Universidade de São Paulo
Escola Politécnica, Engenharia de Produção

PRO3200 - Estatística
06 de abril de 2024

Análise de Dados e Otimização de Markowitz para Construção de Portfólios de Ações

Felipe Carneiro Rodrigues – 12566585
Henrique Fuga Duran – 12553570
Gabriel Corteletti Prezotti Palassi – 11820242
Miguel Shiniti Aguena – 12554230

Prof. Dra. Celma de Oliveira Ribeiro

Contents

I. Objetivos e Delimitação do Escopo

O presente projeto tem como objetivo realizar uma análise e otimização de um portfólio de investimento utilizando a Otimização de Markowitz. O escopo da análise inclui a seleção de quatro ações da bolsa de valores brasileira (PETR4, WEGE3, ABEV3 e VALE3), abrangendo o período de 01 de janeiro de 2000 até 05 de abril de 2024, conforme os dados disponíveis no banco de dados gerado.

A. Objetivos

O objetivo principal deste estudo é implementar computacionalmente a Otimização de Markowitz para identificar as distribuições que oferecem o melhor equilíbrio entre risco e retorno, conhecida como Fronteira Eficiente. Além disso, busca-se entender os movimentos históricos individuais dos ativos através de análises de dados detalhadas.

Para atingir esse objetivo, serão realizadas as seguintes atividades:

- Coleta e organização dos dados históricos de preços de fechamento ajustados para PETR4, WEGE3, ABEV3 e VALE3.
- Cálculo de métricas estatísticas descritivas, incluindo média, mediana, desvio padrão, variância, moda e retorno médio anual.
- Análise individual de cada ativo, utilizando gráficos de evolução temporal e histogramas de preços e retornos, além de testes de hipótese sobre os retornos esperados.
- Realização de ANOVA de dois fatores para avaliar os efeitos de contextos políticos e sociais nos preços dos ativos, dividindo os dados em períodos relevantes da história brasileira.
- Aplicação de regressão linear nos preços dos ativos para cada período analisado na ANOVA de dois fatores.
- Construção das matrizes de correlação e covariância para entender a relação entre os ativos.
- Realização de ANOVA de um fator para identificar diferenças significativas entre as médias dos retornos diários dos ativos.
- Construção de múltiplas Fronteiras Eficientes de Markowitz utilizando o método de Monte Carlo para gerar valores aleatórios dentro das distribuições das covariâncias, criando um intervalo de confiança para as fronteiras de distribuições otimizadas.

1. Definições

a. Média

A média é uma medida de tendência central que representa o valor médio dos preços de fechamento do ativo ao longo do período analisado. É calculada pela soma de todos os valores dividida pelo número total de observações. Matematicamente, a média ($\bar{x}$) é expressa como:

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

b. Mediana

A mediana é o valor central de um conjunto de dados ordenados. Ela é menos sensível a valores extremos do que a média e oferece uma visão mais robusta da distribuição dos preços de fechamento do ativo.

c. Variância

A variância é o quadrado do desvio padrão e representa a dispersão dos preços de fechamento em torno da média. Ela fornece uma medida mais direta da volatilidade do ativo. A variância (σ^2) é dada por:

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

d. Desvio Padrão

O desvio padrão (σ) é uma medida de dispersão que indica a magnitude da variação dos preços de fechamento em relação à média. Quanto maior o desvio padrão, maior a volatilidade do ativo. A fórmula do desvio padrão é:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}} \tag{3}$$

Ou, simplesmente a raiz quadrada da variância.

e. Moda

A moda é o valor que aparece com maior frequência no conjunto de dados. No contexto dos preços de fechamento do ativo, a moda pode fornecer insights sobre os valores mais comuns observados durante o período analisado.

f. Histograma

Um histograma é uma representação gráfica de dados que permite visualizar a distribuição de frequência de um conjunto de observações. No caso de um histograma de retornos diários, ele mostra a frequência com que cada faixa de retorno ocorre.

g. Gráfico de Evolução Temporal do Preço do Ativo

O gráfico de evolução temporal do preço do ativo é uma representação visual da variação dos preços de fechamento ao longo do tempo. No eixo horizontal, são plotadas as datas ou períodos de tempo, enquanto no eixo vertical são representados os valores dos preços de fechamento.

h. Teste de Hipótese

Um teste de hipótese é uma técnica estatística utilizada para tomar uma decisão entre duas hipóteses concorrentes baseadas em dados amostrais. Ele envolve a formulação de uma hipótese nula (H_0) e uma hipótese alternativa (H_1), onde o objetivo é determinar se há evidência suficiente nos dados para rejeitar H_0 em favor de H_1.

i. ANOVA (Análise de Variância)

A ANOVA é uma técnica estatística utilizada para comparar as médias de três ou mais grupos independentes. Ela avalia se há diferenças estatisticamente significativas entre as médias dos grupos, através da análise da variabilidade dentro dos grupos e entre os grupos. A ANOVA parte do pressuposto de que as amostras são independentes e que as populações subjacentes seguem uma distribuição normal. Existem diferentes tipos de ANOVA, como ANOVA de um fator (para comparar médias de dois ou mais grupos) e ANOVA de dois fatores (para examinar o efeito simultâneo de duas variáveis categóricas nos dados). O resultado principal da ANOVA é o valor F, que compara a variabilidade entre os grupos com a variabilidade dentro dos grupos. Se o valor F calculado for maior do que o valor crítico da distribuição F, então concluímos que há pelo menos um par de médias de grupo que são estatisticamente diferentes umas das outras.

B. Delimitação do Escopo

A análise será focada nas quatro ações mencionadas, representando setores importantes da economia brasileira: energia (PETR4), industrial (WEGE3), consumo (ABEV3) e mineração (VALE3). O período de estudo é de 01 de janeiro de 2000 até 05 de abril de 2024, permitindo uma análise aprofundada dos dados históricos.

Serão utilizados diversos métodos estatísticos e financeiros, incluindo:

- Retorno esperado e risco.
- Correlação entre ativos.
- ANOVA para análise de variâncias.
- Testes de hipótese para validar suposições sobre os dados.
- Construção da Fronteira Eficiente para otimização do portfólio.

Para a análise individual dos ativos, serão utilizadas métricas descritivas e análises de tendências, bem como gráficos representativos de comportamento, como média, mediana, desvio padrão, histogramas e evolução temporal de retorno e preço.

C. Resultados Esperados

Espera-se que, ao final do projeto, seja possível:

- Desenvolver aparelhagem embasada em dados históricos que permita a escolha fundamentada dos ativos para uma carteira fictícia de ações.
- Utilizar os ativos escolhidos para montar um portfólio otimizado, maximizando o retorno esperado para um determinado nível de risco ou minimizando o risco para um determinado nível de retorno.
- Compreender os movimentos passados dos ativos em resposta a eventos específicos, auxiliando na previsão de comportamentos futuros.

Esta abordagem fornecerá uma base sólida para decisões de investimento, utilizando a Otimização de Markowitz e outras técnicas estatísticas e econométricas para maximizar a eficiência do portfólio.

II. Introdução Teórica

A. Otimização de Markowitz e Teoria Moderna de Portfólio

A Otimização de Markowitz, também conhecida como Teoria Moderna de Portfólio (MPT, do inglês Modern Portfolio Theory), é uma abordagem desenvolvida pelo economista Harry Markowitz em 1952 para otimizar a alocação de ativos em um portfólio, considerando o equilíbrio entre risco e retorno esperado. Essa teoria revolucionou a maneira como os investidores pensam sobre diversificação de portfólio e continua sendo um dos conceitos fundamentais da gestão de investimentos.

Para entender a Otimização de Markowitz, é essencial primeiro compreender alguns conceitos-chave:

1. **Ativo Financeiro**: Qualquer instrumento financeiro que tenha valor monetário, como ações, títulos, commodities, entre outros.
2. **Retorno**: O ganho ou perda de um investimento ao longo de um determinado período de tempo. Pode ser expresso em termos percentuais ou absolutos.
3. **Risco**: A incerteza associada ao retorno de um investimento. Geralmente, é medido pela variabilidade dos retornos em relação à média. Quanto maior a variabilidade, maior o risco.
4. **Retorno Esperado**: O retorno médio que um investidor espera obter de um ativo ao longo de um período de tempo. É calculado ponderando os possíveis retornos pelo grau de probabilidade de ocorrência de cada um.
5. **Covariância**: Uma medida estatística que indica como duas variáveis se movem em relação uma à outra. No contexto financeiro, a covariância entre dois ativos é usada para medir como seus retornos se movem juntos.
6. **Correlação**: Uma medida estatística normalizada da relação entre duas variáveis. No contexto financeiro, a correlação entre dois ativos indica o grau de interdependência de seus retornos. Os valores variam de -1 a 1, onde:
   - 1 significa uma correlação positiva perfeita.
   - -1 significa uma correlação negativa perfeita.
   - 0 significa ausência de correlação linear.

1. *Retorno Esperado de um Portfólio (μ)*

O retorno esperado de um portfólio é uma média ponderada dos retornos dos ativos individuais, onde os pesos representam a proporção de cada ativo no portfólio.

A fórmula é dada por:

$$\mu_p = \sum_{i=1}^{n} w_i \times \mu_i$$

Onde:

- μ_p: Retorno esperado do portfólio.
- w_i: Peso do ativo i no portfólio.
- μ_i: Retorno esperado do ativo i.
- n: Número de ativos no portfólio.

2. *Risco de um Portfólio (σ)*

O risco de um portfólio é representado pela volatilidade dos retornos do portfólio. Markowitz considerou o desvio padrão dos retornos como uma medida de risco.

A fórmula é dada por:

$$\sigma_p = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} w_i \times w_j \times \sigma_{ij}}$$

Onde:

- σ_p: Risco do portfólio.
- w_i e w_j: Pesos dos ativos i e j no portfólio, respectivamente.
- σ_{ij}: Covariância entre os ativos i e j.
- n: Número de ativos no portfólio.

Mas por que se mede risco pela volatilidade ou desvio padrão? A volatilidade é uma medida fundamental de risco nos mercados financeiros. Ela captura a magnitude das flutuações dos preços dos ativos em torno de sua média histórica. A principal razão pela qual a volatilidade é usada como medida de risco na Otimização de Markowitz é que os investidores geralmente desejam evitar grandes oscilações no valor de seus investimentos. A volatilidade reflete precisamente essa característica, indicando o quanto os retornos de um ativo podem se desviar de seu retorno médio.

3. Fronteira Eficiente (Curva de Markowitz)

A Fronteira Eficiente é uma representação gráfica de todas as possíveis combinações de portfólios que oferecem o maior retorno esperado para um determinado nível de risco, ou o menor risco para um determinado nível de retorno. Em outras palavras, mostra a melhor relação possível entre risco e retorno que um investidor pode alcançar, dadas as opções de investimento disponíveis.

Para construir a Fronteira Eficiente, é necessário considerar a relação entre os diferentes ativos no portfólio. A diversificação é fundamental aqui, pois a combinação de ativos com baixa correlação reduzirá o risco total do portfólio, permitindo que o investidor alcance um determinado nível de retorno com menos volatilidade.

A Fronteira Eficiente é geralmente curva e convexa para cima. Isso significa que, para alcançar um retorno mais alto, um investidor precisa assumir mais risco. No entanto, uma vez que a diversificação eficaz pode reduzir o risco sem sacrificar o retorno, os investidores podem encontrar portfólios na Fronteira Eficiente que ofereçam um equilíbrio ideal entre risco e retorno para suas preferências individuais.

Para formalizar a fronteira eficiente, é necessário resolver o problema de otimização para encontrar os pesos ideais dos ativos na carteira. Supondo n ativos, cada um com um retorno esperado μ_i e um desvio padrão σ_i, uma matriz de covariância Σ que contém as covariâncias entre os retornos dos ativos. O problema de otimização é expresso em 4 como:

$$\text{Maximizar } \mu_p = w^T\mu \quad \text{sujeito a:} \quad w^T 1 = 1 \text{ e } w^T\Sigma w = \sigma_p^2 \tag{4}$$

Em que:

- $\mathbf{w}$ é um vetor de pesos dos ativos na carteira,
- $\boldsymbol{\mu}$ é um vetor de retornos esperados dos ativos,
- $\mathbf{1}$ é um vetor de uns,
- $\mathbf{\Sigma}$ é a matriz de covariância dos retornos dos ativos, e
- σ_p^2 é a variância da carteira.

A primeira restrição ($w^T 1 = 1$) garante que a soma dos pesos dos ativos seja igual a 1, refletindo o fato de que a carteira é totalmente investida. A segunda restrição ($w^T\Sigma w = \sigma_p^2$) limita a variância da carteira ao nível desejado.

Resolvendo este problema de maximização sujeito a restrições, pode-se obter os pesos ótimos dos ativos na carteira eficiente em termos de Markowitz. Esses pesos definem um ponto na fronteira eficiente que representa o máximo retorno que pode ser alcançado para um dado nível de risco.

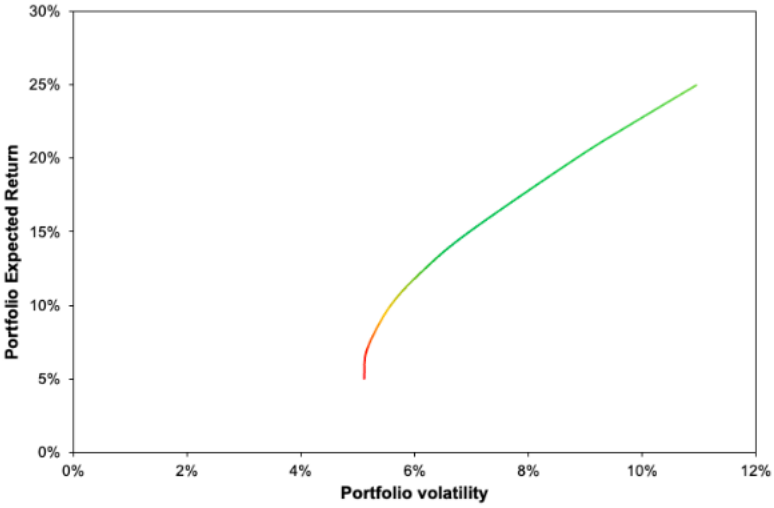


Figura 1: Exemplo de fronteira eficiente de Markowitz

Assim, a fronteira eficiente proporciona uma visão clara das escolhas ótimas de carteira disponíveis, permitindo tomar decisões informadas sobre como equilibrar risco e retorno nos investimentos.

A Otimização de Markowitz permite que os investidores construam portfólios que maximizem o retorno esperado para um nível de risco aceitável ou minimizem o risco para um determinado retorno esperado. Essa abordagem enfatiza a diversificação para reduzir o risco total do portfólio, explorando a relação entre os ativos para encontrar a combinação ótima que equilibra risco e retorno.

III. Metodologia

Este estudo segue uma abordagem sistemática e estruturada para analisar as ações da bolsa de valores brasileira. A metodologia foi desenvolvida para garantir uma análise abrangente e precisa dos dados de mercado, utilizando diversas técnicas estatísticas e econométricas. A seguir, detalhamos cada etapa do processo metodológico.

A. Coleta de Dados

Inicialmente, desenvolveu-se um script em Python utilizando a biblioteca `yfinance` para extrair os preços de fechamento de quatro ações brasileiras: PETR4 (Petrobras), WEGE3 (Weg S.A.), ABEV3 (Ambev) e VALE3 (Vale S.A.). A escolha dessas ações se deu pela representatividade das mesmas em diferentes setores da economia brasileira: energia, industrial, consumo e mineração, respectivamente. O período de dados abrange de 01 de janeiro de 2000 até 05 de abril de 2024. Este amplo intervalo temporal permite uma análise robusta das dinâmicas de mercado e da performance dos ativos ao longo de diversas fases econômicas.

B. Construção do DataFrame e Armazenamento

Os dados coletados foram organizados em um *DataFrame* contendo as séries temporais dos preços de fechamento ajustados para cada ativo. Essa estrutura de dados foi escolhida por sua eficiência e flexibilidade na manipulação e análise de grandes volumes de dados temporais. Em seguida, os dados foram armazenados em um arquivo CSV denominado `assets.csv`, assegurando que as informações estejam prontamente disponíveis para análises subsequentes.

C. Cálculo de Métricas Estatísticas

Com o `assets.csv` em mãos, calculou-se uma série de métricas estatísticas iniciais para cada ativo, incluindo média, mediana, desvio padrão, variância, moda e retorno médio anual. Essas métricas fornecem uma visão geral das características estatísticas dos dados, ajudando a entender a distribuição e a volatilidade dos preços, além de estabelecer uma base para comparações futuras.

D. Análise Individual dos Ativos

Cada ativo foi analisado individualmente, começando por um entendimento do setor de atuação e do desempenho histórico da empresa. Foram gerados gráficos de evolução temporal dos preços, bem como histogramas de preço e retorno, para visualizar as tendências e a distribuição dos dados. Adicionalmente, realizou-se um teste de hipótese sobre o retorno esperado, permitindo avaliar se os retornos médios observados diferem significativamente de um valor de referência pré-determinado, e um teste de hipótese de 2 parâmetros, a fim de validar se a variância dos ativos de setores semelhantes (PETR4 e VALE3, ambos commodities) possuiam variância estatisticamente igual.

E. ANOVA de Dois Fatores

Para analisar os efeitos dos contextos políticos e sociais no preço dos ativos, realizou-se uma ANOVA de dois fatores, dividindo os dados em períodos relevantes da história brasileira. Esta análise permite identificar se mudanças nos contextos políticos e sociais estão associadas a variações significativas nos preços dos ativos, fornecendo insights sobre a sensibilidade dos ativos às mudanças externas.

F. Regressão Linear

Complementando a ANOVA, realizou-se uma regressão linear nos preços dos ativos para cada período dividido na ANOVA de dois fatores. A regressão linear ajuda a quantificar a relação entre as variáveis dependentes e independentes, permitindo uma análise mais detalhada dos efeitos observados.

G. Matrizes de Correlação e Covariância

Para entender a relação entre os ativos, foram construídas as matrizes de correlação e covariância. A matriz de correlação mostra a direção e a intensidade da relação linear entre os preços dos ativos, enquanto a matriz de covariância quantifica o grau de variação conjunta entre os ativos. Estas matrizes são fundamentais para a construção de portfólios otimizados.

H. ANOVA de Um Fator

Realizou-se uma ANOVA de um fator para identificar se existem diferenças significativas entre as médias dos retornos diários dos ativos. Esta análise ajuda a entender se algum ativo possui um comportamento diferenciado em termos de retorno médio diário quando comparado aos demais.

I. Fronteiras Eficientes de Markowitz

Por fim, foram construídas múltiplas fronteiras eficientes de Markowitz para criar um intervalo de confiança das fronteiras de distribuições otimizadas. Utilizou-se o método de Monte Carlo para gerar valores aleatórios dentro das distribuições das covariâncias, permitindo a montagem de diversas fronteiras eficientes. Este procedimento visa capturar a variabilidade e a incerteza na construção de portfólios otimizados, proporcionando uma visão mais abrangente das possíveis combinações de ativos.

Esta metodologia detalhada garante uma análise rigorosa e abrangente dos dados financeiros, utilizando uma combinação de técnicas estatísticas e econométricas para fornecer insights profundos sobre o comportamento dos ativos analisados.

IV. Análise dos Ativos

Na Tabela I abaixo, constam os valores das estatísticas descritivas calculadas:

Tabela I: Medidas de posição dos ativos

	PETR4	WEGE3	ABEV3	VALE3
Média (R$)	7.77	8.03	7.55	23.1
Mediana (R$)	6.45	2.39	6.46	17.9
Desvio Pad. (R$)	6.66	11.7	5.79	20.9
Variância (R$^{2})	44.4	138	33.5	435
Moda (R$)	1.75	0.794	0.735	1.75
Retorno médio anual (%)	13.5%	24%	15.5%	15.3%

Observa-se que a média dos preços de fechamento das ações PETR4, WEGE3, ABEV3 e VALE3 é de 7.77, 8.03, 7.55 e 23.1 reais, respectivamente. Entretanto, é importante ressaltar que a dispersão dos dados em relação à média é significativa, como indicado pelos desvios padrão maiores (em proporção com o preço médio) para PETR4, WEGE3, ABEV3 e VALE3, respectivamente.

A mediana, que representa o valor central dos dados, mostra uma imagem ligeiramente diferente, sugerindo uma distribuição assimétrica dos preços de fechamento, especialmente evidente nas ações WEGE3 e VALE3, em que a diferença entre a média e a mediana é mais pronunciada.

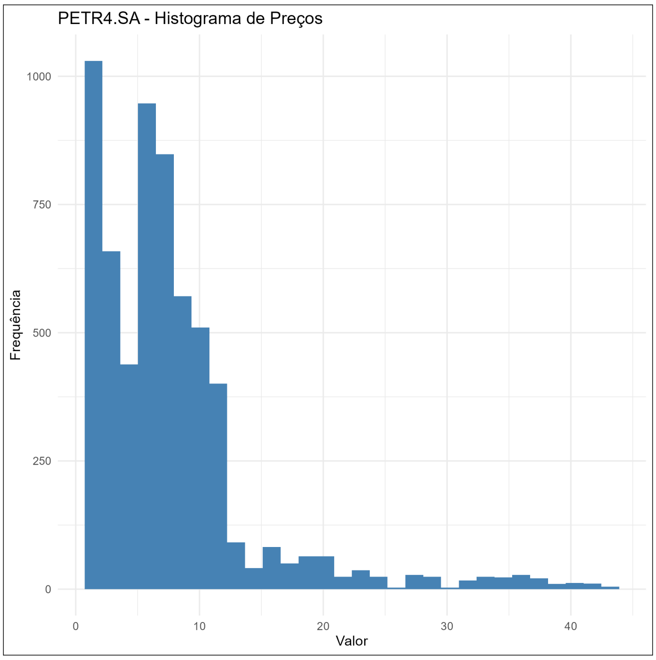
Ao considerar a moda, que representa o valor mais frequente nos dados, é possível verificar que os valores mais comuns de fechamento das ações são 1.75, 0.794, 0.735 e 1.75 reais, respectivamente para Petrobras, WEG, Ambev e Vale. Isso pode indicar a presença de picos de concentração de preços em determinados níveis.

Em termos de retorno médio anual, todas as ações apresentam valores positivos, com PETR4, ABEV3 e VALE3 mostrando retornos próximos, enquanto WEGE3 apresenta o maior retorno médio anual. Isso sugere que - no período considerado - os investidores obtiveram retornos positivos em suas posições nessas ações.

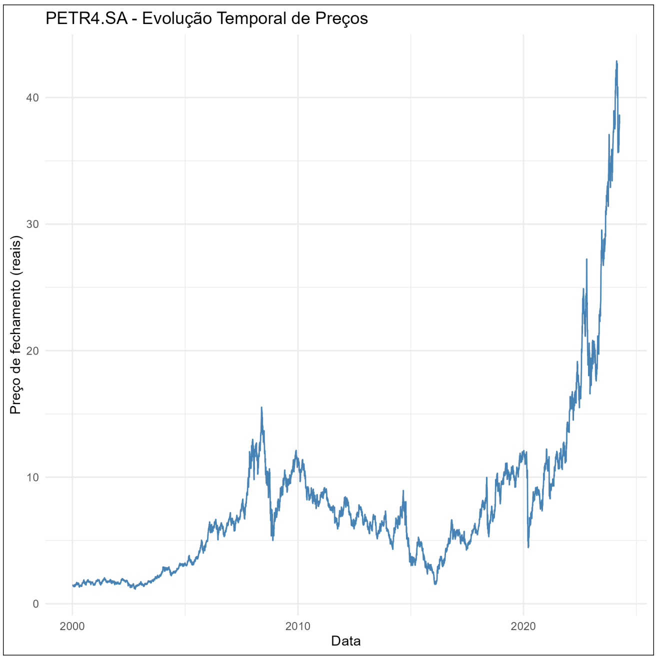
Por fim, ao considerar a variância dos preços de fechamento, observamos que as ações WEGE3 e VALE3 apresentam as maiores variâncias, indicando uma maior dispersão dos dados em torno da média, o que pode ser interpretado como maior volatilidade nos preços dessas ações. Em contrapartida, as ações PETR4 e ABEV3 exibem variâncias menores, sugerindo uma menor volatilidade em comparação com as outras ações analisadas.

A. PETR4.SA

A Petrobras é uma das maiores empresas de energia do mundo e uma das maiores companhias do Brasil. Fundada em 1953, é uma empresa estatal de capital aberto, com sede no Rio de Janeiro. Atua em diversos segmentos da indústria de energia, incluindo exploração, produção, refino, transporte e distribuição de petróleo, gás natural e outros produtos relacionados. A empresa desempenha um papel fundamental na economia brasileira, sendo responsável por uma parte significativa da produção nacional de energia, bem como por importantes investimentos em pesquisa e desenvolvimento no setor de energia. Seu impacto abrange desde a geração de empregos até a contribuição para a balança comercial do país.



(a) Histograma de preços



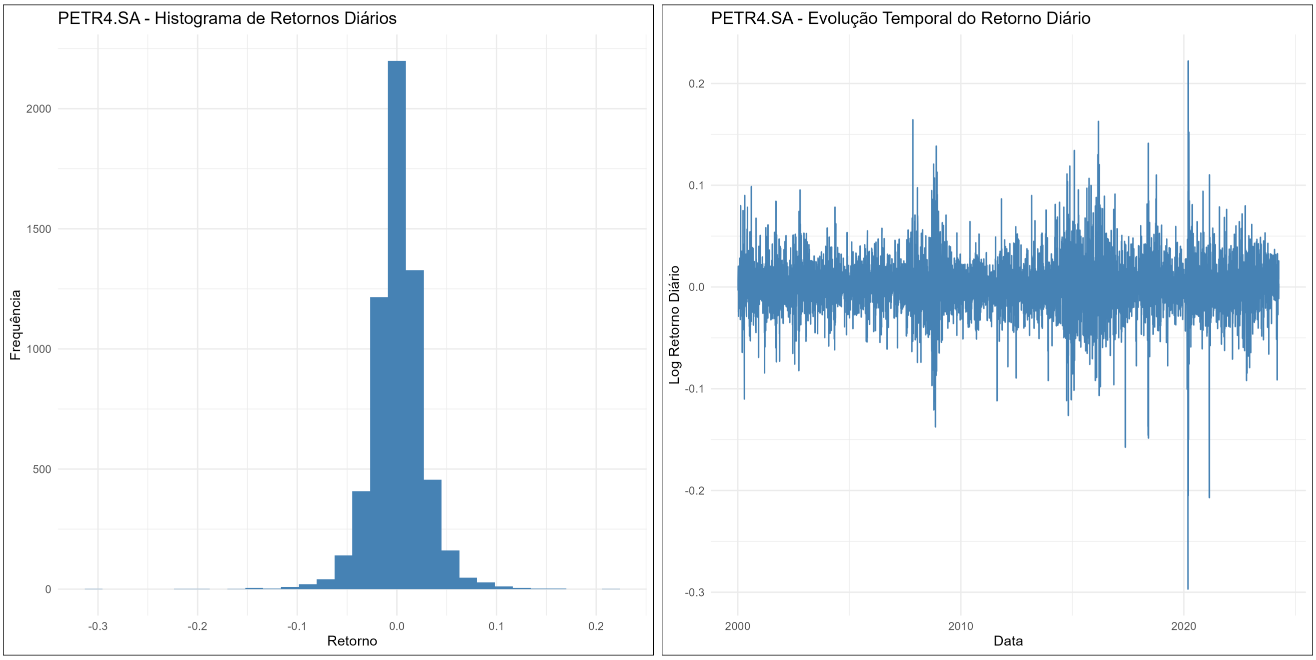
(b) Evolução temporal de preços

Figura 2: Gráficos de preço - PETR4.SA

Primeiramente, ao observar o formato geral do histograma, pode-se identificar que os preços de fechamento não seguem uma distribuição normal. Além disso, existe uma concentração maior de frequências ao redor dos preços mais baixos das ações, menores do que vinte reais, no caso. Isto se deve ao fato de a Petrobras ser uma das empresas mais antigas na B3, realizando seu IPO (Oferta Pública Inicial) em 1957 após a privatização parcial.

Nesse sentido, os efeitos de contingência da inflação ao longo dos governos na redemocratização – com as políticas de trocas de moedas – colaboram para manter preços antigos historicamente mais baixos em comparação com os períodos de inflação recentes.

Em complemento, os picos significativos no histograma indicam momentos de alta volatilidade ou eventos específicos que impactaram os preços das ações da Petrobrás. Por exemplo, durante períodos de instabilidade econômica global, crises geopolíticas do petróleo, mudanças regulatórias no setor de *oil and gas* ou eventos específicos relacionados à empresa, como descobertas de reservas significativas de petróleo, escândalos corporativos ou mudanças na gestão.



(a) Histograma de retornos diários

(b) Evolução temporal de retornos diários

Figura 3: Gráficos de retorno - PETR4.SA

O histograma dos retornos diários da PETR4 parece apresentar uma distribuição aproximadamente normal, concentrada em torno da mediana. A concentração na mediana indica que os retornos diários tendem a ser próximos desse valor na maioria dos dias.

No entanto, observa-se uma assimetria à esquerda, indicando uma maior frequência de retornos negativos em comparação com os positivos. Isso sugere que, em média, os retornos negativos são mais comuns do que os retornos positivos para a PETR4, o que ainda sim não implica retorno negativo no longo prazo.

Além disso, a extensão da cauda direita do histograma indica que existem poucos dias com retornos muito positivos. Por outro lado observa-se que a frequência dos intervalos parecem maiores do que aqueles da cauda esquerda. Isso pode sugerir a ocorrência de eventos ou notícias específicas que causam grandes movimentos de alta nos preços das ações da PETR4 em determinados dias.

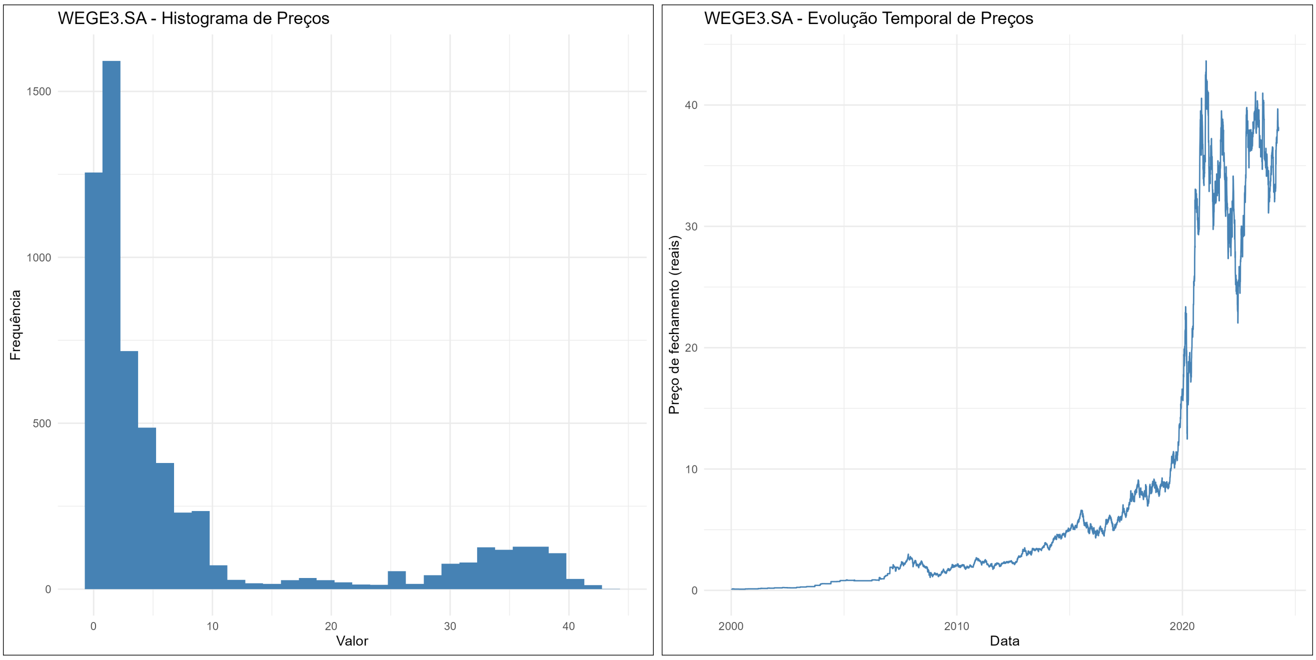
A evolução temporal dos retornos diários da PETR4 mostra uma tendência consistente ao longo do tempo, com exceção de alguns períodos de grande volatilidade. Notavelmente, esses períodos incluem o início do ano de 2020 e meados de 2008.

Durante o início de 2020, observam-se grandes picos tanto de alta quanto de baixa nos retornos diários da PETR4. Isso pode estar relacionado à volatilidade causada pela pandemia de COVID-19 e às incertezas associadas à sua impacto nos mercados financeiros.

Da mesma forma, os picos durante meados de 2008 podem ser atribuídos à crise financeira global que ocorreu naquele ano. Esses períodos de alta volatilidade podem indicar momentos de maior incerteza ou estresse no mercado, levando a movimentos bruscos nos preços das ações da PETR4.

B. WEGE3.SA

A WEG S.A. é uma multinacional brasileira fabricante de equipamentos eletroeletrônicos, incluindo motores elétricos e equipamentos industriais, e realizou sua entrada na B3 (então BM&FBOVESPA) através de um IPO em 1971.



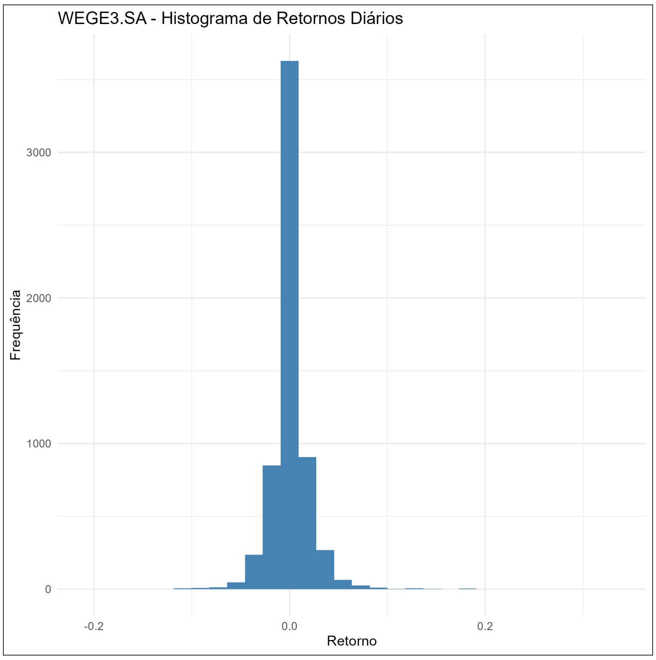
(a) Histograma de preços

(b) Evolução temporal de preços

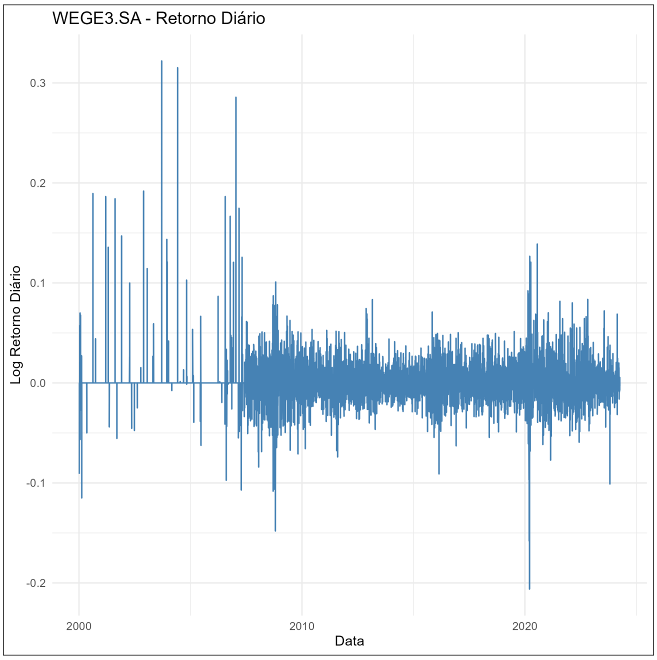
Figura 4: Gráficos de preço - WEGE3.SA

O histograma parece denotar um formato de curva trimodal. O primeiro pico em torno de R$4, 00 indica um período de valorização inicial das ações da WEGE3, que pode estar relacionado ao crescimento orgânico da empresa, expansão de sua base de clientes ou lançamento bem-sucedido de novos produtos no mercado. Essa fase provavelmente reflete a consolidação da empresa como um *player* significativo no setor industrial elétrico brasileiro.

Por outro lado, o último pico menor em torno de R$35, 00 sugere uma valorização posterior das ações da WEGE3. Esse aumento de preço pode ser atribuído a uma variedade de fatores, incluindo o desempenho financeiro sólido da empresa, a entrada em novos mercados ou a adoção de tecnologias inovadoras. Além disso, mudanças nas condições macroeconômicas e regulatórias do Brasil, bem como eventos específicos da empresa, como aquisições ou parcerias estratégicas, podem ter impulsionado esse movimento de alta nos preços das ações, mas mantendo uma relativa estabilidade nos preços recentes.



(a) Histograma de retornos diários



(b) Evolução temporal de retornos diários

Figura 5: Gráficos de retorno - WEGE3.SA

O histograma dos retornos diários da WEGE3 sugere uma distribuição muito concentrada em torno da mediana, ainda mais do que o observado para a PETR4. Isso indica que a maioria dos dias tem retornos próximos de zero, sugerindo uma baixa volatilidade nos retornos diários da WEGE3 em comparação com a PETR4.

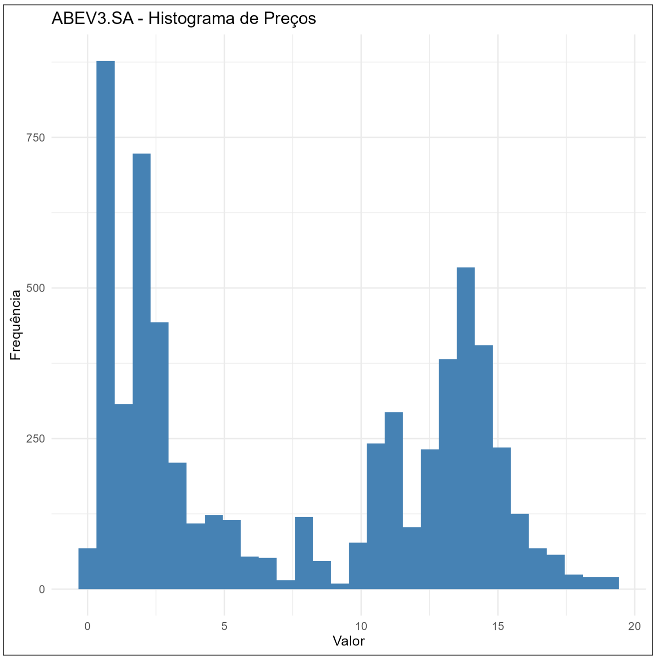
Além disso, a cauda direita do histograma parece ter um pouco mais de extensão do que a cauda esquerda, indicando a presença de alguns dias com retornos muito positivos, embora em menor frequência. Isso sugere que existem dias em que os preços das ações da WEGE3 experimentam ganhos significativos, possivelmente devido a eventos específicos ou notícias positivas relacionadas à empresa.

A evolução temporal dos retornos diários da WEGE3 mostra grandes altas dos anos 2000 até meados de 2004, com poucas baixas durante esse período. Isso pode estar associado a um período de crescimento ou boas perspectivas para a empresa nesse período específico.

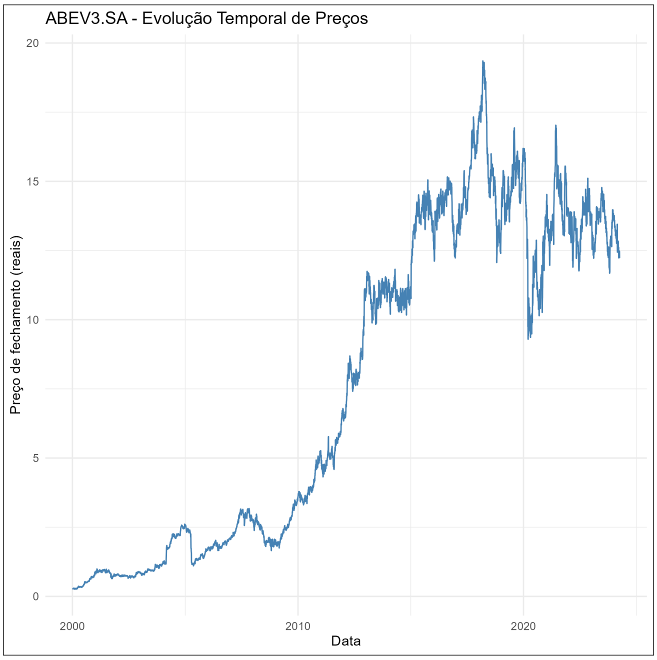
Além desse período, o gráfico parece ser consistente, indicando uma tendência geral de estabilidade nos retornos diários da WEGE3. No entanto, assim como observado para a PETR4, o início de 2020 parece ter sido marcado por uma maior volatilidade e picos nos retornos diários, possivelmente devido à pandemia de COVID-19 e às incertezas associadas a ela.

C. ABEV3.SA

A Ambev se destaca como uma das líderes no mercado de bebidas, tanto no Brasil quanto globalmente. Sua influência abrange uma ampla variedade de marcas reconhecidas, como Guaraná Antarctica e Pepsi, bem como uma extensa rede de distribuição. Desde sua oferta inicial na B3 em 1998, quando era conhecida como Companhia de Bebidas das Américas (AmBev), o desempenho de suas ações é moldado por diversos fatores, incluindo resultados financeiros, estratégias de negócios, cenário macroeconômico e mudanças nas tendências de consumo.



(a) Histograma de preços

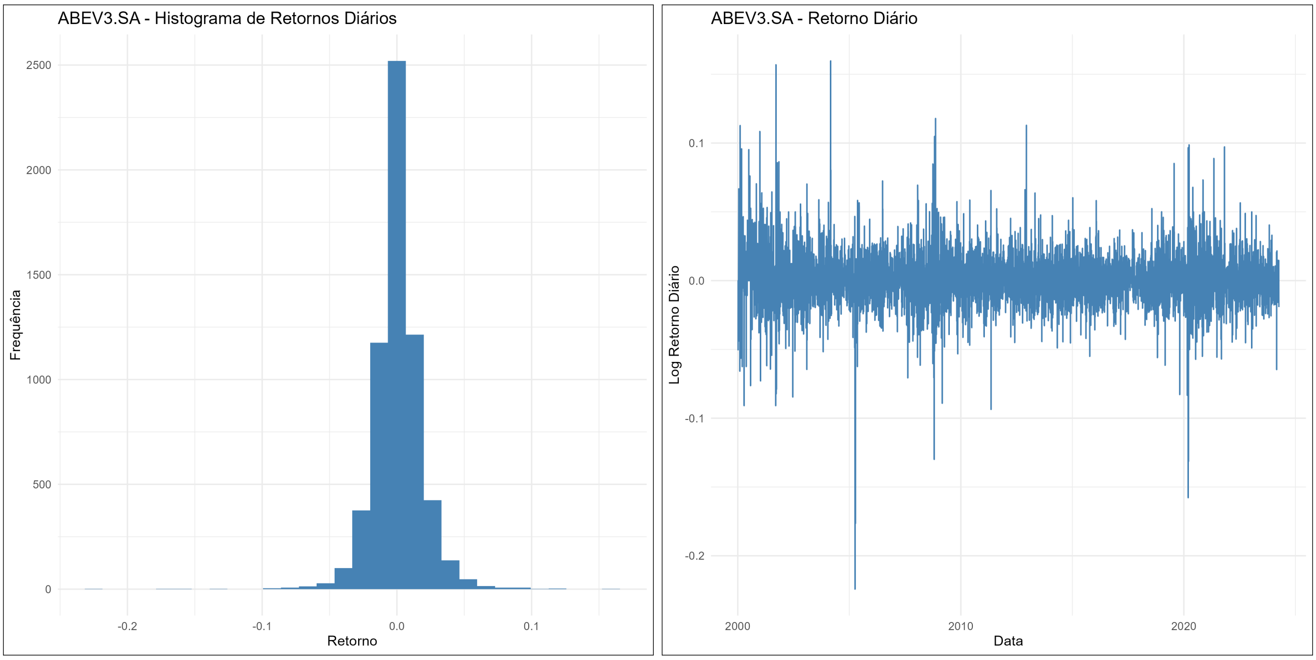


(b) Evolução temporal de preços

Figura 6: Gráficos de preço - ABEV3.SA

O histograma denota um formato de curva bimodal. O primeiro pico, na esquerda do gráfico, sugere um período inicial de volatilidade ou estabilização dos ativos ABEV3 após sua formação, possivelmente relacionado ao processo de *M&A* (fusão e aquisição) de várias marcas que deram origem à empresa. Essa fase inicial pode ter sido marcada por ajustes no mercado em resposta à consolidação da Ambev como uma das principais empresas de bebidas do Brasil.

O segundo pico em torno de R$12,00 indica um período posterior de valorização das ações. Essa elevação nos preços pode ser atribuída a diversos fatores, incluindo o lançamento bem-sucedido de novos produtos, expansão para mercados internacionais ou melhorias no desempenho financeiro da empresa. Além disso, mudanças nas políticas econômicas, flutuações cambiais e eventos específicos da empresa, como a introdução de estratégias de marketing inovadoras, podem ter contribuído para esse aumento no valor das ações.



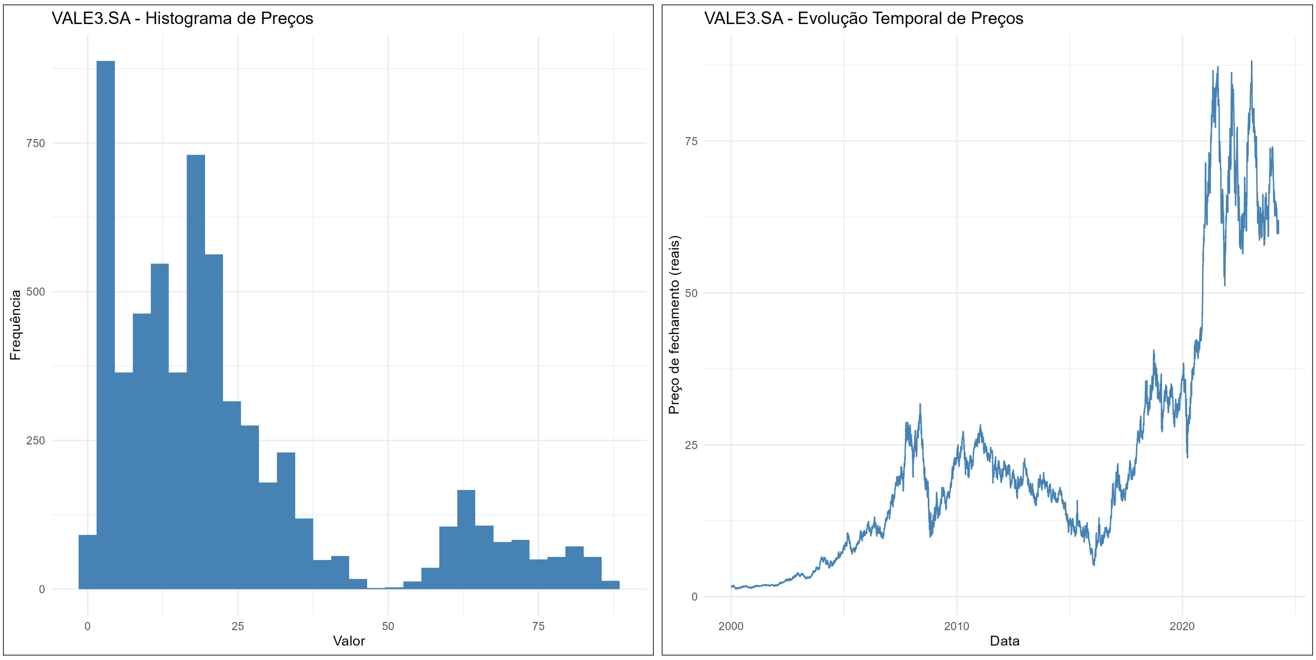
(a) Histograma de retornos diários

(b) Evolução temporal de retornos diários

Figura 7: Gráficos de retorno - ABEV3.SA

D. VALE3.SA

A Vale, anteriormente conhecida como Companhia Vale do Rio Doce, é uma das maiores empresas de mineração do mundo e uma das principais produtoras de minério de ferro. Fundada em 1942, a Vale tem sua sede no Brasil e possui operações em diversos países ao redor do globo. Além da mineração de minério de ferro, a empresa também está envolvida na produção de outros metais e minerais, como níquel, cobre, carvão, manganês e fosfatos. A empresa desempenha um papel crucial na economia brasileira e global, fornecendo matérias-primas essenciais para diversas indústrias, desde a construção civil até a produção de aço.



(a) Histograma de preços

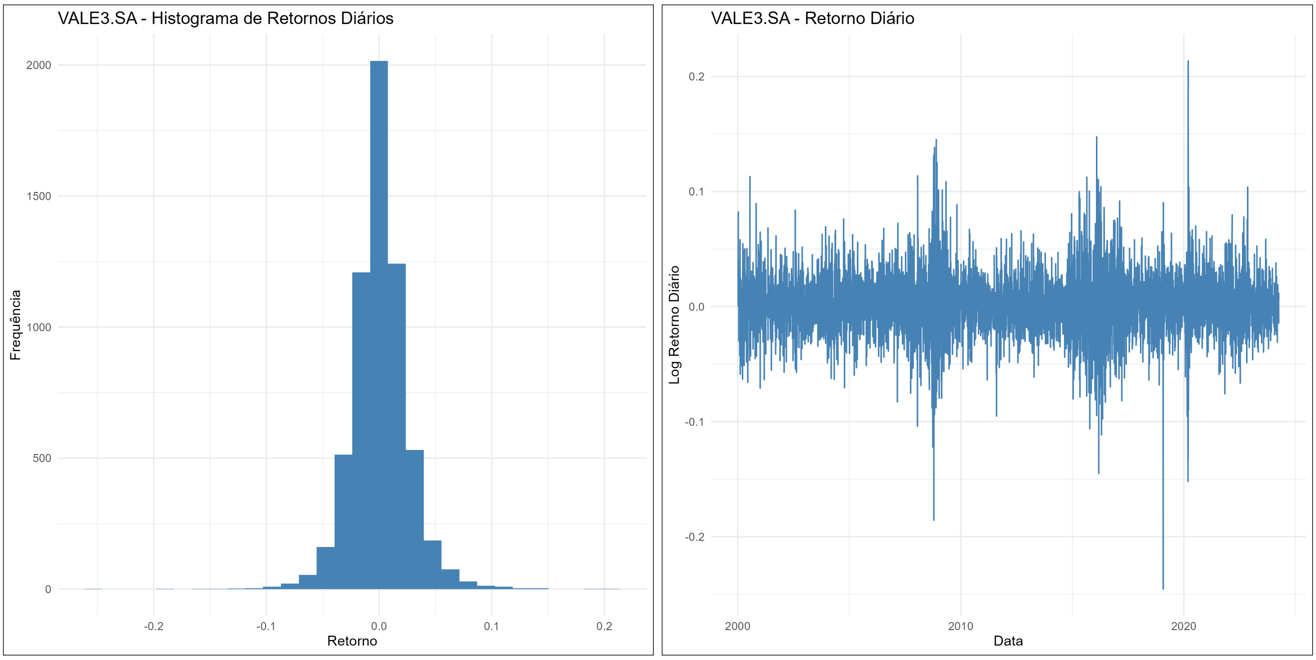
(b) Evolução temporal de preços

Figura 8: Gráficos de preço - VALE3.SA

Analogamente ao histograma de preços da ABEV3.SA esse histograma parece denotar uma curva bimodal. Inicialmente, o primeiro pico da curva do histograma em torno de R$5,00 pode ser associado ao período após o IPO da Vale na B3, em meados de 2000, quando as ações estavam sendo precificadas e havia grande interesse dos investidores. Os desvios subsequentes, como o de R$24,00, podem estar relacionados a momentos de alta demanda por minério de ferro, um dos principais produtos da Vale, impulsionando os preços das ações.

As oscilações posteriores podem refletir eventos específicos que afetaram a empresa. Por exemplo, o período após os trágicos acidentes com barragens em Mariana (2015) e Brumadinho (2019), que resultaram em perdas humanas e ambientais significativas, além de implicações legais e financeiras para a Vale. Esses desastres abalaram a confiança dos investidores e afetaram negativamente a reputação da empresa, levando a uma queda nos preços das ações.

Por outro lado, o seguinte crescimento de frequências em torno de R$62,00 pode refletir períodos de recuperação da Vale, à medida que a empresa implementava medidas para mitigar os riscos associados à segurança de suas operações e buscava reconstruir a confiança do mercado. O platô final poderia indicar uma estabilização subsequente, onde os investidores avaliaram a capacidade da empresa de lidar com os desafios enfrentados e anteciparam um cenário de retorno gradual à normalidade.



(a) Histograma de retornos diários

(b) Evolução temporal de retornos diários

Figura 9: Gráficos de retorno - VALE3.SA

O histograma dos retornos diários da VALE3 parece ser quase idêntico ao observado para a PETR4, com uma distribuição que se assemelha a uma curva normal, concentrada em torno da mediana. Isso indica que a maioria dos dias tem retornos próximos de zero, sugerindo uma distribuição simétrica dos retornos diários da VALE3.

A análise supracitada do hisograma de retornos da PETR4 é aqui também aplicável.

A evolução temporal dos retornos diários da VALE3 mostra picos de altas e baixas em momentos-chave da história recente. Em particular, observam-se picos de altas e baixas em meados de 2008, 2017 e início de 2020.

Os picos em meados de 2008 estão associados à crise financeira global daquele ano, que teve um impacto significativo nos mercados financeiros globais, incluindo o setor de mineração e recursos naturais, no qual a VALE3 está inserida.

Os picos em 2015 podem estar relacionados a eventos políticos, como eleições, que muitas vezes introduzem incertezas nos mercados financeiros e podem levar a movimentos bruscos nos preços das ações, ou então ao desastre de Mariana.

O forte pico de baixa em meados de 2019 está relacionado ao desastre de Brumadinho, que teve um impacto devastador nas operações da Vale e resultou em grandes perdas para a empresa.

2020 parece novamente estar relacionado a pandemia de COVID-19.

E. Testes de Hipótese para Retornos Diários dos Ativos Financeiros

Ao analisar a performance histórica individual dos ativos financeiros, investigamos agora se os retornos diários médios são estatisticamente maiores que zero. Para isso, realizamos testes de hipótese com um nível de significância de 5%.

1. Resultados dos Testes de Hipótese

Os testes foram conduzidos assumindo as seguintes hipóteses:

- **Hipótese Nula (H_0):** O retorno diário médio do ativo é menor ou igual a zero ($\mu \leq 0$).
- **Hipótese Alternativa (H_1):** O retorno diário médio do ativo é maior que zero ($\mu > 0$).

Os resultados dos testes estão resumidos na Tabela II.

2. Interpretação dos Resultados

Os testes de hipótese foram realizados para verificar se o retorno diário médio de cada ativo é maior que zero. Com um nível de significância de 5%, os resultados indicam que para todos os ativos (PETR4.SA, WEGE3.SA, ABEV3.SA e VALE3.SA), o valor-p é significativamente menor que α, levando à rejeição da hipótese nula em favor da hipótese alternativa.

Tabela II: Resultados dos Testes de Hipótese para Retornos Diários de Ativos Financeiros

Ativo	**Valor-p**	**Teste**	**Média**	H_0 **Rejeitada**
PETR4.SA	4.390802e-03	One Sample t-test	0.0008927126	Sim
WEGE3.SA	7.616696e-06	One Sample t-test	0.0011658924	Sim
ABEV3.SA	4.454978e-04	One Sample t-test	0.0007846840	Sim
VALE3.SA	2.153364e-03	One Sample t-test	0.0009246200	Sim

Esses resultados sugerem fortemente que os ativos apresentaram retornos diários médios positivos durante o período analisado. Investidores podem considerar essa informação ao formular estratégias de investimento, destacando que historicamente esses ativos têm tendência a gerar ganhos diários em média.

É importante observar que além dos retornos médios, outros fatores como volatilidade, tendências de mercado e condições econômicas devem ser considerados ao tomar decisões de investimento.

F. Análises de Variâncias para o Portfólio Selecionado - ANOVA Um Fator

A fim de identificar se existem diferenças signficativas entre a média dos retornos diários das diferentes ações. O primeiro passo é calcular os retornos diários de cada empresa dentre as listadas, usando a fórmula descrita em 5:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \tag{5}$$

com R_t sendo o retorno no dia t, P_t o preço de fechamento ajustado naquele mesmo dia, e P_{t-1} no dia anterior (para $t > 1$).

Nesse sentido, pode-se elaborar uma hipótese nula (H_0) de que **as médias dos retornos diários das ações são iguais** e uma hipótese alternativa (H_1) de que **pelo menos uma média dos retornos diários é diferente**. Matematicamente, isso pode ser expresso por 6:

$$\begin{cases} H_0 : \mu_{PETR4} = \mu_{WEGE3} = \mu_{ABEV3} = \mu_{VALE3} \\ H_1 : \exists\, i \neq j \text{ tal que } \mu_i \neq \mu_j \end{cases} \tag{6}$$

A **regra de decisão** é que, se o *p-value* da Análise de Variância for menor do que o nível de signficância – considerado $\alpha = 0.05 = 5\%$ –, então rejeita-se a hipótese nula H_0 e conclui-se que existe uma diferença significativa na média dos retornos diários entre pelo menos dois ativos. Caso contrário, não rejeita H_0. [1]

A validar as proposições, utiliza-se a Análise de Variância de 1 fator, descrita na Tabela III a seguir. Nela, estão descritos os valores da estatística F, calculada por:

$$F = \frac{\frac{n.\sum_{i=1}^{k}\left(\overline{X_i} - \overline{X}^2\right)}{k-1}}{\frac{\sum_{i=1}^{k}\sum_{j=1}^{n}\left(X_i j - \overline{X_i}\right)^2}{N-k}} \tag{7}$$

onde o *numerador* é a variância entre grupos, e o *denominador* é a variância dentro dos grupos. Também define-se:

- k é o número de grupos (ações).
- n é o número de observações em cada grupo.
- N é o número total de observações.
- $\overline{X_i}$ é a média dos retornos do grupo i.
- $\overline{X_i j}$ é a observação j do grupo i.

---

[1] Também poderia ser utilizada uma regra de decisão baseada no Valor Crítico: se valor-F calculado for *maior* do que o valor-$F_{\text{crítico}}$, obtido da tabela F com $\alpha = 0.05 = 5\%$, rejeita-se H_0.

- $\overline{X}$ é a média geral dos retornos.

A Soma de Quadrados Total (SST) é dada por:

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n}\left(X_i j - \overline{X_i}\right)^2 \tag{8}$$

A Soma de Quadrados Entre Grupos (SSB) é dada por:

$$SSB = \sum_{i=1}^{k} n\left(\overline{X_i} - \overline{X}^2\right) \tag{9}$$

A Soma dos Quadrados Dentro dos Grupos (SSW) é dada por:

$$SSW = \sum_{i=1}^{k}\sum_{j=1}^{n}\left(X_i j - \overline{X_i}\right)^2 \tag{10}$$

Os graus de liberdade são definidos abaixo, e permitem obter o valor-p a partir da distribuião F com df_B e df_W:

- Entre grupos: $df_B = k - 1 = 4 - 1$
- Dentro dos grupos: $df_W = N - k = (n_{PETR4} + n_{WEGE3} + n_{ABEV3} + n_{VALE3}) - 4$
- Total: $df_T = N - 1$

O Quadrado Médio entre grupos é $MS_B = \frac{SSB}{df_B}$, e o dentro dos grupos vale $MS_W = \frac{SSW}{df_W}$, possibilitando o cálculo da estatística F por:

$$F = \frac{MS_B}{MS_W} \tag{11}$$

Fonte de variação	**Soma dos Quadrados**	**Graus de Liberdade**	**Quadrado Médio**	valor-p	**valor-F**	**F crítico**
Entre linhas	0.0002381327	3	7.937758e-05			
Residual	2.8993055479	7228	4.011214e-04	0.8978767	0.1978892	2.606138
Total	2.8995436806	7231				

Tabela III: Análise de variâncias de 1 fator para as ações PETR4, WEGE3, ABEV3, VALE3

Como o valor-p=0.8978767 é maior do que $\alpha = 0.05$ [2], **não se pode rejeitar a hipótese nula**. Ou seja, as médias dos retornos diários de todas as ações são iguais, para o nível de signficância adotado com a ANOVA de um fator.

G. Efeitos Político-Sociais da História Brasileira - ANOVA Dois Fatores

Com o objetivo de analisar os efeitos político-sociais da história brasileira no preço de fechamento de ativos dessas empresas nacionais, buscou-se dividir o conjunto de dados em quatro grupos temporais. Os períodos escolhidos para análise econômica no Brasil foram selecionados com base em eventos históricos significativos – tais quais mudanças políticas, crises econômicas e eventos globais relevantes – que impactaram diretamente o mercado financeiro e, consequentemente, a variação dos preços diários das ações da Petrobras, WEG, Ambev e Vale.

---

[2] Além disso, neste caso, tem-se que valor-F=0.1978892 é menor do que $F_{\text{crítico}} = 2.606138$, indicando que não se pode rejeitar a hipótese nula

1. Divisão de Momentos Históricos do Brasil

No primeiro período, de 01/01/2000 e 31/12/2006, o Brasil experimentou a estabilidade econômica após a implementação bem-sucedida do Plano Real, seguida pela ascensão do governo de Luiz Inácio Lula da Silva. Este período é caracterizado por um ambiente de maior previsibilidade econômica e políticas que influenciaram diretamente empresas como as mencionadas, que são peças-chave na economia brasileira, especialmente no setor de *commodities* e industrial.

Na próxima fase, de 01/01/2007 a 31/12/2012, houve um período de expansão econômica robusta, seguido pela crise financeira global de 2008, que teve um impacto profundo nos mercados financeiros mundiais. A recuperação subsequente e as políticas econômicas adotadas influenciaram significativamente a dinâmica das ações dessas empresas, refletindo-se nos preços das ações durante esse período.

Entre 01/01/2013 e 31/08/2016, o Brasil enfrentou uma crise política e econômica intensa, que culminou no *impeachment* da presidente Dilma Rousseff. Esta fase foi marcada por instabilidade política e econômica, o que teve repercussões diretas nos mercados financeiros e, consequentemente, nos preços das ações das empresas analisadas.

Finalmente, de 01/09/2016 até os dias atuais (05/04/2024), o país atravessou um período de recuperação econômica gradual, intercalado com desafios significativos como a pandemia de COVID-19. Este período inclui a gestão de diferentes governos e políticas econômicas, que continuam a moldar o ambiente de investimento e influenciar a volatilidade nos preços das ações das empresas consideradas.

2. Cálculos para Análise de Variância de 2 Fatores

Para outra perspectiva da verificação do portfólio, deve-se verificar se existe interação significativa entre os intervalos temporais e os retornos das ações selecionadas. Dessa forma, formula-se três hipóteses nulas: para as ações, períodos e interações. A hipótese das ações (H_{0_A}), afirma que **as médias dos retornos das diferentes ações são iguais**. A dos períodos (H_{0_P}), sustenta que **as médias dos retornos nos diferentes períodos são iguais**. Finalmente, é feita uma hipótese da interação (H_{0_I}), segundo a qual **não há interação entre o valor do retorno da ação e o período**.

As *regras de decisão* são dadas pela fórmula 12 abaixo, para o nível de signficância $\alpha = 0.05 = 5\%$:

$$\begin{cases} \text{Rejeitar } H_{0_A} \text{ se } F_A \geq F_{\alpha,k-1,N-k} \\ \text{Rejeitar } H_{0_P} \text{ se } F_P \geq F_{\alpha,m-1,N-k} \\ \text{Rejeitar } H_{0_I} \text{ se } F_I \geq F_{\alpha,(k-1)(m-k),N-k} \end{cases} \tag{12}$$

Devem-se, portanto, calcular:

1. Soma dos Quadrados Total (SST): conforme a equação 8 supracitada;
2. Soma dos Quadrados para Ação (SSA): $SSA = \sum_{i=1}^{k} n\left(\overline{X_i} - \overline{X}\right)^2$
3. Soma dos Quadrados para Período (SSP): $SSP = \sum_{j=1}^{m} n\left(\overline{X_j} - \overline{X}\right)^2$, em que m é o número de períodos e $\overline{X_j}$ é a média dos retornos do período j.
4. Soma dos Quadrados da Interação (SSI): $SSI = SST - SSA - SSP$
5. Soma dos Quadrados Residual (SSR): $SSR = SST - SSA - SSP - SSI$

Para determinar os graus de liberdade, pode-se escrever:

- Para a ação: $df_A = k - 1$
- Para o período: $df_P = m - 1$
- Para a interação: $df_I = (k-1)(m-1)$
- Residual: $df_R = N - k.m$

Os quadrados médios, por sua vez, são definidos por:

- Para a ação: $MS_A = \frac{SSA}{df_A}$
- Para o período: $MS_P = \frac{SSP}{df_P}$
- Para a interação: $MS_I = \frac{SSI}{df_I}$
- Residual: $MS_R = \frac{SSR}{df_R}$

As estatísticas F, portanto, são calculadas por $F_A = \frac{MS_A}{MS_R}$ para a ação, $F_P = \frac{MS_P}{MS_R}$ para o período e – finalmente – $F_I = \frac{MS_I}{MS_R}$ para a interação.

Com as expressões elaboradas acima, pode-se construir a Tabela IV a seguir:

Fonte de variação	**Soma de Quadrados**	**Graus de liberdade**	**Quadrado Médio**	**valor-F**	**valor-p**
Período	3.851015e-03	3	0.0012836715	2.4205562	0.064035642
Ações	6.298873e-04	3	0.0002099624	0.3959158	0.755945570
Interação	1.229363e-02	9	0.0013659587	2.5757210	0.005817588
Residual	1.290801e+01	24340	0.0005303209		
Total	1.292479e+01	24355			

Tabela IV: Tabela ANOVA de dois fatores para interação entre períodos e retornos dos ativos PETR4, WEGE3, ABEV3 e VALE3

Tomando em consideração as informações da Tabela IV, o valor-p associado ao fator "Ações"é 0.756, muito acima do nível de significância de $\alpha = 0.05$. Assim, **não há evidência suficiente para rejeitar a hipótese nula** (H_{0_A}) de que os diferentes tipos de ações (PETR4, WEGE3, ABEV3, VALE3) não têm um efeito significativo sobre os retornos. Portanto, pode-se concluir que, independentemente do período, as ações analisadas têm retornos médios semelhantes.

Analogamente, o valor-p associado ao fator "Período-- por sua vez – é $0.064 > \alpha = 0.05$. Isto sugere que, ao nível de significância de 5%, **não rejeita a hipótese nula** (H_{0_P}) de que os períodos não têm um efeito significativo sobre os retornos das ações. Isto é, os períodos não têm um efeito razoável sobre os retornos das ações.

No caso da "Interação", o valor-p associado é 0.006, muito abaixo do $\alpha = 0.05$ estabelecido. Portanto, existe evidência estatisticamente relevante para **rejeitar a hipótese nula** (H_{0_I}) de que não há interação entre os períodos e os retornos das ações. Em outras palavras, a variação dos retornos das ações difere de maneira significativa dependendo do período.

O componente residual mostra a variação dos retornos não explicada pelos fatores considerados – no caso, períodos e tipos de ações. O quadrado médio residual é relativamente pequeno, o que é esperado dado o grande número de observações.

3. Conclusões da ANOVA de 2 Fatores

Os resultados da ANOVA indicam que os retornos dos ativos PETR4, WEGE3, ABEV3 e VALE3 são influenciados por interações complexas entre o tipo de ação e o período de tempo. Isto é consistente com o mercado financeiro, onde fatores como mudanças nas políticas econômicas e crises internacionais podem ter impactos diferentes sobre os vários setores em momentos distintos.

Nesse sentido, a variação nos retornos das ações depende do período considerado, sugerindo que eventos históricos e econômicos os influenciam fortemente. Por outro lado, não há evidência estatística suficiente para afirmar que os diferentes tipos de ações têm retornos médios diferentes ou que os períodos – por si só – influenciam os retornos em peso.

H. Regressão Linear Simples

A regressão linear é usada para aproximar a relação entre uma variável dependente Y e uma independente X por uma reta. O modelo é representado pela equação 13:

$$Y = \beta_0 + \beta_1 X + \epsilon \tag{13}$$

em que β_0 é o intercepto (ou coeficiente linear), β_1 é o *slope* (ou coeficiente de inclinação) e ϵ é o termo de erro.

O intercepto (β_0) representa o valor médio de Y quando $X = 0$, enquanto a inclinação (β_1) indica a mudança média em Y para uma unidade de mudança em X. Nos casos em que $\beta_1 = 0$, não há relação linear entre as variáveis.

Também é possível estimar um Coeficiente de Determinação (R^2), responsável por apontar a proporção da variabilidade total de Y explicada pelo modelo.[3]

Para determinar os valores dos coeficientes linear e de inclinação, utilizam-se as fórmulas dadas por 14 e 15 a seguir:

$$\beta_1 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2} \tag{14}$$

Considerando as médias de X e Y – respectivamente, $\overline{X}$ e $\overline{Y}$ – o *intercept*, por sua vez:

$$\beta_0 = \overline{Y} - \beta_1 \overline{X} \tag{15}$$

Em complemento a isso, para o coeficiente de determinação R^2, pode-se utilizar a Soma dos Quadrados dos Resíduos (SRR) e a Soma Total dos Quadrados (SST), resultando na equação descrita em 16:

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2} \tag{16}$$

Finalmente, a estatística F é calculada em 17:

$$F = \frac{MSR}{MSE} \tag{17}$$

em que $MSR = \frac{SST - SSR}{df_{modelo}}$ é o Quadrado Médio do modelo, e $MSE = \frac{SSR}{df_{residual}}$ é o Quadrado Médio do erro.

Na análise deste portfólio, a variável X é representada pelo tempo de cada período, convertido em um formato numérico, como dias desde um ponto de referência, enquanto a variável Y corresponde aos retornos diários ajustados logarítmicos de cada ação (PETR4, WEGE3, ABEV3, VALE3), cujos resultados são exibidos a seguir.

1. Regressão Linear dos Períodos para a Petrobras

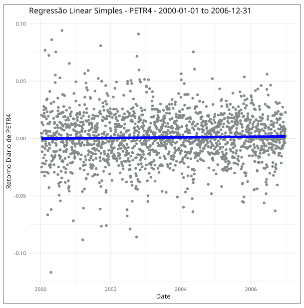


Figura 10: Regressão linear para retornos da PETR4 de 2000 até 2006

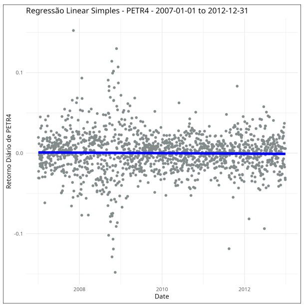


Figura 11: Regressão linear para retornos da PETR4 de 2007 até 2012

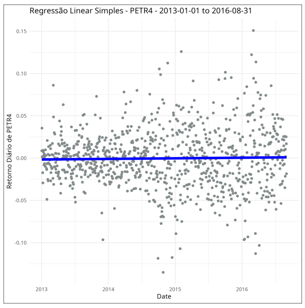


Figura 12: Regressão linear para retornos da PETR4 de 2013 até 2016

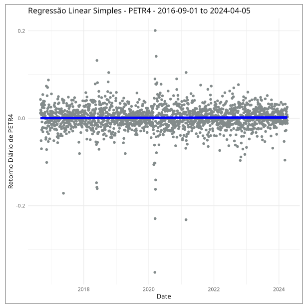


Figura 13: Regressão linear para retornos da PETR4 de 2016 até 2024

Analisando os gráficos da Figura **??**, percebe-se que os interceptos junto ao eixo das ordenadas indicam um retorno médio próximo de zero, com um coeficiente angular baixo que aponta para uma relação fraca entre o tempo e o retorno diário. Além disso, o R^2 próximo de zero indica que a variabilidade dos retornos diários não é explicada pelo tempo, com a curva – praticamente – horizontal confirmando a ausência de tendências signficativas, como visto na Figura 14.

---

[3] Para verificar a significância do modelo usado, utiliza-se a estatística F. Se o valor de F for alto, e for acompanhado de um baixo valor-p, então o modelo é válido para uma boa aproximação. Além disso, a hipótese de que não há relação linear entre X e Y – isto é, $\beta_1 = 0$ – é rejeitada se o valor-p for menor do que o nível de signficância (α).

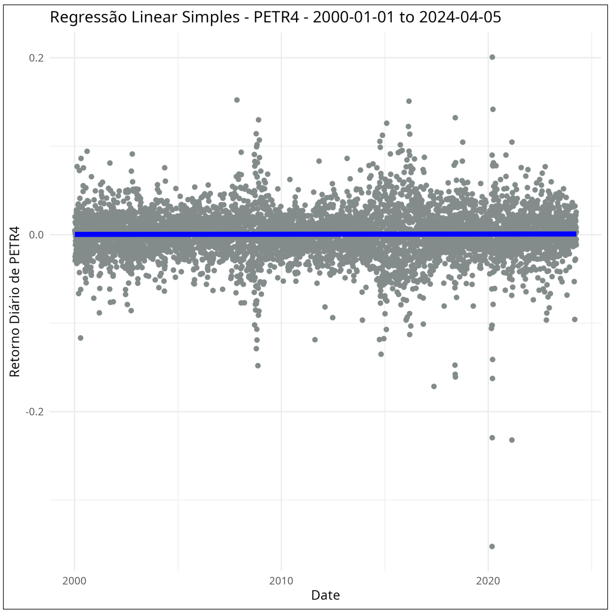


Figura 14: Regressão linear simples para retornos da PETR4, no período completo dos dados, de 2000 até 2024

Resultados para PETR4 no período de 2000-01-01 a 2006-12-31:

Coeficiente	**Estimativa**	**R-quadrado**	**Valor F**	**Valor p**
Intercepto Inclinação	$-4.283611e-05$ $7.227872e-07$	0.0006604334	1.19287	0.2748971

Resultados para PETR4 no período de 2007-01-01 a 2012-12-31:

Coeficiente	**Estimativa**	**R-quadrado**	**Valor F**	**Valor p**
Intercepto Inclinação	$8.391065e-04$ $-8.127948e-07$	0.0004170193	0.6178632	0.4319682

Resultados para PETR4 no período de 2013-01-01 a 2016-08-31:

Coeficiente	**Estimativa**	**R-quadrado**	**Valor F**	**Valor p**
Intercepto Inclinação	$-1.648933e-03$ $1.913024e-06$	0.0004649218	0.4218802	0.5161648

Resultados para PETR4 no período de 2016-09-01 a 2024-04-05:

Coeficiente	**Estimativa**	**R-quadrado**	**Valor F**	**Valor p**
Intercepto Inclinação	$5.225336e-04$ $4.238149e-07$	0.0001393793	0.2630454	0.6080956

A Petrobras, sendo uma empresa de energia e petróleo, é fortemente influenciada por preços internacionais de petróleo, políticas governamentais, descobertas de novos campos de petróleo, e eventos geopolíticos. Assim, o aumento nos preços das *commodities* pode ter tido um impacto positivo, mas as flutuações diárias seriam altamente voláteis devido a fatores exógenos e incontroláveis.

2. Regressão Linear dos Períodos para a WEG

A observação dos gráficos da Figura **??** para a WEG revela que o *intercept* $\beta_0 \approx 0$, sugere um retorno médio diário insignificante. O coeficiente angular também é pequeno, indicando uma fraca correlação entre o tempo e o retorno diário. O coeficiente R^2 – por sua vez – é muito próximo de zero, o que significa que a variação nos retornos

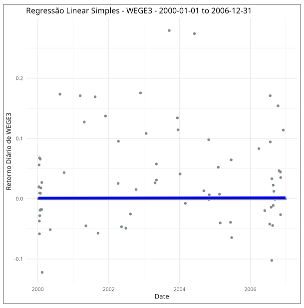


Figura 15: Regressão linear para retornos da WEGE3 de 2000 até 2006

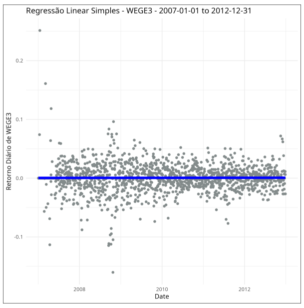


Figura 16: Regressão linear para retornos da WEGE3 de 2007 até 2012

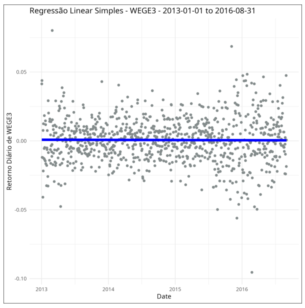


Figura 17: Regressão linear para retornos da WEGE3 de 2013 até 2016

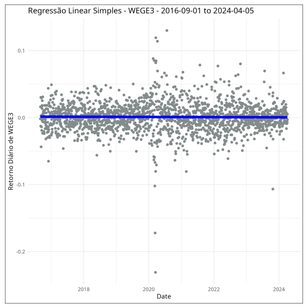


Figura 18: Regressão linear para retornos da WEGE3 de 2016 até 2024

diários não é explicada pelo tempo. Além disso, o gráfico mostra uma linha de regressão praticamente horizontal, corroborando a inexistência de uma tendência ao longo do tempo, como pode ser percebido na Figura 19.

Em complemento a isso, percebe-se uma assimetria na densidade da distribuição de pontos no início do gráfico devido ao processo de expansão e internacionalização da WEG nos anos 2000. Esta assimetria inicial pode ser atribuída à menor liquidez das ações durante o período de consolidação da presença global da empresa e ao esforço contínuo para conquistar a confiança dos investidores. Simultaneamente, a transição gradual para práticas modernas de governança corporativa pode ter contribuído para uma menor transparência e divulgação restrita de dados financeiros nos primeiros anos da década.

Resultados para WEGE3 no período de 2000-01-01 a 2006-12-31:

Coeficiente	**Estimativa**	**R-quadrado**	**Valor F**	**Valor p**
Intercepto	$1.110040e-03$	0.0001062087	0.1917271	0.6615354
Inclinação	$2.427328e-07$			

Resultados para WEGE3 no período de 2007-01-01 a 2012-12-31:

Coeficiente	**Estimativa**	**R-quadrado**	**Valor F**	**Valor p**
Intercepto	$4.145426e-04$	$1.428431e-05$	0.02115537	0.8843763
Inclinação	$1.458925e-07$			

Resultados para WEGE3 no período de 2013-01-01 a 2016-08-31:

Coeficiente	**Estimativa**	**R-quadrado**	**Valor F**	**Valor p**
Intercepto	$8.178473e-04$	$5.246491e-05$	0.04758817	0.8273635
Inclinação	$-3.099111e-07$			

Resultados para WEGE3 no período de 2016-09-01 a 2024-04-05:

Coeficiente	**Estimativa**	**R-quadrado**	**Valor F**	**Valor p**
Intercepto	$1.527566e-03$	0.0001861298	0.3512924	0.5534535
Inclinação	$-3.722518e-07$			

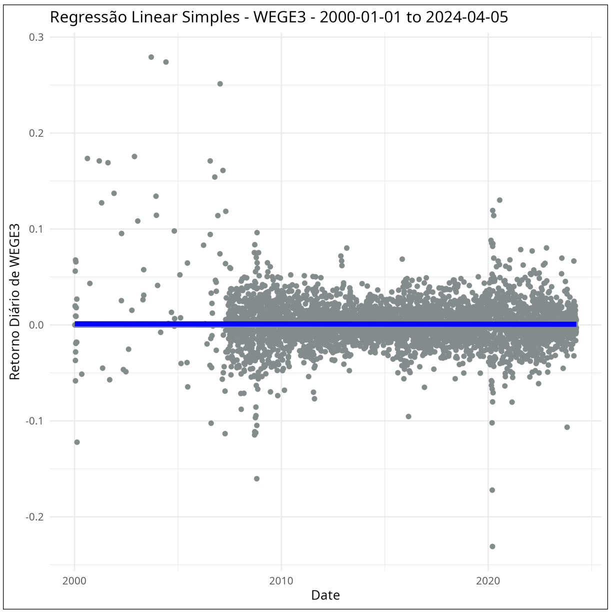


Figura 19: Regressão linear simples para retornos da WEGE3, no período completo dos dados, de 2000 até 2024

3. Regressão Linear dos Períodos para a Vale

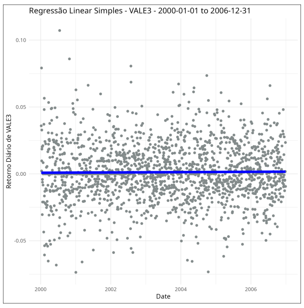


Figura 20: Regressão linear para retornos da VALE3 de 2000 até 2006

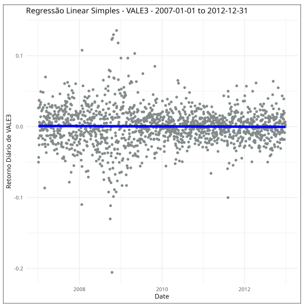


Figura 21: Regressão linear para retornos da VALE3 de 2007 até 2012

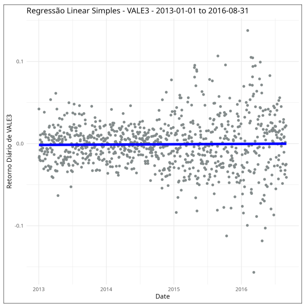


Figura 22: Regressão linear para retornos da VALE3 de 2013 até 2016

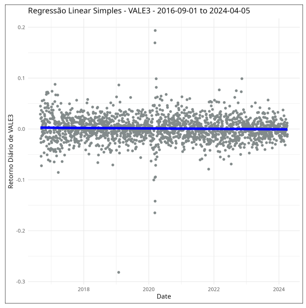


Figura 23: Regressão linear para retornos da VALE3 de 2016 até 2024

Examinando os gráficos da Figura **??**, pode-se notar que os interceptos próximos ao eixo x sugerem um retorno médio próximo de zero para a Vale, juntamente com um coeficiente angular baixo, indicativo de uma relação fraca entre o tempo e os retornos diários. Adicionalmente, o baixo valor de R^2 aponta que a variabilidade dos retornos diários não é adequadamente explicada pelo tempo – compatível com a curva aproximadamente horizontal observada na Figura 24 – o que confirma a ausência de tendências significativas.

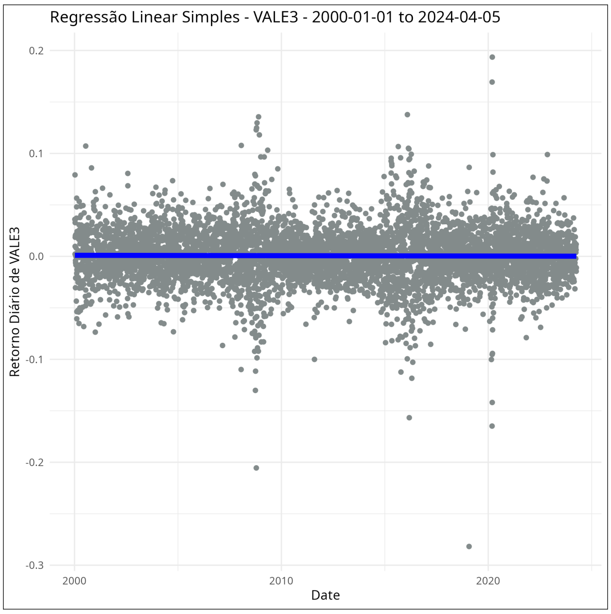


Figura 24: Regressão linear simples para retornos da VALE3, no período completo dos dados, de 2000 até 2024

Resultados para VALE3 no período de 2000-01-01 a 2006-12-31:

Coeficiente	**Estimativa**	**R-quadrado**	**Valor F**	**Valor p**
Intercepto	$8.352896e-04$	0.0001237994	0.2234856	0.6364539
Inclinação	$3.138840e-07$			

Resultados para VALE3 no período de 2007-01-01 a 2012-12-31:

Coeficiente	**Estimativa**	**R-quadrado**	**Valor F**	**Valor p**
Intercepto	$9.375663e-04$	0.0001855181	0.2748033	0.6002061
Inclinação	$-5.808666e-07$			

Resultados para VALE3 no período de 2013-01-01 a 2016-08-31:

Coeficiente	**Estimativa**	**R-quadrado**	**Valor F**	**Valor p**
Intercepto	$-1.582699e-03$	0.000189715	0.1721041	0.6783466
Inclinação	$1.102628e-06$			

Resultados para VALE3 no período de 2016-09-01 a 2024-04-05:

Coeficiente	**Estimativa**	**R-quadrado**	**Valor F**	**Valor p**
Intercepto	$2.546483e-03$	0.001414251	2.672472	0.1022642
Inclinação	$-1.160009e-06$			

Essas oscilações podem ser explicadas em razão dos preços do minério de ferro – que são determinados por oferta e demanda global –, que impactam diretamente os retornos de Vale. Além disso, políticas fiscais e regulamentações ambientais de mineração do governo brasileiro também exercem grande poder sobre a operação e os lucros da empresa.

4. Regressão Linear dos Períodos para a Ambev

A análise da Ambev, conduzida na Figura **??**, mostra que o intercepto apresenta um valor muito reduzido, mostrando indícios de que o retorno médio diário é desprezível. O *slope* também é próximo de zero, indicando uma fraca interferência do tempo no retorno diário. O coeficiente de determinação (R^2) é aproximandaemnte zero, o que significa que a variação nos retornos diários não é explicada pelo tempo. Além disso, a Figura 29 mostra uma linha de regressão quase horizontal, corroborando a inexistência de uma tendência significativa ao longo do tempo.

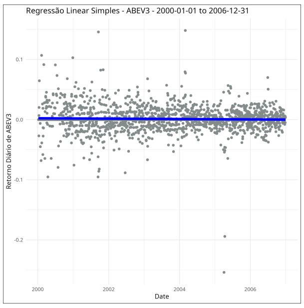


Figura 25: Regressão linear para retornos da ABEV3 de 2000 até 2006

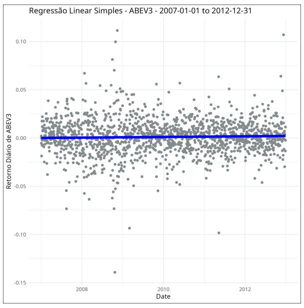


Figura 26: Regressão linear para retornos da ABEV3 de 2007 até 2012

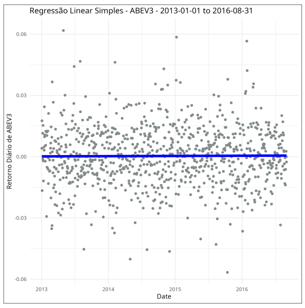


Figura 27: Regressão linear para retornos da ABEV3 de 2013 até 2016

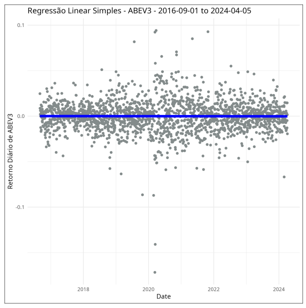


Figura 28: Regressão linear para retornos da ABEV3 de 2016 até 2024

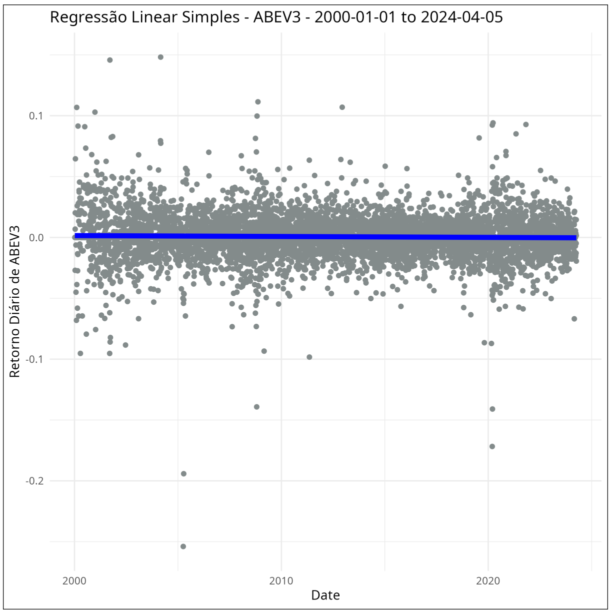


Figura 29: Regressão linear simples para retornos da ABEV3, no período completo dos dados, de 2000 até 2024

Resultados para ABEV3 no período de 2000-01-01 a 2006-12-31:

Coeficiente	**Estimativa**	**R-quadrado**	**Valor F**	**Valor p**
Intercepto Inclinação	$2.230736e-03$ $-8.683661e-07$	0.0009263544	1.67362	0.1959399

Resultados para ABEV3 no período de 2007-01-01 a 2012-12-31:

Coeficiente	**Estimativa**	**R-quadrado**	**Valor F**	**Valor p**
Intercepto Inclinação	$6.721223e-05$ $9.465526e-07$	0.001031692	1.529514	0.2163815

Resultados para ABEV3 no período de 2013-01-01 a 2016-08-31:

Coeficiente	**Estimativa**	**R-quadrado**	**Valor F**	**Valor p**
Intercepto Inclinação	$1.318410e-04$ $2.901909e-07$	$6.069259e-05$	0.05505152	0.8145484

Resultados para ABEV3 no período de 2016-09-01 a 2024-04-05:

Coeficiente	Estimativa	R-quadrado	Valor F	Valor p
Intercepto	$4.40914e-05$	$1.996192e-05$	0.0376689	0.8461304
Inclinação	$-9.72327e-08$			

O setor de bebidas e alimentos, onde a Ambev atua, é mais influenciado por fatores sazonais e de mercado local – como o consumo doméstico e mudanças na preferência de consumidores – além de políticas fiscais e regulamentações específicas do setor. Os retornos diários mostram uma volatilidade que pode ser atribuída a eventos internos da empresa, mudanças na gestão, aquisições, e outros fatores microeconômicos.

5. Conclusão da Regressão Linear Simples

Em todas as curvas apresentadas nesta seção, os pontos cinzas no gráfico representam os retornos diários da ação ao longo do tempo, enquanto a linha azul é a linha de tendência gerada pela regressão linear.

Observando a dispersão dos pontos, é evidente que a maioria dos retornos diários está concentrada em torno de zero, com uma variabilidade relativamente constante ao longo dos anos, apesar de exibierem uma alta volatilidade nos retornos diários. Isto aponta que os ativos deste portfólio estão sujeitos a flutuações diárias consideráveis, que podem ser resultado de eventos específicos da empresa, macroeconômicos, ou notícias de mercado, sem uma forte relação linear entre período de tempo e retorno diário de cada ativo.

I. Correlação

A matriz de correlação nos fornece informações sobre como os retornos de diferentes ativos estão relacionados entre si.

Tabela V: Matriz de Correlação entre os Ativos

	PETR4	WEGE3	ABEV3	VALE3
PETR4	1.0000000	0.2290674	0.2877517	0.4732419
WEGE3	0.2290674	1.0000000	0.2163808	0.2155345
ABEV3	0.2877517	0.2163808	1.0000000	0.2654501
VALE3	0.4732419	0.2155345	0.2654501	1.0000000

Observando a matriz de correlação fornecida:

- As correlações entre os retornos da PETR e da VALE (Petrobras e Vale) e da PETR e da ABEV (Petrobras e Ambev) são relativamente altas (0.4732419 e 0.2877517, respectivamente). Isso sugere que os retornos dessas empresas tendem a se mover juntos em certa medida.
- A correlação entre os retornos da PETR e da WEGE (Petrobras e WEG) é moderada (0.2290674).
- As correlações entre os retornos da VALE e da ABEV e da VALE e da WEGE são baixas, em torno de 0.2654501 e 0.2155345, respectivamente. Isso sugere uma menor relação linear entre os retornos dessas empresas.

As correlações entre os ativos sendo todas positivas indicam uma tendência geral de movimento conjunto dos retornos desses ativos ao longo do período analisado. Isso significa que, em geral, quando o retorno de um ativo é positivo (ou negativo), os retornos dos outros ativos também tendem a ser positivos (ou negativos), embora em diferentes magnitudes.

Existem algumas razões possíveis para essa observação:

1. **Mercado Geral**: Durante o período analisado, pode ter havido uma tendência geral de alta nos mercados financeiros, o que influenciaria positivamente os retornos de todos os ativos considerados.
2. **Setor Econômico**: Os ativos analisados podem pertencer ao mesmo setor econômico ou ter alguma correlação intrínseca devido a fatores macroeconômicos comuns que afetam o desempenho de todas as empresas do setor.
3. **Fatores Externos Comuns**: Pode haver fatores externos que afetam todas as empresas analisadas de maneira semelhante, como mudanças nas taxas de juros, políticas governamentais ou eventos geopolíticos.
4. **Comportamento do Investidor**: O comportamento dos investidores também pode influenciar as correlações entre os ativos. Por exemplo, se os investidores tendem a comprar ou vender todos os ativos simultaneamente em resposta a notícias ou eventos específicos, isso pode levar a correlações positivas entre os retornos.

J. Covariância

A matriz de covariâncias nos fornece informações sobre como os retornos de diferentes ativos variam juntos. Ela é uma medida da dispersão conjunta desses retornos. Os valores são expressos em unidades de variação (ou desvio padrão) ao quadrado.

Observa-se que a diagonal principal da Tabela VI abaixo mostra as variâncias individuais de cada ativo, enquanto os valores fora da diagonal representam as covariâncias entre os pares de ativos.

Vale destacar que a soma dos valores da diagonal principal não precisa ser 1, pois eles representam as variâncias individuais dos ativos.

Tabela VI: Matriz de Covariância Anual entre os Ativos

	PETR4	WEGE3	ABEV3	VALE3
PETR4	0.17796736	0.03224028	0.03549190	0.08006722
WEGE3	0.03224028	0.11130908	0.02110693	0.02883922
ABEV3	0.03549190	0.02110693	0.08548352	0.03112616
VALE3	0.08006722	0.02883922	0.03112616	0.16084325

Observando a matriz de covariâncias anualizada fornecida:

- As covariâncias entre os retornos da PETR e da VALE e da PETR e da ABEV são relativamente altas (0.08006722 e 0.03549190, respectivamente). Isso indica que os retornos dessas empresas tendem a variar juntos em uma quantidade significativa ao longo do tempo.
- A covariância entre os retornos da PETR e da WEGE é menor (0.03224028), mas ainda indica uma relação positiva entre os retornos dessas empresas.
- As covariâncias entre os retornos da VALE e da ABEV e da VALE e da WEGE são baixas, em torno de 0.03112616 e 0.02883922, respectivamente. Isso sugere uma menor dispersão conjunta entre os retornos dessas empresas.

K. Testes de Hipótese para Igualdade de variâncias entre PETR4 e VALE3

Devido a se tratar de dois ativos com forte influência federal, dentre outros fatores em comum, levantou-se a suspeita de que pudessem ter variâncias parecidas. Para tanto, elaborou-se o seguinte teste de hipótese com um nível de significância de 5%:

- **Hipótese Nula (H_0):** A variância dos dois ativos é igual.
- **Hipótese Alternativa (H_1):** A variância dos dois ativos é diferente.

No cálculo do teste de hipótese, os seguintes resultados foram encontrados:

- **a** (Primeiro valor crítico): 0.9587214
- **b** (Segundo valor crítico): 1.043056
- **quociente das variâncias amostrais**: 0.10194

Analisando os resultados, devido ao valor do quociente das variâncias amostrais não estar entre os dois valores críticos do teste F, não é possível concluir que as variâncias são diferentes.

V. Fronteira Eficiente de Markowitz

São feitos aqui alguns experimentos com o conceito de Fronteira Eficiente de Markowitz a partir dos dados obtidos referente aos ativos estudados. Foi desenvolvida uma aplicação em Python capaz de gerar a Fronteira Eficiente de forma gráfica a partir de um *dataset*, além de apresentar as proporções de investimento a serem aplicadas em cada ativo para se atingir o risco ou o retorno esperado, os quais devem ser informados pelo usuário. Se este inserir um risco esperado, a aplicação deverá maximizar o retorno. Caso o usuário insira um retorno esperado, a aplicação deverá minimizar o risco.

A. Fronteira Eficiente a partir dos dados obtidos

A forma mais simples de se calcular a Fronteira Eficiente envolve simplesmente obter a média dos retornos e a matriz de covariâncias do portfólio informado e realizar a otimização do retorno e do risco conforme as restrições em 4. Isso resulta em uma curva que indica todas as combinações de pesos possíveis para aquele portfólio, bem como cada percentual risco associado a cada percentual de retorno.

Apresenta-se a seguir exemplos desta execução. Gerou-se primeiro uma Fronteira Eficiente com um risco esperado igual a 25%, e depois gerou-se outra, mas com retorno esperado de 21% (valores escolhidos arbitrariamente):

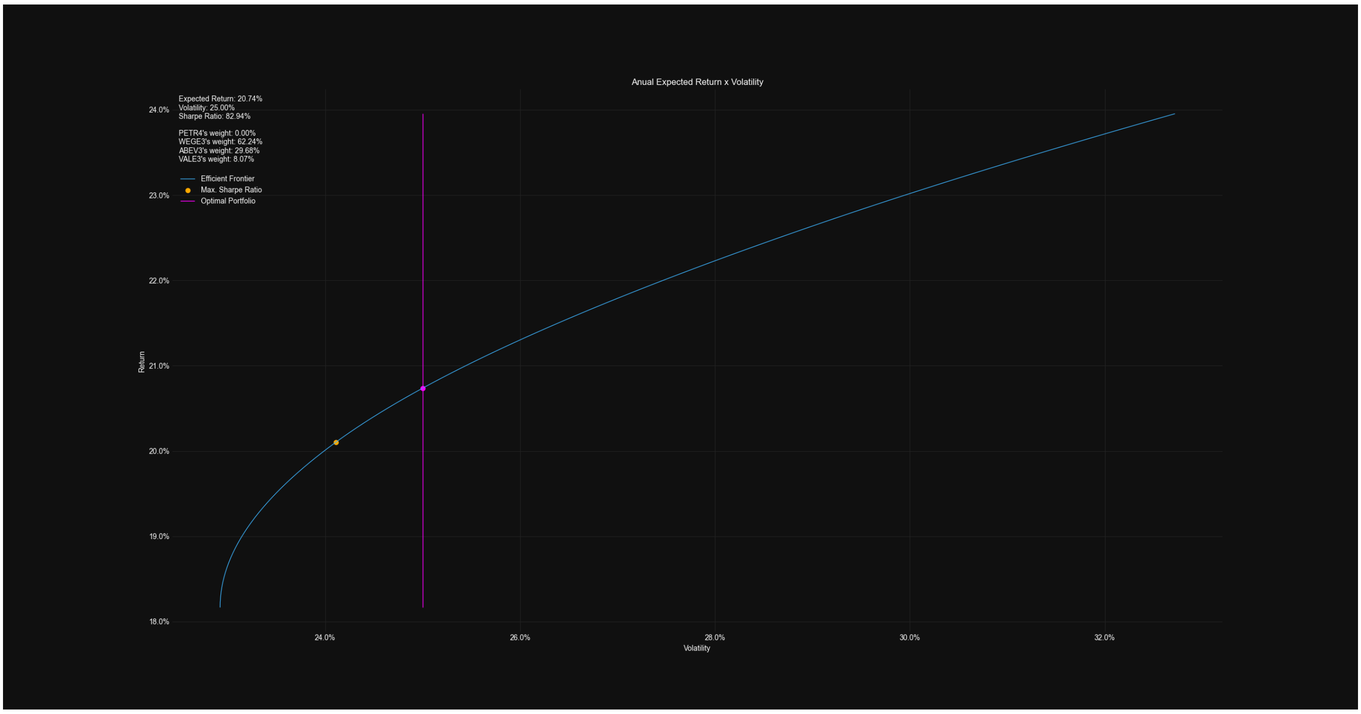


Figura 30: Fronteira Eficiente de Markowitz para um risco esperado de 25%

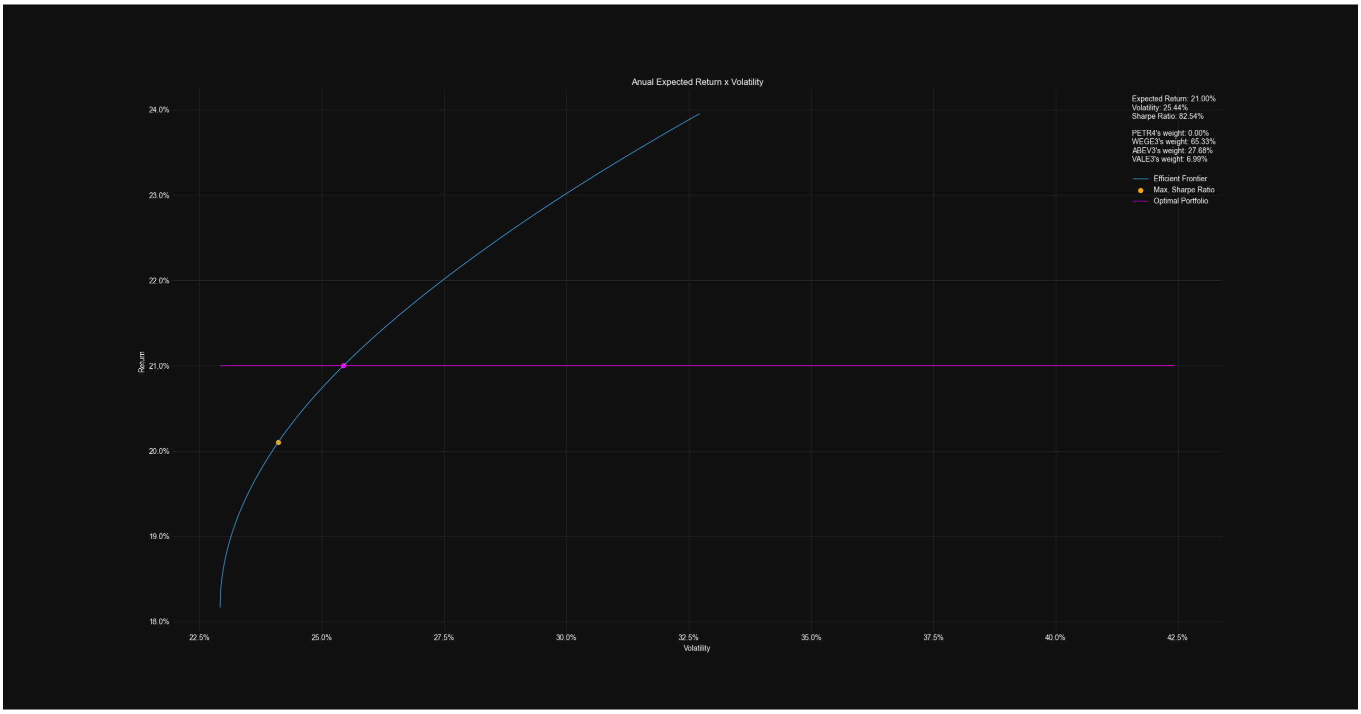


Figura 31: Fronteira Eficiente de Markowitz para um retorno esperado de 21%

B. Amostragem Aleatória e Intervalo de Confiança

Objetiva-se agora aplicar técnicas de amostragens aleatórias para construir múltiplas Fronteiras Eficientes, bem como um Intervalo de Confiança sobre elas. Sabe-se que o comportamento de retornos de ativos em geral apresenta grande variabilidade ao longo do tempo, e mesmo que se obtenha grandes *datasets* com valores históricos de retornos, diversos incidentes reais que tenham acontecido ao longo desse período podem impactar significativamente tais valores. Portanto, é extremamente pouco provável que as ações analisadas neste trabalho se comportem por longos períodos da maneira como muitas medidas estatísticas aqui obtidas apontam, mesmo que o *dataset* utilizado possua valores de fechamento de desde os anos 2000.

Em especial, dá-se atenção às covariâncias entre os ativos. Elas determinam diretamente o risco do portfólio, e são uma das características que mais sofrem grandes variabilidades ao longo do tempo. Devido a isso, decidiu-se gerar múltiplas matrizes de covariância, a partir de um Método de Monte Carlo com base em medidas estatísticas já obtidas, para se construir várias Fronteiras Eficientes amostrais. Ao fazer um Intervalo de Confiança sobre elas, objetiva-se construir uma região gráfica na qual a Fronteira Eficiente populacional deve estar localizada, com 95% de confiança.

1. Geração de múltiplas matrizes de covariância

O primeiro passo realizado foi obter as correlações amostrais entre os ativos. Como o portfólio contém quatro ativos ao todo, existem seis valores de correlação distintos, além de suas cópias simétricas e as autocorrelações (todas das quais valem 1).

Obteve-se as correlações porque, para se gerar de múltiplas covariâncias, é mais fácil manipular as primeiras. Apesar de ambas não terem funções de distribuição de probabilidade conhecidas, pode-se aplicar na correlação a Transformação de Fisher para se obter uma variável (no caso, a arco tangente hiperbólica da correlação, z) cuja distribuição de probabilidade é conhecidamente Normal. Para cada correlação amostral ρ, obtida a partir de N observações, a distribuição de probabilidade de z tem o seguintes parâmetros de média e desvio padrão:

$$\mu_z = \frac{1}{2} ln(\frac{1+\rho}{1-\rho}) = arctanh(\rho) \tag{18}$$

$$\sigma_z = \frac{1}{\sqrt{N-3}} \tag{19}$$

Com a distribuição de z para cada ρ em mãos, pode-se utilizá-la em um Método de Monte Carlo para se obter mais valores de z a partir de amostragens aleatórias. A ideia é gerar e dispersar aleatoriamente valores ao longo do espaço da distribuição de z e obter aqueles que caírem dentro da mesma. Pode-se executar o método quantas vezes for necessário, para se obter uma quantidade suficiente de valores de z para cada ρ.

Da mesma forma que um valor de ρ pode ser transformado em z, o inverso também é válido. A equação que descreve isso é justamente o inverso da Transformação de Fisher:

$$\rho = \frac{exp(2z) - 1}{exp(2z) + 1} = tanh(z) \tag{20}$$

Aplica-se o inverso da Transformação de Fisher para se obter novos valores de correlação a partir de cada z obtido previamente. Como estes foram gerados via Método de Monte Carlo para cada z obtido a partir de uma correlação amostral, os novos valores de correlação estarão agrupados em relação aos valores originais.

Finalmente, transforma-se as correlações obtidas em covariâncias. Isso é feito ao multiplicar cada correlação pelos desvios padrões das duas amostras que a geraram. Ao fim de todo este processo, obteve-se um grande conjunto de covariâncias agrupadas em relação às originais, as quais podem ser utilizadas para análises posteriores. A seguir, explora-se a partir desse conjunto de covariâncias a seguinte abordagem: construir uma Fronteira Eficiente para cada vez em que o Método de Monte Carlo foi executado, e construir um Intervalo de Confiança para as curvas obtidas.

2. Construção de Múltiplas Fronteiras Eficientes

A partir da mesma metodologia descrita anteriormente, gerou-se uma Fronteira Eficiente para cada conjunto de covariâncias amostradas. Modificou-se a aplicação em Python para gerar múltiplas matrizes de covariâncias via Método de Monte Carlo de acordo com os procedimentos explicados anteriormente, e então realizar os cálculos de otimização de 4 para cada simulação executada.

Adicionalmente, também se calculou o que seria um Fronteira Eficiente média: como a única diferença efetiva entre uma fronteira e outra é a matriz de covariâncias (os pesos, apesar de também serem uma diferença, são resultantes desta), pode-se calcular uma matriz média e então traçar uma fronteira a partir da mesma.

A seguir, apresenta-se os resultados deste processo. Para um risco esperado de 25% (arbitrariamente escolhido), primeiro executou-se o programa para gerar quatro Fronteiras Eficientes, e depois para cem delas:

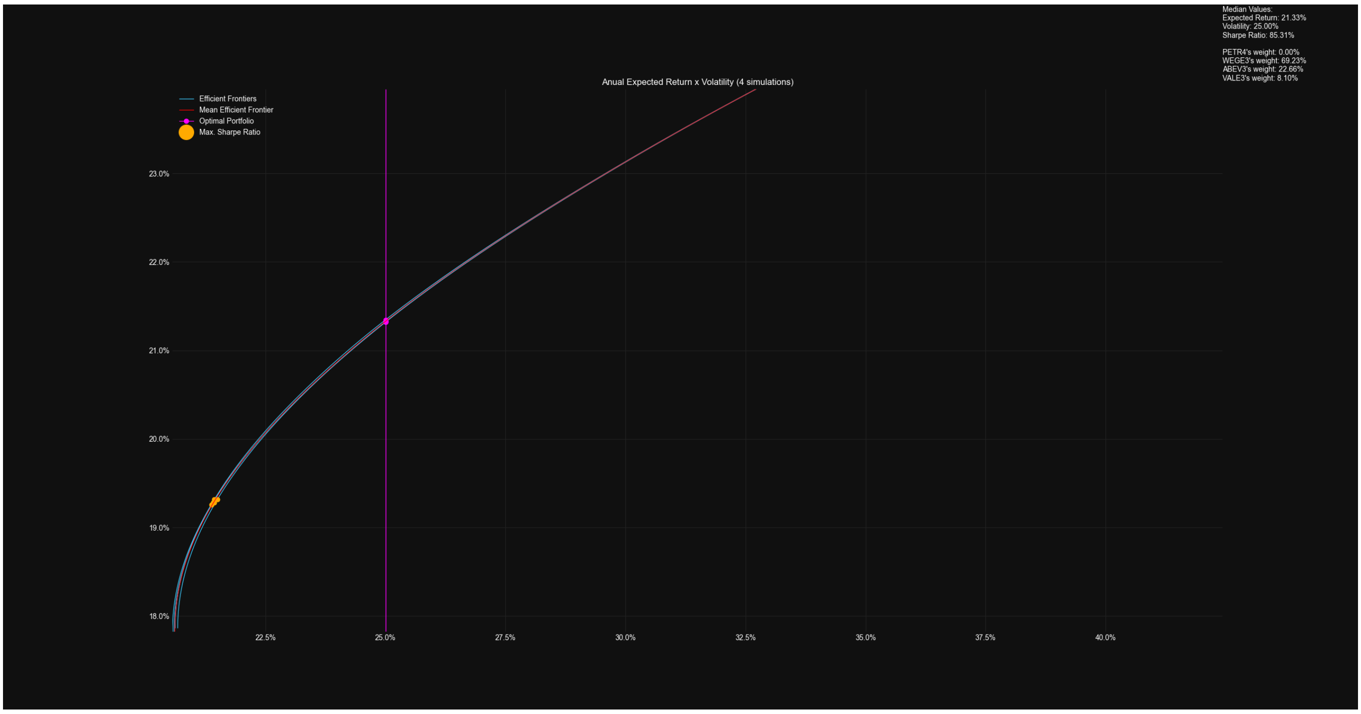


Figura 32: Fronteiras Eficientes de Markowitz para um risco esperado de 25% - 4 simulações

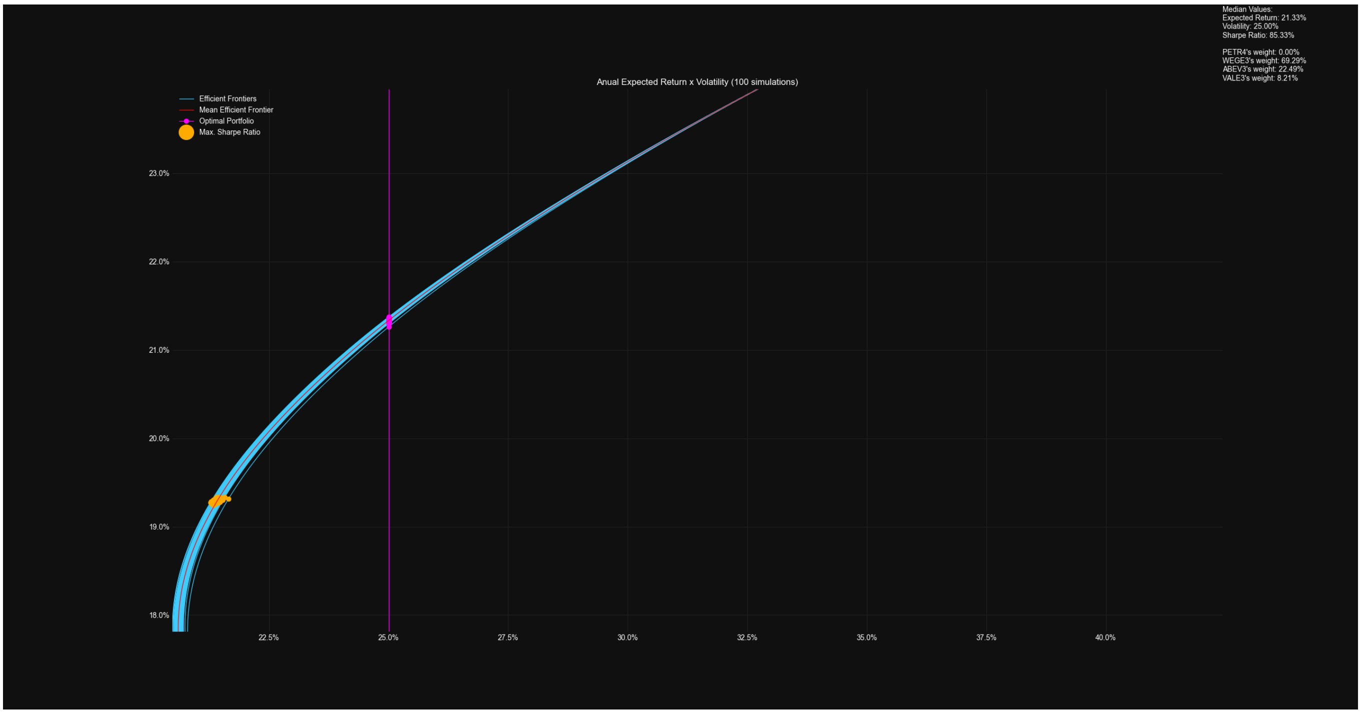


Figura 33: Fronteiras Eficientes de Markowitz para um risco esperado de 25% - 100 simulações

É interessante notar o espalhamento das várias fronteiras ao passo que se aumenta o número de simulações. De fato, é interessante realizar uma boa quantia de simulações justamente para se ter uma melhor percepção da variação entre as várias fronteiras. Um maior número de simulações também torna Intervalos de Confiança mais precisos e restritivos, como será detalhado a seguir.

3. Construção do Intervalo de Confiança

Além da Fronteira Eficiente média, aproveitou-se da natureza da distribuição da variável z, discutida anteriormente, para se estabelecer um Intervalo de Confiança entre as covariâncias obtidas. Novamente com base no pressuposto de que a única diferença efetiva entre fronteiras é a matriz de covariâncias, pode-se utilizar dos passos anteriores (cálculo das correlações e da Transformação de Fisher para elas) para se estabelecer um Intervalo de Confiança para os valores z, os quais possuem distribuição Normal.

O Intervalo de Confiança para uma variável com distribuição Normal, com um nível de significância α, é definido pela equação:

$$P(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1 - \alpha \tag{21}$$

Os limites deste intervalo então são:

$$min = \bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{22}$$

$$max = \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{23}$$

Para o caso da variável z, pode-se calcular mais dois conjuntos de dados - mínimo e máximo - a partir de seu valor médio e desvio padrão (este conforme 19). O conjunto mínimo e máximo são os delimitadores do Intervalo de Confiança para toda a amostra de dados coletada pelo Método de Monte Carlo.

Adicionou-se então na aplicação Python este procedimento, para se gerar um Intervalo de Confiança com nível de significância $\alpha = 5\%$. Traçou-se as Fronteiras Eficientes dos limites junto à Fronteira Eficiente média. Os resultados, para quatro e cem simulações, são apresentados a seguir:

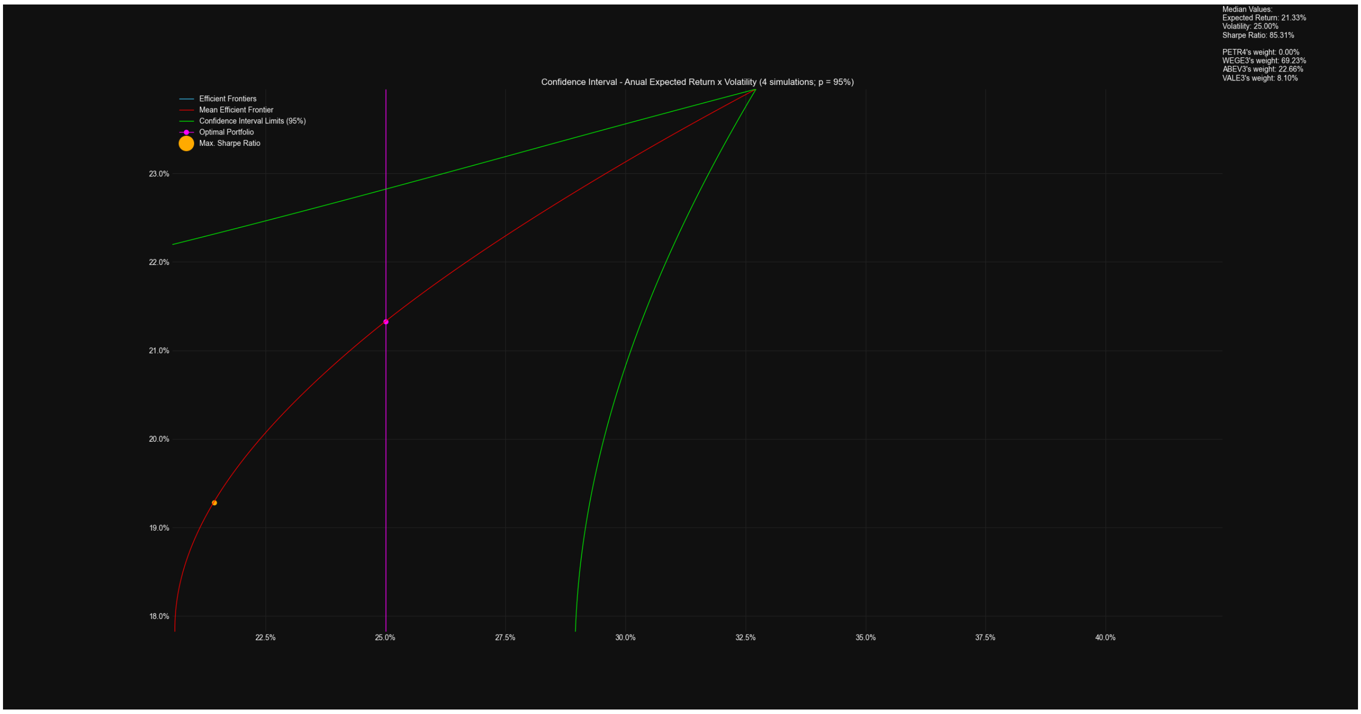


Figura 34: Intervalo de confiança para as Fronteiras Eficientes geradas - 95% de confiança. 4 simulações para um risco esperado de 25%

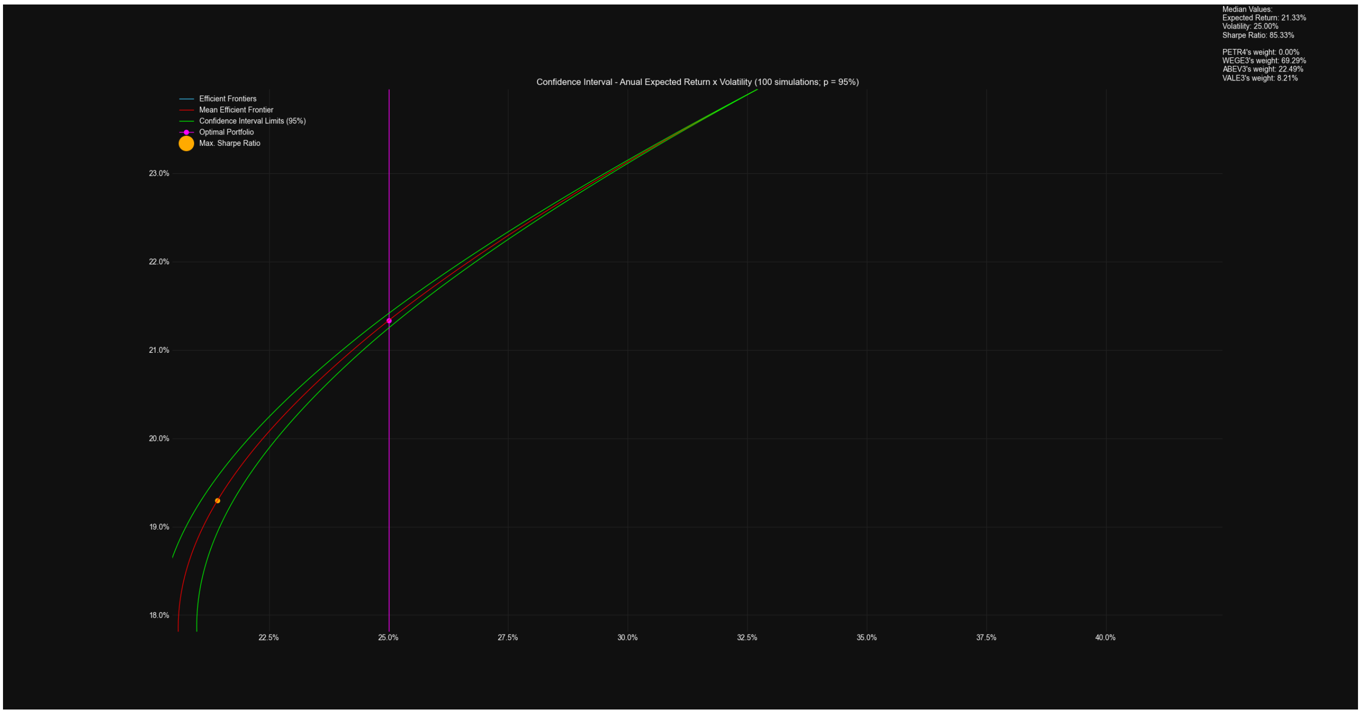


Figura 35: Intervalo de confiança para as Fronteiras Eficientes geradas - 95% de confiança. 100 simulações para um risco esperado de 25%

Como esperado, nota-se que o Intervalo de confiança para quatro simulações é muito mais largo que para o de cem - isso se deve ao desvio padrão de z, que é muito maior para um $N = 4$. Conforme se retira mais amostras, a tendência é que o intervalo se torne mais preciso e restritivo. Junto ao resultado do gráfico com cem Fronteiras

Eficientes, percebe-se que fronteiras tendem a se concentrar na área delimitado pelo Intervalo de Confiança para 100 simulações, o que comprova sua precisão.

VI. Conclusões

Os resultados obtidos neste estudo revelam insights significativos sobre o comportamento dos preços de fechamento das ações analisadas, proporcionando uma visão abrangente das dinâmicas de mercado ao longo do tempo. A análise estatística e econométrica aplicada permitiu não apenas entender a performance histórica dos ativos, mas também identificar oportunidades e desafios potenciais para investidores.

Inicialmente, observamos que os preços de fechamento das ações não seguem uma distribuição normal, evidenciando assimetria e a presença de valores atípicos. Esta característica sublinha a importância de técnicas robustas de análise estatística para compreender a verdadeira natureza dos dados financeiros.

Ao examinar a evolução temporal dos preços, destacaram-se momentos de volatilidade e tendências de valorização ao longo do período estudado. Essas flutuações são reflexo das condições econômicas e eventos específicos que impactaram os mercados durante os últimos anos.

No que diz respeito aos retornos, todas as ações apresentaram retornos médios anuais positivos, indicando que os investidores, em média, obtiveram ganhos ao longo do período analisado. No entanto, cada ação exibiu diferentes níveis de volatilidade, como evidenciado pelas variâncias dos preços de fechamento. WEGE3 e VALE3 mostraram as maiores variâncias, refletindo uma maior volatilidade, enquanto PETR4 e ABEV3 exibiram variâncias menores, indicando uma volatilidade relativamente menor em comparação.

Um ponto relevante identificado foi a alta correlação entre os retornos das ações PETR4 e VALE3. Essa correlação positiva pode ser explicada pelo fato de ambas as empresas serem estatais e atuarem em setores de commodities, como petróleo e mineração, respectivamente. Essa similaridade de contexto setorial pode aumentar a sensibilidade dessas ações a fatores econômicos e políticos comuns, tanto no mercado doméstico quanto global.

Após uma análise detalhada dos dados e gráficos fornecidos, fica claro que as características individuais de cada ativo desempenham um papel crucial na construção de uma carteira de investimentos eficiente. A consideração de métricas como retorno médio, volatilidade e correlações com outros ativos é essencial para a seleção criteriosa dos componentes da carteira.

A otimização de Markowitz emerge como uma ferramenta fundamental neste contexto. Ao levar em conta não apenas os retornos esperados, mas também as volatilidades e as correlações entre os ativos, a otimização de Markowitz permite construir carteiras que buscam maximizar o retorno esperado para um determinado nível de risco, ou minimizar o risco para um determinado nível de retorno. Um aspecto particularmente relevante foi a criação de múltiplas Fronteiras Eficientes de Markowitz, que proporcionaram um Intervalo de Confiança para as distribuições otimizadas das carteiras. Esse método, baseado em uma amostragem aleatória via Método de Monte Carlo, permitiu capturar a incerteza associada às estimativas de covariâncias, oferecendo uma abordagem mais realista e robusta para a construção de carteiras diversificadas.

Em resumo, enquanto a análise individual dos ativos é crucial para identificar oportunidades e compreender suas características específicas, a otimização de Markowitz oferece uma abordagem sistemática e eficaz para a construção de carteiras de investimentos diversificadas e eficientes. Integrando análise detalhada com técnicas avançadas de otimização, os investidores podem tomar decisões mais informadas, alinhando suas estratégias de investimento com seus objetivos de retorno e tolerância ao risco.

VII. Apêndices

A. Código-fonte Python para Extração dos Dados da B3 e Geração da Base de Dados

```python
import logging
from datetime import datetime, date
import yfinance as yf
import pandas as pd

# Set the logging level for yfinance to CRITICAL to reduce noise
logging.getLogger('yfinance').setLevel(logging.CRITICAL)

# Define a function to validate the input date
def validate_date(input_date):
    try:
        # Check if the input matches the desired format (YYYY-MM-DD)
        parsed_date = datetime.strptime(input_date, '%Y-%m-%d')
        year, month, day = map(str, input_date.split('-'))
        if len(year) == 4 and len(month) == 2 and len(day) == 2:
            if parsed_date.date() < date.today():
                return True
            else:
                print('The start date should be before today\'s date.')
                return False
        else:
            return False
    except:
        return False

# Define a function to validate the input asset tickers and retrieve its data
def validate_assets(asset_inputs, start_date):
    assets = {}
    asset_tickers = asset_inputs.split(',')
    for ticker in asset_tickers:
        # Remove leading/trailing spaces
        ticker = ticker.strip()
        # Download asset data using yfinance
        asset_data = yf.download(ticker, start_date)['Adj Close']
        if len(asset_data) > 0:
            # Store asset data if successfully downloaded
            assets[ticker] = asset_data

    assets = pd.DataFrame(assets)
    return assets

# Get user input for the analysis start date
start_date = None
while start_date is None:
    start_date = input('Please input the analysis start date (YYYY-MM-DD): ')
    if not validate_date(start_date):
        print('Invalid date. Please use YYYY-MM-DD format.')
        start_date = None

# Get user input for the asset tickers
asset_tickers = None
while asset_tickers is None:
    asset_tickers = input('Specify the asset ticker symbols (comma-separated): ')
    assets = validate_assets(asset_tickers, start_date)
    if assets.empty:
```

```
            print('No valid assets found. Please enter at least one valid asset ticker
            ↪  symbol.')
            asset_tickers = None

    # Define the file path for the CSV file
    csv_file_path = "assets.csv"

    # Save the DataFrame to a CSV file
    assets.to_csv(csv_file_path)

    print("DataFrame saved to", csv_file_path)
```

B. Código-fonte Python para Geração de Múltiplas Fronteiras Eficientes e Intervalo de Confiança entre Fronteiras

```
import pandas as pd
from pathlib import Path
import statistics as st
import numpy as np
from scipy import optimize, stats
import matplotlib.pyplot as plt
import matplotlib.ticker as mplticker
import mplcursors
from matplotlib.lines import Line2D

def normalMonteCarlo(avgs, std_dev, num_reps):
    i = 0
    arr = []
    while(i < len(avgs)):
        arr_i = np.random.normal(avgs[i], std_dev, num_reps)
        arr.append(arr_i)
        i += 1
    return arr

def calcZ(r):
    return 0.5 * np.log((1 + r)/(1 - r))

def calcR(z):
    return ((np.exp(2 * z) - 1) / (np.exp(2 * z) + 1))

def metrics(weights, log_mean, covariance):
    weights = np.array(weights)
    returns = np.sum(np.array([
        log_mean[0] * weights[0],
        log_mean[1] * weights[1],
        log_mean[2] * weights[2],
        log_mean[3] * weights[3]
    ]))

    volatility = np.sqrt(np.abs(weights.T.dot(covariance.dot(weights))))

    sharpe_ratio = returns / volatility
    return [returns, volatility, sharpe_ratio]

def __main__():
    df = pd.read_csv(Path('assets.csv')).dropna()
    amount = len(df.index)

    petr = df['PETR4.SA'].array
```

```
wege = df['WEGE3.SA'].array
abev = df['ABEV3.SA'].array
vale = df['VALE3.SA'].array

ret_petr  = np.delete(np.log(petr / petr.shift(1)), 0)
ret_wege  = np.delete(np.log(wege / wege.shift(1)), 0)
ret_abev  = np.delete(np.log(abev / abev.shift(1)), 0)
ret_vale  = np.delete(np.log(vale / vale.shift(1)), 0)

ret_means = np.array([st.mean(ret_petr), st.mean(ret_wege), st.mean(ret_abev),
 ↪  st.mean(ret_vale)]) * 252

print("Return means")
print(ret_means)

petr_var = st.variance(ret_petr)
wege_var = st.variance(ret_wege)
abev_var = st.variance(ret_abev)
vale_var = st.variance(ret_vale)

print("Variances")
print(petr_var)
print(wege_var)
print(abev_var)
print(vale_var)

corr_petr_wege = st.correlation(ret_petr, ret_wege)
corr_petr_abev = st.correlation(ret_petr, ret_abev)
corr_petr_vale = st.correlation(ret_petr, ret_vale)
corr_wege_petr = st.correlation(ret_wege, ret_petr)
corr_wege_abev = st.correlation(ret_wege, ret_abev)
corr_wege_vale = st.correlation(ret_wege, ret_vale)
corr_abev_petr = st.correlation(ret_abev, ret_petr)
corr_abev_wege = st.correlation(ret_abev, ret_wege)
corr_abev_vale = st.correlation(ret_abev, ret_vale)
corr_vale_petr = st.correlation(ret_vale, ret_petr)
corr_vale_wege = st.correlation(ret_vale, ret_wege)
corr_vale_abev = st.correlation(ret_vale, ret_abev)

correlations = [
    corr_petr_wege,
    corr_petr_abev,
    corr_petr_vale,
    corr_wege_petr,
    corr_wege_abev,
    corr_wege_vale,
    corr_abev_petr,
    corr_abev_wege,
    corr_abev_vale,
    corr_vale_petr,
    corr_vale_wege,
    corr_vale_abev
]

print("Correlations")
print(correlations)

z_calc_all = calcZ(np.array(correlations))
```

```
print("CALC z all")
print(z_calc_all)

z_desvpad = 1 / np.sqrt(amount - 3)
print("Z desvpad")
print(z_desvpad)

calculation_type = input("Calculation type: ")
expected_value = input("Expected variable value: ")
sims = int(input("Number of simulations: "))

z_all = normalMonteCarlo(z_calc_all, z_desvpad, sims)
print("z all")
print(z_all)

r_rand = calcR(z_all)
print("Monte Carlo")
print(r_rand)

cov_petr_wege = r_rand[0 ] * np.sqrt(petr_var) * np.sqrt(wege_var)
cov_petr_abev = r_rand[1 ] * np.sqrt(petr_var) * np.sqrt(abev_var)
cov_petr_vale = r_rand[2 ] * np.sqrt(petr_var) * np.sqrt(vale_var)
cov_wege_petr = r_rand[0 ] * np.sqrt(wege_var) * np.sqrt(petr_var)
cov_wege_abev = r_rand[4 ] * np.sqrt(wege_var) * np.sqrt(abev_var)
cov_wege_vale = r_rand[5 ] * np.sqrt(wege_var) * np.sqrt(vale_var)
cov_abev_petr = r_rand[1 ] * np.sqrt(abev_var) * np.sqrt(petr_var)
cov_abev_wege = r_rand[4 ] * np.sqrt(abev_var) * np.sqrt(wege_var)
cov_abev_vale = r_rand[8 ] * np.sqrt(abev_var) * np.sqrt(vale_var)
cov_vale_petr = r_rand[2 ] * np.sqrt(vale_var) * np.sqrt(petr_var)
cov_vale_wege = r_rand[5 ] * np.sqrt(vale_var) * np.sqrt(wege_var)
cov_vale_abev = r_rand[8 ] * np.sqrt(vale_var) * np.sqrt(abev_var)

covs = [
    np.full(sims, petr_var),
    cov_petr_wege,
    cov_petr_abev,
    cov_petr_vale,
    cov_petr_wege,
    np.full(sims, wege_var),
    cov_wege_abev,
    cov_wege_vale,
    cov_petr_abev,
    cov_wege_abev,
    np.full(sims, abev_var),
    cov_abev_vale,
    cov_petr_vale,
    cov_wege_vale,
    cov_abev_vale,
    np.full(sims, vale_var),
]

corrs = [
    np.full(sims, 1),
    r_rand[0 ],
    r_rand[1 ],
    r_rand[2 ],
    r_rand[0 ],
    np.full(sims, 1),
```

```
    r_rand[4 ],
    r_rand[5 ],
    r_rand[1 ],
    r_rand[4 ],
    np.full(sims, 1),
    r_rand[8 ],
    r_rand[2 ],
    r_rand[5 ],
    r_rand[8 ],
    np.full(sims, 1)
]

print("Covariances")
print(covs)

## Markowitz
print("#####################")

total_cov_to_test = []
total_corrs = []
linspace_min = 100
linspace_max = 0
horizontal_min = 100
horizontal_max = 0

plt.figure(1)
plt.style.use('./mplstyles/financialgraphs.mplstyle')
plt.xlabel('Volatility')
plt.ylabel('Return')
markowitz_optimization, axes = plt.subplots(figsize=(14, 8))
markowitz_optimization2, axes2 = plt.subplots(figsize=(14, 8))
axes.set_title('Anual Expected Return x Volatility (' + str(sims) + '
↪   simulations)')
axes2.set_title('Confidence Interval - Anual Expected Return x Volatility (' +
↪   str(sims) + ' simulations; p = 95%)')

print("PLEASE WAIT")

i = 0
while(i < sims):
    cov_to_test = [[covs[0 ][i],
                   covs[1 ][i],
                   covs[2 ][i],
                   covs[3 ][i]],
                   [covs[4 ][i],
                   covs[5 ][i],
                   covs[6 ][i],
                   covs[7 ][i]],
                   [covs[8 ][i],
                   covs[9 ][i],
                   covs[10][i],
                   covs[11][i]],
                   [covs[12][i],
                   covs[13][i],
                   covs[14][i],
                   covs[15][i]]]

    corrs_to_add = [[corrs[0 ][i],
```

```
                corrs[1 ][i],
                corrs[2 ][i],
                corrs[3 ][i]],
                [corrs[4 ][i],
                corrs[5 ][i],
                corrs[6 ][i],
                corrs[7 ][i]],
                [corrs[8 ][i],
                corrs[9 ][i],
                corrs[10][i],
                corrs[11][i]],
                [corrs[12][i],
                corrs[13][i],
                corrs[14][i],
                corrs[15][i]]]

total_cov_to_test.append(cov_to_test)
total_corrs.append(corrs_to_add)

cov_to_test = np.array(cov_to_test) * 252

# Define the bounds for portfolio weights (between 0 and 1)
bounds = [(0, 1)] * (len(df.columns) - 1)

# Define an initial guess for portfolio weights (equal weights)
initial_guess = [(1 / (len(df.columns) - 1))] * (len(df.columns) - 1)

# Define an initial constraint that ensures the sum of weights equals 1
constraints = [{'type':'eq','fun':lambda weights: np.sum(weights) - 1}]

# Calculate weights and returns for the optimizations and input filtering
maximum_return_weights = optimize.minimize(lambda weights: metrics(weights,
↪   np.array(ret_means), np.array(cov_to_test))[0] * -1, initial_guess,
↪   method='SLSQP', bounds=bounds, constraints=constraints,).x
maximum_return = metrics(maximum_return_weights, np.array(ret_means),
↪   np.array(cov_to_test))[0]

minimum_risk_weights = optimize.minimize(lambda weights: metrics(weights,
↪   np.array(ret_means), np.array(cov_to_test))[1], initial_guess,
↪   method='SLSQP', bounds=bounds, constraints=constraints).x
minimum_risk = metrics(minimum_risk_weights, np.array(ret_means),
↪   np.array(cov_to_test))[1]
minimum_risk_return = metrics(minimum_risk_weights, np.array(ret_means),
↪   np.array(cov_to_test))[0]

maximum_risk_weights = optimize.minimize(lambda weights: metrics(weights,
↪   np.array(ret_means), np.array(cov_to_test))[1] * -1, initial_guess,
↪   method='SLSQP', bounds=bounds, constraints=constraints).x
maximum_risk = metrics(maximum_risk_weights, np.array(ret_means),
↪   np.array(cov_to_test))[1]

sharpe_ratio_optimal_weights = optimize.minimize(lambda weights:
↪   metrics(weights, np.array(ret_means), np.array(cov_to_test))[2] * -1,
↪   initial_guess, method='SLSQP', bounds=bounds, constraints=constraints).x

if calculation_type == 'risk':
    risk_tolerance = expected_value
    # Find optimal weights that maximize return (minimize return * -1)
```

```
    constraints.append({'type': 'ineq', 'fun': lambda weights:
    ↪  float(risk_tolerance) - metrics(weights, np.array(ret_means),
    ↪  np.array(cov_to_test))[1]})
    optimal_weights = optimize.minimize(lambda weights: metrics(weights,
    ↪  np.array(ret_means), np.array(cov_to_test))[0] * -1, initial_guess,
    ↪  method='SLSQP', bounds=bounds, constraints=constraints).x

elif calculation_type == 'return':
    expected_return = expected_value
    # Find optimal weights that minimize risk
    constraints.append({'type': 'eq', 'fun': lambda weights: metrics(weights,
    ↪  np.array(ret_means), np.array(cov_to_test))[0] -
    ↪  float(expected_return)})
    optimal_weights = optimize.minimize(lambda weights: metrics(weights,
    ↪  np.array(ret_means), np.array(cov_to_test))[1], initial_guess,
    ↪  method='SLSQP', bounds=bounds, constraints=constraints).x

else:
    exit(1)

if(linspace_min > minimum_risk_return) :
    linspace_min = minimum_risk_return

if(linspace_max < maximum_return) :
    linspace_max = maximum_return

if(horizontal_min > minimum_risk) :
    horizontal_min = minimum_risk

if(horizontal_max < maximum_risk) :
    horizontal_max = maximum_risk

# Generate a range of target returns
target_returns = np.linspace(linspace_min, linspace_max, 100)

efficient_frontier_volatility = []
efficient_frontier_return = []

# Calculate the efficient frontier
for target_return in target_returns:
    # Optimize for minimum volatility given the target return
    constraints = [
        {'type': 'eq', 'fun': lambda weights: np.sum(weights) - 1},
        {'type': 'eq', 'fun': lambda weights: metrics(weights,
        ↪  np.array(ret_means), np.array(cov_to_test))[0] - target_return}
    ]
    result = optimize.minimize(lambda weights: metrics(weights,
    ↪  np.array(ret_means), np.array(cov_to_test))[1], initial_guess,
    ↪  method='SLSQP', bounds=bounds, constraints=constraints)

    efficient_frontier_volatility.append(result.fun)
    efficient_frontier_return.append(target_return)

#
# Graph
#
```

```
    axes.plot(efficient_frontier_volatility, efficient_frontier_return,
    ↪  '#3fccff', zorder=2)
    axes.scatter(metrics(optimal_weights, np.array(ret_means),
    ↪  np.array(cov_to_test))[1], metrics(optimal_weights, np.array(ret_means),
    ↪  np.array(cov_to_test))[0], marker='o', color='#ff00ff', zorder=3)
    axes.scatter(metrics(sharpe_ratio_optimal_weights, np.array(ret_means),
    ↪  np.array(cov_to_test))[1], metrics(sharpe_ratio_optimal_weights,
    ↪  np.array(ret_means), np.array(cov_to_test))[0], marker='o',
    ↪  color='#ffaa00', zorder=3)

    axes.xaxis.set_major_formatter(mplticker.PercentFormatter(1.0))
    axes2.xaxis.set_major_formatter(mplticker.PercentFormatter(1.0))
    axes.yaxis.set_major_formatter(mplticker.PercentFormatter(1.0))
    axes2.yaxis.set_major_formatter(mplticker.PercentFormatter(1.0))

    # Enable cursor interaction on the graph
    cursor = mplcursors.cursor()
    @cursor.connect("add")
    def on_add(sel):
        sel.annotation.get_bbox_patch().set(fc='gray', alpha=0.8)
        sel.annotation.get_bbox_patch().set_edgecolor('gray')
        sel.annotation.arrow_patch.set_color('white')
        sel.annotation.arrow_patch.set_arrowstyle('-')

    i += 1

i = 0

covs_mean = []
corrs_mean = []

while(i < 4):
    cov_0 = st.mean(np.transpose(total_cov_to_test)[i][0])
    cov_1 = st.mean(np.transpose(total_cov_to_test)[i][1])
    cov_2 = st.mean(np.transpose(total_cov_to_test)[i][2])
    cov_3 = st.mean(np.transpose(total_cov_to_test)[i][3])

    corrs_0 = st.mean(np.transpose(total_corrs)[i][0])
    corrs_1 = st.mean(np.transpose(total_corrs)[i][1])
    corrs_2 = st.mean(np.transpose(total_corrs)[i][2])
    corrs_3 = st.mean(np.transpose(total_corrs)[i][3])

    covs_mean.append([cov_0, cov_1, cov_2, cov_3])
    corrs_mean.append([corrs_0, corrs_1, corrs_2, corrs_3])
    i += 1

covs_mean = np.array(covs_mean) * 252

print(np.transpose(total_corrs))

corrs_mean_0 = np.delete(np.array(corrs_mean)[0], 0)
corrs_mean_1 = np.delete(np.array(corrs_mean)[1], 1)
corrs_mean_2 = np.delete(np.array(corrs_mean)[2], 2)
corrs_mean_3 = np.delete(np.array(corrs_mean)[3], 3)
print(np.array(corrs_mean))
corrs_mean = np.array([corrs_mean_0, corrs_mean_1, corrs_mean_2, corrs_mean_3])

z_mean_calc = calcZ(corrs_mean)
```

```
z_sd_calc = 1 / np.sqrt(sims - 3)

z_stat = (-1) * stats.norm.ppf(0.025)

z_min = z_mean_calc - (z_stat * z_sd_calc / np.sqrt(sims))
z_max = z_mean_calc + (z_stat * z_sd_calc / np.sqrt(sims))

r_min = calcR(np.array(z_min)).flatten()
r_max = calcR(np.array(z_max)).flatten()

cov_min_petr_wege = r_min[0 ] * np.sqrt(petr_var) * np.sqrt(wege_var)
cov_min_petr_abev = r_min[1 ] * np.sqrt(petr_var) * np.sqrt(abev_var)
cov_min_petr_vale = r_min[2 ] * np.sqrt(petr_var) * np.sqrt(vale_var)
cov_min_wege_petr = r_min[3 ] * np.sqrt(wege_var) * np.sqrt(petr_var)
cov_min_wege_abev = r_min[4 ] * np.sqrt(wege_var) * np.sqrt(abev_var)
cov_min_wege_vale = r_min[5 ] * np.sqrt(wege_var) * np.sqrt(vale_var)
cov_min_abev_petr = r_min[6 ] * np.sqrt(abev_var) * np.sqrt(petr_var)
cov_min_abev_wege = r_min[7 ] * np.sqrt(abev_var) * np.sqrt(wege_var)
cov_min_abev_vale = r_min[8 ] * np.sqrt(abev_var) * np.sqrt(vale_var)
cov_min_vale_petr = r_min[9 ] * np.sqrt(vale_var) * np.sqrt(petr_var)
cov_min_vale_wege = r_min[10] * np.sqrt(vale_var) * np.sqrt(wege_var)
cov_min_vale_abev = r_min[11] * np.sqrt(vale_var) * np.sqrt(abev_var)

covs_min = [
    [petr_var,
    cov_min_petr_wege,
    cov_min_petr_abev,
    cov_min_petr_vale],
    [cov_min_wege_petr,
    wege_var,
    cov_min_wege_abev,
    cov_min_wege_vale],
    [cov_min_abev_petr,
    cov_min_abev_wege,
    abev_var,
    cov_min_abev_vale],
    [cov_min_vale_petr,
    cov_min_vale_wege,
    cov_min_vale_abev,
    vale_var],
]

covs_min = np.array(covs_min) * 252
print(covs_min)
print("\n\n")

cov_max_petr_wege = r_max[0 ] * np.sqrt(petr_var) * np.sqrt(wege_var)
cov_max_petr_abev = r_max[1 ] * np.sqrt(petr_var) * np.sqrt(abev_var)
cov_max_petr_vale = r_max[2 ] * np.sqrt(petr_var) * np.sqrt(vale_var)
cov_max_wege_petr = r_max[3 ] * np.sqrt(wege_var) * np.sqrt(petr_var)
cov_max_wege_abev = r_max[4 ] * np.sqrt(wege_var) * np.sqrt(abev_var)
cov_max_wege_vale = r_max[5 ] * np.sqrt(wege_var) * np.sqrt(vale_var)
cov_max_abev_petr = r_max[6 ] * np.sqrt(abev_var) * np.sqrt(petr_var)
cov_max_abev_wege = r_max[7 ] * np.sqrt(abev_var) * np.sqrt(wege_var)
cov_max_abev_vale = r_max[8 ] * np.sqrt(abev_var) * np.sqrt(vale_var)
cov_max_vale_petr = r_max[9 ] * np.sqrt(vale_var) * np.sqrt(petr_var)
cov_max_vale_wege = r_max[10] * np.sqrt(vale_var) * np.sqrt(wege_var)
cov_max_vale_abev = r_max[11] * np.sqrt(vale_var) * np.sqrt(abev_var)
```

```
covs_max = [
    [petr_var,
    cov_max_petr_wege,
    cov_max_petr_abev,
    cov_max_petr_vale],
    [cov_max_wege_petr,
    wege_var,
    cov_max_wege_abev,
    cov_max_wege_vale],
    [cov_max_abev_petr,
    cov_max_abev_wege,
    abev_var,
    cov_max_abev_vale],
    [cov_max_vale_petr,
    cov_max_vale_wege,
    cov_max_vale_abev,
    vale_var]
]

covs_max = np.array(covs_max) * 252
print(covs_max)

# Generate a range of target returns
target_returns = np.linspace(linspace_min, linspace_max, 100)

medium_efficient_frontier_volatility = []
medium_efficient_frontier_return = []

min_efficient_frontier_volatility = []
min_efficient_frontier_return = []

max_efficient_frontier_volatility = []
max_efficient_frontier_return = []

bounds = [(0, 1)] * (len(df.columns) - 1)
initial_guess = [(1 / (len(df.columns) - 1))] * (len(df.columns) - 1)

# Calculate the efficient frontier
for target_return in target_returns:
    # Optimize for minimum volatility given the target return
    constraints = [
        {'type': 'eq', 'fun': lambda weights: np.sum(weights) - 1},
        {'type': 'eq', 'fun': lambda weights: metrics(weights,
         ↪ np.array(ret_means), np.array(covs_mean))[0] - target_return}
    ]
    result = optimize.minimize(lambda weights: metrics(weights,
    ↪ np.array(ret_means), np.array(covs_mean))[1], initial_guess,
    ↪ method='SLSQP', bounds=bounds, constraints=constraints)

    medium_efficient_frontier_volatility.append(result.fun)
    medium_efficient_frontier_return.append(target_return)

    constraints = [
        {'type': 'eq', 'fun': lambda weights: np.sum(weights) - 1},
        {'type': 'eq', 'fun': lambda weights: metrics(weights,
         ↪ np.array(ret_means), np.array(covs_min))[0] - target_return}
    ]
```

```
    result = optimize.minimize(lambda weights: metrics(weights,
    ↪  np.array(ret_means), np.array(covs_min))[1], initial_guess,
    ↪  method='SLSQP', bounds=bounds, constraints=constraints)

    min_efficient_frontier_volatility.append(result.fun)
    min_efficient_frontier_return.append(target_return)

    constraints = [
        {'type': 'eq', 'fun': lambda weights: np.sum(weights) - 1},
        {'type': 'eq', 'fun': lambda weights: metrics(weights,
        ↪  np.array(ret_means), np.array(covs_max))[0] - target_return}
    ]
    result = optimize.minimize(lambda weights: metrics(weights,
    ↪  np.array(ret_means), np.array(covs_max))[1], initial_guess,
    ↪  method='SLSQP', bounds=bounds, constraints=constraints)

    max_efficient_frontier_volatility.append(result.fun)
    max_efficient_frontier_return.append(target_return)

axes.plot(medium_efficient_frontier_volatility, medium_efficient_frontier_return,
↪  '#ff0000', zorder=4)
axes2.plot(medium_efficient_frontier_volatility,
↪  medium_efficient_frontier_return, '#ff0000', zorder=4)
axes2.plot(min_efficient_frontier_volatility, min_efficient_frontier_return,
↪  '#00ff00', zorder=4)
axes2.plot(max_efficient_frontier_volatility, max_efficient_frontier_return,
↪  '#00ff00', zorder=4)

optimal_weights_mean = []
sharpe_ratio_optimal_weights_mean = []

if calculation_type == 'risk':
    risk_tolerance = expected_value
    constraints = [{'type':'eq','fun':lambda weights: np.sum(weights) - 1},
    ↪  {'type': 'ineq', 'fun': lambda weights: float(risk_tolerance) -
    ↪  metrics(weights, np.array(ret_means), np.array(covs_mean))[1]}]
    optimal_weights_mean = optimize.minimize(lambda weights: metrics(weights,
    ↪  np.array(ret_means), np.array(covs_mean))[0] * -1, initial_guess,
    ↪  method='SLSQP', bounds=bounds, constraints=constraints).x
    sharpe_ratio_optimal_weights_mean = optimize.minimize(lambda weights:
    ↪  metrics(weights, np.array(ret_means), np.array(covs_mean))[2] * -1,
    ↪  initial_guess, method='SLSQP', bounds=bounds, constraints=constraints).x

    axes.vlines(float(expected_value), linspace_min, linspace_max, linewidth=1,
    ↪  color='#ff00ff')
    axes2.vlines(float(expected_value), linspace_min, linspace_max, linewidth=1,
    ↪  color='#ff00ff')

elif calculation_type == 'return':
    expected_return = expected_value
    constraints = [{'type':'eq','fun':lambda weights: np.sum(weights) - 1},
    ↪  {'type': 'eq', 'fun': lambda weights: metrics(weights,
    ↪  np.array(ret_means), np.array(covs_mean))[0] - float(expected_return)}]
    optimal_weights_mean = optimize.minimize(lambda weights: metrics(weights,
    ↪  np.array(ret_means), np.array(covs_mean))[1], initial_guess,
    ↪  method='SLSQP', bounds=bounds, constraints=constraints).x
```

```
    sharpe_ratio_optimal_weights_mean = optimize.minimize(lambda weights:
    ↪  metrics(weights, np.array(ret_means), np.array(covs_mean))[2] * -1,
    ↪  initial_guess, method='SLSQP', bounds=bounds, constraints=constraints).x


    axes.hlines(float(expected_value), horizontal_min, horizontal_max,
    ↪  linewidth=1, color='#ff00ff')
    axes2.hlines(float(expected_value), horizontal_min, horizontal_max,
    ↪  linewidth=1, color='#ff00ff')

else:
    exit(1)

axes.scatter(metrics(optimal_weights_mean, np.array(ret_means),
↪  np.array(covs_mean))[1], metrics(optimal_weights_mean, np.array(ret_means),
↪  np.array(covs_mean))[0], marker='o', color='#ff00ff', zorder=3)
axes2.scatter(metrics(optimal_weights_mean, np.array(ret_means),
↪  np.array(covs_mean))[1], metrics(optimal_weights_mean, np.array(ret_means),
↪  np.array(covs_mean))[0], marker='o', color='#ff00ff', zorder=3)
axes.scatter(metrics(sharpe_ratio_optimal_weights_mean, np.array(ret_means),
↪  np.array(covs_mean))[1], metrics(sharpe_ratio_optimal_weights,
↪  np.array(ret_means), np.array(covs_mean))[0], marker='o', color='#ffaa00',
↪  zorder=3)
axes2.scatter(metrics(sharpe_ratio_optimal_weights_mean, np.array(ret_means),
↪  np.array(covs_mean))[1], metrics(sharpe_ratio_optimal_weights,
↪  np.array(ret_means), np.array(covs_mean))[0], marker='o', color='#ffaa00',
↪  zorder=3)

legend_text = "Median Values:\n"
legend_text += '\n'.join([f'{metric}: {metrics(optimal_weights_mean,
↪  np.array(ret_means), np.array(covs_mean))[j]:.2%}' for j, metric in
↪  enumerate(['Expected Return','Volatility','Sharpe Ratio'])]) + '\n\n'
legend_text += '\n'.join([f'{asset}\'s weight: {j:.2%}' for asset, j in
↪  zip(['PETR4', 'WEGE3', 'ABEV3', 'VALE3'], optimal_weights_mean)]) + '\n'
axes.text(horizontal_max, linspace_max, legend_text, color="#ffffff")
axes2.text(horizontal_max, linspace_max, legend_text, color="#ffffff")
axes.set_xlim(horizontal_min, horizontal_max)
axes2.set_xlim(horizontal_min, horizontal_max)
axes.set_ylim(linspace_min, linspace_max)
axes2.set_ylim(linspace_min, linspace_max)


legend_elements = [Line2D([0], [0], color='#3fccff', label='Efficient
↪  Frontiers'),
                   Line2D([0], [0], color='#ff0000', label='Mean Efficient
                   ↪  Frontier'),
                   Line2D([0], [0], marker='o', color='#ff00ff', label='Optimal
                   ↪  Portfolio'),
                   Line2D([0], [0], marker='o', markersize=20, color='#ffaa00',
                   ↪  label='Max. Sharpe Ratio')]

axes.legend(handles=legend_elements, loc='upper left')

legend_elements = [Line2D([0], [0], color='#3fccff', label='Efficient
↪  Frontiers'),
                   Line2D([0], [0], color='#ff0000', label='Mean Efficient
                   ↪  Frontier'),
```

```
                              Line2D([0], [0], color='#00ff00', label='Confidence Interval
                              ↪  Limits (95%)'),
                              Line2D([0], [0], marker='o', color='#ff00ff', label='Optimal
                              ↪  Portfolio'),
                              Line2D([0], [0], marker='o', markersize=20, color='#ffaa00',
                              ↪  label='Max. Sharpe Ratio')]

    axes2.legend(handles=legend_elements, loc='upper left')

    plt.show()

__main__()
```

C. Código-fonte R para Análise de Dados e Geração de Gráficos e Histogramas

```
library(ggplot2)
library(tidyverse)
library(readr)
library(quadprog)

df_assets <- read.csv("C:/Users/gabri/Downloads/assets.csv")
df_assets <- na.omit(df_assets)

#
# Análise quantitativa
#

# Petrobras
mean_petr <- mean(df_assets$PETR4.SA)
median_petr <- median(df_assets$PETR4.SA)
devPad_petr <- sd(df_assets$PETR4.SA)
var_petr <- var(df_assets$PETR4.SA)
mode_petr <- as.numeric(names(sort(-table(df_assets$PETR4.SA)))[1])

# WEG
mean_wege <- mean(df_assets$WEGE3.SA)
median_wege <- median(df_assets$WEGE3.SA)
devPad_wege <- sd(df_assets$WEGE3.SA)
var_wege <- var(df_assets$WEGE3.SA)
mode_wege <- as.numeric(names(sort(-table(df_assets$WEGE3.SA)))[1])

# Ambev
mean_abev <- mean(df_assets$ABEV3.SA)
median_abev <- median(df_assets$ABEV3.SA)
devPad_abev <- sd(df_assets$ABEV3.SA)
var_abev <- var(df_assets$ABEV3.SA)
mode_abev <- as.numeric(names(sort(-table(df_assets$ABEV3.SA)))[1])

# Vale
mean_vale <- mean(df_assets$VALE3.SA)
median_vale <- median(df_assets$VALE3.SA)
devPad_vale <- sd(df_assets$VALE3.SA)
var_vale <- var(df_assets$VALE3.SA)
mode_vale <- as.numeric(names(sort(-table(df_assets$VALE3.SA)))[1])

#
# Gráficos
#
```

```r
#Histograma de preços
ggplot(df_assets, aes(x = PETR4.SA)) +
  geom_histogram(fill = "steelblue", bins = 30) +
  labs(title = "PETR4.SA - Histograma de Preços", x = "Valor", y = "Frequência") +
  theme_minimal() +
  theme(plot.background = element_rect(fill = "white"))
ggsave("C:/Users/gabri/Downloads/petrobras_prices_histogram.png")

ggplot(df_assets, aes(x = WEGE3.SA)) +
  geom_histogram(fill = "steelblue", bins = 30) +
  labs(title = "WEGE3.SA - Histograma de Preços", x = "Valor", y = "Frequência") +
  theme_minimal() +
  theme(plot.background = element_rect(fill = "white"))
ggsave("C:/Users/gabri/Downloads/weg_prices_histogram.png")

ggplot(df_assets, aes(x = ABEV3.SA)) +
  geom_histogram(fill = "steelblue", bins = 30) +
  labs(title = "ABEV3.SA - Histograma de Preços", x = "Valor", y = "Frequência") +
  theme_minimal() +
  theme(plot.background = element_rect(fill = "white"))
ggsave("C:/Users/gabri/Downloads/ambev_prices_histogram.png")

ggplot(df_assets, aes(x = VALE3.SA)) +
  geom_histogram(fill = "steelblue", bins = 30) +
  labs(title = "VALE3.SA - Histograma de Preços", x = "Valor", y = "Frequência") +
  theme_minimal() +
  theme(plot.background = element_rect(fill = "white"))
ggsave("C:/Users/gabri/Downloads/vale_prices_histogram.png")

# Gráfico de evolução temporal do preço
ggplot(df_assets, aes(x = as.Date(Date), y = PETR4.SA)) +
  geom_line(color = "steelblue") +
  labs(title = "PETR4.SA - Evolução Temporal de Preços", x = "Data", y = "Preço de
  ↪ fechamento (reais)") +
  theme_minimal() +
  theme(plot.background = element_rect(fill = "white"))
ggsave("C:/Users/gabri/Downloads/evolucao_temporal_petr.png")

ggplot(df_assets, aes(x = as.Date(Date), y = WEGE3.SA)) +
  geom_line(color = "steelblue") +
  labs(title = "WEGE3.SA - Evolução Temporal de Preços", x = "Data", y = "Preço de
  ↪ fechamento (reais)") +
  theme_minimal() +
  theme(plot.background = element_rect(fill = "white"))
ggsave("C:/Users/gabri/Downloads/evolucao_temporal_wege.png")

ggplot(df_assets, aes(x = as.Date(Date), y = ABEV3.SA)) +
  geom_line(color = "steelblue") +
  labs(title = "ABEV3.SA - Evolução Temporal de Preços", x = "Data", y = "Preço de
  ↪ fechamento (reais)") +
  theme_minimal() +
  theme(plot.background = element_rect(fill = "white"))
ggsave("C:/Users/gabri/Downloads/evolucao_temporal_abev.png")

ggplot(df_assets, aes(x = as.Date(Date), y = VALE3.SA)) +
  geom_line(color = "steelblue") +
  labs(title = "VALE3.SA - Evolução Temporal de Preços", x = "Data", y = "Preço de
  ↪ fechamento (reais)") +
```

```
  theme_minimal() +
  theme(plot.background = element_rect(fill = "white"))
ggsave("C:/Users/gabri/Downloads/evolucao_temporal_vale.png")

# Calcular os retornos diários
df_assets_returns <- df_assets %>%
  mutate(
    return_petr = PETR4.SA / lag(PETR4.SA) - 1,
    return_wege = WEGE3.SA / lag(WEGE3.SA) - 1,
    return_abev = ABEV3.SA / lag(ABEV3.SA) - 1,
    return_vale = VALE3.SA / lag(VALE3.SA) - 1
  )
df_assets_returns <- na.omit(df_assets_returns)

# Histograma de retornos
ggplot(df_assets_returns, aes(x = return_petr)) +
  geom_histogram(fill = "steelblue", bins = 30) +
  labs(title = "PETR4.SA - Histograma de Retornos Diários", x = "Retorno", y =
  ↪  "Frequência") +
  theme_minimal() +
  theme(plot.background = element_rect(fill = "white"))
ggsave("C:/Users/gabri/Downloads/petrobras_returns_histogram.png")

ggplot(df_assets_returns, aes(x = return_wege)) +
  geom_histogram(fill = "steelblue", bins = 30) +
  labs(title = "WEGE3.SA - Histograma de Retornos Diários", x = "Retorno", y =
  ↪  "Frequência") +
  theme_minimal() +
  theme(plot.background = element_rect(fill = "white"))
ggsave("C:/Users/gabri/Downloads/weg_returns_histogram.png")

ggplot(df_assets_returns, aes(x = return_abev)) +
  geom_histogram(fill = "steelblue", bins = 30) +
  labs(title = "ABEV3.SA - Histograma de Retornos Diários", x = "Retorno", y =
  ↪  "Frequência") +
  theme_minimal() +
  theme(plot.background = element_rect(fill = "white"))
ggsave("C:/Users/gabri/Downloads/ambev_returns_histogram.png")

ggplot(df_assets_returns, aes(x = return_vale)) +
  geom_histogram(fill = "steelblue", bins = 30) +
  labs(title = "VALE3.SA - Histograma de Retornos Diários", x = "Retorno", y =
  ↪  "Frequência") +
  theme_minimal() +
  theme(plot.background = element_rect(fill = "white"))
ggsave("C:/Users/gabri/Downloads/vale_returns_histogram.png")

# Gráfico de linha da evolução temporal do retorno diário
ggplot(df_assets_returns, aes(x = as.Date(Date), y = return_petr)) +
  geom_line(color = "steelblue") +
  labs(title = "PETR4.SA - Evolução Temporal do Retorno Diário", x = "Data", y = "Log
  ↪  Retorno Diário") +
  theme_minimal() +
  theme(plot.background = element_rect(fill = "white"))
ggsave("C:/Users/gabri/Downloads/evolucao_temporal_retorno_petr.png")

ggplot(df_assets_returns, aes(x = as.Date(Date), y = return_wege)) +
  geom_line(color = "steelblue") +
```

```
  labs(title = "WEGE3.SA - Retorno Diário", x = "Data", y = "Log Retorno Diário") +
  theme_minimal() +
  theme(plot.background = element_rect(fill = "white"))
ggsave("C:/Users/gabri/Downloads/evolucao_temporal_retorno_wege.png")

ggplot(df_assets_returns, aes(x = as.Date(Date), y = return_abev)) +
  geom_line(color = "steelblue") +
  labs(title = "ABEV3.SA - Retorno Diário", x = "Data", y = "Log Retorno Diário") +
  theme_minimal() +
  theme(plot.background = element_rect(fill = "white"))
ggsave("C:/Users/gabri/Downloads/evolucao_temporal_retorno_abev.png")

ggplot(df_assets_returns, aes(x = as.Date(Date), y = return_vale)) +
  geom_line(color = "steelblue") +
  labs(title = "VALE3.SA - Retorno Diário", x = "Data", y = "Log Retorno Diário") +
  theme_minimal() +
  theme(plot.background = element_rect(fill = "white"))
ggsave("C:/Users/gabri/Downloads/evolucao_temporal_retorno_vale.png")

#
# Matrizes de cor e cov
#

# Calcular a matriz de correlação
correlation_matrix <- cor(df_assets_returns[, c("return_petr", "return_wege",
↪  "return_abev", "return_vale")])

# Visualizar a matriz de correlação
print("Matriz de Correlação:")
print(correlation_matrix)

# Calcular a matriz de covariâncias
cov_matrix <- cov(df_assets_returns[, c("return_petr", "return_wege", "return_abev",
↪  "return_vale")])

# Número de dias de negociação no ano
dias_no_ano <- 252  # Considerando 252 dias úteis de negociação em um ano

# Anualizar a matriz de covariâncias
cov_matrix_anualizada <- cov_matrix * dias_no_ano

# Visualizar a matriz de covariâncias anualizada
print("Matriz de Covariâncias Anualizada:")
print(cov_matrix_anualizada)
```

D. Código-fonte R para Teste de Hipótese de 1 Parâmetro (Retorno Diário Positivo)

```
library(tidyverse)
library(readr)

# Carregar os dados
df_assets <- read.csv("C:/Users/gabri/Downloads/assets.csv")
df_assets <- na.omit(df_assets)

# Calcula os retornos diários
df_assets_returns <- df_assets %>%
  mutate(
    return_petr = (PETR4.SA / lag(PETR4.SA)) - 1,
    return_wege = (WEGE3.SA / lag(WEGE3.SA)) - 1,
```

```r
    return_abev = (ABEV3.SA / lag(ABEV3.SA)) - 1,
    return_vale = (VALE3.SA / lag(VALE3.SA)) - 1
  )
df_assets_returns <- na.omit(df_assets_returns)

# Nível de confiança e significância
nivelConfianca <- 0.95
alpha <- 1 - nivelConfianca

# Função para realizar o teste de hipótese
one_sample_t_test_greater_than_zero <- function(data, alpha = 0.05) {
  result <- t.test(data,
                   alternative = "greater",
                   mu = 0,
                   conf.level = 1 - alpha)

  return(result)
}

# Realizando o teste para cada ativo
test_petr <- one_sample_t_test_greater_than_zero(df_assets_returns$return_petr)
test_wege <- one_sample_t_test_greater_than_zero(df_assets_returns$return_wege)
test_abev <- one_sample_t_test_greater_than_zero(df_assets_returns$return_abev)
test_vale <- one_sample_t_test_greater_than_zero(df_assets_returns$return_vale)

# Extraindo os resultados
results <- data.frame(
  Ativo = c("PETR4.SA", "WEGE3.SA", "ABEV3.SA", "VALE3.SA"),
  p_valor = c(test_petr$p.value, test_wege$p.value, test_abev$p.value,
  ↪  test_vale$p.value),
  Teste = c(test_petr$method, test_wege$method, test_abev$method, test_vale$method),
  Média = c(mean(df_assets_returns$return_petr), mean(df_assets_returns$return_wege),
  ↪  mean(df_assets_returns$return_abev), mean(df_assets_returns$return_vale))
)

# Interpretando os resultados
results <- results %>%
  mutate(
    `Hipótese Nula Rejeitada` = ifelse(p_valor < alpha, "Sim", "Não")
  )

print(results)
```

E. Código-fonte R para ANOVA de 1 e 2 Fatores e Regressão Linear Simples

```r
# remove.packages("rlang")
# install.packages("rlang")
# install.packages("devtools");
# install.packages("tidyverse");
# install.packages("ggplot2");
# install.packages("readr")
# install.packages("gridExtra")
# install.packages("dplyr")

# utilidades:
populational_stdev <- function(data) {
  mean <- mean(data)
  square_sum <- sum((data - mean) ^ 2)
  result <- sqrt(square_sum/length(data))
```

```r
  return(result)
}

sample_stdev <- function(data) {
  mean <- mean(data)
  square_sum <- sum((data - mean) ^ 2)
  result <- sqrt(square_sum / (length(data) - 1))
  return(result)
}

populational_var <- function(data) {
  mean <- mean(data)
  square_sum <- sum((data - mean) ^ 2)
  result <- square_sum/length(data)
  return(result)
}

sample_var <- function(data) {
  mean <- mean(data)
  square_sum <- sum((data - mean) ^ 2)
  result <- square_sum / (length(data) - 1)
  return(result)
}


print_test <- function(test, sample_value, crit_value, crit_value2=NULL) {
  if(test == ">") {
    if(sample_value > crit_value) {
      print(paste("REJECT H0 - ", sample_value, " > ", crit_value))
    }

    else {
      print(paste("ACCEPT H0 - ", sample_value, " <= ", crit_value))
    }
  }

  else if(test == "<") {
    if(sample_value < crit_value) {
      print(paste("REJECT H0 - ", sample_value, " < ", crit_value))
    }

    else {
      print(paste("ACCEPT H0 - ", sample_value, " >= ", crit_value))
    }
  }

  else if(test == "!=" && !is.null(crit_value2)) {
    if(sample_value < crit_value || sample_value > crit_value2) {
      print(paste("REJECT H0 - ", sample_value, " is not between ", crit_value, " and
      ↪  ", crit_value2))
    }

    else {
      print(paste("ACCEPT H0 - ", crit_value, " <= ", sample_value, " <= ",
      ↪  crit_value2))
    }
  }
}
```

```
# Carregar bibliotecas necessárias
library(tidyverse)
library(ggplot2)
library(readr)
library(gridExtra)
library(dplyr)

# Carregar dados
data <- read.csv("assets.csv")
data <- na.omit(data)

# Converter a coluna Date para o formato de data
data$Date <- as.Date(data$Date, format="%Y-%m-%d")

# Calcular retornos diários
data$ret_PETR4 <- c(NA, diff(log(data$PETR4)))
data$ret_WEGE3 <- c(NA, diff(log(data$WEGE3)))
data$ret_ABEV3 <- c(NA, diff(log(data$ABEV3)))
data$ret_VALE3 <- c(NA, diff(log(data$VALE3)))

# Remover linhas com NA
data <- na.omit(data)

# Dividir o dataframe em quatro períodos
df_periodo1 <- subset(data, Date >= as.Date('2000-01-01') & Date <=
↪  as.Date('2006-12-31'))
df_periodo2 <- subset(data, Date >= as.Date('2007-01-01') & Date <=
↪  as.Date('2012-12-31'))
df_periodo3 <- subset(data, Date >= as.Date('2013-01-01') & Date <=
↪  as.Date('2016-08-31'))
df_periodo4 <- subset(data, Date >= as.Date('2016-09-01') & Date <=
↪  as.Date('2024-04-05'))


# --------------------------- ANOVA DE 1 FATOR ---------------------------
# Nível de significância
alpha <- 0.05

# Selecionar os dados do período 1
df <- df_periodo1

# Calcular médias dos grupos
mean_PETR4 <- mean(df$ret_PETR4)
mean_WEGE3 <- mean(df$ret_WEGE3)
mean_ABEV3 <- mean(df$ret_ABEV3)
mean_VALE3 <- mean(df$ret_VALE3)

# Calcular média geral
mean_total <- mean(c(df$ret_PETR4, df$ret_WEGE3, df$ret_ABEV3, df$ret_VALE3))

# Calcular Soma de Quadrados Entre Grupos (SSB)
n_PETR4 <- length(df$ret_PETR4)
n_WEGE3 <- length(df$ret_WEGE3)
n_ABEV3 <- length(df$ret_ABEV3)
n_VALE3 <- length(df$ret_VALE3)

SSB <- n_PETR4 * (mean_PETR4 - mean_total)^2 +
```

```
        n_WEGE3 * (mean_WEGE3 - mean_total)^2 +
        n_ABEV3 * (mean_ABEV3 - mean_total)^2 +
        n_VALE3 * (mean_VALE3 - mean_total)^2

# Calcular Soma de Quadrados Dentro dos Grupos (SSW)
SSW_PETR4 <- sum((df$ret_PETR4 - mean_PETR4)^2)
SSW_WEGE3 <- sum((df$ret_WEGE3 - mean_WEGE3)^2)
SSW_ABEV3 <- sum((df$ret_ABEV3 - mean_ABEV3)^2)
SSW_VALE3 <- sum((df$ret_VALE3 - mean_VALE3)^2)

SSW <- SSW_PETR4 + SSW_WEGE3 + SSW_ABEV3 + SSW_VALE3

# Calcular graus de liberdade
df_B <- 4 - 1  # número de grupos - 1
df_W <- (n_PETR4 + n_WEGE3 + n_ABEV3 + n_VALE3) - 4  # número total de observações -
↪  número de grupos

# Calcular Quadrados Médios
MSB <- SSB / df_B
MSW <- SSW / df_W

# Calcular estatística F
F <- MSB / MSW

# Calcular valor-p
p_value <- pf(F, df_B, df_W, lower.tail = FALSE)

# Exibir resultados
anova_table <- data.frame(
  Source = c("Between Groups", "Within Groups", "Total"),
  Sum_of_Squares = c(SSB, SSW, SSB + SSW),
  Degrees_of_Freedom = c(df_B, df_W, df_B + df_W),
  Mean_Square = c(MSB, MSW, NA),
  F_Value = c(F, NA, NA),
  P_Value = c(p_value, NA, NA)
)
anova_table$F_Critical <- qf(1 - alpha, df_B, df_W)

print(anova_table)

# Regra de decisão
if (p_value < alpha) {
  decision <- "\nRejeitar H0 para ANOVA de 1 fator.\n"
} else {
  decision <- "\nNão rejeita H0 para ANOVA de 1 fator\n"
}

print(paste("Decisão:", decision))

# -------------------------- ANOVA DE 2 FATORES --------------------------

# Concatenar todos os períodos em um dataframe
df <- rbind(df_periodo1, df_periodo2, df_periodo3, df_periodo4)

# Adicionar coluna de período
df$period <- ifelse(df$Date <= as.Date('2006-12-31'), "Period 1",
                    ifelse(df$Date <= as.Date('2012-12-31'), "Period 2",
```

```
                           ifelse(df$Date <= as.Date('2016-08-31'), "Period 3",
                           ↪  "Period 4")))

# Transformar dados em formato longo
df_long <- data.frame(
  period = rep(df$period, 4),
  stock = rep(c("PETR4", "WEGE3", "ABEV3", "VALE3"), each = nrow(df)),
  return = c(df$ret_PETR4, df$ret_WEGE3, df$ret_ABEV3, df$ret_VALE3)
)

# Calcular médias
mean_total <- mean(df_long$return)
mean_period <- tapply(df_long$return, df_long$period, mean)
mean_stock <- tapply(df_long$return, df_long$stock, mean)
mean_interaction <- tapply(df_long$return, list(df_long$period, df_long$stock), mean)

# Calcular Soma de Quadrados
SSP <- sum(table(df_long$period) * (mean_period - mean_total)^2)
SSS <- sum(table(df_long$stock) * (mean_stock - mean_total)^2)
SSI <- sum(table(df_long$period, df_long$stock) * (mean_interaction -
↪  rep(mean_period, each = 4) - rep(mean_stock, times = 4) + mean_total)^2)
SSR <- sum((df_long$return - rep(mean_interaction, each = nrow(df)/4))^2)

# Calcular graus de liberdade
df_P <- length(unique(df_long$period)) - 1
df_S <- length(unique(df_long$stock)) - 1
df_I <- df_P * df_S
df_R <- nrow(df_long) - (length(unique(df_long$period)) *
↪  length(unique(df_long$stock)))

# Calcular Quadrados Médios
MSP <- SSP / df_P
MSS <- SSS / df_S
MSI <- SSI / df_I
MSR <- SSR / df_R

# Calcular estatísticas F
F_P <- MSP / MSR
F_S <- MSS / MSR
F_I <- MSI / MSR

# Calcular valores-p
p_value_P <- pf(F_P, df_P, df_R, lower.tail = FALSE)
p_value_S <- pf(F_S, df_S, df_R, lower.tail = FALSE)
p_value_I <- pf(F_I, df_I, df_R, lower.tail = FALSE)

# Exibir resultados
anova_table_2factors <- data.frame(
  Source = c("Period", "Stock", "Interaction", "Residual", "Total"),
  Sum_of_Squares = c(SSP, SSS, SSI, SSR, SSP + SSS + SSI + SSR),
  Degrees_of_Freedom = c(df_P, df_S, df_I, df_R, df_P + df_S + df_I + df_R),
  Mean_Square = c(MSP, MSS, MSI, MSR, NA),
  F_Value = c(F_P, F_S, F_I, NA, NA),
  P_Value = c(p_value_P, p_value_S, p_value_I, NA, NA)
)

print(anova_table_2factors)
```

```
decision_period <- ifelse(p_value_P < alpha, "Rejeitar H0: Há diferença significativa
↪  entre os períodos", "Não rejeitar H0: Não há diferença significativa entre os
↪  períodos")
decision_stock <- ifelse(p_value_S < alpha, "Rejeitar H0: Há diferença significativa
↪  entre os tipos de ações", "Não rejeitar H0: Não há diferença significativa entre
↪  os tipos de ações")
decision_interaction <- ifelse(p_value_I < alpha, "Rejeitar H0: Há interação
↪  significativa entre os períodos e os retornos das ações", "Não rejeitar H0: Não
↪  há interação significativa entre os períodos e os retornos das ações")

cat("Regra de Decisão da ANOVA de Dois Fatores:\n")
cat("Período:", decision_period, "\n")
cat("Ações:", decision_stock, "\n")
cat("Interação:", decision_interaction, "\n")

# -------------------------- REGRESSÃO LINEAR SIMPLES --------------------------

# Converter a coluna Date para o formato de data
data$Date <- as.Date(data$Date, format="%Y-%m-%d")

# Calcular retornos diários
data$ret_PETR4 <- c(NA, diff(log(data$PETR4)))
data$ret_WEGE3 <- c(NA, diff(log(data$WEGE3)))
data$ret_ABEV3 <- c(NA, diff(log(data$ABEV3)))
data$ret_VALE3 <- c(NA, diff(log(data$VALE3)))

# Remover linhas com NA
data <- na.omit(data)

# Função para calcular regressão linear simples
regressao_linear <- function(data, start_date, end_date, stock_name) {
  # Filtrar dados pelo período
  df <- subset(data, Date >= as.Date(start_date) & Date <= as.Date(end_date))

  # Converter a coluna Date para numérica (dias desde start_date)
  df$Date_num <- as.numeric(df$Date - as.Date(start_date))

  # Variável dependente (Y) e independente (X)
  Y <- df[[paste0("ret_", stock_name)]]
  X <- df$Date_num

  # Calcular médias
  mean_X <- mean(X)
  mean_Y <- mean(Y)

  # Calcular coeficientes
  beta1 <- sum((X - mean_X) * (Y - mean_Y)) / sum((X - mean_X)^2)
  beta0 <- mean_Y - beta1 * mean_X

  # Prever valores
  Y_hat <- beta0 + beta1 * X

  # Calcular soma dos quadrados dos resíduos (SSR) e soma total dos quadrados (SST)
  SSR <- sum((Y - Y_hat)^2)
  SST <- sum((Y - mean_Y)^2)

  # Calcular coeficiente de determinação (R^2)
  R_squared <- 1 - (SSR / SST)
```

```
  # Calcular estatísticas F
  df_model <- 1  # Graus de liberdade do modelo
  df_residual <- length(Y) - 2  # Graus de liberdade residual
  MSR <- (SST - SSR) / df_model  # Quadrado médio do modelo
  MSE <- SSR / df_residual  # Quadrado médio do erro
  F_value <- MSR / MSE

  # Calcular valor-p
  p_value <- pf(F_value, df_model, df_residual, lower.tail = FALSE)

  # Exibir resultados
  results <- data.frame(
    Coefficient = c("Intercept", "Slope"),
    Estimate = c(beta0, beta1),
    `R-squared` = c(R_squared, NA),
    `F-value` = c(F_value, NA),
    `p-value` = c(p_value, NA)
  )

  print(paste("Resultados para", stock_name, "no período de", start_date, "a",
  ↪  end_date))
  print(results)

  # Plotar os dados e a linha de regressão usando ggplot
  ggplot(df, aes(x = Date, y = Y)) +
    geom_point(color = "azure4") +
    geom_line(aes(y = Y_hat), color = "blue1", linewidth = 2) +
    labs(title = paste("Regressão Linear Simples -", stock_name, "-", start_date,
    ↪  "to", end_date),
         x = "Date",
         y = paste("Retorno Diário de", stock_name)) +
    theme_minimal() +
    theme(plot.background = element_rect(fill = "white"))

  # Salvar o gráfico em um arquivo PNG
  ggsave(paste("regressao_linear_", stock, "_", period$start_date, "_",
  ↪  period$end_date, ".png", sep = ""))
}

# Executar regressão linear para cada empresa e período
periods <- list(
  list(start_date="2000-01-01", end_date="2006-12-31"),
  list(start_date="2007-01-01", end_date="2012-12-31"),
  list(start_date="2013-01-01", end_date="2016-08-31"),
  list(start_date="2016-09-01", end_date="2024-04-05"),
  list(start_date="2000-01-01", end_date="2024-04-05") # total
)

stocks <- c("PETR4", "WEGE3", "ABEV3", "VALE3")

for (period in periods) {
  for (stock in stocks) {
    regressao_linear(data, period$start_date, period$end_date, stock)
  }
}

petr4_var <- sample_var(assets$PETR4)
```

```
vale3_var <- sample_var(assets$VALE3)
print(petr4_var)
print(vale3_var)

gl = length(assets$PETR4) - 1
a <- qf(0.05, gl, gl)
b <- qf(1-0.05, gl, gl)
print(a)
print(b)

value <- petr4_var/vale3_var
print(value)
```