

Análise de Séries Temporais

Fernando Lucambio

Departamento de Estatística
Universidade Federal do Paraná

Novembro, 2020

Aqui cobrimos o modelo de regressão clássico com erros não correlacionados W_t . Nesta seção, discutimos as modificações que podem ser consideradas quando os erros são correlacionados. Ou seja, considere o modelo de regressão

$$Y_t = \sum_{j=1}^r \beta_j z_{t,j} + X_t,$$

onde X_t é um processo com função de covariância $\gamma_x(s, t)$. Em mínimos quadrados ordinários, a suposição é que X_t é ruído branco gaussiano, em que

$$\gamma_x(s, t) = 0 \quad \text{para } s \neq t \quad \text{e} \quad \gamma_x(t, t) = \sigma^2,$$

independente de t . Se este não for o caso, então os mínimos quadrados ponderados devem ser usados.

Exemplo III.44. Mortalidade, Temperatura e Poluição.

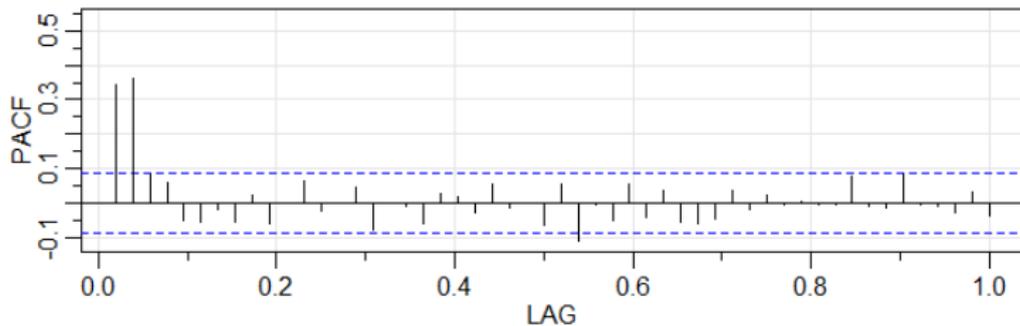
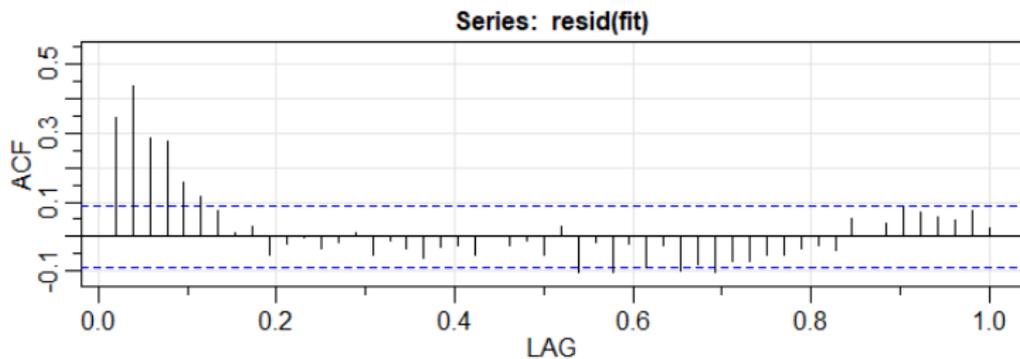
Consideramos as análises apresentadas no Exemplo II.2, relacionando a temperatura média ajustada T_t e os níveis de partículas P_t com a mortalidade cardiovascular M_t .

Consideramos o modelo de regressão

$$M_t = \beta_1 + \beta_2 t + \beta_3 T_t + \beta_4 T_t^2 + \beta_5 P_t + X_t,$$

onde, por enquanto, assumimos que X_t é ruído branco.

O ACF e o PACF amostrais dos resíduos do ajuste de mínimos quadrados ordinários são mostrados na figura abaixo e os resultados sugerem um modelo $AR(2)$ para os resíduos.



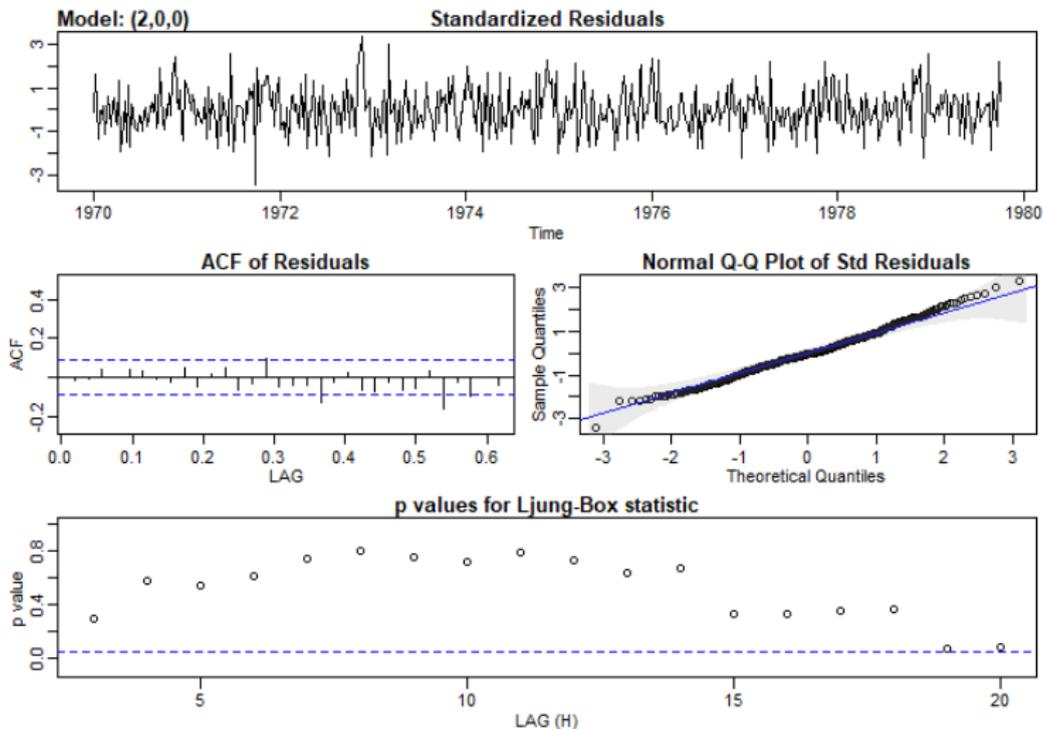
Exemplo III.44. Mortalidade, Temperatura e Poluição (continuação).

Nossa próxima etapa é ajustar o modelo de erro correlacionado mostrado acima, mas onde X_t é AR(2),

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + W_t,$$

e W_t .

```
> trend = time(cmort)
> temp = tempr - mean(tempr)
> temp2 = temp^2
> fit <- lm(cmort~trend + temp + temp2 + part, na.action=NULL)
> acf2(resid(fit), 52) # implies AR2
> sarima(cmort, 2,0,0, xreg=cbind(trend,temp,temp2,part))
```



Exemplo III.45. Regressão com variáveis defasadas.

No Exemplo II.29 ajustamos o modelo

$$R_t = \beta_0 + \beta_1 S_{t-6} + \beta_2 D_{t-6} + \beta_3 D_{t-6} S_{t-6} + W_t,$$

onde R_t é o Recrutamento, S_t é SOI e

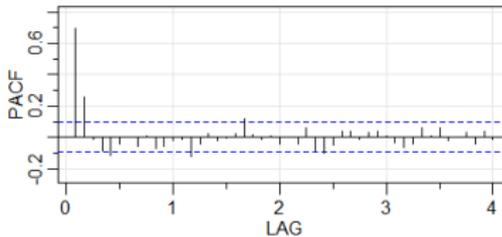
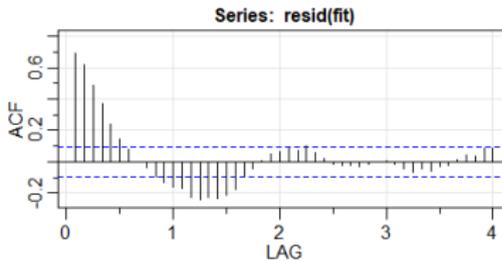
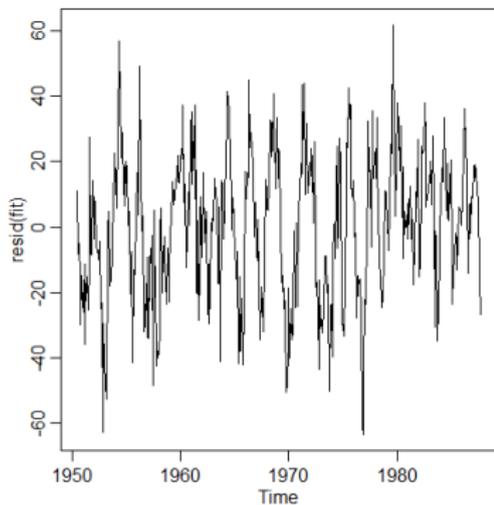
$$D_t = \begin{cases} 0, & \text{se } S_t < 0, \\ 1, & \text{se } S_t \geq 0, \end{cases}$$

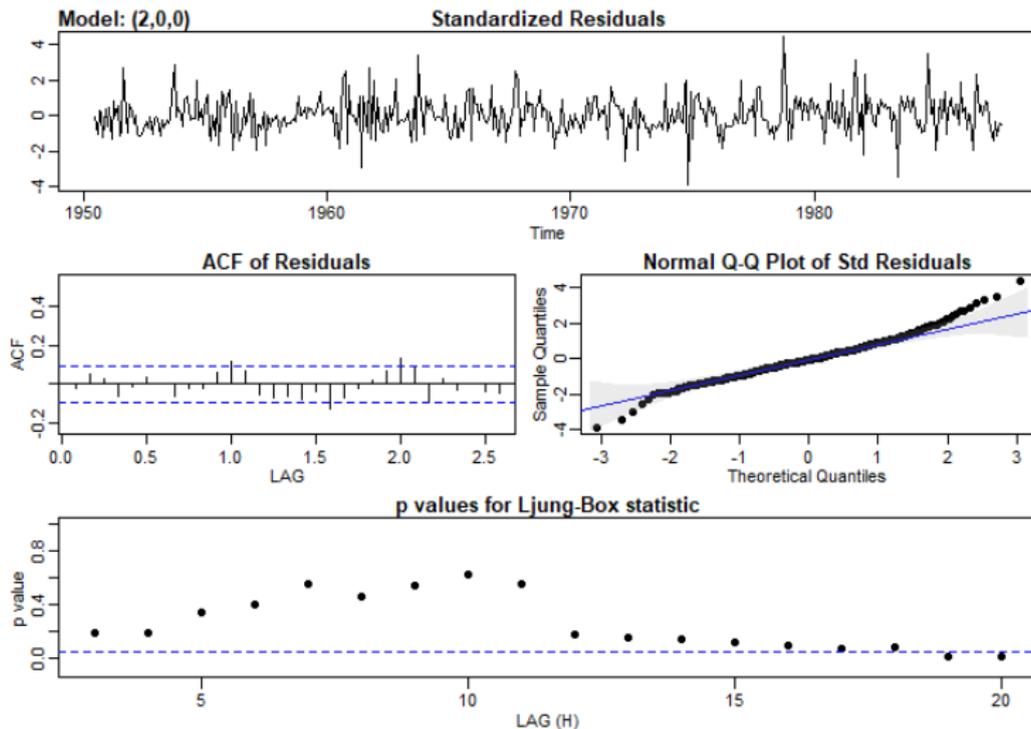
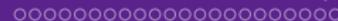
é uma variável fictícia.

No entanto, a análise dos resíduos indica que os resíduos não são ruído branco. O ACF e o PACF amostrais dos resíduos indicam que um modelo $AR(2)$ pode ser apropriado, o que é semelhante aos resultados do Exemplo III.44.

Exemplo III.45. Regressão com variáveis defasadas (continuação).

```
> dummy = ifelse(soi<0, 0, 1)
> fish = ts.intersect(rec, soiL6=lag(soi,-6),
                      dL6=lag(dummy,-6), dframe=TRUE)
> summary(fit <- lm(rec ~soiL6*dL6, data=fish, na.action=NULL))
> par(mfrow = c(1,1), mar=c(3,3,1,1), mgp=c(1.6,.6,0), pch=19)
> plot(resid(fit))
> acf2(resid(fit)) # indica AR(2)
> intract = soiL6*dL6 # termo de intercepto
> sarima(rec,2,0,0, xreg = cbind(soiL6, dL6, intract),
         details = FALSE)$ttable
```





Introduzimos várias modificações feitas no modelo *ARIMA* para considerar o comportamento sazonal e não-estacionário. Muitas vezes, a dependência do passado tende a ocorrer mais fortemente em múltiplos de alguns lags sazonais subjacentes. Por exemplo, com dados econômicos mensais, há um forte componente anual ocorrendo em lags que são múltiplos de $s = 12$, devido às fortes conexões de todas as atividades ao ano civil. Os dados obtidos trimestralmente exibirão o período anual repetitivo em $s = 4$ trimestres. Fenômenos naturais como a temperatura também têm componentes fortes correspondentes às estações do ano. Assim, a variabilidade natural de muitos processos físicos, biológicos e econômicos tende a combinar com as flutuações sazonais. Por causa disso, é apropriado introduzir polinômios autorregressivos de médias móveis que se identifiquem com as defasagens sazonais.

O modelo resultante de médias móveis autorregressivo puro resultante, digamos, $ARMA(P, Q)$, então assume a forma

$$\Phi_P(B^s)X_t = \Theta_Q(B^s)W_t,$$

onde os operadores

$$\Phi_P = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps}$$

e

$$\Theta_Q = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \dots + \Theta_Q B^{Qs},$$

são o operador autoregressivo sazonal e o operador de médias móveis sazonal das ordens P e Q , respectivamente, com o período sazonal s .

Exemplo III.46. Uma série AR sazonal.

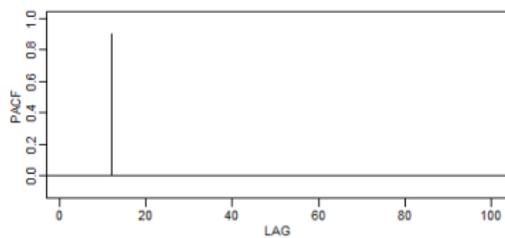
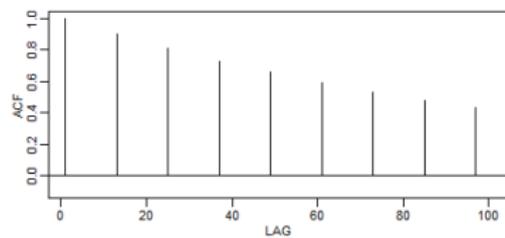
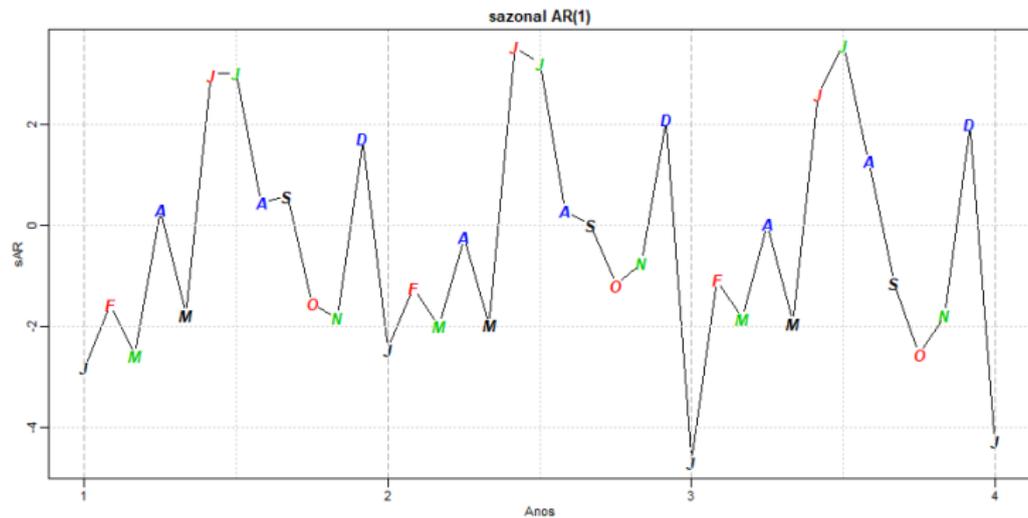
Uma série autorregressiva sazonal de primeira ordem que pode ser executada ao longo de meses pode ser escrita como

$$(1 - \Phi B^{12})X_t = W_t$$

ou

$$X_t = \Phi X_{t-12} + W_t.$$

Este modelo exhibe a série X_t em termos de atrasos passados no múltiplo do período sazonal anual $s = 12$ meses. Fica claro, a partir da forma acima, que a estimação e previsão para tal processo envolve apenas modificações diretas do caso de defasagem unitária já tratado.



Como critério inicial de diagnóstico, podemos usar as propriedades das séries autorregressivas sazonais e de médias móveis sazonais listadas na Tabela abaixo.

O * significa que os valores em atrasos não sazonais $h \neq ks$, para $k = 1, 2, \dots$, são zero.

Tabela III.3. Comportamento do ACF e PACF para modelos SARMA puros

	$AR(P)_s$	$MA(Q)_s$	$ARMA(P, Q)_s$
ACF^*	Cauda em retardos ks $k = 1, 2, \dots$	Corta depois lag Q_s	Cauda em retardos lags ks
$PACF^*$	Corta depois lag P_s	Cauda em retardos ks $k = 1, 2, \dots$	Cauda em retardos lags ks

Em geral, podemos combinar os operadores sazonais e não sazonais em um modelo de média móvel autorregressivo sazonal multiplicativo, denotado por $ARMA(p, q) \times (P, Q)_s$ e escrever

$$\Phi_p(B^s)\phi(B)X_t = \Theta_q(B^s)\theta(B)W_t,$$

como o modelo geral.

Embora as propriedades de diagnóstico na Tabela III.3 não sejam estritamente verdadeiras para o modelo global misto, o comportamento do ACF e do PACF tende a mostrar padrões aproximados da forma indicada. De fato, para modelos mistos, temos a ver uma mistura dos fatos listados na Tabela III.1 e na Tabela III.3. Na adaptação de tais modelos, a focalização nos componentes de média regressiva e média móvel sazonal geralmente leva a resultados mais satisfatórios.

A persistência sazonal ocorre quando o processo é quase periódico na temporada. Por exemplo, com temperaturas médias mensais ao longo de diversos anos, cada janeiro seria aproximadamente o mesmo, cada fevereiro seria aproximadamente o mesmo e assim por diante.

Nesse caso, podemos pensar na temperatura média mensal X_t como sendo modelada como

$$X_t = S_t + W_t,$$

onde S_t é um componente sazonal que varia um pouco de um ano para o outro, de acordo com um passeio aleatório,

$$S_t = S_{t-12} + V_t.$$

W_t e V_t são processos de ruído branco não correlacionados. A tendência dos dados para seguir este tipo de modelo será exibida em uma amostra de ACF que é grande e decai muito lentamente nas defasagens $h = 12k$, para $k = 1, 2, \dots$. Se subtrairmos o efeito de anos sucessivos um do outro, descobriremos que

$$(1 - B^{12})X_t = X_t - X_{t-12} = V_t + W_t - W_{t-12}.$$

Este modelo é um $MA(1)_{12}$ estacionário e o seu ACF terá um pico apenas na defasagem 12. Em geral, a diferenciação sazonal pode ser indicada quando o ACF decai lentamente em múltiplos de algumas estações, mas é insignificante entre os períodos. Então, uma diferença sazonal da ordem D é definida como

$$\nabla_s^D X_t = (1 - B^s)^D X_t,$$

onde $D = 1, 2, \dots$ assume valores inteiros positivos. Normalmente, $D = 1$ é suficiente para obter estacionariedade sazonal.

Definição III.12. Modelo SARIMA.

A modelo multiplicativo sazonal autoregressivo integrado de médias móveis ou modelo SARIMA é dado por

$$\Phi_p(B^s)\phi(B)\nabla_s^D\nabla^dX_t = \delta + \Theta_q(B^s)\theta(B)W_t,$$

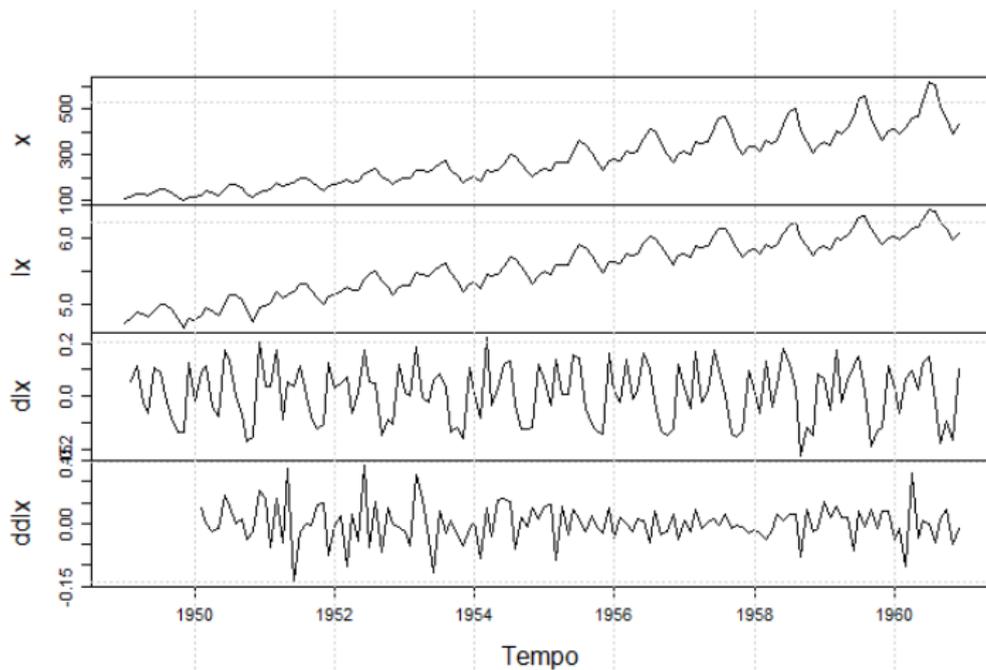
onde W_t é o processo habitual de ruído branco gaussiano. O modelo geral é denotado como $ARIMA(p, d, q) \times (P, D, Q)_s$. As componentes autorregressiva e de médias móveis são representadas pelos polinômios $\phi(B)$ e $\theta(B)$ de ordem p e q , respectivamente, e as componentes autorregressiva e de médias móveis sazonal por $\Phi_p(B^s)$ e $\Theta_q(B^s)$ de ordem P e Q assim como as componentes das diferenças ordinárias e sazonal por $\nabla^d = (1 - B)^d$ e $\nabla_s^D = (1 - B^s)^D$.

Selecionar o modelo *SARIMA* apropriado, para um dado conjunto de dados de todos aqueles representados pela expressão geral, é uma tarefa difícil e geralmente pensamos primeiro em termos de encontrar operadores de diferenças que produzam uma série aproximadamente estacionária e então em termos de encontrar um conjunto autorregressivo de médias móveis simples sazonal multiplicativo para se ajustar à série residual resultante. As operações de diferenciação são aplicadas primeiro e, em seguida, os resíduos são construídos a partir de uma série de comprimentos reduzidos. Em seguida, o ACF e o PACF desses resíduos são avaliados. Os picos que aparecem nessas funções podem ser eliminados com o ajuste de um componente autorregressivo ou de médias móveis de acordo com as propriedades gerais na Tabela III.1 e na Tabela III.3. Ao considerar se o modelo é satisfatório, as técnicas de diagnóstico discutidas se aplicam.

Exemplo III.49. Passageiros aéreos.

Consideramos o conjunto de dados R **AirPassengers**, que são os totais mensais de passageiros de linhas aéreas internacionais entre 1949 a 1960, retirados do livro de Box & Jenkins (1970). Vários gráficos dos dados e dados transformados são mostrados nas figuras a seguir e foram obtidos da forma apresentada.

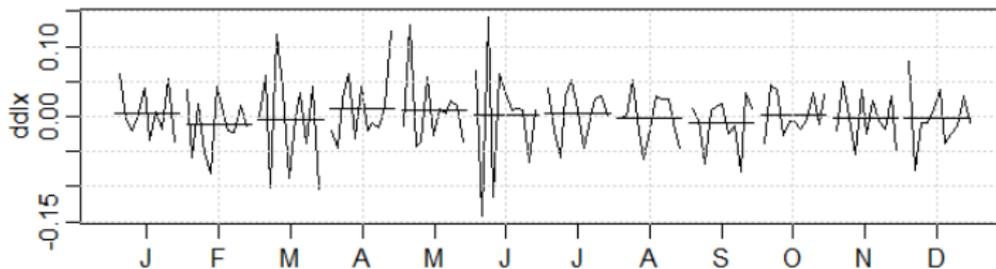
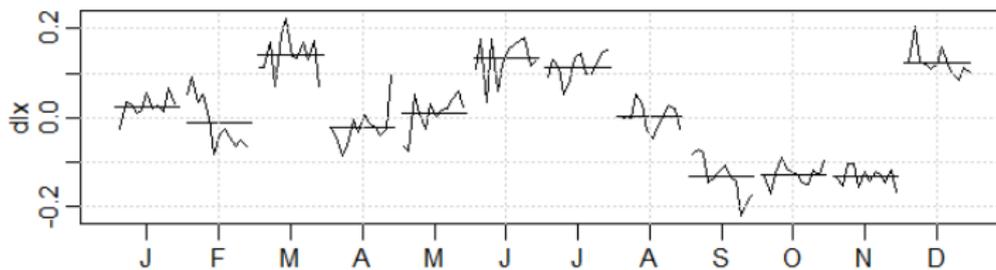
Conjunto de dados R **AirPassengers**, que são os totais mensais dos passageiros aéreos internacionais x e os dados transformados: $lx = \log(x_t)$, $dlx = \nabla \log(x_t)$ e $ddlx = \nabla_{12} \nabla \log(x_t)$.



Exemplo III.49. Passageiros aéreos (continuação).

```
> x = AirPassengers
> lx = log(x); dlx = diff(lx); ddlx = diff(dlx, 12)
> par(mfrow=c(1,1), mar=c(1,1,1,1), mgp=c(1.6,0.8,0))
> plot.ts(cbind(x,lx,dlx,ddl), main="", xlab="Tempo")
> grid()

> # abaixo o interesse é mostrar a componente sazonal
> par(mfrow=c(2,1), mar=c(3,3,2,1), mgp=c(1.6,.6,0))
> monthplot(dlx)
> grid()
> monthplot(ddlx)
> grid()
```

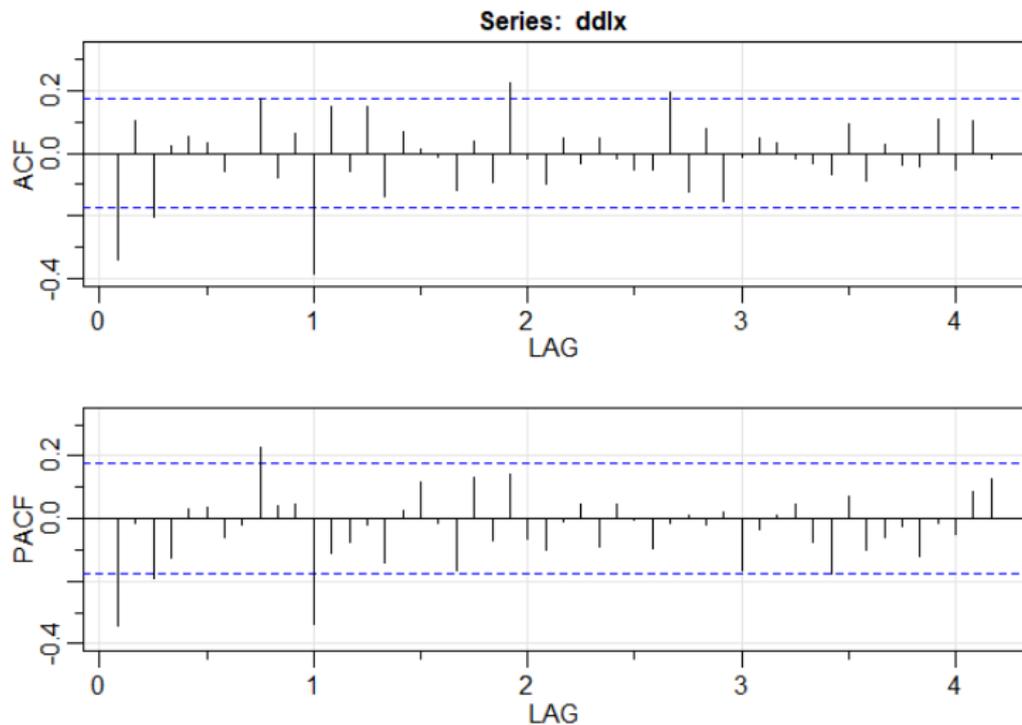


Exemplo III.49. Passageiros aéreos (continuação).

Observe que x é a série original, que mostra tendência mais variação crescente. O logaritmo dos dados estão registrados em lx e a transformação estabiliza a variação. O logaritmo dos dados são então diferenciados para remover a tendência e são armazenados em dlx . É claro que ainda há persistência nas estações, ou seja, $dlx_t \approx dlx_{t-12}$, de modo que uma diferença de décima segunda ordem seja aplicada e armazenada em $ddlx$. Os dados transformados parecem estar estacionários e agora estamos prontos para estimar o modelo.

As funções ACF e PACF amostrais de $ddlx$, $\nabla_{12}\nabla \log(x_t)$ são mostradas na figura a seguir. O código R é:

```
> acf2(ddlx, 50)
```



Exemplo III.49. Passageiros aéreos (continuação).

Componente sazonal: parece que nas estações do ano o ACF está cortando um atraso ou lag de $1s$, $s = 12$, enquanto o PACF está diminuindo nos intervalos $1s, 2s, 3s, 4s, \dots$. Estes resultados implicam um modelo $SMA(1)$, $P = 0$, $Q = 1$, na estação $s = 12$.

Componente não sazonal: inspecionando a amostra ACF e PACF nos defasamentos inferiores, parece que ambos estão diminuindo. Isso sugere um modelo $ARMA(1, 1)$ dentro das estações $p = q = 1$.

Assim, primeiro tentamos um modelo $ARIMA(1, 1, 1) \times (0, 1, 1)_{12}$ no logaritmo dos dados.

Exemplo III.49. Passageiros aéreos (continuação).

```
> modelo = sarima(lx, 1,1,1, 0,1,1,12, details = FALSE)
> modelo$tttable
      Estimate      SE t.value p.value
ar1    0.1960 0.2475  0.7921  0.4298
ma1   -0.5784 0.2132 -2.7127  0.0076
sma1  -0.5643 0.0747 -7.5544  0.0000
```

No entanto, o parâmetro AR não é significativo, então devemos tentar eliminar um parâmetro da parte dentro das estações.

Neste caso, tentamos ambos os modelos

$ARIMA(0, 1, 1) \times (0, 1, 1)_{12}$ e $ARIMA(1, 1, 0) \times (0, 1, 1)_{12}$.

Exemplo III.49. Passageiros aéreos (continuação).

```

> modelo1 = sarima(lx, 0,1,1, 0,1,1,12, details = FALSE)
> modelo2 = sarima(lx, 1,1,0, 0,1,1,12, details = FALSE)
> modelos = matrix(c(modelo[[4]],modelo[[5]],modelo[[6]],
                    modelo1[[4]],modelo1[[5]],modelo1[[6]],
                    modelo2[[4]],modelo2[[5]],modelo2[[6]]),
                    ncol=3, byrow=TRUE)
> colnames(modelos) = c("AIC","AICc","BIC")
> rownames(modelos) = c("modelo","modelo1","modelo2")
> modelos = as.table(modelos)
> modelos

```

	AIC	AICc	BIC
modelo	-3.393658	-3.392434	-3.312667
modelo1	-3.404219	-3.403611	-3.343475
modelo2	-3.390772	-3.390164	-3.330028

```

> sarima(lx, 0,1,1, 0,1,1,12)

```

Exemplo III.49. Passageiros aéreos (continuação).

Todos os critérios de informação preferem o modelo

$$ARIMA(0, 1, 1) \times (0, 1, 1)_{12},$$

que é o modelo ajustado no objeto modelo1.

Os diagnósticos dos resíduos são mostrados na figura abaixo e, com exceção de um ou dois outliers, o modelo parece se encaixar bem.

