

# Trabalho 01 - CE017 - Análise de Séries Temporais

Professor Fernando Lucambio Perez

*Willian Meira Schlichta/ GRR:20159077*

*16 de Setembro de 2019*

## Conteúdo

<b>1</b>	<b>PARTE I - Característica da Série Temporal</b>	<b>2</b>
1.1	Questão 1 . . . . .	2
1.2	Questão 2 . . . . .	3
1.3	Questão 4 . . . . .	8
1.4	Questão 5 . . . . .	9
1.5	Questão 17 . . . . .	10
1.6	Questão 18 . . . . .	14
<b>2</b>	<b>PARTE II - Análise Exploratórios de Dados</b>	<b>16</b>
2.1	Questão 6 . . . . .	16
2.2	Questão 9 . . . . .	16
2.3	Questão 10 . . . . .	24

# 1 PARTE I - Característica da Série Temporal

## 1.1 Questão 1

Considere um modelo sinal-mais-ruído da forma geral  $X_t = S_t + W_t$ , onde  $W_t$  é o ruído branco gaussiano com  $\sigma_W^2 = 1$ . Simule e mostre graficamente o resultado de  $n=200$  observações de cada um dos dois modelos a seguir.

(a)  $X_t = S_t + W_t$ , para  $t = 1, \dots, 200$ , onde

$$S_t = \begin{cases} 0, & t=1,2,\dots,100 \\ 10\exp(-t-100/20)\cos(2\pi t/4), & t=101,\dots,200 \end{cases}$$

(b)  $X_t = S_t + W_t$ , para  $t = 1, \dots, 200$ , onde

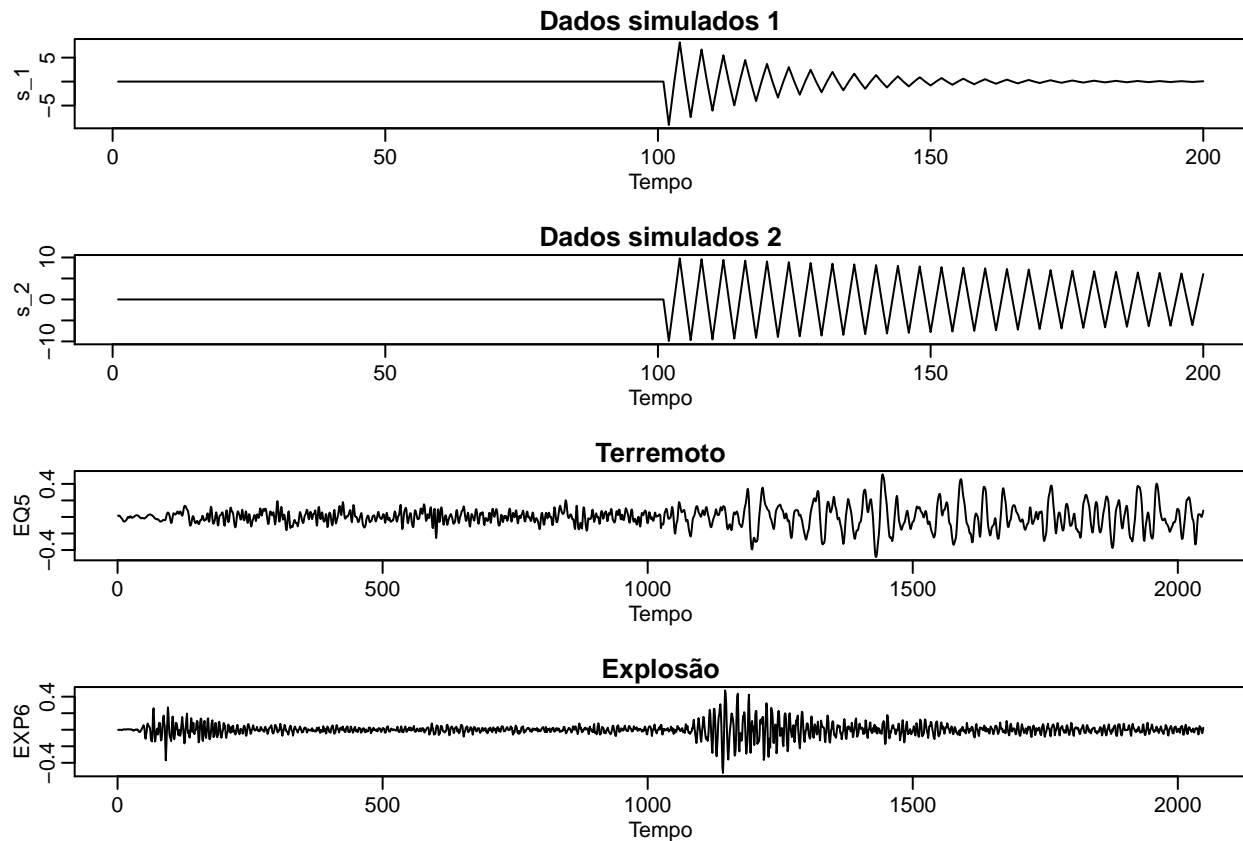
$$S_t = \begin{cases} 0, & t=1,2,\dots,100 \\ 10\exp(-t-100/200)\cos(2\pi t/4), & t=101,\dots,200 \end{cases}$$

### Resoluções

```
# Equação 1 - Dados simulados 1
s_1 = c(rep(0,100), 10*exp(-((101:200)-100)/20)*cos(2*pi*101:200/4))
set.seed(11)
x_1 = s_1 + rnorm(200)

# Equação 2 - Dados simulados 2
s_2 = c(rep(0,100), 10*exp(-((101:200)-100)/200)*cos(2*pi*101:200/4))
set.seed(22)
x_2 = s_2 + rnorm(200)

# Gráfico comparação entre diferente equações dadas pelo exercício:
par(mfrow=c(4,1), mar=c(3,2,1,0)+.5, mgp=c(1.6,.6,0))
plot(s_1, xlab="Tempo", main="Dados simulados 1", type = "l")
plot(s_2, xlab="Tempo", main="Dados simulados 2", type = "l")
plot(EQ5, xlab="Tempo", main="Terremoto")
plot(EXP6, xlab="Tempo", main="Explosão")
```



Percebe-se que o gráfico de dados simulados 1 tem comportamento próximo ao gráfico de explosão. Após o tempo 100 para dados simulados 1 e tempo 1000 para explosão ambos possuem uma variação mais intensa e depois voltam a variar pouco em torno da média. Já os dados simulados 2 tem comportamento mais próximo aos dados de terremoto. Após o tempo 100 para simulados 2 e aproximadamente tempo 1000 para terremoto, ambos aumentam a variação ao redor da média até o fim do período analisado.

## 1.2 Questão 2

Compare a aparência geral das séries em (a) e em (b) com a série sísmica e a série de explosões mostradas no Exemplo I.7. Além disso, mostrar e comparar os moduladores de sinal (a)  $\exp(-t/20)$  e (b)  $\exp(-t/200)$ , para  $t = 1, 2, \dots, 100$ .

- (a) Gerar  $n = 100$  observações da autorregressão  $X_t = -0.9X_{t-2} + W_t$  com  $\sigma_W^2 = 1$ , usando o método descrito no Exemplo I.10. Em seguida, aplique o filtro médio móvel  $V_t = (X_t + X_{t-1} + X_{t-2} + X_{t-3})/4$  para  $X_t$ , aos dados que você gerou. Agora mostre  $X_t$  como uma linha e sobreponha  $V_t$  como uma linha tracejada. Comente sobre o comportamento de  $X_t$  e como aplicar o filtro de média móvel altera esse comportamento.

Sugestão: `> v = filter(x, rep(1/4, 4), sides = 1)`.

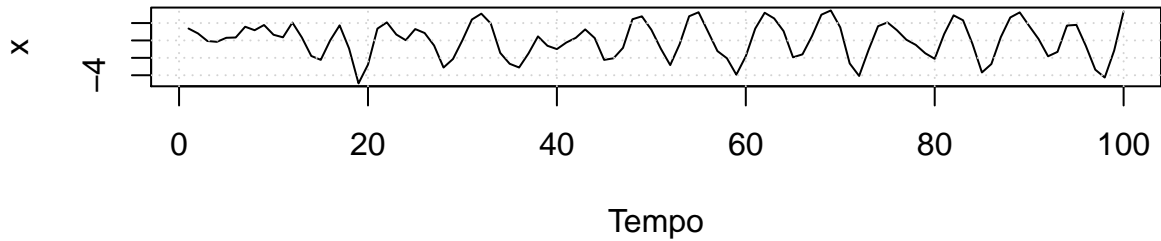
```
set.seed(42)
w = rnorm(100,0,1)
x = filter(w, filter=c(1,-.9), method="recursive") #

v = filter(x, sides=1, rep(1/4,4)) # médias móveis

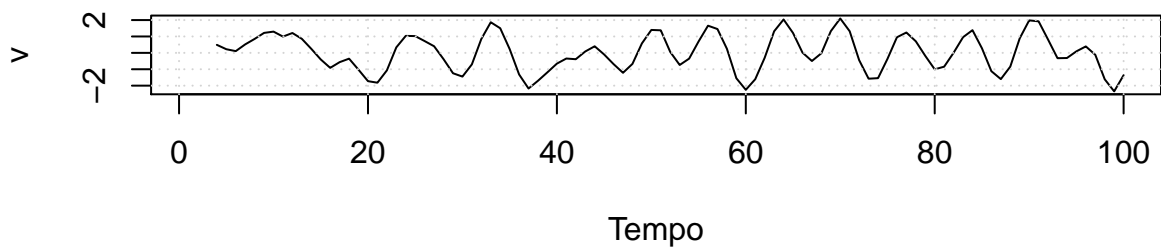
par(mfrow=c(2,1))
plot.ts(x, xlab="Tempo", main="Autoregressão")
```

```
grid()
plot.ts(v, xlab="Tempo", main="Média Move1")
grid()
```

## Autoregressão

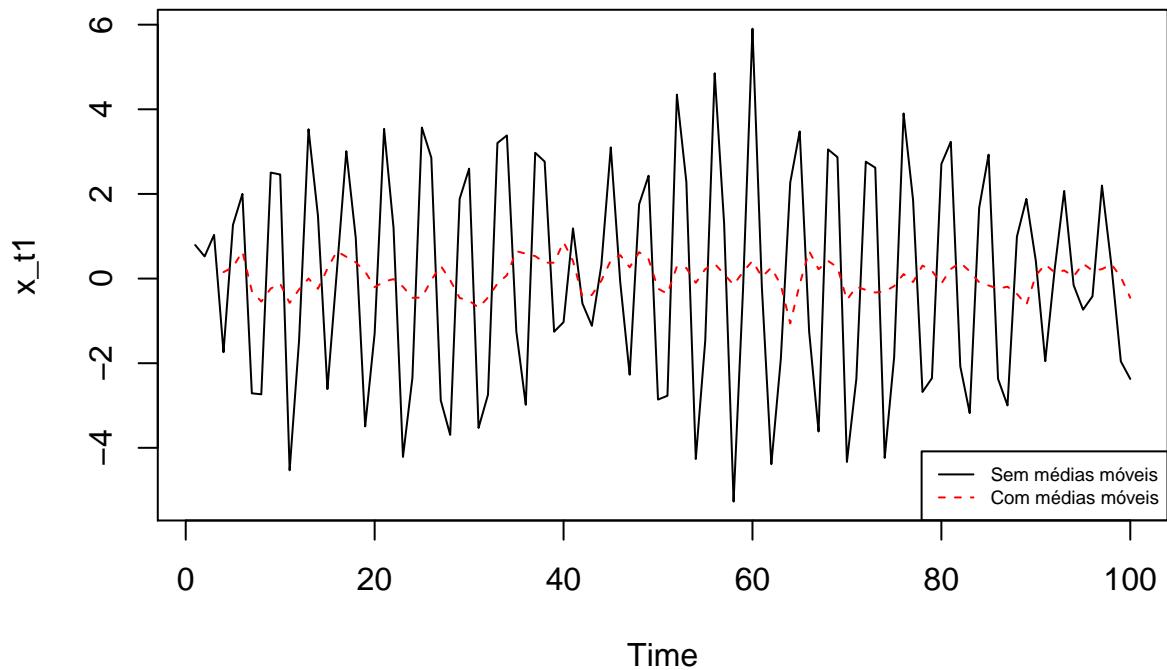


## Média Move1



```
set.seed(21)
w_1 = rnorm(100,0,1)
x_t1 = filter(w_1, filter = c(0,-0.9), method = "recursive")
v_t1 = filter(x_t1, rep(1/4, times = 4), sides=1)

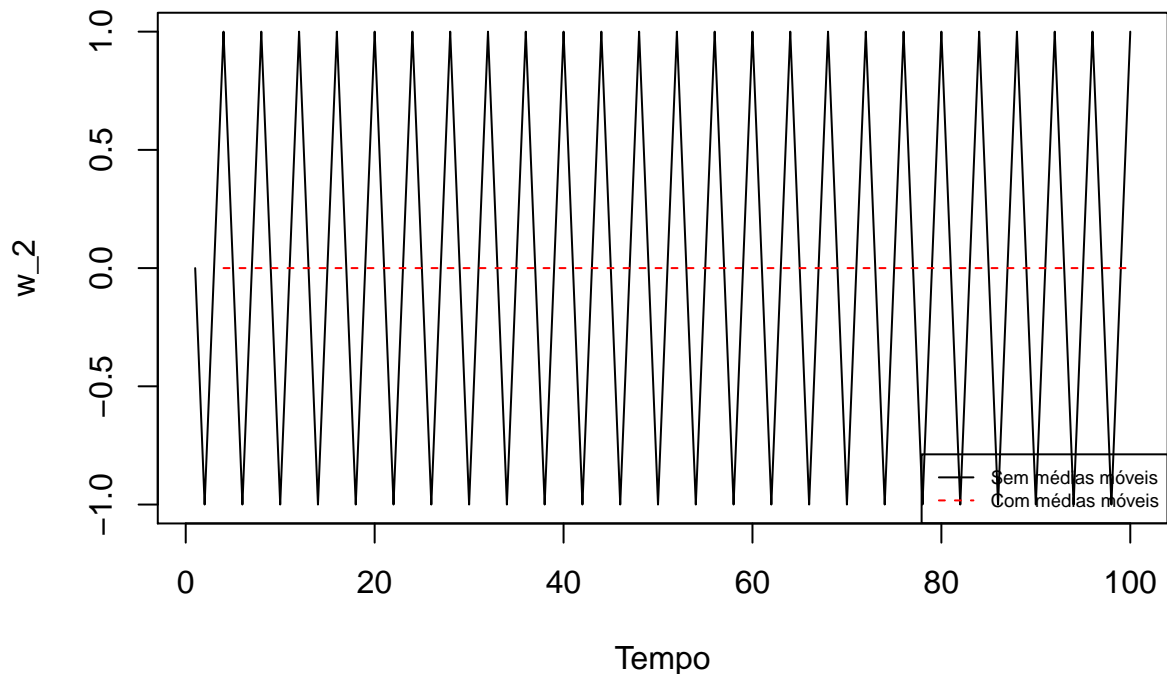
par(mfrow=c(1,1))
plot(x_t1, type = "l")
lines(v_t1, lty=2, col=2)
legend("bottomright", legend=c("Sem médias móveis", "Com médias móveis"),
col = c(1,2), lty = c(1,2), cex=0.6)
```



As médias móveis (linha vermelha tracejada) diminuiram consideravelmente a oscilação dos dados, ou seja, como esperado, após aplicada as médias móveis nos dados gerados seu comportamento fica mais próximo da média.

(b) Repetir (a) mas com  $X_t = \cos(2\pi t/4)$ .

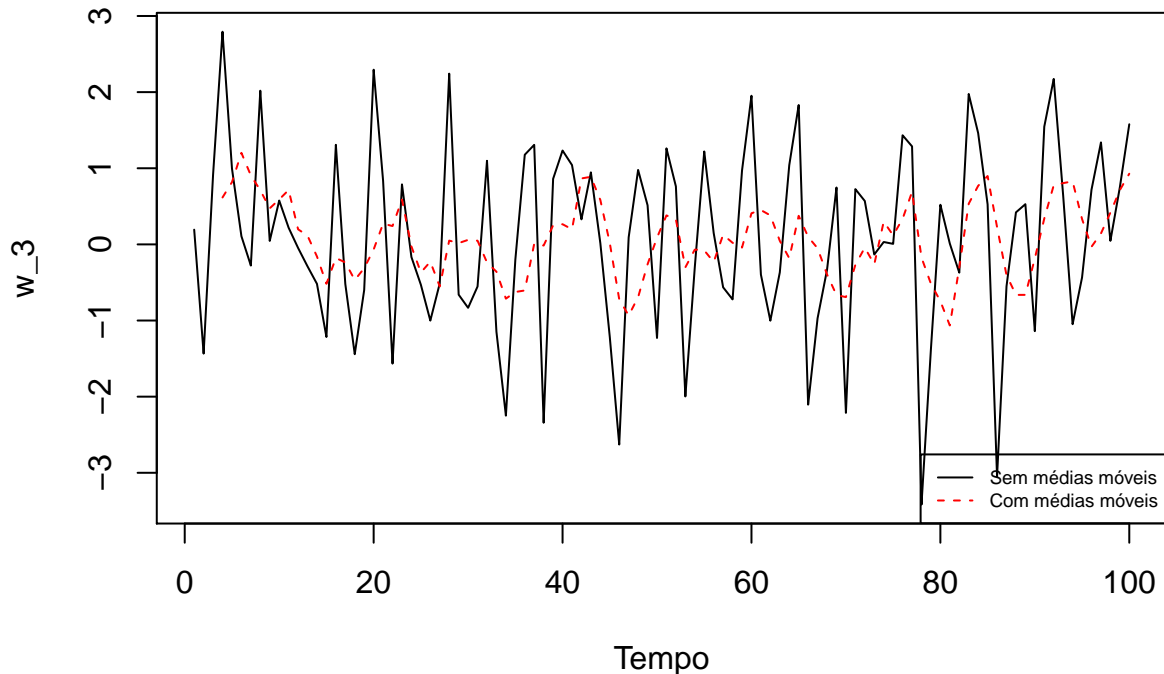
```
w_2 = cos((2*pi*1:100)/4)
v_t2 = filter(w_2, rep(1/4, times = 4), sides=1)
plot(w_2, type = 'l', xlab = "Tempo")
lines(v_t2, lty=2, col=2)
legend("bottomright", legend=c("Sem médias móveis", "Com médias móveis"),
col = c(1,2), lty = c(1,2), cex=0.6)
```



Devido a ausência de ruído branco na geração dos dados, a distribuição apresenta aspecto sistemático e as médias móveis comportamento praticamente constante ao redor de zero.

(c) Repetir (b) mas com adição de ruído  $N(0,1)$ ,  $X_t = \cos(2\pi t/4) + W_t$ .

```
set.seed(23)
w_3 = cos((2*pi*1:100)/4) + rnorm(100,0,1)
v_t3 = filter(w_3, rep(1/4, times = 4), sides=1)
plot(w_3, type = 'l', xlab = "Tempo")
lines(v_t3, lty=2, col=2)
legend("bottomright", legend=c("Sem médias móveis", "Com médias móveis"),
col = c(1,2), lty = c(1,2), cex=0.6)
```



Devido a presença de ruído branco na geração dos dados, a distribuição deixou de apresentar aspecto sistemático, assim como as médias móveis perderam o comportamento praticamente constante ao redor de zero.

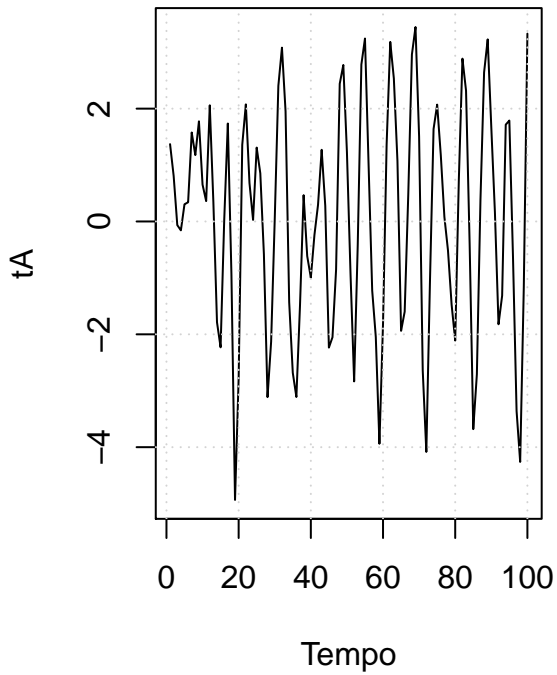
(d) Compare e contraste (a) - (c); isto é, como a média móvel muda cada série.

```
set.seed(42)
t = rnorm(100,0,1)
tA = filter(w, filter=c(1,-.9), method="recursive") #

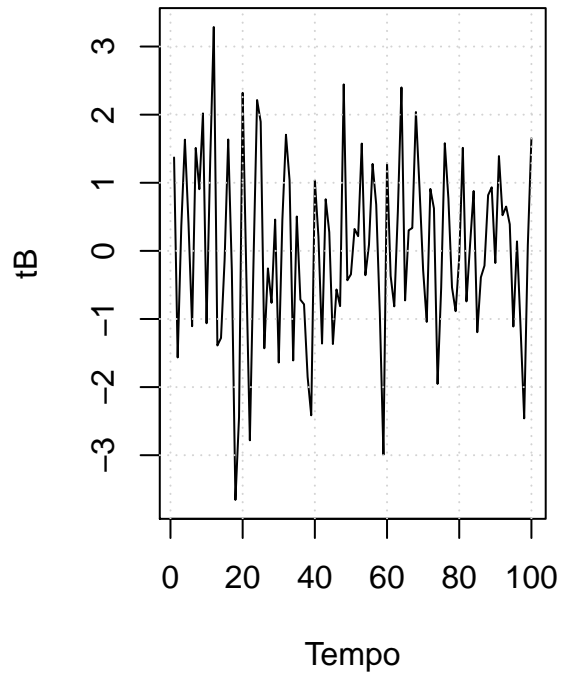
tB = cos(2*pi*1:100/4) + t

par(mfrow=c(1,2))
plot.ts(tA, xlab="Tempo", main="Autoregressão")
grid()
plot.ts(tB, xlab="Tempo", main="Média Movel")
grid()
```

## Autoregressão

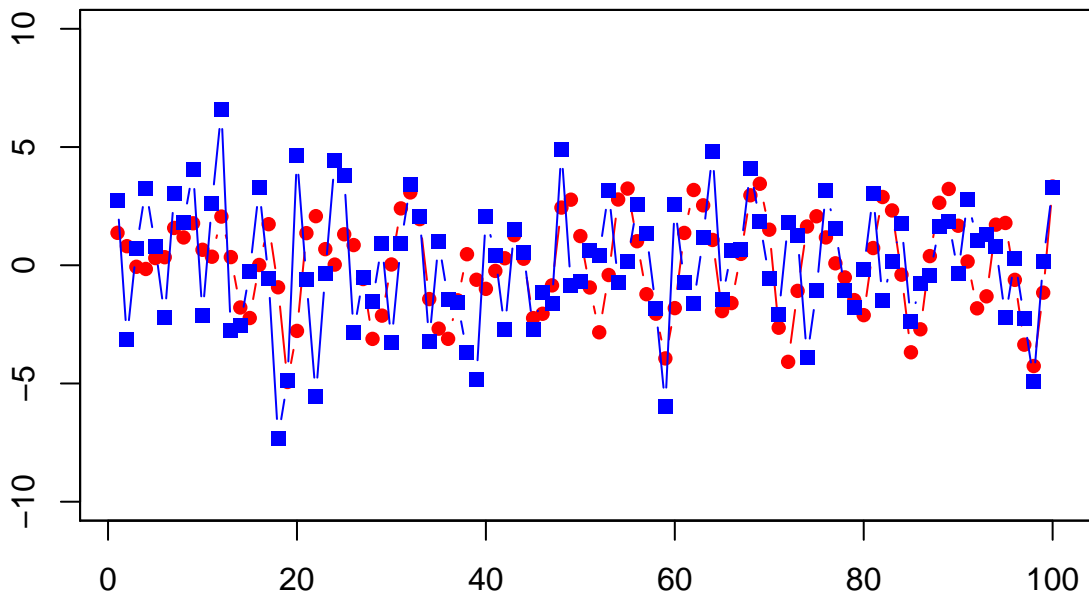


## Média Movel



```
par(mfrow=c(1,1))
plot(tA, pch=16, axes=TRUE, ylim=c(-10,10), xlab="", ylab="", type="b", col="red", main="")
box()

## Plotando o segundo gráfico e colocando a escala do eixo à direita
par(new=TRUE)
plot(tB, pch=15, axes=FALSE, ylim=c(-5,5), xlab="", ylab="", type="b", col="blue")
```



A média móvel atua da mesma maneira em cada uma das séries - reduzindo a oscilação das observações, deixando-as mais próximas da média global da série.

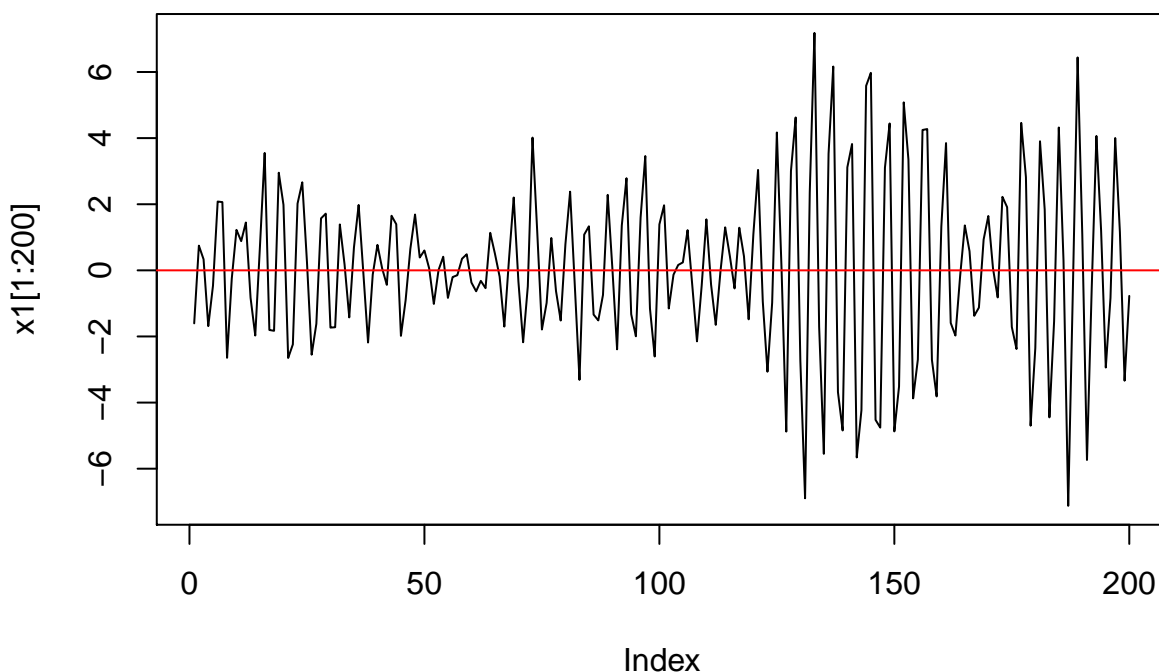
### 1.3 Questão 4

Para as duas séries  $X_t$ , no Exercício 2 (a) e (b):

(a) Calcular e traçar as funções de médias  $\mu_X(t)$ , para  $t = 1, 2, \dots, 200$ .

```
# Exercício 4
set.seed(1234)
w = rnorm(550,0,1) # 50 extras para evitar problemas de inicialização
# 'w' é o ruído
#----- Primeira Série
x1 = filter(w, filter=c(0,-.9), method="recursive")[-(1:50)] # removendo os primeiros 50
#----- Segunda Série
x2 = cos(2*pi*1:550/4)

plot(x1[1:200],type="l")
abline(h=0, col="red")
```



$$((a)) \quad X_t = -0.9X_{t-2} + W_t$$

$$\mu_{x_t} = E(X_t) = E(-0.9X_{t-2}) + E(W_t)$$

$$= -0.9 E(X_{t-2})$$

$$P/ \quad X_1 = W_1 \rightarrow X_1 = 0$$

$$X_2 = W_2 \rightarrow X_2 = 0$$

$$X_3 = -0.9E(X_1) = 0 \quad X_4 = -0.9E(X_2) = 0 \quad \vdots \quad X_{100} = -0.9E(X_{98}) = 0$$

$$((b)) \quad X_t = \cos(2\pi t/4)$$

$$\mu = E(X_t) = E(\cos(\frac{2\pi t}{4})) = \frac{\cos \pi t}{4}$$

(b) Calcular a função de autocovariância  $\gamma(s, t)$ , para  $s, t = 1, \dots, 200$ .

$$((a)) \quad X_t = -0.9X_{t-2} + W_t \quad p/ \quad h=2$$

$$E(X_s \cdot X_3) = E(-0.9X_3 + W_s) \cdot (-0.9X_1 + W_3)$$



$$= E(-0.9^2 X_1 X_3 - 0.9 X_3 W_3 - 0.9 X_1 W_s + W_s W_3)$$

$$E(0.9^2 X_1 X_3) = 0$$

$\Downarrow$

$$+0.9^2 E(X_3 X_1)$$

$\Downarrow$

$$+0.9^2 . E[(-0.9 X_1 + W_3) . X_1]$$

$$+0.9^2 . E[(-0.9 X_1^2 + X_1 W_3)]$$

$$-0.9^3 E(X_1^2) + E(X_1 W_3)$$

$\Downarrow$

$$\sigma_w^2$$

$$-0.9^3 . \sigma_w^2$$

$$-0.9^{h+1} \sigma_w^2$$

$$((b)) \quad X_t = \cos(2\pi t/4)$$

$$\gamma(s, t) = 0 \text{ porque } \acute{e} \text{ uma constante } E(X_t) = \cos \frac{2\pi t}{4}$$

## 1.4 Questão 5

Considere a s rie temporal  $X_t = \beta_0 + \beta_1 t + W_t$ , onde  $\beta_0$  e  $\beta_1$  s o constantes e  $W_t$    um processo de ru do branco com vari ncia  $\sigma_W^2$ .

(a) Determine se  $X_t$    estacion rio.

$$X_t = \beta_0 + \beta_1 t + W_t \therefore W_t \sim N(0, \sigma_w^2)$$

$$E(X_t) = E(\beta_0 + \beta_1 t + W_t) = E(\beta_0) + E(\beta_1 t) + E(W_t) =$$

$$= \beta_0 + \beta_1 t, \text{ n o estacion rio, pois } E(X_t) \text{ depende de } t.$$

(b) Mostre que o processo  $Y_t = X_t - X_{t-1}$    estacion rio.

$$E(Y_t) = E[X_t - X_{t-1}] = E(\beta_0 + \beta_1 t + W_t - \beta_0 - \beta_1(t-1) - W_{t-1}) = E[\beta_1(t-t+1)] + E[W_t] + E[W_{t-1}] = E[\beta_1(1)] = \beta_1$$

$$E[Y_t] \text{ n o depende de } t$$

$$COV[Y_s, Y_t] = (1)(1)COV(X_{t+h}, X_t) + (1)(-1)COV(X_{t+h}, X_{t-1}) + (-1)(1)COV(X_{t+h-1}, X_t) + (-1)(-1)COV(X_{t+h-1}, X_{t-1})$$

$$Y_{t+s} = X_{t+h} - X_{t+h-1}, \quad Y_t = X_t - X_{t-1} \quad \text{Onde; } Y_{t+s} = s; X_{t+h} \rightarrow a_1 = 1; X_{t+h-1} \rightarrow a_2 = -1 \text{ e } X_t \rightarrow b_1 = 1; X_{t-1} \rightarrow b_2 = -1$$

$$\text{Sabemos que: } X_t \sim N(\beta_0 + \beta_1 t, I\sigma_w^2)$$

$$Var(X_t) = I\sigma_w^2, \text{ ent o:}$$

$$Cov(X_t) = \begin{pmatrix} \sigma_w^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_w^2 \end{pmatrix}$$

Logo:

$$COV(Y_{t+h}, Y_t) = \begin{cases} \sigma_w^2, & h=0 \\ 0, & h>0 \end{cases}$$

Como  $E[Y_t]$  não depende do tempo, e a função de autocorrelação vale 0 ou  $\sigma_w^2$ , então a série  $Y_t$  é dita estacionária.

- (c) Mostre que a média do processo de média móvel  $V_t = \frac{1}{2q+1} \sum_{j=-q}^q X_{t-j}$ , é  $\beta_0 + \beta_1 t$  e apresente uma expressão simplificada para a função de autocovariância.

$$V_t = \frac{1}{2q+1} \sum_{j=-q}^q X_{t-j}$$

$$E[V_t] = E\left[\frac{1}{2q+1} \sum_{j=-q}^q X_{t-j}\right] = \frac{1}{2q+1} E[X_{t+q} + X_{t+q+1} + \dots + X_t + \dots + X_{t-q-1} + X_{t-q}]$$

$\therefore$  como  $E(X_t) = \beta_0 + \beta_1 t$ , então:

$$E[V_t] = \frac{1}{2q+1} ((2q+1)\beta_0 + \beta_1((t+q) + (t+q+1) + \dots + t + \dots + (t-q-1) + (t-q))) \text{ cancela os termos sem } t$$

$$= \frac{1}{2q+1} ((2q+1)\beta_0 + \beta_1(2q+1)t)$$

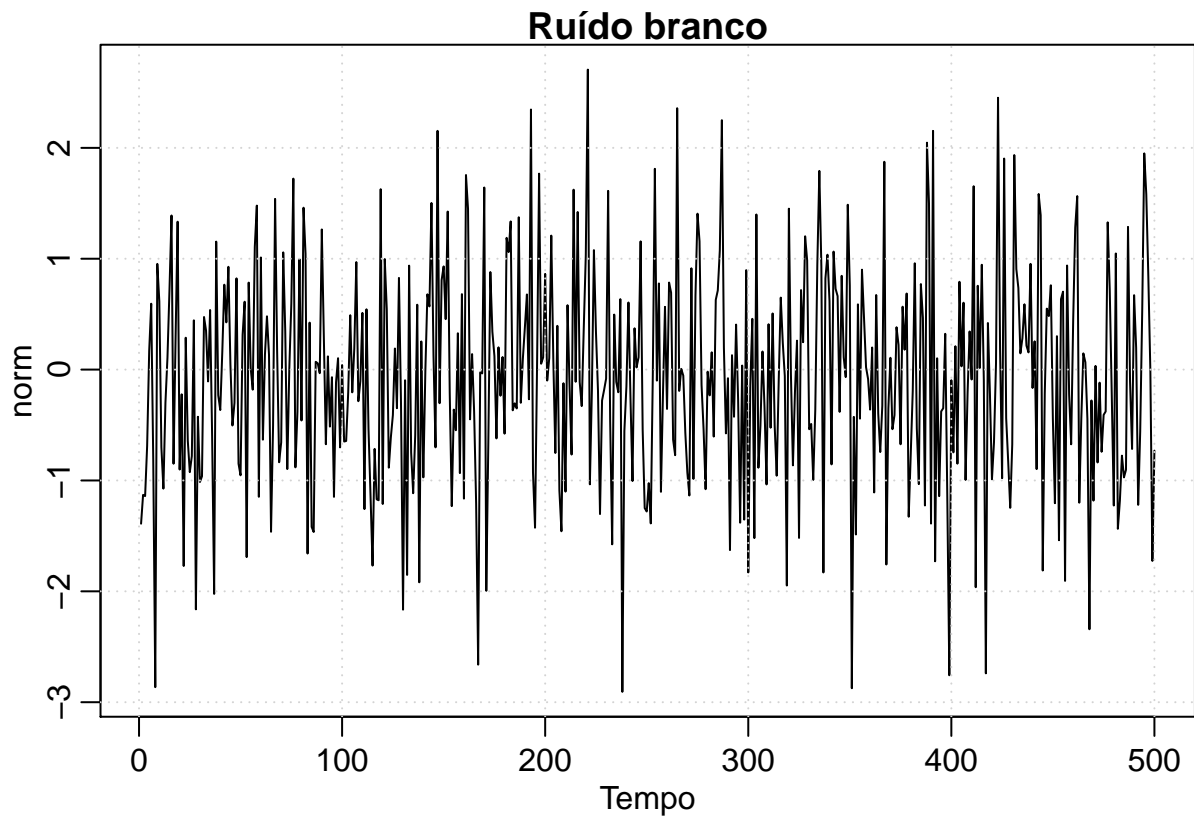
$$= \frac{2q+1}{2q+1} (\beta_0 + \beta_1 t)$$

$$= \beta_0 + \beta_1 t$$

## 1.5 Questão 17

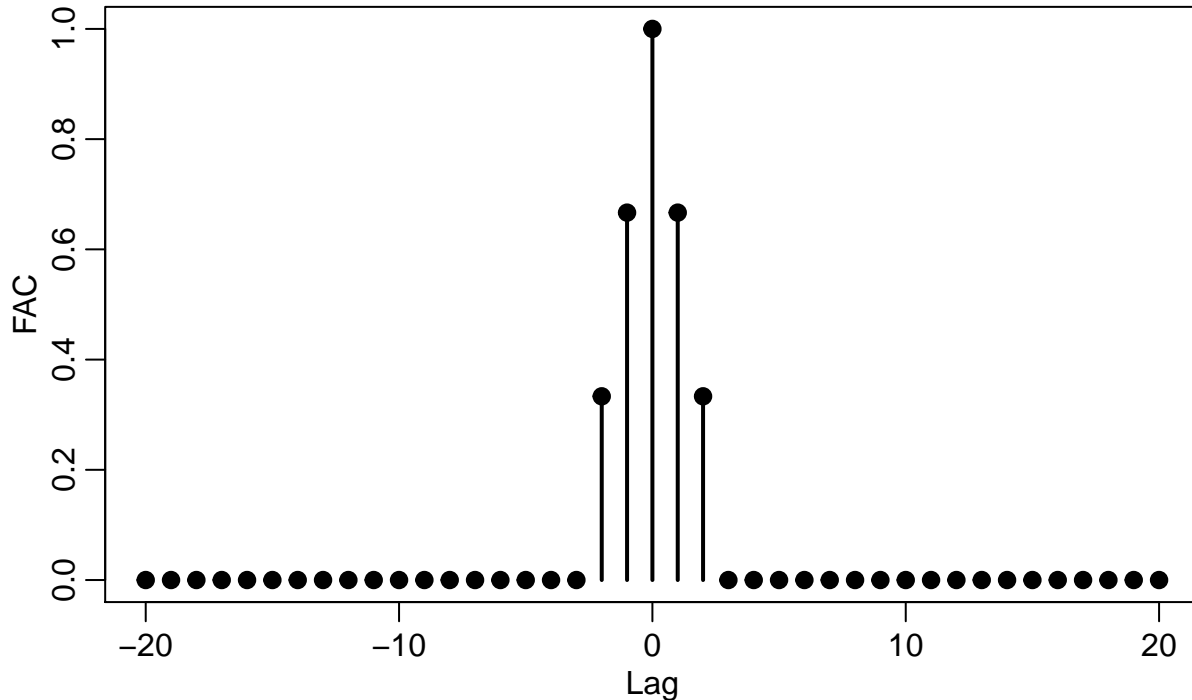
- (a) Simule uma série de  $n = 500$  observações de um processo ruído branco Gaussiano como no Exemplo I.8 e calcule a função ACF mostral  $\hat{\rho}(h)$ , para até o lag 20. Compare a ACF amostral que você obteve com o ACF real  $\rho(h)$ .

```
norm = rnorm(500,0,1)
par(mfrow=c(1,1),mar=c(4,3,1,1),mgp=c(1.6,.6,0))
plot.ts(norm, xlab="Tempo", main="Ruído branco")
grid()
```



```
LAG = seq(-20,20, by=1)
FAC = c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1/3,2/3,1,2/3,1/3,
        0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
par(mfrow=c(1,1), mar=c(4,3,3,1), cex.main=1.5)
plot(LAG,FAC, main="Função de autocorrelação para até lag 20",
     type="h", lwd=2, xlab="Lag")
points(LAG,FAC, lwd=2, pch=19)
```

## Função de autocorrelação para até lag 20



```
acf(norm, lag.max=20, plot=FALSE)
```

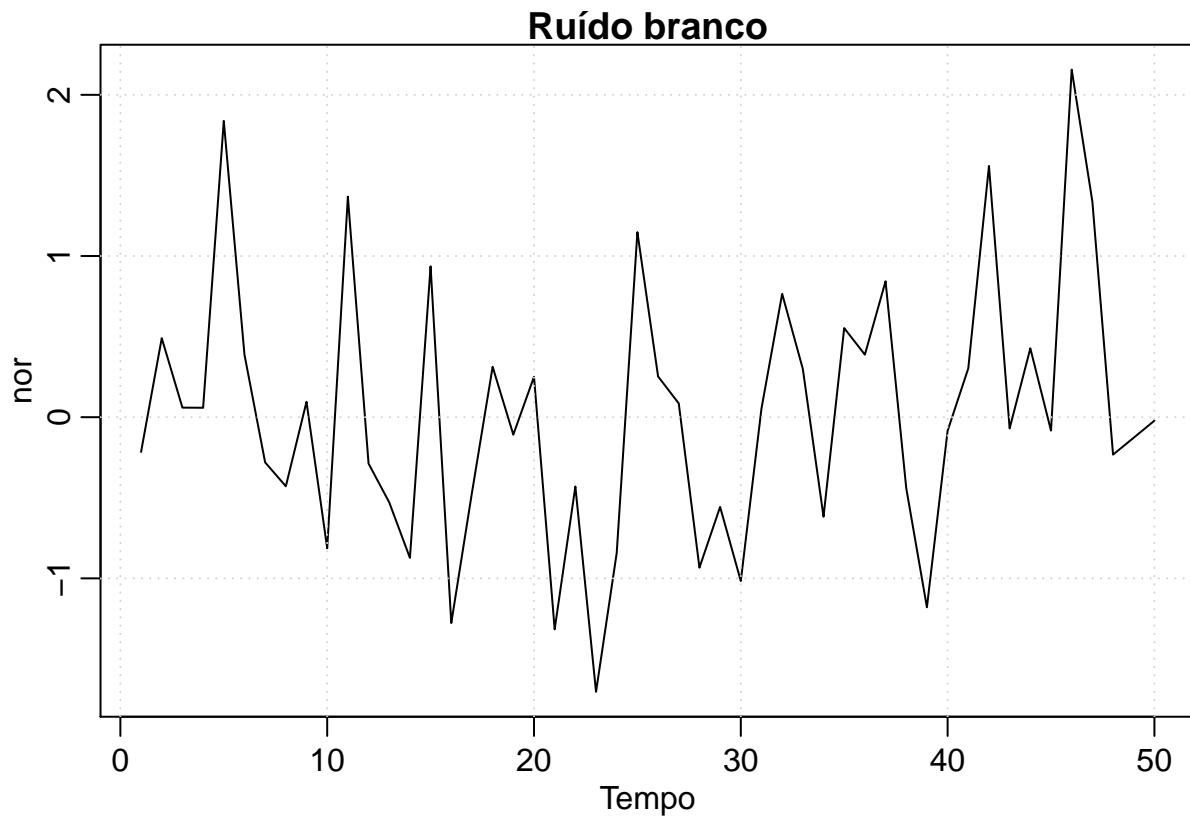
```
##
## Autocorrelations of series 'norm', by lag
##
##      0      1      2      3      4      5      6      7      8      9
## 1.000 -0.042 -0.038  0.038 -0.026  0.017 -0.020 -0.023 -0.014  0.024
##     10     11     12     13     14     15     16     17     18     19
## 0.022 -0.017 -0.011  0.007 -0.015 -0.008 -0.023 -0.014  0.080  0.006
##     20
## 0.013
```

```
ARMAacf(norm, lag.max = 20)
```

```
##           0           1           2           3           4           5
## 1.000000000 -0.78458989 -0.19853726  0.21992183  0.12776468  0.39667536
##           6           7           8           9          10          11
## -1.30004547  0.48203805  1.02962092 -0.55099361 -0.08109556  0.66955944
##          12          13          14          15          16          17
## -0.03345251 -0.81367623 -0.08820092  0.03282553 -0.45336723  0.22937227
##          18          19          20
## -0.12735071  0.54767223 -0.29436442
```

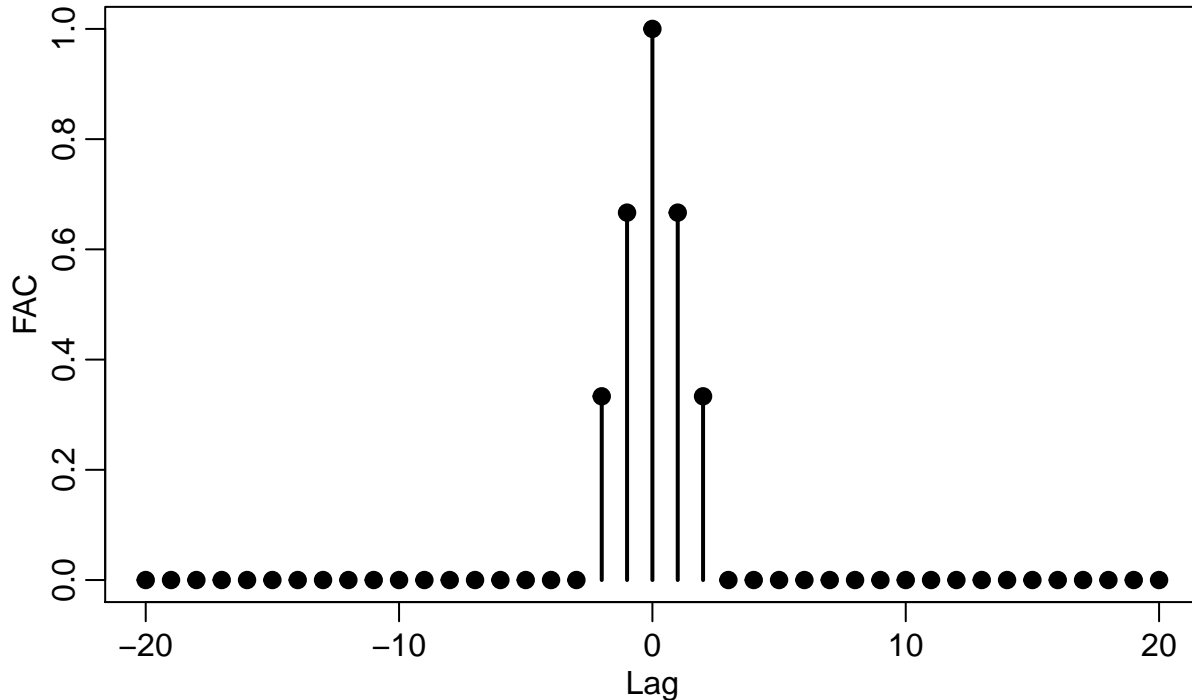
(b) Repita a parte (a) usando apenas  $n = 50$ . Como mudar  $n$  afeta os resultados?

```
nor = rnorm(50,0,1)
par(mfrow=c(1,1),mar=c(4,3,1,1),mgp=c(1.6,.6,0))
plot.ts(nor, xlab="Tempo", main="Ruído branco")
grid()
```



```
LAG = seq(-20,20, by=1)
FAC = c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1/3,2/3,1,2/3,1/3,
        0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
par(mfrow=c(1,1), mar=c(4,3,3,1), cex.main=1.5)
plot(LAG,FAC, main="Função de autocorrelação para até lag 20",
     type="h", lwd=2, xlab="Lag")
points(LAG,FAC, lwd=2, pch=19)
```

## Função de autocorrelação para até lag 20



```
acf(nor, lag.max=20, plot=FALSE)
```

```
##
## Autocorrelations of series 'nor', by lag
##
##      0      1      2      3      4      5      6      7      8      9
## 1.000 0.123 -0.027 -0.161 0.074 0.110 0.152 0.100 -0.159 0.083
##     10     11     12     13     14     15     16     17     18     19
## 0.166 0.032 -0.157 -0.045 0.007 0.067 -0.059 -0.041 -0.189 -0.245
##     20
## 0.085
```

```
ARMAacf(nor, lag.max = 20)
```

```
##      0      1      2      3      4      5      6      7
## 1.000000 3.000107 2.483385 1.597308 2.307047 2.465405 2.007186 2.405151
##      8      9     10     11     12     13     14     15
## 1.891194 2.712067 3.552147 1.800846 2.291362 3.690432 2.448213 1.837600
##     16     17     18     19     20
## 2.847190 2.726548 2.877587 1.577110 1.535878
```

O gráfico do processo ruído branco Gaussiano fica mais aberto por conter menos dados, a ACF não se altera pois por definição ela continua simétrica.

### 1.6 Questão 18

- (a) Simule uma série de  $n = 500$  observações de um processo de médias móveis como no Exemplo I.9 e calcule a função ACF mostrada  $\hat{\rho}(h)$ , para até o lag 20. Compare a ACF amostral que você obteve com o ACF real  $\rho(h)$ .

```
w = rnorm(500,0,1) # 500 N(0,1) variáveis
v = filter(w, sides=2, rep(1/3,3))[-c(1,500)] # médias móveis
acf(v, lag.max=20, plot=FALSE)
```

```
##
## Autocorrelations of series 'v', by lag
##
##      0      1      2      3      4      5      6      7      8      9
## 1.000 0.672 0.330 -0.009 -0.055 -0.112 -0.161 -0.181 -0.153 -0.094
##     10     11     12     13     14     15     16     17     18     19
## -0.017 0.033 0.048 0.057 0.029 0.021 0.009 0.010 -0.011 -0.033
##      20
## -0.038
```

```
ARMAacf(v, ma = rep(1/3,3), lag.max = 20)
```

```
##           0           1           2           3           4           5
## 1.00000000 -0.69921887 -0.03775574  0.74249374 -1.00401661  0.63618997
##           6           7           8           9          10          11
## 0.10644802 -0.79895388  0.98446699 -0.59362512 -0.17618228  0.82021787
##          12          13          14          15          16          17
## -0.98797383  0.53957881  0.22942223 -0.87156359  0.97854970 -0.48492361
##          18          19          20
## -0.29635128  0.90721387 -0.94612146
```

(b) Repita a parte (a) usando apenas  $n = 50$ . Como mudar  $n$  afeta os resultados?

```
w = rnorm(50,0,1) # 500 N(0,1) variáveis
v = filter(w, sides=2, rep(1/3,3))[-c(1,50)] # médias móveis
acf(v, lag.max=20, plot=FALSE)
```

```
##
## Autocorrelations of series 'v', by lag
##
##      0      1      2      3      4      5      6      7      8      9
## 1.000 0.705 0.388 0.062 0.070 0.066 0.084 0.024 -0.081 -0.188
##     10     11     12     13     14     15     16     17     18     19
## -0.336 -0.393 -0.475 -0.369 -0.346 -0.209 -0.212 -0.093 -0.038  0.085
##      20
## 0.135
```

```
ARMAacf(v, ma = rep(1/3,3), lag.max = 20)
```

```
##           0           1           2           3           4           5
## 1.00000000 -0.36825540 -0.54578681  0.76981216 -0.04473121 -0.62985033
##           6           7           8           9          10          11
## 0.57448528 -0.01242429 -0.59611871  0.30157620  0.04487710 -0.21713585
##          12          13          14          15          16          17
## -0.01716814  0.14105334  0.20421749 -0.22497907  0.27023960  0.06660501
##          18          19          20
## 0.02435060  0.04566907 -0.12782767
```

O gráfico do processo de médias móveis fica mais aberto por conter menos dados, a ACF não se altera pois por definição ela continua simétrica.

## 2 PARTE II - Análise Exploratórios de Dados

### 2.1 Questão 6

Considere um processo que consiste em uma tendência linear com um termo de ruído aditivo que consiste em variáveis aleatórias independentes  $W_t$  com médias zero e variância  $\alpha_W^2$ , ou seja,

$$X_t = \beta_0 + \beta_1 t + W_t, W_t \sim N(0, \alpha_W^2),$$

onde  $\beta_0$  e  $\beta_1$  são constantes.

a) Prove que  $X_t$  é não-estacionário.

$$E(X_t) = E(\beta_0 + \beta_1 t + W_t) = E(\beta_0) + E(\beta_1 t) + E(W_t)$$

Como  $E(W_t) = 0$  então:

$$E(X_t) = \beta_0 + \beta_1 t$$

Como  $E(X_t)$  depende do tempo, a série  $X_t$  não é estacionária.

b) Prove que a série da primeira diferença  $\Delta X_t = X_t - X_{t-1}$  é estacionária, encontrando sua função de média e autocovariância.

$$E(\Delta X_t) = E(\beta_0 + \beta_1 t + W_t - \beta_0 - \beta_1(t-1) - W_{t-1}) = E(\beta_1 t + W_t - \beta_1(t-1) - W_{t-1})$$

Como  $E(W_t) = 0$  e  $E(W_{t-1}) = 0$ , então:

$$E(\Delta X_t) = E(\beta_1(t - t + 1)) = \beta_1$$

Sabemos que:  $X_t \sim N(\beta_0 + \beta_1 t, I\sigma_w^2)$

$Var(X_t) = I\sigma_w^2$ , então:

$$Cov(X_t) = \begin{pmatrix} \sigma_w^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_w^2 \end{pmatrix}$$

$\Delta X_{t-1} = X_{t-1} - X_{t-2}$  em que  $a_1 = 1$  e  $a_2 = -1$

$\Delta X_t = X_t - X_{t-1}$  em que  $b_1 = 1$  e  $b_2 = -1$

$$Cov(\Delta X_{t-1}, \Delta X_t) = a_1 b_1 Cov(X_{t-1}, X_t) + a_1 b_2 Cov(X_{t-1}, X_{t-1}) + a_2 b_1 Cov(X_{t-2}, X_t) + a_2 b_2 Cov(X_{t-2}, X_{t-1})$$

$$Cov(\Delta X_{t-1}, \Delta X_t) = a_1 b_2 Cov(X_{t-1}, X_{t-1}) = -1\sigma_w^2 = -\sigma_w^2$$

Como  $E(\Delta X_t)$  e  $Cov(\Delta X_t, \Delta X_{t-1})$  são constantes no tempo, então a série  $\Delta X_t$  é estacionária.

### 2.2 Questão 9

Considere as duas séries temporais de petróleo oil e gás gas. A série de petróleo é em dólares por barril, enquanto a série de gás é em centavos por galão.

- (a) Mostre os dados no mesmo gráfico. Quais das séries simuladas exibidas na Seção I.2 essas séries mais se parecem? Você acredita que as séries são estacionárias, explique sua resposta?

```
# carregando os dados
library(astsa)
head(oil)
```

```
## [1] 26.20 26.07 26.34 24.95 26.27 29.37
```

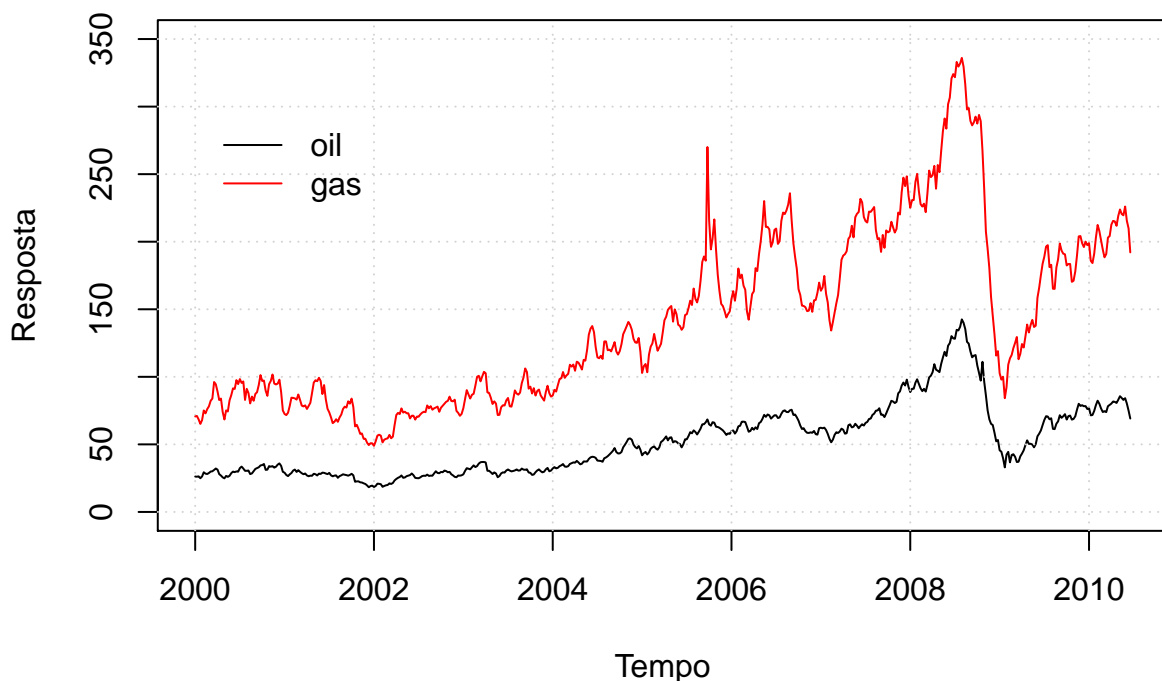


```
head(gas)
```

```
## [1] 70.636 71.040 68.490 65.137 67.918 75.117
```

- a) A série temporal *oil* e *gas* são parecidas com um passeio aleatório com tendência. Eu acredito que as séries não são estacionárias porque a média depende do tempo e, aparentemente, há uma dependência temporal da autocorrelação. Ainda, se considerar que a série se aproxima de um passeio aleatório, então é possível afirmar que a série não é estacionária, uma vez que o passeio aleatório não é estacionário porque sua autocovariância depende diretamente do tempo.

```
plot(oil,  
     ylab = "Resposta", xlab = "Tempo",  
     ylim = c(0, 350), col = "black")  
grid()  
lines(x = gas, add = TRUE, col = "red")  
legend(2000, 300,  
      lty = c(1, 1),  
      col = c("black", "red"),  
      legend = c("oil", "gas"),  
      bty = "n",  
      cex = 1)
```



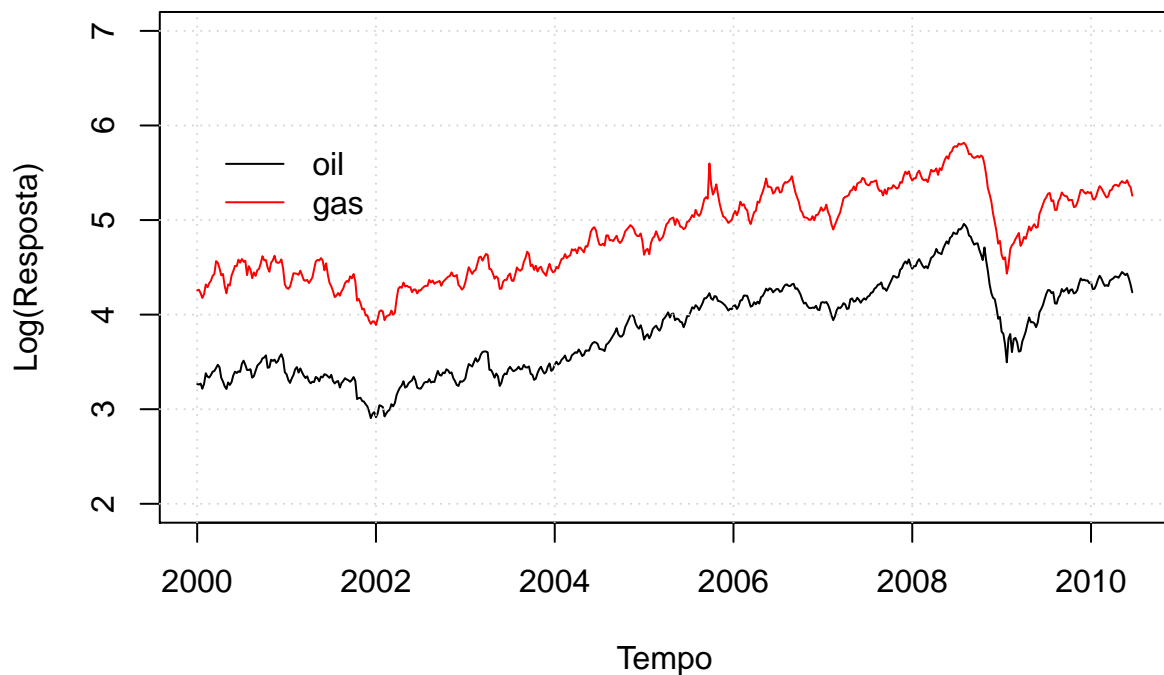
- (b) Em economia, muitas vezes é a variação percentual no preço, denominada taxa de crescimento ou retorno, em vez da mudança absoluta de preço, que é importante. Argumente que uma transformação da forma  $Y_t = \nabla \log(X_t)$  pode ser aplicada aos dados, onde  $X_t$  é a série de preços do petróleo ou gás.

Uma série temporal pode apresentar um padrão não estacionário, onde a variabilidade da série cresce ao longo do tempo. Uma transformação nos valores da série pelo logaritmo tende a tornar homogêneo o processo, devido as características da função de transformação. Assim, caso a série apresente valores bastante elevados ao longo do tempo, a função logarítmica tende a tornar os valores mais próximos e, então, reduzir as discrepância entre os valores na série. Ainda, a operação de diferenciação permite eliminar tendências na série temporal e tem como vantagem evitar a necessidade de estimar parâmetros desconhecidos. A primeira diferenciação permite eliminar tendências lineares, enquanto a segunda diferenciação elimina tendências quadráticas.

- (c) Transforme os dados conforme descrito na parte (b), mostre os dados no mesmo gráfico, observe os exemplos de ACFs dos dados transformados e comente.

Os dados transformados pela função logarítmica está apresentada na Figura dada na sequência.

```
plot(log(oil),
     ylab = "Log(Resposta)", xlab = "Tempo",
     ylim = c(2, 7), col = "black")
grid()
lines(x = log(gas), add = TRUE, col = "red")
legend(2000, 6,
      lty = c(1, 1),
      col = c("black", "red"),
      legend = c("oil", "gas"),
      bty = "n",
      cex = 1)
```

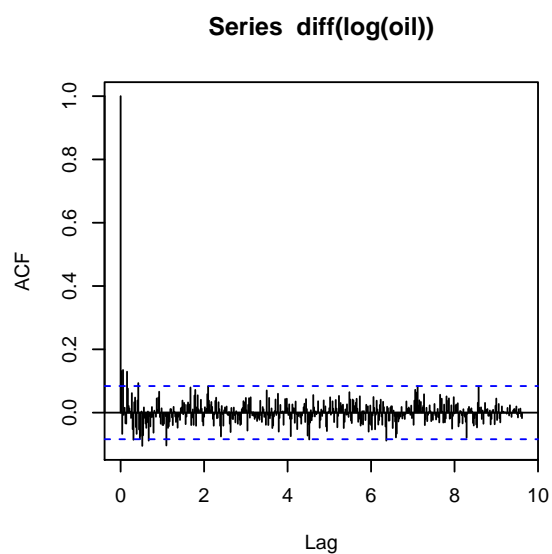
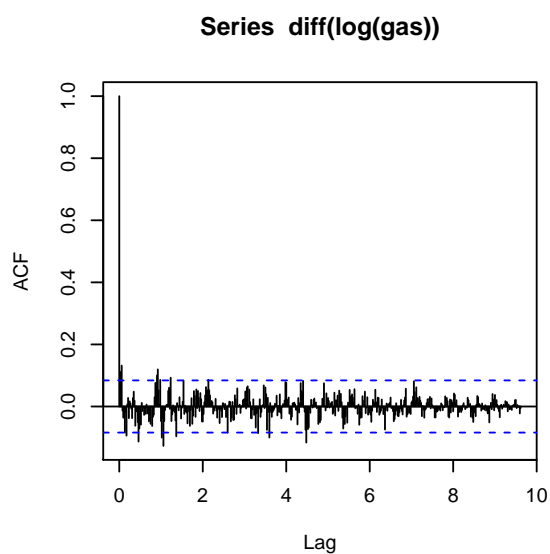
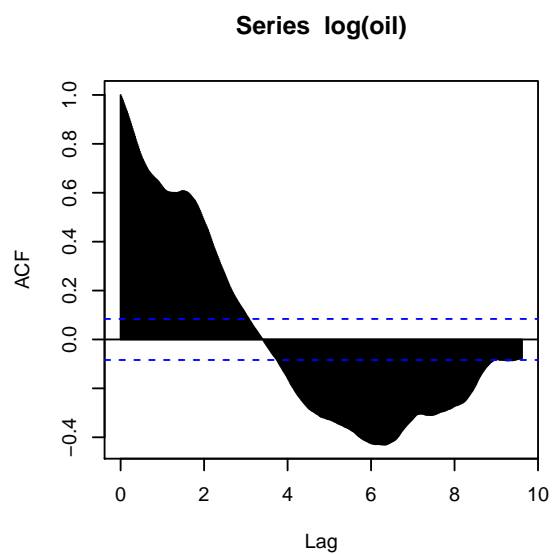
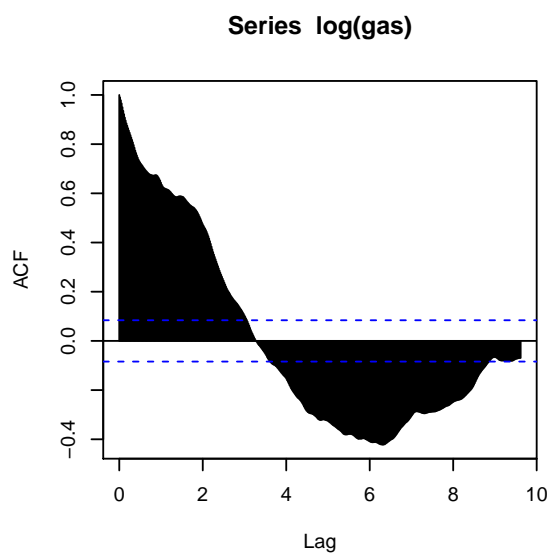
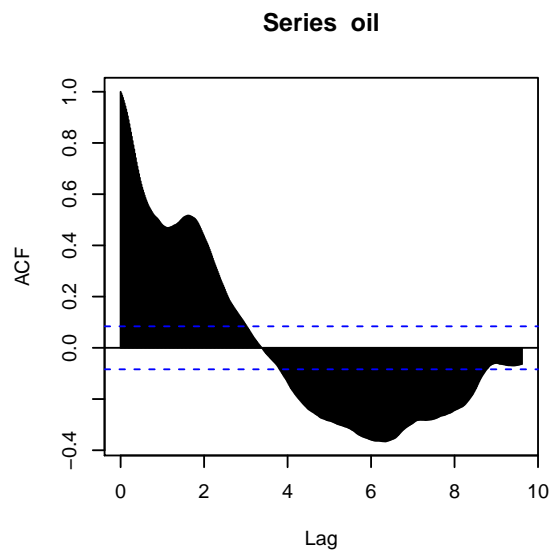
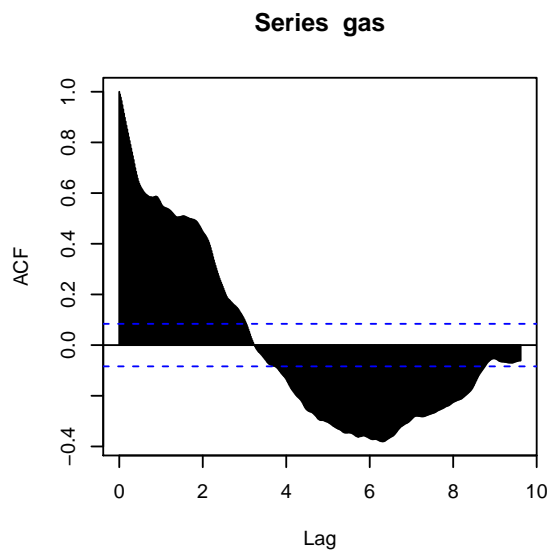


A Figura apresentada na sequência corresponde a função de autocorrelação amostral, para os dados transformação pelo logaritmo e sem transformação. A transformação reduziu ligeiramente a autocorrelação, mas é quase imperceptível essa diferença nos gráficos apresentados para ambas as séries temporais. Ainda, nota-se que a diferenciação foi capaz de eliminar a correlação entre as observações.

```
par(mfrow = c(3,2))
acf(gas, 500)
acf(oil, 500)

acf(log(gas), 500)
acf(log(oil), 500)

acf(diff(log(gas)), 500)
acf(diff(log(oil)), 500)
```

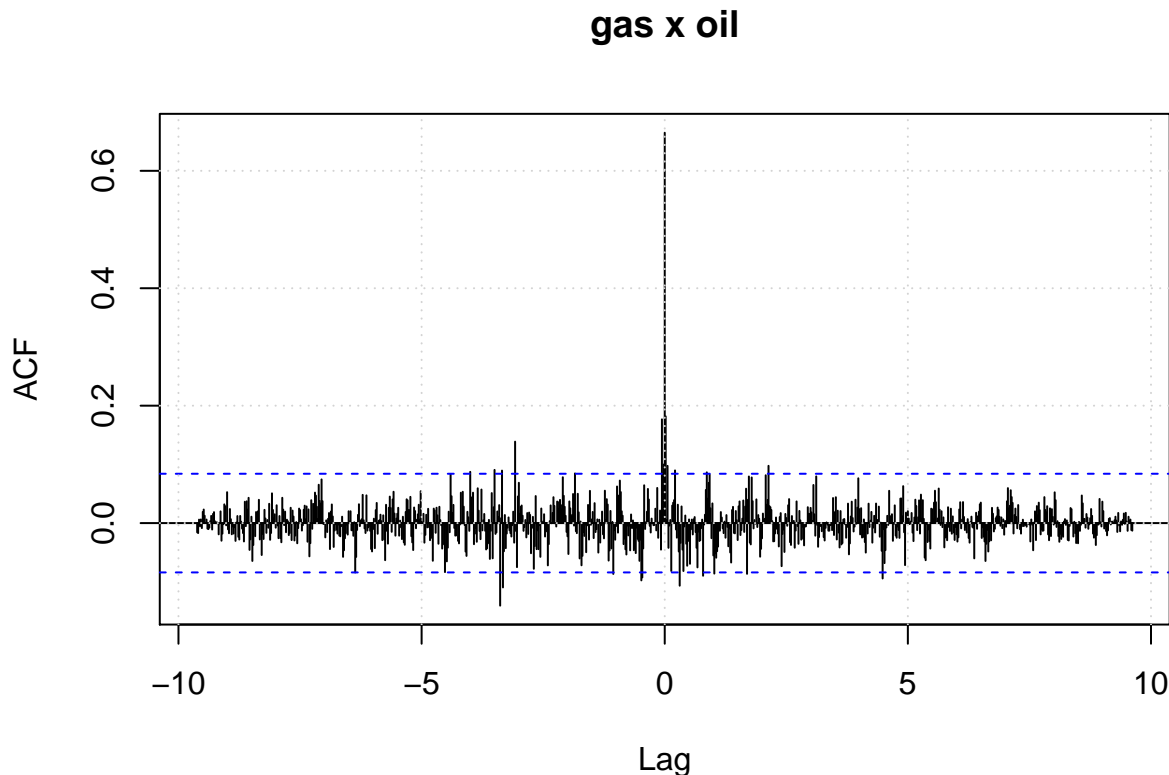


```
par(mfrow = c(1,1))
```

- (d) Grafique o CCF dos dados transformados e comente. Os valores pequenos, mas significativos, quando o gás conduz o petróleo podem ser considerados como explicativos.

Os dados transformados pela função logarítmica e a diferenciação estão apresentados na Figura dada na sequência. É possível notar que a transformação foi capaz de eliminar a correlação entre as duas variáveis.

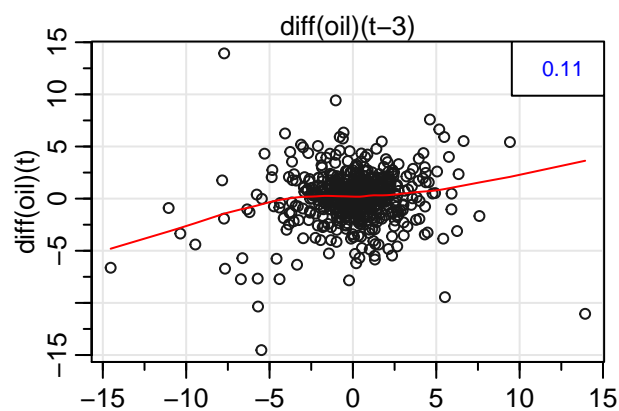
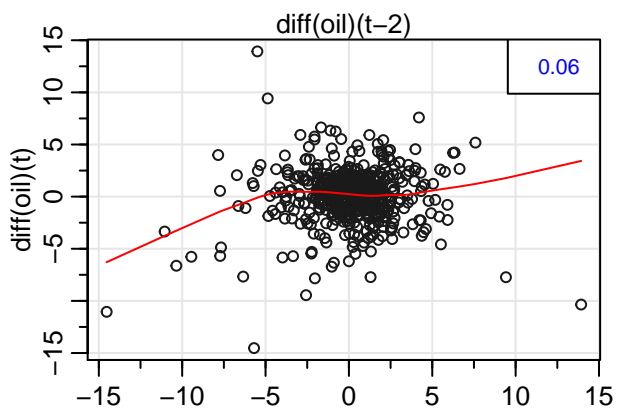
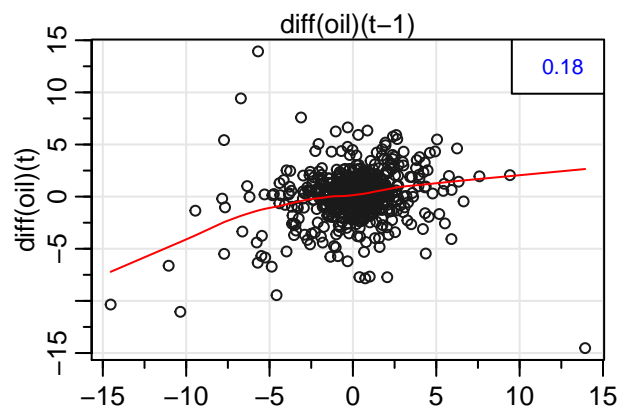
```
ccf(diff(log(gas)), diff(log(oil)), 500, main = "gas x oil")  
grid()
```



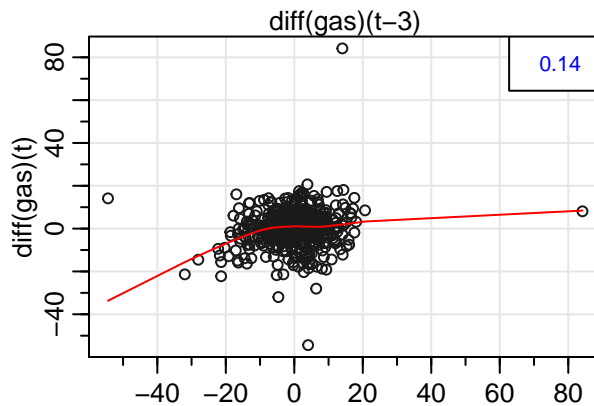
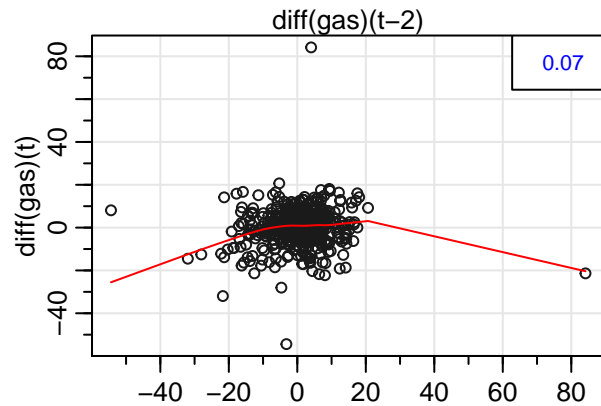
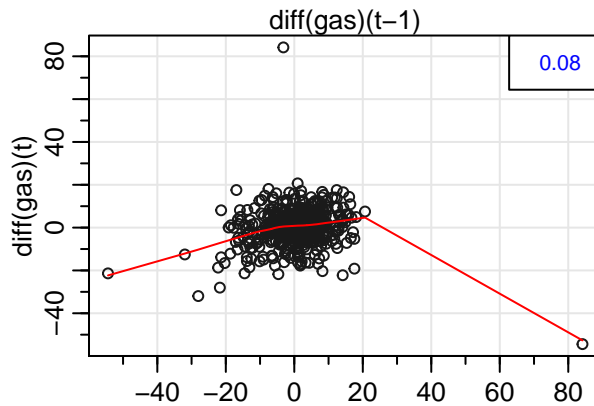
- (e) Exiba gráficos de dispersão da série de taxas de crescimento de petróleo e gás por até três semanas do prazo de entrega dos preços do petróleo; inclua um suavizador não paramétrico em cada gráfico e comente os resultados, por exemplo, existem outliers? As relações são lineares?

Para variável *oil* e *gas*, nota-se que não há tendências aparentes em todas as três semanas. Porém, a curva suavizada (linha vermelha) resalta algumas tendências, as quais estão relacionadas com a presença de observações discrepantes, o que afeta o suavizador. Assim, é possível afirmar que não há correlação entre observações transformadas para os três períodos considerados, ou seja, não há comportamento de autoregressividade.

```
library(astsa)  
lag1.plot(diff(oil), max.lag = 3)
```



```
lag1.plot(diff(gas), max.lag = 3)
```



- (f) Tem havido uma série de estudos questionando se os preços da gasolina respondem mais rapidamente quando os preços do petróleo estão subindo do que quando os preços do petróleo estão caindo, chamamos esse efeito de assimetria dos preços. Vamos tentar explorar essa questão aqui com regressão retardada simples. Ignorando alguns problemas óbvios, como outliers e erros autocorrelacionados, por isso não será uma análise definitiva. Considere  $G_t$  e  $O_t$  denotar as taxas de crescimento de gás e petróleo.
- (g) Ajuste o seguinte modelo de regressão e comente os resultados:  $G_t = \alpha_1 + \alpha_2 I_t + \beta_1 O_t + \beta_2 O_{t-1} + W_t$ , onde  $I_t = 1$  se  $O_t \geq 0$  e 0 (zero) caso contrário, ou seja,  $I_t$  é o indicador de ausência de crescimento ou crescimento positivo no preço do petróleo. Dica: O seguinte código R pode ser útil.

O modelo ajustado apresentou todas as variáveis significativas ao nível de confiança de 95%. Ainda, Nota-se que as covariáveis estão relacionadas positivamente com a variável resposta, isto é, fixando as variáveis que não são de interesse, quando aumenta-se o valor da covariável de interesse, espera-se que a variável resposta aumente. Ainda, o coeficiente de determinação ajustado indica que as covariáveis explicam aproximadamente 0.45 da variação da variável resposta.

```
poil = diff(log(oil))
pgas = diff(log(gas))
indi = ifelse(poil < 0, 0, 1)
mess = ts.intersect(pgases, poil, poilL = lag(poil,-1), indi)
fit <- lm(pgases ~ poil + poilL + indi, data=mess)
summary(fit)
```

```
##
## Call:
## lm(formula = pgases ~ poil + poilL + indi, data = mess)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -0.18451 -0.02161 -0.00038 0.02176 0.34342
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.006445  0.003464  -1.860  0.06338 .
## poil        0.683127  0.058369  11.704 < 2e-16 ***
## poill       0.111927  0.038554   2.903  0.00385 **
## indi        0.012368  0.005516   2.242  0.02534 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04169 on 539 degrees of freedom
## Multiple R-squared:  0.4563, Adjusted R-squared:  0.4532
## F-statistic: 150.8 on 3 and 539 DF,  p-value: < 2.2e-16
```

- (ii) Qual é o modelo ajustado quando há um crescimento negativo no preço do petróleo no tempo  $t$ ? Qual é o modelo ajustado quando não há crescimento positivo do preço do petróleo? Esses resultados suportam a hipótese de assimetria?

O modelo ajustado quando há um crescimento no preço do óleo no tempo é dado por

$$\hat{G}_t = \hat{\alpha}_1 + \hat{\alpha}_2 I_t + \hat{\beta}_1 O_t + \hat{\beta}_2 O_{t-1}$$

$$\hat{G}_t = -0.0064 + 0.0124 I_t + 0.6831 O_t + 0.1120 O_{t-1},$$

quando  $O_t \geq 0$ . O modelo ajustado quando não há crescimento é dado por

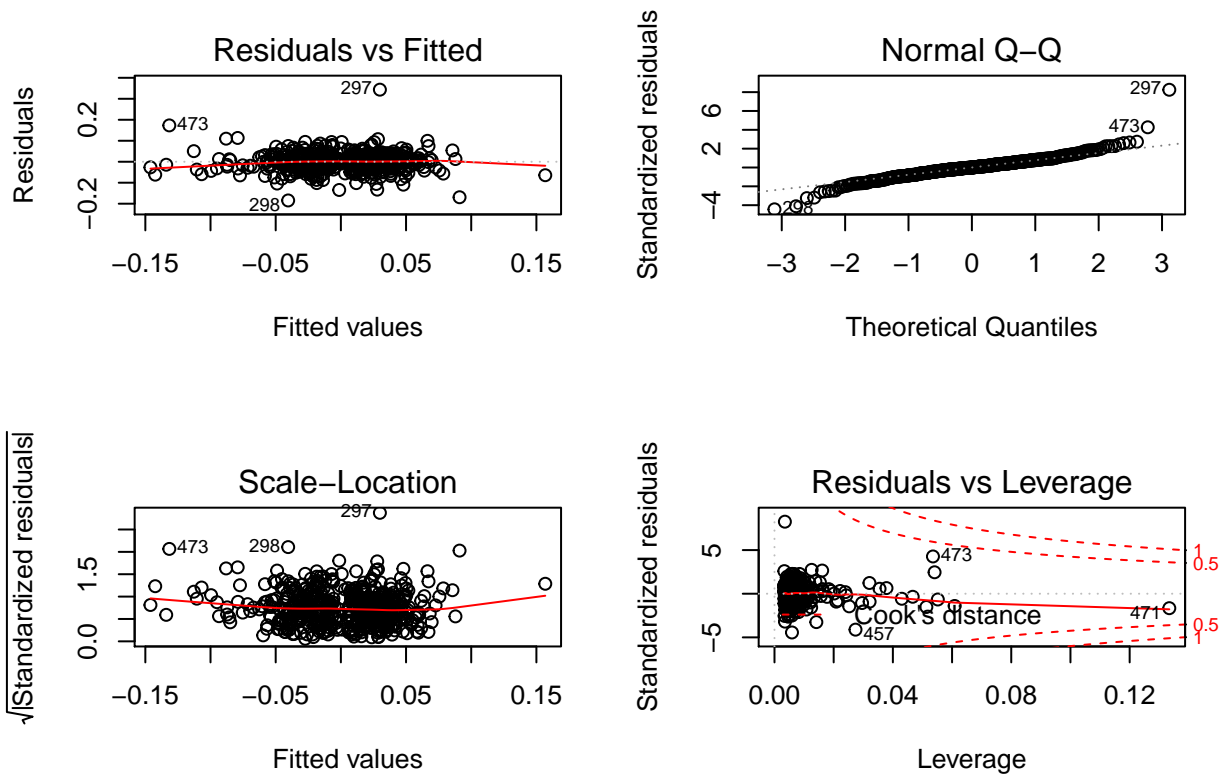
$$\hat{G}_t = \hat{\alpha}_1 + \hat{\beta}_1 O_t + \hat{\beta}_2 O_{t-1}$$

$$\hat{G}_t = -0.0064 + 0.6831 O_t + 0.1120 O_{t-1}.$$

- (iii) Analise os resíduos do ajuste e comente.

Os resíduos do modelo de regressão linear ajustado está apresentado na Figura a seguir. É possível notar que não há relação entre os valores ajustados e o resíduos do modelo. O gráfico qqplot evidencia um padrão de normalidade dos resíduos do modelo, pois há uma evidente relação linear entre os resíduos padronizados e os quantis teóricos. Apenas uma observação apresenta valor um pouco mais discrepante que os demais, o qual deve ser investigado com mais detalhe. Por fim, nenhuma observação aparenta influenciar de modo significativo o ajuste do modelo. Logo, é possível assumir que o modelo é confiável.

```
par(mfrow = c(2,2))
plot(fit)
```



```
par(mfrow = c(1,1))
```

## 2.3 Questão 10

Use duas técnicas diferentes de suavização descritas na Seção II.3 para estimar a tendência na série de temperatura global **globtemp**. Comente.

- Suavização LOWESS: é um método de suavização considerando vizinhos próximos. Neste exemplo, considerou-se os ajustes separando os dados em conjuntos de 10%, com a finalidade de obter curvas suaves para cada grupo. É importante ressaltar que o modelo ajustado é contínuo.

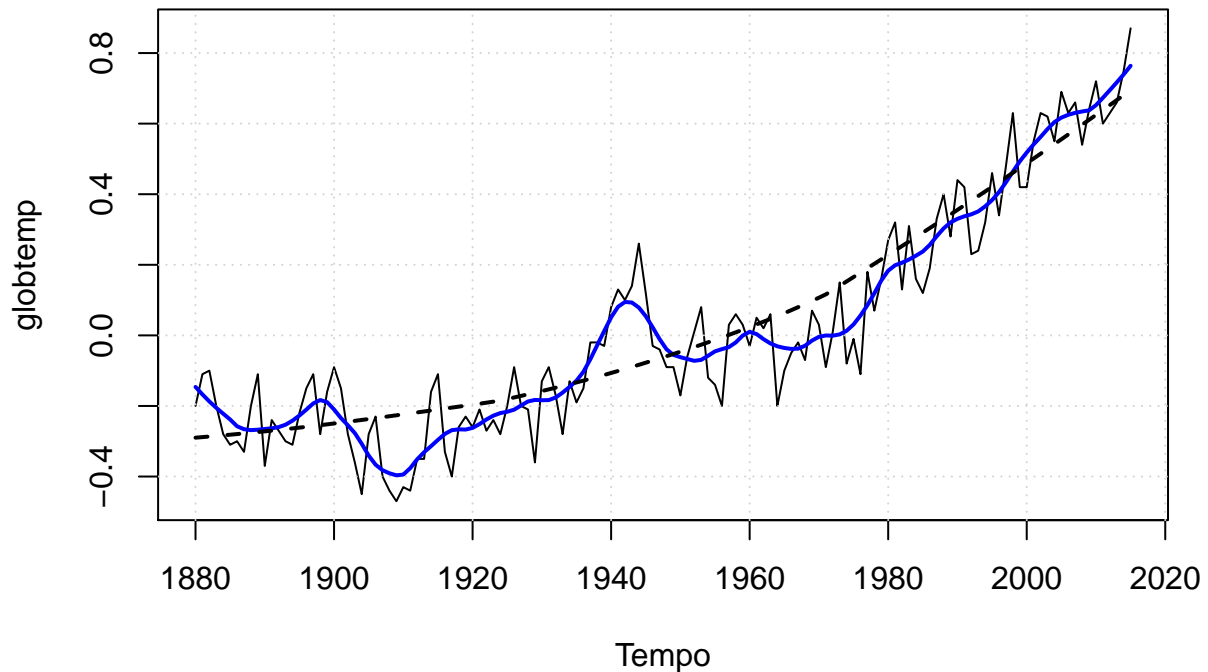
```
#install.packages("astsa")
library(astsa)

# Dados
#str(globtemp)

plot(globtemp, main = "Desvios de temperatura global anual", xlab = "Tempo")
grid()
lines(lowess(globtemp, f = 0.1), lwd = 2, col = 4)
lines(lowess(globtemp), lty = 2, lwd = 2, col = 1)
```



## Desvios de temperatura global anual



- Suavização por Splines: é um método de suavização considerando o ajuste de modelos polinomiais em função do tempo. No entanto, os dados são estratificados em função da variável tempo e esses componentes são conectados por nós. Assim, um modelo polinomial é ajustado para cada intervalo, os quais são conectados de forma contínua. Neste exemplo, ajustou-se modelos polinomiais de ordem três com 10 nós. A escolha por dividir em 10 nós (curva em azul) foi definida empiricamente. Porém, é possível otimizar o número de nós. A curva em verde foi ajustada considerando 25 nós, enquanto a curva em vermelha foi ajustada com 100 nós.

```
#install.packages("astsa")
library(astsa)

# Dados
#str(globtemp)

plot(globtemp, main = "Desvios de temperatura global anual", xlab = "Tempo")
grid()
lines(smooth.spline(time(globtemp), globtemp, nknots = 10), lwd = 2, col = 5)
lines(smooth.spline(time(globtemp), globtemp, nknots = 25), lty = 2, lwd = 2, col = 6)
lines(smooth.spline(time(globtemp), globtemp, nknots = 50), lty = 2, lwd = 2, col = 2)
```

## Desvios de temperatura global anual

