Universidade de Évora - Escola de Ciências e Tecnologia

Mestrado em Modelação Estatística e Análise de Dados

Dissertação

Modelação e previsão das taxas de mortalidade infantil nos países da CPLP entre 1985 e 2020 usando modelos de Séries Temporais

Teotónio Gaspar Feliciano

Orientador(es) | Dulce Maria de Oliveira Gomes

Évora 2025

A dissertação foi objeto de apreciação e discussão pública pelo seguinte júri nomeado pelo Diretor da Escola de Ciências e Tecnologia:

Presidente	|	Dulce Gamito Pereira (Universidade de Évora)
Vogais	|	Dulce Maria de Oliveira Gomes (Universidade de Évora) (Orientador) Maria do Rosário Ramos (Universidade Aberta) (Arguente)

Évora 2025

Aos eternos insatisfeitos.

Agradecimentos

Em primeiro lugar agradeço a Deus, Jeová, pela sua infinita bondade e poder em conceder-me saúde para poder realizar este trabalho. Posteriormente, venho a agradecer a minha familia que abriu mãos de seu tempo e recursos permitindo-me assim realizar este sonho de mestrado. A minha Esposa Guiomar Feliciano, a companheira de todas horas que privou-se de mim para eu puder me dedicar ao estudo e consequentemente a realização desta dissertação. A minha mãe pelo carinho e mesmo distante dando me forças para continuar, como ela mesma me ensinou: " A tempo para tudo, para rir, para chorar, para semear, para colher, a tempo para tudo", aos meus sogros pelo apoio moral, as várias palavras de conforto que me foram dadas ao longo deste período.

Agradeço ainda aos meus irmãos: Peterson, Magda, Mirian, Oseias por fazerem parte de mim e permitirem que eu fizesse também parte deles, como temos dito: WE ARE FELICIANO.

Venho ainda agradecer os meus colegas de trabalho no Instituto Superior Politécnico de Tecnologias e Ciências, poderia citar vários nomes, mas, gostaria de destacar o grande companheiro de batalha pela vida, Adilson Andrade, Criciano Jerónimo e Pedro António. Agradecer ainda a Direcção Geral e dos Laboratórios que permitiram-me abdicar temporariamente das actividades laborais para a realização do mestrado.

Gostaria de incluir neste agradecimento aos meus amigos que também ajudaram e em várias ocasiões também deram-me apoio para poder continuar.

A minha lista de gratidão é vasta, por isso continuo a agradecer aos distintos colegas do Mestrado pelo apoio que me forneceram para o realizar do mestrado, em especial o Luis Ferreira e Fligêncio Gomes.

A minha orientadora, a Professora Doutora Dulce Gomes, pelo incentivo e a pronta resposta em atender as dúvidas na elaboração desta dissertação. Muito obrigado Professora Doutora Dulce. Contudo, de um modo extensivo a todos os Professores que me ajudaram a fazer esta caminhada do mestrado.

Conteúdo

Conteúdo vii

Lista de Figuras xi

Lista de Tabelas xv

Lista de Acrónimos xix

Sumário xxi

Abstract xxiii

1 Introdução 1

1.1 Objetivos . . . 2

1.2 Resumo das características gerais dos países da CPLP . . . 2

2 Séries Temporais 9

2.1 Conceitos . . . 9

2.1.1 Objetivos da análise de uma série temporal . . . 10

2.1.2 Componentes de uma série temporal . . . 10

2.2 Processos estocásticos . . . 11

2.2.1 Processos estocásticos estacionários . . . 12

2.2.2 Processos estocásticos não estacionários . . . 13

2.2.3 Transformação de processos não estacionários em estacionários . . . 14

2.2.3.1 Diferenciação . . . 14

2.2.3.2 Transformação de Box-Cox . . . 15

2.2.4 Função de autocovariância e autocorrelação . . . 16

2.3 Métodos de análises de séries temporais 16

3 Modelos ARIMA 19

3.1 Introdução ao modelo ARIMA 19
3.2 Identificação dos modelos ARIMA 20
3.3 Formulação do modelo ARIMA 21
3.3.1 Modelos AutoRegressivos 21
3.3.2 Modelos de médias móveis 22
3.3.3 Modelos autoregressivos e de médias móveis 25
3.3.4 Modelos autoregressivos integrados e de médias móveis 26
3.4 Estimação dos modelos ARIMA 27
3.4.1 Métodos de estimação 28
3.4.1.1 Método de máxima verossimilhança 28
3.4.1.2 Método dos mínimos quadrados 29
3.4.1.3 Método de Yule-Walker 29
3.4.2 Seleção do modelo 30
3.5 Diagnóstico dos modelos ARIMA 30
3.6 Previsão dos modelos ARIMA 31
3.7 Mudanças de estrutura em séries temporais 33
3.7.1 Tipos de mudanças estruturais 34
3.7.2 Identificação e modelação das mudanças estruturais 34
3.7.3 Testes estatísticos para mudanças estruturais 35

4 Análise dos dados 37

4.1 Análise exploratória da base de dados 38
4.2 Mudança de estrutura 41
4.3 Modelação das taxas de mortalidade infantil 46
4.3.1 Teste da estacionariedade e identificação do modelo 46
4.3.2 Estimação do modelo 48
4.3.3 Testes de diagnóstico aos modelos 60
4.3.4 Discussão das previsões 61
4.3.5 Análise de desempenho com divisão treino-teste 64

5 Considerações finais 71

6 Apêndice 75

A Gráficos ACF e PACF dos dados transformados dos países da CPLP para a modelação 77

B Tabelas **83**

C Resíduos **87**

D Gráficos de previsão do conjunto de dados de treino com os dados de teste **91**

E Gráficos das quebras de estruturas nas séries **95**

Bibliografia **99**

Lista de Figuras

4.1 Evolução da Taxa de Mortalidade Infantil nos Países da CPLP 38

4.2 Boxplot das taxas de mortalidade infantil para indivíduos do sexo masculino nos países da CPLP (1985–2020). 40

4.3 Boxplot das taxas de mortalidade infantil para indivíduos do sexo feminino nos países da CPLP (1985–2020). 40

4.4 Evolução temporal das taxas de mortalidade infantil nos países da CPLP, divididos em dois clusters, para indivíduos do sexo masculino (1985–2020). 41

4.5 Evolução temporal das taxas de mortalidade infantil nos países da CPLP, divididos em dois clusters, para indivíduos do sexo feminino (1985–2020). 41

4.6 Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo masculino - Angola. 49

4.7 Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo feminino - Angola. 49

4.8 Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo masculino - Brasil. 50

4.9 Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo feminino - Brasil. 50

4.10 Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo masculino - Cabo Verde. 51

4.11 Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo feminino - Cabo Verde. 51

4.12 Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo masculino - Guiné-Bissau. 52

4.13 Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo feminino - Guiné-Bissau. 52

4.14 Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo masculino - Guiné Equatorial. 54

4.15 Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo feminino - Guiné Equatorial. 54

4.16 Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo masculino - Moçambique. 55
4.17 Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo feminino - Moçambique. 55
4.18 Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo masculino Portugal. 56
4.19 Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo feminino Portugal. 56
4.20 Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo masculino - São Tomé e Príncipe. 57
4.21 Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo feminino - São Tomé e Príncipe. 57
4.22 Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo masculino - Timor-Leste. 58
4.23 Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo feminino - Timor-Leste. 58

A.1 Gráficos de ACF e PACF de Angola para ambos os géneros 78
A.2 Gráficos de ACF e PACF do Brasil para ambos os géneros 78
A.3 Gráficos de ACF e PACF de Cabo Verde para ambos os géneros 79
A.4 Gráficos de ACF e PACF de Guiné-Bissau para ambos os géneros 79
A.5 Gráficos de ACF e PACF de Guiné Equatorial para ambos os géneros 80
A.6 Gráficos de ACF e PACF de Moçambique para ambos os géneros 80
A.7 Gráficos de ACF e PACF de Portugal para ambos os géneros 81
A.8 Gráficos de ACF e PACF de São Tomé e Principé para ambos os géneros 81
A.9 Gráficos de ACF e PACF de Timor-Leste para ambos os géneros 82

C.1 Análise de autocorrelação de resíduos dos indivíduos do sexo masculino de Angola. 88
C.2 Análise de autocorrelação de resíduos dos indivíduos do sexo feminino de Angola. 88
C.3 Análise de autocorrelação de resíduos dos indivíduos do sexo masculino do Brasil. 88
C.4 Análise de autocorrelação de resíduos dos indivíduos do sexo feminino do Brasil. 88
C.5 Análise de autocorrelação de resíduos dos indivíduos do sexo masculino de Cabo Verde. . . 88
C.6 Análise de autocorrelação de resíduos dos indivíduos do sexo feminino de Cabo Verde. . . 88
C.7 Análise de autocorrelação de resíduos dos indivíduos do sexo masculino de Guiné-Bissau. . 89
C.8 Análise de autocorrelação de resíduos dos indivíduos do sexo feminino de Guiné-Bissau. . . 89
C.9 Análise de autocorrelação de resíduos dos indivíduos do sexo masculino da Guiné Equatorial. 89
C.10 Análise de autocorrelação de resíduos dos indivíduos do sexo feminino da Guiné Equatorial. 89
C.11 Análise de autocorrelação de resíduos dos indivíduos do sexo masculino de Moçambique. . 89
C.12 Análise de autocorrelação de resíduos dos indivíduos do sexo feminino de Moçambique. . . 89
C.13 Análise de autocorrelação de resíduos dos indivíduos do sexo masculino de Portugal. . . . 90

C.14 Análise de autocorrelação de resíduos dos indivíduos do sexo feminino de Portugal. 90

C.15 Análise de autocorrelação de resíduos dos indivíduos do sexo masculino de São Tomé e Príncipe. 90

C.16 Análise de autocorrelação de resíduos dos indivíduos do sexo feminino de São Tomé e Príncipe. 90

C.17 Análise de autocorrelação de resíduos dos indivíduos do sexo masculino de Timor-Leste. . 90

C.18 Análise de autocorrelação de resíduos dos indivíduos do sexo feminino de Timor-Leste. . . 90

D.1 Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Angola (Masculino). 92

D.2 Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Angola (Feminino). 92

D.3 Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Brasil (Masculino). 92

D.4 Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Brasil (Feminino). 92

D.5 Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Cabo Verde (Masculino). 92

D.6 Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Cabo Verde (Feminino). 92

D.7 Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Guiné-Bissau (Masculino). 93

D.8 Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Guiné-Bissau (Feminino). 93

D.9 Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Guiné Equatorial (Masculino). 93

D.10 Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Guiné Equatorial (Feminino). 93

D.11 Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Moçambique (Masculino). 93

D.12 Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Moçambique (Feminino). 93

D.13 Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Portugal (Masculino). 94

D.14 Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Portugal (Feminino). 94

D.15 Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - São Tomé e Príncipe (Masculino). 94

D.16 Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - São Tomé e Príncipe (Feminino). 94

D.17 Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Timor-Leste (Masculino). 94

D.18 Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Timor-Leste (Feminino). 94

E.1 Quebras de estruturas na série - Angola (Masculino). 96
E.2 Quebras de estruturas na série - Angola (Feminino). 96
E.3 Quebras de estruturas na série - Brasil (Masculino). 96
E.4 Quebras de estruturas na série - Brasil (Feminino). 96
E.5 Quebras de estruturas na série - Cabo Verde (Masculino). 96
E.6 Quebras de estruturas na série - Cabo Verde (Feminino). 96
E.7 Quebras de estruturas na série - Guiné-Bissau (Masculino). 97
E.8 Quebras de estruturas na série - Guiné-Bissau (Feminino). 97
E.9 Quebras de estruturas na série - Guiné Equatorial (Masculino). 97
E.10 Quebras de estruturas na série - Guiné Equatorial (Feminino). 97
E.11 Quebras de estruturas na série - Moçambique (Masculino). 97
E.12 Quebras de estruturas na série - Moçambique (Feminino). 97
E.13 Quebras de estruturas na série - Portugal (Masculino). 98
E.14 Quebras de estruturas na série - Portugal (Feminino). 98
E.15 Quebras de estruturas na série - São Tomé e Príncipe (Masculino). 98
E.16 Quebras de estruturas na série - São Tomé e Príncipe (Feminino). 98
E.17 Quebras de estruturas na série - Timor Leste (Masculino). 98
E.18 Quebras de estruturas na série - Timor Leste (Feminino). 98

Lista de Tabelas

1.1 Dados populacionais e de PIB dos estados membros da CPLP para o ano de 2022. Fonte: Countrymeters (`https://countrymeters.info/em`) e `https://www.thetruesize.com/`. 3

4.1 Estatísticas Descritivas por País e Género 39

4.2 Agrupamento de países por clusters. Fonte: Autor, extraído do RStudio. 41

4.3 Comparação das médias de mortalidade infantil antes e depois da mudança de estrutura para Angola 42

4.4 Comparação das médias antes e depois da mudança de estrutura para o Brasil 42

4.5 Comparação das médias antes e depois da mudança de estrutura para Cabo Verde 43

4.6 Comparação das médias antes e depois da mudança de estrutura para Guiné-Bissau . . . 43

4.7 Comparação das médias antes e depois da mudança de estrutura para o Guiné Equatorial . 44

4.8 Comparação das médias antes e depois da mudança de estrutura para Moçambique 44

4.9 Comparação das médias de mortalidade infantil em Portugal antes e depois das mudanças de estrutura 45

4.10 Comparação das médias de mortalidade infantil em São Tomé e Príncipe antes e depois das mudanças de estrutura 45

4.11 Comparação das médias de mortalidade infantil em Timor-Leste antes e depois das mudanças de estrutura 46

4.12 Estacionariedade das Séries pelo teste ADF 47

4.13 Modelo ARIMA ajustado para os indivíduos do sexo masculino em Angola no período compreendido entre 1985 a 2020 48

4.14 Modelo ARIMA ajustado para os indivíduos do sexo feminino em Angola no período compreendido entre 1985 a 2020 48

4.15 Modelo ARIMA ajustado para os indivíduos do sexo masculino no Brasil no período compreendido entre 1985 a 2020 49

4.16 Modelo ARIMA ajustado para os indivíduos do sexo feminino no Brasil no período compreendido entre 1985 a 2020 49

4.17 Modelo ARIMA ajustado para os indivíduos do sexo masculino e feminino de Cabo Verde no período compreendido entre 1985 a 2020 51

4.18 Modelo ARIMA ajustado para os indivíduos do sexo masculino e feminino da Guiné-Bissau no período compreendido entre 1985 a 2020 52
4.19 Modelo ARIMA ajustado para os indivíduos do sexo masculino em Guiné Equatorial no período compreendido entre 1985 a 2020 53
4.20 Modelo ARIMA ajustado para os indivíduos do sexo feminino em Guiné Equatorial no período compreendido entre 1985 a 2020 53
4.21 Modelo ARIMA ajustado para os indivíduos do sexo masculino em Moçambique no período compreendido entre 1985 a 2020 54
4.22 Modelo ARIMA ajustado para os indivíduos do sexo feminino em Moçambique no período compreendido entre 1985 a 2020 54
4.23 Modelo ARIMA ajustado para os indivíduos do sexo masculino em Portugal no período compreendido entre 1985 a 2020 55
4.24 Modelo ARIMA ajustado para os indivíduos do sexo feminino em Portugal no período compreendido entre 1985 a 2020 55
4.25 Modelo ARIMA ajustado para os indivíduos do sexo masculino em São Tomé e Príncipe no período de 1985 a 2020. 56
4.26 Modelo ARIMA ajustado para os indivíduos do sexo feminino em São Tomé e Príncipe no período de 1985 a 2020. 56
4.27 Modelo ARIMA ajustado para os indivíduos do sexo masculino e feminino em Timor-Leste no período compreendido entre 1985 a 2020 57
4.28 Valores das métricas de desempenho dos modelos ajustados para indivíduos do sexo masculino e feminino nos países da CPLP. 58
4.29 Resultados dos testes de Ljung-Box para os modelos ajustados dos indivíduos do sexo masculino e feminino nos países da CPLP 60
4.30 Resultados dos testes de Lilliefors para os modelos ajustados dos indivíduos do sexo masculino e feminino nos países da CPLP. 60
4.31 Resultados das previsões de mortalidade infantil para Angola 61
4.32 Resultados das previsões de mortalidade infantil para Brasil 61
4.33 Resultados das previsões de mortalidade infantil para Cabo Verde 61
4.34 Resultados das previsões de mortalidade infantil para Guiné-Bissau 62
4.35 Resultados das previsões de mortalidade infantil para Guiné Equatorial 62
4.36 Resultados das previsões de mortalidade infantil para Moçambique 62
4.37 Resultados das previsões de mortalidade infantil para Portugal 62
4.38 Resultados das previsões de mortalidade infantil para São Tomé e Príncipe 63
4.39 Resultados das previsões de mortalidade infantil para Timor-Leste 63
4.40 Resultados de RMSE e MAE para Angola de acordo com a previsão de 2016-2020 64
4.41 Resultados de RMSE e MAE para o Brasil de acordo com a previsão de 2016-2020 64
4.42 Resultados de RMSE e MAE para Cabo Verde de acordo com a previsão de 2016-2020 . . 64
4.43 Resultados de RMSE e MAE para Guiné-Bissau de acordo com a previsão de 2016-2020 . 65
4.44 Resultados de RMSE e MAE para Guiné Equatorial de acordo com a previsão de 2016-2020 65
4.45 Resultados de RMSE e MAE para Moçambique de acordo com a previsão de 2016-2020 . 65

4.46 Resultados de RMSE e MAE para Portugal de acordo com a previsão de 2016-2020 65
4.47 Resultados de RMSE e MAE para São Tomé e Príncipe de acordo com a previsão de 2016-2020 65
4.48 Resultados de RMSE e MAE para Timor-Leste de acordo com a previsão de 2016-2020 . . 66
4.49 revisão, valores observados e intervalos de confiança (Lo 95 e Hi 95) para Angola (2016–2020) 66
4.50 Previsão, valores observados e intervalos de confiança (Lo 95 e Hi 95) para o Brasil (2016–2020) 66
4.51 Previsão, valores observados e intervalos de confiança (Lo 95 e Hi 95) para Cabo Verde (2016–2020) 67
4.52 Previsão, valores observados e intervalos de confiança (Lo 95 e Hi 95) para Guiné-Bissau (2016–2020) 67
4.53 Previsão, valores observados e intervalos de confiança (Lo 95 e Hi 95) para Guiné Equatorial (2016–2020) 67
4.54 Previsão, valores observados e intervalos de confiança (Lo 95 e Hi 95) para Moçambique (2016–2020) 67
4.55 Previsão, valores observados e intervalos de confiança (Lo 95 e Hi 95) para Portugal (2016–2020) 68
4.56 Previsão, valores observados e intervalos de confiança (Lo 95 e Hi 95) para São Tomé e Príncipe (2016–2020) 68
4.57 Previsão, valores observados e intervalos de confiança (Lo 95 e Hi 95) para Timor-Leste (2016–2020) 68

B.1 Valores de AIC para Modelos ARIMA - Masculino e Feminino (Antes e Depois da Remoção de Parâmetros Não Significativos) para Angola 84
B.2 Valores de AIC para Modelos ARIMA - Masculino e Feminino (Antes e Depois da Remoção de Parâmetros Não Significativos) para o Brasil 84
B.3 Valores de AIC para Modelos ARIMA - Masculino e Feminino (Antes e Depois da Remoção de Parâmetros Não Significativos) para Cabo Verde 84
B.4 Valores de AIC para Modelos ARIMA - Masculino e Feminino (Antes e Depois da Remoção de Parâmetros Não Significativos) para Guiné-Bissau 85
B.5 Valores de AIC para Modelos ARIMA - Masculino e Feminino (Antes e Depois da Remoção de Parâmetros Não Significativos) para Guiné Equatorial 85
B.6 Valores de AIC para Modelos ARIMA - Masculino e Feminino (Antes e Depois da Remoção de Parâmetros Não Significativos) para Moçambique 85
B.7 Valores de AIC para Modelos ARIMA - Masculino e Feminino (Antes e Depois da Remoção de Parâmetros Não Significativos) para Portugal 85
B.8 Valores de AIC para Modelos ARIMA - Masculino e Feminino (Antes e Depois da Remoção de Parâmetros Não Significativos) para São Tomé e Principe 86
B.9 Valores de AIC para Modelos ARIMA - Masculino e Feminino (Antes e Depois da Remoção de Parâmetros Não Significativos) paraTimor-Leste 86

Lista de Acrónimos

ACF Autocorrelation Function

ADF Augmented Dickey-Fuller

AIC Akaike Information Criterion

AR AutoRegressive

ARIMA AutoRegressive Integrated Moving Average

ARMA AutoRegressive Moving Average

BIC Bayesian Information Criterion

CPLP Comunidade dos Países de Lingua Portuguesa

MA Moving Average

MAE Mean Absolute Error

PACF Partial Autocorrelation Function

PIB Produto Interno Bruto

RMSE Root Mean Square Error

UE Universidade de Évora

URL *Uniform Resource Locator*

Sumário

Esta dissertação aborda, de forma abrangente, a evolução das taxas de mortalidade infantil nos países da CPLP, no período de 1985 a 2020, recorrendo à modelação de séries temporais com o emprego dos modelos ARIMA. Ao longo do estudo, procurou-se não só descrever e analisar os dados históricos, mas também prever o comportamento futuro destes indicadores, num contexto onde a mortalidade infantil se configura como um dos principais indicadores de saúde pública.

Num primeiro momento, a investigação delineou os objetivos e a motivação do trabalho, que se fundamenta na necessidade de reduzir as desigualdades entre os países da comunidade, nomeadamente entre os estados africanos e os de indicadores socioeconómicos mais favoráveis, como Portugal e Brasil. Neste sentido, o estudo procurou analisar as taxas de mortalidade infantil por género, permitindo identificar padrões e tendências que se revelam cruciais para a compreensão do impacto das políticas de saúde implementadas ao longo do tempo.

A aplicação dos modelos ARIMA revelou-se bastante robusta, tendo sido confirmada através de testes de diagnóstico, como o de Ljung−Box e o de Lilliefors, bem como a análise de indicadores de desempenho (RMSE, MAE e AIC). Estes procedimentos permitiram validar a adequação dos modelos ajustados, evidenciando, por exemplo, a capacidade preditiva dos mesmos e a existência de quebras estruturais que coincidem com mudanças significativas nas políticas de saúde.

Num diálogo fluido com os resultados, conclui-se que, apesar da tendência global de redução das taxas de mortalidade infantil, subsistem disparidades marcantes entre os países da CPLP. Em particular, os países africanos, tais como Angola, Guiné-Bissau, Guiné Equatorial e Moçambique, continuam a registar taxas significativamente superiores às de países como Portugal e Brasil. Este facto sublinha a importância de se desenvolverem estratégias de intervenção específicas e bem fundamentadas, que permitam atenuar estas desigualdades e promover uma melhoria nos indicadores de saúde.

Em suma, o uso dos modelos ARIMA revelou-se útil para acompanhar a evolução das taxas de mortalidade infantil ao longo do tempo nos países de língua portuguesa. Ainda que este tipo de modelo não permita compreender as causas por detrás das variações observadas, oferece uma leitura clara das tendências e flutuações que marcam o fenómeno.

Palavras chave: CPLP, Modelação, Mortalidade infantil, Previsão, Séries temporais

Abstract

Modeling and forecasting infant mortality rates in CPLP countries over the period 1985-2020 using time series models

This dissertation provides a comprehensive analysis of the evolution of infant mortality rates in the CPLP countries over the period 1985 to 2020, using time series modelling with ARIMA models. Throughout the study, the aim was not only to describe and analyse the historical data, but also to forecast the future behaviour of these indicators, in a context where infant mortality is one of the most important public health indicators.

The study began by outlining the objectives and motivation of the work, which was based on the need to reduce inequalities between countries in the Community, particularly between African countries and those with more favourable socio-economic indicators, such as Portugal and Brazil. In this context, the study aimed to analyse infant mortality rates by gender, thus allowing the identification of patterns and trends that are crucial for understanding the impact of health policies implemented over time.

The application of ARIMA models proved to be highly robust, as confirmed by rigorous diagnostic tests such as the Ljung−Box and Lilliefors tests, and by the analysis of performance indicators (RMSE, MAE and AIC). These procedures validated the suitability of the fitted models, highlighting, for example, their predictive power and the existence of structural breaks coinciding with significant changes in health policy.

In a fluid dialogue with the results, it is concluded that despite the overall trend of decreasing infant mortality rates, marked disparities persist among CPLP countries. In particular, African countries such as Angola, Guinea-Bissau, Equatorial Guinea and Mozambique continue to have significantly higher rates than countries such as Portugal and Brazil. This fact underlines the importance of developing specific and well-founded intervention strategies that can reduce these inequalities and promote improvements in health indicators.

In conclusion, the use of ARIMA models has proven useful for tracking the evolution of infant mortality rates over time in Portuguese-speaking countries. Although this type of model does not allow for an understanding of the underlying causes behind the observed variations, it provides a clear reading of the trends and fluctuations that characterise the phenomenon.

Keywords: CPLP, Forecast, Infant mortality, Models, Time series

1

Introdução

A mortalidade infantil tem sido alvo de grande preocupação por parte das entidades governamentais em todo o mundo, inclusive, da Organização das Nações Unidas (ONU). Apesar da diminuição significativa da taxa de mortalidade infantil no mundo ao longo dos anos, ainda assim se verifica em alguns países elevadas taxas de mortalidade infantil, que deveriam ser devidamente combatidas. A Organização Mundial da Saúde (OMS), define a mortalidade infantil como o número de óbitos de crianças com menos de 1 ano de idade observado durante um determinado período de tempo, normalmente um ano civil, referente ao número de nados vivos do mesmo período (habitualmente expressa em número de óbitos de crianças com menos de 1 ano por 1000 nados-vivos). De acordo com estimativas da Organização Mundial da Saúde (OMS) e do Fundo das Nações Unidas para a Infância (UNICEF), aproximadamente 2,8 milhões de mulheres grávidas e recém-nascidos morrem a cada ano, ou seja, uma morte a cada 11 segundos, principalmente devido a causas evitáveis (UNICEF, 2019). No entanto, segundo o relatório Níveis e Tendências da Mortalidade Infantil, publicado pelo Grupo Interagencial das Nações Unidas para a Estimativa da Mortalidade Infantil (UN IGME), a taxa global de mortalidade de crianças menores de cinco anos diminuiu 51% desde 2000, atingindo um mínimo histórico de 4,9 milhões de óbitos em 2022. Esses dados indicam que uma criança nascida hoje tem significativamente mais chances de sobreviver até os cinco anos de idade em comparação com há três décadas (UNICEF, 2024).

A redução da mortalidade infantil tem sido uma prioridade para os líderes mundiais, reconhecendo-a como um indicador fundamental da saúde de uma nação. Essa preocupação foi formalizada em 2000, quando as Nações Unidas estabeleceram os Objetivos de Desenvolvimento do Milénio (ODM), consistindo em oito metas globais para promover o desenvolvimento humano. O quarto objetivo desses ODM era "Reduzir a mortalidade infantil", com a meta específica de diminuir em dois terços, entre 1990 e 2015, a taxa de mortalidade de crianças com menos de cinco anos (Nações Unidas, 2000).

Diante do exposto, surgiu o interesse em realizar este estudo que visa perceber como tem sido a evolução da mortalidade infantil, por sexo, dentro da Comunidade dos Países de Língua Portuguesa (CPLP), nomeadamente Angola, Brasil, Cabo Verde, Guine Bissau e Equatorial, Moçambique, Portugal, São Tomé e Príncipe e Timor-Leste, pois, em geral a taxa de mortalidade infantil é um indicador (visto externamente) como um indicador de saúde de um país. Taxas elevadas de mortalidade infantil geralmente refletem condições precárias de vida e saúde, bem como um baixo nível de desenvolvimento socioeconómico (Ministério da Saúde, 2021).

De modo a perceber a evolução das taxas de mortalidade infantil, serão utilizados para o tratamento dos dados a modelação das séries temporais, que se baseia na análise de uma variável ao longo de um certo período de tempo, cujo o objetivo é encontrar modelos matemáticos que consigam descrever a sua evolução, em termos da sua tendência e/ou sazonalidade (ou ciclos) e, a partir destes, prever o comportamento futuro, casos as condições atuais permaneçam relativamente as mesmas (Pires, 2021).

1.1 Objetivos

O objetivo desta dissertação consiste num estudo aprofundado das taxas de mortalidade infantil, por sexo, nos países da CPLP, no período de 1985 a 2020. Especificamente, pretende-se: 1. Realizar uma análise descritiva da evolução das taxas de mortalidade infantil, identificando padrões e variações ao longo do tempo; 2. Aplicar técnicas estatísticas para avaliar a existência de quebras estruturais nas séries temporais, nomeadamente alterações de nível e/ou de tendência, com o intuito de compreender dinâmicas subjacentes à evolução dos dados; 3. Ajustar e estimar modelos SARIMA, garantindo uma modelação adequada à estrutura das séries temporais analisadas; 4. Efetuar previsões de valores futuros com base nos modelos estimados, permitindo inferir sobre possíveis trajetórias futuras da mortalidade infantil nos países em estudo.

1.2 Resumo das características gerais dos países da CPLP

A CPLP foi oficialmente constituída a 17 de julho de 1996, em Lisboa, Portugal, durante a I Conferência de Chefes de Estado e de Governo da CPLP. Nesta ocasião, estiveram presentes os Chefes de Estado e de Governo de Angola, Brasil, Cabo Verde, Guiné-Bissau, Moçambique, Portugal e São Tomé e Príncipe (CPLP, 1996). Atualmente, Timor-Leste e a Guiné Equatorial integram a CPLP, tendo aderido em maio de 2002 e julho de 2014, respetivamente.

Por outro lado, de modo a perceber melhor a realidade de cada país membro da CPLP, foi elaborada a tabela abaixo 1.1.

País	População (2022)	Dimensão territorial (km^2)	PIB (USD)
Angola	34.511.514	1.246.700	121,6 mil milhões
Brasil	215.680.045	8.515.767	1,92 bilhões
Cabo Verde	566.252	4.033	2,283 mil milhões
Guiné-Bissau	2.039.724	36.125	1,633 mil milhões
Guiné Equatorial	1.462.408	28.051	12,825 mil milhões
Moçambique	32.560.330	801.590	17,919 mil milhões
Portugal	10.467.366	92.090	239 mil milhões
São Tomé e Príncipe	226.575	964	518 milhões
Timor-Leste	1.365.995	14.874	3,041 mil milhões
Total	**298.533.713**	**10.740.194**	–

Tabela 1.1: Dados populacionais e de PIB dos estados membros da CPLP para o ano de 2022. Fonte: Countrymeters (`https://countrymeters.info/em`) e `https://www.thetruesize.com/`.

De acordo aos dados tabelados, no ano de 2022, se compararmos a CPLP a uma única nação, ela é composta por cerca de 298 milhões de habitantes numa área de quase 11 milhões km^2. É digno de realce, que quase todos os estados membros da CPLP, com exceção de Guiné Equatorial, estiverem sob o domínio do Estado português, durante muitos anos, até alcançarem a sua independência. No entanto, os membros da CPLP estão distribuídos em quatro continentes, nomeadamente, África, América, Ásia e Europa. O Brasi destaca-se como o maior estado membro da CPLP quer em termos populacionais com cerca de 215 milhões de habitantes como também em extensão territorial, 8.515.767 km^2, o que perfaz aproximadamente 80% da superfície territorial da CPLP. Além disso, apresenta o maior Produto Interno Bruto (PIB) da CPLP – indicador do estado de um país, do seu nível de produção com contrapartida monetária (Jacquinet , 2019). Por outro lado, o estado de São Tomé e Príncipe representa o menor estado da CPLP com cerca de 226.575 habitantes para uma extensão territorial de 964 km^2 e o menor PIB referente ao ano de 2022.

A seguir faremos uma breve caracterização de cada estado membro da CPLP, começando com Angola.

Angola é um país da costa ocidental de África e com uma população de cerca de 34 milhões, sendo o segundo estado membro da CPLP com a maior população e extensão territorial, perdendo apenas para o Brasil. A esperança de vida dos homens à nascença é de 59 anos e da mulher é de 64 anos (Country Economy, 2022).

A economia angolana é baseada na extração de recursos minerais como o petróleo, o gás natural e diamantes. Além disso, uma boa parte da população tem o setor agrícola na geração de renda, sobretudo para a população rural (CIA, 2022).

Segundo os dados do Banco Mundial (World Bank, 2018), em 2015, Angola apresentou uma taxa de alfabetização de 73,3%. No entanto, havia uma disparidade de género, com taxas mais altas entre os homens (82%) do que entre as mulheres (60,7%). Por outro lado, em 2017, apenas metade da população angolana tinha acesso ao saneamento básico, um outro indicador importante para medir o desenvolvimento de um país (UNICEF, 2017).

Relativamente à taxa de mortalidade infantil, em 2017, Angola apresentava uma das maiores taxas no

mundo e, consequentemente, dentro da CPLP. De acordo com a UNICEF (2022), a taxa de mortalidade infantil no país era de 44 mortes por cada 1000 nascidos vivos. Contudo, a taxa de mortalidade infantil em 2020 foi de 48%, e, mesmo apresentando valores mais baixos em relação ao passado, a situação ainda é preocupante em Angola, assim como em vários países da comunidade.

Em sequência temos o **Brasil**, o maior estado membro da CPLP em território e população, sendo o quinto maior país do mundo em superfície territorial, localizado na América do Sul. De acordo com dados do Instituto Brasileiro de Geografia e Estatística (IBGE), a esperança de vida à nascença no Brasil em 2021 era de 73,6 anos para os homens e 80,5 anos para as mulheres, tornando o país no segundo com a maior esperança de vida da CPLP.

Quanto ao tipo de economia, o Brasil possui uma economia emergente, caracterizada por uma grande diversificação nos setores energéticos, industriais, minerais e agrícolas. De acordo com o IBGE (2023), o setor de serviços representou 62,6% do PIB do país, seguido pela indústria com 18,4% e pela agricultura com 4,4%. A soma de todos os bens e serviços produzidos pelo Brasil no ano de 2022 foi cerca de 1,92 trilião de dólares, apresentando assim o maior PIB da comunidade e a nona maior na economia mundial. Não obstante, o Brasil ainda enfrenta uma grande desigualdade social e uma má distribuição do rendimento.

Segundo o IBGE, em 2019, a taxa de alfabetização no Brasil era de 93,4%, o que indica que aproximadamente 11,3 milhões de pessoas com 15 anos ou mais ainda eram analfabetas no país. No que se refere ao saneamento básico, o Sistema Nacional de Informações sobre Saneamento (SNIS) reportou que, em 2019, cerca de 83,3% da população brasileira tinha acesso a esses serviços. A insuficiência desses serviços resultou em 273.403 internamentos por doenças relacionadas à falta de saneamento no mesmo ano.

No que se refere à mortalidade infantil, no ano de 2020, o Brasil apresentou uma taxa de 10,1 óbitos por 1000 nascidos vivos, posicionando-se como a segunda menor taxa dentro da CPLP. Esse índice representa uma redução significativa em relação às décadas anteriores e reflete avanços nos cuidados de saúde infantil no país. No entanto, apesar da melhoria, a situação ainda exige atenção contínua para garantir a saúde e o bem-estar das crianças brasileiras (UNICEF, 2022).

Cabo Verde, situado na África Ocidental, é composto por dez ilhas e possui uma população de aproximadamente 491.233 habitantes (INE Cabo Verde, 2022). De acordo com o Instituto Nacional de Estatística de Cabo Verde (INE Cabo Verde, 2021), a esperança de vida à nascença era de 73,0 anos para os homens e 80,5 anos para as mulheres, refletindo um aumento significativo ao longo das últimas décadas. Por outro lado, relativamente à taxa de alfabetização, cerca de 88,9% da população cabo-verdiana com 15 ou mais anos era alfabetizada, um dado satisfatório, especialmente considerando que, há quatro décadas, o analfabetismo no país superava os 60%, colocando agora Cabo Verde numa posição favorável dentro da CPLP e do continente africano (Expresso das Ilhas, 2022).

A economia de Cabo Verde é predominantemente orientada para o setor de serviços, com destaque para o turismo, que representa uma parcela significativa do PIB. De acordo com o Banco de Cabo Verde (2022), em 2021, o PIB de Cabo Verde foi de aproximadamente 1,94 mil milhões de dólares, registando um crescimento de 7% em relação ao ano anterior. No ano seguinte, o país apresentou um crescimento econômico de 17,7%, impulsionado pela recuperação do turismo e outras atividades econômicas (Banco de Cabo Verde, 2023). Além do turismo, outros setores importantes incluem os serviços e as remessas dos emigrantes, que contribuem significativamente para a economia nacional.

Segundo a Organização Mundial da Saúde (OMS, 2017), Cabo Verde apresentava uma boa cobertura de acesso ao saneamento básico dentro da CPLP, sendo a segunda melhor taxa, atrás apenas de Portugal, com cerca de 83%. Esse dado reflete os avanços do país em infraestrutura sanitária e acesso a serviços essenciais de saúde pública.

Relativamente à mortalidade infantil, Cabo Verde apresenta a terceira menor taxa dentro da CPLP, sendo a melhor entre os Estados-membros do continente africano. Além disso, destaca-se por ter uma das mais baixas taxas de mortalidade infantil em toda a África (INE Cabo Verde, 2024).

Depois de ver como Cabo Verde está nesses indicadores, vamos dar uma olhada na Guiné-Bissau, que vive uma realidade bem mais dura. A **Guiné-Bissau**, está localizada na África Ocidental, enfrenta desafios significativos no setor da saúde, refletidos nos indicadores de esperança de vida e mortalidade infantil. De acordo com o Instituto Nacional de Estatística da Guiné-Bissau (INE Guiné-Bissau, 2020), a esperança de vida ao nascer é de 50,10 anos, sendo 49,24 anos para os homens e 51,17 anos para as mulheres, colocando o país entre os de menor esperança de vida dentro da CPLP. Além disso, a taxa de mortalidade infantil continua elevada. Segundo o Relatório Anual do UNICEF Guiné-Bissau (2023), 51 crianças por 1000 nascidos vivos morrem antes de completar cinco anos de idade, enquanto 22 bebés por 1000 nascidos vivos falecem durante os primeiros 28 dias de vida. Esses números evidenciam a necessidade urgente de investimentos em infraestrutura de saúde e políticas de proteção materno-infantil para melhorar as condições de vida da população.

Com uma população estimada de aproximadamente 2 milhões de habitantes, apenas cerca de 33% têm acesso ao saneamento básico, conforme destacado por Salla et al. (2019). Essa situação é alarmante, pois a falta de saneamento adequado representa uma ameaça significativa à saúde pública, aumentando a incidência de doenças de veiculação hídrica. Investimentos em saneamento básico são fundamentais para a melhoria das condições de saúde da população. Estudos indicam que, para cada dólar investido em saneamento básico, há uma economia aproximada de 4,3 dólares em despesas de saúde, reforçando a importância desse setor como indicador do desenvolvimento social de uma sociedade.

De acordo com dados do Banco Mundial de 2017, o PIB do país é de aproximadamente 2 bilhões de dólares. A economia guineense é predominantemente agrícola, com destaque para a exportação de castanha de caju, que representa uma parcela significativa das receitas de exportação. Além disso, produtos como peixe, camarão, amendoim, palmito e madeira serrada também contribuem para a pauta de exportações do país.

A caracterização dos países da CPLP, continua, desta feita com a recente nação à aderir a CPLP, a **Guiné Equatorial**, localizado na África Ocidental, é o mais recente país a aderir à CPLP, tendo se tornado membro em 2014 (CPLP, s.d.). Além do português, adotado como língua oficial em 2010, o país também reconhece o espanhol e o francês como línguas oficiais. A população da Guiné Equatorial é estimada em aproximadamente 1,4 milhões de habitantes, de acordo com os dados do Banco Mundial (2021).

A economia da Guiné Equatorial é fortemente dependente da exploração de petróleo e serviços correlatos - atividades económicas auxiliares ligadas diretamente à cadeia de valor do setor petrolífero, de forma semelhante a outro Estado-membro da CPLP, Angola. Apesar da riqueza gerada pelo setor petrolífero, a maioria da população vive abaixo do limiar da pobreza, refletindo uma elevada desigualdade social (CIA, 2023). No entanto, a exportação de petróleo e outros bens, juntamente com o setor de serviços, contribuiu para que o PIB do país atingisse aproximadamente 12 mil milhões de dólares no ano de 2022 (INEGE, 2023).

Além disso, enfrenta desafios significativos no acesso a serviços essenciais. Apenas 48% da população tem acesso a água potável, sendo esta uma das taxas mais baixas do continente, o que representa uma preocupação grave para a saúde pública (Mundiconsulting, 2017). No entanto, a cobertura de saneamento básico melhorado abrange cerca de 75% da população, um indicador positivo para a região, pois reflete esforços no desenvolvimento de infraestrutura sanitária (Mundiconsulting, 2017).

Outrossim, segundo os dados da UNICEF (2022), a taxa de mortalidade infantil da Guiné Equatorial em 2020 foi de 59,68 mortes por 1000 nascidos vivos, situando-se atrás apenas de Angola na CPLP.

Moçambique, localizado na costa oriental de África, possui uma população estimada em aproximadamente 32 milhões de habitantes. A esperança de vida ao nascer no país é de 56,5 anos para os homens e 62,7 anos para as mulheres, colocando Moçambique entre os países com menor esperança de vida no mundo e, consequentemente, na CPLP (Banco Mundial, 2024).

Segundo o Portal do Governo Moçambicano, o potencial económico do país passa pelos setores das agro-indústrias, agricultura, pescas, mineração e turismo. Moçambique, também é um país rico em recursos minerais, destacando-se o carvão, o gás natural, o titânio e outros minerais pesados. Em 2022, o PIB moçambicano cresceu 4,4% em relação ao ano de 2021, atingindo aproximadamente 18,9 mil milhões de dólares (INE Moçambique, 2023). No entanto, apesar do crescimento económico, cerca de 65% da população ainda vive abaixo do limiar da pobreza, evidenciando a necessidade de políticas mais eficazes de inclusão social e distribuição de renda (O País, 2024).

De acordo com a Organização Mundial da Saúde (OMS, 2017), Moçambique tinha uma cobertura de saneamento básico de apenas 31%, um nível bastante preocupante para este indicador essencial, tornando-o o país com a pior referência nesta área entre os membros da CPLP. Por outro lado, o Instituto Nacional de Estatística de Moçambique tem destacado esforços para melhorar o acesso ao saneamento básico, incluindo a construção de latrinas e fontes de abastecimento de água potável. Essas iniciativas estão alinhadas com os Objetivos de Desenvolvimento Sustentável (ODS) de Moçambique até 2030, que visam garantir a disponibilidade e a gestão sustentável da água potável e do saneamento para todos. A intervenção no saneamento básico é essencial para a redução da desnutrição, pois cerca de 38% das crianças moçambicanas com menos de 5 anos sofrem de desnutrição crónica, uma condição que pode ser agravada pela falta de saneamento adequado (UNICEF Moçambique, 2022).

No que tange à mortalidade infantil em Moçambique, observa-se uma redução ao longo dos anos; contudo, os índices ainda são considerados elevados. Em 2020, a taxa de mortalidade infantil foi de 64,72 óbitos por mil nascimentos (UNICEF Moçambique, 2022). Segundo a UNICEF, fatores como pobreza, desnutrição, acesso limitado a serviços de saúde de qualidade e a baixa escolaridade materna desempenham um papel crucial na manutenção dessas taxas elevadas.

Portugal, situado no sudoeste da Europa, possui uma população de aproximadamente 10,4 milhões de habitantes, incluindo os arquipélagos da Madeira e dos Açores (INE, 2023). Em relação ao saneamento básico, mais de 90% da população portuguesa tem acesso a esses serviços essenciais, estabelecendo um padrão elevado na CPLP (ERSAR, 2023). Além disso, a esperança de vida à nascença em 2021 era de 78,05 anos para os homens e 83,52 anos para as mulheres, posicionando Portugal como uma referência tanto na CPLP quanto a nível mundial (PORDATA, 2023).

Segundo o Portal Diplomático dos Negócios Estrangeiros, a economia portuguesa é caracterizada pela sua abertura e dependência do comércio internacional, com grande ênfase no turismo, serviços financeiros e indústrias como têxteis, madeira e cortiça. Em 2022, o PIB de Portugal foi de aproximadamente 239,4 mil milhões de euros, refletindo a recuperação económica do país (PORDATA, 2023).

No que diz respeito à mortalidade infantil, apesar do ligeiro aumento em 2022 para 2,6 óbitos por cada 1000 nascimentos vivos, Portugal continua a apresentar os valores mais baixos dentro da CPLP e é uma referência a nível mundial (INE, 2023).

Em sequência, abordamos São Tomé e Príncipe, situado na costa ocidental do continente africano, sendo o estado membro da CPLP com a menor população, estimada em aproximadamente 226.575 habitantes em 2022 (PopulationPyramid, 2022).

Segundo os dados do Programa das Nações Unidas para o Desenvolvimento (PNUD), em 2018, São Tomé e Príncipe tinha uma taxa de alfabetização de 90,1% (INE, 2017). No que se refere à esperança de vida,

em 2021, um homem são-tomense esperava viver, em média, até 64 anos, enquanto uma mulher já rondava os 70 anos (INE, 2021).

De acordo com o Banco Mundial (2023), a economia de São Tomé e Príncipe é vulnerável a choques externos e às alterações climáticas. Todavia, o país possui uma vasta riqueza natural que não é plenamente explorada, sendo a sua biodiversidade um recurso subutilizado para o fomento do turismo local.

Segundo o Fundo Monetário Internacional (FMI, 2023), o PIB reduziu de 1,9% em 2021 para 1,4% em 2022, devido à pandemia da COVID-19, que impactou a economia global. Além disso, o abrandamento do crescimento económico foi agravado por falhas energéticas, inundações no final de 2021 e início de 2022, e pelo aumento dos preços globais de alimentos e combustíveis.

No que se refere ao acesso ao saneamento básico, em 2010, apenas 30% da população urbana e 19,2% da população rural tinham acesso a esses serviços, refletindo uma baixa cobertura que representa um grave problema de saúde pública (Ministério da Saúde de São Tomé e Príncipe, 2022). A deficiência nesse setor resulta em doenças de origem hídrica, como diarreias e doenças renais, afetando principalmente as populações mais vulneráveis.

A taxa de mortalidade infantil em São Tomé e Príncipe, em 2021, foi estimada em 35 óbitos por 1000 nascimentos, colocando-o na lista dos países com baixas taxas de mortalidade infantil dentro da comunidade (INE, 2021).

Por último, caracterizamos Timor-Leste, localizado no sudeste asiático e sendo o único Estado-membro da CPLP nesse continente. O país conquistou sua independência em 2002, após um longo período de ocupação indonésia, tendo sido anteriormente uma colónia portuguesa até 1975 (Ministério dos Negócios Estrangeiros de Portugal, 2023).

A população de Timor-Leste é estimada em aproximadamente 1,3 milhão de habitantes, sendo classificado entre os países mais pobres do mundo. A esperança de vida à nascença em 2022 foi de 66,10 anos para homens e 69,54 anos para mulheres (Country Economy, 2022). Apesar dos desafios socioeconómicos, o país tem avançado na melhoria dos padrões de vida da sua população (Banco Mundial, 2022).

A economia timorense baseia-se predominantemente na agricultura de subsistência, com destaque para a produção de arroz, café e milho. No entanto, o país possui um forte potencial para o turismo e conta com recursos naturais significativos, como petróleo e gás natural, que representam cerca de 85% da renda nacional (Banco Mundial, 2021). Segundo dados do Banco Mundial, em 2021, o PIB foi estimado em 3,62 mil milhões de dólares, colocando Timor-Leste entre as economias menos expressivas no cenário mundial.

No que se refere ao acesso ao saneamento básico, de acordo com os dados obtidos pelo *Joint Monitoring Programme – JMP* (2023), apenas 56,8% da população timorense tinha acesso a instalações sanitárias adequadas em 2020, sendo um dos valores mais baixos da CPLP e um grande desafio para a saúde pública no país.

Em relação à mortalidade infantil, segundo os dados da UNICEF (2021), apesar dos avanços nos últimos anos, os números ainda são preocupantes para a população timorense. A taxa de mortalidade infantil foi estimada em aproximadamente 43,07 óbitos por 1000 nascimentos vivos.

Contudo, embora a situação de cada Estado-membro da CPLP seja única, problemas como o acesso ao saneamento básico e as elevadas taxas de mortalidade infantil são comuns, principalmente nos países africanos do bloco. Além disso, a pandemia da COVID-19 agravou a vulnerabilidade económica dessas nações, impactando significativamente suas trajetórias de desenvolvimento (Banco Mundial, 2022).

2

Séries Temporais

2.1 Conceitos

Segundo Morettin e Toloi (2004), uma série temporal pode ser considerada como uma das várias trajetórias possíveis de um processo estocástico, sendo que os modelos utilizados para descrever séries temporais são, por sua vez, representações desses processos, os quais são governados por leis probabilísticas. Neste contexto, Hamilton (1994), destaca que, ao contrário dos dados transversais, os valores de uma série temporal em momentos subsequentes estão intrinsecamente relacionados, tornando a preservação da ordem cronológica um fator essencial para a sua análise. De acordo com Chatfield (2003), esta dependência ao longo do tempo é a principal característica que distingue as séries temporais de outros tipos de dados.

Além disso, as séries temporais podem ser classificadas de duas maneiras: primeiramente, quanto à natureza dos instantes de tempo em que são observadas podendo ser discretas ou contínuas e, em segundo lugar, quanto ao conjunto de valores que podem assumir também podendo ser discretos ou contínuos (Pires, 2001). Esta distinção está diretamente associada à forma como os dados são recolhidos e analisados, influenciando a escolha dos modelos mais adequados para a sua representação e previsão. Assim, compreender a estrutura subjacente das séries temporais, bem como as suas propriedades estocásticas, é fundamental para garantir a precisão das inferências e previsões geradas a partir desses dados (Box et al., 2015; Shumway & Stoffer, 2017).

2.1.1 Objetivos da análise de uma série temporal

Os principais objetivos da análise de séries temporais variam consoante a aplicação, mas, de forma geral, agrupam-se em três categorias fundamentais: descrição, modelação e previsão (Morettin e Toloi, 2004). Box et al. (2015), destacam que a descrição tem como propósito resumir as características essenciais da série, tais como a média, a variância e outros momentos estatísticos. Além disso, esta etapa inclui a identificação de padrões visuais, como tendências ou sazonalidades, que ajudam a compreender o comportamento da série ao longo do tempo. Shumway & Stoffer (2017), salientam mesmo que, através da descrição, o analista pode detetar irregularidades ou mudanças estruturais que devem ser tidas em conta antes de avançar para as fases seguintes da análise.

O segundo objetivo é a modelação, que consiste em definir um modelo estatístico capaz de descrever com precisão a estrutura da série temporal. Conforme explicado por Montgomery (2015), a modelação envolve a escolha de um modelo que capte as dependências temporais entre os dados, como o ARIMA (AutoRegressive Integrated Moving Average), amplamente utilizado em diferentes domínios. A modelação permite compreender os mecanismos que regem a série, fornecendo uma estrutura matemática útil para prever valores futuros. A identificação de padrões como autocorrelação, tendências e sazonalidade é essencial nesta fase, pois influencia diretamente a seleção do modelo mais adequado.

Outro objetivo fundamental da análise de séries temporais, conforme apontado por Chatfield (2003), é a previsão. Esta consiste em extrapolar os valores futuros com base nos dados observados até ao momento, sendo uma ferramenta essencial para o planeamento em áreas como a economia, a saúde e o clima. Segundo Hyndman & Athanasopoulos (2018), a previsão pode abranger diferentes horizontes temporais, podendo ser de curto, médio ou longo prazo, dependendo das necessidades do analista e da série em questão. Além disso, as previsões podem ser de um único passo à frente, quando se estima apenas o próximo valor, ou de múltiplos passos à frente, antecipando vários períodos futuros. A fiabilidade dessas previsões depende da qualidade do modelo ajustado e da estabilidade dos padrões observados ao longo do tempo.

Por fim, um objetivo complementar, mas não menos importante, é o controlo. Em certos casos, a análise de séries temporais é utilizada para monitorizar e ajustar processos em tempo real, como no controlo de qualidade em indústrias ou na gestão de políticas públicas. O objetivo do controlo é garantir que o processo se mantenha dentro de limites aceitáveis, permitindo intervenções rápidas sempre que a série apresente desvios inesperados. Este objetivo assume especial relevância em séries que são atualizadas continuamente, pois possibilita ajustes dinâmicos para otimizar os resultados futuros (Box et al., 2015).

2.1.2 Componentes de uma série temporal

A análise de séries temporais baseia-se na identificação de diferentes componentes que influenciam o comportamento da variável ao longo do tempo. Esses componentes são essenciais para compreender a estrutura da série e determinar a melhor abordagem para modelar e prever os seus valores futuros. Segundo Delurgio (1998), uma série temporal pode ser decomposta em quatro componentes principais: tendência, sazonalidade, ciclicidade e aleatoriedade. Cada um destes desempenha um papel crucial na análise e na construção de modelos preditivos.

A tendência refere-se ao movimento de longo prazo da série, podendo ser crescente, decrescente ou estável. Kendall & Ord (1990), explicam que a tendência capta a direção geral dos dados ao longo do tempo, muitas vezes refletindo mudanças estruturais ou padrões de crescimento e declínio em variáveis como o PIB ou a mortalidade infantil. Em muitas séries temporais, a tendência é o primeiro componente a ser identificado, pois descreve o comportamento subjacente que não se deve a flutuações de curto prazo. Por exemplo, uma tendência ascendente nas taxas de mortalidade infantil pode indicar a necessidade de políticas de

saúde pública mais eficazes, enquanto uma tendência descendente pode refletir melhorias em intervenções sanitárias.

A sazonalidade é outro componente importante, representando padrões regulares que ocorrem em intervalos específicos de tempo, como meses ou estações do ano. Para Shumway & Stoffer (2017), a sazonalidade reflete flutuações repetitivas que surgem em ciclos inferiores a um ano, influenciadas por fatores como o clima, festividades ou comportamentos sociais. Em séries temporais de saúde, como a mortalidade infantil, a sazonalidade pode capturar variações associadas a epidemias sazonais, como gripes, que afetam a mortalidade de forma previsível a cada ano. A inclusão da sazonalidade nos modelos de séries temporais permite prever essas variações recorrentes com maior precisão.

A componente cíclica distingue-se da sazonalidade, uma vez que os seus padrões de repetição são mais irregulares e ocorrem, geralmente, em intervalos superiores a um ano. Segundo Enders (2014), os ciclos representam flutuações económicas ou demográficas de médio e longo prazo, que não têm uma duração fixa, como recessões económicas ou ciclos de migração populacional. Estes ciclos podem ter efeitos significativos na série temporal, especialmente em séries económicas ou demográficas, como a mortalidade infantil, que pode variar em função das condições económicas ou das políticas de saúde pública.

Por fim, a componente irregular, ou ruído, representa a parte da série que não pode ser explicada pelos componentes anteriores. Além disso, Wei (2006), destaca que esta componente é imprevisível e está associada a variações aleatórias que não seguem nenhum padrão ou regularidade identificável. Embora o ruído possa ser minimizado através de técnicas de alisamento ou filtragem, é impossível eliminá-lo completamente. Em muitos casos, esta variação irregular reflete eventos imprevisíveis, como desastres naturais ou crises políticas, que podem causar flutuações temporárias nas taxas de mortalidade infantil.

Essas quatro componentes tendência, sazonalidade, ciclos e irregularidade formam a base da análise de séries temporais, permitindo aos analistas compreender melhor as forças que influenciam a evolução de uma variável ao longo do tempo. Ao decompor uma série temporal nestes componentes, torna-se possível modelar e prever o seu comportamento futuro com maior precisão, ajustando-se às particularidades de cada componente para maximizar a fiabilidade das previsões (Hyndman & Athanasopoulos, 2018).

2.2 Processos estocásticos

Os processos estocásticos são fundamentais na análise de séries temporais, pois permitem modelar e compreender a aleatoriedade e a incerteza que frequentemente caracterizam dados observados ao longo do tempo (Pires, 2001). Um processo estocástico é uma coleção de variáveis aleatórias que são indexadas no tempo, onde cada variável representa um estado do processo em um momento específico.

Os processos estocásticos podem ser utilizados para descrever e prever a evolução de séries temporais que não são determinísticas. Segundo Shumway & Stoffer (2017), eles podem ser classificados em diferentes tipos, dependendo das suas propriedades, como estacionariedade e dependência temporal. Um exemplo comum é o *passeio aleatório*, que é um tipo de processo estocástico em que o próximo valor é uma função do valor atual e de um erro aleatório. A equação que descreve um passeio aleatório é dada por:

$$X_t = X_{t-1} + \varepsilon_t \tag{2.1}$$

onde ε_t são erros puramente aleatórios (ou seja, uma sucessão de variáveis aleatórias independentes e identicamente distribuídas (i.i.d), em geral de média nula) no tempo t ($t \in \mathbb{Z}$). Essa abordagem é útil para modelar a natureza imprevisível de muitos fenómenos, como por exemplo, as flutuações de preços no

mercado financeiro.

Um outro exemplo importante de processo estocástico é o *ruído branco*. De acordo com Cowpertwait & Metcalfe (2009), um processo estocástico ε_t é um ruído branco discreto *(DWN - Discrete White Noise)* se as variáveis ε_t forem não correlacionadas, de média e variância constantes, usualmente considera-se que a média é nula.

2.2.1 Processos estocásticos estacionários

Os processos estocásticos estacionários são fundamentais na análise de séries temporais, pois a estacionariedade é uma propriedade desejável que permite simplificar a modelação e a estimação dos parâmetros de uma série temporal. Além disso, um processo estocástico é considerado estacionário se suas propriedades estatísticas não dependem do tempo Hamilton (1994).

Os processos estocásticos estacionários podem ser classificados em dois tipos principais: *estacionários fortes* e *estacionários fracos*. O processo é estacionário forte quando todas as distribuições conjuntas de qualquer subconjunto da sequência temporal são idênticas em relação ao tempo (Shumway & Stoffer, 2017). Ou seja, para quaisquer instantes $t_1, t_2, \ldots, t_k$ e quaisquer deslocamentos h, a distribuição conjunta:

$$(X_{t_1}, X_{t_2}, \ldots, X_{t_k})$$

é a mesma que

$$(X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_k+h})$$

A estacionaridade forte é uma propriedade exigente e, em geral, de difícil verificação. Sabe-se que os momentos caracterizam parcialmente uma variável aleatória ou um processo estocástico. Assim, é razoável definir um conceito de estacionaridade baseado nos momentos de uma série temporal, tornando a sua verificação mais acessível do que no caso da estacionaridade forte (Pires, 2001).

Por outro lado, a estacionariedade fraca requer que a média $E[X_t]$ e a variância $Var[X_t]$ sejam constantes, ou seja:

$$E[X_t] = \mu, \quad Var(X_t) = \sigma^2, \quad \forall t \in \mathbb{Z} \tag{2.2}$$

e que a autocovariância deve depender apenas da distância entre os pontos no tempo, e não dos próprios tempos:

$$\mathsf{Cov}(X_t, X_{t+h}) = \mathsf{Cov}(X_s, X_{s+h}), \quad \forall t, s \in \mathbb{Z}, \forall h \in \mathbb{Z} \tag{2.3}$$

onde h representa o intervalo de tempo (Morettin & Toloi, 2004; Hyndman & Athanasopoulos, 2018). A análise de autocovariância é uma ferramenta valiosa para verificar a estacionariedade de uma série temporal, pois se as autocovariâncias forem constantes ao longo do tempo, a série pode ser considerada estacionária.

Uma maneira comum de transformar uma série não estacionária em uma série estacionária é a diferenciação. O processo de diferenciação envolve calcular as diferenças entre valores consecutivos da série, o que muitas vezes estabiliza a média, como demonstrado por Box et al. (2015). Essa técnica é especialmente útil em séries que apresentam tendências ou sazonalidade, por exemplo, na a aplicação de modelos como o ARIMA que pressupõem estacionariedade.

2.2.2 Processos estocásticos não estacionários

Os **processos estocásticos não estacionários** são aqueles cujas propriedades estatísticas, como média e variância, variam ao longo do tempo. Além disso a não estacionariedade também é devida a sazonalidade. Essa característica pode dificultar a análise e a modelação de séries temporais (Wei, 2006; Shumway & Stoffer, 2017).

Uma das principais razões para a não estacionariedade é a presença de tendências na série temporal. Uma tendência é uma mudança sistemática no nível da série ao longo do tempo, que pode ser crescente ou decrescente. Por exemplo, as séries de dados de crescimento populacional frequentemente exibem uma tendência de aumento ao longo dos anos, conforme descrito por Hamilton (1994). Neste contexto, a média da série não se mantém constante, o que é um requisito fundamental para a estacionariedade. Um dos modelos mais simples para representar séries temporais com tendência linear, é definido da seguinte forma:

$$X_t = \mu + \beta t + \varepsilon_t, \quad t \in \mathbb{Z} \tag{2.4}$$

onde:

- X_t representa a variável da série temporal no instante t;
- μ é uma constante que define o nível médio da série;
- β é o coeficiente de tendência, que indica a taxa de crescimento ou decréscimo ao longo do tempo;
- ε_t é um termo de erro aleatório com esperança nula e variância finita, geralmente assumido como um processo de ruído branco com distribuição normal $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$.

O valor esperado da série pode ser obtida da seguinte forma:

$$E[X_t] = E[\mu + \beta t + \varepsilon_t]. \tag{2.5}$$

Dado que μ e βt são determinísticos e que o termo de erro satisfaz $E[\varepsilon_t] = 0$, segue que:

$$E[X_t] = \mu + \beta t. \tag{2.6}$$

Esta equação demonstra que a média da série depende do tempo t, o que implica que a série não é estacionária, uma vez que a sua esperança não é constante ao longo do tempo.

Outra forma de não estacionariedade é a sazonalidade, que se refere a padrões recorrentes em intervalos regulares de tempo. Por exemplo, as vendas de produtos em determinados meses do ano podem ser consistentemente mais altas devido a fatores sazonais, como feriados ou climas favoráveis. A presença de sazonalidade implica que a série não é estacionária em nível ou variância. Além disso, Pollock (1999) refere que a sazonalidade pode ser modelada através de um modelo aditivo, onde a componente sazonal é representada por uma soma de funções trigonométricas. Um dos métodos mais comuns para isso é a decomposição da série temporal numa soma de senos e cossenos, captando as variações periódicas nos dados.

A componente sazonal de uma série temporal pode ser modelada como:

$$X_t = \sum_{j=1}^{n} \left(\alpha_j \cos\left(\frac{2\pi j}{s} t\right) + \beta_j \sin\left(\frac{2\pi j}{s} t\right) \right) + \varepsilon_t \tag{2.7}$$

onde:

- s é o número de períodos por ciclo sazonal;
- j representa a ordem harmónica da decomposição, permitindo captar variações sazonais mais complexas;
- α_j e β_j são coeficientes a serem estimados, representando a amplitude das oscilações;
- ε_t é um termo de erro aleatório, frequentemente assumido como ruído branco gaussiano.

Os coeficientes α_j e β_j podem ser estimados através de regressão linear, ajustando a função trigonométrica aos dados. Esse método é útil pois permite descrever a sazonalidade sem necessidade de introduzir variáveis dummy para cada período.

Os processos estocásticos não estacionários podem ser classificados em duas categorias principais: não estacionariedade em nível e não estacionariedade em variância. A não estacionariedade em nível ocorre quando a média da série varia ao longo do tempo, enquanto a não estacionariedade em variância verifica-se quando a variância da série também se altera.

Para além destas, existe uma terceira categoria relevante: a não estacionariedade sazonal, que ocorre quando os padrões da série temporal variam periodicamente ao longo do tempo. Este tipo de não estacionariedade é comum em séries económicas, climatológicas e financeiras, onde a sazonalidade reflete ciclos naturais ou comportamentais previsíveis.

Um exemplo clássico de não estacionariedade em nível é o passeio aleatório (Enders, 2014; Box et al., 2015). Já a não estacionariedade sazonal pode ser observada em séries temporais cujos padrões se repetem a cada período fixo, como nas vendas a retalho, que aumentam no final do ano, ou nas temperaturas, que oscilam sazonalmente ao longo do ano. Modelos como o SARIMA (Seasonal ARIMA) e técnicas de decomposição sazonal são frequentemente utilizados para lidar com este tipo de não estacionariedade.

2.2.3 Transformação de processos não estacionários em estacionários

A transformação de processos não estacionários em estacionários é um passo fundamental na análise de séries temporais, especialmente para modelação e previsão. Para lidar com a não estacionariedade, é comum aplicar técnicas de diferenciação ou transformação.

2.2.3.1 Diferenciação

A diferenciação é uma técnica essencial na análise de séries temporais, utilizada para converter séries não estacionárias em estacionárias, eliminando tendências e padrões sazonais, condição primordial para a modelação estatística. Matematicamente, a diferenciação é representada pelo operador:

$$\nabla^d X_t = (1 - B)^d X_t, \tag{2.8}$$

onde B é o operador de defasagem, responsável por deslocar os valores da série no tempo, e d representa a ordem da diferenciação. O operador de defasagem é definido como $BX_t = X_{t-1}$, significando que a aplicação de B a uma variável resulta no seu valor no período anterior. Assim, a diferenciação de primeira ordem ($d = 1$) corresponde à remoção de uma tendência linear, pois calcula a variação entre valores consecutivos da série:

$$\nabla X_t = Y_t - X_{t-1} = (1 - B)X_t. \tag{2.9}$$

Se a série apresentar uma tendência quadrática (crescimento acelerado), a diferenciação de segunda ordem ($d = 2$) é necessária para remover esse comportamento:

$$\nabla^2 X_t = (1 - B)^2 X_t = X_t - 2X_{t-1} + X_{t-2}. \tag{2.10}$$

Nos modelos ARIMA, a diferenciação desempenha um papel fundamental, garantindo a estacionariedade da série temporal antes da aplicação dos componentes autorregressivos (AR) e de médias móveis (MA). A escolha da ordem d da diferenciação é essencial para transformar a série de forma adequada, garantindo que as propriedades estatísticas necessárias para modelação e previsão sejam satisfeitas (Box et al., 2015).

A escolha da ordem de diferenciação adequada pode ser feita através de testes estatísticos, como o teste de Dickey-Fuller Aumentado (ADF) e o teste de Kwiatkowski Phillips Schmidt Shin (KPSS), que avaliam a presença de raiz unitária. Apesar da sua eficácia na remoção de tendências, a aplicação excessiva da diferenciação pode aumentar a variabilidade dos dados e prejudicar a capacidade preditiva dos modelos (Shumway & Stoffer, 2017). Assim, é fundamental encontrar um equilíbrio entre a necessidade de estabilizar a série e a preservação da sua estrutura original (Wei, 2006; Hyndman & Athanasopoulos, 2018)

2.2.3.2 Transformação de Box-Cox

A transformação de Box-Cox é um método utilizado para estabilizar a variância e tornar a distribuição dos dados mais próxima da normalidade, facilitando a aplicação de modelos estatísticos e melhorando a interpretação dos resultados (Pires, 2001).

A transformação de Box-Cox com parâmetro λ (real) para a série $\{X_t, t \in \mathbb{Z}\}$ é definida por:

$$T_\lambda(X_t) = \begin{cases} \frac{X_t^\lambda - 1}{\lambda}, & \text{se } \lambda \neq 0 \\ \ln(X_t), & \text{se } \lambda = 0 \end{cases} \tag{2.11}$$

onde Y_t assume valores positivos. Usualmente, considera-se λ no intervalo $[-1, 1]$.

É importante notar que a aplicação da transformação de Box-Cox apenas a séries temporais com valores positivos não representa, na prática, uma restrição significativa. Sempre que necessário, pode-se adicionar uma constante positiva a todos os valores observados da série, garantindo que a transformação possa ser aplicada sem comprometer a integridade dos dados (Pires, 2001).

2.2.4 Função de autocovariância e autocorrelação

A função de autocorrelação, a função de autocorrelação parcial e a função de autocovariância são fundamentais na análise de séries temporais, pois fornecem informações sobre a dependência entre as observações ao longo do tempo (Box et al., 2015; Shumway & Stoffer, 2017).

A função de autocovariância mede a dependência linear entre valores de uma série temporal separados por um determinado intervalo de tempo (defasagem). Para uma série temporal estacionária $\{X_t\}$, a autocovariância para uma defasagem k é definida como (Pires, 2001):

$$\gamma_k = Cov(X_t, X_{t+k}) = E[(X_t - \mu)(X_{t+k} - \mu)]$$

onde $\mu = E[X_t]$ é a média da série temporal. A autocovariância depende apenas da defasagem k e não do tempo t, característica essencial para séries estacionárias (Shumway & Stoffer, 2017).

A função de autocorrelação (ACF), ou correlação serial, é a normalização da autocovariância para facilitar a comparação entre diferentes defasagens. Define-se como:

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

onde γ_0 é a variância da série, ou seja, $\gamma_0 = Var[X_t]$. A função de autocorrelação assume valores entre -1 e 1 e permite identificar padrões de dependência na série, sendo amplamente utilizada na identificação de modelos ARIMA (Pires, 2001).

A função de autocorrelação parcial (PACF) mede a correlação entre X_t e X_{t+k}, eliminando o efeito dos valores intermediários. A PACF é obtida resolvendo um sistema de equações de Yule-Walker e pode ser calculada iterativamente usando o algoritmo de Durbin-Levinson. Para um modelo autoregressivo de ordem p AR(p)), os coeficientes da PACF são nulos para defasagens superiores a p (Shumway & Stoffer, 2017).

A importância destas funções na modelação de séries temporais é notória, pois permitem identificar padrões cíclicos e tendências na série. A ACF é utilizada para analisar a presença de padrões recorrentes, enquanto a PACF auxilia na escolha da ordem p de modelos autoregressivos. A função de autocovariância fornece uma medida direta da dependência linear entre valores defasados da série (Box et al., 2015).

Estas funções são amplamente aplicadas na identificação de modelos AR, MA, ARMA e ARIMA, sendo essenciais para a construção de modelos de previsão robustos (Pires, 2001; Shumway & Stoffer, 2017).

2.3 Métodos de análises de séries temporais

A análise de séries temporais envolve uma variedade de métodos estatísticos projetados para identificar padrões nos dados ao longo do tempo e fazer previsões. De acordo com Tsay (2010), esses métodos podem ser classificados em três categorias principais: métodos clássicos de decomposição, métodos baseados em modelos paramétricos, como os modelos autorregressivos e métodos de suavização (sendo que existem métodos de suavização não-paramétricos).

Um dos métodos mais amplamente utilizados é o modelo de médias móveis e autorregressivo integrado, que foi formalizado por Box et al. (2015). O ARIMA é particularmente útil para séries não estacionárias em tendências, permitindo a modelação da relação entre os valores passados e presentes de uma série.

Este modelo baseia-se na suposição de que a série pode ser descrita como uma combinação linear dos seus valores anteriores e de erros passados.

Outro método relevante é o de suavização exponencial, proposto inicialmente por Brown (1959), Holt (1957), e expandido por Winters (1960) para lidar com séries temporais que possuem sazonalidade. Este método atribui pesos decrescentes às observações passadas, o que o torna eficaz para capturar tanto a tendência quanto a sazonalidade em séries de dados complexas. A versão de Holt-Winters é amplamente utilizada para séries que apresentam comportamento sazonal, como observado em aplicações de previsão de procura e análise de vendas Gardner (1985).

Por outro lado, os métodos baseados em transformações espectrais, como a análise de Fourier, são frequentemente empregados para detetar e modelar ciclos e periodicidades em séries temporais. Estes métodos são particularmente úteis para séries que apresentam padrões cíclicos que não são diretamente observáveis, como fenómenos astronómicos ou oscilações económicas de longo prazo (Priestley,1981) .

Em termos de modelos mais recentes, Zhang (2003) discute a combinação de redes neurais artificiais com modelos ARIMA para melhorar a precisão das previsões. Essa abordagem híbrida é especialmente útil em situações em que os dados possuem componentes não lineares, que podem não ser capturados adequadamente pelos modelos tradicionais de séries temporais.

Outro método emergente é o uso de modelos de estados espaciais, que fornecem uma estrutura flexível para incorporar tanto a componente determinística quanto estocástica das séries temporais. De acordo com Durbin & Koopman (2012) esses modelos são amplamente aplicados em áreas como economia, finanças e controlo de processos industriais, uma vez que permitem a decomposição da série em componentes de tendência, sazonalidade e ciclo.

3

Modelos ARIMA

Os modelos para séries temporais podem ser classificados em duas classes e isso depende do número de parâmetros envolvidos, nomeadamente: os modelos paramétricos, onde o número de parâmetros é finito e os modelos não paramétricos, em que existe um número infinito de parâmetros (Morettin e Toloi, 2004).

Nos modelos paramétricos a análise é feita tendo em conta a variável tempo. Os modelos que fazem parte desta classe são: os modelos de regressão, os autoregressivos e de médias móveis, os autoregressivos integrados e os autoregressivos sazonais.

3.1 Introdução ao modelo ARIMA

O modelo ARIMA, que como já referimos anteriormente é o acrónimo de Autoregressive Integrated Moving Average, é um dos modelos mais populares e amplamente utilizados na análise e previsão de séries temporais. Desenvolvido por George E. P. Box e Gwilym M. Jenkins na década de 1970, o modelo ARIMA combina três componentes principais: autoregressivo (AR), integração (I) e média móvel (MA) (Box et al., 2015). Essa combinação permite que o modelo capture tanto a dependência temporal dos dados quanto as características de não estacionariedade frequentemente encontradas nas séries temporais do mundo real.

Uma das principais vantagens do modelo ARIMA é sua flexibilidade em modelar diferentes tipos de séries temporais. A parte autoregressiva do modelo considera os valores passados da série como preditores do

seu valor atual, enquanto a parte de média móvel incorpora os erros passados do modelo como influências no valor atual. O componente de integração é utilizado para transformar uma série não estacionária em estacionária através de diferenciação, um processo que elimina tendências e sazonalidades que podem distorcer a análise (Shumway & Stoffer, 2017).

A aplicabilidade do modelo ARIMA estende-se a diversas disciplinas, incluindo economia, finanças, meteorologia e ciências sociais, onde a previsão de valores futuros com base em dados históricos é crucial para a tomada de decisões. Os modelos ARIMA são amplamente utilizados para modelar séries temporais, especialmente quando os dados exibem padrões lineares e apresentam dependência temporal bem definida. No entanto, estes modelos possuem limitações na captação de estruturas não lineares, uma vez que são construídos com base em relações lineares entre os valores passados da série e os erros do modelo.

Em séries temporais que apresentam padrões não lineares complexos, como heterocedasticidade, mudanças abruptas ou efeitos de limiar, os modelos ARIMA podem não ser a melhor escolha. Nestes casos, abordagens alternativas, como modelos GARCH (para volatilidade variável), redes neuronais e modelos híbridos que combinam ARIMA com métodos não lineares, podem oferecer um melhor desempenho na modelação e previsão.

Assim, ao contrário do que é frequentemente assumido, os modelos ARIMA não são os mais eficazes para capturar dinâmicas não lineares complexas (Hyndman & Athanasopoulos, 2018). A sua eficácia está mais associada a séries temporais que seguem um comportamento linear ou podem ser linearizadas através de transformações apropriadas.

A construção de um modelo ARIMA envolve um processo sistemático que inclui a identificação, estimativa e verificação do modelo. Essa abordagem sistemática é essencial para garantir que o modelo seja adequado para os dados em questão, minimizando o erro de previsão. A identificação do modelo é frequentemente realizada através da análise dos gráficos da ACF e da PACF, que ajudam a determinar os parâmetros apropriados p, d e q (Brockwell & Davis, 2002).

3.2 Identificação dos modelos ARIMA

A identificação dos modelos ARIMA é um passo crucial na análise de séries temporais, pois envolve a determinação dos parâmetros p, d e q do modelo. O processo de identificação é fundamental para garantir que o modelo se ajuste adequadamente aos dados e seja capaz de realizar previsões precisas (Hyndman & Athanasopoulos, 2018). Os três parâmetros têm funções distintas:

- **Parâmetro** p: Representa a ordem do componente autorregressivo. O valor de p indica quantas observações passadas são utilizadas para prever o valor atual da série. Para determinar o valor de p, utilizam-se as ACF e PACF, que ajudam a identificar a dependência entre as observações (Box et al., 2015).

- **Parâmetro** d: Indica o número de diferenciações necessárias para tornar a série estacionária. A estacionariedade é uma propriedade essencial para a aplicação de modelos ARIMA, pois séries estacionárias têm médias e variâncias constantes ao longo do tempo. Para identificar o valor de d, os analistas costumam aplicar testes de estacionariedade, como o teste de Dickey-Fuller aumentado (ADF) (Dickey & Fuller, 1981).

- **Parâmetro** q: Refere-se à ordem do componente de média móvel. Assim como o parâmetro p, o valor de q é determinado a partir das funções de autocorrelação, que mostram a influência de erros passados na previsão do valor atual (Shumway & Stoffer, 2017).

O processo de identificação pode ser resumido em algumas etapas fundamentais. Primeiro, os dados devem ser representados graficamente para visualizar padrões, tendências e/ou sazonalidades. Em seguida, realiza-se a diferenciação dos dados para alcançar a estacionariedade. Depois, as funções ACF e PACF são analisadas para ajudar na escolha dos valores de p e q. Finalmente, o modelo ARIMA é estimado e validado, utilizando critérios de informação, como o Critério de Informação de Akaike (AIC) ou o Critério de Informação Bayesiano (BIC), para avaliar o ajuste do modelo (Akaike, 1974; Schwarz, 1978).

Como se pode verificar, um bom modelo ARIMA deve equilibrar a complexidade e a capacidade preditiva, evitando o sobreajuste. A identificação eficaz é, portanto, fundamental para garantir que o modelo final seja robusto e confiável (Lutkepohl, 2005). Em resumo, a identificação dos modelos ARIMA requer uma combinação de análise gráfica, testes estatísticos e validação criteriosa para se obter um modelo que se adapte bem aos dados e seja capaz de fornecer previsões úteis.

3.3 Formulação do modelo ARIMA

3.3.1 Modelos AutoRegressivos

O modelo autorregressivo é fundamental na modelação de séries temporais, sendo utilizado para descrever processos onde os valores futuros de uma variável dependem linearmente dos seus valores passados (Cryer & Chan, 2008; Cowpertwait & Metcalfe, 2009).

Um modelo autorregressivo de ordem p, denotado por AR(p), é definido pela seguinte equação:

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t, \quad t \in \mathbb{Z} \tag{3.1}$$

onde:

- X_t representa a variável no instante t;
- μ é a constante, representando o nível médio da série temporal;
- $\phi_1, \phi_2, \ldots, \phi_p$ são os coeficientes do modelo, onde $\phi_p \neq 0$.
- ε_t é um ruído branco de média nula.

Para garantir a estacionariedade do processo, analisamos o polinômio característico associado:

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$

onde B representa o operador de defasagem, definido como $BX_t = X_{t-1}$. A condição necessária e suficiente para a estacionariedade do processo $AR(p)$ é que todas as raízes da equação característica $\Phi(B) = 0$ estejam fora do círculo unitário, ou seja, satisfaçam:

$$|B| > 1$$

Se alguma raiz estiver dentro ou sobre o círculo unitário ($|B| \leq 1$), o processo não será estacionário.

Além disso, sob essa condição, a série temporal AR(p) pode ser representada como uma expansão infinita de (MA(∞)), onde a soma dos coeficientes da representação MA infinita é absolutamente somável. Isso

significa que o modelo AR(p) pode ser escrito como uma combinação linear infinita de termos de erro ε_t, garantindo a estabilidade do processo.

Os modelos AR apresentam características estatísticas bem definidas quanto à média, variância, ACF e PACF, conforme destacam Box et al. (2015) e Enders (2014) em suas análise sobre séries temporais.

Para um processo AR(p) estacionário, Hamilton (1994) demonstra que a média μ pode ser expressa como:

$$\mu = \frac{\phi_0}{1 - \sum_{i=1}^{p} \phi_i} \tag{3.2}$$

onde ϕ_0 representa o termo constante e $\phi_1, \phi_2, \ldots, \phi_p$ são os coeficientes autorregressivos (Lutkepohl, 2005).

A variância γ_0 de um processo AR(p) pode ser obtida através do sistema de equações de Yule-Walker, conforme detalhado por (Shumway & Stoffer, 2017).:

$$\gamma_0 = \sigma_\epsilon^2 + \sum_{i=1}^{p} \phi_i \gamma_i \tag{3.3}$$

onde σ_ϵ^2 denota a variância do termo de erro (Brockwell & Davis, 2002).

Em relação a ACF, Cryer & Chan (2008), destacam que a ACF de um processo AR(p) apresenta um decaimento exponencial ou oscilatório, seguindo a relação:

$$\rho_k = \sum_{i=1}^{p} \phi_i \rho_{k-i}, \quad \text{para } k > 0 \tag{3.4}$$

Este comportamento é fundamental para identificação do processo (Tsay, 2010).

A PACF de um processo AR(p) apresenta uma característica distintiva: valores não nulos até a defasagem p e zero para defasagens superiores. Formalmente, como demonstrado por Box et al. (2015), a PACFϕ_{kk} é dada por:

$$\phi_{kk} = \begin{cases} \phi_k & \text{para } k \leq p \\ 0 & \text{para } k > p \end{cases} \tag{3.5}$$

Esta propriedade fundamental, destacada por Brockwell & Davis (2002), permite identificar a ordem p do processo autorregressivo através do exame da última defasagem significativa na PACF amostral (Tsay, 2010).

3.3.2 Modelos de médias móveis

Os modelos de médias móveis (MA) é uma das abordagens fundamentais na análise de séries temporais, utilizado para descrever e prever dados que apresentam comportamento estocástico. Os modelos MA são caracterizados pela inclusão de termos de erro passados na formulação do modelo, permitindo que a média da série em um determinado instante de tempo seja influenciada pelos erros passados de previsões (Morettin

e Toloi, 2004; Cryer & Chan, 2008).

Formalmente, um modelo MA de ordem q, denotado como MA(q), é representado pela seguinte equação:

$$X_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ldots + \theta_q\varepsilon_{t-q}, \quad t \in \mathbb{Z} \tag{3.6}$$

onde:

- y_t é o valor da série temporal no tempo t,
- μ é a constante,
- ε_t é o ruído branco de média nula,
- θ_i são os coeficientes do modelo que indicam a influência dos erros passados,onde $i = 1, \ldots, q$ $(\theta_q \neq 0)$

De acordo com Box et al. (2015), os modelos MA são particularmente eficazes para captar a dinâmica de séries temporais em que os choques temporários afetam os valores futuros. Esses modelos são geralmente utilizados em conjunto com outros componentes, como os modelos AR, para formar o modelo ARMA, que combina os elementos de ambos os modelos.

Note-se que toda a série $AR(p)$ estacionária é uma série de médias móveis $MA(\infty)$ e que, reciprocamente, toda a série $MA(q)$ invertível é uma série autoregressiva $AR(\infty)$ (Pires, 2001).

Portanto, um modelo MA(q) é sempre estacionário, independentemente dos valores dos coeficientes $\theta_1, \ldots, \theta_q$ (Pires, 2021). Isso ocorre porque um modelo MA(q) é uma combinação linear finita de termos de erro ε_t, que são ruído branco com média e variância constantes. Dessa forma, a média, variância e autocovariância do processo não dependem do tempo, garantindo a estacionariedade de segunda ordem.

Além disso, para um modelo MA(q) ser considerado invertível, é necessário que seja possivel ser reescrito como uma expansão infinita de um processo AR(∞), ou seja:

$$\varepsilon_t = X_t - \theta_1 X_{t-1} - \theta_2 X_{t-2} - \ldots \tag{3.7}$$

Para garantir esta unicidade, as raízes do polinómio característico associado ao modelo devem satisfazer a seguinte condição:

$$\Theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q = 0. \tag{3.8}$$

A condição de invertibilidade exige que todas as raízes da equação característica estejam fora do círculo unitário (Morettin e Toloi, 2004), ou seja:

$$|B| > 1, \quad \forall B \text{ tal que } \Theta(B) = 0. \tag{3.9}$$

Se esta condição não for satisfeita, o modelo MA(q) pode ser representado de múltiplas formas, dificultando a interpretação estatística e a estimação dos parâmetros. A invertibilidade assegura que o modelo MA(q) possa ser descrito de forma única, sem ambiguidade.

Por outro lado, Hyndman & Athanasopoulos (2018), a autocorrelação pode distorcer as previsões se não for adequadamente modelada.

Portanto, ACF é uma ferramenta essencial na identificação e modelação de processos de séries temporais. Para um MA(q), a ACF apresenta um comportamento característico, sendo não nula até a defasagem q e anulando-se para zero para defasagens maiores que q (Montgomery et al, 2015).

A variância do processo MA(q) é dada por:

$$\gamma_X(0) = \text{Var}(X_t) = \sigma^2 \left(1 + \theta_1^2 + \cdots + \theta_q^2\right). \tag{3.10}$$

A **função de autocovariância** para uma defasagem k ($k \leq q$) pode ser obtida através da expressão:

$$\gamma_X(k) = \sigma^2 \left(-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q\right), \quad k = 1, 2, \ldots, q. \tag{3.11}$$

Para defasagens maiores que q, a **autocovariância é exatamente zero**:

$$\gamma_X(k) = 0, \quad k > q. \tag{3.12}$$

A partir dessas expressões, a ACF do processo MA(q) é dada por:

$$\rho_X(k) = \frac{\gamma_X(k)}{\gamma_X(0)} = \begin{cases} \frac{-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \cdots + \theta_q^2}, & \text{se } k = 1, 2, \ldots, q, \\ 0, & \text{se } k > q. \end{cases} \tag{3.13}$$

Esse comportamento da ACF é extremamente útil na **identificação de processos MA(q)**, pois permite determinar a ordem q do modelo através da análise da ACF empírica da série temporal.

Na prática, ao estimarmos a ACF amostral $r(k)$, esta não será exatamente zero após a defasagem q, devido à variabilidade amostral. No entanto, espera-se que os valores da ACF para $k > q$ sejam muito pequenos em termos absolutos. Para verificar se os valores são estatisticamente significativos, utilizamos limites de confiança baseados no tamanho da amostra N, sendo uma aproximação comum:

$$\pm \frac{1.96}{\sqrt{N}}. \tag{3.14}$$

Caso os valores da ACF para $k > q$ caiam dentro desses limites, considera-se que podem ser desprezados, confirmando que um modelo MA(q) pode ser apropriado para os dados.

No entanto, a principal limitação dos modelos de média móvel é que eles dependem de um número finito de termos de erro passados. Isso significa que, em séries com longas dependências, um modelo MA pode não ser suficiente. Nesse caso, os autores sugerem a utilização de modelos ARMA, que permitem uma modelação mais flexível e abrangente das séries temporais (Box et al., 2015).

Em termos de aplicação prática, a seleção da ordem do modelo MA, ou seja, o valor de q, é uma etapa crucial. Métodos como o critério de informação de Akaike (AIC) e o critério de informação Bayesiano (BIC) são frequentemente utilizados para determinar a ordem mais apropriada para o modelo (Hyndman & Athanasopoulos, 2018).

3.3.3 Modelos autoregressivos e de médias móveis

Os modelos autorregressivos e de médias móveis são amplamente utilizados na análise de séries temporais, pois combinam as propriedades dos AR e dos MA para captar a dinâmica temporal de uma série. Um modelo ARMA de ordem (p, q) é definido como uma combinação linear de p termos autorregressivos e q termos de média móvel, permitindo uma modelação mais flexível e robusta (Pires, 2001; Morettin e Toloi, 2004).

Matematicamente, o modelo ARMA pode ser expresso pela seguinte equação:

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t, \quad t \in \mathbb{Z}, \tag{3.15}$$

onde:

- X_t representa a variável da série temporal no instante t;
- μ é a constante do modelo, que reflete o nível médio de oscilação da série quando esta é estacionária;
- $\phi_1, \phi_2, \ldots, \phi_p$ são os coeficientes do componente autoregressivo (AR), que refletem a influência de valores passados da própria série;
- $\theta_1, \theta_2, \ldots, \theta_q$ são os coeficientes do componente de MA, que captam o impacto de choques aleatórios anteriores;
- ε_t é um ruído branco de média nula;

Uma forma alternativa de representar este modelo é através do operador de defasagem B, que já foi definido anteriormente como $BX_t = X_{t-1}$, o que permite reescrever o modelo como:

$$\Phi(B) X_t = \mu + \Theta(B) \varepsilon_t, \tag{3.16}$$

onde:

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p, \tag{3.17}$$

$$\Theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q. \tag{3.18}$$

Assim sendo, para que o modelo seja estacionário, é necessário que as raízes do polinómio $\Phi(B)$ estejam fora do círculo unitário, ou seja:

$$|B| > 1, \quad \forall B \text{ tal que } \Phi(B) = 0. \tag{3.19}$$

Por outro lado, a condição de invertibilidade garante que o modelo MA(q) pode ser reescrito de forma única como um modelo AR(∞). Para isso, as raízes do polinómio $\Theta(B)$ também devem estar fora do círculo unitário:

$$|B| > 1, \quad \forall B \text{ tal que } \Theta(B) = 0. \tag{3.20}$$

Em todo o caso, para garantir uma modelação adequada, é essencial que as condições de estacionariedade e invertibilidade sejam verificadas, de modo a assegurar a estabilidade do modelo e a unicidade da representação do modelo.

De acordo Tsay (2010), a identificação da ordem do modelo p e q pode ser realizada utilizando gráficos de ACF e autocorrelação parcial PACF, o que ajuda a determinar quantos termos anteriores devem ser incluídos no modelo.

A ACF de um modelo ARMA(p, q) satisfaz a seguinte equação de recorrência:

$$\rho(k) - \phi_1\rho(k-1) - \cdots - \phi_p\rho(k-p) = 0, \quad k \geq q+1. \tag{3.21}$$

Para um modelo ARMA(p, q) estacionário, a sequência de valores da ACF apresenta as seguintes propriedades (Morettin e Toloi, 2004):

1. A sequência das funções de autocorrelação de um processo ARMA(p, q) tende para zero de forma exponencial e/ou sinusoidal, à medida que o valor de k aumenta.
2. Para defasagens $k > q$, a ACF satisfaz uma equação de diferença de ordem p, garantindo que os seus valores são combinações lineares de valores anteriores.

A PACF de um modelo ARMA(p, q) apresenta a seguinte característica:

1. A sequência de valores da PACF tende para zero de forma exponencial e/ou sinusoidal, à medida que k aumenta.
2. Ao contrário da ACF, que pode ser expressa como uma combinação linear dos seus valores passados, a PACF mede a correlação direta entre X_t e X_{t-k}, eliminando o efeito das defasagens intermédias.

Portanto, a análise conjunta da ACF e da PACF permite distinguir modelos AR(p), MA(q) e ARMA(p, q).

3.3.4 Modelos autoregressivos integrados e de médias móveis

Os modelos ARIMA são uma extensão dos modelos ARMA que incorporam a diferenciação para tornar uma série temporal estacionária. Um modelo ARIMA é especificado por três parâmetros: p, d, e q, onde p representa a ordem do componente autorregressivo, d é o número de diferenciações necessárias para alcançar a estacionaridade, e q indica a ordem do componente de média móvel. Essa estrutura permite uma modelação mais eficaz de séries temporais que exibem tendências ou padrões sazonais (Box et al., 2015).

Matematicamente, um modelo ARIMA pode ser representado como:

$$\Delta^d X_t = \mu + \phi_1\Delta^d X_{t-1} + \phi_2\Delta^d X_{t-2} + \cdots + \phi_p\Delta^d X_{t-p} + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q} + \varepsilon_t, \quad t \in \mathbb{Z} \tag{3.22}$$

onde:

- Δ^d representa a operação de diferenciação de ordem d,

- ϕ_i são os coeficientes do modelo autorregressivo,
- θ_j são os coeficientes do modelo de média móvel,
- ε_t é um ruído branco de média nula.

Conforme visto, os modelos ARIMA são uma extensão dos modelos ARMA que incorporam a diferenciação para tornar uma série temporal estacionária. Um modelo ARIMA é especificado por três parâmetros: p, d, e q, onde p representa a ordem do componente autorregressivo, d é o número de diferenciações necessárias para alcançar a estacionaridade, e q indica a ordem do componente de média móvel (Lutkepohl, 2005; Enders, 2014). A escolha dos parâmetros p, d, e q pode ser realizada através de análises gráficas, como os gráficos de ACF e PACF, que ajudam na identificação das ordens adequadas dos componentes do modelo (Shumway & Stoffer, 2017).

Adicionalmente, a modelação ARIMA continua a evoluir, com a introdução de variantes que incluem componentes sazonais, como o modelo SARIMA, que permite uma modelação ainda mais precisa de séries temporais que apresentam padrões sazonais (Pankratz, 1991).

3.4 Estimação dos modelos ARIMA

A estimação dos modelos ARIMA é uma etapa crucial no processo de modelação de séries temporais. Para além da escolha dos métodos de estimação, esta etapa envolve a escolha prévia dos parâmetros de modelo. A escolha adequada destes valores, e, por conseguinte, dos parâmetros a eles associados, é fundamental para garantir a precisão das previsões (Box et al., 2015).

A primeira etapa para estimar um modelo ARIMA é a identificação dos valores de p, q e d. A diferenciação, representada por d , é aplicada aos dados até que a série se torne estacionária. Isso pode ser avaliado através de testes de estacionariedade, como o teste de Dickey-Fuller aumentado (ADF), que ajuda a determinar se a série apresenta raízes unitárias (Dickey & Fuller, 1981).

Uma vez que a série esteja estacionária, os valores de p e q podem ser identificados por meio da análise dos gráficos ACF e PACF. O gráfico ACF é utilizado para identificar a ordem q do modelo de média móvel, enquanto o gráfico PACF é usado para determinar a ordem p do modelo autoregressivo. A combinação dessas análises fornece uma base sólida para a especificação inicial do modelo (Hyndman & Athanasopoulos, 2018).

Após a identificação dos parâmetros, a estimação dos coeficientes Φ (autoregressivos) e Θ (médias móveis) é frequentemente realizada através do Método da Máxima Verosimilhança (MMV) ou do Método dos Mínimos Quadrados (MMQ), que serão detalhados a seguir. O MMV, em particular, é amplamente utilizado devido à sua capacidade de fornecer estimativas eficientes, sobretudo em amostras de grande dimensão.

Os vetores de coeficientes são definidos como:

$$\Phi = (\phi_1, \phi_2, \ldots, \phi_p)$$

$$\Theta = (\theta_1, \theta_2, \ldots, \theta_q)$$

onde Φ e Θ representam os vetores dos coeficientes AR e MA com os parâmetros p e q, respetivamente.

A avaliação do modelo ajustado é a última etapa para estimação dos modelos ARIMA. Isso envolve a verificação da adequação do modelo por meio da análise dos resíduos. Os resíduos devem ser ruído branco, de média nula e variância contante ou seja, não devem apresentar autocorrelações significativas. O gráfico da função de autocorrelação e testes de hipóteses, como o teste de Ljung-Box, permitem verificar a presença de autocorrelação nos resíduos (Shumway & Stoffer, 2017).

3.4.1 Métodos de estimação

A estimação dos parâmetros em modelos ARIMA é uma etapa crítica na modelação de séries temporais, pois os valores estimados influenciam diretamente a precisão das previsões. Os métodos de estimação mais comuns para os modelos ARIMA incluem o MMV, o MMQ e o método de Yule-Walker ou dos momentos. Cada um desses métodos possui características distintas que os tornam adequados para diferentes contextos e requisitos dos dados.

3.4.1.1 Método de máxima verossimilhança

O MMV é um dos principais métodos para a estimação dos parâmetros dos modelos ARMA(p, q). Este método baseia-se na ideia de encontrar os valores dos parâmetros que maximizam a função de verosimilhança, ou seja, que tornam os dados observados mais prováveis dado o modelo ajustado (Morettin & Toloi, 2004).

Seja Θ o vetor dos parâmetros desconhecidos do modelo ARMA, definido como:

$$\Theta = (\Phi, \Theta, \sigma^2) = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \sigma^2)$$

onde σ^2 é a variância do termo de erro ε_t.

Dada uma amostra de tamanho T, representada pelo vetor de observações:

$$X = (x_1, x_2, \ldots, x_T),$$

a função de verosimilhança é dada pela função densidade conjunta dos dados, tratada como uma função dos parâmetros Θ:

$$L(\Theta \mid X) = f(x_T, x_{T-1}, \ldots, x_1; \Theta).$$

O estimador de máxima verosimilhança (EMV) é obtido resolvendo:

$$\hat{\Theta}_{MV} = \arg \max_{\Theta \in \mathbb{R}^{p+q+1}} L(\Theta \mid X).$$

Como a função $L(\Theta \mid X)$ pode ser de difícil manipulação, é comum trabalhar com o logaritmo da verosimilhança, definido por:

$$\log L(\Theta \mid X) = l(\Theta \mid X).$$

Uma vez que o logaritmo é uma transformação monotónica, maximizar $L(\Theta \mid X)$ é equivalente a maximizar $l(\Theta \mid X)$:

$$\hat{\Theta}_{MV} = \arg \max_{\Theta \in \mathbb{R}^{p+q+1}} l(\Theta \mid X).$$

De modo que, para encontrar os valores ótimos dos parâmetros, resolvemos a equação da verosimilhança, que é dada por:

$$\frac{\delta l(\Theta \mid X)}{\delta \Theta} = 0.$$

No entanto, o vetor de estimadores $\hat{\Theta}_{MV}$ será a solução desta equação.

Conforme visto por Brockwell & Davis (2002), a maximização da função de verossimilhança pode ser realizada utilizando algoritmos de otimização, como o método de Newton-Raphson ou o método dos scores de Fisher. A vantagem desse método é que ele fornece estimativas que são assintoticamente não tendenciosas e eficientes, além de permitir a construção de intervalos de confiança para os parâmetros estimados (Hyndman & Athanasopoulos, 2018).

3.4.1.2 Método dos mínimos quadrados

O MMQ é outro método popular para a estimação de parâmetros em modelos ARIMA. Nesse método, os parâmetros são estimados minimizando a soma dos quadrados dos resíduos, ou seja, a diferença entre os valores observados e os valores previstos pelo modelo. A função de soma dos quadrados dos resíduos é expressa como:

$$S(\theta) = \sum_{t=1}^{n} (X_t - \hat{X}_t(\theta))^2$$

onde X_t são os valores observados e $\hat{X}_t(\theta)$ são os valores estimados pelo modelo. Embora o método dos mínimos quadrados seja mais simples de implementar, ele pode não ser tão eficiente quanto o método de máxima verossimilhança, especialmente em amostras pequenas ou em modelos com características complexas (Box et al., 2015).

3.4.1.3 Método de Yule-Walker

As equações de Yule-Walke são um conjunto de equações fundamentais usualmente referenciadas para a estimação dos parâmetros em modelos autoregressivos AR(p). Estas equações estabelecem uma relação entre os coeficientes autoregressivos Φ, a função de autocovariância e a função de autocorrelação do processo.

As equações de Yule-Walker relacionam a função de autocovariância $\gamma(h)$ do processo com os coeficientes autoregressivos Φ:

$$\gamma(h) = \sum_{j=1}^{p} \phi_j \gamma(h-j), \quad h = 1, 2, \ldots, p. \tag{3.23}$$

Escrevendo na forma matricial:

$$\begin{bmatrix} \gamma(0) & \gamma(1) & \ldots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \ldots & \gamma(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(p-1) & \gamma(p-2) & \ldots & \gamma(0) \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(p) \end{bmatrix}.$$

Esta matriz simétrica permite estimar os coeficientes ϕ_j através da inversão do sistema linear.

Assim, a solução das equações de Yule-Walker fornece estimadores consistentes para os coeficientes Φ. Além disso, a variância do erro ε_t pode ser estimada por:

$$\sigma^2 = \gamma(0) - \sum_{j=1}^{p} \phi_j \gamma(j). \tag{3.24}$$

De maneira geral, o método é particularmente útil para modelos AR(p) estacionários, pois permite obter estimações eficientes dos coeficientes autoregressivos.

3.4.2 Seleção do modelo

Tal como já referimos, a seleção do modelo ARIMA é uma etapa crucial na análise de séries temporais, pois a escolha adequada do modelo influencia diretamente a precisão das previsões e a compreensão da dinâmica subjacente dos dados. Um dos principais critérios utilizados na seleção de modelos é o AIC, que penaliza a complexidade do modelo em função da qualidade do ajuste. O AIC é definido pela fórmula:

$$AIC = -2\ln(L) + 2k \tag{3.25}$$

onde L é a verossimilhança do modelo e k é o número de parâmetros estimados. De acordo Akaike (1974), um modelo que minimiza o AIC é preferido, pois indica um bom equilíbrio entre a quantidade de parâmetros estimados e a verossimilhança do ajuste. Outro critério amplamente utilizado é o Critério de Informação Bayesiano (BIC), que aplica uma penalização ainda maior para modelos mais complexos e é definido como:

$$BIC = -2\ln(L) + k\ln(n) \tag{3.26}$$

onde n é o tamanho da amostra. Esses critérios permitem a comparação entre diferentes especificações de modelos, auxiliando na identificação da configuração mais apropriada.

3.5 Diagnóstico dos modelos ARIMA

O diagnóstico dos modelos ARIMA é outra etapa fundamental no processo de análise de séries temporais, pois permite avaliar a adequação do modelo ajustado e a qualidade das previsões. Segundo Hyndman

& Athanasopoulos (2018), após a estimação do modelo, é necessário verificar se os resíduos do modelo se comportam como um processo estocástico adequado (ruído branco de média nula, possivelmente, mas não obrigatoriamente, normal). Um dos principais objetivos é assegurar que os resíduos não apresentem autocorrelação significativa, o que indicaria que o modelo não capturou toda a estrutura temporal dos dados.

É importante realçar que uma ferramenta comum para realizar esse diagnóstico é o teste de Ljung-Box. Este teste avalia se os resíduos são não correlacionados até uma defasagem m (hipótese nula). Conforme Box et al.(2015), o teste é formulado como:

$$Q = n(n+2)\sum_{k=1}^{m}\frac{\hat{\rho}_k^2}{n-k} \tag{3.27}$$

onde n é o número de observações, m é o número de defasagens consideradas e $\hat{\rho}_k$ é a autocorrelação estimada dos resíduos em k defasagens. Um valor de Q que excede a distribuição o quantil $(1-\alpha)$ da distribuição qui-quadrado χ^2 com m graus de liberdade, indica a rejeição da hipótese nula, sugerindo que existe autocorrelação nos resíduos e que, portanto, outro modelo deverá ser testado (Hyndman & Athanasopoulos, 2018).

Além disso, a análise gráfica dos resíduos, como a ACF e a PACF, pode ser utilizada para validar a hipótese de não correlação dos resíduos. Para a normalidade temos o histograma e/ou testes de hipótese (por exemplo, Shapiro-Wilk ou Kolmogorov-Smirnov com correção de Lilliefors) (Shumway & Stoffer, 2017). Para além destes, também há o teste t-Student para média nula – verifica a hipótese $H_0 : \mu_{\text{resíduos}} = 0$ contra $H_1 : \mu_{\text{resíduos}} \neq 0$. A não rejeição de H_0 sugere que o modelo capturou adequadamente a componente sistemática da série (Ender, 2014). O testes de estacionariedade KPSS é aplicado aos resíduos para avaliar se estes constituem um processo estacionário. Quando o teste não rejeita a hipótese nula de estacionariedade, podemos inferir que: a média dos resíduos permanece constante ao longo do tempo, a variância apresenta homocedasticidade e a estrutura de autocorrelação é estável.

3.6 Previsão dos modelos ARIMA

A previsão utilizando modelos ARIMA é uma das aplicações mais importantes na análise de séries temporais, permitindo antecipar comportamentos futuros com base em dados passados. A capacidade de modelar dados complexos, que podem exibir tendências e sazonalidades, torna os ARIMA (ou SARIMA) uma escolha popular em diversas áreas, como economia, finanças, saúde e ciências sociais (Hyndman & Athanasopoulos, 2018).

Para realizar previsões com modelos ARIMA, primeiramente é necessário ajustar o modelo aos dados históricos. Uma vez que o modelo está ajustado, as previsões podem ser feitas utilizando a equação do modelo.

A previsão em modelos ARIMA(p, d, q) baseia-se na extrapolação da série temporal a partir das equações do modelo, assumindo que os padrões históricos persistem no futuro. O processo de previsão é definido iterativamente e pode ser dividido em previsão de curto prazo ($h = 1$) e previsão de longo prazo ($h > 1$).

Para prever um período à frente, previsão de curto prazo, utiliza-se a equação do modelo ARIMA, onde os valores passados da série e os erros anteriores são combinados para estimar X_{T+1}:

$$\hat{X}_{T+1|T} = \phi_1 X_T + \phi_2 X_{T-1} + \cdots + \phi_p X_{T-p+1} + \theta_1 \varepsilon_T + \theta_2 \varepsilon_{T-1} + \cdots + \theta_q \varepsilon_{T-q+1}. \tag{3.28}$$

onde:

- $\hat{X}_{T+1|T}$ é a previsão de Y no instante $T+1$;
- ϕ_i e θ_j são os coeficientes autoregressivos e de médias móveis, respetivamente;
- Assume-se que os erros futuros são zero, reduzindo a incerteza no horizonte imediato.

Quando se fala em prever além de um período, previsão de longo prazo, os valores previstos substituem as observações reais na equação do modelo:

$$\hat{X}_{T+h|T} = \phi_1 \hat{X}_{T+h-1|T} + \phi_2 \hat{X}_{T+h-2|T} + \cdots + \phi_p \hat{X}_{T+h-p|T}. \tag{3.29}$$

À medida que h aumenta:

- As previsões tornam-se menos precisas devido à acumulação de incerteza;
- Modelos com componentes MA incluem a influência de erros passados, dificultando a previsão para horizontes longos.

Por outro lado, a incerteza da previsão associada ao modelo é quantificada através dos intervalos de previsão. Para um nível de confiança de 95%, a previsão é dada por:

$$\hat{X}_{T+h|T} \pm 1.96\sqrt{v_{T+h|T}}. \tag{3.30}$$

A variância da previsão, $v_{T+h|T}$, cresce com h, sendo dada por:

$$v_{T+h|T} = \sigma^2 \left(1 + \sum_{j=1}^{h} \psi_j^2\right), \tag{3.31}$$

onde σ^2 é a variância do erro e ψ_j representa os coeficientes da expansão infinita do modelo ARIMA. A expansão infinita do modelo ARIMA É uma maneira alternativa de representar o mesmo modelo, mas que permite perceber como os efeitos dos choques passados se propagam ao longo do tempo.

Pode-se então constatar que, o processo de previsão em modelos ARIMA permite estimar valores futuros com base na estrutura temporal dos dados. No entanto, a fiabilidade das previsões depende da estabilidade do padrão da série e da correta especificação do modelo. Além disso, os intervalos de previsão aumentam com h, refletindo a incerteza crescente em previsões de longo prazo (Shumway & Stoffer, 2017; Hyndman & Athanasopoulos, 2018).

Além disso, as previsões realizadas a partir de modelos ARIMA são frequentemente avaliadas em termos de precisão, utilizando métricas como o Erro Quadrático Médio (MSE), Erro Absoluto Médio (MAE) e Raiz do Erro Quadrático Médio (RMSE):

O *MSE* mede a média dos erros elevados ao quadrado, penalizando desproporcionalmente desvios maiores e sendo sensível a outliers:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(X_t - \hat{X}_t)^2 \tag{3.32}$$

onde X_t representa o valor real da série temporal no instante t, $\hat{X}_t$ é o valor previsto pelo modelo e n é o número total de observações.

O *MAE* é uma métrica mais robusta que o MSE em relação a valores extremos e representa a média dos erros absolutos:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|X_t - \hat{X}_t| \tag{3.33}$$

O MAE fornece uma indicação clara da grandeza do erro médio esperado, sendo útil para avaliar a precisão geral do modelo sem amplificar discrepâncias extremas.

Por último, a *RMSE* expressa a precisão do modelo na mesma unidade da variável analisada, tornando-se mais intuitiva para interpretação:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(X_t - \hat{X}_t)^2} \tag{3.34}$$

O RMSE mantém a penalização para grandes erros, como o MSE, mas apresenta o erro na mesma escala da variável predita.

A comparação das previsões com os dados reais ajuda a determinar a eficácia do modelo e a ajustar a modelação conforme necessário (Pires, 2001).

A escolha entre estas métricas depende do contexto da aplicação. O MSE e o RMSE são mais úteis quando grandes erros precisam ser penalizados, enquanto o MAE é uma alternativa mais equilibrada para cenários onde a distribuição dos erros não apresenta desvios significativos extremos (Pires, 2001). Além disso, a escolha do modelo ARIMA mais apropriado pode ser feita com base em critérios de informação, como o AIC ou o BIC, que penalizam a complexidade do modelo em relação à qualidade do ajuste.

Em séries temporais, essas métricas são essenciais para validar a adequação dos modelos ARIMA e apoiar decisões sobre ajustes e refinamentos nos parâmetros do modelo. Assim, a minimização dessas métricas, em conjunto com critérios de informação como o AIC e BIC, permite selecionar o modelo mais eficiente, garantindo previsões mais precisas e confiáveis.

3.7 Mudanças de estrutura em séries temporais

As séries temporais podem ser afetadas por mudanças estruturais, ou seja, por alterações significativas nos padrões da própria série ao longo do tempo. Estas mudanças refletem transformações nas condições subjacentes ao fenómeno em análise e podem modificar a relação entre a variável estudada e o tempo, comprometendo a estabilidade estatística da série (Palma, 2016). Conforme visto por Muller, compreende-se por mudança de estrutura da série temporal, uma alteração do comportamento das relações existentes entre as observações que levam a que a série possa ser descrita por um modelo diferente a partir do momento em que essa alteração ocorre. Estas mudanças podem resultar de eventos económicos, mudanças de política, crises financeiras ou fatores externos como avanços tecnológicos e desastres naturais.

Quando ocorrem mudanças estruturais, os modelos estatísticos tradicionais podem tornar-se inadequados para a previsão, exigindo abordagens que incorporem essas alterações, ou seja os pontos de quebra estruturais são fundamentais para identificar quando uma variável ou tendência sofre uma mudança que não pode ser capturada por um único modelo. Quando quebras estruturais são identificadas, é necessário ajustar o modelo para se adaptar corretamente aos dados, muitas vezes dividindo-os em segmentos e aplicando

modelos diferentes a cada um deles (Perron, 2006).

Diferentes métodos foram desenvolvidos para identificar e modelar essas mudanças. Os modelos com quebras estruturais consideram que há momentos específicos em que a estrutura da série se altera, enquanto modelos de mudança gradual assumem uma transição progressiva entre regimes distintos.

3.7.1 Tipos de mudanças estruturais

Existem diferentes tipos de mudanças estruturais que podem ocorrer em uma série temporal (Bai & Perron, 1998). Os principais incluem:

- **Mudanças de nível**: Alteração súbita na média da série.
- **Mudanças na tendência**: Modificação na inclinação da série ao longo do tempo.
- **Mudanças na variância**: Alterações na amplitude das flutuações dos dados.

3.7.2 Identificação e modelação das mudanças estruturais

A identificação e modelação das mudanças estruturais podem ser feitas por diferentes métodos estatísticos, usando modelos de intervenção e modelos estruturais.

A identificação de mudanças estruturais em séries temporais constitui um desafio particular, sobretudo quando a análise incide sobre séries longas ou envolve múltiplas variáveis interdependentes. Nestes casos, a complexidade aumenta significativamente, uma vez que é necessário detetar não apenas o momento exato em que ocorrem as quebras, mas também compreender em que dimensão do sistema essas alterações se manifestam. Quando se trata de séries multivariadas, esta tarefa exige especial atenção, dado que uma mudança pode afetar apenas uma parte do conjunto de variáveis, ou alterar as relações entre elas, comprometendo a estabilidade do modelo ao longo do tempo. Uma abordagem comum para detectar essas mudanças é utilizar testes estatísticos de ruptura estrutural. Existem diversos métodos para identificar esses pontos de ruptura, sendo o mais utilizado o teste de Bai e Perron (1998), que propõe um método para detectar múltiplos pontos de quebras estruturais em séries temporais.

O modelo proposto por Bai e Perron é um modelo de regressão segmentada, onde os dados são divididos em diferentes segmentos com base nos pontos de ruptura, e as estimativas dos coeficientes são ajustadas para cada segmento separadamente (Casini et al., 2018). Considerando m quebras, ou m+1 segmentos, o modelo de regressão segmentada pode ser expresso da seguinte forma:

$$y_t = x_t'\beta + z_t'\delta_j + u_t, \quad t = T_{j-1} + 1, \ldots, T_j, \quad j = 1, \ldots, m+1$$

Onde:

- y_t é a variável dependente,
- x_t e z_t são vetores das variáveis explicativas,
- β e δ são os coeficientes a serem estimados,
- u_t é o erro,

- $T_1, T_2, \ldots, T_m$ são os pontos de ruptura a serem identificados.
- $j = 1, \ldots, m+1$: número total de regimes, sendo m o número de quebras, o que implica $m+1$ regimes distintos.

Este modelo permite identificar mudanças estruturais na série temporal, sendo particularmente apropriado para detetar alterações na média e/ou na tendência da variável dependente ao longo do tempo. Através da definição de segmentos temporais distintos, este modelo acomoda alterações nos coeficientes associados a determinadas variáveis explicativas, revelando pontos de rutura nos padrões históricos do sistema. O objetivo do modelo é identificar os valores de $T_1, T_2, \ldots, T_m$ que minimizam a soma dos quadrados dos resíduos em cada segmento, com a condição de que os intervalos entre os pontos de ruptura sejam suficientemente grandes para garantir que cada segmento tenha um número suficiente de observações.

Além disso, quando se trabalha com uma única série temporal, a análise é classificada como univariada, ainda que seja possível integrar variáveis explicativas no modelo. Estas covariáveis podem incluir termos de tendência, *dummies* sazonais, variáveis de intervenção ou até variáveis exógenas economicamente relevantes. A sua inclusão visa aumentar a capacidade explicativa e preditiva do modelo, permitindo captar efeitos sistemáticos que influenciam o comportamento da série principal ao longo do tempo.

Quando realizamos a análise com quebras estruturais, detectamos mudanças no coeficiente da regressão (β) ao longo do tempo. Isso é feito usando testes de hipótese para determinar se o coeficiente permanece o mesmo ($H_0 : \beta_0 = \beta_i$) ou muda ($H_1 : \beta_0 \neq \beta_i$) em diferentes pontos no tempo.

Na prática, o processo envolve estimar o número de quebras e então analisar os coeficientes nos segmentos dos dados divididos pelos pontos de quebra (Zeileis et al., 2002)

Esse método é eficaz porque permite detectar várias mudanças ao longo do tempo, o que é crucial para analisar dados que podem ser impactados por diferentes políticas ou eventos. Além disso, Bai e Perron propõem o uso de programação dinâmica para tornar o processo computacionalmente mais eficiente, dada a grande quantidade de possíveis divisões nos dados.

3.7.3 Testes estatísticos para mudanças estruturais

A detecção de mudanças estruturais em séries temporais é um aspecto fundamental da modelação. Diversos testes foram desenvolvidos para esta finalidade, cada um com suas especificidades:

O teste de Chow, proposto originalmente por Chow (1960), examina a hipótese de existência de uma única quebra estrutural conhecida a priori em modelos de regressão. Baseia-se na teoria de que, se os parâmetros forem constantes, as previsões fora da amostra devem ser imparciais. Ele testa a hipótese nula de que não há quebra estrutural contra a alternativa de que há uma quebra estrutural conhecida no tempo . O teste considera um modelo linear dividido em amostras em um ponto de quebra predeterminado. Como destacam Hansen (2001), este método é particularmente útil quando o ponto de mudança é conhecido, mas apresenta limitações em casos de múltiplas quebras ou quando o ponto de ruptura é desconhecido. O modelo assume que a relação entre as variáveis explicativas e a variável dependente pode mudar num ponto de quebra estrutural conhecido, T_b:

$$y_t = \begin{cases} X_t^\top \beta_1 + u_t, & \text{se } t \leq T_b \\ X_t^\top \beta_2 + u_t, & \text{se } t > T_b \end{cases}$$

onde:

- y_t é a variável dependente no instante t;
- $X_t^\top$ é o vetor linha de variáveis explicativas;
- β_1 e β_2 são os vetores de coeficientes antes e depois de T_b;
- $u_t \sim \mathcal{N}(0, \sigma^2)$ é o termo de erro.

A hipótese nula e a alternativa do teste de Chow são:

$$H_0 : \beta_1 = \beta_2 \quad \text{(sem quebra estrutural)}$$
$$H_1 : \beta_1 \neq \beta_2 \quad \text{(existe quebra estrutural)}$$

A estatística F do teste de Chow é:

$$F = \frac{(S_C - (S_1 + S_2))/k}{(S_1 + S_2)/(n_1 + n_2 - 2k)} \sim F_{k,\, n_1+n_2-2k}$$

com:

- S_C: soma dos quadrados dos resíduos da regressão conjunta (sem quebra);
- S_1, S_2: somas dos quadrados dos resíduos das regressões separadas;
- k: número de parâmetros estimados (incluindo o intercepto);
- n_1, n_2: tamanhos das amostras antes e depois de T_b.

Por outro lado, o teste CUSUM, desenvolvido por Brown (1975), é eficaz para detectar instabilidades globais em modelos de regressão, particularmente alterações no nível médio ou na variabilidade dos resíduos, como demonstrado por Ploberger(1992). A estatística do teste CUSUM é calculada a partir do Resíduos de um passo à frente do modelo de mínimos quadrados recursivo. Baseia-se na intuição de que, se houver mudanças de um período para o outro, a previsão de um passo à frente não será precisa e o erro de previsão será maior que zero. Isso significa que quanto maior a estatística do teste CUSUM, maior o erro de previsão e maior a evidência estatística em favor da instabilidade dos parâmetrosNo entanto, este método apresenta menor sensibilidade para identificar mudanças específicas na tendência (inclinação) da série temporal. Este teste, baseia-se nos resíduos recursivos w_t:

$$W_t = \sum_{k+1}^{t} \frac{w_k}{\hat{\sigma}}, \quad t = k+1, ..., T \tag{3.35}$$

com banda crítica $\pm\lambda\sqrt{T - k + 2(t-k)/(T-k)}$ (Ploberger, 1992).

Para análise de mudanças na tendência, os modelos de regressão segmentada (Muggeo, 2003) e os testes de Bai e Perron (1998, 2003) são metodologicamente mais robustos. Portanto, as mudanças estruturais representam um desafio significativo na análise de séries temporais. A utilização de modelos apropriados, como os de intervenção e estruturais, permitem identificar as alterações na estrutura de séries temporais e melhorar a precisão das previsões. Métodos estatísticos como os testes de Bai e Perron (1998, 2003) e Perron (2006) são fundamentais para a deteção dessas quebras.

4

Análise dos dados

Neste capítulo, apresentamos a modelação e previsão da taxa de mortalidade infantil nos países da Comunidade dos Países de Língua Portuguesa (CPLP) de 1985 a 2020. Utilizamos técnicas de séries temporais para analisar a evolução histórica das taxas de mortalidade infantil e prever tendências futuras, com o objetivo de fornecer informações que sejam práticas para políticas públicas e intervenções de saúde.

Os dados utilizados nesta análise foram recolhidos junto do Banco Mundial e da UNICEF. A base de dados abrange as taxas de mortalidade infantil dos seguintes países da CPLP: Angola, Brasil, Cabo Verde, Guiné-Bissau, Guiné Equatorial, Moçambique, Portugal, São Tomé e Príncipe e Timor-Leste.

A variável principal analisada é a taxa de mortalidade infantil (número de mortes de crianças do sexo masculino e feminino com menos de um ano de idade por mil nascidos vivos) no período de 1985 a 2020 nos países da CPLP.

As análises estatísticas desenvolvidas ao longo deste trabalho foram realizadas com recurso ao software RStudio, na versão 2024.09.0+375, em articulação com a linguagem R, na sua versão 4.3.2. A escolha deste ambiente deveu-se à sua flexibilidade e à ampla disponibilidade de bibliotecas especializadas para séries temporais. Entre os pacotes mais relevantes destacam-se o forecast, utilizado para ajustar e estimar modelos ARIMA; o tseries, que facilitou a aplicação de testes como o ADF; e o strucchange, essencial para

identificar quebras estruturais nas séries. Para as previsões, adoptou-se um intervalo de confiança de 95%, o qual permite avaliar a margem de incerteza em torno das estimativas futuras. Esta opção metodológica confere maior segurança à interpretação dos resultados, respeitando as boas práticas da estatística aplicada.

4.1 Análise exploratória da base de dados

A análise exploratória dos dados é uma etapa fundamental no processo de análise de dados, pois permite entender a distribuição, variabilidade e principais características da base de dados utilizada neste estudo. Nesta seção, são apresentadas as principais técnicas e resultados obtidos na análise exploratória dos dados, fornecendo uma base sólida para as análises subsequentes.

Realizamos uma análise exploratória das séries temporais para identificar tendências, padrões sazonais e possíveis outliers. A Figura 4.1 mostra a evolução da taxa de mortalidade infantil nos países da CPLP ao longo do tempo.

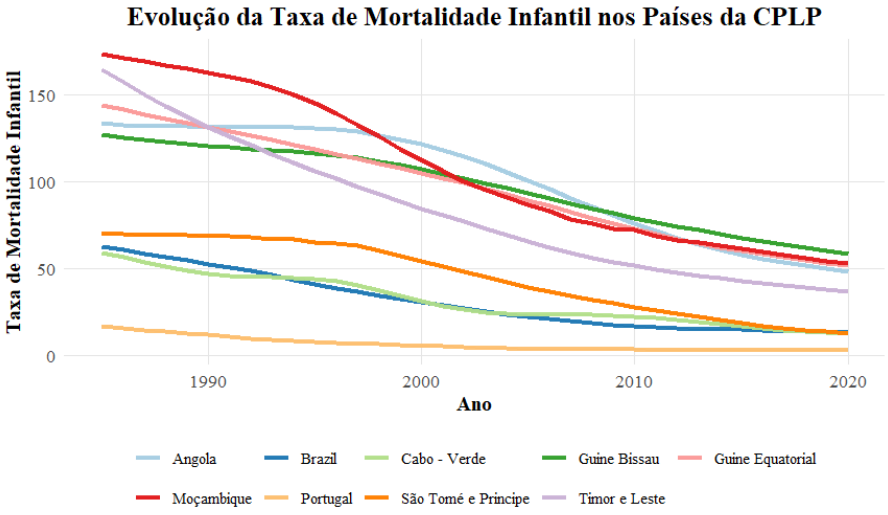

Figura 4.1: Evolução da Taxa de Mortalidade Infantil nos Países da CPLP

Como se pode ver pela figura 4.1 houve de modo geral um decréscimo acentuado nas taxas de mortalidade infantil da CPLP, com alguns países como, Brasil, Cabo Verde e Portugal com decréscimos mais acentuados neste período, em relação aos outros países do grupo..

A tabela 4.1 apresenta as estatísticas descritivas das taxas de mortalidade infantil dos distintos países da CPLP, considerando as populações masculina e feminina. As medidas apresentadas incluem a média, mediana, desvio padrão, e os valores mínimo e máximo para ambos os géneros.

País	Média		Mediana		Desvio Padrão		Mínimo		Máximo	
	Masculino	Feminino	Masculino	Feminino	Masculino	Feminino	Masculino	Feminino	Masculino	Feminino
Angola	109.3	92.55	121.40	102.75	33.17	29.86	53.4	43.10	142.80	123.70
Brasil	34.16	27.34	28.90	22.80	17.91	14.70	14.60	11.60	69.10	56.20
Cabo Verde	34.10	28.57	27.75	23.25	15.05	12.85	13.40	11.00	63.90	53.90
Guiné-Bissau	104.62	87.67	106.45	88.15	31.17	28.08	56.90	45.80	154.50	133.60
Guiné-Equatorial	104.10	88.59	108.20	92.05	23.74	21.37	63.70	52.60	136.10	117.70
Moçambique	113.06	101.80	102.65	92.75	43.86	41.21	56.70	48.70	181.70	165.30
Portugal	6.93	5.65	4.85	4.10	4.62	3.63	3.00	2.50	18.40	14.50
São Tomé e Príncipe	49.37	40.31	51.20	41.60	22.75	19.12	14.10	11.20	76.00	63.00
Timor-Leste	88.71	77.65	80.05	69.40	40.43	37.17	39.80	33.00	172.70	155.10

Tabela 4.1: Estatísticas Descritivas por País e Género

De acordo a tabela 4.1 observa-se claramente diferenças entre os valores médios e medianos, desvios, bem como os limites mínimo e máximo das taxas de mortalidade. De maneira geral, observa-se que Angola e Moçambique apresentam as taxas mais elevadas de mortalidade infantil, tanto para meninos quanto para meninas, o que sugere desafios persistentes em termos de saúde pública e acesso a cuidados adequados nessas regiões. Angola, em particular, revela uma média de mortalidade masculina de 109,3, a contrastar fortemente com os 6,93 em Portugal, o que pôe em evidência uma disparidade acentuada entre países em desenvolvimento e desenvolvidos. Além disso, a mortalidade infantil no sexo masculino é, de um modo geral, quase sempre superior à do sexo feminino.

Além disso, à análise comparativa das caixas de bigodes referentes às taxas de mortalidade infantil para o sexo masculino figura 4.2 e para o sexo feminino figura 4.3, confirmam estas disparidades nos países da CPLP, bem como diferentes padrões de variação ao longo do período. No caso de Angola e Moçambique, por exemplo, os valores medianos e as variações das taxas são elevadas, confirmando mais uma vez as vulnerabilidades persistentes na saúde infantil, enquanto Portugal e Brasil registam níveis mais baixos e estáveis, sugerindo políticas de saúde materno-infantil mais consolidadas. Por seu turno, países como Guiné-Bissau, Guiné Equatorial, São Tomé e Príncipe, Cabo Verde e Timor-Leste situam-se em patamares intermédios, com dispersões que podem indicar avanços pontuais ou retrocessos em determinados anos.

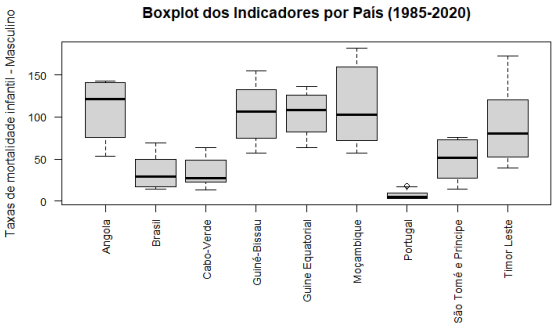

Figura 4.2: Boxplot das taxas de mortalidade infantil para indivíduos do sexo masculino nos países da CPLP (1985–2020).

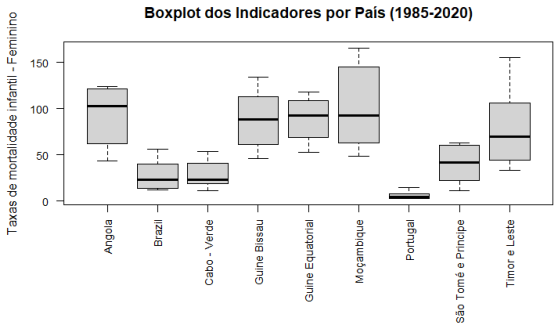

Figura 4.3: Boxplot das taxas de mortalidade infantil para indivíduos do sexo feminino nos países da CPLP (1985–2020).

A partir dos resultados do agrupamento *k-means* [1], que classificou os países da CPLP em dois clusters distintos (tabela 4.2), observou-se que Angola, Guiné-Bissau, Guiné Equatorial, Moçambique e Timor-Leste foram alocados ao Cluster 2, ao passo que Brasil, Cabo Verde, Portugal e São Tomé e Príncipe integraram o Cluster 1, tanto para indivíduos do sexo masculino como para indivíduos do sexo feminino. Importa referir que o K-means é um dos algoritmos de clustering mais populares e amplamente utilizados na análise de dados. O algoritmo K-means de clustering baseia-se no princípio de que existem um número especificado de K populações multivariadas normais presentes nos dados, e essas populações diferem apenas em termos de suas médias. Os membros dos clusters e suas médias são estimados de maneira a minimizar a variabilidade dentro de cada cluster (Zelterman, 2015). Esse processo resulta na formação de clusters onde as distâncias dentro de cada grupo são minimizadas, enquanto as distâncias entre os grupos são maximizadas, proporcionando uma boa separação entre os clusters (Everitt & Hothorn, 2011).

[1] Esta análise tem um carácter exploratório e deve ser interpretada com cautela. Os dados utilizados são séries temporais e, por isso, não cumprem pressupostos fundamentais da análise de clusters formal, como a normalidade das observações, a homogeneidade das variâncias (esfericidade) e a independência temporal. Não foi feita uma validação sistemática desses pressupostos, nem se corrigiram efeitos da autocorrelação ou de tendências. O objetivo desta análise é apenas reforçar visualmente a existência de dois grupos de países com comportamentos distintos, tal como já sugerido pela leitura gráfica das séries.

Tabela 4.2: Agrupamento de países por clusters. Fonte: Autor, extraído do RStudio.

Cluster	Países
1	Brasil, Cabo Verde, Portugal, São Tomé e Príncipe
2	Angola, Guiné-Bissau, Guiné Equatorial, Moçambique, Timor-Leste

Esta segmentação constituiu o ponto de partida para a análise mais aprofundada da evolução temporal dos indicadores, representada nas figuras agora apresentados (Figura 4.4 e Figura 4.5).

Nos referidos gráficos, as linhas a azul correspondem aos países do Cluster 1, enquanto as linhas a vermelho representam os países do Cluster 2. De maneira análoga, para os indivíduos masculino e feminino, esta diferenciação visual confirma, de forma inequívoca, a separação identificada na análise de clusters: os países de cada grupo evidenciam comportamentos semelhantes, quer ao nível dos valores registados quer na tendência de evolução ao longo das décadas. Assim, fica patente que os países do Cluster 1 se situam, de forma consistente, em patamares mais baixos de mortalidade infantil, enquanto os do Cluster 2 apresentam valores notoriamente mais elevados, ainda que ambos os conjuntos registem um declínio global ao longo do período em estudo. Esta convergência entre os resultados do k-means e a observação directa dos gráficos reforça as disparidades estruturais na saúde infantil entre os países lusófonos.

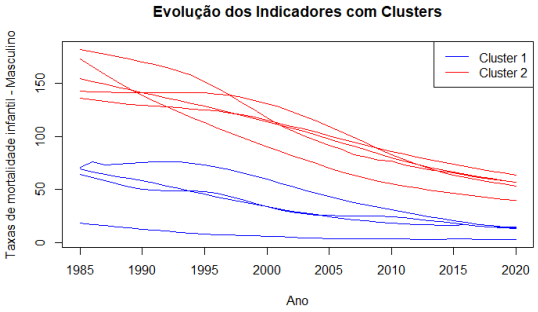

Figura 4.4: Evolução temporal das taxas de mortalidade infantil nos países da CPLP, divididos em dois clusters, para indivíduos do sexo masculino (1985–2020).

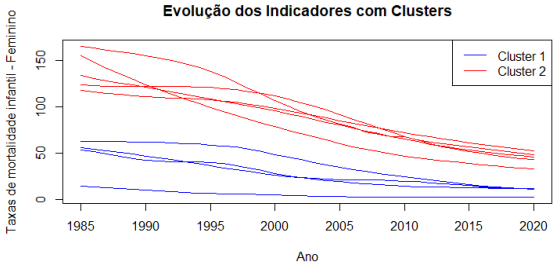

Figura 4.5: Evolução temporal das taxas de mortalidade infantil nos países da CPLP, divididos em dois clusters, para indivíduos do sexo feminino (1985–2020).

Agora passamos à análise dos pontos de ruptura, cuja identificação permite compreender as mudanças estruturais subjacentes aos dados.

4.2 Mudança de estrutura

Com o objetivo de identificar alterações estruturais relevantes nas séries temporais das taxas de mortalidade infantil dos países da CPLP, foi aplicada a metodologia de Bai e Perron, considerando uma regressão linear simples com termo temporal ($y_t = \alpha_j + \beta_j \cdot t + u_t$) e matriz de covariância de Newey-West. Embora este modelo permita detetar simultaneamente mudanças de nível e de tendência, optou-se por uma interpretação centrada nas mudanças de nível, que refletem variações estatisticamente significativas nas médias das taxas ao longo do tempo. Esta escolha foi motivada por razões de clareza e estabilidade interpretativa, dado que o foco da análise recai sobre alterações marcantes no comportamento médio das séries, mais do que em variações no seu declive.

As quebras identificadas foram posteriormente utilizadas para segmentar os períodos das séries temporais,

sendo então calculadas as médias da taxa de mortalidade infantil em cada segmento, por país e por sexo. Estas mudanças podem estar associadas a fatores diversos, como políticas públicas de saúde, intervenções internacionais, contextos de instabilidade política ou alterações demográficas significativas. Os resultados que se seguem devem, portanto, ser lidos como uma aproximação descritiva e exploratória das fases distintas na evolução da mortalidade infantil ao longo das últimas décadas nos países da CPLP. A seguir, são apresentados os resultados da identificação de possíveis quebras estruturais nas séries temporais dos países em análise.

Angola

País	Ano Masculino	Média Masculino	Ano Feminino	Média Feminino
Angola	1996	141,2750	1994	118,55
	2001	103,7556	1999	106,03
	2010	70,4000	2004	84,53
	2015	57,0600	2009	62,80
	-	-	2014	48,19

Tabela 4.3: Comparação das médias de mortalidade infantil antes e depois da mudança de estrutura para Angola

Brasil

A análise dos resultados da tabela 4.3 apontam uma quebra significativa na taxa de mortalidade masculina entre 1996 e 2001, período em que a média caiu de 141,28 para 103,76, representando uma redução de 37,52. Este declínio acentuado sugere a implementação eficaz de políticas de saúde pública, talvez como programas de vacinação e intervenções sociais, que podem ter contribuído para melhorar as condições de vida e de saúde infantil. Para as meninas, a maior diminuição ocorreu entre 2004 e 2009, com uma redução de 21,73, mas essa queda foi menos pronunciada em comparação com os rapazes.

País	Masculino		Feminino	
	Ano	Média	Ano	Média
Brasil	1992	44,32	1990	39,71
	1998	30,03	1997	27,77
	2005	20,70	2002	20,12
	2010	16,78	2007	15,10
	2015	15,30	2012	12,50

Tabela 4.4: Comparação das médias antes e depois da mudança de estrutura para o Brasil

Cabo Verde

De acordo a tabela 4.4 as médias de mortalidade infantil no Brasil entre os anos de 1990 e 2015 demonstra uma tendência geral de diminuição tanto para meninos quanto para meninas. No caso dos meninos, a maior redução foi observada entre 1992 e 1998, quando a média caiu de 44,32 para 30,03, representando uma diminuição de 14,29. Para as meninas, a descida mais acentuada ocorreu entre 1990 e 1997, com a média diminuindo de 39,71 para 27,77, resultando numa redução de 11,94. Embora ambas as categorias tenham mostrado progressos significativos na redução das taxas de mortalidade infantil, os dados sugerem que as intervenções de saúde pública implementadas nesse período tiveram um impacto positivo, contribuindo

para melhores condições de saúde e sobrevivência infantil no Brasil. A tendência contínua de declínio nas médias também indica a eficácia de políticas direcionadas ao bem-estar infantil ao longo do tempo.

País	Masculino		Feminino	
	Ano	Média	Ano	Média
Cabo Verde	1990	48,80	1990	48,80
	1995	38,52	1995	38,52
	2002	25,79	2002	25,79
	2010	21,45	2010	21,45
	2015	15,58	2015	15,58

Tabela 4.5: Comparação das médias antes e depois da mudança de estrutura para Cabo Verde

A análise das médias de mortalidade infantil em Cabo Verde entre 1990 e 2015, segundo os resultados apresentados na tabela 4.5, revela uma tendência de queda constante e significativa, tanto para os meninos quanto para as meninas. A partir de 1990, com uma média de 48,80, observa-se uma diminuição progressiva, alcançando 38,52 em 1995, 25,79 em 2002 e 21,45 em 2010. O declínio mais acentuado ocorre entre 2002 e 2015, quando a média cai para 15,58, indicando uma redução substancial nas taxas de mortalidade infantil. Essa tendência positiva reflete o impacto das políticas de saúde pública implementadas em Cabo Verde, que visam melhorar a saúde infantil e garantir melhores condições de vida. O fato de que as médias para meninos e meninas se mantiveram idênticas ao longo dos anos sugere uma equidade nas intervenções de saúde e no acesso aos serviços, resultando em melhorias significativas na sobrevivência infantil no país.

Guiné-Bissau

País	Masculino		Feminino	
	Ano	Média	Ano	Média
Guiné-Bissau	1985	143,84	1985	143,84
	1993	120,97	1993	120,97
	2002	94,07	2002	94,07
	2010	72,53	2010	72,53
	2015	61,30	2015	61,30

Tabela 4.6: Comparação das médias antes e depois da mudança de estrutura para Guiné-Bissau

A tabela 4.6 apresenta a análise das médias de mortalidade infantil na Guiné-Bissau entre 1993 e 2015, revelando uma diminuição acentuada nas taxas tanto para meninos quanto para meninas. A média em 1993, que era de 120,97, mostra um cenário crítico. A partir desse ano, observa-se uma queda gradual, atingindo 94,07 em 2002 e 72,53 em 2010. A descida mais significativa ocorre entre 2010 e 2015, quando a média cai para 61,30. Essa redução aponta para uma melhoria nas condições de saúde infantil no país, possivelmente impulsionada por iniciativas de saúde pública e/ou outros programas de intervenção. Apesar de ainda existirem desafios, a tendência de queda nas taxas de mortalidade infantil na Guiné-Bissau é um indicativo positivo de progresso em direção a melhores resultados de saúde para as crianças.

Guiné Equatorial

País	Masculino		Feminino	
	Ano	Média	Ano	Média
Guiné Equatorial	1989	127,08	1989	109,36
	1996	118,62	1996	101,55
	2001	100,64	2001	85,21
	2009	80,91	2009	68,73
	2015	68,50	2014	57,74

Tabela 4.7: Comparação das médias antes e depois da mudança de estrutura para o Guiné Equatorial

Segundo os resultados exibidos na tabela 4.7 as médias para meninos em 1989 era de 127,08, enquanto para meninas era de 109,36, evidenciando um cenário preocupante. A partir desse ano, nota-se uma tendência de redução contínua nas taxas de mortalidade, com os valores caindo para 118,62 em 1996 e, subsequentemente, para 100,64 em 2001. A descida se torna ainda mais pronunciada em anos posteriores, alcançando 80,91 em 2009 e 68,50 em 2015 para meninos. Para meninas, as médias também diminuem, atingindo 68,73 em 2009 e 57,74 em 2014. Essa evolução reflete uma melhoria significativa nas condições de saúde infantil, possivelmente resultante de esforços governamentais e iniciativas de saúde pública. A Guiné Equatorial apresenta uma trajetória encorajadora na redução das taxas de mortalidade infantil, sinalizando avanços importantes na promoção da saúde e bem-estar das crianças.

Moçambique

País	Masculino		Feminino	
	Ano	Média	Ano	Média
Moçambique	1989	165.53	1989	109,36
	1994	134.68	1996	101,55
	2001	98.63	2001	85,21
	2006	79.80	2009	68,73
	2011	64.57	2014	57,74

Tabela 4.8: Comparação das médias antes e depois da mudança de estrutura para Moçambique

A tabela 4.8 apresenta uma análise das médias de mortalidade infantil em Moçambique, abrangendo os anos de 1989 a 2011. Inicialmente, em 1989, a taxa de mortalidade infantil para meninos era alarmante, com uma média de 165,53. Em contraste, a média para meninas nesse mesmo ano era de 109,36, evidenciando uma discrepância significativa nas taxas de mortalidade entre os sexos. A partir de 1989, as taxas de mortalidade infantil começam a diminuir de forma acentuada. Em 1994, a média para meninos caiu para 134,68, e, em 2001, houve uma descida significativa para 98,63. A tendência de queda continua, alcançando 79,80 em 2006 e 64,57 em 2011. Para as meninas, a média também diminuiu de 101,55 em 1996 para 57,74 em 2014. Essa evolução positiva nas taxas de mortalidade infantil sugere melhorias nas condições de saúde pública e nas intervenções sociais em Moçambique, refletindo um progresso importante na redução da mortalidade infantil e na promoção da saúde infantil ao longo dos anos.

Portugal

País	Masculino		Feminino	
	Ano	Média	Ano	Média
Portugal	1989	11,27	1992	7,08
	1994	6,43	1997	4,63
	2004	3,72	2005	2,94
	2009	3,27	2011	2,66
	2014	3,19	-	-

Tabela 4.9: Comparação das médias de mortalidade infantil em Portugal antes e depois das mudanças de estrutura

Por outro lado a tabela4.9 apresenta a comparação das médias de mortalidade infantil em Portugal, abrangendo os anos de 1989 a 2014. Em 1989, a taxa de mortalidade infantil para meninos era de 11,27, enquanto para meninas, em 1992, a média era significativamente mais baixa, com 7,08. Ao longo dos anos, observa-se uma tendência contínua de redução nas taxas de mortalidade infantil em ambos os sexos. A média para meninos diminuiu para 6,43 em 1994, 3,72 em 2004, 3,27 em 2009 e 3,19 em 2014. Para as meninas, a mortalidade infantil apresentou uma queda semelhante, passando de 7,08 em 1992 para 4,63 em 1997 e, subsequentemente, para 2,94 em 2005 e 2,66 em 2011. Essa diminuição acentuada nas taxas de mortalidade infantil em Portugal reflete a eficácia das políticas de saúde pública e dos avanços nos cuidados médicos e na qualidade de vida, evidenciando um progresso significativo na proteção da infância ao longo das décadas. A ausência de dados para meninas em 2014 não diminui a relevância desta análise, que demonstra uma trajetória positiva em termos de saúde infantil no país.

São Tomé e Principe

País	Masculino		Feminino	
	Ano	Média	Ano	Média
São Tomé e Príncipe	1992	73,22	1992	59,18
	1997	54,59	1997	45,88
	2006	30,02	2005	29,45
	2015	16,98	2010	20,18
	-	-	2015	13,43

Tabela 4.10: Comparação das médias de mortalidade infantil em São Tomé e Príncipe antes e depois das mudanças de estrutura

Entretanto, a tabela 4.10 apresenta as médias de mortalidade infantil em São Tomé e Príncipe, abrangendo os anos de 1992 a 2015, com dados separados por sexo. Para os meninos, a média de mortalidade infantil começou em 73,22 em 1992, apresentando uma queda significativa ao longo dos anos. Em 1997, a taxa caiu para 54,59, e em 2006, a média atingiu 30,02. A tendência de diminuição continuou, com uma média de 16,98 em 2015. No que diz respeito às meninas, a taxa de mortalidade infantil também diminuiu, passando de 59,18 em 1992 para 45,88 em 1997 e, posteriormente, para 20,18 em 2010, antes de alcançar 13,43 em 2015. A análise desses dados indica uma clara melhoria nas condições de saúde infantil em São Tomé e Príncipe ao longo do período analisado. As quedas acentuadas nas médias refletem avanços nas políticas de saúde, acesso a cuidados médicos e melhorias nas condições socioeconómicas, ressaltando um progresso significativo na redução da mortalidade infantil no país.

Timor-Leste

País	Masculino		Feminino	
	Ano	Média	Ano	Média
Timor-Leste	1989	130,75	1989	130,75
	1994	103,61	1994	103,618
	2000	76,36	2000	76,36
	2007	56,88	2007	56,88
	2012	45,20	2012	45,20

Tabela 4.11: Comparação das médias de mortalidade infantil em Timor-Leste antes e depois das mudanças de estrutura

Por último, a tabela 4.11 apresenta as médias de mortalidade infantil em Timor-Leste, com dados que vão de 1989 a 2012, separados por sexo. Nos dados masculinos, observa-se que a taxa de mortalidade infantil era de 130,75 em 1989, mostrando uma tendência de queda significativa ao longo dos anos. Em 1994, a média caiu para 103,61, e, a partir daí, continuou a declinar, atingindo 76,36 em 2000. A descida prosseguiu com 56,88 em 2007 e, finalmente, 45,20 em 2012.

As informações refletem uma melhoria substancial nas condições de saúde infantil em Timor-Leste. A diminuição contínua das taxas de mortalidade infantil pode ser atribuída a vários fatores, incluindo o aumento do acesso a serviços de saúde, melhorias nas condições de vida, bem como políticas governamentais voltadas para a saúde pública e a educação. Essas mudanças indicam um progresso positivo no contexto da saúde infantil, evidenciando o impacto de iniciativas eficazes em saúde e desenvolvimento social no país.

Contudo, de modo geral, essas reduções podem ser atribuídas a uma combinação de fatores, por exemplo à:

Intervenções em saúde pública: Melhoria do acesso a cuidados de saúde, vacinação, nutrição e educação. Condições socioeconómicas: Avanços nas condições de vida e saneamento básico. Políticas governamentais: Implementação de programas específicos voltados para a redução da mortalidade infantil. A eficácia das intervenções e políticas varia de país para país, mas a tendência geral aponta para uma melhoria nas condições de saúde infantil, refletida na redução das taxas de mortalidade. Esta análise serve como base para a identificação de melhores práticas e políticas que podem ser replicadas em contextos semelhantes.

4.3 Modelação das taxas de mortalidade infantil

4.3.1 Teste da estacionariedade e identificação do modelo

A análise das séries temporais de mortalidade infantil nos países da CPLP foi realizada com base no teste ADF para avaliar a estacionariedade e na função ndiffs para determinar quantas diferenciações são necessárias para tornar as séries estacionárias. Além disso, foram aplicadas as transformções de Box Cox. A tabela 4.12 resume os resultados do teste ADF e as diferenciações necessárias para garantir a estacionariedade das séries. Como essas séries não exibem sazonalidade (evidenciado pela ausência de padrões sazonais significativos nos testes de autocorrelação e análise gráfica), optou-se por não utilizar modelos SARIMA, na nossa abordagem. Em vez disso, modelos ARIMA convencionais foram considerados adequados para a análise.

País	Dickey-Fuller		p-value		ndiffs	
	Masculino	Feminino	Masculino	Feminino	Masculino	Feminino
Angola	-4,1347	-3,8140	0,0158	0,0308	2	2
Brasil	-0,9929	-1,4746	0,9252	0,7776	1	2
Cabo Verde	-0,2849	-0,4098	0,9857	0,9803	1	1
Guiné-Bissau	-1,6175	-4,8361	0,7220	0,0100	1	1
Guiné-Equatorial	-3,5221	-3,5417	0,0555	0,0527	2	2
Moçambique	-1,9562	-2,0419	0,5902	0,5569	2	2
Portugal	-3,4087	-3,0979	0,0720	0,1459	2	2
São Tomé e Príncipe	-2,6856	-3,6403	0,3064	0,0439	2	2
Timor-Leste	-0,2409	-0,3945	0,9876	0,9810	2	2

Tabela 4.12: Estacionariedade das Séries pelo teste ADF

A análise revela uma complexidade significativa na avaliação da estacionariedade das séries temporais de mortalidade infantil nos países da CPLP. Embora o Teste ADF tenha indicado estacionariedade em alguns países (como Angola e São Tomé e Príncipe), a função ndiffs() sugere que são necessárias diferenciações adicionais. Esta discrepância pode ser explicada por fatores como a presença de tendências subtis ou sazonalidade nas séries temporais, que o Teste ADF pode não captar totalmente.

Por outro lado, países como Brasil, Moçambique e Timor-Leste apresentaram séries claramente não estacionárias, necessitando de duas diferenciações para estabilizá-las. Além disso, a Guiné-Bissau apresentou um comportamento divergente entre géneros, com as séries temporais de raparigas a serem estacionárias e as de rapazes a não serem, sugerindo diferenças de comportamento nas tendências de mortalidade infantil entre géneros.

Esses resultados indicam que a simples aplicação do Teste ADF não é suficientemente robusto para testar a estacionariedade.. O uso complementar de métodos como ndiffs é essencial para garantir uma análise mais clara, uma vez que a diferenciação pode remover padrões de longo prazo que não são captados pelo ADF, garantindo assim uma série adequada para modelação e previsão.

Uma vez que a estacionariedade foi alcançada, procurámos identificar um modelo apropriado para obter previsões confiáveis. Para esse propósito, recorremos à conhecida metodologia Box-Jenkins para a identificação da ordem do modelo ARIMA. Conforme a abordagem descrita por Box et al. (2015), a ordem do modelo ARIMA é determinada pelos componentes p, d e q.

Para a estimativa dos parâmetros p e q, utilizámos os gráficos de ACF e PACF com base nos valores de atraso. Os valores de d foram estimados de acordo ao numero de diferenciações necessárias para estacionar as respetivas séries. Esses gráficos foram adicionados ao Apêndice (ver Apêndice A para mais detalhes).

Para evitar qualquer conclusão equivocada, baseada apenas na avaliação gráfica de um modelo provisório, considerámos outros modelos candidatos próximos. A seleção do modelo final foi baseado de acordo aos menores valores de AIC, conforme propostos por Akaike (1974) e Schwarz (1978). Este critério fornece uma base sólida para a escolha do modelo que melhor se ajusta aos dados, penalizando a complexidade do modelo e promovendo a parcimónia, o que é fundamental para garantir a validade das previsões geradas.

No entanto, os modelos concorrentes, juntamente com os seus valores de AIC são mostrados em colunas separadas, indicando os resultados obtidos antes e depois da remoção dos parâmetros não significativos. Para mais detalhes sobre os valores do AIC, poderão ser consultados no Apêndice B. No corpo do texto, consideramos apenas os modelos que apresentaram os menores valores de AIC, sendo estes os que consideramos

devidamente ajustados ao nosso conjunto de dados.

4.3.2 Estimação do modelo

Os modelos ARIMA foram estimados para cada país. As tabelas abaixo apresentam os modelos ajustados e os parâmetros estimados dos modelos ARIMA para cada pais da CPLP, diferenciado pelos géneros.

Angola

País	Modelo	Parâmetros	Coeficiente	Desvio padrão	t-teste	p-value
Angola	ARIMA(4,2,1)	$\hat{\phi}_1$	1,2373	0,0378	32,7019	<0,05
		$\hat{\phi}_4$	-0,3148	0,0287	-8,3200	<0,05
		$\hat{\theta}_1$	-1,0987	0,1499	-38,2730	<0,05

Tabela 4.13: Modelo ARIMA ajustado para os indivíduos do sexo masculino em Angola no período compreendido entre 1985 a 2020

País	Modelo	Parâmetros	Coeficiente	Desvio padrão	t-teste	p-value
Angola	ARIMA(2,2,0)	$\hat{\phi}_1$	0,4711	0,1586	2,9704	0,0054
		$\hat{\phi}_2$	0,3838	0,1591	2,4119	0,0214

Tabela 4.14: Modelo ARIMA ajustado para os indivíduos do sexo feminino em Angola no período compreendido entre 1985 a 2020

As tabelas apresentadas contêm os resultados dos modelos ARIMA ajustados para as taxas de mortalidade infantil em Angola, separados por sexo, entre 1985 e 2020. A tabela 4.13 reflete o modelo ARIMA(4,2,1) para indivíduos do sexo masculino, cuja equação geral pode ser expressa como:

$$\Delta^2 X_t = 1,2373\Delta^2 X_{t-1} - 0,3148\Delta^2 X_{t-4} - 1,0987\varepsilon_{t-1}$$

onde Y_t representa a taxa de mortalidade infantil em tempo t, $\Delta^2 X_t$ indica a segunda diferença da série temporal, e ε_t é o erro aleatório. Os parâmetros estimados demonstram que $\hat{\phi}_1 = 1,2373$ apresenta uma relação positiva e forte com a série temporal, enquanto $\hat{\phi}_4 = -0,3148$ e $\hat{\theta}_1 = -1,0987$ sugerem efeitos opostos na dinâmica da mortalidade infantil, ambos com significância estatística, com p-values inferiores a 0,05.

Por outro lado, a tabela 4.14 mostra o modelo ARIMA(2,2,0) ajustado para indivíduos do sexo feminino, que pode ser representado pela equação:

$$\Delta^2 X_t = 0,4711\Delta^2 X_{t-1} + 0,3838\Delta^2 X_{t-2}$$

Os parâmetros estimados para este modelo são $\hat{\phi}_1 = 0,4711$ e $\hat{\phi}_2 = 0,3838$. Ambos os coeficientes são estatisticamente significativos, com p-values de 0,0054 e 0,0214, respetivamente, indicando que, tal como no modelo masculino, há uma influência significativa destes parâmetros sobre a mortalidade infantil feminina. Em suma, os resultados obtidos sugerem que ambos os modelos ARIMA para Angola revelam

comportamentos dinâmicos distintos na mortalidade infantil, evidenciando a importância de se considerar as diferenças entre os sexos na análise das taxas de mortalidade.

Para ilustrar de forma mais clara esses achados, apresentam-se a seguir os gráficos dos modelos ajustados para ambos os sexos, possibilitando uma comparação visual dos padrões de mortalidade ao longo do tempo.

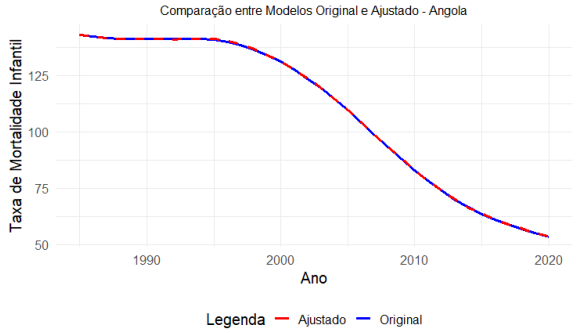

Figura 4.6: Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo masculino - Angola.

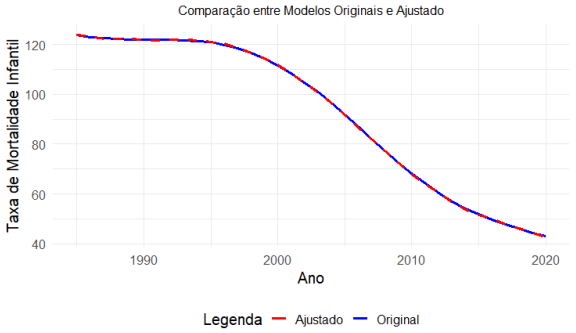

Figura 4.7: Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo feminino - Angola.

Como mostrado nas Figuras 4.6 e 4.7 apresentam a comparação entre a série temporal original das taxas de mortalidade infantil em Angola (linha azul) e o modelo ajustado (linha vermelha), para os indíviduos dos sexo masculino e feminino, entre 1985 a 2020. Observa-se que ambas as curvas praticamente se sobrepõem, indicando que o modelo proposto consegue representar de forma consistente a evolução histórica dos dados. Além disso, verifica-se uma tendência decrescente ao longo do tempo, sugerindo melhorias graduais no indicador de mortalidade infantil nesse intervalo analisado.

Brasil

País	Modelo	Parâmetros	Coeficiente	Desvio padrão	t-teste	p-value
Brasil	ARIMA(0,2,2)	$\hat{\theta}_1$	-0,7889	0,1719	-4,5885	<0,05
		$\hat{\theta}_2$	0,4983	0,1794	2,7785	<0,05

Tabela 4.15: Modelo ARIMA ajustado para os indivíduos do sexo masculino no Brasil no período compreendido entre 1985 a 2020

País	Modelo	Parâmetros	Coeficiente	Desvio padrão	t-teste	p-value
Brasil	ARIMA(3,2,1)	$\hat{\phi}_1$	-0,7063	0,1476	-4,7843	<0,05
		$\hat{\phi}_2$	0,3549	0,1762	2.0143	0,0524
		$\hat{\phi}_3$	0,5158	0,1461	3.5308	<0,05
		$\hat{\theta}_1$	0,9625	0,1755	5,4860	<0,05

Tabela 4.16: Modelo ARIMA ajustado para os indivíduos do sexo feminino no Brasil no período compreendido entre 1985 a 2020

As tabelas apresentadas contêm os resultados dos modelos ARIMA ajustados para as taxas de mortalidade

infantil no Brasil, separados por sexo, entre 1985 e 2020. A tabela 4.15 reflete o modelo ARIMA(0,2,2) para indivíduos do sexo masculino, cuja equação geral pode ser expressa como:

$$\Delta^2 X_t = -0,7889\varepsilon_{t-1} + 0,4983\varepsilon_{t-2}$$

onde Y_t representa a taxa de mortalidade infantil em tempo t, $\Delta^2 Y_t$ indica a segunda diferença da série temporal, e ε_t é o erro aleatório. Os parâmetros estimados demonstram que $\hat{\theta}_1 = -0,7889$ apresenta uma relação negativa com a série temporal, enquanto $\hat{\theta}_2 = 0,4983$ indica um efeito positivo, ambos com significância estatística, com p-values inferiores a 0,05.

Por outro lado, a tabela 4.16 mostra o modelo ARIMA(3,2,1) ajustado para indivíduos do sexo feminino, que pode ser representado pela equação:

$$\Delta^2 X_t = -0,7063\Delta^2 X_{t-1} + 0,3549\Delta^2 X_{t-2} + 0,5158\Delta^2 X_{t-3} + 0,9625\varepsilon_{t-1}$$

Os parâmetros estimados para este modelo são $\hat{\phi}_1 = -0,7063$, $\hat{\phi}_2 = 0,3549$, $\hat{\phi}_3 = 0,5158$ e $\hat{\theta}_1 = 0,9625$. Todos esses coeficientes são estatisticamente significativos, com p-values de 0,0524, 0,0012 e valores inferiores a 0,05 para os demais, indicando que, tal como no modelo masculino, há uma influência significativa destes parâmetros sobre a mortalidade infantil feminina. Em suma, os resultados obtidos sugerem que ambos os modelos ARIMA para o Brasil revelam comportamentos dinâmicos distintos na mortalidade infantil, evidenciando a importância de se considerar as diferenças entre os sexos na análise das taxas de mortalidade. Apresentam-se a seguir os gráficos (4.8 e 4.9) dos modelos ajustados para ambos os sexos:

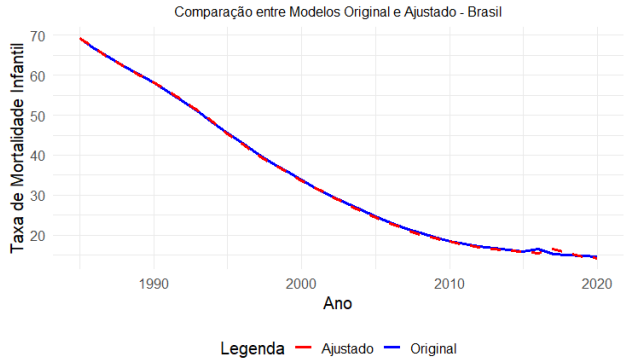

Figura 4.8: Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo masculino - Brasil.

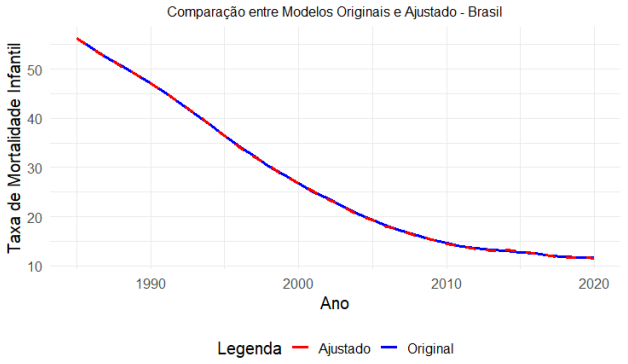

Figura 4.9: Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo feminino - Brasil.

É possível observar uma convergência muito próxima entre as duas linhas ao longo do período analisado (1985 a 2020), indicando que o modelo ajustado segue de forma eficaz a tendência dos dados originais. Isso sugere que o modelo é bem-sucedido na captura das dinâmicas subjacentes à evolução das taxas de mortalidade infantil, sem grandes desvios entre os valores observados e ajustados.

Este comportamento reforça a confiança na precisão do modelo ajustado, pois ele reflete adequadamente as flutuações das taxas de mortalidade infantil ao longo dos anos, mantendo uma consistência com os dados históricos. A aproximação das duas linhas sugere, portanto, que o modelo utilizado foi eficaz em capturar a tendência de queda nas taxas de mortalidade infantil.

Cabo Verde

País	Modelo	Parâmetros	Coeficiente	Desvio padrão	t-teste	p-value
Cabo Verde	ARIMA(2,1,0)	$\hat{\phi}_1$	1,7363	0,0497	34,9554	<0,05
		$\hat{\phi}_2$	-0,9268	0,0487	-19.0470	<0,05
		c	-1,3759	0,1424	-9.6599	<0,05

Tabela 4.17: Modelo ARIMA ajustado para os indivíduos do sexo masculino e feminino de Cabo Verde no período compreendido entre 1985 a 2020

A tabela apresentada contém os resultados do modelo ARIMA ajustado para as taxas de mortalidade infantil em Cabo Verde, considerando ambos os sexos, entre 1985 e 2020. A tabela 4.17 reflete o modelo ARIMA(2,1,0), cuja equação geral pode ser expressa como:

$$\Delta X_t = 1,7363\Delta X_{t-1} - 0,9268\Delta X_{t-2} - 1,3759$$

onde Y_t representa a taxa de mortalidade infantil em tempo t, ΔX_t indica a primeira diferença da série temporal. Os parâmetros estimados demonstram que $\hat{\phi}_1 = 1,7363$ apresenta uma relação positiva e forte com a série temporal, enquanto $\hat{\phi}_2 = -0,9268$ sugere um efeito negativo. O termo constante $c = -1,3759$ também é significativo, todos com p-values inferiores a 0,05. Esses resultados indicam uma influência estatisticamente significativa dos parâmetros sobre a mortalidade infantil em Cabo Verde. Em suma, o modelo ARIMA ajustado para Cabo Verde evidencia uma dinâmica complexa nas taxas de mortalidade infantil, ressaltando a importância de se considerar as características específicas do contexto nacional.

A análise comparativa entre os modelos original e ajustado das taxas de mortalidade infantil em Cabo Verde, como observado nas figuras 4.10 e 4.11, para ambos os sexos, onde a linha azul representa os dados originais, enquanto a linha vermelha reflete os dados ajustados pelo modelo, ao longo do período de 1985 a 2020, revelam uma convergência significativa entre as duas linhas, indicando que o modelo ajustado segue de perto a tendência dos dados originais, po que torna evidente a redução das taxas de mortalidade infantil no país.

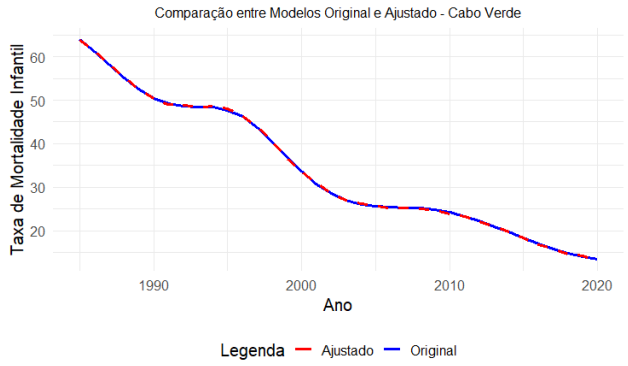

Figura 4.10: Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo masculino - Cabo Verde.

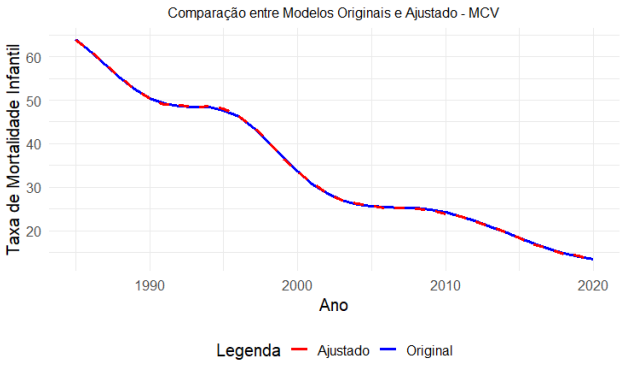

Figura 4.11: Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo feminino - Cabo Verde.

Guiné-Bissau

País	Modelo	Parâmetros	Coeficiente	Desvio padrão	t-teste	p-value
Guiné-Bissau	ARIMA(3,1,0)	$\hat{\phi}_1$	1,3268	0,0686	19,3344	<0,05
		$\hat{\phi}_2$	0	0	-	-
		$\hat{\phi}_3$	-0,3943	0,0723	-5.4500	<0,05
		c	-2,6807	0,3110	-8.6189	<0,05

Tabela 4.18: Modelo ARIMA ajustado para os indivíduos do sexo masculino e feminino da Guiné-Bissau no período compreendido entre 1985 a 2020

A tabela 4.18 apresenta o modelo ARIMA(3,1,0) ajustado para as taxas de mortalidade infantil na Guiné--Bissau entre 1985 e 2020, para ambos os sexos, o que indica que a série temporal foi diferenciada uma vez para estabilizar a variância e a média. A equação correspondente é dada por:

$$\Delta X_t = -2,6807 + 1,3268X_{t-1} - 0,3943X_{t-3} \tag{4.1}$$

onde ΔX_t representa a variação na taxa de mortalidade infantil de um período para outro. Neste contexto, o termo constante ($c = -2,6807$) reflete a tendência média da série quando os valores dos períodos anteriores estão em zero. O coeficiente $\hat{\phi}_1 = 1,3268$ indica que um aumento de uma unidade no valor de X_{t-1} está associado a um aumento de aproximadamente 1,33 unidades na taxa de mortalidade infantil, sugerindo uma relação direta entre os dados do período anterior e o atual. Por outro lado, o coeficiente $\hat{\phi}_3 = -0,3943$ sugere que um aumento de uma unidade no valor de X_{t-3} leva a uma diminuição de cerca de 0,39 unidades na taxa de mortalidade, indicando uma relação inversa com o valor observado três períodos antes. Esta estrutura autorregressiva demonstra a importância dos valores passados na previsão das taxas de mortalidade infantil, o que permite uma análise mais robusta das dinâmicas temporais presentes na série estudada. A seguir apresentamos os gráficos dos modelos original e ajustados para a série temporal de Guiné-Bissau.

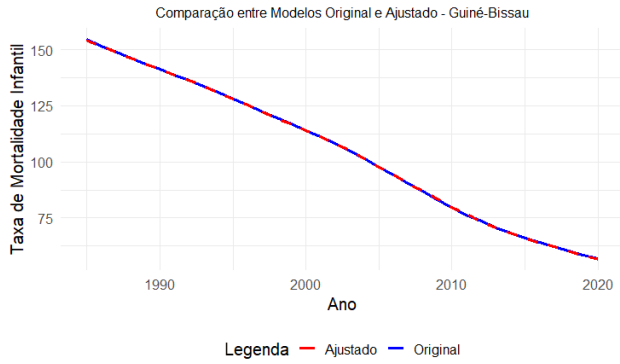

Figura 4.12: Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo masculino - Guiné-Bissau.

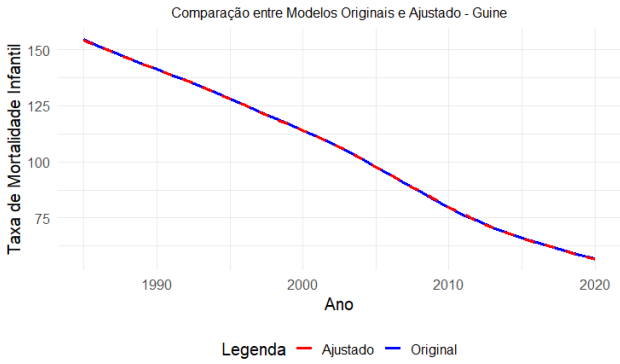

Figura 4.13: Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo feminino - Guiné-Bissau.

Os gráficos 4.12 e 4.13 fornecem uma comparação entre os modelos original e ajustado das taxas de mortalidade infantil na Guiné-Bissau,para ambos os sexos. Observa-se uma tendência de redução contínua das taxas de mortalidade infantil ao longo do período analisado (1985-2020), tanto nos dados originais quanto nos ajustados. A proximidade entre os dois modelos indica que o ajuste realizado reflete de forma

eficaz a tendência observada nos dados reais, sem grandes desvios.

Guine Equatorial

País	Modelo	Parâmetros	Coeficiente	Desvio padrão	t-teste	p-value
Guiné Equatorial	ARIMA(2,2,0)	$\hat{\phi}_1$	0,4776	0,1607	2,9717	0,0054
		$\hat{\phi}_2$	0,2784	0,1726	1,6124	0,1161

Tabela 4.19: Modelo ARIMA ajustado para os indivíduos do sexo masculino em Guiné Equatorial no período compreendido entre 1985 a 2020

País	Modelo	Parâmetros	Coeficiente	Desvio padrão	t-teste	p-value
Guiné Equatorial	ARIMA(2,2,0)	$\hat{\phi}_1$	0,2943	0,1493	1,9708	0.0569
		$\hat{\phi}_2$	0,4828	0,1570	3,0762	0,0041

Tabela 4.20: Modelo ARIMA ajustado para os indivíduos do sexo feminino em Guiné Equatorial no período compreendido entre 1985 a 2020

Os resultados das tabelas 4.19 e 4.20 indicam que, para a Guiné Equatorial, os modelos ARIMA ajustados mostram que as taxas de mortalidade infantil são influenciadas por períodos anteriores. Entretanto, os modelos ARIMA ajustados para a Guiné Equatorial nos períodos de 1985 a 2020 para os indivíduos do sexo masculino e feminino são representados pela forma ARIMA(2,2,0), o que indica que a série temporal foi diferenciada duas vezes para estabilizar a variância e a média. Para os indivíduos do sexo masculino, a equação correspondente é dada por

$$\Delta^2 X_t = c + 0,4776\Delta X_{t-1} + 0,2784\Delta X_{t-2}$$

enquanto que, para os indivíduos do sexo feminino, a equação assume a forma

$$\Delta^2 X_t = c + 0,2943\Delta X_{t-1} + 0,4828\Delta X_{t-2}$$

Nessas equações, $\Delta^2 X_t$ representa a variação na taxa de mortalidade infantil, sendo influenciada pelas variações nos períodos anteriores. Os coeficientes $\hat{\phi}_1$ e $\hat{\phi}_2$ refletem a relação entre a variação da mortalidade em períodos consecutivos, indicando que mudanças nas taxas passadas impactam significativamente as taxas atuais.

Como mostram as figuras 4.14 e 4.15, as taxas de mortalidade infantil na Guiné Equatorial, para ambos os sexos, ao longo do período de 1985 a 2020, revelam uma redução consistente em ambos os sexos. A análise dos modelos original e ajustado, representados pelas linhas azul e vermelha, respectivamente, indica que o modelo ajustado segue de perto a tendência observada nos dados originais, o que confirma a precisão e a eficácia do ajuste realizado.

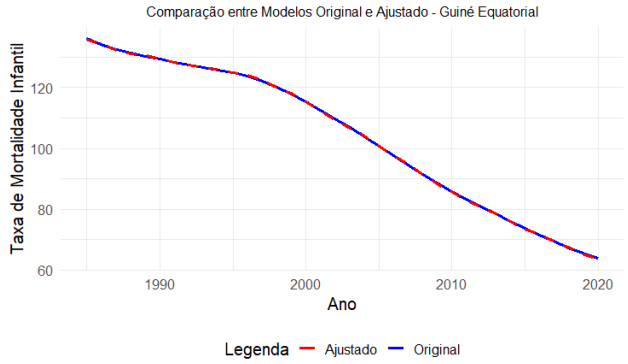

Figura 4.14: Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo masculino - Guiné Equatorial.

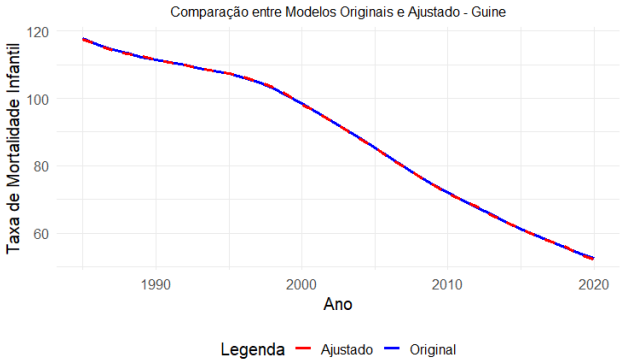

Figura 4.15: Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo feminino - Guiné Equatorial.

Moçambique

País	Modelo	Parâmetros	Coeficiente	Desvio padrão	t-teste	p-value
Moçambique	ARIMA(1,2,0)	$\hat{\phi}_1$	-0.4290	0.1510	-2,8422	0,0074

Tabela 4.21: Modelo ARIMA ajustado para os indivíduos do sexo masculino em Moçambique no período compreendido entre 1985 a 2020

País	Modelo	Parâmetros	Coeficiente	Desvio padrão	t-teste	p-value
Moçambique	ARIMA(1,2,0)	$\hat{\phi}_1$	-0,3912	0,1543	-2,5361	0,0158

Tabela 4.22: Modelo ARIMA ajustado para os indivíduos do sexo feminino em Moçambique no período compreendido entre 1985 a 2020

No caso de Moçambique as tabelas 4.21 e 4.22, mostram que modelo ARIMA ajustado para a mortalidade infantil em Moçambique, tanto para os indivíduos do sexo masculino quanto para os do sexo feminino, revela padrões significativos nas dinâmicas temporais das taxas de mortalidade. Para o sexo masculino, a equação $\Delta^2 X_t = -0,4290$ sugere que um aumento na taxa de mortalidade em um período é seguido por uma diminuição na taxa no período seguinte, com um coeficiente de $-0,4290$, que se mostrou estatisticamente significativo ($p < 0,05$). De maneira semelhante, a análise para o sexo feminino resultou na equação $\Delta^2 X_t = -0,3912$, indica uma relação inversa com um coeficiente de $-0,3912$, também significativo ($p < 0,05$).

Além disso, como mostram as figuras 4.16 e 4.17, as taxas de mortalidade infantil em Moçambique, ao longo do período de 1985 a 2020, apresentam uma redução constante. A comparação entre os modelos original e ajustado revela uma boa correspondência entre os dois, indicando que o modelo ajustado é eficaz em capturar a tendência dos dados reais.

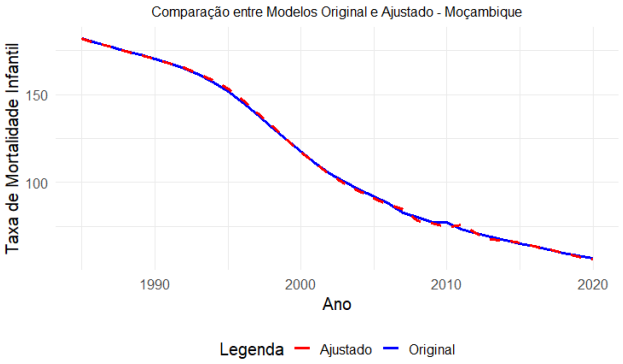

Figura 4.16: Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo masculino - Moçambique.

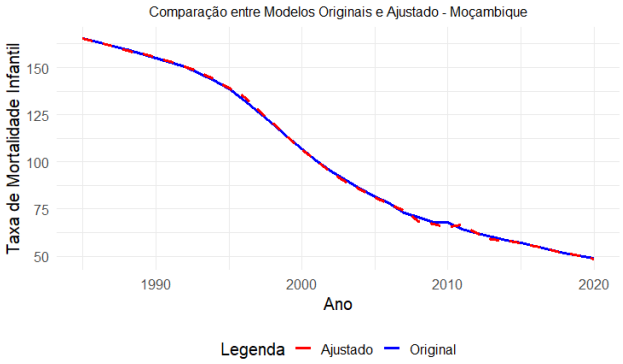

Figura 4.17: Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo feminino - Moçambique.

Portugal

País	Modelo	Parâmetros	Coeficiente	Desvio padrão	t-teste	p-value
Portugal	ARIMA(2,2,0)	$\hat{\phi}_1$	0	0	-	-
		$\hat{\phi}_2$	0,5183	0,1457	3,5580	0.0011

Tabela 4.23: Modelo ARIMA ajustado para os indivíduos do sexo masculino em Portugal no período compreendido entre 1985 a 2020

País	Modelo	Parâmetros	Coeficiente	Desvio padrão	t-teste	p-value
Portugal	ARIMA(1,2,2)	$\hat{\phi}_1$	0,5130	0,2098	2,4449	0,0199
		$\hat{\theta}_1$	-0,8699	0,2206	-3.9432	<0,05
		$\hat{\theta}_2$	0,6970	0,1689	4.1261	<0,05

Tabela 4.24: Modelo ARIMA ajustado para os indivíduos do sexo feminino em Portugal no período compreendido entre 1985 a 2020

Os modelos ARIMA ajustados para Portugal, como se pode verificar nas tabelas 4.23 e 4.24, tanto para indivíduos do sexo masculino quanto do sexo feminino, evidenciam padrões relevantes sobre as dinâmicas temporais dessas taxas.

Para os indivíduos do sexo masculino, a equação do modelo ARIMA(2,2,0) é especificada como $\Delta^2 X_t = 0{,}5183 \cdot \Delta X_{t-2}$, onde o coeficiente $\hat{\phi}_2 = 0{,}5183$ é estatisticamente significativo (p = 0,0011). Este resultado sugere uma relação positiva com a variação nas taxas de mortalidade, o que implica que um aumento em períodos anteriores tende a estar associado a aumentos subsequentes, refletindo possíveis padrões persistentes ao longo do tempo.

Já para os indivíduos do sexo feminino, o modelo ARIMA(1,2,2) é representado pela equação $\Delta^2 X_t = 0{,}5130 \cdot \Delta X_{t-1} - 0{,}8699 \cdot \varepsilon_{t-1} + 0{,}6970 \cdot \varepsilon_{t-2}$. O coeficiente $\hat{\phi}_1 = 0{,}5130$ é também estatisticamente significativo (p = 0,0199), assim como os coeficientes $\hat{\theta}_1 = -0{,}8699$ (p < 0,05) e $\hat{\theta}_2 = 0{,}6970$ (p < 0,05), indicando que a taxa de mortalidade feminina é igualmente influenciada por variações anteriores e por choques provenientes de períodos passados.

Como se observam nas figuras 4.18 e 4.19, as taxas de mortalidade infantil em Portugal, entre 1985 e 2020, apresentam uma redução acentuada e contínua. A comparação entre os modelos original e ajustado revela uma forte correspondência entre ambos, com os dados ajustados a seguirem de perto a tendência dos dados originais. Esta proximidade entre os dois modelos indica que o modelo ajustado é eficaz em representar a evolução das taxas de mortalidade infantil.

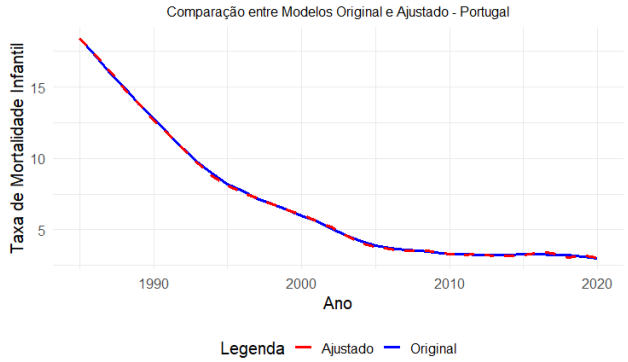

Figura 4.18: Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo masculino Portugal.

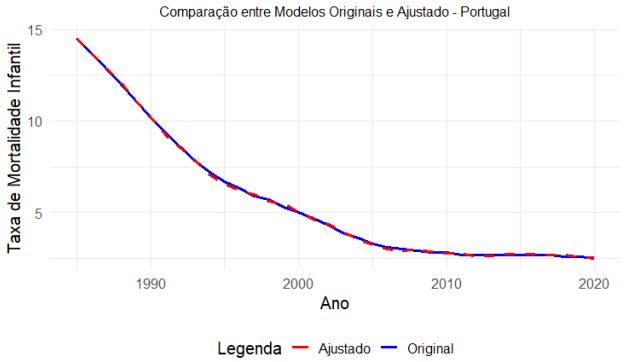

Figura 4.19: Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo feminino Portugal.

São Tomé e Príncipe

País	Modelo	Parâmetros	Coeficiente	Desvio padrão	t-teste	p-value
São Tomé e Príncipe	ARIMA(1,2,3)	$\hat{\phi}_1$	-0,9178	0,1170	-8,8537	$< 0,05$
		$\hat{\theta}_1$	-0,6916	0,1033	-6,6952	0,05
		$\hat{\theta}_2$	1,0000	0,1337	7,4765	$< 0,05$
		$\hat{\theta}_3$	0	0	–	–

Tabela 4.25: Modelo ARIMA ajustado para os indivíduos do sexo masculino em São Tomé e Príncipe no período de 1985 a 2020.

País	Modelo	Parâmetros	Coeficiente	Desvio padrão	t-teste	p-value
São Tomé e Príncipe	ARIMA(1,2,3)	$\hat{\phi}_1$	0	0	–	–
		$\hat{\theta}_1$	0	0	–	–
		$\hat{\theta}_2$	0	0	–	–
		$\hat{\theta}_3$	0,5108	0,1375	3,7150	$< 0,05$

Tabela 4.26: Modelo ARIMA ajustado para os indivíduos do sexo feminino em São Tomé e Príncipe no período de 1985 a 2020.

Entretanto, os resultados dos modelos ARIMA para São Tomé e Príncip(tabelas 4.25 e 4.26), tanto para indivíduos do sexo masculino como do sexo feminino, oferece uma visão detalhada das dinâmicas destas taxas no período entre 1985 e 2020. Para os indivíduos do sexo masculino, o modelo ARIMA(1,2,3) é expresso pela equação:

$$\Delta^2 X_t = -0.9178\Delta X_{t-1} - 0.6916\varepsilon_{t-1} + 1.0000\varepsilon_{t-2} \tag{4.2}$$

O coeficiente $\hat{\phi}_1 = -0.9178$ é estatisticamente significativo ($p < 0.05$), sugerindo que um aumento na taxa de mortalidade infantil no período anterior está associado a uma redução nas taxas atuais, o que pode

indicar uma tendência de autocorrelação negativa. O coeficiente $\hat{\theta}_1 = -0.6916$ também é significativo, indicando que choques passados têm um impacto negativo na taxa atual, enquanto $\hat{\theta}_2 = 1.0000$ implica que choques do segundo período anterior estão positivamente correlacionados com a taxa atual.

Para os indivíduos do sexo feminino, o mesmo modelo ARIMA(1,2,3) apresenta uma equação:

$$\Delta^2 X_t = 0 \cdot \Delta X_{t-1} + 0 \cdot \varepsilon_{t-1} + 0 \cdot \varepsilon_{t-2} + 0,5108 \cdot \varepsilon_{t-3}$$

Como se pode ver apenas $\hat{\theta}_3 = 0.5108$ é significativo ($p < 0.05$). Isso indica que choques passados são relevantes para a previsão da mortalidade infantil feminina, embora não haja influência direta de valores anteriores.

Por outro lado, como ilustrado nas figuras baixo, 4.20 e 4.21, as taxas de mortalidade infantil em São Tomé e Príncipe, para ambos os sexos, entre 1985 e 2020, evidenciam uma tendência de redução contínua. A comparação entre os modelos original e ajustado, representados pelas linhas azul e vermelha, respectivamente, revela uma excelente correspondência entre ambos, o que indica que o modelo ajustado reflete de forma eficaz as flutuações dos dados originais.

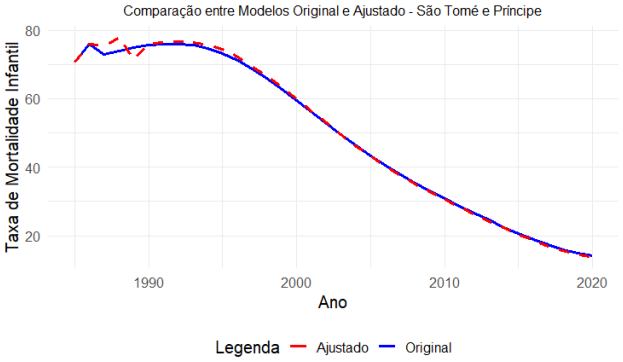

Figura 4.20: Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo masculino - São Tomé e Príncipe.

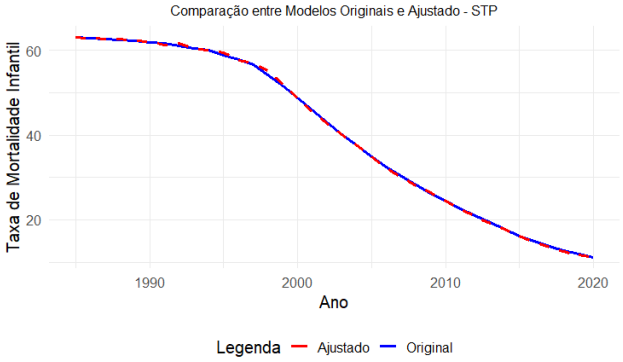

Figura 4.21: Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo feminino - São Tomé e Príncipe.

Timor-Leste

País	Modelo	Parâmetros	Coeficiente	Desvio padrão	t-teste	p-value
Timor-Leste	ARIMA(2,2,0)	$\hat{\phi}_1$	0,3653	0,1445	2,5289	0,0162
		$\hat{\phi}_2$	0,5318	0,1471	3,6164	<0,05

Tabela 4.27: Modelo ARIMA ajustado para os indivíduos do sexo masculino e feminino em Timor-Leste no período compreendido entre 1985 a 2020

A tabela 4.27 apresenta os resultados do modelo ARIMA(2,2,0) ajustado para a mortalidade infantil em Timor-Leste no período de 1985 a 2020, abrangendo tanto indivíduos do sexo masculino quanto feminino. Os parâmetros do modelo revelam que o coeficiente $\hat{\phi}_1 = 0,3653$ indica uma relação positiva significativa entre a variação da mortalidade infantil do período anterior (ΔX_{t-1}) e a variação atual, enquanto $\hat{\phi}_2 = 0,5318$ sugere que a variação da mortalidade infantil do segundo período anterior (ΔX_{t-2}) também exerce um efeito positivo na mortalidade infantil presente. Ambos os coeficientes são estatisticamente

significativos, com valores de t-teste de $2,5289$ e $3,6164$ (p $<$ 0,05), corroborando a relevância dessas influências. Assim, a equação do modelo pode ser expressa como

$$\Delta^2 X_t = 0,3653 \cdot \Delta X_{t-1} + 0,5318 \cdot \Delta X_{t-2}$$

Foi realizada uma comparação entre os modelos original e ajustado das taxas de mortalidade infantil em Timor-Leste, com o objetivo de avaliar a precisão do modelo ajustado e identificar possíveis desvios ou melhorias nas projeções, como se pode ver nas figuras abaixo:

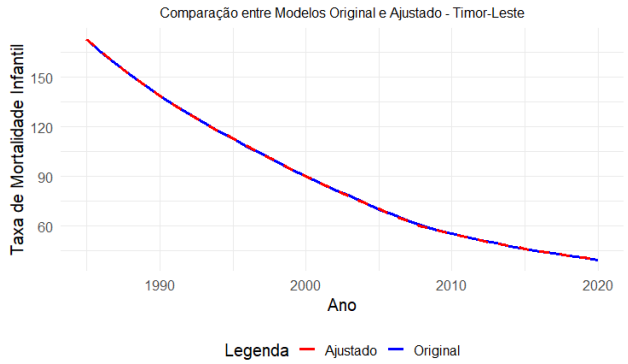

Figura 4.22: Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo masculino - Timor-Leste.

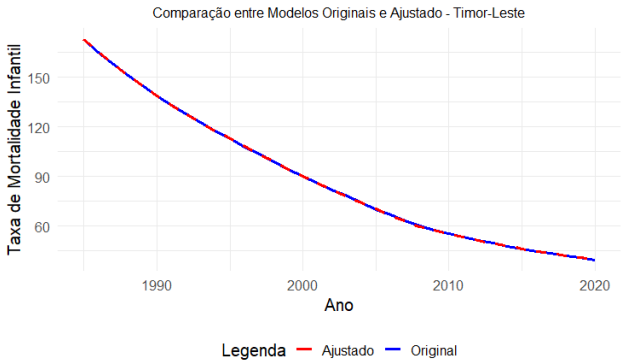

Figura 4.23: Comparação entre o modelo ajustado e o modelo estimado para os indivíduos do sexo feminino - Timor-Leste.

As figuras 4.22 e 4.23, onde a linha azul representa os dados originais e a linha vermelha corresponde aos valores ajustados, observa-se uma redução consistente nas taxas de mortalidade infantil ao longo dos anos. Ambos os modelos original e ajustado seguem uma trajetória semelhante, o que indica que o modelo estimado reflete adequadamente a tendência observada nos dados reais.

No entanto, para avaliar a precisão dos modelos ARIMA ajustados e a sua capacidade preditiva das taxas de mortalidade infantil, foram utilizadas duas métricas de erro amplamente reconhecidas: o Erro Médio Absoluto (MAE) e a Raiz do Erro Quadrático Médio (RMSE). Ambas as métricas permitem aferir a qualidade preditiva dos modelos, complementando a análise estatística dos parâmetros estimados e reforçando a validade dos resultados obtidos. Os valores correspondentes encontram-se sintetizados na tabela 4.28.

País	**Modelo ARIMA**		**RMSE**		**MAE**	
	Masculino	**Feminino**	**Masculino**	**Feminino**	**Masculino**	**Feminino**
Angola	(4,2,1)	(2,2,0)	0,1774	0,1453	0,1791	0,1552
Brasil	(0,2,2)	(3,2,1)	0,3317	0,7776	0,1046	0,0835
Cabo Verde	(2,1,0)	(2,1,0)	0,1555	0,1285	0,1555	0,1285
Guiné-Bissau	(3,1,0)	(3,1,0)	0,1256	0,1022	0,1256	0,1022
Guiné-Equatorial	(2,2,0)	(2,2,0)	0,1491	0,1223	0,1492	0,1226
Moçambique	(1,2,0)	(1,2,0)	0,9666	0,6821	0,8898	0,6269
Portugal	(2,2,0)	(1,2,2)	0,0804	0,0656	0,0723	0,0594
São Tomé e Príncipe	(1,2,3)	(1,2,3)	0,7544	0,4929	0,2685	0,1864
Timor-Leste	(2,2,0)	(2,2,0)	0,1357	0,1023	0,1357	0,1023

Tabela 4.28: Valores das métricas de desempenho dos modelos ajustados para indivíduos do sexo masculino e feminino nos países da CPLP.

A análise dos modelos ARIMA ajustados para os países da CPLP evidencia diferenças assinaláveis na capacidade preditiva em termos de mortalidade infantil. Angola, Cabo Verde, Guiné-Bissau e Timor-Leste apresentam métricas de RMSE e MAE bastante satisfatórias, o que sugere um bom ajustamento dos modelos. Em contrapartida, Moçambique e São Tomé e Príncipe revelam valores consideravelmente elevados, apontando para possíveis limitações na qualidade das previsões. Portugal destaca-se com os melhores valores de acurácia, refletindo a robustez dos modelos ajustados. Importa ainda referir que, de acordo com a análise de resíduos, não foi detetada autocorrelação nos resíduos de Moçambique e São Tomé e Príncipe, reforçando a adequação estatística dos modelos ajustados para esses países.

Por fim, para confirmar a validade do ajustamento dos modelos às séries temporais, foi realizada uma avaliação detalhada dos resíduos, conforme apresentado na secção seguinte.

4.3.3 Testes de diagnóstico aos modelos

Para os testes de diagnóstico aos modelos, recorreu-se aos testes de Ljung–Box e de Lilliefors. Os resultados do teste de Ljung–Box sugerem que os resíduos não apresentam autocorrelação. Já o teste de normalidade de Lilliefors, uma adaptação do teste de Kolmogorov–Smirnov, indica que os resíduos seguem uma distribuição aproximadamente normal, com exceção do Brasil (sexo masculino) e de São Tomé e Príncipe (ambos os sexos). No entanto, os gráficos de ACF e PACF dos resíduos mostram que estes se distribuem aleatoriamente em torno de zero, sugerindo comportamento de ruído branco. Assim, os modelos ajustam-se bem aos dados de mortalidade infantil, sendo considerados adequados para fins preditivos (tabelas 4.29 e 4.30).

País	$\chi^2_{(6)}$		p-**value**	
	Masculino	**Feminino**	**Masculino**	**Feminino**
Angola	4,8443	4,2692	0,5639	0,6403
Brasil	0,7342	1,0176	0,9937	0,9849
Cabo Verde	7,6594	7,6594	0,2641	0,2641
Guiné-Bissau	5,0724	5,0724	0,5346	0,5346
Guiné-Equatorial	3,7312	3,5476	0,7130	0,7376
Moçambique	7,2319	7,4230	0,2999	0,2835
Portugal	1,8170	2,8537	0,9357	0,8270
São Tomé e Príncipe	2,3461	2,9505	0,8853	0,8150
Timor-Leste	4,5347	4,5347	0,6047	0,6047

Tabela 4.29: Resultados dos testes de Ljung-Box para os modelos ajustados dos indivíduos do sexo masculino e feminino nos países da CPLP

.

País	D		p-**value**	
	Masculino	**Feminino**	**Masculino**	**Feminino**
Angola	0,0559	0,1111	0,9951	0,3140
Brasil	0,1985	0,0943	$<$0,005	0,5761
Cabo Verde	0,0780	0,0780	0,8377	0,8377
Guiné-Bissau	0,0945	0,0945	0,5734	0,5734
Guiné-Equatorial	0,0819	0,0998	0,7831	0,4847
Moçambique	0,1399	0,1398	0,0722	0,0727
Portugal	0,1308	0,0955	0,1233	0,5556
São Tomé e Príncipe	0,2399	0,2152	$<$0,05	$<$0,05
Timor-Leste	0,1344	0,1344	0,0980	0,0980

Tabela 4.30: Resultados dos testes de Lilliefors para os modelos ajustados dos indivíduos do sexo masculino e feminino nos países da CPLP.

Por outro lado, com base nos gráficos ACF e PACF dos resíduos (ver Apêndice C), observa-se que os modelos se ajustam bem às séries temporais, uma vez que os resíduos não evidenciam autocorrelação significativa e estão distribuídos aleatoriamente em torno de zero, comportando-se como ruído branco.

4.3.4 Discussão das previsões

As previsões indicam que a taxa de mortalidade infantil continuará a diminuir na maioria dos países da CPLP, seguindo a tendência observada nos dados históricos. As tabelas a seguir apresentam as previsões das taxas de mortalidade infantil dos países da CPLP no período de 2021 a 2025.

País	Ano	Previsão		Observado[2]		Lo 95		Hi 95	
		Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.
Angola	2021	51,55	41,84	45,493	36,058	51,13	41,47	51,96	42,21
	2022	49,64	40,76	44,037	34,929	48,59	39,77	50,68	41,76
	2023	47,55	39,83	42,634	33,824	45,58	37,86	49,53	41,79
	2024	45,29	39,03	–	–	42,04	35,73	48,55	42,32
	2025	42,83	38,34	–	–	37,95	33,35	47,71	43,34

Tabela 4.31: Resultados das previsões de mortalidade infantil para Angola

País	Ano	Previsão		Observado		Lo 95		Hi 95	
		Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.
Brasil	2021	14,20	11,60	14,102	11,237	13,51	11,37	14,88	11,82
	2022	13,96	11,56	13,967	11,140	12,87	11,01	15,04	12,12
	2023	13,72	11,63	13,810	11,041	12,01	10,64	15,43	12,63
	2024	13,48	11,67	–	–	10,99	10,06	15,97	13,28
	2025	13,24	11,75	–	–	9,85	9,45	16,64	14,06

Tabela 4.32: Resultados das previsões de mortalidade infantil para Brasil

País	Ano	Previsão		Observado		Lo 95		Hi 95	
		Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.
Cabo Verde	2021	12,66	12,66	13,052	10,827	12,34	12,34	12,98	12,98
	2022	11,77	11,77	12,510	10,362	10,83	10,83	12,71	12,71
	2023	10,64	10,64	12,025	9,958	8,82	8,82	12,46	12,46
	2024	9,25	9,25	–	–	6,38	6,38	12,11	12,11
	2025	7,61	7,61	–	–	3,66	3,66	11,56	11,56

Tabela 4.33: Resultados das previsões de mortalidade infantil para Cabo Verde

[2] À data da modelação, os valores observados entre 2021 e 2023 ainda não se encontravam disponíveis. Após a sua publicação oficial, optou-se por integrá-los nas tabelas comparativas para todos os países, permitindo uma leitura mais crítica da proximidade entre os valores previstos e os efetivamente observados.

País	Ano	Previsão		Observado		Lo 95		Hi 95	
		Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.
Guiné-Bissau	2021	55,48	55,48	50,63	40,453	55,22	55,22	55,74	55,74
	2022	54,12	54,12	49,276	39,128	53,46	53,46	54,78	54,78
	2023	52,73	52,73	47,759	38,009	51,47	51,47	53,99	53,99
	2024	51,26	51,26	—	—	49,25	49,25	53,28	53,28
	2025	49,67	49,67	—	—	46,76	46,76	52,59	52,59

Tabela 4.34: Resultados das previsões de mortalidade infantil para Guiné-Bissau

País	Ano	Previsão		Observado		Lo 95		Hi 95	
		Masc.	Fem.	Masc.	Fem.	Masc.		Masc.	Fem.
Guiné Equatorial	2021	62.15	51.26	57.683	47.096	61.84	50.95	62.46	51.57
	2022	60.73	50.05	55.880	45.487	59.90	49.28	61.56	50.83
	2023	59.41	48.97	54.196	43.976	57.80	47.46	61.02	50.47
	2024	58.18	47.99	—	—	55.52	45.51	60.84	50.46
	2025	57.02	47.09	—	—	53.05	43.39	60.98	50.79

Tabela 4.35: Resultados das previsões de mortalidade infantil para Guiné Equatorial

País	Ano	Previsão		Observado		Lo 95		Hi 95	
		Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.
Moçambique	2021	55,11	47,22	51,53	43,739	53,14	45,40	57,09	49,04
	2022	53,57	45,77	50,231	42,519	49,88	42,32	57,25	49,22
	2023	52,00	44,31	49,052	41,478	46,10	38,79	57,90	49,84
	2024	50,44	42,86	—	—	42,07	34,99	58,81	50,73
	2025	48,88	41,40	—	—	37,75	30,93	60,02	51,88

Tabela 4.36: Resultados das previsões de mortalidade infantil para Moçambique

País	Ano	Previsão		Observado		Lo 95		Hi 95	
		Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.
Portugal	2021	2.85	2.43	2.90	2.415	2.68	2.28	3.01	2.58
	2022	2.70	2.36	2.836	2.363	2.33	2.06	3.06	2.65
	2023	2.52	2.28	2.787	2.319	1.83	1.76	3.20	2.80
	2024	2.34	2.20	—	—	1.26	1.37	3.42	3.03
	2025	2.15	2.12	—	—	0.59	0.91	3.70	3.33

Tabela 4.37: Resultados das previsões de mortalidade infantil para Portugal

País	Ano	Previsão		Observado		Lo 95		Hi 95	
		Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.
São Tomé e Príncipe	2021	13.10	10.60	11.347	9.072	10.91	10.05	15.30	11.14
	2022	12.20	10.05	10.874	8.692	9.03	8.83	15.37	11.28
	2023	11.21	9.54	10.419	8.359	5.80	7.48	16.62	11.60
	2024	10.30	9.03	—	—	3.21	5.80	17.39	12.25
	2025	9.31	8.51	—	—	-0.37	3.89	18.99	13.13

Tabela 4.38: Resultados das previsões de mortalidade infantil para São Tomé e Príncipe

País	Ano	Previsão		Observado		Lo 95		Hi 95	
		Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.
Timor-Leste	2021	38.77	38.77	41.676	33.988	38.49	38.49	39.06	39.06
	2022	37.88	37.88	40.579	33.005	37.15	37.15	38.60	38.60
	2023	37.07	37.07	39.581	32.159	35.64	35.64	38.51	38.51
	2024	36.37	36.37	—	—	33.96	33.96	38.78	38.78
	2025	35.75	35.75	—	—	32.07	32.07	39.43	39.43

Tabela 4.39: Resultados das previsões de mortalidade infantil para Timor-Leste

A análise das previsões de mortalidade infantil para diferentes países da CPLP revela variações significativas entre as taxas para ambos os sexos ao longo dos anos. Em países como Angola e Moçambique, as previsões mostram uma redução contínua nas taxas de mortalidade para ambos os sexos, embora os níveis ainda sejam elevados em comparação com outros países. No caso de Angola, por exemplo, as previsões para 2025 indicam uma mortalidade de 42,83 para rapazes e 38,34 para raparigas, com uma diminuição expressiva em relação aos anos anteriores.

No Brasil, embora as taxas de mortalidade sejam mais baixas, observa-se também uma tendência de decréscimo gradual, com previsões de 13,24 para rapazes e 11,75 para raparigas em 2025. Já em Cabo Verde e São Tomé e Príncipe, a redução é ainda mais acentuada, com Cabo Verde a projetar uma queda para 7,61 para ambos os sexos em 2025. Países como Portugal apresentam as menores taxas, com uma previsão de 2,15 para rapazes e 2,12 para raparigas em 2025. Contudo, São Tomé e Príncipe destaca-se pela maior amplitude nos intervalos de confiança, sugerindo maior incerteza nas previsões, especialmente no caso dos rapazes, cujas previsões para 2025 variam entre -0,37 e 18,99.

A integração dos valores observados entre 2021 e 2023 permitiu uma comparação direta com as previsões geradas pelos modelos ARIMA ajustados para cada país. De forma geral, observa-se uma boa proximidade entre os valores previstos e os efetivamente registados, com variações dentro dos intervalos de confiança estabelecidos. No caso de países como o Brasil e Portugal, os desvios foram mínimos, reforçando a capacidade preditiva dos modelos. Em contrapartida, em contextos como o da Guiné-Bissau, notaram-se discrepâncias mais acentuadas, particularmente entre os dados femininos, o que poderá indicar a presença de fatores exógenos não captados pelo modelo ou alterações estruturais recentes nas condições de saúde pública.

Essa análise permite identificar tendências de redução global na mortalidade infantil, mas também revela desafios persistentes nos países com maiores taxas, como Angola, Guiné Equatorial e Guiné-Bissau. No geral, todas as regiões apresentam uma tendência positiva de redução das taxas de mortalidade infantil, o que talvez sugere avanços nas intervenções de saúde pública e melhorias das condições de vida em alguns

casos.

4.3.5 Análise de desempenho com divisão treino-teste

Na análise das taxas de mortalidade infantil, a divisão dos dados em conjuntos de treino e teste é realizada de forma a otimizar a precisão e a eficácia de precisão do modelo. O conjunto de treino é composto pelas observações de 1985 a 2015, um período que fornece uma base de dados históricos para o ajuste do modelo. Este intervalo permite ao modelo identificar padrões e tendências ao longo de três décadas, aproveitando as variações e flutuações nas taxas de mortalidade infantil, influenciadas por diversos fatores socioeconómicos e de saúde pública.

Em contrapartida, o conjunto de teste abrange as observações de 2016 a 2020, um período mais recente que permite avaliar a capacidade do modelo em prever taxas de mortalidade infantil em dados não vistos anteriormente. Essa abordagem assegura que as previsões do modelo sejam testadas em um contexto temporal diferente, oferecendo uma avaliação da sua eficácia em capturar tendências emergentes e mudanças nas dinâmicas de mortalidade infantil. Ao dividir os dados dessa forma, conseguimos garantir que as previsões não apenas refletem os padrões históricos, mas também possuem relevância e aplicabilidade no cenário atual.

Portanto, as tabelas abaixo mostram os resultados de acurácia (RMSE e MAE) para o conjunto de treino e teste dos países da CPLP em estudo.

Angola	**RMSE**	**MAE**
Masculino - Treino	0,1812	0,1478
Masculino - Teste	1,0478	0,7599
Feminino - Treino	0,1744	0,1511
Feminino - Teste	2,5226	2,0371

Tabela 4.40: Resultados de RMSE e MAE para Angola de acordo com a previsão de 2016-2020

Brasil	**RMSE**	**MAE**
Masculino - Treino	0,1194	0,1031
Masculino - Teste	0,6552	0,5802
Feminino - Treino	0,1138	0,0931
Feminino - Teste	0,4309	0,4103

Tabela 4.41: Resultados de RMSE e MAE para o Brasil de acordo com a previsão de 2016-2020

Cabo Verde	**RMSE**	**MAE**
Masculino - Treino	0,1553	0,1285
Masculino - Teste	1,8532	1,5081
Feminino - Treino	0,1553	0,1285
Feminino - Teste	1,8532	1,5081

Tabela 4.42: Resultados de RMSE e MAE para Cabo Verde de acordo com a previsão de 2016-2020

Guiné-Bissau	RMSE	MAE
Masculino - Treino	0,1137	0,0920
Masculino - Teste	1,5742	1,2574
Feminino - Treino	0,1137	0,0920
Feminino - Teste	1,5742	1,2574

Tabela 4.43: Resultados de RMSE e MAE para Guiné-Bissau de acordo com a previsão de 2016-2020

Guiné Equatorial	RMSE	MAE
Masculino - Treino	0,1546	0,1264
Masculino - Teste	0,5934	0,5338
Feminino - Treino	0,1474	0,1238
Feminino - Teste	0,5397	0,4600

Tabela 4.44: Resultados de RMSE e MAE para Guiné Equatorial de acordo com a previsão de 2016-2020

Moçambique	RMSE	MAE
Masculino - Treino	1,0403	0,7732
Masculino - Teste	0,2359	0,1692
Feminino - Treino	0,9569	0,7056
Feminino - Teste	0,2354	0,2065

Tabela 4.45: Resultados de RMSE e MAE para Moçambique de acordo com a previsão de 2016-2020

Portugal	RMSE	MAE
Masculino - Treino	0,0731	0,0598
Masculino - Teste	0,6914	0,5904
Feminino - Treino	0,0750	0,0616
Feminino - Teste	0,1598	0,1336

Tabela 4.46: Resultados de RMSE e MAE para Portugal de acordo com a previsão de 2016-2020

São Tomé e Príncipe	RMSE	MAE
Masculino - Treino	1,1438	0,6794
Masculino - Teste	1,8836	1,5338
Feminino - Treino	0,2849	0,1955
Feminino - Teste	1,2438	0,9998

Tabela 4.47: Resultados de RMSE e MAE para São Tomé e Príncipe de acordo com a previsão de 2016--2020

Timor-Leste	RMSE	MAE
Masculino - Treino	0,1426	0,1083
Masculino - Teste	0,3890	0,3252
Feminino - Treino	0,1426	0,1083
Feminino - Teste	0,3890	0,3252

Tabela 4.48: Resultados de RMSE e MAE para Timor-Leste de acordo com a previsão de 2016-2020

A análise global dos resultados de RMSE e MAE para os países em estudo, considerando as previsões de mortalidade infantil entre 2016 e 2020, revela importantes pontos fortes no desempenho dos modelos ajustados. Em primeiro lugar, a capacidade preditiva dos modelos em países como Brasil, Portugal e Timor--Leste se destacam, ao apresentar os menores valores de RMSE e MAE, o que indica uma precisão notável nas previsões de mortalidade infantil. Além disso, o desempenho dos modelos em Angola e Moçambique, embora evidencie algumas limitações, como um RMSE elevado para o grupo feminino, os testes aos resíduos, que indicam a presença de ruído branco, são encorajadores, pois sugerem que os modelos capturam a aleatoriedade dos dados.

Além disso testou-se a capacidade de previsão do modelo e obteve-se os seguintes resultados, para todos os países da CPLP:

País	Ano	Previsão		Observado[3]		Lo 95		Hi 95	
		Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.
Angola	2016	60,9929	50,0076	60,9	49,7	60,5985	49,6413	61,3873	50,3739
	2017	59,0094	48,7153	58,9	47,9	58,0229	47,7248	59,9959	49,7059
	2018	57,4258	47,8825	56,9	46,2	55,5809	45,9001	59,2707	49,8648
	2019	56,2485	47,4852	55,2	44,5	53,2243	44,1248	59,2727	50,8457
	2020	55,4231	47,4950	53,4	43,1	50,9212	42,3400	59,9251	52,6501

Tabela 4.49: revisão, valores observados e intervalos de confiança (Lo 95 e Hi 95) para Angola (2016–2020)

País	Ano	Previsão		Observado		Lo 95		Hi 95	
		Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.
Brasil	2016	15,5123	12,6610	16,6	13,3	15,2661	12,4262	15,7584	12,8957
	2017	15,1160	12,5219	15,3	12,1	14,5656	11,9970	15,6665	13,0469
	2018	14,7198	12,3292	15,1	11,9	13,7039	11,4508	15,7357	13,2076
	2019	14,3236	12,2229	14,9	11,7	12,7396	10,8424	15,9076	13,6034
	2020	13,9274	12,1165	14,6	11,6	11,6923	10,1320	16,1625	14,1011

Tabela 4.50: Previsão, valores observados e intervalos de confiança (Lo 95 e Hi 95) para o Brasil (2016–2020)

[3]Os valores observados correspondem ao período de teste (2016–2020), utilizado para validar os modelos ajustados com dados de 1985 a 2015.

País	Ano	Previsão		Observado		Lo 95		Hi 95	
		Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.
Cabo Verde	2016	16,7950	16,4688	17	14	16,4688	16,4688	17,1212	17,1212
	2017	15,2006	14,2513	15,8	13	14,2513	14,2513	16,1498	16,1498
	2018	13,5487	11,7152	14,9	12,2	11,7152	11,7152	15,3822	15,3822
	2019	11,8799	9,0019	14,1	11,6	9,0019	9,0019	14,7580	14,7580
	2020	10,2353	6,2774	13,4	11	6,2774	6,2774	14,1931	14,1931

Tabela 4.51: Previsão, valores observados e intervalos de confiança (Lo 95 e Hi 95) para Cabo Verde (2016–2020)

País	Ano	Previsão		Observado		Lo 95		Hi 95	
		Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.
Guiné-Bissau	2016	63,8662	63,8662	64,1	51,8	63,6275	63,6275	64,1049	64,1049
	2017	61,5909	61,5909	62,1	50,2	61,0013	61,0013	62,1805	62,1805
	2018	59,2218	59,2218	60,2	48,7	58,1315	58,1315	60,3121	60,3121
	2019	56,7073	56,7073	58,4	47,2	55,0154	55,0154	58,3991	58,3991
	2020	54,0268	54,0268	56,9	45,8	51,6688	51,6688	56,3849	56,3849

Tabela 4.52: Previsão, valores observados e intervalos de confiança (Lo 95 e Hi 95) para Guiné-Bissau (2016–2020)

País	Ano	Previsão		Observado		Lo 95		Hi 95	
		Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.
Guiné Equatorial	2016	71,6446	59,3000	71,5	59,4	71,3198	58,9959	71,9694	59,6040
	2017	69,7262	57,3666	69,4	57,7	68,8377	56,6868	70,6148	58,0464
	2018	67,9106	55,4333	67,3	55,8	66,1766	54,1277	69,6447	56,7389
	2019	66,1777	53,5444	65,4	54,1	63,3191	51,4627	69,0363	55,6260
	2020	64,5098	51,6555	63,7	52,6	60,2534	48,5806	68,7661	54,7303

Tabela 4.53: Previsão, valores observados e intervalos de confiança (Lo 95 e Hi 95) para Guiné Equatorial (2016–2020)

País	Ano	Previsão		Observado		Lo 95		Hi 95	
		Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.
Moçambique	2016	63,5856	55,2782	63,5	55,1	61,4402	53,3049	65,7309	57,2515
	2017	61,7345	53,6258	61,6	53,3	57,7372	49,8874	65,7319	57,3642
	2018	59,8991	51,9854	59,9	51,7	53,5031	45,9966	66,2952	57,9742
	2019	58,0571	50,3403	58,2	50,1	48,9768	41,8174	67,1373	58,8632
	2020	56,2178	48,6970	56,7	48,7	44,1364	37,3468	68,2993	60,0472

Tabela 4.54: Previsão, valores observados e intervalos de confiança (Lo 95 e Hi 95) para Moçambique (2016–2020)

País	Ano	Previsão		Observado		Lo 95		Hi 95	
		Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.
Portugal	2016	3,4000	2,7293	3,3	2,7	3,2492	2,5688	3,5508	2,8898
	2017	3,5632	2,7467	3,2	2,7	3,2260	2,4342	3,9004	3,0591
	2018	3,7264	2,7575	3,2	2,6	3,0832	2,2048	4,3697	3,3103
	2019	3,9296	2,7647	3,1	2,6	2,9078	1,8815	4,9514	3,6478
	2020	4,1327	2,7698	3,0	2,5	2,6312	1,4751	5,6343	4,0644

Tabela 4.55: Previsão, valores observados e intervalos de confiança (Lo 95 e Hi 95) para Portugal (2016–2020)

País	Ano	Previsão		Observado		Lo 95		Hi 95	
		Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.
São Tomé e Príncipe	2016	18,6051	14,7536	18,9	14,9	16,2463	14,1659	20,9640	15,3412
	2017	16,7003	13,3078	17,3	13,7	13,2776	11,9937	20,1229	14,6219
	2018	14,7100	11,8771	16,0	12,7	8,8765	9,6783	20,5436	14,0760
	2019	12,8008	10,4465	15,0	11,9	5,1474	7,0086	20,4541	13,8844
	2020	10,8147	9,0158	14,1	11,2	0,3788	4,0937	21,2506	13,9380

Tabela 4.56: Previsão, valores observados e intervalos de confiança (Lo 95 e Hi 95) para São Tomé e Príncipe (2016–2020)

País	Ano	Previsão		Observado		Lo 95		Hi 95	
		Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.
Timor-Leste	2016	44,8373	44,8373	44,9	37,7	44,5378	44,5378	45,1369	45,1369
	2017	43,3404	43,3404	43,5	36,5	42,5690	42,5690	44,1119	44,1119
	2018	41,8874	41,8874	42,2	35,2	40,3592	40,3592	43,4157	43,4157
	2019	40,4849	40,4849	40,9	34,1	37,9206	37,9206	43,0491	43,0491
	2020	39,1239	39,1239	39,8	33,0	35,2151	35,2151	43,0327	43,0327

Tabela 4.57: Previsão, valores observados e intervalos de confiança (Lo 95 e Hi 95) para Timor-Leste (2016–2020)

Entre os anos de 2016 e 2020, foi possível comparar as previsões geradas pelos modelos ARIMA com os valores reais observados da mortalidade infantil nos países da CPLP. Estes cinco anos correspondem ao período de teste, utilizado para validar a capacidade preditiva dos modelos ajustados com base em dados históricos de 1985 a 2015. De forma geral, observa-se uma boa proximidade entre os valores previstos e os observados, com a maioria das observações a situar-se dentro dos intervalos de confiança calculados (Lo 95 e Hi 95). Esta coerência reforça a fiabilidade dos modelos utilizados, embora em alguns casos sobretudo nos países com maior variabilidade ou contextos mais instáveis surjam pequenas discrepâncias que também refletem a complexidade inerente aos fenómenos demográficos. O alinhamento geral entre previsão e realidade sugere que os modelos conseguem capturar tendências significativas, mesmo num contexto social e económico diverso como o da comunidade lusófona.

Em geral, as previsões de mortalidade infantil para os países da CPLP entre 2016 e 2020 indicam uma tendência geral de redução, com diferenças significativas na magnitude e na incerteza das projeções entre os países. Angola, Guiné Equatorial e Moçambique mostram uma diminuição contínua nas taxas de mortalidade para ambos os sexos, embora as previsões para os anos mais recentes apresentem maior

incerteza, evidenciada pelos intervalos de confiança mais amplos. No Brasil, observa-se uma redução mais moderada e estável, com previsões que demonstram maior precisão. Cabo Verde destaca-se com uma queda acentuada, o que sugere melhorias substanciais nas condições de saúde infantil. Na Guiné-Bissau, a redução é mais lenta e acompanhada de maior incerteza, sobretudo no final do período. Em São Tomé e Príncipe, a tendência de queda é evidente, especialmente entre os rapazes, mas a incerteza é elevada, particularmente em 2020, quando o intervalo de confiança para as raparigas aproxima-se de níveis muito baixos, apontando desafios na precisão das previsões para este grupo. Por outro lado, Timor-Leste apresenta uma diminuição mais gradual nas taxas, com intervalos de confiança consistentes, a refletir maior estabilidade nas previsões. Finalmente, Portugal exibe as taxas mais baixas de mortalidade infantil dentro da CPLP, com pequenas variações, incluindo um ligeiro aumento previsto para o sexo masculino, o que sugere que o país já alcançou um elevado nível de controlo da mortalidade infantil, com flutuações mínimas nas previsões.

Além disso, pode observar-se no Apêndice D, o conjunto de gráficos de previsão dos dados de treino e teste, que as previsões, representadas pela linha verde, mostram uma continuidade dessa tendência de redução, enquanto o intervalo de confiança (representado pela área sombreada) indica a margem de incerteza de 95% associada a essas projeções, para todos os países da CPLP.

5

Considerações finais

A análise das taxas de mortalidade infantil nos países da Comunidade dos Países de Língua Portuguesa entre 1985 e 2020 revelou um panorama complexo e multifacetado, qo refletir as desigualdades socioeconómicas e os diferentes contextos políticos e sociais em que essas nações estão inseridas. De maneira geral, observou-se uma redução significativa das taxas de mortalitadade infantil em todos os países da CPLP no período em análise. No entanto, ainda assim os dados recolhidos evidenciam que, enquanto alguns países conseguiram reduzir significativamente as taxas de mortalidade infantil, outros ainda permanecem com índices elevados, o que põe em evidência a necessidade de uma atenção contínua e de esforços direcionados para enfrentar essas disparidades.

Países como Brasil e Portugal, onde políticas de saúde pública são mais eficazes e com maior acessibilidade a serviços de saúde por parte da população, registaram as menores taxas de mortalidade infantil. Já países como Angola e Moçambique ainda apresentam uma abordagem crítica, uma vez que as taxas de mortalidade infantil continuam a ser muito elevadas, refletindo, em parte, a falta de infraestrutura adequada e a persistência de problemas socioeconómicos graves, como a pobreza e a insegurança alimentar.

Além disso, é importante destacar as diferenças nas taxas de mortalidade entre os sexos. Os dados indicam que, em muitos países da CPLP, as taxas de mortalidade infantil para meninos são geralmente mais altas do que para meninas. Esta tendência, além de observada empiricamente, encontra respaldo na literatura

científica. Segundo Pongou (2012), tal disparidade resulta não apenas de fatores biológicos, como a maior vulnerabilidade imunológica masculina, mas também de influências ambientais anteriores à conceção, que explicam uma parte significativa desta diferença. Este enquadramento reforça a necessidade de políticas de saúde sensíveis ao género, que considerem os múltiplos determinantes da mortalidade infantil. Para além disso, impõe-se o desenvolvimento de mais estudos orientados à compreensão aprofundada das causas contextuais e estruturais desta disparidade.

As intervenções de saúde pública desempenham um papel crucial na redução das taxas de mortalidade infantil. Estudando a eficácia de programas de vacinação, nutrição infantil e educação em saúde, é possível identificar as práticas bem-sucedidas que podem ser replicadas em contextos similares. A experiência de países como Portugal e Brasil, que implementaram programas abrangentes de saúde infantil, pode servir como modelo para outras nações da CPLP que ainda enfrentam desafios significativos.

A modelação preditiva e o uso de técnicas estatísticas avançadas também são ferramentas valiosas para compreender melhor os padrões de mortalidade infantil. A aplicação de modelos de séries temporais, como ARIMA, pode revelar tendências e ciclos, permitindo que os formuladores de políticas antecipem as necessidades de saúde pública e implementem intervenções proativas. A análise das quebras estruturais nas séries temporais também pode fornecer informações relevantes sobre a eficácia de mudanças nas políticas de saúde.

Em relação às políticas de saúde pública, a colaboração entre os países da CPLP é vital. A troca de conhecimentos e experiências pode contribuir para o desenvolvimento de estratégias mais eficazes e adaptadas às necessidades locais. A criação de redes de colaboração entre instituições de pesquisa, organizações não governamentais e governos pode facilitar a disseminação de melhores práticas e fomentar uma abordagem coletiva para enfrentar os desafios da mortalidade infantil.

Além disso, é essencial incluir as vozes das comunidades afetadas nas pesquisas sobre mortalidade infantil. Estudos qualitativos que capturem as percepções e experiências das famílias podem revelar barreiras que não são evidentes nas análises quantitativas. A participação da comunidade no desenvolvimento e na implementação de políticas de saúde pode aumentar a eficácia das intervenções e garantir que atendam às necessidades reais da população.

Outra área a ser explorada em pesquisas futuras é o impacto das condições socioeconómicas e ambientais nas taxas de mortalidade infantil. A relação entre fatores como a pobreza, a educação, o acesso a serviços de saúde e a qualidade da água e do saneamento precisa ser melhor compreendida. A interseccionalidade desses fatores pode ajudar a identificar as populações mais vulneráveis e a desenvolver políticas direcionadas para atendê-las.

É recomendável que, em pesquisas futuras, sejam exploradas outras técnicas de modelação e previsão, como métodos de *machine learning*, aplicadas aos dados, bem como a utilização de uma linguagem de programação alternativa para comparar com os resultados obtidos neste estudo. Além disso, seria proveitoso aprofundar a análise das mudanças estruturais observadas ao longo das séries temporais, com o objetivo de compreender melhor as intervenções realizadas que levaram à redução acentuada nesses períodos.

A questão da mortalidade infantil também não pode ser dissociada do contexto global mais amplo. A pandemia de COVID-19, por exemplo, teve um impacto significativo em muitos sistemas de saúde, exacerbando as desigualdades existentes. A análise de como eventos globais afetam as taxas de mortalidade infantil deve ser uma prioridade em pesquisas futuras, pois as lições aprendidas podem informar estratégias de resiliência e preparação para crises futuras.

Além disso, a consideração das influências culturais e sociais é fundamental para entender as taxas de mortalidade infantil. Estudos que investiguem a percepção da saúde e da mortalidade infantil em diferentes

contextos culturais podem proporcionar uma compreensão mais profunda dos comportamentos e atitudes que afetam a saúde infantil. Essa abordagem pode ajudar a desenvolver campanhas de sensibilização e educação mais eficazes.

Os dados apresentados neste estudo devem servir como uma base para futuras investigações sobre mortalidade infantil na CPLP. A recolha contínua e a análise de dados são essenciais para monitorar o progresso e identificar áreas que requerem atenção urgente.

Por fim, a promoção de uma abordagem multidisciplinar na pesquisa sobre mortalidade infantil é essencial. A colaboração entre profissionais de saúde, sociólogos, economistas e decisores políticos podem enriquecer a compreensão dos fatores que influenciam a mortalidade infantil e contribuir para a formulação de políticas mais eficazes. A pesquisa deve ser guiada pela necessidade de garantir que todas as crianças, independentemente de onde nasçam, tenham a oportunidade de crescer e prosperar em um ambiente saudável e seguro.

Em suma, a mortalidade infantil na CPLP é uma questão complexa que requer uma abordagem integrada e colaborativa. O entendimento aprofundado das tendências, determinantes e contextos locais é crucial para a formulação de intervenções eficazes. O futuro da saúde infantil nos países da CPLP depende de um compromisso contínuo com a pesquisa, a ação colaborativa e a inclusão das vozes das comunidades na construção de soluções que visem à equidade e à melhoria das condições de vida de todas as crianças.

6

Apêndice

A

Gráficos ACF e PACF dos dados transformados dos países da CPLP para a modelação

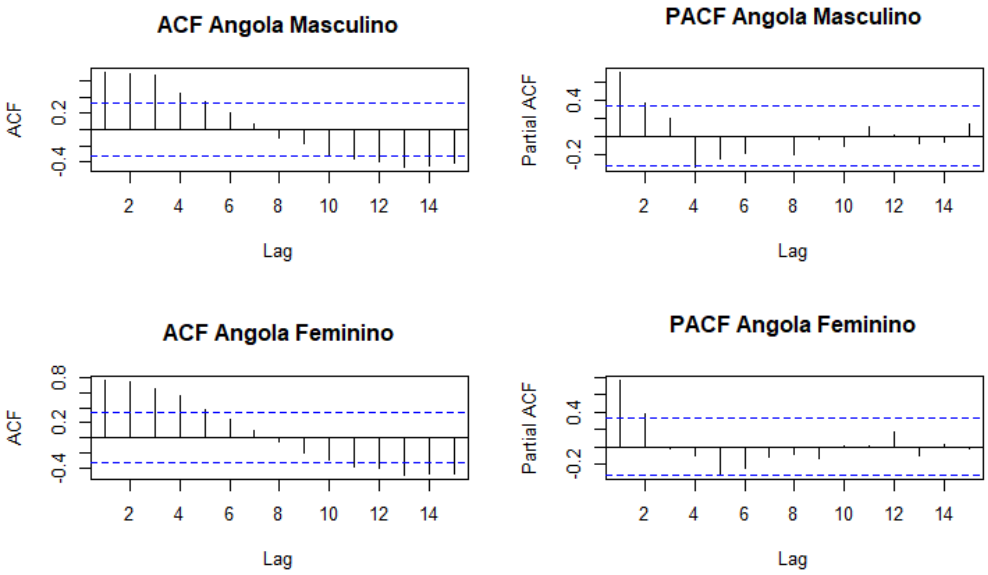

Figura A.1: Gráficos de ACF e PACF de Angola para ambos os géneros

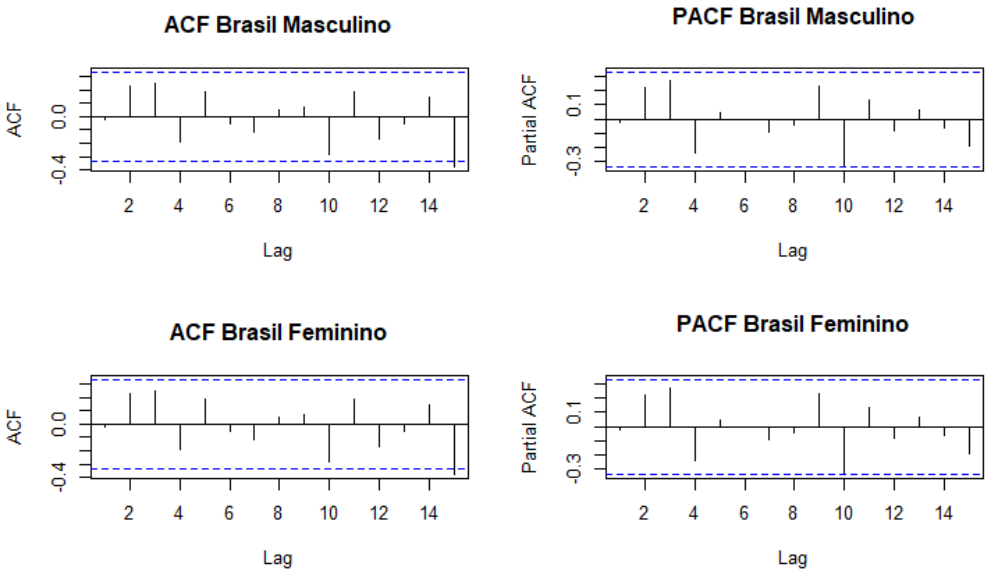

Figura A.2: Gráficos de ACF e PACF do Brasil para ambos os géneros

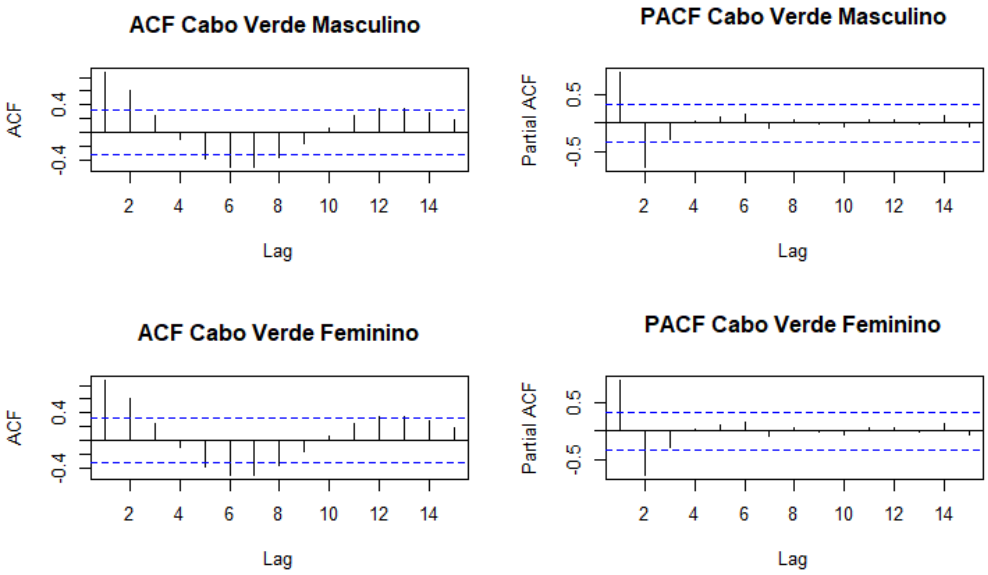

Figura A.3: Gráficos de ACF e PACF de Cabo Verde para ambos os géneros

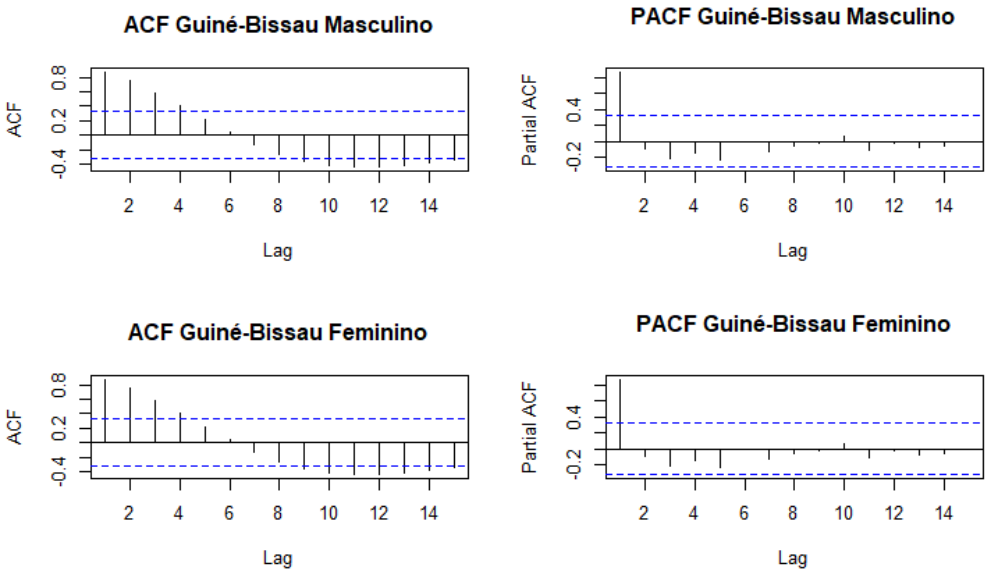

Figura A.4: Gráficos de ACF e PACF de Guiné-Bissau para ambos os géneros

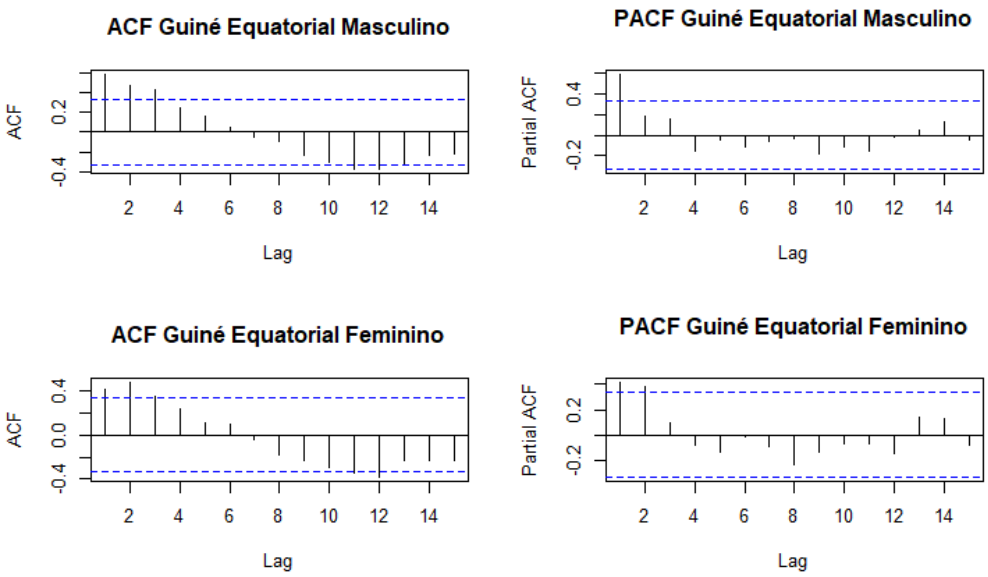

Figura A.5: Gráficos de ACF e PACF de Guiné Equatorial para ambos os géneros

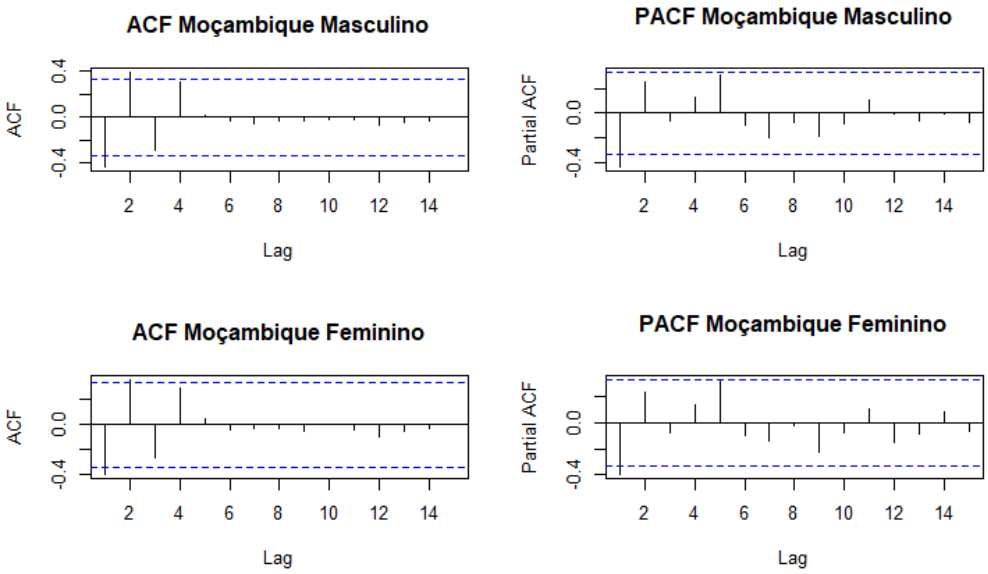

Figura A.6: Gráficos de ACF e PACF de Moçambique para ambos os géneros

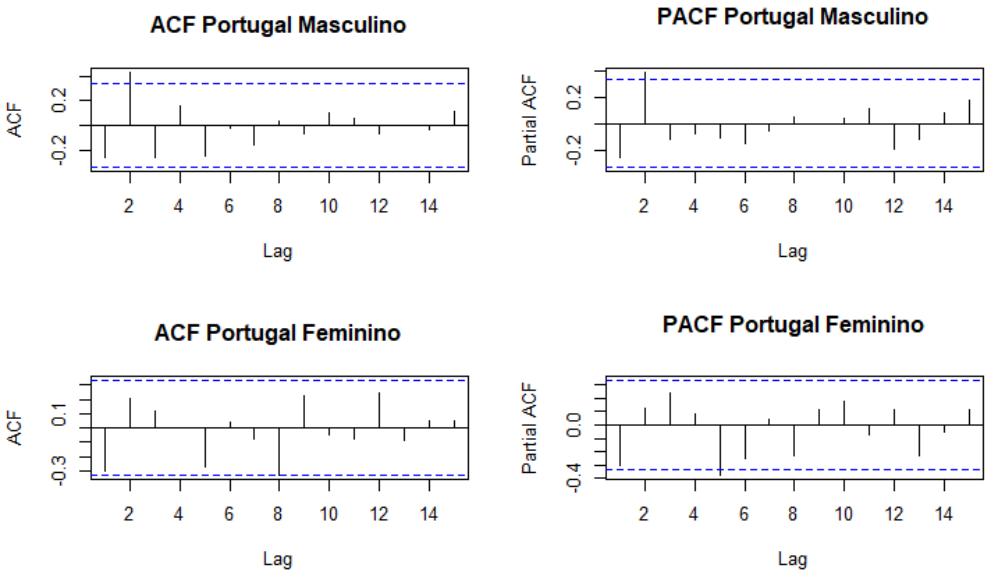

Figura A.7: Gráficos de ACF e PACF de Portugal para ambos os géneros

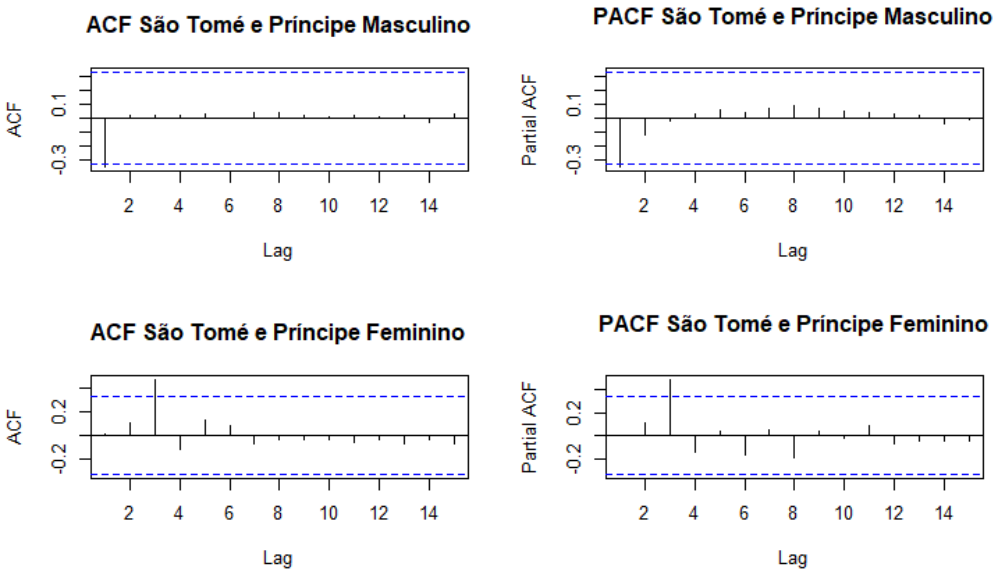

Figura A.8: Gráficos de ACF e PACF de São Tomé e Principé para ambos os géneros

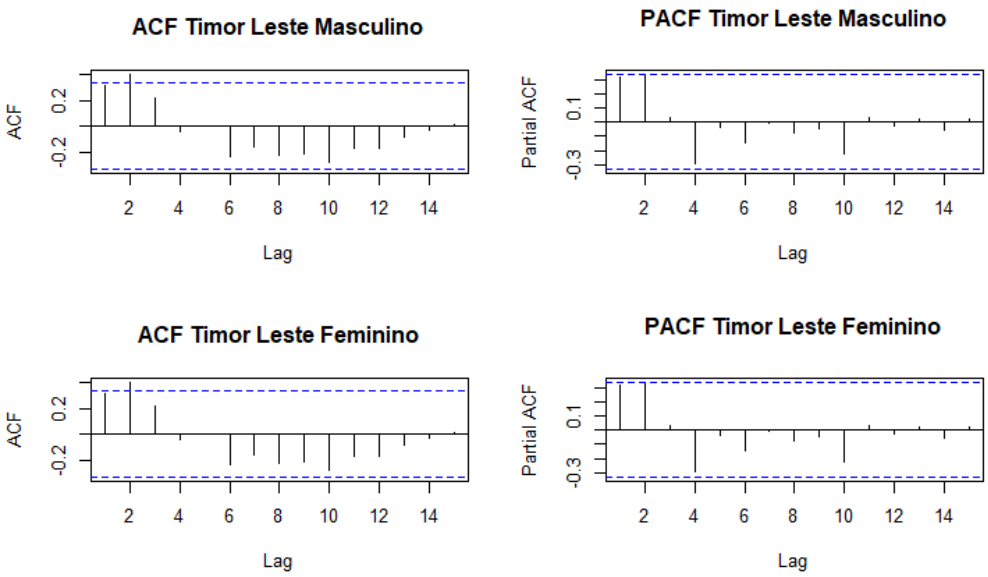

Figura A.9: Gráficos de ACF e PACF de Timor-Leste para ambos os géneros

B

Tabelas

Modelo ARIMA		AIC - Antes		AIC - Depois	
Masculino	**Feminino**	**Masculino**	**Feminino**	**Masculino**	**Feminino**
(1,2,1)	(1,2,0)	6,24	-8,95	-	-
(1,2,2)	(1,2,1)	1,92	-10,62	-	-
(2,2,0)	(2,2,0)	5,37	-12,22*	-	-
(3,2,0)	(2,2,1)	5,77	-10,26	5,37	-12,22
(3,2,1)	-	5,38	-	4,15	-
(3,2,2)	-	2,65	-	1,45	-
(4,2,0)	-	2,31	-	-	-
(4,2,1)	-	-1,92	-	-3,43*	-
(4,2,2)	-	0,05	-	-3,43	-

Tabela B.1: Valores de AIC para Modelos ARIMA - Masculino e Feminino (Antes e Depois da Remoção de Parâmetros Não Significativos) para Angola

Modelo ARIMA		AIC - Antes		AIC - Depois	
Masculino	**Feminino**	**Masculino**	**Feminino**	**Masculino**	**Feminino**
(0,2,0)	(0,2,0)	44,90	-38,78	-	-
(0,2,1)	-	35,11	-	-	-
(0,2,2)	-	30,2*	-	-	-
(1,2,0)	(1,2,0)	32,35	-37,78	-	-38,78
(1,2,1)	(1,2,1)	33,45	-39,47	32,35	-
-	(2,2,0)	-	-39,20	-	-40,90
-	(2,2,1)	-	-39,00	-	-40,90
-	(3,2,0)	-	- 40,84	-	-41,66
-	(3,2,1)	-	-43,11*	-	-

Tabela B.2: Valores de AIC para Modelos ARIMA - Masculino e Feminino (Antes e Depois da Remoção de Parâmetros Não Significativos) para o Brasil

Modelo ARIMA		AIC - Antes		AIC - Depois	
Masculino	**Feminino**	**Masculino**	**Feminino**	**Masculino**	**Feminino**
(2,1,0)	(2,1,0)	-16,53	-16,53	-	-
(2,1,1)	(2,1,1)	-14,76	-16,53	-14,76	-16,53
(7,1,0)	(7,1,0)	-16,23	-16,23	-16,53	-16,53
(8,1,0)	(8,1,0)	-14,58	-14,58	-16,53	-16,53

Tabela B.3: Valores de AIC para Modelos ARIMA - Masculino e Feminino (Antes e Depois da Remoção de Parâmetros Não Significativos) para Cabo Verde

Modelo ARIMA		AIC - Antes		AIC - Depois	
Masculino	**Feminino**	**Masculino**	**Feminino**	**Masculino**	**Feminino**
(2,1,1)	(2,1,1)	-31,04	-31,04	-	-
(3,1,0)	(3,1,0)	-32,71	-32,71	-34,15	-34,15
(3,1,1)	(3,1,1)	-30,85	-30,85	-34,15	-34,15
(4,1,0)	(4,1,0)	-30,75	-30,75	-34,15	-30,91

Tabela B.4: Valores de AIC para Modelos ARIMA - Masculino e Feminino (Antes e Depois da Remoção de Parâmetros Não Significativos) para Guiné-Bissau

Modelo ARIMA		AIC - Antes		AIC - Depois	
Masculino	**Feminino**	**Masculino**	**Feminino**	**Masculino**	**Feminino**
(0,2,0)	(0,2,0)	-11,46	-13,82	-	-
(1,2,0)	(1,2,0)	-25,44	-19,39	-	-
(2,2,0)	(2,2,0)	-25,90*	-25,26*	-	-
(2,2,1)	(2,2.1)	-24,13	-23,30	-25,44	-25,26

Tabela B.5: Valores de AIC para Modelos ARIMA - Masculino e Feminino (Antes e Depois da Remoção de Parâmetros Não Significativos) para Guiné Equatorial

Modelo ARIMA		AIC - Antes		AIC - Depois	
Masculino	**Feminino**	**Masculino**	**Feminino**	**Masculino**	**Feminino**
(0,2,0)	(0,2,0))	105,41	-13,82	-	-
(0,2,1)	(0,2,1)	102,95	96,67	-	-
(1,2,0)	(1,2,0)	100,26	94,59	-	-
(1,2,1)	(1,2,1)	100,25	94,71	100,26	94,59
(2,2,0)	(2,2,0)	100,16	94,68	100,26	94,59
(2,2,1)	(2,2,1)	102	-	100,26	-

Tabela B.6: Valores de AIC para Modelos ARIMA - Masculino e Feminino (Antes e Depois da Remoção de Parâmetros Não Significativos) para Moçambique

Modelo ARIMA		AIC - Antes		AIC - Depois	
Masculino	**Feminino**	**Masculino**	**Feminino**	**Masculino**	**Feminino**
(0,2,0)	(0,2,0))	-60,15	-60,15	-	-
(0,2,1)	(0,2,1)	-58,42	-65,03	-60,15	-65,93
-	(0,2,2)	-	-70,87	-	-
(1,2,0)	-)	-58,69	-	-60,15	-
(1,2,1)	-	-61,29	-	-	-
(1,2,2)	(1,2,2)	-64,14	-70,87*	-	-
(2,2,0)	(2,2,0)	-66,71	-65,56	-68,45*	-65,93
(2,2,1)	(2,2,1)	-65,13	-65,23	-68,45	-65,93
-	(2,2,2)	-	-69,40	-	-70,87

Tabela B.7: Valores de AIC para Modelos ARIMA - Masculino e Feminino (Antes e Depois da Remoção de Parâmetros Não Significativos) para Portugal

Modelo ARIMA		AIC - Antes		AIC - Depois	
Masculino	**Feminino**	**Masculino**	**Feminino**	**Masculino**	**Feminino**
(1,2,0)	(1,2,0)	109,45	23,28	-	21,28
(1,2,1)	(1,2,1)	109,91	20,63	109,45	-34,15
(1,2,2)	(1,2,2)	113,66	22,20	109,45	-
(1,2,3)	(1,2,3)	99,15	16,93	96,77*	13,83*
(2,2,0)	(2,2,0)	108,49	24,83	109,45	21,28
-	(3,2,0)	-	17,95	-	14,64
-	(4,2,0)		19,42	-	14,64

Tabela B.8: Valores de AIC para Modelos ARIMA - Masculino e Feminino (Antes e Depois da Remoção de Parâmetros Não Significativos) para São Tomé e Principe

Modelo ARIMA		AIC - Antes		AIC - Depois	
Masculino	**Feminino**	**Masculino**	**Feminino**	**Masculino**	**Feminino**
(1,2,0)	(1,2,0)	-24,87	-24,87	-	-
(1,2,1)	(1,2,1)	-31,87	-31,87	-	-
(2,2,0)	(2,2,0)	-33,41*	-33,41*	-	-
(2,2,1)	(2,2,1)	-32,00	-32,00	-33,41	-33,41

Tabela B.9: Valores de AIC para Modelos ARIMA - Masculino e Feminino (Antes e Depois da Remoção de Parâmetros Não Significativos) paraTimor-Leste

C

Resíduos

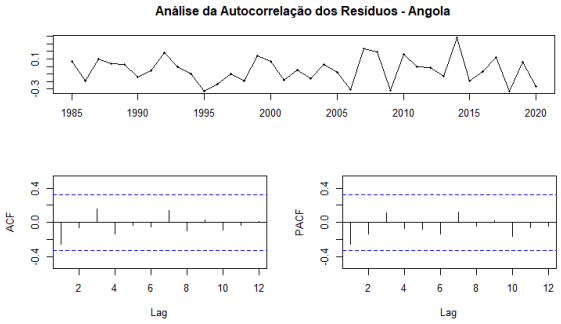

Figura C.1: Análise de autocorrelação de resíduos dos indivíduos do sexo masculino de Angola.

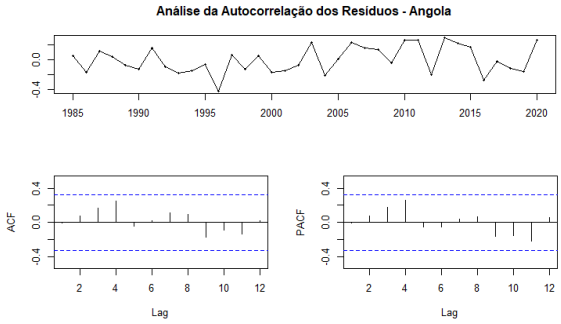

Figura C.2: Análise de autocorrelação de resíduos dos indivíduos do sexo feminino de Angola.

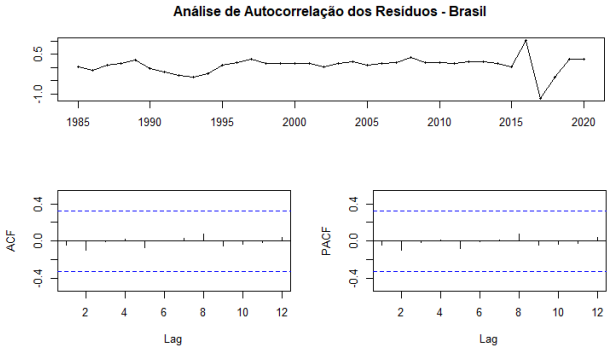

Figura C.3: Análise de autocorrelação de resíduos dos indivíduos do sexo masculino do Brasil.

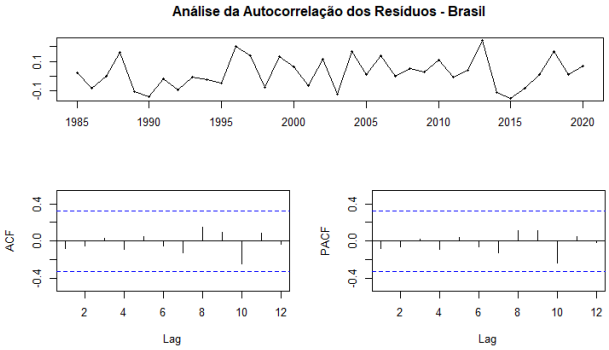

Figura C.4: Análise de autocorrelação de resíduos dos indivíduos do sexo feminino do Brasil.

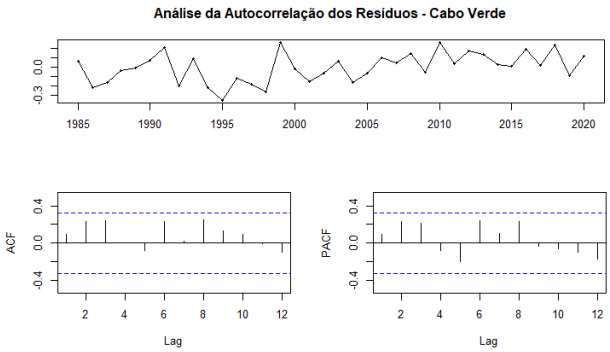

Figura C.5: Análise de autocorrelação de resíduos dos indivíduos do sexo masculino de Cabo Verde.

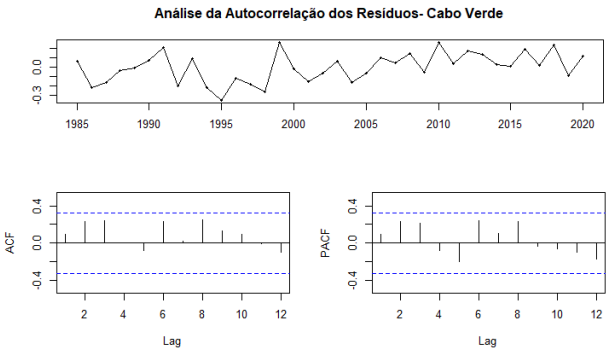

Figura C.6: Análise de autocorrelação de resíduos dos indivíduos do sexo feminino de Cabo Verde.

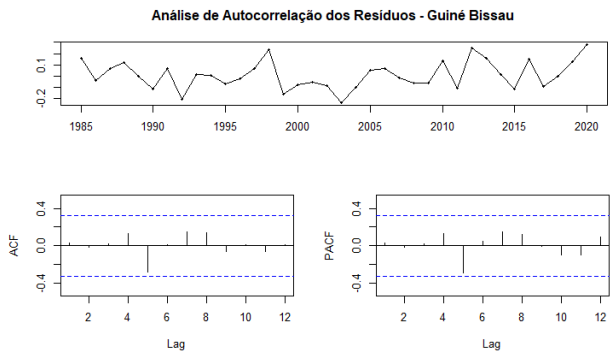

Figura C.7: Análise de autocorrelação de resíduos dos indivíduos do sexo masculino de Guiné-Bissau.

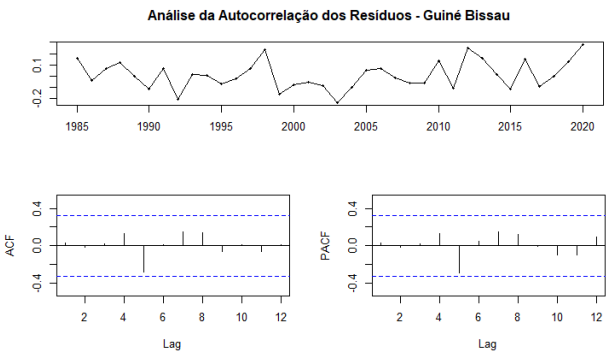

Figura C.8: Análise de autocorrelação de resíduos dos indivíduos do sexo feminino de Guiné-Bissau.

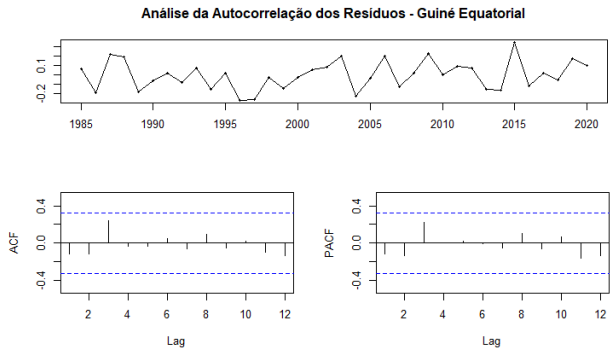

Figura C.9: Análise de autocorrelação de resíduos dos indivíduos do sexo masculino da Guiné Equatorial.

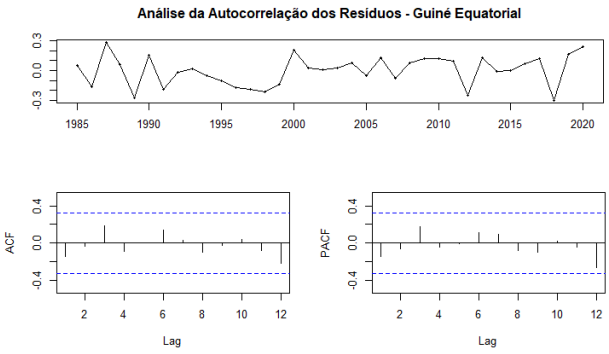

Figura C.10: Análise de autocorrelação de resíduos dos indivíduos do sexo feminino da Guiné Equatorial.

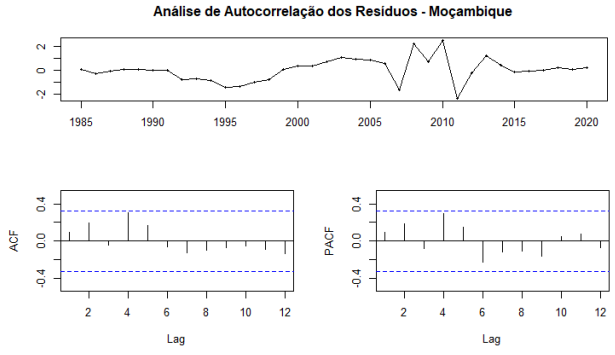

Figura C.11: Análise de autocorrelação de resíduos dos indivíduos do sexo masculino de Moçambique.

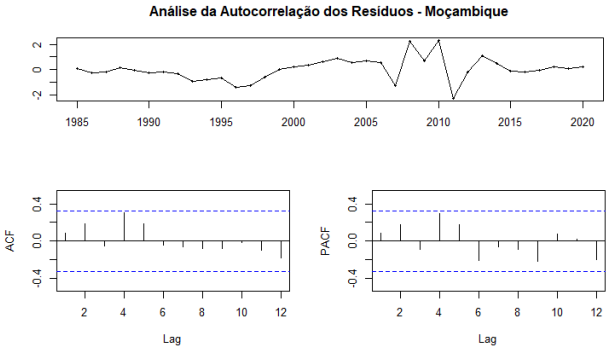

Figura C.12: Análise de autocorrelação de resíduos dos indivíduos do sexo feminino de Moçambique.

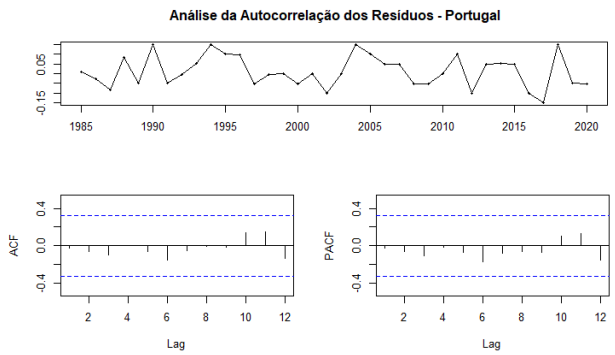

Figura C.13: Análise de autocorrelação de resíduos dos indivíduos do sexo masculino de Portugal.

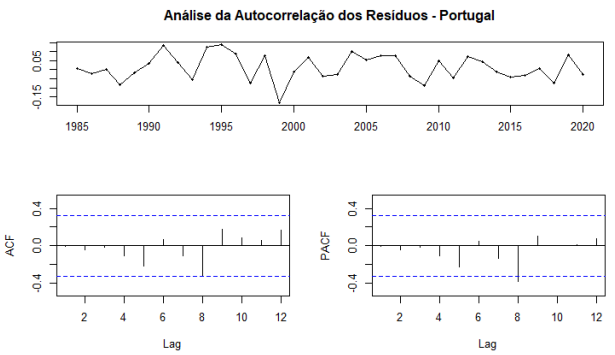

Figura C.14: Análise de autocorrelação de resíduos dos indivíduos do sexo feminino de Portugal.

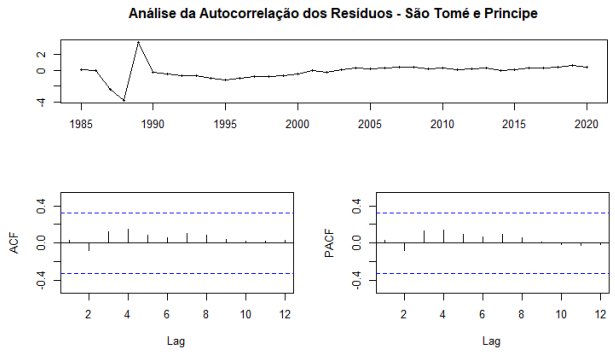

Figura C.15: Análise de autocorrelação de resíduos dos indivíduos do sexo masculino de São Tomé e Príncipe.

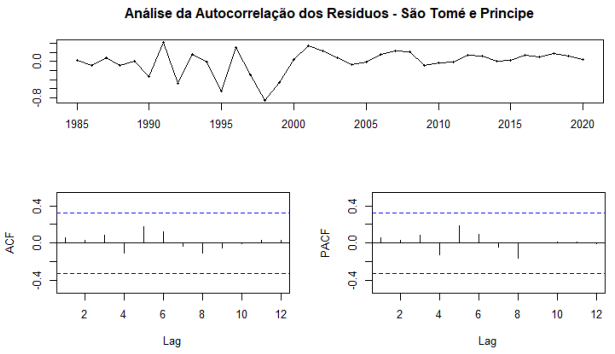

Figura C.16: Análise de autocorrelação de resíduos dos indivíduos do sexo feminino de São Tomé e Príncipe.

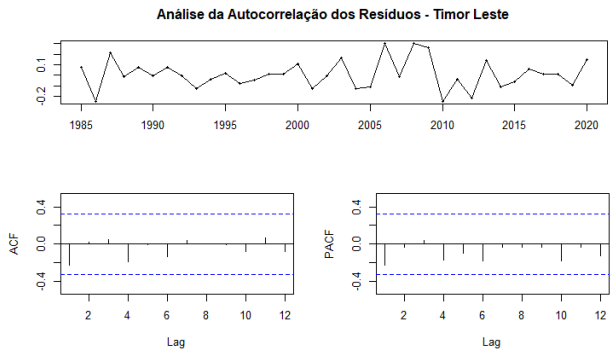

Figura C.17: Análise de autocorrelação de resíduos dos indivíduos do sexo masculino de Timor-Leste.

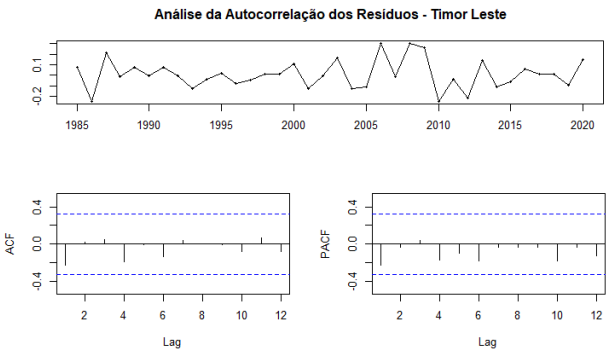

Figura C.18: Análise de autocorrelação de resíduos dos indivíduos do sexo feminino de Timor-Leste.

D

Gráficos de previsão do conjunto de dados de treino com os dados de teste

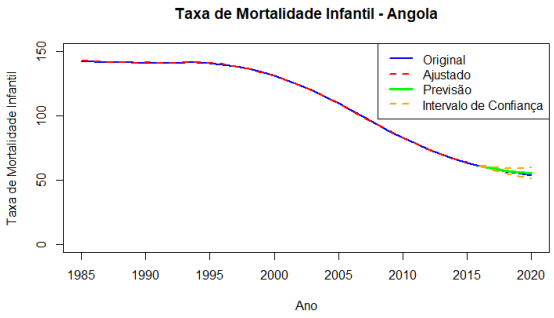

Figura D.1: Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Angola (Masculino).

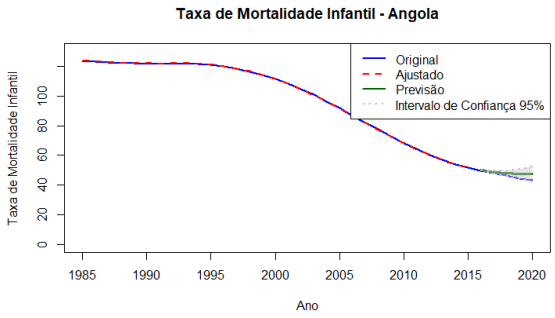

Figura D.2: Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Angola (Feminino).

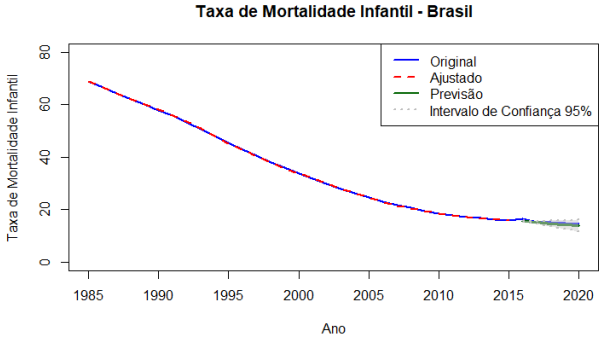

Figura D.3: Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Brasil (Masculino).

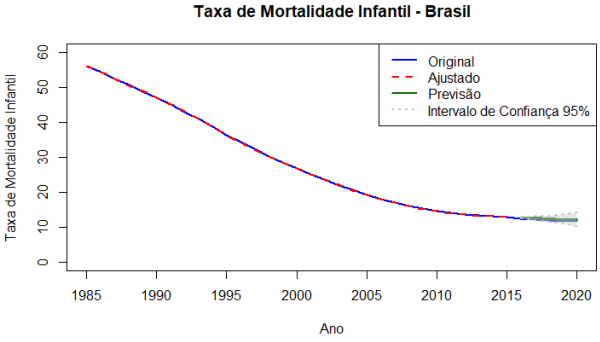

Figura D.4: Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Brasil (Feminino).

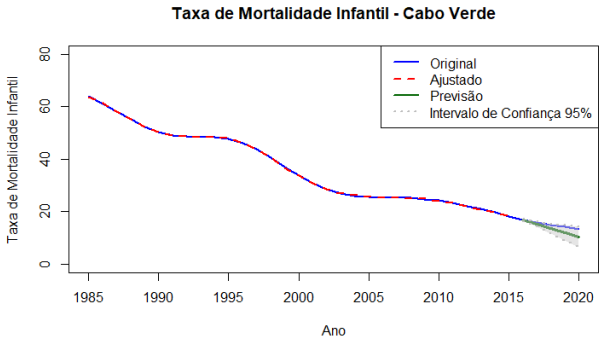

Figura D.5: Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Cabo Verde (Masculino).

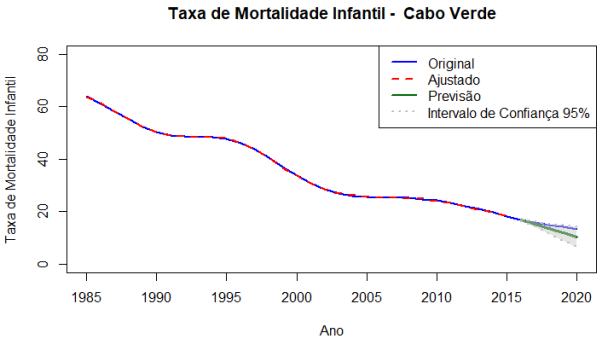

Figura D.6: Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Cabo Verde (Feminino).

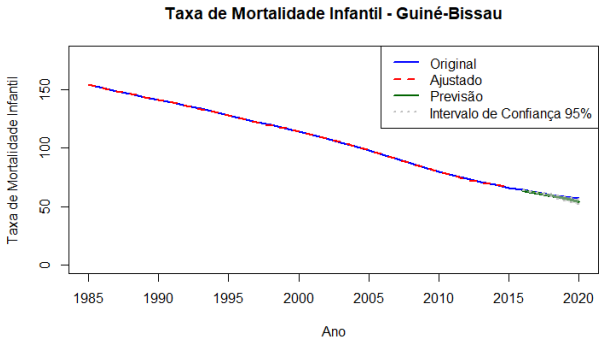

Figura D.7: Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Guiné-Bissau (Masculino).

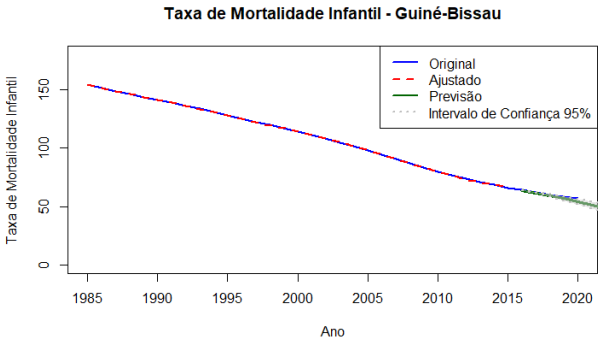

Figura D.8: Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Guiné-Bissau (Feminino).

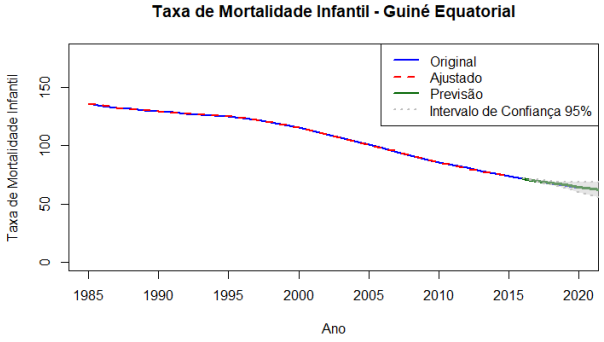

Figura D.9: Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Guiné Equatorial (Masculino).

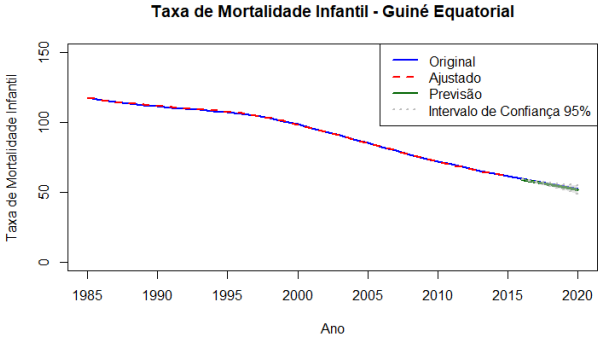

Figura D.10: Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Guiné Equatorial (Feminino).

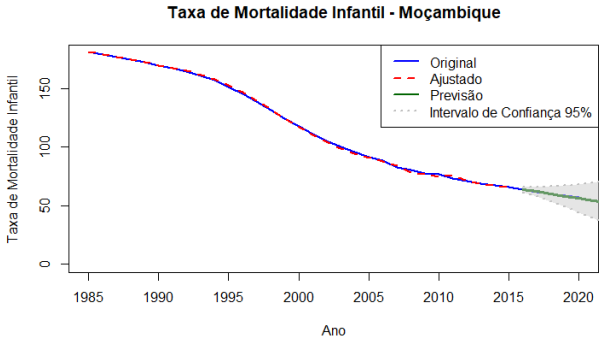

Figura D.11: Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Moçambique (Masculino).

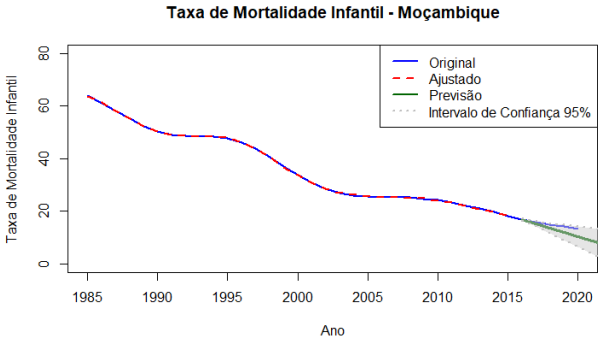

Figura D.12: Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Moçambique (Feminino).

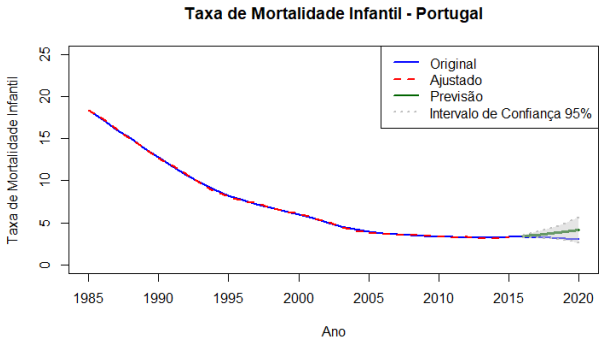

Figura D.13: Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Portugal (Masculino).

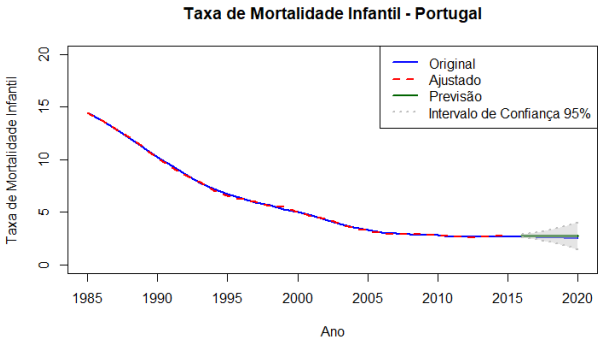

Figura D.14: Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Portugal (Feminino).

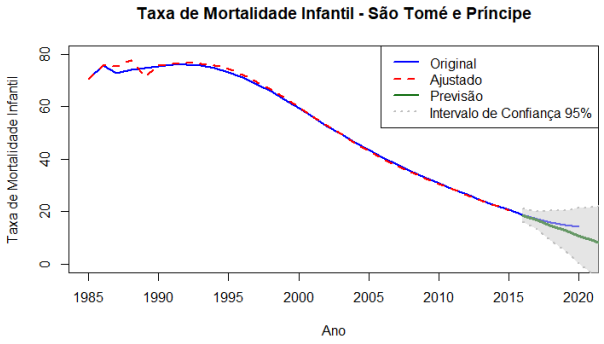

Figura D.15: Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - São Tomé e Príncipe (Masculino).

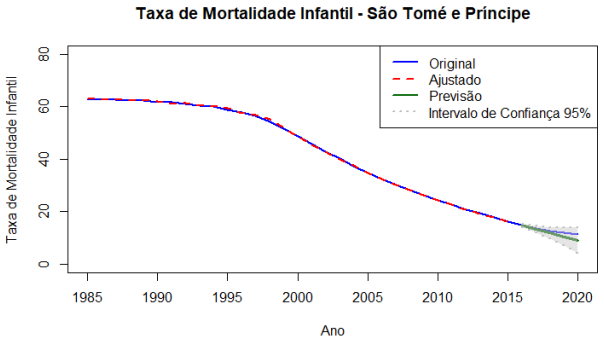

Figura D.16: Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - São Tomé e Príncipe (Feminino).

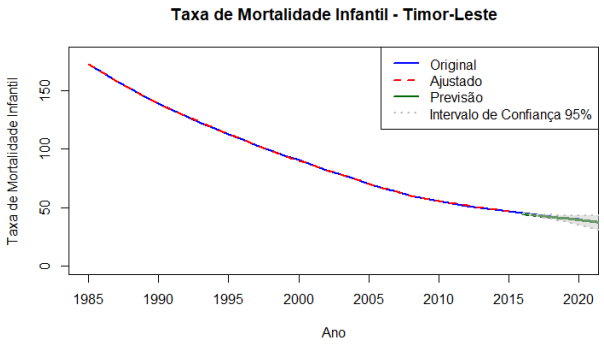

Figura D.17: Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Timor-Leste (Masculino).

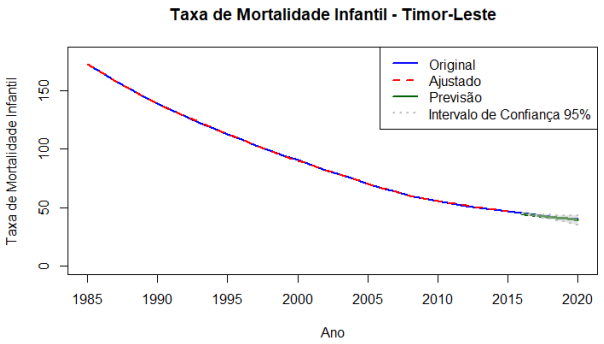

Figura D.18: Previsão das taxas de mortalidade infantil do conjunto de treino com os dados de Teste - Timor-Leste (Feminino).

E

Gráficos das quebras de estruturas nas séries

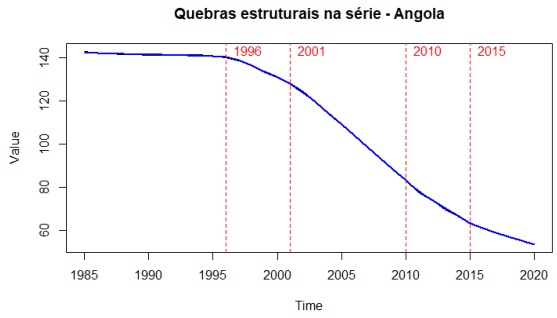

Figura E.1: Quebras de estruturas na série - Angola (Masculino).

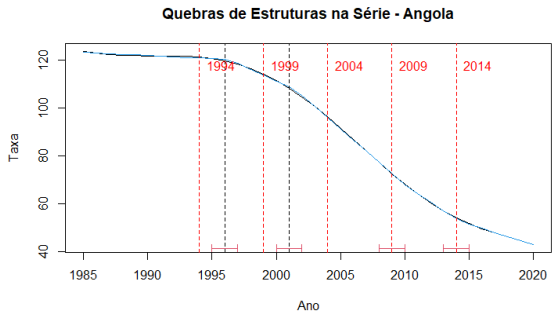

Figura E.2: Quebras de estruturas na série - Angola (Feminino).

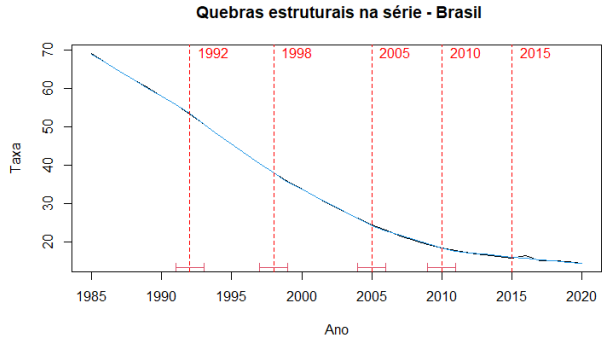

Figura E.3: Quebras de estruturas na série - Brasil (Masculino).

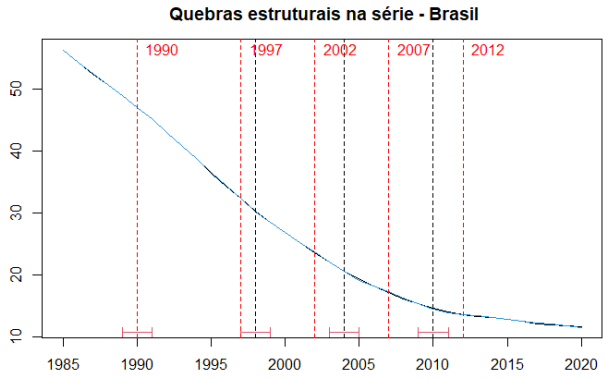

Figura E.4: Quebras de estruturas na série - Brasil (Feminino).

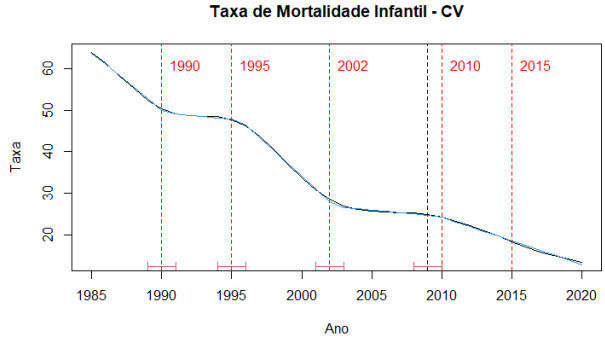

Figura E.5: Quebras de estruturas na série - Cabo Verde (Masculino).

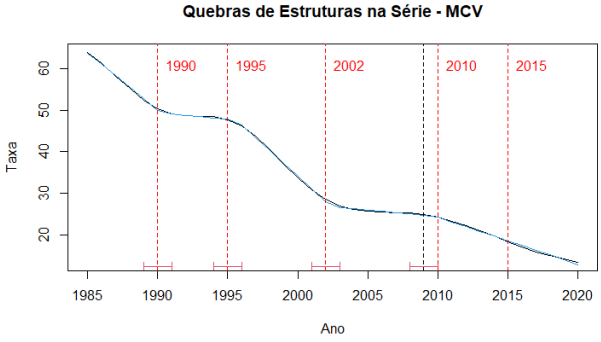

Figura E.6: Quebras de estruturas na série - Cabo Verde (Feminino).

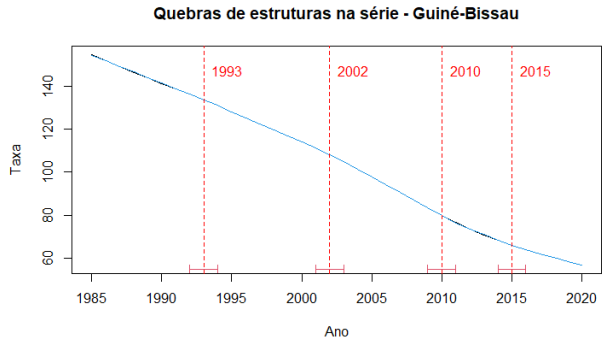

Figura E.7: Quebras de estruturas na série - Guiné--Bissau (Masculino).

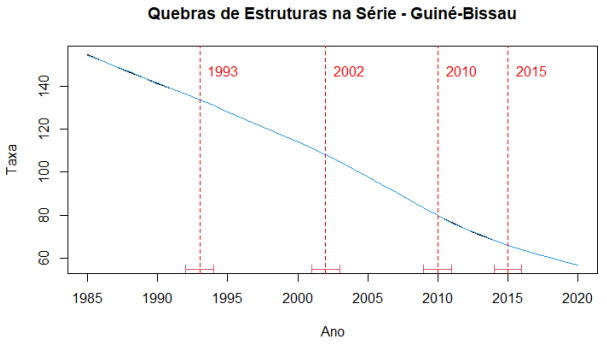

Figura E.8: Quebras de estruturas na série - Guiné--Bissau (Feminino).

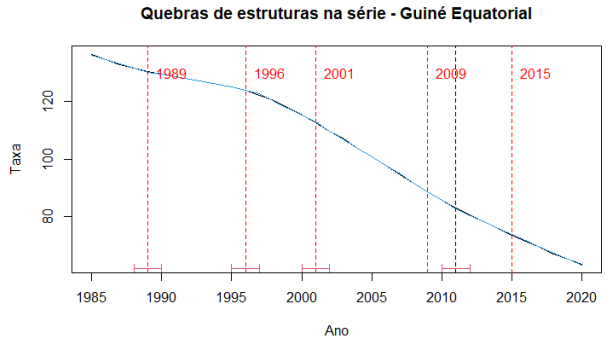

Figura E.9: Quebras de estruturas na série - Guiné Equatorial (Masculino).

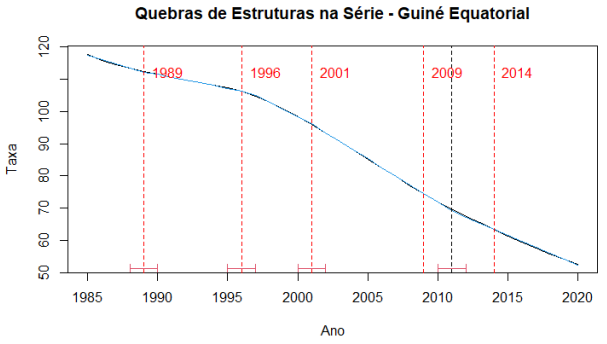

Figura E.10: Quebras de estruturas na série - Guiné Equatorial (Feminino).

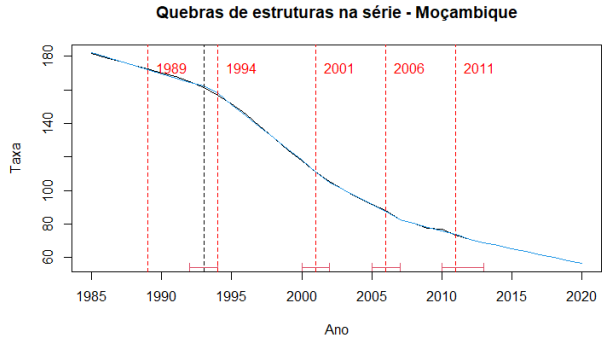

Figura E.11: Quebras de estruturas na série - Moçambique (Masculino).

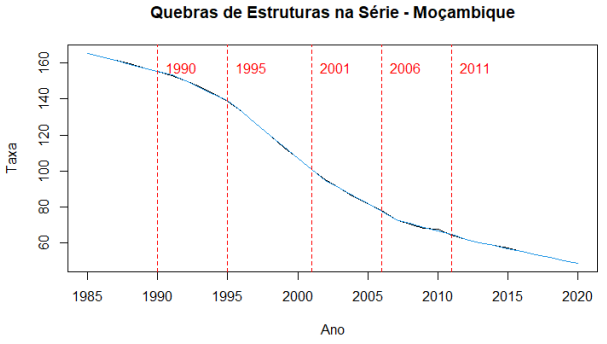

Figura E.12: Quebras de estruturas na série - Moçambique (Feminino).

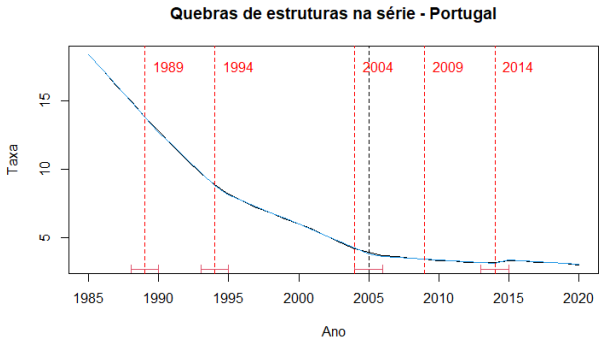

Figura E.13: Quebras de estruturas na série - Portugal (Masculino).

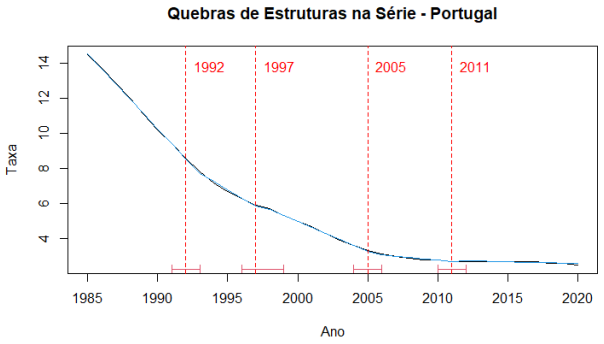

Figura E.14: Quebras de estruturas na série - Portugal (Feminino).

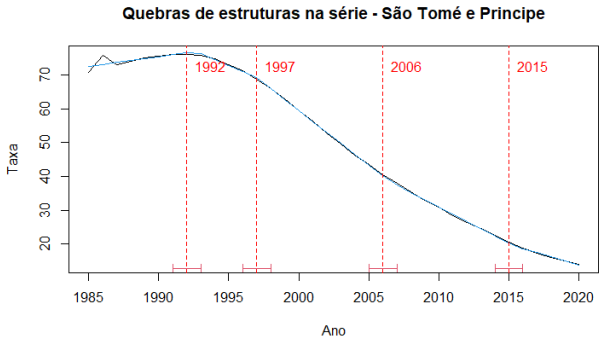

Figura E.15: Quebras de estruturas na série - São Tomé e Príncipe (Masculino).

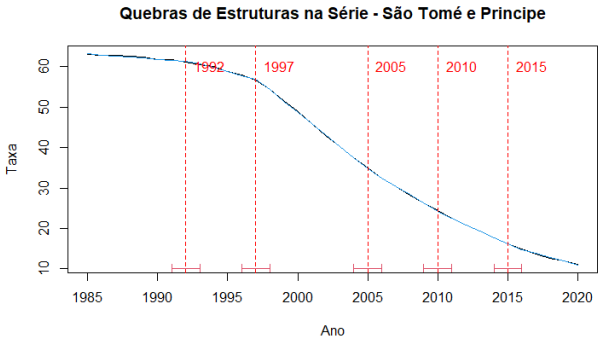

Figura E.16: Quebras de estruturas na série - São Tomé e Príncipe (Feminino).

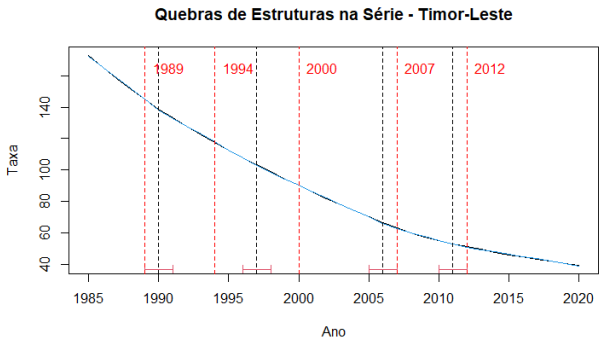

Figura E.17: Quebras de estruturas na série - Timor Leste (Masculino).

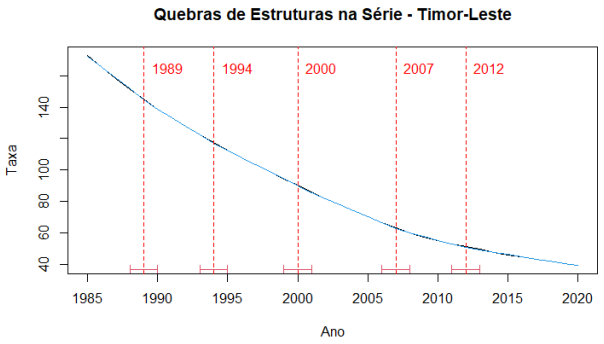

Figura E.18: Quebras de estruturas na série - Timor Leste (Feminino).

Bibliografia

[1] AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, v. 19, n. 6, p. 716723, 1974.

[2] BAI, J.; PERRON, P. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, v. 18, n. 1, p. 122, 2003.

[3] BAI, J.; PERRON, P. Estimating and testing linear models with multiple structural changes. *Econometrica*, v. 66, n. 1, p. 4778, 1998.

[4] BANCO MUNDIAL. Dados sobre o Produto Interno Bruto da Guiné-Bissau. Disponível em: `https://data.worldbank.org/country/guinea-bissau`. Acesso em: 10 jan. 2023.

[5] BANCO MUNDIAL. Dados sobre a população da Guiné Equatorial. Disponível em: `https://data.worldbank.org/indicator/SP.POP.TOTL?locations=GQ`. Acesso em: 15 fev. 2023.

[6] BANCO MUNDIAL. Timor-Leste - Visão Geral. Disponível em: `https://www.worldbank.org/pt/country/timor-leste/overview`. Acesso em: 5 mar. 2023.

[7] BANCO MUNDIAL. Relatório Econômico de São Tomé e Príncipe. Disponível em: `https://www.worldbank.org/en/country/saotomeandprincipe`. Acesso em: 20 abr. 2023.

[8] BANCO MUNDIAL. Moçambique: Indicadores de Desenvolvimento. Disponível em: `https://data.worldbank.org/country/mozambique`. Acesso em: 10 mai. 2024.

[9] BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. *Time Series Analysis: Forecasting and Control*. 5ł ed. Hoboken: Wiley, 2015.

[10] BROCKWELL, P. J.; DAVIS, R. A. *Introduction to Time Series and Forecasting*. 2ł ed. New York: Springer, 2002.

[11] BROWN, R. G. *Statistical Forecasting for Inventory Control*. New York: McGraw-Hill, 1959.

[12] CASINI, A.; PERRON, P. (2018). *Structural Breaks in Time Series*. Preparado para a *Oxford Research Encyclopedia of Economics and Finance*. Disponível em: `https://www.oxfordresearchencyclopedia.com/`

[13] PONGOU, R. (2012). *Why is infant mortality higher in boys than in girls? A new hypothesis based on preconception environment and evidence from a natural experiment in Africa*. *Demography*, 50(2), 421–444. https://doi.org/10.1007/s13524-012-0156-5

[14] JACQUINET, M. (2019). *PIB - Produto Interno Bruto: Breve Introdução*. Lisboa: Universidade Aberta. Versão estendida para uso exclusivo dos alunos da Universidade Aberta.

[15] PLOBERGER, W., & Krämer, W. (1992). *The CUSUM test with OLS residuals*. Econometrica, 60(2), 271-285.

[16] HANSEN, B. E. (2001). *The new econometrics of structural change: Dating breaks in U.S. labor productivity*. Journal of Economic Perspectives, 15(4), 117-128.

[17] BROWN, R. L., Durbin, J., & Evans, J. M. (1975). *Techniques for testing the constancy of regression relationships over time*. Journal of the Royal Statistical Society: Series B, 37, 149-163.

[18] CENTRAL INTELLIGENCE AGENCY. *The World Factbook: Equatorial Guinea*. Disponível em: `https://www.cia.gov/the-world-factbook/countries/equatorial-guinea/`. Acesso em: 25 jun. 2023.

[19] KUROZUMI, E., & TUVAANDORJI, P. (2021). *Testing for structural breaks in cointegrated VAR models*. Journal of Econometrics, 222(1), 423-442.

[20] CHOI, J., & SHIN, D. W. (2019). *Robust tests for structural breaks in linear models*. Econometric Reviews, 38(8), 871-894.

[21] MUGGEO, V. M. R. (2003). *Estimating regression models with unknown break-points*. Statistics in Medicine, 22(19), 3055-3071.

[22] ZEILEIS, A., SHAH, A., & PATNAIK, I. (2013). *Testing, monitoring, and dating structural changes in exchange rate regimes*. Computational Statistics & Data Analysis, 60, 212-228.

[23] CHOW, G. C. (1960). *Tests of equality between sets of coefficients in two linear regressions*. Econometrica, 28(3), 591-605.

[24] CHATFIELD, C. *The Analysis of Time Series: An Introduction*. 6ł ed. Boca Raton: Chapman & Hall/CRC, 2003.

[25] CENTRAL INTELLIGENCE AGENCY. *The World Factbook: Angola Economy*. Disponível em: `https://www.cia.gov/the-world-factbook/countries/angola/`. Acesso em: 30 jul. 2023.

[26] COMUNIDADE DOS PAÍSES DE LÍNGUA PORTUGUESA. Guiné Equatorial. Disponível em: `https://www.cplp.org/id-2769.aspx`. Acesso em: 12 jan. 2024.

[27] COUNTRY ECONOMY. Timor-Leste - Esperança de vida ao nascer. Disponível em: `https://countryeconomy.com/demography/life-expectancy/timor-leste`. Acesso em: 8 fev. 2024.

[28] COWPERTWAIT, P. S. P.; METCALFE, A. V. *Introductory Time Series with R*. New York: Springer, 2009.

[29] CPLP. I Conferência de Chefes de Estado e de Governo da CPLP. Disponível em: `https://www.cplp.org/Default.aspx?ID=3667`. Acesso em: 20 fev. 2024.

[30] CRYER, J. D.; CHAN, K.-S. *Time Series Analysis: With Applications in R*. 2ł ed. New York: Springer, 2008.

[31] DELURGIO, S. A. *Forecasting Principles and Applications*. Boston: Irwin/McGraw-Hill, 1998. ISBN 978-0-256-13433-9.

[32] DICKEY, D. A.; FULLER, W. A. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, v. 49, n. 4, p. 10571072, 1981.

[33] DURBIN, J.; KOOPMAN, S. J. *Time Series Analysis by State Space Methods*. 2ł ed. Oxford: Oxford University Press, 2012.

[34] ENDERS, W. *Applied Econometric Time Series*. 4ł ed. Hoboken: Wiley, 2014.

[35] ENTIDADE REGULADORA DOS SERVIÇOS DE ÁGUAS E RESÍDUOS (ERSAR). Relatório Anual dos Serviços de Águas e Resíduos em Portugal 2022. Disponível em: `https://www.ersar.pt/pt/consumidor/relatorios-anuais`. Acesso em: 15 mar. 2024.

[36] EVERITT, B.; HOTHORN, T. (2011).. *An Introduction to Applied Multivariate Analysis with R*. Springer-Verlag New York. DOI: 10.1007/978-1-4419-9650-3.

[37] FUNDO MONETÁRIO INTERNACIONAL (FMI). Relatório de Crescimento Econômico de São Tomé e Príncipe. Disponível em: `https://www.imf.org/en/Countries/STP`. Acesso em: 22 mar. 2024.

[38] GARDNER, E. S. Exponential smoothing: The state of the art. *Journal of Forecasting*, v. 4, n. 1, p. 128, 1985.

[39] HAMILTON, J. D. *Time Series Analysis*. Princeton: Princeton University Press, 1994.

[40] HOLT, C. C. Forecasting trends and seasonals by exponentially weighted averages. Carnegie Institute of Technology, ONR Memorandum 52, 1957.

[41] HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting: Principles and Practice*. 2ł ed. Melbourne: OTexts, 2018. Disponível em: `https://otexts.com/fpp2/`.

[42] INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Taxa de alfabetização das pessoas de 15 anos ou mais de idade. Disponível em: `https://www.ibge.gov.br/estatisticas/sociais/educacao/9662-pesquisa-nacional-por-amostra-de-domicilios.html?edicao=21830&t=resultados`. Acesso em: 18 abr. 2024.

[43] INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Tábuas completas de mortalidade para o Brasil 2021: Breve análise da evolução da mortalidade no Brasil. Disponível em: `https://ftp.ibge.gov.br/Tabuas_Completas_de_Mortalidade/Tabuas_Completas_de_Mortalidade_2021/notastecnicas.pdf`. Acesso em: 10 jul. 2024.

[44] INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Produto Interno Bruto PIB. Disponível em: `https://www.ibge.gov.br/explica/pib.php`. Acesso em: 12 jul. 2024.

[45] INSTITUTO NACIONAL DE ESTATÍSTICA (INE). Relatório Anual de Indicadores Educacionais de São Tomé e Príncipe. Disponível em: `https://www.ine.st/index.php/o-pais/indicadores`. Acesso em: 15 jul. 2024.

[46] INSTITUTO NACIONAL DE ESTATÍSTICA (INE). Indicadores Demográficos. Disponível em: `https://www.ine.st/index.php/o-pais/indicadores`. Acesso em: 18 jul. 2024.

[47] INSTITUTO NACIONAL DE ESTATÍSTICA (INE). Censos 2021: Resultados Definitivos. Disponível em: `https://www.ine.pt/xurl/pub/586678547`. Acesso em: 20 abr. 2024.

[48] INSTITUTO NACIONAL DE ESTATÍSTICA (INE). Estatísticas Demográficas 2022. Disponível em: https://www.ine.pt/xurl/pub/586678547. Acesso em: 20 jul. 2024.

[49] INSTITUTO NACIONAL DE ESTATÍSTICA DA GUINÉ EQUATORIAL. Dados Económicos de 2022. Disponível em: https://www.inege.gq/. Acesso em: 25 abr. 2024.

[50] INSTITUTO NACIONAL DE ESTATÍSTICA DA GUINÉ-BISSAU. Mortalidade. Disponível em: https://www.stat-guinebissau.com/Menu_principal/IV_RGPH/rgph1/Mortalidade.pdf. Acesso em: 30 abr. 2024.

[51] INSTITUTO NACIONAL DE ESTATÍSTICA DE CABO VERDE. Taxa de Mortalidade Infantil (por 1000) (M.saúde). Disponível em: https://ine.cv/indicadores/taxa-mortalidade-infantil-1000-m-saude/. Acesso em: 22 jul. 2024.

[52] INSTITUTO NACIONAL DE ESTATÍSTICA DE MOÇAMBIQUE. Síntese de Conjuntura Económica IV Trimestre de 2022. Disponível em: https://ine.gov.mz/web/guest/b/sintese-de-conjuntura-economica-ivtrimestre-de-2022. Acesso em: 10 mai. 2024.

[53] JOINT MONITORING PROGRAMME (JMP). Progress on household drinking water, sanitation and hygiene 2000-2020: Five years into the SDGs. Disponível em: https://washdata.org/reports. Acesso em: 15 mai. 2024.

[54] KENDALL, M.; ORD, J. K. *Time Series*. 3ł ed. London: Edward Arnold, 1990.

[55] LUTKEPOHL, H. *New Introduction to Multiple Time Series Analysis*. Berlin: Springer, 2005.

[56] MINISTÉRIO DA SAÚDE DE SÃO TOMÉ E PRÍNCIPE. Anuário Estatístico da Saúde 2021. Disponível em: https://minsaude.st/wp-content/uploads/2022/09/Anuario-de-SAUDE-STP-2021-versao-FINAL_22.pdf. Acesso em: 20 mai. 2024.

[57] MINISTÉRIO DA SAÚDE. Mortalidade infantil no Brasil. *Boletim Epidemiológico*, 52(37). Disponível em: https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/epidemiologicos/edicoes/2021/boletim_epidemiologico_svs_37_v2.pdf. Acesso em: 22 mai. 2024.

[58] MINISTÉRIO DOS NEGÓCIOS ESTRANGEIROS DE PORTUGAL. História de Timor-Leste. Disponível em: https://www.portaldiplomatico.mne.gov.pt/relacoesbilaterais/paises-geral/timor-leste. Acesso em: 25 mai. 2024.

[59] MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. *Introduction to Time Series Analysis and Forecasting*. 2ł ed. Hoboken: Wiley, 2015.

[60] MORETTIN, P. A.; TOLOI, C. M. C. *Análise de Séries Temporais*. São Paulo: Edgard Blucher, 2004. ISBN 8521203896.

[61] MULLER, D. Como Lidar com as Mudanças de Estrutura em Séries Temporais? Instituto Superior de Economia e Gestão, Lisboa.

[62] NAÇÕES UNIDAS. Objetivos de Desenvolvimento do Milênio. Disponível em: https://brasil.un.org/pt-br/66851-os-objetivos-de-desenvolvimento-do-milenio. Acesso em: 30 mai. 2024.

[63] O PAÍS. Número de pobres aumentou de 46,1% para 65% da população nos últimos 10 anos. Disponível em: https://opais.co.mz/numero-de-pobres-aumentou-de-461-para-65-da-populacao-nos-ultimos-10-anos/. Acesso em: 1 jun. 2024.

[64] ORGANIZAÇÃO MUNDIAL DA SAÚDE. Plano Nacional de Desenvolvimento Sanitário 2017-2021. Disponível em: https://extranet.who.int/countryplanningcycles/sites/default/files/public_file_rep/CAV_Cape_Verde_PNDS%202017-2021%20-VOL.1.pdf. Acesso em: 5 jun. 2024.

[65] PALMA, W. *Time Series Analysis*. Hoboken: Wiley, 2016.

[66] PANKRATZ, A. *Forecasting with Dynamic Regression Models*. New York: Wiley, 1991.

[67] PERRON, P. Dealing with structural breaks. In: MILLS, T. C.; PATTERSON, K. (Eds.). *Palgrave Handbook of Econometrics: Volume 1, Econometric Theory*. Basingstoke: Palgrave Macmillan, 2006. p. 278352.

[68] PIRES, A. P. *Notas de Séries Temporais*. Lisboa: Instituto Superior Técnico, 2001. Disponível em: https://www.math.tecnico.ulisboa.pt/~apacheco/CML/Notas.pdf. Acesso em: 10 jun. 2024.

[69] POLLOCK, D. S. G. *Smoothings and Forecasting*. Cambridge: Cambridge University Press, 1999.

[70] POPULATIONPYRAMID.NET. Pirâmides Populacionais: São Tomé e Príncipe - 2022. Disponível em: https://www.populationpyramid.net/pt/s%C3%A3o-tom%C3%A9-e-pr%C3%ADncipe/2022/. Acesso em: 25 jul. 2024.

[71] PORDATA. Esperança de Vida à Nascença em Portugal. Disponível em: https://www.pordata.pt/Portugal/Esperan%C3%A7a+de+vida+%C3%A0+nascen%C3%A7a+total+e+por+sexo-418. Acesso em: 28 jul. 2024.

[72] PORDATA. PIB a preços correntes (base 2016) - Portugal. Disponível em: https://www.pordata.pt/Portugal/PIB+a+pre%C3%A7os+correntes+(base+2016)-2394. Acesso em: 30 jul. 2024.

[73] PRIESTLEY, M. B. *Spectral Analysis and Time Series*. London: Academic Press, 1981.

[74] SALLA, M. R.; SÁ, E.; FERREIRA, P. A. S. C.; MELO, N. A. Relação entre saneamento básico e saúde pública em Bissau, Guiné-Bissau. *Saúde e Sociedade*, v. 28, n. 4, 2019. Disponível em: https://www.scielo.br/j/sausoc/a/8kF8X3t6h8gQp9V4J5w9VtF/?lang=pt. Acesso em: 1 ago. 2024.

[75] SCHWARZ, G. Estimating the dimension of a model. *Annals of Statistics*, v. 6, n. 2, p. 461464, 1978.

[76] SHUMWAY, R. H.; STOFFER, D. S. *Time Series Analysis and Its Applications: With R Examples*. 4ł ed. Cham: Springer, 2017.

[77] SISTEMA NACIONAL DE INFORMAÇÕES SOBRE SANEAMENTO. Diagnóstico dos Serviços de Água e Esgotos 2019. Disponível em: https://www.gov.br/snis/pt-br/diagnosticos/agua-e-esgotos/2019. Acesso em: 5 ago. 2024.

[78] TSAY, R. S. *Analysis of Financial Time Series*. 3ł ed. Hoboken: Wiley, 2010.

[79] UNICEF GUINÉ-BISSAU. Relatório Anual 2023. Disponível em: https://www.unicef.org/guineabissau/media/3216/file/Relatorio-Anual-do-UNICEF-Guine-Bissau-2023.pdf. Acesso em: 15 jun. 2024.

[80] UNICEF MOÇAMBIQUE. Contexto: Desnutrição em Moçambique. Disponível em: https://www.unicef.org/mozambique/contexto. Acesso em: 18 jun. 2024.

[81] UNICEF MOÇAMBIQUE. Situação da Criança em Moçambique. Disponível em: https://www.unicef.org/mozambique/relatorios/situa%C3%A7%C3%A3o-da-crian%C3%A7a-em-mo%C3%A7ambique. Acesso em: 20 jun. 2024.

[82] UNICEF. Água e Saneamento no OGE 2017. Disponível em: https://www.unicef.org/angola/media/946/file/%C3%81gua%20e%20Saneamento%20no%20OGE%202017.pdf. Acesso em: 22 jun. 2024.

[83] UNICEF. Timor-Leste: Dados sobre mortalidade infantil. Disponível em: https://data.unicef.org/country/tls/. Acesso em: 25 jun. 2024.

[84] UNICEF. Desigualdades e impactos da COVID-19 na atenção à primeira infância. Disponível em: https://www.unicef.org/brazil/media/20221/file/desigualdades-e-impactos-da-covid-19-na-atencao-a-primeira-infancia.pdf. Acesso em: 28 jun. 2024.

[85] UNICEF. Equatorial Guinea: Infant Mortality Rate. Disponível em: https://data.unicef.org/country/gnq/. Acesso em: 1 jul. 2024.

[86] UNICEF. Sobrevivência e Desenvolvimento da Criança. Disponível em: https://www.unicef.org/angola/sobrevivencia-e-desenvolvimento-da-crianca-0. Acesso em: 3 jul. 2024.

[87] UNICEF. Mortalidade infantil atinge a mínima histórica em 2022 relatório da ONU. Disponível em: https://www.unicef.org/brazil/comunicados-de-imprensa/mortalidade-infantil-atinge-minima-historica-em-2022-relatorio-da-onu. Acesso em: 5 jul. 2024.

[88] UNICEF. Mortalidade materna e na infância: mulheres e crianças estão sobrevivendo cada vez mais. Fundo das Nações Unidas para a Infância, 19 de setembro de 2019. Disponível em: https://www.unicef.org/brazil/comunicados-de-imprensa/mortalidade-materna-e-na-infancia-mulheres-e-criancas-estao-sobrevivendo-cada-vez-mais. Acesso em: 13 mar. 2025.

[89] WEI, W. W. S. *Time Series Analysis: Univariate and Multivariate Methods*. 2ł ed. Boston: Pearson/Addison Wesley, 2006.

[90] WINTERS, P. R. Forecasting sales by exponentially weighted moving averages. *Management Science*, v. 6, n. 3, p. 324342, 1960.

[91] WORLD BANK. Literacy rate, adult total (% of people ages 15 and above) Angola. Disponível em: https://data.worldbank.org/indicator/SE.ADT.LITR.ZS?locations=AO. Acesso em: 8 jul. 2024.

[92] ZHANG, G. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, v. 50, p. 159175, 2003.

[93] ZEILEIS, A., LEISCH, F., HORNIK, K. ; KLEIBER, C. (2002). *strucchange: An R Package for Testing for Structural Change in Linear Regression Models*. *Journal of Statistical Software*, 7(2), 1–38. https://doi.org/10.18637/jss.v007.i02.

[94] ZELTERMAN, D. (2015). Applied Multivariate Statistics with R. *Springer International Publishing*.