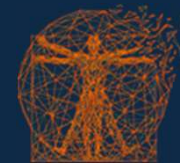


PREVISÃO DE DEMANDA EM SÉRIES TEMPORAIS (EMBASAMENTO TEÓRICO)

Grupo Sydney - 2020




INSTITUTO
D'VINCI³





AGOSTINHO ALONSO

Desenvolvedor

 [linkedin.com/in/agostinho-f-alonso-jr-a27b6327](https://www.linkedin.com/in/agostinho-f-alonso-jr-a27b6327)



ALINE DA SILVA

UX Design

 [linkedin.com/in/alinedasilva](https://www.linkedin.com/in/alinedasilva)



ANA PAULA

Desenvolvedora

 [linkedin.com/in/ana-oliveira-39b54254](https://www.linkedin.com/in/ana-oliveira-39b54254)



JULIO SILVA

Scrum Master

 [linkedin.com/in/jcsilvaxxi](https://www.linkedin.com/in/jcsilvaxxi)



NÁDIA ALENCAR

Product Owner

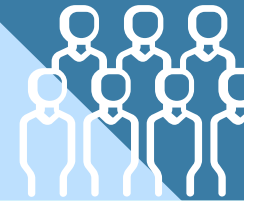
 [linkedin.com/in/nadiaalencar](https://www.linkedin.com/in/nadiaalencar)



VANESSA ROSA

UX Design

 [linkedin.com/in/vanessa-rosa-278167204/](https://www.linkedin.com/in/vanessa-rosa-278167204/)

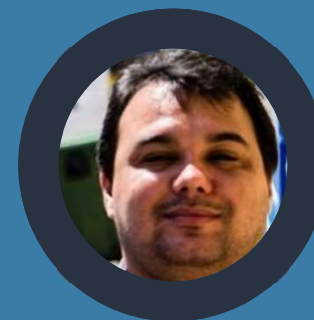


MENTORES



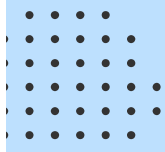
CRISTINA AUGUSTO

[linkedin.com/in/cristina-maria-martinez-augusto-b74379](https://www.linkedin.com/in/cristina-maria-martinez-augusto-b74379)



JHEAN CAMARGO

[linkedin.com/in/jhean-camargo-44a52937](https://www.linkedin.com/in/jhean-camargo-44a52937)



AIKNOW

Soluções que a vida pede



ROGÉRIO ZERO

CEO e sócio fundador

[linkedin.com/in/rogerio-zero-647533b](https://www.linkedin.com/in/rogerio-zero-647533b)



LUIS CABANHAS

Chief Revenue Officer (CRO) & Partner

[linkedin.com/in/luiscabanhas/](https://www.linkedin.com/in/luiscabanhas/)

> CLIENTE

O QUE É A AIKNOW?

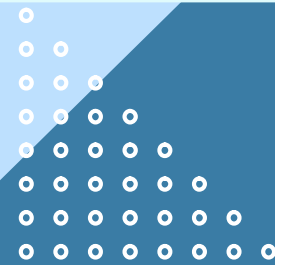
A Aiknow integra a tecnologia de Inteligência Artificial e Computação Cognitiva (reconhecimento facial e corporal biométrico), desenvolvendo soluções para otimizar processos e automatizar funções em sua empresa.

Através de soluções customizadas e inovadoras, a Aiknow utiliza algoritmos para cada necessidade do negócio, trazendo melhoras significativas na produtividade, faturamento e relacionamento com o cliente, tornando a empresa mais atrativa e eficiente.



AIKNOW

SÉRIES TEMPORAIS



PROCESSOS ESTOCÁSTICOS

Dentro da teoria das probabilidades, um processo estocástico é uma família de variáveis aleatórias representando a evolução de um sistema de valores com o tempo. É a contraparte probabilística de um processo determinístico. Ao invés de um processo que possui um único modo de evoluir, em um processo estocástico há uma indeterminação: mesmo que se conheça a condição inicial, existem várias, por vezes infinitas, direções nas quais o processo pode evoluir.

Em casos de tempo discreto, em oposição ao tempo contínuo, o processo estocástico é uma sequência de variáveis aleatórias.

Apesar dos valores aleatórios de um processo estocástico, em momentos diferentes, parecerem variáveis aleatórias independentes, nas situações mais comuns, eles exibem uma complexa dependência estatística.

Exemplo de processos estocásticos incluem flutuações nos mercados de ações e nas taxas de câmbio, dados médicos como temperatura, pressão sanguínea e variações nos potenciais elétricos do cérebro registrados em um eletroencefalograma, fluxo turbulento de um líquido ou gás, variações no campo magnético da Terra, mudanças aleatórias no nível de sinais de rádio sintonizados na presença de distúrbios meteorológicos, entre outros.



SÉRIES TEMPORAIS

Pode-se definir uma série temporal como sendo um conjunto de dados observados e ordenados segundo parâmetros de tempo e com dependência serial, sendo esse espaço de tempo entre os dados disponíveis equidistantes (horários, diário, semanal, mensal, trimestral, anual, etc.) (Souza & Camargo, 2004).

Para que uma determinada série seja classificada como uma série temporal, é necessário que ela preencha outro pré-requisito: os dados também devem apresentar uma dependência serial entre eles. Por exemplo: os dados de uma variável aleatória z no instante t , com t variando de 1 até N , possa, de certa maneira, conter informações necessárias para que seja determinado o valor dessa variável no instante $t + 1$.

A maioria dos métodos de previsão baseia-se na ideia de que as observações passadas contêm informações sobre o padrão de comportamento da série temporal.

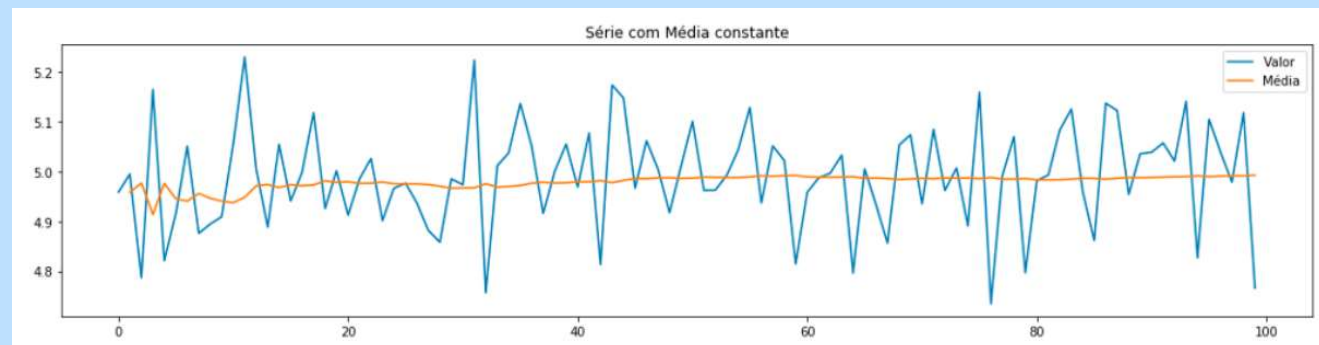


SÉRIES TEMPORAIS ESTACIONÁRIAS

Uma suposição comum em muitas técnicas de séries temporais é que os dados sejam estacionários.

Uma série temporal é dita estacionária quando ela se desenvolve no tempo aleatoriamente ao redor de uma média constante, refletindo alguma forma de equilíbrio estável.

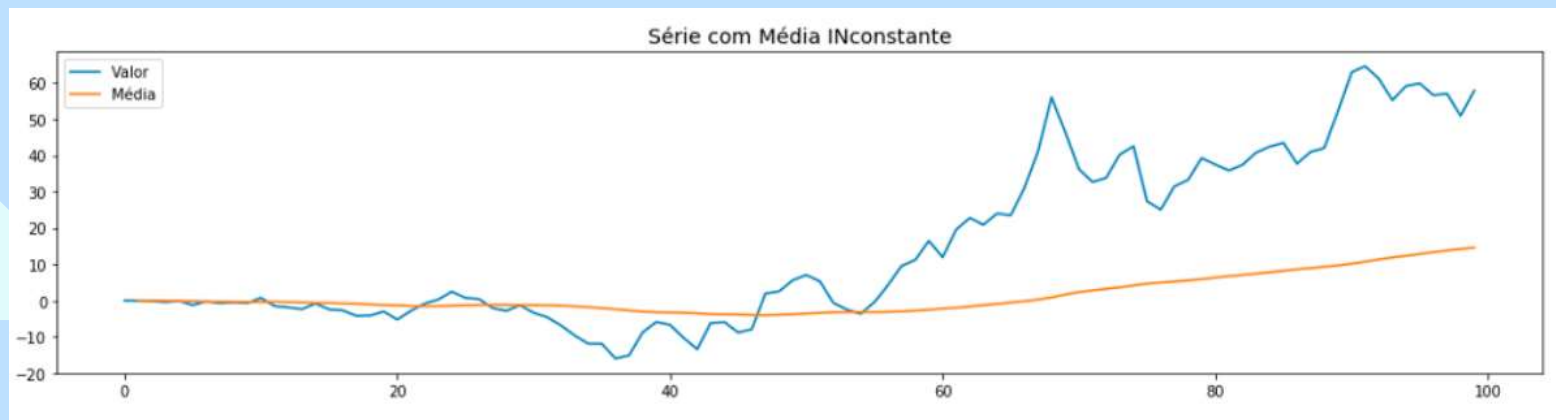
Um processo estacionário tem a propriedade de que a média, variância e estrutura de autocorrelação não mudam no decorrer do tempo, uma série parecida com um plano liso, sem tendência, com variância e estrutura de autocorrelação constante no decorrer do tempo e sem nenhuma flutuação periódica (sazonalidade).



SÉRIES TEMPORAIS EXPLOSIVAS

Uma série temporal explosiva é o contrário de uma série temporal estacionária. Suas propriedades estatísticas mudam com o tempo e possuem tendência e/ou sazonalidade.

Caso a série não seja estacionária, a previsão para o futuro não será eficiente, pois as variações de valores ao redor da média se afastam significativamente.

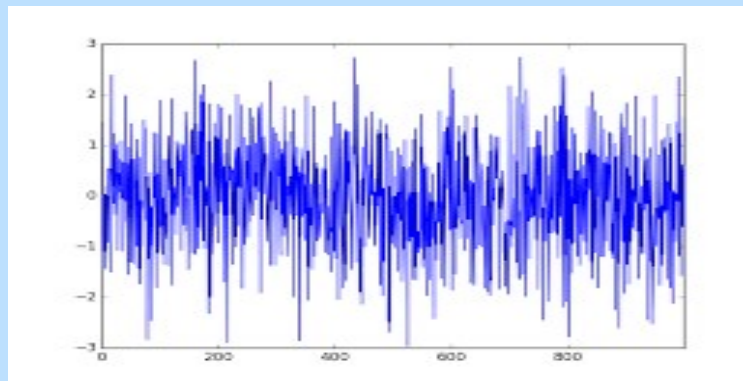


RUÍDO BRANCO

Uma série temporal é ruído branco se as variáveis são independentes e identicamente distribuídas com média zero.

O ruído branco é um sinal discreto cujas amostras são vistas como uma sequência de variáveis aleatórias não auto correlacionadas com média zero e variância finita.

Uma única ocorrência de ruído branco é um choque aleatório. Dependendo do contexto, pode-se também exigir que as amostras sejam independentes e tenham distribuição de probabilidade idêntica (em outras palavras, variáveis aleatórias independentes e identicamente distribuídas são a representação mais simples de ruído branco).

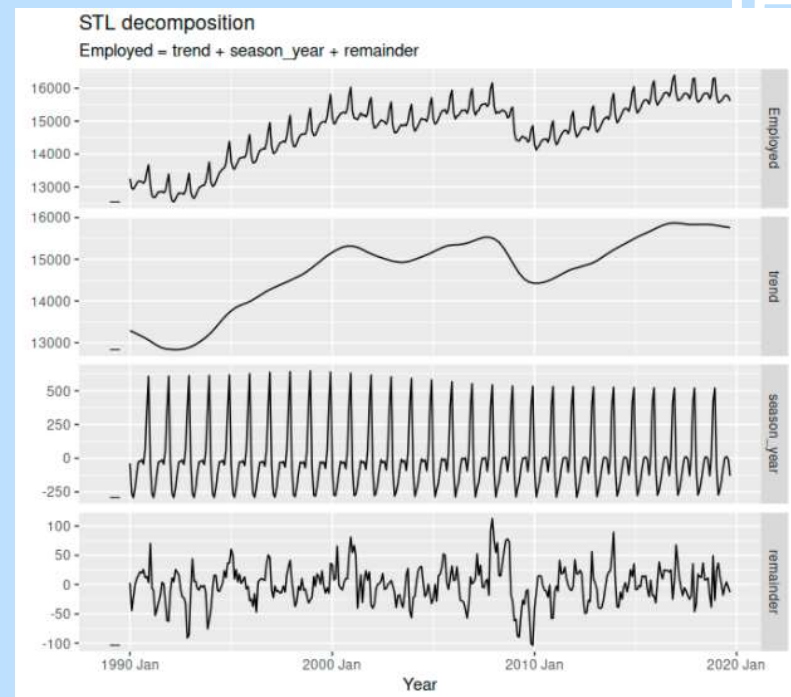


DECOMPOSIÇÃO DE SÉRIES TEMPORAIS

O processo de decomposição de uma série temporal é feito para melhorar sua compreensão ou para melhorar a precisão da previsão.

É a maneira tradicional de analisar uma série temporal, decompondo-a em seus componentes de Tendência, Ciclo, Sazonalidade e Ruído.

Quando decompomos uma série temporal em componentes, geralmente combinamos a tendência e o ciclo em um único componente: “ciclo de tendência” (geralmente chamado apenas de tendência para simplificar). Assim, pensamos em uma série temporal como compreendendo três componentes: um componente de ciclo de tendência, um componente sazonal e um componente ruído (contendo qualquer outra coisa na série temporal).

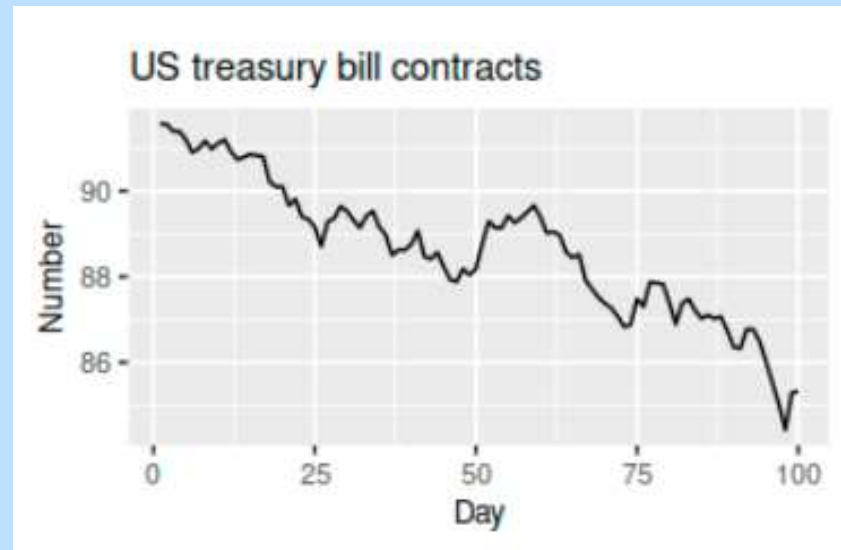


DECOMPOSIÇÃO DE SÉRIES TEMPORAIS

(Tendência – “Trend”)

A tendência de uma série indica o seu comportamento “de longo prazo”, isto é, se ela cresce, decresce ou permanece estável, e qual a velocidade destas mudanças.

Mais especificamente, podemos dizer que uma tendência existe quando há um aumento ou redução de longo prazo associados aos dados.

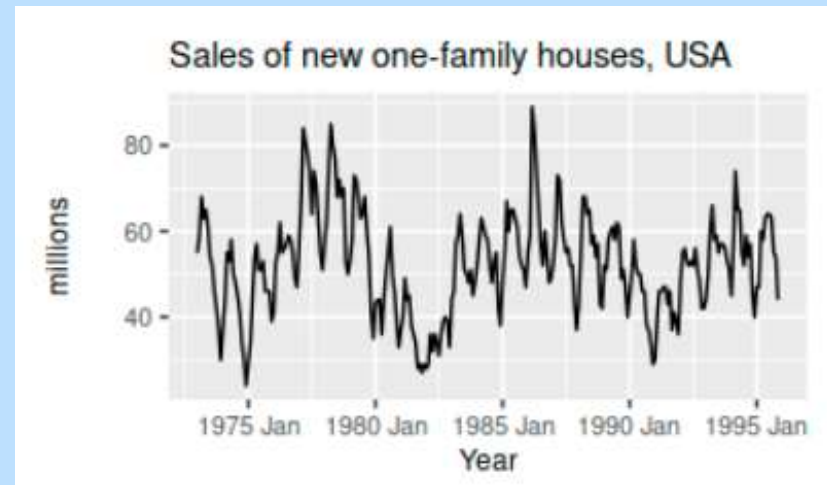


Os contratos de letras do tesouro dos EUA mostram os resultados do mercado de Chicago por 100 dias consecutivos de negociação em 1981. Aqui não há sazonalidade, mas uma tendência de queda óbvia.

DECOMPOSIÇÃO DE SÉRIES TEMPORAIS (Ciclos)

Os ciclos são caracterizados pelas oscilações de subida e de queda nas séries, de forma suave e repetida, ao longo da componente de tendência. Por exemplo, ciclos relacionados à atividade econômica ou ciclos meteorológicos.

Muitas pessoas confundem comportamento cíclico com comportamento sazonal, mas eles são bem diferentes. Se as flutuações não forem de frequência fixa, serão cíclicas; se a frequência for imutável e associada a algum aspecto do calendário, o padrão é sazonal. Em geral, a duração média dos ciclos é maior do que a duração de um padrão sazonal, e as magnitudes dos ciclos tendem a ser mais variáveis do que as magnitudes dos padrões sazonais.

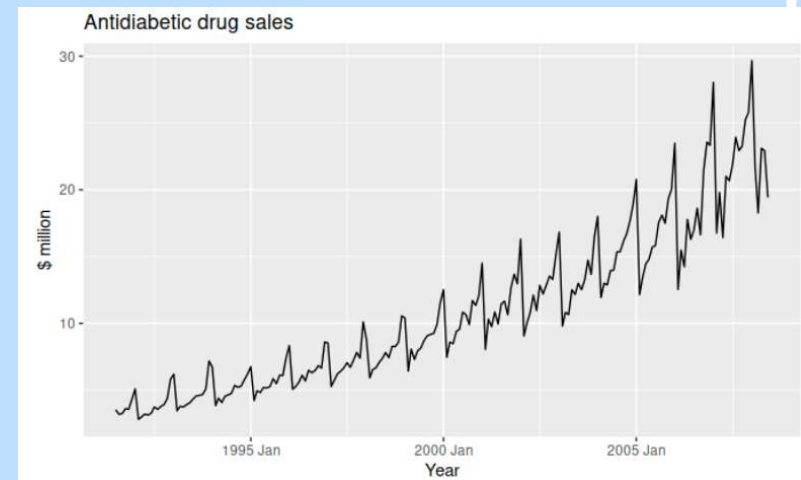


As vendas mensais de habitação mostram forte sazonalidade dentro de cada ano, bem como algum forte comportamento cíclico com um período de cerca de 6 a 10 anos.

DECOMPOSIÇÃO DE SÉRIES TEMPORAIS (Sazonalidade)

As componentes *sazonais* em uma série são aquelas oscilações de subida e de queda que sempre ocorrem em um determinado período do ano, do mês, da semana, do dia ou do horário.

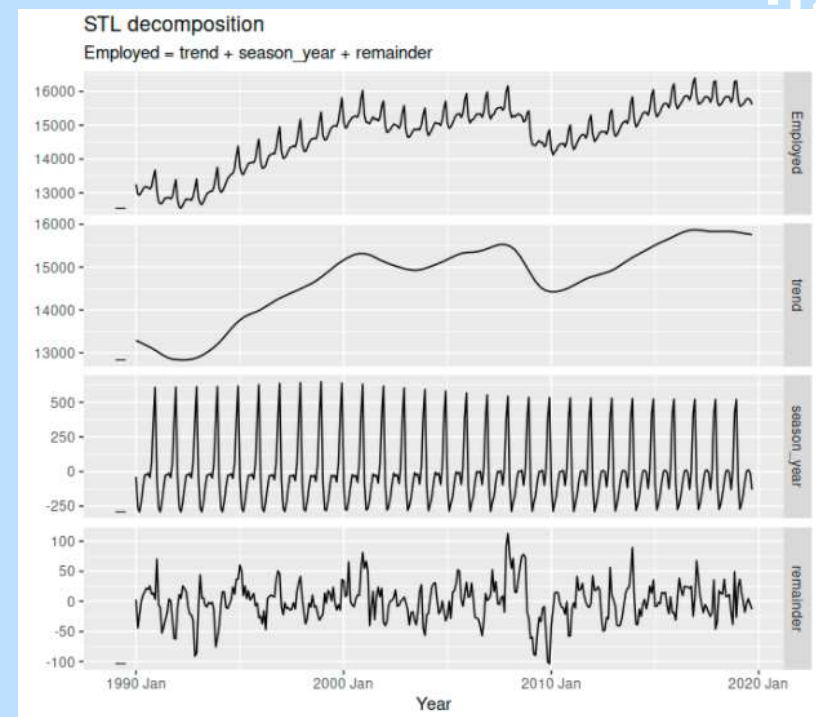
A diferença essencial entre as componentes sazonais e cíclicas é que a primeira possui movimentos facilmente previsíveis, ocorrendo em intervalos regulares de tempo, por exemplo, ano a ano, mês a mês, semana a semana, ou mesmo dia a dia. Já os movimentos cíclicos tendem a ser irregulares, ocorrendo sobre um período de muitos anos.



Vendas mensais de medicamentos antidiabéticos na Austrália. Aqui, há também um forte padrão sazonal que aumenta de tamanho à medida que o nível da série aumenta. A queda repentina no início de cada ano é causada por um esquema de subsídios do governo que torna mais econômico para os pacientes estocar medicamentos no final do ano civil.

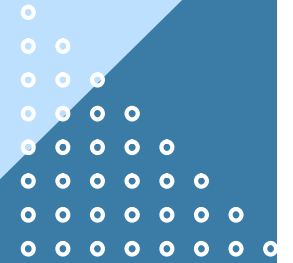
DECOMPOSIÇÃO DE SÉRIES TEMPORAIS (Ruído)

A quarta componente da série, chamada de *Ruído* (*Random*), é o que resta após retirar da série os outros componentes: Tendência e Sazonalidade.





METODOLOGIA BOX & JENKINS



INTRODUÇÃO

A metodologia Box-Jenkins foi proposta por George Box e Gwilym Jenkins no livro: Time Series Analysis: Forecasting and Control de 1970.

O modelo Box-Jenkins é um modelo matemático projetado para prever dados futuros com base nos valores de uma série temporal especificada.

O modelo Box-Jenkins pode analisar vários tipos diferentes de séries temporais para previsão.

O modelo Box-Jenkins prevê dados usando três princípios: autorregressão, diferenciação e média móvel.

Esses três princípios são conhecidos como p , d e q , respectivamente. Cada princípio é usado na análise de Box-Jenkins e juntos são mostrados coletivamente como ARIMA (p , d , q).

Os termos ARIMA e modelo Box-Jenkins podem ser usados alternadamente.





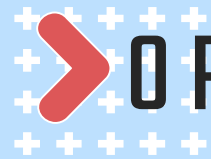
O MODELO ARIMA

O modelo ARIMA é comumente representado como ARIMA (p,d,q) onde (p) representa o número de parâmetros auto-regressivos, (d) o número de diferenciações para que a série torne-se estacionária e (q) o número de parâmetros de médias móveis.

O processo de autorregressão (p) testa os dados quanto ao seu nível de estacionariedade. Se os dados usados forem estacionários, isso pode simplificar o processo de previsão. Se os dados usados não forem estacionários, será necessário diferenciá-los (d).

Os dados também são testados quanto ao seu ajuste de média móvel, o que é feito na parte q do processo de análise. No geral, a análise inicial dos dados os prepara para a previsão, determinando os parâmetros (p, d e q) que são aplicados para desenvolver uma previsão.





O PROCESSO DE CONSTRUÇÃO DO MODELO

A aplicação dos modelos de Box & Jenkins para fins de previsão de séries temporais segue as seguintes etapas: identificação, estimação e verificação.

Na primeira fase o que se deseja é *identificar* o processo aleatório que gerou os dados, para em seguida *estimar* os parâmetros que o caracterizam e *verificar* se as hipóteses do modelo foram cumpridas. Caso negativo, uma nova fase de identificação deve ser considerada até que a verificação das hipóteses seja finalmente positiva, permitindo então a realização da *previsão*.





IDENTIFICAÇÃO DO MODELO



Na etapa de identificação de um modelo ARIMA empregam-se procedimentos que possam identificar a estrutura do modelo, isto é, permitam conhecer os valores dos parâmetros d , p e q que caracterizam o processo estocástico.

Os procedimentos de identificação consistem de duas partes:

- a) inicialmente diferencia-se a série temporal original tantas vezes quantas necessárias para obtenção de uma série estacionária, de modo a possibilitar a análise do processo com o modelo ARMA(p, q); o número de diferenciações d é aquele necessário para que a função de autocorrelação amostral (ACF) da série transformada decresça rapidamente para zero;
- b) a identificação de um processo AR(p), MA(q) ou ARMA(p, q) é feita através da análise das funções de autocorrelação simples (ACF) e da autocorrelação parcial (PACF), com determinação dos valores dos parâmetros p , q



AUTOCORRELAÇÃO SIMPLES E PARCIAL

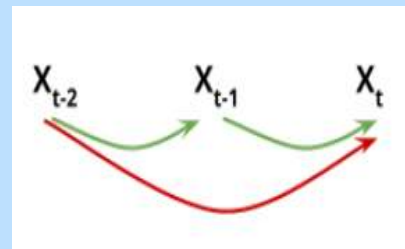
(Testes ACF e PACF)

Depois de uma série temporal ser “estacionarizada”, o passo seguinte é determinar os parâmetros para os termos autorregressivos e de médias móveis de forma a corrigir qualquer autocorrelação que permaneça na série.

Autocorrelação é a correlação entre os valores da série em um determinado período de tempo e os valores da mesma série em um outro momento.

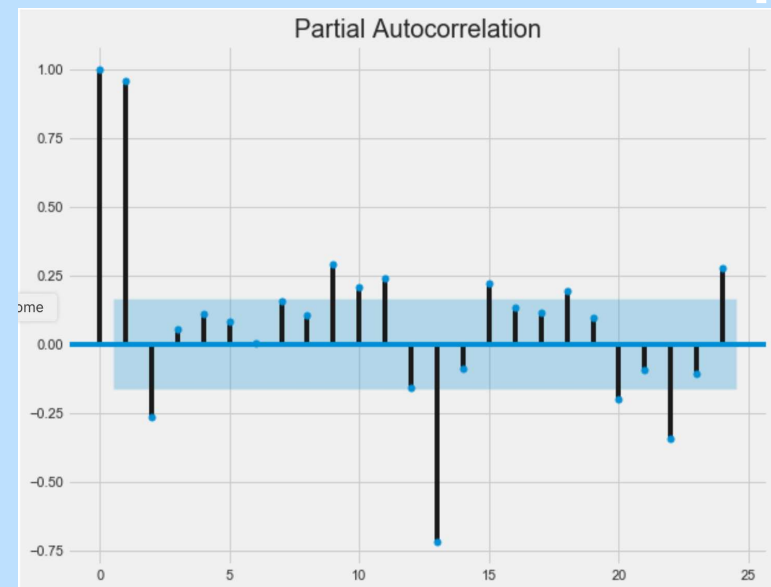
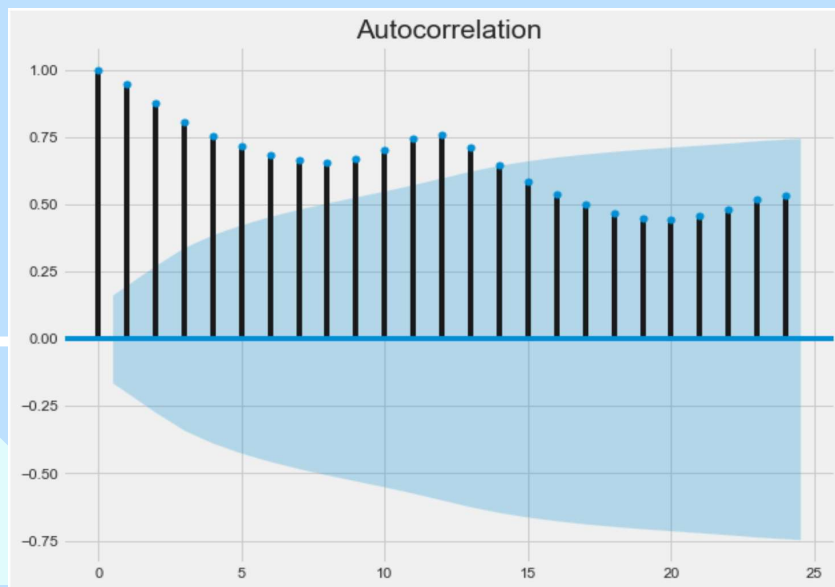
As funções de autocorrelação (FAC ou ACF em inglês) e de autocorrelação parcial (FACP ou PACF) da série contêm os valores de autocorrelação em função do intervalo de tempo (lag) em que foi calculado. O PACF tem como objetivo filtrar correlações de outros lags e manter apenas a correlação pura entre duas observações. Eles calculam a correlação de uma observação com os valores de atraso.

A diferença entre os dois métodos é que o PACF considera os efeitos diretos entre as defasagens X_{t-2} — X_t (setas vermelhas).



AUTOCORRELAÇÃO SIMPLES E PARCIAL

(Testes ACF e PACF)



+++++

AUTOCORRELAÇÃO SIMPLES E PARCIAL




(Testes ACF e PACF)

Modelo	Padrão teórico típico ACF	Padrão teórico típico PACF
AR(p)	Declina exponencialmente e/ou com padrão de onda senoidal amortecida	Apresentam picos significativos até defasagens
MA (q)	Apresentam picos significativos até q defasagens	Declínio exponencial
ARMA (p,q)	Declínio exponencial	Declínio exponencial

(GUAJARATI, 2000)

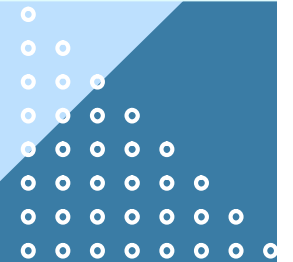
AUTOCORRELAÇÃO SIMPLES E PARCIAL

(Resultados esperados ACF e PACF)

Padrão	O que o padrão indica	Exemplo
Um grande pico no lag 1 que diminui depois de alguns lags.	Um termo auto-regressivo nos dados. Use a função de autocorrelação parcial para determinar a ordem do termo auto-regressivo.	
Um grande pico no lag 1 seguido por uma onda decrescente que alterna entre correlações positivas e negativas.	Um termo auto-regressivo de ordem superior nos dados. Use a função de autocorrelação parcial para determinar a ordem do termo auto-regressivo.	
Correlações significativas no primeiro ou segundo lag, seguidas por correlações que não são significativas.	Um termo de média móvel nos dados. O número de correlações significativas indica a ordem do termo da média móvel.	



ESTIMATIVA DOS PARAMÊTROS





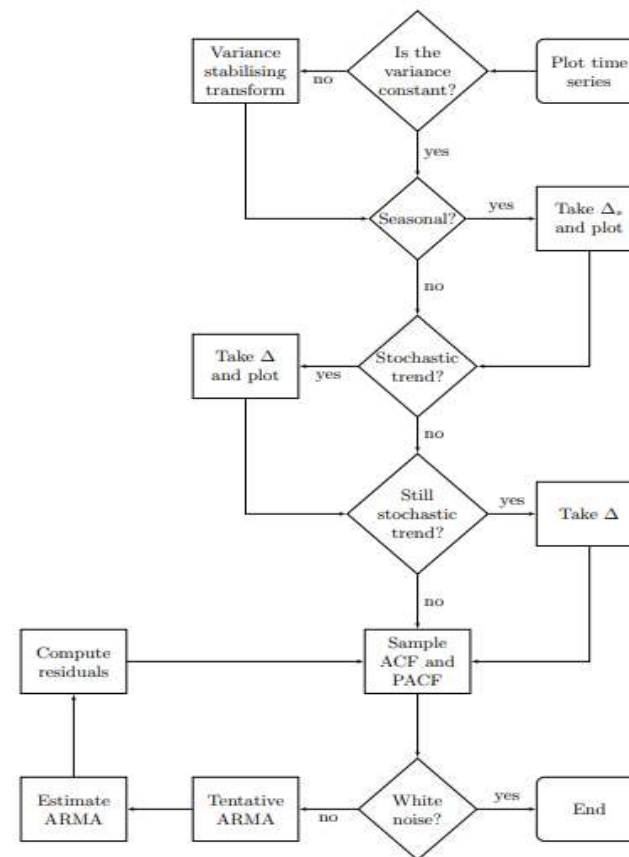
Após a identificação dos valores d , p , q passa-se à estimativa dos parâmetros do modelo. Esta etapa é geralmente executada através de “software” específico para análise de séries temporais.

Existem diferentes métodos para a estimativa dos parâmetros do modelo, e todos produzem resultados semelhantes, embora possam existir diferenças quanto à eficiência na implementação computacional. De maneira geral, durante a etapa de estimativa usa-se algum procedimento iterativo de estimação de mínimos quadrados não linear ou através dos métodos de máxima verossimilhança aproximada.



ESTIMATIVA DOS PARÂMETROS

Box & Jenkins indicam um procedimento para descobrir o modelo ARIMA que melhor se ajusta aos dados:





Teste PP (Phillips-Perron) / Teste de Dickey-Fuller (ADF)

Os testes de Phillips-Perron e Dickey-Fuller são testes de raiz unitária. O teste de Dickey-Fuller (DF) é o mais antigo e famoso método formal para verificar se uma série temporal é ou não estacionária.

O desenvolvimento do teste DF foi motivado pela necessidade de verificar se uma série precisa ser diferenciada para se tornar estacionária. Essa verificação é a primeira etapa da metodologia proposta por Box Jenkins (1970) para construção de modelos ARIMA para séries temporais. Esses autores recomendaram que, se o gráfico da série temporal indicar que ela é não estacionária, então devemos diferenciá-la até apresentar um padrão estacionário.

Nesse sentido, o primeiro aspecto a destacar sobre o teste DF é que ele procura verificar se uma série é estacionária ou não testando se ela possui uma raiz unitária.





Teste PP (Phillips-Perron) / Teste de Dickey-Fuller (ADF)

Em modelos de séries temporais em econometria (aplicação de métodos estatísticos à economia), a unidade de raiz é uma característica dos processos que evoluem ao longo do tempo e que podem causar problemas na inferência estatística, se não for tratada adequadamente.

A maioria das séries econômicas possuem raiz unitária. Quando a hipótese de raiz unitária for verdadeira para uma série, os choques aleatórios que ela sofre geram na mesma um efeito permanente. As flutuações neste caso não são transitórias, derrubando, por exemplo, as teorias de que os ciclos econômicos seriam flutuações temporárias em torno de uma tendência.

Portanto, quando uma variável apresenta raiz unitária, os pressupostos estatísticos de que a média e a variância devem ser constantes ao longo do tempo são violados comprometendo, dessa forma, os resultados obtidos com a utilização de modelos econométricos. A regressão, neste caso, é considerada espúria (sem significado econômico).



Teste KPSS

Um teste amplamente utilizado em econometria é o Kwiatkowski – Phillips – Schmidt – Shint ou abreviado como teste KPSS. Este teste também é bastante semelhante ao ADF e pode ajudar a validar a hipótese nula de que uma série temporal observável é estacionária em torno de uma tendência determinística. No entanto, há uma grande desvantagem de ter uma alta taxa de erros do tipo I. Nesses casos, muitas vezes é recomendado combiná-lo com o teste ADF e verificar se ambos retornam os mesmos resultados. O código é semelhante ao teste ADF, conforme mostrado abaixo:

NON-STATIONARY		STATIONARY	
FIGURE A		FIGURE B	
KPSS Statistic	1.052175	KPSS Statistic	0.152538
p-value	0.010000	p-value	0.100000
#Lags Used	14.000000	#Lags Used	13.000000
Critical Value (10%)	0.347000	Critical Value (10%)	0.347000
Critical Value (5%)	0.463000	Critical Value (5%)	0.463000
Critical Value (2.5%)	0.574000	Critical Value (2.5%)	0.574000
Critical Value (1%)	0.739000	Critical Value (1%)	0.739000
dtype: float64		dtype: float64	

Podemos ver na imagem ao lado que antes de aplicar a transformação (figura A) o valor p dos dados é $<0,05$ e, portanto, não é estacionário. Após a transformação (figura B), o valor p torna-se 0,1 a. confirmar a estacionariedade dos dados.



VERIFICAÇÃO DO MODELO



A ideia da verificação é procurar evidências de que o modelo representa bem os dados. Duas áreas úteis para investigar diagnósticos são:

- a) Overfitting
- b) Erros residuais

1. **Overfitting**

A primeira verificação é validar se o modelo super ajusta-se aos dados. Geralmente, isso significa que o modelo é mais complexo do que precisa ser e captura ruído branco nos dados de treinamento.

Esse é um problema para a previsão de séries temporais porque afeta negativamente a capacidade do modelo de generalizar, resultando em um desempenho de previsão ruim em dados fora da amostra.





2. Erros Residuais

Resíduos de previsão fornecem uma grande oportunidade para diagnósticos.

Uma revisão da distribuição de erros pode ajudar a descobrir o viés do modelo. Os erros de um modelo ideal seriam semelhantes ao **ruído branco**, que é uma distribuição gaussiana com uma média de zero e uma variância simétrica. Para isso, podem ser usados gráficos de densidade, histogramas e gráficos Q-Q que comparam a distribuição de erros com a distribuição esperada. Uma distribuição não gaussiana pode sugerir uma oportunidade para o pré-processamento de dados. Uma distorção na distribuição ou uma média diferente de zero pode sugerir um viés nas previsões que podem estar corretas.

Além disso, um modelo ideal não deixaria estrutura temporal na série temporal dos resíduos da previsão. Isso pode ser verificado criando gráficos ACF e PACF da série temporal de erro residual.

A presença de correlação serial nos erros residuais sugere maior oportunidade de uso dessa informação no modelo.





TESTE DE JARQUE-BERA

O teste de Jarque-Bera é um teste de adequação para determinar se os resíduos têm a **assimetria** e a **curtose** correspondentes a uma distribuição normal. O teste recebeu o nome de Carlos Jarque e Anil K. Bera.

Em geral, se

$JB(Pvalue < 0.10) = \text{Aceita } H_0 \text{ (distribuição normal)}$

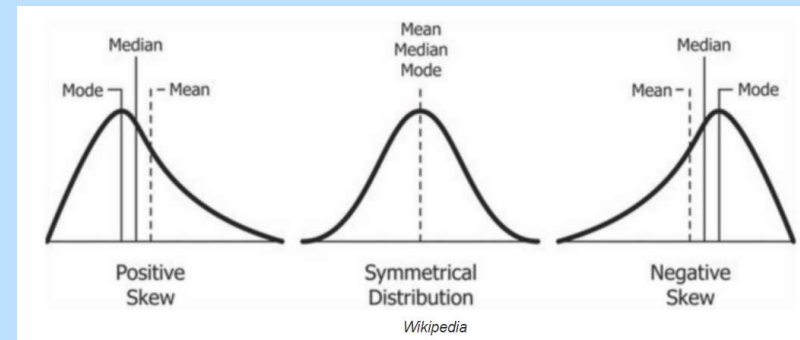
$JB(Pvalue \geq 0.10) = \text{Rejeita } H_0 \text{ (distribuição não normal)}$



➤ ASSIMETRIA

Podemos dizer que a assimetria indica o quanto nossa distribuição subjacente se desvia da distribuição normal, pois a distribuição normal tem assimetria 0. Geralmente, temos três tipos de assimetria.

- Simétrico : quando a assimetria é próxima de 0 e a média é quase igual à mediana
- Inclinação negativa : Quando a cauda esquerda do histograma da distribuição é mais longa e a maioria das observações está concentrada na cauda direita. Nesse caso, podemos usar também o termo “enviesado para a esquerda” ou “caudal para a esquerda”. e a mediana é maior que a média.
- Inclinação positiva : quando a cauda direita do histograma da distribuição é mais longa e a maioria das observações está concentrada na cauda esquerda. Nesse caso, podemos usar também o termo “enviesado à direita” ou “caudal à direita”. e a mediana é menor que a média.



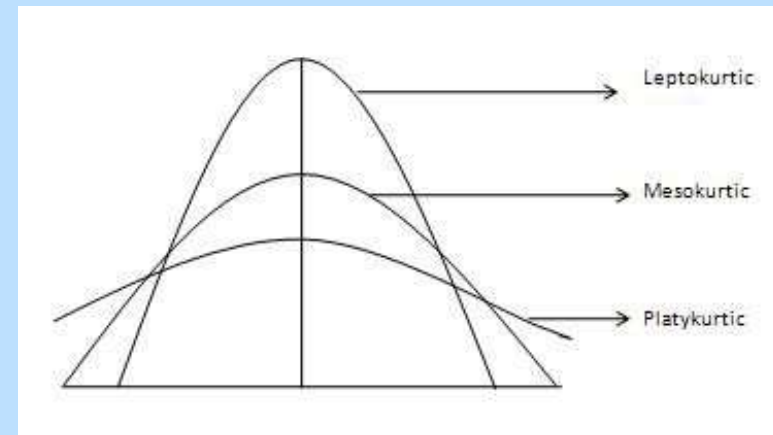
++++
++++
++++
++++

CURTOSE

Em estatística, usamos a medida de curtose para descrever a “cauda” da distribuição, uma vez que descreve a forma dela. É também uma medida do “pico” da distribuição. Uma distribuição de curtose alta tem um pico mais nítido e caudas mais longas e mais grossas, enquanto uma distribuição de curtose baixa tem um amendoim mais arredondado e caudas mais curtas e finas.

Vamos ver os três principais tipos de curtose.

- Mesocúrtico : Esta é a distribuição normal
- Leptocúrtico : esta distribuição tem caudas mais grossas e um pico mais nítido. A curtose é "positiva" com um valor maior que 3
- Platicúrtico : A distribuição tem um pico mais baixo e mais largo e caudas mais finas. A curtose é "negativa" com um valor inferior a 3



+++++

ANÁLISE DE RESÍDUOS

Tanto na Regressão Linear Simples quanto na Regressão Múltipla, as suposições do modelo ajustado precisam ser validadas para que os resultados sejam confiáveis. Chamamos de **Análise dos Resíduos** um conjunto de técnicas utilizadas para investigar a adequabilidade de um modelo de regressão com base nos resíduos. Como visto anteriormente, o resíduo (e_i) é dado pela diferença entre a variável resposta observada (Y_i) e a variável resposta estimada (\hat{Y}_i), isto é

$$e_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \dots - \hat{\beta}_p x_{pi} \quad i = 1, \dots, n$$

A ideia básica da análise dos resíduos é que, se o modelo for apropriado, os resíduos devem refletir as propriedades impostas pelo termo de erro do modelo. Tais suposições são $Y = X\beta + \varepsilon$,

em que $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$, com

- i. ε_i e ε_j são independentes ($i \neq j$);
- ii. $\text{Var}(\varepsilon_i) = \sigma^2$ (constante);
- iii. $\varepsilon_i \sim N(0, \sigma^2)$ (normalidade);
- iv. Modelo é linear;
- v. Não existir outliers (pontos atípicos) influentes.



ANÁLISE DE RESÍDUOS

As técnicas utilizadas para verificar as suposições descritas acima podem ser informais (como gráficos) ou formais (como testes). As técnicas gráficas, por serem visuais, podem ser subjetivas e por isso técnicas formais são mais indicadas para a tomada de decisão. O ideal é combinar as técnicas disponíveis, tanto formais quanto informais, para o diagnóstico de problemas nas suposições do modelo.

Algumas técnicas gráficas para análise dos resíduos são:

- Gráfico dos resíduos versus valores ajustados: verifica a homoscedasticidade do modelo, isto é, σ^2 constante.
- Gráfico dos resíduos versus a ordem de coleta dos dados: avaliar a hipótese de independência dos dados.
- Gráfico dos Resíduos Studentizados versus valores ajustados: verifica se existem outliers em Y.
- Gráfico dos Resíduos Padronizados versus valores ajustados: verifica se existem outliers em Y.
- Gráfico do Leverage (Diagonal da Matriz H): verifica se existem outliers em X.



ANALISE DE RESÍDUOS

Para a análise formal dos resíduos, podemos realizar os seguintes testes:

- Testes de Normalidade em que detalhes estão contidos no conteúdo de Inferência.
- Teste de Durbin-Watson para testar independência dos resíduos.
- Teste de Breusch-Pagan e Goldfeld-Quandt para testar se os resíduos são homoscedásticos.
- Teste de falta de ajuste para verificar se o modelo ajustado é realmente linear.





Teste de Durbin-Watson

O teste de Durbin-Watson verifica a presença de autocorrelação nos erros de um modelo de regressão.

A autocorrelação significa que os erros de observações adjacentes são correlacionados. Se os erros estiverem correlacionados, a regressão de mínimos quadrados pode subestimar o erro padrão dos coeficientes. Os erros padrão subestimados podem fazer com que seus preditores pareçam significativos quando eles não são.

Por exemplo, os erros de um modelo de regressão dos dados de preços diários de ações podem depender da observação anterior porque o preço das ações em um dia afeta o preço do dia seguinte.

O teste de Durbin-Watson (D) está condicionado à ordem das observações (linhas). Supõem-se que as observações estão em uma ordem significativa, como ordem de tempo. O teste de Durbin-Watson determina se ou não a correlação entre os termos de erro adjacente é zero.





Teste de Durbin-Watson

Para obter uma conclusão do teste, é possível comparar os valores exibidos para a estatística de Durbin-Watson com os limites superior e inferior corretos na tabela em anexo, elaborada por Savin e White¹. Se $D > DU$, não existe correlação; se $D < DL$, existe uma correlação positiva; se D estiver entre os dois limites, o teste é inconclusivo. A tabela fornece valores a serem testados quanto à autocorrelação positiva de primeira ordem. O nível de significância para o teste é 0,05. A tabela se destina a modelos com um intercepto.

Também é possível usar esta tabela para realizar testes quanto à autocorrelação negativa de primeira ordem. O teste estatístico é $(4 - D)$. Se $(4 - D) > DU$, não existe correlação; se $(4 - D) < DL$, existe uma correlação negativa; se $(4 - D)$ estiver entre os dois limites, o teste é inconclusivo.

¹ Savin, N. E. e White, K. J. (1977). *The Durbin-Watson test for serial correlation with extreme sample sizes or many regressors*. *Econometrica*, 45(8), p. 1989-1996.





REFERÊNCIAS BIBLIOGRÁFICAS

<http://www.bertolo.pro.br/MetodosQuantitativos/Previsao/pmc442.htm>

APOSTILA SOBRE MODELAGEM BOX AND JENKINS (Glaura C. Franco, Depto. Estatística – UFMG)

[https://www.investopedia.com/terms/b/box-jenkins-](https://www.investopedia.com/terms/b/box-jenkins-model.asp#:~:text=The%20Box%2DJenkins%20Model%20is,data%20points%20to%20determine%20outcomes.)

[model.asp#:~:text=The%20Box%2DJenkins%20Model%20is,data%20points%20to%20determine%20outcomes.](https://www.investopedia.com/terms/b/box-jenkins-model.asp#:~:text=The%20Box%2DJenkins%20Model%20is,data%20points%20to%20determine%20outcomes.)

<http://wilsonfreitas.github.io/estrategias-de-trading/pdf/unit-root-tests.pdf>

https://www.ufjf.br/wilson_rotatori/files/2011/05/Tendencias-e-Raizes-Unitarias-2018.pdf

[https://support.minitab.com/pt-br/minitab/18/help-and-how-to/modeling-](https://support.minitab.com/pt-br/minitab/18/help-and-how-to/modeling-statistics/regression/supporting-topics/model-assumptions/test-for-autocorrelation-by-using-the-durbin-watson-statistic/)

[statistics/regression/supporting-topics/model-assumptions/test-for-autocorrelation-by-using-the-durbin-watson-statistic/](https://support.minitab.com/pt-br/minitab/18/help-and-how-to/modeling-statistics/regression/supporting-topics/model-assumptions/test-for-autocorrelation-by-using-the-durbin-watson-statistic/)

REFERÊNCIAS BIBLIOGRÁFICAS

[http://www.icad.puc-](http://www.icad.puc-rio.br/cfeijo/pdf/revis%C3%A3o%20b%C3%A1sica%20s%C3%A9ries%20temporais_material%20de%20apoio_curso%20teoria%20macroeconomica_PPGE%20UFF.pdf)

[rio.br/cfeijo/pdf/revis%C3%A3o%20b%C3%A1sica%20s%C3%A9ries%20temporais_material%20de%20apoio_curso%20teoria%20macroeconomica_PPGE%20UFF.pdf](http://www.icad.puc-rio.br/cfeijo/pdf/revis%C3%A3o%20b%C3%A1sica%20s%C3%A9ries%20temporais_material%20de%20apoio_curso%20teoria%20macroeconomica_PPGE%20UFF.pdf)

<http://www.portalaction.com.br/series-temporais/142-teste-de-phillips-perron>

<http://www.portalaction.com.br/analise-de-regressao/analise-dos-residuos>

<https://people.bath.ac.uk/hssjrh/Phillips%20Perron.pdf>

<https://www.mathworks.com/help/econ/pptest.html>

<https://medium.com/wwblog/stationarity-testing-using-the-augmented-dickey-fuller-test-8cf7f34f4844>

<https://ibape-nacional.com.br/biblioteca/wp-content/uploads/2020/02/PE-26-Teste-de-Durbin-Watson.pdf>



REFERÊNCIAS BIBLIOGRÁFICAS

<https://estudar.com.vc/conceitos/ruído-branco-series-estacionarias-e-funcao-de-autocorrelacao-fac/8820-ruído-branco-grafico-e-correlograma>

<https://rpubs.com/hudsonchavs/ruarima>

http://jaguar.fcav.unesp.br/RME/fasciculos/v30/v30_n4/A3_Lais.pdf

<https://epge.fgv.br/we/MFEE/Econometria/2008?action=AttachFile&do=get&target=autocorr.pdf>



ANEXO – Tabela de resultados do teste de Durbin-Watson



Planilha do
Microsoft Excel



OBRIGADO!

