

UNIVERSIDADE FEDERAL DE ALFENAS

BRUNA MANCILHA OLIVEIRA RODRIGUES

**ECONOFÍSICA, ANÁLISE DE AGRUPAMENTOS E SELEÇÕES DE
CARTEIRA PELO MODELO DE MARKOWITZ**

POÇOS DE CALDAS

2024

BRUNA MANCILHA OLIVEIRA RODRIGUES

**ECONOFÍSICA, ANÁLISE DE AGRUPAMENTOS E SELEÇÕES DE
CARTEIRA PELO MODELO DE MARKOWITZ**

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Física pelo Programa de Pós- Graduação em Física da Universidade Federal de Alfenas.

Área de concentração: Econofísica

Orientador: Prof. Dr. Cássius Anderson Miquele de Melo.

Co-orientador: Prof. Roni Antônio Mendes

POÇOS DE CALDAS

2024

Sistema de Bibliotecas da Universidade Federal de Alfenas
Biblioteca Campus Poços de Caldas

Rodrigues, Bruna Mancilha Oliveira.

Econofísica, Análise de Agrupamentos e Seleções de Carteira pelo Modelo de Markowitz / Bruna Mancilha Oliveira Rodrigues. - Poços de Caldas, MG, 2024.

92 f. : il. -

Orientador(a): Cassius Anderson Miquele de Melo.

Dissertação (Mestrado em Física) - Universidade Federal de Alfenas, Poços de Caldas, MG, 2024.

Bibliografia.

1. Agrupamento hierárquico. 2. Agrupamento não hierárquico. 3. Teoria Eficiente de Markowitz. 4. Seleção de carteiras. 5. Carteira ótima. I. de Melo, Cassius Anderson Miquele , orient. II. Título.

Ficha gerada automaticamente com dados fornecidos pelo autor.

BRUNA MANCILHA OLIVEIRA RODRIGUES

ECONOFÍSICA, ANÁLISE DE AGRUPAMENTOS E SELEÇÕES DE CARTEIRA PELO MODELO DE MARKOWITZ

O Presidente da banca examinadora abaixo assina a aprovação da Dissertação apresentada como parte dos requisitos para a obtenção do título de Mestre em Física pela Universidade Federal de Alfenas. Área de concentração: Econofísica

Aprovada em: 23 de abril de 2024

Prof. Dr. Cássius Anderson Miquele de Melo
Presidente da Banca Examinadora
Instituição: Universidade Federal de Alfenas

Prof. Dr. Marcos de Mendonça Passini
Instituição: Universidade Federal de Juiz de Fora

Prof. Dr. Marcelo Martins de Oliveira
Instituição: Universidade Federal de São João Del Rei



Documento assinado eletronicamente por **Cassius Anderson Miquele de Melo, Professor do Magistério Superior**, em 26/04/2024, às 14:39, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#)



A autenticidade deste documento pode ser conferida no site https://sei.unifal-mg.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1238667** e o código CRC **0B59AE69**.

*Dedico este trabalho,
à minha família por sempre me apoiarem
em todos os momentos de minha vida.*

AGRADECIMENTOS

Tendo consciência de que nada disso seria possível sozinha, agradeço primeiramente ao meu noivo Nicolas Pessoa Miguel, por seu apoio incondicional e incentivo durante todo caminho percorrido, pela sua presença e paciência em todos os obstáculos superados ao longo dela.

Agradeço ao meu orientador, Prof. Dr. Cássius Anderson Miquele de Melo, pela paciência e empenho ao me orientar na escrita dessa dissertação. Também agradeço ao meu co-orientador Prof. Roni Antônio Mendes pela disponibilidade e por seu apoio.

Mesmo não estando mais entre nós, agradeço à minha irmã Aléxia Mancilha Oliveira Rodrigues, cujo seu amor e sua memória continuarão me motivando em toda minha jornada.

Quero agradecer também à minha família pelo apoio que me deram, em especial minha mãe Alexandrina Helena O. Santos e meu pai Neiwtton Louis Rodrigues, pela força e pelo carinho.

Por fim, agradeço a todos aqueles que, de uma forma ou de outra, possibilitaram a realização dessa dissertação.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

LISTA DE TABELAS

Tabela 1 – Elementos por grupo - três meses	33
Tabela 2 – Elementos por grupo - um ano	35
Tabela 3 – Elementos por grupo - três anos	37
Tabela 4 – Elementos por grupo - cinco anos	39
Tabela 5 – Comparativo das métricas	42
Tabela 6 – Amplitude gráfica - Menor volatilidade	43
Tabela 7 – Comparativo das métricas	44
Tabela 8 – Amplitude gráfica - Maior retorno	44
Tabela 9 – Comparativo das métricas	45
Tabela 10 – Amplitude gráfica - Maior índice Sharpe	46
Tabela 11 – Comparativo das métricas	47
Tabela 12 – Amplitude gráfica - Menor volatilidade	48
Tabela 13 – Comparativo das métricas	49
Tabela 14 – Amplitude gráfica - Maior retorno	49
Tabela 15 – Comparativo das métricas	50
Tabela 16 – Amplitude gráfica - Maior índice Sharpe	50
Tabela 17 – Comparativo das métricas	52
Tabela 18 – Amplitude gráfica - Menor volatilidade	52
Tabela 19 – Comparativo das métricas	53
Tabela 20 – Amplitude gráfica - Maior Retorno	53
Tabela 21 – Comparativo das métricas	54
Tabela 22 – Amplitude gráfica - Maior índice Sharpe	54
Tabela 23 – Comparativo das métricas	55
Tabela 24 – Amplitude gráfica - Menor volatilidade	55
Tabela 25 – Comparativo das métricas	56
Tabela 26 – Amplitude gráfica - Maior retorno	57
Tabela 27 – Comparativo das métricas	57
Tabela 28 – Amplitude gráfica - Maior índice Sharpe	58
Tabela 29 – Elementos por grupo - um ano	62
Tabela 30 – Amplitude gráfica - Menor volatilidade	64
Tabela 31 – Amplitude gráfica - Maior retorno	65
Tabela 32 – Amplitude gráfica - Maior índice Sharpe	66
Tabela 33 – Elementos por grupo - três meses	68
Tabela 34 – Amplitude gráfica - Menor volatilidade	70
Tabela 35 – Amplitude gráfica - Maior retorno	71
Tabela 36 – Amplitude gráfica - Maior índice Sharpe	72

Tabela 37 – Elementos por grupo - um ano	73
Tabela 38 – Amplitude gráfica - Menor volatilidade	75
Tabela 39 – Amplitude gráfica - Maior retorno	76
Tabela 40 – Amplitude gráfica - Maior índice Sharpe	77
Tabela 41 – Elementos por grupo - três anos	78
Tabela 42 – Amplitude gráfica - Menor volatilidade	80
Tabela 43 – Amplitude gráfica - Maior retorno	81
Tabela 44 – Amplitude gráfica - Maior índice Sharpe	82
Tabela 45 – Elementos por grupo - cinco anos	83
Tabela 46 – Amplitude gráfica - Menor volatilidade	85
Tabela 47 – Amplitude gráfica - Maior retorno	86
Tabela 48 – Amplitude gráfica - Maior índice Sharpe	87

LISTA DE FIGURAS

Figura 1 – Dendrograma	17
Figura 2 – Exemplo de diagrama de dispersão com $k=3$	18
Figura 3 – Método de ligação simples. Figura retirada da dissertação Análise Multivariada	21
Figura 4 – Método de ligação completa. Figura retirada da dissertação Análise Multivariada	21
Figura 5 – Método de ligação média. Figura retirada da dissertação Análise Multi- variada	22
Figura 6 – Exemplo fronteira eficiente	27
Figura 7 – Lista com os 78 ativos selecionados inicialmente	31
Figura 8 – Seleção de ativos para três meses	32
Figura 9 – Dendrograma - Período de três meses	33
Figura 10 – Seleção de ativos para um ano	34
Figura 11 – Dendrograma - Período de um ano	35
Figura 12 – Seleção de ativos para três anos.	36
Figura 13 – Dendrograma - Período de três anos	37
Figura 14 – Seleção de ativos para cinco anos.	38
Figura 15 – Dendrograma - Período de cinco anos	39
Figura 16 – Gráfico menor volatilidade - três meses	42
Figura 17 – Gráfico maior retorno - três meses	43
Figura 18 – Gráfico maior Sharpe Ratio - três meses	45
Figura 19 – Menor volatilidade - Um ano	47
Figura 20 – Maior retorno - Um ano	48
Figura 21 – Gráfico maior Sharpe Ratio - um ano	50
Figura 22 – Menor volatilidade - três anos	51
Figura 23 – Maior retorno	52
Figura 24 – Maior Sharpe Ratio	53
Figura 25 – Menor Volatilidade	55
Figura 26 – Maior Retorno	56
Figura 27 – Maior Sharpe Ratio	57
Figura 28 – Quantidade de grupos para as análises de tempo	59
Figura 29 – Lista com 76 ativos	61
Figura 30 – Diagrama de dispersão	62
Figura 31 – Comparativo menor volatilidade - ANH	64
Figura 32 – Comparativo maior retorno - ANH	65
Figura 33 – Comparativo maior índice Sharpe - ANH	66

Figura 34 – Diagrama de dispersão - três meses	68
Figura 35 – Comparativo menor volatilidade com 6 grupos - 3 meses	70
Figura 36 – Comparativo maior retorno com 6 grupos - 3 meses	71
Figura 37 – Comparativo maior índice Sharpe com 6 grupos - 3 meses	72
Figura 38 – Diagrama de dispersão - um ano	73
Figura 39 – Comparativo menor volatilidade com 6 grupos - 1 ano	75
Figura 40 – Comparativo maior retorno com 6 grupos - 1 ano	76
Figura 41 – Comparativo maior índice Sharpe com 6 grupos - 1 ano	77
Figura 42 – Diagrama de dispersão - três anos	78
Figura 43 – Comparativo menor volatilidade com 6 grupos - 3 anos	80
Figura 44 – Comparativo maior retorno com 6 grupos - 3 anos	81
Figura 45 – Comparativo maior retorno com 6 grupos - 3 anos	82
Figura 46 – Diagrama de dispersão - cinco anos	83
Figura 47 – Comparativo menor volatilidade com 6 grupos - 5 anos	85
Figura 48 – Comparativo maior retorno com 6 grupos - 5 anos	86
Figura 49 – Comparativo maior índice Sharpe com 6 grupos - 5 anos	87

RESUMO

Devido ao grande número de ativos disponíveis no mercado atualmente, a construção de um portfólio ótimo têm se tornado uma tarefa complexa. Com a evolução computacional, avanços e facilidades no que se refere à análise de dados em diversas áreas, é possível perceber a necessidade de técnicas exploratórias e interpretação adequada para obter informações, principalmente para tomada de decisões ótimas, de modo que possibilite um gerenciamento eficiente. Dentro deste cenário, foi proposta a análise e resolução do problema de seleção de carteiras utilizando técnicas de econofísica. Essa dissertação trata-se de um algoritmo para a seleção de carteiras e auxílio no processo de tomada de decisão. O estudo será realizado acerca da aplicação da análise de Agrupamento combinada com a Teoria Eficiente de Markowitz, para auxiliar na compreensão da alocação dos ativos, e em busca de propiciar uma estratégia sólida de investimento. Somado a isso, para encontrar os resultados, foi realizada a construção de um modelo matemático baseado em programação em *Python*. O presente estudo busca avaliar se o modelo pode ser mais uma ferramenta significativa para auxiliar na gestão de carteiras.

Palavras-chave: agrupamento hierárquico; agrupamento não hierárquico; teoria eficiente de Markowitz; seleção de carteiras; carteira ótima; agrupamento não hierárquico.

ABSTRACT

Due to the large number of assets currently available on the market, building an optimal portfolio has become a complex task. With computational evolution, advances and facilities regarding data analysis in different areas, it is possible to perceive the need for exploratory techniques and adequate interpretation to obtain information, mainly for making optimal decisions, in a way that enables efficient management. Within this scenario, the analysis and resolution of the portfolio selection problem using econophysics techniques was proposed. This dissertation is an algorithm for selecting portfolios and assisting in the decision-making process. The study will be carried out on the application of Cluster analysis combined with Markowitz Efficient Frontier, to assist in understanding asset allocation, and in search of providing a solid investment strategy. In addition, to find the results, a mathematical model was built based on programming in *Python*. The present study seeks to evaluate whether the model can be another significant tool to assist in portfolio management.

Keywords: hierarchical cluster; non hierarchical clustering; The Markowitz efficient set; portfolio selection; optimum portfolio.

SUMÁRIO

1	INTRODUÇÃO	14
2	REFERENCIAL TEÓRICO	16
2.1	Análise de Agrupamento	16
2.1.1	Distância euclidiana	19
2.1.2	Distância de Manhattan	19
2.1.3	Distância de Minkowski	20
2.1.4	Técnicas para a construção de <i>clusters</i>	20
2.1.4.1	Ligação simples	20
2.1.4.2	Ligação completa	21
2.1.4.3	Ligação média	22
2.1.4.4	Método K-Médias	23
2.2	Teoria de Markowitz	23
2.2.1	Covariância e Correlação	26
2.2.2	Fronteira Eficiente	27
2.2.3	Índice Sharpe	28
3	AGRUPAMENTO HIERÁRQUICO	29
3.1	Seleção de Ativos por Agrupamento Hierárquico	29
3.2	Otimização da Carteira de Markowitz	30
3.3	Resultados	31
3.3.1	Observação 1: período de três meses	32
3.3.2	Observação 2: período de um ano	34
3.3.3	Observação 3: período de três anos	36
3.3.4	Observação 4: período de cinco anos	38
3.4	Resultados e Comparativos	40
3.4.1	Observação 1: período de três meses	41
3.4.1.1	Comparativo menor volatilidade	41
3.4.1.2	Comparativo maior retorno	43
3.4.1.3	Comparativo maior índice Sharpe	44
3.4.2	Observação 2: período de um ano	46
3.4.2.1	Comparativo menor volatilidade	46
3.4.2.2	Comparativo maior retorno	48
3.4.2.3	Comparativo maior índice Sharpe	49
3.4.3	Observação 3: período de três anos	51

3.4.3.1	Comparativo menor volatilidade	51
3.4.3.2	Comparativo maior retorno	52
3.4.3.3	Comparativo maior índice Sharpe	53
3.4.4	Observação 4: período de cinco anos	54
3.4.4.1	Comparativo menor volatilidade	54
3.4.4.2	Comparativo maior retorno	56
3.4.4.3	Comparativo maior índice Sharpe	57
4	AGRUPAMENTO NÃO HIERÁRQUICO	59
4.1	Aplicação com Agrupamento Não Hierárquico	60
4.2	Comparativo com Agrupamento Hierárquico	60
4.2.1	Comparativo menor volatilidade	63
4.2.2	Comparativo maior retorno	64
4.2.3	Comparativo maior índice Sharpe	65
4.3	Comparativos com Agrupamento Não Hierárquico em Múltiplos	
	Períodos	67
4.3.1	Observação 1: período de três meses	67
4.3.1.1	Comparativo menor volatilidade	69
4.3.1.2	Comparativo maior retorno	70
4.3.1.3	Comparativo maior índice Sharpe	71
4.3.2	Observação 2: período de um ano	72
4.3.2.1	Comparativo menor volatilidade	74
4.3.2.2	Comparativo maior retorno	75
4.3.2.3	Comparativo maior índice Sharpe	76
4.3.3	Observação 3: período de três anos	77
4.3.3.1	Comparativo menor volatilidade	79
4.3.3.2	Comparativo maior retorno	80
4.3.3.3	Comparativo maior índice Sharpe	81
4.3.4	Observação 4: período de cinco anos	82
4.3.4.1	Comparativo menor volatilidade	84
4.3.4.2	Comparativo maior retorno	85
4.3.4.3	Comparativo maior índice Sharpe	86
5	CONCLUSÃO	89
	REFERÊNCIAS	90

1 Introdução

Com a evolução da tecnologia computacional, avanços e facilidades no que se refere à análise de dados, têm sido proporcionados. Assim é possível perceber a necessidade de técnicas de análise e interpretação adequada para obter informações, principalmente para tomada de decisões ótimas, de modo que possibilitem um gerenciamento eficiente.

A abordagem isolada, pode não ser suficiente para compreensão e solução eficiente de um problema. Somado a isso, com a rápida evolução tecnológica, criaram-se oportunidades para abordagens que englobam mais de uma matéria. No cenário do mundo moderno, a interdisciplinaridade é essencial para o desenvolvimento de soluções inovadoras. É um campo de estudo que busca fazer com que diferentes disciplinas interajam entre si, abordando problemas complexos, como é o caso do presente trabalho [1].

A união entre a Economia, Física e Ciências da Computação, faz com que a econofísica seja uma matéria interdisciplinar, trazendo consigo conceitos de grande importância para a economia, salientando formas de estudo por interações, como a formação de preços e a dinâmica de mercados financeiros. O surgimento da econofísica ocorreu quando pesquisadores da área de física começaram a introduzir métodos e conceitos físicos para estudar fenômenos da economia, proporcionando novas perspectivas para compreensão dos dados [2].

A Teoria de Markowitz foi revolucionária para a gestão de portfólios ótimos. Porém, sua aplicação prática pode ser um desafio, devido a complexidade dos dados. As dificuldades apresentadas no método tradicional com centenas de ativos foi o que motivou o desenvolvimento desse estudo, que busca otimizar a gestão de carteiras utilizando uma nova abordagem.

Dentro deste cenário, a teoria de carteira eficiente de Markowitz será aplicada à econofísica, em união com o métodos de agrupamento para encontrar os ativos que devem ser incluídos na carteira. Tal aplicação pode ajudar na análise de dados e a desenvolver estratégias de investimento mais eficientes.

Este estudo trata-se de uma proposta de um algoritmo para a seleção de carteiras. Embasado na Análise de agrupamento, e em busca da Moderna Teoria de Carteiras, este trabalho tem a finalidade de aplicar a técnica de estatística multivariada para encontrar automaticamente uma carteira ótima.

O objetivo deste trabalho é desenvolver um método sistemático utilizando modelos da econofísica, conceitos e técnicas que utilizam da física estatística possibilitando que o investidor investigue e analise dados financeiros e econômicos, e realize todo seu processo

de aplicação a partir da escolha de quais empresas devem compor seu portfólio, e por meio de toda análise, o investidor obtenha seu portfólio ótimo.

A análise de agrupamento é uma técnica da estatística multivariada que busca agrupar um conjunto de dados em subgrupos. Essa abordagem, focada na identificação de estruturas inerentes aos dados busca basicamente identificar padrões entre si, organizando os elementos semelhantes em um mesmo grupo, ficando os elementos dissemelhantes em grupos distintos, simplificando a estrutura dos dados [3].

A fim de aprofundar a análise de agrupamento foram realizadas tentativas de seleção de ativos usando parâmetros fundamentalistas. Esses parâmetros são critérios ou indicadores relacionados à saúde financeira de uma empresa, que fornecem informações sobre as ações listadas na bolsa de valores, podendo incluir lucros, receitas, dividendos, dentre outros. Entretanto, os agrupamentos formados não foram satisfatórios para análise, pois geraram um grande grupo contendo grande parte das ações. A partir disso, foi decidido levar em conta as cotações em diferentes períodos, considerando os princípios da análise gráfica, de que o preço de fechamento ajustado incorpora as expectativas do mercado.

Ao conectar a análise de agrupamento à Teoria de Markowitz, pode ser observado que a análise de agrupamento fornece uma base inicial para entender a estrutura dos dados. Ao identificar padrões e agrupar elementos semelhantes, ela cria clusters que representam grupos naturais dentro do conjunto de dados.

A teoria de Markowitz por sua vez, fará um complemento estratégico, abordando a alocação de ativos em uma carteira, em busca do seu resultado ótimo. Ela consiste em um modelo matemático, na qual o cálculo do risco leva em consideração a correlação entre seus ativos para encontrar o risco e retorno de cada um, buscando aprimorar a compreensão sobre como os diferentes grupos de ativos podem ser otimamente combinados para atender aos objetivos de retorno e gerenciamento de risco [4].

A análise de agrupamento pode ser vista como uma fase exploratória, identificando padrões e agrupamentos iniciais nos dados, enquanto a Teoria de Markowitz atua como uma etapa subsequente, refinando os agrupamentos gerados de forma mais precisa e orientada por métricas financeiras. Essa integração entre a análise exploratória e a otimização quantitativa pode resultar em estratégias de investimento sólido.

Esta dissertação foi estruturada da seguinte forma, no segundo capítulo será fundamentado alguns cálculos relevantes para compreensão da aplicação. Em seguida, no capítulo três, será apresentada uma aplicação prática do agrupamento hierárquico, explorando sua implementação. O quarto capítulo é dedicado à aplicação do agrupamento não hierárquico. Para finalizar, nas conclusões, são reunidos os principais pontos de descobertas do trabalho.

2 Referencial Teórico

2.1 Análise de Agrupamento

A análise agrupamento é uma técnica de estatística multivariada que tem como objetivo agrupar elementos segundo suas características, utilizando-se de medidas de similaridade e dissimilaridade para distinguí-las. A análise de agrupamento, também conhecida como análise de cluster, pode ser dividida em dois métodos, hierárquico e não hierárquico [3].

Existe uma ampla variedade de aplicações de agrupamento de dados que permitem estudos em diversas áreas, como na medicina, abrangendo trabalhos como diversidade genética, agrupamentos de redes neurais, bem como na área empresarial, onde realiza agrupamentos por similaridade entre clientes, para priorizar a qualidade de serviços, ou no agronegócio, onde pode ser feita análise de agrupamentos para estabelecer oportunidades de redução de custos em uma rede de distribuição [5], [6].

Somado a isso, diversos trabalhos de pesquisa mostram a eficiência da aplicação dos métodos de agrupamento, que variam de domínios. Algumas dessas metodologias podem ser vistas em aplicações que tem como foco encontrar estratégias de agrupamentos para organizar dados e construir dicionários de circuito eletrônico de memória (ROM - Read-Only Memory) [7], construção de agrupamentos sobre as interações e classificações de partículas de alta energia [8], bem como aplicação de clusterização como uma solução para otimização binária em um conjunto de dados grandes e complexos como é o agrupamento assistido por quantum [9], e, também, agrupamentos que buscam entender o comportamento sobre propriedades e estruturas das redes de ligações de hidrogênio [10]. É possível perceber que a utilização de métodos de agrupamento enriquecem a compreensão sobre as propriedades e comportamentos dos sistemas físicos.

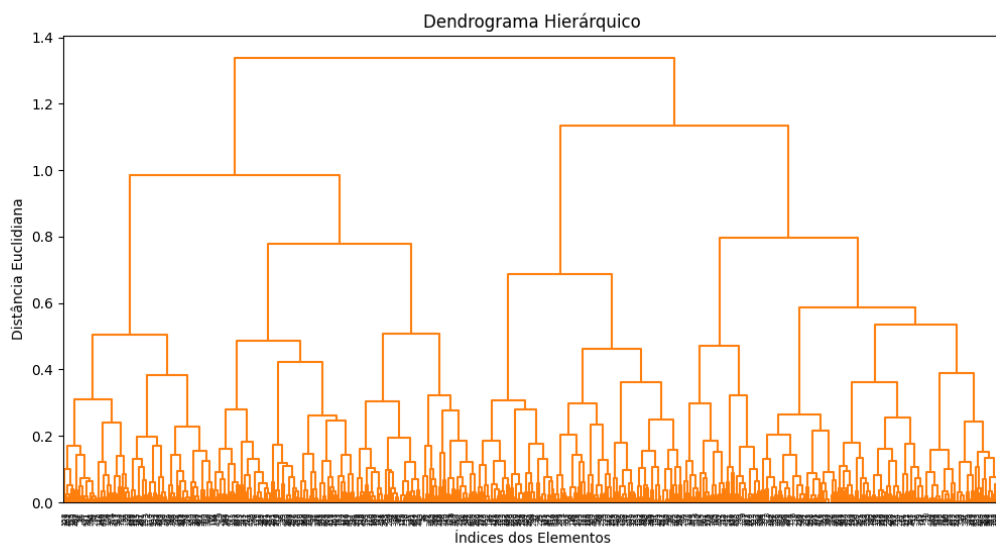
Em particular, alguns trabalhos dentro da área de otimização de portfólios podem ser citados, como León [11], que apresenta em seu artigo o desempenho de sete carteiras criadas utilizando técnicas de agrupamento com interpretações visuais para análise em um conjunto de dados levando em consideração o período de dois meses. Gubu e Rosadi, (2020) [12], em seu trabalho, mostram o desempenho do portfólio desenvolvido com a clusterização KAMILA, usado para agrupar dados com base em critérios de semelhança, unido a estimativa robusta FMCD, método que leva em consideração a presença de pontos atípicos para análise.

Os algoritmos hierárquicos diferem dos não-hierárquicos essencialmente em dois pontos. Nos algoritmos hierárquicos, uma vez que observações foram combinadas, elas não

se separarão. Já o não hierárquico permite que esses elementos sejam realocados de um grupo para outro. O segundo ponto em que se diferem é a escolha do número de grupos, onde para o primeiro algoritmo, uma forma de se estimar o número de grupos da partição é analisar um gráfico chamado dendrograma, já para o segundo, esse número deve ser previamente estabelecido [3].

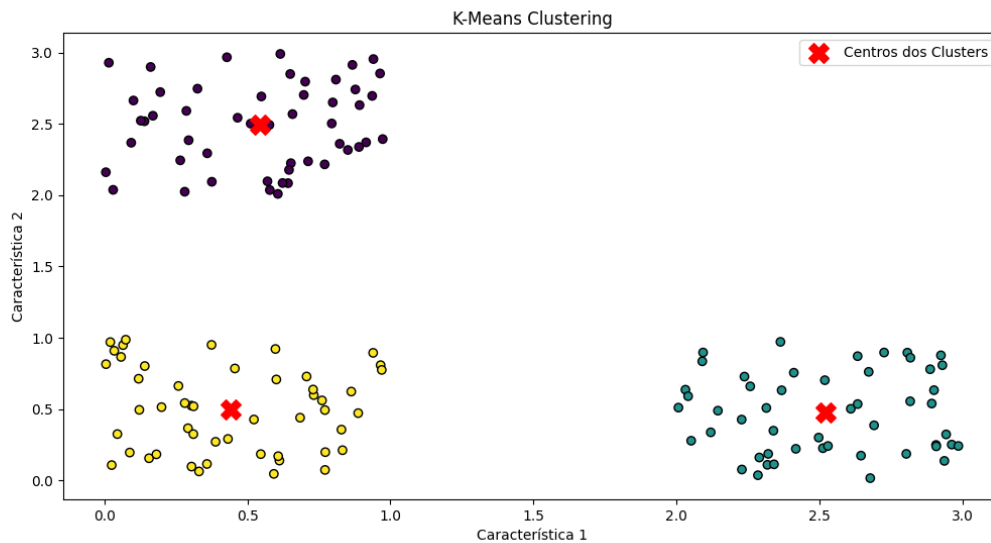
Uma característica importante de avaliar para a tomada de decisão entre aplicar o método hierárquico e não hierárquico é a quantidade de elementos que se têm. Um problema com muitos elementos pode acabar implicando negativamente para análise com o método hierárquico, pois muitos objetos podem ser difíceis de serem visualizados em um dendrograma, dificultando para se estabelecer um corte, conseqüentemente sendo prejudicial para uma análise concreta dos dados. Como o exemplo abaixo, apenas para demonstrar graficamente, a dificuldade de ver todos os elementos.

Figura 1 – Dendrograma



Fonte: Autor (2023)

Já com o agrupamento não hierárquico, também conhecido como *k-means* ou *k-médias*, possuir muitos elementos não é problema, visto que a quantidade de grupo k é escolhida anteriormente, e sua forma gráfica é um gráfico de dispersão. Porém, é necessário ter conhecimento acerca dos dados para poder determinar um número de grupos ideal para o agrupamento. Na figura 2, é mostrado um exemplo de agrupamento para $k = 3$.

Figura 2 – Exemplo de diagrama de dispersão com $k=3$ 

Fonte: Autor (2023)

A vantagem de se utilizar a análise de agrupamento é poder descrever de forma eficiente as características de cada grupo identificado, e somado a isso, descobrir possíveis correlações entre os dados.

Sendo assim, como os elementos e a forma que se comportam são desconhecidos, inicialmente foi escolhido o método hierárquico para aplicação, em busca de compreender e explorar a estrutura dos dados de maneira mais detalhada, identificando padrões, sem a necessidade de pré determinar um número de grupos. Em seguida, no capítulo quatro será usado o método k-médias com o número de grupos definido pelos resultados do capítulo três.

Existem duas maneiras de executar a clusterização hierárquica: a divisiva e a aglomerativa. O algoritmo divisivo compara cada elemento contido inicialmente considerando um único grupo e, em seguida, é feita rodadas de interações entre si para que sejam agrupados de acordo com suas características. Já o aglomerativo é o inverso, na qual cada elemento é considerado um único grupo inicialmente [13].

As etapas do processo de agrupamento consistem em determinar a métrica que será utilizada, e qual técnica será adotada para formação dos grupos.

Para agrupar esses elementos, medidas de similaridade e dissimilaridade devem ser tomadas. Alguns elementos não podem ser mensurados em variáveis quantitativas, quando isso acontece, é usada a medida de similaridade, onde as variáveis são transformadas em binárias [13].

Já para as medidas de dissimilaridade, vários coeficientes podem ser utilizados, a

escolha pode depender do estudo e do seu objetivo. Algumas dessas distâncias são definidas, como a euclidiana, de Manhattan e a de Minkowski.

2.1.1 Distância euclidiana

A distância Euclidiana é comumente utilizada na literatura, e pode ser provada através do Teorema de Pitágoras, que calcula a diferença entre dois pontos [14]. Conforme a equação 2.1:

$$d(X_1, X_2) = \sqrt{(X_{11} - X_{21})^2 + (X_{12} - X_{22})^2 + \dots + (X_{1n} - X_{2n})^2} \quad (2.1)$$

- $d(X_1, X_2)$ representa a distância entre dois pontos.
- $X_{11}, X_{22}, \dots, X_{1n}$ são os componentes do primeiro ponto X_1 ;
- $(X_{11}, X_{22}, \dots, X_{2n})$ são os componentes do segundo ponto X_2 ;
- n é o número de dimensões no espaço euclidiano.

2.1.2 Distância de Manhattan

A distância de Manhattan, conhecida também como distância da cidade, é uma medida de distância entre dois pontos em um espaço euclidiano bidimensional ou tridimensional [14]. A distância é apresentada a seguir:

$$d(X_1, X_2) = \sum_{i=1}^n |x_{1i} - x_{2i}| \quad (2.2)$$

- $d(X_1, X_2)$ é a distância de Manhattan.
- X_{1i} é o componente i -ésimo do ponto X_1 ;
- X_{2i} é o componente i -ésimo do ponto X_2 ;
- n assim como na distância anterior, representa o número de dimensões que definem os pontos $(X_1 e X_2)$.

2.1.3 Distância de Minkowski

Pode ser observado que a distância de Minkowski é a generalização da distância Euclidiana e de Manhattan. A fórmula pode ser definida com um parâmetro "m" que pode ser ajustado para se adequar a diferentes situações [14].

$$d(X_1, X_2) = \left(\sum_{i=1}^n |x_{1i} - x_{2i}|^m \right)^{\frac{1}{m}} \quad (2.3)$$

- Se "m" = 2, a fórmula representa a distância Euclidiana.
- Se "m" = 1, a fórmula representa a distância de Manhattan.

2.1.4 Técnicas para a construção de *clusters*

Definida a distância a ser utilizada, armazenam-se as distâncias amostrais em uma matriz n x n, definida como:

$$D = \begin{bmatrix} 0 & d_{12} & d_{13} & \dots & d_{1n} \\ & 0 & d_{23} & \dots & d_{2n} \\ & & 0 & \dots & d_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ & & & \dots & 0 \end{bmatrix} \quad (2.4)$$

Onde d_{ij} representa o elemento na linha i e na coluna j da matriz D.

Formada a matriz, é necessário selecionar o método que será utilizado para a formação dos grupos. Alguns dos mais utilizados dentro do agrupamento hierárquico são o de ligação simples, completa e média.

2.1.4.1 Ligação simples

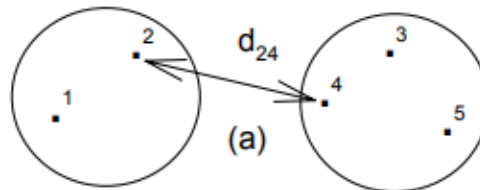
Também conhecido como método do vizinho mais próximo, este tipo de ligação utiliza a menor distância entre os elementos do grupo.

$$d_{ij} = \min(d(X_i, X_k), d(X_j, X_k)) \quad (2.5)$$

A mínima distância d_{ij} se dá usando uma medida de distância escolhida previamente. Em cada passo, os dois grupos com a menor distância são fundidos. Esse processo é repetido até restar somente um grupo.

A figura 3 ilustra como se dá a ligação pelo método de ligação simples, na qual a distância é representada em d_{24} como a menor entre o par de pontos pertencente aos grupos.

Figura 3 – Método de ligação simples. Figura retirada da dissertação Análise Multivariada



Fonte: Ferreira (1996, p. 298) [15].

2.1.4.2 Ligação completa

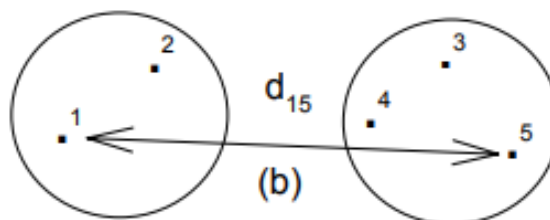
Conhecido como método dos vizinhos mais distantes, esse tipo de ligação difere da anterior somente na utilização da maior distância entre grupos.

$$d(A, B) = \max_{i \in A, j \in B} d(x_i, x_j) \quad (2.6)$$

- $d(X_A, X_B)$ é a distância completa;
- $d(X_i, X_j)$ é a distância entre o ponto pertencente ao grupo A e ao grupo B.

Para representar o método de ligação completa, os elementos mais distantes devem se agrupar, conforme a figura 4, em que d_{15} representa a maior distância.

Figura 4 – Método de ligação completa. Figura retirada da dissertação Análise Multivariada



Fonte: Ferreira (1996, p. 298) [15].

2.1.4.3 Ligação média

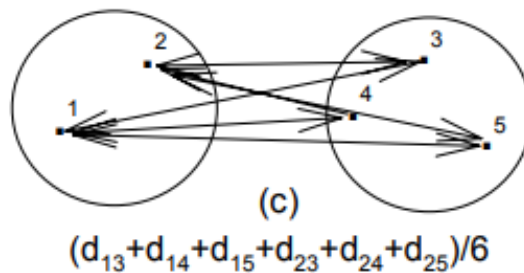
Neste método, a distância é dada pelo cálculo da média das distâncias.

$$d_{(A,B)} = \frac{1}{|A| \cdot |B|} \sum_{i \in A} \sum_{j \in B} d(x_i, x_j) \quad (2.7)$$

- $d(X_A, X_B)$ representa a distância média;
- $d(X_i, X_j)$ é a distância entre o ponto pertencente ao grupo A e ao grupo B;
- $|A|$ e $|B|$ é o número de pontos nos grupos A e B.

O método de ligação média pode ser mostrado abaixo:

Figura 5 – Método de ligação média. Figura retirada da dissertação Análise Multivariada



Fonte: Ferreira (1996, p. 298) [15].

Depois de gerar os dados necessários para a construção do dendrograma, é possível analisá-lo em busca de identificar a distribuição dos dados e obter o número ideal de agrupamentos para o modelo.

O método de ligação a ser escolhido para análise, influencia significativamente na estrutura do dendrograma e interpretação dos grupos. Sua escolha provém da natureza dos dados e o que se deseja obter com o agrupamento [3].

A validação dos resultados obtidos, geralmente é feita de forma subjetiva, analisando critérios de otimização [3].

Após explorar alguns métodos hierárquicos, será apresentado o método k-médias, técnica de agrupamento não hierárquico amplamente utilizada para análise de dados.

2.1.4.4 Método K-Médias

Dentre os métodos não hierárquicos, um dos mais usados, é o k-médias, que foi proposto por MacQueen 1967, método que busca minimizar a distância dos objetos utilizando k centros de modo iterativo [16].

Este é um método de partição que caracteriza-se por assumir um k fixo e um centro que tem como valor sua média. Um ponto a ser abordado é o conceito de similaridade, quanto menor a distância entre dois pontos, mais similares são [17].

A aplicação do k-médias se dá de uma forma simples. Inicialmente é preciso determinar o número k de grupos. Em seguida são selecionados pontos do conjunto para serem os centróides iniciais. Esses centróides são definidos pelo usuário, podendo ser escolhido de forma aleatória ou com base em algum critério [15].

Feito isso, cada ponto do conjunto será atribuído ao centróide mais próximo. Por fim, o centróide de cada grupo é atualizado até que permaneça o mesmo depois de duas iterações consecutivas [15].

Para dar início aos cálculos, é determinado o número de grupos. O algoritmo pode utilizar diferentes distâncias como centroide. Distâncias distintas podem gerar diferentes agrupamentos. A distância mais utilizada é a distância euclidiana, apresentada anteriormente através do Teorema de Pitágoras e utilizada a seguir, na equação 2.8.

$$SQE(k) = \sum_{i=1}^n \sum_{j=1}^p (x_{ij} - y_{ij})^2 \quad (2.8)$$

onde n é o número de objetos, p é o número de variáveis e k é o número de grupos [18].

Um dos problemas associado ao método é que nem todos os grupos sempre apresentam valores satisfatórios. Dessa forma, o método pode ser aplicado várias vezes com valores diferentes para k, até que os resultados sejam melhor representados [19].

2.2 Teoria de Markowitz

A Pesquisa Operacional (PO) proporcionou grandes avanços e desenvolvimento em várias áreas do conhecimento. Seu principal objetivo é auxiliar na tomada de decisões. Sendo assim, a PO trata de problemas de otimização que envolvem modelos matemáticos, aplicando métodos computacionais para encontrar suas soluções ótimas [4].

A Teoria de Markowitz, desenvolvida pelo economista Harry Markowitz em 1952, representa um marco na aplicação da matemática e da estatística na área de finanças.

Markowitz propôs uma abordagem revolucionária para a gestão de investimentos, que se baseia na combinação de ativos financeiros de maneira a otimizar o retorno esperado e minimizar o risco de perdas. Esta utiliza-se da modelagem matemática e de alguns parâmetros estatísticos para a formulação, análise e seleção de carteiras de investimentos [4].

Na formulação de uma carteira de investimentos, comumente, o investidor busca obter o maior retorno possível com o menor risco de perda, ou seja, a teoria propõe a criação de uma carteira, maximizando o retorno esperado ou minimizando o risco do portfólio. A Teoria Moderna de Carteiras pode ser considerada uma ferramenta que permite ao investidor compor sua carteira de investimento com diferentes ativos financeiros, otimizando a relação entre retorno e risco do investimento [20].

Markowitz propõe que as carteiras sejam selecionadas com base no desempenho agregado dos seus ativos. A diversificação da carteira pode levar a composições mais eficientes, em razão da correlação entre seus ativos. Para isso, é necessário ficar atento ao mercado em que esses ativos estão inseridos também [20].

Os trabalhos de Markowitz inspiraram estudos subsequentes na área financeira, envolvendo seleção de carteiras. Muitos pesquisadores utilizaram e expandiram as ideias de Markowitz, como Sharpe (1963) que desenvolveu "*A simplified model for portfolio analysis*", modelo que propõe a utilização do índice Sharpe como uma ferramenta estratégica para avaliar o desempenho da carteira [21]. Roll e Ross (1980) que publicaram o trabalho "*An Empirical Investigation of the Arbitrage Pricing Theory*", que busca analisar a relação entre os retornos esperados e os fatores de riscos implícitos [22].

A Teoria de Markowitz continua sendo uma ferramenta valiosa na gestão de carteiras e na análise de risco e retorno em investimentos, influenciando de forma significativa como os investidores e profissionais financeiros abordam a alocação de ativos e a busca por estratégias de investimento eficazes.

Na busca pelo equilíbrio entre retorno e risco em uma carteira de investimentos, a teoria moderna de portfólios utiliza métodos de otimização para encontrar soluções que maximizem o retorno esperado, minimizando, ao mesmo tempo, a volatilidade associada [20].

O retorno esperado de uma carteira é o quanto se espera obter de retorno com o investimento feito. A taxa de retorno esperada da carteira pode ser obtida por meio do somatório da multiplicação dos valores esperados das taxas de retorno de cada componente pela sua respectiva participação na composição da carteira. O retorno é calculado através da média ponderada dos retornos dos ativos que a compõe. A equação 2.8 mostra como calcular o retorno:

$$R_p = \sum_{i=1}^n w_i \cdot R_i \quad (2.9)$$

Onde:

- R_p é o retorno esperado da carteira.
- w_i representa o peso do ativo i na carteira.
- R_i é o retorno esperado do ativo i .
- n é o número de ativos na carteira.

Já o risco de uma carteira pode ser definido como a probabilidade de ocorrer determinados resultados em relação a um valor médio esperado. É um conceito que mostra possibilidades de perda. Sendo assim, o risco não pode ser totalmente eliminado, mas é desejável que seja minimizado. O risco pode ser dado por meio do cálculo do desvio padrão, que representa a divergência dos resultados em função do retorno esperado. Também está relacionado diretamente à volatilidade do seu rendimento ao passar do tempo. O risco é medido pela sua variância, conforme equação a seguir:

$$\sigma_p^2 = \sum_{i=1}^n \sum_{j=1}^n w_i \cdot w_j \cdot \text{Cov}(R_i, R_j) \quad (2.10)$$

Nesta fórmula:

- σ_p^2 representa a variância da carteira.
- n é o número de ativos na carteira.
- w_i e w_j são os pesos dos ativos i e j na carteira, respectivamente.
- $\text{Cov}(R_i, R_j)$ é a covariância entre os retornos dos ativos i e j .

Portanto o risco pode ser definido como desvio padrão das variações do retorno de um ativo.

2.2.1 Covariância e Correlação

A covariância e a correlação são elementos cruciais para a avaliação de como os ativos se comportam entre si. A covariância mede a tendência dos retornos dos ativos a variarem juntos, enquanto a correlação normaliza essa medida, fornecendo uma interpretação padronizada da relação.

Ao adicionar o cálculo desses parâmetros estatísticos, a Teoria de Markowitz ganha maior precisão. A covariância entre ativos influencia diretamente a variância da carteira, um importante componente no cálculo do risco. Quando dois ativos têm uma covariância positiva, seus retornos tendem a se mover na mesma direção, aumentando o risco total da carteira. Por outro lado, uma covariância negativa indica uma relação inversa, proporcionando uma oportunidade de diversificação eficiente [23].

A correlação, ao normalizar a covariância, oferece uma medida padronizada que varia de -1 a 1. Uma correlação próxima a 1 sugere uma forte relação positiva, enquanto uma correlação próxima a -1 indica uma forte relação negativa. Isso é de extrema importância na identificação de ativos que não apenas oferecem retornos atraentes, mas que também se movem de maneira complementar, permitindo uma diversificação mais eficiente [23].

Assim, a fórmula para calcular a covariância entre dois ativos é:

$$\text{Cov}(X, Y) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \quad (2.11)$$

Onde:

- $\text{Cov}(X, Y)$ é a covariância entre as variáveis X e Y.
- X_i e Y_i são os valores das observações.
- \bar{X} e \bar{Y} são as médias X e Y.
- n representa o número de observações

A correlação pode ser dada:

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y} \quad (2.12)$$

- $\rho(X, Y)$ representa o coeficiente de correlação.
- σ_X e σ_Y são os respectivos desvios padrão de X e Y.

Portanto, a precisão encontrada através do cálculo da covariância e da correlação não apenas aprimora a construção da carteira, mas também destaca a importância da compreensão das relações interativas entre ativos.

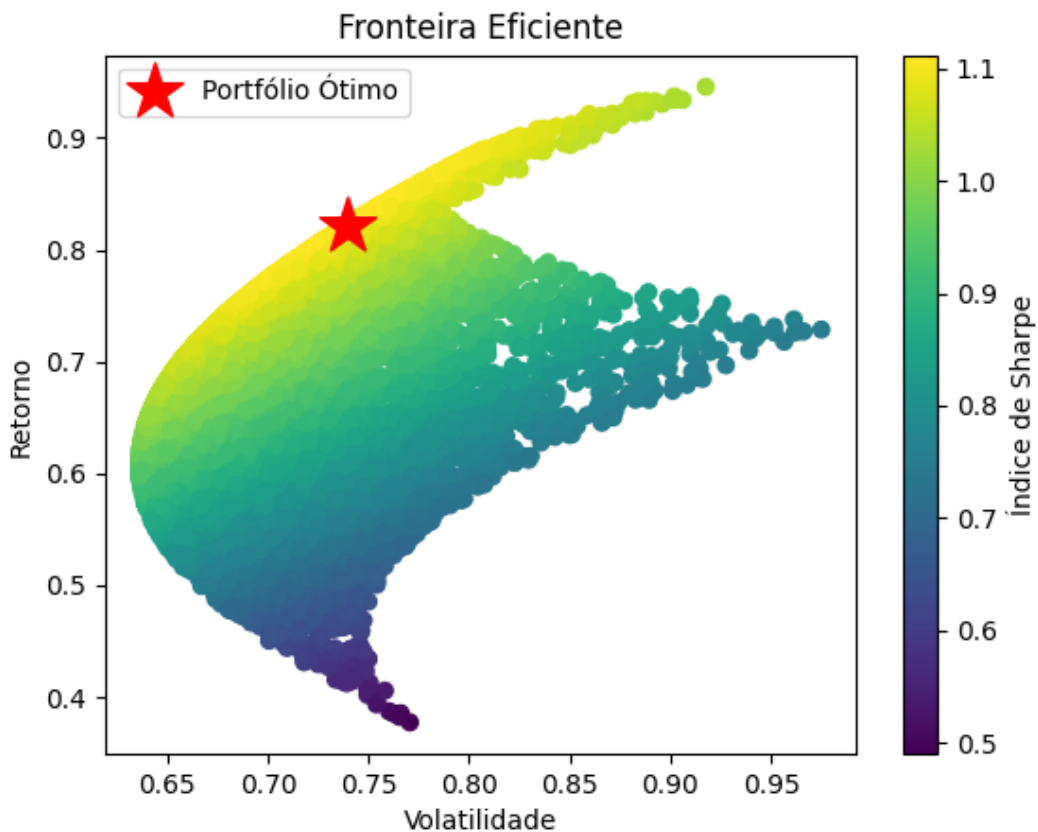
2.2.2 Fronteira Eficiente

A fim de maximizar a precisão, a Fronteira Eficiente representa a culminação da otimização de carteiras, incorporando o equilíbrio entre retorno e risco, e, também, a diversificação inteligente e estratégica [24].

A Fronteira Eficiente representa de uma forma gráfica todas as possíveis combinações de ativos contidos na carteira. A precisão estatística fornecida pela avaliação da covariância e correlação entre os papéis desempenha uma tarefa importante para a determinação da posição relativa de uma carteira nessa fronteira [24].

A figura 6 mostra um exemplo gráfico sobre o comportamento de uma fronteira eficiente e a localização do portfólio ótimo.

Figura 6 – Exemplo fronteira eficiente



Fonte: Autor (2023)

A visualização da fronteira eficiente permite uma compreensão clara entre risco e retorno. A partir dela, é possível que os investidores tomem decisões mais completas sobre como alocar seus recursos, o que é crucial para otimizar o desempenho da carteira.

2.2.3 Índice Sharpe

Ampliando a avaliação de desempenho na Teoria de Markowitz, Willian Sharpe desenvolveu uma métrica chamada Índice Sharpe, que desempenha um papel crucial na análise de portfólios. Enquanto a teoria moderna de portfólio foca na maximização do retorno para um determinado nível de risco, o Índice Sharpe avalia o retorno em relação ao risco total [21].

Obter uma carteira com um Índice Sharpe elevado indica um melhor retorno ajustado ao risco. Portfólios com o Índice Sharpe mais altos são considerados mais eficientes, pois permite uma avaliação mais completa sobre o desempenho da carteira em comparação com a alternativa de investimento livre de risco. Sendo assim, o índice Sharpe é calculado da seguinte forma:

$$\text{Sharpe Ratio} = \frac{R_p - R_f}{\sigma_p} \quad (2.13)$$

Onde:

- R_p é o retorno médio da carteira ou investimento.
- R_f é a taxa livre de risco.
- σ_p é o desvio padrão (volatilidade) do retorno da carteira ou investimento

3 Agrupamento hierárquico

Será apresentada uma abordagem inovadora na construção de carteiras de Markowitz, adicionando as vantagens do agrupamento hierárquico para selecionar os ativos e reduzir os dados, implementadas através da linguagem de programação Python. Serão investigadas a eficiência dessas abordagens através da realização de quatro seleções diferentes, permitindo uma análise abrangente do comportamento da carteira.

Para construir um portfólio, uma etapa importante é a obtenção de dados financeiros confiáveis, que serão obtidos através de uma implementação prática. O Yahoo Finance, é uma biblioteca que fornece acesso a um amplo conjunto de dados de mercado, como preços de ações, preço de fechamento, histórico de preços, tornando-o uma escolha ideal para análise.

Por meio do Python, serão demonstrados a seleção, o processamento e a análise de dados de ações da bolsa de valores.

3.1 Seleção de Ativos por Agrupamento Hierárquico

A metodologia escolhida foi o agrupamento hierárquico, o conjunto de elementos selecionados foi um número limitado de ações negociada na B3 (Bolsa de Valores) em busca de simplificar a análise, avaliar o comportamento dos agrupamentos, além de testar a viabilidade da abordagem. Esse estudo inicial, apesar de limitado, fornece uma base para futuros aprimoramentos na aplicação de agrupamentos no campo financeiro.

Inicialmente foi utilizado um determinado grupo de ações, contendo 78 ativos selecionados do índice Ibovespa, em quatro períodos distintos para construção da carteira. O Ibovespa é o principal índice da bolsa de valores brasileira, é um indicador que avalia o desempenho médio das cotações das ações negociadas na B3.

A distância euclidiana foi a escolhida para a construção da matriz, que é uma medida eficaz ao lidar com grupos de tamanhos diferentes. A escolha da medida de distância pode influenciar em como os elementos interagem de acordo com suas características financeiras [25]. Somado a isso, o preço de fechamento ajustado foi a métrica escolhida para análise dos dados, pois ele representa o ajuste do valor levando em consideração eventos corporativos, como dividendos e bonificações. A partir disso, os cálculos para gerar o dendrograma foram realizados entre o desvio-padrão e a média dessa métrica dos ativos selecionados, buscando uma visão geral da distribuição dos preços.

É necessário padronizar os dados para que eles tenham uma comparação justa,

devido a diferença de seus tamanhos. Para isso, foi utilizada a transformação dos dados, subtraindo sua média e dividindo pelo desvio padrão cada valor. Isto resulta em uma distribuição com média zero e desvio padrão igual a um [26].

Em seguida, para realizar o agrupamento hierárquico, foi definido que o método utilizado será o de ligação simples, pois ele pode levar a agrupamentos mais compactos [3].

Posteriormente, perante análise dos grupos gerados pelos seus dendrogramas, o grupo inicial foi reduzido a grupos diferentes para cada período. Isso acontece porque à medida que as condições de mercado se alteram ao passar do tempo, sua estrutura pode se modificar, mostrando que o agrupamento hierárquico possui uma grande adaptabilidade, que permite a construção de carteiras que se ajustam às condições do mercado.

3.2 Otimização da Carteira de Markowitz

Uma vez que os ativos foram agrupados, é preciso entender também, como funciona o método de Monte Carlo, que é uma técnica estatística e computacional usada para solucionar problemas complexos por meio de simulações de números aleatórios, e que permite avaliar o comportamento de um processo [27].

A cada iteração realizada para gerar os pesos aleatórios, seu resultado é armazenado, e ao final das repetições é possível criar uma distribuição para realizar os cálculos estatísticos para otimização. Toda iteração representa uma combinação única dentro do portfólio [27]. Esse processo iterativo, busca analisar a diversificação e desempenho em diversos cenários, garantindo uma vasta análise.

Esta têm sido uma alternativa muito utilizada para estudos de simulação de dados, que tem como objetivo gerar números pseudo-aleatórios segundo uma distribuição de probabilidade para resolver problemas. Este método possui a vantagem de reduzir análises tendenciosas, além de que, o uso da simulação permite definir a natureza das distribuições de probabilidade, o que seria difícil de ser determinado por procedimentos estatísticos convencionais [28].

Vale ressaltar a importância de uma restrição fundamental durante a realização do método, que o somatório dos pesos de cada ativo presente na carteira seja igual a um, assegurando a diversificação e equilíbrio da carteira. Essa técnica é aplicada através da normalização dos dados, que ajusta seus valores e assegura que a alocação seja proporcional e adequada [29].

O processo de geração de carteiras é influenciado pelo dendrograma, visto que a relação entre clusters hierárquicos e a geração de pesos aleatórios é estabelecida a partir da estrutura formada, concluindo que os elementos dentro de um mesmo conjunto possuam características que reflitam a similaridade dentro de um grupo.

Para dar continuidade na construção do código, serão criadas listas vazias que serão preenchidas com os resultados calculados para cada seleção. Nestas listas estarão informações sobre várias carteiras de investimento, incluindo seus retornos, retornos anuais, matrizes de covariância e matrizes de correlação. Isso é de grande utilidade para análises e gestão de ativos.

Em seguida, com os cálculos citados anteriormente realizados, é possível realizar simulações para criar carteiras de investimento aleatórias e aplicar métricas essenciais para a interpretação do portfólio, como calcular a carteira de menor volatilidade, a de maior retorno e a de maior *Sharpe Ratio*.

O menor risco do portfólio é considerado a menor volatilidade da carteira, com o cálculo sendo feito por meio da medida de dispersão dos retornos. Os retornos das carteiras serão avaliados para determinar o maior ganho. Também será calculado para todas as carteiras, o índice Sharpe, que faz relação entre o risco e o retorno, que é fundamental para compreensão do equilíbrio retorno e risco.

3.3 Resultados

A teoria eficiente da carteira de Markowitz se mostra uma ferramenta valiosa para a construção de portfólios e sua implementação prática, e para melhor demonstrar os resultados ótimos obtidos neste trabalho, além da análise de múltiplos períodos, foi feito também, um comparativo entre carteiras geradas de forma aleatória e as geradas a partir do agrupamento hierárquico.

A lista de ações proposta inicialmente foi selecionada de forma aleatória, pegando uma pequena parcela de ativos da bolsa de valores, que pode ser vista na figura seguinte:

Figura 7 – Lista com os 78 ativos selecionados inicialmente

```
acoes = ['ABEV3.SA', 'AGRO3.SA', 'ALSO3.SA', 'AMER3.SA', 'AZUL4.SA', 'B3SA3.SA',  
'BBAS3.SA', 'BBDC4.SA', 'BBSE3.SA', 'BEEF3.SA', 'BPAC11.SA', 'BRFS3.SA',  
'BRKM5.SA', 'CCRO3.SA', 'CIEL3.SA', 'CMIG4.SA', 'COGN3.SA', 'CPFE3.SA',  
'CRFB3.SA', 'CSAN3.SA', 'CSMG3.SA', 'CSNA3.SA', 'CVCB3.SA', 'CYRE3.SA',  
'ECOR3.SA', 'EGIE3.SA', 'ELET6.SA', 'EMBR3.SA', 'ENBR3.SA', 'EQL3.SA',  
'EVEN3.SA', 'EZTC3.SA', 'FLRY3.SA', 'GGBR4.SA', 'GOLL4.SA', 'GRND3.SA',  
'HAPV3.SA', 'HYPE3.SA', 'IRBR3.SA', 'ITSA4.SA', 'JBSS3.SA', 'JHSF3.SA',  
'JSLG3.SA', 'KLBN11.SA', 'LEVE3.SA', 'LREN3.SA', 'MDIA3.SA', 'MGLU3.SA',  
'MRFG3.SA', 'MRVE3.SA', 'MULT3.SA', 'NTCO3.SA', 'PCAR3.SA', 'PETR4.SA',  
'PRIO3.SA', 'POM04.SA', 'PSSA3.SA', 'PTBL3.SA', 'QUAL3.SA', 'RADL3.SA',  
'RAPT4.SA', 'RAIL3.SA', 'RENT3.SA', 'SANB11.SA', 'SBSP3.SA', 'SLCE3.SA',  
'SHUL4.SA', 'SUZB3.SA', 'TAE11.SA', 'TIMS3.SA', 'TOTS3.SA', 'TRPL4.SA',  
'UGPA3.SA', 'USIM5.SA', 'VALE3.SA', 'VIVT4.SA', 'YDUQ3.SA', 'WEGE3.SA']
```

Fonte: Autor (2023)

A partir dela, as quatro listas para cada período foram geradas. Essas listas foram reduzidas, devido à análise de agrupamento, que mostrou que seus elementos possuíam características distintas. Para os métodos utilizados, vários dos grupos gerados continham somente um ativo. Com isso, foi decidido eliminar os elementos com características discrepantes, a fim de refinar a seleção e concentrar nas combinações mais homogêneas e relevantes para o processo de otimização do portfólio.

Para análise dos períodos seguintes, será utilizado o desvio-padrão e a média dos ativos, conforme citado anteriormente, para construção dos dendrogramas. Após definir os grupos, o desenvolvimento segue com a realização das simulações que podem gerar inúmeras possibilidades de carteiras, mas que para o presente trabalho, serão feitas 10000, pois é um número relativamente grande de iterações que permite explorar uma ampla variedade de combinações de ativos.

3.3.1 Observação 1: período de três meses

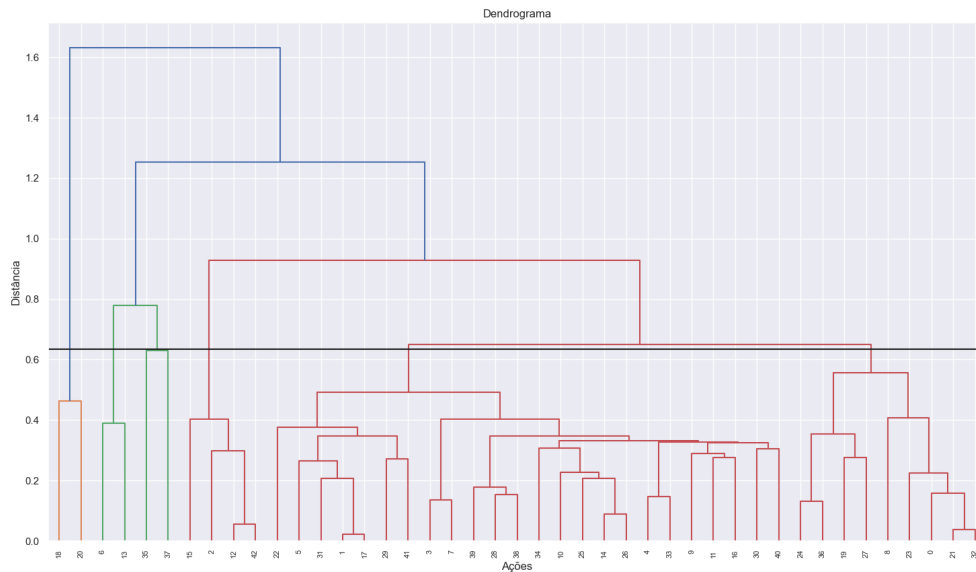
Para a análise em três meses, o grupo selecionado contém 43 ativos e foi realizada da janela de tempo do dia 01/03/2023 a 31/05/2023.

Figura 8 – Seleção de ativos para três meses

```
start = dt.date (2023, 3, 1)
end = dt.date (2023, 5, 31)
acoes = ['ABEV3.SA', 'AGRO3.SA', 'GGBR4.SA', 'JBSS3.SA', 'PCAR3.SA', 'B3SA3.SA',
'BRKM5.SA', 'ELET6.SA', 'SBSP3.SA', 'SUZB3.SA', 'CIEL3.SA', 'GRND3.SA',
'JSLG3.SA', 'SHUL4.SA', 'CMIG4.SA', 'ALSO3.SA', 'CYRE3.SA', 'EZTC3.SA',
'UGPA3.SA', 'CSAN3.SA', 'FLRY3.SA', 'HAPV3.SA', 'ITSA4.SA', 'JHSF3.SA',
'MGLU3.SA', 'PTBL3.SA', 'TAE11.SA', 'BBSE3.SA', 'CPFE3.SA', 'EQL3.SA',
'LREN3.SA', 'NTCO3.SA', 'PSSA3.SA', 'SANB11.SA', 'TOTS3.SA', 'TUPY3.SA',
'BBDC4.SA', 'CCRO3.SA', 'GOLL4.SA', 'KLB11.SA', 'MULT3.SA', 'RAIL3.SA',
'TRPL4.SA']
```

Fonte: Autor (2023)

Em seguida, o dendrograma gerado foi:

Figura 9 – Dendrograma - Período de três meses

Fonte: Autor (2023)

O corte no dendrograma foi realizado na altura de 0,635, de modo que o agrupamento foi dividido em seis grupos. É importante ressaltar que a altura do corte é baseada em critérios específicos, que visam criar grupos significativos e homogêneos.

Segue abaixo na tabela 1 quantos elementos e quais são eles para cada grupo gerado.

Tabela 1 – Elementos por grupo - três meses

Grupo	Ativos
Grupo 1	2
Grupo 2	2
Grupo 3	2
Grupo 4	4
Grupo 5	24
Grupo 6	9

Fonte: Autor (2023)

Os grupos contém os seguintes ativos:

1. Grupo 1: 'GOLL4.SA', 'HAPV3.SA'
2. Grupo 2: 'BRKM5.SA', 'ELET6.SA'
3. Grupo 3: 'SBSP3.SA', 'SUZB3.SA'

4. Grupo 4: 'ALSO3.SA', 'CYRE3.SA', 'EZTC3.SA', 'UGPA3.SA'
5. Grupo 5: 'AGRO3.SA', 'B3SA3.SA', 'BBDC4.SA', 'BBSE3.SA', 'CCRO3.SA', 'CMIG4.SA', 'CPFE3.SA', 'CSAN3.SA', 'EQTL3.SA', 'FLRY3.SA', 'GGBR4.SA', 'JBSS3.SA', 'KLBN11.SA', 'LREN3.SA', 'MULT3.SA', 'NTCO3.SA', 'PCAR3.SA', 'PSSA3.SA', 'RAIL3.SA', 'SANB11.SA', 'TAEE11.SA', 'TOTS3.SA', 'TRPL4.SA', 'TUPY3.SA'
6. Grupo 6: 'ABEV3.SA', 'CIEL3.SA', 'GRND3.SA', 'ITSA4.SA', 'JHSF3.SA', 'JSLG3.SA', 'MGLU3.SA', 'PTBL3.SA', 'SHUL4.SA'

Os grupos gerados revelam que seus ativos mostram comportamentos similares em relação ao seu histórico.

Com a quantidade de grupos definida, é possível dar início nas simulações para construção da carteira, iniciando pelo loop das possibilidades de carteiras que podem ser geradas a partir do agrupamento.

Em seguida são realizados os cálculos de retornos, matriz covariante e correlação para criar as listas necessárias, que são fundamentais para o processo de seleção de carteiras.

O objetivo ao conduzir simulações para criar carteiras de investimento, é encontrar aquelas que minimizam a volatilidade, maximizem o retorno ou obtenham um sharpe ratio elevado.

3.3.2 Observação 2: período de um ano

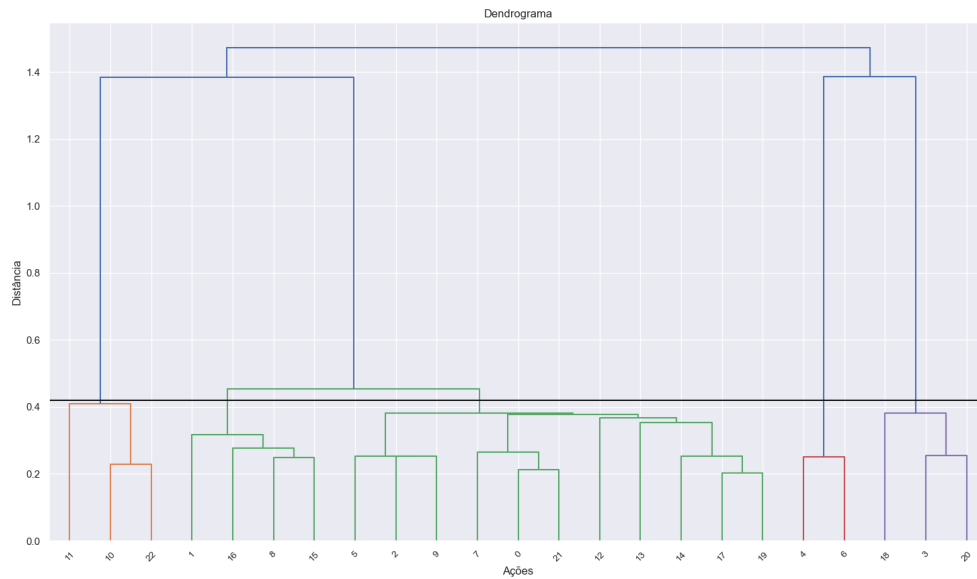
A janela de tempo selecionada foi 01/05/2022 a 31/05/2023, composta por 23 ativos.

Figura 10 – Seleção de ativos para um ano

```
start = dt.date (2022, 5, 1)
end = dt.date (2023, 5, 31)
acoes = ['ABEV3.SA', 'ECOR3.SA', 'ITSA4.SA', 'MGLU3.SA', 'RAPT4.SA',
          'TIMS3.SA', 'TRPL4.SA', 'ALSO3.SA', 'EQTL3.SA', 'MRVE3.SA',
          'MULT3.SA', 'GGBR4.SA', 'HYPE3.SA', 'YDUQ3.SA', 'B3SA3.SA',
          'CCRO3.SA', 'FLRY3.SA', 'KLBN11.SA', 'BBDC4.SA', 'TAEE11.SA',
          'TOTS3.SA', 'BEEF3.SA', 'CSAN3.SA']
```

Fonte: Autor (2023)

O dendrograma a seguir resultou em cinco grupos.

Figura 11 – Dendrograma - Período de um ano

Fonte: Autor (2023)

O ajuste da altura de corte é um ponto importante para o processo de agrupamentos, pois ele afeta diretamente a estrutura formada. Para essa seleção, o corte foi em 0,42.

Abaixo serão mostradas alguma informações sobre os grupos gerados.

Tabela 2 – Elementos por grupo - um ano

Grupo	Ativos
Grupo 1	3
Grupo 2	4
Grupo 3	11
Grupo 4	2
Grupo 5	3

Fonte: Autor (2023)

Os ativos presentes nos grupos são:

1. Grupo 1: 'GGBR4.SA', 'HYPE3.SA', 'YDUQ3.SA'
2. Grupo 2: 'ALSO3.SA', 'EQTL3.SA', 'MRVE3.SA', 'MULT3.SA'
3. Grupo 3: 'ABEV3.SA', 'B3SA3.SA', 'CCRO3.SA', 'ECOR3.SA', 'FLRY3.SA', 'ITSA4.SA', 'KLBN11.SA', 'MGLU3.SA', 'RAPT4.SA', 'TIMS3.SA', 'TRPL4.SA'

4. Grupo 4: 'BEEF3.SA', 'CSAN3.SA'
5. Grupo 5: 'BBDC4.SA', 'TAEE11.SA', 'TOTS3.SA'

Comparando essa seleção com a do período anterior, nota-se que os grupos não têm muita compatibilidade. Com exceção do grupo cinco do período de um ano, que está contido no grupo 5 do período de três meses, os elementos mostraram comportamentos diferentes para períodos diferentes, indicando variação histórica dos ativos.

3.3.3 Observação 3: período de três anos

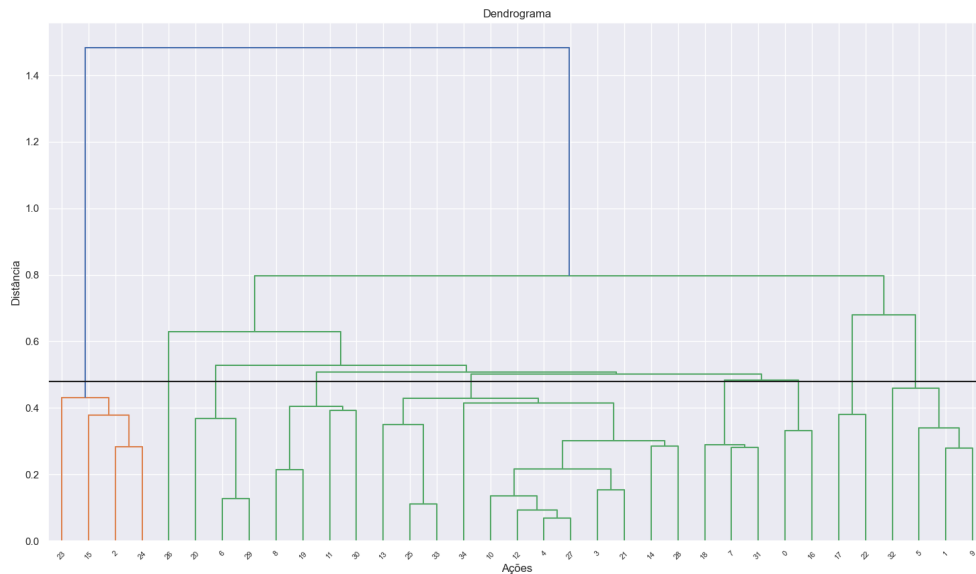
Para a análise de três anos, o grupo selecionado contém 35 ativos e foi realizada da janela de tempo do dia 01/05/2020 a 31/05/2023.

Figura 12 – Seleção de ativos para três anos.

```
start = dt.date (2020, 5, 1)
end = dt.date (2023, 5, 31)
acoes = ['ABEV3.SA', 'CCR03.SA', 'GRND3.SA', 'ITSA4.SA', 'TIMS3.SA', 'B3SA3.SA',
          'BBDC4.SA', 'CSAN3.SA', 'EGIE3.SA', 'EVEN3.SA', 'KLBN11.SA', 'PSSA3.SA',
          'RADL3.SA', 'UGPA3.SA', 'AGRO3.SA', 'BBSE3.SA', 'CPFE3.SA', 'HYPE3.SA',
          'MDIA3.SA', 'POMO4.SA', 'RAPT4.SA', 'TRPL4.SA', 'ALSO3.SA', 'FLRY3.SA',
          'MRFG3.SA', 'MRVE3.SA', 'TOTS3.SA', 'BEEF3.SA', 'JSLG3.SA', 'RAIL3.SA',
          'CIEL3.SA', 'CSMG3.SA', 'EQTL3.SA', 'JHSF3.SA', 'MULT3.SA']
```

Fonte: Autor (2023)

Onde o dendrograma foi gerado.

Figura 13 – Dendrograma - Período de três anos

Fonte: Autor (2023)

Pode ser observado que facilmente este agrupamento poderia ser dividido em dois grupos, notando a semelhança entre seus elementos demonstrados em verde. Porém, a altura escolhida para analisar, foi o corte em 0,48, dividindo em oito grupos para este período, com as seguintes quantidades:

Tabela 3 – Elementos por grupo - três anos

Grupo	Ativos
Grupo 1	4
Grupo 2	3
Grupo 3	3
Grupo 4	2
Grupo 5	4
Grupo 6	12
Grupo 7	1
Grupo 8	2

Fonte: Autor (2023)

Os grupos possuíam os elementos mostrados abaixo:

1. Grupo 1: 'ALSO3.SA', 'FLRY3.SA', 'MRFG3.SA', 'MRVE3.SA'
2. Grupo 2: 'BEEF3.SA', 'JSLG3.SA', 'RAIL3.SA'

3. Grupo 3: 'CCRO3.SA', 'ITSA4.SA', 'TIMS3.SA'
4. Grupo 4: 'ABEV3.SA', 'GRND3.SA'
5. Grupo 5: 'CIEL3.SA', 'CSMG3.SA', 'JHSF3.SA', 'RAPT4.SA', 'CSAN3.SA', 'EQTL3.SA', 'FLRY3.SA', 'GGBR4.SA', 'JBSS3.SA', 'KLBN11.SA', 'LREN3.SA', 'MULT3.SA', 'NTCO3.SA', 'PCAR3.SA', 'PSSA3.SA', 'RAIL3.SA', 'SANB11.SA', 'TAE11.SA', 'TOTS3.SA', 'TRPL4.SA', 'TUPY3.SA'
6. Grupo 6: 'B3SA3.SA', 'BBDC4.SA', 'CSAN3.SA', 'EGIE3.SA', 'EQTL3.SA', 'EVEN3.SA', 'KLBN11.SA', 'MULT3.SA', 'PSSA3.SA', 'RADL3.SA', 'TRPL4.SA', 'UGPA3.SA'
7. Grupo 7: 'POMO4.SA'
8. Grupo 8: 'HYPE3.SA', 'MDIA3.SA'

Pode ser visto que essa estrutura de agrupamentos se apresentou de forma diferente, onde um dos grupos possui apenas um elemento. Isso acontece porque um ativo pode apresentar um comportamento único ou diferente o suficiente para que ele não se encaixe em nenhum conjunto.

3.3.4 Observação 4: período de cinco anos

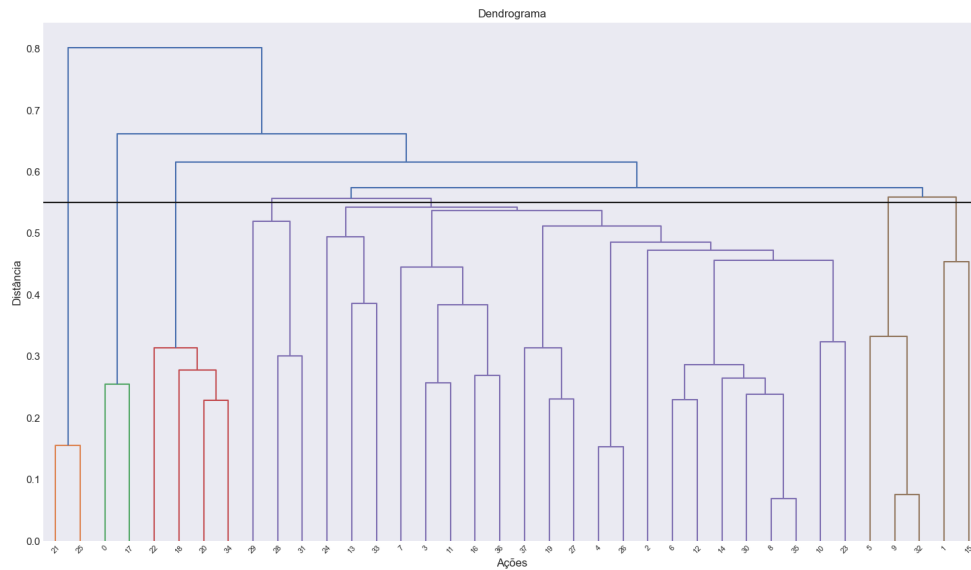
Para a investigação ao longo de cinco anos, o grupo selecionado engloba 38 ativos, e vai do período de 01/05/2018 a 31/05/2023.

Figura 14 – Seleção de ativos para cinco anos.

```
start = dt.date (2018, 5, 1)
end = dt.date (2023, 5, 31)
acoes = [ 'AGRO3.SA', 'CYRE3.SA', 'GGBR4.SA', 'JBSS3.SA', 'LEVE3.SA',
           'MGLU3.SA', 'B3SA3.SA', 'BEEF3.SA', 'CSMG3.SA', 'EVEN3.SA',
           'RAPT4.SA', 'BBAS3.SA', 'CPFE3.SA', 'CSAN3.SA', 'EGIE3.SA',
           'EQTL3.SA', 'RADL3.SA', 'TRPL4.SA', 'BBDC4.SA', 'CRFB3.SA',
           'ECOR3.SA', 'FLRY3.SA', 'UGPA3.SA', 'GRND3.SA', 'ITSA4.SA',
           'JSLG3.SA', 'TIMS3.SA', 'BBSE3.SA', 'MRFG3.SA', 'MULT3.SA',
           'PSSA3.SA', 'RAIL3.SA', 'EMBR3.SA', 'MDIA3.SA', 'SANB11.SA',
           'HAPV3.SA', 'MRVE3.SA', 'USIM5.SA' ]
```

Fonte: Autor (2023)

O dendrograma foi gerado:

Figura 15 – Dendrograma - Período de cinco anos

Fonte: Autor (2023)

A altura escolhida para o corte foi de 0,55, altura que seguiu o padrão das anteriores. A escolha para todos os pontos de cortes acima foi ajustada para obter um número próximo de grupos semelhantes para todas as janelas de tempo de análise. Essa escolha para o corte está relacionada à busca por elementos que sejam dissemelhantes o suficiente que apresente características significativas, mas para que, também, não crie grupos muito pequenos ou insignificantes.

Abaixo será mostrado a quantidades e ativos de cada grupo.

Tabela 4 – Elementos por grupo - cinco anos

Grupo	Ativos
Grupo 1	2
Grupo 2	2
Grupo 3	5
Grupo 4	3
Grupo 5	22
Grupo 6	3
Grupo 7	2

Fonte: Autor (2023)

Para cada grupo, os ativos:

1. Grupo 1: 'JBSS3.SA', 'MGLU3.SA'

2. Grupo 2: 'AGRO3.SA', 'GGBR4.SA'
3. Grupo 3: 'GRND3.SA', 'ITSA4.SA', 'JSLG3.SA', 'TIMS3.SA'
4. Grupo 4: 'MULT3.SA', 'PSSA3.SA', 'RAIL3.SA'
5. Grupo 5: 'BBAS3.SA', 'BBDC4.SA', 'BBSE3.SA', 'CPFE3.SA', 'CRFB3.SA', 'CSAN3.SA', 'CYRE3.SA', 'ECOR3.SA', 'EGIE3.SA', 'EMBR3.SA', 'EQTL3.SA', 'FLRY3.SA', 'HAPV3.SA', 'LEVE3.SA', 'MDIA3.SA', 'MRFG3.SA', 'MRVE3.SA', 'RADL3.SA', 'SANB11.SA', 'TRPL4.SA', 'UGPA3.SA', 'USIM5.SA'
6. Grupo 6: 'BEEF3.SA', 'CSMG3.SA', 'RAPT4.SA'
7. Grupo 7: 'B3SA3.SA', 'EVEN3.SA'

De modo geral, pode ser notado que houve grande variação para os agrupamentos entre os ativos, devido a influência da avaliação do período histórico para cada janela temporal. A formação de grupos de tamanhos tão discrepantes é devido a seleção dos elementos selecionados inicialmente, que foi um grupo reduzido para análise.

Depois de avaliar todos os dendrogramas e grupos, serão detalhados os gráficos comparativos entre a aplicação com agrupamentos e sem, ou seja, de forma aleatória, além das métricas utilizadas para para destacar suas diferenças.

3.4 Resultados e Comparativos

O propósito da criação desse comparativo é fundamental para análises e discussões do trabalho, que procura comparar ambas abordagens, onde as carteiras geradas de forma aleatórias não apresentam estratégia e seus grupos se formam de maneira aleatória, ao contrário da carteira construída com o método de agrupamento.

A análise comparativa permitirá avaliar a utilidade e relevância do método de agrupamento em conjunto com a otimização de portfólio, a fim de mostrar a estabilidade e consistência da aplicação.

Para construir a carteira aleatória comparativa, o período e as ações utilizadas foram os mesmos, e a quantidade de ações definidas para o loop foi referente aos grupos gerados para cada período avaliado anteriormente, assumindo que todos os ativos pertençam a um único grupo.

O método de Monte Carlo é crucial para análise, visto que é uma abordagem flexível para encontrar a carteira ótima, é possível aplicá-lo a Teoria de Markowitz e realizar a comparação com e sem agrupamento, possibilitando comparar os resultados de diferentes estratégias.

Também foi proposto a consideração do tempo de vencimento da carteira como um parâmetro para avaliar o método. Essa escolha é baseada na ideia de que a variação temporal mostre sua dinamicidade, o que reflete na aplicabilidade dos agrupamentos e na tomada de decisão para escolha da carteira.

Nesta sessão, o objetivo é analisar a performance com os gráficos de dispersão, e aprofundando um pouco mais, serão utilizadas métricas estatísticas em busca de ampliar o estudo acerca da eficiência do método. Métricas como média, desvio padrão, coeficiente de variação e amplitude gráfica.

O desvio-padrão e a média, conforme já citado anteriormente, são métricas importantes. Para análise gráfica, a média fornece um ponto de referência para a distribuição dos dados, enquanto o desvio-padrão proporciona um entendimento sobre a variabilidade dos dados. Quanto menor o desvio-padrão, mais próximos os valores estão da média, então o desvio-padrão alto sugere que os dados estão mais dispersos em relação à média [30].

Já o coeficiente de variação busca avaliar a instabilidade das variáveis em relação à média, que pode ser calculado com a fórmula a seguir [30].

$$CV = \left(\frac{\text{Desvio Padrão}}{\text{Média}} \right) \times 100 \quad (3.1)$$

Por fim, a amplitude gráfica dará uma noção referente ao intervalo em que os dados estão distribuídos no gráfico, que é simplesmente calculado pela diferença entre o valor máximo e o mínimo representados. [30].

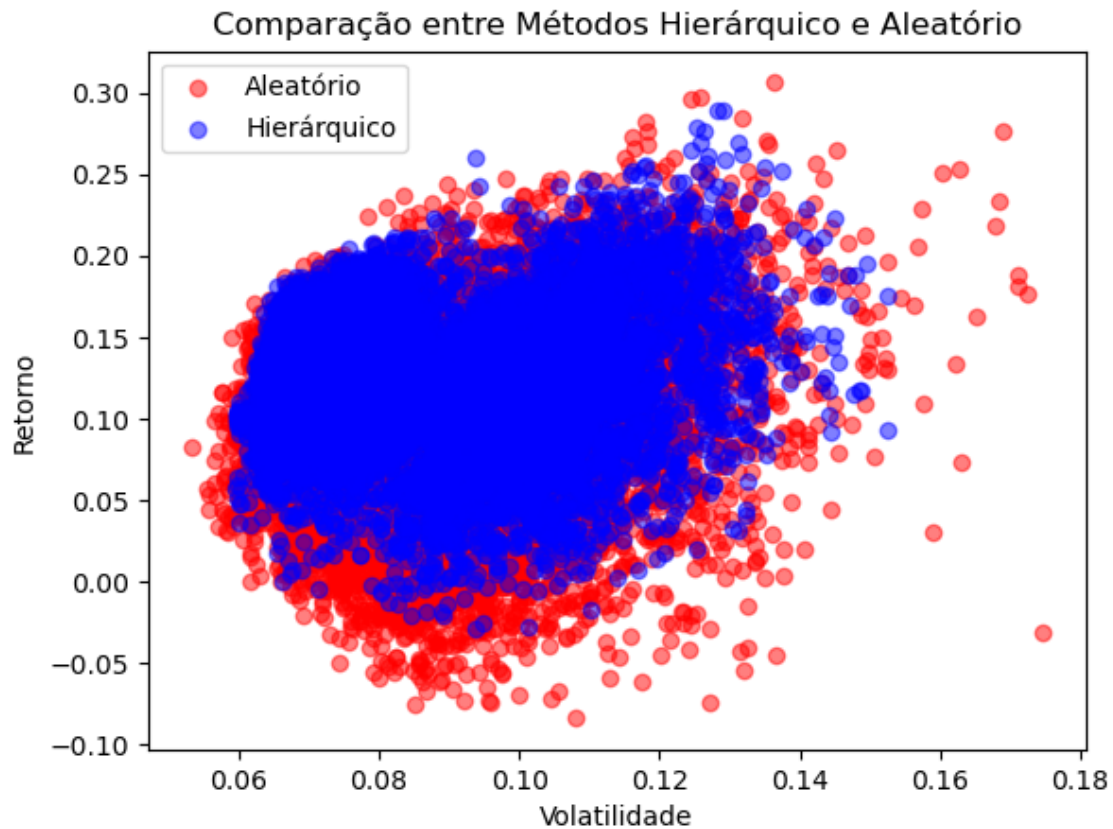
3.4.1 Observação 1: período de três meses

3.4.1.1 Comparativo menor volatilidade

A análise comparativa entre as carteiras geradas a partir do agrupamento e as aleatórias mostra diferenças significativas na eficiência.

A seguir será mostrado o gráfico de dispersão referente às 10000 carteiras geradas buscando a menor volatilidade.

Figura 16 – Gráfico menor volatilidade - três meses



Fonte: Autor (2023)

Enquanto as carteiras geradas pelo agrupamento hierárquico ocupam um espaço mais definido do gráfico, mostradas em azul, as aleatórias, que são mostradas em vermelho, exibem um comportamento mais disperso e amplo. Os resultados das métricas para esse exemplo foram:

Tabela 5 – Comparativo das métricas

Métrica	Agrupamento	Aleatório
Desvio padrão	0.473152	0.388564
Média	0.297049	0.253546
Coef. de var.	1.592840	1.532519

Fonte: Autor (2023)

As métricas mostraram resultados um pouco mais elevados para o método de agrupamento, isso pode ocorrer devido a alguns pontos do gráfico em azul terem ficado mais afastados da mancha compacta, influenciando diretamente nos resultados.

Em seguida, a amplitude gráfica para esta análise é a seguinte:

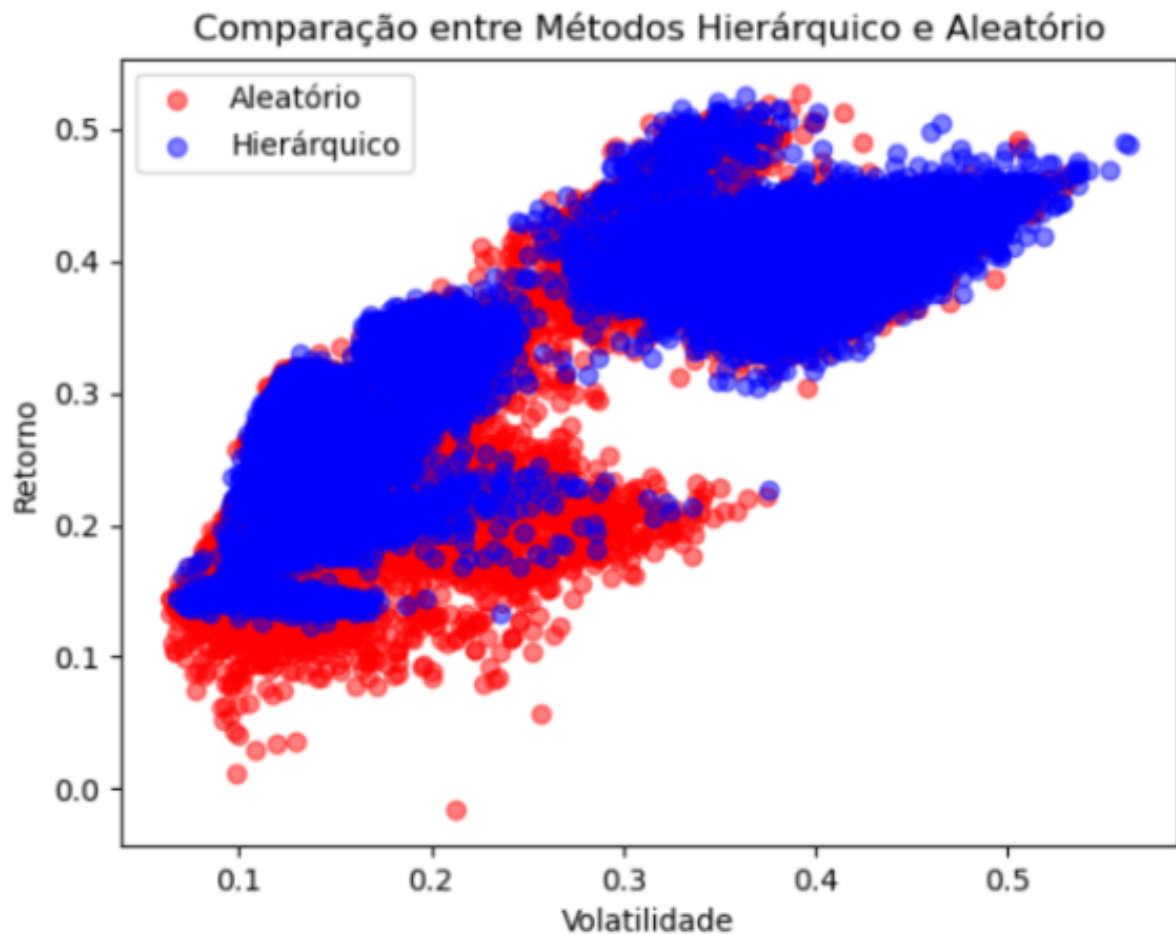
Tabela 6 – Amplitude gráfica - Menor volatilidade

Amplitude	Agrupamento	Aleatório
Retorno	0.318840	0.390212
Volatilidade	0.092918	0.121538

Fonte: Autor (2023)

3.4.1.2 Comparativo maior retorno

Para o loop em relação ao ao maior retorno têm-se o gráfico:

Figura 17 – Gráfico maior retorno - três meses

Fonte: Autor (2023)

Os resultados de suas métricas são:

Tabela 7 – Comparativo das métricas

Métrica	Agrupamento	Aleatório
Desvio padrão	0.448970	0.448715
Média	0.335114	0.324392
Coef. de var.	1.339753	1.383248

Fonte: Autor (2023)

Sua amplitude gráfica pode ser observada:

Tabela 8 – Amplitude gráfica - Maior retorno

Amplitude	Agrupamento	Aleatório
Retorno	0.402075	0.542798
Volatilidade	0.496024	0.472029

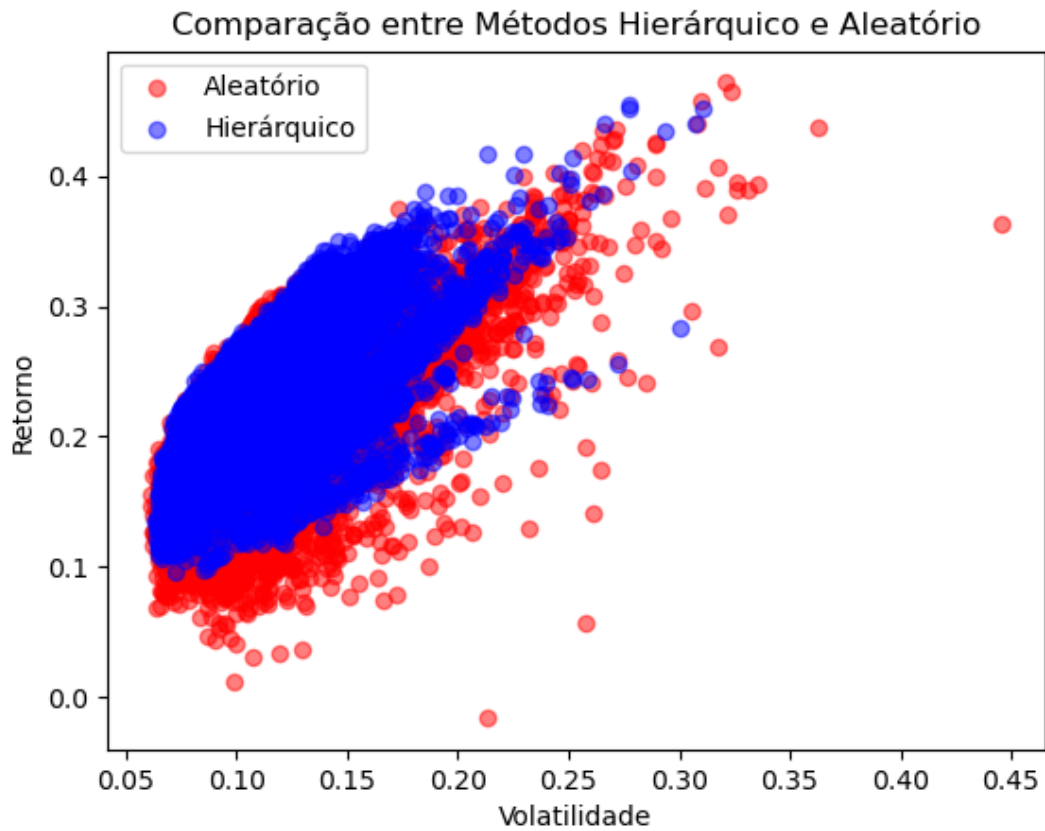
Fonte: Autor (2023)

Os dados acima apresentaram comportamento semelhante. Essas observações métricas juntamente com as gráficas, auxiliam na interpretação acerca das diferenças entre aplicações.

3.4.1.3 Comparativo maior índice Sharpe

A seguir, da mesma forma que anteriormente, serão mostrados os comparativos para o maior Sharpe Ratio observado para esse período.

Figura 18 – Gráfico maior Sharpe Ratio - três meses



Fonte: Autor (2023)

As métricas calculadas resultaram em:

Tabela 9 – Comparativo das métricas

Métrica	Agrupamento	Aleatório
Desvio padrão	.632152	0.547646
Média	0.378929	0.346198
Coef. de var.	1.668259	1.581883

Fonte: Autor (2023)

Houve algumas diferenças entre as métricas observadas, indicando pequenas dispersões de dados em relação a média.

Para complementar, a amplitude gráfica para cada um:

Tabela 10 – Amplitude gráfica - Maior índice Sharpe

Amplitude	Agrupamento	Aleatório
Retorno	0.358517	0.487882
Volatilidade	0.247149	0.385282

Fonte: Autor (2023)

Os resultados de amplitude gráfica ajudam a reforçar a ideia de que os dados aleatórios apresentam maior dispersão comparados com o agrupamento.

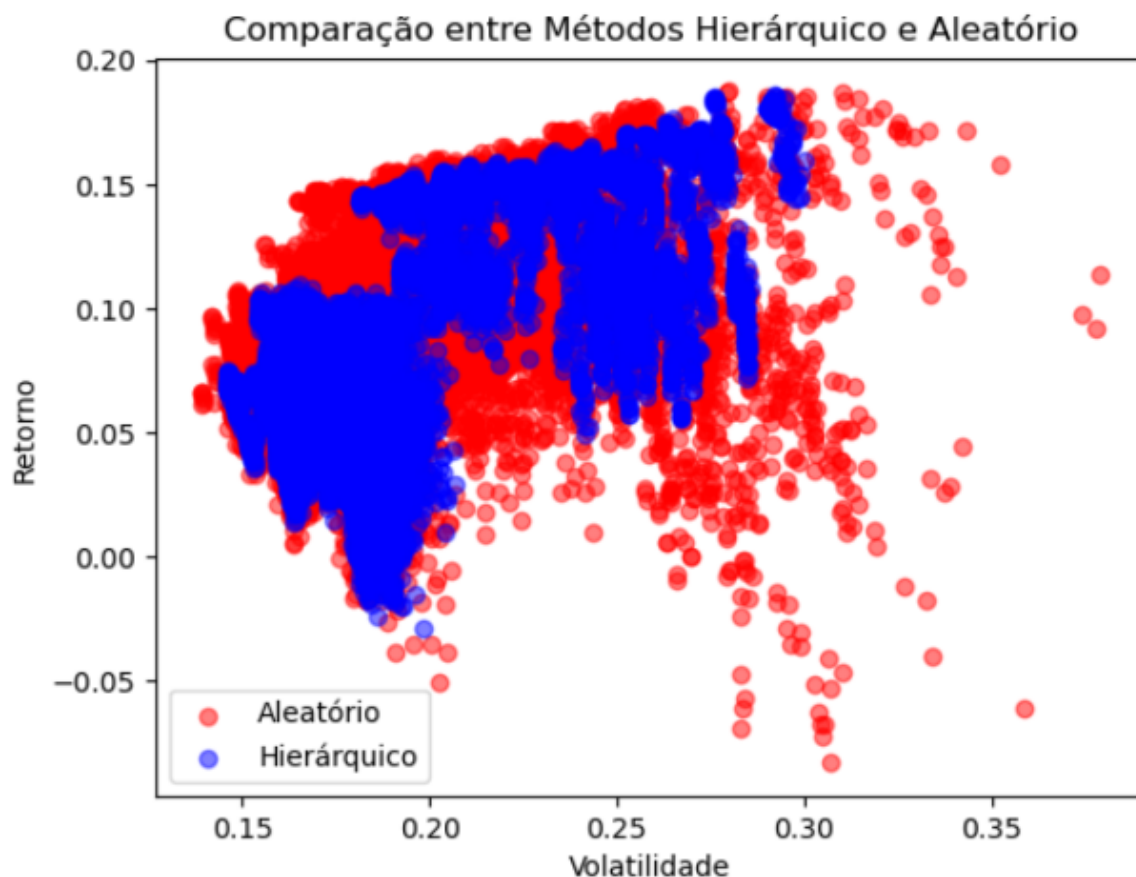
Foi observado que as carteiras formadas pelo método de clusters demonstram maior consistência gráfica, refletindo em um melhor equilíbrio quando comparada com a aleatória para um determinado nível de volatilidade. Em seguida serão feitos os mesmos comparativos para os outros períodos estudados.

3.4.2 Observação 2: período de um ano

3.4.2.1 Comparativo menor volatilidade

Para esse período, o gráfico comparativo abaixo foi gerado.

Figura 19 – Menor volatilidade - Um ano



Inicialmente, já é possível observar que na representação em azul, os dados estão mais concentrados, apresentando dados menos dispersos com a análise de agrupamentos. Algumas métricas foram calculadas para auxiliar na avaliação do gráfico.

Tabela 11 – Comparativo das métricas

Métrica	Agrupamento	Aleatório
Desvio padrão	0.198335	0.202774
Média	0.208829	0.221893
Coef. de var.	0.949750	0.913833

Fonte: Autor (2023)

Os valores de média e desvio-padrão do agrupamento resultaram em valores ligeiramente menores do que os valores aleatórios, indicando uma menor variação nos dados apresentados. Já o coeficiente de variação, para o agrupamento, resultou em um valor um pouco maior do que para os dados aleatórios, podendo indicar maior variabilidade em relação a média. Adicionado a isso, para auxiliar na interpretação, também é representado de forma mais clara a amplitude gráfica.

Tabela 12 – Amplitude gráfica - Menor volatilidade

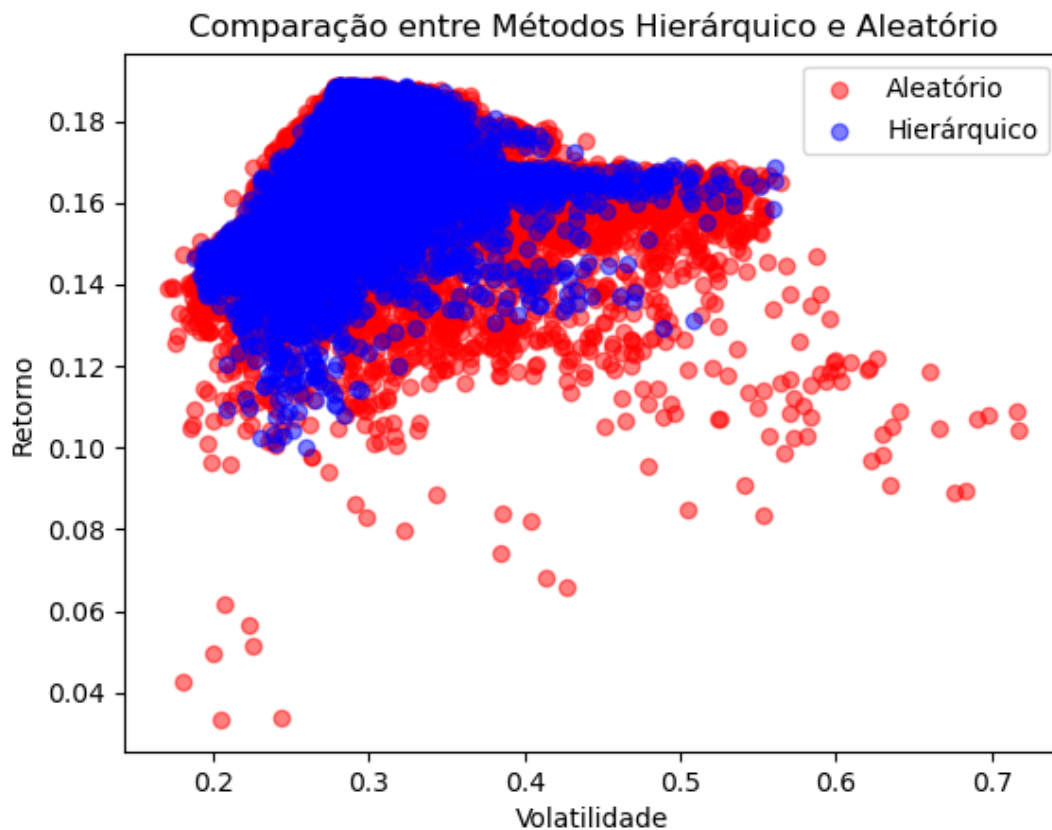
Amplitude	Agrupamento	Aleatório
Retorno	0.215368	0.270563
Volatilidade	0.153891	0.240121

Fonte: Autor (2023)

A amplitude gráfica auxilia na avaliação em cima da variação dos dados. Usando como referência o agrupamento, observa-se que para o retorno, a amplitude ficou 0,055195 a baixo em comparação com o gráfico aleatório, e para a volatilidade, ficou 0,08623 a baixo em comparação ao aleatório, que, conforme já observado anteriormente, a representação com o agrupamento, possui maior consistência nos dados.

3.4.2.2 Comparativo maior retorno

Em seguida, é apresentado o gráfico comparando o maior retorno.

Figura 20 – Maior retorno - Um ano

Fonte: Autor (2023)

Os resultados das métricas são:

Tabela 13 – Comparativo das métricas

Métrica	Agrupamento	Aleatório
Desvio padrão	0.238396	0.235859
Média	0.254031	0.251887
Coef. de var.	0.938452	0.936367

Fonte: Autor (2023)

As métricas calculadas para o gráfico de maior retorno para o método de agrupamento, obtiveram todos resultados um pouco maior do que para as simulações aleatórias. Porém, a diferença é sutil, sugerindo alguma similaridade entre os métodos aplicados. Em seguida, a tabela de amplitude gráfica é apresentada:

Tabela 14 – Amplitude gráfica - Maior retorno

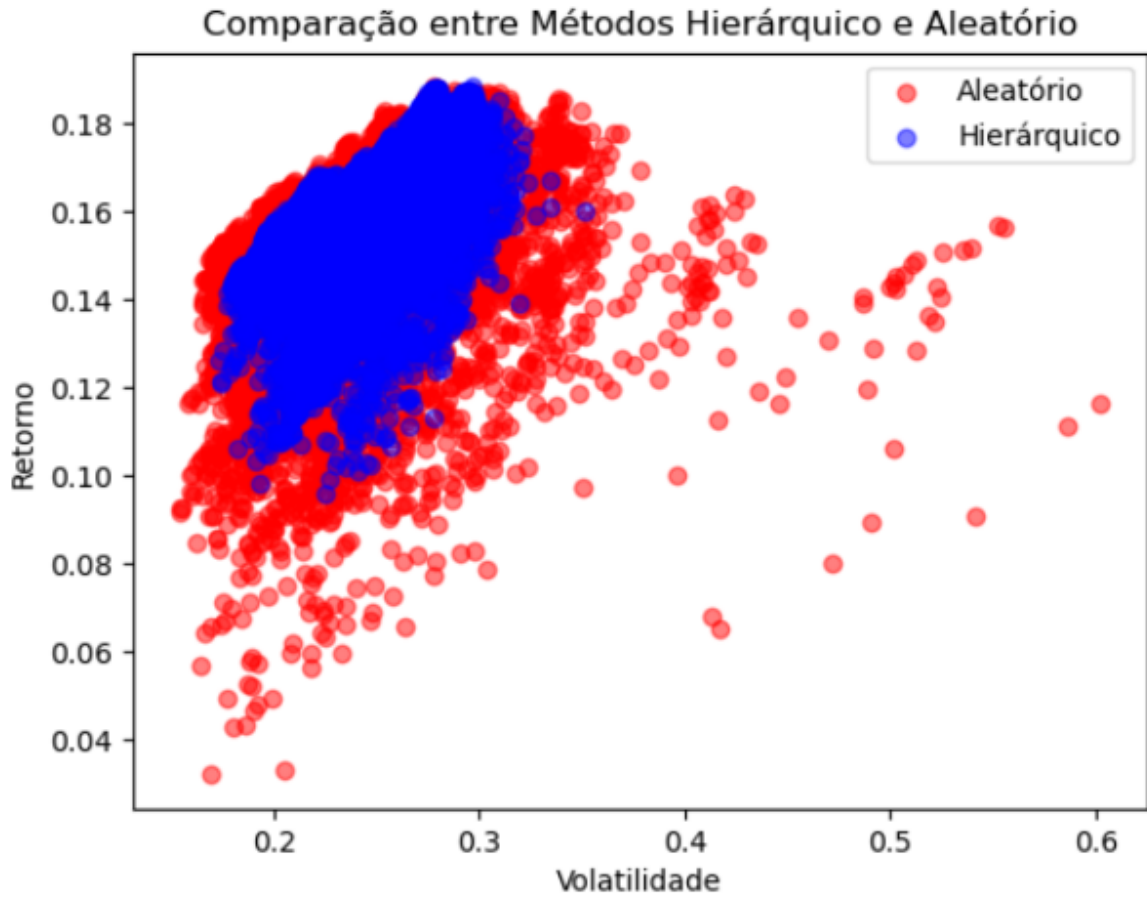
Amplitude	Agrupamento	Aleatório
Retorno	0.088741	0.155645
Volatilidade	0.372345	0.547597

Fonte: Autor (2023)

3.4.2.3 Comparativo maior índice Sharpe

A seguir será feita a análise comparativa em relação ao Sharpe Ratio. Além disso, serão apresentadas algumas métricas para complementar a compreensão da variabilidade entre os elementos analisados.

Figura 21 – Gráfico maior Sharpe Ratio - um ano



Fonte: Autor (2023)

Tabela 15 – Comparativo das métricas

Métrica	Agrupamento	Aleatório
Desvio padrão	0.224257	0.221772
Média	0.255141	0.252941
Coef. de var.	0.878954	0.876774

Fonte: Autor (2023)

Para finalizar, a amplitude gráfica é:

Tabela 16 – Amplitude gráfica - Maior índice Sharpe

Amplitude	Agrupamento	Aleatório
Retorno	0.092154	0.156213
Volatilidade	0.177735	0.448748

Fonte: Autor (2023)

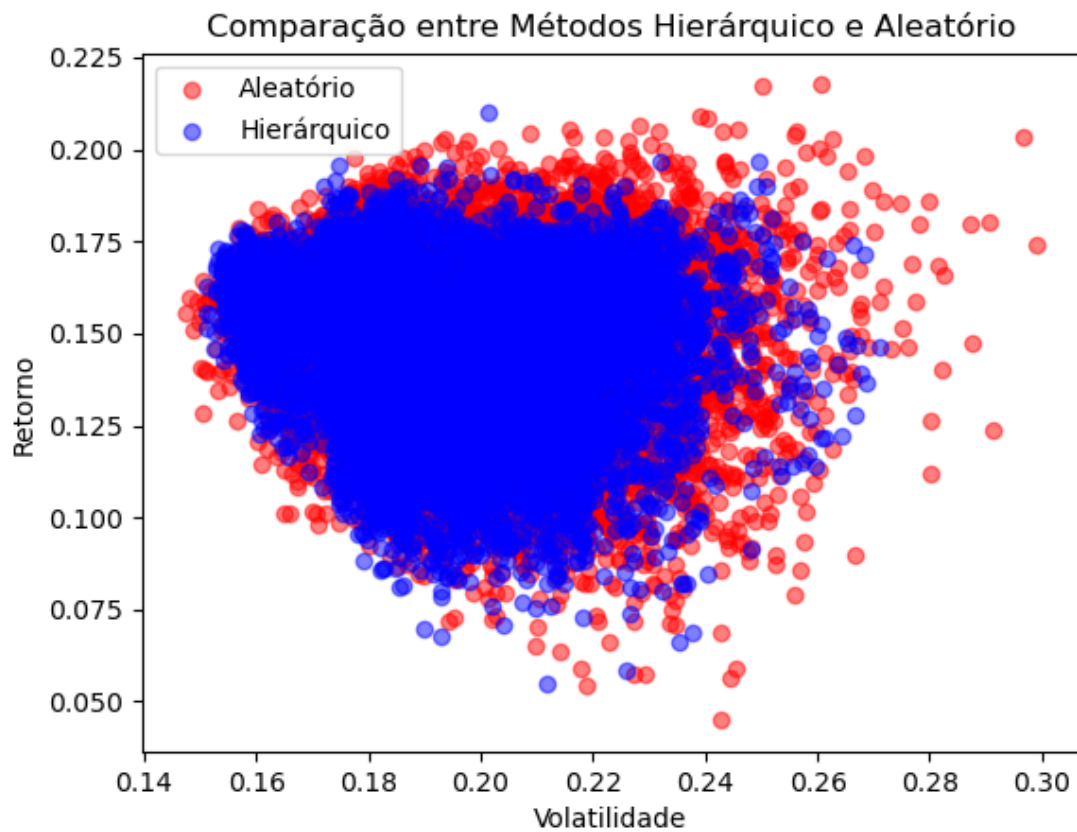
Apesar da grande proximidade entre as análises sobre desvio-padrão, média e coeficiente de variância, graficamente é possível observar grande dissemelhança entre os dados, análise essa que pode ser observada para todos os comparativos.

3.4.3 Observação 3: período de três anos

3.4.3.1 Comparativo menor volatilidade

Os gráficos comparativos para esse período serão mostrados abaixo, seguidos de suas métricas.

Figura 22 – Menor volatilidade - três anos



Fonte: Autor (2023)

Suas métricas são observadas a seguir nas tabelas 17 e 18.

Tabela 17 – Comparativo das métricas

Métrica	Agrupamento	Aleatório
Desvio padrão	0.201597	0.203545
Média	0.189029	0.189802
Coef. de var.	1.066483	1.072405

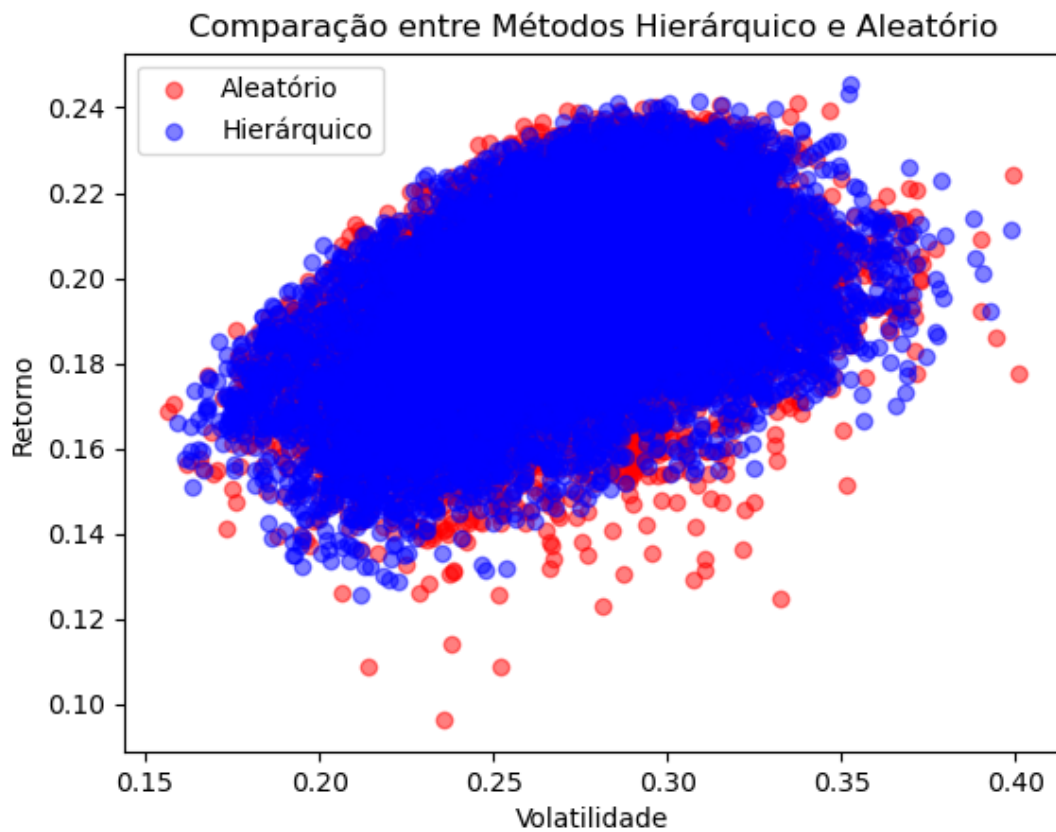
Fonte: Autor (2023)

Tabela 18 – Amplitude gráfica - Menor volatilidade

Amplitude	Agrupamento	Aleatório
Retorno	0.155263	0.172665
Volatilidade	0.120021	0.151974

Fonte: Autor (2023)

3.4.3.2 Comparativo maior retorno

Figura 23 – Maior retorno

Fonte: Autor (2023)

Tabela 19 – Comparativo das métricas

Métrica	Agrupamento	Aleatório
Desvio padrão	0.210747	0.210068
Média	0.199620	0.199814
Coef. de var.	1.055742	1.051314

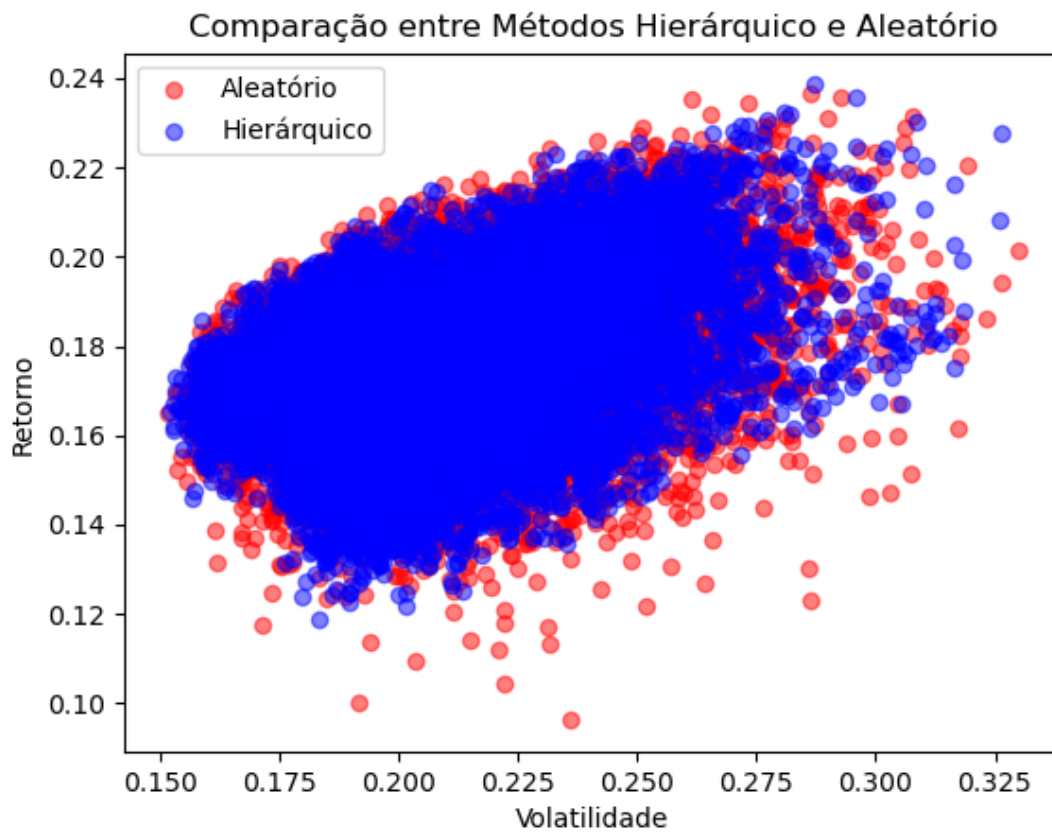
Fonte: Autor (2023)

Tabela 20 – Amplitude gráfica - Maior Retorno

Amplitude	Agrupamento	Aleatório
Retorno	0.119929	0.144869
Volatilidade	0.239587	0.244858

Fonte: Autor (2023)

3.4.3.3 Comparativo maior índice Sharpe

Figura 24 – Maior Sharpe Ratio

Fonte: Autor (2023)

Tabela 21 – Comparativo das métricas

Métrica	Agrupamento	Aleatório
Desvio padrão	0.225596	0.227009
Média	0.201975	0.202733
Coef. de var.	1.116947	1.119743

Fonte: Autor (2023)

Tabela 22 – Amplitude gráfica - Maior índice Sharpe

Amplitude	Agrupamento	Aleatório
Retorno	0.119931	0.140556
Volatilidade	0.174547	0.178684

Fonte: Autor (2023)

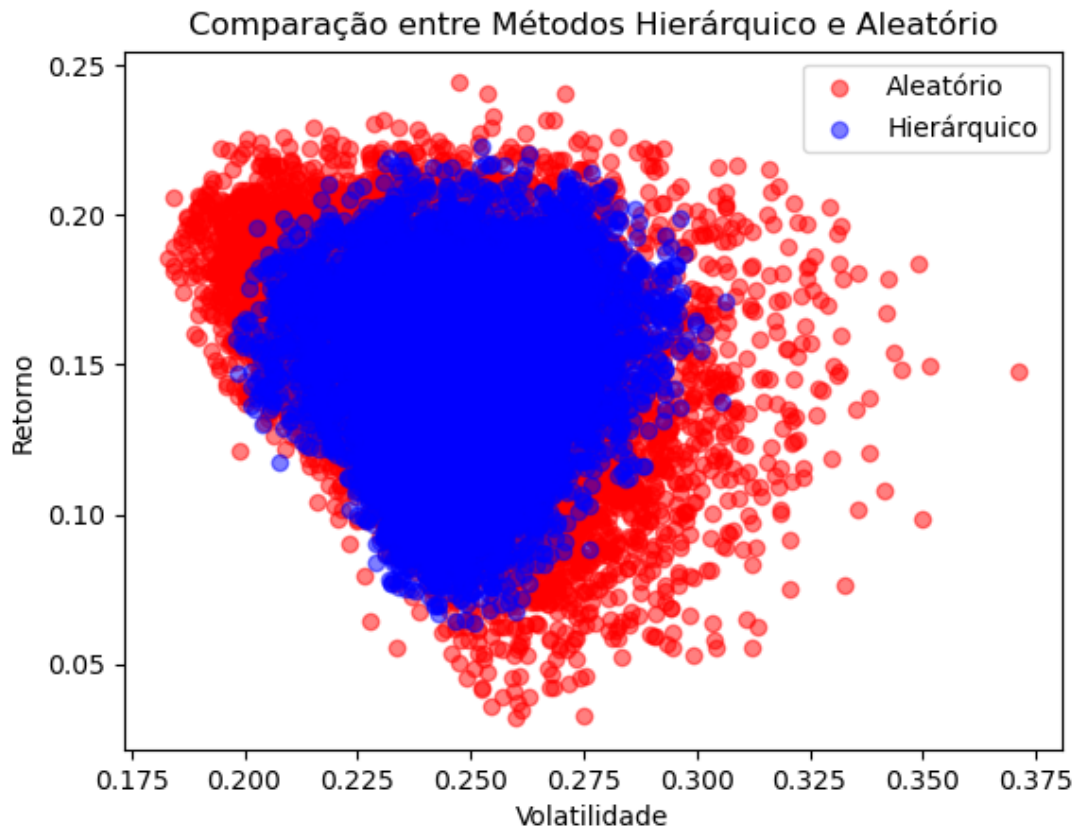
Apesar dos resultados das métricas serem bem semelhantes para todas as seções analisadas para o período de três anos, é importante destacar que a análise gráfica mostra diferenças súteis, porém de grande importância, visto que, novamente, os dados gerados pelo agrupamento mostram maior estabilidade no comportamento e na distribuição dos dados. Além disso, pode ser observado a influência do tempo avaliado,

3.4.4 Observação 4: período de cinco anos

3.4.4.1 Comparativo menor volatilidade

A seguir serão apresentados o gráfico de dispersão e os resultados de suas métricas referentes a menor volatilidade.

Figura 25 – Menor Volatilidade



Fonte: Autor (2023)

Tabela 23 – Comparativo das métricas

Métrica	Agrupamento	Aleatório
Desvio padrão	0.177069	0.194758
Média	0.197801	0.204103
Coef. de var.	0.895189	0.954213

Fonte: Autor (2023)

Tabela 24 – Amplitude gráfica - Menor volatilidade

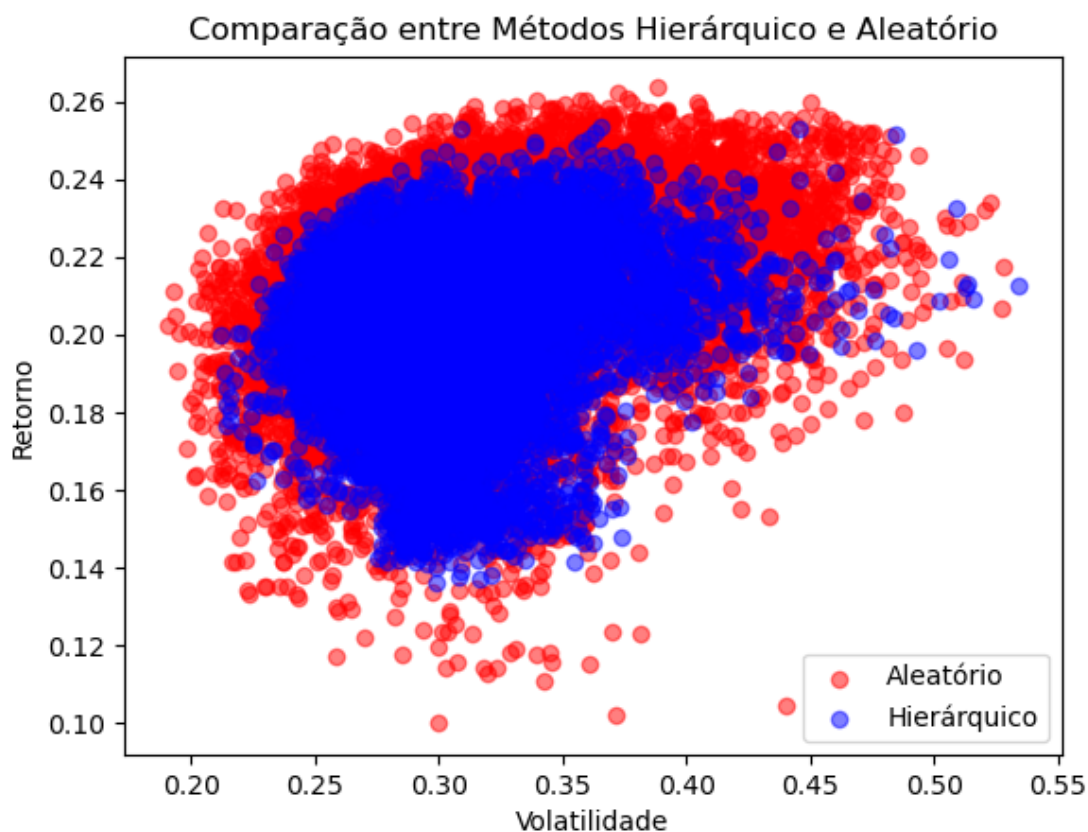
Amplitude	Agrupamento	Aleatório
Retorno	0.159135	0.212367
Volatilidade	0.108389	0.188495

Fonte: Autor (2023)

3.4.4.2 Comparativo maior retorno

A figura 26 apresenta o último gráfico comparativo acerca do melhor retorno para a análise de agrupamento hierárquico, seguidos de seus comparativos entre as métricas aplicadas.

Figura 26 – Maior Retorno



Fonte: Autor (2023)

Tabela 25 – Comparativo das métricas

Métrica	Agrupamento	Aleatório
Desvio padrão	0.207521	0.213310
Média	0.215920	0.220840
Coef. de var.	0.961100	0.965901

Fonte: Autor (2023)

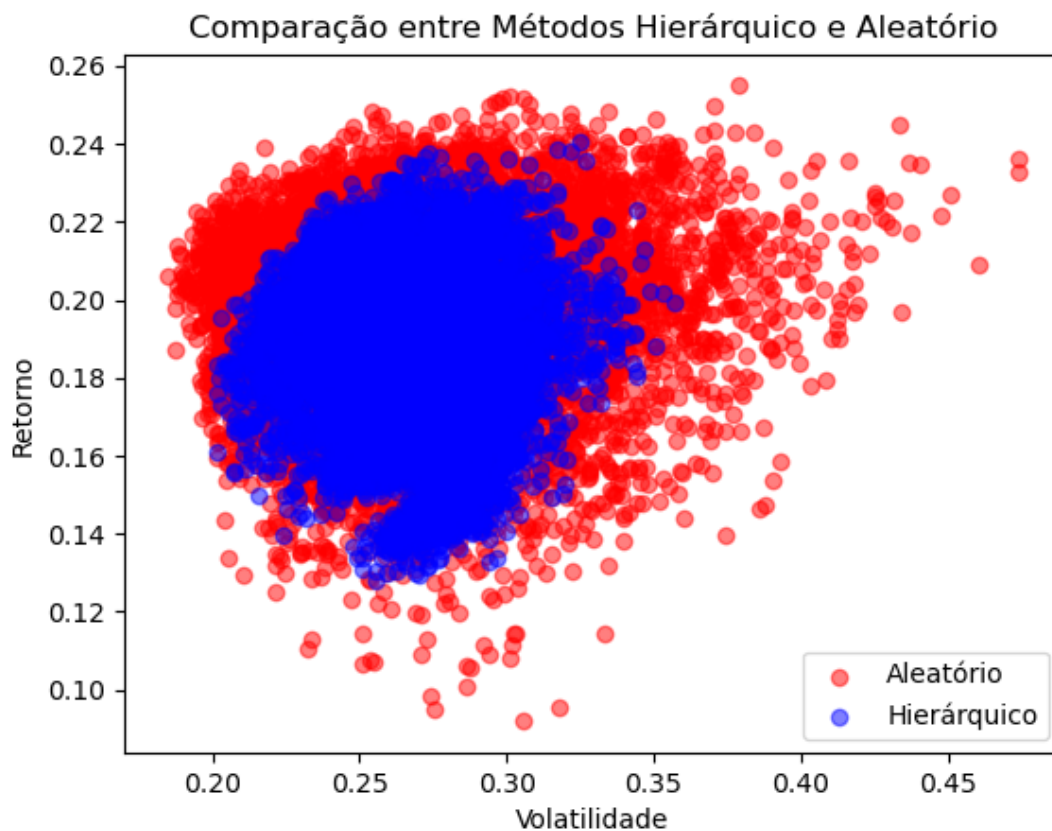
Tabela 26 – Amplitude gráfica - Maior retorno

Amplitude	Agrupamento	Aleatório
Retorno	0.117594	0.163594
Volatilidade	0.322336	0.337437

Fonte: Autor (2023)

3.4.4.3 Comparativo maior índice Sharpe

Abaixo será mostrada a otimização de carteiras referentes ao índice Sharpe

Figura 27 – Maior Sharpe Ratio

Fonte: Autor (2023)

Tabela 27 – Comparativo das métricas

Métrica	Agrupamento	Aleatório
Desvio padrão	0.208928	0.224555
Média	0.215888	0.221891
Coef. de var.	0.967761	1.0120067

Fonte: Autor (2023)

Tabela 28 – Amplitude gráfica - Maior índice Sharpe

Amplitude	Agrupamento	Aleatório
Retorno	0.112432	0.163016
Volatilidade	0.155694	0.289815

Fonte: Autor (2023)

Por fim, de modo geral, é possível observar que todas as comparações avaliadas obtiveram uma amplitude gráfica menor do que a aplicada de forma aleatória, o que indica que as carteiras obtidas através do método de agrupamento possuem uma variação menor para os valores obtidos. É visível que, quando a amplitude é menor, os resultados tendem a ser mais concentrados, sugerindo maior estabilidade na análise.

As carteiras geradas com agrupamentos apresentaram uma tendência mais focalizada para todas as seções dos períodos analisados. Essas diferenças, mesmo não sendo refletidas diretamente nas métricas aplicadas, são indispensáveis para a compreensão sobre a vantagem da utilização do método de clusterização, pois destaca a capacidade de identificar conjuntos de dados com menor volatilidade e maior estabilidade.

É possível observar também, que as diferentes janelas temporais estudadas carregam consigo informações sobre sua influência em relação ao mercado.

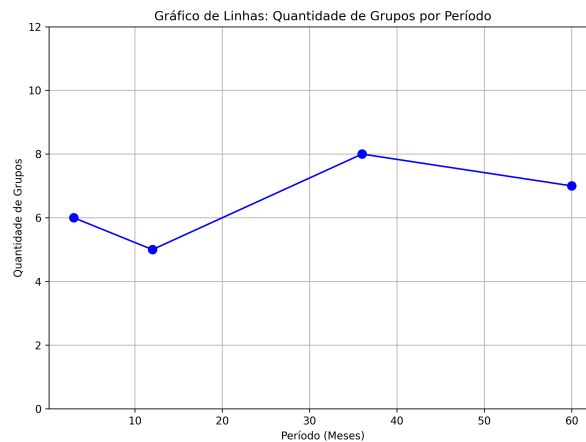
Destacando as observações significativas vistas anteriormente, observa-se o potencial da junção dos métodos em cenários mais complexos, envolvendo análises com mais ativos.

4 Agrupamento Não Hierárquico

Na gestão de portfólios, lidar com vários dados financeiros pode ser uma tarefa difícil. Com isso, este trabalho introduziu a aplicação de técnicas de agrupamento em conjunto com a teoria de Markowitz a fim de reduzir a complexidade dos dados e otimizar os resultados da carteira.

Curiosamente, a abordagem da clusterização hierárquica resultou em dendrogramas que geraram entre cinco e oito grupos para os diferentes períodos. Essa colocação é um fator significativo para o estudo.

Figura 28 – Quantidade de grupos para as análises de tempo



Fonte: Autor (2024)

A constância mostrada acima pode ser utilizada como uma pré análise para estudos referentes a agrupamento não hierárquico. A consistência vista anteriormente pode ser utilizada para definir o número de k clusters, tornando-se um ponto de partida para expandir o escopo e adicionar um conjunto de dados maiores, buscando se tornar uma aplicação mais abrangente e eficiente para o mercado financeiro.

O agrupamento não hierárquico é utilizado para agrupar elementos com características semelhantes. Para aplicar esse método é necessário que se tenha critérios pré definidos para dar início. Um dos mais utilizados é o método das k -médias, que será utilizado nesse trabalho.

4.1 Aplicação com Agrupamento Não Hierárquico

Durante o início do desenvolvimento desse trabalho foram utilizados 78 ativos para aplicação do agrupamento não hierárquico na construção de carteiras eficientes. Porém, ao evoluir da aplicação, foi constatado que dois dos 78 ativos não constavam mais na lista do yfinance, biblioteca que fornece a base de dados do trabalho. Com isso, foi preciso ajustar para 76 a quantidade de ativos na lista. O preço de fechamento ajustado também foi a métrica escolhida para análise de dados, e a distância será calculada entre sua média e desvio padrão.

Assim como no agrupamento hierárquico, para aplicação do método não hierárquico é preciso normalizar os dados para garantir que as variáveis tenham a mesma escala. Em seguida foi determinado o número de grupos para realizar o agrupamento, que, conforme demonstrado pelo corte dos dendrogramas mostrados na aplicação anterior, sugeriram quantidades de grupos de cinco a oito. Escolha que influencia na diversificação e estrutura da carteira.

Para realizar essa aplicação, os ativos selecionados são considerados um grupo inicial, em seguida, é aplicado o algoritmo k-média, pré selecionando aleatoriamente um centroide para cada grupo. Posteriormente, a distância euclidiana foi utilizada para reagrupar os ativos. Esse processo se repete até que o processo atinja a convergência. O gráfico resultante desse processo retorna em um diagrama de dispersão, que mostra a distribuição dos dados em relação aos agrupamentos formados.

Depois de executar o algoritmo, cada ativo é atribuído a um *cluster*, dando continuidade com o método de Monte Carlo que utiliza de simulações para dar pesos aleatórios para cada iteração da carteira, que serão 10000. Em seguida, realizam-se os cálculos estatísticos para otimização do portfólio em cima dessas 10000 simulações, a fim de obter uma amostra significativa de combinações. Essa abordagem permite uma adaptação dinâmica na alocação dos ativos.

4.2 Comparativo com Agrupamento Hierárquico

Nesse capítulo será explorado uma breve comparação de como as técnicas de agrupamento hierárquico e não hierárquico influenciam na eficácia da abordagem de construção de carteiras eficientes.

Por meio de um comparativo, será feita uma análise destacando a eficiência dos métodos. Lembrando que, na aplicação hierárquica, o número de ativos utilizados foi menor. Isso porque no processo de agrupamento foi decidido eliminar grupos com apenas um ativo, a fim de equilibrar o resultado.

Figura 29 – Lista com 76 ativos

```

start = dt.date (2020, 5, 1)
end = dt.date (2023, 5, 31)
acoes = [ 'ABEV3.SA', 'AGRO3.SA', 'AMER3.SA', 'AZUL4.SA', 'B3SA3.SA', 'BBAS3.SA',
'BBDC4.SA', 'BBSE3.SA', 'BEEF3.SA', 'BPAC11.SA', 'BRFS3.SA', 'BRKM5.SA',
'CCRO3.SA', 'CIEL3.SA', 'CMIG4.SA', 'COGN3.SA', 'CPFE3.SA', 'CRFB3.SA',
'CSAN3.SA', 'CSMG3.SA', 'CSNA3.SA', 'CVCB3.SA', 'CYRE3.SA', 'ECOR3.SA',
'EGIE3.SA', 'ELET6.SA', 'EMBR3.SA', 'EQTL3.SA', 'EVEN3.SA', 'EZTC3.SA',
'FLRY3.SA', 'GGBR4.SA', 'GOLL4.SA', 'GRND3.SA', 'HAPV3.SA', 'HYPE3.SA',
'IRBR3.SA', 'ITSA4.SA', 'JBSS3.SA', 'JHSF3.SA', 'JSLG3.SA', 'KLBN11.SA',
'LEVE3.SA', 'LREN3.SA', 'MDIA3.SA', 'MGLU3.SA', 'MULT3.SA', 'NTCO3.SA',
'PCAR3.SA', 'PETR4.SA', 'MRFG3.SA', 'MRVE3.SA', 'PRIO3.SA', 'POMO4.SA',
'PSSA3.SA', 'PTBL3.SA', 'QUAL3.SA', 'RADL3.SA', 'RAPT4.SA', 'RAIL3.SA',
'RENT3.SA', 'SANB11.SA', 'SBSP3.SA', 'SLCE3.SA', 'SHUL4.SA', 'SUZB3.SA',
'TAEE11.SA', 'TIMS3.SA', 'TOTS3.SA', 'TRPL4.SA', 'UGPA3.SA', 'USIM5.SA',
'VALE3.SA', 'VIVT3.SA', 'YDUQ3.SA', 'WEGE3.SA' ]

```

Fonte: Autor (2024)

No agrupamento hierárquico os dendrogramas podem ser particionados em diferentes níveis para obter o número de grupos. Por outro lado, no agrupamento não hierárquico é preciso pré determinar esse número.

Os experimentos a seguir consistem na aplicação do método k-médias, onde usando as técnicas de simulações apresentadas anteriormente. Serão gerados cenários de aplicação para o período de um ano, que obteve como resultado o dendrograma com cinco grupos diferentes. A escolha dessa janela de tempo foi porque ela resultou em uma quantidade de grupos menor, escolha que permite explorar melhor possíveis relações entre os métodos.

A fim de estabelecer um centro inicial para cada um desses grupos, será utilizado o primeiro elemento de cada *cluster*. Esses elementos servirão como ponto de referência para iniciar o agrupamento. A lista com os primeiros ativos é mostrada abaixo:

Ativos iniciais = ['GGBR4.SA', 'EQTL3.SA', 'ABEV3.SA', 'BEEF3.SA', 'BBDC4.SA']

A definição desse ponto inicial pode influenciar diretamente na construção dos *clusters*. No caso desta aplicação, a escolha pelo primeiro elemento foi de forma intuitiva, de modo a ter um ponto fixo como centroide.

O diagrama resultante do k-médias, realizado com os cinco centroides também pré definidos, pode ser visto a seguir:

Figura 30 – Diagrama de dispersão

Fonte: Autor (2024)

Cada grupo ficou com a seguinte quantidade de elementos:

Tabela 29 – Elementos por grupo - um ano

Grupo	Ativos
Grupo 1	9
Grupo 2	19
Grupo 3	23
Grupo 4	6
Grupo 5	19

Fonte: Autor (2024)

Cada grupo contém os ativos:

- Grupo 1: 'BRKM5.SA', 'ELET6.SA', 'IRBR3.SA', 'MDIA3.SA', 'PRIO3.SA', 'RENT3.SA', 'SBSP3.SA', 'VALE3.SA', 'WEGE3.SA'
- Grupo 2: 'BBAS3.SA', 'BBSE3.SA', 'BPAC11.SA', 'CIEL3.SA', 'CSMG3.SA', 'CYRE3.SA', 'EMBR3.SA', 'EZTC3.SA', 'GGBR4.SA', 'HYPE3.SA', 'JSLG3.SA', 'LEVE3.SA', 'PETR4.SA', 'POMO4.SA', 'PSSA3.SA', 'RADL3.SA', 'RAIL3.SA', 'UGPA3.SA', 'YDUQ3.SA'

3. Grupo 3: 'ABEV3.SA', 'B3SA3.SA', 'CCRO3.SA', 'CMIG4.SA', 'COGN3.SA', 'CPFE3.SA', 'ECOR3.SA', 'EGIE3.SA', 'EQTL3.SA', 'EVEN3.SA', 'FLRY3.SA', 'GRND3.SA', 'ITSA4.SA', 'KLBN11.SA', 'MGLU3.SA', 'MRVE3.SA', 'MULT3.SA', 'PTBL3.SA', 'RAPT4.SA', 'SANB11.SA', 'SHUL4.SA', 'TIMS3.SA', 'TRPL4.SA'
4. Grupo 4: 'AMER3.SA', 'CRFB3.SA', 'CVCB3.SA', 'JBSS3.SA', 'MRFG3.SA', 'QUAL3.SA'
5. Grupo 5: 'AGRO3.SA', 'AZUL4.SA', 'BBDC4.SA', 'BEEF3.SA', 'BRFS3.SA', 'CSAN3.SA', 'CSNA3.SA', 'GOLL4.SA', 'HAPV3.SA', 'JHSF3.SA', 'LREN3.SA', 'NTCO3.SA', 'PCAR3.SA', 'SLCE3.SA', 'SUZB3.SA', 'TAE11.SA', 'TOTS3.SA', 'USIM5.SA', 'VIVT3.SA'

Avaliando os grupos gerados pelos métodos, é possível ver que alguns clusters gerados pelos agrupamentos correspondem entre si. Por exemplo, ao olhar para os grupos hierárquicos é possível ver que eles estão sobrepostos nos não hierárquicos. Essa similaridade entre grupos gerados indica certa consistência na estrutura dos dados.

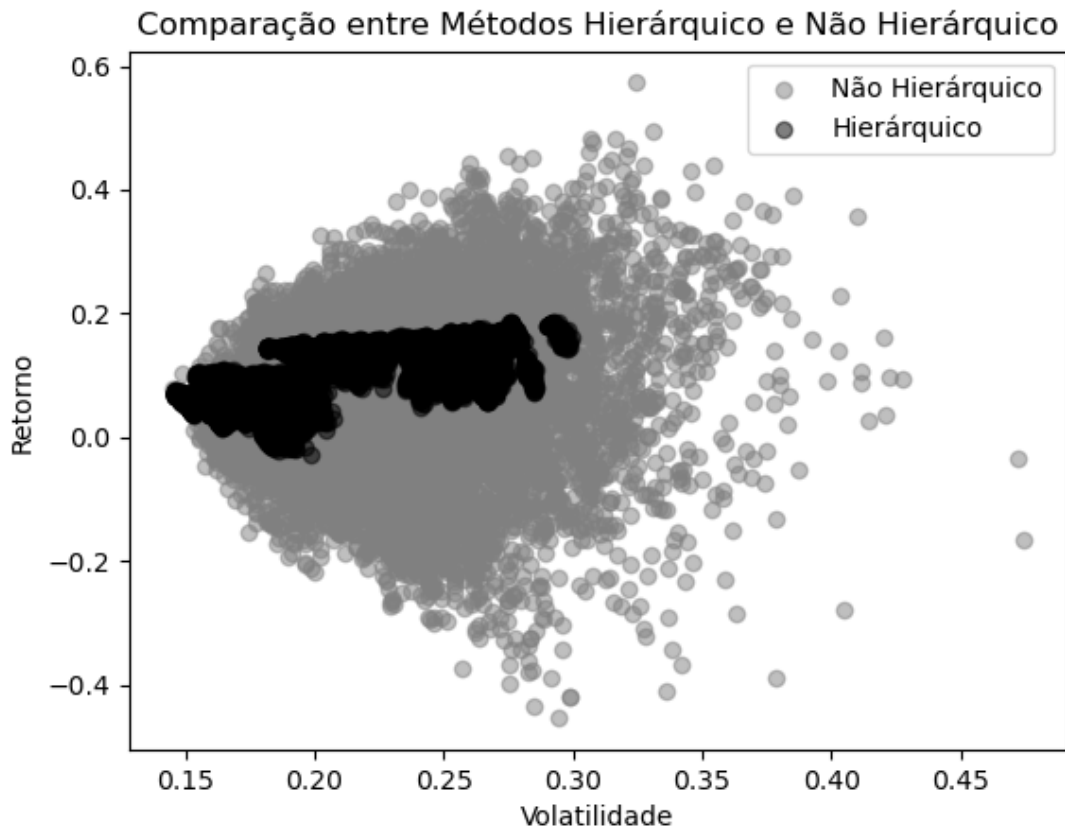
Observa-se que os grupos quatro e cinco gerados pelo agrupamento hierárquico estão contidos no quinto grupo do não hierárquico. Essa análise sugere que os métodos convergem para um resultado semelhante para os dados em questão.

Visto a formação dos grupos, serão apresentados os comparativos da otimização da carteira de Markowitz com agrupamento não hierárquico.

4.2.1 Comparativo menor volatilidade

A seguir será mostrado o comportamento dos resultados para menor volatilidade:

Figura 31 – Comparativo menor volatilidade - ANH



Fonte: Autor (2024)

A partir dele, a amplitude gráfica é:

Tabela 30 – Amplitude gráfica - Menor volatilidade

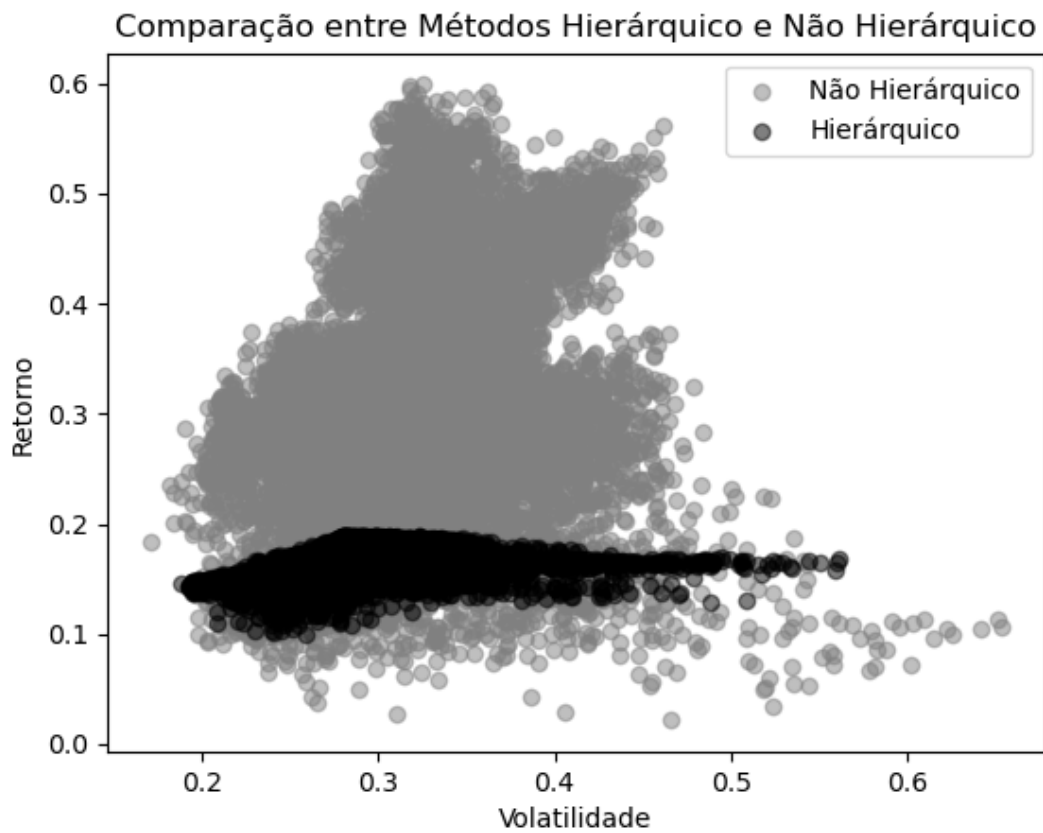
Amplitude	Não Hierárquico	Hierárquico
Retorno	1.027644	0.215368
Volatilidade	0.329909	0.153891

Fonte: Autor (2024)

4.2.2 Comparativo maior retorno

O resultado da otimização para encontrar o maior retorno é mostrado na figura abaixo:

Figura 32 – Comparativo maior retorno - ANH



Fonte: Autor (2024)

Segue na tabela 31, o resultado da amplitude gráfica:

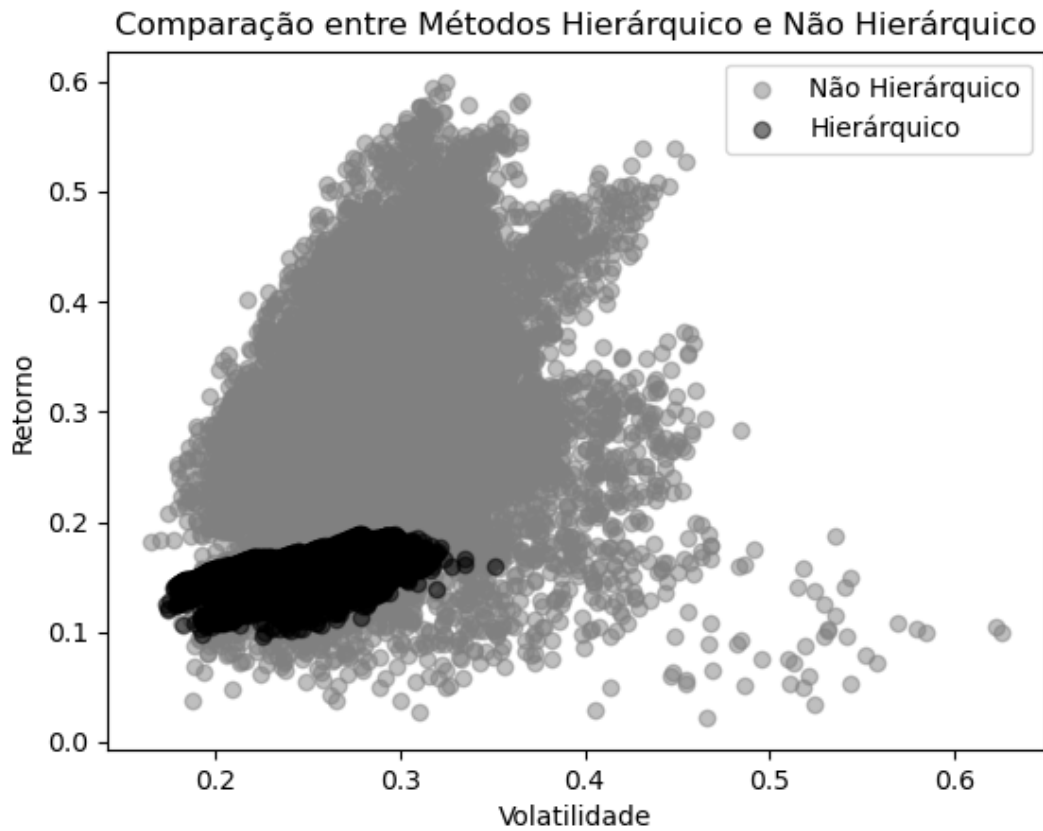
Tabela 31 – Amplitude gráfica - Maior retorno

Amplitude	Não Hierárquico	Hierárquico
Retorno	0.577339	0.088741
Volatilidade	0.483314	0.372345

Fonte: Autor (2024)

4.2.3 Comparativo maior índice Sharpe

Em seguida, o comparativo em relação ao *Sharpe Ratio*:

Figura 33 – Comparativo maior índice Sharpe - ANH

Fonte: Autor (2024)

Por fim, sua amplitude:

Tabela 32 – Amplitude gráfica - Maior índice Sharpe

Amplitude	Não Hierárquico	Hierárquico
Retorno	0.577339	0.092154
Volatilidade	0.461164	0.177735

Fonte: Autor (2024)

Essa diferença apresentada nas amplitudes gráficas vistas anteriormente pode ser atribuída à quantidade de ativos. Isso porque a inclusão de novas ações tende a aumentar a diversidade do conjunto de dados, influenciando diretamente nos resultados.

Por outro lado, é importante avaliar a disparidade entre as amplitudes gráficas. Esse resultado pode indicar um conjunto mais restrito de possíveis soluções.

Essas observações necessitam de uma avaliação mais profunda para entender a natureza e o comportamento das carteiras ótimas geradas. Além disso, é importante

considerar as necessidades do investidor. Em cenários onde se prefere uma maior variedade de ativos em diferentes níveis, o agrupamento não hierárquico pode ser mais recomendado para análise dos dados, enquanto o hierárquico pode ser utilizado para simplificar o processo de seleção de carteiras.

4.3 Comparativos com Agrupamento Não Hierárquico em Múltiplos Períodos

Nesta sessão será explorado a aplicação do agrupamento não hierárquico, onde o k a ser utilizado será dado a partir da média entre os grupos obtidos com a técnica hierárquica. As quantidades de grupos foram 5, 6, 7 e 8, sendo sua média igual a 6,5. De forma intuitiva, esse valor foi arredondado para 6 grupos para ser utilizado como k .

Conforme utilizado anteriormente, serão utilizados como centros iniciais os primeiros elementos do agrupamento hierárquico que gerou seis grupos. Nesse caso foi o período de três meses. Os centros que serão utilizados inicialmente são:

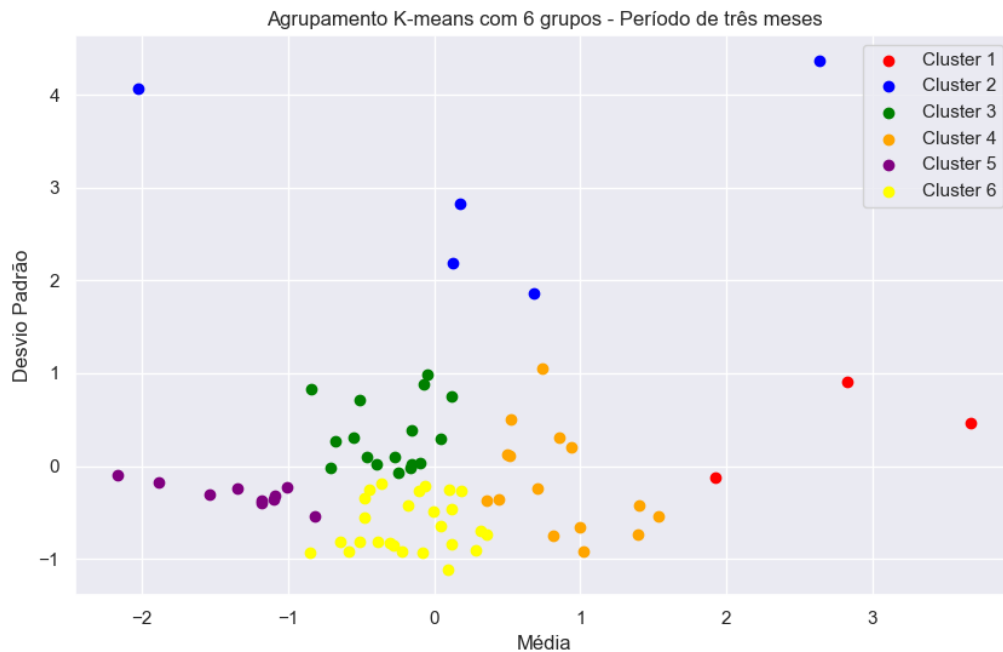
Ativos iniciais = ['GOLL4.SA', 'BRKM5.SA', 'SBSP3.SA', 'CYRE3.SA', 'AGRO3.SA', 'ABEV3.SA']

Essa sessão visa analisar a estabilidade e eficácia de uma carteira com uma quantidade k de grupos para os quatro períodos de tempo avaliados, além de explorar métricas como retorno, risco e índice Sharpe.

Embora na aplicação do método hierárquico tenham sido aplicadas outras métricas, em busca de simplificar a análise, a amplitude gráfica foi a única utilizada como comparativo para avaliar o desempenho do portfólio. Isso aconteceu, pois elas não forneceram informações significativas na análise inicial.

4.3.1 Observação 1: período de três meses

Inicialmente serão avaliados os resultados referentes a três meses, o diagrama de dispersão para esse período foi:

Figura 34 – Diagrama de dispersão - três meses

Fonte: Autor (2024)

Em seguida a quantidade de ativos para cada um dos seis grupos são:

Tabela 33 – Elementos por grupo - três meses

Grupo	Ativos
Grupo 1	3
Grupo 2	5
Grupo 3	17
Grupo 4	26
Grupo 5	10
Grupo 6	26

Fonte: Autor (2023)

As ações presentes em cada grupo estão a seguir:

1. Grupo 1: 'AZUL4.SA', 'MRVE3.SA', 'YDUQ3.SA'
2. Grupo 2: 'IRBR3.SA', 'LEVE3.SA', 'MDIA3.SA', 'RENT3.SA', 'VALE3.SA'
3. Grupo 3: 'BBAS3.SA', 'BRKM5.SA', 'CPFE3.SA', 'EGIE3.SA', 'ELET6.SA', 'EMBR3.SA', 'EQTL3.SA', 'HYPE3.SA', 'KLBN11.SA', 'LREN3.SA', 'NTCO3.SA', 'PETR4.SA', 'PRIO3.SA', 'SANB11.SA', 'SBSP3.SA', 'SUZB3.SA', 'VIVT3.SA'

-
4. Grupo 4: 'B3SA3.SA', 'BPAC11.SA', 'CCRO3.SA', 'COGN3.SA', 'CSMG3.SA', 'CYRE3.SA', 'ECOR3.SA', 'EVEN3.SA', 'EZTC3.SA', 'GOLL4.SA', 'HAPV3.SA', 'POMO4.SA', 'RADL3.SA', 'RAPT4.SA', 'UGPA3.SA'
 5. Grupo 5: 'AGRO3.SA', 'BBSE3.SA', 'BEEF3.SA', 'CRFB3.SA', 'CSNA3.SA', 'GGBR4.SA', 'JBSS3.SA', 'PSSA3.SA', 'SLCE3.SA', 'WEGE3.SA'
 6. Grupo 6: 'ABEV3.SA', 'AMER3.SA', 'BBDC4.SA', 'BRFS3.SA', 'CIEL3.SA', 'CMIG4.SA', 'CSAN3.SA', 'CVCB3.SA', 'FLRY3.SA', 'GRND3.SA', 'ITSA4.SA', 'JHSF3.SA', 'JSLG3.SA', 'MGLU3.SA', 'MRFG3.SA', 'MULT3.SA', 'PCAR3.SA', 'PTBL3.SA', 'QUAL3.SA', 'RAIL3.SA', 'SHUL4.SA', 'TAEE11.SA', 'TIMS3.SA', 'TOTS3.SA', 'TRPL4.SA', 'USIM5.SA'

Assim como no agrupamento hierárquico, um dos grupos formados contém mais ativos dos que o restante, indicando alguma concentração dentro desse subconjunto em específico.

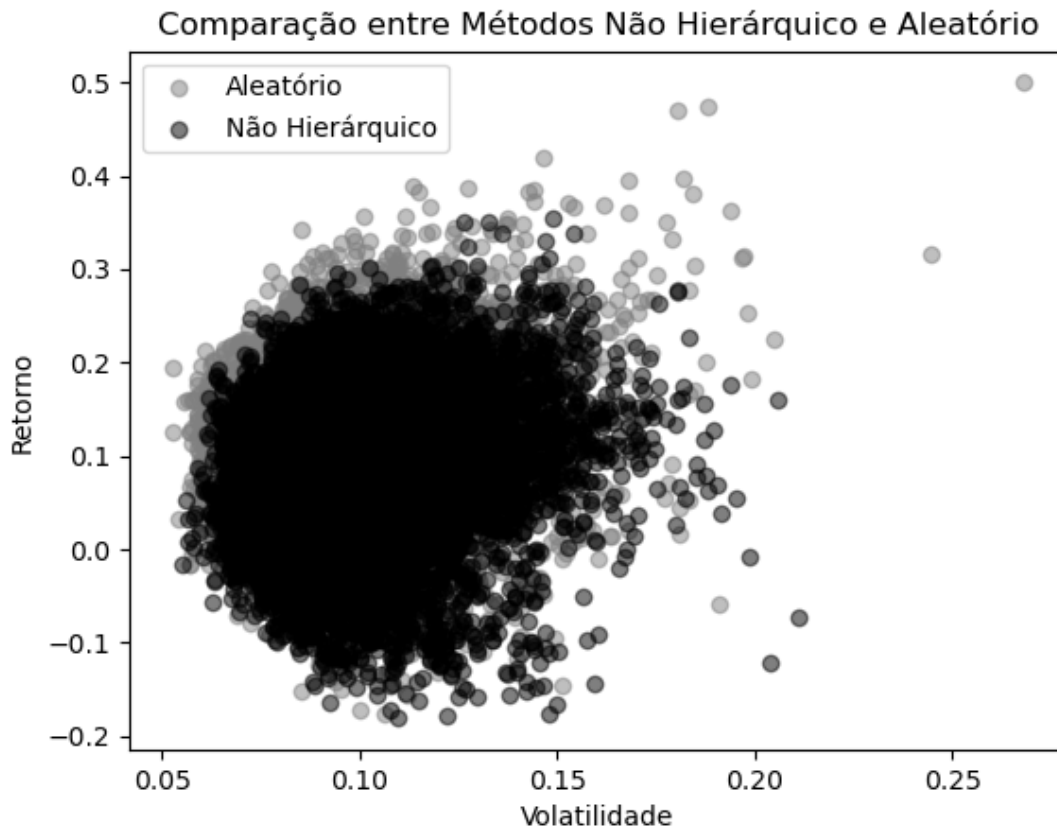
Foi observado que os grupos obtiveram pelo menos três ativos contidos. Isso pode ser uma interpretação positiva dos resultados, obter um número mínimo de elementos, possibilita a realização de análises estatísticas para melhor visualizar a distribuição dos dados dentro dos *clusters*.

Em seguida serão apresentados os gráficos de dispersão comparativos para aprofundar a análise.

4.3.1.1 Comparativo menor volatilidade

O exemplo abaixo representa o comparativo da menor volatilidade:

Figura 35 – Comparativo menor volatilidade com 6 grupos - 3 meses



Fonte: Autor (2024)

Em seguida, a tabela mostra a amplitude gráfica para aplicação com agrupamento não e hierárquico e de forma aleatória.

Tabela 34 – Amplitude gráfica - Menor volatilidade

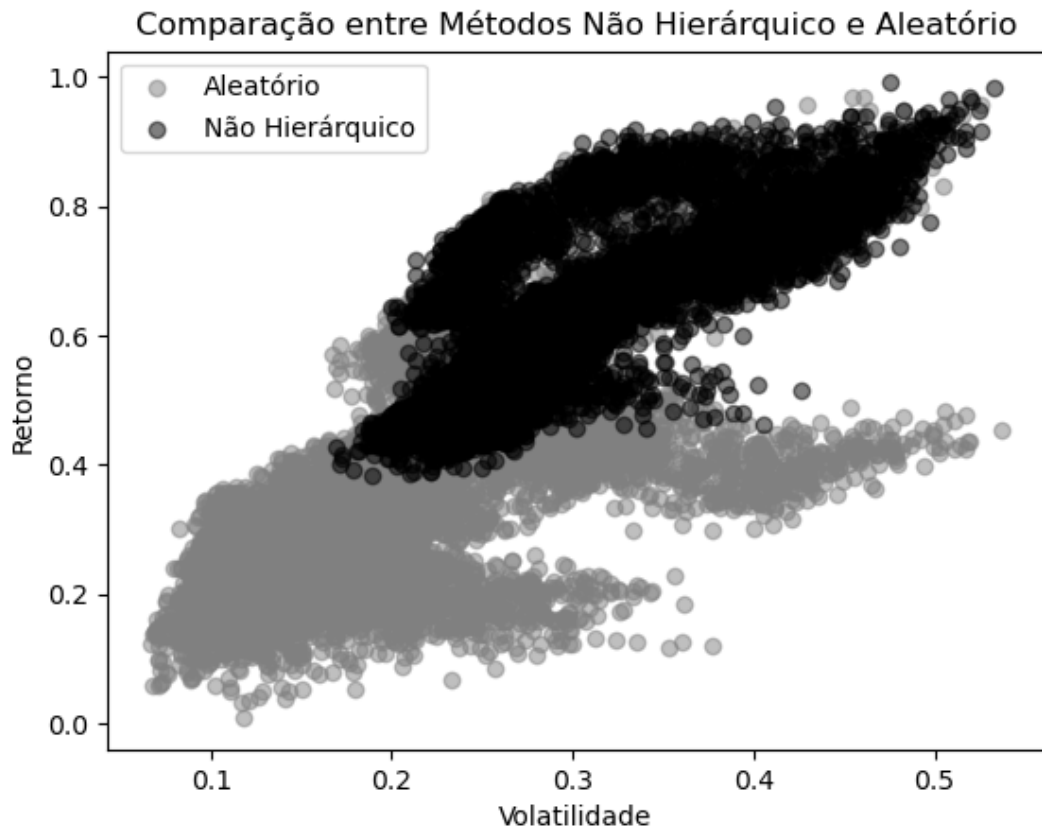
Amplitude	Agrupamento	Aleatório
Retorno	0.535087	0.675227
Volatilidade	0.155688	0.215627

Fonte: Autor (2024)

4.3.1.2 Comparativo maior retorno

As carteiras ótimas geradas com a otimização do retorno são mostradas na figura abaixo:

Figura 36 – Comparativo maior retorno com 6 grupos - 3 meses



Fonte: Autor (2024)

Sua amplitude gráfica resultou em:

Tabela 35 – Amplitude gráfica - Maior retorno

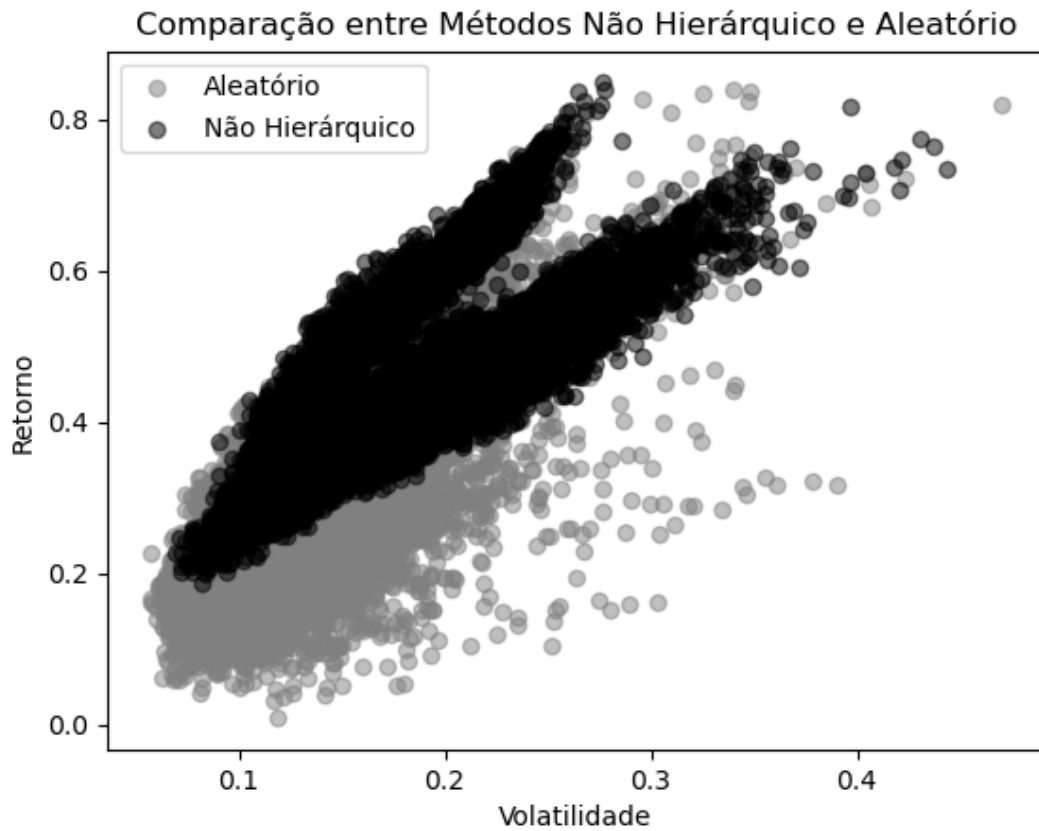
Amplitude	Agrupamento	Aleatório
Retorno	0.607898	0.960788
Volatilidade	0.363878	0.471092

Fonte: Autor (2024)

4.3.1.3 Comparativo maior índice Sharpe

O gráfico apresentado na figura 37 mostra o comparativo entre as carteiras com o maior *Sharpe Ratio*.

Figura 37 – Comparativo maior índice Sharpe com 6 grupos - 3 meses



Fonte: Autor (2024)

Para o gráfico anterior têm-se a amplitude:

Tabela 36 – Amplitude gráfica - Maior índice Sharpe

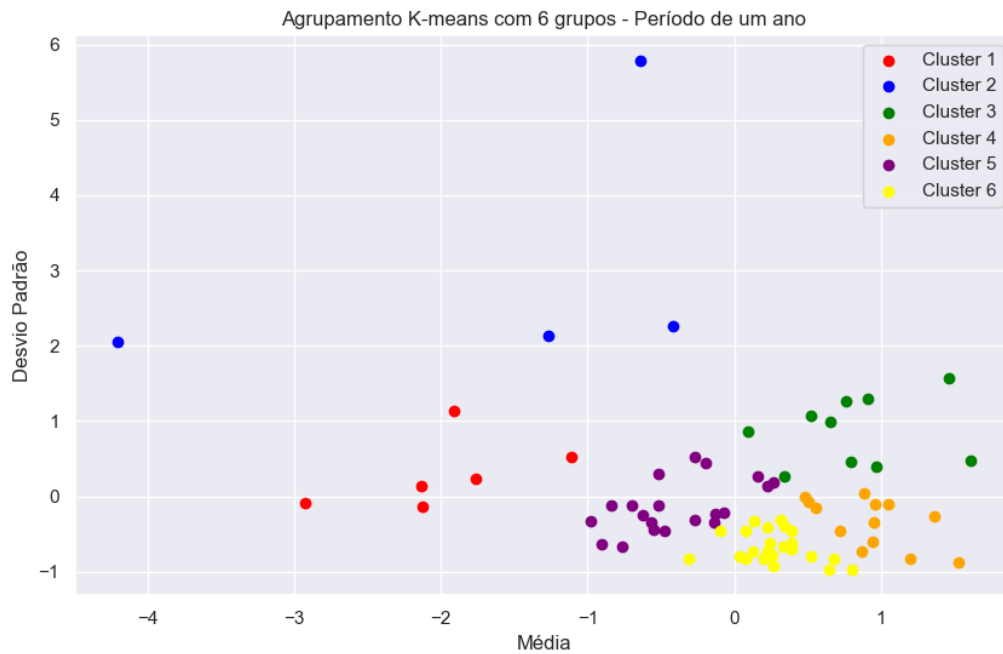
Amplitude	Agrupamento	Aleatório
Retorno	0.660582	0.830070
Volatilidade	0.374211	0.413715

Fonte: Autor (2024)

Os resultados referentes ao período de três meses, se mostraram bem compactos. A faixa colorida em preto pelas carteiras geradas a partir da otimização com o agrupamento não hierárquico apresentaram maior similaridade em seu comportamento em todas as avaliações, em especial para o maior retorno e maior índice Sharpe.

4.3.2 Observação 2: período de um ano

A figura 38 mostra o agrupamento correspondente ao período de um ano.

Figura 38 – Diagrama de dispersão - um ano

Fonte: Autor (2024)

A tabela abaixo contém a quantidade de elementos de cada grupo:

Tabela 37 – Elementos por grupo - um ano

Grupo	Ativos
Grupo 1	6
Grupo 2	4
Grupo 3	10
Grupo 4	13
Grupo 5	20
Grupo 6	23

Fonte: Autor (2023)

Em seguida, a base de ativos de cada *clusters* ilustrado.

1. Grupo 1: 'BRFS3.SA', 'CRFB3.SA', 'CVCB3.SA', 'JBSS3.SA', 'MRFG3.SA', 'QUAL3.SA'
2. Grupo 2: 'AMER3.SA', 'BRKM5.SA', 'IRBR3.SA', 'VALE3.SA'
3. Grupo 3: 'BBAS3.SA', 'BBSE3.SA', 'ELET6.SA', 'HYPE3.SA', 'LEVE3.SA', 'MDIA3.SA', 'PRIO3.SA', 'RENT3.SA', 'SBSP3.SA', 'WEGE3.SA'

4. Grupo 4: 'BPAC11.SA', 'CIEL3.SA', 'CSMG3.SA', 'CYRE3.SA', 'EMBR3.SA', 'EZTC3.SA', 'JSLG3.SA', 'PETR4.SA', 'POMO4.SA', 'PSSA3.SA', 'RADL3.SA', 'RAIL3.SA', 'UGPA3.SA'
5. Grupo 5: 'AGRO3.SA', 'AZUL4.SA', 'BBDC4.SA', 'BEEF3.SA', 'CSAN3.SA', 'CSNA3.SA', 'GGBR4.SA', 'GOLL4.SA', 'HAPV3.SA', 'JHSF3.SA', 'LREN3.SA', 'NTCO3.SA', 'PCAR3.SA', 'SLCE3.SA', 'SUZB3.SA', 'TAE11.SA', 'TOTS3.SA', 'USIM5.SA', 'VIVT3.SA', 'YDUQ3.SA'
6. Grupo 6: 'ABEV3.SA', 'B3SA3.SA', 'CCRO3.SA', 'CMIG4.SA', 'COGN3.SA', 'CPFE3.SA', 'ECOR3.SA', 'EGIE3.SA', 'EQTL3.SA', 'EVEN3.SA', 'FLRY3.SA', 'GRND3.SA', 'ITSA4.SA', 'KLBN11.SA', 'MGLU3.SA', 'MRVE3.SA', 'MULT3.SA', 'PTBL3.SA', 'RAPT4.SA', 'SANB11.SA', 'SHUL4.SA', 'TIMS3.SA', 'TRPL4.SA'

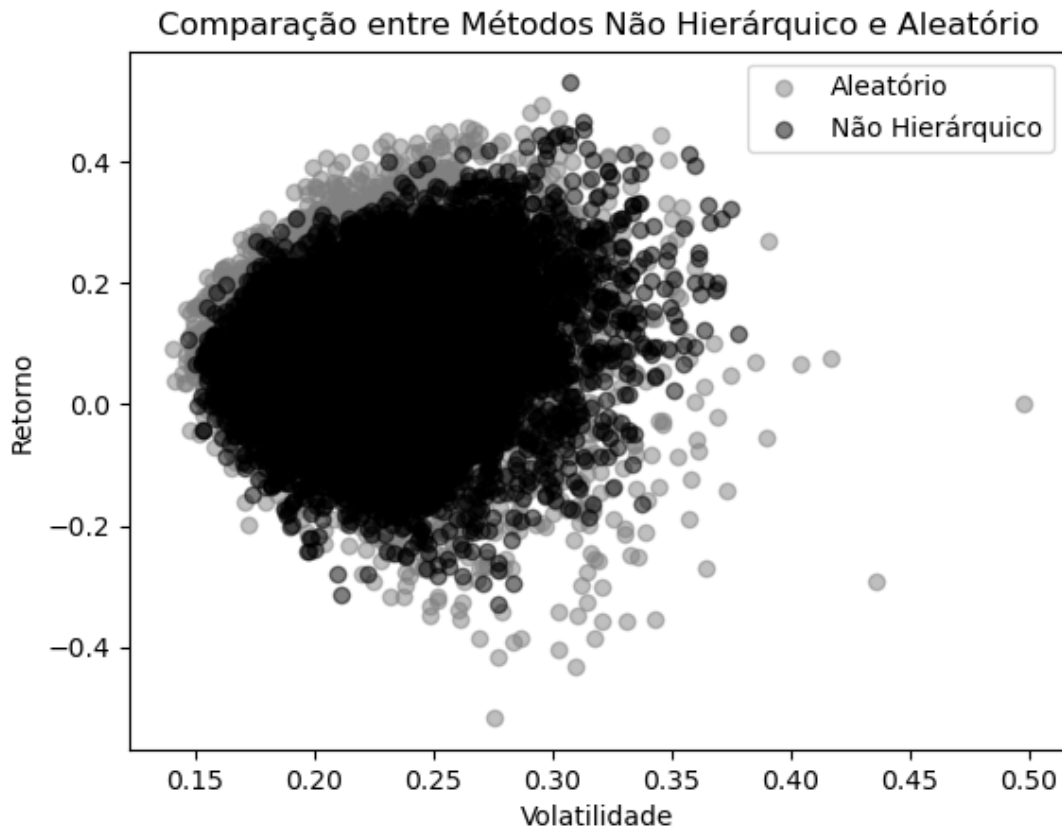
Comparando a formação desse período com o analisado anteriormente, de três meses, pode ser visto que se assemelharam em relação ao comportamento dos ativos, sugerindo estabilidade na sua distribuição. É possível observar que no segundo *cluster*, dois dos ativos do período anterior permanecem no mesmo grupo, sendo eles o 'IRBR3.SA' e 'VALE3.SA'.

Nos tópicos seguintes, serão mostrados os diagramas correspondentes para cada métrica avaliada para a otimização do portfólio.

4.3.2.1 Comparativo menor volatilidade

A figura 39 apresenta o resultado comparativo para a otimização da menor volatilidade.

Figura 39 – Comparativo menor volatilidade com 6 grupos - 1 ano



É possível observar diferenças na amplitude das comparações, mostradas abaixo:

Tabela 38 – Amplitude gráfica - Menor volatilidade

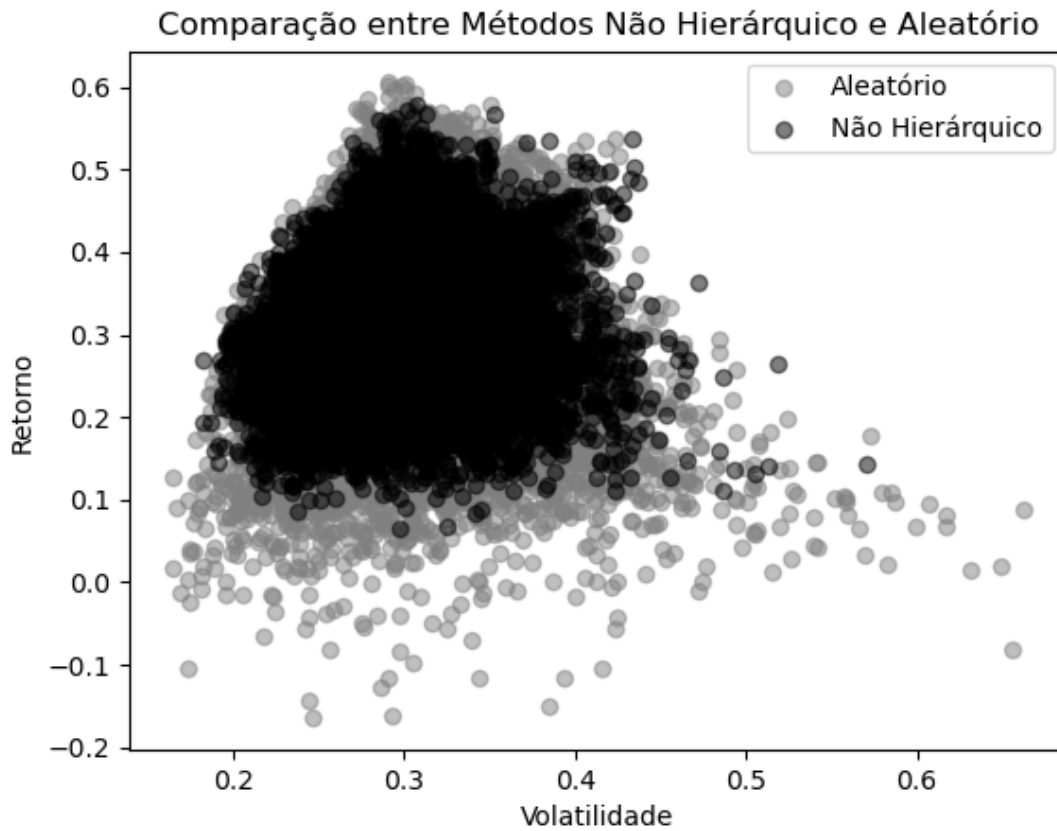
Amplitude	Agrupamento	Aleatório
Retorno	0.859936	1.011699
Volatilidade	0.230970	0.358206

Fonte: Autor (2024)

4.3.2.2 Comparativo maior retorno

A seguir, o diagrama contém informações referentes a maximização dos retornos da carteira.

Figura 40 – Comparativo maior retorno com 6 grupos - 1 ano



Fonte: Autor (2024)

A tabela abaixo oferece uma comparação acerca da sua amplitude.

Tabela 39 – Amplitude gráfica - Maior retorno

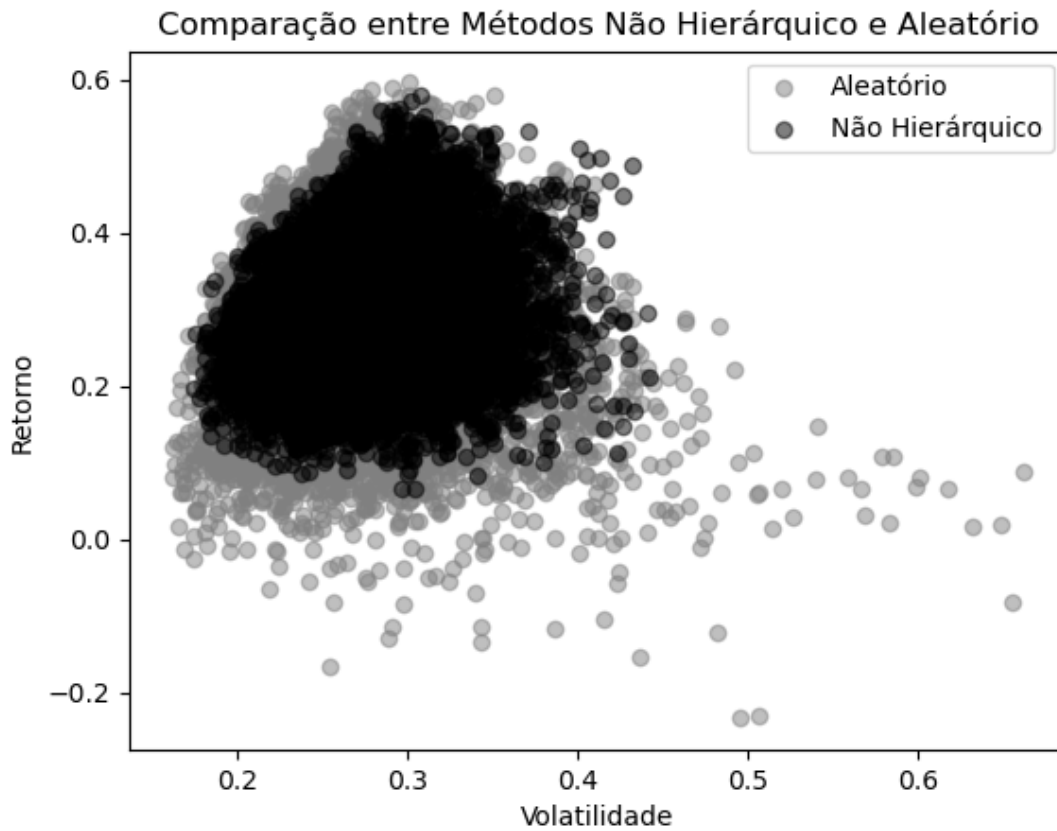
Amplitude	Agrupamento	Aleatório
Retorno	0.513088	0.769822
Volatilidade	0.388410	0.498537

Fonte: Autor (2024)

4.3.2.3 Comparativo maior índice Sharpe

Neste comparativo, foi explorada a distribuição dos dados acerca do índice Sharpe.

Figura 41 – Comparativo maior índice Sharpe com 6 grupos - 1 ano



Fonte: Autor (2024)

A tabela 40 destaca a diferença entre as amplitudes.

Tabela 40 – Amplitude gráfica - Maior índice Sharpe

Amplitude	Agrupamento	Aleatório
Retorno	0.513088	0.829761
Volatilidade	0.268350	0.501319

Fonte: Autor (2024)

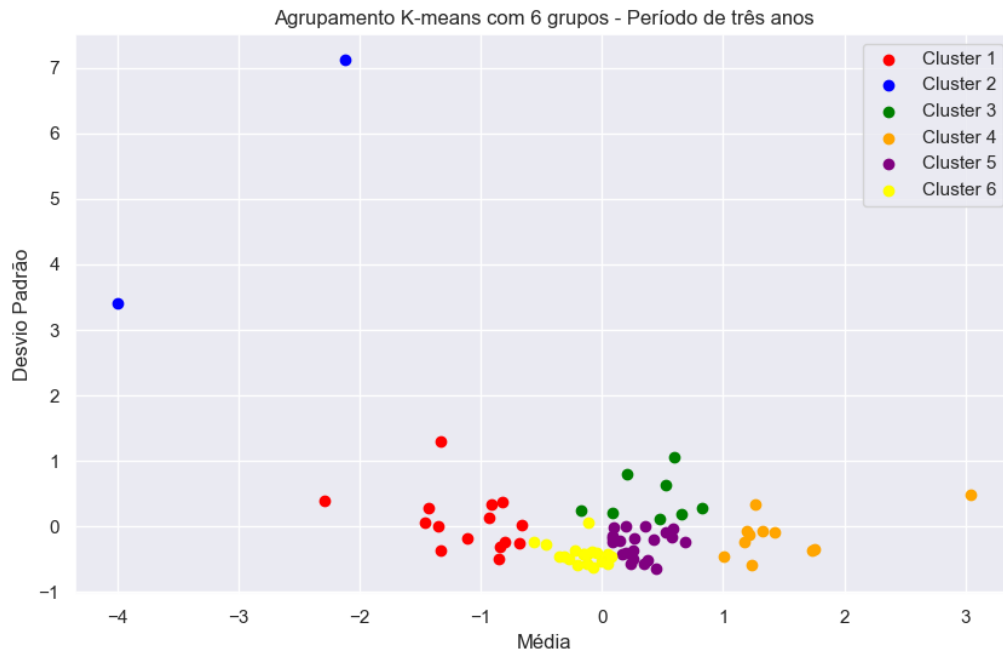
Os resultados anteriores ressaltam uma redução significativa na dispersão dos dados, reafirmando a funcionalidade da aplicação do agrupamento em relação ao método tradicional de Markowitz.

4.3.3 Observação 3: período de três anos

Nessa sessão será apresentada uma análise detalhada da aplicação do agrupamento não hierárquico com seis grupos para o período de três anos. Inicialmente o diagrama é

representado na figura 42.

Figura 42 – Diagrama de dispersão - três anos



Fonte: Autor (2024)

A seguir será destacado a quantidade de elementos e quais elementos compõem os grupos dessa janela temporal analisada.

Tabela 41 – Elementos por grupo - três anos

Grupo	Ativos
Grupo 1	15
Grupo 2	2
Grupo 3	8
Grupo 4	11
Grupo 5	21
Grupo 6	19

Fonte: Autor (2023)

- Grupo 1: 'BRFS3.SA', 'COGN3.SA', 'CRFB3.SA', 'CVCB3.SA', 'ECOR3.SA', 'EZTC3.SA', 'FLRY3.SA', 'HAPV3.SA', 'JBSS3.SA', 'LREN3.SA', 'MGLU3.SA', 'MRFG3.SA', 'NTCO3.SA', 'QUAL3.SA', 'YDUQ3.SA'
- Grupo 2: 'AMER3.SA', 'IRBR3.SA'
- Grupo 3: 'AZUL4.SA', 'BRKM5.SA', 'CSNA3.SA', 'ELET6.SA', 'RENT3.SA', 'SBSP3.SA', 'SUZB3.SA', 'VALE3.SA'

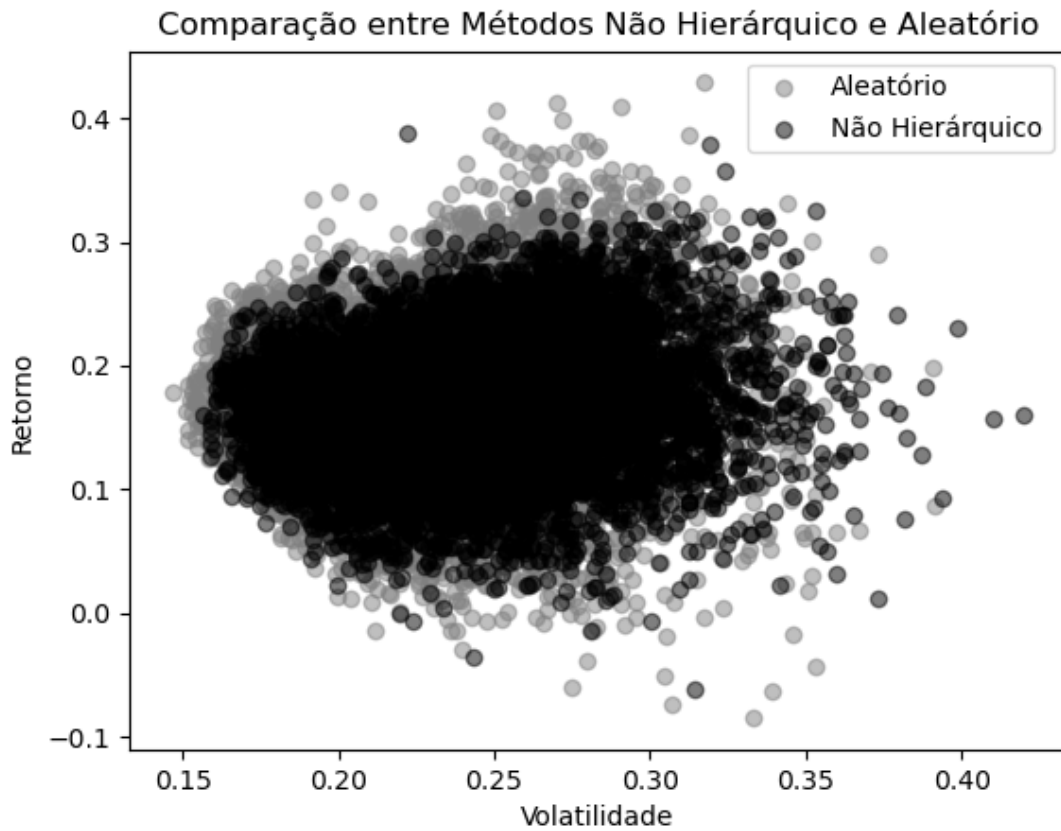
-
4. Grupo 4: 'BPAC11.SA', 'CMIG4.SA', 'EMBR3.SA', 'GGBR4.SA', 'LEVE3.SA', 'PCAR3.SA', 'PETR4.SA', 'PRIO3.SA', 'PTBL3.SA', 'SHUL4.SA', 'UGPA3.SA'
 5. Grupo 5: 'AGRO3.SA', 'BBAS3.SA', 'BBSE3.SA', 'CIEL3.SA', 'CPFE3.SA', 'CSMG3.SA', 'CYRE3.SA', 'EQTL3.SA', 'HYPE3.SA', 'JHSF3.SA', 'MDIA3.SA', 'MULT3.SA', 'POMO4.SA', 'RAPT4.SA', 'SANB11.SA', 'SLCE3.SA', 'TAEE11.SA', 'TOTS3.SA', 'TRPL4.SA', 'USIM5.SA', 'WEGE3.SA'
 6. Grupo 6: 'ABEV3.SA', 'B3SA3.SA', 'BBDC4.SA', 'BEEF3.SA', 'CCRO3.SA', 'CSAN3.SA', 'EGIE3.SA', 'EVEN3.SA', 'GOLL4.SA', 'GRND3.SA', 'ITSA4.SA', 'JSLG3.SA', 'KLBN11.SA', 'MRVE3.SA', 'PSSA3.SA', 'RADL3.SA', 'RAIL3.SA', 'TIMS3.SA', 'VIVT3.SA'

É possível ver que um ativo evidenciado anteriormente permaneceu no mesmo *cluster*, o segundo, além de que os grupos foram gradualmente se juntando ao decorrer dos anos.

4.3.3.1 Comparativo menor volatilidade

Nesta análise, o diagrama apresenta a distribuição dos ativos após a otimização da carteira para a menor volatilidade.

Figura 43 – Comparativo menor volatilidade com 6 grupos - 3 anos



Fonte: Autor (2024)

A tabela 42 esclarece a amplitude referente ao gráfico anterior.

Tabela 42 – Amplitude gráfica - Menor volatilidade

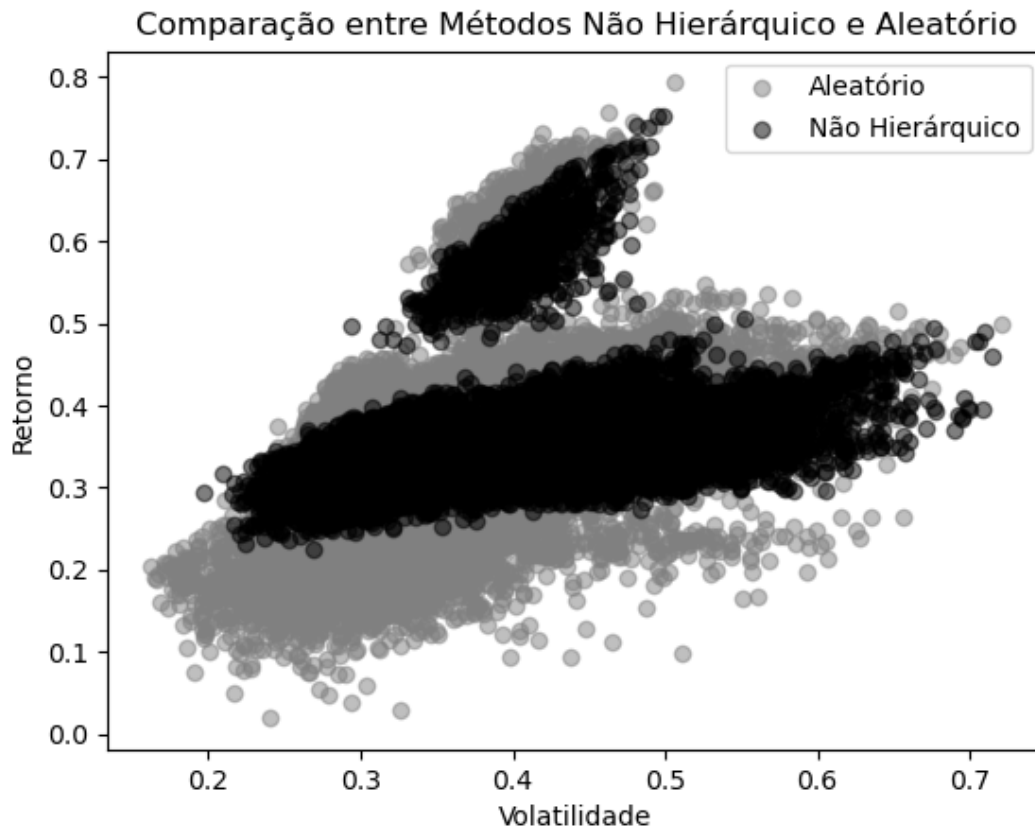
Amplitude	Agrupamento	Aleatório
Retorno	0.449224	0.513307
Volatilidade	0.263425	0.244934

Fonte: Autor (2024)

4.3.3.2 Comparativo maior retorno

A solução do diagrama seguinte demonstra a alocação das carteiras ótimas para o maior retorno.

Figura 44 – Comparativo maior retorno com 6 grupos - 3 anos



Fonte: Autor (2024)

Os resultados somados à amplitude gráfica, revelam uma alocação mais concentrada das carteiras nesse período.

Tabela 43 – Amplitude gráfica - Maior retorno

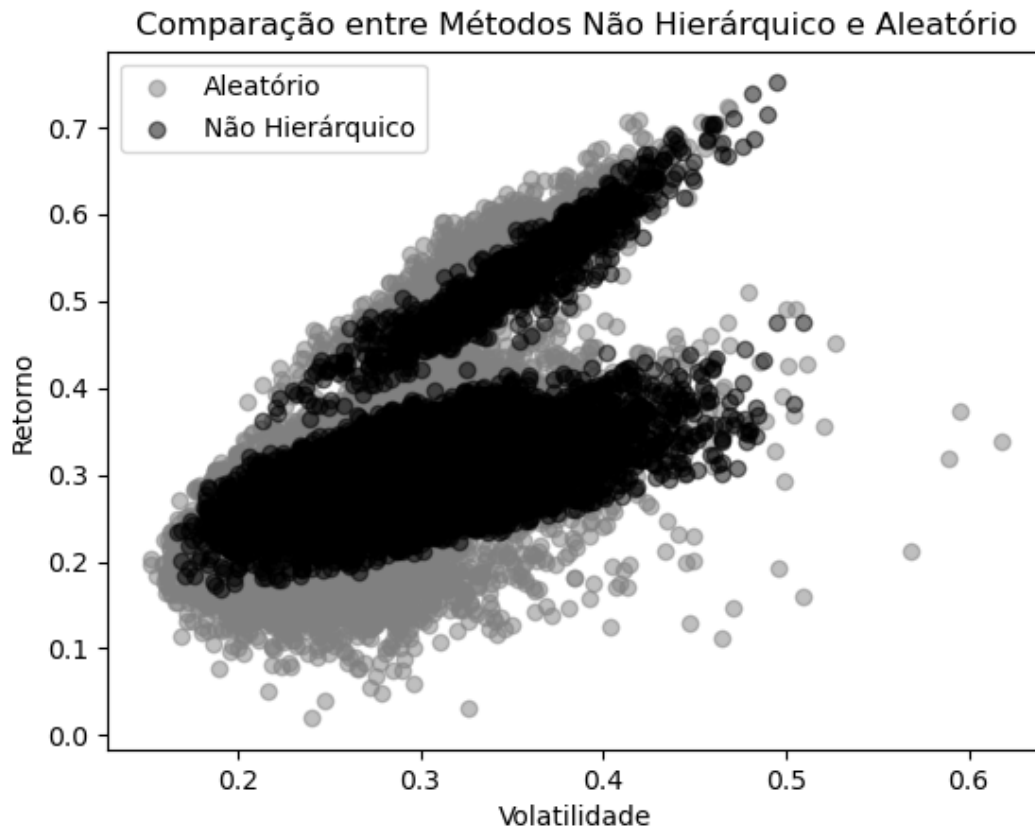
Amplitude	Agrupamento	Aleatório
Retorno	0.527070	0.773283
Volatilidade	0.516728	0.560026

Fonte: Autor (2024)

4.3.3.3 Comparativo maior índice Sharpe

O comparativo seguinte ilustra os resultados para otimização do índice Sharpe.

Figura 45 – Comparativo maior retorno com 6 grupos - 3 anos



Fonte: Autor (2024)

Segue a tabela com uma comparação mais detalhada da amplitude gráfica entre os métodos.

Tabela 44 – Amplitude gráfica - Maior índice Sharpe

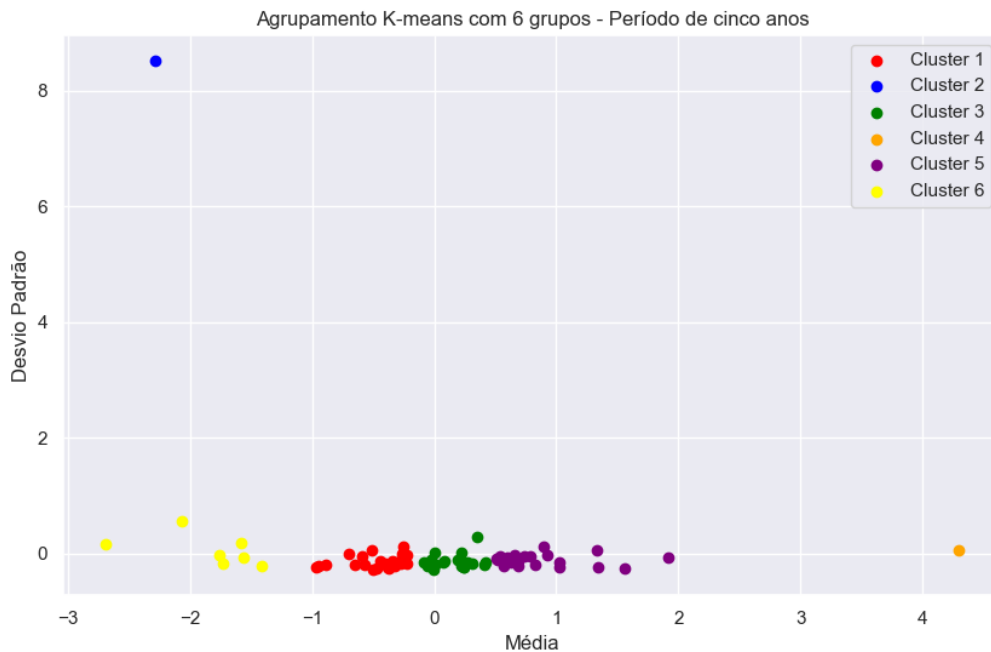
Amplitude	Agrupamento	Aleatório
Retorno	0.583660	0.704775
Volatilidade	0.343021	0.466543

Fonte: Autor (2024)

Mesmo com valores diferentes, é possível identificar os mesmos padrões vistos nos comparativos anteriores. Tanto pelas figuras, quanto pelas tabelas, é visível a funcionalidade da aplicação desse novo modelo para gestão de carteiras ótimas.

4.3.4 Observação 4: período de cinco anos

Para essa janela temporal, observa-se a formação dos seis grupos.

Figura 46 – Diagrama de dispersão - cinco anos

Fonte: Autor (2024)

A tabela 45 mostra a quantidade de ativos de cada grupo, seguido de quais são eles.

Tabela 45 – Elementos por grupo - cinco anos

Grupo	Ativos
Grupo 1	24
Grupo 2	1
Grupo 3	21
Grupo 4	1
Grupo 5	22
Grupo 6	7

Fonte: Autor (2023)

- Grupo 1: 'ABEV3.SA', 'AZUL4.SA', 'BBDC4.SA', 'BRKM5.SA', 'CRFB3.SA', 'ECOR3.SA', 'EMBR3.SA', 'FLRY3.SA', 'GOLL4.SA', 'GRND3.SA', 'HAPV3.SA', 'ITSA4.SA', 'JSLG3.SA', 'KLBN11.SA', 'LREN3.SA', 'MDIA3.SA', 'MRVE3.SA', 'PSSA3.SA', 'SANB11.SA', 'SUZB3.SA', 'TIMS3.SA', 'USIM5.SA', 'VIVT3.SA', 'YDUQ3.SA'
- Grupo 2: 'IRBR3.SA'

3. Grupo 3: 'B3SA3.SA', 'BBAS3.SA', 'BBSE3.SA', 'BEEF3.SA', 'CCRO3.SA', 'CPFE3.SA', 'CSAN3.SA', 'CSMG3.SA', 'EGIE3.SA', 'EQTL3.SA', 'EZTC3.SA', 'HYPE3.SA', 'MRFG3.SA', 'MULT3.SA', 'POMO4.SA', 'RADL3.SA', 'RAIL3.SA', 'RAPT4.SA', 'SBSP3.SA', 'TRPL4.SA', 'VALE3.SA'
4. Grupo 4: 'PRIO3.SA'
5. Grupo 5: 'AGRO3.SA', 'BPAC11.SA', 'CMIG4.SA', 'CSNA3.SA', 'CYRE3.SA', 'ELET6.SA', 'EVEN3.SA', 'GGBR4.SA', 'JBSS3.SA', 'JHSF3.SA', 'LEVE3.SA', 'MGLU3.SA', 'PCAR3.SA', 'PETR4.SA', 'PTBL3.SA', 'RENT3.SA', 'SHUL4.SA', 'SLCE3.SA', 'TAEE11.SA', 'TOTS3.SA', 'UGPA3.SA', 'WEGE3.SA'
6. Grupo 6: 'AMER3.SA', 'BRFS3.SA', 'CIEL3.SA', 'COGN3.SA', 'CVCB3.SA', 'NTCO3.SA', 'QUAL3.SA'

Na janela de cinco anos nota-se que 'IRBR3.SA', ativo citado anteriormente, continuou mostrando um comportamento destoante dos demais, ficando sozinho no grupo dois representado em azul.

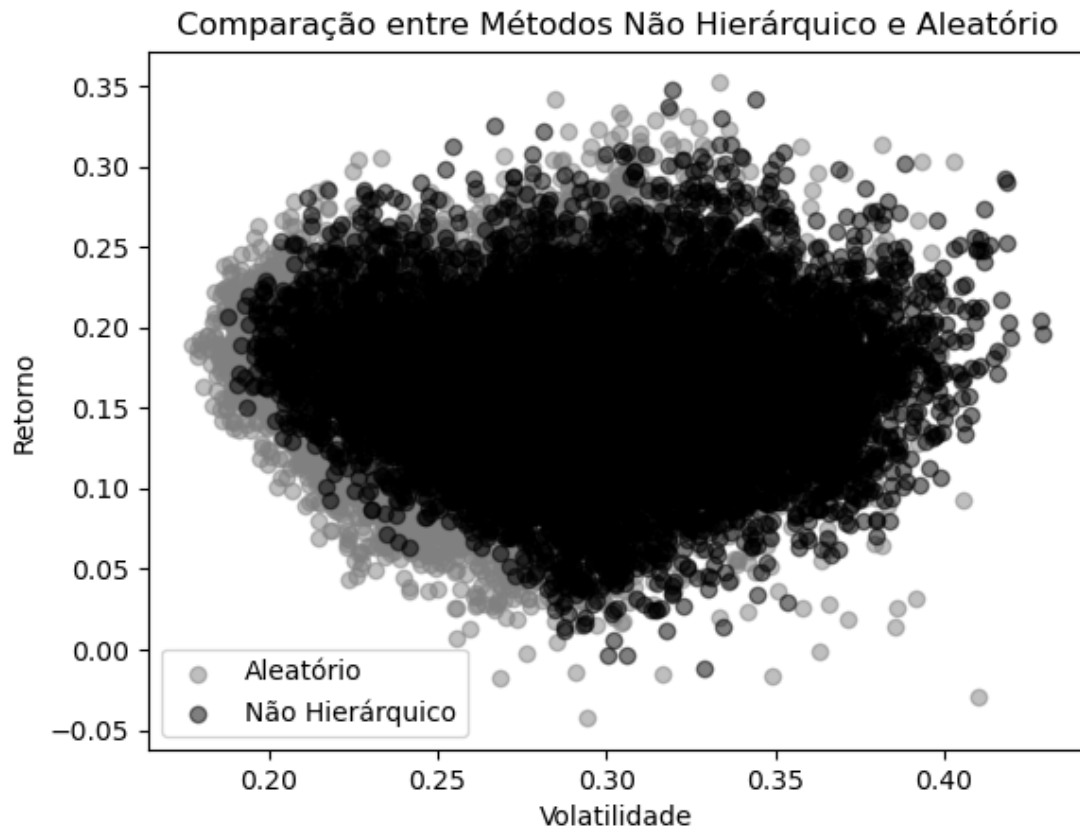
Além disso, a tendência de convergência das médias do preço de fechamento ajustado dos ativos também indica uma estabilidade ao passar do tempo, valores indicados com o desvio padrão próximo a zero no diagrama de dispersão mostrado anteriormente. Essa homogeneidade apresentada pode ser uma informação útil para os investidores que buscam indentificar padrões de comportamento de mercado.

Por fim, os gráficos de dispersão para essa análise serão apresentados, gráficos de grande importância para auxiliar a identificar oportunidades de mercado.

4.3.4.1 Comparativo menor volatilidade

O comparativo a seguir evidencia a distribuição dos dados para a otimização das carteiras de menor volatilidade.

Figura 47 – Comparativo menor volatilidade com 6 grupos - 5 anos



Fonte: Autor (2024)

Para complementar a interpretação, têm-se:

Tabela 46 – Amplitude gráfica - Menor volatilidade

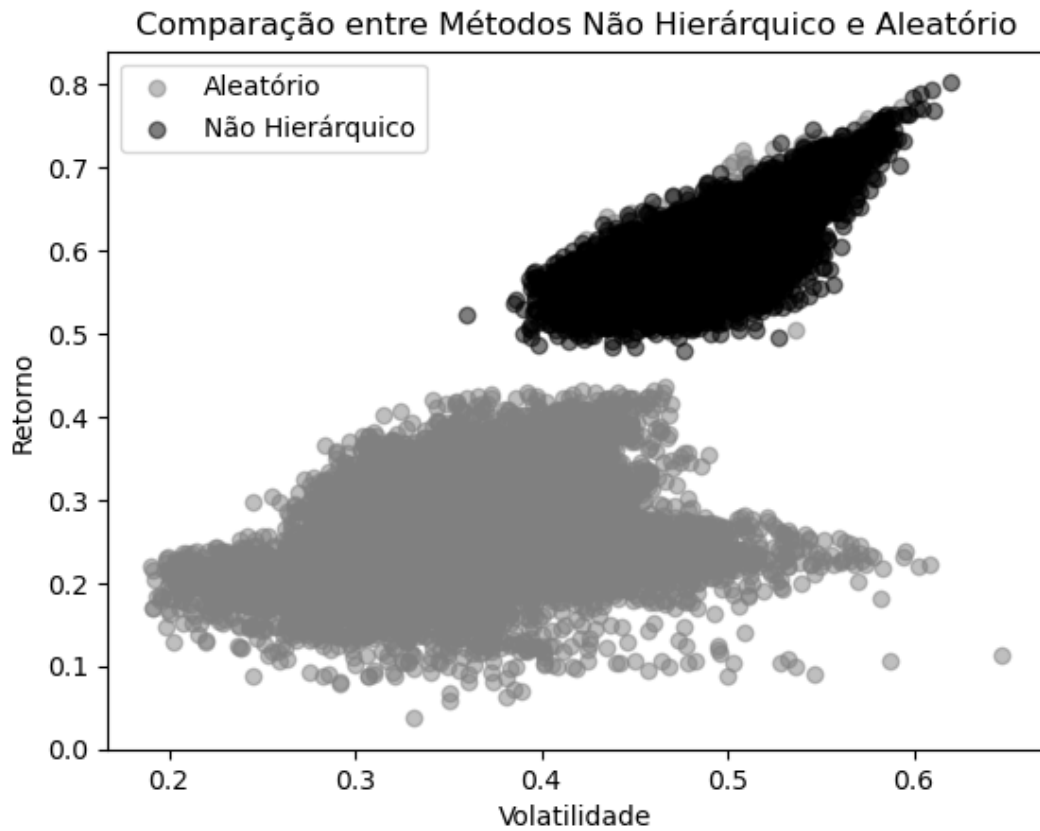
Amplitude	Agrupamento	Aleatório
Retorno	0.359281	0.395062
Volatilidade	0.241443	0.239976

Fonte: Autor (2024)

4.3.4.2 Comparativo maior retorno

Esta sessão identifica as carteiras de maior retorno ao longo de cinco anos.

Figura 48 – Comparativo maior retorno com 6 grupos - 5 anos



Fonte: Autor (2024)

Além do diagrama, a tabela 47 auxilia na interpretação, onde os resultados indicam o impacto positivo das carteiras apresentadas em preto.

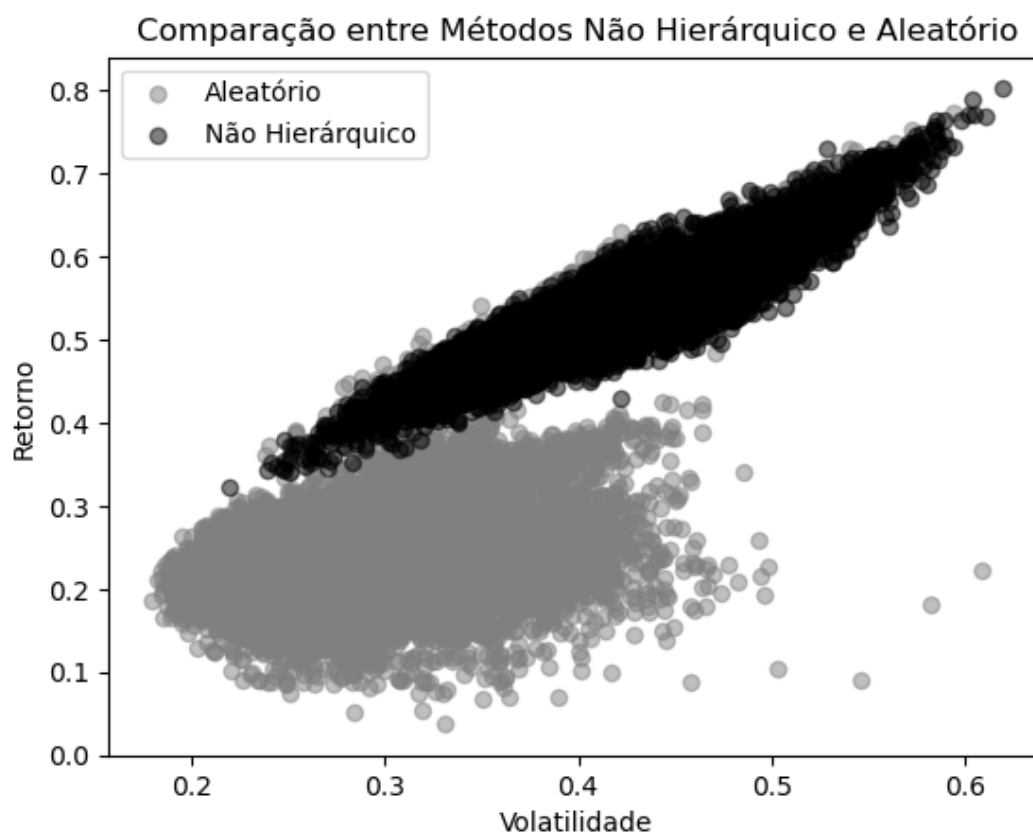
Tabela 47 – Amplitude gráfica - Maior retorno

Amplitude	Agrupamento	Aleatório
Retorno	0.321799	0.736154
Volatilidade	0.260330	0.457711

Fonte: Autor (2024)

4.3.4.3 Comparativo maior índice Sharpe

A figura 49 mostra o resultado da carteira de otimização do índice Sharpe seguida de sua amplitude gráfica, apresentada na tabela 48.

Figura 49 – Comparativo maior índice Sharpe com 6 grupos - 5 anos

Fonte: Autor (2024)

Tabela 48 – Amplitude gráfica - Maior índice Sharpe

Amplitude	Agrupamento	Aleatório
Retorno	0.479756	0.736154
Volatilidade	0.400395	0.429803

Fonte: Autor (2024)

De modo geral, ao comparar ambas amplitudes gráficas dessas aplicações, foi observada uma redução consistente na dispersão dos dados, comparado a aplicação da técnica tradicional de Markowitz.

Ao longo desse estudo, as carteiras apresentadas graficamente demonstraram uma alocação estratégica dos ativos. Ao comparar as análises foi possível perceber que os resultados foram promissores, revelando estruturas mais concentradas em todos os quatro períodos avaliados. Além disso, apresentou-se uma incerteza reduzida para a avaliação com o método de agrupamento.

Ao comparar os resultados dos agrupamentos, foram observadas diferenças na estrutura e na distribuição dos grupos formados ao longo do tempo. Porém, com base nesta

análise comparativa, foi possível ver que ambos métodos de agrupamentos têm influências positivas em relação a otimização da carteira de Markowitz, onde mantiveram padrões nas estruturas de otimização formadas, mostrando maior eficiência e estabilidade.

Esses resultados reforçam a eficiência e utilidade de agrupamentos na construção de carteiras eficientes de Markowitz.

5 Conclusão

No cenário atual, demarcado por crescentes avanços na tecnologia, a análise de dados pode ser uma tarefa árdua. A gestão de carteiras se encontra com um volume cada vez maior de informações financeiras, mostrando a necessidade de ferramentas sofisticadas para simplificar a análise e aumentar a eficiência dos portfólios.

Ao utilizar métodos de agrupamentos é possível observar padrões e estruturas nos elementos, agrupando ativos com características semelhantes. Ao combinar essa metodologia com a teoria de Markowitz, a complexidade dos dados é simplificada, mas também otimiza a eficiência na alocação dos ativos.

Ao longo desse estudo foi possível observar que a combinação dessas técnicas fornece uma estrutura eficaz para identificar oportunidades de investimentos, evidenciando a capacidade do método de formar conjuntos de dados mais concentrados, mostrando também maior estabilidade.

Pontos positivos podem ser indicados, como a quantidade de grupos obtida utilizando o agrupamento hierárquico foi similar em todos as janelas de tempo avaliadas, apontando robustez na metodologia. Isso possibilitou o experimento com o método não hierárquico também, onde adicionando mais ativos, foi reafirmado o resultados ótimo das carteiras em comparação com as carteiras tradicionais de Markowitz.

Este estudo é uma proposta de uma nova metodologia para construção de portfólios ótimos. A abordagem integra técnicas de agrupamento como alternativa aprimorada que reduz a incerteza relacionada à composição de carteiras.

Para trabalhos futuros, uma questão interessante é verificar a técnica em outros contextos. Como a relação entre o tamanho do mercado avaliado e o número de grupos, pois é possível que a quantidade de grupos mude conforme o número de ações contidas no mercado analisado.

Por fim, a análise de agrupamentos possibilita uma nova forma de aplicar estratégias e organizar os dados financeiros, que adicionada à carteira eficiente de Markowitz pode contribuir positivamente acerca do entendimento e gestão de portfólios.

Referências

- 1 COIMBRA, J. d. Á. A. et al. Considerações sobre a interdisciplinaridade. *Interdisciplinaridade em ciências ambientais*, Signus São Paulo, p. 52–70, 2000. Disponível em: [https://www.unievangelica.edu.br/files/images/Interdisciplinaridade%20e%20Ci%C3%A7ncias%20Ambientais%20\(3\).pdf#page=62](https://www.unievangelica.edu.br/files/images/Interdisciplinaridade%20e%20Ci%C3%A7ncias%20Ambientais%20(3).pdf#page=62). Acesso em: 01 out. 2023. 14
- 2 COSTA, L. A. N. d. et al. *Econofísica e a economia*. 2017. Monografia, Universidade Federal de Santa Catarina. 14
- 3 MINGOTI, S. A. *Análise de dados através de métodos estatísticos multivariada: uma abordagem aplicada*. [S.l.]: Editora UFMG, 2007. 295–295 p. 15, 16, 17, 22, 30
- 4 GUTH, G. D. et al. Métodos e técnicas para resolução de modelos de programação inteira de grande complexidade: uma revisão sistemática da literatura. *Revista de Gestão e Secretariado (Management and Administrative Professional Review)*, v. 14, n. 2, p. 2087–2101, 2023. 15, 23, 24
- 5 KARASAWA, M. et al. Aplicação de métodos de agrupamento na quantificação da divergência genética entre acessos de tomateiro. *Horticultura Brasileira*, SciELO Brasil, v. 23, p. 1000–1005, 2005. Disponível em: <https://www.scielo.br/j/hb/a/N4nFd879WHKzNPcnf3MXQbH/>. Acesso em: 07 out. 2023. 16
- 6 KANEGAE, Y. et al. Aplicação de agrupamento de dados para a identificação de perfis gerais entre clientes. In: *V Encontro de Iniciação Científica e Tecnológica-EnICT (ISSN: 2526-6772)*. [S.l.: s.n.], 2020. 16
- 7 DANIEL, T. et al. Physics-informed cluster analysis and a priori efficiency criterion for the construction of local reduced-order bases. *Journal of Computational Physics*, Elsevier, v. 458, p. 111120, 2022. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0021999122001826>. Acesso em: 02 nov. 2023. 16
- 8 MJAHEDE, M. The use of clustering techniques for the classification of high energy physics data. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 559, p. 199–202, 04 2006. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0168900205022692>. Acesso em: 02 nov. 2023. 16
- 9 NEUKART, F.; DOLLEN, D. V.; SEIDEL, C. Quantum-assisted cluster analysis on a quantum annealing device. *Frontiers in Physics*, Frontiers Media SA, v. 6, p. 55, 2018. Disponível em: <https://www.frontiersin.org/journals/physics/articles/10.3389/fphy.2018.00055/full>. Acesso em: 02 nov. 2023. 16
- 10 FENG, Y. et al. Hierarchical clustering analysis of hydrogen bond networks in aqueous solutions. *Physical Chemistry Chemical Physics*, Royal Society of Chemistry, v. 24, n. 16, p. 9707–9717, 2022. 16

-
- 11 LEÓN, D. et al. Clustering algorithms for risk-adjusted portfolio construction. *Procedia Computer Science*, Elsevier, v. 108, p. 1334–1343, 2017. Disponível em: [⟨https://www.sciencedirect.com/science/article/pii/S187705091730772X⟩](https://www.sciencedirect.com/science/article/pii/S187705091730772X). Acesso em: 14 out. 2023. 16
- 12 GUBU, L.; ROSADI, D. et al. Robust mean–variance portfolio selection using cluster analysis: a comparison between kamila and weighted k-mean clustering. *Asian Economic and Financial Review*, Asian Economic and Social Society, v. 10, n. 10, p. 1169, 2020. Disponível em: [⟨https://d1wqtxts1xzle7.cloudfront.net/88508379/AEFR202010101169-1186-libre.pdf?1657637804=&response-content-disposition=inline%3B+filename%3DRobust_Mean_Variance_Portfolio_Selection.pdf&Expires=1718843230&Signature=Yxw-4SHUfMkW3sF7k9H-70UgBEGqjVjsbgYyoWXijJRjL1xvKbebJqZOImOmej7sFDiQQLw0aT481-&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA⟩](https://d1wqtxts1xzle7.cloudfront.net/88508379/AEFR202010101169-1186-libre.pdf?1657637804=&response-content-disposition=inline%3B+filename%3DRobust_Mean_Variance_Portfolio_Selection.pdf&Expires=1718843230&Signature=Yxw-4SHUfMkW3sF7k9H-70UgBEGqjVjsbgYyoWXijJRjL1xvKbebJqZOImOmej7sFDiQQLw0aT481-&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA). Acesso em: 14 out. 2023. 16
- 13 METZ, J. *Interpretação de clusters gerados por algoritmos de clustering hierárquico*. Tese (Doutorado) — Universidade de São Paulo, 2006. Disponível em: [⟨https://www.teses.usp.br/teses/disponiveis/55/55134/tde-14092006-090701/en.php⟩](https://www.teses.usp.br/teses/disponiveis/55/55134/tde-14092006-090701/en.php). Acesso em: 07 out. 2023. 18
- 14 GOWER, J. C. Properties of euclidean and non-euclidean distance matrices. *Linear algebra and its applications*, Elsevier, v. 67, p. 81–97, 1985. Disponível em: [⟨https://www.sciencedirect.com/science/article/pii/0024379585901879⟩](https://www.sciencedirect.com/science/article/pii/0024379585901879). Acesso em: 01 out. 2023. 19, 20
- 15 FERREIRA, D. F. *Análise multivariada*, 1996. Lavras: Editora da UFLA. Disponível em: [⟨https://docs.ufpr.br/~niveam/micro%20da%20sala/Aulas%20-%20Internet/multivariada/multivariada.pdf⟩](https://docs.ufpr.br/~niveam/micro%20da%20sala/Aulas%20-%20Internet/multivariada/multivariada.pdf). Acesso em: 07 out. 2023. 21, 22, 23
- 16 MACQUEEN, J. et al. Some methods for classification and analysis of multivariate observations. In: OAKLAND, CA, USA. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. [S.l.], 1967. v. 1, n. 14, p. 281–297. 23
- 17 DAMASCENO, F. F. R. *Uma melhoria no algoritmo K-médias utilizando o estimador de James-Stein*. Dissertação (Mestrado) — Universidade Federal do Ceará, 2016. Disponível em: [⟨https://repositorio.ufc.br/handle/riufc/21491⟩](https://repositorio.ufc.br/handle/riufc/21491). Acesso em: 10 jan. 2024. 23
- 18 LIMA, I. M. B. D. *Uso de algoritmo RO dados categ.* Tese (Doutorado) — Universidade Federal da Paraíba, 2013. Disponível em: [⟨http://www.de.ufpb.br/graduacao/tcc/TCC2013Isis.pdf⟩](http://www.de.ufpb.br/graduacao/tcc/TCC2013Isis.pdf). Acesso em: 07 out. 2023. 23
- 19 PALMA, L. Agrupamento de dados: K-médias. *Universidade Federal do Recôncavo da Bahia Centro de Ciências Exatas e Tecnológicas*, p. 27, 2018. Disponível em: [⟨https://www2.ufrb.edu.br/bcet/components/com_chronoforms5/chronoforms/uploads/tcc/20190604200511_2018_2_TCC_Luann_Farias_Palma-_Agrupamento_de_dados_-_K_medias.pdf⟩](https://www2.ufrb.edu.br/bcet/components/com_chronoforms5/chronoforms/uploads/tcc/20190604200511_2018_2_TCC_Luann_Farias_Palma-_Agrupamento_de_dados_-_K_medias.pdf). Acesso em: 10 jan. 2024. 23
- 20 MARKOWITZ, H. M. Foundations of portfolio theory. *The journal of finance*, JSTOR, v. 46, n. 2, p. 469–477, 1991. 24

-
- 21 SHARPE, W. F. A simplified model for portfolio analysis. *Management science*, INFORMS, v. 9, n. 2, p. 277–293, 1963. 24, 28
- 22 ROLL, R.; ROSS, S. A. An empirical investigation of the arbitrage pricing theory. *The journal of finance*, Wiley Online Library, v. 35, n. 5, p. 1073–1103, 1980. 24
- 23 NETO, A. A. Mercado financeiro 8. ed. *São Paulo: Atlas*, 2008. 26
- 24 SANTOS, D. R. et al. Escolha de carteira de investimento: aplicação no mercado financeiro brasileiro. *Brazilian Applied Science Review*, v. 5, n. 1, p. 568–583, 2021. Disponível em: <https://ojs.brazilianjournals.com.br/ojs/index.php/BASR/article/view/25077>. Acesso em: 07 out. 2023. 27
- 25 HAN, J.; PEI, J.; TONG, H. *Data mining: concepts and techniques*. [S.l.]: Morgan kaufmann, 2022. 29
- 26 Python DS. *Python machine learning (10): pré-processamento - normalização*. 2022. Disponível em: <https://www.youtube.com/watch?v=kRoKTqDAiHs>. Acesso em: 05 jun. 2023. 30
- 27 PAIXÃO, J. L. da et al. Métodos matemáticos de modelagem e otimização: teoria e aplicações do método de monte carlo. *Salão do Conhecimento*, v. 7, n. 7, 2021. 30
- 28 ABENSUR, E. O. Avaliação de risco de crédito com base no modelo mdb, na teoria de opções reais e na simulação de monte carlo. *Revista de Economia e Administração*, v. 9, n. 2, 2010. 30
- 29 Brenno Sullivan - VAROS Quant. *Como fazer a otimização de Markowitz e a fronteira eficiente com o python — python e finanças*. 2022. Disponível em: <https://www.youtube.com/watch?v=BchQuTJvRAs>. Acesso em: 06 jun. 2023. 30
- 30 GUEDES, T. A. et al. Estatística descritiva. *Projeto de ensino aprender fazendo estatística*, Universidade Estadual de Maringá Maringá, p. 1–49, 2005. Disponível em: https://www.ime.usp.br/~rvicente/Guedes_et al_Estatistica_Descritiva.pdf. Acesso em: 01 out. 2023. 41