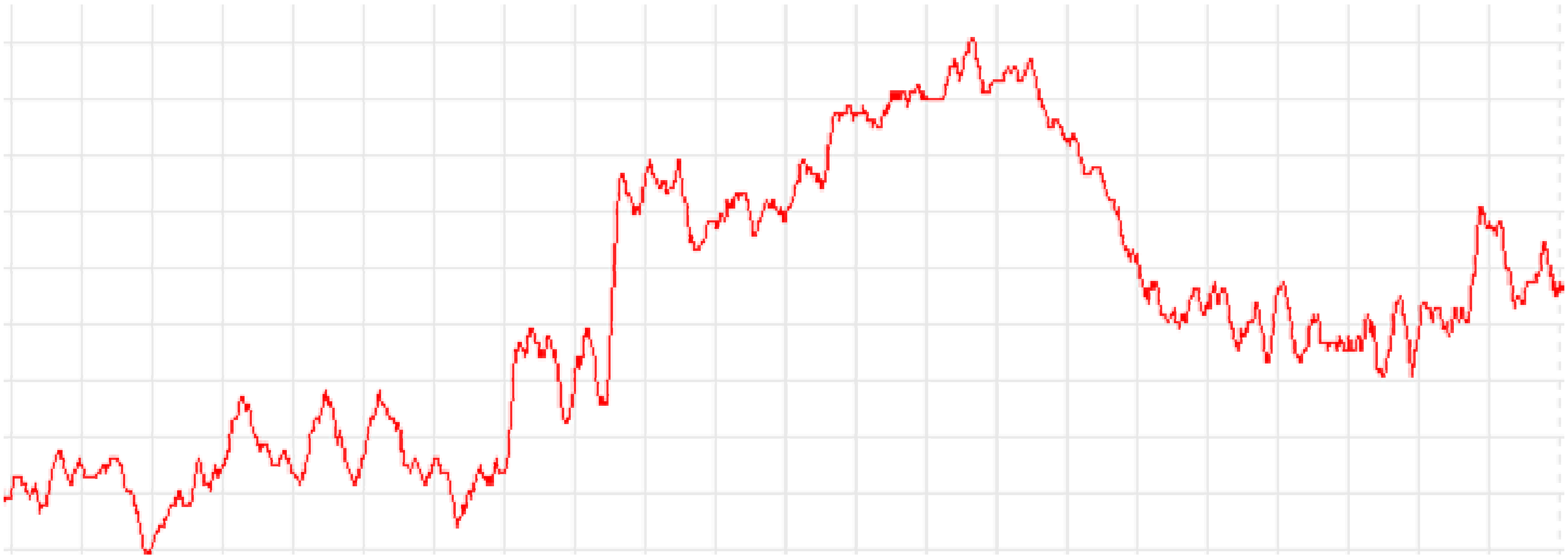

Análise de Séries Temporais (AST)

Prof. Lorí Viali, Dr. – UFRGS – Instituto de Matemática - Departamento de Estatística

Conceito

Na Estatística existem situações onde os dados de interesse são obtidos em instantes sucessivos de tempo (minuto, hora, dia, mês ou ano), ou ainda num período contínuo de tempo, como acontece num eletrocardiograma ou sismógrafo.

Esses dados ordenados no tempo são denominados de série temporal.

Exemplos

(a) O valor diário do índice Bovespa;

(b) A temperatura mínima diária de Porto Alegre;

(c) A quantidade mensal de chuva numa determinada região;

(d) O índice mensal da inflação brasileira;

(e) A altura da maré no Porto de Rio Grande;

(f) O registro do eletrocardiograma (ECG) de uma pessoa;

(g) O movimento da crosta terrestre, medido através de um sismógrafo.

Os exemplos de (a) a (d) mostram séries discretas, enquanto que os de (e) a (g) ilustram séries contínuas.

As séries discretas são observadas em instantes eqüiespaçados de tempo, isto é, em intervalos de tempo t = 1, 2, ..., n, enquanto que as séries contínuas são observadas em intervalos de tempo contínuos

As séries contínuas são discretizadas para poderem ser analisadas. Os valores de uma série também podem ser obtidos por agregação, como, por exemplo, as temperaturas diárias são somadas e é obtido a média, ou no caso, de chuvas as precipitações diárias são somadas para se obter um valor mensal.

Objetivos

Dada a ST

$Y_1, Y_2, \ldots, Y_n$, observada nos instantes:

$t_1, t_2, \ldots, t_n$, pode-se estar interessado em:

(a) Investigar o mecanismo gerador da série;

(b) Fazer previsões sobre valores futuros;

(c) Descrever apenas o seu comportamento;

(d) Procurar por periodicidades.

Análise

Existem basicamente dois enfoques utilizados para se analisar uma ST.

O primeiro enfoque é através do domínio temporal com o uso de modelos paramétricos.

No segundo enfoque é utilizado o domínio das freqüências através de modelos não paramétricos.

Modelos

Os modelos ARMA e ARIMA estão entre os paramétricos.

A análise espectral está incluída entre os modelos não paramétricos.

Estacionariedade

É uma das suposições mais freqüentes sobre uma série temporal. Significa que ela evolui no tempo em torno de um valor constante (médio), refletindo uma estabilidade.

Na prática a grande maioria das ST apresentam algum tipo de tendência, sendo o caso mais simples a tendência linear.

A maioria das técnicas de análise supõe que a série é estacionária, assim é necessário retirar a tendência caso a série não seja estacionária.

Diferenças

A maneira mais simples de tornar uma série não estacionário em uma estacionária é através de diferenças sucessivas da série original.

A primeira diferença é definida por:

$$\Delta Y_t = Y_t - Y_{t-1}$$

A segunda diferença seria:

$$\Delta^2 Y_t = \Delta[\Delta\ Y_t\ ] = \Delta[Y_t - Y_{t-1}] =$$

$$= Y_t - 2Y_{t-1} + Y_{t-2}$$

Em geral, tem-se:

$$\Delta^n Y_t = \Delta[\Delta^{n-1}\ Y_t\ ]$$

Sazonalidade

É comum um conjunto de dados ser medido em subperíodos do tempo apresentar sazonalidade. É possível fazer um ajustamento sazonal dos dados e depois usar um modelo não sazonal. Outra opção é fazer a previsão usando diretamente o modelo sazonal ao invés de remover a sazonalidade da série e utilizar um modelo não sazonal.

Objetivos

Apresentar os principais métodos de previsão (*forecasting*) em séries temporais. Existem duas categorias:

(i) Automáticos: aplicados com o auxílio de um software;

(ii) Não automático: exigem a intervenção de pessoal especializado.

Modelos

Um modelo tradicional de uma ST é que ela possa ser escrita como a soma de três componentes:

$$Y_t = T_t + S_t + A_t,\ t = 1, 2, \ldots, n.$$

Onde:

T_t = tendência;

S_t = a sazonalidade e

A_t = componente aleatória.

Após removidas as componentes da tendência (T_t) e a sazonal (S_t), sobra apenas a componente aleatória (A_t). A suposição normal é que A_t seja um ruído branco, isto é, com média zero e variância constante.

Esse modelo é denominado de aditivo que é adequado quando S_t não depende das outras componentes.

Se a sazonalidade varia com a tendência um modelo mais adequado é o multiplicativo:

$$Y_t = T_t.S_t.A_t$$

Que pode ser transformado em um aditivo aplicando logaritmos:

$$ln(Y_t) = ln(T_t) + ln(S_t) + ln(A_t)$$

Tendências

Supondo uma ST com modelo aditivo:

$Y_t = T_t + S_t + A_t$, $t = 1, 2, \ldots, n$.

Para retirar a sazonalidade é necessário estimar S_t e determinar a série sazonalmente ajustada:

$$\hat{Y}_t = Y_t - \hat{S}_t$$

Para estimar a tendência é necessário supor que a componente sazonal não exista ou que tenha sido removida.

Existem várias formas de se estimar a tendência T_t. Os mais utilizados são:

(i) Ajustar uma função do tipo polinomial, exponencial ou outra função suave de t;

(ii) Suavizar ou filtrar os valores da série ao redor de um ponto para estimar a tendência naquele ponto.

Supondo que a tendência estimada seja $\hat{T}_t$ *removendo esssa componente podemos obter a série ajustada livre da tendência:*

$$\hat{Y}_t = Y_t - \hat{T}_t$$

Outra opção para eliminar a tendência é tomar a série das diferenças como já foi visto anteriormente.

Tendência Polinomial

O problema maior com este tipo de ajuste é que embora ele possa ter um boa aderência nem sempre fornece boas previsões.

Suponha que:

$T_t = \alpha_0 + \alpha_1 t + \ldots + \alpha_m t^m$, onde o grau do polinômio deve ser bem menor do que o número de observações "n".

Para estimar os parâmetros α_i deve-se utilizar o MMQ, procedimento já conhecido.

Para $k = 1$, tem-se uma equação linear com as equações normais já conhecidas. Para $k \geq 2$, deve-se utilizar procedimentos matriciais.

Para estimar a tendência vamos supor que a componente sazonal não esteja presente e que o modelo é aditivo, isto é:

$$Y_t = T_t + A_t$$

onde A_t é a componente irregular.

A estimação da tendência de uma série temporal é feita através da regressão entre as variáveis "t" e Y_t.

Para estimar os parâmetros α_i utiliza-se o método dos mínimos quadrados, isto é, minimiza-se a função:

$$f(\alpha_0, \alpha_1, \ldots, \alpha_m) = \sum_{t=1}^{n} (T_t - \alpha_0 - \alpha_1 t - \ldots - \alpha_m t^m)^2$$

obtendo-se os estimadores dos mínimos quadrados:

De onde, segue, então que a tendência estimada será: $\hat{T}_t = a_0 + a_1 t + \ldots + a_m t^m$

Assim se tivermos o conjunto de n pontos

$(t,\ Y_t)$, então o ajuste do polinômio:

$$f(\alpha_0,\alpha_1,\ldots,\alpha_m)=\sum_{t=1}^{n}(Y_t-\alpha_0-\alpha_1 t-\ldots-\alpha_m t^m)^2$$

Será dado por:

$$-2\sum_{t=1}^{n}(Y_t-\sum_{j=0}^{m}a_j t^j)t^j=0$$

Para que esse valor seja mínimo, devemos ter:

$$\frac{\partial f}{\partial a_k} = 0, \ \ para\ k = 1, 2, ..., m.$$

Ou seja:

$$-2\sum_{t=1}^{n}\left(Y_t - \sum_{j=0}^{m} a_j t^j\right)t^k$$

Para cada k = 0, 1, 2, ...,m, pois:

$$\frac{\partial f}{\partial a_k}\sum_{j=0}^{m} a_j t^j) = \frac{\partial}{\partial a_k}(a_0 + a_1 t + \ldots + a_k t^k + \ldots + a_m t^m)$$

Que pode ser escrita como:

$$\sum_{t=1}^{n}\sum_{j=0}^{m} a_j t^{j+k} = \sum_{t=1}^{n} T_t t^k$$

Ou ainda:

$$\sum_{t=1}^{n}(\sum_{j=0}^{m} t^{j+k}) a_j = \sum_{t=1}^{n} T_t t^k$$

Para cada $k = 0, 1, 2, \ldots, m$.

Isso resulta em um sistema de m + 1 equações lineares cujos valores desconhecidos são os m + 1 coeficientes a_i do polinômio sendo ajustado.

Esse sistema pode ser escrito na forma matricial da seguinte maneira:

$X^T X A = X^T T_t$, onde:

$$X = \begin{pmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^m \\ 1 & t_1 & t_2^2 & \cdots & t_2^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_n & t_n^2 & \cdots & t_n^m \end{pmatrix}$$

X^T *é a sua transposta e:*

$$T_t = \begin{pmatrix} T_1 \\ T_2 \\ \cdots \\ Tn \end{pmatrix} \quad e \quad A = \begin{pmatrix} a_0 \\ a_1 \\ \cdots \\ a_m \end{pmatrix}$$

Essas equações são denominadas de equações normais.

O sistema anterior pode ser escrito como:

$X^T(XA - T_t) = 0$, *onde o vetor no interior do parênteses são os resíduos. Esse vetor é normal (ortogonal) aos vetores formados pelas linhas dos elementos da matriz* X^T.

Exercício

Determinar a tendência da série "vendas - em milhões" para os dados da planilha "Exercício_um"

(a) Ajustando uma polinômio de grau 1.

(b) Ajustando um polinômio de grau 2.

(c) Determine qual das duas opções proporciona o melhor ajuste.

Tendência Linearizável

Algumas funções, que a primeira vista não são lineares, podem ser linearizadas por alguma transformação sobre uma ou mesmo sobre as duas variáveis (t e Y_t). Em geral esta transformação consiste em trabalhar com os logaritmos de uma ou de ambas as variáveis.

Tendência Exponencial

A tendência exponencial pode ser caracterizada por uma equação do tipo:

$$T_t = \alpha\beta^t$$

Aplicando-se logaritmos aos dois lados desta equação vem:

$$ln(T_t) = ln(\alpha\beta^t) = ln(\alpha) + ln(\beta).t,$$

que é uma equação do tipo linear.

Desta forma para se estimar "α" e "β", ajusta-se a $ln(T_t)$ e t uma equação linear com $\beta_0 = ln(\alpha)$ e $\beta_1 = ln(\beta)$. Os valores estimados de "α" e "β" serão:

$$a = e^{\hat{\beta}_0} \quad e \quad b = e^{\hat{\beta}_1}$$

Tendência Geométrica

A tendência geométrica de uma série temporal pode ser avaliada por uma equação do tipo:

$$T_t = \alpha t^{\beta}$$

Aplicando-se logaritmos aos dois lados da expressão acima vem:

$$ln(T_t) = ln(\alpha t^{\beta}) = ln(\alpha) + \beta . ln(t)$$

Neste caso para estimar "α" e "β", ajusta-se aos logaritmos de Y_t e aos logaritmos de t uma equação linear com β_0 = $ln(\alpha)$ e $\beta_1 = \beta$. Os valores estimados "a" e "b" serão: $a = e^{\hat{\beta}_0}$ e $b = \hat{\beta}_1$

Tendência Logística

A tendência logística de uma série temporal pode ser avaliada por uma equação do tipo:

$T_t = 1/\alpha+\beta t$

Nesse caso para linearizar a série basta fazer $T' = 1/T$

Sazonalidade

A sazonalidade para uma série avaliada em sub-períodos do ano pode ser eliminada por meio das seguintes operações:

$$Y_t^s = Y_t - \hat{S}_t$$

Se o modelo for aditivo e

$$Y_t^s = Y_t / \hat{S}_t$$

Se o modelo for multiplicativo.

O modelo multiplicativo é, muitas vezes, adequado para séries econômicas que apresentam uma tendência exponencial.

Aplicando-se logaritmos podemos obter o modelo aditivo.

Seja l o total de anos e k o número de sub-períodos do ano tal que $n = k.l$, conforme tabela.

$$\bar{Y}_i = \frac{1}{k}\sum_{j=1}^{k} Y_{ij} \quad i = 1, 2, \ldots, l$$

$$\bar{Y} = \frac{1}{lk}\sum_{i=1}^{l}\sum_{j=1}^{k} Y_{ij} \quad i = 1, 2, \ldots, l \text{ e } j = 1, 2, \ldots, k$$

Anos	1	2	...	k	Média
1	Y_{11}	Y_{12}	...	Y_{1k}	$\bar{Y}_1$
2	Y_{21}	Y_{22}	...	Y_{2k}	$\bar{Y}_2$
...	...	...	...	...	...
l	Y_{l1}	Y_{l2}	...	Y_{lk}	$\bar{Y}_l$
	Média Geral				$\bar{Y}$

Os índices estacionais serão:

$$S_i = \frac{\overline{Y}_i}{\overline{\overline{Y}}}$$

Para i = 1, 2, ..., k

Assim se o modelo for multiplicativo a estacionalidade pode ser retirada através desses índices.

Autocorrelação

Séries temporais tendem a apresentar autocorrelação entre o valor t e o valor t-k.

A função de autocorrelação pode ser estimada por:

$$r_k = \frac{\sum_{t=k+1}^{n} (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2}, \quad k = 1, 2, \ldots, n$$