

# Estudo de Caso 2 - Avaliação e comparação do retorno médio de ações

Ana Júlia Martins, Antônio Carlos da Anunciação, Melchior Augusto Syrio de Melo

25 de novembro de 2024

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##   select

##
## Attaching package: 'TH.data'

## The following object is masked from 'package:MASS':
##
##   geyser
```

## Resumo

This document provides a template for the Case Study reports. Reports should always start a short executive summary (call it an *Abstract* if you want) to give the reader a general idea of the topic under investigation, the kind of analysis performed, the results obtained, and the general recommendations of the authors.

## 1. Design do Experimento

Para realizar a análise estatística, estabelecemos as seguintes hipóteses:

- **Hipótese Nula ( $H_0$ ):** Não existe diferença no valor médio dos retornos das ações, isto é, não existe diferença no tamanho dos efeitos  $\tau_i$ .

$$\left\{ H_0 : \tau_i = 0, \forall i \in \{1, 2, \dots, a\} \right.$$

- **Hipótese Alternativa ( $H_a$ ):** Existe ao menos uma ação com um efeito significativamente diferente de zero, ou seja, com um retorno médio mensal superior às demais.

$$\left\{ H_1 : \exists \tau_i \neq 0 \right.$$

Para testar a hipótese definida, será utilizado o teste ANOVA. Se decidirmos pela rejeição da hipótese nula e as premissas do teste ANOVA forem validadas, precisaremos determinar quais ações possuem retorno significativamente maior do que as demais. Para isso, vamos realizar comparações múltiplas para identificar quais pares de ação diferem significativamente, utilizando o método de Tukey.

## 2. Descrição do conjunto de dados

Os dados utilizados neste estudo são referentes aos preços de fechamento de cinco ações, extraídos de um arquivo CSV. Cada arquivo contém 36 linhas e 5 colunas, onde:

- Cada linha representa o preço de fechamento mensal das ações, com a linha 1 correspondendo ao mês mais recente e a linha 36 ao mês mais distante.
- Cada coluna representa uma das cinco ações analisadas.
- O conteúdo da posição (i, j) refere-se ao preço de fechamento da ação j no mês i.

Com base nesse conjunto de dados, foi calculado o retorno mensal de cada ação utilizando a fórmula:

$$\text{Retorno} = \frac{\text{Preço no mês atual} - \text{Preço no mês anterior}}{\text{Preço no mês anterior}}$$

O primeiro mês de cada coluna foi descartado, pois não há um valor anterior para o cálculo do retorno. O processo resultou em uma tabela com os retornos mensais das cinco ações.

```
# Ler o arquivo CSV com os preços de fechamento das ações
precos <- read.csv("DadosAcoesGrupoH.csv", header=FALSE)
colnames(precos) <- paste0("Acao_", 1:5)

# Calcular o retorno mensal de cada ação
# Retorno mensal = (Preço atual - Preço anterior) / Preço anterior
retornos <- precos
for (j in 1:ncol(precos)) {
  for (i in 2:(nrow(precos))) {
    retornos[i-1, j] <- ((precos[i-1, j] - precos[i, j])) / precos[i, j]
  }
}

# Remover a ultima linha, porque a primeira linha da tabela de preços foi descartada
retornos <- retornos[-36, ]

retornos <- retornos[nrow(retornos):1, ]
```

### 3. Análise Exploratória

Antes de realizar os testes de hipótese, foi realizada uma análise exploratória para obter uma visão geral dos dados. A Figura @ref{fig:boxplot} fornece um boxplot para comparar as distribuições dos retornos das ações.

```
# Transforma os dados para visualização
retornos_melt <- melt(retornos, variable.name = "Acoes", value.name = "Retorno")

# Boxplot dos retornos
ggplot(retornos_melt,
       aes(x = Acoes, y = Retorno, fill = Acoes)) +
  geom_boxplot() +
  geom_point(alpha = 0.5) +
  ggtitle("Retorno das Ações",
         "(dados originais + boxplots)") +
  theme(legend.position = "none")
```

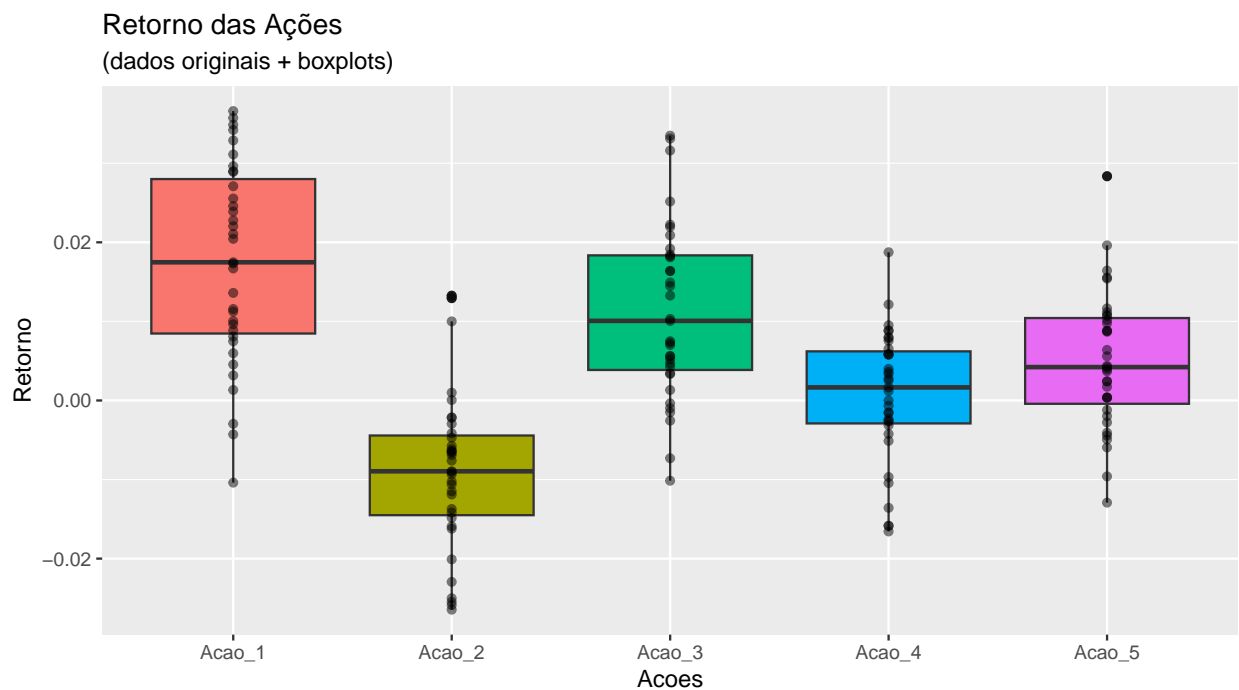


Figure 1: Retorno das Ações (dados originais + boxplots)

A análise gráfica indica que a Ação 2 aparenta apresentar o maior retorno médio entre as ações analisadas.

### 4. Análise Estatística

Para validar nossas observações iniciais, realizamos um teste ANOVA para avaliar as diferenças nos retornos das ações.

```
retornos_melt$Acoes <- as.factor(retornos_melt$Acoes)
modelo <- aov(Retorno ~ Acoes, data = retornos_melt)
summary(model)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Acoes      4 0.01434 0.003586   34.45 <2e-16 ***
## Residuals 170 0.01769 0.000104
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Como resultado do teste, obteve-se uma estatística  $F = 34,66$  e um p-valor  $< 2e-16$ .

Diante do resultado do teste ANOVA, podemos rejeitar, com nível de significância  $\alpha = 0,05$ , a hipótese nula de que as médias de retorno das cinco ações são iguais.

Esse resultado indica que nem todas as ações têm retornos médios semelhantes e, para identificar quais ações possuem maior retorno médio, realizamos um teste de Tukey. Os resultados do teste de Turkey são mostrados na Figura @ref{fig:turkey}.

```
# Comparações múltiplas
library(multcomp)

mc1 <- glht(model,
             linfct = mcp(Acoes = "Tukey"))
#summary(mc1)
mc1_CI <- confint(mc1, level = 0.95)

par(mar = c(5, 10, 4, 2)) # margens
plot(mc1_CI,
     xlab = "Retornos",
     cex.axis = 1.2,
     cex = 2,
     main = "Comparações Múltiplas - Tukey")
```

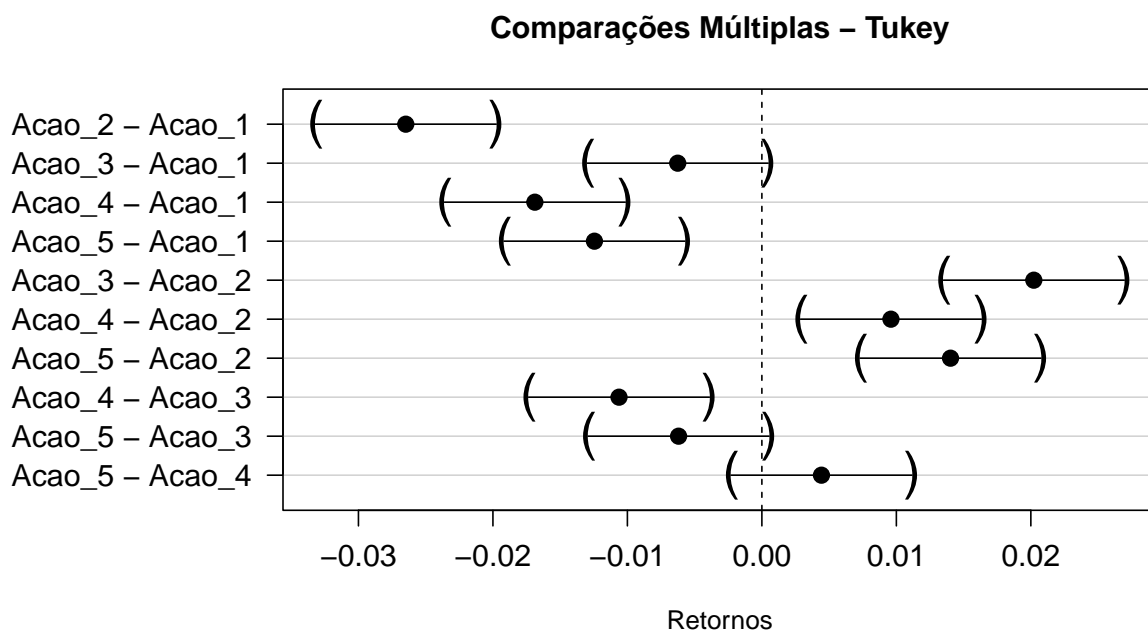


Figure 2: Comparações múltiplas - Turkey

As comparações mais significativas são entre a Ação 2 e as Ações 1, 3, 4 e 5, todas com p-valor  $< 0,001$ , indicando que essas ações têm médias de retorno significativamente diferentes. As Ações 1, 3, 4 e 5, por sua vez, apresentam pelo menos uma diferença não significativa quando comparadas entre si.

Confirmando a suspeita inicial, o teste de Tukey indica que a Ação 2 tem o maior retorno médio entre as ações analisadas, visto que as comparações entre a Ação 2 e as outras ações todas mostram diferenças significativas.

## 5. Verificação das Premissas do Modelo

The assumptions of your test should also be validated, and possible effects of violations should also be explored.

```
par(mfrow=c(2,2), mai=.3*c(1,1,1,1))
plot(model,pch=16,lty=1,lwd=2)
```

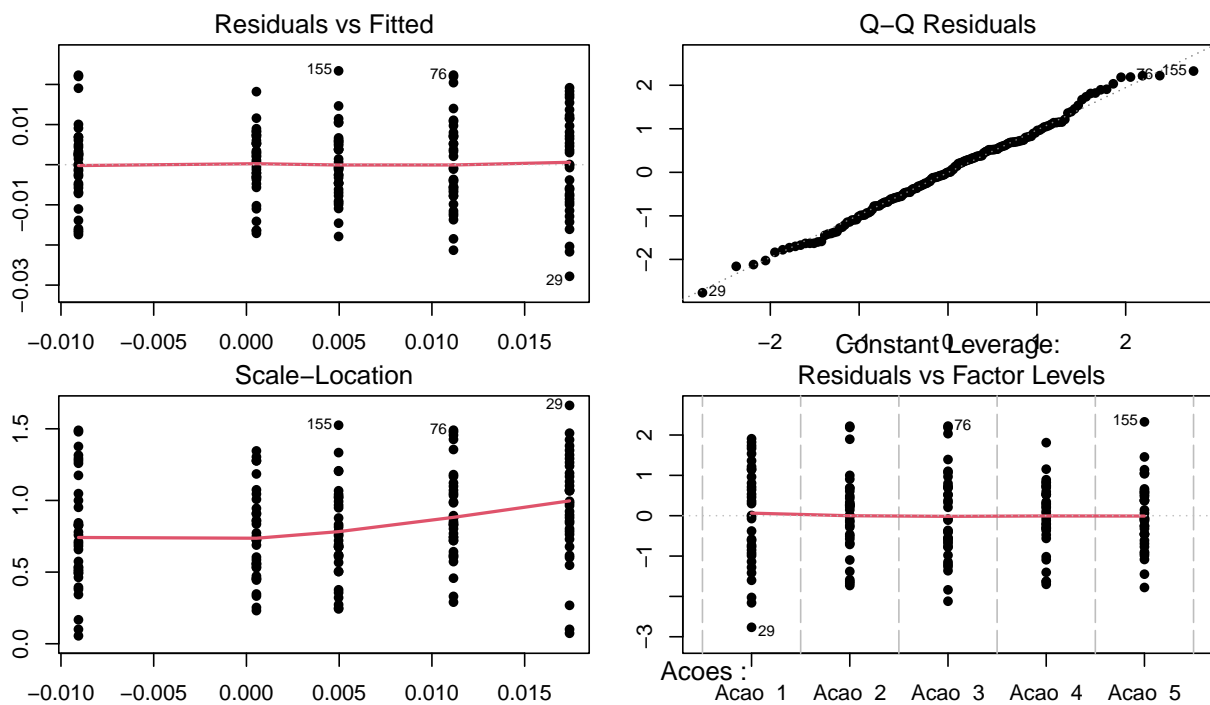


Figure 3: Residual plots for the anova model

## 6. Conclusões e Recomendações

The discussion of your results, and the scientific/technical meaning of the effects detected, should be placed here. Always be sure to tie your results back to the original question of interest!

## 7. Atividades Específicas

- Ana Julia: Design do experimento, análise estatística, análise exploratória e descrição do conjunto de dados (redação e código)
- Antônio:
- Melchior: