considerações Finais</b>	<b>11</b>
<b>Referências</b>	<b>11</b>

# Introdução

Este relatório visa analisar e desenvolver um projeto utilizando modelagem e inferência estatística. Na primeira parte do trabalho, será feita uma análise exploratória dos dados e visualização. Na segunda parte será feita a modelagem e inferência estatística desses dados. Os dados a serem analisados e explorados são sobre causas de morte dos anos de 1990 até 2019, que foram retirados do website Kaggle.com, de autoria de Willian Oliveira Gibin.

## População e amostra:

Os dados utilizados nesta análise foram obtidos do website Kaggle.com e representam uma amostra global de óbitos ocorridos entre os anos de 1990 e 2019. A população alvo compreende todas as mortes causadas por algum tipo das principais doenças do mundo, registradas nesse período em todas as regiões do mundo.

A amostra foi selecionada com base em critérios que visam garantir representatividade e relevância para o estudo das causas de morte. Foram incluídos dados de óbitos de diferentes faixas etárias, gêneros e localidades geográficas, a fim de abranger uma variedade de cenários e contextos epidemiológicos.

## Tratamento dos dados:

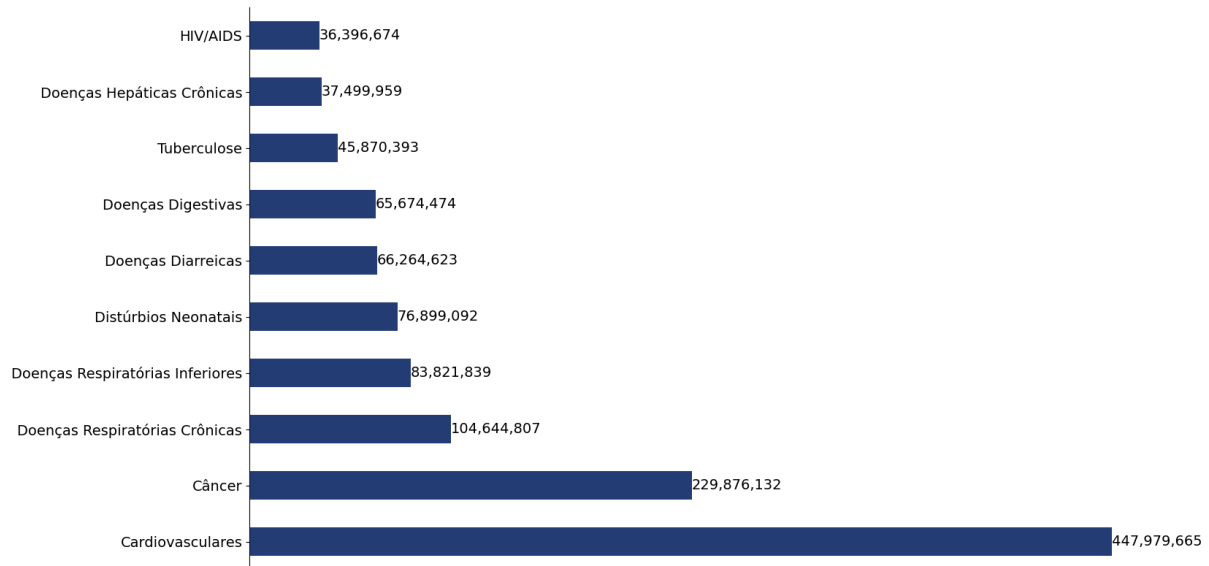
Nesta seção, será descrito como os dados foram preparados para análise, incluindo limpeza, organização e transformações aplicadas, garantindo a qualidade e a consistência dos dados. Quando os dados foram coletados, não havia nenhuma linha vazia (NULL), o que facilitou parte do trabalho, porém havia algumas linhas julgadas não necessárias para a análise (ex: “African Region who” e “OECD Countries”). Esses dados foram retirados da base de dados, pois alteravam os resultados de maneira que não era esperado, já que eram regiões ou blocos econômicos, ou seja, englobam diversos países.

Como os dados foram coletados em inglês, foi necessário traduzir as colunas (formas de mortes) para português. Além disso, duas colunas em específico tiveram seus nomes trocados após a tradução para haver melhor entendimento; as mudanças foram: “Neoplasmas” para “Câncer” e “Conflitos” para “Guerras”. Por ser uma base de dados pequena, o autor conseguiu deixá-la bem limpa e simples, facilitando assim o tratamento dos dados. Não foi necessário nenhum tipo de tratamento de dados faltantes.

Além dessas etapas, foi necessário corrigir os nomes das colunas que continham espaços em branco indesejados. Essa correção garantiu que todas as colunas estivessem devidamente formatadas e prontas para a análise subsequente.

# Análise Exploratória dos Dados

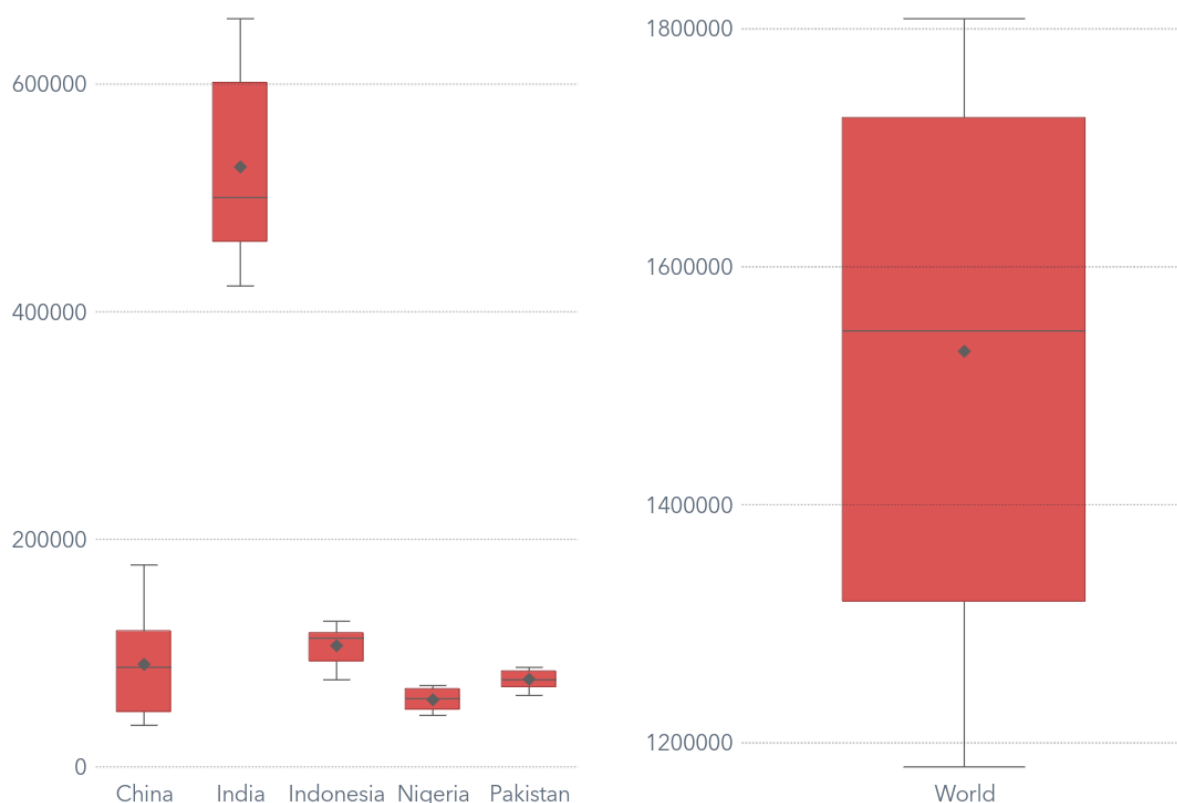
## Óbitos mundiais e Tuberculose



O gráfico anterior mostra as 10 (dez) causas de morte mais comuns no mundo inteiro entre os anos de 1990 e 2019. É possível perceber que doenças foram as maiores causadoras neste período, com as cardiovasculares e o câncer liderando. É evidente a grande diferença entre o número de óbitos a qual as doenças cardiovasculares causam e os outros tipos de enfermidades, é preciso juntar os falecimentos que câncer, doenças respiratórias e HIV/AIDS causam para morrerem o mesmo número de pessoas.

Outro ponto que chama a atenção é que apesar dos antibióticos terem sido descobertos a mais de 80 anos, comorbidades como a tuberculose ainda são uma das principais razões de óbitos no mundo inteiro. Causada pela bactéria *Mycobacterium Tuberculosis*, essa doença afeta principalmente os pulmões, mas também pode afetar outras partes do corpo. Apesar dos avanços na prevenção, diagnóstico e tratamento, a tuberculose continua sendo uma das principais causas de morte por doenças infecciosas em escala global.

Os óbitos relacionados à tuberculose são um reflexo direto das deficiências nos sistemas de saúde, desigualdades socioeconômicas e acesso limitado aos cuidados médicos adequados em muitas partes do mundo. O gráfico a seguir compara os 5 países que mais tiveram óbitos devido à peste branca com o total do mundo.



Analizando os dados, é possível observar uma notável discrepância entre a situação da Índia e a de outros países estudados, assim como em comparação com os números globais da tuberculose ao longo de um período de 29 anos. Apesar de ter uma população maior que a China na época, a Índia enfrentou um volume significativamente maior de casos de tuberculose em relação aos seus vizinhos. Este fenômeno pode ser atribuído a uma série de fatores, incluindo extrema pobreza, deficiências em medidas básicas de higiene e uma densidade populacional desproporcionalmente alta em relação ao território disponível.

Os números revelam uma realidade complexa. A incidência mínima de casos de tuberculose na Índia atingiu 422.634, enquanto a média se estabeleceu em 527.364,06 casos. Em comparação, os dados globais indicam uma incidência mínima de 1.179.766 casos, com uma média de 1.529.013 casos ao longo do período analisado. Esses números ilustram uma situação preocupante, destacando a gravidade do desafio enfrentado pela Índia em relação ao resto do mundo.

A mediana dos casos de tuberculose na Índia ficou em 500.358,5 casos, enquanto a mediana global foi de 1.546.172 casos, sugerindo uma distribuição relativamente uniforme dos dados em ambos os conjuntos. No entanto, o número máximo de casos registrados na Índia chegou a assombrosos 657.515, enquanto globalmente esse número atingiu 1.884.780 casos, evidenciando a escala alarmante do problema em ambos os cenários.

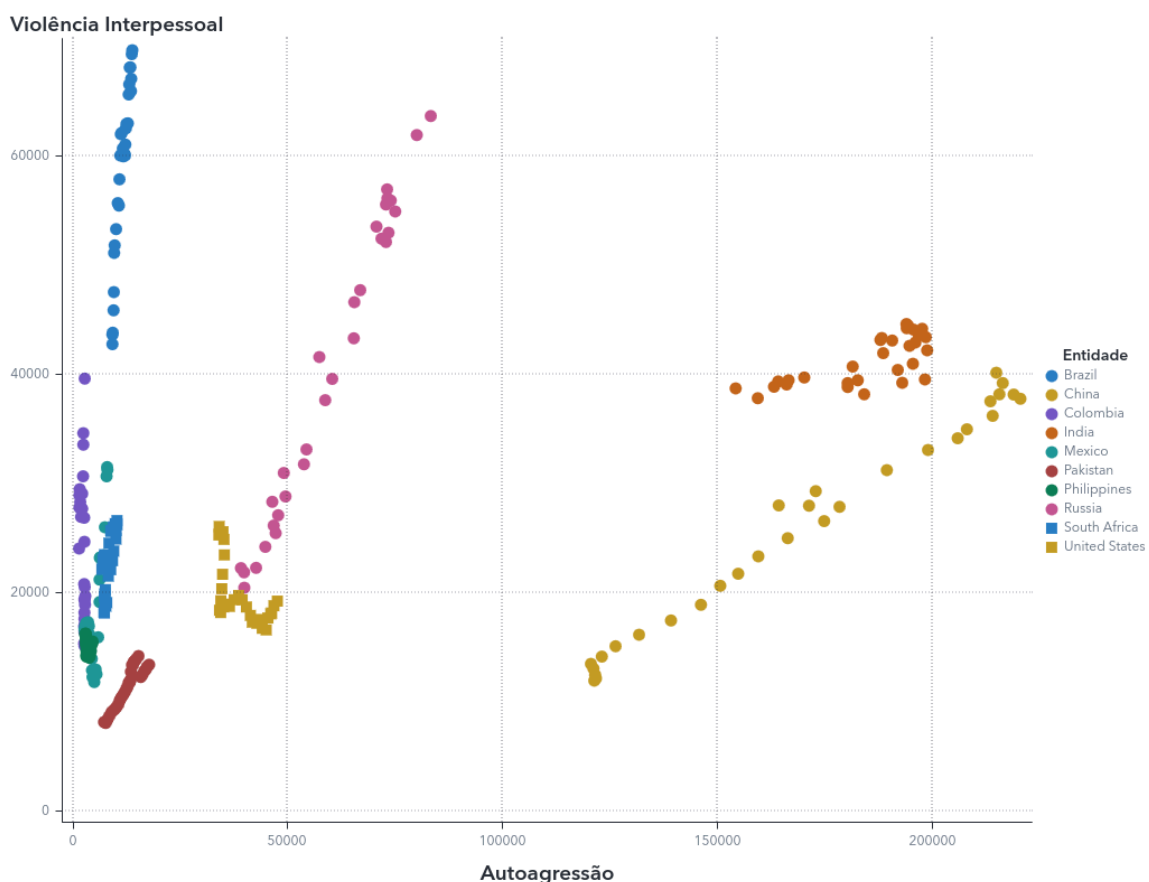
Por fim, o desvio padrão dos casos de tuberculose na Índia foi calculado em 73.298,6, enquanto globalmente foi de 211.228,29. Esses valores ressaltam a variabilidade dos dados em ambos os contextos, sublinhando a complexidade do

desafio que o país e o mundo enfrentam em relação à tuberculose. Estes números apontam não apenas para a urgência de ações eficazes ao nível nacional, mas também para a necessidade de uma abordagem holística e multifacetada em escala global para lidar com essa doença.

## Violência Interpessoal e Autoagressão

A violência interpessoal e a autoagressão são fenômenos complexos que envolvem comportamentos prejudiciais dirigidos a outros indivíduos. A violência interpessoal refere-se a qualquer comportamento intencional que cause dano, sofrimento ou morte a outra pessoa. Pode assumir diversas formas, incluindo violência física, psicológica, sexual e verbal, e ocorre em uma variedade de contextos, como relacionamentos familiares, escolares, comunitários e sociais.

Por outro lado, a autoagressão, também conhecida como automutilação ou autolesão deliberada, envolve comportamentos autodestrutivos nos quais uma pessoa inflige dano a si mesma como uma forma de lidar com emoções intensas, aliviar o sofrimento emocional ou expressar angústia interna. Esses comportamentos podem incluir cortes, queimaduras, arranhões, espancamento da própria cabeça ou corpo, entre outros. O gráfico a seguir mostra a relação entre os dois tipos de violência mencionados.



O Brasil é e sempre foi um dos países com as maiores taxas de violência do mundo e esse gráfico demonstra isso, sendo o país com as maiores taxas de violência interpessoal entre todos os outros. Apesar desses altíssimos números, poucas pessoas falecem devido à autoagressão (por volta de 10000 ao ano), diferentemente da Rússia, que mantém números altos tanto de violência interpessoal quanto de autoagressão.

A China e Índia acabam tendo altos números de suicídios e homicídios por serem países altamente populosos. Além disso, a alta densidade populacional em países como China e Índia pode contribuir para uma variedade de fatores que afetam a saúde mental e a segurança pública. Por exemplo, a competição por recursos limitados, pressões socioeconômicas, superlotação urbana e desigualdades sociais podem aumentar o estresse e a tensão nas comunidades, potencialmente contribuindo para um aumento nos comportamentos autodestrutivos e violentos.

Portanto, embora o tamanho da população seja certamente um fator a ser considerado ao analisar dados de suicídio e homicídio, é igualmente importante examinar as taxas desses comportamentos para entender melhor a extensão do problema e desenvolver intervenções eficazes de prevenção e apoio. Abordar as causas subjacentes desses comportamentos, como questões de saúde mental, desigualdades sociais e acesso a serviços de apoio, é fundamental para reduzir o impacto dessas tragédias em comunidades altamente populosas.

## **Modelagem e Inferência dos Dados**

A modelagem e inferência dos dados neste estudo são cruciais para extrair visões significativas das tendências de mortalidade por doenças ao longo de três décadas e em múltiplos países. Para entender melhor a média de casos de tuberculose no mundo, foram realizadas duas análises estatísticas: o cálculo do intervalo de confiança para a média e a realização de um teste t para verificar se a média de casos de tuberculose difere significativamente de um valor de referência. Além desses dois testes, será feita uma regressão linear simples da variável “Tuberculose”.

### **Intervalos de confiança**

O intervalo de confiança para a média de casos de tuberculose no mundo foi calculado e os resultados são:

(1.450.139,1586; 1.607.887,0414)

Este intervalo de confiança de 95% significa que, com 95% de confiança, a verdadeira média de casos de tuberculose no mundo está entre 1.450.139 e 1.607.887. Esse intervalo oferece uma estimativa razoável da média populacional, fornecendo uma faixa na qual é esperada que a média real esteja.

## Teste de hipótese

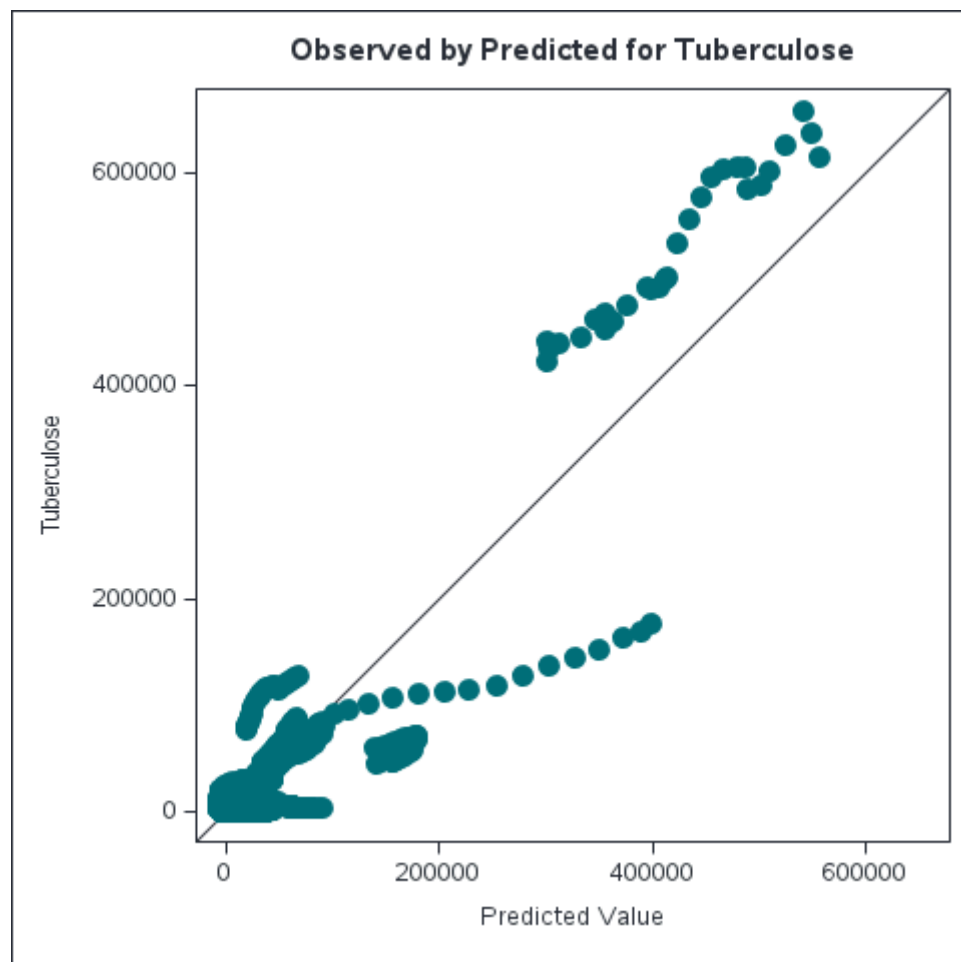
Para determinar se a média observada de casos de tuberculose é significativamente diferente de um valor de referência (0), foi realizado um teste t. Os resultados são:

Estatística t: 39,6479

p-valor: 8,1657e-27

A estatística t de 39,6479 é extremamente alta, indicando uma diferença substancial entre a média observada e o valor de referência. O p-valor de 8,1657e-27 é muito menor que qualquer nível de significância comum (como 0,05), sugerindo ser praticamente impossível que a média observada tenha ocorrido por acaso.

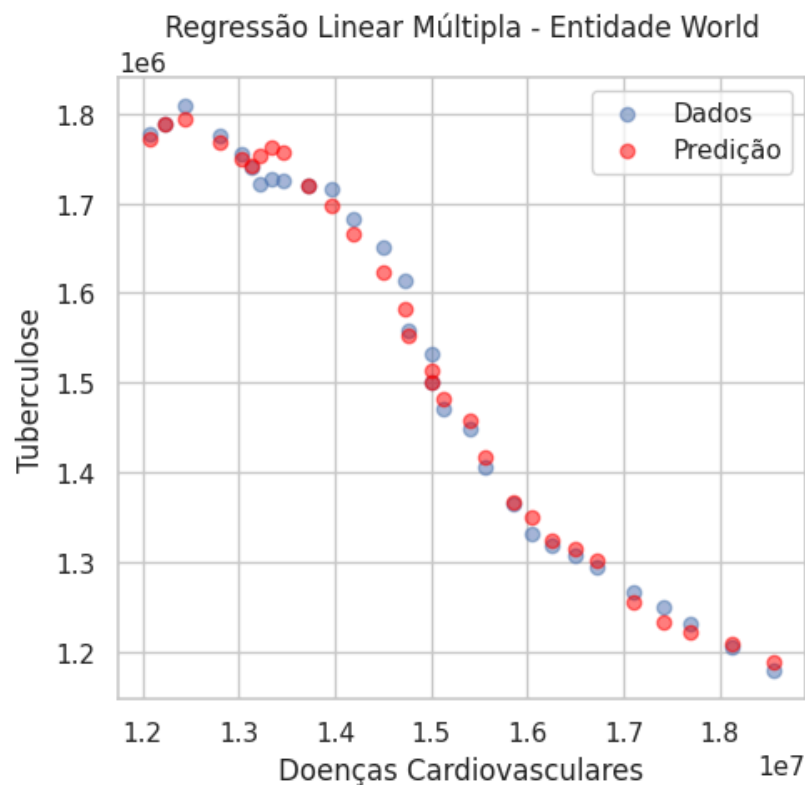
## Regressão Linear Simples/Múltipla



O gráfico comparativo mostra os valores observados de tuberculose no eixo Y contra os valores preditos no eixo X. Idealmente, se o modelo fosse perfeito, todos os pontos estariam alinhados sobre a linha diagonal ( $y = x$ ), indicando uma correspondência exata entre os valores observados e preditos. No entanto, a distribuição dos pontos revela uma disparidade significativa, especialmente em valores preditos elevados.



A maioria dos pontos está concentrada na parte inferior esquerda do gráfico, sugerindo que a maioria dos valores observados e preditos são baixos. Além disso, o gráfico revela que, embora o modelo explique bem uma parte significativa da variabilidade dos dados, há dispersão significativa, especialmente em valores preditos altos.



O modelo de regressão linear múltipla, utilizando doenças cardiovasculares, doenças respiratórias crônicas e doenças respiratórias inferiores como variáveis explicativas, demonstra ser altamente eficaz na predição dos casos de tuberculose, com o alto R-quadrado ajustado (0,993, ou 99,3%) apoiando a eficácia do modelo. O gráfico confirma visualmente a adequação do modelo, com os valores preditos seguindo de perto os dados observados. Além disso, há uma tendência decrescente dos casos de tuberculose à medida que aumentam os casos de doenças cardiovasculares, isso é consistente com o coeficiente negativo para doenças cardiovasculares (-0.2134), indicando uma relação inversa. Porém, a presença de alguma autocorrelação dos resíduos sugere que pode haver padrões nos dados não capturados pelo modelo.

## Conclusão

Este relatório examinou detalhadamente os dados sobre causas de morte entre 1990 e 2019, destacando a importância da análise exploratória e da modelagem estatística para compreender padrões e tendências globais. A investigação focou-se

nas principais causas de mortalidade, como tuberculose, violência interpessoal e autoagressão.

A análise exploratória dos dados revelou visões cruciais sobre as causas de morte mais prevalentes. Doenças cardiovasculares e câncer se destacaram como as principais causas de óbitos no período estudado. Este achado é consistente com a literatura epidemiológica que aponta essas doenças como os maiores desafios de saúde pública globalmente. A análise também ressaltou a persistência da tuberculose como uma causa significativa de morte, mesmo décadas após a introdução de tratamentos eficazes. Este fato evidencia deficiências contínuas nos sistemas de saúde, desigualdades socioeconômicas e o acesso limitado a cuidados médicos em várias regiões.

A discrepância entre os casos de tuberculose na Índia e em outros países, bem como os dados globais, foi um ponto de destaque. A Índia apresentou uma incidência média e um desvio padrão significativamente altos, indicando um desafio substancial em termos de controle e prevenção da tuberculose. Fatores como extrema pobreza, falta de higiene básica e alta densidade populacional foram identificados como contribuintes principais para essa situação.

O estudo também analisou a violência interpessoal e a autoagressão, destacando o Brasil como um dos países com as maiores taxas de violência interpessoal. Este fenômeno é alarmante e sublinha a necessidade de políticas públicas eficazes para combater a violência. Por outro lado, a Rússia se destacou por ter altas taxas tanto de violência interpessoal quanto de autoagressão, sugerindo um complexo panorama de saúde mental e social que requer intervenção multifacetada.

A análise dos dados de China e Índia indicou que, apesar da alta densidade populacional contribuir para números absolutos elevados de suicídios e homicídios, outros fatores, como pressões socioeconômicas e desigualdades sociais, também desempenham um papel crucial. Este ponto reforça a importância de abordar as causas subjacentes dos comportamentos autodestrutivos e violentos via intervenções abrangentes e direcionadas.

Na segunda parte do estudo, a modelagem e inferência estatística foram empregadas para analisar os dados de tuberculose. A utilização de intervalos de confiança e testes de hipótese permitiu conclusões robustas sobre a média global de casos de tuberculose. O intervalo de confiança calculado indicou que, com 95% de confiança, a média verdadeira de casos de tuberculose situa-se entre 1.450.139 e 1.607.887. Esse intervalo fornece uma estimativa confiável da carga da doença ao nível global.

O teste t realizado indicou uma diferença significativa entre a média observada de casos de tuberculose e o valor de referência (0), com uma estatística t extremamente alta (39,6479) e um p-valor praticamente nulo ( $8,1657 \times 10^{-27}$ ). Esses resultados confirmam a gravidade da tuberculose como uma preocupação global de saúde pública, ressaltando a necessidade de ações contínuas e intensificadas para controle e prevenção.

## Considerações Finais

Os resultados deste estudo têm implicações importantes para a saúde pública e a formulação de políticas. A alta mortalidade devido a doenças cardiovasculares e câncer requer investimentos contínuos em prevenção, diagnóstico precoce e tratamento. A persistência da tuberculose como uma causa significativa de morte, especialmente em países com a população geral mais pobre, como a Índia, sublinha a necessidade de fortalecer os sistemas de saúde e abordar as desigualdades socioeconômicas.

Em suma, este relatório oferece uma compreensão abrangente das principais causas de morte globalmente, destacando a importância de abordagens baseadas em evidências para enfrentar esses desafios de saúde pública. A aplicação de análises estatísticas rigorosas e a consideração de contextos socioeconômicos e geográficos são essenciais para o desenvolvimento de intervenções eficazes e sustentáveis.

## Referências

Willian Oliveira Gibin (2024). Annual cause death numbers [.https://www.kaggle.com/datasets/willianoliveiragibin/annual-cause-death-numbers](https://www.kaggle.com/datasets/willianoliveiragibin/annual-cause-death-numbers)

World Bank Country and Lending Groups (2024). <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>

Centro Cultural do Ministério da Saúde (2016). Imagens da Peste Branca: Memória da Tuberculose. <http://www.ccs.saude.gov.br/peste-branca/tb-historia.php>

Depois de sofrer com a Covid-19, Índia está passando por surto de tuberculose (2022). <https://redetb.org.br/depois-de-sofrer-com-a-covid-19-india-esta-passando-por-surto-de-tuberculose/>

Rússia. <https://www.oecdbetterlifeindex.org/pt/paises/russia/>