UFPR: Análise de Séries Temporais - Lista 01

Luiz Paulo Tavares

2024-08-27

1) Defina variável aleatória, processo estocástico e série temporal.

De forma simples, uma variável aleatória é uma função que associa um número real a cada resultado. Formalmente, uma **variável aleatória** Z é uma função que mapeia o espaço amostral Ω de um experimento aleatório para o conjunto dos números reais $\mathbb{R}$, ou seja, $Z : \Omega \rightarrow \mathbb{R}$. Para cada resultado $\omega \in \Omega$, a variável Z associa um valor real $Z(\omega)$. Por outro lado, **processo estocástico** representa uma coleção de variáveis aleatórias $Z(t)_{t \in T}$ indexados para um conjunto T. Por fim, série temporal é um sequência de observações de uma variável Z_t ao longo do tempo t. Em outras palavras: uma **série temporal** é uma coleção de valores $Z(t)_{t \in T}$, onde cada Z_t é uma realização da variável aleatória Z no tempo t.

2) Defina processo estocástico ergódico e processo estocástico estacionário (no sentido amplo (fraco) e no sentido estrito (forte)).

Um processo estocástico é ergódico se, e somente se, as propriedades estatísticas do processo podem ser inferidas a partir de uma realização ao longo do tempo. Ou seja, processo ergódico tem relação com a representatividade amostral.

Por sua vez, processo estacionário divide-se em duas tipologias: no sentido amplo (fraco) ou no sentido estrito (forte). Pois bem, em sentido amplo é quando a média e variância são constantes ao longo do tempo e a função de autocovariância $Cov(Z(t), Z(t+h))$ depende apenas do deslocamento h, e não dos tempos absolutos t e $t+h$. Propriedades:

$$\mathbb{E}(Z_t) = \mu$$

$$\mathbb{E}(Z_t - \mu)^2 = \mu^2 < \infty$$

$$\mathbb{E}[(Z_t - \mu)(Z_{t+k} - \mu)] = \gamma k \forall t \in T$$

O sentido estrito (forte), por outro lado, é quando a distribuição conjunta de qualquer número finito de variáveis aleatórias do processo não depende do tempo. Formalmente: para quaisquer tempos $t_1, t_2, \ldots, t_n$ e para qualquer deslocamento h, a distribuição conjunta das variáveis $Z(t_1), Z(t_2), \ldots, Z(t_n)$ é idêntica à distribuição conjunta das variáveis $Z(t_1 + h), Z(t_2 + h)$, …, $Z(t_n + h)$. Isso pode ser expresso como:

$$(Z(t_1), Z(t_2), \ldots, Z(t_n)) \stackrel{d}{=} (Z(t_1 + h), Z(t_2 + h), \ldots, Z(t_n + h))$$

onde $\stackrel{d}{=}$ indica igualdade em distribuição.

3) Qual a diferença entre correlação serial e correlação simples? Exemplifique.

A diferença é a estrutura temporal subjacente na computação da correlação. Na correlação simples, como na correlação de Pearson, por exemplo, há o cálculo de uma simples associação linear entre as variáveis Z e Y; por outro lado, na correlação serial (autocorrelação), busca-se computar a correlação entre as observações da série temporal Z ao longo do tempo t.

Formalmente, a correlação de Pearson pode ser expressa como segue:

$$\rho_{Z,Y} = \frac{\sum_{i=1}^{n}(Z_i - \bar{Z})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(Z_i - \bar{Z})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

Podemos compactar a notação como segue:

$$r = \frac{Cov(Z,Y)}{\sqrt{Var(z)Var(y)}}$$

Ou ainda:

$$r = \frac{Cov(Z,Y)}{\sigma(z)\sigma(y)}$$

Enquanto a autocorrelação amostral de k-ésima ordem:

$$\hat{p_k} = \frac{\sum_{t=1}^{n-k}(Z_t - \bar{Z})(Z_{t+k} - \bar{Z})}{\sum_{t=1}^{n}(Z_t - \bar{Z})^2}$$

4) Determine a expressão da função de autocovariância (FACV) de um processo autoregressivo de ordem 3, AR(3)

Pode ser expresso como segue:

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \phi_3\gamma_{k-3} + \mathbb{E}(a_t\omega_{t-k})$$

5) Determine a expressão da variância de um processo autoregressivo de ordem 3 AR(3)

Dado um processo AR(3):

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \phi_3 Z_{t-3} + a_t$$

Podemos agora substituir a equação do processo AR(3) na definição de variância:

$$Var(Z_t) = \mathbb{E}[Z_t^2]$$
$$Z_t^2 = (\phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \phi_3 Z_{t-3} + a_t)^2$$

Expandindo os termos quadrados:

$$Z_t^2 = \phi_1^2 Z_{t-1}^2 + \phi_2^2 Z_{t-2}^2 + \phi_3^2 Z_{t-3}^2 + 2\phi_1\phi_2 Z_{t-1}Z_{t-2} + 2\phi_1\phi_3 Z_{t-1}Z_{t-3} + 2\phi_2\phi_3 Z_{t-2}Z_{t-3} + a_t^2$$

Como a_t é um ruído branco:

$$\mathbb{E}[a_t^2] = \sigma_a^2$$
$$Var(Z_t) = \phi_1^2 Var(Z_{t-1}) + \phi_2^2 Var(Z_{t-2}) + \phi_3^2 Var(Z_{t-3}) + \sigma_a^2$$

Usando a propriedade de estacionariedade:

$$Var(Z_t) = Var(Z_{t-1}) = Var(Z_{t-2}) = Var(Z_{t-3}) = \gamma(0)$$

Assim temos:

$$\gamma(0) = \phi_1^2\gamma(0) + \phi_2^2\gamma(0) + \phi_3^2\gamma(0) + \sigma_a^2$$

Reorganizando:

$$\gamma(0) - \phi_1^2\gamma(0) - \phi_2^2\gamma(0) - \phi_3^2\gamma(0) = \sigma_a^2$$

Colocando $\gamma(0)$ em evidência:

$$\gamma(0)(1 - \phi_1^2 - \phi_2^2 - \phi_3^2) = \sigma_a^2$$

Resolvendo, temos a variância do processo AR(3):

$$Var(Z_t) = \gamma(0) = \frac{\sigma_a^2}{1 - \phi_1^2 - \phi_2^2 - \phi_3^2}$$

6) Determine a expressões da FAC, da variância das estruturas indicadas a seguir, em função dos parâmetros do modelo de cada estrutura.

AR(1)

$$Z_t = \phi_1 Z_{t-1} + a_t$$

Expandindo usando a forma geral das equações de Yule-Walker. Para AR(1):

$$\rho = \phi_1 \rho_{k-1}$$

No lag k = 1:

$$\rho_1 = \phi_1$$

No lag k = 2, usando a relação recursiva:

$$\rho_2 = \phi_1 \rho_1 = \phi_1 \phi_1 = \phi_1^2$$

Para k = 3:

$$\rho_3 = \phi_1 \rho_2 = \phi_1 \phi_1^2 = \phi_1^3$$

Forma geral para o AR(1):

$$\rho_k = \phi_1^k$$

A variância, por sua vez:

$$\gamma_0 = \sigma_\omega^2 = \frac{\sigma_a^2}{1 - \phi_1 \rho_1}$$

AR(2) De forma semelhante para AR(2):

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}$$

A variância:

$$\gamma_0 = \sigma_\omega^2 = \frac{\sigma_a^2}{1 - \phi_1 \rho_1 - \phi_2 \rho_2}$$

AR(3)

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \phi_3 \rho_{k-3}$$
$$\gamma_0 = \sigma_\omega^2 = \frac{\sigma_a^2}{1 - \phi_1 \rho_1 - \phi_2 \rho_2 - \phi_3 \rho_3}$$

7) Calcule: as séries diferenciadas seguintes:

a) $\omega\nabla Z_t$

b) $\omega\nabla^2 Z_t$

A resolução das letras A e B podem ser visualizadas na tabela a seguir:

```
library(pacman)
pacman::p_load(tidyverse)

number_07 <- data.frame(t = c(1:15),
                        z = c(47,44,50,62,68,64,80,71,44,38,23,55,56,64,50)) %>%
             dplyr::mutate(mean = mean(z),
                           "diff_1" = c(NA, diff(z)),
                           "diff_2" = c(NA, NA, diff(z, differences = 2))) %>%
                    mutate(across(everything(), ~ifelse(is.na(.), "-", .)))

knitr::kable(number_07)
```

t	z	mean	diff_1	diff_2
1	47	54.4	-	-
2	44	54.4	-3	-
3	50	54.4	6	9
4	62	54.4	12	6
5	68	54.4	6	-6
6	64	54.4	-4	-10
7	80	54.4	16	20
8	71	54.4	-9	-25
9	44	54.4	-27	-18
10	38	54.4	-6	21
11	23	54.4	-15	-9
12	55	54.4	32	47
13	56	54.4	1	-31
14	64	54.4	8	7
15	50	54.4	-14	-22

c) A estimativa da autocovariância γ_k quando k = 1:

Autocovariância:

$$\hat{\gamma} k = \frac{1}{n}\sum_{t=1}^{n}(z_t - \bar{z})(z_{t+k} - \bar{z})$$

```
autocov = (47 - 54.44) * (44 - 54.44) + (44 - 54.44) * (50 - 54.44) +
          (50 - 54.44) * (62 - 54.44) + (62 - 54.44) * (68 - 54.44) +
          (68 - 54.44) * (64 - 54.44) + (64 - 54.44) * (80 - 54.44) +
          (80 - 54.44) * (71 - 54.44) + (71 - 54.44) * (44 - 54.44) +
          (44 - 54.44) * (38 - 54.44) + (38 - 54.44) * (23 - 54.44) +
          (23 - 54.44) * (55 - 54.44) + (55 - 54.44) * (56 - 54.44) +
          (56 - 54.44) * (64 - 54.44) + (64 - 54.44) * (50 - 54.44)
```

```
autocov = autocov/14
print(autocov)
```

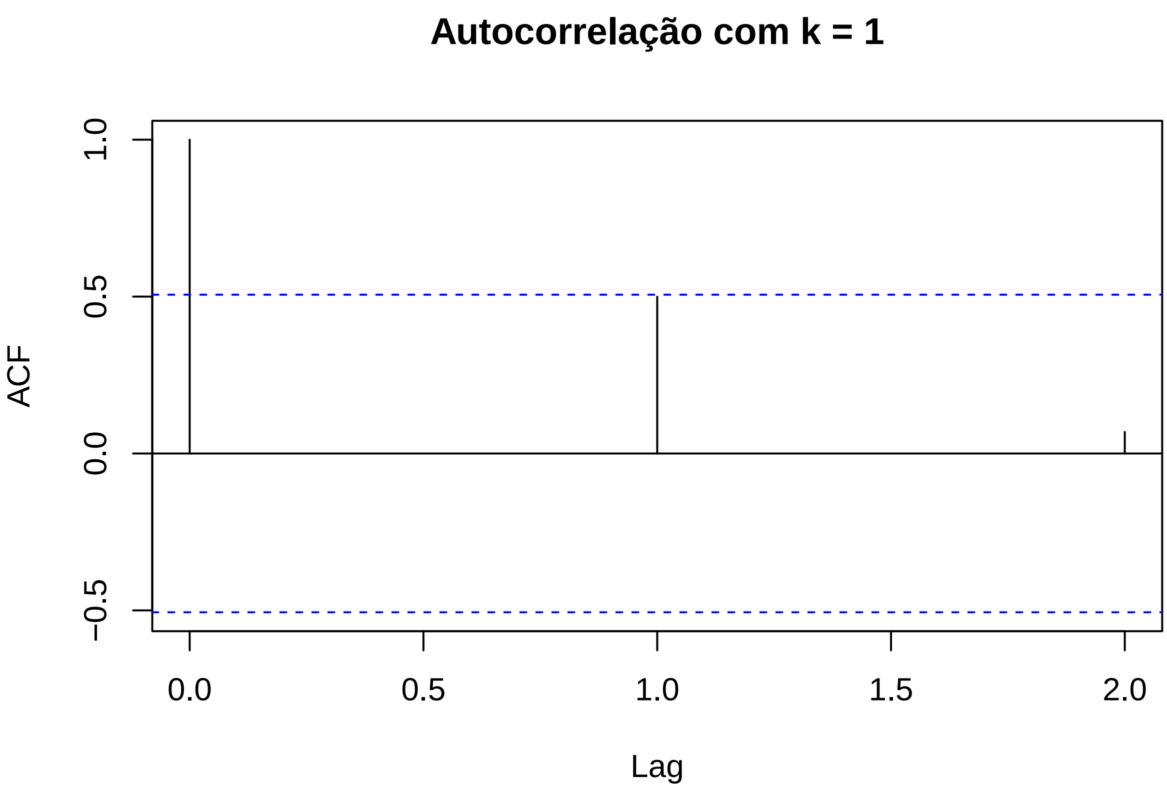
```
## [1] 104.3993
```

d) a estimativa da autocorrelação γ_k quando k = 1.

```
autocor = (autocov/var(number_07$z))
print(autocor)
```

```
## [1] 0.4995865
```

```
acf(number_07$z, lag.max = 2, main = "Autocorrelação com k = 1")
```



8) Para os dados da série A do livro B&J (final do livro) pede-se:

a) a série diferenciada uma vez ;

```
data_raw = readxl::read_excel("db_exercicio.xlsx") %>%
           janitor::clean_names()

data_trans <- data_raw %>%
              dplyr::mutate(diff = c(NA, diff(z)))

knitr::kable(data_trans)
```

t	z	diff
1	17.0	NA
2	16.6	-0.4
3	16.3	-0.3
4	16.1	-0.2
5	17.1	1.0
6	16.9	-0.2
7	16.8	-0.1
8	17.4	0.6
9	17.1	-0.3
10	17.0	-0.1
11	16.7	-0.3
12	17.4	0.7
13	17.2	-0.2
14	17.4	0.2
15	17.4	0.0
16	17.0	-0.4
17	17.3	0.3
18	17.2	-0.1
19	17.4	0.2
20	16.8	-0.6
21	17.1	0.3
22	17.4	0.3
23	17.4	0.0
24	17.5	0.1
25	17.4	-0.1
26	17.6	0.2
27	17.4	-0.2
28	17.3	-0.1
29	17.0	-0.3
30	17.8	0.8
31	17.5	-0.3
32	18.1	0.6
33	17.5	-0.6
34	17.4	-0.1
35	17.4	0.0
36	17.1	-0.3
37	17.6	0.5
38	17.7	0.1
39	17.4	-0.3
40	17.8	0.4
41	17.6	-0.2
42	17.5	-0.1
43	16.5	-1.0
44	17.8	1.3
45	17.3	-0.5
46	17.3	0.0
47	17.1	-0.2
48	17.4	0.3
49	16.9	-0.5
50	17.3	0.4
51	17.6	0.3
52	16.9	-0.7

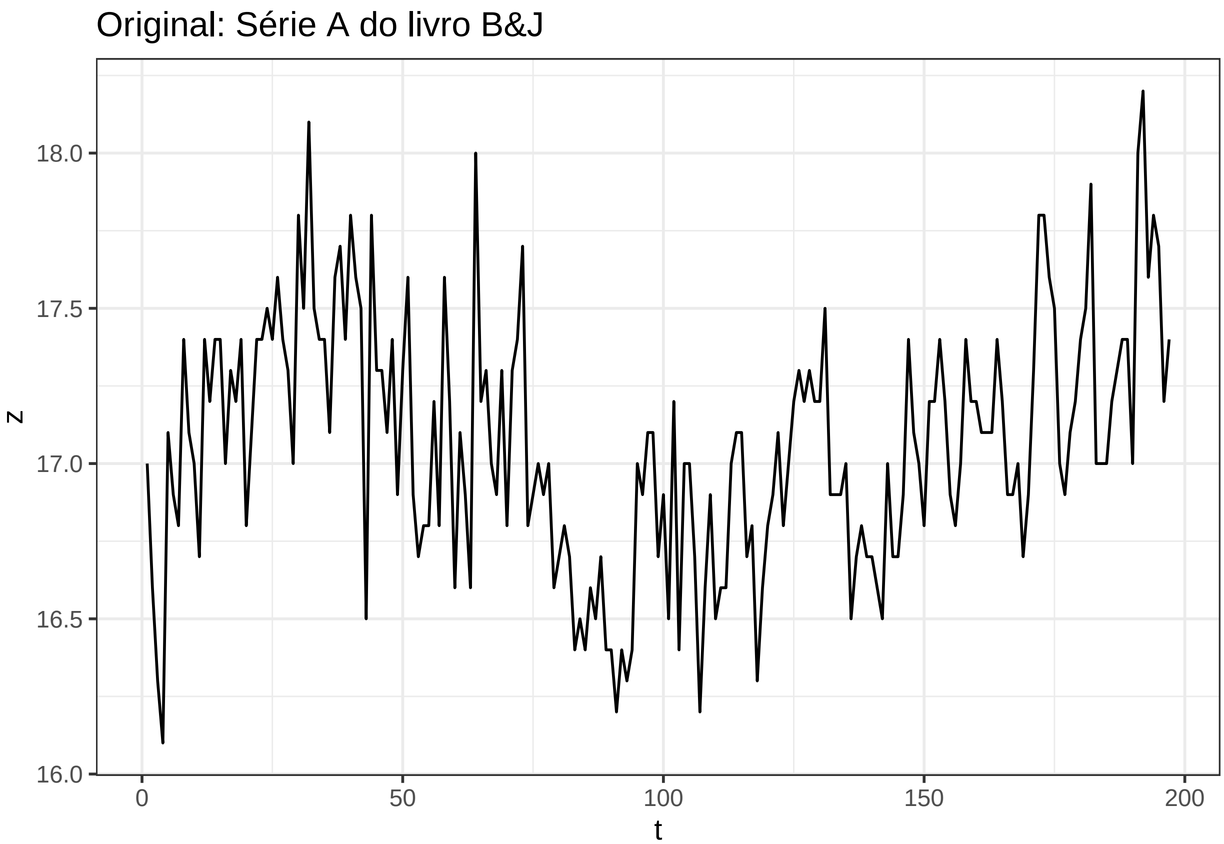
t	z	diff
53	16.7	-0.2
54	16.8	0.1
55	16.8	0.0
56	17.2	0.4
57	16.8	-0.4
58	17.6	0.8
59	17.2	-0.4
60	16.6	-0.6
61	17.1	0.5
62	16.9	-0.2
63	16.6	-0.3
64	18.0	1.4
65	17.2	-0.8
66	17.3	0.1
67	17.0	-0.3
68	16.9	-0.1
69	17.3	0.4
70	16.8	-0.5
71	17.3	0.5
72	17.4	0.1
73	17.7	0.3
74	16.8	-0.9
75	16.9	0.1
76	17.0	0.1
77	16.9	-0.1
78	17.0	0.1
79	16.6	-0.4
80	16.7	0.1
81	16.8	0.1
82	16.7	-0.1
83	16.4	-0.3
84	16.5	0.1
85	16.4	-0.1
86	16.6	0.2
87	16.5	-0.1
88	16.7	0.2
89	16.4	-0.3
90	16.4	0.0
91	16.2	-0.2
92	16.4	0.2
93	16.3	-0.1
94	16.4	0.1
95	17.0	0.6
96	16.9	-0.1
97	17.1	0.2
98	17.1	0.0
99	16.7	-0.4
100	16.9	0.2
101	16.5	-0.4
102	17.2	0.7
103	16.4	-0.8
104	17.0	0.6

t	z	diff
105	17.0	0.0
106	16.7	-0.3
107	16.2	-0.5
108	16.6	0.4
109	16.9	0.3
110	16.5	-0.4
111	16.6	0.1
112	16.6	0.0
113	17.0	0.4
114	17.1	0.1
115	17.1	0.0
116	16.7	-0.4
117	16.8	0.1
118	16.3	-0.5
119	16.6	0.3
120	16.8	0.2
121	16.9	0.1
122	17.1	0.2
123	16.8	-0.3
124	17.0	0.2
125	17.2	0.2
126	17.3	0.1
127	17.2	-0.1
128	17.3	0.1
129	17.2	-0.1
130	17.2	0.0
131	17.5	0.3
132	16.9	-0.6
133	16.9	0.0
134	16.9	0.0
135	17.0	0.1
136	16.5	-0.5
137	16.7	0.2
138	16.8	0.1
139	16.7	-0.1
140	16.7	0.0
141	16.6	-0.1
142	16.5	-0.1
143	17.0	0.5
144	16.7	-0.3
145	16.7	0.0
146	16.9	0.2
147	17.4	0.5
148	17.1	-0.3
149	17.0	-0.1
150	16.8	-0.2
151	17.2	0.4
152	17.2	0.0
153	17.4	0.2
154	17.2	-0.2
155	16.9	-0.3
156	16.8	-0.1

t	z	diff
157	17.0	0.2
158	17.4	0.4
159	17.2	-0.2
160	17.2	0.0
161	17.1	-0.1
162	17.1	0.0
163	17.1	0.0
164	17.4	0.3
165	17.2	-0.2
166	16.9	-0.3
167	16.9	0.0
168	17.0	0.1
169	16.7	-0.3
170	16.9	0.2
171	17.3	0.4
172	17.8	0.5
173	17.8	0.0
174	17.6	-0.2
175	17.5	-0.1
176	17.0	-0.5
177	16.9	-0.1
178	17.1	0.2
179	17.2	0.1
180	17.4	0.2
181	17.5	0.1
182	17.9	0.4
183	17.0	-0.9
184	17.0	0.0
185	17.0	0.0
186	17.2	0.2
187	17.3	0.1
188	17.4	0.1
189	17.4	0.0
190	17.0	-0.4
191	18.0	1.0
192	18.2	0.2
193	17.6	-0.6
194	17.8	0.2
195	17.7	-0.1
196	17.2	-0.5
197	17.4	0.2

b) faça o gráfico da série original e outro da série diferenciada uma vez:

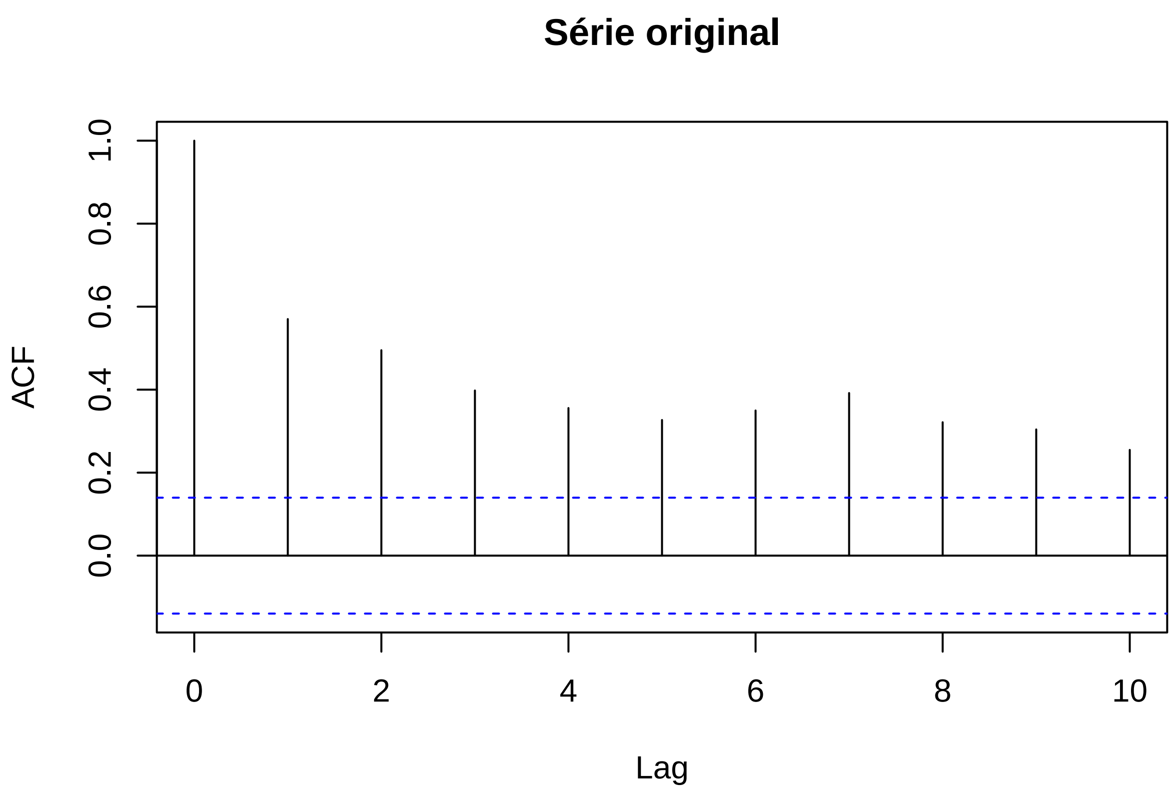
```
ggplot2::ggplot(data = data_trans)+
         aes(x = t, y = z)+
         geom_line()+
         labs(title = "Original: Série A do livro B&J")+
         theme_bw()
```



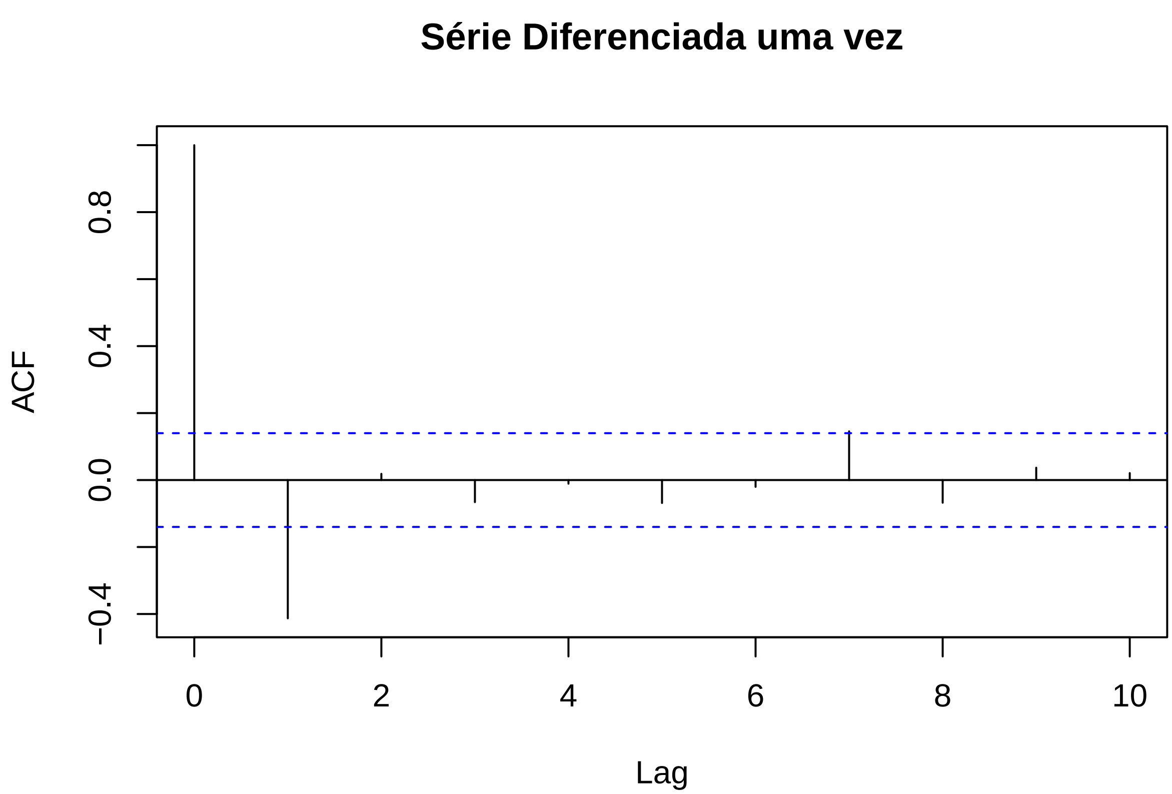

```
data_clean = data_trans %>% na.omit()

  ggplot2::ggplot(data = data_clean)+
         aes(x = t, y = diff)+
         geom_line()+
         labs(title = "Diferenciada uma vez: Série A do livro B&J")+
         theme_bw()
```

c) calcule a FAC da série original e a FAC da série diferenciada (bastam 10 lags);

```
acf(data_trans$z,
    lag.max = 10, main = "Série original")
```




```r
acf(data_clean$diff,
    lag.max = 10, main = "Série Diferenciada uma vez")
```



9) Dada a série temporal a seguir pede-se:

a) a série diferenciada uma vez (d=1);

```
number_09 = data.frame(t = 1:5,
                       z = c(12,14,16,14,18)) %>%
            dplyr::mutate(diff = c(NA, diff(z)))

knitr::kable(number_09)
```

t	z	diff
1	12	NA
2	14	2
3	16	2
4	14	-2
5	18	4

b) A média e a variância

```
cat("A média de Zt é:", mean(number_09$z))
```

```
## A média de Zt é: 14.8
```

```
cat("A Variância de Zt é:", var(number_09$z))
```

```
## A Variância de Zt é: 5.2
```

Manualmente, podemos calcular como segue:

$$\bar{Z} = \frac{Z_1 + Z_2 + Z_3 + ... + Z_n}{n} = \frac{\sum_{i=1}^{n} Z_i}{n}$$

```
média = (12 + 14 + 16 + 14 + 18)/5
print(média)
```

```
## [1] 14.8
```

Variância:

$$\hat{\sigma_z}^2 = \hat{\gamma_0} = \frac{1}{n}\sum_{t=1}^{n}(z_t - \bar{z})^2$$

```
var = (12 - 14.8)^2 + (14 - 14.8)^2 + (16 - 14.8)^2+ (14 - 14.8)^2 + (18 - 14.8)^2

var = var/5
print(var)
```

```
## [1] 4.16
```

c) o valor da autocovariância de defasagem 1 (lag) k = 1 da série original

$$\hat{\gamma}k = \frac{1}{n}\sum_{t=1}^{n}(z_t - \bar{z})(z_{t+k} - \bar{z})$$

```
auto = (12 - 14.8) * (14 - 14.8) + (14 - 14.8) * (16 - 14.8)+
       (16 - 14.8) * (14 - 14.8) + (14 - 14.8) * (18 - 14.8)

auto = auto/5
print(auto)
```

```
## [1] -0.448
```

d) o valor da autocorrelação de defasagem (lag) k=1 da série original

Autocorrelação:

$$\hat{\rho}k = \frac{\frac{1}{n}\sum_{t=1}^{n}(z_t - \bar{z})(z_{t+k} - \bar{z})}{\sum_{t=1}^{n}(z_t - \bar{z})^2} = \frac{\gamma k}{\gamma 0}$$

É apenas dividir:

```
autocorrelacao = (auto/var)
print(autocorrelacao)
```

```
## [1] -0.1076923
```

10) Escreva as equações de Yule-Walker na forma matricial com as estimativas dos coeficientes de autocorrelação

Primeiro, considere uma série temporal estacionária X_t e a função de autocorrelação $\rho(k)$, onde k representa o número de defasagens (lags). A autocorrelação para uma defasagem k é dada por:

$$\rho(k) = \frac{\text{Cov}(Z_t, Z_{t-k})}{\text{Var}(Z_t)}$$

onde $\text{Cov}(X_t, X_{t-k})$ é a covariância entre X_t e X_{t-k}, e $\text{Var}(X_t)$ é a variância de X_t.

Modelo AR(p)

Um modelo autoregressivo de ordem p, denotado por AR(p), é dado por:

$$X_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_p Z_{t-p} + \epsilon_t$$

onde $\phi_1, \phi_2, \ldots, \phi_p$ são os coeficientes autoregressivos, e ϵ_t é um termo de erro branco com média zero e variância constante.

Equações de Yule-Walker

As equações de Yule-Walker são um conjunto de p equações que relacionam as autocorrelações $\rho(k)$ com os coeficientes autoregressivos ϕ_k. Essas equações são dadas por:

$$\begin{aligned}
\rho(1) &= \phi_1 \rho(0) + \phi_2 \rho(1) + \cdots + \phi_p \rho(p-1) \\
\rho(2) &= \phi_1 \rho(1) + \phi_2 \rho(0) + \cdots + \phi_p \rho(p-2) \\
&\vdots \\
\rho(p) &= \phi_1 \rho(p-1) + \phi_2 \rho(p-2) + \cdots + \phi_p \rho(0)
\end{aligned}$$

Forma Matricial das Equações de Yule-Walker

Essas equações podem ser escritas na forma matricial:

$$\begin{pmatrix} \rho(0) & \rho(1) & \dots & \rho(p-1) \\ \rho(1) & \rho(0) & \dots & \rho(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(p-1) & \rho(p-2) & \dots & \rho(0) \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(p) \end{pmatrix}$$

Solução das Equações

Para encontrar os coeficientes $\boldsymbol{\phi} = (\phi_1, \phi_2, \dots, \phi_p)^T$, resolvemos o sistema linear:

$$\boldsymbol{\phi} = \mathbf{R}^{-1}\mathbf{r}$$

onde $\mathbf{R}^{-1}$ é a inversa da matriz de autocorrelações $\mathbf{R}$.

Exemplo para AR(2)

Para um modelo AR(2), as equações de Yule-Walker se tornam:

$$\begin{pmatrix} \rho(0) & \rho(1) \\ \rho(1) & \rho(0) \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \rho(2) \end{pmatrix}$$