

Análise de Séries Temporais

Fernando Lucambio

Departamento de Estatística
Universidade Federal do Paraná

Novembro, 2020

Começamos nossa discussão de regressão linear no contexto de séries temporais, assumindo alguma saída ou séries temporais dependentes, digamos que X_t , para $t = 1, \dots, n$, está sendo influenciado por uma coleção de entradas possíveis ou séries independentes, digamos $Z_{t_1}, Z_{t_2}, \dots, Z_{t_q}$, onde primeiro consideramos as entradas como fixas e conhecidas. Essa suposição, necessária para aplicar a regressão linear convencional, será relaxada posteriormente.

Expressamos essa relação através do modelo de regressão linear

$$X_t = \beta_1 Z_{t_1} + \beta_2 Z_{t_2} + \dots + \beta_q Z_{t_q} + W_t$$

onde $\beta_1, \beta_2, \dots, \beta_q$ são coeficientes de regressão fixos desconhecidos e $\{W_t\}$ é um erro aleatório ou processo de ruído consistindo em variáveis normais independentes e identicamente distribuídas (iid) com média zero e variância σ_w^2 .

Exemplo II.1. Estimando uma tendência linear.

Considere o preço mensal, em libras, de um frango nos EUA entre meados de 2001 e meados de 2016, ou seja, por 180 meses, digamos X_t , mostrado na figura abaixo. Há uma tendência ascendente óbvia na série, e podemos usar a regressão linear simples para estimar essa tendência ajustando o modelo

$$X_t = \beta_0 + \beta_1 z_t + W_t, \quad z_t = 2001 \frac{7}{12}, 2001 \frac{8}{12}, \dots, 2016 \frac{6}{12}.$$

Em mínimos quadrados ordinários (OLS), minimizamos a soma dos quadrados de erros

$$Q = \sum_{t=1}^n w_t^2 = \sum_{t=1}^n \left(x_t - (\beta_0 + \beta_1 z_t) \right)^2,$$

com respeito a $\beta_i, i = 0, 1$.

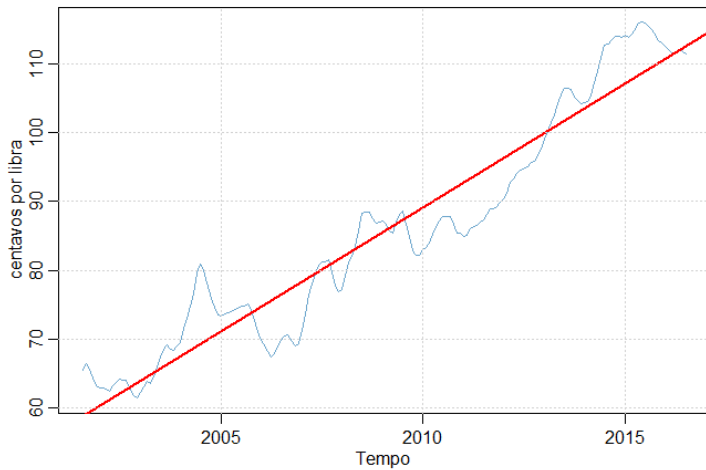
Exemplo II.1. Estimando uma tendência linear (continuação).

Neste caso, podemos usar cálculos simples para avaliar $\partial Q/\partial\beta_i = 0$ para $i = 0, 1$, e assim obter duas equações para resolver os β s.

As estimativas OLS dos coeficientes são explícitas e dadas por

$$\hat{\beta}_1 = \frac{\sum_{t=1}^n (x_t - \bar{x})(z_t - \bar{z})}{\sum_{t=1}^n (z_t - \bar{z})^2}, \quad \text{e} \quad \hat{\beta}_0 = \bar{x} - \hat{\beta}_1 \bar{z},$$

onde $\bar{x} = \sum_{t=1}^n x_t$ e $\bar{z} = \sum_{t=1}^n z_t$ são as respectivas médias amostrais.



Definição II.1. Critério de Informação de Akaike (AIC)

O Critério de Informação de Akaike ou AIC é definido como

$$AIC = \log(\hat{\sigma}_k^2) + \frac{n + 2k}{n},$$

onde $\hat{\sigma}_k^2 = SSE_k/n$, k é o número de parâmetros no modelo e n é o tamanho da amostra.

Formalmente, o AIC é definido como $-2 \log(\ell_k) + 2k$, em que ℓ_k é a log-verossimilhança maximizada e k é o número de parâmetros no modelo. Para o problema de regressão normal, o AIC pode ser reduzido para a forma dada acima. O AIC é uma estimativa da discrepância de Kullback-Leibler entre o modelo verdadeiro e um modelo candidato.

Uma forma corrigida, sugerida por Sugiura (1978) e expandida por Hurvich e Tsai (1989), pode ser baseada em resultados da distribuição de amostras pequenas para o modelo de regressão linear.

Definição II.2. Critério de Informação de Akaike Corrigido (AICc)

O Critério de Informação de Akaike Corrigido ou AICc é definido como

$$AICc = \log(\hat{\sigma}_k^2) + \frac{n+k}{n-k-2},$$

onde $\hat{\sigma}_k^2 = SSE_k/n$, k é o número de parâmetros no modelo e n é o tamanho da amostra.

Podemos também derivar um termo de correção baseado em argumentos bayesianos, como em Schwarz (1978).

Definição II.3. Critério de Informação Bayesiana (BIC)

O Critério de Informação de Akaike Corrigido ou AICc é definido como

$$BIC = \log(\hat{\sigma}_k^2) + \frac{k \log(n)}{n},$$

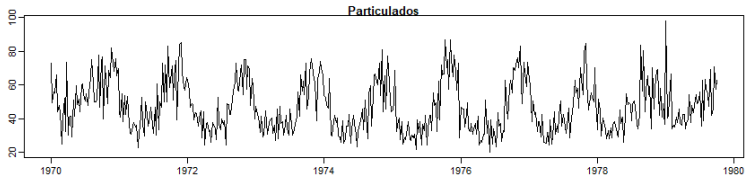
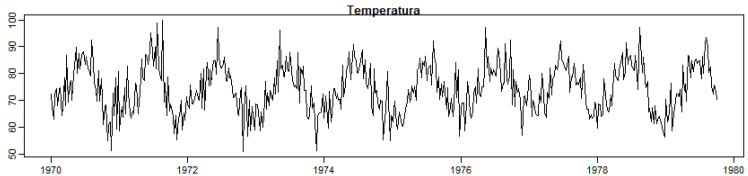
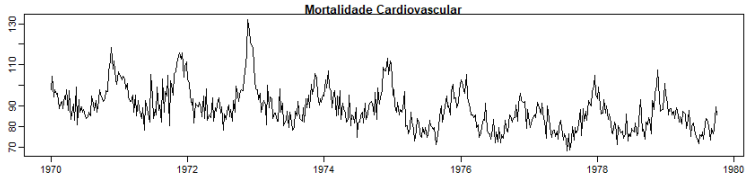
onde $\hat{\sigma}_k^2 = SSE_k/n$, k é o número de parâmetros no modelo e n é o tamanho da amostra.

O BIC também é chamado de Critério de Informação Schwarz (SIC). Vários estudos de simulação tendem a verificar que o BIC faz bem em obter a ordem correta em amostras grandes, enquanto o AICc tende a ser superior em amostras menores, onde o número relativo de parâmetros é grande. Veja McQuarrie e Tsai (1998) para comparações detalhadas.

Exemplo II.2. Poluição, Temperatura e Mortalidade

Os dados mostrados na figura abaixo foram extraídos de um estudo de Shumway et al. (1988) dos possíveis efeitos da temperatura e a poluição na mortalidade semanal no Condado de Los Angeles. Observe os fortes componentes sazonais em todas as séries, correspondentes às variações inverno-verão no hemisfério norte e a tendência de queda na mortalidade cardiovascular no período de 10 anos.

Nesta figura mostra-se o gráfico da mortalidade cardiovascular média semanal (topo), a temperatura média e a poluição por partículas ao fundo, no condado de Los Angeles. Existem 508 médias suavizadas de seis dias obtidas pelo lastreamento dos valores diários ao longo do período de 10 anos, de 1970 a 1979.



Uma matriz de dispersão, mostrada na figura a seguir, indica uma possível relação linear entre a mortalidade e as partículas poluentes e uma possível relação com a temperatura. Observe a forma curvilínea da curva de mortalidade por temperatura, indicando que temperaturas mais altas e temperaturas mais baixas estão associadas ao aumento da mortalidade cardiovascular.

Com base na matriz do gráfico de dispersão, entretemos, experimentalmente, quatro modelos em que M_t denota a mortalidade cardiovascular, T_t denota temperatura e P_t indica os níveis de partículas. Eles são

$$(A) \quad M_t = \beta_1 + \beta_2 t + W_t$$

$$(B) \quad M_t = \beta_1 + \beta_2 t + \beta_3 (T_t - T.) + W_t$$

$$(C) \quad M_t = \beta_1 + \beta_2 t + \beta_3 (T_t - T.) + \beta_4 (T_t - T.)^2 + W_t$$

$$(D) \quad M_t = \beta_1 + \beta_2 t + \beta_3 (T_t - T.) + \beta_4 (T_t - T.)^2 + \beta_5 P_t + W_t$$

onde ajustamos a temperatura por sua média, $T. = 74.26041$.

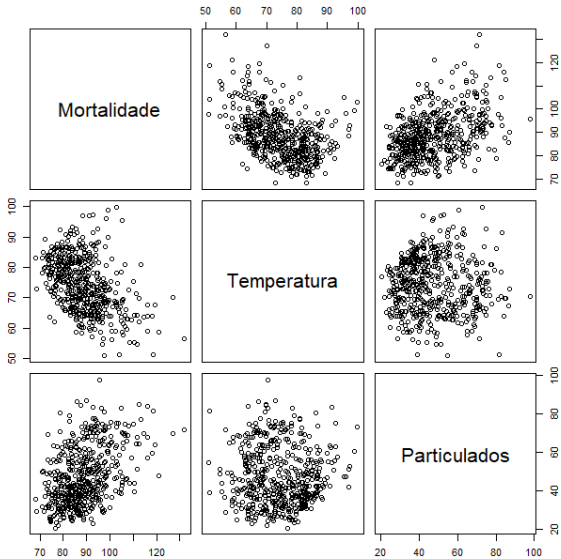


Tabela II.2. Estatísticas resumidas para modelos de mortalidade

Modelo	k	SSE	df	MSE	R^2	AIC	BIC
(A)	2	40020	506	79.0	0.21	5.38	5.40
(B)	3	31413	505	62.2	0.38	5.14	5.17
(C)	4	27985	504	55.5	0.45	5.03	5.07
(D)	5	20508	503	40.8	0.60	4.72	4.77

Exemplo II.3. Regressão com variáveis com defasagem

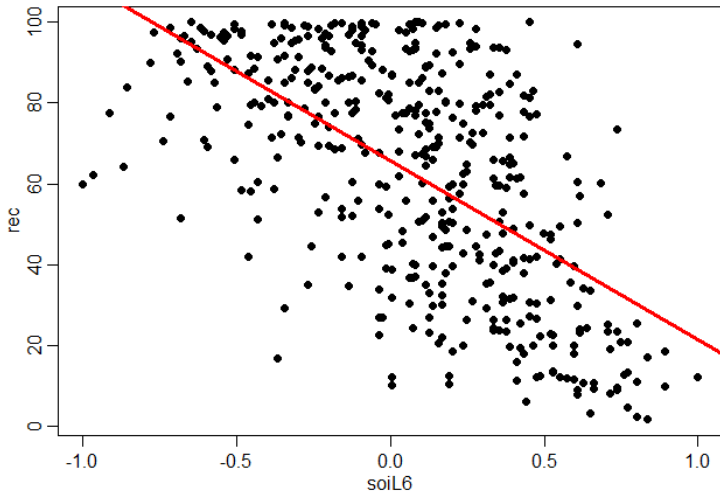
No Exemplo I.25, descobrimos que o Índice de Oscilação do Sul (SOI) medido no tempo $t - 6$ meses está associado à série Recrutamento no tempo t , indicando que o SOI lidera a série Recrutamento por seis meses. Embora haja evidências de que a relação não é linear, podemos considerar a seguinte regressão

$$R_t = \beta_1 + \beta_2 S_{t-6} + W_t,$$

onde R_t denota o Recrutamento no tempo t e S_{t-6} denota o SOI seis meses antes. Assumindo que a sequência W_t é um ruído branco, o modelo estimado é

$$\hat{R}_t = 65.790 - 44.283_{(2.781)} S_{t-6},$$

com $\hat{\sigma}_w = 22.5$ e 445 graus de liberdade.



Para obter qualquer análise estatística significativa dos dados de séries temporais será crucial que, ao menos, a média e a autocovariância satisfazem as condições de estacionariedade, pelo menos durante um período razoável de tempo. Muitas vezes, esse não é o caso, e mencionaremos alguns métodos para minimizar os efeitos da não-estacionariedade, para que as propriedades estacionárias da série possam ser estudadas.

Talvez a forma mais fácil de não-estacionariedade seja trabalhar com o modelo estacionário de tendência, em que o processo tem um comportamento estacionário em torno de uma tendência. Podemos escrever esse tipo de modelo como

$$X_t = \mu_t + Y_t,$$

onde X_t são as observações, μ_t denota a tendência e Y_t é um processo estacionário.

Muitas vezes, a tendência forte μ_t , obscurecerá o comportamento do processo estacionário. Assim, há alguma vantagem em remover a tendência como primeiro passo em uma análise exploratória de tais séries temporais.

As etapas envolvidas são obter uma estimativa razoável da componente de tendência, digamos $\hat{\mu}_t$, e depois trabalhar com os resíduos

$$\hat{Y}_t = X_t - \hat{\mu}_t.$$

Observação: as vezes, quando aplicado este procedimento o resultado \hat{Y}_t é um processo estacionário.

Exemplo II.4. Preços do Frango.

Aqui supomos que o modelo é da forma

$$X_t = \mu_t - Y_t,$$

onde, como sugerimos na análise dos dados de temperatura global apresentados no Exemplo II.1, uma linha reta pode ser um modelo razoável para a tendência, ou seja,

$$\mu_t = \beta_1 + \beta_2 t.$$

Nesse exemplo, estimamos a tendência usando mínimos quadrados ordinários e encontramos

$$\hat{\mu}_t = -7131 + 3.592t,$$

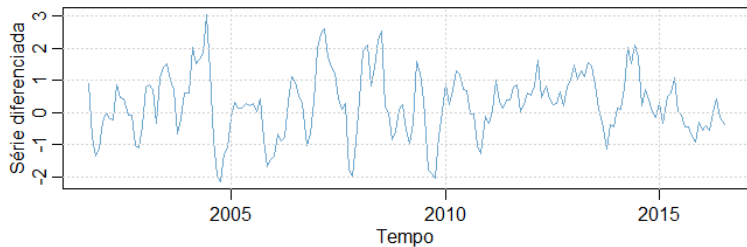
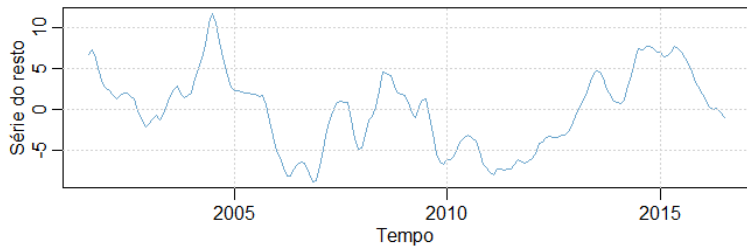
onde utilizamos t em vez de Z_t para o tempo.

Exemplo II.4. Preços do Frango (continuação).

A figura mostra os dados com a linha de tendência estimada sobreposta. Para obter as séries deduzidas, simplesmente subtraímos $\hat{\mu}_t$ das observações x_t , para obter a série

$$\hat{Y}_t = x_t + 7131 - 3.592t.$$

Como o termo de erro, Y_t , não é assumido como independente identicamente distribuído, o leitor pode achar que os mínimos quadrados ponderados deveriam ser aplicados neste caso. O problema é que não sabemos o comportamento de Y_t e é precisamente isso que estamos tentando avaliar neste estágio. Um resultado notável de Grenander e Rosenblatt (1957, Ch7), entretanto, é que sob condições moderadas em Y_t , para a regressão polinomial ou regressão periódica, assintoticamente, mínimos quadrados ordinários são equivalentes a mínimos quadrados ponderados em relação à eficiência.



No Exemplo I.11 vimos que um passeio aleatório também poderia ser um bom modelo para tendência. Ou seja, em vez de modelar a tendência como fixa, como no Exemplo II.4, poderíamos modelar a tendência como um componente estocástico usando o modelo de passeio aleatório com deslocamento aleatório,

$$\mu_t = \delta + \mu_{t-1} + W_t,$$

onde W_t é um ruído branco independente de Y_t . Se o modelo apropriado é $X_t = \mu_t + Y_t$, então, diferenciar os dados X_t , produz um processo estacionário; isso é,

$$\begin{aligned} X_t - X_{t-1} &= (\mu_t + Y_t) - (\mu_{t-1} + Y_{t-1}) \\ &= \delta + W_t + Y_t - Y_{t-1}. \end{aligned}$$

Pode-se mostrar $Z_t = Y_t - Y_{t-1}$ é estacionário. Isto é, porque Y_t é estacionário,

$$\begin{aligned}\gamma_Z(h) &= \text{Cov}(Z_{t+h}, Z_t) = \text{Cov}(Y_{t+h} - Y_{t+h-1}, Y_t - Y_{t-1}) \\ &= 2\gamma_Y(h) - \gamma_Y(h+1) - \gamma_Y(h-1),\end{aligned}$$

é independente do tempo.

Uma vantagem da diferenciação em remover a tendência é que nenhum parâmetro é estimado na operação de diferenciação. Uma desvantagem, no entanto, é que a diferenciação não produz uma estimativa do processo estacionário Y_t . Se uma estimativa de Y_t é essencial, então a desvantagem pode ser mais apropriada. Se o objetivo é coagir os dados para a estacionaridade, então a diferenciação pode ser mais apropriada. A diferenciação também é uma ferramenta viável se a tendência for fixa, como visto em exemplos.

Como a diferenciação desempenha um papel central na análise de séries temporais, ela recebe sua própria notação. A primeira diferença é denotada como

$$\nabla X_t = X_t - X_{t-1}.$$

Como vimos, a primeira diferença elimina uma tendência linear. Uma segunda diferença, isto é, a diferença da primeira diferença, pode eliminar uma tendência quadrática e assim por diante. Para definir diferenças maiores, precisamos de uma variação na notação que usaremos com frequência em nossa discussão dos modelos ARIMA na Parte III.

Definição II.4. Operador de retardo.

Definimos o operador de retardo como

$$BX_t = X_{t-1},$$

e o estendemos à potência $B^2X_t = B(BX_t) = BX_{t-1} = X_{t-2}$, e assim por diante. Então,

$$B^kX_t = X_{t-k}.$$

A ideia de um operador inverso também pode ser dada se precisarmos $B^{-1}B = 1$, de maneira que

$$X_t = B^{-1}BX_t = B^{-1}X_{t-1}.$$

Isto é, B^{-1} é o operador de deslocamento para a frente.

É claro que podemos então reescrever a operação de diferenciação como

$$\nabla X_t = (1 - B)X_t,$$

e podemos ampliar a noção ainda mais. Por exemplo, a segunda diferença se torna

$$\begin{aligned}\nabla^2 X_t &= (1 - B)^2 X_t = (1 - 2B + B^2)X_t \\ &= X_t - 2X_{t-1} + X_{t-2},\end{aligned}$$

pela linearidade do operador. Para verificar, basta ter a diferença da primeira diferença

$$\nabla(\nabla X_t) = \nabla(X_t - X_{t-1}) = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}).$$

Definição II.5. Diferença de ordem d

Definimos o operador de diferença de ordem d como

$$\nabla^d X_t = (1 - B)^d,$$

onde podemos expandir o operador $(1 - B)^d$ algebricamente para avaliar valores inteiros mais altos de d .

A primeira diferença é um exemplo de um filtro linear aplicado para eliminar uma tendência. Outros filtros, formados pela média dos valores próximos de X_t , podem produzir séries ajustadas que eliminam outros tipos de flutuações indesejadas, como na Parte III. A técnica de diferenciação é um componente importante do modelo ARIMA de Box e Jenkins (1970) e Box et. al. (1994), a ser discutido na Parte III.

Exemplo II.5. Diferenciando os preços do frango.

A primeira diferença da série dos preços do frango, também mostrada abaixo, produz resultados diferentes do que remover a tendência, prejudicando a regressão. Por exemplo, as séries diferenciadas não contém o ciclo longo, de cinco anos, que observamos nas séries selecionadas. O ACF desta série também é mostrado na figura mais embaixo. Nesse caso, as séries diferenciadas exibem um ciclo anual que foi obscurecido nos dados originais ou desmembrados.

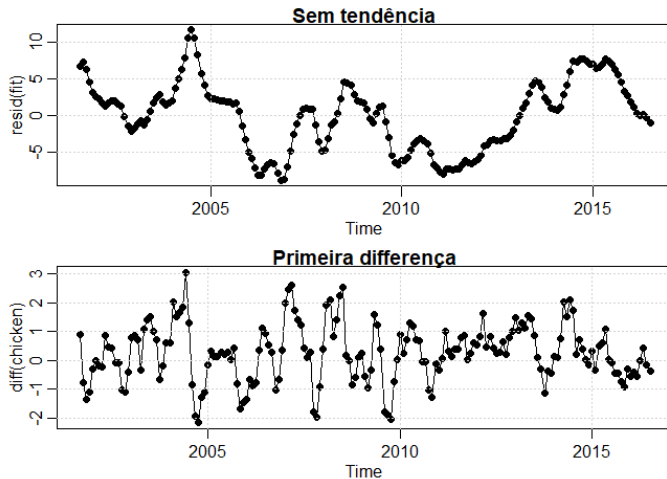


Figure 1: Série dos restos (topo) e diferenciada (inferior) de preços de frango. Os dados originais são mostrados no Exemplo II.1.

Frequentemente, há aberrações óbvias que podem contribuir para o comportamento não-estacionário e não-linear das séries. Em tais casos, transformações podem ser úteis para equalizar a variabilidade ao longo do comprimento de uma única série.

Uma transformação particularmente útil é

$$Y_t = \log(X_t),$$

que tende a suprimir flutuações maiores que ocorrem em partes da série em que os valores subjacentes são maiores.

Outras possibilidades são as transformações da família Box-Cox, definidas como

$$Y_t = \begin{cases} \frac{X_t^\lambda - 1}{\lambda}, & \text{caso } \lambda \neq 0, \\ \log(X_t), & \text{caso } \lambda = 0 \end{cases} .$$

Exemplo II.9. Regressão com variáveis defasadas.

No Exemplo II.3 fizemos a regressão de variáveis com defasagem do Recrutamento em SOI atrasado

$$R_t = \beta_0 + \beta_1 S_{t-6} + W_t.$$

No entanto, no Exemplo II.8, vimos que o relacionamento é não-linear e diferente quando o SOI é positivo ou negativo. Nesse caso, podemos considerar a adição de uma variável fictícia para considerar essa alteração. Em particular, ajustamos o modelo

$$R_t = \begin{cases} \beta_0 + \beta_1 S_{t-6} + W_t, & \text{caso } S_{t-6} < 0 \\ (\beta_0 + \beta_2) + (\beta_1 + \beta + 3)S_{t-6} + W_t, & \text{caso } S_{t-6} \geq 0 \end{cases}$$

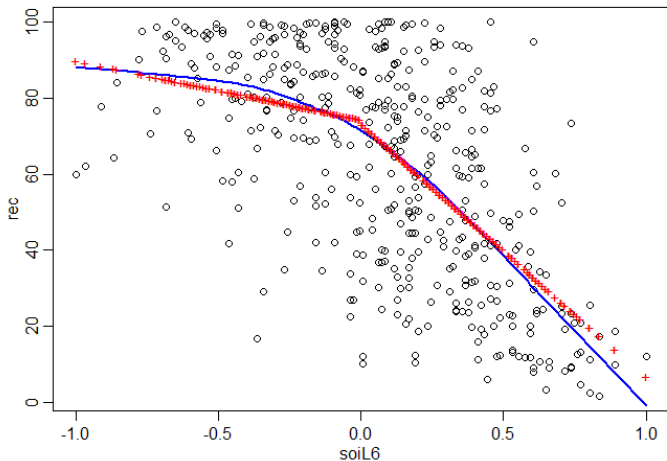


Figure 2: Gráfico do Recrutamento (R_t) vs SOI com defasagem de 6 meses (S_{t-6}) com os valores ajustados da regressão como pontos (+) e ajuste de baixa (-)

Na Seção 1.2, introduzimos o conceito de filtragem ou suavização de uma série temporal e, no Exemplo 1.9, discutimos o uso de uma média móvel para suavizar o ruído branco. Esse método é útil para descobrir certas características em uma série temporal, como componentes sazonais e tendências de longo prazo.

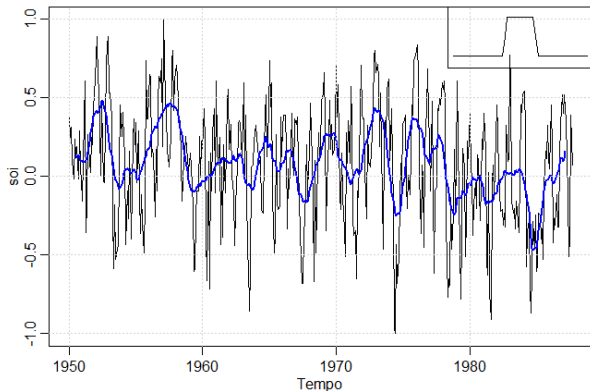
Em particular, se X_t representa as observações, então

$$m_t = \sum_{j=-k}^k a_j X_{t-j},$$

onde $a_j = a_{-j} \geq 0$ e $\sum_{j=-k}^k a_j = 1$ é uma média móvel simétrica dos dados.

Exemplo II.11. Suavização por Médias Móveis.

Série mensal de SOI suavizada utilizando a função m_t . Esse método específico remove (filtra) o ciclo de temperatura anual óbvio e ajuda a enfatizar o ciclo El Niño.



Exemplo II.12. Suavizamento por Kernel.

A suavização kernel é um suavizador de média móvel que usa uma função de peso ou kernel, para calcular a média das observações. A figura mostra a suavização kernel da série SOI, onde m_t é agora

$$m_t = \sum_{i=1}^n w_i(t) x_i,$$

onde

$$w_i(t) = K\left(\frac{t-i}{b}\right) / \sum_{j=1}^n K\left(\frac{t-j}{b}\right),$$

são os pesos e $K(\cdot)$ é uma função kernel. Este estimador, que foi originalmente explorado por Parzen (1962) e Rosenblatt (1956b). Neste exemplo, e normalmente, o kernel normal, $K(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)$, é utilizado.

