UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS SOCIAIS APLICADAS
DEPARTAMENTO DE CIÊNCIAS CONTÁBEIS E ATUARIAIS
CURSO DE CIÊNCIAS ATUARIAIS

RAFAEL FIGUEIREDO RIBEIRO

Agrupamento de Ações na Bovespa com base no Excesso de Perda Média Empírica

Recife

2023

RAFAEL FIGUEIREDO RIBEIRO

Agrupamento de Ações na Bovespa com base no Excesso de Perda Média Empírica

Trabalho de Conclusão de Curso apresentado à Coordenação do Curso de Ciências Atuariais do Campus Recife da Universidade Federal de Pernambuco, na modalidade de monografia, como requisito parcial para obtenção do grau de bacharel em Ciências Atuariais.

Orientador (a): Prof. Dr. Wilton Bernardino da Silva

Recife

2023

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Ribeiro, Rafael Figueiredo.
Agrupamento de Ações na Bovespa com base no Excesso de Perda Média Empírica / Rafael Figueiredo Ribeiro. - Recife, 2023.
30 p. : il., tab.

Orientador(a): Wilton Bernardino da Silva
Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro de Ciências Sociais Aplicadas, Ciências Atuariais, 2023.

1. Bovespa. 2. K-Means. 3. Análise de Grupamento . 4. Análise de Cluster. 5. Excesso de Perda Média Empírico . I. Silva, Wilton Bernardino da. (Orientação). II. Título.

310 CDD (22.ed.)

RAFAEL FIGUEIREDO RIBEIRO

AGRUPAMENTO DE AÇÕES NA BOVESPA COM BASE NO EXCESSO DE PERDA MÉDIA EMPÍRICA

Trabalho de Conclusão de Curso apresentado à Coordenação do Curso de Ciências Atuariais do Campus Recife da Universidade Federal de Pernambuco, na modalidade de monografia, como requisito parcial para obtenção do grau de bacharel em Ciências Atuariais.

Aprovada em: 22/09/2023

BANCA EXAMINADORA

gov.br Documento assinado digitalmente
WILTON BERNARDINO DA SILVA
Data: 02/10/2023 18:22:44-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. WILTON BERNARDINO DA SILVA (Orientadora)
Universidade Federal de Pernambuco

gov.br Documento assinado digitalmente
ALESSANDRA PRAZERES CEZARIO
Data: 02/10/2023 21:45:21-0300
Verifique em https://validar.iti.gov.br

Profª. Drª. ALESSANDRA PRAZERES CESÁRIO
Universidade Federal de Pernambuco

gov.br Documento assinado digitalmente
RENATA GOMES ALCOFORADO
Data: 02/10/2023 19:07:23-0300
Verifique em https://validar.iti.gov.br

Profª. Drª. RENATA GOMES ALCOFORADO
Universidade Federal de Pernambuco

gov.br Documento assinado digitalmente
VINICIUS QUINTAS SOUTO MAIOR
Data: 02/10/2023 18:39:45-0300
Verifique em https://validar.iti.gov.br

Profª. Drª. VINÍCIUS QUINTAS SOUTO MAIOR
Universidade Federal de Pernambuco

RESUMO

Este estudo analisa o nível de risco de ações listadas na bolsa de valores brasileira (B3), utilizando para isso o excesso de perda média empírico calculado a partir da distribuição dos retornos diários. Cada ação é submetida a uma análise individual de risco. Em seguida, uma abordagem de análise de agrupamento K-Means é utilizada para agrupar automaticamente essas ações com base em seus níveis de risco. Todas as análises computacionais são conduzidas por meio da linguagem de programação R. Este estudo desempenha um papel fundamental na gestão de risco e no desenvolvimento de estratégias de investimento no mercado financeiro brasileiro, fornecendo informações essenciais para investidores e tomadores de decisão.

Palavras-chave: B3. risco financeiro. excesso de perda média empírico. análise de cluster. K-means.

ABSTRACT

This study analyzes the risk level of stocks listed on the Brazilian stock exchange (B3), using the empirical mean excess loss calculated from the distribution of daily returns. Each stock undergoes an individual risk analysis. Then, a K-Means clustering analysis approach is used to automatically group these stocks based on their risk levels. All computational analyses are conducted using the R programming language. This study plays a crucial role in risk management and the development of investment strategies in the Brazilian financial market, providing essential information for investors and decision-makers.

Keywords: B3. financial risk. mean excess loss. cluster analysis. K-means.

LISTA DE FIGURAS

Figura 1 – print do R script 19
Figura 2 – print do R script 19
Figura 3 – Grafico de Barras para Quantidade de ações por dia 25
Figura 4 – Gráfico de barras para quantidade de dias fora do estudo 26
Figura 5 – Gráfico de Barras quantidade de ativos para um determinado numero de transições 27
Figura 6 – Boxplot da Mediana dos Grupos 28

SUMÁRIO

1 INTRODUÇÃO 8

2 FUNDAMENTAÇÃO TEÓRICA 10

2.1 TEORIA DO VALOR EXTREMO E MEDIDAS DE RISCO 10

2.2 ANÁLISE DE GRUPAMENTO 13

2.3 K-MEANS 15

3 DADOS E MÉTODOS 16

3.1 DADOS 16

3.2 MÉTODO GARCH-VAR 16

3.3 A FUNÇÃO EMPÍRICA DO MEAN EXCESS LOSS 18

3.4 ANÁLISE DE GRUPAMENTOS 18

3.5 ALGORITMO HARTIGAN-WONG 19

4 RESULTADOS 23

5 CONCLUSÃO 29

REFERÊNCIAS 30

1 INTRODUÇÃO

As crises no mercado de capitais são eventos que ocorrem regularmente ao longo da história financeira global e, nas últimas décadas, têm se tornado ainda mais frequentes. Cabe observar as grandes crises que impactaram a economia mundial: a Crise Financeira Asiática (1997-1998), a Bolha da Internet (2000-2002), a Crise Financeira Global de 2008 e, mais recentemente, a Crise decorrente da Pandemia do Covid-19. Além dessas crises globais, no contexto brasileiro, também podemos destacar crises locais, como o Joesley Day e a Greve dos Caminhoneiros.

As crises financeiras frequentemente estão ligadas a eventos extremos que resultam em grandes perdas. Nesse contexto, a Teoria do Valor Extremo (EVT), *Extreme Value Theory*, surge como uma ferramenta estatística crucial e fornece a base teórica para as principais medidas de risco, como Valor em Risco (VaR), Valor em Risco Condicional (TVaR) e Excesso de Perda Média (MEL), *Mean Excess Loss*. Essas medidas de risco desempenham um papel essencial, uma vez que quantificam de forma numérica o conceito de risco, permitindo análises mais objetivas no que diz respeito à exposição financeira. Isso demonstra que a sua utilização no mercado financeiro pode conferir vantagens significativas tanto para investidores quanto para tomadores de decisão.

O objetivo principal deste trabalho consiste em desenvolver um sistema de classificação para uma lista de ações cotadas na bolsa de valores de São Paulo, com base em seus níveis de risco, os quais serão mensurados através da variável excesso de perda média empírica (*empirical mean excess loss*). A proposta é agrupar esses ativos em categorias distintas, de modo que aqueles pertencentes ao mesmo grupo apresentem níveis de risco próximos, enquanto os ativos em grupos diferentes exibam diferenças significativas em seus níveis de risco.

Para otimizar e automatizar o processo de classificação, será utilizado a técnica de análise de cluster, K-Means, a perda média em excesso será calculada diariamente para cada uma das ações selecionadas neste estudo, permitindo assim a criação diária dos grupos, que será realizada ao longo do segundo semestre de 2021. No âmbito deste trabalho, uma parte fundamental do estudo será destinada à análise da dinâmica da composição dos grupos formados ao longo do tempo. Isso incluirá a observação da movimentação dos ativos entre esses grupos, bem como a avaliação do tempo de permanência e do tamanho dos agrupamentos.

Além disso, para viabilizar o estudo, optamos por utilizar a linguagem de programação *R*, a qual possui uma vasta gama de pacotes e bibliotecas, que nos proporcionaram a capacidade de

realizar os processamentos de dados e cálculos computacionais de maneira eficiente e robusta.

Além da presente introdução, este trabalho conta com mais 4 capítulos numerados de 2 a 5: 2 - Fundamentação Teórica, 3 - Metodologia, 4 - Resultados e 5 - Conclusão. Na Fundamentação Teórica (capítulo 2), serão introduzidos os conceitos da Teoria dos Valores Extremos (EVT), tais como Value at Risk (VaR), Tail Value at Risk (TVar) e Excesso de Perda Média (MEL). Esse mesmo capítulo abordará também a Análise de Agrupamento e o K-Means. A Metodologia (capítulo 3) apresentará a expressão do Excesso de Perda Média Empírico $e^{EMP}(u)$ e explicará o algoritmo Hartingan-Wong para o K-Means. Além disso, será explicado como os ativos financeiros foram selecionados por meio do backtesting do VaR-GARCH. Os Resultados (capítulo 4) se limitarão a uma análise expositiva do comportamento das ações entre os grupos formados. Por fim, a Conclusão (capítulo 5) descreverá as conclusões e discussões finais do estudo.

2 FUNDAMENTAÇÃO TEÓRICA

O presente capítulo tem como objetivo apresentar de forma concisa a Teoria dos Valores Extremos, abordando as medidas de risco VaR, TVaR e Excesso de Perda Média. Além disso, será discutido como essas medidas estão relacionadas e suas propriedades. Também será introduzida a Análise de Agrupamento, explorando os diversos tipos de agrupamentos que podem ser formados. Ao final, será apresentada a ideia central da técnica K-means.

2.1 TEORIA DO VALOR EXTREMO E MEDIDAS DE RISCO

A Teoria dos Valores Extremos, (*Extreme Value Theory* - EVT), é uma área da estatística que tem como foco principal entender as propriedades dos valores situados na cauda da distribuição de probabilidade, ou seja, se concentra no estudo de eventos raros e extremos, os chamados cisnes negros (SINGH; ALLEN; ROBERT, 2013). A aplicação da Teoria dos Valores Extremos abrange uma ampla gama de domínios. Destaca-se, contudo, o seu emprego recorrente no âmbito do mercado financeiro, contribuindo para a tomada de decisões relacionadas ao gerenciamento de riscos.

O *Value at Risk* (VaR) é amplamente reconhecido e utilizado em trabalhos acadêmicos que abordam o tema Finanças Empíricas, devido à sua relativa simplicidade. De fato, o VaR é a métrica mais conhecida, no contexto de risco de mercado, e serve como base para a conceituação de outra medida a ser mencionada posteriormente, o TVaR.

Embora a EVT e o VaR já existissem em períodos anteriores, foi somente no final do século XX que seu uso ganhou considerável expressividade, visto que nesse período a economia mundial passou por algumas crises, à exemplo a crise de 1987 e a crise financeira asiática (1997-1998). Outro marco importante frequentemente citado em trabalhos acadêmicos é a metodologia para cálculo do VaR elaborada em 1994 pelo banco norte-americano JP Morgan, que popularizou a análise do VaR no mercado financeiro. Ainda na década de 1990, a medida de risco VaR foi reconhecida pelo Comitê de Supervisão Bancária da Basiléia, que autorizou o seu uso pelos bancos para embasar os cálculos de Capital Regulatório (HOLTON, 2002).

A definição do *Value at Risk* é simples, o VaR nada mais é do que um percentil da distribuição, ver equação (2.2), porém geralmente o VaR é introduzido em contexto de grandes perdas, piores resultados, onde o uso de variáveis de perda é comumente usado. Assim o VaR

pode ser interpretado como o montante de capital necessário para cobrir um percentual das perdas (KLUGMAN; PANJER; WILLMOT, 2019). Estatisticamente podemos expressar o VaR da seguinte forma:

$$VaR_p(X) = \inf_{x \geq 0}[x|F_X(x) \geq p], \quad 0 < p < 1. \tag{2.1}$$

Para distribuições contínuas, o VaR pode ser escrito de forma simplificada satisfazendo:

$$F_X[VaR_p(X)] = p \tag{2.2}$$

(ARTZNER et al., 1999) introduziram o conceito de medida de risco coerente, tais medidas obedecem a quatro propriedades que são desejadas quando se estuda risco, elas são:

1. **Subaditividade:** $\rho(X + Y) \leq \rho(X) + \rho(Y)$
2. **Monotonicidade:** se $X \leq Y$ para todos os possíveis valores, então $\rho(X) \leq \rho(Y)$
3. **Homogeneidade positiva;** para qualquer constante c, $\rho(cX) = c\rho(X)$
4. **Invariância na translação:** para qualquer constante c, $\rho(X + c) = \rho(X) + c$.

Subaditiva, a medida de risco da soma não pode ser superior à soma das medidas dos riscos, a obediência à subaditividade respeita a ideia de que a diversificação, em geral, diminui o risco. Monotonicidade, se um evento sempre apresenta maiores valores de perda do que um outro, a medida de risco deve ser sempre maior também. Homogeneidade positiva, a medida de risco independe da ocorrência, um exemplo prático: ao multiplicar por K a posição em uma ação, é esperado que o risco também seja multiplicado por K, visto que não houve diversificação. Invariância na translação, a adição de uma constante não adiciona incerteza, sendo assim o acréscimo ao risco equivale ao valor da constante (KLUGMAN; PANJER; WILLMOT, 2019).

(ARTZNER et al., 1999) também mostra que o VaR não é uma medida coerente, devido a não obediência ao critério da subaditividade. Além do VaR existe outra medida de risco, que contornar essa limitação da incoerência, medida essa denominada *Tail Value at Risk* (TVaR), que pode recebe outros nomes tais quais: *Conditional Value at Risk* (CVaR), *Conditional Tail Expectation* (CTE) e *Expected Shortfall* (ES).

TVaR no percentil p nada mais é do que a média de todos os valores de VaR com nível de segurança acima de p, onde 0<p<1.

$$TVaR_p(X) = \frac{\int_1^p VaR_u(X)\, du}{1 - p} \tag{2.3}$$

(KLUGMAN; PANJER; WILLMOT, 2019) expoem que no contexto de variáveis de perda o VaR é uma medida crítica, uma vez que o valor da variável aleatória ultrapassa o valor do VaR, não se pode mais honrar com as perdas. O TVaR é uma média dos piores resultados, O VaR que aparece na fórmula do TVaR, olhar equação 2.3, serve como uma referência para determinar tais piores resultados, assim caso o limiar definido pelo o VaR seja ultrapassado podemos ou não arcar com as perdas.

O TvaR pode ser expresso em termos do excedente de perda média Mean Excess Loss (MEL), conforme demonstrado em: (KLUGMAN; PANJER; WILLMOT, 2019). Caso X seja constante, o que temos na equação 2.2 é valido, e o TVaR pode ser escrito da seguinte forma:

$$TVaR_p(X) = E[X|X > \text{VaR}_p(X)] \tag{2.4}$$

Ainda podemos escrever o TVaR em função do Excedente de Perda Médio (MEL) aqui representado por (π_p).

$$\begin{aligned} TVaR_p(X) &= E(X \,|\, X > \pi_p) \\ &= \pi_p + E(X - \pi_p \,|\, X > \pi_p) \\ &= VaR_p(X) + e(\pi_p) \end{aligned} \tag{2.5}$$

Antes de apresentar o MEL, vamos primeiramente mostrar como o Excesso de Perda (EL), *Excess Loss*, é formulado.

$$eX(d) = e(d) = E(X - d \,|\, X > d) \tag{2.6}$$

O MEL nada mais é que o primeiro momento, a média do EL, os momentos de EL podem ser obtidos da seguinte forma:

$$e_X^k(d) = \frac{\int_d^\infty (x-d)^k f(x)\, dx}{1 - F(d)}, \text{ caso a variável seja contínua}$$

$$= \frac{\sum\limits_{x_j > d} (x_j - d)^k p(x_j)}{1 - F(d)}, \text{ caso a variável seja discreta} \tag{2.7}$$

O EL, pode ser definido como um variável transladada e truncada à esquerda. Truncada à esquerda porque os valores abaixo de $X+d$ não são considerados, e transladada porque d é subtraído de todos os valores remanescentes (KLUGMAN; PANJER; WILLMOT, 2019).

A perda média excedente é tal que, quando crescente em d a função distribuição de probabilidade (fdp) apresenta um cauda pesada, já quando a perda média excedente em d é decrescente fdp terá uma cauda leve (KLUGMAN; PANJER; WILLMOT, 2019).

A cauda da distribuição nada mais é do que a região onde os eventos extremos estão situados, a cauda esquerda corresponde a região onde a variável aleatória apresenta os menores valores, enquanto a cauda direita consiste na região onde a variável assume os maiores valores. Variáveis aleatórias que apresentam altas probabilidade em suas caudas recebem a denominação de variáveis com cauda pesada.

A métrica MEL, embora não seja uma medida de risco coerente, pode ser útil para avaliar a magnitude das nossas perdas mais significativas. Portanto, ela ainda representa uma ferramenta valiosa na análise das caudas de uma distribuição. Além disso, a análise do MEL é mais assertiva quando considerado um cenário de investimentos a longo prazo no mercado de capitais (ALBRECHT; MAURER; RUCKPAUL, 2001). Por esse motivo ela será a nossa variável a ser utilizada na Análise de Agrupamentos.

2.2 ANÁLISE DE GRUPAMENTO

Análise de grupamento é um conjunto de técnicas estatísticas, que consiste em dividir os elementos de uma amostra em subgrupos, com a finalidade de que objetos pertencentes a um mesmo grupo sejam similares entre si, e objetos de grupos distintos tenham características diferentes (JAMES et al., 2013). A qualidade de um agrupamento é medida pela menor diferença entre elementos de um mesmo grupo e pela maior diferença com elementos de outros grupos (PATTARIN; PATERLINI; MINERVA, 2004).

No entanto, a utilização da noção de similaridade como critério de classificação dos grupos pode ser problemática devido à sua natureza abstrata. A fim de tornar a análise mais objetiva é desejável que os objetos da amostra sejam definidos como similares ou diferentes de forma quantitativa. Uma abordagem simples para mensurar o quão semelhantes os objetos são é analisar o quão distantes eles estão entre si, dessa forma objetos que apresentam um valor para distância suficientemente pequeno serão alocados no mesmo grupo, enquanto aqueles com distâncias consideravelmente grandes serão alocados em grupos distintos.(DURAN; ODELL, 1974)

Uma outra interpretação que pode ser feita é analisar a distância como uma medida de dispersão, variância. Ao agrupar as observações da amostra conforme as suas distâncias,

estamos basicamente diminuindo a variância entre elementos de um mesmo grupo (JAMES et al., 2013).

O livro de (TAN et al., 2019) traz um panorama geral a respeito dos Agrupamentos, os classificando em: hierárquicos , particionais , exclusivos , com sobreposição , fuzzy, completos e parciais.

Agrupamentos particionais são obtidos a partir de um simples divisão dos dados, em uma quantidade de partes previamente informada, obedecendo o critério da não sobreposição, ou seja, cada objeto da amostra pertencerá a apenas um grupo.

Agrupamentos hierárquicos, não necessitam que seja informado de antemão a quantidade de grupos no qual a amostra será dividida, ao final da análise obtemos uma árvore de possibilidades, dendograma, que permite a visualização da composição dos agrupamentos conforme a quantidade de grupos desejados, quantidade essa que pode variar de 1 até n, onde n é igual ao tamanho da amostra. (JAMES et al., 2013). Essa forma de visualização dos agrupamentos permite analisar como as observações de um mesmo grupo divergem entre si, e como elas podem vir a ser divididas caso a análise requeira um número maior de grupos a serem formados, também nos possibilita observar a semelhança entre os agrupamentos, que podem ser unidos caso o número de grupos necessite ser menor.

Agrupamentos exclusivos são aqueles em que, cada objeto da amostra pertence a um e somente um grupo, já nos sobrepostos (ou não-exclusivos), um objeto da amostra pode pertencer a mais de um grupo, seja porque a observação recebe duas classificações, ou seja porque se encontra indefinida entre dois grupos.

Nos agrupamentos fuzzy, a classificação de um objeto da amostra com um determinado grupo, não é mais uma variável dicotômica, pertence ou não pertence, essa relação entre objeto e grupo é quantificada por um peso que varia de 0 a 1, a quantidade de pesos que um objeto tem é igual a quantidade de agrupamentos no qual a amostra está sendo sujeita a análise, assim um objeto tem um peso para todos os agrupamentos, de tal forma que a soma dos pesos resulta em um.

Com relação à totalidade dos objetos da amostra a serem utilizados na análise, os agrupamentos podem ser classificados como completos, quando toda a amostra é utilizada, ou parciais, quando a amostra passa por um processo de tratamento e alguns objetos são removidos da análise. O motivo para não utilização de todos os objetos da amostra se deve ao fato de que por vezes alguns objetos representam interferências (*noise*) ou *outliers*, que quando não excluídos podem comprometer a qualidade do estudo.

2.3 K-MEANS

Segundo (JAMES et al., 2013) K-means é uma técnica de Análise de Grupamento, simples e elegante, que necessita de que o número de grupos K a serem formados seja previamente definido. Quando aplicado proporciona a divisão do conjunto de dados em k grupos, não sobrepostos, com essa descrição podemos classificar os agrupamentos formados pelo K-means como Particionais Exclusivos, para isso devem obedecer simultaneamente as duas restrições abaixo

- $C_1 \cup C_2 \cup \ldots \cup C_K = \{1, \ldots, n\}$, ou seja, a união dos cluster deve resultar na amostra.
- $C_k \cap C_{k'} = \emptyset$ for all $k \neq k'$, em outras palavras, a interseção entre quaisquer dois diferentes clusters deve ser vazia.

O problema do K-means se resume a um problema de minimização, queremos diminuir a variância entre elementos de um mesmo grupo, variância intragrupo, em inglês Within-Cluster Variation. W(Ck) é um operador que mensura esta variância, geralmente uma distância, assim podemos expressar nosso problema da seguinte forma.

$$\text{Minimizar } \left\{ \sum_{k=1}^{K} W(C_k) \right\} \tag{2.8}$$

3 DADOS E MÉTODOS

No presente capítulo, serão discutidos aspectos relacionados à seleção de ações pelo processo de Backtesting. Será detalhado como o excesso de perda média é calculado para os ativos, utilizando-se da fórmula empírica. Além disso, será apresentado a lógica por trás da classificação dos grupos bem como o algoritmo K-means de Hartigan-Wong, utilizado para a formação dos agrupamentos.

3.1 DADOS

Os ativos financeiros analisados neste estudo estão listados na bolsa de valores brasileira, B3 (<https://www.b3.com.br/pt_br/b3/institucional/quem-somos/>), e fazem parte do principal índice do mercado de capitais brasileiro o IBOVESPA. As informações a respeito de tais ativos, foram obtidas por meio do *Website* Yahoo Finance (<https://finance.yahoo.com/>), o período de abrangência do estudo compreende o segundo semestre de 2021, e os dados foram coletados diariamente. Todas as análises de dados foram realizadas utilizando o software estatístico R, o qual possui integração com o banco de dados do Yahoo Finance, ou seja a coleta e análise dos dados poderam ser realizadas no mesmo ambiente de programação.

3.2 MÉTODO GARCH-VAR

Nessa seção será demonstrado como podemos estimar o VaR a partir do GARCH, toda a metodologia aqui apresentada foi primeiramente apresentada em (BERNARDINO et al., 2018). O modelo GARCH em si não é utilizado para estimar diretamente o VaR, porém é utilizado para modelar a volatilidade condicional em uma série temporal, que neste estudo em questão será a série dos retornos diferidos da média, demeaned returns, $\{R_t^{dm}\}$, que por sua vez pode ser relacionado com o VaR, de tal forma que, moldando o comportamento do $\{R_t^{dm}\}$ podemos estimar o VaR.

$\{R_t^{dm}\}$ será modelado da seguinte forma, o seu comportamento será divido em duas partes, uma estocástica μ_t e um desvio padrão dependente do tempo σ_t, assim temos que a equação para a volatilidade é dada por, $R_t^{dm} = \mu_t = \sigma_t z_t$, caracterizando um modelo heteroscedásticos centrado em σ_t e com um forte ruído branco z_t. Usando o Generalized Autoregressive

Conditional Heteroskedasticity (GARCH), a variância condicional pode ser modelada por um GARCH(p, q), da seguinte forma:

$$\sigma_t^2 = \gamma_0 + A(\omega)(R_t^{dm})^2 + B(\omega)\sigma_t^2 \tag{3.1}$$

onde $\gamma_i > 0$, $i = 1, 2, ..., q$. $A(\omega) = \gamma_1\omega + \cdots + \gamma_q\omega^q$ e $B(\omega) = \beta_1\omega + \cdots + \beta_p\omega^p$ em que A e B são polinômios de ordem p e q, e ω é um componente de defasagem.

Para estimar o VaR, primeiro precisamos assumir uma distribuição $\mathbf{Pr}$ para os retornos, umas vez que a distribuição for escolhida o VaR resumirar-se ao calculo de um quantil. Considerando que $\{R_t^{dm}\}$ segue a lei de probabilidades $\mathbf{Pr}(\{R_t^{dm}\} \leq \text{VaR}_\alpha^t|\mathcal{F}_{t-1}) = \mathcal{F}_t(\text{VaR}_\alpha^t)$, e que $\{R_t^{dm}\}$ segue o processo estocástico $R_t^{dm} = \mu_t + u_t = \mu_t + \sigma_t z_t$, onde $\mu_t = E(R_t^{dm}|\mathcal{F}_{t-1}) = 0$ e $\sigma_t^2 = E(u_t^2|\mathcal{F}_{t-1})$. A distribuiçao condicional do processo $\{z_t\} = \{u_t/\sigma_t\}$ e dada por $G_t(z)$ $= \mathbf{Pr}(z_t \leq z|\mathcal{F}_{t-1})$.

Relembrado a equação 2.2, temos que o VaR pode ser escrito da seguinte maneira, $F_X[VaR_p(X)] = p$, ou seja, $F_X^{-1}(p) = VaR_p(X)$, trazendo a notação para o contexto de séries temporais temos, $F_t(\text{VaR}_\alpha^t) = (1-\alpha)$ e $\text{VaR}_\alpha^t = F_t^{-1}(1-\alpha)$. Por fim podemos escrever o VaR sendo, $\text{VaR}_\alpha^t = F_t^{-1}(1-\alpha) = \sigma_t G_t^{-1}(z)(1-\alpha)$. usando o que foi apresentado em (3.1) podemos em fim estimar o VaR a partir dos $\{R_t^{dm}\}$ modelados através do GARCH.

O processo de backtesting do VaR envolve a comparação das perdas reais com as perdas estimadas pelo VaR, em um determinado período, para isso considera um nível de confiança específico α e uma janela de estimação, Window Estimation, (WE). Essa análise permite avaliar a qualidade da estimativa do VaR, pois quanto mais precisa for a estimativa, mais próximos as perdas estimadas estarão das perdas reais.

A seleção da WE é baseada na Razão de Violação, Violation Ratio,(VR). A VR é calculada como a divisão entre o número real de violações do VaR e o número esperado de violações. No caso de um $\text{VaR}_{98\%}$, por exemplo, espera-se que em 100 observações, ocorram 2 violações do VaR. Um VR ideal é igual a 1, indicando que a estimativa do VaR coincide com as observações reais. Um VR inferior a 1 sugere que estamos estimando um número de violações do VaR maior do que o realmente ocorrido, já um VR maior que 1 indica que estimamos um número de violações menor do que o número real. O desejável é manter o VR dentro da faixa de 0,5 a 1,5 para garantir uma estimativa eficaz do VaR.

O backtesting é essencial, pois desempenha um papel crucial na otimização da análise ao reduzir a quantidade de ativos a serem examinados. Isso, por sua vez, resulta em uma diminuição dos custos computacionais. Além disso, o backtesting é valioso para mitigar os

efeitos adversos de ativos altamente voláteis e com difícil previsão, melhorando a qualidade de nossas análises.

3.3 A FUNÇÃO EMPÍRICA DO MEAN EXCESS LOSS

A função *mean excess loss* (KLUGMAN; PANJER; WILLMOT, 2019) é definida como o valor esperado da perda excedente a um valor limiar u, ou seja, $e(u) = \mathrm{E}(X - u|X > u)$, e foi melhor detalhada na seção 2.1 (Teoria do Valor Extremo e Medidas de Risco). Como pode ser visto na expressão abaixo, o excesso de perda média empírico, nada mais é do que a média de todos os valores que excedem o limiar μ.

$$e^{EMP}(u) = \frac{\sum_{i=1}^{n} \max(X_i - u, 0)}{\sum_{i=1}^{n} I\{X_i > u\}} \tag{3.2}$$

Para o cálculo do Mean Excess Loss Empírico $e^{EMP}(u)$ da distribuição dos retornos, foram utilizados os dados da série temporal de retornos a partir do primeiro dia útil de 2015, em 01/01/2015. Por exemplo, ao realizar a análise de clusters no dia 13/08/2021, foram consideradas as informações de retorno desde 01/01/2015 até 13/08/2021.

3.4 ANÁLISE DE GRUPAMENTOS

A análise de grupamento, possibilitada através da técnica K-means, foi feita com o intuito de agrupar os ativos financeiros conforme o seu nível de risco, o qual foram calculados por meio da medida de risco excesso de perda média empírico e^{EMP}.

A formação de grupos por meio da técnica do K-means foi prontamente aplicada utilizando a função kmeans do pacote clusters do R. Para esse propósito, foi necessário inicialmente definir o número desejado de grupos, que foi estabelecido como 4, e também especificar o número de repetições do algoritmo, que foi definido para 25. As linhas de código das Figuras 1 e 2, são um recorte do script do código que foi utilizado, e expõe como as configurações iniciais do K-means foram definidas.

O número de grupos a ser formados pode ser maior do que o informado inicialmente, isso se deve ao fato de que as funções do pacote cluster do R definem qual a melhor quantidade de grupos a serem formada, assim poderão ter dias em que o número de clusters exceda 4. O número de vezes que o algoritmo do K-means irá ser repetido foi definido como 25, um

```r
avg_sil <- function(k, df) {
  km.res <- kmeans(df, centers = k, nstart = 25)
  ss <- silhouette(km.res$cluster, dist(df))
  return(mean(ss[, 3]))
}
```

Figura 1 – print do R script

Fonte: Autor, 2023

```r
start_cluster <- 4
k.values <- start_cluster:(nrow(df)-1)
avg_sil_val <- map_dbl(.x = k.values, df = df, .f = avg_sil)
best_k <- which.max(avg_sil_val)+(start_cluster-1)

df_kmean <- df
km <- kmeans(df, centers = best_k, nstart = 25)
```

Figura 2 – print do R script

Fonte: Autor, 2023

número maior de repetições torna a análise melhor, porém aumenta os recursos computacionais necessários para tal e consequentemente o tempo de execução.

Na Figura 2, durante a aplicação da função 'kmeans', não foi especificado qual algoritmo específico foi utilizado. Portanto, o algoritmo padrão da função, que é o Hartigan-Wong, foi empregado. Maiores detalhes sobre o algoritmo serão apresentados no na seção 3.5 (Algoritmo Hartigan-Wong).

Após a conclusão da divisão dos agrupamentos, calculou-se a mediana dos valores de e^{EMP} para cada grupo. Os grupos foram numerados de maneira que aqueles com as medianas mais altas receberam os números mais baixos. Em outras palavras, o grupo 1 concentra os ativos com os maiores valores de e^{EMP}, indicando menor nível de risco. À medida que o número do grupo aumenta, os ativos incluídos nele exibem um e^{EMP} menor, refletindo um maior nível de risco.

3.5 ALGORITMO HARTIGAN-WONG

A linguagem de programação R, através da função kmeans do pacote cluster, nos possibilita escolher qual algoritmo estaremos utilizando, eles são: Hartigan-Wong, Llyod, Forgy e

MacQueen, a seguir será apresentada uma brevemente descrição do algoritmos Hartigan-Wong.

J. A. Hartigan e M. A. Wong, desenvolveram em 1979 um algoritmo para análise de agrupamento K-means, o qual recebe o mesmo nome de seus criadores, Hartigan-Wong, que foi detalhado em (HARTIGAN; WONG, 1979). Antes de entrar mais profundamente no algoritmo é necessário apresentar uma breve nomenclatura que foi utilizada pelos pesquisadores no artigo:

- NC(L): números de grupos no cluster L;
- D(I,L): distância euclidiana entre o ponto I e o cluster L;
- IC1(I): cluster mais próximo de I;
- IC2(I): segundo cluster mais próximo de I.

O algoritmo usa duas métricas para agrupar os pontos, R1 e R2, que em sua formulação utiliza a distância euclidiana. R1 e R2 podem ser interpretadas como variâncias intragrupal, e o objetivo do algoritmo é minimizá-las.

O algoritmo divide os clusters em dois grupos, os clusters que pertencem ao *live set*, e os clusters não pertencentes ao *live set*. O live set consiste no conjunto de clusters que poderão ter sua composição alterada, ou seja, a massa dos possíveis pontos a serem transferidos. O *live set* também serve para determinar quando o algoritmo será finalizado, uma vez que quando o mesmo estiver vazio, significa que nenhum cluster irá necessitar de alterações, caracterizando assim a seleção ótima.

O algoritmo é aplicado em um ponto de cada vez, e a cada rodada os clusters são atualizados, caso seja necessário. Os pontos podem ser transferidos de um cluster para outro em duas etapas denominadas QTRAN (quick transfer, transferência rápida) e OPTRA (optimal transfer, transferência ótima). No QTRAN, é feita uma breve análise do ponto com 2 clusters, esses cluster são o IC1 e o IC2, que são os 2 clusters que estão mais próximos do ponto. No OPTRA é feito algo parecido ao que ocorre no QTRAN porém a análise agora é extrapolada para outros clusters, não se restringindo ao IC1 e IC2.

Para que o Algoritmo possa ser iniciado é necessário informar a quantidade de cluster e seus centros iniciais, que irão ser alterados conforme o andar do algoritmo, tais centros podem ser definidos aleatoriamente. O algoritmo é dividido em nove passos: 1, 2, 3, 4, 4a, 4b, 5, 6 e 7 os quais serão descritos abaixo e foram inicialmente apresentados em (HARTIGAN; WONG, 1979).

Passo 1: para cada ponto I(I = 1,2,..., M), encontre o cluster mais próximo e o segundo mais próximo, IC1(I) E IC2(I) respectivamente. Associe o ponto I ao cluster IC1(I).

Passo 2: atualize o centro do cluster para que seja igual a média dos pontos contidos nele.

Passo 3: inicialmente, todos os pontos pertencem ao live set

Passo 4: transferência ótima OPTRA. Para cada ponto I (1,2,...,M). Se o cluster L(L = 1,2,..,M) foi atualizado no último estágio da transferência rápida QTRAN, então o cluster pertence ao live set para todo esse estágio. Durante o estágio do OPTRA, caso o cluster não seja atualizado ele não pertencerá mais ao live set. Assuma L1 como sendo o cluster de I. caso L1 esteja no live set faça 4a, caso contrário vá para 4b.

Passo 4a: calcule R2 para todos os cluster L(L=/1, 1,2,..,M), $R2 = \frac{NC(L2)\times D(I,L2)^2}{NC(L2)+1}$. Assuma L2 como sendo o cluster com o menor R2. Se esse valor for menor ou igual a $\frac{NC(L1)\times D(I,L1)^2}{NC(L1)-1}$, não é necessário nenhuma transferência e L2 é o novo IC2(I). Caso contrário o ponto I é alocado no cluster L2 e L1 é o novo IC2(I). No caso de haver transferência os centros dos clusters é atualizado para que seja igual a média dos pontos.

Passo 4b: essa etapa é igual a 4a, exceto que o R2 mínimo é calculado apenas nos clusters pertencentes ao live set.

Passo 5: caso o live set seja vazio pare o algoritmo. No contrário, faça o passo 6.

Passo 6: transferência rápida QTRAN, para cada ponto I (I = 1,2, ..., M). Assuma L1 = IC1(I) e L2 = IC2(I). não é necessário análisar o ponto I se ambos os pontos L1 e L2 não foram alterados nos últimos M passos. calcule R1 e R2, $R1 = \frac{NC(L1)\times D(I,L1)^2}{NC(L1)-1}$ e $R2 = \frac{NC(L2)\times D(I,L2)^2}{NC(L2)+1}$. Se R1 for menor do que R2, o ponto I permanece no cluster L1. Caso contrário troque IC1(I) por IC2(I) e atualize os centros dos clusters L1 e L2.

Passo 7: se nenhuma transferência ocorreu nos últimos M passos, faça o passo 4, caso contrário faça o passo 6.

O algoritmo Hartigan-Wong é notavelmente acessível em termos de implementação e compreensão. Ele se destaca por sua simplicidade, dispensando cálculos complexos. No entanto, é importante ressaltar algumas das desvantagens associadas a esse algoritmo, que foram apresentadas em (MORISSETTE; CHARTIER, 2013). Essas limitações são igualmente aplicáveis a outros métodos, como Forgy, Lloyd e MacQueen.

Embora o algoritmo sempre convirja para uma solução, nem sempre essa solução é o mínimo global desejado. O algoritmo pode ficar preso em mínimos locais, o que pode resultar em agrupamentos subótimos. Outra limitação notável está relacionada à sensibilidade à escolha dos centróides iniciais. A performance do algoritmo pode variar significativamente com base na

seleção inicial dos pontos representativos dos clusters, o que torna crucial a escolha cuidadosa desses centróides.

Além disso, o algoritmo Hartigan-Wong não oferece uma metodologia para determinar automaticamente o número ideal de clusters a serem formados, deixando essa tarefa para o usuário ou exigindo a aplicação de métodos externos.

Outra questão relevante é a susceptibilidade a problemas causados por outliers, uma vez que o algoritmo atualiza os centros dos clusters através de médias. Isso pode resultar na distorção dos centróides por valores extremos, afetando negativamente a qualidade dos agrupamentos.

Finalmente, o algoritmo tende a criar clusters com tamanhos aproximadamente iguais, o que pode não refletir adequadamente a estrutura intrínseca dos dados, caso ela seja desigual. Assim, embora o algoritmo Hartigan-Wong tenha suas vantagens em termos de simplicidade e ampla utilização, é importante estar ciente de suas limitações.

4 RESULTADOS

Após a divisão dos ativos nos Grupos, foram formadas 3 *datas frames* (tabelas) que possibilitaram as análises, essas tabelas foram nomeadas da seguinte forma: Tabela 1, Tabela 2 e Tabela 3. Em todas as tabelas a primeira coluna é uma coluna índice contendo as datas nas quais as informações financeiras foram retiradas, as 3 tabelas tem 121 linhas, as Tabela 1 e 2 têm a mesma quantidade de colunas 212; em cada coluna, com exceção da primeira que contém as datas, temos informações a respeito de uma ação; a Tabela 1 foi preenchida com o número do cluster em que a ação se encontrava no respectivo dia; já a Tabela 2 informa para cada ação qual era a mediana do excesso de perda média empírico $e^{EMP}(u)$ do grupo no qual a ação pertencia no dia. Abaixo as Tabelas 1 e 2 podem ser observadas de forma parcial, uma vez que mostrar todas as linhas e colunas das tabelas é visualmente inviável.

Tabela 1 – Ativos e Grupo.

Data	BBAS3	CLSC4	JBSS3	LPSB3	PETR4	VALE3
15/12/2021	2	*NA*	3	3	2	3
16/11/2021	2	1	3	3	2	3
14/10/2021	2	1	2	2	2	3
15/09/2021	2	1	3	*NA*	2	3
16/08/2021	3	*NA*	3	*NA*	3	2
15/07/2021	2	*NA*	2	*NA*	2	3

Tabela 2 – Ativos e Média do Grupo.

Data	BBAS3	CLSC4	JBSS3	LPSB3	PETR4	VALE3
15/12/2021	-6.597650	*NA*	-6.771221	-6.771221	-6.597650	-6.771221
16/11/2021	-6.542546	-4.999907	-6.721180	-6.721180	-6.542546	-6.721180
14/10/2021	-6.173489	-4.484767	-6.173489	-6.173489	-6.173489	-6.275071
15/09/2021	-6.384294	-4.844934	-6.406605	*NA*	-6.384294	-6.406605
16/08/2021	-6.372259	*NA*	-6.372259	*NA*	-6.372259	-6.357423
15/07/2021	-6.119355	*NA*	-6.119355	*NA*	-6.119355	-6.287956

A Tabela 3 possibilita observar a variação da mediana de um grupo ao longo dos dias, bem como a variação da mediana entre diferentes grupos em um mesmo dia. Segue abaixo um recorte da Tabela 3.

Tabela 3 – Evolução das Medianas.

Data	1	2	3	4	5	6
15/12/2021	-4.803596	*-6.597650*	-6.771221	-12.909517	*NA*	*NA*
16/11/2021	-4.999907	-6.542546	-6.721180	-13.555371	*NA*	*NA*
14/10/2021	-4.484767	-6.173489	-6.275071	-6.947383	-9.478787	-13.13282
15/09/2021	-4.844934	-6.384294	-6.406605	*-15.080500*	*NA*	*NA*
16/08/2021	-4.538869	*-6.357423*	-6.372259	*-6.629817*	-14.247115	*NA*
15/07/2021	-4.566965	*-6.119355*	-6.287956	*-9.554145*	-12.892168	*NA*

Conforme mencionado anteriormente neste trabalho, o período de análise correspondeu ao segundo semestre de 2021, abrangendo um total de 128 dias úteis. No entanto, durante a extração dos dados do Yahoo Finance, foi inadvertidamente omitido o registro de 7 datas: 01/07/2021, 09/07/2021, 06/09/2021, 12/11/2021, 24/12/2021, 27/12/2021 e 31/12/2021. Portanto, a análise foi conduzida com base em 121 dias.

O fato de as tabelas 1, 2 e 3 possuírem 212 colunas pode levar a um entendimento equivocado de que diariamente estão sendo analisados 211 ativos, lembrar da presença da coluna Data. Essa suposição está incorreta, uma vez que os ativos são submetidos ao processo de backtesting, sobre o qual não temos controle quanto ao número de ativos que passam por ele diariamente. Portanto, em todos os dias, foram analisadas menos de 211 ações. A média diária da quantidade de ativos analisados é de 196,6 ativos por dia. O dia com menor quantidade de ações analisadas, 170, foi em 12/07/2021. Por outro lado, a maior quantidade de ações analisadas em um dia foi, 198, que ocorreu em 6 dias: 06/10/2021, 07/10/2021, 17/11/2021, 22/11/2021, 23/11/2021, 24/11/2021. Na Figura 3, apresentamos um gráfico de barras que ilustra a distribuição da quantidade de ativos analisados por dia.

A variação na quantidade de ativos analisados a cada dia também explica a ocorrência de valores 'NA' (informação não disponível) nas tabelas 1 e 2. Isso indica que, naquele dia, o ativo não foi validado no backtesting e, portanto, foi excluído da amostra de ativos. Na Tabela 3, a presença de valores 'NA' é atribuída à variação na quantidade de grupos formados em cada dia, que pode variar entre 4, 5 e 6 grupos. O método K-Means trabalha

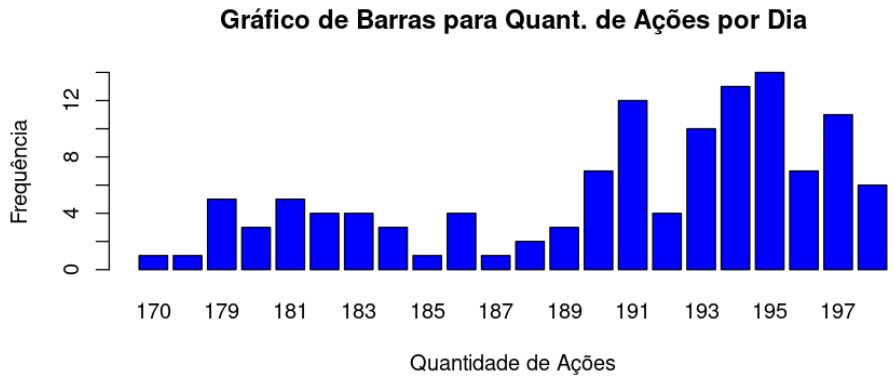

Figura 3 – Grafico de Barras para Quantidade de ações por dia

Fonte: Autor, 2023

com a quantidade de clusters previamente definidas, com o intuito de melhorar a qualidade dos agrupamentos, foram utilizadas funções do pacote cluster para calcular a quantidade ideal de clusters a serem formados, essa quantidade não é fixa e varia durante o andamento dos dias, muito porque as informações financeiras a respeito dos ativos podem ser bastante dinâmicas, além da quantidade de ativos não ser constante, com dias tendo mais ativos avaliados do que outros.

Utilizando os dados apresentados na Tabela 1, é possível realizar uma análise das alterações de grupos ocorridas em relação a um ativo específico. Além disso, é viável identificar os ativos que mais frequentemente mudam de grupo, assim como aqueles que mantêm uma permanência mais prolongada em um mesmo grupo. Porém devemos nos atentar aos NA, por vezes alguns ativos encontram em um grupo e no dia seguinte não passam pelo backing test, a seguir será mostrado na Tabela 4 o que ocorreu com o ativo DIRR3.

Nesse caso consideramos o *NA* como sendo um outro grupo, assim no período em questão foi considerado que o ativo trocou 4 vezes de grupo.

Em relação aos dados ausentes (NA), ao analisar os 211 ativos, notamos que 126 deles, correspondendo a 59,72% do total de ativos, não apresentaram quaisquer dados ausentes, isso sugere que esses ativos estiveram disponíveis para análise durante todos os 121 dias considerados. Por outro lado, dos 85 ativos restantes que em algum momento não estavam disponíveis para análise, 63,53% deles não foram aprovados no backtesting em até 25% dos dias analisados, permanecendo fora da análise de agrupamento por no máximo 30 dias. Para uma melhor visualização desses dados, o gráfico em barras da Figura 4 proporciona uma representação

Tabela 4 – Tabela Evolução DIRR3.

Data	DIRR3
26/11/2021	3
25/11/2021	*NA*
24/11/2021	3
23/11/2021	3
22/11/2021	3
21/11/2021	*NA*
20/11/2021	3

mais clara. Ele nos ajuda a responder à seguinte pergunta: para os ativos que estiveram ausentes em algum momento da análise, quantos dias estiveram ausentes? É importante ressaltar que esses dias não necessariamente ocorreram de forma consecutiva [4].

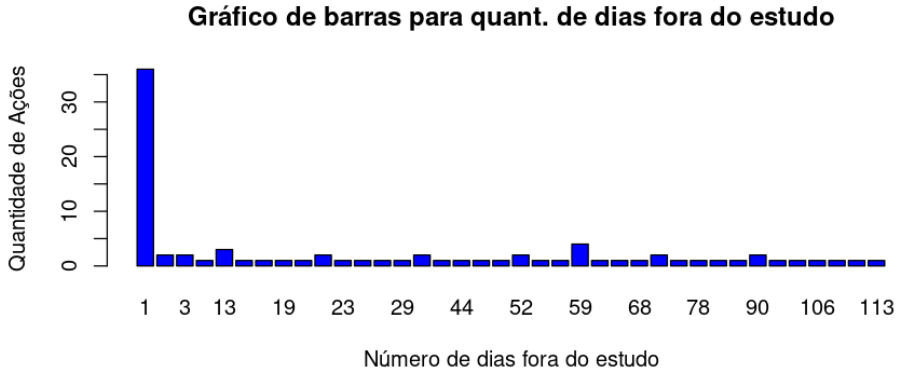

Figura 4 – Gráfico de barras para quantidade de dias fora do estudo

Fonte: Autor, 2023

A seguir na Tabela [6] é mostrado um ranking dos que mais trocaram de grupos, os quais representam uma maior imprevisibilidade na previsão do nível de risco, e por isso devem ser destacados. Na Figura pode-se observa a frequencia de ativos para um número determinado de transições, por exemplo 10 ativos trocaram de grupo 44 vezes.

Os ativos que aparecem com asterisco pertencem ao grupo de 85 ações que em algum momento não passaram pelo backtesting, ou seja, na contagem de mudança de grupo estão sendo considerada as vezes em que o ativo passo pelo estado NA. Segue abaixo gráfico

Tabela 5 – Ranking Maiores Mudanças de Grupo

Ação	Nº de mudanças de Grupo
FRIO3*	47
OIBR4	45
BAUH4*, BDLL4, BMIN4, BOVA11*, GOVE11*, OIBR3, PIBB11*, TCNO3*, TIPS3 e VULC3	44
FHER3*, GOLL4 MATB11* e PDTC3*	43
AZEV4, DASA3, FRTA3 INEP311 e POSI3	42

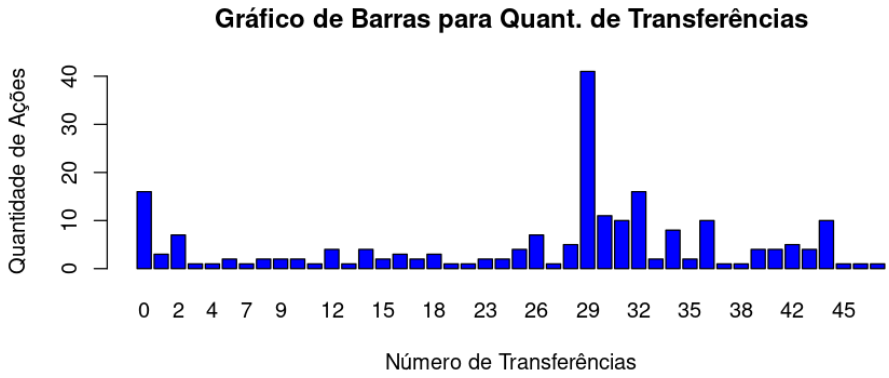

Figura 5 – Gráfico de Barras quantidade de ativos para um determinado numero de transições

Fonte: Autor, 2023

A partir da Tabela 3, podemos elaborar um box-plot, Figura 4, que nos permite avaliar com os diferentes valores das medianas estão dispersos para cada grupo. Ao sobrepor os limites dos boxplots, é possível observar três regiões distintas. A primeira é compreendida pelo grupo 1, a segunda engloba os grupos 2 e 3, e a terceira inclui os grupos 4, 5 e 6.

A partir dos dados da Tabela 1 podemos analisar o tamanho dos grupos, o número de ações que foram alocadas em cada grupo, em uma determinada data, na Tabela 6 temos um exemplo para o dia 10/12/2021.

Vale ressaltar que no dia em questão foram gerados 5 grupos, porém essa tabela poderia ter mais ou menos linhas, a depender se no dia foram formados 4 ou 6 grupos.

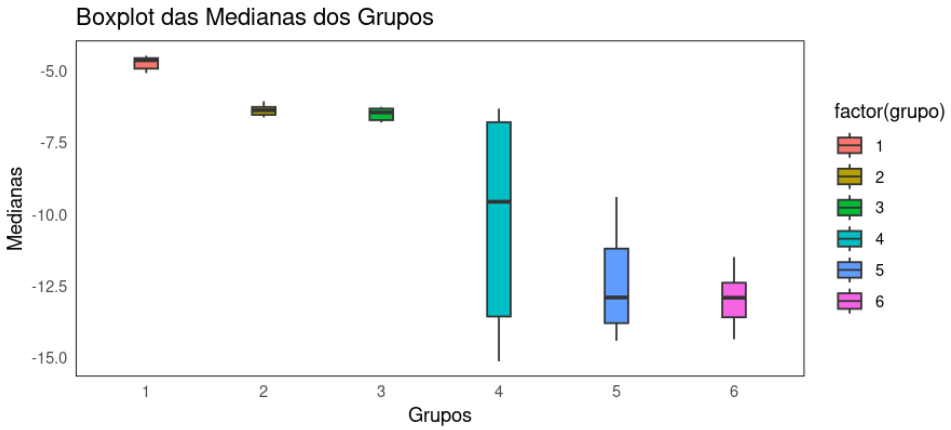

Figura 6 – Boxplot da Mediana dos Grupos

Fonte: Autor, 2023

Tabela 6 – Quantidade de Ações por grupo

Grupo	Quantidade de Ações
1	35
2	61
3	85
4	8
5	6
NA	16

5 CONCLUSÃO

O presente trabalho teve como objetivo apresentar uma metodologia para quantificar o risco financeiro de ativos financeiros, tendo como base a utilização da medida de risco excesso de perda média empírico, bem como versou a respeito da metodologia K-Means para o agrupamento dos ativos financeiros conforme o seu nível de risco.

As metodologias apresentadas neste trabalho fornecem uma importante ferramenta que pode ser utilizada por gestores de fundos e tomadores de decisão. Uma vez que possibilita uma visão resumida de uma vasta quantidade de ativos, contribuindo assim para o entendimento do cenário econômico.

As dificuldades encontradas neste trabalho envolvem principalmente o grande número de dias e ativos envolvidos, necessitando assim de um grande espaço para armazenamento, e aumentando consideravelmente o tempo necessário para a realização dos cálculos em termos computacionais.

Trabalhos futuros poderão utilizar-se das metodologias apresentadas para fazer um estudo mais aprofundado, analisando como a formação dos grupos ocorre em um período de tempo maior, possibilitando assim a análise de agrupamentos em diferentes momentos da economia, em especial em momentos de estressamento do mercado financeiro, as maiores variações podem ser observadas.

REFERÊNCIAS

ALBRECHT, P.; MAURER, R.; RUCKPAUL, U. On the risks of stocks in the long run: A probabilistic approach based on measures of shortfall risk. *None*, 2001.

ARTZNER, P.; DELBAEN, F.; EBER, J.-M.; HEATH, D. Coherent measures of risk. *Mathematical finance*, Wiley Online Library, v. 9, n. 3, p. 203–228, 1999.

BERNARDINO, W.; BRITO, L.; OSPINA, R.; MELO, S. *A GARCH-VaR investigation on the Brazilian sectoral stock indices. Brazilian Review of Finance, 16 (4), 573-610*. 2018.

DURAN, B. S.; ODELL, P. L. *A Cluster Analysis Survey*. ? ?: Springer-Verlag, 1974.

HARTIGAN, J. A.; WONG, M. A. A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applaied Statistics)*, Wiley, v. 28, n. 1, p. 100–108, 1979.

HOLTON, G. A. *History of Value-at-Risk: 1992-1998*. Boston, 2002. (CESIfo Working Paper).

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. et al. *An introduction to statistical learning*. [S.l.]: Springer, 2013. v. 112.

KLUGMAN, S. A.; PANJER, H. H.; WILLMOT, G. E. *Loss Models, from data to decision*. 5. ed. 111 River Street, Hoboken, NJ 07030, USA: John Wiley and Sons, 2019.

MORISSETTE, L.; CHARTIER, S. The k-means clustering technique: General considerations and implementation in mathematica. *Tutorials in Quantitative Methods for Psychology*, v. 9, n. 1, p. 15–24, 2013.

PATTARIN, F.; PATERLINI, S.; MINERVA, T. Clustering financial time series: an application to mutual funds style analysis. *Computational Statistics & Data Analysis*, Elsevier, v. 47, n. 2, p. 353–372, 2004.

SINGH, A. K.; ALLEN, D. E.; ROBERT, P. J. Extreme market risk and extreme value theory. *Mathematics and computers in simulation*, Elsevier, v. 94, p. 310–328, 2013.

TAN, P.-N.; STEINBACH, M.; KUMAR, V.; KARPATNE, A. *Introduction to Data Mining*. 2. ed. 330 Hudson Street. New York, NY, Estados Unidos: Pearson Education, Inc., 2019.