

**Matheus Franco Pereira Carneiro**

**Modelagem dos casos de dengue em  
Florianópolis - SC, no período de 2017 a  
2021**

Niterói - RJ, Brasil

16 de dezembro de 2022

**Matheus Franco Pereira Carneiro**

**Modelagem dos casos de dengue em  
Florianópolis - SC, no período de  
2017 a 2021**

**Trabalho de Conclusão de Curso**

Monografia apresentada para obtenção do grau de Bacharel em  
Estatística pela Universidade Federal Fluminense.

Orientador(a): Prof. Ana Beatriz Monteiro Fonseca

Niterói - RJ, Brasil

16 de dezembro de 2022

**Matheus Franco Pereira Carneiro**

**Modelagem dos casos de dengue em  
Florianópolis - SC, no período de 2017 a  
2021**

Monografia de Projeto Final de Graduação sob o título “*Modelagem dos casos de dengue em Florianópolis - SC, no período de 2017 a 2021*”, defendida por Matheus Franco Pereira Carneiro e aprovada em 16 de dezembro de 2022, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

---

**Prof<sup>a</sup>. Dr<sup>a</sup>. Ana Beatriz Monteiro Fonseca**  
Departamento de Estatística – UFF

---

**Prof. Dr. Victor Eduardo Leite de Almeida Duca**  
Departamento de Estatística – UFF

---

**Prof<sup>a</sup>. Dr<sup>a</sup>. Patrícia Lusié Velozo da Costa**  
Departamento de Estatística – UFF

Niterói, 16 de dezembro de 2022

Ficha catalográfica automática - SDC/BIME  
Gerada com informações fornecidas pelo autor

C289m Carneiro, Matheus Franco Pereira  
Modelagem dos casos de dengue em Florianópolis - SC, no período de 2017 a 2021 / Matheus Franco Pereira Carneiro. - 2022.  
70 f.: il.

Orientador: Ana Beatriz Monteiro Fonseca.  
Trabalho de Conclusão de Curso (graduação)-Universidade Federal Fluminense, Instituto de Matemática e Estatística, Niterói, 2022.

1. Séries temporais. 2. Box e Jenkins. 3. Dengue. 4. ARIMA. 5. Produção intelectual. I. Fonseca, Ana Beatriz Monteiro, orientadora. II. Universidade Federal Fluminense. Instituto de Matemática e Estatística. III. Título.

CDD - XXX

# Resumo

A dengue é uma arbovirose causada pelo vírus DEN e possui quatro sorotipos diferentes: DENV-1, DENV-2, DENV-3 e DENV-4, sendo o mosquito *Aedes aegypti* o principal transmissor da doença no Brasil. A doença se manifesta predominantemente em locais urbanizados de clima tropical, e seus ciclos endêmicos e epidêmicos costumam ocorrer a cada 4 ou 5 anos, totalizando mais de 7 milhões de casos notificados em território nacional. Por ser um país de dimensões continentais, com diferentes cenários populacionais, econômicos e climáticos, o estudo da dengue em cada localidade apresenta particularidades e, conseqüentemente, medidas de enfrentamento diferentes. A região Sul, tratada como uma região de improvável disseminação da doença, tem reportado uma progressão da atividade do mosquito vetor e um aumento do número de casos da doença, registrando, pela primeira vez, em 2020, taxa de incidência da dengue maior que a média nacional. O presente trabalho buscou avaliar o comportamento dos casos de dengue em uma grande metrópole da região Sul com histórico recente de epidemias. Sendo assim, analisou-se, no período de janeiro de 2017 a dezembro de 2021, o número de casos de dengue em Florianópolis, além de fazer previsões para as seis primeiras semanas epidemiológicas do ano de 2022. Para isso, foram realizadas análises de séries temporais pela metodologia Box e Jenkins, investigando a existência de correlação temporal entre a série de contagem de casos e as séries de índice pluviométrico e temperatura mínima, com objetivo de encontrar modelos com boa capacidade preditiva para a incidência de novos casos semanais da doença. Nos resultados obtidos, ao incorporar nos modelos variáveis climáticas, não foi observada uma melhora no ajuste dos dados. O modelo que mostrou-se mais adequado utilizou somente os dados passados da contagem de casos da doença. As previsões para as seis primeiras semanas epidemiológicas do ano de 2022 mostraram-se razoáveis comparadas àquelas observadas na cidade de Florianópolis.

Palavras-chave: Séries temporais. Box e Jenkins. Dengue. ARIMA.

# Dedicatória

Dedico esta monografia à minha família que sempre esteve ao meu lado em toda a minha caminhada e, especialmente, ao meu avô Alcides Baptista Franco (in memoriam) que sempre me apoiou nos meus sonhos e projetos.

# Agradecimentos

Primeiramente agradeço a Deus por me proporcionar perseverança durante toda a minha vida.

Aos meus pais, Aureliano e Alcinete, e meu irmão Felipe, pelo apoio e incentivo que serviram de alicerce para as minhas realizações.

Aos meus avós, Alinete e Alcides, que sempre estiveram ao meu lado me apoiando ao longo de toda a minha trajetória.

À minha companheira, Patrícia, por estar ao meu lado em todos os momentos.

À minha orientadora, Ana Beatriz, pela sua dedicação e atenção durante o projeto.

Aos professores, Victor e Patrícia, por aceitarem avaliar o meu trabalho.

À Universidade Federal Fluminense e ao Departamento de Estatística, pelo comprometimento com a qualidade e excelência do ensino.

*"A arte da previsão consiste em antecipar o que acontecerá e depois explicar o porque não aconteceu."*

*Winston Churchill*

# Sumário

## Lista de Figuras

## Lista de Tabelas

<b>1</b>	<b>Introdução</b>	p. 13
1.1	A dengue	p. 13
1.2	A dengue no mundo	p. 14
1.3	A dengue no Brasil	p. 16
1.3.1	Região Sul e os casos de dengue	p. 17
1.4	Fundamentação Teórica	p. 18
1.5	Objetivos	p. 19
1.5.1	Objetivos específicos	p. 19
1.6	Organização	p. 20
<b>2</b>	<b>Materiais e Métodos</b>	p. 21
2.1	Local de estudo e base de dados	p. 21
2.1.1	Casos de dengue e temperatura mínima	p. 22
2.1.2	Índice pluviométrico	p. 22
2.2	Modelagem de séries temporais	p. 23
2.2.1	Componentes da série temporal	p. 26
2.2.1.1	Tendência	p. 26
2.2.1.2	Sazonalidade	p. 28
2.2.2	Metodologia de Box e Jenkins	p. 29

2.2.2.1	ARIMA . . . . .	p. 30
2.2.2.2	SARIMA . . . . .	p. 31
2.2.2.3	Função de Autocovariância . . . . .	p. 33
2.2.2.4	Função de Autocorrelação e Função de Autocorrelação Parcial . . . . .	p. 34
2.2.2.5	Correlação Cruzada . . . . .	p. 35
2.2.2.6	Função de transferência . . . . .	p. 36
2.2.2.7	Identificação . . . . .	p. 38
2.2.2.8	Estimação . . . . .	p. 39
2.2.2.9	Critérios de comparação de modelos . . . . .	p. 40
2.2.2.10	Análise de resíduos . . . . .	p. 41
2.2.2.11	Previsão <i>out-of-sample</i> . . . . .	p. 42
<b>3</b>	<b>Resultados</b>	p. 44
3.1	Análise exploratória . . . . .	p. 44
3.2	Modelo para as séries temporais . . . . .	p. 50
3.2.1	Série de Casos de Dengue . . . . .	p. 50
3.2.2	Série de Índice Pluviométrico . . . . .	p. 53
3.2.3	Série de Temperatura Mínima . . . . .	p. 54
3.2.4	Correlação cruzada e Função de Transferência . . . . .	p. 56
3.2.5	Comparação de modelos para a série de casos de dengue . . . . .	p. 57
<b>4</b>	<b>Conclusões</b>	p. 60
	<b>Referências</b>	p. 62
	<b>Apêndice 1 – Comparação entre os valores da série e da série ajustada pelos modelos</b>	p. 67
	<b>Apêndice 2 – Comparação de modelos para séries individuais</b>	p. 69

Apêndice 3 – Análise de resíduos para os modelos ajustados de cada série	p. 70
Apêndice 4 – Comparação dos modelos ARIMA com função de transferência	p. 71

# Lista de Figuras

1	Ciclo de vida do <i>Aedes aegypti</i> . . . . .	p. 14
2	Série do número de casos semanais de dengue em Florianópolis . . . . .	p. 45
3	Série semanal do índice pluviométrico em Florianópolis . . . . .	p. 46
4	Série semanal do índice pluviométrico em Florianópolis por ano . . . . .	p. 47
5	Série semanal de temperatura mínima (C°) em Florianópolis . . . . .	p. 48
6	Série semanal de temperatura mínima (C°) em Florianópolis por ano . . . . .	p. 49
7	Correlogramas da FAC e da FACP do número de casos de dengue em Florianópolis . . . . .	p. 51
8	Correlogramas da FAC e FACP do número de casos de dengue em Florianópolis após uma diferenciação . . . . .	p. 52
9	Correlogramas da FAC e FACP do índice pluviométrico em Florianópolis . . . . .	p. 53
10	Correlogramas da FAC e da FACP da temperatura mínima (C°) em Florianópolis . . . . .	p. 54
11	Correlogramas da FAC e da FACP da temperatura mínima (C°) em Florianópolis após uma diferenciação . . . . .	p. 55
12	Função de correlação cruzada para cada série . . . . .	p. 56
13	Série do número de casos de dengue em Florianópolis com previsão para 2022 . . . . .	p. 58
14	Valores observados e valores previstos para o número de casos de dengue em Florianópolis para 2022 . . . . .	p. 59
15	Valores efetivos e valores ajustados para a série do número de casos de dengue em Florianópolis . . . . .	p. 67
16	Valores efetivos e valores ajustados para a série do índice pluviométrico em Florianópolis . . . . .	p. 68

17	Valores efetivos e valores ajustados para a série de temperatura mínima (C°) em Florianópolis . . . . .	p. 68
----	---	-------

# Lista de Tabelas

1	Comportamento da FAC e FACP nos modelos $AR(p)$ , $MA(q)$ e $ARMA(p, q)$	p. 39
2	Comportamento da FAC e FACP em modelos sazonais . . . . .	p. 39
3	Estatísticas descritivas do número de casos semanais de dengue em Florianópolis entre 2017 e 2021 . . . . .	p. 45
4	Estatística Descritiva do índice pluviométrico semanal em Florianópolis de 2017 a 2021 . . . . .	p. 47
5	Estatística Descritiva da temperatura mínima ( $C^\circ$ ) em Florianópolis de 2017 a 2021 . . . . .	p. 49
6	Comparação dos modelos ARIMA para série temporal dos casos de dengue em Florianópolis, através do AIC . . . . .	p. 52
7	Comparação dos modelos . . . . .	p. 57
8	Previsão para o número de casos de dengue semanal na cidade de Florianópolis segundo o modelo $ARIMA(2, 1, 1)$ para as seis primeiras semanas epidemiológicas de 2022 . . . . .	p. 58
9	Comparação dos modelos ARIMA para série temporal do índice pluviométrico em Florianópolis, através do AIC . . . . .	p. 69
10	Comparação dos modelos SARIMA para série temporal de temperatura mínima ( $C^\circ$ ) em Florianópolis, através do AIC . . . . .	p. 69
11	Teste Ljung-Box para resíduos dos modelos . . . . .	p. 70
12	Teste Shapiro-Wilk para resíduos dos modelos . . . . .	p. 70
13	Comparação dos modelos ARIMA com função de transferência com diferentes covariáveis (FT1), através do AIC . . . . .	p. 71

# 1 Introdução

Neste capítulo são apresentadas as características da dengue, suas origens e a forma de disseminação ao longo do tempo, no mundo e no Brasil, com especial foco no estado de Santa Catarina e em sua capital Florianópolis.

## 1.1 A dengue

A dengue é uma arbovirose cujo vírus tem sua origem associada a um vírus que circulava entre os primatas, em locais próximos à península da Malásia. Com a expansão populacional, as habitações regionais ficaram cada vez mais próximas da selva, criando um ambiente propício à transmissão de vírus ancestrais dos primatas aos humanos, através dos mosquitos. Acredita-se que as mutações originaram os quatro tipos de vírus da dengue (UJVARI, 2012).

A dengue é transmitida pela picada da fêmea de mosquitos infectados do gênero *Aedes*, pois o macho se alimenta apenas de seiva de plantas. Suas principais espécies são o *Aedes aegypti*, *Aedes albopictus*, *Aedes scutellaris*, *Aedes africanus* (Theobald) e o *Aedes luteocephalus* (Newstead). Nas Américas, o *Aedes aegypti*, originário da África Subsaariana, é o único transmissor com importância epidemiológica (TIMERMAN; NUNES; LUZ, 2012).

O arbovírus da dengue pode ser um dos quatro sorotipos diferentes: DENV-1, DENV-2, DENV-3 e DENV-4, que pertencem à família *Flaviviridae* (FIOCRUZ, 2010). O indivíduo infectado pelo vírus pode apresentar sintomas como febre, dor de cabeça, dores pelo corpo, náuseas ou até mesmo não apresentar qualquer sintoma. Em casos extremos, o aparecimento de manchas vermelhas na pele, sangramentos, dores abdominais contínuas e vômitos persistentes podem indicar a forma mais grave da doença, a febre hemorrágica. Esse é um quadro grave que necessita intervenção médica urgente, pois pode ser fatal (MINISTÉRIO DA SAÚDE, 2020).

Segundo Beserra et al. (2006), a temperatura favorável para o desenvolvimento do *Aedes aegypti* encontra-se entre 21°C e 29°C, e para a longevidade e fecundidade dos adultos, entre 22°C e 30°C. Fatores como pluviosidade e, principalmente, a temperatura afetam significativamente a longevidade, a fecundidade e a atividade hematofágica do *Aedes* (CALADO; NAVARRO-SILVA, 2002).

Tendo em vista que o ciclo do mosquito transmissor da dengue, apresentado na Figura 1, em geral, leva até 9 dias e o mosquito vive cerca de 40 dias, um total de 49 dias, ou seja, aproximadamente 7 semanas, adota-se neste trabalho 7 tempos passados de cada variável para testagem dos resultados (DENGUETECH, 2020).



Figura 1: Ciclo de vida do *Aedes aegypti*

## 1.2 A dengue no mundo

A dengue chegou ao continente americano a partir do “Velho Mundo”, no final do século 18. O tráfico de escravos e o comércio podem ser os principais responsáveis pela inserção e ampla disseminação do mosquito vindo da África, o *Aedes aegypti*, no novo continente. A espécie mostrou-se altamente adaptável aos ambientes humanos urbanizados e, através da grande circulação marítima à época, acabou se disseminando e chegando as regiões tropicais do mundo. As espécies primeiro infestaram as regiões portuárias e, com os ciclos de urbanização, alcançaram o interior dos continentes. Com a evolução do mosquito, a íntima associação com habitats humanos fez com que essa espécie viesse a se tornar vetor eficiente do vírus da dengue (TIMERMAN; NUNES; LUZ, 2012).

Os primeiros registros de grandes epidemias de doenças que se assemelham à dengue aconteceram nos anos de 1779 e 1780, em três continentes: África, América e Ásia. A principal suspeita é que uma mesma cepa e sorotipo do vírus, chegou aos continentes através das embarcações. Segundo os registros, houve oito pandemias com duração variando entre 3 e 7 anos, entre os anos de 1779 e 1916, e são vistas como consequências de um cenário criado indiretamente pelo desenvolvimento da indústria naval, aumento do tráfego oceânico e pelos acontecimentos referentes à urbanização das áreas de porto (HOWE et al., 1977; TIMERMAN; NUNES; LUZ, 2012).

O *Aedes aegypti* foi erradicado do Mediterrâneo, na década de 1950, e de grande parte das Américas, nos anos 1950 e 1960. No entanto, houve reinfestação na maioria das áreas e, hoje, este vetor é considerado uma espécie “cosmotropical”, e permaneceu em áreas como Venezuela, sul dos Estados Unidos, Guianas e Suriname, além de toda a extensão insular que engloba Caribe e Cuba (FIOCRUZ, 2008; RODHAIN, 1997). O *Aedes aegypti*, que se caracteriza como o vetor primário plenamente adaptado à vida urbana, tem sido detectado de forma cada vez mais ampla nas mais diversas latitudes tropicais e subtropicais (TIMERMAN; NUNES; LUZ, 2012).

O impacto da circulação do vírus da dengue pelo mundo é enorme. Estima-se que cerca de 70% dos casos de dengue estejam associados a uma sintomatologia branda ou sejam assintomáticos. Desta forma, muitas pessoas não procuram os serviços de saúde, levando a um grande número de subnotificações e dificultando a identificação da situação da doença (STANAWAY et al., 2016). Nas Américas, o número de casos de dengue aumentou nas últimas quatro décadas. Enquanto o número de casos acumulados na década de 1980 foi de 1,5 milhão, entre 2010 e 2019 o número chegou a 16,2 milhões. No ano epidêmico de 2013, foram registrados pela primeira vez mais de 2 milhões de casos e uma incidência de aproximadamente 430 casos por cada 100 mil habitantes, além de 37.692 notificações de dengue grave e 1280 mortes (PAHO/WHO, 2022).

Segundo o *CDC (Center for Disease Control)*, existem quatro principais razões para o crescimento do número de casos de dengue pelo mundo:

- Significativas mudanças demográficas globais, tendo como consequência uma urbanização incontrolada e adensamento populacional em grandes cidades, que levaram a um aumento na quantidade de moradias inadequadas e saneamento básico precário.
- O baixo investimento em programas de prevenção e vigilância epidêmica e uma ênfase na implantação de métodos de controle de emergência.

- O aumento de viagens aéreas e a redução da distância geográfica entre os locais é um mecanismo ideal de transporte do vírus por pessoas infectadas, com consequente intercâmbio destes vírus da dengue com outros patógenos.
- Na maioria dos países com epidemia de dengue não existem programas de fato eficazes para o controle do mosquito.

### 1.3 A dengue no Brasil

No Brasil, a doença chegou durante o século 18. Há referências sobre a dengue desde 1916, em São Paulo, e em 1923, em Niterói, ambas sem diagnóstico laboratorial (MEIRA, 1916; PEDRO, 1923). Na década de 1980, quase todas as cidades do litoral apresentaram registro do mosquito *Aedes aegypti* que, nos anos seguintes, já se encontravam no interior do país. A primeira evidência de epidemia de dengue é de 1982, em Boa Vista, capital do estado de Roraima, quando foram isolados os sorotipos DENV1 e DENV4, e, após inquérito sorológico, foram contabilizadas 11 mil pessoas infetadas pelo vírus (OSANAI et al., 1983).

A partir de 1990, houve aumento significativo da incidência de casos de dengue, reflexo do aumento da circulação do sorotipo DENV1 e da introdução do DENV2 no Rio de Janeiro. O número de casos no estado atingiu 613,8 por 100 mil habitantes em 1991. Neste mesmo período surgem os primeiros registros de dengue hemorrágica, onde foram registrados 462 casos e 8 óbitos (TEIXEIRA et al., 2005). No último ano da década, a sazonalidade das infecções pelo vírus tornou-se muito evidente, e este padrão sazonal em território nacional, que nem sempre é observado em outros países, tem sido explicado pelo aumento da densidade das populações de *Aedes aegypti* em virtude das condições de temperatura e umidade propícias para ploriferação, observadas em grandes extensões do país (TEIXEIRA; BARRETO; GUERRA, 1999).

No ano 2000, o sorotipo DENV3 foi observado pela primeira vez no Brasil, no Rio de Janeiro, e no ano seguinte, no estado de Roraima. Em 2002, foi registrada até então a maior incidência da doença, com aproximadamente 800 mil casos notificados. Nesta epidemia, 80% dos casos do continente americano aconteceram em território nacional, o que levou a uma dispersão rápida do novo sorotipo (NOGUEIRA et al., 2005).

A repercussão das epidemias de dengue é percebida diretamente nas comunidades onde ocorrem, geram gastos elevados com hospitalizações e implicam em uma sobrecarga no sistema de saúde (SUAYA et al., 2009). Nos anos de 2000 a 2007, entre todos os países do

ocidente, o Brasil apresentou os maiores gastos anuais agregados induzidos pela dengue, com média de US\$ 1,35 bilhão/ano, considerando custos diretos médicos e não médicos, e custos indiretos decorrentes da perda de produtividade (SHEPARD et al., 2011).

Os ciclos endêmicos e epidêmicos da doença costumam ocorrer a cada 4 ou 5 anos, e mais de 7 milhões de casos já foram notificados no território brasileiro (FIOCRUZ, 2010). As dinâmicas dos fatores que incidem na circulação do vírus da dengue são complexas e frequentemente estudadas (BHATT et al., 2013). Porém, é de consenso nas obras literárias que há a influência de fatores como variações no clima, crescimento da população e do fluxo de pessoas e mercadorias entre países (GUBLER, 2011). Segundo Timerman, Nunes e Luz (2012), os fatores condicionantes da propagação da dengue nas Américas e no Brasil são semelhantes e referem-se, em grande parte, ao modelo de crescimento econômico implementado, caracterizado pelo crescimento desordenado dos centros urbanos.

Segundo Codeço, Falcão e Granato (2022), o ano de 2019 foi marcado por uma grande epidemia de dengue, e com o conhecimento do comportamento cíclico da doença, era esperado um novo aumento a partir de 2022. No entanto, o sistema de vigilância foi muito prejudicado nos últimos anos, tendo em vista o foco quase absoluto na pandemia de COVID-19 e uma epidemia mais recente de influenza H3N2, dificultando as projeções. Observa-se uma atividade expressiva da dengue em algumas partes do país, em particular o eixo que vai do Tocantins até Santa Catarina, passando pelo Centro-Oeste e por porção oeste de São Paulo. O espalhamento da doença na região Sul do país, onde ela era pouco ativa ou quase inexistente, pode estar associada as mudanças climáticas e a própria adaptação do mosquito.

### 1.3.1 Região Sul e os casos de dengue

Vários estudos de modelagem tentaram prever o futuro global da dengue. No entanto, as projeções resultantes são difíceis de comparar e às vezes são contraditórias porque os modelos diferem em sua abordagem, na qualidade dos dados da doença que eles usam e na escolha das variáveis que direcionam a distribuição da doença (MESSINA et al., 2015). Portanto, cada localidade apresentará particularidades na qual um determinado modelo se ajustará melhor, e na busca por um modelo enxuto surgirão covariáveis diferentes, por exemplo, variáveis territoriais, sociais, meteorológicas, dentre outras.

A região Sul do Brasil, por apresentar um clima predominantemente subtropical, com temperaturas mais amenas e uma estrutura social mais inclusiva, sempre contribuiu com poucos casos de dengue e era tratada como uma região de improvável disseminação da

doença (OLIVEIRA; OLIVEIRA, 2019). Após a reintrodução da dengue no Brasil, casos da doença, só voltaram a surgir a partir do ano 2000, o que demonstra disparidade em relação ao restante do país (CÂMARA et al., 2007). Segundo Kudrna et al. (2021), a taxa de incidência de dengue na região era menor que a taxa no Brasil até 2019 e, pela primeira vez, em 2020, a incidência da doença ali foi maior que a média nacional, evidenciando a disseminação do mosquito vetor e risco epidemiológico.

O estado de Santa Catarina, tem notificado continuamente casos de dengue a partir do ano de 2011, e registrou epidemias nos anos de 2015, 2016, 2019, 2020 e 2021. O ano de 2022 registra o maior número de casos de dengue de todos os tempos no estado. Já são 32.206 casos confirmados nos primeiros cinco meses do ano, onde 28.752 são autóctones, ou seja, foram contraídos dentro do estado. Sua capital, Florianópolis, que é considerada infestada pelo mosquito transmissor da dengue desde 2015, decretou estado de emergência em saúde pública e epidemia em abril de 2022, quando já teriam sido registrados mais de 3 mil focos do mosquito *Aedes aegypti* e 329 pessoas diagnosticadas com a doença (DIVE-SC, 2022; G1-SC, 2022).

## 1.4 Fundamentação Teórica

A ocorrência de dengue foi analisada em várias cidades brasileiras através da modelagem de séries temporais com frequência mensal. Esse levantamento trouxe à luz a informação sobre algumas possíveis covariáveis que estariam associadas à quantidade de casos de dengue.

Bicalho, Safadi e Charret (2014) realizaram ajuste dos dados para as cidades de Lavras e Cuiabá através de um modelo SARIMA. Foi observado que a precipitação e a temperatura máxima influenciavam o número de casos de dengue na cidade de Cuiabá. Já em Lavras, além dessas variáveis, a temperatura média e a umidade também foram identificadas como fatores importantes para a ocorrência da dengue.

Secches (2019) analisou *clusters* espaço-temporais no estado de Minas Gerais em casos confirmados de dengue em todas as 853 cidades, de janeiro de 2008 até março de 2018. A sazonalidade e a não estacionaridade nos casos de dengue foram bem ajustadas através de um modelo SARIMA, onde a doença apresentava maior número de casos no verão e variáveis como a quantidade de chuva e temperatura favoreciam o aumento de casos. Neste trabalho, foi encontrado um modelo bastante eficiente para detectar municípios com casos confirmados de dengue, cerca de 61% de todos os municípios de Minas Gerais no

período de estudo, mostrando sensibilidade de detecção de *clusters* espaço-temporais.

Santos e Tavares (2014) realizaram um estudo no estado de Minas Gerais utilizando modelos de alisamento exponencial para previsão dos números de casos de dengue nos municípios para o ano de 2009, tomando como base os dados coletados no período de 2001 a 2008. O estudo também apresentou mapas de riscos obtidos através do modelo idealizado e, quando comparados com os dados reais do ano de 2009, foram encontradas boas estimativas para os casos de dengue.

Lizzi (2012) realizou um estudo em Campinas e Ribeirão Preto, municípios do estado de São Paulo, analisando os casos notificados e confirmados de dengue no período de 1998 a 2009. Utilizou-se modelos SARIMA com extensões que permitem incorporar covariáveis e mostraram-se eficientes ao incluir a variável precipitação observada nos meses anteriores aos casos mensais de dengue. A análise utilizou uma transformação logarítmica para melhor adequação ao modelo, e trouxe previsões razoáveis para um período de previsão de um ano além do que havia sido considerado na modelagem. Além disso, foi comentado nesse estudo que componentes de média móvel e autoregressivas podem variar de uma população para outra e definir diferentes padrões temporais, de acordo com as características de cada local.

## 1.5 **Objetivos**

O objetivo do presente trabalho é investigar, através de modelos Box e Jenkins, uma possível associação entre a quantidade de casos notificados de dengue em Florianópolis - SC no período de 2017 a 2021 e algumas variáveis descritoras do clima da região, como temperatura mínima e índice pluviométrico. A partir dos modelos construídos espera-se poder, também, avaliar a sua capacidade preditiva.

### 1.5.1 **Objetivos específicos**

- Identificar modelos da família ARIMA para cada uma das séries temporais analisadas;
- Investigar a existência de correlação temporal entre a série de contagem de casos de dengue e as demais séries temporais existentes na base de dados;
- Comparar a capacidade preditiva de modelos de Box e Jenkins, incorporando ou não covariáveis através do uso de função de transferência.

## 1.6 Organização

No Capítulo 2 são apresentadas características da área de estudo (cidade de Florianópolis), seguido das variáveis incluídas nas bases de dados utilizadas. Ainda no Capítulo 2, são apresentados conceitos básicos da modelagem de séries temporais e os pressupostos para utilização dos modelos de Box e Jenkins, tendo como caso especial a apresentação dos modelos ARIMA e SARIMA, além da correlação cruzada e função de transferência. No Capítulo 3 encontram-se os resultados derivados da aplicação dos modelos propostos e no Capítulo 4 são apresentadas as conclusões.

## 2 Materiais e Métodos

Neste capítulo são apresentadas as características da cidade de Florianópolis e as variáveis das bases de dados utilizadas para o estudo. Além disso, são apresentados os conceitos básicos da modelagem de séries temporais e os pressupostos para utilização da metodologia de Box e Jenkins.

### 2.1 Local de estudo e base de dados

A cidade de Florianópolis está localizada na região Sul do Brasil, no estado de Santa Catarina, entre as coordenadas 27°10' a 27°50'S e 48°25' a 48°35'O. A capital catarinense apresenta área territorial de aproximadamente 675 km<sup>2</sup> e uma população estimada de 521.524 habitantes. Na parte econômica e social, ocupa posição de destaque no estado, sendo a cidade com maior Índice de Desenvolvimento Humano Municipal (IDHM), 0,847 e Produto Interno Bruto (PIB) per capita de aproximadamente R\$43.842. Além disso, apresenta 87,8% de domicílios com esgotamento sanitário adequado (IBGE, 2021).

O principal bioma da cidade é a Mata Atlântica, e o clima predominante é o subtropical úmido. As estações do ano são bem caracterizadas, apresentando verão e inverno marcantes e a temperatura média anual de 20° C. O índice pluviométrico anual apresenta uma precipitação em torno de 1.600 mm no norte da ilha e 1400 mm no sul da ilha, o que justifica um clima com umidade relativa anual de 85%, apresentando uma média de 140 dias de chuvas por ano (ANDRADE et al., 1996; RÉGIS et al., 2016).

Neste estudo foram utilizadas duas bases de dados distintas, e estão representadas através de três séries temporais que serão descritas a seguir. Os dados referem-se a cidade de Florianópolis para cada semana epidemiológica, no período de 2017 a 2021, além das seis primeiras semanas epidemiológicas do ano de 2022, sendo que estas últimas não participaram do processo de modelagem e serviram de base para a avaliação da capacidade preditiva dos modelos.

### 2.1.1 Casos de dengue e temperatura mínima

Essas duas primeiras séries temporais foram obtidas através do sistema de alerta para arboviroses baseado em dados híbridos gerados por meio da análise integrada de dados minerados a partir da *web social* e de dados climáticos e epidemiológicos.

Em 2021, o sistema ganhou amplitude nacional com o apoio do Ministério da Saúde realizando análises em nível estadual. Com isso, mais secretarias passaram a receber semanalmente os boletins do InfoDengue. Implementado em 2015, o sistema foi desenvolvido por pesquisadores do Programa de Computação Científica (Fundação Oswaldo Cruz, RJ) e da Escola de Matemática Aplicada (Fundação Getúlio Vargas) com a forte colaboração da Secretaria Municipal de Saúde do Rio de Janeiro, do Observatório da Dengue/UFMG e pesquisadores da Universidade Federal do Paraná e da Universidade Estadual do Oeste do Paraná (INFODENGUE, 2022).

Os dados do sistema InfoDengue incluem casos de dengue, chikungunya e zica, que são doenças de notificação obrigatória, isso é, o profissional de saúde que diagnostica um caso suspeito precisa preencher uma ficha de notificação que alimenta um banco de dados municipal que depois é consolidado a nível estadual e finalmente, a nível federal pelo Ministério da Saúde. Apenas uma fração desses casos são confirmados laboratorialmente, a maioria recebe classificação final com base em critérios clínico-epidemiológicos. A partir dos casos notificados, são calculados os indicadores de incidência que alimentam o InfoDengue. Os dados de temperatura são obtidos das estações meteorológicas de aeroportos assim como de imagens de satélite (CODECO et al., 2018).

### 2.1.2 Índice pluviométrico

A série de índice pluviométrico foi obtida através do Instituto Nacional de Meteorologia (INMET) baseado em dados de estações convencionais da região das quais destacam-se: [D2913] Areias Campeche, [D2915] Rodovia SC406, [D2916] Costeira do Pirajubae, [D2917] Canasvieiras, [D2918] Rio Vermelho e [83899] Aeroporto.

A rede de estações meteorológicas automáticas utiliza o que há de mais moderno internacionalmente. Os dados coletados por essa rede são disseminados, de forma democrática e gratuita, em tempo real, e têm aplicação em todos os setores da economia, de modo especial no agropecuário e em apoio à Defesa Civil. O sistema de coleta e distribuição de dados meteorológicos do instituto (temperatura, umidade relativa do ar, direção e velocidade do vento, pressão atmosférica, precipitação, entre outras variáveis) é

dotado de estações de sondagem de ar superior (radiossonda); estações meteorológicas de superfície, operadas manualmente; e a maior rede de estações automáticas da América do Sul (INMET, 2022).

## 2.2 Modelagem de séries temporais

Segundo Morettin e Tolo (2006), uma série temporal é qualquer conjunto de observações ordenadas no tempo. As observações em instantes próximos costumam possuir dependência e, em algumas séries, podem ocorrer, também, dependências relacionadas a ciclos ou efeitos periódicos, muitas vezes relacionados a fenômenos que se repetem a cada ano. (SOUZA; ASSIS, 1989).

As séries temporais podem ser classificadas quanto ao tempo de observação, entre contínuas e discretas. No primeiro caso, as observações são feitas continuamente em um determinado período, enquanto as realizadas em tempos distintos, geralmente equiespaçadas, são discretas (CHATFIELD, 2019).

Além disso, a classificação da série temporal também pode ser feita quanto ao número de variáveis de interesse modeladas, sendo univariada quando houver apenas uma variável sob estudo e multivariada, quando houver interesse na modelagem conjunta de duas ou mais séries.

O objetivo da análise de séries temporais é descrever matematicamente o comportamento passado da série, verificando periodicidades e o seu mecanismo gerador, com intuito de realizar previsões de movimentos futuros (SOUZA; ASSIS, 1989).

O modelo de uma série temporal,  $W_t$ , pode ser descrito, em muitos casos, pela seguinte soma:

$$W_t = T_t + S_t + a_t \quad , \quad t = 1, 2, \dots, n, \quad (2.1)$$

onde  $T_t$  representa a tendência,  $S_t$  é identificada como componente sazonal (ou sazonalidade) e  $a_t$  é um choque aleatório.

Um processo estocástico é caracterizado como uma família de variáveis aleatórias controladas pelas leis da Teoria de Probabilidades. Esse modelo é utilizado para descrever as séries observadas em um determinado espaço de tempo.

**Definição 2.1** *Seja  $N$  um conjunto arbitrário de período de tempos. Um processo estocástico é uma família  $W = \{W_t, t \in T\}$ , tal que, para cada  $t \in N$ ,  $W_t$  é uma variável aleatória.*

Um processo estocástico  $W$  que gerou as observações é considerado estacionário se as características do processo não se alteram com o decorrer do tempo.

**Definição 2.2** *Um processo estocástico  $W = \{W_t, t \in N\}$  diz-se estritamente estacionário se todas as distribuições finito-dimensionais permanecem as mesmas sob translações no tempo, ou seja,*

$$F(w_1, \dots, w_n; t_1 + r, \dots, t_n + r) = F(w_1, \dots, w_n; t_1, \dots, t_n), \quad (2.2)$$

para quaisquer  $t_1, \dots, t_n, r \in \mathbb{N}$ .

Sendo assim, todas as distribuições unidimensionais não variam ao longo do tempo, portanto a média e a variância são constantes, isto é,

$$\mu_t = \mu, \quad V_t = \sigma^2, \quad \forall t \in \mathbb{N}.$$

A suposição que normalmente é feita é que,  $\mu = 0$ , mas, pode-se sempre considerar o processo  $\{W_t - \mu\}$  como tendo média zero (MORETTIN; TOLOI, 2006).

Um exemplo de processo estacionário é o ruído branco, que consiste em uma sequência de variáveis aleatórias que não possuem correlação temporal, apresentam média zero e variância finita e constante ao longo do tempo.

A existência de tendência e/ou de sazonalidade implicam automaticamente que a série observada é não-estacionária.

Caso a série temporal seja não-estacionária, pode ser necessário transformar os dados. Segundo Morettin e Tolo (2006), a maneira mais usual de obter estacionariedade, é através da tomada de diferenças sucessivas da série original. A primeira diferença de  $W_t$  é dada pela expressão:

$$\Delta W_t = W_t - W_{t-1} \quad (2.3)$$

Caso a série ainda seja não-estacionária, é realizada a segunda diferença, expressa por:

$$\Delta^2 W_t = \Delta[\Delta W_t] = \Delta[W_t - W_{t-1}] = W_t - 2W_{t-1} + W_{t-2} \quad (2.4)$$

De modo geral, a  $j$ -ésima diferença de  $W_t$  é:

$$\Delta^j W_t = \Delta[\Delta^{j-1} W_t] \quad (2.5)$$

A verificação da necessidade de transformação da série, para a obtenção da desejada propriedade de estacionariedade através da diferenciação, pode ser executada de formas diferentes. No presente trabalho, foi utilizado o teste aumentado de Dickey-Fuller (ADF), apresentado a seguir.

A verificação da presença de raízes unitárias por meio do teste ADF, é equivalente à validação da existência ou não de estacionariedade na série. A metodologia utilizada no teste é baseada na avaliação da significância dos parâmetros da seguinte equação:

$$\Delta w_t = \beta_0 + \beta_1 t + \delta w_{t-1} + \sum_{i=1}^m a_i \Delta w_{t-i} + \varepsilon_t \quad (2.6)$$

onde

- $\Delta w_t$ : diferenciação da série  $W_t$ ;
- $\beta_0$ : intercepto da série;
- $\beta_1$ : coeficiente de tendência;
- $\delta$ : coeficiente de existência de raiz unitária;
- $m$ : número de defasagens tomadas na série;
- $\varepsilon_t$ : processo de ruído branco

O teste verifica as seguintes hipóteses para a série:

$$\begin{cases} \mathbf{H}_0 : \text{Existência de raiz unitária (a série não é estacionária)} \\ \mathbf{H}_1 : \text{Não existência de raiz unitária (a série é estacionária)} \end{cases}$$

O nível de significância adotado para este teste é, em geral, igual a 5%. Se o p-valor for menor ou igual a 0,05 rejeita-se a hipótese nula do teste e a série pode ser considerada estacionária.

A estatística de teste é descrita pela razão entre o estimador de mínimos quadrados de  $\delta$  e o seu erro padrão, representados por  $\hat{\delta}$  e  $se(\hat{\delta})$  respectivamente:

$$T = \frac{\hat{\delta}}{se(\hat{\delta})} \quad (2.7)$$

Os valores críticos da estatística T foram tabelados por Dickey e Fuller (1979) através da simulação de Monte Carlo e variam em casos de presença somente do intercepto ( $\beta_0$ ), presença somente do coeficiente de tendência e presença de ambos.

Se a série tiver sido considerada estacionária, a modelagem pode ser feita utilizando a abordagem proposta por Box e Jenkins (1970), que será apresentada posteriormente.

## 2.2.1 Componentes da série temporal

Nas Subseções seguintes são apresentadas as componentes de tendência e sazonalidade de uma série temporal. Além disso, são apresentados testes utilizados para verificar a presença de tais componentes na formação de um processo.

### 2.2.1.1 Tendência

A tendência de uma série temporal, indicada neste trabalho por  $T_t$ , é definida como um padrão de crescimento/decrescimento da variável estudada em um certo período de tempo. Segundo Oliveira (2007), os objetivos básicos na identificação da tendência são: avaliar o comportamento observado para auxiliar em previsões, removê-la da série para facilitar a visualização de outros componentes, ou ainda, identificar a faixa de valor específica que a variável assume, caso não seja observado comportamento sistematicamente crescente ou decrescente.

Supondo que somente a componente de tendência  $T_t$  está presente na série temporal deste trabalho, a equação 2.1 pode ser reescrita como:

$$W_t = T_t + a_t \quad t = 1, 2, \dots, n.$$

Segundo Morettin e Toloi (2006), a modelagem de  $T_t$  pode ser feita, por exemplo, através dos seguintes métodos:

- Ajuste de uma função linear exponencial ou polinomial do tempo.
- Suavização ou filtragem dos valores da série ao redor de um ponto para estimar a tendência naquele ponto (através de métodos baseados em médias móveis).
- Suavização dos valores da série através de sucessivos ajustes de retas de mínimos quadrados.

O teste de Mann-Kendall é um teste não-paramétrico, sugerido pela Organização Meteorológica Mundial (OMM) para avaliação da tendência em séries temporais de dados ambientais. O teste verifica se uma série temporal possui uma tendência temporal, testando as seguintes hipóteses (HIPEL; MCLEOD, 1994):

$$\begin{cases} \mathbf{H}_0 : \text{Não existe tendência na série de dados} \\ \mathbf{H}_1 : \text{Existe tendência na série de dados} \end{cases}$$

Sendo assim, a estatística do teste é dada por:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sng}(w_j - w_i), \quad (2.8)$$

sendo  $n$  o número de observações,  $w_i$  e  $w_j$  os respectivos valores das observações em séries de tempo  $i$  e  $j$  ( $j > i$ ) e  $\text{sng}(w_j - w_i)$  dada por:

$$\text{sng}(w_j - w_i) = \begin{cases} +1, & \text{se } (w_j - w_i) > 0 \\ 0, & \text{se } (w_j - w_i) = 0 \\ -1, & \text{se } (w_j - w_i) < 0 \end{cases}$$

Segundo Barbieri et al. (2017), para  $n$  grande, a distribuição de  $S$  tende para a distribuição Normal com média zero e variância dada por:

$$\text{Var}(S) = \frac{n(n-1)(2n+5) - \sum_{t=1}^n t(t-1)(2t+5)}{18} \quad (2.9)$$

em que  $t$  é o número de dados com valores iguais em certo grupo.

Identificados os valores da média e variância, a estatística do teste de Mann-Kendall é baseada no valor da variável  $Z$ , calculada conforme a seguinte equação:

$$Z = \begin{cases} \frac{S-1}{\sqrt{Var(S)}}, & \text{se } S > 0 \\ 0, & \text{se } S = 0 \\ \frac{S+1}{\sqrt{Var(S)}}, & \text{se } S < 0 \end{cases} .$$

Realizando um teste bilateral, ao nível de significância 5%, rejeita-se a hipótese nula  $H_0$  (não existe tendência na série) quando o p-valor for menor ou igual ao nível de significância. Caso haja a rejeição de  $H_0$ , analisa-se o sinal da estatística  $Z$ , concluindo que a tendência é positiva ( $Z > 0$ ) ou negativa ( $Z < 0$ ).

### 2.2.1.2 Sazonalidade

A sazonalidade de uma série temporal, indicada neste trabalho por  $S_t$ , é definida como um fenômeno que ocorre ao longo da série e que se repete a cada período fixo de tempo.

Segundo Morettin e Toloí (2006), existem dois tipos de sazonalidade:

- Determinística: Quando se pressupõe a existência de um padrão sazonal regular e estável no tempo.
- Estocástica: Quando a componente sazonal da série varia com o tempo.

De acordo com Morettin e Toloí (2006) para verificar a presença da componente sazonal na série pode-se utilizar, entre outros, o teste não paramétrico de Kruskal-Wallis, que verifica as seguintes hipóteses:

$$\begin{cases} H_0 : \text{As amostras (grupos) são de populações idênticas} \\ H_1 : \text{Pelo menos uma das amostras (grupos) vem de uma população diferente das outras} \end{cases}$$

Tem-se que as observações são dadas por:

$$Z_{ij}, \quad j = 1, \dots, k, \quad i = 1, \dots, n_j, \quad (2.10)$$

e que a quantidade total de observações é dada por  $N = \sum_{j=1}^k n_j$ .

Tem-se que  $k$  é o número de amostras. As  $N$  observações das  $k$  amostras são ordenadas e considera-se  $R_{.j}$  como sendo a soma dos postos associados à  $j$ -ésima amostra, i.e.

$$R_{.j} = \sum_{i=1}^{n_j} R_{ij}, j = 1, \dots, k. \quad (2.11)$$

Sendo assim, a estatística do teste é dada por:

$$T_1 = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_{.j}^2}{n_j} - (3N+1). \quad (2.12)$$

Rejeita-se  $H_0$  quando essa estatística for maior ou igual ao valor crítico  $T_{1c}$ , onde  $T_{1c}$  é tal que  $P_{H_0}(T_1 \geq T_{1c}) = \alpha$ , onde  $\alpha$  é o nível de significância adotado.

Para  $n_j$  suficientemente grande, ou  $k \geq 4$ , sob  $H_0$ , a distribuição de  $T_1$  pode ser aproximada por uma distribuição  $\chi^2$  com  $(k-1)$  graus de liberdade.

## 2.2.2 Metodologia de Box e Jenkins

A metodologia proposta por Box e Jenkins (1970) foi utilizada neste trabalho para a identificação e estimação de modelos paramétricos para séries temporais, em especial modelos  $ARIMA(p, d, q)$  e  $SARIMA(p, d, q) \times (P, D, Q)_s$ , de tal forma que seja possível encontrar após a modelagem, o melhor modelo e avaliar sua capacidade preditiva.

O método proposto possui três principais etapas:

- **Identificação:** Através da análise de correlograma e funções de autocorrelação e autocorrelação parcial, são avaliadas as correlações em  $k$  lags e identificados possíveis valores para os parâmetros  $p$ ,  $d$  e  $q$ ;
- **Validação do modelo:** Nesta segunda avaliação, é feita a estimação e verificação da significância dos parâmetros da parte autoregressiva e de médias móveis, a fim de obter o melhor modelo para descrever o comportamento da série. Além disso, o modelo também pode ser escolhido a partir da qualidade de seu ajuste à série observada, depois checando se os resíduos assemelham-se a ruídos brancos. Este último passo consiste em verificar a independência e a normalidade dos resíduos. Uma forma de verificar a independência dos resíduos é através do teste Ljung e Box (1978), enquanto a verificação da normalidade pode ser feita através do teste Shapiro e Wilk (1965), ambos testes e hipóteses serão apresentados posteriormente;

- **Previsão:** A fim de testar a capacidade preditiva do modelo, o último passo da metodologia consiste em obter estimativas para valores futuros da série estudada.

Caso o modelo não seja considerado adequado, em termos de validação ou de previsão, todo o ciclo recomeça e um novo modelo é construído.

### 2.2.2.1 ARIMA

Segundo Morettin e Tolo (2006), os modelos para séries temporais podem ser classificados em duas classes, segundo o número de parâmetros envolvidos. A primeira classe, são os modelos paramétricos, que contêm um número finito de parâmetros e análise feita no domínio do tempo, dos quais destacam-se os modelos autoregressivo integrado de média móvel (ARIMA) e sazonal autorregressivo integrado de média móvel (SARIMA). A segunda classe, compreende os modelos não paramétricos, que envolvem um número infinito de parâmetros e análise feita no domínio de frequências.

Os modelos autoregressivos  $AR(p)$  foram criados de forma que o valor mais recente da série  $W_t$  possa ser explicado como função das  $p$  observações anteriores,  $W_{t-1}, W_{t-2}, \dots, W_{t-p}$  (PADILHA et al., 2013). Logo, diz-se que a série temporal  $W_t$  segue um processo autoregressivo, se puder se escrito da seguinte maneira:

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + \phi_p W_{t-p} + \alpha_t, \quad (2.13)$$

sendo  $\phi_1, \dots, \phi_p$  os coeficientes do modelo e  $\alpha_t$  um ruído branco com média 0 e variância  $\sigma^2$ .

Os modelos de médias móveis  $MA(q)$  são formados por combinações lineares de ruídos brancos ocorridos no período corrente e em períodos passados (SOUZA et al., 2007). A estrutura da série temporal  $W_t$  é expressa por:

$$W_t = \alpha_t - \theta_1 \alpha_{t-1} - \dots - \theta_q \alpha_{t-q}, \quad (2.14)$$

onde  $\theta_i$  ( $i = 1, 2, \dots, q$ ) são os coeficientes do modelo,  $q$  a estrutura de média móvel e  $\alpha_t$  um ruído branco com média 0 e variância  $\sigma_a^2$ .

O modelo  $ARMA(p, q)$  utiliza a junção dos modelos de classe autoregressivos,  $AR(p)$ , e de médias móveis,  $MA(q)$ , com objetivo de oferecer uma descrição parcimoniosa de um processo  $W_t$ .

A estrutura da série temporal estacionária  $W_t$ , função de suas  $p$  observações anteriores e dos ruídos brancos correntes e passados, é descrita da seguinte forma:

$$W_t = \phi_1 W_{t-1} + \dots + \phi_p W_{t-p} + \alpha_t - \theta_1 \alpha_{t-1} - \dots - \theta_q \alpha_{t-q}. \quad (2.15)$$

A utilização desse modelo é recorrente em análise de séries temporais, e tem por objetivo, auxiliar no entendimento do comportamento da série ao longo do tempo e, dependendo da sua qualidade, realizar previsão de valores futuros.

A aplicação do filtro de integração I, dependente do parâmetro  $d$ , é realizada somente nos casos em que a série temporal que está sendo modelada não satisfaz à condição de estacionariedade, passando do uso de um modelo ARMA para um modelo ARIMA. A ideia por detrás desse tipo de filtro é que a série pode se tornar estacionária pela aplicação de sucessivas diferenças.

O modelo autoregressivo integrado de média móvel, conhecido pela sigla ARIMA( $p, d, q$ ), é uma generalização do modelo ARMA( $p, q$ ) onde a série temporal  $W_t$  é não estacionária e, portanto, é realizada a aplicação do filtro de integração. O processo consiste em diferenciar-la  $d$  vezes até se obter uma série  $\Delta^d W_t$  estacionária.

Dessa forma, diz-se que  $W_t$  segue um ARIMA( $p, d, q$ ), se  $Z_t = \Delta^d W_t$  for estacionário e  $Z_t$  for da forma ARMA( $p, q$ ), onde:

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + \alpha_t - \theta_1 \alpha_{t-1} - \dots - \theta_q \alpha_{t-q}. \quad (2.16)$$

### 2.2.2.2 SARIMA

O modelo de série temporal, visto na Equação (2.1), pode vir à apresenta em sua composição componentes que influenciam seu comportamento ao longo do tempo. A componente sazonal  $S_t$ , se repete na série a cada  $s$  períodos de tempo.

O modelo sazonal autorregressivo integrado de média móvel, conhecido pela sigla SARIMA( $p, d, q$ )  $\times$  ( $P, D, Q$ ) $_s$ , e introduzido por Box e Jenkins (1970), é uma variação sazonal do modelo ARIMA que considera a presença da sazonalidade na modelagem da série ao invés de eliminá-la.

Quando um processo  $W_t$  possui comportamento sazonal estocástico de período 12, por exemplo, admite-se que:

$$W_t = \mu_t + N_t, \quad (2.17)$$

onde  $\mu_t$  corresponde

$$\mu_t = \mu + \sum_{j=1}^6 \left[ \alpha_j \cos\left(\frac{2\pi jt}{12}\right) + \beta_j \sin\left(\frac{2\pi jt}{12}\right) \right]$$

$$\mu_t = \mu_{t-12}$$

$$\mu_t - \mu_{t-12} = (1 - B^{12})\mu_t$$

assim,

$$(1 - B^{12})W_t = (1 - B^{12})\mu_t + (1 - B^{12})N_t$$

$$(1 - B^{12})W_t = Y_t + (1 - B^{12})N_t$$

e, por  $Y_t$  e  $N_t$  serem processos estacionários

$$\phi_Y(B)Y_t = \phi_Y(B)\alpha_t,$$

$$\phi_N(B)N_t = \phi_N(B)\epsilon_t,$$

sendo,  $\alpha_t$  e  $\epsilon_t$  ruídos brancos independentes.

A Equação (2.17) pode ser descrita, ainda, como:

$$\Phi(B^{12})\Delta_{12}^D W_t = \Theta(B^{12})\alpha_t \quad (2.18)$$

onde

- $B$ : é o operador retardo, aplicado a função, atrasa a função em um período;
- $\Phi(B^{12})$ : é o operador autorregressivo sazonal de ordem  $P$ ;
- $\Theta(B^{12})$ : é o operador de médias móveis sazonal de ordem  $Q$ ;
- $\Delta_{12}$ : é o operador diferença sazonal tal que o valor  $D$  (em  $\Delta_{12}^D$ ) indica o número de diferenças sazonais que foram aplicadas;
- $\alpha_t$ : ruído branco,

Neste caso:

$$\Phi(B^{12}) = (1 - \Phi_1 B^{12} - \Phi_2 B^{12*2} - \dots - \Phi_P B^{12*P});$$

$$\Theta(B^{12}) = (1 - \Theta_1 B^{12} - \Theta_2 B^{12*2} - \dots - \Theta_Q B^{12*Q});$$

$$\Delta_{12} = (1 - B^{12})$$

Supondo-se, agora, que o processo  $\alpha_t$  na Equação (2.18) satisfaça um modelo ARIMA( $p, d, q$ )

$$\varphi(B)\alpha_t = \theta(B)\alpha_t, \quad (2.19)$$

onde,  $\varphi(B) = (1 - B)^d\phi(B)$  e  $\alpha_t$  é um processo ruído branco.

Então,  $W_t$  satisfaz o modelo

$$\phi(B)\Phi(B^{12})(1 - B^{12})^D(1 - B)^dW_t = \theta(B)\Theta(B^{12})\alpha_t, \quad (2.20)$$

onde

$$\begin{aligned} \theta(B) &= 1 - \theta_1B - \dots - \theta_qB^q, \\ \phi(B) &= 1 - \phi_1B - \dots - \phi_pB^p \end{aligned}$$

O modelo descrito na Equação (2.20) pode ser denominado ARIMA sazonal multiplicativo (SARIMA) de ordem  $(p, d, q) \times (P, D, Q)_{12}$ .

### 2.2.2.3 Função de Autocovariância

**Definição 2.3** *Seja  $\{W_t, t \in \mathbb{Z}\}$ , um processo estacionário real discreto, de média zero.*

*A função de autocovariância (FACV) é dada por:*

$$\gamma_k = E(W_tW_{t+k}) = Cov(W_t, W_{t+k})$$

**Proposição 2.1** *A FACV  $\gamma_k$  satisfaz as seguintes propriedades:*

- (i)  $\gamma_0 > 0$ ;
- (ii)  $\gamma_{-k} = \gamma_k$ ;
- (iii)  $|\gamma_k| \leq \gamma_0, \forall k \in \mathbb{Z}$ ;
- (iv) *para quaisquer valores  $t_1, t_2, \dots, t_n \in T$ , tem-se que*

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \gamma_{(|t_i - t_j|)} \geq 0$$

*em que  $\alpha_1, \alpha_2, \dots, \alpha_n$  são constantes reais.*

### 2.2.2.4 Função de Autocorrelação e Função de Autocorrelação Parcial

A função de autocorrelação (FAC) é uma ferramenta utilizada para identificar as propriedades de uma série temporal e medir a correlação entre as observações que são separadas por  $k$  períodos de tempo. Esta função é definida por:

$$\rho_k = \text{Corr}\{W_t, W_{t+k}\} = \frac{\gamma(|(t+k) - t|)}{\sqrt{\gamma_0}\sqrt{\gamma_0}} = \frac{\gamma_k}{\gamma_0} \quad (2.21)$$

**Proposição 2.2** A FAC  $\rho_k$  satisfaz as propriedades:

- (i)  $\rho_0 = 1$ ;
- (ii)  $\rho_{-k} = \rho_k$ ;
- (iii)  $|\rho_k| \leq 1, \forall k \in \mathbb{Z}$ ;
- (iv) para quaisquer valores  $t_1, t_2, \dots, t_n \in T$ , temos que

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \rho(|t_i - t_j|) \geq 0,$$

em que  $\alpha_1, \alpha_2, \dots, \alpha_n$  são constantes reais.

Na prática, obtém-se apenas a realização amostral de um processo, o que implica a obtenção FAC amostral, da forma (GUJARATI; PORTER, 2011; SOUZA; CAMARGO, 1996):

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \quad (2.22)$$

onde  $\hat{\gamma}_k$  e  $\hat{\gamma}_0$  representam respectivamente a covariância amostral e variância amostral com  $k$  defasagens.

A função de autocorrelação parcial (FACP), mede a correlação de  $W_t$  e  $W_{t-k}$ , controlando os efeitos de observações intermediárias  $W_{t-1}, W_{t-2}, \dots, W_{t-k+1}$  e pode ser denominada por:

$$\phi_{kk} = \text{Cov}(W_t, W_{t-k} | W_{t-1}, W_{t-2}, \dots, W_{t-k+1}) \quad (2.23)$$

De modo geral, a FACP,  $\phi_{kk}$ , pode ser obtida através do sistema de equações de Yule-Walker, e a FACP amostral,  $\hat{\phi}_{kk}$ , pode ser obtida substituindo  $\phi_i$  por  $\hat{\phi}_i$  nas equações de Yule-Walker.

Matricialmente,

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \rho_k \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{k-3} \\ \vdots & \vdots & & & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & 1 \end{bmatrix} \begin{bmatrix} \hat{\phi}_{k1} \\ \hat{\phi}_{k2} \\ \hat{\phi}_{k3} \\ \vdots \\ \hat{\phi}_{kk} \end{bmatrix} \quad (2.24)$$

A autocorrelação parcial amostral  $\hat{\phi}_{kk}$  é dada pela solução das equações:

$$\begin{cases} \hat{\phi}_{k1} + \hat{\phi}_{k2}\hat{\rho}_1 + \hat{\phi}_{k3}\hat{\rho}_2 + \dots + \hat{\phi}_{kk}\hat{\rho}_{k-1} = \hat{\rho}_1 \\ \hat{\phi}_{k1}\hat{\rho}_1 + \hat{\phi}_{k2} + \hat{\phi}_{k3}\hat{\rho}_1 + \dots + \hat{\phi}_{kk}\hat{\rho}_{k-2} = \hat{\rho}_2 \\ \vdots \\ \hat{\phi}_{k1}\hat{\rho}_{i-1} + \hat{\phi}_{k2}\hat{\rho}_{i-2} + \hat{\phi}_{k3}\hat{\rho}_{i-3} + \dots + \hat{\phi}_{kk}\hat{\rho}_{i-k} = \hat{\rho}_i \\ \vdots \\ \hat{\phi}_{k1}\hat{\rho}_{k-1} + \hat{\phi}_{k2}\hat{\rho}_{k-2} + \hat{\phi}_{k3}\hat{\rho}_{k-3} + \dots + \hat{\phi}_{kk} = \hat{\rho}_k \end{cases}$$

### 2.2.2.5 Correlação Cruzada

A função de correlação cruzada mensura o grau de associação entre essas séries. A partir disso, são analisadas as possíveis relações e efeitos ao longo do tempo.

Supondo  $k$  séries temporais, denotadas  $w_{1,t}, w_{2,t}, \dots, w_{k,t}$ , então,  $\mathbf{W}_t = (w_{1,t}, w_{2,t}, \dots, w_{k,t})$ , é definido como vetor de séries temporais no tempo  $t$ .

O vetor  $\{\mathbf{W}_t\}$  é estacionário se as distribuições de probabilidade dos vetores aleatórios,  $(\mathbf{W}_{t_1}, \mathbf{W}_{t_2}, \dots, \mathbf{W}_{t_m})$  e  $(\mathbf{W}_{t_1+l}, \mathbf{W}_{t_2+l}, \dots, \mathbf{W}_{t_m+l})$ , são as mesmas para todo  $m$  e todo  $lag$   $l$  nos tempos  $t_1, t_2, \dots, t_m$ .

Sendo  $\{\mathbf{W}_t\}$  um processo estacionário, a covariância entre  $w_{i,t}$  e  $w_{j,t+l}$  deve depender apenas do lag  $l$ . As funções  $\gamma_{ij}(l)$  e  $\rho_{ij}(l)$  que representam a função de covariância cruzada e a função de correlação cruzada entre as séries  $w_{i,t}$  e  $w_{j,t+l}$  no lag  $l$ , respectivamente, são dadas por (BOX et al., 2015):

$$\gamma_{ij}(l) = cov[w_{i,t}, w_{j,t+l}] = \mathbb{E}[(w_{i,t} - \mu_i)(w_{j,t+l} - \mu_j)] \quad (2.25)$$

e

$$\rho_{ij}(l) = corr[w_{i,t}, w_{j,t+l}] = \frac{\gamma_{ij}(l)}{\{\gamma_{ii}(0)\gamma_{jj}(0)\}^{1/2}} \quad (2.26)$$

Deste modo, a função de correlação cruzada mede a intensidade de uma associação e sua direção.

### 2.2.2.6 Função de transferência

Os modelos Box e Jenkins permitem a incorporação de outras variáveis auxiliares à modelagem ARIMA. A forma como os movimentos dessas variáveis auxiliares influenciam na variável de interesse ao longo do tempo, é dada por uma função de transferência  $f(X_t)$  (BOX et al., 2015):

$$Y_t = f(X_t) + \epsilon_t \quad (2.27)$$

onde

$$f(X_t) = \frac{(w_0 + w_1L + w_2L^2 + \dots + w_sL^s)}{(1 - \delta_1L - \delta_2L^2 - \dots - \delta_rL^r)} X_{t-b}$$

A função  $f(X_t)$  pode aglutinar valores observados em qualquer momento do tempo  $t$  de uma ou mais séries temporais auxiliares (covariáveis).

Para identificar e estimar uma função de transferência seguindo a metodologia Box e Jenkins são tomados os seguintes passos (FERREIRA; SOUZA; MATTOS, 2015; BOX et al., 2015):

1. Ajustar um modelo ARIMA para a série independente  $X_t$ 
  - Para identificar os ordens do modelo ARIMA para  $X_t$ , utiliza-se a FAC e FACP, como visto nas Subseções 2.2.2.4.
2. Filtrar  $Y_t$  pelo modelo encontrado para  $X_t$ 
  - Para filtrar  $Y_t$  por meio do modelo de  $X_t$ , também utiliza-se um modelo ARIMA para a série temporal de interesse, acrescentando o modelo estimado anteriormente ajustado a  $Y_t$ .

## 3. Calcular a correlação cruzada entre os resíduos dos dois modelos

- A obtenção da função de correlação cruzada entre os resíduos dos modelos estimados gera um gráfico que mostra o relacionamento entre  $Y_t$  e os *lags* defasados de  $X_t$  a partir dos coeficientes de correlação.

4. Identificar  $r$ ,  $s$  e  $b$ 

- A partir do gráfico gerado pela correlação cruzada, é possível estimar visualmente valores para os parâmetros  $r$ ,  $s$  e  $b$ . Essa identificação pode ser feita da seguinte maneira:
  - $b$ : Representa o primeiro *lag* significativo
  - $r$ : Representa o comportamento da correlação cruzada, onde:
    - \*  $r = 0$ : Correlação constante para todos os *lags* após  $b$
    - \*  $r = 1$ : Correlação crescente após  $b$
    - \*  $r = 2$ : Correlação com comportamento ondular após  $b$
  - $s$ : Representa o número de *lags* crescentes após  $b$

## 5. Estimar o modelo Box e Jenkins com função de transferência

- Identificada a forma de  $f(X_t)$ , inicia-se a etapa de estimação do modelo com função de transferência. Inicialmente, é preciso identificar a ordem do modelo ARIMA para a série  $Y_t$  assim como também é feito para  $X_t$ .

## 6. Verificar a qualidade do modelo

- Para avaliar o modelo ajustado, é realizada uma análise dos resíduos para identificar se esses são independentes e identicamente distribuídos, através de testes como o teste idealizado por Ljung-Box descrito na Subsubseção 2.2.2.10. Se houver padrões de correlação, novos modelos devem ser propostos.

## 7. Verificar o desempenho do modelo

- Por fim, verificar a capacidade preditiva do modelo obtido com função de transferência e compará-lo, através de medidas de seleção como as descritas na Subsubseção 2.2.2.9, com outros modelos tentativos.

### 2.2.2.7 Identificação

A identificação é a etapa fundamental na metodologia Box e Jenkins, ela tem por objetivo determinar os parâmetros dos modelos a serem implementados. Esse processo é constituído por três partes (MORETTIN; TOLOI, 2006):

- O passo inicial consiste em averiguar se a série original precisa de transformação para estabilização da variância da série ao longo do tempo, ou para a transformação de tendências explosivas em lineares, ou ainda para a adequação da distribuição da série e dos resíduos ao modelo Normal, entre outros pressupostos. Nesta etapa, a utilização de ferramentas de análise exploratória e testes de adequação à distribuição Normal são de grande auxílio;
- A segunda etapa consiste em verificar a suposição de estacionariedade da série, através da aplicação sucessiva do teste ADF e, havendo não-estacionariedade, é realizada a primeira diferença da série com intuito de eliminar a tendência. Este processo deve ser repetido quantas vezes forem necessárias para obtenção de uma série estacionária. O número de diferenças necessárias identifica o parâmetro  $d$  do modelo;
- Por fim, é feita a identificação do modelo autoregressivo de média móvel (ARMA) e de seus parâmetro  $p$  e  $q$ , através da análise dos coeficientes das funções de autocorrelação e de autocorrelação parcial, estimados para vários *lags* e seus respectivos intervalos de confiança.

A FAC e FACP, são representadas graficamente pelo correlograma, que apresenta, para  $k$  *lags*, os valores dos coeficientes amostrais para essas funções. Através do correlograma é possível verificar na série características de aleatoriedade e indícios de tendência ou sazonalidade. Segundo Ehlers (2009), os limites aproximados do intervalo de confiança de 95% para ambas funções para esses coeficientes de autocorrelação (da FAC e FACP) são dados por  $\pm 1,96/\sqrt{n}$ .

A relação da FAC e FACP com o modelo autoregressivo (AR), de médias móveis (MA) e autoregressivo de média móvel (ARMA) é representada na Tabela 1 (PINTO, 2006).

O decaimento exponencial acelerado para zero na FAC e FACP indica que o número de diferenças necessárias foi obtido. Caso esse comportamento não seja observado, deve ser verificada a estacionariedade da série, assim como, a necessidade de diferencia-la mais

Tabela 1: Comportamento da FAC e FACP nos modelos AR( $p$ ), MA( $q$ ) e ARMA( $p, q$ )

Modelo	Autocorrelação (FAC)	Autocorrelação parcial (FACP)
AR( $p$ )	Decaimento exponencial	Corte brusco para zero após o <i>lag</i> $p$
MA( $q$ )	Corte brusco para zero após o <i>lag</i> $q$	Decaimento exponencial
ARMA( $p, q$ )	Decaimento exponencial após o <i>lag</i> $q$	Decaimento exponencial após o <i>lag</i> $p$

vezes. Com isso, novas autocorrelações e autocorrelações parciais são estimadas para a nova série, exigindo uma nova etapa de identificação a cada transformação realizada.

Em modelos sazonais, o comportamento da FAC e FACP para identificação dos parâmetros  $p$  e  $q$ , segue o descrito na Tabela 1, além da etapa de identificação dos parâmetros  $P$  e  $Q$  apresentado na Tabela 2

Tabela 2: Comportamento da FAC e FACP em modelos sazonais

Modelo	Autocorrelação (FAC)	Autocorrelação parcial (FACP)
AR( $p$ ) <sub><math>s</math></sub>	Decaimento nos <i>lags</i> $ks$	Zero após o <i>lag</i> $P_s$
MA( $q$ ) <sub><math>s</math></sub>	Zero após o <i>lag</i> $Q_s$	Decaimento nos <i>lags</i> $ks$
ARMA( $p, q$ ) <sub><math>s</math></sub>	Decaimento nos <i>lags</i> $ks$	Decaimento nos <i>lags</i> $ks$

No entanto, a identificação do melhor modelo e estimativa dos parâmetros  $p$  e  $q$  através da FAC e FACP pode ser, na prática, muitas vezes confusa. A partir disso, uma técnica que reduz a subjetividade envolvida na interpretação das funções de autocorrelação é, segundo Ehlers (2009), a utilização de medidas de qualidade de ajuste, incluindo critérios de informação, com objetivo não apenas de considerar a qualidade do ajuste, mas de penalizar a inclusão de parâmetros desnecessários.

### 2.2.2.8 Estimação

Nesta etapa da validação do modelo, são obtidas estimativas para os valores dos coeficientes  $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)$  e  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)$ . Segundo Morettin e Tolo (2006), os estimadores  $\phi$  e  $\theta$  podem ser encontrados através de alguns métodos, como o método dos momentos, máxima verossimilhança ou estimação não linear.

O método dos momentos (MM), por exemplo, consiste em substituir, nas equações que relacionam as autocorrelações (ou autocovariâncias) e os parâmetros do modelo, os momentos teóricos (média, variância e autocorrelações) pelos correspondentes momentos amostrais. No modelo ARMA( $p, q$ ) os estimadores são obtidos em duas etapas:

(i) Obter o estimador de  $\phi$

$$r_j = \hat{\phi}_1 MM r_{j-1} + \cdots + \hat{\phi}_p MM r_{j-p}, \quad j = q+1, \dots, q+p \quad (2.28)$$

onde

$$\hat{\phi}_{MM} = R_p^{-1} r_p$$

$$r_p = (r_1, r_2, \dots, r_p)'$$

$R_p$  é a matriz formada pelos elementos  $(r_{i-j})_{i,j}^p$

observa-se que  $r_i$  é a estimativa da FAC  $\rho_i$ ,  $\forall i = 1, 2, \dots, p$

(ii) Obter o estimador de  $\theta$

$$\gamma_j = \phi_1 \gamma_{j-1} + \cdots + \phi_p \gamma_{j-p} + \gamma_{wa}(j) - \theta_1 \gamma_{wa}(j-1) - \cdots - \theta_q \gamma_{wa}(j-q) \quad (2.29)$$

onde  $\gamma_{wa}(j)$  é a covariância cruzada entre  $W_t$  e  $a_t$ , definida por:

$$\gamma_{wa}(j) = \mathbb{E}(a_t W_{t-j})$$

### 2.2.2.9 Critérios de comparação de modelos

Para a seleção do modelo que fornece o melhor ajuste a um determinado conjunto de dados, busca-se aquele que mais se aproxima do observado com a menor perda de informação. Para isso, são calculadas medidas de aderência e escolhido o modelo mais parcimonioso, adotadas como medidas o AIC (especialmente adotado para avaliar a qualidade do ajuste) e o MAPE (para poder mensurar de forma percentual o erro de previsão cometido), onde:

- **Critério de Informação de Akaike (AIC)**

O critério de informação de Akaike (AIC) minimiza a seguinte medida:

$$AIC = -2\log(L) + 2k \quad (2.30)$$

em que  $L$  é a função de máxima verossimilhança e  $k$  é o número de parâmetros, sendo  $k = p + q + 1$  se o modelo contiver intercepto e  $k = p + q$  em caso contrário (AKAIKE, 1974).

- **Erro Absoluto Percentual Médio (MAPE)**

O MAPE apresenta, em forma de porcentagem, o quanto, em média, os valores ajustados pelo modelo diferenciam-se dos valores efetivos da série. O valor desta estatística varia de 0 a 100%, e quanto menor, melhor o resultado obtido pelo modelo estimado.

$$MAPE = \frac{100}{n} \sum_{i=1}^n \frac{|W_t - \hat{W}_t|}{|W_t|} \quad (2.31)$$

### 2.2.2.10 Análise de resíduos

De acordo com Ehlers (2009), após um modelo ter sido ajustado a uma série temporal, deve-se verificar se ele satisfaz às hipóteses básicas do modelo proposto, a partir do comportamento de seus resíduos, calculados como a diferença entre os valores observados e os respectivos valores ajustados pelo modelo. Espera-se que eles formem um processo aleatório correspondente a um ruído branco, tendo média igual a zero e variância constante, além de serem observações independentes e identicamente distribuídas.

Esta etapa consiste em verificar a independência e normalidade residual, sendo realizada através de testes de hipóteses.

Uma forma de verificar a independência dos resíduos é através do teste idealizado por Ljung e Box (1978), que consiste em testar se os coeficientes da FAC dos resíduos podem ser considerados significativamente iguais a zero, de forma que não seja identificada qualquer estrutura de dependência entre eles.

Desta forma, o teste acaba verificando as seguintes hipóteses:

$$\begin{cases} \mathbf{H}_0 : \text{Os resíduos são independentes} \\ \mathbf{H}_1 : \text{Os resíduos não são independentes} \end{cases}$$

A estatística do teste é dada por:

$$Q = n(n+2) \sum_{k=1}^m \frac{\hat{r}_k^2}{(n-k)} \sim \chi_{m-p-q}^2 \quad (2.32)$$

sendo  $n$  o tamanho da série temporal,  $m$  o número de *lags* que está sendo testado na função de autocorrelação,  $p$  e  $q$  as ordens do modelo ajustado, e  $\hat{r}_k^2$  a estimativa dos valores da FAC (MORETTIN; TOLOI, 2006).

Se o modelo for considerado eficaz, a estatística  $Q$  se aproximará de uma qui-quadrado com  $m - p - q$  graus de liberdade.

Para análise da adequação à distribuição Normal como modelo para a distribuição dos resíduos, o teste idealizado por Shapiro e Wilk (1965) e, utilizado neste trabalho, verifica as seguintes hipóteses:

$$\begin{cases} \mathbf{H}_0 : \text{Os resíduos seguem distribuição normal} \\ \mathbf{H}_1 : \text{Os resíduos não seguem distribuição normal} \end{cases}$$

A estatística do teste é definida por:

$$S = \frac{b^2}{s^2} = \frac{\left(\sum_{i=1}^n a_i x_{(i)}\right)^2}{\sum_{i=1}^n (x_i - \bar{x}_i)^2} \quad (2.33)$$

onde  $x_i$  são as realizações da variável aleatória,  $x_{(i)}$  é a  $i$ -ésima estatística de ordem, ou seja, o  $i$ -ésimo menor número da amostra e  $a_i$  são coeficientes tabelados (SNEC, 1974).

Primeiro, ordena-se as observações em ordem decrescente  $x_1 \leq x_2 \leq \dots \leq x_n$ , em seguida, calcula-se  $b^2$  e  $s^2$ .

Se  $n$  é par,

$$b = \sum_{i=1}^{\frac{n}{2}} a_{n-i+1} (x_{(n-i+1)} - x_{(i)}), \quad (2.34)$$

Se  $n$  é ímpar,

$$b = \sum_{i=1}^{\frac{n+1}{2}} a_{n-i+1} (x_{(n-i+1)} - x_{(i)}) \quad (2.35)$$

Observa-se que  $a_{n-i+1}$  são coeficientes gerados pelas variâncias, covariâncias e médias das estatísticas de ordem de uma amostra de tamanho  $n$  de uma distribuição Normal (EHLERS, 2009).

### 2.2.2.11 Previsão *out-of-sample*

Com a verificação das etapas anteriores, o modelo escolhido é utilizado para fazer previsões da série temporal estudada. O objetivo principal da previsão é antecipar eventos futuros e, a partir disso, realizar o planejamento e tomar melhores decisões.

Uma das formas de se realizar previsão é através da metodologia *out-of-sample*, que possui maior aplicação prática por aproximar melhor a restrição de informação que observamos na vida real. O processo consiste em utilizar informações de uma parte da amostra para estimar o modelo e, então, obter estimativas para o comportamento futuro da série a partir do modelo estimado.

Considerando o modelo estudado,  $ARIMA(p, d, q)$ , tem-se por objetivo prever um valor  $W_{t+h}$ ,  $h \geq 1$ , dado que são conhecidas as observações até o instante  $t$ . Dessa maneira, a previsão da variável  $W_{t+h}$ , de origem  $t$  e horizonte  $h$ , é obtida através da sua esperança condicional (MORETTIN; TOLOI, 2006):

$$\hat{W}_{t+h} = \mathbb{E}(W_{t+h}|w_t, w_{t-1}, \dots) \quad (2.36)$$

Podemos expressar a previsão  $\hat{W}_{t+h}$ , dentre outras formas, utilizando a equação de diferenças.

Logo, tomando a esperança condicional, decrita na Equação 2.36, obtém-se

$$\hat{W}_{t+h} = \phi_1[W_{t+h-1}] + \dots + \phi_{p+d}[W_{t+h-p-d}] - \theta_1[a_{t+h-1}] - \dots - \theta_q[a_{t+h-q}] + [a_{t+h}] \quad (2.37)$$

Realizada as previsões, e verificada sua qualidade, determina-se um intervalo de confiança pra  $W_{t+h}$ . Dado os valores passados e presentes da série  $W_t, W_{t-1}, \dots$  a distribuição condicional de  $W_{t+h}$  será  $N(\hat{W}_{t+h}, V(h))$ .

Dessa forma,

$$U = \frac{W_{t+h} - \hat{W}_{t+h}}{|V(h)|^{\frac{1}{2}}} \sim N(0, 1) \quad (2.38)$$

O valor  $\sigma^2$  em  $V(h)$ , é desconhecido e, portanto, substituído por sua estimativa,  $\hat{\sigma}^2$ , encontrada na fase de estimação do modelo.

Por fim, fixado o coeficiente de confiança  $\lambda$  e encontrando  $u_\lambda$ , tal que  $P(-u_\lambda < U < u_\lambda) = \lambda$  pode-se expressar o intervalo de confiança de  $W_{t+h}$  por:

$$\hat{W}_{t+h} - u_\lambda \sigma_\alpha^2 \left[ 1 + \sum_{j=1}^{h-1} \psi_j^2 \right]^{\frac{1}{2}} \leq W_{t+h} \leq \hat{W}_{t+h} + u_\lambda \sigma_\alpha^2 \left[ 1 + \sum_{j=1}^{h-1} \psi_j^2 \right]^{\frac{1}{2}} \quad (2.39)$$

Observe que a amplitude deste intervalo aumentará à medida que se afasta da origem  $t$ , tornando as previsões incertas e imprecisas para horizontes  $h$  muito grandes.

## 3 Resultados

Neste capítulo, são apresentados os resultados da análise exploratória e da aplicação da modelagem de Box e Jenkins para descrever e prever valores futuros da série temporal semanal de casos de dengue em Florianópolis. Estas análises foram realizadas utilizando a linguagem R (TEAM, 2021).

### 3.1 Análise exploratória

De janeiro de 2017 a dezembro de 2021, houve, ao total, 3.389 casos confirmados de dengue na cidade de Florianópolis distribuídos ao longo das 261 semanas epidemiológicas. A Figura 2 ilustra o comportamento da série, onde é possível observar picos não padronizados de casos de dengue nos primeiros meses de alguns anos, o que pode sugerir que em períodos de verão há uma tendência a uma maior atividade do mosquito vetor da doença e um conseqüente aumento do número de notificações. De 2017 a 2018, o número de casos da série apresentou um comportamento estável e pouca variabilidade. No ano de 2019, o número de casos cresceu abruptamente, contendo o segundo maior registro de casos semanal da doença de toda a série estudada. Em 2020, o patamar da série estava acima do que foi observado ao longo de 2017 e 2018, e em meados de 2021 atingiu sua marca histórica, ano recorde em número de casos de dengue na cidade.

A Tabela 3 apresenta algumas estatísticas descritivas dos casos de dengue em Florianópolis. Pode-se observar que ocorrem na cidade em média 12,9 novos casos de dengue por semana. Na semana 93 do estudo, em outubro de 2018, não foi registrado nenhum caso de dengue, e na semana 225, em abril de 2021, foram registrados 54 casos, sendo este o maior valor assumido da série.

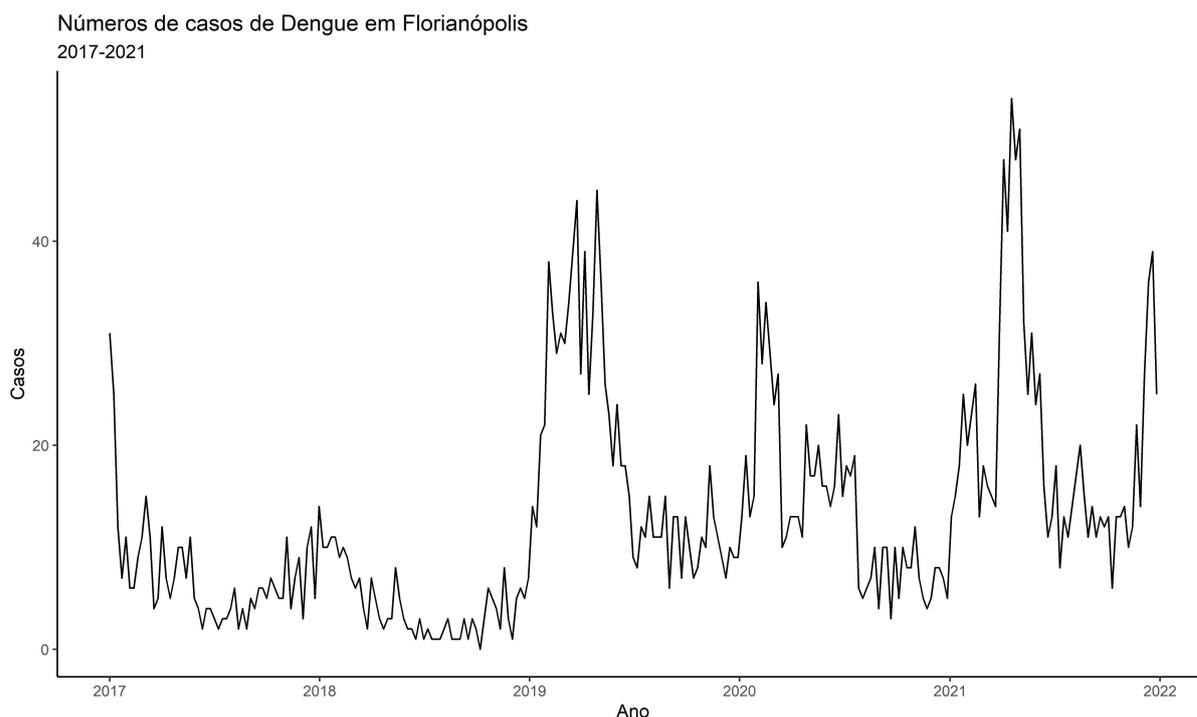


Figura 2: Série do número de casos semanais de dengue em Florianópolis

Tabela 3: Estatísticas descritivas do número de casos semanais de dengue em Florianópolis entre 2017 e 2021

Estatística	Valor
Mínimo	0,0
Máximo	54,0
Média	12,9
Mediana	11,0
Desvio padrão	10,5
Coefficiente de Variação (%)	81,1
Percentil de 5%	2,0
Percentil de 95%	36,0

Foi verificada, ao nível de significância 5%, a presença de tendência na série de casos de dengue ( $p < 0,01$ ) através do teste Mann-Kendall, apresentado na Subsubseção 2.2.1.1, indicando que deve-se rejeitar a hipótese nula do teste e considerar que há tendência na série. Além disso, foi verificada a presença de sazonalidade através do teste Kruskal-Wallis, apresentado na Subseção 2.2.1.2, mas, como  $p = 0,48$ , aceita-se a hipótese nula que estipula que essa série não possui sazonalidade.

Para averiguar se a série possui estacionariedade, utilizou-se o teste ADF, ao nível de significância 5%, descrito na Equação 2.6. Para a série dos casos de dengue em Florianópolis, rejeita-se a hipótese nula de que a série possui raiz unitária ( $p = 0,04$ ).

Portanto, a série deveria ser considerada estacionária observando-se apenas o resultado do teste.

Após a primeira diferenciação da série, com objetivo de torna-la efetivamente estacionária, foi realizado novamente o teste ADF ( $p < 0,01$ ), indicando que a série após uma diferença ainda é estacionária.

O acumulado do índice pluviométrico para a cidade de Florianópolis foi de 8.445,1 mm ao final das 261 semanas epidemiológicas compreendidas entre janeiro de 2017 e dezembro de 2021. A Figura 3 ilustra o comportamento desta série temporal, onde é possível observar uma variabilidade grande no registro de chuvas. Na Figura 4, observa-se uma concentração dos maiores valores de precipitação nos primeiros meses de cada ano (em geral, até a 12<sup>a</sup> semana epidemiológica), o que pode sugerir uma influência das estações do ano no registro do índice, em especial o do verão, com maiores volumes de janeiro a março de cada ano.

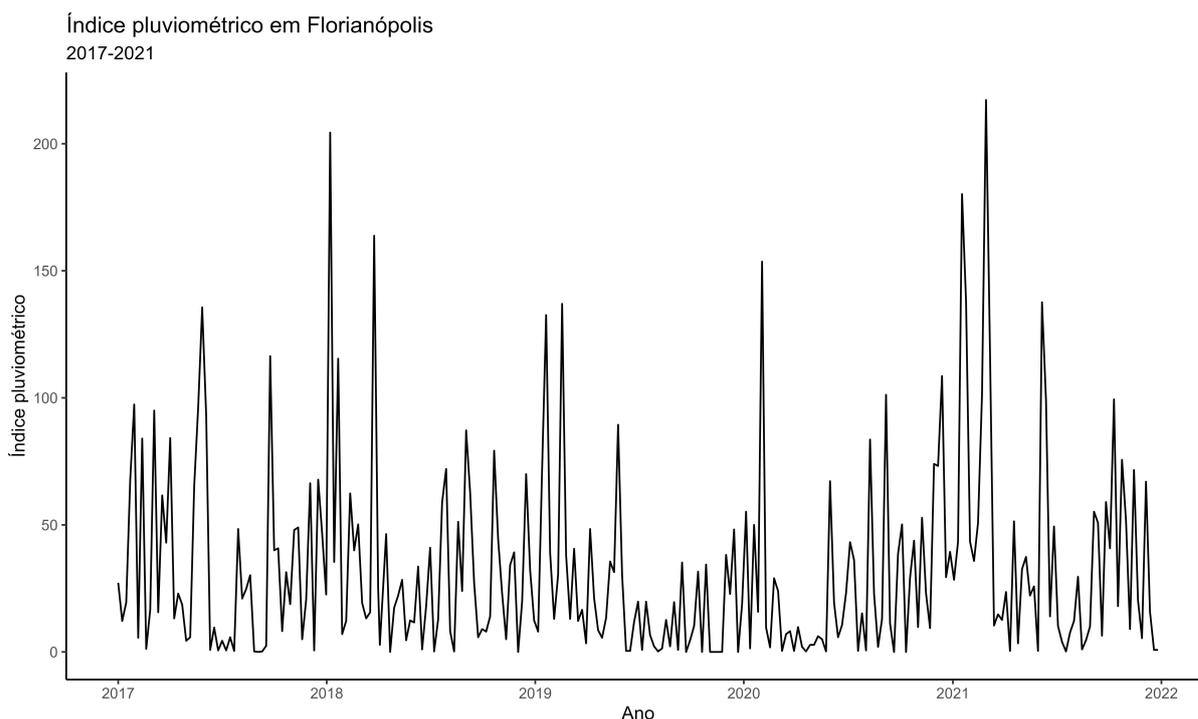


Figura 3: Série semanal do índice pluviométrico em Florianópolis

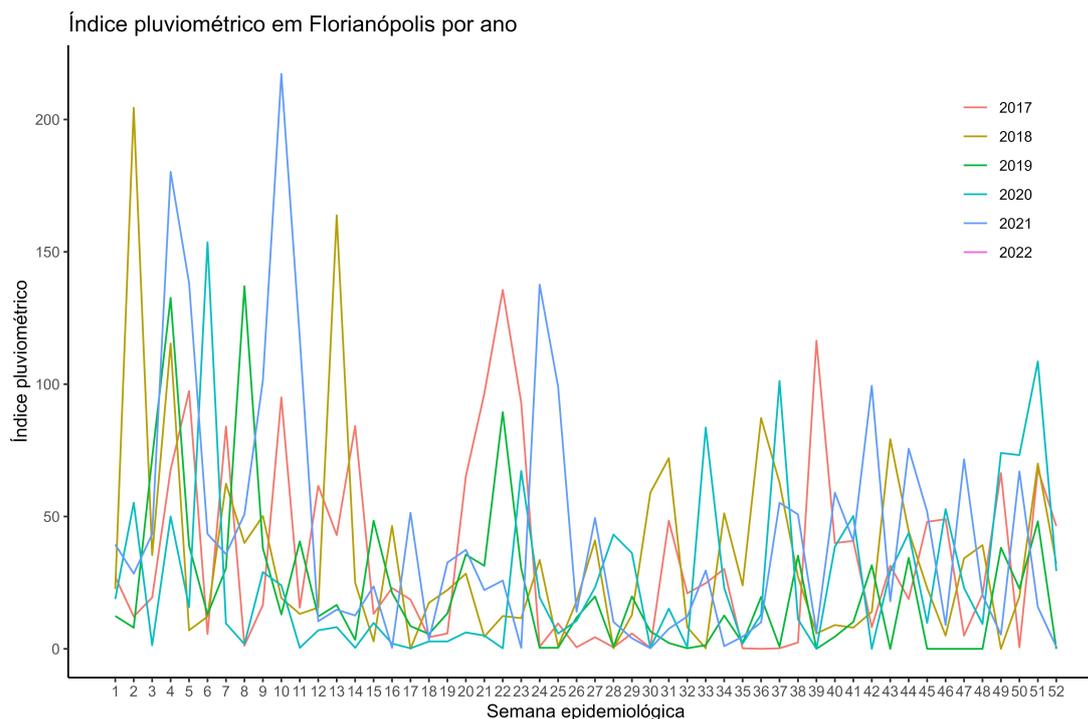


Figura 4: Série semanal do índice pluviométrico em Florianópolis por ano

O ano de 2019, apresentou o menor índice acumulado da série (1.176,2 mm), enquanto o ano de 2021, o maior (2.218,4 mm).

Na Tabela 4 são apresentadas estatísticas descritivas relevantes do índice pluviométrico para a cidade de Florianópolis. É possível observar que chove em média 32,7 mm por semana. Em 12 semanas da série, não foi registrada chuva na cidade, e na semana 218 do estudo, em fevereiro de 2021, foi registrado 217,2 mm de chuva, sendo este o maior valor observado na série.

Tabela 4: Estatística Descritiva do índice pluviométrico semanal em Florianópolis de 2017 a 2021

Estatística	Valor
Mínimo	0,0
Máximo	217,2
Média	32,7
Mediana	19,6
Desvio padrão	37,8
Coefficiente de Variação (%)	116,9
Percentil de 5%	0,2
Percentil de 95%	108,6

A presença de tendência na série de índice pluviométrico foi testada ( $p = 0,09$ ), ao nível de significância 5%, com isso, aceitou-se a hipótese nula de que a série não possui tendência. Em seguida, foi testada a presença de sazonalidade, mas como  $p = 0,38$ , aceitou-se a hipótese nula que estipula que a série não possui sazonalidade.

A estacionariedade da série foi testada ( $p < 0,01$ ), ao nível de significância 5%, indicando que fosse rejeitada a hipótese nula de que a série possui raiz unitária. Logo, a série é estacionária.

A temperatura mínima da cidade de Florianópolis foi observada durante as 261 semanas epidemiológicas no período de janeiro de 2017 a dezembro de 2021. A Figura 5 ilustra o comportamento da série, onde é possível observar uma grande variação térmica no registro das temperaturas mínimas e aparentemente a presença de sazonalidade. Na Figura 6, observa-se um comportamento semelhante das temperaturas mínimas ao longo dos anos estudados, deixando evidente a existência de uma componente sazonal estocástica da série, que possivelmente sofre influência direta das estações do ano. Destaca-se o período das semanas epidemiológicas 20 a 35 de cada ano, correspondentes, aproximadamente, aos meses de maio a junho, onde foram observadas as menores temperaturas da série.

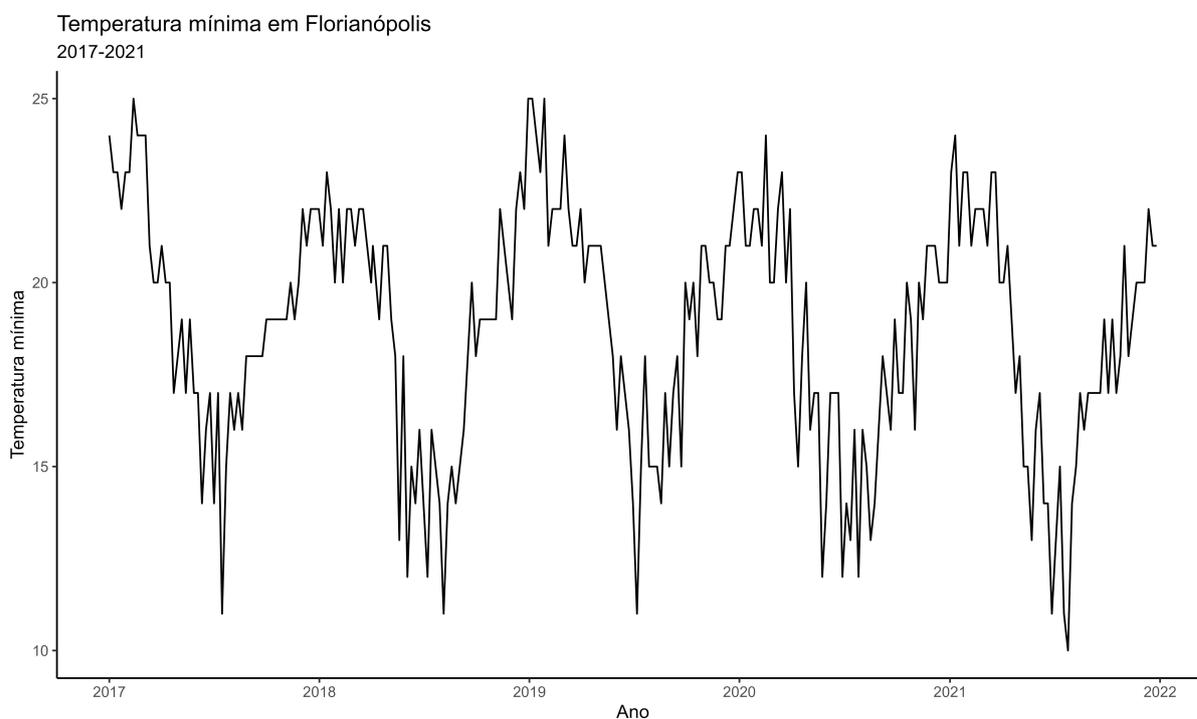


Figura 5: Série semanal de temperatura mínima (C°) em Florianópolis

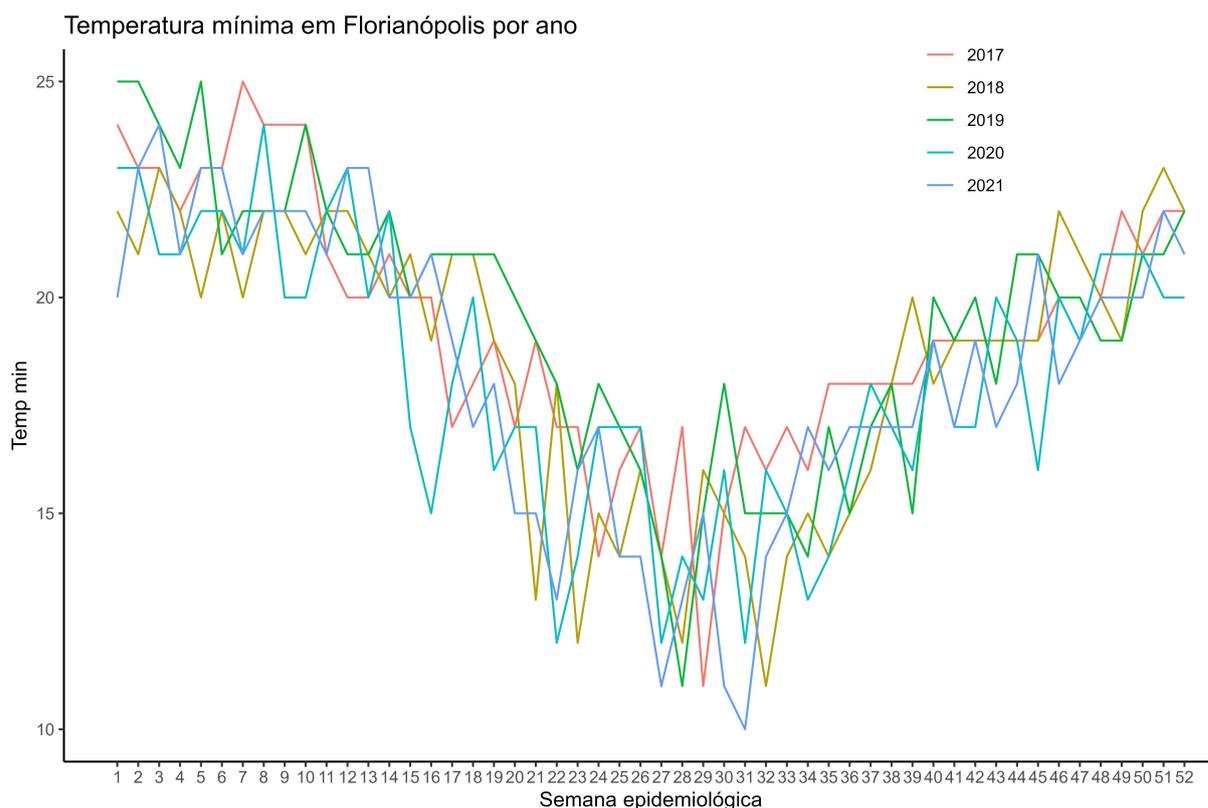


Figura 6: Série semanal de temperatura mínima (C°) em Florianópolis por ano

As estatísticas descritivas da temperatura mínima em Florianópolis são apresentadas na Tabela 5. Observa-se que a temperatura mínima semanal é, em média, igual a 18,7°C. Na semana 239 da série, foi registrada a menor temperatura mínima na cidade, 10°C, enquanto a maior temperatura foi registrada em quatro semanas da série, 25°C.

Tabela 5: Estatística Descritiva da temperatura mínima (C°) em Florianópolis de 2017 a 2021

Estatística	Valor
Mínimo	10,0
Máximo	25,0
Média	18,7
Mediana	19,0
Desvio padrão	3,24
Coefficiente de Variação (%)	17,3
Percentil de 5%	13,0
Percentil de 95%	23,0

A presença de tendência na série foi testada ( $p < 0,01$ ), ao nível de significância 5%, indicando de que a série possui tendência. Além disso, foi testada a presença da componente sazonal ( $p < 0,01$ ), sugerindo que a série possui sazonalidade, como observado graficamente.

Para a série de temperatura mínima em Florianópolis, foi testada a presença de raiz unitária ( $p = 0,09$ ), ao nível de significância 5%, indicando que não se deve rejeitar a hipótese nula de que a série possui raiz unitária. Portanto, conclui-se que a série é não estacionária.

Após a primeira diferenciação da parte sazonal da série, com objetivo de torná-la estacionária, foi testada novamente a estacionariedade da série ( $p < 0,01$ ), mostrando que a série com uma diferença é estacionária.

## 3.2 Modelo para as séries temporais

Nesta Seção são apresentados os processos para estimação dos parâmetros dos modelos, assim como os melhores modelos obtidos para as séries temporais estudadas. Além disso, é apresentada a correlação cruzada e função de transferência utilizando a série de casos de dengue e as séries climáticas e a previsão do número de casos da doença para as seis primeiras semanas epidemiológicas de 2022.

### 3.2.1 Série de Casos de Dengue

Analisando as funções de autocorrelação e autocorrelação parcial, descritas na Subsubseção 2.2.2.4, pode-se observar através da Figura 7 a presença de uma tendência decrescente no gráfico da FAC. Seus *lags* estão acima do intervalo de confiança representado pela linha tracejada em azul, indicando que a autocorrelação é significativamente diferente de zero. Desta forma, se faz necessário diferenciar a série para torná-la de fato estacionária.

A avaliação de estacionariedade foi feita de forma conjunta, olhando o resultado do teste ADF e o correlograma da FAC.

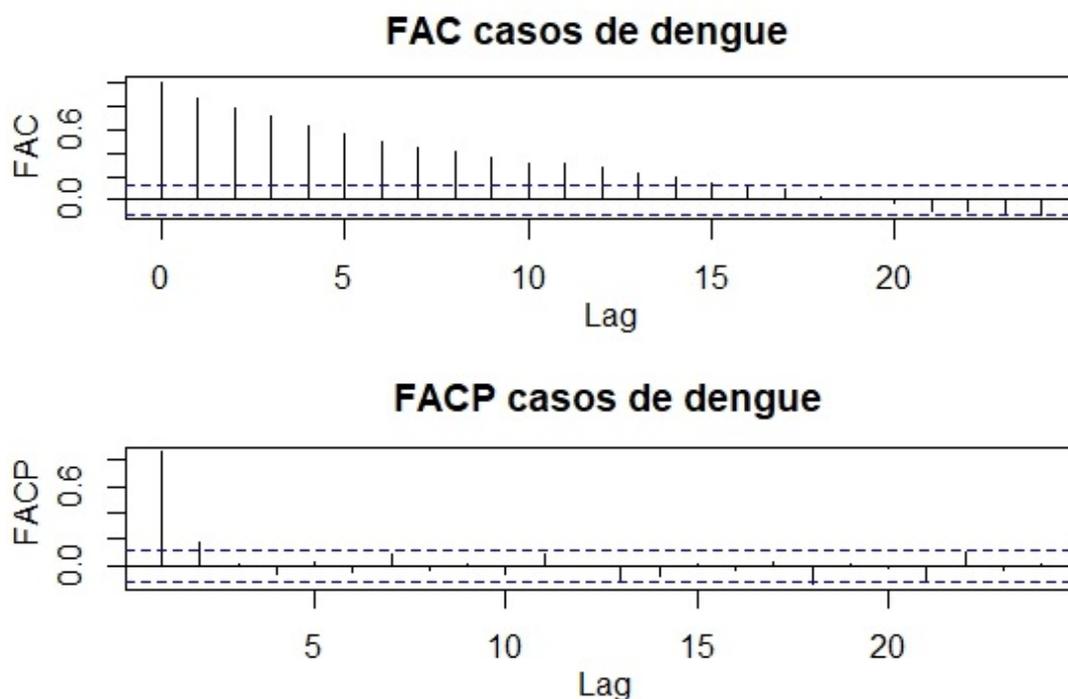


Figura 7: Correlogramas da FAC e da FACP do número de casos de dengue em Florianópolis

Na Figura 8, com a análise dos correlogramas da série diferenciada, é possível observar no gráfico da FAC que o primeiro *lag* foi significativo, ou seja, para a parte de médias móveis,  $q = 1$ . No gráfico da FACP, o primeiro *lag* também mostrou-se significativo sugerindo usar a parte autorregressiva  $p = 1$ . Como foi aplicada uma diferenciação para que a série se tornasse estacionária, o modelo pode ser descrito como um possível ARIMA(1, 1, 1). Observe que embora, pontualmente, alguns outros *lags* tenham sido considerados significativos, eles não seriam efetivamente levados em consideração na identificação do modelo (escolha de  $p$  e  $q$ ).

Para fins de comparação com a identificação realizada pelos correlogramas, foi também feita a identificação adotando como critério o AIC, apresentado na Subsubseção 2.2.2.9, e os resultados foram apresentados na Tabela 6.

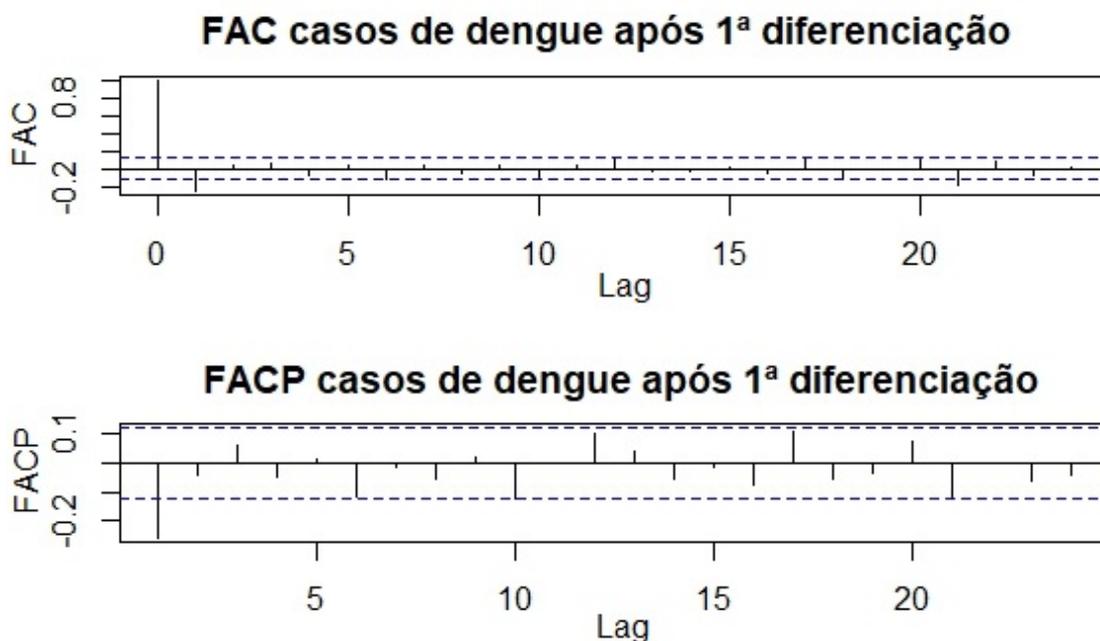


Figura 8: Correlogramas da FAC e FACP do número de casos de dengue em Florianópolis após uma diferenciação

Tabela 6: Comparação dos modelos ARIMA para série temporal dos casos de dengue em Florianópolis, através do AIC

Modelo	Identificação	AIC
1	ARIMA(2,1,0)	1611,402
2	ARIMA(1,1,1)	1611,559
<b>3</b>	<b>ARIMA(2,1,1)</b>	<b>1609,387</b>
4	ARIMA(3,1,1)	1611,296
5	ARIMA(2,1,2)	1614,642
6	ARIMA(3,1,2)	1613,400

Neste caso, incluindo a comparação com o modelo ARIMA(1,1,1) proposto anteriormente, através do AIC, identificou-se que o modelo com menor valor para este critério foi o ARIMA(2,1,1), e a comparação entre os valores observados da série e os valores previstos por esse modelo, foi apresentada no Apêndice 1. Pode-se observar que os valores obtidos pelo modelo ARIMA se ajustam bem ao comportamento da série ao longo das semanas epidemiológicas.

### 3.2.2 Série de Índice Pluviométrico

Analisando as funções de autocorrelação e autocorrelação parcial, na Figura 9, com a análise dos correlogramas, é possível observar no gráfico da FAC que o primeiro *lag* é significativo, ou seja, para a parte de médias móveis,  $q = 1$ . No gráfico da FACP, o primeiro *lag* também mostrou-se significativo, sugerindo usar a parte autorregressiva  $p = 1$ . Como não foi necessário diferenciar a série para que se tornasse estacionária, o modelo pode ser descrito como um possível ARIMA(1,0,1).

A avaliação de estacionariedade foi feita de forma conjunta, olhando o resultado do teste ADF e o correlograma da FAC.

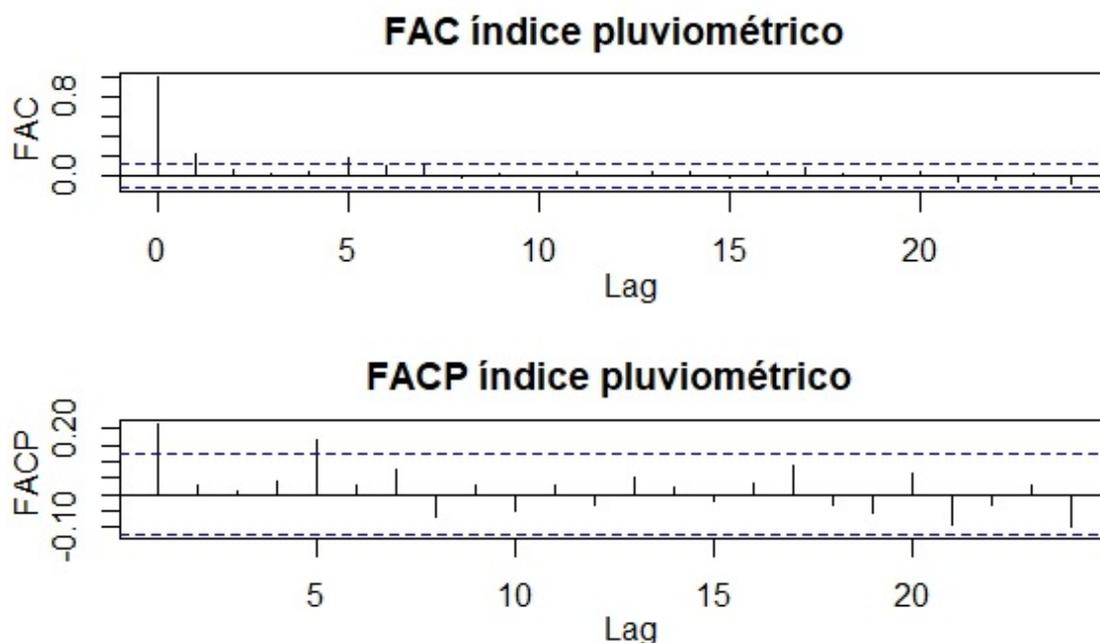


Figura 9: Correlogramas da FAC e FACP do índice pluviométrico em Florianópolis

Para fins de comparação com a identificação realizada pelos correlogramas, foi também feita a identificação adotando como critério o AIC, e os resultados descritos no Apêndice 2. O modelo com menor AIC foi o mesmo identificado nos correlogramas, ARIMA(1,0,1), e a comparação dos valores previstos por esse modelo com os valores efetivos da série, foi descrita no Apêndice 1. Pode-se observar que pela aleatoriedade e natureza da série trabalhada, a alta variabilidade dos dados possivelmente faz com que o modelo ARMA não apresente uma boa qualidade de ajuste, proporcionando erros maiores e valores bem diferentes daqueles observados ao longo das semanas epidemiológicas.

### 3.2.3 Série de Temperatura Mínima

Analisando as funções de autocorrelação e autocorrelação parcial, pode-se observar através da Figura 10 no gráfico da FAC, que há uma interferência sazonal na série, pois há autocorrelações fora do intervalo de confiança, ou seja, estatisticamente significativas, aparecendo de forma sistemática e na mesma periodicidade da série indicando a necessidade de diferenciação da parte série sazonal, concordando assim com o resultado encontrado no teste ADF.

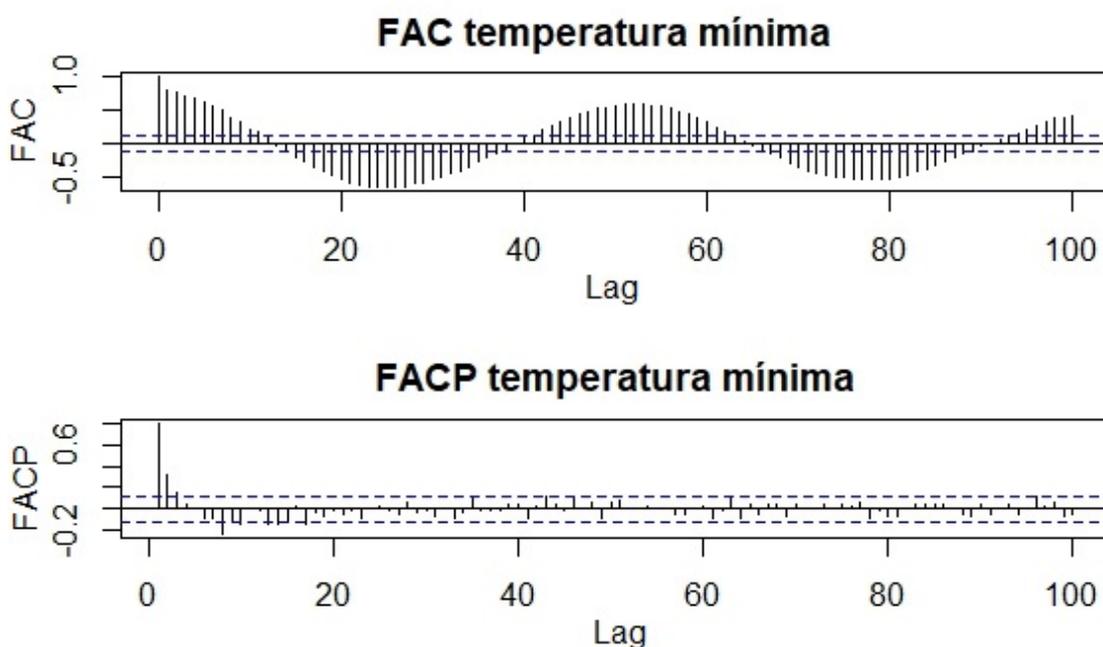


Figura 10: Correlogramas da FAC e da FACP da temperatura mínima (C°) em Florianópolis

Na Figura 11, com a análise dos correlogramas da série diferenciada, é possível observar no gráfico da FAC que os dois primeiros *lags* foram significativos, ou seja, para a parte de médias móveis, deveria-se escolher,  $q = 1$  ou  $2$ . No gráfico da FACP, o primeiro *lag* mostrou-se significativo sugerindo usar a parte autorregressiva  $p = 1$ . Existe um componente sazonal significativo no *lag* 52 da FAC e FACP, o que indica um parâmetro  $P = 1$  e  $Q = 1$ . Como foi aplicada uma diferenciação na parte sazonal da série, para eliminar este componente e torná-la estacionária, o modelo pode ser descrito como um possível SARIMA(1,0,1)(1,1,1)<sub>52</sub>.

Observe que aparecem *lags* pouco significativos da FAC e FACP em *lags* maiores, não trazendo nenhuma grande e significativa contribuição para o modelo, sendo desprezados para fins de parcimônia.

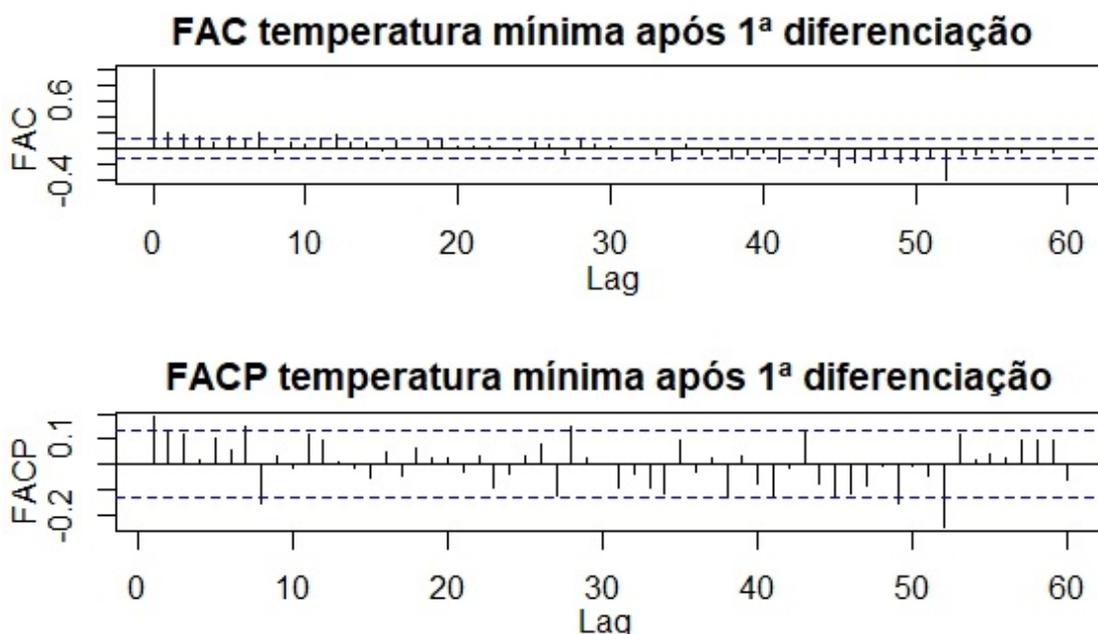


Figura 11: Correlogramas da FAC e da FACP da temperatura mínima ( $^{\circ}$ ) em Florianópolis após uma diferenciação

Para fins de comparação com a identificação realizada pelos correlogramas, foi também feita a identificação adotando como critério o AIC, e os resultados descritos no Apêndice 2. Sendo assim, o melhor modelo com menor AIC, foi o SARIMA(1,0,1)(1,1,0)<sub>52</sub>, e a comparação entre os valores observados da série, e a série ajustada pelo modelo, foi apresentada no Apêndice 1. Pode-se observar que os valores obtidos pelo modelo SARIMA se ajustam bem ao comportamento dos dados originais. Nas primeiras 50 semanas epidemiológicas, os valores ajustados tiveram erros mínimos comparados aos observados. Posteriormente, as diferenças entre o que foi observado e o que foi previsto aumentam em magnitude, mas não de forma sistemática, embora, em geral, a qualidade do ajuste seja aparentemente boa.

Após a escolha dos melhores modelos em cada série, foi realizada a análise dos resíduos, e os resultados apresentados no Apêndice 3. Foi utilizado o teste de Ljung-Box, descrito na Subsubseção 2.2.2.10, e observa-se que todos os modelos possuem resíduos independentes. Verificou-se, ainda, através do teste de Shapiro-Wilk, descrito na Subsubseção 2.2.2.10, se os resíduos seguem uma distribuição Normal. Porém, em todas as séries ( $p < 0,01$ ) a hipótese de normalidade dos resíduos foi rejeitada.

Considerando a natureza dos dados das séries estudadas e a literatura revisada, justifica-se a não normalidade observando-se que há restrições sobre os valores dos da-

dos. Segundo Pino (2014), uma das restrições mais comuns aos valores que as observações podem assumir é que elas sejam estritamente positivas (ou pelo menos, não negativas), além de casos ainda mais restritivos como, por exemplo, dados de contagem, que devem ser estritamente inteiros e não negativos. Dessa forma, os modelos obtidos foram considerados e utilizados para realização de previsões.

### 3.2.4 Correlação cruzada e Função de Transferência

A fim de comparar o modelo ARIMA, que utiliza apenas valores passados da própria série de casos de dengue com outros modelos concorrentes, utilizou-se a função de transferência para estimar o número de casos de dengue para as seis primeiras semanas epidemiológicas de 2022, a partir das séries temporais auxiliares referentes ao índice pluviométrico e à temperatura mínima em Florianópolis.

Através da função de correlação cruzada, apresentada na Figura 12, observa-se que não há correlação significativa entre a série de casos de dengue e as séries de índice pluviométrico e temperatura mínima. Dessa forma, preferiu-se então tentar ajustar um modelo com função de transferência que minimizasse o critério de informação de Akaike (AIC) a partir de diversas combinações entre as duas covariáveis defasadas do *lag* 1 até o *lag* 7, e que todas fossem significativas no modelo, onde o resultado das melhores combinações foi descrito no Apêndice 4.

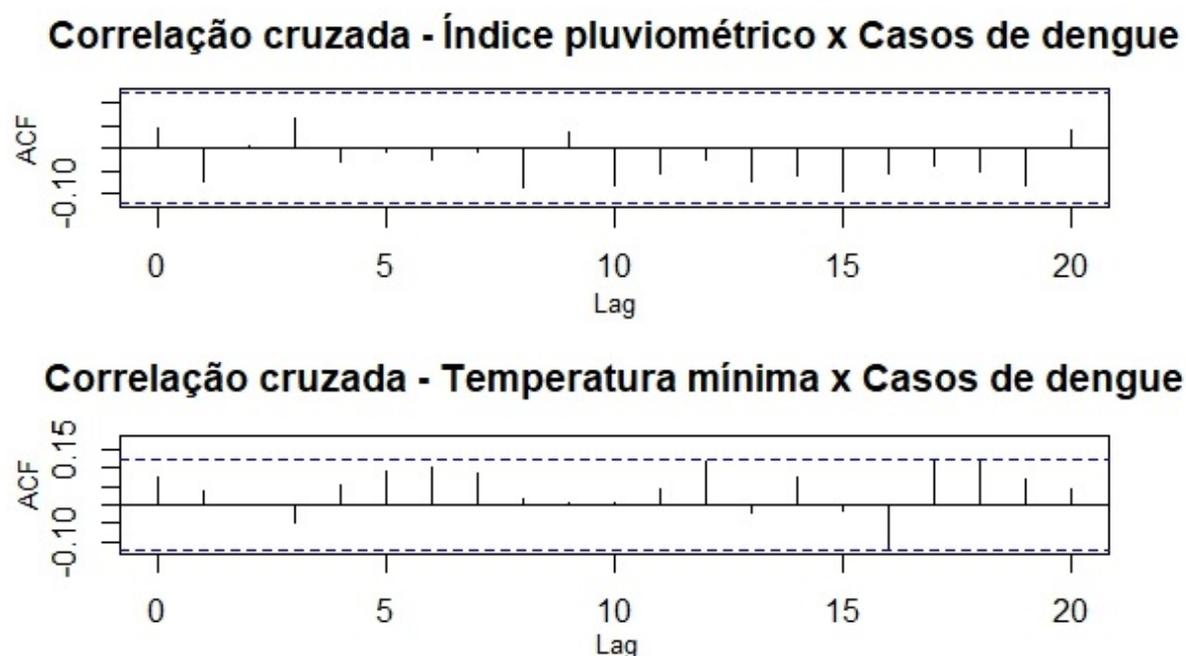


Figura 12: Função de correlação cruzada para cada série

### 3.2.5 Comparação de modelos para a série de casos de dengue

A série dos números de casos de dengue foi ajustada considerando 4 modelos ARIMA diferentes. A denominação dos modelos considerados é a seguinte:

- ARIMA = Sem usar função de transferência;
- FT1 = Modelo ARIMA com função de transferência utilizando como covariáveis as séries índice pluviométrico( $\ell = 7$ ) e temperatura mínima( $\ell = 5$ ).
- FT2 = Modelo ARIMA com função de transferência utilizando como covariável a série índice pluviométrico( $\ell = 7$ );
- FT3 = Modelo ARIMA com função de transferência utilizando como covariável a série temperatura mínima( $\ell = 5$ ).

A seguir na Tabela 7, podemos verificar as medidas AIC e MAPE para comparar a adequabilidade dos modelos e das previsões, respectivamente.

Tabela 7: Comparação dos modelos

Modelo	AIC	MAPE
ARIMA	1609,23	<b>25,67</b>
FT1	<b>1554,70</b>	33,07
FT2	1558,57	30,79
FT3	1569,69	31,53

O primeiro modelo foi obtido considerando minimizar o critério de informação de Akaike (AIC) e a significância dos parâmetros. O segundo e o terceiro modelo, foram obtidos apenas separando as covariáveis do primeiro modelo.

Ao utilizar a função de transferência, não foi observada uma melhora na qualidade de previsão. Na comparação dos modelos pelos AICs, observa-se uma pequena diferença entre os resultados e, portanto, utiliza-se o MAPE para a escolha do melhor modelo. Desta forma, o modelo ARIMA inicial será utilizado como base para as previsões.

Logo, as previsões de valores futuros do número de casos de dengue na cidade de Florianópolis e o intervalo de confiança de 95%, para as seis primeiras semanas epidemiológicas de 2022, são os mesmos valores descritos na Tabela 8, e apresentados nas Figuras 13 e 14.

Tabela 8: Previsão para o número de casos de dengue semanal na cidade de Florianópolis segundo o modelo ARIMA(2, 1, 1) para as seis primeiras semanas epidemiológicas de 2022

Semana/Ano	Observado	Previsão	IC 95%
1/2022	44	28	18-39
2/2022	32	29	16-42
3/2022	22	27	12-42
4/2022	23	29	11-46
5/2022	17	27	9-46
6/2022	32	29	8-49

Observa-se a influência do *outlier* registrado na semana 1/2022 no resultado do MAPE. O valor acaba viesando a análise negativamente, uma vez que, se estivesse próximo a mediana dos valores observados, a medida seria minimizada e forneceria um melhor ajuste aos valores observados.

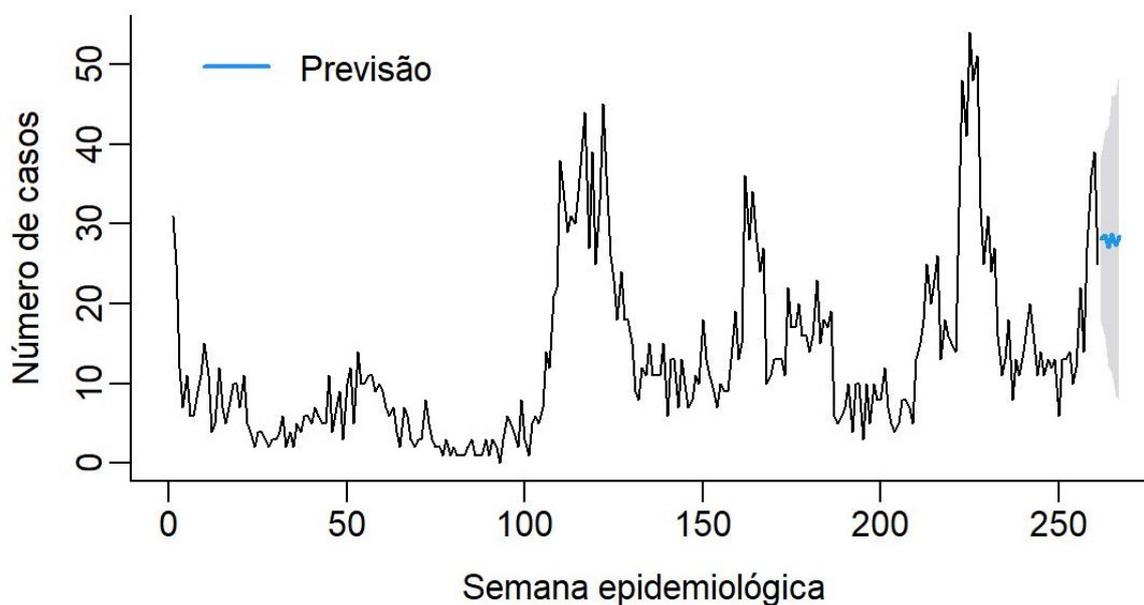


Figura 13: Série do número de casos de dengue em Florianópolis com previsão para 2022

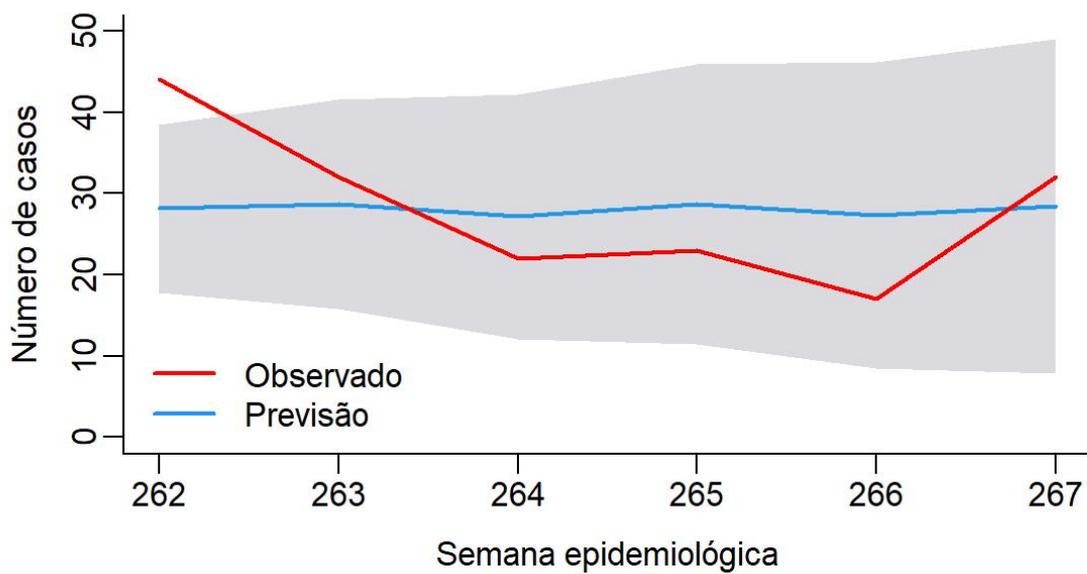


Figura 14: Valores observados e valores previstos para o número de casos de dengue em Florianópolis para 2022

## 4 Conclusões

Para a cidade de Florianópolis não foi possível identificar uma correlação temporal entre a série de casos de dengue e as séries climáticas de índice pluviométrico e temperatura mínima. Na realização de modelos utilizando função de transferência, incorporando essas covariáveis, observou-se ainda que não foram geradas previsões significativamente melhores que o modelo ARIMA(2,1,1), que utiliza apenas valores passados da série de casos de dengue na cidade.

O valores previstos para as seis primeiras semanas epidemiológicas do ano de 2022, pelo modelo ARIMA(2,1,1), mostraram-se boas estimativas comparadas aos valores observados, resultando em um MAPE de 25%. O intervalo de confiança gerado para o número de casos semanais, abrangeu o número real de casos de dengue em cinco das seis semanas onde foram realizadas previsões. Porém, o modelo não apresentou precisão frente a crescimentos/decrescimentos repentinos no número de casos da doença.

Apesar da revisão bibliográfica sugerir correlação de séries climáticas com a incidência de dengue em algumas cidades/estados do Brasil, considera-se, para fins de comparação, a influência direta das particularidades climáticas e econômicas de cada localidade, que podem fornecer fatores condicionantes para a adaptação e propagação do mosquito vetor. Observa-se o ineditismo do presente estudo sobre a doença na cidade, tendo em vista que a região Sul como um todo, era tida como uma região brasileira de improvável disseminação da dengue, além do crescimento do número de casos autóctones e frequentes epidemias registradas no estado de Santa Catarina.

Para trabalhos futuros, sugere-se a utilização de Modelos Dinâmicos Bayesianos para modelar a série e da utilização de Modelos de Poisson ou Binomiais Negativos para modelar dados de contagem, com o objetivo de melhorar o poder explicativo e identificação da relação entre a série de casos de dengue e séries climáticas. Além disso, a avaliação de *outliers* e de seu impacto, uso da transformação na série de dengue (para se aproximar mais à Normal) e nas covariáveis (como no caso do índice pluviométrico) para diminuir a

variabilidade e a realização de uma análise espaço-temporal para ver a difusão dos casos em microrregiões da cidade de Florianópolis ao longo do tempo.

# Referências

- AKAIKE, H. A new look at the statistical model identification. *IEEE transactions on automatic control*, Ieee, v. 19, n. 6, p. 716–723, 1974.
- ANDRADE, S. F. d. et al. Estudo de estratégias, bioclimáticas no clima de Florianópolis. 1996.
- BARBIERI, L. F. P. et al. Impact of climate variations and land use change: A mann-kendall application. *Revista Geama*, p. 127–135, 2017.
- BESERRA, E. B. et al. Biologia e exigências térmicas de *Aedes aegypti* (L.) (Diptera: Culicidae) provenientes de quatro regiões bioclimáticas da Paraíba. *Neotropical Entomology*, SciELO Brasil, v. 35, p. 853–860, 2006.
- BHATT, S. et al. The global distribution and burden of dengue. *Nature*, Nature Publishing Group, v. 496, n. 7446, p. 504–507, 2013.
- BICALHO, C. C.; SAFADI, T.; CHARRET, I. The influence of climatic factors on dengue epidemics in the cities Cuiabá (Mato Grosso State) and Lavras (Minas Gerais State), Brazil, using statistical methods. *Rev Bras Biom São Paulo*, v. 32, n. 2, p. 308–22, 2014.
- BOX, G.; JENKINS, G. 1970: Time series analysis, forecasting and control. San Francisco: Holden-Day, 1970.
- BOX, G. E. et al. *Time series analysis: forecasting and control*. [S.l.]: John Wiley & Sons, 2015.
- CALADO, D. C.; NAVARRO-SILVA, M. A. Influência da temperatura sobre a longevidade, fecundidade e atividade hematofágica de *Aedes (Stegomyia) albopictus* Skuse, 1894 (Diptera, Culicidae) sob condições de laboratório. *Revista Brasileira de Entomologia*, SciELO Brasil, v. 46, p. 93–98, 2002.
- CÂMARA, F. P. et al. Regional and dynamics characteristics of dengue in Brazil: a retrospective study. *Revista da Sociedade Brasileira de Medicina Tropical*, Sociedade Brasileira de Medicina Tropical, v. 40, n. 2, 2007.
- CHATFIELD, C. *The Analysis of Time Series: An Introduction with R*. 7. ed. [S.l.]: CRC Press, 2019. ISBN 9781498795630.
- CODECO, C. et al. Infodengue: A nowcasting system for the surveillance of arboviruses in Brazil. *Revue d'Épidémiologie et de Santé Publique*, v. 66, p. S386, 2018. ISSN 0398-7620. European Congress of Epidemiology “Crises, epidemiological transitions and the role of epidemiologists”. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0398762018311088>>.

- CODEÇO, C.; FALCÃO, M.; GRANATO, C. O que esta por tras de nova epidemia de dengue no brasil. 2022. Disponível em: <https://www.bbc.com/portuguese/brasil-61099201>.
- DENGUETECH. *Ciclo de vida do Aedes aegypti*. 2020. Disponível em: <https://denguetech.com.br/como-matar-larvas-mosquito-dengue/>.
- DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, Taylor & Francis, v. 74, n. 366a, p. 427–431, 1979.
- DIVE-SC. *Orientações para a integração das ações da Vigilância Epidemiológica e Vigilância Sanitaria diante da infestação do mosquito Aedes aegypti e transmissão da dengue no estado de Santa Catarina*. 2022. Disponível em: <https://www.dive.sc.gov.br/phocadownload/notas-tecnicas/notas-tecnicas-2022/NT031-2022>.
- EHLERS, R. S. *Análise de séries temporais*. [s.n.], 2009. Disponível em: <http://www.icmc.usp.br/ehlers/stemp/stemp.pdf.i>.
- FERREIRA, P. G. C.; SOUZA, R. C.; MATTOS, D. M. de. Box & jenkins com função de transferência usando o r: Um estudo para as series de ena do subsistema sul. 2015.
- FIOCRUZ. Dengue. 2010. Disponível em: <https://www.cpqrr.fiocruz.br/pg/dengue/>.
- FIOCRUZ, A. *Pesquisadores esclarecem dúvidas sobre a dengue e comentam hábitos do mosquito*. 2008. Disponível em: <https://agencia.fiocruz.br/pesquisadores-esclarecem-duvidas-sobre-a-dengue-e-comentam-habitos-do-mosquito>.
- G1-SC. *Florianópolis decreta situação de emergência por causa da dengue*. 2022. Disponível em: <https://g1.globo.com/sc/santa-catarina/noticia/2022/04/13/florianopolis-decreta-situacao-de-emergencia-por-cao-da-dengue.ghtml>.
- GUBLER, D. J. Dengue, urbanization and globalization: the unholy trinity of the 21st century. *Tropical medicine and health*, Japanese Society of Tropical Medicine, v. 39, n. 4SUPPLEMENT, p. S3–S11, 2011.
- GUJARATI, D. N.; PORTER, D. C. *Econometria básica-5*. [S.l.]: Amgh Editora, 2011.
- HIPEL, K. W.; MCLEOD, A. I. *Time series modelling of water resources and environmental systems*. [S.l.]: Elsevier, 1994.
- HOWE, G. M. et al. *A world geography of human diseases*. [S.l.]: Academic Press Inc.(London) Ltd., 24/28 Oval Road, London NW1 7DX., 1977.
- IBGE, C. Instituto brasileiro de geografia e estatística-ibge. 2021. Disponível em: <https://cidades.ibge.gov.br/brasil/sc/florianopolis/panorama>.
- INFODENGUE. Situação de casos estimados de arboviroses. 2022. Disponível em: <https://info.dengue.mat.br/>.
- INMET. Banco de dados meteorológicos do instituto nacional de meteorologia. 2022. Disponível em: <https://bdmep.inmet.gov.br/>.

- KUDRNA, G. A. et al. Análise temporal do número de casos de dengue na região sul do Brasil: relação entre a incidência e a letalidade. Florianópolis, SC, 2021.
- LIZZI, E. A. d. S. *Predição do número mensal de casos de dengue por modelos de séries temporais*. Tese (Doutorado) — Universidade de São Paulo, 2012.
- LJUNG, G. M.; BOX, G. E. On a measure of lack of fit in time series models. *Biometrika*, Oxford University Press, v. 65, n. 2, p. 297–303, 1978.
- MEIRA, R. Urucubaca” gripe ou dengue? dengue. .... *Clínica médica. São Paulo: Gráfica O Estado de S. Paulo*, p. 273–85, 1916.
- MESSINA, J. P. et al. The many projected futures of dengue. *Nature Reviews Microbiology*, Nature Publishing Group, v. 13, n. 4, p. 230–239, 2015.
- MINISTÉRIO DA SAÚDE, M. Dengue. 2020. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/d/dengue>.
- MORETTIN, P.; TOLOI, C. Análise de séries temporais—2ª edição revista e ampliada. *ABE-Projeto Fisher, Editora Edgar Blücher*, 2006.
- NOGUEIRA, R. M. R. et al. Dengue virus type 3, Brazil, 2002. *Emerging infectious diseases*, Centers for Disease Control and Prevention, v. 11, n. 9, p. 1376, 2005.
- OLIVEIRA, F. E. M. d. Spss básico para análise de dados. *Rio de Janeiro: Ciência Moderna*, 2007.
- OLIVEIRA, R. M.; OLIVEIRA, L. R. M. Epidemiologia da dengue: análise em diversas regiões do Brasil. *EsSEX: Revista Científica*, v. 2, n. 2, p. 32–44, 2019.
- OSANAI, C. H. et al. Surto de dengue em Boa Vista, Roraima, nota prévia. *Rev. Inst. Med. Trop. São Paulo*, p. 53–4, 1983.
- PADILHA, J. B. et al. Modelagem box-jenkins (arima) para gestão de risco de preço na pecuária de corte do estado do Paraná. Congresso Virtual Brasileiro de Administração—CONVIBRA., p. 12, 2013.
- PAHO/WHO. *Dengue - Principais fatos*. 2022. Disponível em: <https://www.paho.org/pt/topicos/dengue>.
- PEDRO, A. O dengue em Nictheroy. *Brazil Médico*, v. 1, n. 13, p. 174–7, 1923.
- PINO, F. A. A questão da não normalidade: Uma revisão. *Revista de economia agrícola*, v. 61, n. 2, p. 17–33, 2014.
- PINTO, R. B. *Finanças comportamentais e metodologia Box-Jenkins: uma aplicação no mercado brasileiro*. Dissertação (Mestrado) — 81 f. Dissertação (Mestrado em Administração) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Administração, Rio de Janeiro, 2006.
- RÉGIS, T. d. C. et al. Um estudo para elaboração de atlas municipal na perspectiva da educação geográfica inclusiva: o atlas adaptado do município de Florianópolis. 2016.

- RODHAIN, F. R. Mosquito vectors and dengue virus-vector relationships. *Dengue and dengue hemorrhagic fever*, Cab International, p. 45–60, 1997.
- SANTOS, T. R. dos; TAVARES, R. Detecção de hot spots futuros de casos de dengue em minas gerais (pp. 68-78). *Revista da Estatística da Universidade Federal de Ouro Preto*, v. 3, n. 2, 2014.
- SECCHES, M. d. S. Previsão de casos de dengue no estado de minas gerais via análise de clusters e modelo sarima. 2019.
- SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). *Biometrika*, JSTOR, v. 52, n. 3/4, p. 591–611, 1965.
- SHEPARD, D. S. et al. Economic impact of dengue illness in the americas. *The American journal of tropical medicine and hygiene*, The American Society of Tropical Medicine and Hygiene, v. 84, n. 2, p. 200, 2011.
- SNEC, R. D. *Biometrika tables for statisticians, volume 2*. [S.l.]: Taylor & Francis, 1974.
- SOUZA, E. d.; ASSIS, S. d. Violência e mortalidade na américa latina: um estudo das causas externas de 1968 a 1986. *Rio de Janeiro: Centro Latino Americano de Estudos Sobre Violência e Saúde, Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz*, 1989.
- SOUZA, R. C.; CAMARGO, M. E. Análise e previsão de séries temporais: os modelos arima. *Ijuí: Sedigraf*, 1996.
- SOUZA, V. et al. Utilização da metodologia de box & jenkins na previsão do preço futuro pago as exportações paranaenses de madeira serrada. *V Encontro de Economia Paranaense*, 2007.
- STANAWAY, J. D. et al. The global burden of dengue: an analysis from the global burden of disease study 2013. *The Lancet infectious diseases*, Elsevier, v. 16, n. 6, p. 712–723, 2016.
- SUAYA, J. A. et al. Cost of dengue cases in eight countries in the americas and asia: a prospective study. Citeseer, 2009.
- TEAM, R. C. *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2012*. 2021.
- TEIXEIRA, M. d. G.; BARRETO, M. L.; GUERRA, Z. Epidemiologia e medidas de prevenção do dengue. *Informe epidemiológico do SUS*, Centro Nacional de Epidemiologia/Fundação Nacional de Saúde/Ministério da Saúde, v. 8, n. 4, p. 5–33, 1999.
- TEIXEIRA, M. d. G. et al. Dengue and dengue hemorrhagic fever epidemics in brazil: what research is needed based on trends, surveillance, and control experiences? *Cadernos de Saúde Pública*, SciELO Public Health, v. 21, p. 1307–1315, 2005.
- TIMERMAN, A.; NUNES, E.; LUZ, K. *Dengue no Brasil: doença urbana*. 1. ed. [S.l.]: Limay Editora, 2012. ISBN 9788566138009.

UJVARI, S. C. *A história da humanidade contada pelo vírus*. [S.l.]: Editora Contexto, 2012.

## APÊNDICE 1 – Comparação entre os valores da série e da série ajustada pelos modelos

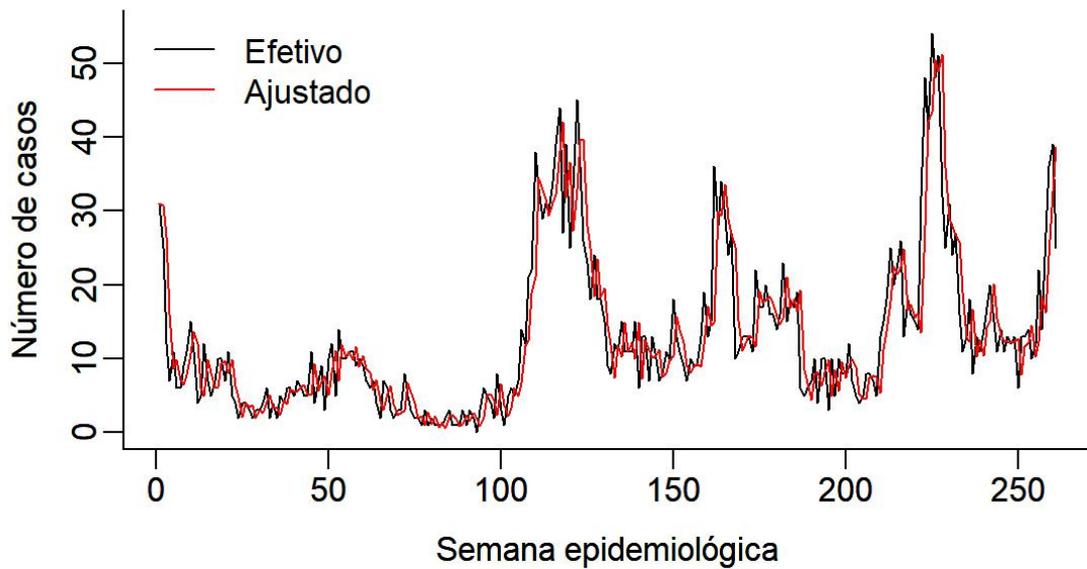


Figura 15: Valores efetivos e valores ajustados para a série do número de casos de dengue em Florianópolis

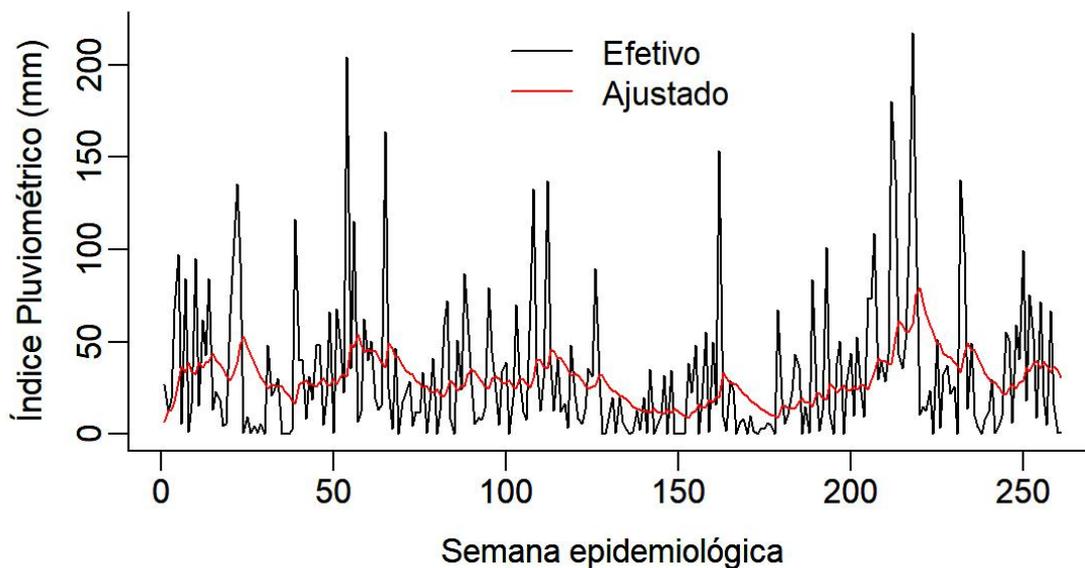


Figura 16: Valores efetivos e valores ajustados para a série do índice pluviométrico em Florianópolis

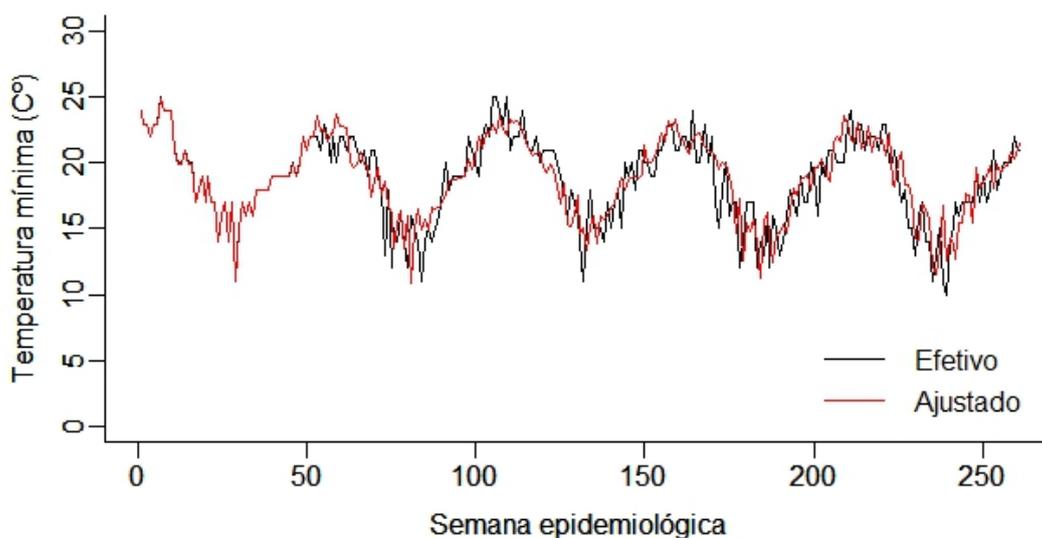


Figura 17: Valores efetivos e valores ajustados para a série de temperatura mínima (C°) em Florianópolis

## APÊNDICE 2 – Comparação de modelos para séries individuais

Tabela 9: Comparação dos modelos ARIMA para série temporal do índice pluviométrico em Florianópolis, através do AIC

Modelo	Identificação	AIC
1	ARIMA(2,0,2)	2631,644
2	ARIMA(0,0,1)	2630,913
<b>3</b>	<b>ARIMA(1,0,1)</b>	<b>2630,824</b>
4	ARIMA(2,0,0)	2631,681
5	ARIMA(2,0,1)	2631,325
6	ARIMA(1,0,0)	2690,988

Tabela 10: Comparação dos modelos SARIMA para série temporal de temperatura mínima (C°) em Florianópolis, através do AIC

Modelo	Identificação	AIC
1	SARIMA(2,0,1)(1,1,0) <sub>52</sub>	886,412
2	SARIMA(2,0,0)(1,1,0) <sub>52</sub>	888,432
<b>3</b>	<b>SARIMA(1,0,1)(1,1,0)<sub>52</sub></b>	<b>884,531</b>
4	SARIMA(1,0,1)(1,1,1) <sub>52</sub>	928,039
5	SARIMA(0,0,1)(1,1,0) <sub>52</sub>	890,029
6	SARIMA(0,0,2)(1,1,0) <sub>52</sub>	890,198

## APÊNDICE 3 – Análise de resíduos para os modelos ajustados de cada série

Tabela 11: Teste Ljung-Box para resíduos dos modelos

Modelo	Identificação	p-valor
Casos de dengue	ARIMA(2,1,1)	0,61
Índice pluviométrico	ARIMA(1,0,1)	0,09
Temperatura mínima	SARIMA(1,0,1)(1,1,0) <sub>52</sub>	0,08

Tabela 12: Teste Shapiro-Wilk para resíduos dos modelos

Modelo	Identificação	p-valor
Casos de dengue	ARIMA(2,1,1)	< 0,01
Índice pluviométrico	ARIMA(1,0,1)	< 0,01
Temperatura mínima	SARIMA(1,0,1)(1,1,0) <sub>52</sub>	< 0,01

## APÊNDICE 4 – Comparação dos modelos ARIMA com função de transferência

Tabela 13: Comparação dos modelos ARIMA com função de transferência com diferentes covariáveis (FT1), através do AIC

Modelo*	Identificação	AIC
1	Índ.Pluv. ( $\ell = 4$ ) + Temp. Mín. ( $\ell = 5$ )	1592,32
2	Índ.Pluv. ( $\ell = 3$ ) + Temp. Mín. ( $\ell = 2$ )	1582,00
3	Índ.Pluv. ( $\ell = 2$ ) + Temp. Mín. ( $\ell = 3$ )	1573,90
4	<b>Índ.Pluv. (<math>\ell = 7</math>) + Temp. Mín. (<math>\ell = 5</math>)</b>	<b>1554,70</b>
5	Índ.Pluv. ( $\ell = 7$ ) + Temp. Mín. ( $\ell = 4$ )	1568,73
6	Índ.Pluv. ( $\ell = 7$ ) + Temp. Mín. ( $\ell = 6$ )	1580,42

\*Na notação utilizada,  $\ell$  indica o número de observações pretéritas incorporadas ao modelo.