Unidade 1 – Conceitos Básicos de Séries Temporais

Uma série temporal é uma sequência de dados pontos coletados ou registrados ao longo do tempo, geralmente em intervalos uniformes. Cada ponto na série representa o valor de uma variável medida em um momento específico. Esses dados são usados para analisar como a variável evolui ao longo do tempo, identificar padrões e fazer previsões futuras. Séries temporais são amplamente aplicadas em diversas áreas, como economia, finanças, climatologia, entre outras.

Características Principais de Séries Temporais

1. **Sequência Cronológica**:
   - ✓ Os dados são organizados em ordem temporal, sendo essencial que a sequência dos pontos respeite a cronologia dos eventos.
2. **Intervalos de Tempo Uniformes**:
   - ✓ As observações são coletadas em intervalos regulares, como diariamente, semanalmente, mensalmente etc.
3. **Dependência Temporal**:
   - ✓ Os valores da série temporal podem ser autocorrelacionados, ou seja, o valor em um ponto no tempo pode depender dos valores em pontos anteriores.

Objetivos da Análise de Séries Temporais

1. **Descrever Dados**: Identificar e descrever padrões de tendência, sazonalidade e ciclos.
2. **Previsão**: Usar dados históricos para prever valores futuros.
3. **Detecção de Anomalias**: Identificar comportamentos inesperados ou fora do padrão.
4. **Controle e Ajuste**: Monitorar e ajustar processos com base em dados temporais.

Componentes das Séries Temporais

1. **Tendência (Trend)**:
   - ✓ Movimento de longo prazo na série que mostra uma direção geral, seja crescente, decrescente ou estável.

2. **Sazonalidade (Seasonality)**:
    - ✓ Padrões ou flutuações que se repetem em intervalos regulares, como ciclos diários, mensais ou anuais.
3. **Ciclos (Cycles)**:
    - ✓ Flutuações de longo prazo causadas por fatores econômicos ou outros, que não possuem uma periodicidade fixa.
4. **Ruído (Noise)**:
    - ✓ Variabilidade aleatória que não pode ser explicada por tendência, sazonalidade ou ciclos. Representa a parte imprevisível dos dados.

Exemplos de Séries Temporais

1. **Econômicas**: Preços de ações, taxa de desemprego, PIB trimestral, taxas de câmbio
2. **Financeiras**: Receita mensal de uma empresa, vendas diárias de um produto.
3. **Climatológicas**: Temperatura diária, níveis de precipitação mensal.
4. **Demográficas**: População anual, taxas de natalidade e mortalidade.
5. **Vendas**: Vendas diárias ou mensais de produtos.
6. **Tráfego**: Número de visitantes em um site ao longo do tempo

Métodos de Análise de Séries Temporais

1. **Modelos ARIMA (Autoregressive Integrated Moving Average)**:
    - ✓ Usados para modelar e prever séries temporais, especialmente úteis para séries não estacionárias.
2. **Decomposição**:
    - ✓ Técnica para separar uma série temporal nos seus componentes (tendência, sazonalidade e ruído).
3. **Suavização Exponencial**:
    - ✓ Métodos como Holt-Winters que capturam tanto a tendência quanto a sazonalidade nos dados.
4. **Análise de Fourier**:
    - ✓ Utilizada para identificar frequências dominantes e padrões periódicos na série.

Aplicações de Séries Temporais

- **Previsão de Vendas**: Utilizar dados históricos para prever futuras vendas e ajustar a produção.
- **Análise Financeira**: Avaliar tendências de mercado e tomar decisões de investimento.
- **Meteorologia**: Prever condições climáticas com base em dados históricos.
- **Planejamento de Recursos**: Estimar a demanda futura para otimizar a alocação de recursos.

Ferramentas e Softwares

- **R**: Pacotes como forecast, tseries, e zoo são amplamente utilizados.
- **Python**: Bibliotecas como pandas, statsmodels, prophet e matplotlib.
- **Excel**: Funções de análise de dados e gráficos.

Estacionariedade

A estacionariedade é um conceito fundamental em séries temporais, referindo-se a uma série cujas propriedades estatísticas, como a média e a variância, são constantes ao longo do tempo. Analisar e garantir a estacionariedade é crucial porque muitos métodos de análise e previsão de séries temporais, como os modelos ARIMA, assumem que os dados são estacionários.

Características da Estacionariedade

Uma série temporal é considerada estacionária se satisfizer as seguintes condições:

1. **Média Constante**:
   - ✓ A média da série não muda ao longo do tempo.
   - ✓ $E[X_t] = \mu$ para todo t (o valor médio esperado para a variável em estudo é constante para todo o tempo)
2. **Variância Constante**:
   - ✓ A variabilidade dos dados em torno da média é constante ao longo do tempo.
   - ✓ $Var(X_t) = \sigma^2$ para todo t

3. **Covariância Constante**:
    - ✓ A covariância entre dois pontos na série depende apenas do intervalo entre eles, não do tempo específico em que são medidos.
    - ✓ $Cov(X_t, X_{t+h}) = \gamma(h)$ para todo t

Tipos de Estacionariedade

1. **Estacionariedade Estrita**:
    - ✓ Uma série é estritamente estacionária se a distribuição conjunta de qualquer subconjunto de valores da série é a mesma, independentemente de quando a amostra é tomada.
2. **Estacionariedade Fraca**:
    - ✓ Uma série é estacionária em sentido amplo se a média, a variância e a covariância são constantes ao longo do tempo. Este é o tipo de estacionariedade mais frequentemente utilizado na prática.

Testes para Estacionariedade

Existem vários testes estatísticos para verificar a estacionariedade de uma série temporal:

1. **Teste de Dickey-Fuller Aumentado (ADF)**:
    - ✓ Testa a presença de uma raiz unitária na série temporal. A hipótese nula do teste é que a série tem uma raiz unitária (não é estacionária).
2. **Teste KPSS (Kwiatkowski-Phillips-Schmidt-Shin)**:
    - ✓ Testa a hipótese nula de que a série é estacionária em torno de uma média ou tendência linear.
3. **Teste de Phillips-Perron**:
    - ✓ Semelhante ao ADF, testa a presença de uma raiz unitária, mas ajusta para a autocorrelação e heterocedasticidade nos resíduos.

Transformações para Estacionariedade

Se uma série não é estacionária, várias transformações podem ser aplicadas para torná-la estacionária:

1. **Diferença (Differencing)**:
   - ✓ Subtrair cada valor da série pelo valor anterior. Esta é a técnica mais comum para remover tendência.
   - ✓ $Y_t = X_t - X_{t-1}$
2. **Transformação Logarítmica**:
   - ✓ Usada para estabilizar a variância.
   - ✓ $Y_t = \log(X_t)$
3. **Suavização (Smoothing)**:
   - ✓ Técnicas como médias móveis podem ser aplicadas para suavizar flutuações de curto prazo.

Exemplos

- **Exemplo Não Estacionário**: Uma série temporal que apresenta uma tendência ascendente ou descendente ao longo do tempo, como o PIB de um país ao longo de décadas.
- **Exemplo Estacionário**: Dados de temperatura diários que flutuam em torno de uma média constante sem uma tendência de longo prazo.

Importância da Estacionariedade

- **Modelagem e Previsão**: Muitos modelos estatísticos, como ARIMA, requerem dados estacionários para produzir previsões precisas.
- **Análise de Componentes**: Facilita a separação dos componentes da série temporal, como tendência e sazonalidade.

<u>Função de Autocorrelação</u>

A função de autocorrelação é uma ferramenta estatística usada para medir o grau de correlação entre valores de uma série temporal em diferentes pontos no tempo. Ela é essencial para identificar a dependência temporal nos dados, o que é fundamental para a modelagem e previsão de séries temporais.

Definição

A autocorrelação de uma série temporal X_t em uma defasagem h (também chamada de lag) é a correlação entre X_t e X_{t+h}. A função de autocorrelação (FAC), denotada por $\rho(h)$ é definida como:

$$\rho(h) = \frac{E[(X_t - \mu)(X_{t+h} - \mu)]}{\sigma^2}$$

em que: μ é a média da série temporal; σ^2 é a variância da série temporal; E denota o valor esperado.

Interpretação

- **Autocorrelação Positiva**: Indica que altos valores em X_t tendem a ser seguidos por altos valores em X_{t+h}, e baixos valores tendem a ser seguidos por baixos valores.
- **Autocorrelação Negativa**: Indica que altos valores em X_t tendem a ser seguidos por baixos valores em X_{t+h}, e vice-versa.
- **Autocorrelação Zero**: Indica que não há dependência <u>linear</u> entre X_t e X_{t+h}.

Função de Autocorrelação Parcial (FACP)

A função de autocorrelação parcial mede a correlação entre X_t e X_{t+h} removendo o efeito das autocorrelações para lags menores que h. É útil para identificar o número adequado de lags a serem incluídos em modelos autoregressivos (AR).

Gráfico da Função de Autocorrelação (ACF)

O gráfico da função de autocorrelação, ou correlograma, é uma ferramenta visual que exibe a autocorrelação $\hat{\rho}(h)$ em diferentes defasagens h. Ele ajuda a identificar padrões, sazonalidades e o grau de dependência temporal.

Exemplos de Uso

- **Identificação de Modelos ARIMA**: A ACF e a PACF são usadas para determinar os parâmetros de ordem dos modelos ARIMA.
- **Detecção de Sazonalidade**: Padrões repetitivos na ACF podem indicar a presença de sazonalidade nos dados.
- **Validação de Modelos**: Analisar os resíduos de um modelo ajustado para verificar se eles são ruído branco (sem autocorrelação significativa).

Ferramentas para Cálculo e Visualização

- **Python**: A biblioteca statsmodels oferece funções como acf e pacf para calcular e plotar a função de autocorrelação.
- **R**: O pacote forecast fornece funções como Acf e Pacf para análise de autocorrelação.
- **Excel**: Análise de dados pode ser feita usando funções incorporadas e add-ins específicos para séries temporais.

Modelos no Domínio do Tempo e da Frequência

A análise de séries temporais pode ser realizada tanto no domínio do tempo quanto no domínio da frequência. Cada abordagem oferece insights diferentes e complementares sobre os dados.

Modelos no Domínio do Tempo

Os modelos no domínio do tempo analisam a relação entre os valores da série temporal ao longo do tempo. Os principais modelos incluem:

1. **Modelos AR (Autoregressive)**:

Baseiam-se na suposição de que os valores passados da série têm uma influência linear nos valores futuros.

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t$$

em que: c é uma constante, ϕ_i são os coeficientes autoregressivos, p é a ordem do modelo e ε_t é o termo de erro (ruído branco).

2. **Modelos MA (Moving Average)**:

Modelam a série como uma função linear de erros passados.

$$Y_t = \mu + \varepsilon_t + \sum_{j=1}^{q} \theta_i \varepsilon_{t-j}$$

em que: θ_i são os coeficientes de média móvel, μ é a média da série (quando a série é estacionária), q é a ordem do modelo; ε_t é o termo de erro.

3. **Modelos ARMA (Autoregressive Moving Average)**:

E a combinação das abordagens AR e MA.

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_i \varepsilon_{t-j}$$

4. **Modelos ARIMA (Autoregressive Integrated Moving Average)**:

Extensão dos modelos ARMA para séries não estacionárias, incluindo uma componente de diferenciação

$$(1 - B)^d Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_i \varepsilon_{t-j}$$

em que: $(1-B)^d Y_t$ representa a série temporal Y_t diferenciada d vezes; c é o termo constante, ϕ_i são os coeficientes da parte autoregressiva (AR); θ_i são os coeficientes da parte de média móvel (MA), ε_t são os termos de erro branco.

5. **Modelos SARIMA (Seasonal ARIMA)**:

Extensão dos modelos ARIMA para capturar sazonalidades

$$\left(1-\sum_{i=1}^{p}\phi_i B^i\right)\left(1-\sum_{j=1}^{p}\Phi_j B^{js}\right)(1-B)^d(1-B^s)^D Y_t = c+(1-\sum_{i=1}^{q}\theta_i B^i)(1-\sum_{k=1}^{Q}\Theta_k B^s)\varepsilon_t$$

em que: ϕ_i coeficientes autorregressivos não sazonais; Φ_j são os coeficientes autoregressivos sazonais; θ_i são os coeficientes de média móvel não sazonais; Θ_k são os coeficientes de média móvel sazonais; B operador de defasagem; d ordem de diferenciação não sazonal; D ordem de diferenciação sazonal; s período sazonal, ε_t termos de erro branco (ruído).

6. **Modelos de Alisamento Exponencial**:

Aplicam pesos decrescentes exponencialmente para dados passados. Esses modelos são úteis para fazer previsões de curto prazo, especialmente quando os dados apresentam flutuações sem grandes mudanças estruturais.

Método de Holt-Winters:

⇒ **Alisamento Simples**: para séries sem tendência ou sazonalidade.

$$F_{t+1} = \alpha y_t + (1-\alpha)F_t$$

em que: F_{t+1} previsão para o próximo período; y_t valor observado no período atual; F_t previsão no período atual; α fator de suavização

⇒ **Alisamento com Tendência (Holt)**: para séries com tendência

$$F_{t+h} = L_t + hT_t$$

$$L_t = \alpha y_t + (1-\alpha)(L_{t-1} + T_{t-1})$$

$$T_t = h[\beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}]$$

em que: L_t o nível estimado (valor médio ou ponto central da série); T_t tendência estimada; β fator de suavização da tendência; F_{t+h} previsão para h períodos a frente

⇒ **Alisamento com Tendência e Sazonalidade (Holt-Winters)**: para séries com tendência e sazonalidade.

$$F_{t+h} = (L_t + hT_t)S_{t+h-p}$$

$$L_t = \alpha \frac{y_t}{S_{t-p}} + (1-\alpha)(L_{t-1} + T_{t-1})$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}$$

$$S_t = \gamma \frac{y_t}{L_t} + (1-\gamma)S_{t-p}$$

em que: S_t componente sazonal estimado; γ fator de suavização sazonal; p é o período sazonal; L_t o nível estimado; T_t tendência estimada; β fator de suavização da tendência; F_{t+h} previsão para h períodos a frente

Modelos no Domínio da Frequência

Os modelos no domínio da frequência analisam a série temporal em termos de suas componentes frequenciais, utilizando transformações como a Transformada de Fourier. Os principais conceitos incluem:

1. **Análise de Fourier**:

Decompõe uma série temporal em uma soma de senóides e cossenoides de diferentes frequências.

$$Y(f) = \sum_{t=0}^{n-1} Y_t e^{-2\pi i f t/n}$$

em que: $Y(f)$ representa a amplitude da frequência f.

2. **Espectro de Frequência**:

Mostra a amplitude ou potência das diferentes frequências presentes na série temporal. É útil para identificar ciclos e padrões periódicos nos dados.

3. **Análise de Ondaletas (Wavelet Analysis)**:

Decompõe a série temporal em componentes de frequência e tempo simultaneamente, permitindo a análise de variações locais em diferentes escalas.

$$w(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} X(t)\psi^* \left(\frac{t-b}{a}\right) dt$$

em que: ψ é a função de ondaleta, a é a escala e b é a posição no tempo.

Aplicações Práticas

Domínio do Tempo

- **Modelos ARIMA** são amplamente usados em previsões de séries temporais, como vendas, estoque e indicadores econômicos.
- **Modelos SARIMA** são usados para séries com componentes sazonais, como consumo de energia elétrica e temperatura.

Domínio da Frequência

- **Análise de Fourier** é usada para identificar padrões cíclicos em dados financeiros, como ciclos econômicos.
- **Análise de Ondaletas** é útil em estudos de sinais biomédicos e geofísicos, onde as variações locais e a análise multi-resolução são importantes.