Cléber da Costa Figueiredo

Simulações de modelos ARIMA(p, d, q)

1

Programa de Aperfeiçoamento de Ensino

Supervisora: Profª Clélia Maria de Castro Toloi

2

ARIMA(0, 0, 0)

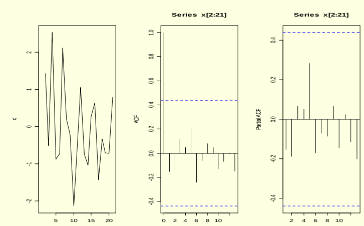
- Se simularmos um modelo ARIMA (0, 0, 0), teremos apenas um ruído
- arima.sim(list(order = c(0,0,0)), n = 200 , sd = sqrt(1))

Time Series:
Start = 1
End = 201
Frequency = 1

- Há um pequeno "bug" e para termos o correlograma e o correlograma parcial, fazemos:
- acf(x[2:201])
- pacf(x[2:201])

3

Resultado



4

Lembremo-nos dos modelos AR.

- Processo AR(p) será estacionário se e somente se ϕ(B)=0 tiver todas as suas raízes, em módulo, maiores que 1.
- Será invertível se e somente se puder ser escrito na forma

$$\pi_0 \tilde{z}_t = \pi_1 \tilde{z}_{t-1} + \pi_2 \tilde{z}_{t-2} + ... + a_t$$

- Com

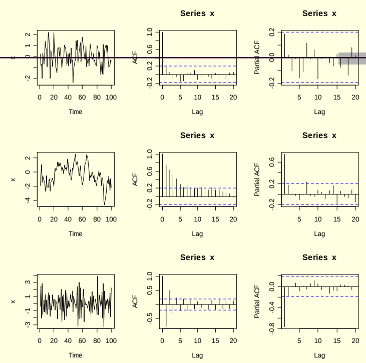
$$\sum_{j=0}^{\infty} \pi_j < \infty$$

5

AR(1)

- O modelo AR(1) pode ser simulado como um ARIMA(1,0,0)
- par(mfrow=c(3,3))
- x=arima.sim(n=100,list(ar= 0.2)); plot(x); acf(x); pacf(x)
- x=arima.sim(n=100,list(ar= 0.8)); plot(x); acf(x); pacf(x)
- x=arima.sim(n=100,list(ar=-0.8)); plot(x); acf(x); pacf(x)
- Para exibir as auto correlações usamos acf(x)$acf

6



7

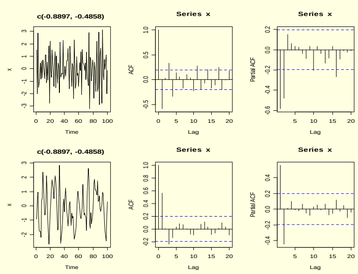
AR(p)

- Basta indicar quem é p
- Observe estas séries com p=2:

- par(mfrow=c(2,3))
- x<-arima.sim(n = 200, list(ar = c(0.8897, 0.4858))); plot(x); acf(x); pacf(x)
- x<-arima.sim(n = 200, list(ar = c(-0.8897, -0.4858))); plot(x,main='c(-0.8897, -0.4858) '); acf(x); pacf(x)
- x<-arima.sim(n = 200, list(ar = c(0.8897, -0.4858))); plot(x, main='c(0.8897, -0.4858) '); acf(x); pacf(x)

8

Ops! Cadê a série ar = c(0.8897, 0.4858) ?



9

Não vá simulando qualquer coisa!

- É necessário verificar se

$$\phi_1 + \phi_2 < 1$$
$$\phi_2 - \phi_1 < 1$$
$$|\phi_2| < 1$$

- A série ar = c(0.8897, 0.4858) não satisfaz a primeira condição, logo o R avisará que

Erro em arima.sim(n = 200, list(ar = c(0.8897, 0.4858))) : parte 'ar' do modelo é não-estacionária

10

Lembremo-nos dos modelos MA.

- Z_t é um processo MA(q) se e somente se puder ser escrito como

$$z_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}$$

- Será invertível se e somente se as raízes de $\theta(B)=0$ tiver todas as suas raízes, em módulo, maiores que 1.
- Com

$$\theta_1 + \theta_2 < 1$$
$$\theta_2 - \theta_1 < 1$$
$$|\theta_2| < 1$$

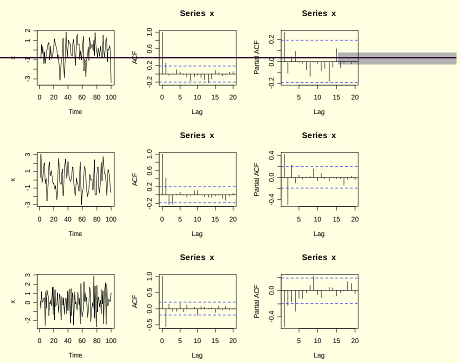
11

MA(1)

- O modelo MA(1) pode ser simulado como um ARIMA(0,0,1)

- par(mfrow=c(3,3))
- x=arima.sim(n=200,list(ma= 0.2)); plot(x); acf(x); pacf(x)
- x=arima.sim(n=200,list(ma= 0.8)); plot(x); acf(x); pacf(x)
- x=arima.sim(n=200,list(ma=-0.8)); plot(x); acf(x); pacf(x)

12

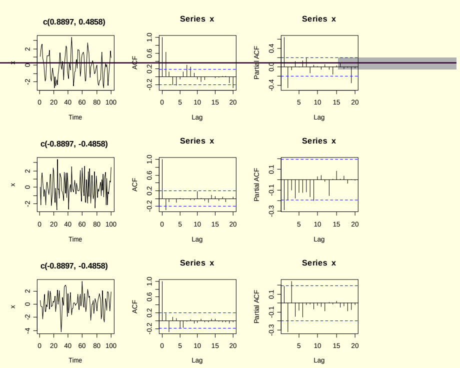


13

MA(q)

- Basta indicar quem é q
- Observe estas séries com q=2:

- par(mfrow=c(3,3))
- x<-arima.sim(n = 200, list(ma = c(0.8897, 0.4858))); plot(x,main='c(0.8897, 0.4858) '); acf(x); pacf(x)
- x<-arima.sim(n = 200, list(ma = c(-0.8897, -0.4858))); plot(x,main='c(-0.8897, -0.4858) '); acf(x); pacf(x)
- x<-arima.sim(n = 200, list(ma = c(0.8897, -0.4858))); plot(x, main='c(0.8897, -0.4858) '); acf(x); pacf(x)



15

Ueeeé!?

- A primeira condição de

$$\theta_1 + \theta_2 < 1$$
$$\theta_2 - \theta_1 < 1$$
$$|\theta_2| < 1$$

não está satisfeita, logo ma = c(0.8897, 0.4858) não será invertível.

- Há uma nota que diz que PARA TODO θ_1 e θ_2 MA(2) é estacionário!!!

16

Ajuste de modelos

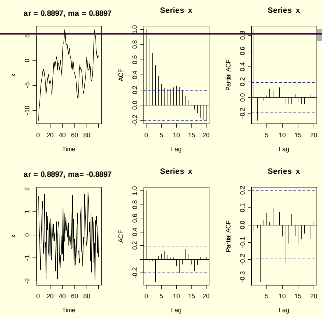
- x<-arima.sim(n = 200, list(ma = c(0.8897, 0.4858))); plot(x,main='c(0.8897, 0.4858) '); acf(x); pacf(x)
- m1 = arima(x, order = c(0,0,2)) aic = 537.1
- m2 = arima(x, order = c(0,0,1)) aic = 597.83
- m3 = arima(x, order = c(1,0,1)) aic = 556.49
- m4 = arima(x, order = c(1,0,2)) aic = 538.94
- m5 = arima(x, order = c(0,1,2)) aic = 591.37
- m6 = arima(x, order = c(1,0,1), include.mean = F) aic = 555.24
- m7 = arima(x, order = c(0,0,2), include.mean = F) aic = 536.29 (MENOR!)
- m8 = arima(x, order = c(0,0,1), include.mean = F) aic = 598.06
- m9 = arima(x, order = c(1,0,1), include.mean = F) aic = 555.24
- m10 = arima(x, order = c(1,0,2), include.mean = F) aic = 538.19
- m11 = arima(x, order = c(0,1,2), include.mean = F) aic = 591.37

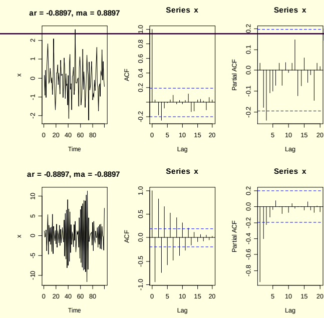
ARMA(1,1) = ARIMA(1, 0, 1)

- O processo ARMA(1, 1) será estacionário se e somente se $|\theta_1|$ < 1 e será invertível se $|\phi_1|$ < 1
- par(mfrow=c(2,3))
- x<-arima.sim(n = 200, list(ar = 0.8897, ma = 0.8897)); plot(x,main='ar = 0.8897, ma = 0.8897 '); acf(x); pacf(x)
- x<-arima.sim(n = 200, list(ar = 0.8897, ma = -0.8897)); plot(x,main='ar = 0.8897, ma= -0.8897 '); acf(x); pacf(x)
- par(mfrow=c(2,3))
- x<-arima.sim(n = 200, list(ar = -0.8897, ma = 0.8897)); plot(x,main='ar = -0.8897, ma = 0.8897'); acf(x); pacf(x)
- x<-arima.sim(n = 200, list(ar = -0.8897, ma = -0.8897)); plot(x,main='ar = -0.8897, ma = -0.8897 '); acf(x); pacf(x)

OBS, haverá cancelamento de raízes, pq os coeficientes são iguais!!!

18





ARIMA(1, 0, 1) – Estes funcionam!

- par(mfrow=c(2,3))
- x<-arima.sim(n = 200, list(ar = 0.8897, ma = 0.5897)); plot(x,main='ar = 0.8897, ma = 0.5897 '); acf(x); pacf(x)
- x<-arima.sim(n = 200, list(ar = 0.8897, ma = -0.5897)); plot(x,main='ar = 0.8897, ma = -0.5897 '); acf(x); pacf(x)
- par(mfrow=c(2,3))
- x<-arima.sim(n = 200, list(ar = -0.8897, ma = 0.5897)); plot(x,main='ar = -0.8897, ma = 0.5897'); acf(x); pacf(x)
- x<-arima.sim(n = 200, list(ar = -0.8897, ma = -0.5897)); plot(x,main='ar = -0.8897, ma = -0.5897 '); acf(x); pacf(x)

21

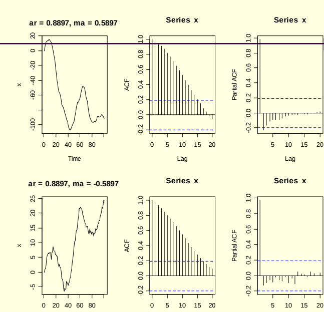
Ajuste uns modelos!

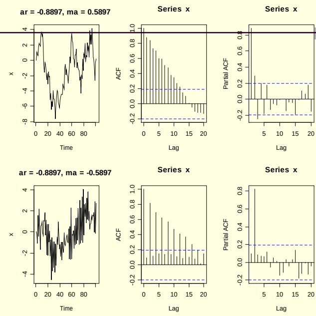
- x<-arima.sim(n = 200, list(ma = 0.8897, ar = 0.5897));
- m1 = arima(x, order = c(2,0,2))
- m2 = arima(x, order = c(1,0,1))
- m3 = arima(x, order = c(2,0,1))
- m4 = arima(x, order = c(3,0,1))
- m5 = arima(x, order = c(2,1,0))
- m6 = arima(x, order = c(2,0,2), include.mean = F)
- m7 = arima(x, order = c(1,0,1), include.mean = F)
- m8 = arima(x, order = c(2,0,1), include.mean = F)
- m9 = arima(x, order = c(3,0,1), include.mean = F)
- m10 = arima(x, order = c(2,1,0), include.mean = F)
- m11 = arima(x, order = c(3,0,2), include.mean = F)

ARIMA(1, 1, 1) – Até que enfim!

- par(mfrow=c(2,3))
- x<-arima.sim(list(order = c(1,1,1), ar = 0.8897, ma = 0.5897) , n = 200); plot(x,main='ar = 0.8897, ma = 0.5897 '); acf(x); pacf(x)
- x<-arima.sim(list(order = c(1,1,1), ar = 0.8897, ma = -0.5897) , n = 200); plot(x,main='ar = 0.8897, ma = -0.5897 '); acf(x); pacf(x)
- par(mfrow=c(2,3))
- x<-arima.sim(list(order = c(1,1,1), ar = -0.8897, ma = 0.5897) , n = 200); plot(x,main='ar = -0.8897, ma = 0.5897 '); acf(x); pacf(x)
- x<-arima.sim(list(order = c(1,1,1), ar = -0.8897, ma = -0.5897) , n = 200); plot(x,main='ar = -0.8897, ma = -0.5897 '); acf(x); pacf(x)

23



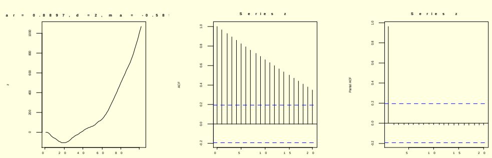


Ajuste uns modelos!

- x<-arima.sim(list(order = c(1,1,1), ar = 0.8897, ma = 0.5897) , n = 200); plot(x,main='ar = 0.8897, ma = 0.5897 '); acf(x); pacf(x)
- m1 = arima(x, order = c(1,0,1))
- m2 = arima(x, order = c(2,0,0))
- m3 = arima(x, order = c(1,1,0))
- m4 = arima(x, order = c(2,1,1))
- m5 = arima(x, order = c(1,1,1))
- m6 = arima(x, order = c(1,2,1))
- m7 = arima(x, order = c(1,3,1))
- m8 = arima(x, order = c(1,0,1), include.mean = F)
- m9 = arima(x, order = c(2,0,0), include.mean = F)
- m10 = arima(x, order = c(1,1,0), include.mean = F)
- m11 = arima(x, order = c(2,1,1), include.mean = F)
- m12 = arima(x, order = c(1,1,1) , include.mean = F)
- m13 = arima(x, order = c(1,2,1), include.mean = F)
- m14 = arima(x, order = c(1,3,1), include.mean = F)

ARIMA(1,2,1)

- z<-arima.sim(list(order = c(1,2,1), ar = 0.8897, ma = -0.5897) , n = 200); plot(z,main='ar = 0.8897, d =2, ma = -0.5897'); acf(z); pacf(z)



Ajustes

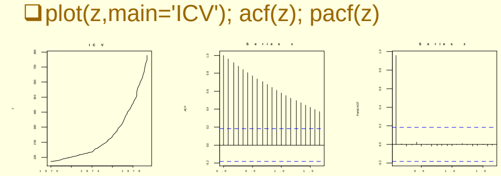
- z<-arima.sim(list(order = c(1,2,1), ar = 0.8897, ma = -0.5897) , n = 200); plot(z,main='ar = 0.8897, d =2, ma = 0.5897'); acf(z); pacf(z)
- m1 = arima(z, order = c(1,0,1))
- m2 = arima(z, order = c(2,0,0))
- m3 = arima(z, order = c(1,1,0))
- m4 = arima(z, order = c(2,1,1))
- m5 = arima(z, order = c(1,1,1))
- m6 = arima(z, order = c(1,2,1))
- m7 = arima(z, order = c(1,3,1))

Exemplo de identificação e estimação usando a série ICV

(Morettin e Toloi, 2004), janeiro de 1970 a junho de 1979.
Disponíveis em
http://www.ime.usp.br/ pam/icv.txt

Leitura dos dados:

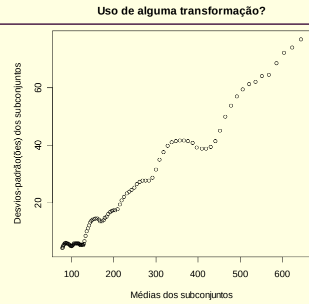
- z = scan(file='http://www.ime.usp.br/~pam/icv.txt')
- z = ts(z,freq=12,start=c(1970,1))
- plot(z,main='ICV'); acf(z); pacf(z)



Verificação da necessidade do uso de alguma transformação

- n<- length(z)
- j<-12 # dados mensais #
- abcissa<- NULL
- ordenada<- NULL
- for (i in 1: (n-j)){
- abcissa[i]<-mean(z[i:(i+j)])
- ordenada[i]<-sd(z[i:(i+j)])
- }
- plot(x=abcissa,y=ordenada, xlab='Médias dos subconjuntos', ylab= 'Desvios-padrão(*ões*) dos subconjuntos', main='Uso de alguma transformação?')

31



y = log(z)

- transformação logarítmica usada para induzir simetria e tentar estabilizar a variância.

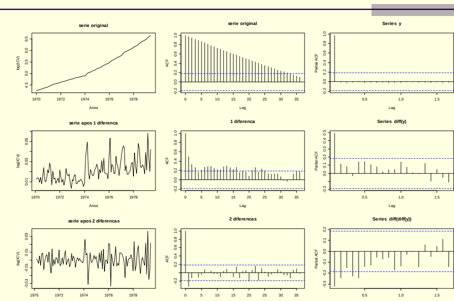
```
par(mfrow=c(3,3))

plot(y,main='serie original',xlab='Anos',ylab='log(ICV)')
m = acf(y,lag.max=36, plot=F)
m$lag = m$lag*12
plot(m,main='serie original')
pacf (y)

plot(diff(y),main='serie apos 1 diferenca',xlab='Anos',ylab='log(ICV)')
m = acf(diff(y),lag.max=36, plot=F)
m$lag = m$lag*12
plot(m,main='1 diferenca')
pacf (diff(y))

plot(diff(diff(y)),main='serie apos 2 diferencas',xlab='Anos',ylab='log(ICV)')
m = acf(diff(diff(y)),lag.max=36, plot=F)
m$lag = m$lag*12
plot(m,main='2 diferencas')
pacf (diff(diff(y)))
```

O que parece melhor?



Candidato:

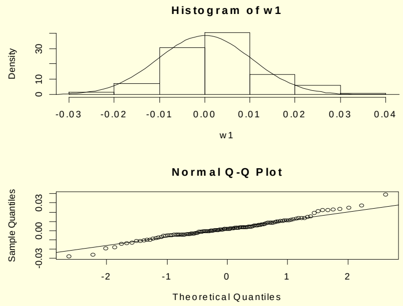
- O modelo candidato é então o ARIMA(1,1,0)
- m1 = arima(y,order=c(1,1,0))

- Observação: Nas simulações os melhores modelos (AIC menores) eram os que não possuíam intercepto, porque o simulador centraliza os dados. Na prática, precisamos do intercepto! É claro que podemos retirar o intercepto do modelo (include.mean = F), mas, em geral, teremos um pior ajuste!

Análise de resíduos (m1)

- tsdiag(m1)
- par(mfrow=c(2,1))
- w1=m1$residuals
- hist(w1,freq=F)
- d = seq(range(w1)[1]-3*sd(w1),range(w1)[2]+3*sd(w1),0.001)
- lines(d,dnorm(d,0,sd(w1)))
- qqnorm(w1)
- qqline(w1)
- shapiro.test(w1)

m1

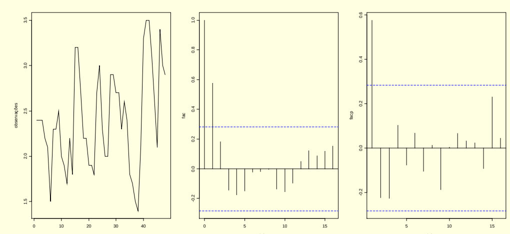


Lembram que...

- Alguns fenômenos necessariamente não devem apresentar tendência.
- No banco de dados do R, a série lh contém as quantidades de um tipo de hormônio em amostras de sangue coletadas a cada 10 minutos de uma pessoa do sexo feminino (Diggle, 1990).

- data(lh)
- lh
- par(mfrow = c(3,1))
- plot(lh,xlab='tempo',ylab='observações',main='')
- acf (lh,xlab='defasagem',ylab='fac',main='')
- pacf(lh,xlab='defasagem',ylab='facp',main='')

A série precisa ser diferenciada?



Candidato:

- O modelo candidato é então o ARIMA(1,0,0)
- m1 = arima(y,order=c(1,0,0))
- Observação: Nas simulações os melhores modelos (AIC menores) eram os que não possuíam intercepto, porque o simulador centraliza os dados. Na prática, precisamos do intercepto! É claro que podemos retirar o intercepto do modelo (include.mean = F), mas, em geral, teremos um pior ajuste!