

Econometria e Séries Temporais com R

Marcio Valk e Guilherme Pumi

2020-08-13

Contents

	5
Prefácio	7
1 Séries temporais	9
1.1 Breve introdução ao R	10
1.2 Apresentação da linguagem R	10
1.3 Instalando o R	10
1.4 Instalando o RStudio	10
1.5 Diretório de trabalho	11
1.6 RStudio Cloud	11
1.7 Instalação de pacotes	13
1.8 Ajuda	13
1.9 Funcionalidades básicas	15
1.10 Estrutura dos dados	20
1.11 Programação básica	32
1.12 Funções	34
1.13 Operador Pipe	35
2 Revisão	37
2.1 Aleatoriedade, a essência da estatística	37
2.2 Variável aleatória	37
2.3 A distribuição Normal e distribuições relacionadas	41
2.4 Parâmetros, estimadores e valores estimados	44
2.5 Propriedades de variáveis aleatórias	44
2.6 Estimadores	48
2.7 Formas funcionais logarítmicas	55
2.8 Exercícios	56

3 Séries temporais no contexto de regressão	59
3.1 Introdução	59
3.2 Exemplos de séries temporais	59
3.3 Regressão com dados de séries temporais	61
3.4 Suposições para modelos com séries temporais	62
3.5 Sazonalidade	65
3.6 Exercícios	68
4 Séries Temporais	71
4.1 Objetivos	71
4.2 Séries temporais: definição formal	71
4.3 Médias e covariâncias	72
4.4 Estacionariedade	73
4.5 Metodologia de Box-Jenkins ou modelagem ARIMA	81
4.6 Exercícios	94
5 Séries temporais não estacionárias	99
5.1 Como lidar com tendências determinísticas	99
5.2 Testes de raiz unitária	102
5.3 Eliminando tendência estocástica	104
5.4 Modelagem ARIMA	105
5.5 Previsão	106
5.6 Regressão espúria e cointegração	110
5.7 Exercícios	111
6 Variáveis instrumentais	121
6.1 Conceito da exogeneidade dos regressores	121
6.2 Variável instrumental	123
6.3 Inferência com o estimador por variável instrumental	125
6.4 Variáveis instrumentais em modelos de regressão múltipla	127
6.5 Testes de endogeneidade	129
6.6 Exercícios	131

Econometria e Séries Temporais com R

Marcio Valk
Guilherme Pumi



Figure 1: book cover

Prefácio

Este livro baseia-se nas notas de aula da Disciplina de Econometria que é oferecida pelo departamento de Estatística da UFRGS. Para cursar a disciplina é necessário ter alguns pré-requisitos, portanto é esperado que o leitor tenha algum tipo de conhecimento sobre estatística básica, inferência e regressão.

Chapter 1

Séries temporais

Dados de séries temporais são observações de um evento ou fenômeno ao longo do tempo. Os intervalos de observações devem ser igualmente espaçados. Geralmente são anos, trimestres, meses, semanas, dias, horas, minutos e segundos. Mas outros tipos de espaçamentos entre observações também são comuns. Como é o caso do Censo Demográfico.

No Figura 1.1 apresentamos alguns exemplos de séries temporais com diferentes intervalos de observações.

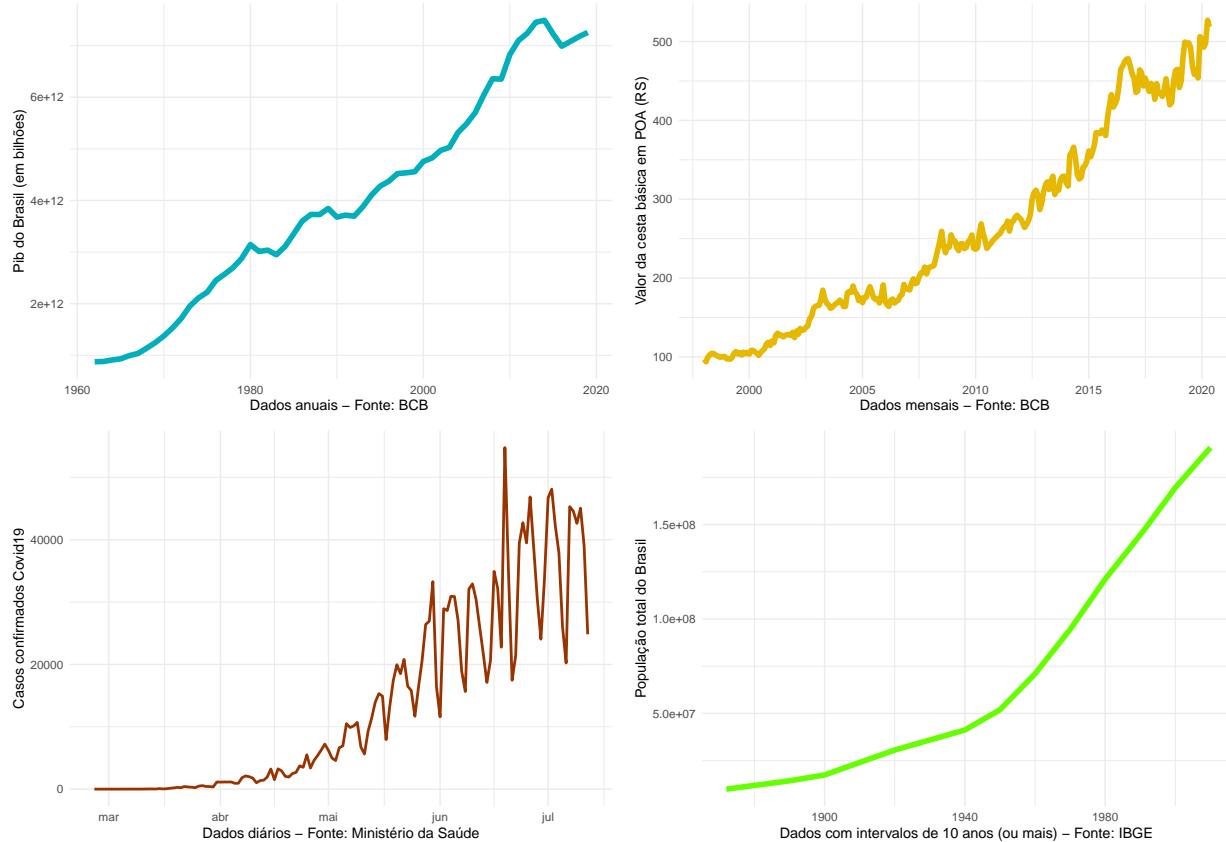


Figure 1.1: Séries temporais com diferentes intervalos de observações

Nesse exemplo temos séries temporais com diferentes intervalos de tempo. A série histórica do Pib tem intervalos de tempo de um ano. Os dados sobre valor da cesta básica em Porto alegre é divulgado mensalmente, os dados sobre Covid19 são divulgados diariamente e uma série temporal com observações a cada dez anos é a da população brasileira.

Séries temporais serão abordadas ao longo do texto, no qual exploraremos as propriedades teóricas das séries, características amostrais, sempre com um enfoque prático usando o R. Por esse motivo apresentamos uma breve introdução ao R que possibilitará ao leitor acompanhar e replicar os exemplos.

1.1 Breve introdução ao R

Das linguagens de programação voltadas a manipulação, vizualização e análises de dados, o R é uma das mais difundidas entre a comunidade Estatística. Outras linguagens, como o Python, tem um apelo maior quando se trata de ciência de dados, no seu sentido mais amplo.

1.2 Apresentação da linguagem R

R é uma linguagem de programação caracterizada como Software Livre sob os termos da *General Public License (GNU)* da *Free Software Foundation* no formato *open source*. É voltada a manipulação, análise e vizualização de dados e tem como característica o aspecto colaborativo, sendo que as ferramentas desenvolvidas são compartilhadas online pelos desenvolvedores, podendo ter acesso a elas qualquer pessoa, sem restrições. Uma breve história do R pode ser encontrada no wikipedia.

1.3 Instalando o R

Para instalar no computador, O R deve ser baixado do CRAN.

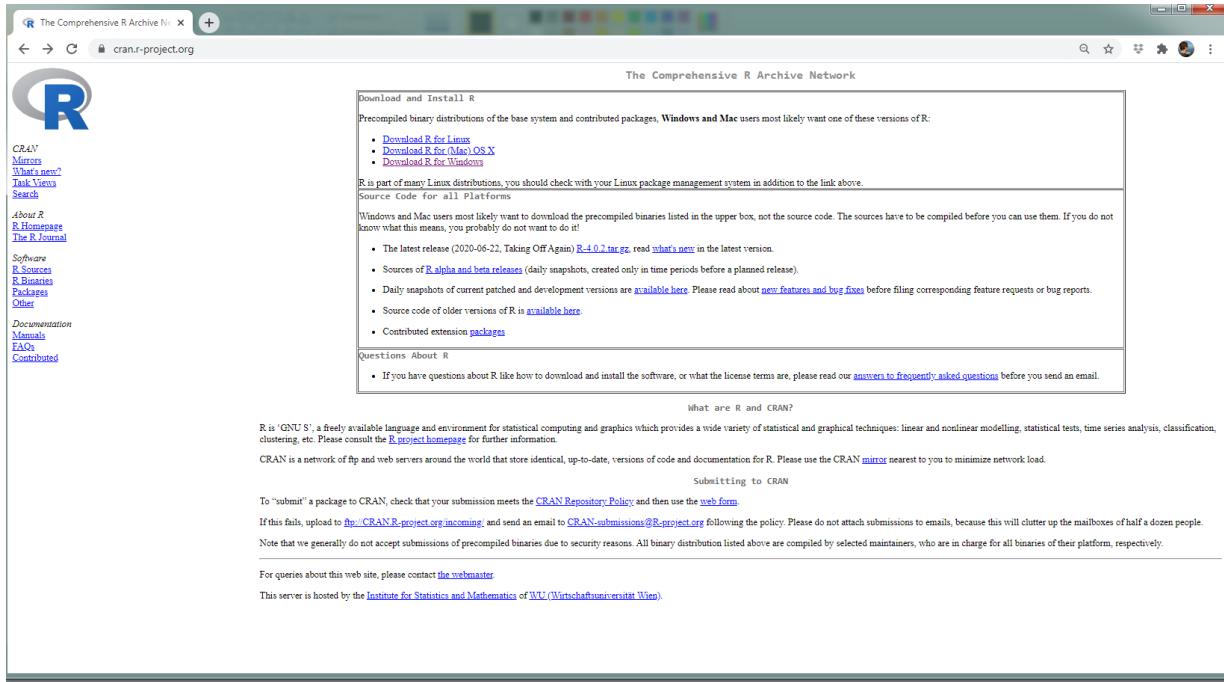


Figure 1.2: Comprehensive R archive network (CRAN)

Se o sistema operacional for Linux, uma versão *base* do R já vem instalada. No caso de outros sistemas operacionais, como o Windows, é necessário instalar o R base.

1.4 Instalando o RStudio

Como quase toda linguagem *Open Source* a utilização se dá por meio de linhas de comando. Para tornar a linguagem mais amigável aos usuários, várias IDEs (*integrated development environment*) são utilizadas. No caso do R, a mais desenvolvida e utilizada é o RStudio.

Uma versão *Free* do RStudio para o seu desktop pode ser baixado de <https://rstudio.com/products/rstudio/download/>.

Depois de instalar o R, o RStudio já estará integrado ao R e terá uma interface intuitiva e amigável ao usuário.

Ainda assim, é importante ressaltar que na linguagem R não encontraremos *botões* para realizar as análises.

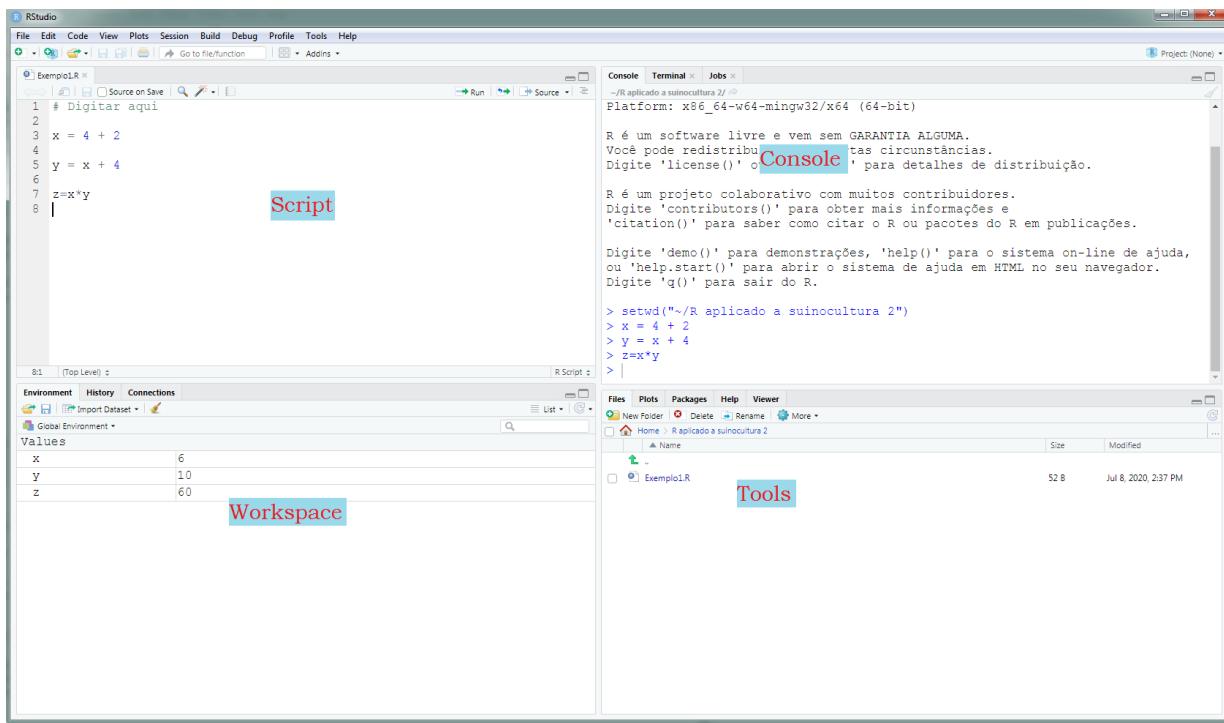


Figure 1.3: Interface do RStudio

1.5 Diretório de trabalho

Um ação importante que deve ser realizada pelo usuário é *setar o diretório* de trabalho. Para isso existem diferentes formas. Uma delas é usando a função *setwd()*.

Outra opção é através do *Go to directory* que está disponível no Workspace do RStudio, conforme figura ???. Nessa opção o usuário escolhe o diretório de trabalho e depois usando a opção *set as working directory* esse diretório será “*settado*” como diretório de trabalho.

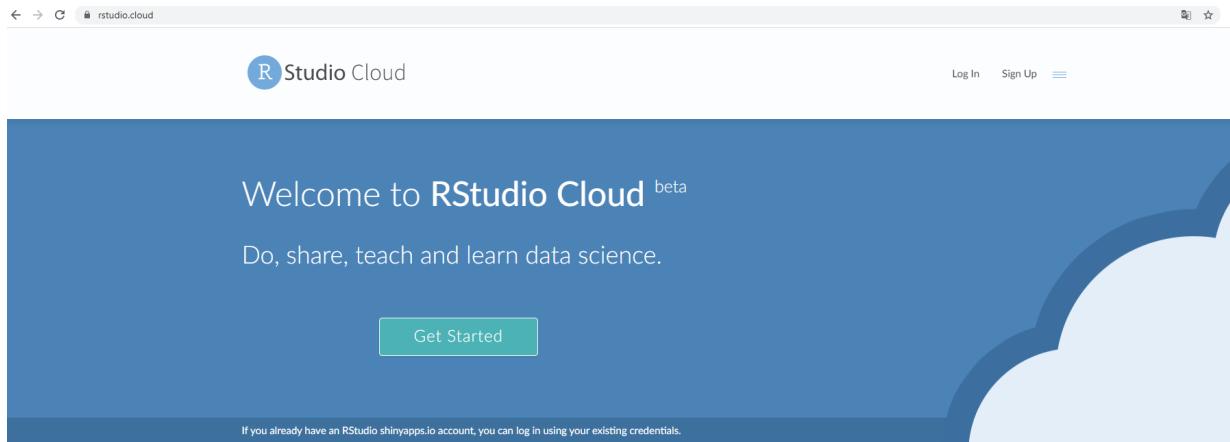
Todas os arquivos gerados, como figuras serão salvos nesse diretório. Para abrir um conjunto de dados, por exemplo, será muito mais simples se este também estiver salvo no mesmo diretório ou um subdiretório. Além disso, toda vez que o RStudio for reinicializado, esse procedimento terá que ser refeito.

1.6 RStudio Cloud

Outra forma simples e prática para usar o R e o RStudio é usar o RStudio Cloud.

THE MISSION We created RStudio Cloud to make it easy for professionals, hobbyists, trainers, teachers and students to do, share, teach and learn data science. (?)

Na nûvem é possível usar o RStudio utilizando um login através da conta google, ou criar gratuitamente uma conta. Uma vez logado, escolhendo a opção *project* o usuário terá uma versão do RStudio perfeitamente funcional, que pode ser utilizada até no smartphone. Obviamente, é necessário ter conexão com a internet para que a ferramenta possa ser utilizada.



THE MISSION

We created RStudio Cloud to make it easy for professionals, hobbyists, trainers, teachers and students to do, share, teach and learn data science.



Figure 1.4: Rstudio Cloud

Figure 1.5: Rstudio Cloud

1.7 Instalação de pacotes

Na versão *base* do R, uma série de ferramentas, funções e métodos estatísticos são disponibilizados. Além disso, alguns pacotes também compõe a versão *base* do R. Depois de instalado, o usuário pode verificar quais pacotes estão instalados acessando o ícone *Packages* em *Tools* ou digitando *installed.packages()* no *console*.

Para instalar um novo pacote, o usuário pode acessar o ícone *Packages* e depois clicar em *Install* ou digitar no console *install.packages("nome do pacote")*. Como exemplo, podemos instalar o pacote usado para vizualização de dados chamado *ggplot2*.

```
install.packages("ggplot2")
```

1.7.1 Carregando pacotes

Importante para usuários iniciantes na linguagem R é entender a diferença entre instalar pacotes e carregar pacotes. Uma vez instalado, o pacote estará à disposição do usuário sempre que ele precisar, mas é necessário carregá-lo. É comum deixar um comando nos *scripts* para que cada vez que seja necessário usar alguma função específica de um pacote, ele primeiro seja carregado. O comando usado é o *library()* mas pode ser feito acessando o espaço que chamando de **Tools**, clicar em *Packages* e marcar o pacote desejado para carregá-lo.

```
library(ggplot2)
```

1.8 Ajuda

Para um usuário iniciante no R é fundamental saber como resolver problemas diversos que certamente vão surgir durante a instalação de um pacote, uso de uma função, criação de um gráfico, manipulação de dados, etc. Muitas coisas no R são feitas por tentativa e erro, mas o conhecimento é acumulativo e problemas similares, poderão ter uma solução mais rápida. Uma das grandes vantagens da utilização do R é que a comunidade é bastante ativa. Existem diferentes formas de conseguir ajuda e vou elencá-las em ordem de importância, segundo a forma utilização desse autor:

- Google
- Stack Overflow
- Help do R (`help()` ou `?`)

1.8.1 Google

Para um iniciante em R coisas simples como calcular a raíz quadrada de um número pode ser difícil. Nesses casos o google é muito útil.

Um pouco de conhecimento de inglês aumenta consideravelmente as opções de ajuda no google.

1.8.2 Stack Overflow

Quando o problema parecer um pouco mais complexo, uma opção é colocar a pergunta no google junto com *Stack Overflow*. Isso provavelmente direcionará o usuário para Stack Overflow em Português ou Stack Overflow que são sites de Pergunta e Resposta utilizados por todas as linguagens de programação.

1.8.3 Help do R

Se o usuário já sabe qual função deve ser usada, então a documentação do R é bem útil.

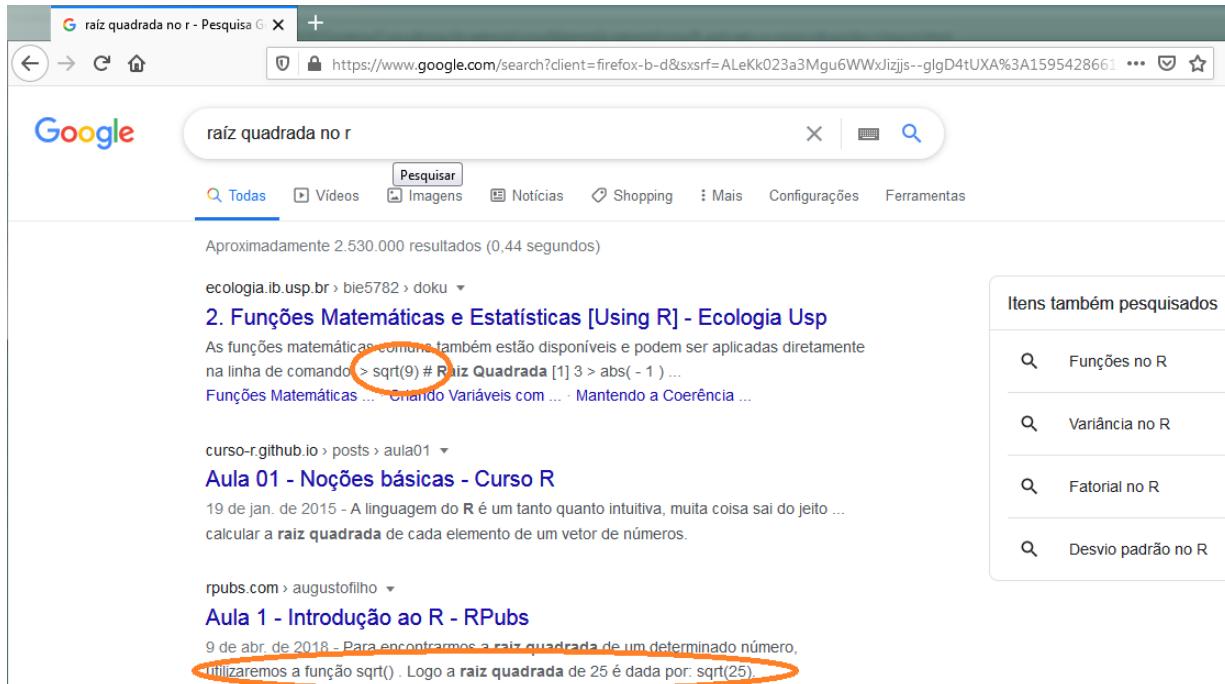


Figure 1.6: Rstudio Cloud

```
> sqrt("2")
Error in sqrt("2") : non-numeric argument to mathematical function
> |
```

Figure 1.7: Rstudio Cloud

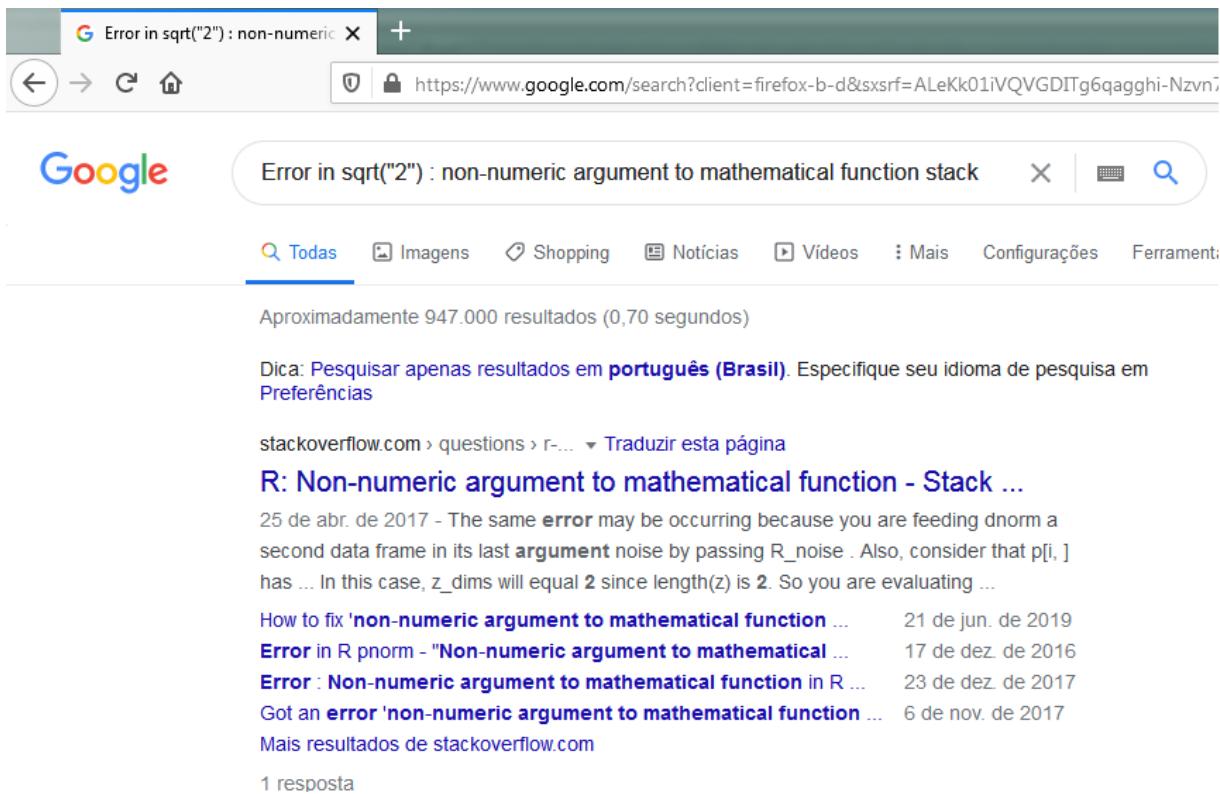


Figure 1.8: Rstudio Cloud

```
?sqrt
help(sqrt)
```

Dicas para uso do Help

- Pode-se ir direto aos exemplos que estão no final;
- Identificar os parâmetros (*Arguments*);
- Funções relacionadas podem ajudar, dependendo da necessida.
- *Vignettes*, que são tutorias mais completas, mas somente alguns pacotes possuem. Esses textos podem ser acessados com a função vignette(package = 'nomeDoPacote'). Por exemplo, vignette(package = 'ggplot2')

1.9 Funcionalidades básicas

Para maior agilidade é importante que o usuário conheça algumas teclas de atalho. Apertando simultaneamente **Alt + Shift + K** o usuário tem acesso à uma grande quantidade de atalhos. Ou clicado em *Tools* no menu de ícones do RStudio. Um comando de atalho que destacaria por ser extremamente útil é o **Ctrl + Enter**. Esse comando executa a linha do *script* em que o cursor está.

Para começar, usaremos o R como uma calculadora simples. Execute o código a seguir diretamente do console do RStudio ou no RStudio, escrevendo-os em um script e executando-os usando **Ctrl + Enter**.

1.9.1 Operações básicas

1.9.1.1 As quatro operações

Operação	código R	Resultado
$3 + 2$	<code>3 + 2</code>	5
$3 - 2$	<code>3 - 2</code>	1
$3 \cdot 2$	<code>3 * 2</code>	6
$3/2$	<code>3 / 2</code>	1.5

```
x <- 10
y <- 2

x+y
```

```
## [1] 12
```

```
x-y
```

```
## [1] 8
```

```
x*y
```

```
## [1] 20
```

```
x/y
```

```
## [1] 5
```

1.9.1.2 Exponenciação

Operação	código R	Resultado
3^2	<code>3 ^ 2</code>	9
$2^{(-3)}$	<code>2 ^ (-3)</code>	0.125
$100^{1/2}$	<code>100 ^ (1 / 2)</code>	10
$\sqrt{100}$	<code>sqrt(100)</code>	10

```
x^y      # x=10 e y=2
```

```
## [1] 100
```

1.9.1.3 Logarítmicos

Observe que não existe `ln()` no R. Usa-se `log()` para significar logaritmo natural. Para as demais bases, é necessário especificar a base desejada.

Operação	código R	Resultado
$\log(e)$	<code>log(exp(1))</code>	1
$\log_{10}(100)$	<code>log10(100)</code>	2
$\log_2(16)$	<code>log2(16)</code>	4
$\log_4(16)$	<code>log(16, base = 4)</code>	2

```
log(x)      # x=10
```

```
## [1] 2.302585
```

```
log(x,base=y) # y=2
```

```
## [1] 3.321928
```

1.9.1.4 Constantes matemáticas

Constante	código R	Resultado
π	pi	3.1415927
e	exp(1)	2.7182818

```
log(exp(1))
```

```
## [1] 1
```

```
exp(1)^y      # y=2
```

```
## [1] 7.389056
```

1.9.2 Operadores lógicos

Operador	Significado	Exemplo	Resultado
$x < y$	x menor do que y	$3 < 42$	TRUE
$x > y$	x maior do que y	$3 > 42$	FALSE
$x \leq y$	x menor ou igual à y	$3 \leq 42$	TRUE
$x \geq y$	x menor ou igual à y	$3 \geq 42$	FALSE
$x == y$	x igual à y	$3 == 42$	FALSE
$x != y$	x diferente de y	$3 != 42$	TRUE
$!x$	não x	$!(3 > 42)$	TRUE
$x y$	x ou y	$(3 > 42) TRUE$	TRUE
$x & y$	x e y	$(3 < 4) & (42 > 13)$	TRUE

1.9.2.1 Operador lógico *if()*

Dentro dessa classe de operadores, podemos destacar o operador *if()*. É comum usar esse operador para testar condições únicas ou múltiplas na instrução *if()* ou *ifelse()*. Operadores lógicos em R podem ser aplicados a vetores numéricos ou complexos ou objetos booleanos, que são **TRUE** ou **FALSE** (à eles são reservados os atalhos **T** e **F**).

```
A=4
B=2
if(A>B){
  print("A é maior do que B")
} else{
  print("A não é maior do que B")
}
```

```
## [1] "A é maior do que B"

A=4
B=6
if(A>B){
  print("A é maior do que B")
}else{
  print("A não é maior do que B")
}
```

```
## [1] "A não é maior do que B"
```

Outra forma simples de usar o operador é:

```
A=4
B=6
A>B
```

```
## [1] FALSE
```

1.9.2.2 Operador lógico *ifelse()*

O operador *ifelse(“1”, “2” , “3”)* possui 3 entradas. Na primeira “1”, deve-se colocar a condição a ser testada. Em “2” o resultado caso a condição testada seja verdadeira e em “3” o resultado, caso a condição testada seja falsa.

```
A=4
B=6
# ifelse(A>=B, "Verdadeiro", "FALSO")
```

1.9.2.3 Operadores de atribuição

Os operadores de atribuição são provavelmente a família de operadores que você mais usará enquanto trabalha com R. Como o nome desse grupo implica, eles são usados para atribuir objetos, como valores numéricos, *strings*, vetores, modelos e plotagens para um nome (variável). Isso inclui operadores como a seta para trás (<-) ou o sinal de igual (=)

```
str <- "Em Brasília, 19 horas!" # String
int <- 10 # Inteiro
vet <- c(1,2,3,4) # Vetor
```

Para visualizar a variável, podemos simplesmente digitá-la ou usar a função *print()*

```
str
```

```
## [1] "Em Brasília, 19 horas!"
```

```
#ou
print(str)
```

```
## [1] "Em Brasília, 19 horas!"
```

Para visualizar mais de uma variável, podemos usar “c()” para “juntar as variáveis”

```
c(str,int)

## [1] "Em Brasília, 19 horas!" "10"

# ou
print(c(str,int))
```

```
## [1] "Em Brasília, 19 horas!" "10"
```

Esse exemplo pode ser repetido usando “=” no lugar de “<-”

1.9.3 Tipos de dados

R possui um número básico de *tipos* de dados. Enquanto o R é uma *linguagem fortemente tipada* (não exige do usuário muito conhecimento sobre diferentes tipos de dados) é útil conhecer os tipos disponíveis.

- Numeric
 - Também conhecido como duplo. O tipo padrão ao lidar com números.
 - Exemplos: 1,1,0, 42,5
- Integer
 - Exemplos: 1L,2L, 42L
- Complex
 - Exemplo: 4 + 2i
- Logical
 - Dois valores possíveis: TRUE e FALSE
 - Você também pode usar T e F, mas isso *não* é recomendado.
 - NA também é considerado lógico.
- Character
 - Exemplos: " a ", " Statistics ", " 1 mais 2. "
- Categorical or factor
 - Uma mistura de número inteiro e caractere. Uma variável **fator** atribui um rótulo a um valor numérico.
 - Por exemplo, **fator** (x = c (0,1), labels = c (" male "," female ")) atribui a string *male* aos valores numéricos 0 e a string *female* ao valor 1.

1.9.4 Valores especiais

Assim como TRUE e FALSE, existem outros *valores* reservados à situações específicas.

- NA (*Not Available*) representa dado faltante (não disponível), ou que é chamado em estatística de *missing data*.
- NaN (*Not a Number*) são gerados quando temos uma indefinições matemáticas, como sqrt(-1) e 0/0.
- Inf (*Infinito*) é usado quando o valor numérico é muito grande (limite). Por exemplo, exp(2000).
- NULL significa ausência de informação. Parecido com o NA, mas conceitualmente mais usado na lógica de programação.

As funções **is.na()**, **is.nan()**, **is.infinite()** e **is.null()** podem ser usadas para verificar se um objeto possui algum valor com essa característica.

1.10 Estrutura dos dados

R também possui um número básico de *estrutura* de dados. Essa estrutura de dados pode ser **homogênea** (setodos os elementos são do mesmo tipo de dados) ou **heterogênea** (se os elementos podem ter mais de um tipo de dados).

Dimensão	Homogênea	Heterogênea
1	Vector	List
2	Matrix	Data Frame
3+	Array	nested Lists

1.10.1 Vetores

Muitas operações em R usam **vetores**. Um vetor contém um conjunto de objetos de tipos identicos e são indexados começando na posição 1. A maneira mais comum para se criar um vetor é usar a função `c()`, o que é uma abreviação de `combine()`. Ela combina uma lista de elementos separados por `,`. Por exemplo,

```
c(1, 2)
```

```
## [1] 1 2
```

Ou podemos combinar outros objetos, desde que sejam do mesmo tipo.

```
A=c(1,2)
B=c(3,4,5)
c(A,B)
```

```
## [1] 1 2 3 4 5
```

Os objetos não precisam ser numéricos. Por exemplo,

```
A="Amarelo" # objeto da classe "character"
class(A)
```

```
## [1] "character"
```

```
B="Azul"
c(A,B)
```

```
## [1] "Amarelo" "Azul"
```

Embora tenham muitas **letras** nesse objeto combinado `c(A,B)`, na verdade só existem dois elementos, {"Amarelo", "Azul"}. Para acessá-los, usamos o colchete `[]`.

1.10.1.1 Subconjuntos de vetores

Para subconjunto de um vetor, ou seja, para escolher apenas alguns elementos dele, usamos colchetes, `[]`. Aqui vemos que `AA[1]` retorna o primeiro elemento e `AA[2]` retorna o segundo elemento:

```
A="Amarelo"
B="Azul"
AA=c(A,B)
AA[1]
```

```
## [1] "Amarelo"
```

```
AA[2]
```

```
## [1] "Azul"
```

Quando o vetor é maior, podemos usar um conjunto de índices para acessar os valores correspondentes no vetor.

```
x=c(35,42,47,54,70,75)
x[c(1,3,6)]      # Acessa as posições 1, 3 e 6 do vetor
```

```
## [1] 35 47 75
```

```
x[-4]           # Acessa todas as posições do vetor, menos a posição 4
```

```
## [1] 35 42 47 70 75
```

Muitas vezes queremos criar um vetor baseado em uma sequência de números. No R o operador : é usado como uma opção de : até. Dessa forma é possível usá-lo em diferentes situações.

```
c(1:10) # vetor com números de 1 à 10
```

```
## [1] 1 2 3 4 5 6 7 8 9 10
```

```
1:10      # vetor com números de 1 à 10
```

```
## [1] 1 2 3 4 5 6 7 8 9 10
```

Outros vetores de objetos são possíveis de criar

```
LETTERS[1:4] # maiúsculas
```

```
## [1] "A" "B" "C" "D"
```

```
letters[1:4] # minuscúlas
```

```
## [1] "a" "b" "c" "d"
```

1.10.1.2 Função paste()

Com a função `paste()` podemos concatenar objetos de tipos diferentes e armazená-los em um vetor.

```
paste("IBGE_", letters[1:4], sep="")    # Underline no IBGE e sem espaço no sep
```

```
## [1] "IBGE_a" "IBGE_b" "IBGE_c" "IBGE_d"
```

```
paste("IBGE", letters[1:4], sep="_")     # Underline no sep
```

```
## [1] "IBGE_a" "IBGE_b" "IBGE_c" "IBGE_d"
```

```
p=paste("IBGE", letters[1:4], sep="_") # Armazenando no vetor p
p[2]                                     # Acessando a posição 2 do vetor p

## [1] "IBGE_b"
```

Note que escalares não existem no R. Eles são vetores de tamanho 1.

```
2
```

```
## [1] 2
```

1.10.1.3 Função seq()

Se quisermos criar uma sequência que não se limite a números inteiros e que aumente 1 por vez, podemos usar a função `seq()`.

```
seq(from = -0.5, to = 1.8, by = 0.1)
```

```
## [1] -0.5 -0.4 -0.3 -0.2 -0.1  0.0  0.1  0.2  0.3  0.4  0.5  0.6  0.7  0.8  0.9
## [16] 1.0  1.1  1.2  1.3  1.4  1.5  1.6  1.7  1.8
```

1.10.1.4 Função rep()

Outra operação comum para criar um vetor é `rep()`, que pode repetir um único valor várias vezes.

```
rep("A", times = 10)
```

```
## [1] "A" "A" "A" "A" "A" "A" "A" "A" "A" "A"
```

```
rep(1, 10)
```

```
## [1] 1 1 1 1 1 1 1 1 1 1
```

A função `rep()` pode ser usada para repetir um vetor várias vezes.

```
x=c("0","A")
rep(x, 3)
```

```
## [1] "0" "A" "0" "A" "0" "A"
```

1.10.1.5 Função length()

Uma função importante é à que identifica o tamanho do vetor, que é a função `length()`.

```
x=1:5
length(x)
```

```
## [1] 5
```

```
y=rep(x,6)
length(y)

## [1] 30
```

1.10.1.6 Resumo

temos quatro formas de criar vetores

- `c()`
- `:`
- `seq()`
- `rep()`

1.10.1.7 Operações com vetores

O R é capaz de executar muitas operações em vetores e escalares.

```
x = 1:10 # Um vetor
x + 1     # Soma um escalar à cada elemento do vetor

## [1] 2 3 4 5 6 7 8 9 10 11

2 * x      # Multiplica todos os elementos por 2

## [1] 2 4 6 8 10 12 14 16 18 20

2 ^ x      # Eleva 2 na potência correspondente a cada elemento de x

## [1] 2 4 8 16 32 64 128 256 512 1024

sqrt(x)    # Calcula a raíz quadrada de cada elemento de x

## [1] 1.000000 1.414214 1.732051 2.000000 2.236068 2.449490 2.645751 2.828427
## [9] 3.000000 3.162278

log(x)     # Calcula o log natural de cada elemento de x

## [1] 0.0000000 0.6931472 1.0986123 1.3862944 1.6094379 1.7917595 1.9459101
## [8] 2.0794415 2.1972246 2.3025851
```

We see that when a function like `log()` is called on a vector `x`, a vector is returned which has applied the function to each element of the vector `x`.

1.10.1.8 Operadores lógicos com vetores

Em R, operadores lógicos também funcionam com vetores:

```
x = c(5,3,1,9,27,90)

x == 9

## [1] FALSE FALSE FALSE  TRUE FALSE FALSE

x != 9

## [1]  TRUE  TRUE  TRUE FALSE  TRUE  TRUE

x > 9

## [1] FALSE FALSE FALSE FALSE  TRUE  TRUE

x < 9

## [1]  TRUE  TRUE  TRUE FALSE FALSE FALSE

x == 9 & x != 9

## [1] FALSE FALSE FALSE FALSE FALSE FALSE

x == 9 | x != 9

## [1]  TRUE  TRUE  TRUE TRUE  TRUE  TRUE
```

Outra operação importante que podemos destacar

```
x[x > 9]

## [1] 27 90

x[x != 9]

## [1]  5  3  1 27 90

sum(x > 9)

## [1] 2

as.numeric(x > 9)

## [1] 0 0 0 0 1 1
```

Usamos a função `sum()` em um vetor de valores lógicos TRUE e FALSE (resultado de `x>3`) e resultado foi um valor numérico. A operação apenas *contou* quantos vezes `x>3` resultou em TRUE. Durante a chamada de `sum()`, o R primeiro *coerce (força)* automaticamente o lógico para numérico, em que TRUE é 1 e FALSE é 0. Essa coerção do lógico para o numérico acontece na maioria das operações matemáticas.

1.10.1.9 Função `which()`

```
# which (dondição de x) retorna verdadeiro / falso
# cada índice de x em que a condição é verdadeira
x = c(5,3,1,9,27,90)
which(x > 9)
```

```
## [1] 5 6
```

```
x[which(x > 9)]
```

```
## [1] 27 90
```

```
max(x)
```

```
## [1] 90
```

```
which(x == max(x))
```

```
## [1] 6
```

```
which.max(x)
```

```
## [1] 6
```

1.10.2 Tarefa 2

1. Crie um vetor preenchido com 10 números sorteados na distribuição uniforme discreta em $\{1,2,3,4,5,6\}$ (dica: use a função `sample()`) e armazene-os em `x`.
2. Usando o subconjunto lógico como acima, obtenha todos os elementos de `x` maiores que 2 e armazene-os em `y`.
3. Usando a função `which`, armazene os *índices* de todos os elementos de `x` que são maiores que 2 em `iy`.
4. Verifique se `y` e `x[iy]` são idênticos.

1.10.3 Matrizes

O R também pode ser usado para cálculos de **matriz**. Matrizes têm linhas e colunas contendo um único tipo de dados. Em uma matriz, a ordem das linhas e colunas é importante. (Isso não se aplica à *dataframe*, que é um outro tipo de dado que veremos mais adiante).

Matrizes podem ser criadas usando a função `matrix`.

```
x = 12:1
x
```

```
## [1] 12 11 10  9  8  7  6  5  4  3  2  1
```

```
X = matrix(x, nrow = 3, ncol = 4)
X
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    12    9    6    3
## [2,]    11    8    5    2
## [3,]    10    7    4    1
```

Note que o R é **case sensitive** (x vs X).

Por padrão, a função `matrix` preenche seus dados na matriz coluna por coluna. Mas também podemos dizer ao R para preencher as linhas:

```
W = matrix(x, nrow = 3, ncol = 4, byrow = TRUE)
W
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    12   11   10    9
## [2,]     8    7    6    5
## [3,]     4    3    2    1
```

Também podemos criar uma matriz de uma dimensão especificada onde cada elemento é o mesmo, neste caso, 0.

```
Y = matrix(0, 2, 3)
Y
```

```
##      [,1] [,2] [,3]
## [1,]     0    0    0
## [2,]     0    0    0
```

1.10.3.1 Matriz diagonal

Para criar uma matriz diagonal podemos usar a função `diag()`

```
diag(4) # cria matriz identidade 4x4
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    1    0    0    0
## [2,]    0    1    0    0
## [3,]    0    0    1    0
## [4,]    0    0    0    1
```

```
diag(4,5) # cria uma matriz digonal 4x4, em que os elementos da diagonal são 5
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    4    0    0    0    0
## [2,]    0    4    0    0    0
## [3,]    0    0    4    0    0
## [4,]    0    0    0    4    0
## [5,]    0    0    0    0    4
```

Como vetores, matrizes podem ser acessadas usando colchetes, `[]`. No entanto, como as matrizes são bidimensionais, precisamos especificar uma linha e uma coluna ao fazer o subconjunto.

```
X
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    12   9    6    3
## [2,]    11   8    5    2
## [3,]    10   7    4    1
```

```
X[1, 2] # primeira linha e na segunda coluna
```

```
## [1] 9
```

Também podemos acessar uma linha ou coluna inteira.

```
x[1, ]
```

```
## [1] 12 9 6 3
```

```
x[, 2]
```

```
## [1] 9 8 7
```

1.10.3.2 Elementos da matriz

Também podemos usar vetores para acessar subconjunto com mais de uma linha ou coluna por vez. Aqui acessamos à primeira e terceira coluna da segunda linha:

```
X[2, c(1, 3)] # segunda linha, primeira e terceira coluna
```

```
## [1] 11 5
```

```
X[c(2,1), c(1, 3)] # segunda e priemira linha, primeira e terceira coluna
```

```
##      [,1] [,2]
## [1,]    11    5
## [2,]    12    6
```

1.10.3.3 Funções rbind() e cbind

As matrizes também podem ser criadas combinando vetores como colunas, usando `cbind`, ou combinando vetores como linhas, usando `rbind`.

```
x = 1:4
rev(x)
```

```
## [1] 4 3 2 1
```

```
rep(1,4)
```

```
## [1] 1 1 1 1
```

```
rbind(x, rev(x), rep(1, 4))
```

```
##      [,1] [,2] [,3] [,4]
## x     1     2     3     4
##        4     3     2     1
##        1     1     1     1
```

```
cbind(col_1 = x, col_2 = rev(x), col_3 = rep(1, 4))

##      col_1 col_2 col_3
## [1,]     1     4     1
## [2,]     2     3     1
## [3,]     3     2     1
## [4,]     4     1     1
```

Ao usar `rbind` ou `ecbind`, você pode especificar nomes de “argumentos” que serão usados como nomes de colunas.

1.10.3.4 Operações com Matrizes

O R pode então ser usado para realizar cálculos de matriz.

```
x = 1:12
y = 12:1
X = matrix(x, 3, 4)
Y = matrix(y, 3, 4)
X
```

```
##      [,1] [,2] [,3] [,4]
## [1,]     1     4     7    10
## [2,]     2     5     8    11
## [3,]     3     6     9    12
```

```
Y
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    12     9     6     3
## [2,]    11     8     5     2
## [3,]    10     7     4     1
```

```
X + Y # Soma elemento por elemento
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    13    13    13    13
## [2,]    13    13    13    13
## [3,]    13    13    13    13
```

```
X - Y # Subtração elemento por elemento
```

```
##      [,1] [,2] [,3] [,4]
## [1,]   -11    -5     1     7
## [2,]    -9    -3     3     9
## [3,]    -7    -1     5    11
```

```
X * Y # Multiplicação elemento por elemento
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    12    36    42    30
## [2,]    22    40    40    22
## [3,]    30    42    36    12
```

```
X / Y # Divisão elemento por elemento
```

```
## [,1] [,2] [,3] [,4]
## [1,] 0.08333333 0.44444444 1.1666667 3.333333
## [2,] 0.18181818 0.62500000 1.600000 5.500000
## [3,] 0.30000000 0.8571429 2.250000 12.000000
```

Note que `X * Y` **não** é multiplicação de matrizes. É multiplicação de *elemento por elemento*. (O mesmo para `X/Y`).

Para a multiplicação de matrizes usa-se `%*%`. Outras funções para operações com matrizes são `t()`, que fornece a transposição de uma matriz e `solve()`, que retorna a inversa de uma matriz quadrada, se for invertível.

```
x = 1:9
y = 9:1
X = matrix(x, 3, 3)
Y = matrix(y, 3, 3)
X %*% Y
```

```
## [,1] [,2] [,3]
## [1,] 90 54 18
## [2,] 114 69 24
## [3,] 138 84 30
```

```
t(X)
```

```
## [,1] [,2] [,3]
## [1,] 1 2 3
## [2,] 4 5 6
## [3,] 7 8 9
```

1.10.4 Arrays

Um vetor é um `array` unidimensional. Uma matriz é um `array` bidimensional. Em R, o usuário pode criar `arrays` de dimensionalidade arbitrária N. Por exemplo:

```
A = 1:16
B = array(data = A, dim = c(4,2,2))
C = array(data = A, dim = c(4,2,2,3)) # Recicla A=1:16
B
```

```
## , , 1
##
## [,1] [,2]
## [1,] 1 5
## [2,] 2 6
## [3,] 3 7
## [4,] 4 8
##
## , , 2
##
## [,1] [,2]
## [1,] 9 13
## [2,] 10 14
## [3,] 11 15
## [4,] 12 16
```

Note que B são simplesmente *duas* matrizes (4,2) armazenadas como *páginas* em um mesmo objeto/variável. Similarmente, C teria duas *páginas* e mais 3 registros em uma quarta dimensão. E assim por diante.

Para acessar esses elementos, o procedimento é similar ao de uma matriz ou um vetor, cuidando para indexar cada dimensão:

```
B[,1,1]      # Todos os elementos da col 1, página 1

## [1] 1 2 3 4

B[2:3, , ]  # Linhas 2:3 de todas as páginas e colunas

## , , 1
##
##      [,1] [,2]
## [1,]    2    6
## [2,]    3    7
##
## , , 2
##
##      [,1] [,2]
## [1,]   10   14
## [2,]   11   15

B[2,2, ]     # linha 2, coluna 2 de todas as páginas.
```

```
## [1] 6 14
```

1.10.4.1 Tarefa 3

1. Crie um vetor contendo 1,2,3,4,5 chamado-o de v .
2. Crie uma matriz (2,5) m contendo os dados 1,2,3,4,5,6,7,8,9,10. A primeira linha deve ser “1,2,3,4,5”.
3. Realize a multiplicação da matriz de m com v . Use o comando $\%*\%$. Qual a dimensão da saída?
4. Por que $v\%*\% m$ não funciona?

1.10.5 Listas

Uma lista é uma estrutura de dados unidimensional *heterogênea*. Portanto, é indexado como um vetor com um único valor inteiro (ou com um nome), mas cada elemento pode conter um elemento de qualquer tipo. As listas são semelhantes a um objeto `python` ou `julia`.

Muitas estruturas e saídas do R são listas. As listas são objetos extremamente úteis e versáteis; portanto, é interessante entender sua utilização:

```
# Listas simples sem nomes nas entradas
list(12, "Bom dia", TRUE)

## [[1]]
## [1] 12
##
## [[2]]
## [1] "Bom dia"
##
## [[3]]
## [1] TRUE
```

```
# Listas com nome para cada entrada
lista1 = list(
  a = c(1, 2, 3, 4),
  b = TRUE,
  c = "Bom dia!",
  d = function(arg = 10) {print("Em Brasília, 19h!")},
  e = diag(4)
)
```

Os elementos das listas podem ser acessados usando duas sintaxes, o operador `$` e colchetes `[]`. O operador `$` retorna um elemento **nomeado** de uma lista. A sintaxe `[]` retorna uma **lista**, enquanto a `[[]]` retorna um **elemento** de uma lista.

- `lista1[1]` retorna uma lista que contém o primeiro elemento
- `lista1[[1]]` retorna o primeiro elemento da lista, neste caso, um vetor.

```
# elementos da lista
lista1$e
```

```
##      [,1] [,2] [,3] [,4]
## [1,]     1     0     0     0
## [2,]     0     1     0     0
## [3,]     0     0     1     0
## [4,]     0     0     0     1
```

```
lista1[[1:2]]
```

```
## [1] 2
```

```
lista1[c("e", "a")]
```

```
## $e
##      [,1] [,2] [,3] [,4]
## [1,]     1     0     0     0
## [2,]     0     1     0     0
## [3,]     0     0     1     0
## [4,]     0     0     0     1
##
## $a
## [1] 1 2 3 4
```

```
lista1["e"]
```

```
## $e
##      [,1] [,2] [,3] [,4]
## [1,]     1     0     0     0
## [2,]     0     1     0     0
## [3,]     0     0     1     0
## [4,]     0     0     0     1
```

```
lista1[["e"]]
```

```

##      [,1] [,2] [,3] [,4]
## [1,]    1    0    0    0
## [2,]    0    1    0    0
## [3,]    0    0    1    0
## [4,]    0    0    0    1

lista1$d

## function(arg = 10) {print("Em Brasília, 19h!")}

lista1$d(arg = 1)

## [1] "Em Brasília, 19h!"

```

Tarefa 4

1. Copie e cole o código acima para `list1` na sua sessão do R. Lembre-se de que `list` pode conter qualquer tipo de objeto R. Como por exemplo, outra lista! Portanto, crie uma nova lista `list2` que tenha dois campos: o primeiro campo chamado "`Isso`" com o conteúdo da string "`é impressionante`" e um segundo campo chamado "`lista`" que contém `list1`.
2. Acessar os elementos é como em uma lista simples, apenas com várias camadas agora. Obtenha o elemento `c` de `list1` em `list2`.
3. Componha uma nova string com primeiro elemento em `list2`, o elemento sob o rótulo "`Isso`". Use a função `paste` para imprimir R `é impressionante` na tela.

1.11 Programação básica

Nesta seção, ilustramos alguns conceitos gerais relacionados à programação.

1.11.1 Variáveis

Na programação, uma variável denota um *objeto*. Outra maneira de dizer isso é que uma variável é um nome ou um *label* para algo:

```

x = 2
y = "doce"
z = function(x){log(x)}

```

Aqui, `x` se refere ao valor 2, `y` refere-se a *string* “doce” e `z` é o nome de uma função que calcula $\log x$. Observe que o argumento `x` da função é diferente do `x` que acabamos de definir.

```

x

## [1] 2

z(9)

## [1] 2.197225

z(x)

## [1] 0.6931472

```

1.11.2 Laços (*Loops*), `for` e `while`

Laços (*Loops*) são uma construção de programação muito importante. Como o nome sugere, em um *loop*, a programação *repetidamente* circula um conjunto de instruções, até que alguma condição diga para parar. Uma construção muito poderosa, porém simples, é que o programa pode *contar quantas etapas* já executou - o que pode ser importante saber para muitos algoritmos. Uma forma de fazer um *loop* é o usando o `for`.

1.11.2.1 Uso do `for()`

Por exemplo, para executar uma tarefa 3 vezes, usamos

```
for (i in 1:3){
  print(i)
}
```

```
## [1] 1
## [1] 2
## [1] 3
```

A iteração pode ser feita sobre um vetor qualquer de objetos

```
for (i in c("IPCA","IGP-M","INPC")){
  print(paste("Índice ",i))  # a função paste() conecta as strings
}
```

```
## [1] "Índice IPCA"
## [1] "Índice IGP-M"
## [1] "Índice INPC"
```

Podemos fazer *loops* dentro de *loops* (*nested loops*)

```
for (i in 2:3){
  # primeiro: para cada i
  for (j in c("IPCA","IGP-M","INPC")){
    # segundo: para cada j
    print(paste("A cada",i,"meses pode-se calcular o Índice",j))
  }
}

## [1] "A cada 2 meses pode-se calcular o Índice IPCA"
## [1] "A cada 2 meses pode-se calcular o Índice IGP-M"
## [1] "A cada 2 meses pode-se calcular o Índice INPC"
## [1] "A cada 3 meses pode-se calcular o Índice IPCA"
## [1] "A cada 3 meses pode-se calcular o Índice IGP-M"
## [1] "A cada 3 meses pode-se calcular o Índice INPC"
```

1.11.2.2 Uso do `while()`

A função `while()`, como o próprio nome sugre, é usada para executar uma tarefa até que uma certa condição aconteça, ou enquanto certa condição não aconteça.

```
dias=27
while(dias<30){
  print(paste("Passaram-se",dias,"dias")) # Enquanto dias é menor do que 30, executa
  dias=dias+1 # Enquanto dias é menor do que 30, soma-se 1 ao dia atual
}
```

```
## [1] "Passaram-se 27 dias"
## [1] "Passaram-se 28 dias"
## [1] "Passaram-se 29 dias"
```

1.12 Funções

Até agora, usamos funções, mas na verdade não discutimos alguns de seus detalhes. Uma função é um conjunto de instruções que o R executa para nós, como as coletadas em um arquivo *script*. O bom é que as funções são muito mais flexíveis que os scripts, pois podem depender de *argumentos de entrada*, que alteram a maneira como a função se comporta. Aqui está como definir uma função:

```
nome_da_funcao <- function(arg1,arg2=valor_padrao){
  # corpo da função
  # faça coisas com arg1 e arg2
  # a última linha será retornada
}
```

Um exemplo de uma função simples

```
fun1<- function(seu_nome = "Baby"){
  paste("Seja bem vindo ao R,",seu_nome)
  # pode-se escrever também
  # return(paste("Seja bem vindo ao R,",seu_nome))
}
#chamando a função sem nenhum argumento, o argumento seu_nome padrão será Baby
fun1()
```

[1] "Seja bem vindo ao R, Baby"

Agora tente com o seu nome:

```
fun1("Pato Donald")
```

[1] "Seja bem vindo ao R, Pato Donald"

Para ver a “programação” da função, basta digitar o nome da função:

```
fun1

## function(seu_nome = "Baby"){
##   paste("Seja bem vindo ao R,",seu_nome)
##   # pode-se escrever também
##   # return(paste("Seja bem vindo ao R,",seu_nome))
## }
## <bytecode: 0x00000000c0df3c0>
```

Funções são maneiras de dizer ao R o que deve ser feito. É como se ensinássemos uma tarefa ao software para que ele passe a executá-la sempre que solicitado. Um dica para não acumular erros durante a programação é dividir o *script* em várias funções. Quando se tem um conjunto de funções que realiza um conjunto de tarefas mais amplas, pode-se juntar tudo em um pacote (*package*), submeter ao CRAN para disponibilizar à comunidade.

Tarefa 6

1. Escreva um *loop for* para escrever o seu nome, letra por letra. Comece com um vetor com as letras do seu nome.
2. Modifique esse *loop* para que cada iteração leve aproximadamente um segundo. Você pode conseguir isso adicionando o comando `Sys.sleep(1)` abaixo da linha que imprime “iteração i nome_parcial”, em que `i` é a correspondente iteração no *loop*.

1.13 Operador Pipe

Para manipulação de dados é interessante conhecer a lógica do operador `pipe`, `%>%`. Vários pacotes são construídos em cima dessa lógica, que busca encadear ações do R de uma forma mais compreensível. Vamos instalar o pacote `magrittr` para começar a usar `%>%`.

```
install.packages("magrittr")
library(magrittr)

## Warning: package 'magrittr' was built under R version 3.6.3

##
## Attaching package: 'magrittr'

## The following object is masked from 'package:tidyverse':
## 
##     extract
```

Vejamos como funciona.

```
sin(cos(pi))
pi %>% cos %>% sin
```

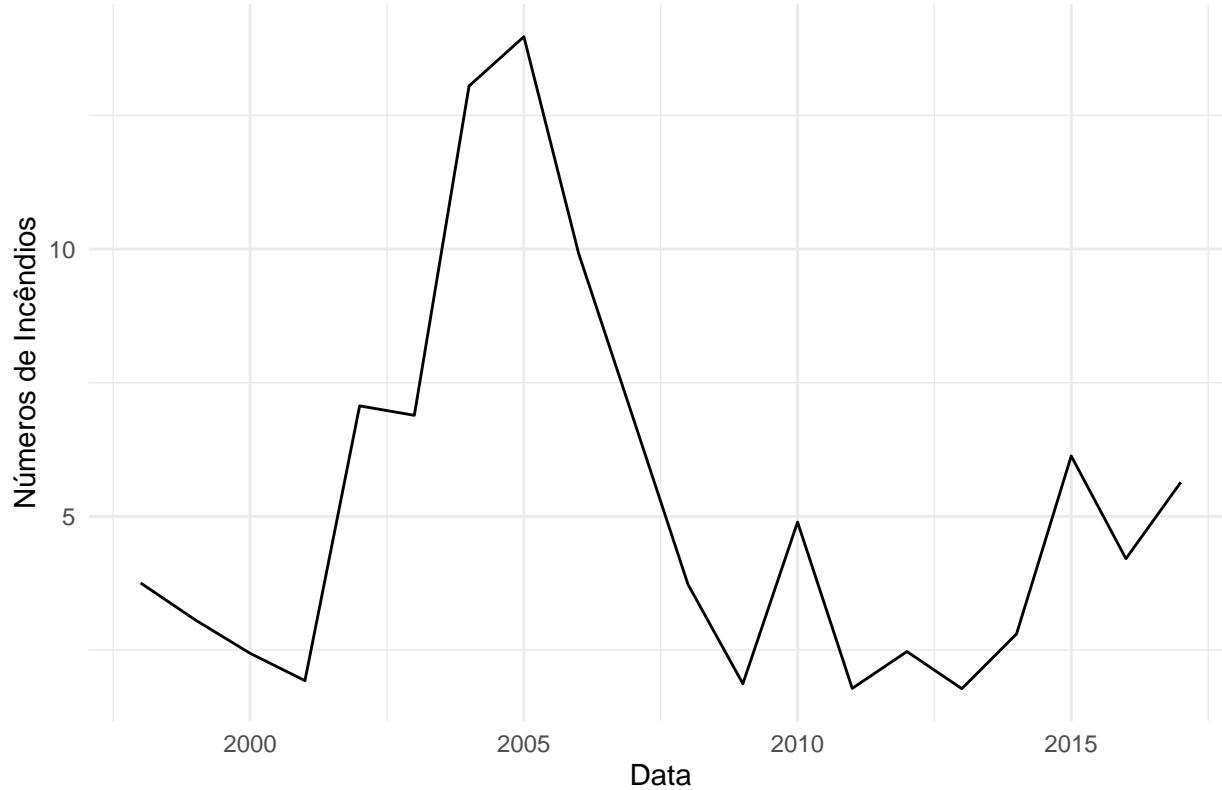
Ambas as expressões acima são equivalentes. Parece redundante e que não se tem nenhuma vantagem ao usar o `%>%`. No entanto, para um caso envolvendo dados reais, podemos verificar sua utilidade

```
library(ggplot2)
library(dplyr)
head(fogo)

##   year state month number      date
## 1 1998  Acre Janeiro       0 1998-01-01
## 2 1999  Acre Janeiro       0 1999-01-01
## 3 2000  Acre Janeiro       0 2000-01-01
## 4 2001  Acre Janeiro       0 2001-01-01
## 5 2002  Acre Janeiro       0 2002-01-01
## 6 2003  Acre Janeiro      10 2003-01-01

fogo$date=as.Date(fogo$date)
fogo %>% na.omit %>% dplyr::filter(month=="Setembro" & state=="Rondonia") %>% ggplot(aes(x = date, y = number)) +
  y = "Números de Incêndios",
  title = "Focos de incêndio em Rondônia nos meses de Setembro")
```

Focos de incêndio em Rondônia nos meses de Setembro



Chapter 2

Revisão

Esse capítulo é uma breve revisão sobre conceitos estatísticos que são fundamentais para entender os conceitos da modelagem econométrica.

2.1 Aleatoriedade, a essência da estatística

Para iniciar qualquer curso em que são utilizadas técnicas estatísticas, é necessário esclarecer/fundamentar bem o conceito de aleatoriedade.

Na história antiga, os conceitos de chance e de aleatoriedade eram interligados ao conceito que era atribuído a destino. Várias pessoas da antiguidade jogavam dados para determinarem o destino, e posteriormente isso se desenvolveu em jogos de azar. A maioria das culturas usaram vários métodos de adivinhações para tentarem contornar a aleatoriedade e o destino, ou mesmo a dita sorte. A palavra aleatoriedade é utilizada para exprimir quebra de ordem, propósito, causa, ou imprevisibilidade em uma terminologia não científica. Um processo aleatório é o processo repetitivo cujo resultado não descreve um padrão determinístico, mas segue uma distribuição de probabilidade Wikipédia.



Figure 2.1: Dados

As técnicas estatísticas surgem para encontrar algum *padrão de variação*. Para tal tarefa é necessário formalizar e definir alguns conceitos, como são os casos de variável aleatória e distribuição de probabilidade.

2.2 Variável aleatória

Denomina-se **variável** uma propriedade (característica) qualquer das unidades da população para a qual foi definida uma unidade de medida, que pode ser quantitativa ou qualitativa. Observe que essa característica é comum a todos os indivíduos e portanto é uma característica da população. Em geral, queremos fazer afirmações sobre características e temos apenas informações de alguns indivíduos (amostra). Assim, toda afirmação feita a partir de uma amostra é passível de erros, ou seja, é uma aproximação. Além disso, em alguns casos não é possível “medir” toda a população e

devemos pensar nessa característica como uma quantidade aleatória. Para isso, é necessário introduzirmos o conceito de **variável aleatória**.

Definição 2.1 (Espaço Amostral). Espaço amostral de um *experimento aleatório* (fenômeno que, mesmo repetidos várias vezes sob condições semelhantes, apresentam resultados imprevisíveis) é **qualquer** conjunto contendo todos os possíveis resultados do experimento. Aqui, sempre que não houver perigo de confusão, o espaço amostral de um experimento em questão será denotado por Ω .

Exemplo 2.1. No seguinte experimento: derrubar o celular e observar a face voltada para cima, o espaço amostral é o conjunto {Frontal, Costas}.

Exemplo 2.2. Se o experimento é lançar um dado de seis faces, o espaço amostral é {1, 2, 3, 4, 5, 6}.

Exemplo 2.3. Poderá perfeitamente existir mais de um espaço amostral adequado para um determinado experimento. No Exemplo 2.2, o conjunto {1, 2, 3, 4, 5, 6, 7} contém todos os possíveis resultados do experimento em questão (lançar um dado de seis faces). Assim, pela definição 2.1, este conjunto é tão adequado como espaço amostral quanto o conjunto mais intuitivo {1, 2, 3, 4, 5, 6}. Até mesmo o conjunto dos números reais \mathbb{R} é adequado. Obviamente, sempre que possível é recomendável utilizar o conjunto mais *natural* como espaço amostral, porém, do ponto de vista teórico, desde que o conjunto escolhido efetivamente contenha todos os possíveis resultados do experimento, não faz diferença alguma qual conjunto se está utilizando.

Exemplo 2.4. Nos exemplos anteriores, é possível (e muito fácil) determinar exatamente quais são todos os possíveis resultados dos experimentos em questão. Porém nem sempre este é o caso. Considere o experimento em que uma pessoa é escolhida ao acaso e sua altura (em metros) medida. Neste caso é difícil determinar precisamente o conjunto contendo exatamente todos os possíveis resultados do experimento. Com certeza o conjunto [0, 10] contém todas as possíveis alturas a serem registradas. O conjunto [0, 3] também. Por outro lado, será que o conjunto [0, 2.7] é apropriado? E (0.3, 2.7)?

Todo subconjunto de um espaço amostral é chamado *evento*. Os subconjuntos de um espaço amostral contendo apenas um elemento são chamados de *eventos elementares*.

Por exemplo, no lançamento de um dado de seis faces, {5} é um evento elementar. Outro evento possível é: *a face superior é ímpar*, o que é equivalente ao subconjunto {1, 3, 5} $\subset \Omega$. Outra possibilidade poderia ser verificar se a face obtida é superior a 3.

Existem ainda experimentos que podem ser vistos como “compostos” por natureza, como por exemplo o lançamento independente de um dado de seis faces e de uma moeda honesta, no qual anotamos a face superior do dado e a face da moeda. Neste caso, é fácil determinar um espaço amostral associado ao experimento que contenha exatamente todos os resultados possíveis. Este constituirá de pares contendo um número inteiro de 0 à 6, correspondente ao lançamento do dado e um elemento do conjunto {Frontal, Costas}, correspondente à queda do celular, ou seja, $\Omega = \{(1, \text{Frontal}), (1, \text{costas}), \dots, (6, \text{Frontal}), (6, \text{Costas})\}$. Uma outra maneira de representar isto é a partir do produto cartesiano dos espaços amostrais de cada um dos experimentos individuais, neste caso $\Omega = \{1, 2, 3, 4, 5, 6\} \times \{\text{Frontal, Costas}\}$.

Espaços amostrais são importantes na definição de um *espaço de probabilidade*. Um espaço de probabilidade $(\Omega, \mathcal{F}, \mathcal{P})$ onde Ω denota um espaço amostral qualquer, \mathcal{F} é um conjunto de eventos associado à Ω satisfazendo certas propriedades (σ -álgebra de eventos), e $\mathcal{P} : \mathcal{F} \rightarrow [0, 1]$ uma medida de probabilidade atribuindo valores em [0, 1] para cada evento de interesse em \mathcal{F} (a probabilidade dos eventos).

Uma variável aleatória é uma função do espaço amostral Ω nos reais, para a qual é possível calcular a probabilidade de ocorrência de seus valores. Em geral, as variáveis aleatórias são representadas por letras maiúsculas do fim do alfabeto. Temos, para cada elemento $\omega \in \Omega$, um número real $X(\omega)$ conforme a Figura 2.1.

Garantimos o cálculo de probabilidades com variáveis aleatórias ao exigir que, para qualquer $I \subset \mathbb{R}$, o conjunto $X^{-1}(I)$ seja um evento. Em outras palavras, o conjunto $X^{-1}(I)$ é um elemento de \mathcal{F} , ou seja, $X^{-1}(I) \in \mathcal{F}$. Lembremos que apenas os elementos de \mathcal{F} têm atribuição de probabilidade. Em linguagem mais matemática, dizemos que uma variável aleatória é qualquer função mensurável em (Ω, \mathcal{F}) . Isto justifica dizer que a variável X é \mathcal{F} -mensurável. Com frequência, faz-se menção ao espaço de probabilidade $(\Omega, \mathcal{F}, \mathcal{P})$, para deixar claro o espaço amostral, a σ -álgebra e a probabilidade envolvidas. Formalmente, definimos

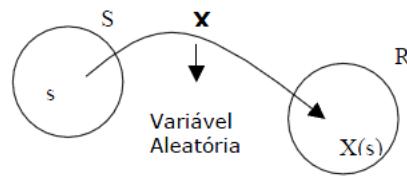


Figure 2.2: Variável aleatória

Definição 2.2. Seja $(\Omega, \mathcal{F}, \mathcal{P})$ um espaço de probabilidade. Denominamos de variável aleatória, qualquer função $X : \Omega \rightarrow \mathbb{R}$ tal que

$$X^{-1}(I) = \{\omega \in \Omega : X(\omega) \in I\} \in \mathcal{F},$$

para todo intervalo $I \subset \mathbb{R}$. Em palavras, X é tal que sua imagem inversa de intervalos $I \subset \mathbb{R}$ pertencem a σ -álgebra \mathcal{F} .

No que segue precisamos do conceito de cardinalidade de um conjunto. Em palavras simples, a cardinalidade de um conjunto é uma maneira de expressar a “quantidade” de elementos que este contém. Um conjunto ordenado A é dito *finito* se contém um número finito de elementos. A cardinalidade de um conjunto finito nada mais é que o número de elementos que este contém. Por exemplo o conjunto $A = \{1, 2, 9, 15\}$ é finito e tem cardinalidade 4.

Por outro lado, a definição de cardinalidade para conjuntos infinitos é matematicamente muito mais complexa pois, no final das contas, a ideia é impor uma hierarquia, uma “ordem”, no “tamanho” de conjuntos infinitos. Obviamente a cardinalidade de um conjunto infinito não pode ser expressa em números. Estamos interessados apenas em distinguir entre dois “tamanhos” de conjuntos infinitos: enumerável e não-enumerável. Por sorte, na maioria das vezes é possível utilizar apenas a intuição para resolver o problema. Intuitivamente, um conjunto ordenado A é dito ser infinito enumerável (ou ainda, *contável*) se dado um elemento qualquer de A , podemos determinar quem é o próximo elemento do conjunto. Caso contrário, o conjunto é dito ser *não-enumerável*. Por exemplo, o conjunto dos números naturais \mathbb{N} é infinito enumerável. De fato, dado qualquer número natural x , o próximo é $x + 1$, obviamente. Já o conjunto $[0, 1]$ é infinito não-enumerável. Por exemplo, dado o número $0.5 \in [0, 1]$, qual é próximo elemento de $[0, 1]$? Poderíamos dizer 0.6, mas e 0.51? Este ainda está mais longe de 0.5 que 0.501. De fato 0, 5001, 0.50001 etc. é uma sequência infinita de números em $[0, 1]$ cada vez mais próxima de 0.5 de forma que não é possível determinar o próximo elemento na ordenação do conjunto. Os conjuntos enumeráveis mais conhecidos são \mathbb{N} , \mathbb{Z} e \mathbb{Q} , sendo que este último é um pouco mais difícil de aplicar a regra intuitiva acima. Os conjuntos não enumeráveis mais conhecidos são \mathbb{R} , $\mathbb{R} \setminus \mathbb{Q}$.

Definição 2.3 (Variável Aleatória Discreta). Se o conjunto dos possíveis valores da variável aleatória é finito ou infinito enumerável.

Definição 2.4 (Variável Aleatória Contínua). Se o conjunto dos possíveis valores da variável aleatória é infinito não-enumerável.

Na prática, é comum a utilização de variáveis aleatórias contínuas pois estas são matematicamente mais simples de se tratar. Quando, por exemplo, falamos que a renda é uma v.a. contínua (na verdade ela é discreta) é pela conveniência da aproximação.

2.2.1 Distribuição de probabilidade

A função que descreve as probabilidades da variável aleatória discreta X assumir os diferentes valores do espaço amostral é chamada de função massa de probabilidade. No caso de uma variável contínua, a probabilidade de uma variável aleatória assumir qualquer valor específico é 0. Neste caso o análogo da função massa de probabilidade é a função de densidade de probabilidade (abreviado f.d.p. ou ainda, do inglês, p.d.f.) que, em poucas palavras, descreve a variação instantânea da probabilidade no ponto. Para que uma função qualquer f seja uma densidade de probabilidade é necessário que

$$\begin{aligned} f(x) &\geq 0 \quad \text{para todo } x \in \mathbb{R}, \\ \int_{\mathbb{R}} f(x) dx &= \int_{-\infty}^{\infty} f(x) dx = 1. \end{aligned} \tag{2.1}$$

Como a probabilidade de ocorrência de um valor em particular de uma variável aleatória contínua é sempre 0, probabilidades são discutidas em termos de intervalos, ou mesmo outros tipos de conjuntos. Essas probabilidades são obtidas por meio de integração da função densidade no intervalo especificado. Por exemplo, seja X uma variável aleatória com densidade $f(x)$. Então $P(a \leq X \leq b)$ é dada por

$$P(a \leq X \leq b) = \int_a^b f(x)dx.$$

Analogamente, para um conjunto $A \subseteq \mathbb{R}$ qualquer,

$$P(X \in A) = \int_A f(x)dx.$$

A probabilidade de que a variável aleatória X assuma valores inferiores ou igual a um número $x \in \mathbb{R}$, $P(X \leq x)$, possui importância intrínseca pois representa a probabilidade acumulada até o ponto x . Por isso, para cada $x \in \mathbb{R}$ fixo, denotamos esta probabilidade por

$$F(x) = P(X \leq x)$$

e a função assim definida $F : \mathbb{R} \rightarrow [0, 1]$ é chamada de função de distribuição acumulada (denotada por f.d.a.), ou somente função de distribuição. Note que se X é uma variável aleatória contínua com densidade f ,

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt.$$

2.2.2 Distribuições conjunta, marginal e condicional

Geralmente estamos interessados não apenas numa variável aleatória mas na relação entre algumas variáveis aleatórias. Suponha que temos duas variáveis aleatórias, X e Y . Agora além do comportamento probabilístico individual de X e Y , caracterizado por suas funções de distribuições, digamos F_X e F_Y , respectivamente, precisamos alguma forma de descrever o comportamento probabilístico conjunto de X e Y . Para isso definimos a função de distribuição acumulada de X e Y , denotada por $F_{X,Y}$, por

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y).$$

Se X e Y são ambas contínuas, podemos definir a densidade conjunta de X e Y denotada por $f_{X,Y}$, como sendo a função que satisfaz

$$F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(z, w)dzdw.$$

A função de distribuição conjunta de um par de variáveis aleatórias X e Y caracteriza também os comportamentos probabilísticos de X e Y individualmente. De fato

$$F_X(x) = \lim_{y \rightarrow \infty} F_{X,Y}(x, y) \quad \text{e} \quad F_Y(y) = \lim_{x \rightarrow \infty} F_{X,Y}(x, y)$$

e também

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y)dy \quad \text{e} \quad f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y)dx.$$

Quando temos a função de distribuição conjunta de um par X e Y de variáveis aleatórias, dizemos que as densidades/distribuições individuais de X e Y são as densidades/distribuições marginais de X e Y .

A função de distribuição condicional de X dado $Y = y$ é descrita por

$$F_{X|Y}(x|y) = P(X \leq x|Y = y) = \begin{cases} \frac{P(X \leq x, Y = y)}{P(Y = y)}, & \text{se } X \text{ é discreta e } P(Y = y) \neq 0 \\ \frac{\int_{-\infty}^x f_{X,Y}(t, y)dt}{f_Y(y)}, & \text{se } X \text{ é contínua e } f_Y(y) \neq 0 \end{cases}$$

- As densidades condicionais são:

- a) $f_{X|Y}(x|y)$, que é a densidade de X dado $Y = y$;
- b) $f_{Y|X}(y|x)$, que é a densidade de Y dado $X = x$.

Formalmente, temos a relação

$$F_{X|Y}(x|y) = \int_{-\infty}^x f_{X|Y}(t|y) dt \quad \text{e} \quad F_{Y|X}(y|x) = \int_{-\infty}^y f_{Y|X}(t|x) dt,$$

no caso em que X e Y são contínuas. Relações parecidas valem no caso em que X e Y são discretas, trocando-se integrais por somas e densidades por função massa de probabilidade.

A densidade conjunta pode ser escrita como o produto das densidades marginal e condicional da seguinte forma:

$$\begin{aligned} f_{X,Y}(x,y) &= f_X(x)f_{Y|X}(y|x) \\ &= f_Y(y)f_{X|Y}(x|y). \end{aligned}$$

Se $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ para todo x e y , então X e Y são chamadas de variáveis *independentes*. Note que, se eles são independentes,

$$f_{X|Y}(x|y) = f_X(x) \quad \text{e} \quad f_{Y|X}(y|x) = f_Y(y),$$

isto é, as distribuições condicionais são as mesmas que as marginais. Intuitivamente, quando X e Y são independentes X não carrega nenhuma informação útil a respeito de Y , assim o fato de Y ser ou não conhecido é irrelevante para a determinação de X .

2.3 A distribuição Normal e distribuições relacionadas

Existem algumas distribuições de probabilidade cujas probabilidades que, devido à sua utilização em diversas aplicações, valores de suas funções de distribuição são tabuladas. Dentre estas distribuições notáveis, podemos citar distribuição normal e as distribuições χ^2 , t e F , as quais discutiremos juntamente com as distribuições lognormal e normal bivariada. Existem diversas outras distribuições para as quais tabelas extensivas estão disponíveis. Como exemplos citamos as distribuições gama e beta. Na verdade, a distribuição χ^2 é um caso particular da distribuição gama, e as distribuições t e F são casos particulares da distribuição beta. Trataremos aqui apenas das citadas.

Existe um grande critionismo sobre a adequação da distribuição normal para descrever variáveis econômicas. Muitas vezes a distribuição normal de fato não é apropriada. Contudo, dois fatos tornam o estudo da distribuição normal importantes: primeiramente, embora existam problemas em que o uso da distribuição normal é questionável, existe um número muito maior de problemas em que o uso desta é totalmente apropriado. Segundo, mesmo que as variáveis não sejam normalmente distribuídas, pode-se considerar transformações de variáveis que façam com que as variáveis transformadas se tornem normalmente distribuídas.

2.3.1 A distribuição Normal

A distribuição normal, cuja densidade possui um formato que lembra um sino, é a distribuição mais amplamente utilizada em aplicações estatísticas numa grande variedade de áreas. Dizemos que X tem distribuição normal com média $\mu \in \mathbb{R}$ e variância $\sigma^2 > 0$, denotado compactamente por $X \sim N(\mu, \sigma^2)$, se sua função de densidade de probabilidade for dada por

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right], \quad \text{para } x \in \mathbb{R}.$$

Os parâmetros μ e σ^2 são também chamados de parâmetros de locação e escala, respectivamente.

Se $\mu = 0$ e $\sigma = 1$, a distribuição é chamada de “distribuição normal padrão” e a função de densidade de probabilidade reduz-se a,

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Uma propriedade importante da distribuição normal é que qualquer combinação linear de variáveis normalmente distribuídas também é normalmente distribuída. De fato, pode-se mostrar que, se

$$X_1 \sim N(\mu_1, \sigma_1^2) \quad \text{e} \quad X_2 \sim N(\mu_2, \sigma_2^2)$$

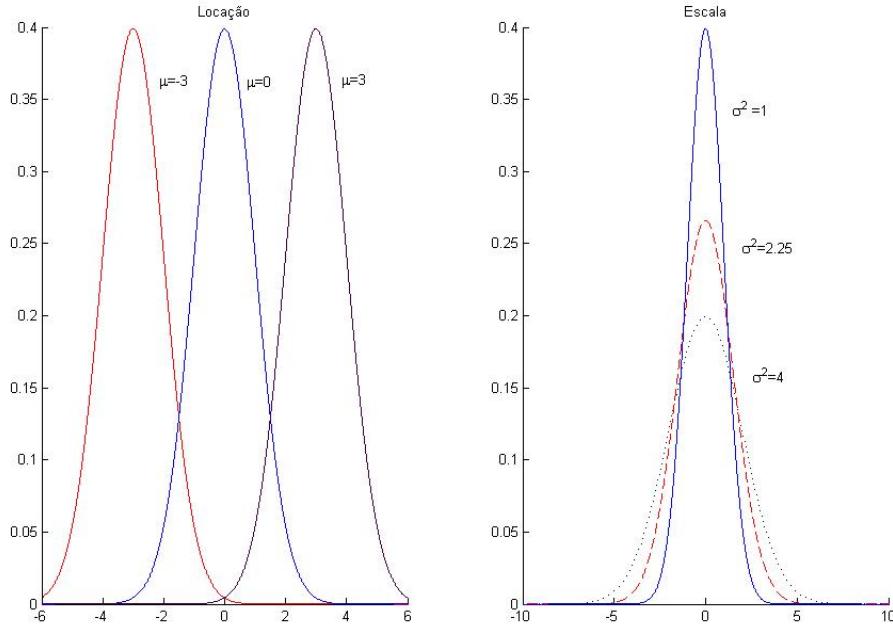


Figure 2.3: Função densidade Normal com diferentes parâmetros de locação e escala.

e a correlação entre X_1 e X_2 é ρ , então

$$a_1 X_1 + a_2 X_2 \sim N(a_1 \mu_1 + a_2 \mu_2, a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + 2\rho a_1 a_2 \sigma_1 \sigma_2).$$

Em particular,

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\rho \sigma_1 \sigma_2)$$

e

$$X_1 - X_2 \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2 - 2\rho \sigma_1 \sigma_2).$$

2.3.2 Distribuições relacionadas

Além da distribuição normal, há outras distribuições de probabilidade que usaremos com frequência. São elas as distribuições χ^2 , t e F , tabuladas no apêndice. Estas distribuições são derivadas da distribuição normal e definidas como descrito a seguir.

2.3.2.1 Distribuição χ^2

A distribuição χ^2 é bastante importante em aplicações e é definida a partir da soma dos quadrados de variáveis normais. Mais especificamente, se X_1, X_2, \dots, X_n são variáveis aleatórias independentes com distribuição normal padrão então

$$Q = \sum_{i=1}^n X_i^2$$

tem distribuição χ^2 com n graus de liberdade (g.l.), e escrevemos isso compactamente como $Q \sim \chi_n^2$.

Se $X_i \sim N(\mu, \sigma^2)$, então Q deve ser definido por

$$Q = \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2}.$$

A distribuição χ^2 também satisfaz uma determinada “propriedade de adição”, no seguinte sentido: se $Z_1 \sim \chi_n^2$ e $Z_2 \sim \chi_m^2$ e Z_1 e Z_2 são independentes, então $Z_1 + Z_2 \sim \chi_{n+m}^2$. Note que esta propriedade de adição é bem mais restritiva que aquela da distribuição normal, já que exige independência para que a simples soma das variáveis satisfaçam à propriedade (para normal, a propriedade vale para combinações lineares quaisquer), mas ainda assim é muito útil na prática.

2.3.2.2 Distribuição t

Se $X \sim N(0, 1)$, $Y \sim \chi_n^2$, e X e Y são independentes, a variável

$$T = \frac{X}{\sqrt{Y/n}} = \frac{\sqrt{n}X}{\sqrt{Y}}$$

possui distribuição t com n g.l. Escrevemos isso como $T \sim t_n$. O subscrito n novamente denota os g.l. Assim como a distribuição normal, a distribuição t é uma distribuição de probabilidade simétrica, com forma lembrando um sino, sendo porém mais achatada e com caudas mais “pesadas” que a normal. Quando o número de graus de liberdade n de uma variável t_n tende ao infinito, obtemos a distribuição normal. Em outras palavras, quando os graus de liberdade de uma variável aleatória com distribuição t_n for grande, esta tem comportamento probabilístico muito similar ao de uma normal.

2.3.2.3 Distribuição F

Se $Y_1 \sim \chi_{n_1}^2$, $Y_2 \sim \chi_{n_2}^2$ e Y_1 e Y_2 são independentes, a variável

$$F = \frac{Y_1/n_1}{Y_2/n_2} = \frac{n_2 Y_1}{n_1 Y_2}$$

é dita possuir distribuição F com n_1 e n_2 g.l. Escrevemos isso como $F \sim F_{n_1, n_2}$. O primeiro subscrito n_1 , refere-se aos g.l. do numerador, e o segundo subscrito, n_2 , refere-se aos g.l. do denominador.

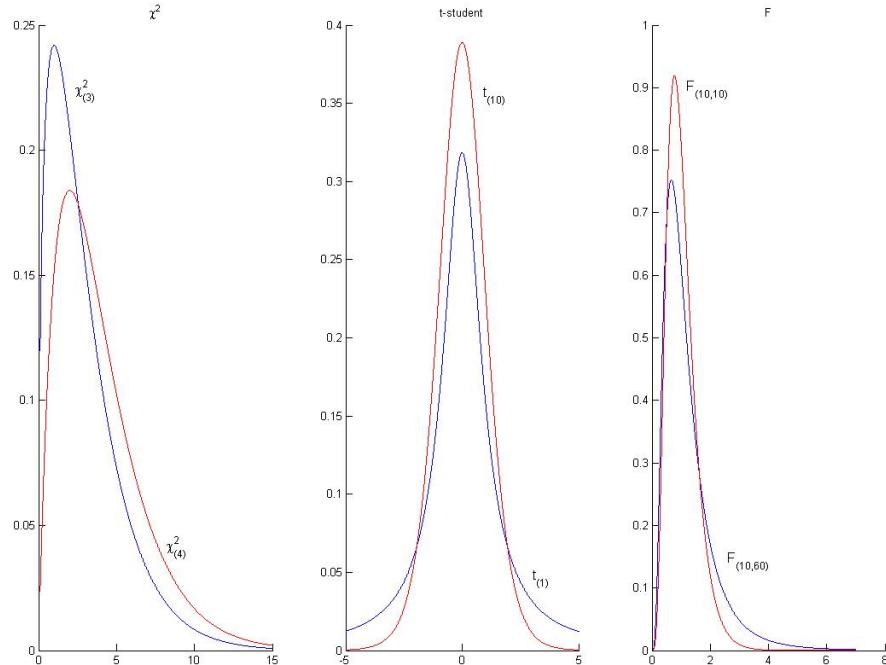


Figure 2.4: Função densidade Qui-Quadrado, t-Student e F-Snedecor. Em parênteses os graus de liberdade.

2.4 Parâmetros, estimadores e valores estimados

Considere o deslocamento de uma partícula no vácuo, em superfície sem atrito. Aprendemos cedo que a velocidade da partícula num instante de tempo t , v_t , é dada por $v_t = v_0 + at$, onde v_0 é a velocidade inicial da partícula, $a > 0$ é a aceleração aplicada na partícula, neste caso assumida constante. Neste modelo idealizado, a velocidade de uma partícula é uma função linear do tempo, cujo gráfico é apresentado na Figura 2.5(a).

Um grupo de pesquisadores realizou o seguinte experimento: numa superfície lisa, porém não absolutamente sem atrito, ao ar livre (isto é, na presença de vento, partículas de poeira, etc.) uma partícula foi acelerada à uma determinada aceleração desconhecida, mas constante em cada repetição do experimento, à partir de uma velocidade inicial desconhecida, mas também constante em cada repetição do experimento. Após um determinado tempo t a velocidade da partícula foi medida. Como resultados obtemos pares (v_i, t_i) representando a i -ésima observação da velocidade da partícula, medida no tempo t_i . Os resultados estão apresentados na Figura 2.5(b). Nossa interesse é determinar a velocidade inicial da partícula e a aceleração, que são chamados de parâmetros populacionais. Note que devido às condições não serem ideais, os dados não estão perfeitamente alinhados em uma reta como o estipulado na teoria, mas estão aproximadamente alinhados. Os desvios da reta “esperada” podem ser interpretados como sendo aleatórios, e são devidos aos vários fatores que estão fora de nosso controle, como atrito, vento, partículas em suspensão no ar, etc, fatores que estão em desalinho com a teoria.

Para estimar os parâmetros a e v_0 , que denotaremos por \hat{a} e \hat{v}_0 , podemos utilizar os estimadores de Mínimos Quadráticos Ordinários que conhecemos, neste caso, dados por (mais detalhes serão fornecidos adiante)

$$\hat{a} = \frac{\sum_{i=1}^n (v_i - \bar{v})(t_i - \bar{t})}{\sum_{i=1}^n (t_i - \bar{t})^2} \quad \text{e} \quad \hat{v}_0 = \bar{v} - \hat{a}\bar{t},$$

onde \bar{v} denota a média das velocidades e \bar{t} denota a média dos tempos observados. Note que, fornecidos os dados para o estimador, ele retorna dois valores sendo eles a estimativa dos parâmetros a e v_0 baseados nos dados. Note que mudando os dados, o estimador continua sendo o mesmo, mas os valores retornados por ele, as estimativas, mudarão. À partir dessas estimativas obtemos a reta apresentada na Figura 2.5(c)

Na resolução do problema aparecem 3 objetos eminentemente diferentes, cada um deles fundamental na solução do problema e que devem ser entendidos com clareza. Primeiramente temos os parâmetros populacionais, que são os valores de interesse, mas que nos são desconhecidos. Baseado numa amostra, gostaríamos, de alguma forma identificar, esses parâmetros. Segundo temos um estimador, que é uma função dos dados. Quando alimentado de dados estes estimadores retornam valores. Os valores retornados pelo estimador compreendem o terceiro objeto mencionado: são os valores estimados dos parâmetros populacionais.

Esta distinção entre parâmetro, estimador e valor estimado é essencial e está no coração das aplicações de estatística à dados reais.

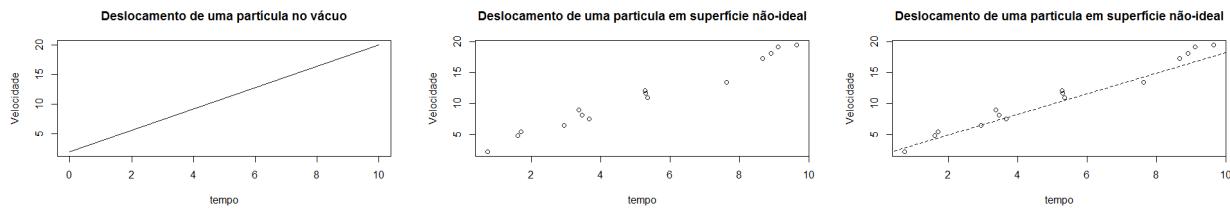


Figure 2.5: Partícula

2.5 Propriedades de variáveis aleatórias

2.5.1 Média, valor esperado ou esperança matemática

A Média ou valor esperado, ou ainda a esperança matemática de uma variável aleatória representa o valor médio assumido pela variável em questão. Esta pode ser interpretada como a média ponderada de cada valor assumido pela variável ponderado pela sua probabilidade de ocorrência.

Definição 2.5 (Média, valor esperado ou esperança matemática de variáveis aleatórias discretas.). Suponha que X seja uma variável aleatória discreta assumindo n valores diferentes x_1, \dots, x_n com probabilidades p_1, \dots, p_n , respectivamente. Então a média, ou valor esperado ou anda a esperança da variável X é definida por

$$\mathbb{E}(X) = x_1p_1 + x_2p_2 + \dots + x_np_n = \sum_{i=1}^n x_i p_i.$$

Observe que, no caso discreto, a esperança de uma variável X nada mais é do que a média ponderada de cada valor assumido pela variável pela sua probabilidade de ocorrência.

Exemplo 2.5. Seja X o valor da face superior obtida no lançamento de um dado equilibrado. Neste caso temos $P(X = 1) = P(X = 2) = P(X = 3) = P(X = 4) = P(X = 5) = P(X = 6) = \frac{1}{6}$, ou seja $p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = \frac{1}{6}$. Segue que

$$\begin{aligned}\mathbb{E}(X) &= \sum_{i=1}^6 p_i x_i = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 \\ &= \frac{1}{6} (1 + 2 + 3 + 4 + 5 + 6) = \frac{1}{6} \cdot \frac{6(6+1)}{2} \\ &= \frac{21}{6} = \frac{7}{2} = 3,5.\end{aligned}$$

O valor 3,5 obtido no resultado deve ser interpretado da seguinte forma: se jogarmos um dado equilibrado um número grande de vezes e calcularmos a média dos valores obtidos, ele será próximo à 3,5. De fato, se fosse possível repartir o experimento um número infinito de vezes, a média dos resultados convergiria para 3,5.

Definição 2.6 (Valor Esperado de $g(X)$). Seja X uma variável aleatória discreta assumindo n valores diferentes x_1, \dots, x_n com probabilidades p_1, \dots, p_n , respectivamente. Seja g uma função definida na imagem da variável aleatória de X . Então $\mathbb{E}(g(X))$ é dado por

$$\mathbb{E}(g(X)) = g(x_1)p_1 + \dots + g(x_n)p_n = \sum_{i=1}^n g(x_i)p_i.$$

Exemplo 2.6. Para o Exemplo considere $g(X) = X^2$. Obtemos

$$\begin{aligned}\mathbb{E}(X^2) &= \sum_{i=1}^6 p_i x_i^2 = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 9 + \frac{1}{6} \cdot 16 + \frac{1}{6} \cdot 25 + \frac{1}{6} \cdot 36 \\ &= \frac{1}{6} (1 + 4 + 9 + 16 + 25 + 36) = \frac{1}{6} \cdot \frac{6(6+1)(12+1)}{6} \\ &= \frac{91}{6} = 15,16666.\end{aligned}$$

Note que $\mathbb{E}(X^2) \neq \mathbb{E}(X)^2$.

Definição 2.7 (Esperança de variáveis aleatórias contínuas.). Supondo que X seja uma variável aleatória contínua com função de densidade de probabilidade f , definimos a esperança de X por

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx.$$

O valor esperado de uma função integrável qualquer de X , digamos $g(X)$ é definido por

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx.$$

Exemplo 2.7. Se $X \sim N(\mu, \sigma^2)$, então $\mathbb{E}(X) = \mu$, como pode ser facilmente computado.

2.5.2 Propriedades da Esperança

No que segue, assumimos que X, Y são variáveis aleatórias e a, b, c são constantes reais.

- **E1** - $\mathbb{E}(a) = a$;
- **E2** - $\mathbb{E}(a + X) = a + \mathbb{E}(X)$;
- **E3** - $\mathbb{E}(bX) = b\mathbb{E}(X)$;
- **E4** - $\mathbb{E}(a + bX) = a + b\mathbb{E}(X)$;
- **E5** - $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$;
- **E6** - $\mathbb{E}(a + bX + cY) = a + b\mathbb{E}(X) + c\mathbb{E}(Y)$;

Estas propriedades podem ser generalizadas para qualquer número de variáveis aleatórias.

Em particular, segue a esperança de uma combinação linear de variáveis aleatórias é a combinação linear das suas esperança, isto é, se X_1, \dots, X_n são variáveis aleatórias e a_1, \dots, a_n são constantes reais,

- **E7** - $\mathbb{E}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i \mathbb{E}(X_i)$.

Por esse motivo, a função $\mathbb{E}(\cdot)$ que associa a cada variável aleatória o seu valor esperado é um *operador linear*, chamado de *operador esperança*.

Em geral, temos que $\mathbb{E}(XY) \neq \mathbb{E}(X)\mathbb{E}(Y)$. Porém, no caso particular em que X e Y são variáveis aleatórias independentes, a igualdade é válida, isto é,

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y) \quad \text{se, e somente se, } X \text{ e } Y \text{ são independentes.}$$

2.5.3 Variância

Seja X uma variável aleatória (contínua ou discreta) e defina $\mu = \mathbb{E}(X)$. Então a variância de X é definida por

$$\text{Var}(X) = \mathbb{E}[(X - \mu)^2] = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2. \quad (2.2)$$

Podemos interpretar a variância como sendo o valor esperado do quadrado do desvio de X da sua própria média. Em linguagem comum isto pode ser expresso como *A média do quadrado da distância de cada ponto até a média*. É assim a *média do quadrado dos desvios*. A variância da variável aleatória X é geralmente designada por $\text{Var}(X)$, σ_X^2 , ou simplesmente σ^2 . A variância é uma medida de dispersão dos dados e sua unidade é a unidade dos dados elevada ao quadrado. Lembramos que a raiz quadrada positiva da variância determina o chamado desvio padrão de X .

2.5.4 Covariância

A covariância entre duas variáveis aleatórias X e Y com $\mathbb{E}(X) = \mu_X$ e $\mathbb{E}(Y) = \mu_Y$ é definida por

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)].$$

Desenvolvendo a expressão para a covariância, temos:

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] \\ &= \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] \\ &= \mathbb{E}[XY - X\mathbb{E}(Y) - Y\mathbb{E}(X) + \mathbb{E}(X)\mathbb{E}(Y)]. \end{aligned}$$

Usando a propriedade de que a esperança da soma entre duas variáveis aleatórias é igual a soma das esperanças, segue que

$$\begin{aligned}\text{Cov}(X, Y) &= \mathbb{E}(XY) - \mathbb{E}[X\mathbb{E}(Y)] - \mathbb{E}[Y\mathbb{E}(X)] + \mathbb{E}[\mathbb{E}(X)\mathbb{E}(Y)] \\ &= \mathbb{E}(XY) - \mathbb{E}(Y)\mathbb{E}(X) - \mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y) \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\end{aligned}\tag{2.3}$$

Note que quando X e Y são independentes, temos que $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ de onde segue que $\text{Cov}(X, Y) = 0$. A recíproca, porém, não é verdadeira pois existem exemplos de variáveis dependentes que possuem covariância nula. Observe ainda que da expressão (2.3) podemos concluir que a covariância é uma forma de medir o quão “distante” X e Y estão de ser independentes.

2.5.5 Correlação

A correlação, também chamada de coeficiente de correlação, indica a força e a direção do relacionamento linear entre duas variáveis aleatórias, se existir. A correlação entre duas variáveis X e Y com $0 < \text{Var}(X) < \infty$ e $0 < \text{Var}(Y) < \infty$, denotado por $\text{Cor}(X, Y)$ ou $\rho_{x,y}$, é definida como

$$\text{Cor}(X, Y) = \rho_{x,y} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)}{\sqrt{\mathbb{E}(X^2) - \mathbb{E}^2(X)} \sqrt{\mathbb{E}(Y^2) - \mathbb{E}^2(Y)}}.$$

Note que a correlação entre X e Y nada mais é do que a covariância entre X e Y normalizada por seus desvios padrões. Esta normalização acaba dando à correlação uma interpretabilidade ausente na covariância como veremos a seguir.

Observe ainda que, quando $\text{Cov}(X, Y) = 0$, temos $\text{Cor}(X, Y) = 0$ também e X e Y são ditos ser variáveis não-correlacionadas.

2.5.6 Propriedades da variância, covariância e correlação

Se a e b forem constantes reais e X uma variável aleatória cuja variância está definida, então:

- **V1** - $\text{Var}(aX + b) = a^2\text{Var}(X)$;
- **V2** - $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$.

Da propriedade V1 segue que a variância de uma constante é zero. Além disso, se a variância de uma variável aleatória é zero, então esta variável assume um único valor com probabilidade 1. Da propriedade V2 segue que se X e Y são não-correlacionados, então a variância da soma é a soma das variâncias.

Suponha agora que X e Y são variáveis aleatórias e a, b, c e d são constantes reais. Então

- **Cv1** - $\text{Cov}(X, X) = \text{Var}(X)$;
- **Cv2** - $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;
- **Cv3** - $\text{Cov}(aX + b, cY + d) = ac\text{Cov}(X, Y)$;
- **Cv4** - $\text{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m \text{Cov}(X_i, Y_j)$.

Como mencionado anteriormente, se X e Y são independentes, então $\text{Cov}(X, Y) = 0$.

A correlação, por sua vez, possui as seguintes propriedades:

- **Cr1** - $|\text{Cor}(X, Y)| \leq 1$;

- **Cr2** - $\text{Cor}(X, Y) = 1$ se, e somente se, X é diretamente proporcional a Y no sentido de que $X = a + bY$ para $a \in \mathbb{R}$ e $b > 0$;
- **Cr3** - $\text{Cor}(X, Y) = -1$ se, e somente se, X é inversamente proporcional a Y no sentido de que $X = a + bY$ para $a \in \mathbb{R}$ e $b < 0$;
- **Cr4** - $\text{Cor}(X, Y) = \text{Cor}(Y, X)$;
- **Cr5** - $\text{Cor}(aX + b, cY + d) = \text{sign}(ac)\text{Cor}(X, Y)$, onde a função $\text{sign}(x)$ é a função sinal de x , sendo igual a -1 , se $x < 0$, 1 se $x > 0$ e 0 se $x = 0$;
- **Cr6** - Se X e Y são independentes, então $\text{Cor}(X, Y) = 0$. A reciproca, porém, não é verdadeira.

2.6 Estimadores

Dada uma amostra x_1, x_2, \dots, x_n de uma variável aleatória X , o estimador de $\mathbb{E}(X)$ é simplesmente a média aritmética dos dados:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Com relação à variância de X , existem dois estimadores muito utilizados na prática. O estimador da variância de X obtido pelo método de máxima verossimilhança é dado por

$$\hat{\sigma}_X^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right).$$

Pode-se mostrar que, embora consistente, este estimador é viesado em amostras finitas. Um estimador consistente e não-viesado em amostras finitas é dado por

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right).$$

Observe que para n grandes, a diferença entre os estimadores $\hat{\sigma}^2$ e S^2 é irrelevante. Em amostras pequenas, porém, o estimador S^2 apresenta uma performance melhor.

Seja x_1, x_2, \dots, x_n e y_1, y_2, \dots, y_n amostras aleatórias das variáveis aleatórias X e Y . Então um estimador para a covariância entre X e Y é dado por

$$\hat{\gamma}_{x,y} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1} \left(\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} \right).$$

Um estimador para a correlação entre X e Y é dado por

$$\hat{\rho}_{x,y} = \frac{\hat{\gamma}_{x,y}}{S_X S_Y}.$$

2.6.1 Propriedades dos estimadores

Dado que temos alguns estimadores definidos acima, é interessante estudar algumas das propriedades qualitativas dos estimadores que nos permitam determinar qual estimador é “bom” e qual não é. É também importante definir critérios para comparar diversos estimadores.

2.6.2 Vício/Viés

Seja $\hat{\theta}$ um estimador do parâmetro θ . o vício/viés (bias, em inglês) é definido como

$$b(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta. \quad (2.4)$$

Se $b(\hat{\theta}) = 0$ segue que $\mathbb{E}(\hat{\theta}) = \theta$ e, neste caso, dizemos que $\hat{\theta}$ é não-viciado ou não-viesado para o parâmetro θ .

2.6.3 Consistência

Em estatística, uma seqüência de estimadores para o parâmetro θ é dito ser consistente (ou assintoticamente consistente) se esta seqüência converge em probabilidade para θ . Isso significa que as distribuições dos estimadores tornar-se mais e mais concentrados perto do verdadeiro valor do parâmetro a ser estimado, de modo que a probabilidade do estimador ser arbitrariamente perto de θ converge para um.

2.6.4 Eficiência

Um estimador de θ é dito ser eficiente se for não viesado e sua variância for menor ou igual a variância de qualquer outro estimador $\hat{\theta}$, ou seja,

$$\text{Var}(\hat{\theta}_0) \leq \text{Var}(\hat{\theta}), \text{ para qualquer outro estimador } \hat{\theta} \text{ de } \theta.$$

Na figura abaixo podemos observar a diferença entre vício e eficiência. Estes conceitos estão relacionados à média e à variância, respectivamente.

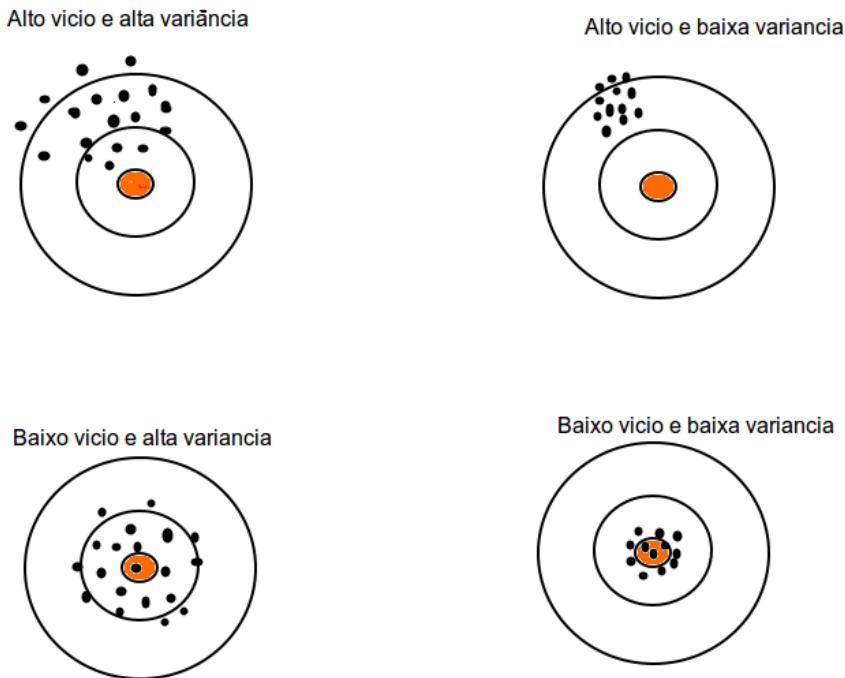


Figure 2.6: Diferença entre vício e eficiência.

2.6.5 Erro quadrático médio (EQM)

O erro quadrático médio de um estimador $\hat{\theta}$ de θ é definido como

$$EQM(\hat{\theta}) = \mathbb{E}(\hat{\theta} - \theta)^2. \quad (2.5)$$

Podemos reescrever esta ultima expressão como

$$EQM(\hat{\theta}) = \text{Var}(\theta) + [\mathbb{E}(\theta) - \theta]^2 = \text{Var}(\hat{\theta}) + b(\hat{\theta}).$$

Assim, o erro quadrático médio é definido como a variância do estimador mais o quadrado do seu viés. Podemos entender o EQM como sendo uma medida da performance de um estimador em relação ao seu vício e variância. Note que $EQM(\theta) = \text{Var}(\theta)$ sempre que o estimador for não-viciado.

2.6.6 Vício versus Variância mínima

O erro quadrático médio utilizado na comparação entre um ou mais estimadores para um mesmo parâmetro θ . Podemos observar de (2.5) que, no cálculo do EQM, existe um balanço entre vício e variância. Naturalmente, estimadores eficientes apresentarão um EQM mínimo dentre os estimadores não-viciados de θ . Muitas vezes, porém, pode ser mais vantajoso do ponto de vista prático a utilização de um estimador viciado mas com variância pequena em detrimento a um estimador de maior variância, mas que seja não-viciado. Isto ocorre por que se a variância de um estimado é muito grande, é grande a chance de uma estimativa esteja longe do verdadeiro valor do parâmetro, mesmo que o estimador seja não-viciado. Este é um ponto importante a ser observado quando da escolha de um estimador para um determinado problema.

2.6.7 Método de mínimos quadrados (MQO)

Considere o modelo

$$Y = \alpha + \beta X + U$$

onde Y é a variável dependente, X é a variável independente e U denota o termo de erro do modelo. Suponhamos que temos uma amostra $(x_1, y_1), \dots, (x_n, y_n)$ provindo deste modelo.

2.6.7.1 Qual critério devo utilizar para obter os estimadores dos parâmetros α e β ?

Podemos minimizar:

- Soma dos erros: não é um bom critério pois pode anular positivos e negativos.
- Soma Absoluta dos Resíduos: é um critério válido e intuitivo, porém seu estudo é de alta complexidade. Devido a isso, o estimador obtido por este critério, denominado LAD (Least Absolute Deviations), é pouco utilizado na prática.
- Soma dos Quadrados dos Erros: possui propriedades estatísticas de simples utilização e interpretação o que o tornam bastante atrativo. É este o critério que dá origem ao estimador de mínimos quadráticos ordinários (MQO).

Utilizando a soma dos quadrados dos erros como critério, devemos resolver o seguinte problema de optimização:

$$\min_{\{\alpha, \beta\}} \left\{ \sum_{i=1}^n u_i^2 \right\} = \min_{\{\hat{\alpha}, \hat{\beta}\}} \left\{ \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \right\}. \quad (2.6)$$

As condições de primeira ordem (CPO's) são obtidas diferenciando-se o argumento do lado direito de (2.6) em relação à α e β . Em α , a solução do problema de optimização será o valor $\hat{\alpha} \in \mathbb{R}$ que satisfaz

$$-2 \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i) = 0 \implies \sum_{i=1}^n \hat{u}_i = 0.$$

Esta CPO nos mostra que a escolha do intercepto ótimo implica que a soma dos resíduos será zero. Continuando com essa CPO

$$\begin{aligned} \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i) = 0 &\iff n\bar{y} - n\hat{\alpha} - \hat{\beta}n\bar{x} = 0 \\ &\iff \hat{\alpha}_{MQO} = \bar{y} - \hat{\beta}\bar{x}. \end{aligned} \quad (2.7)$$

Assim, o estimador de MQO do intercepto α é dado por (2.7).

Difereciando-se o argumento do lado direito de (2.6) em relação à β obtemos que a solução do problema de optimização será o valor $\hat{\beta} \in \mathbb{R}$ que satisfaz

$$\begin{aligned} \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2 = 0 &\iff \sum_{i=1}^n y_i x_i - \hat{\alpha} \sum_{i=1}^n x_i - \hat{\beta} \sum_{i=1}^n x_i^2 = 0 \\ &\iff \sum_{i=1}^n y_i x_i = (\bar{y} - \hat{\beta}\bar{x}) \sum_{i=1}^n x_i + \hat{\beta} \sum_{i=1}^n x_i^2 \\ &\iff \sum_{i=1}^n y_i x_i = \bar{y} \sum_{i=1}^n x_i + \hat{\beta} \left(\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i \right), \end{aligned}$$

onde a última igualdade obtém-se dividindo-se o numerador e denominador por $n - 1$.

2.6.8 Regressão liner múltipla (RML)

Considere o modelo de regressão linear múltipla

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + u_i$$

em que temos k variáveis explicativas x_1, \dots, x_k . Definindo

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix},$$

e

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad U = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

obtemos o modelo de regressão em forma matricial $Y = X\beta + U$. A matriz X é chamada de matriz de design do modelo. Pode-se mostrar que o estimador de MQO para β é dado por:

$$\hat{\beta} = (X'X)^{-1}X'Y.$$

2.6.9 Hipóteses do modelo de regressão

- **Hipótese 1 (Linearidade dos Parâmetros):** A relação entre a variável dependente Y e as explicativas X_1, \dots, X_k é linear

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + U.$$

Definição 2.8. Um modelo de regressão é linear nos parâmetros se as CPOs associadas ao problema de obtenção dos EMQ (Estimadores de MQO) gerarem um sistema linear nos parâmetros.

Exemplo 2.8. Seja o seguinte modelo

$$Y = \alpha + \beta X + U.$$

e (x_i, y_i) , para $i = 1, \dots, n$, uma amostra do modelo. De acordo com o que foi visto anteriormente, o problema de optimização a ser resolvido para a obtenção dos estimadores de MQO para α e β será

$$\min_{\{\alpha, \beta\}} \left\{ \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \right\}.$$

As CPOs serão

$$\hat{\alpha} : -2 \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta} x_i) = 0 \implies \sum_{i=1}^n y_i = n \hat{\alpha} + \hat{\beta} \sum_{i=1}^n x_i$$

$$\hat{\beta} : -2 \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta} x_i) x_i = 0 \implies \sum_{i=1}^n y_i x_i = \hat{\alpha} \sum_{i=1}^n x_i + \hat{\beta} \sum_{i=1}^n x_i^2$$

$$\begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix} \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i x_i \end{bmatrix}.$$

Logo é o sistema linear e o modelo é linear nos parâmetros.

Exemplo 2.9. Seja o seguinte modelo

$$Y = \alpha + \beta X^\gamma + U$$

e seja (x_i, y_i) , para $i = 1, \dots, n$, uma amostra do modelo. O problema de minimização neste caso resume-se a

$$\min_{\{\alpha, \beta, \gamma\}} \left\{ \sum_{i=1}^n (y_i - \alpha - \beta x_i^\gamma)^2 \right\}.$$

A CPO em α é dada por

$$\alpha : -2 \sum_i (y_i - \alpha - \beta x_i^\gamma) = 0,$$

que não é linear por causa do γ .

Exemplo 2.10. Seja o seguinte modelo

$$Y = \alpha X_1^{\beta_1} X_2^{\beta_2} e^U.$$

Este modelo é claramente não-linear, porém, ao tomarmos o logaritmo obtemos

$$\ln(Y) = \ln(\alpha) + \beta_1 \ln(X_1) + \beta_2 \ln(X_2) + U,$$

que é linear nos parâmetros.

- **Hipótese 2 (Amostragem Aleatória):** Podemos extrair uma amostra aleatória

$$\{(x_{1i}, \dots, x_{ki}, y_i), i = 1, \dots, n\}$$

da população.

Observação. Nos livros-texto esta hipótese é geralmente substituída por uma hipótese de que X é determinístico (não aleatório) e seus valores podem ser escolhido de antemão.

- **Hipótese 3 (Média Condisional Zero):** $\mathbb{E}(U|X) = 0$
- **Hipótese 4 (Não há Multicolinearidade Perfeita):** As variáveis explicativas X_1, \dots, X_k são linearmente independentes. Logo, $X_j, j = 1, \dots, k$ não podem ser constantes. Lembrando que o posto de uma matriz X é a dimensão do subspaço gerado pelas colunas da matriz, esta hipótese implica que a matriz de design associada ao modelo,

$$X = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix}_{n \times (k+1)}$$

tem posto máximo, isto é, $\text{posto}(X) = k + 1$, pois $n \geq k + 1$. Relembre das propriedades de álgebra matricial que

$$\text{posto}(X'X) = \text{posto}(X) = k + 1,$$

e assim, $(X'X)$ é uma matriz invertível.

- **Hipótese 5 (Homocedasticidade):** Se U_1, \dots, U_n é a sequência de erros relativa ao modelo linear $Y = X\beta + U$ baseado numa amostra de tamanho n do modelo. Então $\text{Var}(U_i|X) = \sigma^2$, para todo i , ou seja, a variância do erro é constante.
- **Hipótese 6 (Ausência de (Auto)Correlação (Serial) Condisional):** $\text{Cov}(U_i, U_j|X) = 0$, para todo i e j com $i \neq j$.
- **Hipótese 7 (Normalidade):** $U_i \sim N(0, \sigma^2)$ para todo i . Tal hipótese será necessária para inferência.

Teorema 2.1 (de Gauss-Markov). *Dentro da classe dos estimadores lineares e não-viesados, e dadas as hipóteses do MCRL, os EMQs são estimadores que possuem a menor variância (BLUE - Best Linear Unbiased Estimator).*

2.6.10 O coeficiente de determinação

Existe alguma medida que mostre que um determinado modelo apresenta um bom poder preditivo? Ou seja, se o regressor (X) que eu inclui no meu modelo explica bem a variável dependente (Y)? Para construirmos tal medida, primeiramente definimos

$$\begin{aligned} \sum_{i=1}^n (y_i^*)^2 &= \text{Soma dos Quadrados Totais (SQT)} \\ \sum_{i=1}^n (\hat{y}_i^*)^2 &= \text{Soma dos Quadrados Explicados (SQE)} \\ \sum_{i=1}^n \hat{u}_i^2 &= \text{Soma dos Quadrados dos Resíduos (SQR)} \end{aligned}$$

Pode-se mostrar facilmente que

$$SQT = SQE + SQR.$$

Dividindo a expressão por SQT , teremos

$$1 = \frac{SQE}{SQT} + \frac{SQR}{SQT}.$$

O R^2 mede o quanto (em porcentagem) da variação da variável dependente pode ser explicado pela introdução do regressor no modelo. Pode-se mostrar que $R^2 \in [0, 1]$. Expressões alternativas para R^2 são as que segue:

$$R^2 = \frac{SQE}{SQT} = 1 - \frac{SQR}{SQT} = \frac{\sum_i (\hat{y}_i^*)^2}{\sum_i (y_i^*)^2} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sum_i \hat{u}_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

Uma deficiência do R^2 é que este nunca diminui quando adicionamos regressores, o que implica que o R^2 favorece modelos mais complexos. Para minimizar esta deficiência, uma alternativa é penalizar, em certo grau, a inclusão de regressores. Um coeficiente muito utilizado na prática e que faz exatamente isso é o chamado R^2 **ajustado** definido por

$$\begin{aligned}\bar{R}^2 &= 1 - \frac{[SQR/(n-k-1)]}{[SQT/(n-1)]} \\ &= 1 - \frac{\sigma^2}{[SQT/(n-1)]}, \quad \left(\sigma^2 = \frac{SQR}{n-k-1} \right).\end{aligned}$$

O R^2 ajustado também recebe o nome de R^2 corrigido ou, em inglês, de *R-bar squared*.

Pode-se mostrar que $SQR/(n-k-1)$ é um estimador não-viesado de σ^2 , a variância populacional do erro, e $SQT/(n-1)$ é um estimador não-viesado de σ_Y^2 , a variância de Y .

Proposição 2.1. Se adicionamos um novo regressor à regressão, então \bar{R}^2 aumenta e a estatística t deste novo regressor é maior que 1, em módulo.

Proposição 2.2. Adicionando um grupo de variáveis à regressão, então \bar{R}^2 aumenta e a estatística F deste novo grupo de regressores é maior que 1.

Uma fórmula alternativa para o \bar{R}^2 é

$$\bar{R}^2 = 1 - \frac{(1-R^2)(n-1)}{(n-k-1)}.$$

Além de permitir a comparação entre modelos ao se incluir/excluir regressores, o \bar{R}^2 serve também para a escolha dentre modelos *nonnested* (não encaixantes). Por exemplo, o modelo 1 que tem X_1 , X_2 e X_3 como variáveis explicativas e um outro modelo 2 que tem X_1 , X_2 e X_4 . Mas o \bar{R}^2 não serve para escolher dentre formas funcionais diferentes da variável **dependente**.

2.6.11 Propriedade de não-viés dos estimadores MQO

Assumindo X não estocástico, tomando a esperança dos estimadores MQO em versão matricial, obtemos:

$$\begin{aligned}\mathbb{E}(\hat{\beta}) &= \mathbb{E}[(X'X)^{-1}X'y] = \mathbb{E}[(X'X)^{-1}X'(X\beta + U)] \\ &= \mathbb{E}[(X'X)^{-1}X'X\beta] + \mathbb{E}[(X'X)^{-1}X'U] \\ &= \beta + (X'X)^{-1}\mathbb{E}[X'U] = \beta,\end{aligned}$$

pois $\mathbb{E}[X'U] = 0$ por hipótese. Ou seja, se as variáveis regressoras são não-correlacionadas com U , o estimador MQO será não-viesado.

2.6.11.1 Variância dos estimadores MQO

Para um modelo de regressão linear múltipla, a variância do estimador de cada β_j é dado por

$$\text{Var}(\hat{\beta}_j) = \begin{cases} \frac{\sigma_u^2}{\text{Var}(X_j)}, & \text{se a variância de } U, \sigma_U^2 \text{ é conhecida;} \\ \frac{1}{n-1} \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\text{Var}(X_j)}, & \text{se } \sigma_U^2 \text{ é desconhecida.} \end{cases}$$

2.6.12 Testes de hipóteses

2.6.12.1 Teste t

Se queremos testar individualmente a significância ($H_0 : \beta_j = 0$) do modelo

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + u_i$$

, a estatística de teste é dada por

$$t = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\text{Var}\hat{\beta}_j}} \sim t_{n-k-1}$$

Observação. Se houver problema de multicolineariedade, R_j^2 será alto, a variância será alta, e a estatística de teste t será baixa, e os estimadores serão pouco significativos (neste caso assumindo $\beta_j = 0$).

2.6.12.2 Teste F

A estatística F para um modelo com intercepto, que serve para testar se o modelo é significante, ou seja se todos os regressores são conjuntamente significantes, i.e. $H_0 : \beta_0 = \beta_1 = \cdots = \beta_k = 0$ vs. $H_1 : \text{pelo menos um } \beta_j \neq 0$, é dada por

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} \sim F_{k,n-k-1}.$$

Observação. Se temos um problema de multicolineariedade, ainda assim a estatística F e R^2 do modelo de y contra x não depende da correlação entre os regressores (apenas do SQR e SQT, ou seja, da soma dos quadrados dos resíduos e da variável dependente) e, assim, se tivermos regressores relevantes para explicar y , então F e R^2 indicarão que o modelo como um todo terá um alto poder explicativo.

2.7 Formas funcionais logarítmicas

Considere o seguinte modelo:

$$\widehat{\log y} = \hat{\beta}_0 + \hat{\beta}_1 \log x_1 + \hat{\beta}_2 x_2.$$

- Ele é log-log de y em relação a x_1 e é log-linear em relação a x_2 .
- β_1 mede a elasticidade de y em relação a x_1 , fixado x_2 .
- A interpretação de $\hat{\beta}_1$ é que para o aumento de 1% em x_1 temos um aumento de $\beta_1\%$ em y .
- $\hat{\beta}_2$ pode ser interpretado como: um aumento de uma unidade em x_2 dá um aumento exato de $100[\exp \beta_2 - 1]\%$ em y .
- Uma medida aproximada, para uma mudança pequena em x_2 seria $100\hat{\beta}_2\%$. Este coeficiente é denominado muitas vezes como semi-elasticidade.

2.8 Exercícios

Exercício 2.1. O custo de produção de certo bem é uma variável aleatória com função densidade de probabilidade:

$$f(x) = kx^2, \quad 1 \leq x \leq 4.$$

- a) Calcule o valor de k ;
- b) Calcule o custo médio do produto;
- c) Calcule a probabilidade do custo ser menor do que 2;
- d) Calcule a variância do custo do produto;
- e) Calcule a probabilidade do custo ser maior do que 3;

Exercício 2.2. Sejam X e Y duas variáveis aleatórias independentes com média $\mu_X = \mathbb{E}(X) = 4$, $\mu_Y = \mathbb{E}(Y) = 5$, $\sigma_X^2 = \text{Var}(X) = 1$ e $\sigma_Y^2 = \text{Var}(Y) = 2$.

- a) Calcule $\mathbb{E}(X^2)$ e $\mathbb{E}(Y^2)$;
- b) Calcule $\text{Var}(4X - 2Y)$;
- c) Calcule $\text{Cov}(X, Y)$;
- d) Calcule $\text{Cov}(X, 2X - 3Y)$;

e) Suponha que X_1, X_2, \dots, X_n são variáveis aleatórias independentes entre si e independentes de X , mas com a mesma distribuição de probabilidade de X , ou seja, X_1, X_2, \dots, X_n e X são variáveis aleatórias independentes e identicamente distribuídas (i.i.d) com média $\mu = 4$ e variância $\sigma^2 = 1$. Calcule:

- i. $\mathbb{E}(\bar{X}) = \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n X_i\right)$;
- ii. $\text{Var}(\bar{X})$;
- iii. $\text{Cov}(\bar{X}, X)$.

Exercício 2.3. Suponha o seguinte modelo linear: $y = X\beta + \varepsilon$, em que y e ε são vetores $n \times 1$, $X < \infty$ é uma matriz $n \times k$ e β é um vetor $k \times 1$.

- a) Determine a(s) hipótese(s) necessária(s) para estimar esse modelo por MQO.
- b) Determine a(s) hipótese(s) necessária(s) para que o β estimado, $\hat{\beta}$, exista e seja único.
- c) Determine a(s) hipótese(s) necessária(s) para que $\hat{\beta}$ seja não viesado.
- d) Determine a(s) hipótese(s) necessária(s) para que $\hat{\beta}$ seja eficiente.
- e) Determine a(s) hipótese(s) necessária(s) para que se possa fazer inferência estatística.

Exercício 2.4. Os dados da tabela relacionam o peso de plantas, Y (em gramas) com o percentual de matéria orgânica na terra, X_1 e os Kilogramas de nitrogênio suplementares agregados a terra por $1000m^2$, X_2 :

	y	x1	x2
	78.5	7	2.6
	74.3	1	2.9
	104.3	11	5.6
	87.6	11	3.1
	95.9	7	5.2
	109.2	11	5.5
	102.7	3	7.1
Soma:	652.5	51	32.0
média:	93.21	7.29	4.57

- a) Defina a equação de regressão com intercepto em que y é a variável dependente e x_1 e x_2 são variáveis explicativas. Não esqueça da suposição para o termo de erro do modelo.

b) Se

$$(X^T X)^{-1} = \begin{bmatrix} 1.80 & -0.07 & -0.25 \\ -0.07 & 0.01 & -0.00 \\ -0.25 & -0.00 & 0.06 \end{bmatrix}, \text{ e } X^T Y = \begin{bmatrix} 652.50 \\ 4915.30 \\ 3103.66 \end{bmatrix},$$

determine $\hat{\beta}$ via MQO.

Resposta: $\hat{\beta} = (51.56, 1.49, 6.72)$.

- c) Se $SQ_{res} = 27.58$ e $SQ_{total} = 28.30$, calcule o coeficiente de determinação. Resposta: $R^2 = 0.9745$,
- d) Teste $\beta_0 = \beta_1 = \beta_2 = 0$, ou seja, a significância do modelo.
- e) Se $dp(\hat{\beta}_1) = 0.2636$, (dp =desvio padrão), teste se a variável X_1 é relevante para o modelo.
- f) Se $dp(\hat{\beta}_2) = 0.6274$, teste a hipótese $H_0 : \beta_2 = 1$.

Exercício 2.5. Adão Ismiti queria verificar se a produtividade aumentava com a divisão do trabalho. Para isso, fez a seguinte experiência: regrediu a produtividade (p) de n trabalhadores de fábricas de alfinetes contra o número de funções exercidas pelo trabalhador (F), os anos de escolaridade (E), o salário (w) e o número de filhos (N). Formalmente, a regressão foi:

$$p_i = \beta_1 + \beta_2 F_i + \beta_3 E_i + \beta_4 w_i + \beta_5 N_i + u_i$$

Usando o teste t -Student, Ismiti não rejeitou a hipótese nula de parâmetro igual a zero para β_3 . Retirou a variável E da regressão e estimou o modelo restrito, observando que $\hat{\beta}_5$ se tornou também, estatisticamente não significativo. Finalmente, retirou N da regressão e estimou o modelo novamente.

- a) Por que não foi preciso fazer o teste F em $\hat{\beta}_3$ para retirar E do modelo?
- b) Justifique se o procedimento adotado por Ismiti está correto ou equivocado, para ter eliminado a variável N do modelo.

Exercício 2.6. Suponha um modelo de regressão linear múltiplo em que $\hat{\beta}$ exista, seja não viesado e eficiente, pois u é homocedástico. Suponha que você imponha falsas restrições sobre os parâmetros do modelo.

- a) Mostre que as estimativas nesse caso são viesadas.
- b) Mostre que a variância das estimativas do modelo com restrições é menor que a variância das estimativas do modelo sem restrições.
- c) Qual é a implicação desse resultado em termos de previsão? Qual é a intuição desse resultado?

Sugestão: Lembre o que é o EQM, ou seja, o erro quadrático médio.

Exercício 2.7. Responda:

- a) Cite pelo menos dois testes para a hipótese de homocedasticidade.
- b) Cite pelo menos um teste para a hipótese de autocorrelação dos resíduos.
- c) Em caso de rejeição da hipótese nula em (a), por qual método você estimaria o modelo?
- d) Em caso de rejeição da hipótese nula em (b), por qual método você estimaria o modelo?

Exercício 2.8. Desafio: Faça os seguinte exercícios.

- a) Suponha que $\sum_{i=0}^{\infty} |x_i| < \infty$. Mostre que $\sum_{i=0}^{\infty} x_i^2 < \infty$.
- b) Prove (ou não) que $\lim_{n \rightarrow \infty} \sum_{x=1}^n \frac{1}{x} = \infty$.
- c) Prove (ou não) que $\lim_{n \rightarrow \infty} \sum_{x=1}^n \frac{1}{x^2} = \infty$.
- d) Prove (ou não) que, se $\sum_{i=0}^{\infty} x_i^2 < \infty$, então $\sum_{i=0}^{\infty} |x_i| < \infty$.

Chapter 3

Séries temporais no contexto de regressão

Neste capítulo abordamos regressão no contexto de séries temporais. Começamos definindo o que é uma série temporal e introduzimos algumas propriedades teóricas.

3.1 Introdução

Uma série temporal é qualquer conjunto de observações ordenadas no tempo. Alguns exemplos são citados abaixo:

- a) Estimativas trimestrais do Produto Interno Bruto (PIB);
- b) Valores diários da temperatura em Campo Bom;
- c) Índices diários da bolsa de valores de São Paulo;
- d) Quantidade anual de chuva na cidade do Porto Alegre;
- e) Um registro de marés no porto de Santos.

Nos exemplos de a) a d) temos séries temporais discretas, enquanto que e) é um exemplo de série contínua. Podemos obter uma série temporal discreta a partir da amostragem de uma série temporal contínua considerando intervalos de tempos iguais, Δt . Assim para analisar a série e) será necessário amostrá-la, convertendo-a e observando-a no intervalo de tempo $[0, T]$, supondo uma série discreta com N pontos, em que $N = \Delta t/T$ (T horas). Existem dois enfoques utilizados na análise de séries temporais. Em ambos, o objetivo é construir modelos para estas séries. No primeiro enfoque, a análise é feita no domínio temporal e os modelos propostos são modelos paramétricos (com um número finito de parâmetros). No segundo, a análise é conduzida no domínio de frequências e os modelos propostos são modelos não-paramétricos. Dentre os modelos paramétricos temos, por exemplo, os modelos ARIMA, que serão estudados neste curso nos próximos capítulos. No domínio de frequências temos a análise espectral, que tem inúmeras aplicações em ciências físicas e engenharia, principalmente na engenharia elétrica, e que consiste em decompor a série dada em componentes de frequências e onde a existência do espectro é a característica fundamental. Este tipo de análise não será estudado nestas notas de aulas, para detalhes o aluno deve consultar Watts and Jenkins (1968), Koopmans (1974), Morettin (1978) e Marple (1980).

3.2 Exemplos de séries temporais

Exemplo 3.1. Vamos supor que desejamos medir a temperatura média do ar, de um local, durante 24 horas, poderíamos obter um gráfico semelhante a figura abaixo:

Cada curva do gráfico é chamada de trajetória ou série temporal ou função amostral. No gráfico acima $Z_{(j)}(t)$ é o valor da temperatura no instante t , para a j -ésima trajetória (j -ésimo ponto de observação). Para cada t fixo, teremos os valores de uma variável aleatória $Z(t)$ que terá certa distribuição de probabilidade. Na realidade o que chamamos de série temporal, é uma parte de uma trajetória, dentre muitas que poderiam ter sido observadas. O parâmetro t pode ser função de algum outro parâmetro físico como por exemplo: espaço e volume.

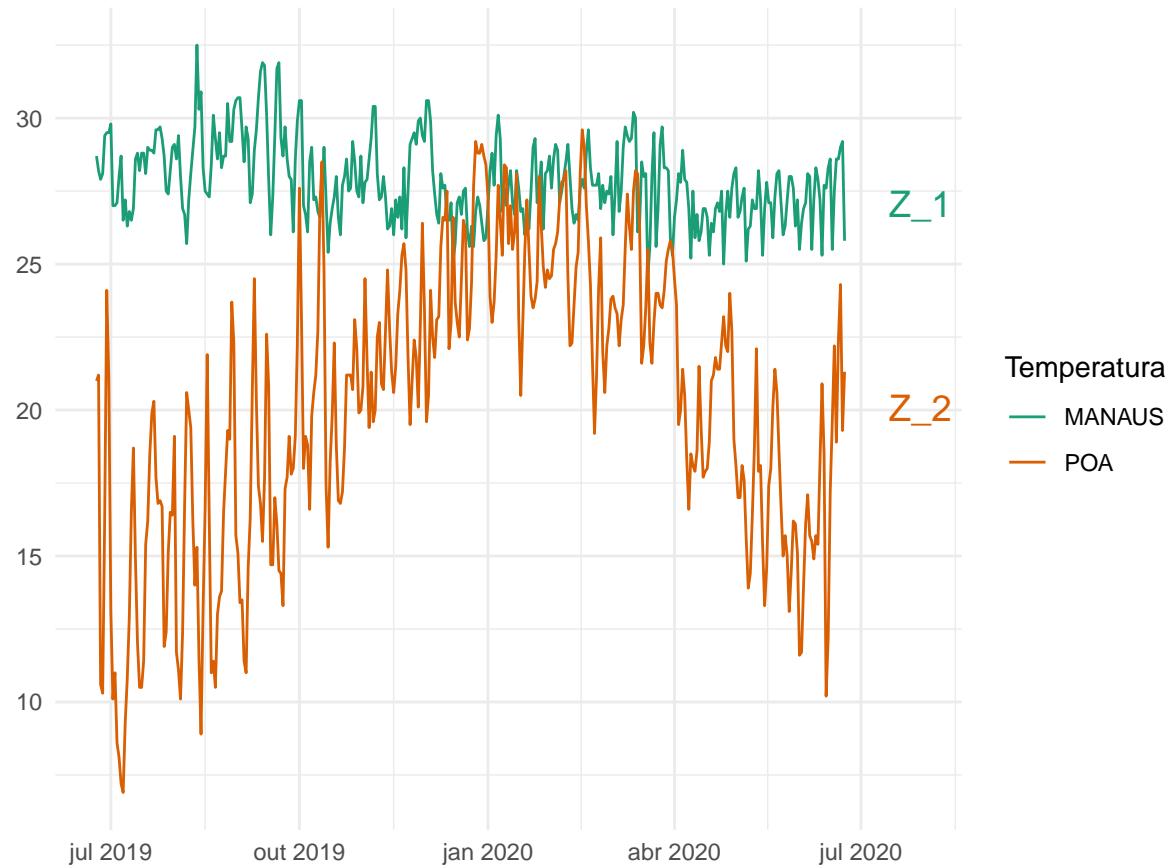


Figure 3.1: Temperaturas médias diárias nas cidades de Manaus (MA) e Porto Alegre (POA)

3.3 Regressão com dados de séries temporais

Nesta seção estudaremos modelos de regressão cujas variáveis são séries temporais. O interesse principal recai sobre as condições necessárias para que o estimador de MQO apresente boas propriedades.

3.3.1 Diferença entre dados de séries temporais e dados de corte transversal

A primeira diferença entre dados de séries temporais e dados de corte transversal é que uma série temporal tem uma ordenação temporal. Outra característica, é que não temos mais independência entre as observações, ou seja, não temos mais uma amostra aleatória de indivíduos. Logo, para estimar um modelo do tipo

$$y_t = \beta_0 + \beta_1 + \beta_2 x_{t1} + x_{t2} + \dots + \beta_k x_{tk} + u_t, \quad (3.1)$$

são necessárias novas suposições para que o estimador de MQO tenha boas propriedades.

3.3.2 Modelos de regressão de séries temporais

3.3.2.1 Modelos estáticos

Suponha que temos dados de séries temporais disponíveis para duas variáveis, digamos y e z , em que y_t e z_t são datadas contemporaneamente. Um modelo que relaciona y a z é:

$$y_t = \beta_0 + \beta_1 z_t + u_t, \quad t = 1, 2, \dots, n. \quad (3.2)$$

O nome “Modelo Estático” deriva do fato de relacionar as variáveis de forma contemporânea.

Exemplo 3.2. Um exemplo de modelo estático é a **curva de Phillips estática**, representada por:

$$inf_t = \beta_0 + \beta_1 desemp_t + u_t, \quad (3.3)$$

em que inf_t é a inflação anual e $desemp_t$ é a taxa de desemprego.

Este modelo é usado para estudar a relação de trocas contemporânea entre inf_t e $desemp_t$ pressupondo uma *tasa natural de desemprego* e expectativas inflacionárias constantes.

3.3.3 Modelos de defasagem distribuída finita

Em um *modelo de defasagem distribuída finita* (MDD) permite-se que uma ou mais variáveis afetem y com defasagens

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t, \quad (3.4)$$

que é um MDD de *ordem 2*. De modo mais geral, um modelo de defasagem distribuída de ordem q incluirá q defasagens de z .

Para interpretar os coeficientes em (3.4) suponha que z seja constante igual a c , em todos os períodos de tempo antes de t ($\dots, z_{t-2} = c, z_{t-1} = c$). Em t , z aumenta em uma unidade, ou seja, $z_t = c + 1$, e, em seguida, retorna ao seu nível anterior em $t + 1$, isto é, $z_{t+1} = c$.

Para enfatizar o efeito *ceteris paribus* de z sobre y , suponhamos que o termo de erro em cada período seja zero. Então,

$$\begin{aligned} y_{t-1} &= \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 c \\ y_t &= \alpha_0 + \delta_0(c+1) + \delta_1 c + \delta_2 c \\ y_{t+1} &= \alpha_0 + \delta_0 c + \delta_1(c+1) + \delta_2 c \\ y_{t+2} &= \alpha_0 + \delta_0 c + \delta_1 c + \delta_2(c+1) \\ y_{t+3} &= \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 c, \end{aligned}$$

e assim por diante. Das duas primeiras equações temos

$$y_t - y_{t-1} = \delta_0,$$

mostra que δ_0 é a mudança imediata em y em razão do aumento de uma unidade em z no tempo t . Denomina-se δ_0 como *propensão de impacto* ou *multiplicador de impacto*.

Da mesma forma,

$$\delta_1 = y_{t+1} - y_{t-1},$$

é a mudança em y após a mudança temporária e

$$\delta_2 = y_{t+2} - y_{t-1},$$

é a mudança em y dois períodos após a mudança. Em $t + 3$, y retornou ao seu nível inicial $y_{t+3} = y_{t-1}$. Isso ocorre porque presumimos que apenas duas defasagens de z aparecem em (3.4).

Quando traçamos um gráfico de δ_j como uma função de j obtemos a *distribuição de defasagem*, que resume o efeito dinâmico que um aumento temporário em z tem em y .

No entanto, o aumento em z pode ser permanente. Suponhamos que antes do tempo t z é constante igual a c , ou seja, $z_s = c$, para $s < t$, e no tempo t , z sofre um aumento permanente de uma unidade no tempo t , ou seja, $z_s = c + 1$ para $s \geq t$. Novamente, fazendo os erros iguais a zero, temos

$$\begin{aligned} y_{t-1} &= \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 c \\ y_t &= \alpha_0 + \delta_0(c+1) + \delta_1 c + \delta_2 c \\ y_{t+1} &= \alpha_0 + \delta_0(c+1) + \delta_1(c+1) + \delta_2 c \\ y_{t+2} &= \alpha_0 + \delta_0(c+1) + \delta_1(c+1) + \delta_2(c+1) \end{aligned}$$

e assim por diante. Com o aumento permanente em z , depois de um período y aumentou $\delta_0 + \delta_1$, e depois de dois períodos, y aumentou $\delta_0 + \delta_1 + \delta_2$. Isso mostra que a soma dos coeficientes de z atual e defazadas,

$$\delta_0 + \delta_1 + \delta_2 \tag{3.5}$$

é a *mudança de longo prazo* em y quando há um aumento permanente em z . A equação (3.5) é chamada *propensão de longo prazo (PLP)*.

A generalização para q defasagens é imediata.

3.4 Suposições para modelos com séries temporais

Nesta seção o objetivo é mostrar como as hipóteses clássicas devem ser alteradas para cobrir regressão de séries temporais.

3.4.1 Inexistência de viés do MQO

Para que as estimativas via MQO dos parâmetros de um modelo de regressão com séries temporais não sejam viesadas são necessárias as seguintes hipóteses:

Suposição TS.1: (*linearidade nos parâmetros*). O processo estocástico

$$\{(x_{t1}, x_{t2}, \dots, y_t) : t = 1, 2, \dots, n\}$$

segue o modelo linear:

$$y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_k x_{tk} + u_t,$$

em que $\{u_t : t = 1, 2, \dots, n\}$ é a sequência de erros ou perturbações.

Suposição TS.2: (*Inexistência de colineariedade Perfeita*). Na amostra, nenhuma das variáveis independentes é constante ou combinação linear perfeita das outras.

As hipóteses **TS.1** e **TS.2** são essencialmente as mesmas daquelas usadas no contexto de dados de cortes transversais.

Suposição TS.3: (*Média condicional zero ou exogeneidade estrita*). O termo de erro em qualquer dado período é não correlacionado com as variáveis explicativas em todos os períodos de tempo, ou seja $\mathbb{E}(u_t|X) = 0$, para $t = 1, 2, \dots, n$.

Analizando-se a hipótese **TS.3**, percebemos que ela difere da hipótese clássica. Observe que a hipótese **TS.3** exige que o erro no tempo t , u_t seja não correlacionado com cada variável explicativa em *todos* os períodos de tempo.

Se em termos de média condicional, temos somente a condição de não correlação somente no tempo t , da forma

$$\mathbb{E}(u_t|x_{1t}, \dots, x_{tk}) = \mathbb{E}(u_t|X_t) = 0, \quad (3.6)$$

diz-se que vale a *exogeneidade contemporânea* das variáveis explicativas. Exogeneidade contemporânea só será suficiente em grandes amostras. (3.6)

A hipótese **TS.3** é muito forte e muitas vezes não verificada. Nos seguintes exemplos podemos ver como ela pode ser verificada na prática.

Exemplo 3.3. Suponha que a taxa de homicídios (hom_i_t) em uma cidade em termos do número de policiais per capita ($polpc_t$)

$$hom_i_t = \beta_0 + \beta_1 polpc_t + u_t.$$

O termo de erro u precisaria ser não correlacionados com os valores atuais, os valores passados e futuros de $polpc_t$. Podemos aceitar que u não é correlacionado com valores corrente e valores passados do regressor. Mas é evidente que um aumento em u hoje, provavelmente, levará a políticas que tentem aumentar $polpc_t$ no futuro. Logo **TS.3** falha.

Quando u é correlacionado com o passado dos regressores, podemos resolver o problema incluindo defasagens dos regressores e utilizando um modelo de defasagem distribuída. Mas não podemos ter, de forma alguma, a influência de u no futuro dos regressores.

Teorema 3.1. *Sob as Hipóteses **ST.1**, **ST.2** e **ST.3** os estimadores de MQO são não viesados condicionados a X e, portanto, também incondicionalmente:*

$$\mathbb{E}(\hat{\beta}_j) = \beta_j, \quad j = 1, \dots, k.$$

3.4.2 Variância dos estimadores MQO

É necessário mais duas hipóteses para completar o conjunto de hipóteses de Gauss-Markov para regressões de séries temporais. A primeira delas é familiar da análise de corte transversal.

Suposição TS.4: (*Homoscedasticidade*). Condisional a X , a variância de u_t é a mesma para todo t :

$$Var(u_t|X) = Var(u_t) = \sigma^2,$$

para $t = 1, 2, \dots, n$.

Suposição TS.5: (*Inexistência de Correlação Serial*). Condisional a X , os erros em dois períodos de tempos diferentes são não correlacionados:

$$Corr(u_t, u_s|X) = 0,$$

para todo $t \neq s$.

Com este conjunto de condições podemos enunciar o teorema de Gauss-Markov no contexto de séries temporais.

Teorema 3.2 (Teorema de Gauss-Markov). *Sob as Hipóteses **ST.1** a **ST.5** os estimadores de MQO são os melhores estimadores lineares não viesados condicionais a X , ou seja, são **BLUE**.*

3.4.3 Inferência sob as hipóteses do modelo linear clássico

Para que sejam válidos os testes t , F e outros testes estatísticos baseadas nos erros padrões é necessário adicionar mais uma hipótese a respeito da distribuição dos erros. Esta hipótese é análoga à hipótese de normalidade usada para análise de corte transversal.

Suposição TS.6: (Normalidade). Os erros u_t são independentes de X e são i.i.d. com distribuição normal com média zero e variância σ^2

$$u_t \sim \mathcal{N}(0, \sigma^2),$$

para $t = 1, 2, \dots, n$.

Teorema 3.3. Sob as hipóteses **TS.1** a **TS.6**, as hipóteses do modelo linear clássico para séries temporais, os estimadores MQO são normalmente distribuídos, condicional em X . Além disso, a estatística t tem uma distribuição t , e cada estatística F tem uma distribuição F .

3.4.4 Tendência

Quando trabalhamos com séries temporais é necessário saber reconhecer se estas séries contém uma tendência temporal. Ignorar o fato de que duas séries temporais podem ser correlacionadas somente porque ambas estão apresentando uma mesma tendência ao longo do tempo, em vez de uma relação causal, pode levar a conclusões errôneas e a possibilidade de uma regressão espúria. Vejamos o exemplo de uma série temporal com tendência temporal:

```
#library(readxl)
#library(ggplot2)
# Dados baixados em 25/06/2020 de
# https://seriesestatisticas.ibge.gov.br/series.aspx?vcodigo=PRECO415&t=custo-medio-m2

custo_medio <- read_excel("Dados/custo_medio_m2.xls")
ggplot(custo_medio, aes(x = data, y = cm)) +
  geom_line() + theme_light() + labs(x = "", y = "Custo médio")
```

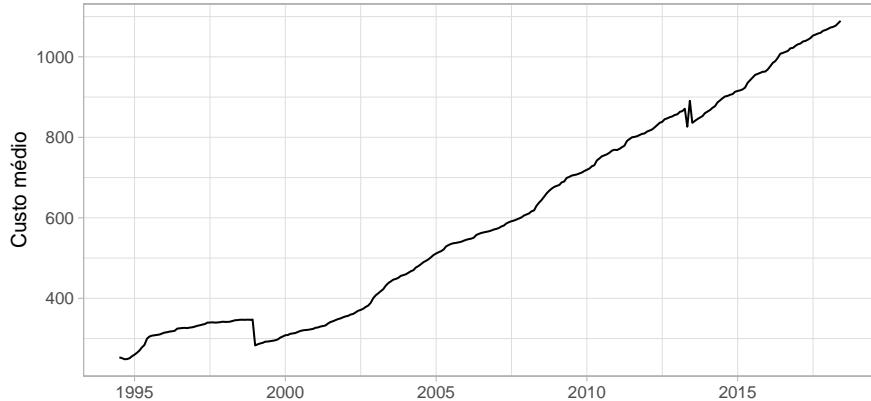


Figure 3.2: Custo médio do m^2 no Brasil a partir do plano real

Um modelo que captura tendência temporal é:

$$y_t = \alpha_0 + \alpha_1 t + e_t, \quad t = 1, 2, \dots, \quad (3.7)$$

em que assume-se que $\{e_t\}$ é i.i.d. com $\mathbb{E}(e_t) = 0$ e $\text{Var}(e_t) = \sigma^2$. Observe que o parâmetro α_1 multiplica o tempo, resultando em uma *tendência temporal linear*. Assim, α_1 mede a mudança em y_t , de um período para o próximo, motivado pela passagem do tempo, mantendo-se todos os outros fatores fixos.

Outros modelos podem ser usados para capturar tendências temporais, dependendo da situação. No modelo em que o logaritmo natural de y_t (presumindo que $y_t > 0$) apresenta uma tendência temporal linear,

$$\log(y_t) = \beta_0 + \beta_1 t + e_t, \quad t = 1, 2, \dots, \quad (3.8)$$

diz-se que y_t tem uma **tendência exponencial**.

Outra possibilidade é que em vez de uma tendência temporal linear, poderíamos ter uma **tendência temporal quadrática**,

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + e_t, \quad t = 1, 2, \dots. \quad (3.9)$$

3.4.5 Usando variáveis de tendência na análise de regressão

Suponha que existam dois fatores observados, x_{t1} e x_{t2} que afetam y_t . Além disso, existem fatores não observados que estão sistematicamente crescendo ou decrescendo ao longo do tempo. Um modelo que captura isso é:

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 t + u_t. \quad (3.10)$$

Permitindo uma tendência temporal no modelo, reconhece-se que y_t pode estar crescendo ou decrescendo ao longo do tempo por razões essencialmente não relacionadas a x_{t1} e x_{t2} .

A omissão da variável t pode levar ao viés por omissão de variável, especialmente se x_{t1} ou x_{t2} apresentarem algum tipo de tendência, pois elas podem ser altamente correlacionadas com t .

Adicionando um termo de tendência linear em um modelo de regressão é a mesma coisa que usar série “destendenciada” numa regressão. Os estimadores β_1 e β_2 do modelo (3.10) podem ser obtidos através de um procedimento de “remoção da tendência temporal” das séries originais:

Destendenciar uma série envolve regredir cada variável do modelo em t e uma constante (no caso de (3.10), regredir y_t , x_{t1} e x_{t2} contra t e uma constante).

Os resíduos destas regressões, \ddot{y}_t , \ddot{x}_{t1} e \ddot{x}_{t2} , constituem uma série temporal sem tendência.

Em seguida, realizar a regressão com variáveis retificadas,

$$\ddot{y}_t = \delta_1 \ddot{x}_{t1} + \delta_2 \ddot{x}_{t2} + v, \quad (3.11)$$

(não precisa intercepto, será igual a 0). As estimativas via MQO, $\hat{\delta}_1$ e $\hat{\delta}_2$ serão iguais as estimativas $\hat{\beta}_1$ e $\hat{\beta}_2$ da regressão (3.10).

3.5 Sazonalidade

Sazonalidade ocorre quando uma série exibe comportamentos semelhantes em determinados períodos. Um exemplo é o **Índice de produção física industrial de bens intermediários*. O objetivo do índice é servir como uma medida aproximada da evolução de curto prazo do valor adicionado da indústria, dado um determinado período de referência (base: média de 2012).

```
#library(readxl)
#library(plotly)
#Fonte: "IBGE - Pesquisa Industrial Mensal - Produção Física"
prodind <- read_excel("Dados/ProducaoIndustrial.xls")
ggplot(prodind,aes(x>Data, y=intermediarios))+
  geom_line(col="#D95F02")+
  labs(x = "",y="Índice")+
  theme_minimal()
```

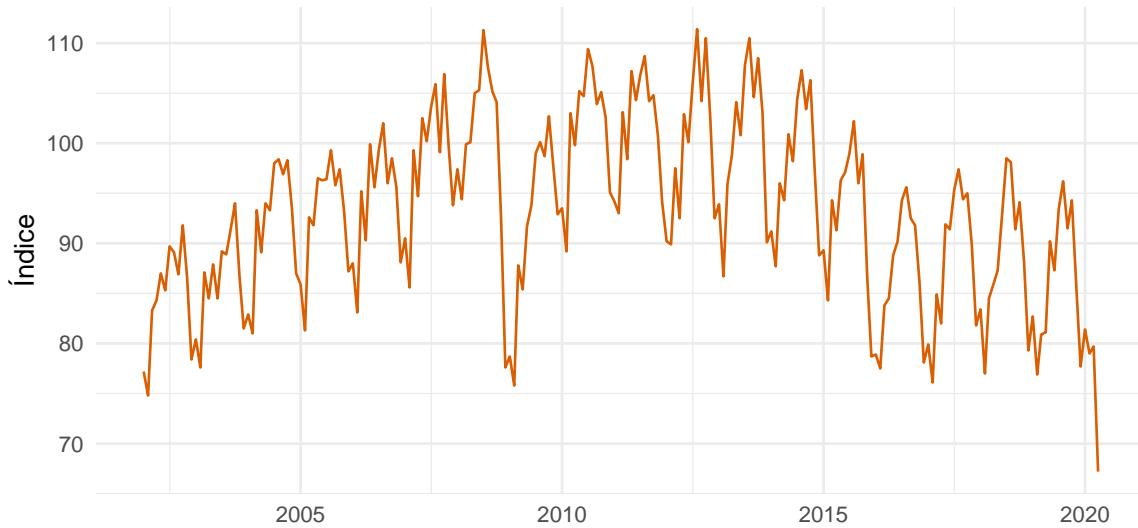


Figure 3.3: Produção física industrial de Bens Intermediários

É comum que as séries de dados mensais e trimestrais exibam padrões sazonais, mas isso não é uma regra. Por exemplo, não existe padrão sazonal observável nas taxas de juros ou de inflação. Além disso, séries que exibem padrões sazonais são **ajustadas sazonalmente** antes de serem informadas para o público.

Uma série ajustada sazonalmente é a série que teve os fatores sazonais removidos. Existem vários métodos para isso. Um dos métodos mais simples é incluir um conjunto de variáveis dummies sazonais. Seja o seguinte modelo para dados mensais:

$$y_t = \beta_0 + \delta_1 fev_t + \delta_2 mar_t + \cdots + \delta_{11} dez_t + \beta_1 x_{t1} + \cdots + \beta_k x_{tk} + u_t. \quad (3.12)$$

em que $fev_t, mar_t, \dots, dez_t$ são variáveis dummy indicando se o período de tempo t corresponde ao mês apropriado. Nesta formulação, janeiro é o mês-base e β_0 seu intercepto. Se colocarmos janeiro no modelo e um intercepto, teremos um problema de multicolineariedade.

Se não existir sazonalidade em y_t , dado que controlamos os regressores x_{tj} , então os coeficientes $\delta_1; \dots; \delta_{11}$ devem ser todos iguais a zero, o que pode ser testado através de um teste F .

Considere o modelo (3.12), para $k = 2$, ou seja 2 regressores. Podemos obter os seus estimadores, $\hat{\beta}_1$ e $\hat{\beta}_2$, através do seguinte procedimento:

Regrida a variável dependente, e cada um dos regressores, separadamente, contra uma constante e as dummies mensais e guarde os resíduos, digamos \ddot{y}_t , \ddot{x}_{t1} e \ddot{x}_{t2} .

Por exemplo,

$$\ddot{y}_t = y_t - \hat{\alpha}_0 - \hat{\alpha}_1 fev_t - \hat{\alpha}_2 mar_t - \cdots - \hat{\alpha}_{11} dez_t.$$

Este é o método para dessazonalizar uma série temporal mensal.

Roda a regressão de \ddot{y}_t contra \ddot{x}_{t1} e \ddot{x}_{t2} sem as dummies mensais.

3.5.1 Processos de covariância estacionária

Um processo estocástico é covariância estacionária se $\mathbb{E}(x_t)$ é constante, $\text{Var}(x_t)$ é constante e para qualquer $t, h \geq 1$, $\text{Cov}(x_t, x_{t+h})$ depende apenas em h , e não em t . Mais adiante abordaremos essa definição com maior profundidade.

3.5.2 Processos fracamente dependentes

Uma série temporal estacionária é fracamente dependente se x_t e x_{t+h} são “quase independentes”, quando h aumenta.

Se, para um processo de covariância estacionária $\text{Cor}(x_t, x_{t+h}) \rightarrow 0$ quando $h \rightarrow \infty$, dizemos que este processo de covariância estacionária é fracamente dependente.

Essa definição é necessária para usar *Leis dos Grandes Números e Teorema Central do Limite*.

Exemplo 3.4. MA(1) pg 356 Wooldridge.

Exemplo 3.5. Exemplo: AR(1) pg 356 Wooldridge.

3.6 Exercícios

Exercício 3.1. Sobre regressão com séries temporais responda:

- Quais as principais diferenças entre dados transversais e séries temporais?
- Explique o que é exogeneidade contemporânea e exogeneidade estrita.
- Comente sobre a diferença entre homocedasticidade e correlação serial.
- A suposição de normalidade dos erros é necessária para se obter estimadores consistentes via MQO? Qual é o objetivo ao se fazer uma suposição para distribuição dos erros?

Exercício 3.2 (anpec-2010). Considere o modelo de regressão linear múltipla com regressores estocásticos $y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t$, no qual ε_t não é autocorrelacionado e tem média e variância condicionais a x_{1t} e x_{2t} iguais a zero e s^2 , respectivamente. Por simplicidade, suponha que as variáveis são expressas como desvios com relação às respectivas médias.

Responda:

- Se $\beta_2 = 0$ e incluirmos x_{2t} na regressão, o estimador de mínimos quadrados ordinários de β_1 será viesado?
- Se não conseguirmos observar x_{1t} , mas apenas $x_{1t}^* = x_{1t} + u_t$, em que u_t é um erro de medida, e se substituirmos x_{1t} por x_{1t}^* na regressão, o estimador de mínimos quadrados ordinários de β_1 ainda assim será consistente?
- Se $x_{2t} = y_{t-1}$ e relaxarmos a hipótese de que os erros ε_t 's não são autocorrelacionados, o estimador de mínimos quadrados ordinários de β_2 será consistente, porém não será eficiente?
- Seja c uma constante diferente de zero. Defina $\tilde{y} = cy_t$, $\tilde{x}_{1t} = cx_{1t}$ e $\tilde{x}_{2t} = cx_{2t}$. Os estimadores de mínimos quadrados ordinários (MQO) em uma regressão de \tilde{y} contra \tilde{x}_{1t} e \tilde{x}_{2t} coincidem com os estimadores de MQO em uma regressão de y_t contra x_{1t} e x_{2t} ?
- A variância do estimador de mínimos quadrados ordinários diverge para infinito à medida que a correlação entre x_{1t} e x_{2t} aproxima-se de 1;
- Denote por $\hat{\varepsilon}_t$ o resíduo da regressão de mínimos quadrados ordinários. A hipótese de que o erro é correlacionado com x_{1t} pode ser testada utilizando a estatística $\frac{1}{T} \sum_{i=1}^T x_{1i} \hat{\varepsilon}_i$?

Exercício 3.3. Em uma equação de dados anuais, supondo que

$$jur_t = 1,6 + 0,48inf_t - 0,15inf_{t-1} + 0,32inf_{t-2} + u_t,$$

em que jur é a taxa de juros e inf é a taxa de inflação.

- Supondo válida a hipótese de exogeneidade estrita, como deve ter sido estimado o modelo acima? Justifique?
- Qual é o efeito de curto prazo (propensão de impacto) da taxa de inflação sobre a taxa juros? Qual é o efeito de longo prazo da taxa de inflação sobre a taxa de juros?

Exercício 3.4. Considere uma série temporal de 10 anos contendo PIB (em R\$) e número de homicídios (em unidades) em um determinado país. O primeiro modelo estimado foi $pib_t = \beta_0 + \beta_1 homic_t + u_t$. Os resultados da estimação se encontram na tabela 1. Um segundo modelo foi $pib_t = \beta_0 + \beta_1 homic_t + \beta_2 t + u_t$, em que t é um termo de tendência. Os resultados da estimação desse modelo se encontram na tabela 2:

Tabela 1	Estimate	Std. Error	t-value	Pr(>)
(Intercept)	-3461194.26	314948.06	-10.99	0.00
homic	102.63	6.12	16.76	0.00

Tabela 2	Estimate	Std. Error	t-value	Pr(>)
(Intercept)	5564710.45	2539866.04	2.19	0.06
homic	-123.64	63.59	-1.94	0.09
t	423054.01	118647.95	3.57	0.01

- a) O coeficiente de $homic$ é significativo no primeiro modelo a 5% de significância? Interprete o valor desse coeficiente.
- b) O coeficiente de $homic$ é significativo no segundo modelo a 5% de significância? Interprete o valor desse coeficiente.
- c) O coeficiente de t é significativo no segundo modelo a 5% de significância? Interprete o valor desse coeficiente.
- d) Explique o resultado (surpreendente) encontrado no primeiro modelo, ressaltando a importância do procedimento adotado no segundo modelo.

Exercício 3.5. Considere uma série do PIB brasileiro com início no primeiro trimestre 1996 e fim no segundo bimestre de 2010. Essa série foi decomposta em sua tendência (t) e variáveis dummy para a sazonalidade, em que $S_i = 1$, se a observação pertence ao trimestre i e $S_i = 0$, caso contrário.

a) Se tentarmos estimar o modelo $pib_t = \beta_0 + \beta_1 S_1 + \beta_2 S_2 + \beta_3 S_3 + \beta_4 S_4 + \gamma t + u_t$, qual problema encontraremos? Explique porque isso ocorre.

- b) No modelo $pib_t = \beta_1 S_1 + \beta_2 S_2 + \beta_3 S_3 + \beta_4 S_4 + \gamma t + u_t$, o que mede cada um dos β 's?
- c) No modelo $pib_t = \beta_0 + \beta_2 S_2 + \beta_3 S_3 + \beta_4 S_4 + \gamma t + u_t$, o que mede β_2 ?
- d) No modelo $pib_t = \beta_0 + \gamma t + \beta_2 S_2 + \beta_3 S_3 + \beta_4 S_4 + u_t$, foi estimado e apresentou a seguinte tabela ANOVA (Tabela 3). Faça um teste F para a hipótese nula de que não há sazonalidade. Use

$$\alpha = 5\%$$

Tabela 3	Df	Sum Sq	Mean Sq
t	1	2287298699531.79	2287298699531.79
s2	1	1216754395.49	1216754395.49
s3	1	31129772.60	31129772.60
s4	1	5037536508.88	5037536508.88
Residuals	53	106216397798.70	2004082977.33

Exercício 3.6. Considere o modelo $y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t$.

- a) Por que devemos considerar a possibilidade de multicolinearidade nesse modelo?
- b) Reparametrize o modelo de modo a isolar o efeito de longo prazo como coeficiente da variável z_t .
- c) Qual o benefício dessa reparametrização se estivermos interessados em testar a significância do efeito de LP da z sobre y ?

Exercício 3.7. Considere o seguinte modelo estático $crime_t = \beta_0 + \beta_1 + pol_t + u_t$, em que $crime_t$ é um índice de criminalidade no período t e pol_t é o número de policiais em t .

- a) Supondo que pol seja estritamente exógeno na equação, como você estimaria β_0 e β_1 . Quais as propriedades do estimador proposto em termos de viés e consistência?
- b) Suponha agora que o número de policiais em t seja definido em função do índice de criminalidade do período anterior. A hipótese de exogeneidade estrita continua válida? Justifique.

Exercício 3.8. Um modelo de ajustamento parcial é dado por:

$$y_t^* = \beta_0 + \beta_1 x_t + e_t$$

$$y_t - y_{t-1} = \lambda(y_t^* - y_{t-1}),$$

em que y_t^* é o nível desejável ou ótimo de y , e y_t é o nível efetivo (observado). Por exemplo, y_t^* é o crescimento desejável nos estoques de uma firma e x_t é o crescimento de vendas da firma. O parâmetro λ mede a velocidade do ajustamento e satisfaz $0 < \lambda < 1$.

- a) Insira a primeira equação na segunda equação e mostre que podemos escrever $y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 x_t + u_t$. Quem são os α s em termos dos β s e λ ? Quem é u_t em termos de e_t ?
- b) Supondo que $\mathbb{E}(e_t | x_t, y_{t-1}) = 0$ e todas as séries sejam fracamente dependentes, como você estimaria os α s? É consistente? Justifique sua resposta. O estimador proposto é viciado?
- c) Seja $\hat{\alpha}_1 = 0,7$ e $\hat{\alpha}_2 = 0,2$.

(i) Qual o coeficiente de ajustamento estimado?

(ii) Qual o efeito de CP (curto prazo) de um crescimento das vendas da firma sobre o crescimento de estoques?

(iii) Qual é o efeito de LP (longo prazo)?

Exercício 3.9. Imagine o seguinte modelo: $Y_t = \beta_0 + \beta_1 X_t + u_t$, onde Y é a demanda por moeda, X^* é a taxa de juros esperada no longo prazo e u é um termo de erro clássico, não correlacionado com X^* . Como a variável de expectativa X^* não é diretamente observável, proporemos a seguinte hipótese para formação de expectativas (adaptativas): $X_t^* - X_{t-1}^* = \gamma(X_t - X_{t-1})$, em que γ , tal que $0 < \gamma < 1$, é conhecido como coeficiente de expectativas.

a) Mostre que podemos escrever esse modelo como $Y_t = \alpha_0 + \alpha_1 X_t + \alpha_2 Y_{t-1} + v_t$. Quem são os α s em termos dos β s e γ ? Quem é v_t em termos de u_t ?

b) O que podemos dizer a respeito dos estimadores de MQO nesse caso? Justifique.

c) Imagine que no modelo original u_t siga o esquema auto-regressivo de primeira ordem, i.e., $u_t = \rho u_{t-1} + \varepsilon_t$, em que ρ é o coeficiente de autocorrelação e onde ε_t satisfaz as premissas clássicas. Se $\rho = \lambda$, como você estimaria o modelo? Justifique.

d) As estimativas obtidas no item anterior são não-viciadas? Consistentes? Justifique sua resposta.

Exercício 3.10. Seja o processo $y_t = e_t + \alpha_1 e_{t-1}$, em que $e_t \sim iid(0, \sigma^2)$.

a) Calcule $\mathbb{E}(y_t)$, $\text{Var}(y_t)$ e $\text{Cov}(y_t, y_{t-h})$, $h = 1, 2, 3, \dots$. O processo y_t é de covariância estacionária?

b) Calcule as autocorrelações de primeira ordem e de segunda ordem para esse processo. Podemos dizer que o processo é fracamente dependente? Justifique.

c) Faça o correlograma (gráfico da função de autocorrelação em função das defasagens) para esse processo.

Exercício 3.11. Seja o processo $y_t = c + \rho y_{t-1} + e_t$, em que $e_t \sim iid(0, \sigma^2)$.

a) Qual é a condição de estabilidade para esse processo? Calcule $\mathbb{E}(y_t)$ e $\text{Var}(y_t)$ considerando válida a condição de estabilidade.

b) Para o processo y_t acima temos que $\text{Cov}(y_t, y_{t-h}) = \frac{\rho^h \sigma^2}{1-\rho^2}$, $h = 1, 2, 3, \dots$. O processo y_t é de covariância estacionária? Justifique.

c) Calcule a autocorrelação de ordem h para o processo y_t . Faça o correlograma até quatro defasagens para esse processo considerando $\rho = 0, 5$.

Chapter 4

Séries Temporais

O estudo de séries temporais tem por objetivos principais definir o processo gerador de dados, fazer previsões futuras da série, identificar ciclos, tendências e/ou sazonalidades de forma que a decisão que envolve as variáveis em questão seja a mais acurada possível.

4.1 Objetivos

Na seção 4.1

Dada uma série temporal $\{Z(t_1), \dots, Z(t_N)\}$, observada nos instantes t_1, \dots, t_N , podemos estar interessados em:

- i) Investigar o mecanismo gerador da série temporal;
- ii) Fazer previsões de valores futuros da série; podendo ser a curto ou longo prazo;
- iii) Descrever apenas o comportamento da série através de gráficos;
- iv) Procurar periodicidades relevantes nos dados. Em todos estes casos podemos construir modelos probabilísticos ou estocásticos, tanto no domínio do tempo como no domínio da freqüência, por exemplo: um sinal aleatório com frequência medida em Hz. Devemos construir modelos simples e com menor número de parâmetros possíveis.

4.2 Séries temporais: definição formal

Neste capítulo vamos descrever os conceitos básicos utilizados dentro da teoria dos modelos de séries temporais. Inicialmente vamos introduzir os conceitos de processos estocásticos, média e função de covariância, processo estacionário e função de autocorrelação.

4.2.1 Processos estocásticos

Seja T um conjunto arbitrário de índices. Um `\textcolor{blue}{\{processo estocástico\}}` é uma família de variáveis aleatórias $\{Z_t\}_{t \in T}$ definidas num mesmo espaço de probabilidades, que denotaremos genericamente por (Ω, \mathcal{A}, P) . O conjunto de índices T pode ser o conjunto dos números inteiros $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$, dos naturais $\mathbb{N} = \{1, 2, \dots\}$ ou o conjunto dos números reais \mathbb{R} . Observe ainda que, para cada $t \in T$, Z_t é uma variável aleatória definida sobre Ω , sendo assim de fato uma função de dois argumentos, do índice $t \in T$ e do ponto $\omega \in \Omega$ que determina o valor do processo no tempo t dado por $Z_t(\omega)$.

Uma série temporal, do ponto de vista teórico, nada mais é do que um processo estocástico para o qual o índice T é \mathbb{Z} ou um subconjunto deste. Do ponto de vista prático porém, uma série temporal é um conjunto de dados indexados no tempo. Esta dualidade de nomenclatura será utilizado em todo o trabalho. Invariavelmente, letras maiúsculas, como Z_1, Z_2, \dots denotarão a série temporal do ponto de vista teórico, isto é, como variáveis aleatórias em um processo estocástico indexado pelo tempo, enquanto letras minúsculas como z_1, z_2, \dots denotarão a série temporal do ponto de vista prático, isto é, como uma observação das variáveis aleatórias que compõem o processo estocástico.

Assim, se do ponto de vista teórico temos a série temporal $\{Z_t\}_{t \in \mathbb{Z}}$, um processo estocástico indexado pelo tempo, uma série temporal do ponto de vista prático significa uma realização z_1, \dots, z_n do processo $\{Z_t\}$, observados nos tempos $t = 1, \dots, n$. Neste caso, observamos $z_1 = Z_1(\omega), \dots, z_n = Z_n(\omega)$, para um determinado $\omega \in \Omega$ fixo. Embora existam maneiras mais formais e precisas de definir uma série temporal, o ponto de vista aqui adotado, embora aparentemente ambíguo na nomenclatura, servirá bem a nossos propósitos sem causar confusões.

Chamamos atenção ainda que existem condições para que um processo estocástico exista. Estes resultados dependem de uma discussão bastante técnica, bem além das intenções de nossa exposição.

4.2.2 Especificação de um processo estocástico

Sejam t_1, t_2, \dots, t_n elementos quaisquer de T e consideremos

$$F(Z_1, \dots, Z_n; t_1, \dots, t_n) = P\{Z(t_1) \leq z_1, \dots, Z(t_n) \leq z_n\} \quad (4.1)$$

então, o processo estocástico $Z = \{Z(t), t \in T\}$ estará especificado se as distribuições finito-dimensionais de (4.1), são conhecidas para todo $n \geq 1$. Contudo, em termos práticos, não conhecemos todas essas distribuições finito-dimensionais. Estudaremos então certas características associadas a (4.1) e que sejam simples de calcular e interpretar. Uma maneira de especificar o processo Z seria determinar todos os produtos dos momentos, ou seja,

$$\mu(r_1, \dots, r_n; t_1, \dots, t_n) = \mathbb{E} Z^{r_1}(t_1) \dots Z^{r_n}(t_n) \quad (4.2)$$

ou

$$\mu(\mathbf{r}, \mathbf{t}) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} Z_1^{r_1} \dots Z_n^{r_n} f(z_1, \dots, z_n; t_1, \dots, t_n) dz_1 \dots dz_n \quad (4.3)$$

em que $f(\mathbf{Z}, \mathbf{t})$ é a função de densidade de $F(\mathbf{Z}, \mathbf{t})$. Porém o que vai nos interessar são os momentos de baixa ordem, ou seja, os chamados processos estacionários de segunda ordem. Consideraremos somente os momentos de primeira e segunda ordem, que serão apresentados a seguir.

4.3 Médias e covariâncias

Naturalmente, quando estamos trabalhando com um processo estocástico, cada variável aleatória que o compõe possui sua própria distribuição, assim como sua própria massa/densidade de probabilidade e sua própria média/variação. Para um processo estocástico $\{Z_t\}_{t \in \mathbb{Z}}$ definimos, para cada $t \in \mathbb{Z}$, a \textcolor{blue}{função média}, μ_t e a \textcolor{blue}{função variância} σ_t^2 respectivamente por

$$\mu_t = \mathbb{E}(Z_t) \quad \text{e} \quad \sigma_t^2 = \text{Var}(Z_t), \quad (4.4)$$

desde que as esperanças envolvidas existam. Chamamos a atenção de que embora as esperanças e variações de um processo estocástico existam, estas podem ser infinitas. Este fato trás diversos problemas técnicos na análise de séries temporais e requerem técnicas avançadas de análise que estão fora do escopo deste trabalho. Por este motivo, neste trabalho assumiremos tacitamente que todos os processos estocásticos e variáveis aleatórias possuem esperança e variação finitas.

Outra estrutura importante relacionada a um processo estocástico é o que chamamos de estrutura de dependência do processo. Dependência entre variáveis aleatórias pode ser definida de diversas maneiras diferentes. Neste trabalho estamos especialmente interessados na estrutura de dependência relacionadas com a covariância e a correlação entre as variáveis do processo. Observe que num processo estocástico podemos definir a covariância e a correlação entre quaisquer pares Z_i e Z_j de variáveis, para $i, j \in \mathbb{Z}$. No caso de processos, estas funções recebem o prefixo *auto* para enfatizar o fato de que as covariâncias/correlações estão sendo calculadas entre as variáveis do processo. Definimos a \textcolor{blue}{função de autocovariância}, abreviada \textcolor{blue}{FACV}, como

$$\gamma_Z(t, s) = \text{Cov}(Z_t, Z_s) = E[(Z_t - \mu_t)(Z_s - \mu_s)] = \mathbb{E}(Z_t Z_s) - \mu_t \mu_s, \quad \text{para } t, s \in \mathbb{Z}. \quad (4.5)$$

Analogamente, definimos a \textcolor{blue}{função de autocorrelação}, abreviada \textcolor{blue}{FAC}, por

$$\rho_Z(t, s) = \text{Cor}(Z_t, Z_s) = \frac{\text{Cov}(Z_t, Z_s)}{\sqrt{\text{Var}(Z_t)\text{Var}(Z_s)}} = \frac{\gamma(t, s)}{\sqrt{\gamma(t, t)\gamma(s, s)}}. \quad (4.6)$$

O subscrito “ Z ” nas definições acima são utilizados para reforçar à qual processo estamos nos referindo. Porém, quando não houver perigo de confusão, podemos eliminar a referência ao processo associado e escrever simplesmente $\gamma(t, s)$ e $\rho(t, s)$.

Observe que, em princípio, as funções $\gamma(t, s)$ e $\rho(t, s)$ dependem tanto de t quanto de s . Nos casos em que isto acontece, qualquer tipo de inferência baseada em autocovariâncias/autocorrelações se torna impossível sem tomarmos medidas para tornar esta estrutura de dependência mais simples. Algumas técnicas relevantes para isso serão estudadas adiante. De qualquer forma, a teoria clássica de séries temporais lida com casos em que essas quantidades possuem uma dependência temporal simplificada, permitindo o seu estudo. Processos com estas características são de grande importância e serão estudados em detalhes mais adiante. Isto por que, do ponto de vista matemático, tal estrutura é conveniente e permite um tratamento rigoroso e aprofundado da teoria enquanto que do ponto de vista prático, é de fácil percepção, permite a modelagem, inferência, previsão e outros aspectos aplicados relevantes de maneira simples e rápida. Tudo isso contribuiu para a difusão de métodos baseado em autocovariâncias/autocorrelações.

4.3.1 Propriedades importantes

As seguintes propriedades da função de autocovariância e autocorrelação são análogas às da covariância e correlação ordinárias, mas merecem destaque. Para todo $t, s \in \mathbb{Z}$, com $t \neq s$,

- 1) $\gamma(t, t) = \text{Var}(Z_t)$, $\rho(t, t) = 1$;
- 2) $\gamma(t, s) = \gamma(s, t)$, $\rho(t, s) = \rho(s, t)$.
- 3) $|\gamma(t, s)| \leq \sqrt{\gamma(t, t)\gamma(s, s)}$, $-1 \leq \rho(t, s) \leq 1$.

A propriedade 3 em particular mostra que a covariância entre duas variáveis está bem definida caso estas tenham variância finita.

Como sabemos a correlação é uma medida da dependência linear entre duas variáveis. Se $\text{Cor}(X, Y) = \pm 1$, isto significa que existem constantes β_0 e β_1 tais que $Y = \beta_0 + \beta_1 X$. Ou seja, uma variável é exatamente uma função linear da outra. Valores próximos de ± 1 indicam forte dependência (linear) e valores próximos de 0 indicam fraca dependência (linear). Se $\rho(t, s) = 0$, Z_t e Z_s são ditas não-correlacionadas, mas note que isso não quer dizer que elas são necessariamente independentes. Agora, se Z_t e Z_s são independentes, então $\rho(t, s) = 0$. Por fim, obviamente $\text{Cov}(Z_t, Z_s) = 0$ se, e somente se, Z_t e Z_s são não-correlacionadas.

Para analisar as propriedades da covariância de vários modelos de séries temporais, o seguinte resultado será utilizado: se c_1, c_2, \dots, c_m e d_1, d_2, \dots, d_n são constantes reais e t_1, t_2, \dots, t_m e s_1, s_2, \dots, s_n são índices temporais, então

$$\text{Cov}\left(\sum_{i=1}^m c_i Z_{t_i}, \sum_{j=1}^n d_j Z_{s_j}\right) = \sum_{i=1}^m \sum_{j=1}^n c_i d_j \text{Cov}(Z_{t_i}, Z_{s_j}) \quad (4.7)$$

podemos dizer que, a covariância entre duas combinações lineares é a soma de todas as covariâncias entre termos de suas combinações lineares. Esta expressão pode ser verificada utilizando as propriedades de esperança e covariância. Como caso especial, podemos obter o seguinte resultado

$$\text{Var}\left(\sum_{i=1}^n c_i Z_{t_i}\right) = \sum_{i=1}^n c_i^2 \text{Var}(Z_{t_i}) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n c_i c_j \text{Cov}(Z_{t_i}, Z_{t_j}). \quad (4.8)$$

4.4 Estacionariedade

Nesta seção definiremos o fundamental conceito da estacionariedade de uma série temporal. Existem diversas maneiras de se definir o conceito de estacionariedade, de acordo com as técnicas que se pretendem utilizar na análise das séries temporais. Em poucas palavras, uma série temporal é estacionária quando, com o passar do tempo, a série se desenvolve

aleatoriamente em torno de uma média constante, refletindo alguma forma de equilíbrio estável. A ideia é de que uma série temporal estacionária Y tende a “flutuar” aleatoriamente ao redor de uma média constante. A Figura 4.1 apresenta duas séries estacionárias.

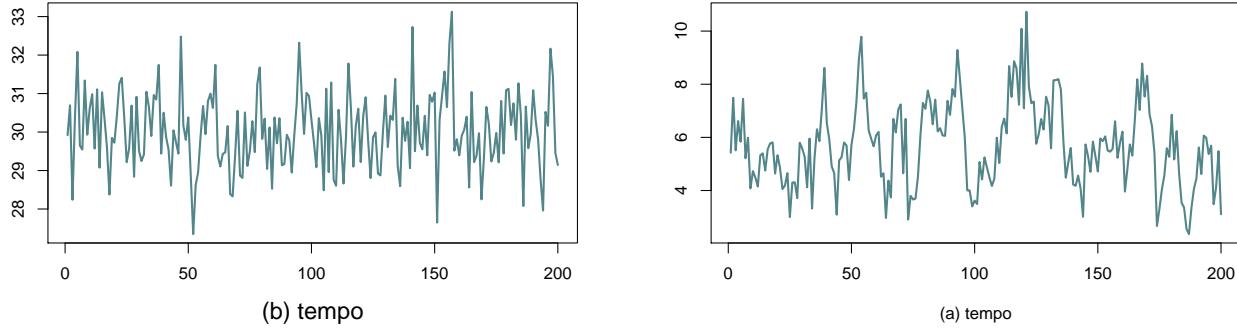


Figure 4.1: Séries estacionárias: (a) Ruído branco, (b) ARMA(1,2).

Entretanto, a maior parte das séries que encontramos na prática apresenta alguma forma de não-estacionariedade. A Figura ?? apresenta algumas séries que apresentam algum tipo de não-estacionariedade que podem resultar de diversas fontes. Algumas das fontes mais comuns de não-estacionariedade de uma série temporal são:

- a) a presença de uma tendência determinística (linear, logarítmica, exponencial, etc.) ao redor da qual a série se desenvolve. Geralmente a presença de uma tendência determinística é facilmente reconhecível através do gráfico. Na Figura ??(a) apresentamos o gráfico de uma série apresentando uma tendência linear.
- b) quebra estrutural na série, que pode ser decorrente de uma mudança na média, como representado na Figura ??(b), ou uma mudança mais sutil, difícil de ser detectada, como por exemplo mudanças na distribuição da série, na variância, no modelo da série, etc.
- c) presença do que chamamos de tendência estocástica, como representado na Figura ??(c). Neste caso a série parece “vagar” por um caminho que apresenta mudanças aleatórias de trajetória, sendo que fica difícil determinar o seu comportamento.
- d) presença de sazonalidade. Neste caso a sazonalidade provoca uma mudança de nível local fazendo com que a média da série se altere nos períodos sazonais. Um exemplo de série sazonal é dado na Figura ??(d).

Mais detalhes serão apresentados adiante. A maior parte das séries que encontramos na prática apresenta alguma forma de não-estacionariedade. As séries econômicas apresentam em geral tendências lineares e muito comumente, tendência estocástica. Podemos ter, também, uma forma de não-estacionariedade explosiva, como o crescimento de uma colônia de bactérias.

4.4.1 Estacionariedade forte ou estrita

Um processo estocástico $Z(t)$ é dito ser um processo **processo fortemente (ou estritamente) estacionário** se a distribuição conjunta de Z_{t_1}, \dots, Z_{t_n} é a mesma de $Z_{t_1-k}, Z_{t_2-k}, \dots, Z_{t_n-k}$, para todas as combinações de tempos t_1, \dots, t_n e para todo $k \in \mathbb{Z}$. Observe que este conceito se traduz em dizer que fixados os tempos t_1, \dots, t_n , ao andarmos k passos à frente homogeneamente no tempo, a distribuição das variáveis não se altera.

Quando $n = 1$, a distribuição de Z_t é igual a distribuição de Z_{t-k} para qualquer k , ou seja, os Z_t 's são identicamente distribuídos. Isto implica que num processo fortemente estacionário, as funções média (μ_t) e variância (σ_t^2) são constantes para todo t , isto é, $\sigma^2 = \text{Var}(Z_t) = \text{Var}(Z_{t-k})$ e $\mu = \mathbb{E}(Z_t) = \mathbb{E}(Z_{t-k})$, independentemente de t e k . Quando $n = 2$, a distribuição de (Z_t, Z_s) é a mesma de (Z_{t-k}, Z_{s-k}) , de onde segue que $\text{Cov}(Z_t, Z_s) = \text{Cov}(Z_{t-k}, Z_{s-k})$, para todo t, s e k .

Fazendo $k = s$ temos:

$$\gamma(t, s) = \text{Cov}(Z_t, Z_s) = \text{Cov}(Z_{t-k}, Z_{s-k}) = \text{Cov}(Z_{t-s}, Z_{s-s}) = \text{Cov}(Z_{t-s}, Z_0) = \gamma(t - s, 0);$$

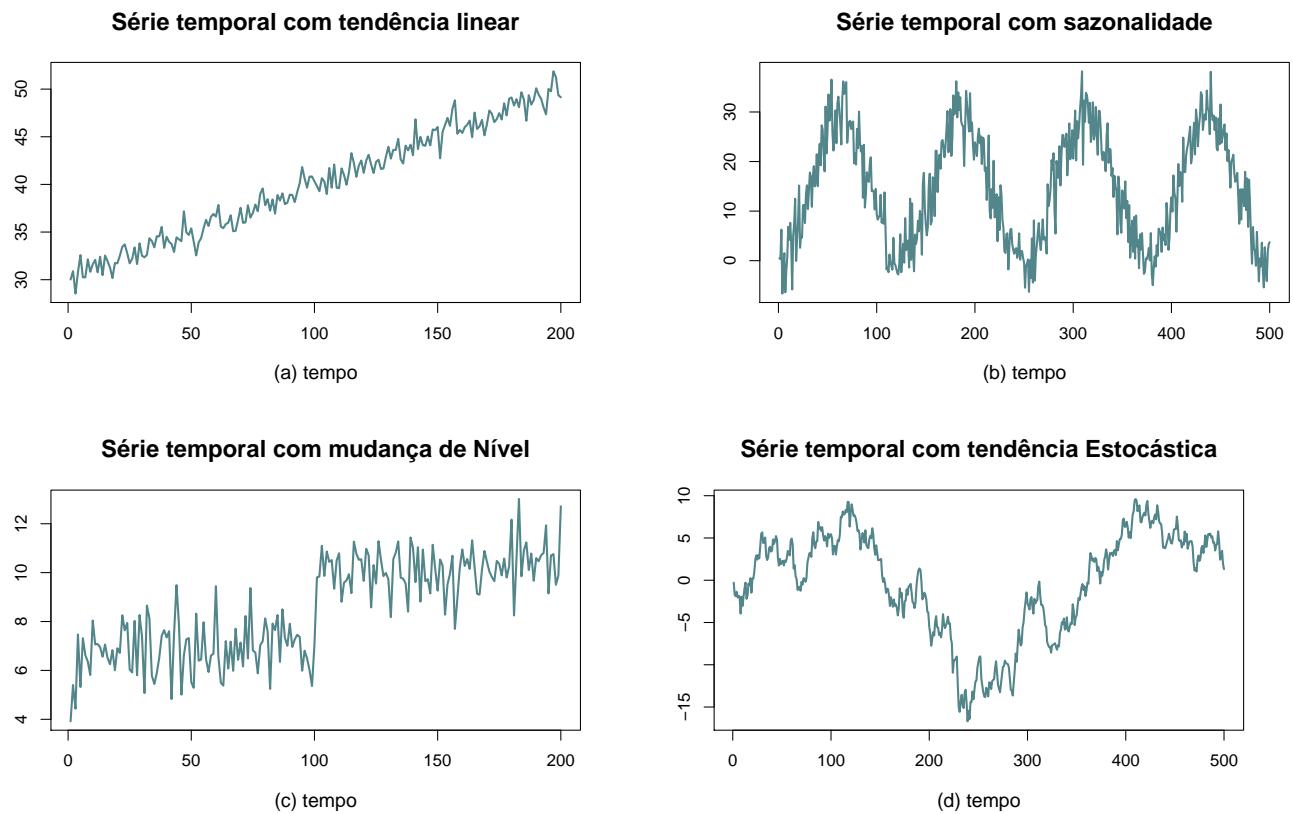


Figure 4.2: Séries não-estacionárias apresentando: (a) Tendência linear, (b) quebra estrutural representada pela mudança de nível da série, (c) tendência estocástica e (d) sazonalidade.

e se $k = t$,

$$\gamma(t, s) = \text{Cov}(Z_{t-t}, Z_{s-t}) = \text{Cov}(Z_0, Z_{s-t}) = \text{Cov}(Z_0, Z_{t-s}) = \gamma(0, s-t),$$

de onde podemos concluir que

$$\gamma(t, s) = \gamma(0, |t-s|), \quad \text{lembrando que } |t-s| = \begin{cases} t-s, & \text{para } t > s; \\ s-t, & \text{para } s > t. \end{cases}$$

Analogamente para a função de autocorrelação. Ou seja, num processo fortemente estacionário a covariância entre Z_t e Z_s depende somente da diferença temporal $|t-s|$ e não dos tempos t e s . Ou ainda, podemos dizer que a $\text{Cov}(Z_t, Z_{t+h})$ depende apenas da distância temporal h entre as variáveis (chamada de \textcolor{blue}{defasagem} ou \textcolor{blue}{lag} entre as variáveis), e não do tempo t . Isto permite simplificar a notação:

$$\begin{aligned} \gamma(h) &= \text{Cov}(Z_t, Z_{t-h}) = \text{Cov}(Z_t, Z_{t+h}) \\ \rho(h) &= \text{cor}(Z_t, Z_{t-h}) = \text{cor}(Z_t, Z_{t+h}), \end{aligned}$$

para todo $t, h \in \mathbb{Z}$. As propriedades gerais da *FAC* e *FACV* para um processo estacionário são:

- 1) $\gamma(0) = \text{Var}(Z_t)$, $\rho(0) = 1$;
- 2) $\gamma(h) = \gamma(-h)$, $\rho(h) = \rho(-h)$;
- 3) $|\gamma(h)| \leq \gamma(0)$, $|\rho(h)| \leq 1$.

Se um processo é estritamente estacionário e tem variância finita, então a *FACV* depende somente do *lag* h .

4.4.2 Estacionariedade fraca ou de segunda ordem

A estacionariedade forte é um conceito na maioria das vezes difícil de ser identificado na prática, mas muito conveniente do ponto de vista matemático. Uma outra maneira de se definir a estacionariedade de uma série de forma que a teoria é matematicamente tratável e de fácil detecção em problemas práticos é a seguinte: um processo estocástico Z_t é dito ser \textcolor{blue}{fracamente estacionário ou estacionário de segunda-ordem} se:

- a) a função média é constante para todo tempo t ;
- b) $\gamma(t, t-h) = \gamma(0, h) = \gamma(h)$ para todo tempo t e lag h .

A condição $\gamma(t, t-h) = \gamma(h)$ para todo tempo t e lag h é equivalente a $\rho(t, t-h) = \rho(h)$. Além disso, $\text{Var}(Z_t) = \gamma(0)$ não depende de t . Como veremos adiante, em processos fracamente estacionários as funções de autocovariância e autocorrelação desempenham papel central no seu estudo. Neste trabalho, sempre que nos referirmos a um processo estacionário, estaremos nos referindo à processos fracamente estacionários.

4.4.3 Teste para significância das autocorrelações

Mais adiante quando estudarmos modelagem ARIMA, precisaremos de ferramentas para decidir se uma dada série é não-correlacionada. Para testar a hipótese conjunta de que $\rho(1) = \dots = \rho(m) = 0$ contra a hipótese de que algum $\rho(k) \neq 0$, pode-se usar a estatística Q_{BP} desenvolvida por [Box](#) e [Pierce](#), ou a estatística Q_{LB} desenvolvida por [Ljung-Box](#), definidas, respectivamente, por

Box e Pierce

$$Q_{BP}(m) = n \sum_{k=1}^m \hat{\rho}_k^2(\hat{\varepsilon})$$

em que n é o tamanho da amostra (série) e m é o maior lag considerado na hipótese. A estatística Q_{BP} em grandes amostras tem distribuição qui-quadrado com m graus de liberdade.

Ljung-Box

$$Q_{LB}(m) = n(n+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2(\hat{\varepsilon})}{n-k}$$

a qual se distribui como uma qui-quadrado com m graus de liberdade em grandes amostras. A estatística Q_{LB} possui maior poder para amostras pequenas que a estatística Q_{BP} .

Observe que a hipótese nula do teste é que todas as correlações de lag $1, \dots, m$ são nulas, para algum m predeterminado, desta forma a escolha do valor de m é fundamental. Quanto maior o m , caso não seja possível rejeitar a hipótese nula, menor é a evidencia de que a série testada é correlacionada. Porém, se m for muito grande, dois problemas poderão acontecer: primeiro, se m é muito próximo de n haverão poucos pontos amostrais com distância temporal m o que torna a estimativa de $\hat{\rho}_k(\hat{\varepsilon})$ problemática, deteriorando a qualidade do teste; segundo, o poder do teste decresce com o aumento de m . Embora não haja consenso na literatura sobre o valor ideal de m , sugerimos utilizar $m = 20$ para séries com $n \geq 50$. Se a série for curta, na literatura encontra-se a sugestão $m = \min(10, n/5)$.

4.4.4 Função de autocorrelação parcial (*FACP*)

A função de autocorrelação parcial (*FACP*) entre as variáveis Y_t e Y_{t+k} , denotada por $\alpha(k)$ em processos estacionários, é a correlação entre as variáveis Y_t e Y_{t+k} removida a influência das variáveis intermediárias $Y_{t+1}, Y_{t+2}, \dots, Y_{t+k-1}$. Dada uma série temporal $\{Y_t\}_{t=1}^{\infty}$ estacionária e uma variável aleatória X , denotemos por $\Pi_{r,s}(X)$ a projeção de X no subespaço gerado pelas variáveis $Y_{r+1}, \dots, Y_{r+s-1}$. A FACP entre Y_t e Y_{t+k} é dada por

$$\alpha(k) = \text{Cor}(Y_t - \Pi_{t,k}(Y_t), Y_{t+k} - \Pi_{t,k}(Y_{t+k})), \text{ para } k \geq 2,$$

e $\alpha(1) = \rho(1)$.

A *FACP* para um processo estacionário com média zero pode ser obtida a partir da regressão

$$y_{t+k} = \phi_{k1}y_{t+k-1} + \phi_{k2}y_{t+k-2} + \dots + \phi_{kk}y_t + \varepsilon_{t+k}, \quad (4.9)$$

da qual podem ser obtidas as equações de *Yule-Walker*.

Multiplicando ambos os lados por y_{t+k-j} e calculando o valor dividindo pela variância, tem-se

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \dots + \phi_{kk}\rho_{k-j}.$$

Então para $j = 1, 2, \dots, k$, temos:

$$\begin{aligned} \rho_1 &= \phi_{k1}\rho_0 + \phi_{k2}\rho_1 + \dots + \phi_{kk}\rho_{k-1}; \\ \rho_2 &= \phi_{k1}\rho_1 + \phi_{k2}\rho_0 + \dots + \phi_{kk}\rho_{k-2}; \\ &\vdots \\ \rho_k &= \phi_{k1}\rho_{k-1} + \phi_{k2}\rho_{k-2} + \dots + \phi_{kk}\rho_0; \end{aligned}$$

Para $k = 1 \rightarrow \hat{\phi}_{11} = \rho_1$.

Para $k = 2 \rightarrow \rho_1 = \phi_{21} + \phi_{22}\rho_1$ e $\rho_2 = \phi_{21}\rho_1 + \phi_{22}$.

Ou podemos escrever a ultima equação em notação matricial:

$$\begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} \begin{bmatrix} \phi_{21} \\ \phi_{22} \end{bmatrix}.$$

cuja solução para o estimador de ϕ_{22} é dada pela regra de *Cramer*:

$$\hat{\phi}_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}$$

Para $k = 3$ temos as equações:

$$\begin{aligned} \rho_1 &= \phi_{31} + \phi_{32}\rho_1 + \phi_{33}\rho_2 \\ \rho_2 &= \phi_{31}\rho_1 + \phi_{32} + \phi_{33}\rho_1 \\ \rho_3 &= \phi_{31} + \phi_{32}\rho_1 + \phi_{33}. \end{aligned}$$

Em notação matricial temos:

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{bmatrix} \begin{bmatrix} \phi_{31} \\ \phi_{32} \\ \phi_{33} \end{bmatrix}.$$

cuja solução para o estimador de ϕ_{33} é dada por:

$$\hat{\phi}_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}},$$

e assim sucessivamente.

4.4.5 Operador de defasagem ou operador *lag*

Em séries temporais é usual trabalhar com operadores que defasam a variável. Definimos então o operador de defasagem L como um operador linear tal que:

Operador defasagem

$$L^j y_t = y_{t-j}$$

As seguintes propriedades do operador L serão úteis no que segue:

- 1) O operador lag aplicado a uma constante resulta na própria constante, isto é, $Lc = c$;
- 2) O operador lag segue a propriedade distributiva em relação à soma

$$(L^i + L^j)Y_t = L^i Y_t + L^j Y_t = Y_{t-i} + Y_{t-j};$$

3) É válida a propriedade associativa da multiplicação

$$L^i L^j Y_t = L^i (L^j Y_t) = L^i (Y_{t-j}) = Y_{t-i-j}.$$

Ou ainda $L^i L^j Y_t = L^{i+j} Y_t = Y_{t-i-j}$;

4) Potências negativas de L significam um operador de avanço,

$L^{-i} Y_t = L^j Y_t$, fazendo $j = -i$. Então $L^{-i} Y_t = L^j Y_t = Y_{t-j} = Y_{t+i}$;

5) Se $|a| < 1$ definimos o operador inverso

$$(1 - aL)^{-1} = 1 + aL + a^2 L^2 + \dots = \sum_{j=0}^{\infty} (aL)^j.$$

A ação do operador $(1 - aL)^{-1}$ em uma variável Y_t é a seguinte:

$$(1 - aL)^{-1}(Y_t) = (1 + aL + a^2 L^2 + \dots)(Y_t) = Y_t + aY_{t-1} + a^2 Y_{t-2} + \dots = \sum_{j=1}^{\infty} a^j Y_{t-j}.$$

Note ainda que, para uma constante $c \in \mathbb{R}$

$$(1 - aL)^{-1}(c) = (1 + aL + a^2 L^2 + \dots)(c) = (c + aL(c) + a^2 L^2(c) + \dots) = c \sum_{j=1}^{\infty} a^j = \frac{c}{1 - a}.$$

6) Se $|a| > 1$, definimos o operador

$$-aL(1 - aL)^{-1} = (1 + (aL)^{-1} + (aL)^{-2} + \dots)$$

A ação do operador $-aL(1 - aL)^{-1}$ em uma variável Y_t é a seguinte:

$$\begin{aligned} (-aL(1 - aL)^{-1})(Y_t) &= (1 + (aL)^{-1} + (aL)^{-2} + \dots)(Y_t) = Y_t + \frac{1}{a} Y_{t+1} + \frac{1}{a^2} Y_{t+2} + \dots \\ &= \sum_{j=1}^{\infty} \frac{1}{a^j} Y_{t+j}. \end{aligned}$$

Para uma constante $c \in \mathbb{R}$, a ação do operador $-aL(1 - aL)^{-1}$ é dada por

$$(-aL(1 - aL)^{-1})(c) = (1 + (aL)^{-1} + (aL)^{-2} + \dots)(c) = c + \frac{c}{a} + \frac{c}{a^2} + \dots = \frac{ca}{1 - a}.$$

4.4.6 Ruído branco

Um importante exemplo de processo estacionário é o ruído branco, o qual é definido como uma sequência de variáveis aleatórias $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ com as seguintes propriedades:

Ruído Branco

- 1) $\mathbb{E}(\varepsilon_t) = 0$, para todo $t \in \mathbb{R}$;
- 2) $\mathbb{E}(\varepsilon_t^2) = \sigma^2$ para todo $t \in \mathbb{R}$;
- 3) $\mathbb{E}(\varepsilon_t \varepsilon_s) = 0$, para todo $t \neq s$, com $t, s \in \mathbb{R}$.

Em outras palavras, um ruído branco é uma sequência de variáveis não-correlacionadas com média constante. Denotaremos um processo ruído branco por $RB(0, \sigma^2)$. Escreveremos ainda $\varepsilon_t \sim RB(0, \sigma_\varepsilon^2)$ para dizer que $\{\varepsilon_t\}_t$ é um processo ruído branco com média 0 e variância σ_ε^2

Para um ruído branco $\varepsilon_t \sim RB(0, \sigma_\varepsilon^2)$, $\mu_t = \mathbb{E}(\varepsilon_t) = 0$ é constante com *FACV* e *FAC* dadas por

$$\gamma_\varepsilon(h) = \begin{cases} \sigma_\varepsilon^2, & \text{se } h = 0; \\ 0, & \text{se } h \neq 0. \end{cases} \quad \rho_\varepsilon(h) = \begin{cases} 1, & \text{se } h = 0; \\ 0, & \text{se } h \neq 0. \end{cases}$$

O termo ruído branco resulta do fato que em uma análise de frequência do modelo, podemos mostrar que todas as frequências são iguais. As características de um processo ruído branco ficam explícitas quando analisamos o seguinte gráfico

```
set.seed(5647)
eps=rnorm(200)

plot(eps,type="l",lwd=2,cex.lab=1.4, xlab="Ruído Branco",ylab="", cex.main=1.7,col="cadetblue4")
acf(eps,lwd=2,cex.lab=1.4, xlab=" lag ", main=" ", cex.main=1.7,col="cadetblue4")
pacf(eps,lwd=2,cex.lab=1.4, xlab=" lag ", main=" ", cex.main=1.7,col="cadetblue4",ylim=c(-0.9,0.9))
```

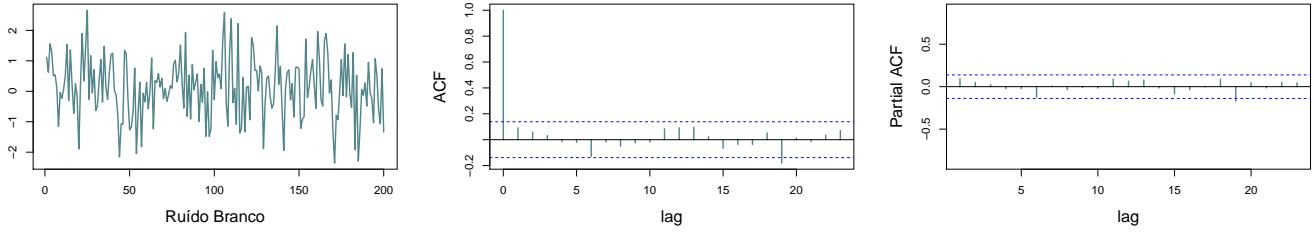


Figure 4.3: Ruído branco gaussiano simulado, FAC amostral e FACP amostral

Exemplo 4.1 (Média-Móvel de ordem 1). Este é um exemplo simples de um processo estacionário que não é um ruído branco. Suponha que

Processo MA(1)

$$Y_t = \varepsilon_t - 0.5\varepsilon_{t-1},$$

onde ε_t é um $RB(0, \sigma_\varepsilon^2)$.

Média do MA(1)

$$\mu_t = \mathbb{E}(Y_t) = \mathbb{E}(\varepsilon_t) - 0.5\mathbb{E}(\varepsilon_{t-1}) = 0$$

Variância do MA(1)

$$\text{Var}(Y_t) = \text{Var}(\varepsilon_t - 0.5\varepsilon_{t-1}) = \sigma_\varepsilon^2 + 0.5\sigma_\varepsilon^2 = 1.25\sigma_\varepsilon^2.$$

Quanto à estrutura de covariância/correlação de um MA(1), temos

$$\begin{aligned} \text{Cov}(Y_t, Y_{t+h}) &= \text{Cov}(\varepsilon_t - 0.5\varepsilon_{t-1}, \varepsilon_{t+h} - 0.5\varepsilon_{t+h-1}) \\ &= \text{Cov}(\varepsilon_t, \varepsilon_{t+h}) - 0.5\text{Cov}(\varepsilon_t, \varepsilon_{t+h-1}) - 0.5\text{Cov}(\varepsilon_{t-1}, \varepsilon_{t+h}) + 0.25\text{Cov}(\varepsilon_{t-1}, \varepsilon_{t+h-1}) \\ &= \gamma_\varepsilon(h) - 0.5\gamma_\varepsilon(h-1) - 0.5\gamma_\varepsilon(h+1) + 0.25\gamma_\varepsilon(h), \end{aligned} \tag{4.10}$$

onde γ_ε denota a função de autocovariância de ε_t . Da equação ??eq:gppo, percebemos que $\text{Cov}(Y_t, Y_{t+h})$ só é diferente de zero quando algum dos argumentos das funções à direita da igualdade em ??eq:gppo é zero. Isto ocorre somente quando $h = 0$ (resultando em $\text{Var}(Y_t) = 1.25\sigma_\varepsilon^2$) e quando $|h| = 1$ (resultando $-0.5\sigma_\varepsilon^2$). Em outras palavras, $\text{Cov}(Y_t, Y_{t+h})$ não depende de t e

$$\gamma(k) = \begin{cases} -0.5\sigma_\varepsilon^2, & \text{se } |k| = 1; \\ 0, & \text{se } |k| > 1. \end{cases} \quad \text{e} \quad \rho(k) = \begin{cases} -0.4, & \text{se } |k| = 1; \\ 0, & \text{se } |k| > 1. \end{cases}$$

Concluimos que um MA(1) é estacionário.

4.5 Metodologia de Box-Jenkins ou modelagem ARIMA

Na análise de séries temporais, a metodologia de Box-Jenkins, nomeada em homenagem ao estatísticos George Box e Gwilym Meirion Jenkins, é uma metodologia pensada para a modelagem de séries temporais que é suficientemente simples de forma a atingir um grande público e suficientemente flexível para se aplicar a uma gama grande de problemas. Centrais à metodologia Box-Jenkins são os modelos *Autorregressivos Integrados de Média Móvel*, abreviados modelos ARIMA, que representam uma classe grande de modelos capazes de modelar uma variedade de tipos de séries temporais. O intuito é modelar os valores da série temporal em função dos seus valores passados (admitindo um termo de erro) de forma que seja possível fazer previsões para esta série. O procedimento pode ser resumido em três etapas:

- Identificação e seleção do modelo. Nesta etapa verificamos se as variáveis são estacionárias, identificando possíveis tendências e/ou sazonalidades na série, removendo-as quando detectadas. Fazemos o uso das funções de autocorrelação e autocorrelação parcial para decidir qual modelo da classe ARIMA é adequado para uma primeira tentativa de modelagem..
- Estimação dos parâmetros usando algoritmos computacionais para chegar a coeficientes que melhor se adaptam ao modelo ARIMA selecionado. Os métodos mais comuns são a máxima verossimilhança e os mínimos quadráticos não-lineares.
- Verificação do ajuste do modelo por meio de testes. Nesta fase, verificamos se o modelo estimado está em conformidade com as especificações do modelo teórico proposto. De suma importância é a análise residual na qual o objetivo é verificar se os resíduos satisfazem a hipótese de serem não-correlacionados. De grande utilidade é o teste Ljung-Box. Se o modelo proposto é inadequado, devemos voltar para a primeira etapa e propor um modelo alternativo.

Um dos modelos mais simples, útil e intuitivo é o modelo autorregressivo. Consideremos o caso mais simples.

4.5.1 Modelo autorregressivo de ordem 1 AR(1)

Processo AR(1)

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t,$$

em que ε_t é um $RB(0, \sigma_\varepsilon^2)$. Para calcularmos a média e a variância do processo, assumiremos primeiramente que os momentos incondicionais sejam iguais, de forma que $\mathbb{E}Y_t = \mathbb{E}Y_{t-1}$. Com esta simplificação, fica fácil calcular a média de um processo AR(1):

A média do processo AR(1) é

$$\mu = \mathbb{E}(Y_t) = \mathbb{E}(c) + \phi\mathbb{E}(Y_{t-1}) + \mathbb{E}(\varepsilon_t) \iff \mu = c + \phi\mu + 0 \iff \mu = \frac{c}{1 - \phi},$$

desde que $\phi \neq 1$. Se $\phi = 1$ a equação não possui solução ¹.

¹mais tarde veremos que para $\phi = 1$ o processo é não estacionário e de fato a média varia com t , sendo, portanto, falsa a hipótese inicial de que a média do processo é constante, utilizada para derivar as equações.

Desta forma procedemos assumindo que $\phi \neq 1$. Observe ainda que $\mu = 0$, quando $c = 0$. Para $\phi \neq 1$, a variância de um AR(1), por sua vez, é dada por

A variância do AR(1) é

$$\text{Var}(Y_t) = \mathbb{E}(Y_t^2) - \mu^2 = \frac{\sigma^2}{1 - \phi^2}.$$

Observe que se $|\phi| > 1$, a variância será negativa, o que é um absurdo. Neste caso as equações não são compatíveis com nenhum processo. Quando $|\phi| = 1$, a variância de Y_t não está definida pois a média não está.

Deste exemplo, é possível concluir que é necessário estabelecer algumas restrições sobre o modelo para que se possa estimá-lo. Em particular, uma condição necessária para estimar os coeficientes do modelo é que $|\phi| < 1$.

Com um pouco mais de trabalho, podemos encontrar o mesmo resultado sem a suposição de que os momentos incondicionais sejam iguais. Para isso usamos o operador defasagem L e suas propriedades para obtermos

$$\begin{aligned} Y_t = c + \phi Y_{t-1} + \varepsilon_t &\iff (1 - \phi L)Y_t = c + \varepsilon_t \iff Y_t = (1 - \phi L)^{-1}(c) + (1 - \phi L)^{-1}(\varepsilon_t) \\ &\iff Y_t = \mu + \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}, \end{aligned} \tag{4.11}$$

onde escrevemos $\mu = c/(1 - \phi)$ por simplicidade. A partir desta representação podemos facilmente obter

$$\mathbb{E}Y_t = \mu + \sum_{j=0}^{\infty} \phi^j \mathbb{E}(\varepsilon_{t-j}) = \mu$$

e

$$\begin{aligned} \text{Var}(Y_t) &= \mathbb{E}(Y_t - \mu)^2 = \mathbb{E}\left(\sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}\right)^2 = \mathbb{E}\left(\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \phi^k \phi^j \varepsilon_{t-j} \varepsilon_{t-k}\right) \\ &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \phi^{k+j} \mathbb{E}(\varepsilon_{t-j} \varepsilon_{t-k}) = \sum_{j=0}^{\infty} \phi^{2j} \mathbb{E}(\varepsilon_{t-j}^2) = \frac{\sigma_{\varepsilon}^2}{1 - \phi^2}, \end{aligned}$$

onde a última igualdade segue do fato de que $\mathbb{E}(\varepsilon_{t-j} \varepsilon_{t-k}) = \gamma_{\varepsilon}(j - k)$ que é igual a zero se $j \neq k$, e σ_{ε}^2 , se $j = k$. Para $h > 0$, a

\textcolor{blue}{\{ função de autocovariância de lag \}} h é dada por

$$\begin{aligned} \gamma(h) &= \mathbb{E}[(Y_t - \mu)(Y_{t-h} - \mu)] = \mathbb{E}\left[\left(\sum_{s=0}^{\infty} \phi^s \varepsilon_{t-s}\right)\left(\sum_{k=0}^{\infty} \phi^k \varepsilon_{t-k-h}\right)\right] \\ &= \sum_{s=0}^{\infty} \sum_{k=0}^{\infty} \phi^{k+s} \mathbb{E}(\varepsilon_{t-s} \varepsilon_{t-k-h}) = \sum_{s=0}^{\infty} \sum_{k=0}^{\infty} \phi^{k+s} \gamma_{\varepsilon}(s - k - h). \end{aligned}$$

Observe que $\gamma_{\varepsilon}(s - k - h)$ só é diferente de zero quando $s - k - h = 0$, o que é equivalente a $s = k + h$, assim

$$\gamma(h) = \sigma_{\varepsilon}^2 \sum_{k=0}^{\infty} \phi^{2k+h} = \sigma_{\varepsilon}^2 (\phi^h + \phi^{h+2} + \phi^{h+4} + \dots) = \frac{\phi^h}{1 - \phi^2} \sigma_{\varepsilon}^2,$$

Para $h < 0$, analogamente obtemos

$$\gamma(h) = \frac{\phi^{-h}}{1 - \phi^2} \sigma_\varepsilon^2,$$

ou seja, para $h \neq 0$,

$$\gamma(h) = \frac{\phi^{|h|}}{1 - \phi^2} \sigma_\varepsilon^2.$$

Como a média e as covariâncias não são funções do tempo o processo é fracamente estacionário, independente do valor de $\phi \in (-1, 1)$.

\textcolor{blue}{A função de autocorrelação de *lag* } h é dada por

$$\rho(h) = \frac{\frac{\phi^{|h|}}{1-\phi^2} \sigma^2}{\frac{\sigma^2}{1-\phi^2}} = \phi^{|h|}.$$

Além disso, como $|\phi| < 1$, a função de autocorrelação é decrescente em $|h|$.

4.5.2 Passeio aleatório (*Random Walk*)

Quando $\phi = 1$ no caso anterior, temos o processo chamado passeio aleatório. Seja $\{\varepsilon_t\}_{t \in \mathbb{N}}$ um $\text{RB}(0, \sigma_\varepsilon^2)$.

O Passeio aleatório pode ser representado por

$$Z_t = Z_{t-1} + \varepsilon_t,$$

que pode ser reescrito de uma maneira bem simples. Defina inicialmente

$$Z_1 = \varepsilon_1, \quad Z_2 = \varepsilon_1 + \varepsilon_2 \leftrightarrow Z_2 = Z_1 + \varepsilon_2$$

e sucessivamente

$$Z_{k-1} = \varepsilon_1 + \cdots + \varepsilon_{k-1}, \quad Z_k = \varepsilon_1 + \cdots + \varepsilon_{k-1} + \varepsilon_k = Z_{k-1} + \varepsilon_k.$$

Com esta representação, o cálculo da média e da variância de Z_t se tornam simples:

A média do passeio aleatório é

$$\mu_t = \mathbb{E}(Z_t) = \mathbb{E}(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t) = \mathbb{E}(\varepsilon_1) + \mathbb{E}(\varepsilon_2) + \cdots + \mathbb{E}(\varepsilon_t) = 0 + 0 + \cdots + 0 = 0,$$

e a variância do passeio aleatório é

$$\text{Var}(Z_t) = \text{Var}(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t) = \text{Var}(\varepsilon_1) + \cdots + \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2 + \sigma_\varepsilon^2 + \cdots + \sigma_\varepsilon^2 = t\sigma_\varepsilon^2.$$

Assim concluímos que a variância de um passeio aleatório cresce linearmente com o tempo, sendo portanto um processo não-estacionário. Observe ainda que se $1 \leq t \leq s$, a função de autocovariância de um passeio aleatório é dada por

$$\begin{aligned} \gamma(t, s) &= \text{Cov}(Z_t, Z_s) \\ &= \text{Cov}(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t, \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_s) \\ &= \sum_{i=1}^t \sum_{j=1}^s \text{Cov}(\varepsilon_i, \varepsilon_j) \\ &= \text{Cov}(\varepsilon_1, \varepsilon_1) + \text{Cov}(\varepsilon_2, \varepsilon_2) + \cdots + \text{Cov}(\varepsilon_t, \varepsilon_t) \\ &= \sigma_\varepsilon^2 + \sigma_\varepsilon^2 + \cdots + \sigma_\varepsilon^2 = t\sigma_\varepsilon^2 \end{aligned}$$

onde $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ para $i \neq j$. O mesmo argumento mostra que se $1 \leq s \leq t$, teremos $\gamma(t, s) = s\sigma_\varepsilon^2$, de forma que podemos escrever compactamente $\gamma(s, t) = \min(s, t)\sigma^2$. A função de autocorrelação de um passeio aleatório é facilmente obtida

$$\begin{aligned}\rho(t, s) &= \frac{\gamma(s, t)}{\sqrt{\text{Var}(X_t)}\sqrt{\text{Var}(X_s)}} = \frac{\min(s, t)\sigma_\varepsilon^2}{\sqrt{t}\sigma_\varepsilon^2\sqrt{s}\sigma_\varepsilon^2} = \frac{\min(s, t)}{\sqrt{t}\sqrt{s}} = \begin{cases} \frac{s}{\sqrt{t}\sqrt{s}} = \sqrt{\frac{s}{t}}, & \text{se } 1 \leq s \leq t; \\ \frac{t}{\sqrt{t}\sqrt{s}} = \sqrt{\frac{t}{s}}, & \text{se } 1 \leq t \leq s \end{cases} \\ &= \sqrt{\frac{\min(s, t)}{\max(s, t)}}.\end{aligned}$$

Em resumo, a *FACV* e a *FAC* de um passeio aleatório são dadas por

FACV do passeio aleatório

$$\gamma(t, s) = \min(s, t)\sigma_\varepsilon^2,$$

FAC do passeio aleatório

$$\rho(t, s) = \sqrt{\frac{\min(s, t)}{\max(s, t)}}$$

O passeio aleatório é um exemplo simples que serve de aproximação para diversas situações reais, tais como o movimento comum de preços e títulos e também a posição de pequenas partículas suspensas dentro de um fluido, chamado movimento Browniano.

```
set.seed(1231)
eps=rnorm(200)
rw=cumsum(eps)
plot(rw,type="l",lwd=2,cex.lab=1.4, xlab="tempo",ylab="", main="(a)", cex.main=1.7,col="cadetblue4")
acf(rw,type = "correlation",lag.max = 50,cex.main=1.7,col="cadetblue4",ylab="FAC",lwd=4,main="(b)")
acf(rw,type = "partial" ,lag.max = 50,cex.main=1.7,col="cadetblue4",ylab="FACP",lwd=4,main="(c)")
```

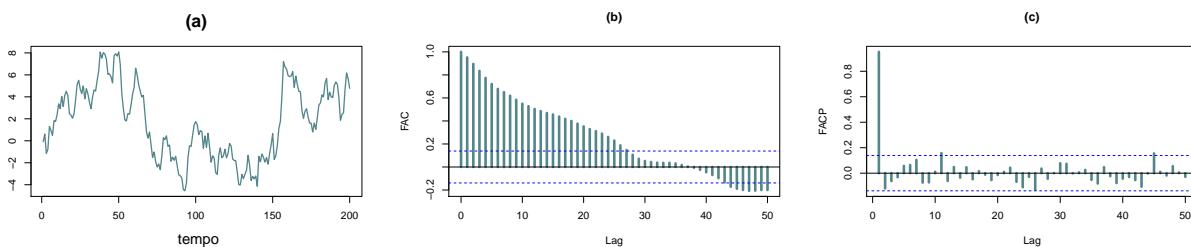


Figure 4.4: Passeio aleatório simulado (a), FACV amostral (b) e FAC amostral (c).

4.5.3 Modelos autorregressivos de ordem p , AR(p)

O processo autorregressivo de ordem p é definido como

\textcolor{blue}{\{ O processo AR(p) \}}

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t = \sum_{j=1}^p \phi_j Y_{t-j} + \varepsilon_t.$$

Escrevendo Y_t em função do operador lag , obtemos

\textcolor{blue}{\{ Definição com o operador lag \}}

$$\begin{aligned} Y_t = \phi_1 L(Y_t) + \phi_2 L^2(Y_t) + \cdots + \phi_p L^p(Y_t) + \varepsilon_t &\iff (1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p) Y_t = \varepsilon_t \\ &\iff \Phi_p(L) Y_t = \varepsilon_t, \end{aligned}$$

onde $\Phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$. O polinômio $\Phi_p(\cdot)$ será importante no estudo da estacionariedade e causalidade de processos ARMA, que veremos adiante.

4.5.3.1 Alguns processos ARMA simulados

```
set.seed(12322)
ar1=arima.sim(n=200,model=list(ar=0.5))
plot(ar1,type="l",lwd=2,cex.lab=1.4, xlab="tempo",ylab="", main="(a)", cex.main=1.7,col="cadetblue4")
acf(ar1,type = "correlation",lag.max = 50,cex.main=1.7,col="cadetblue4",ylab="FAC",lwd=4,main="(b)")
acf(ar1,type = "partial" ,lag.max = 50,cex.main=1.7,col="cadetblue4",ylab="FACP",lwd=4,main="(c)")
```

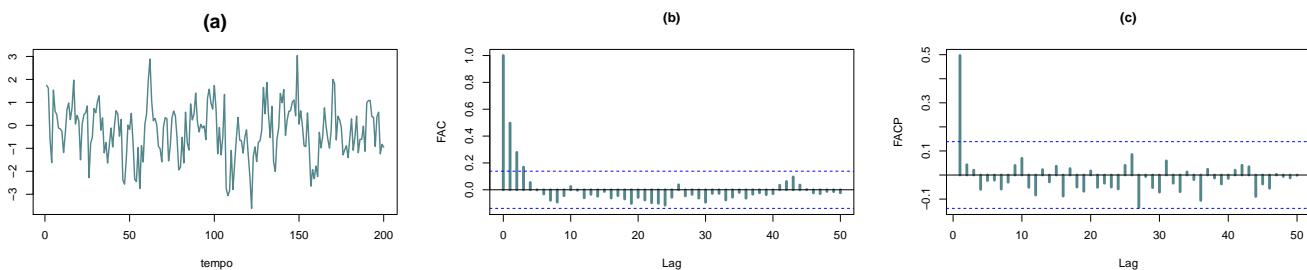


Figure 4.5: AR(1) simulado com coeficiente $\phi_1 = 0.5$ (a), FACV amostral (b) e FACP amostral (c).

```
set.seed(12322)
ar1m=arima.sim(n=200,model=list(ar=-0.5))
plot(ar1m,type="l",lwd=2,cex.lab=1.4, xlab="tempo",ylab="", main="(a)", cex.main=1.7,col="cadetblue4")
acf(ar1m,type = "correlation",lag.max = 50,cex.main=1.7,col="cadetblue4",ylab="FAC",lwd=4,main="(b)")
acf(ar1m,type = "partial" ,lag.max = 50,cex.main=1.7,col="cadetblue4",ylab="FACP",lwd=4,main="(c)")
```

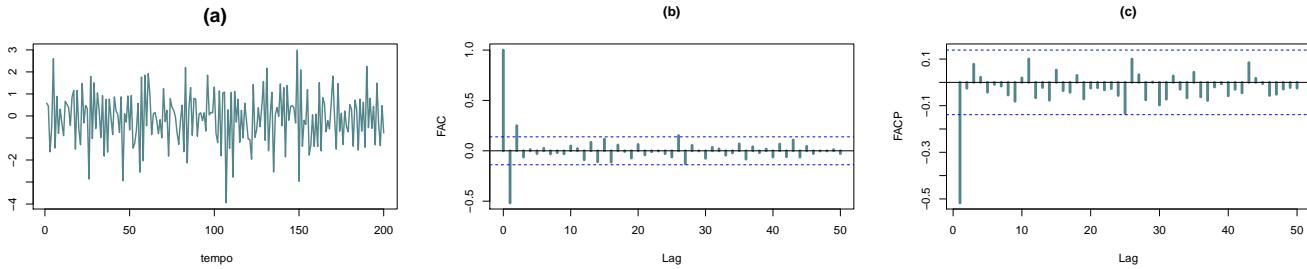


Figure 4.6: AR(1) simulado com coeficiente $\phi_1 = -0.5$ (a), FACV amostral (b) e FACP amostral (c).

```
set.seed(12322)
ar108=arima.sim(n=200,model=list(ar=-0.5))
plot(ar108,type="l",lwd=2,cex.lab=1.4, xlab=" tempo",ylab="", main="(a)", cex.main=1.7,col="cadetblue4")
acf(ar108,type = "correlation",lag.max = 50,cex.main=1.7,col="cadetblue4",ylab="FAC",lwd=4,main="(b)")
acf(ar108,type = "partial" ,lag.max = 50,cex.main=1.7,col="cadetblue4",ylab="FACP",lwd=4,main="(c)")
```

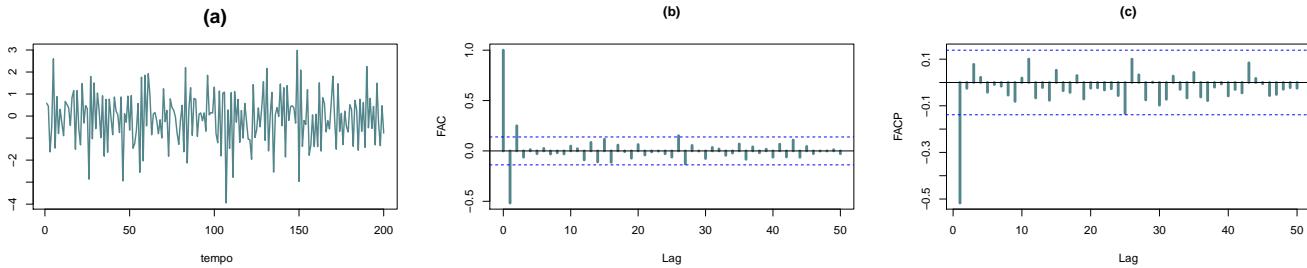


Figure 4.7: AR(1) simulado com coeficiente $\phi_1 = 0.8$ (a), FACV amostral (b) e FACP amostral (c).

```
set.seed(12322)
ar2=arima.sim(n=200,model=list(ar=c(0.5,-0.7)))
plot(ar2,type="l",lwd=2,cex.lab=1.4, xlab=" tempo",ylab="", main="(a)", cex.main=1.7,col="cadetblue4")
acf(ar2,type = "correlation",lag.max = 50,cex.main=1.7,col="cadetblue4",ylab="FAC",lwd=4,main="(b)")
acf(ar2,type = "partial" ,lag.max = 50,cex.main=1.7,col="cadetblue4",ylab="FACP",lwd=4,main="(c)")
```

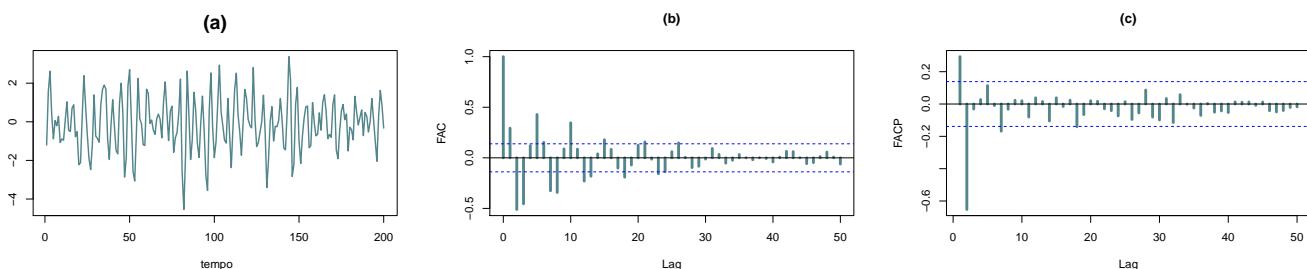


Figure 4.8: AR(2) simulado com coeficiente $\phi_1 = 0.5$ e $\phi_2 = -0.7$ (a), FACV amostral (b) e FACP amostral (c).

```
set.seed(12322)
ar3=arima.sim(n=200,model=list(ar=c(0.5,-0.7,0.6)))
plot(ar3,type="l",lwd=2,cex.lab=1.4, xlab=" tempo",ylab="", main="(a)", cex.main=1.7,col="cadetblue4")
acf(ar3,type = "correlation",lag.max = 50,cex.main=1.7,col="cadetblue4",ylab="FAC",lwd=4,main="(b)")
acf(ar3,type = "partial" ,lag.max = 50,cex.main=1.7,col="cadetblue4",ylab="FACP",lwd=4,main="(c)")
```

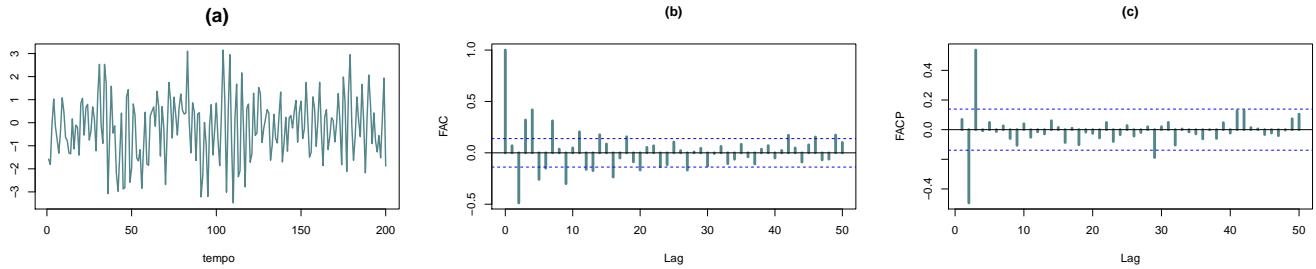


Figure 4.9: AR(3) simulado com coeficiente $\phi_1 = 0.5$, $\phi_2 = -0.7$ e $\phi_3 = 0.6$ (a), FACV amostral (b) e FAC amostral (c).

4.5.4 Modelo de médias-móveis, MA(q)

Chamamos de médias-móveis de ordem q o processo definido por:

\textcolor{blue}{\text{O processo MA}(q)}

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

em que ε_t é um $RB(0, \sigma_\varepsilon^2)$.

Esta terminologia vem do fato que Y_t é obtido aplicando-se os pesos $1, -\theta_1, -\theta_2, \dots, -\theta_q$, às variáveis $\varepsilon_t - \varepsilon_{t-1} - \varepsilon_{t-2} - \cdots - \varepsilon_{t-q}$ e então movendo os mesmos pesos 1 unidade do tempo a frente e aplicando-lhes a $\varepsilon_{t+1} - \varepsilon_t - \varepsilon_{t-1} - \cdots - \varepsilon_{t-q+1}$ para obter Y_{t+1} .

Usando o operador L , podemos reescrever o modelo MA(q) como

$$Y_t = \Theta_q(L)\varepsilon_t, \quad (4.12)$$

$$\Theta_q(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q. \quad (4.13)$$

4.5.5 O modelo MA(1)

Para $q = 1$, obtemos o modelo:

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}, \quad (4.14)$$

onde ε_t é um $RB(0, \sigma_\varepsilon^2)$. Segue que

Média do MA(1)

$$\mathbb{E}(Y_t) = \mathbb{E}(\varepsilon_t + \theta_1 \varepsilon_{t-1}) = \mathbb{E}(\varepsilon_t) + \theta_1 \mathbb{E}(\varepsilon_{t-1}) = 0.$$

A variância do MA(1)

$$\text{Var}(Y_t) = \text{Var}(\varepsilon_t + \theta_1 \varepsilon_{t-1}) = \sigma_\varepsilon^2 + \theta_1^2 \sigma_\varepsilon^2 = (1 + \theta_1^2) \sigma_\varepsilon^2.$$

Temos ainda que a função de autocovariância de *lag* h é:

\textcolor{blue}{\{ FACV do MA(1) \}}

$$\begin{aligned}\gamma(h) &= \text{Cov}(Y_t, Y_{t+h}) \\ &= \text{Cov}(\varepsilon_t + \theta_1 \varepsilon_{t-1}, \varepsilon_{t+h} + \theta_1 \varepsilon_{t+h-1}) \\ &= \text{Cov}(\varepsilon_t, \varepsilon_{t+h}) + \theta_1 \text{Cov}(\varepsilon_t, \varepsilon_{t+h-1}) + \theta_1 \text{Cov}(\varepsilon_{t-1}, \varepsilon_{t+h}) + \theta_1^2 \text{Cov}(\varepsilon_{t-1}, \varepsilon_{t+h-1}) \\ &= \gamma_\varepsilon(h) + \theta_1 \gamma_\varepsilon(h-1) + \theta_1 \gamma_\varepsilon(h+1) + \theta_1^2 \gamma_\varepsilon(h).\end{aligned}$$

Neste caso $\gamma(h)$ só é diferente de 0 quando algum dos argumentos de γ_ε for igual a 0, o que acontece somente quando $h = 0$ ou $h = 1$ ou $h = -1$. Para $h = 0$ obtemos a variância. Para $h = 1$ e $h = -1$ obtemos $\gamma(1) = \gamma(-1) = \theta_1$ e para $|h| \geq 2$ teremos $\gamma(h) = 0$. Desta forma,

\textcolor{blue}{\{ FACV FAC e do MA(1) \}}

$$\gamma(h) = \begin{cases} (1 + \theta_1^2) \sigma_\varepsilon^2 & \text{se } h = 0; \\ \theta_1 & \text{se } |h| = 1; \\ 0 & \text{se } |h| \geq 2, \end{cases} \quad \rho(h) = \begin{cases} 1 & \text{se } h = 0; \\ \frac{\theta_1}{1 + \theta_1^2} & \text{se } |h| = 1; \\ 0 & \text{se } |h| \geq 2. \end{cases}$$

4.5.6 Propriedades do modelo MA(q)

Seja $\varepsilon_t \sim RB(0, \sigma_\varepsilon^2)$ e considere o modelo MA(q)

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}.$$

Definindo $\theta_0 = 1$, podemos reescrever o modelo MA(q) compactamente como

$$Y_t = \sum_{k=0}^q \theta_k \varepsilon_{t-k}.$$

A partir daí podemos obter facilmente a média e a variância de Y_t , assim como sua estrutura de autocovariância e autocorrelação. Primeiramente

\textcolor{blue}{\{ Média do MA(q) \}}

$$\mathbb{E}(Y_t) = \mathbb{E}\left(\sum_{k=0}^q \theta_k \varepsilon_{t-k}\right) = \sum_{k=0}^q \theta_k \mathbb{E}(\varepsilon_{t-k}) = 0$$

\textcolor{blue}{A variância do MA(q)}

$$\begin{aligned}\text{Var}(Y_t) &= \text{Var}\left(\sum_{k=0}^q \theta_k \varepsilon_{t-k}\right) = \sum_{k=0}^q \theta_k^2 \text{Var}(\varepsilon_{t-k}) = \sigma_\varepsilon^2 \sum_{k=0}^q \theta_k^2 \\ &= (1 + \theta_1^2 + \dots + \theta_q^2) \sigma_\varepsilon^2.\end{aligned}$$

\textcolor{blue}{FACV do MA(q)}

$$\begin{aligned}\gamma(h) &= \text{Cov}(Y_t, Y_{t-h}) \\ &= \text{Cov}\left(\sum_{k=0}^q \theta_k \varepsilon_{t-k}, \sum_{j=0}^q \theta_j \varepsilon_{t-h-j}\right) \\ &= \sum_{k=0}^q \sum_{j=0}^q \theta_k \theta_j \text{Cov}(\varepsilon_{t-k}, \varepsilon_{t-h-j}) \\ &= \sum_{k=0}^q \sum_{j=0}^q \theta_k \theta_j \gamma_\varepsilon(k - h - j)\end{aligned}$$

note que $\gamma_\varepsilon(k - h - j) \neq 0$ somente quando $k - h - j = 0$, ou seja, se $j = k - h$.

Além disso, se $|h| > q$, não é possível acontecer $k - h - j = 0$. Desta forma, se $|h| \leq q$,

$$\gamma(h) = \sum_{k=0}^q \theta_k \theta_{k-h} \gamma_\varepsilon(0) = \sigma_\varepsilon^2 \sum_{k=0}^q \theta_k \theta_{k-h}.$$

Concluimos que

\textcolor{blue}{FACV e FAC do MA(q)}

$$\gamma(h) = \begin{cases} \sigma_\varepsilon^2 \sum_{k=0}^q \theta_k \theta_{k-h}, & \text{se } |h| \leq q; \\ 0, & \text{se } |h| > q, \end{cases} \quad \text{e} \quad \rho(h) = \begin{cases} \frac{\sum_{k=0}^q \theta_k \theta_{k-h}}{\sum_{j=0}^q \theta_j^2}, & \text{se } |h| \leq q; \\ 0, & \text{se } |h| > q. \end{cases}$$

```
# MA(1) Simulado
set.seed(12322)
ma1_1=arima.sim(n=200,model=list(ma=c(1)))
plot(ma1_1,type="l",lwd=2,cex.lab=1.4, xlab=" tempo",ylab="", main="(a)", cex.main=1.7,col="cadetblue4")
acf(ma1_1,type = "correlation",lag.max = 50,cex.main=1.7,col="cadetblue4",ylab="FAC",lwd=4,main="(b)")
acf(ma1_1,type = "partial" ,lag.max = 50,cex.main=1.7,col="cadetblue4",ylab="FACP",lwd=4,main="(c)")
```

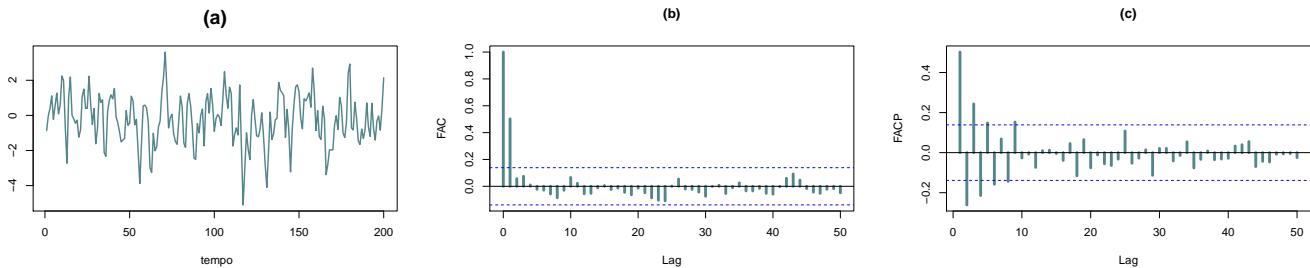


Figure 4.10: MA(1) simulado com coeficiente $\theta_1 = 1$, (a) FACV amostral (b) e FAC amostral (c).

```
# MA(1) Simulado
set.seed(12322)
ma1_m08=arima.sim(n=200,model=list(ma=c(-0.8)))
plot(ma1_m08,type="l",lwd=2,cex.lab=1.4, xlab="tempo",ylab="", main="(a)", cex.main=1.7,col="cadetblue4")
acf(ma1_m08,type = "correlation",lag.max = 50,cex.main=1.7,col="cadetblue4",ylab="FAC",lwd=4,main="(b)")
acf(ma1_m08,type = "partial" ,lag.max = 50,cex.main=1.7,col="cadetblue4",ylab="FACP",lwd=4,main="(c)")
```

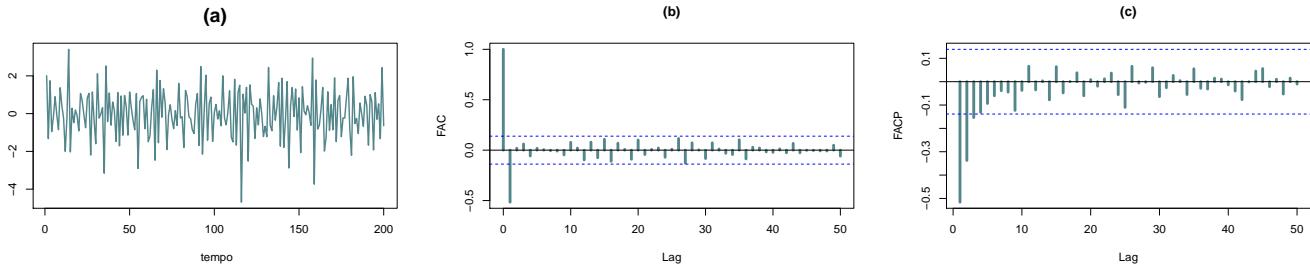


Figure 4.11: MA(1) simulado com coeficiente $\theta_1 = -0.8$, (a) FACV amostral (b) e FAC amostral (c).

```
# MA(1) Simulado
set.seed(1234)
ma1_m0804=arima.sim(n=300,model=list(ma=c(-0.8,0.4)))
plot(ma1_m0804,type="l",lwd=2,cex.lab=1.4, xlab="tempo",ylab="", main="(a)", cex.main=1.7,col="cadetblue4")
acf(ma1_m0804,type = "correlation",lag.max = 50,cex.main=1.7,col="cadetblue4",ylab="FAC",lwd=4,main="(b)")
acf(ma1_m0804,type = "partial" ,lag.max = 50,cex.main=1.7,col="cadetblue4",ylab="FACP",lwd=4,main="(c)")
```

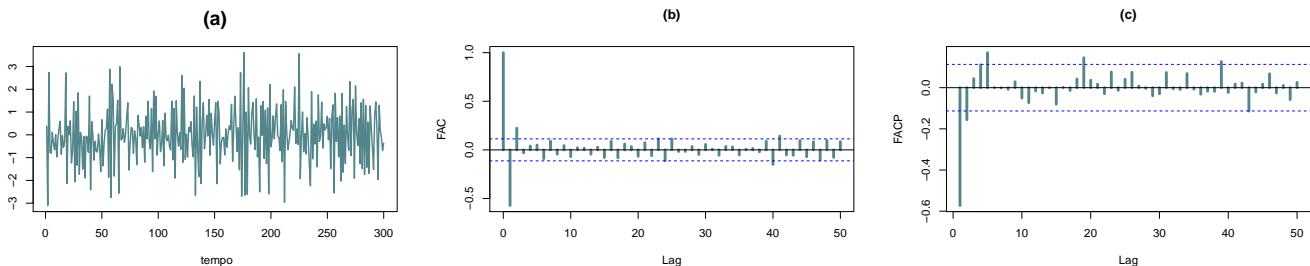


Figure 4.12: MA(2) simulado com coeficiente $\theta_1 = -0.8$ e $\theta_2 = 0.4$, (a) FACV amostral (b) e FAC amostral (c).

```
# MA(1) Simulado
set.seed(12341)
ma1_m080414=arima.sim(n=300,model=list(ma=c(-0.8,0.4,1.4)))
```

```
plot(ma1_m080414,type="l",lwd=2,cex.lab=1.4, xlab="tempo",ylab="", main="(a)", cex.main=1.7,col="cadetblue")
acf(ma1_m080414,type = "correlation",lag.max = 50,cex.main=1.7,col="cadetblue4",ylab="FAC",lwd=4,main="(b)")
acf(ma1_m080414,type = "partial" ,lag.max = 50,cex.main=1.7,col="cadetblue4",ylab="FACP",lwd=4,main="(c)")
```

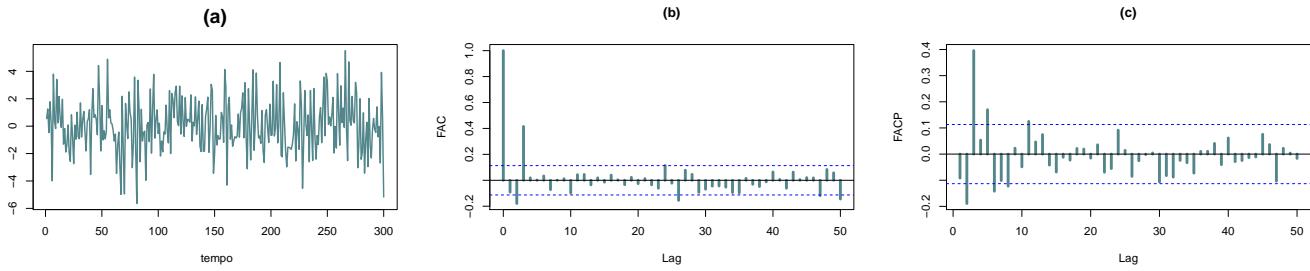


Figure 4.13: MA(2) simulado com coeficiente $\theta_1 = -0.8$, $\theta_1 = 0.4$ e $\theta_1 = 1.4$, (a) FACV amostral (b) e FAC amostral (c).

4.5.7 Modelo ARMA(p,q)

Um modelo mais geral é dado pela aglutinação dos modelos AR e MA em um único modelos, ao qual chamamos Modelos Autoregressivos de Média Móveis, abreviado ARMA. Um processo $\{Y_t\}_t$ é um ARMA(p, q) se pode ser escrito como

onde $\varepsilon_t \sim RB(0, \sigma_\varepsilon^2)$, $\phi_1, \dots, \phi_p \in \mathbb{R}$ são os coeficientes da parte AR e $\theta_1, \dots, \theta_q \in \mathbb{R}$ são os coeficientes da parte MA do processo. Utilizando os polinômios AR e MA vistos anteriormente, podemos escrever um processo ARMA(p, q) de uma forma compacta e elegante.

\textcolor{blue}{O modelo ARMA(p,q)}

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (4.15)$$

em que $\varepsilon_t \sim RB(0, \sigma_\varepsilon^2)$, $\phi_1, \dots, \phi_p \in \mathbb{R}$ são os coeficientes da parte AR e $\theta_1, \dots, \theta_q \in \mathbb{R}$ são os coeficientes da parte MA do processo.

Utilizando os polinômios AR e MA vistos anteriormente, podemos escrever um processo ARMA(p, q) de uma forma compacta e elegante.

$$\Phi_p(L)Y_t = \Theta_q(L)\varepsilon_t,$$

em que ε_t é um $RB(0, \sigma_\varepsilon^2)$, $\Phi_p(L)$ e $\Theta_p(L)$ são polinômios da parte AR e MA (respectivamente) dados por

$$\Phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p \quad \text{e} \quad \Theta_q(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q.$$

Exemplo 4.2 (Modelo ARMA(2,3)). O Modelo ARMA(2,3) é escrito como

$$\begin{aligned} \Phi_2(L)y_t &= \Psi_3(L)\varepsilon_t \\ (1 - \phi_1 L - \phi_2 L^2)y_t &= (1 + \theta_1 L + \theta_2 L^2 + \theta_3 L^3)\varepsilon_t \\ y_t &= \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3}. \end{aligned}$$

4.5.7.1 Exemplos ARMA simulados

Exemplo 4.3 (Causalidade do processo AR(1)). O modelo AR(1):

$$y_t = \phi y_{t-1} + e_t,$$

pode ser escrito como

$$y_t = e_t + \phi e_{t-1} + \phi^2 e_{t-2} + \cdots + \phi^{k-1} e_{t-(k-1)} + \phi^k y_{t-k},$$

em que para k grande tem-se

$$\begin{aligned} y_t &= e_t + \phi e_{t-1} + \phi^2 e_{t-2} + \dots \\ &= \psi_0 e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \dots, \end{aligned}$$

em que $|\phi| < 1$ e $\psi_j = \phi^j$. O que acontece com a variância de y_t ? Assim, essa representação somente faz sentido se $\sum_{j=0}^{\infty} \psi_j < \infty$, o que ocorre se, e somente se, $|\phi| < 1$.

4.5.8 Causalidade

O conceito de causalidade consiste em escrever um processo AR(q) como um MA(∞).

Um processo linear $\{Y_t\}$ é **CAUSAL** (estritamente, uma função causal de $\{\varepsilon_t\}$) se existem reais ψ_0, ψ_1, \dots satisfazendo $\sum_{j=0}^{\infty} |\psi_j| < \infty$ e tais que

$$Y_t = \psi_0 \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k} = \Psi(L) \varepsilon_t, \quad (4.16)$$

onde denotamos

$$\Psi(L) = \psi_0 + \psi_1 L + \psi_2 L^2 + \cdots = \sum_{k=0}^{\infty} \psi_k L^k.$$

O modelo AR(1) é dado por:

$$Y_t = \phi Y_{t-1} + \varepsilon_t,$$

para $|\phi| < 1$ e $\varepsilon_t \sim RB(0, \sigma_{\varepsilon}^2)$. Na Seção ?? obtivemos a representação (com $c = 0$)

$$Y_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}.$$

Identificando-se $\psi_j = \phi^j$, obtemos que Y_t possui a representação (4.16) e observando que, como $|\phi| < 1$,

$$\sum_{j=0}^{\infty} |\psi_j| = \sum_{j=0}^{\infty} |\phi|^j = \frac{1}{1 - |\phi|} < \infty,$$

segue que um AR(1) com $|\phi| < 1$ é causal.

Obviamente todo modelo MA(q) é causal, dado que se

$$Y_t = \varepsilon_t + \theta \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

com $\varepsilon_t \sim RB(0, \sigma_{\varepsilon}^2)$, tomamos $\psi_0 = 1$, $\psi_j = \theta_j$, para $j = 1, \dots, q$ e $\psi_j = 0$ para $j > q$, temos que a representação (4.16) é satisfeita e, evidentemente, $\sum_{j=0}^{\infty} |\psi_j| = 1 + |\theta_1| + \cdots + |\theta_q| < \infty$.

4.5.9 Invertibilidade

Mostramos que um processo AR pode ser reescrito como um processo MA de ordem infinita através de pesos ψ_j 's. Podemos nos perguntar quando (e se) é possível escrever um processo MA como um autorregressivo.

Um processo linear $\{Y_t\}$ é dito ser **INVERTÍVEL** (estritamente, uma função invertível de $\{\varepsilon_t\}$) se existem reais $\text{Varphi}_1, \text{Varphi}_2, \dots$ satisfazendo $\sum_{j=0}^{\infty} |\text{Varphi}_j| < \infty$ e tais que

$$\varepsilon_t = \text{Varphi}_0 Y_t + \text{Varphi}_1 Y_{t-1} + \text{Varphi}_2 Y_{t-2} + \dots = \sum_{k=0}^{\infty} \text{Varphi}_k Y_{t-k} = \Phi(L)Y_t, \quad (4.17)$$

onde denotamos

$$\Phi(L) = \text{Varphi}_0 + \text{Varphi}_1 L + \text{Varphi}_2 L^2 + \dots = \sum_{k=0}^{\infty} \text{Varphi}_k L^k.$$

Exemplo 4.4 (Invertibilidade do processo MA(1)). Considere o modelo MA(1)

$$Y_t = \varepsilon_t - \theta \varepsilon_{t-1},$$

em que ε_t é um $RB(0, \sigma^2)$. Reescrevendo a equação acima como

$$\varepsilon_t = Y_t + \theta \varepsilon_{t-1}$$

e substituindo t por $t-1$ e ε_{t-1} na equação modificada, temos:

$$\begin{aligned} \varepsilon_t &= Y_t + \theta(Y_{t-1} + \theta \varepsilon_{t-2}) \\ &= Y_t + \theta Y_{t-1} + \theta^2 Y_{t-2} \end{aligned}$$

Se $|\theta| < 1$, podemos continuar a substituição e obter:

$$\varepsilon_t = Y_t + \theta Y_{t-1} + \theta^2 Y_{t-2} + \dots = \sum_{j=0}^{\infty} \theta^j Y_{t-j}.$$

Assim, da mesma forma como foi feito para o AR(1), tomando $\text{Varphi}_j = \theta^j$, segue que, se $|\theta| < 1$, a representação (??) é satisfeita e $\sum_{j=0}^{\infty} |\text{Varphi}_j| = \frac{1}{1-|\theta|} < \infty$ de onde concluimos que o modelo MA(1) é invertível. Em outras palavras, um modelo MA(1) pode ser invertido (transformado) para um AR(∞), sempre que $|\theta| < 1$.

4.5.10 Polinômio característico

Nos exemplos mostrados acima tratamos da causalidade e invertibilidade dos casos AR(1) e MA(1) em particular. Para os casos mais gerais AR(p) e MA(q) utilizamos os chamados **polinômios característicos** para decidir se os processos são causais e/ou invertíveis.

Para um modelo geral AR(p), definimos o **polinômios característicos AR** como

$$\Phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p, \quad z \in \mathbb{C}$$

Teorema 4.1. *Uma (única) solução estacionária para $\Phi(L)y_t = e_t$ existe se, e somente, as raízes de $\Phi(z)$ não pertence ao círculo de raio um, ou seja,*

$$|z| = 1 \rightarrow \Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p \neq 0.$$

O processo AR(p) é **causal** se, e somente se as raízes de $\Phi(z)$ estão **fora do círculo unitário**, ou seja,

$$|z| \leq 1 \rightarrow \Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p \neq 0.$$

Para um modelo geral MA(q), definimos o **polinômio característico MA** como

$$\Theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q.$$

Teorema 4.2. Um processo MA(q) é invertível se, e somente se, as raízes de $\Theta(z)$ estão fora do círculo unitário, isto é,

$$z \in \mathbb{C} : |z| \leq 1 \Rightarrow \Theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q \neq 0.$$

Observação. Um processo ARMA será invertível e estacionário se a parte AR o for, e será invertível se a parte MA o for.

4.5.11 Estacionariedade e causalidade de um processo ARMA

Para um processo ARMA, as condições para causalidade, invertibilidade e estacionariedade são dadas no seguinte teorema.

Teorema 4.3. Se $\Phi(\cdot)$ e $\Theta(\cdot)$ não possuem fatores em comum, existe uma única solução estacionária $\{Y_t\}$ para $\Phi(L)Y_t = \Theta(L)\varepsilon_t$ se, e somente se,

$$z \in \mathbb{C} : |z| = 1 \Rightarrow \Phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0.$$

Esse processo ARMA(p, q) é causal se, e somente se,

$$z \in \mathbb{C} : |z| \leq 1 \Rightarrow \Phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0.$$

Será invertível se, e somente se

$$z \in \mathcal{C} : |z| \leq 1 \Rightarrow \Theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q \neq 0.$$

4.6 Exercícios

Exercício 4.1. Defina processo estocástico e ilustre graficamente. Explique o que é a realização de um processo estocástico e por que séries econômicas podem ser entendidas como geradas por um processo estocásticos.

Exercício 4.2. Seja $\{y_t\}_{t=1}^T$ uma série temporal. Quais características essa série deve apresentar para ser considerada uma série de covariância estacionária?

Exercício 4.3. Faça os seguintes items:

- a) Defina o que é um processo ruído branco.
- b) Defina o que é um processo independente e identicamente distribuído (i.i.d.).
- c) Defina ruído branco Gaussiano.
- d) Qual a relação entre ruído branco, ruído branco Gaussiano e processo i.i.d.?
- e) Esses processos são estacionários?

Exercício 4.4. Considere um processo MA(1): $y_t = e_t + \alpha_1 e_{t-1}$; onde $e_t \sim RB(0, \sigma_e^2)$.

- a) Calcule a média e variância de y_t .
- b) Calcule as autocovariâncias de lags 1 e 2 para a série y_t .
- c) Esse processo é estacionário? (Justifique sua resposta usando os valores encontrados nos itens anteriores juntamente com o conceito de estacionariedade definido na Questão 1).
- d) Comente a afirmativa: “Todo processo MA(q), onde $q < \infty$, é estacionário”.
- e) Suponha que $\alpha_1 = 0.5$. O processo é invertível?
- f) Calcule a autocorrelação de ordem 1 para o processo do item anterior e faça o gráfico da FAC com 5 lags.

Exercício 4.5. Considere um processo MA(2): $y_t = e_t + \alpha_1 e_{t-1} + \alpha_2 e_{t-2}$; onde $e_t \sim RB(0, \sigma_e^2)$.

a) Calcule a média e variância de y_t .

b) Calcule as autocovariâncias de *lags* 1, 2 e 3 para a série y_t .

c) Esse processo é estacionário? (Justifique sua resposta usando os valores encontrados nos itens anteriores juntamente com o conceito de estacionariedade definido na Questão 1).

d) Suponha que $\alpha_1 = 0.65$ e que $\alpha_2 = -0.20$. O processo é invertível?

e) Calcule a autocorrelação de ordem 1 e 2 para o processo do item anterior e faça o gráfico da FAC com 5 *lags*.

Exercício 4.6. Considere os seguintes processos

$$y_t = e_t + \theta e_{t-1} \quad \text{e} \quad y_t = e_t + \frac{1}{\theta} e_{t-1},$$

onde $e_t \sim iid(0, \sigma_e^2)$ e $\theta \neq 0$.

a) Os processos acima possuem as mesmas autocorrelações? Verifique.

b) Os processos acima são invertíveis? Verifique.

Exercício 4.7. Considere um processo AR(1): $y_t = 5 + 0.9y_{t-1} + e_t$, onde $e_t \sim RB(0, \sigma_e^2)$.

a) Esse processo é estacionário? Verifique.

b) Calcule as autocorrelações de ordem 1, 2 e 3 para esse processo. Faça um esboço do gráfico da FAC para esse processo com 5 *lags*.

c) O que significa o coeficiente de y_{t-1} num processo AR(1)?

d) Faça um gráfico da FACP desse processo com 5 *lags*.

“`{exercise ar123} Para os seguintes itens, responda:

a) Explique como se comportam os gráficos da FAC e da FACP em processos AR(p) e em processos MA(q).

b) Esboce os gráficos da FAC e FACP para os seguintes processos: AR(1), AR(3), MA(2) e MA(3).


```{exercise exarmafac} Para os seguintes itens, responda:

- | a\`) Supondo que  $E(y_t) = \mu$  e que  $y_t = c_0 + \beta_1 y_{t-1} + e_t + \alpha_1 e_{t-1}$ , calcule o valor de  $c_0$ .
- | b\`) Explique como se comportam os gráficos da FAC e da FACP em processos ARMA( $p, q$ ).
- | c\`) Esboce os gráficos da FAC e FACP para um processos ARMA(1,1).

**Exercício 4.8.** Explique os passos que devem ser seguidos para a modelagem de uma série temporal na metodologia ARMA.

**Exercício 4.9** (ANPEC 2014-5). Suponha que  $Y_t$  seja representado pelo seguinte processo autorregressivo de primeira ordem:

$$Y_t = 10 + 0,6Y_{t-1} + e_t,$$

em que  $e_t$  é um ruído branco que satisfaz as condições:  $E(e_t) = 0$ ,  $E(e_t^2) = \sigma^2$ ,  $E(e_t e_s) = 0$  para  $t \neq s$ . Suponha também que  $Y_0 = 0$ . Obtenha  $E(Y_t)$  para  $t = 2$ .

**Exercício 4.10** (ANPEC 2014-10). Considere o seguinte processo:

$$Y_t = \rho Y_{t-1} + e_t, \quad t = 1, 2, \dots,$$

em que  $Y_0 = 0$  e  $e_t$  é um ruído branco que satisfaz as condições:  $E(e_t) = 0$ ,  $E(e_t^2) = \sigma^2$ ,  $E(e_t e_s) = 0$  para  $t \neq s$ . São corretas as afirmativas:

- 0) Se  $\rho = 1$ ,  $\mathbb{E}(Y_t) = 0$  para todo  $t$ ;
- 1) Se  $\rho = 1$ ,  $\text{Var}(Y_t) = t$  para todo  $t$ ;
- 2) Se  $\rho = 1$ ,  $\mathbb{E}(Y_{t+h}/Y_t) > Y_t$  para todo  $h \geq 1$ ;
- 3) Se  $|\rho| < 1$ ,  $\text{Var}(Y_t) = 1$ ;
- 4) Se  $|\rho| < 1$ ,  $\mathbb{E}(Y_{t+h}/Y_t) = \rho^h Y_t$  para todo  $h \geq 1$ .

**Exercício 4.11** (ANPEC 2013-10). Julgue as seguintes afirmativas:

- 0) O passeio aleatório com drift,  $y_t = c + y_{t-1} + \varepsilon_t$ ,  $y_0 = 0$ , em que  $\varepsilon_t$  é um ruído branco, com média zero e variância  $\sigma^2$ , é um processo estacionário de segunda ordem se  $c = 0$ .
- 1) O processo MA(1),  $y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$ , em que  $\varepsilon_t$  é um ruído branco, com média zero e variância  $\sigma^2$ , é estacionário de segunda ordem se, e somente se, a raiz do polinômio  $1 + \theta_1 x$  cair fora do círculo unitário.
- 2) O processo MA(1),  $y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$ , em que  $\varepsilon_t$  é um ruído branco, com média zero e variância  $\sigma^2$ , é inversível se, e somente se,  $|\theta_1| < 1$ .
- 3) O processo AR(2),  $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$ , em que  $\varepsilon_t$  é um ruído branco, com média zero e variância  $\sigma^2$ , é estacionário de segunda ordem se

$$|\phi_2| < 1, \quad \phi_2 - \phi_1 < 1 \quad \text{e} \quad \phi_2 + \phi_1 < 1.$$

- 4) No passeio aleatório,  $y_t = y_{t-1} + \varepsilon_t$ ,  $y_0 = 0$ , em que  $\varepsilon_t$  é um ruído branco, com média zero e variância  $\sigma^2$ , a variância de  $y_t$  varia com  $t$ .

**Exercício 4.12** (ANPEC 2013-13). Considere o seguinte processo  $x_t = \mu + e_t + \alpha_1 e_{t-1}$ , para  $t = 1, 2, \dots$ , no qual  $e_t$  é uma sequência i.i.d com média 0 e variância  $\sigma_e^2$ . Julgue as seguintes afirmativas:

- 0)  $\text{Var}[x_t] = (1 + \alpha_1^2)\sigma_e^2$ .
- 1)  $\text{Cov}(x_t, x_{t+h}) = 0$ ,  $h > 1$ .
- 2)  $\mathbb{E}[x_t] = \mu + t$ .
- 3) O processo descrito acima é estacionário em covariância.
- 4) A função de autocorrelação deste processo é:  $\rho_1 = \frac{\alpha_1}{1 + \alpha_1^2}$  e  $\rho_j = 0$  para  $j > 1$ .

**Exercício 4.13** (ANPEC 2012-08). Suponha que  $Y_t$  seja descrito por um processo auto-regressivo de ordem 3, isto é,

$$Y_t = Y_{t-1} - 0,50Y_{t-3} + u_t$$

e que

$$u_t | Y_{t-j} \sim N(0, \sigma^2), \quad \forall j > 0.$$

Calcule a correlação entre  $Y_t$  e  $Y_{t-2}$ . Multiplique o resultado por 100.

**Exercício 4.14** (ANPEC 2011-11). Julgue as seguintes afirmativas:

- 0) O processo AR(2),  $y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \varepsilon_t$ , em que  $\varepsilon_t$  é um ruído branco com média zero e variância  $\sigma^2$ , é estacionário de segunda ordem se e somente se as raízes do polinômio  $x^2 - \rho_1 x + \rho_2$  estão fora do círculo unitário.
- 1) No processo MA(2),  $y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$ , em que  $\varepsilon_t$  é um ruído branco com média zero e variância  $\sigma^2$ , a covariância entre  $y_t$  e  $y_{t-3}$  é igual a zero.
- 2) No passeio aleatório com drift,  $y_t = c + y_{t-1} + \varepsilon_t$ ,  $y_0 = 0$ , em que  $\varepsilon_t$  é um ruído branco com média zero e variância  $\sigma^2$ , a média de  $y_t$  varia com  $t$ .
- 3) No processo MA(1),  $y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$ , em que  $\varepsilon_t$  é um ruído branco com média zero e variância  $\sigma^2$ , a correlação entre  $y_t$  e  $y_{t-1}$  é menor ou igual a 0,5 em valor absoluto.
- 4) O processo ARMA(1,1),  $y_t = \rho y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$ , em que  $\varepsilon_t$  é um ruído branco com média zero e variância  $\sigma^2$ , é estacionário de segunda ordem se e somente se  $|\rho| < 1$  e  $|\theta| < 1$ .

**Exercício 4.15** (ANPEC 2009-15). É correto afirmar que:

0) No processo AR(1),  $y_t = \phi_0 + \phi_1 y_{t-1} + e_t$ , em que  $\phi_1 < 1$  e  $e_t$  é um ruído branco de média nula e variância  $\sigma^2$ , a média de  $y_t$  será igual a  $\phi_0$ .

1) O processo MA(1),  $y_t = e_t + \theta e_{t-1}$ , em que  $e_t$  é um ruído branco de média nula e variância constante, será estacionário mesmo que  $\theta > 1$ .

2) Seja a função de autocorrelação do processo AR(1) definido no item (0) dada por  $\rho_j = \phi_1^j$ . É correto afirmar que  $\rho_j = \phi_1^j$ .

3) O processo AR(2),  $y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t$ , em que  $e_t$  é um ruído branco de média nula e variância  $\sigma^2$ , será estacionário de segunda ordem se, e somente se,  $\phi_1 < 1$  e  $\phi_2 < 1$ .

4) No modelo ARMA(1,1),  $y_t = \phi_0 + \phi_1 y_{t-1} + e_t + \theta e_{t-1}$ , em que  $e_t$  é um ruído branco de média nula e variância constante ( $\sigma^2$ ), a variância de  $y_t$  é dada por  $\frac{\sigma^2(1+\theta^2)}{1-\phi^2}$ .

**Exercício 4.16.** Considere uma série temporal com 200 observações. A Figura 4.14 mostra a evolução da série ao longo do tempo. A tabela a seguir fornece as autocorrelações,  $\rho$ 's, e autocorrelações parciais,  $\phi$ 's, estimados a partir dessa série.

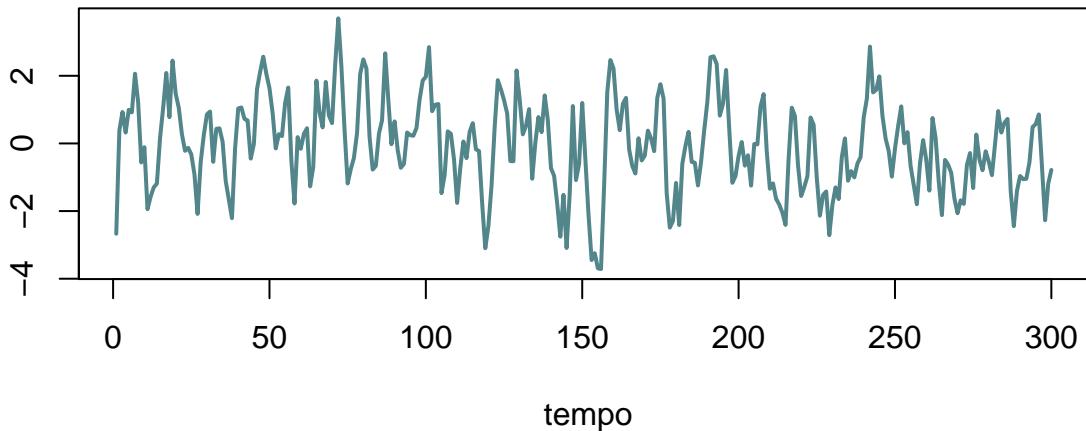


Figure 4.14: Série temporal simulada

| $k$          | 1    | 2     | 3    | 4    | 5     | 6     | 7    | 8     | 9    | 10    |
|--------------|------|-------|------|------|-------|-------|------|-------|------|-------|
| $\rho_k$     | 0.51 | 0.13  | 0.01 | 0.04 | 0.03  | 0.00  | 0.04 | 0.02  | 0.08 | 0.01  |
| $\phi_{k,k}$ | 0.51 | -0.18 | 0.03 | 0.06 | -0.03 | -0.00 | 0.07 | -0.05 | 0.13 | -0.11 |

- a) Analisando a Figura 4.14 a série parece ser estacionária? Explique.
- b) Faça o gráfico da FAC e FACP para esse processo.
- c) Calcule o critério para decisão quanto à significância das autocorrelações estimadas e represente esse critério nos gráficos da FAC e FACP.
- d) Qual(is) modelo(s) você propõe para ajustar essa série temporal? Justifique.

**Exercício 4.17.** Usando a esperança condicional, calcule as previsões 1, 2 e 3 passos a frente ( $\hat{y}_T(1)$ ,  $\hat{y}_T(2)$ ,  $\hat{y}_T(3)$ ) para os seguintes processos:

- a) AR(1);

- b) AR(2);
- c) MA(1);
- d) MA(3);
- e) ARMA(1,1);
- f) ARMA(2,2).

**Exercício 4.18.** Abaixo (Figura 2) encontram-se os gráficos da FAC e FACP calculados para uma série  $\{y_t\}_{t=1}^{200}$ .

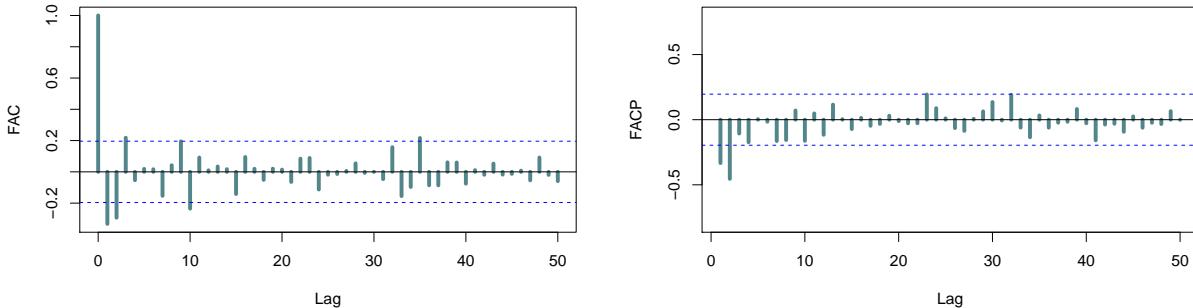


Figure 4.15: FAC e FACP de uma série temporal simulada

- a) Analisando a Figura 4.15 a série parece ser estacionária? Explique.
- b) Usando os gráficos da FAC e FACP, qual(is) modelo(s) você propõe para ajustar essa série temporal? Justifique. (Note que o primeiro *lag* é o 1 em ambos os gráficos).

# Chapter 5

## Séries temporais não estacionárias

No capítulo anterior estudamos processos estacionários, ou seja, processos satisfazendo

$$\begin{cases} \mathbb{E}(Z_t) = 0; \\ \text{Var}(Z_t) = \sigma^2, \text{ para todo } t; \\ \gamma(k) = \text{Cov}(Z_t, Z_{t-k}) \text{ não depende de } t, \text{ somente de } k. \end{cases}$$

No entanto muitas séries temporais econômicas são claramente não estacionárias no sentido de que a média, variância e/ou estrutura de covariância dependem do tempo. Uma série com estas características tende a se afastar permanentemente de qualquer valor à medida que o tempo passa. Fontes comuns de não estacionariedade em séries temporais são tendências, sazonalidades e quebras estruturais diversas. Destas, as mais simples de lidar são as tendências e sazonalidades. Uma série é dita apresentar uma tendência determinística se esta se desenvolve ao redor de uma função determinística, geralmente simples. A Figura 5.1 apresenta alguns diferentes tipos de tendências determinísticas: linear, logarítmica, quadrática e exponencial (veja também a Figura 4.2).

Da Figura 5.1 fica clara que uma série apresentando tendência determinística é não-estacionária: de imediato percebe-se que a média varia com o tempo em todos os casos apresentados. Antes que qualquer tipo de análise adicional possa ser feita, em especial, qualquer tipo de modelagem e previsão utilizando os modelos vistos até aqui, é obrigatória a remoção de tendências. Existem dois tipos fundamentais de tendências que serão estudadas adiante. Nos concentraremos inicialmente na remoção de tendências determinísticas.

### 5.1 Como lidar com tendências determinísticas

Existem várias maneiras de eliminarmos tendências determinísticas. Neste trabalho apresentaremos uma metodologia paramétrica de estimativa da tendência determinística em uma série. Primeiramente é importante observar que, neste contexto, a forma funcional da tendência determinística deve ser identificada e especificada a priori. Uma maneira muito simples e útil para a remoção da tendência é a inclusão de uma função da variável tempo no modelo, geralmente carregando informações sobre o formato da tendência que se quer remover. Assumiremos que a forma funcional da tendência determinística depende de certos parâmetros de forma linear. Podemos dar alguns exemplos de modelos com tendência determinística: o modelo

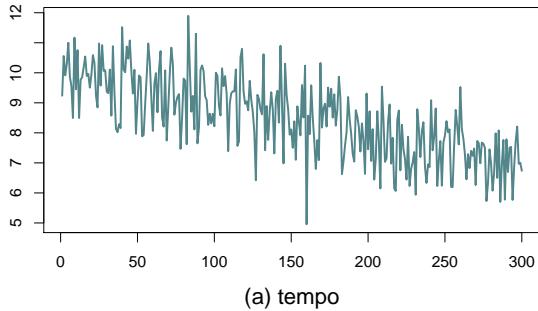
$$Y_t = a + bt + \varepsilon_t \quad (5.1)$$

em que  $\varepsilon_t \sim RB(0, \sigma_\varepsilon^2)$ , torna-se um ruído branco com tendência determinística. O modelo AR(1) com tendência logarítmica pode ser escrito da seguinte forma

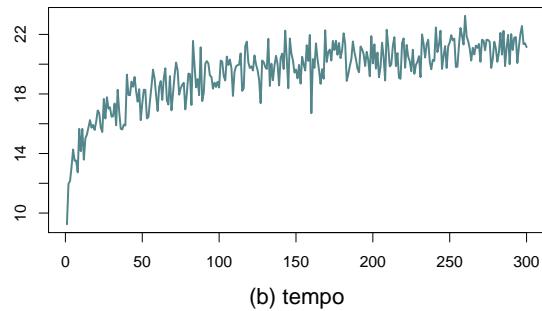
$$Y_t = a + b \ln(t) + \phi Y_{t-1} + \varepsilon_t. \quad (5.2)$$

Nestes casos, acrescentamos uma tendência funcional ao processo, linear nos parâmetros, e procedemos a estimativa desta tendência via MQO. Vejamos alguns exemplos.

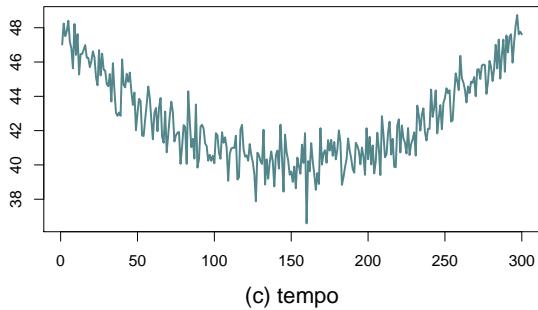
**Exemplo 5.1.** Os dados referentes à receita nominal mensal de vendas do varejo nacional no ramo de combustíveis e lubrificantes (índice de base fixa, sendo o ano de referência 2003 com valor 100) no período de janeiro de 2000 à dezembro de 2011 estão apresentados na Figura 5.2(a) (fonte: IBGE, Pesquisa Mensal do Comércio 2000/jan-2011/dez).

**Série temporal com tendência linear decrescent**

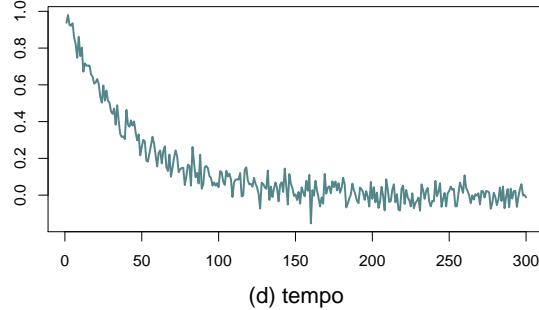
(a) tempo

**Série temporal com tendência logarítmica**

(b) tempo

**Série temporal com tendência quadrática**

(c) tempo

**Série temporal com tendência exponencial**

(d) tempo

Figure 5.1: Séries não-estacionárias apresentando tendências determinísticas: (a) Tendência linear, (b) tendência logarítmica, (c) tendência quadrática e (d) tendência exponencial.

Observe que a série apresenta uma nítida tendência linear crescente. Para removê-la, vamos assumir que a série é da forma

$$Y_t = \alpha_0 + \alpha_1 t + X_t,$$

onde  $Y_t$  denota o receita nominal no tempo  $t$  e  $X_t$  é a série residual apóis removê-la. Denotando por  $x_1, \dots, x_{144}$  os dados, procedemos com a estimação de  $\alpha_0$  e  $\alpha_1$  utilizando MQO que, neste caso, resulta  $\hat{\alpha}_0 = 68.668$  e  $\hat{\alpha}_1 = 0.568$  ambos altamente significativos (p-valores muito próximos de zero). Na Figura 5.2(b) apresentamos os dados e a reta ajustada (o eixo  $x$  foi reescalado para refletir os meses). Para obtermos a série residual  $X_t$  tomamos, naturalmente,  $\hat{X}_t = Y_t - (68.668 + 0.568t)$ . Na Figura 5.2(c) apresentamos a reta residual, com o eixo  $x$  reescalado para refletir as datas da série.

Utilizando MQO podemos ainda remover qualquer tipo de função do tempo que seja linear nos parâmetros, como mostra o exemplo abaixo.

**Exemplo 5.2.** Os dados referentes à série mensal de pessoas desocupadas com idade superior a 11 anos em Porto Alegre entre março de 2002 e outubro de 2015 estão apresentados na Figura 5.3(a). Os dados apresentam o coeficiente de variação mensal relativo para o número de pessoas sem trabalho mas que estavam disponíveis para assumir um trabalho e que tomaram alguma providência efetiva para conseguir trabalho no período de referência de 30 dias, sem terem tido qualquer trabalho ou após terem saído do último emprego que tiveram nesse período (fonte: IBGE, Pesquisa Mensal de Emprego). A série apresenta uma distinta tendência logarítmica ao longo do tempo, que pode ser modelada por

$$Y_t = \alpha_0 + \alpha_1 \ln(t) + X_t,$$

onde  $Y_t$  denota o coeficiente de variação do número de pessoas desocupadas no tempo  $t$  e  $X_t$  é a série residual apóis removê-la. A estimação dos coeficientes via MQO fornecem  $\hat{\alpha}_0 = 28,5875$  e  $\hat{\alpha}_1 = 5,7317$ , altamente significativos (p-valor próximo a zero). A reta ajustada (função) é dada por  $17.4652 + 23.1394 * \log(t)$  e está apresentada na Figura 5.3(b), junto com a série original. Na Figura 5.3(c) apresentamos o residual.

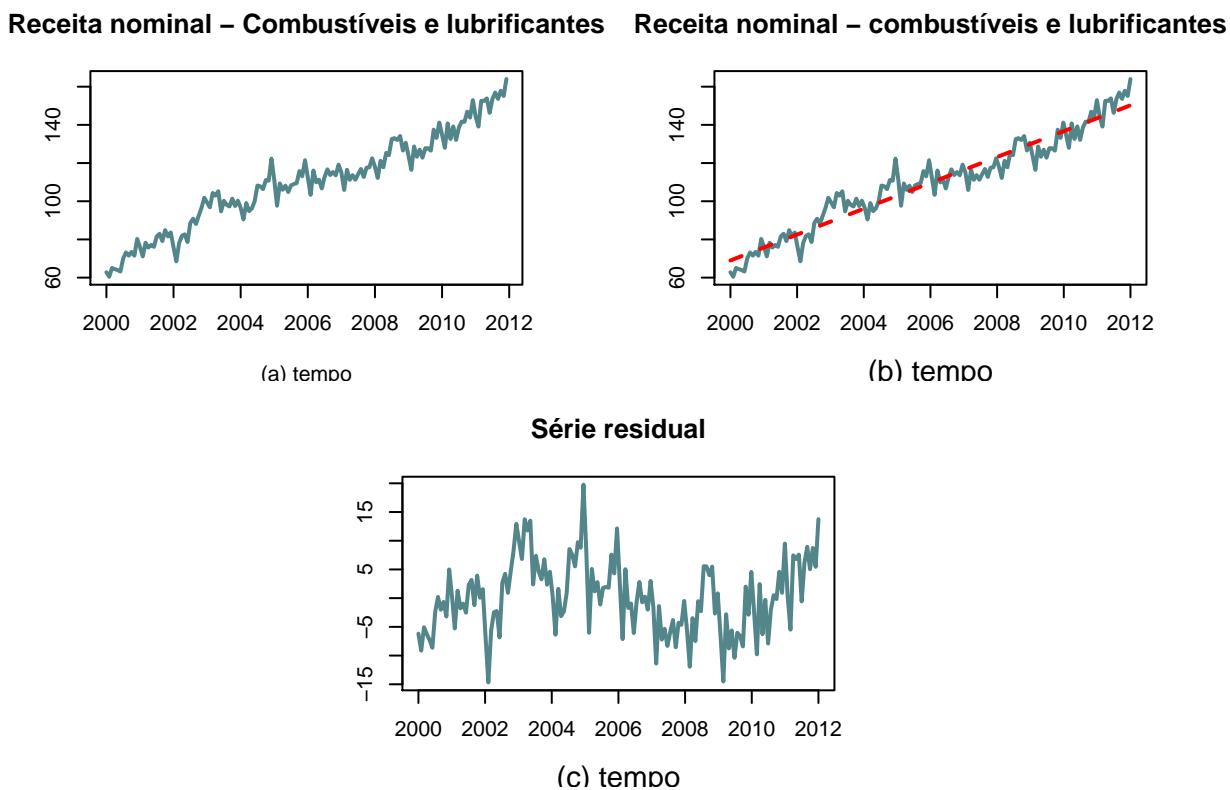


Figure 5.2: Séries da receita nominal mensal de vendas do varejo nacional no ramo de combustíveis e lubrificantes. (a) Série, (b) série e reta ajustada e (c) residual.

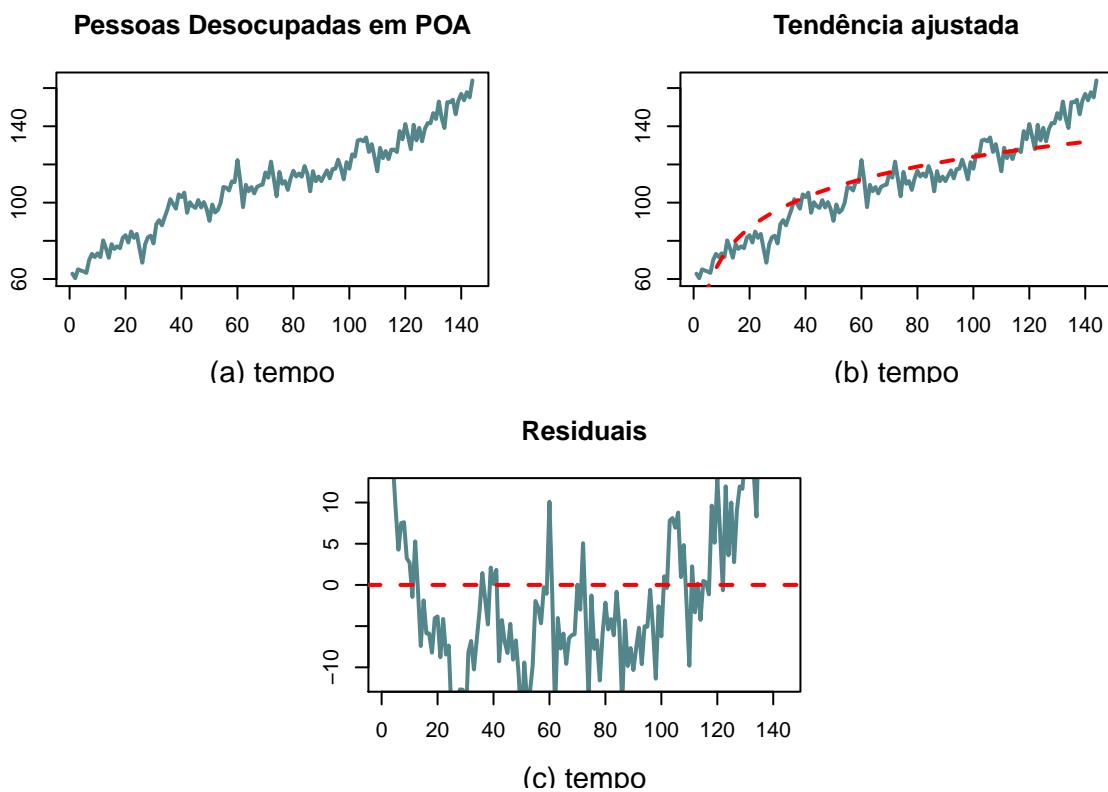


Figure 5.3: Séries do número de pessoas desocupadas em Porto Alegre. (a) Série, (b) série e tendência ajustada e (c) residual

## 5.2 Testes de raiz unitária

### 5.2.1 Identificando tendência estocástica

Uma série com uma tendência estocástica se diferencia de outra com uma tendência determinística, pois as mudanças na mesma deixam de ter um caráter transitório e passam a apresentar um caráter permanente (Gujarati, 2003) e [(Pereira, 1988)].

A presença de uma tendência estocástica implica que flutuações em uma série temporal são o resultado de choques não somente no componente transitório ou cíclico, mas também no componente de tendência. [Balke (1991) e (Gujarati, 2003) ]

Como vimos nas sessões anteriores, para um processo ARMA ser estacionário, o polinômio característico da parte AR não pode conter raízes de módulo igual a um, chamadas de raízes unitárias. Acontece que a presença de raízes unitárias no polinômio AR resulta na presença de tendência estocástica na série. A identificação de raízes unitárias é de grande importância na análise de séries temporais, e este fato se reflete na literatura relativamente longa tratando do assunto. Várias abordagens para a detecção de raízes unitárias estão à nosso dispor. Um dos testes mais utilizados na literatura é o teste de Dickey Fuller que veremos a seguir.

### 5.2.2 Teste de Dickey Fuller (*DF*)

Considere o modelo autorregressivo de ordem 1, AR(1)

$$Y_t = a_0 + \rho Y_{t-1} + \varepsilon_t \quad (5.3)$$

em que  $Y_t$  é a variável de interesse,  $t$  é o índice temporal,  $\rho$  é coeficiente e  $\varepsilon_t$  é o termo de erro. Uma raiz unitária está presente se  $\rho = 1$  implicando que o modelo será não estacionário.

Nota-se que, quando  $\rho = 1$

$$Y_t = a_0 + Y_{t-1} + \varepsilon_t$$

% pode ser reescrito como

$$Y_t = Y_0 + \sum_{i=1}^t \varepsilon_i + a_0 t$$

com uma tendência determinística vindo de  $a_0 t$  e um intercepto estocástico vindo de  $Y_0 + \sum_{i=1}^t \varepsilon_i$ , resultando no que chamamos de tendência estocástica. O teste de Dickey Fuller consiste em fazer um teste t (mas com distribuição de Dickey-Fuller) para a significância do seguinte modelo

#### Teste de Dickey Fuller

$$\Delta Y_t = (\rho - 1)Y_{t-1} + \varepsilon_t = \delta Y_{t-1} + \varepsilon_t,$$

- $H_0$ :  $\delta = 0$  (Não estacionário)
- $H_1$ :  $\delta < 0$  (Estacionário)

em que  $\Delta$  é o operador de diferenciação, dado por  $\Delta Y_t = Y_t - Y_{t-1}$ . Testar a presença de raiz unitária neste modelo ( $\rho = 1$ ) é equivalente a atestar se  $\delta = \rho - 1 = 0$ . Como o teste é feito sobre os resíduos, a distribuição de um teste t usual não será usual, nem mesmo assintoticamente. Para isso existe uma estatística de teste específica,  $\tau$ , cujos valores críticos estão dispostos na tabela de Dickey Fuller.

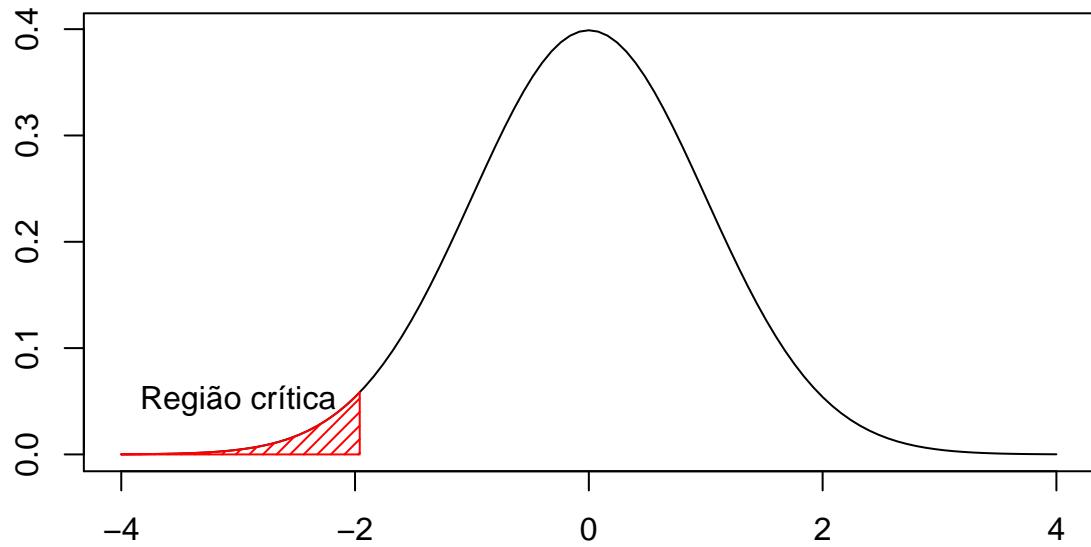


Figure 5.4: Distribuição de Dickey Fuller

Existem três versões principais do teste:

- 1) Teste para raiz unitária

$$\Delta Y_t = \delta Y_{t-1} + \varepsilon_t \rightarrow \tau;$$

- 2) Teste para raiz unitária com drift

$$\Delta Y_t = \mu + \delta Y_{t-1} + \varepsilon_t \rightarrow \tau_\mu;$$

- 3) Teste de raiz unitária com drift e tendência temporal determinística

$$\Delta Y_t = \mu + at + \delta Y_{t-1} + \varepsilon_t \rightarrow \tau_\tau$$

O teste de Dickey Fuller é um teste unilateral a esquerda (veja figura 5.4)

A estatística  $\hat{\tau}$  para cada um dos modelos pode ser obtida da seguinte forma:

$$\hat{\tau} = \frac{\hat{\delta}}{s(\hat{\delta})} \quad (5.4)$$

em que  $s(\hat{\delta})$  é o desvio padrão de

$$\hat{\delta} = \frac{\sum_{t=2}^n Y_{t-1}Y_t}{\sum_{t=2}^n Y_{t-1}^2} - 1,$$

que é a estimativa de mínimos quadráticos de  $\rho$  menos 1, para garantir que, sob  $H_0$ , tenhamos  $\delta = 0$ . O desvio padrão pode ser obtido a partir do cálculo da variância residual, que no caso mais simples se torna

$$S^2 = \frac{1}{n} \sum_{t=1}^n (\Delta Y_t - \hat{\delta} Y_{t-1})^2.$$

Cada versão do teste ( $\tau$ ,  $\tau_\mu$  e  $\tau_\tau$ ) tem sua própria estatística de teste e portanto tem seu próprio valor crítico o qual depende do tamanho amostral. Esses valores foram obtidos a partir de simulações de Monte Carlo.

Em cada caso, a hipótese nula de que *existe raíz unitária* é representada por  $\delta = 0$ . Para estes testes é conhecido que eles têm *baixo poder* no sentido de que frequentemente não conseguem distinguir entre processos com raíz unitária ( $\delta = 0$ ) de processos com raíz quase-unitária ( $\delta$  próximo de zero), ou até mesmo com tendências não lineares.

A tabela a seguir apresenta alguns valores críticos para o teste de Dickey Fuller

| Estatística | n    | 1%    | 2.5%  | 5%    | 10%   |
|-------------|------|-------|-------|-------|-------|
| $\tau$      | 25   | -2.66 | -2.26 | -1.95 | -1.60 |
|             | 50   | -2.62 | -2.25 | -1.95 | -1.61 |
|             | 100  | -2.60 | -2.24 | -1.95 | -1.61 |
|             | 250  | -2.58 | -2.23 | -1.95 | -1.61 |
|             | 500  | -2.58 | -2.23 | -1.95 | -1.61 |
|             | >500 | -2.58 | -2.23 | -1.95 | -1.61 |
| $\tau_\mu$  | 25   | -3.75 | -3.33 | -3.00 | -2.62 |
|             | 50   | -3.58 | -3.22 | -2.93 | -2.60 |
|             | 100  | -3.51 | -3.17 | -2.89 | -2.58 |
|             | 250  | -3.46 | -3.14 | -2.88 | -2.57 |
|             | 500  | -3.44 | -3.13 | -2.87 | -2.57 |
|             | >500 | -3.43 | -3.12 | -2.86 | -2.57 |
| $\tau_\tau$ | 25   | -4.38 | -3.95 | -3.60 | -3.24 |
|             | 50   | -4.15 | -3.80 | -3.50 | -3.18 |
|             | 100  | -4.04 | -3.73 | -3.45 | -3.15 |
|             | 250  | -3.99 | -3.69 | -3.43 | -3.13 |
|             | >500 | -3.98 | -3.68 | -3.42 | -3.13 |

### 5.2.3 Teste ADF

Existe uma extensão do teste de Dickey-Fuller (DF) chamado de Teste de Dickey-Fuller aumentado (ADF) o qual remove todos os efeitos estacionários (autocorrelações) da série temporal e então testa usando o mesmo procedimento.

Existem outros testes bem reconhecidos, que surgiram para resolver o problema de baixo poder do teste de Dickey Fuller. Estes testes devem ser também utilizados em caso de dúvida na hora da modelagem. São os testes de **Phillips-Perron**, **KPSS**, **ERS**, **NG** e **Perron** entre outros. Alguns estão disponíveis no Gretl, na opção *variável – testes de raíz unitária*.

## 5.3 Eliminando tendência estocástica

### 5.3.1 Diferenças sucessivas

O método de diferenciação sucessivas é utilizado para eliminar tendência estocástica. Para isso precisamos definir o operador diferença.

---

Operador Diferença

$$\Delta = 1 - L$$

em que  $L$  é o operador de defasagem.

---

Na figura a seguir temos uma aplicação do operador diferença.

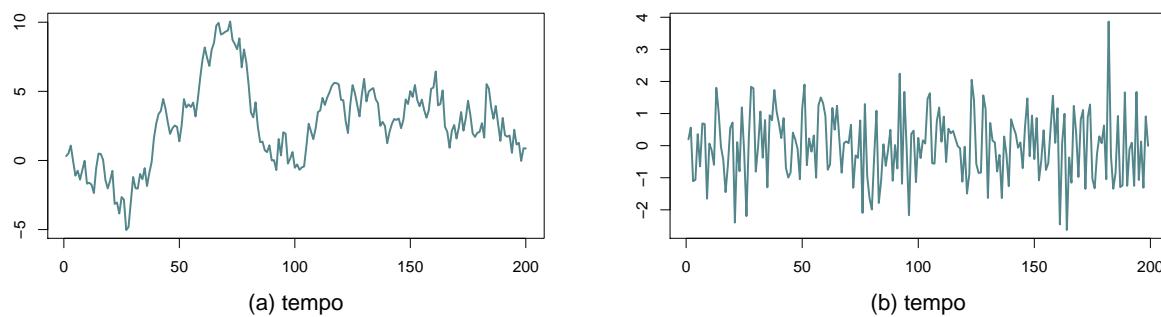


Figure 5.5: Passeio Aleatório (a) e sua diferença (b)

## 5.4 Modelagem ARIMA

Quando uma série temporal apresenta tendência estocástica (não estacionária) diz-se que está é integrada  $I(\cdot)$ . É necessário retirar a tendência para então analisar o ruído. Esse ruído não necessariamente é um ruído branco. Pode ser um modelo ARMA, por exemplo. Como visto anteriormente, a maneira de retirar a tendência estocástica de uma série temporal é diferenciando-a. Algumas vezes, é necessário diferenciar mais do que uma vez a série temporal até torná-la estacionária.

- Diz que uma série sem nenhuma raiz unitária é  $I(0)$ .
- A série é dita  $I(1)$  se for necessário diferenciá-la uma vez para torná-la estacionária.
- A série é dita  $I(d)$  se for necessário diferenciá-la  $d$  vezes para torná-la estacionária.

**Exemplo 5.3.** Na figura 5.6 são apresentados a série sobre dados de vendas *BJSales* de (Box and Jenkins, 1970).

```
layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE), width=1:1, height=1:1)
plot(BJSales-200, type="l", lwd=2, cex.lab=1.4, xlab=" ", ylab="Vendas", main=" ", cex.main=1.3, col="cadetblue")
plot(diff(BJSales), type="l", lwd=2, cex.lab=1.4, xlab=" ", ylab="diff(Vendas)", main=" ", cex.main=1.3, col="darkgreen")
plot(diff(diff(BJSales)), type="l", lwd=2, cex.lab=1.4, xlab=" ", ylab="diff(diff(Vendas))", main=" ", cex.main=1.3, col="darkgreen")
```

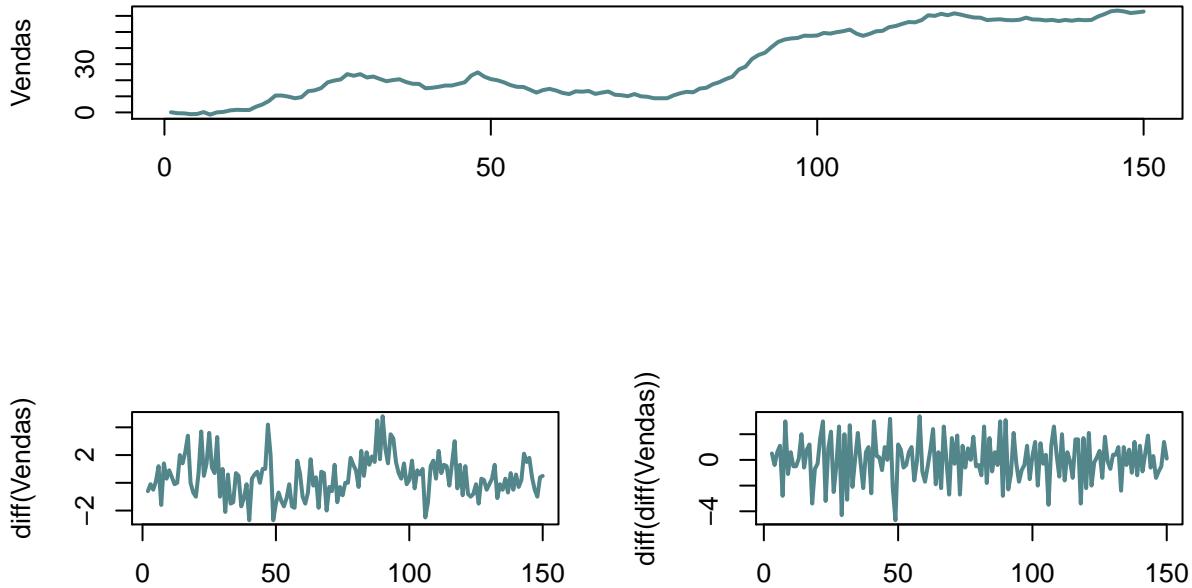


Figure 5.6: Série de vendas, primeira e segunda diferenças

**Exercício 5.1** (ANPEC 2012-07). Suponha que  $\Delta Y_t$  pode ser representado pelo seguinte processo:

$$\Delta Y_t = \varepsilon_t - 0,6\varepsilon_{t-1}, \text{ para } t = 1$$

$$\Delta Y_t = \Delta Y_{t-1} + \varepsilon_t - 0,6\varepsilon_{t-1}, \text{ para } t \geq 2$$

em que  $\varepsilon_t$ ,  $t = 1, 2, \dots$  é uma sequência de variáveis aleatórias independentes e identicamente distribuídas com média igual a 0. Se  $Y_t = 0$ , quando  $t = 0$ , calcule o valor da  $\mathbb{E}[Y_3]$ .

## 5.5 Previsão

Um dos objetivos finais na análise de séries temporais é a previsão. Assim, pode-se usar informações do passado para tomar decisões para o futuro. Existem outros métodos de previsão para séries temporais, como o de **Média Móveis Simples (MMS)**, **Suavizado Exponencial (SE)**, entre outros, mas estes métodos não dependem de um ajuste de um modelo e não são considerados agora. Para uma boa previsão é fundamental que o modelo esteja bem ajustado e por isso deixamos este tópico para o final. Como é feita a previsão na prática?

---

A ideia da previsão é utilizar o conhecimento/observações que se tem até o tempo  $t$ , (digamos que temos observações para uma certa variável durante os últimos 20 anos e, assim,  $t$  seria o último ano observado e  $\dots, Y_{t-2}, Y_{t-1}, Y_t$  as observações). É conveniente definir

$$\mathbb{E}_t(Y_s) = E(Y_s|Y_t, Y_{t-1}, \dots, Y_2, Y_1),$$

Assim,

$$\mathbb{E}_t(Y_s) = Y_s, \text{ se } s \leq t$$


---

**Exemplo 5.4** (Previsão para a série de dados \*AirPassengers\* ). Para essa série o ajuste foi feito para o log dos dados. Além da sazonalidade, incorporada pelo parâmetro *seasonal*, também foi utilizado o ARIMA(0,1,1) para a componente não sazonal do log da série.

```
pass <- ts(AirPassengers, frequency = 12, start=c(1949,1), end=c(1959,12))
m1 <- arima(log(pass), c(0,1,1), seasonal = list(order=c(0,1,1), period=12))
prev <- predict(m1, n.ahead=4*12) # previsão de 1960 até 1964
l <- prev$pred - 1.96 * prev$se
u <- prev$pred + 1.96 * prev$se
ts.plot(AirPassengers, exp(prev$pred), exp(l), exp(u), log = "y", lty = c(1,4, 2, 2), col=c("cadetblue4", 4, 2, 2))
```

Para um exemplo de previsão, consideremos o modelo AR(1):

**Exemplo 5.5** (Previsão para o modelo AR(1)). O modelo AR(1) é representado por:

$$Y_{t+1} = c + \phi Y_t + \varepsilon_t.$$

Assim,

$$\begin{aligned}\mathbb{E}_t(Y_{t+1}) &= c + \phi Y_t = Y_{t+1} - \varepsilon_{t+1} \\ \mathbb{E}_t(Y_{t+2}) &= c + \phi \mathbb{E}_t(Y_{t+1}) = c + \phi(c + \phi Y_t) \\ &\vdots \\ \mathbb{E}_t(Y_{t+h}) &= c \sum_{i=1}^{h-1} \phi^{i-1} + \phi^h Y_t.\end{aligned}$$

Logo

---

\textcolor{blue}{\{Previsão  $h$ -passos à frente para o modelo AR(1)\}}

$$\hat{y}_t(h) = \mathbb{E}_t(Y_{t+h})$$

representa previsão  $h$ -passos à frente, dado que observamos até o tempo  $t$ .

---

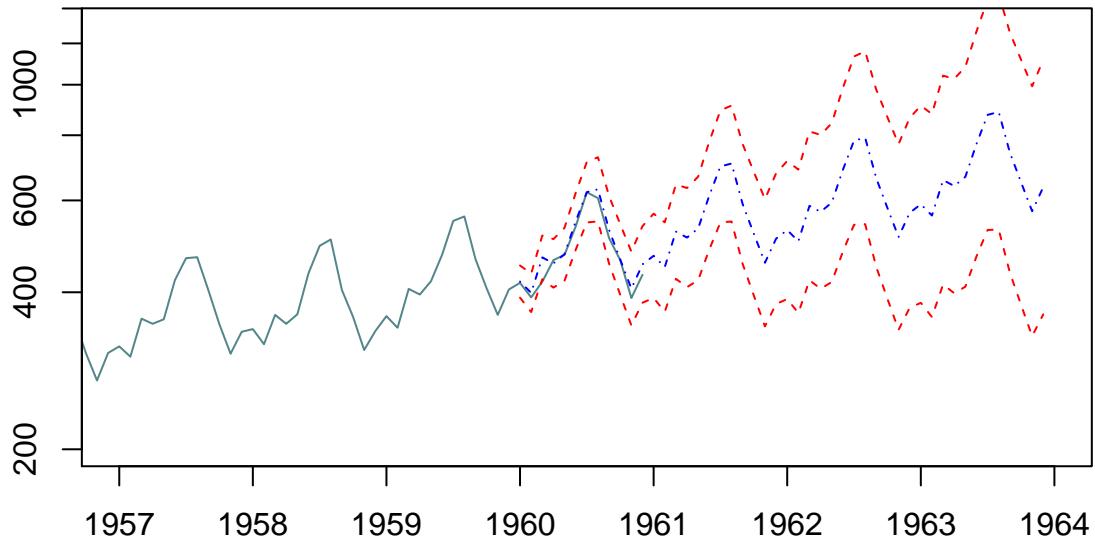


Figure 5.7: Previsão para o log da série de passageiros das companhias aéreas americanas

### 5.5.1 Erro de previsão

O erro de previsão é definido como sendo o valor observado menos o valor previsto. Para um período  $h$ ,  $\varepsilon_t(h)$  é dado por:

#### Erro de previsão

$$\varepsilon_t(h) = Y_{t+h} - \mathbb{E}_t(Y_{t+h})$$

os quais são não viesados, isto é,  $\mathbb{E}(\varepsilon_t(h)) = 0$ ;

$$\begin{aligned}
 \varepsilon_t(1) &= Y_{t+1} - \mathbb{E}_t(Y_{t+1}) = \varepsilon_{t+1} \\
 \varepsilon_t(2) &= Y_{t+2} - \mathbb{E}_t(Y_{t+2}) = c + \rho Y_{t+1} + \varepsilon_{t+2} - c - \rho \mathbb{E}_t(Y_{t+1}) \\
 &= \rho \varepsilon_{t+1} + \varepsilon_{t+2} \\
 \varepsilon_t(3) &= Y_{t+3} - \mathbb{E}_t(Y_{t+3}) = c + \rho Y_{t+2} + \varepsilon_{t+3} - c - \rho \mathbb{E}_t(Y_{t+2}) \\
 &= \rho^2 \varepsilon_{t+1} + \rho \varepsilon_{t+2} + \varepsilon_{t+3} \\
 &\vdots \\
 \varepsilon_t(h) &= Y_{t+h} - \mathbb{E}_t(Y_{t+h}) = \rho^{h-1} \varepsilon_{t+1} + \rho^{h-2} \varepsilon_{t+2} \dots + \rho \varepsilon_{t+h-1} + \varepsilon_{t+h}
 \end{aligned}$$

Tomando-se a [esperança do erro de previsão](#), podemos observar que estes são não viesados,  $\mathbb{E}(\varepsilon_t(h)) = 0$ ;

A variância do erro de previsão é dada por:

$$\begin{aligned}\text{Var}(\varepsilon_t(h)) &= \text{Var}(\rho^{h-1}\varepsilon_{t+1} + \rho^{h-2}\varepsilon_{t+2} \cdots + \rho\varepsilon_{t+h-1} + \varepsilon_{t+h}) \\ &= \sigma_\varepsilon^2 (\phi^{2(h-1)} + \phi^{2(h-2)} + \cdots + \phi^2 + 1)\end{aligned}$$

Note que a variância converge para uma constante, quando  $h \rightarrow \infty$ , que é  $\frac{\sigma_\varepsilon^2}{1-\rho^2}$  que é a variância não condicional da série  $Y_t$ .

Se a distribuição dos resíduos  $\varepsilon_t$  é a Normal, então o intervalo de confiança para os resíduos é dado portanto

$$c \sum_{i=1}^{h-1} \rho^{i-1} + \rho^h y \pm 2\sigma_\varepsilon (\phi^{2(h-1)} + \phi^{2(h-2)} + \cdots + \phi^2 + 1)^{\frac{1}{2}}$$

### 5.5.2 Medidas de desempenho

Diferentes modelos produzem previsões distintas, o que torna necessários avaliar essas previsões. Para isso são utilizadas algumas medidas de desempenho. As estatísticas mais conhecidas são:

- **MSE** - Mean Square Error (erro quadrático médio)

$$MSE_{t,H} = \sqrt{\frac{\sum_{h=1}^H \varepsilon_t^2(h)}{H}}$$

Para calculá-los, deve-se deixar algumas observações fora da amostra. Por exemplo, em uma série com  $n$  observações, deixa-se as  $H$  últimas observações fora da amostra e estima-se o modelo agora com  $n - H$  observações restantes.

- **MAE** - Mean Absolute Error (erro absoluto médio)

$$MAE_{t,H} = \frac{\sum_{h=1}^H |\varepsilon_t(h)|}{H}$$

- **MAPE** - Mean Absolute Percentual Error (erro absoluto percentual médio)

$$MAPE_{t,H} = \sum_{h=1}^H \left| \frac{\varepsilon_t(h)}{H y_{t+h}} \right|$$

### 5.5.3 Previsão dinâmica e estática

Quando faz-se previsões  $h$  passos a frente,  $\hat{y}_t(h)$ , usando somente a informação até o tempo  $t$ , tem-se a previsão dinâmica cuja variância acaba sendo maior. Quando, para prever algum passo a frente usa-se as observações até o tempo imediatamente anterior, tem-se a previsão estática. A previsão estática só é útil para efeito de comparação de modelos. Na prática, a previsão dinâmica é a única que interessa de fato.

## 5.6 Regressão espúria e cointegração

A utilização dos modelos de regressão envolvendo séries temporais não estacionárias pode conduzir ao problema que se convencionou chamar de regressão espúria, isto é quando temos um alto  $R^2$  sem uma relação significativa entre as variáveis (Harris et al., 1995). Assim, na presença de raiz unitária podem-se encontrar relações econométricas entre duas variáveis econômicas sem qualquer relação de causalidade entre uma e outra por puro acaso. Por exemplo, a regressão de uma variável I(1) com outra I(1) obtida independentemente gera alto  $R^2$  e estatística  $t$  significante. No entanto o resultado não tem significado econômico.

**Exemplo 5.6** (Correlação entre duas séries independentes). Fizemos a seguinte experiência. Geramos duas séries I(1) independentes entre si e regredimos um contra a outra. O resultado segue.

```
set.seed(76543210)
s1=cumsum(rnorm(100)) # Série I(1), s1
set.seed(01234567)
s2=cumsum(rnorm(100)) # Série I(1), s2 independente de s1
m1=lm(s1~s2)
summary(m1)

##
Call:
lm(formula = s1 ~ s2)
##
Residuals:
Min 1Q Median 3Q Max
-6.5336 -1.6894 0.1274 1.7391 6.6388

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.36738 0.47727 7.056 2.47e-10 ***
s2 0.76312 0.08868 8.606 1.27e-13 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.955 on 98 degrees of freedom
Multiple R-squared: 0.4304, Adjusted R-squared: 0.4246
F-statistic: 74.06 on 1 and 98 DF, p-value: 1.273e-13
```

Como podemos observar, encontramos um  $R^2 = 0.43$  alto e estatísticas significativas. No entanto, as séries são independentes. O resultado disso, é que quando colocamos no mesmo gráfico, a série  $Y$  e o predito, podemos observar que o predito não é nem de perto razoável.

Isto ocorre devido ao fato de que a presença de uma tendência, decrescente ou crescente, em ambas as séries leva a um alto valor do  $R^2$  mas não necessariamente, a presença de uma relação verdadeira entre séries (Gujarati, 2003)

Detectada a presença de raiz unitária, então deve-se trabalhar com as séries temporais diferenciadas e não em nível, ou seja, a tendência precisa ser removida. Assim, quando uma série econômica apresentar uma tendência estocástica tornar-se-á estacionária após a aplicação de uma ou mais diferenças, pois terá pelo menos uma raiz unitária. No entanto, ao se remover a tendência, elementos de longo prazo entre as variáveis são eliminados.

A interpretação econômica da cointegração é que se duas (ou mais) variáveis possuem uma relação de equilíbrio de longo prazo, então mesmo que as séries possam conter tendências estocásticas (isto é, serem não estacionárias), elas irão mover-se juntas no tempo e a diferença entre elas será estável (isto é, estacionária). Em suma, o conceito de cointegração indica a existência de um equilíbrio de longo prazo, para o qual o sistema econômico converge no tempo (Harris et al., 1995).

### 5.6.1 Quando é possível regredir duas séries I(d)

Para que a regressão entre duas séries temporais não seja espúria, elas devem satisfazer uma das seguintes situações:

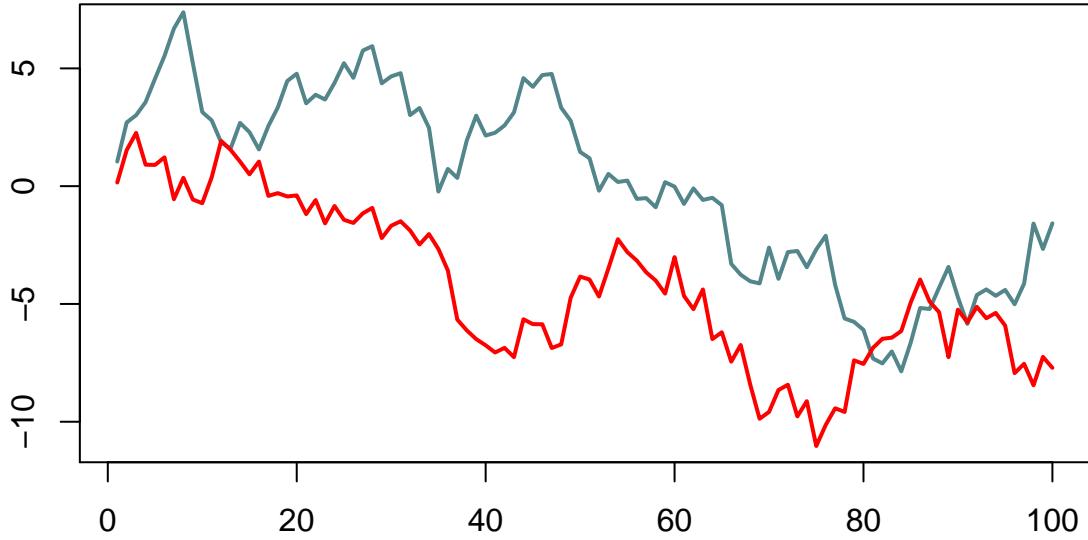


Figure 5.8: Correlação expúria. Dois passeios aleatórios independentes

Séries que cointegram, 1)  $\{Y_t\}$  e  $\{X_t\}$  devem ser estacionárias.

2)  $\{Y_t\}$  e  $\{X_t\}$  devem ser integradas de mesma ordem e o resíduo deve ser estacionário.

Se  $\{Y_t\}$  e  $\{X_t\}$  são integrados de ordens diferentes ou se  $\{Y_t\}$  e  $\{X_t\}$  são integrados de mesma ordem e o resíduo não é estacionário, então a regressão é espúria.

Um teste utilizado para detectar cointegração é o teste de Durbin-Watson.

## 5.7 Exercícios

**Exercício 5.2** (ANPEC 2013-05). Um pesquisador corretamente postula o seguinte modelo de regressão:

$$y_t = \beta_1 + \beta_2 t + u_t, \quad t = 1, \dots, T;$$

em que  $u_t$  é uma variável aleatória independente e identicamente distribuída ao longo do tempo, com média zero e variância finita.

Julgue as afirmativas:

- 0)  $y_t$  é um processo estacionário.
- 1)  $\Delta y_t = y_t - y_{t-1}$  é um processo estacionário de segunda ordem.
- 2) Mínimos quadrados ordinários aplicado à equação (??) produz uma estimativa não viésada de  $\beta_2$ .

- 3) Seja  $\hat{\beta}_2 = \sum_{t=2}^T (y_t - y_{t-1})/(T-1)$ .  $\hat{\beta}_2$  é um estimador consistente de  $\beta_2$ .  
 4) Suponha que  $u_t = \rho u_{t-1} + \varepsilon_t$ ,  $\rho < 1$  e que  $\varepsilon_t$  seja uma variável aleatória

independente e identicamente distribuída ao longo do tempo, com média zero e variância finita. O estimador de mínimos quadrados ordinários de  $\beta_2$  no modelo é não viesado.

**Exercício 5.3** (ANPEC 2007-07). Sejam  $Y_t$  e  $X_t$  duas séries temporais. Considere os resultados dos seguintes modelos de regressão estimados por mínimos quadrados ordinários (MQO):

$$\Delta Y_t = 4,8788 - 0,1512Y_{t-1} \text{ e } \Delta X_t = 0,1094 - 0,1807X_{t-1}$$

(1,70)                  (-1,97)                  (1,26)                  (-2,21)

Considere também os resultados da regressão de  $Y_t$  em  $X_t$

$$Y_t = 23,3924 + 14,4006X_t + \hat{e}_t,$$

(1,70)                  (-1,97)

em que  $\hat{e}_t$  é o resíduo. Finalmente, considere a seguinte regressão:

$$\Delta \hat{e}_t = 0,0730 - 0,4157\hat{e}_{t-1}.$$

(0,06)                  (-3,43)

Os números entre parênteses são os valores do teste t de significância individual dos parâmetros. Dado que o valor crítico a 5% da estatística de Dickey-Fuller é -2,938, é correto afirmar que:

- 0)  $Y_t$  e  $X_t$  são séries temporais integradas de ordem 1.
- 1) A regressão de  $Y_t$  em  $X_t$  é espúria.
- 2) A hipótese de cointegração entre  $Y_t$  e  $X_t$  é rejeitada pois os resíduos da regressão de  $Y_t$  em  $X_t$  são não-estacionários.
- 3) Para que duas variáveis sejam cointegradas é necessário que ambas tenham a mesma ordem de integração.
- 4) A rejeição da hipótese nula do teste Dickey-Fuller implica que a variável em questão é não-estacionária.

**Exercício 5.4** (ANPEC 2007-09). Julgue as proposições:

- 0) A soma de dois processos estocásticos independentes e estacionários de segunda ordem será estacionária de segunda ordem.
- 1) A soma de dois processos estocásticos não-estacionários será não-estacionária.
- 2) Seja  $L$  o operador defasagem tal que  $LY_t = Y_{t-1}$ . Se  $Y_t$  segue um processo AR(1) estacionário de segunda ordem, então  $(1-L)^2Y_t$  é um processo ARMA(2,2).
- 3) O processo ARMA(2,2) definido na forma  $(1-L-0,25L^2)Y_t = (1-0,5L-0,06L^2)u_t$  é não estacionário, em que  $u_t$  é o erro aleatório com média nula e variância constante.
- 4) Todo processo MA é estacionário de segunda ordem.

**Exercício 5.5.** Para este exercício consideremos uma série temporal de vendas  $BJsales$ .

```
Teste de Dickey - Fuller para BJsales
adf1=adf.test(BJsales)
```

```
Augmented Dickey-Fuller Test
alternative: stationary
##
Type 1: no drift no trend
lag ADF p.value
[1,] 0 3.52 0.990
[2,] 1 2.45 0.990
[3,] 2 1.94 0.986
[4,] 3 1.71 0.978
```

```

[5,] 4 1.42 0.960
Type 2: with drift no trend
lag ADF p.value
[1,] 0 -0.172 0.935
[2,] 1 -0.478 0.880
[3,] 2 -0.664 0.814
[4,] 3 -0.837 0.754
[5,] 4 -1.010 0.693
Type 3: with drift and trend
lag ADF p.value
[1,] 0 -0.986 0.937
[2,] 1 -1.316 0.861
[3,] 2 -1.606 0.739
[4,] 3 -1.789 0.662
[5,] 4 -2.077 0.541

Note: in fact, p.value = 0.01 means p.value <= 0.01

Teste de Dickey - Fuller para diff(BJsales)
adf2=adf.test(diff(BJsales))

Augmented Dickey-Fuller Test
alternative: stationary
##
Type 1: no drift no trend
lag ADF p.value
[1,] 0 -8.25 0.01
[2,] 1 -5.34 0.01
[3,] 2 -4.24 0.01
[4,] 3 -3.38 0.01
[5,] 4 -3.18 0.01
Type 2: with drift no trend
lag ADF p.value
[1,] 0 -8.77 0.01
[2,] 1 -5.76 0.01
[3,] 2 -4.64 0.01
[4,] 3 -3.72 0.01
[5,] 4 -3.50 0.01
Type 3: with drift and trend
lag ADF p.value
[1,] 0 -8.74 0.0100
[2,] 1 -5.74 0.0100
[3,] 2 -4.62 0.0100
[4,] 3 -3.70 0.0262
[5,] 4 -3.48 0.0461

Note: in fact, p.value = 0.01 means p.value <= 0.01

```

- a) O que podemos afirmar a respeito da tendência da série *BJsales*? Use os resultados dos testes de hipóteses para justificar a sua resposta.
- b) O que podemos afirmar a respeito da tendência da primeira diferença da série *BJsales*? Use os resultados dos testes de hipóteses para justificar a sua resposta.
- c) Dos gráficos apresentados na figura 5.9, qual(is) pode(m) representar a série *BJsales*? E qual(is) pode(m) representar a primeira diferença da série *BJsales*? Explique.
- d) Na figura 5.10 qual(is) dos gráficos de FAC e FACP pode(m) corresponder à FAC e FACP de um ruído branco? Justifique.

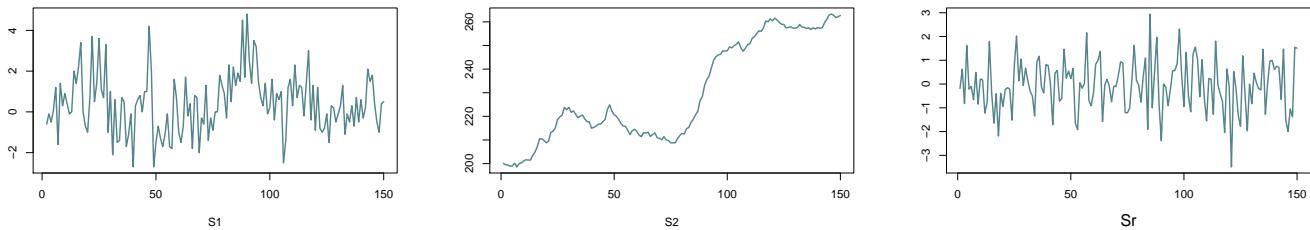
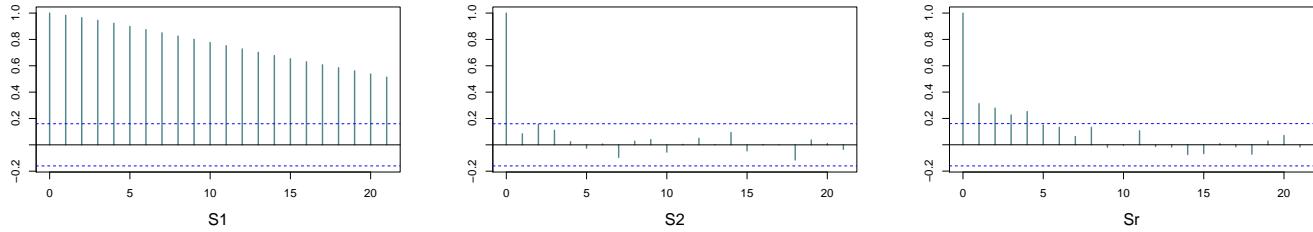


Figure 5.9: Séries Temporais S1, S2 e S3

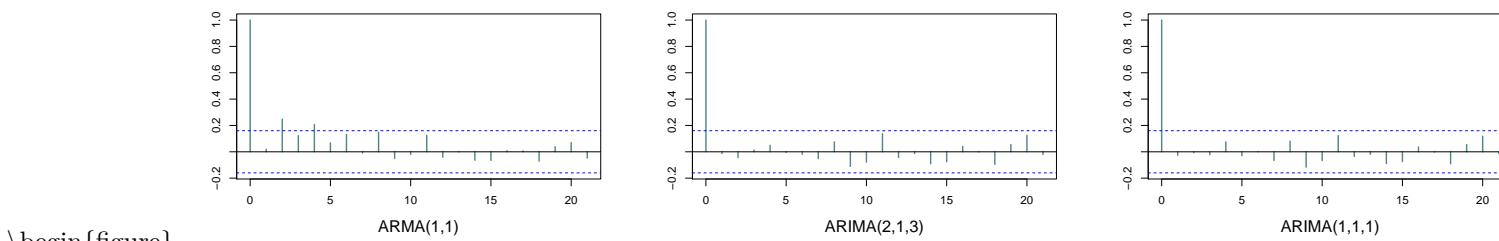
Figure 5.10: FAC e FACP para três séries temporais distintas  $X_1$ ,  $X_2$  e  $X_3$ .

e) Na figura 5.10 qual(is) dos gráficos de FAC e FACP pode(m) corresponder à FAC e FACP de um ruído branco? Justifique.

f) Na figura 5.10 qual(is) dos gráficos de FAC e FACP pode(m) corresponder à FAC e FACP da Série *BJsales*? Justifique.

g) Dos seguintes modelos: AR(1), MA(1), ARMA(1,1), ARIMA(1,1,1), ARIMA(3,1,2) e ARIMA(1,2,1), qual(is) poderiam ajustar corretamente a série temporal *BJsales*? Justifique.

h) Foram ajustados 3 modelos para a série *BJsales*: ARMA(1,1) (AIC = 546.7917), ARIMA(2,1,3)(AIC = 520.1139) e ARIMA(1,1,1) (AIC = 514.736). A FAC e FACP dos resíduos dos ajustes são apresentados na figura ???. Qual é o melhor modelo? Justifique.



\begin{figure}  
 \caption{FAC e FACP dos resíduos do ajuste de três modelos a série *BJsales*,} \end{figure}

i) Faça a correspondência dos teste de Teste *LJUNG-BOX* na Tabela a seguir com a figura 5.7 explicando o seu raciocínio.

|   | Teste<br>1 |       |      |            |               | Teste<br>2 |       |            |               |              | Teste<br>3 |            |               |             |  |
|---|------------|-------|------|------------|---------------|------------|-------|------------|---------------|--------------|------------|------------|---------------|-------------|--|
|   | Def        | ACF   | PACF | Q-<br>stat | [p-<br>valor] | ACF        | PACF  | Q-<br>stat | [p-<br>valor] | ACF          | PACF       | Q-<br>stat | [p-<br>valor] |             |  |
| 1 | -          | ***   | -    | ***        | 47.49 [0.00]  | -          | ***   | -          | ***           | 31.42 [0.00] | -          | -          | 0.19 [0.65]   |             |  |
|   | 0.483      | 0.483 |      |            |               | 0.406      | 0.406 |            |               | 0.031        | 0.031      |            |               |             |  |
| 2 | -          | -     | ***  | 48.77      | [0.00]        | 0.044      | -     | **         | 31.79 [0.00]  | -            | *          | -          | *             | 3.22 [0.19] |  |
|   | 0.079      | 0.408 |      |            |               |            | 0.145 |            |               | 0.121        | 0.122      |            |               |             |  |

|    |       | Teste 1 |     |       |        | Teste 2 |       |       |        | Teste 3 |       |             |
|----|-------|---------|-----|-------|--------|---------|-------|-------|--------|---------|-------|-------------|
| 3  | 0.089 | -       | *** | 50.40 | [0.00] | 0.016   | -     | 31.85 | [0.00] | 0.089   | 0.082 | 4.88 [0.18] |
|    |       | 0.254   |     |       |        |         | 0.026 |       |        |         |       |             |
| 4  | -     | -       | *** | 50.58 | [0.00] | 0.030   | 0.042 | 32.03 | [0.00] | 0.038   | 0.029 | 5.19 [0.26] |
|    | 0.029 | 0.216   |     |       |        |         |       |       |        |         |       |             |
| 5  | 0.044 | -       |     | 50.98 | [0.00] | 0.008   | 0.052 | 32.04 | [0.00] | 0.066   | 0.091 | 6.12 [0.29] |
|    |       | 0.098   |     |       |        |         |       |       |        |         |       |             |
| 6  | -     | -       | *** | 52.87 | [0.00] | -       | 0.008 | 32.12 | [0.00] | 0.034   | 0.040 | 6.37 [0.38] |
|    | 0.095 | 0.189   |     |       |        | 0.020   |       |       |        |         |       |             |
| 7  | 0.072 | -       | *   | 53.99 | [0.00] | 0.027   | 0.023 | 32.27 | [0.00] | 0.053   | 0.070 | 6.96 [0.43] |
|    |       | 0.121   |     |       |        |         |       |       |        |         |       |             |
| 8  | -     | -       |     | 53.99 | [0.00] | 0.045   | 0.075 | 32.68 | [0.00] | -       | -     | 7.18 [0.51] |
|    | 0.002 | 0.100   |     |       |        |         |       |       |        | 0.031   | 0.035 |             |
| 9  | -     | -       | *** | 56.49 | [0.00] | -       | -     | 34.52 | [0.00] | 0.043   | 0.045 | 7.58 [0.57] |
|    | 0.108 | 0.249   |     |       |        | 0.096   | 0.056 |       |        |         |       |             |
| 10 | 0.167 | **      | -   | 62.44 | [0.00] | 0.122 * | 0.073 | 37.54 | [0.00] | 0.048   | 0.023 | 8.09 [0.61] |
|    |       | 0.090   |     |       |        |         |       |       |        |         |       |             |

**Exercício 5.6.** Escreva a equação do modelo para a seguinte saída do R

```
##
Call:
arima(x = X, order = c(2, 1, 1))
##
Coefficients:
ar1 ar2 ma1
0.8305 0.0360 -0.607
s.e. 0.1774 0.1178 0.160
##
sigma^2 estimated as 1.774: log likelihood = -254.32, aic = 516.64
```

**Exercício 5.7.** Seja  $\{y_t\}_{t=1}^{440}$  uma série temporal. Essa série foi ajustada de acordo com um modelo AR(2). A equação estimada foi:  $y_t = 14.62 - 0.61y_{t-1} + 0.15y_{t-2}$ . Os seguintes dados estão disponíveis:

| $t$         | 436   | 437   | 438   | 439   | 440   |
|-------------|-------|-------|-------|-------|-------|
| $y_t$       | 9.88  | 10.42 | 11.08 | 8.12  | 11.71 |
| $\hat{e}_t$ | -0.21 | 0.40  | 1.33  | -1.30 | 0.38  |

a) Calcule a previsão um passo a frente e dois passos a frente para a série  $y_t$ , ou seja,  $\hat{y}_{440}(1)$  e  $\hat{y}_{440}(2)$ . R:  $\hat{y}_{440}(1) = 8.6949$  e  $\hat{y}_{440}(2) = 11.07261$ .

b) Calcule o erro de previsão um e dois passos a frente,  $e_{440}(1)$  e  $e_{440}(2)$ , sabendo-se que  $y_{441} = 8.83$  e  $y_{442} = 12.24$ . R:  $e_{440}(1) = 0.1351$  e  $e_{440}(2) = 1.167389$ .

**Exercício 5.8.** Seja  $\{y_t\}_{t=1}^{450}$  uma série temporal. Essa série foi ajustada de acordo com um modelo MA(2). A equação estimada foi:  $y_t = 10.01 + e_t - 0.64e_{t-1} + 0.22e_{t-2}$ . Os seguintes dados estão disponíveis:

| $t$         | 446   | 447   | 448   | 449   | 450   |
|-------------|-------|-------|-------|-------|-------|
| $y_t$       | 9.79  | 10.22 | 7.43  | 12.41 | 8.35  |
| $\hat{e}_t$ | -0.52 | 0.21  | -2.34 | 0.87  | -0.60 |

a) Calcule a previsão um, dois e três passos a frente para a série  $y_t$ , ou seja,  $\hat{y}_{450}(1)$ ,  $\hat{y}_{450}(2)$  e  $\hat{y}_{450}(3)$ . R:  $\hat{y}_{450}(1) = 10.5854$ ,  $\hat{y}_{450}(2) = 9.878$  e  $\hat{y}_{450}(3) = 10.01$ .

b) Calcule o erro de previsão um, dois e três passos a frente,  $e_{450}(1)$ ,  $e_{450}(2)$  e  $e_{450}(3)$ , sabendo-se que  $y_{451} = 9.80$ ,  $y_{452} = 8.78$  e  $y_{453} = 9.33$ . **R:**  $e_{450}(1) = -0.7767$ ,  $e_{450}(2) = -1.098$  e  $e_{450}(3) = -0.68$ .

**Exercício 5.9.** Escreva cada um dos seguintes processos usando o operador de defasagem  $\mathbf{L}$ .

- a)  $X_t = 0.3X_{t-1} + a_t$ ;
- b)  $X_t = \sum_{j=1}^t a_t$ ,  $t \geq 1$ ;
- c)  $X_t = a_t + 0.4a_{t-1} - 0.2a_{t-2} + 0.17a_{t-3}$ ;
- d)  $X_t = 1.5X_{t-1} - 0.75X_{t-2} + a_t + 4.0$ ;
- e)  $X_t = 0.5X_{t-1} + a_t + 0.4a_{t-1} - 0.2a_{t-2}$ ;
- f)  $X_t - X_{t-1} = -0.3X_{t-1} + a_t + 0.4a_{t-1}$ ;

**Exercício 5.10.** Seja  $\{y_t\}_{t=1}^{450}$  uma série temporal. Essa série foi ajustada de acordo com um modelo ARMA(2,2). A equação estimada foi:  $y_t = 1.61 + 1.39y_{t-1} - 0.55y_{t-2} + e_t - 0.81e_{t-1} + 0.25e_{t-2}$ . Os seguintes dados estão disponíveis:

| $t$         | 446   | 447   | 448   | 449   | 450   |
|-------------|-------|-------|-------|-------|-------|
| $y_t$       | 12.16 | 11.69 | 11.56 | 10.32 | 10.87 |
| $\hat{e}_t$ | 0.56  | -0.07 | 0.19  | -0.75 | 0.62  |

- a) Calcule a previsão um, dois e três passos a frente para a série  $y_t$ , ou seja,  $\hat{y}_{450}(1)$ ,  $\hat{y}_{450}(2)$  e  $\hat{y}_{450}(3)$ . **R:**  $\hat{y}_{450}(1) = 10.3536$ ,  $\hat{y}_{450}(2) = 10.178$  e  $\hat{y}_{450}(3) = 10.06295$ .
- b) Calcule o erro de previsão um, dois e três passos a frente,  $e_{450}(1)$ ,  $e_{450}(2)$  e  $e_{450}(3)$ , sabendo-se que  $y_{451} = 9.80$ ,  $y_{452} = 8.78$  e  $y_{453} = 9.33$ . **R:**  $e_{450}(1) = 1.53$ ,  $e_{450}(2) = 2.052$  e  $e_{450}(3) = 0.69$ .

**Exercício 5.11.** Considere o modelo autorregressivo de primeira ordem, AR(1), definido por

$$Y_t = a + bY_{t-1} + u_t,$$

em que  $a$  e  $b$  são parâmetros e  $u_t$  é uma sequência de variáveis aleatórias independentes e igualmente distribuídas, com média nula e variância  $\sigma^2$ . Suponha que  $|b| < 1$ . A previsão  $n$  passos-à-frente para a variável  $Y$  convergirá para

- a).
- b) a média de  $u_t$ .
- c)  $\frac{a}{1-b}$ .
- d)  $E(Y_t)$ .
- e)  $\infty$ .

**Exercício 5.12.** As vendas mensais de um certo produto são representadas pelo modelo

$$Z_t = 3 + a_t + 0.5a_{t-1} - 0.25a_{t-2}, \quad \sigma_a^2 = 4.$$

- a) Obtenha  $\hat{Z}(\ell)$ ,  $\ell = 1, 2, 3, 100$ ;
- b) Calcule  $\text{Var}[e_t(\ell)]$ ,  $\ell = 1, 2, 3, 100$ ;
- c) Dados  $Z_1 = 3.25$ ,  $Z_2 = 4.75$ ,  $Z_3 = 2.25$  e  $Z_4 = 1.75$ , calcule  $\hat{Z}_4(\ell)$  para  $\ell = 1, 2, 3, 100$ ;

**Exercício 5.13.** Explique os passos que devem ser seguidos para a modelagem de uma série temporal na metodologia ARIMA. Considere a possibilidade de não-estacionariedade da série.

**Exercício 5.14.** Usando a esperança condicional, calcule as previsões 1, 2 e 3 passos a frente ( $\hat{y}_T(1)$ ,  $\hat{y}_T(2)$ ,  $\hat{y}_T(3)$ ) para os seguintes processos:

- a) ARIMA(1,1,0);

- b) ARIMA(1,1,1);
- c) ARIMA(1,2,1);
- d) ARIMA(2,1,2);

**Exercício 5.15.** Seja  $\{y_t\}_{t=1}^{440}$  uma série temporal. Essa série foi ajustada de acordo com um modelo ARIMA(1,1,1). O coeficiente estimado para o componente auto-regressivo foi 0,6347 e o coeficiente estimado referente à parte MA foi 0,3711. As seguintes informações estão disponíveis:

| $t$         | 436    | 437   | 438   | 439   | 440   |
|-------------|--------|-------|-------|-------|-------|
| $y_t$       | 20.52  | 20.04 | 20.52 | 19.64 | 16.13 |
| $\hat{e}_t$ | -0.092 | -1.29 | 1.27  | -1.66 | -2.33 |

- a) Escreva o modelo usando a notação do operador  $lag$ .
- b) Calcule a previsão um passo a frente e dois passos a frente para a série  $y_t$ , ou seja,  $\hat{y}_{440}(1)$  e  $\hat{y}_{440}(2)$ . R:  $\hat{y}_{440}(1) = 13.05$  e  $\hat{y}_{440}(2) = 11.09$ .
- c) Calcule o erro de previsão um e dois passos a frente,  $e_{440}(1)$  e  $e_{440}(2)$ , sabendo-se que  $y_{441} = 12.57$  e  $y_{442} = 9.93$ . R:  $e_{440}(1) = 0.478$  e  $e_{440}(2) = -1.157$ .

**Exercício 5.16.** Seja  $\{y_t\}_{t=1}^{440}$  uma série temporal. Essa série foi ajustada de acordo com um modelo ARIMA(1,2,1). O coeficiente estimado para o componente auto-regressivo foi 0,6364 e o coeficiente estimado referente a parte MA foi 0,3599. As seguintes informações estão disponíveis:

| $t$         | 436    | 437    | 438    | 439    | 440    |
|-------------|--------|--------|--------|--------|--------|
| $y_t$       | 782.78 | 803.30 | 823.34 | 843.86 | 863.50 |
| $\hat{e}_t$ | 1.34   | -0.08  | -1.30  | 1.26   | -1.65  |

- a) Escreva o modelo usando a notação do operador  $lag$ .
- b) Calcule a previsão um passo a frente e dois passos a frente para a série  $y_t$ , ou seja,  $\hat{y}_{440}(1)$  e  $\hat{y}_{440}(2)$ . R:  $\hat{y}_{440}(1) = 881.99$  e  $\hat{y}_{440}(2) = 899.74$ .
- c) Calcule o erro de previsão um e dois passos a frente,  $e_{440}(1)$  e  $e_{440}(2)$ , sabendo-se que  $y_{441} = 879.64$  e  $y_{442} = 892.21$ . R:  $e_{440}(1) = -2.35$  e  $e_{440}(2) = -7.53$ .

**Exercício 5.17.** Seja  $y_t$  o logaritmo de taxa de câmbio iene/US \$. A seguinte regressão foi proposta:  $\Delta y_t = \beta_0 + \beta_1 y_{t-1} + u_t$ . As estimativas seguem abaixo:

|                 | Estimativa | dp( $\cdot$ ) |
|-----------------|------------|---------------|
| $\hat{\beta}_0$ | 0.162      | 0.435         |
| $\hat{\beta}_1$ | 0.099      | 0.025         |

Sabendo-se que  $n = 777$ , faça o teste DF e responda se a série  $inf$  apresenta raiz unitária.

**Nota:** A tabela com os valores críticos para o teste de DF se encontra no final da lista. Note que  $\tau$  se refere ao modelo sem constante,  $\tau_\mu$  ao modelo com constante e  $\tau_\tau$  ao modelo com tendência.

**Exercício 5.18.** Utilizando os dados anuais (1959-1995) de log(PIB) norte americano, a seguinte regressão foi proposta:  $|\log(\text{PIB})|_{t=0+1} t + 2 \log(\text{PIB})_{\{t-1\}} + 3 \log(\text{PIB})_{\{t-1\}} + u_{-t}|$ . As estimativas seguem abaixo:

|                 | Estimativa | dp( $\cdot$ ) |
|-----------------|------------|---------------|
| $\hat{\beta}_0$ | 1.650      | 0.670         |
| $\hat{\beta}_1$ | 0.0059     | 0.003         |

|                 | Estimativa | dp(·) |
|-----------------|------------|-------|
| $\hat{\beta}_2$ | -0.320     | 0.087 |
| $\hat{\beta}_3$ | 0.264      | 0.126 |

$n = 35$

- a) Faça o teste ADF e responda se a série  $\inf$  apresenta raiz unitária.
- b) A inclusão da variável  $\Delta \log(PIB)_{t-1}$  no modelo acima parece ser necessária? Justifique.

**Exercício 5.19.** Utilizando os dados anuais (1948-1996) de inflação norte americana, a seguinte regressão foi proposta:

$$\Delta \inf_t = \beta_0 + \beta_1 \inf_{t-1} + \beta_2 \Delta \inf_{t-1} + u_t.$$

As estimativas seguem abaixo:

|                 | Estimativa | dp(·) |
|-----------------|------------|-------|
| $\hat{\beta}_0$ | 1.360      | 0.517 |
| $\hat{\beta}_1$ | -0.310     | 0.103 |
| $\hat{\beta}_2$ | 0.138      | 0.126 |

$n = 47$

- a) Faça o teste ADF e responda se a série  $\inf$  apresenta raiz unitária.
- b) A inclusão da variável  $\Delta \inf_{t-1}$  no modelo acima parece ser necessária? Justifique.

**Exercício 5.20.** Responda V ou F, justificando sua resposta: Seja o processo auto-regressivo:  $y_t = \phi_1 y_{t-1} + \varepsilon_t$ . Pode-se afirmar que:

- a) O processo é estacionário para  $\phi_1 < 1$ . F
- b) Se  $\phi_1 = 1$ , o processo é dito um passeio aleatório. V
- c) O estimador de MQO do parâmetro  $\phi_1$  é não-viciado. F
- d) A estatística t-Student pode ser usada para testar a presença de raiz unitária. F
- e) O processo pode ser escrito em uma forma alternativa como  $\Delta y_t = \delta y_{t-1} + \varepsilon_t$  em que  $\delta = -1 - \phi_1$  e  $\Delta y_t = y_t - y_{t-1}$ . V

**Exercício 5.21.** Responda V ou F, justificando sua resposta: Um economista estimou uma função consumo usando 25 observações anuais da renda pessoal disponível e consumo, a partir do modelo:  $C_t = \beta_0 + \beta_1 Y_t + u_t$  em que  $C_t$  representa consumo em  $t$ ;  $Y_t$  representa renda pessoal disponível em  $t$  e  $u_t$  é um erro aleatório. O economista fez o teste de Dickey-Fuller aumentado (ADF) para as séries de renda e de consumo, obtendo estimativas para a estatística do teste menores que os valores críticos tabelados, a 1%, 5% e 10%. Consequentemente, o economista:

- a) Aceitou a hipótese nula do teste ADF, concluindo que as séries de renda e consumo são não-estacionárias. V
- b) Concluiu que o teste t não é válido. V
- c) Concluiu que a regressão estimada é espúria. F
- d) Necessita fazer mais outros testes para verificar se a regressão estimada é espúria. V

**Exercício 5.22.** Responda V ou F, justificando sua resposta. Considere o modelo de regressão linear  $C_t = \beta_0 + \beta_1 Y_t + u_t$ . As variáveis são definidas como na questão anterior.

- a) se  $C_t$  e  $Y_t$  são I(1), então  $u_t$  será obrigatoriamente estacionário. F
- b) se  $C_t$  e  $Y_t$  são integradas, mas com ordens de integração diferentes, então a regressão será inválida. V
- c) se  $C_t$  e  $Y_t$  são I(1), então o teste ADF aplicado aos resíduos da regressão poderá identificar a presença de co-integração entre as variáveis. V

- d) se  $C_t$  e  $Y_t$  são I(1), mas os resíduos são I(0), então há co-integração entre as variáveis. V  
e) se  $C_t$  e  $Y_t$  são I(1) e os resíduos também são I(1), então a regressão de  $\Delta C_t$  em  $\Delta Y_t$  é inválida. F

**Exercício 5.23.** Responda V ou F, justificando sua resposta. Considere a seguinte regressão entre  $y_t$  e  $z_t$ :  $y_t = \alpha z_t + u_t$ , em que  $u_t$  é o erro. São corretas as afirmativas:

- a) se  $y_t$  for I(1) e  $z_t$  for I(0), então  $y_t$  e  $z_t$  são co-integradas. F  
b) se  $y_t$  for I(0) e  $z_t$  for I(1), então  $y_t$  e  $z_t$  são co-integradas. F  
c) se  $y_t$  for I(1) e  $z_t$  for I(1), então  $y_t$  e  $z_t$  são co-integradas. F  
d) se  $y_t$  for I(1),  $z_t$  for I(1) e  $u_t$  for I(0), então  $y_t$  e  $z_t$  são co-integradas. V

**Exercício 5.24.** Responda V ou F, justificando sua resposta. Com respeito à teoria das séries temporais, são corretas as afirmativas:

- a) Considere uma série temporal  $Y_t$  auto-regressiva de ordem 1 com parâmetro  $\rho$ . No modelo:  $Y_t - Y_{t-1} = \delta Y_{t-1} + u_t$ , em que  $u_t$  é um ruído branco e  $\delta = \rho - 1$ , se  $\delta$  for de fato igual a zero, a série  $Y_t$  será não estacionária. V  
b) Numa regressão linear simples de duas séries temporais não estacionárias de ordem 1, o teste usual  $t$  de Student ainda é válido. F  
c) Numa regressão linear múltipla de séries temporais de ordem 1, mas cointegráveis, não se corre o risco de os resultados serem espúrios. V  
d) Numa regressão linear múltipla de séries temporais de ordem 1, mas cointegráveis, os resíduos da regressão são estacionários. V  
e) Se uma série temporal tiver que ser diferenciada  $n$  vezes antes de se tornar estacionária, a série original é integrada de ordem  $n - 1$ . F

**Exercício 5.25.** Sejam  $Y_t$  e  $X_t$  duas séries temporais. Considere os resultados dos seguintes modelos de regressão estimados por mínimos quadrados ordinários (MQO):

$$\Delta \hat{Y}_t = 4,8788 - 0,1512Y_{t-1} \quad e \quad \Delta \hat{X}_t = 0,1094 - 0,1807X_{t-1}.$$

Considere também os resultados da regressão de  $Y_t$  em  $X_t$ .

$$Y_t = 23,3924 + 14,4006X_t + \hat{e}_t,$$

em que  $\hat{e}_t$  é o resíduo. Finalmente, considere a seguinte regressão:

$$\Delta \hat{e}_t = 0,0730 - 0,4157\hat{e}_{t-1}$$

Os números entre parênteses são os valores do teste t de significância individual dos parâmetros. Dado que o valor crítico a 5% da estatística de Dickey-Fuller é -2,938, é correto afirmar que:

- a)  $Y_t$  e  $X_t$  são séries temporais integradas de ordem 1.  
b) A regressão de  $Y_t$  em  $X_t$  é espúria.  
c) A hipótese de cointegração entre  $Y_t$  e  $X_t$  é rejeitada pois os resíduos da regressão de  $Y_t$  em  $X_t$  são não-estacionários.  
d) Para que duas variáveis sejam cointegradas é necessário que ambas tenham a mesma ordem de integração.  
e) A rejeição da hipótese nula do teste Dickey-Fuller implica que a variável em questão é não-estacionária.

**Exercício 5.26 (ANPEC 2013-10).** Julgue as seguintes afirmativas:

- 0) O passeio aleatório com drift,  $y_t = c + y_{t-1} + \varepsilon_t$ ,  $y_0 = 0$ , em que  $\varepsilon_t$  é um ruído branco, com média zero e variância  $\sigma^2$ , é um processo estacionário de segunda ordem se  $c = 0$ .  
1) O processo MA(1),  $y_t = \varepsilon_t + \theta_1\varepsilon_{t-1}$ , em que  $\varepsilon_t$  é um ruído branco, com média zero e variância  $\sigma^2$ , é estacionário de segunda ordem se, e somente se, a raiz do polinômio  $1 + \theta_1x$  cair fora do círculo unitário.

- 2) O processo MA(1),  $y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$ , em que  $\varepsilon_t$  é um ruído branco, com média zero e variância  $\sigma^2$ , é inversível se, e somente se,  $|\theta_1| < 1$ .
- 3) O processo AR(2),  $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$ , em que  $\varepsilon_t$  é um ruído branco, com média zero e variância  $\sigma^2$ , é estacionário de segunda ordem se

$$|\phi_2| < 1, \quad \phi_2 - \phi_1 < 1 \quad \text{e} \quad \phi_2 + \phi_1 < 1.$$

- 4) No passeio aleatório,  $y_t = y_{t-1} + \varepsilon_t$ ,  $y_0 = 0$ , em que  $\varepsilon_t$  é um ruído branco, com média zero e variância  $\sigma^2$ , a variância de  $y_t$  varia com  $t$ .



# Chapter 6

## Variáveis instrumentais

Em estatística, econometria, epidemiologia e áreas relacionadas, o método de variáveis instrumentais é usado para estimar relações causais, quando experimentos controlados não são viáveis. O método de **variáveis instrumentais** (*VI*) permite estimativas consistentes quando as variáveis explicativas são endógenas.

### 6.1 Conceito da exogeneidade dos regressores

Nos modelos econométricos estruturais de uma equação a variável dependente (**endógena**) é explicada através de um conjunto de variáveis explicativas (não estocásticas) e do termo de erro. Nos modelos econométricos estruturais de uma equação, as variáveis explicativas são a causa que explicam a variação da variável dependente e a variável endógena reflete o efeito provocado pela variação das variáveis explicativas. Uma das hipóteses básicas iniciais da regressão MQO admite a ausência de correlação das variáveis explicativas  $X_1, \dots, X_k$  com o termo de erro, isto é  $\text{Cov}(X_i, \varepsilon) = 0$ , para todo  $i = 1, \dots, k$ . Desta forma as variáveis explicativas são consideradas exógenas no modelo tradicional de regressão. Contudo, a hipótese da exogeneidade das variáveis explicativas é muitas vezes infringida. Entre as causas mais comuns da endogeneidade das variáveis explicativas podemos citar *erros de mensuração das variáveis*, *variável omitida* e *simultaneidade*.

Quando temos um problema de endogeneidade causado por uma *variável não-observável*, uma possível solução é a utilização de uma variável **proxy**.

#### 6.1.1 Variável proxy

No exemplo apresentado por (Wooldridge, 2016) é considerada a regressão para explicar a **renda** em função do **nível de escolaridade** (*educ*) e a aptidão para determinada tarefa (*aptid*),

$$\log(\text{salario}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{aptid} + u. \quad (6.1)$$

Observe que *aptid* é uma variável inata, não observável. Podemos esperar que *aptid* seja correlacionada com *educ* e, nesse caso, a omissão da variável *aptid* da regressão (6.1) (por motivos que veremos logo mais) torna a estimativa via MQO de  $\beta_1$  viciada e inconsistente. Uma maneira de contornar este problema é substituir a variável *aptid* por uma variável observável altamente correlacionada com *aptid* que será usada como proxy para aptidão. Tal variável é denominada variável **proxy** e idealmente deve satisfazer as seguintes condições:

- Ser altamente correlacionada com a variável não observada que pretende substituir;
- Não ser correlacionada com o termo de erro da regressão.

Com uma boa proxy, podemos estimar o modelo por MQO e obter “boas” estimativas para  $\beta_1$  (6.1). No que segue assumiremos que variáveis proxys não estão à disposição.

### 6.1.2 Variável omitida

O modelo (6.1) pode ser reescrito como

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon, \quad (6.2)$$

$Y$  representa o logaritmo do salário,  $X_1$  representa a variável educação,  $X_2$  a aptidão e  $\varepsilon$  representa o termo de erro do modelo, com  $\text{Cov}(X_i, \varepsilon) = 0$ ,  $i=1,2$ . Como já discutimos, a variável aptidão, não é observável e vamos assumir que uma variável proxy para aptidão não está disponível. Omitindo a variável  $X_2$  no modelo obtemos

$$Y = \beta_0 + \beta_1 X_1 + \xi \quad (6.3)$$

onde  $\xi = \beta_2 X_2 + \varepsilon$  representa o erro na regressão acima. Estimando  $\beta = (\beta_0, \beta_1)$  por MQO obtemos

$$\begin{aligned} \mathbb{E}(\hat{\beta}) &= \mathbb{E}\left(\frac{\text{Cov}(X_1, Y)}{\text{Var}(X_1)}\right) = \mathbb{E}\left(\frac{\text{Cov}(X_1, \beta_0 + \beta_1 X_1 + \xi)}{\text{Var}(X_1)}\right) \\ &= \mathbb{E}\left(\frac{\beta_1 \text{Var}(X_1) + \text{Cov}(X_1, \xi)}{\text{Var}(X_1)}\right) = \mathbb{E}\left(\frac{\beta_1 \text{Var}(X_1) + \text{Cov}(X_1, \xi)}{\text{Var}(X_1)}\right) \\ &= \beta_1 + \mathbb{E}\left(\frac{\text{Cov}(X_1, \xi)}{\text{Var}(X_1)}\right), \end{aligned}$$

de onde segue que o estimador de MQO é não viciado se, e somente se,  $\text{Cov}(X_1, \xi) = 0$ . Pode-se mostrar ainda que se  $\text{Cov}(X_1, \xi) \neq 0$ , o estimador de MQO será inconsistente. Agora observe que

$$\text{Cov}(X_1, \xi) = \text{Cov}(X_1, \beta_2 X_2 + \varepsilon) = \beta_2 \text{Cov}(X_1, X_2) + \text{Cov}(X_1, \varepsilon) = \beta_2 \text{Cov}(X_1, X_2)$$

que é zero se, e somente se,  $\text{Cov}(X_1, X_2) = 0$ , o que nem sempre acontece na prática. Em outras palavras, olhando (6.2) como um modelo geral de onde foi omitida a variável  $X_2$  resultando no modelo (6.3), se  $X_1$  e  $X_2$  são correlacionados então o estimador de MQO é viciado e inconsistente!

**Exemplo 6.1.** Suponha que um pesquisador deseja estimar o efeito causal do tabagismo sobre a saúde geral, como em Leigh and Schembri (2004). Em princípio, a existência de correlação entre a saúde e o hábito de fumar não implica necessariamente que o fumo piora a saúde, porque outras variáveis podem afetar tanto a saúde quanto o hábito de fumar. Por exemplo, pode ocorrer por acaso que pessoas de uma certa cidade exposta à poluição radioativa fumem muito, mas é a poluição que realmente causa problemas de saúde à esta população em estudo. Mesmo que o tabagismo cause realmente problemas, a saúde em si pode afetar o hábito de fumar (digamos, um paciente muito doente pode se sentir instigado a fumar mais). Fazer estudos controlados (por exemplo, colocar uma pessoa num laboratório, sem exposição à poluição, fumando quantidades controladas) pode ser difícil, caro ou antiético. Uma opção alternativa, portanto, seria o pesquisador tentar estimar o efeito causal do tabagismo sobre a saúde a partir de dados observacionais, utilizando, por exemplo, a alíquota de imposto sobre o tabaco como um instrumento para fumar em uma regressão de saúde. Se as alíquotas de imposto sobre o tabaco afetam apenas (positivamente, imagina-se) a saúde porque eles afetam o hábito de fumar (mantendo as outras variáveis do modelo fixas), a correlação entre impostos sobre o tabaco e a saúde é uma evidência de que o tabagismo provoca alterações na saúde. Uma estimativa do efeito do tabagismo sobre a saúde podem ser feita também fazendo uso da correlação entre os impostos e os hábitos de fumar.

### 6.1.3 Erros de mensuração

O problema de erros de mensuração em estudos práticos é bastante comum. Porém, dependendo da natureza do problema pode trazer consequências nefastas a estimação via MQO.

**Exemplo 6.2.** Para explicar o rendimento de um aluno da UFRGS, podemos estar interessados em usar como variáveis explicativas (entre outras): renda familiar, número de horas dedicadas ao estudo, tempo necessário para o trajeto casa-UFRGS, etc. Todas essas variáveis estão sujeitas a erros de mensuração, pois os alunos podem errar (deliberadamente ou não) ao responder à pesquisa. Se os erros forem puramente aleatórios, isto é, não estiverem correlacionados com outras variáveis relevantes, as hipóteses do modelo acima serão satisfeitas.

**Exemplo 6.3.** A Meta-análise é uma técnica que visa agregar a informação contida em várias fontes. Em áreas como música, cinema e televisão são muito comuns os sites que apresentam resumos das críticas de um determinado álbum/filme/série em uma única avaliação agregada. Exemplos são os sites Rotten Tomatoes, Metacritic, Allmusic, dentre muitos outros. A idéia estatística destes sites é obter uma avaliação para um determinado assunto a partir

da análise agragada de avaliações dadas por críticos, espectadores, mídia, blogs, sites, etc. Cada avaliação obtida é agregada ao total, de onde uma avaliação única é calculada. Para exemplificar vamos supor que estamos interessados na avaliação do último álbum do Metallica (Hard Wired... to Self-Destruct, 2016). Para isso, estabelecemos uma escala a avaliação do álbum e passamos a vasculhar a internet por informações a respeito do álbum. Vamos supor que encontramos 20 sites com avaliações do álbum. Cada uma dessas avaliações é baseada em diversas avaliações individuais de pessoas que visitaram o site. Por exemplo, a avaliação final de um determinado site é dada pela média das avaliações dos visitantes do site. No final, teremos coletado 20 avaliações de sites especializados, onde cada avaliação representa a média dos indivíduos que avaliaram o álbum. Desta forma os dados que coletamos não representam a opinião de ouvintes individuais (a população neste caso), mas a média dessas opiniões, que pode ser diferente de todas as avaliações dadas e que certamente varia de site pra site. Esta avaliação, portanto, considerada como uma observação de indivíduos da população, traz consigo um erro aleatório de medição gerada pelo agregamento das informações individuais.

#### 6.1.4 Variável com erro de medição

Considere o modelo de regressão simples:

$$Y = \beta_0 + \beta_1 X + \varepsilon, \quad (6.4)$$

em que  $\text{Cov}(X, \varepsilon) = 0$  e  $\text{Var}(\varepsilon) = \sigma_\varepsilon^2$ . A princípio, nesse contexto a estimativa por MQO deveria gerar estimadores consistentes dos parâmetros. Vamos supor que, por algum motivo, a variável  $X$  seja observada com um erro aleatório. Isto é, suponha que observamos na prática observamos

$$X^* = X + e, \quad (6.5)$$

onde assumimos que

$$\mathbb{E}(e) = 0; \quad \text{Cov}(X, e) = 0; \quad \text{Cov}(e, \varepsilon) = 0; \quad \text{e } \text{Var}(e) = \sigma_e^2 > 0.$$

A aleatoriedade dos erros é fundamental na análise que segue. Reescrevendo o modelo ?? em função da variável observada  $X^*$ :

$$Y = \beta_0 + \beta_1 X + \varepsilon = \beta_0 + \beta_1 (X^* - e) + \varepsilon = \beta_0 + \beta_1 X^* + \xi, \quad (6.6)$$

onde  $\xi = \varepsilon - \beta_1 e$  faz o papel de erro na regressão. Observe que os estimadores via MQO de  $\beta_0$  e  $\beta_1$  em ?? são exatamente os mesmos de ??, embora seja esta última regressão que estamos efetivamente estimando dada a presença de erros aleatórios nas observações. Agora

$$\begin{aligned} \text{Cov}(X^*, \xi) &= \text{Cov}(X + e, \varepsilon - \beta_1 e) = \text{Cov}(X, \varepsilon) - \beta_1 \text{Cov}(X, e) + \text{Cov}(e, \varepsilon) - \beta_1 \text{Var}(e) \\ &= -\beta_1 \sigma_e^2 \neq 0. \end{aligned}$$

Em outras palavras, a estimativa via MQO na presença de variáveis com erro (aleatório) de medição resulta em estimativas inconsistentes. Observe ainda que, dada uma amostra de tamanho  $n$  do modelo ??,

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = \beta + \frac{\sum_{i=1}^n (x_i - \bar{x}) \varepsilon_i}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

e note que  $\text{Var}(X^*) = \text{Var}(X) + \text{Var}(e) = \sigma_X^2 + \sigma_e^2$ , de onde segue que

$$\text{Plim}(\hat{\beta}_1) = \beta_1 + \frac{\text{Cov}(X^*, \varepsilon)}{\text{Var}(X^*)} = \beta_1 - \frac{\beta_1 \sigma_e^2}{\sigma_X^2 + \sigma_e^2} = \beta_1 \left(1 - \frac{\sigma_e^2}{\sigma_X^2 + \sigma_e^2}\right) \neq 0.$$

## 6.2 Variável instrumental

Considere o modelo

$$Y = \beta_0 + \beta_1 X + \varepsilon, \quad (6.7)$$

em que  $\text{Cov}(X, \varepsilon) \neq 0$ . Neste caso, os estimadores de MQO para  $\beta_0$  e  $\beta_1$  são viciados e inconsistentes e o problema é obter um estimador para estes parâmetros que apresentem propriedades melhores que o estimador de MQO. Para isso estudaremos o método de Variáveis Instrumentais.

Suponha que temos uma variável  $Z$  satisfazendo

$$\text{Cov}(Z, \varepsilon) = 0 \quad (6.8)$$

$$\text{Cov}(Z, X) \neq 0. \quad (6.9)$$

Uma variável  $Z$  satisfazendo (6.8) e (6.9) é chamada de *Variável Instrumental*, ou ainda, um *instrumento* para a variável  $X$ . Por razões que ficarão claras na prática, é desejável que  $|\text{Cor}(Z, X)|$  seja a mais alta possível.

A condição (6.9) é chamada de *relevância do instrumento* enquanto (??) é chamada de *exogeneidade do instrumento*.

Uma das maiores críticas em relação ao método de variáveis instrumentais é que a condição (6.8) não pode ser testada diretamente, pois o erro é não-observável. Desta forma, para justificarmos a validade da condição (6.8) precisamos recorrer à introspecção econômica e/ou argumentos filosóficos. É necessário uma boa “historinha” para justificar o instrumento! A condição (6.9), porém, pode ser testada indiretamente via uma regressão de  $X$  em  $Z$  (teste de significância de qual coeficiente?) ou ainda de um teste de hipótese direto do tipo  $H_0 : \text{Cor}(Z, X) = 0$  vs.  $H_1 : \text{Cor}(Z, X) \neq 0$

Infelizmente, é muito difícil encontrar instrumentos válidos para as variáveis de um determinado problema. Uma das razões dessa dificuldade reside no fato de que as duas condições requeridas de um instrumento são muitas vezes conflitantes, pois temos 3 variáveis  $X$ ,  $Z$  e  $\varepsilon$  tais que  $Z$  é correlacionado com  $X$  que por sua vez é correlacionado com  $\varepsilon$ , mas  $Z$  não pode ser correlacionado com  $\varepsilon$ .

**Exemplo 6.4** (Estimação de equação de salário em função da educação). | Variável omitida: **aptidão** do indivíduo - torna viesado e inconsistente o coeficiente da educação.

Possível instrumento: **educação da mãe** (correlacionada com a educação do indivíduo).

Mas, educação da mãe também deve ser correlacionada com a habilidade do indivíduo, presente no erro? Nesse caso, não seria um bom instrumento.

**Um possível instrumento seria o número de irmãos** - não correlacionado com *aptid* - correlacionado com *educ* (negativamente).

### 6.2.1 Diferença entre proxy e variável instrumental

A **variável proxy** é caracterizada por:

não ser correlacionada com o termo de erro do modelo;

correlacionada com a variável não observada.

A **variável instrumental** é caracterizada por

i) *Não ser correlacionada com o termo de erro do modelo, ou seja, não correlacionada com aptid no exemplo 6.4;*

ii) *Correlacionada com a variável endógena, ou seja, no caso do exemplo 6.4, ela deve ser correlacionada com educ.*

No exemplo 6.4 uma boa proxy seria:

$$QI = \begin{cases} \text{correlacionada com } \text{aptid}; \\ \text{não correlacionada com o erro}, \end{cases}$$

e um bom instrumento é o *número de irmãos (nir)*, como visto anteriormente.

$$nir = \begin{cases} \text{não correlacionada com } \text{aptid} \text{ e, portanto, não correlacionada com o erro}; \\ \text{correlacionada com educação (negativamente)}, \end{cases}$$

### 6.2.2 O método de estimação por variável instrumental

Para descrever o método de estimação por variável instrumental, considere o seguinte modelo:

$$Y = \beta_0 + \beta_1 X + \varepsilon. \quad (6.10)$$

em que  $\text{Cov}(X, \varepsilon) \neq 0$  e seja  $Z$  um instrumento para  $X$ . Para obtermos os *Estimadores de Variável Instrumental* para  $\beta_0$  e  $\beta_1$ , começamos calculando a covariância entre  $Z$  e  $Y$  como segue:

$$\begin{aligned} \text{Cov}(Z, Y) &= \text{Cov}(Z, \beta_0 + \beta_1 X + \varepsilon) = \underbrace{\text{Cov}(Z, \beta_0)}_{=0} + \text{Cov}(Z, \beta_1 X) + \underbrace{\text{Cov}(Z, \varepsilon)}_{=0} \\ &= \beta_1 \text{Cov}(Z, X) \end{aligned}$$

de onde segue que

$$\beta_1 = \frac{\text{Cov}(Z, Y)}{\text{Cov}(Z, X)}. \quad (6.11)$$

Dada uma amostra de tamanho  $n$  de  $X$ ,  $Y$  e  $Z$ , obtemos o estimador de VI para  $\beta_1$  substituindo-se as covariâncias que aparecem em (6.11) por suas versões amostrais, isto é,

$$\hat{\beta}_{1VI} = \frac{\sum_{i=1}^n (z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^n (z_i - \bar{z})(x_i - \bar{x})} = \frac{\sum_{i=1}^n z_i y_i - n\bar{z}\bar{y}}{\sum_{i=1}^n z_i x_i - n\bar{z}\bar{x}}.$$

Da mesma maneira como procedemos no caso de MQO, o estimador de VI para  $\beta_0$  é dado por

$$\hat{\beta}_{0VI} = \bar{Y} - \hat{\beta}_{1VI} \bar{X}.$$

### 6.3 Inferência com o estimador por variável instrumental

Pode-se mostrar que para  $n$  grande

$$\hat{\beta}_{1VI} \approx N \left( \beta_1, \frac{\sigma_\varepsilon^2}{n\sigma_X^2\rho_{XZ}^2} \right), \quad (6.12)$$

ou seja, para  $n$  grande a variância do estimador é dada por

$$\text{Var}(\hat{\beta}_{1VI}) = \frac{\text{Var}(\varepsilon)}{n\text{Var}(X)\text{Cor}(X, Z)^2} = \frac{\sigma_\varepsilon^2}{n\sigma_X^2\rho_{XZ}^2}. \quad (6.13)$$

A equação ??eq:bla nos revela algo fundamental na escolha de um bom instrumento: quanto maior a correlação entre o instrumento e a variável, menor é a variância do estimador de VI. Por esta razão, devemos procurar um instrumento que tenha a mais alta correlação possível com  $X$ . Sabemos que

$$\text{Var}(\hat{\beta}_{1MQO}) = \frac{\sigma_\varepsilon^2}{n\sigma_X^2}, \quad (6.14)$$

e desta forma obtemos que

$$\text{Var}(\hat{\beta}_{1VI}) = \frac{\sigma_\varepsilon^2}{n\sigma_X^2\rho_{XZ}^2} = \frac{\text{Var}(\hat{\beta}_{1MQO})}{\rho_{XZ}^2}$$

de onde concluímos que  $\text{Var}(\hat{\beta}_{1VI}) \geq \text{Var}(\hat{\beta}_{1MQO})$  com igualdade ocorrendo se, e somente se,  $|\text{Cor}(X, Z)| = 1$ . Ou seja, a menos que  $Z$  seja um instrumento perfeito para  $X$  (o que não ocorre na prática), o estimador de MQO sempre possui menor variância que o estimador de VI.

O viés assintótico de VI e MQO pode ser estudado usando os respectivos limites de probabilidade, que são:

$$\text{plim}(\hat{\beta}_{1VI}) = \beta_1 + \frac{\text{Cor}(z, \varepsilon)}{\text{Cor}(X, Z)} \frac{\sigma_\varepsilon}{\sigma_X} \quad (6.15)$$

e

$$\text{plim}(\hat{\beta}_{1MQO}) = \beta_1 + \text{Corr}(X, \varepsilon) \frac{\sigma_\varepsilon}{\sigma_X}. \quad (6.16)$$

### 6.3.1 Teste de hipóteses

Consideremos o seguinte problema

$$\log(\text{salário}) = \beta_0 + \beta_1 \text{educ} + u, \quad (6.17)$$

em que  $\text{Cov}(\text{educ}, u) \neq 0$ . Seja o *número de irmãos*, (*nir*) um instrumento. Queremos testar a significância de  $\hat{\beta}_{1VI}$ . Os passos para esse teste são:

i) Obter

$$\hat{\beta}_{1VI} = \frac{\sum_{i=1}^n (\text{nir}_i - \bar{\text{nir}})(y_i - \bar{y})}{\sum_{i=1}^n (\text{nir}_i - \bar{\text{nir}})(x_i - \bar{x})};$$

ii) Obter

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n \hat{u}_{iVI}^2}{n - 2},$$

em que  $\hat{u}_i^{VI} = y_i - \hat{\beta}_{0VI} - \hat{\beta}_{1VI}\text{educ}$ ;

iii) Obter  $R_{\text{educ}, \text{nir}}^2$ , que é o  $R^2$  de

$$\text{educ} = \alpha_0 + \alpha_1 \text{nir} + v;$$

vi) Obter a estimativa do desvio padrão

$$S(\hat{\beta}_{1VI}) = \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{x})^2 R_{\text{educ}, \text{nir}}^2}};$$

v) Teste:  $\begin{cases} H_0 : \beta_{1VI} = 0; \\ H_1 : \beta_{1VI} \neq 0. \end{cases}$

Um exemplo para ver a diferença entre as estimativas obtidas pelos dois métodos de estimação (MQO e VI), é o exemplo abaixo, reproduzido com pequenas adaptações do exemplo 15.1 do Wooldridge (2016).

**Exemplo 6.5.** Utilizamos os dados sobre mulheres casadas que trabalham contidos no arquivo MROZ.RAW para estimar o retorno da educação no modelo de regressão simples

$$Y = \beta_0 + \beta_1 X + \varepsilon,$$

onde  $Y$  é o logaritmo do salário das mulheres casadas por hora e  $X$  representa a educação. Para comparação, obtemos a reta estimada via MQO para  $Y$ ,

$$\begin{aligned} \hat{Y} &= -0.185 + 0.109X \\ &\quad (0.185) \quad (0.014) \end{aligned} \quad (6.18)$$

onde  $n = 428$ . Obtemos ainda  $R^2 = 0.118$ . A estimativa de  $\beta_1$  implica um retorno de perto de 11% para um ano a mais de educação. Em seguida utilizaremos a variável  $\$Z = \$$  educação dos pais como variável instrumental para  $X$ . Primeiramente, para determinarmos que  $\text{Cov}(Z, X) \neq 0$ , fazemos uma regressão simples de  $X$  em  $Z$  de onde obtemos a reta de ajustada

$$\begin{aligned} \hat{X} &= 10.24 + 0.269Z \\ &\quad (0.28) \quad (0.029) \end{aligned}$$

com  $R^2 = 0.173$ . A utilização de  $Z$  como uma VI para  $X$  produz

$$\begin{aligned} \hat{Y} &= -0.441 + 0.059X \\ &\quad (0.446) \quad (0.035) \end{aligned} \quad (6.19)$$

com  $R^2 = 0.093$ . A estimativa de VI do retorno da educação é de 5.9%, que é pouco mais da metade da estimativa de MQO. Isso sugere que a estimativa de MQO é alta demais. Porém devemos lembrar que como as estimativas são

baseadas em dados amostrais, não temos como saber qual está mais próximo do verdadeiro valor. Note ainda que o erro padrão da estimativa de VI para  $\beta_1$  é duas vezes maior que a de MQO. O intervalo de confiança à 95% de  $\beta_1$  utilizando MQO é muito mais estreito do que utilizando VI; de fato, o intervalo de confiança da VI contém a estimativa de MQO. Portanto, embora as diferenças entre (6.19) e (??) sejam grandes na prática, não podemos afirmar com certeza que as diferenças são estatisticamente significantes. O fato de o intervalo de VI conter a estimativa de MQO é uma evidência contra essa hipótese.

## 6.4 Variáveis instrumentais em modelos de regressão múltipla

O estimador de VI para o modelo de regressão simples é facilmente estendido para o caso de regressão múltipla. Consideremos inicialmente o caso em que somente um dos regressores é endógeno. A equação estrutural do modelo linear padrão com duas variáveis explicativas é

$$Y_1 = \beta_0 + \beta_1 Y_2 + \beta_2 Z + \varepsilon. \quad (6.20)$$

Usamos essa notação para identificar as variáveis endógenas ( $Y$ 's) e as variáveis exógenas ( $Z$ 's). Um exemplo para (6.20) é

$$\log(\text{salario}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \varepsilon, \quad (6.21)$$

em que  $Y_1 = \log(\text{salário})$ ,  $Y_2 = \text{educ}$  e  $Z_1 = \text{exper}$ . Assim, estamos supondo que  $\text{exper}$  é exógeno e, por razões habituais, presumimos que  $\text{educ}$  seja correlacionado com  $\varepsilon$ .

Se estimarmos (6.20) por MQO, todos os estimadores serão viesados e inconsistentes. Assim, seguimos a estratégia da seção anterior para encontrar uma VI para  $Y_2$ . Como  $Z_1$  aparece como variável explicativa, devemos procurar outra variável exógena (vamos chamá-la de  $Z_2$ ) que seja correlacionada com  $Y_2$ .

A correlação de  $Z_2$  e  $Y_2$  pode ser afetada por  $Z_1$  que aparece na regressão (6.20). Devemos nos certificar que na presença de  $Z_1$ ,  $Y_2$  e  $Z_2$  ainda são correlacionados. Essa condição pode ser verificada a partir da regressão

$$Y_2 = \pi_0 + \pi_1 Z_1 + \pi_2 Z_2 + \epsilon, \quad (6.22)$$

em que  $\mathbb{E}(\epsilon) = 0$ ,  $\text{Cov}(Z_1, \epsilon) = 0$ ,  $\text{Cov}(Z_2, \epsilon) = 0$  e os  $\pi_j$  são parâmetros desconhecidos. A condição de identificação fundamental é que

$$\pi_2 \neq 0,$$

a qual pode ser testada usando-se um teste  $t$ .

*Observação.* A equação (??) é um exemplo de uma *equação na forma reduzida*, significando que escrevemos uma variável endógena em função de exógenas. O nome ajuda a distingui-lá da equação estrutural (6.20).

A adição de mais variáveis explicativas exógenas ao modelo é direta. O modelo estrutural com  $k - 1$  variáveis exógenas ( $Z_1, Z_2, \dots, Z_{k-1}$ ) e uma variável endógena ( $Y_2$ ) é

$$Y_1 = \beta_0 + \beta_1 Y_2 + \beta_2 Z_1 + \beta_3 Z_2 + \dots + \beta_k Z_{k-1} + \varepsilon. \quad (6.23)$$

Se  $Z_k$  é uma variável exógena e queremos usá-la como uma VI para  $Y_2$ , devemos proceder de forma semelhante a (??), escrevendo a equação na forma reduzida

$$Y_2 = \pi_0 + \pi_1 Z_1 + \pi_2 Z_2 + \dots + \pi_{k-1} Z_{k-1} + \pi_k Z_k + v_2, \quad (6.24)$$

para a qual é necessário que exista alguma correlação parcial entre  $Z_k$  e  $Y_2$ :

$$\pi_k \neq 0.$$

### 6.4.1 Mínimos quadrados em dois estágios (MQ2E)

Nesta subseção vamos mostrar a equivalência do estimador VI com MQ2E. Para isso consideremos a equação estrutural com uma variável endógena

$$Y = \beta_0 + \beta_1 X + \varepsilon, \quad (6.25)$$

em que  $\text{Cov}(X, \varepsilon) \neq 0$ . Seja  $Z$  um instrumento para  $X$ .

O procedimento MQ2E consiste em:

- i) Estimar a equação reduzida

$$X = \alpha_0 + \alpha_1 Z + \epsilon, \quad (6.26)$$

encontrando as estimativas de  $\alpha_0$  e  $\alpha_1$  via MQO.

- ii) Obter a reta ajustada

$$\hat{x}_i = \hat{\alpha}_0 + \hat{\alpha}_1 z_i.$$

- iii) Estimar  $\beta_0$  e  $\beta_1$  via MQO no seguinte modelo:

$$y_i = \beta_0 + \beta_1 \hat{x}_i + \eta_i, \quad (6.27)$$

para  $i = 1, \dots, n$ , ou seja, usar  $\hat{X}$  como uma VI para  $X$ .

Esse procedimento resulta em uma estimativa para  $\beta_1$  equivalente a  $\hat{\beta}_{1VI}$ , como será demonstrado a seguir.

Note que:

$$\begin{aligned} \hat{x}_i &= \hat{\alpha}_0 + \hat{\alpha}_1 z_i \\ &= \bar{x} - \hat{\alpha}_1 \bar{z} + \hat{\alpha}_1 z_i \\ &= \bar{x} + \hat{\alpha}_1 (z_i - \bar{z}). \end{aligned}$$

De (6.27), segue que

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (\hat{x}_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (\hat{x}_i - \bar{x})^2}.$$

Pode-se mostrar facilmente que  $\bar{\hat{x}} = \bar{x}$ . Logo,

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_{i=1}^n \hat{\alpha}_1 (z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^n (\hat{\alpha}_1 (z_i - \bar{z}))^2} \\ &= \hat{\alpha}_1 \frac{\sum_{i=1}^n (z_i - \bar{z})(y_i - \bar{y})}{\hat{\alpha}_1^2 \sum_{i=1}^n (z_i - \bar{z})^2} \\ &= \frac{\sum_{i=1}^n (z_i - \bar{z})(y_i - \bar{y})}{\frac{\sum_{i=1}^n (z_i - \bar{z})(x_i - \bar{x})}{\sum_{i=1}^n (z_i - \bar{z})^2} \sum_{i=1}^n (z_i - \bar{z})^2} \\ &= \frac{\sum_{i=1}^n (z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^n (z_i - \bar{z})(x_i - \bar{x})} \\ &= \hat{\beta}_{1VI}. \end{aligned}$$

Quando existir mais de uma VI disponível, digamos  $Z_1, Z_2, Z_3$ , qualquer combinação linear destas VI's também será uma VI válida. Devemos escolher a combinação linear com a mais alta correlação com a variável endógena  $Y_2$ . Isto acaba sendo fornecido pela equação na forma reduzida de  $Y_2$ , que é dada por

$$Y_2 = \pi_0 + \pi_1 Z_1 + \pi_2 Z_2 + \pi_3 Z_3 + \epsilon, \quad (6.28)$$

em que  $\mathbb{E}(\epsilon) = 0$ ,  $\text{Cov}(Z_1, \epsilon) = 0$ ,  $\text{Cov}(Z_2, \epsilon) = 0$  e  $\text{Cov}(Z_3, \epsilon) = 0$ . Portanto, a melhor VI de  $Y_2$  é a combinação linear dos  $Z_j$ , que denotaremos por

$$Y_2^* = \pi_0 + \pi_1 Z_1 + \pi_2 Z_2 + \pi_3 Z_3. \quad (6.29)$$

Para que esta VI não seja perfeitamente correlacionada com  $Z_1$ , precisamos que

$$\pi_2 \neq 0 \text{ ou } \pi_3 \neq 0. \quad (6.30)$$

Se essa hipótese de identificação, a qual pode ser testada usando-se um teste  $F$ , for satisfeita, então podemos usar  $Y_2^*$  como VI para  $Y_2$ .

*Observação.* Quando o problema é a existência de variáveis explicativas endógenas múltiplas, será necessária que existam pelo menos tantas variáveis exógenas excluídas do modelo quantas forem as endógenas incluídas.

*Observação.* Se para cada endógena existe uma única exógena, então podemos chamar o método de estimação de VI ou MQ2E.

*Observação.* Os testes de hipóteses múltiplas em um modelo estimado por MQ2E devem ser feitos com bastante cuidado. É possível que na estimação por VI ou MQ2E o  $R^2$  seja negativo. Isso influencia diretamente a estatística  $F$  e consequentemente qualquer conclusão a respeito das hipóteses.

## 6.5 Testes de endogeneidade

Como visto anteriormente, a condição (6.8) não pode ser testada, logo não podemos ter certeza se os regressores (variáveis explicativas) estão ou não correlacionados com o erro. Se não houver correlação, é melhor utilizar mínimos quadrados ordinários que o estimador de variáveis instrumentais ou MQ2E. No entanto, pela estimação pura e simples é impossível descobrir se há correlação entre os regressores e o erro.

### 6.5.1 Teste de Hausmann

O teste de especificação proposto por Hausman (1978) é um teste utilizado para avaliar a consistência de um estimador comparado a um outro estimador alternativo e pode ser utilizado no contexto de endogeneidade.

A lógica de Hausmann é a seguinte: sob a hipótese nula (ausência de correlação entre os regressores e o termo de erro), o econometrista tem em mãos dois estimadores consistentes para a matriz de parâmetros: o estimador de mínimos quadrados ordinários  $\hat{\beta}_{MQO}$  e o estimador de variáveis instrumentais  $\hat{\beta}_{VI}$ . Sob a hipótese alternativa, no entanto, somente um destes,  $\hat{\beta}_{VI}$ , é consistente. Portanto, a sugestão foi examinar a diferença  $d = \hat{\beta}_{VI} - \hat{\beta}_{MQO}$ . O resultado desta diferença converge em probabilidade para zero apenas sob a hipótese nula. Podemos testar esta hipótese usando o teste de Wald.

### 6.5.2 Teste de regressão

Para ilustrar, suponha que temos uma única variável suspeita de ser endógena,

$$Y_1 = \beta_0 + \beta_1 Y_2 + \beta_2 Z_1 + \beta_3 Z_2 + \varepsilon, \quad (6.31)$$

em que  $Z_1$  e  $Z_2$  são exógenos. Temos duas outras variáveis exógenas,  $Z_3$  e  $Z_4$ , que não aparecem no modelo (6.31). Se  $y_2$  for não correlacionado com  $\varepsilon$ , devemos estimar (6.31) por MQO. Wooldridge (2016) sugere que é mais fácil usar um teste de regressão para testar se  $Y_2$  é endógena. Isto é feito com base na estimação da forma reduzida de  $Y_2$ , que neste caso é

$$Y_2 = \pi_0 + \pi_1 Z_1 + \pi_2 Z_2 + \pi_3 Z_3 + \pi_4 Z_4 + \epsilon. \quad (6.32)$$

Agora, como cada  $Z_j$  é não correlacionado com  $\varepsilon$ ,  $Y_2$  será não correlacionado com  $\varepsilon$  se, e somente se,  $\epsilon$  for correlacionado com  $\varepsilon$ ; isso é o que queremos testar. Deveríamos escrever o modelo

$$\varepsilon = \delta_1 \epsilon + \eta_1 \quad (6.33)$$

e testar se  $\delta_1 = 0$ . No entanto, não temos  $\varepsilon$  nem  $\epsilon$ . A sugestão é usar  $\hat{\epsilon}_2$  como um regressor em (6.31).

O método pode ser resumido da seguinte forma:

- i) Obter via MQO os resíduos

$$\hat{\epsilon} = Y_2 - \hat{\pi}_0 + \hat{\pi}_1 Z_1 + \hat{\pi}_2 Z_2 + \hat{\pi}_3 Z_3 + \hat{\pi}_4 Z_4;$$

- ii) Estimar via MQO a regressão

$$Y_1 = \beta_0 + \beta_1 Y_2 + \beta_2 Z_1 + \beta_3 Z_2 + \delta_1 \hat{Y}_2 + e;$$

- iii) Testar  $H_0 : \delta_1 = 0$  via teste  $t$  (robusto a heterocedasticidade). Se rejeitamos  $H_0$  a um pequeno nível de significância, concluímos que  $Y_2$  é endógeno porque  $\epsilon$  e  $\varepsilon$  são correlacionados.

## 6.6 Exercícios

**Exercício 6.1.** Qual a motivação para o uso da estimação por variáveis instrumentais? Como se pode definir um bom instrumento?

**Exercício 6.2.** O que são variáveis endógenas e exógenas?

**Exercício 6.3.** Quais são as características de uma boa variável proxy e um bom instrumento?

**Exercício 6.4.** Seja  $X$  a matriz cujas colunas são compostas pelas variáveis  $x_1, \dots, x_k$ . As condições  $\mathbb{E}[u/X] = 0$  e  $\text{Cov}(u, x_j) = 0$ , para todo  $j = 1, \dots, k$ , são equivalentes? Explique o que estas condições significam na prática.

**Exercício 6.5.** Dê um exemplo em que a condição  $\mathbb{E}[u/X]$  não é válida.

**Exercício 6.6.** Quais as propriedades do estimador de variáveis instrumentais quando se possui uma variável instrumental fraca (fraco instrumento)?

**Exercício 6.7.** Qual a diferença da estimação de variáveis instrumentais para a estimação de mínimos quadrados em dois estágios? Explique e mostre as hipóteses de identificação necessárias. ““

**Exercício 6.8.** Em um modelo com duas variáveis regressoras,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u,$$

em que  $\text{Cov}(x_1, x_2) \neq 0$ , as estimativas dos parâmetros  $\beta_1$  e  $\beta_2$  obtidas via MQO serão viesadas?

**Exercício 6.9.** Suponha que você queira estimar o efeito da frequência às aulas sobre o desempenho dos alunos (*respad* - resultado padronizado em um exame final). Um modelo básico é

$$\text{respad} = \beta_0 + \beta_1 \text{taxafreq} + \beta_2 \text{prsGPA} + \beta_3 \text{ACT} + u,$$

em que *taxafreq* é a taxa de frequência, *prsGPA* é a média geral das notas em curso superior no último semestre.

a) Defina *dist* como a distância da residência do aluno até o local de estudos. Você considera que *dist* é não correlacionada com *u*?

b) Suponha que *dist* e *u* não sejam correlacionadas, que outra hipótese *dist* terá que satisfazer para ser uma VI válida de *taxafreq*?

c) Suponha que adicionemos o termo de interação *prsGPA.taxafreq*

$$\text{respad} = \beta_0 + \beta_1 \text{taxafreq} + \beta_2 \text{prsGPA} + \beta_3 \text{ACT} + \beta_4 \text{prsGPA.taxafreq} + u.$$

Se *taxafreq* for correlacionada com *u*, então em geral, *prsGPA.taxafreq* também será. O que poderia ser uma boa VI para *prsGPA.taxafreq*?

**Exercício 6.10.** Em um modelo com duas variáveis regressoras,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u,$$

em que  $x_2$  é endógena, a estimativa do parâmetro  $\beta_1$  obtida via MQO é viesada mesmo que  $\text{Cov}(x_1, x_2) = 0$ ?

**Exercício 6.11.** No modelo

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u,$$

a omissão de uma variável explicativa relevante  $x_2$ , para explicar a variável dependente  $y$  torna a estimativa dos coeficientes  $\beta_0$  e  $\beta_1$  obtidas via MQO viesadas?

**Exercício 6.12** (Experimento de Monte Carlo). Esse exercício encontra-se no capítulo 8 de Dougherty (2011) e relata um experimento de Monte Carlo que investiga o desempenho dos métodos MQO e IV ao estimar a equação de inflação dos preços no modelo de inflação de preços/inflação do salário. Valores numéricos foram atribuídos aos parâmetros das equações da seguinte forma:

$$p = 1.5 + 0.5w + u_p \tag{6.34}$$

$$w = 2.5 + 0.5p - 0.4U + u_w,$$

em que para *U* foi atribuído os valores 2, 2.25, 2.50, ... aumentando em passos de 0.25 até 6.75.  $u_w$  foi gerado como uma variável aleatória normal com média 0 e variância unitária, escalado por um fator 0.8. O termo distúrbio  $u_w$  foi gerado de forma independente, como uma variável aleatória normal com média 0 e variância unitária. Cada repetição do experimento utilizou uma amostra de 20 observações. A tabela a seguir mostra as estimativas via MQO e VI do intercepto,  $b_1$ , e do coeficiente de *w*,  $b_2$ , da equação (6.34), e os respectivos erros padrões (e.p.).

| Sample | MQO   |               |       |               | VI    |               |       |
|--------|-------|---------------|-------|---------------|-------|---------------|-------|
|        | $b_1$ | e.p.( $b_1$ ) | $b_2$ | e.p.( $b_2$ ) | $b_1$ | e.p.( $b_1$ ) | $b_2$ |
| 1      | 0.36  | 0.39          | 1.11  | 0.22          | 2.33  | 0.97          | 0.16  |
| 2      | 0.45  | 0.38          | 1.06  | 0.17          | 1.53  | 0.57          | 0.53  |
| 3      | 0.65  | 0.27          | 0.94  | 0.12          | 1.13  | 0.32          | 0.70  |
| 4      | 0.41  | 0.39          | 0.98  | 0.19          | 1.55  | 0.59          | 0.37  |
| 5      | 0.92  | 0.46          | 0.77  | 0.22          | 2.31  | 0.71          | 0.06  |
| 6      | 0.26  | 0.35          | 1.09  | 0.16          | 1.24  | 0.52          | 0.59  |
| 7      | 0.31  | 0.39          | 1.00  | 0.19          | 1.52  | 0.62          | 0.33  |
| 8      | 1.06  | 0.38          | 0.82  | 0.16          | 1.95  | 0.51          | 0.41  |
| 9      | -0.08 | 0.36          | 1.16  | 0.18          | 1.11  | 0.62          | 0.45  |
| 10     | 1.12  | 0.43          | 0.69  | 0.20          | 2.26  | 0.61          | 0.13  |

Compare as estimativas via MQO e VI em termos de vício e variância.

**Exercício 6.13** (ANPEC 2008). Suponha que o modelo abaixo descreva as relações entre quatro variáveis aleatórias escalares:  $y, X, Z$ , e  $V$ .

$$\mathbb{E}(y/X, Z) = \beta_0 + \beta_1 X + \beta_2 Z$$

$$X = \alpha_0 + \alpha_1 Z + v, \quad \mathbb{E}(v/Z, X) = E(v/Z) = \mathbb{E}(v/X) = \mathbb{E}(v) = 0.$$

Suponha, ainda, que  $\beta_0 \neq 0$ ,  $\beta_1 \neq 0$ ,  $\beta_2 \neq 0$ ,  $\alpha_0 \neq 0$  e  $\alpha_1 \neq 0$ .

- a) Calcule  $\mathbb{E}(y/Z)$ .
- b) Seja  $y = \beta_0 + \beta_1 X + \beta_2 Z + u$ . Calcule  $\mathbb{E}(u/X, Z)$ .
- c) Calcule  $\mathbb{E}(X/Z)$ .
- d) Seja

$$y = \theta_0 + \theta_1 Z + \varepsilon$$

em que  $\theta_0 = \beta_0 + \beta_1 \alpha_0$  e  $\theta_1 = \beta_1 \alpha_1 + \beta_2$ . Calcule  $\mathbb{E}(\varepsilon/Z)$ .

- e) Considere uma amostra de tamanho  $n$  das variáveis  $y$ ,  $X$  e  $Z$ . O estimador

$$T = \frac{\sum_{i=1}^n ny_i(z_i - \bar{z})}{\sum_{i=1}^n n(z_i - \bar{z})^2}$$

é um estimador não tendencioso para  $\theta_1 = \beta_1 \alpha_1 + \beta_2$ ?

**Exercício 6.14.** Mostre que o  $R^2$  da estimativa via VI pode ser negativo.

**Exercício 6.15.** Suponha que o modelo

$$y = \beta_0 + \beta_1 x + u,$$

em que  $x$  é endógena e  $z$  é um instrumento para  $x$ . Para comparar a variância de VI com MQO suponha que  $\sigma_x = \sigma_u$ , ou seja, a variação populacional contida no termo de erro seja a mesma contida em  $x$ . Suponha que a variável instrumental  $z$  seja levemente correlacionada com  $u$ ,  $\text{Cor}(z, u) = 0, 1$ . Suponha também que  $z$  e  $x$  tenha uma correlação um pouco maior:  $\text{Cor}(z, x) = 0, 2$ .

- a) Qual será o viés assintótico no estimador de VI?
- b) Quanta correlação deverá existir entre  $x$  e  $u$  antes que o MQO tenha mais viés assintótico que o MQ2E?

# Bibliography

- Box, G. E. and Jenkins (1970). *Time series analysis: forecasting and control*. San Francisco: Holden-Day.
- Dougherty, C. (2011). *Introduction to econometrics*. Oxford University Press.
- Gujarati, D. N. (2003). Basic econometrics. forth edition. *Singapura: McGraw-Hill*.
- Harris, F. H. d., McInish, T. H., Shoesmith, G. L., and Wood, R. A. (1995). Cointegration, error correction, and price discovery on informationally linked security markets. *JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS*, 30:N0.
- Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the econometric society*, pages 1251–1271.
- Koopmans, L. (1974). The spectral analysis of time series academic press. *New York, San Francisco, London*.
- Leigh, J. P. and Schembri, M. (2004). Instrumental variables technique: cigarette price provided better estimate of effects of smoking on sf-12. *Journal of clinical epidemiology*, 57(3):284–293.
- Marple, L. (1980). A new autoregressive spectrum analysis algorithm. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(4):441–454.
- Morettin, P. (1978). Estimation of the spectrum and of the covariance function of a dyadic-stationary series. *Boletim da Sociedade Brasileira de Matemática-Bulletin/Brazilian Mathematical Society*, 9(2):83–88.
- Watts, D. G. and Jenkins, G. (1968). Spectral analysis and its applications. *San Francisco*.
- Wooldridge, J. M. (2016). *Introductory econometrics: A modern approach*. Nelson Education.