

Planejamento e Análise de Experimentos (EEE933)

Estudo de Caso 3

Pedro Vinícius, Samara Silva e Savio Vieira

14 de Setembro de 2020

Introdução

O mercado de ações é uma forma de investimento onde uma das maneiras de se obter lucratividade é vender as ações quando seus preços atuais estiverem acima do preço no qual foram adquiridas [8]. Embora a rentabilidade passada não seja garantia de rentabilidade futura, normalmente o histórico de rentabilidade é um fator a ser considerado [7].

Neste estudo de caso deseja-se verificar, qual ação dentre cinco candidatas possui maior potencial de **ganho mensal** ao investidor quando escolhida com exclusividade, isto é, quando o montante é investido por completo em uma única ação. Para isso, considera-se o fechamento das cinco ações nos últimos 36 meses.

Planejamento dos Experimentos

Dado que o investidor deseja a maior rentabilidade mensal possível, a série temporal dos preços de fechamento das ações foram utilizadas para calcular a taxa de retorno das cinco ações em cada um dos meses. Tal índice permite avaliar o lucro de uma ação durante um mês, expresso como uma proporção do investimento no mês anterior. A taxa de retorno de uma ação i em um determinado mês j é dada por (1):

$$r_j^i = \frac{p_j^i - p_{j-1}^i}{p_{j-1}^i} \quad (1)$$

onde p_j^i é o preço de fechamento da ação i no mês atual j e p_{j-1}^i é o preço de fechamento da ação i no mês anterior $j - 1$.

Quando o preço de fechamento da ação no mês corrente é superior ao preço de fechamento da ação no mês anterior ($p_j^i > p_{j-1}^i$), o retorno expressará o ganho líquido obtido do mês $j - 1$ para o mês j . Em contrapartida, se o preço de fechamento da ação no mês corrente é inferior ao preço de fechamento da ação no mês anterior ($p_j^i < p_{j-1}^i$), o retorno expressará a perda líquida obtida nessa mesma transição mensal. Uma vez que o presente cálculo considera diferenças entre cotações adjacentes, o tamanho amostral é reduzido em uma unidade ($n = 35$).

As hipóteses estatísticas foram definidas com objetivo de verificar as seguintes proposições:

- Dentre o grupo de ações candidatas, qual delas possui maior potencial de rentabilidade mensal para o investidor?
- Caso haja mais de uma ação em destaque, quais são elas e qual deveria ser escolhida entre elas?

Considerando as questões propostas, foram estabelecidas as hipóteses de teste sobre o retorno médio mensal das ações:

$$\begin{cases} H_0 : \mu_{A_1} = \mu_{A_2} = \mu_{A_3} = \mu_{A_4} = \mu_{A_5} \\ H_1 : \text{duas ou mais médias são diferentes das demais} \end{cases}$$

onde a hipótese nula H_0 implica na igualdade entre os retornos médios mensais das cinco ações ao nível de confiança $1 - \alpha$ e a hipótese alternativa bilateral H_1 na diferença entre duas ou mais ações em relação ao retorno médio mensal também ao nível de confiança $1 - \alpha$.

Os parâmetros experimentais considerados para realização dos testes são nível de significância $\alpha = 0,05$ e potência desejada do teste $\pi^* = 1 - \beta = 0,80$.

Análise Exploratória de Dados

Algumas primeiras propriedades amostrais referentes ao retorno mensal das cinco ações, como média, moda, mediana, valores extremos, variância e desvio podem ser obtidas de imediato. A fim de facilitar a interpretabilidade dessas estatísticas amostrais, os retornos médios mensais são apresentados em porcentagem, ou seja, os valores originais obtidos pela Equação (1) foram multiplicados por 100.

##	Variância	Média	Moda	Mediana	Mínimo	Máximo	Desvio
## A1	0.012973071	-0.7592262	-1.3568984	-0.6938616	-3.5420789	1.521483	1.1389939
## A2	0.012600939	1.0617409	1.6718373	1.1192959	-1.4696377	4.245581	1.1225390
## A3	0.011422153	-1.1738187	-1.7089391	-1.3162823	-3.7987805	1.852466	1.0687447
## A4	0.008625187	1.1996236	0.1444583	1.0788613	-0.2639517	3.038714	0.9287189
## A5	0.012415407	0.3024411	0.7222587	0.3571470	-1.6374871	2.895227	1.1142444

A princípio, é possível observar que, durante os últimos 3 anos, os retornos mensais de duas dentre as cinco ações candidatas apresentaram, em média, retorno negativo, são elas A_1 e A_3 . No entanto, é interessante ressaltar que todas as ações tiveram retornos mensais negativos em algum momento da série histórica. O maior retorno negativo constatado foi o da ação A_3 com uma perda de $-3,79\%$ no 20º mês e o menor retorno negativo foi $-0,26\%$ da ação A_4 no 17º mês. A ação A_4 obteve o maior ganho líquido médio, cerca de $1,20\%$ e a ação A_2 apresentou o maior retorno máximo mensal, $4,24\%$ no 16º mês. O pior rendimento máximo foi de $1,85\%$ pela ação A_3 no 9º mês.

Um fator interessante é que as diferenças absolutas entre os retornos amostrais mensais médios e os retornos amostrais mensais medianos de todas as ações foram bem próximas de zero, isto é, $\Delta_i = |\bar{\mu}_i - \bar{m}_i| \approx 0$ para $i \in \{1, 5\}$. A maior diferença obtida, por exemplo, foi de $0,14\%$ para a ação A_3 . Apesar de ser uma suposição fracamente sustentada, isso sugere que as distribuições de probabilidade amostrais podem ser aproximadas, a priori, por uma distribuição normal. No que se refere às variâncias amostrais das ações, percebe-se que os valores são bastante similares e aproximadamente iguais a $0,01\%$, o que também sugere que as amostras são homocedásticas.

Com o intuito de compreender melhor os dados em estudo e, posteriormente, inferir sobre as populações de onde as amostras provêm, serão analisadas algumas representações gráficas. No que tange a distribuição de frequência das observações, com exceção da ação A_4 , todas as demais ações apresentam curvas em formato de sino, o que fortalece o pressuposto de normalidade mencionado anteriormente.

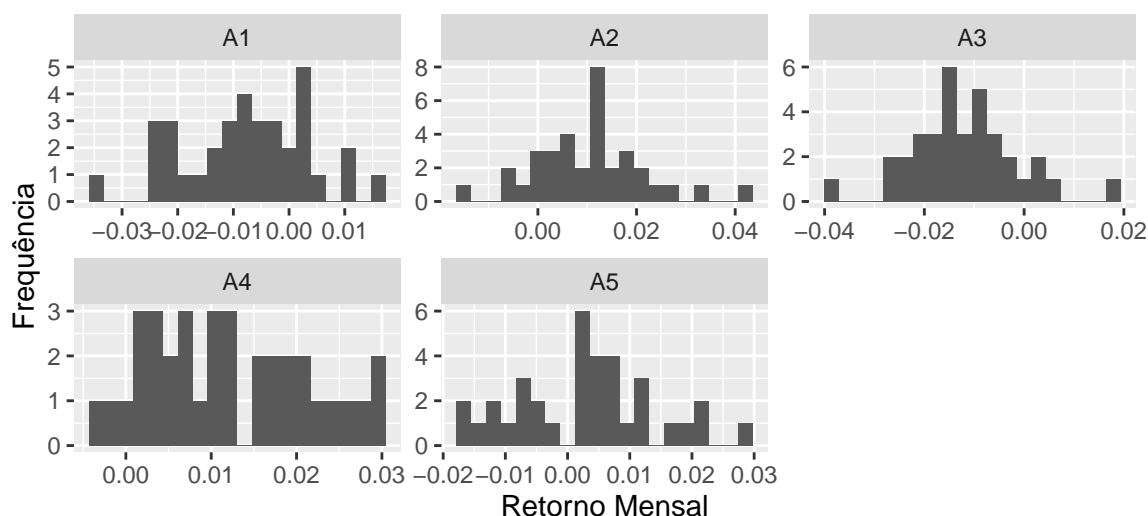


Figura 1: Histogramas.

Ainda a partir dos histogramas, pode-se evidenciar que durante os 36 meses coletados, as ações A_1 e A_3 tiveram a maioria dos rendimentos mensais no domínio negativo (perda líquida) e as ações A_2 , A_4 e A_5 apresentaram a maioria dos rendimentos mensais no domínio positivo (ganho líquido).

Segundo os diagramas de caixa, apenas a distribuição de retornos mensais da amostra A_1 é visualmente simétrica e, portanto apresenta evidências de normalidade. No caso das demais ações, as medianas não estão tão próximas da média, uma vez que os segundos quartis estão deslocados dos respectivos centros das caixas. Essa dubiedade de interpretações entre os histogramas e os *boxplots* acerca da normalidade dos dados instiga análises mais aprofundadas, como serão realizadas a seguir. Além dessas verificações, outra circunstância notável é a presença de *outliers* apenas nas distribuições amostrais das ações A_2 e A_3 .

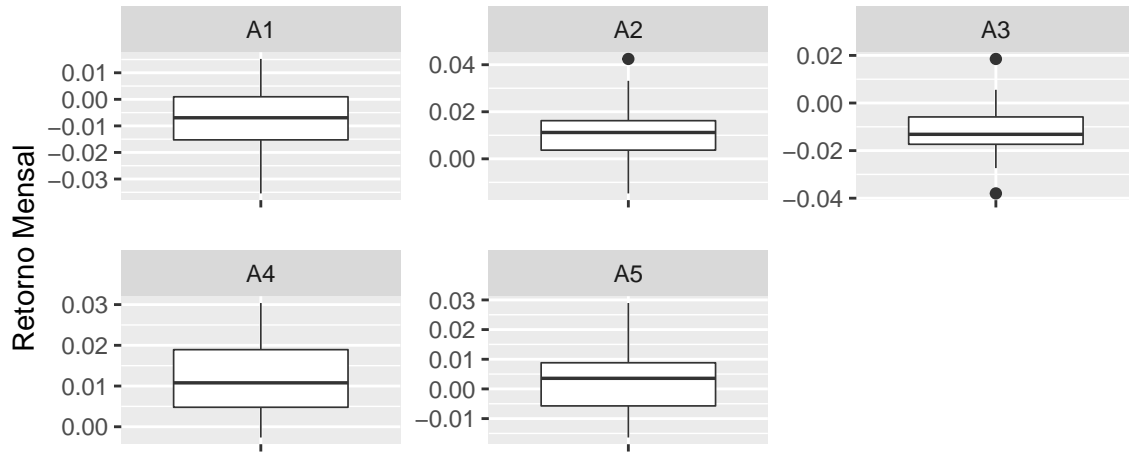


Figura 2: Boxplots.

Outra possibilidade para examinar a normalidade é considerar gráficos quantil-quantil. Essa representação permite comparar as distribuições de probabilidade de cada uma das ações (eixo das ordenadas) com uma distribuição normal (eixo das abcissas). Mediante a boa qualidade do ajuste da reta nos pontos para todas as ações, pode-se concluir que há fortes indícios de que os resíduos das distribuições são normais.

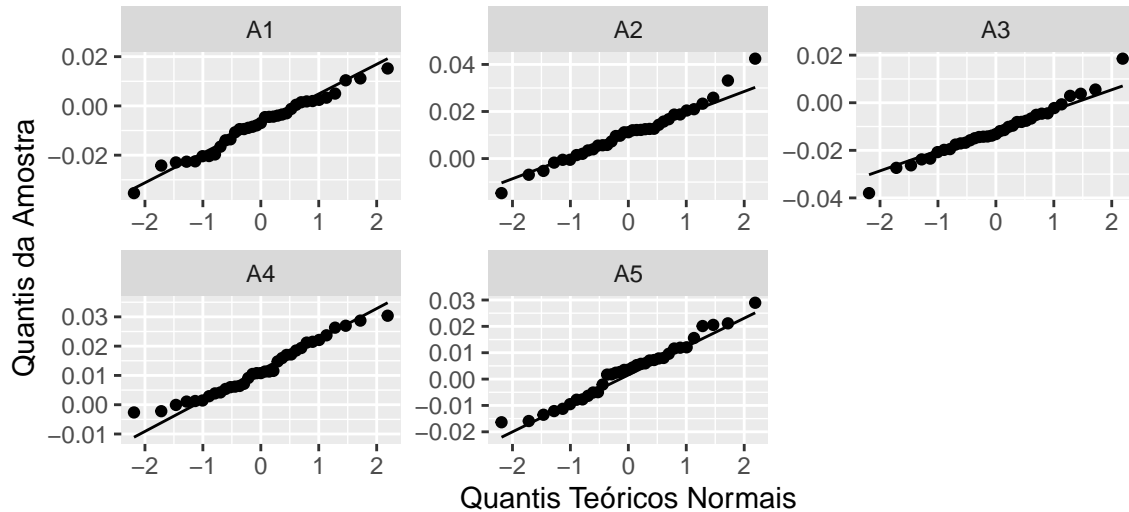


Figura 3: QQ-Plots.

Validação de Premissas

A premissa de normalidade será validada pelo teste não-paramétrico de Shapiro-Wilk. A hipótese nula desse teste afirma que a amostra proveio de uma população com distribuição normal e a hipótese alternativa é de que a amostra não proveio de uma distribuição normal [12]. O nível de significância utilizado para cada teste foi de $\alpha = 0,10$.

```
##      statistic p.value  method      data.name
## [1,] 0.9805968 0.7784242 "Shapiro-Wilk normality test" "sample$A1"
## [2,] 0.9770649 0.6619515 "Shapiro-Wilk normality test" "sample$A2"
## [3,] 0.9808721 0.7871903 "Shapiro-Wilk normality test" "sample$A3"
## [4,] 0.960207  0.2319528 "Shapiro-Wilk normality test" "sample$A4"
## [5,] 0.9753498 0.6053638 "Shapiro-Wilk normality test" "sample$A5"
```

Não foi possível rejeitar a hipótese nula em nenhum dos casos, visto que os p-valores obtidos são maiores que o nível de significância pré-estabelecido ($p > \alpha$). Assim, pode-se afirmar ao nível de confiança de 90% que as amostras de todas as ações provêm de populações com distribuição normal. Como a premissa de normalidade foi validada, não é necessário evocar o Teorema Central do Limite (TCL), que ainda assim seria viável devido o tamanho amostral ser suficiente ($n = 35$) > 30 .

Alguns testes estatísticos de múltiplas amostras, cujas hipóteses nulas são construídas a partir do parâmetro média amostral, são altamente sensíveis a premissa de independência das amostras, como é o caso do ANOVA. Embora não exista um procedimento específico para testar a independência das amostras no caso geral, o caso especial de autocorrelações em série nos dados pode ser testado pelo Teste de Durbin-Watson [4]. A hipótese nula desse teste declara que não há correlação entre os resíduos, isto é, os resíduos são independentes. A hipótese alternativa, por sua vez, afirma que os resíduos são correlacionados [14]. Assim, a correlação entre os resíduos foi tomada segundo a abordagem de todos contra todos, cujo intuito é avaliar as correlações entre todas as 10 possíveis combinações das ações par a par ($A_1 \times A_2$, $A_1 \times A_3$, $A_1 \times A_4$, $A_1 \times A_5$, $A_2 \times A_3$, $A_2 \times A_4$, $A_2 \times A_5$, $A_3 \times A_4$, $A_3 \times A_5$ e $A_4 \times A_5$). O nível de significância utilizado para esse teste foi $\alpha = 0,05$.

```
##      data.name      method  p.value statistic
## 1 sample$A1 ~ sample$A2 Durbin-Watson test 0.4645289  1.954171
## 2 sample$A1 ~ sample$A3 Durbin-Watson test 0.4627024  1.976153
## 3 sample$A1 ~ sample$A4 Durbin-Watson test  0.382042  1.911501
## 4 sample$A1 ~ sample$A5 Durbin-Watson test 0.5138884  2.011896
## 5 sample$A2 ~ sample$A3 Durbin-Watson test 0.8754208  2.39103
## 6 sample$A2 ~ sample$A4 Durbin-Watson test 0.9553317  2.568211
## 7 sample$A2 ~ sample$A5 Durbin-Watson test 0.9360123  2.507087
## 8 sample$A3 ~ sample$A4 Durbin-Watson test 0.2192108  1.754212
## 9 sample$A3 ~ sample$A5 Durbin-Watson test 0.2288992  1.749523
## 10 sample$A4 ~ sample$A5 Durbin-Watson test 0.3379215  1.858596
```

Não é possível rejeitar a hipótese nula ao nível de confiança estabelecido para nenhuma das combinações par a par das ações, uma vez que os p-valores foram todos superiores ao nível de significância. Assim, pode-se afirmar que os resíduos amostrais entre todos os pares de ações são independentes ao nível de confiança de 95%.

Com a finalidade de avaliar a premissa de homocedasticidade, o teste paramétrico de Bartlett foi empregado. Esse teste permite comparar as variâncias de três ou mais grupos e possui como premissa tanto a independência quanto a normalidade das amostras, que foram anteriormente validadas. A hipótese nula desse teste afirma que as variâncias de todas as populações das quais vieram as amostras são iguais e a hipótese alternativa afirma que há pelo menos uma diferença entre as variâncias dessas populações [11]. O nível de significância foi definido em $\alpha = 0,05$.

```
##
## Bartlett test of homogeneity of variances
##
## data:  sample
## Bartlett's K-squared = 1.7598, df = 4, p-value = 0.7798
```

A partir do p-valor substancialmente maior que o nível de significância ($0,7798 \gg 0,05$), não se pode refutar a hipótese nula e, portanto, ao nível de confiança de 95%, tem-se que as variâncias populacionais de todas as ações são iguais.

Análise Estatística

Para prosseguimento do estudo a respeito dos rendimentos das ações, foi proposto a comparação de médias utilizando-se o Anova. Apesar de ser considerado robusto às violações moderadas da normalidade e homoscedasticidade [5], foram provados através dos testes anteriores que as amostras preenchem esses requisitos. Quanto à independência, a ANOVA é sensível a essa violação, no entanto, essa premissa também foi testada e validada. Assim, pode-se utilizar esse teste para comparação das médias.

O teste proposto, também conhecido como *One Way*, tem como hipótese nula que as médias dos grupos em análise é a mesma, e como hipótese alternativa de pelo menos duas ou mais médias se distingue das demais. Dessa forma, este método verifica se as diferenças observadas na média são grandes o suficiente para ser o resultado da seleção aleatória, e pra isso além de considerar a diferença absoluta entre as amostras, também é considerado a variabilidade dentro de cada tratamento. Assim, verifica-se se a diferença entre os tratamentos é muito maior que a diferença no tratamento, e isso é possível fazendo a comparação da variância entre os grupos com a variância dentro do próprio grupo, o que dá nome ao teste [3]. A seguir, é possível ver o resultado do emprego deste teste, por meio da função `aov()` do pacote `stats` do R [1].

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## ind         4 0.01586 0.003966   34.17 <2e-16 ***
## Residuals   170 0.01973 0.000116
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

No experimento em questão, o p valor retornado foi $p = 2 \times 10^{-16}$, o que significa que com 95% de confiança, a hipótese nula de que as médias dos rendimentos são iguais, é rejeitada. Ou seja, foram detectadas diferenças significativas entre as medias dos rendimentos das ações em estudo.

Além do valor p, também é retornado $F = 34.17$, que é calculado a partir do quociente do quadrado médio entre os tratamentos (0,003966) pelo quadrado dos resíduos (0,000116). Nota-se que o numerador desta operação é muito maior que o denominador, que consequentemente resulta em um F grande. Nesse sentido, quanto maior essa proporção, maior a confiança em rejeitar a hipótese nula.

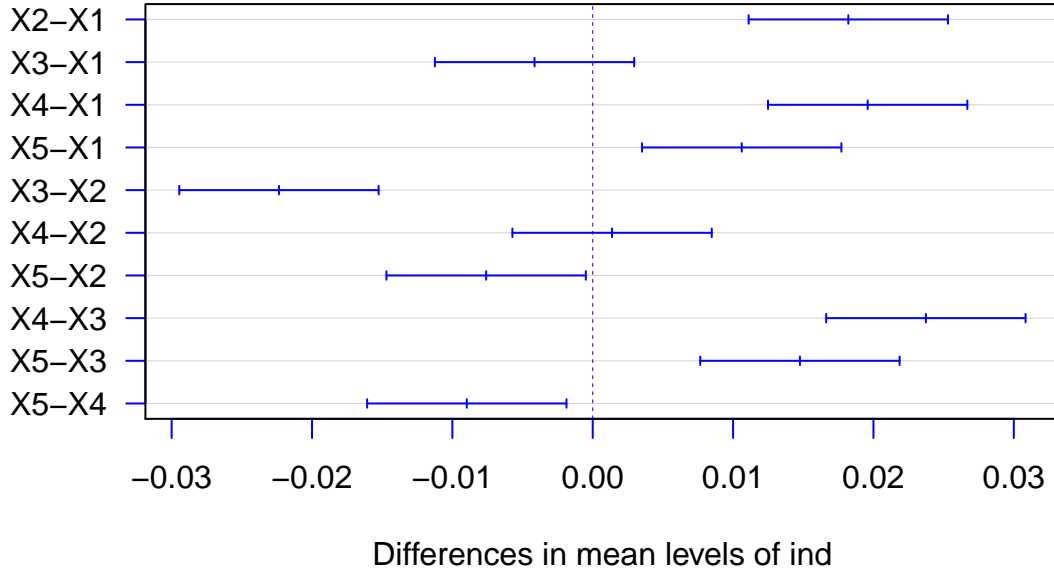
Todavia, embora o Anova seja capaz de identificar a existência dessa distinção entre médias, quando a hipótese nula é rejeita, é preciso identificar quais foram diferenças significativas, que não é uma das atribuições deste método. Para isso são utilizados ferramentas de comparações múltiplas ou ferramentas de correções de significância.

Para encontrar quais ações tem média de rendimento significativamente diferente será aplicado o teste de Tukey, uma ferramenta de comparações múltiplas. Este executa comparações do tipo todos com todos e demonstra ser interessante para este propósito, pois além de ser de fácil aplicação, é considerado um método poderoso. Em grupos amostrais balanceado ele é um teste exato, ou seja, a taxa de erro do grupo testado é exatamente o parâmetro α . Isso constitui uma grande vantagem do método, pois esse tipo de precisão é considerado raro dentre os testes para essa finalidade [10].

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = values ~ ind, data = grupos)
##
## $ind
##           diff           lwr           upr           p adj
## X2-X1 0.018209671 0.011108533 0.025310807 0.0000000
## X3-X1 -0.004145924 -0.011247062 0.0029552134 0.4932074
## X4-X1 0.019588498 0.012487361 0.0266896360 0.0000000
```

```
## X5-X1  0.010616674  0.003515536  0.0177178113  0.0005553
## X3-X2 -0.022355595 -0.029456733 -0.0152544576  0.0000000
## X4-X2  0.001378827 -0.005722310  0.0084799650  0.9835507
## X5-X2 -0.007592997 -0.014694135 -0.0004918598  0.0295782
## X4-X3  0.023734423  0.016633285  0.0308355603  0.0000000
## X5-X3  0.014762598  0.007661460  0.0218637355  0.0000004
## X5-X4 -0.008971825 -0.016072962 -0.0018706871  0.0056131
```

95% family-wise confidence level



Os dados extraídos do teste de Tukey são apresentados graficamente. Baseado no intervalo de confiança das diferenças das médias dos rendimentos, verifica-se que as ações ($X3$ e $X1$) e ($X4$ e $X2$) não são significativamente diferentes, pois estes dois intervalos contêm o zero (0). Já os demais intervalos ($X2$ e $X1$), ($X4$ e $X1$), ($X5$ e $X1$), ($X3$ e $X2$), ($X5$ e $X2$), ($X4$ e $X3$), ($X5$ e $X3$) e ($X5$ e $X4$) têm rendimentos significativamente diferentes, como mostrado graficamente.

Pela análise do p valor, tem-se a mesma conclusão, visto que os intervalos dos rendimentos das ações ($X3$ e $X1$) e ($X4$ e $X2$) retornou p valor maior que 0,05. Os demais p valores são menores que o nível de significância igual a 0,05, enfatizando que têm rendimentos são significativamente diferentes.

Tamanho de Efeito

A fim de investigar o tamanho de efeito para as diferenças de médias de três grupos ou mais, diversas medidas foram desenvolvidas, entre elas *Eta-Squared* (η^2), *Partial Eta-Squared* (η_p^2), *Omega-Squared* (ω^2) e *Epsilon-Squared* (ϵ^2). Dada a preferência de uso na literatura pela *Eta-Squared*, essa será a medida utilizada para estimar o tamanho de efeito no presente experimento [13]. Além disso, para o ANOVA de uma única via (*One Way*), *Eta-Squared* é equivalente ao *Partial Eta-Squared* e é dado por (2) [13]:

$$\eta^2 = \frac{SS_{Effect}}{SS_{Total}} \quad (2)$$

onde SS_{Total} é igual a somas dos quadrados totais e SS_{Effect} é a soma dos quadrados dos efeitos. O pacote `effectsize` do R permite calcular diretamente o η^2 a partir da tabela de resultados do ANOVA [9].

```
## Parameter | Eta2 (partial) |          90% CI
## -----
## ind       |          0.45 | [0.35, 0.52]
```

A estimativa pontual obtida para o tamanho de efeito foi $\eta^2 = 0,45$. Ao nível de confiança de 90%, tem-se ainda que o intervalo de confiança bilateral do tamanho de efeito é $[0,35; 0,52]$.

Uma das maneiras mais usuais de analisar o tamanho de efeito *Eta-Squared* é convertê-lo para a estimativa f de Cohen, cujos valores são tabelados segundo interpretações qualitativas [2]:

- Pequeno: $f \leq 0,10$;
- Médio: $0,10 < f \leq 0,25$;
- Grande: $f \geq 0,25$.

A relação entre η^2 e f é dada pela Equação (3) [13]:

$$\eta^2 = \frac{f^2}{1 + f^2} \quad (3)$$

Como a relação inversa é necessária, basta manipular algebricamente a Equação (3), de modo a obter a Equação (4):

$$f = \sqrt{\frac{\eta^2}{1 - \eta^2}} \quad (4)$$

Assim, tem-se uma estimativa pontual de tamanho de efeito grande e igual a $f = 0,4978$. Quanto ao intervalo de confiança bilateral ao nível de confiança de 90%, tem-se $f \in [0,3745; 0,5941]$. A relação entre os tamanhos de efeito das medidas f e η^2 no intervalo de confiança pode ser visualizada na Figura 4.

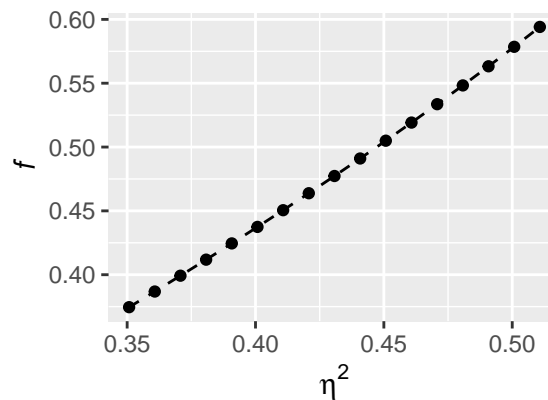


Figura 4: Relação entre o f de Cohen e o η^2 no intervalo de confiança.

Poder do Teste

Tendo em vista o número de grupos $k = 5$, o número de amostras em cada grupo $n = 35$, o nível de significância previamente estabelecido $\alpha = 0,05$ e o tamanho de efeito $f = 0,4978$, torna-se possível verificar se a potência desejada do teste foi alcançada. Para tal, foi utilizado o método `pwr.anova.test` do pacote `pwr` [6].

```
##
##      Balanced one-way analysis of variance power calculation
##
##      k = 5
##      n = 35
##      f = 0.4978336
```

```
##      sig.level = 0.05
##      power = 0.9999097
##
## NOTE: n is number in each group
```

O poder do teste foi bastante alto para o experimento conduzido ($\pi = 0,9999$), sendo substancialmente superior ao valor desejado ($\pi^* = 0,80$). Obter exatamente a potência desejada implicaria em detectar ganhos maiores ou iguais a $f = 0,2648$, isto é, ter um tamanho de efeito mais próximo do que é considerado um tamanho de efeito médio.

```
##
##      Balanced one-way analysis of variance power calculation
##
##      k = 5
##      n = 35
##      f = 0.2648026
##      sig.level = 0.05
##      power = 0.8
##
## NOTE: n is number in each group
```

Conclusões

Discussão de Melhorias

Atividades Desempenhadas

A hipótese de teste foi definida em concordância com os três autores, bem como a definição de conversão dos preços de fechamento mensais em taxas de retorno mensais para melhor condução do experimento. O pré-processamento dos dados, a análise exploratória e a validação das premissas foram conduzidas pelo Pedro. A análise estatística, por sua vez, que inclui tanto a verificação da existência de diferenças entre as médias das cinco ações quanto o teste de comparações múltiplas a posteriori, foi realizada pelo Savio e pela Samara. A estimativa do tamanho de efeito e o cálculo do poder de teste foram elaborados pelo Pedro. Por fim, a Samara e o Savio concluíram o trabalho com as considerações finais e discussões de melhoria.

Referências

- [1] AOV: Fit An Analysis Of Variance Model. <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/aov>, 2020. stats version v3.6.2.
- [2] Mohammad Al-Oudat, Miad Faezipour, and Ahmed El-Sayed. A Smart Intraocular Pressure Risk Assessment Framework Using Frontal Eye Image Analysis. *EURASIP Journal on Image and Video Processing*, 2018(1):90, 2018.
- [3] Stan Brown. Comparing More Than Two Means: One-Way ANOVA. <https://brownmath.com/stat/anova1.htm>, 2016. Acesso em: 03 set.
- [4] Felipe Campelo. Lecture Notes on Design and Analysis of Experiments. <http://git.io/v3Kh8>, 2018. Version 2.12; Creative Commons BY-NC-SA 4.0.
- [5] Eduardo Gontijo Carrano. ANALYSIS OF VARIANCE, 2020. Notas de aula.
- [6] Helios de Rosario. Power Calculations For Balanced One-Way Analysis Of Variance Tests. <https://www.rdocumentation.org/packages/pwr/versions/1.3-0/topics/pwr.anova.test>, 2020. Documentation reproduced from package pwr, version 1.3-0.

- [7] Banco do Brasil S.A. Tabelas de rentabilidades. <https://www.bb.com.br/pbb/pagina-inicial/voce/produtos-e-servicos/investimentos/tabela-de-rentabilidade/>, 2020. Acesso em 31 ago.
- [8] Toro Investimentos. Como operar vendido: aprenda a investir em ações que estão em queda. <https://blog.toroinvestimentos.com.br/operar-vendido>, 2020. Acesso em 31 ago.
- [9] Lüdecke, Waggoner, and Makowski. `effectsize`: Indices of Effect Size and Standardized Parameters. <https://CRAN.R-project.org/package=effectsize>, 2019. Version 0.3.2.
- [10] Bruno Oliveira. TESTE DE TUKEY PARA COMPARAÇÕES MÚLTIPLAS. <https://operdata.com.br/blog/comparacoes-multiplas-teste-de-tukey/>, 2019. Acesso em: 03 set.
- [11] R Development Core Team. Bartlett Test Of Homogeneity Of Variances. <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/bartlett.test>. Documentation reproduced from package stats, version 3.6.2, License: Part of R 3.6.2.
- [12] R Development Core Team. Shapiro-Wilk Normality Test. <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/shapiro.test>. Documentation reproduced from package stats, version 3.6.2, License: Part of R 3.6.2.
- [13] Soner Yigit and Mehmet Mendes. Which effect size measure is appropriate for one-way and two-way ANOVA models? A Monte Carlo simulation study. *Revstat Statistical Journal*, 16:295–313, 2018.
- [14] Achim Zeileis. Durbin-Watson Test. <https://www.rdocumentation.org/packages/lmtest/versions/0.9-37/topics/dwtest>. Documentation reproduced from package lmtest, version 0.9-37.