

Estudo 2

Jonatan Almeida e Helbert Paulino

2023-10-22

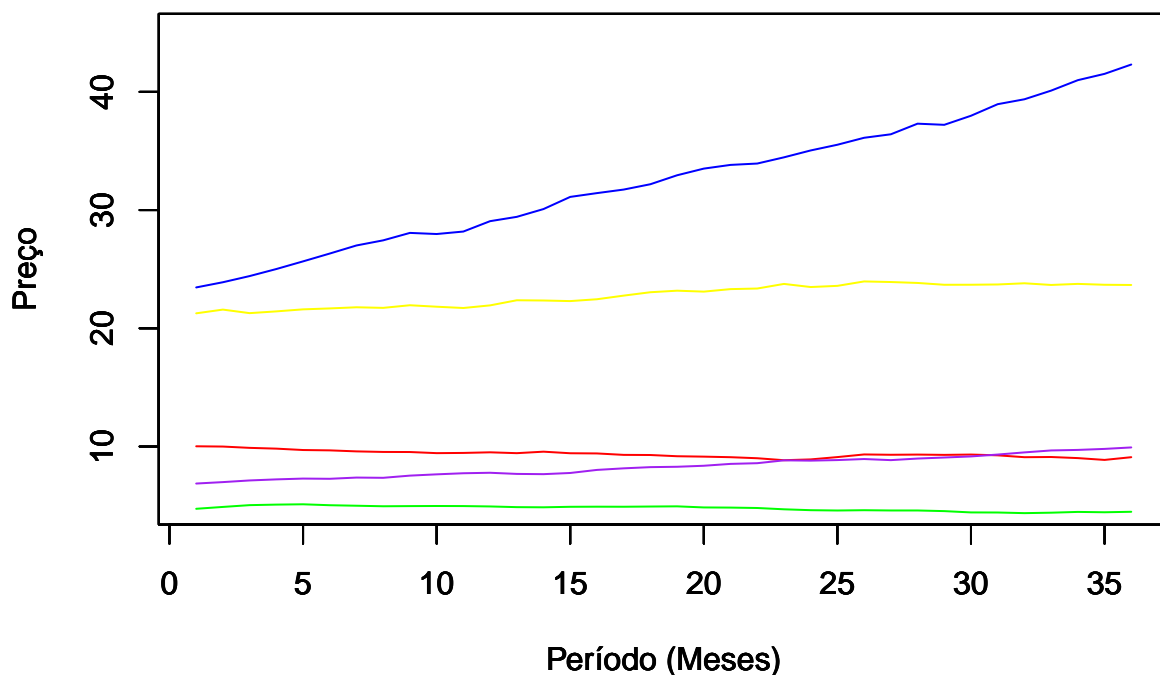
Resumo

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 quais carteiras ofecerem maior variação positiva no preço das ações e, para isso, analisaremos o histórico delas.

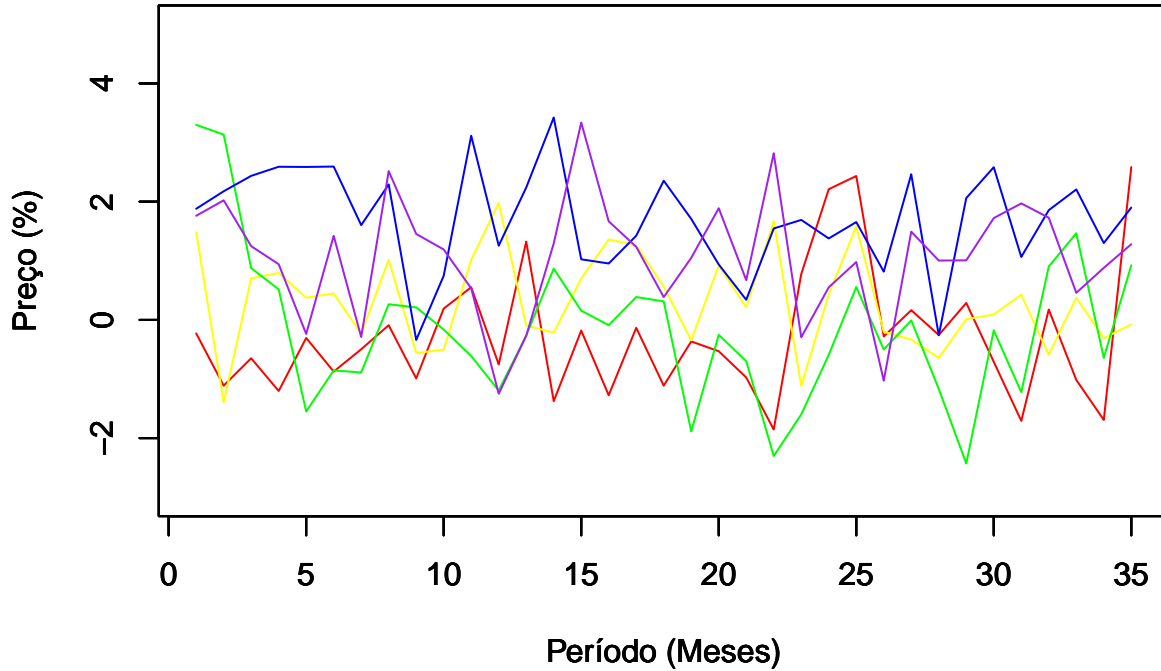
Análise exploratória dos dados

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 sobre a situação de cada uma, podendo avaliar um balanço financeiro positivo ou negativo, 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 avaliar se as ações variaram mesmo de forma significativa 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 inferirmos que as médias seguirão uma distribuição normal.

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

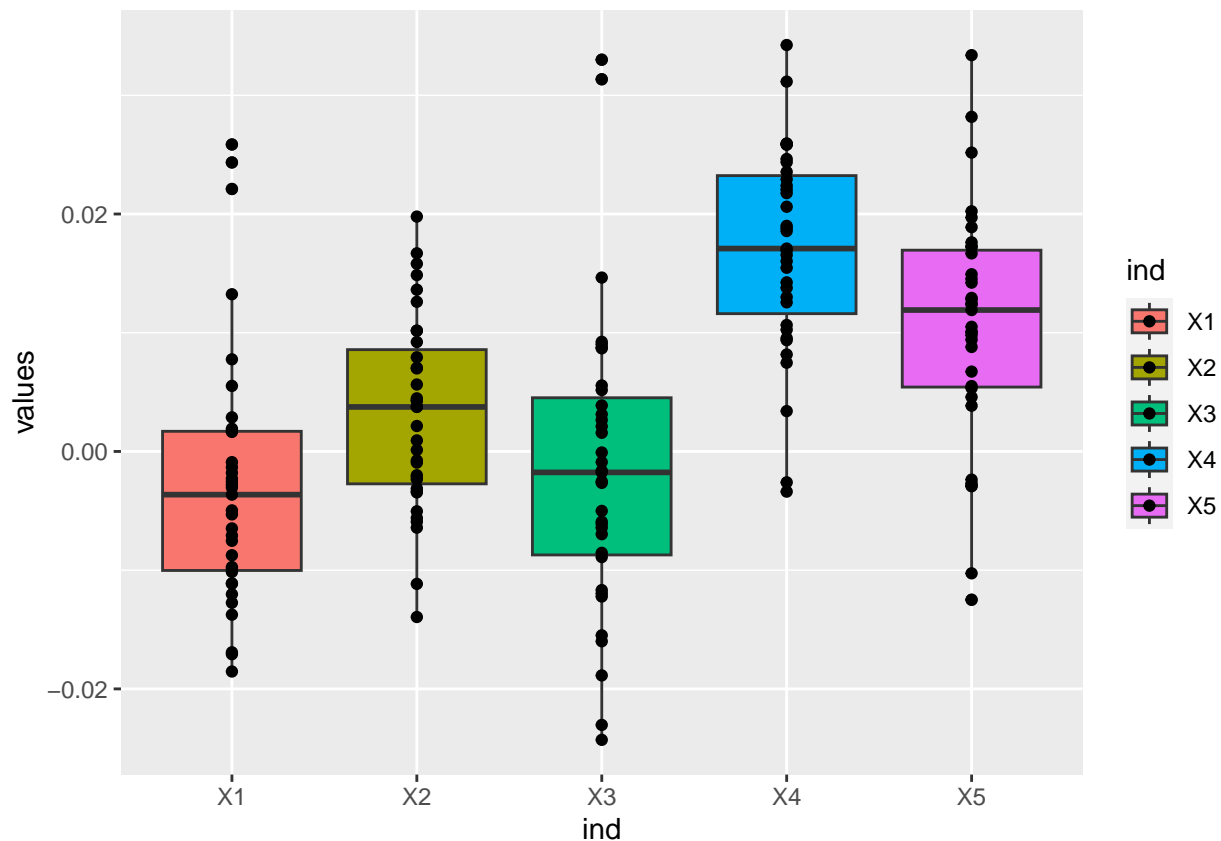
$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

em que $i = 1, \dots, a$ (número de níveis) e $j = 1, \dots, n$ (número de observações). Em que μ é a média global, τ_i é o efeito do nível i e ϵ_{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.

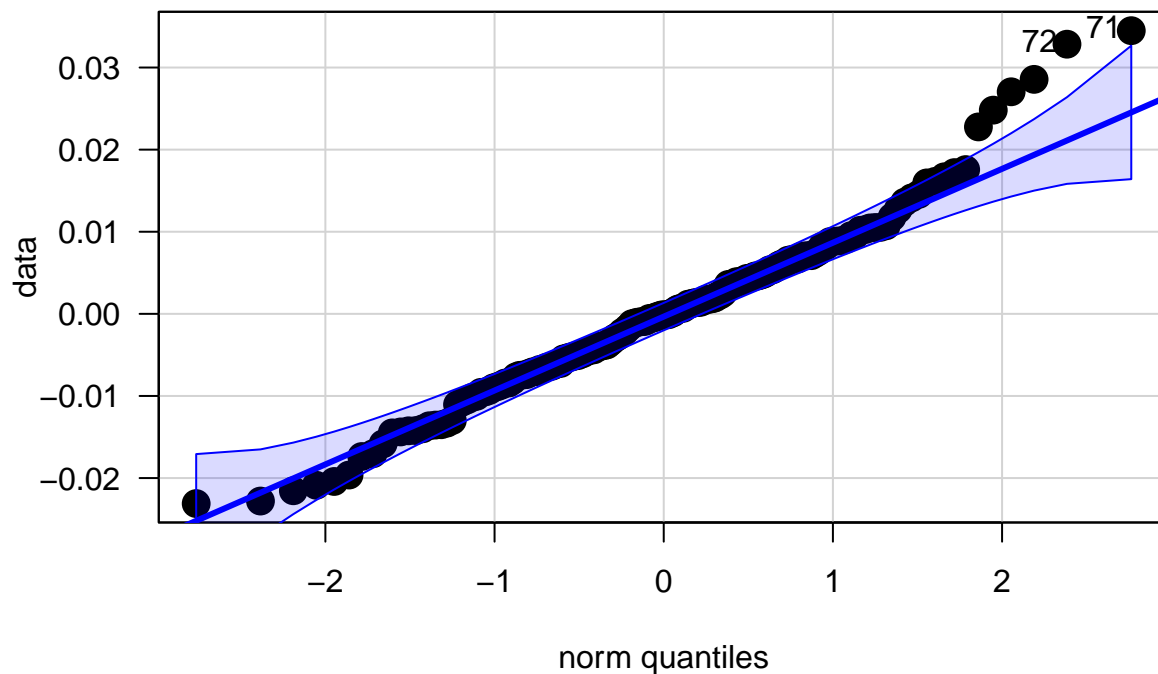


Como pode-se observar, as medianas das ações em amarelo, azul e roxo, são superiores a 0, o que demonstra que essas ações apresentam potencial de valorização. Além disso, o quartil inferior das ações em azul e roxo mostram essas ações apresentaram, no período avaliado, maior valorização. Podemos avaliar, 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 1.2×10^{-15} , que é significativamente menos que um α tipicamente escolhido de 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.

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



```
## [1] 71 72
```

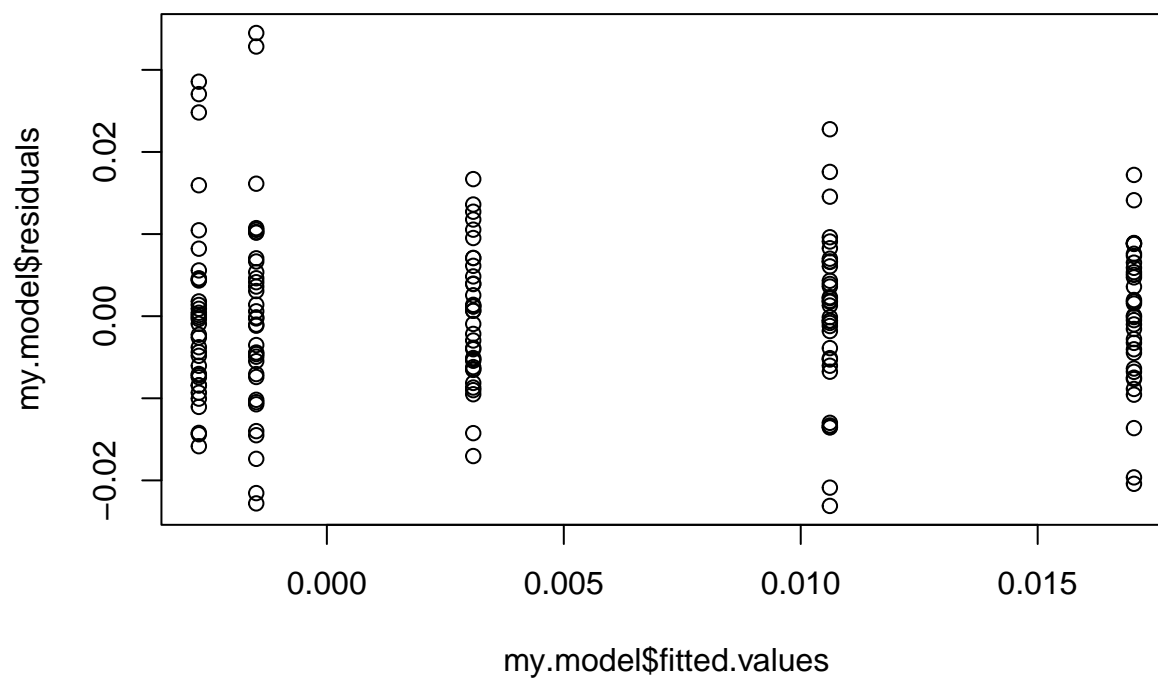
```
##
```

```
## Fligner-Killeen test of homogeneity of variances
```

```
##
```

```
## data: values by ind
```

```
## Fligner-Killeen:med chi-squared = 2.7619, df = 4, p-value = 0.5984
```



```
## Carregando pacotes exigidos: mvtnorm
```

```
## Carregando pacotes exigidos: survival
```

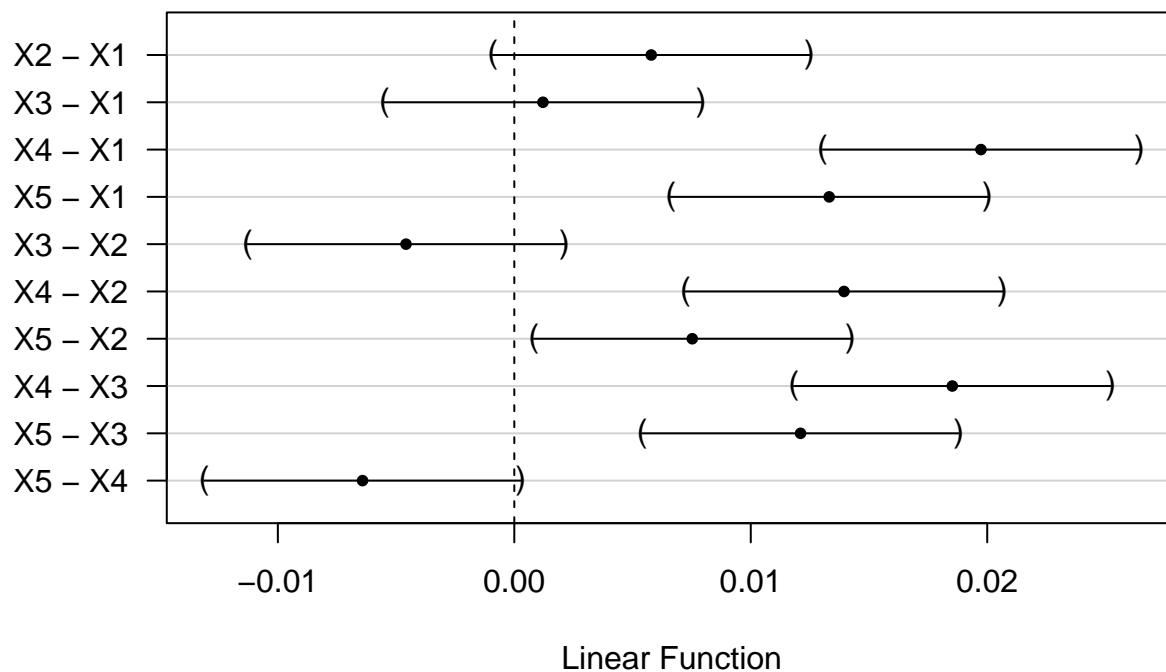
```
## Carregando pacotes exigidos: TH.data
```

```

## Carregando pacotes exigidos: MASS
##
## Attaching package: 'TH.data'
##
## The following object is masked from 'package:MASS':
##
##     geyser
##
## Simultaneous Confidence Intervals
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: aov(formula = values ~ ind, data = returns)
##
## Quantile = 2.7574
## 95% family-wise confidence level
##
##
## Linear Hypotheses:
##           Estimate    lwr      upr
## X2 - X1 == 0  0.0057907 -0.0009068  0.0124882
## X3 - X1 == 0  0.0012114 -0.0054860  0.0079089
## X4 - X1 == 0  0.0197343  0.0130368  0.0264318
## X5 - X1 == 0  0.0133169  0.0066194  0.0200144
## X3 - X2 == 0 -0.0045792 -0.0112767  0.0021183
## X4 - X2 == 0  0.0139436  0.0072461  0.0206411
## X5 - X2 == 0  0.0075262  0.0008287  0.0142237
## X4 - X3 == 0  0.0185229  0.0118254  0.0252204
## X5 - X3 == 0  0.0121054  0.0054079  0.0188029
## X5 - X4 == 0 -0.0064174 -0.0131149  0.0002801

```

95% family-wise confidence level



Determinação do poder de teste

xxx

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

xxx