Autocorrelação

1

Quem sou?

- Licenciado em matemática pelo IME-USP;
- Iniciação Científica sob orientação da Profª Drª Mônica Carneiro Sandoval;
- *Transformação de Box-Cox*.
- Mestre em estatística sob a orientação da Profª Drª Mônica Carneiro Sandoval e co-orientação da Profª Denise Aparecida Botter;
- *Incorporação do plano amostral na análise de regressão.*
- Atualmente, doutorando em estatística...sob a orientação do Profº Drº Heleno Bolfarine e Co-orientação da Profª Drª Mônica Carneiro Sandoval;
- *Calibração linear em modelos com assimetria*.

- Interesse:
  - Análise de Regressão,
  - Inferência Bayesiana e
  - Calibração Linear.

2

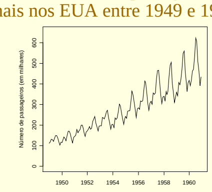
Programa de Aperfeiçoamento de Ensino

Supervisora: Profª Clélia Maria de Castro Toloi

3

Alguns exemplos do R - Passageiros

- Figura 1. Totais mensais de passageiros em linhas aéreas internacionais nos EUA entre 1949 e 1960.



- Existe uma clara tendência de crescimento bem como um padrão sazonal ao longo dos anos.

4

Comandos

- > data(AirPassengers)
- # para obter informações sobre os dados #
- > help(AirPassengers)

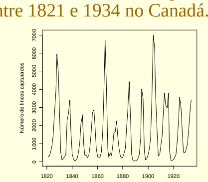
- > x = AirPassengers
- # testa se um objeto é uma série temporal #
- > is.ts(x)

- > plot(x,xlab='Anos',ylim=c(0,650),ylab='Número de passageiros (em milhares)')

5

Número de Linces

- Figura 2. Número anual de linces capturados em armadilhas entre 1821 e 1934 no Canadá.



- Observe o padrão um padrão cíclico em torno de 10 ou 11 anos.

6

Comandos

- > data(lynx)
- # para obter informações sobre os dados #
- > help(lynx)

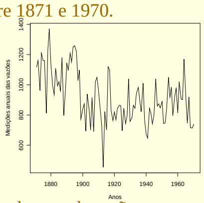
- > linces = lynx
- # testa se o objeto é uma série temporal #
- > is.ts(linces)

- > plot(linces,xlab='Anos', ylab='Número de linces capturados')

7

Vazões do Rio Nilo

- Figura 3. Medições anuais de vazões do Rio Nilo em Ashwan entre 1871 e 1970.



- Parece haver alguma alteração estrutural em torno do ano de 1900.

8

Comandos

- > data(Nile)
- # para obter informações sobre os dados #
- > help(Nile)

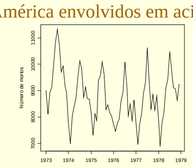
- > Nilo = Nile
- # testa se o objeto é uma série temporal #
- > is.ts(Nilo)

- > plot(Nilo,xlab='Anos', ylab='Medições Anuais das vazões')

9

Total mensal de mortos em acidentes

- Figura 4. Mostra o total mensal de mortos nos Estados Unidos da América envolvidos em acidentes.



- Parece que o número de mortos é maior durante os verões e esse padrão é observado ano a ano. Além disso, o número de mortos não parece aumentar, mantendo-se estacionário.

10

Comandos

- > data(USAccDeaths)
- # para obter informações sobre os dados #
- > help(USAccDeaths)

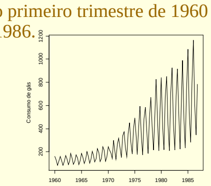
- > mortes = USAccDeaths
- # testa se o objeto é uma série temporal #
- > is.ts(mortes)

- > plot(mortes,xlab='Anos', ylab='Número de mortos')

11

Gás no Reino Unido

- Figura 5. Série trimestral do consumo de gás no Reino Unido entre o primeiro trimestre de 1960 e o quarto trimestre de 1986.



- Há uma tendência de crescimento porém a amplitude do padrão sazonal aumenta bastante a partir de 1971.

12

Comandos

- > data(UKgas)
- # para obter informações sobre os dados #
- > help(UKgas)

- > gas = UKgas
- # testa se o objeto é uma série temporal #
- > is.ts(gas)

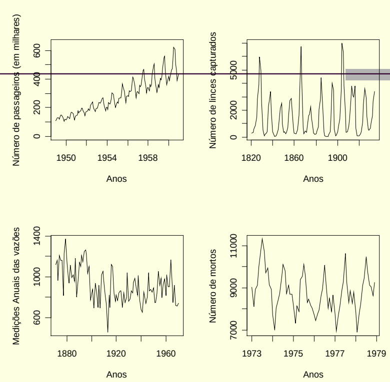
- > plot(gas,xlab='Tempo (em quadrimestres)', ylab='Consumo de gás')

13

Colocando quatro deles lado a lado.

- > par(mfrow=c(2,2))
- > plot(x,xlab='Anos',ylim=c(0,650),ylab='Número de passageiros (em milhares)')
- > plot(linces,xlab='Anos', ylab='Número de linces capturados')
- > plot(Nilo,xlab='Anos', ylab='Medições Anuais das vazões')
- > plot(mortos,xlab='Anos', ylab='Número de mortos')

14



15

Autocorrelação

- *Coeficientes de Autocorrelação*: Ferramenta para se identificar as propriedades de uma série temporal.

- Coeficiente de correlação linear...

$$r = \frac{S_{XY}}{S_X S_Y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

$(x_1,y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), (x_5, y_5), (x_6, y_6)$ . . .

16

Idéia de autocorrelação

- Queremos medir a correlação entre as observações de uma mesma variável em diferentes momentos de tempo.

- Seria ótimo se conseguíssemos entender a relação entre observações defasadas 1, 2, ..., períodos de tempo.

17

r_1

- Será que há correlação entre uma e outra observação?

$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, \ldots$

$(x_1,x_2), (x_2, x_3), (x_3, x_4), (x_4, x_5), (x_5, x_6), (x_6, x_7)$ . . .

18

r_2

- E entre a observação e a segunda vizinha?

$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, \ldots$

$(x_1,x_3), (x_2, x_4), (x_3, x_5), (x_4, x_6), (x_5, x_7), (x_6, x_8) \ldots$

19

r_3

- E com a terceira vizinha?...

$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, \ldots$

$(x_1,x_4), (x_2, x_5), (x_3, x_6), (x_4, x_7), (x_5, x_8), \ldots$

20

A idéia é criar novos pares de variáveis!

- Em r_1 os pares são $(x_1,x_2), (x_2, x_3), \ldots$
- Em r_2 os pares são $(x_1,x_3), (x_2, x_4), \ldots$
- Em r_3 os pares são $(x_1,x_4), (x_2, x_5), \ldots$

- E calcular as correlações dos novos pares de variáveis.

21

O coeficiente pode ser simplificado!

- Considerando $x_1, \ldots, x_{n-1}$ e $x_2, \ldots, x_n$ como duas variáveis o coeficiente de correlação entre x_t e x_{t+1}, por exemplo, é dado por:

$$r_1 = \frac{\sum_{t=1}^{n-1}(x_t - \overline{x}_1)(x_{t+1} - \overline{x}_2)}{\sqrt{\sum_{t=1}^{n-1}(x_t - \overline{x}_1)^2 \sum_{t=1}^{n-1}(x_{t+1} - \overline{x}_2)^2}}$$

22

Mais simplificações!

- Onde as médias amostrais são

$$\overline{x}_1 = \sum_{t=1}^{n-1} x_t/(n-1) \quad \text{e} \quad \overline{x}_2 = \sum_{t=2}^{n} x_t/(n-1).$$

- Se utilizarmos a média amostral total $\overline{x}$ e considerarmos que as duas médias acima são aproximadamente as mesmas.

23

- A versão simplificada do coeficiente de correlação fica:

$$r_1 = \frac{\sum_{t=1}^{n-1}(x_t - \overline{x})(x_{t+1} - \overline{x})}{(n-1)\sum_{t=1}^{n}(x_t - \overline{x})^2/n}$$

- Além disso, podemos excluir o fator n/(n-1) já que está próximo de 1 e temos um estimador bem mais simples!

24

Um estimador da autocorrelação

- A expressão anterior pode ser generalizada para calcular a correlação entre observações defasadas em k períodos de tempo.

$$r_k = \frac{\sum_{t=1}^{n-k}(x_t - \overline{x})(x_{t+k} - \overline{x})}{\sum_{t=1}^{n}(x_t - \overline{x})^2}$$

25

Observações

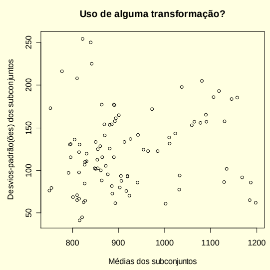
- Assim como o coeficiente de correlação usual, as autocorrelações são adimensionais e $-1 < r_k < 1$.
- Na prática é mais usual calcular primeiro os coeficientes de autocovariância $\{c_k\}$, definidos por analogia com a fórmula usual de covariância

$$c_k = \sum_{t=1}^{n-k}(x_t - \overline{x})(x_{t+k} - \overline{x})/n.$$

- Os coeficientes de autocorrelação são então obtidos como $r_k = c_k/c_0$.

26

Estudo da série de vazões do Nilo



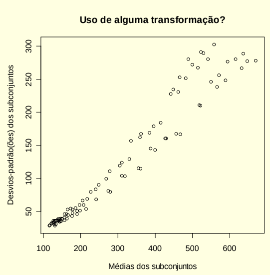
27

Verificação da necessidade do uso de alguma transformação

- j<-floor(n/20) # para garantir L = 20 #
- abcissa<- NULL
- ordenada<- NULL
- for (i in 1: (n-j)){
- abcissa[i]<-mean(Nilo[i:(i+j)])
- ordenada[i]<-sd(Nilo[i:(i+j)])
- }
- plot(x=abcissa,y=ordenada, xlab='Médias dos subconjuntos', ylab= 'Desvios-padrão(*ões*) dos subconjuntos', main='Uso de alguma transformação?')

28

Gás no Reino Unido – Log_e?



29

Estudo das autocorrelações

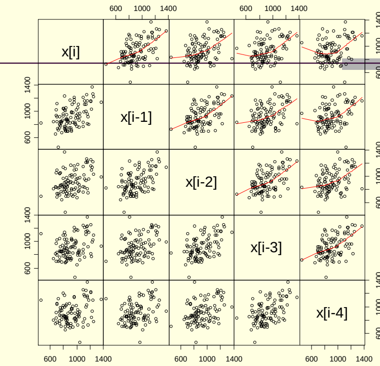
- Como exemplo do cálculo das autocorrelações vamos utilizar os dados das medições das vazões do Nilo disponível no R.
- Vamos montar uma matriz de correlogramas de valores defasados de 1, 2, 3 e 4 períodos.

30

Comandos

- Nilo <- Nile
- n <- length(Nilo)
- k <- 5
- m = NULL
- for (i in 1:k) m=cbind(m,lag(Nilo,k=i-1))
- m=as.data.frame(m)
- colnames(m) <- c("Nilo[i]", "Nilo[i-1]", "Nilo[i-2]", "Nilo[i-3]", "Nilo[i-4]")
- pairs(m, gap = 0, upper.panel = panel.smooth)

31

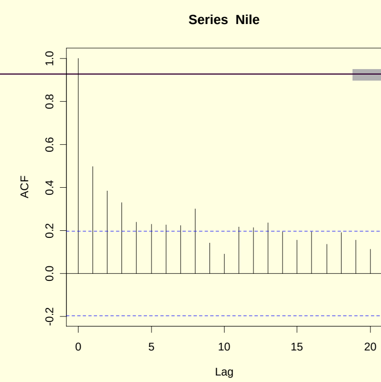


32

A função de autocorrelação - acf

- Para calcularmos todas as possíveis autocorrelações existe uma função chamada acf.
- No exemplo anterior, se tivéssemos digitado o comando:
- acf(Nile)
- Teríamos o correlograma de toda a série.

33



34

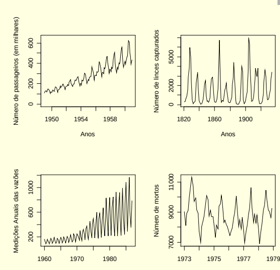
Limites de confiança

- Numa série aleatória esperamos que os valores defasados sejam não correlacionados e espera-se que $r_k = 0$.
- Podemos mostrar que r_k tem distribuição assintótica

  $r_k \sim N(-1/n\ ;\ 1/n)$.
- Assim, os limites de confiança aproximados de 95% são dados por $-1/n \pm 1,96/\sqrt{n}$, que são frequentemente ainda mais aproximados para $\pm 1,96/\sqrt{n}$.

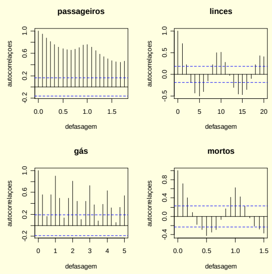
35

Correlogramas dos Exemplos



36

Correlogramas



37

Comandos

```
par(mfrow=c(2,2))
plot(x,xlab='Anos',ylim=c(0,650),ylab='Número de passageiros (em
    milhares)')
plot(linces,xlab='Anos', ylab='Número de linces capturados')
plot(gas,xlab='Anos', ylab='Medições Anuais das vazões')
plot(mortes,xlab='Anos', ylab='Número de mortos')

par(mfrow=c(2,2))
acf(x, main=
    'passageiros',xlab='defasagem',ylab='autocorrelaçoes')
acf(linces, main='linces',xlab='defasagem',ylab='autocorrelaçoes')
acf(gas, main='gás',xlab='defasagem',ylab='autocorrelaçoes')
acf(mortes,
    main='mortos',xlab='defasagem',ylab='autocorrelaçoes')
```

38

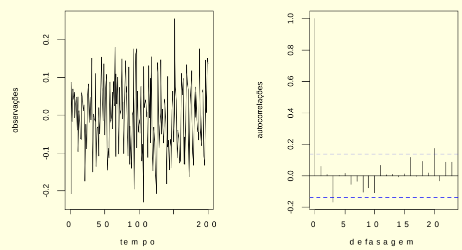
Gerando séries temporais artificiais

1. Série aleatória, observações iid da distribuição N(0,1)

```
x = rnorm(200,0,0.1)
par(mfrow=c(1,2))
plot.ts(x,xlab='tempo',ylab='observações')
acf(x,main='',xlab='defasagem',ylab='autocorrelações')
```

39

Saídas - ruído



40

Gerando séries temporais artificiais

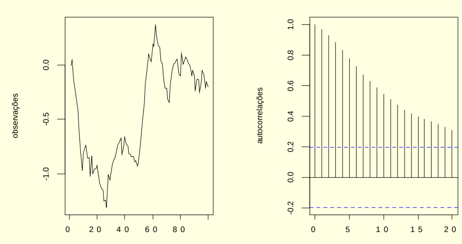
2. Série com tendência,

$$x_t = x_{t-1} + N(0;0{,}01)$$

```
e = rnorm(100,0,0.1)
x = cumsum(e)
par(mfrow=c(1,2))
plot.ts(x,xlab='tempo',ylab='observações')
acf(x,main='',xlab='defasagem',ylab='autocorrelações')
```

41

Saída – passeio aleatório



42

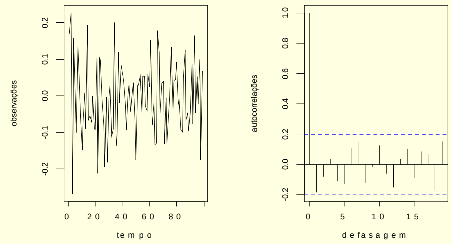
A função diff(...)

- É interessante notar que a primeira diferença de um passeio aleatório é estacionária!

z<-diff(x)
par(mfrow=c(1,2))
plot.ts(z,xlab='tempo',ylab='observações')
acf(z,main='',xlab='defasagem',ylab='autocorrelações')

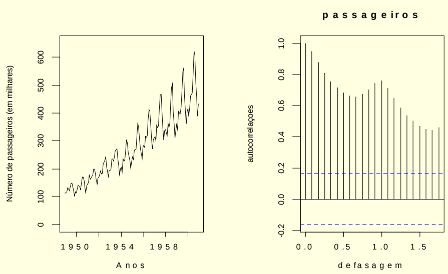
43

Agora é estacionária!



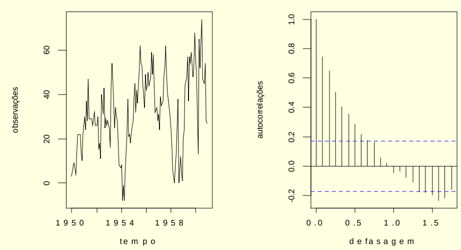
44

Vamos usar diff(...) - Passageiros



45

z<-diff(x,lag=12)
par(mfrow=c(1,2))
plot.ts(z,xlab='tempo',ylab='observações')
acf(z,main='',xlab='defasagem',ylab='autocorrelações')



46

Gerando séries temporais artificiais

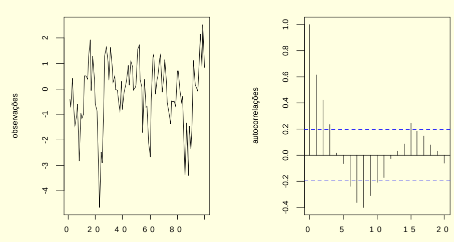
3. Série com correlação de curto-prazo

$$x_t = 0,7x_{t-1} + e_t$$

Curto<-arima.sim(n = 400, list(ar = 0.7))
par(mfrow=c(1,2))
plot.ts(Curto,xlab='tempo',ylab='observações')
acf(Curto,main='',xlab='defasagem',ylab='autocorrelações')

47

Saída – memória de curto prazo



48

Gerando séries temporais artificiais

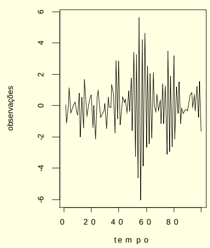
4. Série com correlação negativa

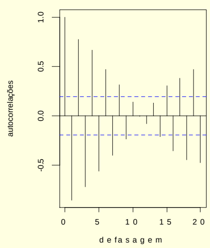
$$x_t = -0{,}8x_{t-1} + e_t$$

```
negativo<-arima.sim(n = 100, list(ar = -0.8))
par(mfrow=c(1,2))
plot.ts(negativo,xlab='tempo',ylab='observações')
acf(negativo,main='',xlab='defasagem',ylab='autocorrelações')
```

49

Saída – correlações negativas





50