

ANA RAQUEL

POSTECH

DATA ANALYTICS

ANÁLISE DE SÉRIES TEMPORAIS

AULA 04

SUMÁRIO

O QUE VEM POR AÍ?	3
CONHEÇA SOBRE O ASSUNTO	4
HANDS ON	10
O QUE VOCÊ VIU NESTA AULA?	11
REFERÊNCIAS.....	12
PALAVRAS-CHAVE	13

EMANIP

O QUE VEM POR AÍ?

Você já imaginou quais são as possíveis análises que podem ser feitas para compreender problemas de séries temporais?

Na aula de hoje, você irá aprender as principais técnicas para trabalhar com esse tipo de desafio. Vamos lá?

EMANO

CONHEÇA SOBRE O ASSUNTO

Para iniciar uma análise exploratória sobre dados de séries temporais, você deve realizar algumas perguntas para os dados, assim como em qualquer conjunto novo de dados:

- Existem correlações fortes entre algumas colunas?
- Qual é a média geral de uma variável relevante? Qual é a sua variância?
- Qual é o intervalo de valores que você visualiza nos dados? Os dados variam por período de tempo?

Para responder essas perguntas, você pode utilizar técnicas de estatística, tais como histogramas e gráficos de dispersão.

Histogramas

Em séries temporais, um histograma da diferença dos dados costuma ser mais interessante do que em histogramas não transformados, pois em análises de séries temporais (principalmente na área de finanças), o interessante é como o valor muda de uma direção para a próxima ao invés de mudar para a medição real desse mesmo valor.

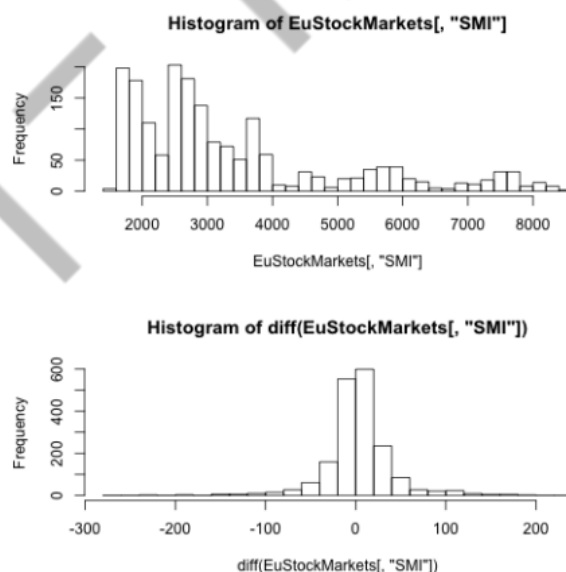


Figura 1 – Histograma

Fonte: Practical Time Series Analysis Prediction with Statistics and Machine (2021)

Gráficos de dispersão

Podemos utilizar gráficos de dispersão para determinar como duas ações estão vinculadas em um tempo específico e como suas mudanças de preço estão relacionadas ao longo do tempo.

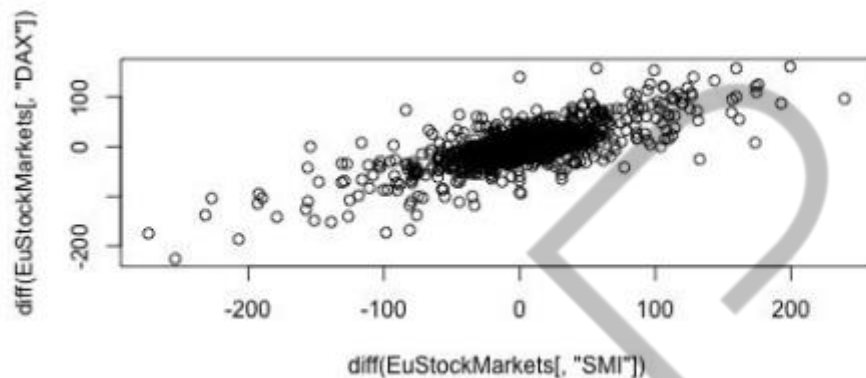


Figura 2 – Gráfico de dispersão

Fonte: Practical Time Series Analysis Prediction with Statistics and Machine (2021)

Métodos específicos para explorar dados de séries temporais

Como já vimos nas aulas de análise exploratória de dados, tanto o histograma quanto o gráfico de dispersão são utilizados para as análises de dados estruturados, e para séries temporais, mas quando se trata de séries temporais, podemos listar alguns conceitos específicos que podem ser muito úteis na análise exploratória em relação ao tempo. Vamos explorar alguns desses conceitos:

- **Estacionaridade:** uma série temporal que possui propriedades estatísticas razoavelmente estáveis ao longo do tempo, sobretudo no que se diz a respeito à média e variância dos dados.

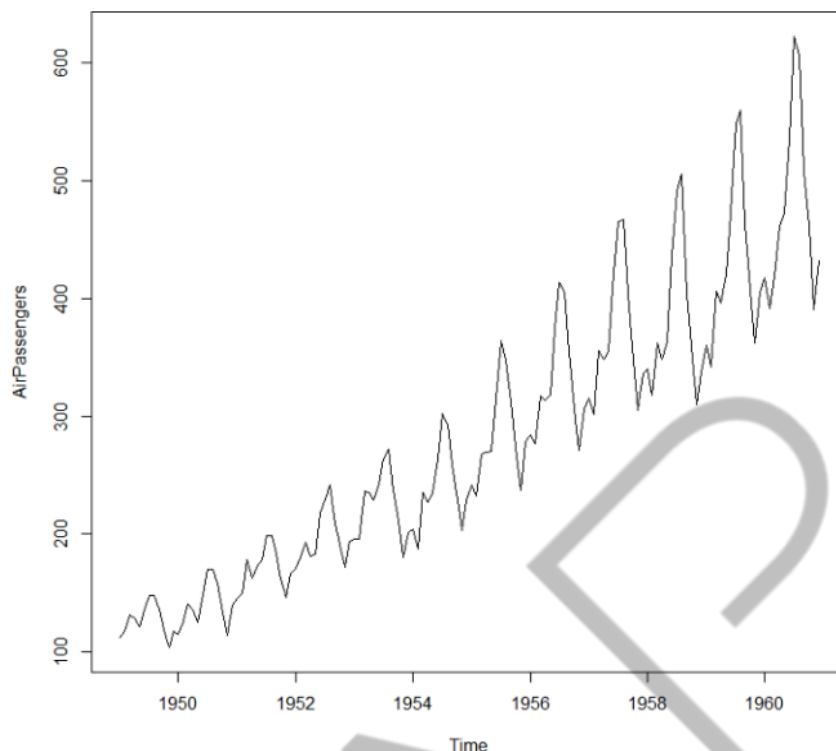


Figura 3 – Estacionaridade

Fonte: Practical Time Series Analysis Prediction with Statistics and Machine (2021)

Como aprendemos na aula 1, às vezes pode ser difícil definir a estacionariedade dos dados. Para validar o conceito de estacionariedade dos dados de uma série temporal, podemos utilizar testes estatísticos, como, por exemplo, o teste de Dickey-Fuller aumentado.

O teste de Dickey-Fuller aumentado, também conhecido como ADF, é um tipo de teste de hipótese para validar uma série temporal quando se trata de problemas de estacionariedade. Esse teste postula uma hipótese nula de que uma raiz unitária está presente em uma série temporal. Dependendo do resultado do teste, essa hipótese nula pode ser rejeitada.

Pensando ainda em solucionar a estacionariedade dos dados (o que pode ser muito difícil), podemos partir para algumas outras técnicas, como transformações sobre os dados.

- **Transformações logarítmicas e transformações com raízes:** podemos fazer com que uma série temporal possa se tornar estacionária o suficiente realizando esses tipos de transformações na escala dos dados. A remoção de uma tendência é normalmente feita por diferenciação.
- **Janela rolante (Rolling Windows):** qualquer tipo de função em que você agrega dados para compactá-los ou para suavizá-los pode ser considerado uma técnica de janela rolante. Os dados suavizados ou agregados por janelas de tempo proporcionam visualizações exploratórias informativas. Podemos calcular a média móvel e fazer outros cálculos de função linear para detectar tendências ou mudanças significativas na média ao longo do tempo. Por exemplo, em vez de calcular a média para todo o conjunto de dados, podemos usar uma técnica de Rolling Windows para calcular a média para cada janela de 30 dias consecutivos ao longo da série temporal.

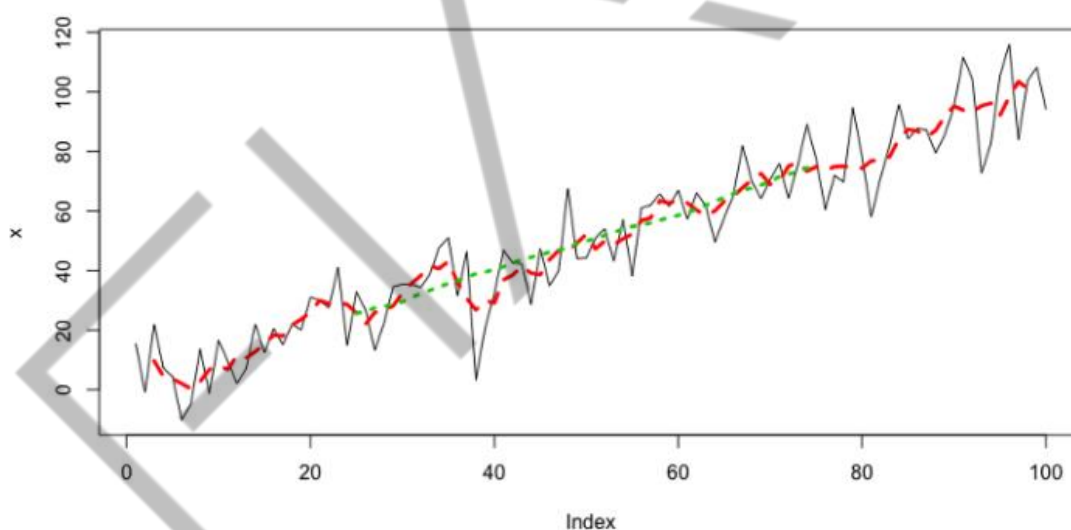


Figura 4 – Gráfico 1

Fonte: Practical Time Series Analysis Prediction with Statistics and Machine (2021)

Self-correlation: podemos definir esse termo à essência de que um valor em uma série temporal, em um determinado ponto no tempo, pode estar correlacionado com outro valor em outro ponto no tempo.

Função de autocorrelação: a autocorrelação dá uma ideia de como os pontos de dados em diferentes pontos no tempo estão linearmente relacionados entre si em função de sua diferença de tempo. Na função ACF (Função de autocorrelação), podemos identificar a correlação dos dados, realizando uma análise sobre o gráfico da figura 5 – Gráfico 2.

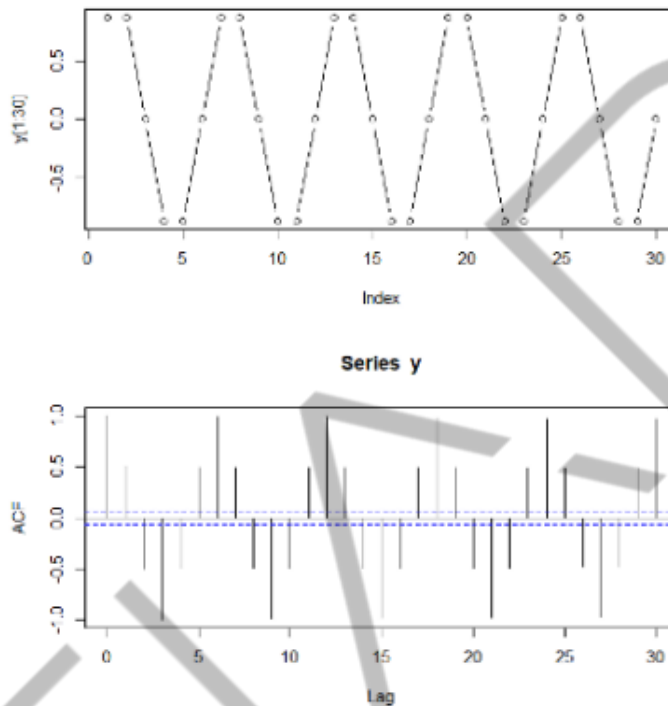


Figura 5 – Gráfico 2

Fonte: Practical Time Series Analysis Prediction with Statistics and Machine (2021)

O gráfico de autocorrelação é um gráfico de barras que mostra o valor da ACF em diferentes defasagens dos dados sobre o tempo (atrasos). A defasagem zero (lag 0) é a correlação da série consigo mesma (correlação 1 no gráfico). O número de intervalos entre duas medições é conhecido como lag. As defasagens subsequentes medem a correlação entre a série e seus valores passados em incrementos de tempo cada vez maiores. Os valores fora da banda de confiança são considerados estatisticamente significativos e indicam que a correlação é provavelmente real e não resultada do acaso, como podemos analisar com o gráfico da figura 6 – ACF.

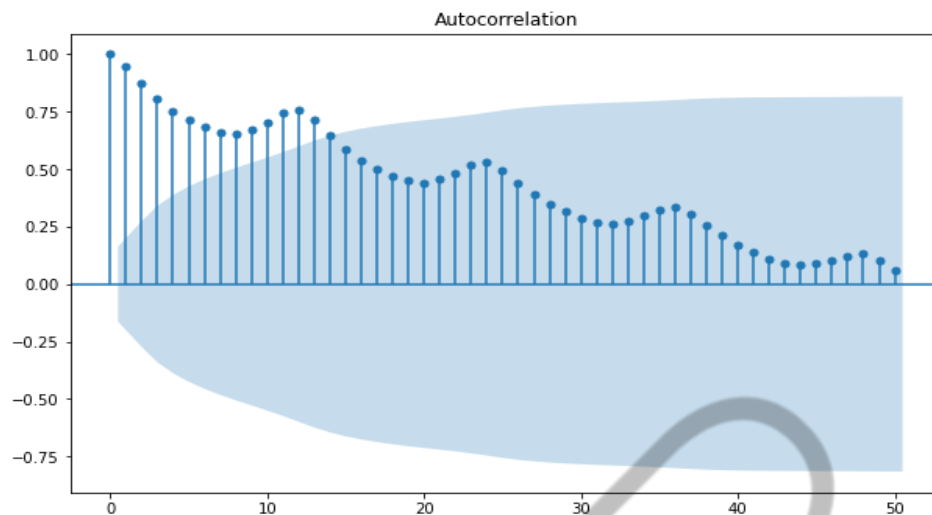


Figura 6 - ACF
Fonte: Elaborado pela autora (2023)

A ACF é simétrica em relação aos lags negativos e positivos, assim, apenas os lags positivos precisam ser considerados explicitamente.

- **Autocorrelação parcial:** A autocorrelação parcial (PACF) de uma série temporal para determinado lag é a correlação parcial desta série temporal com ela mesma nesse lag, dadas as informações entre os dois tempos.

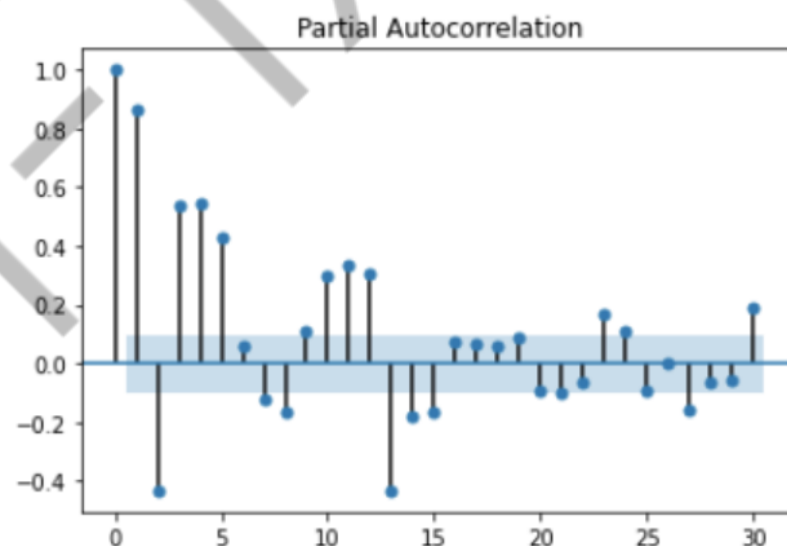


Figura 7 – Autocorrelação parcial.
Fonte: <https://analyticsindiamag.com> (s.d.)

HANDS ON

Agora chegou o momento de ver, na prática, como realizar análise exploratória sobre dados de séries temporais.

EMAP

O QUE VOCÊ VIU NESTA AULA?

Aprendemos a como realizar análise exploratória com séries temporais.

Daqui em diante, é importante que você replique os conhecimentos adquiridos para fortalecer mais suas bases e conhecimentos.

IMPORTANTE: não esqueça de praticar com o desafio da disciplina, para que assim você possa aprimorar os seus conhecimentos!

Você não está sozinho(a) nesta jornada! Te esperamos no Discord e nas *lives* com os(as) professores(as) especialistas, onde você poderá tirar dúvidas, compartilhar conhecimentos e estabelecer conexões!

REFERÊNCIAS

LENDAVE, V. **What are autocorrelation and partial autocorrelation in time series.** 2022. Disponível em: <<https://analyticsindiamag.com/what-are-autocorrelation-and-partial-autocorrelation-in-time-series-data/>>. Acesso em: 09 mai 2023.

NIELSEN, Aileen. **Análise Prática de Séries Temporais:** Predição com Estatística e Aprendizado de Máquina. [s.l.]: O'Reilly Media, Inc., 2021.

PALAVRAS-CHAVE

Palavras-Chave: Acf, Pacf, Rolling Windows, Estacionaridade.

EMSE

The background is a dark blue field filled with numerous small, light blue dots. Overlaid on this are several large, wavy, translucent lines in shades of blue and yellow. A vertical line with a small 'x' at the bottom is on the left. A circle containing the number '7' is in the upper center. A hexagon is in the lower right.

POSTECH