

# Estudo 2

Jonatan Almeida e Helbert Paulino

2023-10-25

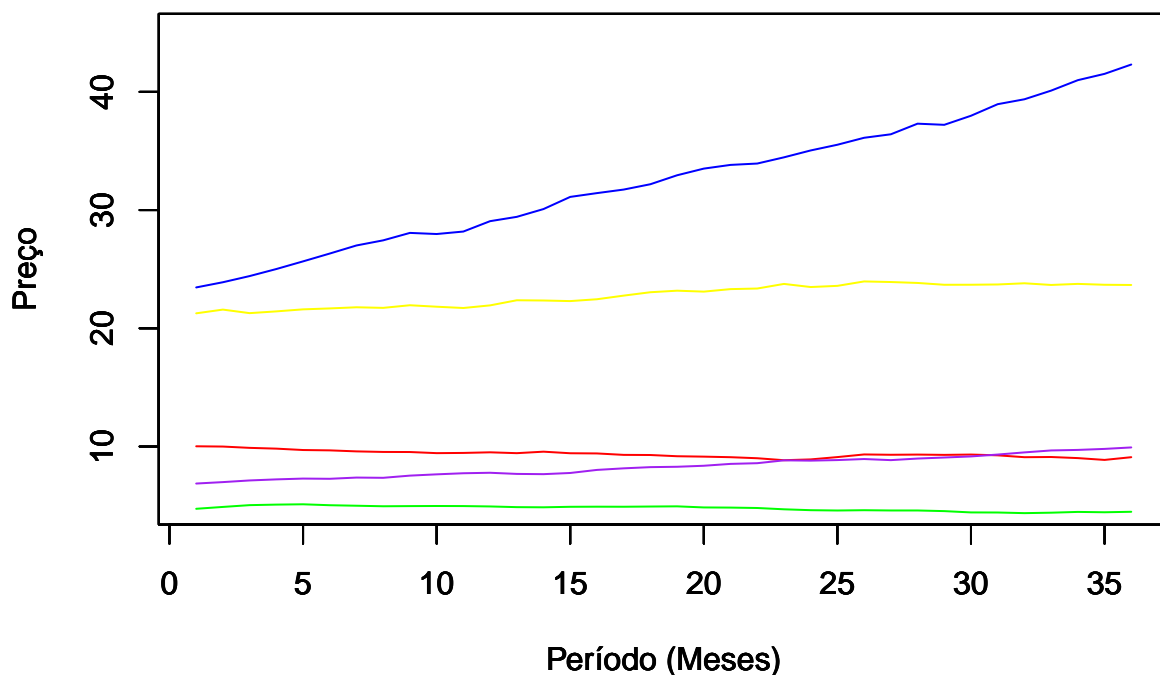
## Resumo

Este estudo de caso tem por objetivo realizar comparações estatísticas entre os dados coletados de ações na bolsa de valores e determinar quais delas irão produzir melhor retornos financeiros para um investidor, que deseja investir todo o seu montante em uma dessas ações. Dessa forma, torna-se crucial determinar qual ação ofecere maior variação positiva no preço e, para isso, analisaremos o histórico delas.

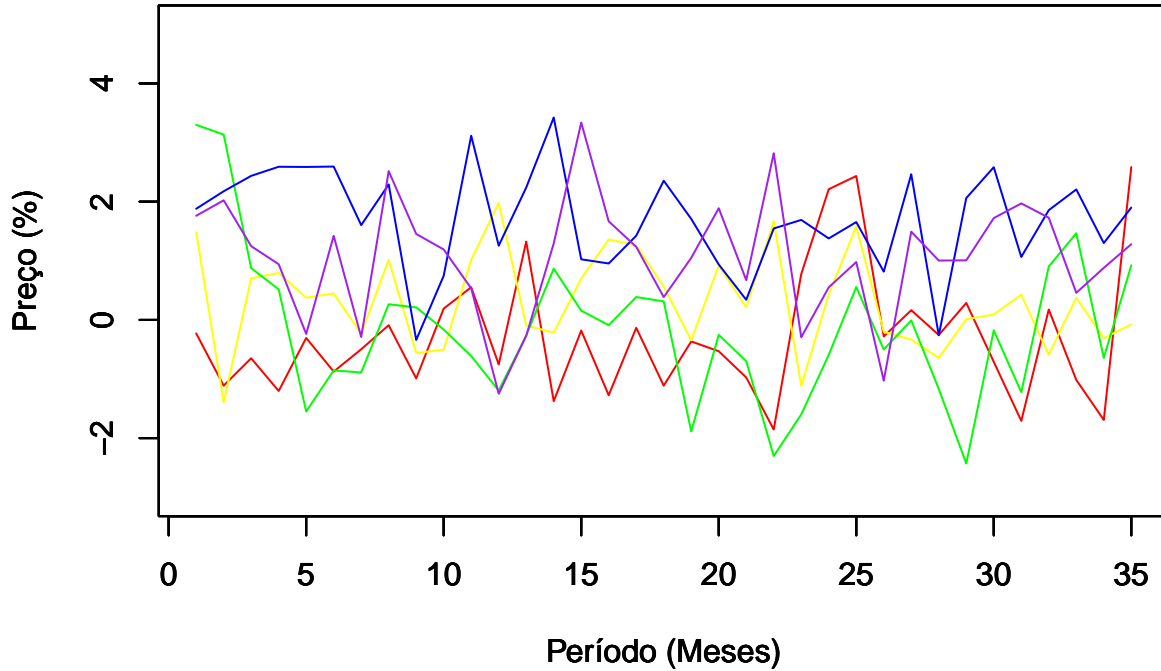
## Análise exploratória dos dados

Os dados que analisaremos consistem no conjunto de fechamento de preços de cinco ações nos últimos 36 meses. Com isso, através do seu histórico, podemos analisar o desempenho de diferentes companhias, observando a condição de independência dos dados. Conforme a tabela obtida, a primeira linha é referente ao período mais atual (ou seja, a variação de preço desse mês para o anterior) até a última linha. Contudo, antes de realizar alguns testes estatísticos, vale a pena analisar graficamente os preços em nessas ações a fim de se obter uma consciência situacional de cada uma. Podendo avaliar um balanço financeiro e postular sobre quais seriam mais rentáveis, de forma hipotética. Para isso, vamos plotar para cada companhia o histórico das ações. Antes, para a plotagem temporal, torna-se necessário reordenar a tabela antes de fazer a plotagem. As figuras a seguir ilustram o processo para cada grupo de ações.

### Historico da ação



## Varição mensal



Pode-se observar em uma primeira análise que a ação 4 (representado pela linha azul) apresenta maior indicativo de valorização, bem como é a que apresenta melhor variação percentual mensal. As ações em amarelo e roxo apresentam também tendência de crescimento, enquanto as ações em verde e vermelho demonstram tendência de queda. Isso, no entanto, precisa ser avaliado com outras técnicas para inferir sobre a variação ao longo dos meses, o que será objeto de investigação nos testes de hipóteses.

### Análise Estatística

Para poder classificar e determinar quais ações seriam as melhores a serem investidas (e se vale a pena investir nelas), temos que definir as hipóteses do experimento. Ao analisarmos as amostras, tendo em vista que são pertencentes a diferentes companhias, que, teoricamente, não possuem relação de dependência explícita, então podemos assumir que os dados são independentes (ou seja, iid). Além disso, tendo em vista que temos um  $N > 30$ , variância e média finitas, então podemos usar do TCL (Teorema Central do Limite) para inferir que as médias seguirão uma distribuição normal.

Se considerarmos o modelo estatístico baseado nos efeitos para cada nível, dado por:

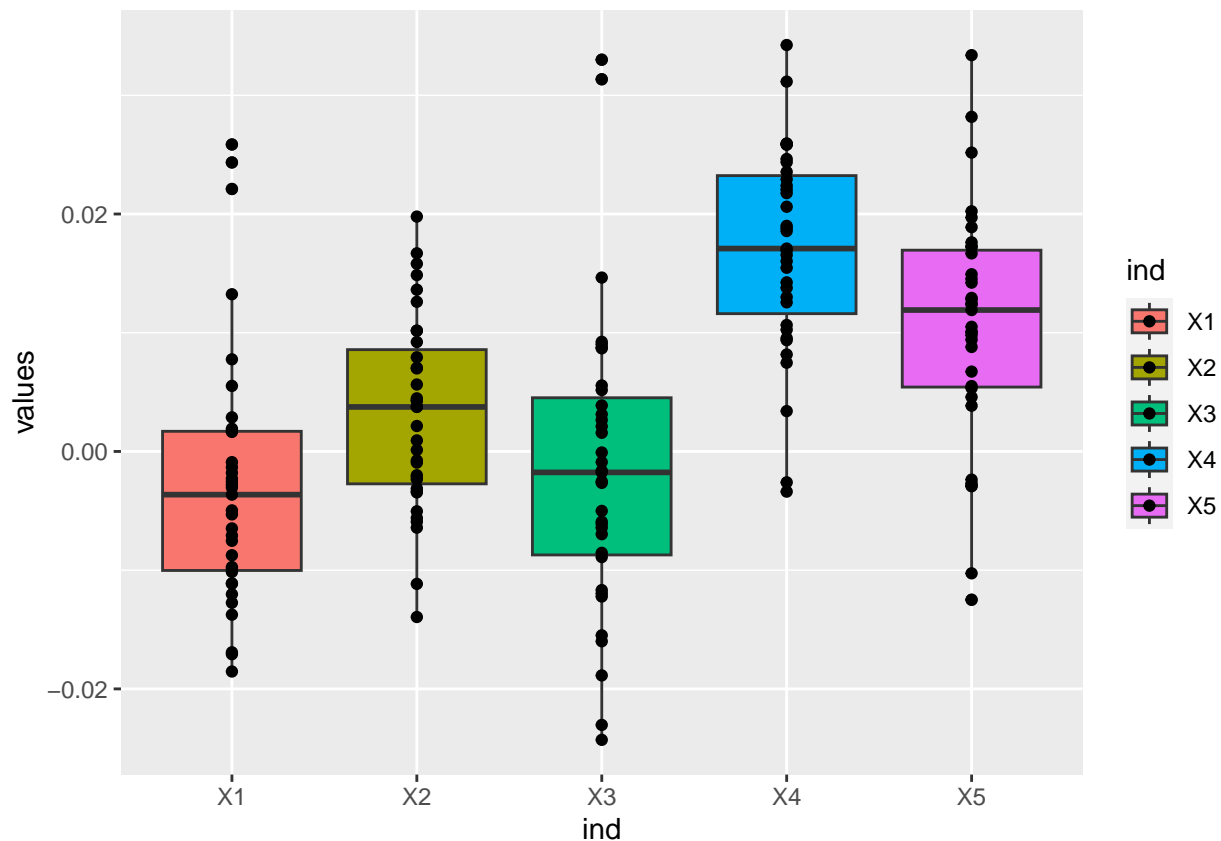
$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

em que  $i = 1, \dots, a$  (número de níveis) e  $j = 1, \dots, n$  (número de observações). Em que  $\mu$  é a média global,  $\tau_i$  é o efeito do nível  $i$  e  $\epsilon_{ij}$  é o resíduo.

Dessa forma, a pergunta de interesse nos leva a definir os seguintes testes de hipóteses:

$$\begin{cases} H_0 : \tau_i = 0, \forall i \in [1, a] \\ H_1 : \exists \tau_i \neq 0 \end{cases}$$

Ou seja, dentro de um determinado valor de tolerância, os efeitos dos níveis são nulos ou para algum dos níveis ele varia de forma significativa? Para avaliar isso, utilizaremos o box plot para observarmos a variabilidade para cada nível, bem como eventuais assimetrias. Para garantir confiabilidade em nosso teste, será definido um  $\alpha = 0.05$ .



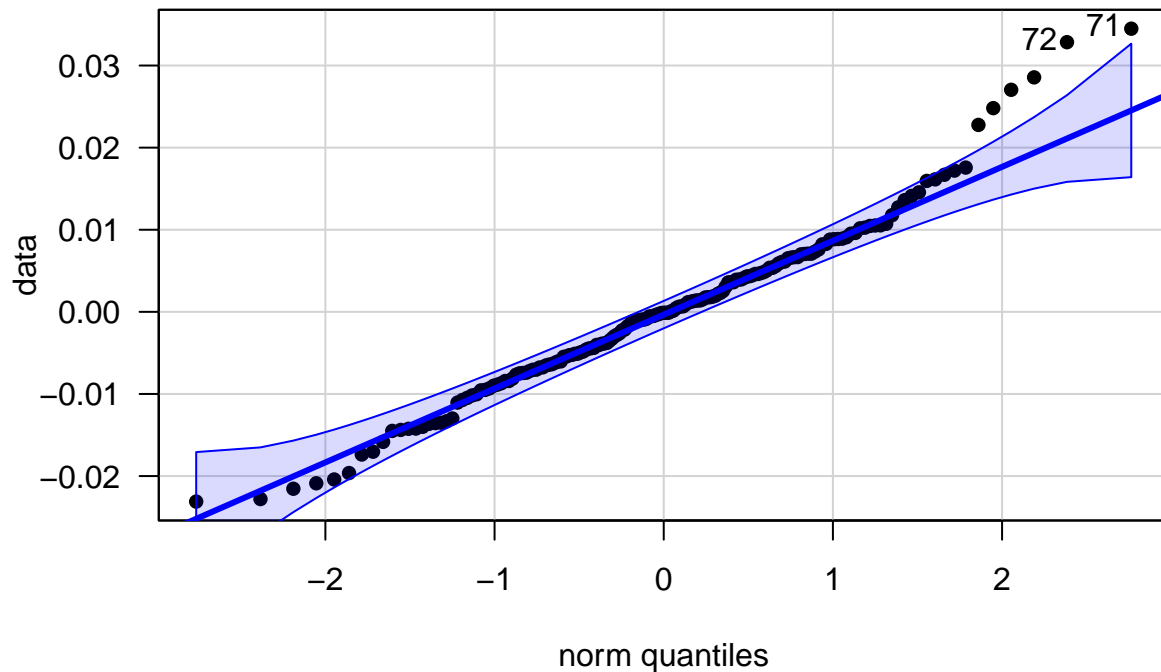
Como pode-se observar, as medianas das ações em amarelo, azul e roxo, são superiores a 0, o que demonstra que essas ações apresentam potencial de valorização. Além disso, o quartil inferior das ações em azul e roxo mostram essas ações apresentaram, no período avaliado, maior valorização. Podemos avaliar, também, com o teste ANOVA, a relação entre os efeitos e as hipóteses levantas.

```
##           Df  Sum Sq  Mean Sq F value  Pr(>F)
## ind           4 0.009832 0.0024581    23.81 1.2e-15 ***
## Residuals    170 0.017552 0.0001032
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Nesse teste, obtemos o p-valor de  $1.2 \times 10^{-15}$ , que é significativamente menos que um  $\alpha$  (p-valor  $< 0.05$ ). Nesse caso, há pelo menos um dos níveis em que temos um efeito significativamente maior que 0. Contudo, vale a pena checar se as premissas adotadas (independência e normalidade) foram atendidas. Para isso, aplicaremos o teste de Shapiro-Wilk com  $\alpha = 0.05$ .

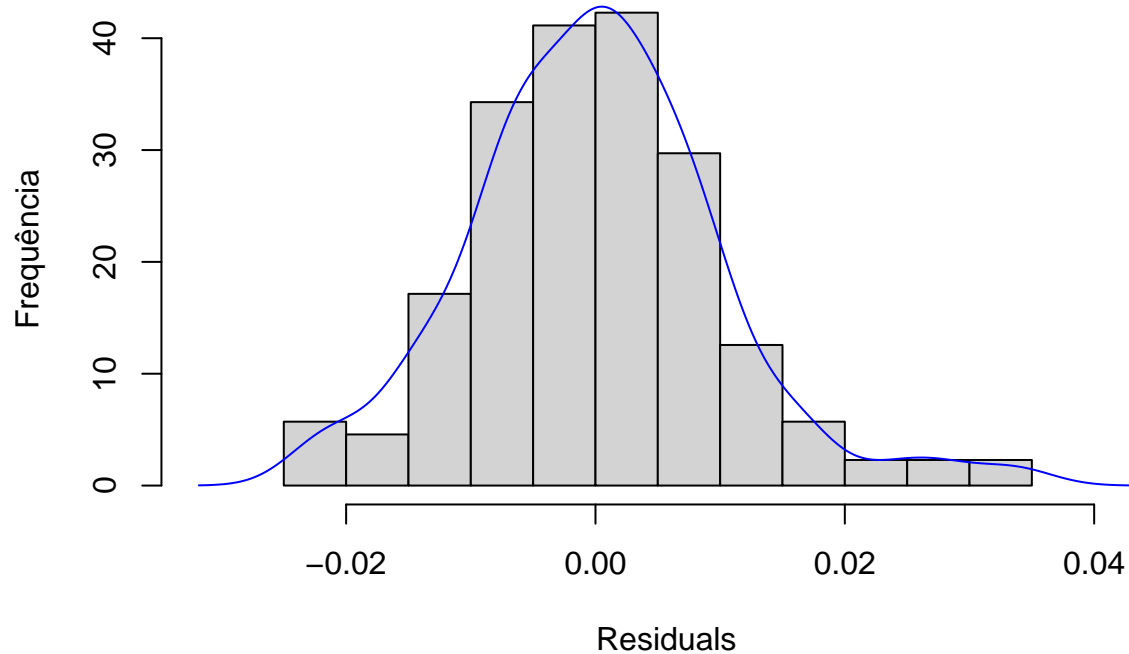
```
##
## Shapiro-Wilk normality test
##
## data:  my.model$residuals
## W = 0.97713, p-value = 0.005541
```

Segundo o teste Shapiro-Wilk, obtemos um valor de p-valor superior a um  $\alpha=0.05$ , o que indica que, dentro do intervalo de confiança estabelecido, não podemos garantir sobre a normalidade dos dados. Contudo, o teste ANOVA é robusto a violações moderadas na normalidade. Então vamos analisar graficamente.



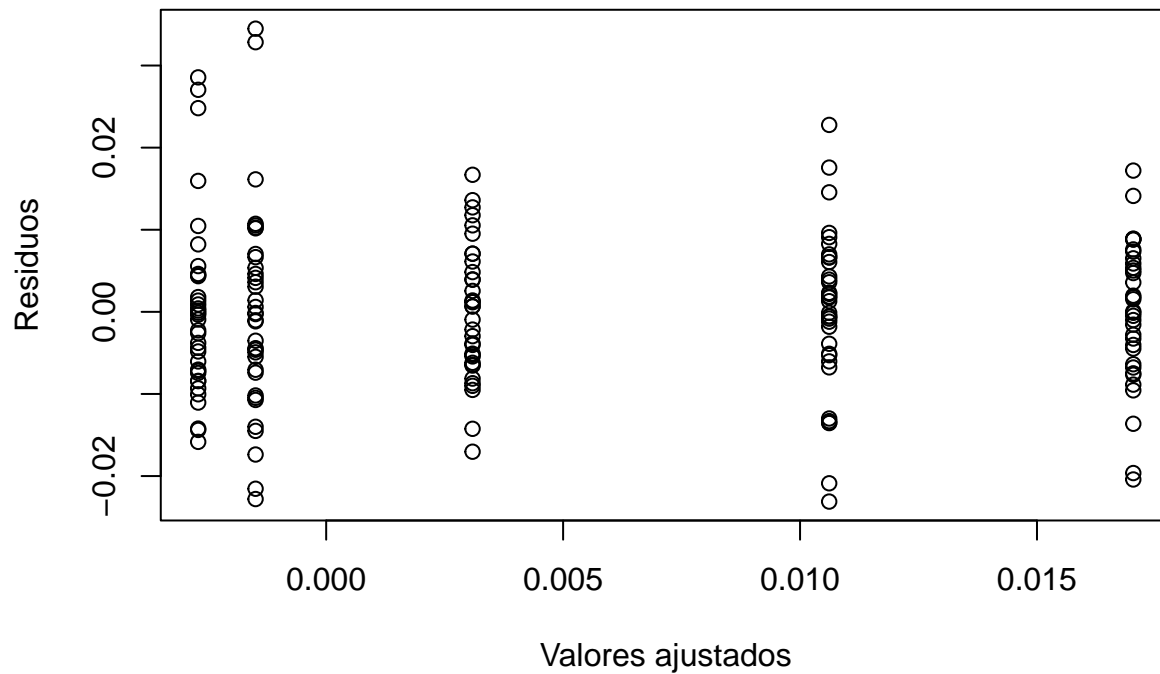
## [1] 71 72

### Frequency of residuals



Observado o QQPlot, observa-se indícios de normalidade nos resíduos do modelo. Quando olhamos o histograma dos resíduos, notamos também uma similaridade com a distribuição normal. Portanto, vamos considerar normalidade nos dados e avaliar outras condições de teste.

Um teste importante é o teste de homoscedasticidade, que assume como hipótese nula que as variâncias são homogêneas. Para isso utilizaremos o teste de Fligner com  $\alpha=0.05$ .



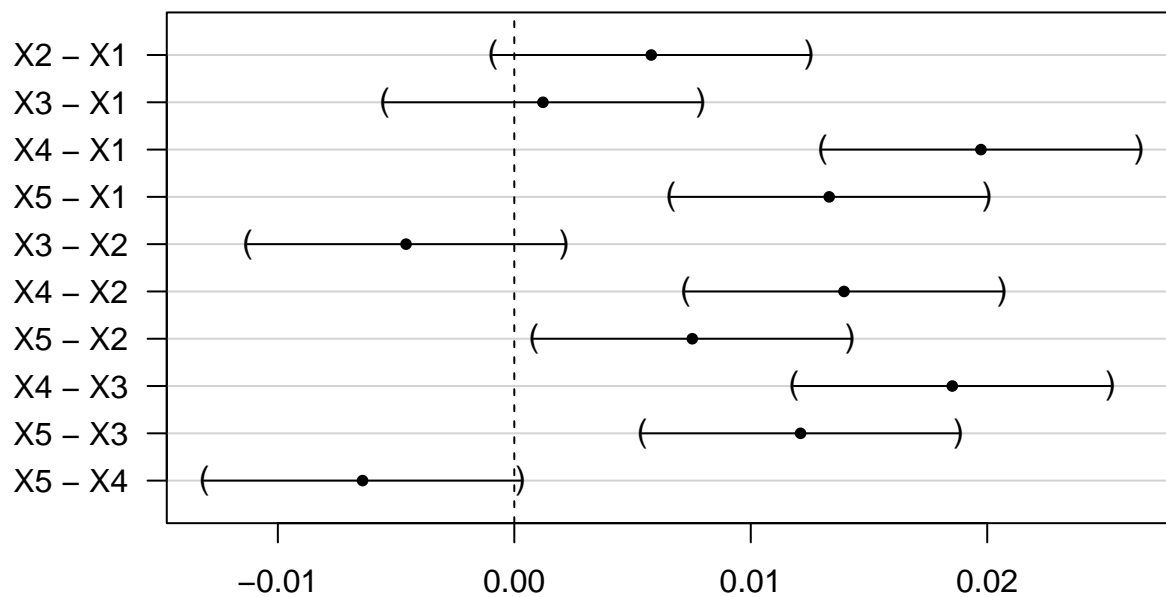
Com um  $p\text{-valor} > \alpha$ , temos como resultado da homoscedasticidade, a não rejeição da hipótese nula, ou seja, os dados possuem variâncias homogêneas.

Por fim, tendo em vista que as premissas do teste ANOVA foram suficientemente atendidas, podemos rejeitar  $H_0$  com o nível de confiança desejado. Contudo, ainda é interessante analisar quais os níveis do fator que apresenta efeito significativo para contribuir com esse resultado, o que definirá o poder do teste.

#### Determinação do poder de teste

Após toda essa análise sobre as hipóteses, podemos agora realizar uma comparação estatística de cada ação para dizer realmente qual é a melhor ação para investir. Para isso utilizaremos uma abordagem denominada de “Todos contra todos” através do método de Tukey, que é conhecido como Abordagem de Diferença Significativa Honesta. Para isso será escolhido um  $\alpha = 0,05$ , ou seja, um nível de confiança 95% para o teste.

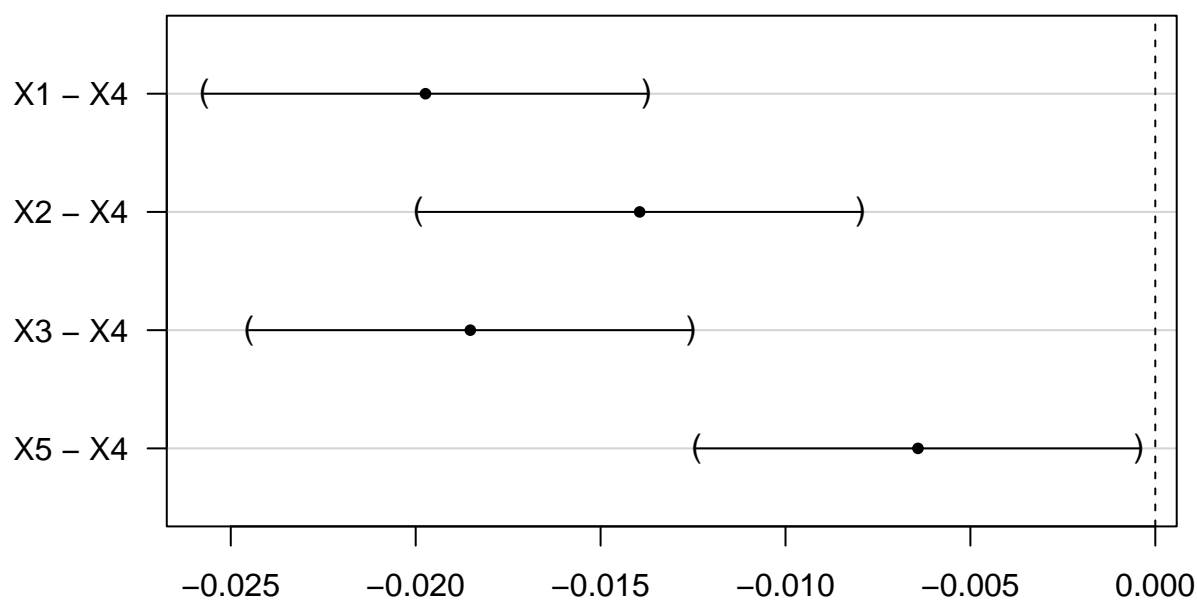
### 95% family-wise confidence level



Linear Function

Podemos observar no gráfico que a ação X4 obteve um melhor resultado em relação às demais. Seguido da ação 5, que também teve um resultado muito bom. Quando olhamos para o BoxPlot, observamos exatamente este comportamento. Para complementar a análise, utilizaremos o teste de Dunnett, que possui sensibilidade mais alta que o teste de Tukey.

### 95% family-wise confidence level



Linear Function

Como pode-se observar, as diferenças significativas foram maiores entre as ações 1 e 4, e que entre as ações 5 e 4 as variações foram menores, isso é esperado, conforme os testes anteriores e os gráficos dos dados

### **Atividades específicas**

Ambos os autores realizaram a avaliação dos dados estatísticos, pesquisaram sobre a ferramenta utilizada para os cálculos, realizaram correções nos trabalhos, implementações em R e sugestão de testes.

### **Conclusões**

A partir dos testes, podemos concluir que existe sim uma ação melhor para investimento. A ação 4 obteve o melhor resultado no teste de Tukey e Dunnett, enquanto a ação 5 também obteve um resultado interessante. Porém analisando o BoxPlot, esta última é uma ação que possui uma variabilidade maior que a 4. Logo, com um resultado maior e uma variabilidade menor, a ação com melhor retorno é a 4. Em contrapartida, a ação que apresentou menor potencial de retorno foi a ação 1, possuindo baixo retorno. Já a ação 3, pelo BoxPlot e pelo gráfico de variação das ações, é possível perceber que foi o que apresentou maior variabilidade, demonstrando ser uma ação de maior risco que as demais, com uma desvalorização média próxima de 1% ao mês, nos 5 primeiros meses.