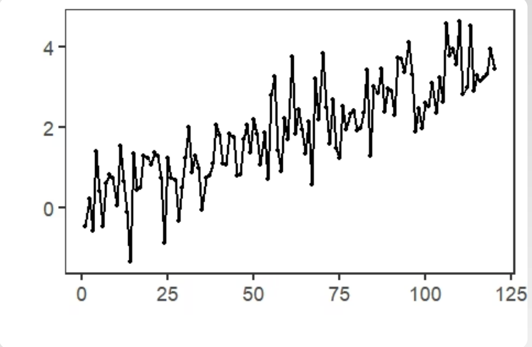
Fundamentos de Séries Temporais e Processos Estocásticos

Uma jornada pelos conceitos essenciais que fundamentam a análise de dados temporais e a modelagem de processos aleatórios ao longo do tempo.

Eduardo Ogasawara

eduardo.ogasawara@cefet-rj.br

https://eic.cefet-rj.br/~eogasawara



CONCEITO FUNDAMENTAL

Série Temporal como Sequência de Observações

Uma série temporal é uma sequência ordenada de valores no tempo, onde cada observação está associada a um instante temporal específico. A ordem das observações é parte essencial da informação.

A série pode ser representada como:

$$x_1, x_2, \ldots, x_T$$

Onde x_t é o valor observado no instante t, e T é o número total de observações.

Também podemos representar como função do tempo: $x(t)$

Por que a ordem importa?

Diferentemente de dados comuns, não podemos reorganizar os valores sem alterar o significado da informação. Cada observação está ligada a um momento específico do tempo.

Série Temporal como Função Discreta do Tempo

Função do Tempo

A série pode ser vista como uma função matemática que mapeia instantes de tempo em valores reais.

Tempo Discreto

O tempo assume valores discretos, como dias, meses ou anos, não contínuos.

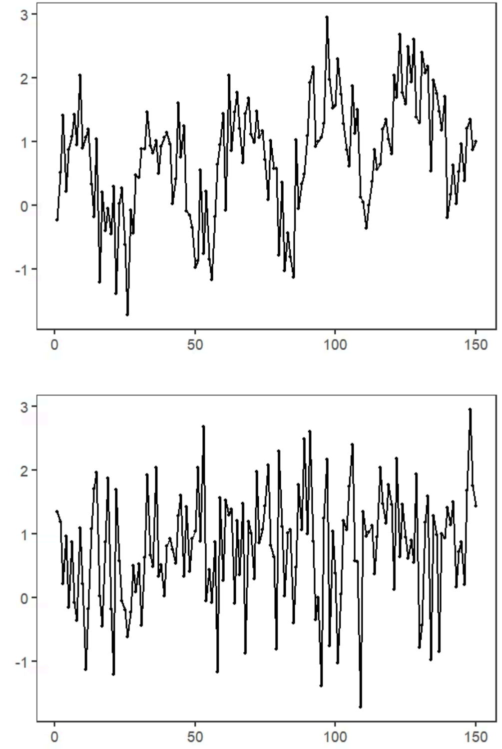
Valor Associado

Cada instante tem um único valor associado, formando a sequência temporal.

Matematicamente, definimos a série como uma função:

$$x : \mathcal{T} \to \mathbb{R}$$

Onde $\mathcal{T}$ é o conjunto de instantes de tempo (valores inteiros) e $\mathbb{R}$ é o conjunto dos números reais. Essa visão funcional permite tratar séries temporais com ferramentas matemáticas e probabilísticas.



A Importância Crítica da Ordem Temporal

Sequência Original

$$(x_1, x_2, \ldots, x_T)$$

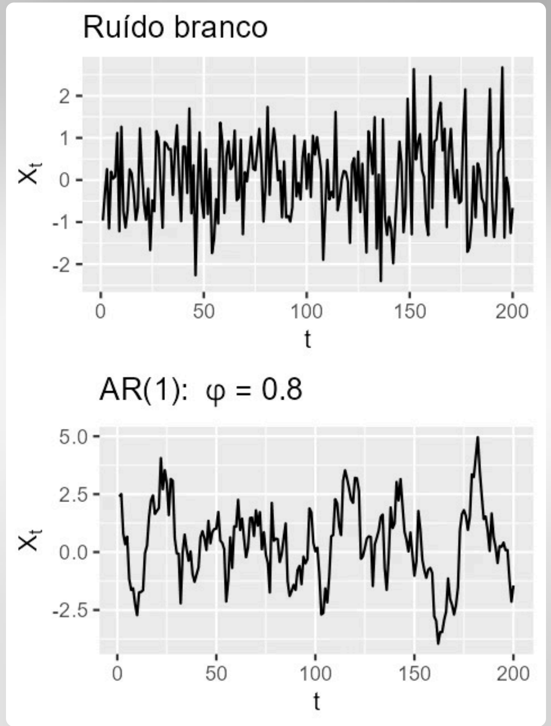
A ordem natural preserva a estrutura temporal e o significado dos dados.

Após Permutação

$$(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(T)})$$

Trocar a ordem destrói a estrutura temporal e altera completamente o significado.

Se trocarmos a ordem das observações, perdemos a estrutura temporal. Em dados tradicionais, a ordem não importa; em séries temporais, ela é fundamental. Isso significa que métodos estatísticos usuais precisam ser adaptados, pois a dependência entre observações consecutivas é parte do fenômeno estudado.



DEPENDÊNCIA

Dependência Temporal entre Observações

1

Valor Passado

Observações anteriores influenciam o presente

2

Valor Presente

Depende dos valores em instantes anteriores

3

Valor Futuro

Será influenciado pelo presente e passado

Em séries temporais, observações ao longo do tempo tendem a ser dependentes. O valor presente pode depender de valores passados, e a hipótese de independência geralmente não é válida.

A dependência pode ser expressa como:

$$X_t \text{ depende de } X_{t-1}, X_{t-2}, \ldots$$

Por exemplo, preços, temperaturas ou demanda raramente mudam de forma totalmente aleatória entre períodos consecutivos. Essa dependência temporal é a razão pela qual precisamos de modelos específicos, diferentes dos usados em estatística clássica.

Série Observada e Processo Estocástico

Conceitos Fundamentais

Processo Estocástico

Conjunto de variáveis aleatórias ao longo do tempo: $\{X_t\}$

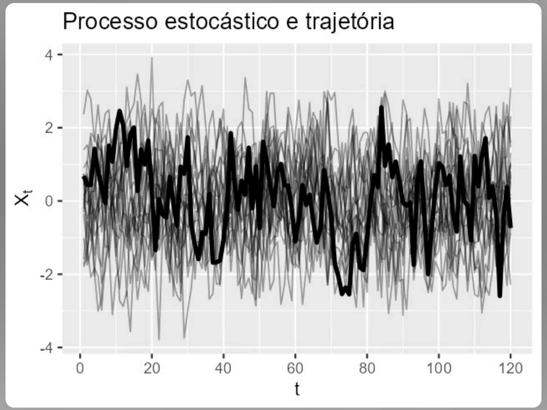
Série Observada

Uma realização específica do processo: x_t

Variável Aleatória

Cada instante t corresponde a uma variável aleatória X_t

A ideia central é separar "processo" e "dados". O processo estocástico representa o mecanismo gerador dos dados, enquanto a série observada é apenas uma trajetória desse processo. Essa distinção é fundamental: modelar séries temporais significa tentar descrever matematicamente o processo que gerou os dados observados.



Processo Estocástico e Trajetória Observada

A série temporal é modelada como um processo aleatório, onde cada instante corresponde a uma variável aleatória. A série observada é apenas uma trajetória possível desse processo.

Processo Estocástico

$$\{X_t\}_{t=1}^{T}$$

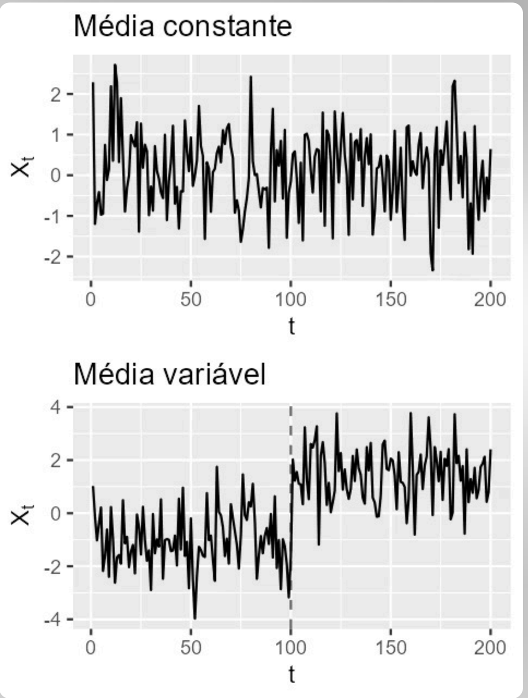
Todas as trajetórias possíveis

Trajetória Observada

$$x_t = X_t(\omega)$$

Uma realização específica

Aqui formalizamos a ideia de que os dados observados são apenas uma realização possível de um processo aleatório. O símbolo ω representa um elemento do espaço probabilístico (resultado possível). Essa visão é essencial para justificar o uso de probabilidades e modelos estatísticos em séries temporais.



ESTATÍSTICAS

Média do Processo Temporal

Média Teórica

Cada instante possui uma média teórica que pode variar ao longo do tempo:

$$\mu_t = \mathbb{E}[X_t]$$

Onde μ_t é a média do processo no instante t, e $\mathbb{E}[X_t]$ é o valor esperado.

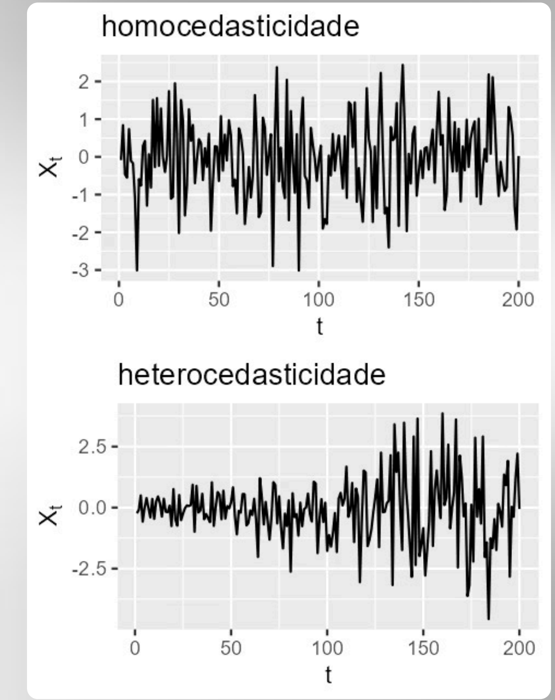
Média Amostral

A média observada é uma estimativa da média teórica:

$$\bar{x} = \frac{1}{T}\sum_{t=1}^{T} x_t$$

Onde $\bar{x}$ é a média da série observada.

A média do processo descreve o comportamento esperado da série em cada instante do tempo. Como essa média não é observável diretamente, usamos a média da série observada como estimativa. Em muitos modelos, assume-se que a média não depende do tempo, o que simplifica a análise e leva ao conceito de estacionariedade.



Variância do Processo Temporal

Variância Teórica

$$\sigma_t^2 = \mathrm{Var}(X_t)$$

Mede a dispersão da série no instante t

Variância Amostral

$$s^2 = \frac{1}{T-1}\sum_{t=1}^{T}(x_t - \bar{x})^2$$

Estimativa calculada a partir dos dados observados

A variância descreve o grau de variabilidade da série ao longo do tempo. Em alguns processos, a variância é constante; em outros, ela muda, indicando instabilidade. Essa diferença é crucial para a modelagem, pois muitos modelos clássicos assumem variância constante.

Covariância Temporal

A covariância temporal mede a relação entre valores em instantes diferentes, expressando a dependência temporal. É uma função do tempo e da defasagem.

1 — Instante t-h

Valor passado

2 — Defasagem h

Intervalo temporal

3 — Instante t

Valor presente

A covariância entre dois instantes é definida como:

$$\mathrm{Cov}(X_t, X_{t-h})$$

Onde h representa a defasagem temporal (lag). Se a covariância é diferente de zero, existe dependência temporal. Esse conceito é a base para entender autocorrelação e para construir modelos que capturam a estrutura de dependência ao longo do tempo.

Função de Autocovariância

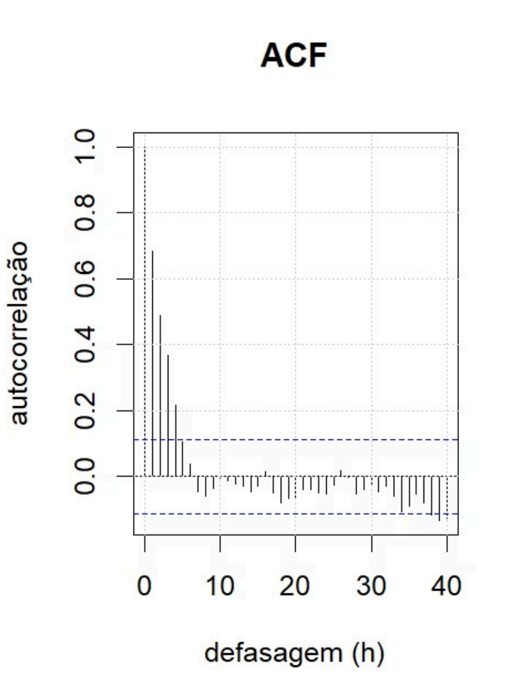
A função de autocovariância descreve a dependência em função da defasagem, sendo central na análise de séries temporais:

$$\gamma(h) = \mathrm{Cov}(X_t, X_{t-h})$$

Onde $\gamma(h)$ é a autocovariância para defasagem h.

> Em processos simples, a dependência diminui à medida que o intervalo entre observações aumenta.

A função de autocovariância mostra como a dependência entre observações varia com a defasagem temporal. É a base para definir a autocorrelação e identificar padrões estruturais na série.



NORMALIZAÇÃO

Autocorrelação

01

Autocovariância

Medida de dependência que depende da escala dos dados

02

Normalização

Dividimos pela variância para remover o efeito da escala

03

Autocorrelação

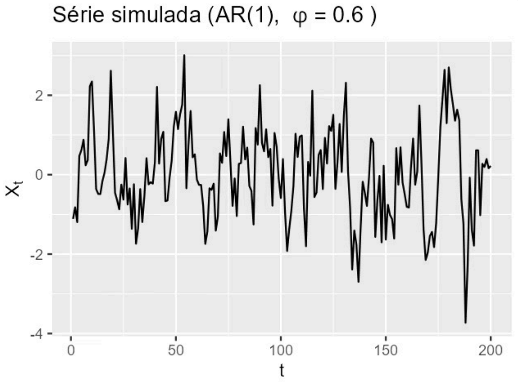
Medida padronizada que varia entre -1 e 1

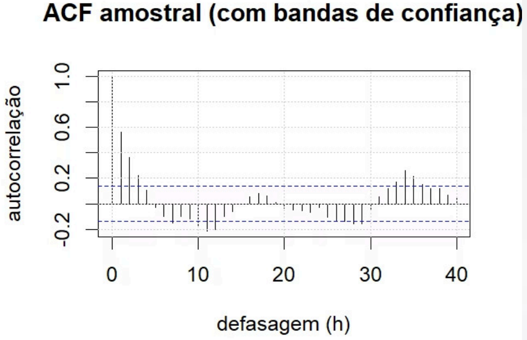
A função de autocorrelação é definida como:

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}$$

Onde $\rho(h)$ é a autocorrelação na defasagem h, $\gamma(h)$ é a autocovariância na defasagem h, e $\gamma(0)$ é a variância do processo.

A autocovariância depende da escala da série, dificultando comparações. Por isso, normalizamos pela variância, obtendo a autocorrelação. Esta mede a força da relação entre valores separados por uma defasagem temporal e é uma das ferramentas mais importantes para entender a estrutura de uma série temporal.





Autocorrelação Amostral

A autocorrelação teórica não é observável diretamente. Usamos estimativas a partir dos dados observados, onde a ACF amostral aproxima a ACF do processo.

Autocovariância Amostral

$$\hat{\gamma}(h) = \frac{1}{T} \sum_{t=h+1}^{T} (x_t - \bar{x})(x_{t-h} - \bar{x})$$

Autocorrelação Amostral

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

Como não conhecemos o processo gerador dos dados, calculamos a autocorrelação a partir da série observada. A ACF amostral fornece uma aproximação da dependência temporal do processo e é usada na prática para identificar modelos e padrões na série.

Conceito de Estacionariedade

Estacionária

Média, variância e autocorrelação constantes ao longo do tempo

Não estacionária

Propriedades mudam; tendência, variância e autocorrelação variáveis

As propriedades estatísticas de uma série podem variar no tempo.

Estacionariedade significa invariância ao longo do tempo e é um conceito central da modelagem de séries temporais.

> Um processo é estacionário se suas propriedades estatísticas não dependem do instante temporal.

A estacionariedade é uma hipótese fundamental em séries temporais. Se a média, variância e dependência temporal mudam ao longo do tempo, muitos modelos clássicos deixam de ser válidos. Por isso, grande parte da análise consiste em verificar ou induzir estacionariedade na série.

✓ ESTACIONARIEDADE FRACA

Condições de Estacionariedade Fraca

1

Média Constante

$$\mathbb{E}[X_t] = \mu$$

A média não varia ao longo do tempo

2

Variância Constante

$$\mathrm{Var}(X_t) = \sigma^2$$

A dispersão permanece estável

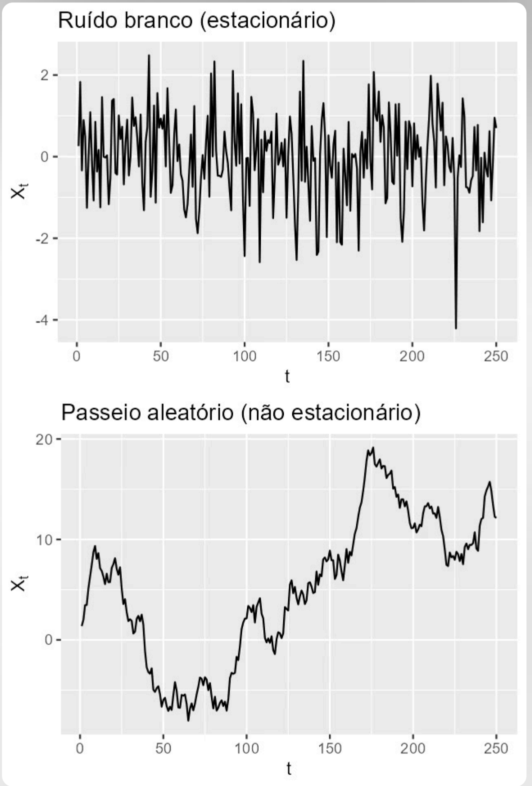
3

Autocovariância Dependente Apenas da Defasagem

$$\mathrm{Cov}(X_t, X_{t-h}) = \gamma(h)$$

A dependência depende só do intervalo, não do tempo absoluto

A estacionariedade fraca é a forma mais usada na prática. Ela exige apenas que média, variância e estrutura de dependência sejam invariantes no tempo. Modelos clássicos como ARMA são construídos sob essa hipótese, o que explica sua importância teórica e prática.



Exemplos: Processos Estacionários e Não Estacionários

Processo Estacionário

Ruído Branco

$$X_t = \varepsilon_t$$

Sem memória temporal, propriedades constantes no tempo

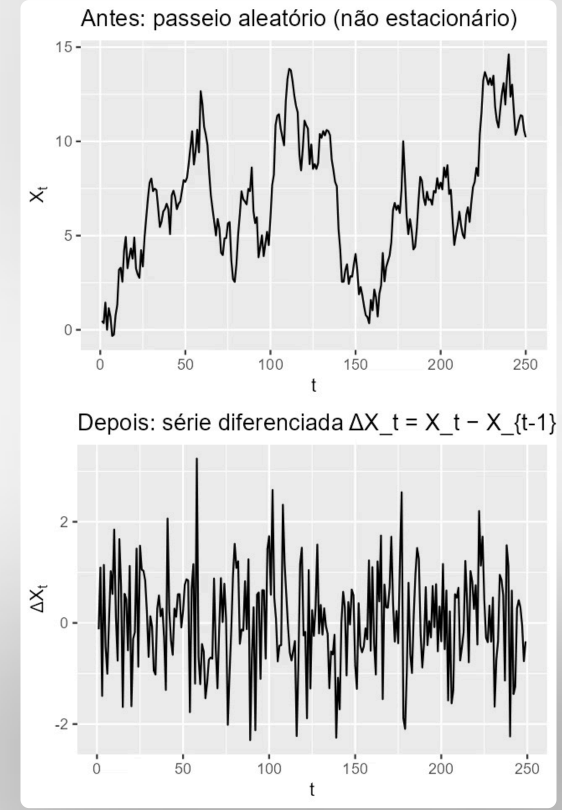
Processo Não Estacionário

Passeio Aleatório

$$X_t = X_{t-1} + \varepsilon_t$$

Acumula choques, variância crescente ao longo do tempo

Onde ε_t é um termo aleatório com média zero e variância constante. O ruído branco representa um processo simples, mas com propriedades constantes. Já o passeio aleatório acumula choques, gerando variância crescente e ausência de estacionariedade. Esse contraste ajuda a entender por que muitas séries reais precisam ser transformadas antes da modelagem.



Transformações para Induzir Estacionariedade

Série Original

Não estacionária, com tendência ou variância crescente

Diferenciação

$$Y_t = X_t - X_{t-1}$$

Remove variações sistemáticas de longo prazo

Série Transformada

Estacionária, adequada para modelagem

Séries reais frequentemente não são estacionárias. Transformações podem estabilizar média e variância, sendo a diferenciação a mais comum. Quando uma série apresenta tendência ou crescimento ao longo do tempo, a diferenciação remove essas variações sistemáticas, transformando a série original em uma nova série mais estável. Esse procedimento é fundamental em modelos como ARIMA.

Interpretação da Diferenciação

A diferenciação mede a variação entre períodos consecutivos, eliminando tendência determinística simples e alterando a estrutura de dependência temporal.

A diferença de primeira ordem é representada por:

$$Y_t = \Delta X_t$$

Onde ΔX_t representa a variação da série entre t e $t-1$.

> **Por que diferenciar?**
>
> Muitas séries econômicas, financeiras e ambientais apresentam crescimento acumulativo que impede a estacionariedade.

A diferenciação transforma a série em variações ao longo do tempo. Em vez de analisar níveis absolutos, passamos a analisar mudanças entre períodos consecutivos.

MODELO AR(1)

Modelo Autorregressivo de Ordem 1

O modelo AR(1) descreve situações onde o valor atual depende linearmente do valor passado. É um modelo linear simples de dependência temporal e serve como base para modelos mais complexos.

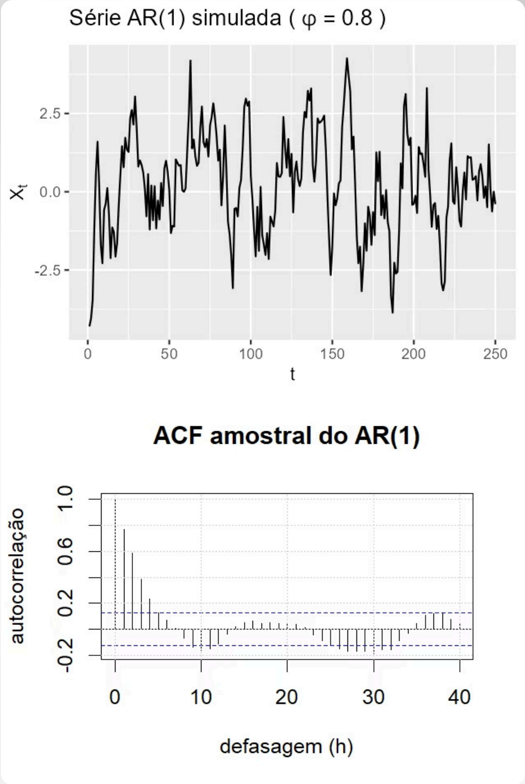
Estrutura do Modelo

$$X_t = \phi X_{t-1} + \varepsilon_t$$

Componentes

- X_t: valor no instante t
- ϕ: parâmetro de dependência
- X_{t-1}: valor no instante anterior
- ε_t: termo aleatório com média zero

O parâmetro ϕ mede a intensidade da dependência temporal. Esse modelo é a estrutura mais simples capaz de capturar memória temporal e serve como ponto de partida para modelos ARMA e ARIMA.



Condição de Estacionariedade no AR(1) e Autocorrelação

Condição de Estacionariedade

A estacionariedade depende do parâmetro do modelo. Valores altos de dependência geram instabilidade.

$$|\phi| < 1$$

Quando $|\phi| < 1$, os efeitos de choques passados diminuem ao longo do tempo, garantindo estabilidade estatística.

Padrão de Autocorrelação

Modelos geram padrões específicos de autocorrelação. A ACF revela a estrutura do processo.

No modelo AR(1), a autocorrelação é:

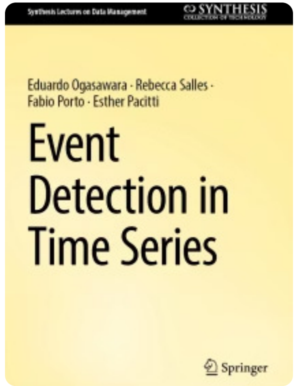
$$\rho(h) = \phi^h$$

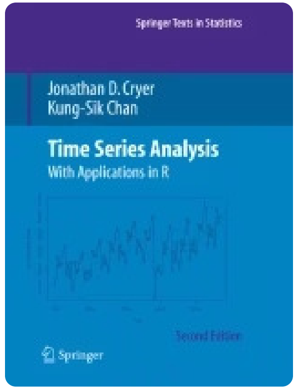
A autocorrelação decai exponencialmente com a defasagem h. Essa "assinatura" permite identificar modelos a partir dos dados observados.

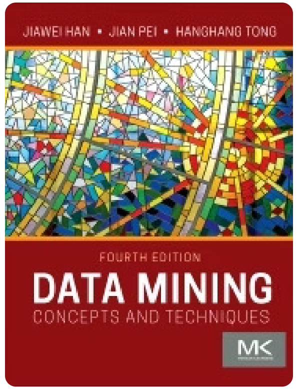
RECURSOS ACADÊMICOS

Referências Bibliográficas

Uma coleção cuidadosamente selecionada de obras fundamentais que abordam análise de séries temporais e mineração de dados.







Event Detection in Time Series

Ogasawara, E.; Salles, R.; Porto, F.; Pacitti, E. (2025). Publicação recente da Springer Nature Switzerland que explora técnicas avançadas de detecção de eventos em séries temporais.

Time Series Analysis: With Applications in R

Cryer, J. D.; Chan, K.-S. (2008). Obra clássica da Springer que combina fundamentação teórica sólida com implementações práticas.

Data Mining: Concepts and Techniques

Han, J.; Pei, J.; Tong, H. (2022). Quarta edição publicada pela Morgan Kaufmann que consolida conceitos fundamentais e técnicas avançadas de mineração de dados