**Series temporais**

**Componentes de uma série temporal**

Uma série temporal pode ser genericamente decomposta em:

* **Tendência:** Capta comportamento de longo prazo relacionados com a série de tempo
* **Ciclo:** são caracterizados pelas oscilações de subida e de queda nas séries
* [**Sazonalidade**](https://pt.wikipedia.org/w/index.php?title=Sazonalidade&action=edit&redlink=1)**:** Capta os padrões regulares da série de tempo
* **Aleatório (ruido):** Capta todos os efeitos que não foram incorporados pela série de tempo via os três componentes anteriormente citados, ou seja, é o resíduo

**Série não estacionária**

Tem tendência e sazonalidade, por isso tentamos transformar a serie em não estacionário fazemos a decomposição da série

**Série estacionária**

Nosso objetivo é ter series estacionárias e elas possuem:

* Média constante
* Variância constante
* Auto covariância independente.

**Covariância**

A covariância é uma medida estatística onde é possível comparar duas variáveis, permitindo entender como elas se relacionam entre si. Já a **correlação** é similar mas varia de -1 a 1

Lag

Uma "**lag**" (atraso, em tradução livre) refere-se à observação de uma variável em um momento anterior no tempo. Em outras palavras, é o deslocamento ou defasagem de uma série temporal em relação a si mesma.

Por exemplo, se você está analisando a série temporal de vendas mensais de um produto e deseja calcular a autocorrelação com um atraso de 1, você estará comparando os dados de vendas de um mês com os dados do mês anterior. Isso pode ajudar a identificar padrões sazonais ou tendências que se repetem a cada mês.

**Métodos para tira tendência**

Valor\_sem\_tendencia = valor – media\_movel\_do\_valor

Valor\_sem\_tendencia = (valor X +1) – (valor X)

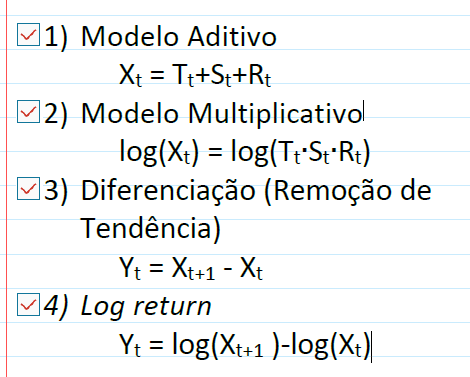
**ANALISE DESCRITIVA EM SERIES TEMPORAIS**

**Tipos de Decomposição de series temporais**

A decomposição de séries temporais é um processo analítico que desagrega uma série temporal em seus componentes subjacentes, geralmente em três componentes principais: tendência, sazonalidade e ruído.

1. Decomposição Clássica:

* Modelo Aditivo: Assume que a série temporal é a soma de tendência, sazonalidade e ruído.
* Modelo Multiplicativo: Assume que a série temporal é o produto desses três componentes.

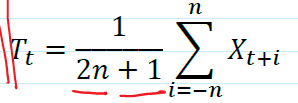


* A diferenciação é muito útil para remover tendência e podemos fazer dezXdez, JanXJan

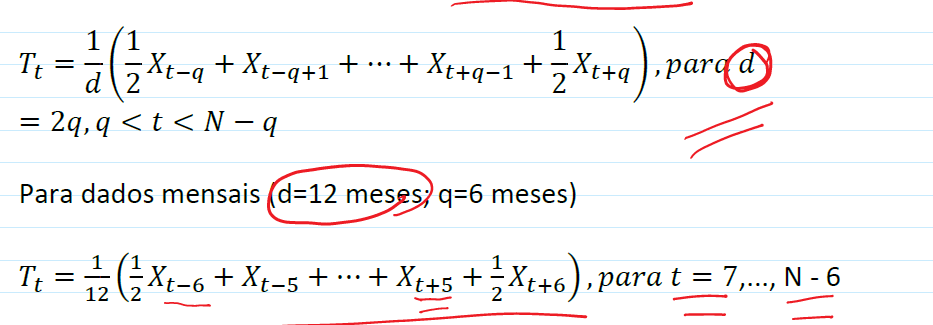
1. Médias Móveis:

* Média Móvel Simples: Calcula a média de um número fixo de observações consecutivas para suavizar a série e identificar a tendência.
* Média Móvel Exponencial: Atribui pesos exponencialmente decrescentes às observações passadas, dando mais importância às observações recentes

1 ) **Na ausência de efeitos sazonais**, a tendência de uma série temporal pode ser estimada pela média móvel



2 ) **No caso da série temporal apresentar tendência e efeito sazonal**, a tendência pode ser estimada pela média móvel *sazonal*(filtro):



1. Decomposição Estocástica:

* Utiliza técnicas estatísticas avançadas, como o uso de modelos autoregressivos e de médias móveis para decompor a série em tendência, sazonalidade e ruído.

1. Decomposição de Wavelets:

* Usa a análise de wavelets para decompor a série em diferentes escalas e frequências, permitindo uma representação mais detalhada dos componentes da série.

1. Decomposição de Componentes Independentes (ICA):

* Aplica o ICA para separar os componentes independentes da série temporal, como tendência, sazonalidade e ruído.

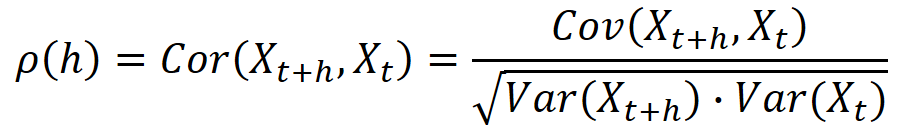
1. Decomposição de Espaço de Estado:

* Modela a série temporal como um sistema de espaço de estado, onde os estados ocultos representam os componentes subjacentes, que podem ser extraídos por meio de técnicas de filtragem de Kalman, por exemplo.

**Autocorrelação**

Autocorrelação porque é a correlação com a própria serie, se fosse outra serie seria crossrelação.

Uma importante característica de uma série temporal é a sua correlação serial. A autocorrelação é uma adimensional da associação linear entre duas variáveis aleatórias



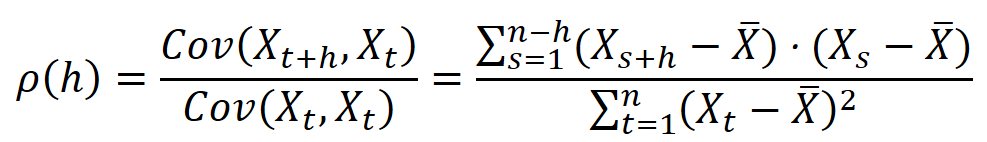
* A autocorrelação depende apenas do intervalo entre as medidas (h).
* Um valor de autocorrelação para um certo hé chamado de Coeficiente de Correlação (de Pearson).
* A autocorrelação varia entre -1 e 1 para uma perfeita correlação linear negativa e positiva, respectivamente.
* Uma autocorrelação igual à ZERO significa nenhuma correlação ou seja as variáveis são independentes!

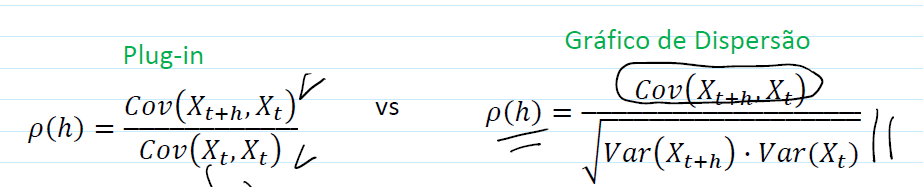
O quadrado do coeficiente de correlação (de Pearson) é chamado de **Coeficiente de Determinação. Assim, se autocorrelação (R) é -0,43, então -0,43^2 (R^2)= 0,189, que significa que Xt explica aproximadamente 18,9% da variabilidade observada em Xt+1.**

Problemas do Método do Gráfico de Dispersão

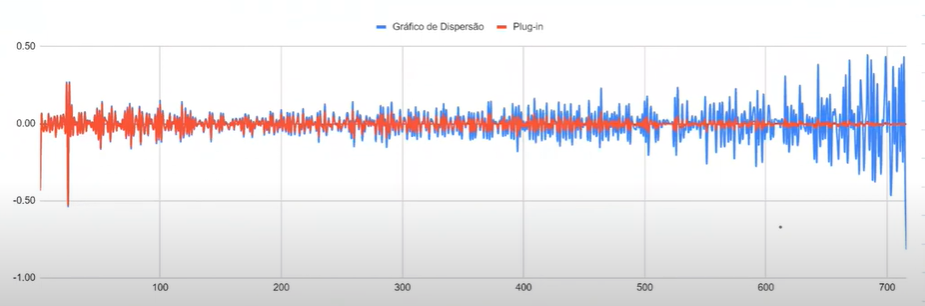
* Quando o intervalo entre as medidas (*h*) aumenta, menos pares (Xt+h, Xt) existirão.
* A variância de *ρ(h)*irá aumentar com um maior *h*, o que pode levar a resultados instáveis e a presença de artefatos.
* As correlações respectivas serão menos precisas.

Método alternativo: Método Plug-inpara a Função de Autocorrelação.





Quando comparamos os Plug-in com a autocorelação (gráfico dispersão) vemos que os artefactos (variação) aumenta qdo aumento h (tempo) já o plug-in rende a zero.



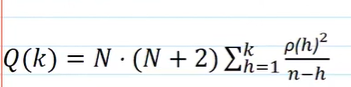
**Intervalo de confiança**

Para qualquer série temporal estacionária, os coeficientes da função de autocorrelação *ρ(h)*que caírem no intervalo de confiança ±1,96/ serão considerados ZEROcom 95% de probabilidade, enquanto que aqueles coeficientes que caírem fora do intervalo, serão considerados diferentes de ZERO



**Ljung-Box teste**

Este teste avalia se existe autocorrelação significantena série temporal. O teste estatístico é escrito como

 n é maiúsculo

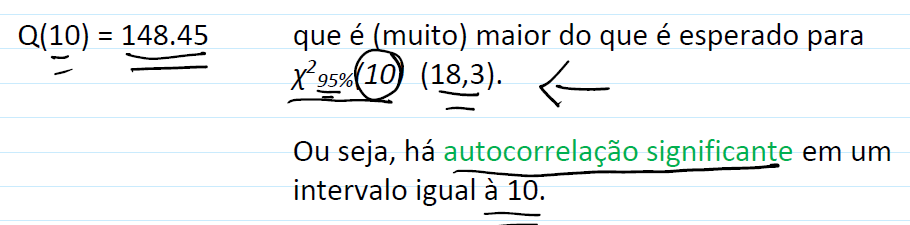
Em que, *N*é comprimento da série temporal, *ρ(h)*é o coeficiente de correlação em *h,*e *k*é o intervalo máximo que o teste é realizado.

Geralmente, utiliza-se *k* igual à 1, 3, 5, 10 ou 20 (aplica o teste para os K primeiros valores do coeficiente de correlação

O teste estatístico segue assintoticamenteuma distribuição Qui-quadrado *χ2*com *k*graus de liberdade.

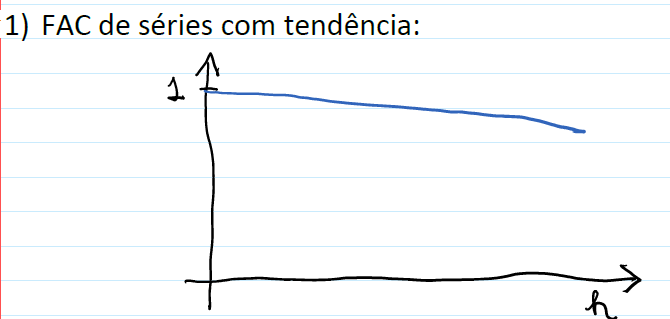
Ex:

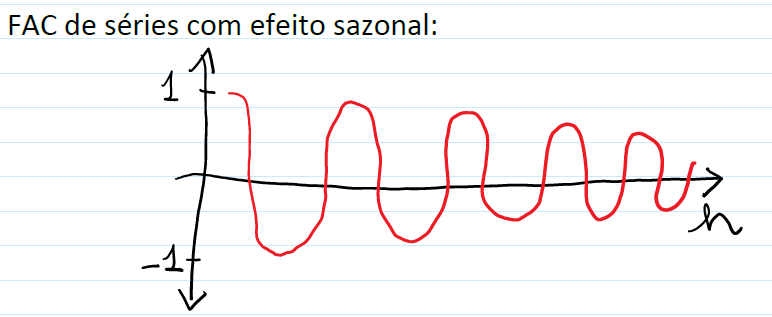


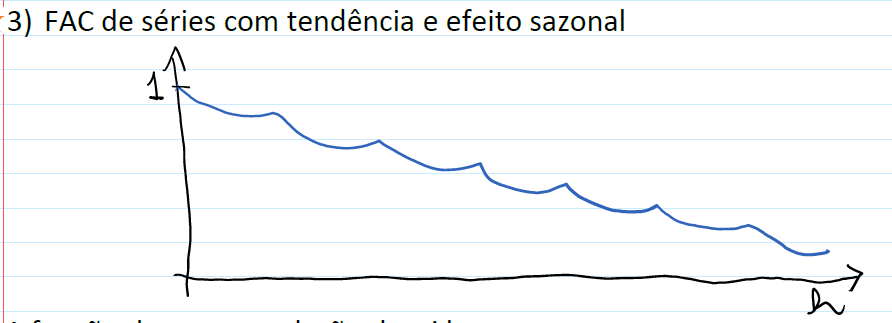


Q(10) é muito maior que 18.3. Se fosse menor não haveria auto correlação.

**Função de autocorrelação (FAC) de séries não estacionárias**







**Ruído branco**

Uma série de ruído branco é importante porque ela normalmente surge como uma série residual (ε), depois que um modelo da série temporal estacionária [f(∙)] é ajustado:

Xt= f(∙) + ε

Em que Xté uma série temporal estacionária

Definição

Uma série temporal (ε1, ε2, ..., εN) é considerada um ruído branco, se as variáveis aleatórias ε1, ε2, ... são **independentes e identicamente distribuídas (iid) com média zero**, que implica que:

1. Todas as variáveis aleatórias εtpossuem a mesma variância
2. Não há autocorrelação ou autocorrelação parcial significativa para qualquer passo *h*.
3. Todas as frequências estão em mesma proporção.

Caso as variáveis aleatórias ε1, ε2, ... também sigam uma distribuição Normal ou Gaussiana, i.e. εt≈N(0, σ2), a série é chamada de **ruído branco gaussiano.**

**Modelos Autoregressivos:**

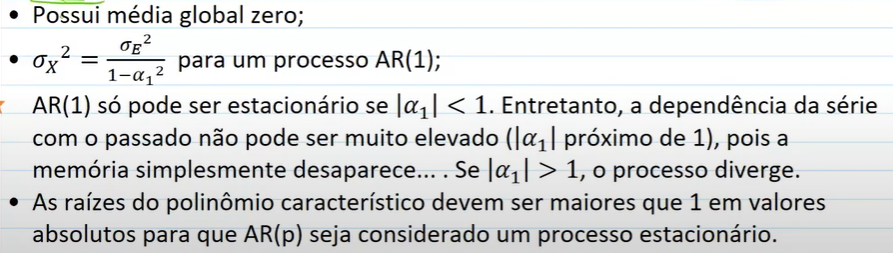
A formulação mais natural de um modelo de séries temporais estacionáriasé uma regressão linear dos valores passadosda própria série



Em que, o termo de pertubação Etvem de um processo de ruído branco (iid), sendo também (estocasticamente) independente de Xt-1, Xt-2... . Assim, Etpode ter o poder de "comandar" a série em uma nova direção (!).



**Atenção**: Modelos de AR(p) devem ser ajustados apenas para séries temporais estacionárias, i.e. qualquer tendência e efeitos sazonais potenciais precisam ser primeiramente removidos. Precisamos também certificar que o processo AR(p) é estacionário, ou seja



A maioria das séries temporais possuem uma média global diferente de zero (!). Neste caso

A maioria das séries temporais possuem uma média global diferente de zero (!). Neste caso, defini-se o seguinte processo:

Yt= m + Xt= m + α1Xt-1+ α2Xt-2+ ... + αpXt-p+ Et

Yt é um processo AR(p) deslocado em m, i.e. possui todas as propriedades de dependência de AR(p), mas possui uma média global diferente de zero.

De fato, quando se fala em modelo autoregressivo, a maioria das pessoas estão pensando em Yte não em Xt(!).

Existem várias formas para se ajustar um Modelo Autoregressivo, como pelo Método dos Mínimos Quadrados, pelo Algoritmo de Burg, pelas Equações de Yuler-Walkere pela Estimativa da Máxima Verossimilhança, mas não tem muita diferença no resultado desses modelos.

Invés do método em si, se preocupe com:

1) Outliers não são desejáveis!

2) Melhor performance com dados (aprox.) Gaussianos!

**Modelos de Média Móvel**

Modelos de média móvel são uma extensão do processo de ruído branco:

Xt= Et + β1Et-1+ β2Et-2+ ... + βqEt-q

Isto é, uma combinação linear da corrente pertubação Etmais as mais recentes Et-1, Et-2, ... Et-q.

Em muitos aspectos, os modelos de média móvel são complementares aos modelos autoregressivos.

**Modelos de Média Móvel VS Modelo autoregressivo**

Modelo autoregressivo: a função de autocorrelação tem um comportamento igual FAC parcial da média movel e que a FAC do modelo regressivo tem um comportamento igual a FAC da media movel

**O ajustamento (ou fitting) de modelos MA(q) a séries temporais é mais difícil que de modelos AR(p), porque não há estimadores explícitos (e eficientes) e, assim, é necessária otimização numérica (!).**

Na prática, dois métodos são utilizados para ajustar um modelo MA(q):

1. A Soma Condicional dos Quadrados (ou Conditional Sum of Squares) (CSS)   
 Um método numérico de minimização da soma dos quadrados dos resíduos

2. A Estimativa da Máxima Verossimilhança (MLE)

Baseado em que a pertubação Et segue uma distribuição normal (Xt= *m*+ Et + β1Et-1+ β2Et-2+ ... + βqEt-q);

- Para um melhor desempenho da MLE, os valores iniciais dos coeficientes são obtidos do CSS.

**Evite**uma quantidade massiva de extremos (*outliers*) ou dados extremamente assimétricos (!). Nesses casos, uma transformação-log é ideal antes de modelar

**Modelos ARMA**

Modelos ARMA são modelos que possuem uma estrutura de dependência tanto em relação às observações passadas Xt, Xt-1, ..., assim como em relação às pertubações Et, Et-1, ... .

Então, esses modelos são um híbrido de modelos Autoregressivos AR(p) e modelos de Média Móvel MA(q):

ARMA(p,q) = Xt= α1Xt-1+ ... + αpXt-p + Et+ β1Et-1+ ... + βqEt-q

com uma vantagem dupla:

(1) capacidade de simular uma quantidade muito maior de estruturas de dependência das séries temporais; e

(2) frenquentemente, um ARMA(p,q) necessita uma quantidade menor de parâmetros que um AR(p) ou um MA(q) sozinhos.

Na prática, iremos considerar um processo ARMA Xt (estacionário e média zero) deslocado em *m*, que é a média global da série temporal em análise:

Yt= m + Xt

AIC menor melhor é o modelo

Os modelos ARIMA E SARIMA são variações do ARMA

O modelo AR**I**MA incorpora a tendência determinística da série temporal:

A série Xt segue um modelo AR**I**MA(p, **d**, q), se a diferença **d** de Xt seguir um processo ARMA(p, q).

O modelo **S**AR**I**MA incorpora a tendência determinística e os efeitos sazonais da série temporal

Vantagem: conferem rapidez e conveniência na previsão de séries temporais!

**Previsão com series temporais**

Quais fatores podem ameaçar a confiabilidade de qualquer procedimento de previsão?

1. O processo governante, i.e. gerador dos dados observados, não pode mudar ao longo do tempo (é o mesmo do passado ao futuro);
2. Não temos garantia que o modelo ajustado é o correto
3. Além disso, há incertezas adicionais da estimativa dos parâmetros

Models

1. AR, MA e ARMA (séries estacionárias);
2. ARIMA e SARIMA (séries com tendência e/ou sazonalidade);
3. Séries não-estacionárias decompostas em tendência, sazonalidade e resíduo.
4. metodologia de Suavização Exponencial, que pode ser considerada um esquema intuitivo "model-free" para previsão de séries temporais

<https://www.youtube.com/watch?v=UFr0nLil1tE&list=PLSDVadsSlXTCVcg95hQsEOVRnVwgaPTRC&index=42>

# Repos for Databricks

<https://www.youtube.com/watch?v=x8Bu58GNPus>

<https://www.youtube.com/watch?v=Op8PbbRAQtw>

<https://github.com/databrickslabs/tempo/blob/master/Tempo%20QuickStart%20-%20Python.ipynb>

[https://www.databricks.com/notebooks/recitibikenycdraft/time-series.html#](https://www.databricks.com/notebooks/recitibikenycdraft/time-series.html)

<https://www.databricks.com/notebooks/#02_2_Data_Ingest.html>

<https://www.analisedeseriestemporais.com/post/aula-01-apresenta%C3%A7%C3%A3o>

<https://www.youtube.com/watch?v=21hH37HAkEg&list=RDCMUCm09WHp3LHA5Wlf2lH-Oisg&index=2>

<https://www.youtube.com/watch?v=RSOgBXMMrNw&list=PLNcS6_7VaJ3lWCs6cCXBAOAJVOItrACIY&index=2>

teste de normalidade

<https://www.statskingdom.com/shapiro-wilk-test-calculator.html>

<https://www.youtube.com/watch?v=KLsdmSMcyJk&list=PLSDVadsSlXTCVcg95hQsEOVRnVwgaPTRC&index=6>