MINISTÉRIO DA EDUCAÇÃO
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DA BAHIA
***CAMPUS* VALENÇA**
CURSO DE TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

VICTOR NASCIMENTO DOS PASSOS

EVOLUÇÃO DAS DOENÇAS CRÔNICAS: UM ESTUDO DE CIÊNCIAS DE DADOS APLICADO À SAÚDE PÚBLICA

Valença

2025

VICTOR NASCIMENTO DOS PASSOS

EVOLUÇÃO DAS DOENÇAS CRÔNICAS: UM ESTUDO DE CIÊNCIAS DE DADOS APLICADO À SAÚDE PÚBLICA

Trabalho de Conclusão de Curso de graduação em Tecnologia em Análise e Desenvolvimento de Sistemas do Instituto Federal de Educação, Ciência e Tecnologia da Bahia (IFBA), *campus* Valença, como requisito para obtenção do grau de Tecnólogo em Análise e Desenvolvimento de Sistemas.

Orientador(a): André Luiz Romano Madureira

Valença

2025

DEDICATÓRIA

Dedico esse trabalho a minha família e a todos que me apoiaram no meu processo de graduação.

AGRADECIMENTOS

Em primeiro lugar, agradeço a Deus por me dar forças e sabedoria. Em segundo, a minha mãe e meu pai por tudo que fizeram por mim. Agradeço também ao meu orientador, André Romano, por seu cuidado em me orientar e por todo conhecimento que compartilhou comigo, e também a todos os professores do instituto que contribuíram para a minha formação. Por fim, expresso minha gratidão por todos os meus familiares e amigos.

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (BIBLIOTECA)

FOLHA DE APROVAÇÃO

VICTOR NASCIMENTO DOS PASSOS

EVOLUÇÃO DAS DOENÇAS CRÔNICAS: UM ESTUDO DE CIÊNCIAS DE DADOS APLICADO À SAÚDE PÚBLICA

Trabalho de Conclusão de Curso de graduação apresentado como requisito parcial para obtenção do título de Bacharel em Tecnologia em Análise e Desenvolvimento de Sistemas, Instituto Federal de Educação, Ciência e Tecnologia da Bahia (IFBA), *campus* Valença.

Valença, 27 de Agosto de 2025

Banca examinadora

André Luiz Romano Madureira - Orientador(a)

Instituto Federal de Educação, Ciência e Tecnologia - Campus Valença

Hortevan Marrocos Frutuoso - Membro da banca

Instituto Federal de Educação, Ciência e Tecnologia - Campus Valença

Bernardo Peters Menezes Silva - Membro da banca

Instituto Federal de Educação, Ciência e Tecnologia - Campus Valença

LISTA DE FIGURAS

LISTA DE SIGLAS

AVC - Acidente Vascular Cerebral

DRC - Doença Renal Crônica

MAE - Erro Médio Absoluto

MSE - Erro Médio Quadrático

OMS - Organização Mundial de Saúde

RMSE - Raíz do Erro Médio Quadrático

R^2 - Coeficiente de Determinação

WHO - World Health Organization

RESUMO

Este trabalho propõe a aplicação de técnicas de ciência de dados para a análise da mortalidade por doença renal crônica ao longo do tempo, de acordo com a influência de fatores como Diabetes Mellitus, doenças cardiovasculares e o fator do tempo. A partir de uma base de dados histórica contendo causas de morte por país e por ano, foi realizado um processo de tratamento, análise exploratória e implementação de dois modelos estatísticos: regressão linear múltipla e regressão polinomial de segundo grau. Ambos os modelos foram avaliados com base em métricas como R^2, MAE, MSE e RMSE. Os resultados mostraram que a regressão polinomial apresentou o melhor desempenho, com um R^2 de 0,9999 e um erro médio absoluto de apenas 1.645 mortes, enquanto o modelo linear ficou com 0,9975 de R^2 e 11.287 mortes de MAE. A análise dos coeficientes e da estrutura dos modelos apontou a forte correlação entre Diabetes Mellitus e a evolução da doença renal crônica, considerando também as curvaturas nos dados temporais. Concluiu-se que a aplicação de técnicas de ciência de dados, por meio de modelos estatísticos e avaliação preditiva, é uma ferramenta interessante para investigar padrões complexos de dados de saúde pública, contribuindo com evidências quantitativas para a formulação de políticas preventivas e de monitoramento das doenças crônicas.
Palavras-chave: ciência de dados, regressão linear, regressão polinomial, doença renal crônica, Diabetes Mellitus, doenças cardiovasculares.

ABSTRACT

This study proposes the application of data science techniques to analyze chronic kidney disease mortality over time, considering the influence of factors such as Diabetes Mellitus, cardiovascular diseases, and the time variable. Based on a historical dataset containing causes of death by country and year, a process of data cleaning, exploratory analysis, and implementation of two statistical models was carried out: multiple linear regression and second-degree polynomial regression. Both models were evaluated using performance metrics such as R^2, MAE, MSE, and RMSE. The results showed that the polynomial regression model performed best, with an R^2 of 0.9999 and a mean absolute error of only 1,645 deaths, while the linear model yielded an R^2 of 0.9975 and a MAE of 11,287 deaths. The analysis of the coefficients and model structure indicated a strong correlation between Diabetes Mellitus and the progression of chronic kidney disease, also capturing the curvatures present in the temporal data. It is concluded that the application of data science techniques, through statistical modeling and predictive evaluation, is a valuable tool for investigating complex patterns in public health data, providing quantitative evidence to support the formulation of preventive policies and chronic disease monitoring strategies.

Keywords: data science, linear regression, polynomial regression, chronic kidney disease, Diabetes Mellitus, cardiovascular diseases.

Sumário

1. INTRODUÇÃO

A saúde pública é um dos pilares fundamentais da vida em sociedade. Nas últimas décadas, o estudo de padrões de mortalidade tem sido essencial para o campo da saúde global. Infelizmente, o crescimento expressivo de doenças crônicas não transmissíveis, como Diabetes Mellitus, doenças cardiovasculares e doença renal crônica, traz novos desafios para a população. Segundo a Organização Mundial da Saúde (2022), as doenças crônicas já são responsáveis por mais de 70% das mortes no mundo, e sua prevalência continua aumentando, especialmente em países de baixa e média renda, onde o acesso à prevenção e ao tratamento é mais limitado. Essas enfermidades estão interligadas por fatores de risco comuns, como estilo de vida, alimentação inadequada, sedentarismo e envelhecimento populacional.

A doença renal crônica, em especial, tem se tornado uma preocupação crescente no cenário mundial, não apenas por seu impacto direto na qualidade de vida, mas também por sua forte associação com outras doenças crônicas, como Diabetes e doenças cardiovasculares. Diversos estudos já indicaram que a presença de Diabetes aumenta significativamente o risco de comprometimento do sistema renal, criando um ciclo de agravamento entre comorbidades. De acordo com Afkarian et al (2016), cerca de 40% dos adultos com diabetes nos Estados Unidos apresentam alguma forma de doença renal, evidenciando a relação direta entre essas condições. Essa interdependência reforça a necessidade de abordagens integradas na prevenção e controle de doenças crônicas. Além disso, pacientes com doença renal frequentemente desenvolvem doenças cardiovasculares, devido a alterações metabólicas e inflamatórias provocadas pelo mau funcionamento renal.

Neste cenário, a ciência de dados se apresenta como uma ferramenta essencial para analisar grandes volumes de informações de forma eficiente e gerar conhecimento. Através de técnicas como modelagem estatística, análise de regressão e avaliação preditiva, é possível extrair padrões, correlações e tendências a partir de dados históricos, que muitas vezes seriam imperceptíveis por métodos tradicionais.

Diante disso, este trabalho propõe uma análise quantitativa da evolução das mortes por doença renal crônica ao longo do tempo, com base em dados históricos internacionais. Utilizando técnicas de regressão linear múltipla e regressão polinomial, buscou-se modelar a influência de variáveis como o tempo, número de mortes por Diabetes Mellitus e por doenças cardiovasculares na mortalidade renal. A partir das regressões, foram analisadas métricas de desempenho R^2, MAE, MSE e RMSE, com o objetivo de identificar padrões relevantes e validar qual o melhor modelo.

Os resultados indicam que a regressão polinomial obteve desempenho superior em termos de apresentação e erro médio, mostrando que conseguiu se ajustar melhor aos dados reais. Esse tipo de modelo consegue capturar padrões mais complexos de comportamento, como variações não lineares. Este estudo apresenta, portanto, o potencial da ciência de dados como aliada estratégica na investigação e compreensão de grandes problemas de saúde contemporâneos.

1.1 OBJETIVOS

1.1.1 OBJETIVO GERAL

Aplicar técnicas de Ciências de Dados para modelar, analisar e prever a evolução das mortes por doença renal crônica, avaliando a influência de fatores como Diabetes Mellitus, doenças cardiovasculares e o tempo (ano), através de modelos estatísticos de regressão linear e polinomial.

1.1.2 OBJETIVOS ESPECÍFICOS

1. Analisar a correlação entre as doenças crônicas (doença renal crônica, Diabetes Mellitus e doenças cardiovasculares).
2. Desenvolver um modelo de Regressão Linear para medir o impacto das variáveis Diabetes Mellitus, Doenças Cardiovasculares e Ano sobre a mortalidade por Doença Renal Crônica.
3. Desenvolver um modelo de Regressão Polinomial de segundo grau, incorporando interações e termos quadráticos, com o objetivo de aprimorar a capacidade preditiva e explicar padrões não lineares.
4. Comparar o desempenho estatísticos dos modelos, utilizando métricas de ciências de dados como R^2, MAE, MSE e RMSE, identificando qual a método apresenta melhor precisão nas informações.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 CIÊNCIA DE DADOS

A Ciência de Dados é um campo que combina conhecimentos de estatística, matemática, e computação para extrair informações relevantes a partir de grandes volumes de dados. Seu objetivo principal é transformar dados brutos em conhecimento aplicável, auxiliando na tomada de decisões estratégicas e na identificação de padrões que seriam imperceptíveis por métodos convencionais (Fadillah et al., 2024).

Com o crescimento do volume e da complexidade dos dados disponíveis nas mais diversas áreas, a ciência de dados tornou-se essencial para setores como finanças, indústria, educação e saúde pública. Por meio de técnicas de análise exploratória, modelagem preditiva e *machine learning*, é possível entender comportamentos históricos e até realizar projeções futuras com base em tendências observadas.

Os passos comuns de um projeto de ciência de dados seguem uma série de etapas bem definidas, que inclui:

1. Coleta de dados
2. Limpeza e pré-processamento
3. Análise exploratória
4. Modelagem estatística ou preditiva
5. Avaliação de desempenho
6. Comunicação dos resultados

Neste trabalho, a ciência de dados foi aplicada como a ferramenta analítica central para investigar e modelar a evolução de doenças crônicas ao longo dos anos, utilizando análise de correlação e modelos estatísticos de regressão, com o objetivo de gerar insights quantitativos e contribuir com o entendimento da progressão de causas de morte associadas a comorbidades como diabetes, doenças cardiovasculares e doença renal crônica.

2.2 REGRESSÃO LINEAR E POLINOMIAL

A regressão é uma técnica estatística fundamental em projetos de Ciências de Dados, utilizada para modelar a relação entre variáveis independentes (explicativas) e uma variável dependente (prevista). No contexto de dados de saúde, ela permite analisar tendências, detectar associações entre causas e efeitos, e fazer previsões com base em padrões históricos.

A Regressão Linear Simples, considera apenas uma variável independente, assumindo uma relação linear com a variável dependente. Já a Regressão Linear Múltipla amplia a quantidade de variáveis independentes para duas ou mais, possibilitando representar

comportamentos mais complexos, como o impacto combinado do tempo e de comorbidades crônicas sobre taxas de mortalidade (Trunfio et al., 2022).

Modelos não lineares, como a regressão polinomial, são úteis para representar relações complexas entre variáveis, especialmente quando há curvaturas nos dados que não podem ser captadas por modelos lineares simples (Mattos, 2013). Portanto, a regressão polinomial é uma extensão da linear, que permite ajustar modelos quando a relação entre a variável dependente e a(s) independente(s) não é constante.

A escolha entre o modelo linear e polinomial depende da natureza dos dados e do comportamento da variável alvo. Neste projeto, ambos modelos foram aplicados para estimar a evolução da Doença Renal Crônica com base em variáveis como Ano, Diabetes Mellitus e Doenças Cardiovasculares. A comparação de desempenho entre os modelos, por meio de métricas como R^2, MAE e RMSE, permitiu avaliar qual abordagem melhor representava a realidade dos dados.

2.3 DOENÇA RENAL CRÔNICA

A Doença Renal Crônica (DRC) é caracterizada pela perda progressiva da função renal, que é quando os rins deixam de filtrar adequadamente o sangue, levando ao acúmulo de resíduos tóxicos no organismo. Segundo o National Kidney Foundation (NKF, 2020), a enfermidade é classificada em cinco estágios com base na taxa de filtração glomerular (TFG). Nos estágios mais avançados, os rins perdem o controle do meio interno, tornando-se alterado e incompatível com a vida, levando o paciente a buscar um nefrologista, pois é necessário medicamentos adicionais para substituir a perda de certas funções renais, e também o planejamento para um transplante renal, diálise ou cuidados de suporte.

A doença é frequentemente imperceptível nos estágios iniciais, o que dificulta o diagnóstico. E quando finalmente aparece, a função renal já costuma estar comprometida de forma significativa. Os principais fatores de risco incluem:

- Diabetes Mellitus
- Hipertensão Arterial
- Doenças cardiovasculares
- Idade avançada
- Histórico familiar

Estudos mostram que a DRC está intimamente relacionada a outras doenças crônicas. Segundo a National Kidney Foundation (2020), cerca de 40% dos pacientes diabéticos desenvolvem algum grau de comprometimento renal. Além disso, indivíduos com DRC têm

risco elevado de eventos cardiovasculares fatais, como infarto agudo do miocárdio e acidente vascular cerebral, mesmo antes de atingirem o estágio terminal da doença (Go et al., 2004).

2.4 DIABETES MELLITUS

Segundo Brutsaert (2025), o diabetes mellitus é uma doença caracterizada pela produção insuficiente de insulina ou pela resistência do organismo à sua ação, o que leva ao acúmulo excessivo de glicose no sangue (hiperglicemia), podendo causar complicações graves ao longo do tempo.

Com o passar dos anos, a hiperglicemia crônica afeta diversos sistemas do organismo, especialmente os sistemas cardiovascular, renal e nervoso. O Diabetes é reconhecido como um dos principais fatores de risco para Doença Renal Crônica (DRC), uma vez que níveis altos de glicose danificam os vasos sanguíneos dos rins. Segundo Afkarian et al (2016), cerca de 40% dos adultos com diabetes apresentam algum grau de comprometimento renal.

O Diabetes Mellitus também está fortemente associado a doenças cardiovasculares, como infarto agudo do miocárdio e acidente vascular cerebral. Beckman, Creager e Libby (2002) demonstraram que pacientes diabéticos têm risco duas a quatro vezes maior de desenvolver doenças cardiovasculares em comparação com indivíduos não diabéticos.

Neste trabalho, o Diabetes Mellitus é tratado como variável explicativa central, dada sua comprovada relação causal com a Doença Renal Crônica e sua contribuição significativa para o aumento das taxas de mortalidade global.

2.5 DOENÇAS CARDIOVASCULARES

As doenças cardiovasculares constituem um grupo de distúrbios do coração e dos vasos sanguíneos, incluindo hipertensão arterial, insuficiência cardíaca, acidente vascular cerebral (AVC), entre outras. Essas condições representam uma das principais causas de morte no mundo, especialmente em países com envelhecimento populacional e hábitos de vida não saudáveis. Segundo a Organização Mundial da Saúde (OMS), estima-se que cerca de 17,9 milhões de pessoas morrem todos os anos por causas relacionadas a doenças cardiovasculares, representando aproximadamente 32% de todas as mortes globais (WHO, 2023). Fatores como sedentarismo, tabagismo, consumo excessivo de sal e gorduras, estresse crônico, obesidade e histórico familiar contribuem diretamente para o surgimento e agravamento dessas enfermidades.

As doenças cardiovasculares também estão fortemente relacionadas a outras doenças crônicas, como o diabetes mellitus e a doença renal crônica. Estudos indicam que pessoas com diabetes possuem risco elevado de desenvolver aterosclerose e outras complicações

cardiovasculares, uma vez que a hiperglicemia crônica danifica os vasos sanguíneos e o coração (Beckman; Creager; Libby, 2002). Além disso, indivíduos com doença renal crônica frequentemente desenvolvem hipertensão arterial e disfunções cardiovasculares, o que cria um ciclo de agravamento mútuo entre essas comorbidades.

2.6 CORRELAÇÃO E CAUSALIDADE ENTRE DOENÇAS

No campo da análise de dados, é essencial compreender a diferença entre os conceitos de correlação e causalidade. A correlação representa uma associação estatística entre duas variáveis — ou seja, quando uma varia, a outra tende a variar também, positiva ou negativamente. Já a causalidade implica que uma variável influencia diretamente o comportamento da outra, ou seja, existe uma relação de causa e efeito. Segundo Cupani e Tesser (2021), “a correlação representa uma associação entre variáveis, mas não implica necessariamente uma relação de causa e efeito. A causalidade exige evidências adicionais que sustentem uma relação direta entre os fenômenos observados”.

Na área da saúde pública, muitas doenças crônicas apresentam correlações elevadas entre si devido a fatores de risco compartilhados, como estilo de vida, predisposição genética e envelhecimento. No entanto, há casos em que a causalidade é bem estabelecida, com base em evidências clínicas e estudos científicos. Por exemplo, o diabetes mellitus tipo 2 é um fator de risco amplamente documentado para o desenvolvimento de doença renal crônica (DRC). De acordo com Afkarian et al (2016), cerca de 40% dos pacientes com diabetes nos Estados Unidos apresentam algum grau de disfunção renal, confirmando a existência de uma relação causal entre essas duas condições. Isso se deve ao fato de que níveis elevados de glicose no sangue danificam progressivamente os vasos sanguíneos dos rins, comprometendo sua função ao longo do tempo.

Outra relação causal importante ocorre entre a diabetes e as doenças cardiovasculares. Beckman, Creager e Libby (2002) mostraram que o diabetes acelera o processo de aterosclerose, aumentando o risco de infarto e acidente vascular cerebral. Pacientes diabéticos têm de duas a quatro vezes mais chances de morrer por causas cardiovasculares do que indivíduos não diabéticos.

A Doença Renal Crônica também compartilha uma via causal bidirecional com as doenças cardiovasculares. Pessoas com DRC têm maior risco de desenvolver hipertensão e insuficiência cardíaca, enquanto indivíduos com doenças cardíacas podem sofrer redução da perfusão renal, agravando o quadro renal (Sarnak et al., 2003).

Na presente pesquisa, as análises de regressão foram utilizadas como ferramentas para investigar essas relações e medir o impacto estatístico entre as variáveis. Embora a regressão não seja suficiente para afirmar a causalidade por si só, quando associada à literatura científica, ela se torna um recurso valioso para validar hipóteses e apoiar decisões baseadas em dados.

3. METODOLOGIA

Este trabalho foi desenvolvido com base em uma abordagem quantitativa, utilizando técnicas de análise estatística e modelagem preditiva para compreender os fatores associados ao crescimento das mortes por Doença Renal Crônica ao longo do tempo. O ambiente de análise escolhido foi o Anaconda Python, uma plataforma muito utilizada em Ciência de Dados e aprendizado de máquina, por ter uma instalação simplificada e possuir mais de 250 pacotes populares na área de dados, como NumPy, Pandas e Matplotlib, além de entregar o ambiente do Jupyter Notebook.

A metodologia foi dividida em quatro etapas principais: coleta e processamento de dados, análises exploratórias e correlações, modelagem estatística (regressão linear e polinomial) e avaliação de desempenho. Essas etapas serão descritas em detalhe nas seções a seguir.

3.1 COLETA E PROCESSAMENTO DE DADOS

A base de dados utilizada neste trabalho foi obtida por meio da plataforma Kaggle, reconhecida por disponibilizar bases de dados públicas voltadas à Ciência de Dados. Para garantir a confiabilidade da base utilizada neste trabalho, foram considerados critérios internos da própria plataforma, como o sistema de selos (bronze, prata e ouro) e a nota de usabilidade atribuída pela comunidade de usuários. A base escolhida, *cause_of_deaths* [1], apresenta selo Ouro e pontuação máxima de 10.0, indicando alto nível de organização, confiabilidade e aplicabilidade, o que favoreceu a qualidade das análises. A base contém informações parciais sobre causas de morte ao redor do mundo, organizadas por ano e por país, abrangendo o período de 1990 a 2019. Cada linha representa os dados de um país em um determinado ano, e as colunas correspondem a diversas causas específicas de mortalidade, totalizando mais de 30 causas distintas.

3.2 ANÁLISES EXPLORATÓRIAS E CORRELAÇÕES

Nesta etapa inicial, foi realizada uma análise exploratória dos dados com o objetivo de compreender a estrutura, distribuição e padrões presentes no conjunto de dados. Essa etapa é essencial em projetos de ciência de dados, pois é ela que dá o direcionamento para qual o projeto deve seguir.

[1] https://www.kaggle.com/datasets/iamsouravbanerjee/cause-of-deaths-around-the-world

Foram desenvolvidas diversas visualizações gráficas com auxílio das bibliotecas Pandas, Matplotlib e Seaborn no ambiente Jupyter Notebook, incluindo:

- Matriz de Correlação de Pearson com Mapa de Calor (Heatmap): Utilizada para identificar possíveis correlações lineares entre as causas de morte. As cores foram configuradas com gradientes que facilitam a visualização de relações positivas (em tons de azul) e negativas (em tons de vermelho), variando em intensidade de acordo com o valor da correlação (de -1 a +1).
- Gráficos de Linha Temporais: Aplicados para investigar o comportamento das causas de morte por doenças crônicas ao longo dos anos (1990 a 2019), observando tendências de crescimento ou queda.
- Gráficos de Barras Agrupadas com Eixos Y Duplos: Desenvolvidos para comparar variáveis com escalas diferentes, como Diabetes Mellitus, Doenças Cardiovasculares e Doença Renal Crônica. A utilização de dois eixos verticais possibilitou a visualização simultânea das séries, sem distorções causadas por diferenças de magnitude.

3.3 MODELAGEM ESTATÍSTICA (REGRESSÃO LINEAR E POLINOMIAL)

A etapa de modelagem estatística teve como objetivo investigar o impacto das variáveis independentes (Ano, Diabetes Mellitus e Doenças Cardiovasculares) sobre a variável dependente (Doença Renal Crônica). Foram desenvolvidos dois modelos principais:

- Regressão Linear Simples e Múltipla: aplicada para verificar a relação linear entre as variáveis. A regressão múltipla considerou o efeito conjunto das variáveis Ano, Diabetes Mellitus e Doenças Cardiovasculares sobre a Doença Renal Crônica. O modelo foi treinado utilizando a biblioteca Linear Regression da linguagem Python, e seus resultados foram avaliados com base nas métricas estatísticas:

  R^2 (coeficiente de determinação)

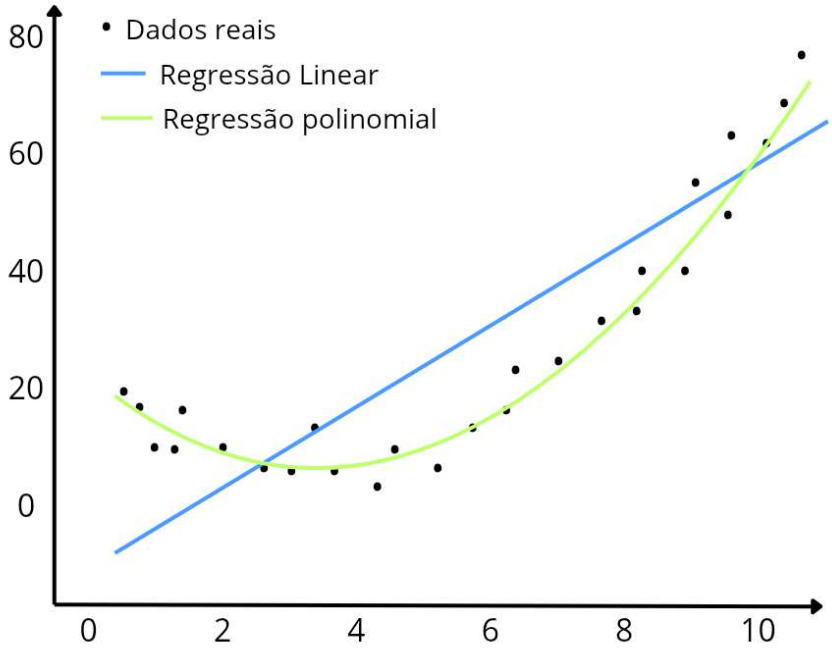
  MAE (Erro Médio Absoluto)

  MSE (Erro Médio Quadrático)

  RMSE (Raíz do Erro Médio Quadrático)
- Regressão Polinomial (grau 2): utilizado para capturar padrões e interações não lineares entre variáveis, incluindo termos quadráticos, como Ano^2 e $Diabetes^2$, e interações como Ano × Diabetes A regressão polinomial foi avaliada com as mesmas métricas e posteriormente comparada com o modelo linear, a fim de

verificar se o aumento da complexidade do modelo gerava melhor desempenho preditivo.

Figura 1: Representação das regressões



Fonte: elaborado pelo autor

3.4 AVALIAÇÃO DE DESEMPENHO

A avaliação de desempenho foi realizada com base em critérios estatísticos que quantificam o grau de ajuste dos modelos aos dados observados. As métricas consideradas foram:

- R^2 (Coeficiente de Determinação) – indica a proporção da variância explicada pelo modelo.
- MAE (Erro Médio Absoluto) – média dos erros absolutos entre os valores previstos e reais.
- MSE (Erro Médio Quadrático) – média dos quadrados dos erros.
- RMSE (Raiz do Erro Médio Quadrático) – raiz quadrada do MSE, mantendo as unidades originais da variável.
- Estatística F e valor-p – utilizados para avaliar a significância estatística do modelo como um todo).

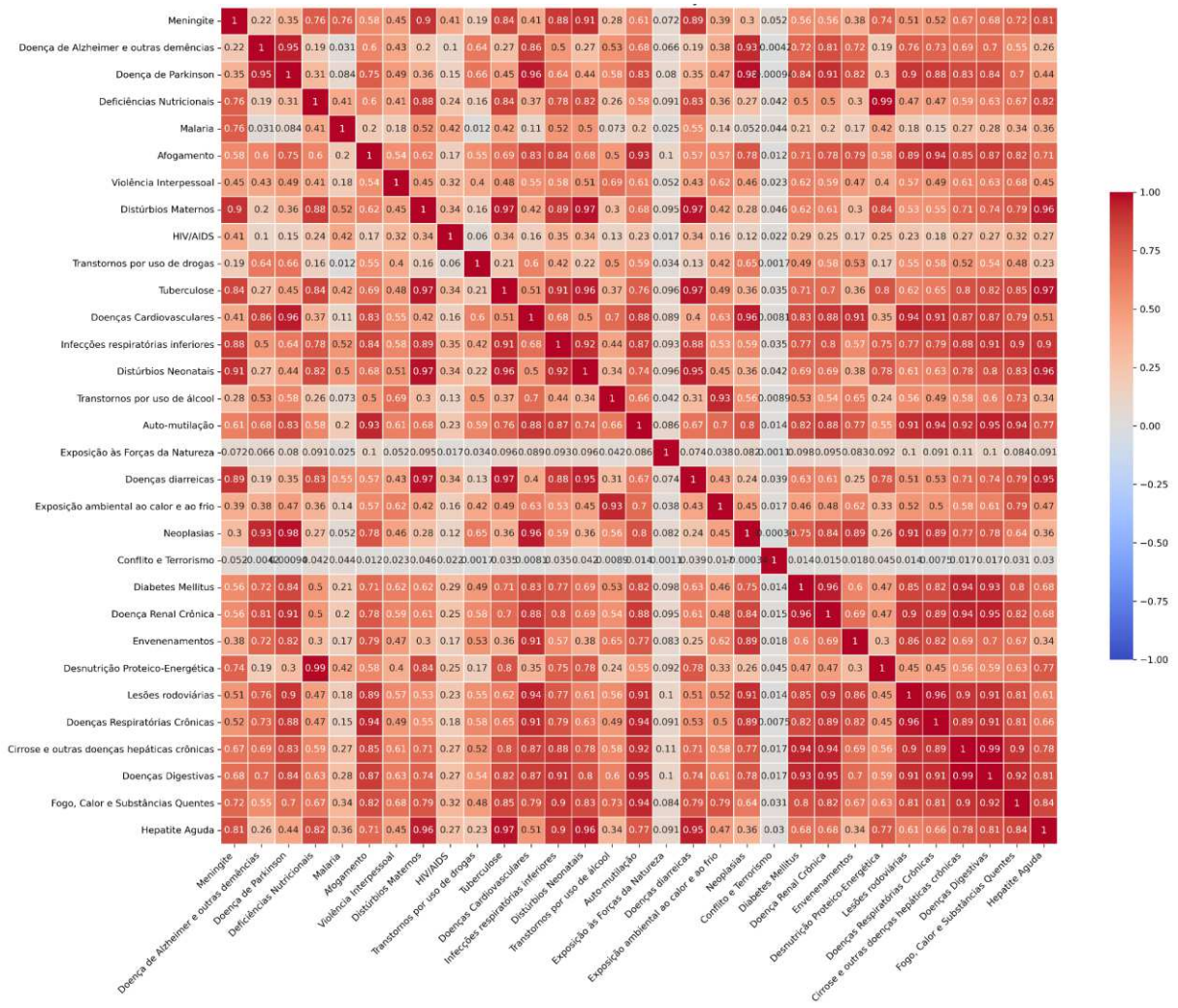
Além disso, a equação final do melhor modelo foi interpretada matematicamente, e gráficos foram utilizados para ilustrar a superfície de regressão e as relações identificadas entre as variáveis.

4. RESULTADOS DA PESQUISA

4.1 ANÁLISE EXPLORATÓRIA

Com o Anaconda Python instalado e pronto para uso, importou-se as bibliotecas fundamentais para o estudo: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn e statsmodels.api. Carregando o arquivo cause_of_deaths.csv, foi realizada inicialmente uma análise exploratória, com a biblioteca Pandas, para confirmar se os tipos de dados estavam de fato corretos e se não havia valores nulos. Em seguida, utilizando as bibliotecas Matplotlib e Seaborn, criou-se uma matriz de correlação com *heatmap* (mapa de calor) para verificar a correlação entre as variáveis, especificamente as utilizadas no projeto. Confira na Figura 2 a seguir.

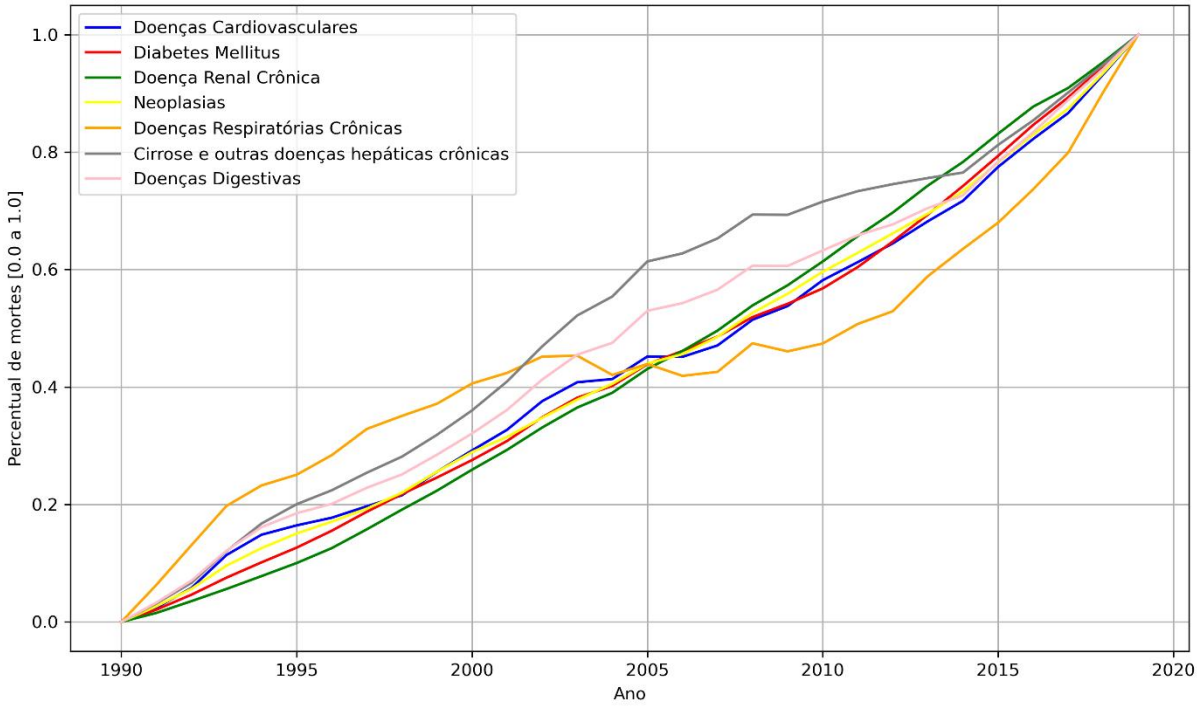
Figura 2: Matriz de correlação com mapa de calor



Fonte: elaborado pelo autor

A análise inicial revelou que diversas variáveis possuem fortes correlações entre si, mostrando como a Ciência de Dados pode ser uma ferramenta poderosa na detecção de padrões em grandes volumes de dados. Esta abordagem possibilitou uma compreensão mais ampla sobre a relação entre diferentes causas de morte e seus comportamentos ao longo do tempo. Observa-se a forte correlação entre Doença Renal Crônica e Diabetes (0,96), Doenças Cardiovasculares e Doença Renal Crônica (0,88) e por fim Doença Cardiovascular e Diabetes (0,83). Além da matriz de correlação, foi desenvolvido um gráfico de linha com base nas séries temporais das doenças crônicas não transmissíveis, o que permitiu visualizar com clareza o avanço anual dessas enfermidades. Esses gráficos iniciais (matriz de correlação e gráfico de linhas) serviram como base para as próximas etapas do projeto, reforçando como a análise exploratória é uma etapa fundamental para projetos de Ciência de Dados em geral. Confira o gráfico de linhas do avanço das causas de morte por doenças crônicas ao longo dos anos.

Figura 3: Mortes Globais por Ano - Doenças crônicas.
Escala de 0,0 a 1,0: valores normalizados, onde 0,0 representa o menor valor absoluto ao longo da série temporal, e 1,0 representa o maior valor.



Fonte: elaborado pelo autor

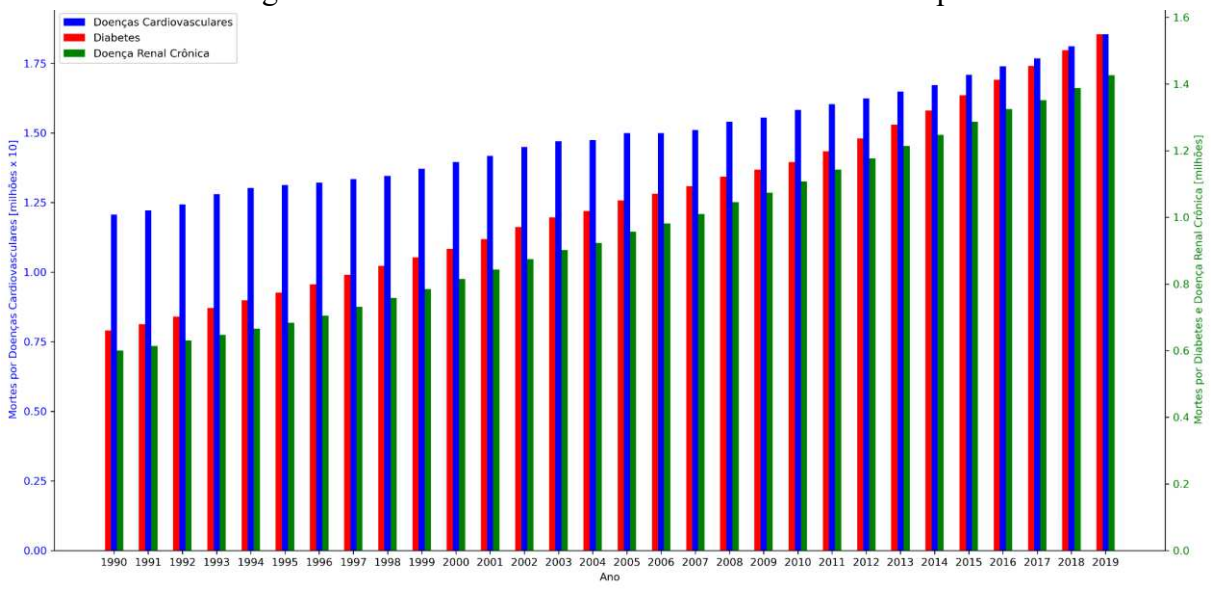
Observa-se uma crescente evolução nas taxas de mortalidade por doenças crônicas ao longo dos anos. Dentre elas, destacam-se a Doença Renal Crônica, o Diabete Mellitus e as Doenças Cardiovasculares, que foram selecionadas como foco deste projeto. A escolha dessas

enfermidades se dá pelo fato de estarem interligadas historicamente, formando o que diversos estudos clínicos denominam como um “eixo-renal-metabólico”, no qual a presença de uma dessas condições frequentemente contribui para o desenvolvimento das outras. Segundo Ronco et al. (2008), “a interação entre insuficiência cardíaca, doença renal e distúrbios metabólicos, como a diabete, caracteriza uma síndrome clínica interdependente, na qual a falha de um sistema acarreta sobrecarga e disfunção dos demais”. Isso evidencia não apenas a correlação entre essas causas de morte, mas também a complexidade do tratamento e prevenção quando afetam múltiplos sistemas orgânicos simultaneamente.

Para verificar os gráficos dos grupos de outras causas de mortes, como por doenças infecciosas, causas maternas e neonatais, lesões e causas externas, uso de substâncias, dentre outras, o leitor pode conferir os códigos do trabalho no repositório do projeto no GitHub [1]

Em seguida, para visualizar separadamente o comportamento das doenças crônicas escolhidas, foi desenvolvido um gráfico de colunas agrupadas, como visto na imagem a seguir.

Figura 4: Causalidade Anual de Mortes com eixo Y duplo



Fonte: elaborado pelo autor

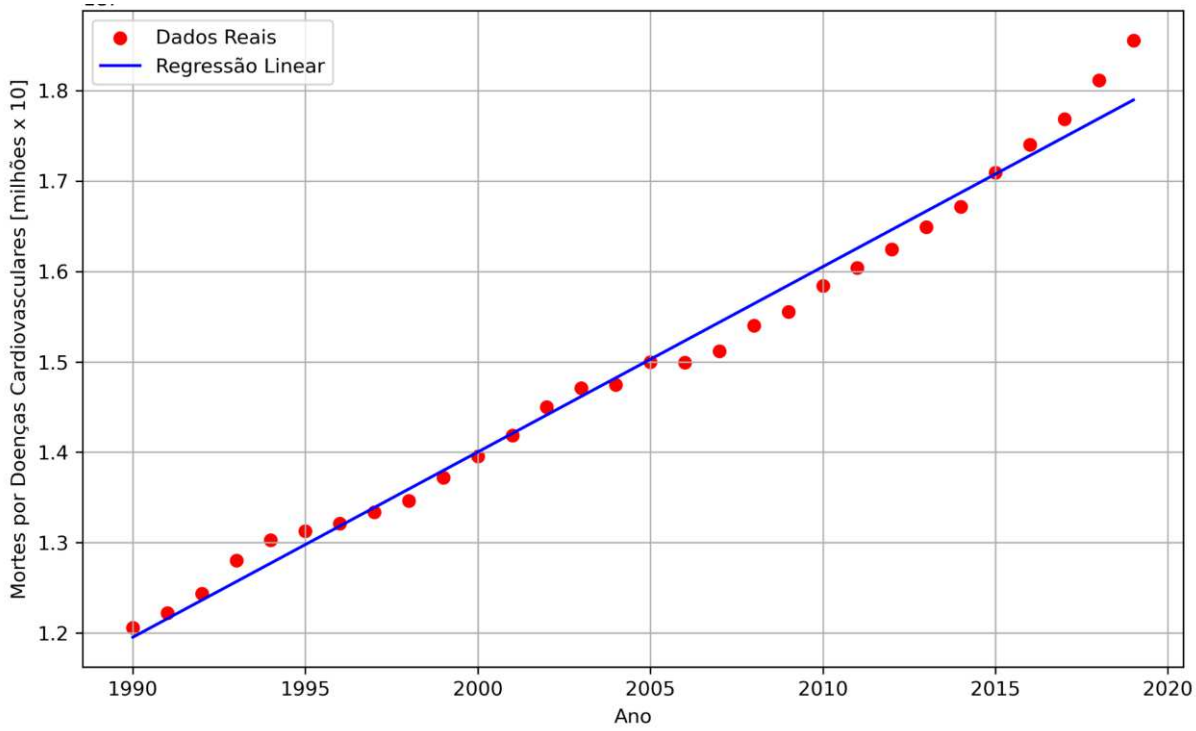
Assim como anteriormente, o gráfico de colunas agrupadas com eixos Y duplo também mostra com clareza a tendência crescente das mortes por doenças crônicas ao longo dos anos. Observa-se durante todo o período analisado que as Doenças Cardiovasculares (eixo Y à esquerda) lideram expressivamente o número de óbitos durante praticamente todo o período analisado. A Diabetes Mellitus apresenta um crescimento constante e acelerado (conforme observado no eixo Y à direita), é esperado que em algumas décadas ela chegue a números de

óbitos parecidos com os das doenças cardiovasculares, ou até mesmo ultrapasse. Já a Doença Renal Crônica, apesar de apresentar números absolutos menores, também revela uma tendência crescente (eixo Y à direita). O uso do eixo Y duplo foi fundamental para melhorar a visualização, permitindo comparar variáveis com escalas distintas em um mesmo gráfico, funcionando como uma forma de normalização visual.

4.2 MODELO DE REGRESSÃO LINEAR SIMPLES E MÚLTIPLA

Antes de partir para modelos de regressão mais complexos e para um dos objetivos específicos do projeto — descobrir o impacto das variáveis Diabetes Mellitus, Doenças Cardiovasculares e ano sobre as mortes por Doença Renal Crônica —, foi feito um modelo de regressão linear simples, a fim de visualizar o comportamento do número de mortes por Doenças Cardiovasculares ao longo dos Anos. Este modelo possui apenas duas variáveis, chamadas de dependente e independente, onde a primeira é a que será prevista (Doenças Cardiovasculares) e a segunda a explicativa para o comportamento da primeira (Anos).

Figura 5: Regressão linear: Doenças cardiovasculares ao longo dos anos

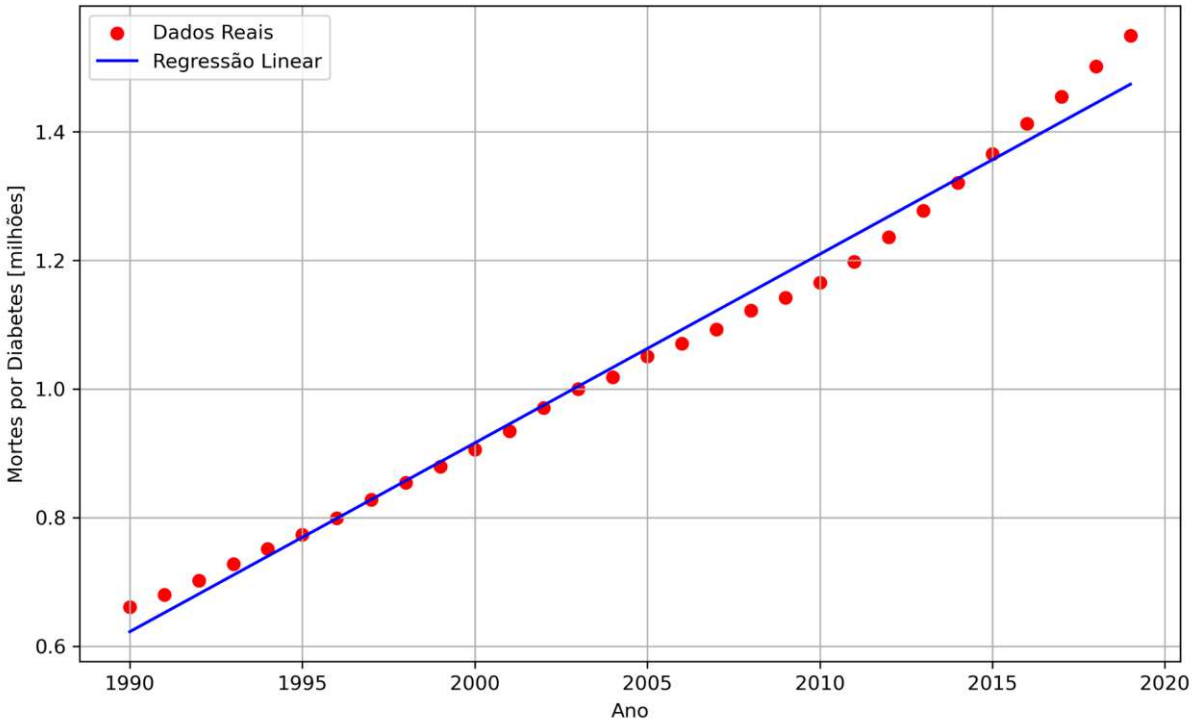


Fonte: elaborado pelo autor

O gráfico apresenta uma tendência de crescimento no número de mortes por Doenças Cardiovasculares ao longo dos anos (1990 a 2019). A linha azul representa o modelo de regressão linear ajustado aos dados históricos (pontos vermelhos), permitindo visualizar a

tendência central da evolução da mortalidade ao longo do tempo. O mesmo comportamento ocorre quando se apresenta o modelo para a Diabetes Mellitus.

Figura 6: Regressão linear: Diabetes Mellitus ao longo dos anos



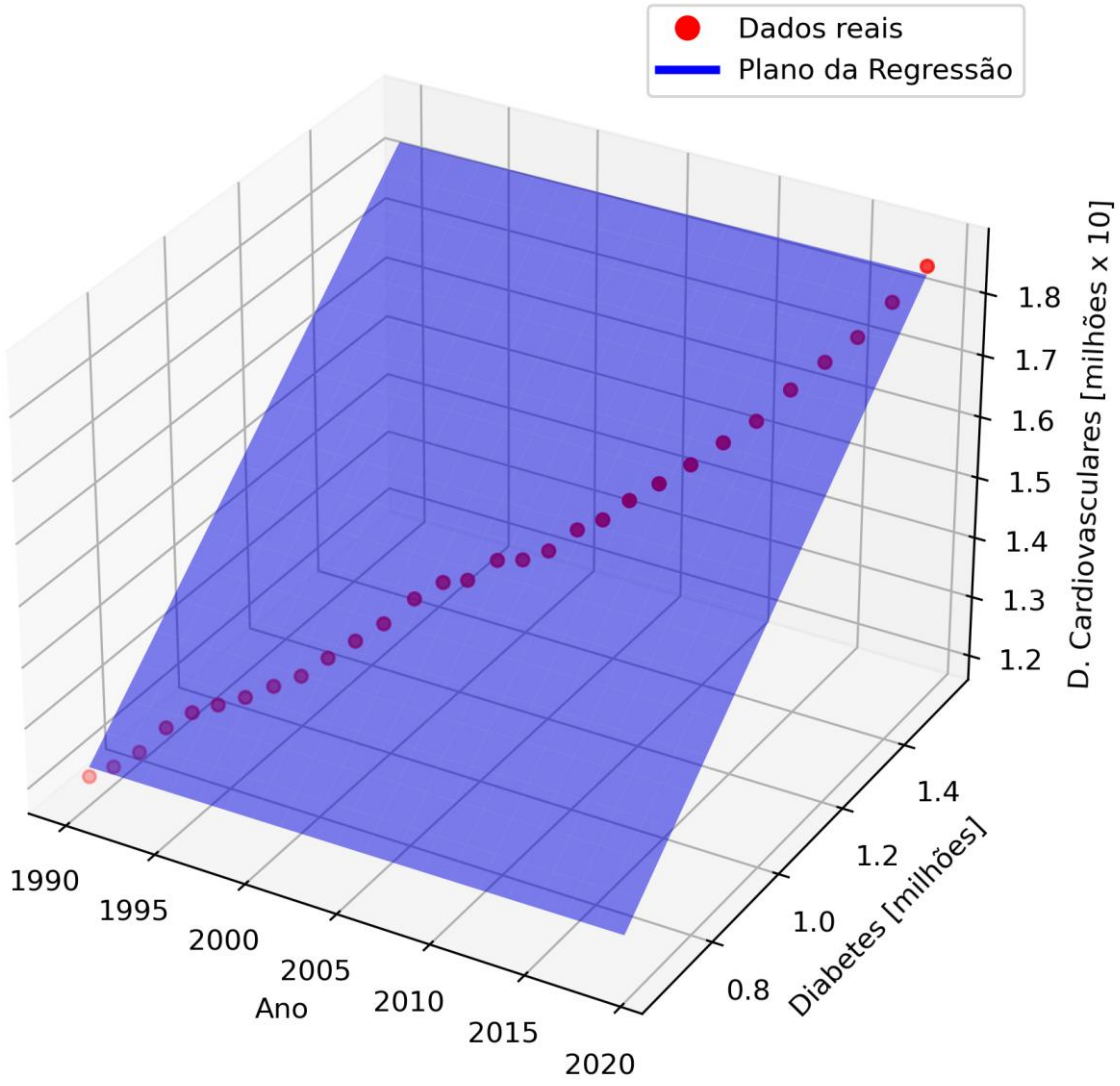
Fonte: elaborado pelo autor

Avançando nas análises, foi utilizada a regressão linear múltipla, uma extensão da regressão simples que permite a utilização de múltiplas variáveis independentes. Nesse caso, foi necessário mudar o cenário 2d (bidimensional) para o 3d(tridimensional, possibilitando a representação gráfica de três dimensões: o ano (tempo), as mortes por Diabetes Mellitus e as mortes por Doenças Cardiovasculares (a variável dependente).

Esta é a fórmula da Regressão Linear Múltipla para este caso:

Doenças cardiovasculares = a.Ano + b.Diabetes Mellitus + c

Figura 7: Regressão Linear Múltipla (Ano + Diabetes -> Cardiovascular)



Fonte: elaborado pelo autor

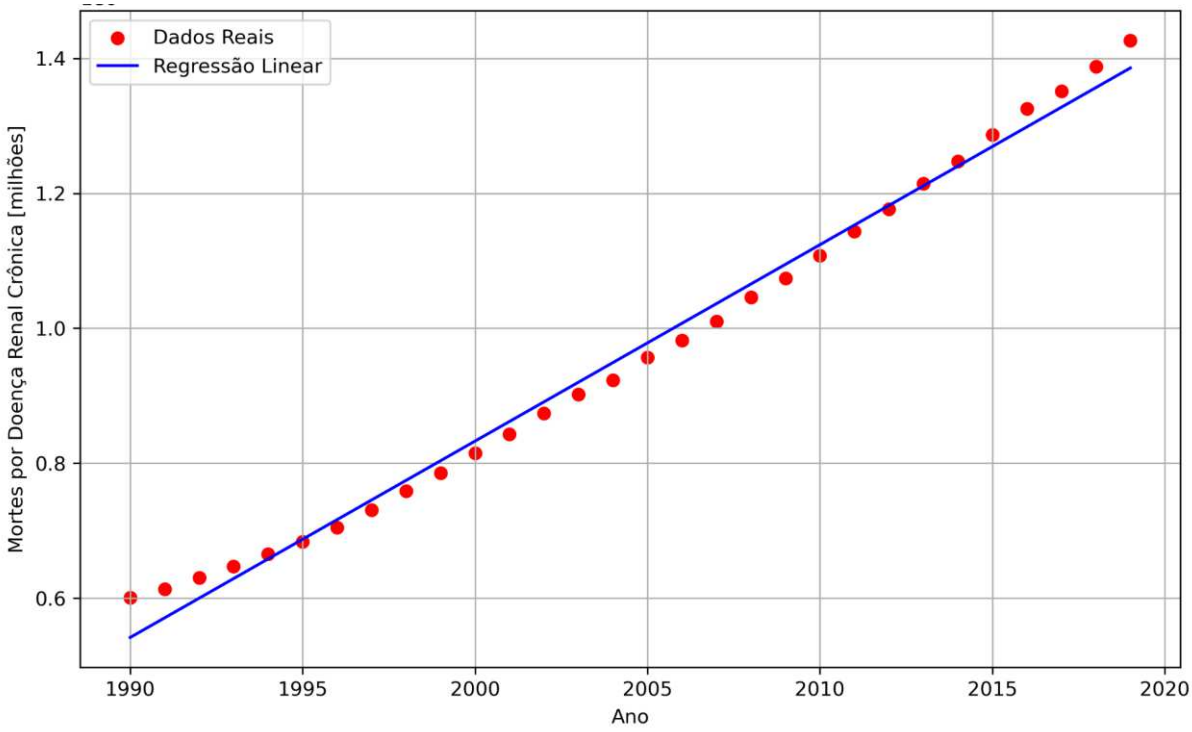
Observa-se que os pontos vermelhos (dados reais) estão perto do plano azul, centralizados, o que mostra que o modelo consegue realizar boas previsões. A superfície azul, o plano da regressão, está subindo com diabetes e anos, mostrando como contribuem positivamente para as mortes por doenças cardiovasculares.

Com este modelo, foi possível investigar como a evolução temporal e o crescimento dos casos de diabetes influenciam diretamente no aumento das mortes por doenças cardiovasculares. Essa abordagem se justifica pelo fato de que a diabetes é um fator de risco reconhecido para doenças cardiovasculares, uma vez que acelera processos como aterosclerose e hipertensão, contribuindo para o comprometimento vascular. Segundo Beckman, Creager e Libby(2002), “Pacientes com diabetes apresentam risco duas a quatro vezes maior de

desenvolver doenças cardiovasculares em comparação com indivíduos não diabéticos", o que reforça a relevância dessa variável na modelagem estatística das causas de morte.

Trabalhando com as causas de morte por Doença Renal Crônica, criou-se o mesmo modelo linear simples, a fim de observar como o tempo afeta essa enfermidade de acordo com os dados.

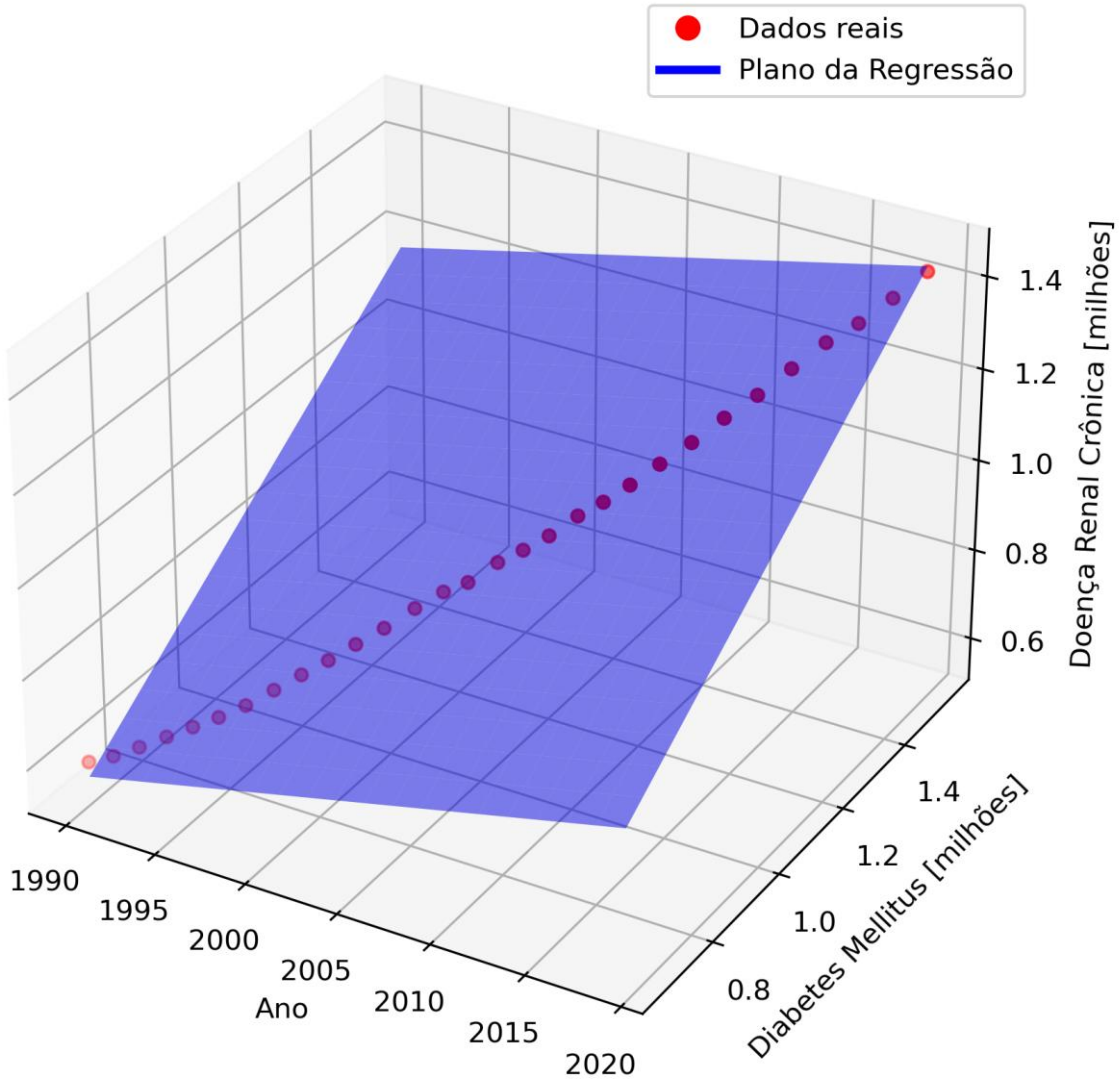
Figura 8: Regressão linear: Doença renal crônica ao longo dos anos



Fonte: elaborado pelo autor

Assim como na Diabetes Mellitus e Doenças Cardiovasculares, para cada ano que passa, o número de mortes por doença renal crônica aumenta consideravelmente. Por fim, com a regressão linear múltipla também, ficou visível a influência da diabetes e tempo no comportamento do número de mortes por doença renal. Confira a Regressão Linear Múltipla para prever a Doença Renal Crônica com base na Diabetes Mellitus e Doenças Cardiovasculares.

Figura 9: Regressão Linear Múltipla (Ano + Diabetes -> Doença renal crônica)



Fonte: elaborado pelo autor

Assim como no primeiro modelo de Regressão Linear Múltipla, o plano cresce ao longo dos anos e também com o aumento do número de mortes por diabetes, mostrando a tendência crescente e a forte relação da diabetes com a doença renal crônica. Os pontos vermelhos, que representam os dados reais, estão bem alinhados ao plano linear e seguem a direção crescente, mas existem pequenas distâncias entre os dados e a superfície, indicando possíveis desvios e erros.

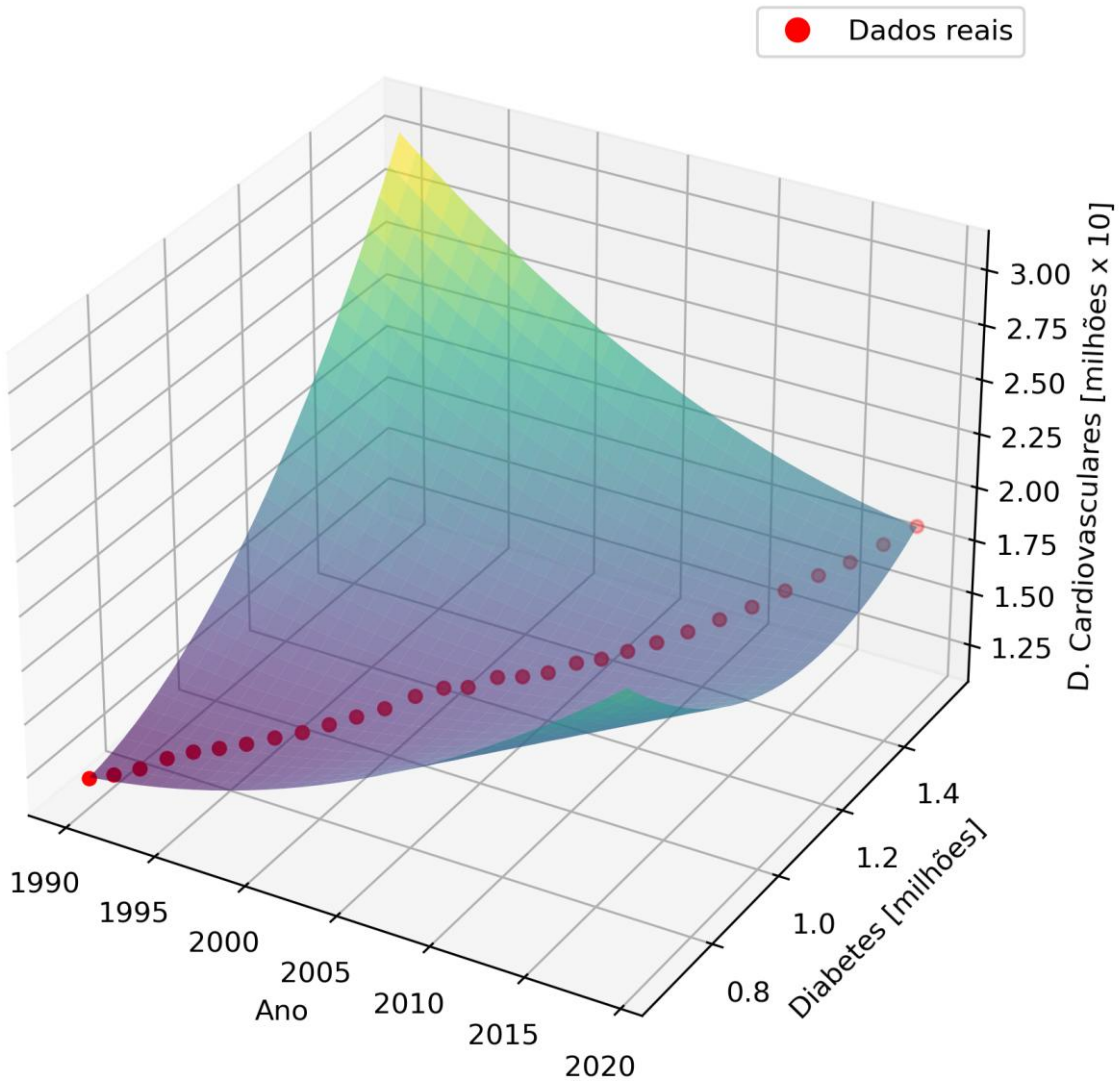
A equação para o modelo de Regressão Linear Múltipla, considerando o tempo (Ano) e as mortes por Diabetes Mellitus como variáveis independentes, é:

Doença Renal Crônica = a.Ano + b.Diabetes Mellitus + c

4.3 MODELO DE REGRESSÃO POLINOMIAL

Esse modelo permite capturar relações não lineares entre as variáveis, ou seja, curvaturas que o modelo linear não conseguiria modelar. Para começar, foi criado o modelo de regressão polinomial para doença cardiovascular com base em diabetes e anos, disposto a seguir.

Figura 10: Regressão Polinomial de Grau 2(Ano + Diabetes -> Cardiovascular)



Fonte: elaborado pelo autor

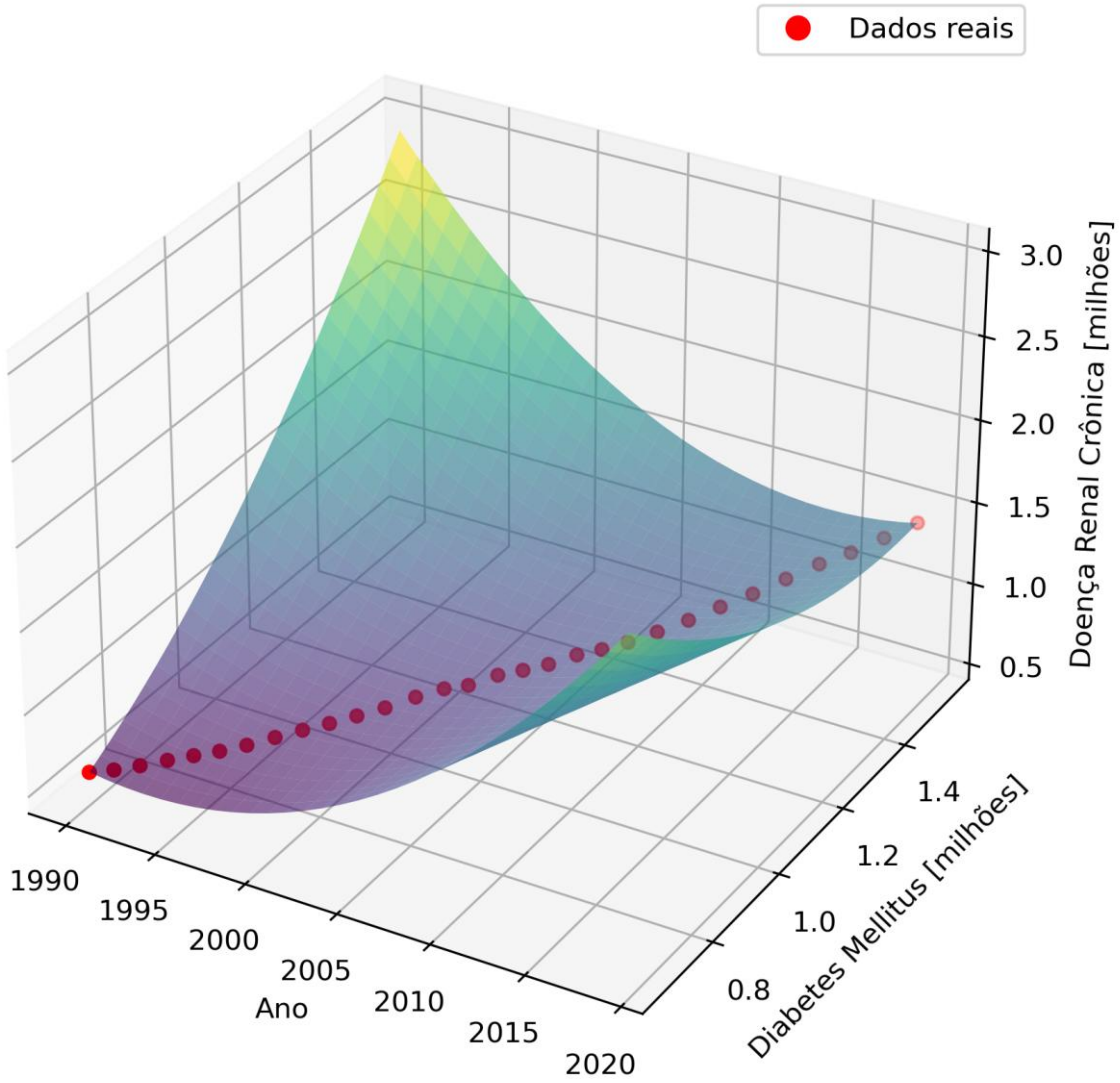
Diferente do plano linear, este possui um plano curvado, pois trabalha com a função quadrática (parabólico). É possível observar que no canto superior esquerdo ocorre uma forte elevação, indicando que quanto mais recente o ano e maior o número de mortes por diabetes, maior o risco de mortes por doenças cardiovasculares. Este tipo de regressão permite capturar

padrões mais complexos entre os anos e os registros de mortes por Diabetes e doenças cardiovasculares. A superfície se adapta melhor a curvatura dos dados reais, proporcionando uma melhor flexibilidade e melhor precisão na modelagem dos dados.

Doenças Cardiovasculares = $a \cdot Ano + b \cdot Diabetes + c \cdot (Ano \times Diabetes) + d \cdot Ano^2 + e \cdot Diabetes^2 + f$

Em seguida foi criado o mesmo modelo polinomial para visualizar a tendência da doença renal crônica de acordo com a diabetes e o tempo.

Figura 11: Regressão Polinomial de Grau 2(Ano + Diabetes -> Doença renal crônica)



Fonte: elaborado pelo autor

Neste caso, os dados reais estão ainda mais próximos da superfície curva, tendo uma alta aderência, embora a amplitude da variável dependente seja menor do que no caso da doença cardiovascular.

Doença Renal Crônica = a·Ano + b·Diabetes + c·(Ano×Diabetes) + d·Ano² + e·Diabetes² + f

Chegando ao objetivo central do projeto — medir o impacto das variáveis Diabetes Mellitus, Doenças Cardiovasculares e Ano sobre a mortalidade por Doença Renal Crônica —, criou-se um modelo de regressão baseado em três variáveis independentes. Utilizou-se primeiramente a Regressão Linear Múltipla, cuja equação tem a seguinte fórmula:

Doença Renal Crônica = a.Ano + b.Diabetes +c.Doenças Cardiovasculares + d

No entanto, por se tratar de um modelo com quatro variáveis no total (uma dependente e três independentes), sua representação gráfica completa exigiria um espaço de quatro dimensões, o que inviabiliza a visualização convencional. Por isso, optou-se por exibir somente a função do modelo em si.

Em seguida, foi ajustado um modelo mais sofisticado: a Regressão Polinomial de segundo grau, que permite capturar relações não lineares e interações entre as variáveis. Sua fórmula estimada é:

Doença Renal Crônica = a.Ano + b.Diabetes + c.Cardiovascular + d.Ano² + e.(Ano.Diabetes) + f.(Ano.Cardiovascular) + g.Diabetes² + h.(Diabetes.Cardiovascular) + i.Cardiovascular² + j

O modelo polinomial apresentou desempenho superior, conforme evidenciado pelas métricas de avaliação, o que reforça a importância de se considerar interações e curvaturas ao analisar fenômenos complexos como a mortalidade por doenças crônicas.

4.4 MÉTRICAS DE DESEMPENHO

As métricas de desempenho ajudam a comparar os modelos criados e a selecionar o melhor desempenho através de métricas estatísticas. Vamos entender as principais métricas de avaliação para modelos de regressão — aquelas que normalmente são utilizadas em modelos preditivos baseados em valores numéricos contínuos, como número de mortes, preços, temperaturas, etc. Com a importação da biblioteca statsmodels.api, foi possível analisar R^2, MAE, MSE e RMSE dos modelos lineares e polinomiais.

Figura 12: Tabela de métricas de desempenho (modelo de regressão linear múltipla)

OLS Regression Results

Dep. Variable:	Doença Renal Crônica	R-squared:	0.997
Model:	OLS	Adj. R-squared:	0.997
Method:	Least Squares	F-statistic:	3411.
Date:	Thu, 26 Jun 2025	Prob (F-statistic):	7.45e-34
Time:	14:33:00	Log-Likelihood:	-326.14
No. Observations:	30	AIC:	660.3
Df Residuals:	26	BIC:	665.9
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-1.746e+07	5.16e+06	-3.386	0.002	-2.81e+07	-6.86e+06
Ano	9044.7711	2618.589	3.454	0.002	3662.184	1.44e+04
Diabetes Mellitus	1.0677	0.212	5.029	0.000	0.631	1.504
Doenças Cardiovasculares	-0.0552	0.028	-1.975	0.059	-0.113	0.002

Omnibus:	5.675	Durbin-Watson:	0.236
Prob(Omnibus):	0.059	Jarque-Bera (JB):	2.211
Skew:	0.303	Prob(JB):	0.331
Kurtosis:	1.816	Cond. No.	3.11e+10

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 3.11e+10. This might indicate that there are
strong multicollinearity or other numerical problems.

Fonte: elaborado pelo autor

O modelo de regressão linear múltipla apresentou um R^2 de 0,997, indicando que quase toda a variação nas mortes por Doença Renal Crônica pode ser explicada pelas variáveis Ano, Diabetes Mellitus e Doenças Cardiovasculares. A significância estatística do modelo é reforçada pela estatística $F = 3411$ ($p < 0,001$).

Partindo para os coeficientes, é notório o impacto dos anos e da Diabetes Mellitus sobre a Doença Renal Crônica. O coeficiente atribuído à variável Ano alcançou 9044,77, o que significa que a cada novo ano aumenta em média 9.045 mortes por doenças renais. Já a diabetes, com o coeficiente de 1,067, evidencia que cada morte por diabetes se associa a 1,07 mortes por doenças renais.

As doenças cardiovasculares ficaram com um coeficiente baixo de -0,06, mostrando pouca contribuição para o número de mortes por doença renal crônica. Contudo, é importante também observar a coluna p>[t], onde todas as variáveis se mostraram significativas para o modelo.

Agora, a seguir, as métricas para o Modelo de Regressão Polinomial:

Figura 13: Tabela de métricas de desempenho (modelo de regressão polinomial)

OLS Regression Results

Dep. Variable:	Doença Renal Crônica	R-squared:	1.000
Model:	OLS	Adj. R-squared:	1.000
Method:	Least Squares	F-statistic:	4.488e+04
Date:	Thu, 26 Jun 2025	Prob (F-statistic):	3.46e-44
Time:	14:33:00	Log-Likelihood:	-272.30
No. Observations:	30	AIC:	560.6
Df Residuals:	22	BIC:	571.8
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0004	3.5e-05	11.576	0.000	0.000	0.000
Ano	0.3937	0.034	11.603	0.000	0.323	0.464
Diabetes Mellitus	-195.9399	16.917	-11.583	0.000	-231.023	-160.857
Doenças Cardiovasculares	11.8969	1.402	8.483	0.000	8.989	14.805
Ano^2	0.1843	1.078	0.171	0.866	-2.050	2.419
Ano Diabetes Mellitus	0.0996	0.008	11.980	0.000	0.082	0.117
Ano Doenças Cardiovasculares	-0.0061	0.001	-7.936	0.000	-0.008	-0.004
Diabetes Mellitus^2	-3.903e-06	3.49e-06	-1.120	0.275	-1.11e-05	3.33e-06
Diabetes Mellitus Doenças Cardiovasculares	3.265e-07	9.94e-07	0.328	0.746	-1.74e-06	2.39e-06
Doenças Cardiovasculares^2	-1.516e-09	7.2e-08	-0.021	0.983	-1.51e-07	1.48e-07

Omnibus:	2.349	Durbin-Watson:	1.053
Prob(Omnibus):	0.309	Jarque-Bera (JB):	1.388
Skew:	0.514	Prob(JB):	0.500
Kurtosis:	3.232	Cond. No.	2.60e+21

Fonte: elaborado pelo autor

A regressão polinomial (grau 2) obteve um R^2 de 1.000 e uma estatística F de 44.880, superando o modelo linear (R^2 = 0.997 e F = 3411). Isso mostra que incluir termos quadráticos melhora significativamente o ajuste do modelo às mortes por Doença Renal Crônica.

Observando as variáveis e seus coeficientes, observam-se vários pontos interessantes. O ano com coeficiente 0.3937 e Valor-p de 0.000 apresentou um impacto significativo e positivo. A Diabete Mellitus com o coeficiente –195.94 e Valor-p de 0.000 tem forte impacto negativo e é altamente significativo. As Doenças Cardiovasculares (coeficiente 11.90 e valor-p 0.000) é significativo e também contribui positivamente. E as interações entre Ano e Diabetes, Ano e Doenças Cardiovasculares e Diabetes e Cardiovasculares se mostraram também significativas e de interação relevante.

Para finalizar, foi feito os cálculos de MAE (Erro Absoluto Médio), MSE (Erro Quadrático Médio e RMSE (Raiz do Erro Quadrático Médio). O Erro Absoluto Médio mede a média da diferença absoluta entre o valor real e o valor previsto, quanto menor o valor, melhor o desempenho do modelo, pois indica que as previsões do modelo se desviam do valor real por X mortes. O Erro Quadrático Médio sinaliza erros maiores, diferente do MAE. E o RMSE é o mesmo que o MSE, porém na mesma escala da variável.

Figura 14: Tabela comparativa de métricas

Métrica	Regressão linear	Regressão polinomial
R^2	0.9975	0.9999
MAE	11287.49	1645.20
MSE	162229657.08	4358355.95
RMSE	12736.94	2087.67

Fonte: elaborado pelo autor

O modelo polinomial apresentou desempenho superior em todos os critérios avaliados. O coeficiente de determinação (R^2) foi de 0.9999, indicando que o modelo é capaz de explicar 99,99% da variabilidade da variável dependente com base nas variáveis independentes — um valor praticamente perfeito. Em comparação, o modelo linear teve R^2 de 0.9975, o que ainda representa um bom ajuste, mas inferior ao modelo polinomial. Além disso, as métricas de erro médio também foram significativamente menores no modelo polinomial.

Esses resultados indicam que o modelo polinomial cometeu menos erros em média e com menor variabilidade em relação aos dados reais, o que é essencial ao trabalhar com dados de saúde pública, onde pequenas variações podem representar milhares de vidas. A melhora no desempenho sugere que a inclusão de termos quadráticos e interações entre variáveis foi eficaz para capturar curvaturas e padrões não lineares que o modelo linear não conseguiu representar.

5. CONCLUSÃO (CONSIDERAÇÕES FINAIS E RECOMENDAÇÕES)

Através da aplicação de técnicas de ciência de dados, este trabalho teve como objetivo principal analisar a influência do tempo, da Diabetes Mellitus e das doenças cardiovasculares na mortalidade por doença renal crônica. Para isso, foram desenvolvidos dois modelos de regressão — linear múltipla e polinomial de segundo grau — e aplicados sobre dados históricos extraídos da plataforma Kaggle, utilizando o *dataset cause_of_deaths* e as ferramentas como Python e bibliotecas especializadas na área de dados.

Os resultados obtidos demonstraram que o modelo de regressão polinomial apresentou desempenho superior, atingindo um R^2 de 0,9999 e erro médio absoluto (MAE) de 1.645,20, contra R^2 de 0,9975 e MAE de 11.287,49 da regressão linear. Isso indica que o modelo polinomial consegue se ajustar melhor à complexidade dos dados, especialmente por considerar interações entre variáveis e comportamentos não lineares, o que ocorre em situações relacionadas à saúde pública.

Além das métricas, as equações resultantes dos modelos permitiram compreender de forma quantitativa a relevância de cada variável. O tempo (ano) e o número de mortes por Diabetes Mellitus mostraram forte impacto sobre a doença renal crônica, enquanto a variável Doenças Cardiovasculares também se mostrou significativa, principalmente no modelo polinomial, reforçando a ligação entre as três doenças crônicas, já apresentados na literatura médica.

Portanto, este estudo evidencia que a ciência de dados pode contribuir para o reconhecimento de padrões ocultos, na antecipação de comportamentos e, principalmente, na construção de ferramentas preditivas capazes de apoiar os órgãos públicos de saúde no manejo e prevenção de doenças crônicas com correlações.

Como recomendação, estudos futuros podem ampliar este projeto inserindo a variável País para promover uma análise específica entre regiões desenvolvidas e subdesenvolvidas. Também, sugere-se uma reorganização temporal da base de dados por décadas ao invés de anos individuais (aglutinamento), possibilitando modelos estatísticos com métricas mais consistentes com a realidade, o que pode favorecer a interpretação dos dados. Desenvolver a matriz de correlação de Spearman e compará-la com a de Pearson, visto que os resultados finais mostraram um melhor cenário em comportamentos não lineares. E, por fim, utilizar os modelos criados para prever eventos futuros.

REFERÊNCIAS

AFKARIAN, afkarian et al. **Adults With Diabetes, 1988-2014.** 2016. Disponível em: https://jamanetwork.com/journals/jama/fullarticle/2542635

BECKMAN, Joshua A., CREAGER, Mark A., e LIBBY, Peter. **Diabetes and Atherosclerosis Epidemiology, Pathophysiology, and Management.** 2002. Disponível em: https://jamanetwork.com/journals/jama/article-abstract/194930

BRUTSAERT, Erika F. **Diabetes mellitus (DM). 2025. Disponível em:** https://www.msdmanuals.com/pt/casa/dist%C3%BArbios-hormonais-e-metab%C3%B3licos/diabetes-mellitus-dm-e-dist%C3%BArbios-do-metabolismo-da-glicose-no-sangue/diabetes-mellitus-dm

FADILLAH, Muhammad Aizri et al. **Bibliometric mapping of data science in education: Trends, benefits, challenges, and future directions.** 2024. Disponível em: https://www.sciencedirect.com/science/article/pii/S2590291125003286?utm_source

GO, Alan S. et al. **Chronic Kidney Disease and the Risks of Death, Cardiovascular Events, and Hospitalization.** 2004. Disponível em: https://www.nejm.org/doi/full/10.1056/NEJMoa041031

MATTOS, Thalita do Bem. **Modelos Não Lineares e suas Aplicações.** 2013. Disponível em: https://www2.ufjf.br/cursoestatistica/files/2014/04/Modelos-N%C3%A3o-Lineares-e-suas-Aplica%C3%A7%C3%B5es.pdf

National Kidney Foundation (NKF). **Chronic kidney disease.** 2025. Disponível em: https://www.kidney.org.uk/chronic-kidney-disease?gad_source=1&gad_campaignid=1066913884&gbraid=0AAAADMbTRg86jhpHGnoGYSzGLWpR1rJS&gclid=EAIaIQobChMIz77z8OvYjgMVgDNECB1MOC9vEAAYASAAEgJSD_D_BwE

RONCO, C. et al. **Cardiorenal syndrome**. 2008. Disponível em: https://doi.org/10.1016/j.jacc.2008.07.051

(SARNAK, Mark J. et al. **Kidney Disease as a Risk Factor for Development of Cardiovascular Disease: A Statement From the American Heart Association Councils on Kidney in Cardiovascular Disease, High Blood Pressure Research, Clinical Cardiology, and Epidemiology and Prevention.** 2003. Disponível em: https://www.ahajournals.org/doi/10.1161/01.CIR.0000095676.90936.80

TRUNFIO, Teresa Angela et al. **Multiple regression model to analyze the total LOS for patients undergoing laparoscopic appendectomy.** 2022. Disponível em:

https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-022-01884-9?utm_source

World Health Organization (WHO). **Noncommunicable diseases.** 2024. Disponível em: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases

World Health Organization (WHO). **Cardiovascular diseases (CVDs).** 2021. Disponível em: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)