UNIVERSIDADE FEDERAL DA PARAÍBA (UFPB)
CENTRO DE CIÊNCIAS SOCIAIS APLICADAS (CCSA)
DEPARTAMENTO DE FINANÇAS E CONTABILIDADE (DFC)
CURSO DE GRADUAÇÃO EM CIÊNCIAS ATUARIAIS (CCA)

CLEO DECKER ANACLETO

PREVISÃO DE MORTALIDADE NO BRASIL COM SÉRIES TEMPORAIS CURTAS:
Uma Abordagem com Combinação de Preditores e Transferência de Aprendizado

João Pessoa - PB
2025

CLEO DECKER ANACLETO

PREVISÃO DE MORTALIDADE NO BRASIL COM SÉRIES TEMPORAIS CURTAS: Uma Abordagem com Combinação de Preditores e Transferência de Aprendizado

Trabalho de Conclusão de Curso como requisito parcial à obtenção do título de Bacharel em Ciências Atuariais pela Universidade Federal da Paraíba.**Área de concentração:** Demografia.

Orientador: Prof. Dr. Filipe Coelho de Lima Duarte.

João Pessoa - PB
2025

Catalogação na publicação
Seção de Catalogação e Classificação

A532p Anacleto, Cleo Decker.
Previsão de mortalidade no Brasil com séries temporais curtas: uma abordagem com combinação de preditores e transferência de aprendizado / Cleo Decker Anacleto. - João Pessoa, 2025.
96 f. : il.

Orientação: Filipe Coelho de Lima Duarte.
TCC (Graduação) - UFPB/CCSA.

1. Previsão de mortalidade. 2. Séries temporais curtas. 3. Deep learning. 4. Transferência de aprendizado. 5. Combinação de preditores. I. Duarte, Filipe Coelho de Lima. II. Título.

UFPB/CCSA CDU 368(043)

Elaborado por ANDRE DOMINGOS DA SILVA FERNANDES - CRB-15/00730

CLEO DECKER ANACLETO

PREVISÃO DE MORTALIDADE NO BRASIL COM SÉRIES TEMPORAIS CURTAS:
Uma Abordagem com Combinação de Preditores e Transferência de Aprendizado

Trabalho de Conclusão de Curso como requisito parcial à obtenção do título de Bacharel em Ciências Atuariais pela Universidade Federal da Paraíba.

João Pessoa - PB, 6 de outubro de 2025:

BANCA EXAMINADORA

Prof. Dr. Filipe Coelho de Lima Duarte
Orientador

Prof. Dr. Adriano Nascimento da Paixão
Examinador

Prof. Dr. Herick Cidarta Gomes de Oliveira
Examinador

João Pessoa - PB
2025

AGRADECIMENTOS

Agradeço primeiramente a Deus, fonte de força e sabedoria em todos os momentos desta caminhada.

À minha família, minha esposa Jordana, meu filho Davi e meus pais, João e Analise, pela compreensão, apoio incondicional e paciência diante das ausências impostas pelas exigências acadêmicas.

Aos amigos Alice, Andressa, Brenda e Valdenilson, que compartilharam cada etapa desta jornada, oferecendo companhia, incentivo e amizade verdadeira.

Aos professores do curso, pelo empenho em transmitir conhecimentos valiosos e fundamentais para minha formação.

Em especial, registro minha profunda gratidão ao professor Filipe, orientador desta pesquisa, pela constante disponibilidade, pelo apoio firme e pela análise minuciosa em todas as fases do trabalho.

Aos professores Herick e Adriano, membros da banca avaliadora, pela leitura atenta, análise justa e contribuições relevantes que enriqueceram este trabalho.

RESUMO

Este trabalho aborda o desafio da previsão de taxas de mortalidade no Brasil com séries temporais curtas, explorando e comparando metodologias estatísticas, de aprendizado de máquina e híbridas. O objetivo foi desenvolver e avaliar um sistema de previsão que aprimorasse a acurácia das projeções demográficas, considerando as particularidades dos dados brasileiros. Foram aplicados modelos estatísticos (Lee-Carter, FDM), clássicos de séries temporais (ARIMA, ETS), arquiteturas de deep learning com transferência de aprendizado (CNN, GRU e um híbrido CNN-GRU) e técnicas de combinação de preditores. Os modelos de deep learning foram pré-treinados na Human Mortality Database e ajustados (fine-tuning) com dados do Brasil (2000–2015) para prever 2016–2019. Os resultados mostram complementaridade entre abordagens: ETS (e ARIMA, de forma próxima) estabeleceu-se como benchmark robusto, liderando nas métricas agregadas (RMSE, MAE e sMAPE), apoiado pela natureza suavizada dos dados oficiais. As redes neurais com transferência não superaram os modelos clássicos no agregado, mas apresentaram ganhos localizados em faixas etárias críticas, como idade 0 e o accident hump masculino. Quanto às combinações, observou-se nuance: a ponderação pelo inverso dos erros reduziu RMSE/MAE em homens e no total, enquanto a média simples manteve-se competitiva em sMAPE, especialmente em mulheres e no agregado. Conclui-se que, embora modelos de séries temporais consagrados sejam altamente eficazes para dados brasileiros, estratégias de transferência de aprendizado e combinação oferecem valor complementar ao capturar dinâmicas complexas em segmentos específicos.

Palavras-chave: Previsão de Mortalidade, Séries Temporais Curtas, *Deep Learning*, Transferência de Aprendizado, Combinação de Preditores

ABSTRACT

This study addresses the challenge of mortality rate forecasting in Brazil using short time series, exploring and comparing statistical, machine learning, and hybrid methodologies. The central objective was to develop and evaluate a forecasting system that enhances the accuracy of demographic projections, considering the specific characteristics of Brazilian data. Statistical models (Lee-Carter, FDM), classical time series models (ARIMA, ETS), deep learning architectures with transfer learning (CNN, GRU, and a hybrid CNN-GRU), and forecast combination techniques were applied. Deep learning models were pre-trained on the Human Mortality Database and fine-tuned with Brazilian data from 2000 to 2015 to forecast the period 2016 to 2019. The results indicate a notable complementarity between approaches. The classical time series models, ETS (and ARIMA, closely), emerged as robust benchmarks, consistently leading in aggregate metrics (RMSE, MAE, and sMAPE), particularly benefiting from the smoothed nature of official data. Deep learning models with transfer learning did not surpass classical models in overall performance but showed localized gains in critical age groups, such as age 0 and the male accident hump. Regarding combinations, a nuanced picture emerged: inverse error weighting reduced RMSE/MAE for males and the total group, while simple averaging remained competitive in sMAPE, especially for females and in aggregate. The study concludes that while established time series models are highly effective for Brazilian mortality data, transfer learning and combination strategies offer complementary value in capturing complex dynamics within specific segments.

Keywords: Mortality Forecasting, Short Time Series, Deep Learning, Transfer Learning, Predictor Combination

LISTA DE QUADROS

LISTA DE FIGURAS

SUMÁRIO

1 INTRODUÇÃO

Este capítulo apresenta uma visão introdutória sobre o tema e o direcionamento da pesquisa. A Seção 1.1 traz a contextualização da importância e dos desafios enfrentados na previsão das taxas de mortalidade. A Seção 1.2 traz o Problema de Pesquisa. Na sequência, a Seção 1.3 apresenta os objetivos geral e específicos do trabalho. A Seção 1.4 expõe as justificativas da pesquisa. Por fim, a Seção 1.5 indica a estrutura do trabalho.

1.1 CONTEXTUALIZAÇÃO

A previsão de mortalidade é fundamental para o planejamento atuarial, a gestão de riscos de longevidade e a formulação de políticas públicas de saúde e previdência. O aumento contínuo da expectativa de vida e a redução das taxas de mortalidade em todas as idades impõem desafios crescentes para a sustentabilidade dos sistemas previdenciários e para a precificação de produtos de seguro (Júnior; Azevedo; Tsunemi, 2019; Perla *et al.*, 2020). A necessidade de estimativas precisas é ainda mais relevante em contextos de baixa taxa de juros, nos quais pequenas variações nas hipóteses de mortalidade podem impactar significativamente o equilíbrio financeiro de fundos de pensão e seguradoras (Perla *et al.*, 2020).

Modelos tradicionais de previsão de taxas de mortalidade, como o de Lee-Carter, são amplamente utilizados devido à sua robustez e simplicidade, mas apresentam limitações importantes, especialmente em populações heterogêneas e com séries históricas curtas, como é o caso brasileiro (Júnior; Azevedo; Tsunemi, 2019; Levantesi; Pizzorusso, 2019; Bravo; Ayuso, 2020). De modo geral, todos os modelos de coorte exigem uma grande quantidade de dados temporais, sendo que se o objetivo é analisar toda a faixa etária, seria necessário que os dados abrangessem mais de um século, a fim de que os dados cobrissem mais de uma coorte completa (Booth; Tickle, 2008).

No contexto de previsão de séries temporais, séries curtas são aquelas compostas por um número reduzido de observações históricas, o que representa um desafio significativo para a modelagem e a obtenção de previsões acuradas (Fawaz *et al.*, 2018). Não há um consenso sobre o número exato de observações que caracteriza uma série como curta, mas, de modo geral, quanto menor o tamanho amostral, maior a dificuldade em capturar padrões e tendências, aumentando o risco de viés e incerteza nas projeções (Duarte, 2024). Estudos recentes mostram que, em séries com menos de sessenta observações, modelos estatísticos tendem a superar métodos de aprendizado de máquina, enquanto a combinação de diferentes técnicas pode mitigar limitações individuais e reduzir a incerteza de seleção de modelos, especialmente em cenários com dados limitados (Cerqueira; Torgo; Soares, 2022; Cruz-Nájera *et al.*, 2022; Duarte; Neto; Firmino,

2024).

A escassez de séries históricas longas no Brasil configura um desafio central para a modelagem de mortalidade. Conforme dados do Instituto Brasileiro de Geografia e Estatística (IBGE), as tábuas de mortalidade nacionais possuem disponibilidade consolidada apenas a partir do ano 2000 (Instituto Brasileiro de Geografia e Estatística (IBGE), 2025), limitando significativamente o horizonte temporal para análises de tendências. Essa restrição amostral, inferior a vinte e cinco anos, contrasta com bases internacionais como o Human Mortality Database (HMD), que oferece séries temporais, algumas centenárias, de mais de quarenta países, constituindo um corpus demográfico robusto para capturar padrões de longo prazo (Human Mortality Database, 2025).

Cabe destacar que, em séries temporais curtas, modelos de aprendizado de máquina estão particularmente sujeitos ao sobreajuste (*overfitting*), ajustando-se excessivamente ao ruído dos dados de treinamento e prejudicando a capacidade de generalização das previsões (Duarte, 2024; Levantesi; Pizzorusso, 2019; Duarte; Neto; Firmino, 2024). Por outro lado, modelos mais simples (e.g., estatísticos lineares) podem incorrer em subajuste (*underfitting*), não capturando adequadamente os padrões relevantes dos dados. A escolha da complexidade do modelo e o uso de técnicas de regularização e validação cruzada são, portanto, essenciais para garantir previsões robustas (Levantesi; Pizzorusso, 2019).

Diante dessas dificuldades, estratégias inovadoras têm sido exploradas para superar a limitação de dados. Dentre elas, destaca-se a transferência de aprendizado (*transfer learning*), que consiste em utilizar o conhecimento adquirido por um modelo treinado em uma tarefa ou domínio com abundância de dados para melhorar o desempenho em outra tarefa ou domínio com dados escassos (Larsson; Vermeire; Verhelst, 2023). No contexto de previsão de taxas de mortalidade, essa abordagem permite adaptar modelos desenvolvidos em bases internacionais ou em populações distintas para contextos nacionais ou específicos, como o caso brasileiro, onde as séries históricas de mortalidade são curtas. Estudos recentes demonstram que a transferência de aprendizado pode aumentar significativamente a acurácia das previsões de taxas de mortalidade, ao possibilitar o aproveitamento de padrões globais e a adaptação a realidades locais, mesmo diante de restrições de dados (Perla *et al.*, 2020; Nalmpatian *et al.*, 2025).

Na previsão de taxas de mortalidade, sistemas híbridos têm se mostrado superiores aos modelos individuais, tanto em termos de acurácia quanto de estabilidade das previsões. Pesquisas indicam que a utilização de sistemas híbridos podem mitigar as limitações individuais de cada técnica, reduzir a incerteza na seleção de modelos e proporcionar maior acurácia às projeções, especialmente em cenários de séries temporais curtas (Duarte, 2024; Gyamerah *et al.*, 2023; Hong *et al.*, 2020) . Além disso, a combinação de modelos, especialmente por meio de abordagens bayesianas e médias ponderadas pelo inverso dos erros, tem se mostrado eficaz para reduzir o risco de modelo e melhorar o poder preditivo das projeções (Bravo; Ayuso, 2020; Gyamerah *et al.*, 2023; Hong *et al.*, 2020; García-Aroca *et al.*, 2023).

A literatura recente destaca o potencial de técnicas de aprendizado de máquina e *deep learning* para superar essas limitações, permitindo a identificação de padrões complexos e a integração de múltiplas variáveis explicativas (Levantesi; Pizzorusso, 2019; Nascimento; Escovedo, 2021; Perla *et al.*, 2020; Gyamerah *et al.*, 2023; Hong *et al.*, 2020). Modelos baseados em redes neurais, recorrentes e convolucionais, têm apresentado boa aderência às séries históricas e resultados promissores na previsão de mortalidade, mesmo em contextos de dados limitados (Nascimento; Escovedo, 2021; Perla *et al.*, 2020).

No Brasil, a aplicação dessas técnicas ainda é incipiente, mas estudos recentes já evidenciam ganhos significativos na precisão das projeções ao empregar métodos de aprendizado profundo no país (Morais; Escovedo; Kalinowski, 2024). A transferência de conhecimento teve sucesso em experimento para dados de mortalidade no Reino Unido (Nalmpatian *et al.*, 2025). A combinação de preditores teve bom desempenho para previsões de mortalidade e expectativa de vida na população portuguesa (Bravo; Ayuso, 2020). Tais avanços são especialmente relevantes para a construção de tábuas de mortalidade mais aderentes à realidade nacional, contribuindo para a sustentabilidade dos regimes previdenciários e para a tomada de decisão em políticas públicas (Júnior; Azevedo; Tsunemi, 2019).

1.2 PROBLEMA DE PESQUISA

Diante das limitações dos modelos tradicionais e da escassez de séries históricas longas no Brasil, surge a seguinte questão:

Como a combinação de preditores estatísticos, demográficos, de *deep learning* e de transferência de aprendizado influenciam a acurácia das projeções de mortalidade brasileira, em comparação com métodos tradicionais e modelos monolíticos?

1.3 OBJETIVOS

A pesquisa se apresenta estruturada em um objetivo geral e quatro objetivos específicos, conforme observados a seguir.

1.3.1 Objetivo geral

Avaliar a eficiência da combinação de preditores, de *deep learning* e de transferência de aprendizado na melhoria da acurácia das projeções de mortalidade brasileira, em comparação com métodos tradicionais e monolíticos.

1.3.2 Objetivos específicos

a) implementar modelos estatísticos para previsão de mortalidade brasileira;

b) desenvolver modelos de *deep learning* para a previsão de mortalidade brasileira, explorando a transferência de aprendizado a partir de dados do HMD;

c) construir algoritmo com combinação de preditores mediante média ponderada pelo inverso das métricas de erros, integrando as projeções dos modelos individuais;

d) avaliar e comparar o desempenho preditivo dos modelos por meio das métricas de erro RMSE (*Root Mean Square Error*), MAE (*Mean Absolute Error*) e sMAPE (*Symmetric Mean Absolute Percentage Error*).

1.4 JUSTIFICATIVA

A estimativa de padrões futuros de mortalidade constitui um pilar fundamental para a atuária, a gestão de sistemas de previdência social e o desenho de políticas governamentais. Projeções que apresentem desvios em relação à realidade demográfica impactam diretamente a solvência de esquemas de aposentadoria e a correta avaliação de instrumentos financeiros, a exemplo de seguros e rendas vitalícias. Este cenário torna-se particularmente sensível em ambientes econômicos de juros estruturalmente baixos, nos quais oscilações mínimas nas premissas sobre mortalidade e longevidade produzem variações significativas no cálculo de provisões técnicas e obrigações futuras (Rabitti; Borgonovo, 2020; Cocco; Gomes, 2012).

Embora o modelo de Lee–Carter permaneça amplamente difundido na literatura especializada, em virtude de sua simplicidade e relativa adaptabilidade, suas restrições metodológicas tornam-se progressivamente mais evidentes. Dentre tais limitações, destacam-se a dificuldade em incorporar heterogeneidades populacionais, a sensibilidade a eventos de mortalidade catastróficos e a dependência de longas séries históricas para garantir estabilidade paramétrica (Basellini; Camarda; Booth, 2023). Tais fragilidades mostram-se especialmente relevantes em realidades como a brasileira, marcada pela relativa escassez de dados longitudinais robustos, o que potencialmente compromete a confiabilidade de qualquer exercício projetivo. A comparação com bancos de dados internacionais consolidados, a exemplo do Human Mortality Database (2025) que oferece séries históricas seculares para inúmeras populações, realça a carência de abordagens alternativas capazes de operar com escassez de dados.

Eventos recentes de impacto global, como a pandemia de COVID-19, expuseram a vulnerabilidade inerente aos modelos tradicionais de projeção e os respectivos reflexos financeiros decorrentes de oscilações bruscas nos índices de mortalidade para seguradoras e entidades de previdência (Schnürch *et al.*, 2022). A conjugação de diferentes perspectivas metodológicas revela

potencial para produzir estimativas mais confiáveis e aplicáveis à gestão de risco, paralelamente ao avanço do arcabouço teórico da demografia atuarial.

Investigações contemporâneas sugerem avanços expressivos na acurácia projetiva por meio da combinação de metodologias estatísticas convencionais com algoritmos de aprendizado de máquina e *deep learning*, incluindo redes neurais recorrentes e estratégias de ensemble. Esses modelos híbridos demonstraram capacidade não apenas de contornar limitações amostrais, mas também de ampliar a estabilidade e a precisão das projeções em cenários complexos e sujeitos a rupturas (Nigri; al., 2019; Petneházi; Gáll, 2019; Gyamerah *et al.*, 2023).

Diante desse contexto, o presente estudo justifica-se pela necessidade de desenvolver abordagens metodológicas adequadas à previsão de mortalidade no Brasil, onde a curta extensão das séries históricas contrasta com a urgência de projeções robustas para garantir a sustentabilidade financeira de longo prazo. Propõe-se, assim, a integração entre modelos estatístico-demográficos, técnicas de *deep learning* e estratégias de transferência de aprendizado, visando aproveitar padrões globais de mortalidade e adaptá-los à realidade nacional. Espera-se que a avaliação dessa articulação metodológica, conforme delineado nos objetivos deste trabalho, proporcione maior rigor, confiabilidade e utilidade prática às projeções de mortalidade.

1.5 ESTRUTURA DO TRABALHO

O trabalho está organizado em cinco capítulos, incluindo esta introdução. O Capítulo 2 apresenta a revisão da literatura, abordando os principais modelos de previsão de mortalidade, suas limitações e os avanços recentes com aprendizado de máquina, combinação de preditores e transferência de aprendizado. O Capítulo 3 descreve a metodologia do estudo, descreve as fontes, os critérios de seleção e a preparação dos dados. Detalha ainda a implementação dos modelos, as técnicas de combinação de preditores e a adaptação do modelo de *deep learning* para transferência de aprendizado. O Capítulo 4 apresenta e discute os resultados obtidos, comparando o desempenho das diferentes abordagens e analisando suas implicações práticas. Por fim, o Capítulo 5 traz as conclusões, limitações do estudo e sugestões para pesquisas futuras.

2 REVISÃO DE LITERATURA

Este capítulo está estruturado de modo a proporcionar uma visão abrangente sobre o tema. Inicialmente, a Seção 2.1 discute a importância da projeção de mortalidade e os desafios inerentes ao trabalho com séries temporais curtas. Em seguida, na Seção 2.2 são apresentados os principais modelos tradicionais, como o Lee-Carter, FDM, ARIMA e ETS, destacando suas características, aplicações e limitações. O capítulo avança para a Seção 2.3 para os modelos de aprendizado de máquina, abordando redes neurais convolucionais e recorrentes, com ênfase em suas potencialidades e restrições diante de bases de dados reduzidas. Na sequência, a Seção 2.4 explora o conceito de transferência de aprendizado, ressaltando sua relevância para cenários com poucas observações. Na Seção 2.5 são discutidas estratégias de combinação de preditores, com foco na ponderação pelo inverso dos erros, analisando como essas técnicas podem contribuir para a precisão das projeções. Por fim, a Seção 2.6 apresenta as considerações do capítulo.

2.1 PROJEÇÃO DE MORTALIDADE

2.1.1 Importância da projeção de mortalidade em atuária

Os estudos de projeção da mortalidade são explorados pela ciência atuarial, haja vista constituírem elementos altamente relevantes para o adequado gerenciamento de riscos e análise de sustentabilidade de regimes previdenciários e planos de seguros. Previsões precisas das taxas de mortalidade são elementos chave para a definição de preços de produtos de seguro, determinação de pagamentos de pensão e gerenciamento eficiente das reservas técnicas, influenciando diretamente a solvência e a competitividade das instituições (Apicella *et al.*, 2024; Hong *et al.*, 2020).

Nesse contexto, a previsão de mortalidade refere-se ao processo de estimar as taxas de mortalidade humana ao longo do tempo, sendo a taxa de mortalidade, usualmente denotada por $m_{x,t}$ ou ${}_tm_x$, a medida central desse processo. A taxa de mortalidade representa a frequência de óbitos para uma determinada idade x em um período específico t, quantificando o risco de morte para cada faixa etária em determinado intervalo de tempo. Bravo (2007) expressa a taxa central de mortalidade matematicamente por:

$$ {}_tm_x = \frac{{}_td_x}{{}_tE_x} \tag{2.1} $$

em que ${}_td_x$ corresponde ao número de óbitos registrados e ${}_tE_x$ à quantidade de pessoas expostas ao risco para a idade x no ano t. A partir dessas taxas, constroem-se as tábuas de mortalidade, instrumentos que sintetizam funções demográficas utilizadas para a análise da longevidade, quantificação das probabilidades de falecimento e sobrevivência por faixas etárias e cálculo

da expectativa de vida de determinada população em estudo (Bravo, 2007). Essas tábuas são fundamentais para a precificação e quantificação de riscos em seguros de vida, bem como para o gerenciamento financeiro e atuarial dos fundos de pensão (Duarte; Neto; Firmino, 2024).

A precisão dos modelos utilizados, sejam eles tradicionais ou baseados em aprendizado de máquina, acarretam avaliações de risco mais adequadas, além de contribuir para a eficiência de capital das seguradoras e para a sustentabilidade dos sistemas previdenciários (Levantesi; Nigri; Piscopo, 2021; Apicella *et al.*, 2024). Por outro lado, desafios como a limitação de dados e a ocorrência de choques externos, como a pandemia de COVID-19, exigem o desenvolvimento e a aplicação de métodos cada vez mais robustos e adaptáveis (Nalmpatian; Heumann; Pilz, 2023; Levantesi; Nigri; Piscopo, 2021). Dessa forma, estudos tem avançado na busca por metodologias inovadoras que possam contribuir com a ciência atuarial (Júnior, 2022; Nascimento; Escovedo, 2021).

2.1.2 Desafios em séries temporais curtas

A modelagem de séries temporais de mortalidade com poucas observações representa um grande desafio para a prática atuarial. Séries temporais curtas, caracterizadas por um número reduzido de unidades amostrais, são comuns em contextos como pequenas populações, países subdesenvolvidos, faixas etárias específicas ou períodos recentes de observação (Cerqueira; Torgo; Soares, 2022; Cruz-Nájera *et al.*, 2022). A escassez de dados dificulta a identificação de padrões subjacentes, aumenta a sensibilidade a ruídos e pode comprometer a acurácia das previsões (Thomakos *et al.*, 2022). Nessas situações, cada observação individual assume um peso maior, o que pode resultar em previsões enviesadas e instáveis (Weigand; Lange; Rauschenberger, 2021).

Entre os principais desafios enfrentados, destaca-se o sobreajuste (*overfitting*), que ocorre quando modelos excessivamente complexos capturam o ruído dos dados em vez das tendências reais, tornando-se especialmente problemático em séries pequenas (Li, 2014; Booth; Tickle, 2008). A instabilidade dos parâmetros é outro obstáculo relevante, pois a limitação de dados pode levar a estimativas pouco confiáveis, dificultando a generalização dos modelos para novos cenários (Venter; Şahin, 2018; Ekheden; Hä¶ssjer, 2015). Essa instabilidade é agravada pela variabilidade inerente aos dados de mortalidade, frequentemente influenciados por efeitos de idade, período e coorte (Booth; Tickle, 2008). Além disso, limitações estatísticas surgem da dificuldade em estimar parâmetros com precisão, resultando em intervalos de confiança amplos e previsões incertas (Li, 2014).

A literatura aponta que o tamanho amostral exerce influência direta sobre o desempenho dos modelos de previsão. Cerqueira, Torgo e Soares (2022) demonstraram que a acurácia dos modelos de aprendizado de máquina tende a aumentar com o crescimento do conjunto de dados, enquanto modelos estatísticos tradicionais superam os métodos de aprendizado de

máquina quando as séries possuem menos de sessenta observações. Resultados semelhantes foram encontrados por Cruz-Nájera *et al.* (2022), que observaram desempenho superior de modelos lineares em séries muito curtas, com aproximadamente duas dúzias de observações. De modo geral, as maiores reduções nos erros de previsão ocorrem até cerca de cem observações, com ganhos mais graduais a partir desse ponto e uma tendência à equalização entre os diferentes tipos de modelos (Cerqueira; Torgo; Soares, 2022).

Para mitigar as limitações impostas pelos dados escassos, diversas estratégias têm sido propostas. Destaca-se que a combinação de diferentes técnicas pode ser especialmente vantajosa em cenários de séries temporais curtas. Sistemas híbridos, que integram modelos estatísticos e de aprendizado de máquina, têm apresentado resultados superiores aos modelos individuais em projeções de taxas de mortalidade (Duarte; Neto; Firmino, 2024; Li, 2022; Gyamerah *et al.*, 2023; Nigri; al., 2019). Modelos com transferência de aprendizado também se mostraram promissoras (Nalmpatian *et al.*, 2025). Esses achados reforçam a importância de estratégias que aproveitem técnicas sofisticadas de modelagem para superar as limitações impostas pelos dados esparsos de mortalidade.

2.2 MODELOS TRADICIONAIS DE PROJEÇÃO DE MORTALIDADE

2.2.1 Lee-Carter

O modelo introduzido por Lee e Carter (1992) é um modelo estocástico amplamente usado para prever taxas de mortalidade específicas por idade. O modelo é baseado em uma estrutura simples, que representa o registro das taxas de mortalidade específicas por idade como uma combinação de componentes específicos por idade e variáveis no tempo. O modelo é expresso matematicamente como:

$$ln({}_tm_x) = a_x + \beta_x k_t + \varepsilon_{x,t} \tag{2.2}$$

sendo ${}_tm_x$ a taxa central de mortalidade para a idade x no tempo t, a_x a média dos logaritmos naturais de ${}_tm_x$ para cada idade x durante o período de treinamento, β_x o coeficiente que aponta a velocidade de variação da taxa de mortalidade em função das variações de k_t, que é o fator que registra a tendência linear de ${}_tm_x$. Por fim, $\varepsilon_{x,t}$ é o termo de erro não capturado pelo modelo (Duarte, 2024; McQuire; Kume, 2022).

Uma vez que o modelo é inerentemente subdeterminado, ou seja, possui várias soluções para os parâmetros que poderiam satisfazer a equação, a fim de obter uma solução única, são aplicadas normalizações aos parâmetros. Os parâmetros b_x – normalizados para somar um – e k_t – normalizados para somar zero. Assim, os parâmetros do modelo são estimados usando o método de decomposição de valor singular (SVD), que é uma técnica de álgebra linear que decompõe uma matriz em três outras matrizes. No contexto deste modelo, ela é aplicada à matriz

dos logaritmos das taxas de mortalidade, após a remoção da média temporal. Os primeiros vetores à direita e à esquerda e o valor principal do SVD, após a normalização descrita, fornecem uma solução única para os parâmetros (Lee; Carter, 1992).

Após o desenvolvimento e ajuste do modelo estatístico às taxas de mortalidade históricas, usando o SVD, passa-se a prever o índice de mortalidade k_t para o futuro. Para tanto, é usado um modelo autorregressivo integrado de médias móveis (*Autoregressive Integrated Moving Average* ou ARIMA, em inglês), sendo no caso do artigo original proposto por Lee e Carter (1992), um modelo denominado passeio aleatório com derivação, ARIMA (0, 1, 0) com uma constante. Isso implica que o índice k_t tende a diminuir a uma taxa aproximadamente constante ao longo do tempo, mas incorporando um componente aleatório (Lee; Carter, 1992). O modelo ARIMA permite a construção de intervalos de confiança probabilísticos para as previsões de k_t, que são usadas para gerar intervalos de confiança para as taxas de mortalidade específicas previstas (Lee; Carter, 1992).

O modelo Lee-Carter tem vantagens que contribuíram para a sua adoção generalizada. O modelo é simples. Os componentes específicos da idade e variáveis no tempo fornecem uma clara decomposição das tendências de mortalidade (McQuire; Kume, 2022). O uso de modelos ARIMA para previsão do parâmetro k_t permite projeções de longo prazo com bom desempenho para países desenvolvidos e em desenvolvimento (Zili; Mardiyati; Lestari, 2018).

Por outro lado, o modelo tem limitações, Lee e Carter (1992) já reconheceram inconsistências em taxas de mortalidade de jovens, apresentando um ajuste menos preciso para estes grupos etários. Os autores também apontaram que o modelo pode apresentar instabilidade quando o período da base de dados é reduzido para dez ou vinte anos, afetando as estimativas e previsões. O modelo assume que o parâmetro variável no tempo k_t evolui linearmente ao longo do tempo. Isso pode limitar sua capacidade de capturar tendências não lineares ou quebras estruturais nas taxas de mortalidade (Andrs-Sánchez; Puchades, 2019).

2.2.2 FDM

O *Functional Demographic Model* (FDM) trata cada ano t como uma curva suave de mortalidade sobre a idade, estimando primeiro uma função subjacente $f_t(x)$ por meio de suavização não parametrizada antes de qualquer decomposição. Essa etapa de *smoothing*, geralmente feita com *splines* penalizados ponderados e, quando necessário, com restrições qualitativas (e.g. monotonicidade em idades altas), reduz ruído e torna o procedimento robusto a anos atípicos (e.g. guerras e pandemias) (Hyndman; Ullah, 2007). A representação funcional central do FDM é:

$$f_t(x) \;=\; \mu(x) \;+\; \sum_{k=1}^{K} \beta_{t,k}\, \phi_k(x) \;+\; e_t(x) \tag{2.3}$$

onde $\mu(x)$ é a média funcional ao longo do tempo, $\{\phi_k(x)\}$ são funções-base (autofunções ortonormais) que capturam os modos de variação por idade, $\beta_{t,k}$ são os escores temporais (coeficientes por ano) e $e_t(x)$ é o resíduo funcional. Para $K = 1$ essa formulação reconduz-se à estrutura do Lee–Carter; para $K > 1$ ela permite múltiplos modos de variação por idade (Hyndman; Ullah, 2007).

Note-se que o modelo assume que observamos uma realização ruidosa de $f_t(x)$. Seja $y_t(x)$ o logaritmo da taxa observada de mortalidade ou fertilidade no ano t e idade x. Então:

$$y_t(x_i) = f_t(x_i) + \sigma_t(x_i)\epsilon_{t,i} \tag{2.4}$$

onde $\epsilon_{t,i} \sim N(0,1)$, independente e identicamente distribuído, reflete a variabilidade observacional, que é maior onde a população exposta é menor ou a taxa é muito baixa (Hyndman; Ullah, 2007).

Na prática o *pipeline* é modular e direto: (i) suaviza-se cada curva anual $y_t(x)$ para obter $\hat{f}_t(x)$; (ii) aplica-se Análise por Componentes Principais Funcionais (FPCA) às curvas suavizadas para estimar $\hat{\mu}(x)$, $\{\hat{\phi}_k(x)\}$ e os escores $\{\hat{\beta}_{t,k}\}$; (iii) ajustam-se modelos de série temporal (e.g. ARIMA) separadamente a cada série $\{\hat{\beta}_{t,k}\}_t$ e projetam-se esses escores para o horizonte h; (iv) reconstrói-se $\hat{f}_{n+h}(x)$ com a soma da média e das bases ponderadas pelos escores previstos; (v) obtêm-se intervalos de previsão combinando a variabilidade de suavização, a incerteza dos escores e o termo residual. Essa sequência é explícita e repetível, facilitando diagnóstico e construção de intervalos de confiança (Hyndman; Ullah, 2007).

A robustez e a acurácia desse *pipeline* dependem criticamente de uma série de escolhas práticas: os pesos no *smoothing* costumam ser proporcionais ao inverso da variância observacional por idade (para corrigir heterocedasticidade), e as restrições (monotonicidade/concavidade) melhoram estabilidade em idades extremas; a decomposição usa FPCA robusta para neutralizar anos-*outliers* - anos em que eventos atípicos distorcem o padrão etário usual; e a escolha de K deve equilibrar variância explicada e parcimônia (critério orientado por desempenho de previsão *out-of-sample*).

Além disso, para múltiplas populações existem extensões coerentes (*product–ratio*) - método que combina previsões independentes de populações relacionadas atra´ves de um modelo de razão para evitar divergências não realista no longo prazo - e *multilevel FPCA* que tratam dependência entre subpopulações sem permitir divergência indesejada nas previsões (Hyndman; Ullah, 2007; Hyndman; Booth; Yasmeen, 2013; Shang, 2016).

Comparando com Lee–Carter o FDM é, em essência, uma generalização funcional e mais robusta, pois incorpora *smoothing* explícito, múltiplos componentes funcionais e estimativas robustas, podendo produzir previsões e intervalos de incerteza melhores quando há variabilidade complexa por idade ou anos atípicos. Para problemas de coerência entre subgrupos (sexos, estados, países) as variantes *product–ratio* e *multilevel* oferecem soluções práticas para manter relações históricas entre séries sem sacrificar acurácia (Hyndman; Ullah, 2007; Hyndman; Booth; Yasmeen, 2013; Shang, 2016).

No entanto, o modelo Lee-Carter permanece competitivo em certos cenários, particularmente com dados estáveis e de alta qualidade ou horizontes de previsão mais curtos (Fang; Härdle, 2015). Algumas pesquisas sugerem que o aprendizado profundo ou os aprimoramentos do aprendizado de máquina do Lee-Carter podem rivalizar ou superar a precisão do FDM, indicando que o FDM não é universalmente superior (Marino; Levantesi; Nigri, 2023). A sensibilidade à qualidade dos dados, potencial sobreajuste, intensidade computacional e desafios na captura de mudanças abruptas de mortalidade ou efeitos de coorte não bem representados por funções suaves podem ser desvantagens do modelo (Fang; Härdle, 2015). A dependência do FDM em componentes principais funcionais lineares pode limitar sua capacidade de modelar totalmente a dinâmica de mortalidade não linear (Yap; Pathmanathan; Dabo-Niang, 2025).

2.2.3 Modelos de Séries Temporais

2.2.3.1 ARIMA

O modelo autorregressivo integrado de médias móveis (*Autoregressive Integrated Moving Average* ou ARIMA), também conhecido como metodologia Box & Jenkis, nomeada em razão da publicação pelos estatísticos Box e Jenkins da obra *Time series analysis: forecasting and control* (1970), é uma técnica estatística amplamente empregada na projeção de dados temporais (Gujarati; Porter, 2011). Este modelo consiste na combinação de três componentes:

i) O componente autorregressivo (AR) que usa os valores passados da série temporal para prever valores futuros. Indicado pelo parâmetro p, representa o número de defasagens utilizadas no modelo. Assim, o valor previsto Y no período t é simplesmente uma proporção (α) mais um choque aleatório no mesmo período t. Um processo AR de ordem p-ésima pode ser expresso pela seguinte expressão (Gujarati; Porter, 2011):

$$(Y_t - \delta) = \alpha_1 (Y_{t-1} - \delta) + \alpha_2 (Y_{t-2} - \delta) + \ldots + \alpha_p \left(Y_{t-p} - \delta\right) + u_t \tag{2.5}$$

em que δ é a média de Y, α o coeficiente autorregressivo associado, medindo o impacto de Y_{t-n} sobre Y_t e u_t é o erro aleatório, ruído branco, então Y_t segue um processo autorregressivo estocástico de ordem p ou AR(p).

ii) O componente de média móvel (MA) apresenta uma combinação linear de termos de erros de ruído branco. É indicado pelo parâmetro q, que representa o número de termos de erros passados usados no modelo. Sua equação geral, que representa um processo MA(q) pode ser expresso da seguinte forma (Gujarati; Porter, 2011):

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \ldots + \beta_q u_{t-q} \tag{2.6}$$

em que μ é uma constante, β_0 coeficiente de média móvel associado, medindo o impacto de u_{t-n} sobre Y_t e u é um termo de erro estocástico de ruído branco. Assim Y_t corresponde a uma constante mais uma média móvel dos termos de erros atuais e passados, ou seja, segue um processo de média móvel de ordem q.

iii) O terceiro componente aponta sobre a necessidade de diferenciar a série temporal a fim de deixá-la estacionária (Gujarati; Porter, 2011). É o componente integrado (I), indicado pelo parâmetro d, que aponta o número de vezes que a série deve ser diferenciada a fim de torná-la estacionária para aplicar o modelo ARMA. Portanto, um processo ARIMA (p, d, q) é uma série temporal autorregressiva integrada de médias móveis, em que p aponta o número de termos autorregressivos, d a quantidade de vezes que a série deve ser diferenciada para se tornar estacionária e q o número de termos incluídos na média móvel.

A metodologia *Box & Jenkis* para estimar os parâmetros de um modelo ARIMA envolve quatro etapas, segundo Gujarati e Porter (2011): Etapa 1 – Identificação. A descoberta dos valores apropriados de p, d e q, com auxílio de correlograma e correlograma parcial resultantes da aplicação das funções de correlação amostral e de correlação amostral parcial; Etapa 2 – Estimação. Após a identificação dos valores p e q, estima-se os parâmetros dos termos autorregressivos e os de média móvel incluídos no modelo. Pode ser aplicado por mínimos quadrados simples ou por métodos não lineares; Etapa 3 – Verificação do diagnóstico. A ordem do modelo ARIMA e os parâmetros calculados devem se ajustar razoavelmente bem aos dados. Esta análise pode ser feita mediante a aplicação de um teste que verifica se os resíduos são ruídos brancos. Caso não sejam, o processo deve recomeçar, o que torna a metodologia um verdadeiro processo iterativo; Etapa 4 – Previsão. Ajustado o modelo passa-se a gerar previsões. Em muitos casos, as previsões geradas por este método são mais confiáveis do que modelos mais complexos, especialmente para o curto prazo, o que tornou a modelagem ARIMA popular em diversos campos de aplicação (Rizvi, 2024).

Um dos campos de uso dos modelos ARIMA é a previsão de mortalidade, sendo aplicado em diversos contextos, inclusive em taxas de mortalidade por COVID-19, mortes relacionadas a medicamentos e taxas de mortalidade infantil. Um estudo na Malásia usou o modelo ARIMA para prever as taxas de mortalidade por COVID-19 (Nor *et al.*, 2024). No Irã, o modelo ARIMA foi aplicado na projeção de mortes relacionadas a drogas (Zarghami *et al.*, 2023). Na Nigéria, os modelos ARIMA foram utilizados para prever taxas de mortalidade infantil (Ifeanyichukwu *et al.*, 2022).

2.2.3.2 ETS

Segundo Hyndman e Athanasopoulos (2021), o método de suavização exponencial foi proposto no final da década de 1950 (Brown, 1959; Holt, 1957; Winters, 1960), e acabou se mostrando como um dos métodos de previsão mais prósperos. As previsões produzidas usando

métodos de suavização exponencial são médias ponderadas de observações passadas, com os pesos em decaindo exponencialmente à medida que as observações envelhecem. A estrutura do modelo pode ser aplicada em diversos contextos de séries temporais, gerando previsões precisas de forma rápida.

Há diversos métodos e variações de suavizações exponenciais, sendo a primeira a chama Suavização Exponencial Simples (SES), utilizado em séries temporais sem tendência ou sazonalidade evidentes. Fundamenta-se na atribuição de pesos decrescentes exponencialmente às observações passadas, onde um parâmetro α, entre 0 e 1, controla a velocidade do decaimento dos pesos atribuídos às observações passadas y_{t-n}, conforme a fórmula a seguir (Hyndman; Athanasopoulos, 2021):

$$\hat{y}_{t+1} = \alpha y_t + \alpha (1-\alpha) y_{t-1} + \alpha (1-\alpha)^2 y_{t-2} + \ \ldots \tag{2.7}$$

A aplicação da SES requer a escolha de α, que pode ser estimado minimizando a soma dos quadrados dos erros.

Holt, em 1957 estendeu a suavização exponencial simples para permitir a previsão de dados com tendência, cujo método envolve três equações, uma para previsão e duas de suavização, sendo a primeira para o nível e outra para a tendência (Hyndman; Athanasopoulos, 2021):

$$\text{Equação de previsão: } \hat{y}_{t+h|t} = l_t + hb_t \tag{2.8}$$

$$\text{Equação de nível: } l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1}) \tag{2.9}$$

$$\text{Equação de tendência: } b_t = \beta^* (l_t - l_{t-1}) + (1-\beta^*) b_{t-1} \tag{2.10}$$

onde l_t é a estimativa do nível da série no tempo t; b_t é a estimativa da tendência da série no tempo t; α é o parâmetro de suavização do nível ($0 \leq \alpha \leq 1$); e β^* o parâmetro de suavização da tendência ($0 \leq \beta^* \leq 1$) (Hyndman; Athanasopoulos, 2021).

Posteriormente, o modelo foi estendido para capturar a sazonalidade, ficando conhecido como método Holt-Winters, que inclui uma terceira equação de suavização, cujo componente é denotado por s_t e com parâmetro γ. Há duas variações do método que diferem entre si na natureza do componente sazonal. O método aditivo, preferido quando as variações sazonais se apresentam constantes no período, e o método multiplicativo, utilizado quando as variações sazonais se alteram proporcionalmente ao nível da série (Hyndman; Athanasopoulos, 2021).

Conforme explicam Hyndman e Athanasopoulos (2021), a combinação dos componentes e suas variações possibilita nove métodos de suavização exponencial, que podem ser resumidas, respectivamente, nos quadros e figura abaixo:

Quadro 2.1 – *Classificação bidirecional de métodos de suavização exponencial*

Componente de Tendência	Componente Sazonal		
	N (Nenhum)	A (Aditivo)	M (Multiplicativo)
N (Nenhum)	(N, N)	(N, A)	(N, M)
A (Aditivo)	(A, N)	(A, A)	(A, M)
A_d (Aditivo amortecido)	(A_d, N)	(A_d, A)	(A_d, M)

Fonte: Adaptado de Hyndman e Athanasopoulos (2021)

Quadro 2.2 – *Classificação dos métodos por nome*

Abreviação	Método
(N, N)	Suavização exponencial simples
(A, N)	Método linear de Holt
(A_d, N)	Método de tendência aditiva amortecida
(A, A)	Método de Holt-Winters aditivo
(A, M)	Método de Holt-Winters multiplicativo
(A_d, M)	Método de Holt-Winters amortecido

Fonte: Adaptado de Hyndman e Athanasopoulos (2021)

Quadro 2.3 – *Equações dos métodos de suavização exponencial*

CT	Sazonal		
	N	A	M
N	$\hat{y}_{t+h\|t} = \ell_t$ $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = \ell_t s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t / s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t / \ell_{t-1}) + (1-\gamma)s_{t-m}$
A	$\hat{y}_{t+h\|t} = \ell_t + hb_t$ $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + hb_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = (\ell_t + hb_t)s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t / s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t / (\ell_{t-1} + b_{t-1})) + (1-\gamma)s_{t-m}$
A_d	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t$ $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - \phi b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = (\ell_t + \phi_h b_t)s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t / s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t / (\ell_{t-1} + \phi b_{t-1})) + (1-\gamma)s_{t-m}$

Fonte: Adaptado de Hyndman e Athanasopoulos (2021)

Estudos indicaram potencial uso do ETS na projeção de taxas de mortalidade, superando modelos multivariados, como Lee-Carter, de forma consistente em diferentes configurações, incluindo faixas etárias variáveis, períodos, horizontes de tempo e suavidade de dados, apontando que informações adicionais, como o de idades vizinhas podem adicionar ruído em vez de poder preditivo (Feng; Shi, 2018). Contudo, em algumas situações o modelo ETS falha em garantir coerência etária, gerando previsões de longo prazo cujo resultado diverge da característica biológica, como por exemplo uma pessoa de 50 anos com uma taxa de mortalidade maior do que uma de 60 anos (Shi, 2020).

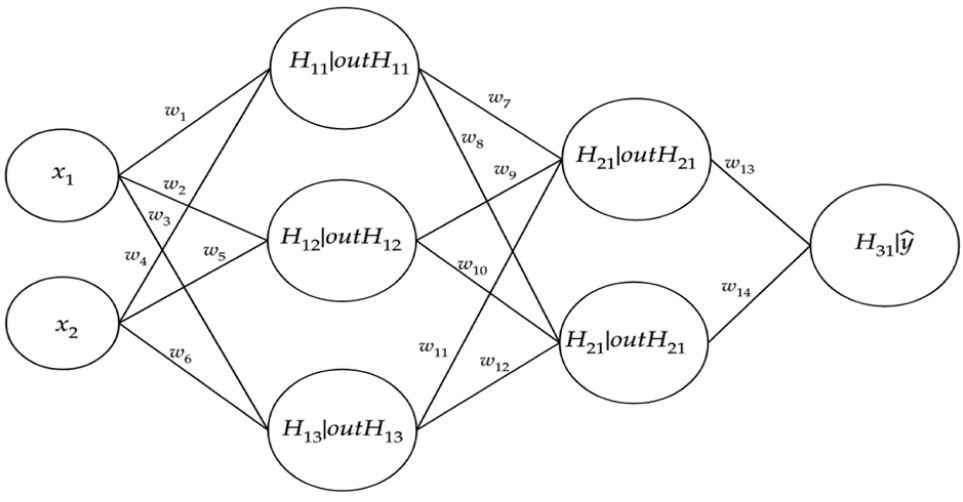
2.3 MODELOS DE APRENDIZADO DE MÁQUINA

Embora uma gama de modelos estocásticos tenha sido proposta para essa finalidade, a previsão da mortalidade futura a partir de tendências históricas permanece sendo um grande desafio. Nesse cenário, as técnicas de aprendizado de máquina têm ganhado destaque, assumindo um papel cada vez mais central em diversas áreas de pesquisa, incluindo a ciência atuarial, possuindo como principal contribuição sua capacidade de identificar padrões em dados que não são facilmente detectáveis ou identificáveis por métodos tradicionais, além de revelar correlações ocultas (Levantesi; Pizzorusso, 2019).

As redes neurais artificiais são grafos computacionais compostos por neurônios organizados em camadas sequenciais, cujas conexões são ponderadas por parâmetros aprendidos através de processos de otimização eficientes (Marino; Levantesi; Nigri, 2021). Essa característica é particularmente valiosa na previsão de séries temporais, onde relações complexas e dependências temporais desafiam métodos estatísticos tradicionais, frequentemente limitados por suposições de linearidade e estacionariedade (Zhang, 2003).

A figura a seguir mostra a representação de rede neural típica *feedfoward*. Cada nó do grafo representa um neurônio, conectado entre si por arcos que representam as sinapses. Os círculos representam neurônios e linhas representam sinapses. As sinapses pegam a entrada e a multiplicam por um "peso" (a "força" da entrada na determinação da saída). Os neurônios adicionam as saídas de todas as sinapses e aplicam uma função de ativação (Nigri; al., 2019).

Figura 1 – Representação de uma Rede Neural Típica



Fonte: Nigri e al. (2019)

Cada saída $H \in \mathbb{R}^{n_h}$ de uma camada oculta com n_h neurônios é definida:

$$H = \phi\left(W^T x + b\right) \tag{2.11}$$

onde $W \in \mathbb{R}^{d \times n_h}$ é uma matriz de peso e $b \in \mathbb{R}^{n_h}$ é um vetor de vieses. No esquema, a saída de uma camada oculta torna-se a entrada para a próxima camada. Considerando um problema de regressão, onde $n \in \mathbb{N}$ o número de camadas ocultas, a saída $\hat{y} \in \mathbb{R}$ é obtida por:

$$H_1 = \phi_1 \left(W_1^T x + b_1 \right) \tag{2.12}$$

$$H_2 = \phi_2 \left(W_2^T H_1 + b_2 \right) \tag{2.13}$$

$$\vdots$$

$$\hat{y} = \phi_n \left(W_n^T H_{n-1} + b_n \right) \tag{2.14}$$

onde $W_1, W_2, \ldots, W_n$ são as matrizes de pesos, $b_1, b_2, \ldots, b_n$ os vetores de vieses, e $\phi_1, \phi_2, \ldots, \phi_n$ as funções de ativação, que podem ser iguais. A dimensão das matrizes de peso e os vetores de viés dependem do número de unidades nas camadas ocultas. Uma observação crucial é que aumentar o número de camadas ocultas em uma rede neural leva a um maior nível de abstração dos dados de entrada (Nigri; al., 2019). Isso significa que redes mais profundas podem aprender características mais complexas e hierárquicas da entrada bruta, transformando-as em representações mais abstratas que são progressivamente mais úteis para a tarefa final. Cada camada oculta adicional permite que a rede se baseie nos recursos aprendidos pela camada anterior, criando uma compreensão mais rica e diferenciada dos dados.

Dentre as arquiteturas de redes neurais, as Redes Neurais Recorrentes (RNNs) se destacam naturalmente para dados sequenciais, como séries temporais. Isso ocorre porque as RNNs possuem uma memória interna, ou estado oculto, que é atualizada a cada passo de tempo, permitindo que a informação de entradas passadas influencie o processamento da entrada atual (Gupta *et al.*, 2019).

Estudos relatam que as variantes de redes neurais recorrentes (RNNs), incluindo suas variantes *Log Short-Term Memory* (LSTM) e *Gated Recurrent Unit* (GRU), geralmente superam os modelos tradicionais de previsão de mortalidade, como Lee-Carter, especialmente na captura de dependências temporais em dados de mortalidade (Chen; Khaliq, 2022; Petneházi; Gáll, 2019; Bravo; Santos, 2022). Redes neurais convolucionais (CNNs) também mostraram fortes capacidades preditivas em taxas de mortalidade, quando projetado especificamente para processar dados sequenciais (Perla *et al.*, 2020). Modelos híbridos, que combinam preditores aumentam ainda mais a precisão da previsão, aproveitando os pontos fortes complementares de diferentes arquiteturas (Li, 2022; Gyamerah *et al.*, 2023).

2.3.1 Redes Neurais Recorrentes (RNNs)

As redes neurais recorrentes sofrem com o problema de *vanishing gradient*, que dificulta o aprendizado de dependências de longo prazo em sequências. Para resolver isso, arquiteturas mais

complexas com mecanismos de portas (*gated units*) foram desenvolvidas. Esses mecanismos utilizam funções de ativação não lineares para regular dinamicamente o fluxo de informação. As funções sigmoidal (*sigmoid* e tangente hiperbólica (*tanh*) são as mais frequentemente empregadas para essa finalidade (Chen; Khaliq, 2022):

$$\text{Sigmóide: } sigmoid(x) = \frac{1}{1+e^{-x}} \tag{2.15}$$

$$\text{Tangente hiperbólica: } tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{2.16}$$

A função sigmoidal, cujos valores variam entre 0 e 1, é utilizada nas portas para decidir se a informação recebida deve ser atualizada e retida ou deve ser descartada e esquecida. A função de ativação tangente hiperbólica, cujos valores variam entre -1 e 1, é usada para regular e normalizar os valores que fluem através da rede, comprimindo-os para uma escala específica.

Redes neurais recorrentes do tipo LSTM operam como uma unidade de memória com portas (*gated memory unit*) dentro de redes neurais recorrentes. Seu núcleo funcional baseia-se em três mecanismos de controle: a porta de entrada (que regula quais informações atualizam a memória), a porta de esquecimento (responsável por descartar dados irrelevantes) e a porta de saída (que define o acesso ao estado interno). Essas portas, implementadas através de funções *sigmoid* gerenciam dinamicamente o conteúdo da célula de memória c_t e o estado oculto h_t, permitindo que a rede retenha dependências temporais de longo prazo. Pode ser descrita por cinco equações (Petneházi; Gáll, 2019):

$$i_t = sigmoid\left(W_i x_t + U_i h_{t-1} + b_i\right) \tag{2.17}$$

$$f_t = sigmoid\left(W_f x_t + U_f h_{t-1} + b_f\right) \tag{2.18}$$

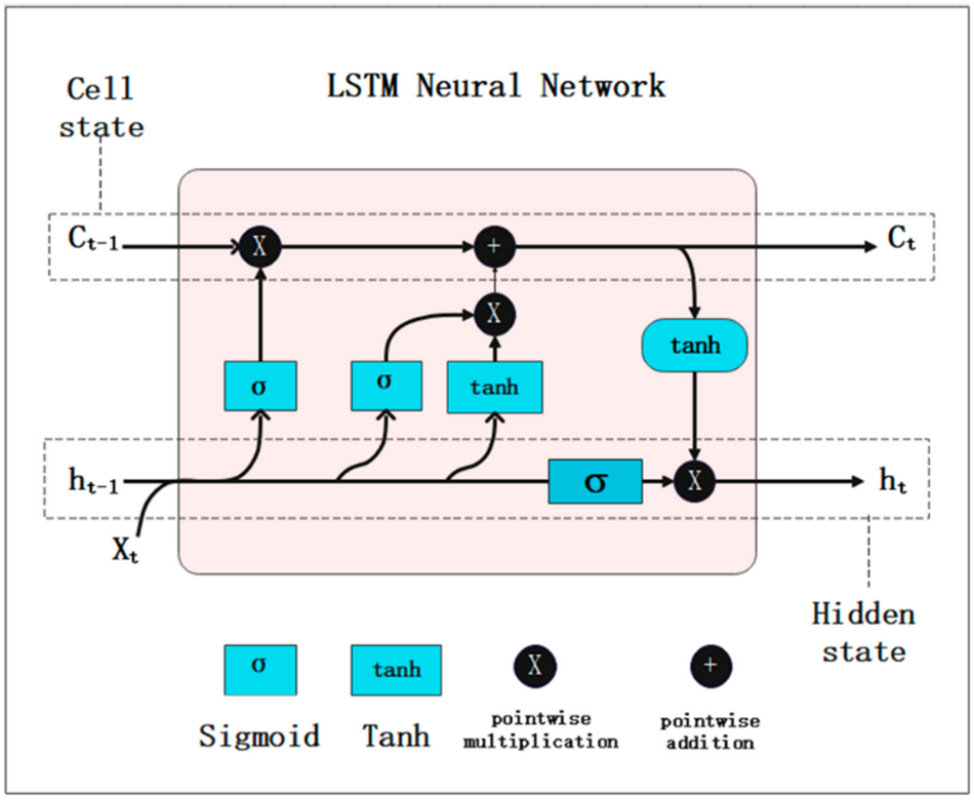
$$o_t = sigmoid\left(W_o x_t + U_o h_{t-1} + b_o\right) \tag{2.19}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot tanh\left(W_c x_t + U_c h_{t-1} + b_c\right) \tag{2.20}$$

$$h_t = o_t \odot tanh\left(c_t\right) \tag{2.21}$$

onde x_t são os dados inseridos na unidade em um intervalo de tempo t, h_t representa a saída no tempo t, ou estado oculto com informações que serão passadas para a etapa seguinte, c_t é o componente de memória principal do LSTM, que transporta informações em intervalos de tempo, i_t, f_t, o_t são vetores de porta, sendo que i_t determina o quanto de novos dados de entrada podem atualizar o estado da célula, f_t decide quais informações devem ser descartadas, o_t controla qual parte do estado da célula é exposta como o estado oculto. O operador $\odot$ denota a multiplicação elementos a elemento (também conhecida como produto de *Hadamard*), na qual cada valor do vetor à esquerda é multiplicado pelo valor correspondente no vetor à direita, permitindo que as portas ajam como filtros que regulam seletivamente a informação que flui pela célula. W, U, b são parâmetros dimensionais que o modelo aprende durante o processo de treinamento (Petneházi; Gáll, 2019).

Figura 2 – Estrutura da unidade LSTM



Fonte: Chen e Khaliq (2022)

A unidade recorrente fechada (GRU) é um tipo de rede neural recorrente semelhante às redes de memória de longo prazo (LSTM), mas como uma arquitetura simplificada, possuindo menos parâmetros, portas e equações, que podem levar a um treinamento com menos sobrecarga computacional, mas mantendo um bom desempenho de modelagem. As GRUs trabalham com as seguintes equações (Chen; Khaliq, 2022):

$$z_t = sigmoid\left(W_z x_t + U_z h_{t-1}\right) \tag{2.22}$$

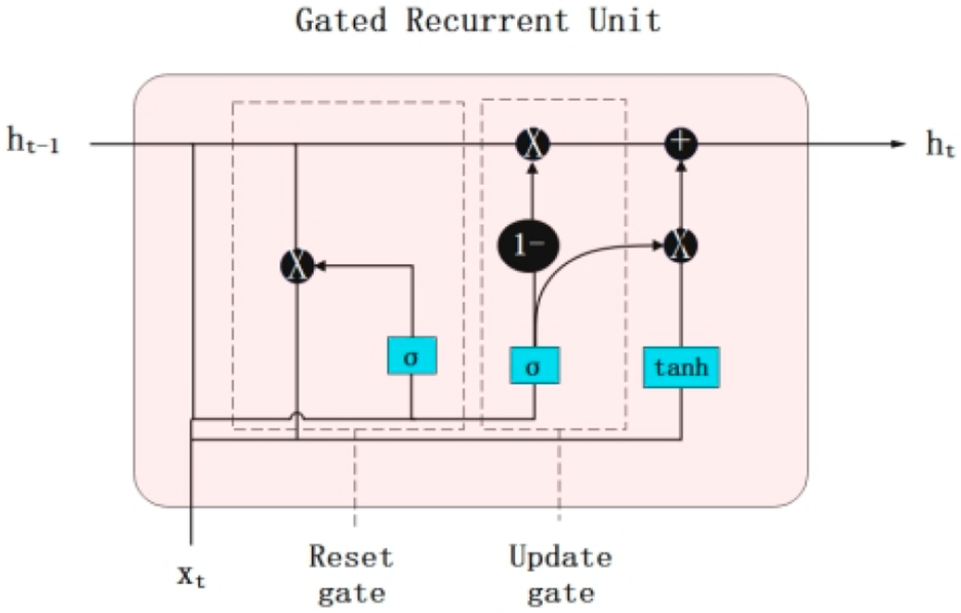
$$r_t = sigmoid\left(W_r x_t + U_r h_{t-1}\right) \tag{2.23}$$

$$\widetilde{h}_t = tanh\left(W_h x_t + U_h\left(r_t \odot h_{t-1}\right)\right) \tag{2.24}$$

$$h_t = z_t \odot \widetilde{h}_t + (1 - z_t) \odot h_{t-1} \tag{2.25}$$

onde h_t representa a informação de saída para a próxima unidade na etapa de tempo atual t, z_t indica o portão de atualização que controla a mistura entre o estado oculto anterior e o novo candidato, $\widetilde{h}_t$ é o estado da célula atual, calculado com base na entrada atual x_t e no estado oculto anterior h_{t-1}, r_t indica o portão de reinicialização que decide quanta informação anterior deve ser esquecida, W e U são as matrizes de pesos, ou seja, os parâmetros aprendidos que serão usados no cálculo dos valores e estados (Chen; Khaliq, 2022).

Figura 3 – Estrutura da unidade GRU



Fonte: Chen e Khaliq (2022)

2.3.2 Redes Neurais Convolucionais (CNNs)

As Redes Neurais Convolucionais (CNNs), originalmente projetadas para processamento de imagens, adaptaram-se eficientemente à análise de séries temporais, como as taxas de mortalidade (Perla *et al.*, 2020; Schnürch; Korn, 2021). Sua estrutura explora padrões locais em dados sequenciais por meio de operações de convolução, que filtram informações relevantes sem a necessidade de conexões densas entre todas as camadas. Essa característica é particularmente vantajosa para dados atuariais, que frequentemente exibem tendências temporais e sazonalidades complexas. Uma camada convolucional *1D* transforma uma série temporal multivariada de entrada $x_t \in R^d$, onde $t = 0,...,T$ e d são as dimensões do vetor a cada tempo, em um novo conjunto de características. Para q núcleos de tamanho m, a operação é definida por (Perla *et al.*, 2020):

$$z : R^{1\times(T+1)\times d} \to R^{1\times(T+2-m)\times q}, x = (x_0,\ldots,x_T) \to z(x) = \left(z_{s,j}(x)\right)_{0\le s\le T+1-m, 1\le j\le q} \tag{2.26}$$

onde cada componente $z_{s,j}$ é calculado como:

$$x \to z_{s,j} = z_{s,j}(x) = \phi\left(w_{j,0} + \left(W^{(j)} \times x\right)_s\right) = \phi\left(w_{j,0} + \sum_{t=1}^{m} \left\langle w_{j,l}, x_{\{m+s-1\}}\right\rangle\right) \tag{2.27}$$

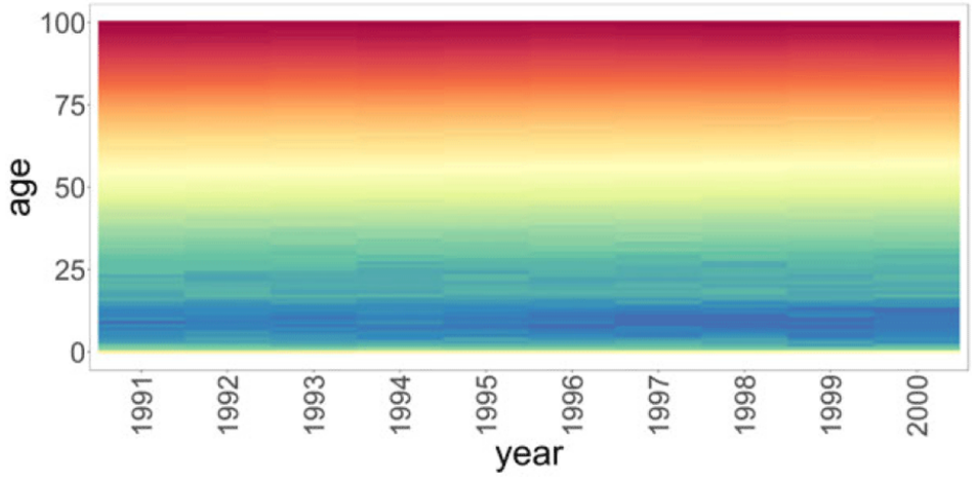
onde ϕ é uma função de ativação, $w_{j,0}$ é o viés do j-ésimo filtro, $w_{j,l}$ são os pesos do filtro j na posição l. A operação de somatória corresponde à convolução discreta entre o filtro W e a janela

deslizante de tamanho *m* da série *x* (Perla *et al.*, 2020).

Perla *et al.* (2020) ainda explica que as aplicações das camadas da CNN são eventualmente seguidas por camadas de *pooling*, cuja função principal é agregar características de camadas de alta dimensão para gerar uma estrutura de características de menor dimensão. Eles realizam operações simples e predefinidas, como selecionar o valor máximo, mínimo ou médio dentro de uma área específica, em vez de aprender parâmetros para convolução. O objetivo principal dessas camadas é "extrair e compactar informações" de camadas de alta dimensão, tornando-as mais valiosas para a tarefa de previsão.

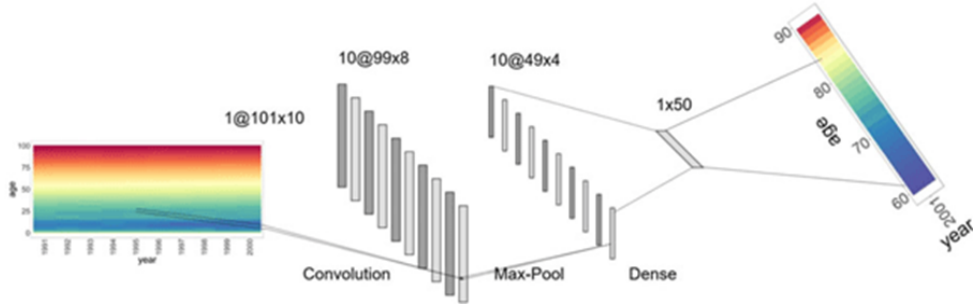
A estrutura das Redes Neurais Convolucionais (CNNs) é composta por camadas especializadas que processam dados complexos de forma sequencial, como a matriz idade-tempo de mortalidade ilustrada na Figura 5, os dados de entrada, neste caso, taxas de mortalidade log-transformadas de mulheres inglesas e galesas (ilustradas em um mapa de calor) são organizados em formatos matriciais onde padrões espaciais são essenciais. As camadas convolucionais atuam como detectores iniciais, identificando características locais como tendências etárias ou efeitos de coorte. Em seguida, as camadas de *pooling* condensam essas informações, reduzindo redundâncias e destacando os sinais dominantes. Por fim, camadas densas integram os padrões extraídos para gerar previsões. A Figura 6 exemplifica essa arquitetura hierárquica, mostrando o fluxo desde a entrada (1x101x10) até as camadas convolucionais (10x99x8) e de *pooling* (10x49x4), culminando nas camadas densas para saída. Esse design permite que a CNN atue como um extrator de características automatizado, transformando dados brutos em padrões significativos para predição (Schnürch; Korn, 2021).

Figura 4 – Mapa de calor das taxas de mortalidade para entrada CNN



Fonte: Schnürch e Korn (2021)

Figura 5 – Arquitetura CNN



Fonte: Schnürch e Korn (2021)

2.4 TRANSFERÊNCIA DE APRENDIZADO

O aprendizado por transferência é definido como o processo de aproveitar o conhecimento de um domínio de origem para melhorar o aprendizado em um domínio de destino com dados limitados (Pan; Yang, 2010). Em séries temporais de mortalidade emergiu como uma área crítica de investigação devido ao seu potencial de aumentar a precisão preditiva em ambientes com disponibilidade limitada de dados e dinâmica temporal complexa (Nalmpatian *et al.*, 2025; Gupta *et al.*, 2019). Modelos, por exemplo, de redes neurais profundas, embora promissoras para tarefas de predição de informações clínicas, inclusive de mortalidade, requerem dados substanciais, ajustes significativos nos hiperparâmetros e considerável poder computacional para treinamento, podendo levar ao sobreajuste quando há poucos dados (Gupta *et al.*, 2019). O aprendizado por transferência oferece uma maneira de mitigar esses problemas transferindo conhecimento de um modelo treinado em dados amplos de origem, para uma tarefa de destino relacionada com dados limitados, tornando o ajuste fino de modelos pré-treinados mais rápido e ajustado (Gupta *et al.*, 2019).

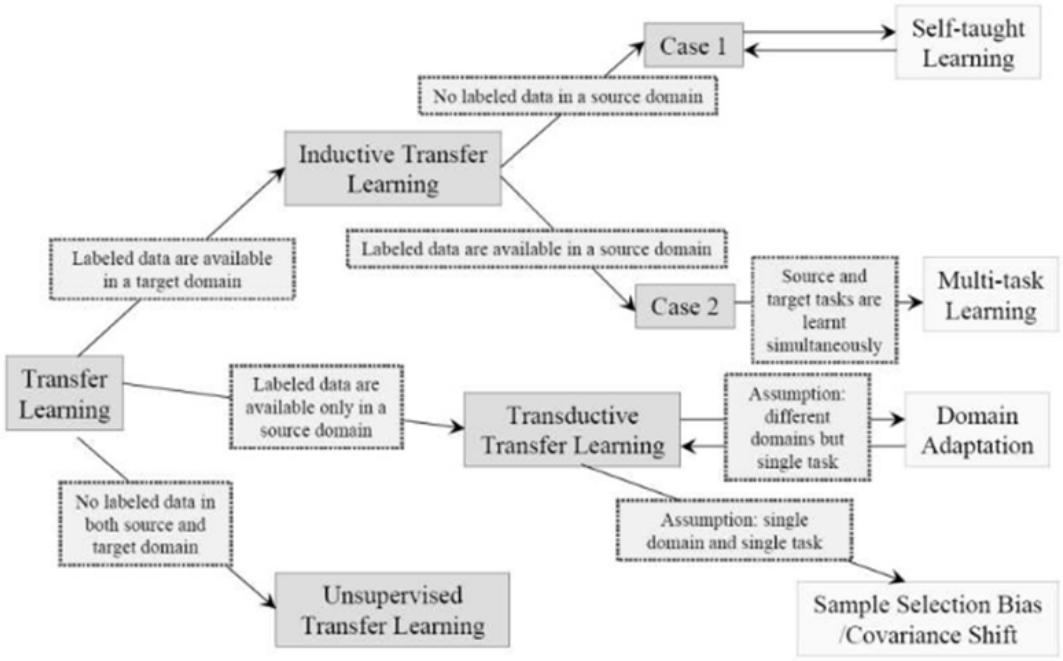
Segundo Pan e Yang (2010), o aprendizado por transferência é categorizado em três configurações principais com base na disponibilidade de dados rotulados e na relação entre domínios de origem e de destino:

i) aprendizado por transferência indutiva: nessa categoria, a tarefa de destino difere da tarefa de origem e alguns dados rotulados são necessários no domínio de destino. Isso pode envolver casos em que muitos dados rotulados estão disponíveis no domínio de origem ou em que nenhum dado rotulado está disponível no domínio de destino;

ii) aprendizado por transferência transdutiva: aqui, as tarefas de origem e de destino são as mesmas, mas seus domínios são diferentes. Nenhum dado rotulado está disponível no domínio de destino, mas uma quantidade significativa de dados rotulados está disponível no domínio de origem. Essa configuração está relacionada à adaptação do domínio e ao viés de seleção da amostra; e

iii) aprendizado por transferência não supervisionada: essa configuração se concentra em tarefas de aprendizado não supervisionado no domínio de destino (por exemplo, agrupamento, redução de dimensionalidade) em que a tarefa de destino é diferente, mas relacionada à tarefa de origem. Nenhum dado rotulado está disponível nos domínios de origem ou de destino para treinamento.

Figura 6 – Visão geral de diferentes configurações de transferência



Fonte: Pan e Yang (2010)

Diante das dessas três configurações, Pan e Yang (2010) identificam quatro abordagens principais de transferência de aprendizado:

i) transferência de instância: faz novas ponderações de partes dos dados de origem para torná-los reutilizáveis para aprendizado no domínio de destino. Técnicas como reponderação de instâncias e amostragem por importância são usadas;

ii) representação-transferência de funcionalidades: visa aprender uma boa representação de recursos que minimize a diferença entre os domínios de origem e de destino e reduza o erro do modelo. O conhecimento é codificado na representação da característica aprendida. Isso pode ser feito por meio da construção de recursos supervisionada ou não supervisionada;

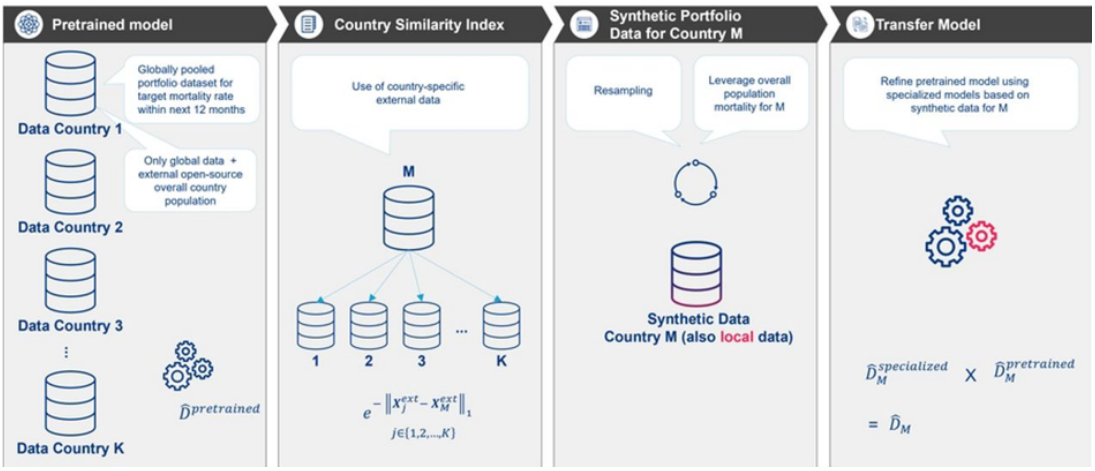
iii) transferência de paramétricos: pressupõe que as tarefas de origem e de destino compartilhem alguns parâmetros ou distribuições anteriores de hiperparâmetros. O conhecimento transferido é codificado nesses parâmetros compartilhados ou anteriores; e

iv) transferência de conhecimento relacional: ocorre em domínios relacionais em que os dados não são independentes e distribuídos de forma idêntica e podem ser representados por várias relações. Ele se concentra na transferência de relacionamentos entre os dados da origem para o domínio de destino.

Estudos recentes desenvolveram ou aplicaram estruturas teóricas específicas para motivar a aprendizagem por transferência em séries temporais de mortalidade, incluindo adaptação de domínio, aprendizagem multitarefa, modelos hierárquicos e abordagem de mudanças distribucionais e deriva posterior por meio de ajustes lineares ou conjuntos de reforço (Nalmpatian *et al.*, 2025; Liu; al., 2021; Li; Li; Panagiotelis, 2024). A adaptabilidade do modelo entre populações ou distribuições de dados foi demonstrada, sendo que estruturas de aprendizagem por transferência podem generalizar bem com técnicas apropriadas de adaptação de domínio (Nalmpatian *et al.*, 2025). Grandes conjuntos de dados disponíveis publicamente, como *Human Mortality Database*, foram utilizados, permitindo reprodutibilidade e *benchmarking* (Diao *et al.*, 2023; Perla *et al.*, 2020). A inclusão de fontes de dados multi-institucionais e multinacionais aumenta a representatividade dos domínios de origem e apoia a transferência interpopulacional (Nalmpatian *et al.*, 2025). Alguns estudos incorporam modelos de mortalidade específicos do domínio e princípios da ciência atuarial para melhorar a interpretabilidade e a relevância do modelo (Nalmpatian *et al.*, 2025; Li; Li; Panagiotelis, 2024).

A Figura 7 ilustra a metodologia proposta por Nalmpatian *et al.* (2025) para adaptação de modelos pré-treinados em análises de mortalidade, combinando dados globais, sintéticos e índices de similaridade entre países. De acordo com as classificações propostas por Pan e Yang (2010) trata-se de uma aprendizagem por transferência indutiva, uma vez que o modelo aprende com dados rotulados nos domínios de origem (países com dados de mortalidade disponíveis) e, em seguida, aplica esse conhecimento aprendido para melhorar o desempenho em uma tarefa alvo diferente, mas relacionada. A abordagem de transferência de aprendizado utilizada por Nalmpatian *et al.* (2025) pode ser classificada principalmente como um método transferência de paramétricos, combinado com elementos de transferência de instâncias (por meio de geração de dados sintéticos) e transferência de representação.

Figura 7 – Ilustração de segmentos populacionais-alvo em diferentes conjuntos de dados e modelos



Fonte: Nalmpatian *et al.* (2025)

A ilustração descreve um *pipeline* para adaptação de modelos de mortalidade a contextos locais, seguindo a abordagem de Nalmpatian *et al.* (2025). Inicialmente, um modelo global (q) é treinado com dados agregados de múltiplos países (*Country* 1, 2, 3), enquanto modelos locais h_j refinam as previsões usando características específicas de cada país. Um índice de similaridade, calculado pela métrica $e^{-\left\|X_j^{ext} - X_m^{ext}\right\|}$ guia a reamostragem proporcional de exposições para gerar um *dataset* sintético $\hat{D}_M$ para o país-alvo M. Esse *dataset* incorpora tanto *features* globais quanto ajustes locais, além de substituir a mortalidade pela do país M, assegurando contextualização. Por fim, as previsões são obtidas combinando as saídas dos modelos global e locais, resultando em uma estimativa contextualizada, alinhada às condições demográficas e econômicas de M.

2.5 COMBINAÇÃO DE PREDITORES

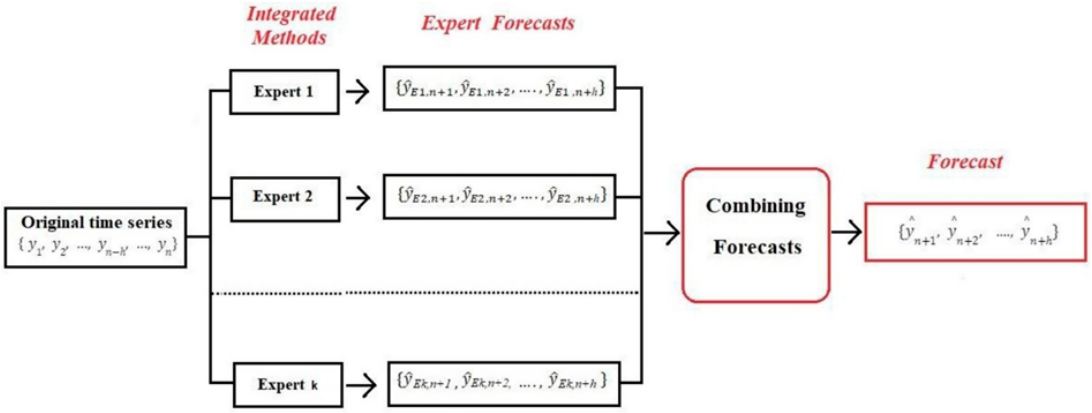
A combinação de modelos surge como estratégia fundamental para superar limitações de abordagens únicas em projeções de mortalidade, equilibrando viés e variância. Estudos detectaram que a combinação de modelos, por diferentes técnicas, como por meio de empilhamento, métodos bayesianos ou abordagens híbridas, melhora significativamente a precisão da previsão em comparação com modelos individuais, com ganhos variando de 13% a mais de 90% em alguns casos (Kessy *et al.*, 2021; Bimonte *et al.*, 2024). Combinações podem mitigar o risco de seleção de modelos e podem adaptar os pesos dinamicamente em todos os horizontes de previsão (Kessy *et al.*, 2021). Por exemplo, conjuntos bayesianos fornecem intervalos confiáveis e coerentes que levam em conta a incerteza dos parâmetros e do modelo (Bravo; Ayuso, 2020). Modelos híbridos que integram métodos clássicos e de aprendizado de máquina mostram maior precisão em configurações de dados limitadas (Gyamerah *et al.*, 2023; Duarte; Neto; Firmino, 2024).

A combinação de preditores na previsão de séries temporais envolve basicamente

três estágios: geração, seleção e agregação, cada um desempenhando um papel crucial no aprimoramento da precisão da previsão (García-Aroca *et al.*, 2023). A primeira etapa, de geração, cria um conjunto de modelos, sendo importante que estes tenham diversidade (Valentini; Dietterich, 2002). A etapa seguinte envolve a seleção, por algum critério, dos modelos que irão compor a combinação. Pode-se escolher apenas um modelo, definindo-se como seleção de preditor, sendo que caso se escolha mais de um modelo, denomina-se seleção de *Ensemble* (Silva, 2021).

A seleção também pode acontecer de forma estática ou dinâmica, sendo que a primeira utiliza um ou mais modelos para prever todo o conjunto de teste, e a segunda usa para cada padrão de teste (Júnior, 2022). No processo de seleção os métodos podem ser ordenados com base na proximidade com o 'melhor' modelo, e utilizado um ponto de corte, para selecionar quantos métodos serão combinados (García-Aroca *et al.*, 2023). A última etapa, de agregação, que é a fase responsável por integrar os modelos selecionados e criar a previsão final (Júnior, 2022). Diferentes podem ser as técnicas de agregação, seja por aplicação de medidas de tendência central ou atribuição de pesos ponderados baseados em desempenho nos conjuntos de treino e teste (García-Aroca *et al.*, 2023).

Figura 8 – Combinação de previsões por métodos especialistas



Fonte: García-Aroca *et al.* (2023)

2.5.1 Ponderação pelo Inverso do Erro

A combinação de preditores com pesos atribuídos conforme métricas de erro configura uma estratégia robusta para aprimorar a acurácia de modelos de séries temporais, conforme demonstrado por García-Aroca *et al.* (2023). No algoritmo *AlpCA*, proposto pelos autores, a ponderação dos modelos individuais fundamenta-se no inverso dos erros observados nas fases de treinamento ($T1$) e validação ($T2$), priorizando modelos com menor discrepância preditiva (García-Aroca *et al.*, 2023).

O cerne da metodologia do *AlpCA* para atribuição de pesos reside na análise das métricas de erro obtidas nas fases de treinamento e validação, especialmente através da Análise de

Componentes Principais (PCA). O algoritmo executa um PCA na matriz de erros, que contêm várias funções de perda: erro percentual absoluto (*sMAPE*) e erro médio absoluto em escala (*MASE*) para T1, e *sMAPE*, *MASE*, raiz do erro quadrático médio (*RMSE*) e média ponderada geral (*OWA*) para T2. O objetivo do PCA é reduzir a dimensionalidade dessa matriz de erros para um ou dois componentes principais (CP1 ou CP1 e CP2), que capturam a maior parte da variabilidade dos erros. Esses componentes principais fornecem uma 'pontuação de erro' para cada método de previsão, que é uma combinação linear das variáveis de erro originais (García-Aroca *et al.*, 2023). A seleção dos pesos considera três alternativas, todas baseadas no princípio da ponderação inversa ao erro:

i) Peso w_1: derivados do primeiro componente principal (CP1) da matriz de erros, refletindo a contribuição marginal de cada modelo na variabilidade total dos erros. Modelos com menores escores absolutos no CP1 recebem maior ponderação:

$$w_{1,i} = \frac{|x_j|}{\sum_{j=1}^{m} |x_j|} \tag{2.28}$$

ii) Peso w_2: atribuídos conforme o erro *sMAPE* na fase de treinamento (T1), privilegiando modelos com melhor ajuste histórico:

$$w_{2,i} = \frac{\frac{1}{sMAPE_{T_{1j}}}}{\sum_{j=1}^{m} \frac{1}{sMAPE_{T_{1j}}}} \tag{2.29}$$

iii) Peso w_3: semelhante a w_2, mas utilizam o erro *sMAPE* da fase de validação (T2), alinhando-se ao horizonte preditivo:

$$w_{3,i} = \frac{\frac{1}{sMAPE_{T_{2j}}}}{\sum_{j=1}^{m} \frac{1}{sMAPE_{T_{2j}}}} \tag{2.30}$$

A ordenação dos modelos baseia-se na distância de *Manhattan* entre suas métricas e o "melhor especialista", seguida de seleção via distribuições paramétricas (Exponencial, Gama). A escolha entre w_1, w_2 e w_3 depende do contexto: enquanto w_2 e w_3 destacam-se em horizontes curtos, w_1 mostra robustez em cenários com maior heterogeneidade de erros (García-Aroca *et al.*, 2023).

Em séries temporais reduzidas, com poucas observações, a combinação mitiga riscos de sobreajuste e incerteza paramétrica através de estratégias como validação *leave-future-out* (Barigou *et al.*, 2022). Bravo e Ayuso (2020) comprovaram sua eficácia em dados portugueses. Modelos hierárquicos (Nalmpatian *et al.*, 2024) e estruturas bayesianas (Berkum; Antonio; Vellekoop, 2017) emergem como soluções, atendendo à necessidade de quantificação de incertezas em contextos regulatórios.

2.6 CONSIDERAÇÕES DO CAPÍTULO

Este capítulo revisou criticamente as principais abordagens para projeção de mortalidade, com ênfase nos desafios impostos por séries temporais curtas. Inicialmente, discutiu-se a relevância atuarial das projeções e as limitações inerentes a bases de dados reduzidas, como sobreajuste e instabilidade. Em seguida, analisaram-se modelos tradicionais Lee-Carter, ARIMA e ETS, destacando suas aplicações e restrições em contextos de escassez amostral. Avançando para técnicas contemporâneas, explorou-se o potencial das redes neurais recorrentes e convolucionais, na ordem LSTM, GRU e CNN e estratégias inovadoras como transferência de aprendizado, usada para extrair conhecimento de domínios com dados abundantes, e combinação de preditores, focando em abordagens com ponderação pelo inverso das métricas de erro e por combinação bayesiana, que mitigam incertezas mediante integração de múltiplos modelos. Esta fundamentação teórica norteará a proposta metodológica apresentada no próximo capítulo, que visa aplicar essas estratégias no contexto específico das projeções de mortalidade brasileira. O Quadro 2.4 sintetiza os modelos descritos nesta revisão de literatura, destacando seus objetivos, forças, limitações, razões para uso e evidências já publicadas.

Quadro 2.4 – *Quadro Resumo da Revisão de Literatura*

Modelo	Objetivo	Forças	Limitações	Por que usar	Evidências
Lee-Carter	Tendência idade–tempo via k_t	Simples, interpretável	Instável em séries curtas; linearidade de k_t	Baseline clássico; comparador	Lee e Carter (1992), McQuire e Kume (2022), Andrs-Sánchez e Puchades (2019)
FDM	Curvas funcionais idade-tempo	Suavização robusta; múltiplos modos de variação; coerência multipopulações	Dependência da qualidade dos dados; custo computacional; risco de sobreajuste; limitação para dinâmicas não lineares	Generalização robusta do LC; melhor com variabilidade complexa ou séries curtas	Hyndman e Ullah (2007), Hyndman, Booth e Yasmeen (2013), Shang (2016), Yap, Pathmanathan e Dabo-Niang (2025)
ARIMA	Dinâmica temporal univariada	Bom curto prazo; probabilístico	Exige estacionariedade; ignora estrutura etária	Prever k_t; benchmark temporal	Gujarati e Porter (2011), Nor *et al.* (2024), Ifeanyichukwu *et al.* (2022)
ETS	Suavização com nível, tendência e sazonalidade	Rápido e acurado; robusto	Pode quebrar coerência por idade	Baseline forte; útil em curto–médio prazo	Hyndman e Athanasopoulos (2021), Feng e Shi (2018), Shi (2020)
LSTM/GRU	Dependências longas e não lineares	Capturam padrões complexos	Demandam dados; risco de *overfitting*	Ganhos quando há mais dados	Chen e Khaliq (2022), Petneházi e Gáll (2019), Bravo e Santos (2022)
CNN	Padrões idade–tempo locais	Extração automática; *pooling*	Menos interpretável	Capturar efeitos de coorte/idade	Perla *et al.* (2020), Schnürch e Korn (2021)
Transfer Learning	Adaptar de fonte rica a alvo escasso	Mitiga escassez; acelera treino	*Shift* de domínio	Essencial em séries curtas	Pan e Yang (2010), Nalmpatian *et al.* (2025), Gupta *et al.* (2019)
Combinação (Inverso do Erro)	Agregar por desempenho	Ganhos consistentes; simples	Curadoria de métricas	Robustez em curto e heterogeneidade	García-Aroca *et al.* (2023), Kessy *et al.* (2021)

Fonte: Elaboração própria

3 METODOLOGIA

Este capítulo expõe a estrutura metodológica que orientou a investigação, detalhando os procedimentos adotados para alcance dos objetivos. A organização divide-se em: a descrição dos Sistemas Propostos (Seção 3.1), com subseções dedicadas às abordagens a serem implementadas de transferência de aprendizado e à combinação estratégica de preditores; e o Protocolo (Seção 3.2), estruturado em três etapas: seleção e tratamento das Bases de Dados (3.2.1), operacionalização das técnicas de Previsão das Taxas de Mortalidade (3.2.2) e definição das Métricas de Avaliação (3.2.3) para análise comparativa. As Considerações do Capítulo (Seção 3.3) sintetizam, por fim, o arcabouço metodológico e seus reflexos na robustez do estudo.

3.1 SISTEMAS PROPOSTOS

Esta seção apresenta os sistemas propostos para esta pesquisa e as suas respectivas etapas de execução. A subseção 3.1.1 apresenta o método de Transferência de Aprendizado. A subseção 3.1.2 detalha os métodos de combinação de preditores.

3.1.1 Transferência de Aprendizado

O *Human Mortality Database* (HMD) reúne dados detalhados de mortalidade de diversos países, muitos com séries históricas significativamente mais extensas que as disponíveis para o Brasil. Esta disparidade temporal motivou a adoção de uma estratégia de transferência de aprendizado: modelos treinados com dados internacionais de maior profundidade histórica foram adaptados para suprir a limitação da série brasileira, mais curta. Contudo, para evitar que padrões demográficos discrepantes introduzissem ruídos nas previsões, foi implementado um processo de seleção de países.

A seleção priorizou países com comportamentos de mortalidade semelhantes aos brasileiros, seguindo a premissa de que contextos socioeconômicos e epidemiológicos análogos geram padrões transferíveis com maior eficácia. A similaridade foi mensurada através de métrica de distância de *Manhattan* aplicada às taxas de mortalidade logarítmicas específicas por idade, nos anos comuns entre as séries históricas do HMD e os dados brasileiros. A escolha da distância de *Manhattan* se alinha ao trabalho de Nalmpatian *et al.* (2025) e ao fato de possuir baixo custo computacional, exprimindo a noção de similaridade de forma simples, gerando resultados aceitáveis quando da impossibilidade de se desenvolver um estudo empírico para determinar a melhor medida (Parmezan, 2016; Junior, 2012). Considerando que os comportamentos das taxas de mortalidade podem diferenciar por sexo, essa avaliação de similaridade foi realizada para cada

conjunto de dados separadamente. A distância foi calculada por meio da expressão:

$$d(c) = \frac{1}{|T| * |X|} \sum_{t \in T} \sum_{x \in X} \left| log\left(m_{c,t,x}\right) - log\left(m_{br,t,x}\right) \right| \tag{3.1}$$

onde X representa as faixas etárias analisadas, T os anos a serem comparados e $m_{c,t,x}$ a taxa de mortalidade do país c, no ano t e idade x, seguindo a abordagem de (Nalmpatian *et al.*, 2025). Com a finalidade de aumentar a base de treinamento, contudo, sem contaminá-la com dados ruidosos, foram selecionados os dez países com menores valores de $d(c)$, recebendo pesos diferenciados por meio de uma função exponencial:

$$w(c) = \frac{exp\left(-\alpha * d(c)\right)}{\sum_{k=1}^{10} exp\left(-\alpha * d(k)\right)} \tag{3.2}$$

A função exponencial de ponderação atribui diferentes níveis de influência aos países, controlados pelo parâmetro α. Este parâmetro define a taxa de decaimento dos pesos conforme aumenta a distância do contexto de cada país em relação ao Brasil. Valores elevados de α resultam em um decaimento abrupto, restringindo a influência apenas aos países muito similares ao Brasil. Valores baixos de α permitem um decaimento mais lento, atribuindo pesos significativos mesmo a países moderadamente distintos. Esta seletividade, alinhada à proposta de Nalmpatian *et al.* (2025), atua como um filtro contra a contaminação por ruídos de contextos dissimilares, buscando garantir que apenas padrões compatíveis com os dados brasileiros alimentem o treinamento.

Os dados dos países selecionados foram treinados com três arquiteturas de aprendizado profundo: rede neural recorrente do tipo GRU, para capturar tendências temporais de longo prazo (Bravo; Santos, 2022), redes convolucionais, para identificar padrões etários locais (Perla *et al.*, 2020; Schnürch; Korn, 2021), e um modelo híbrido combinando ambas. Posteriormente, um *fine-tuning* reajustou cada modelo com dados brasileiros, congelando os pesos das camadas iniciais, que codificam padrões gerais, e otimizando apenas camadas finais com taxa de aprendizado reduzida.

Portanto, a transferência de aprendizado proposta seguiu uma abordagem indutiva (Zhuang *et al.*, 2021; Emmert-Streib; Dehmer, 2022). A distinção entre as tarefas nos domínios de origem, previsão de mortalidade em diversos países, e de destino, previsão específica para o Brasil, justificou essa classificação, especialmente devido à utilização de dados rotulados no contexto-alvo. A metodologia adotou um ajuste fino supervisionado sobre os dados brasileiros, realizado após um pré-treinamento em países com perfis semelhantes. Essa estratégia foi paramétrica, pois envolveu a reutilização e ajuste de parâmetros previamente aprendidos.

A eficácia do ajuste como mecanismo de viés indutivo explícito foi fundamentada em Li, Grandvalet e Davoine (2018), que destacaram a importância de manter a coerência entre o modelo original e o novo domínio por meio de técnicas de regularização. Além disso, Silver, Poirier e Currie (2008) reforçaram que o sucesso da transferência de aprendizado dependerá essencialmente do compartilhamento de estruturas de representação entre tarefas relacionadas,

mesmo que distintas, como ocorre ao aplicar um modelo treinado em dados do HMD ao contexto brasileiro. Dessa forma, o pipeline proposto alinhou-se com a definição clássica de transferência indutiva paramétrica, na qual o conhecimento foi adaptado a partir de parâmetros já otimizados para uma nova tarefa.

As previsões para períodos futuros foram geradas sequencialmente, adotando uma estratégia de previsão vários passos à frente do tipo recursiva (ou iterativa). Nesta abordagem, a estimativa de cada ano serviu de insumo para prever o seguinte, replicando o mecanismo de *forecasting* autorregressivo (Taieb *et al.*, 2012; Cheng *et al.*, 2006). Apesar de poder sofrer com acúmulo de erros no horizonte de previsão, dependendo do grau de ruído, a abordagem tem sido usada com sucesso na previsão de séries temporais por modelos de aprendizagem de máquina, inclusive por redes neurais recorrentes (Taieb *et al.*, 2012).

3.1.2 Combinação de Preditores

Esta pesquisa avaliou duas abordagens de combinação de preditores, a ponderação pelo inverso dos erros preditivos e a combinação pela média simples. A primeira técnica atribui pesos aos modelos individuais de forma inversamente proporcional aos seus erros de previsão, privilegiando aqueles com maior precisão pontual. Já a segunda, fundamentada na abordagem de conjunto, atribuindo pesos iguais a todos os modelos únicos na agregação.

A metodologia de combinação de preditores pelo inverso dos erros proposta seguiu o raciocínio de ponderação sugerida no algoritmo *AlpCA* (García-Aroca *et al.*, 2023), com a verificação inicial do desempenho individual dos modelos em um conjunto de validação e uma atribuição de pesos inversamente proporcionais a uma matriz normalizada dos valores de erros. Assim, durante o período de validação, foram observadas três métricas críticas para cada modelo: a raiz quadrada do erro quadrático médio (*RMSE*), que quantificou a dispersão dos erros, o erro absoluto médio (*MAE*) que capturou a magnitude média dos desvios; e o erro percentual absoluto médio (*sMAPE*), que contextualizou os erros em termos percentuais. Estas métricas passaram por um processo de normalização do tipo *min-max*, onde cada conjunto de métricas foram transformados para a escala [0,1], com a finalizada de eliminar assimetrias dimensionais permitindo a comparação isonômica entre os modelos. A operação de minimização é dada por:

$$e_{kj} = \frac{e_{k,j} - min\,(e_k)}{max\,(e_k) - min\,(e_k)} \tag{3.3}$$

O processo sintetizou o desempenho multidimensional através do erro composto, atribuindo pesos de ponderação diferenciada conforme o desempenho relativo de cada método no conjunto. Na ordem, as equações que sintetizam o desempenho no erro composto e a forma de atribuição de pesos:

$$E_j = \frac{1}{3}\left(e_{j,RMSE} + e_{j,MAE} + e_{j,sMAPE}\right) \tag{3.4}$$

$$w_j^{pond} = \frac{\frac{1}{E_j}}{\sum_{j=1}^{J} \frac{1}{E_j}} \tag{3.5}$$

A previsão combinada final $\tilde{y}_{T+h|T}$ pelo método de ponderação pelo inverso dos erros é obtida pela agregação das previsões individuais ponderadas pelos respectivos pesos w_j^{pond}:

$$\tilde{y}_{T+h|T} = \sum_{j=1}^{J} w_j^{pond} \cdot \hat{y}_{T+h|T,j} \tag{3.6}$$

onde J é o número total de modelos combinados, $\hat{y}_{T+h|T}$ representa a previsão h-passos à frente do modelo j com as informações até o período T, e (w_j^{pond} são os pesos normalizados calculados conforme a equação anterior.

A combinação de modelos para previsão, utilizando o método de média simples com pesos iguais (também conhecido como "*simple average*"ou "*equal weights*"), é uma abordagem base que consiste em agregar múltiplas previsões individuais de forma que cada modelo contribua igualmente para o resultado final. A previsão combinada $\tilde{y}_{T+h|T}$ usando o método de média simples com pesos iguais é dada por:

$$\tilde{y}_{T+h|T} = \frac{1}{N} \sum_{i=1}^{N} \hat{y}_{T+h|T,i} \tag{3.7}$$

onde N é o número de modelos ou previsões individuais, $\hat{y}_{T+h|T,i}$ é a previsão h-passos à frente do modelo i, baseada em informações até o tempo T. Essa fórmula atribui peso $w_i = \frac{1}{N}$ a cada previsão, somando 1 no total.

A notável eficácia desta técnica, que frequentemente supera métodos mais complexos em estudos empíricos, é um fenômeno conhecido como "*forecast combination puzzle*". Claeskens *et al.* (2016) oferecem uma explicação teórica para este fato, argumentando que a simplicidade do método o torna robusto ao evitar o erro de estimação inerente à otimização de pesos, o que muitas vezes leva ao *overfitting* e a um fraco desempenho fora da amostra. Os autores apoiam a ideia de que médias simples, especificamente a média aritmética, podem fornecer um bom desempenho na combinação de previsões, muitas vezes superando métodos mais complexos que dependem de pesos ideais estimados.

A implementação empregou o ecossistema *Python* com ênfase em: bibliotecas de cálculo científico: *NumPy* (Harris *et al.*, 2020) para operações vetoriais e estruturas de dados da biblioteca *pandas* (The Pandas Development Team, 2020) para gestão de séries temporais. Todo o processo de validação ocorreu em ambiente controlado, com separação temporal estrita entre conjuntos de treino, validação e teste, assegurando avaliação imparcial do desempenho.

3.2 PROTOCOLO EXPERIMENTAL

Esta seção apresenta os procedimentos que serão realizados para alcançar os objetivos do trabalho e está estruturado da seguinte forma: A subseção 3.2.1 apresenta as bases de dados e as fases de pré-processamento necessárias para estruturação dos dados. A subseção 3.2.2 esclarece os métodos de modelagem que serão aplicados, bem como a divisão dos dados em períodos de treino, validação e teste. A subseção 3.2.3 indica as métricas de avaliação que serão observadas para comparação de desempenho dos modelos.

3.2.1 Bases de Dados

Os dados brasileiros foram obtidos diretamente das tábuas de mortalidade anuais publicadas pelo Instituto Brasileiro de Geografia e Estatística (Instituto Brasileiro de Geografia e Estatística (IBGE), 2025), disponíveis em seu portal oficial de projeções populacionais. Esta seção tem como objetivo analisar a evolução histórica das taxas centrais de mortalidade por idade (nMx) no Brasil, fundamentando-se nessas séries de dados oficiais. Para a construção dessas estimativas, adotou-se o Sistema de Informações sobre Mortalidade (SIM) como fonte primária para o registro dos óbitos, que compõem o numerador da taxa. O denominador, por sua vez, é representado pelo tamanho da população residente no ponto médio de cada ano, conforme estabelecido pela conciliação demográfica realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE) (2024).

Do ponto de vista metodológico, é crucial destacar que as taxas (${}_nM_x$) fornecidas não consistem em dados brutos, mas sim em estimativas que passaram por um rigoroso processo de ajuste. Os óbitos, desagregados por sexo e grupos etários, foram suavizados mediante a aplicação de uma média móvel de três anos. Este procedimento visa atenuar flutuações de curto prazo e eventuais sazonalidades, com a ressalva de que tal técnica não foi aplicada de forma homogênea para o período atípico compreendido entre 2019 e 2023. A população de referência, também estratificada por sexo e idade, deriva de uma conciliação demográfica que integra os resultados dos Censos Demográficos de 2000, 2010 e 2022, aliados a registros vitais e informações sobre migração (Instituto Brasileiro de Geografia e Estatística (IBGE), 2024).

Adicionalmente, segundo Instituto Brasileiro de Geografia e Estatística (IBGE) (2024) implementaram-se ajustes para corrigir sub-registros na cobertura de óbitos, os quais variam por Unidade da Federação, grupo etário e sexo. Tais ajustes foram calculados com base em metodologias consagradas, incluindo a Busca Ativa, o método de Captura-Recaptura e consulta à literatura especializada. Para as idades mais avançadas, onde a qualidade dos dados frequentemente se deteriora, aplicou-se uma correção na estrutura etária utilizando o modelo log-quadrático, posteriormente estendido pela metodologia das Nações Unidas. Embora esses procedimentos elevem significativamente a consistência temporal e a comparabilidade dos dados,

é importante reconhecer que eles podem introduzir certos vieses. Especificamente, a suavização reduz a variância natural dos dados e a imposição de um modelo paramétrico nas idades avançadas pode influenciar a forma da curva de mortalidade, aspectos que devem ser considerados em análises de caráter preditivo.

Essas tábuas, organizadas em grupos etários quinquenais (0, 1-4, 5-9, 10-14, ..., 90+), foram analisadas em três categorias distintas: população feminina, masculina e total, permitindo avaliar diferenças estruturais na mortalidade por sexo. O período observado abrange os anos de 2000 a 2023, contudo, considerando o impacto atípico da pandemia de COVID-19 nas taxas de mortalidade e os procedimentos diferenciados de suavização para 2019-2023, a análise foi restrita ao intervalo de 2000 a 2019. Esta delimitação visou garantir a avaliação de padrões demográficos em condições de normalidade, evitando distorções por eventos excepcionais e pela heterogeneidade metodológica do período recente.

Complementarmente, os dados internacionais foram acessados por meio do pacote *HMDHFDplus* (Riffe; al., 2021) no ambiente *R* (Software, 2025), abrangendo os 49 países disponíveis no *Human Mortality Database* (HMD) - ver Apêndice A. As tábuas do HMD, também anuais e já estruturadas em grupos quinquenais, foram truncadas na idade 90+ para uniformização com os dados brasileiros. Este processo de truncamento consistiu em:

i) eliminar todas as idades acima de 94 anos;

ii) recalcular as variáveis demográficas na idade 90 utilizando a expectativa de vida residual (ex) como parâmetro central, seguindo as transformações: a força de mortalidade (m_x) foi definida como o inverso da expectativa de vida ($\frac{1}{e_x}$); o número de pessoas-anos vividos (L_x) foi igualado ao número total de pessoas-anos a viver (T_x); a probabilidade de morte (q_x) foi fixada em 1; o número de mortes (d_x foi igualado ao número de sobreviventes (l_x); e a média de anos vividos no intervalo (a_x) foi igualada à expectativa de vida (e_x).

O recorte temporal para os dados internacionais foi de 1950 a 2019, perfazendo 70 (setenta) observações, seguindo padrão amplamente adotado em estudos de previsão de mortalidade com dados do *Human Mortality Database* (Gersten; Barbieri, 2021; Perla *et al.*, 2020; Santis *et al.*, 2024). Esta janela temporal exclui distorções históricas das grandes guerras e prioriza tendências demográficas contemporâneas, sendo consistente com trabalhos que avaliam transições epidemiológicas modernas (Gersten; Barbieri, 2021) e desenvolvem modelos de séries temporais para mortalidade (Perla *et al.*, 2020; Santis *et al.*, 2024). O período pós-1950 oferece maior estabilidade para análises preditivas, aproveitando a disponibilidade de dados consistentes do HMD até 2019 (Santis *et al.*, 2024).

Para o processamento analítico, as taxas de mortalidade (${}_nm_x$) sofreram transformação logarítmica natural. Este procedimento suaviza as séries temporais, estabiliza a variância e potencializa o desempenho de modelos de previsão, particularmente em abordagens baseadas em estrutura autoregressiva (Lütkepohl; Xu, 2010). Adicionalmente, a transformação garante que

as previsões mantenham valores estritamente positivos quando reconvertidas à escala original, assegurando coerência demográfica. Ressalta-se que os modelos Lee-Carter e FDM realizam internamente essa transformação, dispensando aplicação prévia.

Para as modelagens Lee-Carter e FDM, foram incorporados dados complementares de exposição ao risco (população por grupo etário), igualmente disponibilizados pelo Instituto Brasileiro de Geografia e Estatística (IBGE) (2025). Essas informações permitiram a estimativa dos parâmetros do modelo por máxima verossimilhança, fundamentais para capturar interações entre idade e tendências temporais.

3.2.2 Previsão das Taxas de Mortalidade

A série história de dados de mortalidade brasileira foram segmentadas em três períodos distintos para garantir a robustez na avaliação dos modelos. O conjunto de treinamento compreendeu os anos de 2000 a 2011, totalizando doze observações temporais, que foi utilizado para estimar os parâmetros iniciais dos modelos. O período de validação, de 2012 a 2015, com quatro anos de observação e permitiu o ajuste de hiperparâmetros nos modelos de redes neurais, bem como foi utilizado na avaliação individual dos modelos para abordagem combinatória. O intervalo de 2016 a 2019 foi reservado exclusivamente para avaliação final do desempenho preditivo. Esta estratificação temporal permitiu a análise de capacidade de generalização dos modelos frente a dados não vistos durante o treinamento.

A implementação dos modelos Lee-Carter e Functional Demographic Model foram realizadas por meio dos pacotes *demography* (Hyndman, 2024) e *forecast* (Hyndman *et al.*, 2024) do ambiente *R* (Software, 2025). Para a modelagem ARIMA foi empregada a função 'auto.arima' do mesmo pacote *forecast*, em R, uma vez que este algoritmo combina testes de raiz unitária para determinar a ordem de diferenciação, minimização do critério de Informação de Akaike Corrigido (AIC), máxima verossimilhança para estimação de parâmetros, retornando automaticamente a configuração ótima que minimiza o AIC, garantindo equilíbrio entre complexidade e precisão preditiva (Hyndman; Athanasopoulos, 2021). A modelagem ETS foi conduzida com a utilização da função 'ets' também do pacote *forecast* em R. O método da função decompõe a série temporal em componentes estruturais de erro, tendência e sazonalidade, com seleção automática do modelo baseada na minimização do AIC. A especificação das componentes aditiva e multiplicativa e a inclusão de termos de amortecimento são determinadas pelo próprio algoritmo, dispensando intervenção manual e garantindo adaptabilidade aos dados.

A modelagem de redes neurais artificiais foi implementada mediante protocolo unificado a fim de garantir comparabilidade. Inicialmente, justifica-se a opção pelas Unidades Recorrentes Portais (GRU) em detrimento das LSTM, conforme sustentado por Zhou *et al.* (2016) e Jing *et al.* (2019). As redes recorrentes do tipo GRU apresentam arquitetura simplificada com um menor número de parâmetros treináveis, o que reduz o risco de sobreajuste em cenários de dados

limitados, diminui o tempo de treinamento, e mantém capacidade preditiva comparável as LSTM, em situações com poucas observações.

Para todas as arquiteturas neurais (GRU, CNN e híbridas), foi adotado um fluxo metodológico comum. Os dados, já logaritmizados, foram submetidos à padronização 'StandarScaler' da biblioteca *scikit-learn* (Pedregosa; al., 2011), transformando cada variável para distribuição de média zero e desvio padrão unitário, essencial para estabilizar a convergência em redes profundas. O processo de janelamento temporal foi sistematicamente investigado mediante janelas de 2 a 5 anos, para investigar a capacidade de extração de padrões. A implementação ocorreu no ecossistema *Python* utilizando bibliotecas especializadas: *TensorFlow* (Abadi *et al.*, 2015) para operações de tensor, *Keras* (Chollet *et al.*, 2015) para construção de redes, *KerasTuner* (O'Malley; al., 2019) para otimização de hiperparâmetros, complementadas por *NumPy* (Harris *et al.*, 2020) e *pandas* (The Pandas Development Team, 2020) para manipulação numérica.

A calibração dos hiperparâmetros foi conduzida via busca aleatória 'RandomSearch' com 10 combinações de hiperparâmetros e 10 execuções por cada combinação, otimizando diretamente o *RMSE* na base de validação (Hosseini; Prieto; Álvarez, 2024; Andonie; Florea, 2020; Blume; Benedens; Schramm, 2021). Cada configuração foi treinada por até 500 épocas com *batch size* de 8, incorporando parada antecipada ('EarlyStopping' com paciência=10) para interromper treinos sem melhoria no conjunto de validação, prevenindo *overfitting* e otimizando recursos computacionais (Kandel; Castelli; Popovič, 2020). A inicialização de pesos seguiu com otimizador *Adam* ajustando taxas de aprendizado em escala logarítmica (0,01, 0,001, 0,0001) (Macêdo; Zanchettin; Ludermir, 2024).

3.2.3 Métricas de Avaliação

Conforme apontam Hyndman e Athanasopoulos (2021) a acurácia das previsões deve ser verificada considerando o seu desempenho comparado com dados reais e que não foram vistos durante as fases anteriores de treinamento e validação. Importante mencionar que erros de previsão são diferentes de resíduos, uma vez que estes últimos são calculados nos dados utilizados no treinamento, quando os primeiros são observados em relação aos dados não observados, do período de teste. Portanto, erro de previsão denomina-se a diferença entre o valor previsto e o valor real. Para medir a acurácia da previsão temos diversas abordagens, sendo que nesta pesquisa foram trabalhados erros dependentes de escala e percentuais.

O erro percentual absoluto médio simétrico, dado pela sigla em inglês *Symmetric Mean Absolute Percentage Error* (*sMAPE*), é uma métrica de erro percentual, frequentemente utilizada para avaliar o desempenho de modelos de predição, uma vez que são livres de unidade, apresentando uma métrica de fácil interpretação, pois independe da escala dos dados. É dada pela fórmula abaixo:

$$\text{sMAPE} = \frac{200}{k} \sum_{i=1}^{k} \left| \frac{z_i - \hat{z}_i}{z_i} \right| \tag{3.8}$$

Por sua vez, o erro médio absoluto, em inglês *Mean Absolute Error* (*MAE*), é uma medida de precisão dependente da escala e, portanto, não pode ser utilizada para comparar acurácia em escalas diferentes. É bastante utilizada para avaliação de acurácia, pois é fácil de entender e calcular, sendo que um método de previsão que minimiza o *MAE* conduzirá a previsões da mediana. Sua fórmula é expressa por:

$$\text{MAE} = \frac{\sum_{i=1}^{k} |z_i - \hat{z}_i|}{k} \tag{3.9}$$

A raiz quadrada do erro quadrático médio, em inglês *Root Mean Square Error* (*RMSE*), também é uma métrica dependente da escala, mas por elevar ao quadrado as diferenças entre os valores reais e os previstos, acaba penalizando mais fortemente as grandes diferenças, ou seja, o quanto *outliers* prejudicam a acurácia das previsões. É expresso pela fórmula:

$$\text{RMSE} = \sqrt{\frac{1}{K} \sum_{i=1}^{k} (z_i - \hat{z}_i)^2} \tag{3.10}$$

Em todas as três fórmulas apresentadas K é o número de observações do conjunto de teste, e z_i e $\hat{z}_i$, respectivamente, a i-ésima observação e a i-ésima previsão gerada pelo modelo. As previsões geradas por todos os métodos aplicados nesta pesquisa foram avaliadas pelas três métricas apresentadas, avaliando a sua acurácia individualmente por faixa etária, bem como por medidas de tendência central por conjunto de dados.

3.3 CONSIDERAÇÕES DO CAPÍTULO

Este capítulo apresentou o procedimento metodológico que fundamenta a pesquisa, estruturando-se em dois eixos principais. Primeiramente, detalharam-se os sistemas propostos: na Transferência de Aprendizado, países com padrões de mortalidade análogos ao Brasil selecionados e ponderados exponencialmente treinando redes recorrentes do tipo GRU, convolucionais CNN e híbridas CNN-GRU, para posterior ajuste fino com dados locais. Na combinação de preditores foram adotadas duas técnicas, ponderação pelo inverso dos erros (*RMSE*, *MAE*, *sMAPE* normalizados) e abordagem de média simples. Por fim, o Protocolo detalhou a obtenção das bases de dados e o seu pré-processamento, transformação de escala; a segmentação temporal (treino: 2000–2011, validação: 2012–2015, teste: 2016–2019), a implementação de modelos (Lee-Carter, FDM, ARIMA, ETS em R; redes neurais em Python com otimização via *KerasTurner*), e as Métricas de Avaliação (*sMAPE*, *MAE*, *RMSE*) para análise comparativa. O Quadro 3.1 resume a descrição, procedimentos e referências de da metodologia aplicada no presente trabalho. Para

fins de reprodutibilidade, todos os dados, códigos e resultados intermediários deste estudo estão disponíveis publicamente no repositório: <https://github.com/cleodecker/TCC>.

Quadro 3.1 – *Quadro Resumo da Metodologia*

Eixo	Descrição	Procedimentos	Referências
Sistemas Propostos	*Transfer Learning*: seleção dos 10 países mais próximos (Distância de Manhattan, pesos exponenciais $w(c)$); ajuste aos dados brasileiros. Modelos: GRU, CNN, CNN-GRU.	Pré-treino em países fontes (HMD); *Fine-tuning* em Brasil; Combinação de preditores: inverso do erro e média simples.	Pan e Yang (2010), Nalmpatian *et al.* (2025), García-Aroca *et al.* (2023)
Protocolo	Bases: IBGE (2000–2019) e HMD (1950–2019). Transformação: $\log(m_{x,t})$, padronização.	Segmentação: treino (2000–2011), validação (2012–2015), teste (2016–2019); Modelos tradicionais: LC, FDM, ARIMA, ETS (R: *forecast*, *demography*); Modelos neurais: GRU, CNN, CNN-GRU (Python: TensorFlow/Keras, KerasTuner); Hiperparâmetros via *RandomSearch*, até 500 épocas, *early stopping*.	Hyndman e Athanasopoulos (2021), Gujarati e Porter (2011), Petneházi e Gáll (2019)
Métricas de Avaliação	Comparação da acurácia em conjunto de teste.	**sMAPE**: simétrica, independente de escala; **MAE**: dependente de escala; **RMSE**: penaliza grandes erros; Resultados por faixa etária + agregados (tendência central).	Hyndman e Athanasopoulos (2021)

Fonte: Elaboração própria

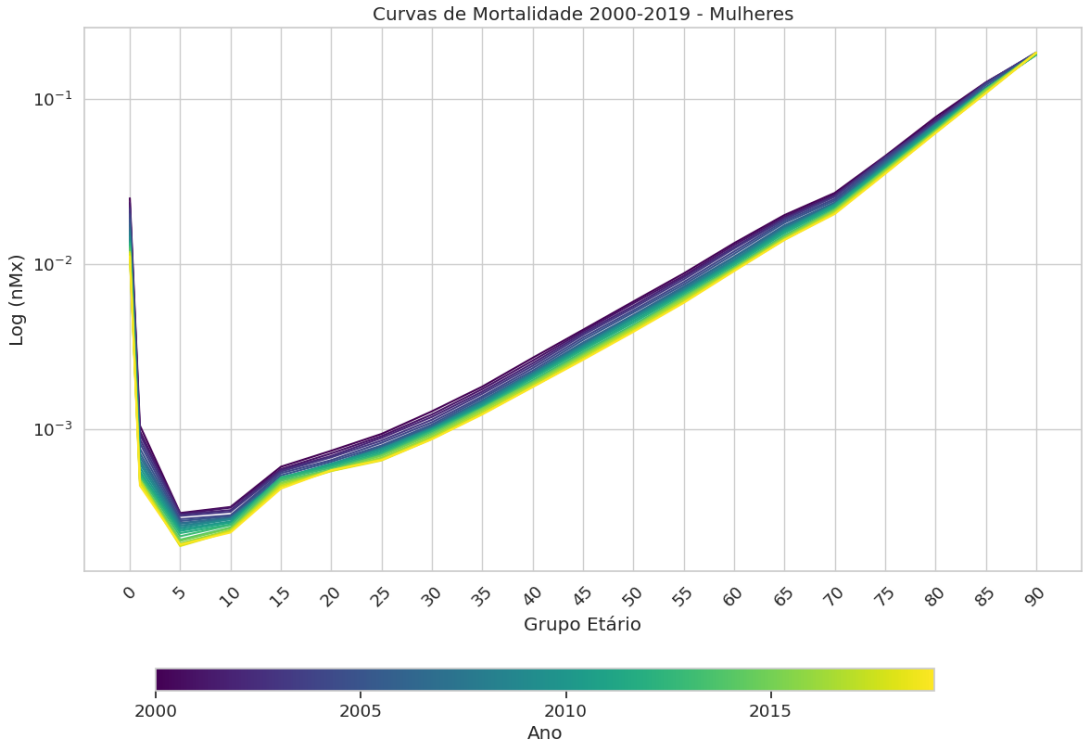
4 RESULTADOS

Este capítulo apresenta os resultados mais relevantes da pesquisa. Inicialmente a Seção 4.1 apresenta as características da base de dados de mortalidade brasileira. A Seção 4.2 descreve as características das séries temporais trabalhados e o diagtnóstico dos resíduos das modelagens ETS e ARIMA. Na sequência, a Seção 4.3 aponta os países selecionados e os respectivos pesos utilizados nos treinamentos do sistema proposto de Transferência de Aprendizado. A Seção 4.4 mostra um comparativo de desempenho preditivo dos modelos utilizados na pesquisa: Lee-Carter, FDM, ARIMA, ETS, Redes Neurais com Transferência de Aprendizado e previsões combinadas. Por fim, os achados desta pesquisa são discutidos na Seção 4.5, onde são confrontados com a literatura existente.

4.1 DADOS DE MORTALIDADE BRASILEIROS

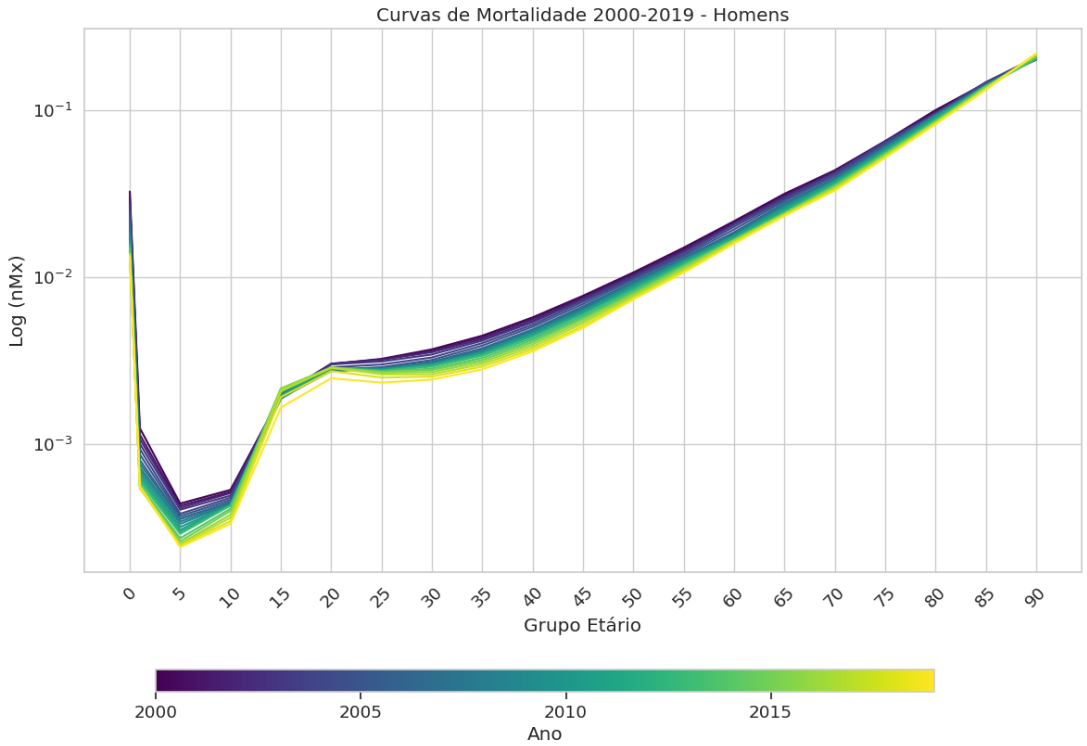
Esta seção tem como objetivo analisar a evolução histórica das taxas centrais de mortalidade por idade (nMx) no Brasil. A observação das séries temporais das taxas de mortalidade, apresentadas graficamente, permite identificar uma tendência de declínio generalizado ao longo do período estudado, com particularidades dignas de nota. Notadamente, a curva de mortalidade feminina demonstra um comportamento mais estável e suave, com tendência consistente de redução das taxas em todas as idades (Figura 9), em nítido contraste com a curva masculina (Figura 10), que apresenta padrão mais irregular com manutenção elevada das taxas em idades jovens (Malta *et al.*, 2020; Murray; Cerqueira; Kahn, 2013). Estudos confirmam essa disparidade de gênero, destacando que as taxas de mortalidade por causas externas são significativamente mais altas em homens jovens brasileiros, com picos entre 15-24 anos atribuídos a fatores de risco como violência urbana e criminalidade (Souza, 2005; Duarte *et al.*, 2015).

Figura 9 – Curva de Mortalidade para Mulheres



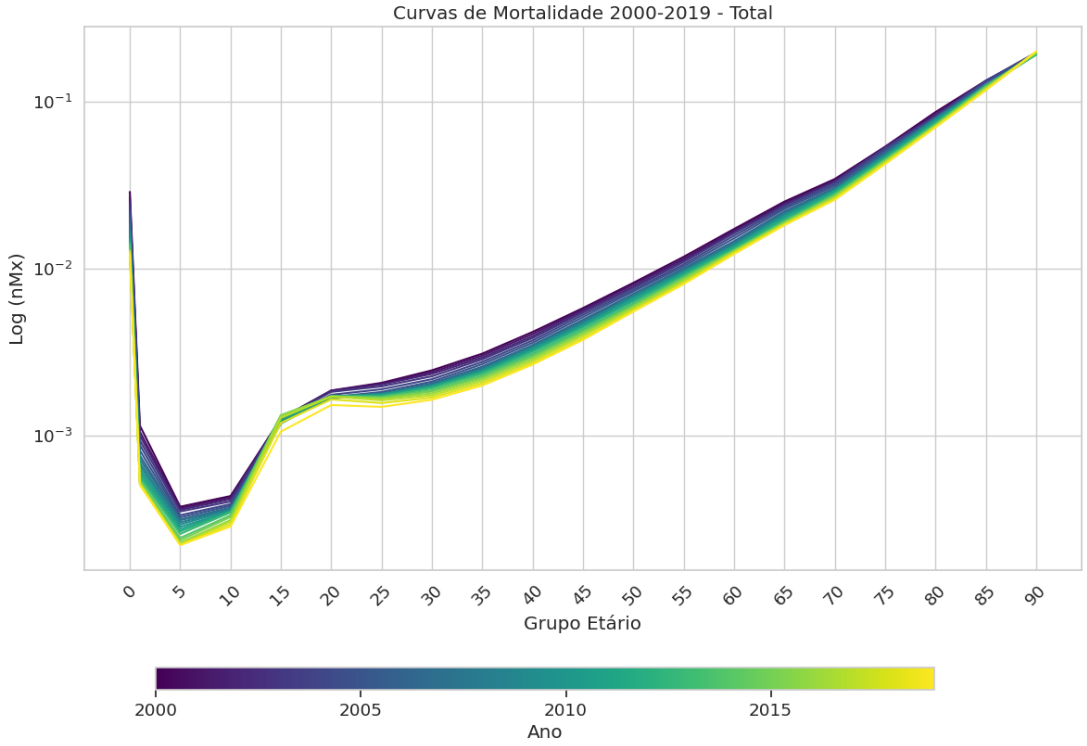
Fonte: Elaboração própria

Figura 10 – Curva de Mortalidade para Homens



Fonte: Elaboração própria

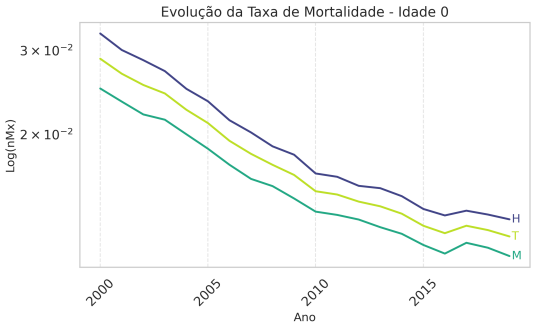
Figura 11 – Curva de Mortalidade Total



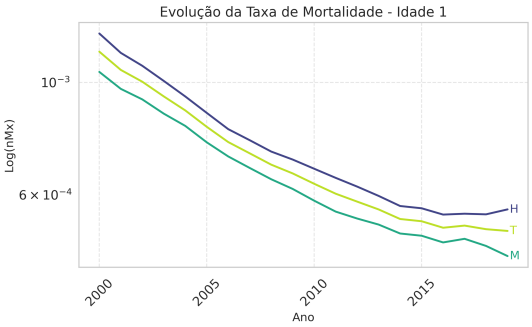
Fonte: Elaboração própria

Figura 12 – Evolução das taxas centrais de mortalidade para os grupos etários 0 a 14 anos

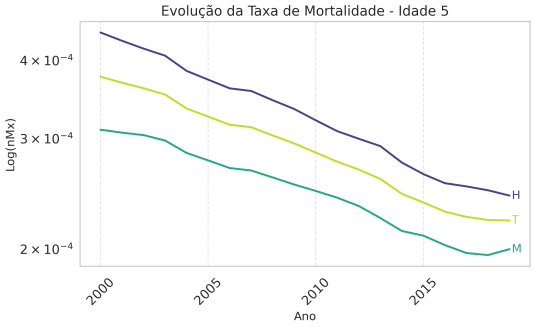
(a) Ln $({}_nM_x)$ para o grupo etário 0 ano



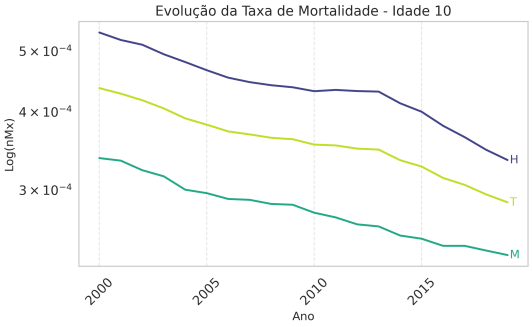
(b) Ln $({}_nM_x)$ para o grupo etário 1 a 4 anos



(c) Ln $({}_nM_x)$ para o grupo etário 5 a 9 anos



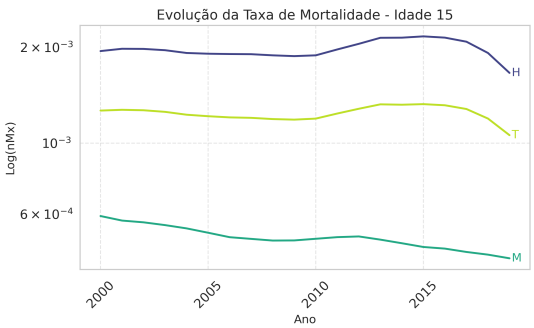
(d) Ln $({}_nM_x)$ para o grupo etário 10 a 14 anos



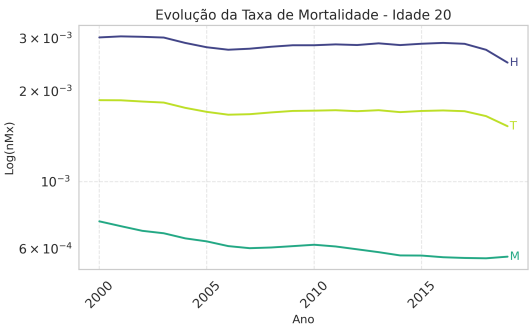
Fonte: Elaboração própria

Figura 13 – Evolução das taxas centrais de mortalidade para os grupos etários 15 a 34 anos

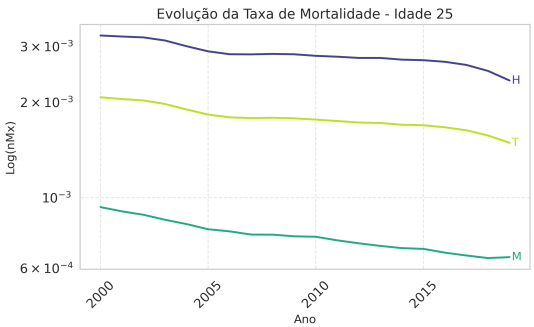
(a) Ln $({}_nM_x)$ para o grupo etário 15 a 19 anos



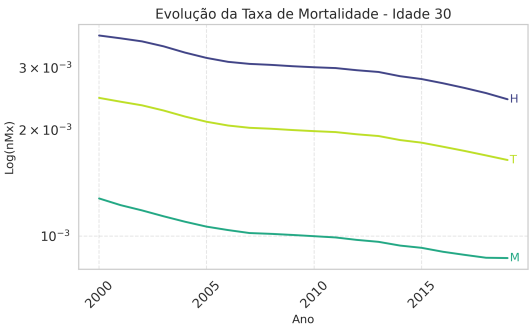
(b) Ln $({}_nM_x)$ para o grupo etário 20 a 24 anos



(c) Ln $({}_nM_x)$ para o grupo etário 25 a 29 anos



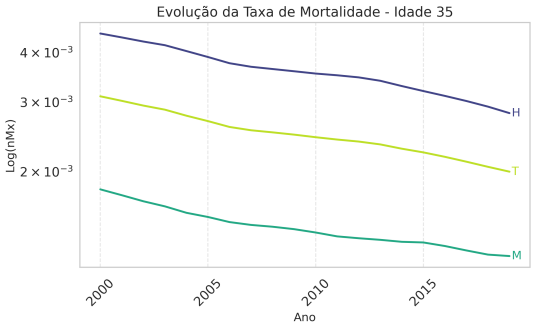
(d) Ln $({}_nM_x)$ para o grupo etário 30 a 34 anos



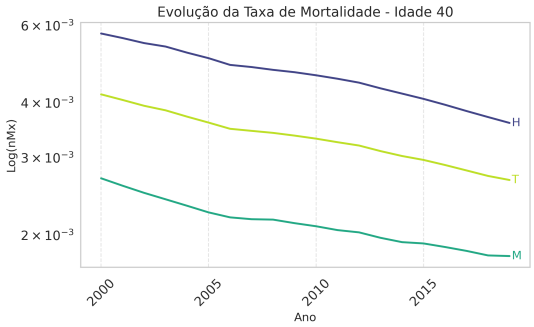
Fonte: Elaboração própria

Figura 14 – Evolução das taxas centrais de mortalidade para os grupos etários 35 a 54 anos

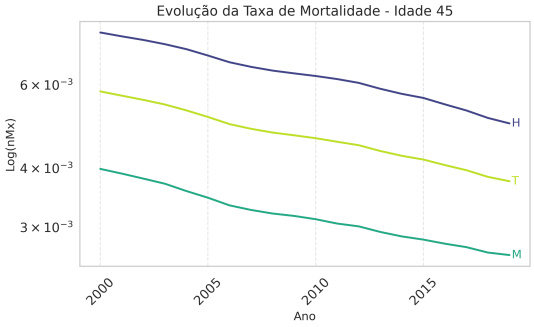
(a) Ln $({}_nM_x)$ para o grupo etário 35 a 39 anos



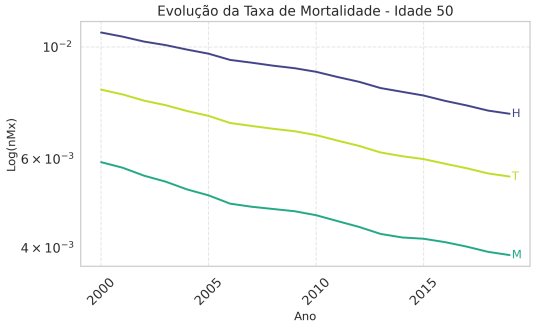
(b) Ln $({}_nM_x)$ para o grupo etário 40 a 44 anos



(c) Ln $({}_nM_x)$ para o grupo etário 45 a 49 anos



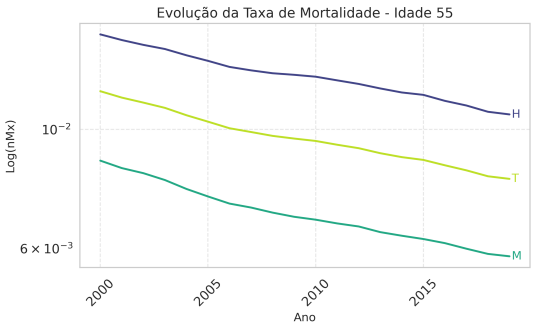
(d) Ln $({}_nM_x)$ para o grupo etário 50 a 54 anos



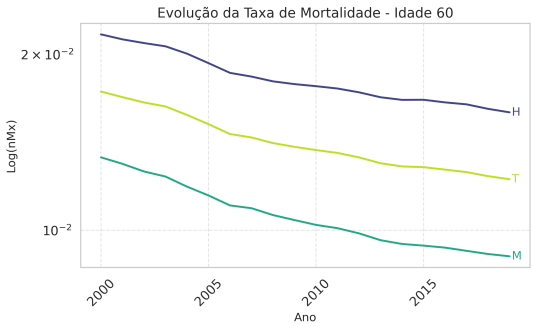
Fonte: Elaboração própria

Figura 15 – Evolução das taxas centrais de mortalidade para os grupos etários 55 a 74 anos

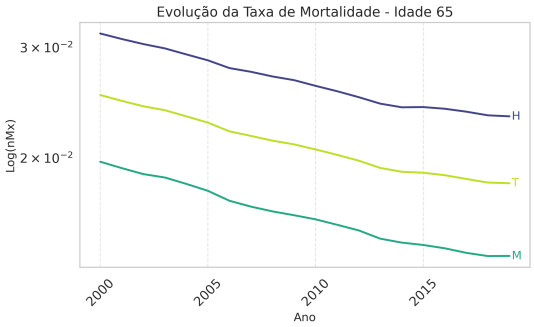
(a) Ln $({}_nM_x)$ para o grupo etário 55 a 59 anos



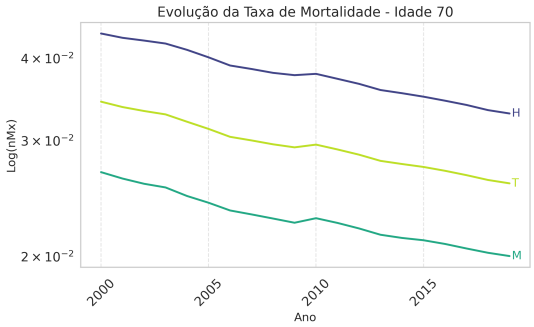
(b) Ln $({}_nM_x)$ para o grupo etário 60 a 64 anos



(c) Ln $({}_nM_x)$ para o grupo etário 65 a 69 anos



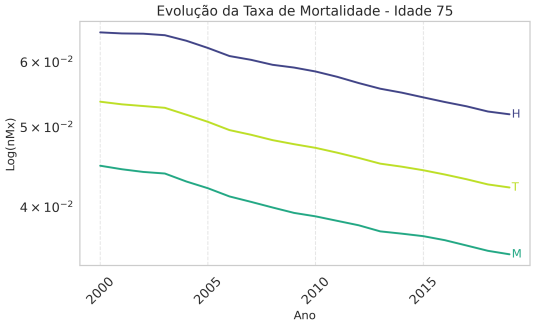
(d) Ln $({}_nM_x)$ para o grupo etário 70 a 74 anos



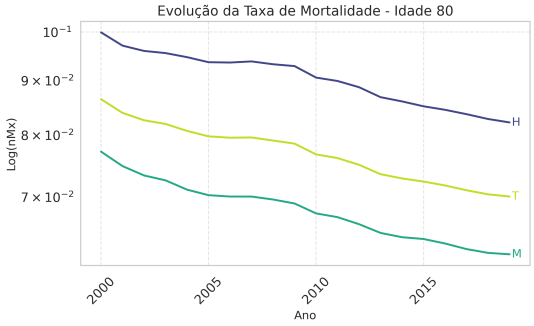
Fonte: Elaboração própria

Figura 16 – Evolução das taxas centrais de mortalidade para os grupos etários 75 a 90+ anos

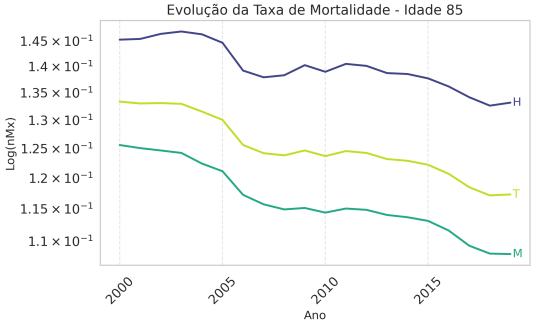
(a) Ln $({}_nM_x)$ para o grupo etário 75 a 79 anos



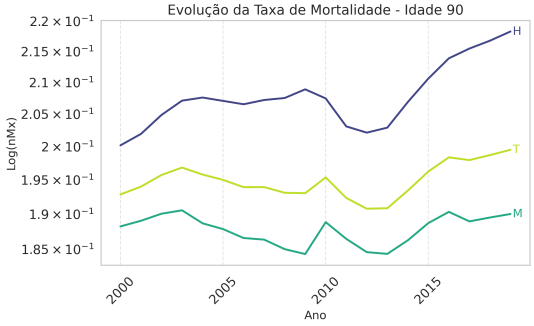
(b) Ln $({}_nM_x)$ para o grupo etário 80 a 84 anos



(c) Ln $({}_nM_x)$ para o grupo etário 85 a 89 anos



(d) Ln $({}_nM_x)$ para o grupo etário 90 anos e mais



Fonte: Elaboração própria

A redução mais pronunciada ocorreu efetivamente nos grupos etários iniciais, especialmente nas idades de 0 ano e 1 a 4 anos (Figuras 12a e 12b, respectivamente). Entretanto, os grupos etários masculinos de 15 a 19 e 20 a 24 anos apresentaram relativa estabilidade das taxas, com a primeira faixa registrando determinada elevação a partir de 2010 e brusca redução em 2019 (Figuras 13a e 13b). Essa dinâmica alinha-se com padrões observados em populações brasileiras, onde causas externas representam parcela significativa das mortes em homens jovens (Fórum Brasileiro de Segurança Pública, 2019; Instituto de Pesquisa Econômica Aplicada; Fórum Brasileiro de Segurança Pública, 2020). A redução brusca em 2019 é corroborada por relatórios que documentam queda nos homicídios dolosos nacionais, influenciada por armistícios entre facções criminosas e políticas estaduais de segurança (Cerqueira; Coelho, 2019; Feltran, 2020). Por fim, observa-se maior oscilação no grupo etário de 90 anos e mais, o que pode ser atribuído à menor representatividade populacional nesta faixa etária extrema (Figura 16d).

4.2 DIAGNÓSTICO DAS SÉRIES TEMPORAIS

A primeira etapa do diagnóstico das séries foi a verificação da estacionariedade, condição fundamental para a adequada modelagem via processos autorregressivos integrados de médias móveis (ARIMA). Essa etapa foi conduzida por meio do teste de raiz unitária de Dickey-Fuller aumentado (ADF), implementado no *software* R a partir da função 'adf.test()' disponível no pacote *tseries* (Trapletti; Hornik; LeBaron, 2024). Esse teste possui como hipótese nula a não estacionariedade da série, de modo que a rejeição desta hipótese indica uma série estacionária.

Tabela 1 – Diagnóstico de estacionariedade das séries temporais segundo o teste ADF por sexo

Classificação	**Ambos**	**Homens**	**Mulheres**	**Total**
Estacionária	2	3	2	7
Não Estacionária	18	17	18	53

Fonte: Elaboração própria

Os resultados apresentados na Tabela 1 revelam que a grande maioria das séries não se mostrou estacionária. Constatou-se estacionariedade apenas em alguns grupos etários específicos: para os grupos de Mulheres e para a população Total, apenas os grupos de 10–14 anos e 30–34 anos foram classificados como estacionários. Para o grupo de Homens, além desses dois grupos etários, o grupo de 75–79 anos também apresentou estacionariedade. Esse padrão sugere que, de forma geral, o comportamento temporal das séries possui tendência ou variabilidade de longo prazo, dificultando um ajuste direto de modelos ARIMA sem diferenciação. Os poucos casos estacionários podem sinalizar estabilização precoce da série ou níveis menos suscetíveis a variações externas.

A identificação das ordens dos modelos ARIMA foi realizada por meio da função 'auto.arima()' do pacote *forecast* (Hyndman *et al.*, 2024) no R (Software, 2025), que avalia

diferentes combinações de parâmetros (p, d, q) e seleciona aquele que minimiza critérios de informação, como AIC e BIC.

Tabela 2 – Frequência das ordens ARIMA identificadas nas séries temporais por sexo

Ordem ARIMA	Ambos	Homens	Mulheres	Total
ARIMA(0,1,0)	8	10	10	28
ARIMA(0,2,0)	7	5	6	18
ARIMA(1,1,0)	3	3	1	7
ARIMA(0,2,1)	1	0	1	2
ARIMA(1,2,0)	0	0	2	2
ARIMA(0,0,1)	1	0	0	1
ARIMA(2,0,0)	0	1	0	1
ARIMA(2,1,0)	0	1	0	1
Total	**20**	**20**	**20**	**60**

Fonte: Elaboração própria

A Tabela 2 indica que a predominância foi de modelos ARIMA simples, em especial o ARIMA(0,1,0) e o ARIMA(0,2,0), evidenciando que diferenças de primeira e segunda ordem foram suficientes para deixar estacionária a maioria das séries. A quase ausência de parâmetros autorregressivos (p) e de médias móveis (q) demonstra que as séries se comportam essencialmente como processos de passeio aleatório diferenciado. Isso reforça que as séries apresentam trajetória persistente, com pouca estrutura de dependência de curto prazo, o que sugere dificuldade para capturar padrões determinísticos mais complexos, refletindo maior aleatoriedade nos dados.

Para avaliar a adequação dos ajustes, aplicou-se o teste de Ljung-Box, que tem como hipótese nula a ausência de autocorrelação nos resíduos em diferentes defasagens (Hyndman; Athanasopoulos, 2021). Esse teste foi conduzido no R a partir da função 'Box.test()', pertencente ao pacote base, considerando ajuste para graus de liberdade de acordo com os parâmetros do modelo ARIMA.

Tabela 3 – Diagnóstico dos modelos ARIMA segundo o teste de Ljung-Box para ruído branco por sexo

Classificação	Ambos	Feminino	Masculino	Total
Ruído Branco	20	18	20	58
Não Ruído Branco	0	2	0	2
Total	**20**	**20**	**20**	**60**

Fonte: Elaboração própria

Os resultados do teste de Ljung-Box (Tabela 3) reforçam que os resíduos dos modelos ARIMA se aproximaram em grande medida de ruído branco, critério essencial para validar o ajuste. Apenas para Mulheres, os grupos etários 10–14 e 65–69 anos não apresentaram resíduos caracterizados como ruído branco, indicando persistência de autocorrelação não capturada. Tal achado sugere que, para esses grupos, o ARIMA pode não ter sido suficiente para modelar

toda a estrutura temporal, possivelmente pela presença de componentes sazonais ou de choques específicos mais marcantes.

Em complemento à análise dos modelos ARIMA, procedeu-se à avaliação dos modelos de Suavização Exponencial (ETS), que oferecem uma abordagem alternativa para a modelagem de séries temporais, focando na decomposição em componentes de erro, tendência e sazonalidade. A identificação dos métodos ETS mais adequados para cada série foi realizada de forma automática por meio da função 'ets()' do pacote *forecast* no R, que avalia diferentes especificações e seleciona aquela que melhor descreve os dados segundo critérios de informação.

Tabela 4 – Frequência dos métodos ETS identificados nas séries temporais por sexo

Método ETS	**Ambos**	**Feminino**	**Masculino**	**Total**
ETS(A,A,N)	16	15	13	44
ETS(A,Ad,N)	1	4	3	8
ETS(A,N,N)	3	1	4	8
Total	**20**	**20**	**20**	**60**

Fonte: Elaboração própria

A análise apontou predominância do método ETS(A,A,N), caracterizado por erros aditivos, tendência aditiva e ausência de sazonalidade, como mostra a Tabela 4. Esse achado indica que boa parte das séries apresenta tendência linear e erros com comportamento simétrico, sem indicar sazonalidade significativa. Os poucos casos em que surgem modelos ETS com tendência amortecida ou sem tendência revelam séries mais estáveis ou com evolução suavizada ao longo do tempo. Em termos interpretativos, a preponderância de modelos com tendência simples sugere trajetórias consistentes, mas difíceis de prever a longuíssimo prazo, pois a ausência de sazonalidade e amortecimento limita a captura de padrões cíclicos.

Após a seleção dos modelos mais adequados, avaliou-se a qualidade dos ajustes via diagnóstico dos resíduos. Novamente, aplicou-se o teste de Ljung-Box, cuja hipótese nula estabelece que os resíduos constituem um processo de ruído branco, ou seja, sem autocorrelação serial significativa. Esse teste foi implementado no R por meio da função 'Box.test()', ajustando-se os graus de liberdade de acordo com os parâmetros dos modelos ETS. O objetivo desse procedimento é verificar se a parcela não explicada pelo modelo se comporta de forma aleatória, condição essencial para sustentar a validade dos ajustes propostos.

Tabela 5 – Diagnóstico dos modelos ETS segundo o teste de Ljung-Box para ruído branco por sexo

Classificação	**Ambos**	**Feminino**	**Masculino**	**Total**
Ruído Branco	19	19	12	50
Não Ruído Branco	1	1	8	10
Total	**20**	**20**	**20**	**60**

Fonte: Elaboração própria

O diagnóstico dos resíduos dos modelos ETS, avaliado também pelo teste de Ljung-Box (Tabela 5), indicou ajuste satisfatório em boa parte das séries, embora com desempenho inferior ao observado para os modelos ARIMA. No sexo feminino, apenas o grupo 10–14 anos não apresentou resíduos como ruído branco; no grupo ambos os sexos, a falha ocorreu no grupo de 45–49 anos. Já no sexo masculino, oito grupos etários (15–19, 25–29, 35–39, 40–44, 45–49, 65–69, 75–79 e 90+) não satisfizeram a condição de ruído branco. Esse resultado sugere uma maior dificuldade dos modelos ETS em capturar a estrutura temporal de determinadas séries, especialmente entre os homens, o que pode refletir trajetórias mais heterogêneas e menos lineares nesse grupo. Em termos aplicados, tal limitação aponta para possíveis ganhos no uso de modelos híbridos ou ao considerar covariáveis externas que expliquem choques não sistemáticos.

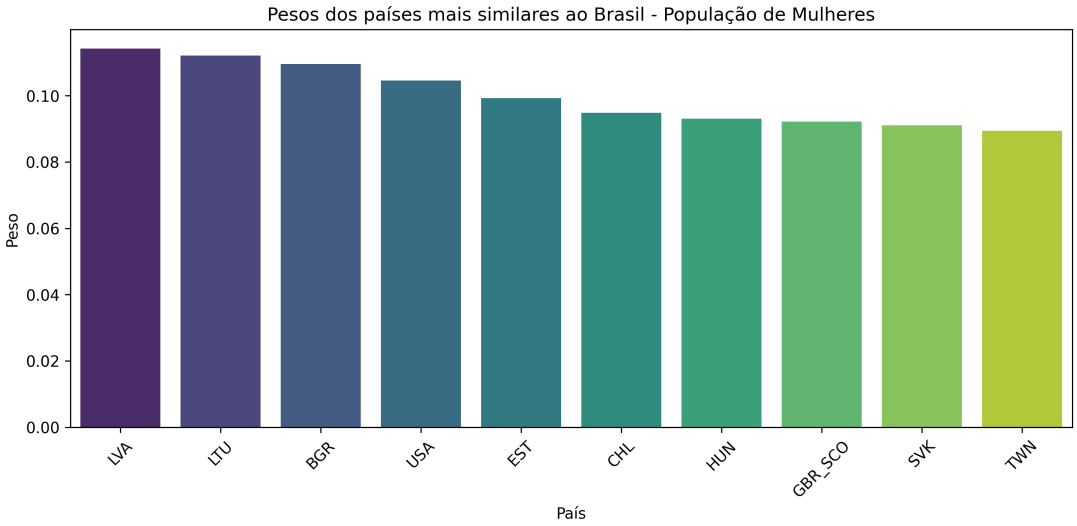
4.3 TRANSFERÊNCIA DE APRENDIZADO

Com o propósito de aprimorar o ajuste fino do modelo, procedeu-se à seleção de um subconjunto de países cujas séries temporais de mortalidade mais se assemelham à brasileira no período de 2000 a 2015. A mensuração da similaridade foi realizada por meio da Distância de Manhattan, métrica que, alinhada a Nalmpatian *et al.* (2025), oferece baixo custo computacional e expressa a proximidade entre as séries de forma eficaz, constituindo uma abordagem robusta na ausência de um estudo empírico para a seleção da medida ótima (Parmezan, 2016; Junior, 2012). Reconhecendo que as tendências de mortalidade podem exibir dinâmicas distintas por sexo, esta avaliação foi conduzida de forma independente para as populações feminina, masculina e total, a partir do conjunto de dados detalhado no Apêndice A.

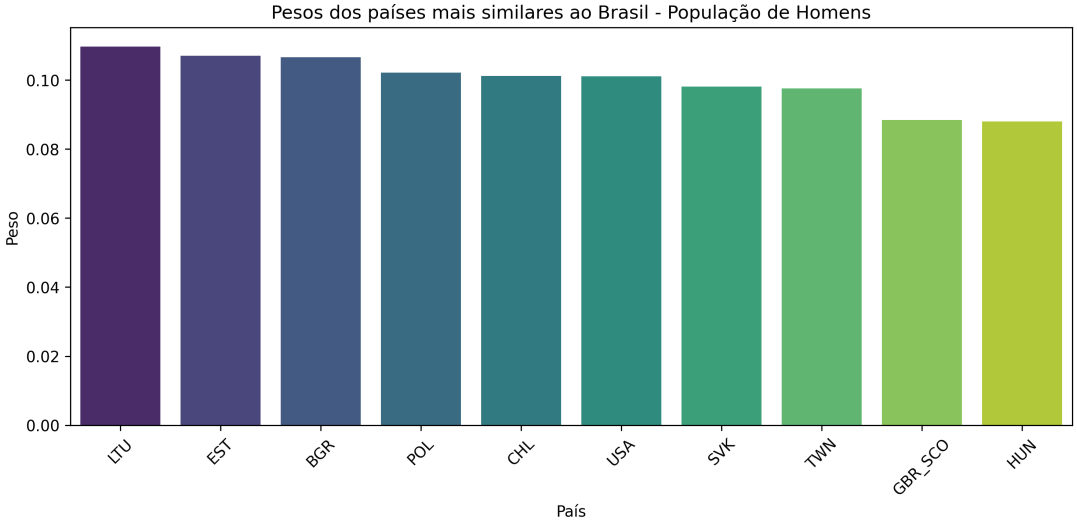
A decisão de restringir a análise aos dez países mais próximos fundamenta-se em um duplo objetivo: mitigar a introdução de variabilidade espúria, que poderia comprometer a capacidade do modelo de aprender padrões generalizáveis para o contexto brasileiro, e otimizar a eficiência computacional do processo de treinamento.

Figura 17 – Países selecionados e seus respectivos pesos por tipo de população

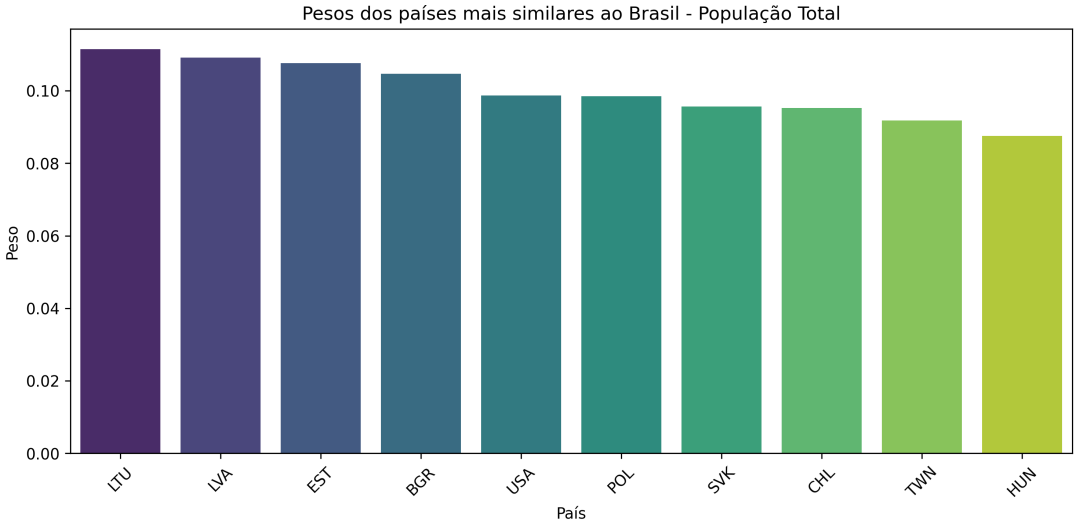
(a) Mulheres



(b) Homens



(c) Total



Fonte: Elaboração própria

Os resultados desta seleção, apresentados na Figura 17, revelam uma notável predominância de nações do Leste Europeu. Para a população feminina (Figura 17a), a Letônia (LVA) e a Lituânia (LTU) lideraram a lista de similaridade. No grupo masculino (Figura 17b), a Lituânia (LTU) e a Estônia (EST) ocuparam as primeiras posições. Por fim, para a população total (Figura 17c), Lituânia (LTU) e Letônia (LVA) novamente se destacam. Países como Bulgária (BGR), Estados Unidos (USA) e Chile (CHL) também demonstram alta similaridade, figurando nos três rankings e reforçando a consistência da seleção.

4.4 DESEMPENHO PREDITIVO

A avaliação do desempenho preditivo no período de teste (2016 a 2019) foi conduzida com base em três métricas complementares: a Raiz do Erro Quadrático Médio (RMSE), o Erro Absoluto Médio (MAE) e o Erro Percentual Absoluto Simétrico (sMAPE). Cada métrica oferece uma perspectiva distinta: o RMSE penaliza grandes desvios com maior severidade, o MAE captura a magnitude média dos erros de forma mais estável, e o sMAPE avalia a acurácia relativa por meio de um percentual de erro simétrico.

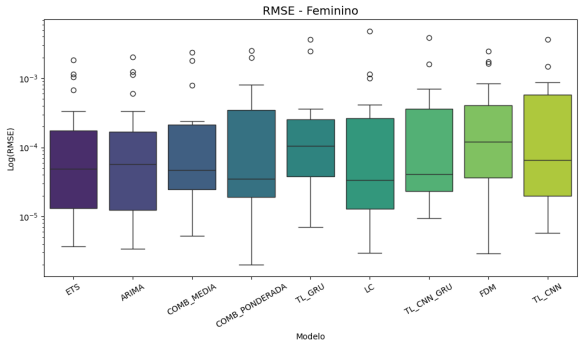
Foram comparados modelos de diferentes classes: os estatísticos tradicionalmente utilizados em demografia (Lee-Carter e FDM), os clássicos de séries temporais univariadas (ETS e ARIMA), a arquitetura proposta de redes neurais profundas com transferência de aprendizado (CNN, GRU e o modelo híbrido CNN-GRU), além de métodos de combinação de previsões (média simples e média ponderada pelo inverso dos erros).

4.4.1 Grupo Feminino

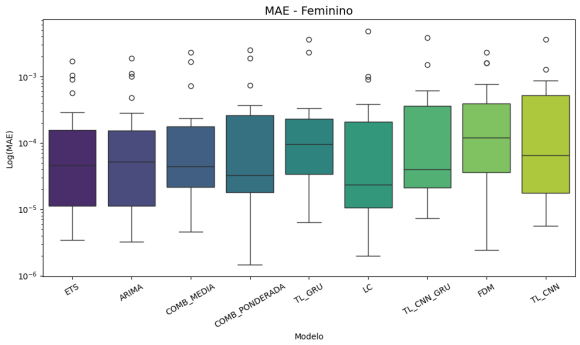
Nos resultados femininos, o modelo ETS foi o que apresentou o melhor desempenho geral, com os menores valores de RMSE (0,000282) e MAE (0,000252), além do menor sMAPE (1,839040) entre todos os modelos individuais. O modelo ARIMA obteve o segundo melhor desempenho nas métricas absolutas. Quanto às combinações de modelos, a abordagem por média simples superou a combinação ponderada em RMSE e MAE, ficando muito próximo em sMAPE. Entre os modelos de aprendizado profundo com transferência, o GRU foi o mais bem-sucedido, alcançando o menor RMSE (0,000409) e MAE (0,000389) dessa categoria. O modelo híbrido CNN-GRU, por sua vez, obteve o melhor sMAPE (2,774955) entre os modelos de *deep learning*. Os modelos clássicos LC e FDM, embora superados pelo GRU em RMSE e MAE, mantiveram um desempenho competitivo. A Tabela 6 apresenta os valores médios das métricas de erro para cada modelo no conjunto de teste feminino, com os menores erros de cada métrica em destaque. A Figura 18 ilustra a distribuição dessas métricas para cada modelo, ordenadas da esquerda para a direita pelo menor valor médio.

Figura 18 – Distribuição dos Erros para Mulheres

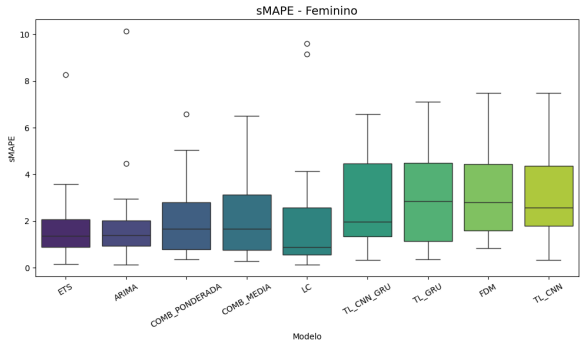
(a) Erros RMSE (log)



(b) Erros MAE (log)



(c) Erros sMAPE (%)



Fonte: Elaboração própria

Tabela 6 – Resultados médios das métricas por modelo - Mulheres

Modelo	RMSE	sMAPE	MAE
ETS	**0,000282**	**1,839040**	**0,000252**
ARIMA	0,000296	1,945806	0,000265
LC	0,000413	2,144690	0,000389
FDM	0,000439	3,125700	0,000411
TL_CNN	0,000474	3,317890	0,000443
TL_GRU	0,000409	3,097805	0,000389
TL_CNN_GRU	0,000421	2,774955	0,000399
COMB_PONDERADA	0,000356	2,056544	0,000331
COMB_MEDIA	0,000310	2,057255	0,000290

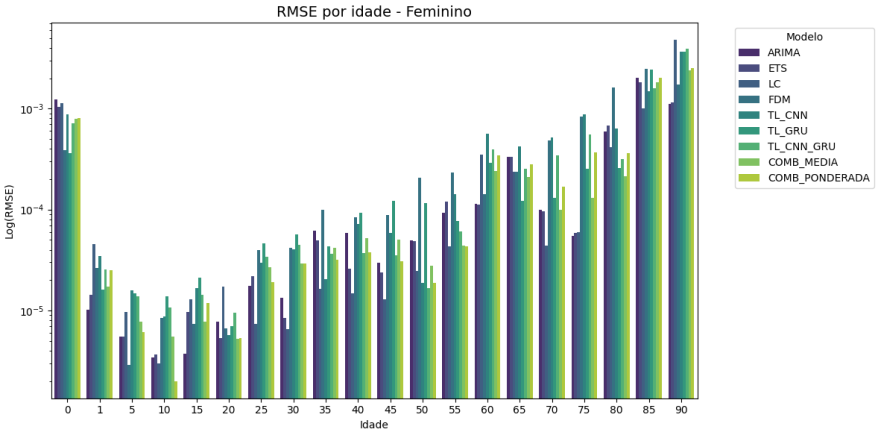
Fonte: Elaboração própria

A análise por idade revela as diferenças de acurácia bem definidas. Os modelos de séries temporais ETS e ARIMA performam adequadamente em quase todas as idades, porém geraram os maiores erros, juntamente com o modelo LC na primeira faixa etária, idade 0, grupo no qual os modelos propostos de redes neurais tiveram o melhor desempenho, inclusive do que as as abordagens por combinação. Percentualmente, com base na análise do sMAPE, os grupos etários entre 40 a 84 anos tiveram os menores erros em praticamente todos os modelos, principalmente

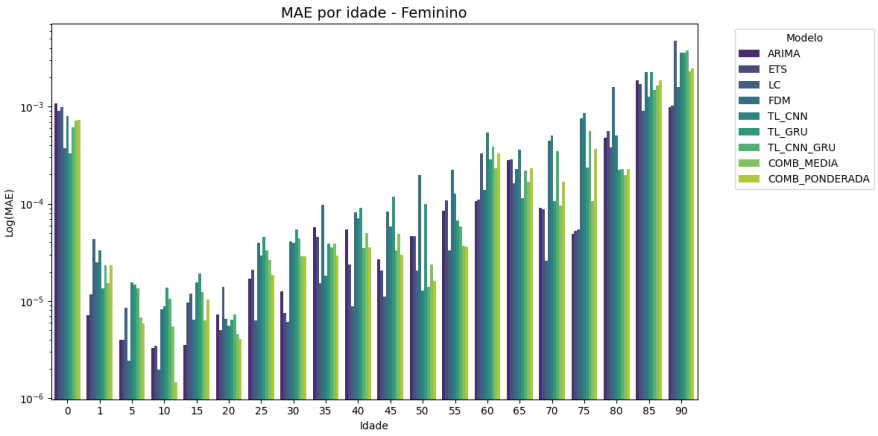
para as abordagens por combinação de preditores. A Figura 19 apresenta as métricas dos erros para cada grupo etário do grupo feminino.

Figura 19 – Erros por Grupo Etário para Mulheres

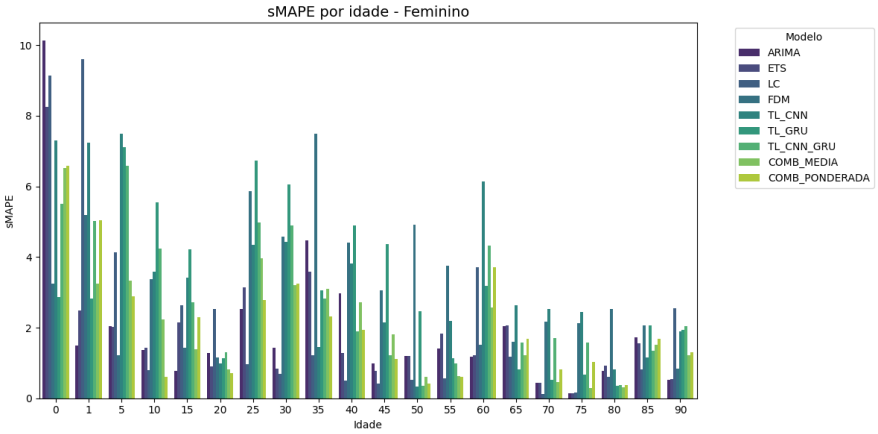
(a) Erros RMSE (log)



(b) Erros MAE (log)



(c) Erros sMAPE (%)



Fonte: Elaboração própria

Como é possível observar na Figura 19, a distribuição dos erros absolutos RMSE e MAE (ambos em escala logarítmica) por grupo etário não é uniforme, mas segue um padrão em forma de "J"que se assemelha notavelmente ao formato clássico da curva de mortalidade humana na mesma escala.

Este padrão indica que a magnitude dos erros de previsão está intrinsecamente ligada à complexidade inerente de se modelar a mortalidade em diferentes fases da vida. Os maiores erros se concentram nas idades extremas: no início da vida (idade 0), onde a dinâmica da mortalidade infantil é singular e volátil, reduzem significativamente nas idades jovens e crescem conforme os grupos etários avançam.

Desta forma, a performance dos modelos não pode ser avaliada apenas por suas médias gerais, mas deve considerar esta heterogeneidade etária. É notável que, mesmo diante deste desafio, os modelos de combinação e os de séries temporais (ETS e ARIMA) conseguiram manter erros percentuais (sMAPE) consistentemente baixos ao longo de quase toda a distribuição, enquanto os modelos de redes neurais demonstraram uma vantagem particular na modelagem da idade 0, um dos pontos mais complexos da curva.

4.4.2 Grupo Masculino

No grupo masculino, o modelo ETS demonstrou superioridade clara, registrando os menores valores em RMSE (0,000694), MAE (0,000644) e sMAPE (3,127033). O modelo ARIMA obteve o segundo melhor desempenho em sMAPE (3,670705), porém com valores de RMSE e MAE significativamente superiores ao ETS. Quanto às combinações de modelos, a abordagem ponderada superou a combinação por média simples em RMSE e MAE, mas apresentou um sMAPE ligeiramente superior. Entre os modelos de aprendizado profundo, o GRU foi o que apresentou melhor desempenho geral nesta categoria, com os menores valores de RMSE (0,001319), sMAPE (4,679003) e MAE (0,001285). O modelo híbrido CNN-GRU teve desempenho similar, enquanto o CNN apresentou os piores resultados entre todos os modelos de *deep learning*. No entanto, todos os modelos de *deep learning* apresentaram sMAPE acima de 4,6%, desempenho inferior ao dos modelos de séries temporais e Lee-Carter. Os modelos tradicionais LC e FDM mostraram diferenças significativas. Enquanto o LC teve desempenho competitivo em RMSE (0,000865) e MAE (0,000832), seu sMAPE foi consideravelmente superior (4,446590). O modelo FDM, por sua vez, apresentou os piores resultados em RMSE (0,001265) e sMAPE (6,092141) entre todos os modelos testados, com exceção do CNN que registrou o maior RMSE geral (0,001807). A Tabela 7 apresenta os desempenhos médios em cada métrica para o grupo masculino (com os menores erros de cada métrica em destaque). A Figura 20 demonstra a distribuição dos erros para cada um dos modelos aplicados na pesquisa.

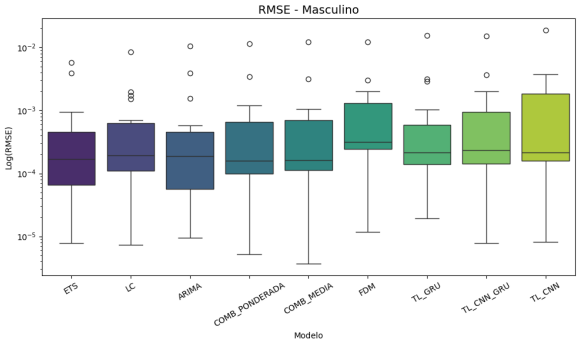
Tabela 7 – Resultados médios das métricas por modelo - Homens

Modelo	RMSE	sMAPE	MAE
ETS	**0,000694**	**3,127033**	**0,000644**
ARIMA	0,000970	3,670705	0,000866
LC	0,000865	4,446590	0,000832
FDM	0,001265	6,092141	0,001215
TL_CNN	0,001807	6,232026	0,001696
TL_GRU	0,001319	4,679003	0,001285
TL_CNN_GRU	0,001345	4,963714	0,001302
COMB_MEDIA	0,001049	3,774574	0,001001
COMB_PONDERADA	0,001015	3,929188	0,000971

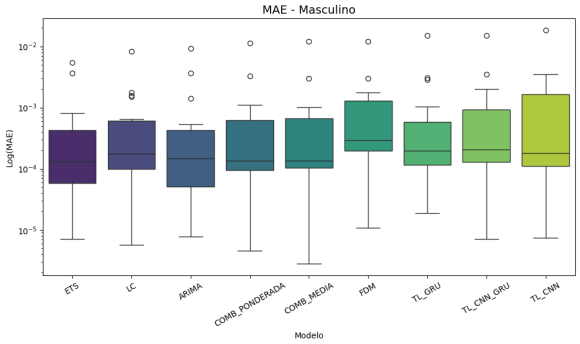
Fonte: Elaboração própria

Figura 20 – Distribuição dos Erros para Homens

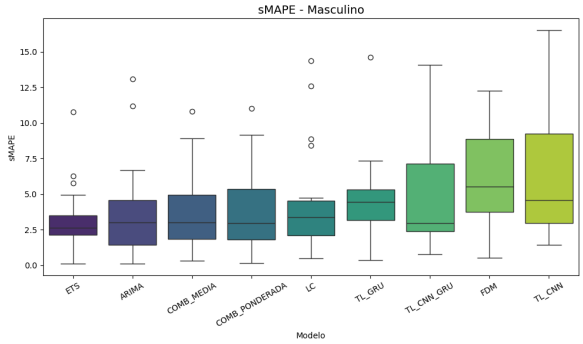
(a) Erros RMSE (log)



(b) Erros MAE (log)



(c) Erros sMAPE (%)



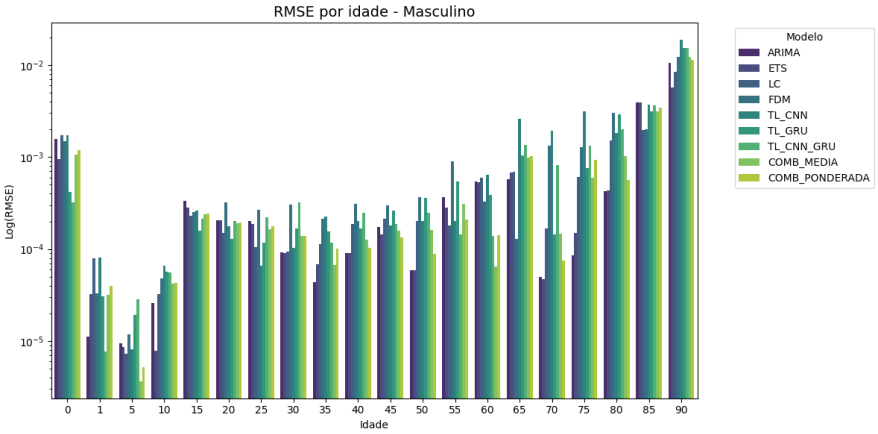
Fonte: Elaboração própria

Na estratificação etária, o padrão foi semelhante ao observado entre mulheres, com erros absolutos elevados nas idades iniciais, estabilização em adultos jovens e crescimento acentuado após 60 anos. Os modelos de redes neurais tiveram também desempenho superior aos modelos estatísticos na primeira idade, bem como boa performance em adultos de meia-idade, entre 20 e 60 anos. A análise do erro percentual apontou a dificuldade de captura dos padrões de mortalidade nos homens jovens, com métricas percentuais relativamente altas para todos os

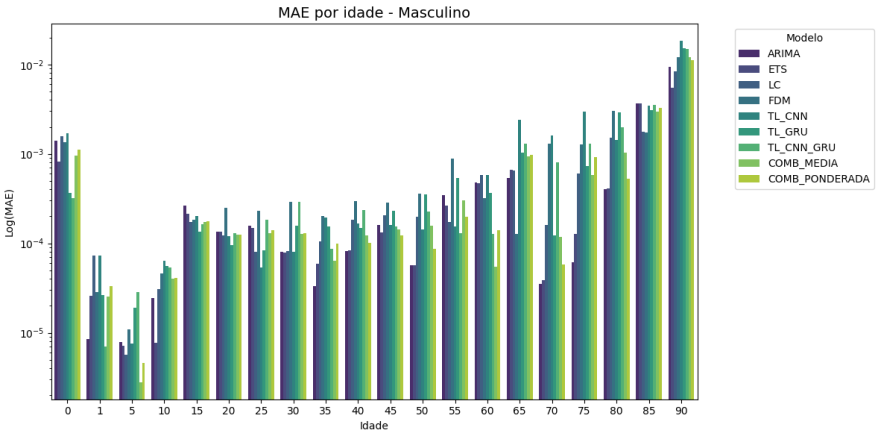
modelos no grupo etário entre 10 e 19 anos. A Figura 21 apresenta as métricas dos erros para cada grupo etário do grupo homens.

Figura 21 – Erros por Grupo Etário para Homens

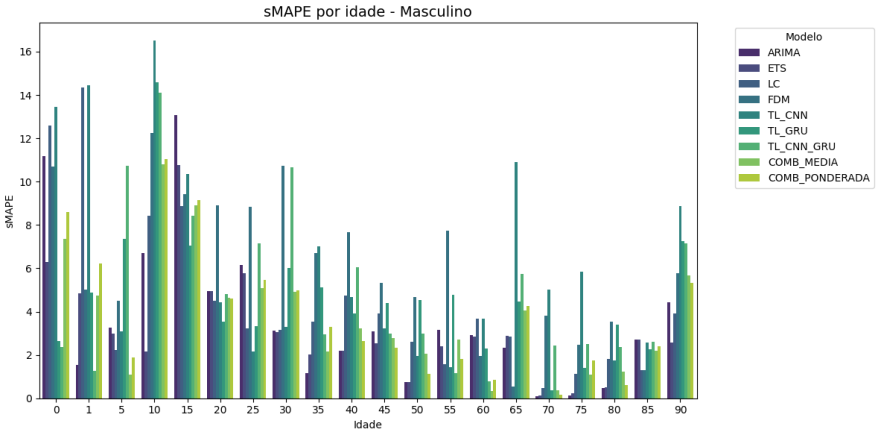
(a) Erros RMSE (log)



(b) Erros MAE (log)



(c) Erros sMAPE (%)



Fonte: Elaboração própria

A análise dos erros por grupo etário, apresentada na Figura 21, revela um padrão distinto e informativo para a população masculina. Conforme observado no grupo feminino, os erros absolutos (RMSE e MAE em log) também seguem a forma geral da curva de mortalidade, com picos nas idades extremas. No entanto, observa-se um pronunciado aumento dos erros, tanto absolutos quanto percentuais, no grupo etário dos 15 aos 25 anos.

Este pico de erro coincide perfeitamente com o fenômeno demográfico conhecido como "*accident hump*"(protuberância dos acidentes), uma elevação típica e bem documentada nas curvas de mortalidade masculina, causada principalmente por mortes violentas (acidentes, homicídios e suicídios) (Heligman; Pollard, 1980; Goldstein, 2011). A volatilidade e as complexas causas externas por trás desses óbitos tornam-nos notoriamente difíceis de prever por modelos puramente estatísticos, que se baseiam em padrões históricos de tendência e sazonalidade.

É justamente neste ponto de dificuldade que os modelos de aprendizado profundo demonstram seu valor, alcançando os melhores desempenhos na métrica percentual sMAPE. Sua capacidade de capturar relações não lineares e padrões complexos parece ser mais adequada para modelar a dinâmica peculiar desta faixa etária. O modelo ETS, embora líder em métricas absolutas gerais, assim como os demais modelos clássicos, apresenta uma piora relativa neste período.

4.4.3 Grupo Total

A análise global confirma a robustez dos modelos estatísticos tradicionais. O modelo ETS obteve os melhores resultados em RMSE (0,000486) e MAE (0,000449), além do menor sMAPE (2,673830) entre os modelos individuais, consolidando-se como referência em todas as métricas. O modelo ARIMA apresentou desempenho muito próximo, com sMAPE de 2,696482. Dentre as combinações de modelos, a abordagem ponderada pelo inverso do erro superou consistentemente a combinação por média simples em todas as métricas, registrando RMSE de 0,000547, MAE de 0,000525 e sMAPE de 3,091398. O modelo híbrido CNN-GRU apresentou desempenho intermediário em sMAPE (3,120521), próximo ao das combinações de modelos. Entre os modelos de aprendizado profundo, o CNN teve o melhor sMAPE (4,459182), enquanto o GRU registrou os piores valores nessa métrica (5,862313) nesta categoria. O modelo Lee-Carter apresentou desempenho intermediário, com métricas superiores às dos modelos de *deep learning*, mas inferiores aos modelos estatísticos e combinações. O modelo FDM, por sua vez, confirmou sua limitação com o maior sMAPE entre os métodos tradicionais (5,000871). A Tabela 8 descreve as métricas médias obtidas por modelo para o grupo total (em destaque os menores erros de cada métrica). A Figura 22 demonstra a distribuição dos erros para cada um dos modelos aplicados na pesquisa.

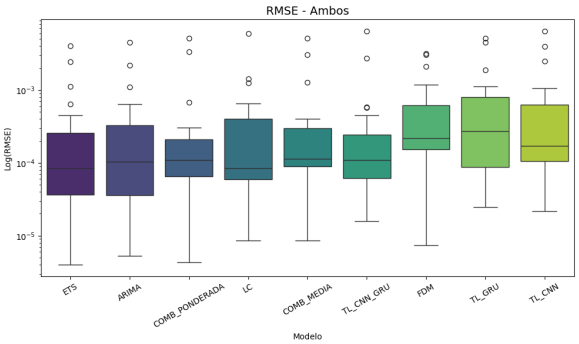
Tabela 8 – Resultados médios das métricas por modelo - Total

Modelo	RMSE	sMAPE	MAE
ETS	**0,000486**	**2,673830**	**0,000449**
ARIMA	0,000507	2,696482	0,000474
LC	0,000556	3,389550	0,000527
FDM	0,000667	5,000871	0,000637
TL_CNN	0,000858	4,459182	0,000814
TL_GRU	0,000843	5,862313	0,000819
TL_CNN_GRU	0,000600	3,120521	0,000570
COMB_MEDIA	0,000564	3,259116	0,000543
COMB_PONDERADA	0,000547	3,091398	0,000525

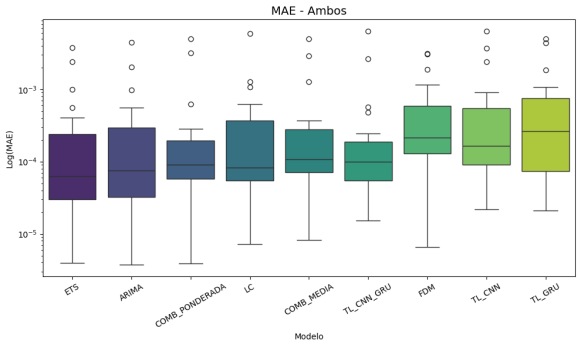
Fonte: Elaboração própria

Figura 22 – Distribuição dos Erros para Total

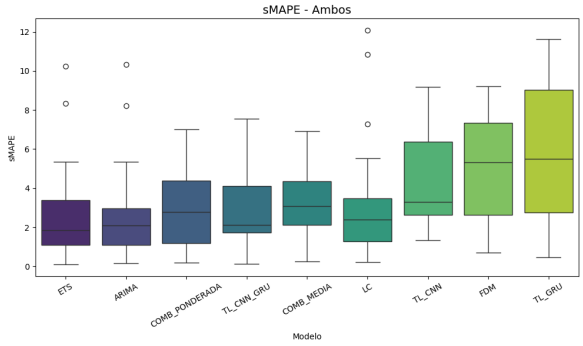
(a) Erros RMSE (log)



(b) Erros MAE (log)



(c) Erros sMAPE (%)



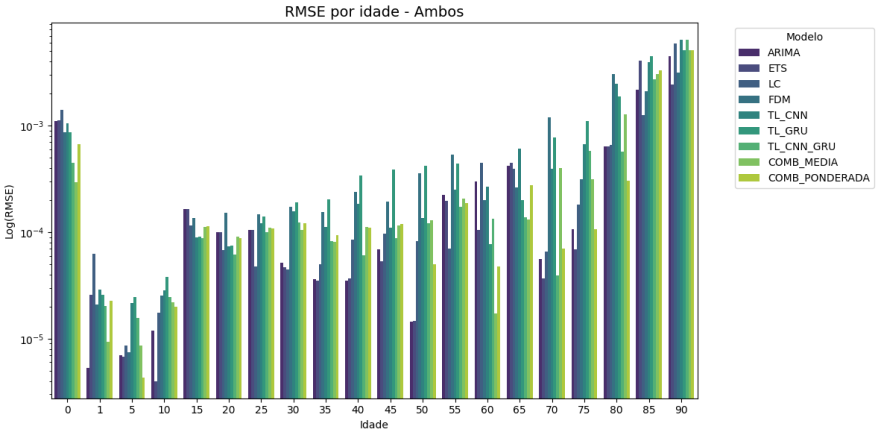
Fonte: Elaboração própria

A análise por grupo etário apresenta-se com semelhança ao observado nas previsões para o grupo de homens. Erros percentuais sMAPE maiores nas faixas de idades jovens, até 34 anos, com declínio dos erros percentuais conforme as idades avançam, até o grupo 75 a 79 anos. Observa-se que as abordagens propostas de transferência de aprendizado por meio de redes neurais e de combinação de preditores tiveram melhor performance percentual que os estatísticos

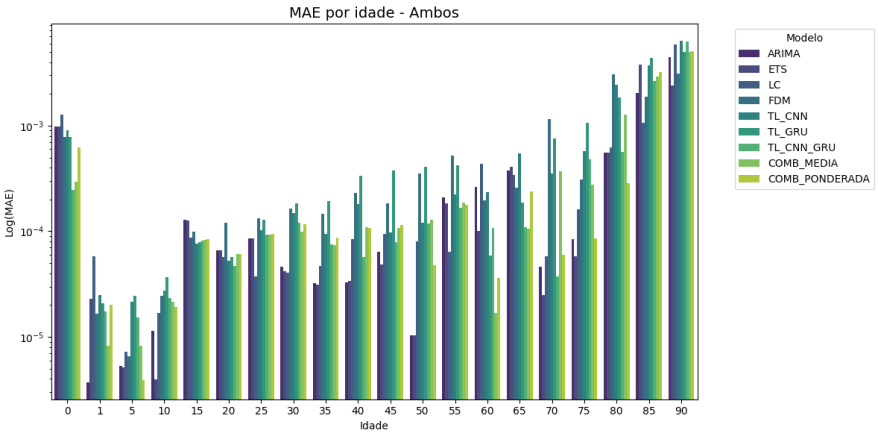
nos grupos etários 0, 15, 20 e 65. A Figura 23 demonstra a distribuição das métricas dos erros de cada modelo, conforme os grupos etários.

Figura 23 – Erros por Grupo Etário para Total

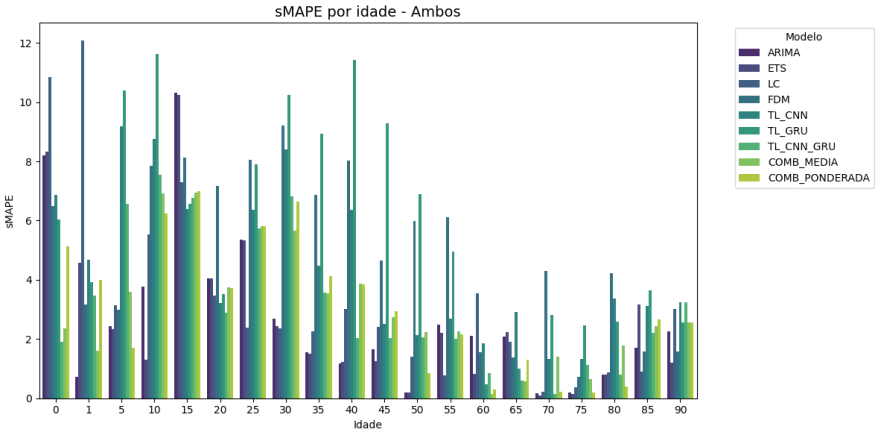
(a) Erros RMSE (log)



(b) Erros MAE (log)



(c) Erros sMAPE (%)



Fonte: Elaboração própria

A análise por grupo etário, apresentada na Figura 23, revela que o padrão de erros do grupo total é uma amalgama dos comportamentos observados em homens e mulheres, mas com uma nítida influência do perfil masculino, que tipicamente apresenta maiores taxas de mortalidade e variabilidade. O característico *accident hump*, marcado por erros percentuais elevados entre 15 e 34 anos, é claramente visível, confirmando a dificuldade universal dos modelos em prever a mortalidade por causas externas nesta faixa etária. Como observado nos grupos individuais, as abordagens mais modernas (redes neurais com transferência de aprendizado e combinações ponderadas) demonstraram resiliência superior nestas idades problemáticas.

4.5 DISCUSSÃO

A aplicação de modelos de previsão à realidade demográfica de um país deve considerar não apenas seu desempenho algorítmico, mas também as particularidades dos dados e as limitações teóricas dos modelos clássicos. Nesse sentido, as limitações inerentes ao modelo Lee-Carter, amplamente reconhecidas na literatura, são um ponto de partida crucial. Como destacado por Lee e Carter (1992), o modelo pode produzir inconsistências em taxas de mortalidade de jovens e apresentar instabilidade quando alimentado com séries históricas muito curtas, o que impacta diretamente a confiabilidade de suas estimativas.

Os resultados obtidos no presente estudo corroboram e ampliam essa percepção, demonstrando que a natureza peculiar dos dados de mortalidade no Brasil, submetidos a criteriosos ajustes metodológicos pelo IBGE, cria um ambiente onde tais limitações se manifestam claramente, ao mesmo tempo que abre espaço para abordagens mais robustas. A técnica de suavização por média móvel trianual, ao minimizar a variância intrínseca das séries e amortecer oscilações de curto prazo, propicia um cenário mais adequado para algoritmos que demandam estabilidade temporal. Esse fator contribui para elucidar a performance superior dos modelos ETS e ARIMA, cuja arquitetura é intrinsecamente voltada para a identificação de tendências suavizadas e de sazonalidade. Em nossos achados, o ETS estabelece-se como baseline de referência em todos os grupos (feminino, masculino e total), liderando simultaneamente em RMSE, MAE e sMAPE, enquanto o ARIMA aparece consistentemente como segundo colocado, com desempenho muito próximo em sMAPE no agregado.

Paralelamente, as correções aplicadas para mitigar o sub-registro, as quais variam conforme a Unidade da Federação, a faixa etária e o sexo, incorporam uma camada de complexidade que beneficia modelos dotados de maior flexibilidade para padrões heterogêneos. Neste aspecto, as redes neurais exibiram vantagem localizada, sobretudo na modelagem da mortalidade no primeiro ano de vida (idade 0), etapa em que os ajustes são mais expressivos e onde os modelos convencionais manifestaram maiores limitações. Em termos percentuais, as arquiteturas com transferência de aprendizado também mostraram resiliência em faixas etárias críticas, ainda que, no cômputo geral, tenham permanecido aquém do ETS e do ARIMA nas métricas agregadas.

A assimetria entre os padrões de mortalidade masculina e feminina, com esta última exibindo um perfil mais estável e homogêneo, repercute diretamente na acurácia preditiva. No grupo masculino, observou-se o "*accident hump*" com elevação de erros percentuais entre 15 e 34 anos, enquanto no feminino os erros percentuais foram mais contidos ao longo da distribuição etária. As redes neurais, em particular as arquiteturas GRU e CNN-GRU, mostraram vantagem relativa nessas idades problemáticas (idade 0 e *accident hump*), mas, de forma geral, não superaram os modelos ETS/ARIMA nas métricas globais por grupo. Assim, sua contribuição configura-se como complementar, reforçando a utilidade em pontos críticos da curva etária.

A consistência com que o modelo híbrido CNN-GRU superou as abordagens isoladas de CNN e GRU não se verificou como um padrão dominante em todas as métricas e grupos. Em nossos resultados, o desempenho entre as variantes de *deep learning* variou por sexo e métrica: por exemplo, o GRU se destacou entre as arquiteturas com transferência nas mulheres (RMSE/MAE), enquanto o híbrido CNN-GRU obteve o melhor sMAPE feminino entre modelos de *deep learning*. No masculino, o GRU foi o mais competitivo dentro da classe. Esses achados dialogam com literatura emergente sobre arquiteturas híbridas para séries temporais demográficas, que reporta ganhos contextuais ao combinar extração de padrões locais (CNN) com dependências temporais (GRU/LSTM), mas sem garantia de dominância global em todos os cenários.

Os índices de erro percentual relativos às arquiteturas de *deep learning* no presente estudo foram, em geral, superiores aos dos modelos ETS/ARIMA no agregado, embora apresentassem vantagens localizadas em idades críticas. Isso recomenda uma leitura cautelosa ao comparar com aplicações de outro domínio (por exemplo, COVID-19), nas quais (Zain; Alturki, 2021) relataram ganhos com CNN-LSTM. Tais evidências são congruentes quanto ao potencial de arquiteturas híbridas capturarem relações complexas, mas nossos achados indicam que, na mortalidade brasileira e no horizonte analisado, ETS e ARIMA preservam superioridade em sMAPE agregado, enquanto as redes neurais oferecem melhorias pontuais em segmentos etários desafiadores.

O desempenho robusto das redes neurais com transferência de aprendizado, embora não tenha superado o ETS/ARIMA no cômputo global, indica um avanço relevante na capacidade de captar não linearidades e padrões idiosiossincráticos em faixas etárias específicas. Se por um lado os modelos ETS e ARIMA mantiveram vantagem em RMSE, MAE e sMAPE nos resultados agregados (com destaque para mulheres e total), por outro, as redes neurais mostraram capacidade de generalização em idades iniciais e em segmentos masculinos de maior volatilidade, sugerindo um papel complementar valioso.

Uma análise crítica do desenho experimental empregado oferece insights para interpretar o desempenho relativo dos modelos. Conforme salientado, os modelos estatísticos clássicos (Lee-Carter, FDM, ETS, ARIMA) foram calibrados utilizando a totalidade do período de treino (2000–2015), enquanto as redes neurais demandaram a subdivisão em treino (2000–2011) e validação (2012–2015) para *fine-tuning* e prevenção de *overfitting*. Essa assimetria — que

confere maior volume de dados aos modelos clássicos para estimação — ajuda a contextualizar os resultados: onde as redes neurais superaram abordagens tradicionais localmente (idade 0; *accident hump* masculino), sua flexibilidade não linear compensou a desvantagem de menor janela de treino; já em cenários com dinâmica mais suave (particularmente no feminino), a parcimônia e a maior amostra efetiva favoreceram ETS e ARIMA, resultando em liderança nas métricas agregadas.

A superação de Lee-Carter por redes neurais não se configurou de forma ampla nos resultados agregados deste estudo, mas observou-se vantagem das arquiteturas de *deep learning* em janelas etárias específicas. Isso nuança a evidência de Chen e Khaliq (2022), segundo a qual GRU, LSTM e BiLSTM superaram Lee-Carter em parcela substancial das séries em MAE e RMSE. No nosso caso, Lee-Carter teve desempenho intermediário, frequentemente inferior ao ETS/ARIMA e, em várias situações, comparável ou superior às redes neurais nos agregados, embora perdesse localmente em idades críticas. Em linha com Levantesi, Nigri e Piscopo (2021), a integração de componentes neuronais à estrutura de Lee-Carter pode melhorar a capacidade preditiva, mas nossos resultados sugerem que, para o horizonte curto e o contexto brasileiro, a superioridade não é generalizada e depende da faixa etária e do sexo. Em síntese, os achados confirmam parcialmente a literatura: as RNAs superam LC em contextos específicos, mas ETS/ARIMA permanecem referências no agregado.

A estratégia de transferência de aprendizado, que utilizou dados de nações com perfis de mortalidade similares para inicializar os modelos no contexto brasileiro, mostrou-se útil para estabilizar o treinamento e melhorar o desempenho relativo das RNAs em idades críticas. Esta abordagem encontra respaldo em desenvolvimentos recentes, como o *framework* de *transfer learning* para mortalidade proposto por Nalmpatian *et al.* (2025), e nos métodos de "*borrowing strength*" entre unidades com perfis demográficos semelhantes em Ludkovski e Padilla (2023). Tais referências são coerentes com nossos resultados ao apontarem benefícios em cenários de dados limitados ou heterogêneos, ainda que, no presente estudo, tais ganhos não tenham sido suficientes para superar ETS/ARIMA nas métricas globais.

Um achado particularmente relevante para a discussão metodológica diz respeito ao desempenho comparativo entre o *Functional Demographic Model* (FDM) e o Lee-Carter no horizonte de curto prazo. Conforme documentado, o FDM representa uma generalização funcional do paradigma Lee-Carter, incorporando decomposição por componentes principais funcionais e procedimentos explícitos de suavização ao longo das idades (Gao; Shi, 2021). Esta arquitetura confere maior robustez para capturar padrões estruturais de longo prazo e lidar com variabilidade complexa e anos atípicos (Booth; Tickle, 2008). Contudo, o mesmo mecanismo de suavização pode representar desvantagem em janelas curtas, amortecendo sinais recentes (Gao; Shi, 2021). Nossos resultados confirmam essa leitura: no horizonte de quatro anos, o FDM apresentou desempenho inferior a LC e, notadamente, a ETS/ARIMA em todas as métricas agregadas.

Em contrapartida, a estrutura paramétrica relativamente simples do Lee-Carter, teoricamente equivalente a um passeio aleatório com *drift* (Girosi; King, 2007), confere ao modelo maior sensibilidade a variações recentes. Essa característica pode favorecer previsões de curto prazo, ainda que com maior suscetibilidade a ruídos. A evidência empírica reforça que a seleção do modelo ótimo deve considerar o horizonte temporal de interesse (Bergeron-Boucher; Kjærgaard, 2022): para curtos prazos, modelos sensíveis a variações recentes (como LC/ETS/ARIMA) tendem a levar vantagem; para horizontes mais longos, abordagens com suavização e componentes funcionais (como FDM) podem ser preferíveis. Esse princípio alinha-se com o padrão observado neste estudo.

Por fim, os resultados relativos à combinação de modelos merecem análise à luz da literatura. Não observamos uma dominância uniforme de um esquema de combinação sobre o outro: no grupo feminino, a média simples superou a ponderada em RMSE e MAE e foi praticamente equivalente em sMAPE; no masculino, a ponderada superou a média em RMSE/MAE, porém com sMAPE ligeiramente superior; no total, a ponderada foi melhor que a média simples em todas as métricas. À primeira vista, isso dialoga com o conhecido "*forecast combination puzzle*" — a robustez da média simples frente a ponderações mais complexas (Claeskens *et al.*, 2016), atribuída ao risco de erros de estimação e *overfitting* (Qian *et al.*, 2019; Frazier *et al.*, 2023). Em nosso contexto, a heterogeneidade por idade/sexo e a complexidade das séries tornam a informação contida no desempenho histórico potencialmente útil, mas sua exploração via pesos não se traduziu em melhorias percentuais consistentes no agregado. Em suma, não há "quebra" generalizada do *puzzle*: os ganhos da ponderação foram mais evidentes em erros absolutos (RMSE/MAE) em homens e no total, enquanto a média simples mostrou competitividade em termos percentuais (sMAPE), especialmente em mulheres e no agregado.

Esse quadro ainda é compatível com evidências de efetividade da ponderação por erro inverso em domínios específicos, como meteorologia e macroeconomia (Sun; Yin; Zhao, 2017). A nuance principal é contextual: em séries de mortalidade com alta variabilidade seletiva (idade 0; *accident hump*) e forte heterogeneidade, a ponderação pode capturar sinais úteis e reduzir erros absolutos, sem necessariamente otimizar métricas percentuais. Isso sugere diferenciar objetivos de "combinação para adaptação" vs. "combinação para melhoria" (Qian *et al.*, 2019). A metodologia aqui implementada combinou todos os modelos disponíveis, abstendo-se de seleção prévia, o que pode ser um fator, uma vez modelos com baixo desempenho compuseram a agregação.

5 CONCLUSÃO

Este capítulo apresenta as Considerações Finais do trabalho (Seção 5.1), com os seus principais achados diante da pergunta de pesquisa formulada. Também são mencionadas as limitações metodológicas da pesquisa (Seção 5.2. Por fim, são listados potenciais trabalhos futuros que envolvam o tema e enfrentem as limitações já citadas (Seção 5.3).

5.1 CONSIDERAÇÕES FINAIS

Este trabalho propôs abordagens para superar os desafios inerentes à previsão de taxas de mortalidade no contexto brasileiro, caracterizado por séries temporais curtas. Os resultados demonstraram que modelos estatísticos clássicos, como o ETS e o ARIMA, estabeleceram-se como o baseline mais robusto, liderando consistentemente nas métricas agregadas (RMSE, MAE e sMAPE) para todos os grupos (feminino, masculino e total). Métodos de aprendizado profundo com transferência de aprendizado, embora não superassem os modelos clássicos no desempenho global, mostraram-se particularmente eficazes na captura de padrões não-lineares e na modelagem de grupos populacionais e etários mais heterogêneos, como a mortalidade infantil (idade 0) e o fenômeno do "*accident hump*"nas idades dos homens jovens, onde modelos estatísticos tradicionais apresentaram maiores dificuldades de predição.

A combinação de preditores heterogêneos também se mostrou uma ferramenta relevante, mas com nuances. A ponderação pelo inverso dos erros foi promissora na redução de erros absolutos (RMSE e MAE) para os grupos masculino e total, superando a média simples. Contudo, a média simples manteve-se competitiva em termos de sMAPE, especialmente no grupo feminino e no agregado, indicando que a superioridade da ponderação não foi uniforme em todas as métricas e grupos. Esta estratégia, no entanto, apontou um caminho para captar os pontos fortes de cada modelo individual em contextos específicos, mitigando suas limitações.

Os resultados indicam que a integração metodológica proposta, com o uso de transferência de aprendizado e redes neurais recorrentes e convolucionais, complementa a robustez dos modelos estatísticos tradicionais. A combinação de preditores, com abordagem de ponderação baseada em desempenho, oferece um caminho promissor para produzir projeções de mortalidade mais eficientes para contextos de séries temporais curtas, como os dados brasileiros, podendo contribuir na melhoria de processos atuariais e de formulação de políticas públicas, especialmente ao considerar a heterogeneidade etária e de sexo.

5.2 LIMITAÇÕES DA PESQUISA

É importante reconhecer que este trabalho, como qualquer investigação científica, apresenta limitações que devem ser consideradas na interpretação de seus resultados. Em primeiro lugar, a análise esteve restrita ao uso da distância Manhattan para os cálculos de similaridade, não tendo sido realizados testes de sensibilidade com outras métricas de similaridade utilizadas em séries temporais, o que poderia influenciar a seleção das populações de referência. Além disso, não foi investigado como o desempenho do modelo se comporta em função do número de países selecionados para compor a população de referência, deixando de explorar esse parâmetro crítico para a calibração do método. Ademais, o trabalho não incorporou modelos demográficos teóricos de previsão para comparação, restringindo-se a modelos estatísticos de séries temporais e estatísticos aplicados a dados demográficos. Também não foram empregadas técnicas de seleção de modelos ou outras abordagens de ponderação antes da etapa de combinação das previsões. Outra limitação reside no horizonte temporal analisado, que se concentrou em um período de teste de curto, não permitindo avaliar a coerência e o comportamento das projeções em horizontes de longo prazo. Aplicações práticas no contexto atuarial, como a construção de tábuas de mortalidade dinâmicas para cálculo de seguros e anuidades, também ficaram fora do escopo deste trabalho. Por fim, vale destacar que a pesquisa utilizou dados massivamente ajustados e suavizados disponibilizados pelo IBGE, os quais, embora conferirem robustez, podem mascarar a volatilidade inerente aos dados vitais brutos e alterar o desempenho preditivo dos modelos. O custo computacional do método proposto também não foi mensurado formalmente, uma métrica relevante para sua viabilidade operacional.

5.3 TRABALHOS FUTUROS

Para trabalhos futuros, que possam aprofundar na análise dos achados da presente pesquisa, propõe-se um amplo leque de investigações. Inicialmente, pode-se replicar estas metodologias utilizando dados brutos de óbitos e expostos aos riscos, coletados diretamente dos registros, a fim de verificar o seu desempenho em um cenário com maior variabilidade. É possível também ampliar o escopo de análise para incluir um maior número de populações e subpopulações, como regiões, unidades federativas ou ainda outros países com diferentes tamanhos de séries temporais, a fim de testar a robustez e a generalidade do modelo em diversos contextos demográficos. Um caminho natural seria a construção de tábuas de mortalidade regionais ou mesmo tábuas por causa específica de morte, o que permitiria uma análise mais granular dos padrões de mortalidade.

A aplicação em outras populações com séries temporais de taxas de mortalidade com mais observações é recomendada para investigar a acurácia das previsões em horizontes de tempo mais longos. Especificamente, recomenda-se investigar o uso de outras métricas de distância ou mesmo um *ensemble* de métricas para a seleção das populações de referência, a fim de verificar

a sensibilidade dos resultados em relação a esse critério fundamental. Adicionalmente, para uma avaliação de previsão mais robusta, sugere-se avaliar outras métricas de acurácia, tais como o MASE (*Mean Absolute Scaled Error*) ou intervalos de previsão (e.g., *Mean Interval Score - MIS*), que capturam diferentes aspectos do erro e da incerteza preditiva.

Adicionalmente, um caminho frutífero de investigação seria tratar o número de países similares como um hiperparâmetro a ser otimizado, analisando empiricamente a curva de aprendizado e o *trade-off* entre a quantidade de populações incluídas e o ganho marginal de desempenho preditivo. Para enriquecer a modelagem, a incorporação de modelos hierárquicos bayesianos surge como uma alternativa promissora, pois permite modelar naturalmente a estrutura de agrupamento dos dados (como países ou regiões) e incorporar covariáveis socioeconômicas diretamente no processo de previsão.

Além disso, os resultados levantam uma questão fundamental: quais *drivers* socioeconômicos, de saúde ou ambientais explicam a similaridade entre as trajetórias de mortalidade? Cruzar os *clusters* de países similares com esses indicadores permitiria transcender a análise preditiva, investigando as causas subjacentes dos padrões de declínio da mortalidade e conferindo profundidade analítica ao método.

Recomenda-se também a expansão do uso de técnicas de *ensemble* para as previsões. O presente trabalho já deu um passo inicial nessa direção ao empregar uma combinação por ponderação pelo inverso dos erros no conjunto de validação. Futuras investigações podem explorar a constituição de conjuntos de confiança (MCS) para descartar a priori modelos com performance inferior, seguida da combinação ponderada dos remanescentes, buscando ganhos adicionais de robustez e acurácia com base em significância estatística (Chang; Shi, 2023; García-Aroca *et al.*, 2023; Hansen; Lunde; Nason, 2011; Bravo; Ayuso, 2020). Destaca-se também o uso de *stacking* (ou *blending*) com um meta-modelo treinável (como uma regressão linear ou regularizada, ou mesmo uma árvore de decisão), que, ao invés de usar uma regra fixa de ponderação, aprende a combinar de forma ótima e potencialmente não-linear as previsões dos modelos base. Adicionalmente, outras técnicas como *bagging* e *boosting* (ex: *XGBoost*, *LightGBM*) aplicadas diretamente à série temporal de mortalidade ou aos dados funcionais merecem investigação. Uma análise comparativa abrangente entre essas diferentes abordagens de *ensemble* e o método já implementado aprofundaria o entendimento sobre a melhor forma de agregar a informação de múltiplos modelos para aumentar a acurácia e a robustez das previsões de mortalidade.

Por fim, tábuas biométricas geracionais produzidas por meio dos diversos modelos preditivos podem ser usadas para avaliar o impacto atuarial em precificações de produtos e cálculos das provisões técnicas, permitindo uma comparação prática entre as abordagens propostas.

REFERÊNCIAS

ABADI, M. *et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. 2015. Software available from tensorflow.org. Disponível em: <https://www.tensorflow.org/>.

ANDONIE, R.; FLOREA, A.-C. Weighted random search for cnn hyperparameter optimization. *INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS amp; CONTROL*, Agora University of Oradea, v. 15, n. 2, mar. 2020. ISSN 1841-9836. Disponível em: <http://dx.doi.org/10.15837/ijccc.2020.2.3868>.

ANDRS-SÁNCHEZ, J. D.; PUCHADES, L. G.-V. A fuzzy-random extension of the lee–carter mortality prediction model. *International Journal of Computational Intelligence Systems*, 2019. Disponível em: <https://doi.org/10.2991/ijcis.d.190626.001>.

APICELLA, G. *et al.* Neural network lee–carter model and the actuarial relevance of longevity risk assessment. *Scandinavian Actuarial Journal*, p. 1–25, 2024.

BARIGOU, K. *et al.* Bayesian model averaging for mortality forecasting using leave-future-out validation. *International Journal of Forecasting*, 2022.

BASELLINI, U.; CAMARDA, C. G.; BOOTH, H. Thirty years on: A review of the lee–carter method for forecasting mortality. *International Journal of Forecasting*, v. 39, n. 3, p. 1033–1049, 2023. ISSN 0169-2070. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0169207022001455>.

BERGERON-BOUCHER, M.-P.; KJæRGAARD, S. Mortality forecasting at age 65 and above: an age-specific evaluation of the lee-carter model. *Scandinavian Actuarial Journal*, Taylor & Francis, v. 2022, n. 1, p. 64–79, 2022. Disponível em: <https://doi.org/10.1080/03461238.2021.1928542>.

BERKUM, F. V.; ANTONIO, K.; VELLEKOOP, M. A bayesian joint model for population and portfolio-specific mortality. *ASTIN Bulletin*, v. 47, n. 3, p. 681–713, 2017.

BIMONTE, G. *et al.* Mortality models ensemble via shapley value. *Decisions in Economics and Finance*, 2024.

BLUME, S.; BENEDENS, T.; SCHRAMM, D. Hyperparameter optimization techniques for designing software sensors based on artificial neural networks. *Sensors*, v. 21, n. 24, 2021. ISSN 1424-8220. Disponível em: <https://www.mdpi.com/1424-8220/21/24/8435>.

BOOTH, H.; TICKLE, L. Mortality modelling and forecasting: a review of methods. *Annals of Actuarial Science*, v. 3, n. 1–2, p. 3–43, 2008.

BRAVO, J. *Tábuas de mortalidade contemporâneas e prospectivas: Modelos estocásticos, aplicações actuariais e cobertura do risco de longevidade*. Tese (Doutorado) — Universidade de Évora, Évora, 2007. Tese de Doutoramento.

BRAVO, J. M.; AYUSO, M. Previsões de mortalidade e de esperança de vida mediante combinação bayesiana de modelos: Uma aplicação à população portuguesa. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, n. 40, p. 128–145, 2020.

BRAVO, J. M.; SANTOS, V. Backtesting recurrent neural networks with gated recurrent unit: probing with chilean mortality data. In: BRAVO, J. M.; SANTOS, V. (Ed.). *Lecture Notes in Networks and Systems*. Cham: Springer International Publishing, 2022. p. 159–174. ISBN 9783030977184.

CERQUEIRA, D.; COELHO, D. *O que pode explicar a queda de homicídios no Brasil em 2019?* 2019. <https://blogdoibre.fgv.br/posts/o-que-pode-explicar-queda-de-homicidios-no-brasil-em-2019>. Blog do IBRE - FGV, acessado em 31 de agosto de 2025.

CERQUEIRA, V.; TORGO, L.; SOARES, C. A case study comparing machine learning with statistical methods for time series forecasting: size matters. *Journal of Intelligent Information Systems*, 2022. Online 16 May.

CHANG, L.; SHI, Y. Forecasting mortality rates with a coherent ensemble averaging approach. *ASTIN Bulletin: The Journal of the IAA*, Cambridge University Press, v. 53, n. 1, p. 2–28, 2023. Disponível em: <https://www.cambridge.org/core/journals/astin-bulletin-journal-of-the-iaa/article/forecasting-mortality-rates-with-a-coherent-ensemble-averaging-approach/00AFDF6884D321EEE6DF99A471D3DC8E>.

CHEN, Y.; KHALIQ, A. Q. M. Comparative study of mortality rate prediction using data-driven recurrent neural networks and the lee–carter model. *Big Data and Cognitive Computing*, v. 6, n. 4, p. 134, 2022.

CHENG, H. *et al.* Multistep-ahead time series prediction. In: *Advances in Knowledge Discovery and Data Mining*. Springer Berlin Heidelberg, 2006. p. 765–774. ISBN 9783540332060. Disponível em: <https://doi.org/10.1007/11731139_89>.

CHOLLET, F. *et al. Keras*. 2015. <https://keras.io>.

CLAESKENS, G. *et al.* The forecast combination puzzle: A simple theoretical explanation. *International Journal of Forecasting*, v. 32, n. 3, p. 754–762, 2016. ISSN 0169-2070. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0169207016000327>.

COCCO, J. F.; GOMES, F. J. Longevity risk, retirement savings, and financial innovation. *Journal of Financial Economics*, Elsevier, v. 103, n. 3, p. 507–529, 2012.

CRUZ-NÁJERA, M. A. *et al.* Short time series forecasting: Recommended methods and techniques. *Symmetry*, v. 14, n. 6, p. 1231, 2022. Disponível em: <https://doi.org/10.3390/sym14061231>.

DIAO, L. *et al.* Enhancing mortality forecasting through bivariate model–based ensemble. *North American Actuarial Journal*, p. 1–20, 2023. Disponível em: <https://doi.org/10.1080/10920277.2023.2167832>.

DUARTE, E. C. *et al.* Mortalidade de jovens no brasil: perfil e tendências no período 2000-2012. *Epidemiologia e Serviços de Saúde*, v. 24, n. 4, p. 595–606, 2015. Disponível em: <http://scielo.iec.gov.br/scielo.php?script=sci_arttext&pid=S1679-49742015000400002&lng=en&nrm=is&tlng=en>.

DUARTE, F. C. de L. *Um sistema híbrido baseado em combinação de preditores para previsão de vários passos à frente de séries temporais de taxas de mortalidade*. Tese (Doutorado) —

Universidade Federal de Pernambuco, Recife, 2024. Tese de doutorado, 137 p. Disponível em: <https://repositorio.ufpe.br/bitstream/123456789/59159/1/TESE%20Filipe%20Coelho%20de%20Lima%20Duarte.pdf>.

DUARTE, F. C. de L.; NETO, P. S. G. de M.; FIRMINO, P. R. A. A hybrid recursive direct system for multi-step mortality rate forecasting. *The Journal of Supercomputing*, v. 13, p. 18430–18463, 2024.

EKHEDEN, E.; HÄ¶SSJER, O. Multivariate time series modeling, estimation and prediction of mortalities. *Insurance: Mathematics and Economics*, v. 65, n. C, p. 156–171, None 2015. Disponível em: <https://ideas.repec.org/a/eee/insuma/v65y2015icp156-171.html>.

EMMERT-STREIB, F.; DEHMER, M. Taxonomy of machine learning paradigms: A data-centric perspective. *WIREs Data Mining and Knowledge Discovery*, v. 12, n. 5, 2022.

FANG, L.; HÄRDLE, W. K. *Stochastic population analysis: A functional data approach*. Berlin, 2015. Disponível em: <https://hdl.handle.net/10419/107909>.

FAWAZ, H. I. *et al.* Transfer learning for time series classification. In: *2018 IEEE International Conference on Big Data (Big Data)*. Seattle, WA, USA: IEEE, 2018. ISBN 9781538650356.

FELTRAN, G. Variações nas taxas de homicídios no brasil: Uma explicação centrada na dinâmica do conflito criminal. *Dilemas: Revista de Estudos de Conflito e Controle Social*, v. 13, n. 3, p. 567–589, 2020. Disponível em: <https://www.scielo.br/j/dilemas/a/37drXYFwTK9hD9rysdFzqzm/>.

FENG, L.; SHI, Y. Forecasting mortality rates: multivariate or univariate models? *Journal of Population Research*, v. 35, n. 3, p. 289–318, 2018.

FRAZIER, D. T. *et al. Solving the Forecast Combination Puzzle*. arXiv, 2023. Disponível em: <https://arxiv.org/abs/2308.05263>.

Fórum Brasileiro de Segurança Pública. *Anuário Brasileiro de Segurança Pública 2019*. [S.l.], 2019. Acessado em 31 de agosto de 2025. Disponível em: <https://forumseguranca.org.br/wp-content/uploads/2019/10/Anuario-2019-FINAL_21.10.19.pdf>.

GAO, G.; SHI, Y. Age-coherent extensions of the lee–carter model. *Scandinavian Actuarial Journal*, Taylor & Francis, v. 2021, n. 10, p. 998–1016, 2021. Disponível em: <https://doi.org/10.1080/03461238.2021.1918578>.

GARCÍA-AROCA, C. *et al.* An algorithm for automatic selection and combination of forecast models. *Expert Systems with Applications*, p. 121636, 2023.

GERSTEN, O.; BARBIERI, M. Evaluation of the cancer transition theory in the us, select european nations, and japan by investigating mortality of infectious- and noninfectious-related cancers, 1950-2018. *JAMA Network Open*, v. 4, n. 4, p. e215322–e215322, 04 2021. ISSN 2574-3805. Disponível em: <https://doi.org/10.1001/jamanetworkopen.2021.5322>.

GIROSI, F.; KING, G. Understanding the lee-carter mortality forecasting method. *Gking. Harvard. Edu*, 2007.

GOLDSTEIN, J. R. A secular trend toward earlier male sexual maturity: Evidence from shifting ages of male young adult mortality. *PloS one*, Public Library of Science San Francisco, USA, v. 6, n. 8, p. e14826, 2011.

GUJARATI, D. N.; PORTER, D. C. *Econometria básica*. 5. ed. São Paulo: McGraw-Hill, 2011.

GUPTA, P. *et al.* Transfer learning for clinical time series analysis using deep neural networks. *Journal of Healthcare Informatics Research*, v. 4, n. 2, p. 112–137, 2019.

GYAMERAH, S. A. *et al.* Improving mortality forecasting using a hybrid of lee–carter and stacking ensemble model. *Bulletin of the National Research Centre*, v. 47, n. 1, 2023.

HANSEN, P. R.; LUNDE, A.; NASON, J. M. The model confidence set. *Econometrica*, Wiley Periodicals LLC on behalf of The Econometric Society, v. 79, n. 2, p. 453–497, 2011. Disponível em: <https://www.econometricsociety.org/publications/econometrica/2011/03/01/model-confidence-set>.

HARRIS, C. R. *et al.* Array programming with numpy. *Nature*, v. 585, n. 7825, p. 357–362, 2020.

HELIGMAN, L.; POLLARD, J. H. The age pattern of mortality. *Journal of the Institute of Actuaries*, Cambridge University Press, v. 107, n. 1, p. 49–80, 1980.

HONG, W. H. *et al.* Forecasting mortality rates using hybrid lee–carter model, artificial neural network and random forest. *Complex & Intelligent Systems*, 2020.

HOSSEINI, F.; PRIETO, C.; ÁLVAREZ, C. Hyperparameter optimization of regional hydrological lstms by random search: A case study from basque country, spain. *Journal of Hydrology*, v. 643, p. 132003, 2024. ISSN 0022-1694. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0022169424013994>.

Human Mortality Database. *The Human Mortality Database*. 2025. <https://www.mortality.org>. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany).

HYNDMAN, R. *demography: Forecasting Mortality, Fertility, Migration and Population Data*. 2024. R package version 2.0.0.9000, <https://github.com/robjhyndman/demography>, <https://pkg.robjhyndman.com/demography/>.

HYNDMAN, R. *et al. forecast: Forecasting functions for time series and linear models*. 2024. R package version 8.23.0, <https://pkg.robjhyndman.com/forecast/>.

HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting: principles and practice*. 3. ed. Melbourne: OTexts, 2021. Disponível em: <https://OTexts.com/fpp3>.

HYNDMAN, R. J.; BOOTH, H.; YASMEEN, F. Coherent mortality forecasting: the product-ratio method with functional time series models. *Demography*, Duke University Press, v. 50, n. 1, p. 261–283, 2013.

HYNDMAN, R. J.; ULLAH, M. S. Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis*, v. 51, n. 10, p. 4942–4956, 2007.

IFEANYICHUKWU, U. C. *et al.* On forecasting infant mortality rate by sex using arima model: a case of nigeria. *European Journal of Statistics and Probability*, v. 10, n. 2, p. 29–38, 2022. Disponível em: <https://doi.org/10.37745/ejsp.2013/vol10n22938>.

Instituto Brasileiro de Geografia e Estatística (IBGE). *Projeções da População: Brasil e Unidades da Federação: Estimativas e Projeções — Revisão 2024. Notas metodológicas 01/2024*. Revisão 2024. Rio de Janeiro, 2024. Diretoria de Pesquisas; Coordenação de População e Indicadores Sociais. © IBGE 2024. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv102111.pdf>.

Instituto Brasileiro de Geografia e Estatística (IBGE). *Projeção da população*. 2025. <https://www.ibge.gov.br/estatisticas/sociais/populacao/9109-projecao-da-populacao.html?edicao=41053>.

Instituto de Pesquisa Econômica Aplicada; Fórum Brasileiro de Segurança Pública. *Atlas da Violência 2020*. [S.l.], 2020. Acessado em 31 de agosto de 2025. Disponível em: <https://www.ipea.gov.br/atlasviolencia/arquivos/artigos/3519-atlasdaviolencia2020completo.pdf>.

JING, L. *et al.* Gated orthogonal recurrent units: on learning to forget. *Neural Computation*, v. 31, n. 4, p. 765–783, 2019.

JÚNIOR, D. S. de O. S. *Método de ensemble para correção de modelos ARIMA: uma abordagem de sistema híbrido para previsão de séries temporais*. 80 p. Tese (Tese (Doutorado em Estatística)) — Universidade Federal de Pernambuco, Recife, PE, Brasil, 2022. Disponível em: <https://repositorio.ufpe.br/bitstream/123456789/45606/1/TESE%20Domingos%20S%C3%A1vio%20de%20Oliveira%20Santos%20J%C3%BAnior.pdf>.

JUNIOR, J. A. *Estudo da influência de diversas medidas de similaridade na previsão de séries temporais utilizando o algoritmo kNN-TSP*. Dissertação (Mestrado) — Universidade Estadual do Oeste do Paraná, Programa de Pós-Graduação em Engenharia de Sistemas Dinâmicos e Energéticos, Foz do Iguaçu, 2012. 143 f. — Dissertação (Mestrado). Disponível em: <http://tede.unioeste.br/handle/tede/1084>.

JÚNIOR, L. C. S.; AZEVEDO, F. I. X. de; TSUNEMI, M. H. Efeitos da mortalidade geral brasileira sobre o cálculo atuarial: uma comparação entre modelos preditivos. *Revista Evidenciação Contábil & Finanças*, Universidade Federal da Paraíba, v. 7, n. 2, p. 79–101, 2019.

KANDEL, I.; CASTELLI, M.; POPOVIč, A. Comparative study of first order optimizers for image classification using convolutional neural networks on histopathology images. *Journal of Imaging*, v. 6, n. 9, 2020. ISSN 2313-433X. Disponível em: <https://www.mdpi.com/2313-433X/6/9/92>.

KESSY, S. R. *et al.* Mortality forecasting using stacked regression ensembles. *Scandinavian Actuarial Journal*, p. 1–36, 2021.

LARSSON, T.; VERMEIRE, F.; VERHELST, S. Machine learning for fuel property predictions: A multi-task and transfer learning approach. In: *WCX SAE World Congress Experience*. Detroit, Michigan, United States: SAE International, 2023.

LEE, R. D.; CARTER, L. R. Modeling and forecasting u.s. mortality. *Journal of the American Statistical Association*, v. 87, n. 419, p. 659–671, 1992.

LEVANTESI, S.; NIGRI, A.; PISCOPO, G. Improving longevity risk management through machine learning. In: *The essentials of machine learning in finance and accounting*. [S.l.]: Routledge, 2021. p. 37–56.

LEVANTESI, S.; PIZZORUSSO, V. Application of machine learning to mortality modeling and forecasting. *Risks*, v. 7, n. 1, p. 26, 2019.

LI, J. An application of mcmc simulation in mortality projection for populations with limited data. *Demographic Research*, v. 30, p. 1–48, 2014.

LI, J. A model stacking approach for forecasting mortality. *North American Actuarial Journal*, p. 1–16, 2022.

LI, L.; LI, H.; PANAGIOTELIS, A. Boosting domain-specific models with shrinkage: An application in mortality forecasting. *International Journal of Forecasting*, May 2024.

LI, X.; GRANDVALET, Y.; DAVOINE, F. Explicit inductive bias for transfer learning with convolutional networks. In: *International Conference on Machine Learning*. [S.l.]: PMLR, 2018. p. 2825–2834.

LIU, L.; AL. et. Multi-task learning via adaptation to similar tasks for mortality prediction of diverse rare diseases. In: *AMIA Annual Symposium Proceedings*. [s.n.], 2021. p. 763. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/33936451/>.

LUDKOVSKI, M.; PADILLA, D. Analyzing state-level longevity trends with the u.s. mortality database. *Annals of Actuarial Science*, Cambridge University Press, v. 17, n. 1, p. 1–25, 2023.

LÜTKEPOHL, H.; XU, F. The role of the log transformation in forecasting economic variables. *Empirical Economics*, v. 42, n. 3, p. 619–638, 2010.

MACÊDO, D.; ZANCHETTIN, C.; LUDERMIR, T. Sigmoidal learning rate optimizer for deep neural network training using a two-phase adaptation approach. *Applied Soft Computing*, v. 167, p. 112264, 2024. ISSN 1568-4946. Disponível em: <https://www.sciencedirect.com/science/article/pii/S156849462401038X>.

MALTA, D. C. *et al.* Association between firearms and mortality in brazil, 1990 to 2017: a global burden of disease brazil study. *Population Health Metrics*, v. 18, n. Suppl 1, p. 19, 2020. Disponível em: <https://pophealthmetrics.biomedcentral.com/articles/10.1186/s12963-020-00222-3>.

MARINO, M.; LEVANTESI, S.; NIGRI, A. *Deepening Lee-Carter for longevity projections with uncertainty estimation*. 2021.

MARINO, M.; LEVANTESI, S.; NIGRI, A. A neural approach to improve the lee-carter mortality density forecasts. *North American Actuarial Journal*, Taylor & Francis, v. 27, n. 1, p. 148–165, 2023.

MCQUIRE, P.; KUME, A. *R Programming for Actuarial Science*. [S.l.]: Wiley & Sons, Incorporated, John, 2022. ISBN 9781119755005.

MORAIS, G.; ESCOVEDO, T.; KALINOWSKI, M. Machine learning applied in the construction of a disability retirement entry table of the general social security regime (rgps) of brazil. In: *SBSI '24: XX Brazilian Symposium on Information Systems*. New York, NY, USA: ACM, 2024.

MURRAY, J.; CERQUEIRA, D. R. C.; KAHN, T. Crime and violence in brazil: Systematic review of time trends, prevalence rates and risk factors. *Aggression and Violent Behavior*, v. 18, n. 5, p. 471–483, 2013. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC3763365/>.

NALMPATIAN, A.; HEUMANN, C.; PILZ, S. Advanced techniques in mortality trend estimation: Integrating generalized additive models and machine learning to evaluate the covid-19 impact. *arXiv preprint arXiv:2311.15401*, 2023.

NALMPATIAN, A. *et al.* Local and global mortality experience: A novel hierarchical model for regional mortality risk. *medRxiv*, p. 2024.10.17.24315673, 2024.

NALMPATIAN, A. *et al.* Transfer learning for mortality risk: A case study on the united kingdom. *PLOS One*, v. 20, n. 5, p. e0313378, 2025.

NASCIMENTO, J. D.; ESCOVEDO, T. Construction of mortality tables using lstm neural networks. In: *SBSI 2021: XVII Brazilian Symposium on Information Systems*. Uberlândia, Brazil: ACM, 2021.

NIGRI, A.; AL. et. A deep learning integrated lee–carter model. *Risks*, v. 7, n. 1, p. 33, 2019.

NOR, S. R. B. M. *et al.* Modelling and forecasting the covid-19 mortality rates in malaysia by using arima model. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, v. 45, n. 1, p. 215–223, 2024.

O'MALLEY, T.; AL. et. *KerasTuner*. 2019. Disponível em: <https://github.com/keras-team/keras-tuner>.

PAN, S. J.; YANG, Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, v. 22, n. 10, p. 1345–1359, 2010.

PARMEZAN, A. R. S. *Predição de séries temporais por similaridade*. 219 p. Dissertação (Dissertação (Mestrado em Ciências de Computação e Matemática Computacional)) — Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, SP, Brasil, 2016.

PEDREGOSA, F.; AL. et. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011. Disponível em: <https://jmlr.org/papers/v12/pedregosa11a.html>.

PERLA, F. *et al.* Time-series forecasting of mortality rates using deep learning. *Available at SSRN 3595426*, 2020. Disponível em: <https://dx.doi.org/10.2139/ssrn.3595426>.

PETNEHÁZI, G.; GÁLL, J. Mortality rate forecasting: can recurrent neural networks beat the lee-carter model? *arXiv preprint arXiv:1909.05501*, 2019.

QIAN, W. *et al.* On the forecast combination puzzle. *Econometrics*, MDPI, v. 7, n. 3, p. 39, 2019. Disponível em: <https://www.mdpi.com/2225-1146/7/3/39>.

RABITTI, G.; BORGONOVO, E. Is mortality or interest rate the most important risk in annuity models? a comparison of sensitivity analysis methods. *Insurance: Mathematics and Economics*, v. 95, p. 48–58, 2020. ISSN 0167-6687. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0167668720301244>.

RIFFE, T.; AL. et. *HMDHFDplus: Read Human Mortality Database and Human Fertility Database Data from the Web*. [S.l.], 2021. Disponível em: <https://doi.org/10.32614/CRAN.package.HMDHFDplus>.

RIZVI, M. F. Arima model time series forecasting. *International Journal for Research in Applied Science and Engineering Technology*, v. 12, n. 5, p. 3782–3785, 2024.

SANTIS, G. D. *et al.* Long-term impact of mortality on population age structures. *International Journal of Population Studies*, p. 1–11, 2024.

SCHNÜRCH, S. *et al.* The impact of mortality shocks on modelling and insurance valuation as exemplified by covid-19. *Annals of Actuarial Science*, v. 16, n. 3, p. 498–526, 2022.

SCHNÜRCH, S.; KORN, R. Point and interval forecasts of death rates using neural networks. *ASTIN Bulletin*, v. 52, n. 1, p. 333–360, 2021.

SHANG, H. L. Mortality and life expectancy forecasting for a group of populations in developed countries: A robust multilevel functional data method. In: *Recent Advances in Robust Statistics: Theory and Applications*. [S.l.]: Springer, 2016. p. 169–184.

SHI, Y. Forecasting mortality rates with the adaptive spatial temporal autoregressive model. *Journal of Forecasting*, 2020.

SILVA, E. G. da. *Uma abordagem de seleção dinâmica de preditores baseada nas janelas temporais mais recentes*. 86 p. Tese (Tese (Doutorado em Estatística)) — Universidade Federal de Pernambuco, Recife, PE, Brasil, 2021. Disponível em: <https://repositorio.ufpe.br/bitstream/123456789/43511/1/TESE%20Eraylson%20Galdino%20da%20Silva.pdf>.

SILVER, D. L.; POIRIER, R.; CURRIE, D. Inductive transfer with context-sensitive neural networks. *Machine Learning*, v. 73, n. 3, p. 313–336, 2008.

SOFTWARE, P. P. *RStudio: Integrated Development Environment for R*. Versão 2025.5.1.513. Boston, MA, USA, 2025. Disponível em: <http://www.posit.co/>.

SOUZA, E. R. Masculinidade e violência no brasil: contribuições para a reflexão, discussão e ação no campo da saúde. *Ciência Saúde Coletiva*, v. 10, n. Suppl, p. 81–92, 2005. Disponível em: <https://www.scielo.br/j/csc/a/5QrxkHxfMdzwgCRVjPXf8yh/>.

SUN, X.; YIN, J.; ZHAO, Y. Using the inverse of expected error variance to determine weights of individual ensemble members: Application to temperature prediction. *Journal of Meteorological Research*, Springer, v. 31, n. 3, p. 502–513, 2017. Disponível em: <http://jmr.cmsjournal.net/en/article/doi/10.1007/s13351-017-6047-0>.

TAIEB, S. B. *et al.* A review and comparison of strategies for multi-step ahead time series forecasting based on the nn5 forecasting competition. *Expert Systems with Applications*, v. 39, n. 8, p. 7067–7083, 2012.

The Pandas Development Team. *pandas-dev/pandas: Pandas*. 2020. Versão latest. Disponível em: <https://doi.org/10.5281/zenodo.3509134>.

THOMAKOS, D. *et al.* Shots forecasting: Short time series forecasting for management research. *British Journal of Management*, jun 2022.

TRAPLETTI, A.; HORNIK, K.; LEBARON, B. *tseries: Time Series Analysis and Computational Finance*. [S.l.], 2024. R package version 0.10-58, <https://CRAN.R-project.org/package=tseries>. Disponível em: <https://CRAN.R-project.org/web/packages/tseries/index.html>.

VALENTINI, G.; DIETTERICH, T. G. Bias—variance analysis and ensembles of svm. In: ROLI, F.; KITTLER, J. (Ed.). *Multiple Classifier Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002. p. 222–231. ISBN 978-3-540-45428-1.

VENTER, G.; ŞAHIN Şule. *Modeling Mortality of Related Populations via Parameter Shrinkage*. 2018. Available at SSRN 3301981. Disponível em: <https://doi.org/10.7916/d8-mryq-rs54>.

WEIGAND, A. C.; LANGE, D.; RAUSCHENBERGER, M. How can small data sets be clustered? In: *Mensch und Computer 2021-Workshopband*. [S.l.]: Gesellschaft für Informatik eV, 2021.

YAP, Z. J.; PATHMANATHAN, D.; DABO-NIANG, S. Forecasting mortality rates with functional signatures. *ASTIN Bulletin: The Journal of the IAA*, Cambridge University Press, v. 55, n. 1, p. 97–120, 2025.

ZAIN, Z. M.; ALTURKI, N. M. Covid-19 pandemic forecasting using cnn-lstm: A hybrid approach. *Journal of Control Science and Engineering*, v. 2021, n. 1, p. 8785636, 2021. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1155/2021/8785636>.

ZARGHAMI, M. *et al.* Time series modeling and forecasting of drug-related deaths in iran (2014–2016). *Addiction and Health*, v. 15, n. 3, p. 149–155, 2023.

ZHANG, G. P. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, v. 50, p. 159–175, 2003.

ZHOU, G.-B. *et al.* Minimal gated unit for recurrent neural networks. *International Journal of Automation and Computing*, v. 13, n. 3, p. 226–234, 2016.

ZHUANG, F. *et al.* A comprehensive survey on transfer learning. *Proceedings of the IEEE*, v. 109, n. 1, p. 43–76, 2021.

ZILI, A. H. A.; MARDIYATI, S.; LESTARI, D. Forecasting indonesian mortality rates using the lee-carter model and arima method. In: *Proceedings of the 3rd International Symposium on Current Progress in Mathematics and Sciences 2017 (ISCPMS2017), Bali, Indonesia*. [s.n.], 2018. Disponível em: <https://doi.org/10.1063/1.5064209>.

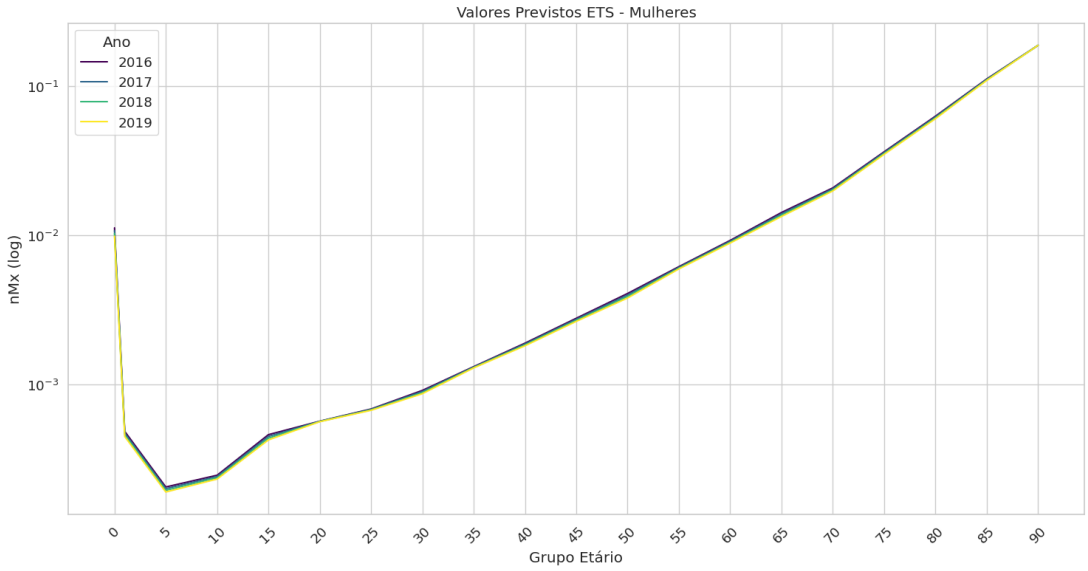
A APÊNDICE A - Lista de Países do HMD

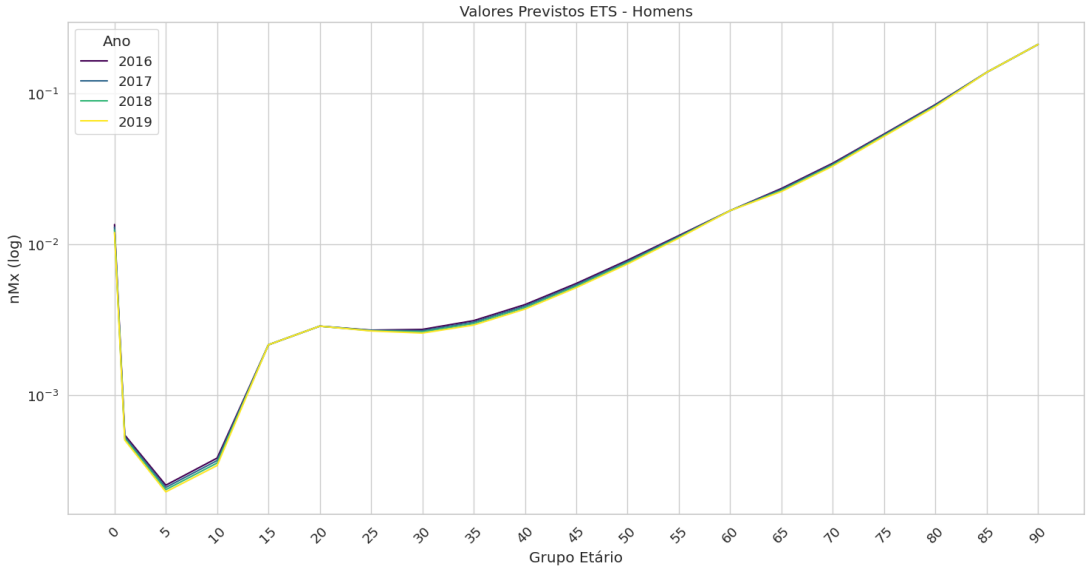
Países e códigos HMD disponíveis

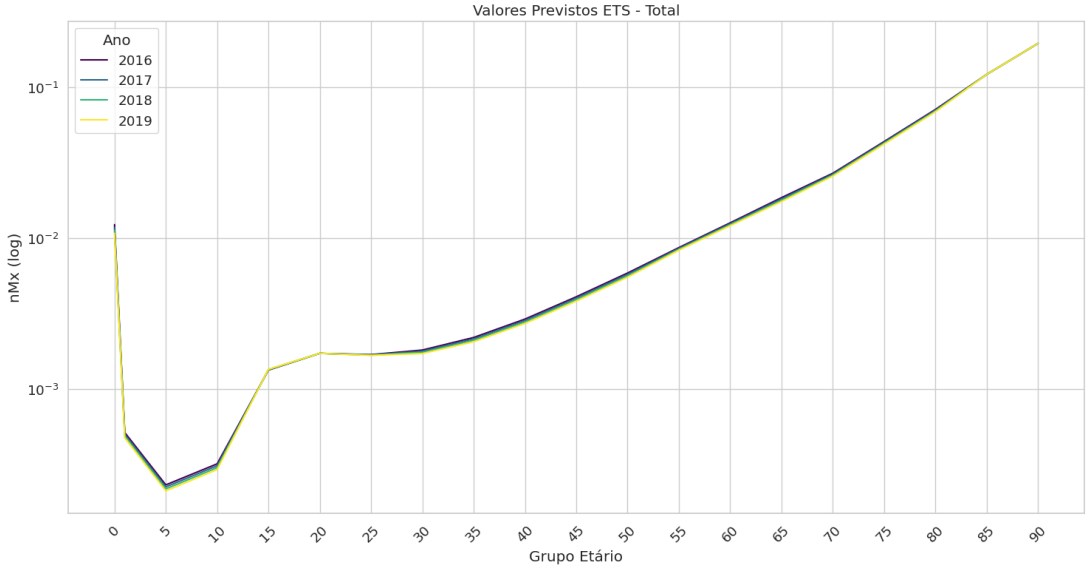
País	**Código HMD**	**Observação**
Australia	AUS	–
Austria	AUT	–
Belarus	BLR	–
Belgium	BEL	–
Bulgaria	BGR	–
Canada	CAN	–
Chile	CHL	–
Croatia	HRV	–
Czechia	CZE	–
Denmark	DNK	–
Estonia	EST	–
Finland	FIN	–
France	FRA	Total
	FRACNP	Civilian
	FRATNP	Inclui territórios
Germany	DEUTNP	Total
	DEUTE	Leste
	DEUTW	Oeste
Greece	GRC	–
Hong Kong	HKG	–
Hungary	HUN	–
Iceland	ISL	–
Ireland	IRL	–
Israel	ISR	–
Italy	ITA	–
Japan	JPN	–
Latvia	LVA	–
Lithuania	LTU	–
Luxembourg	LUX	–
Netherlands	NLD	–
New Zealand	NZL	Total
	NZLMA	Maori

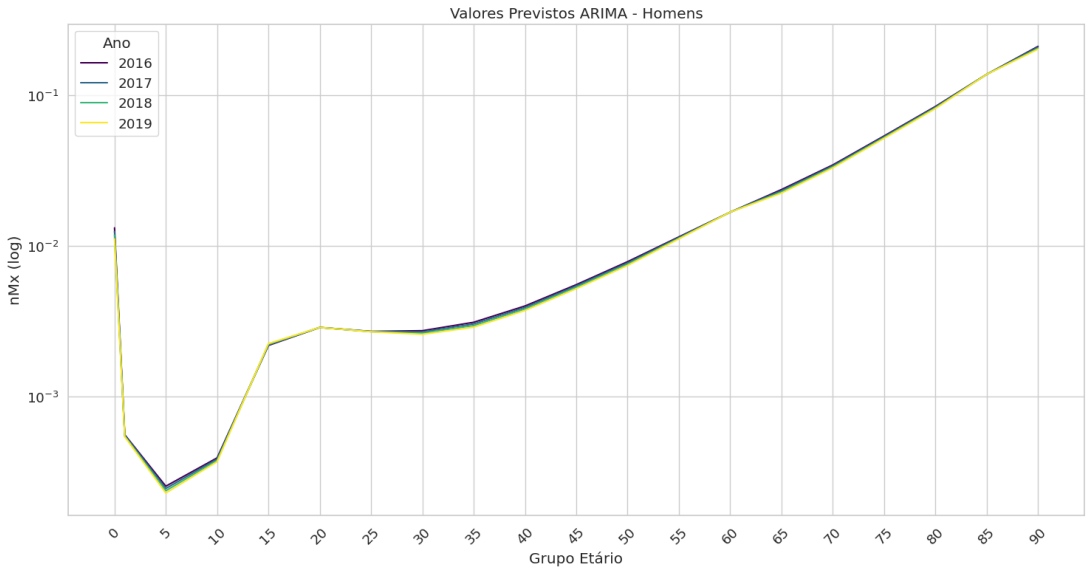
País	Código HMD	Observação
	NZLNM	Non-Maori
Norway	NOR	–
Poland	POL	–
Portugal	PRT	–
Republic of Korea	KOR	–
Russia	RUS	–
Slovakia	SVK	–
Slovenia	SVN	–
Spain	ESP	–
Sweden	SWE	–
Switzerland	CHE	–
Taiwan	TWN	–
United Kingdom	GBR	Inglaterra, País de Gales, etc.
United States	USA	–
Ukraine	UKR	–

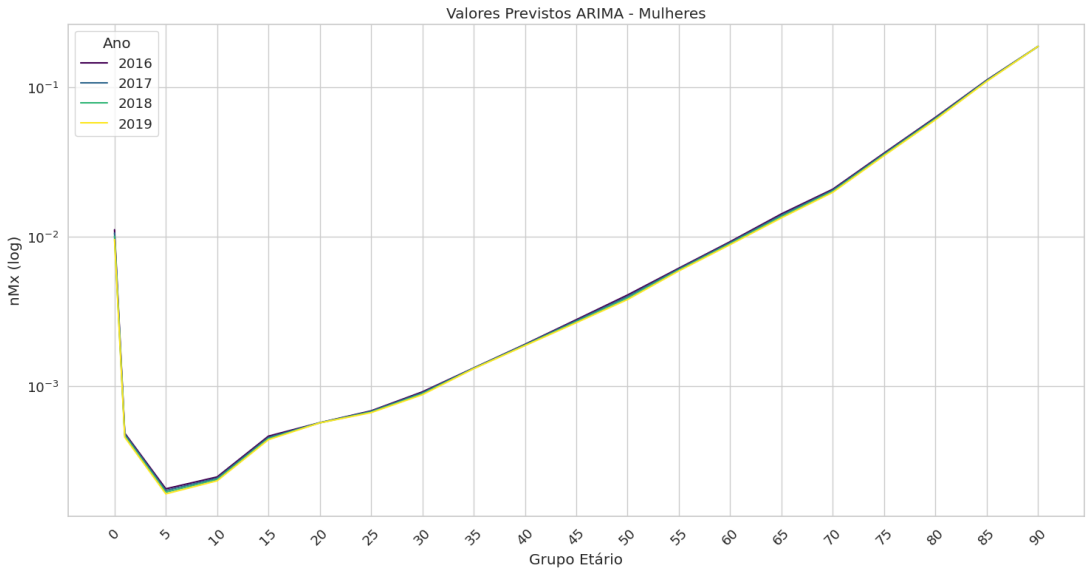
B APÊNDICE B - Curvas de Mortalidade Previstas

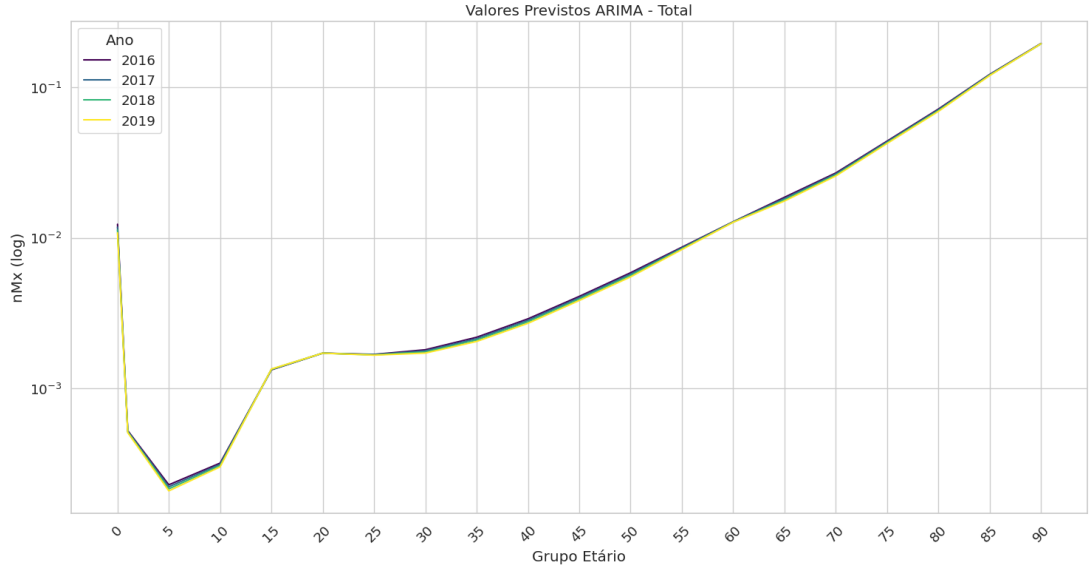


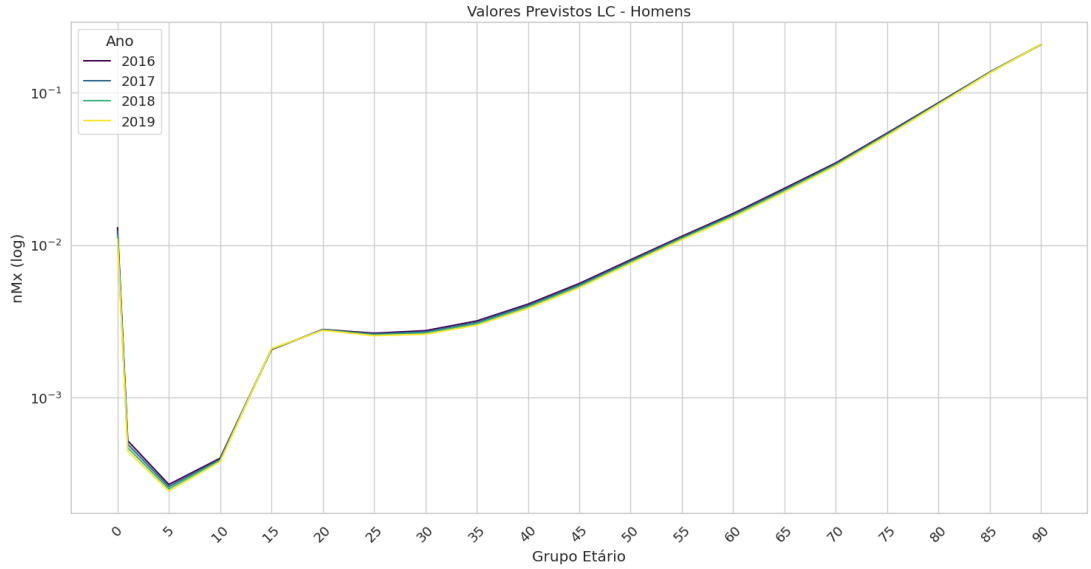


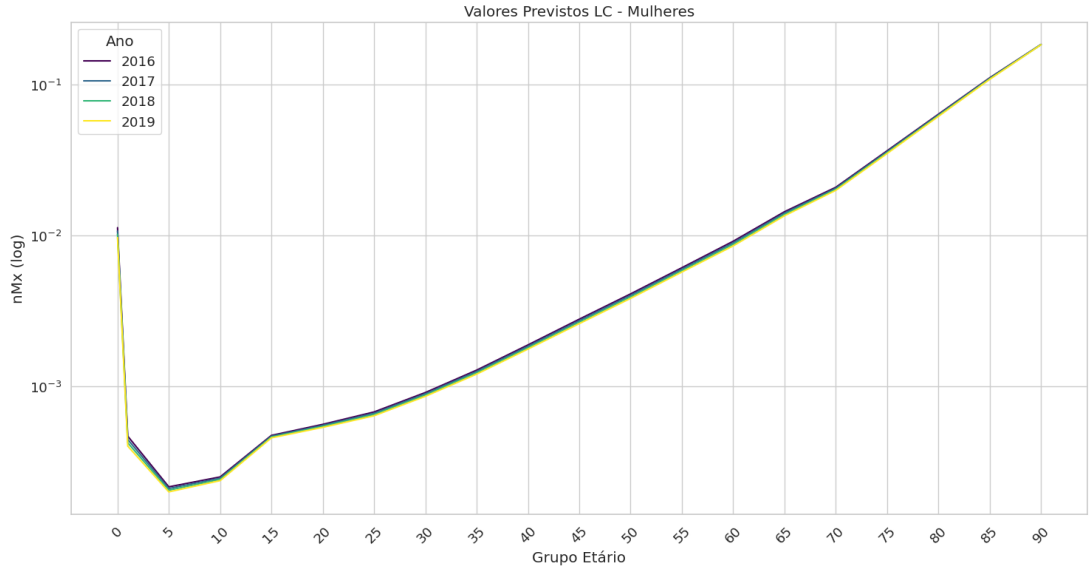

Valores Previstos ETS - Total
Ano
2016
2017
2018
2019
nMx (log)
Grupo Etário

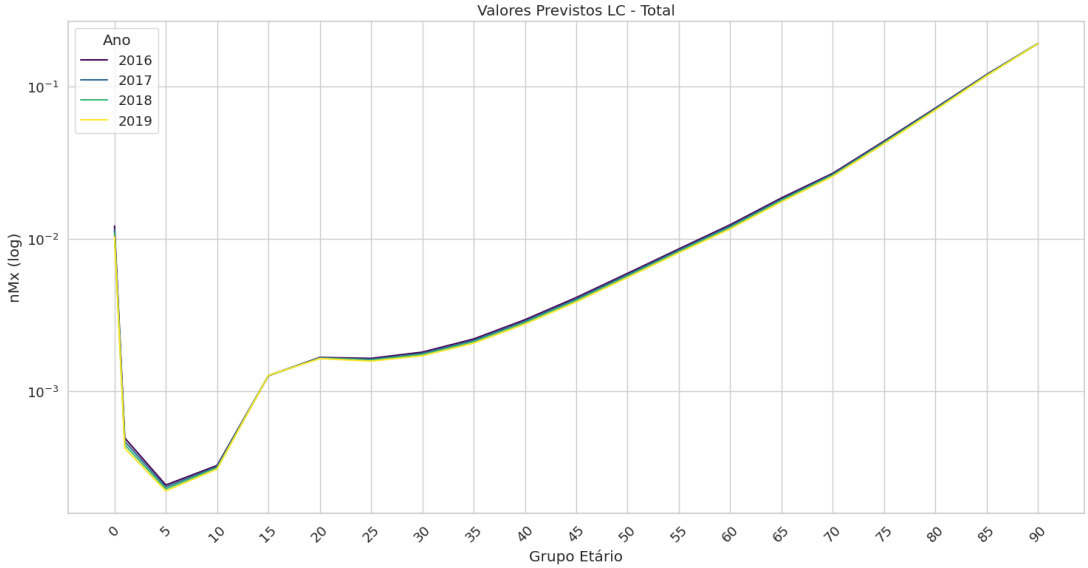

Valores Previstos ARIMA - Homens
Ano
2016
2017
2018
2019
nMx (log)
Grupo Etário

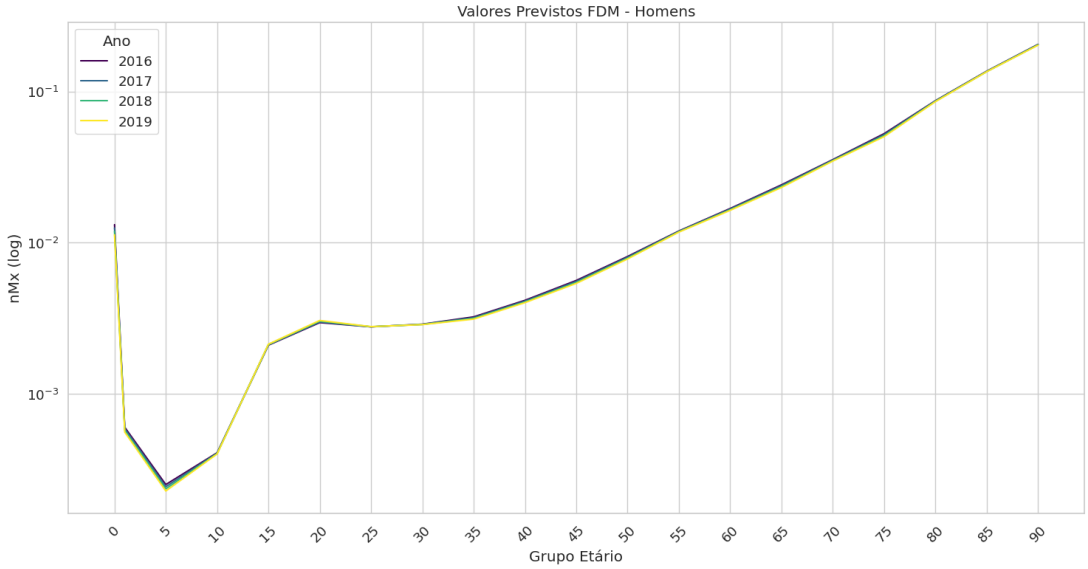

Valores Previstos ARIMA - Mulheres
Ano
2016
2017
2018
2019
nMx (log)
Grupo Etário

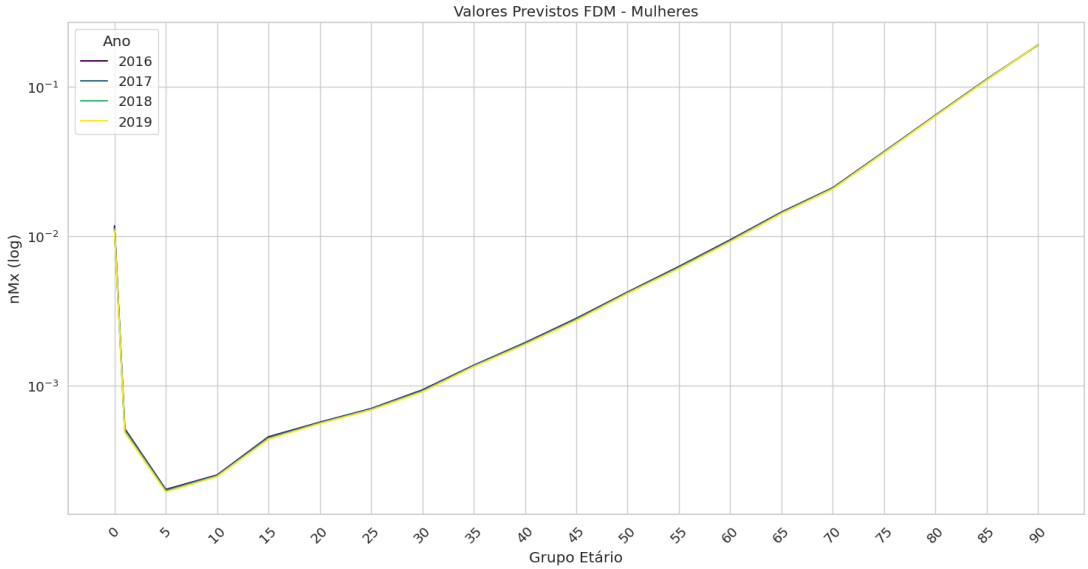

Valores Previstos ARIMA - Total
Ano
2016
2017
2018
2019
nMx (log)
Grupo Etário

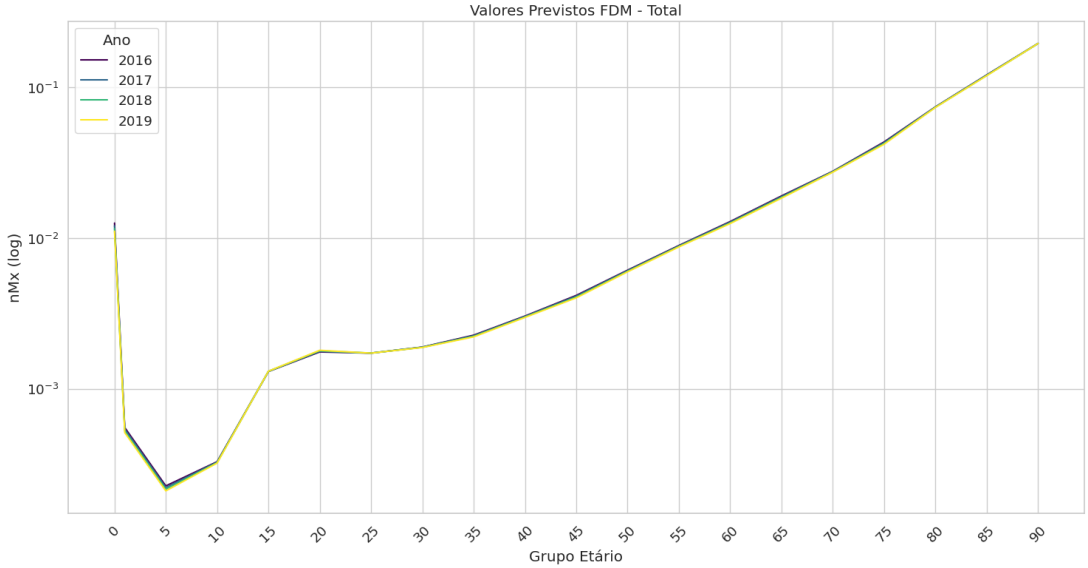

Valores Previstos LC - Homens
Ano
2016
2017
2018
2019
nMx (log)
Grupo Etário

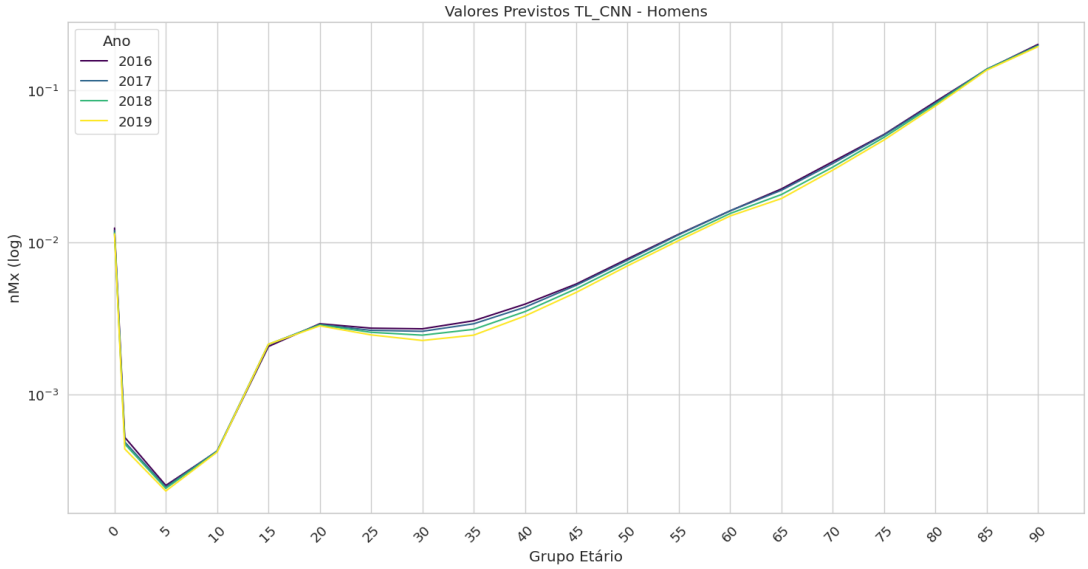

Valores Previstos LC - Mulheres
Ano
2016
2017
2018
2019
nMx (log)
Grupo Etário

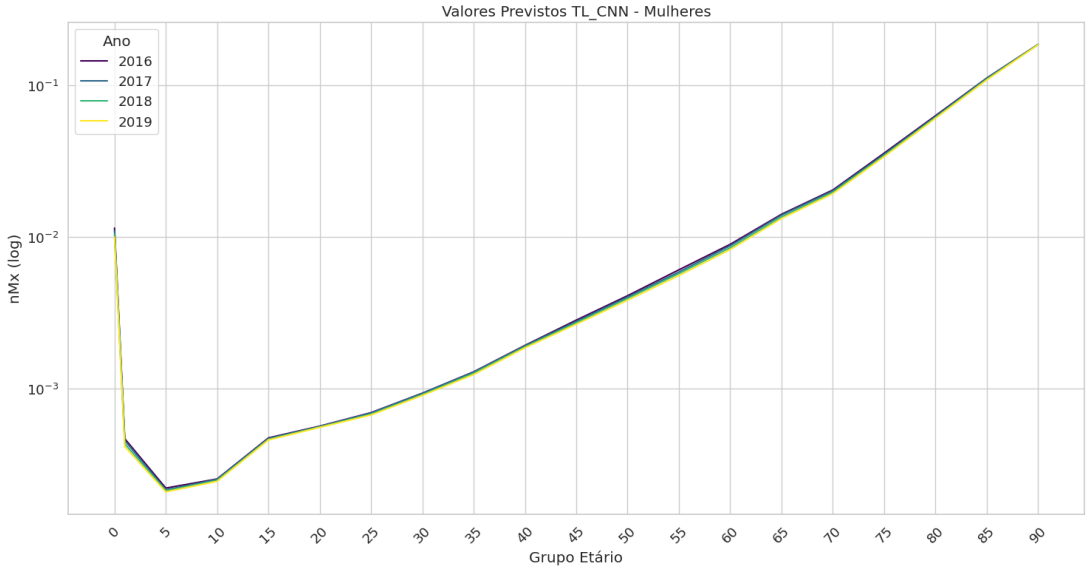

Valores Previstos LC - Total
Ano
2016
2017
2018
2019
nMx (log)
Grupo Etário

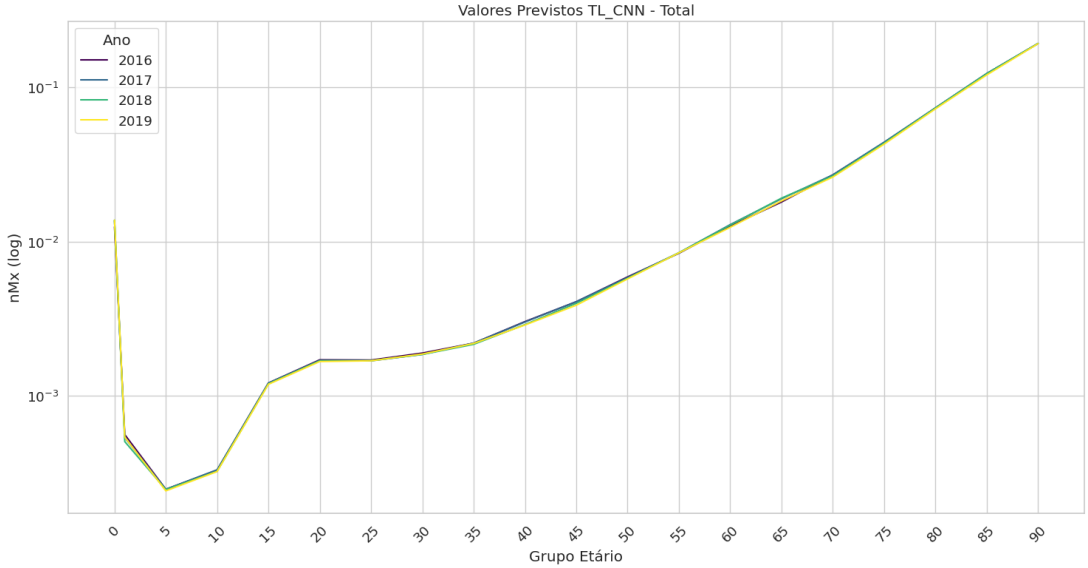

Valores Previstos FDM - Homens
Ano
2016
2017
2018
2019
nMx (log)
Grupo Etário

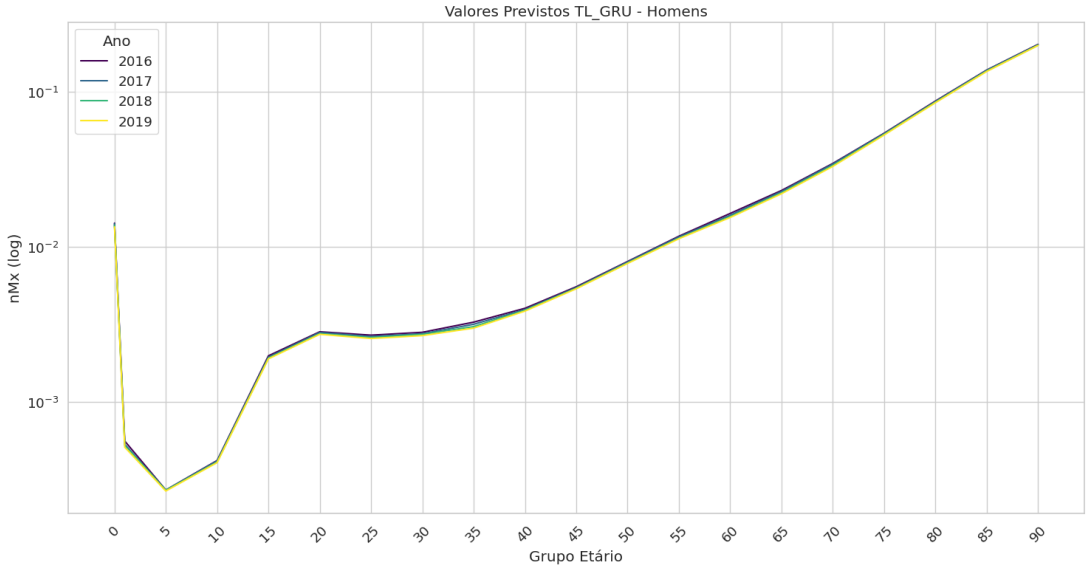

Valores Previstos FDM - Mulheres
Ano
2016
2017
2018
2019
nMx (log)
Grupo Etário

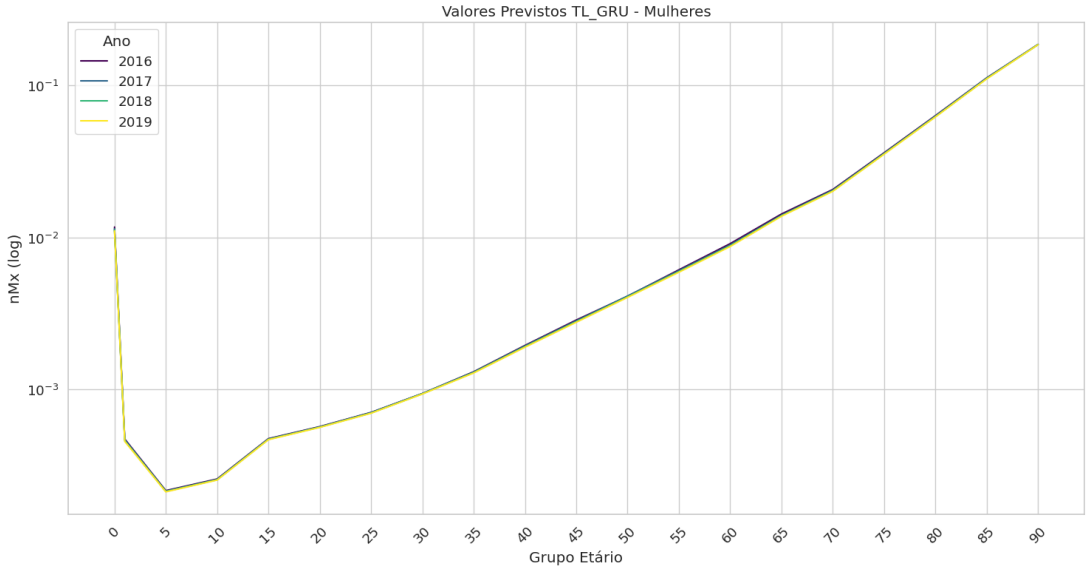

Valores Previstos FDM - Total
Ano
2016
2017
2018
2019
nMx (log)
10^{-1}
10^{-2}
10^{-3}
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90
Grupo Etário

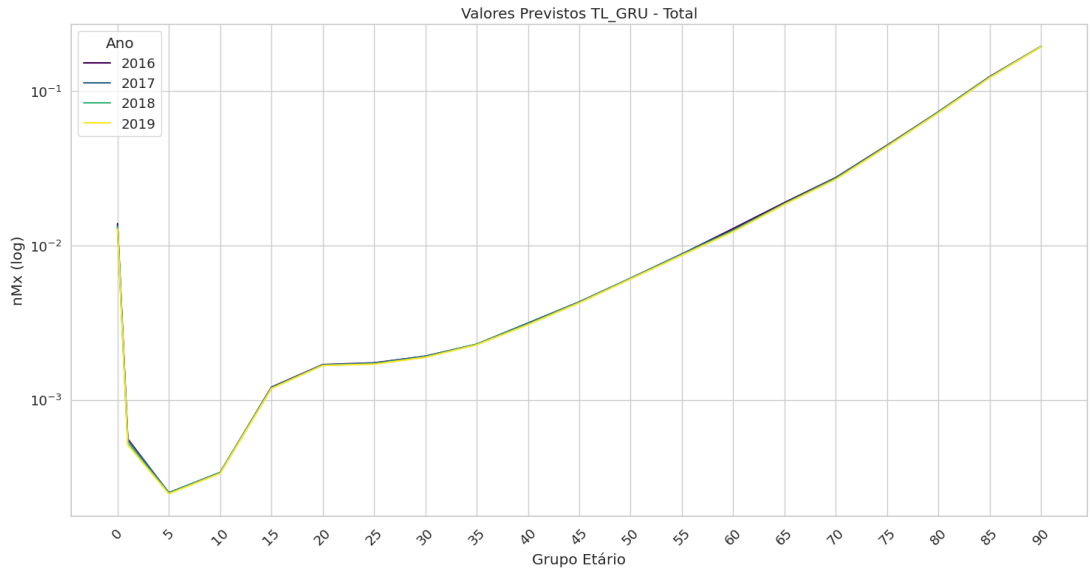

Valores Previstos TL_CNN - Homens
Ano
2016
2017
2018
2019
nMx (log)
10^{-1}
10^{-2}
10^{-3}
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90
Grupo Etário

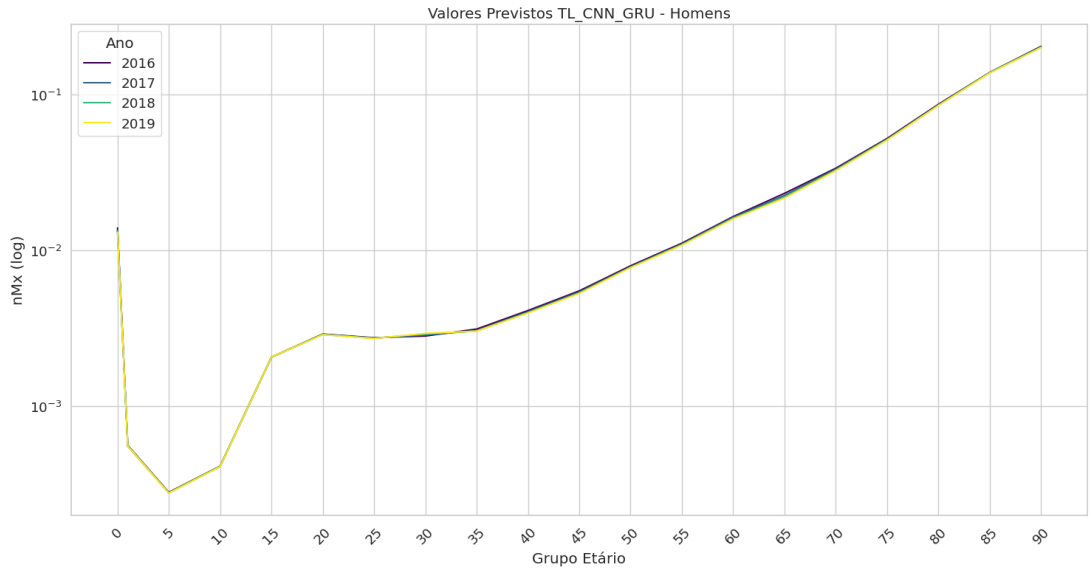

Valores Previstos TL_CNN - Mulheres
Ano
2016
2017
2018
2019
nMx (log)
10^{-1}
10^{-2}
10^{-3}
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90
Grupo Etário

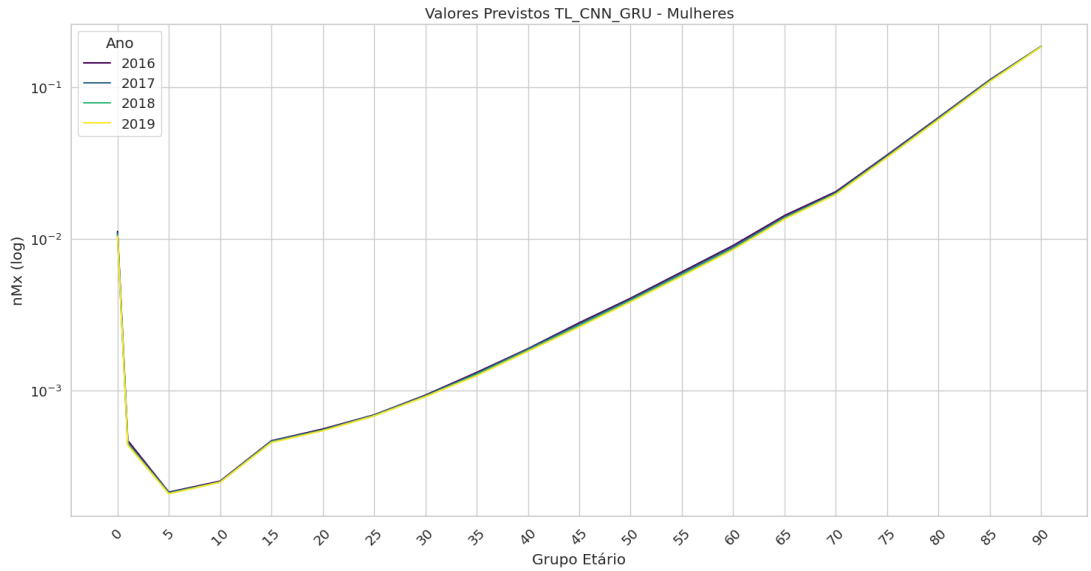

Valores Previstos TL_CNN - Total
Ano
2016
2017
2018
2019
nMx (log)
Grupo Etário

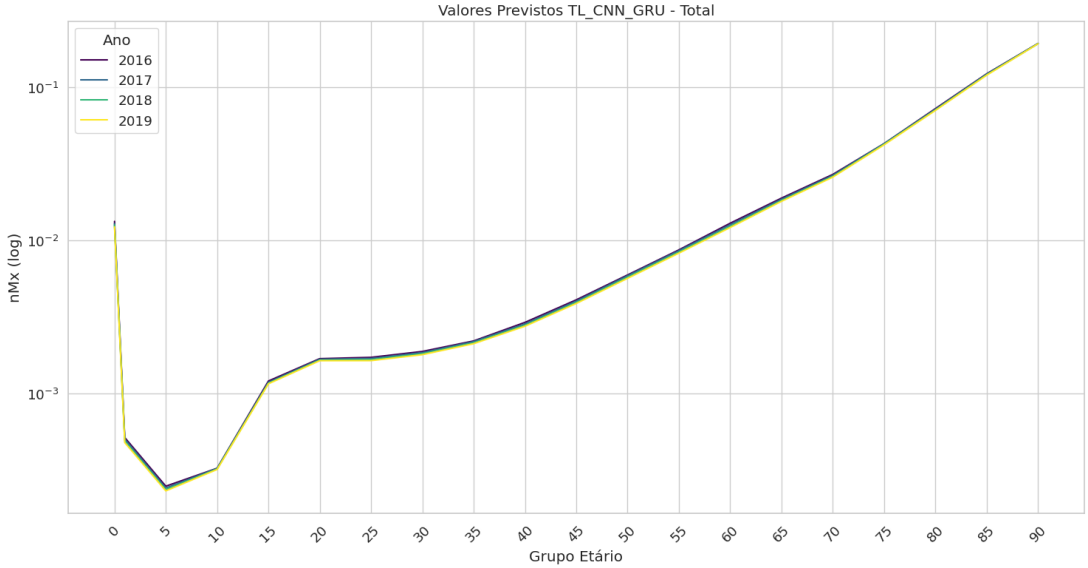

Valores Previstos TL_GRU - Homens
Ano
2016
2017
2018
2019
nMx (log)
Grupo Etário

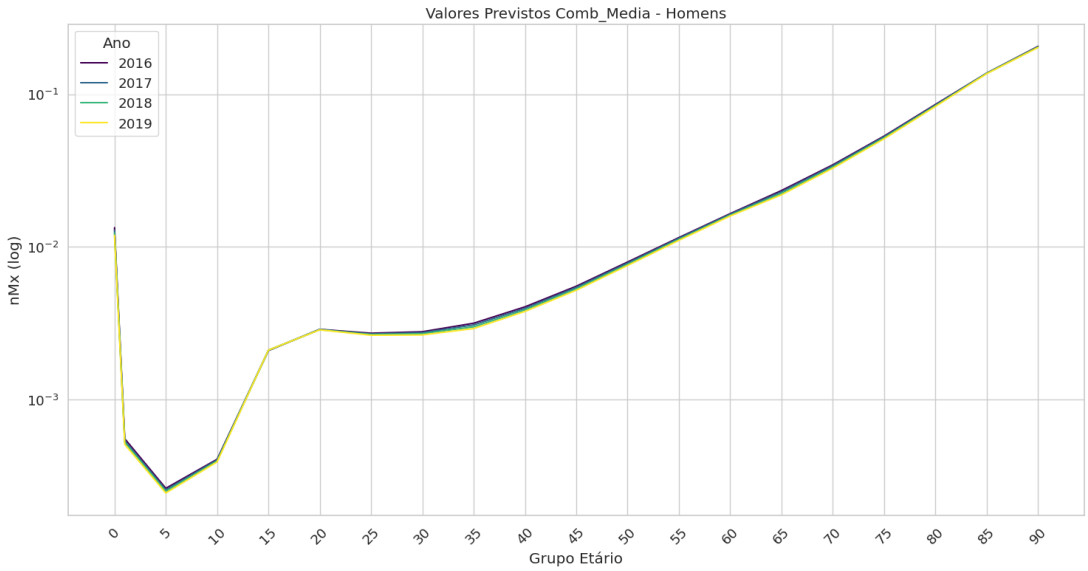

Valores Previstos TL_GRU - Mulheres
Ano
2016
2017
2018
2019
nMx (log)
Grupo Etário

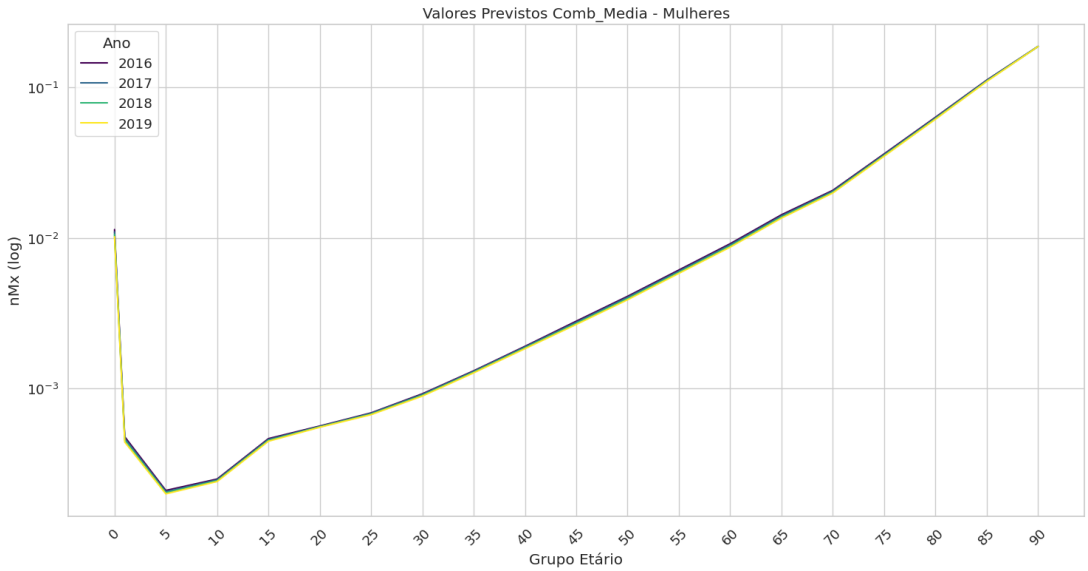

Valores Previstos TL_GRU - Total
Ano
2016
2017
2018
2019
nMx (log)
Grupo Etário

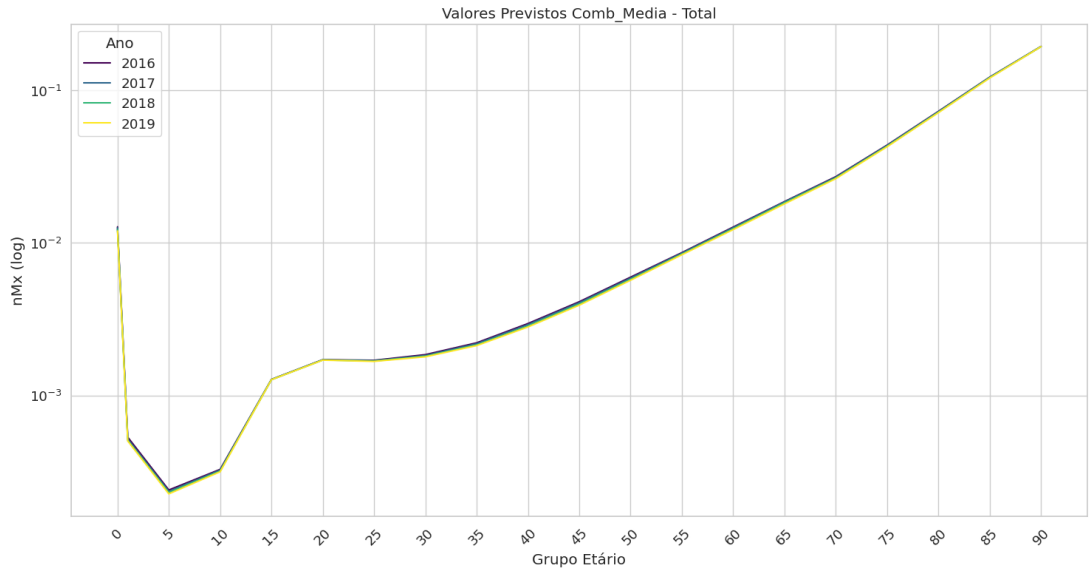

Valores Previstos TL_CNN_GRU - Homens
Ano
2016
2017
2018
2019
nMx (log)
Grupo Etário

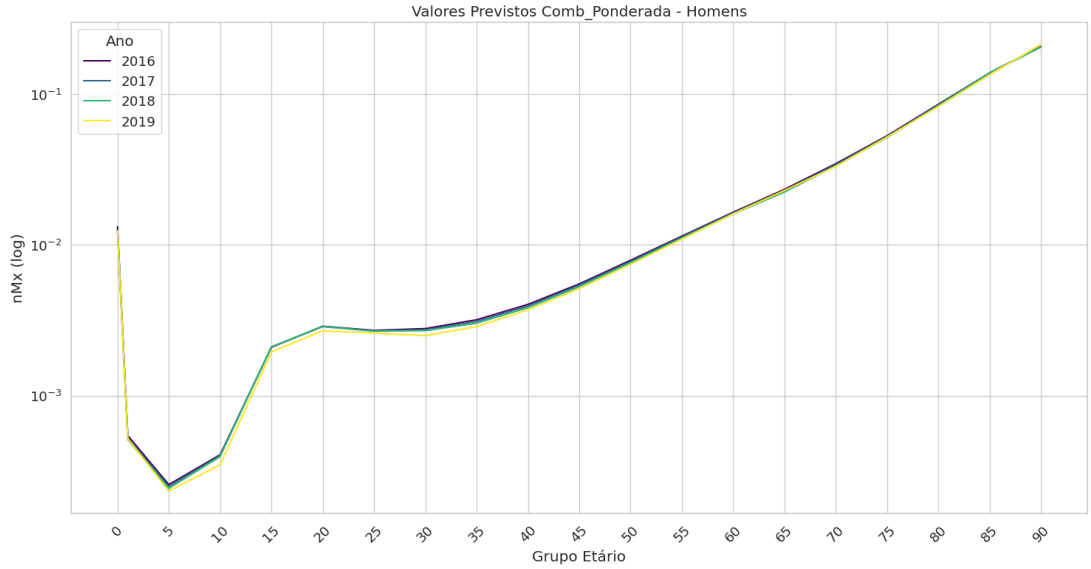

Valores Previstos TL_CNN_GRU - Mulheres
Ano
2016
2017
2018
2019
nMx (log)
Grupo Etário

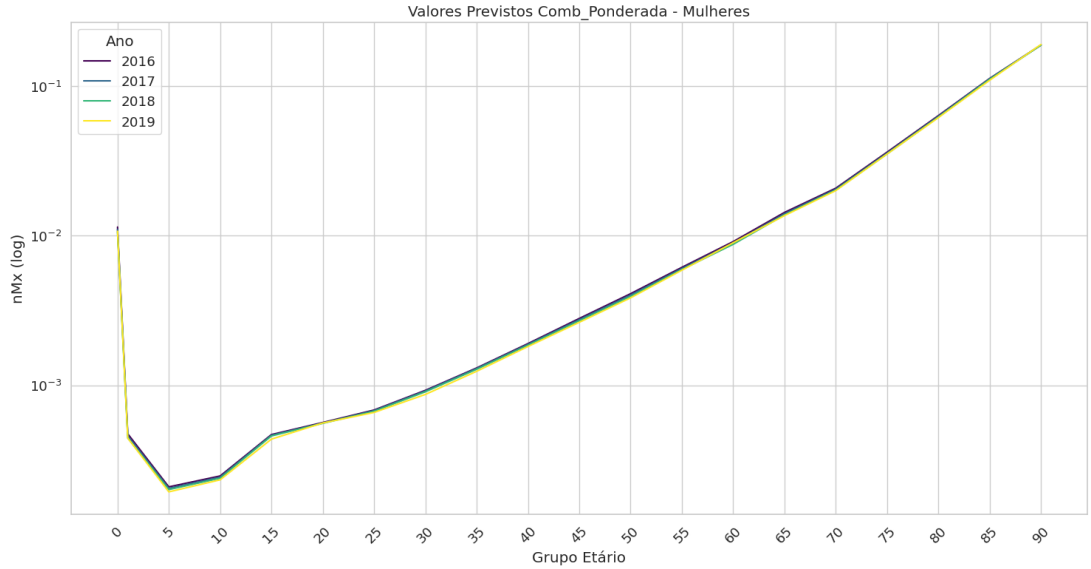

Valores Previstos TL_CNN_GRU - Total
Ano
2016
2017
2018
2019
nMx (log)
10^{-1}
10^{-2}
10^{-3}
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90
Grupo Etário

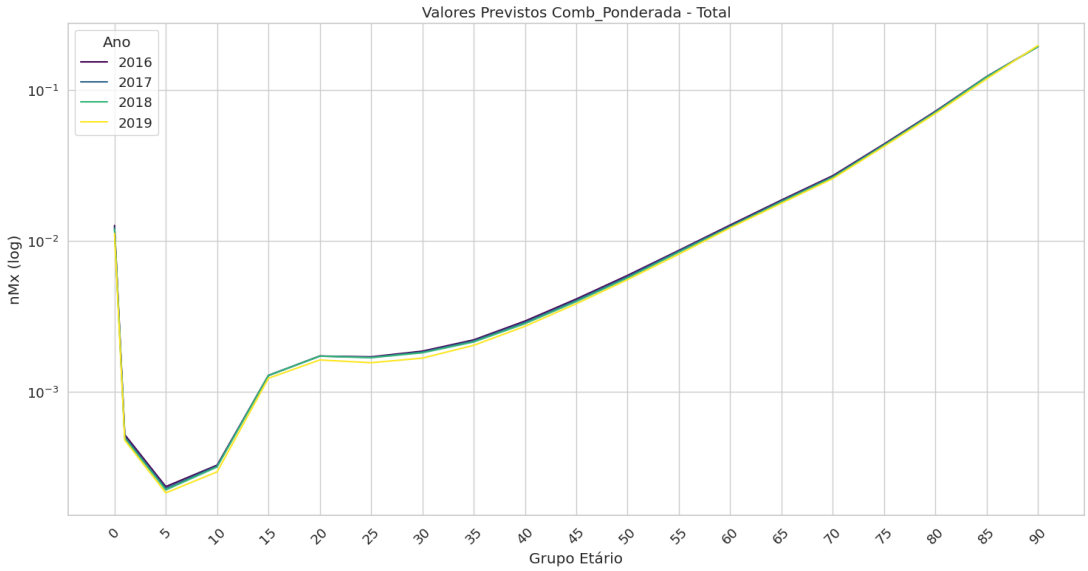

Valores Previstos Comb_Media - Homens
Ano
2016
2017
2018
2019
nMx (log)
10^{-1}
10^{-2}
10^{-3}
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90
Grupo Etário


Valores Previstos Comb_Media - Mulheres
Ano
2016
2017
2018
2019
nMx (log)
10^{-1}
10^{-2}
10^{-3}
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90
Grupo Etário


Valores Previstos Comb_Media - Total
Ano
2016
2017
2018
2019
nMx (log)
Grupo Etário


Valores Previstos Comb_Ponderada - Homens
Ano
2016
2017
2018
2019
nMx (log)
Grupo Etário


Valores Previstos Comb_Ponderada - Mulheres
Ano
2016
2017
2018
2019
nMx (log)
Grupo Etário


Valores Previstos Comb_Ponderada - Total
Ano
2016
2017
2018
2019
nMx (log)
10^{-1}
10^{-2}
10^{-3}
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
Grupo Etário

C APÊNDICE C - Pesos da Combinação Ponderada por idade e sexo