

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

Diego Pontes, Elias Vieira, Matheus Bitarães

Fevereiro, 2021

## Descrição do problema

Neste estudo, deseja-se comparar a rentabilidade de cinco ações diferentes no decorrer dos últimos 36 meses a fim de encontrar aquela com maior rentabilidade para que determinado investidor possa aplicar todo montante financeiro para um maior rendimento para o próximo mês.

## Introdução

Nos últimos anos, o mercado financeiro no Brasil tem ganhado bastante destaque. Devido a queda na taxa básica de juros, que está em sua mínima histórica e, conseqüentemente, a redução nos rendimentos dos investimentos em renda fixa, os investidores estão migrando cada vez mais da renda fixa para a renda variável buscando um maior retorno em seus investimentos.

Esse cenário tem impulsionado o setor financeiro de uma forma como nunca tinha sido vista no país. Com isso, corretoras, casas de análises, agentes de investimentos e outros serviços têm sido cada vez mais procurados e solicitados pelos novos investidores que desejam entrar na renda variável, em especial na bolsa de valores.

Esses novos investidores, são motivados principalmente pela possibilidade de maiores retornos que a renda variável oferece aos seus investimentos. Porém, quanto maior a possibilidade de retorno, geralmente também, maior será o risco envolvido em um investimento. Por isso, cada vez mais os agentes financeiros tem procurado meios de melhorar suas análises, buscando os melhores ativos para seus clientes.

## Design do Experimento

Os dados de entrada do experimento são as informações de preços de fechamento dos últimos 36 meses de 5 ações diferentes. 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 possível realizar uma análise estatística entre as 5 ações e identificar a que apresenta maior incidência de flutuação positiva.

## Análise Estatística

### Importação dos dados

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

```
# 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 mês anterior. Portanto, a linha 1 indica o preço de fechamento do mês atual - 1, a linha 2 representa o mês atual - 2, e assim sucessivamente. Desta forma, foi plotado os gráficos abaixo, nos quais o eixo x, que representa os meses, inicia do mês mais antigo e progride para o mais recente, sendo x igual a zero o mês atual.

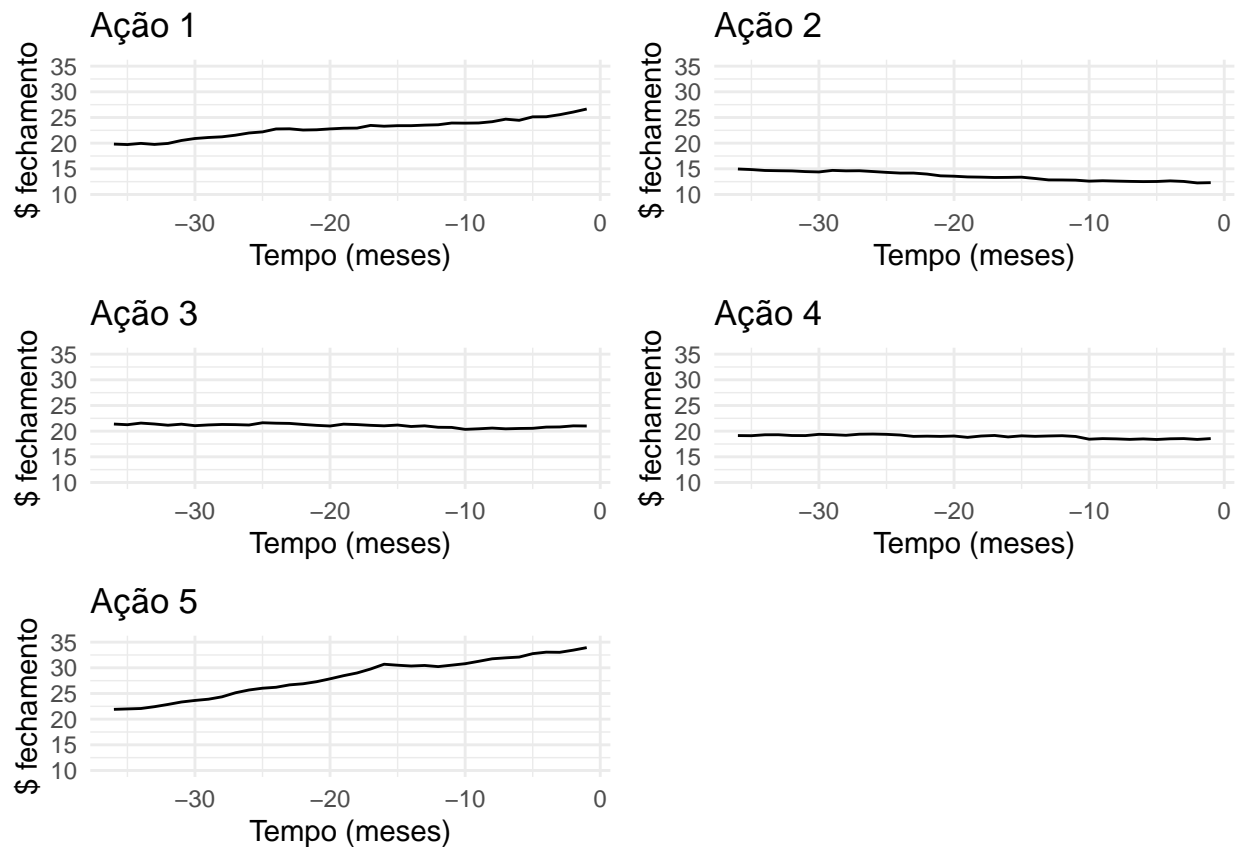
```
theme_set(theme_minimal())
numeracao <- -1:-36 #numerando o eixo x do valor mais antigo (mesatual - 36) ao mês anterior (mesatual)

# Plot

#bkp:nrow(data):1 basta trocar o numeracao e retirar
#o texto do paragrafo acima

plot_a1 <- ggplot(data=data, aes(x=numeracao, y=A1)) + geom_line() +
  labs(title = "Ação 1", x = "Tempo (meses)", y = "$ fechamento") + ylim(10, 35)
plot_a2 <- ggplot(data=data, aes(x=numeracao, y=A2)) + geom_line() +
  labs(title = "Ação 2", x = "Tempo (meses)", y = "$ fechamento") + ylim(10, 35)
plot_a3 <- ggplot(data=data, aes(x=numeracao, y=A3)) + geom_line() +
  labs(title = "Ação 3", x = "Tempo (meses)", y = "$ fechamento") + ylim(10, 35)
plot_a4 <- ggplot(data=data, aes(x=numeracao, y=A4)) + geom_line() +
  labs(title = "Ação 4", x = "Tempo (meses)", y = "$ fechamento") + ylim(10, 35)
plot_a5 <- ggplot(data=data, aes(x=numeracao, y=A5)) + geom_line() +
  labs(title = "Ação 5", x = "Tempo (meses)", y = "$ fechamento") + ylim(10, 35)

grid.arrange(plot_a1, plot_a2, plot_a3, plot_a4, plot_a5, nrow = 3)
```



## 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 fechamento do mes i + 1 em relação ao mes i.
}

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

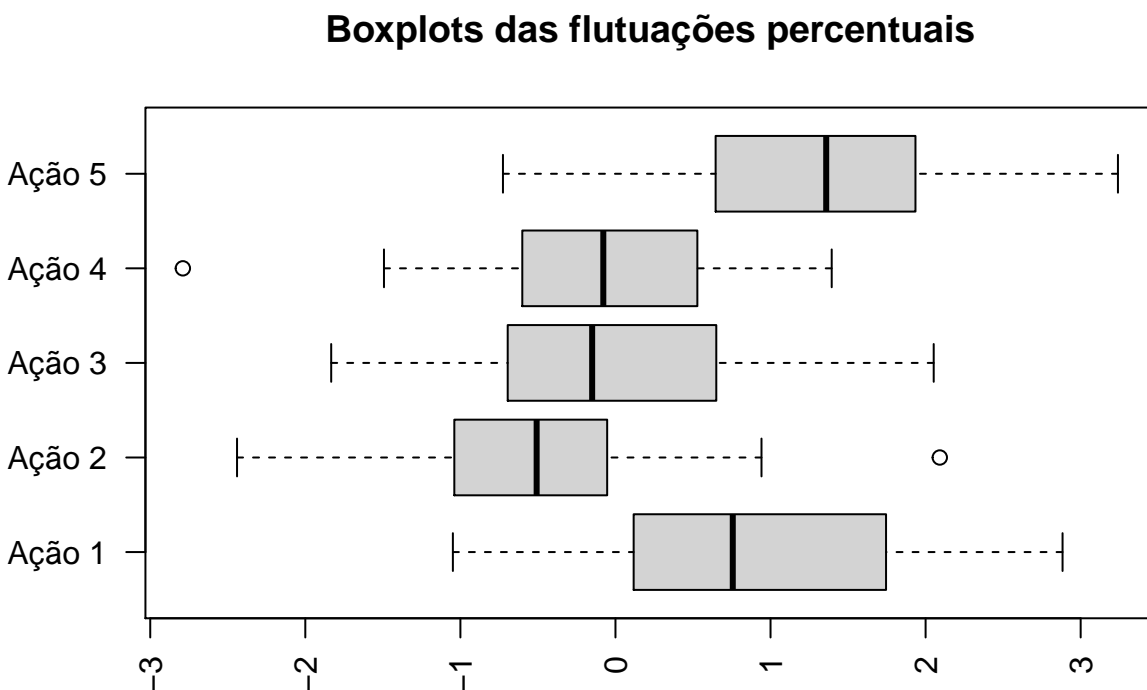
# Criando uma variável para cada amostra
data_a1 <- data_perc$A1
data_a2 <- data_perc$A2
data_a3 <- data_perc$A3
data_a4 <- data_perc$A4
data_a5 <- data_perc$A5
head(data_perc)
```

```
##           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
```

Em posse das flutuações percentuais, foi plotado o seguinte boxplot para uma análise visual das variações destes valores para cada uma das ações.

```
# 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
)
```



Em uma análise visual, é possível perceber que as ações 1 e 5 possuem os maiores valores de variações positivos de um mês para outro nestes 36 meses, enquanto a ação 2 possui a maior variação negativa nesta mesma situação. No entanto, estas inferências não são suficientes visto que, hipoteticamente, uma ação pode ter uma subida abrupta em um determinado momento e uma sequência considerável de quedas de forma a atingir valores menores do que o valor inicial dessa suposição. Da mesma forma, esta mesma ação pode ter uma queda abrupta em um determinado momento e uma sequência considerável de subidas de forma a atingir valores maiores do que o valor inicial dessa suposição. Assim, percebe-se a importância de uma análise estatística dos dados históricos no período em comparação para uma tomada de decisão mais pertinente,

visto que a média do comportamento de uma ação carrega um valor estatístico maior do que uma análise isolada de algum mês específico, por exemplo.

## Dados estatísticos

Para a análise destas ações, foi escolhido o teste da Análise de Variância (ANOVA) capaz de determinar se as médias de três ou mais grupos são diferentes. A ANOVA usa testes F para testar estatisticamente a igualdade entre médias [1].

**Premissas do teste:** Para a realização do teste ANOVA, é necessário que as seguintes premissas sejam cumpridas [2]:

- As amostras devem ser independentes;
- As amostras devem apresentar distribuição normal;
- As variâncias podem ser consideradas iguais (homocedasticidade).

**Verificação da independência:** Como não há maiores informações sobre o ramo de atuação das empresas e suas origens, assume-se que elas não possuem nenhum fator que causa interferência entre elas. Portanto, serão consideradas independentes.

**Verificação de Normalidade:** Para uma avaliação estatística sobre a validação da hipótese de uma distribuição normal para as amostras, pode-se usar o teste Shapiro-Wilk, onde têm-se as seguintes hipóteses [3]:

$$\begin{cases} H_0 : \text{A amostra provém de uma população com distribuição normal} \\ H_1 : \text{A amostra não provém de uma população com distribuição normal} \end{cases}$$

```
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, tem-se que se o p-valor  $< 0.05$  ( $\alpha$ ), deve-se rejeitar a hipótese nula, ou seja, os dados não possuem distribuição normal [3], caso contrário, não há evidências para se rejeitar a hipótese nula. Portanto, analisando os resultados dos testes dispostos acima, pode-se considerar que as amostras de todas as flutuações percentuais das ações seguem distribuição normal.

**Verificação de homocedasticidade:** Para a verificação de homocedasticidade entre dois ou mais grupos de amostras, pode-se utilizar o teste de Levene ou Bartlett, onde têm-se as seguintes hipóteses [4]:

$$\begin{cases} H_0 : \text{As variâncias dos erros são homogêneas} \\ H_1 : \text{As variâncias dos erros não são homogêneas} \end{cases}$$

O teste de Bartlett é sensível em relação a hipótese de normalidade dos dados. Se a hipótese da normalidade tivesse sido rejeitada, seria melhor utilizar o teste proposto por Levene. Porém, se a hipótese de normalidade não for rejeitada, o teste proposto por Bartlett tem um comportamento melhor que o teste proposto por Levene. Com o objetivo de agregar mais valor ao estudo, ambos foram efetuados:

```
# 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
```

```
# teste de Levene
leveneTest(y, group)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group  4  0.7645 0.5497
##      170
```

Como interpretação do teste, tem-se que se o p-valor  $< 0.05$  ( $\alpha$ ), deve-se rejeitar a hipótese nula, ou seja, as variâncias dos grupos de amostras são homogêneas, caso contrário, não há evidências para se rejeitar a hipótese nula. Portanto, analisando os resultados dos testes dispostos acima, onde o p-valor é maior que 0.05, pode-se considerar que as variâncias são homogêneas.

Dada a independência considerada entre os valores, a análise e confirmação da normalidade e homocedasticidade dos dados, pode-se aplicar o teste ANOVA, onde têm-se as seguintes hipóteses:

$$\begin{cases} H_0 : \text{A média de todas as populações são iguais} \\ H_1 : \text{A média de pelo menos uma população é diferente} \end{cases}$$

```
# teste ANOVA
```

```
anova <- aov(y~group)
```

```
anova
```

```
## Call:
```

```
##      aov(formula = y ~ group)
```

```
##
```

```
## Terms:
```

```
##              group Residuals
```

```
## Sum of Squares   78.10652 158.43670
```

```
## Deg. of Freedom      4      170
```

```
##
```

```
## Residual standard error: 0.9653914
```

```
## Estimated effects may be unbalanced
```

```
summary(anova)
```

```
##              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
```

Como interpretação do teste, temos que se o p-valor  $< 0.05$  ( $\alpha$ ), deve-se rejeitar a hipótese nula, ou seja, as médias dos grupos de amostras são homogêneas, caso contrário, não há evidências para se rejeitar a hipótese nula. Portanto, analisando os resultados dos testes dispostos acima, podemos considerar que as médias dos grupos não são iguais. Ou seja, pelo menos uma das amostras tem sua média diferente. Desta forma, é possível concluir que, estatisticamente, haverá diferença em pelo menos uma ação dentre as cinco possíveis para investimento. Essa conclusão, embora importante, não é suficiente, pois ainda não se sabe quantas ações possuem médias diferentes e se são superiores ou inferiores às demais.

Portanto, para dar sequência aos estudos, optou-se por fazer uma análise por pares, para que seja feita uma varredura entre todas as ações em busca daquelas onde a diferença entre as médias é significativa, as chamadas comparações múltiplas. Neste contexto, será aplicado o Teste de Tukey. Este teste foi escolhido, pois quando os tamanhos amostrais dos grupos são iguais, o Teste de Tukey é um teste exato, ou seja, para o conjunto de todas as comparações par a par, a taxa de erro do conjunto dos testes é exatamente  $\alpha$  (nível de significância) e o intervalo de confiança é também exatamente  $1 - \alpha$ . Vale ressaltar que testes de comparações múltiplas exatos são raros, uma vez que a maioria não controla o nível de significância adotado [5].

Para realizar o Teste de Tukey, deve ser levada em conta as seguintes suposições [3]:

- As observações são independentes dentro e entre os grupos;
- Os grupos devem ser normalmente distribuídos;

- A variância dentro do grupo deve ser constante.

Como não houve alteração nas amostras, pode-se concluir que todas as premissas foram atendidas e, portanto, o teste pode ser efetuado.

```
# teste Tukey
library(agricolae)
data3 = data.frame(y,group)
model2 <- aov(y~group, data = data3)
tukey = TukeyHSD(x=model2, "group")
tukey

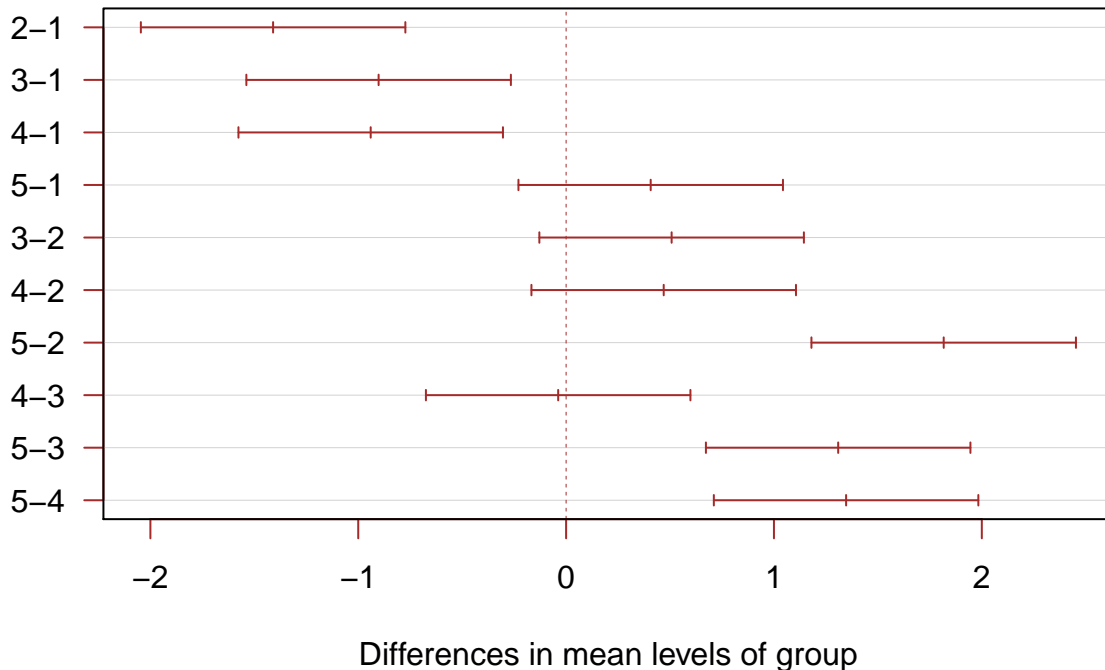
##    Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = y ~ group, data = data3)
##
## $group
##          diff          lwr          upr      p adj
## 2-1 -1.40965768 -2.0459619 -0.7733534 0.0000001
## 3-1 -0.90196859 -1.5382728 -0.2656644 0.0012505
## 4-1 -0.94007992 -1.5763841 -0.3037757 0.0006701
## 5-1  0.40700216 -0.2293021  1.0433064 0.3982636
## 3-2  0.50768908 -0.1286151  1.1439933 0.1847262
## 4-2  0.46957776 -0.1667265  1.1058820 0.2539748
## 5-2  1.81665984  1.1803556  2.4529641 0.0000000
## 4-3 -0.03811132 -0.6744156  0.5981929 0.9998317
## 5-3  1.30897076  0.6726665  1.9452750 0.0000006
## 5-4  1.34708208  0.7107779  1.9833863 0.0000003
```

Para facilitar a análise, o seguinte gráfico foi plotado.

```
#plot teste Tukey
plot(tukey, las=1, col="brown")
```



## 95% family-wise confidence level



Pela análise do teste de Tukey, pode-se fazer algumas observações:

- A ação 1 possui média maior que as ações 2, 3 e 4 e média estatisticamente semelhante a ação 5;
- A ação 2 possui média estatisticamente igual às ações 3 e 4;
- A ação 5 possui média maior que as ações 2, 3 e 4 e média estatisticamente semelhante a ação 1;

Portanto, pode-se concluir que, estatisticamente, as ações 1 e 5 possuem médias semelhantes entre si e maiores do que as outras três ações.

Com isso, percebe-se que, através de testes estatísticos, as ações 1 e 5 seriam boas escolhas para o investidor. No entanto, a proposta é que seja escolhida apenas uma e, portanto, outro parâmetro de análise deve ser utilizado para a tomada de decisão entre as duas ações em questão.

Para escolher entre as duas ações foi levado em consideração a tendência baseada no valor de fechamento mensal histórico de cada uma delas. O gráfico abaixo apresenta os valores de fechamento (pontos) de cada uma das duas ações em análise e a linha de tendência (linha azul) de cada uma.

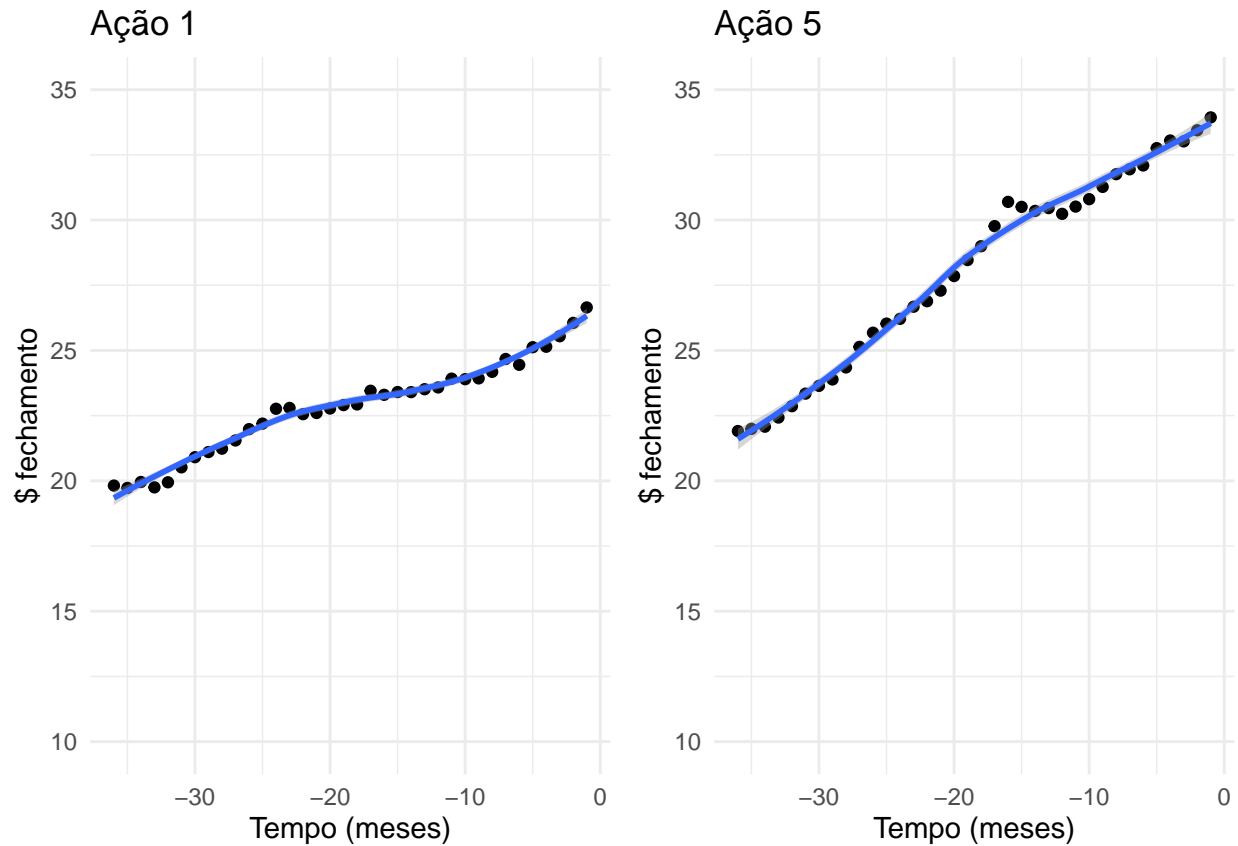
```
#Plot

plot_a1 <- ggplot(data=data, aes(x=numeracao, y=A1)) + geom_point() + geom_smooth() +
  labs(title = "Ação 1", x = "Tempo (meses)", y = "$ fechamento") + ylim(10, 35)

plot_a5 <- ggplot(data=data, aes(x=numeracao, y=A5)) + geom_point() + geom_smooth() +
  labs(title = "Ação 5", x = "Tempo (meses)", y = "$ fechamento") + ylim(10, 35)

grid.arrange(plot_a1, plot_a5, nrow = 1)
```

```
## 'geom_smooth()' using method = 'loess' and formula 'y ~ x'
## 'geom_smooth()' using method = 'loess' and formula 'y ~ x'
```



Com base na observação dos gráficos, infere-se que ambas ações apresentam uma tendência de alta. Porém, a inclinação da linha de tendência da ação 5 é maior quando comparada à ação 1. Isto sugere que, historicamente, a ação 5 apresentou maiores altas nos seus fechamentos quando comparada à ação 1. Portanto, embora ambas apresentem uma tendência de alta, a ação 5 possui um potencial de maior alta, já que sua tendência de alta é maior. Assim, se o investimento busca a maior rentabilidade, esta seria a melhor escolha.

Porém, também é interessante que se faça a análise do risco apresentado por cada uma das duas ações. Para isso, será analisado qual das duas ações apresentou a maior queda ao longo dos 36 meses. Portanto, abaixo são mostrados os valores mínimos das flutuações mensais para as duas ações.

```
min_a1 = min(data_a1)
min_a5 = min(data_a5)
data_min1 <- 1:2
data_min1[1] <- min_a1
data_min1[2] <- min_a5

data_min <- matrix(data_min1,1,2)
colnames(data_min) <- c("A1","A5") # Adicionando nomes às colunas
rownames(data_min) <- c("Minimo")

head(data_min)
```

```
##           A1           A5
## Minimo -1.04843 -0.725655
```

Através da análise dos mínimos, aplicada aos dados de flutuação percentual mensal de cada uma das ações, pode-se inferir que ao longo dos 36 meses a ação 1 apresentou a maior flutuação negativa. Portanto, ao comparar-se as duas, o investidor correrá um risco de maior queda caso opte pela ação 1. Desta maneira, do ponto de vista de risco, também é mais recomendável que o investimento seja realizado na ação 5.

## **Discussão e Conclusão**

O estudo teve como objetivo definir entre cinco ações previamente selecionadas, qual apresenta o maior potencial de ganho mensal para um investidor que deseja investir todo seu capital em uma única ação da bolsa de valores. Foram fornecidos os dados de fechamento dos preços mensais das cinco ações ao longo dos últimos 36 meses.

Para a análise ser realizada, os dados fornecidos foram analisados sob a ótica das flutuações percentuais das ações em cada mês, o que possibilitou por exemplo, determinar quais ações apresentam a maior média nas flutuações dos preços mensais.

Escolheu-se o teste ANOVA para determinar se havia diferença significativa entre a médias das flutuações das ações analisadas. Para utilização da ANOVA foi necessário verificar a independência entre as amostras, a normalidade na distribuição dos dados e, também, a homocedasticidade. Para verificar a independência, como não havia informações sobre as empresas, assumiu-se que elas não possuíam interferência entre si. Para verificação de normalidade, utilizou-se o teste Shapiro-Wilk e para verificação da homocedasticidade utilizou-se os testes de Bartlett e Levene, todos foram apresentados neste relatório.

Após a verificação dos requisitos, aplicou-se a ANOVA e através de seu resultado foi possível inferir que pelo menos uma das amostras possuía sua média diferente das outras. Porém, ainda não haviam dados suficientes para determinar o que foi solicitado. Escolheu-se então o Teste de Tukey para determinar quais ações entre as cinco analisadas, possuíam diferenças significativas entre suas médias.

Através do Teste de Turkey foi possível concluir que as ações 1 e 5 se destacavam das demais, do ponto de vista de rentabilidade, apresentando as melhores flutuações nos preços dos fechamentos mensais. A partir desta conclusão, a análise prosseguiu apenas para as ações 1 e 5. Para escolher a melhor entre as duas, analisou-se as tendências de cada uma delas e pôde-se perceber que, embora ambas apresentassem tendência de alta, a ação 5 apresentou a maior tendência de alta entre as duas e, portanto, do ponto de vista de rentabilidade ela foi a mais recomendada. Optou-se ainda por realizar uma última análise, levando em consideração o risco envolvido em cada uma das opções, para isso analisou-se os valores mínimos das flutuações nos preços de fechamento mensal das duas ações e observou-se que a ação 1 apresentou a maior queda percentual ao longo dos 36 meses. Portanto, considerando o risco, a ação 5 também foi a mais recomendada.

Desta forma, concluiu-se que para os dados fornecidos e através das análises realizadas e apresentadas, a ação 5 é a mais recomendada para o investidor que deseja investir todo seu capital em uma única ação da bolsa.

## **Atividades dos membros**

### **Diego**

Elaboração dos testes e construção textual.

### **Elias**

Elaboração textual, análise dos resultados e tomada de decisão.

### **Matheus**

Elaboração do experimento e estruturação dos códigos.

### **Todos**

Elaboração das hipóteses e definição das premissas.

## Referências Bibliográficas

- [1] Entendendo análise de variância (anova) e o teste f. <https://blog.minitab.com/pt/entendendo-analise-de-variancia-anova-e-o-teste-f>. Acesso em 7 de Fevereiro de 2021.
- [2] "Amanda C. Reiter, Ana M. Barreto, Clara R. Pires, Daniel Gonçalves-Souza, Danilo S.S. Lima, and Rafaela Lampa". "Tutorial ANOVA". [https://rstudio-pubs-static.s3.amazonaws.com/544630\\_7849ef9441ae454e831295b7e2de2d52.html](https://rstudio-pubs-static.s3.amazonaws.com/544630_7849ef9441ae454e831295b7e2de2d52.html), 2019. Acesso em 7 de Fevereiro de 2021.
- [3] Como realizar teste de normalidade no r? <https://rpubs.com/paternogbc/46768>. Acesso em 18 de Janeiro de 2021.
- [4] Teste de igualdade das variâncias. <http://www.portalaction.com.br/anova/161-teste-de-igualdade-das-variancias>. Acesso em 8 de Fevereiro de 2021.
- [5] Teste de tukey para comparações múltiplas. <https://operdata.com.br/blog/comparacoes-multiplas-teste-de-tukey/>, 2019. Acesso em 12 de Fevereiro de 2021.