Estudo de Caso 02: Avaliação e comparação do retorno médio de ações

Diego Pontes, Elias Vieira, Matheus Bitarães

Fevereiro, 2021

Descrição do problema

Introdução

Design do Experimento

Os dados de entrada do experimento são as informações de preços de fechamento mensais de 5 ações diferentes, onde cada coluna representa uma ação e cada linha representa um mes de fechamento. O que se deseja é comparar o potencial de cada ação em gerar maior ganho mensal ao investidor. Como modelos regressivos de previsão não são o escopo deste trabalho, pode-se realizar uma transformação nestes dados, de forma que haja um vetor com as flutuações percentuais das ações em cada mês. Por exemplo, se houver uma ação com preços de fechamento $[10, 11, 12, 10]$, pode-se gerar o seguinte vetor de flutuações percentuais: $[10\%, 9\%, -16\%]$.

Desta forma, é possivel realizar uma análise estatistica entre as 5 ações e identificar a que apresenta maior incidencia de flutuação positiva, o que será considerado como a ação de maior potencial para gerar retornos.

Análise Estatística

Importação dos dados

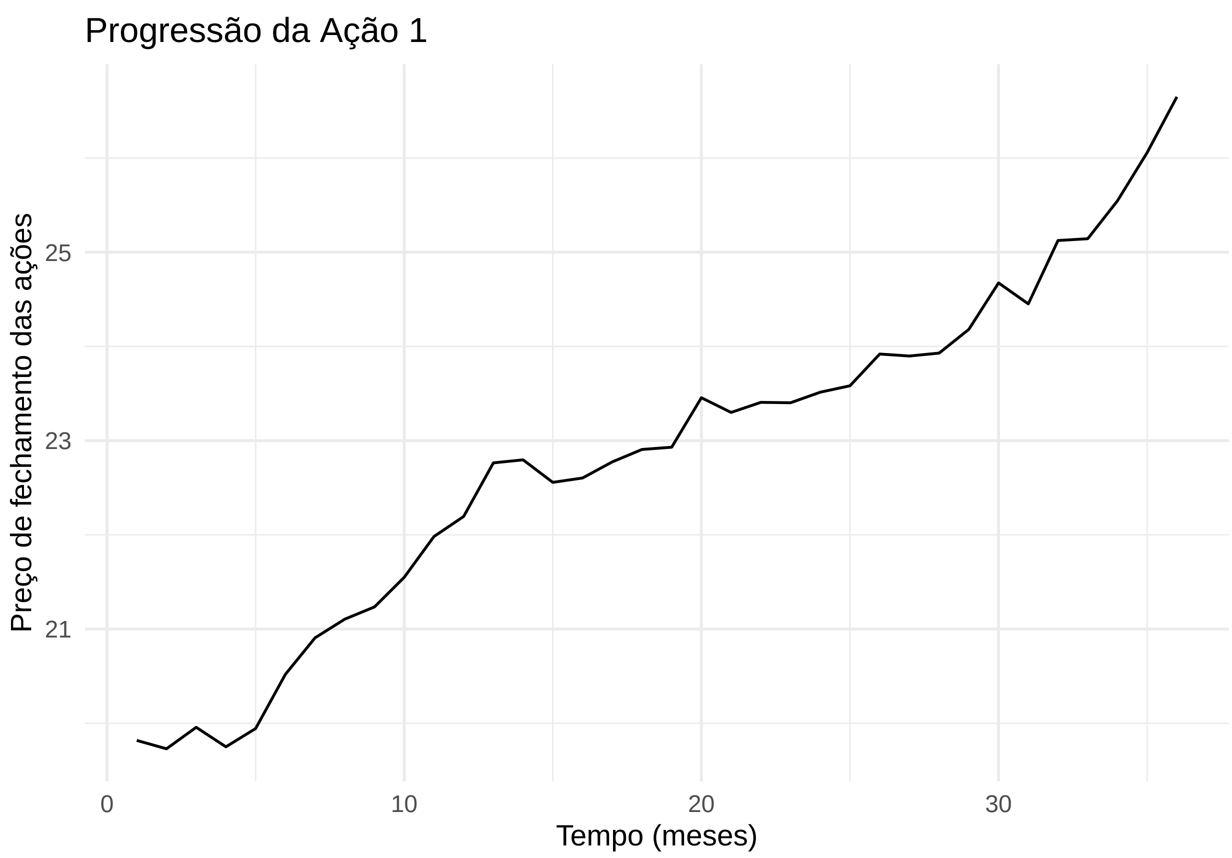
Os dados das ações foram importados do arquivo *DadosAcoesGrupoC.csv*.

```r
# importação dos dados
data <- read.csv(file = 'DadosAcoesGrupoC.csv', header = FALSE)
colnames(data) <- c("A1","A2","A3","A4","A5") # Adicionando nomes às colunas

# plot dos primeiros 6 dados da tabela
head(data)
```

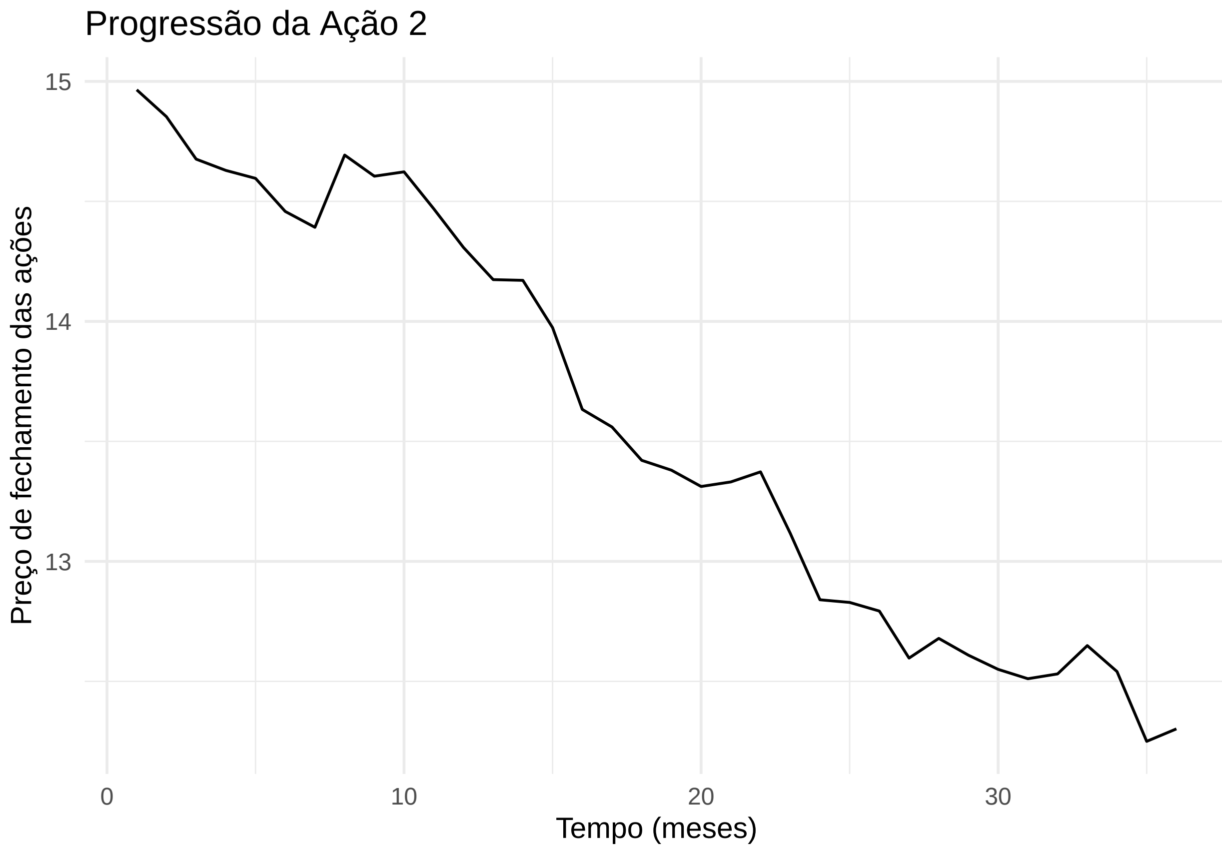
```
##       A1     A2     A3     A4     A5
## 1 26.649 12.302 21.006 18.553 33.937
## 2 26.057 12.250 21.032 18.385 33.443
## 3 25.545 12.541 20.816 18.555 33.018
## 4 25.143 12.649 20.790 18.513 33.050
## 5 25.124 12.531 20.553 18.376 32.758
## 6 24.451 12.511 20.520 18.490 32.094
```

Cada coluna representa uma ação e cada linha representa o preço de fechamento das ações no mes anterior. Portanto, a linha 1 indica o preço de fechamento do mes atual - 1, a linha 2 representa o mes atual - 2, e assim sucessivamente.

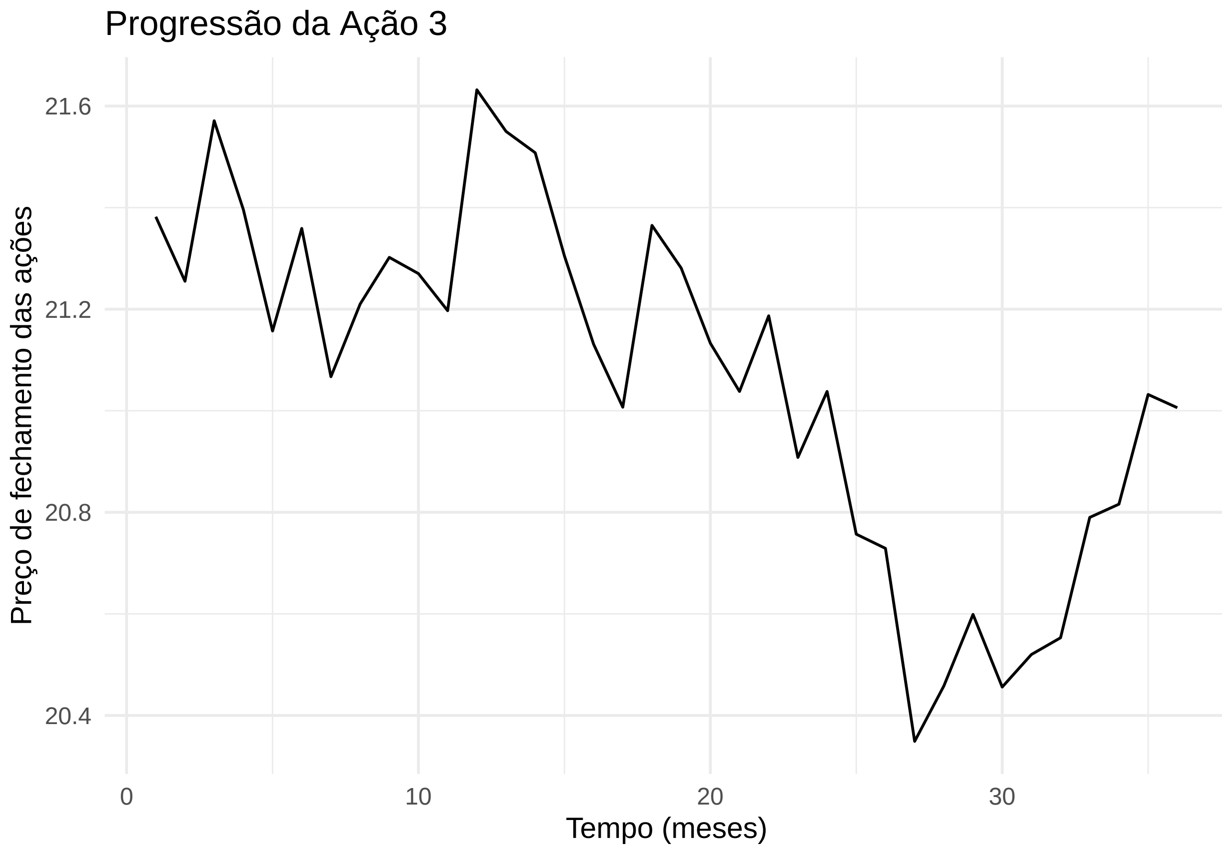
```
theme_set(theme_minimal())

# Plot
plot_a1 <- ggplot(data=data, aes(x=nrow(data):1, y=A1)) + geom_line() +
  labs(title = "Progressão da Ação 1", x = "Tempo (meses)", y = "Preço de fechamento das ações")
plot_a2 <- ggplot(data=data, aes(x=nrow(data):1, y=A2)) + geom_line() +
  labs(title = "Progressão da Ação 2", x = "Tempo (meses)", y = "Preço de fechamento das ações")
plot_a3 <- ggplot(data=data, aes(x=nrow(data):1, y=A3)) + geom_line() +
  labs(title = "Progressão da Ação 3", x = "Tempo (meses)", y = "Preço de fechamento das ações")
plot_a4 <- ggplot(data=data, aes(x=nrow(data):1, y=A4)) + geom_line() +
  labs(title = "Progressão da Ação 4", x = "Tempo (meses)", y = "Preço de fechamento das ações")
plot_a5 <- ggplot(data=data, aes(x=nrow(data):1, y=A5)) + geom_line() +
  labs(title = "Progressão da Ação 5", x = "Tempo (meses)", y = "Preço de fechamento das ações")

plot_a1
```



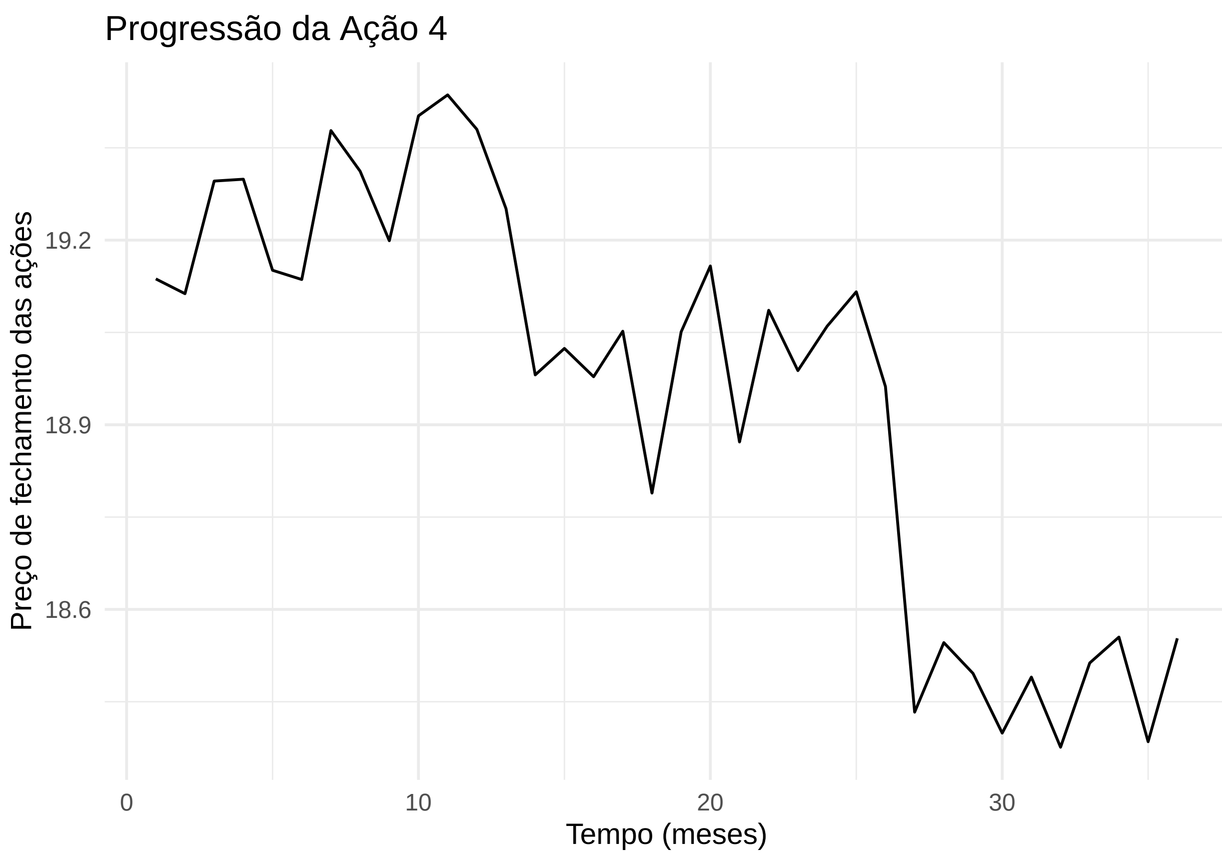

```
plot_a2
```


Progressão da Ação 2
15
14
13
Preço de fechamento das ações
0
10
20
30
Tempo (meses)

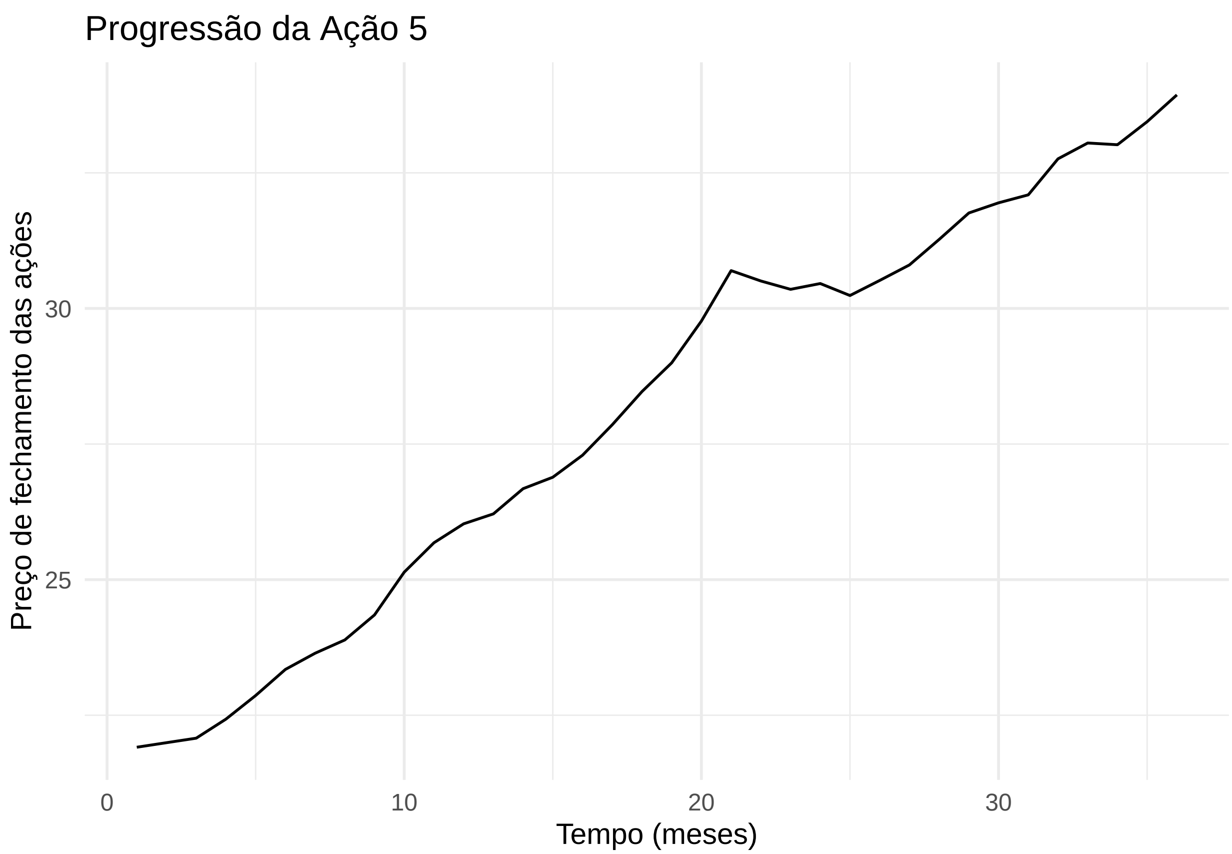
```
plot_a3
```


Progressão da Ação 3
21.6
21.2
20.8
20.4
Preço de fechamento das ações
0
10
20
30
Tempo (meses)

```
plot_a4
```


Progressão da Ação 4
Preço de fechamento das ações
19.2
18.9
18.6
0
10
20
30
Tempo (meses)

```
plot_a5
```



Tratamento dos dados

Os dados brutos serão transformados em flutuações percentuais para que se possa realizar a análise estatística.

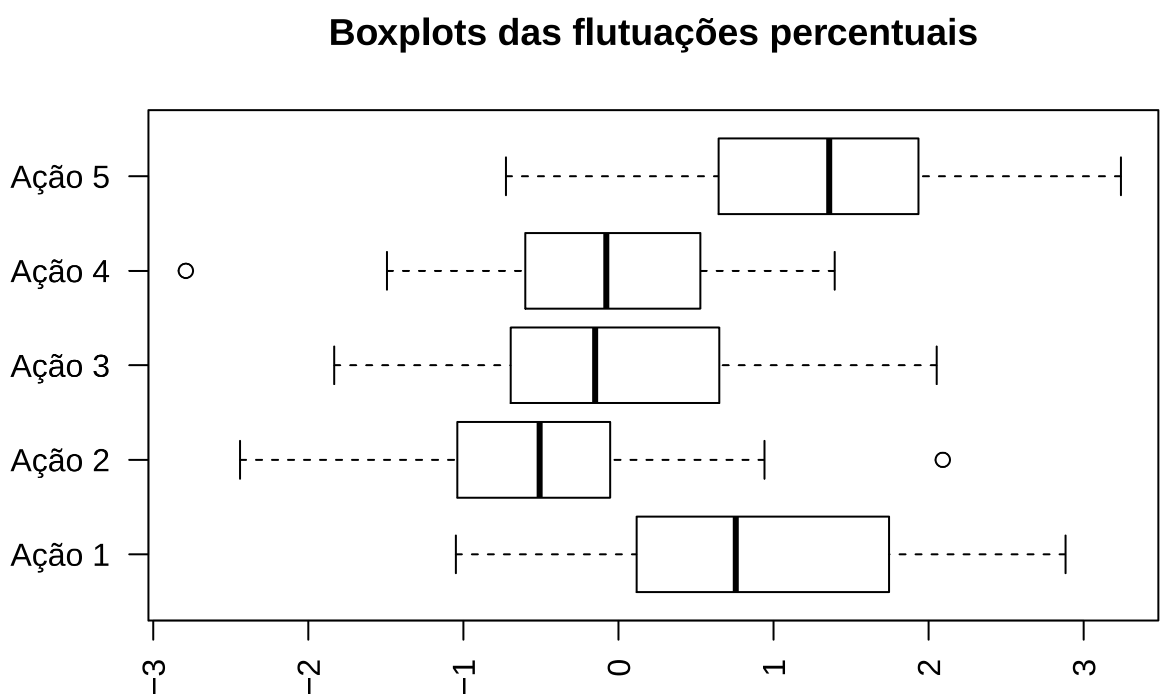
```
# Transforma dados em porcentagens
data_perc <- data
for(i in (nrow(data)-1):1){
        data_perc[i,] <- 100*( (data[i,] - data[i+1,])/data[i+1,] ) # porcentagem de mudança do fechame
}

data_perc <- data_perc[-c(36), ] # removendo ultima linha

# Criando uma variável para cada amostra
data_a1 <- data_perc$A1
data_a2 <- data_perc$A2
data_a3 <- data_perc$A3
data_a4 <- data_perc$A4
data_a5 <- data_perc$A5
head(data_perc)
```

```
##          A1         A2         A3         A4         A5
## 1  2.2719423  0.4244898 -0.1236211  0.9137884  1.4771402
## 2  2.0043061 -2.3203891  1.0376633 -0.9161951  1.2871767
## 3  1.5988546 -0.8538224  0.1250601  0.2268676 -0.0968230
## 4  0.0756249  0.9416647  1.1531163  0.7455377  0.8913853
```

```
## 5  2.7524437  0.1598593  0.1608187 -0.6165495  2.0689225
## 6 -0.9037854 -0.3107570  0.3128666  0.4945921  0.4601371
```

algum comentario aqui

```r
# boxplot
boxplot(data_a1, data_a2, data_a3, data_a4, data_a5,
main = "Boxplots das flutuações percentuais",
at = c(1,2,3,4,5),
names = c("Ação 1", "Ação 2", "Ação 3", "Ação 4", "Ação 5"),
las = 2,
horizontal = TRUE,
notch = FALSE
)
```



algum comentario sobre o boxplot

Dados estatísticos

O teste escolhido foi o ANOVA. . . . . . [elaborar]

Premissas do teste Para a realização do teste ANOVA, é necessário que as seguintes premissas sejam cumpridas: As amostras devem ser independentes, as amostras devem apresentar distribuição normal e as variâncias podem ser consideradas iguais (homocedasticidade).

[falar sobre independencia das amostras]

Para verificação de normalidade, pode-se realizar o teste de Shapiro-Wilk.

```
library(car)

apply(data_perc, 2, shapiro.test)
```

```
## $A1
##
##  Shapiro-Wilk normality test
##
## data:  newX[, i]
## W = 0.96979, p-value = 0.4374
##
##
## $A2
##
##  Shapiro-Wilk normality test
##
## data:  newX[, i]
## W = 0.9722, p-value = 0.5065
##
##
## $A3
##
##  Shapiro-Wilk normality test
##
## data:  newX[, i]
## W = 0.9827, p-value = 0.8428
##
##
## $A4
##
##  Shapiro-Wilk normality test
##
## data:  newX[, i]
## W = 0.96418, p-value = 0.304
##
##
## $A5
##
##  Shapiro-Wilk normality test
##
## data:  newX[, i]
## W = 0.97875, p-value = 0.718
```

Como interpretação do teste, temos que se o p-valor < 0.05 (α), deve-se rejeitar a hipótese nula, ou seja, os dados não possuem distribuição normal [1], caso contrário, não há evidências para se rejeitar a hipótese nula. Portanto, analisando os resultados dos testes dispostos acima, podemos considerar que as amostras de todas as flutuações percentuais das ações seguem distribuição normal

Para a verificação de homocedasticidade entre dois ou mais grupos de amostras, pode-se utilizar o teste de Bartlett [Explicar mais sobre o teste de Bartlett]

```r
# teste de Bartlett
y <- c(data_a1, data_a2, data_a3, data_a4, data_a5)
group <- as.factor(c(rep(1, length(data_a1)), rep(2, length(data_a2)),
                     rep(3, length(data_a3)), rep(4, length(data_a4)), rep(5, length(data_a5))))
bartlett.test(y, group)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  y and group
## Bartlett's K-squared = 1.6615, df = 4, p-value = 0.7977
```

Como interpretação do teste, temos que se o p-valor < 0.05 (α), deve-se rejeitar a hipótese nula, ou seja, os grupos de amotras não possuem equivalencia de variâncias. Como o p-valor > 0.05 , não há evidências para se rejeitar a hipótese nula e portanto podemos considerar homocedasticidade entre as amostras.

Teste de hipóteses

Tendo as premissas validadas, foi realizado o teste one way ANOVA

```r
# transformação dos dados
perc <- c(data_a1, data_a2, data_a3, data_a4, data_a5)
groups <- c(rep(1, length(data_a1)), rep(2, length(data_a2)), rep(3, length(data_a3)), rep(4, length(da
data_t = data.frame(perc, groups)

# ANOVA
model <- aov(perc ~ group, data = data_t)
summary(model)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## group         4  78.11  19.527   20.95 4.66e-14 ***
## Residuals   170 158.44   0.932
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Discussão e Conclusão

Atividades dos membros

Diego

Elias

Matheus

Todos

Elaboração das hipóteses e definição das premissas.

Referências Bibliográficas

[1] Como realizar teste de normalidade no r? https://rpubs.com/paternogbc/46768. Acesso em 18 de Janeiro de 2021.