Universidade Federal do Rio Grande – FURG
Probabilidade e Estatística Aplicada à Engenharia
Rio Grande/RS, Brasil
Engenharia de Computação

Análise exploratória e de Correlação entre duas Variáveis: Volume e Preço Médio com técnicas de estatística descritiva

Orientador(a): Viviane Leite Dias de Mattos
Aluno: Andrey Vinicius Santos Souza

Novembro 2024

1 Introdução

O mercado financeiro conta com diversos tipos de análises, predições e modelagens para auxiliar na tomada de decisão de negócios em todo o mundo, desempenhando um papel crucial na geração de quantidades substanciais de dados [2]. Esses dados precisam passar por um processamento refinado, para serem usados posteriormente para análises técnicas, estatísticas e de retorno.

Uma publicação recente na literatura acadêmica por Cardoso, Malska, Ramiro, Lucca, Borges, Mattos e Berri [1] apresenta o BovDB como um conjunto de dados de referência para pesquisa em previsão do mercado de ações. Este conjunto de dados também passou por uma atualização recente, o BovDBv2[1], e será utilizado neste trabalho. Este novo conjunto de dados é publicamente acessível e pré-processado, contendo informações diárias de todas as ações negociadas na B3 (Brasil Bolsa Balcão)[2] no período de 1995 a 2024.

Análises estatísticas descritivas fornecem ferramentas essenciais para interpretar esses dados e identificar padrões significativos, auxiliando na compreensão de variáveis importantes para o mercado financeiro. Neste trabalho, exploramos duas variáveis principais retiradas do conjunto de dados: - '**average**', que representa o preço médio das ações em determinado dia; - '**volume**', que mede o volume total de negociações. Por meio de técnicas de estatística descritiva como medidas de tendência central, variabilidade, assimetria e curtose, esse estudo busca identificar possíveis valores fora do padrão e identificar a correlação entre essas duas variáveis permitindo uma compreensão detalhada do comportamento, fornecendo *insights* relevantes para estudos financeiros e estratégias de investimento.

Este trabalho está organizado da seguinte forma. A seção 2 apresenta a metodologia adotada neste trabalho. Enquanto a Seção 3 mostra os resultados obtidos. As conclusões e trabalhos futuros são abordados na Seção 4.

[1] `https://github.com/Ginfofinance/BovDbV2repository`
[2] `https://www.b3.com.br`

2 Metodologia

Esta seção apresenta a metodologia adotada no estudo, com o objetivo de fornecer uma compreensão abrangente da abordagem utilizada para a análise estatística descritiva do conjunto de dados BovDBv2. Foram extraídas informações das colunas "Volume" e "Preço Médio" da ação **PETR4** (Petroleo Brasileiro S.A. Petrobras) no período de '2024-01-02' a '2024-06-28'. Esses dados estão presentes na tabela *price*, que contém os dados diários das ações, com colunas representando as variáveis e as linhas representando os dias. Foram realizadas análises exploratórias e descritivas para investigar as propriedades das duas variáveis selecionadas.

2.1 Descrição das Variáveis e Contexto

As variáveis analisadas, "Volume" e "Preço Médio", representam métricas fundamentais para compreender o comportamento do papel **PETR4**. Esta subseção contextualiza essas variáveis, explicando como são usadas e seus respectivos cálculos para formar os valores presentes no banco de dados.

Cálculo do Volume A variável **Volume** é definida como o produto entre o preço de negociação (Preço_i) e a quantidade de ações transacionadas (Quantidade_i) dentro de um intervalo de tempo específico, sendo calculada por:

$$\text{Volume} = \sum_{i=1}^{n} (\text{Preço}_i \times \text{Quantidade}_i).$$

Esta métrica reflete o total financeiro negociado em um período, e no caso dos dados utilizados, o período corresponde a um dia de negociação.

Cálculo do Preço Médio O **Preço Médio** é obtido dividindo o volume financeiro total pela quantidade total de ações negociadas:

$$\text{Preço Médio} = \frac{\text{Volume}}{\text{Quantidade Total}}.$$

Esta variável indica a média ponderada do preço das transações, refletindo o preço médio de negociação das ações em um determinado período.

A ação PETR4, pertencente à Petrobras, foi escolhida por sua alta liquidez e representatividade no mercado brasileiro. Este papel reflete características fundamentais do mercado de capitais, sendo amplamente utilizado em estudos estatísticos e análises de mercado, sendo, portanto, uma variável central para a análise no contexto de ações brasileiras.

2.2 Análise Exploratória de Dados

A análise exploratória foi realizada como uma etapa inicial para compreender as características fundamentais do conjunto de dados [4]. Essa análise incluiu

a verificação de inconsistências, a identificação de valores fora do padrão (*outliers*), além da utilização de técnicas gráficas, como histogramas e diagramas de dispersão, para avaliar a distribuição das variáveis e investigar possíveis relações entre elas.

2.3 Estatísticas Descritivas

A aplicação das técnicas de estatística descritiva foi essencial para resumir as características principais do conjunto de dados. As medidas calculadas incluíram:

- **Medidas de tendência central:**
  - **Média aritmética ($\bar{x}$):**

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

    onde x_i são os valores da variável e n é o número total de observações.
  - **Mediana:** O valor central dos dados ordenados.
    - Para um conjunto de dados com número **ímpar** de observações (n):

$$\text{Mediana} = x_{\left(\frac{n+1}{2}\right)}$$

      onde $x_{(k)}$ representa o k-ésimo valor no conjunto ordenado.
    - Para um conjunto de dados com número **par** de observações (n):

$$\text{Mediana} = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}$$

      ou seja, a média dos dois valores centrais.

- **Medidas de dispersão:**
  - **Valor mínimo ($\min(x)$) e máximo ($\max(x)$):** Os menores e maiores valores do conjunto de dados.
  - **Quartis (Q1, Q3):** Dividem os dados ordenados em quatro partes iguais.
  - **Desvio interquartílico (IQR):**

$$IQR = Q3 - Q1$$

  - **Desvio-padrão (s):**

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

    onde $\bar{x}$ é a média e x_i são os valores individuais.

– **Escore Z (z):**

$$z = \frac{x - \bar{x}}{s}$$

onde x é o valor observado, $\bar{x}$ é a média e s é o desvio-padrão.

- **Medidas de distribuição:**
  - **Coeficiente de assimetria (g_1):**

$$g_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3}$$

onde s é o desvio-padrão e $\bar{x}$ é a média.

  - **Coeficiente de curtose (g_2):**

$$g_2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{s^4} - 3$$

O valor -3 ajusta a curtose para ser comparada com a distribuição normal, que tem curtose igual a 0.

2.4 Representações Gráficas

Foram utilizadas representações gráficas para análise das distribuições individuais de cada variável:

- **Histogramas:** Apresentam a distribuição de frequência das variáveis `average` e `volume`.
- **Boxplots:** Mostram a dispersão e os outliers das variáveis `average` e `volume`.

2.5 Tabela de Comparação

A seguir, apresenta-se uma tabela comparativa entre as duas variáveis em termos de medidas descritivas:

Medidas	**Average**	**Volume**
Média ($\bar{x}$)	μ_{average}	μ_{volume}
Mediana	$\text{Mediana}_{\text{average}}$	$\text{Mediana}_{\text{volume}}$
Desvio-Padrão (s)	σ_{average}	σ_{volume}
Mínimo	min(average)	min(volume)
Máximo	max(average)	max(volume)
Desvio Interquartílico	$\text{average}_{\text{IQR}}$	$\text{volume}_{\text{IQR}}$
Coeficiente de Assimetria	$g_{1,\text{average}}$	$g_{1,\text{volume}}$
Coeficiente de Curtose	$g_{2,\text{average}}$	$g_{2,\text{volume}}$

2.6 Software Utilizado

Todas as análises foram realizadas no ambiente R (*R Programming Language*)[3], que é amplamente utilizado para estatística e visualização de dados [3]. Os pacotes utilizados incluem "e1071" [5] para cálculo de assimetria e curtose, e funções gráficas para geração de histogramas, boxplots e diagramas de dispersão.

3 Resultados

Este capítulo apresenta os resultados e insights obtidos a partir da análise dos dados, organizados em duas subseções principais. Na primeira subseção 3.1, destacam-se as estatísticas descritivas das variáveis analisadas, essas métricas fornecem uma visão detalhada sobre o comportamento individual de cada variável. Na segunda subseção 3.2, é apresentado o diagrama de dispersão, utilizado para investigar a relação entre as variáveis `average` e `volume`. Nesta etapa, discute-se a força e direção da relação linear entre as variáveis, incluindo a interpretação do coeficiente de correlação de Pearson (r). O objetivo é avaliar se a análise sugere a existência de uma correlação nos dados amostrais.

3.1 Análise univariada

Com base na análise exploratória dos dados e na aplicação das técnicas de estatística descritiva descritas na seção 2.3, os resultados resumidos estão apresentados na Tabela 1. Essa tabela contém as principais medidas descritivas calculadas para as variáveis `Volume` e `Preço Médio`(Average), permitindo uma visão clara sobre a tendência central, dispersão, assimetria e curtose dos dados.

Table 1: Estatísticas descritivas das variáveis `Volume` e `Preço Médio`

Medidas	**Average**	**Volume**
Valor Mínimo	34.75	5.03×10^8
Valor Máximo	42.80	8.33×10^9
Quartil 1 ($Q1$)	37.08	1.05×10^9
Mediana	38.46	1.37×10^9
Quartil 3 ($Q3$)	40.71	1.77×10^9
Desvio Interquartílico (IQR)	3.63	7.26×10^8
Média Aritmética	38.89	1.60×10^9
Desvio-Padrão	2.10	1.05×10^9
Coef. Assimetria (g_1)	0.18	3.55
Coef. Curtose (g_2)	-1.17	16.35
Outliers	0	7
Lacunas	0	0

[3] `https://www.r-project.org/`

A partir das estatísticas descritivas apresentadas na Tabela 1, foi possível obter insights relevantes sobre o comportamento das variáveis **Average** e **Volume**. Assim, discute-se cada métrica analisada e as implicações desses resultados em termos de variabilidade, distribuição e características gerais dos dados.

Tendência Central Os valores de média e mediana de **Average** ($\bar{x} = 38.89$, Mediana $= 38.46$) indicam uma distribuição balanceada, com pouca diferença entre esses indicadores, sugerindo que a variável é aproximadamente simétrica. Em contraste, para **Volume**, a média ($\bar{x} = 1.60 \times 10^9$) é significativamente maior que a mediana (1.37×10^9). Essa discrepância, somada ao coeficiente de assimetria ($g_1 = 3.55$), confirma a presença de uma assimetria positiva acentuada, indicando que valores extremos (outliers) elevam a média.

Variabilidade A análise da variabilidade revela comportamentos distintos entre as variáveis analisadas. A variável **Average** apresenta menor dispersão, com desvio-padrão ($s = 2.10$) e intervalo interquartílico ($IQR = 3.63$) relativamente baixos, sugerindo uma maior consistência nos valores. Em contrapartida, a variável **Volume** apresenta alta dispersão, evidenciada pelo desvio-padrão ($s = 1.05 \times 10^9$) e pelo intervalo interquartílico ($IQR = 7.26 \times 10^8$), além de uma amplitude significativa entre os valores mínimo (5.03×10^8) e máximo (8.33×10^9).

Distribuição A análise dos coeficientes de assimetria (g_1) e curtose (g_2) evidencia diferenças marcantes na forma das distribuições. **Average** apresenta um $g_1 = 0.18$, indicando leve assimetria positiva, enquanto o coeficiente de curtose ($g_2 = -1.17$) sugere uma distribuição mais achatada do que a normal. Já **Volume** exibe um $g_1 = 3.55$, refletindo forte assimetria positiva, e $g_2 = 16.35$, evidenciando uma concentração elevada ao redor da média, característica de distribuições leptocúrticas. Essas métricas reforçam que **Volume** possui uma estrutura mais heterogênea e influenciada por valores extremos.

3.1.1 Análise Gráfica

Nesta seção, apresentamos as visualizações gráficas das variáveis analisadas, incluindo histogramas e boxplots, com o objetivo de oferecer uma representação visual das distribuições, tendências e possíveis anomalias presentes nos dados. Estas ferramentas gráficas complementam as estatísticas descritivas apresentadas anteriormente, permitindo uma interpretação mais intuitiva das características do conjunto de dados.

Histograma O histograma de cada variável fornece uma visão geral da distribuição dos dados, indicando se a variável segue uma distribuição simétrica, enviesada ou apresenta múltiplos picos (*multimodalidade*).

- O histograma da variável `average` (preço médio) revela uma distribuição ligeiramente assimétrica, com uma cauda levemente mais longa à direita causada por valores mais altos que estendem a distribuição, mas sem uma concentração tão significativa. Observa-se que a dois picos principais: um na faixa de 36–37 e outro na faixa de 41–42. Isso pode sugerir a presença de grupos distintos ou clusters nos dados.

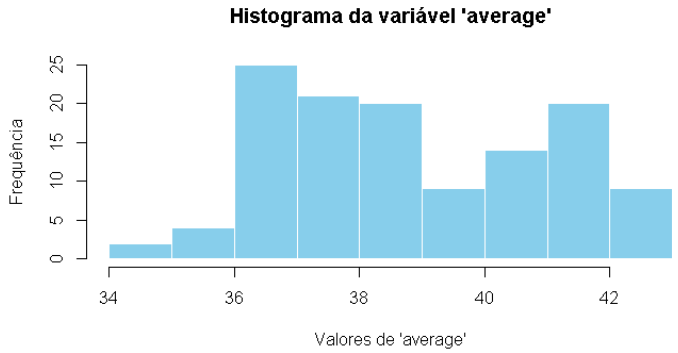

Figure 1: Histograma da variável `average` (preço médio da ação PETR4).

- Para a variável `volume`, o histograma sugere uma alta concentração no início, onde a maioria dos valores está concentrada em um intervalo próximo de 0×10^0 e 2×10^9, sugerindo que a maior parte dos dados está em um intervalo específico, indicando baixa dispersão nesta faixa. Valores fora do padrão, como os que estão entre 4×10^9 e 8×10^9, são representativos de eventos raros ocorridos nas negociações diárias, influenciados pela volatilidade do mercado financeiro.

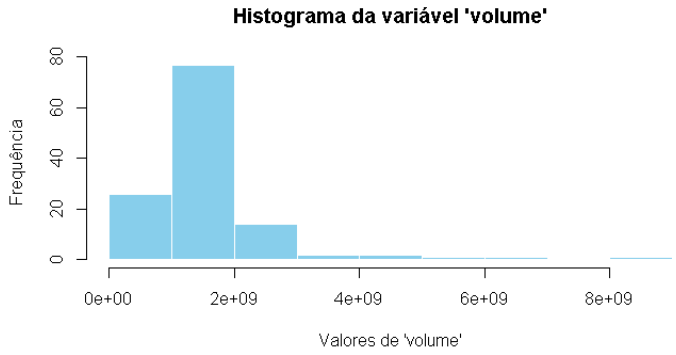

Figure 2: Histograma da variável `volume` (volume de negociações da ação PETR4).

Boxplot Os boxplots das variáveis `average` e `volume` fornecem uma visualização compacta da dispersão dos dados, incluindo os quartis, valores extremos e possíveis *outliers*.

- Para a variável `average`, o boxplot indica que o intervalo interquartílico (IQR), representado pelo retângulo no gráfico, compreende os valores entre o primeiro quartil (Q1), próximo de 36, e o terceiro quartil (Q3), próximo de 40. Isso sugere que 50% dos valores de `average` estão concentrados nesta faixa. Além disso, a linha central do retângulo indica a mediana, aproximadamente igual a 38. Isso demonstra que metade dos valores está abaixo de 38 e a outra metade está acima. Os bigodes do boxplot se estendem para valores próximos de 35 (inferior) e 42 (superior), indicando que a maioria dos valores da variável está dentro deste intervalo, sem grandes discrepâncias.

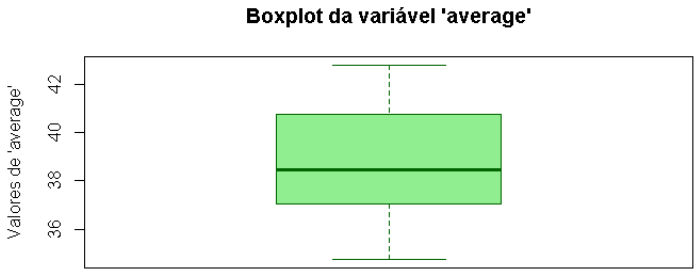

Figure 3: Boxplot da variável `average` (preço médio da ação PETR4).

- Para a variável `volume`, o boxplot sugere a presença de uma quantidade significativa de *outliers* acima do limite superior sendo 7 identificados com sucesso. O corpo principal do boxplot (a área entre o primeiro quartil, $Q1$, e o terceiro quartil, $Q3$) é relativamente pequeno em relação ao eixo dos valores, sugerindo que a maior parte dos dados está concentrada em uma faixa estreita, com menor variação dentro dessa faixa. A mediana, representada pela linha central dentro do boxplot, encontra-se mais próxima do limite inferior do corpo principal, o que também reflete a assimetria positiva da distribuição.

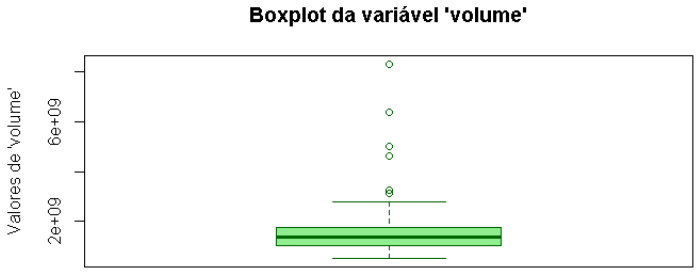

Figure 4: Boxplot da variável `volume` (volume de negociações da ação PETR4).

Essas observações contribuem para os objetivos gerais do estudo, ao destacar áreas de interesse, como volumes atípicos que podem refletir tendências de mercado ou eventos raros, e orientar análises futuras focadas na volatilidade e nas dinâmicas de negociação.

3.2 Análise bivariada

Nesta subseção, realizamos uma análise bivariada entre as variáveis `average` (preço médio das ações) e `volume` (volume de negociações) para investigar a relação existente entre elas. Essa análise foi realizada utilizando o diagrama de dispersão (*scatterplot*) e o cálculo do coeficiente de correlação de Pearson (r).

- **Diagrama de Dispersão:** No gráfico de dispersão, os valores da variável `average` foram plotados no eixo x, enquanto os valores da variável `volume` foram plotados no eixo y. A disposição dos pontos no gráfico foi utilizada para identificar tendências e padrões entre as variáveis.

  A seguir, apresentamos o diagrama de dispersão com as variáveis analisadas:

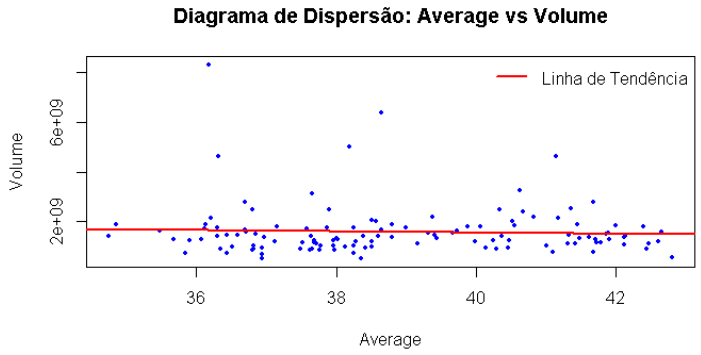

Figure 5: Diagrama de Dispersão `Average` Vs `Volume`

- **Correlação:** O coeficiente de correlação de Pearson (r) foi calculado para avaliar a força e a direção da relação linear entre `average` e `volume`, utilizando a fórmula:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

  - $r > 0$: Indica uma correlação positiva, ou seja, à medida que o preço médio das ações aumenta, o volume de negociações também tende a aumentar.
  - $r < 0$: Indica uma correlação negativa, ou seja, à medida que o preço médio das ações aumenta, o volume de negociações tende a diminuir.
  - $r = 0$: Indica que não há correlação linear entre as variáveis.

Resultados A análise do diagrama de dispersão entre as variáveis `average` (preço médio das ações) e `volume` (volume de negociações) não revelou uma relação linear significativa. A distribuição dos pontos no gráfico demonstrou um padrão amplamente disperso, sem alinhamento claro que indique uma dependência direta entre as variáveis. A linha de tendência ajustada apresentou uma leve inclinação negativa, sugerindo uma correlação inversa de baixa intensidade.

O coeficiente de correlação de Pearson calculado foi $r = -0.053$, um valor próximo de zero, que indica uma associação linear extremamente fraca. O sinal negativo de r aponta para uma relação inversa, sugerindo que, à medida que o preço médio das ações (`average`) aumenta, o volume de negociações (`volume`) tende a apresentar uma redução marginal. No entanto, a intensidade dessa relação é estatisticamente insignificante.

A ausência de uma correlação significativa pode ser atribuída a diversos fatores:

- **Influência de variáveis externas:** O volume de negociações é frequentemente impactado por fatores exógenos, como notícias econômicas, políticas públicas ou eventos relacionados à Petrobras, emissora do papel PETR4. Esses fatores podem obscurecer relações diretas entre preço médio e volume.
- **Características do mercado analisado:** Em mercados financeiros, especialmente em papéis de alta liquidez como PETR4, o volume de negociações nem sempre reflete o comportamento do preço médio. Elementos como volatilidade, liquidez e estratégias de mercado podem ser mais determinantes nesse contexto.
- **Horizonte temporal restrito:** O período analisado (janeiro a junho de 2024) pode ser insuficiente para captar padrões mais robustos ou associações subjacentes. Relações entre preço médio e volume podem emergir em horizontes temporais mais amplos ou sob diferentes condições de mercado.

Em síntese, os resultados indicam que, para o conjunto de dados e o período analisado, o preço médio das ações (`average`) não é um preditor significativo para o volume de negociações (`volume`). Essa conclusão reforça a necessidade de incorporar variáveis adicionais e ampliar o horizonte temporal em análises futuras, com o objetivo de compreender melhor os fatores que influenciam o volume de negociações em ações.

4 Conclusão

Este estudo apresentou uma análise estatística descritiva e exploratória utilizando o conjunto de dados BovDBv2, com foco nas variáveis `average` (preço médio das ações) e `volume` (volume de negociações) da ação **PETR4** no período

de janeiro a junho de 2024. Os resultados indicaram que a variável `average` apresentou uma distribuição relativamente simétrica e sem a presença de *outliers*, enquanto a variável `volume` apresentou uma assimetria positiva, sugerindo a predominância de dias com volumes menores de negociação, com a ocorrência ocasional de valores extremamente altos. Essas observações destacam a importância de considerar as características individuais das variáveis ao realizar análises financeiras.

A análise bivariada revelou uma correlação fraca e negativa ($r = -0.053$) entre as variáveis `average` e `volume`, sugerindo que o preço médio das ações não é um fator determinante para o volume de negociações no curto prazo. Esse resultado, embora esperado em um mercado complexo e influenciado por múltiplos fatores, reforça a necessidade de incorporar outras variáveis e metodologias para compreender melhor a dinâmica entre preço e volume, como análises de volatilidade, fatores macroeconômicos e eventos específicos relacionados à empresa. Para estudos futuros, recomenda-se expandir a análise para incluir um horizonte temporal mais longo, bem como investigar o impacto de outros fatores, como indicadores técnicos, macroeconômicos e notícias de mercado, na relação entre preço médio e volume de negociações. Assim, este trabalho estabelece uma base sólida para pesquisas futuras e reforça a importância da análise de dados financeiros como ferramenta estratégica no mercado de ações.

References

[1] Cardoso, F.C., Malska, J.A.V., Ramiro, P.J., Lucca, G., Borges, E.N., de Mattos, V.L.D., Berri, R.A.: Bovdb: a data set of stock prices of all companies in b3 from 1995 to 2020. Journal of Information and Data Management **13**(1) (2022)

[2] Huang, W.Q., Zhuang, X.T., Yao, S.: A network analysis of the chinese stock market. Physica A: Statistical Mechanics and its Applications **388**(14), 2956–2964 (2009)

[3] Makowski, D., Ben-Shachar, M.S., Lüdecke, D.: bayestestr: Describing effects and their uncertainty, existence and significance within the bayesian framework. Journal of Open Source Software **4**(40), 1541 (2019). https://doi.org/10.21105/joss.01541, `https://joss.theoj.org/papers/10.21105/joss.01541`

[4] Mattos, V.L.D.; Konrath, A.C.A.A.V.: Introdução à estatística: aplicações em ciências exatas. Editora LTC (1) (2017)

[5] Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F.: e1071: Misc functions of the department of statistics, probability theory group (formerly: E1071), tu wien (2024), `https://CRAN.R-project.org/package=e1071`, r package version 1.7-16

Apêndice A: Código R Utilizado na Pesquisa

```r
# Configurar o diretorio de trabalho corretamente
setwd("C:/bovdb/docs")

# Le os dados do arquivo CSV
dados <- read.csv("price_data_107.csv", header = TRUE, sep = ",",
    ↪ dec = ".")

# Verificar a estrutura do dataframe
str(dados)

# Exibir os primeiros registros para confirmar a leitura correta
head(dados)
#--------------------------#

### Analise univariada ###

# Calcular estatisticas descritivas para 'average'
summary_stats <- summary(dados$average)

# Calcular desvio interquartilico (IQR) manualmente
iqr_average <- IQR(dados$average)

# Para coeficientes de assimetria e curtose
library(e1071)

# Media aritmetica
mean_average <- mean(dados$average)

# Desvio-padrao
sd_average <- sd(dados$average)

# Coeficiente de assimetria
skewness_average <- skewness(dados$average)

# Coeficiente de curtose
kurtosis_average <- kurtosis(dados$average)

# Exibir os resultados
cat("Estatisticas descritivas para a variavel 'average':\n")
cat("Valor minimo: ", summary_stats["Min."], "\n")
cat("Quartil 1 (Q1): ", summary_stats["1st Qu."], "\n")
cat("Mediana: ", summary_stats["Median"], "\n")
cat("Quartil 3 (Q3): ", summary_stats["3rd Qu."], "\n")
cat("Valor maximo: ", summary_stats["Max."], "\n")
cat("Desvio Interquartilico (IQR): ", iqr_average, "\n")
cat("Media aritmetica: ", mean_average, "\n")
cat("Desvio-padrao: ", sd_average, "\n")
cat("Coeficiente de assimetria: ", skewness_average, "\n")
cat("Coeficiente de curtose: ", kurtosis_average, "\n")

### Histograma ###
hist(dados$average,
     main = "Histograma da variavel 'average'",
     xlab = "Valores de 'average'",
     ylab = "Frequencia",
```

```
     col = "skyblue",
     border = "white",
     breaks = 10)

### Boxplot ###
boxplot(dados$average,
        main = "Boxplot da variavel 'average'",
        ylab = "Valores de 'average'",
        col = "lightgreen",
        border = "darkgreen")

#----------------------------------------#

# Calcular estatisticas descritivas para 'volume'
summary_stats <- summary(dados$volume)

iqr_volume <- IQR(dados$volume)
mean_volume <- mean(dados$volume)
sd_volume <- sd(dados$volume)
skewness_volume <- skewness(dados$volume)
kurtosis_volume <- kurtosis(dados$volume)

cat("Estatisticas descritivas para a variavel 'volume':\n")
cat("Valor minimo: ", summary_stats["Min."], "\n")
cat("Quartil 1 (Q1): ", summary_stats["1st Qu."], "\n")
cat("Mediana: ", summary_stats["Median"], "\n")
cat("Quartil 3 (Q3): ", summary_stats["3rd Qu."], "\n")
cat("Valor maximo: ", summary_stats["Max."], "\n")
cat("Desvio Interquartilico (IQR): ", iqr_volume, "\n")
cat("Media aritmetica: ", mean_volume, "\n")
cat("Desvio-padrao: ", sd_volume, "\n")
cat("Coeficiente de assimetria: ", skewness_volume, "\n")
cat("Coeficiente de curtose: ", kurtosis_volume, "\n")

### Boxplot ###
boxplot(dados$volume,
        main = "Boxplot da variavel 'volume'",
        ylab = "Valores de 'volume'",
        col = "lightgreen",
        border = "darkgreen")

#----------------------------------------#

### Analise bivariada ###

# Diagrama de Dispersao entre 'average' e 'volume'
plot(dados$average, dados$volume,
     main = "Diagrama de Dispersao: Average vs Volume",
     xlab = "Average",
     ylab = "Volume",
     col = "blue",
     pch = 16,
     cex = 0.6)

# Calcular a correlacao entre 'average' e 'volume'
correlation <- cor(dados$average, dados$volume)
```

```r
cat("O valor da correlacao entre 'average' e 'volume' e:",
    ↪ correlation, "\n")

# Ajustar uma linha de tendencia (regressao linear)
modelo <- lm(volume ~ average, data = dados)
abline(modelo, col = "red", lwd = 2)

legend("topright", legend = "Linha de Tendencia", col = "red", lwd
    ↪ = 2, bty = "n")

if (correlation > 0) {
  cat("A correlacao e positiva.\n")
} else if (correlation < 0) {
  cat("A correlacao e negativa.\n")
} else {
  cat("Nao ha correlacao aparente.\n")
}
```

Listing 1: Código R para análise univariada e bivariada.

$$x_{\text{norm}} = \frac{2 \cdot (x - \min)}{\max - \min} - 1$$