

**MESTRADO EM**  
**MÉTODOS QUANTITATIVOS PARA DECISÕES ECONÓMICAS**  
**E EMPRESARIAIS**

**TRABALHO FINAL DE MESTRADO**  
**DISSERTAÇÃO**

**PREVISÃO DOS INDICADORES DO COVID-19**

**MARGARIDA FARIA CARVALHO LOPES VIEIRA**

**ORIENTAÇÃO:**

**PROFESSOR DOUTOR JORGE CAIADO**

**OUTUBRO - 2022**

## Resumo

A presente dissertação centra-se na modelação e previsão dos principais indicadores da doença Covid-19: número diário de novos casos, óbitos e internamentos. Pretende-se identificar e ajustar os modelos que melhor preveem as séries temporais em estudo. Para tal, apresentamos os principais métodos determinísticos e estocásticos de previsão de séries temporais univariadas: métodos de alisamento exponencial (Holt-Winters) e modelos autoregressivos e de médias móveis (ARMA). Para avaliar a qualidade preditiva dos métodos em estudo, usamos as medidas dos erros de previsão do Erro Quadrático Médio, Erro Absoluto Médio e Erro Percentual Absoluto Médio. Os resultados obtidos apontam para a escolha dos métodos mais simples e com menor esforço computacional para prever qualquer dos indicadores do Covid-19.

**Palavras-chave:** Covid-19, Métodos de previsão, Modelo ARMA, Métodos de Holt-Winters, Erros de previsão.

## **Abstract**

The present dissertation focuses on the modeling and prediction of the main indicators of the Covid-19 disease: daily number of new cases, deaths and hospitalizations. We aim to identify and adjust the models that best predict the time series under study. To accomplish this task, we will present the main deterministic and stochastic methods of forecasting univariate time series: exponential smoothing methods (Holt-Winters) and autoregressive and moving average models (ARMA). To assess the predictive quality of the methods under study, we used the Mean Squared Error, Mean Absolute Error and Mean Absolute Percentage Error measurements. The results obtained show that the choice of the simplest methods and with less computational effort to predict any of the indicators of Covid-19.

**Keywords:** Covid-19, Forecast Methods, ARMA model, Holt-Winters method, Forecast Errors

## Nota prévia

A presente dissertação é o resultado do trabalho de investigação desenvolvido para a obtenção do grau de Mestre em Métodos Quantitativos para Decisões Económicas e Empresariais, junto do Instituto Superior de Economia e Gestão, da Universidade de Lisboa.

O texto, incluindo as referências nele incluídas, encontra-se atualizado a março de 2022. As ligações para textos disponíveis *online* foram consultadas até à mesma data.

É devido o agradecimento ao Professor Doutor Jorge Caiado, pelo apoio, disponibilidade e entusiasmo com que acompanhou este projeto.

Aos meus pais, o agradecimento pelo apoio incondicional e incentivo a concluir mais uma importante etapa.

Aos meus amigos, por me recordarem sempre que com calma, tudo corre melhor, e em especial à Beatriz pelo apoio fundamental já tão perto da meta.

Por fim, quero dedicar este trabalho à minha avó, que apesar de não estar presente, sei que sempre me apoiou e incentivou a conseguir alcançar os meus objetivos.

## Lista de abreviaturas e siglas utilizadas

ACF	Autocorrelation Function
ADF	Augmented Dickley-Fuller
AIC	Akaike's Information Criterion
AR	Autoregressive model
ARMA	Autoregressive moving average model
ARIMA	Autoregressive Integrated Moving Average
BIC	Bayesian Information Criterion
COVID-19	“Corona”, “Vírus” e “Doença”, 2019
DGS	Direção Geral de Saúde
HQIC	Hannan–Quinn Information Criterion
HW	Holt-Winters
HWAAS	Modelo Aditivo de Holt-Winters
I.e.	<i>Id est</i>
MA	Moving average model
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MSE	Mean Squared Error
OMS	Organização Mundial de Saúde
PACF	Partial Autocorrelation Function
RMSE	Root Mean Squared Error
SARS-CoV-2	Síndrome Respiratória Aguda Grave – Coronavírus 2

# Índice

Resumo .....	i
Abstract.....	ii
Nota prévia .....	iii
Lista de abreviaturas e siglas utilizadas.....	iv
Índice Figuras .....	iv
1. Introdução.....	7
1.1 Enquadramento .....	7
1.2 Objetivos do trabalho.....	8
1.3 Estrutura.....	8
2. Revisão da literatura.....	10
2.1 Definição COVID-19.....	10
2.1.1 O que é uma epidemia? .....	11
2.1.2 O que é uma pandemia?.....	11
2.2 Métodos de Previsão .....	12
2.2.1 Séries temporais.....	12
2.2.1.1 Etapas da previsão quantitativa.....	13
2.2.1.2 Modelos de séries temporais.....	14
2.3 Tópicos básicos de previsão.....	17
2.3.1 Estacionaridade.....	17
2.3.2 Função de autocovariância, autocorrelação e autocorrelação parcial.....	20
2.3.3 Função de autocorrelação (ACF).....	21
2.3.4 Função de autocorrelação parcial (PACF).....	23
2.3.5 Ruído Branco.....	24
3. Metodologia .....	25
3.1 Processos Autorregressivos .....	25
3.2 Processo de médias móveis de ordem $q$ , $MA(q)$ .....	26
3.3 Processo autorregressivo e de médias móveis, $ARMA(p, q)$ .....	28
3.4 Processo Integrado Autorregressivo e Médias Móveis.....	29
3.5 Modelos SARIMA .....	30
3.6 Métodos de alisamento exponencial .....	31
3.6.1 Alisamento exponencial simples .....	31
3.6.2 Método de Holt.....	32

3.6.3	Método de Holt-Winters.....	32
3.7	Erros de previsão .....	34
3.8	Seleção dos modelos ARIMA/SARIMA.....	35
4.	Caso de estudo da previsão do Covid-19 .....	37
4.1	Descrição da amostra .....	40
4.2	Modelação.....	40
4.3	Tratamento dos dados .....	44
4.3.1	Conceitos .....	44
4.3.2	Avaliação da estacionaridade .....	46
5.	Análise dos resultados.....	46
5.1	Análise descritiva das séries .....	47
5.2	Estacionariedade da série.....	52
5.3	Avaliação dos erros de previsão .....	58
5.4	Utilização do “melhor” método na previsão futura do Covid-19.....	59
6.	Conclusões .....	61
	Referências Bibliográficas.....	64

# Índice Figuras

Figura 1 - Série não estacionária em média.....	19
Figura 2 - Série não estacionária em média e em variância .....	19
Figura 3 - Série estacionária em média mas não em variância.....	20
Figura 4 - Série estacionária em média e em variância .....	20
Figura 5 - Série com dependência de longa duração .....	22
Figura 6 - Série com sazonalidade.....	23
Figura 7 - Simulação de um processo de médias móveis de ordem 1, AR(1), com $\varphi = 0,7$ , e respectivas ACF e PACF empíricas .....	26
Figura 8 - Simulação de um processo de médias móveis de ordem 1, MA(1), com $\theta = 0,7$ , e respectivas ACF e PACF empíricas .....	27
Figura 9 - Simulação de um processo autorregressivo e de médias móveis ARMA(1,1), com $\varphi = 0,85$ e $\theta = 0,5$ , e respectivas ACF e PACF empíricas .....	29
Figura 10 - Simulação de um modelo sazonal escrito como ARIMA, SARIMA(0,1,0)(1,1,1) <sub>12</sub> , e respectivas ACF e PACF empíricas .....	30
Figura 11 - Diagrama Box-Jenkins.....	36
Figura 12 - Cronograma da série “Novos Casos”.....	47
Figura 13 - Cronograma da série “Óbitos”.....	48
Figura 14 - Cronograma da série “Internamentos”.....	48
Figura 15 - Decomposição da série “Novos Casos”.....	49
Figura 16 - Decomposição da série “Óbitos por dia”.....	49
Figura 17 - Decomposição da série “Internamentos”.....	50
Figura 18 - ACF e PACF da série "Novos Casos" .....	51
Figura 19 - ACF e PACF da série "Óbitos por dia" .....	51
Figura 20 - ACF e PACF da série "Internamentos" .....	52
Figura 21 - Teste Dickey-Fuller da série original “Novos Casos” .....	53
Figura 22 - Teste Dickey-Fuller da série original “Óbitos por dia” .....	53
Figura 23 - Teste Dickey-Fuller da série original “Internamentos” .....	53
Figura 24 - Diferenciação sazonal e não sazonal da série "Novos Casos" .....	54
Figura 25 - Diferenciação sazonal e não sazonal da série "Óbitos por dia" .....	54
Figura 26 - Diferenciação sazonal e não sazonal da série "Internamentos" .....	55
Figura 27 - ACF e PACF da série transformada "Novos Casos" .....	55



Figura 28 - Comparação de valores dos Critérios de Informação para os modelos da série transformada "Novos Casos" .....	56
Figura 29 - ACF e PACF da série transformada "Óbitos por dia" .....	56
Figura 30 - Comparação de valores dos Critérios de Informação para os modelos da série transformada "Óbitos por dia" .....	57
Figura 31 - ACF e PACF da série transformada "Internamentos" .....	57
Figura 32 - Comparação de valores dos Critérios de Informação para os modelos da série transformada "Internamentos" .....	57
Figura 33 - Erros MAPE da série "Novos Casos" .....	58
Figura 34 - Erros MAPE da série "Óbitos por dia" .....	58
Figura 35 - Erros MAPE da série "Internamentos" .....	59
Figura 36 - Previsão a 7 dias para a série "Novos Casos" .....	59
Figura 37 - Previsão a 7 dias para a série "Óbitos por dia" .....	60
Figura 38 - Previsão a 7 dias para a série "Internamentos" .....	60



# 1. Introdução

## 1.1 Enquadramento

Desde dezembro de 2019, o mundo vive alarmado com a evolução do número de casos provocados pela doença Covid-19. Consequentemente o número de novos casos provoca um aumento no número de óbitos e também de internamentos. Durante dois anos, tem sido notícia de abertura em todos os canais de informação, a evolução da pandemia. Vimos países desertos, constantes estados de emergência, aprendemos a trabalhar em casa, vimos a primeira toma da vacina, e vimos também esperança.

Em dois anos, as pessoas viveram uma nova perspectiva de vive, criticada por quase todos, mas que neste momento se tornou um hábito. Tornou-se uma preocupação constante, a prevenção contra os novos casos, quer seja por parte dos governos com a implementação de medidas, quer seja por nós próprios para não nos infetarmos, nem infetarmos quem nos é mais querido. Vimos muita gente partir por causa da doença, mas o mundo não baixou os braços e tentou alcançar o impossível, encontrar uma vacina num curto espaço de tempo, com a pouca informação que se tinha da doença.

No dia 29 de setembro de 2022, o Conselho de Ministros decidiu não prorrogar a situação de alerta devido à Covid-19, a partir do dia 1 de outubro de 2022, e implicou a revogação de vários decretos-lei, como o fim do isolamento para os casos confirmados. O ministro da saúde, Dr. Manuel Pizarro, defendeu que Portugal está numa posição favorável face à pandemia, graças à alta taxa de vacinação do país, à menor agressividade das variantes do vírus e ao menor impacto nos serviços de saúde.

Mas este caminho só foi possível de atingir porque houve vários grupos de investigadores que conseguiram prever com precisão o aumento de novos casos. A pandemia tem causado distúrbios socioeconómicos em todo o mundo, obrigando os governos a introduzir medidas extremas para reduzir a sua propagação. Como forma de prevenção, muitos governos verificaram que era necessário entender o comportamento da disseminação da doença Covid-19, bem como a projeção de infeções e óbitos.

## **1.2 Objetivos do trabalho**

Este TFM tem por base o estudo das variáveis que incorporam a doença Covid-19, número diário de Novos Casos, Óbitos e Internamentos.

Os objetivos desta dissertação levantaram várias questões de investigação que irão ser respondidas no final deste trabalho, designadamente: Como se pode prever o número diário de novos caso, óbitos e internamentos? Será que as previsões vão ao encontro com a realidade?

Para ser possível responder aos objetivos deste trabalho, serão estudadas e analisadas diversas metodologias, nas diferentes séries em estudo, para que no final seja possível prever com qualidade e precisão.

## **1.3 Estrutura**

A estrutura deste TFM encontra-se dividida em seis capítulos. No primeiro está englobado a presente introdução, com a indicação da motivação do problema e os seus objetivos.

No segundo capítulo é feita a revisão da literatura e são apresentados os principais objetivos da previsão.

No terceiro capítulo é descrita a metodologia, ou seja, alguns dos métodos de previsão presentes na literatura e as várias etapas de previsão.

De seguida, no quarto capítulo apresenta-se o tratamento e caracterização do conjunto de dados abordados neste TFM.

O quinto capítulo corresponde à análise dos resultados alcançados e as respetivas conclusões com o intuito de se descobrir qual os melhores modelos de previsão.

Por último, no sexto capítulo, são apresentadas as respetivas conclusões.

## 2. Revisão da literatura

### 2.1 Definição COVID-19

O SARS-CoV-2, também conhecido como Covid-19, é uma doença infectocontagiosa, que provocou uma pandemia de elevado impacto na saúde da população mundial, podendo trazer diversas consequências se não for detetada e tratada a tempo.

Estas infeções afetam o sistema respiratório, sendo assim comparadas às constipações comuns, podendo mesmo evoluir para uma doença mais grave, como a pneumonia. Foi identificada pela primeira vez em dezembro de 2019 na China, na cidade de Wuhan. Este novo agente nunca tinha sido identificado anteriormente em seres humanos e, a sua fonte da infeção é ainda desconhecida, apesar de se desconfiar que tenha sido introduzido na espécie humana por transmissão zoonótica, ou seja, a partir de uma espécie animal.

Segundo o diretor da OMS (Organização Mundial de Saúde), é necessário saber compreender o significado da palavra pandemia, pois *não é uma palavra para ser usada de forma leve ou descuidada. É uma palavra que, se usada incorretamente, pode causar medo irracional ou aceitação injustificada de que a luta acabou, levando a sofrimento desnecessário*<sup>1</sup>. Muitas vezes, esta palavra é usada num contexto incorreto, como por exemplo, quando nos queremos referir a uma epidemia. Assim, é importante saber distinguir uma epidemia de uma pandemia.

---

<sup>1</sup> In [O que distingue uma epidemia de uma pandemia?](https://sicnoticias.pt) - SIC Notícias (sicnoticias.pt), consultado a 20/03/2022

### **2.1.1 O que é uma epidemia?**

Uma epidemia corresponde à *manifestação coletiva de uma doença que rapidamente se espalha, por contágio direto ou indireto, até atingir um grande número de pessoas em um determinado território e que depois se extingue após um período*. Tal acontece nos países onde se registam epidemias de gripe todos os anos, durante a época de outono e inverno os casos aumentam, atinge-se um pico e depois diminui.

A principal diferença entre uma epidemia e uma pandemia é o foco. No caso de uma epidemia, nem todos os países estão em riscos, pois esta pode ocorrer apenas dentro de fronteiras, ou seja, o foco principal é o país onde ocorre. Numa pandemia, o foco não tem fronteiras e as medidas de prevenção com o intuito de ajudar a população são a nível mundial.

### **2.1.2 O que é uma pandemia?**

Não podemos definir uma pandemia como uma doença que apenas se espalha ou que mata um grande número de pessoas, deve também ser infecciosa. Sendo assim, definimos uma pandemia, como uma epidemia de doença infecciosa, mas que se espalha entre a população numa grande região geográfica. Significa isto que, o termo pandemia, está associado a uma disseminação mundial de uma doença, que se espalha por diferentes continentes, afetando um grande número de pessoas ao mesmo tempo.

Segundo a OMS, existem três condições que marcam o início de uma pandemia:

- O surgimento de uma nova doença na população;
- Infeção de humanos, que acarretam sintomas atípicos;
- A doença espalha-se facilmente entre humanos.

No dia 11 de março de 2020, a OMS declarou a Covid-19, como uma pandemia mundial.

## 2.2 Métodos de Previsão

“Uma previsão consiste na tentativa de extrapolar o comportamento futuro a partir das condições atuais, utilizando dados históricos do passado” (Caiado, 2016).

Segundo Gonçalves (2010), os modelos de previsão dividem-se em dois tipos:

- i. Modelos qualitativos. São considerados modelos subjetivos, por se basearem em estimativas e opiniões. São utilizados sempre que não existam dados históricos, ou caso existam, não sejam representativos do futuro. As técnicas qualitativas são utilizadas para previsões a médio e a longo prazo;
- ii. Modelos quantitativos. Baseiam-se na utilização de técnicas e modelos estatísticos, que assentam em dados históricos para estimar ou prever os valores futuros. Identifica padrões nos dados históricos e transpõe-nos para o futuro.

### 2.2.1 Séries temporais

Define-se por uma série temporal (*time series*) um “conjunto de observações de uma variável, feitas em períodos sucessivos, durante um determinado intervalo” (Caiado, 2016).



A representação gráfica de uma série temporal designa-se por cronograma e constitui o ponto de partida para a sua análise. A análise de um cronograma, permite destacar algumas características evidenciadas pela série e propor o tipo de modelação mais adequado ao seu estudo e previsão.

### **2.2.1.1 Etapas da previsão quantitativa**

Segundo Caiado (2016), num processo de previsão de séries temporais, deve seguir-se um conjunto de etapas fundamentais:

- i. Definição do problema;
- ii. Recolha dos dados;
- iii. Análise exploratória;
- iv. Explicação ou modelação;
- v. Previsão.

#### **Definição do problema**

Definir o problema, envolve recolha de dados para o desenvolvimento, manutenção e gestão das bases de dados, e para o manuseamento e utilização dos dados em previsão.

#### **Recolha dos dados**

Uma base de dados sustenta todo o estudo, por isso é importante construir uma com dados fidedignos.

### **Análise exploratória**

A descrição dos dados começa com a inspeção gráfica da série e caracterização do seu andamento geral, procurando identificar os movimentos de tendência, sazonalidade, cíclicos, mudança de estrutura e eventuais observações anómalas (*outliers*). A análise gráfica vai auxiliar na escolha de potenciais modelos candidatos a representar a série observada.

### **Explicação ou modelação**

Consiste na construção e seleção de modelos que permitem explicar o comportamento da série no período observado.

### **Previsão**

Na última fase, tenta-se prever a evolução futura da série com base exclusivamente no seu comportamento passado (modelos univariados ou não-casuais) ou com base no comportamento passado de outras variáveis (modelos multivariados ou explicativos). Este processo nunca é definitivo e requer a avaliação prévia da qualidade preditiva dos modelos para os últimos instantes observados da série em estudo.

#### **2.2.1.2 Modelos de séries temporais**

Os métodos de decomposição utilizam alisadores de médias móveis para eliminar o carácter aleatório das séries temporais. Não podem ser considerados métodos de previsão, mas são úteis para a análise as séries temporais, pois ao isolarmos as suas forças componentes, obtemos uma melhor explicação do seu padrão de comportamento, facilitando assim, a tarefa de previsão (Caiado, 2016).

Segundo Murteira (2000), na decomposição clássica de uma série temporal pode observar-se em quatro componentes: tendência ( $T_t$ ), sazonalidade ( $S_t$ ), componente cíclica ( $C_t$ ) e componente residual ( $\varepsilon_t$ ). Estas podem ser descritas como:

- i. Tendência ( $T_t$ ): refere-se a qualquer mudança sistemática no nível de uma série ao longo do tempo, descrevendo movimentos que se manifestam suave e consistentemente durante períodos longos. Esta componente caracteriza o comportamento mais notório da série (crescente, decrescente ou constante) durante o período considerado.
- ii. Sazonalidade ( $S_t$ ): variações ao longo da tendência que ocorrem com periodicidade curta, ou seja, corresponde a um padrão de aumento e diminuição que ocorre regularmente na série em períodos específicos. O movimento dentro de um período tem duração fixa e, podem ocorrer semanalmente, mensalmente ou trimestralmente no decurso do ano. A sazonalidade pode ser classificada como aditiva, se não depender do nível da série, ou multiplicativa, quando é proporcional ao mesmo (Chatfield, 2000).
- iii. Componente cíclica ( $C_t$ ): define-se por um padrão de flutuação que, ao contrário da componente sazonal, não apresenta qualquer periodicidade definida. Relacionam-se sobretudo com as fases de expansão e recessão dos sistemas económicos, constituindo assim, flutuações positivas e negativas. Em ciclos longos, muitas vezes pode ser difícil separar as componentes da tendência e cíclica, pelo que se pode tomar estas como uma única componente (tendência-cíclica) (Caiado, 2016).

- iv. Componente residual ( $\varepsilon_t$ ): é a componente que contém qualquer variação não explicada pelas restantes componentes e representa o ruído aleatório. É considerada um ruído branco quando é modelada por um processo estocástico de variáveis aleatórias não correlacionadas e identicamente distribuídas.

A série pode ser escrita como função destas quatro componentes, ou seja:

$$Y_t = f(T_t, S_t, C_t, \varepsilon_t) \quad (2.1)$$

Segundo Caiado (2016), o modelo multiplicativo descre a série como o produto das componentes tendência, sazonalidade, cíclica e irregular, *i.e.*,

$$Y_t = T_t \times S_t \times C_t \times \varepsilon_t. \quad (2.2)$$

O modelo de decomposição aditivo descreve cada valor da série temporal como sendo a soma das suas componentes, *i.e.*,

$$Y_t = T_t + S_t + C_t + \varepsilon_t, \quad (2.3)$$

onde  $Y_t$  representa o valor observado no instante  $t$ ,  $T_t$  a componente tendência no instante  $t$ ,  $S_t$  a componente sazonalidade no instante  $t$ ,  $C_t$  a componente cíclica no instante  $t$  e,  $\varepsilon_t$  a componente residual no instante  $t$ . De referir que, muitas vezes torna-se difícil separar a componente cíclica da tendência, sendo por isso referida como componente de tendência-cíclica.

Enquanto o modelo aditivo é o mais apropriado quando a magnitude das oscilações sazonais ou a variação em torno do ciclo de tendência não varia com o nível da série temporal. Quando se verifica o contrário, ou seja, quando se verificam aumentos ou diminuições proporcionais com a tendência da série, devemos optar pelo modelo multiplicativo. Contudo, é importante referir que, o modelo de decomposição multiplicativo não deve ser implementado para séries temporais com valores negativos ou nulos.

## 2.3 Tópicos básicos de previsão

### 2.3.1 Estacionaridade

Um processo estocástico  $\{Y_t, t \in T\}$  é uma sequência de variáveis aleatórias indexadas no tempo. Matematicamente, um processo estocástico é definido como uma coleção de variáveis aleatórias ordenadas no tempo e definidas num conjunto de índices  $t$ , que pode ser contínuo ou discreto. (Brockwell e Davis, 1991)

Os processos estocásticos dividem-se em estacionários e não estacionários.

#### Processos estocásticos estacionários

*Um processo estocástico  $\{Y_t, t \in T\}$  diz-se estritamente estacionário (ou fortemente estacionário) se e só se a distribuição conjunta de  $Y_{t_1}, \dots, Y_{t_n}$  é igual à distribuição conjunta de  $Y_{t_1+k}, \dots, Y_{t_n+k}$ , qualquer que seja o  $n$ -úpla  $(t_1, \dots, t_n)$  e para qualquer  $k$ , i.e.,*

$$F_{(Y_{t_1}, \dots, Y_{t_n})}(y_1, \dots, y_n) = F_{(Y_{t_1+k}, \dots, Y_{t_n+k})}(y_1, \dots, y_n) \quad (2.4)$$

em todos os pontos  $(y_1, \dots, y_n)$ .

A estacionaridade no sentido estrito é uma propriedade demasiado exigente e, na maioria dos casos, de difícil verificação. Os processos estacionários de 2.<sup>a</sup> ordem (ou fracamente estacionários) obedecem a uma propriedade mais fraca, mas que, grosso modo, descreve o mesmo tipo de comportamento (Murteira et al., 1993).

*Um processo  $\{Y_t, t \in T\}$  diz-se estacionário de 2.<sup>a</sup> ordem (ou fracamente estacionário) se e só se todos os momentos até à 2.<sup>a</sup> ordem de  $Y_{t_1}, \dots, Y_{t_n}$  existem e são iguais aos momentos correspondentes até à 2.<sup>a</sup> ordem de  $Y_{t_1+k}, \dots, Y_{t_n+k}$ . Logo, num processo fracamente estacionário:*

1. *O valor médio não depende de  $t$ , i.e.,  $\mu_t = \mu$ ;*
2. *A variância não depende de  $t$ , i.e.,  $\sigma_t^2 = \sigma^2$ ;*
3. *A covariância de  $Y_{t_1}$  e  $Y_{t_2}$  depende apenas do desfasamento  $t_2 - t_1$ , i.e.,*  

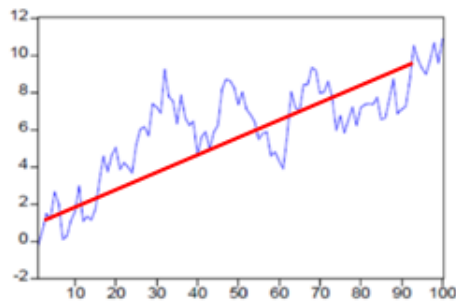
$$Cov[Y_{t_1}, Y_{t_2}] = \gamma(|t_2 - t_1|).$$

Um processo estritamente estacionário é também estacionário de 2.<sup>a</sup> ordem, mas o contrário nem sempre se verifica.

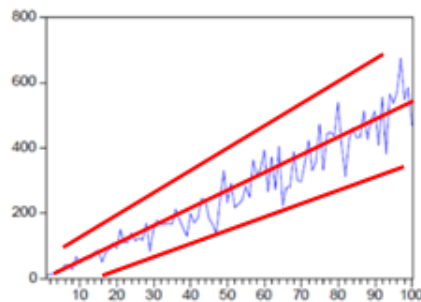
### **Processos estocásticos não estacionários**

Um processo estocástico é considerado não estacionário quando a média e/ou variância variam com o tempo. Caso a estacionaridade não se verifique, poderá ser usada a diferenciação, que consiste em desfasar a série temporal em determinado número de instantes com o objetivo de estabilizar a média da série. Uma série que é não estacionária em média é consequentemente não estacionária em variância e covariância (Wei, 2007).

Da observação das figuras seguintes, pode-se concluir, os diferentes tipos de estacionaridade. Na Figura 1 verifica-se a existência de tendência pelo que a média não é constante, logo é não estacionária em média. Na Figura 2 conclui-se que, quer a média quer a variância vão aumentado ao longo do tempo, pelo que esta série é não estacionária em média e em variância. No entanto, na Figura 3 podemos observar que uma série pode ser estacionária em média, mas não em variância. Por último, podemos observar uma série estacionária tanto em média como em variância (Figura 4).



*Figura 1 - Série não estacionária em média*



*Figura 2 - Série não estacionária em média e em variância*

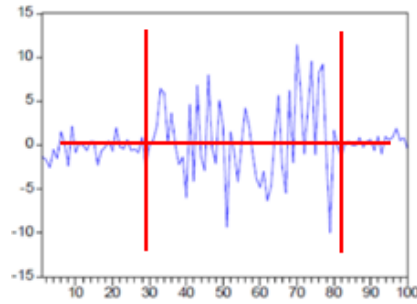


Figura 3 - Série estacionária em média mas não em variância

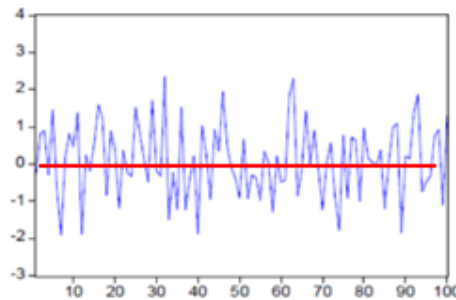


Figura 4 - Série estacionária em média e em variância

### 2.3.2 Função de autocovariância, autocorrelação e autocorrelação parcial

Para um processo estacionário, define-se a função de autocovariância

$$\gamma_k = Cov[Y_t, Y_{t+k}] = E[(Y_t - \mu)(Y_{t+k} - \mu)] \quad (2.5)$$

que mede a intensidade com que covariam pares de valores do processo separados por um intervalo (lag) de amplitude  $k$ .

A função de autocovariância  $\gamma_k$  é definida para  $k \in \mathbb{R}$  se o processo é de tempo contínuo e para  $k \in \mathbb{Z}$  se for de tempo discreto, i.e.,  $k = 0, \pm 1, \pm 2, \dots$ . Além disso esta função satisfaz as seguintes propriedades:

1.  $\gamma_0 = Cov[Y_t, Y_t] = Var[Y_t] = \sigma^2$ ;



2.  $\gamma_k = \gamma_{-k}$ , i.e., a função é par e dispensa a representação gráfica para  $k < 0$ ;
3.  $|\gamma_k| \leq \gamma_0$ , como consequência da desigualdade de Cauchy-Schwarz;
4. A função  $\gamma_k$  é semidefinida positiva, i.e., para qualquer conjunto de números reais  $\alpha_1, \dots, \alpha_n$  e instantes de tempo  $t_1, \dots, t_n$ ,

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \gamma(|t_i - t_j|) \geq 0 \quad (2.6)$$

### 2.3.3 Função de autocorrelação (ACF)

Para um processo estacionário, define-se a função de autocorrelação, do inglês, *Autocorrelation Function (ACF)*

$$\rho_k = \text{Corr}[Y_t, Y_{t+k}] = \frac{\text{Cov}[Y_t, Y_{t+k}]}{\sqrt{\text{Var}[Y_t] \text{Var}[Y_{t+k}]}} = \frac{\text{Cov}[Y_t, Y_{t+k}]}{\text{Var}[Y_t]} = \frac{\gamma_k}{\gamma_0}, \quad (2.7)$$

$$k = 1, \dots, n - 1$$

que mede a correlação entre pares de valores do processo separados por um intervalo (lag) de amplitude  $k$ , e pode se estimada por:

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=k+1}^n (Y_t - \bar{Y})^2}, k = 1, \dots, n - 1 \quad (2.8)$$

À medida que a amplitude do intervalo ( $k$ ) aumenta é de esperar que a capacidade de memória do processo seja limitada, e, portanto, que no momento  $t + k$  se encontre pouco refletido o que se passou no momento  $t$  (Murteira et al., 1993). Assim, é espectável

que a correlação temporal diminua ( $\rho_k \rightarrow 0$ ), quando o desfasamento entre duas observações aumenta ( $k \rightarrow +\infty$ ). Segundo Murteira et al. (1993), refere que pode interpretar-se  $\rho_k$  como uma medida da semelhança entre cada realização e a mesma realização deslocada  $k$  unidades de tempo.

A função de autocorrelação  $\rho_k$  também pode estar definida para  $k \in \mathbb{R}$  ou  $k \in \mathbb{Z}$ , consoante o processo for de tempo contínuo ou discreto, respetivamente. A função satisfaz as seguintes propriedades:

1.  $\rho_0 = \text{Corr}[Y_t, Y_t] = 1$ ;
2.  $\rho_k = \rho_{-k}$ , i.e., a função é par e dispensa a representação gráfica para  $k < 0$ ;
3.  $|\rho_k| \leq 1$ , como consequência da desigualdade de Cauchy-Schwarz;
4. A função  $\rho_k$  é semidefinida positiva, i.e., para qualquer conjunto de números reais  $\alpha_1, \dots, \alpha_n$  e instantes de tempo  $t_1, \dots, t_n$ ,

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \rho(|t_i - t_j|) \geq 0 \quad (2.9)$$

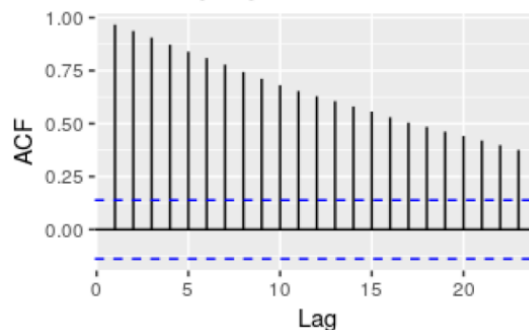


Figura 5 - Série com dependência de longa duração

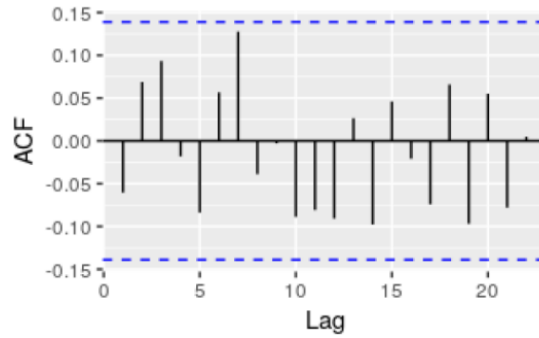


Figura 6 - Série com sazonalidade

### 2.3.4 Função de autocorrelação parcial (PACF)

Para além de se analisar a correlação total entre  $Y_t$  e  $Y_{t-k}$ , interessa investigar a correlação simples entre  $Y_t$  e  $Y_{t-k}$ , depois de eliminar o efeito que sobre elas exercem os valores intermédios  $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$ , ou seja, se o primeiro elemento da série está fortemente relacionado com o segundo, e o segundo com o terceiro, então o primeiro está de alguma forma também relacionado com o primeiro. Esta correlação condicional na análise temporal designa-se por função de autocorrelação parcial, do inglês *Partial Autocorrelation Function* (PACF).

A estimação da função de autocorrelação parcial pode ser feita através de um método recursivo, através da expressão:

$$\varphi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \varphi_{k-1,j} \rho_{k-j}}{1 - \sum_{j=1}^{k-1} \varphi_{k-1,j} \rho_{k-j}} \quad (2.10)$$

Com  $\phi_{11} = \rho_1$  (inicialização) e  $\rho_{kj} = \rho_{k-1,j} - \rho_{kk} \rho_{k-1,k-j}$ ,  $j = 1, 2, \dots, k-1$ .

### 2.3.5 Ruído Branco

Um processo estocástico  $\{\varepsilon_t\}$  é chamado de ruído branco (*white noise*) se for uma sequência de variáveis aleatórias não correlacionadas provenientes da mesma distribuição com média e variâncias constantes. (Murteira, 2000)

1.  $E(\varepsilon_t) = \mu_\varepsilon$ ;
2.  $V(\varepsilon_t) = \sigma_\varepsilon^2$ ;
3.  $Cov(\varepsilon_t, \varepsilon_{t-k}) = 0$ , para todo o  $k \neq 0$ .

As FAC e FACP de um ruído branco são igualmente nulas para todo o  $k \neq 0$ . Embora o ruído branco dificilmente se observe em séries reais, este desempenha um papel fundamental na construção de modelos probabilísticos ou estocásticos. Por outro lado, um bom modelo de previsão deve ser aquele que produz erros de previsão com comportamento análogo a um ruído branco, isto porque um ruído branco é imprevisível.

### 3. Metodologia

No âmbito desta dissertação, propõe-se responder às seguintes questões:

1. Qual o modelo de previsão a adotar por cada uma das séries em estudo, para prever, o número diário de novos casos, óbitos e internamentos?
2. Existe alguma relação entre as variáveis?

Neste capítulo são apresentadas as diferentes metodologias consideradas para a análise dos dados estudados nesta dissertação. Neste caso, será descrita a metodologia subjacente aos modelos integrados autoregressivos e de médias móveis (ARIMA) e aos métodos de alisamento exponencial (simples, Holt e Holt-Winters). No que se refere aos modelos ARIMA, definem-se os seus casos particulares, a saber: AR, MA e ARMA.

Será utilizado o *software* com linguagem Python para fazer a estimação dos modelos e a previsão dos dados.

#### 3.1 Processos Autorregressivos

Baseiam-se no pressuposto de que a observação da variável no instante  $t$  se relaciona, de forma linear, com as observações nos instantes anteriores. Assim, o processo  $Y_t$  diz-se um processo autorregressivo de ordem  $p$ ,  $AR(p)$ , quando satisfaz a equação

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t, \text{ com } \varphi_p \neq 0, \quad (3.1)$$

onde  $\varepsilon_t$  é um ruído branco de média nula, independente de  $Y_{t-k}$  para todo o  $k \geq 1$ .

Se o processo  $Y_t$  é um processo  $AR(p)$ , então a sua função de autocorrelação parcial,  $\varphi_{kk}$ , é igual a zero para todo o  $k > p$ . Assim, a PACF de um processo  $AR(p)$  apresenta, graficamente, uma queda brusca para zero a partir do *lag*  $p + 1$ . No que respeita à ACF, esta tem um decaimento exponencial ou sinusoidal amortecido para zero.

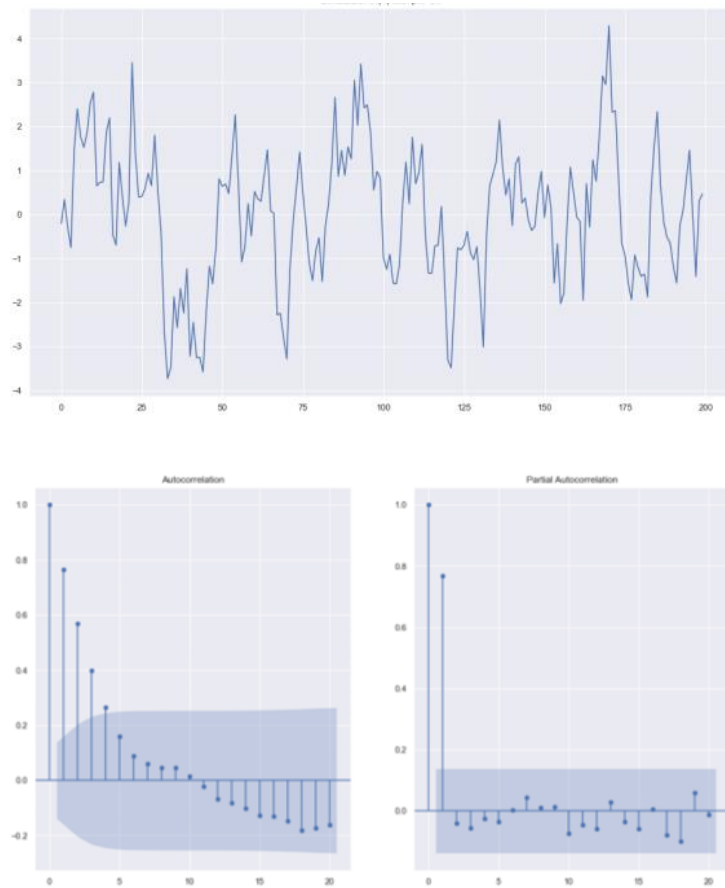


Figura 7 - Simulação de um processo de médias móveis de ordem 1,  $AR(1)$ , com  $\varphi = 0,7$ , e respetivas ACF e PACF empíricas

### 3.2 Processo de médias móveis de ordem $q$ , $MA(q)$

O processo  $Y_t$  diz-se um processo de médias móveis de ordem  $q$ ,  $MA(q)$ , quando assume a expressão

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}, \text{ com } \theta_q \neq 0 \quad (3.2)$$

Onde  $\varepsilon_t$  é um ruído branco de média zero e variância constante.

Se o processo  $Y_t$  é um processo  $MA(q)$ , então a sua função de autocorrelação parcial,  $\rho_k$ , é igual a zero para todo o  $k > q$ , e por isso, a ACF de um processo  $MA(q)$  apresenta, graficamente, uma queda brusca para zero a partir do *lag*  $q + 1$ . No que respeita à PACF, esta tem um decaimento exponencial ou sinusoidal amortecido para zero, tendo, assim, a mesma estrutura que a ACF de um processo  $AR(q)$ .

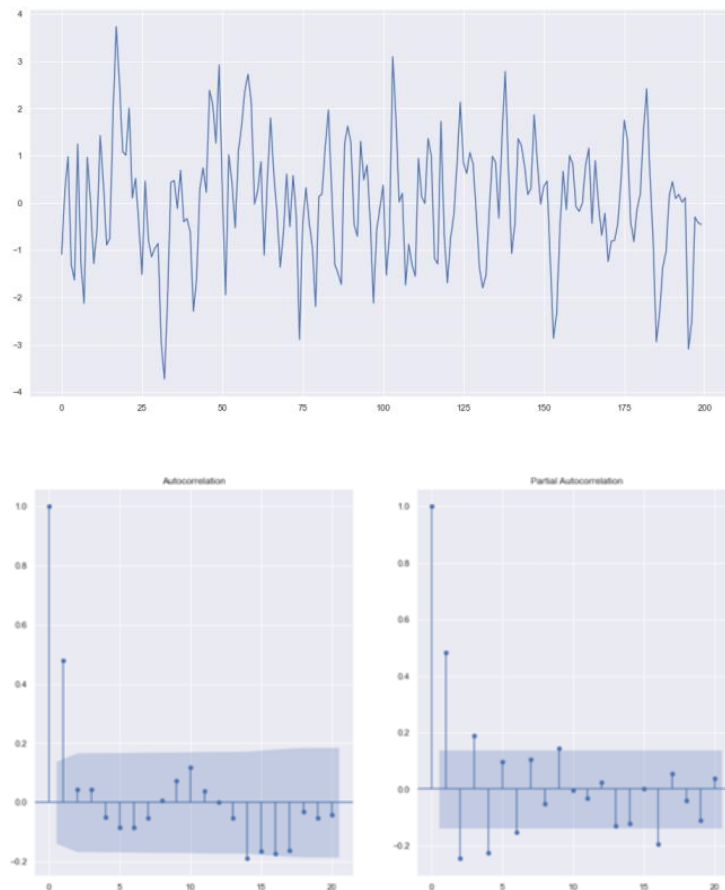


Figura 8 - Simulação de um processo de médias móveis de ordem 1,  $MA(1)$ , com  $\theta = 0,7$ , e respetivas ACF e PACF empíricas

### 3.3 Processo autorregressivo e de médias móveis, ARMA( $p, q$ )

Diz-se um processo autorregressivo e de médias móveis de ordem  $p$  e  $q$ , ARMA( $p, q$ ), se satisfaz a equação:

$$Y_t = \mu + \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (3.3)$$

$$Y_t = \sum_{i=1}^p \varphi_i Y_{t-i} + \mu_t + \sum_{i=0}^q \theta_i \varepsilon_{t-i} \quad (3.3)$$

onde,  $p$  é o grau de AR, e  $q$  representa o grau de MA.

Se  $p = 0$ , então o processo é MA( $q$ ); e, se  $q = 0$ , então o processo é AR( $p$ ). Os coeficientes  $p$  e  $q$  são escolhidos através da análise da autocorrelação global e parcial.

Inicialmente é medida para cada desfasamento de  $k$ , a relação entre  $Y_t$  e  $Y_{t-k}$ , considerando também as variáveis  $Y_{t-1}, \dots, Y_{t-k+1}$ , e também a relação entre  $Y_t$  e  $Y_{t-k}$ , sem considerar outras variáveis.

A FAC e a AFPC de um processo ARMA( $p, q$ ) resultam da combinação das respetivas funções dos processos AR( $p$ ) e MA( $q$ ). Como foi dito anteriormente, a ACF de um processo MA( $q$ ) é insignificante a partir do lag  $q + 1$ , e a PACF de um processo AR( $p$ ) depois do lag  $p$ .



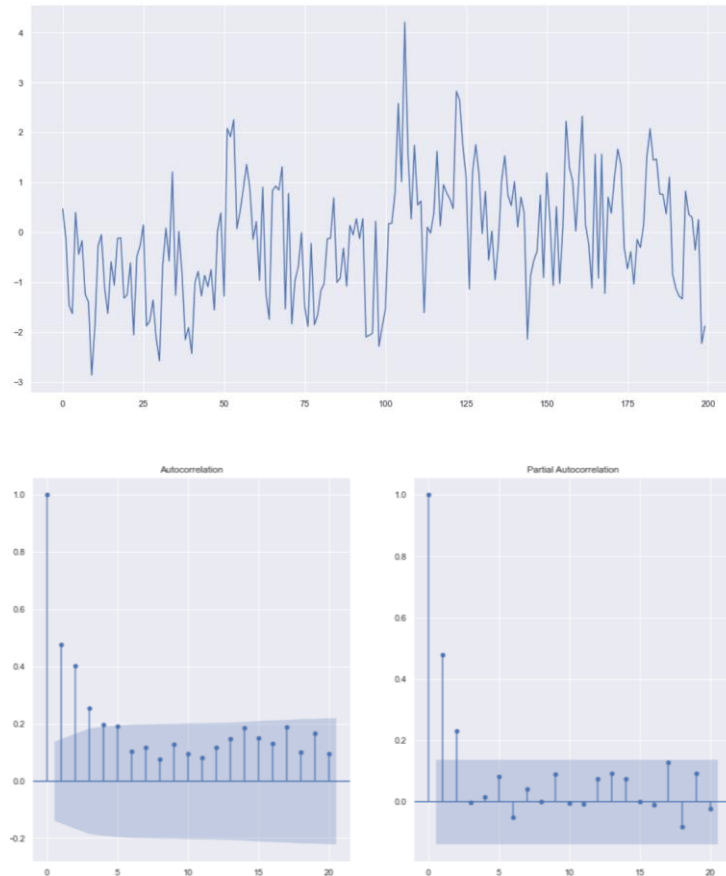


Figura 9 - Simulação de um processo autorregressivo e de médias móveis ARMA(1,1), com  $\varphi = 0,85$  e  $\theta = 0,5$ , e respectivas ACF e PACF empíricas

### 3.4 Processo Integrado Autorregressivo e Médias Móveis

Um processo integrado autorregressivo e médias móveis é uma generalização dos processos ARMA que incorpora séries não estacionárias. Segundo Brockwell e Davis (2002), o modelo ARIMA possui apenas uma variável e descreve o comportamento de uma variável em termos dos seus valores passados, como é referido nos modelos anteriores.

Os modelos ARIMA, introduzidos por Box e Jenkins (1970), são normalmente representados por  $ARIMA(p, d, q)$ , onde o parâmetro  $p$  refere-se à ordem da componente autorregressiva; o parâmetro  $d$  ao número de diferenciações que são necessárias para

transformar a série não estacionária em estacionária; e, o parâmetro  $q$  diz respeito à ordem da componente médias móveis. Assim, verifica-se que o modelo ARIMA resulta da junção de três componentes: (i) autorregressiva, (ii) diferenciação, e (iii) médias móveis.

### 3.5 Modelos SARIMA

Existem ainda modelos mistos que apresentam uma componente não sazonal e uma componente sazonal que são escritos como  $SARIMA(p, d, q)(P, D, Q)_s$ . Os inteiros  $(p, d, q)$  correspondem à parte não sazonal e os inteiros  $(P, D, Q)$  à parte sazonal. O inteiro  $s$  refere-se ao ciclo sazonal.

A etapa de identificação é crucial para a determinação dos valores dos inteiros  $p, d, q, P, D, Q$  que caracterizam o modelo  $SARIMA(p, d, q)(P, D, Q)_s$ . Em geral, nesta fase, o analista procura identificar os modelos candidatos a descrever a série em estudo com base no comportamento das funções de autocorrelação (ACF) e autocorrelação parcial (PACF). Segue-se a etapa da estimação dos seus parâmetros tal como foi demonstrado anteriormente no modelo ARIMA.

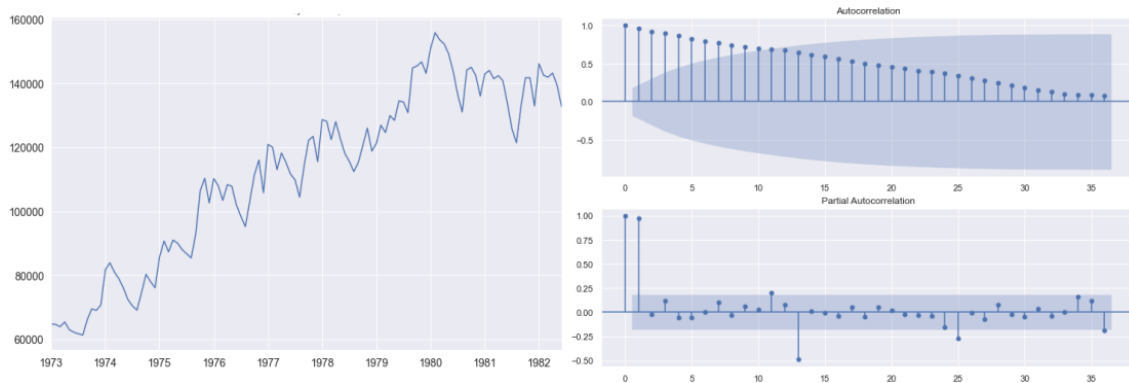


Figura 10 - Simulação de um modelo sazonal escrito como ARIMA,  $SARIMA(0,1,0)(1,1,1)_{12}$ , e respetivas ACF e PACF empíricas

### 3.6 Métodos de alisamento exponencial

Os métodos de alisamento exponencial foram introduzidos por Brown (1963), Holt (1957) e Winters (1960) e são métodos de previsão que utilizam combinações ponderadas das observações passadas para posteriormente preverem os valores futuros.

Hyndman (2012) refere que tanto o modelo ARIMA como o método de alisamento exponencial, constituem métodos de previsão univariados, uma vez que, só usam os valores históricos da própria série temporal para realizar as previsões.

#### 3.6.1 Alisamento exponencial simples

Segundo Caiado (2016), o método de alisamento exponencial simples é adequado para séries com tendência localmente estacionária e sem movimentos sazonais. É um dos métodos mais simples e é definido como uma combinação linear de todas as observações conhecidas, dando maior peso às observações mais recentes.

O alisamento exponencial simples considera a previsão correspondente ao período anterior e faz um acerto para obter a previsão para o período seguinte. O acerto é realizado através da multiplicação do erro de previsão do período anterior por uma constante de alisamento  $0 < \alpha < 1$ . Quanto mais próximo de 1 estiver o valor de  $\alpha$ , maior o efeito de ajustamento do erro de previsão passado no cálculo da nova previsão. Por outro lado, quanto mais próximo de 0 estiver o valor de  $\alpha$ , menor o efeito de ajustamento do erro de previsão.

A representação do método de alisamento exponencial simples é expressa pela equação:

$$\hat{Y}_{t+1} = \hat{Y}_t + \alpha(Y_t - \hat{Y}_t) \quad (3.4)$$

Onde  $\hat{Y}_{t+1}$  é a previsão para o período  $t + 1$ ,  $\hat{Y}_t$  é a previsão para o período  $t$  com origem no período  $t - 1$ ,  $Y_t$  é o valor real observado no período  $t$  e  $\alpha$  a constante de alisamento.

### 3.6.2 Método de Holt

Quando uma série apresenta tendência linear, mas não movimentos de carácter sazonal, é adequado usar o método de Holt (1957). Holt desenvolveu um método mais abrangente, tendo por base o alisamento exponencial simples, incorporando a componente tendência na função de previsão. Este método utiliza as seguintes equações de atualização (Caiado, 2016):

$$a(t) = \alpha Y_t + (1 + \alpha)[a(t - 1) + b(t - 1)], \quad 0 < \alpha < 1 \quad (3.5)$$

$$b(t) = \beta[a(t) - a(t - 1)] + (1 - \beta)b(t - 1), \quad 0 < \alpha < 1, \quad (3.6)$$

onde  $a(t)$  e  $b(t)$  são equações de atualização do nível e do declive da série, respetivamente, e  $\alpha$  e  $\beta$  são as constantes de alisamento. As previsões a  $h$  passos à frente são obtidas através da função:

$$P_{t+h} = a(t) + b(t) \times h, \quad h = 1, 2, \dots \quad (3.7)$$

### 3.6.3 Método de Holt-Winters

O método de Holt foi desenvolvido por Winters, em 1960, para prever séries temporais com tendência e movimentos de carácter sazonal. O método de Holt-Winters é

bastante idêntico ao método de Holt, no entanto, adiciona uma equação para captar a componente sazonal. É, portanto, um método de alisamento exponencial que permite construir previsões, considerando o nível, a tendência e a sazonalidade (Makridakis et al., 1998).

O método de Holt-Winters apresenta dois métodos distintos de construir previsões: um incorpora a sazonalidade de forma aditiva e outro de forma multiplicativa. No nosso trabalho, usamos o método aditivo de Holt-Winters uma vez que não existe a tendência

O método aditivo de Holt-Winters baseia-se nas seguintes equações de atualização:

$$\text{Nível: } a(t) = \alpha(Y_t - S_{t-s}) + (1 - \alpha)[a(t - 1) + b(t - 1)] \quad (3.8)$$

$$\text{Declive: } b(t) = \beta[a(t) - a(t - 1)] + (1 - \beta)b(t - 1) \quad (3.9)$$

$$\text{Sazonalidade: } S_t = \gamma[Y_t - a(t)] + (1 - \gamma)S_{t-s} \quad (3.10)$$

onde  $S_t$  corresponde à estimativa do fator sazonal no período  $t$ ,  $\alpha$  a constante de amortecimento do nível, com  $0 \leq \alpha \leq 1$ ,  $\beta$  a constante de amortecimento da tendência, com  $0 \leq \beta \leq 1$ , e,  $\gamma$  a constante de amortecimento do fator sazonal, com  $0 \leq \gamma \leq 1$ .

As previsões a  $h$  passos à frente do modelo de sazonalidade aditiva são obtidas através da função:

$$P_{T+h} = a(T) + b(T) \times h + S_{T+h-s}, \quad h = 1, 2, \dots \quad (3.11)$$

### 3.7 Erros de previsão

Como avaliar a qualidade preditiva dos métodos usados? Se o futuro é desconhecido, tal só é possível fazendo um exercício de previsão para dentro da amostra e comparando as previsões com os valores reais. Com o intuito de comparar os diferentes métodos de previsão melhor descritos infra, serão usadas três medidas de avaliação dos erros de previsão: Erro Percentual Absoluto Médio (MAPE), Erro Quadrático Médio (MSE) e Erro Absoluto Médio (MAE). De seguida, são apresentadas estas grandezas, onde  $H$  representa o número de observações disponíveis para teste,  $Y_t$  o valor observado no instante  $t$  e  $\hat{Y}_t$  a previsão para o instante  $t$ . (Caiado, 2011; Ramos, 2012).

O *Mean Absolute Percentage Error* (MAPE), mede o tamanho do erro em termos percentuais. Segundo Davydenko e Fildes (2013), o MAPE é uma medida relativa que não depende da escala dos dados, já que divide a diferença ( $Y_t - \hat{Y}_t$ ) pelo valor atual ( $Y_t$ ).

$$MAPE = \frac{1}{H} \times \sum_{t=1}^H \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100 \quad (3.12)$$

O *Mean Square Error* (MSE) é obtido através de uma escala diferente do original, neste caso o quadrado, sendo a sua principal desvantagem a sensibilidade a erros elevados. Assim, assinala um desvio sistemático por defeito ou por excesso.

$$MSE = \frac{1}{H} \times \sum_{t=1}^H (Y_t - \hat{Y}_t)^2 \quad (3.13)$$

Por outro lado, o *Mean Absolute Error* (MAE) é obtido através da equação:

$$MAE = \frac{1}{H} \times \sum_{t=1}^H |Y_t - \hat{Y}_t| \quad (3.14)$$

O MAE e o MSE são úteis para comparar métodos aplicados à mesma série temporal, no entanto o MAE não inflaciona os erros mais elevados ao contrário do MSE.

### 3.8 Seleção dos modelos ARIMA/SARIMA

Na modelação e previsão de séries temporais com os modelos ARIMA/SARIMA é conveniente adotar um procedimento metodológico assente nas etapas de identificação, estimação, avaliação do diagnóstico, seleção de modelo e previsão. Box e Jenkins (1970, 1976), e mais tarde, Box et al. (1994), propuseram uma metodologia para a seleção do “melhor” modelos usando critérios de informação (AIC, BIC) que ajudam a medir a qualidade do ajustamento entre os vários modelos ajustados à série temporal. O melhor modelo, é aquele que otimiza o valor de cada um dos critérios. Assim, os critérios comparam os vários modelos estatísticos construídos com base nos dados e utilizando diferentes estruturas de parâmetros no modelo. Na Figura 11 ilustra-se a construção do modelo através da metodologia de Box-Jenkins.

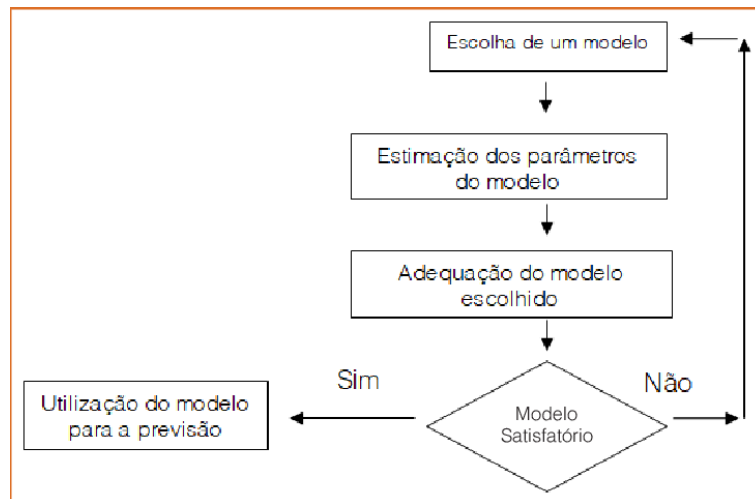


Figura 11 - Diagrama Box-Jenkins



## 4. Caso de estudo da previsão do Covid-19

No dia 2 de março de 2022, fez dois anos desde o início da pandemia, e desde então, Portugal já atravessou cinco vagas, mais de dez estados de emergência, dois confinamentos, e neste momento foram levantadas quase todas as medidas.

O primeiro caso de alerta em Portugal, ocorreu a 26 de janeiro de 2020, quando a Direção-Geral da Saúde (DGS), comunicou o primeiro caso suspeito de infeção pelo novo coronavírus, que após análises laborais obteve um resultado negativo. Tratava-se de um homem que tinha aterrado em Lisboa proveniente do “epicentro da epidemia”, a cidade de Wuhan, e que ligou para a linha do Serviço Nacional de Saúde, com sintomas do foro respiratório, tendo sido de imediato encaminhado para o Hospital de Curry Cabral.

No dia 26 de fevereiro de 2020, Portugal ainda não registava casos confirmados de infeção por coronavírus, mas foram registados 25 casos suspeitos, 18 dos quais tiveram resultado negativo após testes laborais. A informação existente era escassa, e apenas se sabia que a maioria dos casos tinham sido registados no Hospital de São João no Porto, e que todos os cidadãos suspeitos eram provenientes do norte de Itália.

A 2 de março de 2020, foram reportados pela ministra da Saúde, os dois primeiros casos de infeção pelo novo coronavírus. A partir deste primeiro anúncio a DGS começou a emitir boletins diários. A primeira morte provocada pelo Covid-19, foi anunciada quinze dias depois, pela ministra da Saúde. No dia 18 de março, o Presidente da República falava ao país, e anunciava que Portugal iria entrar em estado de emergência.

Como já foi referido, Portugal enfrentou cinco vagas pandémicas. A primeira começou em março de 2020, sendo que o pico foi atingido na primeira semana de abril. Nessa altura, era alarmante o país ter mais do que 1000 casos diários, valor que fica bem

aquém dos registados atualmente. No final do mês de abril, juntamente com a desaceleração dos casos diários, o Governo apresentou um plano de desconfinamento, sendo que no dia 1 de junho, ocorreu a reabertura das grandes superfícies comerciais.

Durante o verão, os números tendiam a estabilizar. No entanto a partir de setembro de 2020, o Governo viu-se obrigado a implementar um novo conjunto de medidas, no qual foi decretado estado de calamidade. Em novembro de 2020, foram apresentados números astronómicos, tanto de novas infeções, como de óbitos. A 27 de dezembro de 2020, chegou a primeira vacina contra o Covid-19 a Portugal.

A terceira vaga, ocorreu logo no primeiro mês de 2021, onde os internamentos em UCI atingiam números recorde. No dia 23 de janeiro de 2021, Portugal superou a barreira dos dez mil mortos por Covid-19. No estado em que se encontrava os internamentos por Covid-19 em Portugal, o Ministério da Saúde, viu-se obrigado a pedir ajuda externa para ajudar no tratamento de doentes nos hospitais portugueses.

Em julho, com o surgimento da variante *Delta*, ocorre a quarta vaga da pandemia. Para fazer face a esta nova variante, o Governo apostou fortemente na vacinação, e a 9 de outubro de 2021, Portugal atingiu a marca dos 85% da população com a vacinação completa, tornando-se no primeiro país do mundo a consegui-lo. Contudo, e com a chegada dos meses de inverno, o número de novos casos disparou. Para piorar a situação vivida, é identificada uma nova variante, a *Ómicron*, que provoca em janeiro de 2022 a quinta vaga.

Depois de atingido o pico, começou a notar-se num decréscimo de novos casos, pelo que o governo decidiu aliviar as medidas de combate à pandemia, mantendo apenas a obrigatoriedade de uso de máscaras na generalidade dos espaços públicos fechados.

No entanto, a 21 de abril de 2022, o governo acabou com a obrigatoriedade da máscara, com algumas exceções. Marta Temido admitia que estavam *reunidas as condições para a não obrigatoriedade do uso de máscara* nos espaços fechados. Afirmava também que *não estamos no patamar ideal, mas entendemos assumir com transparência que o caminho feito permite neste momento alterar o enquadramento que tínhamos. Estamos a assumir que as circunstâncias da pandemia mudaram*. Passadas três semanas, já se afirmava que a sexta vaga se começava a “desenhar” em Portugal. Verificou-se assim que, a eliminação do uso de máscara aumentou as infeções em Portugal, o índice de transmissibilidade ( $R_t$ ) era de 1,17, e a marca dos 20 mil novos casos diários foi ultrapassada pela primeira vez desde 15 de fevereiro.<sup>2</sup>

A 26 de maio, a ministra de Estado e da Presidência, Mariana Vieira da Silva, afirmava que *a análise dos números indica que, muito provavelmente, o pico já terá passado, com algumas regiões já com quedas visíveis e de vários grupos etários também, pelo que a decisão do Conselho de Ministros foi a de manter as medidas em vigor*. No entanto, o relatório de situação referente ao período entre 17 de maio e 23 de maio, indicava 188.970 novos casos de Covid-19, o que corresponde a um aumento de 12.699 casos face aos registados na semana anterior. Era possível verificar também que, com o aumento do número de óbitos, a taxa de mortalidade encontrava-se em 22 óbitos por milhão de habitantes, a sete dias, verificando-se uma subida de 20% em relação à semana anterior. Contudo, o risco de transmissibilidade ( $R_t$ ) diminuía.<sup>3</sup>

---

<sup>2</sup> In [Governo deixa cair obrigatoriedade do uso de máscaras, excepto em lares e transportes coletivos \(com áudio\) – O Jornal Económico \(jornaleconomico.pt\)](https://jornaleconomico.pt), consultado a 30/04/2022

<sup>3</sup> In [Governo mantém medidas Covid. Ministra diz que já “terá passado” o pico da sexta vaga em Portugal – ECO \(sapo.pt\)](https://sapo.pt), consultado a 15/06/2022

## 4.1 Descrição da amostra

Em resultado da pandemia, muitas organizações e investigadores, organizaram conjuntos de dados, acessíveis a todos, sobre os números do Covid-19 para análise e pesquisa.

Para a realização deste TFM, irão ser usados os dados dos boletins diários, disponíveis na página Web da DGS, direcionada para esta doença (<https://covid19.min-saude.pt/>). Neste estudo é feita referência a dados relativos aos números diários de novos casos, óbitos e internamentos. Foram também analisados os dados referentes à vacinação. Esta informação foi compilada e tratada numa folha de cálculo Excel.

## 4.2 Modelação

Atualmente, a prevenção da disseminação da doença Covid-19 tornou-se uma das maiores preocupações mundiais, com governos e empresas de diversos países a investir em iniciativas para garantir o controlo da evolução da pandemia.

Em resposta à pandemia, muitos países implementaram medidas como o isolamento e distanciamento social para evitar maior disseminação.

Tornou-se assim urgente a capacidade de identificar a taxa de propagação da doença, pois estar ciente do nível de disseminação em qualquer momento, permite aos governos ajudar no planeamento da saúde pública, bem como na implementação de políticas para lidar com as diversas consequências causadas pela pandemia.

A componente de metodologia adotada neste TFM baseia-se numa revisão sistemática de literatura, sobre o Covid-19 e métodos de previsão em diferentes países. Foi escolhido este método pelo facto de ter como objetivo estudar a literatura relativa a um tema através da análise duma amostra de artigos equilibrada e imparcial (Nightingale, 2009). Neste sentido, este estudo pretende prever com precisão o número de novos casos, mas também e com a ajuda de variáveis explicativas entender o aumento do número de óbitos e internamentos face ao aumento do número de novos casos.

Como tal, um dos objetivos deste trabalho é disponibilizar uma base de investigação, para que no futuro possa ser possível prevenir com precisão situações semelhantes. Assim, o processo de pesquisa, baseou-se principalmente em artigos escritos por outros investigadores sobre o mesmo tema, mas em países diferentes.

Foi realizado um estudo (Papastefanopoulos, V., Linardatos, P., & Kotsiantis, S., 2020) em dez países diferentes, EUA, Reino Unido, Itália, Espanha, Rússia, França, Turquia, Alemanha, Irão e Brasil, teve como objetivo prever, com precisão, os diversos picos da propagação da doença. Ao conseguir prever novas vagas da doença, permite aos governos de cada país alterar as suas políticas e planear atempadamente as medidas preventivas necessárias. Esta análise foi feita através de diferentes métodos de previsão de séries temporais, como:

1. ARIMA;
2. Modelo aditivo de Holt-Winters («HWAAS»);
3. TBAT, acrónimo para quatro componentes: formulação sazonal trigonométrica, transformação Box-Cox, erros ARIMA e tendência;

4. Facebook's Prophet, modelo de previsão de séries temporais desenvolvido pelo Facebook e, originalmente projetado para lidar com problemas de séries temporais de negócio;
5. DeepAR, método de previsão baseado em redes neuronais recorrentes autoregressivas;
6. N-Beats, análise de expansão de redes neuronais para previsão de séries temporais interpretáveis.

Verificou-se assim que, o resultado de desempenho de cada modelo para cada país indica que, não existe uma abordagem única para prever a percentagem de casos ativos em relação à população total de cada país. Usando a raiz do erro quadrático médio (RMSE) para avaliar o desempenho de cada modelo, constatou-se que as abordagens ARIMA, TBAT e HWAAS foram as que demonstraram um melhor desempenho.

Verificou-se também que é difícil identificar as razões exatas pelas quais certos algoritmos funcionam melhor do que outros em certos países, e entre os fatores observados, podem-se destacar os seguintes:

1. Características climáticas e geográficas de cada país;
2. Diferentes atributos relacionados à população, como densidade populacional;
3. Discrepâncias nos procedimentos de teste e recolha de dados entre os diferentes países;
4. Diferentes medidas implementadas em cada país, como a quarentena ou o distanciamento social.

Wang (2021) comparou diferentes métodos para a previsão da 2.<sup>a</sup> vaga de Covid-19 na Índia e nos EUA. Os desempenhos de cada modelo foram avaliados pelo erro percentual absoluto médio (MAPE).

Neste caso os modelos em estudo foram os seguintes:

1. ARIMA;
2. GRNN, modelos de redes neuronais de regressão generalizada;
3. Modelo híbrido ARIMA-GRNN;
4. Método de alisamento exponencial.

Observou-se que o modelo ARIMA foi o melhor modelo de previsão do número de novos casos diários de Covid-19 na Índia, e o modelo híbrido ARIMA-GRNN teve melhor desempenho de previsão nos EUA.

Mais uma vez verifica-se que, na monitorização e previsão da pandemia, a tendência nos diferentes países pode ser comparada para analisar de forma abrangente a situação pandémica, mas as medidas usadas devem ser reformuladas de acordo com o impacto causado pelo Covid-19 nos diferentes países, de forma a controlar o surto.

Após a análise vários *papers*, concluímos que a grande maioria usa modelos como o ARIMA e o modelo aditivo de *Holt-Winters* para prever com precisão os picos e determinar novas vagas da pandemia.

Na análise descritiva serão retiradas as últimas observações e serão ajustadas aos seguintes modelos para cada série:

1. Modelo Holt-Winters com sazonalidade aditiva;
2. Modelo SARIMA( $p, q, d$ )( $P, Q, D$ )<sub>s</sub> identificado através de testes de raízes unitárias (ADF) e das funções de autocorrelação (ACF) e autocorrelação parcial (PACF).

## 4.3 Tratamento dos dados

### 4.3.1 Conceitos

Quando foram feitos os primeiros boletins diários, estes continham os casos importados, as cadeias de transmissão ativas, o número de casos suspeitos, os sintomas declarados e até o número de portugueses infetados fora do país. Atualmente, os Relatórios de Situação, são muito mais complexos e detalhados, e como tal, é importante compreender certos conceitos:

- Casos ativos: são considerados casos ativos, quando através da realização de um teste, se obtém resultado positivo. Representa assim, o número total de infetados num determinado momento que após confirmação de infeção pelo coronavírus, não evoluíram a óbito, nem se enquadram nos critérios de casos recuperados;
- Recuperados: uma pessoa que está em casa infetada com o novo coronavírus, é considerada um caso recuperado se ao fim de 14 dias tiverem um único teste negativo e não apresentarem sintomas. Se se tratar de um doente internado, são necessários dois testes negativos, com um intervalo de 24 horas. O número apresentado é um total acumulado;
- Óbitos: pacientes que faleceram em consequência da Covid-19, após confirmação laboral ou por critério clínico epidemiológico;



- Contactos em vigilância: um contacto é uma pessoa que teve exposta a um caso confirmado de infeção por Covid-19, dentro de período de transmissibilidade/infecciosidade, ou material biológico infetado. Podem estar em vigilância passiva, ou seja, contactos de baixo risco com pessoas infetadas e que fazem auto-monitorização de sintomas, ou podem estar em vigilância ativa se são contactos de alto risco que são acompanhados diariamente pela Unidade de Saúde Pública. Apenas estes últimos, são apresentados nos boletins diários;
- Casos confirmados: representa o total acumulado de casos confirmados por teste de SARS-CoV-2 desde a chegada da Covid-19 a Portugal;
- Casos suspeitos: são considerados casos suspeitos, as pessoas que apresentam sintomas, ou os casos de vigilância que apresentam sintomas, até obterem o resultado do teste. O número apresentado diz respeito ao número de casos que aguardam resultado num determinado dia.
- Pessoas vacinadas: número de indivíduos com registo de inoculação de dose de vacina contra Covid-19 no sistema VACINAS (Plataforma Nacional de Registo e Gestão da Vacinação).
- Vacinação completa: pessoas vacinadas com o número de doses recomendadas para a vacina Covid-19 que foi administrada, registado no sistema VACINAS.
- Cobertura vacinal: proporção da população residente que tem registo de vacinação contra Covid-19 no sistema VACINAS.

#### **4.3.2 Avaliação da estacionaridade**

A escolha do modelo deverá ter em conta a estacionaridade e a análise das funções ACF e PACF. Se uma série não for estacionária em média, deve ser diferenciada até que verifique o contrário. Neste sentido, para se perceber se a série é ou não estacionária, será realizado um teste de Dickey-Fuller Aumentado (ADF).

## **5. Análise dos resultados**

Neste capítulo são apresentados e analisados os dados em estudo. São ainda aplicados os métodos de previsão aos modelos referidos anteriormente (Holt-Winters

Aditivo e SARIMA), e é indicado o procedimento adotado para a previsão do caso em estudo. Por último, é apresentada uma síntese conclusiva e discussão sobre os resultados obtidos.

## 5.1 Análise descritiva das séries

A fim de se seleccionar a melhor metodologia adequada para cada série temporal, é necessário analisar os diferentes métodos, verificando aqueles que melhor se adaptam aos dados existentes.

Nas Figuras seguintes, são apresentadas as séries originais dos Novos Casos, Internamentos e Óbitos, com os seus valores expressos em unidades. Da observação do comportamento das séries, nota-se algumas oscilações e alterações de comportamento ao longo do tempo que podem sugerir a não estacionariedade (em média e mesmo em variância) das séries em estudo.

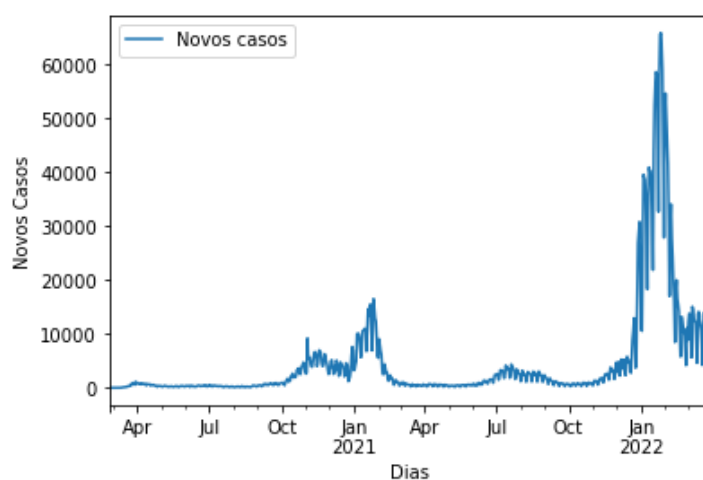


Figura 12 - Cronograma da série "Novos Casos"

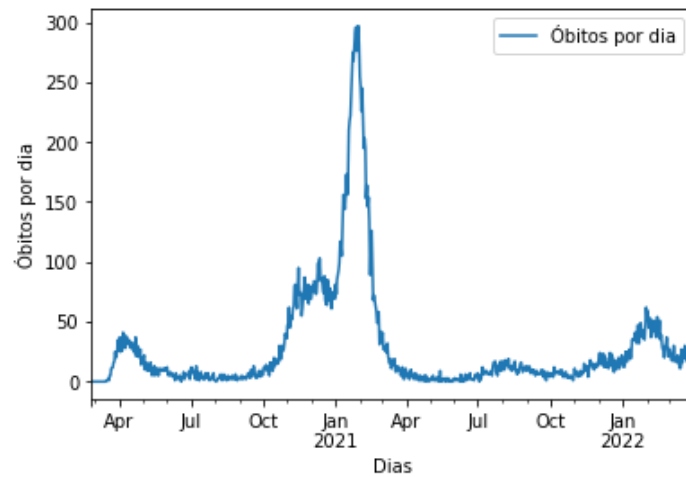


Figura 13 - Cronograma da série "Óbitos"

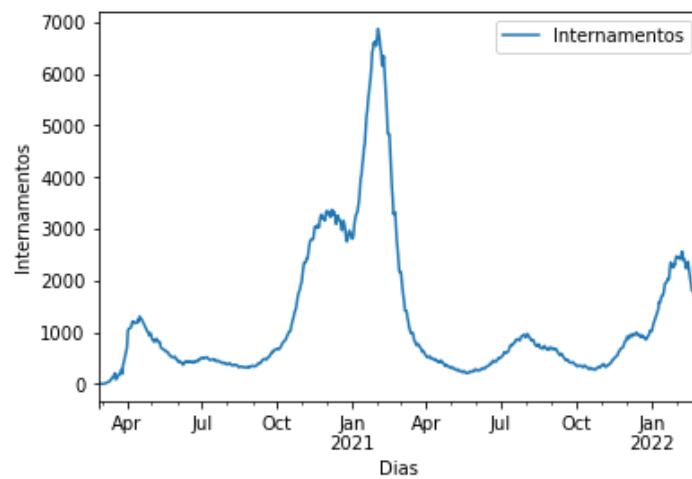


Figura 14 - Cronograma da série "Internamentos"

Com o intuito de perceber melhor as séries, foi feita a decomposição em três componentes: tendência ( $T_t$ ), sazonalidade ( $S_t$ ), e componente residual ( $\varepsilon_t$ ). Como foi descrito anteriormente o método de decomposição utilizado foi a decomposição clássica. Nas figuras seguintes é possível observar a decomposição em componentes de cada uma das séries:

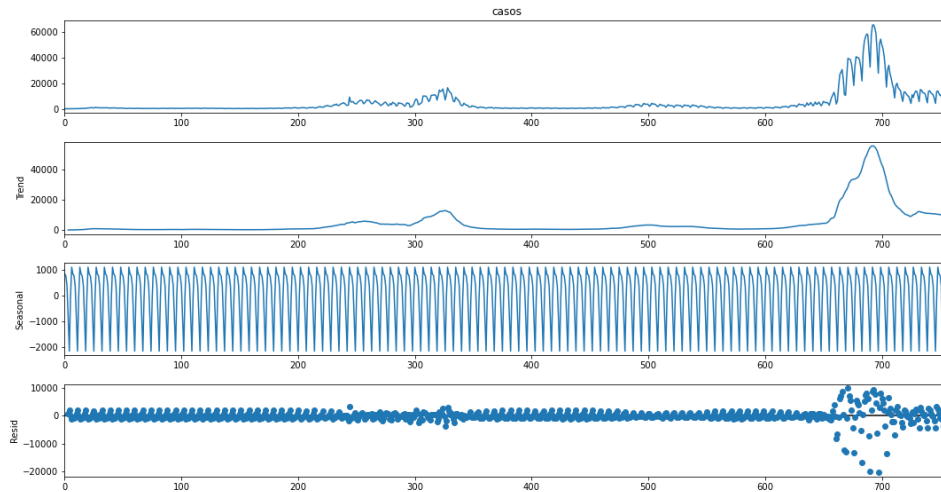


Figura 15 - Decomposição da série "Novos Casos"

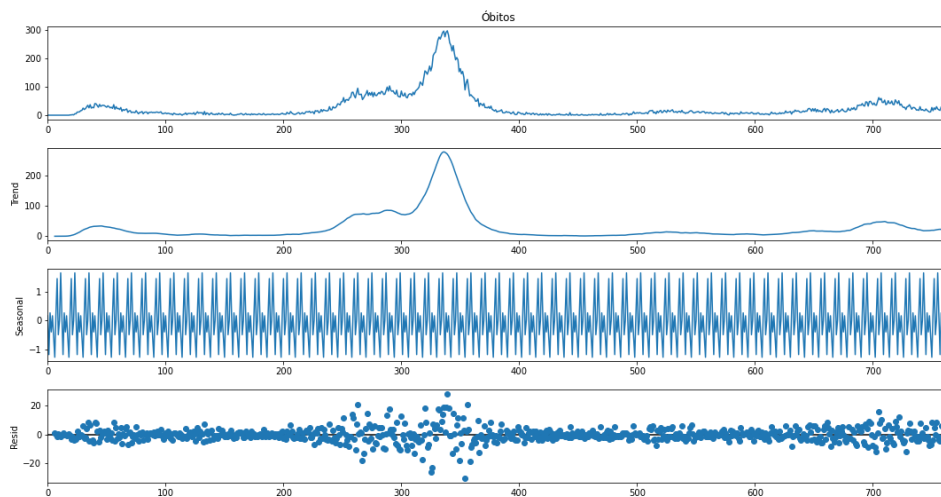


Figura 16 - Decomposição da série "Óbitos por dia"

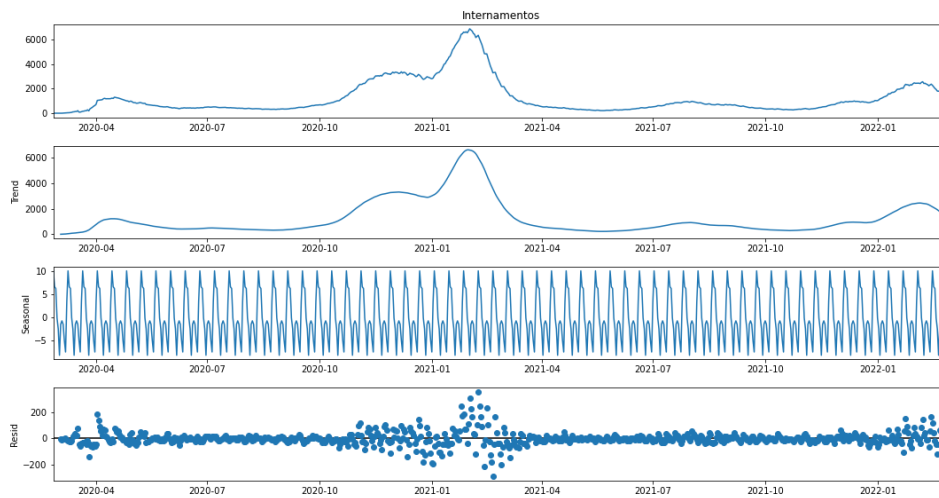


Figura 17 - Decomposição da série "Internamentos"

Através do método de decomposição, é possível observar picos no comportamento de qualquer uma das três séries temporais, que fazem com que as respetivas séries irregulares extraídas exibam oscilações não constantes ao longo do tempo. As componentes sazonais extraídas não parecem ser pronunciadas no sentido de identificar padrões sazonais.

Nas Figuras 18, 19 e 20 apresentam-se as funções ACF e PACF das séries originais. Relativamente ao comportamento da função de autocorrelção (ACF), esta apresenta um decaimento bastante lento nos três casos em estudo, o que revela o comportamento não estacionário das séries. No que toca à função de autocorrelção parcial, verifica-se uma queda mais brusca para zero nos três casos, mas com algumas autocorrelações significativas em *lags* de amplitude semanal (8 e 15 nos "novos casos", 5 a 10 nos "óbitos" e 7 e 11 nos "internamentos").

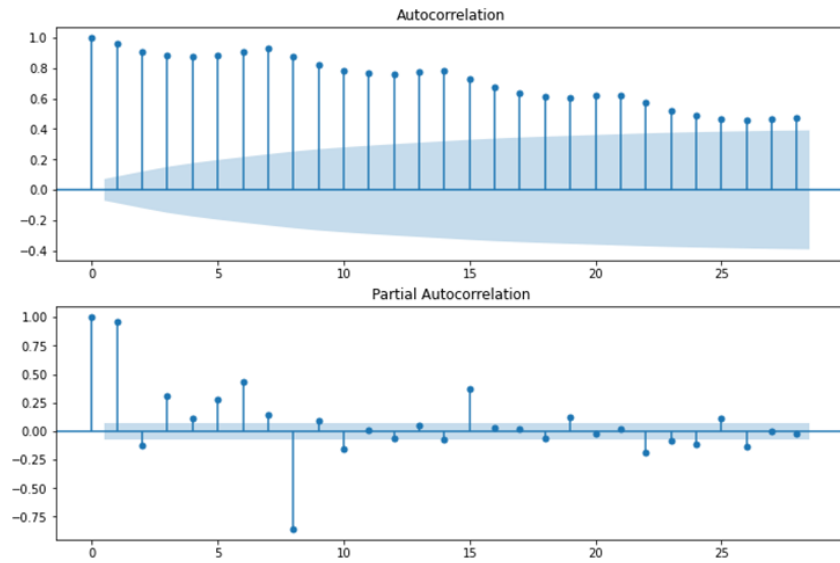


Figura 18 - ACF e PACF da série "Novos Casos"

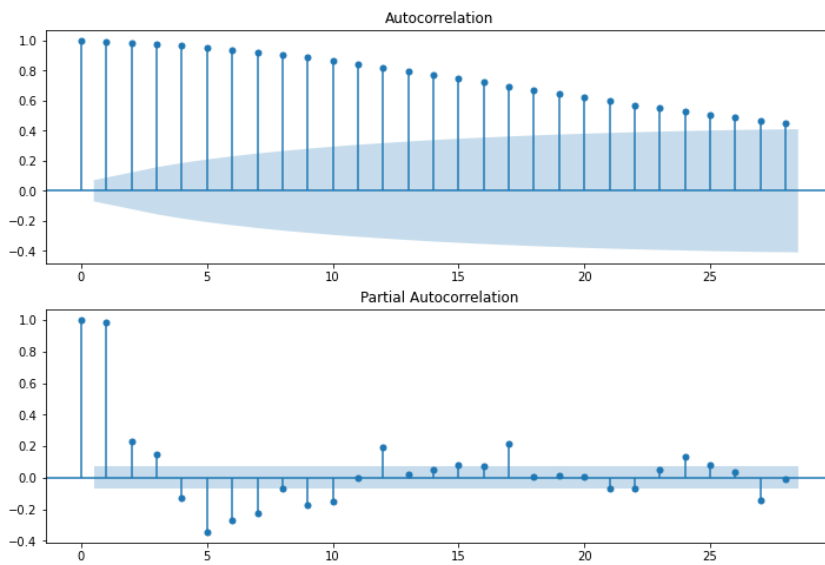


Figura 19 - ACF e PACF da série "Óbitos por dia"

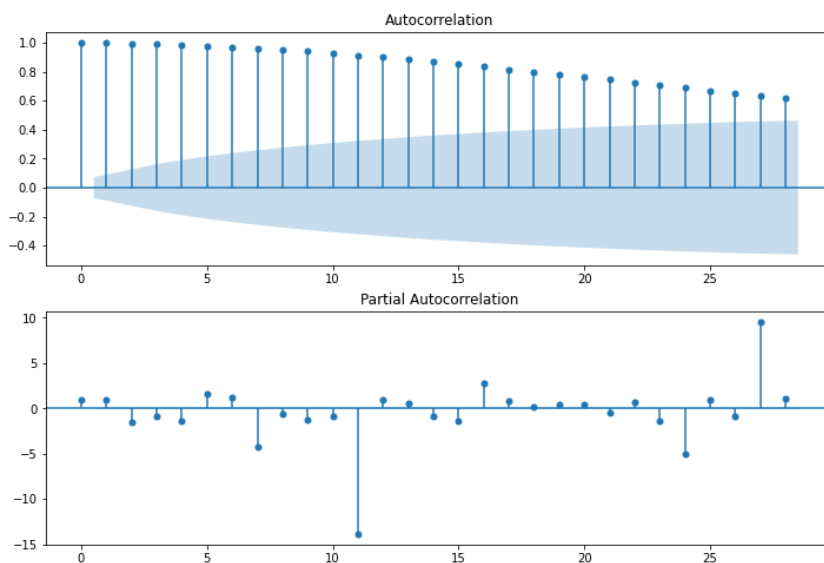


Figura 20 - ACF e PACF da série "Internamentos"

## 5.2 Estacionariedade da série

Com vista a melhorar a previsão das séries em estudo, foram previamente aplicados testes de estacionariedade. Como foi referido anteriormente, a tendência existente na série é um sinal de que a mesma é não estacionária em média, e, por isso, também não o será em variância e em covariância. Uma forma elementar de analisar a estacionariedades de uma série é representar graficamente os dados ao longo do tempo. No entanto, esta análise é subjetiva e, apesar de útil, deve ser confirmada através de testes estatísticos formais. Existem vários testes para realizar este estudo, baseados, em encontrar uma raiz unitária. Um dos testes mais utilizado é o teste de Dickey Fuller Aumentado (ADF). Através deste teste é possível encontrar uma raiz unitária, ou seja, a não estacionariedade, e no caso de não ser rejeitada, são fornecidas informações sobre o número de diferenciações necessárias para atingir a estacionariedade. Nas Figuras 21, 22 e 23 encontra-se os resultados do teste ADF para as três séries em estudo.



```

...: adf = adfuller(df_nc['casos'].dropna())
...: print('ADF Statistic: %f' % adf[0])
...: print('p-value: %f' % adf[1])
ADF Statistic: -3.046449
p-value: 0.030779

```

Figura 21 - Teste Dickey-Fuller da série original "Novos Casos"

```

...: adf = adfuller(df_no['Óbitos'].dropna())
...: print('ADF Statistic: %f' % adf[0])
...: print('p-value: %f' % adf[1])
ADF Statistic: -2.933507
p-value: 0.041588

```

Figura 22 - Teste Dickey-Fuller da série original "Óbitos por dia"

```

...: adf = adfuller(df_ni['Internamentos'].dropna())
...: print('ADF Statistic: %f' % adf[0])
...: print('p-value: %f' % adf[1])
ADF Statistic: -2.255323
p-value: 0.186782

```

Figura 23 - Teste Dickey-Fuller da série original "Internamentos"

Com as hipóteses definidas pelo teste  $H_0$ : existe uma raiz unitária e  $H_1$ : não existe raiz unitária, pode concluir-se que, ao nível de 1%, a hipótese nula da não estacionaridade não é rejeitada ( $p - value > 0.01$ ) em nenhum dos casos, o que nos leva a concluir pela presença de pelo menos uma raiz unitária em cada uma das séries.

Após a realização deste teste, foi realizada uma diferenciação simples, e ao aplicar de novo o teste de Dickey Fuller Aumentado, verificou-se que o  $p - value (0)$  é claramente inferior a 1%, o que sugere a presença de apenas uma raiz unitária. Nas

Figuras 24, 25 e 26<sup>4</sup> são apresentadas as séries em diferenças simples (não sazonais) e em diferenças sazonais.

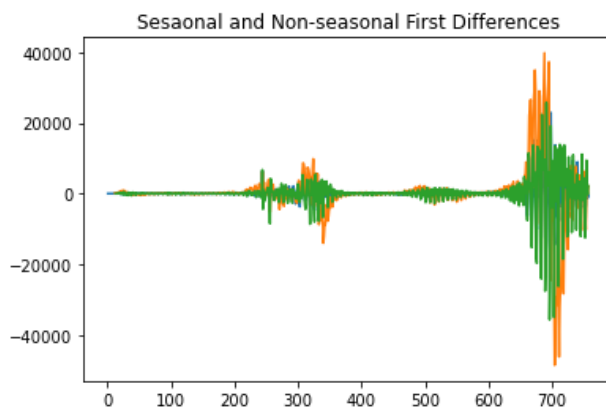


Figura 24 - Diferenciação sazonal e não sazonal da série "Novos Casos"

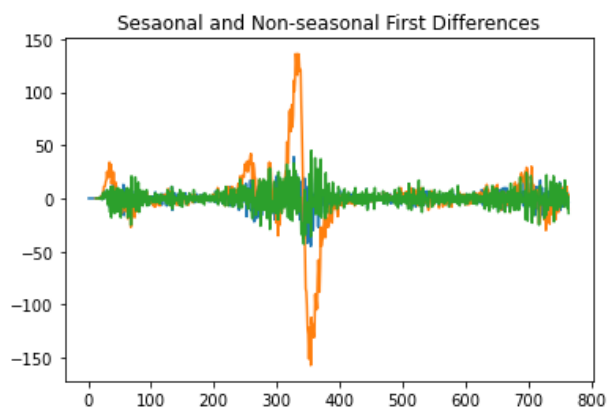


Figura 25 - Diferenciação sazonal e não sazonal da série "Óbitos por dia"

---

<sup>4</sup> A curva de cor azul corresponde ao cronograma da série após uma diferenciação simples. A cor laranja às diferenças sazonais, e a cor verde às diferenças sazonais e não sazonais da série original.

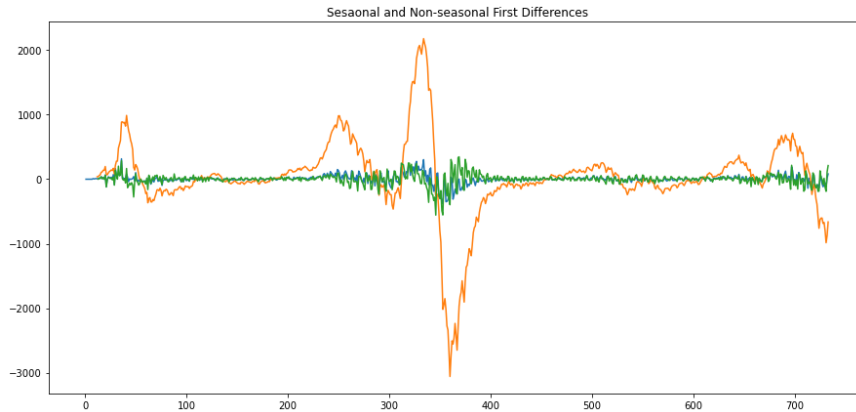


Figura 26 - Diferenciação sazonal e não sazonal da série "Internamentos"

De seguida, procedeu-se à escolha de um modelo SARIMA candidato a descrever as séries. Para tal, estimaram-se as ACF e PACF das séries estacionárias. Nas figuras seguintes encontram-se os respetivos correlogramas. Estes sugerem diferentes estruturas de modelos.

Para a série “Novos Casos”, os modelos sugeridos são SARIMA (0,1,1)(0,1,1)<sub>7</sub> e SARIMA (2,1,0)(1,1,0)<sub>7</sub>.

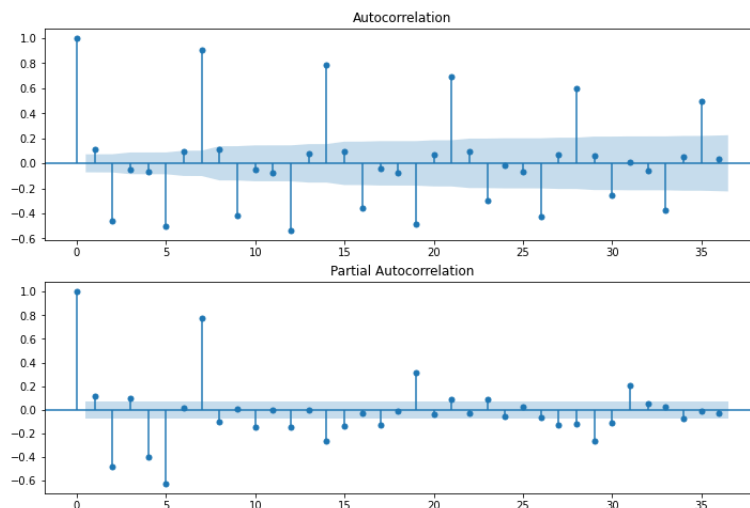


Figura 27 - ACF e PACF da série transformada "Novos Casos"

	<b>SARIMA (0, 1, 1)(0, 1, 1)7</b>	<b>SARIMA (2, 1, 0)(1, 1, 0)7</b>
<b>AIC</b>	12932.871	12933.153
<b>BIC</b>	12951.351	12951.634
<b>HQIC</b>	12939.992	12940.274

Figura 28 - Comparação de valores dos Critérios de Informação para os modelos da série transformada "Novos Casos"

À luz do princípio da parcimónia, foi escolhido com base no critério *Schwartz Bayesian* (BIC), o modelo SARIMA (0,1,1)(0,1,1)7 como o que conduzirá a um melhor ajustamento dos dados, uma vez que, é o que apresenta menor valor. Em relação aos outros Critérios de Informação, este também seria o modelo escolhido, pois apresentam sempre valores inferiores quando comparados com o modelo SARIMA (1,1,0)(1,1,0)7.

Em relação à série "Óbitos por dia", os modelos sugeridos são SARIMA (0,1,1)(0,1,1)7 e SARIMA (4,1,0)(2,1,0)7.

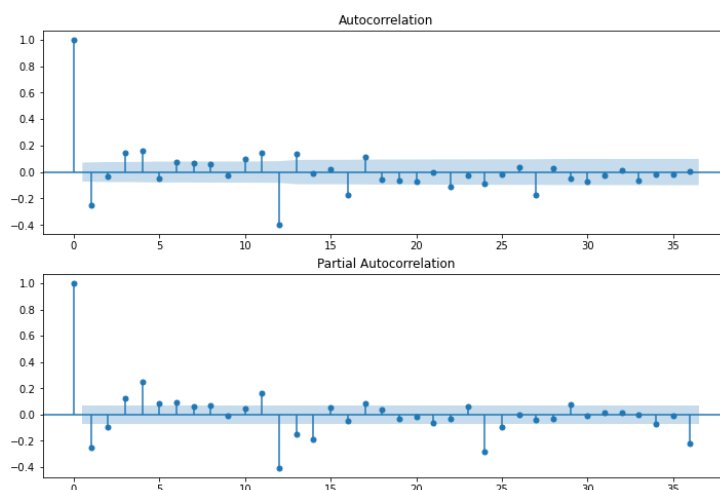


Figura 29 - ACF e PACF da série transformada "Óbitos por dia"

**SARIMA(0, 1, 1)(0, 1, 1)7    SARIMA(4, 1, 0)(2, 1, 0)7**

<b>AIC</b>	5164.036	5295.260
<b>BIC</b>	5177.924	5327.665
<b>HQIC</b>	5169.385	5307.741

Figura 30 - Comparação de valores dos Critérios de Informação para os modelos da série transformada "Óbitos por dia"

Mais uma vez se verifica que, com base no critério BIC, o modelo SARIMA  $(0,1,1)(0,1,1)_7$  é o que apresenta menor valor.

Por último, para a série "Internamentos", os modelos sugeridos são SARIMA  $(0,1,1)(0,1,1)_7$  e SARIMA  $(1,1,0)(1,1,0)_7$ .

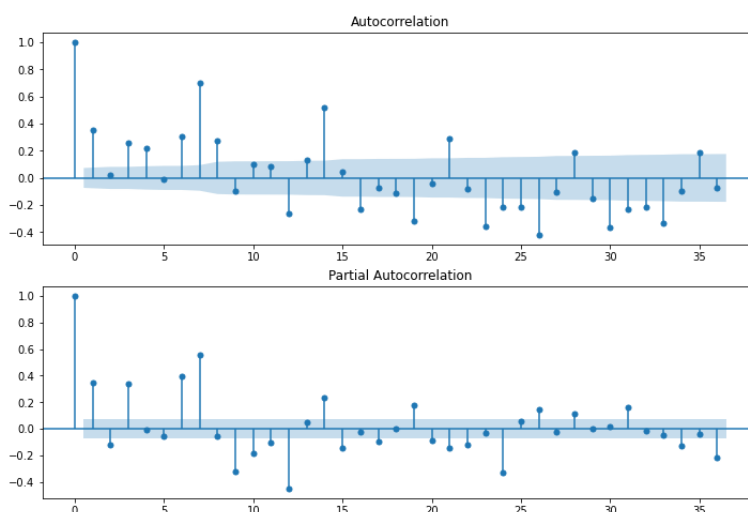


Figura 31 - ACF e PACF da série transformada "Internamentos"

**SARIMA(0, 1, 1)(0, 1, 1)<sub>7</sub>    SARIMA(1, 1, 0)(1, 1, 0)<sub>7</sub>**

<b>AIC</b>	7651.863	7565.274
<b>BIC</b>	7665.626	7579.037
<b>HQIC</b>	7657.174	7570.585

Figura 32 - Comparação de valores dos Critérios de Informação para os modelos da série transformada "Internamentos"

Neste caso verifica-se que o modelo escolhido é SARIMA (1,1,0)(1,1,0)<sub>7</sub>.

Depois de se eleger um modelo, e de este ser ajustado aos dados, foram calculados e comparados os erros MAPE, que são apresentados na tabela da seção seguinte, com o objetivo de avaliar a qualidade preditiva dos modelos.

### 5.3 Avaliação dos erros de previsão

Como referido anteriormente, nesta seção são analisados os resultados de previsão obtidos com os três métodos de previsão. Para a avaliação e comparação dos métodos foi usada a medida do erro percentual absoluto médio (MAPE).

	<b>MAPE</b>
<b>HOLT-WINTERS ADITIVO</b>	6.37
<b>SARIMA (0, 1, 1)(0, 1, 1)<sub>7</sub></b>	68.59
<b>SARIMA (2, 1, 0)(1, 1, 0)<sub>7</sub></b>	65.90

*Figura 33 - Erros MAPE da série "Novos Casos"*

	<b>MAPE</b>
<b>HOLT-WINTERS ADITIVO</b>	18.84
<b>SARIMA (0, 1, 1)(0, 1, 1)<sub>7</sub></b>	65.92
<b>SARIMA (4, 1, 0)(2, 1, 0)<sub>7</sub></b>	132.18

*Figura 34 - Erros MAPE da série "Óbitos por dia"*

	MAPE
<b>HOLT-WINTERS ADITIVO</b>	10.07
<b>SARIMA (0, 1, 1)(0, 1, 1)7</b>	10.62
<b>SARIMA (1, 1, 0)(1, 1, 0)7</b>	15.60

Figura 35 - Erros MAPE da série "Internamentos"

### 5.4 Utilização do “melhor” método na previsão futura do Covid-19

Como vimos anteriormente o modelo com melhor resultado para todas as séries foi o Holt-Winters Aditivo. Segue-se assim a previsão a 7 dias para cada uma das séries temporais.

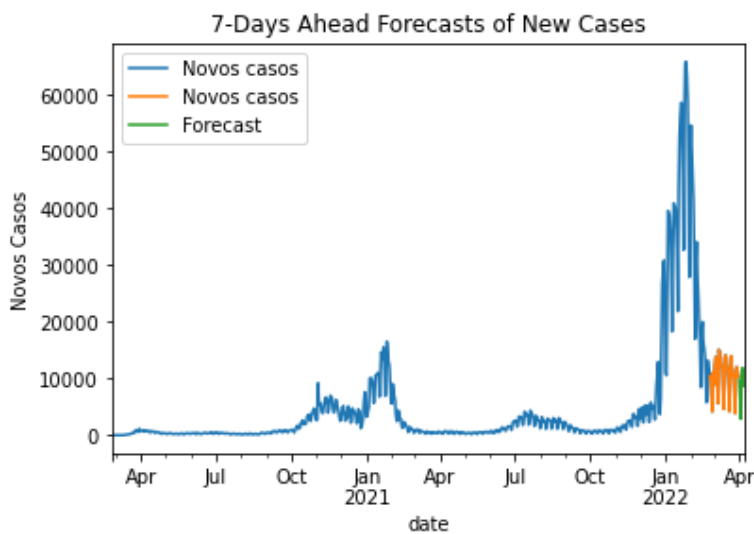


Figura 36 - Previsão a 7 dias para a série "Novos Casos"

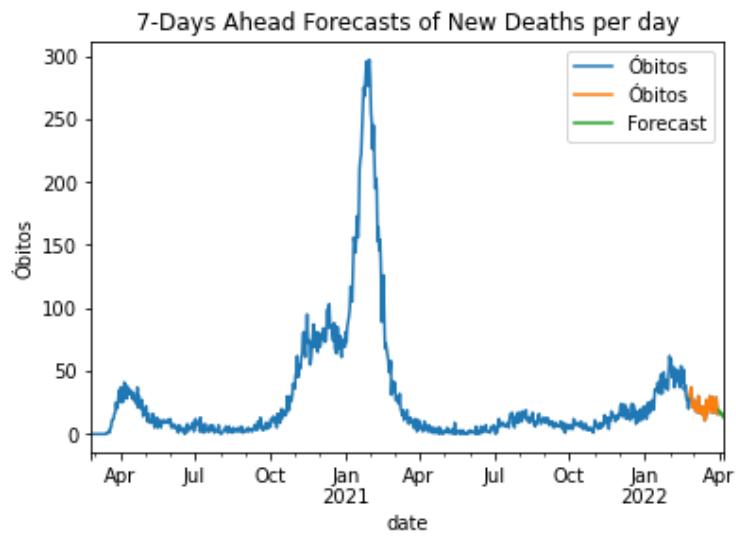


Figura 37 - Previsão a 7 dias para a série "Óbitos por dia"

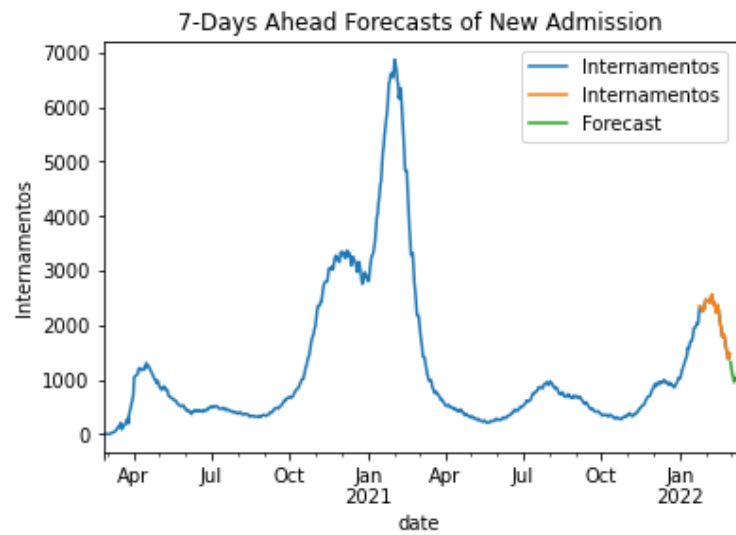


Figura 38 - Previsão a 7 dias para a série "Internamentos"



## 6. Conclusões

A crescente procura por níveis cada vez mais elevados de serviço e de qualidade pressiona as organizações a terem maior agilidade e direcionarem os seus estudos para prever, com precisão, os diversos picos da propagação do Covid-19, pois ao prevermos novas vagas desta doença, é possível delinear todos os seus sistemas e procedimentos com o objetivo principal de melhorar a velocidade e qualidade da resposta.

A literatura existente neste domínio, tornou-se bastante extensa, pois verificou-se a necessidade de existir uma abordagem que consiga prevenir a disseminação da doença Covid-19, ou outras que futuramente se adequam aos mesmos critérios. Tornou-se uma das maiores preocupações mundiais, uma vez que, a sua prevenção permite aos governos alterar as suas políticas e planear atempadamente as medidas preventivas necessárias.

Neste trabalho foram aplicados métodos de previsão determinísticos (Holt-Winters) e estocásticos (ARIMA/SARIMA) para prever o número de “Novos Casos”, “Óbitos por dia” e “Internamentos” associados à Covid-19. Por curiosidade, o método mais simples e de menor esforço computacional – o método de alisamento exponencial (Holt-Winters Aditivo) foi o que conseguiu obter melhores previsões para qualquer uma

das séries, de acordo com a medida de avaliação do erro de previsão MAPE (Erro Percentual Absoluto Médio).

Uma outra conclusão interessante que retiramos deste estudo foi a existência de uma relação entre o número diário de novos casos e de internamentos, uma vez que, é notável que quando uma das curvas aumenta/diminui, a outra também aumenta/diminui, ainda que não na mesma proporção. Significa isto que, o aumento do número de novos casos faz aumentar o número de internamentos e vice-versa, mas dependendo da variante da doença em curso, esse número pode ser maior ou menor. Em relação ao número de óbitos verifica-se o mesmo, isto é, quando a curva de internamentos aumenta/diminui, a curva dos óbitos também aumenta/diminui, mas neste caso é visível a mesma proporção. Pelo que, quando o número de internamentos aumenta, conseqüentemente aumenta o número de óbitos, e vice-versa.

De referir que, existe a importância de perceber a importância da vacina para conviver com este vírus e voltar à vida normal. A primeira dose da vacina começou no final de dezembro de 2020, e é perceptível nos cronogramas *supra* apresentados, que após o início do programa VACINAS, houve uma queda acentuada dos indicadores do Covid-19. No entanto, o maior pico ocorreu no início do ano de 2022, com uma nova variante, *Ómicron*. De realçar que depois de ultrapassado o pico, a situação começou a estabilizar.

Tendo em conta uma reflexão a posteriori dos temas desenvolvidos anteriormente, assim como os diversos aspetos metodológicos estudados ao longo deste trabalho, surge a necessidade de ponderar sobre algumas questões específicas como a relação entre os indicadores, e os melhores modelos para compreender tanto essa relação, como prever com precisão.

Como possíveis desenvolvimentos futuros, devemos considerar a realização do mesmo estudo recorrendo a dados com ajustamentos de calendário, de forma a aferir se, nesse caso, os resultados diferem. De facto, as previsões de indicadores de doenças, dependem das variantes em curso, características climáticas, discrepâncias nos procedimentos de teste e recolha de dados, e também diferentes medidas implementadas. Portanto, as previsões devem ser ajustadas de modo a captar as variações, caso contrário, podem ser enviesadas.

Relativamente às metodologias adotadas, é necessário avaliar de novo os erros de cada modelo, pois pode ocorrer alterações nesse sentido. Apesar dos resultados terem por base a análise das três séries temporais com comportamentos idênticos, a aplicação do modelo SARIMA, pode também ser útil para avaliar a sazonalidade e tendência das séries.

## Referências Bibliográficas

- Box, G., & Jenkins, G. (1970). *Time Series Analysis, Forecasting and Control*, San Francisco: Holden-Day.
- Box, G. P. E., Jenkins, G. M. e Reinsel, G. (1994). *Time Series Analysis: Forecasting and Control*, 3rd Edition, Prentice-Hall.
- Brockwell, P. and R. Davis (1991). *Time Series: Theory and Methods*, 2nd Edition, New York: Springer-Verlag.
- Brockwell, P.J. and Davis, R.A. (2002) *Introduction to Time Series and Forecasting*. Springer, New York.
- Brown, R. (1963). *Smoothing, Forecasting, and prediction of Discrete Time series*, Prentice-Hall, New Jersey.
- Caiado, J. (2016). *Métodos de Previsão em Gestão – Com Aplicações em Excel*. Lisboa: Edições Sílabo, 2ª ed.
- Chatfield, C. (2000). *Time-Series Forecasting*. Chapman and Hall/CRC.

- Davydenko, A. e Fildes, R. (2013). *Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts*, International Journal of Forecasting, Vol. 29, N° 3, p. 510–522.
- Gonçalves, J. (2010). *Gestão de Aprovisionamento – Stocks, Previsão, Compras*, Publindústria, Edições Técnicas.
- Holt, C. (1957). *Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages*, ONR Research Memorandum, Carnegie Institute of Technology.
- Hyndman R., Athanasopoulos G. (2012). *Forecasting: Principles and practice*.
- Makridakis, S., Wheelwright, S. C e Hyndman, R. J. (1998). *Forecasting Methods and Applications*, 3ª edição, New York: John Wiley & Sons.
- Murteira, B., Muller, D., & Turkman, K. (1993). *Análise de Sucessões Cronológicas*, McGraw-Hill.
- Murteira, B., Muller, D. e Turkman, K. (2000). *Análise de sucessões cronológicas*, McGrawHill.
- Papastefanopoulos, V., Linardatos, P., & Kotsiantis, S. (2020). *COVID-19: A comparison of time series methods to forecast percentage of active cases per population*. Applied Sciences (Switzerland), 10(11), in <https://doi.org/10.3390/app10113880>
- Wang, G., Wu, T., Wei, W., Jiang, J., An, S., Liang, B., Ye, L., & Liang, H. (2021). *Comparison of ARIMA, ES, GRNN and ARIMA-GRNN hybrid models to forecast the second wave of COVID-19 in India and the United States*. Epidemiology and Infection, 149, in <https://doi.org/10.1017/S0950268821002375>

Wei, W. W. S. (2007). *Time Series Analysis: Univariate and Multivariate Methods*, 2<sup>nd</sup> ed., Addison-Wesley Publishing Company, Redwood City, California.

Winters, P. (1960). *Forecasting sales by exponentially weighted moving averages*. *Management Science*, p. 324–342.