

Modelação e Previsão de Consumo anual de Carnes Brancas em Portugal (kg/hab)

André Catarino | 56788

Diana Salvador | 56370

Diogo Valente | 56662

Inês Caseiro | 48764

Sara Lopes | 52000

Yiqing Zhu | 52219

Lisbon School of Economics and Management - ISEG

Resumo: O presente trabalho analisa e modela a série não sazonal do consumo anual de carne *per capita* em Portugal. Realiza-se uma análise dos dados, com o objetivo de encontrar o melhor modelo entre os possíveis selecionados. Na identificação dos mesmos foram aplicados métodos de Alisamento Exponencial e modelos ARIMA. O modelo em destaque, entre os 8 modelos selecionados, é o modelo de Alisamento Exponencial com Tendência Exponencial. Os resultados da previsão para o futuro são, em parte, os expectáveis.

Palavras-Chave: Consumo de Carnes Brancas, Alisamento Exponencial, ARIMA, Erros de Previsão, Previsão.

1. Introdução

O presente trabalho pretende modelar o comportamento do consumo anual de carnes brancas *per capita* em Portugal, com o objetivo de obter uma previsão para a futura evolução do consumo, num horizonte temporal de 6 anos. Numa primeira abordagem, observaram-se e analisaram-se os dados desta série. Com os conhecimentos adquiridos na Unidade Curricular de Métodos de Previsão, decompôs-se a série em vários fatores, e com as características identificadas na série, foram aplicados modelos de Alisamento Exponencial e modelos ARIMA, de forma a entender qual o melhor entre estes. Por fim, foi realizada a previsão para o futuro, utilizando o modelo selecionado.

1.1. Definição do problema

O tema da série selecionada para este trabalho tem como intuito entender o comportamento e evolução do consumo de carnes brancas *per capita* em Portugal, desde 1981 até 2020. O estudo da evolução deste consumo torna-se pertinente devido a fatores como a alteração do nível de preços da carne, o aumento dos estudos sobre o consumo de carnes brancas em prol da saúde e a sensibilização ambiental.

1.2. Recolha e descrição dos dados

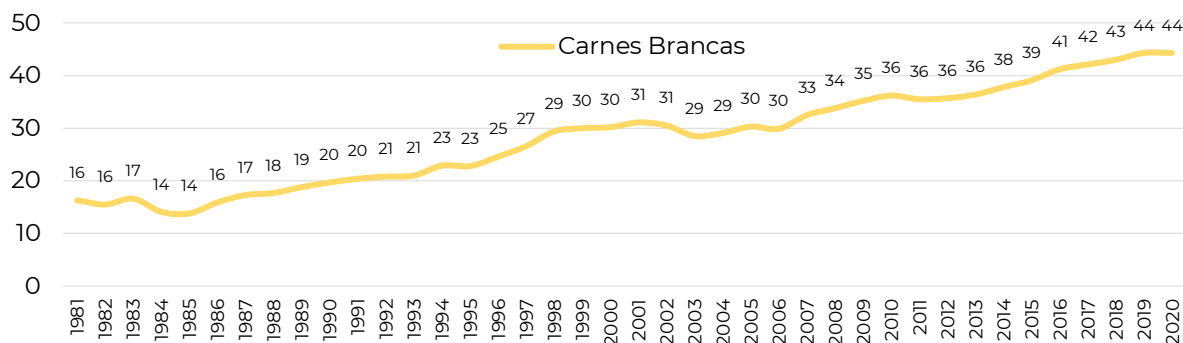


Figura 1- Consumo anual de Carnes Brancas em Portugal (kg/hab). Fonte: INE

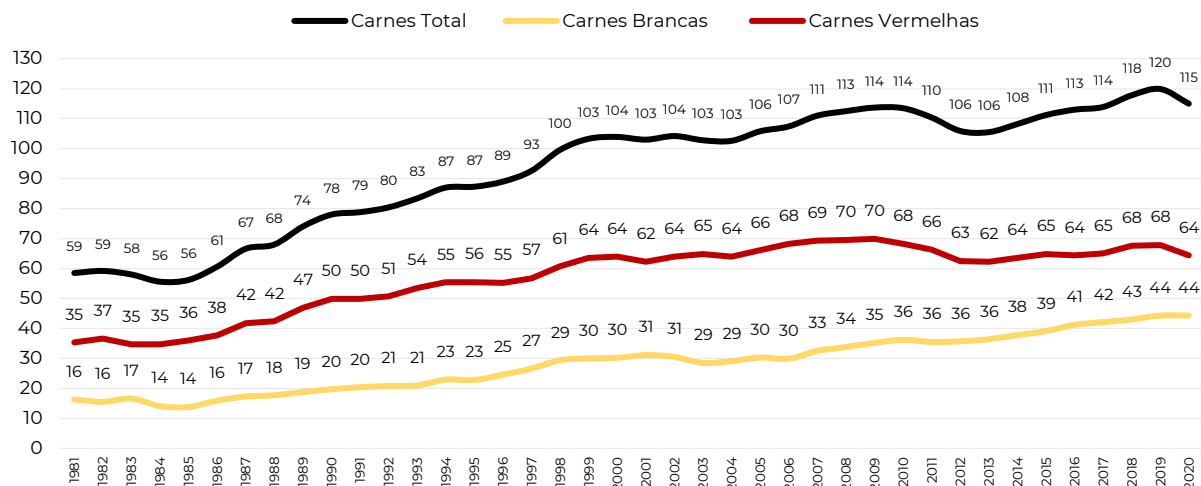


Figura 2- Consumo anual de Carnes em Portugal (kg/hab). Fonte: INE

A figura 1 contém a série representativa do consumo anual de carnes brancas em Portugal, em quilograma por habitante, entre 1981 e 2020. A figura 2 contém, além desta, a série representativa do consumo anual de carnes vermelhas e da totalidade de carnes em Portugal, em quilograma por habitante, entre 1981 e 2020.

Pela observação da figura 1, entende-se que o consumo de carnes brancas tem vindo a aumentar ao longo dos anos. Em 2020 atinge o patamar mais elevado de consumo (44 kg/hab), uma variação de 175% face ao ano de 1981, e de 47% face a 2000. Em média, o crescimento anual do consumo de carnes brancas é de 700 gramas. Este fenómeno pode ser explicado pela troca do consumo de carnes vermelhas por carnes brancas. Observando a figura 2, conclui-se que o consumo de carnes segue uma tendência crescente, atingindo o seu pico em 2019, no entanto, o maior contributo para este acontecimento provém do consumo de carnes brancas, uma vez que ocorre uma estabilização do consumo de carnes vermelhas desde 2008 até 2020. É de destacar a queda significativa do consumo de carnes de 2019 para 2020, provocada sobretudo pela diminuição do consumo de carnes vermelhas.

Os dados utilizados neste trabalho foram extraídos do site do INE (Instituto Nacional de Estatística).

1.3. Revisão da literatura empírica

O consumo de carnes vermelhas prevaleceu durante décadas em relação ao de carnes brancas. Segundo dados do INE, a partir de 2008, pela primeira vez, o consumo de carnes vermelhas entra em declínio. Esta diminuição do consumo de carnes vermelhas ao longo da segunda década do século XXI está diretamente relacionada com fatores de consciencialização nas matérias de saúde alimentar e ambiental. Relativamente à saúde alimentar, a OMS (Organização Mundial de Saúde) confirmou a existência de menores benefícios no consumo de carnes vermelhas em detrimento de brancas, nomeadamente devido à maior possibilidade de desenvolvimento de patologias.

É de salientar o impacto que o vegetarianismo tem vindo a ter em Portugal. Além do aumento da percentagem de vegetarianos, é notável o aumento da percentagem de pessoas que adota medidas para a redução do consumo de carnes. Destes, 41% afirmam pretender reduzir o consumo de carnes vermelhas, e apenas uma minoria pretende limitar o consumo de carnes brancas.

Por fim, o fator geracional também tem relevância para a explicação dos dados, uma vez que a geração atual tem maior consciência acerca dos problemas de saúde alimentar e ambiental, procurando adotar um estilo de vida mais sustentável.

2. Decomposição das forças componentes

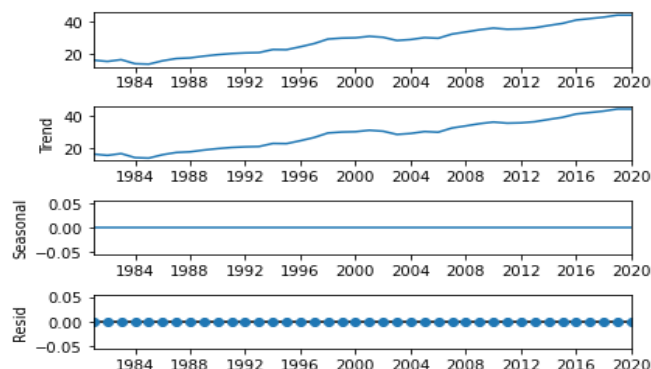


Figura 3- Decomposição da série através do Modelo Aditivo

As séries temporais exibem muitas vezes uma variedade de padrões, e por esse motivo, torna-se útil separar uma série temporal em diversas componentes. Para tal, são aplicados procedimentos de decomposição que compreendem três componentes: uma componente de tendência cíclica, uma componente sazonal, e uma componente residual. Os resultados da decomposição desta série estão evidenciados na figura 3.

Um dos principais objetivos para a decomposição consiste na estimação dos efeitos sazonais, de forma a desenvolver e apresentar valores corrigidos de sazonalidade. Um valor corrigido de sazonalidade remove o efeito sazonal de um valor, para que as tendências possam ser observadas mais claramente.

Existem dois modelos possíveis de decomposição: Modelo Aditivo e Modelo Multiplicativo. O Modelo Multiplicativo utiliza-se quando a componente sazonal aumenta com a tendência, ou seja, há interdependência entre elas. Para esta série optou-se pela aplicação do Modelo Aditivo, devido à ausência de sazonalidade, comprovável pela figura 3.

O modelo aditivo sugere que os dados da série são resultado de uma função que soma as três componentes da série, tendência cíclica (TC_t), sazonalidade (S_t), e resíduo (E_t) da seguinte forma:

$$Y_t = TC_t + S_t + E_t$$

Relativamente à tendência, esta é claramente crescente. Isto é, ao longo prazo, os dados da série seguem uma direção crescente. No que respeita à sazonalidade, este elemento toma sempre um valor nulo ao longo dos anos, tal como era expectável. Não se observam outliers, e os resíduos da série tomam também o valor zero, podendo afirmar-se que esta série é totalmente explicada pela tendência cíclica.

3. Métodos de Alisamento Exponencial

3.1. Revisão da literatura

Os métodos de alisamento exponencial foram propostos por investigadores operacionais no final dos anos 50. Os principais nomes apontados para o desenvolvimento desta matéria foram Holt (1957), Brown (1959), Magee (1958) e Winters (1960). Desde então, o conceito de alisamento exponencial tem vindo a crescer, tornando-se mais popular, sobretudo na área da previsão. Algumas das vantagens são a facilidade, o baixo custo de aplicação e o facto de serem os únicos métodos rápidos o suficiente para implementar previsões com inúmeros itens. Além destes métodos de previsão, podem ainda ser aplicados métodos clássicos de decomposição ou métodos mais sofisticados, como modelos autorregressivos ou de médias móveis. Na implementação de métodos de alisamento exponencial, há três fatores a ter em conta, sendo eles, a inicialização, a otimização e os intervalos de previsão. Estes fatores variam consoante o método de alisamento exponencial aplicado (Makridakis, Wheelwright, & Hyndman, 1998).

Para aplicar estes métodos é necessário, em primeiro lugar, escolher uma série temporal (conjunto de dados) para analisar. De seguida, é necessário separar estes dados na amostra de treino e na amostra de teste. A amostra de treino permite estimar os parâmetros do modelo a usar e a amostra de teste é necessária para avaliar a precisão da previsão. Normalmente, a amostra de treino consiste em 80% dos dados, enquanto que a amostra de teste são os restantes 20%. Esta regra é subjetiva, mas a amostra de teste deve conter dados com, pelo menos, a dimensão do horizonte de previsão pretendido.

De seguida, vamos descrever vários métodos de alisamento exponencial, sendo que se caracterizam por atribuir às observações, pesos exponencialmente decrescentes com a antiguidade.

Alisamento Exponencial Simples:

Para obter a previsão do período seguinte, o alisamento exponencial simples considera a previsão para o período mais recente ajustada usando o erro de previsão registado. Este ajustamento faz-se através da multiplicação do erro pelo parâmetro de alisamento, $0 < \alpha < 1$. Quando este parâmetro se encontra próximo de 1, o ajustamento é total, e o alisamento é nulo. No entanto, se estiver próximo de zero, o ajustamento é mínimo e o alisamento é total (Makridakis, Wheelwright, & Hyndman, 1998).

Previsão para o período $t+1$:

$$\widehat{Y}_{t+1} = \hat{Y}_t + \alpha(Y_t - \hat{Y}_t) = \alpha Y_t + (1 - \alpha)\hat{Y}_t = \sum_{j=0}^{T-1} \alpha(1 - \alpha)^j Y_{t-j}$$

Equação de previsão a k períodos:

$$\hat{Y}_{t+k} = \hat{Y}_t, \quad k > 0 \quad (k \text{ é o horizonte temporal})$$

Assim, pode concluir-se que este método é apropriado para fazer previsão em séries sem tendência ou sazonalidade.

Alisamento Exponencial Duplo:

Tal como o nome sugere, este método requer duas equações para realizar a previsão. Uma delas reflete o alisamento exponencial simples, S_t , e a outra o duplo, D_t :

$$\begin{aligned}S_t &= \alpha Y_t + (1 - \alpha)S_{t-1} \\D_t &= \alpha S_t + (1 - \alpha)D_{t-1}\end{aligned}$$

O parâmetro de alisamento α mantém o significado explicado no método anterior.

Equação de previsão a k períodos:

$$\begin{aligned}\hat{Y}_{t+k} &= a_t + kb_t, \text{ com:} \\a_t &= 2S_t - D_t \text{ (nível)} \\b_t &= (S_t - D_t)\alpha/(1 - \alpha) \text{ (declive)}\end{aligned}$$

Conclui-se que este método pode ser aplicado a séries com tendência linear, mas sem sazonalidade.

Alisamento Exponencial de Holt com Tendência Linear:

Segundo Makridakis, Wheelwright, & Hyndman (1998), Holt, em 1957, estendeu o método de alisamento exponencial simples para um método de alisamento exponencial linear, de forma a permitir prever séries temporais com tendência. Este método recorre a dois parâmetros de alisamento, $0 < \alpha < 1$ e $0 < \beta < 1$.

Equação de previsão a k períodos:

$$\begin{aligned}\hat{Y}_{t+k} &= a_t + kb_t, \text{ com:} \\a_t &= \alpha Y_t + (1 - \alpha)(a_{t-1} + b_{t-1}) \text{ (nível)} \\b_t &= \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1} \text{ (declive)}\end{aligned}$$

Conclui-se que este método, tal como o anterior, pode ser aplicado a séries com tendência linear, mas sem sazonalidade, no entanto a sua utilização é mais simples.

Comparando o Alisamento Exponencial de Holt com Tendência Linear com o Alisamento Exponencial Simples, o primeiro revela-se quase sempre melhor em termos da Raiz do Erro Quadrático Médio (REQM), pois utiliza um parâmetro de alisamento adicional.

Outros Métodos:

Além destes métodos, existe ainda um método de previsão com tendência aditiva/multiplicativa amortecida e outro com tendência exponencial, que podem fornecer melhores resultados.

O Alisamento Exponencial com Tendência Aditiva Amortecida tem como equação de previsão para k períodos, $\hat{Y}_{t+k} = a_t + b_t \sum_{i=1}^k \phi^i$, $0 < \phi < 1$, onde a_t representa uma estimativa do nível, e b_t representa uma estimativa do declive (a_t e b_t representam as fórmulas usadas no Alisamento Exponencial de Holt com Tendência Linear, sendo que quando $\phi = 1$, este método consiste no de Holt).

O Alisamento Exponencial com Tendência Multiplicativa Amortecida tem como equação de previsão para k períodos, $\hat{Y}_{t+k} = a_t \times b_t^{\sum_{i=1}^k \phi^i}$, $0 < \phi < 1$ (a_t e b_t representam fórmulas diferentes das usadas no Alisamento Exponencial de Holt com Tendência Linear, sendo que quando $\phi = 1$, este método consiste no Alisamento Exponencial com Tendência Exponencial).

O Alisamento Exponencial com Tendência Exponencial tem como equação de previsão para k períodos, $\hat{Y}_{t+k} = a_t \times b_t^k$, onde a_t representa uma estimativa do nível, e b_t representa uma estimativa do crescimento relativo da série (a_t e b_t representam fórmulas diferentes das usadas no Alisamento Exponencial de Holt com Tendência Linear).

(Hyndman & Athanasopoulos, 2014)

Os parâmetros de alisamento, para todos os modelos, são encontrados, geralmente, através da minimização do Erro Quadrático Médio (EQM), sendo que também se pode minimizar o Erro Absoluto Médio (EAM) ou o Erro Percentual Absoluto Médio (EPAM) (Makridakis, Wheelwright, & Hyndman, 1998).

3.2. Identificação do método de Alisamento Exponencial

A série temporal não sazonal escolhida consiste num conjunto de dados anuais, de 1981 a 2020, sobre o consumo, em kg, de carnes brancas por habitante em Portugal. Existem 40 observações, sendo que as 32 primeiras constituem a amostra de treino e as 8 últimas a amostra de teste.

Quanto aos métodos de alisamento exponencial utilizados, excluiu-se o Alisamento Exponencial Simples por não ser o mais adequado para a previsão de séries com tendência linear e ausência de sazonalidade. Aplicou-se o método de Alisamento Exponencial de Holt com Tendência Linear, o método de Alisamento Exponencial com Tendência Aditiva Amortecida, e o método de Alisamento Exponencial com Tendência Exponencial.

No *software Python*, obtiveram-se os gráficos que comparam os dados da amostra de teste com os dados da previsão, para cada um dos três métodos utilizados.

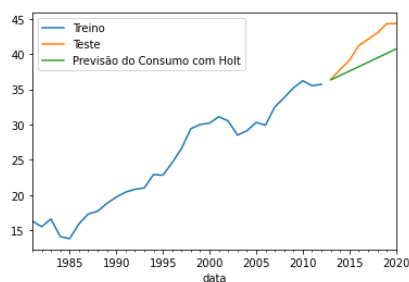


Figura 4- Método de Alisamento Exponencial de Holt com Tendência Linear

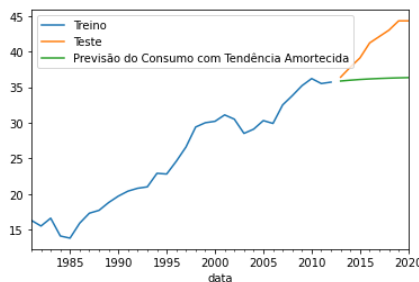


Figura 5- Método de Alisamento Exponencial com Tendência Aditiva Amortecida

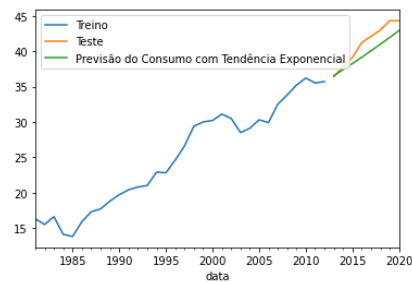


Figura 6- Método de Alisamento Exponencial com Tendência Exponencial

Através da observação gráfica, conclui-se que o melhor método a utilizar é o Alisamento Exponencial com Tendência Exponencial.

4. Modelos ARMA e ARIMA

4.1. Revisão da literatura

Na análise de séries temporais, utilizam-se vários tipos de modelos de previsão com diferentes parâmetros, de modo a entender melhor os dados e prever observações futuras de uma dada série. Estes modelos utilizaram-se na previsão de séries temporais e surgiram na década de 70, impulsionados pelos estatísticos britânicos George Box e Gwilym Jenkins. Os dois matemáticos popularizaram os modelos de previsão através do seu livro lançado em 1970, onde conceberam um método iterativo, para escolher e estimar os modelos, designado Box-Jenkins.

Um caso particular desses modelos, é o modelo autorregressivo integrado de médias móveis ARIMA (*autoregressive integrated moving average*), que corresponde a uma generalização de um modelo autorregressivo de médias móveis, ARMA (*autoregressive moving average*).

Os modelos ARIMA podem ser estudados conforme as suas diferentes componentes. Existe a componente autorregressiva (AR), a componente de médias móveis (MA), e ainda a componente integrada (I). Estes modelos podem também ser utilizados na previsão de séries temporais com sazonalidade, passando a ser designados por modelos SARIMA (*seasonal autoregressive integrated moving average*).

Modelos ARMA(p,q):

Os modelos ARMA podem ser processos estacionários e invertíveis mediante certas condições, são utilizados para séries temporais não sazonais, e caracterizam as próprias séries.

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

Este modelo é acompanhado dos parâmetros ϕ_p , que representa o parâmetro autorregressivo de ordem p, e θ_q , que representa o parâmetro de médias móveis de ordem q.

Os modelos ARMA apresentam a forte limitação de poderem ser praticáveis apenas em condições de estacionariedade dos dados.

Modelos ARIMA(p,d,q):

Além das componentes autorregressivas (AR) e de médias móveis (MA) referente aos modelos ARMA, existe ainda uma terceira componente, a componente integrada (I), que indica que o valor dos dados foi substituído pela diferença entre os valores atuais e os valores prévios (processo que pode ser repetido mais do que uma vez).

Para que os modelos ARMA estejam aptos a dados não estacionários, deve ser adicionada ao modelo a componente integrada (I), dando assim origem aos modelos ARIMA.

$$(1 - \phi_1 L - \dots - \phi_p L^p)(1 - L)^d Y_t = (1 - \theta_1 L - \dots - \theta_q L^q) \varepsilon_t \text{ (L é o operador de atraso)}$$

O operador $(1 - L)^d$ refere-se ao operador de diferenciação de ordem d, $d \geq 1$.

4.2. Identificação dos modelos ARMA ou ARIMA

Correlograma:

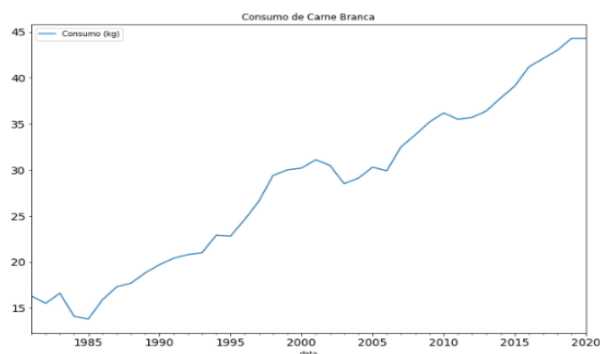


Figura 7- Representação gráfica da série

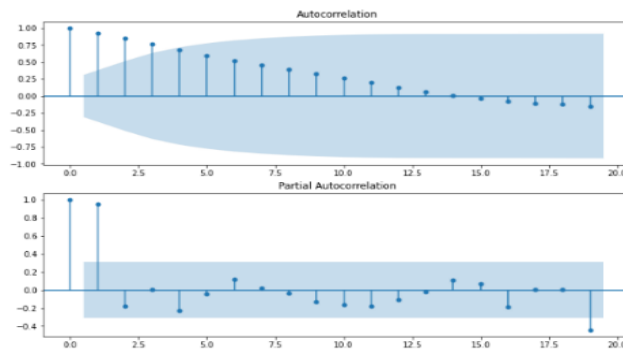


Figura 8- Correlograma da série

De acordo com o gráfico da série temporal apresentado na figura 7, depreende-se que esta é uma série não estacionária em média e em variância.

Através do correlograma da mesma série, presente na figura 8, conclui-se que a FAC apresenta um decaimento exponencial amortecido e a FACP demonstra um decaimento brusco do lag 1 para o lag 2. Deste modo, e apenas com base nos dados originais, entende-se que o modelo mais adequado será um AR(1) ou um passeio aleatório, devido à sua característica de não estacionariedade.

Teste Dickey-Fuller Aumentado:

Para determinar qual o modelo a ser estimado, deve averiguar-se a estacionariedade da série. Existem vários testes de raízes unitárias, porém aquele que prevalece é o teste Dickey-Fuller. Este teste consiste em realizar uma inferência estatística para detetar a existência da raiz unitária.

$$H_0: \text{Existe uma raiz unitária} \quad \text{vs} \quad H_1: \text{Não existe uma raiz unitária}$$

Caso não se rejeite a hipótese nula, então a série possui raiz unitária. A existência de raiz unitária sugere que a série em estudo é não estacionária, pelo que devem aplicar-se transformações ao nível da série original.

Transformações:

Quando se trata de uma série não estacionária em variância, utilizam-se as transformações logarítmicas de Box-Cox de modo a estabilizar a mesma. Este processo será apenas possível realizar antes das diferenciações simples, uma vez que, matematicamente, não se podem aplicar logaritmos a valores negativos. Por este motivo, deve sempre realizar-se primeiro as transformações a nível da variância, e só depois a nível da média.

$$X_t = T(Y_t) = \begin{cases} Y_t^\lambda, \lambda \neq 0 \\ \log(Y_t), \lambda = 0 \end{cases}$$

Caso a série seja não estacionária em média, são aplicadas diferenciações simples.

$$\Delta^d Y_t = (1 - L)^d Y_t$$

Resultados (Teste Dickey-Fuller Aumentado e Transformações):

Resultados do teste Dickey-Fuller:	Original	1ª diferença	2ª diferença	Logarítmo
Test Statistic	-0.465857	-3.425556	-4.196489	-1.690388
p-value	0.898509	0.010118	0.000668	0.436072
#Lags Used	6	5	6	7
Nº of Obs Used	25	25	23	24
Critical Value (1%)	-3.723863	-3.723863	-3.752928	-3.737709
Critical Value (5%)	-2.986489	-2.986489	-2.998500	-2.992216
Critical Value (10%)	-2.632800	-2.632800	-2.638967	-2.635747

Tabela 1- Resultados do teste Dickey-Fuller Aumentado

Resultado do teste para os dados originais:

Aplicando o teste de Dickey-Fuller Aumentado para testar a estacionariedade da série, obtém-se um valor-p de 0,89. Desta forma, não se rejeita a hipótese nula a um nível de significância de 5%, pelo que a série original será não estacionária.

Resultado do teste para os dados transformados:

Após aplicar várias transformações nos dados originais, com o objetivo de tornar a série estacionária para conseguir ajustar os modelos de previsão, o processo que apresenta um valor-p mais favorável a nível de estacionariedade e de invertibilidade das raízes consiste na aplicação de uma diferenciação simples aos dados originais a um nível de significância de 5%. No entanto, para um maior grau de precisão, deve aplicar-se a segunda diferenciação de modo a obter um valor-p inferior a 1 %, tornando assim os dados satisfatórios para a aplicação do modelo ARIMA.

Seleção dos modelos:

	ARIMA (0,2,1)	ARIMA (1,2,1)	ARIMA (0,2,2)	ARIMA (1,2,2)	ARIMA (2,2,1)
ϕ (valor-p)		$\hat{\phi}_1 = 0.1577$ (0.493)		$\hat{\phi}_1 = -0.7886$ (0.004)	$\hat{\phi}_1 = -0.5682$ (0.146) $\hat{\phi}_2 = -0.4280$ (0.052)
θ (valor-p)	$\hat{\theta}_1 = -0.9994$ (0.951)	$\hat{\theta}_1 = -0.9991$ (0.941)	$\hat{\theta}_1 = -0.8416$ (0.973) $\hat{\theta}_2 = -0.1579$ (0.968)	$\hat{\theta}_1 = -0.0004$ (1.000) $\hat{\theta}_2 = -0.9996$ (1.000)	$\hat{\theta}_1 = -0.0066$ (0.989)
Estacionariedade	Sim	Sim	Sim	Sim	Sim
Invertibilidade	Sim	Sim	Sim	Sim	Sim
Resíduos são ruído branco?	Sim	Sim	Sim	Sim	Sim
Valor-p da estatística Q	0.49	0.88	0.89	0.88	0.91
AIC	103.689	104.974	104.989	105.565	109.390
BIC	106.492	109.177	109.192	111.170	114.994

Tabela 2- Características dos modelos selecionados

De acordo com a tabela acima apresentada, podem extrair-se algumas conclusões que permitem compreender quais os melhores modelos de previsão da série em estudo. Essas conclusões são recolhidas através

do teste de significância individual, dos correlogramas dos resíduos de cada modelo (em anexo) onde se verifica se há indícios de estrutura da série dos resíduos, e do teste de hipóteses Ljung-Box, que permite estudar a correlação dos resíduos do modelo, sendo que o objetivo é não rejeitar a hipótese nula (os resíduos têm o comportamento de um ruído branco). Por fim, são apresentados dois critérios de seleção de modelos (AIC e BIC), e quanto menor for o seu valor, melhor será o seu ajustamento do modelo aos dados.

Conclui-se que, para todos os modelos, os parâmetros são estatisticamente não significativos a um nível de confiança de 5%, excepto o parâmetro de AR para o modelo ARIMA(1,2,2), que é estatisticamente significativo. Em relação aos resíduos, de acordo com os correlogramas, estes aproximam-se de um ruído branco para qualquer modelo, o que corrobora as conclusões retiradas através do teste do Ljung-Box, pois como a estatística Q apresenta valores-p superiores a qualquer nível de confiança, procede-se à não rejeição da hipótese nula, confirmando-se assim a ausência de correlação entre os resíduos.

No *software Python*, obtiveram-se os gráficos que comparam os dados da amostra de teste com os dados da previsão, para cada um dos cinco métodos utilizados.

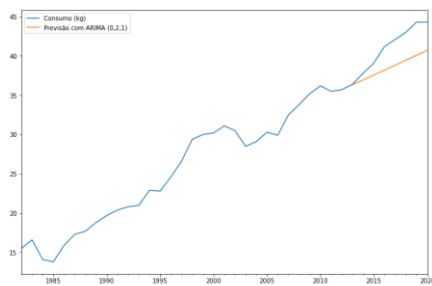


Figura 9-Método ARIMA (0,2,1)

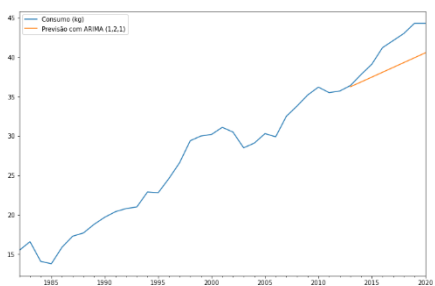


Figura 10-Método ARIMA (1,2,1)

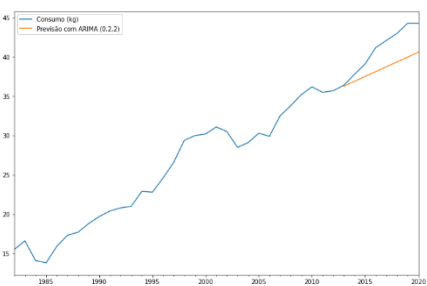


Figura 11-Método ARIMA (0,2,2)

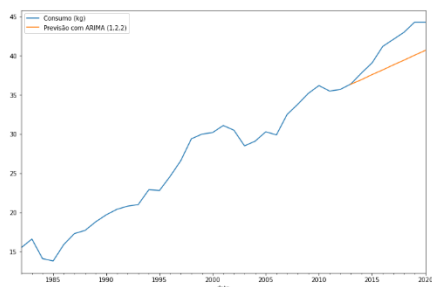


Figura 12-Método ARIMA (1,2,2)

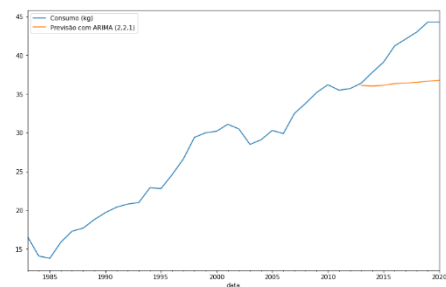


Figura 13-Método ARIMA (2,2,1)

De acordo com o critério de seleção apresentado (AIC), o modelo ARIMA(0,2,1) apresenta um melhor ajustamento, e por isso, será o principal candidato para realizar a previsão.

5. Avaliação dos erros de previsão

5.1. Revisão da literatura

Torna-se agora importante perceber a adequabilidade de um método de previsão a um determinado conjunto de dados. Na maioria das situações, a precisão é o critério chave para seleccionar o método de previsão, referindo-se à qualidade do ajustamento, ou seja, à boa capacidade do modelo para reproduzir os dados (Makridakis, Wheelwright, & Hyndman, 1998).

De forma a escolher o melhor método para prever futuras observações da série temporal, é necessário avaliar os erros de previsão e escolher o método que os minimize. Para este efeito, pode utilizar-se o Erro Quadrático Médio (EQM) ou a sua raiz (REQM), o Erro Absoluto Médio (EAM), o Erro Percentual Absoluto Médio (EPAM), entre outros.

Supondo que a previsão ocorre desde $T + 1$ a $T + k$, sendo Y_t o valor atual no período t , e \hat{Y}_t o valor da previsão.

Erro Quadrático Médio (EQM):

$$EQM = \left(\frac{1}{k}\right) \sum_{t=T+1}^{T+k} (Y_t - \hat{Y}_t)^2$$

Raiz do Erro Quadrático Médio (REQM):

$$REQM = \sqrt{\left(\frac{1}{k}\right) \sum_{t=T+1}^{T+k} (Y_t - \hat{Y}_t)^2}$$

Estas medidas de precisão não permitem comparar séries temporais diferentes nem períodos diferentes da mesma série, pois o EQM não tem unidades válidas, e a unidade de medida do REQM depende das unidades dos dados. Por outro lado, estas medidas são sensíveis a valores anormais (outliers).

Erro Absoluto Médio (EAM):

$$EAM = \left(\frac{1}{k}\right) \sum_{t=T+1}^{T+k} |Y_t - \hat{Y}_t|$$

Tal como as anteriores, esta medida de precisão não permite comparar séries temporais diferentes nem períodos diferentes da mesma série, pois a sua unidade de medida depende das unidades dos dados. No entanto, esta é mais facilmente interpretável face ao EQM.

Erro Percentual Absoluto Médio (EPAM):

$$EPAM = \left(\frac{1}{k}\right) \sum_{t=T+1}^{T+k} |(Y_t - \hat{Y}_t)/Y_t| * 100$$

Esta medida de precisão é relativa, não dependendo da escala dos dados, e permitindo comparar séries temporais diferentes e períodos diferentes da mesma série. Para esta medida, surgem dificuldades quando a série temporal apresenta valores próximos de zero.

(Makridakis, Wheelwright, & Hyndman, 1998).

5.2. Resultados da avaliação dos erros de previsão

Quanto aos métodos de Alisamento Exponencial, no *software Python*, aplicou-se o método de Alisamento Exponencial de Holt com Tendência Linear, registando-se um valor de 2,87 para a Raiz do Erro Quadrático Médio

(REQM), um valor de 2,51 para o Erro Absoluto Médio (EAM) e um valor de 5,91% para o Erro Percentual Absoluto Médio (EPAM). De seguida aplicou-se o método de Alisamento Exponencial com Tendência Aditiva Amortecida registando-se um valor de 5,54 para a Raiz do Erro Quadrático Médio (REQM), um valor de 4,88 para o Erro Absoluto Médio (EAM) e um valor de 11,51% para o Erro Percentual Absoluto Médio (EPAM). Por fim, aplicou-se o método de Alisamento Exponencial com Tendência Exponencial registando-se um valor de 1,62 para a Raiz do Erro Quadrático Médio (REQM), um valor de 1,41 para o Erro Absoluto Médio (EAM) e um valor de 3,33% para o Erro Percentual Absoluto Médio (EPAM).

Através das medidas de precisão da previsão, conclui-se que o melhor método a utilizar é o Alisamento Exponencial com Tendência Exponencial.

Quanto aos modelos ARIMA, no *software Python*, aplicou-se o modelo ARIMA (0,2,1), registando-se um valor de 2,87 para a Raiz do Erro Quadrático Médio (REQM), um valor de 2,51 para o Erro Absoluto Médio (EAM) e um valor de 5,91% para o Erro Percentual Absoluto Médio (EPAM). De seguida aplicou-se o modelo ARIMA (1,2,1) registando-se um valor de 3 para a Raiz do Erro Quadrático Médio (REQM), um valor de 2,63 para o Erro Absoluto Médio (EAM) e um valor de 6,21% para o Erro Percentual Absoluto Médio (EPAM). Aplicou-se também o modelo ARIMA (0,2,2) registando-se um valor de 2,95 para a Raiz do Erro Quadrático Médio (REQM), um valor de 2,58 para o Erro Absoluto Médio (EAM) e um valor de 6,08 % para o Erro Percentual Absoluto Médio (EPAM). De seguida, aplicou-se o modelo ARIMA (1,2,2) registando-se um valor de 2,88 para a Raiz do Erro Quadrático Médio (REQM), um valor de 2,5 para o Erro Absoluto Médio (EAM) e um valor de 5,89 % para o Erro Percentual Absoluto Médio (EPAM). Por fim, aplicou-se o modelo ARIMA (2,2,1) registando-se um valor de 5,31 para a Raiz do Erro Quadrático Médio (REQM), um valor de 4,65 para o Erro Absoluto Médio (EAM) e um valor de 10,96 % para o Erro Percentual Absoluto Médio (EPAM).

Utilizando as medidas de precisão da previsão, conclui-se que o melhor modelo a utilizar é o ARIMA (0,2,1) (semelhante ao ARIMA (1,2,2)).

Erros de previsão dos modelos:

	REQM	EAM	EPAM (%)
Holt com Tendência Linear	2.87	2.51	5.91
Tendência Aditiva Amortecida	5.54	4.88	11.51
Tendência Exponencial	1.62	1.41	3.33
ARIMA (0,2,1)	2.87	2.51	5.91
ARIMA (1,2,1)	3.00	2.63	6.21
ARIMA (0,2,2)	2.95	2.58	6.08
ARIMA (1,2,2)	2.88	2.50	5.89
ARIMA (2,2,1)	5.31	4.65	10.96

Tabela 3- Erros de previsão dos modelos selecionados

Conclui-se que, entre todos os modelos utilizados para fazer a previsão, o método de Alisamento Exponencial com Tendência Exponencial é aquele que apresenta menores erros.

6. Previsão para o futuro com o melhor modelo

Realizou-se a previsão para os 6 anos futuros utilizando o método de Alisamento Exponencial com Tendência Exponencial. Visto que a presente série consiste num conjunto de dados anuais, de 1981 a 2020, sobre o consumo, em kg, de carnes brancas por habitante em Portugal, pretende prever-se o mesmo consumo para os anos de 2021 a 2026.

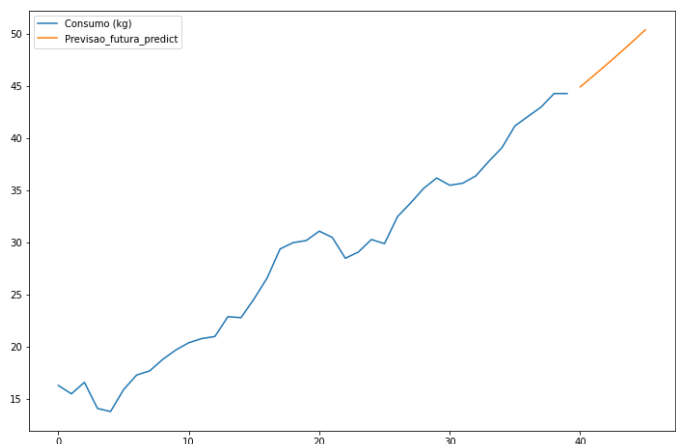


Figura 14- Previsão futura do consumo anual de carnes brancas (kg/hab) em Portugal

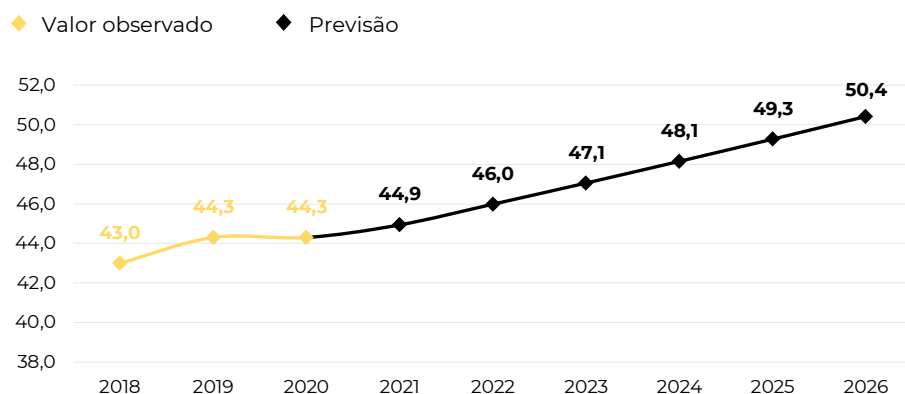


Figura 15- Previsão futura do consumo anual de carnes brancas (kg/hab) em Portugal

Através da observação da figura 14 é possível concluir que a previsão para os 6 anos futuros apresenta um comportamento que segue a mesma tendência crescente dos dados da série. Como é observável, no ano de 2020 existiu uma ligeira estagnação do consumo. Por este motivo, a previsão foi iniciada com valores pouco superiores aos do último ano, o que revela que o método utilizado dá maior destaque às observações mais recentes.

Com base na figura 15, constata-se que em 2021 existirá um aumento no consumo de carnes brancas por habitante em Portugal, face ao ano anterior, no valor de 600 gramas. Esta oscilação pouco relevante deve-se à estagnação que existiu no ano de 2020, observável na figura 14 e 15. A partir de 2021 prevê-se um crescimento anual significativo, no qual em 2026 o consumo deste tipo de carnes atinge os 50,4 kg por habitante, um crescimento percentual de 13,77% em relação a 2020.

Nesta série não é apresentado o intervalo de confiança da previsão para o futuro, uma vez que a biblioteca que calcula o intervalo de confiança não suporta o modelo de Alisamento Exponencial com Tendência Exponencial.

7. Previsão para o futuro com uma combinação dos três melhores modelos

7.1. Revisão da literatura

Podem realizar-se combinações dos modelos de previsão, e, por vezes, estas combinações podem originar um melhor resultado do que o obtido utilizando um modelo individualmente. Assim, existem vários critérios de combinação. Para o presente trabalho foram realizadas combinações dos modelos através de uma média simples e de uma média ponderada. A média ponderada atribui maior importância aos modelos com melhores resultados, isto é, utiliza o EAM (Erro Absoluto Médio) para ponderar inversamente as previsões. (Winkler e Makridakis (1983))

Combinação dos modelos através de uma média simples:

$\hat{Y}_{t+m}^c = \frac{1}{n} \sum_{i=1}^n \hat{Y}_{t+m}^i$, onde \hat{Y}_{t+m}^c representa a previsão da combinação dos modelos para instante $t + m$ e \hat{Y}_{t+m}^i representa a previsão do modelo i para o instante $t + m$.

Combinação dos modelos através de uma média ponderada:

$\hat{Y}_{t+m}^c = \frac{1}{(k-1)M} \sum_{i=1}^k (M - EAM^i) \hat{Y}_{t+m}^i$, com $M = \sum_{i=1}^k EAM^i$, sendo M a soma dos EAM de todos os modelos (EAM^i é o Erro Absoluto Médio do modelo i).

7.2. Resultados da previsão com combinação dos três melhores modelos

Utilizaram-se os três melhores modelos, ou seja, o método de Alisamento Exponencial com Tendência Exponencial, ARIMA (1,2,2) e ARIMA (0,2,1).

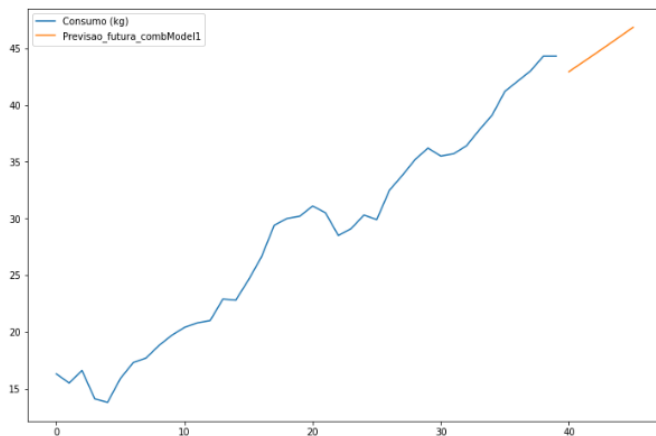


Figura 16- Previsão futura com combinação dos modelos usando uma média simples

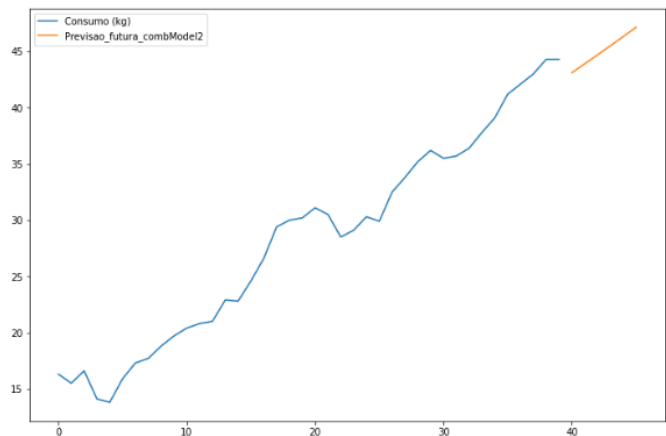


Figura 17- Previsão futura com combinação dos modelos usando uma média ponderada

Para avaliar estas combinações, podem obter-se os erros de previsão. Deste modo, foi realizada a previsão para o período de teste usando ambas as combinações e posteriormente estas foram avaliadas, como é visível na tabela 4.

	REQM	EAM	EPAM (%)
Combinação dos modelos através de uma média simples	2.45	2.13	5.02
Combinação dos modelos através de uma média ponderada	2.38	2.07	4.88
Tendência Exponencial	1.62	1.41	3.33

Tabela 4- Erros de previsão das combinações de modelos e do melhor modelo

Comparando os modelos, é notório que ambas as combinações, tanto através da média simples, como através da média ponderada, apresentam piores resultados do que o modelo escolhido (método de Alisamento Exponencial com Tendência Exponencial) para realizar a previsão. Assim, o modelo de previsão para o futuro mantém-se individual.

8. Conclusão

O objetivo deste estudo consiste na previsão do consumo anual, em kg, de carnes brancas *per capita* em Portugal, para os próximos 6 anos. A previsão é calculada com recurso ao *Software Python*, onde é selecionado o modelo de previsão com melhores resultados. Esta seleção foi feita com base num critério de comparação entre os erros de previsão dos vários modelos selecionados.

Apesar do modelo de Alisamento Exponencial com Tendência Exponencial, ser um modelo menos complexo que os modelos ARIMA, este apresenta melhores resultados nos erros de previsão, ou seja, os valores da previsão aproximam-se mais da série original no período de teste. Assim, pode deduzir-se que nem sempre os modelos mais complexos originam as melhores previsões.

É de salientar que os resultados obtidos com a combinação dos três melhores modelos (um de alisamento exponencial e dois SARIMA) não superaram os alcançados com o melhor modelo anteriormente selecionado. Este acontecimento deve-se ao facto das previsões dos três melhores modelos apresentarem características muito similares, isto é, existe pouca diversidade nas previsões. Deste modo, a diferença apresentada entre os modelos não é suficiente para atingir o objetivo da combinação dos modelos.

Utiliza-se o modelo de Alisamento Exponencial com Tendência Exponencial de modo a projetar uma previsão nos 6 anos seguintes, onde é possível visualizar que existe uma tendência crescente.

O crescimento do consumo de carnes brancas é expectável devido ao efeito ambiental, inerente ao consumo alimentar de carnes vermelhas, e à sensibilização da população portuguesa para as práticas de uma alimentação mais saudável. No entanto, este crescimento regista um comportamento rápido e acentuado, contrário ao esperado. Desta forma, é provável que a previsão efetuada não corresponda ao crescimento real dos próximos anos.

9. Referências

Instituto Nacional de Estatística. Consumo humano de carne per capita (kg/ hab.) por Tipo de carnes. Consultado em 21 mar. 2022. Disponível em

https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&contexto=pi&indOcorrCod=0000211&selTab=tab0.

Gonzaga, S.(2019). Curso de Séries Temporais. Consultado em 14 mai. 2022. Disponível em http://sillasgonzaga.com/material/curso_series_temporais/.

Hyndman, R. e Athanasopoulos, G. (2014): Forecasting: Principles and Practice, OTexts.com

Makridakis, S., Wheelwright, S. e Hyndman, R. (1998): Forecasting: Methods and Applications, 3ª edição, John Wiley & Sons, New York

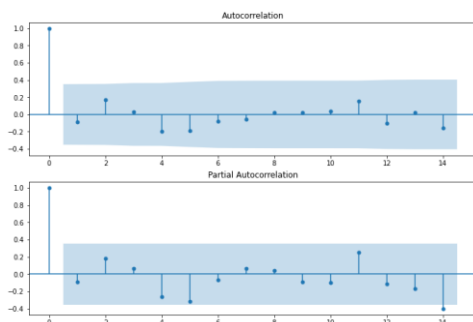
Brockwell, P. and Davis, R. (2016): Introduction to Time Series and Forecasting, 3rd edition, Springer

Associação Vegetariana Portuguesa (2021). Mais de 1 Milhão de Portugueses Está a Deixar de Comer Carne. Consultado em 3 jun. 2022. Disponível em <https://www.avp.org.pt/mais-1-milhao-portugueses-deixar-comer-carne/>

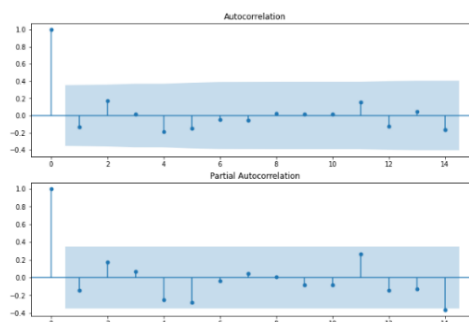
Renascença (2015). Portugueses comem cada vez mais carne de aves e menos carne de vaca e porco. Consultado em 3 jun. 2022. Disponível em <https://rr.sapo.pt/noticia/pais/2015/10/26/portugueses-comem-cada-vez-mais-carne-de-aves-e-menos-carne-de-vaca-e-porco/37767/>

Winkler e Makridakis (1983)] Winkler, R, & Makridakis, S 1983, 'The Combination of Forecasts', Journal of the Royal Statistical Society. Series A (General).

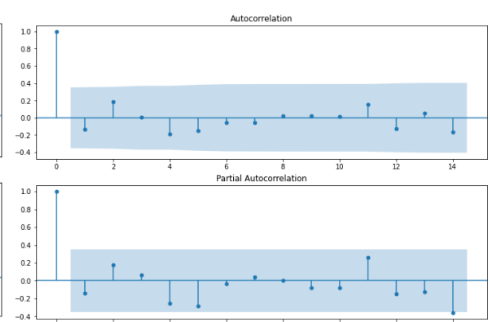
Anexos



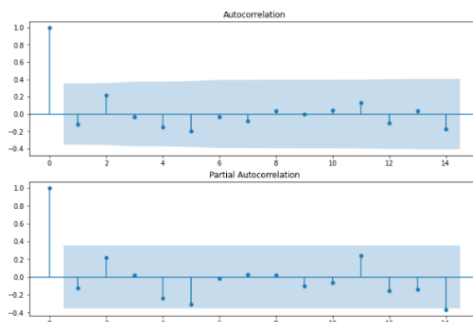
Anexo 1- Correlograma dos resíduos do modelo ARIMA (0,2,1)



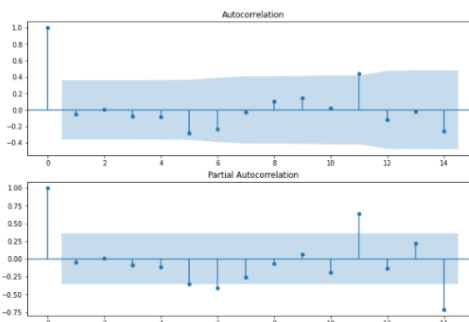
Anexo 2- Correlograma dos resíduos do modelo ARIMA (1,2,1)



Anexo 3- Correlograma dos resíduos do modelo ARIMA (0,2,2)



Anexo 4- Correlograma dos resíduos do modelo ARIMA (1,2,2)



Anexo 5- Correlograma dos resíduos do modelo ARIMA (2,2,1)