

# ANÁLISE ESTATÍSTICA E MODELAGEM PREDITIVA DE SÉRIES TEMPORAIS EM PYTHON (DSA)

## Aula 2 – O Que São Componentes de Séries Temporais?

Agora que você já sabe o que são séries temporais, vamos compreender como elas são formadas, definindo seus componentes.

Qualquer série temporal pode conter alguns ou todos os seguintes componentes:

1. Tendência (T)
2. Ciclo (C)
3. Sazonalidade (S)
4. Irregularidade ou Ruído (I)

Vamos descrever cada um dos componentes e a imagem desta aula resume bem o que é cada um deles.

### Componente Sazonal (Seasonality)

A sazonalidade ocorre quando a série temporal exhibe flutuações regulares durante o mesmo mês (ou meses), todos os anos ou durante o mesmo trimestre de cada ano.

Muitos dados de série temporal exibem uma variação sazonal. Esse tipo de variação é fácil de entender e pode ser facilmente medido ou removido dos dados para fornecer dados não sazonais.

Flutuações sazonais descrevem qualquer variação regular (flutuação) com um período de menos de um ano. Por exemplo: custo de vários tipos de frutas e legumes, custo de roupas, números de desemprego, precipitação média diária, aumento da venda de sorvete no verão, etc... mostram variações sazonais.

As mudanças que se repetem em um período fixo também são chamadas de variações sazonais, por exemplo, tráfego nas estradas nas horas da manhã e da noite, vendas em festivais como Rock in Rio, aumento do número de passageiros no fim de semana, etc. Variações sazonais são causadas pelo clima, costumes sociais, atividades religiosas, atividades de negócio, etc.

### Componente de Ciclo (Cycle)

Qualquer padrão mostrando um movimento para cima e para baixo em torno de uma determinada tendência é identificado como um padrão cíclico. A duração de um ciclo depende do tipo de negócio ou setor que está sendo analisado.

As séries temporais exibem variações cíclicas em um período fixo devido a alguma outra causa física, como variação diária de temperatura. A variação cíclica é um componente não sazonal que varia em um ciclo reconhecível.

Às vezes, as séries exibem oscilação que não tem um período fixo, mas é previsível até certo ponto. Por exemplo, dados econômicos afetados por ciclos de negócios com um período que varia entre 5 e 7 anos. Nos dados semanais ou mensais, o componente cíclico pode descrever qualquer variação regular (flutuações) nos dados de séries temporais. A variação cíclica é de natureza periódica e se repete como um ciclo comercial, que possui quatro fases: Pico, Recessão, Depressão e Expansão.

### Componente de Tendência (Trend)

A tendência é o padrão de longo prazo de uma série temporal. Uma tendência pode ser positiva ou negativa, dependendo se a série temporal exibe um padrão crescente de longo prazo ou um padrão decrescente de longo prazo.

Se uma série temporal não mostrar um padrão crescente ou decrescente, a série será estacionária (falaremos sobre isso mais tarde).

Aqui, levamos em conta o número de observações disponíveis e fazemos uma avaliação subjetiva do que é de longo prazo.

Para entender o significado de longo prazo temos, por exemplo, variáveis climáticas que às vezes exibem variação cíclica por um período muito longo, como 50 anos. Se alguém tivesse apenas 20 anos de dados, essa oscilação de longo prazo pareceria uma tendência, mas se várias centenas de anos de dados estiverem disponíveis, as oscilações de longo prazo seriam visíveis. Esses movimentos são de natureza sistemática, onde os movimentos são amplos, constantes, mostrando um lento aumento ou queda na mesma direção.

A tendência pode ser linear ou não linear (curvilínea). Alguns exemplos das tendências seculares são: aumento dos preços, aumento da poluição, aumento da necessidade de trigo, aumento da taxa de alfabetização, diminuição das mortes devido aos avanços da ciência, etc...

Calcular médias em um determinado período é uma maneira simples de detectar uma tendência nos dados sazonais. A mudança nas médias com o tempo é evidência de uma tendência nas séries dadas, embora existam testes mais formais para detectar uma tendência nas séries temporais.

### Componente Irregular ou Ruído (Noise)

Este componente é imprevisível. Toda série temporal tem algum componente imprevisível que a torna uma variável aleatória. Na previsão, o objetivo é "modelar" todos os componentes até o ponto em que o único componente que permanece inexplicável é o componente aleatório.

Quando tendências e variações cíclicas são removidas de um conjunto de dados de séries temporais, o restante residual pode ou não ser aleatório.

Várias técnicas para analisar séries desse tipo examinam para ver “se a variação irregular pode ser explicada em termos de modelos de probabilidade, como modelos de média móvel (Moving Average) ou autoregressivo (Autoregressive), ou seja, podemos ver se ainda resta alguma variação cíclica nos resíduos. Essas variações ocorrem devido a causas repentinas, chamadas de variação residual (variação irregular ou flutuações acidentais ou erráticas) e são imprevisíveis. Por exemplo: aumento dos preços do aço devido a greve na fábrica, acidente por falha no rompimento de barragem, guerra, etc.

No curso Machine Learning e IA em Ambientes Distribuídos (FED) mostramos como aplicar Moving Average com o Apache Spark.

Conforme veremos mais adiante, a podemos decompor a série temporal de acordo com seus componentes, para então realizar a análise e criar o modelo preditivo.

As principais características de muitas séries temporais são tendências e variações sazonais. Outra característica importante da maioria das séries temporais é que as observações próximas no tempo tendem a ser correlacionadas (serialmente dependentes).

Vamos fazer uma analogia, para você compreender porque os componentes são importantes:

Você lembra a fórmula matemática de um modelo de regressão linear simples (isso é mostrado no curso gratuito de Python Fundamentos)? Não lembra? Então aqui está de forma simplificada:

$$y = a + bx$$

Onde:

y = Variável de saída (também chamada de variável alvo, target ou dependente). É o que queremos prever.

x = Variável de entrada (também chamada de variável explicativa, variável preditora ou variável independente). É o que ajuda a explicar a variável de saída.

a e b = São os coeficientes do modelo.

Em um exemplo simples, x seria o tamanho de uma casa e y o preço de uma casa. Com base no tamanho, podemos tentar prever o preço.

Sendo assim, como treinamos o modelo? Coletamos dados históricos de x e y, alimentamos o modelo, fazemos o treinamento e ele aprende os melhores valores de a e b. Isso é o modelo treinado (e isso vale para qualquer modelo de Machine Learning, apenas aumentando a complexidade entre algoritmos).

Com o modelo treinado (ou seja, com os coeficientes a e b), entregamos ao modelo novos

valores de  $x$  e ele prevê  $y$ . Isso não deveria ser novidade para você, mas de qualquer forma, fizemos uma revisão.

Vale ressaltar que o modelo é uma estimativa. A relação entre  $x$  e  $y$  pode ser explicada por uma função matemática, geralmente indicada por  $f(x)$ . A questão é que NUNCA saberemos qual é o valor exato de  $f(x)$ , (afinal, se soubéssemos, não precisaríamos de Machine Learning). O que temos é  $g(x)$ , ou seja, uma hipótese, uma aproximação da  $f(x)$ , o que nos atende muito bem. Talvez a famosa frase abaixo faça mais sentido agora:

“Essencialmente todos os modelos estão errados, mas alguns são úteis.”

A frase é atribuída a George E. P. Box, Estatístico Britânico.

[https://en.wikipedia.org/wiki/George\\_E.\\_P.\\_Box](https://en.wikipedia.org/wiki/George_E._P._Box)

Depois dessa revisão sobre regressão linear, voltemos às séries temporais.

E como poderíamos fazer previsões em séries temporais?

Os componentes de uma série temporal se combinam para fornecer as séries temporais observadas. E, embora esses componentes possam ser combinados de diferentes maneiras, esta fórmula resume bem a relação entre eles:

$$y_t = T \times C \times S \times I$$
$$y_t = T + C + S + I$$

Onde:

$y_t$  = Valor da variável aleatória  $y$  no tempo  $t$

$T$  = Tendência

$C$  = Ciclo

$S$  = Sazonalidade

$I$  = Irregularidade ou Ruído

É possível fazer suposições sobre esses componentes, tanto no comportamento quanto na maneira como eles são combinados, o que permite que eles sejam modelados usando métodos estatísticos tradicionais.

Esses componentes também podem ser a maneira mais eficaz de fazer previsões sobre valores futuros, mas nem sempre.

Nos casos em que esses métodos clássicos não resultam em desempenho efetivo, esses componentes ainda podem ser conceitos úteis e até contribuir para métodos alternativos.

E o que torna a série temporal especial?

Como o nome sugere, uma série temporal é uma coleção de pontos de dados coletados

em intervalos de tempo, que são analisadas para determinar a tendência de longo prazo, a fim de prever o futuro ou realizar alguma outra forma de análise. Mas o que diferencia uma série temporal de um problema regular de regressão? Existem 2 fatores:

1-Depende do tempo. Portanto, a suposição básica de um modelo de regressão linear de que as observações são independentes não se aplica em séries temporais.

2-Juntamente com uma tendência crescente ou decrescente, a maioria das séries temporais tem algum tipo de tendência de sazonalidade, ou seja, variações específicas para um período de tempo específico. Por exemplo, se você analisar as vendas de sorvete ao longo do tempo, encontrará invariavelmente vendas mais altas no verão.

Devido às propriedades inerentes a uma série temporal, há várias etapas envolvidas no processo de análise.

Achou que ia ser moleza por se tratar de um “cursinho via texto”? Então reavalie seus conceitos. 😊

E ainda estou aquecendo os motores. Na Aula 3 tem mais!

