Lista 1 - Séries Temporais

Table of contents

0.1 Questão 1

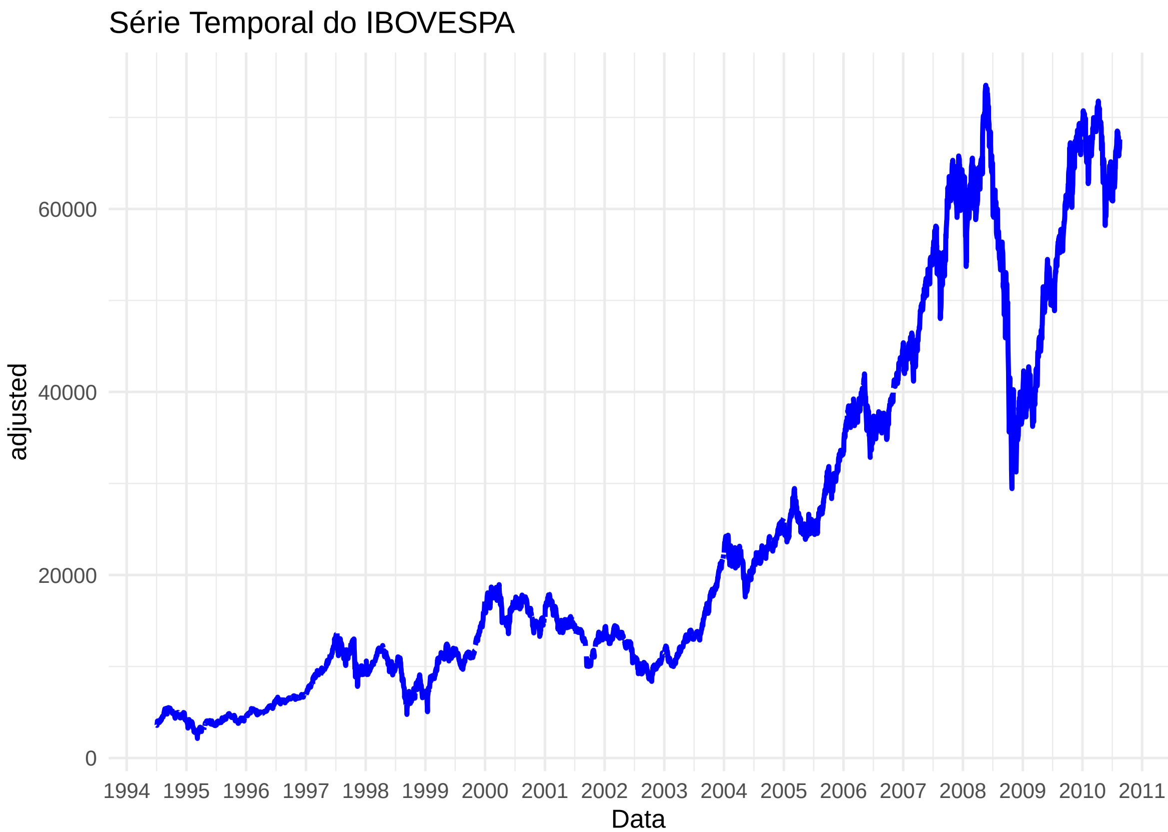
Considere os log retornos diários do IBOVESPA de 4/07/1994 a 19/08/2010:

```
data <- tq_get(
  "^BVSP",
  from = "1994-07-04",
  to = "2010-08-19",
  get = "stock.prices")

data <- data %>%
  select(-symbol) %>%
  mutate(date = as.Date(date, format = "%Y-%m-%d"))
```

- Faça um gráfico da série e da série dos log-retornos, calcule as estatísticas de média, mediana, variância, assimetria e curtose, e comente.

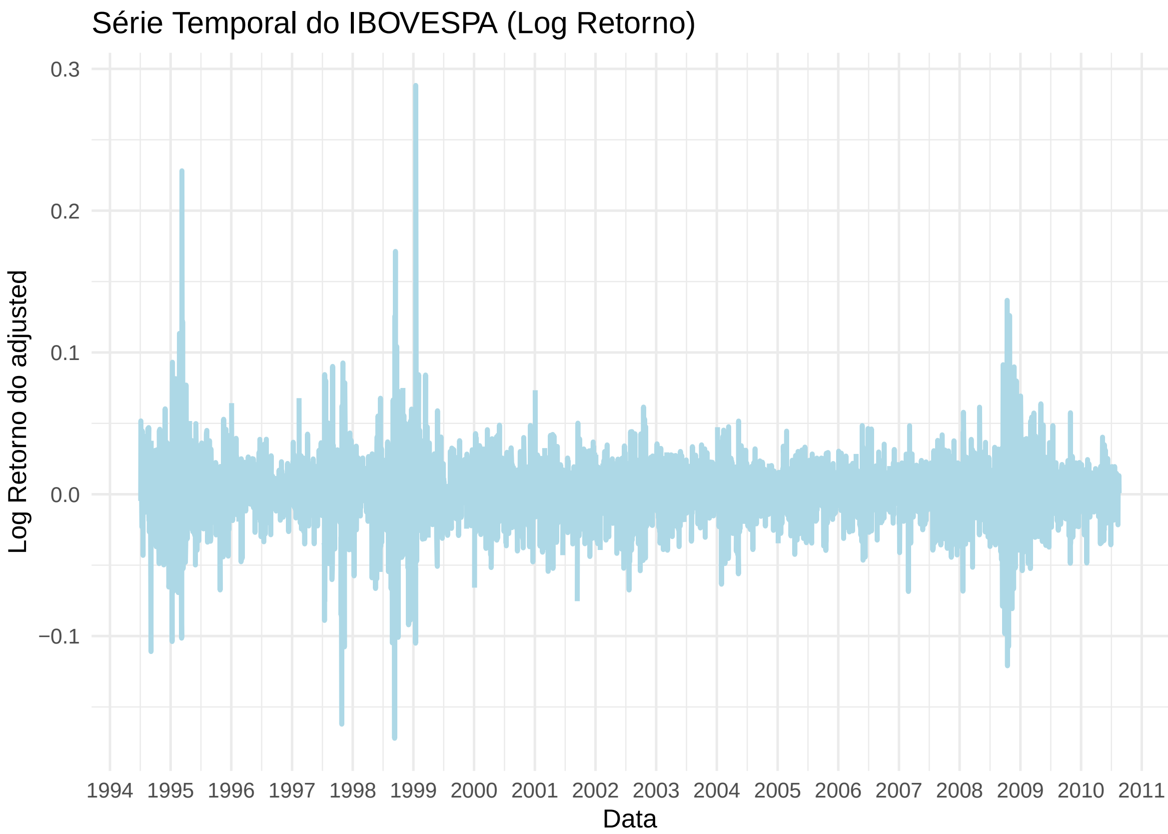
```
# gráfico da série temporal

ggplot(data, aes(x = date, y = adjusted)) +
  geom_line(color = "blue", size = 1) +
```

```r
  labs(
    title = "Série Temporal do IBOVESPA",
    x = "Data",
    y = "adjusted"
  ) +
  theme_minimal() +
  scale_x_date(date_labels = "%Y", date_breaks = "1 year")
```




```r
# calculando log retorno diário

data <- data %>%
  arrange(date) %>%
  mutate(log_return = log(adjusted / lag(adjusted)))

# gráfico do log retorno
```

```r
ggplot(data, aes(x = date, y = log_return)) +
  geom_line(color = "lightblue", size = 1) +
  labs(
    title = "Série Temporal do IBOVESPA (Log Retorno)",
    x = "Data",
    y = "Log Retorno do adjusted"
  ) +
  theme_minimal() +
  scale_x_date(date_labels = "%Y", date_breaks = "1 year")
```




```r
stats <- data.frame(
  Estatística = c("Média", "Mediana", "Variância", "Assimetria", "Curtose"),
  Valor = c(
    mean(data$adjusted, na.rm = TRUE),
    median(data$adjusted, na.rm = TRUE),
    var(data$adjusted, na.rm = TRUE),
```

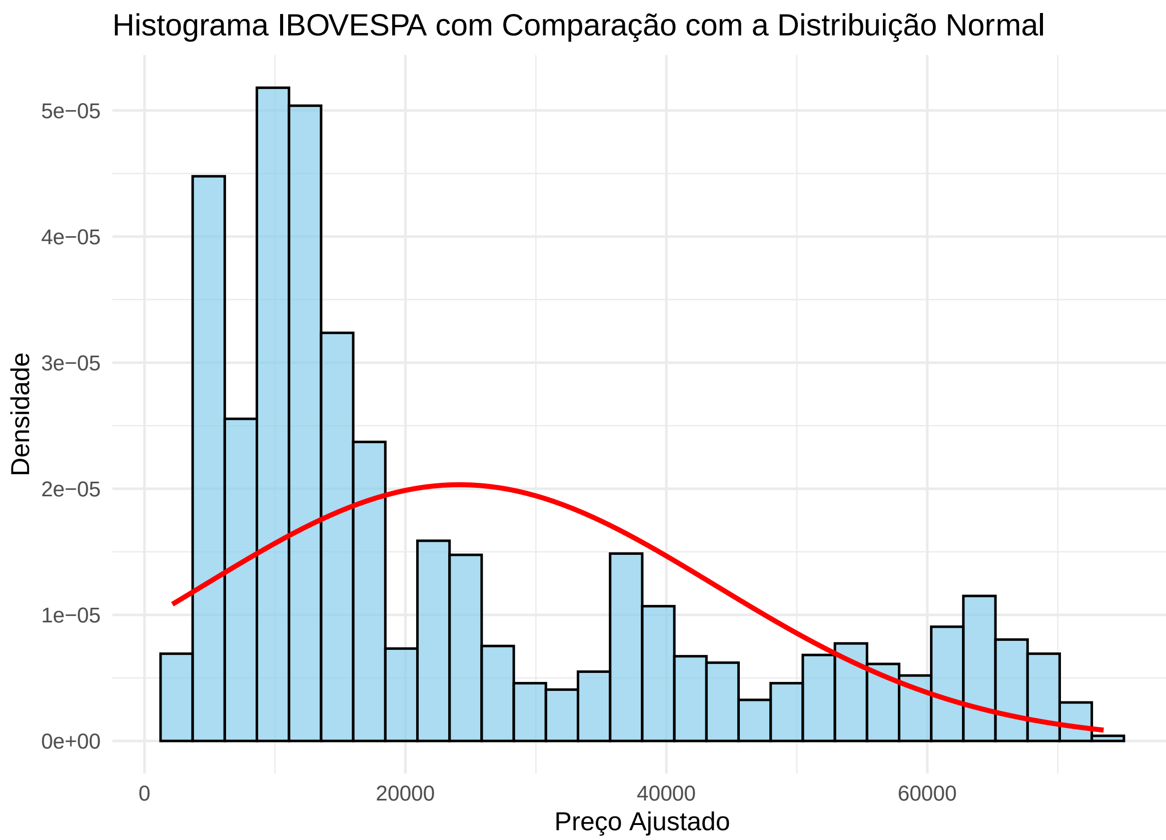
```
    skewness(data$adjusted, na.rm = TRUE),
    kurtosis(data$adjusted, na.rm = TRUE)
  )
)

kable(stats,
      caption = "Estatísticas descritivas do IBOVESPA",
      align = "c") %>%
  kable_styling(bootstrap_options = c("striped"), full_width = FALSE)
```

Table 1: Estatísticas descritivas do IBOVESPA

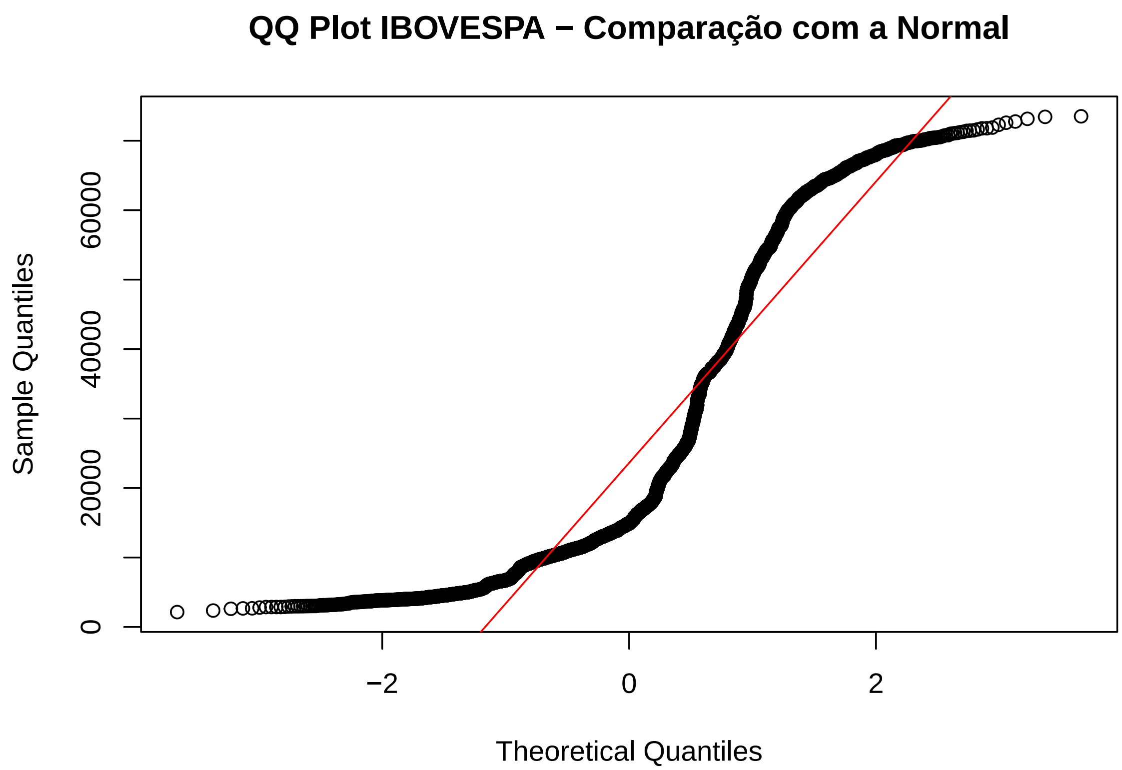
Estatística	Valor
Média	2.414843e+04
Mediana	1.491200e+04
Variância	3.853972e+08
Assimetria	9.963455e-01
Curtose	-3.185480e-01

- Obtenha um histograma dos dados (série) e comente sobre a forma da distribuição, comparando com a distribuição Normal de mesma média e variância. Faça o QQ plot e comente.

```
#histograma

ggplot(data, aes(x = adjusted)) +
  geom_histogram(aes(y = after_stat(density)), bins = 30,
                 fill = "skyblue", color = "black", alpha = 0.7) +
  stat_function(fun = dnorm, args = list(mean = mean(data$adjusted, na.rm = TRUE),
                                        sd = sd(data$adjusted, na.rm = TRUE)),
                color = "red", size = 1) +
  labs(title = "Histograma IBOVESPA com Comparação com a Distribuição Normal",
       x = "Preço Ajustado",
       y = "Densidade") +
  theme_minimal()
```




```r
# qqplots

qqnorm(data$adjusted, main = "QQ Plot IBOVESPA - Comparação com a Normal")
qqline(data$adjusted, col = "red")
```



Observa-se uma assimetria positiva, com uma cauda longa à direita. Isso é esperado, pois os preços financeiros geralmente não seguem uma distribuição normal. A curva em vermelho, representando uma distribuição normal ajustada, não modela bem os dados devido à assimetria e ao comportamento de cauda.

- Comente o significado da media e teste se a serie é ruido branco ou não.

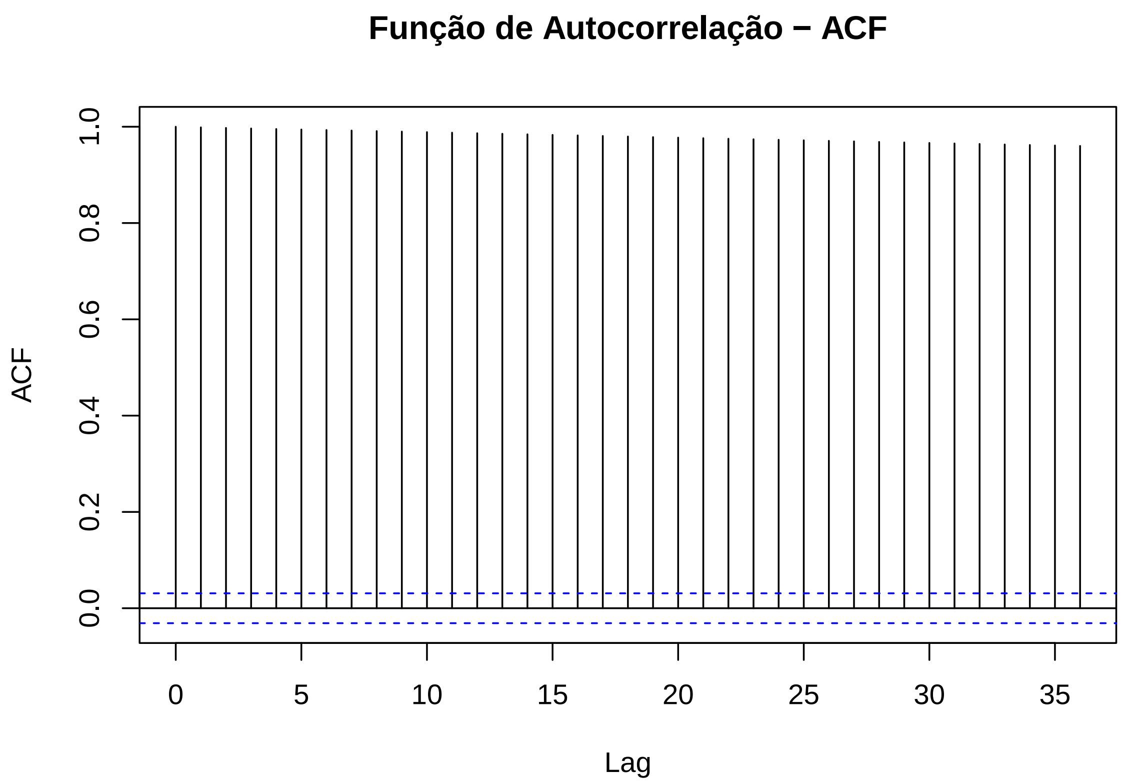
```
stats[1,]
```

```
  Estatística    Valor
1       Média 24148.43
```

Esta medida central pode ser interpretada como o comportamente típico ao longo do tempo.

```
# gráfico da função de autocorrelação

adjusted <- na.omit(data$adjusted)
```

```r
acf(adjusted, main = "Função de Autocorrelação - ACF")
```




```r
# teste de Ljung-box

Box.test(adjusted, lag = 10, type = "Ljung-Box")
```

```
	Box-Ljung test

data:  adjusted
X-squared = 39497, df = 10, p-value < 2.2e-16
```

Observando o gráfico da função de autocorrelaçaõ há um forte indicativo da série não ser um ruido branco, aplicando o teste de Ljung-Box confirma-se que de fato não é o caso.

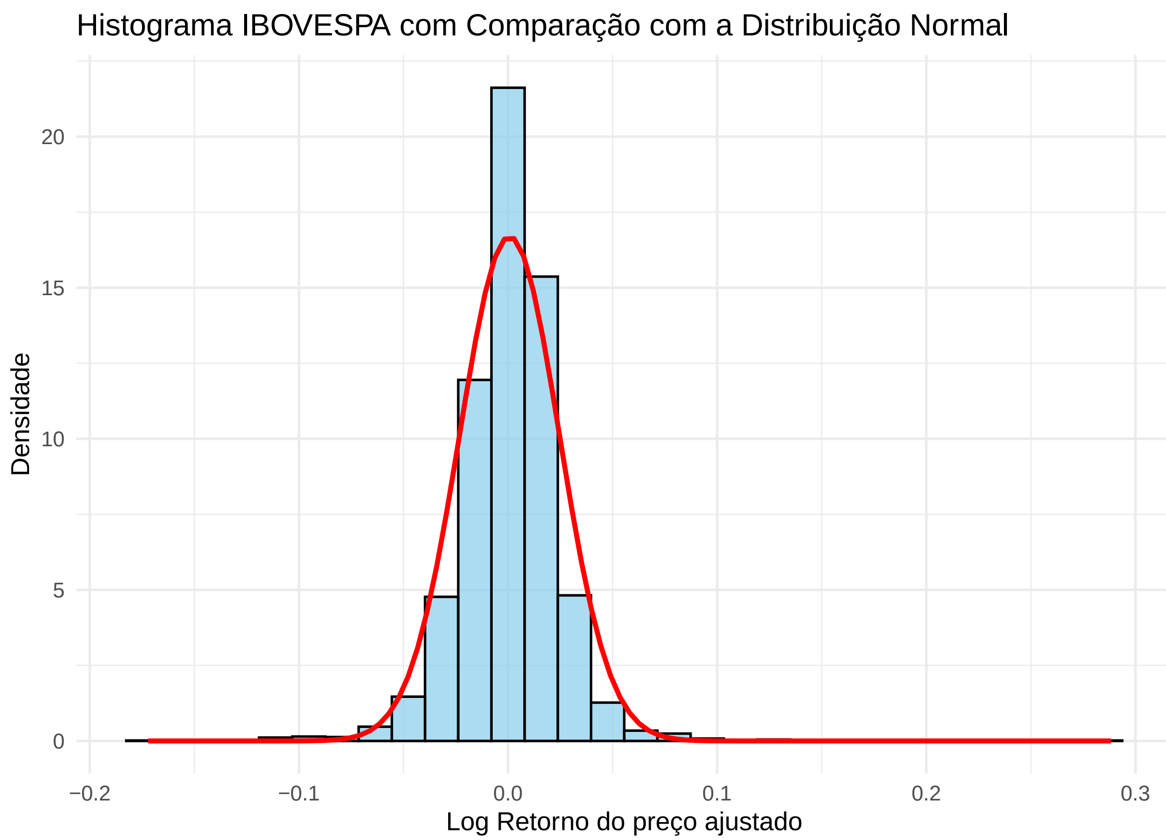
- Faz o item (b)-(c) para os log-retornos.

```r
# medidas descritivas para log retorno

stats_logreturn <- data.frame(
  Mean = mean(data$log_return, na.rm = TRUE),
  Median = median(data$log_return, na.rm = TRUE),
  Variance = var(data$log_return, na.rm = TRUE),
  Skewness = skewness(data$log_return, na.rm = TRUE),
  Kurtosis = kurtosis(data$log_return, na.rm = TRUE)
)

kable(stats_logreturn,
      caption = "Estatísticas descritivas do Log Retorno diário do IBOVESPA",
      align = "c") %>%
  kable_styling(bootstrap_options = c("striped"), full_width = FALSE)
```

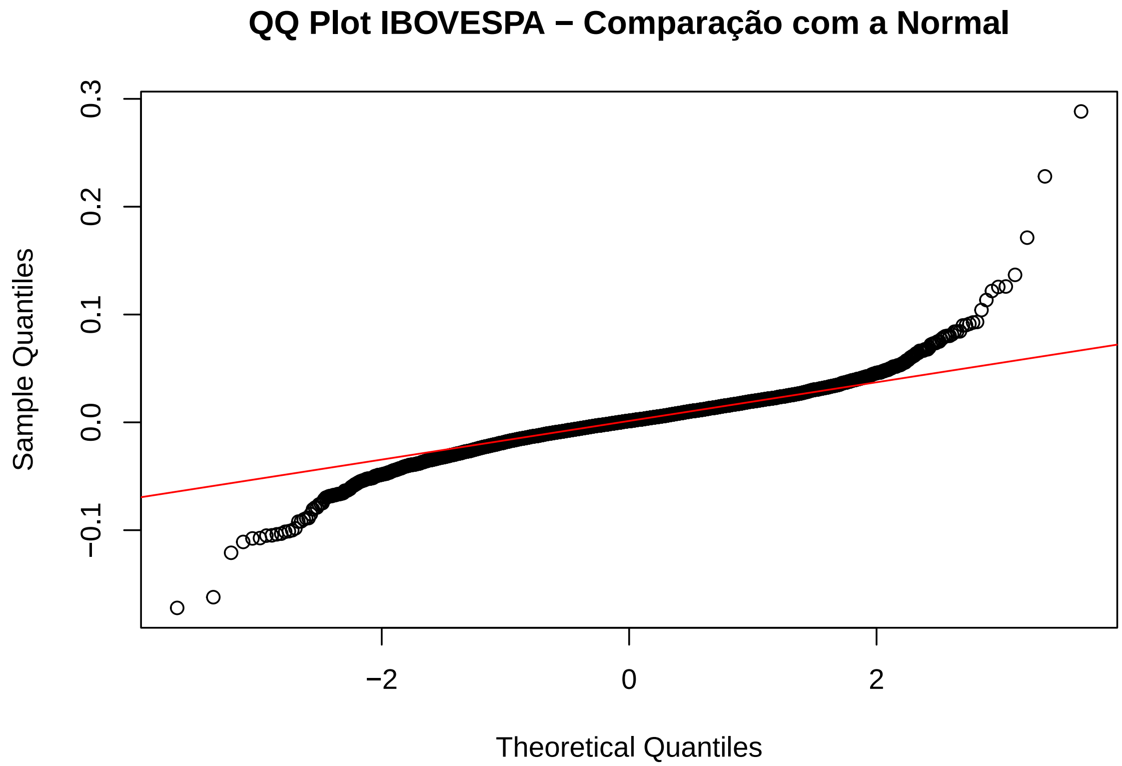
Table 2: Estatísticas descritivas do Log Retorno diário do IBOVESPA

Mean	Median	Variance	Skewness	Kurtosis
0.0007069	0.0013432	0.0005709	0.4509458	11.85277

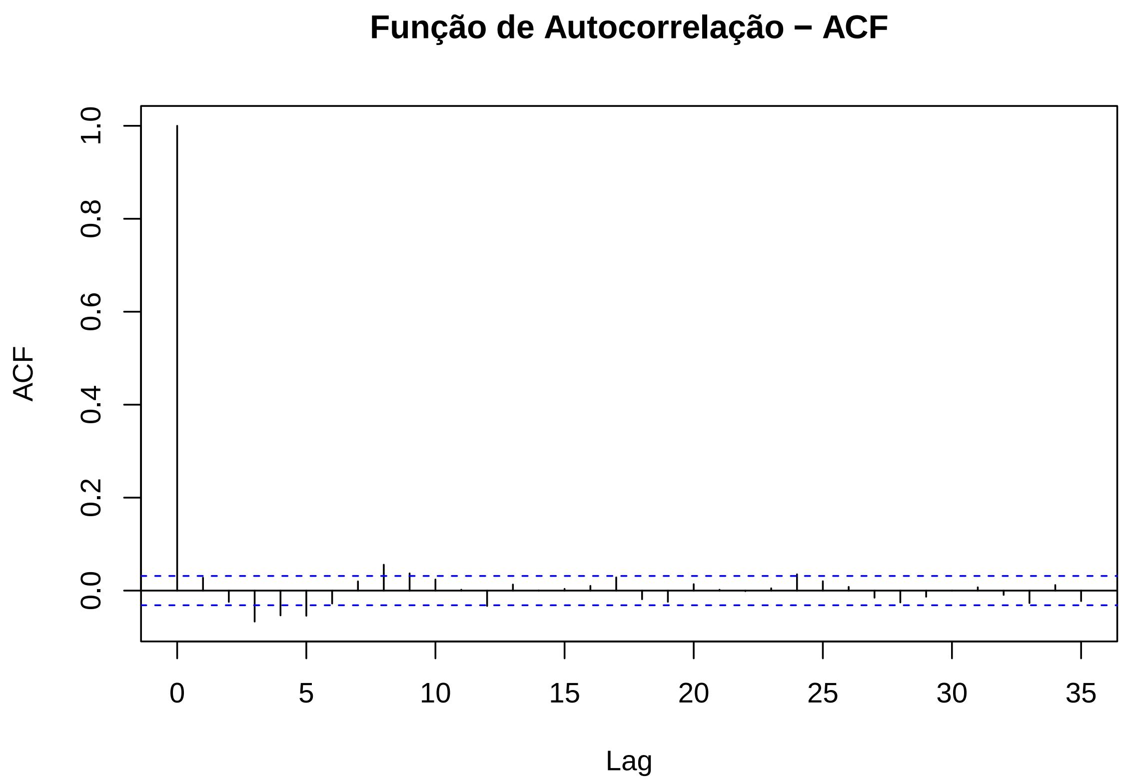
```r
# histograma para log retorno

ggplot(data, aes(x = log_return)) +
  geom_histogram(aes(y = after_stat(density)), bins = 30,
                 fill = "skyblue", color = "black", alpha = 0.7) +
  stat_function(fun = dnorm, args = list(mean = mean(data$log_return, na.rm = TRUE),
                                         sd = sd(data$log_return, na.rm = TRUE)),
                color = "red", size = 1) +
  labs(title = "Histograma IBOVESPA com Comparação com a Distribuição Normal",
       x = "Log Retorno do preço ajustado",
       y = "Densidade") +
  theme_minimal()
```




```
# qqplots log retorno

qqnorm(data$log_return, main = "QQ Plot IBOVESPA - Comparação com a Normal")
qqline(data$log_return, col = "red")
```




```r
# testando ruido branco

log_return <- na.omit(data$log_return)

acf(log_return, main = "Função de Autocorrelação - ACF")
```




```
Box.test(log_return, lag = 10, type = "Ljung-Box")
```

```
	Box-Ljung test

data:  log_return
X-squared = 69.061, df = 10, p-value = 6.729e-11
```

No gráfico de histograma do log retorno do preço ajustado há indicios de que os dados aderem a normalidade após a transformação. De forma similar, observa-se que o qqplot também tem um comportamento normal.

Observando o gráfico da função de autocorrelaçaõ, indica a possibilidade da série ser um ruido branco, mas aplicando o teste de Ljung-Box essa hipótese é rejetada. O que indica que apesar de ter autocorrelação baixa, não é igual a zero.

0.2 Questão 2

Para os dados da série log-retornos mensais do SP500 de Junho/1994 a Agosto /2001:

```
data <- tq_get("^GSPC",
               from = "1994-06-01",
               to = "2001-08-01",
               get = "stock.prices")

data <- data %>%
  select(-symbol) %>%
  mutate(date = as.Date(date, format = "%Y-%m-%d"))
```

- Qual é o retorno medio anual sobre o periodo dos dados?

```
n_anos <- as.numeric(difftime(max(data$date), min(data$date), units = "days")) / 365

valor_inicial <- first(data$adjusted)
valor_final <- last(data$adjusted)

retorno_medio_anual <- (valor_final / valor_inicial)^(1 / n_anos) - 1

retorno_medio_anual
```

```
[1] 0.145399
```

- Qual é o retorno simples médio anualizado?

```
data <- data %>%
  arrange(date) %>%
  mutate(retorno_diario = (adjusted / lag(adjusted)) - 1)

data <- data %>% filter(!is.na(retorno_diario))

media_retorno_diario <- mean(data$retorno_diario)

retorno_simples_anualizado <- media_retorno_diario * 252

retorno_simples_anualizado
```

```
[1] 0.1502934
```

- Se vc investir R/. 1.00 no ativo no final de Dezembro de 1994, qual é o valor do investimento ao final do Dezembro de 2001?

```
log_retornos_periodo <- data %>%
  filter(date >= as.Date("1994-12-31") & date <= as.Date("2001-08-31")) %>%
  pull(retorno_diario)

# Calcular log-retorno total
log_retorno_total <- sum(log_retornos_periodo)

# Converter para retorno acumulado
retorno_acumulado <- exp(log_retorno_total)

# Calcular valor final do investimento
valor_inicial <- 1  # Investimento inicial
valor_final <- valor_inicial * retorno_acumulado

valor_final
```

```
[1] 2.923327
```

0.3 Questão 3

Considere os logretornos diarios da PETR4 de 03/01 a 27/12/2000

```
data <- tq_get("^GSPC",
               from = "2000-03-01",
               to = "2001-12-27",
               get = "stock.prices")

data <- data %>%
  select(-symbol) %>%
  mutate(date = as.Date(date, format = "%Y-%m-%d"))

data <- data %>%
  arrange(date) %>%
  mutate(
    retorno_simples = (adjusted / lag(adjusted)) - 1,
    log_retorno = log(adjusted / lag(adjusted))
  )
```

- Qual é o retorno simples do dia 1 para o dia 2? E do primeiro para o 7mo?

```
retorno_dia1_para_dia2 <- data$retorno_simples[2]
retorno_dia1_para_dia7 <- (data$adjusted[7] / data$adjusted[1]) - 1

retorno_dia1_para_dia2;retorno_dia1_para_dia7
```

```
[1] 0.001863462
```

```
[1] 0.01631392
```

- Quanto é o log retorno do dia 5 para o dia 6? E do 5 dia para o 10mo?

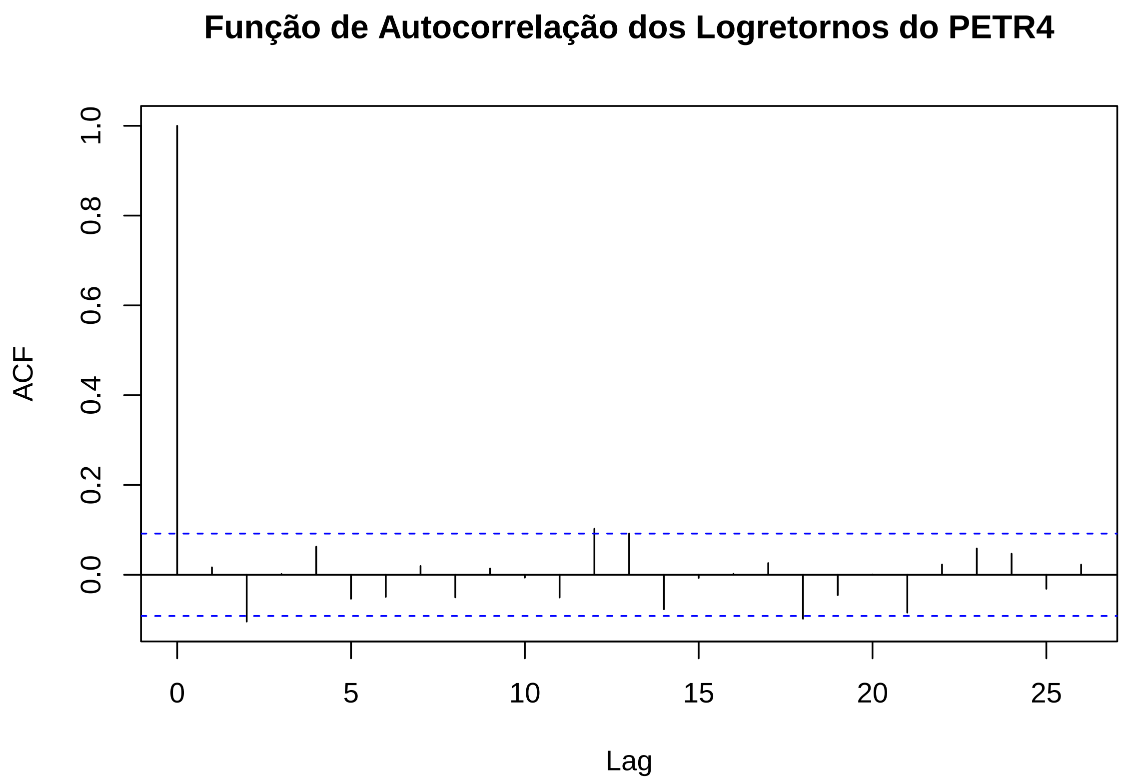
```
logretorno_dia5_para_dia6 <- log(data$adjusted[6] / data$adjusted[5])
logretorno_dia5_para_dia10 <- log(data$adjusted[10] / data$adjusted[5])

logretorno_dia5_para_dia6; logretorno_dia5_para_dia10
```

```
[1] 0.008140129
```

```
[1] 0.002600612
```

- Calcule a função de autocorrelação estimada e comente.

```
logretornos <- na.omit(data$log_retorno)
acf_logretornos <- acf(logretornos,
                       main = "Função de Autocorrelação dos Logretornos do PETR4")
```




```r
acf_logretornos
```

```
Autocorrelations of series 'logretornos', by lag

     0      1      2      3      4      5      6      7      8      9     10 
 1.000  0.017 -0.104  0.002  0.063 -0.053 -0.049  0.020 -0.050  0.014 -0.006 
    11     12     13     14     15     16     17     18     19     20     21 
-0.051  0.103  0.092 -0.077 -0.007  0.002  0.026 -0.098 -0.045  0.001 -0.084 
    22     23     24     25     26 
 0.023  0.059  0.047 -0.031  0.023 
```

0.4 Questão 4

- Quetão 21

A função $\gamma(\tau) = \sin(\tau)$ é uma possível função de autocovariância. Verifique sua resposta.

A função $\gamma(\tau)$ deve satisfazer as seguintes propriedades:

1. $\gamma(0) > 0$
2. $\gamma(-\tau) = \gamma(\tau)$
3. $|\gamma(\tau)| \leq \gamma(0)$
4. $\gamma(\tau)$ é não negativa definida, no sentido de que:

$$\sum_{j=1}^{n}\sum_{k=1}^{n} a_j a_k \gamma(\tau_j - \tau_k) \geq 0$$

para quaisquer números reais $a_1, a_2, \ldots, a_n$ e $\tau_1, \ldots, \tau_n \in \mathbb{R}$.

É necessário verificar se $\gamma(\tau) = \sin(\tau)$ atende essas propriedades.

Verificando as propriedades para $\gamma(\tau) = \sin(\tau)$:

(i) Verificação de $\gamma(0) > 0$:

$$\gamma(0) = \sin(0) = 0,$$

logo, temos que $\gamma(0) = 0$, o que não é maior que zero. Portanto, a função não satisfaz a condição de ser não negativa em $\tau = 0$.

(ii) Verificação da simetria de $\gamma(\tau)$:

$$\gamma(-\tau) = \sin(-\tau) = -\sin(\tau),$$

logo, $\gamma(-\tau) \neq \gamma(\tau)$. Portanto, a função não é simétrica.

(iii) Verificação de $|\gamma(\tau)| \leq \gamma(0)$:

Para $\tau = \pi/2$ temos:

$$\gamma\left(\frac{\pi}{2}\right) = \sin\left(\frac{\pi}{2}\right) = 1, \quad \gamma(0) = \sin(0) = 0.$$

Logo, temos que $|\gamma(\tau)| = 1 \nleq 0$. Portanto, a função não satisfaz a condição de ser limitada por $\gamma(0)$.

(iv) Verificação de não negatividade definida:

A condição de não negatividade definida exige que a soma

$$\sum_{j=1}^{n}\sum_{k=1}^{n} a_j a_k \gamma(\tau_j - \tau_k) \geq 0$$

para qualquer escolha de números reais $a_1, a_2, \ldots, a_n$ e $\tau_1, \ldots, \tau_n$. No entanto, dado que $\gamma(\tau) = \sin(\tau)$ não é uma função simétrica e é oscilante, essa soma pode assumir valores negativos. Portanto, não podemos garantir que a função seja não negativa definida.

- Questão 25

Suponha que $\{X_t, t = 0, \pm 1, ...\}$ seja uma sequência de variáveis aleatórias independentes, todas com a mesma distribuição, com $E\{X_t\} = \mu, \quad \forall t, \quad \text{e} \quad \text{Var}\{X_t\} = \sigma^2, \quad \forall t.$

Considere o processo $\{Y_t, t = 0, \pm 1, ...\}$, onde: $Y_t = \frac{1}{2}X_t + \frac{1}{4}X_{t-1} + \frac{1}{8}X_{t-2}$.

O processo $\{Y_t\}$ é estacionário?

Cálculos do Processo $\{Y_t\}$

Calculando a esperança:

$$E[Y_t] = E\left(\frac{1}{2}X_t + \frac{1}{4}X_{t-1} + \frac{1}{8}X_{t-2}\right) = \frac{1}{2}E[X_t] + \frac{1}{4}E[X_{t-1}] + \frac{1}{8}E[X_{t-2}]$$

Como $E[X_t] = \mu$ para todos os t, temos:

$$E[Y_t] = \frac{1}{2}\mu + \frac{1}{4}\mu + \frac{1}{8}\mu = \left(\frac{1}{2} + \frac{1}{4} + \frac{1}{8}\right)\mu = \frac{7}{8}\mu.$$

Calculando a variância:

$$\text{Var}(Y_t) = \text{Var}\left(\frac{1}{2}X_t + \frac{1}{4}X_{t-1} + \frac{1}{8}X_{t-2}\right)$$

Utilizando a independência das variáveis X_t, temos:

$$\text{Var}(Y_t) = \left(\frac{1}{2}\right)^2 \text{Var}(X_t) + \left(\frac{1}{4}\right)^2 \text{Var}(X_{t-1}) + \left(\frac{1}{8}\right)^2 \text{Var}(X_{t-2})$$

Como $\text{Var}(X_t) = \sigma^2$, temos:

$$\text{Var}(Y_t) = \frac{1}{4}\sigma^2 + \frac{1}{16}\sigma^2 + \frac{1}{64}\sigma^2 = \left(\frac{1}{4} + \frac{1}{16} + \frac{1}{64}\right)\sigma^2 = \frac{21}{64}\sigma^2.$$

Calculando a covariância:

$$\text{Cov}(Y_t, Y_{t+s}) = E[Y_t Y_{t+s}] - E[Y_t]E[Y_{t+s}]$$

Para $s = 0$:

$$\text{Cov}(Y_t, Y_t) = \text{Var}(Y_t)$$

Para $s = 1$:

$$\text{Cov}(Y_t, Y_{t+1}) = \frac{5}{32}\sigma^2$$

Para $s = 2$:

$$\text{Cov}(Y_t, Y_{t+2}) = \frac{1}{16}\sigma^2$$

Para $s \geq 3$:

$$\text{Cov}(Y_t, Y_{t+s}) = 0$$

O processo é estacionário.

0.5 Questão 5

Use um programa computacional para calcular:

```r
# baixando dados

data1 <- read_csv("data/ENERGIA.csv")

data2 <- read_csv("data/OZONIO.csv")

# tratando dados

data1 <- data1 %>%
  mutate(
    Ano = as.numeric(as.character(Ano)),
    Mes = as.numeric(as.character(Mes)),
    date = as.Date(paste(Ano, Mes, "01", sep = "-"), format = "%Y-%m-%d"),
    energia = Energia
  ) %>%
  select(date, energia)

data2 <- data2 %>%
  mutate(
    Ano = as.numeric(as.character(Ano)),
    Mes = as.numeric(as.character(Mes)),
    date = as.Date(paste(Ano, Mes, "01", sep = "-"), format = "%Y-%m-%d"),
    ozonio = Ozonio
  ) %>%
  select(date, ozonio)

# tansformando em formato time-series

data1_ts <- ts(data1$energia,  start = c(1968, 1), frequency = 12)

data2_ts <- ts(data2$ozonio, start = c(1956, 1), frequency = 12)
```

- média amostral

```r
media_amostral1 <- mean(data1$energia)

media_amostral2 <- mean(data2$ozonio)

media_amostral1
```

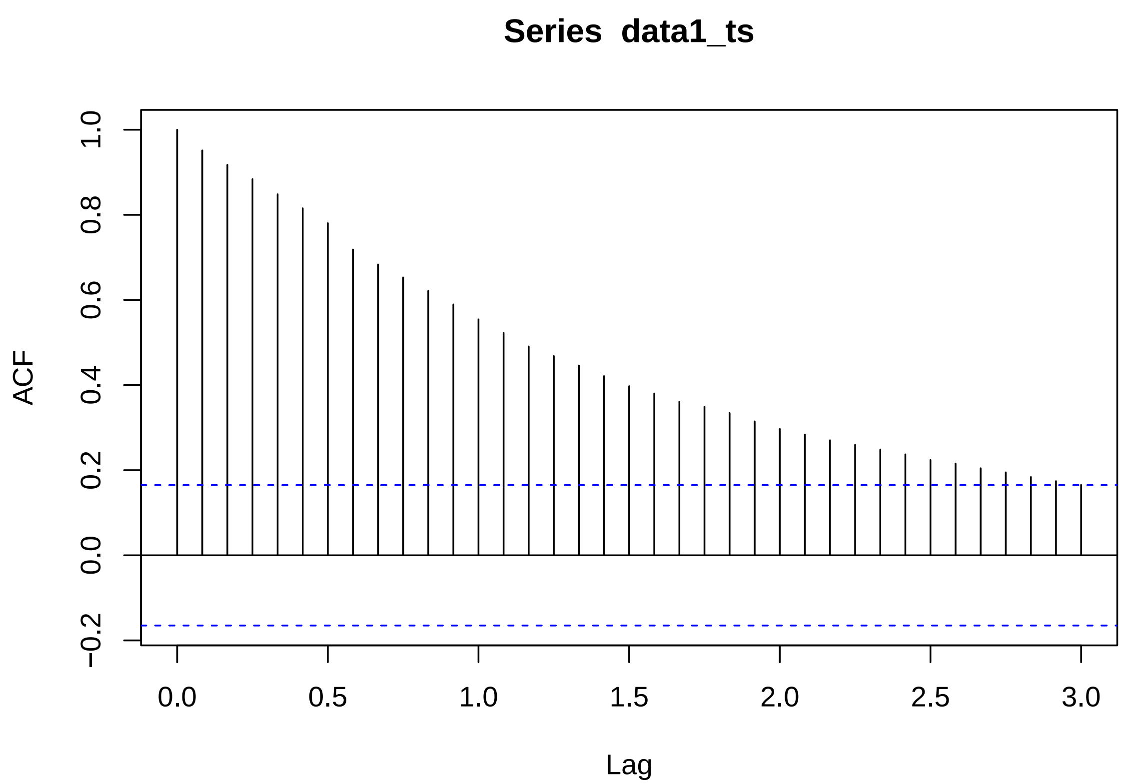
```
[1] 66020.52
```

```r
media_amostral2
```

```
[1] 46.89444
```

- C_k e r_k para k = 1,...,36.

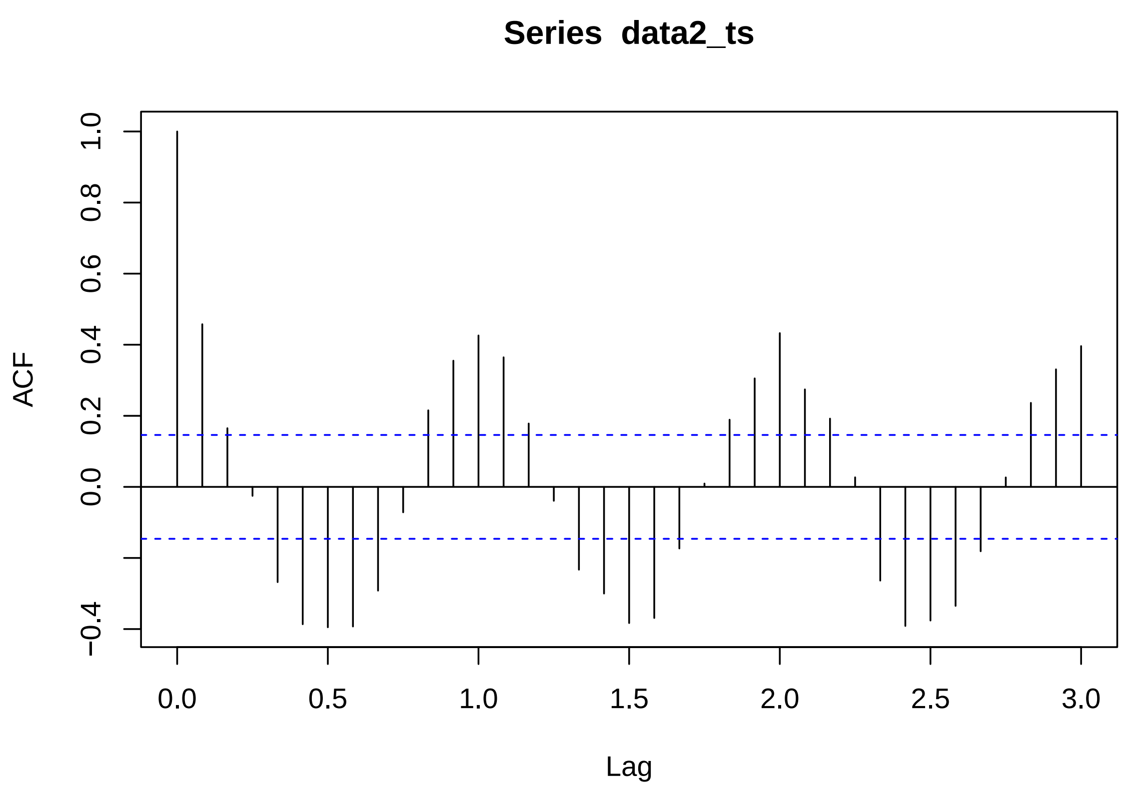
```r
# calculando a autocorrelação (ACF) para K = 1, ..., 36

acf1 <- acf(data1_ts, lag.max = 36, plot = TRUE)$acf[-1] # remove o lag 0
```



```r
acf1
```

```
 [1] 0.9514160 0.9174419 0.8840017 0.8483770 0.8154543 0.7804471 0.7185768
 [8] 0.6834363 0.6528914 0.6214676 0.5895272 0.5543717 0.5225446 0.4908249
[15] 0.4682221 0.4459799 0.4211280 0.3974041 0.3801805 0.3612299 0.3494211
[22] 0.3342753 0.3146517 0.2967974 0.2838737 0.2703261 0.2597187 0.2483536
[29] 0.2371584 0.2239555 0.2157054 0.2045383 0.1949174 0.1838837 0.1742343
[36] 0.1654774
```

```r
acf2 <- acf(data2_ts, lag.max = 36, plot = TRUE)$acf[-1]
```




```r
acf2
```

```
[1]  0.457739053  0.165169955 -0.025204957 -0.267903375 -0.386381327
[6] -0.395016694 -0.392795163 -0.291885862 -0.071538185  0.215495164
```

```
[11]  0.355054343  0.426085941  0.364615242  0.178311373 -0.039130730
[16] -0.232983672 -0.300129749 -0.383149164 -0.368928301 -0.173292411
[21]  0.009378308  0.189056564  0.305343616  0.432648385  0.274358601
[26]  0.192129752  0.026931755 -0.263676474 -0.391256416 -0.376087161
[31] -0.334783444 -0.181118099  0.026688826  0.236411592  0.330787234
[36]  0.396128713
```

```r
# calculando a variância da série

var1 <- var(data1_ts)

var2 <- var(data2_ts)

var1;var2
```

```
[1] 2915802343
```

```
[1] 539.0223
```

```r
# calculando a autocovariância

autocov1 <- acf1 * var1

autocov2 <- acf2 * var2

autocov1
```

```
 [1] 2774141108 2675079164 2577574088 2473699718 2377703470 2275629450
 [7] 2095227879 1992765203 1903702323 1812076624 1718944670 1616438264
[13] 1523636753 1431148457 1365243012 1300389134 1227925872 1158751931
[19] 1108531136 1053274934 1018842942  974680667  917462113  865402597
[25]  827719475  788217584  757288446  724150126  691507011  653009891
[31]  628954284  596393368  568340648  536168404  508032881  482499493
```

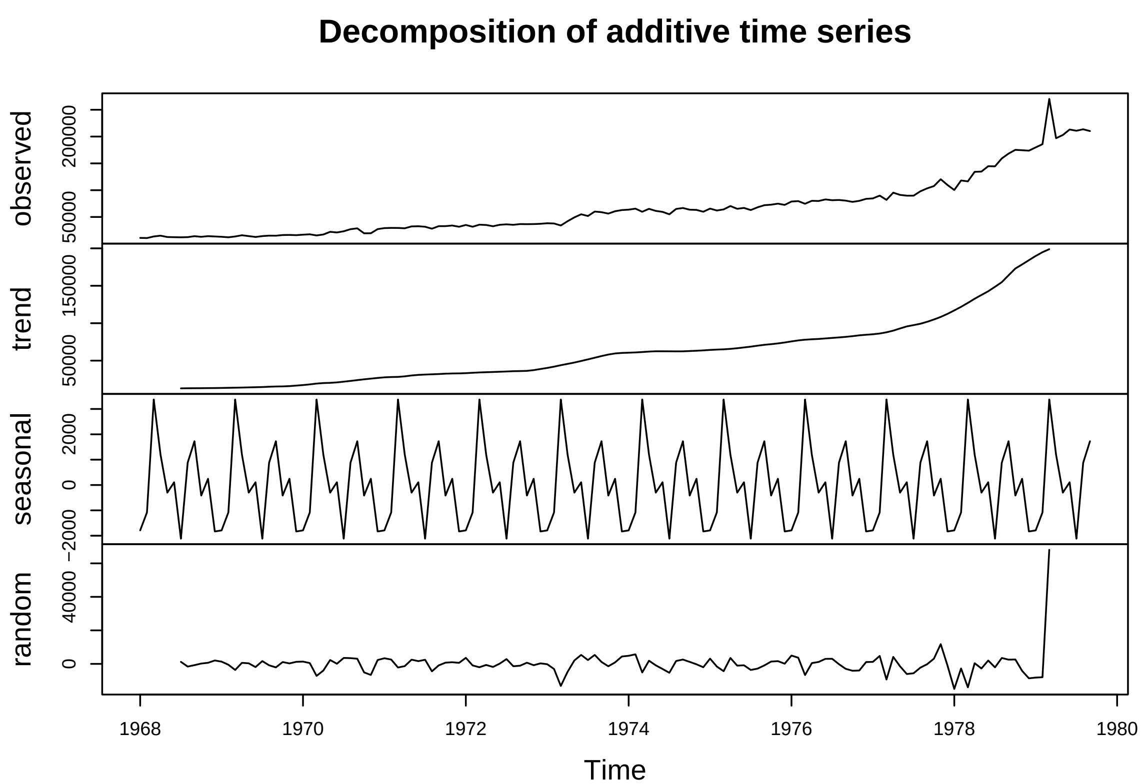
```r
autocov2
```

```
 [1]  246.731564   89.030291  -13.586034 -144.405898 -208.268157 -212.922813
 [7] -211.725358 -157.332993  -38.560678  116.156702  191.382214  229.669830
[13]  196.535752   96.113809  -21.092337 -125.583398 -161.776632 -206.525950
```

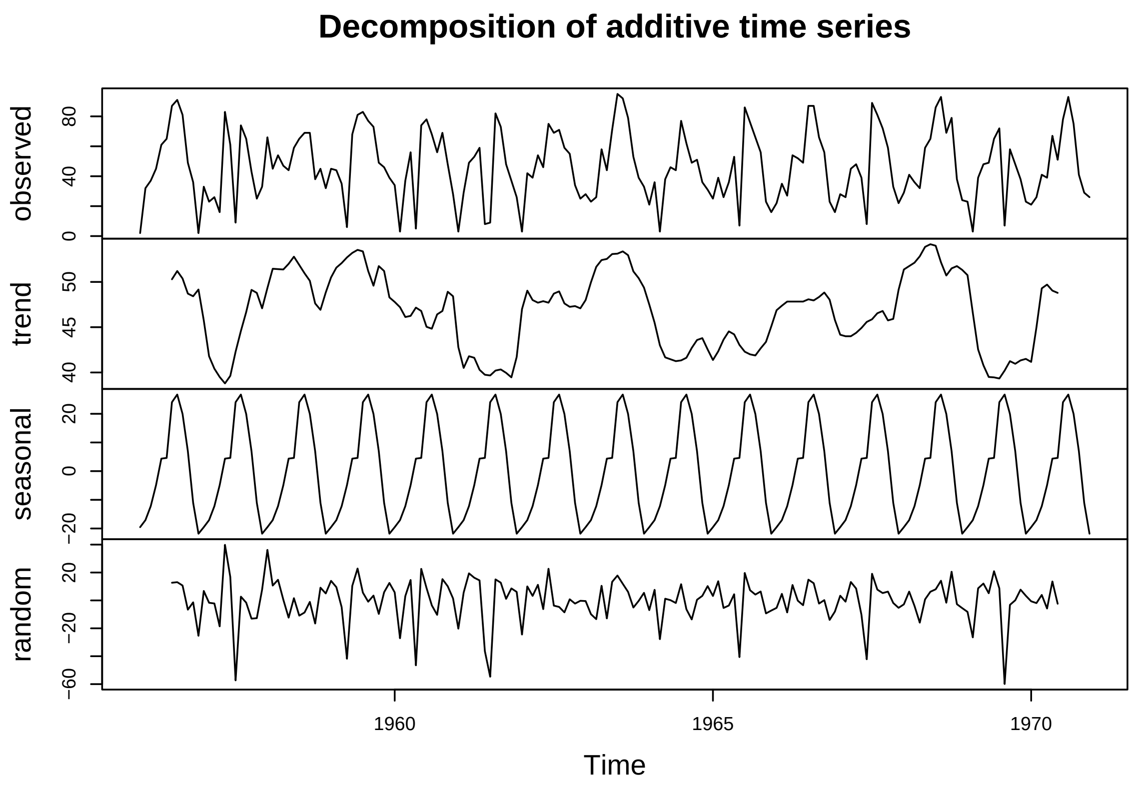
```
[19] -198.860587  -93.408476    5.055117  101.905707  164.587023  233.207134
[25]  147.885408  103.562224   14.516817 -142.127503 -210.895939 -202.719372
[31] -180.455747  -97.626697   14.385873  127.431124  178.301701  213.522216
```

- Faça os gráficos da série e de r_k. Comente sobre a presença de tendencia, sazonalidades, ciclos.

```r
# gráficos da série, sazonalidade e tendencia

decompose_graph1 <- decompose(data1_ts)

decompose_graph2 <- decompose(data2_ts)

plot(decompose_graph1)
```




```r
plot(decompose_graph2)
```



O primeiro gráfico apresenta uma série temporal com uma tendência aparentemente crescente, com um pico acentuado próximo ao final do período, acompanhado por uma variabilidade aleatória mais intensa. Já o segundo gráfico reflete uma série mais estacionária, com a tendência oscilando suavemente ao longo do tempo, sem padrões de crescimento ou declínio definidos. Em ambas as séries, a sazonalidade se mostra regular e consistente. A principal diferença está no comportamento geral: enquanto a primeira série evidencia um crescimento acentuado e eventos fora do padrão, a segunda mantém estabilidade e equilíbrio entre os componentes.

0.6 Questão 6

Use um programa computacional e a série dos log retornos mensais do IBOVESPA para calcular:

```
data <- tq_get(
  "^BVSP",
  from = "1994-07-04",
```

```
  to = "2010-08-19",
  get = "stock.prices")

data <- data %>%
  select(-symbol) %>%
  mutate(date = as.Date(date, format = "%Y-%m-%d"),
         log_return = log(adjusted / lag(adjusted)))

data <- data %>%
  mutate(month = format(date, "%Y-%m")) %>%
  group_by(month) %>%
  summarise(
    log_return_mes = sum(log_return, na.rm = TRUE)
  )
```

- média e variância amostrais, coeficientes de assimetria e curtose, máximo e mínimo, histogramas;

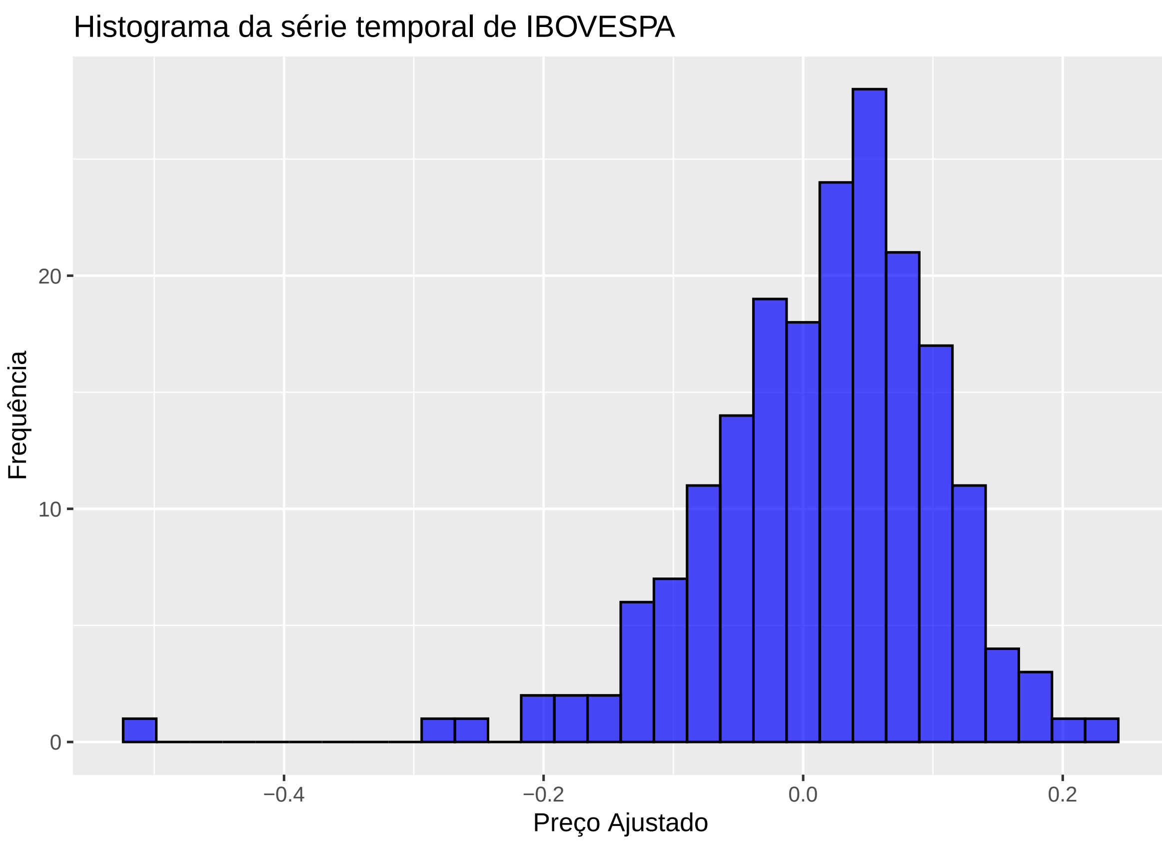
```
stats_data <- data %>%
  summarise(
    media = mean(log_return_mes, na.rm = TRUE),
    variancia = var(log_return_mes, na.rm = TRUE),
    assimetria = skewness(log_return_mes, na.rm = TRUE),
    curtose = kurtosis(log_return_mes, na.rm = TRUE),
    max = max(log_return_mes, na.rm = TRUE),
    min = min(log_return_mes, na.rm = TRUE)
  )

kable(stats_data,
      caption = "Estatísticas descritivas do Log Retorno mensal do IBOVESPA",
      align = "c") %>%
  kable_styling(bootstrap_options = c("striped"), full_width = FALSE)
```

Table 3: Estatísticas descritivas do Log Retorno mensal do IBOVESPA

media	variancia	assimetria	curtose	max	min
0.0140989	0.0086741	-1.231667	4.574243	0.2378635	-0.5034126

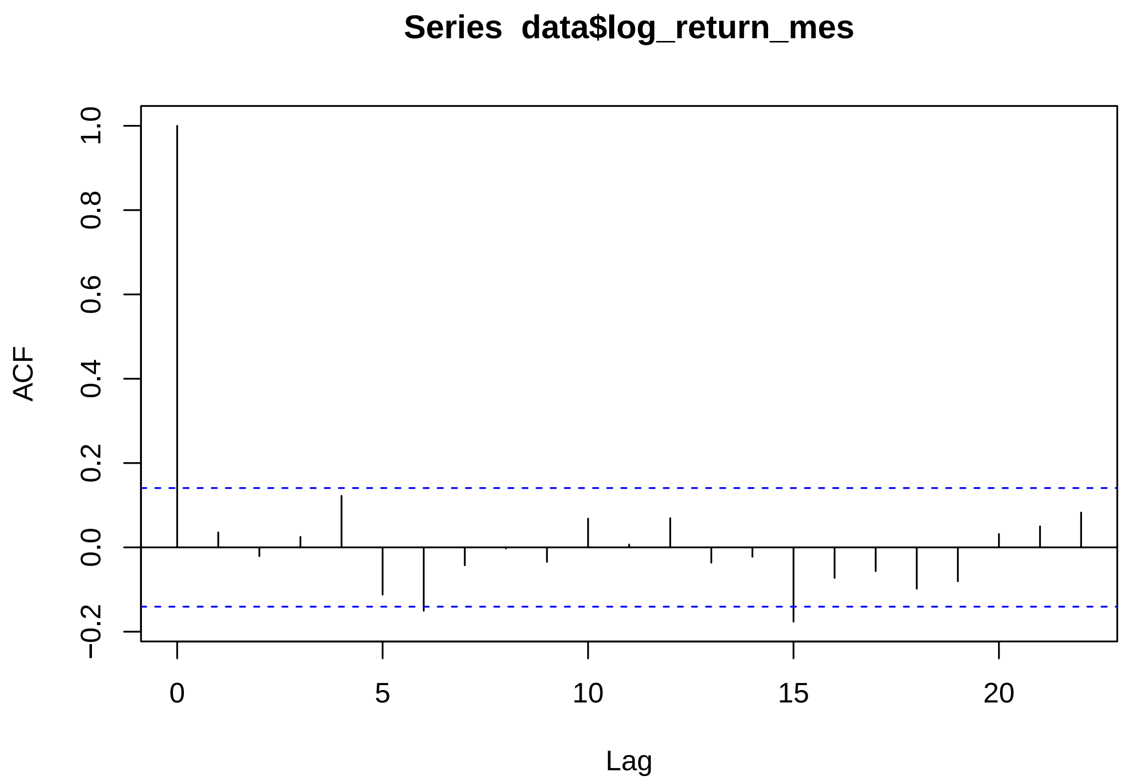
```
ggplot(data, aes(x = log_return_mes)) +
  geom_histogram(bins = 30, fill = "blue", color = "black", alpha = 0.7) +
```

```
  labs(title = "Histograma da série temporal de IBOVESPA",
       x = "Preço Ajustado",
       y = "Frequência")
```



- autocorreções amostrais

```
acf <- acf(data$log_return_mes, plot = TRUE)
```




```r
acf
```

```
Autocorrelations of series 'data$log_return_mes', by lag

     0      1      2      3      4      5      6      7      8      9     10 
 1.000  0.036 -0.021  0.025  0.122 -0.112 -0.150 -0.042 -0.003 -0.034  0.068 
    11     12     13     14     15     16     17     18     19     20     21 
 0.007  0.069 -0.036 -0.022 -0.176 -0.072 -0.056 -0.098 -0.080  0.032  0.050 
    22 
 0.083 
```