

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 mês 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 possível realizar uma análise estatística entre as 5 ações e identificar a que apresenta maior incidência 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*.

```
# 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)
```

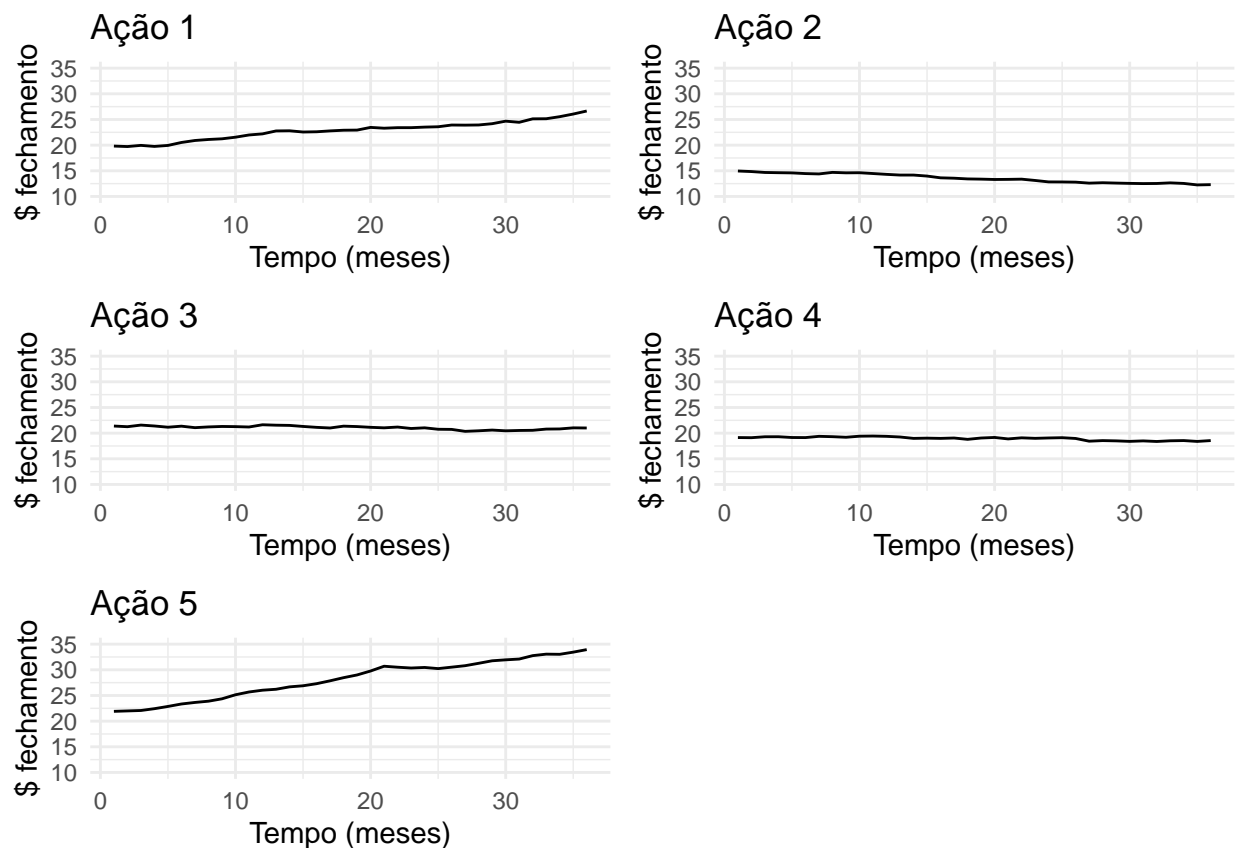
```
##      A1      A2      A3      A4      A5
## 1 26.649 12.302 21.006 18.553 33.937
## 2 26.057 12.250 21.032 18.385 33.443
## 3 25.545 12.541 20.816 18.555 33.018
## 4 25.143 12.649 20.790 18.513 33.050
## 5 25.124 12.531 20.553 18.376 32.758
## 6 24.451 12.511 20.520 18.490 32.094
```

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 = "Ação 1", x = "Tempo (meses)", y = "$ fechamento") + ylim(10, 35)
plot_a2 <- ggplot(data=data, aes(x=nrow(data):1, y=A2)) + geom_line() +
  labs(title = "Ação 2", x = "Tempo (meses)", y = "$ fechamento") + ylim(10, 35)
plot_a3 <- ggplot(data=data, aes(x=nrow(data):1, y=A3)) + geom_line() +
  labs(title = "Ação 3", x = "Tempo (meses)", y = "$ fechamento") + ylim(10, 35)
plot_a4 <- ggplot(data=data, aes(x=nrow(data):1, y=A4)) + geom_line() +
  labs(title = "Ação 4", x = "Tempo (meses)", y = "$ fechamento") + ylim(10, 35)
plot_a5 <- ggplot(data=data, aes(x=nrow(data):1, 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.
}

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

```

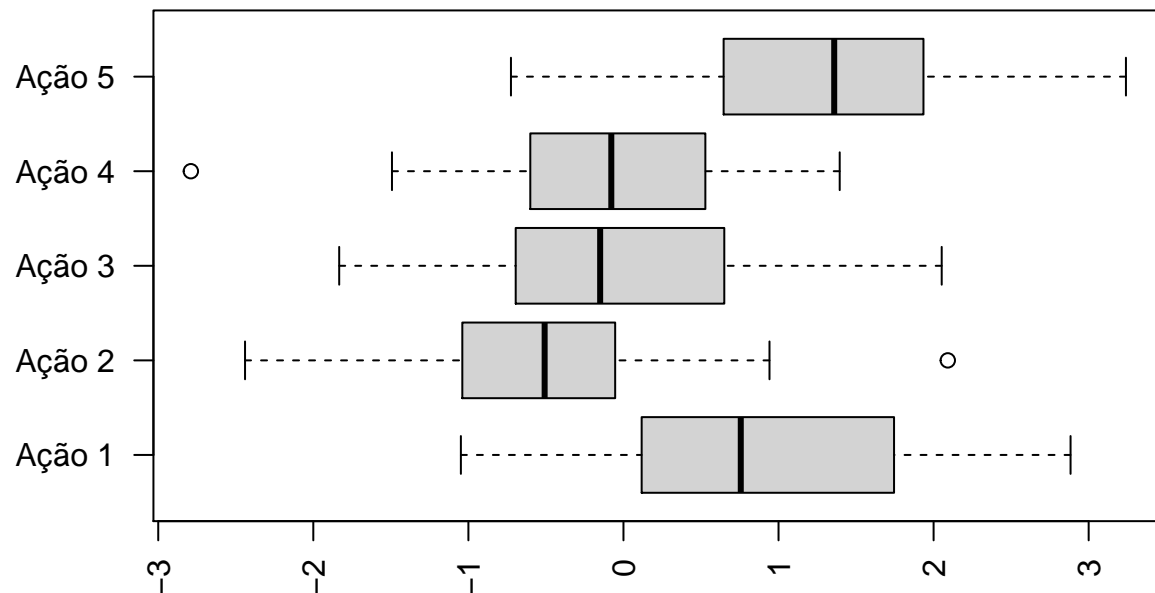
algum comentario aqui

```

# 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
)

```

Boxplots das flutuações percentuais



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]

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:

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

```
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.

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

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

```
# 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 amostras não possuem equivalência 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.

Outro teste apenas para comprovar...

```
# 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, temos que se o p-valor < 0.05 (α), 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, podemos considerar que as variâncias são homogêneas.

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

$$\begin{cases} H_0 : \text{A média de todas as populações são iguais} \\ H_1 : \text{A média de todas as populações não são iguais} \end{cases}$$

```
# 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 (α), 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 iguais. Ou seja, pelo menos uma das amostras tem sua média diferente. Para dar sequência aos estudos é necessário fazer uma análise por pares para encontrar-se... *dar explicação*

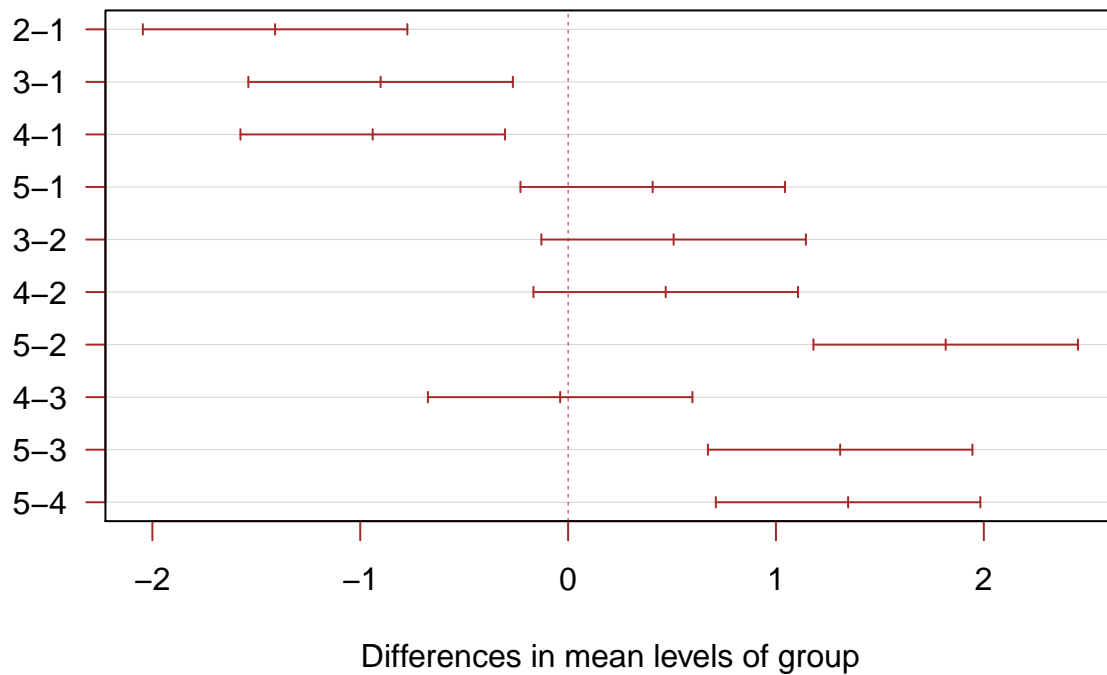
```
# teste Tukey
library(agricolae)
data3 = data.frame(y,group)
model2 <- aov(y~group, data = data3)
#tukey = HSD.test(model2, "group")
tukey = TukeyHSD(x=model2, "group")

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

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

95% family-wise confidence level



```
#library(MASS)
#data("crabs")
#crabs$groups.4 <- with(crabs, paste(sp,sex,sep = "."))
#crab.aov <- aov(FL ~ groups.4, data = crabs)
#crab.Tukey <- TukeyHSD(crab.aov)
#plot(crab.Tukey)

#Gerar o gráfico
#library(tibble)
#tukey$groups %>%
#  rownames_to_column(var = "trt") %>%
#  mutate(trt = reorder(trt, -y, mean)) %>%
#  ggplot(aes(trt, y)) +
#    geom_col(alpha = 0.8, color = "black") +
#    geom_text(aes(label = groups), vjust = 1.8, size = 9, color = "white") +
#    labs(x = "Ação", y = "Médias") +
#    theme_bw(16)
```

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.