

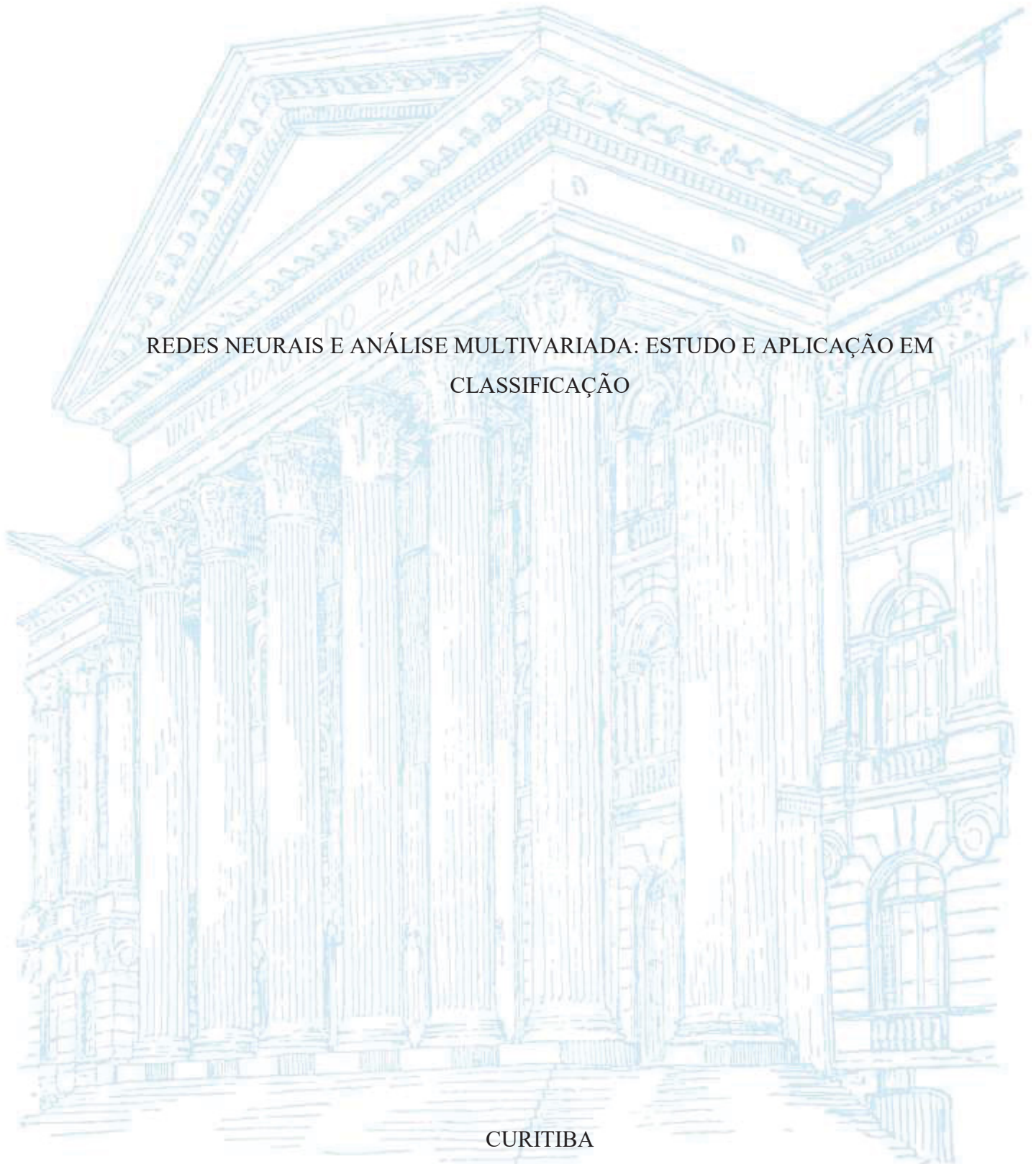
UNIVERSIDADE FEDERAL DO PARANÁ

DANIEL STEFFEN

REDES NEURAIS E ANÁLISE MULTIVARIADA: ESTUDO E APLICAÇÃO EM
CLASSIFICAÇÃO

CURITIBA

2021



DANIEL STEFFEN

REDES NEURAIS E ANÁLISE MULTIVARIADA: ESTUDO E APLICAÇÃO EM
CLASSIFICAÇÃO

Tese apresentada ao Programa de Pós-Graduação em Métodos Numéricos em Engenharia, Setor de Ciências Exatas, Universidade Federal do Paraná, como requisito parcial para a obtenção do título de Doutor em Métodos Numéricos em Engenharia.

Orientador: Prof. Dr. Anselmo Chaves Neto

CURITIBA

2021

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

- S817r Steffen, Daniel,
Redes neurais e análise multivariada: estudo e aplicação em
classificação [recurso eletrônico] / Daniel Steffen – Curitiba, 2021.
- Tese - Universidade Federal do Paraná, Setor de Ciências Exatas
Programa de Pós-Graduação em Métodos Numéricos em
Engenharia.
- Orientador: Prof. Dr. Anselmo Chaves Neto
1. Redes neurais (computação). 2. Kohonen, Mapas de. 3.
Análise Fatorial. I. Universidade Federal do Paraná. II. Chaves
Neto, Anselmo. III. Título.

CDD: 519.5354

Bibliotecária: Roseny Rivelini Morciani CRB-9/1585



MINISTÉRIO DA EDUCAÇÃO
SETOR DE CIÊNCIAS EXATAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO MÉTODOS NUMÉRICOS
EM ENGENHARIA - 40001016030P0

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação MÉTODOS NUMÉRICOS EM ENGENHARIA da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **DANIEL STEFFEN** intitulada: **REDES NEURAIS E ANÁLISE MULTIVARIADA: ESTUDO E APLICAÇÃO EM CLASSIFICAÇÃO**, sob orientação do Prof. Dr. ANSELMO CHAVES NETO, 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 doutor 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, 17 de Dezembro de 2021.

Assinatura Eletrônica
10/01/2022 10:37:53.0
ANSELMO CHAVES NETO
Presidente da Banca Examinadora

Assinatura Eletrônica
17/01/2022 08:18:45.0
GUSTAVO VALENTIM LOCH
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica
27/12/2021 15:03:29.0
ALEXANDRE RASI AOKI
Avaliador Externo (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica
05/01/2022 18:32:46.0
EDSON APARECIDO MITISHITA
Avaliador Externo (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica
17/01/2022 09:51:44.0
JAIR MENDES MARQUES
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica
10/01/2022 15:33:17.0
NELSON HEIN
Avaliador Externo (FUNDAÇÃO UNIVERSIDADE REGIONAL DE BLUMENAU)

AGRADECIMENTOS

Agradeço primeiramente à Deus que permite que tudo isto seja possível.

Ao meu orientador, professor Doutor Anselmo Chaves Neto, pelas aulas de Estatística, pela grande disposição em me atender sempre que necessário e pela paciência nas orientações.

À minha família que sempre me incentivou a concluir mais esta etapa em minha vida.

Ao Programa de Pós-Graduação em Métodos Numéricos em Engenharia pelas aulas disponibilizadas e por todo suporte dado ao decorrer do curso.

Aos professores: Jair Mendes Marques, Liliana Madalena Gramani e Nelson Hein, pelas considerações e recomendações apresentadas durante a qualificação desta tese.

À Universidade Comunitária da Região de Chapecó (UNOCHAPECÓ) pelo auxílio e flexibilização da jornada de trabalho possibilitando assim o cumprimento integral dos créditos em disciplinas do curso.

RESUMO

Este trabalho apresenta uma proposta para a avaliação dos Cursos de Pós-Graduação do Brasil utilizando critérios quantitativos. A metodologia de avaliação proposta também pode ser utilizada em outras atividades em que se pretenda classificar indivíduos ou instituições por desempenho. O trabalho se refere à avaliação de Programas de Pós-Graduação em níveis de Mestrado e Doutorado acadêmico nas áreas de Engenharias III. A CAPES, órgão responsável pela avaliação do ensino de pós-graduação atribui nota/conceito de 2 a 7 aos Programas, seguindo critérios de análise que envolvem dados quantitativos de produção intelectual, bem como, critérios subjetivos, avaliando determinados aspectos dos Programas. A proposta deste trabalho é o desenvolvimento de uma metodologia quantitativa de classificação baseada em Métodos Multivariados e, especificamente, em Análise Fatorial que agrupe as instituições por similaridade utilizando apenas os dados quantitativos dos Programas. Para alcançar este objetivo, aplicou-se a técnica de Redes Neurais Artificiais chamada SOM (*Self Organizing Maps*), também conhecida por Mapas Auto-Organizáveis de Kohonen. A Análise Fatorial é a base do método desenvolvido e sua aplicação se dá em etapa anterior à aplicação do SOM a fim de utilizar os escores fatoriais das variáveis latentes como dados de entrada para o treinamento da rede. Assim, o mapa resultante apresentará o cluster/agrupamento de acordo com a eficiência apresentada na Análise Fatorial. São apresentados os resultados computacionais de ensaios realizados com a metodologia proposta, utilizando 35 variáveis quantitativas próprias da avaliação dos 92 Programas, da avaliação quadrienal (2013 – 2016), para agrupar os cursos similares em desempenho em seis clusters, os quais podem ser atribuídas as notas/conceitos de 2 a 7 de acordo com a classificação obtida.

Palavras-chave: Classificação. Ranking. Redes Neurais. Análise Fatorial. Mapas de Kohonen.

ABSTRACT

This work presents a proposal for evaluating Postgraduate Courses in Brazil using quantitative criteria. The valuation methodology proposed can also be used in other activities where it is intended to rank institutions by performance. The work refers to the evaluation of Postgraduate Programs at Master's and Doctoral levels in the academic areas of Engineering III. The Coordination for the Improvement of Higher Education Personnel (CAPES), is the responsible for evaluating postgraduate education, assigns a grade/concept from 2 to 7 to the Programs, following analysis criteria that involve quantitative data on intellectual production, as well as subjective criteria, evaluating certain aspects of the Programs. The purpose of this work is to develop a quantitative classification methodology based on Multivariate Methods and, specifically, on Factor Analysis, which groups institutions by similarity using only quantitative data from the Programs. To achieve this goal, is applied the Artificial Neural Networks technique called SOM (Self Organizing Maps). Factor Analysis is the basis of the developed method and its application takes place before the application of the SOM in order to use the factor scores of the latent variables as input data for network training. Thus, the resulting map will present the cluster/grouping according to the efficiency presented in the Factor Analysis. The computational results of tests carried out with the proposed methodology are presented, using 35 quantitative variables specific to the evaluation of the 92 Programs, of the quadrennial evaluation (2013 - 2016), to group similar courses in performance into six clusters, which can be attributed to grades/concepts from 2 to 7 according to the classification obtained.

Keywords: Classification. Ranking. Neural Networks. Factor Analysis. Self-Organizing Maps.

LISTA DE FIGURAS

FIGURA 1 – COMPONENTES PRINCIPAIS	25
FIGURA 2 – ILUSTRAÇÃO DE UMA ANÁLISE FATORIAL	29
FIGURA 3 – GRÁFICO DE AUTOVALOR PARA O CRITÉRIO DO TESTE SCREE	36
FIGURA 4 – FLUXOGRAMA DA ANÁLISE FATORIAL POR COMPONENTES PRINCIPAIS E ROTAÇÃO VARIMAX.....	41
FIGURA 5 – FLUXOGRAMA DE EXECUÇÃO DO ALGORÍTMO DAS K-MÉDIAS.....	48
FIGURA 6 – ESTRUTURA DE UMA REDE DE KOHONEN.....	49
FIGURA 7 – EM (A) VÊ-SE A VIZINHANÇA TOPOLÓGICA RETANGULAR E EM (B) O ARRANJO COM VIZINHANÇA HEXAGONAL, ONDE m_i É O I-ÉSIMO NEURÔNIO.....	50
FIGURA 8 – ESTRUTURA DA REDE DE KOHONEN DE DIMENSÃO 4×6	51
FIGURA 9 – ATUALIZAÇÃO DO NEURÔNIO VENCEDOR E DA VIZINHANÇA TOPOLÓGICA EM TORNO DO VETOR DE ENTRADA.....	54
FIGURA 10 – FLUXOGRAMA PARA O ALGORÍTMO QUE PRODUZ OS MAPAS AUTO-ORGANIZÁVEIS (SOM).....	56
FIGURA 11 – FLUXOGRAMA DO RANQUEAMENTO E CLASSIFICAÇÃO POR AF/SOM.....	61
FIGURA 12 – GRÁFICO <i>SCREEPLOT</i> DOS AUTOVALORES OBTIDOS NA ANÁLISE FATORIAL.....	65
FIGURA 13 – CARGAS DOS FATORES 1 E 2 APÓS ROTAÇÃO VARIMAX	66
FIGURA 14 – GRÁFICO DAS VARIÁVEIS CANÔNICAS.....	77
FIGURA 15 – <i>SCRIPT</i> DO ALGORITMO SOM NO SOFTWARE R.....	78
FIGURA 16 – EVOLUÇÃO DO TREINAMENTO DA REDE	79
FIGURA 17 – GRÁFICO VETOR DE PESOS DOS NEURÔNIOS.....	80
FIGURA 18 – FORMAÇÃO DOS CLUSTERS AF/K-MÉDIAS.....	85

LISTA DE TABELAS

TABELA 1 – CLASSIFICAÇÃO DA ESTATÍSTICA KMO.....	39
TABELA 2 – VARIÁVEIS UTILIZADAS NO ESTUDO.....	57
TABELA 3 – ESTATÍSTICA KMO E TESTE DE ESFERICIDADE DE BARTLETT	62
TABELA 4 – MEDIDA DE ADEQUAÇÃO AMOSTRAL ENTRE AS 35 VARIÁVEIS	62
TABELA 5 – AUTOVALORES DA MATRIZ DE DADOS ORIGINAL E PERCENTUAL DA VARIÂNCIA EXPLICADA PELOS MESMOS	63
TABELA 6 – COMUNALIDADES E VARIÂNCIAS ESPECÍFICAS DAS VARIÁVEIS ..	65
TABELA 7 – MATRIZ DE CARREGAMENTOS/PESOS FATORIAIS APÓS ROTAÇÃO VARIMAX	67
TABELA 8 – CARREGAMENTOS/PESOS FATORIAIS COM AS VARIÁVEIS AGRUPADAS	67
TABELA 9 – MATRIZ DOS ESCORES FATORIAIS.....	70
TABELA 10 – RANKING DOS 92 PPG,s AVALIADOS PELA ANÁLISE FATORIAL	73
TABELA 11 – RESULTADO DA APLICAÇÃO DA ANÁLISE DE CORRELAÇÃO CANÔNICA	76
TABELA 12 – PESOS DOS FATORES POR NEURÔNIO/CLUSTER DO SOM.....	79
TABELA 13 – ÍNDICE CLASSIFICATÓRIO DOS CLUSTERS.....	81
TABELA 14 – RESULTADO DA CLASSIFICAÇÃO COM SEIS CLUSTERS (AF/SOM)	82
TABELA 15 – CLASSIFICAÇÃO POR AF/SOM E AF/K-MÉDIAS	86

SUMÁRIO

1	INTRODUÇÃO	12
1.1	DEFINIÇÃO DO PROBLEMA.....	14
1.2	OBJETIVOS.....	15
1.2.1	Objetivo geral	15
1.2.2	Objetivos específicos.....	15
2	REVISÃO DE LITERATURA	16
2.1	A AVALIAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO <i>STRICTO SENSU</i> REALIZADA PELA CAPES.....	16
2.1.1	Quesitos adotados em 2017 na avaliação quadrienal dos Programas de Pós-Graduação das Engenharias III	17
2.1.1.1	Quesito I – Proposta do Programa.....	17
2.1.1.2	Quesito II – Corpo Docente.....	18
2.1.1.3	Quesito III – Corpo Discente, Teses e Dissertações	19
2.1.1.4	Quesito IV – Produção Intelectual	20
2.1.1.5	Quesito V – Inserção Social	22
2.1.2	Quesitos adotados em 2021 na avaliação quadrienal dos Programas de Pós-Graduação das Engenharias III	23
2.1.2.1	Quesito I – Programa.....	23
2.1.2.2	Quesito II – Formação	23
2.1.2.3	Quesito III – Impacto na Sociedade	24
2.2	ANÁLISE MULTIVARIADA.....	24
2.3	ANÁLISE DE COMPONENTES PRINCIPAIS	25
2.3.1	Matriz de Dados	26
2.3.2	Matriz de Covariância Estimada do Vetor $\underline{X}$	27
2.3.3	Determinação das Componentes Principais	27
2.4	ANÁLISE FATORIAL	28
2.4.1	Etapas da Análise Fatorial.....	30
2.4.2	Análise Fatorial pela Matriz de Correlação.....	31
2.4.3	Modelo Fatorial Ortogonal.....	32
2.4.4	Método da Componente Principal.....	34
2.4.5	Estimação do Número de Fatores.....	35
2.4.6	Rotação Ortogonal – Critério Varimax	36

2.4.7	Escores Fatoriais.....	37
2.4.7.1	Método dos Mínimos Quadrados	37
2.4.8	Suposição da Análise Fatorial	38
2.4.8.1	Crítério de Kaiser-Meyer-Olkin (KMO)	39
2.4.8.2	Teste de Esfericidade de Bartlett.....	40
2.5	ANÁLISE DE CORRELAÇÃO CANÔNICA	42
2.6	TÉCNICAS DE AGRUPAMENTO	46
2.6.1	Medidas de Distância e Similaridade	46
2.6.2	Algoritmo K-Médias	47
2.6.3	Mapas Auto-Organizáveis de Kohonen	48
2.6.3.1	Processo Competitivo.....	50
2.6.3.2	Processo Cooperativo	52
2.6.3.3	Processo Adaptativo	53
3	MATERIAL E MÉTODOS	57
3.1	MATERIAL	57
3.2	MÉTODO DE DETERMINAÇÃO DO MODELO DE CLASSIFICAÇÃO	58
4	RESULTADOS E DISCUSSÕES	62
4.1	VERIFICAÇÃO DA ADEQUAÇÃO DA AMOSTRA PARA A ANÁLISE FATORIAL	62
4.2	RESULTADOS DA ANÁLISE FATORIAL	63
4.2.1	Carregamentos Fatoriais e Identificação dos Fatores.....	66
4.2.2	Ranqueamento dos Programas pela Análise Fatorial	73
4.3	RESULTADO DA ANÁLISE DE CORRELAÇÃO CANÔNICA.....	75
4.4	RESULTADO AF/SOM	77
5	CONCLUSÃO E TRABALHOS FUTUROS	87
	REFERÊNCIAS	89
	APÊNDICE 1 – PROGRAMA DESENVOLVIDO EM R PARA O ALGORITMO PROPOSTO	94
	APÊNDICE 2 – MATRIZ DE CORRELAÇÃO R DAS 35 VARIÁVEIS.....	108
	APÊNDICE 3 – GRÁFICO DA CORRELAÇÃO R DAS 35 VARIÁVEIS	109
	APÊNDICE 4 – MATRIZ DE ESCORES FATORIAIS NORMALIZADOS UTILIZADOS PARA O TREINAMENTO DA REDE SOM.....	110
	ANEXO 1 – NÚMEROS E NOMES DOS PROGRAMAS DE PÓS-GRADUAÇÃO (PPG’S)	112

ANEXO 2 – DADOS ORIGINAIS DAS 35 VARIÁVEIS SELECIONADAS..114

1 INTRODUÇÃO

Educação de qualidade é algo primordial para uma nação que almeja o seu desenvolvimento, porque um país só se livra da categoria de “em desenvolvimento” passando para a condição de país desenvolvido com muita tecnologia própria, e isto se alcança começando com uma educação de nível acima do muito bom. Na última década, houve um grande avanço neste sentido, com o aumento na oferta de cursos de Graduação e Pós-Graduação no Brasil, conseqüentemente, um aumento significativo de matrículas nas Universidades. Esse crescimento quantitativo torna cada vez mais importante a avaliação da universidade brasileira sob vários aspectos, em relação ao ensino e também em relação ao desempenho dos Programas. No Brasil, o órgão responsável pelo acompanhamento dos cursos de Pós-Graduação *Stricto Sensu* é a CAPES.

A CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), foi criada em 1951, durante o segundo governo de Getúlio Vargas, com o objetivo de garantir a existência de pessoal especializado em quantidade e qualidade para atender as demandas nacionais.

A CAPES é o órgão responsável pela elaboração do Plano Nacional de Pós-Graduação *Stricto Sensu*, cabendo a ela elaborar, avaliar, acompanhar e coordenar as atividades relativas ao Ensino Superior. Destacam-se as seguintes atribuições da CAPES:

- i) Coordenar e avaliar os cursos de Pós Graduação;
- ii) Estimular mediante bolsa de estudo, auxílios e outros mecanismos, a formação de recursos humanos altamente qualificados para a docência de grau superior, para a pesquisa e para o atendimento da demanda dos setores público e privado.

A partir de 2009 é também atribuição da CAPES, induzir e fomentar a formação inicial e continuada de professores para a Educação Básica.

Em paralelo à atividade exercida pela CAPES, novas técnicas de análise da qualidade em Educação vêm sendo discutidas e implementadas frequentemente. A metodologia da Análise por Envoltória de Dados (DEA), junto com outros métodos da pesquisa operacional vem sendo empregados com sucesso em análises educacionais. (MIRANDA et al. 2010)

No Brasil, os primeiros trabalhos que utilizam DEA na construção de medidas de avaliação de Intuições de Ensino Superior (IES) têm origem em grupos de pesquisa da Universidade Federal de Santa Catarina (UFSC), onde se pode destacar o trabalho de Beloni (2000). Este autor, construiu uma metodologia baseada em DEA para avaliação de eficiência produtiva de Universidades Federais Brasileiras.

Outro trabalho que pode ser citado é o de Borba (2011), que propôs um método para avaliar a eficiência técnica do ensino em Programas de Pós-Graduação, na área de Engenharias III, sob o prisma do ensino, e com foco na formação e o aperfeiçoamento de recursos humanos. Este trabalho foi inovador, pois, a pesquisa foi realizada sob o enfoque do ensino, o que difere da metodologia adotada pela CAPES.

A aplicação de Métodos Multivariados também se mostra eficiente em pesquisas científicas nas mais variadas áreas. Em Economia, por exemplo, a avaliação para concessão de crédito a consumidores é uma prática já realizada por algumas organizações. A decisão dessa concessão é fundamentada em técnicas de classificação, ou seja, de rotulação do cliente como “bom” ou “ruim”. A Análise Fatorial e a Análise de Componentes Principais podem ser empregadas nestes casos para identificar as correspondências entre as variáveis, sendo possível ranquear os consumidores em relação a sua capacidade de pagamento. A hierarquização de gerentes de banco, em função da sua atividade com os produtos do banco, para auxiliar na tomada de decisão de movimentação desses gerentes nas muitas agências da instituição bancária é uma aplicação importante da Análise Fatorial e grandes instituições bancárias a utilizam com frequência. E, ainda, na área da Engenharia Civil foi desenvolvida uma metodologia baseada na Análise Fatorial para ranquear pontes sujeitas à manutenção, identificando as que deveriam ser trabalhadas em primeiro lugar. Essas técnicas de Análise Multivariada podem ser adaptadas para as mais diversas áreas, inclusive para a área da Educação.

Em Redes Neurais, pode-se citar alguns trabalhos direcionados a gestão educacional, assim como no trabalho de Martinho (2014), onde a autora propõe um método para predição da evasão discente em cursos superiores presenciais utilizando uma Rede Neural ARTMAP-Fuzzy. Com a metodologia adotada em seu trabalho pôde-se identificar precocemente, ainda no primeiro semestre do curso, o grupo de estudantes em risco de evasão e suas possíveis causas.

A rede SOM (Redes Neurais Auto-Organizáveis) é um algoritmo de Redes Neurais muito utilizado em aprendizado não supervisionado, agrupamento, classificação e visualização de dados. Várias publicações estão relatadas na literatura e vários projetos comerciais empregam o SOM como um procedimento para resolver problemas considerados difíceis do mundo real (COTTRELL, 2006). Sistemas híbridos envolvendo essa rede neural também foram aplicados com sucesso em estudos tais como em Churilov et al. (2006), no qual se aplica a técnica DEA, juntamente com uma Rede SOM, produzindo um novo ranqueamento em um sistema modelado para ser imparcial e considerado como “justo” pelos grupos envolvidos na análise.

Neste trabalho é proposta a utilização de Redes Neurais Auto-Organizáveis (SOM), em uma aplicação de classificação e comparação de procedimentos em dados da avaliação quadrienal (2013 – 2016) da CAPES referentes à 92 Programas de Pós-Graduação Acadêmicos na Área de Engenharias III. Este estudo propõe uma metodologia de ranqueamento e clusterização conforme desempenho, sem levar em conta critérios subjetivos, diferentemente da avaliação feita pela CAPES que os leva em consideração.

Nesta proposta, a novidade é a caracterização dos agrupamentos de acordo com a hierarquização obtida da aplicação da Análise Fatorial. Para que as propriedades da hierarquização dos indivíduos verificada na Análise Fatorial seja refletida na aplicação do SOM utiliza-se os escores fatoriais das variáveis latentes como dado de entrada para o treinamento da rede, pois esses escores carregam praticamente toda a informação referente às variáveis do problema, perdendo apenas uma parcela insignificante da variabilidade, mas situando-se nas direções de maior variabilidade.

Esta aplicação em conjunto, Análise Fatorial e SOM, se faz necessária para a posterior classificação dos clusters, pois após a Análise Fatorial se torna conhecida a importância das variáveis latentes pelo grau de explicação que cada uma delas carrega. Com isso, é possível ranquear o cluster como eficiente ou não eficiente, podendo inclusive identificar as características necessárias a cada indivíduo para ser bem classificado.

A mesma metodologia é aplicada utilizando o método de agrupamento das K-Médias em substituição ao SOM para fins de comparação.

1.1 DEFINIÇÃO DO PROBLEMA

Várias organizações utilizam critérios de classificação para o ranqueamento de pessoas e/ou instituições. Em diversos setores, os ranqueamentos são utilizados para classificar os melhores indivíduos dentro de um grupo, ou até mesmo determinar que um grupo de indivíduos seja considerado mais eficiente e produtivo que outro.

Nas avaliações de Programas de Pós-Graduação *Stricto Sensu* em específico, o ranqueamento ocorre com a atribuição de conceito que está em uma escala de 2 a 7. Em problemas desta natureza, onde os indivíduos semelhantes recebem a mesma classificação pode-se buscar o ranqueamento com agrupamentos.

Para solucionar este problema, propõe-se o desenvolvimento de um novo método de classificação com agrupamento envolvendo os critérios de classificação de Análise Multivariada e a técnica de clusterização dos Mapas Auto-Organizáveis de Kohonen.

1.2 OBJETIVOS

1.2.1 Objetivo geral

O objetivo geral deste trabalho é propor uma metodologia baseada em Métodos Multivariados e Mapas Auto-Organizáveis (SOM) que elimine subjetividade, que pode estar presente, em avaliações.

1.2.2 Objetivos específicos

- a) Desenvolver uma metodologia baseada em Métodos Multivariados e Mapas Auto-Organizáveis (SOM) para construir um índice classificatório com a formação de clusters;
- b) Implementar computacionalmente um algoritmo de classificação e clusterização para avaliar indivíduos e/ou instituições utilizando Análise Fatorial e SOM;
- c) Aplicar a metodologia proposta utilizando dados da CAPES referentes aos Programas de Pós-Graduação *Stricto Sensu* na Área de Engenharias III.

2 REVISÃO DE LITERATURA

2.1 A AVALIAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO *STRICTO SENSU* REALIZADA PELA CAPES

A Fundação de Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), ligada ao Ministério da Educação (MEC), é o órgão que define a abertura, funcionamento e avaliação dos Programas de Pós-Graduação no Brasil.

Dentre os fatores mais relevantes de se ter um sistema de avaliação está a busca permanente da elevação dos padrões da pesquisa brasileira (CAPES, 2001). Atualmente os Programas de Pós-Graduação (PPG's) são avaliados a cada 4 anos. No entanto, eles são avaliados anualmente com intuito de estabelecer um diálogo, objetivando a melhoria da qualidade dos PPG's a fim de verificar se as metas propostas no projeto inicial estão sendo plenamente atingidas no âmbito do Programa. Caso contrário, o mesmo pode ser descredenciado.

A avaliação quadrienal dos cursos na CAPES é desenvolvida pelos comitês de área, formados por consultores acadêmicos escolhidos dentre profissionais de comprovada competência em pesquisa e ensino de pós-graduação. Os comitês de área também são responsáveis pelo estabelecimento de critérios e diretrizes para essa avaliação, que são divulgados em documentos das respectivas áreas dos comitês. Compete ainda aos comitês de área a qualificação de periódicos, anais, jornais e revistas (QUALIS) em cada área de pesquisa (CAPES, 2017).

Segundo a Portaria Nº 59, de 21 de março de 2017 do MEC, que dispõe sobre o regulamento da Avaliação Quadrienal, pode-se destacar que são submetidos à avaliação os programas recomendados pela CAPES, que estejam em funcionamento de acordo com as seguintes condições mínimas:

- Programas Acadêmicos: 3 anos
- Programas Profissionais: 2 anos

A avaliação dos Programas Acadêmicos ocorre em separado aos Programas Profissionais, utilizando comissões diferentes e fichas de avaliação específicas para cada modalidade.

Segundo a CAPES (2017), os programas recebem nota final na escala de “2 a 7”, baseada em quesitos atribuídos (Muito Bom, Bom, Regular, Fraco e Insuficiente), observada as seguintes orientações:

- 1) O programas com conceito “insuficiente” no Quesito 1, “Proposta do Programa”, poderá alcançar no máximo nota 2 e com conceito “Fraco” poderá alcançar no máximo 3.
- 2) Recomendação da nota 3: corresponde ao padrão mínimo de qualidade para a recomendação do programa ao Conselho Nacional de Educação (CNE) e permanência no Sistema Nacional de Pós-Graduação (SNPG).
- 3) Recomendação da nota 4: é possível para programas que tenham, no mínimo conceito “Bom” em três quesitos, incluindo os quesitos 3 e 4.
- 4) Recomendação da nota 5: obter “Muito Bom” em quatros quesitos, incluindo os quesitos 3 e 4. Esta é a nota máxima para um programa de mestrado.
- 5) Recomendação para notas 6 e 7: exclusivo para programas com doutorado que obtiveram nota 5 e conceitos “Muito Bom” em todos os quesitos e que atendam as condições: a) desempenho equivalente a centros de excelência internacionais; b) desempenho diferenciado no que se refere a formação de doutores e produção intelectual; c) nucleação de novos programas no país ou no exterior.

No caso de programas com doutorado receber nota 3, poderá a critério da comissão decidir pelo descredenciamento do doutorado, atribuindo nota 2 e manter o mestrado com nota 3.

Os quesitos utilizados na avaliação quadrienal serão detalhados nos próximos capítulos, as informações foram retiradas do “Documento de Área (2017)” das Engenharias III, disponível no site da CAPES.

2.1.1 Quesitos adotados em 2017 na avaliação quadrienal dos Programas de Pós-Graduação das Engenharias III

2.1.1.1 Quesito I – Proposta do Programa

De acordo com o conselho técnico científico, em relação a Proposta do Programa três itens prioritários são considerados na avaliação:

- Coerência, consistência, abrangência e atualização das áreas de concentração, linhas de pesquisa, projetos em andamento e proposta curricular (peso 40%): são verificados aspectos como coerência e consistência das linhas de pesquisa com as áreas de concentração, das linhas de pesquisa com os projetos que estão em

andamento, verifica-se também se há atualização da grade curricular e se as disciplinas estão em consonância com o corpo docente permanente.

- Planejamento do programa com vistas a seu desenvolvimento futuro, contemplando os desafios internacionais da área na produção do conhecimento, seus propósitos na melhor formação de seus alunos, suas metas quanto à inserção social mais rica dos seus egressos, conforme os parâmetros da área (peso 40%): é verificado de que forma o curso vê sua trajetória e evolução de seu conceito na avaliação CAPES, que medidas o programa adota para a melhoria do curso no que tange os itens supracitados.
- Infraestrutura para ensino, pesquisa e, se for o caso, extensão (peso 20%): é verificado se a infraestrutura é adequada para a proposta do programa.

2.1.1.2 Quesito II – Corpo Docente

Em relação ao corpo docente, quatro itens prioritários são considerados na avaliação, este quesito tem peso 20% na avaliação total:

- Perfil do corpo docente, consideradas titulação, diversificação na origem de formação, aprimoramento e experiência, e sua compatibilidade e adequação à Proposta do Programa (Peso 30%): Neste item avalia-se o corpo docente, considerando sua posição como pesquisador e aderência dos trabalhos à proposta do programa. Leva-se em conta também, o impacto da produção dos docentes permanentes em termos de indicadores de citação.
- Adequação e dedicação dos docentes permanentes em relação às atividades de pesquisa e de formação do programa (Peso 30%): São considerados, o número de disciplinas ministradas pelo docente na pós-graduação, a qualidade e relevância dos projetos de pesquisa aprovados com financiamento junto aos órgãos de fomento, bem como o engajamento dos professores envolvidos.
- Distribuição das atividades de pesquisa e de formação entre os docentes do programa (Peso 30%): verifica-se a percentagem de docentes envolvidos em atividades de pesquisa e de formação. É verificado se os docentes permanentes (no quadriênio) ministraram no mínimo 4 disciplinas no programa, participaram de pelo menos uma publicação em periódico A1, A2 ou B1 e tiveram pelo menos 2 orientações no programa concluídas.

- Contribuição dos docentes para atividades de ensino e/ou de pesquisa na graduação (Peso 10%): Este item só vale quando o PPG estiver ligado a curso de graduação; se não o estiver, seu peso será redistribuído proporcionalmente entre os demais itens do quesito.

2.1.1.3 Quesito III – Corpo Discente, Teses e Dissertações

Em relação ao corpo discente, teses e dissertações são obtidos 4 índices para compor a avaliação, este quesito tem peso 35% na avaliação total:

- Quantidade de teses e dissertações defendidas no período de avaliação, em relação ao corpo docente permanente e à dimensão do corpo discente (Peso 30%): o indicador de teses e dissertações é dado pela equação:

$$ORI = \frac{MT + 2DT}{TD} \quad (1)$$

onde *MT*: total de mestres titulados, *DT*: total de doutores titulados e *TD*: total de docentes incluídos os colaboradores.

- Distribuição das orientações das teses e dissertações defendidas no período de avaliação em relação aos docentes do programa (Peso 10%): o cálculo deste indicador é dado por:

$$PSA = \frac{DPMD}{DP \cdot 100} \quad (2)$$

onde *PSA*: percentual de docentes permanentes que não tiveram dissertações ou teses orientadas e aprovadas no quadriênio e *DP*: total de docentes permanentes.

- Qualidade das teses e dissertações e da produção de discentes autores da pós-graduação e da graduação (no caso de IES com curso de graduação na área) na produção científica do programa, aferida por publicações e outros indicadores pertinentes à área (Peso 50%): A avaliação deste item é realizada através de dois subitens, um quantitativo e outro qualitativo. A avaliação quantitativa é dada por:

$$PRD = \frac{4PIL + 3PNL + 2PIC + 2PNC + 0,1PID + 0,05PND + 0,05SWR + A1 + 0,9A2 + 0,75B1 + 0,5B2 + 0,3B3 + 0,1B4 + 0,05B5 + 0,1AI + 0,05AN + 0,025REI}{NRD} \quad (3)$$

Onde:

NRD: número de teses e dissertações defendidas no quadriênio;

PIL: número de patentes internacionais licenciadas;

PNL: número de patentes nacionais licenciadas;

PIC: número de patentes internacionais concedidas;

PNC: número de patentes nacionais concedidas;

PID: número de patentes internacionais depositadas;

PND: número de patentes nacionais depositadas;

SWR: software registrado no INPI;

A1: número de publicações classificadas como A1 nos periódicos qualis capes;

A2: número de publicações classificadas como A2 nos periódicos qualis capes;

B1: número de publicações classificadas como B1 nos periódicos qualis capes;

B2: número de publicações classificadas como B2 nos periódicos qualis capes;

B3: número de publicações classificadas como B3 nos periódicos qualis capes;

B4: número de publicações classificadas como B4 nos periódicos qualis capes;

B5: número de publicações classificadas como B5 nos periódicos qualis capes;

AI: anais internacionais;

AN: anais nacionais;

REI: resumos expandidos internacionais.

A avaliação qualitativa é realizada com base em outras produções técnicas não consideradas no subitem quantitativo, o programa deverá listar as 5 produções técnicas mais relevantes, que estejam relacionadas às teses e dissertações aprovadas.

- Eficiência do Programa na formação de mestres e doutores bolsistas (Peso 10%): este item avalia o tempo de formação de mestres e doutores e percentual de bolsistas titulados.

2.1.1.4 Quesito IV – Produção intelectual

Este item tem peso 35% na avaliação total, e leva em conta 3 índices:

- Publicações qualificadas do programa por docente permanente (Peso 50%): este é um indicador referente a qualidade das publicações de docentes permanentes em relação aos docentes totais.

$$PQD = \frac{A1 + 0,9A2 + 0,75B1 + 0,5B2 + 0,3B3 + 0,1B4 + 0,05B5}{TD} \quad (4)$$

Onde:

PQD: indicador de publicações;

A1: número de publicações classificadas como A1 nos periódicos qualis capes;

A2: número de publicações classificadas como A2 nos periódicos qualis capes;

B1: número de publicações classificadas como B1 nos periódicos qualis capes;

B2: número de publicações classificadas como B2 nos periódicos qualis capes;

B3: número de publicações classificadas como B3 nos periódicos qualis capes;

B4: número de publicações classificadas como B4 nos periódicos qualis capes;

B5: número de publicações classificadas como B5 nos periódicos qualis capes;

TD: total de docentes.

- Distribuição de publicações qualificadas em relação ao corpo docente permanente do programa (Peso 30%): neste critério é obtido o índice de participação das publicações por docente do programa.
- Produção técnica, patentes e outras produções consideradas relevantes (Peso 20%): recomenda-se que cada PPG demonstre a existência dessa produção destacando as mais relevantes. A avaliação é feita de forma quantitativa e qualitativa. A quantitativa envolve a obtenção de um índice dado por:

$$PTC = \frac{4PIL + 3PNL + 2PIC + PNC + 0,1PID + 0,05PND + 0,05SWR + 0,2CLI + 0,1CLN + 0,5LID + 0,1AI + 0,05AN + 0,025REI}{TD} \quad (5)$$

Onde:

PTC: índice de produção técnica dos docentes permanentes;

TD: total de docentes;

PIL: número de patentes internacionais licenciadas;

PNL: número de patentes nacionais licenciadas;

PIC: número de patentes internacionais concedidas;

PNC: número de patentes nacionais concedidas;

PID: número de patentes internacionais depositadas;

PND: número de patentes nacionais depositadas;

SWR: software registrado no INPI;

CLI: número de capítulos de livros de circulação internacional;

CLN: número de capítulos de livros de circulação nacional;

LID: número de livros (texto integral) com repercussão tecnológica, de extensão ou didática (não científica);

AI: anais internacionais;

AN: anais nacionais;

REI: resumos expandidos internacionais.

A avaliação qualitativa deste item se dá a partir de uma análise global, devem ser considerados outros itens de produção técnica como: prestação de serviços, desenvolvimento de material didático ou de produto, revisões para periódicos nacionais e internacionais, bem como toda produção técnica que o programa considerar relevante.

2.1.1.5 Quesito V – Inserção Social

Este item tem peso 10% na avaliação total, e leva em conta 3 índices:

- Inserção e impacto regional e (ou) nacional do programa (Peso 60%): Neste item são verificados os diferentes tipos impactos (social, educacional, tecnológico, econômico e profissional) que os produtos originários de teses e dissertações proporcionam à sociedade, inclusive as ações dos egressos.
- Integração e cooperação com outros programas e centros de pesquisa e desenvolvimento profissional relacionados à área de conhecimento do programa, com vistas ao desenvolvimento da pesquisa e da pós-graduação (Peso 20%): Neste item é verificada a efetiva participação dos programas em projetos de cooperação, voltados para inovação e desenvolvimento da pós-graduação.
- Visibilidade ou transparência dada pelo programa a sua atuação (Peso 20%): Neste item é avaliado se o programa disponibiliza as informações de procedimentos em meio eletrônico (inclusive em inglês) e se disponibiliza em rede as teses e dissertações aprovadas.

2.1.2 Quesitos adotados em 2021 na avaliação quadrienal dos Programas de Pós-Graduação das Engenharias III

Uma das novidades na avaliação quadrienal (2017-2020) é a estrutura da ficha de avaliação dos quesitos que foi reduzida de 5 para 3 quesitos principais. Sendo avaliados Programa, Formação e Impacto. O objetivo é enfatizar a avaliação e a avaliação de resultados. No quesito Programa, serão avaliados o funcionamento, a estrutura e o planejamento do programa de pós-graduação. Quanto à Formação, a análise abrangerá aspectos como qualidade das teses, dissertações e produção intelectual de alunos e professores e das atividades de pesquisa, bem como o desempenho do egresso. Já em Impacto na Sociedade, serão verificados o caráter inovador da produção intelectual, os efeitos econômicos e sociais do programa bem como sua internacionalização e visibilidade (CAPES, 2021).

2.1.2.1 Quesito I – Programa

De acordo com o conselho técnico científico, em relação a proposta do programa três itens prioritários são considerados na avaliação:

- Articulação, aderência e atualização das áreas de concentração, linhas de pesquisa, projetos em andamento e estrutura curricular, bem como a infraestrutura disponível, em relação aos objetivos, missão e modalidade do programa (peso 40%).
- Perfil do corpo docente, e sua compatibilidade e adequação à proposta do programa (peso 40%).
- Planejamento estratégico, considerando também articulações com o planejamento estratégico da instituição, com vistas à gestão do seu desenvolvimento futuro, adequação e melhorias da infraestrutura e melhor formação de seus alunos, vinculada à produção intelectual – bibliográfica, técnica e/ou artística (peso 10%).
- Os processos, procedimentos e resultados da autoavaliação, com foco na formação discente e produção intelectual (peso 10%).

2.1.2.2 Quesito II – Formação

- Qualidade e adequação das teses, dissertações ou equivalente em relação às áreas de concentração e linhas de pesquisa do programa (peso 15%).
- Qualidade da produção intelectual de discentes e egressos (peso 25%).

- Destino, atuação e avaliação dos egressos em relação à formação recebida (peso 10%).
- Qualidade das atividades de pesquisa e da produção intelectual do corpo docente no programa (peso 40%).
- Qualidade e envolvimento do corpo docente em relação às atividades de formação no programa (peso 10%).

2.1.2.3 Quesito III – Impacto na Sociedade

- Impacto e caráter inovador da produção intelectual em função da natureza do programa (peso 40%).
- Impacto econômico, social e cultural do programa (peso 30%).
- Internacionalização, inserção (local, regional, nacional) e visibilidade do programa (peso 30%).

Os programas ainda podem preencher o módulo de Destaques. Ali, os PPGs poderão marcar seus produtos mais relevantes, que evidenciem a importância de suas ações. As ferramentas para seleção estão divididas em cinco partes:

- Produções para classificações;
- Teses e Dissertações;
- Produção docente;
- Produções do ciclo avaliativo;
- Egressos.

Nestes itens os Programas poderão indicar trabalhos de destaque como artigos, teses, dissertações e produtos criados por alunos professores.

2.2 ANÁLISE MULTIVARIADA

As técnicas da Análise Multivariada são amplamente utilizadas em várias áreas do conhecimento. Quando o objetivo é medir, explicar e prever o grau de relacionamento entre variáveis estatísticas com a intenção de reduzir o número de variáveis, as técnicas multivariadas podem ser empregadas com sucesso. O termo multivariado, não está relacionado somente ao fato de existir mais de uma variável no problema, o caráter multivariado reside nas múltiplas variáveis estatísticas envolvidas.

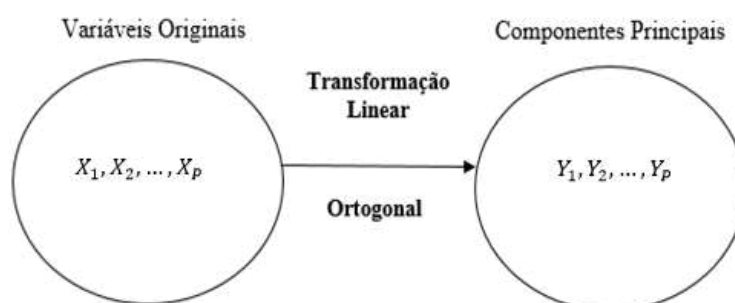
Em Análise Estatística de Dados, os métodos estão dispostos em dois grupos: a Análise Univariada e a Análise Multivariada. A primeira, olha cada variável de maneira isolada e a segunda olha as variáveis de forma conjunta, trabalhando principalmente com o seu relacionamento. Embora estivesse pronta já na década de 30, as técnicas multivariadas só foram aplicadas com eficiência com o advento da computação veloz e barata, do computador. E, foram se ampliando com o surgimento de programas computacionais de fácil uso, tais como Statgraphics, Minitab, R, entre outros.

A Análise Multivariada possui certo número de técnicas que utilizam, simultaneamente, todas as variáveis na interpretação teórica do conjunto de dados obtidos (CHAVES NETO, 2004). Como existem vários métodos, com finalidades diferentes, é necessário saber identificar qual método utilizar. Portanto, é preciso saber onde se quer chegar, isto é, deve-se ter bem definido os objetivos da pesquisa. Se for preciso verificar como as amostras se relacionam, pode-se citar a Análise de Componentes Principais.

2.3 ANÁLISE DE COMPONENTES PRINCIPAIS

A Análise de Componentes Principais (ACP) foi desenvolvida por Pearson (1901) e Hotelling (1933). Trata-se de uma técnica multivariada poderosa que pode ser utilizada para a redução do número de variáveis, eliminando sobreposições e escolhendo a forma mais representativa de dados. Essa técnica consiste em reescrever as variáveis originais em termos de novas variáveis, que são combinações lineares das originais e são chamadas de componentes principais. Segundo Morrison (1976), a ACP surge da necessidade de se conhecer as estruturas de dependência das variáveis que *a priori* não é encontrada. Cada componente principal é uma combinação linear de todas as variáveis originais e que terão peso e importâncias diferentes na análise. A FIGURA 1 a seguir, ilustra esse processo.

FIGURA 1 – COMPONENTES PRINCIPAIS



FONTE: Adaptado de Marques (2008).

A transformação ortogonal que gera as “componentes principais” pode ser vista como uma operação que gera um novo conjunto de dados, ou um novo conjunto de coordenadas que são perpendiculares (ou não) entre si. Estas componentes são então, não correlacionadas entre si, ao contrário dos dados originais (ANDRIOTTI, 1997).

As Componentes Principais podem ser usadas de acordo com os objetivos inicialmente propostos. Deve-se destacar, quais as variáveis ou, quais os conjuntos de variáveis explicam a maior parte da variabilidade total, revelando o tipo de relacionamento existente entre eles.

De acordo com Vasconcelos (2012), os passos para calcular as Componentes Principais são:

- Obter os dados ou as m amostras de vetores de dimensão n ;
- Calcular a média ou o vetor médio destes dados;
- Subtrair a média de todos os itens de dados;
- Calcular a matriz de covariância usando todas as subtrações. Ela é o resultado da média do produto de cada subtração por ela mesma e terá dimensão $n \times n$;
- Calcular os autovalores e autovetores da matriz de covariância.
- Arranjar os dados da matriz da Transformada de Hotelling (cujas linhas são formadas a partir dos autovetores da matriz de covariância arranjados de modo que a primeira linha, o elemento (0, 0), seja o autovetor correspondente ao maior autovalor, e assim sucessivamente até que a última linha corresponda ao menor autovalor.

O autovetor com o maior autovalor associado, corresponde à componente principal do conjunto de dados usado. Isso significa que esse é o relacionamento mais significativo entre as dimensões dos dados.

2.3.1 Matriz de Dados

Considere p características de n indivíduos de uma população. As características observáveis são representadas por $X_1, X_2, X_3, \dots, X_p$ e podem compor o vetor aleatório $\underline{X}$ e, as n observações desse vetor formam a matriz de dados X de ordem $n \times p$,

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1p} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2p} \\ x_{31} & x_{32} & x_{33} & \dots & x_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{np} \end{bmatrix}$$

A estrutura de covariância do vetor $\underline{X}$ pode ser representada pela matriz de covariância Σ ou pela matriz de correlação ρ e, é possível por técnica multivariada transformar essa estrutura em outra estrutura representada por p componentes não correlacionadas $Y_1, Y_2, Y_3, \dots, Y_p$ (FIGURA 1). Essas componentes são chamadas de principais e têm a vantagem de não depender da suposição de Gaussianidade para os dados.

2.3.2 Matriz de Covariância Estimada do Vetor $\underline{X}$

O estimador S da matriz de covariância Σ do vetor $\underline{X}$ é formada com as variâncias estimadas, $\hat{\sigma}_i^2, i = 1, 2, \dots, p$ compondo a diagonal principal e as covariâncias estimadas, $\hat{\sigma}_{ij}, i, j = 1, 2, \dots, p$ ocupando as posições na matriz fora da diagonal principal. A expressão da matriz é dada por:

$$S = \begin{bmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12} & \hat{\sigma}_{13} & \dots & \hat{\sigma}_{1p} \\ \hat{\sigma}_{21} & \hat{\sigma}_2^2 & \hat{\sigma}_{23} & \dots & \hat{\sigma}_{2p} \\ \hat{\sigma}_{31} & \hat{\sigma}_{32} & \hat{\sigma}_3^2 & \dots & \hat{\sigma}_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{n1} & \hat{\sigma}_{n2} & \hat{\sigma}_{n3} & \dots & \hat{\sigma}_p^2 \end{bmatrix}$$

Segundo Regazzi (2000) é conveniente padronizar as variáveis X_j ($j = 1, 2, 3, \dots, p$) quando as escalas das unidades de medidas das características observadas não forem as mesmas. A padronização geralmente é feita com média 0 e variância 1 e, assim, trabalha-se com a matriz de correlação ρ , em que,

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j}, \quad i = 1, 2, \dots, n \quad \text{e} \quad j = 1, 2, \dots, p \quad (6)$$

onde, $\bar{x}_j$ e S_j são, respectivamente, a estimativa da média e do desvio padrão da característica X_j .

2.3.3 Determinação das Componentes Principais

As componentes principais são estimadas resolvendo-se a equação característica da matriz S ou ρ , isto é:

$$\det[S - \lambda I] = 0 \quad (7)$$

Se a matriz S for de posto completo igual a ' p ', a equação (7) terá ' p ' raízes chamadas de autovalores ou raízes características da matriz S . Sejam $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_p$ as raízes da equação característica da matriz S , então, $\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_p$. Para cada autovalor λ_i , existe um autovetor de dimensão p associado, $\underline{e}_i' = [e_{i1} \ e_{i2} \ \dots \ e_{ip}]$. Os autovetores $\underline{e}_i$ são ortonormais, e portanto, apresentam as características: $\sum_{j=1}^p e_{ij}^2 = 1 \ (\langle \underline{e}_i, \underline{e}_i \rangle = 1)$ e $\sum_{j=1}^p e_{ij} \cdot e_{kj} = 0 \ (\langle \underline{e}_i, \underline{e}_k \rangle = 0, \forall k \neq j)$.

Sendo $\underline{e}_i$ o autovetor correspondente ao autovalor λ_i , então o i -ésimo componente principal é dado pela combinação linear:

$$Y_i = e_{i1}X_1 + e_{i2}X_2 + \dots + e_{ip}X_p \quad (8)$$

A primeira componente é a que apresenta a maior variância (autovalor correspondente), o segundo componente tem a segunda maior variância, e assim sucessivamente; $V(Y_1) > V(Y_2) > \dots > V(Y_p)$.

A maior parte da variabilidade pode ser explicada por um número k de variáveis sendo $k < p$. Então, as variáveis $Y_j, (j = 1, 2, \dots, k)$ são chamadas de componentes principais extraídas. A quantidade de informações contida nessas $k < p$ componentes é praticamente a mesma das p variáveis originais, mas estão nas direções de maior variabilidade. Portanto, as p variáveis podem ser substituídas pelas k componentes principais sem maior perda de informação.

2.4 ANÁLISE FATORIAL

A Análise Fatorial é uma técnica multivariada inicialmente proposta no início do século 20, por Charles Spearman e Karl Pearson (MAROCO, 2007), para entender problemas relacionados a psicologia educacional e, particularmente, ao estudo de variáveis latentes como a inteligência. Spearman testou a hipótese de que diferentes testes de habilidade mental poderiam ser explicadas por um fator comum, a inteligência. Thurstone (1935) desenvolveu a ideia de "multiple factor analysis". Hotelling propôs "o Método das Componentes Principais que permite o cálculo da única matriz de fatores ortogonais" (KAPLUNOVSK, 2009). Segundo Marriot (1974), a Análise Fatorial é a principal e a mais antiga técnica de Análise Multivariada.

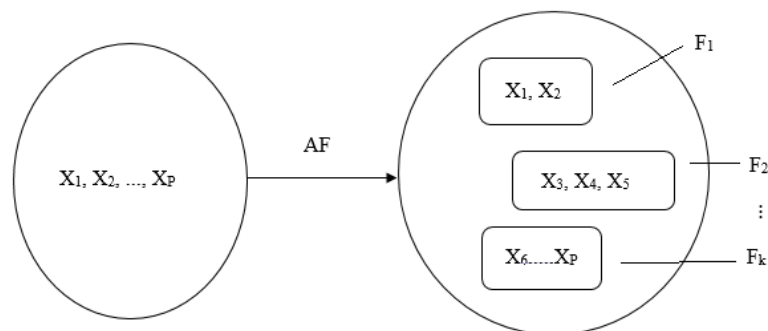
Segundo Menezes et al. (1978), a diferença básica entre o modelo de Análise de Componentes Principais e o modelo clássico de Análise Fatorial é que, no primeiro caso,

considera-se a variável como tal, sem tentar extrair dela o fator único. Na Análise Fatorial, faz-se uma estimativa através da comunalidade, que é inserida na diagonal da matriz de pesos. Como a variância explicada pelo modelo é a relação entre a soma dos valores na diagonal e a soma dos valores fora da diagonal, quanto mais o valor diagonal exceder a comunalidade correta (aproximar-se de um), tanto maior será a parcela da variância não comum sendo tomada como se fosse variância comum.

Kubrusly (1981) aponta que, ao contrário do modelo de Análise das Componentes Principais que busca explicar a variância total dos dados, o modelo de Análise Fatorial procura reproduzir da melhor forma possível a correlação entre as variáveis originais.

O propósito principal da Análise Fatorial é descrever a estrutura de covariâncias entre as variáveis em termos de um número menor de variáveis não observáveis (latentes) chamadas fatores. Em suma, as variáveis são agrupadas levando-se em conta suas associações, logo, as variáveis dentro de cada grupo (fator), terão forte correlação. Em outras palavras a Análise Fatorial explica as correlações entre um conjunto grande de variáveis em termos de um conjunto de poucas variáveis não observáveis chamados fatores. Este processo está ilustrado na FIGURA 2, a seguir, onde no segundo círculo aparecem as variáveis que dominam especificamente cada fator. Deve-se lembrar que cada fator é combinação linear das p variáveis.

FIGURA 2 – ILUSTRAÇÃO DE UMA ANÁLISE FATORIAL



FONTE: Adaptado de Marques (2008).

Gontijo e Aguirre (1988) descrevem os objetivos da Análise Fatorial que são:

- Condensar um grande número de observações em grupos;
- Obter um número reduzido de variáveis em comparação ao original sem grande perda de informação;
- Obter os fatores que reproduzem um padrão separado de relações entre as variáveis;
- Interpretar de forma lógica o padrão de relações entre as variáveis;

- Identificar as variáveis apropriadas para Análise de Regressão e Correlação ou Análise Discriminante.

2.4.1 Etapas da Análise Fatorial

Há basicamente, quatro passos para a realização de uma Análise Fatorial:

1º) Entrada dos dados: Conjunto de observações das variáveis para cada objeto ou indivíduo formando a matriz de dados. Qualquer matriz cujos componentes apresentem uma medida de similaridade entre variáveis (correlação), pode ser utilizada para Análise Fatorial.

2º) Cálculo das Correlações: estima-se a matriz de correlação do vetor aleatório observado com base nos valores que formam a matriz de dados e a partir daí é feita a Análise Fatorial R ou Análise Fatorial Q. Segundo Hair et al. (2005), o tipo mais comum de Análise Fatorial é a do tipo R, obtida da matriz de correlação das variáveis.

3º) Extração de Fatores Iniciais: Costa (2006) descreve que o objetivo da extração de fatores é encontrar um conjunto de fatores, onde cada um forma uma combinação linear das variáveis originais. Se certas variáveis são altamente correlacionadas entre si, elas receberão pesos (coeficientes) maiores na combinação linear que forma um fator. Pode-se definir uma combinação linear entre variáveis da seguinte forma:

$$F_j = C_{1j}X_1 + C_{2j}X_2 + \dots + C_{pj}X_p \quad j = 1, 2, \dots, k. \quad (9)$$

em que, F_j é denominado fator comum e corresponde a combinação linear entre as variáveis $X_1, X_2, \dots, X_p$, sendo que variáveis com maior correlação receberão coeficientes C_{ij} maiores. Cada fator comum F_j é obtido por componentes principais. O objetivo é encontrar um conjunto de valores para os pesos C_{ij} que melhor expliquem a variância das variáveis originais e diferenciem grupos de variáveis de outros. Esse conjunto é composto pelos pesos ou carregamentos das variáveis originais no fator. Assim, encontra-se o primeiro fator principal e os demais $k < p$ fatores, na sequência é subtraída da matriz de correlação a variância explicada pelos fatores resultando em uma matriz chamada de matriz residual. Este procedimento permite extrair fatores que não são correlacionados, chamados então de ortogonais.

4º) Rotação: Para interpretar os resultados da Análise Fatorial utiliza-se um procedimento de rotação de fatores. Existem dois tipos de rotação fatorial: a rotação ortogonal, que pode ser executada pelo método Varimax na qual os eixos são mantidos a 90º,

consequentemente os fatores são não correlacionados; e a rotação oblíqua, que torna os fatores correlacionados entre si. Ambos os métodos de rotação tem o objetivo de simplificar as linhas e colunas da matriz fatorial para facilitar a interpretação. (HAIR et al., 2005).

2.4.2 Análise Fatorial pela Matriz de Correlação

Seja o vetor aleatório $\underline{X}' = [X_1, X_2, \dots, X_p]$, com p componentes e com $E(\underline{X}) = \underline{\mu} = [\mu_1, \mu_2, \dots, \mu_n]'$ e matriz de covariância $V(\underline{X}) = \Sigma_{p \times p}$. O vetor $\underline{X}'$ é linearmente dependente sobre algumas variáveis não observáveis (latentes) $F_1, F_2, \dots, F_m$ ($m < p$), chamados fatores comuns e p fontes de variação aditivas, $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$, chamadas de erros ou fatores específicos. Cada variável padronizada no modelo fatorial pode ser escrita matricialmente como decrito por Johnson e Wichern (1998):

$$\underline{X}_{p \times 1} = L_{p \times m} \cdot \underline{F}_{m \times 1} + \underline{\varepsilon}_{p \times 1} \quad (10)$$

onde L é a matriz de pesos ou carregamentos de ordem $p \times m$, $\underline{F}$ é o vetor de fatores extraídos de dimensão m e $\underline{\varepsilon}$ é o vetor de erros de dimensão p . E, de forma detalhada para as variáveis tem-se:

$$\begin{aligned} X_1 &= \ell_{11}F_1 + \ell_{12}F_2 + \dots + \ell_{1m}F_m + \varepsilon_1 \\ X_2 &= \ell_{21}F_1 + \ell_{22}F_2 + \dots + \ell_{2m}F_m + \varepsilon_2 \\ &\vdots \\ X_i &= \ell_{i1}F_1 + \ell_{i2}F_2 + \dots + \ell_{im}F_m + \varepsilon_i \\ &\vdots \\ X_p &= \ell_{p1}F_1 + \ell_{p2}F_2 + \dots + \ell_{pm}F_m + \varepsilon_p \end{aligned}$$

Na forma matricial tem-se o vetor de fatores $\underline{X}$:

$$\underline{X} = L \cdot \underline{F} + \underline{\varepsilon} = \begin{bmatrix} \ell_{11} & \ell_{12} & \dots & \ell_{1m} \\ \ell_{21} & \ell_{22} & \dots & \ell_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \ell_{p1} & \ell_{p2} & \dots & \ell_{pm} \end{bmatrix} \cdot \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix} \quad (11)$$

onde:

ℓ_{ij} : loading (ou peso) da variável aleatória X_i no fator X_j , para $i = 1, 2, \dots, p$ e $j = 1, 2, \dots, m$.

$L = \begin{bmatrix} \ell_{11} & \dots & \ell_{1m} \\ \vdots & \dots & \vdots \\ \ell_{p1} & \dots & \ell_{pm} \end{bmatrix}$ - matriz de pesos ou cargas fatoriais;

$\underline{F}' = [F_1 \ F_2 \ \dots \ F_m]$ - vetor de variáveis aleatórias não observáveis (latentes) chamados fatores comuns extraídos;

$\underline{\varepsilon}' = [\varepsilon_1 \ \varepsilon_2 \ \dots \ \varepsilon_m]$ - vetor de variáveis aleatórias não observáveis chamadas fatores específicos (erros).

O fator específico ε_i está associado apenas com a variável X_i . As p variáveis X_i são expressas em termos de $m + p$ variáveis não observáveis $[F_1, F_2, \dots, F_m]$ e $[\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p]$.

2.4.3 Modelo Fatorial Ortogonal

Segundo Johnson e Wichern (1998) deve-se assumir no modelo fatorial, $\underline{X} = L \cdot \underline{F} + \underline{\varepsilon}$, que $E(\underline{F}') = \underline{0}$, $V(\underline{F} \cdot \underline{F}') = I_m$, $E(\underline{\varepsilon}) = \underline{0}$, $V(\underline{\varepsilon}) = E(\underline{\varepsilon} \underline{\varepsilon}') = \Psi$ de ordem $p \times p$ e que $\underline{F}$ e $\underline{\varepsilon}$ são independentes, assim $cov(\underline{\varepsilon}, \underline{F}) = E(\underline{\varepsilon}, \underline{F}) = 0$ com $m = p$. E, com essas suposições, o modelo é chamado de Modelo Fatorial Ortogonal. Já o vetor de variáveis $\underline{X} = L \cdot \underline{F} + \underline{\varepsilon}$ tem matriz de covariâncias obtida da expressão:

$$V(\underline{X}) = V(L \cdot \underline{F} + \underline{\varepsilon}) = L \cdot V(\underline{F}) \cdot L' + V(\underline{\varepsilon}) = L \cdot I \cdot L' + \Psi = LL' + \Psi \quad (12)$$

$$E(\underline{\varepsilon}) = \underline{0}_{p \times 1}, \quad V(\underline{\varepsilon}) = E(\underline{\varepsilon} \underline{\varepsilon}') = \Psi_{p \times p} = \begin{bmatrix} \Psi_1 & 0 & 0 & \dots & 0 \\ 0 & \Psi_2 & 0 & \dots & 0 \\ 0 & 0 & \Psi_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \Psi_p \end{bmatrix}$$

Desta forma tem-se para a variável X_i o modelo:

$$X_i = \ell_{i1} \cdot F_1 + \ell_{i2} \cdot F_2 + \dots + \ell_{im} \cdot F_m + \varepsilon_i \quad (13)$$

E, a variância de cada variável X_i é:

$$V(X_i) = \ell_{i1}^2 + \ell_{i2}^2 + \dots + \ell_{im}^2 + \Psi_i, \quad i = 1, 2, \dots, p. \quad (14)$$

Ainda, com Ψ em (12) sendo a matriz diagonal das variâncias específicas e a matriz LL' tem na diagonal principal as communalidades das variáveis aleatórias, ou seja, a parcela da variância correspondente a contribuição dos fatores. Portanto, tem-se que a variância de cada variável X_i é dada por:

$$\sigma_i^2 = h_i^2 + \Psi_i \quad (15)$$

onde h_i^2 é a communalidade que expressa o quanto da variabilidade de X_i é advinda dos fatores; e Ψ_i é a variância específica que expressa o quanto da variabilidade de X_i não é explicada pelo modelo, ou seja, é a variância pura.

Um bom modelo deve apresentar uma communalidade alta para cada variável e, em correspondência, baixa variância específica. Assim, a communalidade representa a importância de cada variável no modelo e variáveis com baixa communalidade e consequentemente alta variância específica têm pouca relação com o fenômeno em estudo e podem ser descartadas.

A estimação dos pesos dos fatores é feita a partir de um conjunto de observações do vetor de variáveis originais $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$, padronizadas ou não. E, quando não se faz a padronização trabalha-se com a matriz de covariância Σ que resume o relacionamento entre as p variáveis originais. Mas, trabalhando-se com dados padronizados tem-se em correspondência esse relacionamento resumido pela matriz de correlação ρ , o que é satisfatório em razão de se evitar o problema de escalas diferentes para as variáveis.

Dentre os métodos de estimação dos fatores pode-se destacar:

- Método das Componentes Principais;
- Método da Máxima Verossimilhança;
- Método dos Mínimos Quadrados;
- Principal-axes Factoring;
- Método Alfa.

Johnson e Wichern (1998) consideram o Método das Componentes Principais e o Método da Máxima Verossimilhança como os mais recomendados para a análise fatorial. O Método da Máxima Verossimilhança, supõe que os dados seguem uma distribuição Normal Multivariada. E no Método das Componentes Principais, não há necessidade da restrição de normalidade. Esse método é baseado no Teorema da Decomposição Espectral que garante que qualquer matriz simétrica Σ de ordem $p \times p$ pode ser escrita como $\Sigma = P\Lambda P' = \sum_{i=1}^p \lambda_i \underline{e}_i \underline{e}_i'$ onde Λ é uma matriz diagonal formada com os autovalores de Σ e P é uma matriz ortogonal

($P'P = I$) cujas colunas são os autovetores normalizados (normalizados $\underline{e}_j' \underline{e}_j = 1$ e $\underline{e}_j' \underline{e}_k = 0, i \neq k$) de Σ .

2.4.4 Método da Componente Principal

A matriz de covariância do vetor de variáveis originais, $V(\underline{X}) = \Sigma$, pode ser fatorada pelo processo de expansão de matrizes simétricas, denominado Teorema da Decomposição Espectral. Dessa forma, Σ será formada por pares de autovalores e autovetores correspondentes normalizados, $(\lambda_i, \underline{e}_j)$, ou seja:

$$\Sigma = \lambda_1 \underline{e}_1 \underline{e}_1' + \lambda_2 \underline{e}_2 \underline{e}_2' + \dots + \lambda_p \underline{e}_p \underline{e}_p' = \begin{bmatrix} \sqrt{\lambda_1} \underline{e}_1 & \sqrt{\lambda_2} \underline{e}_2 & \dots & \sqrt{\lambda_p} \underline{e}_p \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} \underline{e}_1' \\ \sqrt{\lambda_2} \underline{e}_2' \\ \vdots \\ \sqrt{\lambda_p} \underline{e}_p' \end{bmatrix} \quad (16)$$

Os autovalores (λ) e os autovetores ($\underline{e}$) da matriz de covariâncias ou de correlações são a essência da Análise de Componentes Principais. Os autovetores determinam as direções de variabilidade máxima e os autovalores especificam as variâncias.

Se tomar o modelo da Análise Fatorial com $m = p$ fatores e os erros nulos, $e_j = 0$ para todo j , então a matriz de covariâncias com $m = p$ é dada por:

$$\Sigma_{p \times p} = LL'_{p \times p} + 0_{p \times p} = LL'_{p \times p} \quad (17)$$

A representação (17), considerando os p fatores não é muito útil, pois são preferíveis modelos que expliquem a estrutura de covariância levando em conta apenas os $m < p$ fatores comuns mais importantes. Essa aproximação, quando os últimos $p - m$ autovalores são insignificantes ou pequenos, omite a contribuição dos termos $\lambda_{m+1} \cdot \underline{e}_{m+1} \cdot \underline{e}_{m+1}' + \dots + \lambda_p \cdot \underline{e}_p \cdot \underline{e}_p'$ para a matriz Σ (JOHNSON e WICHERN, 1998). Omitindo essa contribuição, obtém-se a aproximação:

$$\Sigma \approx \begin{bmatrix} \sqrt{\lambda_1} \underline{e}_1 & \sqrt{\lambda_2} \underline{e}_2 & \dots & \sqrt{\lambda_p} \underline{e}_p \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} \underline{e}_1' \\ \sqrt{\lambda_2} \underline{e}_2' \\ \vdots \\ \sqrt{\lambda_p} \underline{e}_p' \end{bmatrix} = \begin{matrix} LL' \\ (p \times p) \end{matrix} \quad (18)$$

O que corresponde a expressão (17), omitindo-se Ψ . E, a matriz de variâncias específicas (Ψ), pode ser estimada por $\hat{\Psi} = \Sigma - \hat{L}\hat{L}'$, onde a matriz de variâncias específicas é:

$$\hat{\Psi} = \begin{bmatrix} \hat{\Psi}_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \hat{\Psi}_p \end{bmatrix} \text{ com } \hat{\Psi}_{ij} = s_{ii} - \sum_{i=1}^m \hat{\ell}_{ij}^2 \quad (19)$$

Quando se trabalha com a matriz de correlação, ρ , a variância estimada é $\rho_{ii} = 1$. Já as comunalidades estimadas são obtidas da diagonal do produto matricial $\hat{L}\hat{L}'$, ou seja,

$$\hat{h} = \hat{\ell}_{i1}^2 + \hat{\ell}_{i2}^2 + \dots + \hat{\ell}_{im}^2 = \sum_{j=1}^m \hat{\ell}_{ij}^2 \quad (20)$$

As comunalidades são quantidades das variâncias de cada variável explicada pelos fatores. Quanto maior o valor de $\hat{h}$, maior será o poder de explicação daquela variável pelo fator.

2.4.5 Estimação do Número de Fatores

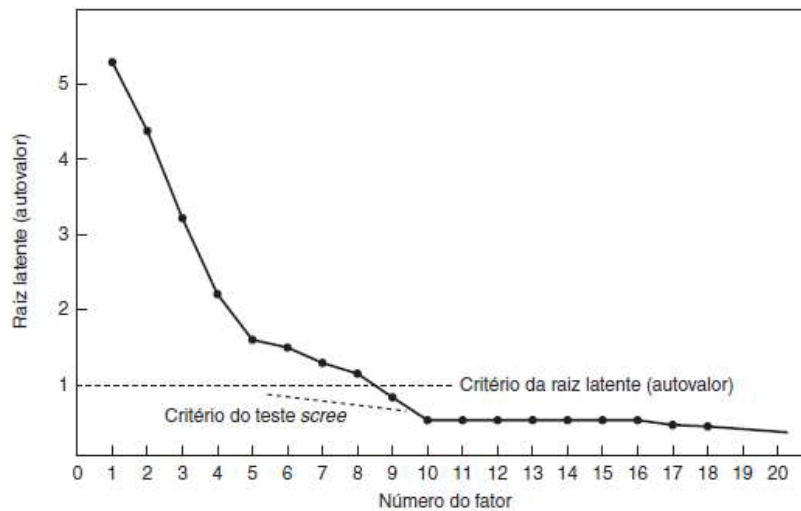
A estimação do número de fatores pode ser feita usando-se o critério de Kaiser, também chamado de critério da “raiz latente”. O critério de Kaiser propõe considerar apenas os autovalores superiores a um, que seriam os valores estatisticamente significativos. Entretanto, nem todos os autovalores superiores a um tem significado evidente. (FLECK, 1998). E, por outro lado, alguns autovalores com magnitude inferior a 1 mas próximos de 1 podem explicar um valor razoável de variabilidade. Então, isto pode ser considerado. Assim, para determinar o número de fatores pode-se basear na proporção da variância amostral devido a cada fator, que é:

$$\frac{\hat{\lambda}_j}{s_{11} + s_{22} + \dots + s_{pp}}, \text{ para análise a partir de } S \quad (21)$$

$$\frac{\hat{\lambda}_j}{p}, \text{ para análise a partir de } \rho \quad (22)$$

O Teste *Scree* corresponde ao gráfico das raízes latentes (autovalores) em relação ao número de fatores em sua ordem de extração e a forma da curva resultante é usada para avaliar o ponto de corte. (HAIR, et al., 2005)

FIGURA 3 – GRÁFICO DE AUTOVALOR PARA O CRITÉRIO DO TESTE SCREE



FONTE: Adaptado de Hair, et al. (2005).

O ponto no qual o gráfico começa a ficar horizontal é considerado indicativo do número máximo de fatores a serem extraídos. No caso do gráfico da FIGURA 3, os primeiros 10 fatores se qualificariam. Mas, pelo Critério de Kaiser seriam escolhidos apenas 8 fatores.

2.4.6 Rotação Ortogonal – Critério Varimax

A necessidade de rotação dos fatores vem da dificuldade de interpretação dos mesmos. O giro dos fatores é para identificar os fatores que possuam variáveis com alta correlação e outros com variáveis de baixa correlação, ou seja, evidenciar os contrastes.

O Critério Varimax, proposto por Kaiser (1960), dá ênfase a simplificação das colunas da matriz de pesos (cargas fatoriais) de forma a obter uma estrutura mais simples, que evidencie os contrastes entre os grupos. Segundo Hair et al. (2005) o objetivo é maximizar a variação entre os pesos de cada componente principal. A simplificação máxima é conseguida se houver apenas cargas fatoriais próximas de + ou - 1 ou 0, facilitando a interpretação dos fatores.

O Critério Varimax define $\tilde{e}_{ij}^* = \tilde{e}_{ij} / \hat{h}_{ij}$ como sendo os coeficientes finais após a rotação corrigidas pela raiz quadrada das comunalidades. De modo que o procedimento Varimax seleciona a transformação ortogonal T que maximiza o valor de V , adiante.

$$V = \frac{1}{p} \sum_{j=1}^m \left[\sum_{i=1}^p \tilde{e}_{ij}^{*4} - \frac{1}{p} \left(\sum_{i=1}^p \tilde{e}_{ij}^{*4} \right)^2 \right] \quad (23)$$

em que se tem $i = 1, 2, \dots, p$ (variáveis) e $j = 1, 2, \dots, m$ (fatores).

Espera-se encontrar grupos de variáveis com coeficientes grandes (pesos) e grupos com coeficientes desprezíveis em outros grupos. Assim, as variáveis com pesos altos dominam o fator. Para se obter a rotação aplica-se uma transformação T a ser determinada, ou seja, $\hat{\underline{f}}_j^* = T' \hat{\underline{f}}_j$, $j = 1, 2, \dots, n$. Os pesos $\hat{e}_{ij}^*$ são multiplicados por $\hat{h}_{ij}$ tal que as comunalidades originais sejam preservadas (CHAVES NETO, 2004). Portanto, espera-se encontrar grupos com coeficientes de elevada magnitude como grupos de magnitude desprezível em cada coluna da matriz de cargas rotacionadas.

2.4.7 Escores Fatoriais

Os valores estimados dos fatores comuns, também chamados de escores fatoriais, podem ser necessários, visto que, frequentemente são utilizados para diagnósticos e também para realização de possíveis análises posteriores (CARVALHO, 2013). É de se salientar que os escores não são estimativas de parâmetros do sentido usual, mas sim estimativas do valor não observado de cada um dos fatores para cada indivíduo da amostra. Um escore f_{ij} , pode ser interpretado como uma estimativa do valor do fator F_i para o j -ésimo indivíduo da amostra, $i = 1, 2, \dots, m$ e $j = 1, 2, \dots, n$.

Mingoti (2005) destaca os Métodos dos Mínimos Quadrados e de Regressão para determinar os escores fatoriais.

2.4.7.1 Método dos Mínimos Quadrados

Segundo Johnson e Wichern (1998), o uso dos Mínimos Quadrados Ponderados para estimar os escores fatoriais foi sugerido por Bartlett. Mas apenas se, μ , L e Ψ forem conhecidos no modelo $\underline{X} - \underline{\mu} = \underline{L}\underline{F} + \underline{\varepsilon}$ e $V(\varepsilon_i) = \Psi_i$. Então, escolhe-se as estimativas $\hat{f}$ de f para minimizar a soma de quadrados dos erros ponderados pela recíproca da variância,

$$\sum_{i=1}^p \frac{\varepsilon_i^2}{\Psi_i} = \underline{\varepsilon}' \Psi^{-1} \underline{\varepsilon} = (\underline{X} - \underline{\mu} - \underline{LF})' \Psi^{-1} (\underline{X} - \underline{\mu} - \underline{LF}) \quad (24)$$

e, usando as estimativas $\hat{L}$, $\hat{\Psi}$ e $\hat{\mu} = \bar{x}$ como verdadeiros, obtem-se o j -ésimo escore fatorial,

$$\hat{f}_j = (\hat{L}' \hat{\Psi}^{-1} \hat{L})^{-1} \hat{L}' \hat{\Psi}^{-1} (\underline{x}_j - \bar{x}), \quad j = 1, 2, \dots, n \quad (25)$$

em que:

$\hat{f}_j$: são os escores fatoriais estimados;

$\hat{L}$: é a matriz de cargas fatoriais estimada;

$\hat{\Psi}$: é a matriz de variâncias específicas estimada.

E, se a análise é feita a partir da matriz de correlação R , tem-se:

$$\hat{f}_j = (\hat{L}'_z \hat{\Psi}_z^{-1} \hat{L}_z)^{-1} \hat{L}'_z \hat{\Psi}_z^{-1} \underline{z}_j, \quad j = 1, 2, \dots, n \quad (26)$$

Mas, se as cargas fatoriais forem estimadas pelo método das componentes principais, os escores não são obtidos por mínimos quadrados ponderados e sim por mínimos quadrados ordinários (MQO). Neste caso, é suposto que as variâncias específicas Ψ_i , são iguais ou quase iguais. Os escores são dados por:

$$\hat{f}_j = (\hat{L}' \hat{L})^{-1} \hat{L}' (\underline{x}_j - \bar{x}), \quad j = 1, 2, \dots, n \quad (27)$$

E, ainda, se a análise é feita a partir de matriz de correlação ρ tem-se:

$$\hat{f}_j = (\hat{L}' \hat{L})^{-1} \hat{L}' \underline{z}_j, \quad j = 1, 2, \dots, n \quad (28)$$

2.4.8 Suposição da Análise Fatorial

A normalidade é necessária quando um teste estatístico deve ser aplicado para revelar a significância dos fatores. Uma estrutura de relacionamento entre as variáveis é necessária (multicolinearidade) para se empregar a Análise Fatorial. Um modo de determinar a adequação

dos dados à Análise Fatorial pode ser feita pelos critérios de Kaiser-Meyer-Olkin (KMO) e pelo teste de Esfericidade de Bartlett (DZIUBAN e SHIRKEY, 1974).

2.4.8.1 Critério de Kaiser-Meyer-Olkin (KMO)

De acordo com Maroco (2007) o critério KMO é uma medida da homogeneidade entre as variáveis. Assim, compara as correlações simples com as correlações parciais observadas entre as variáveis. Em outras palavras, o critério de Kaiser-Meyer-Olkin (KMO) identifica se o modelo de Análise Fatorial será adequadamente ajustado aos dados. Esta medida é dada pela expressão:

$$KMO = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p a_{ij}^2} \quad (29)$$

em que:

r_{ij} : é o coeficiente de variação simples entre X_i e X_j ;

a_{ij} : é o coeficiente de correlação parcial entre X_i e X_j .

O valor da estatística *KMO* varia de 0 a 1 e avalia a adequação da amostra de dados quanto ao grau de correlação parcial entre as variáveis. Segundo Chaves Neto (2004) pequenos valores de KMO, indicam que o uso da análise fatorial não é adequada, quanto mais próximo de 1, mais adequado se torna a aplicação da Análise Fatorial aos dados. F’avero et al. (2009), classifica os resultados da estatística KMO conforme o quadro a seguir:

TABELA 1 – CLASSIFICAÇÃO DA ESTATÍSTICA KMO

KMO	Classificação
0,90 – 1,00	Excelente
0,80 – 0,90	Ótimo
0,70 – 0,80	Bom
0,60 – 0,70	Regular
0,50 – 0,60	Ruim
0,0 – 0,50	Inadequado

FONTE: F’avero et al. (2009)

Conforme Kaiser (1970) a matriz anti-imagem de correlação é obtida por:

$$MSA = (diag R^{-1})^{-1/2} R^{-1} (diag R^{-1})^{-1/2} \quad (30)$$

em que R é a matriz de correlação. Para Hair et al. (2005), valores de MSA entre 0,5 e 1 significa que a Análise Fatorial é adequada, caso contrário, a análise fatorial não é adequada.

Outra forma de determinar se a análise fatorial é adequada aos dados é pelo teste de esfericidade de Bartlett.

2.4.8.2 Teste de Esfericidade de Bartlett

De acordo com Hair et al. (2005), o teste fornece a probabilidade de que a matriz de correlação tenha correlações significantes entre pelo menos algumas variáveis. As hipóteses testadas no teste são:

H_0 : a matriz de correlação é uma matriz identidade

H_1 : a matriz de correlações não é uma matriz identidade.

A hipótese nula a ser testada diz que a matriz de correlação é uma matriz identidade. E, isto significa que o modelo fatorial não é apropriado. A estatística do teste tem uma distribuição Qui-quadrado com $v = \frac{p(p-1)}{2}$ graus de liberdade:

$$\chi^2 = - \left[(n-1) - \frac{2p+5}{6} \right] \ln|R| \quad (31)$$

onde:

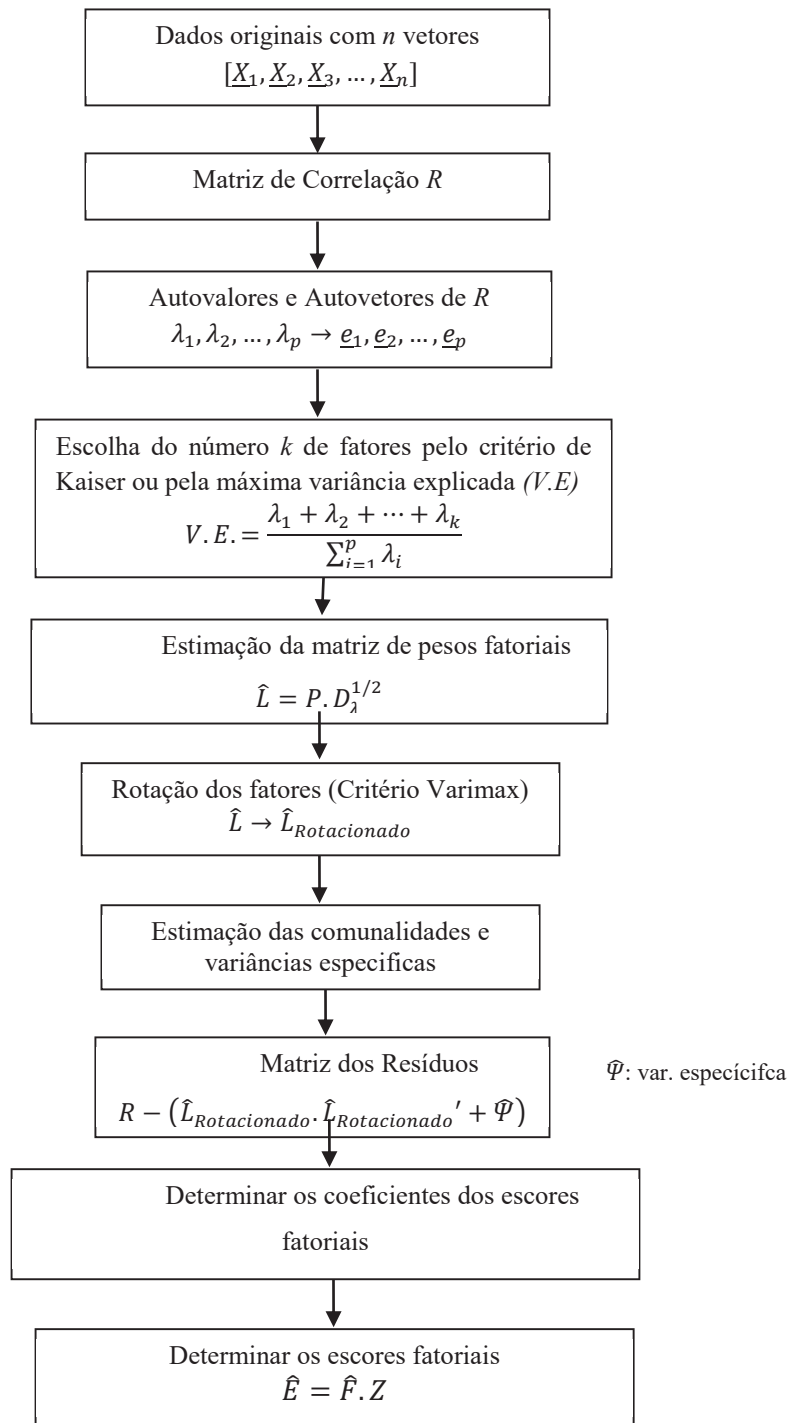
n : é o tamanho da amostra

p : é o número de variáveis

$|R|$: é o determinante da matriz de correlação.

O fluxograma a seguir, FIGURA 4, mostra os passos do algoritmo para Análise Fatorial com extração dos carregamentos fatoriais por Componentes Principais e Rotação Varimax.

FIGURA 4 – FLUXOGRAMA DA ANÁLISE FATORIAL POR COMPONENTES PRINCIPAIS E ROTAÇÃO VARIMAX



FONTE: O autor (2021).

2.5 ANÁLISE DE CORRELAÇÃO CANÔNICA

A Análise de Correlação Canônica é uma técnica de Análise Multivariada que procura identificar e quantificar a relação entre dois grupos de variáveis. Esta técnica pode ser vista, segundo Hair et al. (2005), como uma extensão lógica da Análise de Regressão Múltipla, sendo este último, um dos modelos mais utilizados para se analisar dois conjuntos de dados.

Na Regressão Múltipla, um dos conjuntos é composto por apenas uma variável, enquanto o outro conjunto é composto por várias variáveis. Assim, esta variável única é denominada variável dependente e está em função das $(p-1)$ variáveis independentes do outro conjunto por meio de uma combinação linear. Denota-se o modelo de Regressão Múltipla por:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_{p-1} X_{p-1i} + \varepsilon_i \quad (32)$$

onde $\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_{p-1} X_{p-1i}$ é a parte sistemática do modelo e ε_i ($i = 1, \dots, n$) é a parte estocástica.

Na forma matricial o mesmo modelo é descrito por:

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \quad (33)$$

onde:

$\underline{Y}$: é o vetor resposta de dimensão n ;

X : é a matriz do modelo de ordem $n \times p$;

$\underline{\beta}$: é o vetor de parâmetros de dimensão p ;

$\underline{\varepsilon}$: é o vetor de erros de dimensão n .

A solução da equação (33) é a combinação linear $X\underline{\beta}$ que é altamente correlacionada com $\underline{Y}$.

A Correlação Canônica avalia a associação entre dois conjuntos de variáveis distintas. Como é bem conhecido, a avaliação da associação entre duas variáveis pode ser feita pela Correlação Linear de Pearson ou outra medida adequada e a avaliação da associação entre uma variável e um conjunto de variáveis pode ser feita pelo Coeficiente de Correlação Múltipla ao Quadrado. Mas, quando se tem dois conjuntos de variáveis, o indicado é aplicar-se a Correlação Canônica. A idéia fundamental é desenvolver uma combinação linear em cada um dos conjuntos de variáveis, não existindo distinção entre variável dependente e independente, neste

caso, existem somente dois conjuntos de variáveis em que se busca a máxima correlação entre ambos (MORRISON, 1976).

Sejam dois grupos de variáveis $[X_1, X_2, \dots, X_p]$ e $[Y_1, Y_2, \dots, Y_q]$, que são vetores de dimensão p e q (assume-se que $p < q$). Portanto, tem-se p características (variáveis) no grupo A e q características do grupo B, respectivamente. O problema estatístico consiste em estimar a máxima correlação entre combinações lineares de características do grupo A e do grupo B, bem como estimar os respectivos coeficientes de ponderação das características em cada combinação linear.

Sendo $\underline{X}$ e $\underline{Y}$ uma das combinações lineares das variáveis pertencentes aos dois grupos, procura-se os vetores $\underline{c}_1$ e $\underline{c}_2$ para os quais a correlação entre as combinações lineares $\underline{c}_1 \underline{X}$ e $\underline{c}_2 \underline{Y}$ seja máxima. A forma de expressar as variáveis canônicas é pela combinação linear entre $\underline{X}$ e $\underline{Y}$, ou seja,

$$\begin{cases} U = c_{11}x_1 + c_{12}x_2 + \dots + c_{1p}x_p \\ V = c_{21}y_1 + c_{22}y_2 + \dots + c_{2q}y_q \end{cases} \quad (34)$$

Os pesos canônicos $c_{11}, c_{12}, \dots, c_{1p}$ e $c_{21}, c_{22}, \dots, c_{2q}$ são escolhidos de forma a maximizar a correlação entre as variações canônicas U e V .

Seja $\underline{W}$ um vetor de dimensão $(p + q \times 1)$, o qual possui matriz de covariância Σ e média $\underline{\mu}$. Sejam os vetores $\underline{X}_{p \times 1}$ e $\underline{Y}_{q \times 1}$ definidos como sendo originados de uma partição original do vetor $\underline{W}$, representando um grupo com p variáveis e outro com q variáveis, respectivamente. Para o vetor aleatório $\underline{W}$, tem-se os seguintes resultados:

$$\underline{W}_{p+q \times 1} = \begin{bmatrix} \underline{X} \\ \underline{Y} \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \\ y_1 \\ y_2 \\ \vdots \\ y_q \end{bmatrix}, \quad (35)$$

cujas médias são:

$$\underline{\mu} = E(\underline{W}) = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \quad (36)$$

e matriz de covariância:

$$\Sigma = E(\underline{W} - \underline{\mu})(\underline{W} - \underline{\mu})' = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \quad (37)$$

Assim, para os vetores $\underline{X}_{p \times 1}$ e $\underline{Y}_{q \times 1}$ verifica-se que:

$$\begin{cases} E(\underline{X}) = \underline{\mu}_X & Cov(\underline{X}) = \Sigma_{11} \\ E(\underline{Y}) = \underline{\mu}_Y & Cov(\underline{Y}) = \Sigma_{22} \\ Cov(\underline{X}, \underline{Y}) = \Sigma_{21} = \Sigma'_{12} \end{cases} \quad (38)$$

As covariâncias entre pares de variáveis pertencentes aos dois grupos, estão contidas em Σ_{12} . Dessa forma, os $p \cdot q$ elementos de Σ_{12} medem a associação entre os dois grupos. Se os valores de p e q forem muito grandes a interpretação simultânea das covariâncias se tornam muito complicadas, a ideia então, é concentrar a atenção em poucas combinações lineares de variáveis pertencentes a $\underline{X}$ e $\underline{Y}$, ao invés de utilizar todas as $p \cdot q$ covariâncias de Σ_{12} (CRUZ e REGAZZI, 2014). Para isso tem-se:

$$\begin{cases} U = \underline{c}_1' \cdot \underline{X} \\ V = \underline{c}_2' \cdot \underline{Y} \end{cases} \quad (39)$$

sendo $\underline{c}_1$ e $\underline{c}_2$ vetores dos coeficientes dessas combinações. Assim,

$$\begin{cases} Var(U) = \underline{c}_1' \cdot Cov(\underline{X}) \cdot \underline{c}_1 = \underline{c}_1' \cdot \Sigma_{11} \cdot \underline{c}_1 \\ Var(V) = \underline{c}_2' \cdot Cov(\underline{Y}) \cdot \underline{c}_2 = \underline{c}_2' \cdot \Sigma_{22} \cdot \underline{c}_2 \\ Var(U, V) = \underline{c}_1' \cdot Cov(\underline{X}, \underline{Y}) \cdot \underline{c}_2 = \underline{c}_1' \cdot \Sigma_{12} \cdot \underline{c}_2 \end{cases} \quad (40)$$

O método encontra os vetores $\underline{c}_1$ e $\underline{c}_2$, que maximiza:

$$\begin{cases} \rho(U, V) = \frac{\underline{c}_1' \cdot \Sigma_{12} \cdot \underline{c}_2}{\sqrt{\underline{c}_1' \cdot \Sigma_{11} \cdot \underline{c}_1} \sqrt{\underline{c}_2' \cdot \Sigma_{22} \cdot \underline{c}_2}} \\ s. a. \quad \underline{c}_1' \cdot \Sigma_{11} \cdot \underline{c}_1 = \underline{c}_2' \cdot \Sigma_{22} \cdot \underline{c}_2 = 1 \end{cases} \quad (41)$$

Existirão $\min(p, q) - 1$ pares de variáveis canônicas independentes do par de correlação máxima, que irão expressar a variância total dos dois grupos de variáveis (JOHNSON e WICHERN, 1992). Para o cálculo da correlação máxima $\max_{a,b} Corr(U, V) = \rho_1$, restrita pela combinação linear (primeiro par de variáveis canônicas):

$$U_1 = \underbrace{\underline{e}_1' \Sigma_{11}^{-1/2}}_{\underline{c}_1'} \underline{X} \quad e \quad V_1 = \underbrace{\underline{f}_1' \Sigma_{22}^{-1/2}}_{\underline{c}_2'} \underline{Y}$$

Nesse caso, ρ_1^2 é o autovalor de $\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1/2}$ e $\underline{e}_1$ seu respectivo autovetor. Assim, o primeiro par de variáveis canônicas será definido pelo par de combinações lineares U_1 e V_1 com variância unitária que maximizará (41). Da mesma forma, o k -ésimo par seguinte de variáveis canônicas é o par de combinações lineares U_k e V_k , tendo variâncias unitárias, que também maximiza (41) entre todas as escolhas que sejam não correlacionadas com as $k - 1$ variáveis canônicas anteriores.

Assim, sejam os vetores $\underline{X}$ e $\underline{Y}$ de dimensão p e q e com matrizes de covariância Σ_1 e Σ_2 , respectivamente, e covariância cruzada Σ_{12} , então as combinações lineares em U e V que maximizam a função objetivo (41), são dadas por:

$$\underline{c}'_1 = \underline{e}'_1 \Sigma_1^{-1/2} \quad \text{e} \quad \underline{c}'_2 = \underline{f}'_1 \Sigma_2^{-1/2},$$

onde $\underline{e}_1$ é o autovetor correspondente ao maior autovalor ρ_1^{*2} de $\Sigma_1^{-1/2}\Sigma_{12}\Sigma_2^{-1}\Sigma_{21}\Sigma_1^{-1/2}$ que tem p autovalores $\rho_1^{*2} \geq \rho_2^{*2} \geq \dots \geq \rho_p^{*2}$ e $\underline{e}_1, \underline{e}_2, \dots, \underline{e}_p$ autovetores e $\underline{f}_1$ é o autovetor correspondente ao maior autovalor de $\Sigma_2^{-1/2}\Sigma_{12}\Sigma_1^{-1}\Sigma_{21}\Sigma_2^{-1/2}$ que tem $\underline{f}_1, \underline{f}_2, \dots, \underline{f}_q$ autovetores correspondentes aos autovalores $\rho_1^{*2} \geq \rho_2^{*2} \geq \dots \geq \rho_q^{*2}$.

Segundo Johnson e Wichern (1992), o teste de significância para todas as correlações canônicas é:

$$H_0: \rho_1 = \rho_2 = \dots = \rho_s = 0, s = \min\{p, q\}$$

$$H_1: \rho_i \neq 0, \quad 1 \leq i \leq s$$

A estatística do teste segue a distribuição $\chi^2_{(\alpha, pq)}$:

$$\chi^2 = -[n - 1 - 0,5 \cdot (p + q + 1)] \ell n \left[\prod_{i=1}^s (1 - r_i^2) \right], \quad (42)$$

onde:

n : número de observações;

p : número de variáveis do grupo $[X_1 \quad X_2 \quad \dots \quad X_p]$;

q : número de variáveis do grupo $[Y_1 \quad Y_2 \quad \dots \quad Y_q]$;

r_i^2 : correlação canônica, ao quadrado, da equação a ser testada;

Na especificação dos objetivos da análise devem ser identificados os dois grupos de variáveis de que a análise canônica tratará, determinando quais objetivos se espera alcançar com essa análise (BASNIAK, 2009). A correlação canônica pode ser utilizada quando se quer

identificar e quantificar a relação entre dois grupos de variáveis podendo em uma análise subsequente substituir um grupo grande de variáveis por um grupo menor de variáveis, sem grande perda de informação, caso a correlação entre os dois grupos seja alta.

2.6 TÉCNICAS DE AGRUPAMENTO

Técnicas de Agrupamento ou de Cluster são técnicas de classificação que agrupam n itens em k grupos (clusters) distintos de forma que os elementos de um mesmo grupo sejam homogêneos e os elementos em grupos diferentes sejam heterogêneos, de acordo com uma medida de proximidade (distância) ou de similaridade.

Existem vários métodos de clusterização, o primeiro registro foi em 1948 com o trabalho de Sorensen (1948) sobre o método hierárquico de ligação completa. Os métodos hierárquicos envolvem a construção de uma hierarquia em uma estrutura do tipo árvore.

Todos os métodos de clusterização são fundamentados na ideia de distância ou similaridade entre as observações e definem a pertinência de cada item ao cluster de acordo com o que o item tem de similar aos outros itens do grupo.

2.6.1 Medidas de Distância e Similaridade

Similaridade entre os itens é uma medida empírica de semelhança ou de proximidade entre os elementos a serem agrupados. A proximidade dos itens é usualmente indicada por uma distância.

Sejam os vetores aleatórios de dimensão p , $\underline{x}' = [x_1, x_2, \dots, x_p]$ e $\underline{y}' = [y_1, y_2, \dots, y_p]$, a distância entre $\underline{x}$ e $\underline{y}$ pode ser medida por:

a) Distância Euclidiana: $d(\underline{x}, \underline{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2}$

b) Distância de Mahalanobis: $d(\underline{x}, \underline{y}) = \sqrt{(\underline{x} - \underline{y})' A^{-1} (\underline{x} - \underline{y})}$, onde A é a matriz de variâncias e covariâncias amostrais.

c) Distância de Minkowsky: $d(\underline{x}, \underline{y}) = [\sum_{i=1}^p |x_i - y_i|^m]^{1/m}$, para $m = 2$ essa distância é a Euclidiana. Esta métrica é menos afetada pela presença de *outliers* do que a distância Euclidiana.

Dentre os algoritmos de agrupamento pode-se citar os hierárquicos e os não-hierárquicos. O método hierárquico fornece um gráfico da hierarquização chamado dendrograma. Nesses métodos hierárquicos são fornecidos vários agrupamentos possíveis e pelo dendrograma (diagrama em árvore) se define o número de clusters. Já nos não-hierárquicos é necessário definir o número de clusters a ser formado a priori.

Dentre os métodos não-hierárquicos o algoritmo das K-Médias é o mais utilizado. Esse método exige o conhecimento prévio do número de grupos a ser formado. Geralmente aplica-se um método hierárquico para definir o número de grupos a serem formados e a seguir usa-se esse número no método das K-Médias para obter um certo refinamento na classificação.

2.6.2 Algoritmo K-Médias

O algoritmo particional das K-Médias é uma técnica de clusterização do tipo não-hierárquico e é muitas vezes usado como um dos algoritmos mais simples de aprendizagem não supervisionada. O objetivo é definir k grupos iniciais, e em seguida, utilizar uma técnica de realocação iterativa baseada em similaridade, de forma a melhorar a posição dos centróides dentro de cada grupo (PIMENTEL et al., 2003).

Este algoritmo é muito utilizado devido a sua eficiência, simplicidade e facilidade de implementação. O método é baseado em centróide, que é o ponto geométrico médio obtido pela média dos valores nas coordenadas dos pontos de um cluster.

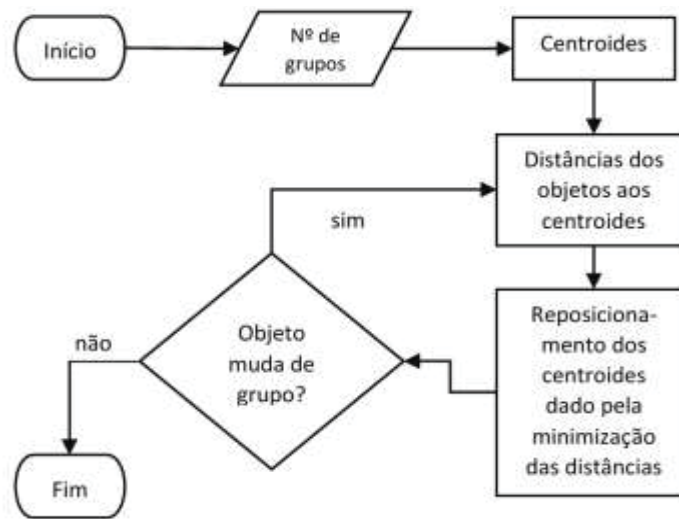
Chaves Neto (2004), descreve algoritmo das K-Médias pelos seguintes passos:

- 1) Partição dos itens em k grupos iniciais;
- 2) Prosseguir com a lista de itens, colocando cada item no grupo cuja média (centróide) está mais próximo, sendo o centróide recalculado pelo cálculo da distância (a, b ou c), sendo o centróide recalculado para o grupo que recebeu um novo item e para o grupo que perdeu o item;
- 3) Repete-se o segundo passo até que não restem realocações a serem feitas.

Na fase final da execução do K-Médias, as fases de “*atribuição ao grupo*” e “*recálculo dos centróides*”, são repetidas até o cluster se tornar estático, isso ocorre quando nenhum dos itens mudam de cluster.

Um fluxograma da execução do Algoritmo das K-Médias pode ser visto na FIGURA 5, a seguir:

FIGURA 5 – FLUXOGRAMA DE EXECUÇÃO DO ALGORITMO DAS K-MÉDIAS



FONTE: Adaptado de Domingues et al. (2018)

Um problema análogo ao das K-Médias, é conhecido como problema dos K-Medoids, sendo este uma variante do K-Médias. Neste problema, substitue-se os centróides do K-Médias por *medoids*, objetivando-se obter agrupamentos mais homogêneos e robustos. Este método foi proposto Kauffamn e Rousseeuw (2005), onde os clusters são representados usando a mediana dos dados em vez da média. Segundo Brito et al. (2010), o método encontra aplicação frequente na fase de estratificação em amostragem probabilística.

2.6.3 Mapas Auto-Organizáveis de Kohonen (SOM)

A Rede SOM (*Self Organizing Maps*) é uma Rede Neural que foi desenvolvida por Teuvo Kohonen na década de 1980 e desde então, o algoritmo original vem sofrendo modificações para atender a outros critérios não abordados inicialmente. A aplicação mais utilizada do SOM é o agrupamento (*clusters*), ou seja, o particionamento do conjunto de dados em subconjuntos de dados “semelhantes”, sem uso de conhecimento prévio sobre esses subconjuntos (KOHONEN, 1991).

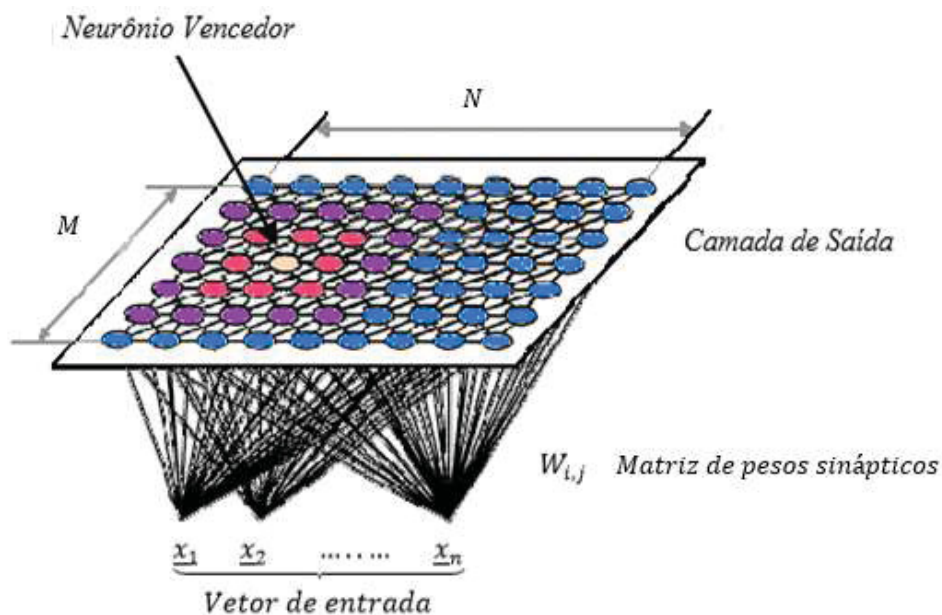
Uma das primeiras demonstrações de agrupamento foi proposta por Ritter e Kohonen (1989). Eles tinham uma lista de 16 animais (pássaros, predadores e presas) e 13 atributos binários para cada um deles (por exemplo, tamanho grande ou não, cabelo ou não, duas pernas ou não, pode voar ou não, ...). Depois de treinar uma rede (10×10) de neurônios com esses vetores e rotulando o neurônio vencedor para cada vetor de código animal, um agrupamento

natural de pássaros, de predadores e de presas apareceu no mapa. Os autores chamaram este mapa de "mapa semântico"

Assim como no método das K-Médias, este tipo de Rede Neural utiliza método de treinamento não-supervisionado. Isso significa, que o resultado é formado pelas propriedades inerentes aos dados em si; não se tem conhecimento prévio do que é verdadeiro ou falso. Portanto, nenhum conhecimento a respeito da estrutura dos dados é necessária para o uso do algoritmo. A própria Rede Neural busca encontrar similaridades, baseando-se apenas nos padrões de entrada (ULTSCH, VETTER, 1994).

O modelo de Rede proposto por Kohonen (1982) é composto por um Mapa Auto Organizável concebido através de duas camadas (camada de entrada e camada de saída), a conexão entre as duas camadas é feita através de vetores de pesos sinápticos $\underline{w}$ (FIGURA 6).

FIGURA 6 – ESTRUTURA DE UMA REDE DE KOHONEN



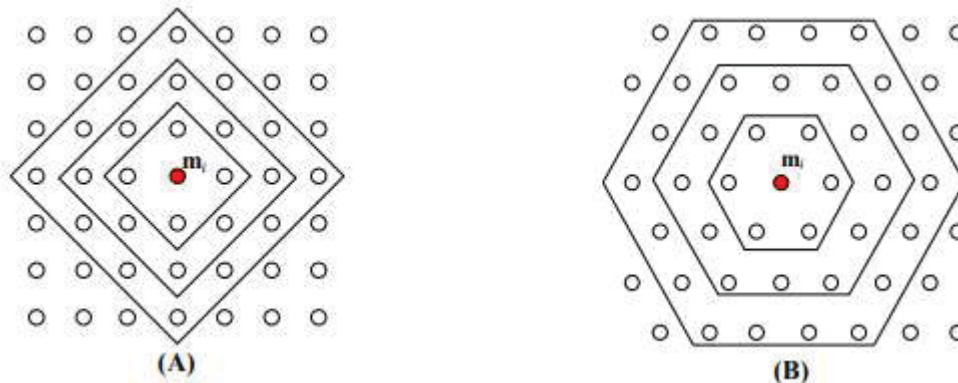
FONTE: Adaptado de Ferreira da Silva (2007)

A primeira camada é composta pelos dados de entrada $[\underline{x}_1 \quad \underline{x}_2 \quad \dots \quad \underline{x}_n]$, onde todos os vetores estão conectados aos neurônios, que estão na segunda camada, que por sua vez, está disposta em uma grade bidimensional. O conjunto de pesos sinápticos (W_{ij}) , são inicialmente gerados aleatoriamente para não formatar o mapa com uma tendência inicial de se agrupar, os valores dos pesos são gerados entre 0 e 1.

A camada ordenada de neurônios é dita competitiva e é gradualmente adaptada para responder seletivamente às entradas similares. Neste processo, a rede utiliza o lema “o vencedor

leva tudo” para reajustar os pesos e assim ensinar à rede um novo conceito. Os neurônios competem entre si para serem ativados e apenas o vencedor terá seus pesos atualizados caso as conexões laterais sejam nulas, caso contrário, seus vizinhos serão proporcionalmente ajustados. Este ajuste proporcional dos vizinhos ao neurônio vencedor irá depender da topologia da camada de saída, que pode ser hexagonal ou retangular. Estes parâmetros de vizinhança são ilustrados na FIGURA 7.

FIGURA 7 – EM (A) VÊ-SE A VIZINHANÇA TOPOLÓGICA RETANGULAR E EM (B) O ARRANJO COM VIZINHANÇA HEXAGONAL, ONDE m_i É O I-ÉSIMO NEURÔNIO



FONTE: Adaptado de Zuchini (2003)

A escolha do tipo de topologia pode influenciar no número de elementos em cada grupo. Considerando a forma hexagonal, esta permite que cada neurônio interaja com até seis vizinhos, enquanto que a retangular só permite a interação com no máximo quatro vizinhos (FIGURA 9), o que diminui a cooperação entre os neurônios, sendo que o modelo hexagonal oferece tradicionalmente resultados melhores que o retangular (RODRIGUES, 2009).

O algoritmo se divide em três etapas: competição, cooperação e adaptação sináptica. Na primeira etapa, um único elemento é declarado vencedor. Em seguida, são determinados os elementos que compõem a vizinhança do neurônio vencedor para ser associado um coeficiente de cooperação. Com base nesses dados, o neurônio vencedor recebe um incremento para correção da sua função de ativação e, conforme o fator de vizinhança, um percentual do incremento recebido pelo neurônio vencedor é aplicado aos neurônios vizinhos (HAYKIN, 1998). As etapas de competição, cooperação e adaptação sinápticas são detalhadas a seguir.

2.6.3.1 Processo Competitivo

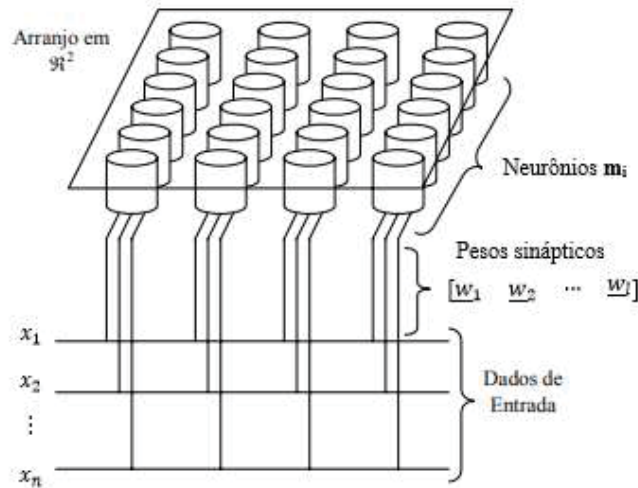
Seja n a dimensão do espaço dos dados de entrada. Considere que um vetor de entrada selecionado aleatoriamente do espaço de entrada seja representado por:

$$\underline{x} = [x_1, x_2, \dots, x_n]^T$$

Considere o vetor peso sináptico de cada neurônio i representado por:

$$\underline{w}_i = [w_1, w_2, \dots, w_n]^T \quad i = 1, 2, \dots, l.$$

FIGURA 8 – ESTRUTURA DA REDE DE KOHONEN DE DIMENSÃO 4×6



FONTE: Adaptado de Zuchini (2003)

onde l é número de neurônios na grade. Cada neurônio tem seu peso sináptico iniciado aleatoriamente. Para encontrar o melhor casamento do vetor de entrada $\underline{x}$ com os vetores de peso sináptico $\underline{w}_i$ deve-se calcular a distância (comumente a Euclidiana) entre o vetor $\underline{x}$ e os vetores $\underline{w}_i$ e escolher a menor.

$$d(\underline{w}_i, \underline{x}_k) = \sqrt{\sum_{j=1}^p [x_{kj}(t) - w_{ij}(t)]^2} \quad i = 1, \dots, m \quad (43)$$

O neurônio c que corresponder a menor distância será o neurônio vencedor. Este neurônio, c , é escolhido de acordo com a equação

$$\|\underline{x} - \underline{w}_c\| = \min_i \{\|\underline{x} - \underline{w}_i\|\} \quad (44)$$

onde $\| \cdot \|$ denota a distância obtida.

2.6.3.2 Processo Cooperativo

Um neurônio classificado como vencedor no processo competitivo, tende a excitar os neurônios em sua vizinhança imediata. Este neurônio determina o centro desta vizinhança topológica de neurônios cooperativos (HAYKIN, 1998).

Estes neurônios excitados, terão seus pesos ajustados juntamente com o neurônio vencedor. O neurônio vencedor dá a localização do centro de vizinhança topológica, definindo junto com o raio (R), quais serão os neurônios que irão cooperar. Se a escolha for, $R = 0$, somente o neurônio vencedor será atualizado, para $R = 1$, serão atualizados o neurônio vencedor e os neurônios excitados que totalizam 4 e 6, para a topologia retangular e hexagonal, respectivamente (FIGURA 7). Durante a fase de treinamento o valor de R deverá diminuir suavemente com o tempo. Na medida em que esta exigência é satisfeita faz com que a largura σ da função de vizinhança topológica $h_{ik}(t)$, diminua. Uma escolha recorrente de σ é o decaimento exponencial (HAYKIN, 1998):

$$\sigma(t) = \sigma(0) \exp\left(-\frac{t}{\tau_1}\right) \quad (45)$$

onde

t : é a iteração atual (número inteiro);

$\sigma(0)$: é o valor da largura inicial para $\sigma(t)$ (taxa de aprendizagem);

τ_1 : é a constante de tempo do SOM.

Para que a taxa de aprendizagem $\sigma(t)$ nunca caia para zero o valor de τ_1 é definido pela expressão (46).

$$\tau_1 = \frac{NIter}{\log \sigma(0)} \quad (46)$$

em que, $NIter$ é o número de iterações.

Após o processo de definição da vizinhança topológica, a próxima etapa realizará as modificações do vetor de peso sináptico.

Há diversas formas de implementar a função de vizinhança, uma escolha típica para a função de vizinhança que satisfaz a condição de convergência é a função Gaussiana, que é definida por:

$$h_{ik}(t) = \exp\left(-\frac{\|r_i - r_k\|}{2\sigma^2(t)}\right) \quad (47)$$

em que o termo $\|r_i - r_k\|$ é a distância entre o neurônio “ i ” vencedor e o neurônio “ k ” da vizinhança que está sendo atualizada. O parâmetro $\sigma(t)$ é definido pela equação (45) e define a largura da função que é decrescente no tempo.

2.6.3.3 Processo Adaptativo

A modificação dos pesos em relação a entrada se dá de forma iterativa, o neurônio vencedor e seus vizinhos serão atualizados usando a expressão:

$$\underline{w}_i(t+1) = \underline{w}_i(t) + \sigma(t)h_{ik}(t)\left(\underline{x}(t) - \underline{w}_i(t)\right) \quad (48)$$

onde:

$\underline{w}_i(t+1)$: é o vetor de peso atualizado;

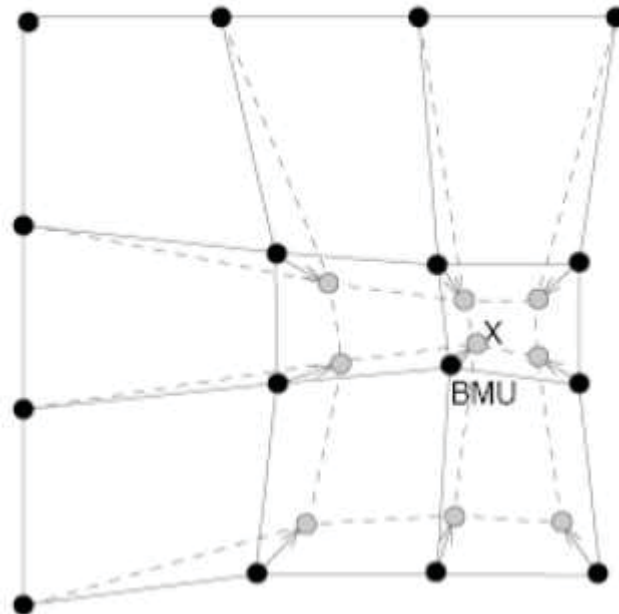
$\sigma(t)$: é a taxa de aprendizagem;

$h_{ik}(t)$: é a função de vizinhança;

$\underline{x}(t)$: é o vetor de entrada.

A equação (48) é aplicada a todos os neurônios da grade dentro da região de vizinhança R , com efeito de mover o vetor peso sináptico $\underline{w}_i$ do neurônio vencedor em direção ao vetor de entrada $\underline{x}$ (FIGURA 9).

FIGURA 9 – ATUALIZAÇÃO DO NEURÔNIO VENCEDOR E DA VIZINHANÇA TOPOLÓGICA EM TORNO DO VETOR DE ENTRADA



FONTE: Vesanto et al. (2000)

As linhas contínuas e pontilhadas correspondem respectivamente, a situação antes e depois da atualização (FIGURA 9), a sigla BMU (*Best Matching Unit*) representa o neurônio vencedor. Através da apresentação repetida dos dados de treinamento os vetores de pesos sinápticos tendem a acompanhá-los devido à atualização da vizinhança topológica, levando desta forma o mapa de características a uma ordenação topológica. Assim sendo, neurônios adjacentes na camada de saída terão vetores de pesos sinápticos similares (HAYKIN, 1998).

Na fase de convergência, ocorre o refinamento do mapa. Kohonen (2001) sugere um número de iterações na ordem de 500 vezes o número de neurônios da camada de saída, no entanto, esta não é a regra, na literatura encontra-se bons resultados obtidos com número menor de iterações ou então, maior.

Oliveira (2001) descreve os seguintes passos do algoritmo:

Passo 1 – Inicia-se a rede: pesos w_{ij} aleatoriamente (entre 0 e 1), raio de vizinhança R , taxa de aprendizagem $\sigma(0)$;

Passo 2 – Apresenta-se os dados de entrada X ;

Passo 3 – Calcula-se a distância Euclidiana do vetor de entrada selecionado a cada neurônio j de saída utilizando a equação (43);

Passo 4 – Seleciona-se o neurônio de menor distância obtido no passo 3;

Passo 5 – Atualizar os pesos sinápticos da unidade vencedora e de seus vizinhos usando a expressão (48);

Passo 6 – Atualizar a função de vizinhança $h_{ik}(t)$ e a taxa de aprendizagem $\sigma(t)$, com:

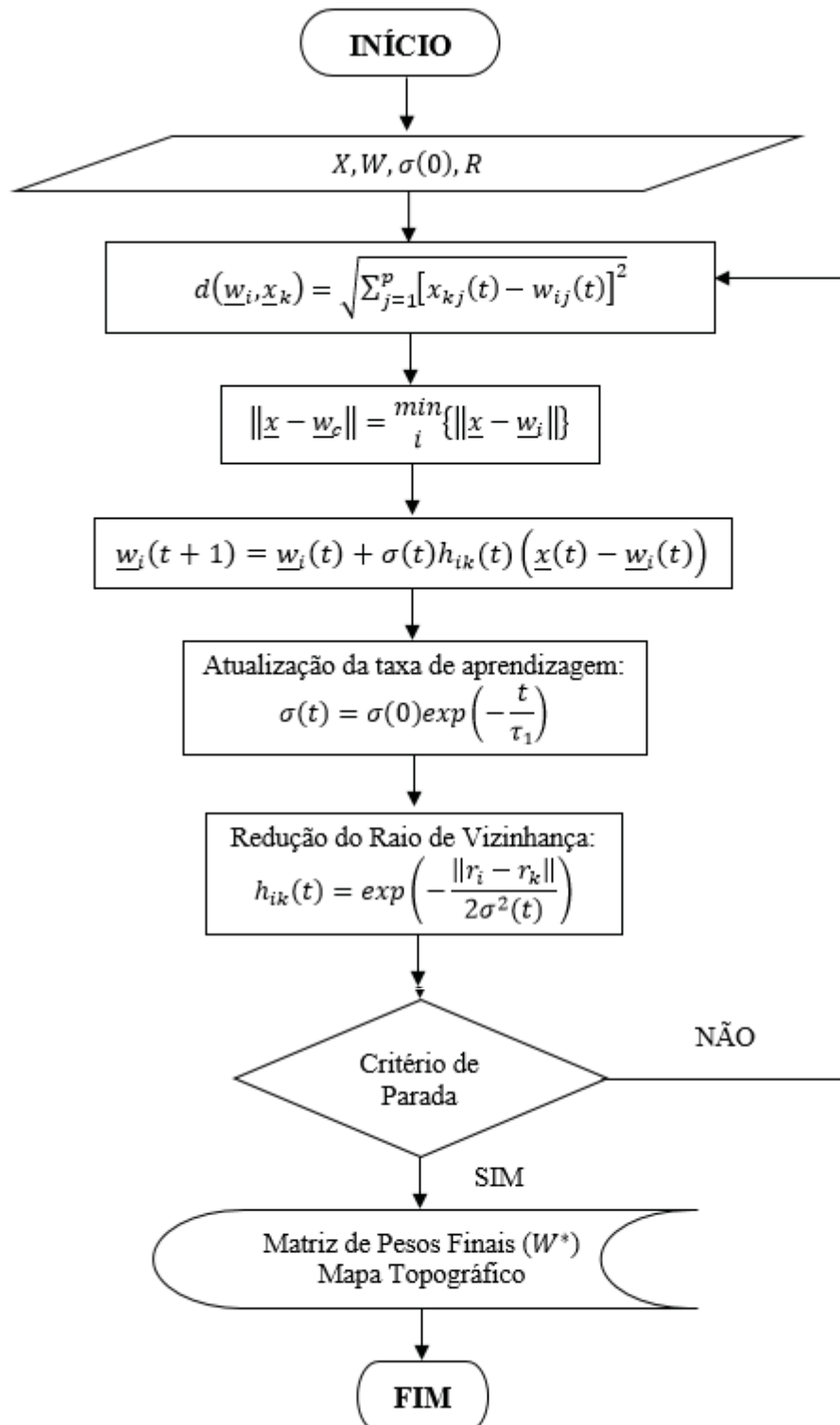
$$h_{ik}(t) = \exp\left(-\frac{\|r_i - r_k\|}{2\sigma^2(t)}\right)$$

e

$$\sigma(t) = \sigma(0)\exp\left(-\frac{t}{\tau_1}\right)$$

Passo 7 – Repetir a partir do passo 3 para cada um dos vetores do conjunto de treinamento até o número de épocas estimadas ou até os pesos se estabilizem.

FIGURA 10 – FLUXOGRAMA PARA O ALGORÍTMO QUE PRODUZ OS MAPAS AUTO-ORGANIZÁVEIS



FONTE: O autor (2021).

3 MATERIAL E MÉTODOS

3.1 MATERIAL

Para o desenvolvimento deste trabalho foi realizada a coleta de dados dos Programas de Pós-Graduação na Área de Engenharias III e, para isto, foram utilizados os dados relativos às atividades desenvolvidas pelos Programas durante os anos do quadriênio 2013-2016.

Para estudar as características dos PPG's, foram selecionadas 35 variáveis, as quais trazem informações sobre: Número de docentes e discentes, Artigos publicados em periódicos, Artigos publicados em anais de eventos, Produção de egressos, Quantidade de alunos titulados no período, entre outras.

Na TABELA 2 consta a descrição das 35 variáveis selecionadas para fazerem parte desse estudo.

TABELA 2 – VARIÁVEIS UTILIZADAS NO ESTUDO

Variável	Descrição	Variável	Descrição
X_1	Número de docentes permanentes do PPG	X_{19}	Artigos B1 completos publicados pelos docentes
X_2	Número de docentes colaboradores do PPG	X_{20}	Artigos B2 completos publicados pelos docentes
X_3	Número total de discentes	X_{21}	Artigos B3 completos publicados pelos docentes
X_4	Artigos A1 completos publicados pelos discentes	X_{22}	Artigos B4 completos publicados pelos docentes
X_5	Artigos A2 completos publicados pelos discentes	X_{23}	Artigos B5 completos publicados pelos docentes
X_6	Artigos B1 completos publicados pelos discentes	X_{24}	Artigos C completos publicados pelos docentes
X_7	Artigos B2 completos publicados pelos discentes	X_{25}	Artigos NP completos publicados pelos docentes
X_8	Artigos B3 completos publicados pelos discentes	X_{26}	Livros – Obra Completa (docentes)
X_9	Artigos B4 completos publicados pelos discentes	X_{27}	Livros – Capítulo (docentes)
X_{10}	Artigos B5 completos publicados pelos discentes	X_{28}	Anais em congresso (docentes)
X_{11}	Artigos C completos publicados pelos discentes	X_{29}	Produção técnica (docentes)
X_{12}	Artigos NP completos publicados pelos discentes	X_{30}	Artigos A1 ou A2 publicados pelos egressos do PPG
X_{13}	Livros – Obra Completa (discente)	X_{31}	Artigos B1 ou B2 publicados pelos egressos do PPG
X_{14}	Livros – Capítulo (discente)	X_{32}	Anais em congresso (egressos)
X_{15}	Anais em congresso (discente)	X_{33}	Produção técnica (egressos)
X_{16}	Produção técnica (discente)	X_{34}	Mestres titulados pelo PPG no quadriênio
X_{17}	Artigos A1 completos publicados pelos docentes	X_{35}	Doutores titulados pelo PPG no quadriênio
X_{18}	Artigos A2 completos publicados pelos docentes		

FONTE: O autor (2021)

Em relação às 35 variáveis, 23 delas trazem informações sobre a publicação de artigos em congresso e revistas especializadas, quatro variáveis correspondem à livros publicados, cinco variáveis referem-se à recursos humanos e três variáveis à produção técnica dos PPG's.

Algumas informações enviadas à CAPES pelos Programas não foram consideradas nesta pesquisa, como por exemplo, informações sobre objetivos e metas dos PPG's, visto que a análise é feita utilizando dados quantitativos apenas. Informações como tempo mediano para titulação, abandono de curso e quantidade de bolsistas também não foram consideradas pois não refletem a qualidade do PPG diretamente.

Os dados utilizados estão no Anexo B deste trabalho. A metodologia utilizada, a construção do algoritmo de execução e a aplicação serão detalhados nas próximas seções deste capítulo. Todos os procedimentos de cálculo, neste trabalho, foram feitos utilizando-se o Sistema Computacional R, cujos *scripts* estão disponíveis no APÊNDICE 1 e o Software *Statgraphics*.

3.2 MÉTODO DE DETERMINAÇÃO DO MODELO DE CLASSIFICAÇÃO

Para alcançar o objetivo proposto, criou-se um método híbrido de classificação utilizando os métodos de Análise Fatorial e Mapas Auto-Organizáveis (SOM) que se convencionou chamar de AF/SOM.

A aplicação da Análise Fatorial corresponde a primeira etapa da metodologia proposta em que são determinados os escores fatoriais para posterior aplicação do SOM. A Análise Fatorial é aplicada sob os dados de entrada da matriz X de ordem $n \times p$ e calcula-se a matriz de correlação $R_{p \times p}$ do vetor aleatório observado n vezes. Para a utilização da Análise Fatorial, é importante que as premissas para sua aplicação sejam observadas (Teste KMO, p.ex.).

Após a obtenção da matriz de correlação R calcula-se os p autovalores e autovetores da matriz R . Para determinar o número de fatores para a Análise Fatorial optou-se por utilizar o critério da percentagem da variância total explicada pelos k autovalores selecionados:

$$\frac{\hat{\lambda}_1 + \hat{\lambda}_2 + \dots + \hat{\lambda}_k}{\sum_{i=1}^p \lambda_i} \quad (49)$$

O número de fatores para esta aplicação é extraído de modo a explicar pelo menos 90% da variância total. A matriz de pesos fatoriais é obtida por Componentes Principais no qual, os modelos estão baseados na variância total:

$$\hat{L} = P \cdot D_{\lambda}^{1/2} \quad (50)$$

onde:

$\hat{L}$: é a Matriz de carregamentos/pesos fatoriais estimada;

P : é a Matriz de autovetores associados aos p autovalores;

$D_\lambda^{1/2}$: é a Matriz diagonal da raiz quadrada dos p autovalores.

Para identificar claramente quais variáveis pertencem a quais fatores aplica-se o método Varimax para rotação dos fatores (JOHNSON e WICHERN, 1998). E, com os fatores rotacionados de pesos compondo a matriz $(\hat{L}^*)$ obtém-se a matriz dos Escores Fatoriais $(\hat{F})$.

Após a verificação de forte correlação entre a matriz de dados originais e a matriz de escores fatoriais, esta, pode substituir a matriz de dados de entrada X , haja vista, que representa o mesmo conjunto de dados, mas com um número menor de variáveis.

A matriz de escores fatoriais será utilizada para o treinamento da Rede SOM. O objetivo disso, está no fato de após a aplicação da análise fatorial, ser conhecido o grau de importância/explicação de cada fator/variável, esta importância está na proporção da variância explicada por cada fator. Esta informação é fundamental para obter o índice classificatório para os clusters.

Para iniciar o treinamento da Rede SOM, primeiramente é preciso padronizar os dados de entrada $(\hat{F})$. Uma das razões para padronização deve-se ao fato de termos muitas vezes uma escala de dados discrepantes entre si.

A padronização ocorre em dois passos; primeiramente padronizamos os dados com média zero e variância um:

$$\hat{F} \rightarrow Z = \frac{E - \mu}{\sigma} \quad (51)$$

A segunda padronização é do tipo exponencial (equação 52), colocando os dados numa escala entre 0 e 1. Esta segunda padronização é necessária pois a matriz de pesos sinápticos da rede SOM é a princípio gerada por pesos aleatórios pertencentes ao intervalo $[0,1]$.

$$Z \rightarrow Z^* = \frac{1}{1 + e^{-x}} \quad (52)$$

Com a matriz Z^* obtida, inicia-se o treinamento da Rede SOM. A Rede é treinada definindo-se os seguintes parâmetros:

a) Dimensão da rede;

- b) Topologia: Hexagonal ou Retangular;
- c) Taxa de Aprendizagem;
- d) Número de Iterações;

O critério de parada será alcançado quando não houver mais alterações significativas na matriz de pesos dos neurônios treinada $W_{n \times p}^*$ (n : neurônios, p : fatores). Os pesos que formam esta matriz, representam o grau de importância que cada fator da análise fatorial tem dentro do cluster formado. Para obter o índice classificatório dos clusters, multiplica-se esta matriz W^* pelo vetor $\underline{VE}$ da variância explicada por cada fator.

$$\underline{CN} = W^* \cdot \underline{VE} \quad (53)$$

onde $\underline{CN}$ é o índice classificatório do neurônio/cluster, $\underline{VE}$ é o vetor das variâncias explicadas pelos p fatores e W^* é a matriz de pesos dos neurônios após o treinamento da rede neural.

Para a hierarquização dos itens pertencentes a cada cluster, utiliza-se os escores classificatórios da Análise Fatorial, estes escores são obtidos pela expressão:

$$Efp_j = \frac{\sum_{i=1}^k \hat{f}_j \cdot \hat{\lambda}_i}{\sum_{i=1}^k \hat{\lambda}_i} \quad (54)$$

onde:

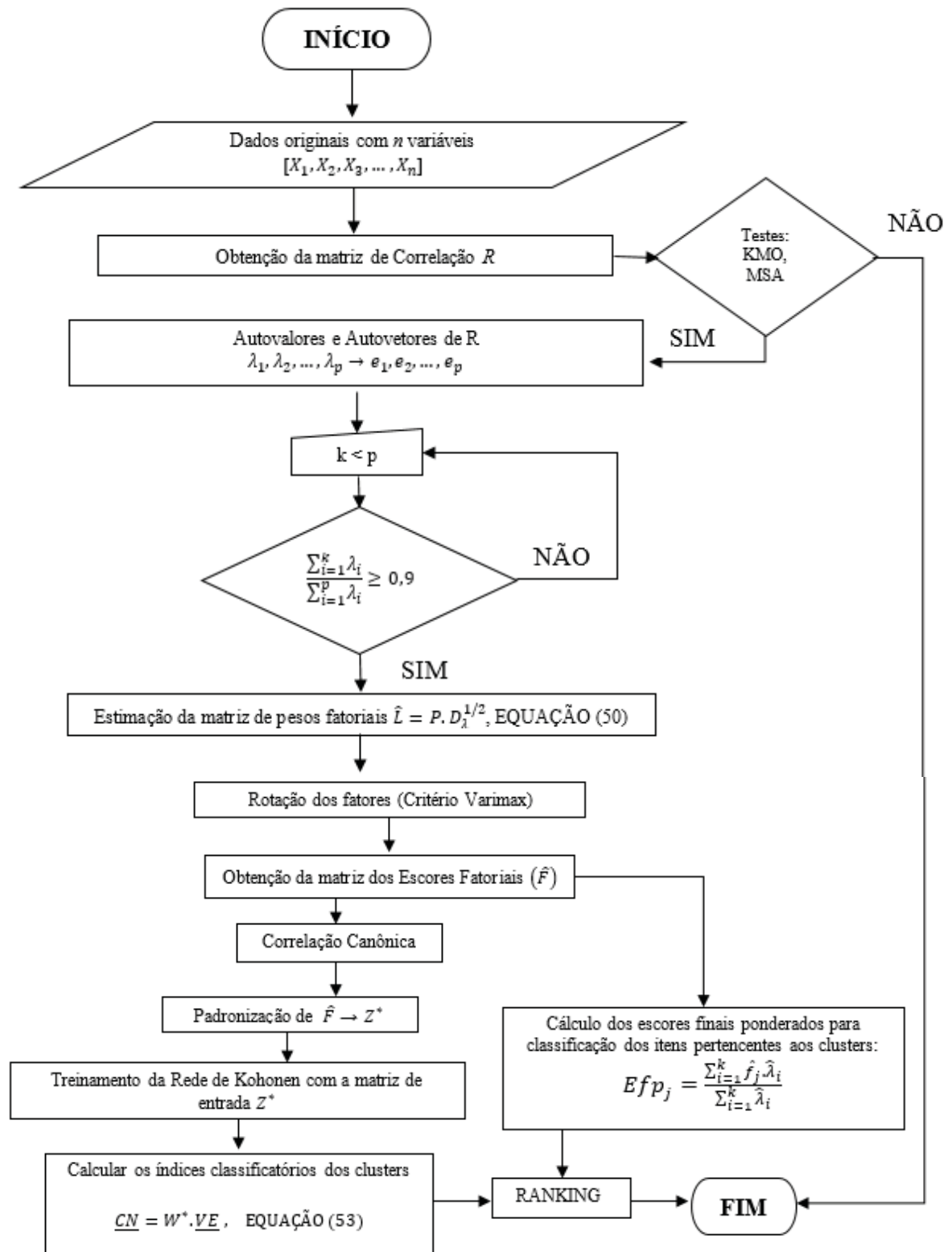
Efp_j : são os escores fatoriais finais ponderados;

$\hat{\lambda}_i$: são os autovalores;

$\hat{f}_j$: é o escore fatorial contido na matriz dos escores fatoriais $\hat{F}$ em correspondência a observação j .

O fluxograma a seguir, FIGURA. 11, mostra os passos do algoritmo proposto:

FIGURA 11 – FLUXOGRAMA DO RANQUEAMENTO E CLASSIFICAÇÃO POR AF/SOM



FONTE: O autor (2021)

4 RESULTADOS E DISCUSSÕES

Nesta seção, são apresentados e discutidos os resultados da aplicação do método proposto de classificação AF/SOM (Análise Fatorial e Mapas Auto-Organizáveis). São encontrados os resultados do ranqueamento da Análise Fatorial e na sequência os agrupamentos desse resultado.

4.1 VERIFICAÇÃO DA ADEQUAÇÃO DA AMOSTRA PARA A ANÁLISE FATORIAL

A estatística *KMO* mede o grau de adequação da amostra quanto ao grau de correlação parcial entre as variáveis, resultando um valor de 0,7879, sendo a medida de adequação da amostra considerada “boa”. Os resultados *KMO* e teste de esfericidade de Bartlett estão na TABELA 3.

TABELA 3 – ESTATÍSTICA *KMO* E TESTE DE ESFERICIDADE DE BARTLETT

Critério de Kaiser-Meyer-Olkin (<i>KMO</i>)		0,7879
Teste de esfericidade de Bartlett	Qui-quadrado aprox.	68562,62
	Gl	595
	Valor-p	< 0,000

FONTE: Software R (2021).

Em geral, os resultados dos testes de *KMO* e de esfericidade de Bartlett tendem a ser uniformes, aceitando ou negando a possibilidade de fatoração da matriz de dados, neste caso, como pode ser visualizado na TABELA 3, apresentou valor- $p < 0,000$, logo, pelo teste de esfericidade constata-se que os dados são adequados para a Análise Fatorial nas variáveis indicadas, rejeitando a hipótese nula de que a matriz de correlação dos dados é uma matriz identidade.

Em relação a Matriz Anti-Imagem (TABELA 4), destaca-se que quanto maior o valor do MSA, mais adequada é a variável para a análise.

TABELA 4 – MEDIDA DE ADEQUAÇÃO AMOSTRAL ENTRE AS 35 VARIÁVEIS

(continua)

Variável	<i>MSA</i>	Variável	<i>MSA</i>	Variável	<i>MSA</i>
X_1	0,8701	X_{13}	0,5059	X_{25}	0,6790
X_2	0,7879	X_{14}	0,7916	X_{26}	0,5475

TABELA 4 – MEDIDA DE ADEQUAÇÃO AMOSTRAL ENTRE AS 35 VARIÁVEIS

(conclusão)

Variável	<i>MSA</i>	Variável	<i>MSA</i>	Variável	<i>MSA</i>
X_3	0,8898	X_{15}	0,8228	X_{27}	0,7698
X_4	0,8282	X_{16}	0,6106	X_{28}	0,8176
X_5	0,7421	X_{17}	0,8275	X_{29}	0,6754
X_6	0,8370	X_{18}	0,7465	X_{30}	0,8913
X_7	0,8185	X_{19}	0,8347	X_{31}	0,9319
X_8	0,7443	X_{20}	0,8163	X_{32}	0,8894
X_9	0,7835	X_{21}	0,7582	X_{33}	0,7760
X_{10}	0,6492	X_{22}	0,7931	X_{34}	0,9006
X_{11}	0,6801	X_{23}	0,6546	X_{35}	0,9549
X_{12}	0,7071	X_{24}	0,6799		

FONTE: Software R (2021).

De acordo com Corrar et al. (2009), as variáveis com valores de *MSA* menor que 0,5 podem ser retiradas da análise, pois não seriam adequadas. No caso em questão, nenhuma das 35 variáveis testadas obteve *MSA* menor que 0,5 portanto, todas as variáveis são mantidas na análise pois são adequadas ao estudo.

4.2 RESULTADOS DA ANÁLISE FATORIAL

Analizada as condições para a aplicação da Análise Fatorial, a partir da matriz de dados originais, obteve-se a matriz de correlação, os autovalores e autovetores correspondentes. A escolha do número de fatores foi determinada através da máxima variância explicada. Assim admitiu-se 12 fatores que explicam 92,65% da variância total, conforme representado na TABELA 5.

TABELA 5 – AUTOVALORES DA MATRIZ DE DADOS ORIGINAIS E PERCENTUAL DA VARIÂNCIA EXPLICADA

(continua)

Fator	Autovalor	Total da Variância Explicada	Total da Variância acumulada
1	15,3103	43,74%	43,74%
2	5,7635	16,47%	60,21%
3	2,0801	5,94%	66,15%
4	1,9291	5,51%	71,67%
5	1,6911	4,83%	76,50%
6	1,3828	3,95%	80,45%
7	0,8419	2,41%	82,85%
8	0,8178	2,34%	85,19%
9	0,7714	2,20%	87,39%

TABELA 5 – AUTOVALORES DA MATRIZ DE DADOS ORIGINAIS E PERCENTUAL DA VARIÂNCIA EXPLICADA

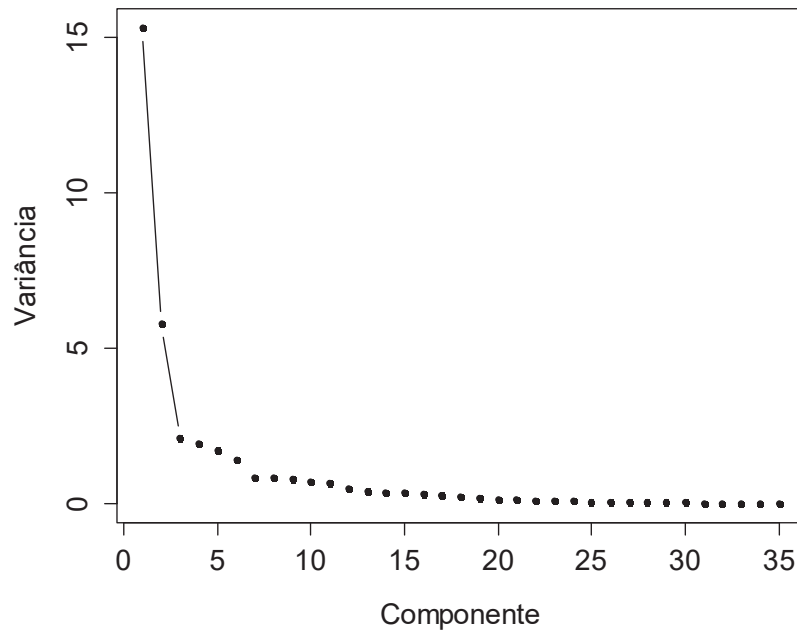
(conclusão)

Fator	Autovalor	Total da Variância Explicada	Total da Variância acumulada
10	0,7072	2,02%	89,42%
11	0,6490	1,85%	91,27%
12	0,4846	1,38%	92,65%
13	0,3766	1,08%	93,73%
14	0,3446	0,98%	94,71%
15	0,3266	0,93%	95,65%
16	0,2778	0,79%	96,44%
17	0,2340	0,67%	97,11%
18	0,1946	0,56%	97,67%
19	0,1693	0,48%	98,15%
20	0,1363	0,39%	98,54%
21	0,1089	0,31%	98,85%
22	0,0902	0,26%	99,11%
23	0,0865	0,25%	99,36%
24	0,0612	0,17%	99,53%
25	0,0369	0,11%	99,64%
26	0,0351	0,10%	99,74%
27	0,0313	0,09%	99,83%
28	0,0171	0,05%	99,87%
29	0,0117	0,03%	99,91%
30	0,0104	0,03%	99,94%
31	0,0065	0,02%	99,96%
32	0,0054	0,02%	99,97%
33	0,0046	0,01%	99,98%
34	0,0032	0,01%	99,99%
35	0,0023	0,01%	100,00%

FONTE: O autor (2021).

O *Screeplot* (FIGURA 12), foi obtido por meio da Análise Fatorial Exploratória, utilizando o método de extração de Componentes Principais e o método de Rotação Varimax na matriz de correlação das $p = 35$ variáveis. Como pode ser observado (FIGURA 12), a partir do 12º autovalor o gráfico de autovalores se torna linear, ou seja, não há contribuição significativa nos demais fatores.

FIGURA 12 – GRÁFICO *SCREEPLOT* DOS AUTOVALORES OBTIDOS NA ANÁLISE FATORIAL



FONTE: Software R (2021)

Estes 12 fatores irão representar as 35 variáveis originais. Assim, consegue-se reduzir a matriz de dados de ordem (92×35) para (92×12) e os fatores são independentes e estão nas direções de maior variabilidade.

Na TABELA 6, estão as comunalidades e variâncias específicas das variáveis.

TABELA 6 – COMUNALIDADES E VARIÂNCIAS ESPECÍFICAS DAS VARIÁVEIS

Variáveis	Comu.	Var. Esp.	Variáveis	Comu.	Var. Esp.	Variáveis	Comu.	Var. Esp.
X_1	0,83	0,17	X_{13}	0,99	0,01	X_{25}	0,96	0,04
X_2	0,97	0,03	X_{14}	0,95	0,05	X_{26}	0,98	0,02
X_3	0,85	0,15	X_{15}	0,91	0,09	X_{27}	0,98	0,02
X_4	0,89	0,11	X_{16}	0,94	0,06	X_{28}	0,93	0,07
X_5	0,89	0,11	X_{17}	0,89	0,11	X_{29}	0,97	0,03
X_6	0,93	0,07	X_{18}	0,94	0,06	X_{30}	0,88	0,12
X_7	0,96	0,04	X_{19}	0,92	0,08	X_{31}	0,89	0,11
X_8	0,93	0,07	X_{20}	0,98	0,02	X_{32}	0,91	0,09
X_9	0,94	0,06	X_{21}	0,94	0,06	X_{33}	0,89	0,11
X_{10}	0,93	0,07	X_{22}	0,92	0,08	X_{34}	0,90	0,10
X_{11}	0,96	0,04	X_{23}	0,92	0,08	X_{35}	0,86	0,14
X_{12}	0,96	0,04	X_{24}	0,96	0,04			

FONTE: O autor (2021).

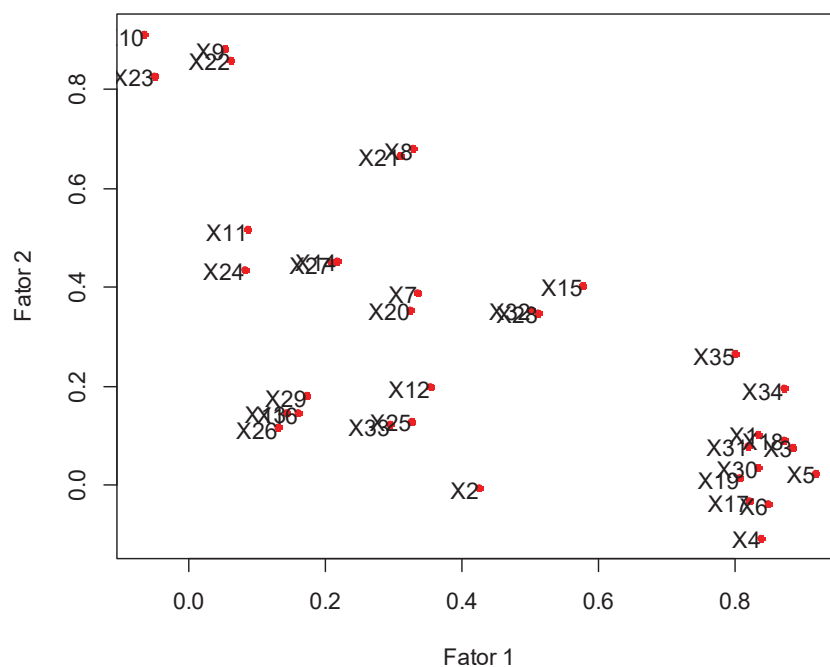
Observa-se que as comunalidades são relativamente altas indicando que todas as $p = 35$ variáveis são importantes na análise. Por outro lado, as variâncias específicas são baixas

variando de 0,03 a 0,17 pois são os complementos das comunalidades para 1. Isto sugere que a maior parte da variabilidade apresentada pelas variáveis são devido aos 12 fatores e, conseqüentemente, o erro devido à aleatoriedade é pequeno. Isto é imprescindível para a verossimilhança da classificação (FURTADO, 2003).

4.2.1 Carregamentos Fatoriais e Identificação dos Fatores

Para definir claramente as variáveis de cada fator procede-se a Rotação Varimax, que maximiza a variância das cargas fatoriais para cada fator por meio do aumento das cargas altas e a diminuição das cargas baixas. Assim, observa-se pela FIGURA 13, que é possível identificar com maior precisão quais variáveis correspondem a cada fator, observando os pesos fatoriais e sua proximidade.

FIGURA 13 – CARGAS DOS FATORES 1 E 2 APÓS ROTAÇÃO VARIMAX



FONTE: Software R (2021).

Com a Rotação Varimax se percebe uma separação das variáveis em relação aos fatores. Na FIGURA 13, as variáveis X_1 , X_3 , X_4 , X_5 , X_6 , X_{17} , X_{18} , X_{19} , X_{30} , X_{31} , X_{34} , X_{35} têm cargas mais altas no fator 1, ao passo que X_9 , X_{10} , X_{22} , X_{23} têm cargas mais altas no fator 2. O

detalhamento completo das variáveis podem ser obtidos a partir da matriz de carregamentos fatoriais (TABELA 7). As variáveis são agrupadas nos fatores através do tamanho de seu carregamento, quanto maior for o carregamento maior é a contribuição da variável dentro do fator.

TABELA 7 – MATRIZ DE CARREGAMENTOS/PESOS FATORIAIS APÓS ROTAÇÃO VARIMAX

Var.	Fator											
	1	2	3	4	5	6	7	8	9	10	11	12
X_1	0,8349	0,1030	-0,0668	0,0837	0,0721	-0,2449	0,1114	0,0524	0,1349	0,0537	0,0678	-0,0463
X_2	0,4268	-0,0063	-0,0018	-0,0417	-0,0188	-0,2247	0,0307	0,0158	0,0125	0,8590	0,0017	-0,0091
X_3	0,8851	0,0764	0,0955	-0,0236	0,0209	0,0051	0,0955	-0,0261	-0,0452	0,1503	0,0660	0,1143
X_4	0,8394	-0,1074	0,1288	-0,1707	0,0326	-0,1194	-0,0091	-0,1601	0,1882	0,2071	0,0519	0,0365
X_5	0,9177	0,0246	0,0328	-0,0758	0,0568	-0,1129	0,0227	0,0506	0,0191	0,0169	-0,0346	-0,1368
X_6	0,8494	-0,0391	-0,0212	0,0747	0,1380	-0,3588	0,0136	-0,0952	0,1476	0,0450	0,1249	0,0152
X_7	0,3353	0,3889	0,0877	-0,0264	0,0775	-0,0872	0,1366	-0,0814	0,7950	0,0094	0,1128	0,0921
X_8	0,3282	0,6803	0,1183	0,1490	0,2976	-0,0020	-0,0160	-0,1200	0,2968	0,0436	0,2850	-0,2254
X_9	0,0536	0,8815	0,0646	-0,1705	0,0603	-0,0670	0,1719	-0,1790	0,1030	-0,0017	0,1434	0,1542
X_{10}	-0,0652	0,9114	0,0795	-0,0571	0,0042	-0,0901	0,2011	-0,0598	0,0485	-0,0382	0,0959	0,1427
X_{11}	0,0867	0,5167	0,0846	-0,1012	-0,0060	-0,1017	0,7776	-0,1349	0,0703	0,0229	0,1765	0,0511
X_{12}	0,3533	0,1973	0,0246	-0,0913	0,1049	-0,8537	0,1126	-0,0498	0,0696	0,0874	0,1271	0,0477
X_{13}	0,1436	0,1466	0,9606	-0,0334	-0,0180	-0,0122	0,0390	-0,0438	0,0185	-0,0030	0,1281	0,0062
X_{14}	0,2183	0,4537	0,0817	-0,0671	0,1757	-0,0227	0,1317	-0,7757	0,0537	-0,0163	0,1622	0,0678
X_{15}	0,5784	0,4039	-0,0155	0,1559	0,2843	-0,4051	0,1685	-0,0889	0,1631	0,1219	0,1995	0,1551
X_{16}	0,1611	0,1454	0,0222	-0,2108	0,8878	-0,1427	0,0124	-0,1305	0,0370	0,0114	0,1490	0,0563
X_{17}	0,8213	-0,0327	0,1747	-0,1198	0,1101	-0,1474	-0,0079	-0,1443	0,2872	0,1528	-0,1032	-0,0284
X_{18}	0,8721	0,0908	0,0826	-0,0124	0,1140	-0,1480	0,0129	0,0830	0,1058	-0,0279	-0,2746	-0,1879
X_{19}	0,8075	0,0139	0,0078	0,1236	0,2112	-0,3736	0,0014	-0,0430	0,2352	-0,0040	-0,0740	-0,0263
X_{20}	0,3249	0,3543	0,1010	0,0608	0,1677	-0,1238	0,1208	-0,0277	0,8113	0,0073	-0,1191	0,0142
X_{21}	0,3106	0,6676	0,1122	0,1253	0,3739	-0,0498	-0,0119	-0,1366	0,3597	0,0050	0,0223	-0,2724
X_{22}	0,0610	0,8605	0,0723	-0,1043	0,0939	-0,0907	0,1285	-0,2679	0,1643	0,0213	-0,1427	0,0605
X_{23}	-0,0492	0,8273	0,0808	0,0987	0,1105	-0,1791	0,1782	-0,0542	0,1155	0,0081	-0,3562	-0,0151
X_{24}	0,0831	0,4368	0,1013	0,0854	0,0839	-0,1579	0,7895	-0,1313	0,1830	0,0209	-0,1887	-0,0368
X_{25}	0,3260	0,1292	0,0281	-0,0348	0,1973	-0,8635	0,0710	-0,0108	0,0586	0,1309	-0,1417	-0,0105
X_{26}	0,1301	0,1153	0,9462	0,0262	0,0992	-0,0235	0,0774	-0,0570	0,1115	0,0060	-0,1386	-0,0247
X_{27}	0,2082	0,4489	0,0848	-0,0238	0,2188	-0,0672	0,1450	-0,7833	0,0712	0,0001	-0,1560	0,0786
X_{28}	0,5115	0,3485	0,0252	0,2247	0,3786	-0,4500	0,1625	-0,0634	0,2658	0,0812	-0,2118	0,0481
X_{29}	0,1721	0,1806	0,0645	-0,1030	0,8863	-0,1773	0,0443	-0,1315	0,1523	-0,0346	-0,1669	0,0475
X_{30}	0,8347	0,0351	-0,0394	-0,2484	0,1247	0,0582	0,0011	-0,1645	-0,0788	-0,0001	-0,0907	0,2350
X_{31}	0,8192	0,0801	-0,0116	-0,1171	0,1229	-0,0814	-0,0165	-0,1664	0,1563	-0,0658	0,0309	0,3519
X_{32}	0,5007	0,3538	-0,0290	-0,0926	0,1986	-0,0851	0,0154	-0,2031	0,1243	-0,0050	0,0059	0,6494
X_{33}	0,2948	0,1221	0,0094	-0,8074	0,3457	-0,0673	0,0088	-0,0655	-0,0356	0,0465	0,0040	0,0604
X_{34}	0,8731	0,1946	0,1627	-0,0812	0,0596	-0,0378	0,1142	-0,0728	0,0758	0,1044	0,0966	0,1239
X_{35}	0,8006	0,2651	0,2243	-0,0067	0,0155	-0,0740	-0,0455	-0,2173	0,1176	0,1170	0,1172	0,0535

FONTE: O autor (2021).

TABELA 8 – CARREGAMENTOS/PESOS FATORIAIS COM AS VARIÁVEIS AGRUPADAS

(continua)

Var.	Fator											
	1	2	3	4	5	6	7	8	9	10	11	12
X_1	0,8349	0,1030	-0,0668	0,0837	0,0721	-0,2449	0,1114	0,0524	0,1349	0,0537	0,0678	-0,0463
X_3	0,8851	0,0764	0,0955	-0,0236	0,0209	0,0051	0,0955	-0,0261	-0,0452	0,1503	0,0660	0,1143
X_4	0,8394	-0,1074	0,1288	-0,1707	0,0326	-0,1194	-0,0091	-0,1601	0,1882	0,2071	0,0519	0,0365
X_5	0,9177	0,0246	0,0328	-0,0758	0,0568	-0,1129	0,0227	0,0506	0,0191	0,0169	-0,0346	-0,1368
X_6	0,8494	-0,0391	-0,0212	0,0747	0,1380	-0,3588	0,0136	-0,0952	0,1476	0,0450	0,1249	0,0152
X_{15}	0,5784	0,4039	-0,0155	0,1559	0,2843	-0,4051	0,1685	-0,0889	0,1631	0,1219	0,1995	0,1551
X_{17}	0,8213	-0,0327	0,1747	-0,1198	0,1101	-0,1474	-0,0079	-0,1443	0,2872	0,1528	-0,1032	-0,0284
X_{18}	0,8721	0,0908	0,0826	-0,0124	0,1140	-0,1480	0,0129	0,0830	0,1058	-0,0279	-0,2746	-0,1879
X_{19}	0,8075	0,0139	0,0078	0,1236	0,2112	-0,3736	0,0014	-0,0430	0,2352	-0,0040	-0,0740	-0,0263
X_{28}	0,5115	0,3485	0,0252	0,2247	0,3786	-0,4500	0,1625	-0,0634	0,2658	0,0812	-0,2118	0,0481
X_{30}	0,8347	0,0351	-0,0394	-0,2484	0,1247	0,0582	0,0011	-0,1645	-0,0788	-0,0001	-0,0907	0,2350

TABELA 8 – CARREGAMENTOS/PESOS FATORIAIS COM AS VARIÁVEIS AGRUPADAS

(conclusão)

Var.	Fator											
	1	2	3	4	5	6	7	8	9	10	11	12
X_{31}	0,8192	0,0801	-0,0116	-0,1171	0,1229	-0,0814	-0,0165	-0,1664	0,1563	-0,0658	0,0309	0,3519
X_{34}	0,8731	0,1946	0,1627	-0,0812	0,0596	-0,0378	0,1142	-0,0728	0,0758	0,1044	0,0966	0,1239
X_{35}	0,8006	0,2651	0,2243	-0,0067	0,0155	-0,0740	-0,0455	-0,2173	0,1176	0,1170	0,1172	0,0535
X_8	0,3282	0,6803	0,1183	0,1490	0,2976	-0,0020	-0,0160	-0,1200	0,2968	0,0436	0,2850	-0,2254
X_9	0,0536	0,8815	0,0646	-0,1705	0,0603	-0,0670	0,1719	-0,1790	0,1030	-0,0017	0,1434	0,1542
X_{10}	-0,0652	0,9114	0,0795	-0,0571	0,0042	-0,0901	0,2011	-0,0598	0,0485	-0,0382	0,0959	0,1427
X_{21}	0,3106	0,6676	0,1122	0,1253	0,3739	-0,0498	-0,0119	-0,1366	0,3597	0,0050	0,0223	-0,2724
X_{22}	0,0610	0,8605	0,0723	-0,1043	0,0939	-0,0907	0,1285	-0,2679	0,1643	0,0213	-0,1427	0,0605
X_{23}	-0,0492	0,8273	0,0808	0,0987	0,1105	-0,1791	0,1782	-0,0542	0,1155	0,0081	-0,3562	-0,0151
X_{13}	0,1436	0,1466	0,9606	-0,0334	-0,0180	-0,0122	0,0390	-0,0438	0,0185	-0,0030	0,1281	0,0062
X_{26}	0,1301	0,1153	0,9462	0,0262	0,0992	-0,0235	0,0774	-0,0570	0,1115	0,0060	-0,1386	-0,0247
X_{33}	0,2948	0,1221	0,0094	-0,8074	0,3457	-0,0673	0,0088	-0,0655	-0,0356	0,0465	0,0040	0,0604
X_{16}	0,1611	0,1454	0,0222	-0,2108	0,8878	-0,1427	0,0124	-0,1305	0,0370	0,0114	0,1490	0,0563
X_{29}	0,1721	0,1806	0,0645	-0,1030	0,8863	-0,1773	0,0443	-0,1315	0,1523	-0,0346	-0,1669	0,0475
X_{12}	0,3533	0,1973	0,0246	-0,0913	0,1049	-0,8537	0,1126	-0,0498	0,0696	0,0874	0,1271	0,0477
X_{25}	0,3260	0,1292	0,0281	-0,0348	0,1973	-0,8635	0,0710	-0,0108	0,0586	0,1309	-0,1417	-0,0105
X_{11}	0,0867	0,5167	0,0846	-0,1012	-0,0060	-0,1017	0,7776	-0,1349	0,0703	0,0229	0,1765	0,0511
X_{24}	0,0831	0,4368	0,1013	0,0854	0,0839	-0,1579	0,7895	-0,1313	0,1830	0,0209	-0,1887	-0,0368
X_{14}	0,2183	0,4537	0,0817	-0,0671	0,1757	-0,0227	0,1317	-0,7757	0,0537	-0,0163	0,1622	0,0678
X_{27}	0,2082	0,4489	0,0848	-0,0238	0,2188	-0,0672	0,1450	-0,7833	0,0712	0,0001	-0,1560	0,0786
X_7	0,3353	0,3889	0,0877	-0,0264	0,0775	-0,0872	0,1366	-0,0814	0,7950	0,0094	0,1128	0,0921
X_{20}	0,3249	0,3543	0,1010	0,0608	0,1677	-0,1238	0,1208	-0,0277	0,8113	0,0073	-0,1191	0,0142
X_2	0,4268	-0,0063	-0,0018	-0,0417	-0,0188	-0,2247	0,0307	0,0158	0,0125	0,8590	0,0017	-0,0091
X_{32}	0,5007	0,3538	-0,0290	-0,0926	0,1986	-0,0851	0,0154	-0,2031	0,1243	-0,0050	0,0059	0,6494

FONTE: O autor (2021).

Analisando as variáveis agrupadas na TABELA 8, pode-se atribuir a seguinte interpretação (identificação) aos fatores:

Fator 1: as variáveis *Número de docentes permanentes do PPG, Número total de discentes, Artigos A1 completos publicados pelos discentes, Artigos A2 completos publicados pelos discentes, Artigos B1 completos publicados pelos discentes, Anais em congresso (discente), Artigos A1 completos publicados pelos docentes, Artigos A2 completos publicados pelos docentes, Artigos B1 completos publicados pelos docentes, Anais em congresso (docentes), Artigos A1 ou A2 publicados pelos egressos do PPG, Artigos B1 ou B2 publicados pelos egressos do PPG, Mestres titulados pelo PPG no quadriênio, Doutores titulados pelo PPG no quadriênio*, foram as variáveis de maior carregamento para o fator 1, denominou-se este fator de “**Publicações A1, A2, B1**”, que explica 43,74% da variabilidade total dos dados.

Fator 2: as variáveis *Artigos B3 completos publicados pelos discentes, Artigos B4 completos publicados pelos discentes, Artigos B5 completos publicados pelos discentes, Artigos B3 completos publicados pelos docentes, Artigos B4 completos publicados pelos docentes, Artigos B5 completos publicados pelos docentes*, foram as variáveis de maior carregamento para o fator 2, denominou-se este fator de “**Publicações B3, B4 e B5**”, que explica 16,46% da variabilidade total dos dados.

Fator 3: as variáveis *Livros – Obra Completa (discentes) e Livros - Obra Completa (docentes)* formam o terceiro fator denominado “**Livro completo**”, que explica 5,94% da variabilidade total dos dados.

Fator 4: a variável *Produção técnica (egressos)*, é o quarto fator, que explica 5,51% da variabilidade total dos dados.

Fator 5: as variáveis *Produção técnica (discente)* e *Produção técnica (docentes)* formam o quinto fator denominado “**Produção técnica**”, que explica 4,83% da variabilidade total dos dados.

Fator 6: as variáveis *Artigos NP completos publicados pelos discentes* e *Artigos NP completos publicados pelos docentes* formam o sexto fator denominado “**Artigos NP**”, que explica 3,95% da variabilidade total dos dados.

Fator 7: as variáveis *Artigos C completos publicados pelos discentes* e *Artigos C completos publicados pelos docentes* formam o sétimo fator denominado “**Artigos C**”, que explica 2,40% da variabilidade total dos dados.

Fator 8: *Livros – Capítulo (discente)* e *Livros – Capítulo (docentes)* formam o oitavo fator denominado “**Capítulo de livro**”, que explica 2,33% da variabilidade total dos dados.

Fator 9: as variáveis *Artigos B2 completos publicados pelos discentes* e *Artigos B2 completos publicados pelos docentes* formam o nono fator denominado “**Artigos B2**”, que explica 2,20% da variabilidade total dos dados.

Fator 10: a variável *Número de docentes colaboradores do PPG* forma o décimo fator, que explica 2,020% da variabilidade total dos dados.

Fator 11: este fator sofre maior influência da variável “*Artigos B5 completos publicados pelos docentes*” e explica cerca de 1,85% da variabilidade total dos dados.

Fator 12: a variável *Anais em congresso (egressos)* forma o décimo segundo fator, que explica 1,38% da variabilidade total dos dados.

A seguir foram obtidos os coeficientes dos Escores Fatoriais, estimados a partir dos carregamentos fatoriais rotacionados.

TABELA 9 – MATRIZ DOS ESCORES FATORIAIS

PPG	Fator 1	Fator 2	Fator 3	Fator 4	Fator 5	Fator 6	Fator 7	Fator 8	Fator 9	Fator 10	Fator 11	Fator 12
1	-4,4231	-2,6088	-1,0861	0,2829	-2,2839	0,6281	-1,2599	1,8558	-1,1056	0,5777	-0,0263	-0,2266
2	-2,1260	-1,4965	-0,6345	0,2664	-1,9575	1,2261	-1,1754	0,6472	-0,6205	-0,3003	-0,0254	0,0551
3	-8,6875	-8,0803	-2,2825	0,2132	-4,6683	4,2989	-3,1274	3,6127	-3,5966	-1,4100	-0,4313	-1,2364
4	2,2786	-2,8677	-0,6134	0,3956	-2,2207	-0,4235	0,4578	1,2905	-1,5269	3,2224	0,2297	-0,3830
5	-6,7297	1,8133	0,6766	0,5369	-0,7981	2,0220	1,9196	-0,7885	-1,4658	-0,4730	-0,1296	-0,5738
6	8,5347	-0,0259	0,4581	-0,2000	-0,9367	-1,2828	-0,4589	-0,1519	3,0355	2,5775	0,0308	-0,7168
7	-13,0076	-8,4696	-2,5787	0,9985	-5,1912	5,1175	-3,3159	3,3342	-4,6333	-1,6856	-0,0111	-1,4222
8	-4,2818	-4,2047	-0,9671	0,2739	-3,6460	2,7122	-0,8411	1,8036	-1,5619	0,1056	0,4539	0,2796
9	-9,7445	-2,7713	-0,2293	0,5670	-3,2581	3,9487	-0,9472	-0,5566	-3,6884	-1,6465	0,0017	-0,4820
10	-10,8778	-7,8300	-2,5619	1,0581	-4,8613	4,0674	-2,6353	3,4737	-3,7849	-0,4655	-0,0300	-1,1896
11	-7,5934	-3,4790	-2,0070	0,7945	-2,3280	0,0941	-1,8003	1,5281	-1,5401	-0,7472	0,1339	-0,2254
12	1,2637	0,1024	-0,4764	-1,3545	6,1719	0,8397	-1,4232	-1,7321	0,8766	-0,8699	0,5821	0,9655
13	-1,7717	-4,5750	-2,0845	0,6881	-2,6200	0,1892	-1,7427	2,1081	-1,9030	-0,0776	0,0697	0,3770
14	-8,5999	-5,9853	-2,3776	1,0095	-2,6956	3,5894	-2,6526	2,0907	-2,9720	-1,3833	0,2746	-1,4965
15	-7,1177	-6,4102	-1,6452	0,5260	-3,1595	3,3171	-2,2750	2,4357	-3,0308	-0,4550	0,2129	-1,2017
16	3,9057	1,3467	-0,2032	-0,0306	0,5031	-0,3763	1,5175	-0,5316	-0,4055	-0,2573	0,1366	-0,1463
17	-3,4225	-5,9896	-1,7314	-0,0791	-2,5860	2,6514	-2,1815	2,1675	-2,3340	-1,1114	0,1706	-0,1184
18	14,0509	-1,4018	0,5230	-0,4017	0,8562	-2,0547	-0,7634	-0,9414	0,7157	1,1726	0,4541	2,2494
19	0,5421	-4,6476	-1,5297	-0,3881	-0,8042	2,5990	-2,2577	1,7633	-3,0532	0,3606	0,4752	-0,1998
20	7,0649	3,4898	3,7493	-1,4136	0,2901	0,2190	-0,1377	-4,0250	0,3759	-0,1043	1,9350	0,9931
21	2,4737	-3,7758	14,7865	-1,1438	-1,5762	2,6386	-1,1575	0,6184	-1,4153	-0,1496	-0,2313	-1,1158
22	1,1752	9,5718	1,7207	-0,1477	3,0050	-5,4361	2,9049	-2,4633	1,8179	-0,1872	0,6407	0,6609
23	2,2914	-3,6343	-0,8493	0,5582	-0,6155	0,1308	-1,1703	0,3818	1,4663	-0,3645	0,1031	-0,8987
24	-9,7550	-6,7941	-2,6693	0,9416	-4,4317	1,3519	-1,6424	3,4710	-3,7182	-1,2431	-0,0178	-1,1949
25	-6,6090	-3,2841	-0,0920	0,8564	-2,4623	1,4289	-1,0655	0,2130	-2,4940	-0,3433	0,2003	-1,0279
26	6,6623	-4,1106	-0,6187	-0,1166	-0,0802	0,0941	-1,9009	1,4438	-1,2694	0,5751	-0,1369	-0,5279
27	-2,6591	-1,4104	-1,3506	-1,2643	2,4736	1,6339	-1,7961	0,4078	-1,2039	-1,2089	0,6822	0,8058
28	-5,7607	-5,7310	-1,8018	-0,5481	-2,0678	2,1909	-2,6293	2,4222	-1,8895	-1,3467	0,0356	-0,5903
29	-13,7542	-8,6385	-3,0007	0,8478	-5,4594	5,3748	-3,5493	3,8225	-4,9871	-1,9498	-0,0981	-1,3163
30	-7,8474	-1,2058	-0,6778	0,3344	-1,5094	2,5606	-1,1506	1,2839	-1,3548	-1,6155	0,3229	-0,3790
31	-11,0503	-1,1245	-1,2740	1,2230	-3,1324	3,3124	-0,7708	1,8385	-2,8283	-0,7757	-0,2525	-0,5552
32	-13,0858	-6,6456	-2,0759	0,3434	-3,7500	4,4410	-2,8331	3,5788	-4,7712	-1,1740	0,0177	-0,9995
33	-6,7934	-7,4431	-2,2040	-0,0044	-3,3063	4,0622	-3,2002	3,1012	-3,6792	-0,7167	-0,0936	-1,2130
34	8,1732	3,2948	0,7563	1,1238	0,7674	-4,0681	3,6107	-0,8856	4,0578	2,1323	0,3900	-0,3569
35	-2,7880	-5,0313	-1,0150	-0,2296	-1,7332	2,2708	-2,2242	0,7971	-1,6869	-0,6993	0,1744	-0,8476
36	1,2418	-2,1944	0,2864	-0,1610	-0,1561	-1,5553	-0,5574	0,4482	-0,6792	1,6038	-0,2832	0,7631
37	2,5369	1,4564	1,9126	0,7907	0,7652	0,0767	-0,7971	0,0488	2,6081	0,7505	-0,1463	-0,1099
38	-6,5037	-3,8074	0,0441	0,4911	-3,1169	1,6322	-1,0468	1,8665	-2,6024	-0,2896	-0,6752	-0,9836
39	18,0317	3,5655	1,6779	-0,4508	3,6816	-4,1271	1,0458	-4,6440	6,7681	2,1152	0,9831	3,2316
40	-5,9290	-4,9907	-1,9664	-0,2285	-2,3371	2,7709	-0,0117	1,9234	-3,3419	-1,1489	-0,0996	-0,1635
41	-5,8037	-4,4274	-1,4300	0,7623	-3,0182	2,2799	-0,6787	1,8894	-1,1939	-0,6129	0,1546	-0,4372
42	-11,4154	-6,8683	-2,5430	0,8088	-4,4382	4,4082	-2,3892	3,1388	-4,2565	-1,5423	0,0769	-1,0378
43	8,8414	13,3315	4,6618	1,6413	3,3929	-1,0187	2,1986	-4,5794	5,9567	-0,2457	2,3508	-0,1295
44	-9,4911	-2,4106	-1,6964	-0,0495	-1,9193	3,6949	-2,0719	2,4804	-1,6437	-1,4311	0,1788	-0,6564
45	33,8754	6,3011	1,9629	-6,1248	11,0133	-10,8222	1,8464	-7,9150	4,8532	4,4640	0,0869	5,1449
46	-5,1280	-6,3809	-2,1977	-0,3367	-2,6891	3,3557	-3,0184	2,6176	-3,6289	-1,2972	-0,1158	-0,4135

(continua)

TABELA 9 – MATRIZ DOS ESCORES FATORIAIS

(continua)

PPG	Fator 1	Fator 2	Fator 3	Fator 4	Fator 5	Fator 6	Fator 7	Fator 8	Fator 9	Fator 10	Fator 11	Fator 12
47	3,8771	9,0968	2,7618	-0,6766	5,7023	-0,9876	1,1137	-4,8298	3,9089	1,6645	1,1369	-0,0826
48	15,5556	1,3741	0,5158	-0,4336	3,0614	-4,6173	0,5722	-2,8751	1,3565	2,2235	0,7174	1,3685
49	4,1191	2,9483	1,8912	-1,3538	4,1716	-0,1862	0,4946	-2,8254	2,0258	0,0041	0,7919	0,9401
50	45,5949	7,8828	2,7588	-2,5358	5,8857	-10,4480	4,7649	-4,7901	4,4853	4,4853	0,2866	3,3642
51	1,0573	-3,8926	-1,9254	-1,7910	-0,2141	1,9620	-2,2230	1,9144	-1,6778	-0,2690	-0,1488	0,4544
52	-4,1156	0,9179	-0,0971	1,0403	0,0928	0,8824	0,0158	0,4127	-0,4185	0,7469	-0,0828	-1,3249
53	1,3275	-0,2158	-0,8747	0,5560	3,7961	-2,6581	0,5526	-0,8862	1,7076	-0,7341	-0,0251	0,1675
54	-0,7209	4,0735	1,1321	-0,3923	5,9475	0,2863	0,6892	-1,5859	2,1107	-1,0292	0,3025	-0,7191
55	28,9417	14,2126	8,3537	0,0295	5,8906	-11,4556	6,0372	-3,5435	11,9332	5,7092	0,3200	0,5764
56	-1,2096	-2,5297	-0,3425	0,0494	-0,2222	-1,0374	0,4199	1,1716	-0,7792	0,3640	0,3025	-0,0599
57	-4,9130	3,7435	0,2962	-1,1175	0,7225	0,5696	2,0573	-0,1810	-1,2429	-1,1926	1,1976	1,7437
58	29,2291	6,6072	1,7805	-2,0380	11,8320	-18,1120	3,0755	-3,2496	4,8714	6,1879	0,2775	1,1393
59	40,2061	19,6447	7,3098	-0,2763	15,9090	-14,3980	5,5118	-7,4663	16,0598	2,3014	-2,0600	1,4258
60	16,1991	11,1806	2,2745	2,3934	8,1265	-9,2804	4,0756	-1,4112	9,6996	1,6128	-1,2414	-1,3089
61	-5,9803	-5,1615	-2,3000	0,9201	-2,9974	3,0674	-2,2302	2,5325	-3,0454	-1,1606	0,2830	-0,2521
62	4,3353	13,5003	3,0756	-0,9798	7,4825	-5,3006	4,0325	-6,7442	5,9680	-0,3906	-1,5185	1,7014
63	-7,5805	-2,5545	-0,8788	1,4656	-1,4864	-0,3892	-1,2908	2,6599	-2,1352	-0,1441	-2,8823	-1,6666
64	-2,2851	3,6168	0,4419	0,3169	3,0482	-2,5123	2,4071	-1,1417	1,2350	-0,2340	-1,6497	1,1404
65	-7,7492	-5,7492	-1,3990	1,2173	-2,8775	1,4387	-1,4263	2,6525	-2,8409	-0,2277	0,2609	-1,3254
66	2,9063	5,9852	1,1682	0,5846	0,2475	-0,1276	2,7606	-1,7234	1,3390	1,0790	0,5590	-0,2835
67	3,2200	-2,5588	-0,0555	-0,1556	-0,7420	0,7822	-1,6056	0,3438	-1,9448	-0,5736	-0,0184	0,4597
68	-10,0646	-0,0032	1,4035	0,9733	-3,1638	2,8393	1,2903	1,5854	-2,0950	-0,8640	0,0392	-0,6739
69	-4,7455	-7,4468	-2,0164	-0,2338	3,2197	2,9147	-2,6522	2,8639	-2,5002	-0,5642	-0,0969	-0,7239
70	-0,7790	3,1515	-0,1022	0,4460	1,7286	-0,0018	1,6827	-3,2226	0,7746	-0,1116	0,0056	-0,2185
71	-8,2980	-7,5234	-2,0030	0,6750	-4,9026	4,1220	-2,9769	3,9388	-2,0746	-1,0774	-0,5443	-1,9129
72	-2,4612	-2,4612	-1,3407	0,9822	-0,9436	0,5056	-0,3771	0,9350	0,1178	-0,8081	0,1069	0,0164
73	-3,8888	8,7514	2,4861	-0,1918	0,1898	0,3993	7,4487	-4,8123	1,6725	-1,0750	-0,0082	0,3254
74	-9,1846	-5,4306	-1,9899	0,8495	-3,6066	0,0834	-1,2567	2,5067	-3,2487	-1,0458	-0,0600	-1,3964
75	-10,9074	-5,4080	-2,6009	1,1707	-4,3292	2,8235	-1,4670	3,0081	-3,3657	-0,9464	0,0370	-1,3540
76	-12,1666	-4,1009	-2,3974	1,1577	-3,5950	4,0265	-1,3874	2,5125	-2,4376	-1,7725	-0,1538	-1,1514
77	30,0250	4,2721	2,3950	-4,7774	5,8526	-5,6734	0,9404	-3,7433	5,5856	4,9387	0,5340	3,3617
78	13,6563	35,9338	4,7232	-0,9140	9,5447	-9,0054	10,9001	-11,7141	11,3499	1,3788	0,7337	4,0962
79	-7,9549	-1,8570	0,3064	1,6656	-0,2711	0,8791	1,5399	1,9653	-0,4593	-1,0320	-3,1919	-1,5412
80	17,8629	10,3399	2,5349	-2,0598	5,2192	-4,8026	3,2818	-7,0440	4,8045	1,4235	0,1887	5,9080
81	10,6119	16,0237	3,5073	-0,9678	5,1547	-2,7760	4,9539	-7,9142	4,7650	0,5069	-2,2295	1,0768
82	1,2761	15,9368	2,8453	-0,7568	2,2868	-3,5104	7,8398	-2,5535	5,5464	-0,5648	1,0853	-0,2185
83	-5,8313	2,8089	-0,1098	0,7747	-1,4216	0,6964	0,5168	-0,2329	-1,1272	-0,9236	-1,0339	-0,2185
84	-12,6630	-7,9783	-2,6265	1,2274	-5,0884	4,9899	-3,2423	3,8500	-3,3689	-1,8874	0,0501	-1,4415
85	-2,2812	9,1901	0,5307	-6,4318	6,0824	-0,3453	3,2776	-3,6079	0,8774	-0,4632	0,3815	0,9591
86	-7,7837	-3,0370	-1,4190	0,3323	-0,1855	1,4550	-1,2726	1,8772	-1,6268	-1,2052	-1,3831	-0,4545
87	-9,7092	-1,6562	-1,7927	0,5280	-0,8560	3,0320	-1,4478	1,8159	-2,3622	-1,2207	-0,2723	-0,7940
88	-9,2251	-5,3023	-2,4166	1,3571	-4,0788	3,2718	-2,0904	2,7681	-3,7469	-1,2656	0,0076	-1,3275
89	-12,5600	-7,1322	-2,5845	0,7055	-4,8747	4,4114	-2,6771	3,4778	-4,8314	-1,5310	0,0534	-1,0158

TABELA 9 – MATRIZ DOS ESCORES FATORIAIS

(conclusão)

PPG	Fator 1	Fator 2	Fator 3	Fator 4	Fator 5	Fator 6	Fator 7	Fator 8	Fator 9	Fator 10	Fator 11	Fator 12
90	-6,7099	-3,6707	-1,8828	0,6890	-2,9201	2,0869	-0,7054	1,4880	-2,2750	0,3467	0,3900	-0,6297
91	-3,6929	-4,2148	-1,7321	1,0036	-3,2214	1,1820	-1,3428	2,2175	-2,3077	0,0388	-0,0348	-1,1211
92	-8,9741	-4,8285	-2,3499	1,3127	-3,8631	2,8539	-1,4256	2,4608	-3,5890	-1,0264	-0,1682	-0,9025

FONTE: O autor (2021).

4.2.2 Ranqueamento dos Programas pela Análise Fatorial

O fator 1 (“Publicações A1, A2, B1”) possui maior autovalor, e portanto, explica a maior parte da variância sendo este o fator mais importante na classificação. Assim, é natural portanto, que os Programas que apresentam um escore maior nesse fator tendem a ser melhor classificados. A TABELA 10 apresenta o Ranking dos Programas conforme essa classificação.

O ranqueamento do vetor (PPG) é feito através do Escore Final que corresponde a média dos escores fatoriais ponderados pelos autovalores que ditam a importância dos fatores. O escore fatorial ponderado para o PPG (vetor) i é dado pela expressão (54). Como os Escores Fatoriais podem apresentar valores negativos, para melhor entendimento, somou-se a cada escore o módulo do valor mínimo dos escores e por regra de três colocou-se os escores em uma escala de 9 a 10, que na TABELA 10 chamou-se de *Escore Padronizado*, de forma a facilitar a interpretação.

TABELA 10 – RANKING DOS 92 PPG's AVALIADOS PELA ANÁLISE FATORIAL

(continua)

Rank AF	N	PPG	Sigla	Escore (E _{fp})	Escore Padronizado
1º	59	ENGENHARIA AERONÁUTICA E MECÂNICA	ITA	23,5087	10,0000
2º	50	ENGENHARIA MECÂNICA	UNICAMP	23,1444	9,8973
3º	45	ENGENHARIA MECÂNICA	USP	17,1273	8,2003
4º	55	ENGENHARIA MECÂNICA	UNESP/GUAR	17,0370	8,1749
5º	77	ENGENHARIA MECÂNICA	UFSC	15,0989	7,6283
6º	58	ENGENHARIA E TECNOLOGIA ESPACIAIS	INPE	15,0829	7,6238
7º	78	ENGENHARIA DE PRODUÇÃO	UFSC	13,5597	7,1942
8º	80	ENGENHARIA MECÂNICA	UFRGS	10,5235	6,3379
9º	60	CIÊNCIAS E TECNOLOGIAS ESPACIAIS	ITA	10,2431	6,2589
10º	39	ENGENHARIA MECÂNICA	UFU	9,4288	6,0292
11º	81	ENGENHARIA DE PRODUÇÃO	UFRGS	8,2007	5,6829
12º	48	ENGENHARIA MECÂNICA	USP/SC	7,6163	5,5181
13º	43	ENGENHARIA DE PRODUÇÃO M/D	UFSCAR	7,1968	5,3997
14º	18	ENGENHARIA MECÂNICA	UFRJ	6,3930	5,1731
15º	62	ENGENHARIA DE PRODUÇÃO	UNIP	4,8124	4,7273
16º	34	ENGENHARIA MECÂNICA	UFMG	4,6431	4,6796
17º	20	ENGENHARIA DE PRODUÇÃO	UFRJ	4,0917	4,5241
18º	6	CIÊNCIA E ENGENHARIA DE PETRÓLEO	UFRN	4,0413	4,5098
19º	47	ENGENHARIA (ENGENHARIA DE PRODUÇÃO)	USP	3,8973	4,4692
20º	82	ENGENHARIA DE PRODUÇÃO	UFSM	3,8631	4,4596
21º	49	ENGENHARIA DE PRODUÇÃO	USP/SC	2,7388	4,1425
22º	66	MÉTODOS NUMÉRICOS EM ENGENHARIA	UFPR	2,6436	4,1157
23º	22	ENGENHARIA DE PRODUÇÃO	UFF	2,3576	4,0350
24º	26	ENGENHARIA MECÂNICA	PUC-RIO	2,3268	4,0263
25º	16	ENGENHARIA INDUSTRIAL	UFBA	2,0899	3,9595
26º	37	ENGENHARIA DE PRODUÇÃO	UNIFEI	1,7238	3,8563
27º	21	PLANEJAMENTO ENERGÉTICO	UFRJ	1,3349	3,7466
28º	67	ENGENHARIA MECÂNICA	UFPR	1,0546	3,6675
29º	12	ENGENHARIA DE PRODUÇÃO	UFPE	0,8087	3,5982
30º	54	ENGENHARIA DE PRODUÇÃO	UNESP/BAU	0,7562	3,5834
31º	53	ENGENHARIA MECÂNICA	UNESP/BAU	0,6686	3,5587
32º	85	ENGENHARIA DE PRODUÇÃO E SISTEMAS	UNISINOS	0,5364	3,5214
33º	4	ENGENHARIA MECÂNICA	UFRN	0,4937	3,5094
34º	23	ENGENHARIA MECÂNICA	UFF	0,3942	3,4813

TABELA 10 – RANKING DOS 92 PPG's AVALIADOS PELA ANÁLISE FATORIAL

(conclusão)

Rank					Escore	Escore
AF	N	PPG	Sigla	(Efp _i)	Padronizado	
35°	72	ENGENHARIA MECÂNICA E DE MATERIAIS	UTFPR	0,3520	3,4694	
36°	70	ENGENHARIA DE PRODUÇÃO E SISTEMAS	PUC/PR	0,2776	3,4484	
37°	36	ENGENHARIA MECÂNICA	UNIFEI	0,1520	3,4130	
38°	73	ENGENHARIA DE PRODUÇÃO	UTFPR	-0,0126	3,3666	
39°	64	ENGENHARIA DE PRODUÇÃO	UNINOVE	-0,2950	3,2869	
40°	51	CIÊNCIAS E ENGENHARIA DE PETRÓLEO	UNICAMP	-0,4016	3,2569	
41°	19	ENGENHARIA OCEÂNICA	UFRJ	-0,6948	3,1742	
42°	56	ENGENHARIA MECÂNICA	UNESP/IS	-1,0605	3,0710	
43°	2	ENGENHARIA DE RECURSOS NATURAIS DA AMAZÔNIA	UFPA	-1,3796	2,9810	
44°	27	ENGENHARIA DE PRODUÇÃO	PUC-RIO	-1,5349	2,9372	
45°	57	ENGENHARIA DE PRODUÇÃO	UNIMEP	-1,5964	2,9199	
46°	52	ENGENHARIA DE PRODUÇÃO E DE MANUFATURA	UNICAMP/LI	-1,6861	2,8946	
47°	13	ENGENHARIA MECÂNICA	UFPE	-1,9029	2,8335	
48°	83	ENGENHARIA OCEÂNICA	FURG	-2,3226	2,7151	
49°	35	ENGENHARIA DE PRODUÇÃO	UFMG	-2,3850	2,6975	
50°	1	ENGENHARIA MECÂNICA	UFPA	-2,7006	2,6085	
51°	91	CIÊNCIAS MECÂNICAS	UNB	-2,7120	2,6053	
52°	5	ENGENHARIA DE PRODUÇÃO	UFRN	-2,7614	2,5914	
53°	8	ENGENHARIA MECÂNICA	UFPB/J,P,	-2,8870	2,5559	
54°	17	ENGENHARIA MECÂNICA	UFES	-2,8980	2,5528	
55°	41	ENGENHARIA MECÂNICA	UFSJ	-3,6487	2,3411	
56°	25	ENGENHARIA MECÂNICA	UERJ	-3,8268	2,2909	
57°	69	ENGENHARIA MECÂNICA	PUC/PR	-3,8320	2,2894	
58°	46	ENGENHARIA NAVAL E OCEÂNICA	USP	-3,8487	2,2847	
59°	38	ENGENHARIA DE ENERGIA	UNIFEI	-3,8846	2,2746	
60°	40	ENGENHARIA MECÂNICA	PUC/MG	-3,8904	2,2730	
61°	61	ENGENHARIA MECÂNICA	FEI	-3,9490	2,2564	
62°	79	ENGENHARIA MECÂNICA	UDESC	-3,9744	2,2493	
63°	30	ENGENHARIA DE PRODUÇÃO E SISTEMAS	CEFET/RJ	-3,9765	2,2487	
64°	28	METROLOGIA	PUC-RIO	-3,9905	2,2447	
65°	90	SISTEMAS MECATRÔNICOS	UNB	-3,9923	2,2442	
66°	63	ENGENHARIA MECÂNICA	IFSP	-4,1992	2,1859	
67°	86	ENGENHARIA MECÂNICA	UNISINOS	-4,3187	2,1522	
68°	11	ENGENHARIA MECÂNICA	UFCEG	-4,4640	2,1112	
69°	68	ENGENHARIA DE PRODUÇÃO	UFPR	-4,6527	2,0580	
70°	15	MECATRÔNICA	UFBA	-4,6906	2,0473	
71°	33	ENGENHARIA DE RESERVATÓRIO E DE EXPLORAÇÃO	UENF	-4,7990	2,0167	
72°	65	ENGENHARIA MECÂNICA	UFABC	-4,8437	2,0041	
73°	87	SISTEMAS E PROCESSOS INDUSTRIAIS	UNISC	-4,9691	1,9687	
74°	44	ENGENHARIA DE PRODUÇÃO M	UFSCAR	-5,0314	1,9512	
75°	9	ENGENHARIA DE PRODUÇÃO	UFPB/J,P,	-5,2451	1,8909	
76°	14	ENGENHARIA DE PRODUÇÃO – CA DO AGRESTE	UFPE	-5,3378	1,8648	
77°	92	INTEGRIDADE DE MATERIAIS DA ENGENHARIA	UNB	-5,3470	1,8622	
78°	71	ENGENHARIA MECÂNICA	UEM	-5,5133	1,8153	
79°	31	ENGENHARIA DE PRODUÇÃO	UCAM	-5,5192	1,8136	
80°	88	ENGENHARIA	UNIPAMPA	-5,5661	1,8004	
81°	74	ENGENHARIA MECÂNICA	UTFPR	-5,6546	1,7754	
82°	3	ENGENHARIA MECÂNICA	UFC	-5,8650	1,7161	
83°	24	ENGENHARIA MECÂNICA - VOLTA REDONDA	UFF	-6,1906	1,6243	
84°	75	ENGENHARIA MECÂNICA	UTFPR	-6,3957	1,5664	
85°	76	ENGENHARIA DE PRODUÇÃO E SISTEMAS (PPGEPS)	UTFPR	-6,6632	1,4910	
86°	10	ENERGIAS RENOVÁVEIS	UFPB/J,P,	-6,8081	1,4501	
87°	42	ENGENHARIA DA ENERGIA - CEFET-MG – UFSJ	CEFET/MG	-6,9003	1,4241	
88°	89	ENGENHARIA DE PRODUÇÃO E SISTEMAS	PUC-GOÍAS	-7,5315	1,2461	
89°	32	ENGENHARIA DE PRODUÇÃO	UENF	-7,6151	1,2225	
90°	84	ENGENHARIA MECÂNICA	FURG	-7,6734	1,2061	
91°	7	ENGENHARIA MECATRÔNICA	UFRN	-7,9754	1,1209	
92°	29	ENGENHARIA MECÂNICA	IME	-8,4050	1,0000	

FONTE: O autor (2021).

Observa-se que o curso Engenharia Aeronáutica e Mecânica - ITA (nº59) apresentou o maior escore padronizado, igual a 10, seguido dos Programas de Engenharia Mecânica da Unicamp e da USP (nº50 e nº45), com escores padronizados iguais a 9,886 e 9,80, respectivamente. Na parte inferior do Ranking em 90º aparece o curso de Engenharia Mecânica da FURG, com escore padronizado igual a 9,0229, seguidos de Engenharia Mecatrônica da UFRN e Engenharia Mecânica do IME, com escore padronizado 9,01135 e 9,00, respectivamente.

A partir do Ranking obtido pela Análise Fatorial pode-se caracterizar os Programas da seguinte forma:

- PPG 59 – Engenharia Aeronáutica e Mecânica – ITA: Este PPG é o 1º colocado no Ranking e apresentou os maiores Escores Fatoriais para os fatores 1 e 2 que são os mais significativos 40,2061 e 19,6447, respectivamente. O menor escore está no fator 12, com valor igual a 1,4258, sendo o fator 12, o fator com menor importância na classificação. Portanto, o PPG 59 é o Programa com maior destaque em publicações com Qualis Capes A1, A2 e B1, sendo este portanto, um critério importante para a obter boa classificação, sendo que na avaliação da CAPES o Programa obteve conceito 6.
- PPG 29 – Engenharia Mecânica – IME: Este PPG é o último colocado no ranking (92º) e apresentou escores extremamente baixos nos fatores 1 e 2, com valores -13.7542 e -8.6385, respectivamente. Na avaliação CAPES o Programa obteve conceito 2.

Pode-se concluir que, quanto maior o valor escore fatorial nos fatores que explicam a maior parte da variação total, melhor classificado tende a ser o PPG.

4.3 RESULTADO DA ANÁLISE DE CORRELAÇÃO CANÔNICA

Neste trabalho aplicou-se a Análise de Correlação Canônica no sentido de verificar se, para os 92 Programas de Pós-Graduação avaliados, a associação entre as 35 variáveis originais e os escores fatoriais obtidos dos 12 fatores extraídos era significativa, ou seja, se as informações são praticamente as mesmas. Isto é mostrado pelos resultados da técnica e estão na TABELA 11.

TABELA 11 – RESULTADO DA APLICAÇÃO DA ANÁLISE DE CORRELAÇÃO CANÔNICA

Número	Autovalor	Correlação Canônica	Lâmbda De Wilks	Valor-p
Fator 1	1,0	1,0	1,80297E-138	0,000
Fator 2	1,0	1,0	2,14244E-125	0,000
Fator 3	1,0	1,0	1,56393E-113	0,000
Fator 4	1,0	1,0	9,9665E-102	0,000
Fator 5	1,0	1,0	5,26326E-90	0,000
Fator 6	1,0	1,0	2,50355E-78	0,000
Fator 7	1,0	1,0	5,86049E-67	0,000
Fator 8	1,0	1,0	1,36671E-55	0,000
Fator 9	1,0	1,0	2,62021E-44	0,000
Fator 10	1,0	1,0	3,35457E-33	0,000
Fator 11	1,0	1,0	3,5826E-22	0,000
Fator 12	1,0	1,0	2,33517E-11	0,000

FONTE: Statgraphics (2021).

Observando o Valor-*p*, verifica-se que todas as correlações canônicas são estatisticamente importantes considerando um nível de significância de 5% pelo teste do *Qui-quadrado*. Analisando ainda, os autovalores bem como as correlações canônicas, são todas perfeitas iguais a 100%.

Assim, pode-se afirmar que existe perfeita correlação entre os dois conjuntos de variáveis, sendo a primeira função canônica do primeiro conjunto representada por:

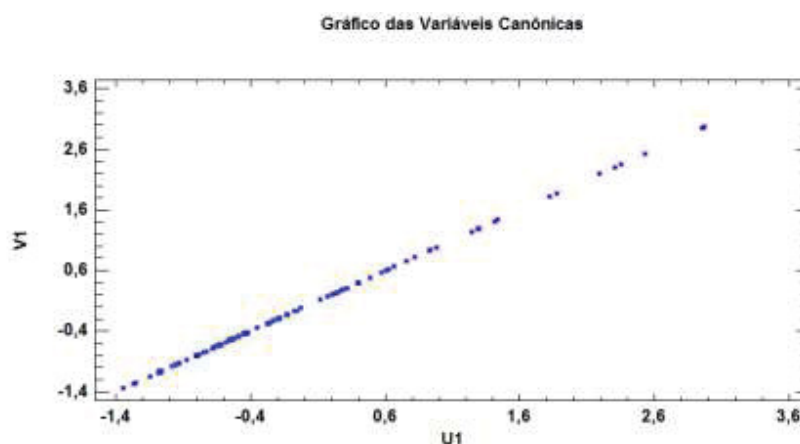
$$\begin{aligned}
 U_1 = & 0,0187439X_1 + 0,0621681X_2 + 0,0356116X_3 + 0,0504784X_4 + 0,00957968X_5 + \\
 & 0,0301942X_6 + 0,0585812X_7 + 0,0217687X_8 + 0,0574738X_9 + 0,0419535X_{10} + 0,06419X_{11} + \\
 & 0,0562208X_{12} + 0,0445091X_{13} + 0,0621114X_{14} + 0,0570109X_{15} + 0,0570925X_{16} + \\
 & 0,0458905X_{17} + 0,00260505X_{18} + 0,0231925X_{19} + 0,0464971X_{20} + 0,0193196X_{21} + \\
 & 0,0498837X_{22} + 0,0244465X_{23} + 0,0501551X_{24} + 0,0462994X_{25} + 0,0432614X_{26} + \\
 & 0,0607444X_{27} + 0,0435587X_{28} + 0,0514624X_{29} + 0,0407987X_{30} + 0,0514217X_{31} + \\
 & 0,0781884X_{32} + 0,0673866X_{33} + 0,0493535X_{34} + 0,0442207X_{35}
 \end{aligned}$$

E a função canônica do segundo conjunto representada por:

$$\begin{aligned}
 V_1 = & 0,114775.Fator1 + 0,113118.Fator2 + 0,0949047.Fator3 - 0,0737122.Fator4 + \\
 & 0,118331.Fator5 - 0,116212.Fator6 + 0,107318.Fator7 - 0,118657.Fator8 + \\
 & 0,118886.Fator9 + 0,0972059.Fator10 + 0,0111923.Fator11 + 0,101353.Fator12
 \end{aligned}$$

A FIGURA 14, apresenta o gráfico da relação entre as primeiras variáveis dos dois conjuntos, U_1 e V_1 , onde se verifica a perfeita correlação, portanto, se pode utilizar os escores fatoriais para o treinamento da Rede SOM em substituição aos dados originais praticamente sem perda de informação.

FIGURA 14 – GRÁFICO DAS VARIÁVEIS CANÔNICAS



FONTE: Statgraphics (2021).

4.4 RESULTADO AF/SOM

A rede SOM, aplicada em conjunto com a Análise Fatorial permite a obtenção de clusters homogêneos, tendo como dado de entrada para o treinamento da Rede Neural os escores fatoriais das variáveis latentes.

A categorização dos clusters torna-se possível, pois as características de cada neurônio/cluster podem ser comparadas com o poder de explicação de cada variável latente (fator), sendo possível, portanto, a distinção dos clusters como “mais eficientes” ou “menos eficientes” adotando os critérios de classificação da Análise Fatorial. A necessidade de conhecer a importância das variáveis no estudo para a classificação do cluster torna o uso das duas técnicas em conjunto necessária.

Para o estudo de caso apresentado, devido a natureza dos dados, optou-se pela formação de um número fixo de seis Clusters, pois como a CAPES atribui conceito de 2 a 7 para os PPG's é de interesse dessa pesquisa, a separação dos mesmos em seis grupos distintos e classificá-los com conceitos 2, 3, 4, 5, 6 ou 7, assim como na CAPES. Em outras aplicações pode ser necessário que

o algoritmo investigue a quantidade apropriada de clusters, isso pode ser obtido mais facilmente com o SOM, utilizando uma malha de dimensão maior na execução do algoritmo e verificar a quantidade de clusters pela análise do mapa topográfico.

O treinamento da Rede SOM ocorre no ambiente do Software R, o *script* de execução permite ajustar todos os parâmetros da Rede. A Rede foi desenhada e testada com diversas topologias de malhas bidimensionais para obtenção dos clusters. A malha considerada para este propósito foi 3×2 . Todas as outras malhas que também resultam em seis neurônios/clusters foram testadas e apresentaram resultados semelhantes à malha.

O primeiro passo para a obtenção dos resultados é a normalização em duas etapas, utilizando as equações (51) e (52). A matriz de dados normalizada utilizada está no APÊNDICE 4.

A base de dados para entrada no algoritmo continua uma matriz de ordem 92×12 , em que as linhas representam os PPG's e as colunas as variáveis latentes (fatores). Os neurônios em conjunto com a matriz de distâncias unificadas (*U-mat*), que contém as distâncias Euclidianas calculadas para os pares dos neurônios vizinhos, possibilitam a definição dos agrupamentos.

O processo de organização topológica ocorre primeiro onde dados semelhantes ocupam regiões específicas do mapa e, diante disso, são verificadas as distâncias em que neurônios próximos são associados a um mesmo grupo (*cluster*). A densidade dos dados também pode ser utilizada para definir as fronteiras entre agrupamentos diferentes, principalmente em mapas maiores, nos quais os neurônios de transição entre um grupo e outro ficam vazios, ou seja, não possuem dados associados ao seu modelo (ZHANG; LI, 1993).

A FIGURA 15 apresenta parte do *script* de execução em que se define o número de iterações e taxa de aprendizagem. A taxa de aprendizagem é que determina o tamanho dos ajustes durante o treinamento, a diminuição é linear e os valores padrão devem começar em 0,05 e parar em 0,01.

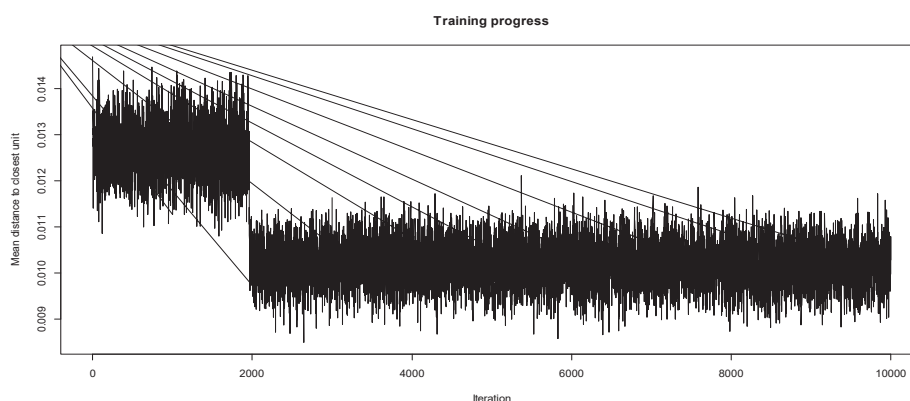
FIGURA 15 – *SCRIPT* DO ALGORÍTMO SOM NO SOFTWARE R

```
# Executar o algoritmo SOM (Self Organizing Map)
#####
#--> Definir o lay-out da camada (grid de neurônios) "Kohonen"
#####
vGrid_som = somgrid(xdim=3, ydim=2, topo="hexagonal")
#--> executa efetivamente o algoritmo...
set.seed(7)
vMapaKohonen = som(vKOH2,
  grid = vGrid_som,
  rlen = 10000,
  alpha = c(0.05, 0.01),
  keep.data = TRUE)
vMapaKohonen
```

FONTE: Software R (2021)

Na fase de treinamento da Rede SOM a estabilidade na atualização dos pesos é obtida a partir de um certo número de iterações (LOIOLA, et al 2007). Uma medida que tem sido adotada por Kohonen é de 500 iterações para cada neurônio de saída. Isto significa que havendo seis neurônios na camada de saída, deverá haver perto de 3000 iterações para a ordenação correta do mapa. Essas 3000 iterações significam passar 3000 vezes o conjunto de treinamento pela Rede Neural. No entanto, conforme discutido na Revisão de Literatura, esses parâmetros não são regras e podem variar conforme a aplicação, neste caso, 10.000 iterações foram necessárias para que não houvesse mais mudanças na malha. O gráfico (FIGURA 16) mostra o processo iterativo.

FIGURA 16 – EVOLUÇÃO DO TREINAMENTO DA REDE



FONTE: Software R (2021)

A TABELA 12 apresenta a matriz de pesos sinápticos da Rede Neural após o treinamento. Esses são os “pesos” das variáveis (fatores) que detém maior influência na formação do perfil dos clusters.

TABELA 12 – PESO DOS FATORES POR NEURÔNIO/CLUSTER DO SOM

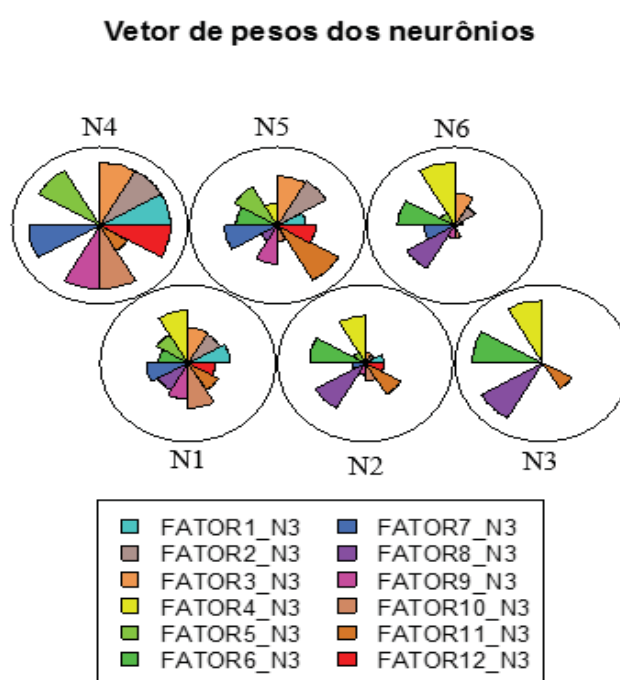
N/F	Fator 1	Fator 2	Fator 3	Fator 4	Fator 5	Fator 6	Fator 7	Fator 8	Fator 9	Fator 10	Fator 11	Fator 12
N1	0,63	0,56	0,56	0,56	0,58	0,38	0,59	0,42	0,62	0,67	0,51	0,53
N2	0,45	0,38	0,39	0,53	0,41	0,58	0,39	0,61	0,41	0,47	0,55	0,46
N3	0,31	0,31	0,31	0,61	0,30	0,69	0,30	0,71	0,30	0,33	0,50	0,33
N4	0,85	0,80	0,76	0,26	0,84	0,16	0,79	0,14	0,84	0,80	0,49	0,83
N5	0,53	0,71	0,67	0,39	0,68	0,46	0,66	0,29	0,63	0,44	0,74	0,60
N6	0,37	0,48	0,55	0,62	0,42	0,58	0,51	0,58	0,41	0,42	0,29	0,36

FONTE: O autor (2021).

O Fator 1 é composto por 14 das 35 variáveis originais e explica a maior parte da variabilidade, cerca de 43,74%, seguido pelo Fator 2, com 6 variáveis originais explicando 16,43% da variabilidade e assim por diante, diminuindo o grau de explicação dos fatores ao passo que, o último fator explica somente 1,4% da variabilidade total. Portanto, as variáveis originais que compõem o primeiro fator tem mais impacto na obtenção dos índices classificatórios, seguido dos demais.

Observa-se na matriz da TABELA 12 que o Neurônio 4 (N4) possui o maior peso para o Fator 1, igual a 0,85 seguido do Neurônio 1 (N1) com peso 0,63. Os pesos desta matriz podem ser visualizados na FIGURA 17, de uma forma melhor.

FIGURA 17 – GRÁFICO VETOR DE PESOS DOS NEURÔNIOS



FONTE: Software R (2021).

A FIGURA 17, chamada de diagrama do ventilador, é uma representação gráfica dos pesos que estão na TABELA 12. A área do setor circular do peso é diretamente proporcional ao valor do peso (que está sempre entre 0 e 1) e pode-se perceber que o Neurônio 4 possui os maiores pesos para todos os fatores, exceto pelo Fator 6 (Artigos NP) e pelo Fator 8 (Capítulo de Livro), isto sugere que o Neurônio 4 tende a ter uma melhor classificação, haja vista, que os pesos desse

neurônio só são menores em relação aos outros, nos Fatores 6 e 8 e que, no entanto, o grau de importância desses fatores não é muito relevante, cerca de 3,9% e 2,3%, respectivamente, não influenciando de forma significativa a classificação.

Para obter uma classificação precisa levando em conta o poder de explicação de cada um dos 12 fatores, multiplica-se essa matriz de pesos (TABELA 12) pelo vetor correspondente a variância explicada pelos 12 fatores,

$$\underline{CN} = W^* . \underline{VE} \quad i = 1, 2, \dots, 12$$

onde $\underline{CN}$ é índice classificatório do neurônio/cluster, $\underline{VE}$ é o vetor da variância explicada pelo i -ésimo fator obtida na análise fatorial e W^* é a matriz de pesos após o treinamento da Rede SOM. Os resultados da aplicação dessa equação são apresentados na TABELA 13.

TABELA 13 – ÍNDICE CLASSIFICATÓRIO DOS *CLUSTERS*

<i>Cluster</i>	<i>CN</i>	<i>Ranking Cluster</i>	<i>CN</i> Ordem decrescente
1	0,543367	4	0,688918
2	0,412926	1	0,543367
3	0,332781	5	0,528907
4	0,688918	2	0,412926
5	0,528907	6	0,405480
6	0,405480	3	0,332781

FONTE: O autor (2021).

Quanto maior o valor de $\underline{CN}$, melhor classificado está o *Cluster*. A classificação dos PPG's pelos escores da Análise Fatorial é feita ordenando-se os escores em ordem decrescente, logo a ordenação dos clusters fica de acordo com a sequência 4, 1, 5, 2, 6 e 3, conforme exposto na 3ª e 4ª coluna da TABELA 13.

A alocação dos PPG's em cada cluster se dá pelo cálculo da distância Euclidiana mínima em relação ao vetor peso dos neurônios (TABELA 12) e a ordenação dos PPG's dentro dos *Clusters* pelos escores fatoriais ponderados (TABELA 10). Os resultados da classificação dos PPG's estão na TABELA 14.

TABELA 14 – RESULTADO DA CLASSIFICAÇÃO COM SEIS *CLUSTERS* (AF/SOM)

(continua)

Rank					Cluster	Score	Conceito	Conceito
AF/SOM	PPG	NOME DO PPG	IES	SOM	Fatorial	Capes	AF/SOM	
1º	59	ENGENHARIA AERONÁUTICA E MECÂNICA	ITA	4	1,0000	6	Conceito 7	
2º	50	ENGENHARIA MECÂNICA	UNICAMP	4	0,9886	5		
3º	45	ENGENHARIA MECÂNICA	USP	4	0,8000	5		
4º	55	ENGENHARIA MECÂNICA	UNESP/GUAR	4	0,7972	4		
5º	77	ENGENHARIA MECÂNICA	UFSC	4	0,7365	6		
6º	58	ENGENHARIA E TECNOLOGIA ESPACIAIS	INPE	4	0,7360	4		
7º	78	ENGENHARIA DE PRODUÇÃO	UFSC	4	0,6883	5		
8º	80	ENGENHARIA MECÂNICA	UFRGS	4	0,5931	6		
9º	39	ENGENHARIA MECÂNICA	UFU	4	0,5588	7		
10º	81	ENGENHARIA DE PRODUÇÃO	UFRGS	4	0,5203	7		
11º	62	ENGENHARIA DE PRODUÇÃO	UNIP	4	0,4142	5		
12º	60	CIÊNCIAS E TECNOLOGIAS ESPACIAIS	ITA	1	0,5843	4	Conceito 6	
13º	48	ENGENHARIA MECÂNICA	USP/SC	1	0,5020	5		
14º	18	ENGENHARIA MECÂNICA	UFRJ	1	0,4637	7		
15º	34	ENGENHARIA MECÂNICA	UFMG	1	0,4089	4		
16º	6	CIÊNCIA E ENGENHARIA DE PETRÓLEO	UFRN	1	0,3900	3		
17º	66	MÉTODOS NUMÉRICOS EM ENGENHARIA	UFPR	1	0,3462	3		
18º	16	ENGENHARIA INDUSTRIAL	UFBA	1	0,3289	5		
19º	37	ENGENHARIA DE PRODUÇÃO	UNIFEI	1	0,3174	4		
20º	53	ENGENHARIA MECÂNICA	UNESP/BAU	1	0,2843	4		
21º	70	ENGENHARIA DE PRODUÇÃO E SISTEMAS	PUC/PR	1	0,2721	5		
22º	36	ENGENHARIA MECÂNICA	UNIFEI	1	0,2681	4		
23º	64	ENGENHARIA DE PRODUÇÃO	UNINOVE	1	0,2541	4		
24º	43	ENGENHARIA DE PRODUÇÃO M/D	UFSCAR	5	0,4889	5	Conceito 5	
25º	20	ENGENHARIA DE PRODUÇÃO	UFRJ	5	0,3916	4		
26º	47	ENGENHARIA (ENGENHARIA DE PRODUÇÃO)	USP	5	0,3855	4		
27º	82	ENGENHARIA DE PRODUÇÃO	UFSM	5	0,3844	4		
28º	49	ENGENHARIA DE PRODUÇÃO	USP/SC	5	0,3492	5		
29º	22	ENGENHARIA DE PRODUÇÃO	UFF	5	0,3372	4		
30º	12	ENGENHARIA DE PRODUÇÃO	UFPE	5	0,2887	7		
31º	54	ENGENHARIA DE PRODUÇÃO	UNESP/BAU	5	0,2871	5		
32º	85	ENGENHARIA DE PRODUÇÃO E SISTEMAS	UNISINOS	5	0,2802	4		
33º	73	ENGENHARIA DE PRODUÇÃO	UTFPR	5	0,2630	4		
34º	57	ENGENHARIA DE PRODUÇÃO	UNIMEP	5	0,2133	4		
35º	26	ENGENHARIA MECÂNICA	PUC-RIO	2	0,3363	7	Conceito 4	
36º	67	ENGENHARIA MECÂNICA	UFPR	2	0,2964	5		
37º	4	ENGENHARIA MECÂNICA	UFRN	2	0,2788	3		
38º	23	ENGENHARIA MECÂNICA	UFF	2	0,2757	5		
39º	72	ENGENHARIA MECÂNICA E DE MATERIAIS	UTFPR	2	0,2744	4		
40º	51	CIÊNCIAS E ENGENHARIA DE PETRÓLEO	UNICAMP	2	0,2508	5		
41º	19	ENGENHARIA OCEÂNICA	UFRJ	2	0,2416	4		
42º	56	ENGENHARIA MECÂNICA	UNESP/IS	2	0,2301	4		
43º	2	ENGENHARIA DE RECURSOS NATURAIS DA AMAZÔNIA	UFPA	2	0,2201	5		
44º	27	ENGENHARIA DE PRODUÇÃO	PUC-RIO	2	0,2153	5		
45º	13	ENGENHARIA MECÂNICA	UFPE	2	0,2037	3		
46º	35	ENGENHARIA DE PRODUÇÃO	UFMG	2	0,1886	4		

TABELA 14 – RESULTADO DA CLASSIFICAÇÃO COM SEIS *CLUSTERS* (AF/SOM)

							(conclusão)	
Rank	AF/SOM	PPG	NOME DO PPG	IES	Cluster SOM	Score Fatorial	Conceito Capes	Conceito AF/SOM
47º	1		ENGENHARIA MECÂNICA	UFPA	2	0,1787	3	Conceito 4
48º	91		CIÊNCIAS MECÂNICAS	UNB	2	0,1784	4	
49º	8		ENGENHARIA MECÂNICA	UFPB/J.P.	2	0,1729	2	
50º	17		ENGENHARIA MECÂNICA	UFES	2	0,1726	4	
51º	41		ENGENHARIA MECÂNICA	UFSJ	2	0,1490	4	
52º	25		ENGENHARIA MECÂNICA	UERJ	2	0,1435	3	
53º	40		ENGENHARIA MECÂNICA	PUC/MG	2	0,1415	4	
54º	30		ENGENHARIA DE PRODUÇÃO E SISTEMAS	CEFET/RJ	2	0,1388	4	
55º	90		SISTEMAS MECATRÔNICOS	UNB	2	0,1383	2	
56º	11		ENGENHARIA MECÂNICA	UFCG	2	0,1235	3	
57º	21		PLANEJAMENTO ENERGÉTICO	UFRJ	6	0,3052	6	Conceito 3
58º	52		ENGENHARIA DE PRODUÇÃO E DE MANUFATURA	UNICAMP/LI	6	0,2105	3	
59º	83		ENGENHARIA OCEÂNICA	FURG	6	0,1906	3	
60º	5		ENGENHARIA DE PRODUÇÃO	UFRN	6	0,1768	3	
61º	38		ENGENHARIA DE ENERGIA	UNIFEI	6	0,1416	3	
62º	79		ENGENHARIA MECÂNICA	UDESC	6	0,1388	3	
63º	63		ENGENHARIA MECÂNICA	IFSP	6	0,1318	3	
64º	86		ENGENHARIA MECÂNICA	UNISINOS	6	0,1280	4	
65º	68		ENGENHARIA DE PRODUÇÃO	UFPR	6	0,1176	3	
66º	69		ENGENHARIA MECÂNICA	PUC/PR	3	0,1433	6	
67º	46		ENGENHARIA NAVAL E OCEÂNICA	USP	3	0,1428	5	
68º	61		ENGENHARIA MECÂNICA	FEI	3	0,1396	3	
69º	28		METROLOGIA	PUC-RIO	3	0,1383	5	
70º	15		MECATRÔNICA	UFBA	3	0,1164	4	
71º	33		ENGENHARIA DE RESERVATÓRIO E DE EXPLORAÇÃO	UENF	3	0,1130	4	
72º	65		ENGENHARIA MECÂNICA	UFABC	3	0,1116	3	
73º	87		SISTEMAS E PROCESSOS INDUSTRIAIS	UNISC	3	0,1077	3	
74º	44		ENGENHARIA DE PRODUÇÃO M	UFSCAR	3	0,1057	3	
75º	9		ENGENHARIA DE PRODUÇÃO	UFPB/J.P.	3	0,0990	2	
76º	14		ENGENHARIA DE PRODUÇÃO – CA DO AGRESTE	UFPE	3	0,0961	3	
77º	92		INTEGRIDADE DE MATERIAIS DA ENGENHARIA	UNB	3	0,0958	3	
78º	71		ENGENHARIA MECÂNICA	UEM	3	0,0906	3	
79º	31		ENGENHARIA DE PRODUÇÃO	UCAM	3	0,0904	2	
80º	88		ENGENHARIA	UNIPAMPA	3	0,0890	3	
81º	74		ENGENHARIA MECÂNICA	UTFPR	3	0,0862	3	
82º	3		ENGENHARIA MECÂNICA	UFC	3	0,0796	3	
83º	24		ENGENHARIA MECÂNICA - VOLTA REDONDA	UFF	3	0,0694	2	
84º	75		ENGENHARIA MECÂNICA	UTFPR	3	0,0630	3	
85º	76		ENGENHARIA DE PRODUÇÃO E SISTEMAS (PPGEPS)	UTFPR	3	0,0546	3	
86º	10		ENERGIAS RENOVÁVEIS	UFPB/J.P.	3	0,0500	3	
87º	42		ENGENHARIA DA ENERGIA - CEFET-MG – UFSJ	CEFET/MG	3	0,0471	2	
88º	89		ENGENHARIA DE PRODUÇÃO E SISTEMAS	PUC-GOÍÁS	3	0,0274	3	
89º	32		ENGENHARIA DE PRODUÇÃO	UENF	3	0,0248	2	
90º	84		ENGENHARIA MECÂNICA	FURG	3	0,0229	3	
91º	7		ENGENHARIA MECATRÔNICA	UFRN	3	0,0135	2	
92º	29		ENGENHARIA MECÂNICA	IME	3	0,0000	2	

FONTE: O autor (2021).

Na 1ª coluna (TABELA 14) está o Ranking dos PPG's, na 2ª coluna, o número de identificação do PPG, na 3ª coluna, a IES do Programa, na 4ª coluna, o número do cluster do SOM no qual o PPG está inserido, na 5ª coluna estão os Escores Finais Classificatórios obtidos na Análise Fatorial, na 6ª coluna, os conceitos atribuídos pela CAPES no quadriênio 2013-2016, e na última coluna, o conceito atribuído aos PPG's de acordo com a metodologia aplicada neste trabalho. De acordo com o ranqueamento feito através do algoritmo híbrido AF/SOM, pode-se concluir que o cluster de melhor desempenho receberia a Nota 7, o segundo melhor cluster a Nota 6, e assim sucessivamente conforme colocado na 7ª coluna da TABELA 14.

Assim, o *Cluster 4*, fica em primeiro lugar com 11 Programas que serão classificados com “*Conceito 7*”, depois vem os demais na classificação:

Cluster 1, com 12 Programas classificados com “*Conceito 6*”;

Cluster 5, com 11 Programas classificados com “*Conceito 5*”;

Cluster 2, com 22 Programas classificados com “*Conceito 4*”;

Cluster 6, com 9 Programas classificados com “*Conceito 3*”;

Cluster 3, com 27 Programas classificados com “*Conceito 2*”.

O Ranking final obtido pelo método híbrido AF/SOM (TABELA 14) e o Ranking inicial obtido por Análise Fatorial (TABELA 10) apresentam Coeficiente de Correlação de Spearman igual a 0,5730 com Valor- $p = 0$ que, pelos padrões normais, a associação entre as duas variáveis (resultados) seria considerada estatisticamente significativa. Existe, portanto, consistência entre os métodos da Análise Fatorial e o híbrido AF/SOM.

Aplicando o algoritmo das K-Médias em substituição ao SOM e estabelecendo a formação de seis clusters, assim como no SOM, o *Cluster 4* melhor classificado (“conceito 7”), tem apenas quatro PPG's, seguido por:

Cluster 1, com 8 Programas classificados com “*Conceito 6*”;

Cluster 5, com 9 Programas classificados com “*Conceito 5*”;

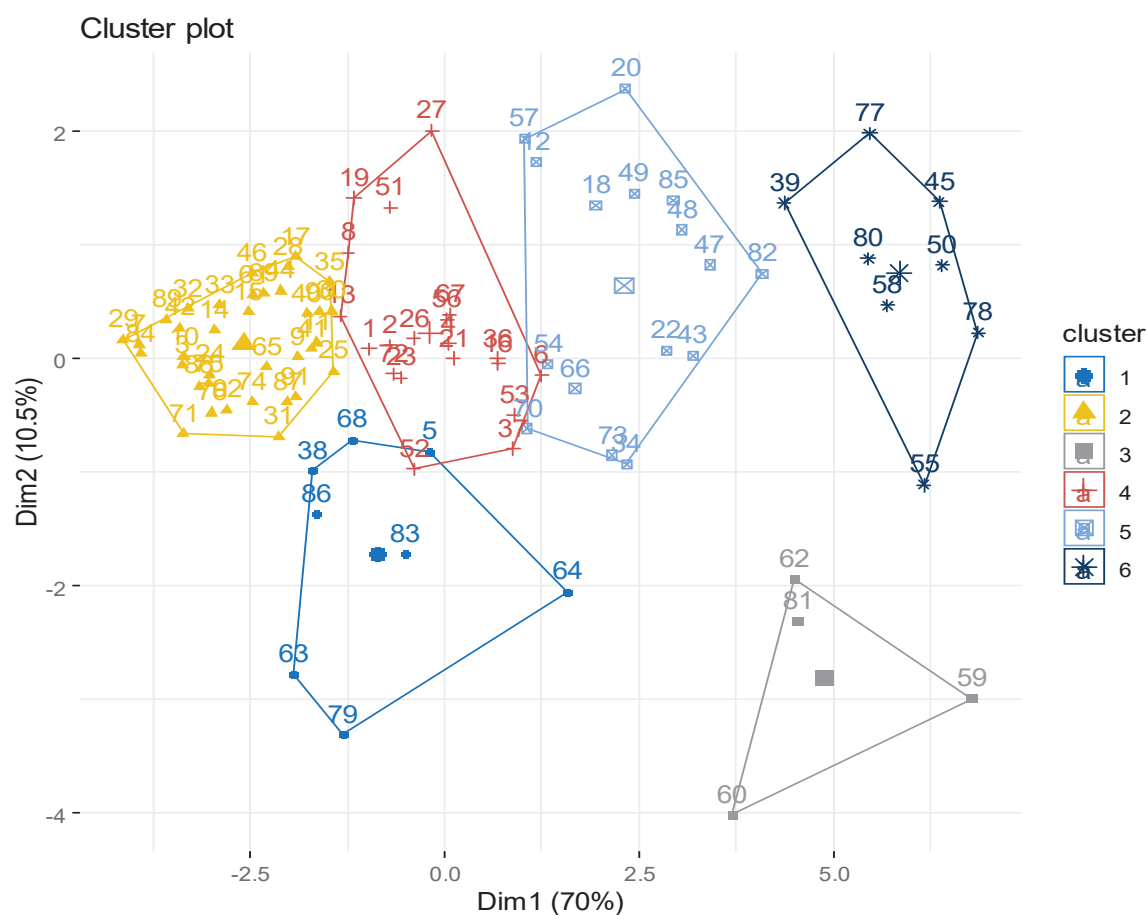
Cluster 2, com 7 Programas classificados com “*Conceito 4*”;

Cluster 6, com 22 Programas classificados com “*Conceito 3*”;

Cluster 3, com 42 Programas classificados com “*Conceito 2*”.

A distribuição dos Programas nos *Clusters* pelo Algoritmo das K-Médias pode ser visualizado no gráfico da FIGURA 18.

FIGURA 18 – FORMAÇÃO DOS *CLUSTERS* AF/K-MÉDIAS



FONTE: Software R (2021).

Para uma comparação dos resultados obtidos com AF/SOM (Análise Fatorial e SOM) e AF/K-Médias (Análise Fatorial e K-Médias) utilizados nesta análise, passou-se a uma avaliação das classificações encontradas.

Na TABELA 15, estão as frequências de Programas classificados pelas duas formas: AF/SOM e AF/K-Médias.

TABELA 15 – CLASSIFICAÇÃO POR AF/SOM E AF/K-MÉDIAS

AF/K- Médias	AF/SOM						Total
	<i>Cluster</i> 1	<i>Cluster</i> 2	<i>Cluster</i> 3	<i>Cluster</i> 4	<i>Cluster</i> 5	<i>Cluster</i> 6	
<i>Cluster</i> 1	0	0	0	8	0	0	8
<i>Cluster</i> 2	2	0	0	0	5	0	7
<i>Cluster</i> 3	0	12	27	0	0	3	42
<i>Cluster</i> 4	1	0	0	3	0	0	4
<i>Cluster</i> 5	2	1	0	0	6	0	9
<i>Cluster</i> 6	7	9	0	0	0	6	22
Total	12	22	27	11	11	9	92

FONTE: O autor (2021).

Pode-se perceber pela TABELA 15 que dos 92 Programas classificados, 42 foram classificados da mesma forma, tanto em AF/SOM como em AF/K-Médias. Nenhum dos clusters são idênticos aos gerados por AF/SOM.

O coeficiente de correlação de Spearman entre a classificação pela Análise Fatorial (pura) e AF/K-Médias é de 0,3629 que é uma correlação menor que a obtida entre Análise Fatorial e AF/SOM que foi de 0,5730. Fica demonstrado nesta aplicação que a utilização de Redes Neurais (SOM) em detrimento do método K-Médias é mais adequado para o agrupamento da classificação obtida pela Análise Fatorial.

5 CONCLUSÃO E TRABALHOS FUTUROS

Este trabalho é formado por dois componentes: uma metodologia de classificação em agrupamento através da proposição de um modelo híbrido baseado em Análise Fatorial e Mapas Auto-Organizáveis (SOM) e a aplicação deste na avaliação da Pós-Graduação *Stricto Sensu*, em particular, os 92 Programas Acadêmicos da Área específica de Engenharias III da CAPES.

A Análise Fatorial por meio da ponderação de seus escores fatoriais, permite a construção de um índice classificatório, em que a unidade com melhor desempenho na atividade fica melhor classificada. A Rede Neural SOM foi treinada utilizando como dado de entrada os valores dos escores fatoriais das variáveis latentes, estas novas variáveis preservam as informações dos dados originais, isto é confirmado e verificado pela Análise da Correlação Canônica. O propósito de se utilizar os escores das variáveis latentes é o de agrupar em clusters os indivíduos que são semelhantes nessa classificação.

Por meio do caso de aplicação, conclui-se que o modelo híbrido desenvolvido apresenta resultados muito bons, visto que, após o treinamento da Rede Neural os resultados obtidos na classificação AF/SOM, se comparado ao obtido da classificação inicial gerada pela Análise Fatorial (AF), apresentaram Coeficiente de Correlação de Spearmann de 57,30%.

Este modelo também pode ser aplicado incorporando outros métodos de clusterização mais simples em substituição ao SOM, como o K-Médias, por exemplo. Mas, apesar do K-Médias ser uma técnica simples de ser implementada e ter custo computacional baixo, apresenta algumas desvantagens em relação ao SOM, que além de ser suscetível a problemas quando os clusters são de diferentes tamanhos e densidades, a quantidade de clusters precisa ser sempre especificada pelo usuário, no caso do SOM a clusterização pode ocorrer de forma automática.

Para comparação dos resultados foi aplicado o método K-Médias em substituição ao SOM, nesta mesma metodologia. Os resultados como se pôde observar são bem distintos, gerando coeficiente de concordância de apenas $K = 0,3174$, o que torna a utilização de um método mais simples como o K-Médias inviável em detrimento ao SOM nesta aplicação, sendo que o Coeficiente de Correlação de Spearman entre AF/K-Médias e AF resultou em apenas 0,3629.

Observa-se que este modelo difere tecnicamente do modelo adotado pela CAPES, haja vista, por exemplo, as variáveis subjetivas presentes nas avaliações da CAPES que nesta proposta não são levadas em consideração, assim como a “Proposta do Programa” e a “Inserção

Internacional”, este último, sendo item necessário ao PPG para obter conceitos 6 e 7 de acordo com critério da CAPES. Entretanto, a metodologia desenvolvida possibilita futuras análises mediante apenas a atualização dos dados e remoção ou inserção de novas variáveis, evitando análises com critérios subjetivos, se baseando apenas em critérios científicos tornando a avaliação mais justa aos PPG's. Com esta aplicação os PPG's podem identificar na análise as variáveis de maior influência positiva na classificação podendo direcionar suas práticas a atingir estes objetivos.

Pode-se concluir portanto, que os resultados obtidos atendem aos objetivos propostos inicialmente e produzem uma metodologia de classificação livre de critérios subjetivos.

Como sugestões para trabalhos futuros a fim de melhorar essa avaliação, pode-se adequar o modelo híbrido testando outras técnicas de agrupamento ao invés do SOM, como por exemplo, as variantes do K-Médias: Forgy e Isodata que faz uma realocação iterativa e auto-organizável dos dados e verificar o desempenho dessas ferramentas. Também pode-se adaptar a metodologia utilizando outro modelo de classificação como base, podendo ser a Análise por Envoltória de Dados (DEA) que em vários trabalhos na literatura é utilizado como metodologia de classificação.

REFERÊNCIAS

- ANDRIOTTI, J. L. S. Análise de Componentes Principais: Fundamentos de uma técnica de análise de dados multivariada aplicável a dados geológicos. **Acta Geológica Leopoldensia**, São Leopoldo, v. 20, n.44, p. 27-50, 1997.
- BANKER, R.D.; CHARNES A.; COOPER, W.W. Some models for estimation technical and scale inefficiencies. **Data Envelopment Analysis. Management Science**, v. 30, n. 9, p. 1078-1092, set. 1984. Disponível em: <https://doi.org/10.1287/mnsc.30.9.1078>. Acesso em: 14 out. 2019.
- BASNIAK, M. I. **Teoria Rough Sets e Análise de Correlação Canônica: Aplicação na Avaliação de Recursos Tecnológicos do Ensino Básico**. 2009. 137 f. Dissertação (Mestrado em Métodos Numéricos em Engenharia) – Setor de Ciências Exatas, Universidade Federal do Paraná, Curitiba (PR), 2009. Disponível em: <http://www.ppge.ufpr.br>. Acesso em: 14 out. 2020.
- BELLONI, I. Uma Metodologia de Avaliação da Eficiência Produtiva de Universidades Federais Brasileiras. 2000. 246 f. Tese (Doutorado em Engenharia de Produção) – Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis (SC), 2000. Disponível em: <https://repositorio.ufsc.br/bitstream/>. Acesso em: 15 jul. 2019.
- BORBA, J. T. Uma Metodologia para Avaliar a Eficiência Técnica do Ensino de Programas de Pós Graduação: uma aplicação aos programas das engenharias III da CAPES. 2011. 173 f. Tese (Doutorado em Engenharia de Produção) – Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis (SC), 2011. Disponível em: <https://repositorio.ufsc.br/bitstream/>. Acesso em: 20 out. 2018.
- BRITO, J. A. M.; OCHI, L. S.; BRITO, L. R.; MONTENEGRO, F. M. T. Um Algoritmo para o Agrupamento Baseado em K-Medoids. **Revista Brasileira de Estatística**, Rio de Janeiro, v. 71, n. 234, p. 75-100, jan./dez. 2010.
- CAPES, Documento de Área (Engenharias III). Ministério da Educação, 2017.
- CARVALHO, F. R. D.; **Análise Fatorial**. 2013. 61 f. Dissertação (Mestrado em Matemática) – Faculdade de Ciências e Tecnologia, Universidade de Coimbra, 2013. Disponível em: <https://estudogeral.sib.uc.pt/bitstream/10316/33696/1/Analise%20fatorialFranciscoCarvalho.pdf>. Acesso em: 03 jan. 2020.
- CHARNES, A.; COOPER, W. W.; RHODES, E. Measuring the efficiency of decision making units. **European Journal of Operational Research**, v. 2, p. 429-444, nov. 1978. Disponível em: [https://doi.org/10.1016/0377-2217\(78\)90138-8](https://doi.org/10.1016/0377-2217(78)90138-8). Acesso em: 03 jan. 2020.
- CHAVES NETO, A. **Análise Multivariada Aplicada a Pesquisa**. Notas de aula. Departamento de Estatística, Universidade Federal do Paraná, Curitiba, 2004.

L. CHURILOV, L.; FLITMAN, A. Towards fair ranking of Olympics achievements: the case of Sidney 2000. **Computers e Operational Research**, v.33, p. 2057-2082, jul. 2006. Disponível em: <https://doi.org/10.1016/j.cor.2004.09.027>. Acesso em: 03 jan. 2020.

COOPER, W. W.; SEIFORD, L. M.; TONE, K. **Data Envelopment Analysis: a comprehensive text with models, applications, references, and DEA-Solver software**. Boston: Kluwer Academic Publishers, 2000.

CORRAR, L. J.; PAULO, E; DIAS FILHO, J. M. D. **Análise Multivariada para Cursos de Administração, Ciências Contábeis e Economia**. São Paulo: Atlas, 2009.

COSTA, G. G. O. Um procedimento inferencial para Análise Fatorial utilizando as técnicas Bootstrap e Jackknife: Construção de intervalos de confiança e teste de hipóteses. 2006. Tese (Doutorado em Engenharia Elétrica) – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro (RJ), 2006. Disponível em: <https://doi.org/10.17771/PUCRio.acad.8741>. Acesso em: 04 fev. 2020.

COTTRELL, M., VERLEYSEN, M., Advances in Self-Organizing Maps. **Neural Networks** 19, v. 1, p. 721-722, 2006. Disponível em: <https://doi.org/10.1016/j.neunet.2006.05.011>. Acesso em: 27 out. 2019.

CRUZ, C. D.; REGAZZI, A. J. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa: UFV, 2014.

DOMINGUES, A. S.; CARVALHO, M. L. S.; MORAIS NETO, J. W. V.; CASTRO, A. F. Comparação entre os algoritmos k-means e Dynamic Cluster em imagens digitais. **Periódicos Ufersa**, Mossoró, n. 2, p. 27-35, mai. 2018. Disponível em: <https://periodicos.ufersa.edu.br/index.php/ecop/issue/view/196> . Acesso em: 27 out. 2019.

DZIUBAN, C. D.; SHIRKEY, E. C. When is a correlation matrix appropriate for factor analysis? Some decision rules. **Psychological Bulletin**, v. 81, n. 6, p. 358-361, 1974. Disponível em: <https://doi.org/10.1037/h0036316>. Acesso em: 07 nov. 2020.

FÁVERO, L. P.; BELFIORE, P.; SILVA, F. L.; CHAN, B. L. **Análise de dados: modelagem multivariada para tomada de decisões**. Rio de Janeiro: Elsevier, 2009.

FARREL, M. J. The measurement of productive efficiency. **Journal of the Royal Statistical Society**. Series A. v.120, n.3, p.253-290, 1957. Disponível em: <https://www.jstor.org/stable/2343100>. Acesso em: 07 nov. 2020.

FAUSET, L. V. **Fundamentals of Neural Networks: Architectures, Algorithms, and Applications**. New Jersey, Prentice Hall, 1994.

FERREIRA DA SILVA, P. **Uso de Rede de Kohonen para a clusterização de objetos de aprendizagem**. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Presbiteriana Mackenzie, São Paulo, 2007. Disponível em: <http://tede.mackenzie.br/jspui/handle/tede/2751>. Acesso em: 05 mai. 2019.

FLECK, M. P.A.; BOURDEL, M. C. Método de simulação e escolha de fatores na análise dos principais componentes. **Revista Saúde Pública** [online]. 1998, vol.32, n.3, p. 267-272. Disponível em: <http://dx.doi.org/10.1590/S0034-89101998000300010>. Acesso em: 05 mai. 2019.

E. M. FURTADO, E.M.; CHAVES NETO, A.; DOMINGUES, Z. H. Ranqueamento de faxinais do estado do Paraná através da Análise Fatorial. **Revista Ciências exatas e Naturais**. Unicentro, v.5, n.1, 2003.

GONTIJO, C.; AGUIRRE, A. Elementos para uma tipologia do uso do solo agrícola no Brasil: uma aplicação da Análise Fatorial. **Revista Brasileira de Economia**, Rio de Janeiro v. 42, n. 1, p. 13-49, jan./mar. 1988. Disponível em: <https://bibliotecadigital.fgv.br/ojs/index.php/rbe/article/view/399>. Acesso em: 25 set. 2019.

GUERREIRO, A. S. **Análise da Eficiência de Empresas de Comércio Eletrônico usando Técnicas da Análise Envoltória de Dados**. Dissertação (Mestrado em Logística). Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro (RJ), 2006. Disponível em: <https://www.maxwell.vrac.puc-rio.br/9973/99731.PDF>. Acesso em: 04 nov. 2018.

HAIR, Joseph F. Jr.; ANDERSON, Rolph E.; TATHAN, Ronald L.; BLACK, William C. **Análise Multivariada de dados**. Tradução Adonai Schlup Sant'Anna e Anselmo Chaves Neto. Porto Alegre: Bookman, 2005.

HAYKIN, S. **Neural Networks – A Comprehensive foundation**. Canadá: Prentice Hall, 1998.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. New Jersey: Prentice Hall, 1998.

KAISER, Henry F. The application of electronic computers to factor analysis. **Educational and Psychological Measurement**, Illinois, v. 20, n. 1, abr. 1960. Disponível em: <https://doi.org/10.1177/001316446002000116>. Acesso em: 23 set. 2020.

KAPLUNOVSKY, A.S. Why using factor analysis? (dedicated to the centenary of factor analysis). Israel, 2009. Disponível em: <http://www.magniel.com/fa/kaplunovsky.pdf>. Acesso em: 23 set. 2020.

KAUFFMAN, L.; ROUSSEEUW, P. J. **Finding Groups in Data – An Introduction to Cluster Analysis**. Wiley-Interscience Publication, 2005.

KOHONEN, T. Self-Organizing formation of topologically correct feature maps. **Biological Cybernetics**, v. 43, p. 59-69, 1989. Disponível em: <https://doi.org/10.1007/BF00337288>. Acesso em: 05 fev. 2019.

KOHONEN, T. Self-Organizing Maps: optimization approaches. In: KOHONEN, T.; MAKISARA, K.; SIMULA, O.; KANGAS, J. (Ed.). **Artificial neural networks**. Amsterdam: North-Holland, 1991. p. 981–990.

KUBRUSLY, L. S. **O Modelo de Análise Fatorial**. Dissertação (Mestrado em Ciências) – Setor de Engenharia, Universidade Federal do Rio de Janeiro, Rio de Janeiro (RJ), 1981. Disponível em: <https://www.cos.ufrj.br/uploadfile/1364229159.pdf>. Acesso em: 20 out. 2019.

LOIOLA, R. R.; M.; BONI, M. H. L.; NEVES, C. E. V. Mapas Auto-Organizáveis Kohonen para Reconhecimento de Padrões: Uma aplicação à segmentação de mercado. **X Encontro de Modelagem Computacional**, Rio de Janeiro, 2007.

MARQUES, J. M. **Análise multivariada**. Notas de Aula. Departamento de Estatística, Universidade Federal do Paraná, Curitiba, 2003.

MARRIOT, F. H. C. **The interpretation of multiple observations**. New York: Academic Press, 1974.

MAROCO, J. **Análise Estatística – Com Utilização do SPSS**. Lisboa: Sílabo, 2007.

MARTINHO, V. R. de C. Sistema inteligente para predição do grupo de risco de evasão discente. Tese (Doutorado em Engenharia Elétrica) – Faculdade de Engenharia, Universidade Estadual Paulista Julio de Mesquita Filho, Ilha Solteira, 2014. Disponível em: <https://repositorio.unesp.br/handle/11449/100340>. Acesso em: 21 mar. 2019.

MENEZES, A. C. F.; FAISSOL, S.; FERREIRA, M. L. Análise da matriz geográfica: estruturas e inter-relações. **IBGE. Tendências atuais na geografia urbano-regional: teorização e quantificação**. Rio de Janeiro, p. 67-109, 1978.

MINGOTI, S. A. **Análise de Dados Através de Métodos de Estatística Multivariada: uma abordagem aplicada**. Belo Horizonte: UFMG, 2005.

MIRANDA, A. C.; RODRIGUES, S. C. O uso da DEA como ferramenta alternativa da Gestão Escolar na Avaliação Institucional. **Educação: Teoria e prática**, São Paulo, v. 10, n.35, jul./dez. 2010.

MORRISON, D. F. **Multivariate statistical methods**. New York: McGraw-Hill Company, 1976.

OLIVEIRA, L. O. L. Mapas Auto-Organizáveis de Kohonen aplicados ao mapeamento de ambientes de robótica móvel. Dissertação (Mestrado em Computação) – Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis (SC), 2001. Disponível em: <http://repositorio.ufsc.br/xmlui/handle/123456789/79539>. Acesso em: 13 dez. 2020.

PIMENTEL, E. P.; FRANÇA, V. F.; OMAR, N. A Identificação de grupos de aprendizes no ensino presencial utilizando técnicas de clusterização. **Simpósio Brasileiro De Informática Na Educação**. n. 14, 2003.

REGAZZI, A.J. **Análise multivariada**. Notas de Aula, Departamento de Informática da Universidade Federal de Viçosa, v. 2, 2000.

THURSTONE, L. L. **The vector of mind**. Chicago: University of Chicago, 1935.

ULTSCH, A.; VETTER, C. Self-Organizing Feature Maps versus Statistical Clustering Methods: A Benchmark. **Research Report**. Marbug, n. 9, jan. 1994.

VASCONCELOS, S. **Análise de Componentes Principais (PCA)**. Notas de Aula. Universidade Federal Fluminense. Disponível em: <http://www.ic.uff.br/~aconci/PCA-ACP.pdf>. Acesso em: 07 jun. 2019.

VESANTO, J.; HIMBERG, J.; ALHONIEMI, E.; PARHANKANGAS, J. **SOM Toolbox for Matlab 5**. Helsinki: Technical Report, 2000.

ZHANG, X.; L. I. Y.; Self-Organizing Map as a New Method for Clustering and Data Analysis. **International Conference on Neural Networks**, Nagoya, v. 3, p. 2448-2451, out. 1993.

ZUCHINI, M. H., **Aplicações de mapas auto-organizáveis em mineração de dados e recuperação de informação**. Dissertação (Mestrado em Engenharia Elétrica), Faculdade de Engenharia Elétrica e Computação, Universidade Estadual de Campinas, Campinas (SP), 2003. Disponível em: <https://www.dca.fee.unicamp.br/~vonzuben/research/zuchinimest.html>. Acesso em: 12 ago. 2019.

APÊNDICE 1: PROGRAMA DESENVOLVIDO EM R PARA O ALGORITMO PROPOSTO

```

initial.dir<-getwd()
setwd("C:/Estatistica/TESE ")
library(corrplot) #biblioteca para visualizar as correlações
library("FactoMineR")
library(ggplot2)
library(factoextra)
library(psych)
library(psychTools)
library(GPArotation)
library(dplyr)
library(stats)
library(psy)
library(corrplot)
library(plotly)

sink("AFSOM.out", append = FALSE)
coolBlueHotRed <- function(n, alpha = 1) {rainbow(n, end=4/6, alpha=alpha)[n:1]}
vDados <- read.csv("PPG_QUADRIENIO2016.csv", sep=";", dec=".", header=TRUE)

summary(vDados) #estatísticas descritivas

# Matriz de correlação dos dados
correl <- cor(vDados) # matriz de correlação
round(correl, 2)

#gráfico da correlação, quanto maior o círculo maior a correlação
corrplot(cor(vDados), method = "pie")

##ESTATÍSTICA KMO PARA ADEQUAÇÃO DA AMOSTRA À AF

## Função para calcular as correlações parciais
#####
p.corr <- function(correl){
  Rinv <- solve(correl)
  Vp <- 1/sqrt(diag(Rinv))
  Vp <- diag(Vp)
  Rp <- -Vp%*%Rinv%*%Vp
  diag(Rp) <- 0
  Rp
}

## Função para calcular o coeficiente KMO
#####
calcula.kmo <- function(correl){
  Rp <- p.corr(correl)
  p <- nrow(correl)
  somar <- 0
  somap <- 0

```

```

somar.j <- rep(0,p)
somap.j <- rep(0,p)
for(i in 1:p){
  somar <- somar + sum(correl[i,]^2)-1
  somap <- somap + sum(Rp[i,]^2)
  somar.j[i] <- sum(correl[i,]^2)-1
  somap.j[i] <- sum(Rp[i,]^2)
}
kmo <- somar/(somar+somap)
msa <- somar.j/(somar.j+somap.j)
cat("\n KMO = ",round(kmo,4),"\n")
list("kmo"=round(kmo,4),"msa"=round(msa,4),
"Rp"=round(Rp,3))
}

kmo <- calcula.kmo(correl)
"Estatística KMO"
kmo

"Teste de Esfericidade de Bartlett"
cortest.bartlett(correl, n=1033)

##Os fatores serão obtidos (ou extraídos) aplicando o método dos componentes principais com
a
#matriz de correlações amostral.

####ANALISE FATORIAL
#####
acpcor <- prcomp(vDados, scale = TRUE)
acpcor
summary(acpcor)

#screeplot
plot(1:ncol(vDados), acpcor$sdev^2, type = "b", xlab = "Componente",
ylab = "Variância", pch = 20, cex.axis = 1.3, cex.lab = 1.3)

res.pca <- prcomp(correl, scale = TRUE)
fviz_eig(res.pca)

fviz_pca_var(res.pca,
  geom = c("arrow", "text"),
  col.var = "cos2", # Color by the quality of representation
  gradient.cols = c("#00AFBB", "#BFB330", "#E7B800", "#FC4E07", "#FF0000"),
  repel = TRUE, # Avoid text overlapping
  col.circle = "grey70",
  col.quanti.sup = "blue",
  legend.title = "Contribuição"
)

```

```

'CARREGAMENTOS FATORIAIS'
#####
k<- 12
carfat <- acpcor$rotation[, 1:k] %*% diag(acpcor$sdev[1:k])
colnames(carfat) <- paste("Fator",1:k, sep = " ")
carfat

"rotação varimax"
#####
carfatr <- varimax(carfat)
carfatr$loadings
write.table(carfatr$loadings, file='carregamentos_fatoriais_rotacionados.csv', sep=';', dec='.',
            row.names=FALSE)

#DADOS ORIGINAIS PADRONIZADOS
#####
score_z = scale(vDados, center = T)

"ESCORES FATORIAIS"
#####
car <- t(carfatr$loadings) %*% t(score_z)
escores<- t(car)
escores

#CARGAS DOS FATORES 1 E 2 APÓS ROTAÇÃO VARIMAX
plot(carfatr$loadings, pch = 20, col = "red", xlab = "Fator 1", ylab = "Fator 2")
text(carfatr$loadings, rownames(carfat), adj = 1)

"COMUNALIDADES"
#####
comun<-rowSums(carfatr$loadings^2)
varEsp<-1-comun
varEsp
plot(acpcor, type = "l", pch = 16, main = "Scree plot")

write.table(comun, file='comunalidades.csv', sep=';', dec='.', row.names=FALSE)

#LINHA DE AUTOVALORES DA MATRIZ DE CORRELAÇÃO
autovalores<- as.data.frame(t(eigen(correl)$values))

autovalor<- autovalores[c(1:k)]
"autovalores"
autovalores
autovalor

"VARIÂNCIA EXPLICADA"
#####
varexp <- (sum(autovalor))/(sum(autovalores))
varexp

```

```

vacu <- autovalor/(sum(autovalores))

write.table(t(vacu), file='varexplicada.csv', sep=';', dec='.', row.names=FALSE)

#TRANFORMANDO DO DATAFRAME EM MATRIZ PARA MULTIPLICAÇÃO
autovalorr<- as.matrix(t(autovalor))

#MULTIPLICAÇÃO DA MATRIZ DE ESCORES PELOS AUTOVALORES
Escore_Fatorial = (escores %*% autovalorr)/(sum(autovalor))

"ESCORES FINAIS PONDERADOS"
#####
ESCPONDERADO<- data.frame(Escore_Fatorial)
ESCPONDERADO

#COLOCAR EM ORDEM DECRESCENTE
n1<- 1:92
n2<- ESCPONDERADO
#escore normalizado entre 0 e 1
n3<-((n2-min(n2))/(max(n2)-min(n2)))
C <- cbind(n1, n2, n3)
C
D <- cbind(n1, n3)

"CLASSIFICAÇÃO PELA ANÁLISE FATORIAL"
#####
D %>%
  arrange(desc(n3))

write.table(C, file='escores padronizado.csv', sep=';', dec='.', row.names=FALSE)

#ESTRUTURAR OS DADOS PARA A NORMALIZAÇÃO PARA O TREINAMENTO DA REDE
SOM
escores = data.frame(escores)

write.table(escores, file='escores fatoriais rotacionados.csv', sep=';', dec='.', row.names=FALSE)

###KOHONEN
#####
library("kohonen")

#NORMALIZAÇÃO EM 2 PASSOS#

#---> Normalização 1: Normalizar os dados de entrada - Padronização das variáveis: (x - xmedia)
/ desvio-padrão
#-----> Para a variavel FATOR1...
vMedia = mean(escores$Fator.1)
vDesvPad = sd(escores$Fator.1) # desvio padrão da amostra
vNormalizacao = ((escores$Fator.1) - vMedia) / vDesvPad
escores = cbind(escores,vNormalizacao)

```

```
names(escores)[names(escores)=="vNormalizacao"] = "FATOR1_N2" # renomear a coluna
```

```
#-----> Para a variavel FATOR2...
```

```
vMedia = mean(escores$Fator.2)
```

```
vDesvPad = sd(escores$Fator.2) # desvio padrão da amostra
```

```
vNormalizacao = ((escores$Fator.2) - vMedia) / vDesvPad
```

```
escores = cbind(escores,vNormalizacao)
```

```
names(escores)[names(escores)=="vNormalizacao"] = "FATOR2_N2" # renomear a coluna
```

```
#-----> Para a variavel FATOR3...
```

```
vMedia = mean(escores$Fator.3)
```

```
vDesvPad = sd(escores$Fator.3) # desvio padrão da amostra
```

```
vNormalizacao = (escores$Fator.3 - vMedia) / vDesvPad
```

```
escores = cbind(escores,vNormalizacao)
```

```
names(escores)[names(escores)=="vNormalizacao"] = "FATOR3_N2" # renomear a coluna
```

```
#-----> Para a variavel FATOR4...
```

```
vMedia = mean(escores$Fator.4)
```

```
vDesvPad = sd(escores$Fator.4) # desvio padrão da amostra
```

```
vNormalizacao = (escores$Fator.4 - vMedia) / vDesvPad
```

```
escores = cbind(escores,vNormalizacao)
```

```
names(escores)[names(escores)=="vNormalizacao"] = "FATOR4_N2" # renomear a coluna
```

```
#-----> Para a variavel FATOR5...
```

```
vMedia = mean(escores$Fator.5)
```

```
vDesvPad = sd(escores$Fator.5) # desvio padrão da amostra
```

```
vNormalizacao = (escores$Fator.5 - vMedia) / vDesvPad
```

```
escores = cbind(escores,vNormalizacao)
```

```
names(escores)[names(escores)=="vNormalizacao"] = "FATOR5_N2" # renomear a coluna
```

```
#-----> Para a variavel FATOR6...
```

```
vMedia = mean(escores$Fator.6)
```

```
vDesvPad = sd(escores$Fator.6) # desvio padrão da amostra
```

```
vNormalizacao = (escores$Fator.6 - vMedia) / vDesvPad
```

```
escores = cbind(escores,vNormalizacao)
```

```
names(escores)[names(escores)=="vNormalizacao"] = "FATOR6_N2" # renomear a coluna
```

```
#-----> Para a variavel FATOR7...
```

```
vMedia = mean(escores$Fator.7)
```

```
vDesvPad = sd(escores$Fator.7) # desvio padrão da amostra
```

```
vNormalizacao = (escores$Fator.7 - vMedia) / vDesvPad
```

```
escores = cbind(escores,vNormalizacao)
```

```
names(escores)[names(escores)=="vNormalizacao"] = "FATOR7_N2" # renomear a coluna
```

```
#-----> Para a variavel FATOR8...
```

```
vMedia = mean(escores$Fator.8)
```

```
vDesvPad = sd(escores$Fator.8) # desvio padrão da amostra
```

```
vNormalizacao = (escores$Fator.8 - vMedia) / vDesvPad
```

```
escores = cbind(escores,vNormalizacao)
```

```
names(escores)[names(escores)=="vNormalizacao"] = "FATOR8_N2" # renomear a coluna
```

```
#-----> Para a variavel FATOR9...
vMedia = mean(escores$Fator.9)
vDesvPad = sd(escores$Fator.9) # desvio padrão da amostra
vNormalizacao = (escores$Fator.9 - vMedia) / vDesvPad
escores = cbind(escores,vNormalizacao)
names(escores)[names(escores)=="vNormalizacao"] = "FATOR9_N2" # renomear a coluna
```

```
#-----> Para a variavel FATOR10...
vMedia = mean(escores$Fator.10)
vDesvPad = sd(escores$Fator.10) # desvio padrão da amostra
vNormalizacao = (escores$Fator.10 - vMedia) / vDesvPad
escores = cbind(escores,vNormalizacao)
names(escores)[names(escores)=="vNormalizacao"] = "FATOR10_N2" # renomear a coluna
```

```
#-----> Para a variavel FATOR11...
vMedia = mean(escores$Fator.11)
vDesvPad = sd(escores$Fator.11) # desvio padrão da amostra
vNormalizacao = (escores$Fator.11 - vMedia) / vDesvPad
escores = cbind(escores,vNormalizacao)
names(escores)[names(escores)=="vNormalizacao"] = "FATOR11_N2" # renomear a coluna
```

```
#-----> Para a variavel FATOR12...
vMedia = mean(escores$Fator.12)
vDesvPad = sd(escores$Fator.12) # desvio padrão da amostra
vNormalizacao = (escores$Fator.12 - vMedia) / vDesvPad
escores = cbind(escores,vNormalizacao)
names(escores)[names(escores)=="vNormalizacao"] = "FATOR12_N2" # renomear a coluna
```

#--> Normalização 2: Normalizar os dados já normalizados (_N2): função sigmoidal $1 / (1 + (\exp(-\text{var_N2})))$

```
#-----> Para a variavel FATOR1...
vNormalizacao = 1 / (1 + exp(-escores$FATOR1_N2))
escores = cbind(escores,vNormalizacao)
names(escores)[names(escores)=="vNormalizacao"] = "FATOR1_N3" # renomear a coluna
```

```
#-----> Para a variavel FATOR2...
vNormalizacao = 1 / (1 + exp(-escores$FATOR2_N2))
escores = cbind(escores,vNormalizacao)
names(escores)[names(escores)=="vNormalizacao"] = "FATOR2_N3" # renomear a coluna
```

```
#-----> Para a variavel FATOR3...
vNormalizacao = 1 / (1 + exp(-escores$FATOR3_N2))
escores = cbind(escores,vNormalizacao)
names(escores)[names(escores)=="vNormalizacao"] = "FATOR3_N3" # renomear a coluna
```

```
#-----> Para a variavel FATOR4...
vNormalizacao = 1 / (1 + exp(-escores$FATOR4_N2))
escores = cbind(escores,vNormalizacao)
names(escores)[names(escores)=="vNormalizacao"] = "FATOR4_N3" # renomear a coluna
```

```

#-----> Para a variavel FATOR5...
vNormalizacao = 1 / (1 + exp(-escores$FATOR5_N2))
escores = cbind(escores,vNormalizacao)
names(escores)[names(escores)=="vNormalizacao"] = "FATOR5_N3" # renomear a coluna

#-----> Para a variavel FATOR6...
vNormalizacao = 1 / (1 + exp(-escores$FATOR6_N2))
escores = cbind(escores,vNormalizacao)
names(escores)[names(escores)=="vNormalizacao"] = "FATOR6_N3" # renomear a coluna

#-----> Para a variavel FATOR7...
vNormalizacao = 1 / (1 + exp(-escores$FATOR7_N2))
escores = cbind(escores,vNormalizacao)
names(escores)[names(escores)=="vNormalizacao"] = "FATOR7_N3" # renomear a coluna

#-----> Para a variavel FATOR8...
vNormalizacao = 1 / (1 + exp(-escores$FATOR8_N2))
escores = cbind(escores,vNormalizacao)
names(escores)[names(escores)=="vNormalizacao"] = "FATOR8_N3" # renomear a coluna

#-----> Para a variavel FATOR9...
vNormalizacao = 1 / (1 + exp(-escores$FATOR9_N2))
escores = cbind(escores,vNormalizacao)
names(escores)[names(escores)=="vNormalizacao"] = "FATOR9_N3" # renomear a coluna

#-----> Para a variavel FATOR10...
vNormalizacao = 1 / (1 + exp(-escores$FATOR10_N2))
escores = cbind(escores,vNormalizacao)
names(escores)[names(escores)=="vNormalizacao"] = "FATOR10_N3" # renomear a coluna

#-----> Para a variavel FATOR11...
vNormalizacao = 1 / (1 + exp(-escores$FATOR11_N2))
escores = cbind(escores,vNormalizacao)
names(escores)[names(escores)=="vNormalizacao"] = "FATOR11_N3" # renomear a coluna

#-----> Para a variavel FATOR12...
vNormalizacao = 1 / (1 + exp(-escores$FATOR12_N2))
escores = cbind(escores,vNormalizacao)
names(escores)[names(escores)=="vNormalizacao"] = "FATOR12_N3" # renomear a coluna

#---> Selecionar somente os campos numéricos normalizados pelo terceiro tipo (N3)
vKOH =
  subset(escores,select=c("FATOR1_N3","FATOR2_N3","FATOR3_N3","FATOR4_N3","FA
    TOR5_N3","FATOR6_N3","FATOR7_N3","FATOR8_N3","FATOR9_N3","FATOR10_N3","
    FATOR11_N3","FATOR12_N3"))
#---> O algoritmo só funciona se os dados estiverem em formato matrix... então vamos
  transformar em matrix...
vKOH2 = as.matrix(vKOH)
vKOH2

```

```

#
# Executar o algoritmo SOM (Self Organizing Mpas)
#####
#--> Definir o lay-out da camada (grid de neurônios) "Kohonen"
#####
vGrid_som = somgrid(xdim=3, ydim=2, topo="hexagonal")
#--> executa efetivamente o algoritmo...
set.seed(7)
vMapaKohonen = som(vKOH2,
grid = vGrid_som,
rlen = 10000,
alpha = c(0.05,0.01),
keep.data = TRUE)
vMapaKohonen

cat("\n\n")
cat("Após a execução do algoritmo de treinamento do mapa de Kohonen temos:")
cat("\n\n")

#EVOLUÇÃO DO TREINAMENTO
dev.new() # Or X11()
dev.1 <- as.integer(dev.cur())
dev.set(dev.1)
plot(vMapaKohonen, type="changes", main = "Desempenho dos Programas - treinamento")
plot(vMapaKohonen, type="changes")

dev.2 <- as.integer(dev.cur())
dev.set(dev.2)
dev.new()
plot(vMapaKohonen, type="counts", main = "Total de objetos (padrões) alocados pela rede")

#MATRIZ U
dev.new()
dev.3 <- as.integer(dev.cur())
dev.set(dev.3)
plot(vMapaKohonen, type="dist.neighbours", main = "U-Matrix - Matriz de distância aos neurônios
vizinhos")

#--> Gráfico do vetor de pesos...
#####
#####vCores = c("#FF0000", "#00FF00", "#0000FF", "#FF0F0F", "#E0E0E0", "#FF00EE")
vCores <- colorRampPalette(
c('#33CDC7','#F2716b','#E1F505','#00FF00', '#574DD8', '#FF00EE', '#BFBC30')
)
dev.new() # Or X11()
dev.4 <- as.integer(dev.cur())
dev.set(dev.4)
plot(vMapaKohonen, type="codes", palette.name=vCores, main = "Vetor de pesos dos
neurônios")

```

```
#####
#----> Resultados impressos....
#####
cat("(0) Resumo geral\n")
cat("-----\n")
print(summary(vMapaKohonen))
cat("\n")
cat("=====\n")
cat("\n")
cat("(1) Neurônio vencedor do padrão apresentado")
cat("\n")
cat("-----")
cat("\n")
vNeuronioVencedor = vMapaKohonen$unit.classif # resultado do algoritmo
vTotInputs = nrow(vKOH)
cat("\n")
cat("PPG | FATOR1 | FATOR2 | FATOR3 | FATOR4 | FATOR5 | FATOR6 | FATOR7 | FATOR8 |
      FATOR9 | FATOR10 | FATOR11 | FATOR12 | Neurônio vencedor"); cat("\n");
cat("----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
      -----"); cat("\n");
for (k in 1:vTotInputs) {
  vAux = escores[k,"PPG"]
  vAux = format(vAux, width = 3, justify="left")
  cat(" "); cat(vAux); cat(" | ")
  vAux = escores[k,"Fator.1"]
  vAux = format(vAux, width = 1, justify="right")
  cat(""); cat(vAux); cat(" | ")
  vAux = escores[k,"Fator.2"]
  vAux = format(vAux, width = 1, justify="right")
  cat(""); cat(vAux); cat(" | ")
  vAux = escores[k,"Fator.3"]
  vAux = format(vAux, width = 1, justify="right")
  cat(""); cat(vAux); cat(" | ")
  vAux = escores[k,"Fator.4"]
  vAux = format(vAux, width = 1, justify="right")
  cat(""); cat(vAux); cat(" | ")
  vAux = escores[k,"Fator.5"]
  vAux = format(vAux, width = 1, justify="right")
  cat(""); cat(vAux); cat(" | ")
  vAux = escores[k,"Fator.6"]
  vAux = format(vAux, width = 1, justify="right")
  cat(""); cat(vAux); cat(" | ")
  vAux = escores[k,"Fator.7"]
  vAux = format(vAux, width = 1, justify="right")
  cat(""); cat(vAux); cat(" | ")
  vAux = escores[k,"Fator.8"]
  vAux = format(vAux, width = 1, justify="right")
  cat(""); cat(vAux); cat(" | ")
  vAux = escores[k,"Fator.9"]
  vAux = format(vAux, width = 1, justify="right")
  cat(""); cat(vAux); cat(" | ")
}
```

```

vAux = escores[k,"Fator.10"]
vAux = format(vAux, width = 1, justify="right")
cat(""); cat(vAux); cat(" | ")
vAux = escores[k,"Fator.11"]
vAux = format(vAux, width = 1, justify="right")
cat(""); cat(vAux); cat(" | ")
vAux = escores[k,"Fator.12"]
vAux = format(vAux, width = 1, justify="right")
cat(""); cat(vAux); cat(" | ")
vAux = vMapaKohonen$unit.classif[k]
vAux = format(vAux, width = 3, justify="right")
cat(" "); cat(vAux); cat(" | ")
cat("\n")
}
cat("\n")
cat("2) A matriz de vetores-peso (code vectors) do mapa de Kohonen")
cat("\n")
cat("-----" )
cat("\n")
cat("\n\n")
print(vMapaKohonen$codes)
cat("3) Gráfico do vetor de pesos")
cat("\n")
cat("-----")
cat("\n")
cat("\n")
cat("-> Em cada neurônio é apresentado seu vetor de pesos que o liga à camada de entrada de
      dados")
cat("\n")
cat("-> A imagem do vetor de pesos dentro do neurônio é conhecida como 'diagrama de
      ventilador'")
cat("\n")
cat("-> Quanto maior o setor circular de um elemento do vetor de peso, maior é o seu valor")
cat("\n")
cat("-> Quanto menor o setor circular de um elemento do vetor de peso, menor é o seu valor")
cat("\n")
cat("-> Note que o vetor de pesos representa, em última instância, um padrão do conjunto de
      dados de entrada")
cat("\n\n")
cat("\n\n")
cat("4) Gráfico de distância dos vizinhos (gráfico U-Matrix)")
cat("\n")
cat("-----" )
cat("\n")
cat("\n")
cat("-> Conhecido como gráfico 'U-Matrix', permite visualizar a distância entre cada neurônio e
      seus vizinhos")
cat("\n")
cat("-> Áreas com baixa distância de vizinhos indica grupo de neurônios similares")
cat("\n")
cat("-> Áreas com alta distância de vizinhos indica grupo de neurônios disimilares")

```

```

cat("\n")
cat("-> Áreas com alta distância de vizinhos indica então limites 'naturais' entre clusters")
cat("\n")
cat("-> Desta forma, a U-Matrix serve para identificar clusters no mapa de Kohonen")
cat("\n\n")
cat("\n\n")
cat("11) Distância dos padrões (PPGs) ao correspondente neurônio vencedor")
cat("\n")
cat("-----")
cat("\n")
vNeuronioVencedor = vMapaKohonen$unit.classif # resultado do algoritmo
vTotInputs = nrow(vKOH)
vDistanciaAoNeuronioVencedor = vMapaKohonen$distances
cat("PPG | FATOR1 | FATOR2 | FATOR3 | FATOR4 | FATOR5 | FATOR6 | FATOR7 | FATOR8 |
    FATOR9 | FATOR10 | FATOR11 | FATOR12 | Neurônio vencedor | Distância |Escore
    Fatorial"); cat("\n");
cat("----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
    -----|-----|-----"); cat("\n");
for (k in 1:vTotInputs) {
vAux = escores[k,"1"]
vAux = format(vAux, width = 7, justify="left")
cat(" "); cat(vAux); cat(" | ")
vAux = escores[k,"Fator.1"]
vAux = format(vAux, width = 2, justify="right")
cat(""); cat(vAux); cat(" | ")
vAux = escores[k,"Fator.2"]
vAux = format(vAux, width = 2, justify="right")
cat(""); cat(vAux); cat(" | ")
vAux = escores[k,"Fator.3"]
vAux = format(vAux, width = 2, justify="right")
cat(""); cat(vAux); cat(" | ")
vAux = escores[k,"Fator.4"]
vAux = format(vAux, width = 2, justify="right")
cat(""); cat(vAux); cat(" | ")
vAux = escores[k,"Fator.5"]
vAux = format(vAux, width = 2, justify="right")
cat(""); cat(vAux); cat(" | ")
vAux = escores[k,"Fator.6"]
vAux = format(vAux, width = 2, justify="right")
cat(""); cat(vAux); cat(" | ")
vAux = escores[k,"Fator.7"]
vAux = format(vAux, width = 2, justify="right")
cat(""); cat(vAux); cat(" | ")
vAux = escores[k,"Fator.8"]
vAux = format(vAux, width = 2, justify="right")
cat(""); cat(vAux); cat(" | ")
vAux = escores[k,"Fator.9"]
vAux = format(vAux, width = 2, justify="right")
cat(""); cat(vAux); cat(" | ")
vAux = escores[k,"Fator.10"]
vAux = format(vAux, width = 2, justify="right")

```

```
cat(""); cat(vAux); cat(" | ")
vAux = escores[k,"Fator.11"]
vAux = format(vAux, width = 2, justify="right")
cat(""); cat(vAux); cat(" | ")
vAux = escores[k,"Fator.12"]
vAux = format(vAux, width = 2, justify="right")
cat(""); cat(vAux); cat(" | ")
vAux = vMapaKohonen$unit.classif[k]
vAux = format(vAux, width = 3, justify="right")
cat(" "); cat(vAux); cat(" | ")
vAux = vMapaKohonen$distances[k]
vAux = sprintf("%3.6f", vAux)
vAux = format(vAux, width = 11, justify="right")
cat(" "); cat(vAux); cat(" | ")
vAux = n3[k,"Escore_Fatorial"]
vAux = format(vAux, width = 2, justify="right")
cat(""); cat(vAux); cat(" | ")

cat("\n")
}
cat("\n")
cat("\n")
cat("-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x-x-")
cat("\n")
cat("\n")
cat("==> Unit distances \n\n ")
vDists <- unit.distances(vMapaKohonen$grid, toroidal=FALSE)
print(vMapaKohonen$grid$n.hood)
print("/n/n acimaa /n")
print(vDists)
cat("\n")
cat("----- \n\n")
#-----
detach("package:kohonen")

"Matriz de pesos após o treinamento (SOM)"
vMapaKohonen$codes
autovalor

write.table(vMapaKohonen$codes, file='matriz_pesos.csv', sep=';', dec='.', row.names=FALSE)

vDa <- read.csv("matriz_pesos.csv", sep=";", dec=".", header=TRUE)
vDa <- as.matrix(vDa)
"teste vda"
vDa
autt <- as.matrix(autovalor)

"Classificação dos clusters (SOM)"
#####
pesos <- vDa%*(t(vacu))
```

```

pesos <- as.data.frame(pesos)
clusterr <- c(1:8)
tabP <- cbind(clusterr, pesos)
tabP

tabPP <- tabP %>%
  arrange(desc(pesos))
tabPP
write.table(tabPP, file='ClassCluster.csv', sep=';', dec='.', row.names=FALSE)

#Neurônio Vencedor
classcluster <- as.matrix(vMapaKohonen$unit.classif)
#Vetor das instituições
ppgs <- read.csv("instituicoes.csv", sep=";", dec=".", header=TRUE)
ppgs <- as.matrix(ppgs)
ies <- read.csv("IES.csv", sep=";", dec=".", header=TRUE)
ies <- as.matrix(ies)
capes <- read.csv("capes.csv", sep=";", dec=".", header=TRUE)
capes <- as.matrix(capes)

tabCLASS <- cbind(n1, ppgs, ies, classcluster, n3, capes)
tabCLASS

rr <- tabCLASS %>%
  arrange(desc(n3))

rr
write.table(rr, file='resultadoAF.csv', sep=';', dec='.', row.names=FALSE)

#CLASSIFICAÇÃO PELO SOM
#####
rr2 <- tabCLASS %>%
  arrange(desc(classcluster))
"rr2"
rr2$n1
write.table(rr2, file='resultadoSOM.csv', sep=';', dec='.', row.names=FALSE)

"TAMANHO DOS CLUSTERS"
#####
cont <- table(classcluster)
cont

"KMEDIAS"
#####
kk=kmeans(vKOH2, 6, iter.max=10000, nstart=1)

tabCLASSK <- cbind(n1, ppgs, ies, kk$cluster, n3, capes)
tabCLASSK

```

```
write.table(tabCLASSK, file='kmeans_cluster.csv', sep=';', dec='.', row.names=FALSE)

df=scale(vKOH2)
head(df, n=8)

km.res=kmeans(df, 6, nstart=25)

fviz_cluster(km.res, data=vKOH2,
  palette = c("#2E9FDF", "#00AFBB", "#E7B800", "#FC4E07", "#1a9b1e", "#210a08"),
  ellipse.type="euclid",
  star.plot=TRUE,
  repel=TRUE,
  ggtheme=theme_minimal()
)

fviz_nbclust(df, kmeans, method="gap_stat")

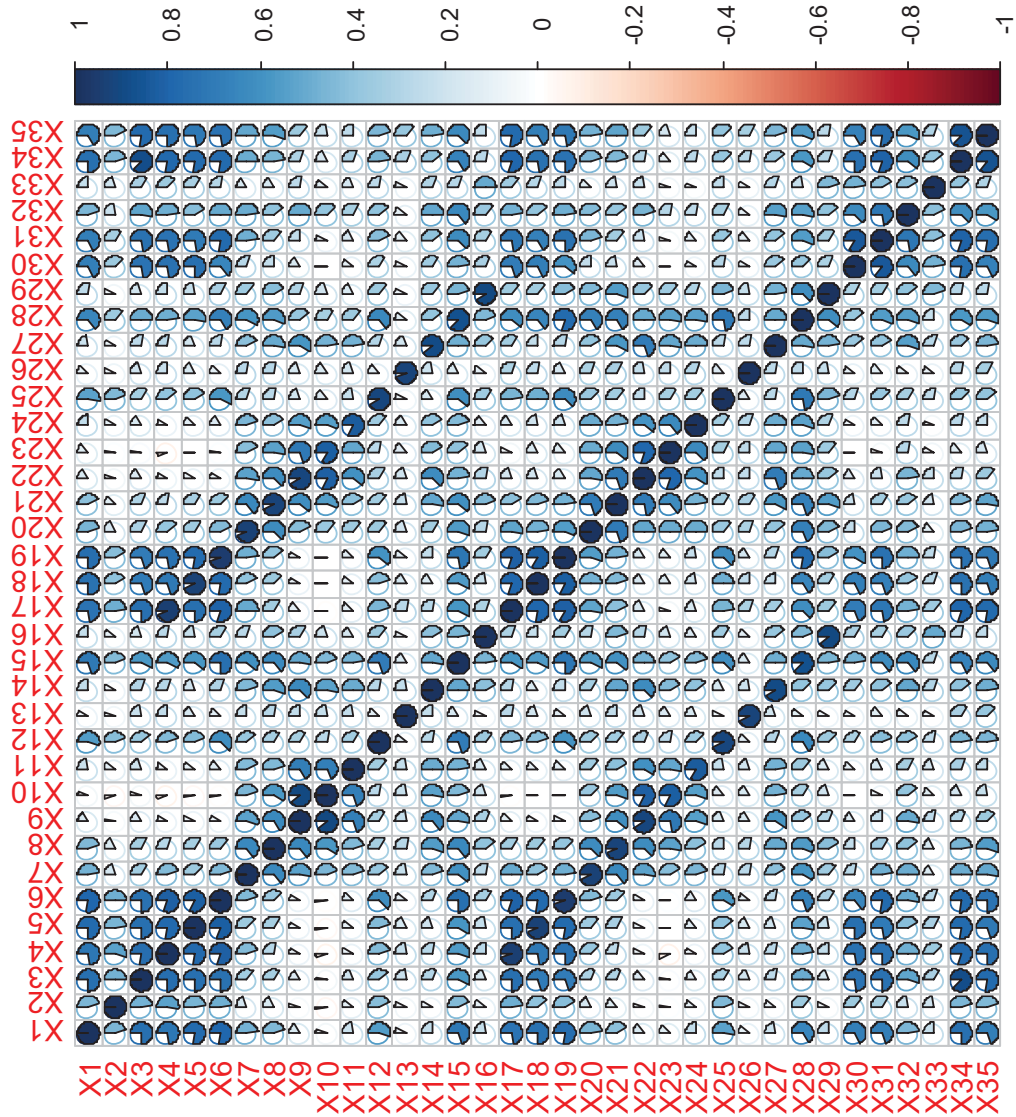
fviz_cluster(km.res, data=df, ellipse.type="convex",
  palette="jco", ggtheme=theme_minimal())

setwd(initial.dir)
sink()
```

APÊNDICE 2 – MATRIZ DE CORRELAÇÃO DAS 35 VARIÁVEIS

1.30	0.45	0.75	0.73	0.77	0.79	0.47	0.44	0.16	0.05	0.22	0.55	0.09	0.24	0.71	0.23	0.73	0.75	0.77	0.47	0.41	0.16	0.13	0.24	0.53	0.11	0.22	0.66	0.26	0.68	0.46	0.23	0.75	0.68		
									-																										
0.45	1.00	0.48	0.54	0.47	0.49	0.19	0.17	0.04	0.04	0.11	0.42	0.07	0.08	0.43	0.10	0.48	0.41	0.43	0.18	0.13	0.06	0.01	0.10	0.44	0.07	0.09	0.37	0.07	0.36	0.22	0.20	0.46	0.44		
0.75	0.48	1.00	0.77	0.76	0.75	0.34	0.33	0.16	0.05	0.20	0.38	0.23	0.29	0.58	0.20	0.76	0.71	0.70	0.31	0.30	0.15	0.03	0.17	0.33	0.21	0.26	0.49	0.19	0.73	0.74	0.53	0.33	0.88	0.78	
									-													-													
0.73	0.54	0.77	1.00	0.78	0.81	0.43	0.28	0.09	0.06	0.10	0.44	0.25	0.29	0.56	0.27	0.95	0.72	0.74	0.38	0.26	0.07	0.07	0.09	0.40	0.23	0.27	0.48	0.25	0.74	0.73	0.49	0.36	0.79	0.75	
									-																										
0.77	0.47	0.76	0.78	1.00	0.82	0.37	0.34	0.08	0.03	0.11	0.44	0.17	0.21	0.59	0.24	0.76	0.93	0.77	0.35	0.32	0.09	0.00	0.13	0.41	0.18	0.19	0.53	0.24	0.75	0.71	0.42	0.32	0.77	0.71	
									-																										
0.79	0.49	0.75	0.81	0.82	1.00	0.44	0.37	0.09	0.02	0.14	0.62	0.13	0.30	0.75	0.33	0.80	0.77	0.95	0.44	0.36	0.10	0.01	0.15	0.57	0.12	0.27	0.68	0.32	0.65	0.77	0.51	0.28	0.77	0.74	
0.47	0.19	0.34	0.43	0.37	0.44	1.00	0.65	0.53	0.45	0.44	0.39	0.23	0.44	0.61	0.28	0.51	0.41	0.48	0.93	0.65	0.53	0.42	0.45	0.29	0.27	0.41	0.58	0.35	0.28	0.48	0.49	0.17	0.49	0.51	
0.44	0.17	0.33	0.28	0.34	0.37	0.65	1.00	0.65	0.38	0.44	0.33	0.29	0.55	0.64	0.41	0.35	0.37	0.40	0.62	0.93	0.63	0.53	0.41	0.27	0.29	0.31	0.56	0.42	0.24	0.36	0.40	0.17	0.47	0.55	
0.16	0.04	0.16	0.09	0.08	0.09	0.53	0.65	1.00	0.90	0.68	0.34	0.24	0.63	0.51	0.32	0.12	0.09	0.09	0.43	0.61	0.91	0.71	0.55	0.20	0.18	0.58	0.39	0.30	0.16	0.25	0.50	0.26	0.31	0.36	
																-																			
0.35	0.04	0.05	0.06	0.03	0.02	0.45	0.58	0.90	1.00	0.68	0.28	0.23	0.53	0.45	0.21	0.01	0.00	0.01	0.37	0.55	0.82	0.82	0.54	0.16	0.17	0.48	0.34	0.21	0.01	0.09	0.39	0.13	0.19	0.23	
0.22	0.11	0.20	0.10	0.11	0.14	0.44	0.44	0.68	0.68	1.00	0.35	0.24	0.50	0.48	0.18	0.12	0.11	0.13	0.36	0.41	0.60	0.52	0.84	0.23	0.19	0.47	0.36	0.17	0.12	0.19	0.33	0.19	0.32	0.26	
0.55	0.42	0.38	0.44	0.44	0.62	0.39	0.33	0.34	0.28	0.35	1.00	0.13	0.28	0.69	0.36	0.46	0.42	0.60	0.37	0.35	0.33	0.27	0.35	0.92	0.12	0.29	0.65	0.37	0.32	0.43	0.40	0.27	0.46	0.45	
0.39	0.07	0.23	0.25	0.17	0.13	0.23	0.29	0.24	0.23	0.24	0.13	1.00	0.24	0.17	0.09	0.28	0.19	0.13	0.20	0.24	0.21	0.16	0.17	0.08	0.93	0.20	0.13	0.09	0.09	0.14	0.11	0.10	0.33	0.38	
0.24	0.08	0.29	0.29	0.21	0.30	0.44	0.55	0.63	0.53	0.50	0.28	0.24	1.00	0.51	0.41	0.29	0.16	0.25	0.35	0.52	0.62	0.43	0.41	0.18	0.22	0.90	0.39	0.38	0.35	0.40	0.51	0.30	0.40	0.49	
0.71	0.43	0.58	0.56	0.59	0.75	0.61	0.64	0.51	0.45	0.48	0.69	0.17	0.51	1.00	0.47	0.58	0.57	0.73	0.56	0.61	0.48	0.41	0.44	0.61	0.16	0.49	0.87	0.46	0.47	0.63	0.65	0.29	0.67	0.64	
0.23	0.10	0.20	0.27	0.24	0.33	0.28	0.41	0.32	0.21	0.18	0.36	0.09	0.41	0.47	1.00	0.31	0.23	0.35	0.28	0.45	0.29	0.20	0.15	0.36	0.12	0.39	0.45	0.90	0.32	0.33	0.38	0.50	0.28	0.26	
									-																										
0.73	0.48	0.76	0.95	0.76	0.80	0.51	0.35	0.12	0.01	0.12	0.46	0.28	0.29	0.58	0.31	1.00	0.79	0.81	0.53	0.42	0.19	0.06	0.18	0.47	0.31	0.34	0.61	0.37	0.70	0.72	0.48	0.35	0.79	0.77	
0.75	0.41	0.71	0.72	0.93	0.77	0.41	0.37	0.09	0.00	0.11	0.42	0.19	0.16	0.57	0.23	0.79	1.00	0.82	0.48	0.45	0.18	0.17	0.21	0.49	0.26	0.25	0.66	0.34	0.69	0.67	0.40	0.29	0.74	0.69	
0.77	0.43	0.70	0.74	0.77	0.95	0.48	0.40	0.09	0.01	0.13	0.60	0.13	0.25	0.73	0.35	0.81	0.82	1.00	0.56	0.48	0.17	0.14	0.22	0.62	0.18	0.30	0.78	0.42	0.60	0.75	0.49	0.26	0.74	0.71	
0.47	0.18	0.31	0.38	0.35	0.44	0.93	0.62	0.43	0.37	0.36	0.37	0.20	0.35	0.56	0.28	0.53	0.48	0.56	1.00	0.71	0.52	0.50	0.50	0.37	0.32	0.39	0.68	0.44	0.24	0.45	0.43	0.13	0.46	0.48	
0.41	0.13	0.30	0.26	0.32	0.36	0.65	0.93	0.61	0.55	0.41	0.35	0.24	0.52	0.61	0.45	0.42	0.45	0.48	0.71	1.00	0.70	0.62	0.48	0.34	0.32	0.57	0.67	0.55	0.23	0.34	0.39	0.21	0.45	0.52	
0.16	0.06	0.15	0.07	0.09	0.10	0.53	0.63	0.91	0.82	0.60	0.33	0.21	0.62	0.48	0.29	0.19	0.18	0.17	0.52	0.70	1.00	0.79	0.60	0.26	0.24	0.70	0.51	0.38	0.16	0.23	0.47	0.24	0.30	0.36	
0.13	0.01	0.03	0.07	0.00	0.01	0.42	0.53	0.71	0.82	0.52	0.27	0.16	0.43	0.41	0.20	0.06	0.17	0.14	0.50	0.62	0.79	1.00	0.64	0.34	0.26	0.32	0.53	0.35	0.01	0.05	0.28	0.06	0.14	0.19	
0.24	0.10	0.17	0.09	0.13	0.15	0.45	0.41	0.55	0.54	0.84	0.35	0.17	0.41	0.44	0.15	0.18	0.21	0.22	0.50	0.48	0.60	0.64	1.00	0.34	0.29	0.49	0.53	0.31	0.08	0.14	0.27	0.06	0.27	0.24	
0.53	0.44	0.33	0.40	0.41	0.57	0.29	0.27	0.20	0.16	0.23	0.92	0.08	0.18	0.61	0.36	0.47	0.49	0.62	0.37	0.34	0.26	0.34	1.00	0.15	0.27	0.72	0.45	0.29	0.39	0.34	0.26	0.38	0.39		
0.11	0.07	0.21	0.23	0.18	0.12	0.27	0.29	0.18	0.17	0.19	0.12	0.93	0.22	0.16	0.12	0.31	0.26	0.18	0.32	0.32	0.24	0.26	0.34	1.00	0.15	1.00	0.27	0.26	0.23	0.10	0.13	0.12	0.09	0.28	0.33
0.22	0.09	0.26	0.27	0.19	0.27	0.41	0.51	0.58	0.48	0.47	0.29	0.20	0.90	0.49	0.39	0.34	0.25	0.30	0.39	0.57	0.70	0.52	0.49	0.27	0.27	1.00	0.51	0.47	0.37	0.40	0.55	0.28	0.38	0.45	
0.56	0.37	0.49	0.48	0.53	0.68	0.58	0.56	0.39	0.34	0.36	0.65	0.13	0.39	0.87	0.45	0.61	0.66	0.78	0.68	0.67	0.51	0.53	0.53	0.72	0.26	0.31	1.00	0.62	0.39	0.55	0.23	0.58	0.56		
0.26	0.07	0.19	0.25	0.24	0.32	0.35	0.42	0.30	0.21	0.17	0.37	0.09	0.38	0.46	0.90	0.37	0.34	0.42	0.44	0.55	0.38	0.35	0.31	0.45	0.23	0.47	0.62	1.00	0.31	0.34	0.41	0.44	0.29	0.28	
0.58	0.36	0.73	0.74	0.75	0.65	0.28	0.24	0.16	0.01	0.12	0.32	0.09	0.35	0.47	0.32	0.70	0.69	0.60	0.24	0.23	0.16	0.01	0.08	0.29	0.10	0.37	0.39	0.31	1.00	0.84	0.63	0.47	0.76	0.67	
0.58	0.33	0.74	0.73	0.71	0.77	0.48	0.36	0.25	0.09	0.19	0.43	0.14	0.40	0.63	0.33	0.72	0.67	0.75	0.45	0.34	0.23	0.05	0.14	0.39	0.13	0.40	0.55	0.34	0.84	1.00	0.73	0.42	0.80	0.77	
0.46	0.22	0.53	0.49	0.42	0.51	0.49	0.40	0.50	0.39	0.33	0.40	0.11	0.51	0.65	0.38	0.48	0.40	0.49	0.43	0.39	0.47	0.28	0.27	0.34	0.12	0.55	0.55	0.41	0.63	0.73	1.00	0.41	0.61	0.58	
0.23	0.20	0.33	0.36	0.32	0.28	0.17	0.17	0.26	0.13	0.19	0.27	0.10	0.30	0.29	0.50	0.35	0.29	0.26	0.13	0.21	0.24	0.06	0.06	0.26	0.09	0.28	0.23	0.44	0.47	0.42	0.41	1.00	0.39	0.30	
0.75	0.46	0.88	0.79	0.77	0.77	0.49	0.47	0.31	0.19	0.32	0.46	0.33	0.40	0.67	0.28	0.79	0.74	0.74	0.46	0.39	0.14	0.27	0.38	0.29	0.38	0.58	0.28	0.39	0.76	0.80	0.61	0.39	1.00	0.86	
0.58	0.44	0.78	0.75	0.71	0.74	0.51	0.55	0.36	0.23	0.26	0.45	0.38	0.49	0.64	0.26	0.77	0.69	0.71	0.48	0.52	0.36	0.19	0.24	0.39	0.33	0.45	0.56	0							

APÊNDICE 3 – GRÁFICO DA CORRELAÇÃO R DAS 35 VARIÁVEIS



APÊNDICE 4 – MATRIZ DE ESCORES FATORIAIS NORMALIZADOS UTILIZADOS PARA O TREINAMENTO DA REDE SOM

PPG	Fator 1	Fator 2	Fator 3	Fator 4	Fator 5	Fator 6	Fator 7	Fator 8	Fator 9	Fator 10	Fator 11	Fator 12
1	0,4084	0,4132	0,4012	0,5501	0,3690	0,5360	0,3876	0,6400	0,4311	0,5859	0,4918	0,4615
2	0,4556	0,4499	0,4418	0,5472	0,3870	0,5699	0,3949	0,5500	0,4611	0,4550	0,4921	0,5094
3	0,3256	0,2524	0,3013	0,5378	0,2504	0,7286	0,2431	0,7286	0,2885	0,3000	0,3687	0,3012
4	0,5476	0,4048	0,4437	0,5699	0,3725	0,4757	0,5415	0,5987	0,4054	0,8740	0,5711	0,4352
5	0,3626	0,5606	0,5620	0,5943	0,6141	0,6141	0,6675	0,4392	0,4091	0,4294	0,4597	0,4036
6	0,6716	0,4991	0,5421	0,4645	0,4452	0,4269	0,4584	0,4882	0,6817	0,8248	0,5096	0,3804
7	0,2516	0,2427	0,2788	0,6705	0,2280	0,7641	0,2308	0,7376	0,2282	0,2664	0,4966	0,2753
8	0,4112	0,3624	0,4118	0,5486	0,2981	0,6509	0,4242	0,6362	0,4032	0,5159	0,6378	0,5474
9	0,3065	0,4080	0,4789	0,5995	0,3175	0,7124	0,4149	0,4570	0,2838	0,2710	0,5005	0,4187
10	0,2866	0,2588	0,2800	0,6798	0,3240	0,7179	0,2775	0,7459	0,2789	0,4305	0,4906	0,3080
11	0,3460	0,3852	0,3231	0,6376	0,3666	0,5054	0,3422	0,6162	0,4046	0,3896	0,5416	0,4617
12	0,5265	0,5034	0,4562	0,2762	0,8100	0,5481	0,3736	0,3689	0,5548	0,3722	0,6739	0,6586
13	0,4629	0,3510	0,3168	0,6200	0,3508	0,5109	0,3469	0,6578	0,3828	0,4883	0,5217	0,5638
14	0,3272	0,3091	0,2939	0,6722	0,3468	0,6952	0,2762	0,6566	0,3217	0,3034	0,5848	0,2653
15	0,3551	0,2971	0,3529	0,5925	0,3225	0,6818	0,3045	0,6803	0,3185	0,4321	0,5660	0,3062
16	0,5811	0,5451	0,4813	0,4946	0,5295	0,4784	0,6344	0,4589	0,4746	0,4614	0,5425	0,4751
17	0,4288	0,3090	0,3457	0,4859	0,3526	0,6477	0,3117	0,6619	0,3576	0,3390	0,5530	0,4799
18	0,7645	0,4530	0,5480	0,4290	0,5501	0,3841	0,4311	0,4276	0,5448	0,6692	0,6379	0,8221
19	0,5114	0,3488	0,3627	0,4314	0,4529	0,6450	0,3058	0,6333	0,3173	0,5540	0,6439	0,4660
20	0,6439	0,6151	0,7993	0,2679	0,5170	0,5126	0,4875	0,2231	0,5236	0,4843	0,9178	0,6628
21	0,5517	0,3758	0,9957	0,3071	0,4085	0,6470	0,3965	0,5478	0,4121	0,4775	0,4284	0,3188
22	0,5246	0,7835	0,6534	0,4737	0,6695	0,2229	0,7417	0,3179	0,6121	0,4719	0,6897	0,6106
23	0,5479	0,3803	0,4224	0,5980	0,4639	0,5075	0,3953	0,5641	0,5910	0,4455	0,5321	0,3517
24	0,3063	0,2864	0,2721	0,6615	0,2609	0,5770	0,3552	0,7457	0,2823	0,3215	0,4945	0,3072
25	0,3649	0,3914	0,4915	0,6478	0,3593	0,5813	0,4045	0,5165	0,3484	0,4486	0,5621	0,3319
26	0,6361	0,3653	0,4432	0,4793	0,4953	0,5054	0,3340	0,6100	0,4210	0,5856	0,4574	0,4111
27	0,4445	0,4528	0,3781	0,2892	0,6413	0,5927	0,3425	0,5316	0,4250	0,3259	0,7007	0,6338
28	0,3816	0,3165	0,3398	0,4037	0,3809	0,6232	0,2779	0,6793	0,3836	0,3080	0,5111	0,4009
29	0,2400	0,2385	0,2486	0,6464	0,2171	0,7746	0,2161	0,7658	0,2224	0,2365	0,4695	0,2899
30	0,3412	0,4596	0,4379	0,5592	0,4123	0,6429	0,3970	0,5982	0,4158	0,2747	0,5993	0,4359
31	0,2837	0,4623	0,3847	0,7047	0,3239	0,6815	0,4305	0,6387	0,3296	0,3855	0,4219	0,4066
32	0,2503	0,2905	0,3175	0,5608	0,2930	0,7350	0,2633	0,7520	0,2319	0,3306	0,5055	0,3362
33	0,3614	0,2689	0,3074	0,4992	0,3150	0,7177	0,2383	0,7234	0,2843	0,3939	0,4709	0,3046
34	0,6649	0,6089	0,5692	0,6898	0,5449	0,2820	0,7877	0,4318	0,7346	0,7827	0,6192	0,4396
35	0,4418	0,3371	0,4076	0,4593	0,3996	0,6275	0,3084	0,5615	0,3957	0,3964	0,5542	0,3597
36	0,5260	0,4268	0,5264	0,4714	0,4908	0,4116	0,4496	0,5347	0,4575	0,7239	0,4126	0,6270
37	0,5530	0,5488	0,6693	0,6370	0,5448	0,5044	0,4281	0,5038	0,6580	0,6109	0,4545	0,4813
38	0,3670	0,3748	0,5041	0,5865	0,3247	0,5926	0,4061	0,6407	0,3423	0,4566	0,3011	0,3386
39	0,8193	0,6175	0,6499	0,4205	0,7037	0,2793	0,5938	0,1916	0,8453	0,7810	0,7731	0,9002
40	0,3783	0,3384	0,3263	0,4595	0,3661	0,6540	0,4989	0,6448	0,3018	0,3339	0,4690	0,4722
41	0,3807	0,3555	0,3712	0,6323	0,3298	0,6280	0,4387	0,6424	0,4257	0,4089	0,5480	0,4262
42	0,2775	0,2844	0,2814	0,6400	0,2607	0,7335	0,2958	0,7257	0,2557	0,2835	0,5240	0,3304
43	0,6772	0,8571	0,8479	0,7627	0,6893	0,4418	0,6896	0,1948	0,8168	0,4632	0,9494	0,4780
44	0,3110	0,4197	0,3486	0,4912	0,3892	0,7003	0,3203	0,6833	0,3983	0,2973	0,5555	0,3901
45	0,9448	0,6999	0,6734	0,0127	0,9300	0,0769	0,6616	0,0792	0,7717	0,9360	0,5271	0,9707
46	0,3942	0,2979	0,3079	0,4404	0,3471	0,6837	0,2505	0,6924	0,2869	0,3144	0,4640	0,4301

PPG	Fator 1	Fator 2	Fator 3	Fator 4	Fator 5	Fator 6	Fator 7	Fator 8	Fator 9	Fator 10	Fator 11	Fator 12
47	0,5805	0,7725	0,7346	0,3819	0,7924	0,4435	0,5997	0,1829	0,7273	0,7311	0,8049	0,4859
48	0,7865	0,5460	0,5474	0,4235	0,6724	0,2572	0,5518	0,2909	0,5843	0,7919	0,7098	0,7174
49	0,5855	0,5978	0,6675	0,2763	0,7271	0,4893	0,5448	0,2941	0,6244	0,5006	0,7286	0,6547
50	0,9786	0,7426	0,9786	0,1414	0,7994	0,4893	0,8494	0,1847	0,8610	0,9368	0,5884	0,9080
51	0,5221	0,7321	0,3297	0,2186	0,4874	0,6108	0,3085	0,6441	0,3963	0,4597	0,5537	0,5767
52	0,4146	0,5308	0,4911	0,6770	0,5054	0,5505	0,5014	0,5319	0,4738	0,6104	0,4742	0,2887
53	0,5278	0,4928	0,4201	0,5976	0,7093	0,3519	0,5500	0,4318	0,6055	0,3914	0,4922	0,5285
54	0,4849	0,6335	0,6028	0,4307	0,8017	0,5164	0,5622	0,3795	0,6294	0,3501	0,5932	0,3800
55	0,9188	0,8710	0,9560	0,5052	0,7996	0,0671	0,8995	0,2501	0,9523	0,9687	0,5984	0,5968
56	0,4747	0,4158	0,4685	0,5088	0,4870	0,4407	0,5380	0,5898	0,4513	0,5545	0,5932	0,4898
57	0,3985	0,6232	0,5273	0,3111	0,5423	0,5327	0,6785	0,4860	0,4226	0,3281	0,8166	0,7662
58	0,9206	0,7084	0,6584	0,1901	0,9416	0,0154	0,7534	0,2675	0,9763	0,9763	0,5856	0,6847
59	0,9668	0,9334	0,9367	0,4510	0,9767	0,0353	0,8809	0,0900	0,9825	0,7995	0,0712	0,7252
60	0,7954	0,8179	0,6981	0,8459	0,8709	0,1061	0,8145	0,3924	0,9194	0,7250	0,1754	0,2909
61	0,3772	0,3332	0,2999	0,6580	0,3309	0,6692	0,3079	0,6868	0,3177	0,3324	0,5873	0,4572
62	0,5899	0,8599	0,7565	0,3325	0,8529	0,2284	0,8122	0,1100	0,8172	0,4416	0,1309	0,7610
63	0,3463	0,4150	0,4197	0,7393	0,4136	0,4777	0,3849	0,6952	0,3692	0,4784	0,0268	0,2434
64	0,4523	0,6192	0,5406	0,5561	0,6717	0,3596	0,7056	0,4124	0,5769	0,4649	0,1134	0,6849
65	0,3431	0,3159	0,3739	0,7039	0,3372	0,5819	0,3733	0,6947	0,3290	0,4658	0,5806	0,2886
66	0,5606	0,6909	0,6060	0,6025	0,5145	0,4927	0,7315	0,3695	0,5832	0,6567	0,6675	0,4519
67	0,5671	0,4149	0,4949	0,4724	0,4565	0,5448	0,3582	0,5266	0,6197	0,4147	0,4943	0,5776
68	0,3008	0,4999	0,6265	0,6665	0,3223	0,6575	0,6150	0,6204	0,3715	0,3730	0,5122	0,3873
69	0,4019	0,2688	0,3223	0,4585	0,3194	0,6614	0,2763	0,7084	0,3481	0,4160	0,4698	0,3793
70	0,4837	0,6043	0,4906	0,5787	0,6001	0,4999	0,6482	0,2692	0,5484	0,4832	0,5018	0,4629
71	0,3328	0,2668	0,3234	0,6178	0,2402	0,7205	0,2533	0,7722	0,3727	0,3435	0,3366	0,2138
72	0,5374	0,4181	0,3789	0,6679	0,4448	0,5290	0,4658	0,5719	0,5074	0,3809	0,5333	0,5028
73	0,4192	0,7642	0,7143	0,4659	0,5111	0,5229	0,9373	0,1837	0,6034	0,3439	0,4975	0,5551
74	0,3165	0,3252	0,3244	0,6466	0,3000	0,5048	0,3879	0,6850	0,3068	0,3478	0,4813	0,2788
75	0,2861	0,3259	0,2771	0,6969	0,2656	0,6567	0,3699	0,7176	0,3006	0,3615	0,5115	0,2846
76	0,2651	0,3656	0,2924	0,6950	0,3006	0,7160	0,3767	0,6854	0,3517	0,2563	0,4522	0,3135
77	0,9253	0,6397	0,7074	0,0323	0,7982	0,2136	0,5845	0,2386	0,8025	0,9511	0,6605	0,9079
78	0,7585	0,9921	0,8508	0,3429	0,9040	0,1122	0,9813	0,0258	0,9452	0,6961	0,7140	0,9420
79	0,3392	0,4379	0,5282	0,7658	0,4841	0,5503	0,6363	0,6477	0,4712	0,3497	0,0183	0,2594
80	0,8172	0,8005	0,7180	0,1877	0,7731	0,2492	0,7670	0,1013	0,7695	0,7017	0,5586	0,9824
81	0,7088	0,8960	0,7846	0,3344	0,7705	0,3458	0,8580	0,0792	0,7678	0,5756	0,0584	0,6754
82	0,5267	0,8949	0,7405	0,3686	0,6312	0,3087	0,9451	0,3119	0,8009	0,4160	0,7946	0,8644
83	0,3802	0,5933	0,4899	0,6344	0,4173	0,5399	0,5468	0,4820	0,4297	0,3647	0,2160	0,4629
84	0,2570	0,2550	0,2753	0,7054	0,2323	0,7588	0,2355	0,7673	0,3004	0,2434	0,5156	0,2727
85	0,4523	0,7747	0,5487	0,0102	0,8067	0,4802	0,7668	0,2463	0,5530	0,4309	0,6167	0,6576
86	0,3424	0,3994	0,3721	0,5588	0,4891	0,5828	0,3865	0,6415	0,3993	0,3264	0,1513	0,4233
87	0,3071	0,4446	0,3406	0,5928	0,4499	0,6674	0,3715	0,6371	0,3560	0,3244	0,4159	0,3681
88	0,3158	0,3290	0,2910	0,7242	0,2772	0,6795	0,3189	0,7022	0,2808	0,3185	0,5024	0,2883
89	0,2587	0,2772	0,2784	0,6229	0,2414	0,7337	0,2745	0,7461	0,2293	0,2849	0,5166	0,3337
90	0,3630	0,3791	0,3331	0,6201	0,3349	0,6176	0,4363	0,6133	0,3610	0,5519	0,6192	0,3945
91	0,4232	0,3621	0,3456	0,6713	0,3194	0,5675	0,3805	0,6654	0,3591	0,5058	0,4892	0,3180
92	0,3203	0,3433	0,2961	0,7178	0,2875	0,6583	0,3734	0,6819	0,2889	0,3505	0,4478	0,3511

ANEXO 1 – NÚMEROS E NOMES DOS PROGRAMAS DE PÓS-GRADUAÇÃO(PPG's)

Número	PPG	IES	Número	PPG	IES
1	ENGENHARIA MECÂNICA	UFPA	41	ENGENHARIA MECÂNICA	UFSJ
2	ENGENHARIA DE RECURSOS NATURAIS DA AMAZÔNIA	UFPA	42	ENGENHARIA DA ENERGIA - CEFET-MG - UFSJ	CEFET/MG
3	ENGENHARIA MECÂNICA	UFC	43	ENGENHARIA DE PRODUÇÃO M/D	UFSCAR
4	ENGENHARIA MECÂNICA	UFRN	44	ENGENHARIA DE PRODUÇÃO M	UFSCAR
5	ENGENHARIA DE PRODUÇÃO	UFRN	45	ENGENHARIA MECÂNICA	USP
6	CIÊNCIA E ENGENHARIA DE PETRÓLEO	UFRN	46	ENGENHARIA NAVAL E OCEÂNICA	USP
7	ENGENHARIA MECATRÔNICA	UFRN	47	ENGENHARIA (ENGENHARIA DE PRODUÇÃO)	USP
8	ENGENHARIA MECÂNICA	UFPB/I.P.	48	ENGENHARIA MECÂNICA	USP/SC
9	ENGENHARIA DE PRODUÇÃO	UFPB/I.P.	49	ENGENHARIA DE PRODUÇÃO	USP/SC
10	ENERGIAS RENOVÁVEIS	UFPB/I.P.	50	ENGENHARIA MECÂNICA	UNICAMP
11	ENGENHARIA MECÂNICA	UFCG	51	CIÊNCIAS E ENGENHARIA DE PETRÓLEO	UNICAMP
12	ENGENHARIA DE PRODUÇÃO	UFPE	52	ENGENHARIA DE PRODUÇÃO E DE MANUFATURA	UNICAMP/LI
13	ENGENHARIA MECÂNICA	UFPE	53	ENGENHARIA MECÂNICA	UNESP/BAU
14	ENGENHARIA DE PRODUÇÃO - CENTRO ACADÊMICO DO AGRESTE	UFPE	54	ENGENHARIA DE PRODUÇÃO	UNESP/BAU
15	MECATRÔNICA	UFBA	55	ENGENHARIA MECÂNICA	UNESP/GUAR
16	ENGENHARIA INDUSTRIAL	UFBA	56	ENGENHARIA MECÂNICA	UNESP/IS
17	ENGENHARIA MECÂNICA	UFES	57	ENGENHARIA DE PRODUÇÃO	UNIMEP
18	ENGENHARIA MECÂNICA	UFRJ	58	ENGENHARIA E TECNOLOGIA ESPACIAIS	INPE
19	ENGENHARIA OCEÂNICA	UFRJ	59	ENGENHARIA AERONÁUTICA E MECÂNICA	ITA
20	ENGENHARIA DE PRODUÇÃO	UFRJ	47	ENGENHARIA (ENGENHARIA DE PRODUÇÃO)	USP
21	PLANEJAMENTO ENERGÉTICO	UFRJ	48	ENGENHARIA MECÂNICA	USP/SC
22	ENGENHARIA DE PRODUÇÃO	UFF	49	ENGENHARIA DE PRODUÇÃO	USP/SC
23	ENGENHARIA MECÂNICA	UFF	50	ENGENHARIA MECÂNICA	UNICAMP
24	ENGENHARIA MECÂNICA - VOLTA REDONDA	UFF	51	CIÊNCIAS E ENGENHARIA DE PETRÓLEO	UNICAMP
25	ENGENHARIA MECÂNICA	UERJ	52	ENGENHARIA DE PRODUÇÃO E DE MANUFATURA	UNICAMP/LI
26	ENGENHARIA MECÂNICA	PUC-RIO	53	ENGENHARIA MECÂNICA	UNESP/BAU
27	ENGENHARIA DE PRODUÇÃO	PUC-RIO	54	ENGENHARIA DE PRODUÇÃO	UNESP/BAU
28	METROLOGIA	PUC-RIO	55	ENGENHARIA MECÂNICA	UNESP/GUAR
29	ENGENHARIA MECÂNICA	IME	56	ENGENHARIA MECÂNICA	UNESP/IS
30	ENGENHARIA DE PRODUÇÃO E SISTEMAS	CEFET/RJ	57	ENGENHARIA DE PRODUÇÃO	UNIMEP
31	ENGENHARIA DE PRODUÇÃO	UCAM	58	ENGENHARIA E TECNOLOGIA ESPACIAIS	INPE
32	ENGENHARIA DE PRODUÇÃO	UENF	59	ENGENHARIA AERONÁUTICA E MECÂNICA	ITA
33	ENGENHARIA DE RESERVATÓRIO E DE EXPLORAÇÃO	UENF	60	CIÊNCIAS E TECNOLOGIAS ESPACIAIS	ITA
34	ENGENHARIA MECÂNICA	UFMG	61	ENGENHARIA MECÂNICA	FEI
35	ENGENHARIA DE PRODUÇÃO	UFMG	62	ENGENHARIA DE PRODUÇÃO	UNIP
36	ENGENHARIA MECÂNICA	UNIFEI	63	ENGENHARIA MECÂNICA	IFSP
37	ENGENHARIA DE PRODUÇÃO	UNIFEI	64	ENGENHARIA DE PRODUÇÃO	UNINOVE
38	ENGENHARIA DE ENERGIA	UNIFEI	65	ENGENHARIA MECÂNICA	UFABC
39	ENGENHARIA MECÂNICA	UFU	66	MÉTODOS NUMÉRICOS EM ENGENHARIA	UFPR
40	ENGENHARIA MECÂNICA	PUC/MG	67	ENGENHARIA MECÂNICA	UFPR

Número	PPG	IES
68	ENGENHARIA DE PRODUÇÃO	UFPR
69	ENGENHARIA MECÂNICA	PUC/PR
70	ENGENHARIA DE PRODUÇÃO E SISTEMAS	PUC/PR
71	ENGENHARIA MECÂNICA	UEM
72	ENGENHARIA MECÂNICA E DE MATERIAIS	UTFPR
73	ENGENHARIA DE PRODUÇÃO	UTFPR
74	ENGENHARIA MECÂNICA	UTFPR
75	ENGENHARIA MECÂNICA	UTFPR
76	ENGENHARIA DE PRODUÇÃO E SISTEMAS (PPGEPS)	UTFPR
77	ENGENHARIA MECÂNICA	UFSC
78	ENGENHARIA DE PRODUÇÃO	UFSC
79	ENGENHARIA MECÂNICA	UDESC
80	ENGENHARIA MECÂNICA	UFRGS
81	ENGENHARIA DE PRODUÇÃO	UFRGS
82	ENGENHARIA DE PRODUÇÃO	UFSC
83	ENGENHARIA OCEÂNICA	FURG
84	ENGENHARIA MECÂNICA	FURG
85	ENGENHARIA DE PRODUÇÃO E SISTEMAS	UNISINOS
86	ENGENHARIA MECÂNICA	UNISINOS
87	SISTEMAS E PROCESSOS INDUSTRIAIS	UNISC
88	ENGENHARIA	UNIPAMPA
89	ENGENHARIA DE PRODUÇÃO E SISTEMAS	PUC-GOÍÁS
90	SISTEMAS MECATRÔNICOS	UNB
91	CIÊNCIAS MECÂNICAS	UNB
92	INTEGRIDADE DE MATERIAIS DA ENGENHARIA	UNB

FONTE: CAPES (2017).

ANEXO 2 – DADOS ORIGINAIS DAS 35 VARIÁVEIS SELECIONADAS

PPG	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁	X ₁₂	X ₁₃	X ₁₄	X ₁₅	X ₁₆	X ₁₇	X ₁₈	X ₁₉	X ₂₀	X ₂₁	X ₂₂	X ₂₃	X ₂₄
1	17	7	105	13	10	15	10	9	19	16	9	23	7	7	309	76	18	14	24	17	13	28	22	11
2	14	3	113	20	15	35	12	10	22	28	11	11	7	13	287	92	29	20	50	14	11	35	29	15
3	9	1	45	15	24	16	6	0	1	7	3	2	2	2	92	17	25	33	17	7	0	1	8	5
4	14	16	224	28	30	47	5	8	6	17	37	17	4	11	304	87	33	31	53	6	10	8	20	45
5	10	4	91	6	10	14	6	15	21	42	45	4	15	33	278	269	6	15	14	7	25	33	62	70
6	17	11	147	80	51	56	21	25	24	22	15	17	5	16	260	44	92	65	64	27	28	26	24	20
7	11	1	23	6	4	12	2	2	4	1	4	0	3	7	79	35	9	4	14	2	3	4	1	4
8	23	5	160	13	3	11	11	3	7	6	23	8	6	6	175	31	12	3	13	16	4	9	7	35
9	13	0	63	7	7	7	1	4	13	28	17	2	13	44	168	92	9	7	9	1	5	17	35	23
10	18	5	42	15	3	5	5	1	2	2	8	3	2	4	172	30	17	6	5	5	1	3	12	3
11	12	2	55	4	1	15	9	13	8	9	5	43	2	13	228	81	4	1	18	16	26	11	14	10
12	14	0	111	31	13	32	12	33	10	15	7	2	4	24	322	975	42	14	42	15	43	9	7	10
13	24	3	209	14	9	32	6	8	2	15	7	24	0	5	206	41	21	10	51	9	10	3	22	13
14	16	0	29	24	10	20	3	19	6	2	6	1	0	13	250	162	29	12	23	3	22	6	1	8
15	15	3	99	41	10	19	3	11	8	5	13	4	4	7	151	175	41	11	21	3	12	9	5	14
16	21	0	312	30	36	37	5	26	14	41	44	17	4	22	350	249	42	47	45	7	31	14	44	56
17	15	0	105	26	15	46	6	10	4	1	11	4	4	5	216	77	29	16	58	6	13	4	1	14
18	26	4	381	54	35	85	6	11	17	8	12	16	9	16	425	171	73	46	116	6	13	19	8	12
19	20	5	327	13	30	31	1	14	13	1	3	1	1	8	204	358	15	34	32	1	15	19	2	4
20	14	0	176	45	32	60	10	32	42	39	15	17	35	56	348	168	42	30	51	5	19	33	19	10
21	12	1	218	52	34	16	4	7	0	7	12	3	106	11	121	164	70	50	20	4	7	0	7	14
22	20	0	110	12	11	28	11	36	57	82	38	77	16	27	457	401	18	12	34	12	45	67	91	46
23	20	0	104	60	24	54	16	14	4	4	16	14	2	13	251	224	78	32	81	24	27	5	5	21
24	20	0	85	7	5	13	4	1	5	5	17	36	0	2	127	23	8	7	16	4	1	5	6	26
25	13	4	71	13	14	24	5	10	10	16	17	22	13	32	269	113	14	15	28	5	10	11	17	17
26	17	3	220	60	51	49	1	16	2	16	7	9	3	4	355	262	69	67	60	1	27	3	16	9
27	13	0	91	22	20	20	8	22	28	30	8	13	3	12	248	720	27	20	21	8	21	13	14	6
28	9	0	56	16	22	30	11	11	3	5	3	16	4	7	184	156	16	23	36	13	11	4	5	4
29	11	0	34	6	3	5	1	0	4	9	1	2	0	3	32	48	6	3	6	2	0	6	11	1
30	12	0	48	9	13	19	11	19	17	55	19	11	11	18	178	269	9	16	23	14	18	17	44	10
31	6	4	45	2	1	3	5	12	13	68	17	0	6	12	326	106	2	1	4	5	14	15	88	23
32	8	3	33	0	2	0	0	4	10	19	6	4	6	0	167	196	0	2	0	0	4	11	19	6
33	13	3	84	23	19	16	2	10	4	1	3	0	1	2	107	135	25	25	24	3	13	5	1	3
34	34	7	208	49	25	48	21	30	21	7	54	31	7	16	557	94	77	34	60	26	36	31	7	94
35	15	1	102	38	23	32	8	12	7	1	6	7	5	23	188	179	43	28	39	9	12	7	1	7
36	14	9	152	25	18	20	7	4	11	5	15	31	11	12	392	214	39	27	23	13	7	19	9	23
37	10	8	129	14	22	39	18	32	18	23	8	6	21	13	396	206	24	42	77	32	47	35	35	8
38	13	4	82	14	17	7	3	5	13	15	14	20	14	10	149	75	28	29	7	9	11	20	27	30
39	25	6	217	75	23	79	33	19	10	7	19	19	8	43	670	244	98	28	99	38	23	11	7	23
40	11	0	94	14	13	19	1	1	2	4	42	3	1	6	288	155	18	19	25	1	1	2	6	55
41	9	4	56	9	11	32	12	15	2	4	28	11	5	8	162	56	9	14	44	15	18	4	3	44
42	17	0	54	15	1	5	1	1	9	11	10	2	1	5	168	93	18	1	6	1	4	10	12	16
43	26	0	163	17	26	47	21	118	44	68	34	8	29	43	563	96	16	28	49	26	123	41	66	37
44	13	1	43	3	7	3	12	21	22	31	5	0	4	3	229	219	4	9	4	14	24	26	34	6
45	48	11	277	112	67	92	11	17	29	6	17	60	7	54	699	737	127	77	100	11	16	25	5	23
46	11	1	77	7	37	24	2	3	10	3	1	2	2	6	224	149	7	39	30	2	4	10	3	1

PPG	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁	X ₁₂	X ₁₃	X ₁₄	X ₁₅	X ₁₆	X ₁₇	X ₁₈	X ₁₉	X ₂₀	X ₂₁	X ₂₂	X ₂₃	X ₂₄
47	14	10	100	23	13	21	17	78	37	33	25	11	13	52	377	542	32	13	27	22	105	51	35	29
48	40	7	270	58	33	98	4	17	18	13	21	30	7	35	552	308	67	40	121	5	22	21	22	23
49	19	3	121	33	16	33	17	31	31	17	28	14	17	32	258	730	43	16	37	20	35	32	22	29
50	71	9	478	98	123	127	18	34	19	20	61	50	7	26	639	139	114	156	162	21	36	25	25	72
51	15	4	137	6	50	27	8	9	10	7	4	1	0	5	301	257	6	58	32	12	10	11	6	4
52	25	9	62	11	25	18	10	25	26	35	20	4	12	21	378	311	16	27	22	10	34	31	52	28
53	16	0	83	16	18	64	14	11	6	11	27	27	3	26	485	647	28	22	87	25	17	13	18	35
54	16	0	68	22	24	20	15	44	34	23	27	5	13	24	371	923	36	34	29	19	63	41	26	38
55	42	15	273	94	53	89	36	58	38	60	59	65	41	6	599	195	139	76	128	53	83	60	85	90
56	17	4	138	28	15	38	8	7	12	7	37	33	10	8	367	295	37	15	48	11	12	13	7	41
57	9	0	83	5	7	15	7	15	37	92	62	18	18	17	434	431	6	6	19	5	14	19	49	46
58	36	17	202	68	72	141	3	35	11	21	31	116	11	20	1059	874	77	89	188	5	45	10	32	42
59	48	0	291	62	68	107	33	70	32	38	28	45	21	25	777	579	148	148	226	64	151	72	94	72
60	56	4	164	33	50	76	33	63	16	37	24	43	5	13	808	445	49	79	112	50	86	29	137	72
61	22	0	119	12	13	11	3	13	7	8	6	4	1	5	357	75	16	15	13	3	16	7	8	7
62	12	0	91	14	10	32	27	19	39	74	28	44	12	59	380	510	40	22	62	34	56	89	121	65
63	14	5	50	3	0	5	0	2	3	23	9	6	6	0	110	123	27	50	40	12	12	8	144	13
64	11	3	51	7	3	10	12	21	16	16	39	32	4	4	367	226	11	20	23	15	27	45	55	77
65	16	5	25	10	8	41	4	9	4	6	22	18	9	7	281	136	13	8	42	4	10	4	6	23
66	27	7	132	23	29	24	10	40	41	45	55	2	11	26	446	141	24	36	30	12	43	49	53	64
67	17	1	120	34	34	43	24	12	3	10	4	9	8	18	323	112	40	41	52	31	15	4	12	4
68	10	3	47	1	4	1	6	13	23	43	50	5	25	9	290	85	2	4	1	8	15	32	52	56
69	10	3	91	28	18	43	7	1	2	0	8	3	2	2	135	98	31	26	55	10	1	2	0	11
70	12	3	118	26	23	23	11	27	17	28	38	13	4	45	385	358	32	26	26	14	38	27	34	54
71	11	3	27	16	38	15	17	2	6	2	3	3	2	0	25	17	19	51	16	15	2	10	2	5
72	25	0	176	18	38	48	13	13	6	7	21	7	3	18	450	114	23	50	55	16	15	6	7	28
73	13	0	92	10	8	12	15	21	32	55	114	10	17	59	306	151	12	12	15	21	24	40	73	144
74	14	1	29	4	15	16	5	6	8	5	18	48	4	11	167	48	4	16	19	4	7	8	4	30
75	11	3	23	4	9	18	5	8	9	17	23	14	0	6	207	25	4	9	18	5	9	11	17	25
76	11	0	44	2	1	5	9	8	13	22	18	2	0	8	174	114	2	2	7	14	14	18	29	26
77	45	13	438	114	58	74	22	27	18	23	13	24	10	23	572	332	137	74	96	26	30	19	24	13
78	24	4	183	18	19	29	33	89	139	215	88	48	12	76	856	387	24	25	37	40	112	175	250	108
79	8	2	37	7	12	12	4	1	3	2	9	2	2	4	75	178	13	22	37	23	15	10	101	131
80	28	5	198	55	37	67	17	18	44	47	39	25	16	51	591	322	60	44	85	23	21	59	57	47
81	13	3	159	32	32	32	9	41	55	77	47	9	9	42	335	174	76	73	70	19	99	115	137	91
82	15	0	81	16	12	24	32	32	93	138	105	37	20	14	613	237	15	13	23	31	35	72	96	110
83	12	1	54	12	16	19	4	11	40	40	18	11	10	16	363	82	19	21	28	6	20	81	79	38
84	12	0	29	2	6	6	9	6	2	1	2	1	2	2	120	18	2	5	7	12	6	2	1	3
85	11	2	96	10	4	4	11	30	56	62	68	18	6	36	195	718	22	8	7	15	57	77	71	55
86	8	1	53	7	2	9	4	2	10	15	6	9	1	3	201	325	18	23	32	16	13	26	47	31
87	9	2	38	6	3	9	3	21	10	47	11	3	2	8	209	366	8	4	12	8	33	16	75	21
88	17	1	41	12	12	22	0	11	6	28	11	4	1	9	271	17	14	13	27	0	13	7	33	13
89	9	1	45	2	4	9	0	0	13	11	9	7	2	3	149	53	2	4	9	0	0	12	11	9
90	12	7	142	5	8	15	8	10	9	14	26	15	1	18	270	116	6	9	18	11	11	9	14	27
91	18	4	127	20	5	24	5	9	7	21	13	18	1	9	271	37	27	33	34	5	14	9	24	23
92	13	2	62	8	12	23	1	4	5	31	17	5	1	11	303	30	9	16	26	2	6	6	34	23

FONTE: CAPES (2017).

PPG	X ₂₅	X ₂₆	X ₂₇	X ₂₈	X ₂₉	X ₃₀	X ₃₁	X ₃₂	X ₃₃	X ₃₄	X ₃₅
1	31	7	12	401	104	5	6	33	11	94	0
2	15	8	18	352	118	19	10	30	6	40	40
3	3	3	2	125	22	7	5	13	11	44	0
4	23	6	12	392	94	7	1	36	4	157	36
5	5	16	35	318	306	3	0	13	2	100	0
6	19	5	18	336	53	7	11	31	6	108	44
7	0	3	10	90	36	0	1	3	0	6	0
8	8	7	11	203	38	4	13	54	10	91	55
9	2	16	47	188	90	2	0	23	1	76	0
10	3	2	5	234	36	0	0	13	0	15	0
11	57	2	22	319	82	0	7	59	4	52	0
12	3	5	34	426	1044	23	11	100	20	89	28
13	48	0	9	364	57	7	16	72	5	90	21
14	1	0	16	271	181	0	0	8	2	29	0
15	6	4	7	174	184	0	4	7	0	37	4
16	23	4	24	430	284	20	8	40	10	84	33
17	5	5	7	249	92	19	8	59	18	87	0
18	18	13	18	577	193	26	46	110	7	159	46
19	1	1	8	234	374	15	9	31	11	177	29
20	8	20	50	190	134	13	19	42	22	226	91
21	5	141	16	180	207	11	4	5	23	133	52
22	83	15	33	534	414	5	10	39	5	116	32
23	17	2	14	344	206	7	9	14	1	86	21
24	51	0	3	190	34	0	1	3	0	32	0
25	22	17	38	308	129	0	0	0	0	40	10
26	16	4	6	451	291	19	16	34	6	155	35
27	8	1	12	201	716	15	15	86	15	75	19
28	20	4	7	207	170	7	9	37	28	51	0
29	2	0	3	35	45	0	3	3	0	18	0
30	9	8	11	158	280	4	3	46	4	40	0
31	0	7	15	374	119	0	2	15	0	61	0
32	4	6	0	168	197	0	0	14	10	32	0
33	0	2	2	160	180	15	3	13	12	54	14
34	49	7	22	764	123	0	7	34	1	117	34
35	11	6	23	236	209	6	6	15	15	73	19
36	54	17	21	595	308	13	14	73	8	101	32
37	6	27	20	521	251	14	24	39	4	92	11
38	37	18	18	261	103	9	1	17	0	72	0
39	21	12	52	802	282	22	48	129	10	201	77
40	3	1	11	414	195	12	5	41	22	85	19
41	13	6	13	206	60	5	19	37	4	68	0
42	2	1	5	187	76	0	0	11	2	48	0
43	6	31	46	538	101	14	22	91	2	151	74
44	0	4	2	250	231	3	4	30	21	45	0
45	101	12	73	892	885	93	62	174	72	196	77
46	1	2	6	231	169	9	13	46	20	58	22

PPG	X ₂₅	X ₂₆	X ₂₇	X ₂₈	X ₂₉	X ₃₀	X ₃₁	X ₃₂	X ₃₃	X ₃₄	X ₃₅
47	21	29	55	502	637	15	10	75	23	93	43
48	36	8	41	670	373	27	26	66	13	175	65
49	16	20	39	286	709	31	27	36	10	118	47
50	60	7	27	743	163	108	68	122	17	391	116
51	1	0	5	364	294	19	18	83	47	110	30
52	5	12	26	511	385	0	0	0	0	22	0
53	43	5	32	656	677	5	15	26	3	72	5
54	6	16	29	454	978	11	3	29	7	68	0
55	90	47	11	863	255	15	24	68	5	267	110
56	37	10	11	426	338	5	4	43	9	93	5
57	16	6	8	325	405	4	11	113	24	76	24
58	180	11	28	1408	807	17	24	85	53	161	60
59	90	40	51	1797	1213	32	49	127	30	291	110
60	71	26	19	1267	730	4	4	34	5	106	11
61	5	1	9	395	78	8	4	61	4	101	0
62	60	28	83	871	976	8	6	97	24	82	35
63	86	12	19	573	340	0	0	0	0	20	0
64	63	24	49	785	864	8	13	117	8	82	0
65	21	9	7	320	159	1	0	1	0	23	0
66	4	12	32	599	167	8	9	31	6	109	62
67	11	13	21	391	132	22	24	55	15	79	28
68	5	26	10	346	93	0	1	15	0	61	0
69	5	2	3	157	119	6	9	24	20	55	14
70	14	5	72	587	396	2	6	34	3	75	17
71	3	5	0	43	19	1	0	0	0	16	0
72	7	3	18	565	132	6	13	68	5	87	6
73	10	25	77	385	155	2	7	25	16	86	5
74	66	4	14	199	60	0	0	1	0	12	0
75	18	0	6	240	26	0	0	1	0	9	0
76	2	0	11	270	132	0	0	0	0	10	0
77	28	11	30	685	379	69	33	145	73	315	67
78	55	12	86	1014	444	12	20	179	17	138	73
79	35	33	9	630	612	3	3	3	0	22	0
80	31	20	73	682	338	44	37	325	19	174	56
81	19	25	117	687	425	37	16	97	19	130	55
82	28	21	14	650	240	6	15	123	16	121	0
83	18	13	33	669	99	2	4	31	4	48	0
84	1	2	2	139	21	0	2	1	0	15	0
85	26	8	42	223	738	8	12	43	125	90	0
86	29	10	14	495	518	2	6	35	11	45	0
87	4	2	10	291	370	4	6	23	8	45	0
88	5	1	9	289	19	1	1	5	0	40	0
89	5	2	3	161	53	0	0	10	4	50	0
90	15	1	19	282	103	3	2	14	4	72	11
91	26	0	10	333	45	2	2	7	0	93	26
92	8	1	13	355	39	0	0	21	0	44	0

FONTE: CAPES (2017).