Universidade Federal do Rio de Janeiro
Instituto de Matemática
Departamento de Métodos Estatísticos

Análise de Séries Temporais

Prof. Hélio Migon

Sumário

1. Introdução

1.1. Noções Básicas

Série temporal: conjunto de observações ordenadas (no tempo). Tempo pode ser: espaço, profundidade, ...

Observações vizinhas são dependentes.

Estudos de séries temporais: modelagens, análise dessa dependência, técnicas específicas a séries temporais.

Exemplos de aplicações:
Economia: preços diários de ações; taxa de desemprego.
Medicina: níveis de eletrocardiograma ou eletroencefalograma.
Epidemiologia: casos semanais de sarampo; casos mensais de AIDS.
Metereologia: temperatura diária; registro de marés, . . .

Classificação:
Série temporal é o conjunto de observações $\{Y(t), t \in T\}$, Y: variável de interesse, T: conjunto de índices

Tipos de séries temporais

1. Discreta: $T = \{t_1, t_2, \ldots, t_n\}$
   Ex: Exportações mensais de 1970 a 1980 $\{01/1970, 02/1970, \ldots, 11/1980, 12/1980\}$.
   Notação: Y_t

2. Contínua: $T = \{t : t_1 < t < t_2\}$
   Ex: Registro da maré no Rio durante 1 ano $T = [0, 24]$ se unidade de tempo é a hora.
   Notação: $Y(t)$

3. Multivariada: Observações são $Y_1(t), \ldots, Y_k(t), t \in T$.
   Ex: Vendas semanais $Y_1(t)$ e gastos com propaganda $Y_2(t)$.

Y pode também ser discreto ou contínuo. Muitas vezes, Y é discreto mas pode ser tratado como contínuo.

Ex: Número de casos notificados de AIDS. Nesse curso, séries são univariadas, discretas e observada em tempos equiespaçados.

Podemos identificar T com $\{1, 2, \ldots, n\}$

Objetivos de uma análise de séries temporais

Os principais objetivos são:

i) Compreender o mecanismo gerador da série;

ii) Predizer o comportamento futuro da série.

Compreender o mecanismo da série possibilita:

- Descrever efetivamente o comportamento da série;
- Encontrar periodicidades na série;
- Tentar obter razões para o comportamento da série (possivelmente através de variáveis auxiliares);
- Controlar a trajetória da série.

Predizer o futuro possibilita:

- Fazer planos a longo, médio e curto prazo;
- Tomar decisões apropriadas.

Objetivos **(i)** e **(ii)** estão ligados. É possível ocorrer bem rotineiramente se o modelo é adequado, a não ser nos raros casos de modelos determinísticos.

$$\text{Futuro envolve incerteza} \Longrightarrow \text{previsões não são perfeitas.}$$

Objetivo é reduzir ao máximo os erros de previsão.

1.2. Fundamentos Probabilísticos

Definição: Seja T um conjunto arbitrário. Um processo estocástico é uma família $\{Y(t), t \in T\}$ tal que, $\forall t \in T$, $Y(t)$ é uma variável aleatória.

Série temporal é um processo estocástico. O conjunto de valores $\{Y(t), t \in T\}$ é chamado de espaço de estados e os valores Y(t) são chamados de estados.

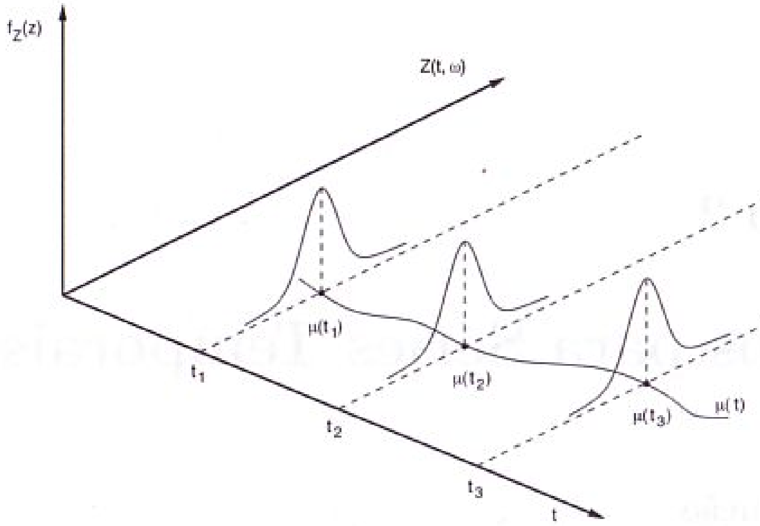


Figura 2.1: Um processo estocástico interpretado como uma família de variáveis aleatórias.

Para cada t, $Y(t)$ tem uma distribuição de probabilidade. Pode ser a mesma ou não.

Um possível valor de um processo estocástico é uma trajetória em t.

Uma forma alternativa de definição de processo estatístico é uma família de v.a. $\{Y(t), t \in T\}$ é um processo estocástico se as v.a. $Y(t_1), Y(t_2), ..., Y(t_n)$ tem f.d. finito-dimensionais

$$F(y_1, ..., y_n; t_1, ..., t_n) = Pr(Y(t_1) \leq y_1, ..., Y(t_n) \leq y_n)$$

conhecidas para todo $n \geq 1$ satisfazendo as condições de:

- Simetria: Para qualquer permutação $j_1, \ldots, j_n$ dos índices $1, 2, \ldots, n$ temos:

$$F(y_{j_1}, \ldots, y_{j_n}; t_{j_1}, \ldots, t_{j_n} = F(y_1, \ldots, y_n; t_1, \ldots, t_n)$$

Ex: $(n = 3)$ $F(y_2, y_1, y_3; t_2, t_1, t_3) = F(y_1, y_2, y_3; t_1, t_2, t_3)$

- Consistência:

$$\lim F(y_1, \ldots, y_n; t_1, \ldots, t_n) = F(y_1, \ldots, y_{n-1}; t_1, \ldots, t_{n-1})$$

Essa definição não é muito útil na prática pois é muito difícil a especificação de todas as distribuições finito-dimensionais Normalmente, o que se faz é concentrar nos primeiros momentos. Estes são:

i) Função média: $\mu(t) = E\{Y(t)\}$

ii) Função auto-covariância **(facv)**:

$$\begin{aligned}\gamma(t_1, t_2) &= E[Y(t_1) - \mu(t_1)][Y(t_2) - \mu(t_2)] \\ &= E\{Y(t_1)Y(t_2)\} - \mu(t_1)\mu(t_2)\end{aligned}$$

Em particular se $t_1 = t_2 = t$,

$\gamma(t, t) = V\{Y(t)\}$ é a variância de $Y(t)$ denotada por $V(t)$ ou $\sigma^2(t)$.

A **(facv)** fornece a forma de dependência temporal do processo $Y(t)$. Ela não traduz a força dessa dependência pois depende da unidade de medição de Y.

Para denotar esse problema, a **(facv)** é comumente substituida pela:

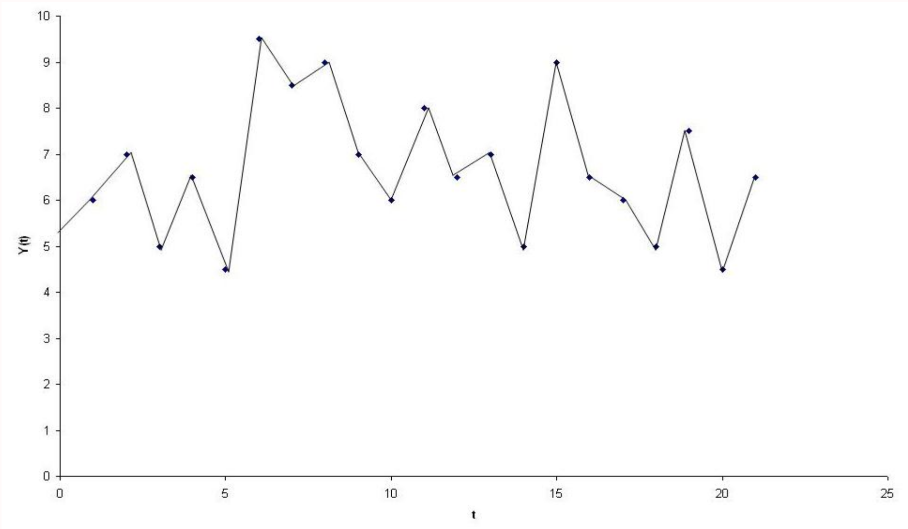
iii) Função de auto-correlação: $\rho(t_1, t_2) = \gamma(t_1, t_2)/\sigma(t_1)\sigma(t_2)$

A importância da função média e da facv deve-se ao fato de que se as distribuições finito-dimensionais de $Y(t)$ são normais estão basta conhecer μ e γ para conhecer todo o processo.

1.3. Processos Estacionários

Como a quantidade de parâmetros é usualmente maior que o número de observações, são necessárias hipóteses simplificadoras.

A mais comum em séries temporais é a de estacionariedade. Basicamente isso significa que o comportamento da série não se altera com o passar do tempo, ou seja, média e facv não mudam se caminharmos no tempo.



Tecnicamente, existem duas formas de estacionariedade:

- estrita (ou forte);
- fraca (ou ampla ou de 2ª ordem).

Um processo estocástico $Y(t)$ é estritamente estacionário se suas distribuições finito-dimensionais são invariantes por translações no tempo, isto é,

$$F(y_1, \ldots, y_n; t_1 + \tau, \ldots, t_n + \tau) = F(y_1, \ldots, y_n; t_1, \ldots, t_n),$$

$\forall t_1, \ldots, t_n, \tau$

O processo deslocado τ unidades no tempo permanece com as mesmas características.

Em particular com $n = 1$ e $t_2 = t_1 + \tau$ temos que:

$$F(y_1; t_2) = F(y_2; t_1)$$

e, portanto,

$$\mu = \mu(t)$$

e

$$\sigma^2 = \sigma^2(t)$$

são constantes.

Além disso,

$$F(y_1, y_2; t_1 + \tau, t_2 + \tau) = F(y_1, y_2; t_1, t_2)$$

Logo,

$$\gamma(t_1, t_2) = \gamma(t_1 + \tau, t_2 + \tau)$$

Fazendo $\tau = -t_2$ e $t = t_1 - t_2$ temos que:

$$\gamma(t_1, t_2) = \gamma(t_1 - t_2, 0) = \gamma(t, 0) = \gamma(t)$$

A **facv** depende apenas da distância entre os pontos considerados.

Um processo estocástico $Y(t)$ é fracamente estacionário se:

1. $E\{Y(t)\} = \mu(t) = \mu$
2. $V\{Y(t)\} = \sigma^2(t) = \sigma^2$
3. $Corr\{Y(t_1), Y(t_2)\} = \gamma(t_1, t_2) = \gamma(t_2 - t_1)$

Se os momentos existem, então:

$$Estacionariedade\ forte \rightarrow Estacionariedade \text{ fraca}$$

A volta só vale se a Distribuição finito-dimensionais de $Y(t)$ são normais (Processo Gaussiano).

Observe que a função de auto-correlação de processos estacionários é:

$$\rho(t_1, t_2) = \frac{\sigma(t_1, t_2)}{\sigma(t_1)\sigma(t_2)} = \frac{\gamma(t_2 - t_1)}{\sigma^2(t_1)} = \frac{\gamma(t)}{\gamma(0)} = \rho(t)$$

De agora em diante, consideraremos apenas processos estacionários ou passíveis de "estacionarização" por transformações.

Conceito não tão importante em modelos dinâmicos.

Após observar a série temporal discreta $Y_1, \ldots, Y_n$, γ_k será estimado por:

$$c_k = \frac{1}{n} \sum_{t=1}^{n-k} (Y_i - \overline{Y})(Y_{i+k} - \overline{Y})$$

onde

$\overline{Y} = \frac{1}{n} \sum_{t=1}^{n} Y_i$ e
$k = 0, 1, \ldots, n-1$

c_k é a **facv** amostral.

ρ_k pode então ser estimado por $r_k = \frac{c_k}{c_0}$, função de autocorrelação amostral.

Como normalmente as séries não apresentam esse comportamento estável, recorre-se a transformações nos dados. As transformações mais comuns são:

- **Transformação Box-Cox**: Considere a função da forma:

$$g(y) = \frac{(y^\lambda - 1)}{\lambda} \tag{1}$$

Se

$\lambda = 1$, g é a identidade,
$\lambda = -1$, g é a transformação inversa,
$\lambda = 1/2$, g transforma y em $\sqrt{y}$,
e $\lim_{\lambda \to 0} g(y) = \log(y)$

Essas transformações são usadas principalmente para estudar a variância.

- **Operação Diferença (Δ)**: Define-se o operador diferença através de:

$$\Delta y_t = y_t - y_{t-1} \tag{2}$$

Aplicando-se o operador novamente obtem-se a 2ª diferença:

$$\begin{aligned}
\Delta^2 y_t &= \Delta(\Delta y_t) \\
&= \Delta(y_t - y_{t-1}) \\
&= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \\
&= y_t - 2y_{t-1} + y_{t-2}
\end{aligned}$$

A n-ésima diferença de y_t é obtida recursivamente por:

$$\Delta^n y_t = \Delta(\Delta^{n-1} y_t) \tag{3}$$

Normalmente, uma ou duas diferenças são suficientes para tornar a série estacionária.

Modelagem, aprendizado e previsão

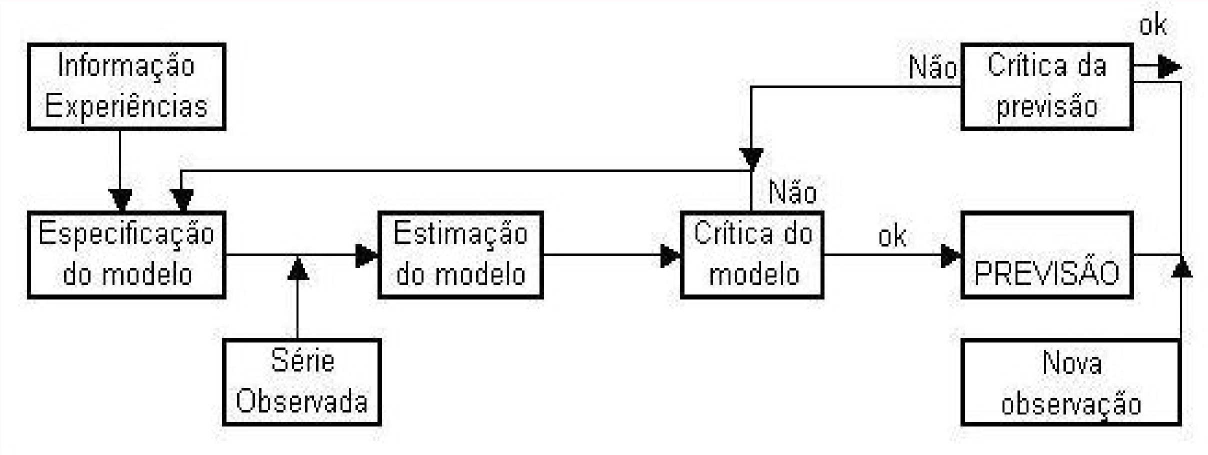
Central à análise de série temporais está a construção de um modelo.

Modelo - esquema de descrição (e explicação) que organiza informação (e experiência) de forma a propiciar aprendizagem e previsão.

Bom **modelo** permite **aprendizado** levando a **previsões** adequadas.

- Devido à incerteza presente, modelo é probabilístico.
- Deve também ser econômico (parsimônia).
- Descrição deve ser relativamente simples e flexível para poder se adaptar ao futuro (incerto) e facilitar aprendizado.
- Aprendizado é processamento de informação através do modelo.
- Previsão é hipótese, conjectiva ou especulação sobre o futuro.

Esquema de sistema de previsão:



Critérios de Previsão

O melhor critério para escolher um modelo de previsão é a sua capacidade preditiva, ou seja, quão perto estão as previsões dos valores posteriormente observados.

Suponha que observamos uma série até o instante t e queremos prever o valor da série no instante $t+h$.

Denotaremos por $\hat{Y}_t(h)$ a previsão de Y_{t+h} no instante t.

Ex: $\hat{Y}_{t-1}(1)$ é a previsão de Y_t no instante $t-1$

t é a origem da previsão

h é o horizonte da previsão

Observe que $\hat{Y}_t(h)$ é uma v.a. conhecida apenas dada a história observada do processo até o tempo t. Associado a $\hat{Y}_t(h)$ temos o erro de previsão $Y_{t+h} - \hat{Y}_t(h)$

Se erros de previsão positivos e negativos são igualmente importantes faz sentido procurar previsões que minimizem o erro absoluto médio $E[Y_{t+h} - \hat{Y}_t(h)]$ e o erro quadrático médio $E[Y_{t+h} - \hat{Y}_t(h)]^2$

Podemos identificar dois procedimentos distintos de construção de modelos de previsão:

i) baseia-se em teorias e inclui muitas variáveis - econométrico;
ii) baseia-se no comportamento observado da série - séries temporais.

Privilegiaremos o segundo.

1.4. Modelos de séries temporais

Modelos podem ser divididos em 2 classes :

(i) paramétricos - Nº finito de parâmetros. Análise é feita no domínio do tempo.

(ii) não-paramétricos - Nº infinito de parâmetros. Análise é feita no domínio da frequência.

O curso será basicamente sobre modelos paramétricos.

Os modelos dessa classe podem ser genericamente escritos como

$$Y_t = S_t + \varepsilon_t$$

ou seja

$$\text{Observação} = \text{Sinal} + \text{Ruído}$$

Assim temos:

(a) Modelos de regressão (CAP.2)

- Ruídos são não-correlacionados
- $S_t = x_t \beta$

Exemplos :

1. Modelo de tendência linear

$$S_t = \alpha + \beta t$$

2. Modelo de curva de crescimento

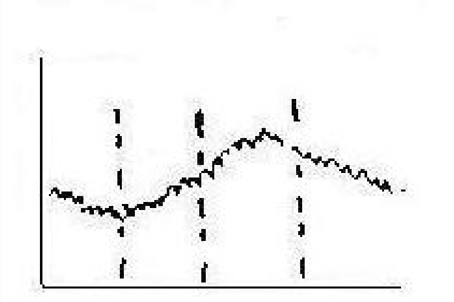
$$Y_t = \alpha e^{\beta t + \varepsilon_t},$$

ou seja

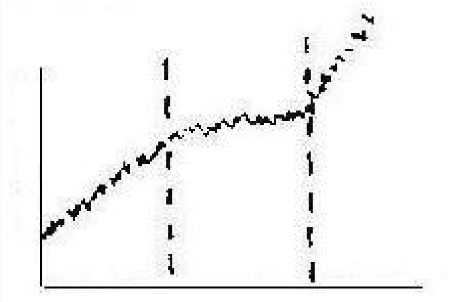
$$\log Y_t = \log \alpha + \beta t + \varepsilon_t$$

(b) Modelos de alisamento exponencial

- O sinal novamente descreve uma função como em (a).
- Relação do sinal vale apenas "localmente".
- Parâmetros sujeitos a pequenas variações temporais



(a) "Localmente" constante.



(b) "Localmente" linear.

(c) Modelos Autoregressivos (AR)

Como em (a) com

$$S_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p}$$

Nível (sinal) atual depende dos níveis passados.

(d) Modelos lineares estacionários

$$S_t = \Psi_1 \varepsilon_{t-1} + \Psi_2 \varepsilon_{t-2} + \ldots = \sum_{i=1}^{\infty} \Psi_i \varepsilon_{t-i}$$

Esse processo é estacionário se $\sum_{i=1}^{\infty}(\Psi^2) < \infty$

Inclui os modelos ARMA (que serão estudados no CAP.4) que inclui os modelos AR em (c).

Os modelos ARIMA são uma generalização dos modelos ARMA que visam basicamente tornar o processo estacionário através de operações diferença.

(e) Modelos de espaço de estados ou dinâmicos (CAP.5)

$$S_t = \underline{X_t}\,\underline{\beta_t}$$

generalizando os modelos de regressão

$$\underline{\beta_t} = G_{t+1}\underline{\beta_t} + \underline{W_t}$$

Esses modelos incluem os modelos ARIMA. Podem ser usados tanto do ponto de vista clássico quanto Bayesiano.

Ex: $\underline{X_t} = 1$, $\underline{\beta_t} = \beta_t$, $G_t = 1$ e $\underline{W_t} = W_t$

logo

$$\begin{aligned} y_t &= \beta_t + \varepsilon_t, \\ \beta_t &= \beta_{t-1} + W_t. \end{aligned}$$

Equivale ao modelo de alisamento exponencial para séries "localmente" constantes.

(f) Modelos de equações simultâneas ou econométricos

$$Y_t^T \Gamma + X_t B = \epsilon_t$$

Y_t - Variáveis endógenas
X_t - Variáveis exógenas

Se Γ tem posto máximo então após observar amostra de tamanho n

$$\underline{Y_t} = \underline{X_t} T + \underline{E}$$

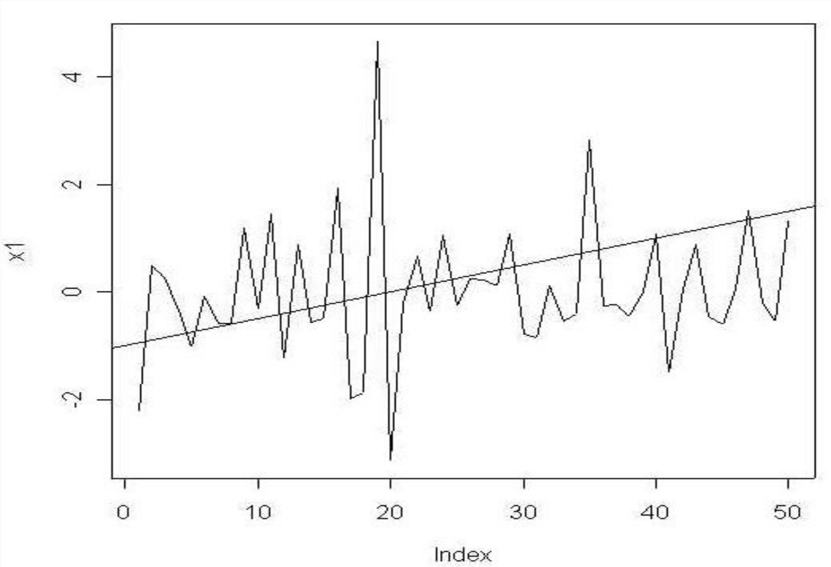
onde:

$$\begin{aligned} \underline{Y}^T &= (Y_1 Y_2 \ldots Y_n) \\ \underline{X}^T &= (X_1 X_2 \ldots X_n) \\ T &= B\Gamma^{-1} \\ \underline{E}^T &= (\epsilon_1 \epsilon_2 \ldots \epsilon_n) \end{aligned}$$

Fenômenos Típicos de Séries Temporais

Boa parte das séries têm características típicas, as principais são:

i) Tendência: é o efeito de longo prazo na média. Especificação de longo prazo é difícil.



A Série acima apresenta tendência de crescimento linear.

ii) Sazonalidade: efeitos ligados à variações periódicas (semanal, mensal, anual, etc.).

Ex: Medidas de Temperatura (aumenta no verão e diminui no inverno).

iii) Ciclos: variações que apesar de periódicas não são associadas automaticamente a nenhuma medida temporal.

Ex: Ciclos Econômicos (5 e 7 anos) e Ciclos de epidemias.

Uma das tarefas mais importantes em **ST** é identificar estas componentes visando a decomposição da série estudada.

2. Modelos de Regressão

Conforme visto no capítulo 1 podemos pensar na Série Temporal como uma coleção de observações determinadas por um sinal dependendo de uma forma determinística do tempo às quais são superpostos erros não correlacionados.

Procura-se neste caso usar modelos de regressão para caracterizar o sinal que controla a série.

EXEMPLOS:

1- Modelo de Tendência Linear: $S_t = a + bt$;

2- Modelo de Crescimento Exponencial: $y_t = ye^{bt+\epsilon t}$;

3- Modelo de Regressão Linear Simples: $S_t = a + bx_t$;

4- Modelo de Regressão Não Linear: $S_t = 1/(a + bx_t)$;

Os modelos 2 e 4 não são lineares nos parâmetros, embora Z possa ser parametrizado através da transformação logarítmica:

$$\log(a) + bt + \epsilon_t$$

Os modelos aqui considerados serão lineares. Importante é a linearidade nos parâmetros.

Ex: Um modelo que pode ser usado para **ST** descrevendo uma curva S.

$$\log(S_t) = a + b/t,$$

que embora não seja linear em t, é linear em a e b, após transformação logarítmica.

2.1. Revisão de Modelos Lineares

Variável dependente: Y
Variáveis explicativas: $X_1, ..., X_p$
Relacionadas através de: $Y_t = \beta_0 + \sum_{i=1}^{p} \beta_i x_{it} + \varepsilon_t$ onde $t = 1, \ldots, n$

Podemos escrever:

$$S_t = \underline{x}_t^T \underline{\beta}$$

$$\underline{x}_t^T = (1, x_{1t}, ..., x_{pt})$$

$$\underline{\beta}_t^T = (\beta_0, \beta_1, ..., \beta_p)$$

Os erros ε_t são não-correlacionados com: $E[\varepsilon_t] = 0 \quad \text{e} \quad V[\varepsilon_t] = \sigma^2$
Comumente se assume também que os ε_t são normais.

Usando notação matricial pode-se escrever:

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$$

$$Y^T = (y_1, ..., y_n) \qquad X^T = (\underline{x}_1, ..., \underline{x}_n) \qquad \underline{\varepsilon}^T = (\varepsilon_1, ..., \varepsilon_n)$$

onde:

$$E[\underline{\varepsilon}] = \underline{0} \qquad e \qquad V[\underline{\varepsilon}] = \sigma^2 I_n$$

O critério para estimação de $\underline{\beta}$ é a minimização de:

$$S(\underline{\beta}) = \sum_{t=1}^{n} (y_t - \underline{x}_t^T \underline{\beta})^2$$

Para $\underline{\beta}$ minimizar $S(\underline{\beta})$ é preciso que $X^T X \hat{\underline{\beta}} = X^T \underline{Y}$
Se X tem posto máximo (var. explicativas são linearmente independentes), então $X^T X$ pode ser invertido fornecendo $\hat{\underline{\beta}} = (X^T X)^{-1} X^T \underline{Y}$

Propriedades:

i) $E[\underline{\hat{\beta}}] = \underline{\beta}$ $\qquad e \qquad$ $V[\underline{\hat{\beta}}] = \sigma^2(X^T X)^{-1}$.

ii) Se os erros tem distribuição normal então $\underline{\hat{\beta}}$ também tem.

iii) Os valores ajustados de Y são $\underline{\hat{Y}} = X\underline{\hat{\beta}}$ e os resíduos $\underline{\varepsilon}$ são dados por $\underline{\varepsilon} = \underline{Y} - \underline{\hat{Y}}$.

iv) Definindo:

- Soma Total dos Quadrados por $STQ = \sum_{t=1}^{n}(Y_t - \bar{Y})^2$
- Soma do Erros Quadrados por $SEQ = \sum_{t=1}^{n}(Y_t - \underline{\hat{Y}_t})^2$
- Soma dos Quadrados da Regressão por $SQR = \sum_{t=1}^{n}(\hat{Y}_t - \bar{Y}])^2$

Temos que $STQ = SEQ + SQR$.
Normalmente mede-se o ajuste da regressão através de

$$R^2 = 1 - \frac{SEQ}{STQ}$$

v) Se σ^2 é desconhecido, ele é estimado por $S^2 = \frac{SEQ}{n-p-1}$ onde $\frac{SEQ}{\sigma^2} \sim \chi^2_{n-p-1}$ e portanto $E[S^2] = \sigma^2$ e a variância de $\underline{\hat{\beta}}$ é estimada por $\hat{V}[\hat{\beta}] = S^2(X^TX)^{-1}$

vi) Usando o fato que $t_i = \frac{\hat{\beta}_i - \beta_i}{S\sqrt{c_{ii}}} \sim t_{n-p-1}(0,1)$ onde c_{ii} é o i-ésimo elemento da diagonal de $(X^TX)^{-1}$ pode-se obter:

- Intervalo de confiança $100(1-\alpha)$ para β_i da forma:

$$[\hat{\beta}_i - t_{\alpha/2,n-p-1} S\sqrt{c_{ii}} \quad ; \quad \hat{\beta}_i + t_{\alpha/2,n-p-1} S\sqrt{c_{ii}}];$$

onde $t_{\alpha/2,n-p-1}$ é o percentil $100(1-\alpha/2)$ da t com $n-p-1$ g.l.

- Teste de nível de α para testar $H : \beta_i = 0$ que rejeita H se $\frac{|\hat{\beta}_i|}{S\sqrt{c_{ii}}} > t_{\alpha/2,n-p-1}$

vii) O teste simultâneo da regressão testa a hipótese

$$H_0 : \beta_1 = \beta_2 = ... = \beta_p \quad \text{vs.} \quad H_1 : \text{algum} \beta_i \neq 0.$$

Ele rejeita H_0 se

$$\frac{\frac{SQR}{p}}{\frac{SEQ}{(n-p-1)}} > F_\alpha(p, n-p-1),$$

onde $F_\alpha(p, n-p-1)$ é o percentil $100(1-\alpha)\%$ da distribuição F com p e $n-p-1$ graus de liberdade.

2.2. Previsão

Suponha que se deseja prever Y_s baseado nos valores $x_{1s}, \ldots, x_{ps}$. Denotando o preditor por $\widehat{Y}_s$ e o erro de previsão é dado por $e_s = Y_s - \widehat{Y}_s$, se utilizarmos como critério a minimização do EQM dado por

$$E[e_s^2] = E[Y_s - \widehat{Y}_s]^2$$

obtemos $\widehat{Y}_s = E[Y_s] = \underline{x}_s^T \underline{B}$ onde $x_s^T = (x_{1s}, \ldots, x_{ps})$.

Como $\underline{B}$ é desconhecido, podemos substituí-lo por seu estimador fornecendo a previsão

$$\widehat{Y}_s = \underline{x}_s^T \underline{B}$$

Propriedades:

i) $E[\widehat{Y}_s] = \underline{x}_s^T \underline{B}$ e $V(\widehat{Y}_s) = \sigma^2 \underline{x}_s^T (X^T X)^{-1} \underline{x}_s$

ii) $E[Y_s - \widehat{Y}_s] = 0$ e $V(Y_s - \widehat{Y}_s) = \sigma^2[1 + \underline{x}_s^T (X^T X)^{-1} \underline{x}_s]$
Como σ^2 é desconhecido, estima-se $V(Y_s - \widehat{Y}_s)$ por $\widehat{V}(Y_s - \widehat{Y}_s)$ dado por $S^2 \left[1 + \underline{x}_s^T (X^T X)^{-1} \underline{x}_s\right]$

iii) Se os erros são normais então $\frac{Y_s - \widehat{Y}_s}{\sqrt{\widehat{V}(Y_s - \widehat{Y}_s)}}\ t_{n-p-1}(0,1)$ O intervalo de confiança $100(1-\alpha)\%$ para Y_s é dado por:

$$\left[\widehat{Y}_s - t_{\alpha/2,n-p-1}\sqrt{\widehat{V}(Y_s - \widehat{Y}_s)} \quad ; \quad \widehat{Y}_s + t_{\alpha/2,n-p-1}\sqrt{\widehat{V}(Y_s - \widehat{Y}_s)}\right]$$

Correlação serial entre os erros

Normalmente em regressão, assume-se que os erros ϵ_t não são correlacionados.

Em dados de séries temporais é razoável esperar que isso não aconteça, i.e., o erro no tempo t, ϵ_t, esteja relacionado aos erros contíguos, ϵ_{t-1} e ϵ_{t+1}.

O não reconhecimento das correlação podem levar a ajustes incorretos. Importante estudar as autocorrelações da série, ρ_k.

Já vimos que ρ_k's dadas por:

$$\frac{\sum_{t=1}^{n-k}(Y_t - \overline{Y})(Y_{t+k} - \overline{Y})}{\sum_{t=1}^{n}(Y_t - \overline{Y})^2}$$

Pode ser mostrado que se $\rho_k = 0$ e n é grande, vale aproximadamente que

$$r_k \sim N(0, n^{-1})$$

e portanto o teste de nível $\alpha = 0,05$ rejeita a hipótese $\rho_k = 0$ se $\sqrt{n} \mid r_k \mid > 1,96$.

Observe que temos de fazer vários testes simultâneos e o nível para um teste global é outro.

2.3. Modelos Sazonais

Séries sazonais ocorrem com frequência em várias áreas:

- consumo mensal de eletricidade em uma dada região;
- produção mensal de leite;
- número de casamentos em cada mês;
- etc.

A sazonalidade muitas vezes tem padrão bem estabelecido e estável (consumo de eletricidade é maior no verão).

Muito da sazonalidade é devido à rotação da Terra em torno do Sol e ás decorrências climáticas.

É importante modelar para melhor compreender e estimar.

Normalmente, além de sazonalidade, a série está sujeita à tendências e ciclos. Isso vai ser visto na próxima seção.

Além disso, o espaço de tempo pra completar um ciclo sazonal (período) pode assumir vários valores. Se a sazonalidade é anual (mais comum) e os dados são trimestrais (período $= 4$), se os dados são mensais (período $= 12$) e se os dados são semanais (período $= 52$).

Vamos assumir aqui que série é modelada apenas pela sazonalidade e o período sazonal (denotado por s) será tomado como 12.

Vamos nos concentrar no caso mais comum: dados mensais e sazonalidade anual.

A sazonalidade pode ser modelada por indicadores sazonais ou por funções trigonométricas.

Modelagem com indicadores sazonais

$$S_t = \sum_{i=1}^{s} \beta_i x_{ti}$$

onde

$$x_{ti} = \begin{cases} 1, & \text{tempo } t \text{ corresponde ao período } i \\ 0, & \text{caso contrario.} \end{cases}$$

ou, mais comumente,

$$S_t = \beta_0 + \sum_{i=1}^{s} \beta_i x_{ti}$$

β_0 representa o nível médio da série
β_i representa o efeito do período i na série, $i = 1, 2, ..., s$

O modelo acima tem $s + 1$ parâmetros mas a matriz não tem posto máximo: a primeira coluna é a soma das outras.

É necessário impor alguma restrição sobre os parâmetros. As mais comuns são:

i) Omitir o nível médio β_0

ii) Fazer com que um dos β's seja zero. Nesse caso, β_0 passa a ser o nível da série para esse período e β_i é o efeito do período i comparado com o nível do período escolhido;

iii) Restringir $\sum_{i=1}^{s} \beta_i = 0$. Nesse caso, a soma dos efeitos é nula , ou equivalentemente $S_t = -\sum_{i=1}^{s} \beta_i x_{ti}$.

Adotando a opção (iii) temos que

$$
\begin{aligned}
S_t &= \beta_0 + \sum_{i=1}^{s-1} \beta_i x_{ti} + \beta_s x_{ts} \\
&= \beta_0 + \sum_{i=1}^{s-1} \beta_i x_{ti} - \sum_{i=1}^{s-1} \beta_i x_{ts} \\
&= \beta_0 + \sum_{i=1}^{s-1} \beta_i (x_{ti} - x_{ts}) \\
&= \beta_0 + \sum_{i=1}^{s-1} \beta_i x_{ti}^*
\end{aligned}
$$

onde

$$
x_{ti}^* = \begin{cases} 1, & t \text{ corresponde ao período } i \\ \text{-}1, & t \text{ corresponde ao período } s \\ 0, & \text{caso contrario.} \end{cases}
$$

E a teoria de regressão pode ser utilizada.

3. Capitulo

4. Modelos ARIMA

Classe muito popular de modelos bastante desenvolvida na década passada. Depende apenas dos dados para a especificação do modelo.

ARIMA - Auto Regressivo Intregado Moving Average (Médias Móveis).

O ciclo de inferência consiste em:

$$\text{Identificação} \longrightarrow \text{Estimação} \longrightarrow \text{Verificação} \longrightarrow \text{Previsão.}$$

4.1. Introdução

Recordação do Capítulo 1 (operadores):

i) $By_t = y_{t-1}$; $B^m y_t = y_{t-m}$ (backward)

ii) $Fy_t = y_{t+1}$; $F^m y_t = y_{t+m}$ (forward)

iii) $\Delta y_t = y_t - y_{t-1} = (1-B)y_t$; $\Delta = (1-B)$

iv) $\mathrm{S}y_t = \sum_{j=0}^{\infty} y_{t-j} = y_t + y_{t-1} + \ldots = (1 + B + B^2 + \ldots)y_t$.
Logo, $S = (1 + B + B^2 + \ldots)$.
Como $(1+B+B^2+\ldots)(1-B) = 1$, temos que $S = (1-B)^{-1} = \Delta^{-1}$

4.2. Modelo de Filtro Linear

Os modelos aqui estudados são supostos gerados por filtros lineares. Esquema de um filtro:

$$a_t \longrightarrow \Psi(B) \longrightarrow y_t$$

onde:

a_t é a entrada do filtro (ruído branco);

Ψ é a função de transferência do filtro;

y_t é a saída do filtro.

Matematicamente, $y_t = \Psi(B)a_t = a_t + \Psi_1 a_{t-1} + \Psi_2 a_{t-2} + \ldots$

Se $\mathrm{E}[y_t] = \mu$, trabalhamos com (y_t-μ) ao invés de y_t.

Como a_t é ruído branco, temos que:

- $\mathrm{E}[a_t] = 0, \forall t$;
- $\mathrm{Var}[a_t] = \sigma^2, \forall t$;
- $\mathrm{E}[a_t a_s] = 0, \forall (t, s)$.

$E[y_t] = \mu + E[a_t + \Psi_1 a_{t-1} + \Psi_2 a_{t-2} + \ldots]$. Como $E[a_t] = 0$, temos que $E[y_t] = \mu$ se $\sum_{j=1}^{\infty} \Psi_j < \infty$

Logo, y_t é estacionária com média μ se a soma acima converge. Caso contrário, μ serve com referência de posição da série não estacionária.

- $\gamma_0 = \mathrm{V}[y_t] = \sum_{j=0}^{\infty} \mathrm{V}[\Psi_j a_{t-j}] = \sigma^2 \sum_{j=0}^{\infty} \Psi_j^2$

- $\gamma_1 = \mathrm{Cov}[y_t, y_{t-j}] = \mathrm{Cov}[\sum_{k=0}^{\infty} \Psi_k a_{t-k}, \sum_{k=0}^{\infty} \Psi_k a_{t-k-j}] = \Psi_0 \Psi_j \sigma^2 + \Psi_1 \Psi_{j+1} \sigma^2 + \ldots = \sigma^2 \sum_{k=0}^{\infty} \Psi_k \Psi_{k+j}$.

Se escrevermos y_t como função de $y_{t-1}, y_{t-2}, \ldots$ ao invés de $a_{t-1}, a_{t-2}, \ldots$, temos

$$y_t = \Pi_1 y_{t-1} + \Pi_2 y_{t-2} + \ldots + a_t$$

Daí,

$$(1 - \sum_{j=1}^{\infty} \Pi_j B^j) y_t = a_t$$

ou Π(B)$y_t = a_t$, com Π(B) = 1 - Π_1B - $\Pi_2 B^2$ - ....

Como $y_t = \Psi(B) a_t$ e $\Pi(B)\Psi(B) a_t = a_t$. Logo:

$$\Pi(B)\Psi(B) = 1 \longrightarrow \Pi(B) = \Psi^{-1}(B).$$

Pode-se obter Π_j's como função dos Ψ_j's e vice-versa.

4.3. Estacionariedade e Invertibilidade

Suponha que $\Psi_j = \Phi^j$

$$\sum \Psi_j = \sum \Phi^j = 1/(1-\Phi)$$

A série é estacionária. (Se Φ=1, a série é não estacionária com $y_t = a_t + a_{t-1} + \ldots$ que é o mesmo que dizer que $y_t = a_t + y_{t-1}$. Nesse caso, y_t é um passeio aleatório. No entanto, $\Delta y_t = a_t$ é um a série estacionária!)

Não é dificil obter que

$$\gamma_0 = \sigma^2/(1 \text{ - } \Phi^2)$$

e

$$\gamma_j = \Phi^j j^2/(1 \text{ - } \Phi^2)$$