

UNIVERSIDADE FEDERAL DO PARANÁ

ALYSON AUGUSTO VIVAN

LÓGICA DIFUSA APLICADA AO AGRUPAMENTO DE CHEIAS HISTÓRICAS
OBSERVADAS NO RIO IGUAÇU NA CIDADE DE UNIÃO DA VITÓRIA:
COMPARAÇÃO ENTRE MÉTODOS DE AGRUPAMENTO *K-MEANS* E *FUZZY C-*
MEANS

CURITIBA

2020

ALYSON AUGUSTO VIVAN

LÓGICA DIFUSA APLICADA AO AGRUPAMENTO DE CHEIAS HISTÓRICAS
OBSERVADAS NO RIO IGUAÇU NA CIDADE DE UNIÃO DA VITÓRIA:
COMPARAÇÃO ENTRE MÉTODOS DE AGRUPAMENTO *K-MEANS* E *FUZZY C-
MEANS*

Dissertação de mestrado apresentada como requisito parcial para a obtenção do grau de Mestre em Engenharia de Recursos Hídricos e Ambiental através do Programa de Pós-Graduação em Engenharia de Recursos Hídricos e Ambiental, setor de tecnologia, Departamento de Hidráulica e Saneamento, Universidade Federal do Paraná.

Orientador: Professor Dr. Júlio Gomes

CURITIBA

2020

Catálogo na Fonte: Sistema de Bibliotecas, UFPR
Biblioteca de Ciência e Tecnologia

V855l Vivan, Alyson Augusto

Lógica difusa aplicada ao agrupamento de cheias históricas observadas no rio Iguaçu na cidade de União da Vitória [recurso eletrônico]: comparação entre métodos de agrupamento k-means e Fuzzy Cmeans / Alyson Augusto Vivan. – Curitiba, 2020.

Dissertação - Universidade Federal do Paraná, Setor de Tecnologia, Programa de Pós-Graduação em Engenharia de Recursos Hídricos e Ambiental, 2020.

Orientador: Júlio Gomes.

1. Lógica difusa. 2. Análise de componentes principais. 3. Inundações. I. Universidade Federal do Paraná. II. Gomes, Júlio. III. Título.

CDD: 363.3493

Bibliotecária: Vanusa Maciel CRB- 9/1928



MINISTÉRIO DA EDUCAÇÃO
SETOR DE TECNOLOGIA
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO ENGENHARIA DE
RECURSOS HÍDRICOS E AMBIENTAL - 40001016021P0

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ENGENHARIA DE RECURSOS HÍDRICOS E AMBIENTAL da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **ALYSON AUGUSTO VIVAN** intitulada: **LÓGICA DIFUSA APLICADA AO AGRUPAMENTO DE CHEIAS HISTÓRICAS OBSERVADAS NO RIO IGUAÇU NA CIDADE DE UNIÃO DA VITÓRIA: COMPARAÇÃO ENTRE MÉTODOS DE AGRUPAMENTO K-MEANS E FUZZY CMEANS.**, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 27 de Outubro de 2020.

Assinatura Eletrônica

11/02/2021 14:50:19,0

JÚLIO GOMES

Presidente da Banca Examinadora

Assinatura Eletrônica

11/02/2021 14:50:19,0

CRISTOVÃO VICENTE SCAPULATEMPO FERNANDES

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

11/02/2021 14:50:19,0

IRANI DOS SANTOS

Avaliador Externo (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

11/02/2021 14:50:19,0

ELOY KAVISKI

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Centro Politécnico, Bloco V - CURITIBA - Paraná - Brasil
CEP 81531-990 - Tel: (41) 3361-3210 - E-mail: ppgerha@ufpr.br

Documento assinado eletronicamente de acordo com o disposto na legislação federal Decreto 8539 de 08 de outubro de 2015.
Gerado e autenticado pelo SIGA-UFPR, com a seguinte identificação única: 79975

Para autenticar este documento/assinatura, acesse <https://www.prppg.ufpr.br/siga/visitante/autenticacaoassinaturas.jsp> e insira o código 79975

Dedico questo lavoro ai miei nonni Giuseppe e Luigia, ancca a la mia spoxa Denise e mia piccola tosa Vitória. Chi va piano se va sano e va lontano¹.

AGRADECIMENTOS

¹ Texto escrito em língua Veneta, com gramática característica da região da província de Treviso, Itália.

A Deus, pelo dom da vida e pela oportunidade da realização deste curso.

À minha esposa, pelos 28 anos de convivência. Posso dizer sem sombra de dúvidas que jamais teria chego até aqui sem o seu auxílio.

À minha filha, por ter me ensinado um amor tão grande, que só os pais e os anjos podem experimentar.

À colega Patrícia Steffen pelas trocas de informações; pela disponibilidade, cordialidade e educação com que sempre me tratou.

Ao amigo Adhemar Romero, jovem de inteligência ímpar, o qual acompanhou de perto os melhores e piores momentos dessa trajetória.

A Giuseppe Verdi, Giacomo Puccini, Antonio Vivaldi, Richard Wagner, Ludwig Van Beethoven, Wolfgang Amadeus Mozart e Enrico Caruso, por terem feito a trilha sonora desta viagem, acalmando e aquecendo meu coração, em incontáveis madrugadas gélidas de Curitiba e São Mateus do Sul.

Aos professores e a todos os funcionários do Programa de Pós-Graduação em Engenharia de Recursos Hídricos e Ambiental, da Universidade Federal do Paraná, que de alguma forma fizeram parte desta conquista.

Ao meu orientador Professor Dr. Júlio Gomes, e aos Professores Drs. Cristóvão Vicente Scapulatempo Fernandes, Eloy Kaviski, José Jungi Ota, Miriam Rita Moro Mine e Tobias Bernward Bleninger, agradeço e parafraseando *Sir Isaac Newton* afirmo: *“Se cheguei até aqui, foi porque me apoiei no ombro de gigantes”*.

“Os nativos que habitavam as margens do rio Iguaçu antes da chegada dos europeus, acreditavam que o mundo era governado por Mboi, um Deus com forma de serpente e filho de Tupã. Igobi, o cacique da tribo, tinha uma filha chamada Naipi, tão bonita que as águas dos rios paravam quando a jovem índia nele se mirava. Devido à sua beleza, Naipi seria consagrada ao deus Mboi, passando a viver somente para seu culto.

Havia, porém, na tribo, um jovem guerreiro chamado Tarobá, que se apaixonou ao ver Naipi. No dia da festa de consagração da jovem índia, enquanto o pajé e os caciques bebiam cauim (bebida feita de milho fermentado) e os guerreiros dançavam, Tarobá fugiu com a linda Naipi numa canoa que seguiu rio abaixo, arrastada pela correnteza.

Ao saber da fuga de Naipi e Tarobá, Mboi ficou furioso. Penetrou as entranhas da terra, retorcendo o seu corpo e produzindo uma enorme fenda que formou as cataratas. Envolvidos pelas águas dessa imensa cachoeira, a piroga (embarcação indígena a remo) e os fugitivos caíram de uma grande altura desaparecendo para sempre.

Naipi foi transformada em uma das rochas centrais das cataratas, perpetuamente fustigada pelas águas revoltas. Tarobá foi convertido em uma palmeira situada à beira de um abismo, inclinada sobre a garganta do rio. Debaixo dessa palmeira acha-se a entrada de uma gruta onde o monstro vingativo vigia eternamente as duas vítimas”.

(Mito Caingangue sobre a formação das Cataratas do Rio Iguaçu. Domínio público)

RESUMO

Enchentes são fenômenos naturais que afetam as sociedades humanas há milênios. Devido à sua posição geográfica e relevo característico, a cidade de União da Vitória, localizada no sudeste do Estado do Paraná, é frequentemente impactada por esse tipo de fenômeno, que causa danos ambientais, compromete drasticamente sua economia, e, não raramente, ocasiona perdas de vidas humanas. Neste cenário, com os objetivos de aprofundar o conhecimento sobre o comportamento hidrológico da bacia hidrográfica do médio Iguaçu e auxiliar na determinação de parâmetros de modelos hidrológicos, se insere o agrupamento de cheias baseado em lógica *Fuzzy*, que busca agrupar cheias com características similares, visando posterior análise dos mecanismos envolvidos na sua formação. Com base no trabalho de Steffen (2017), foram utilizados 85 eventos de cheias anuais máximas do rio Iguaçu, observados na Cidade de União da Vitória no período de 1931 a 2015 e 7 indicadores de cheia, usados como variáveis para o agrupamento. O presente trabalho comparou agrupamento de cheias históricas observadas em União da Vitória produzido por dois métodos distintos: *K-means* e *Fuzzy C-means*. Foi também analisada a influência de 4 métodos de normalização de dados (linear, por desvio padrão, por escala decimal e por valor máximo) sobre o resultado dos agrupamentos. Finalmente, aplicou-se a Análise de Componentes Principais, visando a diminuição do número de indicadores de cheia utilizados. Os resultados obtidos mostraram que a formação dos grupos é mais sensível ao método de normalização do que ao método de agrupamento (*K-means* e *Fuzzy C-means*). A Análise de Componentes Principais permitiu reduzir o conjunto inicial de 7 indicadores para somente dois indicadores de cheia, com uma correlação de 100% entre os grupos formados, mostrando a importância do uso de técnicas de estatística multivariada para a redução da dimensionalidade do problema.

Palavras-chave: Agrupamento de cheias; Lógica difusa; *Fuzzy C-means*; *K-means*; Análise de Componentes Principais.

ABSTRACT

Floods are natural phenomena that have affected human societies for millennia. Due to its geographical position and characteristic relief, the city of União da Vitória, located in the southeast of the State of Paraná, is frequently impacted by this type of phenomenon, which causes environmental damage, drastically compromises its economy, and, not infrequently, causes losses of human lives. In this scenario, with the objective of deepening the knowledge on the hydrological behavior of the hydrographic basin of the middle Iguaçu and assist in the determination of parameters of hydrological models, the flood grouping based on Fuzzy logic is inserted, which seeks to batch floods with similar characteristics, aiming the subsequent analysis of the mechanisms involved in its formation. Based on the work of Steffen (2017), 85 events of maximum annual floods of the Iguaçu River were used, observed in the City of União da Vitória from 1931 to 2015 and 7 flood indicators, used as variables for the batching. The present work has compared the grouping of historical floods observed in União da Vitória produced by two different methods: K-means and Fuzzy C-means. Was also analyzed the influence of 4 data normalization methods (linear, by standard deviation, by decimal scale and by maximum value) on the batch's results. Finally, the Principal Component Analysis was applied, aiming to reduce the number of flood indicators used. The obtained results showed that the formation of groups is more sensitive to the normalization method than to the batching method (K-means and Fuzzy C-means). The Principal Component Analysis allowed to reduce the initial set of 7 indicators to only two flood indicators, with a 100% correlation between the groups formed, showing the importance of using multivariate statistical techniques to reduce the dimensionality of the problem.

Keywords: Flood clustering; Flood indicators; Diffuse Logic; *Fuzzy C-means*; *K-means*; Principal Component Analysis.

SOMMARIO

Le inondazioni sono fenomeni naturali che hanno colpito le società umane per millenni. Per la sua posizione geografica e il caratteristico rilievo, la città di União da Vitória, situata nel sud-est dello Stato del Paraná, è frequentemente interessata da questo tipo di fenomeno, che provoca danni ambientali, ne compromette drasticamente l'economia e, non di rado, causa perdite di vite umane. In questo scenario, con l'obiettivo di approfondire la conoscenza sul comportamento idrologico del bacino idrografico del medio Iguaçu e coadiuvare nella determinazione dei parametri del modello idrologico, si inserisce il raggruppamento delle piene basato sulla logica Fuzzy, che cerca di raggruppare piene con caratteristiche simili, mirando successiva analisi dei meccanismi coinvolti nella sua formazione. Sulla base del lavoro di Steffen (2017), sono stati utilizzati 85 eventi di inondazioni annuali massime del fiume Iguaçu, osservati nella città di União da Vitória dal 1931 al 2015 e 7 indicatori di alluvione, usati come variabili per il raggruppamento. Il presente lavoro ha confrontato il raggruppamento di inondazioni storiche osservate a União da Vitória prodotte con due diversi metodi: K-means e Fuzzy C-means. È stata inoltre analizzata l'influenza di 4 metodi di normalizzazione dei dati (lineare, deviazione standard, scala decimale e valore massimo) sul risultato dei cluster. Infine, è stata applicata l'analisi delle componenti principali, con l'obiettivo di ridurre il numero di indicatori di alluvione utilizzati. I risultati ottenuti hanno mostrato che la formazione di gruppi è più sensibile al metodo di normalizzazione che al metodo di raggruppamento (K-mean e Fuzzy C-means). La Principal Component Analysis ha consentito di ridurre il set iniziale di 7 indicatori a soli due indicatori di piena, con una correlazione del 100% tra i gruppi formati, dimostrando l'importanza di utilizzare tecniche statistiche multivariate per ridurre la dimensionalità del problema.

Parole chiave: Raggruppamento inondazioni; Logica diffusa; *Fuzzy C-Means*; *K-means*; Analisi del Componente Principale.

LISTA DE FIGURAS

FIGURA 1 - LEITOS DE INUNDAÇÃO.	36
FIGURA 2 - NÚMERO DE EVENTOS DE INUNDAÇÃO NO MUNDO ENTRE 1986 E 2015.....	38
FIGURA 3 - EXEMPLO DE MÉTODO HIERÁRQUICO AGLOMERATIVO E DIVISIVO.....	41
FIGURA 4 - EXEMPLO DE EXECUÇÃO DO ALGORITMO DO MÉTODO <i>K-MEANS</i>	46
FIGURA 5 - EXEMPLO DE MÁ INICIALIZAÇÃO DO ALGORITMO <i>K-MEANS</i>	47
FIGURA 6 - EXEMPLO DE CONJUNTOS COM SUPERFÍCIE DE INTERSECÇÃO.	49
FIGURA 7 - AGRUPAMENTO DE DADOS UTILIZANDO MÉTODO <i>FUZZY C- MEANS</i>	49
FIGURA 8 - REPRESENTAÇÃO GRÁFICA DE ACP.	60
FIGURA 9 - MAPA DE SUSCETIBILIDADE À INUNDAÇÃODAS ÁREAS URBANAS DE UNIÃO DA VITÓRIA.....	68
FIGURA 10 - IMAGEM DE UNIÃO DA VITÓRIA 1983, ANTES DA ENCHENTE.	69
FIGURA 11 - IMAGEM DE UNIÃO DA VITÓRIA 1983, DURANTE A ENCHENTE. .	69
FIGURA 12 - LOCALIZAÇÃO DA BACIA HIDROGRÁFICA DO RIO IGUAÇU.	70
FIGURA 13 - LOCALIZAÇÃO DA CIDADE DE UNIÃO DA VITÓRIA.....	71
FIGURA 14 - CÁLCULO DO INSTANTE DE INÍCIO DO EXTRAVASAMENTO	76
FIGURA 15 - CÁLCULO DO INSTANTE DE FIM DO EXTRAVASAMENTO	76
FIGURA 16 - FLUXOGRAMA DO MÉTODO <i>K-MEANS</i>	85
FIGURA 17 - NÚMERO ÓTIMO DE GRUPOS.....	86
FIGURA 18 - REPRESENTAÇÃO GRÁFICA DO CÁLCULO DA DISTÂNCIA DE <i>MAHALANOBIS</i>	89
FIGURA 19 - AUTOVALORES E AUTOVETORES. VALORES NÃO TRATADOS. .	95
FIGURA 20 - SOMA DO QUADRADO DAS DISTÂNCIAS, EM FUNÇÃO DO NÚMERO DE GRUPOS <i>K</i>	98
FIGURA 21 - VALORES NORMALIZADOS DO INDICADOR I_1 (VAZÃO MÁXIMA ANUAL).....	105
FIGURA 22 - PARÂMETROS ESTATÍSTICOS DOS VALORES NORMALIZADOS DO INDICADOR I_1 (VAZÃO MÁXIMA ANUAL).....	105

FIGURA 23 - VALORES NORMALIZADOS DO INDICADOR I_2 (NÍVEL MÁXIMO ANUAL).....	106
FIGURA 24 - PARÂMETROS ESTATÍSTICOS DOS VALORES NORMALIZADOS DO INDICADOR I_2 (NÍVEL MÁXIMO ANUAL)	107
FIGURA 25 - VALORES NORMALIZADOS DO INDICADOR I_3 (VOLUME MÁXIMO ANUAL).....	108
FIGURA 26 - PARÂMETROS ESTATÍSTICOS DOS VALORES NORMALIZADOS DO INDICADOR I_3 (VOLUME MÁXIMO ANUAL)	108
FIGURA 27 - VALORES NORMALIZADOS DO INDICADOR I_4 (TEMPO ACIMA DO LIMAR DE EXTRAVASAMENTO).....	109
FIGURA 28 - PARÂMETROS ESTATÍSTICOS DOS VALORES NORMALIZADOS DO INDICADOR I_4 (TEMPO ACIMA DO LIMAR DE EXTRAVASAMENTO).	109
FIGURA 29 - VALORES NORMALIZADOS DO INDICADOR I_5 (TEMPO ENTRE O INÍCIO DO EVENTO E A ULTRAPASSAGEM DO LIMAR DE EXTRAVASAMENTO).	110
FIGURA 30 - PARÂMETROS ESTATÍSTICOS DOS VALORES NORMALIZADOS DO INDICADOR I_5 (TEMPO ENTRE O INÍCIO DO EVENTO E A ULTRAPASSAGEM DO LIMAR DE EXTRAVASAMENTO).....	110
FIGURA 31 - VALORES NORMALIZADOS DO INDICADOR I_6 (TAXA MÉDIA DE ASCENSÃO DO HIDROGRAMA).....	111
FIGURA 32 - PARÂMETROS ESTATÍSTICOS DOS VALORES NORMALIZADOS DO INDICADOR I_6 (TAXA MÉDIA DE ASCENSÃO DO HIDROGRAMA).	111
FIGURA 33 - VALORES NORMALIZADOS DO INDICADOR I_7	112
FIGURA 34 - PARÂMETROS ESTATÍSTICOS DOS VALORES NORMALIZADOS DO INDICADOR I_7 (TAXA MÉDIA DE RECESSÃO DO HIDROGRAMA).	112
FIGURA 35 – COMPARAÇÃO DAS COORDENADAS DO EVENTO 1997 E DOS CENTROIDES DO AGRUPAMENTO (MÉTODO <i>K-MEANS</i> E NORMALIZAÇÃO PELO MÉTODO DO DESVIO PADRÃO)	114
FIGURA 36 - VARIAÇÃO DOS VALORES NORMALIZADOS PARA O INDICADOR I_5 ORIGINAL E PARA O INDICADOR I_5 ALTERADO	117

FIGURA 37 - VARIAÇÃO DOS VALORES NORMALIZADOS PARA O INDICADOR I_7 ORIGINAL E PARA O INDICADOR I_7 ALTERADO	117
FIGURA 38 — VARIAÇÃO NO AGRUPAMENTO DAS AMOSTRAS EM VIRTUDE DO TIPO DE NORMALIZAÇÃO E DOS INDICADORES UTILIZADOS	129

LISTA DE TABELAS

TABELA 1 -	DEFINIÇÃO DOS EVENTOS HIDROLÓGICOS/HIDRÁULICOS.....	36
TABELA 2 -	REGISTROS DE INUNDAÇÃO NO BRASIL, PERÍODO DE 1940 A 2018.....	38
TABELA 3 -	FORMA DE APRESENTAÇÃO DOS RESULTADOS DA ANOVA.....	94
TABELA 4 -	CENTRÓIDES CALCULADOS PELO MÉTODO <i>K-MEANS</i>	99
TABELA 5 -	AGRUPAMENTO DAS CHEIAS HISTÓRICAS PELO MÉTODO <i>K-MEANS</i>	99
TABELA 6 -	CENTRÓIDES CÁLCULADOS PELO MÉTODO <i>FUZZY C-MEANS</i>	100
TABELA 7 -	AGRUPAMENTO DAS CHEIAS HISTÓRICAS PELO MÉTODO <i>FUZZY C-MEANS</i>	100
TABELA 8 -	DIFERENÇA (EM VALORES ABSOLUTOS) ENTRE AS COORDENADAS E A DISTÂNCIA EUCLIDIANA DOS GRUPOS OBTIDOS PELOS MÉTODOS <i>K-MEANS</i> E <i>FUZZY C-MEANS</i>	101
TABELA 9 -	EVENTOS COM CLASSIFICAÇÃO DISTINTA NOS AGRUPAMENTOS PELOS MÉTODOS <i>K-MEANS</i> E <i>FUZZY C-MEANS</i> (STEFFEN,2017) COM NORMALIZAÇÃO LINEAR.....	102
TABELA 10 -	DISTÂNCIA EUCLIDIANA ENTRE OS EVENTOS 1931, 1950 E 2011 E OS CENTROIDES DOS GRUPOS DEFINIDOS PELOS MÉTODOS <i>K-MEANS</i> E <i>FUZZY C-MEANS</i>	102
TABELA 11 -	CENTROIDES DOS INDICADORES NORMALIZADOS PELO DESVIO PADRÃO	114
TABELA 12 -	AMOSTRAS QUE TIVERAM ALTERAÇÕES NO AGRUPAMENTO: DADOS ORIGINAIS x DADOS NORMALIZADOS PELO MÉTODO DO DESVIO PADRÃO (MÉTODO <i>K-MEANS</i>)	114
TABELA 13 -	AMOSTRAS QUE TIVERAM ALTERAÇÕES NO AGRUPAMENTO: DADOS ORIGINAIS x DADOS NORMALIZADOS PELO MÉTODO DO VALOR MÁXIMO (MÉTODO <i>K-MEANS</i>)	115
TABELA 14 -	CENTROIDES DOS INDICADORES NORMALIZADOS PELO VALOR MÁXIMO	115
TABELA 15 -	AMOSTRAS QUE TIVERAM ALTERAÇÕES NO AGRUPAMENTO: DADOS ORIGINAIS x DADOS NORMALIZADOS PELO MÉTODO DA ESCALA DECIMAL (MÉTODO <i>K-MEANS</i>).....	116

TABELA 16 - CENTROIDES, INDICADORES NORMALIZADOS POR ESCALA DECIMAL	116
TABELA 17 - CENTROIDES DOS GRUPOS PARA OS VALORES DE I_5 E I_7 ALTERADOS, NORMALIZADOS POR NORMALIZAÇÃO LINEAR.	118
TABELA 18 - AMOSTRAS QUE TIVERAM ALTERAÇÕES NO AGRUPAMENTO. DADOS ORIGINAIS x INDICADORS I_5 E I_7 ALTERADOS (K-MEANS E NORMALIZAÇÃO LINEAR).	118
TABELA 19 - AUTOVALORES E AUTOVETORES. DADOS BRUTOS.....	119
TABELA 20 - AUTOVALORES E AUTOVETORES. NORMALIZAÇÃO LINEAR...	119
TABELA 21 - AUTOVALORES E AUTOVETORES. NORMALIZAÇÃO POR DESVIO PADRÃO, I_5 E I_7 ALTERADOS	120
TABELA 22 - AUTOVALORES E AUTOVETORES. NORMALIZAÇÃO EPLO VALOR MÁXIMO DOS ELEMENTOS E I_5 E I_7 ALTERADOS	120
TABELA 23 - AUTOVALORES E AUTOVETORES. NORMALIZAÇÃO POR ESCALA DECIMAL E I_5 E I_7 ALTERADOS.....	121
TABELA 24 - AGRUPAMENTO DOS DADOS DE TODAS AS AMOSTRAS, ATRAVÉS DE TODOS OS MÉTODOS DE NORMALIZAÇÃO E COMPARAÇÕES ANALISADAS NO TRABALHO.....	126
TABELA 24 - AGRUPAMENTO DOS DADOS DE TODAS AS AMOSTRAS, ATRAVÉS DE TODOS OS MÉTODOS DE NORMALIZAÇÃO E COMPARAÇÕES ANALISADAS NO TRABALHO (continuação).....	128
TABELA 25 - LEGENDA DA TABELA 24.....	127

LISTA DE ABREVIATURAS

ACP	- Análise de Componentes Principais;
ANA	- Agência Nacional de águas;
CEPED UFSC	- Centro Universitário de Estudos e Pesquisas sobre Desastres. Universidade Federal de Santa Catarina;
EIRD/ONU	- Estratégia Internacional para a Redução de Desastres da Organização das Nações Unidas;
EM-DAT	- <i>Emergency Disaster Data base</i> ;
FCM	- Algoritmo <i>Fuzzy C-means</i> ;
IBGE	- Instituto Brasileiro de geografia e estatística;
IPT	- Instituto de pesquisa tecnológica;
ISODATA	- <i>Iterative Self-Organizing Data Analysis Technique</i> ;
MIN	- Ministério da Integração Nacional;
MINC	- Ministério da Cultura;
OFDA/CRED	- <i>International Disaster Data</i> ;
RHN	- Rede Hidrometeorológica Nacional;
SEC-CORPRERI	- Sociedade de estudos contemporâneos, União da Vitória Paraná;
SEMA	- Secretaria de meio ambiente e infraestrutura;
SSE	- <i>Sum of Squared Error</i> ;
Wcss	- <i>Within clusters Sum of Squares</i> .

LISTA DE SÍMBOLOS E VARIÁVEIS

a	- Base do logaritmo, igual a 10. Determinação do número de grupos calculado através do algoritmo <i>Fuzzy C-means</i> ;
A	- Área de drenagem (km ²);
C	- Número de dados de um conjunto agrupado através do algoritmo <i>Fuzzy C-means</i> ;
c_j	- Centro de um agrupamento <i>Fuzzy</i> ($j = 1, 2, \dots, c$); algoritmo <i>Fuzzy C-means</i> ;
$d(p_i, X)$	- Distância entre um ponto p_i e um conjunto de clusters. Algoritmo <i>K-means</i> ;
$d(x_i; c_j)$	- Distância entre x_i e c_j ; algoritmo <i>Fuzzy C-means</i> ;
$f(x)$	- Função de normalização por desvio padrão;
F_c	- Média dos quadrados dos elementos da matriz u ; Determinação do número de grupos calculado através do algoritmo <i>Fuzzy C-means</i> ;
N	- Número de dados da amostra do no início do algoritmo aglomerativo;
X	- Distância entre um ponto e o centroide de um agrupamento;
K	- Centróides de agrupamento não hierárquico <i>K-means</i> ;
p_i	- Ponto num espaço vetorial;
N	- Número de clusters, agrupado através do algoritmo <i>Fuzzy C-means</i> , considerados no algoritmo, o qual deve ser decidido antes da execução;
M	- Parâmetro da fuzzificação. Usualmente, m está no intervalo de $[1,25;2]$; algoritmo <i>Fuzzy C-means</i> ;
x_i	- Vetor de dados de treinamento, onde $i = 1, 2, \dots, n$; algoritmo <i>Fuzzy C-means</i> ;
NC	- Número de maneiras para selecionar o centroide para cada agrupamento;
NK	- Número de maneiras para selecionar k centroides.
I	- Número de observações da amostra. Determinação do número de grupos calculado através do algoritmo <i>Fuzzy C-means</i> ;

I_1	- Vazão máxima anual;
I_2	- Nível de água máximo anual;
I_3	- Volume total extravasado;
I_4	- Tempo acima do limiar de extravasamento;
I_5	- Tempo entre o início do evento e a ultrapassagem do limiar de extravasamento;
I_6	- Taxas médias de subida;
I_7	- Taxas médias de recessão do hidrograma de cheia;
K	- Número de agrupamentos agrupados através dos Métodos particionais. Número de grupos. Determinação do número de grupos calculado através do algoritmo <i>Fuzzy C-means</i> ;
k_j	- Menor valor inteiro tal que máximo $ x_{ij} $ resulte ≥ 1 ; calculado através do método de normalização por escala decimal;
k_{max}	- Valor máximo da base de dados; calculado através do método de normalização pelo valor máximo dos elementos;
N	- número total de períodos acima do limiar de extravasamento de uma cheia; calculado no indicador I_4 (tempo total acima do limiar de extravasamento).
$t_{ef}(i)$	- Instante do fim do extravasamento do i-ésimo período acima do limiar de extravasamento do hidrograma de cheias; calculado nos indicadores I_4 (tempo total acima do limiar de extravasamento) e I_5 (tempo entre o início do evento e a ultrapassagem do limiar de extravasamento).
$t_{e0}(i)$	- Instante de início do extravasamento do i-ésimo período acima do limiar de extravasamento do hidrograma de cheia; calculado nos indicadores I_4 (tempo total acima do limiar de extravasamento) e I_5 (Tempo entre o início do evento e a ultrapassagem do limiar de extravasamento).
t_0	- Instante de início do hidrograma de cheia; calculado nos indicadores I_4 (tempo total acima do limiar de extravasamento) e I_5 (Tempo entre o início do evento e a ultrapassagem do limiar de extravasamento);

Q_0	- Vazão no instante de início do hidrograma de cheia (m^3/s); calculado no indicador I_6 (Taxa média de ascensão do hidrograma de cheia);
Q_{ref}	- Vazão de referência igual a $1.387 m^3/s$; calculado nos indicadores I_6 (Taxa média de ascensão do hidrograma de cheia) e I_7 (Taxa média de recessão do hidrograma de cheia);
$Q_{máx}$	- Vazão de pico (m^3/s); calculado no indicador I_7 (Taxa média de recessão do hidrograma de cheia);
$t_{máx}$	- Instante de ocorrência da vazão de pico do hidrograma de cheia; calculado no indicador I_7 (Taxa média de recessão do hidrograma de cheia).
t_{efpico}	- Instante de fim do extravasamento do período acima do limiar de extravasamento que contém a vazão de pico; calculado no indicador I_7 (Taxa média de recessão do hidrograma de cheia);
X	- Valor inicial utilizado nas normalizações;
X_{ij}	- Valor tratado da i-ésima observação do j-ésimo indicador; calculado no método de normalização linear. Valor tratado; calculado através do método de normalização pelo valor máximo dos elementos;
x^{0ij}	- Valor não tratado da i-ésima observação do j-ésimo indicador; calculado no método de normalização linear. Valor não tratado; calculado através do método de normalização pelo valor máximo dos elementos;
$X_{jmáx}$	- Máximo valor não tratado do j-ésimo indicador; calculado no método de normalização linear;
X_{jmn}	- Mínimo valor não tratado do j-ésimo indicador; calculado no método de normalização linear;
H_c	- Média dos quadrados dos elementos da matriz u multiplicados pelos seus respectivos logaritmos na base a . Determinação do número de grupos calculado através do algoritmo <i>Fuzzy C-means</i> ;
Δt_i	- Período de tempo acima do limiar de extravasamento do i-ésimo período acima do limiar de extravasamento do hidrograma de cheias

(dias); calculado no indicador I_4 (tempo total acima do limiar de extravasamento);

ε - Valor de convergência pretendido, definido previamente, utilizado como critério de parada no algoritmo *Fuzzy C-means*;

μ - Média;

σ - Desvio padrão.

SUMÁRIO

1	INTRODUÇÃO	29
1.1	PROBLEMA.....	31
1.2	OBJETIVOS	31
1.2.1	Objetivo geral	31
1.2.2	Objetivos específicos.....	32
1.3	JUSTIFICATIVA.....	32
1.4	ESTRUTURA DO TRABALHO	34
2	REVISÃO DE LITERATURA.....	34
2.1	CLASSIFICAÇÃO DE CHEIAS.....	35
2.2	IMPACTO DAS INUNDAÇÕES.....	37
2.3	INDICADORES DE INUNDAÇÃO	39
2.3.1	Características do canal	39
2.3.2	Área e forma da bacia hidrográfica	39
2.3.3	Impermeabilização e declividade do terreno.....	40
2.4	AGRUPAMENTO DE DADOS	40
2.4.1	Métodos hierárquicos.....	41
2.4.2	Método particionais.....	42
2.4.3	Comparação entre métodos de agrupamento de dados.....	42
2.4.4	Método de agrupamento não hierárquico <i>K-means</i>	45
2.4.5	Método de agrupamento não hierárquico <i>Fuzzy C-means</i>	48
2.4.6	Função de distância de <i>Mahalanobis</i>	50
2.4.7	Comparação entre os métodos de agrupamento <i>K-means</i> e <i>Fuzzy C-means</i>	51
2.5	ANÁLISE DE COMPONENTES PRINCIPAIS (ACP).....	54
2.5.1	Representação gráfica da Análise dos Componentes Principais (ACP).....	59
2.6	APLICAÇÕES DO AGRUPAMENTO DE DADOS EM HIDROLOGIA	60
3	MATERIAIS E MÉTODOS	64
3.1	HIDROGRAFIA DO ESTADO DO PARANÁ.....	64
3.2	ÁREA DE ESTUDO	65
3.2.1	Suscetibilidade a inundações em União da Vitória.....	67
3.2.2	Bacia hidrográfica do rio Iguaçu	69
3.3	MÉTODOS	72

3.3.1	Análise hidrológica preliminar	72
3.3.2	Definição dos eventos de cheias	73
3.3.3	Definição dos indicadores de cheia	77
3.3.4	Métodos de normalização dos dados	81
3.3.5	Implementação computacional do método <i>K-means</i>	84
3.3.6	Análise da influência do método de normalização no agrupamento das cheias.....	90
3.3.7	Determinação dos indicadores de cheia mais representativos	94
3.3.8	Redefinição do tratamento dos indicadores I_5 e I_7	96
4	ANÁLISE DOS RESULTADOS.....	97
4.1	COMPARAÇÃO DOS AGRUPAMENTOS FEITOS PELOS MÉTODOS <i>K-MEANS</i> E <i>FUZZY C-MEANS</i>	97
4.1.1	Determinação do número de grupos	97
4.1.2	Agrupamento pelo método <i>K-means</i>	98
4.1.3	Comparação entre os métodos <i>K-means</i> e <i>Fuzzy C-means</i>	100
4.2	INFLUÊNCIA DOS MÉTODOS DE NORMALIZAÇÃO DOS DADOS NO AGRUPAMENTO DE CHEIAS	103
4.2.1	Análise do efeito da normalização sobre os valores não tratados (dados observados).....	104
4.2.1.1	Normalização da vazão máxima anual (I_1).....	104
4.2.1.2	Normalização do nível de água máximo anual (I_2).....	106
4.2.1.3	Normalização do volume máximo anual (I_3).....	107
4.2.1.4	Normalização do tempo acima do limiar de extravasamento (I_4)	109
4.2.1.5	Normalização do tempo entre o início do evento e a ultrapassagem do limiar de extravasamento (I_5)	110
4.2.1.6	Normalização da taxa média de ascensão do hidrograma (I_6).....	111
4.2.1.7	Normalização da taxa média de recessão do hidrograma (I_7).....	112
4.2.2	Influência do método de normalização sobre o agrupamento das cheias pelo método <i>K-means</i>	113
4.2.2.1	Dados normalizados pelo método do desvio padrão e agrupados por <i>K-means</i>	113
4.2.2.2	Dados normalizados pelo método do valor máximo dos elementos e agrupados por <i>K-means</i>	115

4.2.2.3	Dados normalizados pelo método da escala decimal e agrupados por <i>K-means</i>	116
4.3	INFLUÊNCIA DA ALTERAÇÃO DOS INDICADORES I_5 E I_7 SOBRE O AGRUPAMENTO DE CHEIAS	116
4.4	REDUÇÃO DO NÚMERO DE INDICADORES DE CHEIA PELA ANÁLISE DE COMPONENTES PRINCIPAIS (ACP).....	118
4.4.1	Agrupamento de cheias pelo método <i>K-means</i> com redução do número de indicadores de cheias	121
4.4.2	Comparação dos resultados dos agrupamentos de cheias pelo método <i>k-means</i> com e sem redução do número de indicadores de cheia.....	121
4.4.2.1	Método <i>K-means</i> , normalização linear e indicador I_7 original.....	122
4.4.2.2	Método <i>K-means</i> , normalização linear e indicador I_7 alterado	122
4.5	COMPARATIVO DE TODAS AS ANÁLISES EXECUTADAS	123
5	CONCLUSÕES E RECOMENDAÇÕES	130
	REFERÊNCIAS.....	133
	ANEXO 1 – VAZÃO MÁXIMA ANUAL (I_1) E NÍVEL DE ÁGUA MÁXIMO ANUAL (I_2).....	140
	ANEXO 2 – VOLUMES TOTAIS EXTRAVASADOS (I_3).	141
	ANEXO 3 – TEMPO ACIMA DO LIMIAR DE EXTRAVASAMENTO (I_4).	142
	ANEXO 4 – TEMPOS DECORRIDOS ENTRE O INÍCIO DO EVENTO E ULTRAPASSAGEM DO LIMIAR DE EXTRAVASAMENTO (I_5).....	143
	ANEXO 5 – TAXA MÉDIA DE ASCENSÃO NO TRECHO ENTRE O INÍCIO DA ONDA DE CHEIA E O INSTANTE DE EXTRAVASAMENTO (I_6), ESTIMADOS POR STEFFEN (2017).	144
	ANEXO 6 – TAXA MÉDIA DE RECESSÃO DA ONDA DE CHEIA (I_7)... ..	145
	ANEXO 7 – MATRIZ DE INDICADORES DE CHEIA NORMALIZADOS PELO MÉTODO DA NORMALIZAÇÃO LINEAR.	146
	ANEXO 8 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORÍTMO <i>FUZZY C- MEANS</i>.	148
	ANEXO 9 – VALORES NORMALIZADOS DOS CENTROIDES DOS GRUPOS ENCONTRADOS PELO MÉTODO <i>FUZZY C - MEANS</i>.	149
	APÊNDICE 1 – MATRIZ DE INDICADORES DE CHEIA NORMALIZADOS OBTIDOS ATRAVÉS DO MÉTODO DO DESVIO PADRÃO.....	150

APÊNDICE 2 – MATRIZ DE INDICADORES DE CHEIA NORMALIZADOS OBTIDOS ATRAVÉS DO MÉTODO DA ESCALA DECIMAL.	153
APÊNDICE 3 – MATRIZ DE INDICADORES DE CHEIA NORMALIZADOS OBTIDOS ATRAVÉS DO MÉTODO DO VALOR MÁXIMO DOS ELEMENTOS.	156
APÊNDICE 4 – MATRIZ DE INDICADORES DE CHEIA COM DADOS DE I_5 E I_7 ALTERADOS.	159
APÊNDICE 5 – MATRIZ DE INDICADORES DE CHEIA NORMALIZADOS OBTIDOS ATRAVÉS DO MÉTODO DE NORMALIZAÇÃO LINEAR; VALORES DE I_5 E I_7 ALTERADOS.	161
APÊNDICE 6 – MATRIZ DE INDICADORES DE CHEIA NORMALIZADOS OBTIDOS ATRAVÉS DO MÉTODO DO DESVIO PADRÃO; VALORES DE I_5 E I_7 ALTERADOS.	163
APÊNDICE 7 – MATRIZ DE INDICADORES DE CHEIA NORMALIZADOS OBTIDOS ATRAVÉS DO MÉTODO DA ESCALA DECIMAL; VALORES DE I_5 E I_7 ALTERADOS.	165
APÊNDICE 8 – MATRIZ DE INDICADORES DE CHEIA NORMALIZADOS OBTIDOS ATRAVÉS DO MÉTODO DO VALOR MÁXIMO DOS ELEMENTOS; VALORES DE I_5 E I_7 ALTERADOS.	167
APÊNDICE 9 – VARIAÇÃO ESTATÍSTICA DO INDICADOR I_1 , NORMALIZADO ATRAVÉS DE NORMALIZAÇÃO LINEAR, NORMALIZAÇÃO POR DESVIO PADRÃO, NORMALIZAÇÃO POR ESCALA DECIMAL E NORMALIZAÇÃO POR VALOR MÁXIMO DO ELEMENTO.	169
APÊNDICE 10 – VARIAÇÃO ESTATÍSTICA DO INDICADOR I_2 , NORMALIZADO ATRAVÉS DE NORMALIZAÇÃO LINEAR, NORMALIZAÇÃO POR DESVIO PADRÃO, NORMALIZAÇÃO POR ESCALA DECIMAL E NORMALIZAÇÃO POR VALOR MÁXIMO DO ELEMENTO.	171
APÊNDICE 11 – VARIAÇÃO ESTATÍSTICA DO INDICADOR I_3 , NORMALIZADO ATRAVÉS DE NORMALIZAÇÃO LINEAR, NORMALIZAÇÃO POR DESVIO PADRÃO, NORMALIZAÇÃO POR ESCALA DECIMAL E NORMALIZAÇÃO POR VALOR MÁXIMO DO ELEMENTO.	173

APÊNDICE 12 – VARIAÇÃO ESTATÍSTICA DO INDICADOR I_4 , NORMALIZADO ATRAVÉS DE NORMALIZAÇÃO LINEAR, NORMALIZAÇÃO POR DESVIO PADRÃO, NORMALIZAÇÃO POR ESCALA DECIMAL E NORMALIZAÇÃO POR VALOR MÁXIMO DO ELEMENTO.....	175
APÊNDICE 13 – VARIAÇÃO ESTATÍSTICA DO INDICADOR I_5 , NORMALIZADO ATRAVÉS DE NORMALIZAÇÃO LINEAR, NORMALIZAÇÃO POR DESVIO PADRÃO, NORMALIZAÇÃO POR ESCALA DECIMAL E NORMALIZAÇÃO POR VALOR MÁXIMO DO ELEMENTO.....	177
APÊNDICE 14 – VARIAÇÃO ESTATÍSTICA DO INDICADOR I_6 , NORMALIZADO ATRAVÉS DE NORMALIZAÇÃO LINEAR, NORMALIZAÇÃO POR DESVIO PADRÃO, NORMALIZAÇÃO POR ESCALA DECIMAL E NORMALIZAÇÃO POR VALOR MÁXIMO DO ELEMENTO.....	179
APÊNDICE 15 – VARIAÇÃO ESTATÍSTICA DO INDICADOR I_7 , NORMALIZADO ATRAVÉS DE NORMALIZAÇÃO LINEAR, NORMALIZAÇÃO POR DESVIO PADRÃO, NORMALIZAÇÃO POR ESCALA DECIMAL E NORMALIZAÇÃO POR VALOR MÁXIMO DO ELEMENTO.....	181
APÊNDICE 16 – VARIAÇÃO ESTATÍSTICA DO INDICADOR I_5 ALTERADO NORMALIZADO ATRAVÉS DE NORMALIZAÇÃO LINEAR, NORMALIZAÇÃO POR DESVIO PADRÃO, NORMALIZAÇÃO POR ESCALA DECIMAL E NORMALIZAÇÃO POR VALOR MÁXIMO DO ELEMENTO.....	183
APÊNDICE 17 – VARIAÇÃO ESTATÍSTICA DO INDICADOR I_7 ALTERADO NORMALIZADO ATRAVÉS DE NORMALIZAÇÃO LINEAR, NORMALIZAÇÃO POR DESVIO PADRÃO, NORMALIZAÇÃO POR ESCALA DECIMAL E NORMALIZAÇÃO POR VALOR MÁXIMO DO ELEMENTO.....	185
APÊNDICE 18 – DISTÂNCIA DE <i>MAHALANOBIS</i> CALCULADA ENTRE O VETOR DE SEVERIDADE DO ANO DE 1931 E O VETOR DE BAIXA SEVERIDADE	187

APÊNDICE 19 – DISTÂNCIA DE <i>MAHALANOBIS</i> CALCULADA ENTRE O VETOR DE SEVERIDADE DO ANO DE 1931 E O VETOR DE MÉDIA SEVERIDADE	190
APÊNDICE 20 – CÁLCULO DO NÚMERO ÓTIMO DE AGRUPAMENTOS PARA O ALGORITMO <i>K-MEANS</i> . ATRAVÉS DO MÉTODO DO COTOVELO.	193
APÊNDICE 21 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORITMO <i>K-MEANS</i> ; DADOS ORIGINAIS NORMALIZADOS ATRAVÉS DO MÉTODO DA NORMALIZAÇÃO LINEAR; PROGRAMA EXECUTADO EM LINGUAGEM DE PROGRAMAÇÃO <i>PYTHON</i>	196
APÊNDICE 22 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORITMO <i>K-MEANS</i> ; DADOS ORIGINAIS NORMALIZADOS ATRAVÉS DO MÉTODO DO DESVIO PADRÃO; PROGRAMA EXECUTADO EM LINGUAGEM DE PROGRAMAÇÃO <i>PYTHON</i>	201
APÊNDICE 23 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORITMO <i>K-MEANS</i> ; DADOS ORIGINAIS NORMALIZADOS ATRAVÉS DO MÉTODO DA NORMALIZAÇÃO PELO VALOR MÁXIMO DO ELEMENTO; PROGRAMA EXECUTADO EM LINGUAGEM DE PROGRAMAÇÃO <i>PYTHON</i>	206
APÊNDICE 24 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORITMO <i>K-MEANS</i> ; DADOS ORIGINAIS NORMALIZADOS ATRAVÉS DO MÉTODO DA ESCALA DECIMAL; PROGRAMA EXECUTADO EM LINGUAGEM DE PROGRAMAÇÃO <i>PYTHON</i>	211
APÊNDICE 25 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORITMO <i>K-MEANS</i> ; DADOS DE I_5 E I_7 ALTERADOS. NORMALIZADOS ATRAVÉS DO MÉTODO DA NORMALIZAÇÃO LINEAR; PROGRAMA EXECUTADO EM LINGUAGEM DE PROGRAMAÇÃO <i>PYTHON</i>	216
APÊNDICE 26 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORITMO <i>K-MEANS</i> ; DADOS DE I_5 E I_7 ALTERADOS. NORMALIZADOS ATRAVÉS DO MÉTODO DO DESVIO PADRÃO; PROGRAMA EXECUTADO EM LINGUAGEM DE PROGRAMAÇÃO <i>PYTHON</i>	221

APÊNDICE 27 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORITMO <i>K-MEANS</i> ; DADOS DE I_5 E I_7 ALTERADOS. NORMALIZADOS ATRAVÉS DO MÉTODO DO VALOR MÁXIMO DO ELEMENTO; PROGRAMA EXECUTADO EM LINGUAGEM DE PROGRAMAÇÃO <i>PYTHON</i>	226
APÊNDICE 28 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORITMO <i>K-MEANS</i> ; DADOS DE I_5 E I_7 ALTERADOS. NORMALIZADOS ATRAVÉS DO MÉTODO DA ESCALA DECIMAL; PROGRAMA EXECUTADO EM LINGUAGEM DE PROGRAMAÇÃO <i>PYTHON</i>	231
APÊNDICE 29 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORITMO <i>K-MEANS</i> . CONSIDERANDO APENAS OS INDICADORES DE SEVERIDADE I_2 E I_7 . NORMALIZADOS ATRAVÉS DO MÉTODO DA NORMALIZAÇÃO LINEAR; PROGRAMA EXECUTADO EM LINGUAGEM DE PROGRAMAÇÃO <i>PYTHON</i>	236
APÊNDICE 30 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORITMO <i>K-MEANS</i> . CONSIDERANDO APENAS OS INDICADORES DE SEVERIDADE I_2 E I_7 . NORMALIZADOS ATRAVÉS DO MÉTODO DA NORMALIZAÇÃO LINEAR E COM VALORES DE I_7 ALTERADOS; PROGRAMA EXECUTADO EM LINGUAGEM DE PROGRAMAÇÃO <i>PYTHON</i>	241
APÊNDICE 31 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORITMO <i>K-MEANS</i> . CONSIDERANDO APENAS OS INDICADORES DE SEVERIDADE I_2 E I_7 . NORMALIZADOS ATRAVÉS DO MÉTODO DA NORMALIZAÇÃO POR DESVIO PADRÃO, E COM VALORES DE I_7 ALTERADOS; PROGRAMA EXECUTADO EM LINGUAGEM DE PROGRAMAÇÃO <i>PYTHON</i>	246
APÊNDICE 32 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORITMO <i>K-MEANS</i> . CONSIDERANDO APENAS OS INDICADORES DE SEVERIDADE I_2 E I_7 . NORMALIZADOS ATRAVÉS DO MÉTODO DO VALOR MÁXIMO DOS ELEMENTOS. E COM VALORES DE I_7 ALTERADOS; PROGRAMA EXECUTADO EM LINGUAGEM DE PROGRAMAÇÃO <i>PYTHON</i>	251

APÊNDICE 33 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORITMO <i>K-MEANS</i> . CONSIDERANDO APENAS OS INDICADORES DE SEVERIDADE I_2 E I_7 . NORMALIZADOS ATRAVÉS DO MÉTODO DA ESCALA DECIMAL E COM VALORES DE I_7 ALTERADOS; PROGRAMA EXECUTADO EM LINGUAGEM DE PROGRAMAÇÃO <i>PYTHON</i>	256
APÊNDICE 34 – TESTE <i>T DE STUDENT</i> , EXECUTADO PARA DUAS AMOSTRAS PAREADAS; VALORES ORIGINAIS.....	261
APÊNDICE 35 – TESTE <i>T DE STUDENT</i> , EXECUTADO PARA DUAS AMOSTRAS PAREADAS; VALORES DE I_5 E I_7 ALTERADOS.....	263
APÊNDICE 36 – ANÁLISE ESTATÍSTICA DE VARIÂNCIA (ANOVA), EXECUTADA PARA MULTIVARIÁVEIS; VALORES ORIGINAIS.....	265
APÊNDICE 37 – ANÁLISE ESTATÍSTICA DE VARIÂNCIA (ANOVA), EXECUTADA PARA MULTIVARIÁVEIS; VALORES DE I_5 E I_7 ALTERADOS.	266
APÊNDICE 38 – ANÁLISE DE COMPONENTES PRINCIPAIS APLICADO AOS DADOS NÃO NORMALIZADOS.....	267
APÊNDICE 39 – ANÁLISE DE COMPONENTES PRINCIPAIS APLICADO AOS DADOS NORMALIZADOS PELO MÉTODO DA NORMALIZAÇÃO LINEAR.	272
APÊNDICE 40 – ANÁLISE DE COMPONENTES PRINCIPAIS APLICADO AOS DADOS NORMALIZADOS PELO MÉTODO DO DESVIO PADRÃO.	276
APÊNDICE 41 – ANÁLISE DE COMPONENTES PRINCIPAIS APLICADO AOS DADOS NORMALIZADOS PELO MÉTODO DO VALOR MÁXIMO DOS ELEMENTOS.	280

1 INTRODUÇÃO

No primeiro quartil do século XXI, parece desnecessário discorrer sobre a importância da água no planeta. Contudo, uma análise detalhada do contexto global apresenta uma reflexão paradoxalmente contrária. Mesmo após uma considerável evolução comportamental da sociedade, a água continua sendo um insumo pouco valorizado em alguns pontos do planeta. A causa se deve à heterogeneidade de sua distribuição no globo. Nessa ótica cabe dissertar, de maneira introdutória, sobre a sua importância para os sistemas da biosfera terrestre, além das consequências da sua má gestão no contexto geopolítico mundial.

A água é a substância química mais importante para a vida no planeta. É consenso entre os pesquisadores que a vida teve início na água, foi nela que as primeiras formas biológicas surgiram e tiveram seu desenvolvimento até saírem da água e ganharem os continentes.

Estima-se que 97,5% da água existente no mundo é salgada e não é adequada ao nosso consumo direto, nem à irrigação da plantação. Dos 2,5% de água doce, a maior parte (69%) é de difícil acesso, pois está concentrada nas geleiras, 30% são águas subterrâneas (armazenadas em aquíferos) e 1% encontra-se nos rios. Logo, o uso desse bem precisa ser pensado para que não prejudique nenhum dos diferentes usos que ela tem para a vida humana (BRASIL; ANA, 2019).

A importância da água para a vida, associada às dificuldades para a sua captação, tratamento e distribuição, fez com que, historicamente, as populações mundiais predominantemente se fixassem às margens de grandes cursos de água, lagos ou mares. Isto propiciou o crescimento das sociedades, mas, por outro lado, as deixou suscetíveis à desastres naturais.

As sociedades humanas sofrem com catástrofes ocasionadas por inundações há milênios. Relatos de enchentes de magnitude calamitosa são encontrados nos livros sagrados das religiões mais praticadas no planeta. Exemplo disso é o grande dilúvio, relatado na Bíblia, na Torá judaica e no Alcorão Islâmico.

Estudo elaborado pelo Banco Mundial, em parceria com o Centro de Estudos e Pesquisas sobre Desastres Naturais da Universidade Federal de Santa Catarina (UFSC) em 2016, levantou os custos despendidos pelo Brasil causados por desastres naturais entre os anos de 1995 e 2014, chegando a um montante de R\$ 182,8 bilhões (CEPED UFSC; BANCO MUNDIAL, 2016). Segundo o relatório publicado pelas entidades:

Os desastres mais comumente reportados pelos municípios são aqueles do grupo climatológico, que são aqueles de evolução mais gradual, mais lenta, como chuva e estiagem. Eles representam 48% dos registros, e ocorrem com maior frequência nas regiões Nordeste e Sul do país. Em segundo lugar, vem o grupo hidrológico, que são os desastres relacionados ao excesso de chuvas. E eles ocorrem com maior frequência na região Sudeste do país, representando 39% dos registros da pesquisa. (CEPED UFSC; BANCO MUNDIAL, 2016, p. 34).

A variabilidade de precipitação pluvial na região sul do país é consideravelmente complexa, com atuação de diversas dinâmicas importantes, tais como: massas polares, sistemas convectivos, maritimidade e continentalidade. Com localização abaixo do Trópico de Capricórnio, em zona temperada, sendo influenciada pelo sistema de circulação perturbada de Sul, responsável pelas chuvas, principalmente no verão e pelo sistema de circulação perturbada de Oeste, que acarretam chuvas intensas (JORGE, 2015).

No estudo das ocorrências de inundações no sul do país, destacam-se os episódios ocorridos na cidade de União da Vitória, estado do Paraná.

A cidade de União da Vitória está localizada no sudoeste do estado do Paraná, divisa com o estado de Santa Catarina, e na separação entre o segundo e o terceiro planaltos Paranaense. Nessa área há uma mudança brusca nas características do solo e do relevo, a terra arenosa dá lugar a rocha de basalto impermeável. Além disso, ocorre a formação de uma barragem natural, ocasionada pelo estreitamento e elevação do curso do rio Iguaçu, formando um funil natural, cercado por rochas (CHMYZ, 1969).

Ao chegar a essa região, o rio Iguaçu tem a sua capacidade de descarga diminuída, ocasionando uma elevação do nível de água. Outra complicação é o fato da região ser uma grande planície. Logo, a água tende a sair facilmente do leito do rio e gerar grandes enchentes nos terrenos circunvizinhos (ARAÚJO, 2006).

Em virtude da recorrência dos eventos de inundação na região de União da Vitória e de todos os efeitos danosos para a sociedade, se faz necessário prever com antecedência tais episódios, uma vez que não se possui formas de anular as perdas.

Dentro desta lógica, podem ser utilizados métodos de agrupamento de cheias, com o objetivo de auxiliar no melhor entendimento de tais fenômenos, por meio da análise de variáveis que traduzem sua intensidade, visando à identificação de seus mecanismos de formação, e à modelagem hidrodinâmica e de cheias.

Devido à vasta gama de algoritmos de agrupamento de dados e metodologias para sua utilização, fator preponderante para obtenção de resultados fidedignos é a escolha do método de agrupamento. Cuidados adicionais devem ser tomados na fixação dos parâmetros de inicialização, bem como, dos critérios de parada. Respeitando-se as etapas descritas, será possível definir grupos de eventos definidos por indicadores de cheias, os quais podem, em trabalhos futuros, serem utilizados para análise e previsão de enchentes.

1.1 PROBLEMA

Dentro do contexto do uso de métodos de agrupamentos de cheias para análise hidrológica insere-se o presente trabalho; tendo como base o estudo feito por Steffen (2017) e Steffen e Gomes (2018), os quais utilizaram o algoritmo *Fuzzy C-means* como método de agrupamento de cheias do rio Iguaçu observados na cidade de União da Vitória.

Pretende-se responder as seguintes perguntas: qual a relevância do método (algoritmo) de agrupamento de dados escolhido na definição dos grupos dos eventos de cheias do rio Iguaçu observados na cidade de União da Vitória? Caso utilizemos o algoritmo *K-means*, chegaremos aos mesmos resultados de agrupamento obtido com o uso do algoritmo *Fuzzy C-means* ?

A hipótese adotada no presente estudo é que a mudança no algoritmo de agrupamento utilizado produzirá diferentes resultados, devido às diferenças existentes na métrica de inicialização, na formação dos grupos e no critério de parada dos algoritmos em questão.

1.2 OBJETIVOS

Nesta seção são apresentados os objetivos gerais e específicos deste trabalho.

1.2.1 Objetivo geral

O objetivo geral desta pesquisa foi comparar os resultados obtidos por dois métodos de agrupamento de dados, *Fuzzy C-means* e *K-means*, aplicados às cheias observadas no rio Iguaçu na cidade de União da Vitória.

1.2.2 Objetivos específicos

Como objetivos específicos deste trabalho, podem ser destacados:

- a) Aplicar o método *K-means* para as séries de dados de cheias observadas na cidade de União da Vitória, mediante os indicadores de cheia previamente selecionados por Steffen (2017);
- b) Verificar a aderência entre os resultados obtidos a partir do método *Fuzzy C-means*, usado por Steffen em 2017 e Steffen e Gomes (2018), e o método *K-means usado* no presente trabalho;
- c) Analisar criticamente os resultados obtidos através do algoritmo *K-means*, descrever as diferenças e semelhanças entre os agrupamentos obtidos através das duas metodologias (*Fuzzy C-means* e *K-means*);
- d) Executar o agrupamento de dados através do algoritmo *K-means* para diferentes métricas de normalização, analisando as diferenças encontradas.

1.3 JUSTIFICATIVA

O fato da água ser um insumo indispensável para a vida humana e as dificuldades na sua captação, distribuição e tratamento, fizeram com que os aglomerados urbanos ocupassem espaços próximos aos cursos de água, expondo-os diretamente a inundações catastróficas. A cidade de União da Vitória se insere neste contexto, sofrendo recorrentemente com as consequências destes fenômenos.

Por se tratar de desastres envolvendo causas naturais, que não podem ser evitados, deve haver elevado esforço em sua previsão e mitigação. Segundo Tucci e Bertoni (2003), a política existente no Brasil consiste em drenar a água precipitada, o mais rápido possível, e enviá-la para corpos receptores. Esse princípio foi abandonado nos países desenvolvidos no início da década de 1970. Projetos baseados neste contexto demandam montantes financeiros vultosos. Como exemplo, o custo da canalização do rio Tamanduateí, em São Paulo, foi de U\$50 milhões/km e o município de Curitiba reservou R\$134,8 milhões para obras de macrodrenagem em seu orçamento do ano de 2018, como forma de evitar as inundações ocorridas no verão de 2017. Projetos baseados no conceito de drenagem tem como consequência

imediate, apenas a transferência da enchente, fazendo com que ela ocorra em locais mais à jusante (BRASIL; ANA, 2019).

Eventos meteorológicos extremos e heterogêneos, ocorridos na década de 90, levando seca ao nordeste e severas inundações ao sul do país, levantaram dúvidas sobre a capacidade governamental em gerir um insumo tão importante quanto à água.

Neste contexto, em um esforço para regulamentar e traçar diretrizes eficazes para a minimização dos problemas relacionados aos recursos hídricos, o Brasil promulgou, em 1997, a Política Nacional de Recursos Hídricos, estabelecida pela Lei Nº 9.433, que especifica como um de seus objetivos (BRASIL, 1997): “*a prevenção e a defesa contra eventos hidrológicos críticos de origem natural ou decorrentes do uso inadequado dos recursos naturais*”.

Em julho de 2000, em decorrência da Política Nacional de Recursos Hídricos, visando a implementação do sistema nacional de recursos hídricos (SNRH), foi criada a Agência Nacional de Águas, que, entre outras diretrizes, é responsável pela coordenação da Rede Hidrometeorológica Nacional (RHN), um sistema que hoje abriga 4.641 pontos de monitoramento no país, divididos em estações que monitoram parâmetros relacionados aos rios, tais como: níveis, vazões, qualidade da água e transporte de sedimentos, além de outros que monitoram principalmente as chuvas (BRASIL, 2019).

O sistema se destina à concentração e apresentação de dados; contudo, não executa análises técnicas relativas à interpretação dos dados, que, caso necessária, deve ser obrigatoriamente executada pelo usuário.

Neste contexto está inserido o presente trabalho, que utiliza o algoritmo de agrupamento de dados *K-means*, visando verificar os agrupamentos de cheias formados, comparando os dados com o estudo realizado por Steffen (2017) e Steffen e Gomes (2018), que utilizaram o algoritmo *Fuzzy C-means* com o mesmo propósito.

A ideia central do trabalho é averiguar qual o impacto da técnica de agrupamento de dados utilizado no agrupamento de cheias.

O presente trabalho pretende colaborar para definição da técnica de agrupamento de dados a ser utilizada no agrupamento de eventos de enchente; assim como, o tipo de tratamento inicial a ser utilizado, visando a obtenção de resultados fidedignos que, futuramente, podem colaborar para previsão de eventos de cheias, com consequente redução dos danos sociais e econômicos do local em estudo, em específico, e de diferentes bacias hidrográficas, em geral.

1.4 ESTRUTURA DO TRABALHO

O capítulo 1 apresenta o contexto em que se insere o estudo de cheias, evidenciando sua importância no cenário nacional, bem como, a questão fundamental a ser respondida no desenvolvimento do trabalho. O capítulo 2 traz a revisão de literatura que concentra, basicamente: a classificação de cheias e técnicas de agrupamentos de dados. No capítulo 3, descreve-se o objeto de estudo e os métodos aplicados para se atingir os objetivos estabelecidos. No capítulo 4 são descritos os resultados obtidos a partir dos métodos apresentados anteriormente. O capítulo 5 mostra as conclusões do trabalho, além de recomendações para trabalhos futuros. Por fim, apresentam-se os apêndices referentes ao trabalho desenvolvido.

2 REVISÃO DE LITERATURA

Neste item é apresentada a revisão de literatura desenvolvida para os diversos tópicos relacionados ao estudo de cheias e inundações, bem como, para as técnicas possíveis de serem adotadas para o agrupamento de cheias.

Para a orientação do leitor, no presente trabalho, utilizou-se, sempre que possível, os termos cheia e enchente como sinônimos e relacionados estritamente ao fenômeno hidrológico da ocorrência de altas vazões. De forma diversa, o termo inundação foi utilizado para caracterizar aquelas situações nas quais a ocorrência de uma cheia ou enchente acarreta impactos que atingem também as dimensões social e econômica, portanto, não se restringe apenas à dimensão hidrológica, ou seja, não se restringe à condição de fenômeno natural.

De forma ilustrativa, o uso dos termos cheia, enchente e inundação, no presente trabalho, admite que pode haver uma cheia ou enchente sem necessariamente ocorrer uma inundação, nas situações em que as planícies de inundação não estão ocupadas. Em contrapartida, não se admite que possa haver uma inundação sem que tenha havido uma cheia ou enchente.

Em outras palavras, entendeu-se a inundação como uma possível consequência da ocorrência de uma cheia ou enchente e que o termo inundação não se aplica estritamente à ação de inundar, consequência física de elevação dos níveis de água, sendo usado quando a ação de inundar implica na ocorrência de impactos sociais e econômicos decorrentes da elevação do nível de água.

2.1 CLASSIFICAÇÃO DE CHEIAS

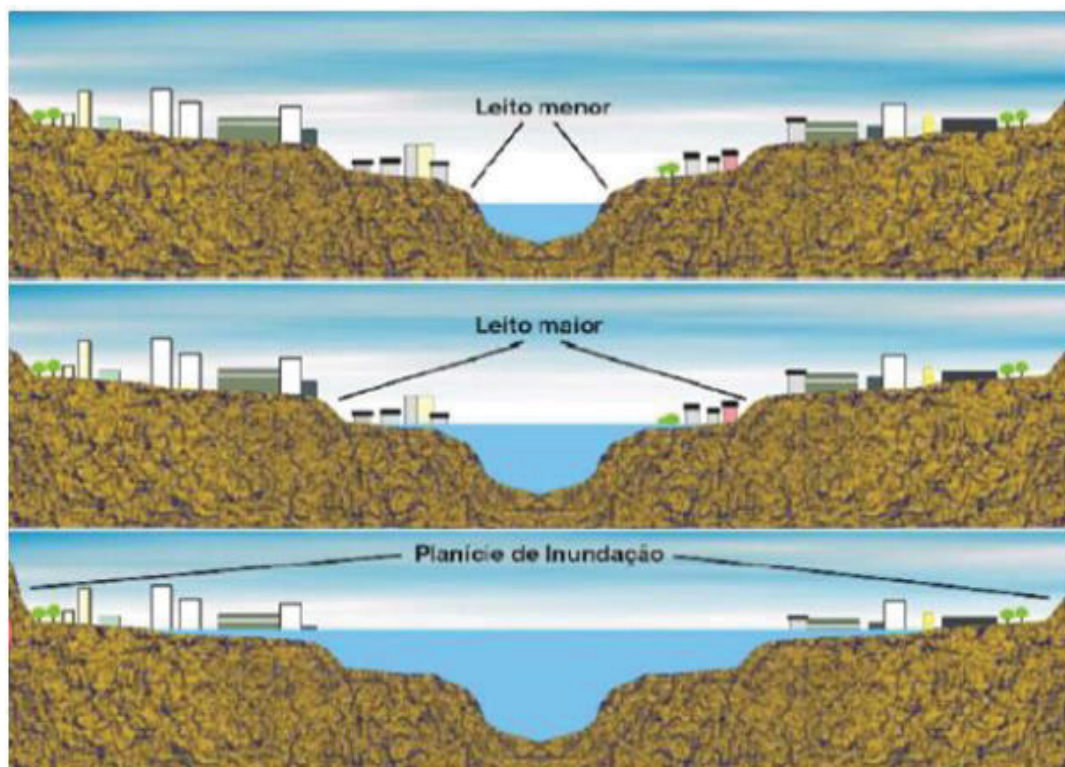
As inundações podem ser definidas como um transbordamento de água proveniente de rios, lagos e açudes (BRASIL, 2003). Segundo Tucci (2002), este é um processo natural que ocorre quando a precipitação é intensa, causada por eventos chuvosos, e o solo perde a capacidade de infiltração, aumentando o escoamento superficial, que rapidamente alcança os cursos d'água. As inundações podem ser deflagradas por chuvas rápidas e fortes, chuvas intensas de longa duração, degelo nas montanhas e outros eventos climáticos tais como furacões e tornados, sendo intensificados pelas alterações ambientais e intervenções urbanas de origem antrópica (BRASIL; IPT, 2007).

Fatores como uso do solo, quantidade de áreas impermeáveis, taxa de cobertura vegetal, retificação de córregos e rios, e outros fatores podem determinar a ocorrência ou não de uma inundação, bem como o seu tipo (GOERL; KOBIYAMA, 2008). A Figura 1 demonstra uma situação normal de inundação, desde o momento em que o rio se encontra dentro de seu leito normal até o preenchimento completo de toda a planície de inundação.

O Atlas Brasileiro de Desastres Naturais 1991 a 2012, desenvolvido pelo Centro de Estudos e Pesquisas em Engenharia e Defesa Civil (CEPED) da Universidade Federal de Santa Catarina (UFSC), especifica que as inundações podem ser classificadas em função da magnitude e da evolução através de dados comparativos de longo prazo, dividindo-se em (CEPED, 2016): inundações excepcionais; inundações de grande magnitude; inundações normais ou regulares e inundações de pequena magnitude. Em função da evolução, as inundações são classificadas em (CEPED, 2016): enchentes ou inundações graduais; enxurradas ou inundações bruscas; alagamentos e inundações litorâneas, provocadas pela brusca invasão do mar.

A Tabela 1 apresenta as principais definições sobre a nomenclatura de eventos hidrológicos/hidráulicos, comumente utilizada no Brasil. Contudo, cabe ressaltar que o País não apresenta uma definição única a respeito do tema inundação. A nomenclatura é estabelecida nos âmbitos federal, estadual e municipal, normalmente pela Defesa Civil, e podem divergir em alguns aspectos.

FIGURA 1 - LEITOS DE INUNDAÇÃO.



FONTE: Adaptado de Ministério de Minas e Energia (2017).

TABELA 1- DEFINIÇÃO DOS EVENTOS HIDROLÓGICOS/HIDRÁULICOS.

EVENTO HIDROLÓGICO/HIDRÁULICO	DEFINIÇÃO
1 – Enchentes ou inundações graduais	Nas enchentes, as águas elevam-se de forma paulatina e previsível, mantém-se em situação de cheia durante algum tempo e, a seguir, escoam-se gradualmente.
2 – Enxurradas ou inundações bruscas	As enxurradas são provocadas por chuvas intensas e concentradas, em regiões de relevo acidentado, caracterizando-se por provocar súbitas e violentas elevações dos caudais, os quais escoam de forma rápida e intensa. Nessas condições, ocorre um desequilíbrio entre o continente (leito de rio) e o conteúdo (volume caudal), provocando transbordamento.
3 – Alagamentos	São águas acumuladas no leito das ruas e nos perímetros urbanos por fortes precipitações pluviométricas em cidades com sistemas de drenagem deficientes. Nos alagamentos, o extravasamento das águas depende muito mais de uma drenagem deficiente, que dificulta a vazão das águas acumuladas, do que das precipitações locais.

2.2 IMPACTO DAS INUNDAÇÕES

De acordo com a Estratégia Internacional para a Redução de Desastres da Organização das Nações Unidas (EIRD; ONU, 2019), desastre é uma *“séria interrupção do funcionamento de uma comunidade ou sociedade, que causa perdas humanas e/ou importantes impactos ou perdas materiais, econômicas ou ambientais que excedem a capacidade da comunidade ou sociedade afetada de lidar com a situação utilizando seus próprios recursos”*.

Segundo Tucci e Bertoni (2003), desastre é definido como resultado de eventos adversos, naturais ou provocados pelo homem, sobre um ecossistema (vulnerável), causando danos humanos, materiais e/ou ambientais e consequentes prejuízos econômicos e sociais. Os desastres são normalmente súbitos e inesperados, de uma gravidade e magnitude capazes de produzir danos e prejuízos diversos, resultando em mortos e feridos. Portanto, exigem ações preventivas e restituidoras, que envolvem diversos setores governamentais e privados, visando uma recuperação que não pode ser alcançada por meio de procedimentos rotineiros.

Os desastres relacionados aos recursos hídricos geralmente estão associados ao excesso de água (inundações graduais e bruscas, rompimento de barragens) ou à sua escassez (estiagem, seca, dificuldades no abastecimento de água potável, impactos na agricultura). Além dos extremos, outros tipos de desastres podem envolver os recursos hídricos de diferentes maneiras, como impactos no ambiente, saúde pública, dinâmica urbana e produção agrícola (SALES, 2015).

O banco de dados *Emergency Database* (EM-DAT) é uma seleção de dados e informações sobre a ocorrência de desastres obtidos de diversas fontes governamentais e não governamentais. A Tabela 2 apresenta as estatísticas históricas de desastres causados por inundações ocorridas no Brasil, da década de 1940 até a atualidade, a partir de informações do EM-DAT.

A Figura 2 mostra o número de eventos de inundação ao redor do mundo, no período compreendido entre os anos de 1986 a 2015, podendo-se ter um comparativo entre o número de eventos ocorridos no Brasil e nos demais países.

Analisando a Tabela 2, é possível constatar a evolução na severidade dos eventos, uma vez que o maior número de eventos, assim como a maior taxa de mortalidade (1/3 do número de mortos), ocorreu na década compreendida entre os anos de 2009 e 2018.

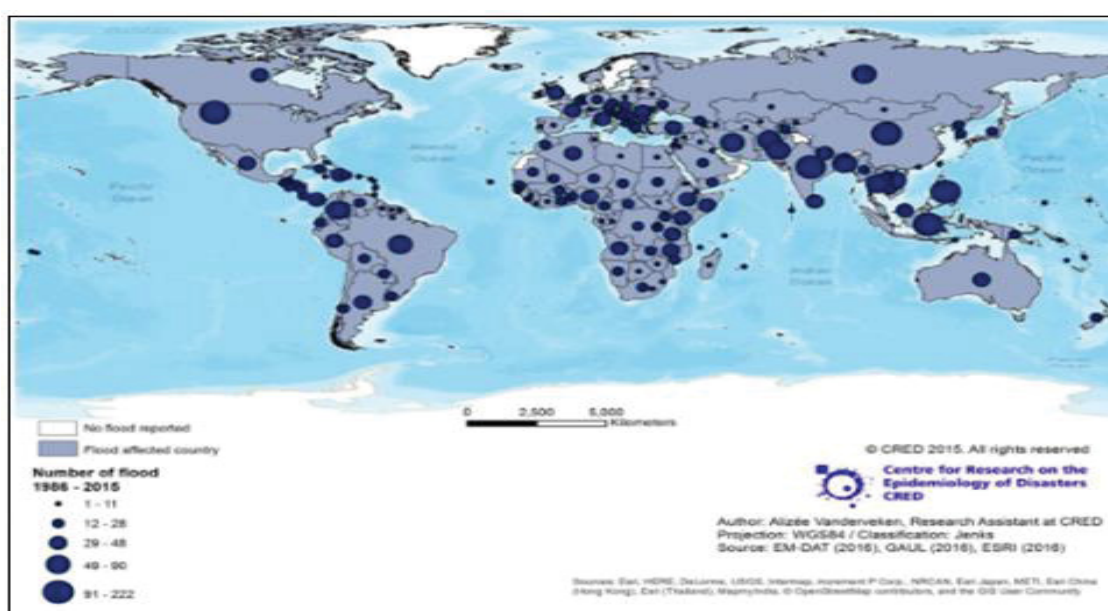
TABELA 2 - REGISTROS DE INUNDAÇÃO NO BRASIL, PERÍODO DE 1940 A 2018.

Período	Nº de Eventos	Nº de Mortes	Nº de Afetados (desabrigados/desalojados)
1940-1949	1	200	Sem registro
1950-1959	2	212	Sem registro
1960-1969	13	1818	825.986
1970-1979	11	1142	2.902.371
1980-1989	23	1598	8.789.613
1990-1999	20	386	317.793
2000-2008	27	776	2.466.592
2009-2018	39	2066	426.908

As informações contidas na Figura 2 colocam o Brasil no grupo de países com os maiores números de eventos de inundação registrados entre 1986 e 2015. Rafaeli Neto (2000) menciona que as inundações são inevitáveis, mas, por outro lado, pode-se reduzir significativamente os danos e prejuízos.

A frequência das inundações se altera devido a modificações na bacia hidrográfica. BARROS e HENRRQUES (1999) afirmaram que a pressão exercida pelo crescimento populacional resulta na exclusão da parcela mais pobre da população, que passa a viver na planície de inundação. Este tipo de problema é comumente presenciado nas áreas urbanas, razão pela qual devem ser introduzidos novos conceitos e práticas para uma melhor convivência com este fenômeno.

FIGURA 2 - NÚMERO DE EVENTOS DE INUNDAÇÃO NO MUNDO ENTRE 1986 E 2015.



FONTE: EMDAT/OFDA/CRED (2018).

2.3 INDICADORES DE INUNDAÇÃO

Os fatores que controlam as formas dos hidrogramas de cheia podem ser considerados como permanentes e temporários (AZEVEDO, 2014). Nos fatores temporários, tem-se as tempestades que podem variar as suas características em função da quantidade e tipo de precipitação. Nos fatores permanentes, os mais relevantes são as características do canal de escoamento, forma e relevo da bacia hidrográfica, além da impermeabilização e inclinação do terreno (WANG *et al.*, 2014).

2.3.1 Características do canal

Se o canal for constituído por uma superfície muito irregular, ocorrerá turbulência, com consequente retardamento do pico de cheia, devido à dissipação de energia, proporcionando menor velocidade de escoamento (STEPHENSON, 2002).

Se o canal for estreito, a probabilidade de ocorrência de cheia é maior, pois a seção da passagem da água é menor, originando velocidades superiores quando comparadas com um canal mais largo, que permite escoamentos mais lentos (STEPHENSON, 2002).

2.3.2 Área e forma da bacia hidrográfica

A área da bacia hidrográfica e a sua relação comprimento/largura têm influência nos valores dos caudais e no tempo que se demora a atingir o pico (STEPHENSON, 2002).

Em relação à sua forma em planta, as bacias podem apresentar (MACIEL, 2017):

- Forma alongada: constituída por um talvegue principal, por onde passa o curso de água mais importante, e afluentes de menor relevância. Para uma chuva intensa, existe um aumento rápido no volume de água do curso hídrico, mas depois existe um período estacionário onde o fluxo do caudal se mantém quase constante;
- Forma arredondada: formada por cursos hídricos com comprimento e relevância semelhante, o que tende a aumentar a probabilidade de

ocorrência de enchentes, visto que os tempos de concentração são semelhantes;

- Forma ramificada: formada por uma série de bacias alongadas reunidas em conjunto num mesmo tronco. A ocorrência de uma chuva em toda a bacia origina cheias em todas as sub bacias que vão se somar, mas não simultaneamente. Esse fenômeno pode ocasionar vários picos do hidrograma em uma mesma cheia.

2.3.3 Impermeabilização e declividade do terreno

A impermeabilização do terreno influencia o escoamento no sentido de acelerá-lo. Uma zona com vegetação tem maior capacidade de retenção e absorção da água, servindo como barreira à sua passagem. Já uma zona menos permeável, como uma cidade, onde existem construções, não oferece tanta resistência ao escoamento, fazendo com que este se dê com maior velocidade, logo, diminuindo o tempo até que o fluxo atinja o caudal, ocasionado picos de cheias maiores, porém com menor duração (BARROS; HENRRQUES, 1999).

A declividade interfere no sentido de que um terreno mais inclinado permite maiores velocidades de escoamento superficial, permitindo que o evento de cheia aconteça mais rapidamente (MACIEL, 2017).

Relacionando os fatores declividade do terreno e impermeabilização, conclui-se que os locais mais propícios à ocorrência de cheias são aqueles onde existem terrenos inclinados e impermeáveis.

2.4 AGRUPAMENTO DE DADOS

Métodos de agrupamento de dados são técnicas utilizadas para analisar, categorizar e agrupar dados (ANDERBEG, 1973). A principal razão para existência de diferentes métodos de agrupamento é o fato de que a noção de grupos não é precisamente definida (ZHAO; KARYPIS, 2002). Como consequência, diferentes métodos de agrupamento foram desenvolvidos, cada um dos quais usando um princípio de formação diferente.

Fraley e Raftery (1998) sugerem dividir os métodos de agrupamento em duas categorias principais: métodos hierárquicos e métodos de particionamento (particionais).

2.4.1 Métodos hierárquicos

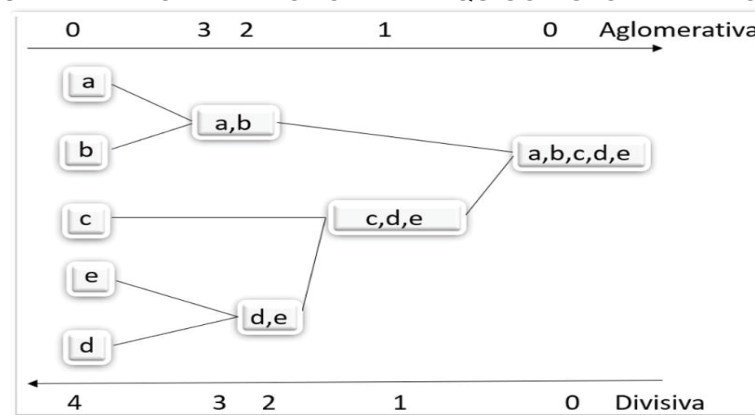
Rokach (2005) explica que esses métodos constroem os grupos (*clusters*) particionando recursivamente as instâncias de cima para baixo ou de baixo para cima. Desta forma, esses métodos podem ser subdivididos da seguinte forma:

- agrupamento hierárquico aglomerativo: cada objeto representa inicialmente um grupo (*cluster*) próprio. Em seguida, os grupos são sucessivamente fundidos até que a estrutura de grupo desejada é obtida.
- agrupamento hierárquico divisivo: todos os objetos pertencem inicialmente a um único grupo (*cluster*). Em seguida, o grupo inicial é dividido em subgrupos que são novamente subdivididos criando os seus próprios subgrupos. Este processo é feito de forma sucessiva até que a estrutura de grupo desejada é obtida.

A representação gráfica dos métodos hierárquicos é um dendrograma, que mostra os sucessivos agrupamentos de objetos em função dos níveis de similaridade para os quais os agrupamentos se formam. A estrutura de grupo desejada é obtida seccionando o dendrograma na posição desejada de nível de similaridade (ROKACH, 2005).

A Figura 3 apresenta um exemplo de dendrograma que ilustra o funcionamento dos agrupamentos hierárquicos aglomerativo e divisivo.

FIGURA 3 - EXEMPLO DE MÉTODO HIERÁRQUICO AGLOMERATIVO E DIVISIVO.



FONTE: Adaptado de Kaufman e Rousseeuw (2005)².

² KAUFMAN, L.; ROUSSEEUW, P. J. *Finding groups in data: an introduction to cluster analysis*. Páginas 205, 206, 246, 257 e 258 (2005).

2.4.2 Método particionais

Métodos de agrupamento particionais estabelecem um agrupamento constituído de k grupos (*clusters*) de um conjunto de n objetos diretamente a partir de uma métrica de comparação, que geralmente é definida com base em uma distância. Portanto, de forma simples, nos métodos particionais em distinção aos métodos hierárquicos, o número de grupos é definido previamente ao próprio processo de agrupamento.

O objetivo principal dos métodos particionais é separar grandes grupos em k grupos menores (ZHAO; KARYPIS, 2002).

2.4.3 Comparação entre métodos de agrupamento de dados

O agrupamento de dados pode ser formulado como um problema de otimização com múltiplos objetivos, onde o algoritmo a se escolher e seus parâmetros, como, por exemplo, função de distância, limiar de densidade ou número esperado de grupos, dependem dos dados e do tipo de resultado procurado (KRONBAUER *et al.*, 2016).

Everitt (2001) apresenta as seguintes definições relativas a um agrupamento de dados:

- conjunto de entidades que são iguais, porém, as entradas de diferentes agrupamentos de dados não são iguais;
- agregação de pontos no espaço de teste, de modo que a distância entre quaisquer dois pontos no agrupamento seja menor que a distância entre qualquer ponto no agrupamento e qualquer ponto que não esteja nele;
- podem ser descritos como regiões conectadas de um espaço multidimensional, contendo uma densidade relativamente alta de pontos, separados de outras regiões, por uma região contendo uma densidade relativamente baixa de pontos.

Quanto à classificação dos métodos, pode-se citar a proposta feita por Jain e Dubes (1999); nela os métodos são classificados de acordo com o algoritmo adotado para definir os grupos, sendo eles: hierárquicos e particionais (não hierárquicos).

Nievola (2015) descreve como principais vantagens dos métodos hierárquicos:

- a) produz um conjunto de grupos aninhados organizado como uma árvore hierárquica;
- b) pode ser visualizado como um dendograma (tipo de árvore que registra a sequência de uniões ou divisões);
- c) não há necessidade de se assumir qualquer número particular de grupos.

Ainda segundo Nievola (2015), as principais limitações dos agrupamentos hierárquicos são:

- a) uma vez que uma decisão tenha sido tomada de combinar dois grupos, ela não pode ser desfeita;
- b) nenhuma função objetivo é minimizada diretamente;
- c) sensibilidade a ruído e *outliers*;
- d) dificuldade de manipular grupos de tamanhos diferentes e formatos convexos;
- e) fragmentação de grandes grupos.

Ao contrário dos métodos hierárquicos, que produzem uma série de agrupamentos relacionados, os métodos particionais (não hierárquicos) produzem agrupamentos simples (ANDERBERG, 1973).

Os algoritmos não hierárquicos de agrupamento de dados dividem um conjunto de dados em um número apropriado de subconjuntos. É uma técnica amplamente utilizada em diversas aplicações devido à sua simplicidade, facilidade de implementação e rápida convergência (LACHI; ROCHA, 2005).

Um algoritmo não hierárquico de agrupamento de dados realiza uma busca para determinar um ponto que represente cada uma das partições, esse ponto é o centro de massa da partição (centroide). Após a definição dos centroides, os dados são comparados com os centroides e são agrupados com a partição de maior similaridade. Depois de todos os dados estarem agrupados, o centroide é recalculado e o processo se repete até atingir determinado critério de parada (MATTEUCCI, 2005).

Matteucci (2005) apresenta como as principais vantagens da utilização dos métodos particionais (não hierárquicos):

- a) técnica de agrupamento mais popular, devido à simplicidade;
- b) algoritmo básico e eficiente com etapas simples de cálculo;
- c) operação chave é o cálculo da proximidade entre dois grupos;
- d) diferentes abordagens para definir a distância entre grupos distinguem os diferentes algoritmos;
- e) similaridade entre dois grupos é baseada nos dois pontos mais similares (próximos) nos diferentes grupos.

Nievola (2015) descreve as principais limitações de agrupamentos particionais (não hierárquicos):

- a) Se há k grupos reais, então a chance de selecionar um centroide de cada grupo é pequena;
- b) Se k é grande, a chance selecionar um centroide de cada grupo é relativamente pequena;
- c) Se os k grupos são de mesmo tamanho (n), então a chance de selecionar um centroide de cada grupo é dada por:

$$P = \frac{NC}{NK} = \frac{k! \frac{n^k}{k!}}{(k \frac{n}{k})^k} = \frac{k!}{k^k}$$

Onde: NC = número de maneiras de selecionar um centroide para cada grupo;

NK = número de maneiras de selecionar k centroides.

P = chance de selecionar um centróide de cada grupo.

- d) os grupos têm sua formação impactada pelos centroides inicialmente escolhidos;
- e) algoritmo básico do *K-means* pode levar a grupos vazios.

Segundo Matteucci (2015), de modo a minorar possíveis problemas ocasionados pela inicialização dos centroides, deve-se tomar os seguintes cuidados:

- a) selecionar mais que k centroides iniciais e então, entre estes centroides iniciais, selecionar os mais largamente separados;
- b) utilizar alguma das metodologias conhecidas para determinar o número correto de grupos;
- c) Se há vários grupos vazios, o procedimento acima pode ser repetido várias vezes;
- d) uma alternativa é atualizar os centroides após cada assinalamento (abordagem incremental).

2.4.4 Método de agrupamento não hierárquico *K-means*

O método de agrupamento *K-means* é uma técnica de agrupamento não hierárquico que busca minimizar a distância dos elementos a um conjunto de k centroides, representado por $\chi = \{x_1, x_2, \dots, x_k\}$ de forma iterativa (LINDEN, 2009).

A distância entre um ponto p_i e o conjunto de k centroides, dada por $d(p_i, \chi)$, é definida como sendo a distância do ponto ao centroide mais próximo dele. Geralmente, o método utiliza a distância Euclidiana. Contudo, no presente trabalho utilizou-se a distância de *Mahalanobis*, descrita de modo detalhado no item 2.4.6 do presente capítulo.

O algoritmo do método exige a definição *a priori* do parâmetro K (número de grupos). Isto costuma ser um problema, uma vez que normalmente não se sabe previamente o número de grupos existentes (LINDEN, 2009).

Ainda segundo Linden (2009), o algoritmo do método *K-means* pode ser descrito da seguinte maneira:

- 1) escolher k posições iniciais para os centroides dos K grupos (possivelmente, de forma aleatória);
- 2) associar cada elemento da amostra ao grupo cujo centroide é mais próximo do referido elemento;
- 3) recalcular a posição do centroide de cada grupo;
- 4) repetir os passos 2 e 3 até que nenhum elemento da amostra mude de grupo entre duas iterações sucessivas, em razão da mudança da posição dos centroides.

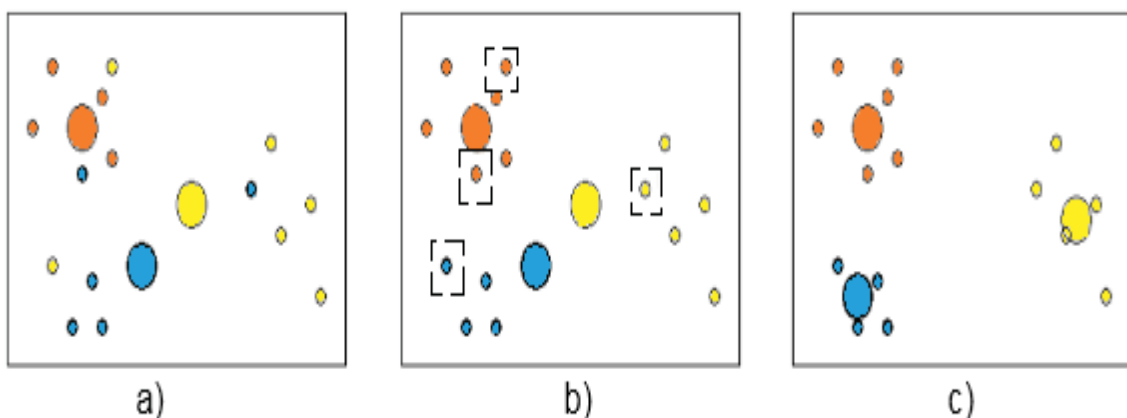
O algoritmo descrito apresenta uma rápida convergência, exigindo geralmente poucas iterações para a obtenção de uma configuração estável, na qual nenhum elemento da amostra estará associado a um grupo cujo centroide não lhe seja o mais próximo.

Um exemplo da execução do algoritmo de *K-means* pode ser visto na Figura 4, adaptada de Linden (2009), descrito resumidamente a seguinte forma:

- a) cada elemento da amostra foi designado de forma aleatória para um dos três grupos e os centroides (círculos maiores) de cada grupo foram então calculados;
- b) com a definição da posição dos centroides, os elementos da amostra podem ser associados aos grupos cujos centroides lhe estão mais próximos;
- c) uma nova posição dos centroides foi calculada, o que permite uma nova associação dos elementos da amostra aos grupos cujos centroides lhe estão mais próximos.

No exemplo da Figura 4, os grupos já estão em sua forma final. Caso não estivessem, repetem-se os passos b) e c) até que não nenhum elemento da amostra mude de grupo como consequência da mudança de posição dos centroides.

FIGURA 4 - EXEMPLO DE EXECUÇÃO DO ALGORITMO DO MÉTODO *K-MEANS*.



FONTE: Adaptado de Linden (2009).

Como mencionado anteriormente, em geral, se utiliza, como métrica de comparação para a determinação dos grupos, a distância Euclidiana entre cada

elemento e o centroide do grupo mais próximo. Novamente, destaca-se que, no presente trabalho, utilizou-se a distância de *Mahalanobis*, dada pela Equação (1).

$$D_{ab}^2 = [x_a - x_b]^T S^{-1} [x_a - x_b] \quad (1)$$

Onde: D_{ab}^2 = distância de *Mahalanobis* entre os indivíduos a e b ;

x_a = vetor de características do indivíduo a ;

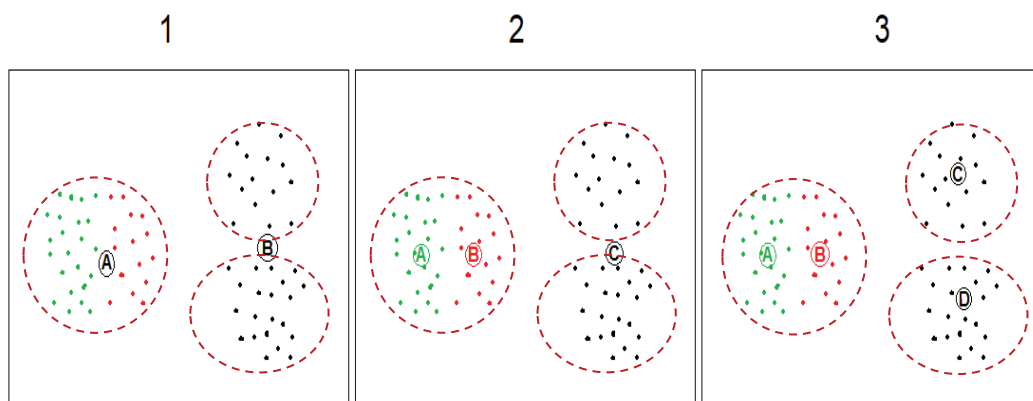
x_b = vetor de características do indivíduo b ;

S = matriz de covariância amostral.

Um eventual problema do método *K-means* é que a minimização de $d(p_i, \chi)$ valoriza a questão da homogeneidade e ignora a importante questão da boa separação dos grupos (clusters). Isto pode causar uma má separação dos conjuntos no caso de uma má inicialização dos centroides, inicialização esta que é feita de forma arbitrária (aleatória) no início da execução (LINDEN, 2009).

A Figura 5 apresenta um exemplo do efeito de uma má escolha do número de grupos e de uma má inicialização na execução do método *K-means*. Destaca-se que os elementos da amostra se agrupam naturalmente em três grupos, circundados em pontilhado nos Quadros 1, 2 e 3 da Figura 5.

FIGURA 5 - EXEMPLO DE MÁ INICIALIZAÇÃO DO ALGORITMO *K-MEANS*.



FONTE: Adaptado de Linden (2009)

Os quadros 1 e 3 da Figura 5, para os quais foram definidos *a priori* dois e quatro grupos, respectivamente, ilustram o efeito da escolha inadequada do número de grupos, uma vez que o algoritmo do método *K-means*, por si próprio, não é capaz

de identificar se a escolha do número de grupos foi adequada. No quadro 1, o algoritmo identifica os dois grupos do lado direito da figura como um grupo único, enquanto no quadro 3, o algoritmo subdivide o grupo do lado esquerdo da figura em dois grupos distintos.

Já o Quadro 2 da Figura 5 ilustra o efeito da má inicialização dos centroides, já que o número de grupos foi escolhido corretamente, mas o algoritmo do método convergiu para uma solução que não representa de forma adequada a classificação dos elementos da amostra em três grupos. O resultado do algoritmo levou a considerar o grupo do lado esquerdo da figura como dois grupos distintos e os dois grupos do lado direito da figura como um grupo único. Nesta situação, o problema é que o terceiro grupo, localizado do lado esquerdo da figura, é bem separado dos outros dois grupos. Sendo assim, se dois centroides e respectivos grupos forem inicializados do lado esquerdo da figura, tenderão a permanecer no referido lado, não havendo migração de um dos centroides e, por consequência, de um dos grupos para o lado direito da figura.

2.4.5 Método de agrupamento não hierárquico *Fuzzy C-means*

O método de agrupamento não hierárquico *Fuzzy C-means* (FCM) é uma adaptação do algoritmo *K-means*, sendo que sua versão final foi proposta por James C. Bezdek em 1973 (BEZDEK *et al.*, 1992). O método foi o primeiro a utilizar o elemento fuzzificador, que determina o grau de fuzzificação entre os elementos da amostra, considerando que, quanto maior o seu valor, mais os elementos serão relacionados.

A função que associa cada elemento a um grau de relevância ou pertencimento é chamada de função de pertinência. Esta função é normalizada assumindo valores reais no intervalo entre 0 e 1, sendo descrita por $\mu_A: X \rightarrow [0,1]$, onde A é o conjunto nebuloso e X é o universo das amostras.

O algoritmo se propõe a resolver um problema comum na abordagem *K-means*, que se refere à intersecção de conjuntos. Quando tal fenômeno ocorre, por vezes, fica difícil definir o grau de pertencimento dos elementos inscritos na região da intersecção a cada um dos conjuntos previamente escolhidos.

A Figura 6 mostra um exemplo de 2 conjuntos com uma superfície de intersecção entre eles e a Figura 7 mostra o resultado do agrupamento gerado pelo

FCM, no qual os dados internos à região da intersecção foram divididos em dois grupos (Grupo A e Grupo B).

FIGURA 6 – EXEMPLO DE CONJUNTOS COM SUPERFÍCIE DE INTERSECÇÃO.

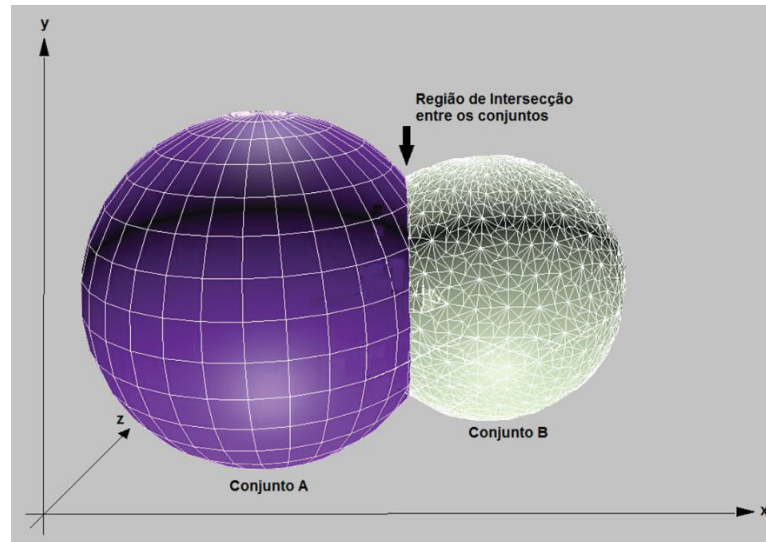
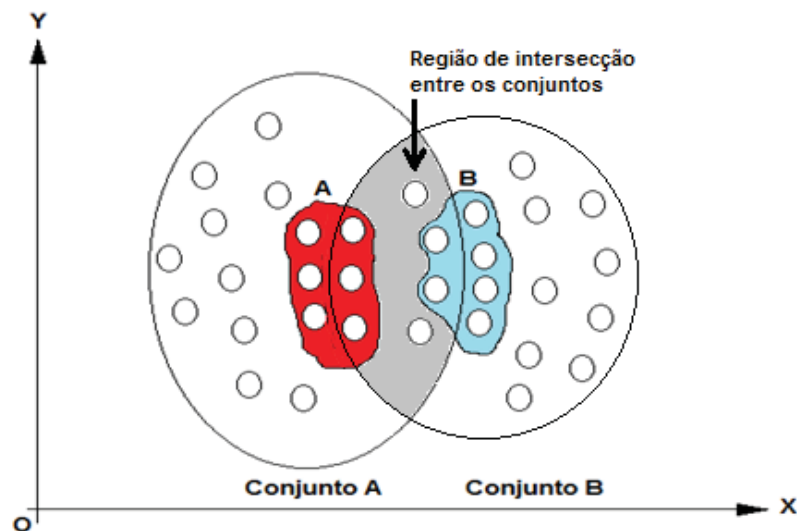


FIGURA 7 - AGRUPAMENTO DE DADOS UTILIZANDO MÉTODO *FUZZY C- MEANS*.



FONTE: Adaptado de COX (2005).

A partir da Figura 7, é possível identificar que os elementos posicionados na intersecção entre os conjuntos (região sombreada) pertencem simultaneamente aos grupos A e B, porém, com graus de pertinência diferente. Um objeto posicionado na região sombreada, mas destacado em fundo vermelho, pertence mais ao grupo A do que ao grupo B. De modo contrário, um objeto posicionado na região sombreada, mas destacado em fundo azul, pertence mais ao grupo B do que ao grupo A.

O método *Fuzzy C-means* pode ser equacionado por meio de um algoritmo iterativo, baseado na minimização de um índice de desempenho, que indica a adequabilidade da pseudopartição gerada (YONAMINE *et al.*, 2002).

O algoritmo do método *Fuzzy C-means* possui funcionamento e estrutura semelhantes ao *K-means*, porém, o cálculo da posição dos centroides utiliza a função de pertinência, que é atualizada constantemente até a convergência do algoritmo.

Os dados de entrada do algoritmo são os n dados, o número de grupos k e a aproximação inicial da matriz de pertinência u . Os passos que definem o algoritmo do método *Fuzzy C-means* são (YONAMINE *et al.*, 2002):

- a) definir o número C de grupos;
- b) estabelecer a posição inicial dos C centroides de forma aleatória;
- c) calcular a matriz de pertinência u , conforme Equação (2):

$$u_{ij} = \frac{\left(\frac{1}{d(x_i; c_j)}\right)^{\frac{2}{m-1}}}{\sum_{k=1}^C \left(\frac{1}{d(x_i; c_k)}\right)^{\frac{2}{m-1}}} \quad (2)$$

Onde $d(x_i; c_j)$ é a distância do elemento i ao centroide j e m é qualquer número real maior que 1;

- d) calcular a nova posição dos C centroides, utilizando a matriz de pertinência obtida no passo anterior, conforme Equação (3):

$$c_j = \frac{\sum_{i=1}^n u_{ij}^m x_i}{\sum_{i=1}^n u_{ij}^m} \quad (3)$$

- e) retornar às etapas c) e d) até que não haja mais modificação na posição dos centroides e dos elementos dos grupos.

2.4.6 Função de distância de *Mahalanobis*

Em estatística, a distância de *Mahalanobis* é uma medida de distância introduzida pelo matemático indiano *Prasanta Chandra Mahalanobis* em 1936, baseada nas correlações entre variáveis com os quais distintos padrões podem ser identificados e analisados. É uma estatística útil para determinar a similaridade entre

uma amostra desconhecida e uma conhecida. Distingue-se da distância Euclidiana já que utiliza as correlações do conjunto de dados e é invariante à escala, ou seja, não depende da escala das medições (YURAS, 2019).

Os dados em ambos os grupos a comparar deverão ter o mesmo número de variáveis (ou seja, o mesmo número de colunas) mas não necessariamente o mesmo número de elementos (o número de linhas pode ser diferente) (YURAS, 2019).

Formalmente a distância de *Mahalanobis* (D_M) entre um grupo de valores com média $\mu = (\mu_1, \mu_2, \mu_3, \dots, \mu_p)^T$ e matriz de covariância S para um vetor multivariado $\vec{x} = (x_1, x_2, x_3, \dots, x_p)^T$ é definida como (MOORE, 1999):

$$D_M(x) = \sqrt{(\vec{x} - \mu)^T S^{-1} (\vec{x} - \mu)} \quad (4)$$

A distância de *Mahalanobis* pode também definir-se como uma média de dissimilaridade entre dois vetores aleatórios $\vec{x}$ e $\vec{y}$ com mesma distribuição e com matriz de covariância S (BEZDEK *et al.*, 1992):

$$d(\vec{x}, \vec{y}) = \sqrt{(\vec{x} - \vec{y})^T S^{-1} (\vec{x} - \vec{y})} \quad (5)$$

Se a matriz de covariância (S) é a matriz identidade, a distância de Mahalanobis coincide com a distância Euclidiana. Se a matriz de covariância é diagonal, então a média de distância resultante é chamada distância Euclidiana normalizada (MOORE, 1999).

2.4.7 Comparação entre os métodos de agrupamento *K-means* e *Fuzzy C-means*

Segundo Fung (2001), o método de agrupamento *K-means* é um dos métodos mais populares das técnicas particionais. Ainda segundo Fung (2001), o método foi estabelecido independentemente em diferentes campos científicos, primeiramente por Steinhaus (1956), Ball e Hall (1965), Macqueen (1967) e Lloyd (1982) (na verdade, proposto em 1957, mas publicado somente em 1982). Mesmo tendo sido proposto há mais de 50 anos, ainda é um dos métodos mais utilizados para agrupamento de dados devido à facilidade de implementação, simplicidade, eficiência e sucesso empírico e possui várias extensões desenvolvidas em várias formas (JAIN; DUBES, 1999).

Este método particiona os dados em K agrupamentos mutuamente exclusivos. O método *K-means* não cria uma estrutura em árvore para descrever o processo de agrupamento, diferentemente dos métodos hierárquicos (KRONBAUER *et al.*, 2016).

Cavalcante (2018) descreve como principais vantagens da utilização do método *K-means*:

- a) o algoritmo do método é largamente utilizado em grande base de dados amostrais;
- b) o método *K-means* é mais adequado para uma grande quantidade de dados, dada o modo com que faz os particionamentos;
- c) o algoritmo do método busca, dentro do limite, a partição em que os elementos de cada grupo são mais parecidos entre si e mais distantes dos elementos de outros agrupamentos;
- d) o algoritmo do método é do tipo iterativo, onde a soma das distâncias de cada elemento ao centroide de cada grupo é minimizada, sobre todos os grupos.

Ainda segundo Cavalcante (2018), as principais desvantagens da utilização do método *K-means* são:

- a) o algoritmo tem alta sensibilidade a ruídos, dado que amostras com valores discrepantes em relação ao conjunto de dados observados podem distorcer a distribuição dos centroides;
- b) devido ao uso do centroide para comparação entre grupos, o algoritmo tem uma tendência a formar grupos esféricos;
- c) não é indicado para descobrir grupos com formas não convexas ou de diferentes tamanhos;
- d) a quantidade de grupos é fixa durante todo o processo.

O método de agrupamento *Fuzzy C-means* foi proposto por Bezdek e Pal (1992). É um método de agrupamento bastante conhecido da abordagem difusa. O algoritmo compartilha características do método *K-means*, como, por exemplo, a determinação do número de grupos *a priori* e o uso da distância Euclidiana para o cálculo da dissimilaridade entre objetos. Contudo, a similaridade de um objeto com

cada grupo é fornecida através de uma função, função de pertinência, cujos valores estão entre zero e um (CAVALCANTE, 2018).

Essa característica faz com que a função objetivo e o cálculo dos centroides sejam diferentes do método *K-means*. Além disso, o algoritmo do método possui uma nova etapa de processamento importante, o cálculo do grau de pertinência (CAVALCANTE, 2018).

Kronbauer *et al.* (2016) salientam como principais vantagens para utilização do algoritmo de agrupamento *Fuzzy C-means*:

- a) o algoritmo do método é semelhante ao algoritmo do método *K-means*, portanto, compartilha a mesma organização em quatro etapas principais: inicialização, representação, alocação e critério de parada;
- b) o algoritmo básico do método tem sido frequentemente utilizado em uma ampla área de aplicações, desde engenharia até economia;
- c) o método é flexível permitindo agrupar dados difusos nos quais um objeto não é apenas um membro de um grupo (*cluster*), mas também um membro de muitos grupos (*clusters*) em vários níveis de associação;
- d) o algoritmo do método é um processo iterativo que é finalizado quando o número máximo de iterações é atingido, ou quando a diferença entre dois valores consecutivos da função objetivo é menor que um valor de convergência pretendido (ε).

Ainda segundo Kronbauer *et al.* (2016), as desvantagens na utilização do algoritmo *Fuzzy C-means* são:

- a) necessidade de estabelecer *a priori* o número de agrupamentos C que o método irá usar para encontrar as c partições;
- b) o método *Fuzzy C-means* produz resultados próximos ao agrupamento pelo método *K-means*, mas requer mais tempo de processamento devido ao cálculo de medidas difusas envolvido no algoritmo;
- c) em agrupamentos subtrativos, deve-se tomar cuidado ao escolher o valor do raio da vizinhança, pois raios muito pequenos resultarão em negligenciar o efeito de dados vizinhos, enquanto raios grandes resultarão em uma vizinhança de todos os dados, cancelando assim o efeito do grupo (*cluster*).

2.5 ANÁLISE DE COMPONENTES PRINCIPAIS (ACP)

A análise de componentes principais (ACP) é uma técnica de análise multivariada que avalia dados interrelacionados por meio de variáveis quantitativas. O propósito da técnica é extrair os componentes mais representativos, simplificando assim os processos de análise (ABDI; WILLIAMS, 2010).

Aitchison (1983) afirma que a análise de componentes principais (PCA- *Principal Component Analysis*) simplifica a complexidade em dados que envolvem muitas dimensões, mantendo tendências e padrões. A técnica permite representar os dados em um número menor de dimensões, resultando em economia de recursos e em simplificação do processo de análise.

Rennó (2020) esclarece as etapas a serem seguidas na implementação de uma ACP:

- a) obter os dados ou as M amostras de vetores de dimensões n ;
- b) calcular a média ou o vetor médio dos dados;
- c) calcular o desvio dos dados em relação à média;
- d) calcular a matriz de covariância dos desvios em relação à média, cuja dimensão será $[n \times n]$
- e) calcular os autovalores e os autovetores da matriz de covariância.

A seguir apresenta-se o procedimento matemático envolvido na utilização da análise dos componentes principais (ACP), com base no descrito por Richardson (2009).

Assumindo-se, por hipótese, um conjunto de dados representado por uma matriz $X[m \times n]$ onde as n colunas representam as amostras (por exemplo, observações) e as m linhas são as variáveis de interesse.

O problema é estabelecido inicialmente como uma transformação linear da matriz $X[m \times n]$ em outra matriz Y , também de dimensões $[m \times n]$, de tal forma que, para alguma matriz $P[m \times m]$:

$$Y = PX \tag{6}$$

A Equação (6) representa uma mudança de base. Se as linhas de $\mathbf{P}$ forem entendidas como vetores $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_m$ e as colunas de $\mathbf{X}$ como vetores $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, então o produto matricial $\mathbf{PX}$ pode ser representado da seguinte maneira:

$$\mathbf{Y} = \mathbf{PX} = (\mathbf{Px}_1 \quad \mathbf{Px}_2 \quad \dots \mathbf{Px}_n) = \begin{pmatrix} \mathbf{p}_1 \cdot \mathbf{x}_1 & \mathbf{p}_1 \cdot \mathbf{x}_2 & \dots & \mathbf{p}_1 \cdot \mathbf{x}_n \\ \mathbf{p}_2 \cdot \mathbf{x}_1 & \mathbf{p}_2 \cdot \mathbf{x}_2 & \dots & \mathbf{p}_2 \cdot \mathbf{x}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{p}_m \cdot \mathbf{x}_1 & \mathbf{p}_m \cdot \mathbf{x}_2 & \dots & \mathbf{p}_m \cdot \mathbf{x}_n \end{pmatrix} \quad (7)$$

Se $\mathbf{p}_i, \mathbf{x}_j \in \mathbb{R}^m$, então $\mathbf{p}_i \cdot \mathbf{x}_j$ representa o produto interno padrão euclidiano entre o vetor linha $\mathbf{p}_i$ e o vetor coluna $\mathbf{x}_j$.

Essa interpretação resulta em que os dados originais da matriz $\mathbf{X}$ estão sendo projetados nas colunas de $\mathbf{P}$. Assim, os vetores linha de $\mathbf{P}$, $\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_m\}$ representam uma nova base para representar as colunas de $\mathbf{X}$.

O próximo passo é estabelecer certas condições para a definição de $\mathbf{P}$ de tal forma que se possa atingir o objetivo principal da análise de componentes principais que é a redução da dimensionalidade do problema de análise da variabilidade do conjunto de dados. Sob determinadas condições específicas, os vetores linha de $\mathbf{P}$ representam as direções dos componentes principais.

Assim, é necessário definir qual deve ser essa nova base. Em outras palavras, qual é a melhor maneira de expressar os dados de $\mathbf{X}$ em outra base, ou seja, como devemos definir a independência entre os componentes principais na nova base.

A Análise de Componentes Principais (ACP) define independência, considerando a variância dos dados na base original. A técnica procura eliminar a correlação entre os dados originais, encontrando as direções em que a variância é maximizada e, em seguida, utilizar essas direções para definir a nova base.

Supondo conhecido um vetor de n medidas discretas, $\tilde{\mathbf{r}} = (\tilde{r}_1, \tilde{r}_2, \dots, \tilde{r}_n)$, com média μ_r . O desvio das observações em relação à sua média (μ_r) pode ser representado por um vetor $\mathbf{r} = (r_1, r_2, \dots, r_n)$, que tem média zero e que representa, em realidade, uma translação do conjunto de dados originais em razão da subtração de cada o valor observado pela média (μ_r).

Com o vetor $\mathbf{r}$ assim definido, a variância das observações pode ser representada por:

$$C_r = \sigma_r^2 = \frac{1}{n} \mathbf{r} \mathbf{r}^T \quad (8)$$

Onde: $C_r = \sigma_r^2$ = variância das observações $\tilde{r}_1, \tilde{r}_2, \dots, \tilde{r}_n$;

$\mathbf{r}$ = vetor que contém o desvio das observações $\tilde{r}_1, \tilde{r}_2, \dots, \tilde{r}_n$ em relação à média μ_r ;

n = número de observações.

Considerando um segundo vetor de n medições $\mathbf{s} = (s_1, s_2, \dots, s_n)$, novamente com média zero, então a Equação (8) pode ser generalizada e utilizada para calcular a covariância de $\mathbf{r}$ e $\mathbf{s}$.

A Equação (9) representa a generalização da Equação (8), onde se utiliza o fator $(n-1)$ como divisor no lugar de n , e cuja explicação se encontra na teoria de estimação, quando se busca estimar os parâmetros populacionais a partir do conhecimento dos parâmetros amostrais:

$$C_{rs} = \sigma_{rs}^2 = \frac{1}{n-1} \mathbf{r} \mathbf{s}^T \quad (9)$$

A covariância pode ser considerada como uma medida da correlação entre duas variáveis.

Retornando ao problema original da matriz $\mathbf{X}[\mathbf{m} \times \mathbf{n}]$, onde as $\mathbf{n}$ colunas representam as amostras (por exemplo, observações) e as $\mathbf{m}$ linhas são as variáveis de interesse. Pode-se, agora, entender a matriz $\mathbf{X}$ como sendo composta por $\mathbf{m}$ vetores linha $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$, cada um com comprimento $\mathbf{n}$, representada como a seguir:

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_m \end{pmatrix} = \begin{pmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,n} \\ x_{2,1} & x_{2,2} & \dots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{m,2} & \dots & x_{m,n} \end{pmatrix} \quad (10)$$

Uma forma de entender a representação dada pela Equação (10) é a de que cada linha $\mathbf{i}$ da matriz $\mathbf{X}$ contém todas as $\mathbf{n}$ observações da variável de interesse x_i .

Considerando-se ainda as Equações (8) e (9), que representam os conceitos de variância e covariância, pode-se definir a matriz c_x [$m \times m$]:

$$C_x = \frac{1}{n-1} X X^T = \frac{1}{n-1} \begin{pmatrix} x_1 x_1^T & x_1 x_2^T & \dots & x_1 x_m^T \\ x_2 x_1^T & x_2 x_2^T & \dots & x_2 x_m^T \\ \vdots & \vdots & \ddots & \vdots \\ x_m x_1^T & x_m x_2^T & \dots & x_m x_m^T \end{pmatrix} \quad (11)$$

A Equação (11) representa o cálculo das covariâncias entre todos os possíveis pares das m variáveis. Os elementos da diagonal principal da matriz c_x armazenam as variâncias das m variáveis de interesse, enquanto os elementos fora da diagonal principal da matriz c_x armazenam as covariâncias. A matriz c_x é, portanto, denominada como matriz de covariância.

A análise dos componentes principais adota, como hipótese fundamental, que as variáveis na matriz transformada Y devem ser o menos correlacionadas possível. Isso é equivalente a dizer que as covariâncias na matriz c_y devem ser o mais próximo de zero possível (matrizes de covariância são sempre positivo definidas ou positivo semi-definidas). Por outro lado, grandes variações nos valores de covariância devem ser investigadas, pois correspondem a dinâmicas interessantes do sistema.

Portanto, os seguintes requisitos são estabelecidos para a determinação de P que transforma X em Y , considerando-se a matriz de covariância c_y :

- a) maximizar o sinal, medido pela variância (maximizar os elementos da diagonal principal de c_y);
- b) minimizar a covariância entre as variáveis (minimizar os elementos fora da diagonal de c_y).

Deste modo, conclui-se que, uma vez que a mínima covariância possível, em termos absolutos, é igual a zero, então condição ótima é representada por uma matriz de covariância c_y , cujos elementos fora da diagonal principal sejam todos nulos, ou seja, que a matriz de covariância c_y resulte em uma matriz diagonal.

Em resumo, a matriz de transformação P que transforma X em Y deve ser tal que a matriz de covariância c_y resulte em uma matriz diagonal. Resta agora definir um

procedimento matemática que permita definir os elementos da matriz de transformação P

Como mencionado anteriormente, as linhas de P podem ser entendidas como vetores $p_1, p_2, \dots, p_m$. Ao se adotar a hipótese de que $p_1, p_2, \dots, p_m$ são ortogonais, a solução do problema pode ser obtida, utilizando-se a teoria de álgebra linear.

Considerando-se Equação (11) para a determinação da matriz de covariância c_Y , a transformação linear dada pela Equação (6) e as propriedades de álgebra matricial, pode-se definir:

$$C_Y = \frac{1}{n-1} Y Y^T = \frac{1}{n-1} (PX)(PX)^T = \frac{1}{n-1} P(XX^T)P^T \quad (12)$$

ou, de forma alternativa:

$$C_Y = \frac{1}{n-1} P S P^T \quad (13)$$

Onde $S = XX^T$.

Por definição, a matriz S resulta em uma matriz simétrica de dimensões $[m \times m]$. Da teoria da álgebra linear, tem-se que toda matriz quadrada simétrica é passível de diagonalização, ou seja, pode ser escrita na forma:

$$S = E D E^T \quad (14)$$

Onde E é uma matriz ortogonal $[m \times m]$, cujas colunas são os autovetores ortogonais de S , e D é uma matriz diagonal $[m \times m]$, cujos elementos da diagonal principal contém os autovalores de S .

O posto r de S é o número de autovetores ortogonais que a matriz S possui. Se S acabar por ser deficiente em termos de posto, ou seja, se o posto r for menor que o tamanho m da matriz S , então será necessário gerar $(m - r)$ vetores ortogonais para preencher as colunas restantes de S .

A escolha da matriz P é feita de forma que as linhas de P sejam os autovetores de S , garantindo que P resulte igual a E^T . Desta forma, a Equação (13) pode ser reescrita como:

$$C_Y = \frac{1}{n-1} E^T (EDE^T) E \quad (15)$$

Como a matriz E é ortogonal, resulta em $E^T E = I$, onde I é uma matriz identidade $[m \times m]$. Finalmente, ao se definir a matriz P de tal forma que as linhas de P sejam os autovetores de S , onde $S = XX^T$ e X contém os valores observados das variáveis de interesse, tem-se.

$$C_Y = \frac{1}{n-1} D \quad (16)$$

Onde D é uma matriz diagonal $[m \times m]$, cujos elementos da diagonal principal contém os autovalores de S .

A análise dos componentes principais (ACP) permite a definição da direção de componentes principais (autovetores de S) e a importância relativa (autovalores de S) de cada componente principal para a explicação da variação dos dados observados. A maior variação corresponde ao primeiro componente principal, a segunda maior variação corresponde ao segundo componente principal e sucessivamente, fornecendo um método para organizar os dados observados na etapa de diagonalização.

Em termos de procedimento, uma vez obtidos os autovalores e autovetores de S , os autovalores são classificados em ordem decrescente e são colocados nessa ordem na diagonal da matriz D . Na sequência, constrói-se a matriz ortogonal E , colocando os autovetores na mesma ordem dos respectivos autovalores para formar as colunas de E .

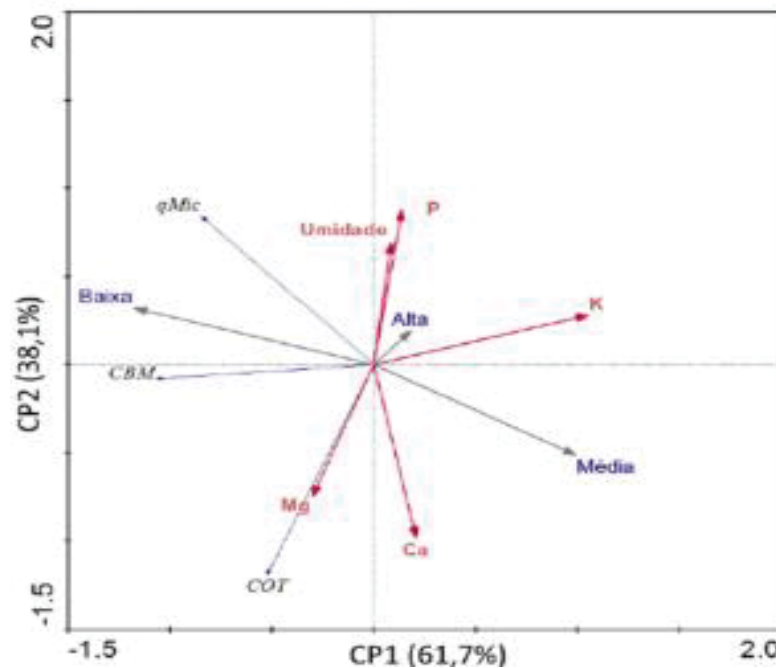
Seguindo o procedimento descrito, se atinge o objetivo de diagonalizar a matriz de covariância C_Y dos dados transformados.

2.5.1 Representação gráfica da Análise dos Componentes Principais (ACP)

Como exemplo da representação gráfica, a Figura 8 mostra a análise de componentes principais (ACP), feita por Kraft *et al.* (2018), correlacionando os níveis de produtividade no plantio direto de Soja, na região oeste do estado de Santa Catarina. Os parâmetros de produtividade foram caracterizados em Alta, Média e Baixa, e as variáveis respostas foram: Carbono da Biomassa Microbiana (CBM),

Carbono Orgânico Total (COT) e quociente microbiano (qMic); e variáveis explicativas: umidade (Umi), fósforo (P), potássio (K), cálcio (Ca) e magnésio (Mg).

FIGURA 8 - REPRESENTAÇÃO GRÁFICA DE ACP.



FONTE: Adaptado de Kraft *et al.* (2018).

A ACP demonstrou clara separação entre os níveis de produtividade (Alta > Baixa) e os atributos microbianos, sendo que os componentes principais 1 e 2 (CP1 e CP2) explicaram 61,7% e 38,1% da variabilidade, respectivamente, totalizando juntos 99,8% da variação total dos dados observados (KRAFT *et al.*, 2018).

2.6 APLICAÇÕES DO AGRUPAMENTO DE DADOS EM HIDROLOGIA

De acordo com Jain e Dubes (1988), o agrupamento de dados é um método que utiliza o aprendizado não supervisionado ou auto organizável, ou seja, não há um professor ou crítico que lhe indique o que cada padrão representa. Basicamente, a previsão será feita através da representação interna montada pelo sistema, utilizando um conjunto de dados de entrada.

Este método está normalmente associado à análise exploratória (formulação de hipóteses e tomada de decisão), pois envolve problemas em que há pouca informação *a priori* acerca dos dados; e poucas hipóteses podem ser sustentadas. A

formação de grupos pode fornecer novas hipóteses a respeito dos inter-relacionamentos dos dados e de sua estrutura intrínseca (MOSCATO; ZUBER, 2008).

A popularização do uso e desenvolvimento de métodos de agrupamento ocorreu devido à grande quantidade de dados coletados nas diversas áreas do conhecimento e atividade que tornam a análise de grupo (*cluster*) um tópico altamente atrativo em várias pesquisas na mineração de dados (LINDEN, 2009).

É exaustivamente difícil listar todas as aplicações que têm utilizado técnicas de agrupamento distintas, bem como, a grande quantidade de algoritmos desenvolvidos. Destacam-se a seguir, de forma resumida, alguns dos estudos correlatos à previsão de enchentes e mitigação de seus efeitos.

Manly (1994) aplicou duas abordagens diferentes de agrupamento para o delineamento de regiões hidrológicas homogêneas, com o objetivo de compará-las. A primeira foi a abordagem hierárquica de agrupamento, amplamente utilizada para estudos de regionalização. A segunda foi o método *Fuzzy C-means*, que permite que uma estação pertença, em diferentes graus, a várias regiões. O número ideal de regiões foi baseado em medidas de validação de cluster nebuloso. Os resultados experimentais de ambas as abordagens foram semelhantes, o que confirmou a homogeneidade das regiões delineadas para o referido estudo de caso.

Vicini (2005) utilizou uma métrica de agrupamento *Fuzzy C-means*, baseada em kernel (KFCM) para uso em situações em que essa separação linear de regiões não pode ser realizada. A abordagem utilizou funções baseadas em algoritmo *Kernel* para mapear os pontos de dados do espaço de atributo, para um espaço de maior dimensão, desta forma, possibilitando a separação em regiões por planos lineares. Foi implementado um procedimento para determinar o número ideal de regiões com a abordagem KFCM (*Kernel Fuzzy C-means*). A eficácia da abordagem foi demonstrada por experimentos de simulação Monte-Carlo e um estudo de caso em bacias hidrográficas nos Estados Unidos. A comparação dos resultados, com os baseados no agrupamento convencional de meios difusos, indicou que a abordagem KFCM superou as outras abordagens na formação de regiões mais próximas de serem estatisticamente homogêneas e na estimativa de quantis de inundação em locais não controlados.

Zahmatkesh *et al.* (2014) utilizaram agrupamento de dados baseado no método *K-means* para desenvolvimento de um modelo de transformação chuva-vazão, visando diminuir a incerteza conceitual (estrutural) do modelo na previsão de escoamento em tempo real. Na primeira etapa do estudo, modelos orientados a dados

foram desenvolvidos e usados para prever chuvas, usando sinais climáticos em larga escala como preditores de chuvas. Na segunda etapa, devido ao alto efeito de diferentes fontes de incerteza nas entradas e saídas de modelos hidrológicos, os parâmetros e a estrutura do modelo foram incorporados à modelagem e simulação de escoamento fluvial. Os parâmetros do modelo e a incerteza de entrada foram investigados usando um método adaptável de Monte Carlo em Cadeia de *Markov*. Finalmente, os modelos calibrados de precipitação e escoamento foram utilizados, usando a precipitação prevista, para prever escoamentos futuros para a bacia hidrográfica. O esquema proposto foi empregado no estudo de caso da bacia do rio *Bronx*, em Nova York, USA. Os resultados indicaram que a metodologia desenvolvida, não apenas forneceu ferramentas confiáveis para modelagem da chuva e da vazão, mas também proporcionou tempo adequado para incorporar medidas de mitigação necessárias para lidar com eventos potencialmente extremos de escoamento e risco de inundação.

Liu *et al.* (2016) utilizaram uma abordagem *Fuzzy* espacial (SFCA) para estimar possíveis riscos de inundação em sistemas de drenagem de águas pluviais e abordar a complexidade e incertezas da avaliação de risco de inundação. A aplicação do método em uma bacia de drenagem natural, demonstrou se tratar de uma técnica apropriada para determinar a distribuição espacial dos riscos de inundações. Os resultados revelam que áreas com alto, moderado e baixo risco de inundação, identificadas pela SFCA correspondem a um certo grau de estresse ambiental, ocasionado pela ocupação irregular das áreas inundáveis (leitos naturais de inundação), podendo servir como base preliminar para orientar os gestores na avaliação dos riscos de inundações.

Steffen (2017) utilizou agrupamento de dados pelo método *Fuzzy C-means*, para agrupar cheias observadas na cidade de União da Vitória, Paraná, a partir de hidrogramas observados, a fim de se identificar características semelhantes entre os eventos de um mesmo grupo, tendo-se por definição um conjunto de indicadores de cheia.

Xu *et al.* (2018) propuseram uma metodologia integrada para previsão de inundações, mesclando o método de modelo de peso de entropia e algoritmo de agrupamento *K-means*, visando avaliar o risco de inundação urbana. A abordagem proposta foi orientada por dados, sem considerar o padrão de classificação de diferentes níveis de risco e, portanto, fornecendo um resultado mais razoável e objetivo. A metodologia foi aplicada na região em Haikou, China. Sete índices de

avaliação foram selecionados, acoplando o sistema de índice de risco natural e modelos hidrológicos. Os pesos do índice foram calculados por um método aprimorado de peso de entropia, o qual integrou o método de peso de entropia e o processo de hierarquia analítica (AHP). Posteriormente, o método *K-means* foi utilizado para desenvolver um mapa de risco de inundação na área de estudo. Os resultados do estudo fornecem uma nova abordagem para a avaliação de riscos de inundação, tendo apresentado similaridade com os dados históricos de eventos de inundação, podendo assim fornecer informações valiosas para previsão e mapeamento de inundações urbanas.

Steffen e Gomes (2018) aplicaram diferentes métodos de pré-tratamento dos dados observados, analisando o consequente comportamento do agrupamento pelo método *Fuzzy C-means* das cheias observadas no município de União da Vitória, Paraná. Os métodos aplicados foram: *autoscaling*, *range scaling*, *Pareto scaling*, *vast scaling* e *level scaling*. Os resultados mostraram que os diferentes métodos produziram agrupamentos distintos das cheias. Verificou-se ainda que o método *vast scaling* apresentou um melhor desempenho, considerando-se os eventos observados no local do estudo.

Uliana *et al.* (2019) apresentaram uma nova estrutura classificada de previsão de inundação em tempo real, integrando um modelo de agrupamento nebuloso e uma rede neural com um modelo hidrológico conceitual. No estudo em questão, uma rede neural de retropropagação foi treinada usando dados de precipitação em tempo real e resultados do modelo de agrupamento difuso. A rede neural forneceu uma classificação *on-line* rápida para eventos de inundação em tempo real. Com base na classificação *on-line*, um conjunto apropriado de parâmetros do modelo hidrológico foi escolhido automaticamente para produzir previsão de inundação em tempo real. A metodologia proposta foi aplicada a uma grande bacia hidrográfica na província de Liaoning, China. Os resultados mostraram que a estrutura classificada forneceu uma previsão mais precisa do que o método tradicional não classificado.

Chen e Wang (2012) utilizaram o método de agrupamento *Fuzzy C-means* para definir indicadores de severidade de cheias no rio Wujiand no sul da China. Foram definidos 5 indicadores de severidade de cheia, sendo eles: vazão de pico, nível de água durante o pico de vazão, volume máximo de 24 horas, o volume máximo de 72 horas e o volume total de inundação. A análise revelou tendências crescentes significativas no pico de descarga e no volume máximo de 24 horas, enquanto que o

nível do pico de inundação, o volume total de inundação e o volume máximo de 72 horas, tiveram apenas uma tendência fraca de crescimento.

3 MATERIAIS E MÉTODOS

Tendo como objetivo comparar dois métodos de agrupamento de dados, *K-means* e *Fuzzy C-means*, aplicados ao agrupamento de cheias previamente selecionadas por Steffen (2017), e posteriormente determinar os indicadores de cheia mais representativos por meio do uso da Análise de Componentes Principais (ACP), apresenta-se, neste capítulo, a área de estudo e a metodologia empregada no desenvolvimento do presente estudo.

Inicialmente, são apresentadas a área de estudo e a contextualização da problemática referente às frequentes enchentes do rio Iguaçu. Na sequência, procurou-se detalhar os métodos utilizados, com ênfase na descrição do método *K-means*, utilizado para análise de agrupamento das cheias máximas anuais observadas na Estação Fluviométrica União da Vitória (65310000), localizada no rio Iguaçu, na cidade de União da Vitória, estado do Paraná.

3.1 HIDROGRAFIA DO ESTADO DO PARANÁ

Paraná, em língua Tupi-guarani antiga³ significa “*confluência de grandes rios, parecidos com o mar*”, devido ao fato do território do estado ser vastamente suprido por recursos hídricos.

A abundância hídrica do estado do Paraná já era conhecida pelos habitantes da América pré-colombiana, sendo os rios do estado largamente utilizados como rotas comerciais pelos índios Tupinambás, Tupi-guaranis, Carijós, Caingangues, além de outras denominações da etnia Jê. Em se tratando especialmente dos rios Paraná e Iguaçu, que eram importantes rotas Sul e Leste da parte final do caminho do Peabiru, que ligava o litoral Brasileiro ao Peru e ao Oceano Pacífico.

O memorando sobre Bacias hidrográficas do Paraná, produzido pela Secretaria de Estado do Meio Ambiente e Recursos Hídricos – SEMA (PARANÁ,

³ Existem diferenças entre o idioma Tupi guarani, Tupimambá, Tupiniquim e o Nheengatu (nova língua Tupi, difundida a partir do século XVIII).

2013), especifica que a hidrografia do estado do Paraná pode ser classificada em cinco bacias hidrográficas:

- Bacia do rio Paraná, cujos afluentes mais importantes são os rios Piquiri e Ivaí;
- Bacia do rio Paranapanema, drenada pelos rios Pirapó, Tibagi, das Cinzas e Itararé;
- Bacia do rio Iguaçu, que tem como principais afluentes o rio Chopim, no sul do estado, e o rio Negro, no limite com o estado de Santa Catarina;
- Bacia do Rio Ribeira do Iguape, cujas águas seguem para o rio Ribeira do Iguape;
- Bacia Atlântica ou do litoral Paranaense, cujas águas seguem diretamente para o Oceano Atlântico.

3.2 ÁREA DE ESTUDO

Devido à frequência da ocorrência de enchentes e à sua relevância no cenário estadual, o presente trabalho enfoca sua análise nas cheias do rio Iguaçu observadas na cidade de União da Vitória, localizada na região sudeste do Estado do Paraná, na divisa com o estado de Santa Catarina

O município de União da Vitória foi fundado em 1769 pelo capitão Peixoto, com nome inaugural de Entrepasto de Nossa Senhora da Vitória. O ano de 1880 marcou a chegada do coronel Amazonas de Araújo Marcondes, o homem que lançou os alicerces da civilização e tradição união-vitoriense. Em 1881, chegou a primeira leva de imigrantes alemães, vindos de Rio Negro e do Vale do Itajaí. Estas famílias dedicavam-se especialmente à agricultura (KARPINSKI, 2012).

Pelo Decreto nº. 54, de 27 de março de 1890, a freguesia foi elevada à categoria de vila, e pelo Decreto nº. 55, de mesma data, foi criado o município de Porto União da Vitória, com território desmembrado de Palmas. Em 1905, foi inaugurada a estação ferroviária (KARPINSKI, 2012).

Em função do Tratado de Limites entre os estados do Paraná e Santa Catarina, em 07 de setembro de 1917, a cidade de Porto União da Vitória dividiu-se em duas: União da Vitória, em território paranaense, e Porto União, em território catarinense, separadas pelo Rio Iguaçu.

Desde a colonização da região de União da Vitória, as enchentes no rio Iguaçu têm sido amplamente documentadas. Registros oficiais do governo imperial detalham a expedição executada pelos irmãos Francisco e José Keller, Engenheiros Alemães, diplomados pela Escola Politécnica de Carlsburg, através do rio Iguaçu. Em 1865, os Kellers chegaram à Província do Paraná a fim de promover a exploração dos rios paranaenses que pudessem servir de comunicação com a Província do Mato Grosso. Além disso, conforme indica o relatório do vice-presidente Ermelino de Leão, esses engenheiros iniciaram a maior expedição realizada no rio Iguaçu até então, com a presença de agrimensores, tradutores e remadores que levariam oito canoas rio abaixo, partindo de São José dos Pinhais até onde as águas permitissem (KARPINSKI, 2012).

Anotações do diário de Francisco Keller demonstram que a expedição teve que ser interrompida logo em seu início, devido a uma severa enchente:

Iniciaram os trabalhos expedicionários no dia 24 de junho de 1866 saindo de Curitiba com destino à ponte que havia sobre o rio Iguaçu nas proximidades de São José dos Pinhais. Devido às intensas chuvas e ao nível do rio estar 2,5 metros acima do que se considerava normal, ali ficaram até o dia 27 fazendo medições altimétricas e observações sobre o perfil transversal do vale do rio (Transcrição do diário de Francisco Keller, Jornal Dezenove de Dezembro, 1867, apud KARPINSKI, 2012).

Assim como grande parte dos assentamentos humanos, a ocupação urbana de União da Vitória se deu próxima ao rio Iguaçu, devido à facilidade para obtenção de água, além de transporte de insumos, etc. Esse processo desordenado de urbanização de áreas alagáveis do Iguaçu teve consequências que puderam ser sentidas já nos primeiros anos da cidade. Silva (2006) apresenta, em linhas gerais, os danos causados pela primeira inundação que se tem registros:

Em junho de 1891, transborda o Rio Iguaçu, em consequência dos grandes temporais, causando sérios prejuízos aos moradores de suas margens. Os terrenos fronteiriços ficaram inteiramente alagados: um pequeno mar a levantar furiosas ondas. O velho depósito que existia à margem esquerda, próximo ao ponto de atracação das embarcações, pouco faltou para ser totalmente coberto pelas águas. (SILVA, 2006, p.64).

Em maio de 1905, uma cheia do rio Iguaçu tornou a atingir o município, causando a interrupção das linhas férreas e deixando as casas mais próximas ao rio praticamente submersas. Em setembro de 1911, outra enchente de grande porte ocorreu na região (ANSBACH, 2008).

No início da década de 1930, teve início a coleta de dados de vazão dos principais rios do estado do Paraná. A partir do ano de 1931, a cidade de União da Vitória passou a realizar os registros da cota e vazão do Rio Iguaçu, fato que possibilitou a produção de uma série histórica de vazões, tornando possível aferir a variação do nível do rio com maior precisão, para fins de estudo e compreensão desse tipo de evento (ANSBACH, 2008).

3.2.1 Suscetibilidade a inundações em União da Vitória

O relevo da cidade de União da Vitória é caracterizado por formações geológicas variadas, com cotas e declividades distintas, apresentando diferentes respostas relacionadas aos problemas ocasionados por inundações, frente à ocupação e uso do solo (SILVEIRA, 2016).

A região central de União da Vitória, onde se localiza a sede do município, além dos bairros Navegantes, Ponte Nova, São Bernardo e Anita Jacobs, encontra-se instalada sobre depósitos aluviais quaternários, com cotas altimétricas entre 750m e 770m, e declividade reduzida que sujeita-se a inundações periódicas, destacando-se que as áreas mais propensas a tais fenômenos situam-se no entorno da região central, ao longo da margem esquerda do rio Iguaçu, constituindo-se de bairros residenciais de classe baixa com focos de ocupação irregulares (PARANÁ, 2013).

Os bairros situados à margem direita do rio Iguaçu possuem como embasamento geológico, depósitos aluviais e rochas sedimentares das formações Rio do Rasto e Botucatu (PARANÁ, 2013).

Alguns bairros, como o Muzzolon, Área Industrial Domício Scaramela, São Judas Tadeu, Ouro Verde, São João, São José e João Paulo II, estão situados em áreas de vertentes com alta declividade onde afloram rochas da Formação Botucatu. Nesses locais, os danos diretos ocasionados por inundações não são tão intensos, sendo as maiores consequências os movimentos de massa como escorregamentos, rolamentos de blocos, dentre outros (PARANÁ, 2007).

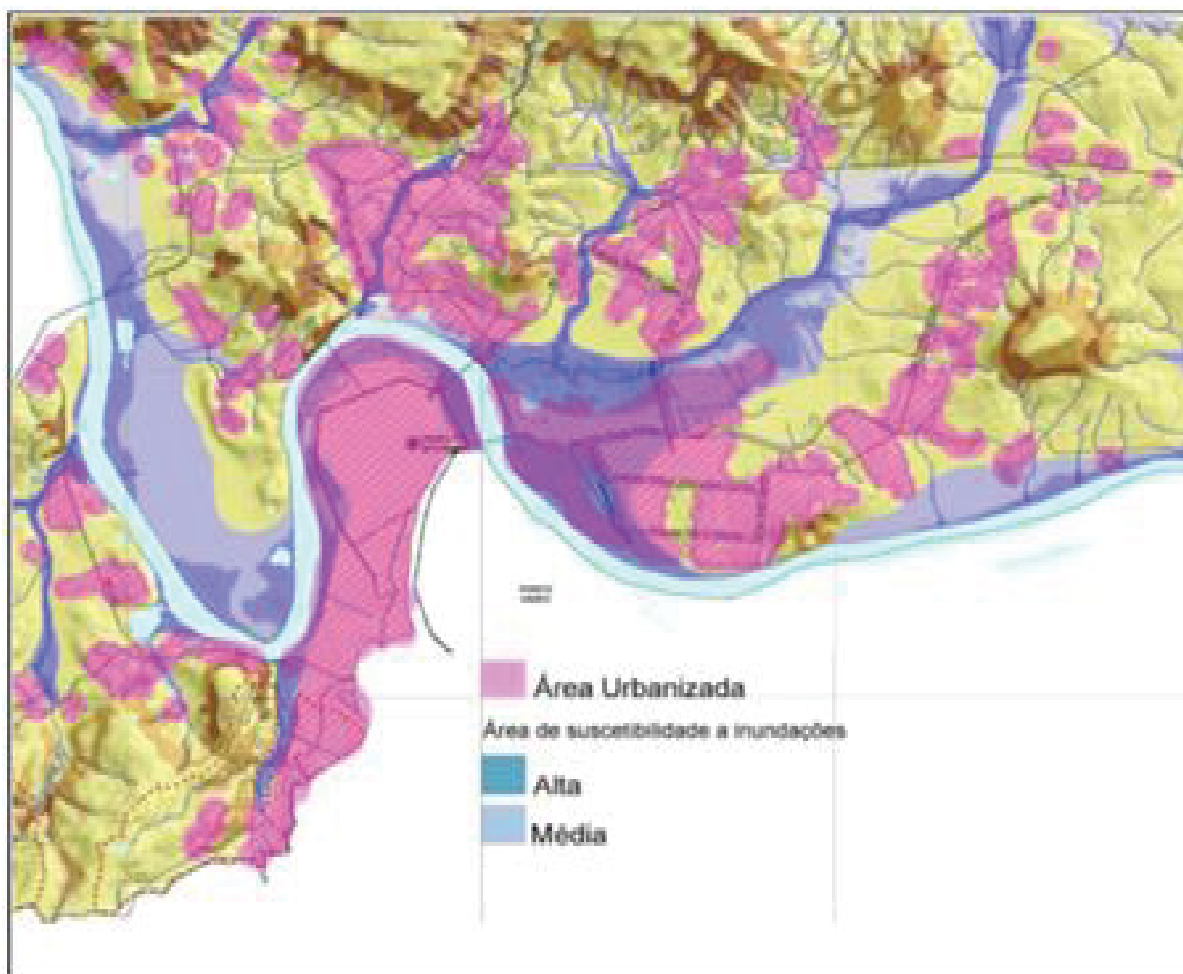
As regiões mais propensas a riscos de enchentes e inundações são as localizadas principalmente às margens dos rios Iguaçu e Vermelho e dos ribeirões do Meio, da Prata e Guariroba, conforme mostrado na Figura 9. Algumas dessas planícies de inundação chegam a atingir mais de 2 km de largura, não constituindo locais adequados para a implantação de edificações. Porém, como mencionado

anteriormente, devido à falta de planejamento urbano, acabam por se tornar regiões com maior adensamento populacional (BRASIL, 2007).

As Figuras 10 e 11 mostram a extensão da catástrofe causada pela enchente de 1983 em União da Vitória, cujos registros oficiais são reproduzidos em parte a seguir:

No mês de julho de 1983, a população de Porto União, União da Vitória e municípios vizinhos passou por um período de grande enchente. A comunidade local ficou sem energia elétrica, água potável e produtos alimentícios entre outros itens indispensáveis. As águas do rio Iguaçu cobriram a cidade. A chuva perdurou por uma semana e fez com que o rio de 2,50 metros chegasse à 10,42 metros, diante de tal realidade a maior parte dos moradores deixou tudo para trás. A média de chuva para os meses junho e julho é de 138 milímetros, mas nesse desastre choveu em torno de 800 milímetros, sendo que mais de 2.400 famílias ficaram em situação vulnerável, muitas foram alojadas em casa de familiares e amigos, mas a maioria teve que ficar em abrigos na região. Dezoito pessoas perderam suas vidas nesse desastre (DEFESA CIVIL; PARANÁ, 2000).

FIGURA 9 - MAPA DE SUSCETIBILIDADE À INUNDAÇÃODAS ÁREAS URBANAS DE UNIÃO DA VITÓRIA.



FONTE: Ministério de Minas e Energia –MME (2015).

FIGURA 10 - IMAGEM DE UNIÃO DA VITÓRIA 1983, ANTES DA ENCHENTE.



FONTE: SEC-CORPRERI (2009).

FIGURA 11 - IMAGEM DE UNIÃO DA VITÓRIA 1983, DURANTE A ENCHENTE.



FONTE: SEC-CORPRERI (2009).

3.2.2 Bacia hidrográfica do rio Iguaçu

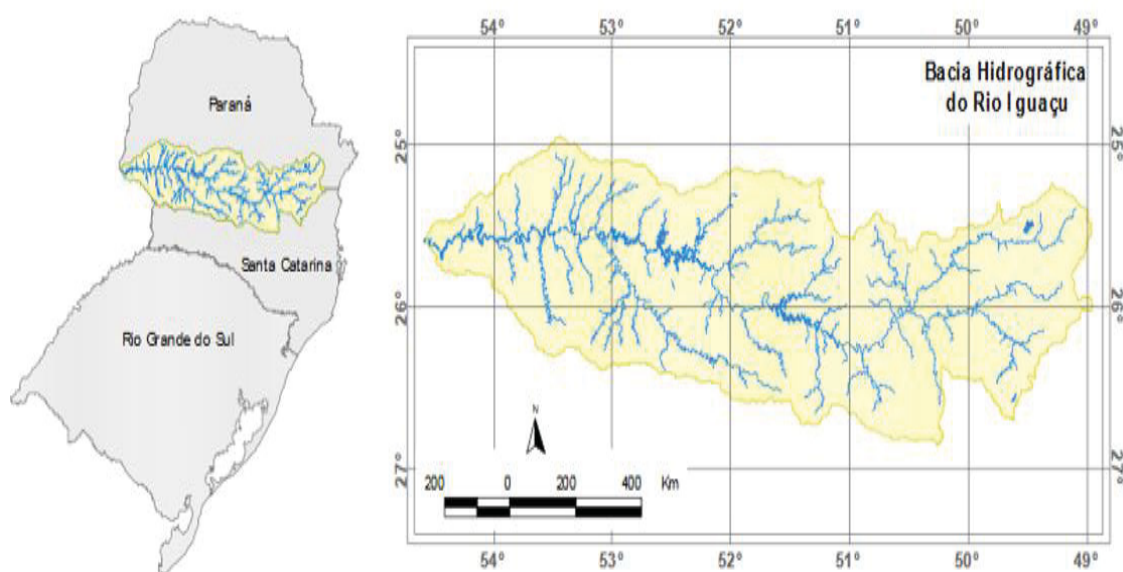
O Sol já havia passado 25 graus de seu zênite, na tarde de 31 de janeiro de 1542. Depois de haver percorrido aproximadamente 16 léguas à jusante, os homens estavam exaustos. Muitos não haviam dormido na noite anterior, em virtude do forte e estrondoso ruído - advindo de rio abaixo - e que se tornava cada vez mais forte à

medida que avançavam, amedrontando até os mais experientes canoeiros. Estes aliás, índios da etnia Tupi-guarani, agenciados para participar da comitiva, que havia partido das terras da ilha de Santa Catarina (hoje conhecida como Florianópolis) seis meses antes, com destino à Assuncion no Paraguai, onde Alvar Nuñez Cabeça de Vaca, Adelantado do Rei da Espanha, assumiria o Governo da Província (VACA, 1542). O diário de Alvar Nuñez Cabeça de Vaca, relata também suas impressões ao ver as cataratas do Iguaçu, sendo esses os primeiros registros conhecidos sobre a região, os quais são reproduzidos a seguir:

O governador comprou algumas canoas dos índios e embarcou com oitenta homens rio Iguaçu abaixo, seguindo o restante por terra, devendo todos se juntarem no rio Paraná. Mas, ao irem rio Iguaçu abaixo, era tão forte a correnteza que as canoas corriam com muita fúria. Logo adiante do ponto onde havia embarcado, o rio dá uns saltos por uns penhascos enormes e a água golpeia a terra com tanta força que de muito longe se ouve o ruído. De modo que foi necessário sair da água, tirar as canoas e conduzi-las por terra até passar por aqueles saltos (VACA, 1542).

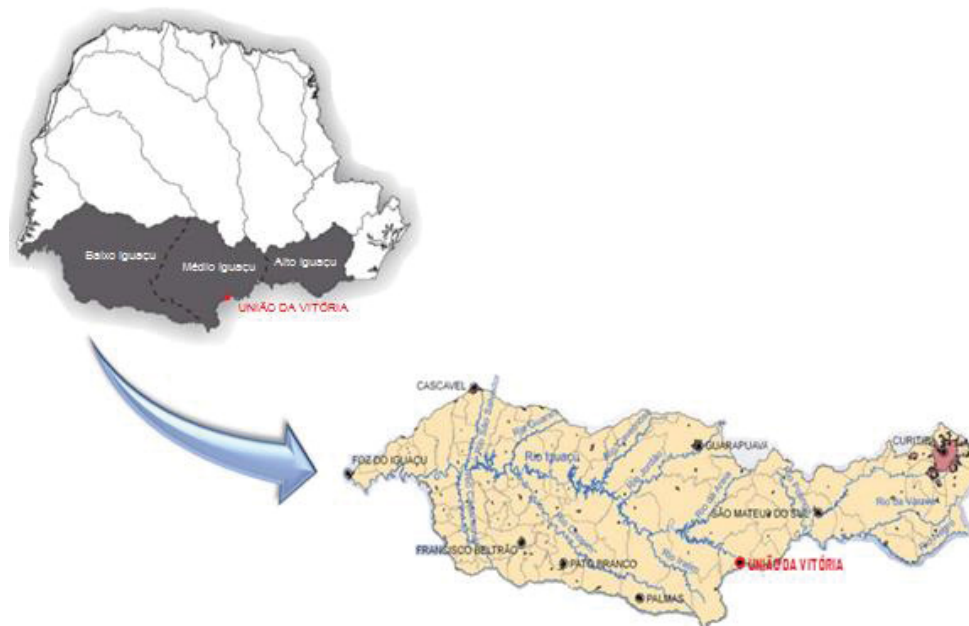
O rio Iguaçu constitui-se no principal curso fluvial do estado do Paraná, com uma extensão total de 1.330 km, desenvolvendo-se no sentido leste-oeste e, assim, se posicionando perpendicularmente ao oceano Atlântico. As Figuras 12 e 13 mostram, respectivamente, a localização da bacia hidrográfica do rio Iguaçu e a localização da cidade de União da Vitória em relação ao estado do Paraná e à bacia hidrográfica do rio Iguaçu.

FIGURA 12 - LOCALIZAÇÃO DA BACIA HIDROGRÁFICA DO RIO IGUAÇU.



FONTE: PARANÁ, SEMA, IBGE (2013).

FIGURA 13 – LOCALIZAÇÃO DA CIDADE DE UNIÃO DA VITÓRIA.



FONTE: Adaptado de PARANÁ, SEMA (2010).

As nascentes do rio Iguaçu se localizam na Serra do Mar, na região metropolitana da cidade de Curitiba, em elevações superiores a 900m, permanecendo com altitude de 600m em praticamente todo o seu trecho médio. No decorrer do seu percurso, o rio Iguaçu recebe uma quantidade considerável de afluentes, sendo 61 rios na margem direita e 42 rios na margem esquerda, cortando o estado do Paraná em uma linha perpendicular ao oceano atlântico (PARANÁ, 2013).

O rio Iguaçu é alimentado pela principal bacia hidrográfica do estado do Paraná, com aproximadamente 57.329 km². Se forem considerados os rios afluentes do Iguaçu, localizados no estado de Santa Catarina, a extensão da bacia sobe para 79.799 km² (ARAÚJO, 2006).

De acordo com o critério de divisão por bacias hidrográficas (BRASIL, 2000), a sub-bacia do Iguaçu pertence à bacia do Rio Paraná, estando compreendida entre 24° 50' e 27° 00' de latitude Sul e 49° 00' e 54° 30' de longitude Oeste.

Os dados observados no rio Iguaçu e utilizados no presente trabalho se referem à estação fluviométrica União da Vitória (65310000), localizada nas coordenadas de 26° 13' 37" S e 51° 04' 48" O, abrangendo uma área de drenagem de aproximadamente 24.200 km².

3.3 MÉTODOS

Neste trabalho, adotou-se o método *K-means* para o agrupamento de cheias históricas observadas na estação fluviométrica União da Vitória (65310000), localizada no rio Iguaçu na cidade de União da Vitória, com o objetivo de comparar os resultados obtidos por Steffen (2017), que utilizou o algoritmo *Fuzzy C-means*, com o mesmo propósito.

Para satisfazer os objetivos do estudo, fez-se necessário o desenvolvimento das seguintes etapas:

Etapa 1 – Análise hidrológica preliminar;

Etapa 2 - Definição dos eventos de cheia a serem analisados;

Etapa 3 - Definição dos indicadores de cheia a serem utilizados;

Etapa 4 - Normalização dos valores dos indicadores de cheia;

Etapa 5 - Desenvolvimento de programa computacional para o agrupamento de dados utilizando o método *K-means*;

Etapa 6 - Agrupamento das cheias pelo método *K-means*;

Etapa 7 - Comparação dos agrupamentos das cheias obtidos pela utilização dos métodos *K-means* e *Fuzzy C-means*;

Etapa 8 - Análise da influência dos métodos de normalização sobre o agrupamento das cheias pelo método *K-means*;

Etapa 9 - Determinação dos indicadores de cheia mais representativos por meio da Análise dos Componentes Principais (ACP);

Etapa 10 - Agrupamento das cheias pelo método *K-means* com o uso dos indicadores de cheia mais representativos;

Etapa 11 - Comparação dos dois agrupamentos das cheias pelo método *K-means* (comparação dos resultados das Etapas 5 e 9).

3.3.1 Análise hidrológica preliminar

Conforme mencionado anteriormente, os dados fluviométricos (nível de água e vazão) são referentes à estação fluviométrica União da Vitória (65310000), localizada no rio Iguaçu, na cidade de União da Vitória, pertencente à Companhia Paranaense de Energia (COPEL). A referida estação apresenta coordenadas 26° 13'

37" de latitude sul e 51° 04' 48" de longitude oeste, e abrange uma área de drenagem de aproximadamente 24.200 km².

O período de dados observados analisados compreendeu os anos de 1931 a 2015 (85 anos). O Anexo I apresenta a data de ocorrência, o nível máximo anual e a vazão máxima anual dos 85 eventos analisados.

Cabe destacar, conforme detalhado em Steffen (2017), que as vazões do período de 1931 a 1979 resultam dos níveis de água observados na estação fluviométrica União da Vitória (65310000). As vazões do período de 1980 a 2015 resultam dos níveis observados na estação fluviométrica União da Vitória (65310000) e na estação fluviométrica R5 Porto Vitória (65365800), localizada a jusante de União da Vitória e a montante do reservatório da Usina Hidrelétrica Foz do Areia, cujo início de operação ocorreu em 1980.

Na etapa de análise hidrológica preliminar, procurou-se analisar, considerando-se os objetivos do trabalho, as séries históricas de nível e vazão observadas na estação fluviométrica União da Vitória (65310000), principalmente em relação aos eventos de cheia ocorridos na bacia em estudo.

São apresentados os valores de vazões mínima, média e máxima anuais no período de 1931 a 2015 e um panorama dos eventos de cheia máxima anual em termos de superação de um limiar de extravasamento, definido conforme descrito a seguir.

3.3.2 Definição dos eventos de cheias

Em razão dos objetivos do trabalho que envolvem a comparação do agrupamento de cheias observadas na estação fluviométrica União da Vitória (65310000), localizada no rio Iguaçu, na cidade de União da Vitória, obtido pelo método *K-means* e obtido por Steffen (2017), que utilizou o método *Fuzzy C-means*, optou-se por usar os mesmos eventos de cheia analisados pela referida autora.

Na sequência, apresenta-se, de forma resumida, o procedimento usado para a definição dos eventos de cheia máxima anual observados na estação fluviométrica União da Vitória (65310000) em Steffen (2017). Maiores informações, recomenda-se ao leitor consultar o referido trabalho.

Cota e vazão de extravasamento

Estudos realizados na década de 1970 definiram a ocorrência de inundações em União da Vitória quando o nível de água do rio Iguaçu, observado na estação fluviométrica União da Vitória (65310000), ultrapassa a cota 744,50m. Essa definição levou a Companhia Paranaense de Energia (COPEL) a realizar desapropriações em cotas inferiores a 744,50m, visando a criação de uma zona de segurança para a operação do reservatório da Usina Hidrelétrica Foz do Areia, pertencente à empresa (ROCHA, 2012).

Considerando a cota do zero da régua da estação fluviométrica União da Vitória (65310000), correspondente a 739,61m, e a cota de desapropriação da área ribeirinha ao rio Iguaçu em União da Vitória, igual a 744,50m, caracterizando a ocorrência de inundação, tem-se que o nível de água (leitura da régua) que representaria a condição de extravasamento é igual a 4,89 m.

Castanharo e Buba (2008) apresentam uma família de curvas de descarga para a estação fluviométrica União da Vitória (65310000) em função do nível de água observado na estação fluviométrica R5 Porto Vitória (65365800). Portanto, para a cota de extravasamento de 744,50m (leitura de régua = 4,89m) na estação fluviométrica União da Vitória (65310000), pode-se definir diferentes valores de vazão de extravasamento.

O valor de vazão de extravasamento adotado por Steffen (2017) foi igual à 1.387 m³/s, que corresponde à menor vazão para a cota 744,50m (leitura de régua = 4,89m) na estação fluviométrica União da Vitória (65310000), considerando-se ainda as regras de operação da Usina Hidrelétrica Foz do Areia.

A necessidade da definição da cota e da vazão de extravasamento teve por finalidade auxiliar na caracterização dos eventos de cheia observados na área de estudo, uma vez que nem todos os eventos observados ultrapassaram a cota 744,50m, ou seja, não provocaram inundações em União da Vitória. Adicionalmente, a cota e a vazão de extravasamento foram utilizadas para a definição de alguns dos indicadores de cheia como descrito no item 3.3.2.

Delimitação das cheias máximas anuais

Steffen (2017) analisou os hidrogramas das vazões médias diárias na estação fluviométrica União da Vitória (65310000), relativos ao período de 1931 a 2015, visando observar quais anos tiveram uma ou mais ocorrências de cheias e quais foram os maiores eventos registrados e em que mês ocorreram.

O procedimento adotado foi selecionar a maior onda de cheia anual, definida como aquela que continha, em geral, a máxima vazão observada no respectivo ano. Para cada onda de cheia selecionada, foram identificados os menores valores de vazões observadas à esquerda e à direita do pico de cheia, que correspondem, respectivamente, às vazões de início e fim do hidrograma de cheia. Delimitado o hidrograma de cheia, identificou-se os pontos de início e fim do extravasamento, nos eventos em que a vazão de 1.387 m³/s foi superada.

Para a determinação dos instantes de tempo de início e fim do extravasamento, Steffen (2017) adotou a hipótese de variação linear da vazão no intervalo de tempo (1 dia) que continha os referidos instantes, hipótese esta que pode ser representada pela Equação (17):

$$\frac{t_{ext} - t_{inf}}{t_{sup} - t_{inf}} = \frac{Q_{ref} - Q_{inf}}{Q_{sup} - Q_{inf}} \quad (17)$$

Onde: Q_{ref} = vazão de referência (m³/s), igual a 1.387 m³/s;

Q_{sup} = vazão observada imediatamente superior à Q_{ref} (m³/s);

Q_{inf} = vazão observada imediatamente inferior à Q_{ref} (m³/s);

t_{sup} = instante de tempo relativo a Q_{sup} (dia);

t_{inf} = instante de tempo relativo a Q_{inf} (dia);

t_{ext} = instante de início ou fim do extravasamento (dia).

A Figura 14 apresenta a descrição das variáveis utilizadas no cálculo do instante do início do extravasamento (t_{ext}), portanto ocorrendo antes da vazão de pico ($Q_{máx}$). Para o cálculo do instante do fim do extravasamento (t_{ext}), também é utilizada

a Equação (17), porém, com valores observados após a vazão de pico ($Q_{m\acute{a}x}$), como ilustra a Figura 15

FIGURA 14 – CÁLCULO DO INSTANTE DE INÍCIO DO EXTRAVASAMENTO

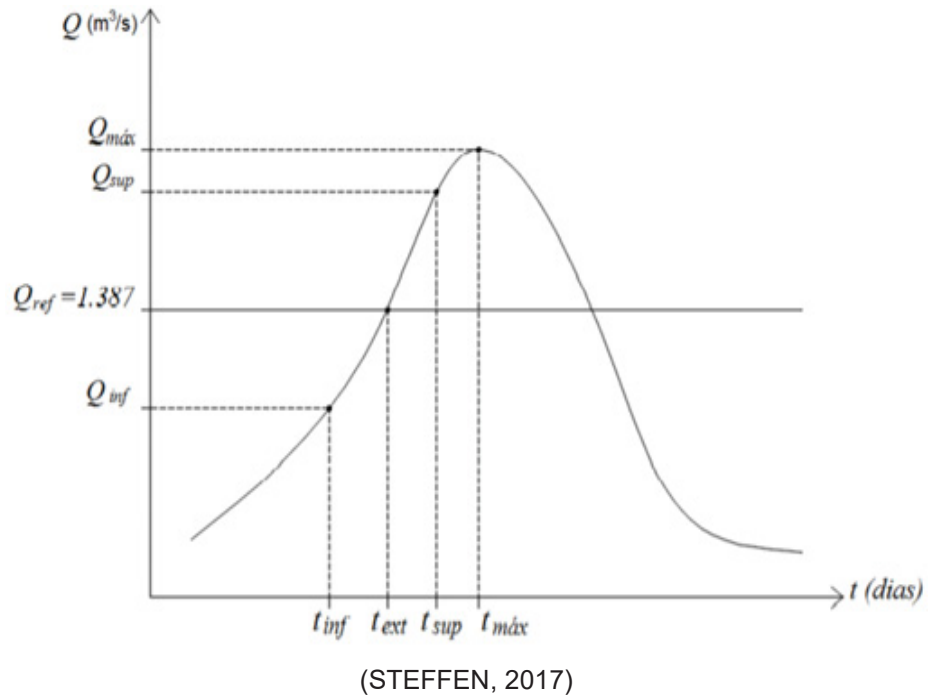
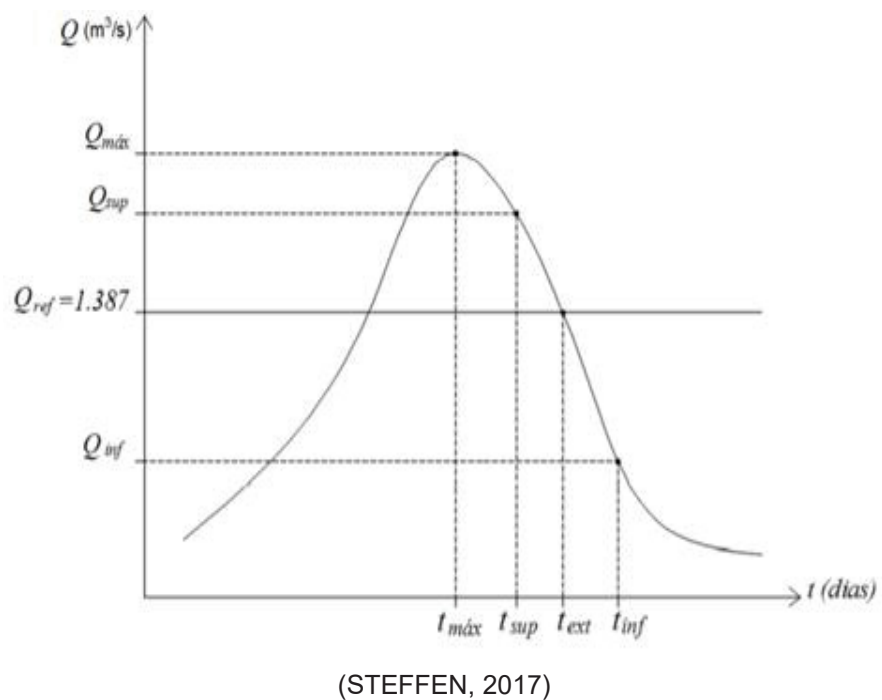


FIGURA 15 – CÁLCULO DO INSTANTE DE FIM DO EXTRAVASAMENTO



3.3.3 Definição dos indicadores de cheia

Novamente, em razão dos objetivos do trabalho que envolvem a comparação do agrupamento de cheias observadas na estação fluviométrica União da Vitória (65310000), localizada no rio Iguaçu, na cidade de União da Vitória, obtido pelo método *K-means* e obtido por Steffen (2017), que utilizou o método *Fuzzy C-means*, optou-se por usar os mesmos indicadores de cheia utilizados pela referida autora, relacionados a seguir:

- 1) vazão máxima anual (I_1);
- 2) nível de água máximo anual (I_2);
- 3) volume total extravasado (I_3);
- 4) tempo acima do limiar de extravasamento (I_4);
- 5) tempo entre o início do evento de cheia e a ultrapassagem do limiar de extravasamento (I_5);
- 6) taxas médias de ascensão do hidrograma de cheia (I_6);
- 7) taxa média de recessão do hidrograma de cheia (I_7).

Na sequência, são descritos de modo resumido cada um dos sete indicadores de cheias utilizados para agrupar as cheias registradas na estação fluviométrica União da Vitória (65310000), conforme Steffen (2017).

Vazão máxima anual (I_1) e Nível de água máximo anual (I_2)

Como já referenciado, os dados de vazão e nível de água utilizados no trabalho são referentes à estação fluviométrica União da Vitória (65310000), pertencente à Companhia Paranaense de Energia (COPEL). Sua operação teve início em 21 de maio de 1930, tendo sido utilizados os dados diários divulgados até dezembro de 2015, resultando, portanto, em uma série de 85 anos completos de dados diários observados.

Como o próprio nome dos indicadores sugere, Steffen (2017) identificou para cada ano, a vazão máxima anual e o nível de água máximo anual, observados na estação fluviométrica União da Vitória (65310000).

Steffen (2017) verificou ainda se os valores máximos anuais não ocorriam próximos do início ou do fim do ano, o que pode caracterizar que os referidos valores

pertenceriam a uma onda de cheia, cujo pico aconteceu em um ano anterior ou em um ano posterior. Nesse caso, buscou-se adotar o maior pico de cheia observado naquele ano. Em outras palavras, pode-se dizer que os indicadores I_1 e I_2 correspondem aos maiores picos de cheia observados em um determinado ano e que, na maioria dos casos, representavam também os máximos valores anuais.

Os valores de I_1 e I_2 , obtidos por Steffen (2017), utilizados no presente trabalho são apresentados no Anexo 1. Cabe ressaltar que, evidentemente, estes dois indicadores (vazão máxima anual e nível máximo anual) são relacionados entre si através da curva de descarga da estação fluviométrica em análise.

Importante aqui também destacar que os indicadores I_1 e I_2 podem ser entendidos como representativos de uma abordagem clássica de avaliação da magnitude de uma cheia, caracterizada somente pelos seus valores máximos de nível e vazão. Evidentemente que maiores valores de I_1 e I_2 estão associados, em tese, a cheias mais severas.

Volume total extravasado (I_3)

O volume total extravasado é obtido pela área do hidrograma acima da vazão de extravasamento (1.387 m³/s).

Para cálculo da área do hidrograma, correspondente ao volume extrapolado, além da definição dos instantes de início (t_{ext0}) e fim do extravasamento (t_{extf}), Steffen (2017) somou todas as áreas relativas a intervalos de tempo em que um determinado hidrograma esteve acima da vazão de extravasamento nas situações em que houve oscilações dos valores da vazão na região próxima à vazão de extravasamento. Os valores de I_3 , obtidos por Steffen (2017), são apresentados no Anexo 2.

Destaca-se que maiores valores de I_3 estão associados, em tese, a cheias mais severas.

Tempo acima do limiar de extravasamento (I_4)

Este indicador de cheia corresponde ao tempo total em que as vazões se mantiveram acima da vazão de extravasamento (1.387 m³/s).

Assim, como no volume total extravasado (I_3), para o cálculo do tempo acima do limiar de extravasamento (I_4), Steffen (2017) somou todos os intervalos de tempo em que um determinado hidrograma esteve acima da vazão de extravasamento nas

situações em que houve oscilações dos valores da vazão na região próxima à vazão de extravasamento. Os valores de I_4 , obtidos por Steffen (2017), são apresentados no Anexo 3.

Este indicador representa o tempo no qual uma determinada onda de cheia provocou inundação na área de estudo. Maiores valores de I_4 estão relacionados, em tese, a cheias mais severas.

Tempo entre o início do evento de cheia e a ultrapassagem do limiar de extravasamento (I_5)

O tempo entre o início do evento de cheia e a ultrapassagem do limiar de extravasamento foi calculado considerando-se o instante da superação da vazão de extravasamento ($1.387\text{m}^3/\text{s}$) e o instante inicial do hidrograma de cheia analisado. Os valores de I_5 , obtidos por Steffen (2017), são apresentados no Anexo 4.

Este indicador representa o tempo no qual uma determinada onda de cheia demora para começar a provocar inundação na área de estudo. De certa forma, representa o tempo disponível para a tomada de ações, visando a mitigação dos impactos iniciais da inundação provocada pela onda de cheia.

De forma diversa, em comparação aos indicadores anteriores, maiores valores de I_5 estão relacionados, em tese, a cheias de menos severas.

Taxa média de ascensão do hidrograma de cheia (I_6)

A taxa média de ascensão do hidrograma de cheia (I_6) foi determinada para o intervalo de tempo entre o início da onda de cheia e o primeiro instante do extravasamento, obtida por meio da Equação (18) (STEFFEN, 2017):

$$I_6 = \frac{Q_{ref} - Q_0}{t_{ext0} - t_0} \quad (18)$$

Onde: I_6 = taxa média de ascensão do hidrograma ($\text{m}^3/\text{s}/\text{dia}$);

Q_{ref} = vazão de referência (m^3/s), igual a $1.387 \text{ m}^3/\text{s}$;

Q_0 = vazão inicial do hidrograma (m^3/s);

t_0 = instante inicial do hidrograma (dia);

t_{ext0} = instante do primeiro extravasamento (dia).

Os valores de I_6 , obtidos por Steffen (2017), são apresentados no Anexo 5.

Este indicador também está associado ao tempo no qual uma determinada onda de cheia demora para começar a provocar inundação na área de estudo. Maiores valores de I_6 estão relacionados, em tese, a cheias mais severas.

Taxa média de recessão do hidrograma de cheia (I_7)

A taxa média de recessão do hidrograma foi calculada como a razão entre a diferença entre a vazão de pico e a vazão de referência, considerando o trecho de recessão do hidrograma, e o intervalo de tempo em que ocorrem, conforme a Equação (19) (STEFFEN, 2017):

$$I_7 = \frac{Q_{m\acute{a}x} - Q_{ref}}{t_{m\acute{a}x} - t_{efpico}} \quad (19)$$

Onde: I_7 = taxa média de recessão do hidrograma (m³/s/dia);

$Q_{m\acute{a}x}$ = vazão de pico (m³/s);

Q_{ref} = vazão de referência (m³/s), igual a 1.387 m³/s;

$t_{m\acute{a}x}$ = instante de ocorrência da vazão de pico do hidrograma (dia);

t_{efpico} = instante de fim do extravasamento do período acima do limiar de extravasamento que contém a vazão de pico (dia).

Importante destacar que t_{efpico} representa o instante no qual o hidrograma atinge novamente, pela primeira vez, a vazão de extravasamento após o pico do hidrograma. Em situações, onde há oscilações dos valores da vazão na região próxima à vazão de extravasamento, o cruzamento limiar de extravasamento pode acontecer em mais de uma vez, mesmo no ramo de recessão do hidrograma.

Os valores de I_7 , obtidos por Steffen (2017), são apresentados no Anexo 6.

Este indicador está relacionado ao tempo no qual uma determinada onda de cheia demora, após atingir o seu valor máximo, para retornar a valores de vazão inferiores à vazão de extravasamento e, como consequência, a uma condição sem inundação. De certa forma, está também relacionada ao tempo em que a área de estudo apresenta inundações provocadas pela onda de cheia.

Assim como o tempo entre o início do evento de cheia e a ultrapassagem do limiar de extravasamento (I_5), de forma diversa, em comparação aos indicadores anteriores, maiores valores de (I_7) estão relacionados, em tese, a cheias menos severas.

3.3.4 Métodos de normalização dos dados

Os processos de mineração de dados (*data mining*), por vezes, apresentam valores impossíveis de serem comparados, devido a diferenças de escala ou variáveis físicas envolvidas; por isso, procedimentos de preparação das amostras são essenciais para garantir a consistência da análise e facilitar a compreensão (MOORE, 1999).

Em geral, de modo a resolver o problema, aplica-se um método de normalização dos dados com objetivo de homogeneizar a variabilidade das variáveis de uma base de dados, criando um intervalo de amplitude similar, onde todas as variáveis irão residir. Em geral, a normalização é necessária no caso de variáveis com unidades diferentes ou dispersões muito heterogêneas (BERNARDI, 2001).

Normalização significa tornar normal, ou seja, tornar a distribuição de frequências dos valores parecida com a distribuição normal ou, pelo menos, torná-la simétrica. A normalidade é um pressuposto de uma série de testes estatísticos univariados e multivariados (MOORE, 1999).

A normalidade plena não é necessária, mas o desempenho dos métodos é melhorado se os dados não são muito assimétricos. A normalização pode ser feita de diferentes maneiras, que requerem o exame da distribuição de frequência dos dados.

A transformação normalizadora geralmente tem a propriedade de estabilizar as variâncias, ou seja, tornar os dados homocedásticos para cada grupo (com a mesma variância), pressuposto para qualquer análise paramétrica (ALVARES, 2017).

Alvares (2017) destaca que, entre as formas mais comuns de normalização, estão:

- Normalização linear;
- Normalização por desvio padrão;
- Normalização pelo valor máximo dos elementos;
- Normalização por escala decimal.

Normalização linear

Visando a normalização dos dados obtidos em seu trabalho, Steffen (2017) aplicou a Equação (20) ao conjunto de valores referente a cada um dos indicadores de cheia utilizados no presente trabalho:

$$x_{ij} = \frac{x_{ij}^0 - x_{jmin}^0}{x_{jmax}^0 - x_{jmin}^0} \quad (20)$$

Onde: x_{ij} = valor tratado da i -ésima observação do j -ésimo indicador;

x_{ij}^0 = valor não tratado da i -ésima observação do j -ésimo indicador;

x_{jmax}^0 = máximo valor não tratado do j -ésimo indicador;

x_{jmin}^0 = mínimo valor não tratado do j -ésimo indicador.

A Equação (20) representa basicamente uma transformação linear, onde os dados originais, que variam em uma escala de x_{jmin}^0 até x_{jmax}^0 , são transformados de modo linear para uma escala entre 0 e 1.

Os valores tratados dos indicadores de cheia, obtidos por Steffen (2017) utilizando a normalização linear, são apresentados no Anexo 7.

Normalização por desvio padrão

A normalização pelo desvio padrão considera a posição média dos valores e o grau de dispersão em relação à posição média. Essa técnica é útil quando se desconhece os valores máximos e mínimos.

A Equação (21) foi utilizada para a normalização pelo desvio padrão dos valores dos indicadores de cheia utilizados no presente estudo:

$$x_{ij} = \frac{(x_{ij}^0 - \mu_j^0)}{\sigma_j^0} \quad (21)$$

Onde: x_{ij} = valor tratado da i -ésima observação do j -ésimo indicador;

x_{ij}^0 = valor não tratado da i -ésima observação do j -ésimo indicador;

μ_j^0 = média dos valores não tratados do j -ésimo indicador;

σ_j^0 = desvio padrão dos valores não tratados do j -ésimo indicador.

A Equação (21) representa basicamente uma transformação linear, onde os dados originais, que variam em uma escala de x_{jmin}^0 até x_{jmax}^0 , são transformados de modo linear para uma escala entre $\frac{(x_{jmin}^0 - \mu_j^0)}{\sigma_j^0}$ e $\frac{(x_{jmax}^0 - \mu_j^0)}{\sigma_j^0}$. Sendo, portanto, uma variação expressa em termos de unidades do desvio padrão de cada indicador.

Os valores tratados dos indicadores de cheia, obtidos no presente trabalho utilizando a normalização pelo desvio padrão, são apresentados no Apêndice 1.

Normalização por escala decimal

A normalização pelo escalonamento decimal move a casa decimal dos valores do atributo x . O número de casas decimais movidas depende do valor máximo absoluto do atributo x .

A normalização por escala decimal foi obtida utilizando-se a Equação (22):

$$x_{ij} = \frac{x_{ij}^0}{10^{k_j}} \quad (22)$$

Onde: x_{ij} = valor tratado da i -ésima observação do j -ésimo indicador;

x_{ij}^0 = valor não tratado da i -ésima observação do j -ésimo indicador;

k_j = menor valor inteiro tal que o máximo valor tratado do j -ésimo indicador

(x_{jmax}), em módulo, resulte ≤ 1 .

A Equação (22) representa basicamente uma transformação linear, onde os dados originais, que variam em uma escala de x_{jmin}^0 até x_{jmax}^0 , são transformados de modo linear para uma escala entre $\frac{x_{jmin}^0}{10^{k_j}} \geq 0$ e $\frac{x_{jmax}^0}{10^{k_j}} \leq 1$, supondo-se $x_{ij}^0 \geq 0$ para qualquer i e j .

Os valores tratados dos indicadores de cheia, obtidos no presente trabalho utilizando a normalização por escala decimal, são apresentados no Apêndice 2.

Normalização pelo valor máximo dos elementos

A normalização pelo valor máximo dos elementos consiste na divisão dos valores observados pelo valor do máximo elemento observado, conforme a Equação (23):

$$x_{ij} = \frac{x_{ij}^0}{x_{jmax}^0} \quad (23)$$

Onde: x_{ij} = valor tratado da i -ésima observação do j -ésimo indicador;

x_{ij}^0 = valor não tratado da i -ésima observação do j -ésimo indicador;

x_{jmax}^0 = máximo valor não tratado do j -ésimo indicador.

A Equação (23) representa basicamente uma transformação linear, onde os dados originais, que variam em uma escala de x_{jmin}^0 até x_{jmax}^0 , são transformados de modo linear para uma escala entre $\frac{x_{jmin}^0}{x_{jmax}^0} \geq 0$ e 1, supondo-se $x_{ij}^0 \geq 0$ para qualquer i e j .

Os valores tratados dos indicadores de cheia, obtidos no presente trabalho utilizando a normalização pelo valor máximo dos elementos, são apresentados no Apêndice 3.

3.3.5 Implementação computacional do método *K-means*

Para implementação do algoritmo do método *K-means* foi escolhido o software de programação *Python*.

Sweigart (2018) explica que *Python* se refere à linguagem de programação (com regras de sintaxe para escrever o que é considerado um código *Python* válido) e ao software do interpretador *Python*, que lê o código fonte (escrito em linguagem Python) e executa suas instruções.

A linguagem de programação *Python* tem uma ampla variedade de construções sintáticas, biblioteca-padrão de funções e recursos de ambiente interativo de desenvolvimento (SWEIGART, 2018).

Ramos (2018), salienta que a simplicidade, a velocidade de aprendizagem, a versatilidade, além do fato de ser multiplataforma, são as principais vantagens da linguagem *Phyton*.

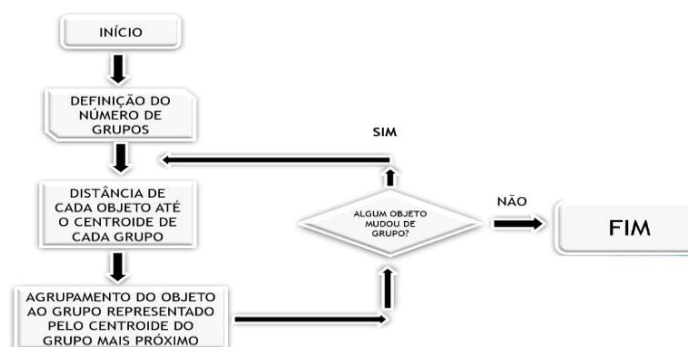
Algoritmo do método *K - means*

A implementação computacional do algoritmo do método *K-means* requer os seguintes passos (LINDEN, 2009):

- 5) escolher k posições iniciais para os centroides dos K grupos (possivelmente, de forma aleatória);
- 6) associar cada elemento da amostra ao grupo cujo centroide é mais próximo do referido elemento;
- 7) recalcular a posição do centroide de cada grupo;
- 8) repetir os passos 2 e 3 até que nenhum elemento da amostra mude de grupo entre duas iterações sucessivas, em razão da mudança da posição dos centroides.

A Figura 16 exemplifica, de forma simplificada, o fluxograma do método *K-means* para o agrupamento de dados.

FIGURA 16 – FLUXOGRAMA DO MÉTODO *K-MEANS*



Determinação do número de grupos para o método *K-means*

Visando a obtenção do número ótimo de agrupamentos, utilizou-se o método da Curva do Cotovelo (*Elbow Curve*), que consiste em encontrar a quantidade ideal

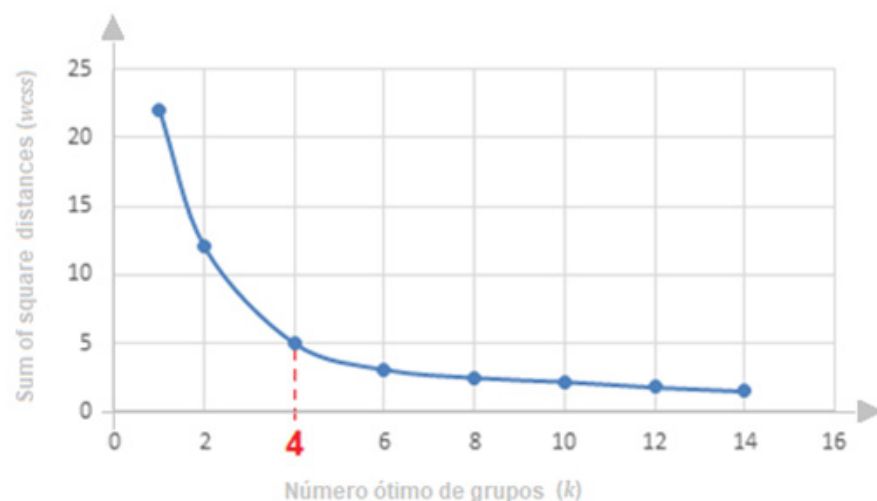
de grupos (k), por meio de testes de variância dos dados em relação ao número de grupos. O valor ideal de número de grupos (k) será aquele em que a soma dos quadrados das distâncias das amostras ao centroide mais próximo (ou do inglês *within-clusters sum-of-squares*, comumente abreviado para *wcss*) seja a menor possível, sendo zero o resultado ótimo.

O método é chamado de Curva do Cotovelo, porque à medida em que aumenta o número de grupos (k), há uma redução da soma dos quadrados das distâncias das amostras ao centroide mais próximo. A referida redução é acentuada para os valores iniciais de k e assume um comportamento linear (mais suave) para valores maiores de k .

O valor ótimo de k representa o ponto no qual a variação da soma dos quadrados das distâncias das amostras ao centroide mais próximo passa de acentuada para um comportamento aproximadamente linear (mais suave), graficamente lembrando a curva de um cotovelo.

A Figura 17 ilustra o método da Curva do Cotovelo para a determinação do valor ideal de número de grupos (k). Na Figura em questão, a qual apresenta os valores das somas dos quadrados das distâncias (*wcss*) no eixo x , e o número de grupos (k) no eixo y , é possível observar claramente tal fenômeno. Observa-se que, quando o número de grupos era apenas 1, a soma dos quadrados das distâncias (*wcss*) foi de 23, e que, quando o número de grupos foi alterado para 4, tal medida (*wcss*) foi apenas 5. Da mesma forma, observa-se que números de grupos superiores a 4 apresentaram comportamento praticamente constante da curva, enfatizando o número de quatro grupos como tendo sido o número de grupos ideal.

FIGURA 17 – NÚMERO ÓTIMO DE GRUPOS



Com o objetivo de definir os valores ótimos do número de grupos (k) e do parâmetro de incerteza (r), Steffen (2017) utilizou o método descrito por Bezdek *et al.* (1983), aplicável ao método *Fuzzy C-means*, estabelecendo um número ótimo de 4 grupos. O referido método não é aplicável ao método *K-means* por depender do próprio parâmetro de incerteza (r) e dos valores da matriz de pertinência u , ou seja, dependem de parâmetros que não definidos para o método *K-means*.

Distância de Mahalanobis, utilizada no algoritmo do método *K-means*

Conforme já relatado, no presente estudo, utilizou-se a distância de *Mahalanobis* para o agrupamento dos dados no lugar da distância Euclidiana, com os objetivos de diminuir a influência dos *outliers*, ou seja, eventos extremos que podem influenciar a composição dos agrupamentos, e considerar a possível correlação existente entre os indicadores de cheia selecionados.

O cálculo da distância matemática, é ,via de regra, executado através da raiz quadrada do quadrado da diferença entre dois pontos, conforme descrito na Equação (24) conhecida como distância Euclidiana.

$$D = \sqrt{(a - b)^2} \quad (24)$$

Onde: D = distância Euclidiana;

a = coordenadas do ponto a;

b = coordenadas do ponto b.

Observa-se que a raiz quadrada tem a função aritmética de fazer com que o resultado da Equação sempre apresente valores positivos, independentemente da posição dos termos.

A distância Euclidiana obtém êxito em calcular as distancias entre pontos, mesmo quando representados por universos multidimensionais (múltiplos planos), contudo, apresenta problemas quando se trata de classificação de dados em lógica difusa, pois não leva em consideração fatores de pertencimento (*Fuzzy*).

Objetivando a resolução de tal problema, o matemático indiano *Prasanta Chandra Mahalanobis* definiu em 1936 uma distância baseada em um conceito de densidade, isto é, o quanto cada elemento pertence ao grupo em questão.

De modo a melhorar a classificação dos grupos, será conveniente, introduzir a ideia de campo estatístico, de modo que em cada ponto nesses campos haja um conjunto específico de parâmetros (α_1, α_2) que definem uma população particular. Em outras palavras, cada ponto em um campo estatístico é o centro de um aglomerado de densidade pertencente a uma determinada população normal completamente especificada pelo valor de (α_i, α_{ij}) naquele ponto específico do campo (MAHALANOBIS, 1936).

O conceito de pertencimento é calculado através da matriz de covariância.

A definição da distância de *Mahalanobis* já foi mostrado no presente trabalho, porém, apenas a título de situar o leitor neste capítulo, além de prepara-lo para explicações futuras, cabe apresenta-lo novamente; sendo ele representado, entre um grupo de valores com média $\alpha = (\alpha_1, \alpha_2, \alpha_3, \dots, \mu_p)^T$ e matriz de covariância S para um vetor multivariado $x = (x_1, x_2, x_3, \dots, x_p)^T$ definio pela Equação (25) (MAHALANOBIS, 1936).

$$D^2 = (x - \alpha)^T C^{-1} (x - \alpha) \quad (25)$$

Onde: D^2 = Distância de *Mahalanobis*.

x = Vetor

α = Vetor médio das variáveis independentes

C^{-1} = Matriz inversa de covariância das variáveis independentes

T = indicativo de transposição da matriz.

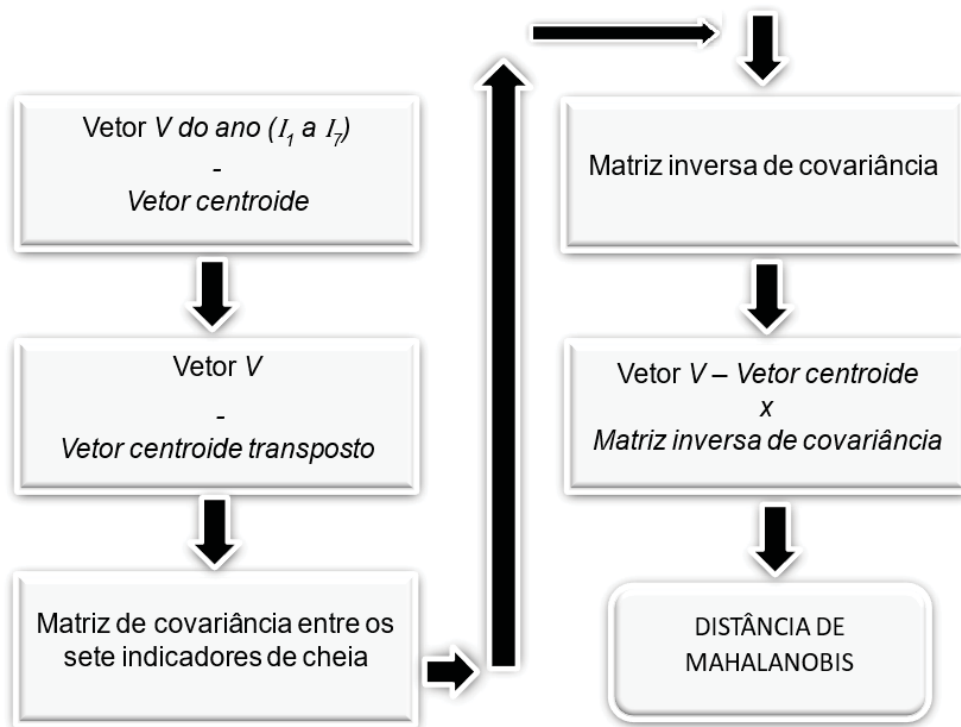
No caso do presente trabalho, a matriz x é uma matriz coluna com 7 linhas, correspondendo aos 7 indicadores de cheia (I_1 a I_7), calculados para um determinado evento de cheia. A matriz μ também é uma matriz coluna com 7 linhas, correspondendo às coordenadas dos centroides de cada grupo. No presente trabalho, cada centroide apresenta 7 coordenadas, correspondentes a cada um dos sete indicadores de cheia.

Se a matriz de covariância é a matriz identidade, a distância de *Mahalanobis* coincide com a distância Euclidiana. Se a matriz de covariância é diagonal, então a medida de distância resultante é chamada distância Euclidiana normalizada, obtida pela Equação (26) (VICINI, 2005):

$$d(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1}^p \left(\frac{x_i - y_i}{\sigma_i^2} \right)^2} \quad (26)$$

O cálculo da distância de *Mahalanobis* seguiu, a título de ilustração, a sequência descrita na Figura 18.

FIGURA 18 – REPRESENTAÇÃO GRÁFICA DO CÁLCULO DA DISTÂNCIA DE MAHALANOBIS.



Cabe ressaltar que a distância de *Mahalanobis* é calculada para cada amostra e em relação aos centroides dos k grupos utilizados no agrupamento. No presente trabalho, tem-se 85 amostras, cada qual apresentando um valor para cada um dos 7 indicadores de cheia.

Para o cálculo da referida distância, utilizou-se a função *Mahalanobis* da biblioteca-padrão de funções da linguagem *Python*. Dois exemplos do cálculo da distância de *Mahalanobis* são apresentados nos Apêndices 18 a 19.

3.3.6 Análise da influência do método de normalização no agrupamento das cheias

Para a análise da influência do método de normalização no agrupamento das cheias, utilizou-se os testes estatísticos *t-Student* e *F* e a análise da variância (ANOVA).

Comparação das médias

As médias obtidas a partir das amostras ($\bar{x}$) são estimativas de parâmetros populacionais (μ) que são desconhecidos. Além disso, estão sujeitas a variação em decorrência do efeito do acaso na seleção dos sujeitos. Portanto, inferir sobre a diferença entre médias populacionais com base em médias amostras é uma situação especial (MARDIA *et al.*, 1979).

O teste *t* de *Student*, proposto em 1908 por William Sealy Gosset, é uma das ferramentas mais empregadas para este tipo de análise. Quando duas amostras independentes são tomadas ao acaso e a variável tem distribuição normal, a comparação de suas médias pode ser realizada usando o teste *t* de *Student* (HAIR *et al.*, 2009).

A estatística *t* é estimada pela Equação (27):

$$t = \frac{\bar{x}_1 - \bar{x}_2}{EP_{(\bar{x}_1 - \bar{x}_2)}} \quad (27)$$

Onde: $\bar{x}_1 - \bar{x}_2$ = diferença entre as médias dos dois grupos;

$EP_{(\bar{x}_1 - \bar{x}_2)}$ = erro padrão da diferença entre as duas médias.

Quando as variâncias populacionais são iguais, o erro padrão da diferença é calculado a partir da média ponderada das variâncias dos grupos, também chamada de variância ponderada (s_p^2), através da Equação (28) (BROWNLEE, 1965):

$$s_p^2 = \frac{s_1^2 (n_1 - 1) + s_2^2 (n_2 - 1)}{n_1 + n_2 - 2} \quad (28)$$

Onde: s_1^2 e n_1 representam a variância e número de casos do grupo 1;

s_2^2 e n_2 representam a variância e número de casos do grupo 2.

O erro padrão da diferença $EP_{(\bar{x}_1 - \bar{x}_2)}$ é então calculado, usando a variância (s_p^2), por meio da Equação (29):

$$EP_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} \quad (29)$$

Quando as variâncias populacionais são diferentes, o erro padrão da diferença é determinado diretamente, a partir das variâncias dos grupos, conforme a Equação (30) (HAIR *et al.*, 2009):

$$EP_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad (30)$$

O valor da estatística t , calculado pela Equação (27), é então comparado ao valor de crítico de t , obtido, por exemplo, em tabelas da distribuição t de *Student*, considerando o número de graus de liberdade, o nível de significância (α) e o tipo de teste de hipótese (unilateral ou bilateral) (MARDIA *et al.*, 1979).

Para variâncias populacionais iguais, o número de graus de liberdade é dado pela Equação (31):

$$gl = n_1 + n_2 \quad (31)$$

Para variâncias populacionais distintas, o número de graus de liberdade é dado pela Equação (32):

$$gl = \frac{(a+b)^2}{\frac{a^2}{n_1-1} + \frac{b^2}{n_2-1}} \quad (32)$$

Onde $a = \frac{s_1^2}{n_1}$ e $b = \frac{s_2^2}{n_2}$.

Comparação das variâncias

Para a comparação das variâncias de dois grupos independentes, emprega-se a estatística F , obtida a partir da Equação (33), definida como a razão entre a variância maior e a variância menor (TABACHICK *et al.*, 2001):

$$F = \frac{s_{maior}^2}{s_{menor}^2} \quad (33)$$

O valor de F calculado é comparado ao valor de F crítico obtido, por exemplo, em tabela da distribuição F , considerando o nível de significância (α) e os números de graus de liberdade do grupo com variância maior e do grupo com variância menor (TABACHICK *et al.*, 2001).

Análise de Variância (ANOVA)

A análise de variância (ANOVA) constitui um método estatístico que permite realizar comparações simultâneas entre duas ou mais médias, ou seja, permite testar hipóteses sobre médias de distintas populações. Basicamente divide a variabilidade em variabilidade entre grupos e variabilidade dentro de grupos, e compara as duas. Quanto maior for a primeira comparada à segunda, maior é a evidência de que existe variabilidade entre grupos, ou seja, médias diferentes (BROWNLEE, 1965).

O procedimento se inicia com o cálculo da soma de quadrados total (SQT), conforme Equação (34) (HAIR *et al.*, 2009):

$$SQT = \sum (x_i - \bar{x})^2 \quad (34)$$

Onde: SQT = soma dos quadrados totais;

x_i = amostra;

$\bar{x}$ = média amostral global.

A estimativa usual de variância amostral é dada por (TABACHNICK *et al.*, 2001):

$$S_{n-1}^2 = SQT/(n - 1) \quad (35)$$

Onde: S_{n-1}^2 = variância amostral global;

SQT = soma dos quadrados totais;

n = número de amostras.

O cálculo de SQT também pode ser entendido, conforme o expresso pela Equação (36):

$$SQT = SQD + SQE \quad (36)$$

Onde SQD é definido por:

$$SQD = \sum_{g1} (x_i - \bar{x}_1)^2 + \sum_{g2} (x_i - \bar{x}_2)^2 + \dots \sum_{gk} (x_i - \bar{x}_k)^2 \quad (37)$$

Onde $\bar{x}_k$ é a média amostral do grupo k , e que n_k é o número de amostras do grupo k . Na Equação (35), SQD representa a soma dos desvios dentro dos grupos e SQE a soma dos desvios entre grupos (TABACHNICK *et al.*, 2001).

Agora tendo separado a variabilidade, é possível obter estimativas independentes da variância populacional comum (σ^2), a partir dos valores de SQD e SQE .

Com os valores de SQD e SQE , pode definir uma variância entre grupos (s_1^2) e uma variância dentro dos grupos, dadas, respectivamente pelas Equações (38) e (39) (HAIR *et al.*, 2009):

$$s_1^2 = SQE/(m - 1) \quad (38)$$

$$s_2^2 = SQD/(N - 1) \quad (39)$$

Onde: m = número de grupos;

N = é o tamanho amostral.

Como estas estimativas de variância são construídas a partir de dois tipos diferentes de variabilidade, quanto mais elas diferirem, mais evidência existe de diferença nas médias entre os grupos (BERTOLDO *et al.*, 2008).

A estatística da ANOVA é definida pela Equação (40):

$$F = s_1^2/s_2^2 \quad (40)$$

O valor da estatística F , calculada a partir da Equação (40), é então comparada a um valor crítico dado por uma distribuição F com $(m-1)$ e $(N-1)$ graus de liberdade, obtendo-se um p - *valor* (BERTOLDO *et al.*, 2008).

Em geral, os resultados da uma ANOVA são mostrados em forma tabular, semelhante à Tabela 3, apresentada a seguir.

TABELA 3 - FORMA DE APRESENTAÇÃO DOS RESULTADOS DA ANOVA

Forma de variação	Soma dos quadrados	Graus de liberdade	Quadrado médio	F	p - <i>valor</i>
Entre grupos	SQE	$gl_e = k-1$	$MQ_0 = \frac{SQ_e}{gl_N}$	$\frac{MQ_e}{MQ_d}$	p - <i>valor</i>
Dentro dos grupos	SQD	$gl_d = n-k$	$MQ_0 = \frac{SQ_N}{gl_D}$		
Total	SQT	$gl_t = n-1$			

3.3.7 Determinação dos indicadores de cheia mais representativos

A determinação dos indicadores de cheia mais representativos utilizou a Análise dos Componentes Principais (ACP), descrita no item 2.5. Neste ponto, cabe descrever as ferramentas utilizadas no presente trabalho:

Utilizou-se as ferramentas disponíveis no programa computacional *Microsoft Excel* e no seu suplemento *Real Statistics*, por ser um suplemento gratuito e de fácil utilização.

Foram analisadas todas as amostras do estudo, sendo estas compostas de:

- valores não tratados;
- valores normalizados pelo método da normalização linear;
- valores normalizados pelo método do desvio padrão;
- valores normalizados pelo método da escala decimal;

e. valores normalizados pelo método do valor máximo dos elementos.

O programa computacional fornece, entre outros, os valores da matriz de correlação, inversa da matriz de correlação, matriz fatorial, matriz de erro, autovalores e autovetores. Os resultados são disponibilizados em planilhas, que estão apresentadas nos Apêndices 38 a 41.

Apenas a título ilustrativo, a Figura 19 apresenta a matriz de autovalores e autovetores correspondente aos valores não tratados.

FIGURA 19- AUTOVALORES E AUTOVETORES. VALORES NÃO TRATADOS.

		65,78%	20,08%	7,9%	4,0%	1,1%	1,0%	0,04%	$\Sigma = 100\%$
AUTOVALORES		4,5981	1,4039	0,5589	0,2843	0,0808	0,0710	0,0029	$\Sigma = 6,99$
AUTOVETORES	I_1 (m³/s)	0,4325	-0,2413	-0,0006	0,4371	-0,1647	-0,0659	0,7296	
	I_2 (cm)	0,4413	-0,1579	0,0326	0,4438	-0,3769	0,0160	-0,6633	
	I_3 (Km³)	0,3629	-0,4902	-0,0747	-0,1825	0,6603	-0,3608	-0,1515	
	I_4 (dias)	0,4065	-0,2268	0,1217	-0,6989	-0,3359	0,4039	0,0635	
	I_5 (dias)	0,3775	0,4539	0,0024	0,1772	0,5059	0,6033	-0,0111	
	I_6 (m³/s/dia)	0,2992	0,4195	-0,7585	-0,1953	-0,1586	-0,3092	0,0140	
	I_7 (m³/s/dia)	0,2987	0,4930	0,6349	-0,1439	-0,0302	-0,4924	0,0215	

A interpretação da Figura 19 deve ser feita da seguinte forma:

- 1) Na parte superior da figura, são apresentados os autovalores e seus pesos proporcionais. Observa-se, por exemplo, que o primeiro autovalor (4,5981) representa 65,78% da representatividade das amostras e o segundo autovalor (1,4039) representa 20,08%; totalizando 85,86% da representatividade das amostras. Como a métrica definida no presente trabalho é definir os indicadores que representem 80% das amostras, as duas primeiras colunas atendem o requisito.
- 2) Posteriormente, deve-se verificar quais autovetores correspondem aos autovalores previamente selecionados (no presente caso, as duas primeiras colunas);
- 3) Para identificar os indicadores de cheia mais representativos, seleciona-se, em cada uma das duas primeiras colunas, que correspondem aos maiores autovalores, o indicador que apresenta o valor mais alto da coluna (em módulo). Neste exemplo, a primeira coluna (autovetor), que

corresponde ao maior autovalor (4,5981), apresenta como maior valor em módulo 0,4413 que corresponde ao indicador I_2 . Na sequência, a segunda coluna (autovetor), que corresponde ao segundo maior autovalor (1,4039), apresenta como maior valor em módulo 0,4930, relacionado ao indicador I_7 .

Pelo exposto, é possível afirmar que os indicadores I_2 e I_7 correspondem, neste caso, a 85,86% da representatividade das amostras.

3.3.8 Redefinição do tratamento dos indicadores I_5 e I_7

Conforme exposto no item 3.3.3, a maioria dos indicadores de cheia, em tese, apresentam uma relação direta com a severidade das cheias, ou seja, quanto maior o indicador, mais severa é a cheia. Exceção feita a dois indicadores, o tempo entre o início do evento de cheia e a ultrapassagem do limiar de extravasamento (I_5) e a taxa média de recessão do hidrograma de cheia (I_7).

Uma análise dos indicadores de cheia e dos agrupamentos propostos por Steffen (2017) resultaram, no presente trabalho, em uma proposição de redefinição dos indicadores I_5 e I_7 que passaram a ser tratados do seguinte modo:

- 1) cálculo do valor do indicador para cada uma das amostras, considerando-se os valores não tratados. Valores obtidos por Steffen (2017), apresentados nos Anexos 1 a 6;
- 2) determinação do mínimo valor e do máximo valor do indicador, a partir dos valores obtidos no passo 1;
- 3) cálculo, para cada amostra, da diferença entre o máximo valor do indicador, obtido no passo 2, e o valor do indicador para a respectiva amostra, obtido no passo 1;
- 4) cálculo, para cada amostra, do valor corrigido do indicador, resultante da diferença obtida no passo 3, acrescida do mínimo valor do indicador, obtido no passo 2.

O procedimento descrito foi aplicado somente aos valores dos indicadores I_5 e I_7 que resultaram diferentes de zero. Nas situações em que os indicadores I_5 e I_7 resultaram iguais a zero, os valores não foram corrigidos, ou seja, foram mantidos os valores obtidos no passo 1.

4 ANÁLISE DOS RESULTADOS

Neste capítulo são apresentados os resultados obtidos no desenvolvimento do presente trabalho.

Inicialmente, é apresentada uma comparação entre os agrupamentos das cheias históricas observadas no rio Iguaçu na estação fluviométrica União da Vitória (6531000), obtidos pelo método *Fuzzy C-means* e pelo método *K-means* com o objetivo de verificar a influência do método sobre os agrupamentos formados. Na sequência apresenta-se uma análise da influência dos métodos de normalização dos dados (pré-tratamento) sobre os agrupamentos obtidos pelo método *K-means*.

Posteriormente são apresentados os agrupamentos obtidos utilizando-se a metodologia *K-means* e os indicadores I_5 e I_7 alterados, conforme já detalhado no capítulo 3.

Em passo seguinte é utilizada a técnica estatística Análise de Componentes Principais, visando a minimização dos indicadores de magnitude de cheia.

Finalmente, apresenta-se um resumo dos resultados obtidos de modo a destacar a influência ocasionada pelos métodos de normalização aplicados, assim como as alterações suscitadas nos indicadores I_5 e I_7 , agrupados através do método *K-means* e as comparações com o método *Fuzzy C-means*, executado por Steffen em 2017, além dos dados obtidos com a redução dimensional no número de indicadores de magnitude de cheias, conseguido através da aplicação da Análise de Componentes principais.

4.1 COMPARAÇÃO DOS AGRUPAMENTOS FEITOS PELOS MÉTODOS *K-MEANS* E *FUZZY C-MEANS*

4.1.1 Determinação do número de grupos

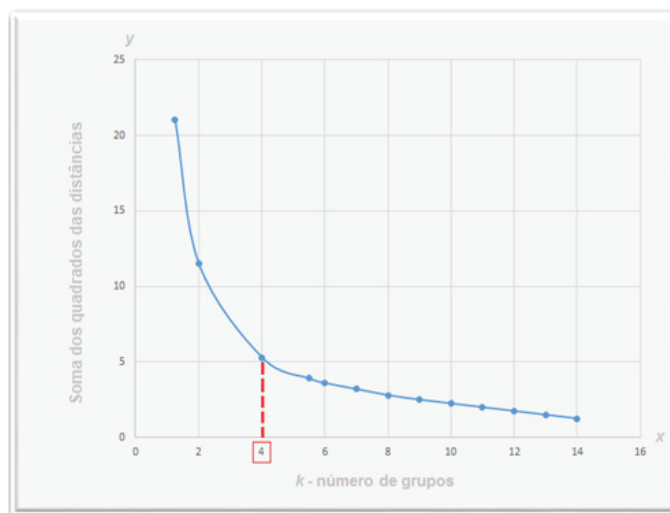
Foi desenvolvido um programa computacional em linguagem de programação *Python*, descrito nos Apêndices 21 a 33. O número de agrupamentos foi calculado pelo método da Curva Cotovelo (*elbow curve*) para o algoritmo *K-means*, obtendo-se em um valor ótimo de quatro grupos, conforme representado na Figura 20 e no cálculo executado no Apêndice 20. Steffen (2017) obteve o mesmo valor ótimo de número de grupos, a partir da análise da matriz de partição difusa.

A Figura 20 apresenta o cálculo do número ótimo dos grupos através do método da curva cotovelo. O método consiste em calcular a soma dos quadrados das distâncias de cada elemento até o centroide de cada grupo. É possível observar

na referida figura, que, à medida que se aumenta o número de grupos, existe uma estabilização e um crescimento progressivo na soma dos quadrados das distâncias, calculada entre cada elemento e o centroide de cada grupo. O número ótimo de grupos é justamente no ponto de inflexão da curva, o qual, no caso em questão, ocorre com 4 grupos.

Conforme já descrito no presente trabalho, o algoritmo *Fuzzy C-means* consiste em uma variação do algoritmo *K-means*, isso explica o fato do número de grupos resultar igual para os dois métodos usados.

FIGURA 20 - SOMA DO QUADRADO DAS DISTÂNCIAS, EM FUNÇÃO DO NÚMERO DE GRUPOS K .



4.1.2 Agrupamento pelo método *K-means*

A primeira análise executada no presente trabalho teve como objetivo agrupar as cheias históricas observadas no rio Iguaçu na estação fluviométrica União da Vitória (6531000), usando o método *K-means* e a mesma base de dados produzida por Steffen (2017), que utilizou o método da normalização linear para o tratamento dos dados brutos. As Tabelas 4 e 5 apresentam, respectivamente, os centroides e os agrupamentos obtidos pelo método *K-means*.

Os valores dos indicadores (I_1 a I_7) normalizados pelo método de normalização linear de todos os eventos analisados no presente trabalho é apresentada no Anexo 7. Os resultados do programa computacional em linguagem

Python para agrupamento pelo método *K-means* e dados normalizados pelo método da normalização linear são apresentados no Apêndice 21.

TABELA 4 – CENTRÓIDES CALCULADOS PELO MÉTODO *K-MEANS*

magnitude da cheia	I_1	I_2	I_3	I_4	I_5	I_6	I_7
baixa	0,1028	0,1567	0,0000	0,0010	0,0033	0,0083	0,0042
Média	0,2302	0,3200	0,0194	0,1288	0,4045	0,1117	0,1651
Alta	0,3422	0,4406	0,0950	0,2693	0,1852	0,2795	0,3971
Excepcional	0,5616	0,6449	0,3842	0,6027	0,4040	0,1439	0,4796

Observa-se na Tabela 4 que, à exceção dos indicadores I_5 (tempo entre o início do evento de cheia e a ultrapassagem do limiar de extravasamento) e I_6 (taxa média de ascensão do hidrograma de cheia), a coordenada do centroide correspondente a cada indicador aumenta no sentido do grupo de cheias de magnitude **baixa** para o grupo de cheias de magnitude **excepcional**.

TABELA 5 - AGRUPAMENTO DAS CHEIAS HISTÓRICAS PELO MÉTODO *K-MEANS*

Magnitude da cheia						
Baixa		Média		Alta		Excepcional
1933	1964	1931	1999	1932	2010	1935
1934	1967	1937	2004	1938	2011	1946
1936	1968	1939	2007	1947	2013	1957
1940	1969	1948	2008	1954	2015	1983
1941	1974	1950	2012	1955		1992
1942	1977	1953		1971		1997
1943	1978	1961		1972		1998
1944	1985	1965		1982		2014
1945	1986	1966		1984		
1949	1991	1970		1987		
1951	1994	1973		1989		
1952	2002	1975		1990		
1956	2003	1976		1993		
1958	2006	1979		1995		
1959		1980		2000		
1960		1981		2001		
1962		1988		2005		
1963		1996		2009		

Considerando-se, portanto, a Tabela 5 mostra que, do total de 85 eventos analisados, 32 eventos foram classificados como magnitude **baixa**, 23 como magnitude **média**, 22 como magnitude **alta** e 8 como magnitude **excepcional**.

4.1.3 Comparação entre os métodos *K-means* e *Fuzzy C-means*

Esta etapa do trabalho visou comparar os resultados obtidos pelos métodos *K-means* e *Fuzzy C-means*, avaliando influência do método (algoritmo de agrupamento) na formação dos grupos.

Conforme citadas anteriormente, as Tabelas 4 e 5 no item anterior apresentam, respectivamente, os centroides e os agrupamentos obtidos pelo método *K-means*. As Tabelas 6 e 7 a seguir mostram, respectivamente, os centroides e os agrupamentos obtidos pelo método *Fuzzy C-means*. (Steffen, 2017). Como já descrito, os valores normalizados através do método de normalização linear, e agrupados através do programa de agrupamento de dados, executado em linguagem de programação *Python*, se encontra no Apêndice 21.

TABELA 6 - CENTRÓIDES CÁLCULADOS PELO MÉTODO *FUZZY C-MEANS*

magnitude da cheia	<i>I₁</i>	<i>I₂</i>	<i>I₃</i>	<i>I₄</i>	<i>I₅</i>	<i>I₆</i>	<i>I₇</i>
Baixa	0,1018	0,1552	0,0001	0,0010	0,0026	0,0046	0,0025
Média	0,2375	0,3273	0,0259	0,1447	0,4041	0,1166	0,1737
Alta	0,3481	0,4471	0,1032	0,2905	0,2048	0,2647	0,3952
Excepcional	0,6310	0,7059	0,4628	0,6102	0,3271	0,1730	0,5655

FONTE: STEFFEN (2017)

TABELA 7 - AGRUPAMENTO DAS CHEIAS HISTÓRICAS PELO MÉTODO *FUZZY C-MEANS*

magnitude da cheia						
Baixa		média		Alta		excepcional
1931	1962	1937	2004	1932	2001	1935
1933	1963	1939	2007	1938	2005	1946
1934	1964	1948	2008	1946	2009	1957
1936	1967	1953	2012	1947	2010	1983
1940	1968	1961		1950	2011	1992
1941	1969	1965		1954	2013	1997
1942	1974	1966		1955	2015	1998
1943	1977	1970		1971		2014
1944	1978	1973		1972		
1945	1985	1975		1982		
1949	1986	1976		1984		
1951	1991	1979		1987		
1952	1994	1980		1989		
1956	2002	1981		1990		
1958	2003	1988		1993		
1959	2006	1996		1995		
1960		1999		2000		

Importante destacar que, no presente trabalho, utilizou-se uma nomenclatura para os grupos distinta da usada por Steffen (2017), que estabeleceu como nome para os 4 (quatro) grupos: severidade baixa; severidade média; severidade alta; e desastres ou catástrofe. De modo equivalente, no presente trabalho, utilizou-se: magnitude baixa; magnitude média; magnitude alta; e magnitude excepcional. De modo sucinto, a principal diferença está em utilizar o termo magnitude para caracterizar a cheia e não severidade, por entender que o termo severidade remete aos impactos socioeconômicos das cheias e que não foram avaliados neste estudo.

A Tabela 8 apresenta a diferença (em valores absolutos) entre as coordenadas e a distância Euclidiana dos centroides dos 4 grupos obtidos pelos métodos *K-means* e *Fuzzy C-means*, a partir dos resultados apresentados nas Tabelas 4 e 6.

TABELA 8 - DIFERENÇA (EM VALORES ABSOLUTOS) ENTRE AS COORDENADAS E A DISTÂNCIA EUCLIDIANA DOS GRUPOS OBTIDOS PELOS MÉTODOS *K-MEANS* E *FUZZY C-MEANS*

magnitude da cheia	Indicador							Distância Euclidiana (centroide)
	<i>I</i> ₁	<i>I</i> ₂	<i>I</i> ₃	<i>I</i> ₄	<i>I</i> ₅	<i>I</i> ₆	<i>I</i> ₇	
Baixa	0,0010	0,0015	0,0001	0,0000	0,0007	0,0037	0,0017	0,0045
Média	0,0073	0,0073	0,0065	0,0159	0,0004	0,0049	0,0086	0,0224
Alta	0,0059	0,0065	0,0082	0,0212	0,0196	0,0148	0,0019	0,0346
Excepcional	0,0694	0,0610	0,0786	0,0075	0,0769	0,0291	0,0859	0,1700
Distância Euclidiana (indicador)	0,0700	0,0618	0,0793	0,0275	0,0794	0,0332	0,0864	

Os resultados apresentados na Tabela 9 mostram que as coordenadas e, consequentemente, a distância Euclidiana entre os centroides dos grupos correspondentes obtidos pelos dois métodos, *K-means* e *Fuzzy C-means*, resultaram relativamente próximas entre si. A maior distância Euclidiana entre os centroides dos grupos correspondentes foi a distância entre os centroides do grupo 4 (magnitude excepcional), cujo valor obtido foi igual a 0,1700.

Em termos de distância entre as coordenadas de cada indicador em específico, as maiores diferenças foram observadas para o indicador *I*₇, cujo valor obtido foi igual a 0,0864. Contudo, destaca-se que a distância entre as coordenadas

de cada indicador foi afetada sobretudo pelas distâncias observadas no grupo 4 (magnitude excepcional).

Como consequência da proximidade entre as coordenadas dos centroides dos 4 grupos definidos por cada um dos métodos de agrupamento, observa-se um comportamento similar, ou seja, à exceção dos indicadores I_5 (tempo entre o início do evento de cheia e a ultrapassagem do limiar de extravasamento) e I_6 (taxa média de ascensão do hidrograma de cheia), a coordenada do centroide correspondente a cada indicador aumenta no sentido do grupo de cheias de magnitude baixa para o grupo de cheias de magnitude excepcional ou, segundo a nomenclatura usada por Steffen (2017), do grupo de severidade baixa para o grupo de desastres ou catástrofes.

A comparação entre os resultados apresentados nas Tabelas 5 e 7 mostra que, dos 85 eventos analisados, somente 3 (três) eventos apresentaram classificação distinta nos dois métodos utilizados, *K-means* e *Fuzzy C-means*, conforme destacados na Tabela 9.

TABELA 9 - EVENTOS COM CLASSIFICAÇÃO DISTINTA NOS AGRUPAMENTOS PELOS MÉTODOS *K-MEANS* E *FUZZY C-MEANS* (STEFFEN, 2017) COM NORMALIZAÇÃO LINEAR

Ano	<i>Fuzzy C-means</i>	<i>K-means</i>
1931	Baixa	Média
1950	Alta	Média
2011	Média	Alta

A Tabela 10 apresenta a distância Euclidiana entre os eventos apresentados na Tabela 09 em relação aos centroides dos 4 grupos definidos pelos métodos *K-means* e *Fuzzy C-means*. Importante destacar que a posição no espaço de um evento é definida pelas suas coordenadas que correspondem aos valores dos sete indicadores calculados para cada evento.

TABELA 10 - DISTÂNCIA EUCLIDIANA ENTRE OS EVENTOS 1931, 1950 E 2011 E OS CENTROIDES DOS GRUPOS DEFINIDOS PELOS MÉTODOS *K-MEANS* E *FUZZY C-MEANS*

Evento	Método	magnitude da cheia			
		baixa	Média	alta	Excepcional
1931	<i>K-means</i>	0,3539	0,3552	0,4277	0,9777
	<i>Fuzzy C-means</i>	0,3583	0,3600	0,4487	1,0744
1950	<i>K-means</i>	0,4558	0,3354	0,2980	0,8725
	<i>Fuzzy C-means</i>	0,4600	0,3331	0,3218	0,9646
2011	<i>K-means</i>	1,2898	0,8246	0,9058	0,6740
	<i>Fuzzy C-means</i>	1,2914	0,8101	0,8752	0,8228

Observa-se, a partir da Tabela 10, que os eventos 1931 e 1950 apresentaram, para os dois métodos analisados, as distâncias Euclidianas mínimas relativamente a grupos correspondentes. Portanto, a diferença de classificação, destacada na Tabela 9, deveu-se ao fato de se utilizar a distância de *Mahalanobis* no método *K-means* e a distância Euclidiana no método *Fuzzy C-means*, e não as diferenças entre os métodos em si. Os Apêndices 18 e 19 demonstram o cálculo da proximidade dos indicadores do ano de 1931 até os centroides de baixa e média severidade utilizando-se a distância de *Mahalanobis*; é possível verificar que, quando se utilizou a distância de *Mahalanobis* (Apêndices 18 e 19), o vetor do ano de 1931 ficou mais próximo do centroide de média severidade.

Em relação ao evento 2011, verifica-se que o método *K-means* classificou o evento como grupo 3 (magnitude alta), mas a mínima distância Euclidiana ao centroide foi observada em relação ao grupo 4 (magnitude excepcional). No método *Fuzzy C-means*, a menor distância Euclidiana foi observada para o centroide do grupo 2 (magnitude média). Contudo, destaca-se que as distâncias Euclidianas para os centroides dos grupos 2 (magnitude média) e 4 (magnitude excepcional) resultaram, respectivamente, 0,8101 e 0,8228, muito próximas entre si para o método *Fuzzy C-means*, o que mostra, em termos de lógica difusa, que a classificação do evento 2011 não seria muito bem definida. Novamente, a diferença entre a classificação para estar mais associada ao método de cálculo das distâncias do que ao método de agrupamento em si.

Os programas descritos nos Apêndices 21 a 33 calculam as distâncias de *Mahalanobis* entre todos os 85 vetores (1931 a 2015) e os quatro vetores dos grupos de magnitude de cheia (baixa, média, alta e excepcional) para todos os métodos de normalização utilizados neste trabalho.

4.2 INFLUÊNCIA DOS MÉTODOS DE NORMALIZAÇÃO DOS DADOS NO AGRUPAMENTO DE CHEIAS

A normalização visa possibilitar a comparação dos valores dos indicadores I_1 a I_7 através da supressão da mensuração dimensional, porém, resta dúvida sobre a influência dos métodos de normalização sobre o agrupamento das amostras. Visando responder essa questão, realizou-se o agrupamento dos dados normalizados por meio de quatro métodos de normalização distintos: linear; por desvio padrão, por escala decimal; e pelo valor máximo do elemento.

Visando o refinamento das análises, foram realizados testes de hipóteses primeiramente baseados em médias (teste t de *Student*), pareados sempre dois a dois, comparando os valores obtidos através de normalização linear e os valores normalizados pelos outros métodos de normalização (Apêndices 34 e 35). Posteriormente, foram efetuados testes estatísticos de variância, conhecidos como ANOVA (Análise de Variância), procedimento usado para comparar a distribuição de três ou mais grupos em amostras independentes (Apêndices 36 e 37).

Os resultados obtidos mostraram que não existe correlação estatística entre os valores normalizados pelos diferentes métodos de normalização utilizados no trabalho. Existem diferenças representativas nos principais indicadores estatísticos, sendo eles minimamente: média, desvio padrão, variância, 1º, 2º e 3º quartil. Em todos os testes estatísticos realizados (teste t e ANOVA), foram encontrados valores de p -valor menores que 0,05, demonstrando a inexistência de correlação estatística entre as amostras de valores normalizados.

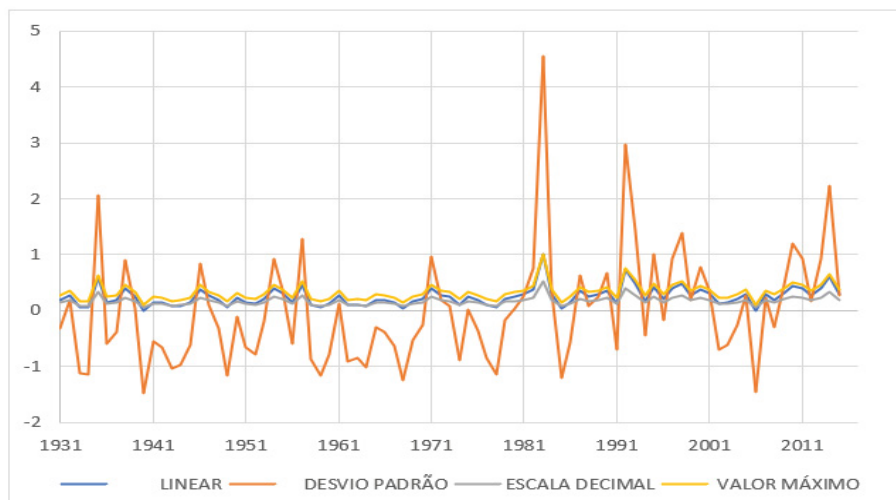
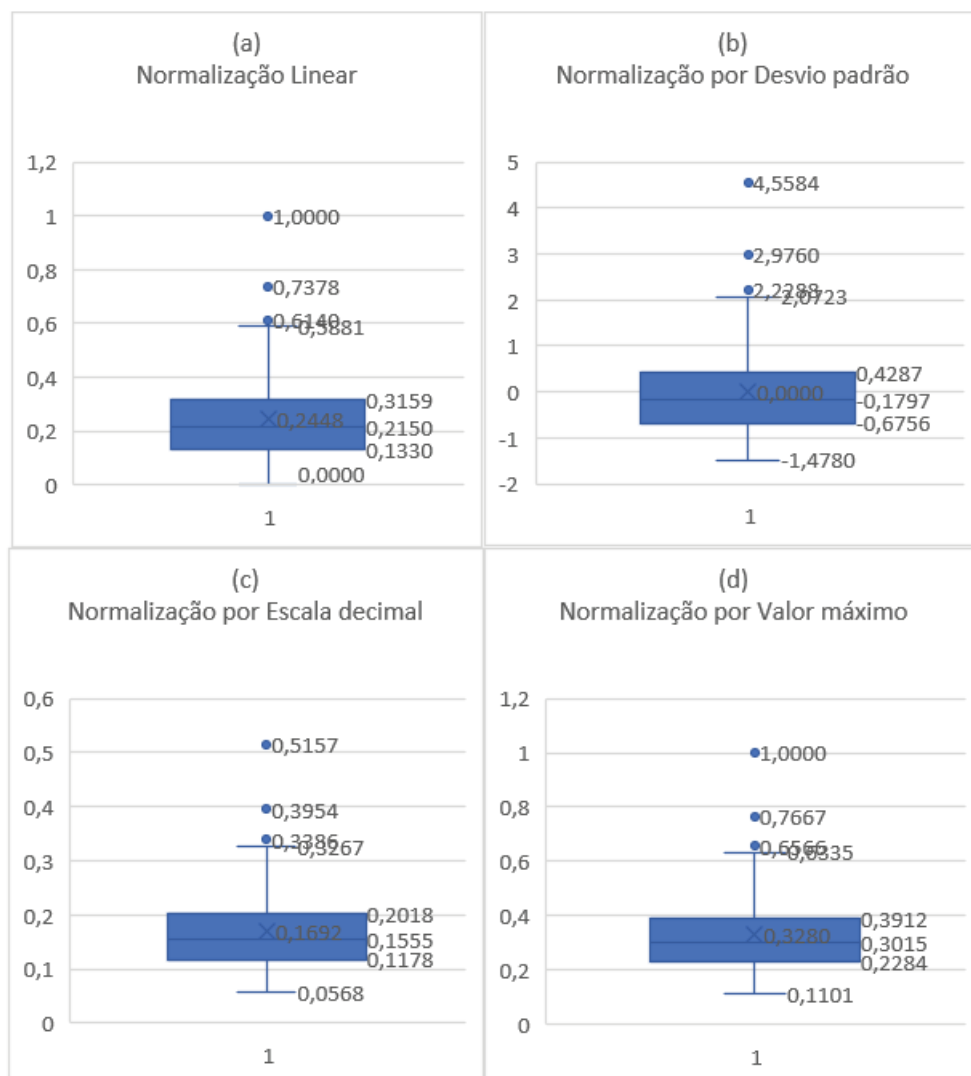
4.2.1 Análise do efeito da normalização sobre os valores não tratados (dados observados)

4.2.1.1 Normalização da vazão máxima anual (I_1)

A Figura 21 mostra o valor do indicador I_1 para cada evento, considerando-se os 4 (quatro) métodos de normalização analisados.

A Figura 22 mostra os principais parâmetros estatísticos: valor mínimo, primeiro quartil (25%), segundo quartil (50%), média, terceiro quartil (75%), limite para outlier superior, outlier superior. É possível constatar grande variação em todos os parâmetros estatísticos, principalmente no valor mínimo, que variou entre -1,47 (normalização por desvio padrão) e 0 (normalização linear), e no valor máximo, que variou entre 0,51 (normalização por escala decimal) e 4,55 (normalização por desvio padrão).

A planilha com os dados de I_1 normalizados através dos 4 métodos de normalização aplicados no presente trabalho se encontra no Apêndice 9.

FIGURA 21 – VALORES NORMALIZADOS DO INDICADOR I_1 (VAZÃO MÁXIMA ANUAL).FIGURA 22 – PARÂMETROS ESTATÍSTICOS DOS VALORES NORMALIZADOS DO INDICADOR I_1 (VAZÃO MÁXIMA ANUAL)

4.2.1.2 Normalização do nível de água máximo anual (I_2)

Os valores normalizados de I_2 são apresentados no Apêndice 10. A Figura 23 apresenta a estratigrafia das amostras do indicador I_2 , normalizadas através dos 4 métodos de normalização propostos no presente trabalho.

A Figura 24 mostra a variação dos parâmetros estatísticos do indicador I_2 . É possível constatar grande variação em todos os parâmetros, principalmente no valor mínimo, que variou entre -1,80 (normalização por desvio padrão) e 0 (normalização linear), e no valor máximo, que variou de 0,10 (normalização por escala decimal) a 3,80 (normalização por desvio padrão).

Os valores de I_2 que mais se aproximaram foram os produzidos pela normalização linear e pela normalização pelo valor máximo do elemento, porém, ainda assim existem diferenças significativas.

FIGURA 23 - VALORES NORMALIZADOS DO INDICADOR I_2 (NÍVEL MÁXIMO ANUAL).

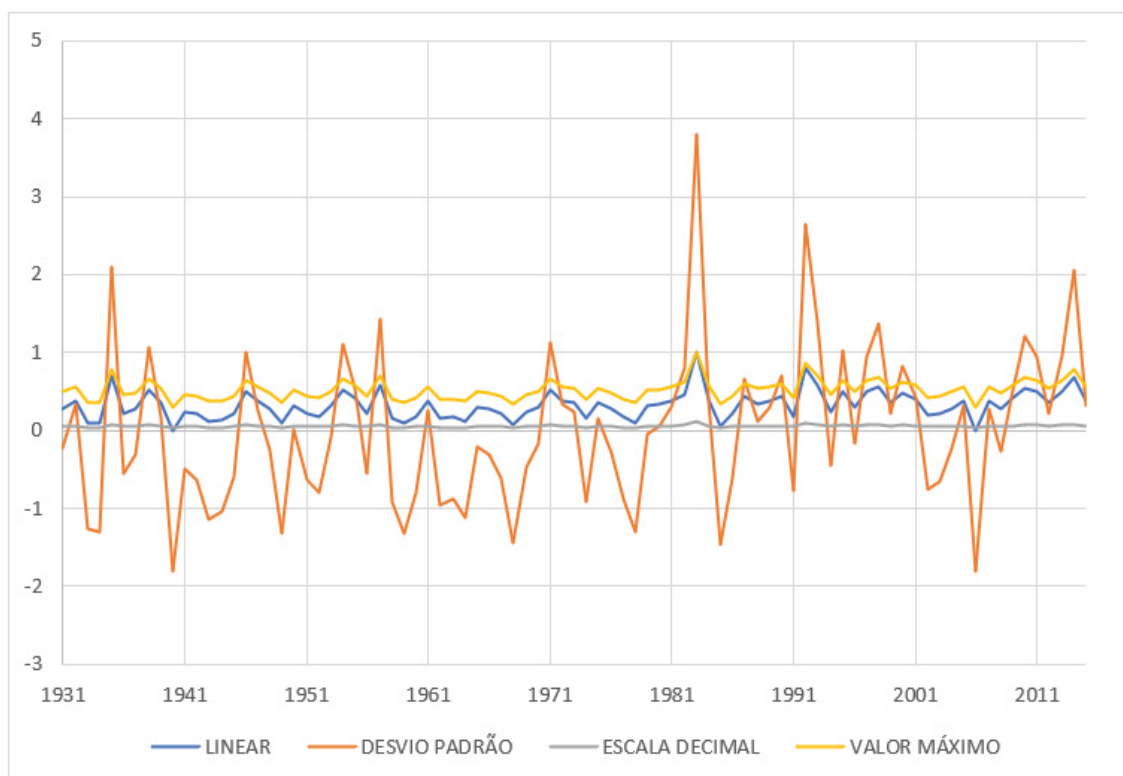
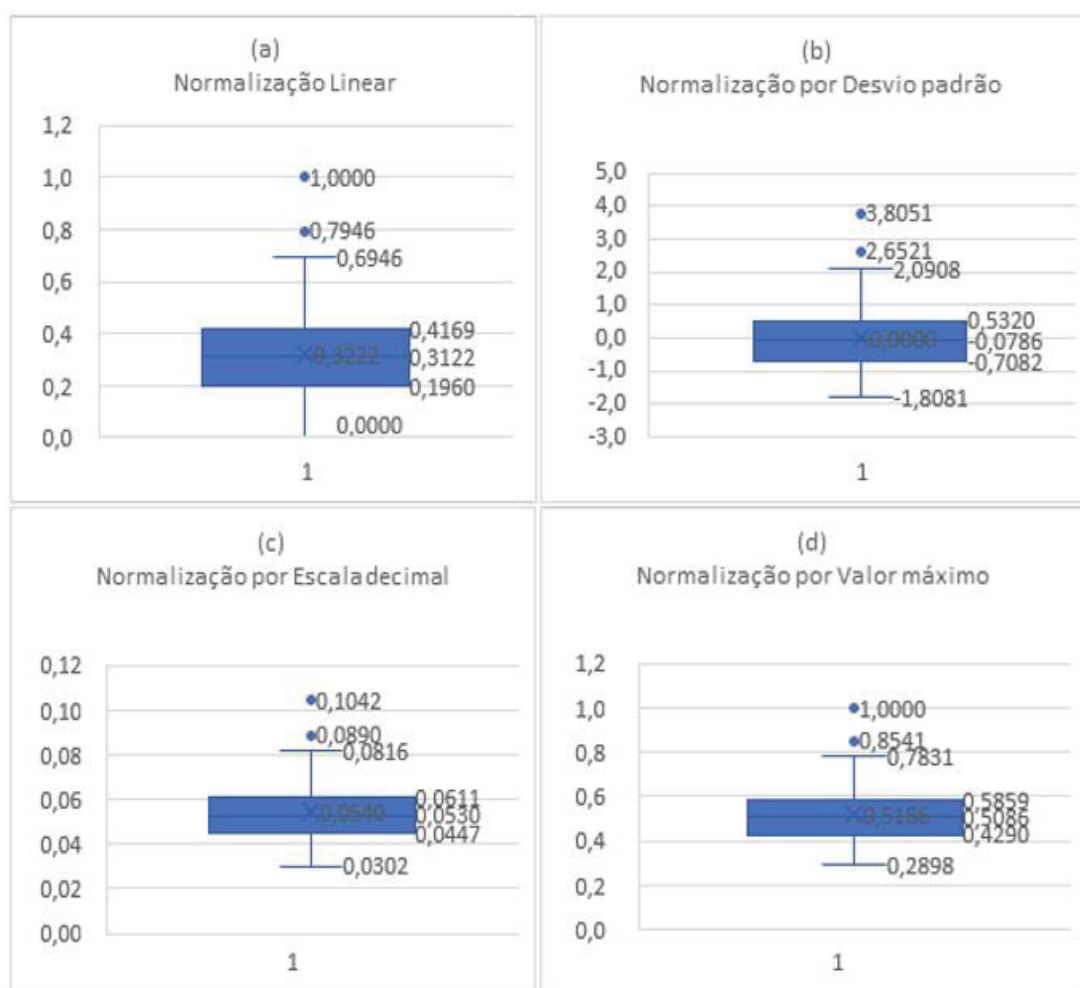


FIGURA 24 - PARÂMETROS ESTATÍSTICOS DOS VALORES NORMALIZADOS DO INDICADOR I_2 (NÍVEL MÁXIMO ANUAL)

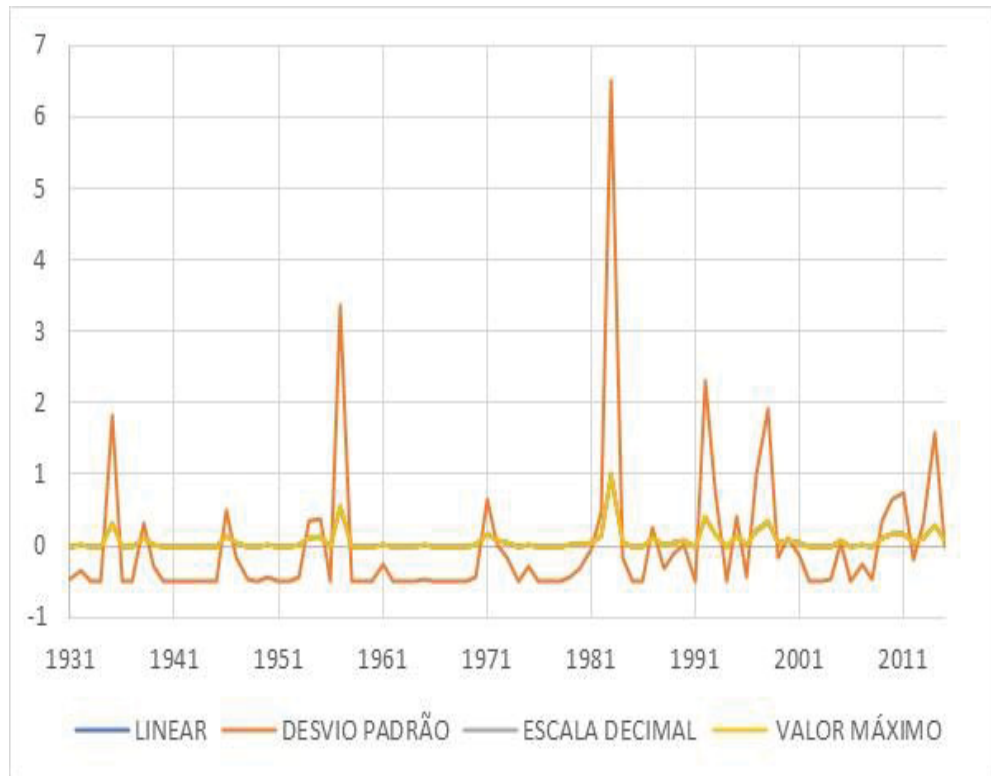
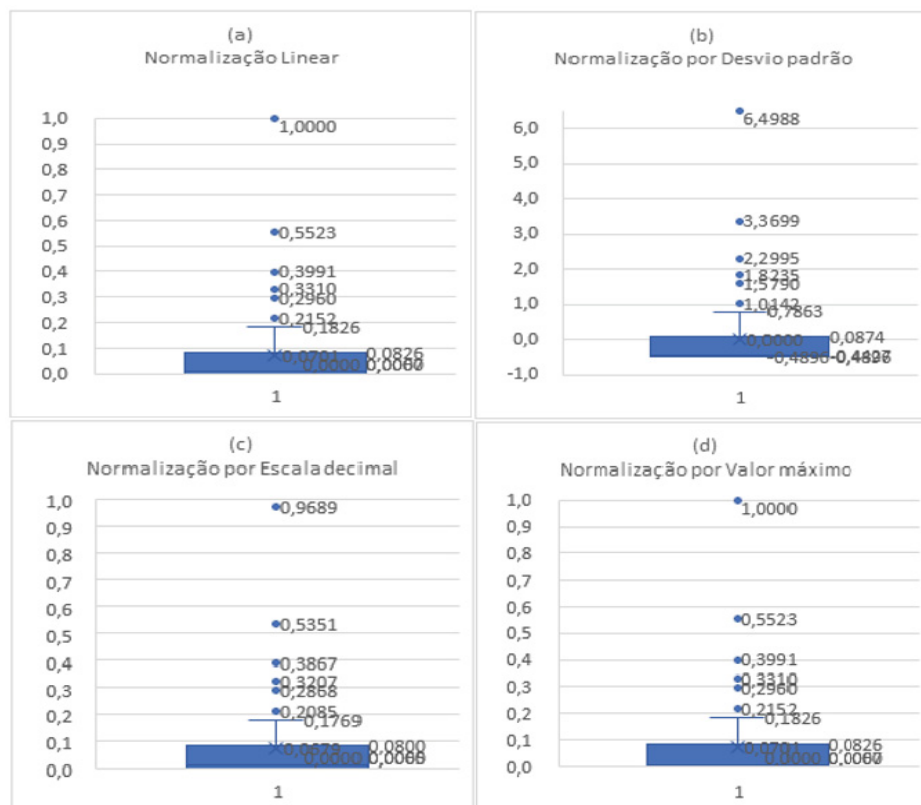


4.2.1.3 Normalização do volume máximo anual (I_3)

Os valores normalizados de I_3 são apresentados no Apêndice 11. A Figura 25 apresenta a amplitude das amostras do indicador I_3 , normalizadas através dos 4 métodos de normalização propostos no presente trabalho.

A Figura 26 mostra a variação dos parâmetros estatísticos do indicador I_3 . É possível constatar grande variação em todos os parâmetros, principalmente no valor mínimo, que variou entre -0,44 (normalização por desvio padrão) e 0 (normalização linear), e no valor máximo, que variou de 0,96 (normalização por escala decimal) a 6,49 (normalização por desvio padrão).

Cabe ressaltar que, dentre dos indicadores analisados, o indicador I_3 (volume máximo anual) foi o que mostrou maior similaridade entre os métodos de normalização aplicados.

FIGURA 25 - VALORES NORMALIZADOS DO INDICADOR I_3 (VOLUME MÁXIMO ANUAL).FIGURA 26 - PARÂMETROS ESTATÍSTICOS DOS VALORES NORMALIZADOS DO INDICADOR I_3 (VOLUME MÁXIMO ANUAL)

4.2.1.4 Normalização do tempo acima do limiar de extravasamento (I_4)

Os valores normalizados do indicador I_4 são apresentados no Apêndice 12. A Figura 27 apresenta a variabilidade observada nos valores normalizados do indicador I_4 (tempo acima do limiar de extravasamento).

A Figura 28 mostra os parâmetros estatísticos do indicador I_4 . É possível constatar grande variação em todos os parâmetros, principalmente no valor mínimo, que variou entre -0,77 (normalização por desvio padrão) e 0 (normalização linear), e no valor máximo, que variou de 0,98 (normalização por escala decimal) a 8,79 (normalização por desvio padrão).

FIGURA 27 - VALORES NORMALIZADOS DO INDICADOR I_4 (TEMPO ACIMA DO LIMAR DE EXTRAVASAMENTO).

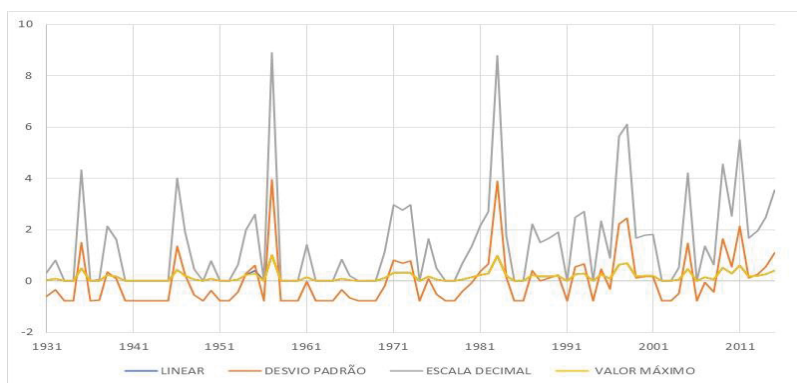
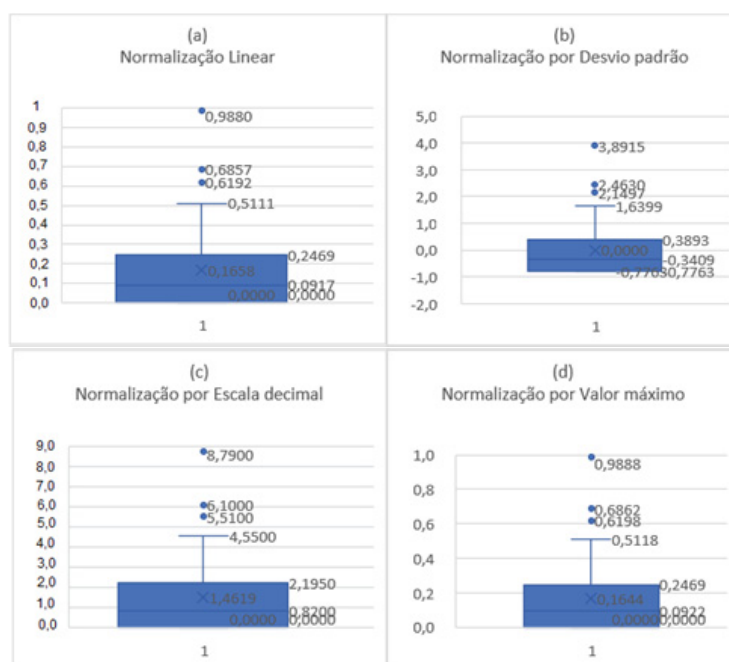


FIGURA 28 - PARÂMETROS ESTATÍSTICOS DOS VALORES NORMALIZADOS DO INDICADOR I_4 (TEMPO ACIMA DO LIMAR DE EXTRAVASAMENTO).



4.2.1.5 Normalização do tempo entre o início do evento e a ultrapassagem do limiar de extravasamento (I_5)

Os valores normalizados do indicador I_5 são apresentados no Apêndice 13. A análise dos valores normalizados do indicador I_5 mostra a diferença na variabilidade das amostras, ocasionada pelo tipo de normalização aplicado, como mostra a Figura 29.

A Figura 30 apresenta os parâmetros estatísticos do indicador I_5 (tempo entre o início do evento e a ultrapassagem do limiar de extravasamento), onde se verificam variações em todos os parâmetros estatísticos, com variações consideráveis nos limites inferiores, que oscilam entre -0,88 e 0 (normalização por desvio padrão); e os limites superiores, que variaram entre 1 e 5,70 (escala decimal).

FIGURA 29 - VALORES NORMALIZADOS DO INDICADOR I_5 (TEMPO ENTRE O INÍCIO DO EVENTO E A ULTRAPASSAGEM DO LIMAR DE EXTRAVASAMENTO).

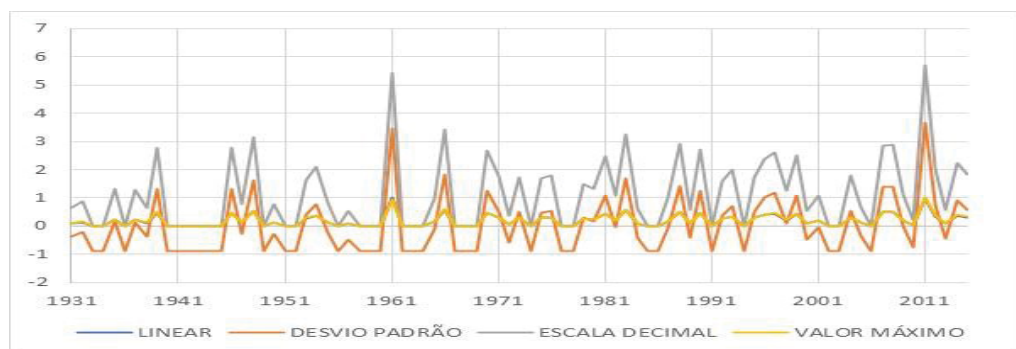
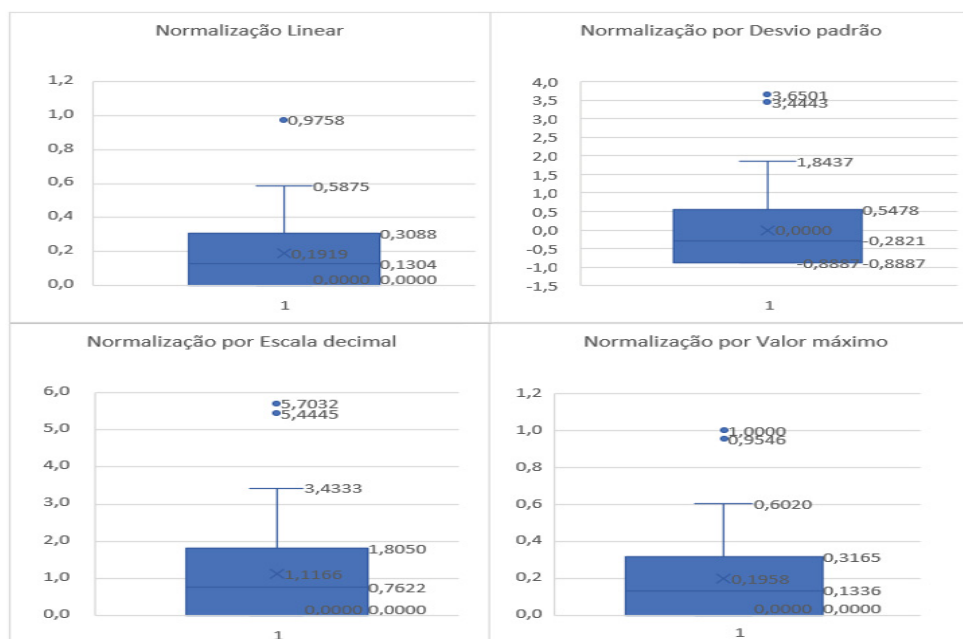


FIGURA 30 - PARÂMETROS ESTATÍSTICOS DOS VALORES NORMALIZADOS DO INDICADOR I_5 (TEMPO ENTRE O INÍCIO DO EVENTO E A ULTRAPASSAGEM DO LIMAR DE EXTRAVASAMENTO).



4.2.1.6 Normalização da taxa média de ascensão do hidrograma (I_6)

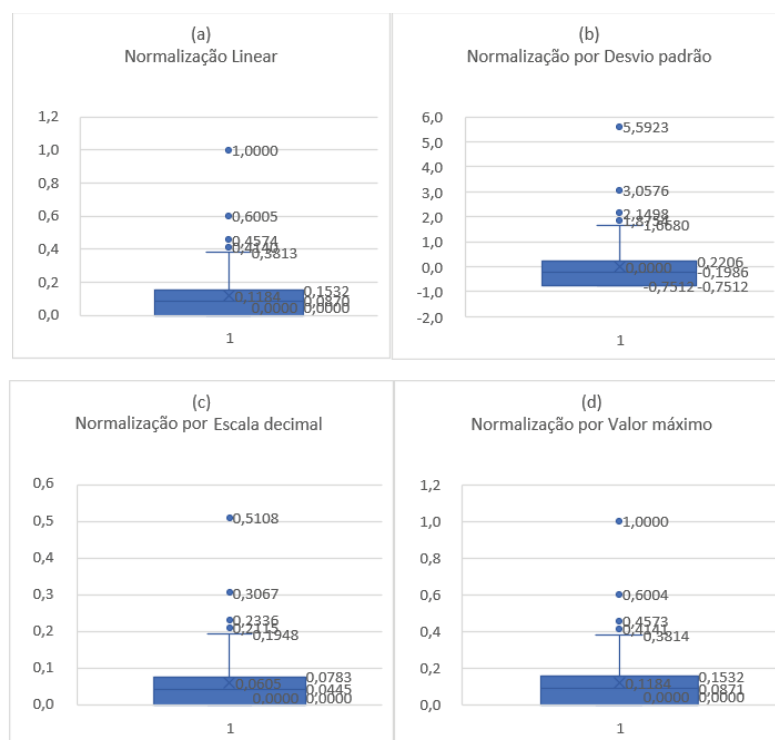
Os valores normalizados do indicador I_6 são apresentados no Apêndice 14. Apesar das variações observadas no conjunto de amostras, como mostra a Figura 31, quando são comparados os parâmetros estatísticos, apresentados na Figura 32, verifica-se similaridade entre as amostras produzidas pela normalização linear e pela normalização por valor máximo dos elementos.

O indicador I_6 (taxa média de ascensão do hidrograma) foi o que mostrou a segunda maior similaridade entre amostras produzidas por tipos distintos de normalização.

FIGURA 31 - VALORES NORMALIZADOS DO INDICADOR I_6 (TAXA MÉDIA DE ASCENSÃO DO HIDROGRAMA).



FIGURA 32 - PARÂMETROS ESTATÍSTICOS DOS VALORES NORMALIZADOS DO INDICADOR I_6 (TAXA MÉDIA DE ASCENSÃO DO HIDROGRAMA).



4.2.1.7 Normalização da taxa média de recessão do hidrograma (I_7)

Os valores normalizados do indicador I_7 são apresentados no Apêndice 15. A Figura 33 apresenta a variabilidade das amostras normalizadas do indicador I_7 em função do método de normalização utilizado.

Assim como no caso dos demais indicadores, observou-se diferenças na variabilidade do indicador I_7 , conforme mostra a Figura 34. Verifica-se diferenças significativas em todos os parâmetros estatísticos, como exemplo, o valor mínimo que variou entre -0,89 (normalização por desvio padrão) e 0 (nos demais métodos) e o valor máximo que variou entre 0,19 (normalização por escala decimal) e 3,64 (normalização por desvio padrão).

FIGURA 33 - VALORES NORMALIZADOS DO INDICADOR I_7 .

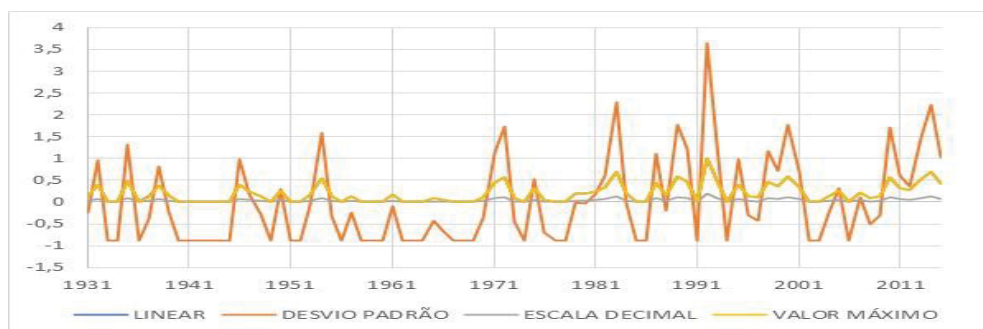
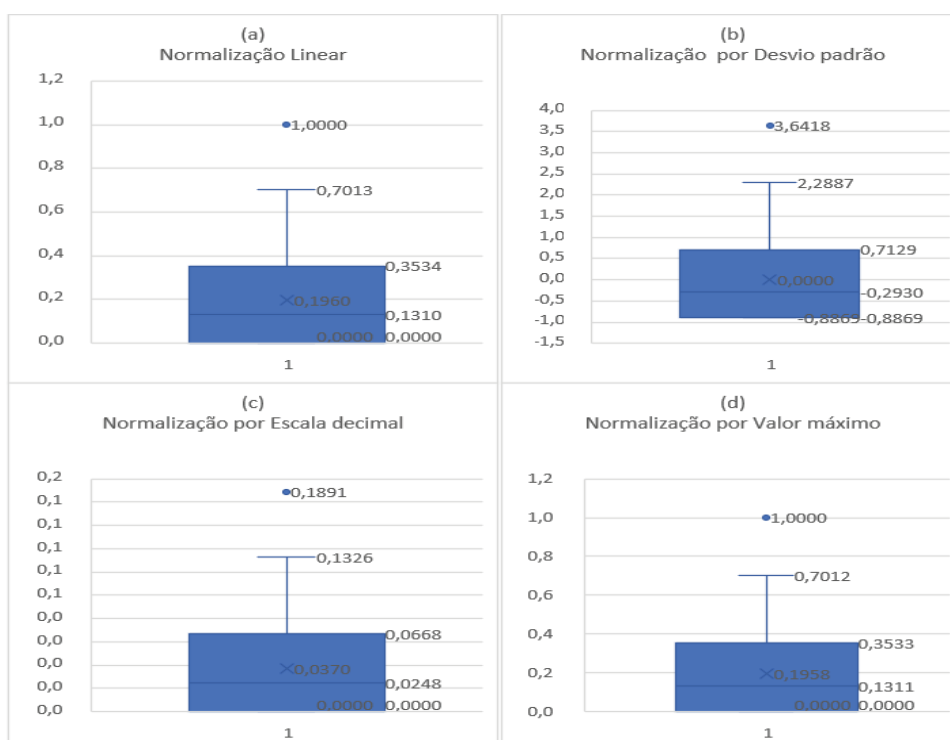


FIGURA 34 - PARÂMETROS ESTATÍSTICOS DOS VALORES NORMALIZADOS DO INDICADOR I_7 (TAXA MÉDIA DE RECESSÃO DO HIDROGRAMA).



4.2.2 Influência do método de normalização sobre o agrupamento das cheias pelo método *K-means*

O subitem 4.2.1 do presente capítulo, apresentou a análise da influência dos métodos de normalização sobre os parâmetros estatísticos das séries normalizadas dos indicadores utilizados no presente trabalho. Como exposto no subitem em questão, o método de normalização provocou diferenças significativas nos parâmetros estatísticos (valor mínimo, primeiro quartil, segundo quartil, média, mediana, terceiro quartil, valor máximo) nos conjuntos de amostras normalizadas e em comparação aos valores brutos.

No presente item, procurou-se verificar se o método de normalização impacta no resultado do agrupamento de cheias históricas observadas na estação fluviométrica União da Vitória (65310000), localizada no rio Iguaçu, na cidade de União da Vitória. Assumiu-se, como base de comparação, o agrupamento obtido utilizando o método *K-means* e o método de normalização linear, cujos resultados já foram apresentados.

4.2.2.1 Dados normalizados pelo método do desvio padrão e agrupados por *K-means*

Tendo os dados brutos sido normalizados pelo método do desvio padrão, foi executado o agrupamento das amostras utilizando-se o método *K-means*, e feita uma comparação com os dados originais normalizados pelo método de normalização linear.

A Tabela 11 apresenta os novos centroides calculados para as amostras normalizadas por desvio padrão. Uma vez que não se pode traçar um comparativo dos novos centroides (calculados por desvio padrão) e dos centroides calculados por normalização linear, pois, as amostras se encontram em escalas diferentes, analisou-se apenas as mudanças ocorridas nos agrupamentos, conforme apresentadas na Tabela 12.

Apenas a título explicativo, foi feita uma comparação visual entre as coordenadas de um evento específico e as coordenadas dos grupos de magnitude de cheia, definidos pelo método de normalização por desvio padrão. Para tanto, foi escolhido o vetor referente ao ano de 1997, que migrou de magnitude excepcional (grupo 4), quando normalizado por normalização linear, para magnitude alta (grupo 3), quando normalizado pelo método do desvio padrão.

TABELA 11– CENTROIDES DOS INDICADORES NORMALIZADOS PELO DESVIO PADRÃO

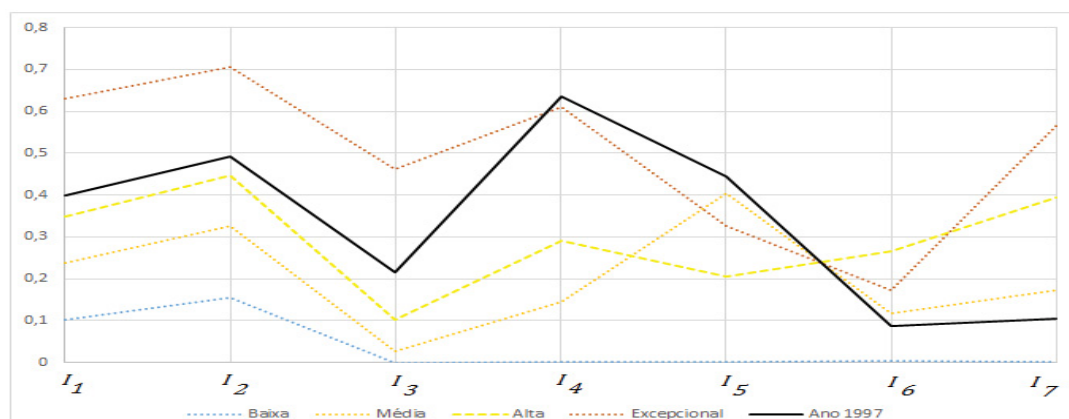
magnitude da cheia	I_1	I_2	I_3	I_4	I_5	I_6	I_7
Baixa	-0,8744	-0,9502	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
Média	-0,0543	0,0180	-0,3118	-0,0885	1,0345	-0,0168	-0,1145
Alta	0,6128	0,6913	0,2231	0,5822	0,0777	0,9209	0,8588
excepcional	2,4206	2,2336	2,9140	2,1505	0,4624	0,3869	1,7338

TABELA 12 – AMOSTRAS QUE TIVERAM ALTERAÇÕES NO AGRUPAMENTO: DADOS ORIGINAIS x DADOS NORMALIZADOS PELO MÉTODO DO DESVIO PADRÃO (MÉTODO K-MEANS)

Ano	Método de Normalização	
	Normalização linear	Desvio padrão
1938	Alta	Média
1946	excepcional	Alta
1950	Média	Alta
1987	Alta	Média
1989	Alta	Média
1997	excepcional	Alta
2011	Alta	Média

A Figura 35 apresenta as coordenadas do evento de 1997, representadas pelos indicadores I_1 , I_2 , I_3 , I_4 , I_5 , I_6 e I_7 , e as coordenadas dos centroides dos grupos de magnitude de cheia (baixa, média, alta e excepcional) definidos pelo método *K-means* e pelo método de normalização pelo desvio padrão.

FIGURA 35 – COMPARAÇÃO DAS COORDENADAS DO EVENTO 1997 E DOS CENTROIDES DO AGRUPAMENTO (MÉTODO K-MEANS E NORMALIZAÇÃO PELO MÉTODO DO DESVIO PADRÃO)



Na Figura 35 é possível verificar (visualmente), que o vetor do ano de 1997, em seus indicadores I_1 , I_2 , I_3 , estão mais próximos das coordenadas do centroide do grupo magnitude alta. O indicador I_4 está mais próximo do centroide de magnitude

excepcional, e os indicadores I_5 , I_6 e I_7 estão mais próximos do centroide de magnitude média. Para o ano em questão, a utilização do método de normalização por desvio padrão, e o cálculo das distancias através da distância de *Mahalanobis*, fez com que o vetor do ano (1997) se posicionasse mais próximo do centroide de magnitude alta.

Salienta-se que o exemplo em questão foi utilizado apenas para situar o leitor. As análises visuais nele praticadas não devem ser tomadas como regra, uma vez que a distância de *Mahalanobis* é calculada entre 2 vetores, não sendo possível sua interpretação num espaço de 7 dimensões.

4.2.2.2 Dados normalizados pelo método do valor máximo dos elementos e agrupados por *K-means*

Analisando-se os dados normalizados através do método do valor máximo dos elementos e agrupados pelo método *K-means*, comparando-os aos dados originais agrupados através da normalização linear e pelo método *K-means* é possível observar que houve alteração em 4 amostras (de um total de 85), conforme demonstrado na Tabela 13.

A Tabela 14 apresenta os novos centroides calculados por *K-means*, referentes aos dados normalizados pelo método do valor máximo do elemento.

TABELA 13 – AMOSTRAS QUE TIVERAM ALTERAÇÕES NO AGRUPAMENTO: DADOS ORIGINAIS x DADOS NORMALIZADOS PELO MÉTODO DO VALOR MÁXIMO (MÉTODO *K-MEANS*)

Ano	Método de Normalização	
	Normalização linear	Valor Máximo
1931	Média	Baixa
1992	Excepcional	Alta
1997	Alta	Excepcional
2011	Alta	Excepcional

TABELA 14 – CENTROIDES DOS INDICADORES NORMALIZADOS PELO VALOR MÁXIMO

magnitude da cheia	I_1	I_2	I_3	I_4	I_5	I_6	I_7
Baixa	0,2015	0,401	0,0000	0,001	0,003	0,0087	0,004
Média	0,3148	0,5168	0,0188	0,1264	0,3986	0,1196	0,1687
Alta	0,4452	0,6257	0,1215	0,2808	0,2161	0,2583	0,4405
excepcional	0,6022	0,7416	0,4365	0,7361	0,4281	0,1594	0,3697

4.2.2.3 Dados normalizados pelo método da escala decimal e agrupados por *K-means*

Analisando-se os grupos formados utilizando-se o método *K-means*, normalizados pelo método da escala decimal, e comparando-os com os dados normalizados pelo método de normalização linear, observou-se que houve alteração em apenas 2 amostras (de um total de 85), conforme demonstrado na Tabela 15. A Tabela 16 apresenta os novos centroides calculados através do método *K-means* para valores normalizados através do método da escala decimal.

TABELA 15 – AMOSTRAS QUE TIVERAM ALTERAÇÕES NO AGRUPAMENTO: DADOS ORIGINAIS x DADOS NORMALIZADOS PELO MÉTODO DA ESCALA DECIMAL (MÉTODO *K-MEANS*).

Ano	Método de Normalização	
	Normalização linear	Escala decimal
1932	Alta	Média
1997	Alta	Excepcional

TABELA 16 – CENTROIDES, INDICADORES NORMALIZADOS POR ESCALA DECIMAL

magnitude da cheia	I_1	I_2	I_3	I_4	I_5	I_6	I_7
Baixa	0,2235	0,3891	0,0000	0,0010	0,0028	0,0084	0,0004
Média	0,3144	0,5132	0,0185	0,1222	0,3333	0,2413	0,4350
Alta	0,4298	0,6009	0,1197	0,2746	0,2205	0,2413	0,4350
Excepcional	0,6011	0,7312	0,4255	0,7213	0,4280	0,1493	0,3581

O programa computacional implementado em linguagem de programação *Python*, para o agrupamento das 85 amostras (anos de 1931 a 2015), normalizadas através do método da escala decimal, é apresentado no Apêndice 24.

4.3 INFLUÊNCIA DA ALTERAÇÃO DOS INDICADORES I_5 E I_7 SOBRE O AGRUPAMENTO DE CHEIAS

O subitem 3.3.8 discorreu sobre os motivos relacionados à alteração dos indicadores I_5 e I_7 e o procedimento utilizado para a referida alteração. Os valores normalizados dos indicadores I_5 e I_7 alterados para os diferentes métodos de normalização se encontram nos Apêndices 16 e 17. No entanto, no presente item, são analisados apenas os resultados relativamente à normalização linear.

Os resultados até aqui apresentados mostraram que a mudança do método *Fuzzy C-means* para o método *K-means* provocou mudanças no agrupamento das cheias. Da mesma forma, verificou-se que os métodos de normalização empregados também provocaram mudanças no agrupamento das cheias, mesmo utilizando-se o mesmo método de agrupamento (método *K-means*).

São apresentadas comparações entre os agrupamentos feitos para os valores de I_5 e I_7 originais e para os valores de I_5 e I_7 alterados, considerando-se o método de normalização linear. As Figuras 36 e 37 apresentam respectivamente uma comparação da variabilidade dos valores dos indicadores I_5 e I_7 originais e alterados.

FIGURA 36 - VARIAÇÃO DOS VALORES NORMALIZADOS PARA O INDICADOR I_5 ORIGINAL E PARA O INDICADOR I_5 ALTERADO

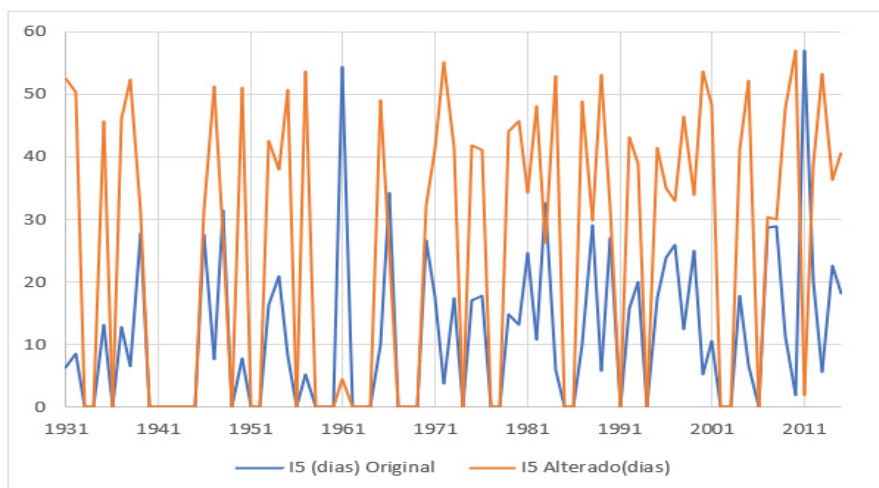
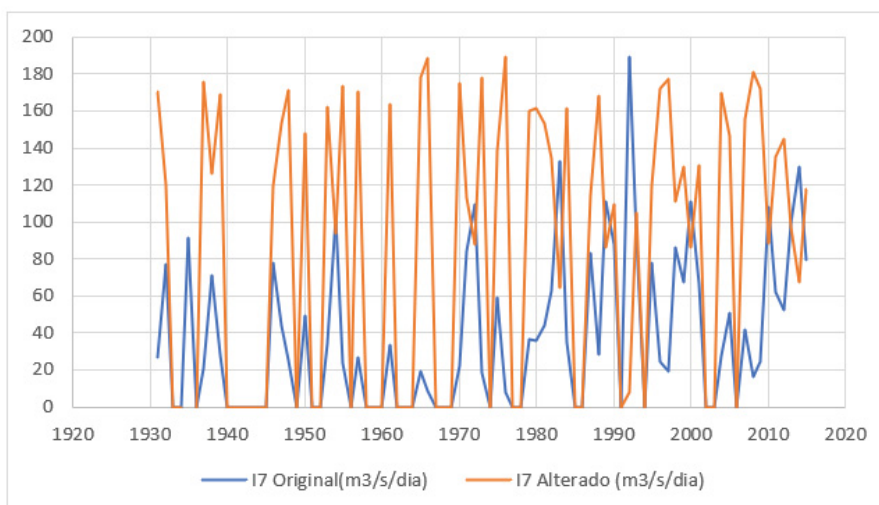


FIGURA 37 - VARIAÇÃO DOS VALORES NORMALIZADOS PARA O INDICADOR I_7 ORIGINAL E PARA O INDICADOR I_7 ALTERADO



A Tabela 17 apresenta os novos centroides calculados para os indicadores I_5 e I_7 alterados. Em análise à tabela em questão, conforme esperado, se observa uma alteração nos centros de gravidade dos referidos indicadores (I_5 e I_7).

A Tabela 18 apresenta as alterações ocasionadas nos agrupamento pela alteração dos indicadores I_5 e I_7 para o método *K-means* e a normalização linear.

TABELA 17 – CENTROIDES DOS GRUPOS PARA OS VALORES DE I_5 E I_7 ALTERADOS, NORMALIZADOS POR NORMALIZAÇÃO LINEAR

Magnitude da cheia	I_1	I_2	I_3	I_4	I_5	I_6	I_7
Baixa	0,0999	0,1528	0,0000	0,0000	0,7356	0,0000	0,8522
Média	0,2547	0,3460	0,0407	0,1850	0,8602	0,1062	0,6602
Alta	0,3349	0,4330	0,0916	0,2611	0,5913	0,1794	0,3709
Excepcional	0,6458	0,7200	0,4870	0,6196	0,0000	0,2878	0,0000

TABELA 18 - AMOSTRAS QUE TIVERAM ALTERAÇÕES NO AGRUPAMENTO. DADOS ORIGINAIS x INDICADORES I_5 E I_7 ALTERADOS (*K-MEANS* E NORMALIZAÇÃO LINEAR).

Ano	<i>K-means</i> e Normalização Linear	
	I_5 e I_7 originais	I_5 e I_7 modificados
1931	Média	Alta
1946	Alta	Média
1950	Média	Alta
1979	Média	Baixa
1982	Alta	Média
1990	Alta	Média
1997	Alta	Média
2011	Alta	Média

O programa computacional implementado em linguagem de programação *Python*, com o agrupamento das 85 amostras (anos de 1931 a 2015), normalizadas através do método de normalização linear, com os indicadores I_5 e I_7 alterados, é apresentado no Apêndice 25.

4.4 REDUÇÃO DO NÚMERO DE INDICADORES DE CHEIA PELA ANÁLISE DE COMPONENTES PRINCIPAIS (ACP)

Como já explanado anteriormente, foi aplicada a técnica estatística de Análise de Componentes Principais (ACP), com o intuito de diminuir o número de indicadores de cheia e, por consequência, a dimensionalidade do problema de classificação. A técnica consistiu basicamente em analisar os autovalores e autovetores, descartando os indicadores menos representativos. Via de regra, são buscados valores de autovetores que expliquem, no mínimo, 80% da amostra.

Apenas a título explicativo, nas Tabelas de 19 a 23 são apresentadas as avaliações referentes à redução no número de indicadores, usando a Análise dos Componentes Principais (ACP), considerando-se os dados brutos (Tabela 19) e os dados normalizados por normalização linear (Tabela 20), desvio padrão (Tabela 21), valor máximo dos elementos (Tabela 22) e escala decimal (Tabela 23). Os valores destacados em fundo amarelo nas Tabelas 19 a 23 representam qual o indicador mais representativo de cada componente principal.

TABELA 19 - AUTOVALORES E AUTOVETORES. DADOS BRUTOS.

		65,78%	20,08%	7,9%	4,0%	1,1%	1,0%	0,04%	$\Sigma = 100\%$
AUTOVALORES		4,5981	1,4039	0,5589	0,2843	0,0808	0,0710	0,0029	$\Sigma = 6,99$
AUTOVETORES	I_1 (m ³ /s)	0,4325	-0,2413	-0,0006	0,4371	-0,1647	-0,0659	0,7296	
	I_2 (cm)	0,4413	-0,1579	0,0326	0,4438	-0,3769	0,0160	-0,6633	
	I_3 (Km ³)	0,3629	-0,4902	-0,0747	-0,1825	0,6603	-0,3608	-0,1515	
	I_4 (dias)	0,4065	-0,2268	0,1217	-0,6989	-0,3359	0,4039	0,0635	
	I_5 (dias)	0,3775	0,4539	0,0024	0,1772	0,5059	0,6033	-0,0111	
	I_6 (m ³ /s/dia)	0,2992	0,4195	-0,7585	-0,1953	-0,1586	-0,3092	0,0140	
	I_7 (m ³ /s/dia)	0,2987	0,4930	0,6349	-0,1439	-0,0302	-0,4924	0,0215	

TABELA 20 - AUTOVALORES E AUTOVETORES. NORMALIZAÇÃO LINEAR

		65,37%	20,31%	8,09%	4,01%	1,13%	1,10%	0,04%	$\Sigma = 100\%$
AUTOVALORES		4,5704	1,4280	0,5621	0,2803	0,0790	0,0773	0,0028	$\Sigma = 6,99$
AUTOVETORES	I_1 (m ³ /s)	0,4344	-0,2400	0,0011	0,4261	-0,1887	0,0569	0,7303	
	I_2 (cm)	0,4427	-0,1608	0,0289	0,4399	-0,2616	0,2776	-0,6620	
	I_3 (Km ³)	0,3680	-0,4765	-0,0519	-0,1987	0,1881	-0,7324	-0,1538	
	I_4 (dias)	0,4097	-0,2171	0,1372	-0,6874	0,0832	0,5314	0,0659	
	I_5 (dias)	0,3788	0,4452	-0,0135	0,2120	0,7817	0,0458	-0,0042	
	I_6 (m ³ /s/dia)	0,2986	0,4182	-0,7545	-0,2184	-0,3310	-0,0966	0,0104	
	I_7 (m ³ /s/dia)	0,2818	0,5189	0,6389	-0,1425	-0,3648	-0,2993	0,0142	

As análises realizadas para os quatro métodos normalização utilizados neste trabalho mostraram que os indicadores mais representativos foram o I_2 (nível de água

máximo) e I_7 (taxa média de recessão do hidrograma), que explicaram, em todas as análises, mais de 85% das amostras.

A sequência completa de cálculos, executada através do *Software Microsoft Excel*, e seu suplemento estatística *Real Statistics*, é apresentada nos Apêndices 38 a 41.

TABELA 21 - AUTOVALORES E AUTOVETORES. NORMALIZAÇÃO POR DESVIO PADRÃO

AUTOVALORES		65,37%	20,31%	8,09%	4,01%	1,13%	1,10%	0,04%	$\Sigma = 100\%$
		4,5704	1,4283	0,5622	0,2808	0,0792	0,0775	0,0029	$\Sigma = 6,99$
AUTOVETORES	I_1 (m ³ /s)	0,4347	-0,2402	0,0012	0,4263	-0,1888	0,0574	0,7304	
	I_2 (cm)	0,4428	-0,1609	0,0292	0,4399	-0,2618	0,2778	-0,6621	
	I_3 (Km ³)	0,3682	-0,4767	-0,0523	-0,1988	0,1882	-0,7325	-0,1539	
	I_4 (dias)	0,4098	-0,2172	0,1374	-0,6875	0,0833	0,5317	0,0662	
	I_5 (dias)	0,3787	0,4453	-0,0137	0,2122	0,7818	0,0459	-0,0043	
	I_6 (m ³ /s/dia)	0,2987	0,4183	-0,7549	-0,2187	-0,3312	-0,0967	0,0105	
	I_7 (m ³ /s/dia)	0,2819	0,5187	0,6392	-0,1426	-0,3655	-0,2998	0,0148	

TABELA 22 - AUTOVALORES E AUTOVETORES. NORMALIZAÇÃO POR VALOR MÁXIMO DOS ELEMENTOS

AUTOVALORES		65,36%	20,31%	8,08%	4,01%	1,13%	1,10%	0,04%	$\Sigma = 100\%$
		4,5812	1,4282	0,5702	0,2812	0,0795	0,0778	0,0029	$\Sigma = 6,99$
AUTOVETORES	I_1 (m ³ /s)	0,4345	-0,2403	0,0013	0,4266	-0,1889	0,0570	0,7312	
	I_2 (cm)	0,4431	-0,1701	0,0295	0,4400	-0,2618	0,2778	-0,6625	
	I_3 (Km ³)	0,3700	-0,4766	-0,0520	-0,1991	0,1885	-0,7328	-0,1539	
	I_4 (dias)	0,4110	-0,2172	0,1374	-0,6876	0,0836	0,5316	0,0660	
	I_5 (dias)	0,3788	0,4458	-0,0137	0,2128	0,7819	0,0459	-0,0045	
	I_6 (m ³ /s/dia)	0,2986	0,4185	-0,7548	-0,2194	-0,3315	-0,0968	0,0109	
	I_7 (m ³ /s/dia)	0,2818	0,5189	0,6389	-0,1426	-0,3702	-0,2994	0,0145	

TABELA 23 - AUTOVALORES E AUTOVETORES. NORMALIZAÇÃO POR ESCALA DECIMAL

AUTOVALORES →		65,35%	20,31%	8,07%	4,00%	1,13%	1,10%	0,04%	$\Sigma = 100\%$
		4,5704	1,4280	0,5621	0,2803	0,0790	0,0773	0,0028	$\Sigma = 6,99$
AUTOVETORES →	I_1 (m ³ /s)	0,4344	-0,2400	0,0011	0,4261	-0,1887	0,0569	0,7303	
	I_2 (cm)	0,4427	-0,1608	0,0289	0,4399	-0,2616	0,2776	-0,6620	
	I_3 (Km ³)	0,3680	-0,4765	-0,0519	-0,1987	0,1881	-0,7324	-0,1538	
	I_4 (dias)	0,4097	-0,2171	0,1372	-0,6874	0,0832	0,5314	0,0659	
	I_5 (dias)	0,3788	0,4452	-0,0135	0,2120	0,7817	0,0458	-0,0042	
	I_6 (m ³ /s/dia)	0,2986	0,4182	-0,7545	-0,2184	-0,3310	-0,0966	0,0104	
	I_7 (m ³ /s/dia)	0,2818	0,5189	0,6389	-0,1425	-0,3648	-0,2993	0,0142	

4.4.1 Agrupamento de cheias pelo método *K-means* com redução do número de indicadores de cheias

Neste item apresenta-se o resultado do agrupamento das cheias, utilizando-se apenas dos indicadores I_2 (volume total extravasado) e I_7 (taxa média de recessão do hidrograma de cheias) como resultado da Análise dos Componentes Principais (ACP).

A aplicação da Análise de Componentes Principais (ACP) demonstrou a abrangência dos indicadores de cheia I_2 e I_7 , anteriormente descritos, como sendo os mais representativos das amostras, tanto na utilização dos dados brutos, quanto na aplicação dos quatro métodos de normalizações investigados no presente trabalho.

O objetivo deste item foi avaliar possíveis mudanças no agrupamento das cheias, ocasionadas pelo uso de apenas dois indicadores (I_2 e I_7), tendo em vista que, de acordo com o explanado no item anterior, conjuntamente tais indicadores explicam mais de 85% da variabilidade das amostras.

4.4.2 Comparação dos resultados dos agrupamentos de cheias pelo método *K-means* com e sem redução do número de indicadores de cheia

Assim como o observado anteriormente, houve modificações nos agrupamentos ocasionadas pelo tipo de normalização utilizada. O uso de apenas dois indicadores para a classificação resultou em menores discrepâncias nos agrupamentos. Contudo, verificou-se variações na distribuição estatística de todas as amostras; sendo percebida maior dispersão, quando da utilização do método de normalização por desvio padrão, e menor dispersão quando foi a normalização pelo método da escala decimal.

4.4.2.1 Método *K-means*, normalização linear e indicador I_7 original

Foram produzidos agrupamentos pelo método *K-means* e normalização linear, usando somente os indicadores I_2 e I_7 , considerando o indicador I_7 original, como definido em Steffen (2017).

O agrupamento obtido foi então comparado ao agrupamento obtido também pelo método *K-means* e normalização linear, porém usando todos os indicadores I_1 a I_7 , conforme apresentado no item 4.1.2 do presente trabalho.

O agrupamento obtido com os indicadores I_2 (volume total extravasado) e I_7 (taxa média de recessão do hidrograma) foram exatamente os mesmos formados quando se utilizou o método *K-means*, normalização linear e todos os indicadores I_1 a I_7 . Tal resultado ratifica as análises executadas no capítulo 4.4, que estabeleceu os indicadores I_2 e I_7 como os mais representativos.

O programa computacional implementado em linguagem de programação *Python*, com o agrupamento para as 85 amostras (anos de 1931 a 2015), normalizados através do método de normalização linear e com os indicadores I_2 e I_7 original, é apresentado no Apêndice 29.

4.4.2.2 Método *K-means*, normalização linear e indicador I_7 alterado

O agrupamento obtido usando somente os indicadores I_2 e I_7 alterado foi comparado ao agrupamento obtido utilizando todos os indicadores os indicadores I_1 a I_7 , considerando os indicadores I_5 e I_7 alterados, conforme discutido no item 4.3.

Comparou-se os resultados obtidos pelo método *K-means* e normalização linear, visando verificar a acurácia da redução dos indicadores de cheia. Assim como o ocorrido no item anterior, o agrupamento resultou o mesmo. Tal resultado evidencia a importância da aplicação da ACP realizada no presente trabalho.

O programa computacional implementado em linguagem de programação *Python*, com o agrupamento para as 85 amostras (anos de 1931 a 2015), normalizadas através do método de normalização linear e com os indicadores I_2 e I_7 alterado, é apresentado no Apêndice 30.

4.5 COMPARATIVO DE TODAS AS ANÁLISES EXECUTADAS

A síntese dos resultados obtidos é por meio da Tabela 24 e da Figura 38, as quais apresentam as mudanças ocorridas no agrupamento das 85 amostras (1931 a 2015), classificadas através de todos os métodos de normalização e variações de indicadores de cheia propostos no presente trabalho.

Conforme já descrito anteriormente, em análise ao capítulo 4 do trabalho acadêmico em tela, é possível identificar as alterações ocasionadas pelo tipo de algoritmo utilizado (*K-means* no lugar de *Fuzzy C-means*) assim como em função dos métodos de normalização empregados. O algoritmo de agrupamento apresentou influência na classificação final dos dados, isto é: houve alterações nos grupos formados; sendo que na maioria das vezes o vetor formado pelos sete indicadores de cheia de cada ano deslocou-se, via de regra, um grau na escala de magnitude da cheia, positiva e negativamente. Os resultados foram demonstrados no capítulo 4, item 4.1.3, onde se constatou alteração na escala de magnitude dos anos de 1931, 1950 e 2011; observa-se que na modelagem em questão os dados foram normalizados através da técnica de normalização linear, mesma técnica utilizada por Steffen (2017), havendo apenas mudança no algoritmo de agrupamento (de *Fuzzy C-means* para *K-means*).

Em etapa posterior, visando observar o efeito provocado pelo método empregado para tratamento inicial dos dados, foi implementado o algoritmo *K-means*, contudo, na etapa em questão, tendo a base de dados sido normalizados através de quatro métodos de normalização distintos, sendo eles: normalização por desvio padrão; normalização por escala decimal e normalização por valor máximo dos elementos.

O capítulo 4.2.1 fez uma análise detalhada da distribuição estatística dos valores brutos dos indicadores de cheia (I_1 a I_7), e, de modo incontestável, evidenciou alterações significativas, tanto nos valores brutos dos dados, quanto nos valores estratificados, sendo eles: valor mínimo, primeiro quartil (25%), segundo quartil (50%), média, terceiro quartil (75%), limite para outlier superior e outlier superior. Etapa posterior executou modelagem das amostras normalizadas através dos 4 métodos de normalização empregados, tendo chegado aos seguintes resultados:

- a. Normalização por desvio padrão; alteração nos anos de 1931, 1946, 1950, 1987, 1989, 1997 e 2011.

- b. Normalização por valor máximo do elemento; alteração nos anos de 1931, 1992, 1997 e 2011.
- c. Normalização pelo método da escala decimal; alteração nos anos de 1932 e 1997.

Analisando o produto das modelagens, observa-se que, tanto na questão da troca do algoritmo (de *Fuzzy C-means* para *K-means*), quanto na aplicação de outras técnicas de normalização, houve repetição de padrões nas alterações, ou seja, os vetores formados pelos indicadores de cheia de I_1 a I_7 , os quais sofreram alterações na escala de magnitude, tiveram bastante similaridade. Tal fato se deve à dissemelhança estatística das amostras, a qual variou em virtude do método de normalização utilizado. Como exemplo, gráficos de quartis são mostrados no capítulo 4, itens 4.2.1: os referidos gráficos (boxplots) mostram a retração estatística das amostras, ocorrida quando foi utilizada normalização por escala decimal, e a expansão estatística das amostras, percebida quando da utilização do método de normalização por desvio padrão.

Estágio terciário do trabalho versou sobre a sugestão de uma alteração nos indicadores I_5 (tempo entre o início do evento e a ultrapassagem do limiar de extravasamento) e I_7 (taxas médias de recessão do hidrograma de cheia). Tendo em vista se tratar de indicadores baseados em tempo, entendeu-se por ser indicada uma inversão no critério de análise destes indicadores, considerando os valores mínimos como sendo mais críticos para as amostras, uma vez que resultariam em eventos mais significativos no quesito magnitude. A metodologia de alteração foi explanada no capítulo 4.3. As modelagens aplicadas aos indicadores I_5 e I_7 alterados, suscitaram mudanças nos graus de magnitude dos vetores, conforme demonstrado, porém, qualquer tentativa de comparação dos agrupamentos produzidos, com outros previamente calculados, é improdutiva, por se tratar de amostras completamente distintas. O intuito das avaliações, neste caso, é observar o efeito causado pelo método de normalização nos agrupamentos de produzidos.

Alteração nos métodos de normalização, aplicado aos indicadores I_5 e I_7 alterados, suscitou modificações na escala de magnitude de vetores de magnitude de cheia, contudo, como já explicado, mesmo não sendo o contexto da avaliação, trouxe novamente vetores já modificados em outros cenários. Tal fato pode surpreender um leitor atento; contudo, cabe esclarecer que a singularidade na maneira de interpretar os indicadores, causa igualmente uma modificação na apresentação e interpretação

dos centroides. Haja vista que nesta nova abordagem o evento mais severo dos indicadores I_5 e I_7 , será aquele com menor valor, uma vez que representará elevações mais bruscas do hidrograma de cheias.

A última verificação processada no trabalho consistiu em executar uma análise estatística de componentes principais (ACP), a qual visou verificar os componentes mais representativos da amostra, e posteriormente averiguar qual o resultado provocado pelo agrupamento apenas destes componentes. Como exposto no capítulo 4, subitem 4.4, os indicadores I_2 (máximo nível de água) e I_7 (taxas médias de recessão do hidrograma de cheia) representam 85,86% das amostras. Por conseguinte, foi implementado agrupamento apenas nas amostras constituídas por esses dois indicadores; os resultados obtiveram correlação de 100% (amostra x magnitude), quando se utilizou a técnica de normalização linear e agrupamento executado através do algoritmo *K-means*. Posteriormente, foram executados agrupamentos através do algoritmo *K-means*, na base de dados com os valores de I_2 e I_7 alterados, normalizados através das quatro técnicas propostas no presente trabalho (normalização linear, normalização por desvio padrão, normalização por valor máximo dos elementos e normalização por escala decimal). Efetuar comparações entre os agrupamentos produzidos por Steffen (2017) e os executados com os valores dos indicadores I_2 e I_7 alterados seria improdutivo, tendo em vista que as alterações, naturalmente, inviabilizam a comparação. Por conseguinte, foram comparados os grupos de I_2 e I_7 alterados, formados a partir das técnicas de normalização citadas. Constatou-se significativa diminuição na alteração dos grupos, tendo as modificações existentes se manifestado nos anos de 1931, 1982 e 2011.

Em suma, as análises executadas no presente trabalho indicaram que o algoritmo de agrupamento utilizado interfere na formação dos grupos, e mesmo cenário se observa em referência às técnicas de normalização utilizadas. Houve 100% de compatibilidade quando foram utilizados apenas os indicadores que representam 85,86% das amostras (I_2 e I_7), e normalizados através do método da normalização linear, o qual foi o mesmo método utilizado por Steffen (2017).

TABELA 24 - AGRUPAMENTO DOS DADOS DE TODAS AS AMOSTRAS, ATRAVÉS DE TODOS OS MÉTODOS DE NORMALIZAÇÃO E COMPARAÇÕES ANALISADAS NO TRABALHO.

Agrupamento (ver legenda)	A	B	C	D	E	F	G	H	I	J
ANO	GRUPO									
1931	1	2	2	2	1	3	2	2	3	1
1932	3	3	3	2	3	3	2	2	3	3
1933	1	1	1	1	1	1	1	1	1	1
1934	1	1	1	1	1	1	1	1	1	1
1935	4	4	4	4	4	4	3	4	4	4
1936	1	1	1	1	1	1	1	1	1	1
1937	2	2	2	2	2	2	2	2	2	2
1938	3	3	2	3	3	3	4	2	3	3
1939	2	2	2	2	2	2	2	2	2	2
1940	1	1	1	1	1	1	1	1	1	1
1941	1	1	1	1	1	1	1	1	1	1
1942	1	1	1	1	1	1	1	1	1	1
1943	1	1	1	1	1	1	1	1	1	1
1944	1	1	1	1	1	1	1	1	1	1
1945	1	1	1	1	1	1	1	1	1	1
1946	3	3	3	3	3	2	3	3	3	3
1947	3	3	3	3	3	3	3	3	3	3
1948	2	2	2	2	2	2	2	2	1	2
1949	1	1	1	1	1	1	1	1	1	1
1950	3	2	3	2	2	3	2	2	3	3
1951	1	1	1	1	1	1	1	1	1	1
1952	1	1	1	1	1	1	1	1	1	1
1953	2	2	2	2	2	2	2	2	2	2
1954	3	3	3	3	3	3	4	3	3	3
1955	3	3	3	3	3	3	4	3	3	3
1956	1	1	1	1	1	1	1	1	1	1
1957	4	4	4	4	4	3	3	4	4	4
1958	1	1	1	1	1	1	1	1	1	1
1959	1	1	1	1	1	1	1	1	1	1
1960	1	1	1	1	1	1	1	1	1	1
1961	2	2	2	2	2	2	2	1	2	2
1962	1	1	1	1	1	1	1	1	1	1
1963	1	1	1	1	1	1	1	1	1	1
1964	1	1	1	1	1	1	1	1	1	1
1965	2	2	2	2	2	2	2	2	2	2
1966	2	2	2	2	2	2	2	2	2	2
1967	1	1	1	1	1	1	1	1	1	1
1968	1	1	1	1	1	1	1	1	1	1
1969	1	1	1	1	1	1	1	1	1	1
1970	2	2	2	2	2	2	2	2	2	2
1971	3	3	3	3	3	3	4	3	3	3

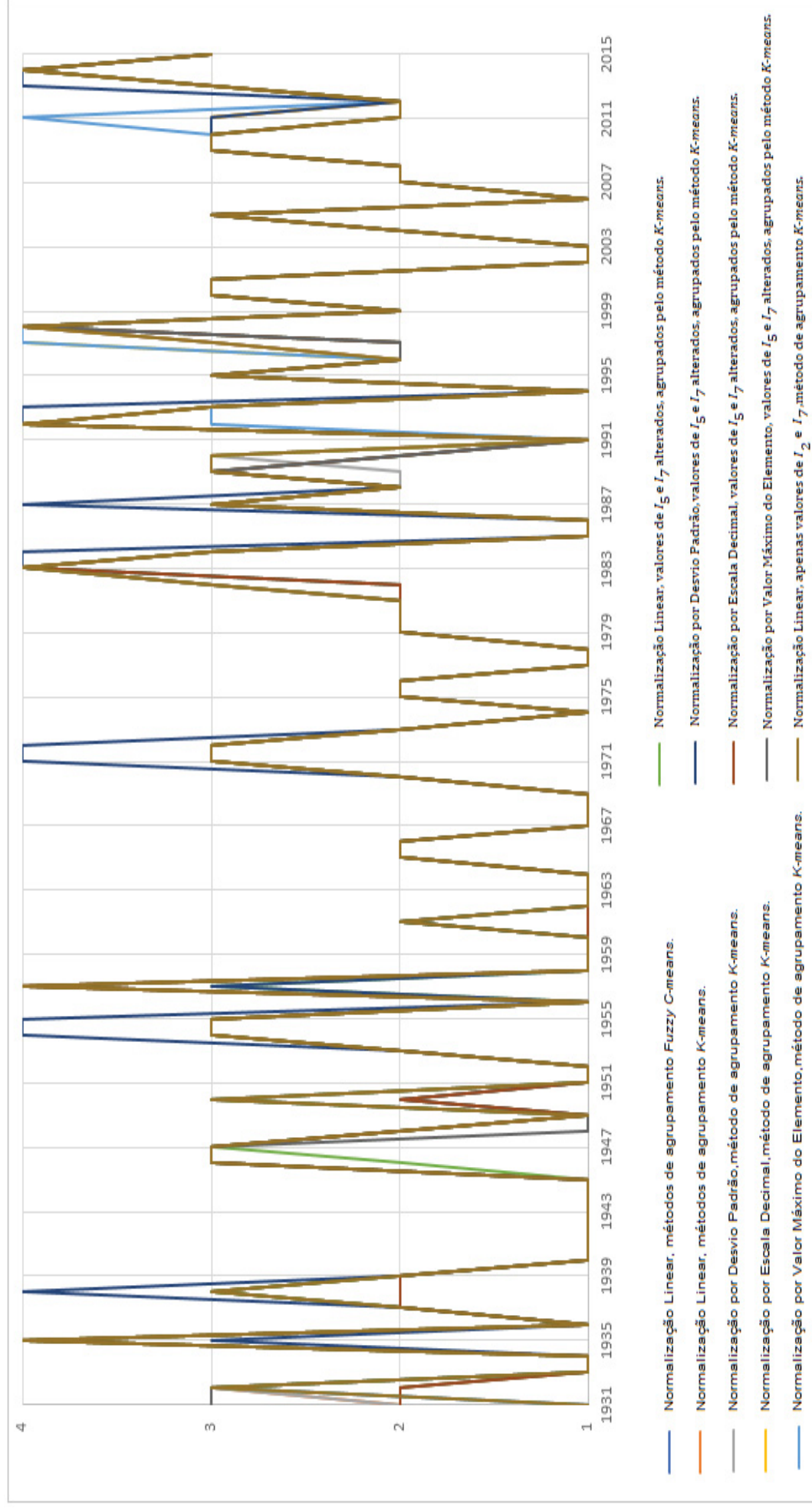
TABELA 25 - AGRUPAMENTO DOS DADOS DE TODAS AS AMOSTRAS, ATRAVÉS DE TODOS OS MÉTODOS DE NORMALIZAÇÃO E COMPARAÇÕES ANALISADAS NO TRABALHO (continuação).

Agrupamento (ver legenda)	A	B	C	D	E	F	G	H	I	J
ANO	GRUPO									
2013	3	3	3	3	3	3	4	3	3	3
2014	4	4	4	4	4	4	4	4	4	4
2015	3	3	3	3	3	3	3	3	3	3

TABELA 25 – LEGENDA DA TABELA 24

LEGENDA	
EVENTO	COMPARAÇÕES ANALISADAS
A	Normalização Linear, método de agrupamento <i>Fuzzy C-means</i> .
B	Normalização Linear, método de agrupamento <i>K-means</i> .
C	Normalização por Desvio padrão, método de agrupamento <i>K-means</i> .
D	Normalização por Escala decimal, método de agrupamento <i>K-means</i> .
E	Normalização por valor máximo do elemento, método de agrupamento <i>K-means</i> .
F	Normalização Linear I_5 e I_7 alterados, método de agrupamento <i>K-means</i> .
G	Normalização por Desvio padrão I_5 e I_7 alterados, método de agrupamento <i>K-means</i> .
H	Normalização por Escala decimal I_5 e I_7 alterados, método de agrupamento <i>K-means</i> .
I	Normalização por Valor máximo do elemento I_5 e I_7 alterados, método de agrupamento <i>K-means</i> .
J	Normalização Linear, apenas I_2 e I_7 , método de agrupamento <i>K-means</i> .
GRUPO	MAGNITUDE DA CHEIA
1	Baixa
2	Média
3	Alta
4	Excepcional

FIGURA 38 – VARIAÇÃO NO AGRUPAMENTO DAS AMOSTRAS EM VIRTUDE DO TIPO DE NORMALIZAÇÃO E DOS INDICADORES UTILIZADOS



5 CONCLUSÕES E RECOMENDAÇÕES

Estudos detalhados de cheias tem se tornado cada vez mais necessários, em virtude da intensidade de tais desastres naturais e seus impactos nas sociedades modernas. Este panorama se aplica ao Brasil e ao estado do Paraná, onde se destaca a situação peculiar da cidade de União da Vitória, foco do presente trabalho.

Neste contexto, cindir dados observados nos eventos de cheias e, corretamente classificá-los, torna-se tarefa primordial, uma vez que corrobora com o claro entendimento do fenômeno, auxiliando sua previsão e consequente mitigação. Destarte, em primazia de raciocínio, questiona-se “qual metodologia empregar na classificação?” Não obstante, inquire-se também se a aplicação de diferentes métodos de agrupamento e/ou diferentes métodos de tratamento inicial dos dados (normalização matemática) produzem alterações na classificação dos dados.

Visando responder tais questões, o presente trabalho empregou o método de agrupamento *K-means*, aplicado à mesma base de dados e indicadores de cheia, usados por Steffen (2017), que utilizou o método *Fuzzy C-means* para a classificação de cheias históricas observadas na estação fluviométrica União da Vitória (65310000), localizada no rio Iguaçu, na cidade de União da Vitória, Paraná, Brasil.

As cheias máximas anuais foram classificadas em 4 grupos, denominados, segundo a magnitude da cheia, em: magnitude baixa, magnitude média, magnitude alta e magnitude excepcional.

De modo geral, os métodos de agrupamento utilizados (*K-means* e *Fuzzy C-means*) produziram grupos muito semelhantes entre si. Houve algumas alterações nos grupos formados; sendo que, na maioria das vezes, o vetor formado pelos sete indicadores de cheia adotados deslocou-se um grau na escala de magnitude da cheia, positiva e negativamente, para os anos onde houve modificação na classificação.

Em referência à técnica de tratamento inicial dos dados (normalização matemática), devem ser levadas em consideração as variações na dispersão dos dados, ocasionadas pelos diferentes tipos de normalizações, tendo o presente trabalho observado um aumento na dispersão das amostras, quando utilizado o método de normalização por desvio padrão, e uma diminuição da dispersão das amostras, quando utilizado o método da escala decimal.

A variação na dispersão das amostras, resultante do uso das técnicas de normalização, foi determinante para discrepâncias observadas nas comparações entre os grupos de magnitude de cheia, formados pelos métodos *K-means* e *Fuzzy C-means*.

Observou-se uma maior similaridade entre os grupos formados, quando se utilizou normalização pelo método da escala decimal, e uma maior discrepância entre as amostras, quando se utilizou o método de normalização por desvio padrão.

Visando investigar as inconsistências eventualmente ocasionadas na formação dos grupos pelos diferentes métodos de normalização (tratamento inicial) de dados, foram empregadas 2 técnicas estatísticas: *teste t de Student* e ANOVA, as quais apresentaram valores de *p-valor* inferiores a 0,05, independentemente do tipo de normalização utilizado.

A um leitor atento, tal fato pode parecer intrigante, contudo, quando observados os *boxplots*, estabelecidos a partir das amostras, para cada método de normalização usado, é possível observar modificações ocorridas em todos os quartis, médias, medianas, *outliers* superiores e inferiores. Pelo exposto, verificou-se que a normalização altera a estratificação estatística, ocasionando modificação nos agrupamentos.

Em etapa final, o presente trabalho empregou a Análise de Componentes Principais (ACP) para avaliar a possibilidade de redução no número de indicadores de cheias utilizados para a classificação das cheias. O uso da ACP permitiu a redução dos indicadores de cheia de um número de 7 indicadores para apenas 2 indicadores: Nível máximo anual (I_2) e Taxa média de recessão do hidrograma de cheias (I_7).

O agrupamento produzido com o uso apenas dos dois referidos indicadores, obteve correlação de 100% entre os agrupamentos produzidos por *K-means* e por *Fuzzy C-means*; pelo exposto, iniciando a etapa de recomendações deste trabalho, a indicação é que, antes de se aplicar um algoritmo de agrupamento de dados à uma base amostral, seja executada uma análise de componentes principais (ACP), visando concentrar o agrupamento dos indicadores mais significativos, excluindo possíveis *outliers*.

Ainda como sugestão para trabalhos futuros, sugere-se a aplicação de outros algoritmos de agrupamento, baseados em coeficientes de *fuzzificação*, voltados a minorar parâmetros de incertezas relacionados às características físicas da bacia, fato que não é abordado pelo método *K-means*, não obstante, o emprego de técnicas de mineração de dados, aprendizado de máquina, inteligência computacional e SVM (*Support Vector Machine*).

Por fim, recomenda-se o uso da classificação das cheias e das conclusões apresentadas neste trabalho na modelagem de eventos de cheia, principalmente no entendimento e previsão da ocorrência de tais fenômenos, principalmente aqueles de

magnitude excepcional, de modo a tornar mais efetivas as medidas de mitigação, como por exemplo, na rápida emissão de alertas de segurança, visando situações de abandono de área.

Outrossim, as conclusões aqui explanadas poderão subsidiar poderes públicos em políticas de ocupação de solos; evitando danos ambientais, financeiros e sociais, colaborando na obtenção do equilíbrio civilização x meio ambiental, equilíbrio este cada vez mais almejado, na medida que o ser humano ao se enxergar como parte do ecossistema, adquire maturidade para enfrentar os desafios do terceiro milênio.

REFERÊNCIAS

ABDI, H; WILLIAMS, L.J. Principal components analysis. **Wiley Interdisciplinary Reviews:Computational Statistics**.V.2.p 433-459, 2010.

AITCHISON, J. **Principal component analysis of compositional data**. Department of Statistics, University of Hong Kong.1983. Disponível em:<<http://biomet.oxfordjournals.org/>>. Acesso em:12 jul.2020.

ANDERBERG, Michael R. **Cluster analysis for applications**. 1.ed. New York, Academic Press, 1973.

ALVARES, L. O. **Normalização de dados**; Notas de aula. Universidade Federal de Santa Catarina. Disponível em:<www.inf.ufsc.br/teaching/INE_5644/aula_DCBD>. Acesso em: 23 jul. 2019.

ANTON, H.; RORRES, C. **Álgebra Linear com Aplicações**. Bookman, Porto Alegre, 2004.

ANSBACH, O. **Navegando na memória: O patrimônio cultural da extinta hidrovía do Rio Iguaçu**.123f. Dissertação (Mestrado em Cartografia) - Setor de Ciências Cartográficas, Universidade estadual de ponta grossa. Ponta Grossa. 2008.

ARAÚJO, A. N. D. **Simulação hidrológica com chuva estimada por satélite**.163f. Dissertação (Mestrado em Engenharia de Recursos Hídricos e Ambiental) - Setor de Ciências Exatas e de Tecnologia. Universidade Federal do Paraná, Curitiba, 2006.

AZEVEDO,J. **Situações de risco: cheias e inundações**. Disponível em: <<http://www.google.pt/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&ved=0CC0QFjAA&url=http%3A%2F%2Fwww.administradores.com.br%2Finforme-ie%2Fproducao-academica%2Fsituacoes-de-risco-cheias-e-nundacoes%2.pdf>>.Acesso em:11 mar. 2019.

BALASKO, B.; ABONYI, J.; FIEL, B. **Fuzzy clustering and data analysis toolbox**. Disponível em: <<http://www.fmt.vein.hu/softcomp/fclusttoolbox/FuzzyClusteringToolbox.pdf>>. Acesso em:11 mar. 2019.

BAO, H. T. **Automatic Clustering Detection**. Disponível em: <<http://netnam.vn/unescocourse/knowlegde/51.html>>. Acesso em: 10 abr. 2019.

BARROS, R.P.; HENRIQUES, R.M.; O combate à pobreza no Brasil: dilemas entre políticas de crescimento e políticas de redução da desigualdade. **Anais do Seminário Desigualdade e Pobreza no Brasil**. Rio de Janeiro, ago.1999.

BERNARDI, J.V.E.; FOWLER, H.G.; LANDIM, P.M.B. Um estudo de impacto ambiental utilizando análises estatísticas espacial e multivariada. **Holos Environmental**, 2001, p 162-172.

BERTOLDO, J. G. et al. Problemas relacionados com o uso de testes de comparação de médias em artigos científicos. **Biotemas**, [S.I.], v. 21, n. 2, p. 145-153, 2008.

BEZDEK, J. C;PAL, S. K. **Fuzzy models for pattern recognition: methods that search for structures in data. Methods that search for structures in data.** 1.ed. IEEE Press. New York, 1992.

BRASIL. MINISTÉRIO DA INTEGRAÇÃO NACIONAL. Agência Nacional de Águas. Situação da Água no Mundo,2019.Disponível em: <http://www.ana.gov.br/panorama_das_aguas/aguas-no-mundo>. Acesso em: 5 jan. 2019.

BRASIL. MINISTÉRIO DA INTEGRAÇÃO NACIONAL. Agência Nacional de Águas. **Conjuntura dos recursos hídricos no brasil - Encarte especial sobre a crise hídrica.**Brasília, 2000.

BRASIL,MINISTÉRIO DAS CIDADES;IPT, INSTITUTO DE PESQUISAS TECNOLÓGICAS.**MAPEAMENTO DE RISCOS EM ENCOSTAS E MARGENS DE RIOS.** 1.ed. Brasília: Ministério das Cidades,2007.

BRASIL. Lei número 9.433,de 8 de Janeiro de 1997.Institui a Política Nacional de Recursos Hídricos,cria o Sistema Nacional de Gerenciamento de Recursos Hídricos,regulamenta o inciso XIX do art.1 da Lei número 8.001,de 13 de março de 1990. **Diário Oficial da União**, Brasília.

BRASIL. Lei número 9.984,de 17 de julho de 2000.Dispõe sobre a criação da Agência Nacional de Águas - ANA, entidade federal de implementação da Política Nacional de Recursos Hídricos e de coordenação do Sistema Nacional de Gerenciamento de Recursos Hídricos. **Diário Oficial da União.**, Brasília, 18 Julho 2000.

BRASIL. MINISTÉRIO DA INTEGRAÇÃO NACIONAL. Secretaria Nacional de Proteção e Defesa Civil. Departamento de Minimização de Desastres. **Manual de desastres naturais.** Brasília, 2003.

BRASIL. MINISTÉRIO DE MINAS E ENERGIA. Secretaria de Planejamento e Desenvolvimento Energético. **Manual de Inventário Hidroelétrico de Bacias Hidrográficas.** Biblioteca da Presidência da república.Ed.2007. Disponível em: <www.biblioteca.presidencia.gov.br>Acesso em: 11/03/2020.

BROWNLEE, K. A. **Statistical theory and methodology in science and engineering.** New York: Wiley, 1965.

CASTANHARO, G.; BUBA, H. Revisão do comportamento hidráulico do rio Iguaçu na região de união da vitória e seu impacto na operação hidráulica dos reservatórios de Foz do Areia e Segredo. **Revista : Espaço Energia**, n 8,p.28-38,abr.2008.

CAVALCANTE, R. D. S. **Aritmética com python.** 2018.80.f.Dissertação (Programa de Mestrado Profissional de Matemática em Rede Nacional).Instituto de Matemática e estatística. Universidade Federal de Goiás.Goiânia, 2018.

CEPET UFSC; BANCO MUNDIAL. **Relatório de Danos Materiais e Prejuízos Decorrentes de Desastres Naturais no Brasil 1995-2014.** Florianópolis: CEPED UFSC.76f. 2016.

CHEN, X.;WANG, L. Flood feature identification and clustering in Wujiang river, South China. **XIX International Conference on Water Resources. 2012.**

CHMYZ, I. **Pesquisas Arqueológicas no alto e médio rio Iguaçu**. Curitiba: Universidade Federal do Paraná, 1969. 29 f. Relatório técnico.

COSTA, J. A. F. **Classificação Automática e Análise de Dados por Redes Neurais Auto-Organizáveis**. Disponível em: <http://www.dee.ufrn.br/~alfredo/index_tese.html>. Acesso em: 13 fev. 2019.

COX, E. **Fuzzy modeling and genetic algorithms for data mining and exploration. Morgan Kaufmann series in data management systems**. 1.ed. New York: Elsevier/Morgan Kaufmann, 2005.

DEFESA CIVIL, PARANÁ. **Enchente de União da Vitória de 1983**. Disponível em: <http://www.defesacivil.pr.gov.br/sites/defesa-civil/arquivos_restritos/files_documento/2019-05/enchente_de_uniao_da_vitoria_de_1983.pdf>. Acesso em 05 dez. 2019.

EIRD/ONU, Estratégia Internacional para Redução de desastres da Organização das Nações Unidas. Terminologia sobre redução de riscos de desastres. Suíça: ONU: 2009. Disponível em: <http://www.unisdr.org/files/7817_UNISDR_Terminology_Spanish.pdf>. Acesso em: 03 set. 2019.

EM-DAT. The International Disaster Database - Center for Research on The Epidemiology of Disaster-CRED. **Floods; South America**. Disponível em: <<https://www.emdat.be/>>. Acesso em: 02 out. 2019.

EVERITT, B. S.; LANDAL, S.; LEESE, M. **Cluster Analysis**. 5.ed. Colorado: A Hodder Arnold Publication, 2001.

FRALEY, C.; RAFTERY, A. **How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis**. Department of Statistics, University of Washington, USA. Disponível em: <<https://sites.stat.washington.edu/raftery/Research/PDF/fraley1998.pdf>> Acesso em: 09 mai. 2020.

FUNG, G. A **Comprehensive Overview of Basic Clustering Algorithms**, 2001. Disponível em: <<http://www.cs.wisc.edu/~gfunf/clustering.pdf>>. Acesso em: 20 mar. 2019.

GOERL, R. F.; KOBAYAMA, M. Considerações sobre as inundações no Brasil. **REGA-Revista de Gestão de Água da América Latina**. v. 1, n.1, p. 14 -15, set. 2008.

HAIR, J. F. et al. **Análise multivariada de dados**. São Paulo: Bookman Editora, 2009.

JAIN, A. K.; DUBES, R. C. **Algorithms for Clustering Data**. 5 ed. New Jersey: Michigan States University, 1988.

JAIN, A. K.; DUBES, R. C. Data clustering: a review. **Association for Computing Machinery**, Fairfax, v. 31,3, p. 264 - 323, 1999.

JORGE, F.V.; **A dinâmica pluvial do clima subtropical: variabilidade e tendências no Sul do Brasil**. 125f. Dissertação (Mestrado em ciências da terra) - Setor de Ciências da Terra, Universidade Federal do Paraná, Curitiba, 2015.

KAINULAINEN, J. **Clustering Algorithms: Basics and Visualization**. Disponível em: <<http://www.niksula.cs.hut.fi/~jkainula/pdfs/clustering.pdf>>. Acesso em: 14 jun. 2019.

KARPINSKI, C. Gentes e paisagens do Rio Iguaçu na viagem expedicionária dos engenheiros Keller em 1886. **Revista História regional**, p. 38-65, 2012.

KAUFMAN, L.; ROUSSEAU, P.J. **Finding Groups in Data: An Introduction to Cluster Analysis**. New Jersey: John Wiley & Sons, Inc, 2005.

KRAFT et al. Atributos microbianos do solo em sistema plantio direto no Oeste Catarinense. **XII Reunião Sul Brasileira de ciência do Solo**. Xanxerê 2018.

KRONBAUER, L.A.M.; FONTANA, L.M.; WINCK, A.T. Um estudo sobre processos para avaliação de algoritmos de agrupamento de dados. **Revista Coming - Communication and innovations gazette**, Santa Maria, v.1, n.1, 2016.

KELLER, José; KELLER, Francisco. Relatório da exploração do rio Iguassú, feita em 1866 pelos engenheiros José e F. Keller. **Jornal Dezenove de Dezembro**, Curitiba, 31 de ago. 1867, n.817, p.3.

KVALHEIM, O.M. Interpretation to latent-variable for projection methods and their use and aims in the interpretation of multicomponent spectroscopic and chromatographic data. **hemometrics and Intelligent Laboratory Systems**, v.4, p.11-15, 1998.

LACHI, R. L.; ROCHA, H. V. **Aspectos básicos de clustering: conceitos e técnicas**. Núcleo de Informática Aplicada à Educação. Instituto de Computação – Universidade Estadual de Campinas. Campinas, São Paulo. 2005.

LINDEN, R. **Um algoritmo híbrido para extração de conhecimento em Bioinformática**. 114f. Tese (Doutorado em Ciência de Dados) - Centro de Tecnologia, Universidade Federal do Rio de Janeiro, 2009.

LINDEN, R. Técnicas de agrupamento. **Revista de Sistemas de Informação da FSMA**. n. 4, 2009, pp.18-36. Disponível em: <http://www.fsma.edu.br/si/edicao4/FSMA_SI_2009_2_Tutorial.pdf>. Acesso em: 14/07/2019.

LIU, Z.D.; LIU, Q.Y.; XU, L.; ZHANG, G.Y. **Jiang Distributed lag effects and vulnerable groups of floods on bacillary dysentery in Huaihua, China**. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/27427387>>. Acesso em: 14 jul. 2019.

MACIEL, S. A. **Análise da relação chuva-vazão na bacia**. 214f. Dissertação (Mestrado em Geografia) - Instituto de Geografia, Universidade Federal de Minas Gerais. Belo Horizonte, 2017.

MAHALANOBIS, P.C., ON THE GENERALIZED DISTANCE IN STATISTICS. **The National Institute of Sciences of India**. Caucatá, v.2, n.1, p.49-55, jan./abr. 1936.

MANLY, B.F.J. **Cluster Analysis**. In: **Multivariate statistical methods**. 2 ed. London: Chapman, Cap. 9, p. 129-145. 1994

MARDIA, K. V.; KENT, J. T.; BIBBY, J. M. **Multivariate analysis**. 7. ed. San Diego: Academic Press, v. 1, 1979. 267 p.

MATTEUCCI, M. A. **Tutorial on Clustering Algorithms**. Disponível em: <http://elet.polimi.it/upload/matteucc/Clustering/tutorial_html/hierarchi.cal.html>. Acesso em: 5 mai. 2019.

MOSCATO, P.; ZUBER, F. J. V. **Um Visão geral de Clusterização de dados**, 2008. Disponível em: <http://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ia368_02/topico5_02.pdf>. Acesso em: 1 mai. 2019.

MOORE, A. Very fast EM-based mixture model clustering using multiresolution kd-trees. **Advances in Neural Information Processing Systems**, San Diego, v. 1, n.5, p. 7 -12, set. 1999.

NIEVOLA, J. C. **Análise de Agrupamentos**. Notas de aula. Pontificia Universidade Católica do Paraná. Curitiba, 23f. 2015. Disponível em: <www.ppgia.pucpr.br/~fabricio/ftp/Aulas/Mestrado/IA/Nievola/MD/MD-06-Agrupamento.pdf>. Acesso em: 20 ago. 2019.

PARANÁ. **Bacias hidrográficas do Paraná**. Secretaria de Estado do Meio Ambiente e Recursos Hídricos - SEMA. Curitiba, p. 140. 2010.

PARANÁ. **Bacias hidrográficas do Paraná - Série histórica**. Secretaria de Estado do Meio Ambiente e Recursos Hídricos - SEMA. Curitiba, p. 154. 2013.

PARANÁ, G. D. E. D. **Plano diretor da cidade de União da Vitória**. Secretaria de Estado do Desenvolvimento Urbano. Curitiba, p. 473. 2007. Disponível em: <<http://www.uniaoavitoria.pr.gov.br/wp-content/uploads/2017/08/polar-uniao-da-vitoria-2.pdf>>. Acesso em: 21 jun. 2019.

RAMOS, V. PORQUE APRENDER PYTHO: 5 MOTIVOS PARA COMEÇAR AGORA. **Python Academy**. 11, abril 2018. Disponível em: <<https://pythonacademy.com.br/blog/porque-aprender-python-5-motivos-para-aprender-agora>>. Acesso em: 09, mai de 2020.

RAMALHO, L. **Python Fluente: Programação clara, concisa e eficaz**. 3.ed. São Paulo: Novatec, v. 1, 2019. 799 p.

RAFAELI NETO, S.L.R. **Um modelo conceitual de sistema de apoio à decisão espacial para gestão de desastres por inundações**. 231p. Tese (Doutorado em Engenharia). Universidade de São Paulo - USP. São Paulo, 2000.

RENNÓ, C.D. **Estatística: Aplicação ao Sensoriamento Remoto**. Notas de aula. INPE - Instituto Nacional de Pesquisas Espaciais. Disponível em: <<http://www.dpi.inpe.br/~camilo/estatistica/index.html>>. Acesso em: 15 jan. 2020.

RIGNEL, D. G.; SALLES, H.L.; TUNESS, F.D. 117f. Uma introdução à Lógica Fuzzy. **Revista Eletrônica de Sistemas de Informação e Gestão Tecnológica**, v. 1, n. 1, p.17-28, 2011. Disponível em: <http://www.logicafuzzy.com.br/wp-content/uploads/2013/04/uma_introducao_a_logica_fuzzy.pdf>. Acesso em: 25 ago. 2019.

ROCHA, P.S.M. **Gestão em áreas de risco de enchentes: estudo de caso para União da Vitória - Paraná**. 134f. Dissertação (Mestrado em Gestão Ambiental). Universidade Positivo, Curitiba, 2012.

ROKACH, L. **CLUSTERING METHODS**. Department of Industrial Engineering Tel-Aviv University. Disponível em: < https://link.springer.com/chapter/10.1007%2F0-387-25465-X_15>. Acesso em: 11 set.2020.

SALES, S. D. S. **Determinação do efeito do regime hidrológico e da regularização do caudal na ecologia trófica do barbo do norte (Luciobarbus bocagei Straindachner 1864)**.157f. Dissertação (Mestrado em Gestão e Conservação de Recursos Naturais) - Setor de ciências da Terra, Universidade de Lisboa. Évora, 2015.

SILVA, C. da. **Apontamentos Históricos de União da Vitória (1769-1933)**. Curitiba: Imprensa Oficial, Curitiba, 2006.

SILVEIRA, P. V. D. **Sob as Águas do Iguaçu: Uma análise dos impactos das enchentes sobre as coletividades do município de União da Vitória**. 80f. Monografia (Licenciatura e Bacharelado em Ciências Sociais) - Setor de Ciências Humanas, Letras e Artes, Universidade Federal do Paraná. Curitiba, 2016.

STEPHENSON, D. **Integrated flood plain management strategy for the Vaal**. 1. ed. [S.I.]: Stbooks, v. 1, 2002. 425-430 p.

STEFFEN, P.C. **Agrupamento de cheias históricas**. 178f. Dissertação (Mestrado em Engenharia de Recursos Hídricos e Ambiental) - Setor de Ciências Exatas e de Tecnologia. Universidade Federal do Paraná, Curitiba, 2017.

STEFFEN, P.C; GOMES, J. Clustering of historical floods observed on Iguaçu River, in União da Vitória. **Brazilian Journal of Water Resources**, Porto Alegre, v.23, n.38, p.8, julho 2018.ISSN 2318-0331.

SWEIGART, A. **AUTOMATIZE TAREFAS MAÇANTES COM PYTHON: PROGRAMAÇÃO PRÁTICA PARA VERDADEIROS INICIANTES**.3.ed.São Paulo: Novatec, v. 1, 2018. 568 p.

TABACHNICK, B. G.; FIDELL, L. S.; OSTERLIND, S. J. **Using multivariate statistics**. New York: Pearson, 2001.

TUCCI, C. E. M. Gerenciamento de drenagem urbana. **RBRH-Revista Brasileira de Recursos Hídricos**, v. 7, n. 1, p. 5-27, jan/mar 2002.

TUCCI, C. E. M.; BERTONI, J. C. **Inundações urbanas na América do Sul**. 1. ed. Porto Alégre: Associação Brasileira de Recursos Hídricos, 2003.

UFSC; CEPET. **Atlas Brasileiro de desastres naturais 1991 A 2012**. Florianópolis:CEPET UFSC, 2013.105 f. Relatório técnico

ULIANA, E. M. et al. Modelo hidrológico híbrido para previsão de vazões na bacia do rio Piracicaba. **Revista Brasileira de Meteorologia**, São Paulo,p.34, Outubro/Dezembro 2019. Acesso em: 25 dez. 2019.

VACA, Á.Ñ.C.de. **Nalfrágios e Comentários**.1.ed.Espanha, 1554. Disponível em:<http://www.dominiopublico.gov.br/pesquisa/PesquisaObraForm.do?select_action=&o_autor=1077>. Acesso em: 21 fev.2020.

VICINI, L. **Análise multivariada da teoria à prática**.115f.Monografia(Departamento de Estatística). Universidade Federal de Santa Maria. Santa Maria, 2005.

XU, H. et al. Urban flooding risk assessment based on an integrated K-means cluster algorithm and improved entropy weight method in the region of Haikou,China. **Journal of Hydrology**, Los Angeles, 35, p.975-986, ago.2018.

WANG, L. N.; CHEN, X. H.; SHAO, Q. X.; LI, Y. Flood indicators and their clustering features in Wujiang River, South China. **Ecological Engineering**, v. 76, p. 66-74, abr. 2014.

YURAS, G. **Minimum Distance Classifiers**. Disponível em: <<http://www.profc.udec.cl/~gabriel/tutoriales/rsnote/cp11-6.html>>. Acesso em: 11 abr. 2019.

YONAMINE, F. S.; SPECIA, L.; CARVALHO, V. O.; NICOLETTI, M. C. **Aprendizado não supervisionado em domínios fuzzy Algoritmo fuzzy c-means**. São Carlos:Universidade Federal de São Carlos,Centro de Ciências Exatas e de Tecnologia, Departamento de computação, 2002.Relatório técnico.

ZAHMATKESH, Z. et al. Analysis of the effects of climate change on urban storm water runoff using statistically downscaled precipitation data and a change factor approach. **Journal of Hydrologic Engineering**, New York, set. 2014.

ZHAO, YING.; KARYPIS, GEORGE. **Comparison of Agglomerative and Partitional Document Clustering Algorithms TR-02-014**. Department of Computer Science and Engineering University of Minnesota 4- 192 EECS Building 200 Union Street SE Minneapolis, MN. April 17, 2002.

ANEXO 1 – VAZÃO MÁXIMA ANUAL (I_1) E NÍVEL DE ÁGUA MÁXIMO ANUAL (I_2).

Ano	Data pico	Q(m³/s)	H (cm)	Ano	Data pico	Q(m³/s)	H (cm)
1931	13/06/31	1453	510	1974	28/01/74	1022	418
1932	22/04/32	1833,2	584	1975	15/12/75	1711	562
1933	10/10/33	836,3	373	1976	15/06/76	1419,4	503
1934	02/03/34	820	369	1977	19/10/77	1048	424
1935	17/10/35	3266,8	816	1978	25/07/78	816	368
1936	09/06/36	1242	468	1979	16/05/79	1665	533
1937	19/11/37	1399	499	1980	06/09/80	1715	548
1938	03/07/38	2383	682	1981	01/01/81	1887,1	579
1939	06/09/39	1716,5	563	1982	27/11/82	2263	645
1940	06/05/40	568	302	1983	18/07/83	5156,7	1042
1941	10/02/41	1268	474	1984	09/08/84	1939,5	594
1942	20/02/42	1198	457	1985	21/04/85	768,5	347
1943	05/08/43	901,9	389	1986	29/12/86	1256,9	459
1944	19/03/44	956,4	402	1987	29/05/87	2167,4	628
1945	20/07/45	1218	462	1988	25/05/88	1758,9	556
1946	04/03/46	2333,6	673	1989	25/09/89	1887,1	579
1947	05/10/47	1777	574	1990	27/01/90	2196	633
1948	18/08/48	1443,4	508	1991	24/06/91	1155,3	437
1949	10/06/49	804	365	1992	08/06/92	3953,6	890
1950	21/10/50	1610	542	1993	06/10/93	2786,2	725
1951	26/10/51	1194	456	1994	08/07/94	1359,8	480
1952	20/10/52	1093	434	1995	23/01/95	2453,3	675
1953	03/11/53	1550	530	1996	12/07/96	1554,5	519
1954	27/05/54	2402,5	685	1997	14/10/97	2395,4	666
1955	28/06/55	1985	612	1998	12/10/98	2751,7	720
1956	09/05/56	1234	466	1999	14/07/99	1832,9	570
1957	20/08/57	2674,7	728	2000	25/09/00	2284	649
1958	16/09/58	1030	420	2001	11/10/01	1975,6	596
1959	08/09/59	804	365	2002	02/10/02	1162,3	440
1960	28/10/60	1093	434	2003	24/12/03	1223,9	454
1961	05/11/61	1782,5	575	2004	27/10/04	1500,7	510
1962	20/10/62	1002	413	2005	20/09/05	1896,7	582
1963	28/03/63	1048	424	2006	27/12/06	580,7	303
1964	31/08/64	914,4	392	2007	25/05/07	1864	577
1965	15/07/65	1471,8	514	2008	31/10/08	1470,6	505
1966	23/02/66	1399	499	2009	03/10/09	2050,2	609
1967	06/03/67	1210	460	2010	08/05/10	2605,6	700
1968	27/01/68	745	350	2011	10/09/11	2388,3	665
1969	08/04/69	1286	478	2012	20/06/12	1813,6	568
1970	09/07/70	1495,3	519	2013	02/07/13	2391,6	667
1971	13/01/71	2428,5	689	2014	17/06/14	3385,5	813
1972	30/08/72	1833,2	584	2015	24/10/15	1901	583
1973	31/08/73	1760,5	571				

FONTE: STEFFEN (2017).

ANEXO 2 – VOLUMES TOTAIS EXTRAVASADOS (*I*₃).

Ano	Volume extravasado (km ³)	Ano	Volume extravasado (km ³)	Ano	Volume extravasado (km ³)	Ano	Volume extravasado (km ³)
1931	0,01	1953	0,06	1975	0,27	1997	2,08
1932	0,02	1954	1,15	1976	0,01	1998	3,33
1933	0,00	1955	1,23	1977	0,00	1999	0,46
1934	0,00	1956	0,00	1978	0,00	2000	0,84
1935	3,21	1957	5,35	1979	0,06	2001	0,52
1936	0,00	1958	0,00	1980	0,23	2002	0,00
1937	0,00	1959	0,00	1981	0,6	2003	0,00
1938	1,11	1960	0,00	1982	1,34	2004	0,02
1939	0,27	1961	0,31	1983	9,69	2005	0,77
1940	0,00	1962	0,00	1984	0,46	2006	0,00
1941	0,00	1963	0,00	1985	0,00	2007	0,31
1942	0,00	1964	0,00	1986	0,00	2008	0,02
1943	0,00	1965	0,04	1987	1,02	2009	1,16
1944	0,00	1966	0,00	1988	0,26	2010	1,57
1945	0,00	1967	0,00	1989	0,55	2011	1,72
1946	1,39	1968	0,00	1990	0,72	2012	0,42
1947	0,46	1969	0,00	1991	0,00	2013	1,13
1948	0,02	1970	0,08	1992	3,87	2014	2,87
1949	0,00	1971	1,6	1993	1,77	2015	0,65
1950	0,09	1972	0,7	1994	0,00		
1951	0,00	1973	0,44	1995	1,24		
1952	0,00	1974	0,00	1996	0,06		

FONTE: STEFFEN (2017).

ANEXO 3 – TEMPO ACIMA DO LIMIAR DE EXTRAVASAMENTO (I_4).

Ano	I_4 (dias)	Ano	I_4 (dias)	Ano	I_4 (dias)	Ano	I_4 (dias)
1931	3,1	1953	6,5	1975	16,4	1997	56,6
1932	8,2	1954	20	1976	5,1	1998	61
1933	0,0	1955	35,9	1977	0,0	1999	16,7
1934	0,0	1956	0,0	1978	0,0	2000	17,8
1935	43,2	1957	88,9	1979	6,9	2001	18,2
1936	0,0	1958	0,0	1980	13,2	2002	0,0
1937	0,7	1959	0,0	1981	21,6	2003	0,0
1938	21,5	1960	0,0	1982	27,1	2004	5,3
1939	16,1	1961	14,3	1983	87,9	2005	42,2
1940	0,0	1962	0,0	1984	17,6	2006	0,0
1941	0,0	1963	0,0	1985	0,0	2007	13,7
1942	0,0	1964	0,0	1986	0,0	2008	6,4
1943	0,0	1965	8,5	1987	22,3	2009	45,5
1944	0,0	1966	2,1	1988	14,9	2010	25,3
1945	0,0	1967	0,0	1989	16,7	2011	55,1
1946	40,2	1968	0,0	1990	19,2	2012	16,9
1947	19,2	1969	0,0	1991	0,0	2013	19,5
1948	4,7	1970	11,2	1992	24,8	2014	24,9
1949	0,0	1971	29,8	1993	27,2	2015	35,4
1950	7,7	1972	27,6	1994	0,0		
1951	0,0	1973	29,6	1995	23,3		
1952	0,0	1974	0,0	1996	8,9		

FONTE: STEFFEN (2017).

ANEXO 4 – TEMPOS DECORRIDOS ENTRE O INÍCIO DO EVENTO E ULTRAPASSAGEM DO LIMIAR DE EXTRAVASAMENTO (I_5).

Ano	I_5 (dias)	Ano	I_5 (dias)	Ano	I_5 (dias)	Ano	I_5 (dias)
1931	6,4	1953	16,3	1975	17,1	1997	26
1932	8,6	1954	20,9	1976	17,9	1998	12,5
1933	0,00	1955	8,5	1977	0,00	1999	24,9
1934	0,00	1956	0,00	1978	0,00	2000	5,3
1935	13,2	1957	5,3	1979	14,9	2001	10,7
1936	0,00	1958	0,00	1980	13,1	2002	0,00
1937	12,8	1959	0,00	1981	24,7	2003	0,00
1938	6,5	1960	0,00	1982	10,8	2004	17
1939	27,8	1961	58,4	1983	32,5	2005	6,8
1940	0,0	1962	0,0	1984	6,0	2006	0,0
1941	0,0	1963	0,0	1985	0,0	2007	28,6
1942	0,0	1964	0,0	1986	0,0	2008	28,8
1943	0,0	1965	9,9	1987	10,1	2009	11,2
1944	0,0	1966	34,3	1988	29,1	2010	1,9
1945	0,0	1967	0,0	1989	5,9	2011	57,0
1946	27,7	1968	0,0	1990	27,0	2012	20,3
1947	7,6	1969	0,0	1991	0,0	2013	5,6
1948	31,5	1970	26,7	1992	15,8	2014	22,5
1949	0,0	1971	17,6	1993	20,0	2015	18,2
1950	7,8	1972	3,7	1994	0,0		
1951	0,0	1973	17,4	1995	17,4		
1952	0,0	1974	0,0	1996	23,8		

ANEXO 5 – TAXA MÉDIA DE ASCENSÃO NO TRECHO ENTRE O INÍCIO DA ONDA DE CHEIA E O INSTANTE DE EXTRAVASAMENTO (I_6), ESTIMADOS POR STEFFEN (2017).

Ano	I_6 (m³/s/dia)	Ano	I_6 (m³/s/dia)	Ano	I_6 (m³/s/dia)	Ano	I_6 (m³/s/dia)
1931	139,90	1953	72,80	1975	60,60	1997	43,80
1932	110,30	1954	58,30	1976	60,10	1998	75,20
1933	0,00	1955	114,60	1977	0,00	1999	48,00
1934	0,00	1956	0,00	1978	0,00	2000	211,50
1935	81,30	1957	233,60	1979	85,70	2001	98,30
1936	0,00	1958	0,00	1980	74,80	2002	0,00
1937	84,00	1959	0,00	1981	44,20	2003	0,00
1938	162,60	1960	0,00	1982	62,10	2004	70,20
1939	44,50	1961	21,80	1983	33,00	2005	171,40
1940	0,00	1962	0,00	1984	191,30	2006	0,00
1941	0,00	1963	0,00	1985	0,00	2007	43,60
1942	0,00	1964	0,00	1986	0,00	2008	40,10
1943	0,00	1965	117,70	1987	122,50	2009	97,50
1944	0,00	1966	30,90	1988	41,70	2010	510,80
1945	0,00	1967	0,00	1989	168,70	2011	21,80
1946	46,00	1968	0,00	1990	47,20	2012	57,40
1947	108,80	1969	0,00	1991	0,00	2013	194,80
1948	40,10	1970	46,50	1992	72,50	2014	54,30
1949	0,00	1971	72,60	1993	61,20	2015	59,50
1950	147,10	1972	306,70	1994	0,00		
1951	0,00	1973	61,20	1995	63,90		
1952	0,00	1974	0,00	1996	52,50		

FONTE: STEFFEN (2017).

ANEXO 6 – TAXA MÉDIA DE RECESSÃO DA ONDA DE CHEIA (I_7).

Ano	I_7 (m³/s/dias)	Ano	I_7 (m³/s/dias)	Ano	I_7 (m³/s/dias)	Ano	I_7 (m³/s/dias)
1931	26,70	1953	34,50	1975	58,90	1997	19,60
1932	77,10	1954	103,00	1976	8,10	1998	86,40
1933	0,00	1955	23,50	1977	0,00	1999	67,20
1934	0,00	1956	0,00	1978	0,00	2000	111,20
1935	91,80	1957	26,90	1979	36,90	2001	66,40
1936	0,00	1958	0,00	1980	35,50	2002	0,00
1937	24,00	1959	0,00	1981	44,00	2003	0,00
1938	71,10	1960	0,00	1982	62,80	2004	27,30
1939	28,20	1961	33,50	1983	132,60	2005	50,60
1940	0,00	1962	0,00	1984	35,40	2006	0,00
1941	0,00	1963	0,00	1985	0,00	2007	42,00
1942	0,00	1964	0,00	1986	0,00	2008	15,90
1943	0,00	1965	19,30	1987	83,30	2009	24,70
1944	0,00	1966	8,60	1988	28,50	2010	108,40
1945	0,00	1967	0,00	1989	111,10	2011	62,20
1946	78,00	1968	0,00	1990	87,80	2012	52,20
1947	44,00	1969	0,00	1991	0,00	2013	99,30
1948	25,60	1970	22,20	1992	189,10	2014	129,80
1949	0,00	1971	83,90	1993	92,00	2015	79,60
1950	49,30	1972	109,40	1994	0,00		
1951	0,00	1973	18,70	1995	78,00		
1952	0,00	1974	0,00	1996	24,80		

FONTE: STEFFEN (2017).

ANEXO 7 – MATRIZ DE INDICADORES DE CHEIA NORMALIZADOS PELO MÉTODO DA NORMALIZAÇÃO LINEAR.

Ano	I_1	I_2	I_3	I_4	I_5	I_6	I_7
1931	0,1929	0,2811	0,0014	0,0347	0,1093	0,2740	0,1409
1932	0,2757	0,3811	0,0204	0,0917	0,1477	0,2160	0,4078
1933	0,0585	0,0959	0,0000	0,0000	0,0000	0,0000	0,0000
1934	0,0549	0,0905	0,0000	0,0000	0,0000	0,0000	0,0000
1935	0,5881	0,6946	0,3310	0,4862	0,2263	0,1592	0,4856
1936	0,1469	0,2243	0,0000	0,0000	0,0000	0,0000	0,0000
1937	0,1811	0,2662	0,0000	0,0079	0,2190	0,1645	0,1269
1938	0,3955	0,5135	0,1143	0,2413	0,1119	0,3183	0,3762
1939	0,2503	0,3527	0,0274	0,1808	0,4758	0,0870	0,1491
1940	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
1941	0,1525	0,2324	0,0000	0,0000	0,0000	0,0000	0,0000
1942	0,1373	0,2095	0,0000	0,0000	0,0000	0,0000	0,0000
1943	0,0728	0,1176	0,0000	0,0000	0,0000	0,0000	0,0000
1944	0,0846	0,1351	0,0000	0,0000	0,0000	0,0000	0,0000
1945	0,1417	0,2162	0,0000	0,0000	0,0000	0,0000	0,0000
1946	0,3848	0,5014	0,1437	0,4518	0,4732	0,0901	0,4123
1947	0,2635	0,3676	0,0475	0,2162	0,1304	0,2130	0,2329
1948	0,1908	0,2784	0,0016	0,0533	0,5384	0,0785	0,1353
1949	0,0514	0,0851	0,0000	0,0000	0,0000	0,0000	0,0000
1950	0,2271	0,3243	0,0091	0,0863	0,1342	0,2880	0,2609
1951	0,1364	0,2081	0,0000	0,0000	0,0000	0,0000	0,0000
1952	0,1144	0,1784	0,0000	0,0000	0,0000	0,0000	0,0000
1953	0,2140	0,3181	0,0058	0,0725	0,2785	0,1425	0,1822
1954	0,3998	0,5176	0,1183	0,2245	0,3575	0,1142	0,5446
1955	0,3088	0,4189	0,1265	0,4036	0,1458	0,2243	0,1244
1956	0,1451	0,2216	0,0000	0,0000	0,0000	0,0000	0,0000
1957	0,4591	0,5757	0,5523	1,0000	0,0901	0,4574	0,1421
1958	0,1007	0,1595	0,0000	0,0000	0,0000	0,0000	0,0000
1959	0,0514	0,0851	0,0000	0,0000	0,0000	0,0000	0,0000
1960	0,1144	0,1784	0,0000	0,0000	0,0000	0,0000	0,0000
1961	0,2647	0,3689	0,0321	0,1613	1,0000	0,0427	0,1774
1962	0,0946	0,1500	0,0000	0,0000	0,0000	0,0000	0,0000
1963	0,1046	0,1649	0,0000	0,0000	0,0000	0,0000	0,0000
1964	0,0755	0,1216	0,0000	0,0000	0,0000	0,0000	0,0000
1965	0,1970	0,2865	0,0045	0,0951	0,1701	0,2304	0,1020
1966	0,1811	0,2662	0,0020	0,0320	0,5875	0,0604	0,0453
1967	0,1399	0,2135	0,0000	0,0000	0,0000	0,0000	0,0000
1968	0,0386	0,0649	0,0000	0,0000	0,0000	0,0000	0,0000
1969	0,1565	0,2378	0,0000	0,0000	0,0000	0,0000	0,0000
1970	0,2021	0,2932	0,0083	0,1254	0,4572	0,0910	0,1175
1971	0,4055	0,5230	0,1646	0,3347	0,3019	0,1421	0,4437

Ano	I_1	I_2	I_3	I_4	I_5	I_6	I_7
1972	0,2757	0,3811	0,0719	0,3107	0,0626	0,6005	0,5787
1973	0,2599	0,3635	0,0453	0,3328	0,2982	0,1198	0,0987
1974	0,0989	0,1568	0,0000	0,0000	0,0000	0,0000	0,0000
1975	0,2491	0,3514	0,0276	0,1848	0,2920	0,1187	0,3115
1976	0,1855	0,2716	0,0010	0,0578	0,3055	0,1176	0,0428
1977	0,1046	0,1649	0,0000	0,0000	0,0000	0,0000	0,0000
1978	0,0540	0,0892	0,0000	0,0000	0,0000	0,0000	0,0000
1979	0,2173	0,3122	0,0062	0,0777	0,2552	0,1678	0,1952
1980	0,2500	0,3324	0,0239	0,1479	0,2241	0,1464	0,1875
1981	0,2875	0,3743	0,0622	0,2434	0,4228	0,0866	0,2328
1982	0,3694	0,4635	0,1380	0,3050	0,1850	0,1216	0,3322
1983	1,0000	1,0000	1,0000	0,9880	0,5569	0,0646	0,7013
1984	0,2989	0,3946	0,0478	0,1981	0,1022	0,3746	0,1874
1985	0,0437	0,0608	0,0000	0,0000	0,0000	0,0000	0,0000
1986	0,1501	0,2122	0,0000	0,0000	0,0000	0,0000	0,0000
1987	0,3483	0,4405	0,1057	0,2503	0,1730	0,2399	0,4402
1988	0,2595	0,3432	0,0270	0,1674	0,4987	0,0816	0,1509
1989	0,2875	0,3743	0,0563	0,1872	0,1001	0,3302	0,5874
1990	0,3548	0,4473	0,0747	0,2163	0,4615	0,0925	0,4642
1991	0,1280	0,1824	0,0000	0,0000	0,0000	0,0000	0,0000
1992	0,7378	0,7946	0,3991	0,2784	0,2706	0,1419	1,0000
1993	0,4834	0,5716	0,1826	0,3060	0,3420	0,1198	0,4866
1994	0,1726	0,2405	0,0000	0,0000	0,0000	0,0000	0,0000
1995	0,4109	0,5041	0,1281	0,2615	0,2981	0,1251	0,4122
1996	0,2150	0,2932	0,0067	0,1003	0,4079	0,1028	0,1310
1997	0,3982	0,4919	0,2152	0,6358	0,4444	0,0858	0,1035
1998	0,4759	0,5649	0,3438	0,6857	0,2135	0,1472	0,4569
1999	0,2757	0,3622	0,0475	0,1876	0,4268	0,0941	0,3555
2000	0,3740	0,4689	0,0862	0,1997	0,0908	0,4140	0,5881
2001	0,3068	0,3973	0,0532	0,2045	0,1825	0,1925	0,3513
2002	0,1295	0,1865	0,0000	0,0000	0,0000	0,0000	0,0000
2003	0,1429	0,2054	0,0000	0,0000	0,0000	0,0000	0,0000
2004	0,2033	0,2811	0,0025	0,0595	0,3059	0,1374	0,1443
2005	0,2896	0,3784	0,0790	0,4748	0,1160	0,3355	0,2676
2006	0,0028	0,0014	0,0000	0,0000	0,0000	0,0000	0,0000
2007	0,2824	0,3716	0,0320	0,1545	0,4897	0,0854	0,2219
2008	0,1967	0,2743	0,0022	0,0724	0,4931	0,0786	0,0841
2009	0,3230	0,4149	0,1193	0,5111	0,1922	0,1909	0,1304
2010	0,4440	0,5378	0,1618	0,2848	0,0327	1,0000	0,5732
2011	0,3967	0,4905	0,1770	0,6192	0,9758	0,0426	0,3287
2012	0,2714	0,3595	0,0429	0,1899	0,3471	0,1125	0,2758
2014	0,6140	0,6905	0,2960	0,2795	0,3856	0,1063	0,6866
2015	0,2905	0,3797	0,0673	0,3980	0,3117	0,1165	0,4210

FONTE: STEFFEN (2017).

ANEXO 8 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORÍTMO *FUZZY C-MEANS*.

AGRUPAMENTO DAS CHEIAS						
GRUPO 1 SEVERIDADE BAIXA		GRUPO 2 SEVERIDADE MÉDIA		GRUPO 3 SEVERIDADE ALTA		DESASTRE OU CATASTROFE
1931	1962	1937	2004	1932	2000	1935
1933	1963	1939	2007	1938	2001	1957
1934	1964	1948	2008	1946	2005	1983
1936	1967	1953	2011	1947	2009	1992
1940	1968	1961	2012	1950	2010	1998
1941	1969	1965		1954	2013	2014
1942	1974	1966		1955	2015	
1943	1977	1970		1971		
1944	1978	1973		1972		
1945	1985	1975		1982		
1949	1986	1976		1984		
1951	1991	1979		1987		
1952	1994	1980		1989		
1956	2002	1981		1990		
1958	2003	1988		1993		
1959	2006	1996		1995		
1960		1999		1997		

FONTE: STEFFEN (2017).

ANEXO 9 – VALORES NORMALIZADOS DOS CENTROIDES DOS GRUPOS ENCONTRADOS PELO MÉTODO *FUZZY C -MEANS*.

	I_1	I_2	I_3	I_4	I_5	I_6	I_7
Centroide 1	0,1018	0,1552	0,0001	0,001	0,0026	0,0046	0,0025
Centroide 2	0,2375	0,3273	0,0259	0,1447	0,4041	0,1166	0,1737
Centroide 3	0,3481	0,4471	0,1032	0,2905	0,2048	0,2647	0,3952
Centroide 4	0,6310	0,7059	0,4628	0,6102	0,3271	0,1730	0,5655

FONTE: STEFFEN (2017).

APÊNDICE 1 – MATRIZ DE INDICADORES DE CHEIA NORMALIZADOS OBTIDOS ATRAVÉS DO MÉTODO DO DESVIO PADRÃO.

Ano	I_1	I_2	I_3	I_4	I_5	I_6	I_7
1931	0,340	0,279	0,485	0,621	0,391	0,977	0,258
1932	0,159	0,282	0,353	0,350	0,212	0,610	0,949
1933	1,152	1,318	0,495	0,786	0,899	0,760	0,897
1934	1,173	1,349	0,495	0,786	0,899	0,760	0,897
1935	2,046	2,042	1,818	1,509	0,153	0,250	1,301
1936	0,618	0,598	0,495	0,786	0,899	0,760	0,897
1937	0,411	0,363	0,495	0,748	0,119	0,283	0,395
1938	0,883	1,026	0,304	0,356	0,379	1,259	0,805
1939	0,007	0,123	0,304	0,069	1,314	0,207	0,222
1940	1,505	1,857	0,495	0,786	0,899	0,760	0,897
1941	0,584	0,552	0,495	0,786	0,899	0,760	0,897
1942	0,676	0,681	0,495	0,786	0,899	0,760	0,897
1943	1,065	1,197	0,495	0,786	0,899	0,760	0,897
1944	0,994	1,098	0,495	0,786	0,899	0,760	0,897
1945	0,649	0,643	0,495	0,786	0,899	0,760	0,897
1946	0,819	0,957	0,509	1,349	1,302	0,189	0,971
1947	0,086	0,206	0,164	0,234	0,293	0,591	0,156
1948	0,354	0,294	0,485	0,536	1,605	0,262	0,284
1949	1,194	1,379	0,495	0,786	0,899	0,760	0,897
1950	0,134	0,036	0,432	0,377	0,275	1,067	0,283
1951	0,681	0,689	0,495	0,786	0,899	0,760	0,897
1952	0,814	0,856	0,495	0,786	0,899	0,760	0,897
1953	0,213	0,127	0,455	0,440	0,396	0,144	0,071
1954	0,909	1,048	0,331	0,277	0,764	0,036	1,569
1955	0,359	0,495	0,388	0,590	0,250	0,663	0,335
1956	0,628	0,613	0,495	0,786	0,899	0,760	0,897
1957	1,267	1,375	3,364	3,935	0,480	2,141	0,253
1958	0,897	0,962	0,495	0,786	0,899	0,760	0,897
1959	1,194	1,379	0,495	0,786	0,899	0,760	0,897
1960	0,814	0,856	0,495	0,786	0,899	0,760	0,897
1961	0,094	0,214	0,271	0,026	3,434	0,489	0,095
1962	0,934	1,015	0,495	0,786	0,899	0,760	0,897
1963	0,873	0,931	0,495	0,786	0,899	0,760	0,897
1964	1,049	1,174	0,495	0,786	0,899	0,760	0,897
1965	0,315	0,249	0,464	0,334	0,108	0,702	0,435
1966	0,411	0,363	0,494	0,674	1,833	0,376	0,692
1967	0,660	0,658	0,495	0,786	0,899	0,760	0,897
1968	1,272	1,493	0,495	0,786	0,899	0,760	0,897

Ano	I_1	I_2	I_3	I_4	I_5	I_6	I_7
1969	0,560	0,522	0,495	0,786	0,899	0,760	0,897
1970	0,285	0,211	0,437	0,191	1,227	0,183	0,366
1971	0,943	1,079	0,655	0,797	0,505	0,141	1,112
1972	0,159	0,282	0,007	0,680	0,608	3,049	1,723
1973	0,065	0,184	0,179	0,786	0,488	0,000	0,450
1974	0,907	0,977	0,495	0,786	0,899	0,760	0,897
1975	0,001	0,115	0,303	0,085	0,459	0,008	0,513
1976	0,385	0,332	0,488	0,515	0,522	0,014	0,704
1977	0,873	0,931	0,495	0,786	0,899	0,760	0,897
1978	1,178	1,356	0,495	0,786	0,899	0,760	0,897
1979	0,193	0,105	0,452	0,419	0,288	0,304	0,014
1980	0,004	0,009	0,328	0,085	0,143	0,169	0,047
1981	0,231	0,244	0,061	0,361	1,067	0,211	0,156
1982	0,725	0,745	0,469	0,654	0,039	0,011	0,607
1983	4,532	3,756	6,493	3,882	1,691	0,350	2,278
1984	0,300	0,358	0,162	0,149	0,424	1,616	0,050
1985	1,240	1,516	0,495	0,786	0,899	0,760	0,897
1986	0,598	0,666	0,495	0,786	0,899	0,760	0,897
1987	0,599	0,616	0,243	0,399	0,095	0,761	1,097
1988	0,062	0,070	0,307	0,006	1,420	0,242	0,215
1989	0,231	0,244	0,102	0,101	0,434	1,335	1,763
1990	0,637	0,654	0,026	0,234	1,247	0,174	1,205
1991	0,732	0,833	0,495	0,786	0,899	0,760	0,897
1992	2,950	2,603	2,294	0,531	0,359	0,140	3,631
1993	1,413	1,352	0,780	0,659	0,691	0,000	1,306
1994	0,463	0,507	0,495	0,786	0,899	0,760	0,897
1995	0,975	0,972	0,400	0,452	0,487	0,033	0,971
1996	0,206	0,211	0,449	0,313	0,998	0,108	0,304
1997	0,899	0,904	1,008	2,220	1,168	0,216	0,428
1998	1,368	1,314	1,908	2,454	0,094	0,174	1,172
1999	0,159	0,176	0,164	0,101	1,086	0,164	0,712
2000	0,753	0,775	0,107	0,160	0,477	1,866	1,766
2001	0,348	0,373	0,124	0,181	0,050	0,461	0,693
2002	0,723	0,810	0,495	0,786	0,899	0,760	0,897
2003	0,642	0,704	0,495	0,786	0,899	0,760	0,897
2004	0,277	0,279	0,478	0,504	0,524	0,112	0,244
2005	0,244	0,267	0,056	1,455	0,359	1,368	0,314
2006	1,487	1,849	0,495	0,786	0,899	0,760	0,897
2007	0,204	0,229	0,272	0,058	1,379	0,219	0,108
2008	0,317	0,317	0,480	0,446	1,394	0,262	0,517
2009	0,445	0,472	0,338	1,631	0,005	0,451	0,306
2010	1,176	1,162	0,636	0,558	0,747	5,583	1,699

Ano	I_1	I_2	I_3	I_4	I_5	I_6	I_7
2011	0,890	0,897	0,742	2,140	3,640	0,489	0,592
2012	0,135	0,161	0,195	0,112	0,715	0,047	0,353
2013	0,895	0,912	0,319	0,250	0,455	1,659	1,481
2014	2,202	2,019	1,573	0,537	0,894	0,086	2,211
2015	0,249	0,275	0,025	1,094	0,551	0,021	1,009

APÊNDICE 2 – MATRIZ DE INDICADORES DE CHEIA NORMALIZADOS OBTIDOS ATRAVÉS DO MÉTODO DA ESCALA DECIMAL.

Ano	I_1	I_2	I_3	I_4	I_5	I_6	I_7
1931	0,1453	0,051	0,0014	0,3100	0,6389	0,1399	0,0267
1932	0,1833	0,0584	0,0198	0,8200	0,8634	0,1103	0,0771
1933	0,0836	0,0373	0,0000	0,0000	0,0000	0,0000	0,0000
1934	0,0820	0,0369	0,0000	0,0000	0,0000	0,0000	0,0000
1935	0,3267	0,0816	0,3207	4,3200	13226	0,0813	0,0918
1936	0,1242	0,0468	0,0000	0,0000	0,0000	0,0000	0,0000
1937	0,1399	0,0499	0,0000	0,0700	1,2797	0,0840	0,0210
1938	0,2383	0,0682	0,1108	2,1500	0,6538	0,1626	0,0711
1939	0,1717	0,0563	0,0265	1,6100	2,7811	0,0445	0,0282
1940	0,0568	0,0302	0,0000	0,0000	0,0000	0,0000	0,0000
1941	0,1268	0,0474	0,0000	0,0000	0,0000	0,0000	0,0000
1942	0,1198	0,0457	0,0000	0,0000	0,0000	0,0000	0,0000
1943	0,0902	0,0389	0,0000	0,0000	0,0000	0,0000	0,0000
1944	0,0956	0,0402	0,0000	0,0000	0,0000	0,0000	0,0000
1945	0,1218	0,0462	0,0000	0,0000	0,0000	0,0000	0,0000
1946	0,2334	0,0673	0,1392	4,0200	2,7654	0,0460	0,0780
1947	0,1777	0,0574	0,0460	1,9200	0,7622	0,1088	0,0440
1948	0,1443	0,0508	0,0015	0,4700	31465	0,0401	0,0256
1949	0,0804	0,0365	0,0000	0,0000	0,0000	0,0000	0,0000
1950	0,1610	0,0542	0,0088	0,7700	0,7844	0,1471	0,0493
1951	0,1194	0,0456	0,0000	0,0000	0,0000	0,0000	0,0000
1952	0,1093	0,0434	0,0000	0,0000	0,0000	0,0000	0,0000
1953	0,1550	0,0530	0,0056	0,6500	1,6279	0,0728	0,0345
1954	0,2403	0,0685	0,1146	2,0000	2,0895	0,0583	0,1030
1955	0,1985	0,0612	0,1225	2,5900	0,8159	0,1146	0,0235
1956	0,1234	0,0466	0,0000	0,0000	0,0000	0,0000	0,0000
1957	0,2675	0,0728	0,5351	8,8900	0,5267	0,2336	0,0269
1958	0,1030	0,0420	0,0000	0,0000	0,0000	0,0000	0,0000
1959	0,0804	0,0365	0,0000	0,0000	0,0000	0,0000	0,0000
1960	0,1093	0,0434	0,0000	0,0000	0,0000	0,0000	0,0000
1961	0,1783	0,0575	0,0311	1,4300	5,4445	0,0218	0,0335
1962	0,1002	0,0413	0,0000	0,0000	0,0000	0,0000	0,0000

Ano	I_1	I_2	I_3	I_4	I_5	I_6	I_7
1963	0,1048	0,0424	0,0000	0,0000	0,0000	0,0000	0,0000
1964	0,0914	0,0392	0,0000	0,0000	0,0000	0,0000	0,0000
1965	0,1472	0,0514	0,0043	0,8500	0,9939	0,1177	0,0193
1966	0,1399	0,0499	0,0002	0,2100	3,4333	0,0309	0,0086
1967	0,1210	0,0460	0,0000	0,0000	0,0000	0,0000	0,0000
1968	0,0745	0,035	0,0000	0,0000	0,0000	0,0000	0,0000
1969	0,1286	0,0478	0,0000	0,0000	0,0000	0,0000	0,0000
1970	0,1495	0,0519	0,0081	1,1200	2,6723	0,0465	0,0222
1971	0,2429	0,0689	0,1595	2,9800	1,7643	0,0726	0,0839
1972	0,1833	0,0584	0,0697	2,7600	0,3661	0,3067	0,1094
1973	0,1761	0,0571	0,0439	2,9600	1,7428	0,0612	0,0187
1974	0,1022	0,0418	0,0000	0,0000	0,0000	0,0000	0,0000
1975	0,1711	0,0562	0,0267	1,6400	1,7066	0,0606	0,0589
1976	0,1419	0,0503	0,0010	0,5100	1,7857	0,0601	0,0081
1977	0,1048	0,0424	0,0000	0,0000	0,0000	0,0000	0,0000
1978	0,0816	0,0368	0,0000	0,0000	0,0000	0,0000	0,0000
1979	0,1565	0,0533	0,0060	0,6900	1,4914	0,0857	0,0369
1980	0,1715	0,0548	0,0232	1,3200	1,3096	0,0748	0,0355
1981	0,1887	0,0579	0,0603	2,1600	2,4709	0,0442	0,0440
1982	0,2263	0,0645	0,1337	2,7100	1,0814	0,0621	0,0628
1983	0,5157	0,1042	0,9689	8,7900	3,2547	0,033	0,1326
1984	0,1940	0,0594	0,0463	1,7600	0,5973	0,1913	0,0354
1985	0,0769	0,0347	0,0000	0,0000	0,0000	0,0000	0,0000
1986	0,1257	0,0459	0,0000	0,0000	0,0000	0,0000	0,0000
1987	0,2167	0,0628	0,1024	2,2300	1,0109	0,1225	0,0833
1988	0,1759	0,0556	0,0261	1,4900	2,9144	0,0417	0,0285
1989	0,1887	0,0579	0,0546	1,6700	0,5851	0,1687	0,1111
1990	0,2196	0,0633	0,0723	1,9200	2,6974	0,0472	0,0878
1991	0,1155	0,0437	0,0000	0,0000	0,0000	0,0000	0,0000
1992	0,3954	0,0890	0,3867	2,4800	1,5813	0,0725	0,1891
1993	0,2786	0,0725	0,1769	2,7200	1,9988	0,0612	0,0920
1994	0,1360	0,0480	0,0000	0,0000	0,0000	0,0000	0,0000
1995	0,2453	0,0675	0,1241	2,3300	1,7426	0,0639	0,0780
1996	0,1555	0,0519	0,0065	0,8900	2,3842	0,0525	0,0248
1997	0,2395	0,0666	0,2085	5,6600	2,5974	0,0438	0,0196

Ano	I_1	I_2	I_3	I_4	I_5	I_6	I_7
1998	0,2752	0,0720	0,3332	6,100	1,2477	0,0752	0,0864
1999	0,1833	0,0570	0,0466	1,6700	2,4947	0,048	0,0672
2000	0,2284	0,0649	0,0835	1,7800	0,5306	0,2115	0,1112
2001	0,1976	0,0596	0,0515	1,8200	1,0667	0,0983	0,0664
2002	0,1162	0,0440	0,0000	0,0000	0,0000	0,0000	0,0000
2003	0,1224	0,0454	0,0000	0,0000	0,0000	0,0000	0,0000
2004	0,1501	0,0510	0,0024	0,5300	1,7879	0,0702	0,0273
2005	0,1897	0,0582	0,0765	4,2200	0,6782	0,1714	0,0506
2006	0,0581	0,0303	0,0000	0,0000	0,0000	0,0000	0,0000
2007	0,1867	0,0577	0,0310	1,3700	2,8621	0,0436	0,0420
2008	0,1471	0,0505	0,0021	0,6400	2,882	0,0401	0,0159
2009	0,2050	0,0609	0,1156	4,5500	1,1231	0,0975	0,0247
2010	0,2606	0,0700	0,1568	2,5300	0,1912	0,5108	0,1084
2011	0,2388	0,0665	0,1715	5,5100	5,7032	0,0218	0,0622
2012	0,1814	0,0568	0,0416	1,6900	2,0286	0,0574	0,0522
2013	0,2392	0,0667	0,1129	1,9500	0,5581	0,1948	0,0993
2014	0,3386	0,0813	0,2868	2,4900	2,2534	0,0543	0,1298
2015	0,1901	0,0583	0,0652	3,5400	1,822	0,0595	0,0796

APÊNDICE 3 – MATRIZ DE INDICADORES DE CHEIA NORMALIZADOS OBTIDOS ATRAVÉS DO MÉTODO DO VALOR MÁXIMO DOS ELEMENTOS.

Ano	I_1	I_2	I_3	I_4	I_5	I_6	I_7
1931	0,2818	0,4894	0,0014	0,0349	0,1120	0,2739	0,1412
1932	0,3554	0,5605	0,0204	0,0922	0,1514	0,2159	0,4077
1933	0,1621	0,3580	0,0000	0,0000	0,0000	0,0000	0,0000
1934	0,1590	0,3541	0,0000	0,0000	0,0000	0,0000	0,0000
1935	0,6335	0,7831	0,3310	0,4859	0,2319	0,1592	0,4855
1936	0,2408	0,4491	0,0000	0,0000	0,0000	0,0000	0,0000
1937	0,2713	0,4789	0,0000	0,0079	0,2244	0,1644	0,1111
1938	0,4621	0,6545	0,1144	0,2418	0,1146	0,3183	0,3760
1939	0,3329	0,5403	0,0274	0,1811	0,4876	0,0871	0,1491
1940	0,1101	0,2898	0,0000	0,0000	0,0000	0,0000	0,0000
1941	0,2459	0,4549	0,0000	0,0000	0,0000	0,0000	0,0000
1942	0,2323	0,4386	0,0000	0,0000	0,0000	0,0000	0,0000
1943	0,1749	0,3733	0,0000	0,0000	0,0000	0,0000	0,0000
1944	0,1854	0,3858	0,0000	0,0000	0,0000	0,0000	0,0000
1945	0,2362	0,4434	0,0000	0,0000	0,0000	0,0000	0,0000
1946	0,4526	0,6459	0,1437	0,4522	0,4849	0,0901	0,4125
1947	0,3446	0,5509	0,0475	0,2160	0,1336	0,2130	0,2327
1948	0,2798	0,4875	0,0015	0,0529	0,5517	0,0785	0,1354
1949	0,1559	0,3503	0,0000	0,0000	0,0000	0,0000	0,0000
1950	0,3122	0,5202	0,0091	0,0866	0,1375	0,2880	0,2607
1951	0,2315	0,4376	0,0000	0,0000	0,0000	0,0000	0,0000
1952	0,2119	0,4165	0,0000	0,0000	0,0000	0,0000	0,0000
1953	0,3006	0,5086	0,0058	0,0731	0,2854	0,1425	0,1824
1954	0,4660	0,6574	0,1183	0,2250	0,3664	0,1141	0,5447
1955	0,3849	0,5873	0,1264	0,2913	0,1431	0,2244	0,1243
1956	0,2393	0,4472	0,0000	0,0000	0,0000	0,0000	0,0000
1957	0,5187	0,6987	0,5523	1,0000	0,0924	0,4573	0,1423
1958	0,1997	0,4031	0,0000	0,0000	0,0000	0,0000	0,0000
1959	0,1559	0,3503	0,0000	0,0000	0,0000	0,0000	0,0000
1960	0,2119	0,4165	0,0000	0,0000	0,0000	0,0000	0,0000
1961	0,3457	0,5518	0,0321	0,1609	0,9546	0,0427	0,1772
1962	0,1943	0,3964	0,0000	0,0000	0,0000	0,0000	0,0000
1963	0,2032	0,4069	0,0000	0,0000	0,0000	0,0000	0,0000
1964	0,1772	0,3762	0,0000	0,0000	0,0000	0,0000	0,0000
1965	0,2854	0,4933	0,0044	0,0956	0,1743	0,2304	0,1021
1966	0,2713	0,4789	0,0002	0,0236	0,6020	0,0605	0,0455
1967	0,2346	0,4415	0,0000	0,0000	0,0000	0,0000	0,0000
1968	0,1445	0,3359	0,0000	0,0000	0,0000	0,0000	0,0000

Ano	I_1	I_2	I_3	I_4	I_5	I_6	I_7
1969	0,2494	0,4587	0,0000	0,0000	0,0000	0,0000	0,0000
1970	0,2899	0,4981	0,0084	0,1260	0,4686	0,0910	0,1174
1971	0,4710	0,6612	0,1646	0,3352	0,3094	0,1421	0,4437
1972	0,3554	0,5605	0,0719	0,3105	0,0642	0,6004	0,5785
1973	0,3415	0,5480	0,0453	0,3330	0,3056	0,1198	0,0989
1974	0,1982	0,4012	0,0000	0,0000	0,0000	0,0000	0,0000
1975	0,3318	0,5393	0,0276	0,1845	0,2992	0,1186	0,3115
1976	0,2752	0,4827	0,0010	0,0574	0,3131	0,1177	0,0428
1977	0,2032	0,4069	0,0000	0,0000	0,0000	0,0000	0,0000
1978	0,1582	0,3532	0,0000	0,0000	0,0000	0,0000	0,0000
1979	0,3035	0,5115	0,0062	0,0776	0,2615	0,1678	0,1951
1980	0,3326	0,5259	0,0239	0,1485	0,2296	0,1464	0,1877
1981	0,3659	0,5557	0,0622	0,2430	0,4332	0,0865	0,2327
1982	0,4388	0,6190	0,1380	0,3048	0,1896	0,1216	0,3321
1983	1,0000	1,0000	1,0000	0,9888	0,5707	0,0646	0,7012
1984	0,3762	0,5701	0,0478	0,1980	0,1047	0,3745	0,1872
1985	0,1491	0,3330	0,0000	0,0000	0,0000	0,0000	0,0000
1986	0,2437	0,4405	0,0000	0,0000	0,0000	0,0000	0,0000
1987	0,4202	0,6027	0,1057	0,2508	0,1773	0,2398	0,4405
1988	0,3411	0,5336	0,0269	0,1676	0,5110	0,0816	0,1507
1989	0,3659	0,5557	0,0564	0,1879	0,1026	0,3303	0,5875
1990	0,4258	0,6075	0,0746	0,2160	0,4730	0,0924	0,4643
1991	0,2240	0,4194	0,0000	0,0000	0,0000	0,0000	0,0000
1992	0,7667	0,8541	0,3991	0,2790	0,2773	0,1419	1,0000
1993	0,5402	0,6958	0,1826	0,3060	0,3505	0,1198	0,4865
1994	0,2637	0,4607	0,0000	0,0000	0,0000	0,0000	0,0000
1995	0,4757	0,6478	0,1281	0,2621	0,3054	0,1251	0,4125
1996	0,3015	0,4981	0,0067	0,1001	0,4180	0,1028	0,1311
1997	0,4644	0,6392	0,2152	0,6367	0,4554	0,0857	0,1036
1998	0,5336	0,6910	0,3439	0,6862	0,2188	0,1472	0,4569
1999	0,3554	0,5470	0,0475	0,1879	0,4374	0,0940	0,3554
2000	0,4429	0,6228	0,0862	0,2002	0,0930	0,4141	0,5880
2001	0,3832	0,5720	0,0532	0,2047	0,1870	0,1924	0,3511
2002	0,2253	0,4223	0,0000	0,0000	0,0000	0,0000	0,0000
2003	0,2373	0,4357	0,0000	0,0000	0,0000	0,0000	0,0000
2004	0,2911	0,4894	0,0025	0,0596	0,3135	0,1374	0,1444
2005	0,3678	0,5585	0,0790	0,4747	0,1189	0,3356	0,2676
2006	0,1127	0,2908	0,0000	0,0000	0,0000	0,0000	0,0000
2007	0,3620	0,5537	0,0320	0,1541	0,5018	0,0854	0,2221
2008	0,2852	0,4846	0,0022	0,0720	0,5053	0,0785	0,0841
2009	0,3975	0,5845	0,1193	0,5118	0,1969	0,1909	0,1306
2010	0,5053	0,6718	0,1618	0,2846	0,0335	1,0000	0,5732

Ano	I_1	I_2	I_3	I_4	I_5	I_6	I_7
2011	0,4631	0,6382	0,1770	0,6198	1,0000	0,0427	0,3289
2012	0,3518	0,5451	0,0429	0,1901	0,3557	0,1124	0,2760
2013	0,4638	0,6401	0,1165	0,2193	0,0978	0,3814	0,5251
2014	0,6566	0,7802	0,2960	0,2801	0,3951	0,1063	0,6864
2015	0,3686	0,5595	0,0673	0,3982	0,3195	0,1165	0,4209

APÊNDICE 4 – MATRIZ DE INDICADORES DE CHEIA COM DADOS DE I_5 E I_7 ALTERADOS.

Ano	I_1	I_2	I_3	I_4	I_5	I_6	I_7
1931	1453	510	0,0140	3,100	52,5230	139,90	170,5
1932	1833	584	0,1980	8,200	50,2780	110,30	120,1
1933	836	373	0,0000	0,0000	0,0000	0,0000	0,0000
1934	820,0	369	0,0000	0,0000	0,0000	0,0000	0,0000
1935	3267	816	3,2070	43,20	45,6860	81,30	105,4
1936	1242	468	0,0000	0,0000	0,0000	0,0000	197,2
1937	1399	499	0,0000	0,7000	46,1150	84,00	176,2
1938	2383	682	1,1080	21,50	52,3740	162,60	126,1
1939	1717	563	0,2650	16,10	31,1010	44,50	169
1940	568,0	302	0,0000	0,0000	0,0000	0,0000	0,0000
1941	1268	474	0,0000	0,0000	0,0000	0,0000	0,0000
1942	1198	457	0,0000	0,0000	0,0000	0,0000	0,0000
1943	902	389	0,0000	0,0000	0,0000	0,0000	0,0000
1944	956	402	0,0000	0,0000	0,0000	0,0000	0,0000
1945	1218	462	0,0000	0,0000	0,0000	0,0000	0,0000
1946	2334	673	1,3920	40,20	31,2580	46,00	119,2
1947	1777	574	0,4600	19,20	51,2900	108,80	153,2
1948	1443	508	0,0150	4,700	27,4470	40,10	171,6
1949	804	365	0,0000	0,0000	0,0000	0,0000	0,0000
1950	1610	542	0,0880	7,700	51,0680	147,10	147,9
1951	1194	456	0,0000	0,0000	0,0000	0,0000	0,0000
1952	1093	434	0,0000	0,0000	0,0000	0,0000	0,0000
1953	1550	530	0,0560	6,5000	42,633	72,8000	162,7
1954	2403	685	1,1460	20,0000	38,017	58,3000	94,2
1955	1985	612	1,2250	25,9000	50,753	114,60	173,7
1956	1234	466	0,0000	0,0000	0,0000	0,0000	0,0000
1957	2675	728	5,3510	88,90	53,645	233,60	170,3
1958	1030	420	0,0000	0,0000	0,0000	0,0000	0,0000
1959	804	365	0,0000	0,0000	0,0000	0,0000	0,0000
1960	1093	434	0,0000	0,0000	0,0000	0,0000	0,0000
1961	1783	575	0,3110	14,3000	4,4670	21,8000	163,7
1962	1002	413	0,0000	0,0000	0,0000	0,0000	0,0000
1963	1048	424	0,0000	0,0000	0,0000	0,0000	0,0000
1964	914	392	0,0000	0,0000	0,0000	0,0000	0,0000
1965	1472	514	0,0430	8,5000	48,9730	117,70	177,9
1966	1399	499	0,0020	2,1000	24,5790	30,90	188,6
1967	1210	460	0,0000	0,0000	0,0000	0,0000	0,0000
1968	745	350	0,0000	0,0000	0,0000	0,0000	0,0000
1969	1286	478	0,0000	0,0000	0,0000	0,0000	0,0000
1970	1495	519	0,0810	11,2000	32,1890	46,50	175
1971	2429	689	1,5950	29,8000	41,2690	72,60	113,3

Ano	I_1	I_2	I_3	I_4	I_5	I_6	I_7
1972	1833	584	0,6970	27,6000	55,251	306,70	87,8
1973	1761	571	0,4390	29,6000	41,484	61,20	178,5
1974	1022	418	0,0000	0,0000	0,0000	0,0000	0,0000
1975	1711	562	0,2670	16,4000	41,846	60,60	138,3
1976	1419	503	0,0100	5,1000	41,055	60,10	189,1
1977	1048	424	0,0000	0,0000	0,0000	0,0000	0,0000
1978	816	368	0,0000	0,0000	0,0000	0,0000	0,0000
1979	1565	533	0,0600	6,9000	43,998	85,70	160,3
1980	1715	548	0,2320	13,2000	45,816	74,80	161,7
1981	1887	579	0,6030	21,6000	34,203	44,20	153,2
1982	2263	645	1,3370	27,1000	48,098	62,10	134,4
1983	5157	1042	9,6890	87,9000	26,365	33,00	64,6
1984	1940	594	0,4630	17,6000	52,939	191,30	161,8
1985	769	347	0,0000	0,0000	0,0000	0,0000	0,0000
1986	1257	459	0,0000	0,0000	0,0000	0,0000	0,0000
1987	2167	628	1,0240	22,3000	48,8030	122,50	113,9
1988	1759	556	0,2610	14,9000	29,768	41,70	168,7
1989	1887	579	0,5460	16,7000	53,061	168,70	86,1
1990	2196	633	0,7230	19,2000	31,938	47,20	109,4
1991	1155	437	0,0000	0,0000	0,0000	0,0000	0,0000
1992	3954	890	3,8670	24,8000	43,099	72,50	8,1
1993	2786	725	1,7690	27,2000	38,924	61,20	105,2
1994	1360	480	0,0000	0,0000	0,0000	0,0000	0,0000
1995	2453	675	1,2410	23,3000	41,492	63,90	119,2
1996	1555	519	0,0650	8,9000	35,070	52,50	172,4
1997	2395	666	2,0850	56,6000	32,938	43,80	177,6
1998	2752	720	3,3320	61,0000	46,435	75,20	110,8
1999	1833	570	0,4600	16,7000	33,965	48,00	130
2000	2284	649	0,8350	17,8000	53,606	211,5	86
2001	1976	596	0,5150	18,2000	48,245	98,30	130,8
2002	1162	440	0,0000	0,0000	0,0000	0,0000	0,0000
2003	1224	454	0,0000	0,0000	0,0000	0,0000	0,0000
2004	1501	510	0,0240	5,3000	41,033	70,20	169,9
2005	1897	582	0,7650	42,20	52,130	171,40	146,6
2006	581	303	0,0000	0,0000	0,0000	0,0000	0,0000
2007	1867	577	0,3100	13,70	30,291	43,60	155,2
2008	1471	505	0,0210	6,400	30,092	40,10	181,3
2009	2050	609	1,1560	45,50	47,6810	97,50	172,5
2010	2606	700	1,5680	25,30	57,00	510,80	88,8
2011	2388	665	1,7150	55,10	1,8800	21,80	135
2012	1814	568	0,4160	16,90	38,626	57,40	145
2013	2392	667	1,1290	19,50	53,332	194,80	97,9
2014	3386	813	2,8680	24,90	36,378	54,30	67,4
2015	1901	583	0,6520	35,40	40,692	59,50	117,6

APÊNDICE 5 – MATRIZ DE INDICADORES DE CHEIA NORMALIZADOS OBTIDOS ATRAVÉS DO MÉTODO DE NORMALIZAÇÃO LINEAR; VALORES DE I_5 E I_7 ALTERADOS.

Ano	I_1	I_2	I_3	I_4	I_5	I_6	I_7
1931	0,1929	0,2811	0,0014	0,0347	0,9214	0,274	0,9015
1932	0,2757	0,3811	0,0204	0,0917	0,8820	0,216	0,6349
1933	0,0585	0,0959	0,0000	0,0000	0,0000	0,0000	0,0000
1934	0,0549	0,0905	0,0000	0,0000	0,0000	0,0000	0,0000
1935	0,5881	0,6946	0,331	0,4862	0,8015	0,1592	0,0000
1936	0,1469	0,2243	0,0000	0,0000	0,0000	0,0000	0,0000
1937	0,1811	0,2662	0,0000	0,0079	0,8090	0,1645	0,9317
1938	0,3955	0,5135	0,1143	0,2413	0,9188	0,3183	0,6666
1939	0,2503	0,3527	0,0274	0,1808	0,5456	0,087	0,8936
1940	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
1941	0,1525	0,2324	0,0000	0,0000	0,0000	0,0000	0,0000
1942	0,1373	0,2095	0,0000	0,0000	0,0000	0,0000	0,0000
1943	0,0728	0,1176	0,0000	0,0000	0,0000	0,0000	0,0000
1944	0,0846	0,1351	0,0000	0,0000	0,0000	0,0000	0,0000
1945	0,1417	0,2162	0,0000	0,0000	0,0000	0,0000	0,0000
1946	0,3848	0,5014	0,1437	0,4518	0,5483	0,0901	0,6301
1947	0,2635	0,3676	0,0475	0,2162	0,8998	0,213	0,8100
1948	0,1908	0,2784	0,0016	0,0533	0,4815	0,0785	0,9074
1949	0,0514	0,0851	0,0000	0,0000	0,0000	0,0000	0,0000
1950	0,2271	0,3243	0,0091	0,0863	0,8959	0,288	0,7820
1951	0,1364	0,2081	0,0000	0,0000	0,0000	0,0000	0,0000
1952	0,1144	0,1784	0,0000	0,0000	0,0000	0,0000	0,0000
1953	0,214	0,3181	0,0058	0,0725	0,7479	0,1425	0,8603
1954	0,3998	0,5176	0,1183	0,2245	0,6669	0,1142	0,4978
1955	0,3088	0,4189	0,1265	0,4036	0,8904	0,2243	0,9185
1956	0,1451	0,2216	0,0000	0,0000	0,0000	0,0000	0,0000
1957	0,4591	0,5757	0,5523	1,0000	0,9411	0,4574	0,9005
1958	0,1007	0,1595	0,0000	0,0000	0,0000	0,0000	0,0000
1959	0,0514	0,0851	0,0000	0,0000	0,0000	0,0000	0,0000
1960	0,1144	0,1784	0,0000	0,0000	0,0000	0,0000	0,0000
1961	0,2647	0,3689	0,0321	0,1613	0,0783	0,0427	0,8656
1962	0,0946	0,15	0,0000	0,0000	0,0000	0,0000	0,0000
1963	0,1046	0,1649	0,0000	0,0000	0,0000	0,0000	0,0000
1964	0,0755	0,1216	0,0000	0,0000	0,0000	0,0000	0,0000
1965	0,197	0,2865	0,0045	0,0951	0,8591	0,2304	0,9407
1966	0,1811	0,2662	0,002	0,032	0,4312	0,0604	0,9973
1967	0,1399	0,2135	0,0000	0,0000	0,0000	0,0000	0,0000
1968	0,0386	0,0649	0,0000	0,0000	0,0000	0,0000	0,0000
1969	0,1565	0,2378	0,0000	0,0000	0,0000	0,0000	0,0000
1970	0,2021	0,2932	0,0083	0,1254	0,5647	0,091	0,9253

Ano	I_1	I_2	I_3	I_4	I_5	I_6	I_7
1971	0,4055	0,523	0,1646	0,3347	0,7240	0,1421	0,5989
1972	0,2757	0,3811	0,0719	0,3107	0,9693	0,6005	0,4640
1973	0,2599	0,3635	0,0453	0,3328	0,7277	0,1198	0,9439
1974	0,0989	0,1568	0,0000	0,0000	0,0000	0,0000	0,0000
1975	0,2491	0,3514	0,0276	0,1848	0,7341	0,1187	0,7312
1976	0,1855	0,2716	0,001	0,0578	0,7202	0,1176	1,0000
1977	0,1046	0,1649	0,0000	0,0000	0,0000	0,0000	0,0000
1978	0,054	0,0892	0,0000	0,0000	0,0000	0,0000	0,0000
1979	0,2173	0,3122	0,0062	0,0777	0,7718	0,1678	0,8476
1980	0,25	0,3324	0,0239	0,1479	0,8037	0,1464	0,8550
1981	0,2875	0,3743	0,0622	0,2434	0,6005	0,0866	0,8100
1982	0,3694	0,4635	0,138	0,305	0,8438	0,1216	0,7105
1983	1,0000	1,0000	1,0000	0,988	0,4625	0,0646	0,3412
1984	0,2989	0,3946	0,0478	0,1981	0,9287	0,3746	0,8555
1985	0,0437	0,0608	0,0000	0,0000	0,0000	0,0000	0,0000
1986	0,1501	0,2122	0,0000	0,0000	0,0000	0,0000	0,0000
1987	0,3483	0,4405	0,1057	0,2503	0,8561	0,2399	0,6021
1988	0,2595	0,3432	0,027	0,1674	0,5222	0,0816	0,8920
1989	0,2875	0,3743	0,0563	0,1872	0,9308	0,3302	0,4550
1990	0,3548	0,4473	0,0747	0,2163	0,5603	0,0925	0,5783
1991	0,128	0,1824	0,0000	0,0000	0,0000	0,0000	0,0000
1992	0,7378	0,7946	0,3991	0,2784	0,7561	0,1419	0,0423
1993	0,4834	0,5716	0,1826	0,306	0,6828	0,1198	0,5560
1994	0,1726	0,2405	0,0000	0,0000	0,0000	0,0000	0,0000
1995	0,4109	0,5041	0,1281	0,2615	0,7279	0,1251	0,6301
1996	0,215	0,2932	0,0067	0,1003	0,6152	0,1028	0,9116
1997	0,3982	0,4919	0,2152	0,6358	0,5778	0,0858	0,9391
1998	0,4759	0,5649	0,3438	0,6857	0,8146	0,1472	0,5857
1999	0,2757	0,3622	0,0475	0,1876	0,5958	0,0941	0,6873
2000	0,374	0,4689	0,0862	0,1997	0,9404	0,414	0,4544
2001	0,3068	0,3973	0,0532	0,2045	0,8464	0,1925	0,6915
2002	0,1295	0,1865	0,0000	0,0000	0,0000	0,0000	0,0000
2003	0,1429	0,2054	0,0000	0,0000	0,0000	0,0000	0,0000
2004	0,2033	0,2811	0,0025	0,0595	0,7198	0,1374	0,8984
2005	0,2896	0,3784	0,079	0,4748	0,9145	0,3355	0,7751
2006	0,0028	0,0014	0,0000	0,0000	0,0000	0,0000	0,0000
2007	0,2824	0,3716	0,032	0,1545	0,5314	0,0854	0,8206
2008	0,1967	0,2743	0,0022	0,0724	0,5279	0,0786	0,9587
2009	0,323	0,4149	0,1193	0,5111	0,8365	0,1909	0,9121
2010	0,444	0,5378	0,1618	0,2848	1,0000	1,0000	0,4693
2011	0,3967	0,4905	0,177	0,6192	0,0329	0,0426	0,7137
2012	0,2714	0,3595	0,0429	0,1899	0,6776	0,1125	0,7666
2013	0,3974	0,4932	0,1165	0,2196	0,9356	0,3813	0,5174
2014	0,614	0,6905	0,296	0,2795	0,6382	0,1063	0,3560
2015	0,2905	0,3797	0,0673	0,398	0,7138	0,1165	0,6216

APÊNDICE 6 – MATRIZ DE INDICADORES DE CHEIA NORMALIZADOS OBTIDOS ATRAVÉS DO MÉTODO DO DESVIO PADRÃO; VALORES DE I_5 E I_7 ALTERADOS.

Ano	I_1	I_2	I_3	I_4	I_5	I_6	I_7
1931	-0,3157	-0,2316	-0,4823	-0,6153	1,2362	0,9920	1,1529
1932	0,1870	0,3329	-0,3488	-0,3428	1,1333	0,6222	0,4674
1933	-1,1321	-1,2770	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1934	-1,1533	-1,3075	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1935	2,0845	2,1032	1,8343	1,5267	0,9227	0,2599	-1,1647
1936	-0,5949	-0,5521	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1937	-0,3871	-0,3156	-0,4925	-0,7435	0,9423	0,2937	1,2305
1938	0,9148	1,0807	0,3113	0,3675	1,2294	1,2756	0,5490
1939	0,0335	0,1727	-0,3002	0,0791	0,2538	-0,1997	1,1325
1940	-1,4867	-1,8188	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1941	-0,5605	-0,5063	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1942	-0,6531	-0,6361	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1943	-1,0448	-1,1549	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1944	-0,9733	-1,0557	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1945	-0,6266	-0,5979	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1946	0,8500	1,0120	0,5174	1,3665	0,2610	-0,1809	0,4552
1947	0,1129	0,2566	-0,1587	0,2447	1,1797	0,6035	0,9176
1948	-0,3289	-0,2469	-0,4816	-0,5298	0,0862	-0,2546	1,1679
1949	-1,1745	-1,3381	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1950	-0,1079	0,0124	-0,4286	-0,3695	1,1695	1,0819	0,8455
1951	-0,6584	-0,6437	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1952	-0,7920	-0,8116	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1953	-0,1873	-0,0790	-0,4519	-0,4337	0,7826	0,1538	1,0468
1954	0,9413	1,1036	0,3389	0,2874	0,5709	-0,0273	0,1151
1955	0,3882	0,5466	0,3962	0,6026	1,1550	0,6759	1,1964
1956	-0,6055	-0,5674	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1957	1,3012	1,4317	3,3899	3,9680	1,2877	2,1625	1,1502
1958	-0,8754	-0,9184	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1959	-1,1745	-1,3381	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1960	-0,7920	-0,8116	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1961	0,1209	0,2642	-0,2668	-0,0170	-0,9677	-0,4833	1,0604
1962	-0,9125	-0,9718	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1963	-0,8516	-0,8879	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1964	-1,0289	-1,1320	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1965	-0,2905	-0,2011	-0,4613	-0,3268	1,073	0,7147	1,2536
1966	-0,3871	-0,3156	-0,4910	-0,6687	-0,0453	-0,3696	1,3991
1967	-0,6372	-0,6132	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1968	-1,2525	-1,4525	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1969	-0,5367	-0,4758	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1970	-0,2601	-0,1630	-0,4337	-0,1826	0,3037	-0,1747	1,2141

Ano	I_1	I_2	I_3	I_4	I_5	I_6	I_7
1971	0,9757	1,1341	0,6647	0,8109	0,7201	0,1513	0,3749
1972	0,1870	0,3329	0,0131	0,6934	1,3613	3,0757	0,02812
1973	0,0918	0,2337	-0,1740	0,8002	0,7299	0,0088	1,2617
1974	-0,8860	-0,9337	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1975	0,0256	0,1650	-0,2988	0,0951	0,7465	0,0013	0,7150
1976	-0,3607	-0,2851	-0,4852	-0,5084	0,7103	-0,0048	1,4059
1977	-0,8516	-0,8879	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1978	-1,1586	-1,3152	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1979	-0,1675	-0,0561	-0,4490	-0,4123	0,8452	0,3149	1,0142
1980	0,0300	0,0582	-0,3242	-0,0757	0,9286	0,17878	1,0332
1981	0,2585	0,2948	-0,0550	0,3729	0,3960	-0,2034	0,9176
1982	0,7560	0,7984	0,4775	0,6667	1,0333	0,0201	0,6619
1983	4,5854	3,8277	6,5373	3,9146	0,0365	-0,3433	-0,2874
1984	0,3286	0,4092	-0,1566	0,1592	1,2553	1,6341	1,0346
1985	-1,2208	-1,4754	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1986	-0,5750	-0,6208	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1987	0,6290	0,6687	0,2504	0,4103	1,0656	0,7746	0,3831
1988	0,0891	0,1193	-0,3031	0,0150	0,1926	-0,2347	1,1284
1989	0,2585	0,2948	-0,0963	0,1111	1,2609	1,3518	0,0050
1990	0,6674	0,7068	0,0320	0,2447	0,2921	-0,1660	0,3219
1991	-0,7100	-0,7887	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1992	2,9936	2,6678	2,3131	0,5438	0,8040	0,1500	-1,0559
1993	1,4481	1,4088	0,7909	0,6720	0,6125	0,0088	0,2647
1994	-0,4387	-0,4606	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
1995	1,0074	1,0273	0,4078	0,4637	0,7303	0,0426	0,4552
1996	-0,1807	-0,1630	-0,4453	-0,3054	0,4358	-0,0997	1,1788
1997	0,9307	0,9586	1,0202	2,2426	0,3380	-0,2084	1,2495
1998	1,4031	1,3707	1,9250	2,4776	0,9570	0,1837	0,3409
1999	0,1870	0,2261	-0,1587	0,1111	0,3851	-0,1560	0,6021
2000	0,7838	0,8289	0,1133	0,1699	1,2859	1,8865	0,0036
2001	0,3763	0,4245	-0,1188	0,1913	1,0400	0,4723	0,6129
2002	-0,7007	-0,7658	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
2003	-0,6187	-0,6590	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
2004	-0,2522	-0,2316	-0,4751	-0,4978	0,7093	0,1213	1,1448
2005	0,2717	0,3176	0,0625	1,4733	1,2182	1,3855	0,8278
2006	-1,4695	-1,8112	-0,4925	-0,7809	-1,1725	-0,7556	-1,1647
2007	0,2320	0,2795	-0,2676	-0,0490	0,2166	-0,2109	0,9448
2008	-0,2919	-0,2698	-0,4772	-0,4390	0,2075	-0,2546	1,2998
2009	0,4742	0,5237	0,3462	1,6496	1,0142	0,4623	1,1801
2010	1,2099	1,2180	0,6451	0,5705	1,4416	5,6254	0,0417
2011	0,9214	0,9510	0,75179	2,1624	-1,0863	-0,4833	0,6701
2012	0,16194	0,2108	-0,1907	0,1218	0,5989	-0,0385	0,8061
2013	0,9267	0,9662	0,3266	0,26074	1,2733	1,6778	0,1655
2014	2,2420	2,0803	1,5883	0,5492	0,4958	-0,0773	-0,2493
2015	0,2770	0,3253	-0,0194	1,1101	0,6936	-0,0123	0,43345

APÊNDICE 7 – MATRIZ DE INDICADORES DE CHEIA NORMALIZADOS OBTIDOS ATRAVÉS DO MÉTODO DA ESCALA DECIMAL; VALORES DE I_5 E I_7 ALTERADOS.

Ano	I_1	I_2	I_3	I_4	I_5	I_6	I_7
1931	0,1453	0,051	0,0014	0,3111	0,52523	0,1399	0,1704
1932	0,1833	0,0584	0,0198	0,8222	0,50278	0,1103	0,12
1933	0,0836	0,0373	0,0000	0,0000	0,0000	0,0000	0,0000
1934	0,082	0,0369	0,0000	0,0000	0,0000	0,0000	0,0000
1935	0,3267	0,0816	0,3207	43,222	0,45686	0,0813	0,0000
1936	0,1242	0,0468	0,0000	0,0000	0,0000	0,0000	0,0000
1937	0,1399	0,0499	0,0000	0,0777	0,46115	0,084	0,1761
1938	0,2383	0,0682	0,1108	21,555	0,52374	0,1626	0,126
1939	0,1717	0,0563	0,0265	16,111	0,31101	0,0445	0,1689
1940	0,0568	0,0302	0,0000	0,0000	0,0000	0,0000	0,0000
1941	0,1268	0,0474	0,0000	0,0000	0,0000	0,0000	0,0000
1942	0,1198	0,0457	0,0000	0,0000	0,0000	0,0000	0,0000
1943	0,0902	0,0389	0,0000	0,0000	0,0000	0,0000	0,0000
1944	0,0956	0,0402	0,0000	0,0000	0,0000	0,0000	0,0000
1945	0,1218	0,0462	0,0000	0,0000	0,0000	0,0000	0,0000
1946	0,2334	0,0673	0,1392	40,222	0,31258	0,046	0,1191
1947	0,1777	0,0574	0,046	19,222	0,5129	0,1088	0,1531
1948	0,1443	0,0508	0,0015	0,4777	0,27447	0,0401	0,1715
1949	0,0804	0,0365	0,0000	0,0000	0,0000	0,0000	0,0000
1950	0,161	0,0542	0,0088	0,7777	0,51068	0,1471	0,1478
1951	0,1194	0,0456	0,0000	0,0000	0,0000	0,0000	0,0000
1952	0,1093	0,0434	0,0000	0,0000	0,0000	0,0000	0,0000
1953	0,155	0,053	0,0056	0,6555	0,42633	0,0728	0,1626
1954	0,2403	0,0685	0,1146	20,000	0,38017	0,0583	0,0941
1955	0,1985	0,0612	0,1225	25,999	0,50753	0,1146	0,1736
1956	0,1234	0,0466	0,0000	0,0000	0,0000	0,0000	0,0000
1957	0,2675	0,0728	0,5351	88,999	0,53645	0,2336	0,1702
1958	0,103	0,042	0,0000	0,0000	0,0000	0,0000	0,0000
1959	0,0804	0,0365	0,0000	0,0000	0,0000	0,0000	0,0000
1960	0,1093	0,0434	0,0000	0,0000	0,0000	0,0000	0,0000
1961	0,1783	0,0575	0,0311	14,333	0,04467	0,0218	0,1636
1962	0,1002	0,0413	0,0000	0,0000	0,0000	0,0000	0,0000
1963	0,1048	0,0424	0,0000	0,0000	0,0000	0,0000	0,0000
1964	0,0914	0,0392	0,0000	0,0000	0,0000	0,0000	0,0000
1965	0,1472	0,0514	0,0043	0,8555	0,48973	0,1177	0,1778
1966	0,1399	0,0499	0,0002	0,2111	0,24579	0,0309	0,1885
1967	0,121	0,046	0,0000	0,0000	0,0000	0,0000	0,0000
1968	0,0745	0,035	0,0000	0,0000	0,0000	0,0000	0,0000
1969	0,1286	0,0478	0,0000	0,0000	0,0000	0,0000	0,0000
1970	0,1495	0,0519	0,0081	11,222	0,32189	0,0465	0,1749

Ano	I_1	I_2	I_3	I_4	I_5	I_6	I_7
1971	0,2429	0,0689	0,1595	29,898	0,41269	0,0726	0,1132
1972	0,1833	0,0584	0,0697	27,666	0,55251	0,3067	0,0877
1973	0,1761	0,0571	0,0439	29,666	0,41484	0,0612	0,1784
1974	0,1022	0,0418	0,0000	0,0000	0,0000	0,0000	0,0000
1975	0,1711	0,0562	0,0267	16,444	0,41846	0,0606	0,1382
1976	0,1419	0,0503	0,001	0,5111	0,41055	0,0601	0,189
1977	0,1048	0,0424	0,0000	0,0000	0,0000	0,0000	0,0000
1978	0,0816	0,0368	0,0000	0,0000	0,0000	0,0000	0,0000
1979	0,1565	0,0533	0,006	0,6999	0,43998	0,0857	0,1602
1980	0,1715	0,0548	0,0232	13,222	0,45816	0,0748	0,1616
1981	0,1887	0,0579	0,0603	21,636	0,34203	0,0442	0,1531
1982	0,2263	0,0645	0,1337	27,111	0,48098	0,0621	0,1343
1983	0,5157	0,1042	0,9689	87,999	0,26365	0,033	0,0645
1984	0,194	0,0594	0,0463	17,666	0,52939	0,1913	0,1617
1985	0,0769	0,0347	0,0000	0,0000	0,0000	0,0000	0,0000
1986	0,1257	0,0459	0,0000	0,0000	0,0000	0,0000	0,0000
1987	0,2167	0,0628	0,1024	22,333	0,48803	0,1225	0,1138
1988	0,1759	0,0556	0,0261	14,999	0,29768	0,0417	0,1686
1989	0,1887	0,0579	0,0546	16,777	0,53061	0,1687	0,086
1990	0,2196	0,0633	0,0723	19,222	0,31938	0,0472	0,1093
1991	0,1155	0,0437	0,0000	0,0000	0,0000	0,0000	0,0000
1992	0,3954	0,089	0,3867	24,888	0,43099	0,0725	0,0080
1993	0,2786	0,0725	0,1769	27,222	0,38924	0,0612	0,1051
1994	0,136	0,048	0,0000	0,0000	0,0000	0,0000	0,0000
1995	0,2453	0,0675	0,1241	23,333	0,41492	0,0639	0,1191
1996	0,1555	0,0519	0,0065	0,8999	0,3507	0,0525	0,1723
1997	0,2395	0,0666	0,2085	56,666	0,32938	0,0438	0,1775
1998	0,2752	0,072	0,3332	6,111	0,46435	0,0752	0,1107
1999	0,1833	0,057	0,046	16,777	0,33965	0,048	0,1299
2000	0,2284	0,0649	0,0835	17,888	0,53606	0,2115	0,0859
2001	0,1976	0,0596	0,0515	18,224	0,48245	0,0983	0,1307
2002	0,1162	0,044	0,0000	0,0000	0,0000	0,0000	0,0000
2003	0,1224	0,0454	0,0000	0,0000	0,0000	0,0000	0,0000
2004	0,1501	0,051	0,0024	0,5354	0,41033	0,0702	0,1698
2005	0,1897	0,0582	0,0765	42,224	0,5213	0,1714	0,1465
2006	0,0581	0,0303	0,0000	0,0000	0,0000	0,0000	0,0000
2007	0,1867	0,0577	0,031	13,777	0,30291	0,0436	0,1551
2008	0,1471	0,0505	0,0021	0,6448	0,30092	0,0401	0,1812
2009	0,205	0,0609	0,1156	45,555	0,47681	0,0975	0,1724
2010	0,2606	0,07	0,1568	25,345	0,5777	0,5108	0,0887
2011	0,2388	0,0665	0,1715	55,115	0,0188	0,0218	0,1349
2012	0,1814	0,0568	0,0416	16,999	0,38626	0,0574	0,1449
2013	0,2392	0,0667	0,1129	19,595	0,53332	0,1948	0,0978
2014	0,3386	0,0813	0,2868	24,949	0,36378	0,0543	0,0673
2015	0,1901	0,0583	0,0652	35,454	0,40692	0,0595	0,1175

APÊNDICE 8 – MATRIZ DE INDICADORES DE CHEIA NORMALIZADOS OBTIDOS ATRAVÉS DO MÉTODO DO VALOR MÁXIMO DOS ELEMENTOS; VALORES DE I_5 E I_7 ALTERADOS.

Ano	I_1	I_2	I_3	I_4	I_5	I_6	I_7
1931	0,2817	0,4894	0,0014	0,0348	0,9214	0,2738	0,9015
1932	0,3554	0,5604	0,0204	0,0922	0,8820	0,2159	0,6349
1933	0,1621	0,3579	0,0000	0,0000	0,0000	0,0000	0,0000
1934	0,1590	0,3541	0,0000	0,0000	0,0000	0,0000	0,0000
1935	0,6335	0,7831	0,3309	0,4859	0,8015	0,1591	0,0000
1936	0,24083	0,4491	0,0000	0,0000	0,0000	0,0000	0,0000
1937	0,2712	0,4788	0,0000	0,0078	0,8090	0,1644	0,9317
1938	0,4620	0,6545	0,1143	0,2418	0,9188	0,3183	0,6666
1939	0,3329	0,5403	0,0273	0,1811	0,5456	0,0871	0,893
1940	0,1101	0,2898	0,0000	0,0000	0,0000	0,0000	0,0000
1941	0,2458	0,4548	0,0000	0,0000	0,0000	0,0000	0,0000
1942	0,2323	0,4385	0,0000	0,0000	0,0000	0,0000	0,0000
1943	0,17490	0,3733	0,0000	0,0000	0,0000	0,0000	0,0000
1944	0,1853	0,3857	0,0000	0,0000	0,0000	0,0000	0,0000
1945	0,2361	0,4433	0,0000	0,0000	0,0000	0,0000	0,0000
1946	0,4525	0,6458	0,1436	0,4521	0,5483	0,0900	0,6301
1947	0,3445	0,5508	0,0474	0,2159	0,8998	0,2129	0,8100
1948	0,2798	0,4875	0,0015	0,05286	0,4815	0,0785	0,9074
1949	0,1559	0,3502	0,0000	0,0000	0,0000	0,0000	0,0000
1950	0,3121	0,5201	0,0090	0,0866	0,8959	0,2879	0,7820
1951	0,2315	0,4376	0,0000	0,0000	0,0000	0,0000	0,0000
1952	0,2119	0,4165	0,0000	0,0000	0,0000	0,0000	0,0000
1953	0,3005	0,5086	0,0057	0,0731	0,7479	0,1425	0,8603
1954	0,4659	0,6573	0,1182	0,2249	0,6669	0,1141	0,4978
1955	0,3849	0,5873	0,1264	0,2913	0,8904	0,2243	0,9185
1956	0,2392	0,4472	0,0000	0,0000	0,0000	0,0000	0,0000
1957	0,5187	0,6986	0,5522	1,0000	0,9411	0,4573	0,9005
1958	0,1997	0,4030	0,0000	0,0000	0,0000	0,0000	0,0000
1959	0,1559	0,3502	0,0000	0,0000	0,0000	0,0000	0,0000
1960	0,2119	0,4165	0,0000	0,0000	0,0000	0,0000	0,0000
1961	0,3457	0,5518	0,0320	0,16085	0,0783	0,0426	0,8656
1962	0,1942	0,3963	0,0000	0,0000	0,0000	0,0000	0,0000
1963	0,2032	0,4069	0,0000	0,0000	0,0000	0,0000	0,0000
1964	0,1772	0,3761	0,0000	0,0000	0,0000	0,0000	0,0000
1965	0,2854	0,4932	0,0044	0,0956	0,8591	0,2304	0,9407
1966	0,2712	0,4788	0,0002	0,0236	0,4312	0,0604	0,9973
1967	0,2346	0,4414	0,0000	0,0000	0,0000	0,0000	0,0000
1968	0,1444	0,3358	0,0000	0,0000	0,0000	0,0000	0,0000
1969	0,2493	0,4587	0,0000	0,0000	0,0000	0,0000	0,0000
1970	0,2898	0,4980	0,0083	0,1259	0,5647	0,0910	0,9253

Ano	I_1	I_2	I_3	I_4	I_5	I_6	I_7
1971	0,4710	0,6612	0,1646	0,3352	0,7240	0,1421	0,5989
1972	0,3554	0,5604	0,0719	0,3104	0,9693	0,6004	0,4640
1973	0,3414	0,5479	0,0453	0,3329	0,7277	0,1198	0,9439
1974	0,1981	0,4011	0,0000	0,0000	0,0000	0,0000	0,0000
1975	0,3317	0,5393	0,0275	0,1844	0,7341	0,1186	0,7312
1976	0,2751	0,4827	0,0010	0,0573	0,7202	0,1176	1,0000
1977	0,2032	0,4069	0,0000	0,0000	0,0000	0,0000	0,0000
1978	0,1582	0,3531	0,0000	0,0000	0,0000	0,0000	0,0000
1979	0,3034	0,5115	0,0061	0,0776	0,7718	0,1677	0,8476
1980	0,3325	0,5259	0,0239	0,1484	0,8037	0,1464	0,8550
1981	0,3659	0,5556	0,0622	0,2429	0,6000	0,0865	0,8100
1982	0,4388	0,6190	0,1379	0,3048	0,8438	0,1215	0,7105
1983	1,0000	1,0000	1,0000	0,9887	0,4625	0,0646	0,3412
1984	0,3761	0,5700	0,0477	0,1979	0,9287	0,3745	0,8555
1985	0,1491	0,3330	0,0000	0,0000	0,0000	0,0000	0,0000
1986	0,2437	0,4404	0,0000	0,0000	0,0000	0,0000	0,0000
1987	0,4202	0,6026	0,1056	0,2508	0,8561	0,2398	0,6021
1988	0,3410	0,5335	0,0269	0,1676	0,5222	0,0816	0,8920
1989	0,3659	0,5556	0,0563	0,1878	0,9308	0,3302	0,4550
1990	0,4258	0,6074	0,0746	0,2159	0,5603	0,0924	0,5783
1991	0,2239	0,4193	0,0000	0,0000	0,0000	0,0000	0,0000
1992	0,7667	0,8541	0,3991	0,2789	0,7561	0,1419	0,0423
1993	0,5402	0,6957	0,1825	0,3059	0,6828	0,1198	0,5560
1994	0,2637	0,4606	0,0000	0,0000	0,0000	0,0000	0,0000
1995	0,4756	0,6477	0,1280	0,2620	0,7279	0,1250	0,6301
1996	0,3015	0,4980	0,0067	0,1001	0,6152	0,1027	0,9116
1997	0,4644	0,6391	0,2151	0,6366	0,5778	0,0857	0,9391
1998	0,5336	0,6909	0,3438	0,6861	0,8146	0,1472	0,5857
1999	0,3554	0,5470	0,0474	0,1878	0,5958	0,0939	0,6873
2000	0,4428	0,6228	0,0861	0,2002	0,9404	0,4140	0,4544
2001	0,3831	0,5719	0,0531	0,2047	0,8464	0,1924	0,6915
2002	0,2253	0,4222	0,0000	0,0000	0,0000	0,0000	0,0000
2003	0,2373	0,4357	0,0000	0,0000	0,0000	0,0000	0,0000
2004	0,2910	0,4894	0,0024	0,0596	0,7198	0,1374	0,8984
2005	0,3678	0,5585	0,0789	0,4746	0,9145	0,3355	0,7751
2006	0,1126	0,2907	0,0000	0,0000	0,0000	0,0000	0,0000
2007	0,3620	0,5537	0,0319	0,1541	0,5314	0,0853	0,8206
2008	0,2852	0,4846	0,0021	0,0719	0,52792	0,0785	0,9587
2009	0,3975	0,5844	0,1193	0,5118	0,8365	0,1908	0,9121
2010	0,5053	0,6717	0,16183	0,2845	1,0000	1,0000	0,4693
2011	0,4630	0,6381	0,1770	0,6197	0,0329	0,0426	0,7137
2012	0,3517	0,5451	0,0429	0,1901	0,6776	0,1123	0,7666
2013	0,4638	0,6401	0,1165	0,2193	0,9356	0,3813	0,5174
2014	0,6565	0,7802	0,2960	0,2800	0,6382	0,1063	0,3560
2015	0,3686	0,5595	0,0672	0,3982	0,7138	0,1164	0,62169

APÊNDICE 9 – VARIAÇÃO ESTATÍSTICA DO INDICADOR I_t , NORMALIZADO ATRAVÉS DE NORMALIZAÇÃO LINEAR, NORMALIZAÇÃO POR DESVIO PADRÃO, NORMALIZAÇÃO POR ESCALA DECIMAL E NORMALIZAÇÃO POR VALOR MÁXIMO DO ELEMENTO.

Ano	I_t LINEAR	I_t DESVIO PADRÃO	I_t ESCALA DECIMAL	I_t VALOR MÁXIMO
1931	0,1929	-0,3138	0,1453	0,2817
1932	0,2757	0,1859	0,1833	0,3554
1933	0,0585	-1,1254	0,0836	0,1621
1934	0,0549	-1,1465	0,082	0,1590
1935	0,5881	2,0722	0,3267	0,6335
1936	0,1469	-0,5914	0,1242	0,24084
1937	0,1811	-0,3849	0,1399	0,2712
1938	0,3955	0,9094	0,2383	0,4620
1939	0,2503	0,0333	0,1717	0,3329
1940	0	-1,4780	0,0568	0,1101
1941	0,1525	-0,5572	0,1268	0,2458
1942	0,1373	-0,6493	0,1198	0,2323
1943	0,0728	-1,0386	0,0902	0,1749
1944	0,0846	-0,9676	0,0956	0,1853
1945	0,1417	-0,6229	0,1218	0,2361
1946	0,3848	0,8450	0,2334	0,4525
1947	0,2635	0,1123	0,1777	0,3445
1948	0,1908	-0,3270	0,1443	0,2798
1949	0,0514	-1,1675	0,0804	0,1559
1950	0,2271	-0,1073	0,161	0,3121
1951	0,1364	-0,6545	0,1194	0,2315
1952	0,1144	-0,7874	0,1093	0,2119
1953	0,214	-0,1862	0,155	0,3005
1954	0,3998	0,9357	0,2403	0,4659
1955	0,3088	0,3859	0,1985	0,3849
1956	0,1451	-0,6019	0,1234	0,2392
1957	0,4591	1,2935	0,2675	0,5187
1958	0,1007	-0,8702	0,103	0,1997
1959	0,0514	-1,1675	0,0804	0,1559
1960	0,1144	-0,7874	0,1093	0,2119
1961	0,2647	0,1202	0,1783	0,3457
1962	0,0946	-0,9071	0,1002	0,1942
1963	0,1046	-0,8466	0,1048	0,2032
1964	0,0755	-1,0228	0,0914	0,1772
1965	0,197	-0,2888	0,1472	0,2854
1966	0,1811	-0,3849	0,1399	0,2712
1967	0,1399	-0,6335	0,121	0,2346
1968	0,0386	-1,2451	0,0745	0,1444
1969	0,1565	-0,5335	0,1286	0,2493
1970	0,2021	-0,2586	0,1495	0,2898
1971	0,4055	0,9699	0,2429	0,4710
1972	0,2757	0,1859	0,1833	0,3554

Ano	I_t LINEAR	I_t DESVIO PADRÃO	I_t ESCALA DECIMAL	I_t VALOR MÁXIMO
1973	0,2599	0,0912	0,1761	0,3414
1974	0,0989	-0,8808	0,1022	0,1981
1975	0,2491	0,0255	0,1711	0,3317
1976	0,1855	-0,3585	0,1419	0,2751
1977	0,1046	-0,8466	0,1048	0,2032
1978	0,054	-1,1517	0,0816	0,1582
1979	0,2173	-0,1665	0,1565	0,3034
1980	0,25	0,0307	0,1715	0,3325
1981	0,2875	0,2570	0,1887	0,3659
1982	0,3694	0,7516	0,2263	0,4388
1983	1	4,5584	0,5157	1
1984	0,2989	0,3267	0,194	0,3761
1985	0,0437	-1,2136	0,0769	0,1491
1986	0,1501	-0,5716	0,1257	0,2437
1987	0,3483	0,6253	0,2167	0,4202
1988	0,2595	0,0886	0,1759	0,3410
1989	0,2875	0,2570	0,1887	0,3659
1990	0,3548	0,6634	0,2196	0,4258
1991	0,128	-0,7058	0,1155	0,2239
1992	0,7378	2,9759	0,3954	0,7667
1993	0,4834	1,4395	0,2786	0,5402
1994	0,1726	-0,4362	0,136	0,2637
1995	0,4109	1,0015	0,2453	0,4756
1996	0,215	-0,1797	0,1555	0,3015
1997	0,3982	0,9252	0,2395	0,4644
1998	0,4759	1,3948	0,2752	0,5336
1999	0,2757	0,1859	0,1833	0,3554
2000	0,374	0,7792	0,2284	0,4428
2001	0,3068	0,3740	0,1976	0,3831
2002	0,1295	-0,6966	0,1162	0,2253
2003	0,1429	-0,6151	0,1224	0,2373
2004	0,2033	-0,2507	0,1501	0,2910
2005	0,2896	0,2701	0,1897	0,3678
2006	0,0028	-1,4609	0,0581	0,1126
2007	0,2824	0,2307	0,1867	0,3620
2008	0,1967	-0,2901	0,1471	0,2852
2009	0,323	0,4714	0,205	0,3975
2010	0,444	1,2028	0,2606	0,5053
2011	0,3967	0,9160	0,2388	0,4630
2012	0,2714	0,1609	0,1814	0,3517
2013	0,3974	0,9213	0,2392	0,4638
2014	0,614	2,2288	0,3386	0,6565
2015	0,2905	0,2754	0,1901	0,3686

APÊNDICE 10 – VARIAÇÃO ESTATÍSTICA DO INDICADOR I_2 , NORMALIZADO ATRAVÉS DE NORMALIZAÇÃO LINEAR, NORMALIZAÇÃO POR DESVIO PADRÃO, NORMALIZAÇÃO POR ESCALA DECIMAL E NORMALIZAÇÃO POR VALOR MÁXIMO DO ELEMENTO

Ano	I_2 LINEAR	I_2 DESVIO PADRÃO	I_2 ESCALA DECIMAL	I_2 VALOR MÁXIMO
1931	0,2811	-0,2303	0,0510	0,4894
1932	0,3811	0,3310	0,0584	0,5605
1933	0,0959	-1,2695	0,0373	0,3580
1934	0,0905	-1,2999	0,0369	0,3541
1935	0,6946	2,0908	0,0816	0,7831
1936	0,2243	-0,5489	0,0468	0,4491
1937	0,2662	-0,3138	0,0499	0,4789
1938	0,5135	1,0744	0,0682	0,6545
1939	0,3527	0,1717	0,0563	0,5403
1940	0,0000	-1,8081	0,0302	0,2898
1941	0,2324	-0,5034	0,0474	0,4549
1942	0,2095	-0,6324	0,0457	0,4386
1943	0,1176	-1,1482	0,0389	0,3733
1944	0,1351	-1,0496	0,0402	0,3858
1945	0,2162	-0,5944	0,0462	0,4434
1946	0,5014	1,0061	0,0673	0,6459
1947	0,3676	0,2551	0,0574	0,5509
1948	0,2784	-0,2455	0,0508	0,4875
1949	0,0851	-1,3302	0,0365	0,3503
1950	0,3243	0,0124	0,0542	0,5202
1951	0,2081	-0,6399	0,0456	0,4376
1952	0,1784	-0,8068	0,0434	0,4165
1953	0,3181	-0,0786	0,0530	0,5086
1954	0,5176	1,0971	0,0685	0,6574
1955	0,4189	0,5434	0,0612	0,5873
1956	0,2216	-0,5641	0,0466	0,4472
1957	0,5757	1,4233	0,0728	0,6987
1958	0,1595	-0,9130	0,0420	0,4031
1959	0,0851	-1,3302	0,0365	0,3503
1960	0,1784	-0,8068	0,0434	0,4165
1961	0,3689	0,2627	0,0575	0,5518
1962	0,1500	-0,9661	0,0413	0,3964
1963	0,1649	-0,8827	0,0424	0,4069
1964	0,1216	-1,1254	0,0392	0,3762
1965	0,2865	-0,2000	0,0514	0,4933
1966	0,2662	-0,3138	0,0499	0,4789
1967	0,2135	-0,6096	0,0460	0,4415
1968	0,0649	-1,4440	0,0350	0,3359
1969	0,2378	-0,4731	0,0478	0,4587
1970	0,2932	-0,1621	0,0519	0,4981
1971	0,5230	1,1275	0,0689	0,6612

Ano	I_2 LINEAR	I_2 DESVIO PADRÃO	I_2 ESCALA DECIMAL	I_2 VALOR MÁXIMO
1972	0,3811	0,3310	0,0584	0,5605
1973	0,3635	0,2324	0,0571	0,5480
1974	0,1568	-0,9282	0,0418	0,4012
1975	0,3514	0,1641	0,0562	0,5393
1976	0,2716	-0,2834	0,0503	0,4827
1977	0,1649	-0,8827	0,0424	0,4069
1978	0,0892	-1,3075	0,0368	0,3532
1979	0,3122	-0,0559	0,0533	0,5115
1980	0,3324	0,0579	0,0548	0,5259
1981	0,3743	0,2931	0,0579	0,5557
1982	0,4635	0,7937	0,0645	0,6190
1983	1,0000	3,8051	0,1042	1,0000
1984	0,3946	0,4068	0,0594	0,5701
1985	0,0608	-1,4668	0,0347	0,3330
1986	0,2122	-0,6172	0,0459	0,4405
1987	0,4405	0,6648	0,0628	0,6027
1988	0,3432	0,1186	0,0556	0,5336
1989	0,3743	0,2931	0,0579	0,5557
1990	0,4473	0,7027	0,0633	0,6075
1991	0,1824	-0,7841	0,0437	0,4194
1992	0,7946	2,6521	0,0890	0,8541
1993	0,5716	1,4005	0,0725	0,6958
1994	0,2405	-0,4579	0,0480	0,4607
1995	0,5041	1,0213	0,0675	0,6478
1996	0,2932	-0,1621	0,0519	0,4981
1997	0,4919	0,9530	0,0666	0,6392
1998	0,5649	1,3626	0,0720	0,6910
1999	0,3622	0,2248	0,0570	0,5470
2000	0,4689	0,8240	0,0649	0,6228
2001	0,3973	0,4220	0,0596	0,5720
2002	0,1865	-0,7613	0,0440	0,4223
2003	0,2054	-0,6551	0,0454	0,4357
2004	0,2811	-0,2303	0,0510	0,4894
2005	0,3784	0,3158	0,0582	0,5585
2006	0,0014	-1,8005	0,0303	0,2908
2007	0,3716	0,2779	0,0577	0,5537
2008	0,2743	-0,2683	0,0505	0,4846
2009	0,4149	0,5206	0,0609	0,5845
2010	0,5378	1,2109	0,0700	0,6718
2011	0,4905	0,9454	0,0665	0,6382
2012	0,3595	0,2096	0,0568	0,5451
2013	0,4932	0,9606	0,0667	0,6401
2014	0,6905	2,0681	0,0813	0,7802
2015	0,3797	0,3234	0,0583	0,5595

APÊNDICE 11 – VARIAÇÃO ESTATÍSTICA DO INDICADOR I_3 , NORMALIZADO ATRAVÉS DE NORMALIZAÇÃO LINEAR, NORMALIZAÇÃO POR DESVIO PADRÃO, NORMALIZAÇÃO POR ESCALA DECIMAL E NORMALIZAÇÃO POR VALOR MÁXIMO DO ELEMENTO.

Ano	I_3 LINEAR	I_3 DESVIO PADRÃO	I_3 ESCALA DECIMAL	I_3 VALOR MÁXIMO
1931	0,0014	-0,4795	0,0014	0,0014
1932	0,0204	-0,3468	0,0198	0,0204
1933	0,0000	-0,4896	0,0000	0,0000
1934	0,0000	-0,4896	0,0000	0,0000
1935	0,3310	1,8235	0,3207	0,3310
1936	0,0000	-0,4896	0,0000	0,0000
1937	0,0000	-0,4896	0,0000	0,0000
1938	0,1143	0,3095	0,1108	0,1144
1939	0,0274	-0,2985	0,0265	0,0274
1940	0,0000	-0,4896	0,0000	0,0000
1941	0,0000	-0,4896	0,0000	0,0000
1942	0,0000	-0,4896	0,0000	0,0000
1943	0,0000	-0,4896	0,0000	0,0000
1944	0,0000	-0,4896	0,0000	0,0000
1945	0,0000	-0,4896	0,0000	0,0000
1946	0,1437	0,5144	0,1392	0,1437
1947	0,0475	-0,1578	0,0460	0,0475
1948	0,0016	-0,4788	0,0015	0,0015
1949	0,0000	-0,4896	0,0000	0,0000
1950	0,0091	-0,4262	0,0088	0,0091
1951	0,0000	-0,4896	0,0000	0,0000
1952	0,0000	-0,4896	0,0000	0,0000
1953	0,0058	-0,4492	0,0056	0,0058
1954	0,1183	0,3370	0,1146	0,1183
1955	0,1265	0,3939	0,1225	0,1264
1956	0,0000	-0,4896	0,0000	0,0000
1957	0,5523	3,3699	0,5351	0,5523
1958	0,0000	-0,4896	0,0000	0,0000
1959	0,0000	-0,4896	0,0000	0,0000
1960	0,0000	-0,4896	0,0000	0,0000
1961	0,0321	-0,2653	0,0311	0,0321
1962	0,0000	-0,4896	0,0000	0,0000
1963	0,0000	-0,4896	0,0000	0,0000
1964	0,0000	-0,4896	0,0000	0,0000
1965	0,0045	-0,4586	0,0043	0,0044
1966	0,0020	-0,4882	0,0002	0,0002
1967	0,0000	-0,4896	0,0000	0,0000
1968	0,0000	-0,4896	0,0000	0,0000
1969	0,0000	-0,4896	0,0000	0,0000
1970	0,0083	-0,4312	0,0081	0,0084
1971	0,1646	0,6608	0,1595	0,1646
1972	0,0719	0,0131	0,0697	0,0719

Ano	I_3 LINEAR	I_3 DESVIO PADRÃO	I_3 ESCALA DECIMAL	I_3 VALOR MÁXIMO
1973	0,0453	-0,1730	0,0439	0,0453
1974	0,0000	-0,4896	0,0000	0,0000
1975	0,0276	-0,2970	0,0267	0,0276
1976	0,0010	-0,4824	0,0010	0,0010
1977	0,0000	-0,4896	0,0000	0,0000
1978	0,0000	-0,4896	0,0000	0,0000
1979	0,0062	-0,4464	0,0060	0,0062
1980	0,0239	-0,3223	0,0232	0,0239
1981	0,0622	-0,0547	0,0603	0,0622
1982	0,1380	0,4747	0,1337	0,1380
1983	1,0000	6,4988	0,9689	1,0000
1984	0,0478	-0,1557	0,0463	0,0478
1985	0,0000	-0,4896	0,0000	0,0000
1986	0,0000	-0,4896	0,0000	0,0000
1987	0,1057	0,2490	0,1024	0,1057
1988	0,0270	-0,3014	0,0261	0,0269
1989	0,0563	-0,0958	0,0546	0,0564
1990	0,0747	0,0319	0,0723	0,0746
1991	0,0000	-0,4896	0,0000	0,0000
1992	0,3991	2,2995	0,3867	0,3991
1993	0,1826	0,7863	0,1769	0,1826
1994	0,0000	-0,4896	0,0000	0,0000
1995	0,1281	0,4055	0,1241	0,1281
1996	0,0067	-0,4427	0,0065	0,0067
1997	0,2152	1,0142	0,2085	0,2152
1998	0,3438	1,9137	0,3332	0,3439
1999	0,0475	-0,1578	0,0460	0,0475
2000	0,0862	0,1126	0,0835	0,0862
2001	0,0532	-0,1182	0,0515	0,0532
2002	0,0000	-0,4896	0,0000	0,0000
2003	0,0000	-0,4896	0,0000	0,0000
2004	0,0025	-0,4723	0,0024	0,0025
2005	0,0790	0,0621	0,0765	0,0790
2006	0,0000	-0,4896	0,0000	0,0000
2007	0,0320	-0,2660	0,0310	0,0320
2008	0,0022	-0,4745	0,0021	0,0022
2009	0,1193	0,3442	0,1156	0,1193
2010	0,1618	0,6413	0,1568	0,1618
2011	0,1770	0,7474	0,1715	0,1770
2012	0,0429	-0,1896	0,0416	0,0429
2013	0,1165	0,3247	0,1129	0,1165
2014	0,2960	1,5790	0,2868	0,2960
2015	0,0673	-0,0194	0,0652	0,0673

APÊNDICE 12 – VARIAÇÃO ESTATÍSTICA DO INDICADOR I_4 , NORMALIZADO ATRAVÉS DE NORMALIZAÇÃO LINEAR, NORMALIZAÇÃO POR DESVIO PADRÃO, NORMALIZAÇÃO POR ESCALA DECIMAL E NORMALIZAÇÃO POR VALOR MÁXIMO DO ELEMENTO.

Ano	I_4 LINEAR	I_4 DESVIO PADRÃO	I_4 ESCALA DECIMAL	I_4 VALOR MÁXIMO
1931	0,0347	-0,6117	0,3100	0,0349
1932	0,0917	-0,3409	0,8200	0,0922
1933	0,0000	-0,7763	0,0000	0,0000
1934	0,0000	-0,7763	0,0000	0,0000
1935	0,4862	1,5178	4,3200	0,4859
1936	0,0000	-0,7763	0,0000	0,0000
1937	0,0079	-0,7391	0,0700	0,0079
1938	0,2413	0,3654	2,1500	0,2418
1939	0,1808	0,0787	1,6100	0,1811
1940	0,0000	-0,7763	0,0000	0,0000
1941	0,0000	-0,7763	0,0000	0,0000
1942	0,0000	-0,7763	0,0000	0,0000
1943	0,0000	-0,7763	0,0000	0,0000
1944	0,0000	-0,7763	0,0000	0,0000
1945	0,0000	-0,7763	0,0000	0,0000
1946	0,4518	1,3585	4,0200	0,4522
1947	0,2162	0,2433	1,9200	0,2160
1948	0,0533	-0,5267	0,4700	0,0529
1949	0,0000	-0,7763	0,0000	0,0000
1950	0,0863	-0,3674	0,7700	0,0866
1951	0,0000	-0,7763	0,0000	0,0000
1952	0,0000	-0,7763	0,0000	0,0000
1953	0,0725	-0,4311	0,6500	0,0731
1954	0,2245	0,2858	2,0000	0,2250
1955	0,4036	0,5991	2,5900	0,2913
1956	0,0000	-0,7763	0,0000	0,0000
1957	1,0000	3,9446	8,8900	1,0000
1958	0,0000	-0,7763	0,0000	0,0000
1959	0,0000	-0,7763	0,0000	0,0000
1960	0,0000	-0,7763	0,0000	0,0000
1961	0,1613	-0,0169	1,4300	0,1609
1962	0,0000	-0,7763	0,0000	0,0000
1963	0,0000	-0,7763	0,0000	0,0000
1964	0,0000	-0,7763	0,0000	0,0000
1965	0,0951	-0,3249	0,8500	0,0956
1966	0,0320	-0,6648	0,2100	0,0236
1967	0,0000	-0,7763	0,0000	0,0000
1968	0,0000	-0,7763	0,0000	0,0000
1969	0,0000	-0,7763	0,0000	0,0000
1970	0,1254	-0,1816	1,1200	0,1260
1971	0,3347	0,8062	2,9800	0,3352

Ano	I_4 LINEAR	I_4 DESVIO PADRÃO	I_4 ESCALA DECIMAL	I_4 VALOR MÁXIMO
1972	0,3107	0,6894	2,7600	0,3105
1973	0,3328	0,7956	2,9600	0,3330
1974	0,0000	-0,7763	0,0000	0,0000
1975	0,1848	0,0946	1,6400	0,1845
1976	0,0578	-0,5055	0,5100	0,0574
1977	0,0000	-0,7763	0,0000	0,0000
1978	0,0000	-0,7763	0,0000	0,0000
1979	0,0777	-0,4099	0,6900	0,0776
1980	0,1479	-0,0753	1,3200	0,1485
1981	0,2434	0,3707	2,1600	0,2430
1982	0,3050	0,6628	2,7100	0,3048
1983	0,9880	3,8915	8,7900	0,9888
1984	0,1981	0,1583	1,7600	0,1980
1985	0,0000	-0,7763	0,0000	0,0000
1986	0,0000	-0,7763	0,0000	0,0000
1987	0,2503	0,4079	2,2300	0,2508
1988	0,1674	0,0149	1,4900	0,1676
1989	0,1872	0,1105	1,6700	0,1879
1990	0,2163	0,2433	1,9200	0,2160
1991	0,0000	-0,7763	0,0000	0,0000
1992	0,2784	0,5407	2,4800	0,2790
1993	0,3060	0,6681	2,7200	0,3060
1994	0,0000	-0,7763	0,0000	0,0000
1995	0,2615	0,4610	2,3300	0,2621
1996	0,1003	-0,3037	0,8900	0,1001
1997	0,6358	2,2294	5,6600	0,6367
1998	0,6857	2,4630	6,1000	0,6862
1999	0,1876	0,1105	1,6700	0,1879
2000	0,1997	0,1689	1,7800	0,2002
2001	0,2045	0,1902	1,8200	0,2047
2002	0,0000	-0,7763	0,0000	0,0000
2003	0,0000	-0,7763	0,0000	0,0000
2004	0,0595	-0,4949	0,5300	0,0596
2005	0,4748	1,4647	4,2200	0,4747
2006	0,0000	-0,7763	0,0000	0,0000
2007	0,1545	-0,0488	1,3700	0,1541
2008	0,0724	-0,4365	0,6400	0,0720
2009	0,5111	1,6399	4,5500	0,5118
2010	0,2848	0,5672	2,5300	0,2846
2011	0,6192	2,1497	5,5100	0,6198
2012	0,1899	0,1211	1,6900	0,1901
2013	0,2196	0,2592	1,9500	0,2193
2014	0,2795	0,5460	2,4900	0,2801
2015	0,3980	1,1036	3,5400	0,3982

APÊNDICE 13 – VARIAÇÃO ESTATÍSTICA DO INDICADOR I_5 , NORMALIZADO ATRAVÉS DE NORMALIZAÇÃO LINEAR, NORMALIZAÇÃO POR DESVIO PADRÃO, NORMALIZAÇÃO POR ESCALA DECIMAL E NORMALIZAÇÃO POR VALOR MÁXIMO DO ELEMENTO.

Ano	I_5 LINEAR	I_5 DESVIO PADRÃO	I_5 ESCALA DECIMAL	I_5 VALOR MÁXIMO
1931	0,1093	-0,3802	0,6389	0,1120
1932	0,1477	-0,2015	0,8634	0,1514
1933	0,0000	-0,8887	0,0000	0,0000
1934	0,0000	-0,8887	0,0000	0,0000
1935	0,2263	0,1639	1,3226	0,2319
1936	0,0000	-0,8887	0,0000	0,0000
1937	0,2190	0,1298	1,2797	0,2244
1938	0,1119	-0,3683	0,6538	0,1146
1939	0,4758	1,3246	2,7811	0,4876
1940	0,0000	-0,8887	0,0000	0,0000
1941	0,0000	-0,8887	0,0000	0,0000
1942	0,0000	-0,8887	0,0000	0,0000
1943	0,0000	-0,8887	0,0000	0,0000
1944	0,0000	-0,8887	0,0000	0,0000
1945	0,0000	-0,8887	0,0000	0,0000
1946	0,4732	1,3121	2,7654	0,4849
1947	0,1304	-0,2821	0,7622	0,1336
1948	0,5384	1,6154	3,1465	0,5517
1949	0,0000	-0,8887	0,0000	0,0000
1950	0,1342	-0,2644	0,7844	0,1375
1951	0,0000	-0,8887	0,0000	0,0000
1952	0,0000	-0,8887	0,0000	0,0000
1953	0,2785	0,4069	1,6279	0,2854
1954	0,3575	0,7742	2,0895	0,3664
1955	0,1458	-0,2393	0,8159	0,1431
1956	0,0000	-0,8887	0,0000	0,0000
1957	0,0901	-0,4695	0,5267	0,0924
1958	0,0000	-0,8887	0,0000	0,0000
1959	0,0000	-0,8887	0,0000	0,0000
1960	0,0000	-0,8887	0,0000	0,0000
1961	1,0000	3,4443	5,4445	0,9546
1962	0,0000	-0,8887	0,0000	0,0000
1963	0,0000	-0,8887	0,0000	0,0000
1964	0,0000	-0,8887	0,0000	0,0000
1965	0,1701	-0,0977	0,9939	0,1743
1966	0,5875	1,8437	3,4333	0,6020
1967	0,0000	-0,8887	0,0000	0,0000
1968	0,0000	-0,8887	0,0000	0,0000
1969	0,0000	-0,8887	0,0000	0,0000
1970	0,4572	1,2381	2,6723	0,4686
1971	0,3019	0,5154	1,7643	0,3094
1972	0,0626	-0,5973	0,3661	0,0642

Ano	I_5 LINEAR	I_5 DESVIO PADRÃO	I_5 ESCALA DECIMAL	I_5 VALOR MÁXIMO
1973	0,2982	0,4983	1,7428	0,3056
1974	0,0000	-0,8887	0,0000	0,0000
1975	0,2920	0,4695	1,7066	0,2992
1976	0,3055	0,5325	1,7857	0,3131
1977	0,0000	-0,8887	0,0000	0,0000
1978	0,0000	-0,8887	0,0000	0,0000
1979	0,2552	0,2983	1,4914	0,2615
1980	0,2241	0,1536	1,3096	0,2296
1981	0,4228	1,0778	2,4709	0,4332
1982	0,1850	-0,0280	1,0814	0,1896
1983	0,5569	1,7015	3,2547	0,5707
1984	0,1022	-0,4133	0,5973	0,1047
1985	0,0000	-0,8887	0,0000	0,0000
1986	0,0000	-0,8887	0,0000	0,0000
1987	0,1730	-0,0841	1,0109	0,1773
1988	0,4987	1,4307	2,9144	0,5110
1989	0,1001	-0,4230	0,5851	0,1026
1990	0,4615	1,2580	2,6974	0,4730
1991	0,0000	-0,8887	0,0000	0,0000
1992	0,2706	0,3698	1,5813	0,2773
1993	0,3420	0,7021	1,9988	0,3505
1994	0,0000	-0,8887	0,0000	0,0000
1995	0,2981	0,4977	1,7420	0,3054
1996	0,4079	1,0088	2,3842	0,4180
1997	0,4444	1,1784	2,5974	0,4554
1998	0,2135	0,1043	1,2477	0,2188
1999	0,4268	1,0967	2,4947	0,4374
2000	0,0908	-0,4664	0,5306	0,0930
2001	0,1825	-0,0397	1,0667	0,1870
2002	0,0000	-0,8887	0,0000	0,0000
2003	0,0000	-0,8887	0,0000	0,0000
2004	0,3059	0,5342	1,7879	0,3135
2005	0,1160	-0,3489	0,6782	0,1189
2006	0,0000	-0,8887	0,0000	0,0000
2007	0,4897	1,3891	2,8621	0,5018
2008	0,4931	1,4049	2,8820	0,5053
2009	0,1922	0,0051	1,1231	0,1969
2010	0,0327	-0,7365	0,1912	0,0335
2011	0,9758	3,6501	5,7032	1,0000
2012	0,3471	0,7258	2,0286	0,3557
2013	0,0955	-0,4446	0,5580	0,0978
2014	0,3856	0,9047	2,2534	0,3951
2015	0,3117	0,5614	1,8220	0,3195

APÊNDICE 14 – VARIAÇÃO ESTATÍSTICA DO INDICADOR I_6 , NORMALIZADO ATRAVÉS DE NORMALIZAÇÃO LINEAR, NORMALIZAÇÃO POR DESVIO PADRÃO, NORMALIZAÇÃO POR ESCALA DECIMAL E NORMALIZAÇÃO POR VALOR MÁXIMO DO ELEMENTO.

Ano	I_6 LINEAR	I_6 DESVIO PADRÃO	I_6 ESCALA DECIMAL	I_6 VALOR MÁXIMO
1931	0,2740	0,9862	0,1399	0,2739
1932	0,2160	0,6186	0,1103	0,2159
1933	0,0000	-0,7512	0,0000	0,0000
1934	0,0000	-0,7512	0,0000	0,0000
1935	0,1592	0,2585	0,0813	0,1592
1936	0,0000	-0,7512	0,0000	0,0000
1937	0,1645	0,2920	0,0840	0,1644
1938	0,3183	1,2681	0,1626	0,3183
1939	0,0870	-0,1986	0,0445	0,0871
1940	0,0000	-0,7512	0,0000	0,0000
1941	0,0000	-0,7512	0,0000	0,0000
1942	0,0000	-0,7512	0,0000	0,0000
1943	0,0000	-0,7512	0,0000	0,0000
1944	0,0000	-0,7512	0,0000	0,0000
1945	0,0000	-0,7512	0,0000	0,0000
1946	0,0901	-0,1799	0,0460	0,0901
1947	0,2130	0,6000	0,1088	0,2130
1948	0,0785	-0,2532	0,0401	0,0785
1949	0,0000	-0,7512	0,0000	0,0000
1950	0,2880	1,0756	0,1471	0,2880
1951	0,0000	-0,7512	0,0000	0,0000
1952	0,0000	-0,7512	0,0000	0,0000
1953	0,1425	0,1529	0,0728	0,1425
1954	0,1142	-0,0272	0,0583	0,1141
1955	0,2243	0,6720	0,1146	0,2244
1956	0,0000	-0,7512	0,0000	0,0000
1957	0,4574	2,1498	0,2336	0,4573
1958	0,0000	-0,7512	0,0000	0,0000
1959	0,0000	-0,7512	0,0000	0,0000
1960	0,0000	-0,7512	0,0000	0,0000
1961	0,0427	-0,4805	0,0218	0,0427
1962	0,0000	-0,7512	0,0000	0,0000
1963	0,0000	-0,7512	0,0000	0,0000
1964	0,0000	-0,7512	0,0000	0,0000
1965	0,2304	0,7105	0,1177	0,2304
1966	0,0604	-0,3674	0,0309	0,0605
1967	0,0000	-0,7512	0,0000	0,0000
1968	0,0000	-0,7512	0,0000	0,0000
1969	0,0000	-0,7512	0,0000	0,0000
1970	0,0910	-0,1737	0,0465	0,0910
1971	0,1421	0,1504	0,0726	0,1421
1972	0,6005	3,0576	0,3067	0,6004

Ano	I_6 LINEAR	I_6 DESVIO PADRÃO	I_6 ESCALA DECIMAL	I_6 VALOR MÁXIMO
1973	0,1198	0,0088	0,0612	0,1198
1974	0,0000	-0,7512	0,0000	0,0000
1975	0,1187	0,0014	0,0606	0,1186
1976	0,1176	-0,0048	0,0601	0,1177
1977	0,0000	-0,7512	0,0000	0,0000
1978	0,0000	-0,7512	0,0000	0,0000
1979	0,1678	0,3131	0,0857	0,1678
1980	0,1464	0,1777	0,0748	0,1464
1981	0,0866	-0,2023	0,0442	0,0865
1982	0,1216	0,0200	0,0621	0,1216
1983	0,0646	-0,3414	0,0330	0,0646
1984	0,3746	1,6245	0,1913	0,3745
1985	0,0000	-0,7512	0,0000	0,0000
1986	0,0000	-0,7512	0,0000	0,0000
1987	0,2399	0,7701	0,1225	0,2398
1988	0,0816	-0,2333	0,0417	0,0816
1989	0,3302	1,3439	0,1687	0,3303
1990	0,0925	-0,1650	0,0472	0,0924
1991	0,0000	-0,7512	0,0000	0,0000
1992	0,1419	0,1492	0,0725	0,1419
1993	0,1198	0,0088	0,0612	0,1198
1994	0,0000	-0,7512	0,0000	0,0000
1995	0,1251	0,0424	0,0639	0,1251
1996	0,1028	-0,0992	0,0525	0,1028
1997	0,0858	-0,2072	0,0438	0,0857
1998	0,1472	0,1827	0,0752	0,1472
1999	0,0941	-0,1551	0,0480	0,0940
2000	0,4140	1,8754	0,2115	0,4141
2001	0,1925	0,4696	0,0983	0,1924
2002	0,0000	-0,7512	0,0000	0,0000
2003	0,0000	-0,7512	0,0000	0,0000
2004	0,1374	0,1206	0,0702	0,1374
2005	0,3355	1,3774	0,1714	0,3356
2006	0,0000	-0,7512	0,0000	0,0000
2007	0,0854	-0,2097	0,0436	0,0854
2008	0,0786	-0,2532	0,0401	0,0785
2009	0,1909	0,4596	0,0975	0,1909
2010	1,0000	5,5923	0,5108	1,0000
2011	0,0426	-0,4805	0,0218	0,0427
2012	0,1125	-0,0384	0,0574	0,1124
2013	0,3813	1,6680	0,1948	0,3814
2014	0,1063	-0,0769	0,0543	0,1063
2015	0,1165	-0,0123	0,0595	0,1165

APÊNDICE 15 – VARIAÇÃO ESTATÍSTICA DO INDICADOR I_7 , NORMALIZADO ATRAVÉS DE NORMALIZAÇÃO LINEAR, NORMALIZAÇÃO POR DESVIO PADRÃO, NORMALIZAÇÃO POR ESCALA DECIMAL E NORMALIZAÇÃO POR VALOR MÁXIMO DO ELEMENTO.

Ano	I_7 LINEAR	I_7 DESVIO PADRÃO	I_7 ESCALA DECIMAL	I_7 VALOR MÁXIMO
1931	0,1409	-0,2475	0,0267	0,1412
1932	0,4078	0,9595	0,0771	0,4077
1933	0,0000	-0,8869	0,0000	0,0000
1934	0,0000	-0,8869	0,0000	0,0000
1935	0,4856	1,3116	0,0918	0,4855
1936	0,0000	-0,8869	0,0000	0,0000
1937	0,1269	-0,3840	0,0210	0,1111
1938	0,3762	0,8158	0,0711	0,3760
1939	0,1491	-0,2116	0,0282	0,1491
1940	0,0000	-0,8869	0,0000	0,0000
1941	0,0000	-0,8869	0,0000	0,0000
1942	0,0000	-0,8869	0,0000	0,0000
1943	0,0000	-0,8869	0,0000	0,0000
1944	0,0000	-0,8869	0,0000	0,0000
1945	0,0000	-0,8869	0,0000	0,0000
1946	0,4123	0,9811	0,0780	0,4125
1947	0,2329	0,1668	0,0440	0,2327
1948	0,1353	-0,2738	0,0256	0,1354
1949	0,0000	-0,8869	0,0000	0,0000
1950	0,2609	0,2938	0,0493	0,2607
1951	0,0000	-0,8869	0,0000	0,0000
1952	0,0000	-0,8869	0,0000	0,0000
1953	0,1822	-0,0607	0,0345	0,1824
1954	0,5446	1,5798	0,1030	0,5447
1955	0,1244	-0,3241	0,0235	0,1243
1956	0,0000	-0,8869	0,0000	0,0000
1957	0,1421	-0,2427	0,0269	0,1423
1958	0,0000	-0,8869	0,0000	0,0000
1959	0,0000	-0,8869	0,0000	0,0000
1960	0,0000	-0,8869	0,0000	0,0000
1961	0,1774	-0,0846	0,0335	0,1772
1962	0,0000	-0,8869	0,0000	0,0000
1963	0,0000	-0,8869	0,0000	0,0000
1964	0,0000	-0,8869	0,0000	0,0000
1965	0,1020	-0,4247	0,0193	0,1021
1966	0,0453	-0,6810	0,0086	0,0455
1967	0,0000	-0,8869	0,0000	0,0000
1968	0,0000	-0,8869	0,0000	0,0000
1969	0,0000	-0,8869	0,0000	0,0000
1970	0,1175	-0,3553	0,0222	0,1174
1971	0,4437	1,1224	0,0839	0,4437
1972	0,5787	1,7331	0,1094	0,5785

Ano	I_7 LINEAR	I_7 DESVIO PADRÃO	I_7 ESCALA DECIMAL	I_7 VALOR MÁXIMO
1973	0,0987	-0,4391	0,0187	0,0989
1974	0,0000	-0,8869	0,0000	0,0000
1975	0,3115	0,5237	0,0589	0,3115
1976	0,0428	-0,6929	0,0081	0,0428
1977	0,0000	-0,8869	0,0000	0,0000
1978	0,0000	-0,8869	0,0000	0,0000
1979	0,1952	-0,0032	0,0369	0,1951
1980	0,1875	-0,0367	0,0355	0,1877
1981	0,2328	0,1668	0,0440	0,2327
1982	0,3322	0,6171	0,0628	0,3321
1983	0,7013	2,2887	0,1326	0,7012
1984	0,1874	-0,0391	0,0354	0,1872
1985	0,0000	-0,8869	0,0000	0,0000
1986	0,0000	-0,8869	0,0000	0,0000
1987	0,4402	1,1080	0,0833	0,4405
1988	0,1509	-0,2044	0,0285	0,1507
1989	0,5874	1,7738	0,1111	0,5875
1990	0,4642	1,2158	0,0878	0,4643
1991	0,0000	-0,8869	0,0000	0,0000
1992	1,0000	3,6418	0,1891	1,0000
1993	0,4866	1,3164	0,0920	0,4865
1994	0,0000	-0,8869	0,0000	0,0000
1995	0,4122	0,9811	0,0780	0,4125
1996	0,1310	-0,2930	0,0248	0,1311
1997	0,1035	-0,4175	0,0196	0,1036
1998	0,4569	1,1823	0,0864	0,4569
1999	0,3555	0,7224	0,0672	0,3554
2000	0,5881	1,7762	0,1112	0,5880
2001	0,3513	0,7033	0,0664	0,3511
2002	0,0000	-0,8869	0,0000	0,0000
2003	0,0000	-0,8869	0,0000	0,0000
2004	0,1443	-0,2331	0,0273	0,1444
2005	0,2676	0,3249	0,0506	0,2676
2006	0,0000	-0,8869	0,0000	0,0000
2007	0,2219	0,1189	0,0420	0,2221
2008	0,0841	-0,5061	0,0159	0,0841
2009	0,1304	-0,2954	0,0247	0,1306
2010	0,5732	1,7091	0,1084	0,5732
2011	0,3287	0,6027	0,0622	0,3289
2012	0,2758	0,3632	0,0522	0,2760
2013	0,5253	1,4912	0,0993	0,5251
2014	0,6866	2,2217	0,1298	0,6864
2015	0,4210	1,0194	0,0796	0,4209

APÊNDICE 16 – VARIAÇÃO ESTATÍSTICA DO INDICADOR I_5 ALTERADO NORMALIZADO ATRAVÉS DE NORMALIZAÇÃO LINEAR, NORMALIZAÇÃO POR DESVIO PADRÃO, NORMALIZAÇÃO POR ESCALA DECIMAL E NORMALIZAÇÃO POR VALOR MÁXIMO DO ELEMENTO.

Ano	I_5 LINEAR	I_5 DESVIO PADRÃO	I_5 ESCALA DECIMAL	I_5 VALOR MÁXIMO
1931	0,9215	1,2363	0,5252	0,9215
1932	0,8821	1,1333	0,5028	0,8821
1933	0,0000	-1,1726	0,0000	0,0000
1934	0,0000	-1,1726	0,0000	0,0000
1935	0,8015	0,9227	0,4569	0,8015
1936	0,0000	-1,1726	0,0000	0,0000
1937	0,8090	0,9424	0,4612	0,8090
1938	0,9188	1,2294	0,5237	0,9188
1939	0,5456	0,2538	0,3110	0,5456
1940	0,0000	-1,1726	0,0000	0,0000
1941	0,0000	-1,1726	0,0000	0,0000
1942	0,0000	-1,1726	0,0000	0,0000
1943	0,0000	-1,1726	0,0000	0,0000
1944	0,0000	-1,1726	0,0000	0,0000
1945	0,0000	-1,1726	0,0000	0,0000
1946	0,5484	0,2610	0,3126	0,5484
1947	0,8998	1,1797	0,5129	0,8998
1948	0,4815	0,0862	0,2745	0,4815
1949	0,0000	-1,1726	0,0000	0,0000
1950	0,8959	1,1695	0,5107	0,8959
1951	0,0000	-1,1726	0,0000	0,0000
1952	0,0000	-1,1726	0,0000	0,0000
1953	0,7479	0,7827	0,4263	0,7479
1954	0,6670	0,5710	0,3802	0,6670
1955	0,8904	1,1551	0,5075	0,8904
1956	0,0000	-1,1726	0,0000	0,0000
1957	0,9411	1,2877	0,5365	0,9411
1958	0,0000	-1,1726	0,0000	0,0000
1959	0,0000	-1,1726	0,0000	0,0000
1960	0,0000	-1,1726	0,0000	0,0000
1961	0,0784	-0,9677	0,0447	0,0784
1962	0,0000	-1,1726	0,0000	0,0000
1963	0,0000	-1,1726	0,0000	0,0000
1964	0,0000	-1,1726	0,0000	0,0000
1965	0,8592	1,0735	0,4897	0,8592
1966	0,4312	-0,0453	0,2458	0,4312
1967	0,0000	-1,1726	0,0000	0,0000
1968	0,0000	-1,1726	0,0000	0,0000
1969	0,0000	-1,1726	0,0000	0,0000
1970	0,5647	0,3037	0,3219	0,5647
1971	0,7240	0,7201	0,4127	0,7240

Ano	I_5 LINEAR	I_5 DESVIO PADRÃO	I_5 ESCALA DECIMAL	I_5 VALOR MÁXIMO
1972	0,9693	1,3614	0,5525	0,9693
1973	0,7278	0,7300	0,4148	0,7278
1974	0,0000	-1,1726	0,0000	0,0000
1975	0,7341	0,7466	0,4185	0,7341
1976	0,7203	0,7103	0,4106	0,7203
1977	0,0000	-1,1726	0,0000	0,0000
1978	0,0000	-1,1726	0,0000	0,0000
1979	0,7719	0,8453	0,4400	0,7719
1980	0,8038	0,9287	0,4582	0,8038
1981	0,6001	0,3961	0,3420	0,6001
1982	0,8438	1,0333	0,4810	0,8438
1983	0,4625	0,0366	0,2637	0,4625
1984	0,9288	1,2554	0,5294	0,9288
1985	0,0000	-1,1726	0,0000	0,0000
1986	0,0000	-1,1726	0,0000	0,0000
1987	0,8562	1,0657	0,4880	0,8562
1988	0,5222	0,1927	0,2977	0,5222
1989	0,9309	1,2609	0,5306	0,9309
1990	0,5603	0,2922	0,3194	0,5603
1991	0,0000	-1,1726	0,0000	0,0000
1992	0,7561	0,8041	0,4310	0,7561
1993	0,6829	0,6126	0,3892	0,6829
1994	0,0000	-1,1726	0,0000	0,0000
1995	0,7279	0,7304	0,4149	0,7279
1996	0,6153	0,4358	0,3507	0,6153
1997	0,5779	0,3381	0,3294	0,5779
1998	0,8146	0,9571	0,4644	0,8146
1999	0,5959	0,3852	0,3397	0,5959
2000	0,9405	1,2859	0,5361	0,9405
2001	0,8464	1,0401	0,4825	0,8464
2002	0,0000	-1,1726	0,0000	0,0000
2003	0,0000	-1,1726	0,0000	0,0000
2004	0,7199	0,7093	0,4103	0,7199
2005	0,9146	1,2183	0,5213	0,9146
2006	0,0000	-1,1726	0,0000	0,0000
2007	0,5314	0,2167	0,3029	0,5314
2008	0,5279	0,2075	0,3009	0,5279
2009	0,8365	1,0142	0,4768	0,8365
2010	1,0000	1,4416	0,5700	1,0000
2011	0,0330	-1,0864	0,0188	0,0330
2012	0,6776	0,5989	0,3863	0,6776
2013	0,9356	1,2734	0,5333	0,9356
2014	0,6382	0,4958	0,3638	0,6382
2015	0,7139	0,6937	0,4069	0,7139

APÊNDICE 17 – VARIAÇÃO ESTATÍSTICA DO INDICADOR I_7 ALTERADO NORMALIZADO ATRAVÉS DE NORMALIZAÇÃO LINEAR, NORMALIZAÇÃO POR DESVIO PADRÃO, NORMALIZAÇÃO POR ESCALA DECIMAL E NORMALIZAÇÃO POR VALOR MÁXIMO DO ELEMENTO.

Ano	I_7 LINEAR	I_7 DESVIO PADRÃO	I_7 ESCALA DECIMAL	I_7 VALOR MÁXIMO
1931	0,9016	1,1530	0,1704	0,9016
1932	0,6349	0,4675	0,1200	0,6349
1933	0,0000	-1,1647	0,0000	0,0000
1934	0,0000	-1,1647	0,0000	0,0000
1935	0,0000	-1,1647	0,0000	0,0000
1936	0,0000	-1,1647	0,0000	0,0000
1937	0,9317	1,2305	0,1761	0,9317
1938	0,6667	0,5491	0,1260	0,6667
1939	0,8937	1,1326	0,1689	0,8937
1940	0,0000	-1,1647	0,0000	0,0000
1941	0,0000	-1,1647	0,0000	0,0000
1942	0,0000	-1,1647	0,0000	0,0000
1943	0,0000	-1,1647	0,0000	0,0000
1944	0,0000	-1,1647	0,0000	0,0000
1945	0,0000	-1,1647	0,0000	0,0000
1946	0,6302	0,4552	0,1191	0,6302
1947	0,8101	0,9177	0,1531	0,8101
1948	0,9074	1,1679	0,1715	0,9074
1949	0,0000	-1,1647	0,0000	0,0000
1950	0,7820	0,8456	0,1478	0,7820
1951	0,0000	-1,1647	0,0000	0,0000
1952	0,0000	-1,1647	0,0000	0,0000
1953	0,8603	1,0469	0,1626	0,8603
1954	0,4979	0,1152	0,0941	0,4979
1955	0,9185	1,1965	0,1736	0,9185
1956	0,0000	-1,1647	0,0000	0,0000
1957	0,9005	1,1503	0,1702	0,9005
1958	0,0000	-1,1647	0,0000	0,0000
1959	0,0000	-1,1647	0,0000	0,0000
1960	0,0000	-1,1647	0,0000	0,0000
1961	0,8656	1,0605	0,1636	0,8656
1962	0,0000	-1,1647	0,0000	0,0000
1963	0,0000	-1,1647	0,0000	0,0000
1964	0,0000	-1,1647	0,0000	0,0000
1965	0,9407	1,2536	0,1778	0,9407
1966	0,9974	1,3992	0,1885	0,9974
1967	0,0000	-1,1647	0,0000	0,0000
1968	0,0000	-1,1647	0,0000	0,0000
1969	0,0000	-1,1647	0,0000	0,0000
1970	0,9254	1,2142	0,1749	0,9254
1971	0,5989	0,3750	0,1132	0,5989

Ano	I_7 LINEAR	I_7 DESVIO PADRÃO	I_7 ESCALA DECIMAL	I_7 VALOR MÁXIMO
1972	0,4640	0,0281	0,0877	0,4640
1973	0,9439	1,2618	0,1784	0,9439
1974	0,0000	-1,1647	0,0000	0,0000
1975	0,7312	0,7150	0,1382	0,7312
1976	1,0000	1,4060	0,1890	1,0000
1977	0,0000	-1,1647	0,0000	0,0000
1978	0,0000	-1,1647	0,0000	0,0000
1979	0,8476	1,0142	0,1602	0,8476
1980	0,8550	1,0333	0,1616	0,8550
1981	0,8101	0,9177	0,1531	0,8101
1982	0,7106	0,6620	0,1343	0,7106
1983	0,3413	-0,2874	0,0645	0,3413
1984	0,8556	1,0346	0,1617	0,8556
1985	0,0000	-1,1647	0,0000	0,0000
1986	0,0000	-1,1647	0,0000	0,0000
1987	0,6021	0,3831	0,1138	0,6021
1988	0,8921	1,1285	0,1686	0,8921
1989	0,4550	0,0050	0,0860	0,4550
1990	0,5783	0,3219	0,1093	0,5783
1991	0,0000	-1,1647	0,0000	0,0000
1992	0,0423	-1,0559	0,0080	0,0423
1993	0,5561	0,2648	0,1051	0,5561
1994	0,0000	-1,1647	0,0000	0,0000
1995	0,6302	0,4552	0,1191	0,6302
1996	0,9116	1,1788	0,1723	0,9116
1997	0,9392	1,2495	0,1775	0,9392
1998	0,5857	0,3410	0,1107	0,5857
1999	0,6873	0,6021	0,1299	0,6873
2000	0,4545	0,0036	0,0859	0,4545
2001	0,6915	0,6130	0,1307	0,6915
2002	0,0000	-1,1647	0,0000	0,0000
2003	0,0000	-1,1647	0,0000	0,0000
2004	0,8984	1,1448	0,1698	0,8984
2005	0,7751	0,8279	0,1465	0,7751
2006	0,0000	-1,1647	0,0000	0,0000
2007	0,8206	0,9449	0,1551	0,8206
2008	0,9587	1,2999	0,1812	0,9587
2009	0,9122	1,1802	0,1724	0,9122
2010	0,4693	0,0417	0,0887	0,4693
2011	0,7138	0,6701	0,1349	0,7138
2012	0,7667	0,8061	0,1449	0,7667
2013	0,5175	0,1655	0,0978	0,5175
2014	0,3561	-0,2493	0,0673	0,3561
2015	0,6217	0,4335	0,1175	0,6217

APÊNDICE 18 – DISTÂNCIA DE MAHALANOBIS CALCULADA ENTRE O VETOR DE SEVERIDADE DO ANO DE 1931 E O VETOR DE BAIXA SEVERIDADE

```
#Mahalanobis distance between 2 vectors; 1931 Vetor anda low sevetity group
Vector

import numpy as np
import pandas as pd
import math

#import data vector 1 - low severity vector

x=np.array([0.1018,0.1552,0.0010,0.0026,0.0046,0.0025]);axis(I1,I2,I3,I4,I5,
I6,I7)]
print(x)

0.1018  0.1552  0.0001  0.0010  0.0026  0.0046  0.0025]
   I1      I2      I3      I4      I5      I6      I7

#import data vector 2 - vector 1931

y=np.array([0.1929,0.2811,0.0014,0.0347,0.1093,0.274,0.1409]);axis(I1,I2,I3
,I4,I5,I6,I7)]
print(y)

0.1929  0.2811  0.0014  0.0347  0.1093  0.274  0.1409]
   I1      I2      I3      I4      I5      I6      I7

#vector; V1931 - low severity vector

XX = x-y
print(XX)

0,091  0,1259  0,0013  0,0337  0,1067  0,2692  0,1384]
   I1      I2      I3      I4      I5      I6      I7

#transpose matrix XX

Trans=np.traspose(XX)
print (Trans)

0,091  I1
0,1259 I2
0,0013 I3
0,0337 I4
0,1067 I5
0,2692 I6
0,1384 I7

#variance matrizz between x and y 7 indicators of severity

VAR =np.var(x,y)
print (VAR)

var(I1)  0,002766403
var(I2)  0,007925405
var(I3)  0,000000845
var(I4)  0,000567845
var(I5)  0,005692445
var(I6)  0,03623432
var(I7)  0,00957728
```

```
#covariance matrix between x and y 7 indicators of severity conjugated
```

```
COV = np.cov(VAR)
print(VAR)
```

```
cov(I1,I2) 0,003823163
cov(I1,I3) 0,000059215
cov(I1,I4) 0,001535035
cov(I1,I5) 0,004860185
cov(I1,I6) 0,01226206
cov(I1,I7) 0,00630412
cov(I2,I3) 0,000081835
cov(I2,I4) 0,002121415
cov(I2,I5) 0,006716765
cov(I2,I6) 0,01694614
cov(I2,I7) 0,00871228
cov(I3,I4) 0,000021905
cov(I3,I5) 0,000069355
cov(I3,I6) 0,00017498
cov(I3,I7) 0,00008996
cov(I4,I5) 0,001797895
cov(I4,I6) 0,00453602
cov(I4,I7) 0,00233204
cov(I5,I6) 0,01436182
cov(I5,I7) 0,00738364
cov(I6,I7) 0,01862864
```

```
#correlation covariance matrix between 7 x and y severity indicators
```

```
correl.cov=np.correl.cov(VAR;COV)
print(correl.cov)
```

	I1	I2	I3	I4	I5	I6	I7
I1	1,0000	0,0038	0,0001	0,0015	0,0049	0,0123	0,0063
I2	0,0038	1,0000	0,0001	0,0021	0,0067	0,0169	0,0087
I3	0,0001	0,0001	1,0000	0,0000	0,0001	0,0002	0,0001
I4	0,0015	0,0021	0,0000	1,0000	0,0018	0,0045	0,0023
I5	0,0049	0,0067	0,0001	0,001	1,0000	0,0144	0,0074
I6	0,0123	0,0169	0,0002	0,0045	0,0144	1,0000	0,0087
I7	0,0063	0,0087	0,0001	0,0023	0,0074	0,0087	1,0000

```
#correlation inverse correlation for 7 x and y severity indicators
```

```
Invert = np.inverse(correl.cov)
print(Invert)
```

-18,3232	-515,568	36767,61	874,526	-65,4049	-4,08305	-19,3625
-612,601	410,5252	160812,1	-4025,58	-483,463	-22,1846	-89,1517
7871	-280	-14067	3522	4264	1773	7808
-11829,3	4820,05	24219	-7603	-27627,8	-2687,72	-12747,2
-4911,67	1779,4	8900	-17720	-41528,6	-1105,86	-4907,74
-80,25	28,86	1444	-3241,0	-426,32	-18,07	-180,74
-6,10	2,7544	12759,5	405,85	-15,42	-102,19	-6,59

```
# avarage vector XX) multiplie by correlation inverse matrix(Invert)
```

```

TMP = np.array( XX.Invert)

print(TMP)

0,0130 -0,0020 -0,8024 -0,0006 6,8569 -0,0125 -0,0012

# transpose vector multiplied for vector TMP = Mahalanobis^2

Mahalanobis^2 = np.array (Trans.TMP)
Print (Mahalanobis^2)
0,000078

# Mahalanobis Distance is sqr Mahalanobis^2

raiz=math.sqrt (Mahalanobi^2)
print (raiz)

0,008832 # Mahalanonis distance between 1931 Vector and Vector of low
severity group

```

APÊNDICE 19 – DISTÂNCIA DE MAHALANOBIS CALCULADA ENTRE O VETOR DE SEVERIDADE DO ANO DE 1931 E O VETOR DE MÉDIA SEVERIDADE

```
#Mahalanobis distance between 2 vectors; 1931 Vetor and avarage severity
group Vector

import numpy as np
import pandas as pd
import math

#import data vector 1 - vector avarage severity

x=np.array([0.2375,0.3273,0.0259,0.1447,0.4041,0.1166,0.1737];axis(I1,I2,I3
,I4,I5,I6,I7)]
print(x)

[0.2375    0.3273    0.0259    0.1447    0.4041    0.1166    0.1737]
   I1         I2         I3         I4         I5         I6         I7

#import data vector 2 - vector of 1931

y=np.array([0.1929,0.2811,0.0014,0.0347,0.1093,0.274,0.1409]);axis(I1,I2,I3
,I4,I5,I6,I7)]
print(y)

[0.1929    0.2811    0.0014    0.0347    0.1093    0.274    0.1409]
   I1         I2         I3         I4         I5         I6         I7

#VECTOR V1931 - Avarage vector

XX = x-y
print(XX)

[0,0446 0,0462 0,0245 0,11 0,2948 0,1574 0,0328]
   I1         I2         I3         I4         I5         I6         I7

#transpose matrix XX

Trans=np.traspose(XX)
print (Trans)

[0.0446,      I1
0.0462,      I2
0.0245,      I3
0.11,        I4
0.2948       I5
0.1574,      I6
0.0328]      I7

#Variance matrix between 7 severity indicators, vectors x and y

VAR =np.var(x,y)
print (VAR)

var(I1) 0,000663
var(I2) 0,001067
var(I3) 0,0003
var(I4) 0,00605
var(I5) 0,043454
var(I6) 0,012387
var(I7) 0,000538
```

Variance matrix between 7 severity indicators, vectors x and y conjugate

```
COV = np.cov(VAR)
print(VAR)
```

```

cov(I1,I2) 0,0007
cov(I1,I3) 0,0005
cov(I1,I4) 0,0025
cov(I1,I5) 0,0066
cov(I1,I6) -
           0,0035
cov(I1,I7) 0,0007
cov(I2,I3) 0,0006
cov(I2,I4) 0,0025
cov(I2,I5) 0,0068
cov(I2,I6) -
           0,0036
cov(I2,I7) 0,0008
cov(I3,I4) 0,0013
cov(I3,I5) 0,0036
cov(I3,I6) -
           0,0019
cov(I3,I7) 0,0004
cov(I4,I5) 0,0162
cov(I4,I6) -
           0,0087
cov(I4,I7) -
           0,0087
cov(I5,I6) -
           0,0232
cov(I5,I7) 0,0048
cov(I6,I7) -
           0,0026
```

#covariance correlation matrix between 7 severity indicators,vectors x and y

```
correl.cov=np.correl.cov(VAR;COV)
print(correl.cov)
```

	I1	I2	I3	I4	I5	I6	I7
I1	1,0000	0,0007	0,0005	0,0025	0,0066	0,0035	0,0007
I2	0,0007	1,0000	0,0006	0,0025	0,0068	0,0036	0,0008
I3	0,0005	0,0006	1,0000	0,0013	0,0036	0,0019	0,0004
I4	0,0025	0,0025	0,0013	1,0000	0,0162	0,0087	0,0087
I5	0,0066	0,0068	0,0036	0,0162	1,0000	0,0232	0,0048
I6	0,0035	0,0036	0,0019	0,0087	0,0232	1,0000	0,0026
I7	0,0007	0,0008	0,0004	0,0087	0,0048	0,0026	1,0000

```

#inverse correlation matrix 7 indicators, vectors x and y

Invert = np.inverse(correl.cov)
print(Invert)

3,1231 -2914,74 -717,868 0,11063 502,1377 25,9357 -0,063
-2914,7 2805,263 10066,61 -0,5865 -801,219 -63,2341 -0,69
-717,86 10066,61 -9332946 250,1903 754234,9 37009,64 670,88
0,1106 -0,5865 250,1903 -0,00602 -36,6797 -0,4955 145,84
502,137 -801,219 754234,9 -36,6797 -60395,9 -4097,73 -108,2
25,9357 -63,2341 37009,64 -0,4955 -4097,73 1926,274 -3,441
-0,0635 -0,69 670,8833 145,8447 -108,22 -3,4414 -0,037

# avarage vector (XX) plus inverse correlation matrix(Invert)

TMP = np.array( XX.Invert)

print(TMP)

-0,0024 -0,009 -35,9 22,7859 -0,800 5,064 0,0125

# transpose vector multipliyed bt vector TMP = Mahalanobis^2

Mahalanobis^2 = np.array (Trans.TMP)
Print (Mahalanobis^2)
0,0000648795

# Mahalanobis Distance is igual raiz Mahalanobis^2

raiz=math.sqrt (Mahalanobi^2)
print (raiz)

0,0080547 # Mahalanonis distance between 1931 Vector and Avarage Vector

```

APÊNDICE 20 – CÁLCULO DO NÚMERO ÓTIMO DE AGRUPAMENTOS PARA O ALGORITMO K-MEANS. ATRAVÉS DO MÉTODO DO COTOVELO.

```
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
data = pd.read_csv("C:/User/55419/Desktop/Kmeans
Python/Valores_prontos2_originais_Normalização Linear.csv")
data.head ( )
```

Ano	I1	I2	I3	I4	I5	I6	I7
1931	0.1929	0.2811	0.0014	0.0347	0.1093	0.274	0.1409
1932	0.2757	0.3811	0.0204	0.0917	0.1477	0.216	0.4078
1933	0.0585	0.0959	0.0000	0.0000	0.0000	0.000	0.0000
1934	0.0549	0.0905	0.0000	0.0000	0.0000	0.000	0.0000
1935	0.5881	0.6946	0.331	0.4862	0.2263	0.1592	0.4856
1936	0.1469	0.2243	0.0000	0.0000	0.0000	0.0000	0.0000
1937	0.1811	0.2662	0.0000	0.0079	0.219	0.1645	0.1269
1938	0.3955	0.5135	0.1143	0.2413	0.1119	0.3183	0.3762
1939	0.2503	0.3527	0.0274	0.1808	0.4758	0.087	0.1491
1940	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1941	0.1525	0.2324	0.0000	0.0000	0.0000	0.0000	0.0000
1942	0.1373	0.2095	0.0000	0.0000	0.0000	0.0000	0.0000
1943	0.0728	0.1176	0.0000	0.0000	0.0000	0.0000	0.0000
1944	0.0846	0.1351	0.0000	0.0000	0.0000	0.0000	0.0000
1945	0.1417	0.2162	0.0000	0.0000	0.0000	0.0000	0.0000
1946	0.3848	0.5014	0.1437	0.4518	0.4732	0.0901	0.4123
1947	0.2635	0.3676	0.0475	0.2162	0.1304	0.213	0.2329
1948	0.1908	0.2784	0.0016	0.0533	0.5384	0.0785	0.1353
1949	0.0514	0.0851	0.0000	0.0000	0.0000	0.0000	0.0000
1950	0.2271	0.3243	0.0091	0.0863	0.1342	0.288	0.2609
1951	0.1364	0.2081	0.0000	0.0000	0.0000	0.0000	0.0000
1952	0.1144	0.1784	0.0000	0.0000	0.0000	0.0000	0.0000
1953	0.214	0.3181	0.0058	0.0725	0.2785	0.1425	0.1822
1954	0.3998	0.5176	0.1183	0.2245	0.3575	0.1142	0.5446
1955	0.3088	0.4189	0.1265	0.4036	0.1458	0.2243	0.1244
1956	0.1451	0.2216	0.0000	0.0000	0.0000	0.0000	0.0000
1957	0.4591	0.5757	0.5523	1.0000	0.0901	0.4574	0.1421
1958	0.1007	0.1595	0.0000	0.0000	0.0000	0.0000	0.0000
1959	0.0514	0.0851	0.0000	0.0000	0.0000	0.0000	0.0000
1960	0.1144	0.1784	0.0000	0.0000	0.0000	0.0000	0.0000
1961	0.2647	0.3689	0.0321	0.1613	1.0000	0.0427	0.1774
1962	0.0946	0.15	0.0000	0.0000	0.0000	0.0000	0.0000
1963	0.1046	0.1649	0.0000	0.0000	0.0000	0.0000	0.0000
1964	0.0755	0.1216	0.0000	0.0000	0.0000	0.0000	0.0000
1965	0.197	0.2865	0.0045	0.0951	0.1701	0.2304	0.102
1966	0.1811	0.2662	0.002	0.032	0.5875	0.0604	0.0453
1967	0.1399	0.2135	0.0000	0.0000	0.0000	0.0000	0.0000
1968	0.0386	0.0649	0.0000	0.0000	0.0000	0.0000	0.0000
1969	0.1565	0.2378	0.0000	0.0000	0.0000	0.0000	0.0000
1970	0.2021	0.2932	0.0083	0.1254	0.4572	0.091	0.1175
1971	0.4055	0.523	0.1646	0.3347	0.3019	0.1421	0.4437
1972	0.2757	0.3811	0.0719	0.3107	0.0626	0.6005	0.5787
1973	0.2599	0.3635	0.0453	0.3328	0.2982	0.1198	0.0987
1974	0.0989	0.1568	0.0000	0.0000	0.0000	0.0000	0.0000
1975	0.2491	0.3514	0.0276	0.1848	0.292	0.1187	0.3115
1976	0.1855	0.2716	0.001	0.0578	0.3055	0.1176	0.0428
1977	0.1046	0.1649	0.0000	0.0000	0.0000	0.0000	0.0000
1978	0.054	0.0892	0.0000	0.0000	0.0000	0.0000	0.0000
1979	0.2173	0.3122	0.0062	0.0777	0.2552	0.1678	0.1952
1980	0.25	0.3324	0.0239	0.1479	0.2241	0.1464	0.1875
1981	0.2875	0.3743	0.0622	0.2434	0.4228	0.0866	0.2328

```

1982  0.3694  0.4635  0.138   0.305  0.185   0.1216  0.3322
1983  1.0000  1.0000  1.0000  0.988  0.5569  0.0646  0.7013
1984  0.2989  0.3946  0.0478  0.1981  0.1022  0.3746  0.1874
1985  0.0437  0.0608  0.0000  0.0000  0.0000  0.0000  0.0000
1986  0.1501  0.2122  0.0000  0.0000  0.0000  0.0000  0.0000
1987  0.3483  0.4405  0.1057  0.2503  0.173   0.2399  0.4402
1988  0.2595  0.3432  0.027   0.1674  0.4987  0.0816  0.1509
1989  0.2875  0.3743  0.0563  0.1872  0.1001  0.3302  0.5874
1990  0.3548  0.4473  0.0747  0.2163  0.4615  0.0925  0.4642
1991  0.128   0.1824  0.0000  0.0000  0.0000  0.0000  0.0000
1992  0.7378  0.7946  0.3991  0.2784  0.2706  0.1419  1.0000
1993  0.4834  0.5716  0.1826  0.306   0.342   0.1198  0.4866
1994  0.1726  0.2405  0.0000  0.0000  0.0000  0.0000  0.0000
1995  0.4109  0.5041  0.1281  0.2615  0.2981  0.1251  0.4122
1996  0.215   0.2932  0.0067  0.1003  0.4079  0.1028  0.131
1997  0.3982  0.4919  0.2152  0.6358  0.4444  0.0858  0.1035
1998  0.4759  0.5649  0.3438  0.6857  0.2135  0.1472  0.4569
1999  0.2757  0.3622  0.0475  0.1876  0.4268  0.0941  0.3555
2000  0.374   0.4689  0.0862  0.1997  0.0908  0.414   0.5881
2001  0.3068  0.3973  0.0532  0.2045  0.1825  0.1925  0.3513
2002  0.1295  0.1865  0.0000  0.0000  0.0000  0.0000  0.0000
2003  0.1429  0.2054  0.0000  0.0000  0.0000  0.0000  0.0000
2004  0.2033  0.2811  0.0025  0.0595  0.3059  0.1374  0.1443
2005  0.2896  0.3784  0.079   0.4748  0.116   0.3355  0.2676
2006  0.0028  0.0014  0.0000  0.0000  0.0000  0.0000  0.0000
2007  0.2824  0.3716  0.032   0.1545  0.4897  0.0854  0.2219
2008  0.1967  0.2743  0.0022  0.0724  0.4931  0.0786  0.0841
2009  0.323   0.4149  0.1193  0.5111  0.1922  0.1909  0.1304
2010  0.444   0.5378  0.1618  0.2848  0.0327  1.0000  0.5732
2011  0.3967  0.4905  0.177   0.6192  0.9758  0.0426  0.3287
2012  0.2714  0.3595  0.0429  0.1899  0.3471  0.1125  0.2758
2013  0.3974  0.4932  0.1165  0.2196  0.0955  0.3813  0.5253
2014  0.614   0.6905  0.296   0.2795  0.3856  0.1063  0.6866
2015  0.2905  0.3797  0.0673  0.398   0.3117  0.1165  0.421

```

```

continuos_features=["I1","I2","I3","I4","I5","I6","I7"]

```

```

data [ continuos_features].describe( )

```

```

      I1      I2      I3      I4      I5      I6      I7
cont  85      85      85      85      85      85      85
mean  0.244845 0.322231 0.072286 0.165759 0.191935 0.118426 0.196022
std   0.165651 0.178144 0.143081 0.212789 0.217895 0.157646 0.220759
min   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000   0.0000
25%   0.136400 0.205400 0.0000   0.0000   0.0000   0.0000   0.0000
50%   0.215000 0.312200 0.006700 0.091700 0.130400 0.087000 0.131000
75%   0.308800 0.414900 0.079000 0.243400 0.305900 0.147200 0.351300
máx   1.0000   1.0000   1.0000   1.0000   1.0000   1.0000   1.0000

```

```

mms = MinMaxScaler ( )

```

```

mms . fit ( data )

```

```

MinMaxScaler ( copy = True, feature_range (0 , 1))

```

```

data_transformed = mms.transform ( data )

```

```

sum_of_squared_distance = []

```

```

K = range (1, 15)

for k in K:
    km = KMeans (n_clusters = k)
    km = KM.fit (data_transformed)
    Sum_of_squared_distances.append (km.inertia_)

plt.plot ( K, Sum_of_squared_distances, "bx-" )

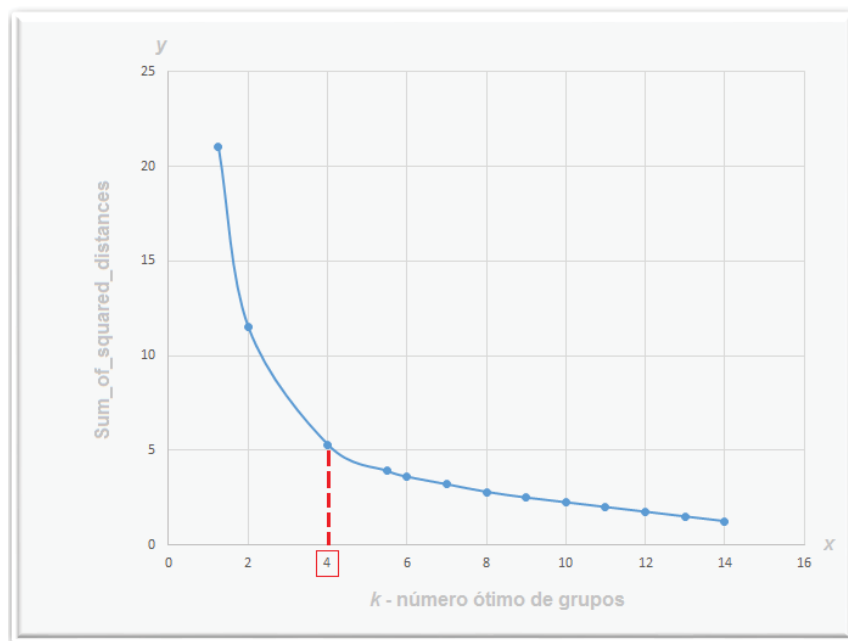
plt.xlabel ('k")

plt.ylabel ( "Sum_of_squared_distances")

plt.title ("Elbow Method For Optimal K")

text (0.5, 1.0, " Elbow Method For Optimal K" )
Plot.show ( )

```



APÊNDICE 21 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORITMO K-MEANS; DADOS ORIGINAIS NORMALIZADOS ATRAVÉS DO MÉTODO DA NORMALIZAÇÃO LINEAR; PROGRAMA EXECUTADO EM LINGUAGEM DE PROGRAMAÇÃO PYTHON.

```
import pandas as pd
import numpy as np
def dist(a,b,ax=1):
    return np.linalg.norm(a-b, axis = ax)
data = pd.read_csv("C:/Users//55419/Desktop/Kmeans
Normalisações/Normalização L_ORIGINAIS/NLVO.csv")
data
```

Ano	I1	I2	I3	I4	I5	I6	I7
1931	0.1929	0.2811	0.0014	0.0347	0.1093	0.274	0.1409
1932	0.2757	0.3811	0.0204	0.0917	0.1477	0.216	0.4078
1933	0.0585	0.0959	0.0000	0.0000	0.0000	0.0000	0.0000
1934	0.0549	0.0905	0.0000	0.0000	0.0000	0.0000	0.0000
1935	0.5881	0.6946	0.331	0.4862	0.2263	0.1592	0.4856
1936	0.1469	0.2243	0.0000	0.0000	0.0000	0.0000	0.0000
1937	0.1811	0.2662	0.0000	0.0079	0.219	0.1645	0.1269
1938	0.3955	0.5135	0.1143	0.2413	0.1119	0.3183	0.3762
1939	0.2503	0.3527	0.0274	0.1808	0.4758	0.087	0.1491
1940	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1941	0.1525	0.2324	0.0000	0.0000	0.0000	0.0000	0.0000
1942	0.1373	0.2095	0.0000	0.0000	0.0000	0.0000	0.0000
1943	0.0728	0.1176	0.0000	0.0000	0.0000	0.0000	0.0000
1944	0.0846	0.1351	0.0000	0.0000	0.0000	0.0000	0.0000
1945	0.1417	0.2162	0.0000	0.0000	0.0000	0.0000	0.0000
1946	0.3848	0.5014	0.1437	0.4518	0.4732	0.0901	0.4123
1947	0.2635	0.3676	0.0475	0.2162	0.1304	0.213	0.2329
1948	0.1908	0.2784	0.0016	0.0533	0.5384	0.0785	0.1353
1949	0.0514	0.0851	0.0000	0.0000	0.0000	0.0000	0.0000
1950	0.2271	0.3243	0.0091	0.0863	0.1342	0.288	0.2609
1951	0.1364	0.2081	0.0000	0.0000	0.0000	0.0000	0.0000
1952	0.1144	0.1784	0.0000	0.0000	0.0000	0.0000	0.0000
1953	0.214	0.3181	0.0058	0.0725	0.2785	0.1425	0.1822
1954	0.3998	0.5176	0.1183	0.2245	0.3575	0.1142	0.5446
1955	0.3088	0.4189	0.1265	0.4036	0.1458	0.2243	0.1244
1956	0.1451	0.2216	0.0000	0.0000	0.0000	0.0000	0.0000
1957	0.4591	0.5757	0.5523	1.0000	0.0901	0.4574	0.1421
1958	0.1007	0.1595	0.0000	0.0000	0.0000	0.0000	0.0000
1959	0.0514	0.0851	0.0000	0.0000	0.0000	0.0000	0.0000
1960	0.1144	0.1784	0.0000	0.0000	0.0000	0.0000	0.0000
1961	0.2647	0.3689	0.0321	0.1613	1.0000	0.0427	0.1774
1962	0.0946	0.15	0.0000	0.0000	0.0000	0.0000	0.0000
1963	0.1046	0.1649	0.0000	0.0000	0.0000	0.0000	0.0000
1964	0.0755	0.1216	0.0000	0.0000	0.0000	0.0000	0.0000
1965	0.197	0.2865	0.0045	0.0951	0.1701	0.2304	0.102

1966	0.1811	0.2662	0.002	0.032	0.5875	0.0604	0.0453
1967	0.1399	0.2135	0.0000	0.0000	0.0000	0.0000	0.0000
1968	0.0386	0.0649	0.0000	0.0000	0.0000	0.0000	0.0000
1969	0.1565	0.2378	0.0000	0.0000	0.0000	0.0000	0.0000
1970	0.2021	0.2932	0.0083	0.1254	0.4572	0.091	0.1175
1971	0.4055	0.523	0.1646	0.3347	0.3019	0.1421	0.4437
1972	0.2757	0.3811	0.0719	0.3107	0.0626	0.6005	0.5787
1973	0.2599	0.3635	0.0453	0.3328	0.2982	0.1198	0.0987
1974	0.0989	0.1568	0.0000	0.0000	0.0000	0.0000	0.0000
1975	0.2491	0.3514	0.0276	0.1848	0.292	0.1187	0.3115
1976	0.1855	0.2716	0.001	0.0578	0.3055	0.1176	0.0428
1977	0.1046	0.1649	0.0000	0.0000	0.0000	0.0000	0.0000
1978	0.054	0.0892	0.0000	0.0000	0.0000	0.0000	0.0000
1979	0.2173	0.3122	0.0062	0.0777	0.2552	0.1678	0.1952
1980	0.25	0.3324	0.0239	0.1479	0.2241	0.1464	0.1875
1981	0.2875	0.3743	0.0622	0.2434	0.4228	0.0866	0.2328
1982	0.3694	0.4635	0.138	0.305	0.185	0.1216	0.3322
1983	1.0000	1.0000	1.0000	0.988	0.5569	0.0646	0.7013
1984	0.2989	0.3946	0.0478	0.1981	0.1022	0.3746	0.1874
1985	0.0437	0.0608	0.0000	0.0000	0.0000	0.0000	0.0000
1986	0.1501	0.2122	0.0000	0.0000	0.0000	0.0000	0.0000
1987	0.3483	0.4405	0.1057	0.2503	0.173	0.2399	0.4402
1988	0.2595	0.3432	0.027	0.1674	0.4987	0.0816	0.1509
1989	0.2875	0.3743	0.0563	0.1872	0.1001	0.3302	0.5874
1990	0.3548	0.4473	0.0747	0.2163	0.4615	0.0925	0.4642
1991	0.128	0.1824	0.0000	0.0000	0.0000	0.0000	0.0000
1992	0.7378	0.7946	0.3991	0.2784	0.2706	0.1419	1.0000
1993	0.4834	0.5716	0.1826	0.306	0.342	0.1198	0.4866
1994	0.1726	0.2405	0.0000	0.0000	0.0000	0.0000	0.0000
1995	0.4109	0.5041	0.1281	0.2615	0.2981	0.1251	0.4122
1996	0.215	0.2932	0.0067	0.1003	0.4079	0.1028	0.131
1997	0.3982	0.4919	0.2152	0.6358	0.4444	0.0858	0.1035
1998	0.4759	0.5649	0.3438	0.6857	0.2135	0.1472	0.4569
1999	0.2757	0.3622	0.0475	0.1876	0.4268	0.0941	0.3555
2000	0.374	0.4689	0.0862	0.1997	0.0908	0.414	0.5881
2001	0.3068	0.3973	0.0532	0.2045	0.1825	0.1925	0.3513
2002	0.1295	0.1865	0.0000	0.0000	0.0000	0.0000	0.0000
2003	0.1429	0.2054	0.0000	0.0000	0.0000	0.0000	0.0000
2004	0.2033	0.2811	0.0025	0.0595	0.3059	0.1374	0.1443
2005	0.2896	0.3784	0.079	0.4748	0.116	0.3355	0.2676
2006	0.0028	0.0014	0.0000	0.0000	0.0000	0.0000	0.0000
2007	0.2824	0.3716	0.032	0.1545	0.4897	0.0854	0.2219
2008	0.1967	0.2743	0.0022	0.0724	0.4931	0.0786	0.0841
2009	0.323	0.4149	0.1193	0.5111	0.1922	0.1909	0.1304
2010	0.444	0.5378	0.1618	0.2848	0.0327	1.0000	0.5732
2011	0.3967	0.4905	0.177	0.6192	0.9758	0.0426	0.3287

```

2012    0.2714  0.3595  0.0429  0.1899  0.3471  0.1125  0.2758
2013    0.3974  0.4932  0.1165  0.2196  0.0955  0.3813  0.5253
2014    0.614   0.6905  0.296   0.2795  0.3856  0.1063  0.6866
2015    0.2905  0.3797  0.0673  0.398   0.3117  0.1165  0.421

```

```

i1 = data['I1'].values
i2 = data['I2'].values
i3 = data['I3'].values
i4 = data['I4'].values
i5 = data['I5'].values
i6 = data['I6'].values
i7 = data['I7'].values

```

```

from sklearn.cluster import KMeans
X = np.array(list(zip(i1,i2,i3,i4,i5,i6,i7)))
kmeans = KMeans(n_clusters = 4)
kmeans = kmeans.fit(X)
labels = kmeans.predict(X)
centroids = kmeans.cluster_centers_
print(centroids)

```

```

[[2.30209524e-01 3.20033333e-01 1.94619048e-02 1.28776190e-01 4.04452381e-
01 1.11728571e-01 1.65128571e-01] [1.02809091e-01 1.56715152e-01
4.24242424e-05 1.05151515e-03 3.31212121e-03 8.30303030e-03 4.26969697e-03]
[3.42277273e-01 4.40600000e-01 9.49590909e-02 2.69359091e-01 1.85240909e-01
2.79581818e-01 3.97104545e-01] [5.61622222e-01 6.44900000e-01 3.84233333e-
01 6.02733333e-01 4.04044444e-01 1.43900000e-01 4.79666667e-01]]

```

```

def mahalanobis(x=None,cov=None)

```

```

    #compute the Mahalanobis Distance Between each of x and the data

```

```

x: vector or matrix of data with.say.p colum

```

```

data:ndarray of the distribution from which Mahalanobis distance of each
observation of x is

```

```

cov:covariance matrix (pxp)of the distribution.If None.will be computed
from data

```

```

"""

```

```

x_minus_mu = x-np.mean(data)

```

```

if not cov:

```

```

    inv_covmat = sp.linalg.inv_covamat

```

```

    left_term = np.dot(x_minus_mu,inv_covmat)

```

```

    mahal = np.dot(left_term,x_minus_mu.T)

```

```

"""

```

```

df_x = df[['I1','I2','I3','I4','I5','I6','I7']].head(85)

```

```

df_x['mahala']=mahalanobis(x=df_x.data=df[['I1','I2','I3','I4','I5','I6','I
7']])

```

```

tabela = df_x['mahala'](np.array(list(zip(i1, i2, i3, i4, i5, i6, i7,
labels)))).

```

```

columns = ['I1', 'I2', 'I3', 'I4', 'I5', 'I6','I7','Centroide'])

```

```

I1.I2.I3.I4.I5.I6.I7.Centroide

```

```

0.0.1929.0.2811.0.0014.0.0347.0.1093.0.27399999999999997.0.1409.2.0

```

```

1.0.2757.0.3811.0.0204.0.0917.0.1477.0.216000000000000003.0.4078.3.0

```

```

2.0.0585.0.0959.0.0.0.0.0.0.0.0.0.0.1.0

```

```

3.0.0549.0.0905.0.0.0.0.0.0.0.0.0.0.1.0

```

```

4.0.5881.0.6946.0.331.0.4862.0.2263.0.1592.0.4856.4.0

```

```

5.0.1469.0.2243.0.0.0.0.0.0.0.0.0.0.1.0

```

```

6.0.1811.0.2662.0.0.0.0.0079.0.21899999999999997.0.1645.0.1269.2.0

```

```

7.0.3955.0.5135.0.1143.0.2413.0.1119.0.3183.0.3762.3.0

```

```

8.0.2503.0.3527.0.0274.0.1808.0.4758.0.087.0.1491.2.0

```



```

59.0.3548.0.4473.0.0747.0.2163.0.4615.0.0925.0.4642.3.0
60.0.128.0.1824.0.0.0.0.0.0.0.0.0.0.1.0
61.0.7378.0.7946.0.3991.0.2784.0.2706.0.1419.1.0.4.0
62.0.4834.0.5716.0.1826.0.306.0.342.0.1198.0.4866.3.0
63.0.1726.0.2405.0.0.0.0.0.0.0.0.0.0.1.0
64.0.4109.0.5041.0.1281.0.2615.0.2981.0.1251.0.4122.3.0
65.0.215.0.2932.0.0067.0.1003.0.4079.0.1028.0.131.2.0
66.0.3982.0.4919.0.2152.0.6358.0.4444.0.0858.0.1035.3.0
67.0.4759.0.5649.0.3438.0.6857.0.2135.0.1472.0.4569.4.0
68.0.2757.0.3622.0.0475.0.1876.0.4268.0.0941.0.3555.2.0
69.0.374.0.4689.0.0862.0.1997.0.0908.0.414.0.5881.3.0
70.0.3068.0.3973.0.0532.0.2045.0.1825.0.1925.0.3513.3.0
71.0.1295.0.1865.0.0.0.0.0.0.0.0.0.0.1.0
72.0.1429.0.2054.0.0.0.0.0.0.0.0.0.0.1.0
73.0.2033.0.2811.0.0025.0.0595.0.3059.0.1374.0.1443.2.0
74.0.2896.0.3784.0.079.0.4748.0.11599.0.3355.0.2676.3.0
75.0.0028.0.0014.0.0.0.0.0.0.0.0.0.0.1.0
76.0.2824.0.3716.0.032.0.1545.0.4897.0.0854.0.2219.2.0
77.0.1967.0.2743.0.0022.0.0724.0.4931.0.0786.0.0841.2.0
78.0.32299999999999995.0.4149.0.1193.0.5111.0.1922.0.1909.0.1304.3.0
79.0.444.0.5378.0.1618.0.2848.0.0327.1.0.0.5732.3.0
80.0.3967.0.4905.0.177.0.6192.0.9758.0.0426.0.3287.3.0
81.0.2714.0.3595.0.0429.0.1899.0.3471.0.1125.0.2758.2.0
82.0.3974.0.4932.0.1165.0.2196.0.0955.0.3813.0.5253.3.0
83.0.614.0.6905.0.296004.0.2795.0.3856.0.1063.0.6866.4.0
84.0.2905.0.3797.0.0673.0.39796.0.3117.0.1165.0.42104.3.0

```

```

tabela.to_csv("C:/Users/55419/Desktop/Arquivos_ULTIMOS/ARQUIVOS
CVS_ORIGINAIS/CSV/NORMALIZAÇÃO LINEAR_ORIGINAIS/Valores
LINEARES_ORIGINAIS_agrupados.csv")

```

```

centroides_df = pd.DataFrame(centroids)

```

```

centroides_df.to_csv("C:/Users/55419/Desktop/Arquivos
10_03_ULTIMOS/ARQUIVOS CVS_ORIGINAIS/CSV/NORMALIZAÇÃO
LINEAR_ORIGINAIS/Valores LINEAR_ORIGINAL_Centróides.csv")

```

APÊNDICE 22 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORITMO K-MEANS; DADOS ORIGINAIS NORMALIZADOS ATRAVÉS DO MÉTODO DO DESVIO PADRÃO; PROGRAMA EXECUTADO EM LINGUAGEM DE PROGRAMAÇÃO PYTHON.

```
import pandas as pd
import numpy as np
def dist(a,b,ax=1):
    return np.linalg.norm(a-b, axis = ax)
data = pd.read_csv("C:/Users//55419/Desktop/Kmeans
Normalisações/Normalização por DP_ORIGINAIS/NPDPVOC.csv")
```

data

Data							
Ano	I1	I2	I3	I4	I5	I6	I7
1931	-0,3138	-0,2303	-0,4795	-0,6116	-0,3801	0,9861	-0,2474
1932	0,1859	0,3309	-0,3468	-0,3408	-0,2015	0,6185	0,9595
1933	-1,1254	-1,2695	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1934	-1,1465	-1,2695	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1935	2,0722	2,098	1,8235	1,5177	0,1639	0,2584	1,3115
1936	-0,5914	-0,5489	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1937	-0,3849	-0,3137	-0,4896	-0,7391	0,1297	0,2919	-0,384
1938	0,9094	1,074	0,3095	0,3654	-0,3683	1,2688	0,8158
1939	0,0333	0,1716	-0,2984	0,0786	1,3246	-0,1985	-0,2115
1940	-1,478	-1,8081	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1941	-0,5572	-0,5034	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1942	-0,6493	-0,6323	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1943	-1,0386	-1,1481	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1944	-0,9676	-1,0495	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1945	-0,6229	-0,5944	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1946	0,84501	1,0061	0,5143	1,3584	1,312	-0,1799	0,981
1947	0,1123	0,2551	-0,1578	0,2432	-0,282	0,5999	0,1668
1948	-0,327	-0,2455	-0,4788	-0,5267	1,615	-0,2531	-0,2738
1949	-1,1675	-1,3302	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1950	-0,1073	0,0124	-0,4261	-0,3674	-0,2644	1,0756	0,2937
1951	-0,6545	-0,6399	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1952	-0,7874	-0,8068	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1953	-0,1862	-0,0786	-0,4492	-0,4311	0,4068	0,1528	-0,0607
1954	0,9357	1,0971	0,3369	0,2857	0,7742	-0,0271	1,5798
1955	0,3859	0,5433	0,3939	0,599	-0,2393	0,6719	-0,3241
1956	-0,6015	-0,5645	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1957	1,2935	1,4233	3,3699	3,9446	-0,4694	2,1498	-0,2427
1958	-0,8702	-0,913	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1959	-1,1675	-1,3302	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1960	-0,7874	-0,8068	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1961	0,1202	0,2627	-0,2653	-0,0169	3,4442	-0,4804	-0,0846
1962	-0,9071	-0,9661	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1963	-0,8466	-0,8826	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1964	-1,0228	-1,1254	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1965	-0,2888	-0,1999	-0,4586	-0,3249	-0,0976	0,7104	-0,4247
1966	-0,3849	-0,3137	-0,4881	-0,6648	1,8436	-0,3674	-0,6809
1967	-0,6335	-0,6096	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869

1968	-1,2451	-1,444	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1969	-0,5335	-0,473	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1970	-0,2586	-0,162	-0,4312	-0,1815	1,238	-0,1737	-0,3552
1971	0,9699	1,1274	0,6608	0,8061	0,5154	0,1504	1,1223
1972	0,1859	0,3309	0,0131	0,6893	-0,5972	3,0576	1,733
1973	0,0912	0,2323	-0,1729	0,7955	0,4983	0,0088	-0,439
1974	-0,8808	-0,9281	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1975	0,0255	0,1641	-0,297	0,09458	0,4695	0,0013	0,5236
1976	-0,3585	-0,2834	-0,4824	-0,5054	0,5324	-0,0048	-0,6929
1977	-0,8466	-0,8826	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1978	-1,1517	-1,3074	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1979	-0,1665	-0,0558	-0,4463	-0,4099	0,2982	0,313	-0,0032
1980	0,0307	0,0579	-0,3222	-0,0753	0,1535	0,1777	-0,0367
1981	0,257	0,293	-0,0546	0,3707	1,0777	-0,202	0,1668
1982	0,7516	0,7937	0,4747	0,6628	-0,028	0,02	0,617
1983	4,558	3,8051	6,4988	3,8915	1,7015	-0,3413	2,2887
1984	0,3267	0,4068	-0,1556	0,1583	-0,4133	1,6245	-0,0391
1985	-1,2136	-1,4667	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1986	-0,5716	-0,6171	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1987	0,6253	0,6647	0,2489	0,4079	-0,0841	0,7701	1,108
1988	0,0886	0,1186	-0,3013	0,0149	1,4307	-0,2333	-0,2043
1989	0,257	0,293	-0,0958	0,1105	-0,423	1,3438	1,7738
1990	0,6634	0,7026	0,0318	0,2432	1,258	-0,165	1,2157
1991	-0,7058	-0,784	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1992	2,9759	2,6521	2,2995	0,5406	0,3697	0,1491	3,6418
1993	1,4395	1,4005	0,7863	0,6681	0,702	0,0088	1,3163
1994	-0,4362	-0,4578	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
1995	1,0015	1,021	0,4054	0,461	0,4976	0,0423	0,981
1996	-0,1797	-0,162	-0,4427	-0,3036	1,0087	-0,0992	-0,2929
1997	0,9252	0,953	1,014	2,2293	1,1784	-0,2072	-0,4175
1998	1,3948	1,3626	1,9136	2,463	0,1043	0,1827	1,1822
1999	0,1859	0,2247	-0,1578	0,1105	1,0967	-0,155	0,7224
2000	0,7792	0,824	0,1126	0,1689	-0,4663	1,8753	1,7762
2001	0,374	0,422	-0,1181	0,1901	-0,0397	0,4695	0,7032
2002	-0,6966	-0,7613	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
2003	-0,6151	-0,6551	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
2004	-0,2507	-0,2303	-0,4723	-0,4948	0,5342	0,1206	-0,2331
2005	0,2701	0,3158	0,0621	1,4646	-0,3489	1,3773	0,3248
2006	-1,4609	-1,8005	-0,4896	-0,7763	-0,8886	-0,7511	-0,8869
2007	0,2307	0,2778	-0,266	-0,0487	1,3891	-0,2097	0,1189
2008	-0,2901	-0,2682	-0,4744	-0,4364	1,4049	-0,2531	-0,5061
2009	0,4714	0,5206	0,3441	1,6399	0,0051	0,4596	-0,2953
2010	1,2028	1,2109	0,6413	0,5672	-0,7364	5,5923	1,7091
2011	0,916	0,9454	0,7473	2,1497	3,6501	-0,4804	0,6026
2012	0,1609	0,2096	-0,1895	0,1211	0,7257	-0,0383	0,3632
2013	0,9213	0,9605	0,3246	0,2592	-0,4445	1,6679	1,4912
2014	2,2288	2,068	1,5789	0,5459	0,9046	-0,0768	2,2216
2015	0,2754	0,3234	-0,0193	1,1035	0,5613	-0,0122	1,0194

```

data['I1'].values
i2 = data['I2'].values
i3 = data['I3'].values
i4 = data['I4'].values
i5 = data['I5'].values
i6 = data['I6'].values
i7 = data['I7'].values

from sklearn.cluster import KMeans
X = np.array(list(zip(i1.i2.i3.i4.i5.i6.i7)))
kmeans = KMeans(n_clusters = 4)
kmeans = kmeans.fit(X)
labels = kmeans.predict(X)
centroids = kmeans.cluster_centers_
print(centroids)

[[-0.874450003,-0.95023875001,-0.48963,-0.77631997,-0.88865,-0.751189996,-
0.8869306]
[0.6128394999,0.691304166,0.223142874,0.582259,0.077790,0.920943417,0.858854125]
[-0.054337,0.018011521,-0.3118177,-0.08850569,1.034572,-0.0168086,-0.11452243]
[2.4206665,2.2336,2.9140751667,2.150598,0.4624593,0.386989834,1.7338934]]

def mahalanobis(x=None,cov=None)
    #compute the Mahalanobis Distance Between each of x and the data
x: vector or matrix of data with.say.p colum
data:ndarray of the distribution from which Mahalanobis distance of each
observation of x is
cov:covariance matrix (pxp)of the distribution.If None.will be computed
from data
"""
x_minus_mu = x-np.mean(data)
if not cov:
    inv_covmat = sp.linalg.inv_covamat
    left_term = np.dot(x_minus_mu,inv_covmat)
    mahal = np.dot(left_term,x_minus_mu.T)
"""

df_x = df[['I1'.'I2'.'I3'.'I4'.'I5'.'I6'.'I7']].head(85)
df_x['mahala']=mahalanobis(x=df_x.data=df[['I1'.'I2'.'I3'.'I4'.'I5'.'I6'.'I
7']])
tabela = df_x['mahala'](np.array(list(zip(i1. i2. i3. i4. i5. i6. i7.
labels)))).
columns = ['I1'. 'I2'. 'I3'. 'I4'. 'I5'. 'I6'.'I7'.'Centroide'])

,I1,I2,I3,I4,I5,I6,I7,Centroide
0,-0.31388,-0.23033,-0.47953,-0.6117,-0.38019,0.986194,-0.247490,2.0
1,0.18598498,0.330994,-0.34681996,-0.34087,-0.20153,0.61850.95954,3.0
2,-1.12548,-1.26954,-0.48963,-0.77632,-0.88865,-0.751189,-0.88692,1.0
3,-1.14654,-1.29988,-0.48963,-0.77632,-0.88865,-0.75118,-0.88692,1.0
4,2.072299,2.090824,1.823505,1.517776,0.163916,0.25456,1.311591,4.0
5,-0.59143,-0.54892,-0.48963,-0.77632,-0.88865,-0.751189,-0.88692,1.0
6,-0.38491,-0.31377,-0.48963,-0.73915,0.129775,0.291986,-0.384,2.0
7,0.909467,1.074371,0.309546,0.365417,-0.36834,1.268099,0.815846,2.0
8,0.033395,0.171699,-0.298490,0.07865,1.324639,-0.19855,-0.21156,2.0
9,-1.47802,-1.8081099,-0.48963,-0.77632,-0.88865,-0.7511,-0.88692,1.0
10,-0.55723,-0.50341,-0.48963,-0.77632,-0.88865,-0.75118,-0.88692,1.0
11,-0.64931,-0.63236,-0.48963,-0.77632,-0.88865,-0.75118,-0.88699,1.0
12,-1.03867,-1.14817001,-0.48963,-0.77632,-0.888,-0.7511,-0.88692,1.0

```

13,-0.96764,-1.04956,-0.48963,-0.77632,-0.88865,-0.7511,-0.886929,1.0
 14,-0.623,-0.59443,-0.48963,-0.77632,-0.88865,-0.7511,-0.886999,1.0
 15,0.84501,1.006101,0.514389,1.358464,1.3121450,-0.17993,0.981094,3.0
 16,0.112322,0.255140,-0.15784,0.243279,-0.28207,0.599971,0.166826,2.0
 17,-0.327030,-0.2455,-0.47881,-0.52673,1.615437,-0.2532,-0.2738,2.0
 18,-1.167579,-1.3302,-0.48963,-0.77632,-0.88865,-0.7511,-0.8869,1.0
 19,-0.10735,0.012404,-0.426160,-0.3674197,-0.2644,1.0756,0.293756,3.0
 20,-0.65457,-0.63995,-0.48963,-0.77632,-0.88865,-0.7511,-0.8869,1.0
 21,-0.78743,-0.80683,-0.48963,-0.77632,-0.88865,-0.7511,-0.8869,1.0
 22,-0.18628,-0.07862,-0.44924,-0.4311,0.406884,0.152897,-0.06069,2.0
 23,0.935775,1.097,0.336955,0.285762,0.774241,-0.0271,1.579821,3.0
 24,0.385929,0.5433,0.393936,0.599076,-0.2393,0.672,-0.3241,3.0
 25,-0.60195,-0.56409,-0.48963,-0.77632,-0.88865,-0.7511,-0.8869,1.0
 26,1.29357,1.423302,3.36992,3.94463,-0.4694,2.14983,-0.2427,4.0
 27,-0.8703,-0.91302,-0.48963,-0.77632,-0.88865,-0.7511,-0.8869299,1.0
 28,-1.1675,-1.3302,-0.48963,-0.77632,-0.88865,-0.7511,-0.8869,1.0
 29,-0.78743,-0.80683,-0.48963,-0.77632,-0.8886,-0.7511,-0.8869,1.0
 30,0.1202,0.262725,-0.26531,-0.01693,3.444261,-0.4804,-0.0846,2.0
 31,-0.9071,-0.9661,-0.48963,-0.77632,-0.88865,-0.7511,-0.8869,1.0
 32,-0.84662,-0.8826,-0.48963,-0.77632,-0.8886,-0.7511,-0.8869,1.0
 33,-1.0228,-1.12542,-0.48963,-0.77632,-0.88865,-0.751189,-0.88692,1.0
 34,-0.28888,-0.19999,-0.45861,-0.32493,-0.09767,0.710498,-0.42472,2.0
 35,-0.38491,-0.31377,-0.48818,-0.6648,1.843682,-0.36745,-0.68097,2.0
 36,-0.63352,-0.6096,-0.48963,-0.77632,-0.88865,-0.7511,-0.8869,1.0
 37,-1.24519,-1.44401,-0.48963,-0.77632,-0.88865,-0.7511,-0.8869,1.0
 38,-0.53355,-0.47307,-0.48963,-0.77632,-0.88865,-0.7511,-0.8869,1.0
 39,-0.25863,-0.16205,-0.4312,-0.18155,1.238053,-0.17372,-0.3552,2.0
 40,0.9699,1.127469,0.6608,0.806182,0.515436,0.1504132,1.1223,3.0
 41,0.1859,0.33099,0.013102,0.689353,-0.5973,3.057639,1.733095,3.0
 42,0.091275,0.232383,-0.17299,0.795561,0.498325,0.008839,-0.43908,2.0
 43,-0.88082,-0.92819,-0.48963,-0.77632,-0.88865,-0.7511,-0.8869,1.0
 44,0.025504,0.164114,-0.29705,0.094588,0.469516,0.00138,0.523667,2.0
 45,-0.3586,-0.28343,-0.48241,-0.50549,0.532467,-0.0048,-0.69294,2.0
 46,-0.84662,-0.8826,-0.48963,-0.77632,-0.88865,-0.7511,-0.8869,1.0
 47,-1.1518,-1.3074,-0.48963,-0.77632,-0.88865,-0.7511,-0.8869,1.0
 48,-0.16655,-0.05586,-0.44635,-0.4099,0.2982,0.313098,-0.00321,2.0
 49,0.0307,0.0579,-0.32229,-0.07535,0.153571,0.1777,-0.036739995,2.0
 50,0.2570,0.293067,-0.0547,0.370729,1.077772,-0.20228,0.166826,2.0
 51,0.7516,0.7937,0.47471,0.662801,-0.02804,0.020016,0.617069,3.0
 52,4.558446,3.80514,6.498818,3.8915,1.701546,-0.34137,2.2887,4.0
 53,0.3267,0.4068,-0.15567,0.15831198,-0.4133,1.624517,-0.03914,4.0
 54,-1.2136,-1.46676,-0.48963,-0.77632,-0.88865,-0.7511,-0.8869,1.0
 55,-0.5717,-0.61719,-0.48963,-0.7763,-0.88865,-0.7511,-0.8869,1.0
 56,0.6253,0.66475,0.2489,0.407901,-0.0841,0.770108,1.108024,2.0
 57,0.0886,0.118601,-0.3013699997,0.0149,1.4307,-0.23333,-0.20438,2.0
 58,0.2570,0.293067,-0.09581,0.110519,-0.42301,1.3438,1.7738,2.0
 59,0.663483,0.7026,0.031855,0.243279,1.2580,-0.16502,1.2157950002,3.0
 60,-0.70587,-0.7840,-0.48963,-0.77632,-0.88865,-0.751189,-0.8869,1.0
 61,2.975994,2.652149,2.299548,0.540662,0.369798,0.149171,3.641835,4.0
 62,1.439582,1.4005,0.7863,0.668111,0.702059,0.008839,1.316381,3.0

```

63,-0.436210,-0.45788,-0.48963,-0.77632,-0.88865,-0.7511,-0.8869,1.0
64,1.00154,1.021272,0.405476,0.461005,0.497689,0.04237,0.981094,3.0
65,-0.1797,-0.16205,-0.44273,-0.30369,1.008773,-0.0992,-0.293,2.0
66,0.9252,0.953003,1.014233,2.22937,1.178445,-0.20725,-0.41753,3.0
67,1.394857,1.362618,1.91366,2.463028,0.104308,0.182702,1.182266,4.0
68,0.1859,0.224798,-0.15784,0.110519,1.0967,-0.15509,0.722445,2.0
69,0.7792,0.82405,0.1126,0.168933,-0.46638,1.875375,1.776203,3.0
70,0.3740,0.42202,-0.1181,0.190175,-0.03974,0.469574,0.703285,3.0
71,-0.6966,-0.76131,-0.48963,-0.77632,-0.88865,-0.7511,-0.8869,1.0
72,-0.61511,-0.65512,-0.48963,-0.77632,-0.88865,-0.7511,-0.8869,1.0
73,-0.25073,-0.23033,-0.4723,-0.49487,0.5342,0.1206,-0.23312,2.0
74,0.2701,0.315823,0.062149,1.464672,-0.34892,1.377384,0.32489,3.0
75,-1.4609,-1.80052,-0.48963,-0.7762,-0.88865,-0.7511899,-0.8869,1.0
76,0.2307,0.2778,-0.2660,-0.04879,1.389102,-0.2097,0.1189,2.0
77,-0.2902,-0.26826,-0.47448,-0.43645,1.404939,-0.2532,-0.5061,2.0
78,0.4714,0.5206,0.344168,1.6399,0.005148,0.459639,-0.29539,3.0
79,1.2028,1.2109,0.6413,0.5672,-0.73649,5.592304,1.7091459,3.0
80,0.9160,0.9454,0.7473,2.149,3.650,-0.4804,0.602699,3.0
81,0.160992,0.209627,-0.18958,0.1211,0.7257,-0.03835,0.3632,2.0
82,0.9213,0.9605,0.3246,0.25921,-0.44458,1.6679,1.4912,3.0
83,2.2288,2.06806,1.5789,0.545972,0.9046,-0.07685,2.22165,4.0
84,0.275434,0.3234,-0.01936,1.1035,0.5613,-0.0122,1.0194,3.0

```

```

tabela.to_csv("C:/Users/55419/Desktop/Arquivos_ULTIMOS/ARQUIVOS
CVS_ORIGINAIS/CSV/NORMALIZAÇÃO DP_ORIGINAIS/Valores
DP_ORIGINAIS_agrupados.csv")

```

```
centroides_df = pd.DataFrame(centroids)
```

```

centroides_df.to_csv("C:/Users/55419/Desktop/Arquivos
10_03_ULTIMOS/ARQUIVOS CVS_ORIGINAIS/CSV/NORMALIZAÇÃO DP_ORIGINAIS/Valores
DP_ORIGINAL_Centróides.csv")

```

APÊNDICE 23 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORITMO K-MEANS; DADOS ORIGINAIS NORMALIZADOS ATRAVÉS DO MÉTODO DA NORMALIZAÇÃO PELO VALOR MÁXIMO DO ELEMENTO; PROGRAMA EXECUTADO EM LINGUAGEM DE PROGRAMAÇÃO PYTHON.

```
import pandas as pd
import numpy as np
def dist(a,b,ax=1):
    return np.linalg.norm(a-b, axis = ax)
data = pd.read_csv("C:/Users//55419/Desktop/Kmeans
Normalisações/Normalização Valor Máximo_ORIGINAIS/NPVM.csv")
data
```

Ano	I1	I2	I3	I4	I5	I6	I7
1931	0.2817	0.4894	0.0014	0.0348	0.1120	0.2738	0.1411
1932	0.3554	0.5604	0.0204	0.0922	0.1513	0.2159	0.4077
1933	0.1621	0.3579	0.0000	0.0000	0.0000	0.0000	0.0000
1934	0.1590	0.3541	0.0000	0.0000	0.0000	0.0000	0.0000
1935	0.6335	0.7831	0.3309	0.4859	0.2319	0.1591	0.4854
1936	0.2408	0.4491	0.0000	0.0000	0.0000	0.0000	0.0000
1937	0.2712	0.4788	0.0000	0.0078	0.2243	0.1644	0.1110
1938	0.4620	0.6545	0.1143	0.2418	0.1146	0.3183	0.3759
1939	0.3329	0.5403	0.0273	0.1811	0.4876	0.0871	0.1491
1940	0.1101	0.2898	0.0000	0.0000	0.0000	0.0000	0.0000
1941	0.2458	0.4548	0.0000	0.0000	0.0000	0.0000	0.0000
1942	0.2323	0.4385	0.0000	0.0000	0.0000	0.0000	0.0000
1943	0.1749	0.3733	0.0000	0.0000	0.0000	0.0000	0.0000
1944	0.1853	0.3857	0.0000	0.0000	0.0000	0.0000	0.0000
1945	0.2361	0.4433	0.0000	0.0000	0.0000	0.0000	0.0000
1946	0.4525	0.6458	0.1436	0.4521	0.4848	0.0900	0.4124
1947	0.3445	0.5508	0.0474	0.2159	0.1336	0.2129	0.2326
1948	0.2798	0.4875	0.0015	0.0528	0.5517	0.0785	0.1353
1949	0.1559	0.3502	0.0000	0.0000	0.0000	0.0000	0.0000
1950	0.3121	0.5201	0.0090	0.0866	0.1375	0.2879	0.2607
1951	0.2315	0.4376	0.0000	0.0000	0.0000	0.0000	0.0000
1952	0.2119	0.4165	0.0000	0.0000	0.0000	0.0000	0.0000
1953	0.3005	0.5086	0.0057	0.0731	0.2854	0.1425	0.1824
1954	0.4659	0.6573	0.1182	0.2249	0.3663	0.1141	0.5446
1955	0.3849	0.5873	0.1264	0.2913	0.1430	0.2243	0.1242
1956	0.2392	0.4472	0.0000	0.0000	0.0000	0.0000	0.0000
1957	0.5187	0.6986	0.5522	1	0.0923	0.4573	0.1422
1958	0.1997	0.4030	0.0000	0.0000	0.0000	0.0000	0.0000
1959	0.1559	0.3502	0.0000	0.0000	0.0000	0.0000	0.0000
1960	0.2119	0.4165	0.0000	0.0000	0.0000	0.0000	0.0000
1961	0.3457	0.5518	0.0320	0.1608	0.9546	0.0426	0.1771
1962	0.1942	0.3963	0.0000	0.0000	0.0000	0.0000	0.0000
1963	0.2032	0.4069	0.0000	0.0000	0.0000	0.0000	0.0000
1964	0.1772	0.3761	0.0000	0.0000	0.0000	0.0000	0.0000
1965	0.2854	0.4932	0.0044	0.0956	0.1742	0.2304	0.1020

1966	0.2712	0.4788	0.0002	0.0236	0.6019	0.0604	0.0454
1967	0.2346	0.4414	0.0000	0.0000	0.0000	0.0000	0.0000
1968	0.1444	0.3358	0.0000	0.0000	0.0000	0.0000	0.0000
1969	0.2493	0.4587	0.0000	0.0000	0.0000	0.0000	0.0000
1970	0.2898	0.4980	0.0083	0.1259	0.4685	0.0910	0.1173
1971	0.4710	0.6612	0.1646	0.3352	0.3093	0.1421	0.4436
1972	0.3554	0.5604	0.0719	0.3104	0.0641	0.6004	0.5785
1973	0.3414	0.5479	0.0453	0.3329	0.3055	0.1198	0.0988
1974	0.1981	0.4011	0.0000	0.0000	0.0000	0.0000	0.0000
1975	0.3317	0.5393	0.0275	0.1844	0.2992	0.1186	0.3114
1976	0.2751	0.4827	0.0010	0.0573	0.3131	0.1176	0.0428
1977	0.2032	0.4069	0.0000	0.0000	0.0000	0.0000	0.0000
1978	0.1582	0.3531	0.0000	0.0000	0.0000	0.0000	0.0000
1979	0.3034	0.5115	0.0061	0.0776	0.2615	0.1677	0.1951
1980	0.3325	0.5259	0.0239	0.1484	0.2296	0.1464	0.1877
1981	0.3659	0.5556	0.0622	0.2429	0.4332	0.0865	0.2326
1982	0.4388	0.6190	0.1379	0.3048	0.1896	0.1215	0.3320
1983	1	1	1	0.9887	0.5706	0.0646	0.7012
1984	0.3761	0.5700	0.0477	0.1979	0.1047	0.3745	0.1872
1985	0.1491	0.3330	0.0000	0.0000	0.0000	0.0000	0.0000
1986	0.2437	0.4404	0.0000	0.0000	0.0000	0.0000	0.0000
1987	0.4202	0.6026	0.1056	0.2508	0.1772	0.2398	0.4405
1988	0.3410	0.5335	0.0269	0.1676	0.5110	0.0816	0.1507
1989	0.3659	0.5556	0.0563	0.1878	0.1025	0.3302	0.5875
1990	0.4258	0.6074	0.0746	0.2159	0.4729	0.0924	0.4643
1991	0.2239	0.4193	0.0000	0.0000	0.0000	0.0000	0.0000
1992	0.7667	0.8541	0.3991	0.2789	0.2772	0.1419	1
1993	0.5402	0.6957	0.1825	0.3059	0.3504	0.1198	0.4865
1994	0.2637	0.4606	0.0000	0.0000	0.0000	0.0000	0.0000
1995	0.4756	0.6477	0.1280	0.2620	0.3054	0.1250	0.4124
1996	0.3015	0.4980	0.0067	0.1001	0.4180	0.1027	0.1311
1997	0.4644	0.6391	0.2151	0.6366	0.4554	0.0857	0.1036
1998	0.5336	0.6909	0.3438	0.6861	0.2187	0.1472	0.4569
1999	0.3554	0.5470	0.0474	0.1878	0.4374	0.0939	0.3553
2000	0.4428	0.6228	0.0861	0.2002	0.0930	0.4140	0.5880
2001	0.3831	0.5719	0.0531	0.2047	0.1870	0.1924	0.3511
2002	0.2253	0.4222	0.0000	0.0000	0.0000	0.0000	0.0000
2003	0.2373	0.4357	0.0000	0.0000	0.0000	0.0000	0.0000
2004	0.2910	0.4894	0.0024	0.0596	0.3134	0.1374	0.1443
2005	0.3678	0.5585	0.0789	0.4746	0.1189	0.3355	0.2675
2006	0.1126	0.2907	0.0000	0.0000	0.0000	0.0000	0.0000
2007	0.3620	0.5537	0.0319	0.1541	0.5018	0.0853	0.2221
2008	0.2852	0.4846	0.0021	0.0719	0.5053	0.0785	0.0840
2009	0.3975	0.5844	0.1193	0.5118	0.1969	0.1908	0.1306
2010	0.5053	0.6717	0.1618	0.2845	0.0335	1	0.5732
2011	0.4630	0.6381	0.1770	0.6197	1	0.0426	0.3289

```

2012    0.3517  0.5451  0.0429  0.1901  0.3556  0.1123  0.2760
2013    0.4638  0.6401  0.1165  0.2193  0.0978  0.3813  0.5251
2014    0.6565  0.7802  0.2960  0.2800  0.3951  0.1063  0.6864
2015    0.3686  0.5595  0.0672  0.3982  0.3194  0.1164  0.4209

```

```

i1 = data['I1'].values
i2 = data['I2'].values
i3 = data['I3'].values
i4 = data['I4'].values
i5 = data['I5'].values
i6 = data['I6'].values
i7 = data['I7'].values

```

```

from sklearn.cluster import KMeans
X = np.array(list(zip(i1,i2,i3,i4,i5,i6,i7)))
kmeans = KMeans(n_clusters = 4)
kmeans = kmeans.fit(X)
labels = kmeans.predict(X)
centroids = kmeans.cluster_centers_
print(centroids)

```

```

[[0.4452625.0.6257874999999999.0.12155833333333332.0.28087916666666667.0.21
61875.0.258325.0.44052916666666664]
[0.20157878787878786.0.40106969696969696.4.2424242424288217e-
05.0.001054545454545455302.0.00339393939393939255.0.008296969696969747.0.00427
5757575757627]
[0.31483636363636364.0.5168772727272727.0.01885.0.12645681818181817.0.3986
4999999999995.0.11968636363636363.0.16873636363636366]
[0.6022.0.7416333333333334.0.4365.0.7361666666666667.0.42815000000000003.0
.15941666666666665.0.36970000000000003]

```

```

def mahalanobis(x=None,cov=None)
    #compute the Mahalanobis Distance Between each of x and the data
    x: vector or matrix of data with.say.p colum
    data:ndarray of the distribution from which Mahalanobis distance of each
    observation of x is
    cov:covariance matrix (pxp)of the distribution.If None.will be computed
    from data
    """
    x_minus_mu = x-np.mean(data)
    if not cov:
        inv_covmat = sp.linalg.inv_covamat
        left_term = np.dot(x_minus_mu.inv_covmat)
        mahal = np.dot(left_term,x_minus_mu.T)
    """
    df_x = df[['I1','I2','I3','I4','I5','I6','I7']].head(85)
    df_x['mahala']=mahalanobis(x=df_x.data=df[['I1','I2','I3','I4','I5','I6','I
7']])
    tabela = df_x['mahala'](np.array(list(zip(i1. i2. i3. i4. i5. i6. i7.
labels)))).
    columns = ['I1'. 'I2'. 'I3'. 'I4'. 'I5'. 'I6'.'I7'.'Centroide'])

```

```

I1.I2.I3.I4.I5.I6.I7.Centroide
0.0.2817.0.4894.0.0014.0.0348.0.112.0.2738.0.1411.1.0
1.0.3554.0.5604.0.0204.0.0922.0.1513.0.2159.0.4077.3.0
2.0.1621.0.3579.0.0.0.0.0.0.0.0.0.0.1.0
3.0.159.0.3541.0.0.0.0.0.0.0.0.0.0.1.0
4.0.6335.0.7831.0.3309.0.4859.0.2319.0.1591.0.4854.4.0
5.0.2408.0.4491.0.0.0.0.0.0.0.0.0.0.1.0
6.0.2712.0.4788.0.0.0.0.0078.0.2243.0.1644.0.111.2.0

```

7.0.462.0.6545.0.1143.0.2418.0.1146.0.3183.0.3759.3.0
8.0.3329.0.5403.0.0273.0.1811.0.4876.0.0871.0.1491.2.0
9.0.1101.0.2898.0.0.0.0.0.0.0.0.0.0.1.0
10.0.2458.0.4548.0.0.0.0.0.0.0.0.0.0.1.0
11.0.2323.0.4385.0.0.0.0.0.0.0.0.0.0.1.0
12.0.1749.0.3733.0.0.0.0.0.0.0.0.0.0.1.0
13.0.1853.0.3857.0.0.0.0.0.0.0.0.0.0.1.0
14.0.2361.0.4433.0.0.0.0.0.0.0.0.0.0.1.0
15.0.4525.0.6458.0.1436.0.4521.0.4848.0.09.0.4124.3.0
16.0.3445.0.5508.0.0474.0.2159.0.1336.0.2129.0.2326.3.0
17.0.2798.0.4875.0.0015.0.0528.0.5517.0.0785.0.1353.2.0
18.0.1559.0.3502.0.0.0.0.0.0.0.0.0.0.1.0
19.0.3121.0.5201.0.009.0.0866.0.1375.0.2879.0.2607.2.0
20.0.2315.0.4376.0.0.0.0.0.0.0.0.0.0.1.0
21.0.2119.0.4165.0.0.0.0.0.0.0.0.0.0.1.0
22.0.3005.0.5086.0.0057.0.0731.0.2854.0.1425.0.1824.2.0
23.0.4659.0.6573.0.1182.0.2249.0.3663.0.1141.0.5446.3.0
24.0.3849.0.5873.0.1264.0.2913.0.143.0.2243.0.1242.3.0
25.0.2392.0.4472.0.0.0.0.0.0.0.0.0.0.1.0
26.0.5187.0.6986.0.5522.1.0.0.0923.0.4573.0.1422.4.0
27.0.1997.0.403.0.0.0.0.0.0.0.0.0.0.1.0
28.0.1559.0.3502.0.0.0.0.0.0.0.0.0.0.1.0
29.0.2119.0.4165.0.0.0.0.0.0.0.0.0.0.1.0
30.0.3457.0.5518.0.032.0.16085.0.9546.0.0426.0.1771.2.0
31.0.1942.0.3963.0.0.0.0.0.0.0.0.0.0.1.0
32.0.2032.0.4069.0.0.0.0.0.0.0.0.0.0.1.0
33.0.1772.0.3761.0.0.0.0.0.0.0.0.0.0.1.0
34.0.2854.0.4932.0.0044.0.0956.0.1742.0.2304.0.102.2.0
35.0.2712.0.4788.0.0002.0.0236.0.6019.0.0604.0.0454.2.0
36.0.2346.0.4414.0.0.0.0.0.0.0.0.0.0.1.0
37.0.1444.0.3358.0.0.0.0.0.0.0.0.0.0.1.0
38.0.2493.0.4587.0.0.0.0.0.0.0.0.0.0.1.0
39.0.2898.0.498.0.0083.0.1259.0.4685.0.091.0.1173.2.0
40.0.471.0.6612.0.1646.0.3352.0.3093.0.1421.0.4436.3.0
41.0.3554.0.5604.0.0719.0.3104.0.0641.0.6004.0.5785.3.0
42.0.3414.0.5479.0.0453.0.3329.0.3055.0.1198.0.0988.2.0
43.0.1981.0.4011.0.0.0.0.0.0.0.0.0.0.1.0
44.0.3317.0.5393.0.0275.0.1844.0.2992.0.1186.0.3114.2.0
45.0.2751.0.4827.0.001.0.0573.0.3131.0.1176.0.0428.2.0
46.0.2032.0.4069.0.0.0.0.0.0.0.0.0.0.1.0
47.0.1582.0.3531.0.0.0.0.0.0.0.0.0.0.1.0
48.0.3034.0.5115.0.0061.0.0776.0.2615.0.1677.0.1951.2.0
49.0.3325.0.5259.0.0239.0.1484.0.2296.0.1464.0.1877.2.0
50.0.3659.0.5556.0.0622.0.2429.0.4332.0.0865.0.2326.2.0
51.0.4388.0.619.0.1379.0.3048.0.1896.0.1215.0.332.3.0
52.1.0.1.0.1.0.0.9887.0.5706.0.0646.0.7012.4.0
53.0.3761.0.57.0.0477.0.1979.0.1047.0.3745.0.1872.3.0
54.0.1491.0.333.0.0.0.0.0.0.0.0.0.0.1.0
55.0.2437.0.4404.0.0.0.0.0.0.0.0.0.0.1.0
56.0.4202.0.6026.0.1056.0.2508.0.1772.0.2398.0.4405.3.0

57.0.341.0.5335.0.0269.0.1676.0.511.0.0816.0.1507.2.0
58.0.3659.0.5556.0.0563.0.1878.0.1025.0.3302.0.5875.3.0
59.0.4258.0.6074.0.0746.0.2159.0.4729.0.0924.0.4643.3.0
60.0.2239.0.4193.0.0.0.0.0.0.0.0.0.0.0.0.0.1.0
61.0.7667.0.8541.0.3991.0.2789.0.2772.0.1419.1.0.3.0
62.0.5402.0.6957.0.1825.0.3059.0.3504.0.1198.0.4865.3.0
63.0.2637.0.4606.0.0.0.0.0.0.0.0.0.0.0.0.0.1.0
64.0.4756.0.6477.0.128.0.262.0.3054.0.125.0.4124.3.0
65.0.3015.0.498.0.0067.0.1001.0.418.0.1027.0.1311.2.0
66.0.4644.0.6391.0.2151.0.6366.0.4554.0.0857.0.1036.4.0
67.0.5336.0.6909.0.3438.0.6861.0.2187.0.1472.0.4569.4.0
68.0.3554.0.547.0.0474.0.1878.0.4374.0.0939.0.3553.2.0
69.0.4428.0.6228.0.0861.0.2002.0.093.0.414.0.588.3.0
70.0.3831.0.5719.0.0531.0.2047.0.187.0.1924.0.3511.3.0
71.0.2253.0.4222.0.0.0.0.0.0.0.0.0.0.0.0.0.1.0
72.0.2373.0.4357.0.0.0.0.0.0.0.0.0.0.0.0.0.1.0
73.0.291.0.4894.0.0024.0.0596.0.3134.0.1374.0.1443.2.0
74.0.3678.0.5585.0.0789.0.4746.0.1189.0.3355.0.2675.3.0
75.0.1126.0.2907.0.0.0.0.0.0.0.0.0.0.0.0.0.1.0
76.0.362.0.5537.0.0319.0.1541.0.5018.0.0853.0.2221.2.0
77.0.2852.0.4846.0.0021.0.0719.0.5053.0.0785.0.084.2.0
78.0.3975.0.5844.0.1193.0.5118.0.1969.0.1908.0.1306.3.0
79.0.5053.0.6717.0.1618.0.2845.0.0335.1.0.0.5732.3.0
80.0.463.0.6381.0.177.0.6197.1.0.0.0426.0.3289.4.0
81.0.3517.0.5451.0.0429.0.1901.0.3556.0.1123.0.276.2.0
82.0.4638.0.6401.0.1165.0.2193.0.0978.0.3813.0.5251.3.0
83.0.6565.0.7802.0.296.0.28.0.3951.0.1063.0.6864.4.0
84.0.3686.0.5595.0.0672.0.3982.0.3194.0.1164.0.4209.3.0

APÊNDICE 24 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORITMO K-MEANS; DADOS ORIGINAIS NORMALIZADOS ATRAVÉS DO MÉTODO DA ESCALA DECIMAL; PROGRAMA EXECUTADO EM LINGUAGEM DE PROGRAMAÇÃO PYTHON.

```
import pandas as pd
import numpy as np
def dist(a,b,ax=1):
    return np.linalg.norm(a-b, axis = ax)
data = pd.read_csv("C:/Users//55419/Desktop/Kmeans
Normalizações/Normalização por Escala Decimal_ORIGINAIS/NPVM.csv")
data
```

Ano	I1	I2	I3	I4	I5	I6	I7
1931	0.1453	0.051	0.0014	0.31	0.6389	0.1399	0.0267
1932	0.1833	0.0584	0.0198	0.82	0.8634	0.1103	0.0771
1933	0.0836	0.0373	0.0000	0.0000	0.0000	0.0000	0.0000
1934	0.082	0.0369	0.0000	0.0000	0.0000	0.0000	0.0000
1935	0.3267	0.0816	0.3207	4.32	1.3226	0.0813	0.0918
1936	0.1242	0.0468	0.0000	0.0000	0.0000	0.0000	0.0000
1937	0.1399	0.0499	0.0000	0.07	1.2797	0.084	0.02156
1938	0.2383	0.0682	0.1108	2.15	0.6538	0.1626	0.07118
1939	0.1717	0.0563	0.0265	1.61	2.7811	0.0445	0.0282
1940	0.0568	0.0302	0.0000	0.0000	0.0000	0.0000	0.0000
1941	0.1268	0.0474	0.0000	0.0000	0.0000	0.0000	0.0000
1942	0.1198	0.0457	0.0000	0.0000	0.0000	0.0000	0.0000
1943	0.0902	0.0389	0.0000	0.0000	0.0000	0.0000	0.0000
1944	0.0956	0.0402	0.0000	0.0000	0.0000	0.0000	0.0000
1945	0.1218	0.0462	0.0000	0.0000	0.0000	0.0000	0.0000
1946	0.2334	0.0673	0.1392	4.02	2.7654	0.046	0.078
1947	0.1777	0.0574	0.046	1.92	0.7622	0.1088	0.044
1948	0.1443	0.0508	0.0015	0.47	3.1465	0.0401	0.0256
1949	0.0804	0.0365	0.0000	0.0000	0.0000	0.0000	0.0000
1950	0.161	0.0542	0.0088	0.77	0.7844	0.1471	0.0493
1951	0.1194	0.0456	0.0000	0.0000	0.0000	0.0000	0.0000
1952	0.1093	0.0434	0.0000	0.0000	0.0000	0.0000	0.0000
1953	0.155	0.053	0.0056	0.65	1.6279	0.0728	0.0345
1954	0.2403	0.0685	0.1146	2	2.0895	0.0583	0.103
1955	0.1985	0.0612	0.1225	2.59	0.8159	0.1146	0.0235
1956	0.1234	0.0466	0.0000	0.0000	0.0000	0.0000	0.0000
1957	0.2675	0.0728	0.5351	8.89	0.5267	0.2336	0.0269
1958	0.103	0.042	0.0000	0.0000	0.0000	0.0000	0.0000
1959	0.0804	0.0365	0.0000	0.0000	0.0000	0.0000	0.0000
1960	0.1093	0.0434	0.0000	0.0000	0.0000	0.0000	0.0000
1961	0.1783	0.0575	0.0311	1.43	5.4445	0.0218	0.0335
1962	0.1002	0.0413	0.0000	0.0000	0.0000	0.0000	0.0000
1963	0.1048	0.0424	0.0000	0.0000	0.0000	0.0000	0.0000
1964	0.0914	0.0392	0.0000	0.0000	0.0000	0.0000	0.0000
1965	0.1472	0.0514	0.0043	0.85	0.9939	0.1177	0.0193

1966	0.1399	0.0499	0.0002	0.21	3.4333	0.0309	0.0086
1967	0.121	0.046	0.0000	0.0000	0.0000	0.0000	0.0000
1968	0.0745	0.035	0.0000	0.0000	0.0000	0.0000	0.0000
1969	0.1286	0.0478	0.0000	0.0000	0.0000	0.0000	0.0000
1970	0.1495	0.0519	0.0081	1.12	2.6723	0.0465	0.0222
1971	0.2429	0.0689	0.1595	2.98	1.7643	0.0726	0.0839
1972	0.1833	0.0584	0.0697	2.76	0.3661	0.3067	0.1094
1973	0.1761	0.0571	0.0439	2.96	1.7428	0.0612	0.0187
1974	0.1022	0.0418	0.0000	0.0000	0.0000	0.0000	0.0000
1975	0.1711	0.0562	0.0267	1.64	1.7066	0.0606	0.0589
1976	0.1419	0.0503	0.001	0.51	1.7857	0.0601	0.0081
1977	0.1048	0.0424	0.0000	0.0000	0.0000	0.0000	0.0000
1978	0.0816	0.0368	0.0000	0.0000	0.0000	0.0000	0.0000
1979	0.1565	0.0533	0.006	0.69	1.4914	0.0857	0.0369
1980	0.1715	0.0548	0.0232	1.32	1.3096	0.0748	0.0355
1981	0.1887	0.0579	0.0603	2.16	2.4709	0.0442	0.044
1982	0.2263	0.0645	0.1337	2.71	1.0814	0.0621	0.0628
1983	0.5157	0.1042	0.9689	8.79	3.2547	0.033	0.1326
1984	0.194	0.0594	0.0463	1.76	0.5973	0.1913	0.0354
1985	0.0769	0.0347	0.0000	0.0000	0.0000	0.0000	0.0000
1986	0.1257	0.0459	0.0000	0.0000	0.0000	0.0000	0.0000
1987	0.2167	0.0628	0.1024	2.23	1.0109	0.1225	0.0833
1988	0.1759	0.0556	0.0261	1.49	2.9144	0.0417	0.0285
1989	0.1887	0.0579	0.0546	1.67	0.5851	0.1687	0.1111
1990	0.2196	0.0633	0.0723	1.92	2.6974	0.0472	0.0878
1991	0.1155	0.0437					
1992	0.3954	0.089	0.3867	2.48	1.5813	0.0725	0.1891
1993	0.2786	0.0725	0.1769	2.72	1.9988	0.0612	0.092
1994	0.136	0.048	0.0000	0.0000	0.0000	0.0000	0.0000
1995	0.2453	0.0675	0.1241	2.33	1.742	0.0639	0.078
1996	0.1555	0.0519	0.0065	0.89	2.3842	0.0525	0.0248
1997	0.2395	0.0666	0.2085	5.66	2.5974	0.0438	0.0196
1998	0.2752	0.072	0.3332	6.1	1.2477	0.0752	0.0864
1999	0.1833	0.057	0.046	1.67	2.4947	0.048	0.0672
2000	0.2284	0.0649	0.0835	1.78	0.5306	0.2115	0.1112
2001	0.1976	0.0596	0.0515	1.82	1.0667	0.0983	0.0664
2002	0.1162	0.044	0.0000	0.0000	0.0000	0.0000	0.0000
2003	0.1224	0.0454	0.0000	0.0000	0.0000	0.0000	0.0000
2004	0.1501	0.051	0.0024	0.53	1.7879	0.0702	0.0273
2005	0.1897	0.0582	0.0765	4.22	0.6782	0.1714	0.0506
2006	0.0581	0.0303	0.0000	0.0000	0.0000	0.0000	0.0000
2007	0.1867	0.0577	0.031	1.37	2.8621	0.0436	0.042
2008	0.1471	0.0505	0.0021	0.64	2.882	0.0401	0.0159
2009	0.205	0.0609	0.1156	4.55	1.1231	0.0975	0.0247
2010	0.2606	0.07	0.1568	2.53	0.1912	0.5108	0.1084
2011	0.2388	0.0665	0.1715	5.51	5.7032	0.0218	0.0622

2012	0.1814	0.0568	0.0416	1.69	2.0286	0.0574	0.0522
2013	0.2392	0.0667	0.1129	1.95	0.558	0.1948	0.0993
2014	0.3386	0.0813	0.2868	2.49	2.2534	0.0543	0.1298
2015	0.1901	0.0583	0.0652	3.54	1.822	0.0595	0.0796

```
i1 = data['I1'].values
i2 = data['I2'].values
i3 = data['I3'].values
i4 = data['I4'].values
i5 = data['I5'].values
i6 = data['I6'].values
i7 = data['I7'].values
```

```
from sklearn.cluster import KMeans
X = np.array(list(zip(i1,i2,i3,i4,i5,i6,i7)))
kmeans = KMeans(n_clusters = 4)
kmeans = kmeans.fit(X)
labels = kmeans.predict(X)
centroids = kmeans.cluster_centers_
print(centroids)
```

```
[[0.2235157894736842,0.3891578947368421,0.000000947368421,0.0010547105
2631579,0.002899999999999996,0.008489473684212,0.0000315789473684],[0.3144757
5757575757,0.5132393939393939,0.01852424242424203,0.122233333333333834,0.3
3333333333,0.1512060606060606,0.1548606060606013],[0.4298000000000006,0.
6009548,0.1197752,0.2746884,0.220540037,0.2413133299999997,0.435011745],[0.
6011262825,0.73120702583333334,0.425533333333336,0.721366666666666,0.42823
8229500000000005,0.1493071425,0.3581833333333333]]
```

```
def mahalanobis(x=None,cov=None)
    #compute the Mahalanobis Distance Between each of x and the data
x: vector or matrix of data with say p colum
data:ndarray of the distribution from which Mahalanobis distance of each
observation of x is
cov:covariance matrix (pxp)of the distribution.If None.will be computed
from data
"""
x_minus_mu = x-np.mean(data)
if not cov:
    inv_covmat = sp.linalg.inv_covamat
    left_term = np.dot(x_minus_mu,inv_covmat)
    mahal = np.dot(left_term,x_minus_mu.T)
"""
df_x = df[['I1','I2','I3','I4','I5','I6','I7']].head(85)
df_x['mahala']=mahalanobis(x=df_x.data=df[['I1','I2','I3','I4','I5','I6','I
7']])
tabela = df_x['mahala'](np.array(list(zip(i1, i2, i3, i4, i5, i6, i7,
labels)))).
columns = ['I1', 'I2', 'I3', 'I4', 'I5', 'I6','I7','Centroide']
```

```
I1.I2.I3.I4.I5.I6.I7.Centroide
0.0.1453.0.0510.0.0014.0.3100.0.6389.0.1399.0.0267.2.0
1.0.1833.0.0584.0.0198.0.8200.0.8634.0.1103.0.0771.2.0
2.0.0836.0.0373.0.0000.0.0000.0.0000.0.0000.0.0000.1.0
3.0.0820.0.0369.0.0000.0.0000.0.0000.0.0000.0.0000.1.0
4.0.3267.0.0816.0.3207.4.3200.1.3226.0.0813.0.0918.4.0
5.0.1242.0.0468.0.0000.0.0000.0.0000.0.0000.0.0000.1.0
```

6.0.1399.0.0499.0.0000.0.0700.1.2797.0.0840.0.0210.2.0
7.0.2383.0.0682.0.1108.2.1500.0.6538.0.1626.0.0711.3.0
8.0.1717.0.0563.0.0265.1.6100.2.7811.0.0445.0.0282.2.0
9.0.0568.0.0302.0.0000.0.0000.0.0000.0.0000.1.0
10.0.1268.0.0474.0.0000.0.0000.0.0000.0.0000.1.0
11.0.1198.0.0457.0.0000.0.0000.0.0000.0.0000.1.0
12.0.0902.0.0389.0.0000.0.0000.0.0000.0.0000.1.0
13.0.0956.0.0402.0.0000.0.0000.0.0000.0.0000.1.0
14.0.1218.0.0462.0.0000.0.0000.0.0000.0.0000.1.0
15.0.2334.0.0673.0.1392.4.0200.2.7654.0.0460.0.0780.3.0
16.0.1777.0.0574.0.0460.1.9200.0.7622.0.1088.0.0440.3.0
17.0.1443.0.0508.0.0015.0.4700.3.1465.0.0401.0.0256.2.0
18.0.0804.0.0365.0.0000.0.0000.0.0000.0.0000.1.0
19.0.1610.0.0542.0.0088.0.7700.0.7844.0.1471.0.0493.2.0
20.0.1194.0.0456.0.0000.0.0000.0.0000.0.0000.1.0
21.0.1093.0.0434.0.0000.0.0000.0.0000.0.0000.1.0
22.0.1550.0.0530.0.0056.0.6500.1.6279.0.0728.0.0345.2.0
23.0.2403.0.0685.0.1146.2.0000.2.0895.0.0583.0.1030.3.0
24.0.1985.0.0612.0.1225.2.5900.0.8159.0.1146.0.0235.3.0
25.0.1234.0.0466.0.0000.0.0000.0.0000.0.0000.1.0
26.0.2675.0.0728.0.5351.8.8900.0.5267.0.2336.0.0269.4.0
27.0.1030.0.0420.0.0000.0.0000.0.0000.0.0000.1.0
28.0.0804.0.0365.0.0000.0.0000.0.0000.0.0000.1.0
29.0.1093.0.0434.0.0000.0.0000.0.0000.0.0000.1.0
30.0.1783.0.0575.0.0311.1.4300.5.4445.0.0218.0.0335.2.0
31.0.1002.0.0413.0.0000.0.0000.0.0000.0.0000.1.0
32.0.1048.0.0424.0.0000.0.0000.0.0000.0.0000.1.0
33.0.0914.0.0392.0.0000.0.0000.0.0000.0.0000.1.0
34.0.1472.0.0514.0.0043.0.8500.0.9939.0.1177.0.0193.2.0
35.0.1399.0.0499.0.0002.0.2100.3.4333.0.0309.0.0086.2.0
36.0.2210.0.0460.0.0000.0.0000.0.0000.0.0000.1.0
37.0.0745.0.0350.0.0000.0.0000.0.0000.0.0000.1.0
38.0.1286.0.0478.0.0000.0.0000.0.0000.0.0000.1.0
39.0.1495.0.0519.0.0081.1.1200.2.6723.0.0465.0.0222.2.0
40.0.2429.0.0689.0.1595.2.9800.1.7643.0.0726.0.0839.3.0
41.0.1833.0.0584.0.0697.2.7600.0.3661.0.3067.0.1094.3.0
42.0.1761.0.0571.0.0439.2.9600.1.7428.0.0612.0.0187.2.0
43.0.1022.0.0418.0.0000.0.0000.0.0000.0.0000.1.0
44.0.1711.0.0562.0.0267.1.6400.1.7066.0.0606.0.0589.2.0
45.0.1419.0.0503.0.0010.0.5100.1.7857.0.0601.0.0081.2.0
46.0.1048.0.0424.0.0000.0.0000.0.0000.0.0000.1.0
47.0.0816.0.0368.0.0000.0,0000.0,0000.0,0000.1.0

```

48.0.1565.0.0533.0.0060.0.6900.1.4914.0.0857.0,0369.2.0
49.0.1715.0.0548.0.0232.1.3200.1.3096.0.0748.0.0355.2.0
50.0.1887.0.0579.0.0603.2.1600.2.4709.0.0442.0.0440.2.0
51.0.2263.0.0645.0.1337.2.7100.1.0814.0.0621.0.0628.3.0
52.0.5157.0.1042.0.9689.8.7900.3.2547.0.0330.0.1326.4.0
53.0.1940.0.0594.0.0463.1.7600.0.5973.0.1913.0.0354.3.0
54.0.0769.0.0347.0.0000.0.0000.0.0000.0.0000.0.0000.1.0
55.0.1257.0.0459.0.0000.0.0000.0.0000.0.0000.0.0000.1.0
56.0.2167.0.0628.0.1024.2.2300.1.0109.0.1225.0.0833.3.0
57.0.1759.0.0556.0.0261.1.4900.2.9144.0.0417.0.0285.2.0
58.0.1887.0.0579.0.0546.1.6700.0.5851.0.1687.0.1111.3.0
59.0.2196.0.0633.0.0723.1.9200.2.6974.0.0472.0.0878.3.0
60.0.1155.0.0437.0.0000.0.0000.0.0000.0.0000.0.0000.1.0
61.0.3954.0.0890.0.3867.2.4800.1.5813.0.0725.0.1891.4.0
62.0.2786.0.0725.0.1769.2.7200.1.9988.0.0612.0.0920.3.0
63.0.1360.0.0480.0.0000.0.0000.0.0000.0.0000.0.0000.1.0
64.0.2453.0.0675.0.1241.2.3300.1.7420.0.0639.0.0780.3.0
65.0.1555.0.0519.0.0065.0.8900.2.3842.0.0525.0.0248.2.0
66.0.2395.0.0666.0.2085.5.6600.2.5974.0.0438.0.0196.4.0
67.0.2752.0.0720.0.3332.6.1000.1.2477.0.0752.0.0864.4.0
68.0.1833.0.0570.0.0460.1.6700.2.4947.0.0480.0.0672.2.0
69.0.2284.0.0649.0.0835.1.7800.0.5306.0.2115.0.1112.3.0
70.0.1976.0.0596.0.0515.1.8200.1.0667.0.0983.0.0664.3.0
71.0.1162.0.0440.0.0000.0.0000.0.0000.0.0000.0.0000.1.0
72.0.1224.0.0454.0.0000.0.0000.0.0000.0.0000.0.0000.1.0
73.0.1501.0.0510.0.0024.0.5300.1.7879.0.0702.0.0273.2.0
74.0.1897.0.0582.0.0765.4.2200.0.6782.0.1714.0.0506.3.0
75.0.0581.0.0303.0.0000.0.0000.0.0000.0.0000.0.0000.1.0
76.0.1867.0.0577.0.0310.1.3700.2.8621.0.0436.0.0420.2.0
77.0.1471.0.0505.0.0021.0.6400.2.8820.0.0401.0.0159.2.0
78.0.2050.0.0609.0.1156.4.5500.1.1231.0.0975.0.0247.3.0
79.0.2606.0.0700.0.1568.2.5300.0.1912.0.5108.0.1084.3.0
80.0.2388.0.0665.0.1715.5.5100.5.7032.0.0218.0.0622.3.0
81.0.1814.0.0568.0.0416.1.6900.2.0286.0.0574.0.0522.2.0
82.0.2392.0.0667.0.1129.1.9500.0.5580.0.1948.0.0993.3.0
83.0.3386.0.0813.0.2868.2.4900.2.2534.0.0543.0.1298.4.0
84.0.1901.0.0583.0.0652.3.5400.1.8220.0.0595.0.0796.3.0
tabela.to_csv("C:/Users/55419/Desktop/Arquivos_ULTIMOS/ARQUIVOS
CVS_ORIGINAIS/CSV/NORMALIZAÇÃO ED/Valores ED_Agrupados.csv")
centroides_df = pd.DataFrame(centroids)

centroides_df.to_csv("C:/Users/55419/Desktop/Arquivos
10_03_ULTIMOS/ARQUIVOS CVS NORMALIZAÇÃO ED_ORIGINAIS/Valores
ED_ORIGINAIS_Centróides.csv")

```

APÊNDICE 25 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORITMO K-MEANS; DADOS DE I_5 E I_7 ALTERADOS. NORMALIZADOS ATRAVÉS DO MÉTODO DA NORMALIZAÇÃO LINEAR; PROGRAMA EXECUTADO EM LINGUAGEM DE PROGRAMAÇÃO PYTHON.

```
import pandas as pd
import numpy as np
def dist(a,b,ax=1):
    return np.linalg.norm(a-b, axis = ax)
data = pd.read_csv("C:/Users//55419/Desktop/Kmeans
Normalizações/Normalização Linear_ALTERADOS/NPLADA.csv")
data
```

ANO	I1	I2	I3	I4	I5	I6	I7
1931	0.1929	0.2811	0.0014	0.0347	0.9214	0.274	0.9015
1932	0.2757	0.3811	0.0204	0.0917	0.8820	0.216	0.6349
1933	0.0585	0.0959	0.0000	0.0000	0.0000	0.0000	0.0000
1934	0.0549	0.0905	0.0000	0.0000	0.0000	0.0000	0.0000
1935	0.5881	0.6946	0.331	0.4862	0.8015	0.1592	0.0000
1936	0.1469	0.2243	0.0000	0.0000	0.0000	0.0000	0.0000
1937	0.1811	0.2662	0.0000	0.0079	0.8090	0.1645	0.9317
1938	0.3955	0.5135	0.1143	0.2413	0.9188	0.3183	0.6666
1939	0.2503	0.3527	0.0274	0.1808	0.5456	0.087	0.8936
1940	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1941	0.1525	0.2324	0.0000	0.0000	0.0000	0.0000	0.0000
1942	0.1373	0.2095	0.0000	0.0000	0.0000	0.0000	0.0000
1943	0.0728	0.1176	0.0000	0.0000	0.0000	0.0000	0.0000
1944	0.0846	0.1351	0.0000	0.0000	0.0000	0.0000	0.0000
1945	0.1417	0.2162	0.0000	0.0000	0.0000	0.0000	0.0000
1946	0.3848	0.5014	0.1437	0.4518	0.5483	0.0901	0.6301
1947	0.2635	0.3676	0.0475	0.2162	0.8998	0.213	0.81
1948	0.1908	0.2784	0.0016	0.0533	0.4815	0.0785	0.9074
1949	0.0514	0.0851	0.0000	0.0000	0.0000	0.0000	0.0000
1950	0.2271	0.3243	0.0091	0.0863	0.8959	0.288	0.782
1951	0.1364	0.2081	0.0000	0.0000	0.0000	0.0000	0.0000
1952	0.1144	0.1784	0.0000	0.0000	0.0000	0.0000	0.0000
1953	0.214	0.3181	0.0058	0.0725	0.7479	0.1425	0.8603
1954	0.3998	0.5176	0.1183	0.2245	0.6669	0.1142	0.4978
1955	0.3088	0.4189	0.1265	0.4036	0.8904	0.2243	0.9185
1956	0.1451	0.2216	0.0000	0.0000	0.0000	0.0000	0.0000
1957	0.4591	0.5757	0.5523	1	0.9411	0.4574	0.9005
1958	0.1007	0.1595	0.0000	0.0000	0.0000	0.0000	0.0000
1959	0.0514	0.0851	0.0000	0.0000	0.0000	0.0000	0.0000
1960	0.1144	0.1784	0.0000	0.0000	0.0000	0.0000	0.0000
1961	0.2647	0.3689	0.0321	0.1613	0.0783	0.0427	0.8656
1962	0.0946	0.15	0.0000	0.0000	0.0000	0.0000	0.0000
1963	0.1046	0.1649	0.0000	0.0000	0.0000	0.0000	0.0000
1964	0.0755	0.1216	0.0000	0.0000	0.0000	0.0000	0.0000
1965	0.197	0.2865	0.0045	0.0951	0.8591	0.2304	0.9407
1966	0.1811	0.2662	0.002	0.032	0.4312	0.0604	0.9973
1967	0.1399	0.2135	0.0000	0.0000	0.0000	0.0000	0.0000

1968	0.0386	0.0649	0.0000	0.0000	0.0000	0.0000	0.0000
1969	0.1565	0.2378	0.0000	0.0000	0.0000	0.0000	0.0000
1970	0.2021	0.2932	0.0083	0.1254	0.5647	0.091	0.9253
1971	0.4055	0.523	0.1646	0.3347	0.7240	0.1421	0.5989
1972	0.2757	0.3811	0.0719	0.3107	0.9693	0.6005	0.464
1973	0.2599	0.3635	0.0453	0.3328	0.7277	0.1198	0.9439
1974	0.0989	0.1568	0.0000	0.0000	0.0000	0.0000	0.0000
1975	0.2491	0.3514	0.0276	0.1848	0.7341	0.1187	0.7312
1976	0.1855	0.2716	0.001	0.0578	0.7202	0.1176	1
1977	0.1046	0.1649	0.0000	0.0000	0.0000	0.0000	0.0000
1978	0.054	0.0892	0.0000	0.0000	0.0000	0.0000	0.0000
1979	0.2173	0.3122	0.0062	0.0777	0.7718	0.1678	0.8476
1980	0.25	0.3324	0.0239	0.1479	0.8037	0.1464	0.855
1981	0.2875	0.3743	0.0622	0.2434	0.6000	0.0866	0.81
1982	0.3694	0.4635	0.138	0.305	0.8438	0.1216	0.7105
1983	1	1	1	0.988	0.4625	0.0646	0.3412
1984	0.2989	0.3946	0.0478	0.1981	0.9287	0.3746	0.8555
1985	0.0437	0.0608	0.0000	0.0000	0.0000	0.0000	0.0000
1986	0.1501	0.2122	0.0000	0.0000	0.0000	0.0000	0.0000
1987	0.3483	0.4405	0.1057	0.2503	0.8561	0.2399	0.6021
1988	0.2595	0.3432	0.027	0.1674	0.5222	0.0816	0.892
1989	0.2875	0.3743	0.0563	0.1872	0.9308	0.3302	0.455
1990	0.3548	0.4473	0.0747	0.2163	0.5603	0.0925	0.5783
1991	0.128	0.1824	0.0000	0.0000	0.0000	0.0000	0.0000
1992	0.7378	0.7946	0.3991	0.2784	0.7561	0.1419	0.0423
1993	0.4834	0.5716	0.1826	0.306	0.6828	0.1198	0.556
1994	0.1726	0.2405	0.0000	0.0000	0.0000	0.0000	0.0000
1995	0.4109	0.5041	0.1281	0.2615	0.7279	0.1251	0.6301
1996	0.215	0.2932	0.0067	0.1003	0.6152	0.1028	0.9116
1997	0.3982	0.4919	0.2152	0.6358	0.5778	0.0858	0.9391
1998	0.4759	0.5649	0.3438	0.6857	0.8146	0.1472	0.5857
1999	0.2757	0.3622	0.0475	0.1876	0.5958	0.0941	0.6873
2000	0.374	0.4689	0.0862	0.1997	0.9404	0.414	0.4544
2001	0.3068	0.3973	0.0532	0.2045	0.8464	0.1925	0.6915
2002	0.1295	0.1865	0.0000	0.0000	0.0000	0.0000	0.0000
2003	0.1429	0.2054	0.0000	0.0000	0.0000	0.0000	0.0000
2004	0.2033	0.2811	0.0025	0.0595	0.7198	0.1374	0.8984
2005	0.2896	0.3784	0.079	0.4748	0.9145	0.3355	0.7751
2006	0.0028	0.0014	0.0000	0.0000	0.0000	0.0000	0.0000
2007	0.2824	0.3716	0.032	0.1545	0.5314	0.0854	0.8206
2008	0.1967	0.2743	0.0022	0.0724	0.5279	0.0786	0.9587
2009	0.323	0.4149	0.1193	0.5111	0.8365	0.1909	0.9121
2010	0.444	0.5378	0.1618	0.2848	1	1	0.4693
2011	0.3967	0.4905	0.177	0.6192	0.0329	0.0426	0.7137
2012	0.2714	0.3595	0.0429	0.1899	0.6776	0.1125	0.7666
2013	0.3974	0.4932	0.1165	0.2196	0.9356	0.3813	0.5174
2014	0.614	0.6905	0.296	0.2795	0.6382	0.1063	0.356
2015	0.2905	0.3797	0.0673	0.398	0.7138	0.1165	0.6216

```

i1 = data['I1'].values
i2 = data['I2'].values
i3 = data['I3'].values
i4 = data['I4'].values
i5 = data['I5'].values
i6 = data['I6'].values
i7 = data['I7'].values

from sklearn.cluster import KMeans
X = np.array(list(zip(i1.i2.i3.i4.i5.i6.i7)))
kmeans = KMeans(n_clusters = 4)
kmeans = kmeans.fit(X)
labels = kmeans.predict(X)
centroids = kmeans.cluster_centers_
print(centroids)

[[0.09999375000000002,0.152828125,4.163336342344337e-17,-
2.7755575615628914e-17,0.7356666666666667,0.0,0.8522399999999999]
[0.254756,0.346072,0.040772,0.185096,0.8602454545454545,0.106292,0.66021818
18181818][0.33491818181818184,0.4330454545454545,0.09162727272727272,0.2611
045454545454,0.59136,0.17943333333333333,0.37095]
[0.6458166666666666,0.7200499999999999,0.4870333333333333,0.619633333333333
4,5.551115123125783e-17,0.2878318181818182,2.220446049250313e-16]]

def mahalanobis(x=None,cov=None)
    #compute the Mahalanobis Distance Between each of x and the data
x: vector or matrix of data with say p column
data: ndarray of the distribution from which Mahalanobis distance of each
observation of x is
cov:covariance matrix (pxp)of the distribution.If None.will be computed
from data
"""
x_minus_mu = x-np.mean(data)
if not cov:
    inv_covmat = sp.linalg.inv_covamat
    left_term = np.dot(x_minus_mu.inv_covmat)
    mahal = np.dot(left_term.x_minus_mu.T)
"""

df_x = df[['I1'. 'I2'. 'I3'. 'I4'. 'I5'. 'I6'. 'I7']].head(85)
df_x['mahala']=mahalanobis(x=df_x.data=df[['I1'. 'I2'. 'I3'. 'I4'. 'I5'. 'I6'. 'I
7']])
tabela = df_x['mahala'](np.array(list(zip(i1. i2. i3. i4. i5. i6. i7.
labels)))).
columns = ['I1'. 'I2'. 'I3'. 'I4'. 'I5'. 'I6'. 'I7'. 'Centroide'])

, I1, I2, I3, I4, I5, I6, I7, Centroide
0,0.1929,0.2811,0.0014,0.0347,0.9214,0.273999997,0.9015,3.0
1,0.2757,0.3811,0.0204,0.0917,0.882,0.2160000003,0.6349,3.0
2,0.0585,0.0959,0.0,0.0,0.0,0.0,0.0,1.0
3,0.0549,0.0905,0.0,0.0,0.0,0.0,0.0,1.0
4,0.5881,0.6946,0.331,0.4862,0.8015,0.1592,0.0,4.0
5,0.1469,0.2243,0.0,0.0,0.0,0.0,0.0,1.0
6,0.1811,0.2662,0.0,0.0079,0.809,0.1645,0.9317,2.0
7,0.3955,0.5135,0.1143,0.2413,0.9188,0.3183,0.6666,3.0
8,0.2503,0.3527,0.0274,0.1808,0.5456,0.087,0.8936,2.0
9,0.0,0.0,0.0,0.0,0.0,0.0,0.0,1.0
10,0.1525,0.2324,0.0,0.0,0.0,0.0,0.0,1.0
11,0.1373,0.2095,0.0,0.0,0.0,0.0,0.0,1.0
12,0.0728,0.1176,0.0,0.0,0.0,0.0,0.0,1.0

```

13,0.0846,0.1351,0.0,0.0,0.0,0.0,0.0,1.0
 14,0.1417,0.2162,0.0,0.0,0.0,0.0,0.0,1.0
 15,0.3848,0.5014,0.1437,0.4518,0.5483,0.0901,0.6301,2.0
 16,0.2635,0.3676,0.0475,0.2162,0.8998,0.213,0.81,3.0
 17,0.1908,0.2784,0.0016,0.0533,0.4815,0.0785,0.9074,2.0
 18,0.0514,0.0851,0.0,0.0,0.0,0.0,0.0,1.0
 19,0.2271,0.3243,0.0091,0.0863,0.8959,0.28800003,0.782,3.0
 20,0.1364,0.2081,0.0,0.0,0.0,0.0,0.0,1.0
 21,0.1144,0.1784,0.0,0.0,0.0,0.0,0.0,1.0
 22,0.214,0.3181,0.0058,0.0725,0.7479,0.1425,0.8603,2.0
 23,0.3998,0.5176,0.1183,0.2245,0.6669,0.1142,0.4978,3.0
 24,0.3088,0.4189,0.1265,0.4036,0.8904,0.2243,0.9185,3.0
 25,0.1451,0.2216,0.0,0.0,0.0,0.0,0.0,1.0
 26,0.4591,0.5757,0.5523,1.0,0.9411,0.4574,0.9005,4.0
 27,0.1007,0.1595,0.0,0.0,0.0,0.0,0.0,1.0
 28,0.0514,0.0851,0.0,0.0,0.0,0.0,0.0,1.0
 29,0.1144,0.1784,0.0,0.0,0.0,0.0,0.0,1.0
 30,0.2647,0.3689,0.0321,0.1613,0.0783,0.0427,0.8656,2.0
 31,0.0946,0.15,0.0,0.0,0.0,0.0,0.0,1.0
 32,0.1046,0.1649,0.0,0.0,0.0,0.0,0.0,1.0
 33,0.0755,0.1216,0.0,0.0,0.0,0.0,0.0,1.0
 34,0.1969999998,0.2865,0.0045,0.0951,0.8591,0.2304,0.9407,2.0
 35,0.1811,0.2662,0.002,0.032,0.4312,0.0604,0.9973,2.0
 36,0.1399,0.2135,0.0,0.0,0.0,0.0,0.0,1.0
 37,0.0386,0.0649,0.0,0.0,0.0,0.0,0.0,1.0
 38,0.1565,0.2378,0.0,0.0,0.0,0.0,0.0,1.0
 39,0.2021,0.2932,0.0083,0.1254,0.5647,0.091,0.9253,2.0
 40,0.4055,0.523,0.1646,0.3347,0.724,0.1421,0.5989,3.0
 41,0.2757,0.3811,0.0719,0.3107,0.9693,0.6005,0.46399999997,3.0
 42,0.2599,0.3635,0.0453,0.3328,0.7277,0.1198,0.9439,2.0
 43,0.0989,0.1568,0.0,0.0,0.0,0.0,0.0,1.0
 44,0.2491,0.3514,0.0276,0.1848,0.7341,0.1187,0.7312,2.0
 45,0.1855,0.2716,0.001,0.0578,0.7202,0.1176,1.0,2.0
 46,0.1046,0.1649,0.0,0.0,0.0,0.0,0.0,1.0
 47,0.05400000006,0.0892,0.0,0.0,0.0,0.0,0.0,1.0
 48,0.2173,0.3122,0.0062,0.0777,0.7718,0.1678,0.8476,1.0
 49,0.25,0.3324,0.0239,0.1479,0.8037,0.1464,0.855,2.0
 50,0.2875,0.3743,0.0622,0.2434,0.6,0.0866,0.81,2.0
 51,0.3694,0.4635,0.138,0.305,0.8438,0.1216,0.7105,2.0
 52,1.0,1.0,1.0,0.988,0.4625,0.0646,0.3412,4.0
 53,0.2989,0.3946,0.0478,0.1981,0.9287,0.3746,0.8555,3.0
 54,0.0437,0.0608,0.0,0.0,0.0,0.0,0.0,1.0
 55,0.1501,0.2122,0.0,0.0,0.0,0.0,0.0,1.0
 56,0.3483,0.4405,0.1057,0.2503,0.8561,0.2399,0.6021,3.0
 57,0.2595,0.3432,0.02700000003,0.1674,0.5222,0.0816,0.892,2.0
 58,0.2875,0.3743,0.0563,0.1872,0.9308,0.3302,0.455,3.0
 59,0.3548,0.4473,0.0747,0.2163,0.5603,0.0925,0.5783,2.0
 60,0.128,0.1824,0.0,0.0,0.0,0.0,0.0,1.0
 61,0.7378,0.7946,0.3991,0.2784,0.7561,0.1419,0.0423,4.0
 62,0.4834,0.5716,0.1826,0.306,0.6828,0.1198,0.556,3.0

```

63,0.1726,0.2405,0.0,0.0,0.0,0.0,0.0,1.0
64,0.4109,0.5041,0.1281,0.2615,0.7279,0.1251,0.6301,3.0
65,0.215,0.2932,0.0067,0.1003,0.6152,0.1028,0.9116,2.0
66,0.3982,0.4919,0.2152,0.6358,0.5778,0.0858,0.9391,2.0
67,0.4759,0.5649,0.3438,0.6857,0.8146,0.1472,0.5857,4.0
68,0.2757,0.3622,0.0475,0.1876,0.5958,0.0941,0.6873,2.0
69,0.374,0.4689,0.0862,0.1997,0.9404,0.414,0.4544,3.0
70,0.3068,0.3973,0.0532,0.2045,0.846,0.1925,0.6915,3.0
71,0.1295,0.1865,0.0,0.0,0.0,0.0,0.0,1.0
72,0.1429,0.2054,0.0,0.0,0.0,0.0,0.0,1.0
73,0.2033,0.2811,0.0025,0.0595,0.7198,0.1374,0.8984,2.0
74,0.2896,0.3784,0.079,0.4748,0.9145,0.3355,0.7751,3.0
75,0.0028,0.0014,0.0,0.0,0.0,0.0,0.0,1.0
76,0.2824,0.3716,0.032,0.1545,0.5314,0.0854,0.8206,2.0
77,0.1967,0.2743,0.0022,0.0724,0.5279,0.0786,0.9587,2.0
78,0.32299999999999995,0.4149,0.1193,0.5111,0.8365,0.1909,0.9121,3.0
79,0.444,0.5378,0.1618,0.2848,1.0,1.0,0.4693,3.0
80,0.3967,0.4905,0.177,0.6192,0.0329,0.0426,0.7137,2.0
81,0.2714,0.3595,0.0429,0.1899,0.6776,0.1125,0.7666,2.0
82,0.3974,0.4932,0.1165,0.2196,0.9356,0.3813,0.5174,3.0
83,0.614,0.6905,0.2960004,0.2795,0.6382,0.1063,0.3560000004,4.0
84,0.2905,0.3797,0.0673,0.3979999996,0.7138,0.1165,0.6216,3.0

```

```

tabela.to_csv("C:/Users/55419/Desktop/Arquivos_ULTIMOS/ARQUIVOS
CVS_ALTERADOS/CSV/NORMALIZAÇÃO LINEAR_I5eI7_ALTERADOS/Valores
LINEARES_ALTERADOS_agrupados.csv")

```

```

centroides_df = pd.DataFrame(centroids)

```

```

centroides_df.to_csv("C:/Users/55419/Desktop/Arquivos
10_03_ULTIMOS/ARQUIVOS CVS_ALTERADOS/CSV/NORMALIZAÇÃO
LINEAR_ALTERADOS/Valores LINEAR_I5eI7_ALTERADOS _Centróides.csv")

```

APÊNDICE 26 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORITMO K-MEANS; DADOS DE I_5 E I_7 ALTERADOS. NORMALIZADOS ATRAVÉS DO MÉTODO DO DESVIO PADRÃO; PROGRAMA EXECUTADO EM LINGUAGEM DE PROGRAMAÇÃO PYTHON.

```
import pandas as pd
import numpy as np
def dist(a,b,ax=1):
    return np.linalg.norm(a-b, axis = ax)
data = pd.read_csv("C:/Users//55419/Desktop/Kmeans
Normalizações/Normalização por DP_ALTERADOS/NPDPDA.csv")
data
```

Ano	I1	I2	I3	I4	I5	I6	I7
1931	-0.3157	-0.2316	-0.4823	-0.6153	1.2362	0.9920	1.1529
1932	0.1870	0.3329	-0.3488	-0.3428	1.1333	0.6222	0.4674
1933	-1.1321	-1.2770	-0.4925	-0.7809	-1.1725	-0.755	-1.1647
1934	-1.1533	-1.3075	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
1935	2.0845	2.1032	1.8343	1.5267	0.9227	0.2599	-1.1647
1936	-0.5949	-0.5521	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
1937	-0.3871	-0.3156	-0.4925	-0.7435	0.9423	0.2937	1.2305
1938	0.9148	1.0807	0.3113	0.3675	1.2294	1.2756	0.5490
1939	0.0335	0.1727	-0.3002	0.0791	0.2538	-0.1997	1.1325
1940	-1.4867	-1.8188	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
1941	-0.5605	-0.5063	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
1942	-0.653	-0.6361	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
1943	-1.0448	-1.1549	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
1944	-0.9733	-1.0557	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
1945	-0.6266	-0.5979	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
1946	0.8500	1.0120	0.5174	1.3665	0.2610	-0.1809	0.4552
1947	0.1129	0.2566	-0.1587	0.2447	1.1797	0.6035	0.9176
1948	-0.3289	-0.2469	-0.4816	-0.52985	0.0862	-0.2546	1.1679
1949	-1.1745	-1.3381	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
1950	-0.1079	0.0124	-0.4286	-0.3695	1.1695	1.0819	0.8455
1951	-0.6584	-0.643	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
1952	-0.7920	-0.8116	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
1953	-0.1873	-0.0790	-0.4519	-0.4337	0.7826	0.1538	1.0468
1954	0.9413	1.10363	0.3389	0.2874	0.5709	-0.0273	0.1151
1955	0.3882	0.5466	0.3962	0.6026	1.1550	0.6759	1.1964
1956	-0.6055	-0.5674	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
1957	1.30124	1.43174	3.38992	3.9680	1.2877	2.1625	1.1502
1958	-0.8754	-0.9184	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
1959	-1.17451	-1.3381	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
1960	-0.7920	-0.8116	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
1961	0.12092	0.2642	-0.2668	-0.0170	-0.9677	-0.4833	1.0604
1962	-0.9125	-0.9718	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
1963	-0.8516	-0.8879	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
1964	-1.02895	-1.1320	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
1965	-0.2905	-0.2011	-0.4613	-0.3268	1.0734	0.7147	1.2536
1966	-0.3871	-0.3156	-0.4910	-0.6687	-0.0453	-0.3696	1.3991
1967	-0.6372	-0.6132	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
1968	-1.25258	-1.4525	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
1969	-0.5367	-0.4758	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647

1970	-0.2601	-0.1630	-0.4337	-0.1826	0.3037	-0.1747	1.2141
1971	0.9757	1.1341	0.6647	0.81096	0.7201	0.1513	0.3749
1972	0.1870	0.3329	0.0131	0.69344	1.3613	307578	0.0281
1973	0.0918	0.2337	-0.1740	0.8002	0.7299	0.0088	1.2617
1974	-0.8860	-0.9337	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
1975	0.0256	0.1650	-0.2988	0.0951	0.7465	0.0013	0.7150
1976	-0.3607	-0.2851	-0.4852	-0.5084	0.7103	-0.0048	1.4059
1977	-0.8516	-0.8879	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
1978	-1.1586	-1.3152	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
1979	-0.1675	-0.0561	-0.4490	-0.4123	0.8452	0.3149	1.0142
1980	0.0309	0.0582	-0.3242	-0.0757	0.9286	0.1787	1.0332
1981	0.2585	0.2948	-0.0550	0.3729	0.3960	-0.2034	0.9176
1982	0.7560	0.7984	0.47753	0.6667	1.0333	0.0201	0.6619
1983	4.58549	3.8277	6.53738	3.9146	0.0365	-0.3433	-0.2874
1984	0.3286	0.4092	-0.1566	0.1592	1.2553	1.6341	1.0346
1985	-1.2208	-1.4754	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
1986	-0.5750	-0.6208	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
1987	0.6290	0.6687	0.2504	0.4103	1.0656	0.7746	0.3831
1988	0.0891	0.1193	-0.3031	0.0150	0.1926	-0.2347	1.1284
1989	0.2585	0.2948	-0.0963	0.1111	1.2609	1.3518	0.0050
1990	0.6674	0.7068	0.0320	0.2447	0.2921	-0.1660	0.3219
1991	-0.7100	-0.788	-0.4925	-0.780	-1.1725	-0.7556	-1.1647
1992	2.99365	2.6678	2.3131	0.5438	0.8040	0.1500	-1.0559
1993	1.4481	1.4088	0.7909	0.6720	0.6125	0.0088	0.2647
1994	-0.4387	-0.4606	-0.4925	-0.780	-1.1725	-0.7556	-1.1647
1995	1.0074	1.0273	0.4078	0.4637	0.7303	0.0426	0.4552
1996	-0.1807	-0.1630	-0.4453	-0.305	0.4358	-0.0997	1.1788
1997	0.9307	0.9586	1.0202	2.2426	0.3380	-0.2084	1.2495
1998	1.4031	1.3707	1.9250	2.4776	0.9570	0.1837	0.3409
1999	0.1870	0.2261	-0.1587	0.1111	0.3851	-0.1560	0.6021
2000	0.7838	0.8289	0.1133	0.1699	1.2859	1.8865	0.0036
2001	0.3763	0.4245	-0.1188	0.1913	1.0400	0.4723	0.6129
2002	-0.7007	-0.765	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
2003	-0.6187	-0.659	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
2004	-0.2522	-0.2316	-0.4751	-0.4978	0.7093	0.1213	1.1448
2005	0.2717	0.3176	0.0625	1.4733	1.2182	1.3855	0.8278
2006	-1.4695	-1.8112	-0.4925	-0.7809	-1.1725	-0.7556	-1.1647
2007	0.2320	0.2795	-0.2676	-0.0490	0.2166	-0.2109	0.9448
2008	-0.2919	-0.2698	-0.4772	-0.4390	0.2075	-0.2546	1.2998
2009	0.4742	0.5237	0.3462	1.6496	1.0142	0.4623	1.1801
2010	1.20994	1.21809	0.6451	0.5705	1.4416	5.6254	0.0417
2011	0.9214	0.9510	0.7517	2.1624	-1.0863	-0.4833	0.6701
2012	0.1619	0.2108	-0.1907	0.1218	0.5989	-0.0385	0.8061
2013	0.9267	0.9662	0.3266	0.2607	1.2733	1.6778	0.1655
2014	2.2420	2.0803	1.5883	0.5492	0.4958	-0.0773	-0.2493
2015	0.2770	0.3253	-0.0194	1.1101	0.6936	-0.0123	0.4334

```

i1 = data['I1'].values
i2 = data['I2'].values
i3 = data['I3'].values
i4 = data['I4'].values
i5 = data['I5'].values
i6 = data['I6'].values
i7 = data['I7'].values

from sklearn.cluster import KMeans
X = np.array(list(zip(i1.i2.i3.i4.i5.i6.i7)))
kmeans = KMeans(n_clusters = 4)
kmeans = kmeans.fit(X)
labels = kmeans.predict(X)
centroids = kmeans.cluster_centers_
print(centroids)

[[-0.8795825,-0.95576562500000002,-0.49250000000000001,-0.78084375000000005,
0.8441047619047619,-0.75558124999999998,1.0366769230769228]
[-0.061929999999999985,0.018934615384615376,-0.3469269230769232,-
0.1725288461538462,0.7061666666666667,0.11056153846153847,0.4960380952380953
][0.72130666666666665,0.7911104761904763,0.34254904761904764,0.7854857142857
142,0.5880769230769232,0.38925,-0.21103333333333335]
[2.4349966666666667,2.2469066666666664,2.9313333333333333,2.163316666666666
8,-1.1725000000000003,0.9020999999999998,-1.1646999999999998]]

def mahalanobis(x=None,cov=None)

    #compute the Mahalanobis Distance Between each of x and the data

x: vector or matrix of data with.say.p colum
data:ndarray of the distribution from which Mahalanobis distance of each
observation of x is
cov:covariance matrix (pxp)of the distribution.If None.will be computed
from data
"""
x_minus_mu = x-np.mean(data)
if not cov:
    inv_covmat = sp.linalg.inv_covamat
    left_term = np.dot(x_minus_mu.inv_covmat)
    mahal = np.dot(left_term.x_minus_mu.T)
"""
df_x = df[['I1'.'I2'.'I3'.'I4'.'I5'.'I6'.'I7']].head(85)
df_x['mahala']=mahalanobis(x=df_x.data=df[['I1'.'I2'.'I3'.'I4'.'I5'.'I6'.'I
7']])
tabela = df_x['mahala'](np.array(list(zip(i1. i2. i3. i4. i5. i6. i7.
labels)))).
columns = ['I1'. 'I2'. 'I3'. 'I4'. 'I5'. 'I6'.'I7'.'Centroide'])

I1.I2.I3.I4.I5.I6.I7.Centroide

0,-0.3157,-0.2316,-0.4823,-0.6153,1.2362,0.992,1.1529,2.0
1,0.187,0.3329,-0.3488,-0.3428,1.1333,0.6222,0.4674,2.0
2,-1.1321,-1.277,-0.4925,-0.7809,-1.1725,-0.755,-1.1647,1.0
3,-1.1533,-1.3075,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
4,2.0845,2.1032,1.8343,1.5267,0.9227,0.2599,-1.1647,3.0
5,-0.5949,-0.5521,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
6,-0.3871,-0.3156,-0.4925,-0.7435,0.9423,0.2937,1.2305,2.0
7,0.9148,1.0807,0.3113,0.3675,1.2294,1.2756,0.549,4.0
8,0.0335,0.1727,-0.3002,0.0791,0.2538,-0.1997,1.1325,2.0
9,-1.4867,-1.8188,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
10,-0.5605,-0.5063,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
11,-0.653,-0.6361,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0

```

12,-1.0448,-1.1549,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
 13,-0.9733,-1.0557,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
 14,-0.6266,-0.5979,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
 15,0.85,1.012,0.5174,1.3665,0.261,-0.1809,0.4552,3.0
 16,0.1129,0.2566,-0.1587,0.2447,1.1797,0.6035,0.9176,2.0
 17,-0.3289,-0.2469,-0.4816,-0.52985,0.0862,-0.2546,1.1679,2.0
 18,-1.1745,-1.3381,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
 19,-0.1079,0.0124,-0.4286,-0.3695,1.1695,1.0819,0.8455,2.0
 20,-0.6584,-0.643,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
 21,-0.792,-0.8116,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
 22,-0.1873,-0.079,-0.4519,-0.4337,0.7826,0.1538,1.0468,2.0
 23,0.9413,1.1036299999999999,0.3389,0.2874,0.5709,-0.0273,0.1151,4.0
 24,0.3882,0.5466,0.3962,0.6026,1.155,0.6759,1.1964,4.0
 25,-0.6055,-0.5674,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
 26,1.30124,1.43174,3.3899199999999996,3.968,1.2877,2.1625,1.1502,3.0
 27,-0.8754,-0.9184,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
 28,-1.17451,-1.3381,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
 29,-0.792,-0.8116,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
 30,0.12092,0.2642,-0.2668,-0.017,-0.9677,-0.4833,1.0604,2.0
 31,-0.9125,-0.9718,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
 32,-0.8516,-0.8879,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
 33,-1.02895,-1.132,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
 34,-0.2905,-0.2011,-0.4613,-0.3268,1.0734,0.7147,1.2536,2.0
 35,-0.3871,-0.3156,-0.491,-0.6687,-0.0453,-0.3696,1.3991,2.0
 36,-0.6372,-0.6132,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
 37,-1.25258,-1.4525,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
 38,-0.5367,-0.4758,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
 39,-0.2601,-0.163,-0.4337,-0.1826,0.3037,-0.1747,1.2141,2.0
 40,0.9757,1.1341,0.6647,0.81096,0.7201,0.1513,0.3749,4.0
 41,0.187,0.3329,0.0131,0.69344,1.3613,3.05,0.0281,4.0
 42,0.0918,0.2337,-0.174,0.8002,0.7299,0.0088,1.2617,2.0
 43,-0.886,-0.9337,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
 44,0.0256,0.165,-0.2988,0.0951,0.7465,0.0013,0.715,2.0
 45,-0.3607,-0.2851,-0.4852,-0.5084,0.7103,-0.0048,1.4059,2.0
 46,-0.8516,-0.8879,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
 47,-1.1586,-1.3152,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
 48,-0.1675,-0.0561,-0.449,-0.4123,0.8452,0.3149,1.0142,2.0
 49,0.0309,0.0582,-0.3242,-0.0757,0.9286,0.1787,1.0332,2.0
 50,0.2585,0.2948,-0.055,0.3729,0.396,-0.2034,0.9176,2.0
 51,0.756,0.7984,0.47753,0.6667,1.0333,0.0201,0.6619,2.0
 52,4.58549,3.8277,6.537380000000001,3.9146,0.0365,-0.3433,-0.2874,4.0
 53,0.3286,0.4092,-0.1566,0.1592,1.2553,1.6341,1.0346,4.0
 54,-1.2208,-1.4754,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
 55,-0.575,-0.6208,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
 56,0.629,0.6687,0.2504,0.4103,1.0656,0.7746,0.3831,4.0
 57,0.0891,0.1193,-0.3031,0.015,0.1926,-0.2347,1.1284,2.0
 58,0.2585,0.2948,-0.0963,0.1111,1.2609,1.3518,0.005,3.0
 59,0.6674,0.7068,0.032,0.2447,0.2921,-0.166,0.3219,2.0
 60,-0.71,-0.7879999999999999,-0.4925,-0.78,-1.1725,-0.7556,-1.1647,1.0
 61,2.99365,2.6678,2.3131,0.5438,0.804,0.15,-1.0559,4.0
 62,1.4481,1.4088,0.7909,0.672,0.6125,0.0088,0.2647,4.0
 63,-0.4387,-0.4606,-0.4925,-0.78,-1.1725,-0.7556,-1.1647,1.0
 64,1.0074,1.0273,0.4078,0.4637,0.7303,0.0426,0.4552,3.0
 65,-0.1807,-0.163,-0.4453,-0.305,0.4358,-0.0997,1.1788,2.0
 66,0.9307,0.9586,1.0202,2.2426,0.338,-0.2084,1.2495,2.0
 67,1.4031,1.3707,1.925,2.4776,0.957,0.1837,0.3409,4.0
 68,0.187,0.2261,-0.1587,0.1111,0.3851,-0.156,0.6021,2.0
 69,0.7838,0.8289,0.1133,0.1699,1.2859,1.8865,0.0036,3.0
 70,0.3763,0.4245,-0.1188,0.1913,1.04,0.4723,0.6129,3.0
 71,-0.7007,-0.765,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
 72,-0.6187,-0.659,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
 73,-0.2522,-0.2316,-0.4751,-0.4978,0.7093,0.1213,1.1448,2.0
 74,0.2717,0.3176,0.0625,1.4733,1.2182,1.3855,0.8278,3.0

```

75,-1.4695,-1.8112,-0.4925,-0.7809,-1.1725,-0.7556,-1.1647,1.0
76,0.232,0.2795,-0.2676,-0.049,0.2166,-0.2109,0.9448,2.0
77,-0.2919,-0.2698,-0.4772,-0.439,0.2075,-0.2546,1.2998,2.0
78,0.4742,0.5237,0.3462,1.6496,1.0142,0.4623,1.1801,4.0
79,1.20994,1.21809,0.6451,0.5705,1.4416,5.6254,0.0417,3.0
80,0.9214,0.951,0.7517,2.1624,-1.0863,-0.4833,0.6701,3.0
81,0.1619,0.2108,-0.1907,0.1218,0.5989,-0.0385,0.8061,2.0
82,0.9267,0.9662,0.3266,0.2607,1.2733,1.6778,0.1655,4.0
83,2.242,2.0803,1.5883,0.5492,0.4958,-0.0773,-0.2493,4.0
84,0.277,0.3253,-0.0194,1.1101,0.6936,-0.0123,0.4334,3.0

```

```

tabela.to_csv("C:/Users/55419/Desktop/Arquivos_ULTIMOS/ARQUIVOS
CVS_ALTERADOS/CSV/NORMALIZAÇÃO DP I5 e I7_ALTERADOS/Valores
DP_ALTERADOS_Agrupados.csv")

```

```

centroides_df = pd.DataFrame(centroids)

```

```

centroides_df.to_csv("C:/Users/55419/Desktop/Arquivos
10_03_ULTIMOS/ARQUIVOS CVS_ALTERADOS/CSV/NORMALIZAÇÃO DP I5 e
I7_ALTERADOS/Valores DP_ALTERADOS_Centróides.csv")

```

APÊNDICE 27 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORITMO K-MEANS; DADOS DE I₅ E I₇ ALTERADOS. NORMALIZADOS ATRAVÉS DO MÉTODO DO VALOR MÁXIMO DO ELEMENTO; PROGRAMA EXECUTADO EM LINGUAGEM DE PROGRAMAÇÃO PYTHON.

```
import pandas as pd

import numpy as np

def dist(a,b,ax=1):
    return np.linalg.norm(a-b, axis = ax)

data = pd.read_csv("C:/Users//55419/Desktop/Kmeans
Normalizações/Normalização por VM/NPVMEL.csv")
data
```

Ano	I1	I2	I3	I4	I5	I6	I7
1931	0.2817	0.4894	0.0014	0.0348	0.1120	0.2738	0.1411
1932	0.3554	0.5604	0.0204	0.0922	0.1513	0.2159	0.4077
1933	0.1621	0.3579	0.0000	0.0000	0.0000	0.0000	0.0000
1934	0.1590	0.3541	0.0000	0.0000	0.0000	0.0000	0.0000
1935	0.6335	0.7831	0.3309	0.4859	0.2319	0.1591	0.4854
1936	0.2408	0.4491	0.0000	0.0000	0.0000	0.0000	0.0000
1937	0.2712	0.4788	0.0000	0.0078	0.2243	0.1644	0.1110
1938	0.4620	0.6545	0.1143	0.2418	0.1146	0.3183	0.3759
1939	0.3329	0.5403	0.0273	0.1811	0.4876	0.0871	0.1491
1940	0.1101	0.2898	0.0000	0.0000	0.0000	0.0000	0.0000
1941	0.2458	0.4548	0.0000	0.0000	0.0000	0.0000	0.0000
1942	0.2323	0.4385	0.0000	0.0000	0.0000	0.0000	0.0000
1943	0.1749	0.3733	0.0000	0.0000	0.0000	0.0000	0.0000
1944	0.1853	0.3857	0.0000	0.0000	0.0000	0.0000	0.0000
1945	0.2361	0.4433	0.0000	0.0000	0.0000	0.0000	0.0000
1946	0.4525	0.6458	0.1436	0.4521	0.4848	0.0900	0.4124
1947	0.3445	0.5508	0.0474	0.2159	0.1336	0.2129	0.2326
1948	0.2798	0.4875	0.0015	0.0528	0.5517	0.0785	0.1353
1949	0.1559	0.3502	0.0000	0.0000	0.0000	0.0000	0.0000
1950	0.3121	0.5201	0.0090	0.0866	0.1375	0.2879	0.2607
1951	0.2315	0.4376	0.0000	0.0000	0.0000	0.0000	0.0000
1952	0.2119	0.4165	0.0000	0.0000	0.0000	0.0000	0.0000
1953	0.3005	0.5086	0.0057	0.0731	0.2854	0.1425	0.1824
1954	0.4659	0.6573	0.1182	0.2249	0.3663	0.1141	0.5446
1955	0.3849	0.5873	0.1264	0.2913	0.1430	0.2243	0.1242
1956	0.2392	0.4472	0.0000	0.0000	0.0000	0.0000	0.0000
1957	0.5187	0.6986	0.5522	1	0.0923	0.4573	0.1422
1958	0.1997	0.4030	0.0000	0.0000	0.0000	0.0000	0.0000
1959	0.1559	0.3502	0.0000	0.0000	0.0000	0.0000	0.0000
1960	0.2119	0.4165	0.0000	0.0000	0.0000	0.0000	0.0000
1961	0.3457	0.5518	0.0320	0.16085	0.9546	0.0426	0.1771
1962	0.1942	0.3963	0.0000	0.0000	0.0000	0.0000	0.0000
1963	0.2032	0.4069	0.0000	0.0000	0.0000	0.0000	0.0000
1964	0.1772	0.3761	0.0000	0.0000	0.0000	0.0000	0.0000
1965	0.2854	0.4932	0.0044	0.0956	0.1742	0.2304	0.1020

1966	0.2712	0.4788	0.0002	0.0236	0.6019	0.0604	0.0454
1967	0.2346	0.4414	0.0000	0.0000	0.0000	0.0000	0.0000
1968	0.1444	0.3358	0.0000	0.0000	0.0000	0.0000	0.0000
1969	0.2493	0.4587	0.0000	0.0000	0.0000	0.0000	0.0000
1970	0.2898	0.4980	0.0083	0.1259	0.4685	0.0910	0.1173
1971	0.4710	0.6612	0.1646	0.3352	0.3093	0.1421	0.4436
1972	0.3554	0.5604	0.0719	0.3104	0.0641	0.6004	0.5785
1973	0.3414	0.5479	0.0453	0.3329	0.3055	0.1198	0.0988
1974	0.1981	0.4011	0.0000	0.0000	0.0000	0.0000	0.0000
1975	0.3317	0.5393	0.0275	0.1844	0.2992	0.1186	0.3114
1976	0.2751	0.4827	0.0010	0.0573	0.3131	0.1176	0.0428
1977	0.2032	0.4069	0.0000	0.0000	0.0000	0.0000	0.0000
1978	0.1582	0.3531	0.0000	0.0000	0.0000	0.0000	0.0000
1979	0.3034	0.5115	0.0061	0.0776	0.2615	0.1677	0.1951
1980	0.3325	0.5259	0.0239	0.1484	0.2296	0.1464	0.1877
1981	0.3659	0.5556	0.0622	0.2429	0.4332	0.0865	0.2326
1982	0.4388	0.6190	0.1379	0.3048	0.1896	0.1215	0.3320
1983	1	1	1	0.9887	0.5706	0.0646	0.7012
1984	0.3761	0.5700	0.0477	0.1979	0.1047	0.3745	0.1872
1985	0.1491	0.3330	0.0000	0.0000	0.0000	0.0000	0.0000
1986	0.2437	0.4404	0.0000	0.0000	0.0000	0.0000	0.0000
1987	0.4202	0.6026	0.1056	0.2508	0.1772	0.2398	0.4405
1988	0.3410	0.5335	0.0269	0.1676	0.5110	0.0816	0.1507
1989	0.3659	0.5556	0.0563	0.1878	0.1025	0.3302	0.5875
1990	0.4258	0.6074	0.0746	0.2159	0.4729	0.0924	0.4643
1991	0.2239	0.4193	0.0000	0.0000	0.0000	0.0000	0.0000
1992	0.7667	0.8541	0.3991	0.2789	0.2772	0.1419	1
1993	0.5402	0.6957	0.1825	0.3059	0.3504	0.1198	0.4865
1994	0.2637	0.4606	0.0000	0.0000	0.0000	0.0000	0.0000
1995	0.4756	0.6477	0.1280	0.2620	0.3054	0.1250	0.4124
1996	0.3015	0.4980	0.0067	0.1001	0.4180	0.1027	0.1311
1997	0.4644	0.6391	0.2151	0.6366	0.4554	0.0857	0.1036
1998	0.5336	0.6909	0.3438	0.6861	0.2187	0.1472	0.4569
1999	0.3554	0.5470	0.0474	0.1878	0.4374	0.0939	0.3553
2000	0.4428	0.6228	0.0861	0.2002	0.0930	0.4140	0.5880
2001	0.3831	0.5719	0.0531	0.2047	0.1870	0.1924	0.3511
2002	0.2253	0.4222	0.0000	0.0000	0.0000	0.0000	0.0000
2003	0.2373	0.4357	0.0000	0.0000	0.0000	0.0000	0.0000
2004	0.2910	0.4894	0.0024	0.0596	0.3134	0.1374	0.1443
2005	0.3678	0.5585	0.0789	0.4746	0.1189	0.3355	0.2675
2006	0.1126	0.2907	0.0000	0.0000	0.0000	0.0000	0.0000
2007	0.3620	0.5537	0.0319	0.1541	0.5018	0.0853	0.2221
2008	0.2852	0.4846	0.0021	0.0719	0.5053	0.0785	0.0840
2009	0.3975	0.5844	0.1193	0.5118	0.1969	0.1908	0.1306
2010	0.5053	0.6717	0.1618	0.2845	0.0335	1	0.5732
2011	0.4630	0.6381	0.1770	0.6197	1	0.0426	0.3289
2012	0.3517	0.5451	0.0429	0.1901	0.3556	0.1123	0.2760
2013	0.4638	0.6401	0.1165	0.2193	0.0978	0.3813	0.5251
2014	0.6565	0.7802	0.2960	0.2800	0.3951	0.1063	0.6864
2015	0.3686	0.5595	0.0672	0.3982	0.3194	0.1164	0.4209

```

i1 = data['I1'].values
i2 = data['I2'].values
i3 = data['I3'].values
i4 = data['I4'].values
i5 = data['I5'].values
i6 = data['I6'].values
i7 = data['I7'].values

from sklearn.cluster import KMeans
X = np.array(list(zip(i1.i2.i3.i4.i5.i6.i7)))
kmeans = KMeans(n_clusters = 4)
kmeans = kmeans.fit(X)
labels = kmeans.predict(X)
centroids = kmeans.cluster_centers_
print(centroids)

[[0.199075,0.398309375,4.163336342344337e-17,-2.7755575615628914e-
17,0.8602636363636365,4.163336342344337e- 17,0.8522399999999999]
[0.3368,0.535264,0.040640000000000001,0.184788,0.7356666666666667,0.106236,0
.6602181818181818]
[0.4081181818181818,0.5973136363636363,0.09156818181818183,0.25616363636363
637,0.59136,0.1794,0.37095]
[0.6848333333333333,0.80115,0.487,0.6199333333333332,3.3306690738754696e-
16,0.28776818181818187,2.220446049250313e-16]]

#compute the Mahalanobis Distance Between each of x and the data

x: vector or matrix of data with.say.p colum
data:ndarray of the distribution from which Mahalanobis distance of each
observation of x is
cov:covariance matrix (pxp)of the distribution.If None.will be computed
from data

"""
x_minus_mu = x-np.mean(data)
if not cov:
    inv_covmat = sp.linalg.inv_covamat
    left_term = np.dot(x_minus_mu.inv_covmat)
    mahal = np.dot(left_term.x_minus_mu.T)
"""
df_x = df[['I1'. 'I2'. 'I3'. 'I4'. 'I5'. 'I6'. 'I7']].head(85)
df_x['mahala']=mahalanobis(x=df_x.data=df[['I1'. 'I2'. 'I3'. 'I4'. 'I5'. 'I6'. 'I
7']])
tabela = df_x['mahala'](np.array(list(zip(i1. i2. i3. i4. i5. i6. i7.
labels)))).
columns = ['I1'. 'I2'. 'I3'. 'I4'. 'I5'. 'I6'. 'I7'. 'Centroide']

I1.I2.I3.I4.I5.I6.I7.Centroide
0,0.2817,0.4894,0.0014,0.0348,0.9214,0.2738,0.9015,3.0
1,0.3554,0.5604,0.0204,0.0922,0.882,0.2159,0.6349,3.0
2,0.1621,0.3579,0.0,0.0,0.0,0.0,0.0,1.0
3,0.159,0.3541,0.0,0.0,0.0,0.0,0.0,1.0
4,0.6335,0.7831,0.3309,0.4859,0.8015,0.1591,0.0,4.0
5,0.2408,0.4491,0.0,0.0,0.0,0.0,0.0,1.0
6,0.2712,0.4788,0.0,0.0078,0.809,0.1644,0.9317,2.0
7,0.462,0.6545,0.1143,0.2418,0.9188,0.3183,0.6666,3.0
8,0.3329,0.5403,0.0273,0.1811,0.5456,0.0871,0.8936,2.0
9,0.1101,0.2898,0.0,0.0,0.0,0.0,0.0,1.0
10,0.2458,0.4548,0.0,0.0,0.0,0.0,0.0,1.0

```

11,0.2323,0.4385,0.0,0.0,0.0,0.0,0.0,1.0
 12,0.1749,0.3733,0.0,0.0,0.0,0.0,0.0,1.0
 13,0.1853,0.3857,0.0,0.0,0.0,0.0,0.0,1.0
 14,0.2361,0.4433,0.0,0.0,0.0,0.0,0.0,1.0
 15,0.4525,0.6458,0.1436,0.4521,0.5483,0.09,0.6301,2.0
 16,0.3445,0.5508,0.0474,0.2159,0.8998,0.2129,0.81,3.0
 17,0.2798,0.4875,0.0015,0.0528,0.4815,0.0785,0.9074,1.0
 18,0.1559,0.3502,0.0,0.0,0.0,0.0,0.0,1.0
 19,0.3121,0.5201,0.009,0.0866,0.8959,0.2879,0.782,3.0
 20,0.2315,0.4376,0.0,0.0,0.0,0.0,0.0,1.0
 21,0.2119,0.4165,0.0,0.0,0.0,0.0,0.0,1.0
 22,0.3005,0.5086,0.0057,0.0731,0.7479,0.1425,0.8603,2.0
 23,0.4659,0.6573,0.1182,0.2249,0.6669,0.1141,0.4978,3.0
 24,0.3849,0.5873,0.1264,0.2913,0.8904,0.2243,0.9185,3.0
 25,0.2392,0.4472,0.0,0.0,0.0,0.0,0.0,1.0
 26,0.5187,0.6986,0.5522,1.0,0.9411,0.4573,0.9005,4.0
 27,0.1997,0.403,0.0,0.0,0.0,0.0,0.0,1.0
 28,0.1559,0.3502,0.0,0.0,0.0,0.0,0.0,1.0
 29,0.2119,0.4165,0.0,0.0,0.0,0.0,0.0,1.0
 30,0.3457,0.5518,0.032,0.1608,0.0783,0.0426,0.8656,2.0
 31,0.1942,0.3963,0.0,0.0,0.0,0.0,0.0,1.0
 32,0.2032,0.4069,0.0,0.0,0.0,0.0,0.0,1.0
 33,0.1772,0.3761,0.0,0.0,0.0,0.0,0.0,1.0
 34,0.2854,0.4932,0.0044,0.0956,0.8591,0.2304,0.9407,2.0
 35,0.2712,0.4788,0.0002,0.0236,0.4312,0.0604,0.9973,2.0
 36,0.2346,0.4414,0.0,0.0,0.0,0.0,0.0,1.0
 37,0.1444,0.3358,0.0,0.0,0.0,0.0,0.0,1.0
 38,0.2493,0.4587,0.0,0.0,0.0,0.0,0.0,1.0
 39,0.2898,0.498,0.0083,0.1259,0.5647,0.091,0.9253,2.0
 40,0.471,0.6612,0.1646,0.3352,0.724,0.1421,0.5989,3.0
 41,0.3554,0.5604,0.0719,0.3104,0.9693,0.6004,0.464,3.0
 42,0.3414,0.5479,0.0453,0.3329,0.7277,0.1198,0.9439,2.0
 43,0.1981,0.4011,0.0,0.0,0.0,0.0,0.0,1.0
 44,0.3317,0.5393,0.0275,0.1844,0.7341,0.1186,0.7312,2.0
 45,0.2751,0.4827,0.001,0.0573,0.7202,0.1176,1.0,2.0
 46,0.2032,0.4069,0.0,0.0,0.0,0.0,0.0,1.0
 47,0.1582,0.3531,0.0,0.0,0.0,0.0,0.0,1.0
 48,0.3034,0.5115,0.0061,0.0776,0.7718,0.1677,0.8476,2.0
 49,0.3325,0.5259,0.0239,0.1484,0.8037,0.1464,0.855,2.0
 50,0.3659,0.5556,0.0622,0.2429,0.6,0.0865,0.81,2.0
 51,0.4388,0.619,0.1379,0.3048,0.8438,0.1215,0.7105,3.0
 52,1.0,1.0,1.0,0.9887,0.4625,0.0646,0.3412,4.0
 53,0.3761,0.57,0.0477,0.1979,0.9287,0.3745,0.8555,3.0
 54,0.1491,0.333,0.0,0.0,0.0,0.0,0.0,1.0
 55,0.2437,0.4404,0.0,0.0,0.0,0.0,0.0,1.0
 56,0.4202,0.6026,0.1056,0.2508,0.8561,0.2398,0.6021,3.0
 57,0.341,0.5335,0.0269,0.1676,0.5222,0.0816,0.892,2.0
 58,0.3659,0.5556,0.0563,0.1878,0.9308,0.3302,0.455,3.0
 59,0.4258,0.6074,0.0746,0.2159,0.5603,0.0924,0.5783,2.0
 60,0.2239,0.4193,0.0,0.0,0.0,0.0,0.0,1.0

```

61,0.7667,0.8541,0.3991,0.2789,0.7561,0.1419,0.0423,4.0
62,0.5402,0.6957,0.1825,0.3059,0.6828,0.1198,0.556,3.0
63,0.2637,0.4606,0.0,0.0,0.0,0.0,0.0,1.0
64,0.4756,0.6477,0.128,0.262,0.7279,0.125,0.6301,3.0
65,0.3015,0.498,0.0067,0.1001,0.6152,0.1027,0.9116,2.0
66,0.4644,0.6391,0.2151,0.6366,0.5778,0.0857,0.9391,2.0
67,0.5336,0.6909,0.3438,0.6861,0.8146,0.1472,0.5857,4.0
68,0.3554,0.547,0.0474,0.1878,0.5958,0.0939,0.6873,2.0
69,0.4428,0.6228,0.0861,0.2002,0.9404,0.414,0.4544,3.0
70,0.3831,0.5719,0.0531,0.2047,0.8464,0.1924,0.6915,3.0
71,0.2253,0.4222,0.0,0.0,0.0,0.0,0.0,1.0
72,0.2373,0.4357,0.0,0.0,0.0,0.0,0.0,1.0
73,0.291,0.4894,0.0024,0.0596,0.7198,0.1374,0.8984,2.0
74,0.3678,0.5585,0.0789,0.4746,0.9145,0.3355,0.7751,3.0
75,0.1126,0.2907,0.0,0.0,0.0,0.0,0.0,1.0
76,0.362,0.5537,0.0319,0.1541,0.5314,0.0853,0.8206,2.0
77,0.2852,0.4846,0.0021,0.0719,0.5279,0.0785,0.9587,2.0
78,0.3975,0.5844,0.1193,0.5118,0.8365,0.1908,0.9121,3.0
79,0.5053,0.6717,0.1618,0.2845,1.0,1.0,0.4693,3.0
80,0.463,0.6381,0.177,0.6197,0.0329,0.0426,0.7137,2.0
81,0.3517,0.5451,0.0429,0.1901,0.6776,0.1123,0.7666,2.0
82,0.4638,0.6401,0.1165,0.2193,0.9356,0.3813,0.5174,3.0
83,0.6565,0.7802,0.296,0.28,0.6382,0.1063,0.356,4.0
84,0.3686,0.5595,0.0672,0.3982,0.7138,0.1164,0.6216,3.0

```

```

tabela.to_csv("C:/Users/55419/Desktop/Arquivos_ULTIMOS/ARQUIVOS
CVS_ALTERADOS/CSV/NORMALIZAÇÃO VM_ALTERADOS/Valores
VM_I5Ei7_ALTERADOS_Agrupados.csv")

```

```
centroides_df = pd.DataFrame(centroids)
```

```

centroides_df.to_csv("C:/Users/55419/Desktop/Arquivos
10_03_ULTIMOS/ARQUIVOS CVS_ALTERADOS/CSV/NORMALIZAÇÃO VM_ALTERADOS/Valores
VM_I5EI7_ALTERADOS_Centróides.csv")

```

APÊNDICE 28 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORITMO K-MEANS; DADOS DE I_5 E I_7 ALTERADOS. NORMALIZADOS ATRAVÉS DO MÉTODO DA ESCALA DECIMAL; PROGRAMA EXECUTADO EM LINGUAGEM DE PROGRAMAÇÃO PYTHON.

```
import pandas as pd

import numpy as np

def dist(a,b,ax=1):
    return np.linalg.norm(a-b, axis = ax)

data = pd.read_csv("C:/Users//55419/Desktop/Kmeans
Normalizações/Normalização p_ESCALA DECIMAL/NPVMEL.csv")
data
```

Ano	I1	I2	I3	I4	I5	I6	I7
1931	0.1453	0.051	0.0014	0.3111	0.52523	0.1399	0.1704
1932	0.1833	0.0584	0.0198	0.8222	0.50278	0.1103	0.12
1933	0.0836	0.0373	0.0000	0.0000	0.0000	0.0000	0.0000
1934	0.082	0.0369	0.0000	0.0000	0.0000	0.0000	0.0000
1935	0.3267	0.0816	0.3207	43.222	0.45686	0.0813	0.0000
1936	0.1242	0.0468	0.0000	0.0000	0.0000	0.0000	0.0000
1937	0.1399	0.0499	0.0000	0.0777	0.46115	0.084	0.1761
1938	0.2383	0.0682	0.1108	21.555	0.52374	0.1626	0.126
1939	0.1717	0.0563	0.0265	16.111	0.31101	0.0445	0.1689
1940	0.0568	0.0302	0.0000	0.0000	0.0000	0.0000	0.0000
1941	0.1268	0.0474	0.0000	0.0000	0.0000	0.0000	0.0000
1942	0.1198	0.0457	0.0000	0.0000	0.0000	0.0000	0.0000
1943	0.0902	0.0389	0.0000	0.0000	0.0000	0.0000	0.0000
1944	0.0956	0.0402	0.0000	0.0000	0.0000	0.0000	0.0000
1945	0.1218	0.0462	0.0000	0.0000	0.0000	0.0000	0.0000
1946	0.2334	0.0673	0.1392	40.222	0.31258	0.046	0.1191
1947	0.1777	0.0574	0.046	19.222	0.5129	0.1088	0.1531
1948	0.1443	0.0508	0.0015	0.4777	0.27447	0.0401	0.1715
1949	0.0804	0.0365	0.0000	0.0000	0.0000	0.0000	0.0000
1950	0.161	0.0542	0.0088	0.7777	0.51068	0.1471	0.1478
1951	0.1194	0.0456	0.0000	0.0000	0.0000	0.0000	0.0000
1952	0.1093	0.0434	0.0000	0.0000	0.0000	0.0000	0.0000
1953	0.155	0.053	0.0056	0.6555	0.42633	0.0728	0.1626
1954	0.2403	0.0685	0.1146	20.000	0.38017	0.0583	0.0941
1955	0.1985	0.0612	0.1225	25.999	0.50753	0.1146	0.1736
1956	0.1234	0.0466	0.0000	0.0000	0.0000	0.0000	0.0000
1957	0.2675	0.0728	0.5351	88.999	0.53645	0.2336	0.1702
1958	0.103	0.042	0.0000	0.0000	0.0000	0.0000	0.0000
1959	0.0804	0.0365	0.0000	0.0000	0.0000	0.0000	0.0000
1960	0.1093	0.0434	0.0000	0.0000	0.0000	0.0000	0.0000
1961	0.1783	0.0575	0.0311	14.333	0.04467	0.0218	0.1636
1962	0.1002	0.0413	0.0000	0.0000	0.0000	0.0000	0.0000

1963	0.1048	0.0424	0.0000	0.0000	0.0000	0.0000	0.0000
1964	0.0914	0.0392	0.0000	0.0000	0.0000	0.0000	0.0000
1965	0.1472	0.0514	0.0043	0.8555	0.48973	0.1177	0.1778
1966	0.1399	0.0499	0.0002	0.2111	0.24579	0.0309	0.1885
1967	0.121	0.046	0.0000	0.0000	0.0000	0.0000	0.0000
1968	0.0745	0.035	0.0000	0.0000	0.0000	0.0000	0.0000
1969	0.1286	0.0478	0.0000	0.0000	0.0000	0.0000	0.0000
1970	0.1495	0.0519	0.0081	11.222	0.32189	0.0465	0.1749
1971	0.2429	0.0689	0.1595	29.898	0.41269	0.0726	0.1132
1972	0.1833	0.0584	0.0697	27.666	0.55251	0.3067	0.0877
1973	0.1761	0.0571	0.0439	29.666	0.41484	0.0612	0.1784
1974	0.1022	0.0418	0.0000	0.0000	0.0000	0.0000	0.0000
1975	0.1711	0.0562	0.0267	16.444	0.41846	0.0606	0.1382
1976	0.1419	0.0503	0.001	0.5111	0.41055	0.0601	0.189
1977	0.1048	0.0424	0.0000	0.0000	0.0000	0.0000	0.0000
1978	0.0816	0.0368	0.0000	0.0000	0.0000	0.0000	0.0000
1979	0.1565	0.0533	0.006	0.6999	0.43998	0.0857	0.1602
1980	0.1715	0.0548	0.0232	13.222	0.45816	0.0748	0.1616
1981	0.1887	0.0579	0.0603	21.636	0.34203	0.0442	0.1531
1982	0.2263	0.0645	0.1337	27.111	0.48098	0.0621	0.1343
1983	0.5157	0.1042	0.9689	87.999	0.26365	0.033	0.0645
1984	0.194	0.0594	0.0463	17.666	0.52939	0.1913	0.1617
1985	0.0769	0.0347	0.0000	0.0000	0.0000	0.0000	0.0000
1986	0.1257	0.0459	0.0000	0.0000	0.0000	0.0000	0.0000
1987	0.2167	0.0628	0.1024	22.333	0.48803	0.1225	0.1138
1988	0.1759	0.0556	0.0261	14.999	0.29768	0.0417	0.1686
1989	0.1887	0.0579	0.0546	16.777	0.53061	0.1687	0.086
1990	0.2196	0.0633	0.0723	19.222	0.31938	0.0472	0.1093
1991	0.1155	0.0437	0.0000	0.0000	0.0000	0.0000	0.0000
1992	0.3954	0.089	0.3867	24.888	0.43099	0.0725	0.0080
1993	0.2786	0.0725	0.1769	27.222	0.38924	0.0612	0.1051
1994	0.136	0.048	0.0000	0.0000	0.0000	0.0000	0.0000
1995	0.2453	0.0675	0.1241	23.333	0.41492	0.0639	0.1191
1996	0.1555	0.0519	0.0065	0.8999	0.3507	0.0525	0.1723
1997	0.2395	0.0666	0.2085	56.666	0.32938	0.0438	0.1775
1998	0.2752	0.072	0.3332	6.111	0.46435	0.0752	0.1107
1999	0.1833	0.057	0.046	16.777	0.33965	0.048	0.1299
2000	0.2284	0.0649	0.0835	17.888	0.53606	0.2115	0.0859
2001	0.1976	0.0596	0.0515	18.224	0.48245	0.0983	0.1307
2002	0.1162	0.044	0.0000	0.0000	0.0000	0.0000	0.0000
2003	0.1224	0.0454	0.0000	0.0000	0.0000	0.0000	0.0000
2004	0.1501	0.051	0.0024	0.5354	0.41033	0.0702	0.1698
2005	0.1897	0.0582	0.0765	42.224	0.5213	0.1714	0.1465
2006	0.0581	0.0303	0.0000	0.0000	0.0000	0.0000	0.0000
2007	0.1867	0.0577	0.031	13.777	0.30291	0.0436	0.1551
2008	0.1471	0.0505	0.0021	0.6448	0.30092	0.0401	0.1812

2009	0.205	0.0609	0.1156	45.555	0.47681	0.0975	0.1724
2010	0.2606	0.07	0.1568	25.345	0.5777	0.5108	0.0887
2011	0.2388	0.0665	0.1715	55.115	0.0188	0.0218	0.1349
2012	0.1814	0.0568	0.0416	16.999	0.38626	0.0574	0.1449
2013	0.2392	0.0667	0.1129	19.595	0.53332	0.1948	0.0978
2014	0.3386	0.0813	0.2868	24.949	0.36378	0.0543	0.0673
2015	0.1901	0.0583	0.0652	35.454	0.40692	0.0595	0.1175

```
i1 = data['I1'].values
i2 = data['I2'].values
i3 = data['I3'].values
i4 = data['I4'].values
i5 = data['I5'].values
i6 = data['I6'].values
i7 = data['I7'].values
```

```
from sklearn.cluster import KMeans
X = np.array(list(zip(i1.i2.i3.i4.i5.i6.i7)))
kmeans = KMeans(n_clusters = 4)
kmeans = kmeans.fit(X)
labels = kmeans.predict(X)
centroids = kmeans.cluster_centers_
print(centroids)
```

```
[[0.10497575757575757,0.041993939393939396,0.0009424242424241891,0.00433333
3333333356,0.42904,0.0006606060606060699,0.004957575757575758][0.1836157894
736842,0.057557894736842106,0.04722894736842106,0.1488421052631579,0.40005,
0.1051578947368421,0.106016666666666666][0.262825,0.070258333333333334,0.2033
58333333333336,0.404916666666666666,0.381975,0.13329999999999997,0.11735]
[0.391600000000000006,0.0885,0.752,0.884,0.0013536363636362791,0.071425,0.14
533157894736842]]
```

```
def mahalanobis(x=None,cov=None)
```

```
#compute the Mahalanobis Distance Between each of x and the data
```

```
x: vector or matrix of data with.say.p colum
data:ndarray of the distribution from which Mahalanobis distance of each
observation of x is
cov:covariance matrix (pxp)of the distribution.If None.will be computed
from data
```

```
"""
```

```
x_minus_mu = x-np.mean(data)
if not cov:
    inv_covmat = sp.linalg.inv_covamat
    left_term = np.dot(x_minus_mu.inv_covmat)
    mahal = np.dot(left_term.x_minus_mu.T)
"""
```

```
df_x = df[['I1'.'I2'.'I3'.'I4'.'I5'.'I6'.'I7']].head(85)
df_x['mahala']=mahalanobis(x=df_x.data=df[['I1'.'I2'.'I3'.'I4'.'I5'.'I6'.'I
7']])
tabela = df_x['mahala'](np.array(list(zip(i1. i2. i3. i4. i5. i6. i7.
labels)))).
columns = ['I1'. 'I2'. 'I3'. 'I4'. 'I5'. 'I6'.'I7'.'Centroide'])
```

```
I1.I2.I3.I4.I5.I6.I7.Centroide
```

0,0.1453,0.051,0.0014,0.031,0.52523,0.1399,0.1704,2.0
1,0.1833,0.0584,0.0198,0.081999999999999,0.50278,0.1103,0.12,2.0
2,0.0836,0.0373,0.0,0.0,0.0,0.0,0.0,1.0
3,0.08199999999999999,0.0369,0.0,0.0,0.0,0.0,0.0,1.0
4,0.3267,0.0816,0.3207,0.4320000000005,0.45686000000000004,0.0813,0.0,4.0
5,0.1242,0.0468,0.0,0.0,0.0,0.0,0.0,1.0
6,0.1399,0.0499,0.0,0.006999999999999999,0.46115,0.084,0.1761,2.0
7,0.2383,0.0682,0.1108,0.215,0.52374,0.1626,0.126,2.0
8,0.1717,0.0563,0.0265,0.161,0.31101,0.0445,0.1689,2.0
9,0.0568,0.0302,0.0,0.0,0.0,0.0,0.0,1.0
10,0.1268,0.0474,0.0,0.0,0.0,0.0,0.0,1.0
11,0.1198,0.0457,0.0,0.0,0.0,0.0,0.0,1.0
12,0.0902,0.0389,0.0,0.0,0.0,0.0,0.0,1.0
13,0.0956,0.0402,0.0,0.0,0.0,0.0,0.0,1.0
14,0.1218,0.0462,0.0,0.0,0.0,0.0,0.0,1.0
15,0.2334,0.0673,0.1392,0.402,0.31258,0.046,0.1191,3.0
16,0.1777,0.0574,0.046,0.192,0.5129,0.1088,0.1531,2.0
17,0.1443,0.0508,0.0015,0.047,0.27447,0.0401,0.1715,2.0
18,0.0804,0.0365,0.0,0.0,0.0,0.0,0.0,1.0
19,0.161,0.0542,0.0088,0.077,0.51068,0.1471,0.1478,2.0
20,0.1194,0.0456,0.0,0.0,0.0,0.0,0.0,1.0
21,0.1093,0.0434,0.0,0.0,0.0,0.0,0.0,1.0
22,0.155,0.053,0.0056,0.065,0.42633000000000004,0.0728,0.1626,2.0
23,0.2403,0.0685,0.1146,0.2,0.38017,0.0583,0.0941,3.0
24,0.1985,0.0612,0.1225,0.259,0.50753,0.1146,0.1736,3.0
25,0.1234,0.0466,0.0,0.0,0.0,0.0,0.0,1.0
26,0.2675,0.0728,0.5351,0.889,0.53645,0.2336,0.1702,4.0
27,0.103000000000000001,0.042,0.0,0.0,0.0,0.0,0.0,1.0
28,0.0804,0.0365,0.0,0.0,0.0,0.0,0.0,1.0
29,0.1093,0.0434,0.0,0.0,0.0,0.0,0.0,1.0
30,0.1783,0.0575,0.0311,0.143000000000000002,0.04467,0.0218,0.1636,1.0
31,0.1002,0.0413,0.0,0.0,0.0,0.0,0.0,1.0
32,0.1048,0.0424,0.0,0.0,0.0,0.0,0.0,1.0
33,0.0914,0.0392,0.0,0.0,0.0,0.0,0.0,1.0
34,0.1472,0.0514,0.0043,0.085,0.48973,0.1177,0.1778,2.0
35,0.1399,0.0499,0.0002,0.021,0.24579,0.0309,0.1885,2.0
36,0.121,0.046,0.0,0.0,0.0,0.0,0.0,1.0
37,0.0745,0.035,0.0,0.0,0.0,0.0,0.0,1.0
38,0.1286,0.0478,0.0,0.0,0.0,0.0,0.0,1.0
39,0.1495,0.0519,0.0081,0.11199999999999999,0.32189,0.0465,0.1749,2.0
40,0.2429,0.0689,0.1595,0.298,0.41268999999999995,0.0726,0.1132,3.0
41,0.1833,0.0584,0.0697,0.276,0.552510000000000001,0.3067,0.0877,2.0
42,0.1761,0.0571,0.0439,0.29600000000000004,0.41484,0.0612,0.1784,2.0
43,0.1022,0.0418,0.0,0.0,0.0,0.0,0.0,1.0
44,0.1711,0.0562,0.0267,0.163999999998,0.41846000000,0.0606,0.1382,2.0
45,0.1419,0.0503,0.001,0.051,0.41055,0.0601,0.188999999997,2.0
46,0.1048,0.0424,0.0,0.0,0.0,0.0,0.0,1.0
47,0.0816,0.0368,0.0,0.0,0.0,0.0,0.0,1.0
48,0.1565,0.0533,0.006,0.069,0.439980000000000004,0.0857,0.1602,2.0
49,0.1715,0.0548,0.0232,0.132,0.45816,0.0748,0.1616,2.0

```

50,0.1887,0.0579,0.0603,0.21600000000000003,0.34203,0.0442,0.1531,2.0
51,0.2263,0.0645,0.1337,0.271,0.48098,0.0621,0.1343,2.0
52,0.5157,0.1042,0.9689,0.879,0.26365,0.033,0.0645,4.0
53,0.19399999998,0.0594,0.0463,0.176000002,0.529389999,0.1913,0.1617,3.0
54,0.0769,0.0347,0.0,0.0,0.0,0.0,0.0,1.0
55,0.1257,0.0459,0.0,0.0,0.0,0.0,0.0,1.0
56,0.2167,0.0628,0.1024,0.223,0.48803,0.1225,0.1138,3.0
57,0.1759,0.0556,0.0261,0.149,0.29768,0.0417,0.1686,2.0
58,0.1887,0.0579,0.0546,0.16699999999999998,0.53061,0.1687,0.086,3.0
59,0.2196,0.0633,0.0723,0.192,0.31938,0.0472,0.1093,2.0
60,0.1155,0.0437,0.0,0.0,0.0,0.0,0.0,1.0
61,0.3954,0.08900000000000001,0.3867,0.248,0.43099,0.0725,0.008,4.0
62,0.2786,0.0725,0.1769,0.272,0.38924000000000003,0.0612,0.1051,3.0
63,0.136,0.048,0.0,0.0,0.0,0.0,0.0,1.0
64,0.2453,0.0675,0.1241,0.233,0.41491999999999996,0.0639,0.1191,3.0
65,0.1555,0.0519,0.0065,0.08900000000000001,0.3507,0.0525,0.1723,2.0
66,0.2395,0.0666,0.2085,0.56600000000000001,0.32938,0.0438,0.1775,2.0
67,0.2752,0.07200000000000001,0.3332,0.61,0.46435,0.0752,0.1107,4.0
68,0.1833,0.057,0.046,0.16699999999999998,0.33965,0.048,0.1299,2.0
69,0.2284,0.0649,0.0835,0.17800000000000002,0.53606,0.2115,0.0859,2.0
70,0.1976,0.0596,0.0515,0.182,0.48245,0.0983,0.1307,3.0
71,0.1162,0.044000000000000004,0.0,0.0,0.0,0.0,0.0,1.0
72,0.1224,0.0454,0.0,0.0,0.0,0.0,0.0,1.0
73,0.1501,0.051,0.0024,0.053,0.41033000000000003,0.0702,0.1698,2.0
74,0.1897,0.0582,0.0765,0.42200000000000004,0.5213,0.1714,0.1465,3.0
75,0.0581,0.0303,0.0,0.0,0.0,0.0,0.0,1.0
76,0.1867,0.0577,0.031,0.13699999999999998,0.30291,0.0436,0.1551,2.0
77,0.1471,0.0505,0.0021,0.064,0.30091999999999997,0.0401,0.1812,2.0
78,0.205,0.0609,0.1156,0.455,0.47681,0.0975,0.1724,3.0
79,0.2606,0.07,0.1568,0.253,0.57,0.5108,0.0887,3.0
80,0.2388,0.0665,0.1715,0.551,0.0188,0.0218,0.1349,2.0
81,0.1814,0.0568,0.0416,0.16899999999999998,0.38626,0.0574,0.1449,2.0
82,0.2392,0.0667,0.1129,0.195,0.53332,0.1948,0.0978,3.0
83,0.3386,0.0813,0.2868,0.249,0.36378,0.0543,0.0673,4.0
84,0.1901,0.0583,0.0652,0.354,0.40692,0.0595,0.1175,3.0

```

```

tabela.to_csv("C:/Users/55419/Desktop/Arquivos_ULTIMOS/ARQUIVOS
CVS_ALTERADOS/CSV/NORMALIZAÇÃO ED_ALTERADOS/Valores
ED_I5eI7_ALTERADOS_Agrupados.csv")

```

```
centroides_df = pd.DataFrame(centroids)
```

```

centroides_df.to_csv("C:/Users/55419/Desktop/Arquivos
10_03_ULTIMOS/ARQUIVOS CVS_ALTERADOS/CSV/NORMALIZAÇÃO ED_ALTERADOS/Valores
ED_I5eI7_ALTERADOS_Centróides.csv")

```

APÊNDICE 29 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORITMO K-MEANS. CONSIDERANDO APENAS OS INDICADORES DE SEVERIDADE I_2 E I_7 . NORMALIZADOS ATRAVÉS DO MÉTODO DA NORMALIZAÇÃO LINEAR; PROGRAMA EXECUTADO EM LINGUAGEM DE PROGRAMAÇÃO PYTHON.

```
import pandas as pd
import numpy as np
def dist(a,b,ax=1):
    return np.linalg.norm(a-b, axis = ax)
data = pd.read_csv("C:/Users//55419/Desktop/Kmeans
Normalisações/Normalização Linear_I2-I7_ORIGINAIS/NPLADO.csv")
data
```

Ano	I2	I7
1931	0.2811	0.1409
1932	0.3811	0.4078
1933	0.0959	0.0000
1934	0.0905	0.0000
1935	0.6946	0.4856
1936	0.2243	0.0000
1937	0.2662	0.1269
1938	0.5135	0.3762
1939	0.3527	0.1491
1940	0.0000	0.0000
1941	0.2324	0.0000
1942	0.2095	0.0000
1943	0.1176	0.0000
1944	0.1351	0.0000
1945	0.2162	0.0000
1946	0.5014	0.4123
1947	0.3676	0.2329
1948	0.2784	0.1353
1949	0.0851	0.0000
1950	0.3243	0.2609
1951	0.2081	0.0000
1952	0.1784	0.0000
1953	0.3181	0.1822
1954	0.5176	0.5446
1955	0.4189	0.1244
1956	0.2216	0.0000
1957	0.5757	0.1421
1958	0.1595	0.0000
1959	0.0851	0.0000
1960	0.1784	0.0000
1961	0.3689	0.1774
1962	0.15	0.0000
1963	0.1649	0.0000
1964	0.1216	0.0000

1965	0.2865	0.102
1966	0.2662	0.0453
1967	0.2135	0.0000
1968	0.0649	0.0000
1969	0.2378	0.0000
1970	0.2932	0.1175
1971	0.523	0.4437
1972	0.3811	0.5787
1973	0.3635	0.0987
1974	0.1568	0.0000
1975	0.3514	0.3115
1976	0.2716	0.0428
1977	0.1649	0.0000
1978	0.0892	0.0000
1979	0.3122	0.1952
1980	0.3324	0.1875
1981	0.3743	0.2328
1982	0.4635	0.3322
1983	1.0000	0.7013
1984	0.3946	0.1874
1985	0.0608	0.0000
1986	0.2122	0.0000
1987	0.4405	0.4402
1988	0.3432	0.1509
1989	0.3743	0.5874
1990	0.4473	0.4642
1991	0.1824	0.0000
1992	0.7946	1.0000
1993	0.5716	0.4866
1994	0.2405	0.0000
1995	0.5041	0.4122
1996	0.2932	0.131
1997	0.4919	0.1035
1998	0.5649	0.4569
1999	0.3622	0.3555
2000	0.4689	0.5881
2001	0.3973	0.3513
2002	0.1865	0.0000
2003	0.2054	0.0000
2004	0.2811	0.1443
2005	0.3784	0.2676
2006	0.0014	0.0000
2007	0.3716	0.2219
2008	0.2743	0.0841
2009	0.4149	0.1304
2010	0.5378	0.5732

```

2011    0.4905  0.3287
2012    0.3595  0.2758
2013    0.4932  0.5253
2014    0.6905  0.6866
2015    0.3797  0.421

```

```

i2 = data['I2'].values
i7 = data['I7'].values

```

```

from sklearn.cluster import KMeans
X = np.array(list(zip(i2,i7)))
kmeans = KMeans(n_clusters = 4)
kmeans = kmeans.fit(X)
labels = kmeans.predict(X)
centroids = kmeans.cluster_centers_
print(centroids)

```

```

[[0.2850310344827586,0.8310965517241378]
 [0.15282812500000004,2.6653375615628914e-15]
 [0.46567499999999995,0.5882698457231]
 [0.79492545608871,0.18487548963125]]

```

```

def mahalanobis(x=None,cov=None)
    #compute the Mahalanobis Distance Between each of x and the data
x: vector or matrix of data with.say.p colum
data:ndarray of the distribution from which Mahalanobis distance of each
observation of x is
cov:covariance matrix (pxp)of the distribution.If None.will be computed
from data
"""
x_minus_mu = x-np.mean(data)
if not cov:
    inv_covmat = sp.linalg.inv_covamat
    left_term = np.dot(x_minus_mu.inv_covmat)
    mahal = np.dot(left_term.x_minus_mu.T)
"""
df_x = df[['I2','I7']].head(85)
df_x['mahala']=mahalanobis(x=df_x.data=df[['I2','I7']])
tabela = df_x['mahala'](np.array(list(zip(i2. i7. labels))).
columns = ['I2','I7','Centroide'])

```

```

I1.I7.Centroide
0.0.1929.0.1409.1.0
1.0.2757.0.4078.3.0
2.0.0585.0.0.1.0
3.0.0549.0.0.1.0
4.0.5881.0.4856.4.0
5.0.1469.0.0.1.0
6.0.1811.0.1269.2.0
7.0.3955.0.3762.3.0
8.0.2503.0.1491.2.0
9.0.0.0.0.1.0
10.0.1525.0.0.1.0
11.0.1373.0.0.1.0
12.0.0728.0.0.1.0

```

13.0.0846.0.0.1.0
14.0.1417.0.0.1.0
15.0.3848.0.4123.3.0
16.0.2635.0.2329.3.0
17.0.1908.0.1353.2.0
18.0.0514.0.0.1.0
19.0.2271.0.2609.3.0
20.0.1364.0.0.1.0
21.0.1144.0.0.1.0
22.0.214.0.1822.2.0
23.0.3998.0.5446.3.0
24.0.3088.0.1244.3.0
25.0.1451.0.0.1.0
26.0.4591.0.1421.4.0
27.0.1007.0.0.1.0
28.0.0514.0.0.1.0
29.0.1144.0.0.1.0
30.0.2647.0.1774.2.0
31.0.0946.0.0.1.0
32.0.1046.0.0.1.0
33.0.0755.0.0.1.0
34.0.19699998.0.102.2.0
35.0.1811.0.0453.2.0
36.0.1399.0.0.1.0
37.0.0386.0.0.1.0
38.0.1565.0.0.1.0
39.0.2021.0.1175.2.0
40.0.4055.0.4437.3.0
41.0.2757.0.5787.3.0
42.0.2599.0.0987.2.0
43.0.0989.0.0.1.0
44.0.2491.0.3115.2.0
45.0.1855.0.0428.2.0
46.0.1046.0.0.1.0
47.0.05400006.0.0.1.0
48.0.2173.0.1952.2.0
49.0.25.0.1875.2.0
50.0.2875.0.2328.2.0
51.0.3694.0.3322.3.0
52.1.0.0.7013.4.0
53.0.2989.0.1874.3.0
54.0.0437.0.0.1.0
55.0.1501.0.0.1.0
56.0.3483.0.4402.3.0
57.0.2595.0.1509.2.0
58.0.2875.0.5874.3.0
59.0.3548.0.4642.3.0
60.0.128.0.0.1.0
61.0.7378.1.0.4.0
62.0.4834.0.4866.3.0

```

63.0.1726.0.0.1.0
64.0.4109.0.4122.3.0
65.0.215.0.131.2.0
66.0.3982.0.1035.3.0
67.0.4759.0.4569.4.0
68.0.2757.0.3555.2.0
69.0.374.0.5881.3.0
70.0.3068.0.3513.3.0
71.0.1295.0.0.1.0
72.0.1429.0.0.1.0
73.0.2033.0.1443.2.0
74.0.2896.0.2676.3.0
75.0.0028.0.0.1.0
76.0.2824.0.2219.2.0
77.0.1967.0.0841.2.0
78.0.322995.0.1304.3.0
79.0.444.0.5732.3.0
80.0.3967.0.3287.2.0
81.0.2714.0.2758.2.0
82.0.3974.0.5253.3.0
83.0.614.0.6866.4.0
84.0.2905.0.421004.3.0

```

```

tabela.to_csv("C:/Users/55419/Desktop/Arquivos_ULTIMOS/ARQUIVOS
CVS_ALTERADOS/CSV/NORMALIZAÇÃO LINEAR_I2-I7_ORIGINAIS/Valores
LINEARES_ORIGINAIS_agrupados.csv")

```

```

centroides_df = pd.DataFrame(centroids)

```

```

centroides_df.to_csv("C:/Users/55419/Desktop/Arquivos
10_03_ULTIMOS/ARQUIVOS CVS_ALTERADOS/CSV/NORMALIZAÇÃO
LINEAR_ORIGINAIS/Valores LINEAR_I2-I7_ORIGINAIS_Centróides.csv")

```

APÊNDICE 30 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORITMO K-MEANS. CONSIDERANDO APENAS OS INDICADORES DE SEVERIDADE I_2 E I_7 . NORMALIZADOS ATRAVÉS DO MÉTODO DA NORMALIZAÇÃO LINEAR E COM VALORES DE I_7 ALTERADOS; PROGRAMA EXECUTADO EM LINGUAGEM DE PROGRAMAÇÃO PYTHON.

```
import pandas as pd
import numpy as np
def dist(a,b,ax=1):
    return np.linalg.norm(a-b, axis = ax)
data = pd.read_csv("C:/Users//55419/Desktop/Kmeans
Normalizações/Normalização Linear_I2-I7_ALTERADOS/NPLADA.csv")
data
```

Ano	I2	I7
1931	0.2811	0.9021
1932	0.3811	0.6354
1933	0.0959	0.0000
1934	0.0905	0.0000
1935	0.6946	0.0000
1936	0.2243	0.0000
1937	0.2662	0.9322
1938	0.5135	0.6671
1939	0.3527	0.8941
1940	0.0000	0.0000
1941	0.2324	0.0000
1942	0.2095	0.0000
1943	0.1176	0.0000
1944	0.1351	0.0000
1945	0.2162	0.0000
1946	0.5014	0.6306
1947	0.3676	0.8105
1948	0.2784	0.9079
1949	0.0851	0.0000
1950	0.3243	0.7825
1951	0.2081	0.0000
1952	0.1784	0.0000
1953	0.3181	0.8608
1954	0.5176	0.4984
1955	0.4189	0.9190
1956	0.2216	0.0000
1957	0.5757	0.9010
1958	0.1595	0.0000
1959	0.0851	0.0000
1960	0.1784	0.0000
1961	0.3689	0.8661
1962	0.15	0.0000
1963	0.1649	0.0000
1964	0.1216	0.0000
1965	0.2865	0.9412
1966	0.2662	0.9978
1967	0.2135	0.0000
1968	0.0649	0.0000

1969	0.2378	0.0000
1970	0.2932	0.9259
1971	0.523	0.5994
1972	0.3811	0.4645
1973	0.3635	0.9444
1974	0.1568	0.0000
1975	0.3514	0.7317
1976	0.2716	1.005
1977	0.1649	0.0000
1978	0.0892	0.0000
1979	0.3122	0.8481
1980	0.3324	0.8555
1981	0.3743	0.8105
1982	0.4635	0.7111
1983	1	0.3417
1984	0.3946	0.8560
1985	0.0608	0.0000
1986	0.2122	0.0000
1987	0.4405	0.6026
1988	0.3432	0.8925
1989	0.3743	0.4555
1990	0.4473	0.5788
1991	0.1824	0.0000
1992	0.7946	0.0428
1993	0.5716	0.5566
1994	0.2405	0.0000
1995	0.5041	0.6306
1996	0.2932	0.9121
1997	0.4919	0.9396
1998	0.5649	0.5862
1999	0.3622	0.6878
2000	0.4689	0.4550
2001	0.3973	0.6920
2002	0.1865	0.0000
2003	0.2054	0.0000
2004	0.2811	0.8989
2005	0.3784	0.7756
2006	0.0014	0.0000
2007	0.3716	0.8211
2008	0.2743	0.9592
2009	0.4149	0.9126
2010	0.5378	0.4698
2011	0.4905	0.7142
2012	0.3595	0.7671
2013	0.4932	0.5179
2014	0.6905	0.3566
2015	0.3797	0.6222

```
i2 = data['I2'].values  
i7 = data['I7'].values
```

```

from sklearn.cluster import KMeans
X = np.array(list(zip(i2.i7)))
kmeans = KMeans(n_clusters = 4)
kmeans = kmeans.fit(X)
labels = kmeans.predict(X)
centroids = kmeans.cluster_centers_
print(centroids)

[[0.3450310344827586, 0.8810965517241378]
[0.15282812500000004, 2.7755575615628914e-16]
[0.46567499999999995, 0.58826]
[0.794925, 0.184875]]

def mahalanobis(x=None, cov=None)
    #compute the Mahalanobis Distance Between each of x and the data
    x: vector or matrix of data with .say.p colum
    data: ndarray of the distribution from which Mahalanobis distance of each
    observation of x is
    cov: covariance matrix (pxp) of the distribution. If None, will be computed
    from data
    """
    x_minus_mu = x - np.mean(data)
    if not cov:
        inv_covmat = sp.linalg.inv_covamat
        left_term = np.dot(x_minus_mu, inv_covmat)
        mahal = np.dot(left_term, x_minus_mu.T)
    """
    df_x = df[['I2', 'I7']].head(85)
    df_x['mahala'] = mahalanobis(x=df_x.data=df[['I2', 'I7']])
    tabela = df_x['mahala'].np.array(list(zip(i2, i7, labels)))
    columns = ['I2', 'I7', 'Centroides']

    I2.I7.Centroides
    0.0.2811.0.902116402116402.2.0
    1.0.3811.0.635449735449735.3.0
    2.0.0959.0.0.1.0
    3.0.0905.0.0.1.0
    4.0.6946.0.0.4.0
    5.0.2243.0.0.1.0
    6.0.2662.0.932275132275132.2.0
    7.0.5135.0.6671957671957671.3.0
    8.0.3527.0.8941798941798941.2.0
    9.0.0.0.0.1.0
    10.0.2324.0.0.1.0
    11.0.2095.0.0.1.0
    12.0.1176.0.0.1.0
    13.0.1351.0.0.1.0
    14.0.2162.0.0.1.0
    15.0.5014.0.6306878306878311.3.0
    16.0.3676.0.8105820105820111.3.0
    17.0.2784.0.9079365079365079.2.0
    18.0.0851.0.0.1.0
    19.0.3243.0.782539682539683.2.0
    20.0.2081.0.0.1.0
    21.0.1784.0.0.1.0

```

22.0.3181.0.8608465608465609.2.0
23.0.5176.0.498412698412698.3.0
24.0.4189.0.919047619047619.3.0
25.0.2216.0.0.1.0
26.0.5757.0.9010582010582009.4.0
27.0.1595.0.0.1.0
28.0.0851.0.0.1.0
29.0.1784.0.0.1.0
30.0.3689.0.866137566137566.2.0
31.0.15.0.0.1.0
32.0.1649.0.0.1.0
33.0.1216.0.0.1.0
34.0.2865.0.9412698412698409.2.0
35.0.2662.0.997883597883598.2.0
36.0.2135.0.0.1.0
37.0.0649.0.0.1.0
38.0.2378.0.0.1.0
39.0.2932.0.9259259259259259.2.0
40.0.523.0.5994708994708989.3.0
41.0.3811.0.46455026455026394.3.0
42.0.3635.0.9444444444444444.2.0
43.0.1568.0.0.1.0
44.0.3514.0.7317460317460319.2.0
45.0.2716.1.0005291005291.2.0
46.0.1649.0.0.1.0
47.0.0892.0.0.1.0
48.0.3122.0.8481481481481479.2.0
49.0.3324.0.8555555555555556.2.0
50.0.3743.0.8105820105820111.2.0
51.0.4635.0.7111111111111109.3.0
52.1.0.0.341798941798942.4.0
53.0.3946.0.8560846560846558.3.0
54.0.0608.0.0.1.0
55.0.2122.0.0.1.0
56.0.4405.0.6026455026455029.3.0
57.0.3432.0.892592592592592.2.0
58.0.3743.0.45555555555555605.3.0
59.0.4473.0.578835978835979.3.0
60.0.1824.0.0.1.0
61.0.7946.0.0428571428571429.4.0
62.0.5716.0.5566137566137571.3.0
63.0.2405.0.0.1.0
64.0.5041.0.6306878306878311.3.0
65.0.2932.0.912169312169312.2.0
66.0.4919.0.93968253968254.3.0
67.0.5649.0.586243386243386.4.0
68.0.3622.0.687830687830688.2.0
69.0.4689.0.455026455026455.3.0
70.0.3973.0.692063492063492.3.0
71.0.1865.0.0.1.0

```

72.0.2054.0.0.1.0
73.0.2811.0.8989417989417992.2.0
74.0.3784.0.7756613756613759.3.0
75.0.0014.0.0.1.0
76.0.3716.0.8211640211640211.2.0
77.0.2743.0.9592592592592591.2.0
78.0.4149.0.912698412698413.3.0
79.0.5378.0.46984126984127.3.0
80.0.4905.0.7142857142857142.3.0
81.0.3595.0.7671957671957672.2.0
82.0.4932.0.517989417989418.3.0
83.0.6905.0.35661375661375605.4.0
84.0.3797.0.622222222222222.3.0

tabela.to_csv("C:/Users/55419/Desktop/Arquivos_ULTIMOS/ARQUIVOS
CVS_ALTERADOS/CSV/NORMALIZAÇÃO LINEAR_I2-I7_ALTERADOS/Valores
LINEARES_ALTERADOS_agrupados.csv")

centroides_df = pd.DataFrame(centroids)

centroides_df.to_csv("C:/Users/55419/Desktop/Arquivos
10_03_ULTIMOS/ARQUIVOS CVS_ALTERADOS/CSV/NORMALIZAÇÃO
LINEAR_ALTERADOS/Valores LINEAR_I2-17_ALTERADOS_Centróides.csv")

```

APÊNDICE 31 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORITMO K-MEANS. CONSIDERANDO APENAS OS INDICADORES DE SEVERIDADE I_2 E I_7 . NORMALIZADOS ATRAVÉS DO MÉTODO DA NORMALIZAÇÃO POR DESVIO PADRÃO, E COM VALORES DE I_7 ALTERADOS; PROGRAMA EXECUTADO EM LINGUAGEM DE PROGRAMAÇÃO PYTHON.

```
import pandas as pd
import numpy as np
def dist(a,b,ax=1):
    return np.linalg.norm(a-b, axis = ax)
data = pd.read_csv("C:/Users//55419/Desktop/Kmeans
Normalizações/Normalização Linear_I2-I7_ALTERADOS/NPLADA.csv")
data
```

ANO	I2	I7
1931	-0.2317	1.1528
1932	0.3330	0.4677
1933	-1.2771	-1.1649
1934	-1.3076	-1.1649
1935	2.1032	-1.1649
1936	-0.5522	-1.1649
1937	-0.3156	1.2303
1938	1.0807	0.5493
1939	0.1727	1.1324
1940	-1.8188	-1.1649
1941	-0.5064	-1.1649
1942	-0.6361	-1.1649
1943	-1.1550	-1.1649
1944	-1.0558	-1.1649
1945	-0.5980	-1.1649
1946	1.0121	0.4555
1947	0.2567	0.9176
1948	-0.2470	1.1677
1949	-1.3381	-1.1649
1950	0.0125	0.8456
1951	-0.6437	-1.1649
1952	-0.8116	-1.1649
1953	-0.0791	1.0468
1954	1.1036	0.1156
1955	0.5466	1.1963
1956	-0.5674	-1.1649
1957	1.4317	1.1501
1958	-0.9184	-1.1649
1959	-1.3381	-1.1649
1960	-0.8116	-1.1649
1961	0.2643	1.0604
1962	-0.9719	-1.1649
1963	-0.8879	-1.1649
1964	-1.1321	-1.1649
1965	-0.2012	1.2534
1966	-0.3156	1.3988
1967	-0.6132	-1.1649

1968	-1.4526	-1.1649
1969	-0.4759	-1.1649
1970	-0.1630	1.2140
1971	1.1342	0.3753
1972	0.3330	0.0286
1973	0.2338	1.2615
1974	-0.9337	-1.1649
1975	0.1651	0.7151
1976	-0.2851	1.4056
1977	-0.8879	-1.1649
1978	-1.3152	-1.1649
1979	-0.0562	1.0141
1980	0.0583	1.0332
1981	0.2948	0.9176
1982	0.7984	0.6621
1983	3.8277	-0.2867
1984	0.4093	1.0345
1985	-1.4755	-1.1649
1986	-0.6209	-1.1649
1987	0.6687	0.3834
1988	0.1193	1.1283
1989	0.2948	0.0055
1990	0.7069	0.3222
1991	-0.7887	-1.1649
1992	2.6679	-1.0548
1993	1.4089	0.2652
1994	-0.4606	-1.1649
1995	1.0273	0.4555
1996	-0.1630	1.1786
1997	0.9587	1.2493
1998	1.3707	0.3413
1999	0.2261	0.6023
2000	0.8289	0.0042
2001	0.4245	0.6131
2002	-0.7658	-1.1649
2003	-0.6590	-1.1649
2004	-0.2317	1.1446
2005	0.3177	0.8279
2006	-1.8112	-1.1649
2007	0.2795	0.9448
2008	-0.2699	1.2996
2009	0.5237	1.1800
2010	1.2181	0.0422
2011	0.9510	0.6702
2012	0.2109	0.8062
2013	0.9663	0.1659
2014	2.0803	-0.2487
2015	0.3253	0.4337

```
i2 = data['I2'].values
```

```

i7 = data['I7'].values

from sklearn.cluster import KMeans
X = np.array(list(zip(i2.i7)))
kmeans = KMeans(n_clusters = 4)
kmeans = kmeans.fit(X)
labels = kmeans.predict(X)
centroids = kmeans.cluster_centers_
print(centroids)

[[0.8750867995714287.0.4169434178095238]
[-0.9558787516249999.-1.1648685439999997]
[2.6697961692499996.-0.6887617212500001]
[0.05471830653571429.1.07536285675]

def mahalanobis(x=None,cov=None)
    #compute the Mahalanobis Distance Between each of x and the data
    x: vector or matrix of data with say p column
    data: ndarray of the distribution from which Mahalanobis distance of each
    observation of x is
    cov: covariance matrix (pxp) of the distribution. If None, will be computed
    from data
    """
    x_minus_mu = x-np.mean(data)
    if not cov:
        inv_covmat = sp.linalg.inv_covamat
        left_term = np.dot(x_minus_mu,inv_covmat)
        mahal = np.dot(left_term,x_minus_mu.T)
    """

    df_x = df[['I2','I7']].head(85)
    df_x['mahala']=mahalanobis(x=df_x.data=df[['I2','I7']])
    tabela = df_x['mahala'](np.array(list(zip(i2.i7, labels))))
    columns = ['I2', 'I7','Centroide']

    I2.I7.Centroide
    0.-0.231697579.1.1527963620000001.3.0
    1.0.332958667.0.467691897.3.0
    2.-1.277074682.-1.164868544.1.0
    3.-1.307596641.-1.164868544.1.0
    4.2.103232301.-1.164868544.4.0
    5.-0.55217815.-1.164868544.1.0
    6.-0.315632966.1.230278414.2.0
    7.1.080746668.0.549251952.3.0
    8.0.172718381.1.132406348.2.0
    9.-1.818839458.-1.164868544.1.0
    10.-0.506395212.-1.164868544.1.0
    11.-0.6361135379999999.-1.164868544.1.0
    12.-1.154986845.-1.164868544.1.0
    13.-1.055790477.-1.164868544.1.0
    14.-0.597961089.-1.164868544.1.0
    15.1.012072259.0.455457888.3.0
    16.0.256653769.0.917631535.3.0
    17.-0.246958558.1.167749039.2.0
    18.-1.3381186.-1.164868544.1.0

```

19.0.012478095.0.84558682.2.0
20.-0.643744028.-1.164868544.1.0
21.-0.8116148040000001.-1.164868544.1.0
22.-0.079087783.1.04676829.2.0
23.1.1036381370000001.0.115624324.3.0
24.0.546612381.1.196295058.3.0
25.-0.56743913.-1.164868544.1.0
26.1.431749199.1.150077693.4.0
27.-0.9184416609999999.-1.164868544.1.0
28.-1.3381186.-1.164868544.1.0
29.-0.8116148040000001.-1.164868544.1.0
30.0.26428425899999997.1.060361632.2.0
31.-0.97185509.-1.164868544.1.0
32.-0.887919702.-1.164868544.1.0
33.-1.132095375.-1.164868544.1.0
34.-0.201175618999999997.1.253387097.2.0
35.-0.315632966.1.3988358619999999.2.0
36.-0.613222069.-1.164868544.1.0
37.-1.452575947.-1.164868544.1.0
38.-0.475873252.-1.164868544.1.0
39.-0.16302317.1.213966403.2.0
40.1.134160096.0.375257167.3.0
41.0.332958667.0.028626932.3.0
42.0.233762298999999998.1.261543102.2.0
43.-0.9337026409999999.-1.164868544.1.0
44.0.165087891.0.715090731.2.0
45.-0.285111006999999997.1.4056325330000001.2.0
46.-0.887919702.-1.164868544.1.0
47.-1.31522713099999999.-1.164868544.1.0
48.-0.056196313.1.014144268.2.0
49.0.058261034.1.033174947.2.0
50.0.294806218.0.917631535.2.0
51.0.798418545.0.662076695.2.0
52.3.827722997.-0.286738615.4.0
53.0.409263565.1.034534281.3.0
54.-1.475467417.-1.164868544.1.0
55.-0.620852559.-1.164868544.1.0
56.0.668700218.0.383413173.3.0
57.0.119304951999999999.1.128328345.2.0
58.0.294806218.0.00551825.3.0
59.0.706852667.0.322243131.3.0
60.-0.788723334.-1.164868544.1.0
61.2.667888547.-1.05476247.4.0
62.1.4088577290000002.0.265151093.4.0
63.-0.460612272999999996.-1.164868544.1.0
64.1.027333239.0.455457888.3.0
65.-0.16302317.1.178623713.2.0
66.0.95865883099999999.1.249309094.3.0
67.1.37070528.0.34127381100000004.4.0
68.0.22613181.0.602265988.2.0

```

69.0.8289405040000001.0.0041589159999999995.3.0
70.0.424524544.0.6131406620000001.3.0
71.-0.765831865.-1.164868544.1.0
72.-0.659005008.-1.164868544.1.0
73.-0.231697579.1.144640356.2.0
74.0.317697687.0.827915475.3.0
75.-1.811208968.-1.164868544.1.0
76.0.27954523800000003.0.9448182209999999.2.0
77.-0.269850028.1.299604461.2.0
78.0.523720912.1.1799830470000001.3.0
79.1.218095484.0.042220275.3.0
80.0.9510283409999999.0.670232701.2.0
81.0.21087083.0.8061661259999999.2.0
82.0.966289321.0.16591969199999998.3.0
83.2.080340832.-0.248677256.4.0
84.0.325328177.0.43370854.3.0

```

```

tabela.to_csv("C:/Users/55419/Desktop/Arquivos_ULTIMOS/ARQUIVOS
CVS_ALTERADOS/CSV/NORMALIZAÇÃO DP_I2-I7_ALTERADOS/Valores
DP_ALTERADOS_agrupados.csv")

```

```

centroides_df = pd.DataFrame(centroids)

```

```

centroides_df.to_csv("C:/Users/55419/Desktop/Arquivos
10_03_ULTIMOS/ARQUIVOS CVS_ALTERADOS/CSV/NORMALIZAÇÃO DP_ALTERADOS/Valores
DP_I2-17_ALTERADOS_Centróides.csv")

```

APÊNDICE 32 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORITMO K-MEANS. CONSIDERANDO APENAS OS INDICADORES DE SEVERIDADE I_2 E I_7 . NORMALIZADOS ATRAVÉS DO MÉTODO DO VALOR MÁXIMO DOS ELEMENTOS. E COM VALORES DE I_7 ALTERADOS; PROGRAMA EXECUTADO EM LINGUAGEM DE PROGRAMAÇÃO PYTHON.

```
import pandas as pd
import numpy as np
def dist(a,b,ax=1):
    return np.linalg.norm(a-b, axis = ax)
data = pd.read_csv("C:/Users//55419/Desktop/Kmeans
Normalizações/Normalização VM_I2-I7_ALTERADOS/NPLADA.csv")
data
```

ANO	I2	I7
1931	0.489443	0.902116
1932	0.560461	0.63545
1933	0.357965	0.0000
1934	0.354127	0.0000
1935	0.783109	0.0000
1936	0.449136	0.0000
1937	0.478887	0.932275
1938	0.654511	0.667196
1939	0.540307	0.89418
1940	0.289827	0.0000
1941	0.454894	0.0000
1942	0.43858	0.0000
1943	0.373321	0.0000
1944	0.385797	0.0000
1945	0.443378	0.0000
1946	0.645873	0.630688
1947	0.550864	0.810582
1948	0.487524	0.907937
1949	0.350288	0.0000
1950	0.520154	0.78254
1951	0.43762	0.0000
1952	0.416507	0.0000
1953	0.508637	0.860847
1954	0.65739	0.498413
1955	0.587332	0.919048
1956	0.447217	0.0000
1957	0.698656	0.901058
1958	0.403071	0.0000
1959	0.350288	0.0000
1960	0.416507	0.0000
1961	0.551823	0.866138
1962	0.396353	0.0000
1963	0.40691	0.0000
1964	0.3762	0.0000
1965	0.493282	0.94127
1966	0.478887	0.997884

1967	0.441459	0.0000
1968	0.335893	0.0000
1969	0.458733	0.0000
1970	0.498081	0.925926
1971	0.661228	0.599471
1972	0.560461	0.46455
1973	0.547985	0.944444
1974	0.401152	0.0000
1975	0.539347	0.731746
1976	0.482726	1.000529
1977	0.40691	0.0000
1978	0.353167	0.0000
1979	0.511516	0.848148
1980	0.525912	0.855556
1981	0.555662	0.810582
1982	0.619002	0.711111
1983	1	0.341799
1984	0.570058	0.856085
1985	0.333013	0.0000
1986	0.440499	0.0000
1987	0.602687	0.602646
1988	0.533589	0.892593
1989	0.555662	0.455556
1990	0.607486	0.578836
1991	0.419386	0.0000
1992	0.854127	0.042857
1993	0.695777	0.556614
1994	0.460653	0.0000
1995	0.647793	0.630688
1996	0.498081	0.912169
1997	0.639155	0.939683
1998	0.690979	0.586243
1999	0.547025	0.687831
2000	0.622841	0.455026
2001	0.571977	0.692063
2002	0.422265	0.0000
2003	0.435701	0.0000
2004	0.489443	0.898942
2005	0.558541	0.775661
2006	0.290787	0.0000
2007	0.553743	0.821164
2008	0.484645	0.959259
2009	0.584453	0.912698
2010	0.671785	0.469841
2011	0.638196	0.714286
2012	0.545106	0.767196
2013	0.640115	0.517989
2014	0.78023	0.356614
2015	0.559501	0.622222

```

i2 = data['I2'].values
i7 = data['I7'].values

from sklearn.cluster import KMeans
X = np.array(list(zip(i2,i7)))
kmeans = KMeans(n_clusters = 4)
kmeans = kmeans.fit(X)
labels = kmeans.predict(X)
centroids = kmeans.cluster_centers_
print(centroids)

[[0.6495921305182342.0.11883125][0.3722221111114444444456.0.0002352223341]
[0.39836252399232247.-4.163336342344337e-17]
[0.5309428255685453.0.15626969696969698]

def mahalanobis(x=None,cov=None)
    #compute the Mahalanobis Distance Between each of x and the data
    x: vector or matrix of data with,say,p colum
    data:ndarray of the distribution from which Mahalanobis distance of each
    observation of x is
    cov:covariance matrix (pxp)of the distribution.If None.will be computed
    from data
    """
    x_minus_mu = x-np.mean(data)
    if not cov:
        inv_covmat = sp.linalg.inv_covamat
        left_term = np.dot(x_minus_mu.inv_covmat)
        mahal = np.dot(left_term.x_minus_mu.T)
    """
    df_x = df[['I2','I7']].head(85)
    df_x['mahala']=mahalanobis(x=df_x.data=df[['I2','I7']])
    tabela = df_x['mahala'](np.array(list(zip(i2,i7. labels)))).

columns = ['I2'. 'I7'.'Centroide'])

I2.I7.Centroide
0.0.48944337811900196.0.1705.3.0
1.0.5604606525911711.0.1201.3.0
2.0.357965451055662.0.0.1.0
3.0.35412667946257204.0.0.1.0
4.0.7831094049904029.0.0.4.0
5.0.449136276391555.0.0.1.0
6.0.478886756238004.0.1762.2.0
7.0.654510556621881.0.1261.3.0
8.0.540307101727447.0.16898.2.0
9.0.28982725527831105.0.0.1.0
10.0.45489443378119.0.0.1.0
11.0.43857965451055697.0.0.1.0
12.0.3733205374280229.0.0.1.0
13.0.385796545105566.0.0.1.0
14.0.44337811900191904.0.0.1.0
15.0.645873320537428.0.1192.2.0
16.0.5508637236084449.0.1532.3.0
17.0.4875239923224571.0.1716.2.0
18.0.35028790786948194.0.0.1.0
19.0.520153550863724.0.1479.3.0

```

20.0.43761996161228395.0.0.1.0
21.0.4165067178502879.0.0.1.0
22.0.508637236084453.0.1627.2.0
23.0.657389635316699.0.0942.3.0
24.0.587332053742802.0.1737.3.0
25.0.44721689059501.0.0.1.0
26.0.6986564299424181.0.1703.4.0
27.0.40307101727447203.0.0.1.0
28.0.35028790786948194.0.0.1.0
29.0.4165067178502879.0.0.1.0
30.0.5518234165067181.0.1637.3.0
31.0.396353166986564.0.0.1.0
32.0.406909788867562.0.0.1.0
33.0.37619961612284103.0.0.1.0
34.0.49328214971209206.0.1779.2.0
35.0.478886756238004.0.1886.2.0
36.0.44145873320537393.0.0.1.0
37.0.33589251439539297.0.0.1.0
38.0.4587332053742801.0.0.1.0
39.0.498080614203455.0.175.2.0
40.0.6612284069097889.0.1133.3.0
41.0.5604606525911711.0.0878.3.0
42.0.547984644913628.0.1785.2.0
43.0.401151631477927.0.0.1.0
44.0.539347408829175.0.1383.2.0
45.0.48272552783109396.0.1891.2.0
46.0.406909788867562.0.0.1.0
47.0.353166986564299.0.0.1.0
48.0.511516314779271.0.1603.2.0
49.0.525911708253359.0.1617.2.0
50.0.5556621880998079.0.1532.2.0
51.0.619001919385797.0.1344.2.0
52.1.0.0.0646.4.0
53.0.570057581573896.0.1618.3.0
54.0.333013435700576.0.0.1.0
55.0.4404990403071021.0.0.1.0
56.0.602687140115163.0.1139.3.0
57.0.533589251439539.0.1687.2.0
58.0.5556621880998079.0.0861.3.0
59.0.607485604606526.0.1094.2.0
60.0.41938579654510605.0.0.1.0
61.0.854126679462572.0.0081.4.0
62.0.6957773512476011.0.1052.3.0
63.0.46065259117082497.0.0.1.0
64.0.647792706333973.0.1192.3.0
65.0.498080614203455.0.1724.2.0
66.0.63915547024952.0.1776.2.0
67.0.6909788867562379.0.1108.4.0
68.0.5470249520153551.0.13.2.0
69.0.6228406909788871.0.086.3.0

```

70.0.571976967370441.0.1308.3.0
71.0.422264875239923.0.0.1.0
72.0.435700575815739.0.0.1.0
73.0.48944337811900196.0.1699.2.0
74.0.558541266794626.0.1466.3.0
75.0.29078694817658396.0.0.1.0
76.0.553742802303263.0.1552.2.0
77.0.48464491362763895.0.1813.2.0
78.0.584452975047985.0.1725.3.0
79.0.671785028790787.0.0888.3.0
80.0.6381957773512479.0.135.2.0
81.0.54510556621881.0.145.2.0
82.0.6401151631477929.0.0979.3.0
83.0.7802303262955849.0.0674.4.0
84.0.5595009596928979.0.1176.3.0

```

```

tabela.to_csv("C:/Users/55419/Desktop/Arquivos_ULTIMOS/ARQUIVOS
CVS_ALTERADOS/CSV/NORMALIZAÇÃO VM_I2-I7_ALTERADOS/Valores
VM_ALTERADOS_agrupados.csv")

```

```

centroides_df = pd.DataFrame(centroids)

```

```

centroides_df.to_csv("C:/Users/55419/Desktop/Arquivos
10_03_ULTIMOS/ARQUIVOS CVS_ALTERADOS/CSV/NORMALIZAÇÃO VM_ALTERADOS/Valores
VM_I2-17_ALTERADOS_Centróides.csv")

```

APÊNDICE 33 – AGRUPAMENTO DE DADOS EXECUTADO ATRAVÉS DO ALGORITMO K-MEANS. CONSIDERANDO APENAS OS INDICADORES DE SEVERIDADE I_2 E I_7 . NORMALIZADOS ATRAVÉS DO MÉTODO DA ESCALA DECIMAL E COM VALORES DE I_7 ALTERADOS; PROGRAMA EXECUTADO EM LINGUAGEM DE PROGRAMAÇÃO PYTHON.

```
import pandas as pd
import numpy as np
def dist(a,b,ax=1):
    return np.linalg.norm(a-b, axis = ax)
data = pd.read_csv("C:/Users//55419/Desktop/Kmeans
Normalizações/Normalização EC_I2-I7_ALTERADOS/NPEDDA.csv")
data
```

ANO	I2	I7
1931	0.051	0.1705
1932	0.0584	0.1201
1933	0.0373	0.0000
1934	0.0369	0.0000
1935	0.0816	0.0000
1936	0.0468	0.0000
1937	0.0499	0.1762
1938	0.0682	0.1261
1939	0.0563	0.169
1940	0.0302	0.0000
1941	0.0474	0.0000
1942	0.0457	0.0000
1943	0.0389	0.0000
1944	0.0402	0.0000
1945	0.0462	0.0000
1946	0.0673	0.1192
1947	0.0574	0.1532
1948	0.0508	0.1716
1949	0.0365	0.0000
1950	0.0542	0.1479
1951	0.0456	0.0000
1952	0.0434	0.0000
1953	0.053	0.1627
1954	0.0685	0.0942
1955	0.0612	0.1737
1956	0.0466	0.0000
1957	0.0728	0.1703
1958	0.042	0.0000
1959	0.0365	0.0000
1960	0.0434	0.0000
1961	0.0575	0.1637
1962	0.0413	0.0000
1963	0.0424	0.0000
1964	0.0392	0.0000
1965	0.0514	0.1779
1966	0.0499	0.1886
1967	0.046	0.0000

1968	0.035	0.0000
1969	0.0478	0.0000
1970	0.0519	0.175
1971	0.0689	0.1133
1972	0.0584	0.0878
1973	0.0571	0.1785
1974	0.0418	0.0000
1975	0.0562	0.1383
1976	0.0503	0.1891
1977	0.0424	0.0000
1978	0.0368	0.0000
1979	0.0533	0.1603
1980	0.0548	0.1617
1981	0.0579	0.1532
1982	0.0645	0.1344
1983	0.1042	0.0646
1984	0.0594	0.1618
1985	0.0347	0.0000
1986	0.0459	0.0000
1987	0.0628	0.1139
1988	0.0556	0.1687
1989	0.0579	0.0861
1990	0.0633	0.1094
1991	0.0437	0.0000
1992	0.089	0.0081
1993	0.0725	0.1052
1994	0.048	0.0000
1995	0.0675	0.1192
1996	0.0519	0.1724
1997	0.0666	0.1776
1998	0.072	0.1108
1999	0.057	0.13
2000	0.0649	0.086
2001	0.0596	0.1308
2002	0.044	0.0000
2003	0.0454	0.0000
2004	0.051	0.1699
2005	0.0582	0.1466
2006	0.0303	0.0000
2007	0.0577	0.1552
2008	0.0505	0.1813
2009	0.0609	0.1725
2010	0.07	0.0888
2011	0.0665	0.135
2012	0.0568	0.145
2013	0.0667	0.0979
2014	0.0813	0.0674
2015	0.0583	0.1176

```
i2 = data['I2'].values
```

```

i7 = data['I7'].values

from sklearn.cluster import KMeans
X = np.array(list(zip(i2.i7)))
kmeans = KMeans(n_clusters = 4)
kmeans = kmeans.fit(X)
labels = kmeans.predict(X)
centroids = kmeans.cluster_centers_
print(centroids)

[[0.0.04408529411764706.0.00023823529411760858]
[0.06284375.0.12560625]
[0.05555.0.1693269230769231]
[0.0716.0.08644444444444445]]

def mahalanobis(x=None,cov=None)
    #compute the Mahalanobis Distance Between each of x and the data
x: vector or matrix of data with say p column
data: ndarray of the distribution from which Mahalanobis distance of each
observation of x is
cov: covariance matrix (pxp) of the distribution. If None, will be computed
from data
"""
x_minus_mu = x-np.mean(data)
if not cov:
    inv_covmat = sp.linalg.inv_covamat
    left_term = np.dot(x_minus_mu,inv_covmat)
    mahal = np.dot(left_term,x_minus_mu.T)
"""
df_x = df[['I2','I7']].head(85)
df_x['mahala']=mahalanobis(x=df_x.data=df[['I2','I7']])
tabela = df_x['mahala'](np.array(list(zip(i2.i7, labels))))
columns = ['I2', 'I7', 'Centroide']

I2.I7.Centroide
0.0.051.0.1705.3.0
1.0.0584.0.1201.3.0
2.0.0373.0.0.1.0
3.0.0369.0.0.1.0
4.0.0816.0.0.4.0
5.0.0468.0.0.1.0
6.0.0499.0.1762.2.0
7.0.0682.0.1261.3.0
8.0.0563.0.168998.2.0
9.0.0302.0.0.1.0
10.0.0474.0.0.1.0
11.0.0457.0.0.1.0
12.0.0389.0.0.1.0
13.0.0402.0.0.1.0
14.0.0462.0.0.1.0
15.0.0673.0.1192.3.0
16.0.0574.0.1532.3.0

```

17.0.0508.0.1716.2.0
18.0.0365.0.0.1.0
19.0.0542.0.1479.2.0
20.0.0456.0.0.1.0
21.0.0434.0.0.1.0
22.0.053.0.1627.2.0
23.0.0685.0.0942.3.0
24.0.0612.0.1737.2.0
25.0.0466.0.0.1.0
26.0.0728.0.1703.4.0
27.0.042.0.0.1.0
28.0.0365.0.0.1.0
29.0.0434.0.0.1.0
30.0.0575.0.1637.2.0
31.0.0413.0.0.1.0
32.0.0424.0.0.1.0
33.0.0392.0.0.1.0
34.0.0514.0.1779.2.0
35.0.0499.0.1886.2.0
36.0.046.0.0.1.0
37.0.035.0.0.1.0
38.0.0478.0.0.1.0
39.0.0519.0.175.2.0
40.0.0689.0.1133.3.0
41.0.0584.0.0878.3.0
42.0.0571.0.1785.2.0
43.0.0418.0.0.1.0
44.0.0562.0.1383.2.0
45.0.0503.0.1891.2.0
46.0.0424.0.0.1.0
47.0.0368.0.0.1.0
48.0.0533.0.1603.2.0
49.0.0548.0.1617.2.0
50.0.0579.0.1532.2.0
51.0.0645.0.1344.2.0
52.0.1042.0.0646.4.0
53.0.0594.0.1618.3.0
54.0.0347.0.0.1.0
55.0.0459.0.0.1.0
56.0.0628.0.1139.3.0
57.0.0556.0.1687.2.0
58.0.0579.0.0861.3.0
59.0.0633.0.1094.3.0
60.0.0437.0.0.1.0
61.0.08901.0.0081.4.0
62.0.0725.0.1052.3.0
63.0.048.0.0.1.0
64.0.0675.0.1192.3.0
65.0.0519.0.1724.2.0
66.0.0666.0.1776.3.0

```

67.0.07201.0.1108.4.0
68.0.057.0.13.2.0
69.0.0649.0.086.3.0
70.0.0596.0.1308.3.0
71.0.0440004.0.0.1.0
72.0.0454.0.0.1.0
73.0.051.0.1699.2.0
74.0.0582.0.1466.3.0
75.0.0303.0.0.1.0
76.0.0577.0.1552.2.0
77.0.0505.0.1813.2.0
78.0.0609.0.1725.3.0
79.0.07.0.0888.3.0
80.0.0665.0.135.2.0
81.0.0568.0.145.2.0
82.0.0667.0.0979.3.0
83.0.0813.0.0674.4.0
84.0.0583.0.1176.3.0

```

```

tabela.to_csv("C:/Users/55419/Desktop/Arquivos_ULTIMOS/ARQUIVOS
CVS_ALTERADOS/CSV/NORMALIZAÇÃO ED_I2-I7_ALTERADOS/Valores
ED_ALTERADOS_agrupados.csv")

```

```

centroides_df = pd.DataFrame(centroids)

```

```

centroides_df.to_csv("C:/Users/55419/Desktop/Arquivos
10_03_ULTIMOS/ARQUIVOS CVS_ALTERADOS/CSV/NORMALIZAÇÃO ED_ALTERADOS/Valores
ED_I2-17_ALTERADOS_Centróides.csv")

```

APÊNDICE 34 – TESTE *T* DE STUDENT, EXECUTADO PARA DUAS AMOSTRAS PAREADAS; VALORES ORIGINAIS.

	<i>I</i> ₁ -NormLinear	<i>I</i> ₁ -Normalização por Desvio Padrão
Média	0,244844706	0,740706947
Variância	0,0274402	0,44482171
Observações	85	85
Correlação de Pearson	0,502383885	
Hipótese da diferença	0	
gl	84	
Stat t	-7,606133128	
P(T<=t) uni-caudal	1,85208E-11	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	3,70416E-11	
t crítico bi-caudal	1,988609667	

	<i>I</i> ₁ -NormLinear	<i>I</i> ₁ -Normalização por Escala Decimal
Média	0,244844706	0,169161176
Variância	0,0274402	0,005779233
Observações	85	85
Correlação de Pearson	0,999999845	
Hipótese da diferença	0	
gl	84	
Stat t	7,785014507	
P(T<=t) uni-caudal	8,15486E-12	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	1,63097E-11	
t crítico bi-caudal	1,988609667	

	<i>I</i> ₂ -Normalização Linear	<i>I</i> ₂ -Normalização por Valor Máximo do Elemento
Média	0,322230588	0,518584171
Variância	0,031735233	0,01600661
Observações	85	85
Correlação de Pearson	0,999981477	
Hipótese da diferença	0	
gl	84	
Stat t	-35,05956796	
P(T<=t) uni-caudal	3,17462E-52	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	6,34924E-52	
t crítico bi-caudal	1,988609667	

	<i>I</i> ₂ -Normalização Linear	<i>I</i> ₂ -Normalização por Desvio Padrão
Média	0,070085882	0,586048925
Variância	0,02047224	0,652457927
Observações	85	85
Correlação de Pearson	0,906011455	
Hipótese da diferença	0	
gl	84	
Stat t	-6,987139829	
P(T<=t) uni-caudal	3,0843E-10	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	6,16859E-10	
t crítico bi-caudal	1,988609667	

	<i>I</i> ₂ -Normalização Linear	<i>I</i> ₂ -Normalização por Escala Decimal
Média	0,070085882	0,067883529
Variância	0,02047224	0,019221916
Observações	85	85
Correlação de Pearson	0,999999062	
Hipótese da diferença	0	
gl	84	
Stat t	4,57075936	
P(T<=t) uni-caudal	8,2804E-06	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	1,65608E-05	
t crítico bi-caudal	1,988609667	

	<i>I</i> ₄ -Normalização Linear	<i>I</i> ₄ -Normalização por Valor Máximo
Média	0,165758824	0,16444121
Variância	0,045279263	0,044868492
Observações	85	85
Correlação de Pearson	0,998355096	
Hipótese da diferença	0	
gl	84	
Stat t	0,994456962	
P(T<=t) uni-caudal	0,161428265	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	0,32285653	
t crítico bi-caudal	1,988609667	

	<i>I</i> ₃ -Normalização Linear	<i>I</i> ₃ -Normalização por Desvio Padrão
Média	0,448894737	0,807572034
Variância	0,148262482	0,34006345
Observações	85	85
Correlação de Pearson	-0,473635768	
Hipótese da diferença	0	
gl	84	
Stat t	-3,949534519	
P(T<=t) uni-caudal	8,11302E-05	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	0,00016226	
t crítico bi-caudal	1,988609667	

	<i>I</i> ₁ -NormLinear	<i>I</i> ₁ -Normalização por Valor Máximo do Elemento
Média	0,244844706	0,328022448
Variância	0,0274402	0,021730812
Observações	85	85
Correlação de Pearson	0,999999845	
Hipótese da diferença	0	
gl	84	
Stat t	-42,04903609	
P(T<=t) uni-caudal	1,67853E-58	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	3,35707E-58	
t crítico bi-caudal	1,988609667	

	<i>I</i> ₂ -Normalização Linear	<i>I</i> ₂ -Normalização por Desvio Padrão
Média	0,322230588	0,775332855
Variância	0,031735233	0,391702523
Observações	85	85
Correlação de Pearson	0,306075861	
Hipótese da diferença	0	
gl	84	
Stat t	-7,009341639	
P(T<=t) uni-caudal	2,79068E-10	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	5,58137E-10	
t crítico bi-caudal	1,988609667	

	<i>I</i> ₂ -Normalização Linear	<i>I</i> ₂ -Normalização por Escala Decimal
Média	0,322230588	0,054036471
Variância	0,031735233	0,000173794
Observações	85	85
Correlação de Pearson	0,999981477	
Hipótese da diferença	0	
gl	84	
Stat t	14,98916249	
P(T<=t) uni-caudal	9,23107E-26	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	1,84621E-25	
t crítico bi-caudal	1,988609667	

	<i>I</i> ₃ -Normalização Linear	<i>I</i> ₃ -Normalização pelo Valor Máximo do Elemento
Média	0,070085882	0,070062472
Variância	0,02047224	0,0204757
Observações	85	85
Correlação de Pearson	0,999999062	
Hipótese da diferença	0	
gl	84	
Stat t	1,099164703	
P(T<=t) uni-caudal	0,137418294	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	0,274836588	
t crítico bi-caudal	1,988609667	

	<i>I</i> ₄ -Normalização Linear	<i>I</i> ₄ -Normalização por Desvio Padrão
Média	0,165758824	0,73698695
Variância	0,045279263	0,450384167
Observações	85	85
Correlação de Pearson	0,688395909	
Hipótese da diferença	0	
gl	84	
Stat t	-9,630435698	
P(T<=t) uni-caudal	1,60016E-15	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	3,20032E-15	
t crítico bi-caudal	1,988609667	

	<i>I</i> ₄ -Normalização Linear	<i>I</i> ₄ -Normalização por Escala Decimal
Média	0,165758824	0,146188235
Variância	0,045279263	0,035460512
Observações	85	85
Correlação de Pearson	0,998355096	
Hipótese da diferença	0	
gl	84	
Stat t	6,673169808	
P(T<=t) uni-caudal	1,2588E-09	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	2,5176E-09	
t crítico bi-caudal	1,988609667	

	<i>I</i> ₃ -Normalização Linear	<i>I</i> ₃ -Normalização por Valor Máximo do Elemento
Média	0,448894737	0,448894737
Variância	0,148262482	0,148262482
Observações	85	85
Correlação de Pearson	1	
Hipótese da diferença	0	
gl	84	
Stat t	-3,00000001	
P(T<=t) uni-caudal	7,5944741E-07	
t crítico uni-caudal	1,5122034	
P(T<=t) bi-caudal	0,0000111	
t crítico bi-caudal	1,677777777	

	<i>I₃-Normalização Linear</i>	<i>I₃-Normalização por Escala Decimal</i>
Média	0,448894737	0,25587
Variância	0,148262482	0,048170481
Observações	85	85
Correlação de Pearson	1	
Hipótese da diferenç	0	
gl	84	
Stat t	10,74826443	
P(T<=t) uni-caudal	9,38349E-19	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	1,8767E-17	
t crítico bi-caudal	1,988609667	

	<i>I₈-Normalização Linear</i>	<i>I₈-Normalização por Valor Máximo do Elemento</i>
Média	0,118425882	0,118418628
Variância	0,024852262	0,02485097
Observações	85	85
Correlação de Pearson	0,999999958	
Hipótese da diferenç	0	
gl	84	
Stat t	1,450070063	
P(T<=t) uni-caudal	0,075381337	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	0,150762674	
t crítico bi-caudal	1,988609667	

	<i>I₇-Normalização Linear</i>	<i>I₇-Normalização por Desvio Padrão</i>
Média	0,448073452	0,819588878
Variância	0,15392249	0,320277333
Observações	85	85
Correlação de Pearson	-0,471164171	
Hipótese da diferenç	0	
gl	84	
Stat t	-4,143243987	
P(T<=t) uni-caudal	4,06334E-05	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	8,12667E-05	
t crítico bi-caudal	1,988609667	

	<i>I₈-Normalização Linear</i>	<i>I₈-Normalização por Desvio Padrão</i>
Média	0,118425882	0,663475124
Variância	0,024852262	0,554560293
Observações	85	85
Correlação de Pearson	0,727280415	
Hipótese da diferenç	0	
gl	84	
Stat t	-7,860836059	
P(T<=t) uni-caudal	5,75546E-12	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	1,15109E-11	
t crítico bi-caudal	1,988609667	

	<i>I₈-Normalização Linear</i>	<i>I₈-Normalização por Escala Decimal</i>
Média	0,118425882	0,060488235
Variância	0,024852262	0,006484032
Observações	85	85
Correlação de Pearson	0,999999958	
Hipótese da diferenç	0	
gl	84	
Stat t	6,926104984	
P(T<=t) uni-caudal	4,05909E-10	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	8,11818E-10	
t crítico bi-caudal	1,988609667	

	<i>I₇-Normalização Linear</i>	<i>I₇-Normalização por Escala Decimal</i>
Média	0,448073452	0,084685882
Variância	0,15392249	0,005498265
Observações	85	85
Correlação de Pearson	1	
Hipótese da diferenç	0	
gl	84	
Stat t	10,52949724	
P(T<=t) uni-caudal	2,55039E-17	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	5,10079E-17	
t crítico bi-caudal	1,988609667	

APÊNDICE 35 – TESTE *T* DE STUDENT, EXECUTADO PARA DUAS AMOSTRAS PAREADAS; VALORES DE I_5 E I_7 ALTERADOS.

	<i>I</i> ₅ -Normalização Linear	<i>I</i> ₅ -Normalização por Desvio Padrão
Média	0,244848935	6,98787E-17
Variação	0,027443159	1,011904762
Observações	85	85
Correlação de Pears	1	
Hipótese da diferenç	0	
gl	84	
Stat t	2,686496254	
P(T<=t) uni-caudal	0,004350748	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	0,008701496	
t crítico bi-caudal	1,988609667	

	<i>I</i> ₅ -Normalização Linear	<i>I</i> ₅ -Normalização por Escala Decimal
Média	0,244848935	0,169161176
Variação	0,027443159	0,005779233
Observações	85	85
Correlação de Pears	1	
Hipótese da diferenç	0	
gl	84	
Stat t	7,784675725	
P(T<=t) uni-caudal	8,16756E-12	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	1,63351E-11	
t crítico bi-caudal	1,988609667	

	<i>I</i> ₅ -Normalização Linear	<i>I</i> ₅ -Normalização por Valor Máximo do Elemento
Média	0,322114467	0,518584171
Variação	0,031737402	0,01600661
Observações	85	85
Correlação de Pears	1	
Hipótese da diferenç	0	
gl	84	
Stat t	-35,08166075	
P(T<=t) uni-caudal	3,02109E-52	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	6,04217E-52	
t crítico bi-caudal	1,988609667	

	<i>I</i> ₅ -Normalização Linear	<i>I</i> ₅ -Normalização por Desvio Padrão
Média	0,070062472	7,02053E-18
Variação	0,0204757	1,011904762
Observações	85	85
Correlação de Pears	1	
Hipótese da diferenç	0	
gl	84	
Stat t	0,748624308	
P(T<=t) uni-caudal	0,22808739	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	0,45617478	
t crítico bi-caudal	1,988609667	

	<i>I</i> ₅ -Normalização Linear	<i>I</i> ₅ -Normalização por Escala Decimal
Média	0,16444121	-8,88178E-17
Variação	0,044868492	1,011904762
Observações	85	85
Correlação de Pears	1	
Hipótese da diferenç	0	
gl	84	
Stat t	1,909140358	
P(T<=t) uni-caudal	0,029829163	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	0,059658326	
t crítico bi-caudal	1,988609667	

	<i>I</i> ₅ -Normalização Linear	<i>I</i> ₅ -Normalização por Desvio Padrão
Média	0,448642867	-1,71105E-16
Variação	0,148096152	1,011904762
Observações	85	85
Correlação de Pears	1	
Hipótese da diferenç	0	
gl	84	
Stat t	6,659585574	
P(T<=t) uni-caudal	1,33728E-09	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	2,67456E-09	
t crítico bi-caudal	1,988609667	

	<i>I</i> ₅ -Normalização Linear	<i>I</i> ₅ -Normalização por Escala Decimal
Média	0,118418628	5,73071E-17
Variação	0,02485097	1,011904762
Observações	85	85
Correlação de Pears	1	
Hipótese da diferenç	0	
gl	84	
Stat t	1,287015193	
P(T<=t) uni-caudal	0,100812135	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	0,20162427	
t crítico bi-caudal	1,988609667	

	<i>I</i> ₅ -Normalização Linear	<i>I</i> ₅ -Normalização por Valor Máximo do Elemento
Média	0,244848935	0,328022448
Variação	0,027443159	0,021730812
Observações	85	85
Correlação de Pears	1	
Hipótese da diferenç	0	
gl	84	
Stat t	-42,0267932	
P(T<=t) uni-caudal	1,75132E-58	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	3,50265E-58	
t crítico bi-caudal	1,988609667	

	<i>I</i> ₅ -Normalização Linear	<i>I</i> ₅ -Normalização por Desvio Padrão
Média	0,322114467	-3,33067E-17
Variação	0,031737402	1,011904762
Observações	85	85
Correlação de Pears	1	
Hipótese da diferenç	0	
gl	84	
Stat t	3,587585263	
P(T<=t) uni-caudal	0,000279932	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	0,000559864	
t crítico bi-caudal	1,988609667	

	<i>I</i> ₅ -Normalização Linear	<i>I</i> ₅ -Normalização por Escala Decimal
Média	0,322114467	0,054036471
Variação	0,031737402	0,000173794
Observações	85	85
Correlação de Pears	1	
Hipótese da diferenç	0	
gl	84	
Stat t	14,98214383	
P(T<=t) uni-caudal	9,50037E-26	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	1,90007E-25	
t crítico bi-caudal	1,988609667	

	<i>I</i> ₅ -Normalização Linear	<i>I</i> ₅ -Normalização por Escala Decimal
Média	0,070062472	0,067883529
Variação	0,0204757	0,019221916
Observações	85	85
Correlação de Pears	1	
Hipótese da diferenç	0	
gl	84	
Stat t	4,514145368	
P(T<=t) uni-caudal	1,02715E-05	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	2,05431E-05	
t crítico bi-caudal	1,988609667	

	<i>I</i> ₅ -Normalização Linear	<i>I</i> ₅ -Normalização por Escala Decimal
Média	0,16444121	0,146188235
Variação	0,044868492	0,035460512
Observações	85	85
Correlação de Pears	1	
Hipótese da diferenç	0	
gl	84	
Stat t	7,157302732	
P(T<=t) uni-caudal	1,43022E-10	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	2,86043E-10	
t crítico bi-caudal	1,988609667	

	<i>I</i> ₅ -Normalização Linear	<i>I</i> ₅ -Normalização por Escala Decimal
Média	0,448642867	0,25587
Variação	0,148096152	0,048170481
Observações	85	85
Correlação de Pears	1	
Hipótese da diferenç	0	
gl	84	
Stat t	10,74826443	
P(T<=t) uni-caudal	9,38349E-18	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	1,8767E-17	
t crítico bi-caudal	1,988609667	

	<i>I</i> ₅ -Normalização Linear	<i>I</i> ₅ -Normalização por Escala Decimal
Média	0,118418628	0,060488235
Variação	0,02485097	0,006484032
Observações	85	85
Correlação de Pears	1	
Hipótese da diferenç	0	
gl	84	
Stat t	6,9256066	
P(T<=t) uni-caudal	4,06819E-10	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	8,13638E-10	
t crítico bi-caudal	1,988609667	

	<i>I₃</i> -Normalização Linear	<i>I₃</i> -Normalização por Desvio Padrão
Média	0,447836501	-0,013704336
Variância	0,153759738	1,01596375
Observações	85	85
Correlação de Pearson	1	
Hipótese da diferença de m	0	
gl	84	
Stat t	6,909713516	
P(T<=t) uni-caudal	4,36937E-10	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	8,73873E-10	
t crítico bi-caudal	1,988609667	

	<i>I₃</i> -Normalização Linear	<i>I₃</i> -Normalização por Escala Decimal
Média	0,447836501	0,084685882
Variância	0,153759738	0,005498265
Observações	85	85
Correlação de Pearson	1	
Hipótese da diferença de m	0	
gl	84	
Stat t	10,52949724	
P(T<=t) uni-caudal	2,55039E-17	
t crítico uni-caudal	1,663196679	
P(T<=t) bi-caudal	5,10079E-17	
t crítico bi-caudal	1,988609667	

APÊNDICE 36 – ANÁLISE ESTATÍSTICA DE VARIÂNCIA (ANOVA), EXECUTADA PARA MULTIVARIÁVEIS; VALORES ORIGINAIS.

ANOVA Indicador I_1 - Vazão máxima anual						
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	16,5906	3	5,53021	44,2618	3,9509E-24	2,6314
Dentro dos grupos	41,9808	336	0,12494			
Total	58,5714	339				

ANOVA Indicador I_2 - Nível de água máximo anual						
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	23,7528	3	7,9176	72,0406	0,0000	2,6315
Dentro dos grupos	36,9279	336	0,1099			
Total	60,6807	339				

ANOVA Indicador I_3 - Volume total extravasado						
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	17,0205	3	5,6735	31,8455	3,8472E-18	2,6315
Dentro dos grupos	59,8607	336	0,1782			
Total	76,8812	339				

ANOVA Indicador I_4 - Tempo acima do limiar de extravasamento						
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	21,3322	3	7,1107	49,3808	1,82512E-26	2,6314
Dentro dos grupos	48,3833	336	0,1439			
Total	69,7156	339				

ANOVA Indicador I_5 - Tempo entre o início do evento e ultrapassagem do limiar de cheia						
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	13,5191	3	4,5064	26,32374	2,56537E-15	2,6314
Dentro dos grupos	57,5197	336	0,1712			
Total	71,0388	339				

ANOVA Indicador I_6 - Taxas média de ascensão do hidrograma						
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	20,4950	3	6,8317	44,7430	2,366E-24	2,6315
Dentro dos grupos	51,3028	336	0,1527			
Total	71,7978	339				

ANOVA Indicador I_7 - Taxas média de recessão dos hidrograma						
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	22,9549	3	7,6516	48,3042	5,577E-26	2,6315
Dentro dos grupos	53,2241	336	0,1584			
Total	76,1790	339				

APÊNDICE 37 – ANÁLISE ESTATÍSTICA DE VARIÂNCIA (ANOVA), EXECUTADA PARA MULTIVARIÁVEIS; VALORES DE I_5 E I_7 ALTERADOS.

ANOVA Indicador I_1 . Vazão máxima anual						
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	4,9735	3	1,6578	6,2158	0,000404	2,6314
Dentro dos grupos	89,6160	336	0,26671			
Total	94,5896	339				

ANOVA Indicador I_2 . Nível de água máximo anual						
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	14,9149	3	4,9716	18,7640	2,822E-11	2,6314
Dentro dos grupos	89,0250	336	0,2649			
Total	103,9400	339				

ANOVA Indicador I_3 . Volume total extravasado						
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	0,3067	3	0,1022	0,3814	0,766E-07	2,631489119
Dentro dos grupos	90,0545	336	0,2680			
Total	90,3613	339				

ANOVA Indicador I_4 . Tempo acima do limiar de extravasamento						
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	1,6175	3	0,5391	1,8966	0,12E-07	2,631489119
Dentro dos grupos	95,5165	336	0,2842			
Total	97,1341	339				

ANOVA Indicador I_5 . Tempo entre o início do evento e ultrapassagem do limiar de cheia						
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	11,5250	3	3,8416	11,3301	4,25E-07	2,6314
Dentro dos grupos	113,9264	336	0,3390			
Total	125,4514	339				

ANOVA Indicador I_6 . Taxas média de ascensão do hidrograma						
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	0,8163	3	0,2721	1,0190	0,38E-09	2,6314
Dentro dos grupos	89,7196	336	0,2670			
Total	90,5359	339				

ANOVA Indicador I_7 . Taxas média de recessão dos hidrograma						
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	14,8638	3	4,9546	14,9125	3,89E-09	2,6314
Dentro dos grupos	111,6344	336	0,3322			
Total	126,4983	339				

APÊNDICE 38 – ANÁLISE DE COMPONENTES PRINCIPAIS APLICADO AOS DADOS NÃO NORMALIZADOS.

Descriptive statistics

	I1 (m3/s)	I2 (cm)	I3 (km3)	I4 (dias)	I5 (dias)	I6 (m3/s/dia)	I7 (m3/s/dia)
Mean	1691.612	540.365	0.679	14.619	25.587	60.488	85.694
Std dev	760.213	131.831	1.38643	18.831	21.948	80.52348432	74.00726169
Skewness	1.5908	0.87133	4.14112	1.9202	-0.1418	2.756707881	-0.10486142
Kurtosis	4.62511	1.70285	22.1584	4.4457	-1.73	11.43019012	-1.72551674

Correlation Matrix

	I1 (m3/s)	I2 (cm)	I3 (km3)	I4 (dias)	I5 (dias)	I6 (m3/s/dia)	I7 (m3/s/dia)
I1 (m3/s)	1	0.98979	0.85754	0.801	0.6093	0.432516769	0.41158087
I2 (cm)	0.98979	1	0.80046	0.7999	0.6731	0.480229211	0.490519697
I3 (km3)	0.85754	0.80046	1	0.8373	0.3197	0.251829296	0.150954121
I4 (dias)	0.80104	0.7999	0.83727	1	0.5297	0.408440974	0.459791032
I5 (dias)	0.60934	0.67312	0.31969	0.5297	1	0.756124312	0.803904095
I6 (m3/s/dia)	0.43252	0.48023	0.25183	0.4084	0.7561	1	0.451293381
I7 (m3/s/dia)	0.41158	0.49052	0.15095	0.4598	0.8039	0.451293381	1

Inverse of Correlation Matrix

	I1 (m3/s)	I2 (cm)	I3 (km3)	I4 (dias)	I5 (dias)	I6 (m3/s/dia)	I7 (m3/s/dia)
I1 (m3/s)	186.881	-167.355	-39.739	15.48	-4.1955	3.823186381	5.703124702
I2 (cm)	-167.355	156.036	31.7039	-14.061	0.6165	-2.93119235	-5.15226354
I3 (km3)	-39.7395	31.7039	15.5618	-7.6092	1.4148	-0.35982705	0.979153079
I4 (dias)	15.4798	-14.0607	-7.6092	6.9174	0.6057	-0.51558043	-1.76024211
I5 (dias)	-4.19546	0.61647	1.41479	0.6057	8.6217	-3.6379424	-4.35691668
I6 (m3/s/dia)	3.82319	-2.93119	-0.3598	-0.5156	-3.6379	3.033799037	1.711057207
I7 (m3/s/dia)	5.70312	-5.15226	0.97915	-1.7602	-4.3569	1.711057207	4.57188044

Partial Correlation Matrix

	I1 (m3/s)	I2 (cm)	I3 (km3)	I4 (dias)	I5 (dias)	I6 (m3/s/dia)	I7 (m3/s/dia)
I1 (m3/s)	1	0.98004	0.7369	-0.4305	0.1045	-0.16056446	-0.19511146
I2 (cm)	0.98004	1	-0.6434	0.428	-0.0168	0.134721839	0.192902442
I3 (km3)	0.7369	-0.64338	1	0.7334	-0.1221	0.052368501	-0.11608414
I4 (dias)	-0.43054	0.42798	0.7334	1	-0.0784	0.112546339	0.313006479
I5 (dias)	0.10452	-0.01681	-0.1221	-0.0784	1	0.711320806	0.69396077
I6 (m3/s/dia)	-0.16056	0.13472	0.05237	0.1125	0.7113	1	-0.45943459
I7 (m3/s/dia)	-0.19511	0.1929	-0.1161	0.313	0.694	-0.45943459	1
	1.76364	1.61323	1.52597	1.0232	1.0198	0.776397492	0.879390116

Eigenvalues and eigenvectors

	4.5981	1.4039	0.5589	0.2843	0.0808	0.0710	0.0029
I1 (m3/s)	0.4325	-0.2413	-0.0006	0.4371	-0.1647	-0.0659	0.7296
I2 (cm)	0.4413	-0.1579	0.0326	0.4438	-0.3769	0.0160	-0.6633
I3 (km3)	0.3629	-0.4902	-0.0747	-0.1825	0.6603	-0.3608	-0.1515
I4 (dias)	0.4065	-0.2268	0.1217	-0.6989	-0.3359	0.4039	0.0635
I5 (dias)	0.3775	0.4539	0.0024	0.1772	0.5059	0.6033	-0.0111
I6 (m3/s/dia)	0.2992	0.4195	-0.7585	-0.1953	-0.1586	-0.3092	0.0140
I7 (m3/s/dia)	0.2987	0.4930	0.6349	-0.1439	-0.0302	-0.4924	0.0215

Full Load Matrix

	1	2	3	4	5	6	7
I1 (m3/s)	0.92733	-0.28586	-0.0005	0.233	-0.0468	-0.01756063	0.039056407
I2 (cm)	0.94638	-0.18714	0.02434	0.2366	-0.1071	0.004276771	-0.03550864
I3 (km3)	0.7781	-0.58087	-0.0559	-0.0973	0.1877	-0.09617497	-0.00811005
I4 (dias)	0.87174	-0.26872	0.09095	-0.3726	-0.0955	0.107661067	0.003399805
I5 (dias)	0.80948	0.53784	0.00177	0.0945	0.1438	0.160794501	-0.00059618
I6 (m3/s/dia)	0.64164	0.49702	-0.5671	-0.1041	-0.0451	-0.08241679	0.000747901
I7 (m3/s/dia)	0.64047	0.58419	0.47469	-0.0767	-0.0086	-0.13124683	0.001148906

Scree Plot

eValue	%	Cum %
4.59811	65.69%	65.69%
1.40394	20.06%	85.74%
0.55893	7.98%	93.73%
0.28429	4.06%	97.79%
0.08083	1.15%	98.94%
0.07104	1.01%	99.96%
0.00287	0.04%	100%

7

Factor Matrix (unrotated)

	1	2	Commun	Specific
I1 (m3/s)	0.92733	-0.28586	0.9417	0.0583
I2 (cm)	0.94638	-0.18714	0.9307	0.0693
I3 (km3)	0.7781	-0.58087	0.9429	0.0571
I4 (dias)	0.87174	-0.26872	0.8321	0.1679
I5 (dias)	0.80948	0.53784	0.9445	0.0555
I6 (m3/s/dia)	0.64164	0.49702	0.6587	0.3413
I7 (m3/s/dia)	0.64047	0.58419	0.7515	0.2485
	4.59811	1.40394	6.0021	0.9979

Factor Matrix (rotated Varimax)

	1	2	Commun	Specific
I1 (m3/s)	0.90196	0.35796	0.9417	0.0583
I2 (cm)	0.85498	0.44685	0.9307	0.0693
I3 (km3)	0.97039	0.03451	0.9429	0.0571
I4 (dias)	0.84788	0.3365	0.8321	0.1679
I5 (dias)	0.29418	0.92627	0.9445	0.0555
I6 (m3/s/dia)	0.18891	0.78933	0.6587	0.3413
I7 (m3/s/dia)	0.1334	0.85655	0.7515	0.2485
	3.34511	2.65694	6.0021	0.9979

**Reproduced Correlation
Matrix**

	I1 (m3/s)	I2 (cm)	I3 (km3)	I4 (dias)	I5 (dias)	I6 (m3/s/dia)	I7 (m3/s/dia)
I1 (m3/s)	0.94166	0.93111	0.88761	0.8852	0.5969	0.452931875	0.426930048
I2 (cm)	0.93111	0.93066	0.84509	0.8753	0.6654	0.514219276	0.496800908
I3 (km3)	0.88761	0.84509	0.94286	0.8344	0.3174	0.210553698	0.159008855
I4 (dias)	0.88521	0.87529	0.8344	0.8321	0.5611	0.425779647	0.401336777
I5 (dias)	0.59691	0.66542	0.31744	0.5611	0.9445	0.786706673	0.832647119
I6 (m3/s/dia)	0.45293	0.51422	0.21055	0.4258	0.7867	0.658723456	0.701301211
I7 (m3/s/dia)	0.42693	0.4968	0.15901	0.4013	0.8326	0.701301211	0.751479504

**Error
Matrix**

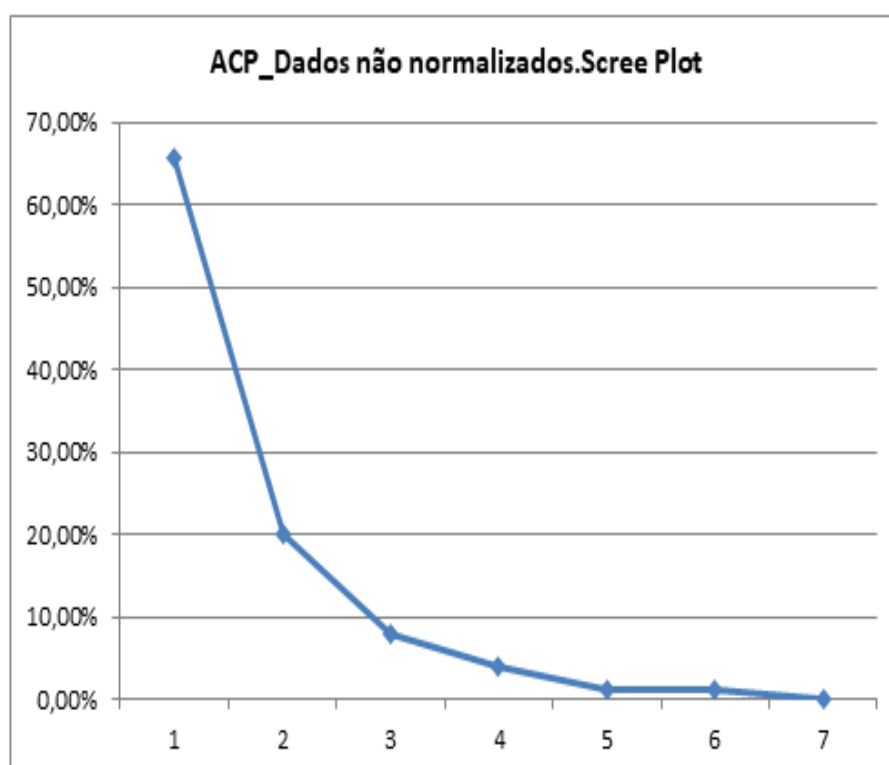
	I1 (m3/s)	I2 (cm)	I3 (km3)	I4 (dias)	I5 (dias)	I6 (m3/s/dia)	I7 (m3/s/dia)
I1 (m3/s)	0.05834	0.05869	-0.0301	-0.0842	0.0124	-0.02041511	-0.01534918
I2 (cm)	0.05869	0.06934	-0.0446	-0.0754	0.0077	-0.03399006	-0.00628121
I3 (km3)	-0.03007	-0.04462	0.05714	0.0029	0.0022	0.041275598	-0.00805473
I4 (dias)	-0.08417	-0.07539	0.00287	0.1679	-0.0315	-0.01733867	0.058454255
I5 (dias)	0.01243	0.00769	0.00225	-0.0315	0.0555	-0.03058236	-0.02874302
I6 (m3/s/dia)	-0.02042	-0.03399	0.04128	-0.0173	-0.0306	0.341276544	-0.25000783
I7 (m3/s/dia)	-0.01535	-0.00628	-0.0081	0.0585	-0.0287	-0.25000783	0.248520496

Factor Scores Matrix - Regression Method

	1	2
I1 (m3/s)	0.28475	0.03242
I2 (cm)	0.24394	0.02499
I3 (km3)	0.39106	0.21655
I4 (dias)	0.26768	0.03047
I5 (dias)	-0.1027	0.40891
I6 (m3/s/dia)	-0.11294	0.36338
I7 (m3/s/dia)	-0.15203	0.41162

I4 (dias)	I5 (dias)	I6 (m3/s/dia)	I7 (m3/s/dia)
0.8852	0.5969	0.452931875	0.426930048
0.8753	0.6654	0.514219276	0.496800908
0.8344	0.3174	0.210553698	0.159008855
0.8321	0.5611	0.425779647	0.401336777
0.5611	0.9445	0.786706673	0.832647119
0.4258	0.7867	0.658723456	0.701301211
0.4013	0.8326	0.701301211	0.751479504

I4 (dias)	I5 (dias)	I6 (m3/s/dia)	I7 (m3/s/dia)
-0.0842	0.0124	-0.02041511	-0.01534918
-0.0754	0.0077	-0.03399006	-0.00628121
0.0029	0.0022	0.041275598	-0.00805473
0.1679	-0.0315	-0.01733867	0.058454255
-0.0315	0.0555	-0.03058236	-0.02874302
-0.0173	-0.0306	0.341276544	-0.25000783
0.0585	-0.0287	-0.25000783	0.248520496



APÊNDICE 39 – ANÁLISE DE COMPONENTES PRINCIPAIS APLICADO AOS DADOS NORMALIZADOS PELO MÉTODO DA NORMALIZAÇÃO LINEAR.

Descriptive statistics

	I1 (m3/s)	I2 (cm)	I3 (km3)	I4 (dias)	I5 (dias)	I6 (m3/s/dia)	I7 (m3/s/dia)
Mean	0.24485	0.32211	0.07006	0.16444	0.44864	0.11842	0.44784
Std dev	0.16566	0.17815	0.14309	0.21182	0.38483	0.15764	0.39212
Skewness	1.59080	0.87133	4.14112	1.92021	-0.14179	2.75671	-0.08003
Kurtosis	4.62511	1.70285	22.15843	4.44572	-1.72998	11.43019	-1.73530

Correlation Matrix

	I1 (m3/s)	I2 (cm)	I3 (km3)	I4 (dias)	I5 (dias)	I6 (m3/s/dia)	I7 (m3/s/dia)
I1 (m3/s)	1.00000	0.99032	0.86369	0.80703	0.61326	0.42746	0.36912
I2 (cm)	0.99032	1.00000	0.80927	0.80590	0.67493	0.47373	0.44490
I3 (km3)	0.86369	0.80927	1.00000	0.84221	0.33178	0.25231	0.12165
I4 (dias)	0.80703	0.80590	0.84221	1.00000	0.53637	0.40730	0.42891
I5 (dias)	0.61326	0.67493	0.33178	0.53637	1.00000	0.75484	0.78089
I6 (m3/s/dia)	0.42746	0.47373	0.25231	0.40730	0.75484	1.00000	0.44400
I7 (m3/s/dia)	0.36912	0.44490	0.12165	0.42891	0.78089	0.44400	1.00000
	3.0730606	3.1630175	2.298725	2.647636	2.4089486	1.403613564	1.339871276

Inverse of Correlation Matrix

	I1 (m3/s)	I2 (cm)	I3 (km3)	I4 (dias)	I5 (dias)	I6 (m3/s/dia)	I7 (m3/s/dia)
I1 (m3/s)	192.50604	171.83606	41.46686	16.47519	-2.66020	3.06095	4.08758
I2 (cm)	171.83606	159.80300	33.05425	15.01727	-1.09405	-2.11467	-3.45426
I3 (km3)	-41.46686	33.05425	16.20371	-7.89073	1.38380	-0.33233	1.08079
I4 (dias)	16.47519	-15.01727	-7.89073	7.08751	0.48820	-0.45606	-1.65896
I5 (dias)	-2.66020	-1.09405	1.38380	0.48820	8.09964	-3.34126	-3.75048
I6 (m3/s/dia)	3.06095	-2.11467	-0.33233	-0.45606	-3.34126	2.87188	1.38104
I7 (m3/s/dia)	4.08758	-3.45426	1.08079	-1.65896	-3.75048	1.38104	3.92359

Partial Correlation Matrix

	I1 (m3/s)	I2 (cm)	I3 (km3)	I4 (dias)	I5 (dias)	I6 (m3/s/dia)	I7 (m3/s/dia)
I1 (m3/s)	1.00000	0.97971	0.74246	-0.44603	0.06737	-0.13018	-0.14873
I2 (cm)	0.97971	1.00000	-0.64957	0.44622	0.03041	0.09871	0.13795
I3 (km3)	0.74246	-0.64957	1.00000	0.73631	-0.12079	0.04872	-0.13555
I4 (dias)	-0.44603	0.44622	0.73631	1.00000	-0.06443	0.10109	0.31459
I5 (dias)	0.06737	0.03041	-0.12079	-0.06443	1.00000	0.69278	0.66529
I6 (m3/s/dia)	-0.13018	0.09871	0.04872	0.10109	0.69278	1.00000	-0.41142
I7 (m3/s/dia)	-0.14873	0.13795	-0.13555	0.31459	0.66529	-0.41142	1.00000

Eigenvalues and eigenvectors

	4.57043	1.4280	0.5621	0.2803	0.0790	0.0773	0.0028
I1 (m3/s)	0.4344	-0.2400	0.0011	0.4261	-0.1887	0.0569	0.7303
I2 (cm)	0.4427	-0.1608	0.0289	0.4399	-0.2616	0.2776	-0.6620
I3 (km3)	0.3680	-0.4765	-0.0519	-0.1987	0.1881	-0.7324	-0.1538
I4 (dias)	0.4097	-0.2171	0.1372	-0.6874	0.0832	0.5314	0.0659
I5 (dias)	0.3788	0.4452	-0.0135	0.2120	0.7817	0.0458	-0.0042
I6 (m3/s/dia)	0.2986	0.4182	-0.7545	-0.2184	-0.3310	-0.0966	0.0104
I7 (m3/s/dia)	0.2818	0.5189	0.6389	-0.1425	-0.3648	-0.2993	0.0142

Full Load Matrix

	1	2	3	4	5	6	7
I1 (m3/s)	0.92860	-0.28681	0.00082	0.22558	-0.05304	0.01581	0.03856
I2 (cm)	0.94643	-0.19219	0.02163	0.23293	-0.07353	0.07717	-0.03496
I3 (km3)	0.78666	-0.56943	-0.03889	-0.10519	0.05287	-0.20361	-0.00812
I4 (dias)	0.87593	-0.25947	0.10287	-0.36395	0.02338	0.14774	0.00348
I5 (dias)	0.80982	0.53203	-0.01014	0.11225	0.21973	0.01272	-0.00022
I6 (m3/s/dia)	0.63839	0.49973	-0.56567	-0.11564	-0.09304	-0.02684	0.00055
I7 (m3/s/dia)	0.60246	0.62007	0.47901	-0.07546	-0.10254	-0.08321	0.00075

Scree Plot

	eValue	%	Cum %
I1 (m3/s)	4.57043	0.65292	0.65292
I2 (cm)	1.42804	0.20401	0.85692
I3 (km3)	0.56210	0.08030	0.93722
I4 (dias)	0.28034	0.04005	0.97727
I5 (dias)	0.07901	0.01129	0.98856
I6 (m3/s/dia)	0.07730	0.01104	0.99960
I7 (m3/s/dia)	0.00279	0.00040	1.00000

Factor Matrix (unrotated)

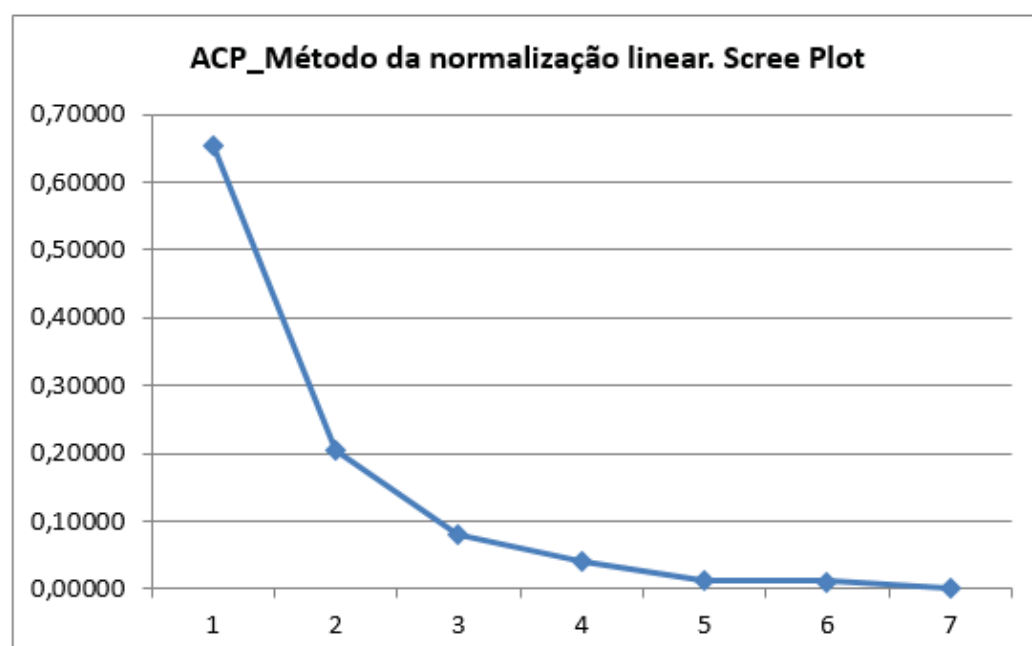
	1	2	Commun	Specific
I1 (m3/s)	0.92860	-0.28681	0.94456	0.05544
I2 (cm)	0.94643	-0.19219	0.93267	0.06733
I3 (km3)	0.78666	-0.56943	0.94307	0.05693
I4 (dias)	0.87593	-0.25947	0.83458	0.16542
I5 (dias)	0.80982	0.53203	0.93886	0.06114
I6 (m3/s/dia)	0.63839	0.49973	0.65727	0.34273
I7 (m3/s/dia)	0.60246	0.62007	0.74744	0.25256
	4.57043	1.42804	5.99847	1.00153

Reproduced Correlation Matrix

	I1 (m3/s)	I2 (cm)	I3 (km3)	I4 (dias)	I5 (dias)	I6 (m3/s/dia)	I7 (m3/s/dia)
I1 (m3/s)	0.94456	0.93398	0.89381	0.88781	0.59941	0.44948	0.38161
I2 (cm)	0.93398	0.93267	0.85396	0.87888	0.66419	0.50815	0.45102
I3 (km3)	0.89381	0.85396	0.94307	0.83680	0.33410	0.21763	0.12085
I4 (dias)	0.88781	0.87888	0.83680	0.83458	0.57130	0.42952	0.36683
I5 (dias)	0.59941	0.66419	0.33410	0.57130	0.93886	0.78285	0.81778
I6 (m3/s/dia)	0.44948	0.50815	0.21763	0.42952	0.78285	0.65727	0.69447
I7 (m3/s/dia)	0.38161	0.45102	0.12085	0.36683	0.81778	0.69447	0.74744

Error Matrix

	I1 (m3/s)	I2 (cm)	I3 (km3)	I4 (dias)	I5 (dias)	I6 (m3/s/dia)	I7 (m3/s/dia)
I1 (m3/s)	0.05544	0.05634	-0.03011	-0.08078	0.01385	-0.02202	-0.01248
I2 (cm)	0.05634	0.06733	-0.04469	-0.07298	0.01074	-0.03441	-0.00612
I3 (km3)	-0.03011	-0.04469	0.05693	0.00540	-0.00232	0.03468	0.00080
I4 (dias)	-0.08078	-0.07298	0.00540	0.16542	-0.03493	-0.02222	0.06208
I5 (dias)	0.01385	0.01074	-0.00232	-0.03493	0.06114	-0.02802	-0.03689
I6 (m3/s/dia)	-0.02202	-0.03441	0.03468	-0.02222	-0.02802	0.34273	-0.25048
I7 (m3/s/dia)	-0.01248	-0.00612	0.00080	0.06208	-0.03689	-0.25048	0.25256



APÊNDICE 40 – ANÁLISE DE COMPONENTES PRINCIPAIS APLICADO AOS DADOS NORMALIZADOS PELO MÉTODO DO DESVIO PADRÃO.

Descriptive statistics

	I1 (m3/s)	I2 (cm)	I3 (km3)	I4 (dias)	I5 (dias)	I6 (m3/s/dia)	I7 (m3/s/dia)
Mean	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	-0.01370
Std dev	1.00593	1.00593	1.00593	1.00593	1.00593	1.00593	1.00795
Skewness	1.59080	0.87133	4.14112	1.92021	-0.14179	2.75671	-0.08003
Kurtosis	4.62511	1.70285	22.15843	4.44572	-1.72998	11.43019	-1.73530

Correlation Matrix

	I1 (m3/s)	I2 (cm)	I3 (km3)	I4 (dias)	I5 (dias)	I6 (m3/s/dia)	I7 (m3/s/dia)
I1 (m3/s)	1.00000	0.99032	0.86369	0.80703	0.61326	0.42746	0.36912
I2 (cm)	0.99032	1.00000	0.80927	0.80590	0.67493	0.47373	0.44490
I3 (km3)	0.86369	0.80927	1.00000	0.84221	0.33178	0.25231	0.12165
I4 (dias)	0.80703	0.80590	0.84221	1.00000	0.53637	0.40730	0.42891
I5 (dias)	0.61326	0.67493	0.33178	0.53637	1.00000	0.75484	0.78089
I6 (m3/s/dia)	0.42746	0.47373	0.25231	0.40730	0.75484	1.00000	0.44400
I7 (m3/s/dia)	0.36912	0.44490	0.12165	0.42891	0.78089	0.44400	1.00000
	3.07306056	3.16301753	2.298725	2.647636	2.4089486	1.403613564	1.339871

Inverse of Correlation Matrix

	I1 (m3/s)	I2 (cm)	I3 (km3)	I4 (dias)	I5 (dias)	I6 (m3/s/dia)	I7 (m3/s/dia)
I1 (m3/s)	192.50604	-171.83606	41.46686	16.47519	-2.66020	3.06095	4.08758
I2 (cm)	-171.83606	159.80300	33.05425	15.01727	-1.09405	-2.11467	-3.45426
I3 (km3)	-41.46686	33.05425	16.20371	-7.89073	1.38380	-0.33233	1.08079
I4 (dias)	16.47519	-15.01727	-7.89073	7.08751	0.48820	-0.45606	-1.65896
I5 (dias)	-2.66020	-1.09405	1.38380	0.48820	8.09964	-3.34126	-3.75048
I6 (m3/s/dia)	3.06095	-2.11467	-0.33233	-0.45606	-3.34126	2.87188	1.38104
I7 (m3/s/dia)	4.08758	-3.45426	1.08079	-1.65896	-3.75048	1.38104	3.92359

Eigenvalues and eigenvectors

	4.5704	1.4280	0.5621	0.2803	0.0790	0.0773	0.0028
I1 (m3/s)	0.4344	-0.2400	0.0011	0.4261	-0.1887	0.0569	0.7303
I2 (cm)	0.4427	-0.1608	0.0289	0.4399	-0.2616	0.2776	-0.6620
I3 (km3)	0.3680	-0.4765	-0.0519	-0.1987	0.1881	-0.7324	-0.1538
I4 (dias)	0.4097	-0.2171	0.1372	-0.6874	0.0832	0.5314	0.0659
I5 (dias)	0.3788	0.4452	-0.0135	0.2120	0.7817	0.0458	-0.0042
I6 (m3/s/dia)	0.2986	0.4182	-0.7545	-0.2184	-0.3310	-0.0966	0.0104
I7 (m3/s/dia)	0.2818	0.5189	0.6389	-0.1425	-0.3648	-0.2993	0.0142

Full Load Matrix

	1	2	3	4	5	6	7
I1 (m3/s)	0.92860	-0.28681	0.00082	0.22558	-0.05304	0.01581	0.03856
I2 (cm)	0.94643	-0.19219	0.02163	0.23293	-0.07353	0.07717	-0.03496
I3 (km3)	0.78666	-0.56943	-0.03889	-0.10519	0.05287	-0.20361	-0.00812
I4 (dias)	0.87593	-0.25947	0.10287	-0.36395	0.02338	0.14774	0.00348
I5 (dias)	0.80982	0.53203	-0.01014	0.11225	0.21973	0.01272	-0.00022
I6 (m3/s/dia)	0.63839	0.49973	-0.56567	-0.11564	-0.09304	-0.02684	0.00055
I7 (m3/s/dia)	0.60246	0.62007	0.47901	-0.07546	-0.10254	-0.08321	0.00075

Scree Plot

	eValue	%	Cum %
I1 (m3/s)	4.57043	0.65292	0.65292
I2 (cm)	1.42804	0.20401	0.85692
I3 (km3)	0.56210	0.08030	0.93722
I4 (dias)	0.28034	0.04005	0.97727
I5 (dias)	0.07901	0.01129	0.98856
I6 (m3/s/dia)	0.07730	0.01104	0.99960
I7 (m3/s/dia)	0.00279	0.00040	1.00000

Factor Matrix (unrotated)

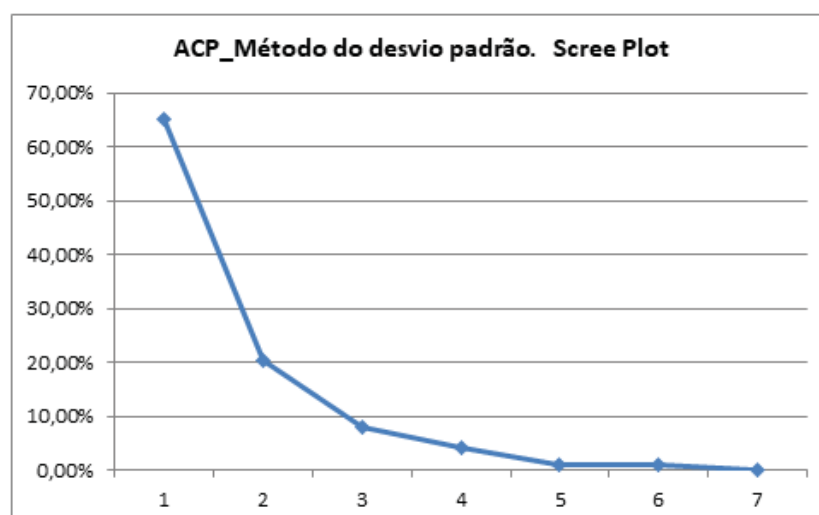
	1	2	Commun	Specific
I1 (m3/s)	0.92860	-0.28681	0.94456	0.05544
I2 (cm)	0.94643	-0.19219	0.93267	0.06733
I3 (km3)	0.78666	-0.56943	0.94307	0.05693
I4 (dias)	0.87593	-0.25947	0.83458	0.16542
I5 (dias)	0.80982	0.53203	0.93886	0.06114
I6 (m3/s/dia)	0.63839	0.49973	0.65727	0.34273
I7 (m3/s/dia)	0.60246	0.62007	0.74744	0.25256
	4.57043	1.42804	5.99847	1.00153

Reproduced Correlation Matrix

	I1 (m3/s)	I2 (cm)	I3 (km3)	I4 (dias)	I5 (dias)	I6 (m3/s/dia)	I7 (m3/s/dia)
I1 (m3/s)	0.94456	0.93398	0.89381	0.88781	0.59941	0.44948	0.38161
I2 (cm)	0.93398	0.93267	0.85396	0.87888	0.66419	0.50815	0.45102
I3 (km3)	0.89381	0.85396	0.94307	0.83680	0.33410	0.21763	0.12085
I4 (dias)	0.88781	0.87888	0.83680	0.83458	0.57130	0.42952	0.36683
I5 (dias)	0.59941	0.66419	0.33410	0.57130	0.93886	0.78285	0.81778
I6 (m3/s/dia)	0.44948	0.50815	0.21763	0.42952	0.78285	0.65727	0.69447
I7 (m3/s/dia)	0.38161	0.45102	0.12085	0.36683	0.81778	0.69447	0.74744

Error Matrix

	I1 (m3/s)	I2 (cm)	I3 (km3)	I4 (dias)	I5 (dias)	I6 (m3/s/dia)	I7 (m3/s/dia)
I1 (m3/s)	0.05544	0.05634	-0.03011	-0.08078	0.01385	-0.02202	-0.01248
I2 (cm)	0.05634	0.06733	-0.04469	-0.07298	0.01074	-0.03441	-0.00612
I3 (km3)	-0.03011	-0.04469	0.05693	0.00540	-0.00232	0.03468	0.00080
I4 (dias)	-0.08078	-0.07298	0.00540	0.16542	-0.03493	-0.02222	0.06208
I5 (dias)	0.01385	0.01074	-0.00232	-0.03493	0.06114	-0.02802	-0.03689
I6 (m3/s/dia)	-0.02202	-0.03441	0.03468	-0.02222	-0.02802	0.34273	-0.25048
I7 (m3/s/dia)	-0.01248	-0.00612	0.00080	0.06208	-0.03689	-0.25048	0.25256



APÊNDICE 41 – ANÁLISE DE COMPONENTES PRINCIPAIS APLICADO AOS DADOS NORMALIZADOS PELO MÉTODO DO VALOR MÁXIMO DOS ELEMENTOS.

Descriptive statistics

	I1 (m3/s)	I2 (cm)	I3 (km3)	I4 (dias)	I5 (dias)	I6 (m3/s/dia)	I7 (m3/s/dia)
Mean	0.32802	0.51858	0.07006	0.16444	0.44864	0.11842	0.44784
Std dev	0.14741	0.12652	0.14309	0.21182	0.38483	0.15764	0.39212
Skewness	1.59080	0.87133	4.14112	1.92021	-0.14179	2.75671	-0.08003
Kurtosis	4.62511	1.70285	22.15843	4.44572	-1.72998	11.43019	-1.73530

Correlation Matrix

	I1 (m3/s)	I2 (cm)	I3 (km3)	I4 (dias)	I5 (dias)	I6 (m3/s/dia)	I7 (m3/s/dia)
I1 (m3/s)	1.00000	0.99032	0.86369	0.80703	0.61326	0.42746	0.36912
I2 (cm)	0.99032	1.00000	0.80927	0.80590	0.67493	0.47373	0.44490
I3 (km3)	0.86369	0.80927	1.00000	0.84221	0.33178	0.25231	0.12165
I4 (dias)	0.80703	0.80590	0.84221	1.00000	0.53637	0.40730	0.42891
I5 (dias)	0.61326	0.67493	0.33178	0.53637	1.00000	0.75484	0.78089
I6 (m3/s/dia)	0.42746	0.47373	0.25231	0.40730	0.75484	1.00000	0.44400
I7 (m3/s/dia)	0.36912	0.44490	0.12165	0.42891	0.78089	0.44400	1.00000
	3.0730606	3.1630175	2.2987249	2.647636	2.40895	1.403614	1.339871

Inverse of Correlation Matrix

	I1 (m3/s)	I2 (cm)	I3 (km3)	I4 (dias)	I5 (dias)	I6 (m3/s/dia)	I7 (m3/s/dia)
I1 (m3/s)	192.50604	-171.83606	-41.46686	16.47519	-2.66020	3.06095	4.08758
I2 (cm)	-171.83606	159.80300	33.05425	-15.01727	-1.09405	-2.11467	-3.45426
I3 (km3)	-41.46686	33.05425	16.20371	-7.89073	1.38380	-0.33233	1.08079
I4 (dias)	16.47519	-15.01727	-7.89073	7.08751	0.48820	-0.45606	-1.65896
I5 (dias)	-2.66020	-1.09405	1.38380	0.48820	8.09964	-3.34126	-3.75048
I6 (m3/s/dia)	3.06095	-2.11467	-0.33233	-0.45606	-3.34126	2.87188	1.38104
I7 (m3/s/dia)	4.08758	-3.45426	1.08079	-1.65896	-3.75048	1.38104	3.92359

Partial Correlation Matrix

	I1 (m3/s)	I2 (cm)	I3 (km3)	I4 (dias)	I5 (dias)	I6 (m3/s/dia)	I7 (m3/s/dia)
I1 (m3/s)	1.00000	0.97971	0.74246	-0.44603	0.06737	-0.13018	-0.14873
I2 (cm)	0.97971	1.00000	-0.64957	0.44622	0.03041	0.09871	0.13795
I3 (km3)	0.74246	-0.64957	1.00000	0.73631	-0.12079	0.04872	-0.13555
I4 (dias)	-0.44603	0.44622	0.73631	1.00000	-0.06443	0.10109	0.31459
I5 (dias)	0.06737	0.03041	-0.12079	-0.06443	1.00000	0.69278	0.66529
I6 (m3/s/dia)	-0.13018	0.09871	0.04872	0.10109	0.69278	1.00000	-0.41142
I7 (m3/s/dia)	-0.14873	0.13795	-0.13555	0.31459	0.66529	-0.41142	1.00000
	1.7536317	1.6105991	1.550683	1.053551	0.94676	0.688489	0.770369

Eigenvalues and eigenvectors

	4.5704	1.4280	0.5621	0.2803	0.0790	0.0773	0.0028
I1 (m3/s)	0.4344	-0.2400	0.0011	0.4261	-0.1887	0.0569	0.7303
I2 (cm)	0.4427	-0.1608	0.0289	0.4399	-0.2616	0.2776	-0.6620
I3 (km3)	0.3680	-0.4765	-0.0519	-0.1987	0.1881	-0.7324	-0.1538
I4 (dias)	0.4097	-0.2171	0.1372	-0.6874	0.0832	0.5314	0.0659
I5 (dias)	0.3788	0.4452	-0.0135	0.2120	0.7817	0.0458	-0.0042
I6 (m3/s/dia)	0.2986	0.4182	-0.7545	-0.2184	-0.3310	-0.0966	0.0104
I7 (m3/s/dia)	0.2818	0.5189	0.6389	-0.1425	-0.3648	-0.2993	0.0142

Full Load Matrix

	1	2	3	4	5	6	7
I1 (m3/s)	0.9286	-0.2868	0.0008	0.2256	-0.0530	0.0158	0.0386
I2 (cm)	0.9464	-0.1922	0.0216	0.2329	-0.0735	0.0772	-0.0350
I3 (km3)	0.7867	-0.5694	-0.0389	-0.1052	0.0529	-0.2036	-0.0081
I4 (dias)	0.8759	-0.2595	0.1029	-0.3640	0.0234	0.1477	0.0035
I5 (dias)	0.8098	0.5320	-0.0101	0.1122	0.2197	0.0127	-0.0002
I6 (m3/s/dia)	0.6384	0.4997	-0.5657	-0.1156	-0.0930	-0.0268	0.0005
I7 (m3/s/dia)	0.6025	0.6201	0.4790	-0.0755	-0.1025	-0.0832	0.0007

Scree Plot

	eValue	%	Cum %
I1 (m3/s)	4.5704	0.6529	0.6529
I2 (cm)	1.4280	0.2040	0.8569
I3 (km3)	0.5621	0.0803	0.9372
I4 (dias)	0.2803	0.0400	0.9773
I5 (dias)	0.0790	0.0113	0.9886
I6 (m3/s/dia)	0.0773	0.0110	0.9996
I7 (m3/s/dia)	0.0028	0.0004	1.0000

7

Factor Matrix (unrotated)

	1	2	Commun	Specific
I1 (m3/s)	0.9286	-0.2868	0.9446	0.0554
I2 (cm)	0.9464	-0.1922	0.9327	0.0673
I3 (km3)	0.7867	-0.5694	0.9431	0.0569
I4 (dias)	0.8759	-0.2595	0.8346	0.1654
I5 (dias)	0.8098	0.5320	0.9389	0.0611
I6 (m3/s/dia)	0.6384	0.4997	0.6573	0.3427
I7 (m3/s/dia)	0.6025	0.6201	0.7474	0.2526
	4.5704	1.4280	5.9985	1.0015

Reproduced Correlation Matrix

	I1 (m3/s)	I2 (cm)	I3 (km3)	I4 (dias)	I5 (dias)	I6 (m3/s/dia)	I7 (m3/s/dia)
I1 (m3/s)	0.9446	0.9340	0.8938	0.8878	0.5994	0.4495	0.3816
I2 (cm)	0.9340	0.9327	0.8540	0.8789	0.6642	0.5081	0.4510
I3 (km3)	0.8938	0.8540	0.9431	0.8368	0.3341	0.2176	0.1208
I4 (dias)	0.8878	0.8789	0.8368	0.8346	0.5713	0.4295	0.3668
I5 (dias)	0.5994	0.6642	0.3341	0.5713	0.9389	0.7829	0.8178
I6 (m3/s/dia)	0.4495	0.5081	0.2176	0.4295	0.7829	0.6573	0.6945
I7 (m3/s/dia)	0.3816	0.4510	0.1208	0.3668	0.8178	0.6945	0.7474

Error Matrix

	I1 (m3/s)	I2 (cm)	I3 (km3)	I4 (dias)	I5 (dias)	I6 (m3/s/dia)	I7 (m3/s/dia)
I1 (m3/s)	0.0554	0.0563	-0.0301	-0.0808	0.0138	-0.0220	-0.0125
I2 (cm)	0.0563	0.0673	-0.0447	-0.0730	0.0107	-0.0344	-0.0061
I3 (km3)	-0.0301	-0.0447	0.0569	0.0054	-0.0023	0.0347	0.0008
I4 (dias)	-0.0808	-0.0730	0.0054	0.1654	-0.0349	-0.0222	0.0621
I5 (dias)	0.0138	0.0107	-0.0023	-0.0349	0.0611	-0.0280	-0.0369
I6 (m3/s/dia)	-0.0220	-0.0344	0.0347	-0.0222	-0.0280	0.3427	-0.2505
I7 (m3/s/dia)	-0.0125	-0.0061	0.0008	0.0621	-0.0369	-0.2505	0.2526

