UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA
FACULDADE DE FILOSOFIA E CIÊNCIAS
Campus de Marília

GIOVANNA DIAS CARDOSO

Um modelo baseado em regras *fuzzy* para análise do desempenho acadêmico dos estudantes da UNESP que ingressaram pelo programa de inclusão

Marília
2026

UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA
FACULDADE DE FILOSOFIA E CIÊNCIAS
Campus de Marília

GIOVANNA DIAS CARDOSO

Um modelo baseado em regras *fuzzy* para análise do desempenho acadêmico dos estudantes da UNESP que ingressaram pelo programa de inclusão

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Filosofia da Faculdade de Filosofia e Ciências da Universidade Estadual Paulista "Júlio de Mesquita Filho", campus de Marília, como um dos requisitos para obtenção do mestrado em filosofia.

Linha de Pesquisa: Filosofia da Informação, da Cognição e da Consciência.

Orientador: Prof. Dr. Luiz Henrique da Cruz Silvestrini
Coorientadora: Profa. Dra. Ana Cláudia de Jesus Golzio

Marília
2026

C268m

Cardoso, Giovanna Dias
Um modelo baseado em regras fuzzy para análise do desempenho acadêmico dos estudantes da UNESP que ingressaram pelo programa de inclusão / Giovanna Dias Cardoso. -- Marília, 2026
147 p.

Dissertação (mestrado) - Universidade Estadual Paulista (UNESP), Faculdade de Filosofia e Ciências, Marília
Orientador: Luiz Henrique da Cruz Silvestrini
Coorientadora: Ana Cláudia de Jesus Golzio

1. Lógica fuzzy. 2. Sistemas fuzzy. 3. Programa de Inclusão. 4. Política de cotas. 5. Modelagem. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Dados fornecidos pelo autor(a).

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Marília

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO:Um modelo baseado em regras fuzzy para análise do desempenho acadêmico dos estudantes da UNESP que ingressaram pelo programa de inclusão

AUTORA: GIOVANNA DIAS CARDOSO
ORIENTADOR: LUIZ HENRIQUE DA CRUZ SILVESTRINICOORIENTADORA: ANA CLAUDIA DE JESUS GOLZIO

Aprovada como parte das exigências para obtenção do Título de Mestra em Filosofia, pela Comissão Examinadora:

Prof(a). Dr(a). LUIZ HENRIQUE DA CRUZ SILVESTRINI (Participaçao Presencial)
Departamento de Matemática / UNESP / Câmpus de Bauru - FC

Prof. Dr. TIAGO AUGUSTO DOS SANTOS BOZA (Participaçao Presencial)
Instituto Federal de Educação, Ciência e Tecnologia de São Paulo

Profa. Dra. MARIANA MATULOVIC DA SILVA RODRIGUEIRO (Participaçao Virtual)
Departamento de Engenharia de Biossistemas / UNESP / Câmpus de Tupã - FCE

Marília, 12 de janeiro de 2026.

Gustavo Henrique Nicolau Alves Gabardo
Supervisor Técnico de Seção
Seção Técnica de Pós-Graduação

Dedico esta vitória aos meus amados pais, Silvana e Vanderlei. Por todo o amor, apoio incondicional e por moldarem, com paciência e sabedoria, a pessoa que me tornei. Vocês sempre foram meu porto seguro e meu maior impulso para ir além.

Agradecimentos

Este trabalho é fruto de um esforço coletivo e do apoio incondicional de pessoas incríveis. A cada uma delas, meu mais sincero obrigada.

Aos meus pais, Silvana e Vanderlei, meu agradecimento mais profundo. Por todo o suporte, por me darem as condições necessárias para me dedicar aos estudos e por acreditarem em mim em cada passo desta jornada. O apoio de vocês foi o alicerce que me manteve firme.

Ao meu orientador, Professor Luiz Henrique, minha eterna gratidão. Sua orientação precisa e seu incentivo constante foram fundamentais para a conclusão desta pesquisa. Obrigada por me guiar com sabedoria, por me ajudar a superar os desafios e por ser um grande mentor. Também agradeço imensamente à minha coorientadora, Ana Claudia, por sua ajuda incansável, seu valioso apoio e por todo o conhecimento que me foi transmitido durante este percurso.

Ao meu namorado, Paulo Henrique, que esteve ao meu lado desde o processo seletivo, com apoio e paciência. Obrigada por todo o suporte, por fazer o possível para que eu pudesse me dedicar e por ser meu parceiro em todas as horas.

Agradeço ao Professor Hércules de Araújo Feitosa por sua enorme sabedoria e por me proporcionar aulas que ampliaram minha visão e contribuíram significativamente para minha formação.

Aos amigos que o mestrado me deu, Rômulo e Beatriz, pela parceria nos eventos, pelo companheirismo e por tornarem o caminho mais leve e divertido. Espero levar a amizade de vocês para a vida toda.

À minha melhor amiga, Rafaella, por me apoiar em cada decisão e por torcer genuinamente por meu sucesso. Aos meus queridos amigos e companheiros de graduação, Matheus e Júlia, que sempre acreditaram em mim e cujas amizades levo para a vida.

À minha família, meu alicerce. Aos meus avós, Antônia, Moacir e Mariede, e à minha prima Bruna e minha madrinha Andréia, pelo carinho e por estarem sempre presentes. Deixo uma homenagem especial ao meu avô, Deolindo (in memoriam), que sempre comemorou minhas conquistas e que, tenho certeza, estaria orgulhoso deste momento.

“Há e haverá muitas tarefas que os homens podem cumprir com facilidade, que vão além da capacidade de qualquer computador, qualquer máquina e qualquer sistema lógico que podemos conceber nos dias de hoje.”

- Lotfi A. Zadeh

RESUMO

Um conjunto *fuzzy* pode ser entendido como uma função de certo domínio V, denominado universo de discurso, no intervalo real [0,1], que possibilita uma passagem gradual da pertinência para a não pertinência. A lógica *fuzzy* desempenha um papel crucial ao abordar a representação do pensamento e raciocínio, pois oferece a capacidade de lidar com incertezas, raciocínio aproximado e termos vagos e ambíguos, que são formas comuns pelas quais as pessoas expressam suas ideias. Assim, a lógica *fuzzy* fundamenta sistemas computacionais que subsidiam algoritmos que "raciocinam" de maneira mais semelhante ao pensamento humano, considerando nuances inerentes à incerteza e processos realísticos. Essa abordagem contribui para tornar os sistemas mais adaptados e receptivos às complexidades do mundo real. A partir dos conjuntos *fuzzy* cresce a possibilidade de interpretar fenômenos imprecisos e não quantitativos e a necessidade de buscar mecanismos de inferências através desses dados, com isso, surgiram os sistemas de inferência baseado em regras *fuzzy* (SBRF). Nesta pesquisa, foi desenvolvido e aplicado um SBRF com o objetivo de analisar possíveis diferenças no desempenho acadêmico de alunos formados no curso de Licenciatura em Matemática da UNESP de Bauru, considerando distintos grupos de ingresso, incluindo estudantes oriundos do programa de inclusão e aqueles admitidos pelo sistema universal. O modelo incorporou múltiplas variáveis acadêmicas e produziu classificações linguísticas de desempenho geral. Os resultados obtidos indicaram predominância da classificação "bom" em todos os grupos analisados, bem como diferenças pontuais nas distribuições das categorias "regular" e "ruim". A validação revelou alto grau de concordância com os resultados do modelo, reforçando a adequação da abordagem adotada.

Palavras-chave: Programa de Inclusão; Sistemas Especialistas; Sistema Baseado em Regras *Fuzzy*.

ABSTRACT

A *fuzzy* set is understood as a function defined over a domain V, called the universe of discourse, whose values belong to the real interval [0,1], allowing a gradual transition between states of membership and non-membership. *Fuzzy* logic plays a fundamental role in representing human thought by dealing with uncertainties, approximations, and vague concepts, which are common characteristics of everyday language. This approach makes computational systems more flexible and capable of handling the complexity of the real world, promoting a kind of reasoning that is closer to the way humans think. In this context, *fuzzy* sets open possibilities for interpreting imprecise and qualitative phenomena, encouraging the development of rule-based inference mechanisms. Thus, *Fuzzy* Rule-Based Systems (FRBS) emerge as tools that enable the modeling of complex situations based on imprecise data. In this research, an FRBS was developed and applied with the aim of analyzing possible differences in the academic performance of students graduating from the Mathematics Degree course at UNESP in Bauru, considering different admission groups, including students from the inclusion program and those admitted through the universal system. The model incorporated multiple academic variables and produced linguistic classifications of overall performance. The results obtained indicated a predominance of the "good" classification in all groups analyzed, as well as specific differences in the distributions of the "fair" and "poor" categories. Validation revealed a high degree of agreement with the model's results, reinforcing the adequacy of the approach adopted.

keywords: Inclusion Program; Knowledge-Based Systems; *Fuzzy* Rule-Based Systems

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

COPE: Coordenadoria de Permanência Estudantil
CR: Coeficiente de Rendimento
FAPESP: Fundação de Amparo à Pesquisa do Estado de São Paulo
IA: Inteligência Artificial
LGPD: Lei Geral de Proteção de Dados
PGD: Plano de Gestão de Dados
PIBIC: Programa Institucional de Bolsas de Iniciação Científica
PIBID: Programa Institucional de Bolsas de Iniciação à Docência
PIMESP: Programa de Inclusão com Mérito no Ensino Superior Público Paulista
PNAES: Programa Nacional de Assistência Estudantil
PPI: Pretos, Pardos e Indígenas
PPISES: Programa Paulista de Inclusão Social no Ensino Superior
PROGRAD: Pró-Reitoria de Graduação
RAs: Registros acadêmicos
SBRF: Sistemas Baseados em Regras *Fuzzy*
SISGRAD: Sistema de Graduação
SRVEBP: Sistema de Reserva de Vagas da Educação Básica Pública
UERJ: Universidade do Estado do Rio de Janeiro
UNICAMP: Universidade Estadual de Campinas
UNESP: Universidade Estadual Paulista "Júlio de Mesquita Filho"
USP: Universidade de São Paulo

SUMÁRIO

1. INTRODUÇÃO

A discussão sobre certeza e verdade, assim como sobre dúvida e incerteza, é central na epistemologia do conhecimento. Na filosofia grega, Platão associa a certeza ao "mundo das ideias", em oposição ao "mundo sensível", onde os resultados seriam apenas aproximações das verdades ideais. Já os sofistas adotavam uma postura cética em relação ao conhecimento absoluto e objetivo, buscando transformar modos de argumentação em verdades relativas, enfatizando a subjetividade na construção do saber. (D'Ottaviano e Feitosa, 2003)

Na busca pela essência do conhecimento, a matemática tem se mostrado uma ferramenta fundamental para esclarecer e aprofundar teorizações. Entretanto, a procura pela verdade ou certeza frequentemente leva à descoberta de evidências contraditórias, o que impulsiona o estudo de diferentes formas de incerteza (Corcoll-Spina, 2010).

Segundo D'Ottaviano e Feitosa (2003), no final do século XIX, em busca de soluções não aristotélicas para questões lógicas em aberto, alguns trabalhos foram os precursores das chamadas lógicas não-clássicas e no início do século XX, matemáticos e filósofos criaram novos sistemas lógicos, que se diferem daqueles que representam a lógica de Aristóteles.

Tempo depois, Jan Łukasiewicz, lógico polonês, considerou sistemas lógicos com mais de dois valores de verdade, a saber, as lógicas polivalentes, multivalentes ou multivaloradas. Em uma lógica multivalorada, as proposições podem assumir três ou mais valores de verdade. Além daqueles que conhecemos da lógica clássica, falso e verdadeiro, a princípio, foi adicionado o valor "possível" como um terceiro valor. Łukasiewicz introduz seus sistemas de lógicas polivalentes com o intuito de investigar proposições modais e as noções de possibilidade e necessidade relacionadas com tais proposições. Essas proposições, são do tipo que retratavam as expressões: "é possível que p", "não é possível que p", "é possível que não-p" ou "é contingente que p" e "não é possível que não-p" ou "é necessário que p". (Łukasiewicz e Tarski, 1930).

D'Ottaviano e Feitosa (2003) observam que na lógica apresentada por Łukasiewicz, uma afirmação poderia ser verdadeira e falsa ao mesmo tempo, o que seria possível somente na condição de não assumirmos apenas sentenças verdadeiras e falsas, mas com algum grau de verdade distinto, o que produziria diversos níveis de possibilidades e não apenas os dois valores clássicos. Ao assumir a existência de sentenças às quais podíamos atribuir um terceiro valor de verdade além do "verdadeiro" e "falso", Łukasiewicz não rejeitou os princípios da não-contradição ou do terceiro excluído. Dessa forma, as concepções de lógicas multivaloradas por ele apresentadas, foram o embrião das designadas lógicas *fuzzy*.

De acordo com D'Ottaviano e Feitosa (2003), a partir dos estudos apresentados por Łukasiewicz, o professor Lotfi Askar Zadeh, em meados da década de 1960, notou que os métodos tradicionais da matemática da época não se mostravam capazes de formalizar algumas situações referentes a problemas que compreendessem posições ambíguas ou não completamente claras. Para contornar a incapacidade de representar esses problemas por tomadas de decisões binárias, Zadeh propõe a aceitação de mais que dois valores possíveis de verdade. O professor, dessa forma, apresentou uma teoria de conjuntos que recebeu o nome de teoria de conjuntos *fuzzy*, caracterizada pela ausência da aplicação tradicional da bivalência. Mais tarde, por volta da metade da década seguinte, ele propôs uma lógica não clássica, fundamentada em sua teoria de conjuntos, a lógica *fuzzy*.

Para Takács (2004), o cérebro humano possui determinadas características especiais que possibilitam aprender a raciocinar em ambientes vagos ou imprecisos. Daí, entendemos a importância da lógica *fuzzy*. Esta lógica, se assemelha, em questão de estrutura, com a lógica aristotélica, porém, se diferencia dela pela obtenção de uma conclusão vaga e imprecisa, que foi deduzida de premissas também imprecisas, representadas por conjuntos *fuzzy*.

A lógica *fuzzy*, foi construída inicialmente através de conceitos já estabelecidos na lógica clássica, mas de modo a ampliá-la e permitir raciocínios aproximados. Segundo Tákacs (2004), o raciocínio é fundamental na lógica *fuzzy*, pois é um sistema baseado em regras. A representação desse conhecimento é dado pela regra "Se…, então…", uma implicação lógica.

Ao modelar situações da realidade, muitas variáveis linguísticas apresentam alto grau de subjetividade, sem suporte em distribuições estatísticas ou padrões aleatórios. Nesse contexto, tais variáveis são avaliadas por meio de graduações ou níveis intermediários de verdade, permitindo uma abordagem mais flexível e adaptada à complexidade dos fenômenos (Tákacs, 2004).

Nesse cenário de incerteza, a lógica *fuzzy* se destaca ao utilizar uma linguagem baseada em conjuntos, na qual cada elemento possui um grau de pertinência ao conjunto. Essa abordagem tem avançado significativamente em suas aplicações práticas, graças à sua capacidade de graduar soluções com base em medidas de credibilidade. Diferentemente da lógica bivalente de Aristóteles, em que uma solução é considerada apenas verdadeira ou falsa, a lógica *fuzzy* permite que as soluções sejam avaliadas com um grau específico de confiança, proporcionando maior flexibilidade e adequação às necessidades do usuário (Tákacs, 2004).

Para Shaw e Simões (1999), a característica principal da lógica *fuzzy* é oferecer aos pesquisadores uma nova maneira de trabalhar com informações imprecisas, e ainda, de forma diferente do que se faz na teoria das probabilidades, por exemplo. A lógica *fuzzy* nos permite

traduzir em valores numéricos as expressões verbais, vagas, imprecisas e qualitativas, encontradas na comunicação humana, o que possibilita a conversão da experiência humana em uma linguagem decodificável por computador. Essa tradução é realizada por meio de um sistema formal que utiliza funções de pertinência, as quais mapeiam elementos do domínio em graus de verdade dentro de um intervalo contínuo. Termos linguísticos são modelados como variáveis linguísticas cujos valores são conjuntos fuzzy, permitindo a representação computacional de predicados vagos. O processamento dessas informações se dá através de operações algébricas sobre esses conjuntos e mecanismos de inferência baseados em regras, possibilitando que máquinas interpretem e manipulem conhecimento expresso em linguagem natural.

O estudo de problemas e situações reais utilizando a linguagem matemática para compreendê-los, simplificá-los e solucioná-los, em busca de uma possível revisão e modificação do objeto em estudo, é parte do processo da modelagem matemática. Entretanto, muitos fenômenos reais não podem ser adequadamente representados por modelos que exigem precisão absoluta e valores discretos. Variáveis como "conforto térmico", "qualidade do produto" ou "nível de risco" são subjetivas e graduais, não se ajustando à lógica binária tradicional. Devido a possibilidade de manipulação de informações incertas e de seu respectivo armazenamento em computadores, o tratamento *fuzzy* de variáveis linguísticas subjetivas, referentes ao sujeito, ganhou espaço na modelagem matemática. Assim, toda tecnologia baseada no "enfoque *fuzzy*" ganha aplicabilidade prática, permitindo que a experiência e a intuição de controladores humanos sejam incorporadas em sistemas de controle computadorizado (SHAW; SIMÕES, 1999).

Pesquisas têm utilizado a teoria *fuzzy* como ferramenta para análise de desempenho acadêmico de estudantes, tais como citadas abaixo, em razão de sua capacidade de lidar com a subjetividade e imprecisão inerentes aos processos avaliativos educacionais. Essa abordagem permite incorporar variáveis linguísticas e considerar nuances que os sistemas tradicionais de avaliação frequentemente ignoram.

Silva et al. (2023) propõem um modelo matemático baseado na teoria dos conjuntos *fuzzy* para auxiliar docentes na avaliação das aprendizagens discentes. Os autores argumentam que avaliar o processo de ensino-aprendizagem e mensurar o desempenho discente constitui uma tarefa complexa e subjetiva, na qual os conjuntos *fuzzy* se apresentam como uma excelente ferramenta.

Barrantes (2011) desenvolve um sistema de avaliação para o ensino básico que emprega a teoria *fuzzy* para trabalhar com variáveis linguísticas (bom, ruim, normal) e

quantificá-las numericamente. A autora critica o sistema tradicional de média aritmética, argumentando que este mascara deficiências específicas de aprendizagem ao permitir que notas excelentes em alguns conteúdos compensem notas insuficientes em outros, incentivando estudantes a focarem apenas em atingir a média mínima. O modelo proposto agrupa o desempenho considerando diversas categorias de avaliações (diagnósticas, processuais e formativas).

Wilges et al. (2010) implementaram um sistema de agentes baseado em lógica *fuzzy* para identificar perfis de aprendizagem em Ambientes Virtuais de Aprendizagem (AVA), distinguindo o desempenho prático do teórico. Ao analisar dados de 57 estudantes, os autores identificaram uma forte correlação entre as atividades no AVA e as avaliações presenciais.

Barros e Lanzillotti (2023) propõem um método baseado em conjuntos *fuzzy* intuicionistas para analisar o desempenho de estudantes de matemática do ensino fundamental. Os autores criticam o sistema de avaliação da rede pública de ensino, argumentando que bônus por atividades complementares mascaram lacunas de conhecimento e geram resultados irreais, evidenciando a discrepância entre planejamento pedagógico e aprendizado efetivo, como demonstrado por estudantes que chegam ao ensino médio com dificuldades em operações básicas.

Neste contexto, a presente pesquisa busca, através de um sistema *fuzzy*, analisar o desempenho de estudantes formados no curso de Licenciatura em Matemática da UNESP Bauru. A escolha deste método justifica-se pela sua capacidade de lidar com a complexidade e subjetividade inerentes à avaliação do desempenho acadêmico, permitindo uma análise mais aprofundada e realista que vai além das limitações dos métodos convencionais. A teoria dos conjuntos *fuzzy* possibilita incorporar múltiplas dimensões do desempenho estudantil e identificar padrões que métodos convencionais poderiam mascarar. Dessa forma, espera-se obter uma compreensão mais abrangente do perfil de formação dos egressos, contribuindo para reflexões sobre o processo formativo e possíveis aprimoramentos curriculares do curso.

1.1. Objetivos

Objetivo geral:

Esta pesquisa possui como objetivo geral a análise do desempenho acadêmico dos estudantes formados no curso de Licenciatura em Matemática da UNESP de Bauru, entre os anos de 2016 a 2024, investigando as diferenças de desempenho entre aqueles que

ingressaram através do sistema universal e os que ingressaram por meio dos sistemas de cotas. Tal análise será realizada a partir da construção de um modelo SBRF utilizando dados fornecidos pela universidade.

Objetivos específicos:

- Investigar a lógica *fuzzy* e os conjuntos *fuzzy*, destacando as diferenças entre conjuntos *fuzzy* e os conjuntos usuais; e abordar os Sistemas Baseados em Regras *Fuzzy* (SBRF);
- Coletar e analisar estatisticamente dados dos estudantes formados no curso de Licenciatura em Matemática da UNESP de Bauru, comparando o desempenho acadêmico daqueles que ingressaram através do sistema universal com os que ingressaram através dos sistemas de cotas, considerando indicadores como: coeficiente de rendimento (CR), índices de desempenho e frequência;
- Realizar uma modelagem via Sistemas Baseados em Regras *Fuzzy* (SBRF) a fim de verificar se existem diferenças significativas entre o desempenho acadêmico dos estudantes que ingressaram no curso de matemática da UNESP de Bauru pelo sistema universal e os que ingressaram pelo Sistema de Reserva de Vagas da Educação Básica Pública (SRVEBP).

1.2 Materiais e métodos

Este trabalho teórico consistiu na realização de um estudo rigoroso e detalhado dos textos indicados na bibliografia, com o objetivo de fundamentar as atividades de pesquisa desta dissertação. O estudo forneceu a base conceitual necessária para o desenvolvimento de um Sistema Baseado em Regras *Fuzzy*, no qual foi adotado o método Mamdani.

A implementação do sistema, bem como a análise estatística, foi realizada com o uso do Google Colab com a linguagem de programação Python. O processo de desenvolvimento envolveu inicialmente a etapa de modelagem conceitual, na qual foram identificadas as variáveis linguísticas de entrada e a variável de saída. Em seguida, procedeu-se ao ajuste de parâmetros dos conjuntos fuzzy, determinando os limites e formas das funções de pertinência que melhor representassem os conceitos linguísticos desejados. A definição das funções de pertinência considerou tanto as características estatísticas dos dados disponíveis quanto a opinião de especialistas sobre a interpretação adequada dos termos linguísticos no contexto do problema. Por fim, foram estabelecidas as regras de inferência fuzzy que governam o

comportamento do sistema. Esse processo iterativo de refinamento assegurou a precisão e a adequação do sistema às demandas da pesquisa.

Esta dissertação foi organizada em 5 capítulos. Introduzimos a pesquisa delineando os objetivos e a metodologia utilizada. O Capítulo 2 introduz o leitor aos conjuntos *fuzzy*, explorando conceitos fundamentais, bem como as operações com conjuntos *fuzzy*, a álgebra dos conjuntos *fuzzy* e as relações *fuzzy*. Além disso, apresentamos a lógica *fuzzy*, discutindo seus elementos construtivos, características e regras de dedução. Por fim, apresenta-se os sistemas baseados em regras *fuzzy* (SBRF), destacando as etapas de modelagem. O Capítulo 3 promove uma análise a respeito do programa de inclusão da Universidade Estadual Paulista (UNESP). Este capítulo tem como objetivo compreender o funcionamento das ações afirmativas voltadas para a permanência estudantil, proporcionando reflexões sobre as políticas de inclusão.

O quarto capítulo descreve detalhadamente o processo de coleta dos dados utilizados na pesquisa, especificando as fontes consultadas, os critérios de seleção e os procedimentos adotados para garantir a confiabilidade das informações obtidas. Além disso, neste capítulo, foi realizada uma análise estatística desses dados, incluindo medidas descritivas, identificação de padrões e visualizações que auxiliam na compreensão das características das variáveis estudadas. O quinto capítulo apresenta o desenvolvimento e a implementação do Sistema Baseado em Regras Fuzzy (SBRF), abordando desde a modelagem conceitual até os testes de validação do sistema proposto. Por fim, encerramos a dissertação com as conclusões e considerações finais, nas quais são discutidas as implicações dos resultados obtidos.

2. A TEORIA *FUZZY*

Na teoria de conjuntos usuais, um elemento pertence ou não a um conjunto, não existindo um meio termo. Mas quando trabalhamos com o mundo real, este conceito nem sempre pode ser aplicado. Por exemplo, uma temperatura média de 25 graus para uma certa cidade no Brasil pertence ao conjunto baixa ou alta? A resposta é relativa e depende do referencial. Se compararmos, por exemplo, com uma cidade do sul do país, essa temperatura pode ser considerada alta, mas, se compararmos com uma cidade do norte do país, pode ser considerada baixa. Por esse motivo, é difícil limitarmos a temperatura apenas em alta ou baixa.

Dessa forma, Loft A. Zadeh, na década de 60, propôs a teoria dos conjuntos *fuzzy*, também conhecida como conjuntos nebulosos, ou ainda, conjuntos difusos, uma teoria alternativa à teoria dos conjuntos clássica e menos rígida do que o habitual. Nesta proposta, a mudança de pertinência para não pertinência de um determinado elemento em um conjunto fuzzy, acontece de maneira gradual.

2.1 Os conjuntos fuzzy

Um conjunto *fuzzy* A_f é dado através de uma função f_A: $V \rightarrow [0,1]$, em que o conjunto V é o conjunto universo ou domínio do conjunto *fuzzy*, [0,1] é um intervalo de números reais e f_A é a função de pertinência de A_f. Denota-se:

$$f_A: V \rightarrow [0,1]$$
$$x \mapsto f_A(x).$$

Dado que um conjunto *fuzzy* é determinado por uma função, e temos da teoria dos conjuntos usuais que uma função pode ser representada por um conjunto de pares ordenados, denotam-se os conjuntos *fuzzy* como a seguir:

Um conjunto *fuzzy* A_f é denotado por um conjunto de pares ordenados, em que o primeiro elemento pertence a V e o segundo elemento indica o seu grau de pertinência em A_f.

$$A_f = \{(a, \mu) / f_A(a) = \mu \text{ e } \mu \in [0,1]\}.$$

A teoria apresentada a seguir acerca dos conjuntos *fuzzy* foi desenvolvida com base na abordagem apresentada por Feitosa e Paulovich (2005).

2.1.1 Operações com conjuntos fuzzy

Definição 2.1.1: Dois conjuntos *fuzzy* A_f e B_f são *iguais* se: $(\forall x \in V)\ f_A(x) = f_B(x)$, denotado por $A_f =_f B_f$.

Definição 2.1.2: Sejam A_f e B_f dois conjuntos *fuzzy* em V, dizemos que B_f é um *subconjunto fuzzy* de A_f se $(\forall x \in V)\ f_B(x) \leq f_A(x)$, denotado por $B_f \subseteq_f A_f$.

Definição 2.1.2: O conjunto *fuzzy vazio* é dado pela função constante zero, denotado por $0_f = \emptyset_f =_{def} f_\emptyset(x) = 0\ ,\ (\forall x \in V)$.

Definição 2.1.4: O conjunto *fuzzy universo* é dado pela função constante um, denotado por $1_f = V =_{def} f_V(x) = 1\ ,\ (\forall x \in V)$.

Definição 2.1.5: A *união* de dois conjuntos *fuzzy* A_f e B_f é um conjunto *fuzzy* $A_f \cup B_f$, tal que, para cada $x \in V$, o seu grau de pertinência no conjunto união é o *valor máximo* entre $f_A(x)$ e $f_B(x)$. Assim,

$$A_f \cup B_f = \{(x, \max\{f_A(x), f_B(x)\}) / x \in V\}.$$

Proposição 2.1.1: A união de dois conjuntos *fuzzy* A_f e B_f é dada pelo menor conjunto *fuzzy* que contém A_f e B_f.

Considere C_f um conjunto *fuzzy* que contém A_f e B_f. Assim,

$$(\forall x \in V) f_C(x) \geq f_A(x) \text{ e } f_C(x) \geq f_B(x).$$

Desse modo,

$$f_C(x) \geq max\{f_A(x), f_B(x)\}, \text{ logo, } A_f \cup B_f \subseteq C_f.$$

Definição 2.1.6: A *intersecção* de dois conjuntos *fuzzy* A_f e B_f é um conjunto *fuzzy* $A_f \cap B_f$, tal que, para cada $x \in V$, o seu grau de pertinência no conjunto união é o *valor mínimo* entre $f_A(x)$ e $f_B(x)$. Logo,

$$A_f \cap B_f = \{(x, \min\{f_A(x), f_B(x)\}) / x \in V\}.$$

Proposição 2.1.2: A intersecção de dois conjuntos *fuzzy* A_f e B_f é dada pelo maior conjunto *fuzzy* que está contido em A_f e B_f.

Considere C_f um conjunto *fuzzy* contido em A_f e em B_f. Assim,

$$(\forall x \in V) f_C(x) \leq f_A(x) \text{ e } f_C(x) \leq f_B(x).$$

Desse modo,

$$(\forall x \in V)\, f_C(x) \leq min\{f_A(x),\, f_B(x)\}, \text{ e então, } C_f \subseteq A_f \cap B_f$$

Pelas Proposições 2.1.1 e 2.1.2, entendemos que, para dois conjuntos *fuzzy* A_f e B_f, as operações de união e intersecção assumem, respectivamente, os valores *supremos* e *ínfimos* das funções de pertinência de A_f e B_f.

Exemplificando:

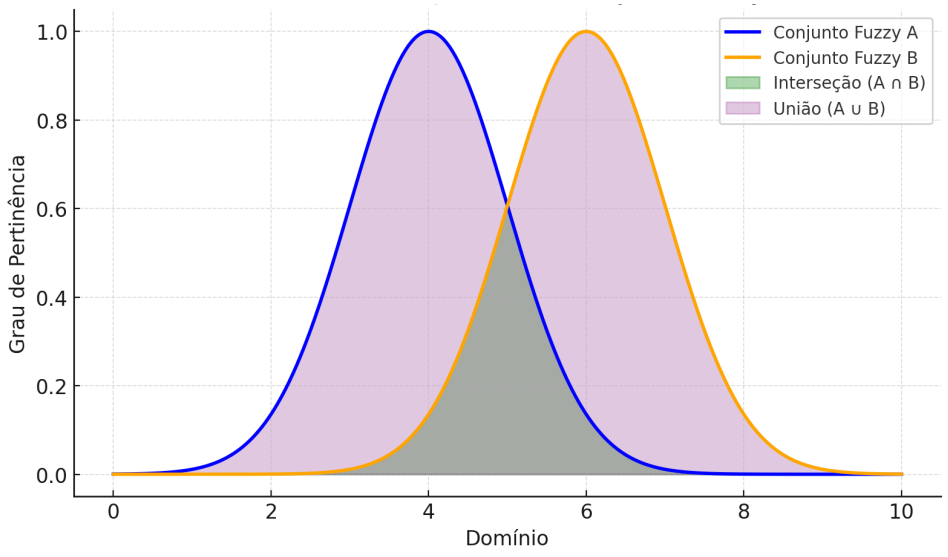
Considerando V = $\{x_1, x_2, x_3, x_4, x_5, x_6\}$, A = $\{(x_1, 0.4); (x_3, 1); (x_4, 0.7); (x_5, 0.5)\}$ e B = $\{(x_1, 0.2); (x_2, 0.9); (x_3, 0.8); (x_4, 0.7); (x_5, 0.5); (x_6, 1)\}$, então:

$$A \cup B = \{(x_1, 0.4); (x_2, 0.9); (x_3, 1); (x_4, 0.7); (x_5, 0.5); (x_6, 1)\} \text{ e}$$
$$A \cap B = \{(x_1, 0.2); (x_3, 0.8); (x_4, 0.7); (x_5, 0.5)\}.$$

Definição 2.1.7: Dado um conjunto *fuzzy* A no domínio V, o seu *complemento fuzzy*, denotado por A', é determinado por $f_{A'}(x) = 1 - f_A(x)$, $(\forall x \in V)$.

Abaixo, ilustramos graficamente a união e a intersecção de dois conjuntos *fuzzy*

Figura 1: União e intersecção dos conjuntos fuzzy



Fonte: Autora

Definição 2.1.8: Dados dois conjuntos *fuzzy* A e B no domínio V, a *diferença fuzzy* entre A e B, denotada por A – B, é definida por:

$$f_{A-B}(x) = \begin{cases} 0, \text{ se } f_A(x) \leq f_B(x) \\ f_A(x) - f_B(x), \text{ se } f_A(x) > f_B(x) \end{cases}$$

Exemplificando:

Considerando V = $\{x_1, x_2, x_3\}$, A = $\{(x_1, 0.4); (x_2, 0.3); (x_3, 0.8)\}$ e B = $\{(x_1, 0.9); (x_2, 0.5); (x_3, 0.3)\}$, então:

A' = $\{(x_1, 0.6); (x_2, 0.7); (x_3, 0.2)\}$
B' = $\{(x_1, 0.1); (x_2, 0.5); (x_3, 0.7)\}$
A – B = $\{(x_1, 0); (x_2, 0); (x_3, 0.5)\} = \{(x_3, 0.5)\}$.

A teoria clássica dos conjuntos pode ser vista como um caso particular da teoria dos conjuntos *fuzzy*. Isso ocorre quando a função de pertinência equivale à função característica, cuja imagem é restrita ao conjunto {0, 1}. Nesse contexto, um elemento x possui grau de pertinência igual a 1 se x pertence ao conjunto A, e grau de pertinência igual a 0 se x não pertence a A.

2.1.2 A álgebra dos conjuntos *fuzzy*

Nesta seção, iniciamos indicando algumas notações que simplificarão o trabalho.

$f_A(u),\ (\forall x \in V)$ será denotado por f_A

Máx $\{f_A, f_B\}$ será denotado por $f_A \vee f_B$

Min $\{f_A, f_B\}$ será denotado por $f_A \wedge f_B$

Considere L = {A / A é conjunto *fuzzy* com universo V} e as definições dadas como axiomas. Daremos uma atenção particular à estrutura algébrica determinada por $(L, \subseteq_f, \cup_f, \cap_f, ')$. Na álgebra dos conjuntos *fuzzy*, são válidas as propriedades:

Propriedade reflexiva:

$A \subseteq_f A$, se, e somente se, $(\forall x \in V) f_A(x) \leq f_A(x)$.

Propriedade anti-simétrica:

$A \subseteq_f B$ e $B \subseteq_f A \Leftrightarrow A =_f B$, pois

$$f_A = f_B \Leftrightarrow f_A \leq f_B \, e \, f_B \leq f_A .$$

Propriedade transitiva:

$A \subseteq_f B$ e $B \subseteq_f C \Rightarrow A \subseteq_f C$, pois

$$\text{se } f_A \leq f_B \, e \, f_B \leq f_C, \text{ então } f_A \leq f_C.$$

Um conjunto não vazio que apresenta uma relação onde são válidas as propriedades reflexiva, anti-simétrica e transitiva é uma *ordem parcial*. Ademais, se para todo x, y em L existir o supremo e o ínfimo do conjunto {x, y} de L, a estrutura é chamada *reticulado*. Assim, $(L, \subseteq_f)$ é parcialmente ordenado e como vimos anteriormente, a união de dois conjuntos *fuzzy* A e B é o menor conjunto *fuzzy* que contém A e B e a intersecção de dois conjuntos *fuzzy* é o maior conjunto *fuzzy* que está contido em A e B. Portanto, segue que $(L, \subseteq_f)$ é um reticulado.

Princípio da dualidade: Todo resultado obtido dos axiomas da estrutura algébrica se mantém válido se nele trocamos $\cup_f$ por $\cap_f$ e $\emptyset_f$ por V e vice-versa.

Há uma simetria entre os operadores $\cup_f$ e $\cap_f$ e os elementos $\emptyset_f$ e V, o que assegura que tanto os operadores $\cup_f$ e $\cap_f$ quanto os elementos $\emptyset_f$ e V podem ser permutados sem alterar a validade dos resultados obtidos.

Propriedade de idempotência:

$A \cup_f A = A$, pois

$$f_{A \cup A} = f_A \vee f_A = f_A \text{ e}$$

$A \cap_f A = A$, pelo princípio da dualidade. Basta trocarmos $\cup_f$ por $\cap_f$ e obtemos

$$f_{A \cap A} = f_A \wedge f_A = f_A$$

Propriedade comutativa:

$A \cup_f B = B \cup_f A$, pois

$$f_{A \cup B} = f_A \vee f_B = f_B \vee f_A = f_{B \cup A} \text{ e}$$

$A \cap_f B = B \cap_f A$, pelo princípio da dualidade.

Propriedade associativa:

$A \cup_f (B \cup_f C) = (A \cup_f B) \cup_f C$, pois

$$f_{A \cup (B \cup C)} = f_A \vee (f_B \vee f_C) = (f_A \vee f_B) \vee f_C = f_{(A \cup B) \cup C} \text{ e}$$

$A \cap_f (B \cap_f C) = (A \cap_f B) \cap_f C$, pelo princípio da dualidade.

Propriedade de absorção:

$A \cap_f (A \cup_f B) = A$, pois

$$f_{A \cap (A \cup B)} = f_A \wedge f_{A \cup B} = f_A \wedge (f_A \vee f_B) = f_A \text{ e}$$

$A \cup_f (A \cap_f C) = A$, pelo princípio da dualidade

Um conjunto não vazio, dotado de duas operações que obedecem às propriedades de idempotência, comutatividade, associatividade e absorção, também recebe o nome de reticulado, portanto (L, $\cup_f$, $\cap_f$) constitui um reticulado.

Proposição 2.1.3:

$B \subseteq_f A \Leftrightarrow A \cup_f B = A$, pois

$$f_{A \cup B} = f_A \Leftrightarrow f_A \vee f_B = f_A \Leftrightarrow f_B \leq f_A \text{ e}$$

$B \subseteq_f A \Leftrightarrow A \cap_f B = B$, pelo princípio da dualidade.

Propriedade distributiva:

$A \cup_f (B \cap_f C) = (A \cup_f B) \cap_f (A \cup_f C)$, pois

$$f_{A \cup (B \cap C)} = f_A \vee f_{B \cap C} = f_A \vee (f_B \wedge f_C) = (f_A \vee f_B) \wedge (f_A \vee f_C) =$$

$$f_{A \cup B} \wedge f_{A \cup C} = f_{(A \cup B) \cap (A \cup C)}$$

$A \cap_f (B \cup_f C) = (A \cap_f B) \cup_f (A \cap_f C)$, pelo princípio da dualidade.

Um reticulado que admite a propriedade distributiva é chamado *reticulado distributivo*. Assim, (L, $\cup_f$, $\cap_f$) constitui um reticulado distributivo.

Proposição 2.1.4:

$A \cup_f \emptyset_f = A$, pois

$$f_{A \cup \emptyset} = f_A \vee f_\emptyset = f_A \text{ e}$$

$A \cap_f \emptyset_f = \emptyset$, pois

$$f_{A \cap \emptyset} = f_A \wedge f_\emptyset = f_\emptyset$$

Proposição 2.1.5:

$A \cup_f V = V$, pois

$$f_{A \cup V} = f_A \vee f_V = f_V \text{ e}$$

$A \cap_f V = A$, pois

$$f_{A \cap V} = f_A \wedge f_V = f_A$$

Com base nas proposições anteriores, temos que $0_f = \emptyset_f$ e $1_f = V$ são, respectivamente, o ínfimo (zero) e o supremo (unidade) de L. Um reticulado que possui zero e unidade, atendendo a essas duas condições, é denominado reticulado completo. Assim, (L, $\cap_f$, $\cup_f$, 0_f, 1_f) caracteriza um reticulado completo.

Proposição 2.1.6:

$(A')' = A$, pois

$$f_{(A')'} = 1 - f_{A'} = 1 - [1 - f_A] = f_A$$

Proposição 2.1.7:

$A \subseteq_f B \Leftrightarrow B' \subseteq_f A'$, pois

$$A \subseteq_f B \Leftrightarrow f_A \leq f_B \Leftrightarrow f_{B'} \leq f_{A'} \Leftrightarrow B' \subseteq_f A'.$$

Leis de De Morgan:

I. $(A \cup_f B)' = A' \cap_f B'$

II. $(A \cap_f B)' = A' \cup_f B'$

Um reticulado distributivo que admite o complemento de De Morgan é chamado *reticulado de De Morgan*. Assim, a estrutura (L, $\subseteq_f$, $\cap_f$, $\cup_f$, ') é um reticulado de De Morgan completo.

2.1.3 Relações fuzzy

Sejam A um conjunto *fuzzy* de domínio U e B um conjunto *fuzzy* de domínio V. O produto cartesiano *fuzzy* de A e B, denotado por $A \times_f B$ é definido por:

$$A \times_f B = \{((u, v), f_A(u) \wedge f_B(v)) / u \in U, v \in V\}.$$

Exemplo:

Sejam U = {a, b} e V = {p, q, r}os domínios dos conjuntos *fuzzy* A = {(a, 0.4); (b,0.6)} e B = {(p, 0.8); (q, 0.2); (r, 0.3)}. Então:

$$A \times_f B = \{((a,p),\ 0.4);\ ((a,q),\ 0.2);\ ((a,r),\ 0.3);\ (b,p),\ 0.6);\ ((b,q),\ 0.2);\ ((b,r),\ 0.3)\}\}$$

Por tratarmos de dois conjuntos finitos, podemos representá-los matricialmente:

$$A \times_f B = \begin{bmatrix} 0.4 & 0.2 & 0.3 \\ 0.6 & 0.2 & 0.3 \end{bmatrix}$$

Uma *relação fuzzy* de A em B, indicada por R_f, é um subconjunto *fuzzy* do produto cartesiano $A \times_f B$, caracterizado por uma função de pertinência f_R que associa a cada par (x, y) o seu grau de pertinência $f_R(x, y)$ em R_f.

Se A e B são conjuntos usuais, então reduzimos o produto cartesiano *fuzzy* ao produto cartesiano usual. Se A e B são conjuntos *fuzzy*, temos que $f_R(x, y) \leq f_A(x) \wedge f_B(y)$.

Uma *relação fuzzy* n-ária é um subconjunto *fuzzy* R_f do produto cartesiano $A_1 \times_f A_2 \times_f \ldots \times_f A_n$, onde:

$$f_R(x_1,\ x_2,\ \ldots,\ x_n) \leq \wedge_{1 \leq i \leq n} f_{Ai}(x_i)$$

Dado A um conjunto *fuzzy*, o *suporte* de A é o conjunto usual:

$$S(A) = \{x/\ x \in V\ e\ f_A(x) > 0\}$$

O *domínio* da relação R_f, denotado por Dom(R_f) é o conjunto *fuzzy* definido por:

$$f_{Dom(R)}(x) = Sup_x\, f_R(x,\ y)$$

A imagem da relação R_f, denotado por Im(R_f) é o conjunto *fuzzy* definido por:

$$f_{Im(R)}(x) = Sup_y\, f_R(x,\ y)$$

O *campo* da relação R_f, denotado por *fl*(R_f), é definido por:

$$fl(R_f) = Dom(R_f) \cup Im(R_f)$$

O *comprimento* da relação *fuzzy* R_f, denotado por h(R_f), é definido por:

$$h(R_f) = Sup\{f_R(x,\ y)/\ (x,\ y) \in S(R_f)\}$$

Uma relação *fuzzy* R_f é *subnormal* se $h(R_f) < 1$, e é *normal* se $h(R_f) = 1$.

Seja A um conjunto *fuzzy* com domínio V e x, y ∈ V:

A *relação identidade* I_f em A é dada por:

$$f_I(x,y) = \begin{cases} 0, \text{ se } x \neq y \\ 1, \text{ se } x = y \end{cases}$$

A *relação nula* 0_f em A é definida por $f_0(x,y) = 0$

Exemplo:

Seja R_f = {((a, p); 0.8), ((a, q); 0.9), ((a, r); 0.5), ((b, p); 0.2), ((b, q); 0.7), ((b, r); 0.6)}

$Dom(R_f) = \{(a, 0.9); (b, 0.7)\}$

$Im(R_f) = \{(p, 0.8); (q, 0.9); (r, 0.6)\}$

$h(R_f) = 0.9$, logo, R_f é subnormal

$S(R_f) = \{(a, p); (a, q); (a, r); (b, p); (b, q); (b, r)\}$.

Se R_f é uma relação *fuzzy* de A em B e S_f é uma relação *fuzzy* de B em C, então a *composição* de R e S é uma relação *fuzzy* de A em C, denotada por RoS e com a função de pertinência definida por:

$$f_{RoS}(x,z) = Sup_y\{f_R(x,y) \wedge f_S(y,z)\}$$

Exemplo:

Consideremos as seguintes relações *fuzzy*:

$R_f = \{((p,r); 0.5), ((p,s); 0.8), ((q,r); 0.4); ((q,s); 0.7)\}$

$S_f = \{((r,c); 0,7), ((r,d); 0.1), ((s,c); (0.2)), ((s,d); 0.3)\}$

Então:

$f_{RoS}(p,c) = Sup\{(p,r) \wedge (r,c); (p,s) \wedge (s,c)\} = Sup\{0.5; 0.2\} = 0.5$

$f_{RoS}(p,d) = Sup\{(p,r) \wedge (r,d); (p,s) \wedge (s,d)\} = Sup\{0.1; 0.3\} = 0.3$

$f_{RoS}(q,c) = Sup\{(q,r) \wedge (r,c); (q,s) \wedge (s,c)\} = Sup\{0.4; 0.2\} = 0.4$

$f_{RoS}(q,d) = Sup\{(q,r) \wedge (r,d); (q,s) \wedge (s,d)\} = Sup\{0.4; 0.3\} = 0.4$

Uma vez estabelecidos os fundamentos dos conjuntos *fuzzy* e sua capacidade de representar incertezas através de graus de pertinência, faz-se necessário estender esses conceitos ao âmbito do raciocínio lógico. Nesse sentido, o capítulo seguinte tratará da lógica *fuzzy* e seus mecanismos de inferência.

2.2 A lógica fuzzy

De maneira semelhante à extensão do conceito do conjunto, Zadeh (1965) também propôs uma extensão da lógica multivalorada, a lógica *fuzzy*, que nos permite presumir soluções e resoluções de problemas, mesmo que encontradas imprecisões nos dados e/ou informações e que nos permite obter conclusões precisas. Nesta lógica, o raciocínio exato corresponde a um caso limite do raciocínio aproximado, sendo interpretado como um processo de composição de relações nebulosas.

Nos sistemas lógicos multivalorados, o valor de verdade de uma proposição pode assumir diferentes formas: pode ser um elemento de um conjunto finito, pertencer a um intervalo numérico ou ainda fazer parte de uma álgebra booleana. Na lógica nebulosa (ou *fuzzy*), os valores de verdade são expressos linguísticamente, por exemplo: "verdade", "muito verdade", "não verdade", "falso", "muito falso", entre outros. Cada um desses termos linguísticos é interpretado como um subconjunto *fuzzy* do intervalo real unitário [0, 1].

Enquanto os sistemas binários trabalham com predicados exatos, como "par", "maior que" e "menor que", a lógica *fuzzy* opera com predicados imprecisos ou graduais, como "alto", "baixo" e "magro". Nos sistemas clássicos, a principal operação de modificação é a negação. Em contraste, a lógica *fuzzy* permite uma ampla gama de modificadores linguísticos, como "muito", "mais ou menos", "um pouco", entre outros, que são fundamentais para a construção de expressões mais próximas da linguagem natural.

Além disso, a lógica clássica se restringe aos quantificadores existencial ($\exists$) e universal ($\forall$). A lógica *fuzzy*, por sua vez, admite uma variedade muito maior de quantificadores linguísticos, como "poucos", "vários", "usualmente", "frequentemente", "em torno de", entre outros (Gomide; Gudwin, 1994)

A lógica *fuzzy* é uma lógica que lida com modelos de raciocínios imprecisos ou aproximados. Ela nos permite trabalhar com problemas de tomada de decisões que não são

facilmente representados nos modelos matemáticos usuais. Um controle *fuzzy* busca "imitar" um agente humano a partir de uma análise descritiva e experimental de um processo específico (Zadeh, 1973).

Fenômenos quantitativos são frequentemente representados por variáveis numéricas, que permitem medições exatas e análises objetivas. No entanto, essas variáveis mostram-se inadequadas para modelar fenômenos qualitativos, que envolvem aspectos mais sutis e subjetivos. É nesse contexto que as variáveis linguísticas se destacam como uma abordagem mais apropriada. Elas permitem descrever fenômenos complexos e imprecisos através de palavras e expressões, sejam elas de linguagem natural ou artificial. Dessa forma, as variáveis linguísticas oferecem uma representação mais abrangente e intuitiva, capturando nuances e interpretações que números por si só não conseguiriam expressar, tornando-as fundamentais para a análise de situações onde a ambiguidade e a complexidade são predominantes (Marro et al., 2010).

Quando falamos de inteligência artificial, a lógica *fuzzy* é um mecanismo fundamental ao buscarmos representar o pensamento e o raciocínio, devido a possibilidade de trabalhar com incertezas, raciocínio aproximado, termos vagos e ambíguos, que são as maneiras pelas quais as pessoas expressam seus pensamentos. Dessa forma, a lógica *fuzzy* permite aos sistemas computacionais inteligentes "raciocinar" considerando aspectos inerentes à incerteza e aos processos realísticos e torná-lo mais "humano" (Marro et al., 2010).

A inteligência artificial depende de uma expressiva utilização da lógica, tendo em vista que a lógica é o principal formalismo de representação do conhecimento, sendo útil no desenvolvimento de sistemas inteligentes, em especial os especialistas e os multiagentes, visto que, de acordo com Luger (2005) e Konar (2000), a imprecisão dos dados e a incerteza do conhecimento são parte do problema em si, sendo o real desafio da inteligência artificial, e raciocinar considerando esses aspectos sem uma fundamentação adequada pode ocasionar imprecisões nas inferências obtidas.

A seguir, será realizada uma comparação entre a lógica clássica e a lógica fuzzy, com o objetivo de destacar suas principais diferenças conceituais e aplicacionais. Para isso, utilizaremos como referência a obra de Barros e Bassanezi (2006).

2.2.1 Conectivos na lógica clássica

A lógica matemática contempla o estudo dos conectivos de conjunção (∧), disjunção (∨), negação (¬) e implicação (→), utilizados em sentenças do tipo:

"Se a está em A **e** b está em B,

então c está em C **ou** d **não** está em D" (2.2)

Os valores lógicos destes conectivos são estudados através de tabelas de verdade. Desse modo, o valor lógico de uma sentença, formada por duas ou mais proposições, é obtido pelas composições das tabelas de verdade dos conectivos presentes.

Na lógica clássica, sentenças verdadeiras possuem valor 1, enquanto sentenças falsas possuem valor 0. A seguir, serão apresentadas as tabelas de verdade dos conectivos para a lógica clássica:

Tabela 1: Negação

p	$\neg p$
1	0
0	1

Tabela 2: Conjunção

p	q	$p \wedge q$
1	1	1
1	0	0
0	1	0
0	0	0

Tabela 3: Disjunção

p	q	$p \vee q$
1	1	1
1	0	1
0	1	1
0	0	0

Tabela 4: Implicação

p	q	$p \rightarrow q$
1	1	1
1	0	0
0	1	1
0	0	1

Cada um dos conectivos apresentados pode ser visto como um operador matemático, cujos valores são definidos como os das respectivas tabelas.

- Conjunção: ∧

$$\wedge : \{0,1\} \times \{0,1\} \Rightarrow \{0,1\}$$

$$(p,q) \Rightarrow \wedge\,(p,q) = p \wedge q = min\{p,q\}$$

Logo,

$$\wedge\ (1,1)\ =\ 1 \wedge 1 = 1$$

$$\wedge\ (1,0)\ =\ 1 \wedge 0 = 0$$

$$\wedge\ (0,1)\ =\ 0 \wedge 1 = 0$$

$$\wedge\ (0,0)\ =\ 0 \wedge 0 = 0$$

- Disjunção: ∨

$$\vee\ :\{0,1\}\ \times\ \{0,1\}\ \Rightarrow\ \{0,1\}$$

$$(p,q) \Rightarrow \vee\ (p,q)\ =\ p\ \vee\ q\ =\ max\{p,q\}$$

Logo,

$$\vee\ (1,1)\ =\ 1 \wedge 1 = 1$$

$$\vee\ (1,0)\ =\ 1 \wedge 0 = 1$$

$$\vee\ (0,1)\ =\ 0 \wedge 1 = 1$$

$$\vee\ (0,0)\ =\ 0 \vee 0 = 0$$

- Negação: ¬

A negação é uma operação unária, em que:

$$\neg : \{0,1\}\ \Rightarrow \{0,1\}$$

$$p \Rightarrow \neg p$$

onde, $\neg 1\ =\ 0\ e\ \neg 0\ \ =\ 1$.

- Implicação: →

$$\rightarrow\ :\{0,1\}\ \times\ \{0,1\}\ \Rightarrow\ \{0,1\}$$

$$(p,q) \Rightarrow\ \ \rightarrow (p,q)\ =\ (p \rightarrow q)$$

A partir desses conectivos é possível obter três fórmulas básicas que reproduzem a tabela de verdade da implicação:

1) $(p \rightarrow q)\ =\ (\neg p)\ \vee\ q$
2) $(p \rightarrow q)\ =\ (\neg p)\ \vee\ (p\ \wedge\ q)$
3) $(p \rightarrow q)\ =\ max\{x\ \in \{0,1\} : p\ \wedge\ x\ \leq q\}$

Agora, voltamos à sentença (2.2). Esta sentença pode ter uma avaliação lógica através dos valores lógicos dos conectivos. Como os conjuntos são a princípio clássicos, essa avaliação obtém valores de verdade 0 ou 1.

Considere:

p = Se a está em A

q = b está em B

r = c está em C

s = d não está em D

P = Se a está em A e b está em B

Q = c está em C ou d não está em D

Os valores de cada uma das expressões p, q, r e s podem ser somente 0 ou 1, a depender da pertinência de cada um deles ao conjunto indicado. Por exemplo, p = 1 se a está em A e p = 0 se a não está em A.

Dessa forma, podemos avaliar a sentença para cada situação. Exemplo:

$$\text{Se } a \in A(p = 1);\ b \notin B(q = 0);\ c \in C(r = 1)\ e\ d \notin D(s = 1)$$

então o valor lógico da sentença é:

$$(0 \wedge 1) \rightarrow (1 \vee 1) = (0 \rightarrow 1) = 1$$

Quando pensamos na lógica *fuzzy*, o valor lógico da sentença p: "a está em A" coincide com o valor obtido com a função característica do conjunto A em a. Em seguida, abordaremos o caso *fuzzy* para a sentença apresentada.

2.2.2 Conectivos na lógica *fuzzy*

Barros e Bassanezi (2006) propõem admitirmos agora que os conjuntos da expressão apresentada sejam conjuntos *fuzzy*. De início, atribui-se um valor que indique o quando a proposição "a está em A" é verdadeira, com A sendo *fuzzy*. Assim, a pode pertencer a A com valores no intervalo [0,1].

As extensões dos conectivos clássicos para conectivos *fuzzy*, são obtidas a partir das normas e conormas triangulares.

Definição 2.2.1 (t-norma). O operador $\Delta : [0,1] \times [0,1] \to [0,1]$, $\Delta(x,y) = x \Delta y$, é uma *t-norma*, se satisfaz as condições:

t_1) *elemento neutro:* $\Delta(1, x) = 1 \Delta x = x$

t_2) *comutativa:* $\Delta(x,y) = x \Delta y = y \Delta x = \Delta(y,x)$

t_3) *associativa:* $x \Delta (y \Delta z) = (x \Delta y) \Delta z$

t_4) *monotonicidade:* se $x \leq u \, e \, y \leq v$, $então \, x \Delta y \leq u \Delta v$

A operação *t-norma* estende o operador ∧ que modela o conectivo "e".

Alguns exemplos de *t-norma* são:

$$\Delta_1(x,y) = min\{x,y\} = x \wedge y$$

$$\Delta_2(x,y) = xy$$

$$\Delta_3(x,y) = max\{0, x + y - 1\}$$

$$\Delta_4(x,y) = \begin{cases} x & \text{se y = 1} \\ y & \text{se x = 1} \\ 0 & \text{caso contrário} \end{cases}$$

Definição 2.2.2 (t-conorma). O operador $\bigtriangledown(x,y) = x \bigtriangledown y$, é uma *t-conorma*, se satisfaz as condições:

c_1) *elemento neutro:* $\bigtriangledown(0, x) = 0 \bigtriangledown x = x$

c_2) *comutativa:* $\bigtriangledown(x,y) = x \bigtriangledown y = y \bigtriangledown x = \bigtriangledown(y,x)$

c_3) *associativa:* $x \bigtriangledown (y \bigtriangledown z) = (x \bigtriangledown y) \bigtriangledown z$

c_4) *monotonicidade:* se $x \leq u \, e \, y \leq v$, $então \, x \bigtriangledown y \leq u \bigtriangledown v$

A operação *t-conorma* estende o operador ∨ que modela o conectivo "ou".

Alguns exemplos de *t-conorma* são:

$$\bigtriangledown_1(x,y) = max\{x,y\} = xy$$

$$\bigtriangledown_2(x,y) = min\{1, x + y\}$$

$$\bigtriangledown_3(x,y) = x + y - xy$$

Definição 2.2.3 (negação). Uma aplicação $\eta: [0,1] \to [0,1]$ é uma negação se satisfaz as condições:

n_1) *fronteiras*: $\eta(0) = 1 \; e \; \eta(1) = 0$;

n_2) *involução*: $\eta(\eta(x)) = x$

n_3) *monotonicidade*: η é decrescente

As aplicações:

$$\eta_1 = 1 - x \quad \text{e} \qquad \eta_2 = \frac{1-x}{1+x}$$

reproduzem a tabela verdade da negação ¬.

Além disso, as operações $\Delta = \wedge$, $\bigtriangledown = \vee$ e $\eta = 1 - x$, satisfazem as leis De Morgan, isto é, para todo par (x, y) de $[0,1] \times [0,1]$ valem:

$$\eta(x \wedge y) = \eta(x) \vee \eta(y)$$

$$\eta(x \vee y) = \eta(x) \wedge \eta(y)$$

Definição 2.2.4 (implicação *fuzzy*). Um operador $\Rightarrow: [0,1] \times [0,1] \to [0,1]$ é considerado uma implicação fuzzy quando satisfaz as seguintes condições fundamentais:

Reprodução da tabela clássica: Deve reproduzir a tabela da implicação clássica nos casos extremos

Decrescimento na primeira variável: Para cada $x \in [0, 1]$, tem-se $(a \Rightarrow x) \leq (b \Rightarrow x)$ quando $a \geq b$

Crescimento na segunda variável: Para cada $x \in [0, 1]$, tem-se $(x \Rightarrow a) \geq (x \Rightarrow b)$ quando $a \geq b$

Portanto, a classe das implicações *fuzzy* é composta por todas as aplicações do quadrado unitário $[0, 1] \times [0, 1]$ em $[0, 1]$, cuja restrição aos vértices coincidem com os valores da implicação clássica, sendo decrescentes em relação às abscissas e crescentes em relação às ordenadas.

Na lógica clássica, a implicação pode ser representada por três fórmulas equivalentes:

1. $(p \Rightarrow q) = (\neg p) \vee q$
2. $(p \Rightarrow q) = (\neg p) \vee (p \wedge q)$

3. $(p \Rightarrow q) = \max\{x \in \{0, 1\} : p \wedge x \leq q\}$

No contexto *fuzzy*, essas fórmulas não produzem as mesmas implicações. Por isso, distinguimos três tipos principais:

I. S-Implicações

$$(x \Rightarrow y) = \eta(x) \bigtriangledown y$$

As S-implicações são construídas a partir de conormas (também chamadas de s-normas).

II. Q-Implicações

$$(x \Rightarrow y) = \eta(x) \bigtriangledown (x \bigtriangleup y)$$

O nome "Q-implicações" deriva de sua origem na mecânica quântica.

III. R-Implicações

$$(x \Rightarrow y) = \sup\{z \in [0, 1] : x \bigtriangleup z \leq y\}$$

O termo "R-implicação" provém de "operação residual". Pode ser interpretada como: $(x \Rightarrow y)$ representa o maior valor com que y supera x segundo $\bigtriangleup$, ou seja, é o resíduo de x em relação a y, segundo $\bigtriangleup$.

Principais Implicações *Fuzzy*

a) Implicação de Gödel

$$(x \Rightarrow y) = g(x, y) = \begin{cases} 1 & se\ x \leq y \\ y & se\ x > y \end{cases}$$

b) Implicação de Goguen

$$(x \Rightarrow y) = g_{\square}(x, y) = \begin{cases} 1 & se\ x \leq y \\ \frac{y}{x} & se\ x > y \end{cases}$$

c) Implicação de Lukasiewicz

$$(x \Rightarrow y) \;=\; \ell(x,\, y) \;=\; min\{(1 \;-\; x \;+\; y),\, 1\}$$

d) Implicação de Kleene-Dienes

$$(x \Rightarrow y) \;=\; k_{\mathrm{a}}(x,\, y) \;=\; max\{(1 \;-\; x),\, y\}$$

e) Implicação de Reichenbach

$$(x \Rightarrow y) \;=\; r(x,\, y) \;=\; (1 \;-\; x \;+\; xy)$$

f) Implicação de Zadeh

$$(x \Rightarrow y) \;=\; z(x,\, y) \;=\; max\{(1 \;-\; x),\, min(x,\, y)\}$$

A implicação de Zadeh é uma Q-implicação: (x ⇒ y) = (1 - x) ∨ (x ∧ y) = η(x) ∨ (x ∧ y)

g) Implicação de Gaines-Rescher

$$(x \Rightarrow y) \;=\; g_{\mathrm{r}}(x,\, y) \;=\; \begin{cases} 1 & se\ x \leq y \\ 0 & se\ x \;>\; y \end{cases}$$

h) Implicação de Wu

$$(x \Rightarrow y) \;=\; w(x,\, y) \;=\; \begin{cases} 1 & se\ x \leq y \\ min\{1 \;-\; x,\, y\} & se\ x \;>\; y \end{cases}$$

Exemplo 2.2.1: Vamos retomar a expressão (2.2) e calcular seu valor lógico considerando:

- △ = ∧ (conjunção como mínimo)
- ▽ = ∨ (disjunção como máximo)
- η(x) = 1 - x (negação padrão)
- Implicação de Gödel

Dados de entrada:

φ(a) = 0,6, φ(b) = 0,7, φ(c) = 0,4 φ(d) = 0,7

Cálculos:

p △ q = min(0,6; 0,7) = 0,6

s = 1 - φ(d) = 1 - 0,7 = 0,3

r $\bigtriangledown$ s = max(0,4; 0,3) = 0,4

Resultado final: O valor lógico de (2.2) é obtido pela aplicação (p $\bigtriangleup$ q) $\Rightarrow$ (r $\bigtriangledown$ s).

Usando a implicação de Gödel

$$(p \bigtriangleup q) \Rightarrow (r \bigtriangledown s) = \begin{cases} 1 & se\ (p \bigtriangleup q) \leq (r \bigtriangledown s) \\ (r \bigtriangledown s) & se\ (p \bigtriangleup q) > (r \bigtriangledown s) \end{cases} = 0{,}4$$

Como $(p \bigtriangleup q) = 0,6 > (r \bigtriangledown s) = 0,4$, o resultado é 0,4.

Portanto, para as pertinências dadas, a expressão (2.2) é verdadeira com grau 0,4.

A sensibilidade da lógica *fuzzy* às diferenças sutis nos valores de entrada permite uma avaliação mais refinada. Um sistema de controle baseado nessa lógica poderia, por exemplo, ajustar sua resposta proporcionalmente ao grau de verdade, enquanto um sistema clássico aplicaria uma ação binária.

Dando continuidade à teoria da lógica *fuzzy*, as próximas seções, fundamentadas na proposta de Feitosa e Paulovich (2005), detalharão seus principais elementos.

2.2.3 Elementos da lógica *fuzzy*

A utilização de variáveis linguísticas nos permitem uma melhor caracterização de fenômenos complexos ou imprecisos, onde seus valores são palavras em uma linguagem natural ou artificial, assumindo como valor sentenças ou palavras nesta linguagem.

Por exemplo, a palavra "idade", bastante presente em nosso cotidiano, não possui em ambientes numéricos uma noção clara de seu significado. Através dos conjuntos *fuzzy*, conseguimos atribuir noções que se aproximam de 'idade', nossa variável linguística, definindo os termos da variável linguística: muito jovem, jovem, meia-idade, velho, muito velho e entre outras, que são interpretados por conjuntos *fuzzy*. Desse modo, cada termo determina um conjunto *fuzzy* cuja função de pertinência seja apropriada ao termo.

Definição 2.2.5: Uma variável linguística é caracterizada por uma quíntupla (N, Gr, U, T(N), S(N)), onde:

N: Nome da variável

Gr: Regra sintática que permite gerar valores linguísticos

U: Universo de discurso

T(N): Conjunto dos termos de N (subvariáveis)

S(N): Regra semântica que associa a cada termo x de N gerado por V o seu significado com valores no intervalo real [0,1]

Exemplificando:

T(estatura) = {muito baixa, baixa, pouco baixa, mais ou menos alta, alta, muito alta, ...} e U = [0.5; 2.5] em metros;

T(sensação térmica) = {muito frio, frio, pouco quente, quente, muito quente, ...} e U = [-30; 50] em graus Celsius.

Os termos de N representam um sumário de várias subclasses de elementos considerados no universo de discurso.

Definição 2.2.6: Seja F um subconjunto *fuzzy* de N, com universo de discurso V. Então F é uma *restrição fuzzy* em N se:

$$F = \{(u, f_F(u))/u \in N\}$$

Definição 2.2.7: Uma *variável fuzzy* é caracterizada por uma terna (N, V, M(N, u)), onde:

N: Nome da variável

V: Universo de discurso

M(N, u): Valoração no intervalo real [0,1] que representa uma restrição *fuzzy* sobre os valores de u impostos por N.

Dos exemplos anteriores, temos:

N = alto e V = [1.5; 2.1], o universo de alto sendo uma restrição do universo de estatura [0; 2.5]

Na variável linguística "idade", as restrições impostas a cada termo estabelecem o significado numérico de cada um deles. Dessa forma, os termos são definidos por suas respectivas funções de pertinência. As funções de pertinência para cada termo (como "muito jovem", "jovem", "meia-idade", "velho" e "muito velho") associados à variável linguística "idade" podem ser representadas da seguinte maneira:

$$f_{muito\ jovem}(x) = \begin{cases} 1 & \text{para } 0 \leq x \leq 5, \\ \frac{(30-x)}{25} & \text{para } 5 \leq x \leq 30, \end{cases}$$

$$f_{jovem}(x) = \begin{cases} \frac{(x-5)}{25} & \text{para } 5 \leq x \leq 30, \\ \frac{(50-x)}{20} & \text{para } 30 \leq x \leq 50, \end{cases}$$

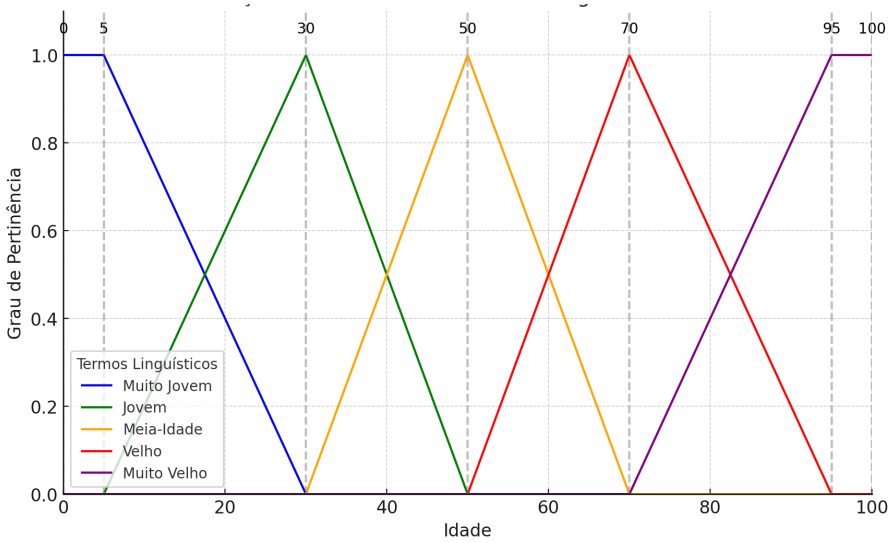
$$f_{meia-idade}(x) = \begin{cases} \frac{(x-30)}{20} & \text{para } 30 \leq x \leq 50, \\ \frac{(70-x)}{20} & \text{para } 50 \leq x \leq 70, \end{cases}$$

$$f_{velho}(x) = \begin{cases} \frac{(x-50)}{20} & \text{para } 50 \leq x \leq 70, \\ \frac{(95-x)}{25} & \text{para } 70 \leq x \leq 95, \end{cases}$$

$$f_{muito\ velho}(x) = \begin{cases} \frac{(x-70)}{25} & \text{para } 70 \leq x \leq 95, \\ 1 & \text{para } 95 \leq x \leq 100. \end{cases}$$

O gráfico da variável linguística "idade" é apresentado considerando o universo de discurso no intervalo [0, 100]. Ele foi construído com base nas funções de pertinência previamente definidas para cada termo linguístico, representando graficamente os graus de pertinência em relação à faixa etária correspondente.

Figura 2: Gráfico da variável linguística ‘idade’



Fonte: Autora

Para M(velho, u), com u = 70, temos $f_{velho}(70) = 1$. Esse valor indica o grau de pertinência de u no conjunto ‘velho’. Isso significa que o pertencimento ao conjunto ‘velho’ é máximo, de acordo com os critérios estabelecidos pela função de pertinência f_{velho}. Por outro lado, para M(jovem, u), com u = 50, temos $f_{jovem}(50) = 0$, que representa o grau de pertinência de u no conjunto ‘jovem’, desse modo, o pertencimento ao conjunto ‘jovem’ é mínimo conforme os critérios definidos pela função de pertinência f_{jovem}.

A lógica *fuzzy* nos permite modelar situações em que os limites entre categorias, como jovem e velho, não são bem definidos, utilizando os valores de pertinência entre 0 e 1 e permitindo representar transições graduais entre essas categorias.

Uma proposição *fuzzy* estabelece um relação entre um elemento u de um domínio U e um termo predicado *fuzzy* F. Esse predicado é representado por um conjunto *fuzzy* definido sobre um universo de discurso V, que pode ser diferente do domínio U. A proposição pode ser expressa por P: “u é F”.

Considere a proposição *fuzzy* “x é F”, onde R(A(x)) representa a restrição *fuzzy* aplicada ao sujeito x em relação ao atributo A, com x pertencendo ao universo de discurso V. Nesse caso, R(A(x)) é uma subvariável *fuzzy* determinada pelo termo predicado F, e temos:

$$R(A(x)) = F$$

Seja F um subconjunto *fuzzy* no universo de discurso V. Suponha que A(X) seja uma subvariável *fuzzy* cujos valores estão em V, e que R(A(x)) seja uma restrição associada a F. Então, a distribuição de possibilidades Π_x, associada com A(x) é definida por:

$$\Pi_x = R(A(x))$$

Assim, a distribuição de possibilidades Π_x é equivalente à restrição *fuzzy* associada à subvariável *fuzzy* A(x). Em outras palavras, a possibilidade de x admitir o atributo A é determinada pela compatibilidade de x com o termo predicado F, expresso como:

$$F = R(A(x)) = \Pi_x$$

Considerando o gráfico do exemplo anterior, da variável linguística idade, consideremos o universo de discurso V = {1, 2, 3, …, 110} e o conjunto *fuzzy* 'muito jovem', indicado por P, que representa a sentença "x é P".

P = {(1, 1); (2, 1); (3, 1); (4, 1); (5, 1); (6, 0.9); (7, 0.9); (8, 0.9); (9, 0.8); (10, 0.8); (11, 0.8); (12, 0.7), …, (29, 0.1); (30, 0.1)}.

Assim, a proposição "x é P" associa x à distribuição de possibilidades $\Pi_x = P$.

Uma proposição *fuzzy* P pode envolver múltiplos atributos associados a um sujeito. Por exemplo, na proposição "Juliana é alta, jovem e morena", o predicado *fuzzy* é composto por três atributos: altura, idade e cor de cabelo. Cada atributo pode ser representado por um conjunto *fuzzy* que define seu grau de pertinência.

Uma proposição *fuzzy* do tipo "x é F", onde F representa uma relação n-ária definida nos universos $V_1 \times V_2 \times \ldots \times V_n$, pode ser expressa em termos dos n atributos de x da seguinte maneira:

$$x \text{ é } F \rightarrow R(A_1(x), A_2(x), \ldots, A_n(x)) = F,$$

e, ampliando o conceito de distribuição de possibilidades, obtemos:

$$x \text{ é } F \rightarrow \Pi_{(A_1(x), A_2(x), \ldots, A_n(x))} = F,$$

trata-se de uma distribuição de possibilidades gerada pela proposição "x é F", onde:

$$\omega_{(A_1(x), A_2(x), \ldots, A_n(x))}(u_1, u_2, \ldots, u_n) = f_R(u_1, u_2, \ldots, u_n)$$

Assim, encontra-se uma proposição *fuzzy* com n atributos $A_1(x), A_2(x), \ldots, A_n(x)$, onde o significado é obtido através de uma distribuição de possibilidades n-ária. Com apenas uma proposição *fuzzy* relaciona-se n atributos interpretados por suas distribuições de possibilidades.

Considere a proposição *fuzzy* "João é alto, velho, moreno e elegante" que pode ser traduzida por uma proposição do tipo:

$$x \text{ é } F \rightarrow \Pi_{(A_1(x), \ldots, A_4(x))},$$

onde $A_1(x)$ = alto, $A_2(x)$ = velho, $A_3(x)$ = moreno, $A_4(x)$ = elegante.

Toda proposição *fuzzy*, seja em linguagem natural ou artificial, pode ser interpretada por uma distribuição de possibilidades. Da mesma forma, se conhecemos uma distribuição de possibilidades, então podemos determinar uma proposição *fuzzy* P:

$$P \leftarrow \Pi_{(x_1, x_2, \dots, x_n)}$$

A informação dada por P, uma proposição "x é F", é transformada em uma distribuição de possibilidades Π_x da variável X, representada como um conjunto *fuzzy*, o que também pode ocorrer reciprocamente.

2.2.4 Valor de verdade *fuzzy*

O conceito de valor de verdade é algo intrigante na lógica *fuzzy*. Zadeh (1965) foi além e determinou que os valores de verdade também podem assumir subvariáveis *fuzzy* como valores. O valor de verdade *fuzzy* é um conjunto *fuzzy* definido em V = [0,1]. Denota-se:

$$f_{v(P)}: [0, 1] \rightarrow [0, 1]$$

Esse conjunto de valores de verdade trata-se de um conjunto enumerável de variáveis *fuzzy*, na forma:

> δ = {verdadeiro, muito verdadeiro, não verdadeiro, falso, mais ou menos falso, não falso, não muito falso, quase falso, absolutamente verdadeiro, …},

em que cada elemento de δ é gerado pelos termos verdadeiro e falso.

Esses valores de verdade de δ são valores linguísticos que são interpretados por funções, o que os diferencia dos valores usuais de verdade da lógica clássica ou das lógicas multivaloradas.

De forma geral, dada uma proposição P, seu valor de verdade $v(P)$ não é um elemento de δ. No entanto, tomamos um valor linguístico $v(P)^*$ que está em δ, denominada uma aproximação linguística de $v(P)$. Neste cenário, falso não equivale a não verdadeiro, logo:

Se

$$verdadeiro = \{(x, f_x)/x \in [0, 1]\} \; e \; f_x \; o \; grau \; de \; verdade \; de \; x\}$$

Então:

$$falso = \{(1 - x, f_x)/x \in [0,1]\}\ e$$

$$não\ verdadeiro = \{(x, 1 - f_x)/x \in [0,1]\}$$

Exemplificando:

Considere o domínio V = {0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1} de uma função de verdade, onde:

verdadeiro = {(0.4, 0.2); (0.5, 0.4); (0.6, 0.6); (0.7, 0.8)}

Temos que:

falso = {(0.6, 0.2); (0.5, 0.4); (0.4, 0.6); (0.3, 0.8)} e

não verdadeiro = {(0, 1), (0.1, 1), (0.2, 1), (0.3, 1), (0.4, 0.8), (0.5, 0.6), (0.6, 0.4), (0.7, 0.2)}

Os valores "verdadeiro", "não verdadeiro" e "falso" são atributos da variável linguística "valor de verdade", e seu conjunto de termos pode ser derivado diretamente a partir do termo atômico "verdadeiro". Dessa forma, eliminamos o princípio do terceiro excluído, uma vez que, na lógica *fuzzy*, uma proposição não precisa ser estritamente verdadeira ou completamente falsa.

2.2.5 Regras de dedução

Nesta seção, apresentaremos o conjunto de regras de dedução da lógica *fuzzy*, iniciando pela equivalência e implicação *fuzzy*.

2.2.5.1 EQUIVALÊNCIA E IMPLICAÇÃO

Sejam P e Q duas proposições *fuzzy*, com as distribuições de possibilidades:

$$P \rightarrow \Pi^{P}_{(x_1, \ldots, x_n)} = F_f$$

$$Q \rightarrow \Pi^{Q}_{(x_1, \ldots, x_n)} = G_f$$

Regra de equivalência (RE): As proposições T e J são equivalentes se suas distribuições de possibilidades são iguais, isto é:

$$P \Leftrightarrow Q\ se,\ e\ somente\ se, \Pi^{P}_{(x_1, \ldots, x_n)} = \Pi^{Q}_{(x_1, \ldots, x_n)}$$

Regra de implicação (RI): A proposição T implica a proposição J se a distribuição de possibilidades de T está contida na distribuição de possibilidades de J, isto é:

$$P \Rightarrow Q \; se, \; e \; somente \; se, \Pi^{P}_{(x_1, \ldots, x_n)} \subseteq \Pi^{Q}_{(x_1, \ldots, x_n)}$$

2.2.5.2 PROJEÇÃO E PARTICULARIZAÇÃO

Seja $S = i_1, i_2, \ldots, i_n$ uma sequência indexada. Indicamos uma subsequência de S por $s = i_{j1}, i_{j2}, \ldots, i_{jk}$, quando $1 \leq j_i \leq n$

Regra da projeção (RPr): A projeção de $\Pi_x(x_1, x_2, \ldots, x_n)$ sobre $U_{(S)}$, em que $U_{(S)} = U_{j1} \times U_{j2} \times \ldots \times U_{jk}$ é a distribuição de possibilidades k-ária cujo domínio é $U_{(S)}$ e seu grau de pertinência é o seu supremo dentre $\Pi_{x(S)}(u_{(S)})$.

Denotamos a projeção de $\Pi_x(x_1, x_2, \ldots, x_n)$ sobre $U_{(S)}$ por:

$$\Pi_{x(S)} = Proj_{U(S)} \Pi_{(x_1, x_2, \ldots, x_n)}$$

No domínio, é realizada a projeção convencional para uma n-upla qualquer. No entanto, pode acontecer que o mesmo elemento esteja associado a múltiplos valores de pertinência. Nessa situação, opta-se pelo maior valor, ou seja, o supremo desses valores.

Desse modo, a distribuição $\Pi_{x(S)}$ na subvariável $X_{(S)} = (x_{j1}, x_{j2}, \ldots, x_{jk})$ é obtida pela projeção de Π_x sobre $U_{(S)}$.

A seguir, dada uma distribuição de possibilidades $\Pi_x = \Pi_{(x_1, x_2, x_3)}$, denotaremos cada termo $((x, y, z), \mu)$ por $xyz\backslash\mu$, com $x \in X_1$, $y \in X_2$, $z \in X_3$ e $\mu \in [0, 1]$.

Exemplo:

Sejam $V_1 = V_2 = V_3 = \{a, b\}$ e $\Pi_x = \Pi_{(x_1, x_2, x_3)}$ como abaixo:

$$\Pi_x = \Pi_{(x_1, x_2, x_3)} = \{aaa\backslash 0.4; \; aab\backslash 0.9; \; aba\backslash 0.3; \; abb\backslash 0.5; \; baa\backslash 0.8; \; bbb\backslash 1\} \text{ e}$$

$$Proj_{(u_1, u_2)} \Pi_x = Sup\{aa\backslash 0.4; \; aa\backslash 0.9; \; ab\backslash 0.3; \; ab\backslash 0.5; \; ba\backslash 0.8; \; bb\backslash 1\} =$$

$$\{aa\backslash 09; \; ab\backslash 0.5; \; ba\backslash 0.8; \; bb\backslash 1\} = \Pi_{(x_1, x_2)}$$

Seja $\Pi_X(u_1, \ldots, u_n) = F$ uma distribuição de possibilidades na variável $X = (x_1, \ldots, x_n)$ e seja $\Pi_{X(s)} = G$ uma distribuição de possibilidades na subvariável $X(s) = (x_{i1}, \ldots, x_{ik})$. A *extensão cilíndrica* de G é o conjunto *fuzzy*, denotado por $\widehat{G}$ em $V_1 \times \ldots \times V_n$, tal que sua projeção $X_{(S)}$ coincide com G, ou seja:

$$f_{\widehat{G}}(u_1, \ldots, u_n) = f_G(u_{i1}, \ldots, u_{ik}), (u_1, \ldots, u_n) \in U_1 \times \ldots \times U_n$$

A extensão cilíndrica faz o caminho inverso da projeção

Regra da Particularização (RPa): Sejam $\Pi_{(x_1, \ldots, x_n)} = F$ uma distribuição de possibilidades na variável $X = (x_1, \ldots, x_n)$ e $\Pi_{X(s)} = G$ uma distribuição de possibilidades na subvariável $X(s) = (x_{i1}, \ldots, x_{ik})$ de X. A particularização de Π_X com relação $\Pi_{X(s)}$ na variável $X = (x_1, \ldots, x_n)$ é definida por:

$$\Pi_{(x_1, \ldots, x_n)}[\Pi_{X(s)} = G] = F \cap \widehat{G}$$

Exemplo:

Considere a distribuição de possibilidades do exemplo anterior, isto é:

$$F = \Pi_{(x_1, x_2, x_3)} = \{aaa\backslash 0.4;\ aab\backslash 0.9;\ aba\backslash 0.3;\ abb\backslash 0.5;\ baa\backslash 0.8;\ bbb\backslash 1\} \text{ e}$$

$$G = \Pi_{(x_2, x_3)} = \{aa\backslash 0.7;\ ba\backslash 0.4;\ bb\backslash 0.9\}$$

Então:

I. A *extensão cilíndrica* de G é:

$$\widehat{G} = \{aaa\backslash 0.7;\ baa\backslash 0.7;\ aba\backslash 0.4;\ bba\backslash 0.4;\ abb\backslash 0.9;\ bbb\backslash 0.9\}$$

II. A *particularização* de F é:

$$\Pi_{(x_1, x_2, x_3)}[G] = F \cap \widehat{G} = \{(aaa\backslash 0.4);\ (aba\backslash 0.3);\ (abb\backslash 0.5);\ (baa\backslash 0.7);\ (bbb\backslash 0.9)\}$$

2.2.5.3 MODIFICAÇÃO, COMPOSIÇÃO E QUALIFICAÇÃO

Essas três regras permitem a análise de proposições *fuzzy* que sejam compostas, modificadas ou qualificadas.

As proposições *fuzzy* compostas são construídas a partir de combinações de proposições *fuzzy* simples usando conectivos lógicos como "e", "ou" e "não". Esses conectivos são tratados de maneira semelhante à lógica clássica, mas os graus de verdade são calculados usando funções específicas.

Exemplo: A temperatura está alta *e* a umidade está baixa.

As proposições *fuzzy* modificadas são proposições que incluem modificadores linguísticos para ajustar o grau de verdade de uma expressão. Esses modificadores são palavras como "muito", "pouco", "extremamente", ou "ligeiramente", que alteram o valor da função de pertinência da proposição.

Exemplo: A temperatura está *muito* alta.

As proposições *fuzzy* qualificadas incluem qualificadores que indicam a confiança ou probabilidade associada à proposição. Esses qualificadores são expressos por termos como "provavelmente", "possivelmente", "quase certo", ou "improvável".

Exemplo: A temperatura está alta com *alta probabilidade.*

A seguir, apresentamos cada uma das regras de tradução com as proposições *fuzzy* mencionadas.

Regras de modificação (RM):

Sejam γ um número real positivo e P um predicado *fuzzy* com domínio V. O conjunto *fuzzy* P^{γ} é definido por:

$$P^{\gamma} = \{(a,\ f_p^{\gamma}(a))/a \in V\}$$

Sejam γ um número real no intervalo [0, 1] e P um predicado *fuzzy* com domínio V. O conjunto *fuzzy* γP é definido por:

$$\gamma P\ = \{(a,\ \gamma f_p(a))/a \in V\}$$

Seja P um predicado *fuzzy* de domínio V. O operador negação é definido por:

$$P'\ = \{(a,\ 1 - f_p(a))/a \in V\}$$

Seja P um predicado *fuzzy* de domínio V. O operador *concentração* é definido por:

$$Con(P)\ =\ P^2$$

Seja P um predicado *fuzzy* de domínio V. O operador *dilatação* é definido por:

$$Dil(P) = P^{1/2}$$

Os operadores de concentração e dilatação, ao serem aplicados a conjuntos *fuzzy*, geram novos conjuntos *fuzzy*. No caso da concentração, o grau de pertinência dos elementos é reduzido, enquanto na dilatação, esse grau é ampliado.

O operador *contraste* é definido para um predicado *fuzzy* P por:

$$Contr(P) = \begin{cases} 2P^2 & , 0 \leq f_p(x) < 0.5 \\ [2(P')^2]' & , 0.5 \leq f_p(x) \leq 1 \end{cases}$$

Exemplo:

Considere a distribuição de possibilidades Π_x que interpreta um predicado *fuzzy* P, com $P = \Pi_x = \{a\backslash 0.5,\ b\backslash 0.6,\ c\backslash 0.3\}$

Assim,

$$P^2 = \{a\backslash 0.25,\ b\backslash 0.36,\ c\backslash 0.09\} \text{ e}$$

$$0.7P = \{a\backslash 0.0,35,\ b\backslash 0.42,\ c\backslash 0.21\}$$

Os modificadores *fuzzy* fornecem descrições aproximadas para os predicados *fuzzy*. Dada uma proposição como "x é P", cuja interpretação é representada por x é $P \rightarrow \Pi_x = P$, é possível gerar uma proposição modificada "x é ηP". Aqui, η é um modificador *fuzzy*, como: pouco, muito, não, mais ou menos, bastante, meio, entre outros. A interpretação da proposição modificada é expressa por:

$$X \text{ é } \eta P \rightarrow P^{\star}, \text{ ou ainda, } f_{\eta p}(x) = f_{P^{\star}}(x)$$

Utilizamos para descrever os modificadores não, muito e mais ou menos:

$$\text{se } \eta = \text{não, então } P^{\star} = P' \text{ , com } f_{\text{não}p}(x) = 1 - f_P(x)$$

$$\text{se } \eta = \text{muito, então } P^{\star} = Con(P) = P^2 \text{ , com } f_{muitop}(x) = [f_P(x)]^2$$

$$\text{se } \eta = \text{mais ou menos, então } P^{\star} = Dil(P) = P^{1/2} \text{ , com } f_{P^{\star}}(x) = [f_P(x)]^{1/2}$$

Os modificadores mencionados são frequentemente encontrados na linguagem oral. No entanto, também é possível criar modificadores artificiais que ajustem (intensifiquem ou reduzam) o grau de pertinência dos elementos no universo V, como:

$$P^{\star} = P^{1.25}, \text{ quando } \eta = \text{mais}$$

$$P^{\star} = P^{0.75}, \text{ quando } \eta = \text{menos}$$

Regras de composição (RC)

As regras de composição são da forma:

$$R = P * Q$$

onde P, Q são proposições *fuzzy* e $*$ denota uma *operação de composição*, bem como a conjunção, a disjunção, a condicional, etc.

Sejam P: x é $E \rightarrow \Pi_{(x_1, \ldots, x_n)} = E$ e Q: y é $G \rightarrow \Pi_{(y_1, \ldots, y_m)} = G$. Podemos definir diversas operações de composição utilizando a álgebra dos conjuntos *fuzzy* acrescida das operações de projeção e particularização.

Conjunção: quando x é $E \rightarrow \Pi_{(x_1, \ldots, x_n)} = E$ e y é $G \rightarrow \Pi_{(x_1, \ldots, x_n)} = G$ (os domínios coincidem), então a conjunção é $E \cap G$.

Conjunção estendida: x é $E \rightarrow \Pi_{(x_1, \ldots, x_n)} = E$ e y é $G \rightarrow \Pi_{(y_1, \ldots, y_n)} = G$ (os domínios são distintos), então a conjunção $\Pi_{(x_1, \ldots, x_n, y_1, \ldots, y_n)} = \widehat{E} \cap \widehat{G}$.

Disjunção: quando x é $E \rightarrow \Pi_{(x_1, \ldots, x_n)} = E$ e y é $G \rightarrow \Pi_{(x_1, \ldots, x_n)} = G$ (os domínios coincidem), então a disjunção é $E \cup G$.

Disjunção estendida: x é $E \rightarrow \Pi_{(x_1, \ldots, x_n)} = E$ e y é $G \rightarrow \Pi_{(y_1, \ldots, y_n)} = G$ (os domínios são distintos), então a disjunção é $\Pi_{(x_1, \ldots, x_n, y_1, \ldots, y_n)} = \widehat{E} \cup \widehat{G}$.

Condicional *fuzzy*: Se x é E, então y é G, a condicional será $\Pi_{(x_1, \ldots, x_n, y_1, \ldots, y_n)} = E \times G \cup E' \times V$.

Uma importante regra de dedução *fuzzy* é dada pela composição de relações.

Sejam $P \rightarrow \Pi_{(x, y)} = E$ e $Q \rightarrow \Pi_{(y, z)} = G$

Inferência composicional:

$$\frac{P \rightarrow \Pi_{(x,\,y)} = E \qquad Q \rightarrow \Pi_{(y,\,z)} = G}{R \rightarrow \Pi_{(x,\,z)} = E \circ G}$$

Regras de qualificação (RQ)

As regras de qualificação visam traduzir proposições do tipo "x é P é ψ", onde ψ pode representar um valor de verdade, probabilidade ou possibilidade. Em qualquer caso; ψ é uma função mapeando [0, 1] em [0, 1].

Considere a qualificação de verdade de uma proposição P:

P(x é F) é ψ, onde F é um termo predicado *fuzzy* e ψ é um valor de verdade linguístico. Essa proposição pode ser dada por outra proposição *fuzzy*, da forma:

Q: x é G, tal que G é definida por $f_G = \psi \circ f_F$, onde f_F é a função de pertinência associada a F.

O próximo passo é explorar como os princípios *fuzzy* podem ser aplicados na prática para resolver problemas reais. No capítulo a seguir, apresentaremos os sistemas baseados em regras *fuzzy*, que constituem uma das implementações mais poderosas e amplamente utilizadas da lógica *fuzzy*, permitindo a criação de sistemas inteligentes capazes de tomar decisões e realizar inferências de forma similar ao raciocínio humano através de regras linguísticas.

2.3 Os sistemas baseados em regras *fuzzy* (SBRF)

Nas abordagens clássica e moderna de controle, o ponto de partida para a implementação do controle de um processo é a obtenção de um modelo matemático capaz de representá-lo com precisão (Dorf e Bishop, 2011).

De modo geral, a teoria de controle consiste no conjunto de métodos e princípios que permitem analisar e projetar sistemas de controle, com o objetivo de regular o comportamento de processos dinâmicos e garantir que apresentem desempenho desejado, como estabilidade e precisão. Em particular, a teoria clássica de controle baseia-se fortemente em modelos

matemáticos bem definidos, geralmente lineares e invariantes no tempo, utilizando ferramentas como funções de transferência e análises no domínio do tempo e da frequência. Esse procedimento exige um conhecimento profundo do funcionamento do processo em questão, uma exigência que nem sempre pode ser atendida, especialmente quando se trata de sistemas complexos. Embora as teorias de controle tenham sido desenvolvidas para uma ampla gama de aplicações, elas pressupõem que o processo seja claramente definido e quantificável, o que pode limitar sua aplicabilidade em situações reais mais complexas (Dorf e Bishop, 2011).

Diversas técnicas sofisticadas foram criadas e aplicadas com êxito a problemas bem estruturados, como o controle linear multivariável (Doyle e Skin, 1981), a estimação de estados em sistemas com medições ruidosas (Anderson e Moore, 1979), o controle ótimo (Sage e White, 1977), o tratamento de sistemas lineares estocásticos (Bertsekas, 1976), além de certas classes de problemas não lineares determinísticos (Holtzman, 1970).

Contudo, essas metodologias enfrentam limitações significativas quando se deparam com sistemas reais cuja modelagem matemática é impraticável ou inviável. Em muitos casos, informações críticas estão disponíveis apenas em termos qualitativos e subjetivos, e os critérios de desempenho são expressos por meio de linguagem natural, em vez de fórmulas precisas. Esse cenário, marcado por incertezas e ausência de exatidão, compromete a aplicação das teorias tradicionais, exigindo abordagens mais flexíveis e tolerantes à imprecisão (Gomide e Gudwin, 1994).

Para Gomide e Gudwin (1994), a modelagem e o controle *fuzzy* abordam a relação entre as variáveis de entrada e saída, integrando múltiplos parâmetros tanto do processo quanto do sistema de controle. Essa abordagem possibilita o tratamento de processos complexos, resultando em sistemas de controle capazes de oferecer maior precisão, além de estabilidade e robustez no desempenho. Uma das principais vantagens dos sistemas *fuzzy* é sua simplicidade de implementação, o que pode reduzir significativamente a complexidade do projeto. Dessa forma, problemas que antes eram considerados intratáveis podem se tornar solucionáveis por meio dessa técnica.

Mc Neil e Thro (1994) promovem a relação de características dos sistemas onde a aplicação da lógica *fuzzy* é necessária ou benéfica, sendo elas: sistemas complexos que são difíceis ou impossíveis de modelar; sistemas controlados por especialistas; sistemas com

entradas e saídas complexas e contínuas; sistemas que se utilizam da observação humana como entradas ou como base para regras e; sistemas que são naturalmente "vagos", como os que envolvem ciências sociais e comportamentais, cuja descrição é extremamente complexa.

Para Castillo e Melin (2008), sistemas especialistas que utilizam lógica *fuzzy* vem sendo aplicados com sucesso nos problemas de decisão, controle diagnóstico e classificação, pois esses sistemas possuem a capacidade de gerenciar o raciocínio complexo presente nas áreas de aplicação. Os sistemas especialistas são compostos por uma *base de dados, ou base de conhecimento*, sendo um banco de informações retiradas de um domínio em estudo por especialistas, onde é representado o conhecimento possuído por eles sobre o domínio do problema, contendo os dados e as formas de condução para identificação e solução de um determinado problema e um *mecanismo de inferência (raciocínio)*, que atua como um processador e trabalha com as informações fornecidas pela base de dados em função dos dados do problema abordado. Além disso, utilizam-se regras (*fuzzy*) através do mecanismo de inferência para lidar com a base de dados.

A utilização da lógica *fuzzy* na modelagem e em ações de tomada de decisão se viu crescente com o trabalho realizado por Mamdani e Assinlian (1975), onde foi proposto um sistema *fuzzy* controlador para sintetizar o processo de tomada de decisão de um operador industrial habilitado, adotando um processo de decisão baseado em regras *fuzzy*. Nas regras, tanto o antecedente quanto o consequente são compostos de proposições que definem valores de variáveis como termos linguísticos. Esses sistemas, são conhecidos como Sistemas *Fuzzy* Baseados em Regras (SFBR) (Klir; Yuan, 1995).

De maneira simplificada, podemos traduzir um sistema *fuzzy* como qualquer sistema que utiliza a teoria de conjuntos *fuzzy* para representar suas variáveis ou suas interações. As variáveis de um sistema *fuzzy* são chamadas de *variáveis linguísticas*, pois seus valores são sentenças expressas na linguagem natural, como temperatura, altura, velocidade, distância, entre outros. Essas variáveis são definidas por *termos linguísticos*, que são rótulos ou valores de uma variável linguística aos quais estão associados conjuntos *fuzzy*, como frio, alto, rápido, longe, entre outros (Lima, 2015).

O *universo de discurso* de uma variável em um sistema *fuzzy* é o conjunto de todos os valores precisos (*crisp*) que uma variável pode assumir, definindo um *conjunto universo*, e sobre o qual os conjuntos *fuzzy* são definidos. O processo de particionar o *universo de discurso* de uma variável em termos linguísticos irá definir uma *partição fuzzy*.

Para Barros e Bassanezi (2006), os principais tipos de SBRF são os do tipo Mamdani e Takagi-Sugeno. A principal característica do método Mamdani é que a condição e a ação são expressas em termos linguísticos, já nos sistemas Takagi-Sugeno, apenas a condição é expressa em termos linguísticos e a ação, por termos funcionais. Para este projeto, optamos pelo método Mamdani, introduzido por Mamdani e Assilian em 1975, que se destaca por ser um método mais intuitivo.

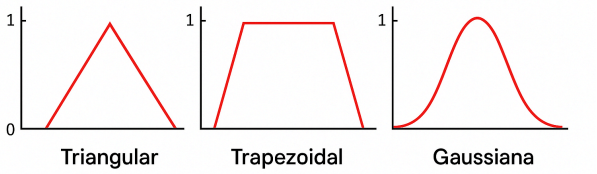
O princípio central do controle *fuzzy* consiste em definir as ações de controle com base no conhecimento de especialistas, em vez de depender exclusivamente da modelagem matemática do processo. Essa abordagem contrasta com os métodos tradicionais de controle de processos, que partem da modelagem precisa do sistema para, então, derivar as leis de controle em função de seu estado. O interesse por essa metodologia alternativa surgiu especialmente em situações em que havia acesso ao conhecimento prático de operadores ou projetistas experientes, mas em que a construção de modelos matemáticos era excessivamente complexa ou dispendiosa. (Mamdani e Assilian, 1975)

2.3.1 Etapas do sistema *fuzzy* de inferência Mamdani

O estilo de inferência Mamdani foi desenvolvido em 1975 pelo professor Ebrahim Mamdani, da Universidade de Londres (Reino Unido), no contexto do avanço dos sistemas *fuzzy*. Seu objetivo era representar experiências da vida real por meio de regras baseadas em conjuntos *fuzzy*. Para a construção desse sistema, foi estabelecido um processo de raciocínio dividido em quatro etapas: (I) fuzzyficação, (II) Base de regras, (III) inferência e (IV) defuzificação.

A etapa de *fuzzificação* (I) é responsável por converter os valores de entrada, mesmo quando são precisos (*crisp*), em valores *fuzzy* por meio das funções de pertinência definidas na fase de construção do sistema. Essas funções podem assumir diferentes formas, como triangular, trapezoidal, gaussiana, entre outras.

Figura 3: Funções de pertinência triangular, trapezoidal e gaussiana



Fonte: Autora

Nessa fase, determina-se o grau de pertinência com que cada valor de entrada pertence a um ou mais conjuntos *fuzzy*. Para isso, cada variável de entrada é previamente delimitada dentro de um universo de discurso específico e associada a um grau de pertinência em relação a diferentes categorias linguísticas, com base no conhecimento fornecido por especialistas.

Assim, para obter o grau de pertinência de uma entrada *crisp* (precisa), basta localizar seu valor correspondente dentro da base de conhecimento do sistema *fuzzy*. Como parte deste processo, será analisada qual tipo de função de pertinência melhor atende à proposta do sistema, considerando as necessidades do programador e os objetivos da aplicação.

Na etapa de regras (II) as informações do fenômeno em estudo são utilizadas. Para cada informação definida por termos linguísticos da variável de entrada é definida uma regra. Desse modo, quanto mais termos linguísticos, mais informações são incorporadas na modelagem.

Para Barros e Bassanezi (2006), a base de regras cumpre o papel de "traduzir" matematicamente as informações que formam a base de conhecimento do sistema *fuzzy*. Quanto mais precisas as informações, menos *fuzzy* será a relação *fuzzy* que representa a base de conhecimento.

Em um sistema *fuzzy*, existe um conjunto de regras para descrever a ação necessária e essas regras, representam as relações entre suas entradas e saídas, formando a base de conhecimento.

Barros e Bassanezi (2006) apresentam a estrutura do método Mamdani, que se baseia na regra de composição de inferência do tipo max-min, conforme descrito a seguir:

Em cada regra R_i da base de regras *fuzzy*, a condicional "se a é A_i então b é B_i"é modelada pela aplicação do mínimo. Cada regra irá satisfazer a seguinte estrutura:

SE *a* está em A_i ENTÃO *b* está em B_i

em que A_i e B_i são conjuntos *fuzzy* que representam termos linguísticos das variáveis de entrada e saída, respectivamente.

O termo *a*, que representa a entrada do sistema, está em Ai, e significa que $\mu A_i(a) \in [0,1]$. Desse modo, A_i e B_i são escritos como produtos cartesianos de conjuntos *fuzzy*.

$$A_i = A_{i1} \times A_{i2} \times A_{i3} \times ... \times A_{im}$$

$$B_i = B_{i1} \times B_{i2} \times B_{i3} \times ... \times B_{in}.$$

Desse modo, cada par de conjuntos *fuzzy* (A_{ij}, B_{ik}) representa um termo linguístico para a j-ésima variável de entrada e a k- ésima variável de saída. Portanto, dizer, que *a* está em A_i é dizer que:

$$\mu A_i\,(a) = \min\{\mu A_{i1}\,,\mu A_{i2}\,,...,\mu A_{im}\} \in [0,1].$$

Caracterizamos a relação existente entre as variáveis linguísticas pelo operador mínimo (min), ou seja, cada regra é considerada uma relação *fuzzy* R_i e obtém-se o grau de pertinência para cada par (*a*,*b*) por:

$$\mu R_i\,(a,b) = \min\{\mu A_i\,(a),\, \mu B_i\,(b)\}.$$

Caracterizamos a relação estabelecida entre cada regra pelo operador máximo (max), ou seja, uma relação *fuzzy* R que representa o modelo determinado pela base de regras, obtida pela união (máximo) de cada regra individual, onde para cada par (a,b) temos:

$$\mu R(a,b) = \max_{1\le i\le n}\{\mu A_i(a)\wedge\mu B_i(b)\},$$

em que o $\wedge$ é o operador min. Portanto, pelo método Mamdani, a função de pertinência de B é representada por:

$$\mu_B(b) = \max_{1\le i\le n}\{\max_a\{\mu A(a) \wedge \mu A_i(a)\} \wedge \mu B_i(b)\},$$

em que *a* é um conjunto clássico unitário, com $\mu A(a) = 1$ e $\mu A(a) \le 1$, obtendo:

$$\mu_B(b) = \max_{1\le i\le n}\{\mu Ai(a) \wedge \mu B_i(b)\}.$$

O exemplo a seguir, apresentado por Barros e Bassanezi (2006), ilustra o funcionamento do método Mamdani para um sistema *fuzzy* com duas entradas e uma saída.

Exemplo: considere um controlador *fuzzy* com duas entradas e uma saída, cuja base de regras é dada abaixo:

R_1: Se x_1 é A_{11} e x_2 é A_{12} então u é B_1

ou

R_2: Se x_1 é A_{21} e x_2 é A_{22} então u é B_2

Logo, para cada terna (x_1, x_2, u) temos

$$\mu R(t) = \{\mu A_{11}\,(x_1) \wedge \mu A_{12}\,(x_2) \wedge \mu B_1\,(u)\} \vee \{\mu A_{21}\,(x_1) \wedge \mu A_{22}\,(x_2) \wedge \mu B_2\,(u)\} = \max$$
$$\{\mu A_{11}\,(x_1) \wedge \mu A_{12}\,(x_2) \wedge \mu B_1\,(u),\, \mu A_{21}\,(x_1) \wedge \mu A_{22}\,(x_2) \wedge \mu B_2\,(u)\}$$

que representa a relação *fuzzy* obtida da base de regras pelo método Mamdani.

Desse modo, para um conjunto *fuzzy* de entrada A = A_1 x A_2, com A_1 e A_2 dois números *fuzzy*, o conjunto de saída, que representa o controle a ser adotado para A, é dado por B = $R \circ A$, que possui a seguinte função de pertinência:

$$\mu B(u) = (R \circ A)(u) = sup_x\{\mu R(x, u) \wedge \mu A(x)\}$$

Como A = A_1 x A_2, então $\mu A(x_1, x_2) = \mu A(x_1) \wedge \mu A(x_2)$

Dessa forma:

$$\mu B(u) = sup_x\{\mu R(x,u) \wedge \mu A(x)\}$$

$$= sup_{(x_1, x_2)}\{\mu R(x_1, x_2, u) \wedge [\mu A_1(x_1) \wedge \mu A_2(x_2)]\}$$

$$= sup_{(x_1, x_2)}\{[(\mu A_{11}(x_1) \wedge \mu A_{12}(x_2) \wedge \mu B_1(u)) \vee$$

$$(\mu A_{21}(x_1) \wedge \mu A_{22}(x_2) \wedge \mu B_2(u))] \wedge [\mu A_1(x_1) \wedge \mu A_2(x_2)]\}$$

$$= sup\{[\mu A_1(x_1) \wedge \mu A_{11}(x_1)] \wedge [\mu A_2(x_2) \wedge \mu A_{12}(x_2)] \wedge \mu B_1(u)\} \vee$$

$$sup\{[\mu A_1(x_1) \wedge \mu A_{21}(x_1)] \wedge [\mu A_2(x_2) \wedge \mu A_{22}(x_2)] \wedge \mu B_2(u)\}$$

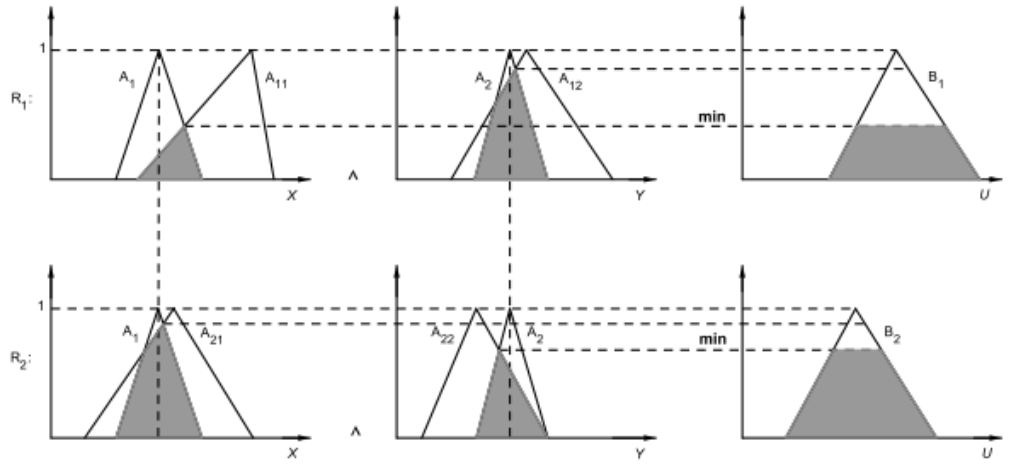
$$= \mu B_{R_1}(u) \vee \mu B_{R_2}(u) .$$

em que B_{R_1} e B_{R_2} são as saídas parciais dadas pelas regras R_1 e R_2.

Observa-se, contudo, que a saída do método de Mamdani é obtida pela união das saídas parciais geradas por cada regra do sistema. Para determinar cada uma dessas saídas parciais, realiza-se inicialmente a interseção entre os valores de entrada e os antecedentes das respectivas regras. Em seguida, é feito o produto cartesiano, considerando universos distintos, entre essas interseções e os consequentes das regras. A projeção desse produto cartesiano sobre o universo de saída U constitui a saída parcial correspondente ao conjunto *fuzzy* de entrada A.

Representa-se graficamente:

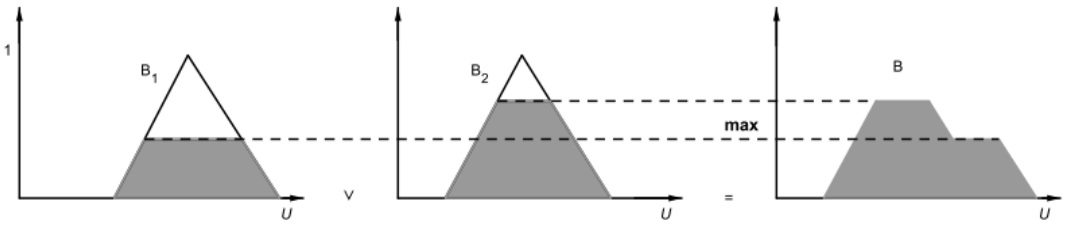
Figura 4: Saídas parciais no método Mamdani



Fonte: Barros e Bassanezi (2006)

A saída geral é determinada pela união das saídas parciais, conforme abaixo:

Figura 5: Saída final do controlador *fuzzy* de Mamdani



Fonte: Barros e Bassanezi (2006)

Observa-se que a última imagem ilustra a função de pertinência do controle B que foi obtido pelo conectivo ∨, que é a t-conorma do maximo.

Na etapa de inferência (III) é onde cada proposição *fuzzy* é traduzida matematicamente através de técnicas da lógica *fuzzy*. É onde se define quais t-normas, t-conormas e regras de inferência (que podem ser implicações *fuzzy*) serão utilizadas para se obter a relação *fuzzy* que modela a base de regras. Esta etapa irá fornecer a saída *fuzzy*, a partir de cada entrada.

Por fim, na etapa de defuzzificação (IV), o valor *fuzzy* obtido como resposta, passa pelo processo de defuzzificação dos valores *fuzzy* obtidos como saída e então, obtemos novamente uma saída precisa (*crisp*). O valor de saída é calculado através do somatório do produto entre o conjunto de entrada pelo seu respectivo grau de pertinência. Em seguida, divide-se esse resultado pelo somatório da pertinência para cada entrada, resultando na ação de controle. Os conjuntos *fuzzy*, são bastante utilizados para construção de modelos computacionais de análise de dados que lidam com informações desse tipo. Lopes, Jafelice e

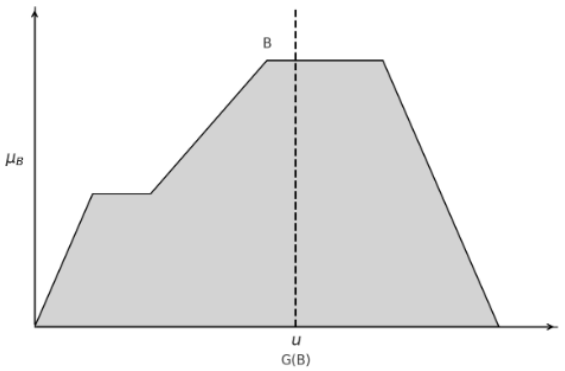
Barros (2005) destacam que um dos principais métodos de defuzzificação é o método centróide, ou centro de massa, que para um conjunto no domínio discreto e no domínio contínuo as respectivas expressões são:

$$m(\mathrm{B}) = \frac{\sum_{i=1}^{n} b.\mu B(b)}{\sum_{i=1}^{n} \mu B(b)} \quad \text{e} \quad m(\mathrm{B}) = \frac{\int_{R} b.\mu B(b)db}{\int_{R} \mu B(b)db}$$

Sendo o caso discreto ou contínuo, o valor m(B) corresponde à projeção do centro de massa da figura definida pelo conjunto de regras sobre o eixo da variável de controle. Dessa maneira, o controlador *fuzzy* pode ser definido como uma função $f: R^n \rightarrow R^m$. Portanto, dado um número real de entrada, existe um único elemento correspondente na saída.

Mostramos a seguir o gráfico do defuzzificador G(B)

Figura 6: Defuzzificador centróide



Fonte: Autora

Além do defuzzificador centróide, Barros e Bassanezi (2006) também apresentam o método de defuzzificação denominado centro dos máximos (C(B)). Esse método é considerado mais radical, pois são levados em conta apenas as regiões de maior possibilidade entre os possíveis valores da variável que modela o conceito *fuzzy* em questão. Temos:

$$C(B) = \frac{i+s}{2}$$

onde

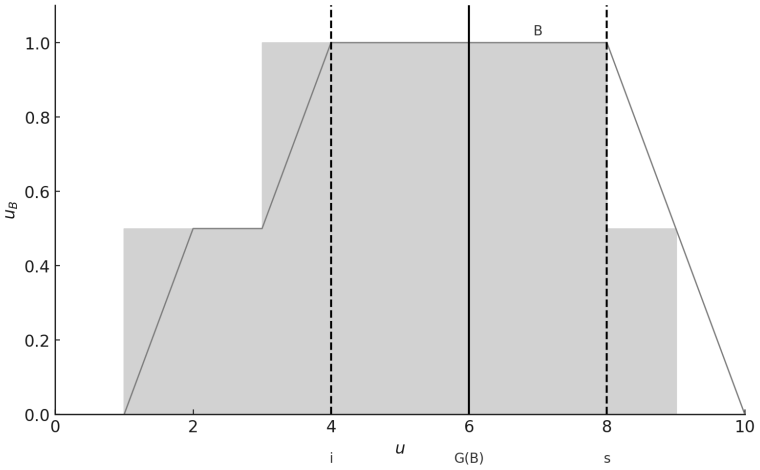
$$i = inf\{u \in \Re: \mu B(u) = max_u \mu B(u)\}$$

e

$$s = sup\{u \in \Re: \mu B(u) = max_u \mu B(u)\}$$

Representa-se o centro dos máximos a seguir:

Figura 7: Centro de máximo C(B)



Fonte: Autora

Para domínios discretos, é comum utilizar a média dos valores com pertinência máxima como método de defuzzificação, denominado média dos máximos, apresentado a seguir:

$$M(B) = \frac{\Sigma u_i}{n}$$

onde n é dado e u_i, com $1 \leq i \leq n$, são os elementos de maior pertinência ao conjunto *fuzzy* B.

O método de inferência *fuzzy* de Mamdani é amplamente utilizado devido à sua interpretabilidade e aproximação com o raciocínio humano, já que utiliza regras linguísticas e funções de pertinência intuitivas. Entre suas principais vantagens estão a simplicidade na construção do sistema e a facilidade de aplicação em problemas que envolvem conhecimento qualitativo. Embora existam outros métodos de inferência, como o modelo de Takagi-Sugeno-Kang (TSK), que se destaca por sua capacidade de representar sistemas com maior precisão matemática e facilitar otimizações, neste trabalho optamos pelo método de Mamdani por sua adequação ao contexto proposto e sua clareza na representação das regras *fuzzy*.

2.3.2 Parâmetros e controladores *fuzzy*

Durante a modelagem *fuzzy*, a definição de determinados parâmetros é essencial. Esses parâmetros podem ser determinados com base na experiência do projetista ou obtidos por meio de experimentação. Em um sistema, alguns desses parâmetros permanecem constantes sob condições normais de operação, enquanto outros precisam ser ajustados periodicamente.

Conforme apresentado por Gomide e Gudwin (1994), os parâmetros que não se alteram são chamados de *parâmetros estruturais*, enquanto aqueles que exigem ajustes frequentes são denominados *parâmetros de sintonização*.

Parâmetros estruturais incluem:

- Número de variáveis de entrada e saída;
- Operações aplicadas aos dados de entrada;
- Variáveis linguísticas;
- Funções de pertinência parametrizadas;
- Intervalos de discretização e normalização;
- Estrutura da base de regras;
- Conjunto básico de regras.

Parâmetros de sintonização abrangem:

- Universo de discurso das variáveis;
- Parâmetros das funções de pertinência, como altura, largura e conjunto suporte;
- Ganhos e *offsets* das variáveis de entrada e saída.

Algumas propriedades da base de regras precisam ser avaliadas cuidadosamente, entre elas: completude, consistência, interação e robustez. Esta última relacionada à sensibilidade do sistema de controle frente a comportamentos anômalos não modelados. Além disso, é necessário tomar diversas decisões importantes durante o projeto do sistema *fuzzy*, como:

- A definição dos operadores utilizados para realizar o *matching* dos conectivos lógicos "e" (*and*) e da implicação;
- A escolha do operador responsável pela agregação das diferentes regras;
- A seleção do método de *defuzzificação* mais adequado ao contexto da aplicação.

A etapa seguinte no desenvolvimento de um controlador *fuzzy* é a sintonização, considerada uma das fases mais desafiadoras e trabalhosas do projeto. Isso se deve à grande quantidade de parâmetros envolvidos, o que confere ao sistema uma notável flexibilidade, mas também impõe ao projetista um esforço significativo para alcançar o melhor desempenho possível.

Embora existam mecanismos automáticos de adaptação e aprendizado capazes de ajustar alguns desses parâmetros, esses métodos ainda não estão completamente consolidados dentro da teoria de controle fuzzy. Por esse motivo, a maior parte do trabalho de ajuste fino e treinamento dos parâmetros ainda recai sobre o projetista. Esse processo de sintonização geralmente envolve técnicas de busca, o que a caracteriza como uma atividade típica da área de Inteligência Artificial (IA). Na teoria clássica de controle, essa questão também tem recebido atenção significativa, dada sua importância para a eficácia e estabilidade dos sistemas controlados.

Até o momento, apresentamos os fundamentos teóricos que sustentam esta pesquisa: os conjuntos *fuzzy*, a lógica *fuzzy* e os sistemas baseados em regras *fuzzy*. Com essa base conceitual estabelecida, podemos agora direcionar o foco para o contexto prático de aplicação desses conhecimentos. No capítulo a seguir, será apresentado o Programa de Inclusão da UNESP, com a exploração de suas características e funcionamento. Esse passo nos conduz aos objetivos centrais do estudo: analisar e comparar o desempenho acadêmico de estudantes ingressantes pelo sistema universal e pelo sistema de cotas, através de sistemas *fuzzy*.

3. O PROGRAMA DE INCLUSÃO DA UNIVERSIDADE ESTADUAL PAULISTA (UNESP)

De acordo com Galhardo e Vasconcelos (2016), ao tratarmos de investimentos públicos voltados para ações afirmativas governamentais dirigidas ao acesso a universidade, busca-se assegurar a inclusão por meio de cotas e dos programas de permanência estudantil, que possui como um dos principais objetivos proporcionar condições que contribuam para evitar a retenção e a evasão dos estudantes, buscando à formação de profissionais qualificados para o mercado de trabalho.

Galhardo e Vasconcelos (2016) apresentam que embora as universidades paulistas estaduais (UNESP, UNICAMP e USP) há mais de três décadas buscam desenvolver ações com o objetivo de favorecer a inclusão e a permanência estudantil, as propostas de inclusão com o perfil de cotas apenas foram concretizadas a partir de 2013, quando o governo do Estado de São Paulo, através do Programa Paulista de Inclusão Social no Ensino Superior (PPISES), solicitou que essas universidades implementassem ações afirmativas para promover a inclusão de estudantes em vulnerabilidade socioeconômica em seus cursos de graduação, a partir do acolhimento, em 50% de suas vagas, de alunos vindos de escolas públicas, onde 35% deveriam ser reservados para pretos, pardos e indígenas. Após estudos realizados por comissões compostas por docentes das três universidades, foi elaborada a proposta do Programa de Inclusão com Mérito no Ensino Superior Público Paulista (PIMESP), em que estava previsto que essas universidades tivessem ao menos 50% das matrículas em cada turno, proveniente de alunos que cursaram integralmente o Ensino Médio em escolas públicas e dentre os 50%, o percentual de preto, pardos e indígenas deveriam ser no mínimo de 35%, com o objetivo de atingir essas metas até o final de 2014.

O conselho universitário da Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), em busca de atender à solicitação do governo do estado, após diversos debates envolvendo o PPISES e o PIMESP, em agosto de 2013 aprovou uma proposta de inclusão diferente do PIMESP e semelhante à USP e UNICAMP. O projeto de inclusão aprovado mantém que para cada curso e turno dos cursos de graduação da universidade devem ser preenchidas 50% das vagas por estudantes que realizaram o Ensino Médio integralmente em escolas públicas, e que destas, 35% devem ser reservadas para pretos, pardos e indígenas. Contudo, prevê um prazo de 5 anos para a universidade atingir a meta de inclusão de 50%, sendo 15% em 2014, 25% em 2015, 35% em 2016, 45% em 2017 e 50% em 2018. Além

disso, a metodologia adotada para esta inclusão foi o processo classificatório do vestibular, através do Sistema de Reserva para a Escola Pública (SRVEBP), com o aproveitamento dos candidatos até o limite de vagas para cada curso e turno (Galhardo e Vasconcelos, 2016).

Compreende-se que existe uma resistência em relação à implantação de programas de inclusão nas universidades públicas e com isso, busca-se desenvolver pesquisas que apontem dados expressivos em relação ao desempenho e a permanência dos ingressantes pelo sistema de cotas. Resultados de análises a partir da comparação de grupos focais compostos por estudantes cotistas e não cotistas em relação ao coeficiente de rendimento (CR), as taxas de graduação e a evasão, demonstram que para os cotistas existe uma tendência a atribuir um alto valor aos cursos que ingressam e por isso persistem em maior proporção na universidade (Mendes Júnior, 2014). Além disso, estuda-se o perfil socioeconômico do estudante ingressante de escolas públicas comparado ao desempenho nas disciplinas da graduação, no vestibular e até mesmo na escolha de carreiras de maior ou menor prestígio social (Souza, 2012; Zago, 2006).

Em uma pesquisa sobre ações afirmativas na Universidade Estadual do Rio de Janeiro (UERJ), Mendes Júnior e Mello e Souza (2012) observaram que os candidatos cotistas obtêm, em média, pontuações inferiores no vestibular em comparação aos não cotistas. Essa discrepância inicial no desempenho pode influenciar a trajetória acadêmica dos cotistas, impactando sua progressão ao longo do curso. Além disso, a maioria desses estudantes provém de famílias com menor renda, o que aumenta a probabilidade de evasão no ensino superior. Esse risco está associado tanto à necessidade de conciliar os estudos com o trabalho para garantir a subsistência quanto à dificuldade de acesso a materiais acadêmicos essenciais, como livros, cópias de textos e recursos tecnológicos.

Mendes Júnior (2014) apresenta os resultados de uma análise com alunos da UERJ, focada na progressão acadêmica dos estudantes que ingressaram em 2005. O estudo revelou que os não cotistas apresentam, em média, um coeficiente de rendimento superior ao dos cotistas. No entanto, ao avaliar diversos indicadores, constatou-se que, apesar das notas médias mais baixas, os estudantes cotistas estão se formando em uma proporção maior do que os não cotistas. Isso sugere que, embora enfrentem mais desafios ao longo do curso, os cotistas atribuem um valor significativo à formação acadêmica, o que se reflete em uma maior persistência e em taxas de conclusão mais elevadas.

Conforme apontam Galhardo, Vasconcelos, Frei e Rodrigues (2020), o programa de inclusão da UNESP tem impactado significativamente o perfil socioeconômico dos

estudantes. Desde 2010, quando a universidade ampliou suas iniciativas de ações afirmativas, até 2018, registrou-se um aumento no número de ingressantes provenientes de famílias com renda per capita de até 1,5 salário mínimo, que passou de 2053 para 3155 estudantes. Um crescimento também foi observado entre alunos de famílias com renda entre 1,6 e 2,0 salários mínimos, de 1029 para 1461 estudantes. Em contrapartida, a quantidade de ingressantes com renda familiar per capita superior a 2 salários mínimos apresentou uma redução, de 3285 para 2646. Embora o Sistema de Reserva de Vagas para a Educação Básica Pública (SRVEBP) tenha sido instituído apenas em 2014, a tendência de mudança no perfil socioeconômico já era evidente.

Não existem dúvidas que a inclusão definiu um novo perfil no quadro de estudantes da UNESP e consequentemente, aumentou anualmente os alunos solicitantes de auxílios de permanência estudantil. Sabendo da mudança no perfil socioeconômico dos estudantes através do programa de inclusão e com isso, o aumento de demandas para a permanência dos estudantes na universidade, em 2013, na mesma sessão do conselho universitário que aprovou o programa de inclusão, a UNESP criou a Coordenadoria de Permanência Estudantil (COPE).

A COPE tem como missão formular, implementar e avaliar iniciativas que promovam a igualdade de oportunidades para estudantes em situação de vulnerabilidade socioeconômica, além de atuar na redução dos índices de retenção e evasão acadêmica. Entre suas principais metas destacam-se: a criação de estratégias para acolher os estudantes que ingressaram na UNESP por meio do SRVEBP e a incorporação das diretrizes do Programa Nacional de Assistência Estudantil (PNAES) à universidade. Instituído em 2010, o PNAES tem como foco principal atender estudantes provenientes da rede pública de educação básica, cuja renda familiar per capita não ultrapasse 1,5 salário mínimo.

Galhardo, Vasconcelos, Frei e Rodrigues (2020) apresentam, que em 2017, considerando a experiência da universidade no processo de implantação do projeto de inclusão, e visando atender as demandas dos estudantes, A COPE reformulou as modalidades de assistência aos estudantes e passou a oferecer os auxílios de permanência, a saber: Moradia estudantil, auxílio socioeconômico, auxílio aluguel, subsídio alimentação e auxílio estágio. De 2014 a 2018, todos os estudantes que solicitaram os auxílios e atenderam aos critérios da universidade, foram atendidos. No total, 17638 estudantes foram contemplados.

A partir de 2014, nota-se um crescimento contínuo no número de estudantes que solicitaram auxílios de permanência após ingressarem na UNESP pelo Programa de Inclusão. Neste ano, 2.877 estudantes fizeram solicitações, enquanto em 2018 o número praticamente

dobrou, atingindo 5.659 pedidos. Em 2017, 61% dos auxílios concedidos pela política de permanência estudantil foram destinados a alunos provenientes do Programa de Inclusão, percentual que subiu para 70% em 2018.

Galhardo e Vasconcelos (2016) nos lembram que, quando falamos do ensino superior brasileiro, não é fácil ingressar em um curso em universidades públicas, em especial para um estudante de escola pública que ao longo dos anos passa por grandes índices de desestabilizações e investimentos abaixo do necessário. Devido a essa realidade, a parcela de estudantes de escolas públicas que ingressam nas universidades ainda é pequena comparada aos estudantes que vieram de colégios privados. Sendo assim, existe uma disparidade que torna desigual a competição por uma vaga nas instituições públicas de ensino, pois os estudantes de escolas privadas, no geral, são munidos de muita preparação para o vestibular e possuem condições para cursos pré-vestibulares. Por isso, é importante reconhecer que fatores socioeconômicos influenciam, de forma direta e indireta, tanto o acesso às universidades públicas quanto o desenvolvimento educacional dos estudantes ao longo de sua formação.

Ao analisar os retornos do ingresso no ensino superior no contexto brasileiro, Neri (2008) conclui que os salários médios de quem possui o ensino superior são mais que o dobro daqueles que possuem apenas ensino médio, enfatizando a relevância de um curso superior na vida da população. Desse modo, quando o governo investe em universidades públicas de alta qualidade, o faz por duas razões. Primeiro, para que os egressos, através do conhecimento adquirido durante o curso, obtenham maior produtividade ao ingressar no mercado de trabalho. Segundo, pois a educação é reconhecida como uma fonte geradora de externalidades positivas em uma nação.

No entanto, para que esses objetivos de política educacional sejam atingidos plenamente, é necessário que os recém-matriculados na universidade sejam capazes de progredir nos períodos até a graduação. Com isso, garantir a permanência dos alunos até a formatura torna-se uma condição necessária para que a missão das instituições de ensino superior seja cumprida integralmente.

Pesquisas têm buscado compreender as resistências, impactos e consequências da implantação de programas de inclusão voltados para estudantes de escolas públicas em universidades públicas, considerando as interações entre fatores socioeconômicos, culturais e políticas afirmativas. Estudos quantitativos e qualitativos conduzidos em instituições que adotaram sistemas de cotas para alunos da rede pública têm apresentado dados relevantes, abordando questões como preconceito, reparação de direitos, desempenho acadêmico,

permanência, retenção e evasão dos estudantes beneficiados por esses programas (Menin et al., 2008; Júnior, 2014).

Desse modo, nesta pesquisa, realizamos uma análise do desempenho acadêmico dos estudantes ingressantes pelo programa de inclusão da UNESP, egressos de escolas públicas e autodeclarados pretos, pardos e indígenas, e dos estudantes provenientes do sistema universal, a fim de inferir, através da construção de um Sistema Baseado em Regras *Fuzzy* (SBRF), se existem diferenças relevantes entre o desempenho acadêmico desses estudantes. O processo de coleta e análise dos dados será relatado no próximo capítulo.

4. COLETA E ANÁLISE DOS DADOS

Inicialmente, esta pesquisa foi concebida com o objetivo de realizar uma análise abrangente do desempenho acadêmico e trajetória de estudantes de graduação em toda a Universidade Estadual Paulista (UNESP), incluindo todos os campus. Para tanto, foi planejada a utilização de dados institucionais, que permitiriam uma investigação ampla sobre os fatores que influenciam o rendimento acadêmico dos discentes.

O projeto original previa a coleta de dados junto aos setores administrativos competentes da UNESP, especificamente através da Pró-Reitoria de Graduação (PROGRAD). Esta abordagem foi fundamentada na premissa de que os dados administrativos institucionais ofereceriam maior confiabilidade, completude e padronização, características essenciais para análises de grande escala.

O conjunto de variáveis inicialmente definido para a pesquisa contempla múltiplas dimensões da experiência acadêmica dos estudantes, organizadas nas seguintes categorias:

Dados de identificação e ingresso:

- Ano de ingresso no curso de graduação
- Tipo de ingresso (Vunesp, ENEM, transferência externa)
- Informações sobre cotas (ingresso por ações afirmativas: escola pública, autodeclaração étnico-racial)
- Sexo/gênero
- Classificação e nota no vestibular

Indicadores de desempenho acadêmico:

- Coeficiente de rendimento (CR) individual e CR da turma
- Frequência média
- Histórico de reprovações ou pendências em disciplinas

Atividades complementares e oportunidades acadêmicas:

- Participação na pós-graduação
- Envolvimento em projetos (iniciação científica, extensão, iniciação à docência)
- Participação em programas de bolsa (PIBIC, PIBID, FAPESP, entre outros)

- Recebimento de auxílios estudantis (auxílio permanência)
- Realização de intercâmbio acadêmico

Esta estrutura de variáveis foi proposta para permitir análises multidimensionais que contemplassem não apenas o desempenho acadêmico, mas também os fatores socioeconômicos, as oportunidades de desenvolvimento acadêmico e as políticas de permanência estudantil.

Em janeiro de 2025, foi realizada uma solicitação formal à Pró-Reitoria de Graduação (PROGRAD) da UNESP para acesso ao banco de dados administrativos necessários à realização da pesquisa. O pedido foi acompanhado do resumo do projeto de pesquisa com justificativa acadêmica e científica para o acesso aos dados. Além disso, foi elaborado um Plano de Gestão de Dados (PGD) conforme as diretrizes institucionais.

O Plano de Gestão de Dados foi desenvolvido em conformidade com as normas da Lei Geral de Proteção de Dados (LGPD) e as diretrizes éticas para pesquisa envolvendo seres humanos. O documento especificava os procedimentos de coleta, armazenamento, tratamento, análise e descarte dos dados, bem como as medidas de segurança da informação a serem adotadas durante todo o ciclo de vida da pesquisa.

Inicialmente, a PROGRAD demonstrou receptividade à solicitação, colocando-se à disposição para o fornecimento dos dados mediante o cumprimento de pré-requisitos específicos. Todos os requisitos solicitados foram devidamente atendidos, incluindo a elaboração do Plano de Gestão de Dados e o cumprimento das exigências éticas e legais estabelecidas pela instituição.

Contudo, após o cumprimento dos pré-requisitos, não obtivemos retorno às comunicações subsequentes, permanecendo as solicitações de acompanhamento sem resposta. Esta ausência de resposta institucional, após período prolongado de aguardo, representou um obstáculo significativo ao desenvolvimento da pesquisa conforme inicialmente planejado, exigindo uma reformulação metodológica do projeto.

4.1 Coleta de dados do curso de Licenciatura em Matemática

Diante da impossibilidade de acesso aos dados administrativos institucionais em escala universitária, foi necessário proceder à redefinição da metodologia da pesquisa. Após avaliação das alternativas disponíveis e considerando os recursos e prazos da investigação,

optou-se por delimitar o estudo ao curso de graduação em Licenciatura em Matemática da UNESP, campus de Bauru.

A delimitação foi fundamentada em critérios que assegurem sua viabilidade prática, relevância científica e coerência metodológica. O primeiro critério considerado foi a viabilidade de acesso aos dados, uma vez que as informações sobre os estudantes do curso selecionado estavam disponíveis por meio de fontes acessíveis, permitindo o levantamento de dados consistentes e suficientes para a realização das análises propostas.

Em segundo lugar, levou-se em conta a representatividade acadêmica do curso de graduação em Matemática. Esse curso apresenta características que possibilitam análises relevantes sobre o desempenho acadêmico, especialmente no contexto da investigação das diferenças entre modalidades de ingresso no ensino superior. Além disso, trata-se do curso de formação da pesquisadora, o que contribui para uma compreensão mais aprofundada do contexto analisado e facilita a interpretação dos dados com base em vivências acadêmicas concretas.

Outro aspecto considerado foi o tempo. A opção por um curso específico e por um método de coleta manual foi orientada pela necessidade de concluir essa etapa dentro dos prazos estabelecidos no cronograma da pesquisa. Dessa forma, garantiu-se que a execução do trabalho ocorresse de maneira viável e organizada, sem prejuízos à profundidade da análise.

Quanto à metodologia de coleta manual, adotou-se uma abordagem sistemática de levantamento de dados diretamente relacionados aos estudantes do curso de Matemática. Embora mais trabalhosa, essa estratégia permite um controle mais rigoroso sobre a qualidade, a integridade e a consistência das informações coletadas. Essa escolha metodológica contribui para o aprimoramento da análise e para a confiabilidade dos resultados obtidos.

O processo de coleta manual foi conduzido seguindo os seguintes critérios de sistematização, garantindo a uniformidade e confiabilidade das informações obtidas:

- Identificação e seleção dos casos a serem incluídos na amostra
- Padronização dos procedimentos de coleta
- Organização e estruturação do banco de dados
- Verificação e validação das informações coletadas

A coleta de dados foi realizada nas dependências da Universidade Estadual Paulista (UNESP), campus de Bauru, especificamente no Departamento de Matemática, utilizando-se os equipamentos computacionais disponibilizados pelo professor orientador desta pesquisa. O acesso aos dados foi viabilizado através do Sistema de Graduação (SISGRAD), sendo empregada a permissão institucional do orientador em sua função de vice-coordenador do curso de Licenciatura em Matemática. Destaca-se que a pesquisadora teve acesso exclusivamente aos dados previamente coletados pelo orientador, os quais já se encontravam anonimizados, não sendo disponibilizadas, em nenhuma etapa do estudo, informações que permitissem a identificação dos estudantes.

O processo de coleta foi conduzido mediante a aplicação de filtros específicos, direcionados aos elementos de interesse para os objetivos desta investigação. O recorte temporal estabelecido abrangeu egressos formados no período de 2016 a 2024, totalizando uma amostra de 111 estudantes graduados em Licenciatura em Matemática. Deste universo, identificaram-se 55 estudantes beneficiários do sistema de cotas, sendo 39 cotistas oriundos de escolas públicas e 16 cotistas autodeclarados "preto, pardo ou indígena".

Os parâmetros de filtragem selecionados para a extração dos dados compreenderam: data de desvínculo, estrutura curricular, forma de ingresso, cor, origem do ensino médio, modalidade de cotas, classificação no vestibular, nota final do vestibular, coeficiente de rendimento (CR), aproveitamento de disciplinas, frequência média, média ponderada e forma de ingresso. Após a coleta, os dados foram organizados em planilhas do Microsoft Excel, sendo excluídas as colunas referentes aos nomes e registros acadêmicos (RAs) dos estudantes, procedimento que assegurou a completa anonimização dos dados e o cumprimento dos requisitos éticos da pesquisa científica.

A redefinição da estratégia de pesquisa acarretou modificações no delineamento do estudo. O universo de análise foi modificado, passando de toda a UNESP para o curso de Licenciatura em Matemática do campus de Bauru. Do ponto de vista metodológico, a principal alteração consistiu na transição de uma coleta automatizada de dados administrativos para um processo manual de obtenção das informações. Esta mudança demandou maior investimento temporal na coleta, resultou na redução do número de variáveis analisadas e exigiu adaptação dos instrumentos analíticos ao novo contexto. Contudo, os objetivos centrais da pesquisa foram preservados, mantendo-se os critérios de rigor metodológico estabelecidos inicialmente.

A delimitação do escopo não compromete a relevância científica nem a validade da investigação. Os resultados obtidos, embora circunscritos a um contexto específico, continuam a oferecer contribuições significativas para a compreensão do desempenho acadêmico no ensino superior. As análises serão adequadamente contextualizadas, com os devidos ajustes nas limitações e possibilidades de generalização dos resultados, assegurando a confiabilidade das conclusões e sua aplicabilidade dentro dos parâmetros estabelecidos. Na próxima seção, será apresentada a análise estatística realizada para caracterizar os dados e extrair informações relevantes para a construção do sistema proposto.

4.2 Análise estatística

Com o propósito de compreender a estrutura e as particularidades do conjunto de dados, foi conduzida uma análise estatística descritiva, buscando entender a distribuição e o comportamento das variáveis. A análise foi realizada no Google Colab, utilizando a linguagem de programação Python, e a metodologia de validação baseou-se no cálculo de estatísticas descritivas (média, mediana, variância, desvio padrão e quartis), complementadas por visualizações gráficas para explorar as distribuições dos dados e identificar padrões distintivos entre os grupos. Para avaliar a significância estatística das diferenças observadas, foram aplicados testes de hipóteses não paramétricos, como o Teste U de Mann-Whitney para comparações entre dois grupos e o Teste de Kruskal-Wallis com análises post-hoc para múltiplas comparações, além de análises de correlação. Para a realização da análise estatística, foi estudada a obra de J. S. Gaspar et al., *Introdução à análise de dados em saúde com Python* (2023). A partir desse estudo, foram identificadas as medidas estatísticas mais adequadas para a pesquisa, bem como aprendidos os comandos em Python necessários para sua aplicação. Com base nesse referencial, o conteúdo assimilado foi utilizado para realizar e interpretar as análises estatísticas desenvolvidas neste trabalho.

Para obter uma compreensão detalhada das características e possíveis diferenças no desempenho acadêmico e nos resultados do vestibular, os alunos foram categorizados em três grupos distintos: Não Cotistas, Cotistas oriundos de Ensino Público e Cotistas Raciais (PPI - Pretos, Pardos e Indígenas). Na sequência, foi realizada uma análise estatística descritiva para cada grupo, considerando as seguintes variáveis numéricas de interesse: Coeficiente de Rendimento (CR), Nota Final do Vestibular, Classificação no Vestibular, Média Ponderada, Aproveitamento de Disciplinas e Frequência Média.

4.2.1 Medidas estatísticas descritivas

Como primeira etapa da análise, foram calculadas as seguintes medidas estatísticas descritivas para cada variável em cada um dos três grupos:

- Média (mean): Medida de tendência central que representa o valor médio aritmético dos dados.
- Mínimo (min): Menor valor observado na distribuição.
- Primeiro Quartil (Q1 - 25%): Valor abaixo do qual se encontram 25% das observações da amostra.
- Mediana (Q2 - 50%): Valor central da distribuição ordenada, que divide os dados em duas metades iguais.
- Terceiro Quartil (Q3 - 75%): Valor abaixo do qual se encontram 75% das observações da amostra.
- Máximo (max): Maior valor observado na distribuição.
- Desvio Padrão (std): Medida de dispersão que quantifica a variabilidade dos dados em torno da média. Valores elevados indicam maior heterogeneidade nos dados.

A seguir, são apresentadas as tabelas com os resultados das medidas descritivas calculadas para cada grupo de alunos.

Tabela 5: Medidas não cotistas

	Classificação Vestibular	Nota Final Vestibular	CR	Aproveitamento Disciplinas	Frequência Média	Média Ponderada
count	52	52	56	56	56	56
mean	33,90	49,34	7,69	95%	91%	7,78
min	1,00	35,46	5,32	63%	77%	5,51
25%	15,00	40,99	7,01	93%	89%	7,22
50%	29,50	48,13	7,88	98%	92%	7,99
75%	48,50	55,29	8,52	100%	95%	8,55
max	110,00	70,18	9,35	100%	100%	9,39
std	24,73	9,04	1,01	8%	5%	0,98

Fonte: Autora

Tabela 6: Medidas Ensino Público

	Classificação Vestibular	Nota Final Vestibular	CR	Aproveitamento Disciplinas	Frequência Média	Média Ponderada
count	39	39	39	39	39	39
mean	46,21	45,54	7,56	94%	91%	7,64
min	5,00	35,66	4,89	67%	74%	4,96
25%	27,50	40,74	6,98	93%	88%	7,21
50%	46,00	44,65	7,81	100%	92%	7,91
75%	59,50	48,38	8,44	100%	96%	8,47
max	109,00	65,62	9,41	100%	99%	9,47
std	25,13	7,01	1,21	10%	6%	1,19

Fonte: Autora

Tabela 7: Medidas Pretos, Pardos e Indígenas

	Classificação Vestibular	Nota Final Vestibular	CR	Aproveitamento Disciplinas	Frequência Média	Média Ponderada
count	16	16	16	16	16	16
mean	52,69	43,51	7,07	89%	90%	7,16
min	21,00	37,05	4,60	61%	69%	4,73
25%	43,50	40,18	6,44	81%	89%	6,59
50%	57,50	42,48	7,25	93%	91%	7,28
75%	63,50	44,87	7,84	100%	93%	7,93
max	82,00	52,87	9,00	100%	97%	9,02
std	16,78	4,64	1,26	12%	7%	1,23

Fonte: Autora

Para complementar a análise, a mediana de cada variável foi calculada nos três grupos, e os resultados são apresentados nas tabelas de comparação, juntamente com a média.

Tabela 8: Média e mediana não cotistas

	mean	median
CR	7,69	7,88
Nota Final Vestibular	49,34	48,13
Classificação Vestibular	33,90	29,50
Média Ponderada	7,78	7,99
Aproveitamento Disciplinas	0,95	0,98
Frequência Média	0,91	0,92

Fonte: Autora

Tabela 9: Média e mediana ensino público

	mean	median
CR	7,56	7,81
Nota Final Vestibular	45,54	44,65
Classificação Vestibular	46,21	46,00
Média Ponderada	7,64	7,91
Aproveitamento Disciplinas	0,94	1,00
Frequência Média	0,91	0,92

Fonte: Autora

Tabela 10: Média e mediana PPI

	mean	median
CR	7,07	7,25
Nota Final Vestibular	43,51	42,48
Classificação Vestibular	52,69	57,50
Média Ponderada	7,16	7,28
Aproveitamento Disciplinas	0,89	0,93
Frequência Média	0,90	0,91

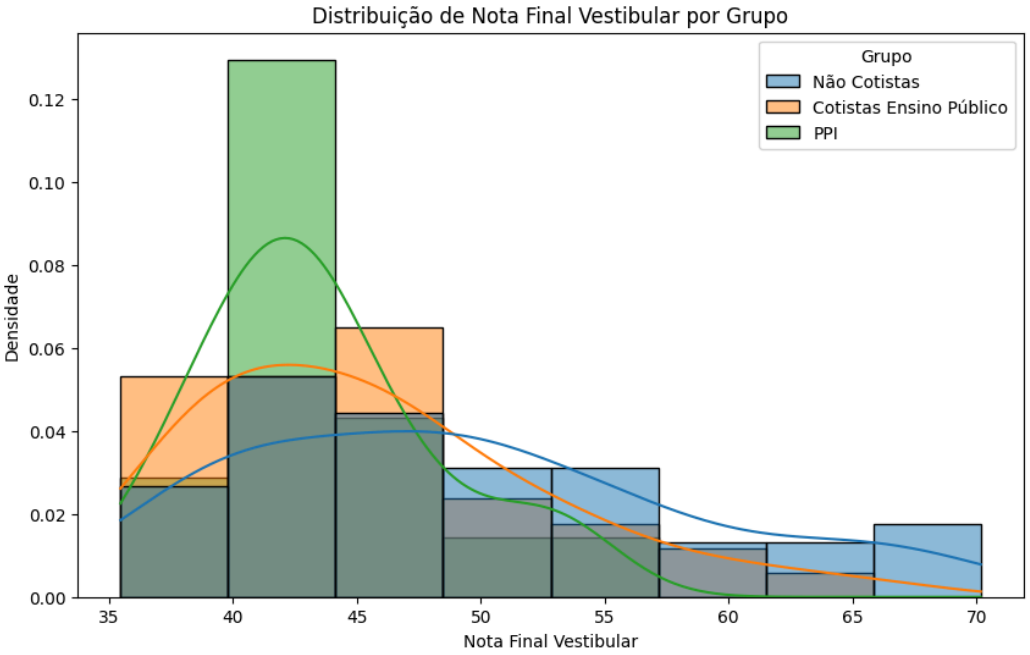
Fonte: Autora

A análise das estatísticas descritivas revela padrões de desempenho que diferenciam os grupos de alunos, com as maiores disparidades concentradas nas métricas de entrada. A Nota Final do Vestibular e a Classificação Vestibular são as que mais evidenciam essa diferença inicial. Os alunos Não Cotistas obtiveram as notas mais altas (média de aproximadamente 49,3), seguidos pelos Cotistas Ensino Público (média de 45,5) e, por fim, pelos PPIs (média de 43,5). Essa hierarquia de desempenho se reflete na classificação, com os não cotistas tendo a melhor média (aproximadamente 33,9) e os cotistas PPI, a pior (aproximadamente 52,7). É interessante notar que os não cotistas apresentam a maior dispersão na nota (desvio padrão de 9,04), enquanto a nota dos cotistas PPI é a mais homogênea (desvio padrão de 4,64). Essa alta variabilidade na classificação para todos os grupos sugere uma grande amplitude nos resultados.

No entanto, as métricas de desempenho contínuo mostram uma aproximação entre os grupos. Embora os não cotistas ainda apresentem médias ligeiramente superiores em indicadores como o CR (Coeficiente de Rendimento) e a Média Ponderada, os grupos de cotistas demonstram um desempenho comparável. O Aproveitamento de Disciplinas é homogêneo, com médias altas e próximas para todos (entre 0,89 e 0,95), indicando que a grande maioria dos alunos, independentemente da categoria, tem um bom desempenho nas disciplinas. A Frequência Média também é praticamente idêntica entre os grupos (entre 0,90 e 0,91). Por outro lado, a variabilidade (desvio padrão) nessas métricas é ligeiramente maior nos grupos de cotistas, o que sugere uma maior dispersão no desempenho acadêmico individual dentro desses grupos.

Para complementar a análise descritiva, a seguir são apresentados histogramas que visualizam a distribuição das variáveis numéricas para cada grupo. Essa representação gráfica nos permite observar como os dados de cada variável se comportam e se agrupam, fornecendo uma base visual para as análises estatísticas subsequentes.

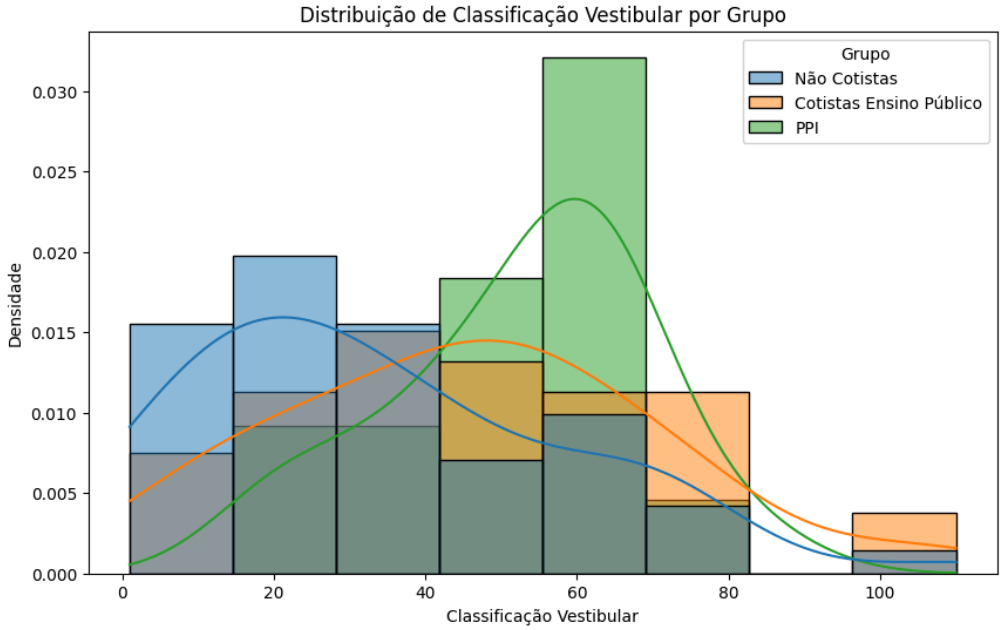
Figura 8: Distribuição nota final vestibular



Fonte: Autora

O gráfico revela uma disparidade entre os grupos logo no processo de ingresso. O grupo de Não Cotistas apresenta a distribuição mais distribuída e alongada, alcançando as pontuações mais altas da escala, acima de 65. Em contrapartida, o grupo PPI se concentra em notas menores, com um pico de densidade entre 40 e 45 pontos, seguido de uma queda brusca. O grupo de Cotistas de Ensino Público ocupa uma posição intermediária, mas ainda com uma frequência maior na metade inferior da pontuação. No geral, as curvas de densidade mostram pouca sobreposição nas notas mais altas, evidenciando pontos de partida bastante distintos para cada categoria de alunos.

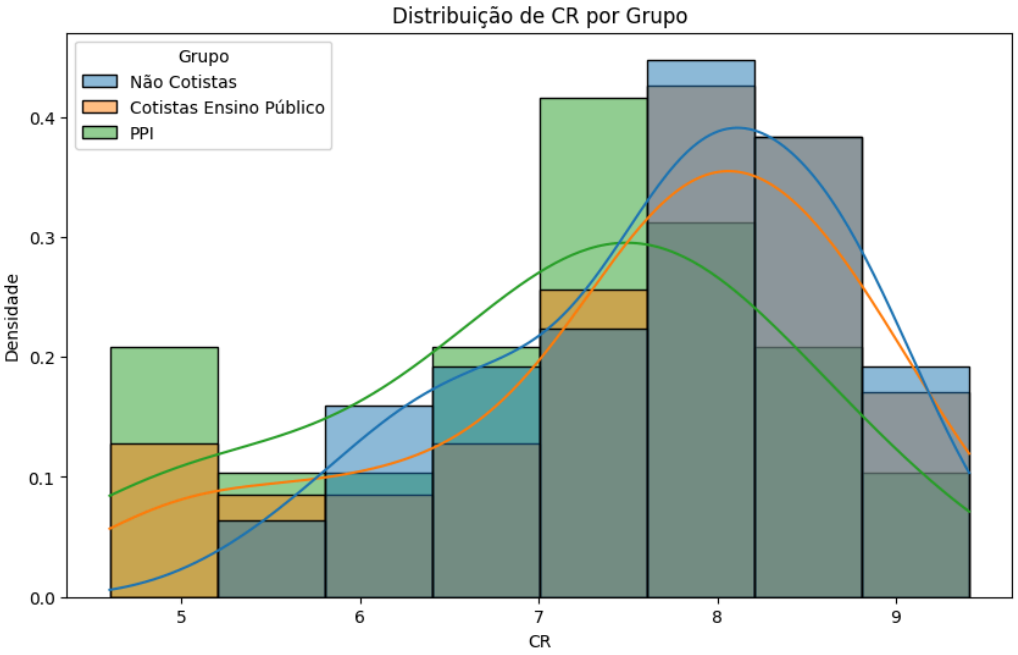
Figura 9: Distribuição classificação vestibular



Fonte: Autora

O gráfico evidencia como os grupos ocupam faixas distintas de posicionamento no ingresso. O grupo de Não Cotistas concentra sua maior densidade nas primeiras classificações (entre 0 e 40), indicando que ocupam as posições de topo no processo seletivo. Em contrapartida, o grupo PPI apresenta um pico de concentração deslocado para a direita, situando-se majoritariamente entre as classificações 60 e 80, o que reflete posições de entrada mais baixas. O grupo de Cotistas de Ensino Público demonstra uma distribuição mais equilibrada, com uma presença constante em posições intermediárias. No geral, as curvas de densidade reforçam uma separação clara entre as posições de liderança e as faixas de classificação ocupadas pelos três grupos.

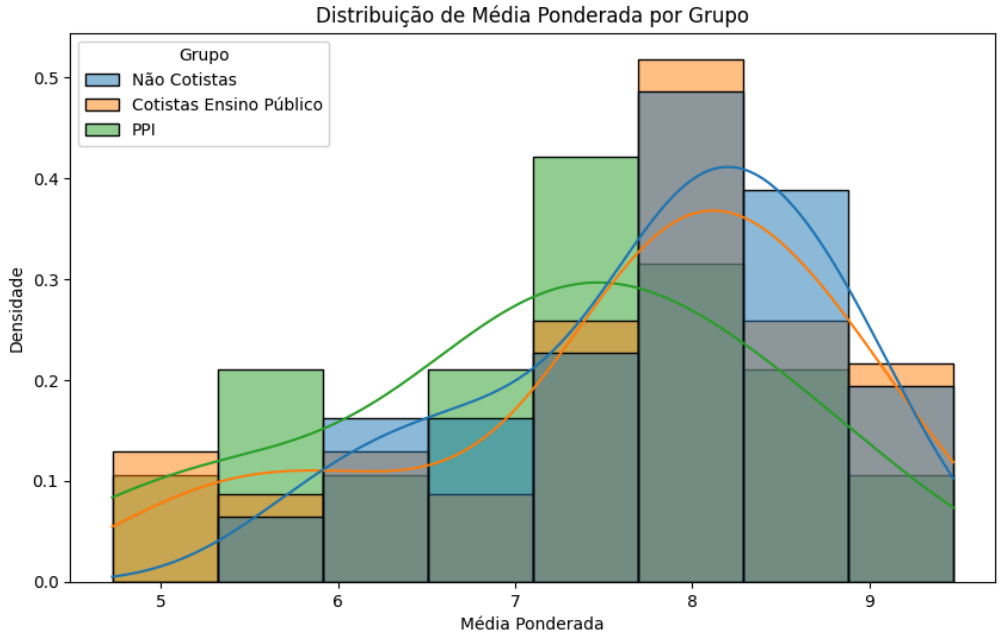
Figura 10: Distribuição CR



Fonte: Autora

O gráfico revela equilíbrio no desempenho entre os grupos, com o volume principal de alunos situado entre os CRs 7 e 9. Enquanto o grupo de Não Cotistas lidera sutilmente nas notas mais altas, os Cotistas PPI apresentam maior dispersão dos dados, evidenciada por uma presença mais marcante na faixa de CR 5. No geral, as curvas de densidade mostram uma sobreposição considerável, indicando comportamentos similares entre as categorias.

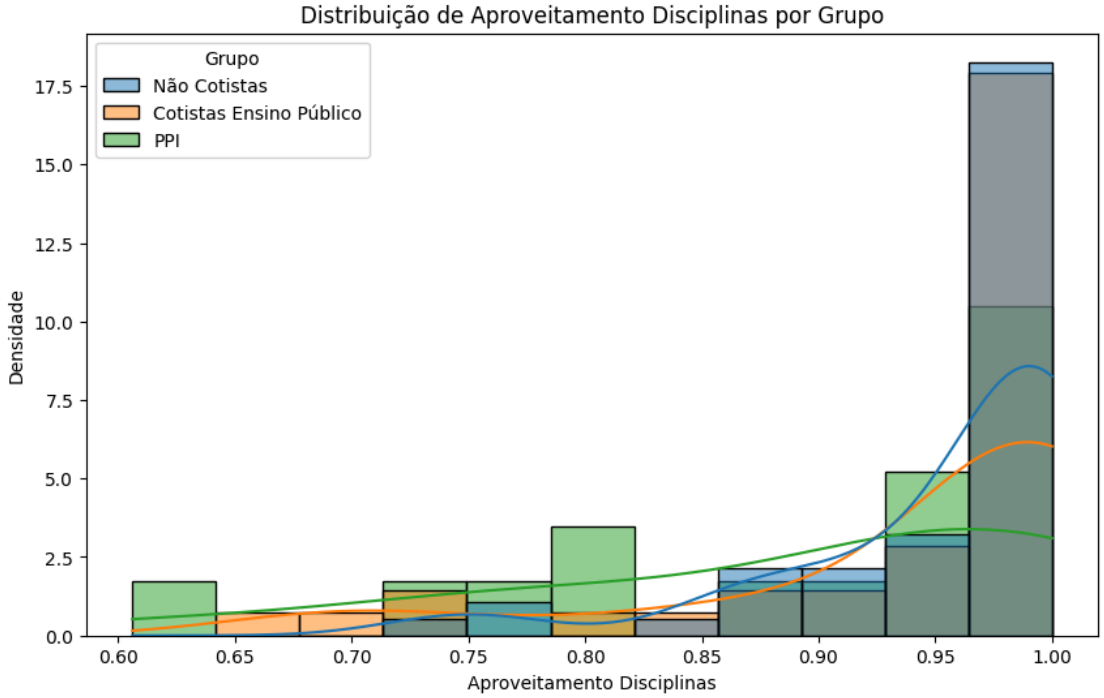
Figura 11: Distribuição média ponderada



Fonte: Autora

A análise da Média Ponderada reafirma a paridade de desempenho entre as categorias, com a grande maioria dos estudantes concentrada no intervalo entre 7,0 e 9,0. O grupo de Não Cotistas apresenta a maior densidade em torno da média 8,2, indicando um desempenho mais homogêneo em notas elevadas. O grupo PPI exibe uma curva mais distribuída e com maior dispersão, evidenciada por uma presença mais acentuada nas faixas inferiores de desempenho (entre 5,0 e 6,0). Apesar dessas variações nos picos de frequência, a sobreposição das curvas de densidade demonstra que alunos de todos os grupos alcançam resultados acadêmicos equivalentes na maior parte da distribuição.

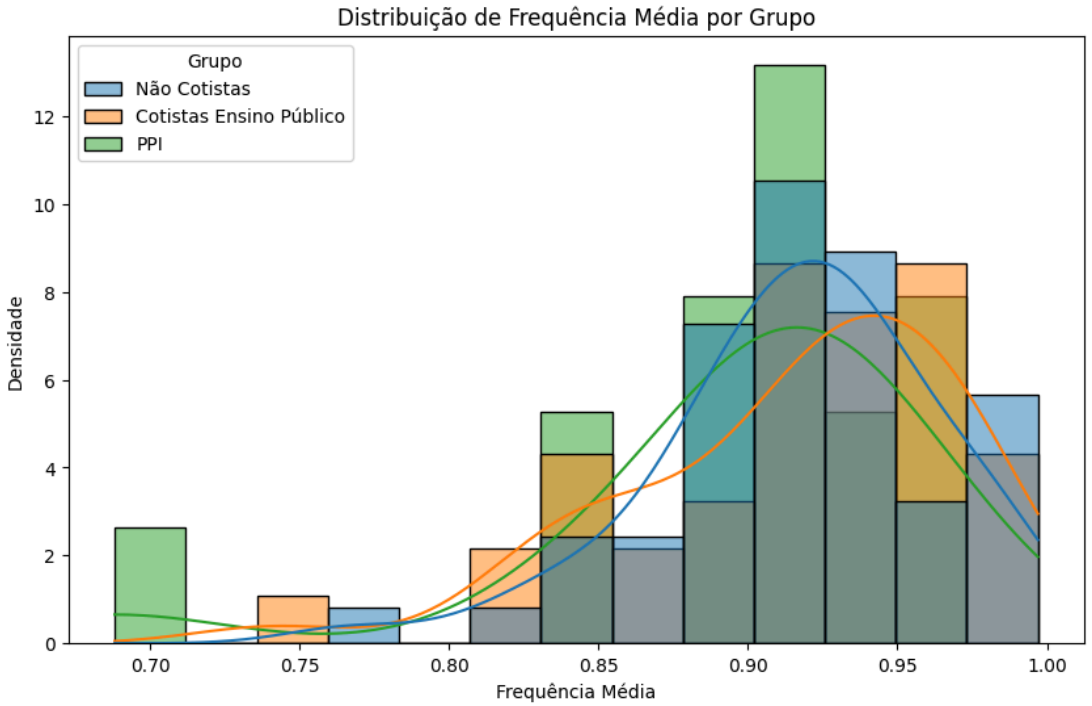
Figura 12: Distribuição aproveitamento disciplinas



Fonte: Autora

O gráfico revela um desempenho acadêmico final positivo para todos os grupos. A característica mais marcante é a forte concentração de estudantes no topo da escala, com a grande maioria apresentando um aproveitamento entre 95% e 100%. Tanto o grupo de Não Cotistas quanto o de Cotistas de Ensino Público apresentam curvas de densidade elevadas e persistentes nesta fase final. Isso demonstra que o aproveitamento integral nas disciplinas é o padrão dominante, refletindo o alto índice de sucesso acadêmico desses grupos. Embora o grupo PPI apresente uma densidade ligeiramente maior em faixas de aproveitamento intermediárias (entre 60% e 85%), ele também possui seu maior volume de alunos no nível máximo de aproveitamento.

Figura 13: Distribuição frequência média

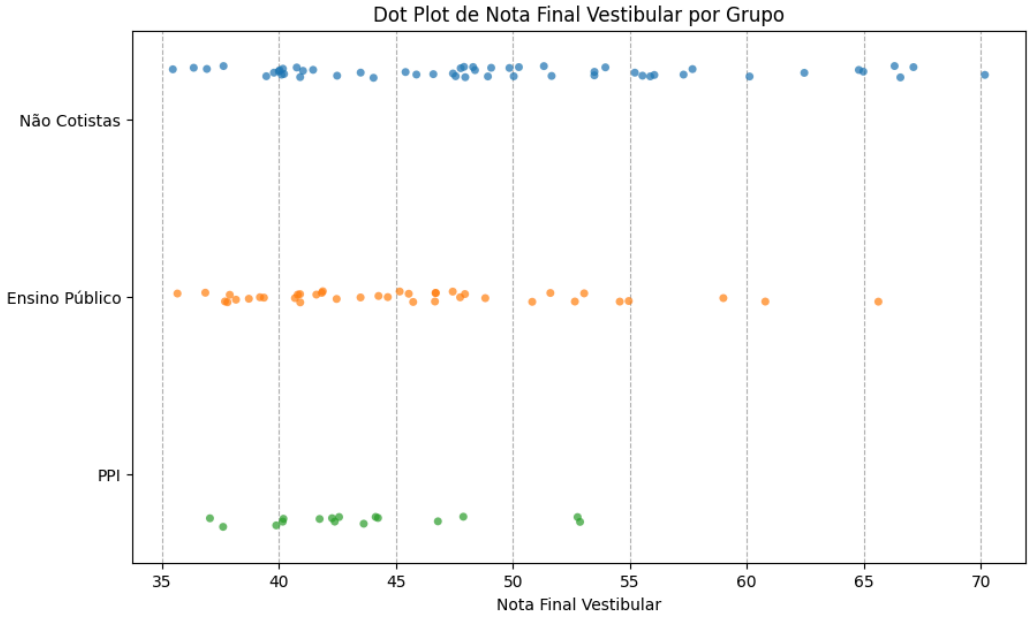


Fonte: Autora

O gráfico de Frequência Média revela um alto nível de engajamento presencial em todos os grupos, com a grande maioria dos alunos mantendo uma assiduidade superior a 90%. Os grupos de Não Cotistas e de Cotistas de Ensino Público apresentam picos de densidade muito próximos, o que indica um comportamento de presença bastante similar. Já o grupo PPI, embora também possua seu maior volume entre 90% e 95%, exibe uma dispersão ligeiramente maior, com uma presença mais notável na faixa de 70% em comparação aos demais.

Também foi realizada uma análise por meio de gráficos de dispersão (*dot plots*), para complementar os testes estatísticos. Esses gráficos, que serão apresentados e comentados a seguir, permitem uma visualização do comportamento das variáveis e das diferenças entre os grupos.

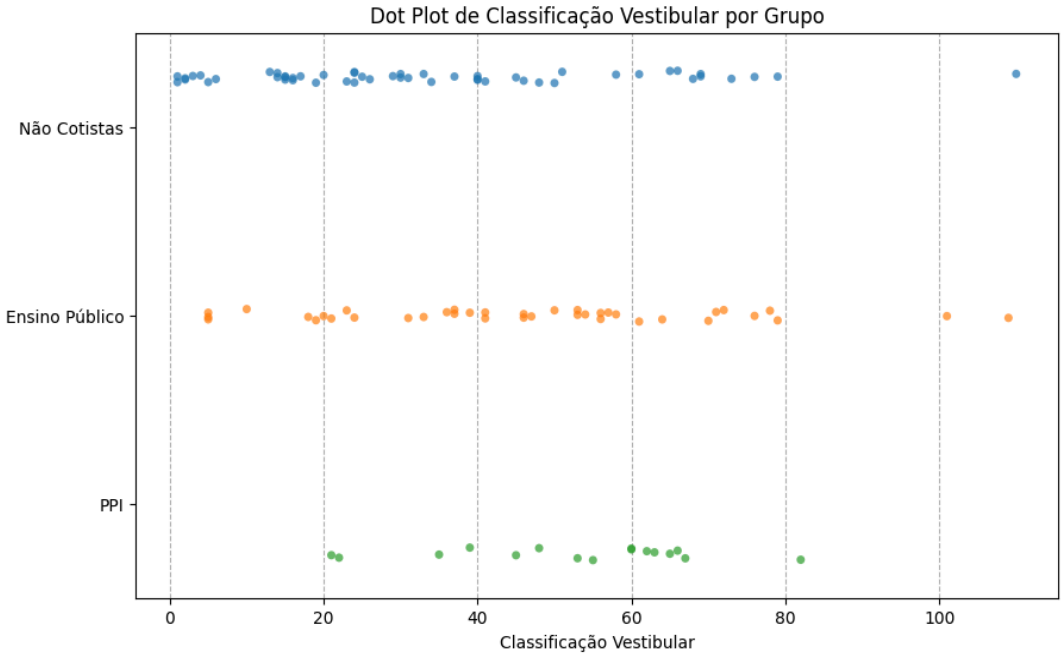
Figura 14: Dot Plot Nota Vestibular



Fonte: Autora

A visualização das notas finais do vestibular revela uma distinção clara entre os grupos. O grupo de não cotistas exibe uma concentração de notas em faixas mais altas. Em contrapartida, as notas dos grupos de Cotistas de Ensino Público e Cotistas PPI parecem mais concentradas em valores inferiores, com o grupo de Cotistas PPI apresentando notas consideravelmente mais baixas.

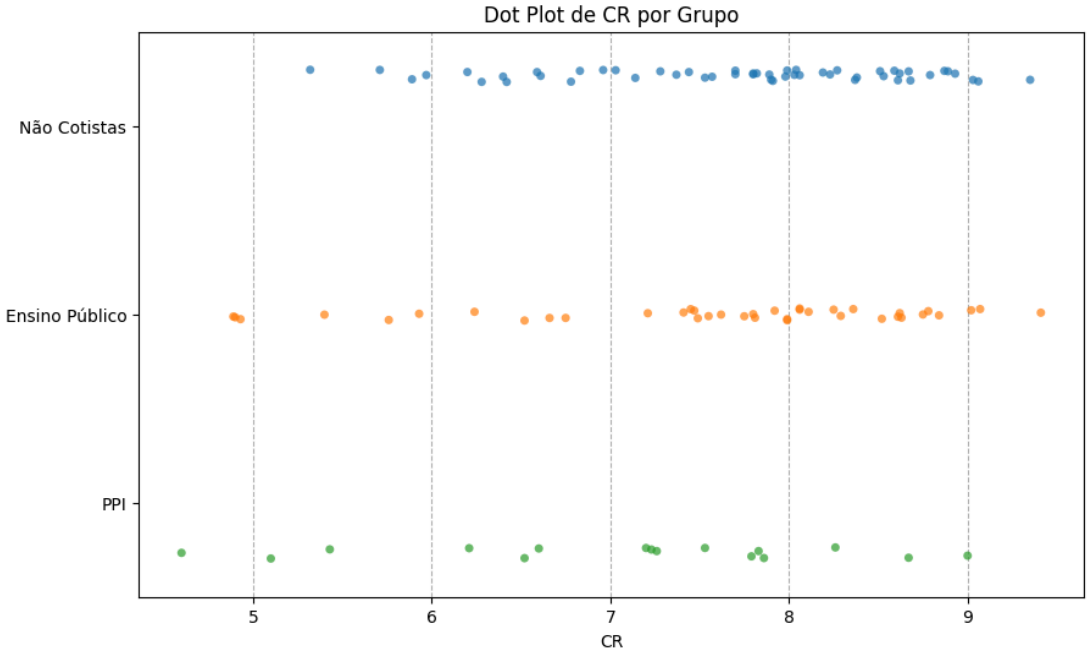
Figura 15: Dot Plot Classificação Vestibular



Fonte: Autora

O *dot plot* da Classificação do Vestibular, onde valores menores indicam uma posição mais favorável, reforça as diferenças observadas. Os pontos do grupo de não cotistas tendem a se agrupar em posições mais vantajosas, enquanto os grupos de Cotistas de Ensino Público e Cotistas PPI apresentam suas classificações em posições mais distantes do início da lista.

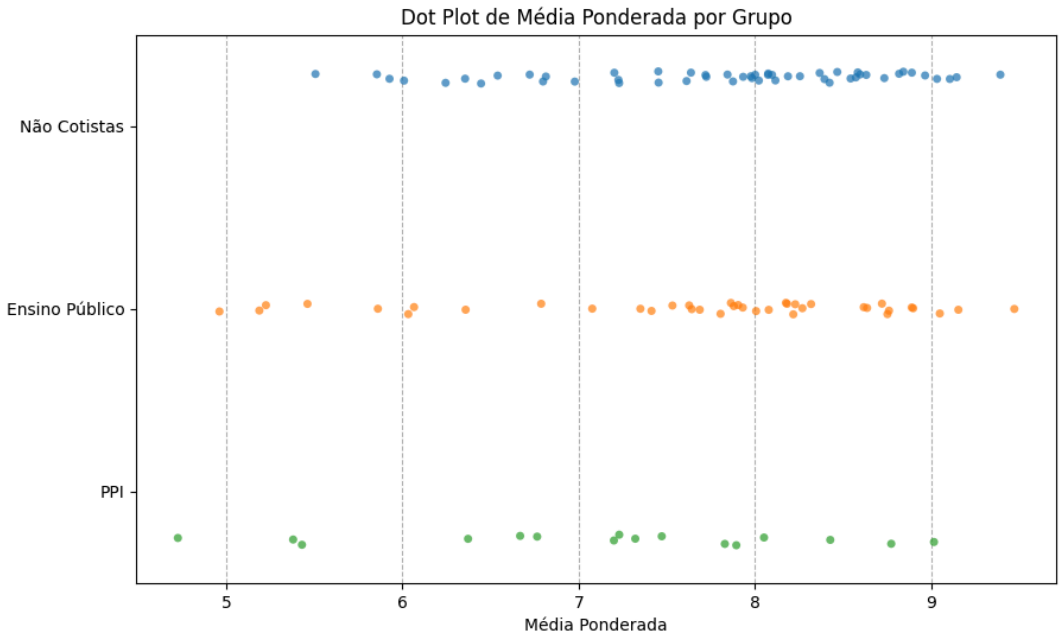
Figura 16: Dot Plot CR



Fonte: Autora

O *dot plot* do CR ilustra a distribuição individual do Coeficiente de Rendimento para cada estudante. A análise visual revela que o grupo de não cotistas demonstra uma maior concentração de pontos em valores mais altos, sugerindo uma menor variabilidade entre seus membros. Em contraste, o grupo Cotistas PPI exibe uma distribuição mais espalhada ao longo do gráfico, o que indica uma maior dispersão de valores. O grupo de Cotistas de Ensino Público se posiciona de forma intermediária, com valores um pouco mais baixos que os não cotistas, mas com uma sobreposição considerável entre todos os grupos.

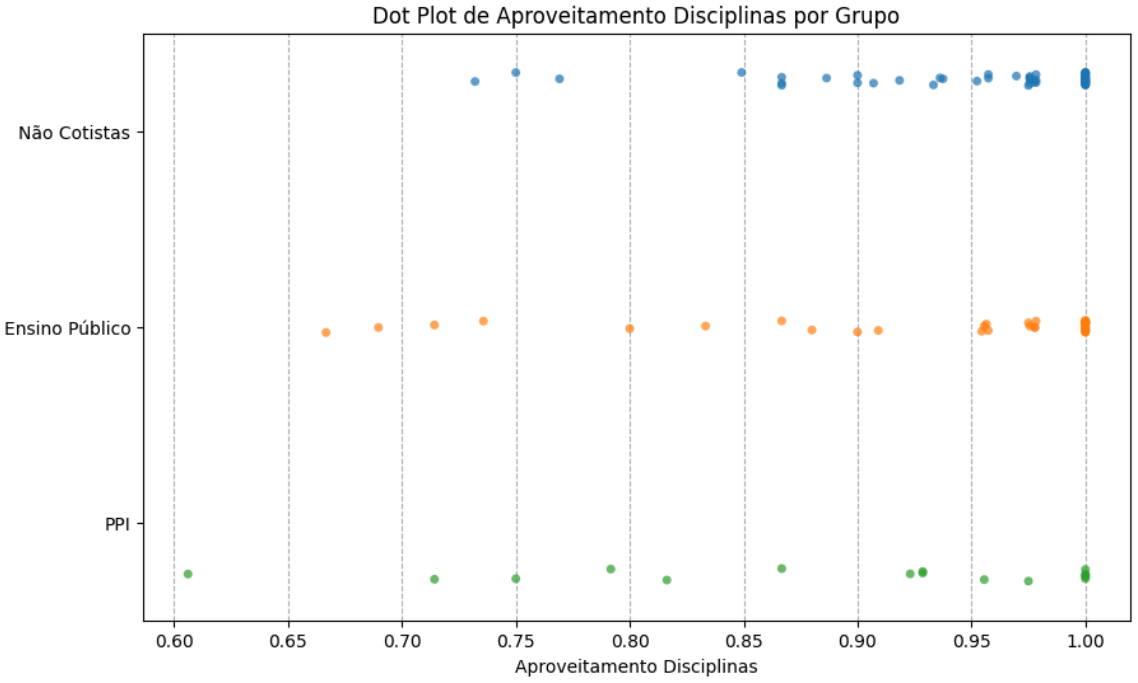
Figura 17: Dot Plot Média Ponderada



Fonte: Autora

O *dot plot* da Média Ponderada demonstra uma distribuição similar à do CR. Os valores do grupo de não cotistas se posicionam, em média, em um patamar um pouco superior. No entanto, a sobreposição entre os três grupos é substancial, com os grupos de Cotistas de Ensino Público e Cotistas PPI exibindo uma dispersão ligeiramente maior.

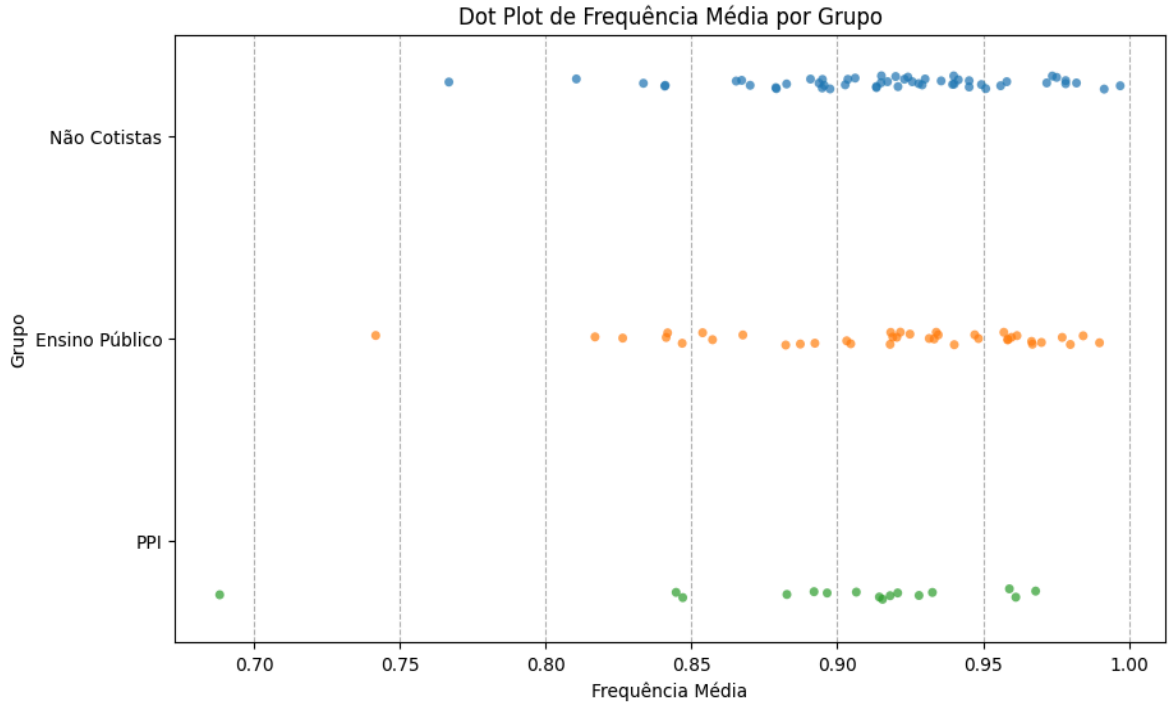
Figura 17: Dot Plot Aproveitamento Disciplinas



Fonte: Autora

A análise visual do aproveitamento em disciplinas mostra que a maioria dos estudantes, independentemente do grupo, alcançou um excelente desempenho. Apesar dessa tendência geral, a distribuição do grupo de Cotistas PPI se diferencia, pois apresenta uma maior proporção de pontos em valores mais baixos, indicando uma dispersão superior e maior variabilidade do que nos outros grupos.

Figura 18: Dot Plot Frequência Média



Fonte: Autora

A distribuição da Frequência Média se mostra bastante homogênea entre os três grupos, com a maioria dos pontos concentrada em valores próximos a 1.0 (100% de frequência). A dispersão é comparável, sem diferenças visuais marcantes que sugiram distinções significativas na frequência média entre os grupos.

4.2.2 Teste de normalidade (Shapiro-Wilk)

Para entender a distribuição dos dados e escolher os testes estatísticos mais adequados, aplicamos o Teste de Shapiro-Wilk nas variáveis numéricas de interesse. Esse teste verifica se uma amostra segue uma distribuição normal, com a hipótese nula de que a distribuição é normal. Rejeitamos a hipótese nula quando o p-valor é inferior a 0,05, indicando que a distribuição não é normal.

A análise revelou que a maioria das variáveis em todos os grupos não segue uma distribuição normal, o que é uma descoberta crucial. Essa constatação fortalece a decisão de usar testes estatísticos não paramétricos para futuras comparações entre os grupos, já que esses testes não exigem a suposição de normalidade dos dados. Analisando os p-valores obtidos nos testes de Shapiro-Wilk para cada variável dentro de cada grupo:

Tabela 11: Testes de Shapiro-Wilk

Variável	Grupo	P-valor	Distribuição
Nota Final Vestibular	Não Cotistas	0,0198	Não Normal
	Ensino Público	0,0194	Não Normal
	PPI	0,1471	Pode ser Normal
Classificação Vestibular	Não Cotistas	0,0077	Não Normal
	Ensino Público	0,5022	Pode ser Normal
	PPI	0,3525	Pode ser Normal
CR (Coeficiente de Rendimento)	Não Cotistas	0,0238	Não Normal
	Ensino Público	0,0060	Não Normal
	PPI	0,6438	Pode ser Normal
Média Ponderada	Não Cotistas	0,0130	Não Normal
	Ensino Público	0,0052	Não Normal
	PPI	0,6902	Pode ser Normal
Aproveitamento Disciplinas	Não Cotistas	0,0000	Não Normal
	Ensino Público	0,0000	Não Normal
	PPI	0,0151	Não Normal
Frequência Média	Não Cotistas	0,0212	Não Normal
	Ensino Público	0,0098	Não Normal
	PPI	0,0013	Não Normal

Fonte: Autora

Os resultados dos testes de normalidade confirmam que a maioria das variáveis numéricas em todos os grupos não segue uma distribuição normal. Notavelmente, algumas variáveis como CR, Nota Final Vestibular e Média Ponderada podem se aproximar da normalidade apenas no grupo de Cotistas Raciais, e a Classificação Vestibular pode ser normal nos grupos de Cotistas. No entanto, a predominância de distribuições não normais (especialmente em variáveis como Aproveitamento Disciplinas e Frequência Média, que não foram normais em nenhum grupo) reforça a adequação do uso de testes estatísticos não paramétricos para comparar as distribuições dessas variáveis entre os três grupos, pois esses testes não assumem normalidade.

4.2.3 Análise comparativa do desempenho acadêmico entre os grupos

A análise comparativa entre os grupos foi conduzida por meio de testes não paramétricos, uma vez que as variáveis de interesse não atenderam aos pressupostos de normalidade, conforme verificado pelos testes de Shapiro-Wilk.

A análise do desempenho acadêmico foi realizada em duas etapas. Primeiramente, comparou-se o grupo de não cotistas com o grupo de cotistas para verificar diferenças gerais. Em seguida, os grupos de cotas foram desagregados para uma análise mais detalhada entre os três grupos distintos: não cotistas, cotistas de ensino público e cotistas PPI.

Para cada teste estatístico, foram obtidos dois valores principais: a estatística do teste e o p-valor. A estatística do teste é um valor calculado a partir dos dados que quantifica a magnitude da diferença entre os grupos. Já o p-valor é a probabilidade de se observar uma diferença tão grande ou maior que a encontrada, sob a suposição de que não há diferença real entre os grupos (hipótese nula). Um p-valor abaixo do nível de significância ($p < 0.05$) é considerado evidência suficiente para rejeitar a hipótese nula.

Para a comparação de desempenho entre o grupo de *não cotistas* e o *grupo cotistas*, empregou-se o Teste U de Mann-Whitney. A hipótese nula (H0) do teste postula que não há diferença na distribuição da variável entre os grupos, enquanto a hipótese alternativa (H1) sugere que as distribuições são diferentes. Os resultados obtidos foram:

Tabela 12: Teste U de Mann-Whitney

Variável	Estatística do Teste (U)	P-valor	Resultado Estatístico ($p < 0.05$)
CR	1713,0	0,3089	Não Significativo
Nota Final Vestibular	1843,0	0,0101	Significativo
Média Ponderada	1725,0	0,2765	Não Significativo
Aproveitamento em Disciplinas	1654,0	0,4811	Não Significativo
Frequência Média	1494,5	0,7907	Não Significativo

Fonte: Autora

A análise revelou uma diferença estatisticamente significativa ($p < 0.05$) apenas para a 'Nota Final do Vestibular'. Para as demais variáveis, os p-valores foram superiores a 0.05, indicando que não há evidência estatística para concluir que as distribuições dessas métricas de desempenho acadêmico diferem significativamente entre os dois grupos.

Para a análise dos três grupos, *não cotistas*, *cotistas de ensino público* e *cotistas PPI*, foi empregado o Teste de Kruskal-Wallis, utilizado para comparar três ou mais populações independentes. Os resultados obtidos foram:

Tabela 13: Teste de Kruskal-Wallis

Variável	Estatística do Teste (H)	P-valor	Resultado Estatístico ($p < 0.05$)
CR	3,5343	0,1708	Não Significativo
Nota Final Vestibular	7,2103	0,0272	Significativo
Média Ponderada	3,6546	0,1608	Não Significativo
Aproveitamento em Disciplinas	3,8793	0,1438	Não Significativo
Frequência Média	1,1153	0,5725	Não Significativo
Classificação Vestibular	10,9268	0,0042	Significativo

Fonte: Autora

Os resultados do Teste de Kruskal-Wallis foram estatisticamente significativos ($p < 0.05$) para as variáveis 'Nota Final do Vestibular' e 'Classificação do Vestibular'. Tal achado

sinaliza que existe uma diferença estatisticamente significativa na distribuição dessas variáveis entre, no mínimo, um par de grupos.

Para identificar quais pares de grupos apresentavam diferenças significativas, foram aplicados testes *post-hoc*. Esses testes são usados para realizar comparações pareadas entre os grupos, a fim de localizar as diferenças específicas que o teste inicial (Kruskal-Wallis) detectou de forma geral. Para essa análise, foi utilizado o Teste de Dunn (com ajuste de Bonferroni). Os resultados para a 'Nota Final do Vestibular' foram os seguintes:

Tabela 14: Teste de Dunn (*post-hoc*) Nota Vestibular

Comparação	P-valor	Resultado Estatístico ($p < 0.05$)
Cotistas Ensino Público vs. Cotistas PPI	1.0000	Não Significativo
Cotistas Ensino Público vs. Não Cotistas	0.1243	Não Significativo
Cotistas PPI vs. Não Cotistas	0.0636	Não Significativo

Fonte: Autora

Apesar de o teste global de Kruskal-Wallis ter indicado uma diferença para a 'Nota Final do Vestibular', a análise *post-hoc* não revelou diferenças significativas em nenhuma das comparações pareadas. Esse achado sugere que, embora a distribuição geral dos grupos seja diferente, as diferenças entre os pares não foram fortes o suficiente para serem consideradas estatisticamente significativas após o ajuste para múltiplas comparações.

Para a 'Classificação do Vestibular', o Teste de Dunn forneceu os seguintes resultados:

Tabela 15: Teste de Dunn (*post-hoc*) Classificação Vestibular

Comparação	P-valor	Resultado Estatístico ($p < 0.05$)
Cotistas Ensino Público vs. Cotistas PPI	0.7908	Não Significativo
Cotistas Ensino Público vs. Não Cotistas	0.0513	Não Significativo
Cotistas PPI vs. Não Cotistas	0,0102	Significativo

Fonte: Autora

Neste caso, a análise *post-hoc* identificou uma diferença estatisticamente significativa ($p < 0.05$) exclusivamente entre os grupos 'Cotistas PPI' e 'Não Cotistas'.

4.2.4 Análise de correlação entre nota do vestibular e coeficiente de rendimento

Para avaliar a relação entre a 'Nota Final do Vestibular' e o 'CR' (Coeficiente de Rendimento), foi realizada uma análise de correlação de Spearman. Este teste não paramétrico é a ferramenta estatística adequada para medir a força e a direção da relação entre duas variáveis, para dados que não seguem uma distribuição normal, como é o caso das variáveis em estudo. A análise foi conduzida separadamente para cada um dos três grupos. Os resultados indicam o coeficiente de correlação de Spearman (ρ), que mede a força e a direção da relação, e o p-valor. O p-valor, por sua vez, é um indicador crucial para a tomada de decisões: se for menor que 0.05, a relação é considerada estatisticamente significativa.

Tabela 16: Correlação de Spearman Nota Vestibular e CR

Grupo	Coeficiente de Correlação (ρ)	P-valor
Não Cotistas	0,5007	0,0002
Cotistas Ensino Público	0,4952	0,0014
Cotistas PPI	0,2529	0,3446

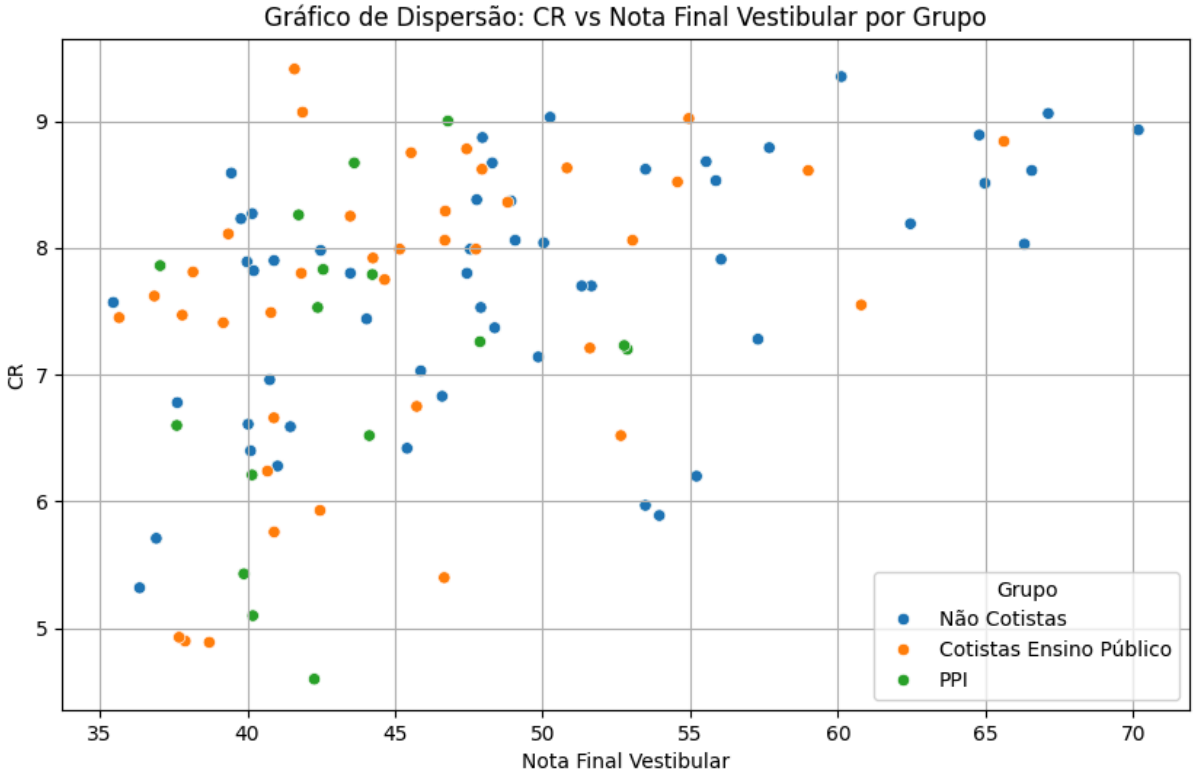
Fonte: Autora

- Não Cotistas: O coeficiente de correlação de Spearman foi de $\rho = 0.5007$, com um p-valor de 0.0002. Este resultado demonstra uma relação monotônica positiva e estatisticamente significativa entre a nota do vestibular e o CR para este grupo. O valor do coeficiente sugere uma força de relação de moderada a forte.
- Cotistas Ensino Público: Para este grupo, o coeficiente de correlação foi de $\rho = 0.4952$, com um p-valor de 0.0014. Similar ao grupo de não cotistas, o resultado indica uma relação monotônica positiva, significativa e de força comparável, evidenciando que, também para os cotistas de ensino público, notas mais altas no vestibular estão associadas a melhores coeficientes de rendimento.
- Cotistas PPI: Neste grupo, o coeficiente de correlação foi de $\rho = 0.2529$, com um p-valor de 0.3446. Como o p-valor é superior a 0.05, não há evidência estatística significativa de uma relação monotônica entre as duas variáveis para os cotistas

raciais. Isso sugere que, para este grupo, a nota do vestibular não possui uma relação monotônica clara com o desempenho no CR.

Visualmente, os gráficos de dispersão corroboram esses achados. A visualização para os grupos de Não Cotistas e Cotistas de Ensino Público mostra uma clara tendência de crescimento, confirmando a correlação positiva observada nos testes. De acordo com o gráfico, podemos observar que alunos que obtêm uma nota no vestibular superior a 55 tendem a possuir um CR maior que 7. Em contrapartida, o gráfico para os Cotistas PPI exibe uma dispersão de pontos muito maior, sem uma tendência monotônica evidente, o que valida a ausência de correlação estatisticamente significativa.

Figura 19: Correlação CR e Nota Vestibular



Fonte: Autora

4.2.5 Síntese dos resultados estatísticos

Com base na análise exploratória e estatística dos dados comparativos entre alunos Não Cotistas, Cotistas de Ensino Público e Cotistas PPI, observamos padrões distintos e similaridades importantes. As estatísticas descritivas iniciais sugeriram que os alunos Não Cotistas tendem a apresentar médias ligeiramente superiores em métricas como CR, Nota

Final Vestibular e Média Ponderada quando comparados ao grupo combinado de Cotistas. No entanto, ao desagregar o grupo de Cotistas, os testes de hipóteses não paramétricos (Mann-Whitney e Kruskal-Wallis) e as visualizações (histogramas e dot plots) revelaram nuances. Especificamente, houve uma diferença estatisticamente significativa na distribuição da Nota Final do Vestibular e da Classificação do Vestibular entre os três grupos. Os testes post-hoc indicaram que a diferença na Classificação do Vestibular residiu principalmente na comparação entre os alunos PPI e Não Cotistas, com os Não Cotistas geralmente obtendo melhores classificações. Para a Nota Final do Vestibular, embora o teste global (Kruskal-Wallis) tenha sido significativo, os testes post-hoc não identificaram pares de grupos com diferenças estatisticamente fortes após ajustes para comparações múltiplas.

No que tange ao desempenho acadêmico contínuo, como CR, Média Ponderada, Aproveitamento de Disciplinas e Frequência Média, os testes de hipóteses não encontraram evidência estatística significativa de diferenças nas distribuições entre os três grupos, sugerindo que, apesar das variações iniciais nas médias, a performance geral nessas métricas ao longo do curso tende a ser mais comparável. A análise de correlação de Spearman entre a Nota Final do Vestibular e o CR revelou uma relação positiva e significativa para os grupos Não Cotistas e Ensino Público, indicando que notas de vestibular mais altas estão associadas a CRs mais altos nestes grupos; contudo, essa relação não foi estatisticamente significativa para o grupo PPI.

Em suma, os dados apontam para diferenças mais marcantes na performance no processo seletivo de vestibular, particularmente para o grupo PPI em relação aos Não Cotistas, do que no desempenho acadêmico subsequente, onde as distribuições das métricas de rendimento contínuo se mostraram mais similares entre todos os grupos, embora a força da associação entre a nota de ingresso e o rendimento acadêmico varie entre os grupos.

Após as análises estatísticas realizadas, que proporcionaram um aprofundado entendimento sobre as características e a distribuição dos dados, foi possível desenvolver o sistema de regras *fuzzy*. O funcionamento e a estrutura desse sistema serão apresentados detalhadamente no próximo capítulo.

5. MODELAGEM VIA SBRF

Para realizar a modelagem via Sistemas Baseados em Regras *Fuzzy* (SBRF), utilizamos o Google Colab com a linguagem de programação Python. Esta plataforma oferece um ambiente de desenvolvimento integrado e acessível para implementação, simulação e análise de sistemas de inferência *fuzzy*.

5.1 As variáveis do sistema

O primeiro passo da modelagem consistiu na definição das variáveis do sistema, selecionadas com base nos dados coletados durante o processo anterior de análise do desempenho acadêmico. Essas variáveis foram escolhidas por apresentarem indicadores fundamentais que caracterizam tanto o perfil de ingresso quanto o desempenho ao longo da trajetória acadêmica do estudante.

As variáveis de entrada selecionadas para compor o sistema *fuzzy* foram:

Classificação no Vestibular: posição relativa do estudante no processo seletivo, indicando seu desempenho comparativo no momento do ingresso na universidade.

Nota Final no Vestibular: pontuação absoluta obtida no exame de seleção, refletindo o nível de conhecimento demonstrado para o acesso ao curso superior.

Coeficiente de Rendimento (CR): índice que sintetiza o desempenho acadêmico global do estudante através de uma média aritmética simples, considerando as notas obtidas em todas as disciplinas cursadas.

Média Ponderada: Média que considera o peso de cada disciplina, onde o peso é o número de créditos. Disciplinas com mais créditos têm maior impacto na sua média.

Aproveitamento de Disciplinas (%): proporção de disciplinas cursadas com aprovação direta, sem necessidade de repetição.

Frequência Média (%): percentual médio de presença nas disciplinas.

A partir dessas entradas, o sistema calcula a variável de saída "Desempenho Geral do Aluno", que representa uma avaliação total do rendimento acadêmico estudantil. A Classificação e a Nota Final no Vestibular oferecem um panorama do perfil de ingresso e do potencial inicial. O CR e a Média Ponderada refletem diretamente a performance nas

avaliações teóricas e práticas. O Aproveitamento de Disciplinas indica a eficiência no processo de aprendizagem e a progressão curricular, enquanto a Frequência Média demonstra o nível de engajamento e participação nas atividades acadêmicas. Essa combinação de indicadores permite uma análise que considera tanto fatores de entrada quanto de processo.

5.2 Definição das funções de pertinência

Após a identificação das variáveis linguísticas de entrada e saída, procedeu-se à definição das funções de pertinência para cada uma delas. Esta etapa é fundamental no desenvolvimento de sistemas *fuzzy*, pois estabelece como os valores numéricos são mapeados em termos linguísticos. Para cada variável, foram estabelecidos conjuntos *fuzzy* utilizando funções triangulares e trapezoidais, considerando as características específicas de cada indicador e a opinião de 4 professores especialistas. O objetivo dessa consulta foi definir os parâmetros para as categorias 'ruim', 'regular', 'bom' e 'muito bom'. Com base nos dados coletados, calculou-se a média aritmética ponderada para cada termo, definindo-se os 'picos' de pertinência. A seguir, detalham-se as definições das funções para cada variável, acompanhadas de suas representações gráficas.

5.2.1. Classificação no Vestibular (1 a 120)

Para esta variável, foi definido um universo de discurso de 1 a 120. A observação da base de dados revelou que a pior classificação registrada foi a 110ª posição, sendo o limite superior estabelecido em 120 para contemplar um intervalo razoável além dos valores observados. É importante notar que valores menores indicam melhor desempenho, ou seja, 1ª posição é melhor que 120ª posição.

Para definir os intervalos das funções de pertinência, levamos em conta a visão dos quatro especialistas e o critério da programadora. As respostas fornecidas pelos especialistas foram organizadas em intervalos de 10 em 10 unidades, de forma a padronizar as avaliações e viabilizar o cálculo da média aritmética ponderada. Cada especialista pôde indicar um ou mais valores dentro dessa escala para representar sua percepção em relação a cada termo linguístico.

Nesse contexto, a indicação do valor 10 representa a consideração do intervalo de 0 a 10; a indicação dos valores 10 e 20 representa a consideração dos intervalos de 0 a 10 e de 10 a 20, isto é, do intervalo total de 0 a 20; de forma análoga, a indicação dos valores 10, 20 e 30

corresponde à consideração dos intervalos de 0 a 10, de 10 a 20 e de 20 a 30, totalizando o intervalo de 0 a 30.

Na tabela abaixo, apresentam-se as respostas obtidas junto aos especialistas.

Tabela 17: Avaliação dos especialistas - Classificação no Vestibular

Especialistas/ Avaliação	**Muito Bom**	**Bom**	**Regular**	**Ruim**
Especialista 1	[0, 10]	]10, 30]	]30, 50]	]50, 110]
Especialista 2	[0, 20]	]10, 30]	]20, 50]	]50, 110]
Especialista 3	[0, 30]	]20, 60]	]50, 90]	]80, 110]
Especialista 4	[0, 10]	]10, 20]	]20, 30]	]30, 110]

Fonte: Autora

Utilizando como exemplo o termo linguístico “muito bom”, as avaliações foram as seguintes: o Especialista 1 indicou o valor 10 (intervalo de 0 a 10); o Especialista 2 indicou os valores 10 e 20 (intervalo de 0 a 20); o Especialista 3 indicou os valores 10, 20 e 30 (intervalo de 0 a 30); e o Especialista 4 indicou novamente o valor 10 (intervalo de 0 a 10).

A partir dessas indicações, foi realizada a contagem das frequências associadas a cada valor, resultando em quatro ocorrências para o valor 10, duas ocorrências para o valor 20 e uma ocorrência para o valor 30. Com base nessa distribuição, a média aritmética ponderada foi calculada considerando os valores indicados e suas respectivas frequências, conforme a expressão

$$\frac{4.10 + 2.20 + 1.30}{4 + 2 + 1} = \frac{110}{7} = 15,71$$

De forma análoga, considerando os mesmos critérios, foi calculada a média ponderada para os demais termos:

- Bom

$$\frac{3.20 + 3.30 + 1.40 + 1.50 + 1.60}{3 + 3 + 1 + 1 + 1} = \frac{300}{9} = 33,33$$

- Regular

$$\frac{2.30 + 2.40 + 2.50 + 60 + 70 + 80 + 90}{2 + 2 + 2 + 1 + 1 + 1 + 1} = \frac{540}{10} = 54$$

- Ruim

$$\frac{40 + 50 + 3.60 + 3.70 + 3.80 + 4.90 + 4.100 + 4.110}{1 + 1 + 3 + 3 + 3 + 4 + 4 + 4} = \frac{1920}{23} = 83,48$$

A partir dos valores obtidos por meio da média ponderada de cada termo linguístico, foram definidos os intervalos das funções de pertinência. Para isso, adotou-se um critério de expansão em torno do valor central de cada termo, acrescentando-se e subtraindo-se 15 unidades, com o objetivo de permitir a sobreposição gradual entre as categorias. Dessa forma, os intervalos ficaram definidos conforme apresentado na tabela a seguir.

Tabela 18: Delimitadores de “Classificação no Vestibular”

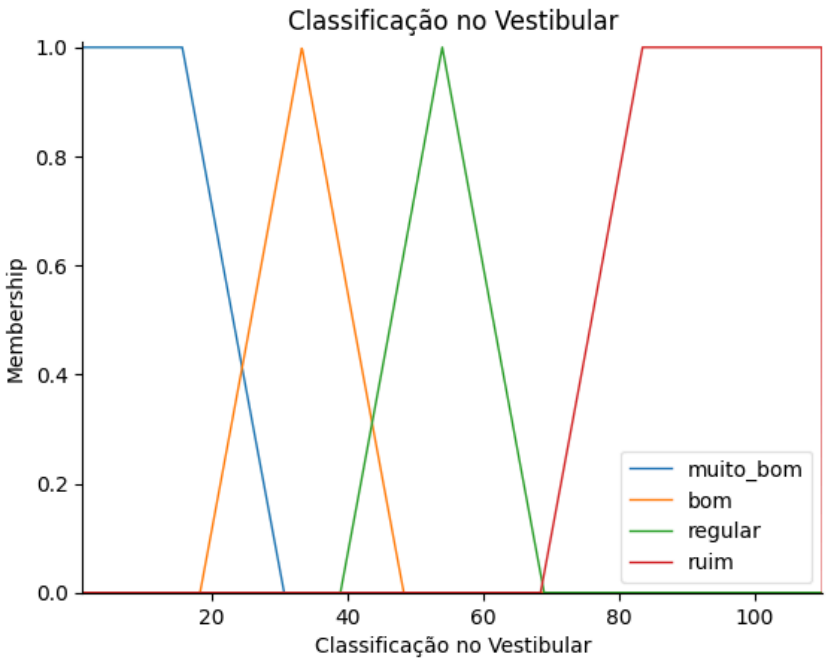
CONJUNTO DE TERMOS	TIPOS	DELIMITADORES
Muito bom	Trapezoidal	[1 1 15,71 30,71]
Bom	Triangular	[18,33 33,33 48,33]
Regular	Triangular	[39 54 69]
Ruim	Trapezoidal	[68,48 83,48 110 110]

Fonte: Autora

- **Muito Bom**: Função trapezoidal [1 1 15,71 30,71] - Pertinência total para classificações de 1 a 15,71, diminuindo linearmente de 15,71 a 30,71.
- **Bom**: Função triangular [18,33 33,33 48,33] - A pertinência aumenta linearmente de 18,33 até 33,33 e depois diminui linearmente até 48,33.
- **Regular**: Função triangular [39 54 69] - A pertinência aumenta linearmente de 39 até 54 e depois diminui linearmente até 69.
- **Ruim**: Função trapezoidal [68,48 83,48 110 110] - A pertinência aumenta linearmente de 68,48 até 83,48. Pertinência total para classificações de 83,48 em diante.

Abaixo, apresenta-se a representação gráfica dessa variável

Figura 20: Classificação no Vestibular



Fonte: Autora

5.2.2 Nota Final no Vestibular (0 a 100)

Para esta variável, foi estabelecido um universo de discurso de 0 a 100, correspondendo ao intervalo de pontuação possível no vestibular da UNESP, onde 100 representa a nota máxima que pode ser obtida considerando o desempenho conjunto na primeira e segunda fase do exame.

A seguir, apresenta-se as percepções acerca do que seriam pontuações “ruim”, “regular”, “bom” e “muito bom”, obtidas na consulta com os especialistas.

Tabela 19: Avaliação dos especialistas - Nota Final no Vestibular

Especialistas/ Avaliação	**Ruim**	**Regular**	**Bom**	**Muito Bom**
Especialista 1	[0, 30]	]30, 60]	]60, 80]	]80, 100]
Especialista 2	[0, 30]	]20, 60]	]40, 70]	]70, 100]
Especialista 3	[0, 30]	]20, 60]	]50, 80]	]70, 100]
Especialista 4	[0, 10]	]10, 30]	]30, 60]	]60, 100]

Fonte: Autora

Os intervalos das funções de pertinência foram definidos da mesma forma que a variável anterior, através do cálculo da média ponderada em cada termo, considerando a visão de cada especialista. Os cálculos estão apresentados a seguir.

- Ruim

$$\frac{4.10 + 3.20 + 3.30}{4 + 3 + 3} = \frac{190}{10} = 19$$

- Regular

$$\frac{20 + 3.30 + 3.40 + 3.50 + 3.60}{1 + 3 + 3 + 3 + 3} = \frac{560}{13} = 43,07$$

- Bom

$$\frac{40 + 2.50 + 3.60 + 3.70 + 2.80}{1 + 2 + 3 + 3 + 2} = \frac{690}{11} = 62,72$$

- Muito Bom

$$\frac{70 + 3.80 + 4.90 + 4.100}{1 + 3 + 4 + 4} = \frac{1070}{12} = 89,16$$

Da mesma forma, adotou-se o critério de expansão em torno do valor central de cada termo, com acréscimo e decréscimo de 15 unidades, de modo a permitir uma sobreposição gradual entre as categorias. Os intervalos obtidos são apresentados na tabela a seguir.

Tabela 20: Delimitadores de “Nota Final no Vestibular”

CONJUNTO DE TERMOS	TIPOS	DELIMITADORES
Ruim	Trapezoidal	[0 0 19 34]
Regular	Triangular	[28,07 43,07 58,07]
Bom	Triangular	[47,72 62,72 77,72]
Muito bom	Trapezoidal	[74,16 89,16 100 100]

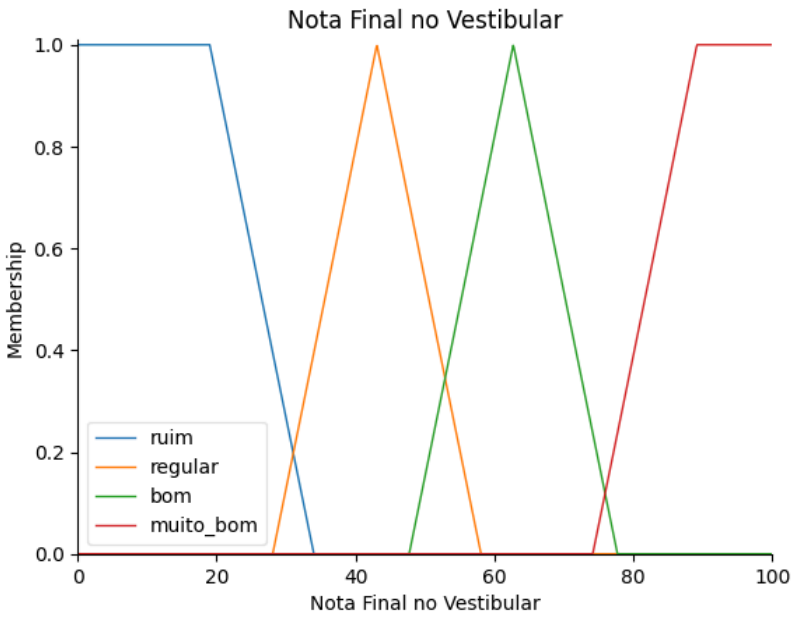
Fonte: Autora

- **Ruim**: Função trapezoidal [0 0 19 34] - Pertinência total para pontuações de 0 a 19, diminuindo linearmente de 19 a 34.
- **Regular**: Função triangular [28,07 43,07 58,07] - A pertinência aumenta linearmente de 28,07 até 43,07 e depois diminui linearmente até 58,07.
- **Bom**: Função triangular [47,72 62,72 77,72] - A pertinência aumenta linearmente de 47,72 até 62,72 e depois diminui linearmente até 77,72.

- **Muito Bom**: Função trapezoidal [74,16 89,16 100 100] - A pertinência aumenta linearmente de 74,16 até 89,16. Pertinência total para pontuações de 89,16 em diante.

A seguir, apresenta-se a representação gráfica dessa variável.

Figura 21: Nota Final Vestibular



Fonte: Autora

5.2.3 Coeficiente de Rendimento - CR (0 a 10)

Para esta variável, foi definido um universo de discurso de 0 a 10, seguindo a escala padrão utilizada pela UNESP para o Coeficiente de Rendimento, que sintetiza o desempenho acadêmico global do estudante durante o curso. A seguir, apresenta-se a avaliação dos especialistas referente ao CR.

Tabela 21: Avaliação dos especialistas - CR

Especialistas/ Avaliação	Ruim	Regular	Bom	Muito Bom
Especialista 1	[0, 4]	]4, 6]	]6, 8]	]8, 10]
Especialista 2	[0, 4]	]3, 6]	]5, 8]	]7, 10]
Especialista 3	[0, 3]	]3, 6]	]6, 8]	]8, 10]
Especialista 4	[0, 4]	]4, 6]	]6, 8]	]8, 10]

Fonte: Autora

Desse modo, o cálculo da média ponderada para cada termo se deu da seguinte forma:

- Ruim

$$\frac{4.1 + 4.2 + 4.3 + 3.4}{4 + 4 + 4 + 3} = \frac{36}{15} = 2,4$$

- Regular

$$\frac{2.4 + 4.5 + 4.6}{10} = \frac{52}{10} = 5,2$$

- Bom

$$\frac{1.6 + 4.7 + 4.8}{1 + 4 + 4} = \frac{66}{9} = 7,3$$

- Muito Bom

$$\frac{1.8 + 4.9 + 4.10}{1 + 4 + 4} = \frac{84}{9} = 9,3$$

Na sequência, foi estabelecido um acréscimo e decréscimo de 2 unidades em torno do pico de cada termo para definir os intervalos das funções.

Tabela 22: Delimitadores de “Coeficiente de Rendimento”

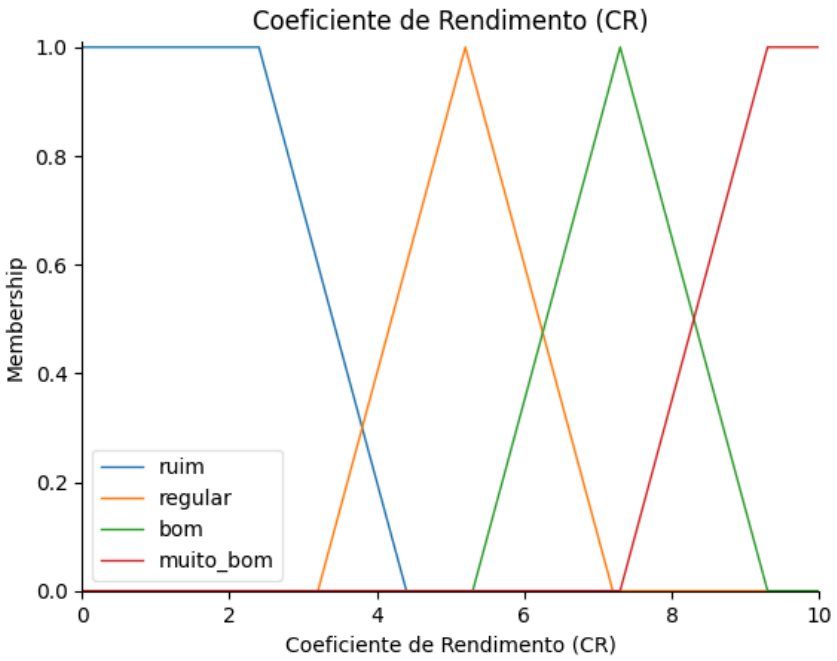
CONJUNTO DE TERMOS	TIPOS	DELIMITADORES
Ruim	Trapezoidal	[0 0 2,4 4,4]
Regular	Triangular	[3,2 5,2 7,2]
Bom	Triangular	[5,3 7,3 9,3]
Muito bom	Trapezoidal	[7,3 9,3 10 10]

Fonte: Autora

- **Ruim**: Função trapezoidal [0 0 2,4 4,4] - Pertinência total para notas de 0 a 2,4, diminuindo linearmente de 2,4 a 4,4.
- **Regular**: Função triangular [3,2 5,2 7,2] - A pertinência aumenta linearmente de 3,2 até 5,2 e depois diminui linearmente até 7,2.
- **Bom**: Função triangular [5,3 7,3 9,3] - A pertinência aumenta linearmente de 5,3 até 7,3 e depois diminui linearmente até 9,3.
- **Muito Bom**: Função trapezoidal [7,3 9,3 10 10] - A pertinência aumenta linearmente de 7,3 até 9,3. Pertinência total para notas de 9,3 em diante.

Abaixo, apresenta-se sua representação gráfica.

Figura 22: Coeficiente de Rendimento



Fonte: Autora

5.2.4 Média Ponderada (0 a 10)

Representa o valor médio das notas obtidas pelo estudante, considerando os pesos das disciplinas. A estrutura das funções de pertinência segue o mesmo padrão do Coeficiente de Rendimento, uma vez que ambas as variáveis utilizam a mesma escala (0 a 10) e possuem interpretação similar quanto ao desempenho acadêmico. Abaixo, apresenta-se as avaliações dos especialistas:

Tabela 23: Avaliação dos especialistas - Média Ponderada

Especialistas/ Avaliação	**Ruim**	**Regular**	**Bom**	**Muito Bom**
Especialista 1	[0, 4]	]4, 6]	]6, 8]	]8, 10]
Especialista 2	[0, 4]	]3, 6]	]5, 8]	]7, 10]
Especialista 3	[0, 3]	]3, 6]	]6, 8]	]8, 10]
Especialista 4	[0, 4]	]4, 6]	]6, 8]	]8, 10]

Fonte: Autora

Assim, a média ponderada de cada termo foi calculada da seguinte maneira:

- Ruim

$$\frac{4.1 + 4.2 + 4.3 + 3.4}{4 + 4 + 4 + 3} = \frac{36}{15} = 2,4$$

- Regular

$$\frac{2.4 + 4.5 + 4.6}{10} = \frac{52}{10} = 5,2$$

- Bom

$$\frac{1.6 + 4.7 + 4.8}{1 + 4 + 4} = \frac{66}{9} = 7,3$$

- Muito Bom

$$\frac{1.8 + 4.9 + 4.10}{1 + 4 + 4} = \frac{84}{9} = 9,3$$

Assim como no caso do Coeficiente de Rendimento, para a Média Ponderada foi estabelecido um acréscimo e um decréscimo de 2 unidades em torno do valor central, com o objetivo de definir os intervalos correspondentes a cada termo.

Tabela 24: Delimitadores de “Média Ponderada”

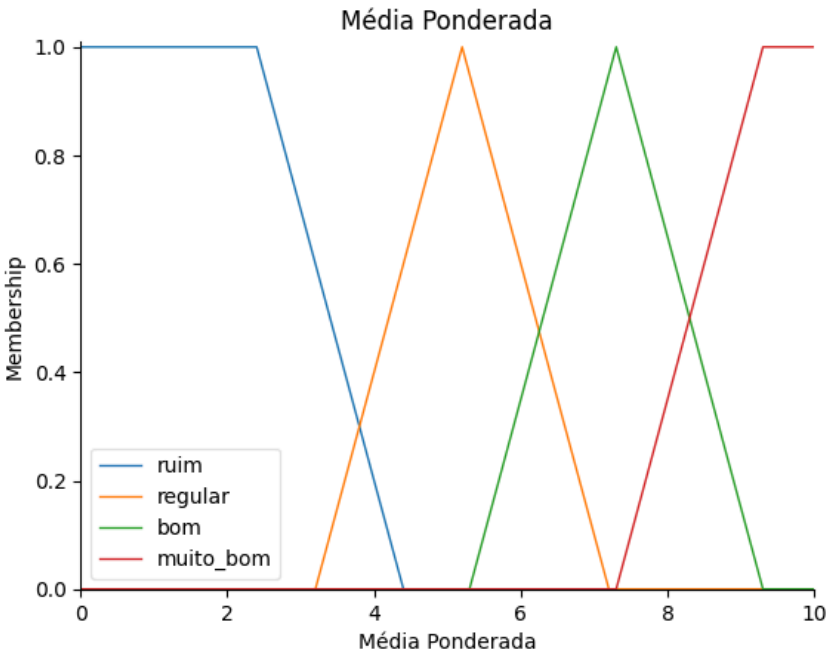
CONJUNTO DE TERMOS	TIPOS	DELIMITADORES
Ruim	Trapezoidal	[0 0 2,4 4,4]
Regular	Triangular	[3,2 5,2 7,3]
Bom	Triangular	[5,3 7,3 9,3]
Muito bom	Trapezoidal	[7,3 9,3 10 10]

Fonte: Autora

- **Ruim**: Função trapezoidal [0 0 2,4 4,4] - Pertinência total para notas de 0 a 2,4, diminuindo linearmente de 2,4 a 4,4.
- **Regular**: Função triangular [3,2 5,2 7,2] - A pertinência aumenta linearmente de 3,2 até 5,2 e depois diminui linearmente até 7,2.
- **Bom**: Função triangular [5,3 7,3 9,3] - A pertinência aumenta linearmente de 5,3 até 7,3 e depois diminui linearmente até 9,3.
- **Muito Bom**: Função trapezoidal [7,3 9,3 10 10] - A pertinência aumenta linearmente de 7,3 até 9,3. Pertinência total para notas de 9,3 em diante.

Na sequência, é apresentada a representação gráfica correspondente.

Figura 23: Média Ponderada



Fonte: Autora

5.2.5 Aproveitamento de Disciplinas (0 a 100%)

Para esta variável, foi definido um universo de discurso de 0 a 100%, representando a proporção de disciplinas em que o estudante obteve aprovação direta, sem necessidade de repetição ou dependência. As avaliações dos especialistas a respeito dessa variável se apresentam na seguinte tabela:

Tabela 25: Avaliação dos especialistas - Aproveitamento de Disciplinas

Especialistas/ Avaliação	**Ruim**	**Regular**	**Bom**
Especialista 1	[0, 30]	]30, 70]	]70, 100]
Especialista 2	[0, 40]	]30, 80]	]80, 100]
Especialista 3	[0, 30]	]30, 70]	]70, 100]
Especialista 4	[0, 30]	]30, 60]	]60, 100]

Fonte: Autora

Assim como nas variáveis anteriores, o pico de cada termo foi estabelecido através do cálculo da média ponderada de cada um deles, apresentados a seguir:

- Ruim

$$\frac{4.10 + 4.20 + 4.30 + 1.40}{4 + 4 + 4 + 1} = \frac{280}{13} = 21,54$$

- Regular

$$\frac{4.40 + 4.50 + 4.60 + 3.70 + 1.80}{4 + 4 + 4 + 3 + 1} = \frac{890}{16} = 55,62$$

- Bom

$$\frac{70 + 3.80 + 4.90 + 4.100}{1 + 3 + 4 + 4} = \frac{1070}{12} = 89,16$$

Considerando que, nesta etapa, passaram a ser utilizados apenas três termos linguísticos “ruim”, “regular” e “bom”, tornou-se necessário ampliar a extensão dos intervalos das funções de pertinência, de modo a manter uma sobreposição adequada entre as categorias. Por essa razão, adotou-se um critério de expansão em torno do valor central de cada termo, com acréscimo e decréscimo de 20 unidades.

Tabela 26: Delimitadores de “Aproveitamento de Disciplinas”

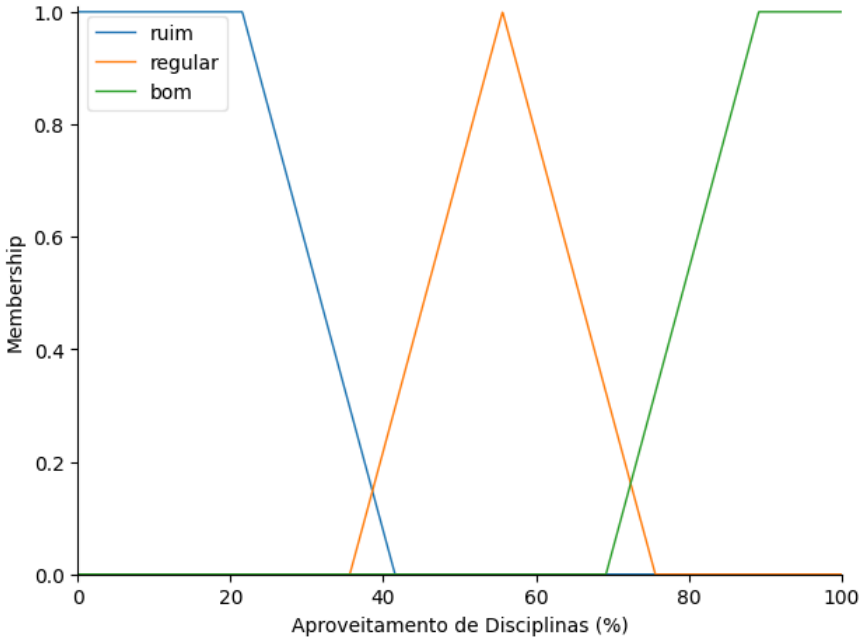
CONJUNTO DE TERMOS	TIPOS	DELIMITADORES
Ruim	Trapezoidal	[0 0 21,54 41,54]
Regular	Triangular	[35,62 55,62 75,62]
Bom	Trapezoidal	[69,16 89,16 100 100]

Fonte: Autora

- **Ruim**: Função trapezoidal [0 0 21,54 41,54] - Representa percentuais entre 0 e 41,54, com adesão máxima entre 0 e 21,54, diminuindo linearmente até 41,54.
- **Regular**: Função triangular [35,62 55,62 75,62] - Abrange percentuais entre 35,62 e 75,62, com adesão máxima em 55,62, aumentando linearmente de 35,62 a 55,62 e diminuindo linearmente de 55,62 a 75,62.
- **Bom**: Função trapezoidal [69,16 89,16 100 100] - Corresponde a percentuais entre 69,16 e 100, com adesão aumentando linearmente de 69,16 a 89,16 e mantendo a adesão máxima de 89,16 a 100.

A seguir, apresenta-se a representação gráfica.

Figura 24: Aproveitamento de Disciplinas



Fonte: Autora

5.2.6 Frequência Média (0 a 100%)

Para esta variável, foi definido um universo de discurso de 0 a 100%, representando o percentual médio de presença do estudante nas disciplinas cursadas, calculado considerando a frequência final em cada disciplina ao longo de todo o curso. As avaliações dos especialistas referentes à frequência são apresentadas na tabela a seguir.

Tabela 27: Avaliação dos especialistas - Frequência Média

Especialistas/ Avaliação	**Ruim**	**Regular**	**Bom**
Especialista 1	[0, 60]	]60, 80]	]80, 100]
Especialista 2	[0, 60]	]50, 80]	]80, 100]
Especialista 3	[0, 60]	]60, 80]	]80, 100]
Especialista 4	[0, 50]	]50, 70]	]70, 100]

Fonte: Autora

O valor de pico de cada termo foi determinado a partir do cálculo da média ponderada correspondente, conforme apresentado a seguir:

- Ruim

$$\frac{4.10 + 4.20 + 4.30 + 4.40 + 4.50 + 3.60}{4 + 4 + 4 + 4 + 4 + 3} = \frac{780}{23} = 33,91$$

- Regular

$$\frac{2.60 + 4.70 + 3.80}{2 + 4 + 3} = \frac{640}{9} = 71,11$$

- Bom

$$\frac{1.80 + 4.90 + 4.100}{1 + 4 + 4} = \frac{840}{9} = 93,33$$

Para a variável frequência, adotou-se o mesmo critério da variável anterior, considerando a utilização de três termos linguísticos “ruim”, “regular” e “bom”. Dessa forma, os intervalos das funções de pertinência também foram definidos por meio da expansão em torno do valor central de cada termo, com acréscimo e decréscimo de 20 unidades.

Tabela 28: Delimitadores de “Frequência Média”

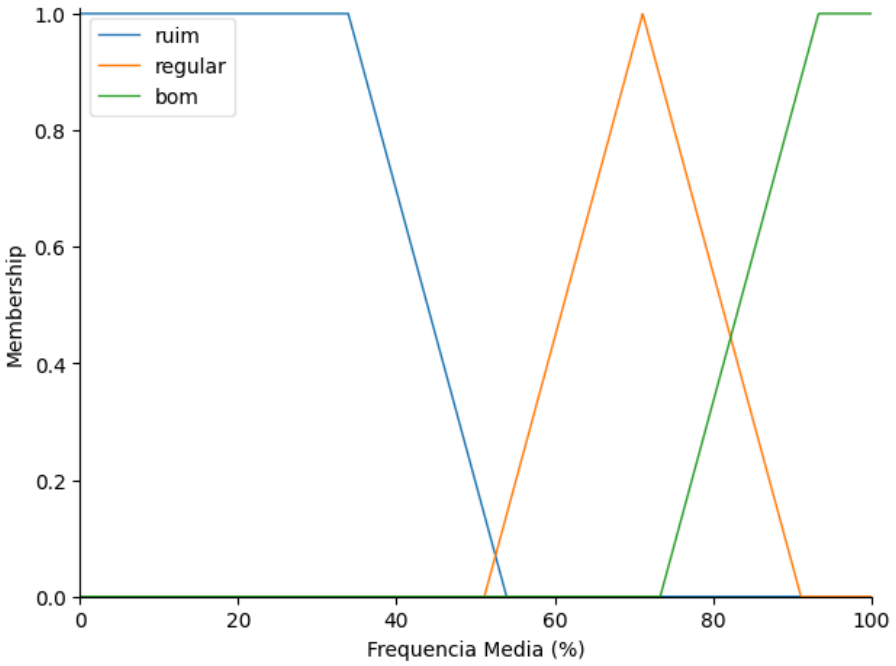
CONJUNTO DE TERMOS	TIPOS	DELIMITADORES
Ruim	Trapezoidal	[0 0 33,91 53,91]
Regular	Triangular	[51,11 71,11 91,11]
Bom	Trapezoidal	[73,33 93,33 100 100]

Fonte: Autora

- **Ruim**: Função trapezoidal [0 0 33,91 53,91] - Representa percentuais entre 0 e 53,91, com adesão máxima entre 0 e 33,91, diminuindo linearmente até 53,91.
- **Regular**: Função triangular [51,11 71,11 91,11] - Abrange percentuais entre 51,11 e 91,11, com adesão máxima em 71,11, aumentando linearmente de 51,11 a 71,11 e diminuindo linearmente de 71,11 a 91,11.
- **Bom**: Função trapezoidal [73,33 93,33 100 100] - Corresponde a percentuais entre 73,33 e 100, com adesão aumentando linearmente de 73,33 a 93,33 e mantendo a adesão máxima a partir de 93,33.

A seguir, é apresentada a representação gráfica correspondente.

Figura 25: Frequência Média



Fonte: Autora

5.2.7 Desempenho Geral do Aluno (Variável de Saída - 0 a 10)

Esta variável representa a saída do sistema *fuzzy*, constituindo a síntese final do desempenho acadêmico do estudante. Diferentemente das variáveis anteriores, que são antecedentes (entradas) do sistema, o Desempenho Geral do Aluno é a variável consequente, obtida através da aplicação das regras de inferência *fuzzy* que combinam as informações das seis variáveis de entrada.

O universo de discurso foi estabelecido de 0 a 10, seguindo uma escala intuitiva onde valores maiores indicam melhor desempenho acadêmico global. Esta variável de saída consolida, através do processamento das regras *fuzzy*, uma avaliação que considera tanto o desempenho no ingresso quanto o desenvolvimento acadêmico durante o curso. As funções de pertinência foram definidas com apenas três categorias linguísticas e foram baseadas nas avaliações especialistas abaixo:

Tabela 29: Avaliação dos especialistas - Desempenho Geral

Especialistas/ Avaliação	Ruim	Regular	Bom
Especialista 1	[0, 4]	]4, 7]	]7, 10]
Especialista 2	[0, 4]	]3, 7]	]7, 10]
Especialista 3	[0, 5]	]5, 8]	]8, 10]
Especialista 4	[0, 4]	]4, 7]	]7, 10]

Fonte: Autora

Para a variável de saída, também foi calculada a média ponderada de cada termo, com o objetivo de determinar os respectivos valores de pico. Os cálculos correspondentes são apresentados a seguir.

- Ruim

$$\frac{4.1 + 4.2 + 4.3 + 4.4 + 1.5}{4 + 4 + 4 + 4 + 1} = \frac{45}{17} = 2,65$$

- Regular

$$\frac{1.4 + 3.5 + 4.6 + 4.7 + 8}{1 + 3 + 4 + 4 + 1} = \frac{79}{13} = 6,08$$

- Bom

$$\frac{3.8 + 4.9 + 4.10}{3 + 4 + 4} = \frac{100}{11} = 9,09$$

Para a variável de saída, referente ao desempenho geral, adotou-se o mesmo critério aplicado às variáveis Coeficiente de Rendimento e Média Ponderada. Assim, os intervalos das funções de pertinência foram definidos a partir de um acréscimo e de um decréscimo de 2 unidades em torno do valor central de cada termo.

Tabela 30: Delimitadores de “Desempenho Geral do Aluno”

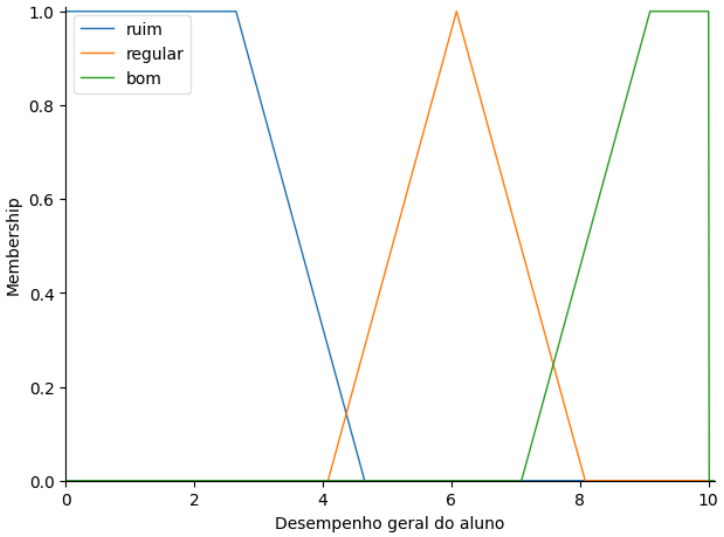
CONJUNTO DE TERMOS	TIPOS	DELIMITADORES
Ruim	Trapezoidal	[0 0 2,65 4,65]
Regular	Triangular	[4,08 6,08 8,08]
Bom	Trapezoidal	[7,09 9,09 10 10]

Fonte: Autora

- **Ruim**: Função trapezoidal [0 0 2,65 4,65] - Desempenho acadêmico global insatisfatório, resultado da combinação de indicadores predominantemente desfavoráveis
- **Regular**: Função triangular [4,08 6,08 8,08] - Desempenho acadêmico mediano, com pico em 6,08, indicando equilíbrio entre aspectos positivos e negativos.
- **Bom**: Função trapezoidal [7,09 9,09 10 10] - Desempenho acadêmico satisfatório, resultante da combinação favorável dos diversos indicadores analisados.

Em seguida, apresenta-se a representação gráfica desta variável.

Figura 26: Desempenho Geral do Aluno



Fonte: Autora

5.3 Regras de Inferência

A etapa de definição das regras de inferência é fundamental para qualquer Sistema Baseado em Regras *Fuzzy*, pois é nela que se estabelece a lógica central que guiará o comportamento do sistema. Em um sistema de inferência *fuzzy*, as regras conectam as variáveis de entrada (antecedentes) às variáveis de saída (consequentes) usando termos linguísticos. A estrutura básica de uma regra *fuzzy* é "SE [antecedente 1] E [antecedente 2] ... ENTÃO [consequente]".

Para determinar o número total de regras necessárias para cobrir todas as combinações possíveis dos termos linguísticos das variáveis de entrada, multiplicamos o número de termos linguísticos definidos para cada variável de entrada.

Neste sistema, as variáveis de entrada e seus respectivos números de termos linguísticos são:

- Classificação no Vestibular: 4 termos ('muito_bom', 'bom', 'regular', 'ruim')
- Nota Final no Vestibular: 4 termos ('ruim', 'regular', 'bom', 'muito_bom')
- Coeficiente de Rendimento (CR): 4 termos ('ruim', 'regular', 'bom', 'muito_bom')
- Média Ponderada: 4 termos ('ruim', 'regular', 'bom', 'muito_bom')
- Aproveitamento de Disciplinas (%): 3 termos ('ruim', 'regular', 'bom')
- Frequência Média (%): 3 termos ('ruim', 'regular', 'bom')

Portanto, o cálculo para o número total de regras que cobririam todas as combinações é dado por:

$$4 \times 4 \times 4 \times 4 \times 3 \times 3 \; = \; 2304$$

Isso significa que, teoricamente, seriam necessárias 2304 regras para considerar cada combinação única dos estados linguísticos de todas as variáveis de entrada.

Mesmo que a lógica descrita nos ajude a definir um sistema de regras completo, na prática, criar todas essas regras nem sempre é o mais eficiente. Isso porque um conjunto de regras excessivamente grande se torna complexo de gerenciar, além de incluir combinações de termos linguísticos que podem ser irrelevantes ou até mesmo nunca acontecer. Além disso, algumas regras podem ser redundantes, levando ao mesmo resultado. Por isso, é mais comum focar na criação de regras que lidam com os cenários mais importantes e frequentes para o problema em questão, como as que definem um desempenho "bom" ou "ruim", em vez de tentar cobrir todas as situações possíveis.

Ao criar um sistema *fuzzy*, o objetivo é capturar o conhecimento especialista ou os padrões importantes nos dados através de um conjunto de regras que seja ao mesmo tempo eficaz e gerenciável.

5.3.1 O Uso dos Operadores "E" (AND) e "OU" (OR) em Regras *Fuzzy*

Em sistemas de inferência *fuzzy*, os operadores lógicos "E" (AND) e "OU" (OR) são usados para combinar as condições dos antecedentes de uma regra. A maneira como eles funcionam é diferente da lógica booleana tradicional (verdadeiro/falso), pois operam com graus de pertinência (valores entre 0 e 1).

As regras definidas neste sistema utilizam principalmente o operador "E" (AND). Por exemplo:

```
      rule1 = ctrl.Rule(
           classification['muito_bom'] & vestibular_score['muito_bom'] &
cr['muito_bom']  &  discipline_success['bom']  &  attendance['bom']  &
weighted_average['muito_bom'],
          overall_performance['bom'])
```

Nesta regra, o operador &, que representa o "E" em scikit-*fuzzy*, exige que todas as condições dos antecedentes sejam consideradas. O grau de ativação dessa regra (o quão "verdadeira" a regra é para um determinado conjunto de entradas) é tipicamente calculado usando o mínimo dos graus de pertinência de cada condição individual. Se qualquer uma das condições tiver um grau de pertinência baixo, a ativação da regra será limitada por esse menor valor. Em essência, o "E" busca o menor valor entre as condições combinadas, refletindo a ideia de que todas as condições devem ser satisfeitas em algum grau.

O operador "OU" (| em scikit-*fuzzy*) é usado quando pelo menos uma das condições dos antecedentes precisa ser verdadeira em algum grau para que a regra seja ativada. A ativação de uma regra com "OU" é tipicamente calculada usando o máximo dos graus de pertinência de cada condição individual. Por exemplo:

```
      rule3 = ctrl.Rule(
      ((classification['ruim'] & vestibular_score['ruim'] & cr['ruim'])
| (discipline_success['ruim'] & attendance['ruim'])) | (cr['regular'] &
weighted_average['regular']),
          overall_performance['ruim'])
```

Nesta regra, se qualquer uma das variáveis de entrada tiver um grau de pertinência significativo no termo 'ruim', essa regra será ativada com um grau igual ao maior desses graus de pertinência. O "OU" busca o maior valor entre as condições combinadas, refletindo a ideia de que a satisfação de qualquer uma das condições contribui para a ativação da regra.

A escolha entre "E" e "OU", ou uma combinação deles, depende da lógica que o especialista quer modelar. Regras com "E" tendem a ser mais restritivas, enquanto regras com "OU" tendem a ser mais permissivas na ativação. A combinação estratégica de "E" e "OU" permite construir um conjunto de regras que represente de forma mais precisa a complexidade do conhecimento especialista ou do sistema que está sendo modelado.

5.3.2 Construção das regras

Para modelar o sistema de inferência *fuzzy* capaz de avaliar o desempenho geral dos alunos, foi desenvolvido um conjunto de 25 regras. A escolha por este número de regras,

embora não exaustiva de todas as combinações possíveis das variáveis de entrada e seus respectivos termos linguísticos, buscou cobrir os cenários de desempenho mais relevantes e representativos para o contexto da avaliação acadêmica, incorporando tanto situações de destaque quanto aquelas que indicam necessidade de atenção. Este conjunto de regras reflete o conhecimento especialista sobre os fatores que influenciam o sucesso discente e serve como base para a tomada de decisão *fuzzy*. As 25 regras criadas são apresentadas a seguir, detalhando as condições antecedentes e o consequente associado a cada uma.

```
                # --- Definição das Regras Fuzzy ---

# Regra 1: Se todas as métricas de entrada são muito boas, o desempenho
geral é bom.
rule1 = ctrl.Rule(
    classification['muito_bom'] & vestibular_score['muito_bom'] &
cr['muito_bom'] & discipline_success['bom'] & attendance['bom'] &
weighted_average['muito_bom'],
    overall_performance['bom'])

# Regra 2: Se as variáveis de ingresso são regulares ou ruins mas o
desempenho no curso é bom, o desempenho geral é bom.
rule2 = ctrl.Rule(
    (classification['regular'] | classification['ruim']) &
(vestibular_score['regular'] | vestibular_score['ruim']) & (cr['bom'] |
cr['muito_bom']) & (weighted_average['bom'] |
weighted_average['muito_bom']) & (discipline_success['bom']),
    overall_performance['bom'])

# Regra 3: Combinações de indicadores fracos (iniciais ou acadêmicos)
levam a desempenho ruim ou regular.
rule3 = ctrl.Rule(
    ((classification['ruim'] & vestibular_score['ruim'] & cr['ruim']) |
(discipline_success['ruim'] & attendance['ruim'])) | (cr['regular'] &
weighted_average['regular']),
    overall_performance['ruim'])

# Regra 4: Se múltiplos indicadores iniciais ou acadêmicos são muito
bons, o desempenho é bom.
rule4 = ctrl.Rule(
  (classification['muito_bom'] | vestibular_score['muito_bom']) &
(cr['muito_bom'] | weighted_average['muito_bom']),
    overall_performance['bom'])
```

```
# Regra 5: Se múltiplos indicadores iniciais ou acadêmicos são ruins, o
desempenho é ruim.
rule5 = ctrl.Rule(
    (cr['ruim'] & weighted_average['ruim'])|(discipline_success['ruim']
& attendance['ruim']),
    overall_performance['ruim'])
# Regra 6: Bom CR, aproveitamento e média ponderada indicam bom
desempenho.
rule6 = ctrl.Rule(
     cr['muito_bom'] & discipline_success['bom'] &
weighted_average['muito_bom'],
    overall_performance['bom'])

# Regra 7: CR ou Média Ponderada ruins indicam desempenho geral ruim.
rule7 = ctrl.Rule(
    cr['ruim'] | weighted_average['ruim'], overall_performance['ruim'])

# Regra 8: bom desempenho inicial com desempenho regular contínuo
resulta em desempenho regular.
rule8 = ctrl.Rule(
    vestibular_score['bom'] & classification['bom'] & (cr['regular'] |
discipline_success['regular']),
    overall_performance['regular'])

# Regra 9: Alta frequência e bons escores contínuos indicam bom
desempenho.
rule9 = ctrl.Rule(
    attendance['bom'] & (cr['bom'] | weighted_average['bom']),
    overall_performance['bom'])

# Regra 10: Baixo aproveitamento e frequência regular/ruim com CR ruim
indicam desempenho ruim.
rule10 = ctrl.Rule(
     discipline_success['ruim'] & (attendance['regular'] |
attendance['ruim']) & cr['ruim'],
    overall_performance['ruim'])

# Regra 11: Boa classificação e escores muito bons indicam bom
desempenho.
rule11 = ctrl.Rule(
    classification['bom'] & vestibular_score['muito_bom'] &
cr['muito_bom'],
    overall_performance['bom'])
```

```
# Regra 12: Início ruim no vestibular com desempenho acadêmico razoável
indica desempenho regular.
rule12 = ctrl.Rule(
    (classification['ruim'] & vestibular_score['ruim']) & (cr['bom'] |
cr['regular']) & (weighted_average['bom'] |
weighted_average['regular']),
    overall_performance['regular'])

# Regra 13: Boa frequência e CR ou bom aproveitamento indicam bom
desempenho.
rule13 = ctrl.Rule(
  (attendance['bom'] & cr['bom']) | discipline_success['bom'],
    overall_performance['bom'])

# Regra 14: Desempenho médio/regular consistente em várias variáveis
indica desempenho regular.
rule14 = ctrl.Rule(
    weighted_average['regular'] & attendance['regular'] & cr['regular']
& discipline_success['regular'],
    overall_performance['regular'])

# Regra 15: Forte início e boa frequência com bons escores contínuos
indicam bom desempenho.
rule15 = ctrl.Rule(
 classification['muito_bom'] & vestibular_score['bom'] &
attendance['bom'] & (cr['bom'] | weighted_average['bom']),
    overall_performance['bom'])

# Regra 16: Forte desempenho inicial mas desempenho contínuo fraco
indica desempenho ruim.
rule16 = ctrl.Rule(
    (classification['muito_bom'] | classification['bom']) &
(vestibular_score['muito_bom'] | vestibular_score['bom']) & (cr['ruim']
| weighted_average['ruim']),
    overall_performance['ruim'])

# Regra 17: Desempenho inicial médio mas desempenho contínuo forte
indica bom desempenho.
rule17 = ctrl.Rule(
      (classification['regular'] | classification['bom']) &
(vestibular_score['regular'] | vestibular_score['bom']) &
(cr['muito_bom'] | weighted_average['muito_bom']),
```

```
      overall_performance['bom'])

# Regra 18: Bons escores acadêmicos mas baixo engajamento (disciplinas,
frequência) indicam desempenho regular.
rule18 = ctrl.Rule(
    cr['bom'] & weighted_average['bom'] & (discipline_success['ruim'] |
attendance['ruim']),
    overall_performance['regular'])

#  Regra 19: Escores acadêmicos fracos/regulares mas bom engajamento
indicam desempenho regular.
rule19 = ctrl.Rule(
    (cr['ruim'] | cr['regular']) & (weighted_average['ruim'] |
weighted_average['regular']) &
(attendance['regular']|attendance['bom']) &
(discipline_success['bom']|discipline_success['regular']) ,
    overall_performance['regular'])

#  Regra 20: Desempenho regular em CR e aproveitamento de disciplinas
indica desempenho regular.
rule20 = ctrl.Rule(
     cr['regular'] & discipline_success['regular']
    overall_performance['regular'])

#  Regra 21: CR muito bom apesar de um início fraco indica desempenho
bom (melhora significativa).
rule21 = ctrl.Rule(
    cr['muito_bom'] & (classification['regular'] |
classification['ruim']) & (vestibular_score['regular'] |
vestibular_score['ruim']),
    overall_performance['bom'])

#  Regra 22: CR ou Média Ponderada ruins, mesmo com aproveitamento
regular, indicam desempenho ruim.
rule22 = ctrl.Rule(
    (cr['ruim'] | weighted_average['ruim']) &
discipline_success['regular'],
    overall_performance['ruim'])

# Regra 23: Boa frequência e aproveitamento e CR/Média bom ou regular,
indicam bom desempenho (engajamento positivo).
rule23 = ctrl.Rule(
```

```
    attendance['bom'] & discipline_success['bom'] & (cr['regular'] |
cr['bom']),
    overall_performance['bom'])

# Regra 24: Boa classificação e nota no vestibular, mas desempenho
contínuo regular, indicam desempenho regular.
rule24 = ctrl.Rule(
    (classification['bom'] | classification['muito_bom']) &
(vestibular_score['bom'] | vestibular_score['muito_bom']) &
(cr['regular'] | weighted_average['regular']),
    overall_performance['regular'])

# Regra 25: Bom CR e Média Ponderada indicam bom desempenho.
rule25 = ctrl.Rule(
    (cr['bom'] | cr['muito_bom']) & (weighted_average['bom'] |
weighted_average['muito_bom']),
    overall_performance['bom'])
```

5.4 Aplicação do modelo e análise dos resultados

A partir do conjunto de 25 regras *fuzzy* cuidadosamente elaboradas, foi possível construir o modelo de inferência *fuzzy* capaz de avaliar o desempenho geral dos alunos. Este modelo foi então aplicado individualmente a cada um dos 111 conjuntos de dados correspondentes aos alunos incluídos na análise. Para cada aluno, o sistema *fuzzy* processou as variáveis de entrada e gerou um valor numérico correspondente à variável de saída "desempenho geral do aluno", representando o desempenho individual. Esse valor foi posteriormente convertido em uma classificação linguística, conforme apresentado na tabela a seguir.

Tabela 31: Desempenho Geral por Aluno

Aluno	Cotas	Desempenho Geral Fuzzy	Desempenho Linguístico
1	Ensino Público	7,96	bom
2	Preto, pardo ou indígena	8,96	bom
3	Ensino Público	8,96	bom
4	Não	8,96	bom
5	Ensino Público	8,96	bom

6	Preto, pardo ou indígena	8,96	bom
7	Não	7,37	regular
8	Ensino Público	8,96	bom
9	Não	8,96	bom
10	Não	8,96	bom
11	Não	6,44	regular
12	Preto, pardo ou indígena	3,65	ruim
13	Preto, pardo ou indígena	6,28	regular
14	Não	8,96	bom
15	Ensino Público	8,96	bom
16	Não	3,70	ruim
17	Ensino Público	2,98	ruim
18	Ensino Público	8,96	bom
19	Não	8,96	bom
20	Ensino Público	8,96	bom
21	Não	8,96	bom
22	Não	5,64	regular
23	Ensino Público	3,17	ruim
24	Preto, pardo ou indígena	8,96	bom
25	Não	5,93	regular
26	Não	8,96	bom
27	Não	8,96	bom
28	Não	8,96	bom
29	Ensino Público	3,84	ruim
30	Preto, pardo ou indígena	8,96	bom
31	Não	8,96	bom

32	Ensino Público	5,64	regular
33	Preto, pardo ou indígena	8,96	bom
34	Não	5,30	regular
35	Não	8,96	bom
36	Preto, pardo ou indígena	6,20	regular
37	Não	8,96	bom
38	Preto, pardo ou indígena	8,96	bom
39	Ensino Público	6,63	regular
40	Não	3,96	ruim
41	Preto, pardo ou indígena	2,98	ruim
42	Ensino Público	4,36	ruim
43	Não	8,96	bom
44	Não	8,95	bom
45	Ensino Público	8,96	bom
46	Ensino Público	8,96	bom
47	Não	8,96	bom
48	Preto, pardo ou indígena	5,36	regular
49	Ensino Público	8,96	bom
50	Ensino Público	8,96	bom
51	Não	8,96	bom
52	Não	8,96	bom
53	Ensino Público	3,03	ruim
54	Não	5,88	regular
55	Ensino Público	8,96	bom
56	Não	4,37	regular
57	Ensino Público	5,12	regular

58	Não	8,96	bom
59	Não	5,60	regular
60	Não	6,77	regular
61	Ensino Público	8,96	bom
62	Não	8,96	bom
63	Ensino Público	6,00	regular
64	Não	8,81	bom
65	Preto, pardo ou indígena	3,61	ruim
66	Não	4,16	ruim
67	Não	8,96	bom
68	Ensino Público	8,96	bom
69	Não	8,96	bom
70	Ensino Público	8,96	bom
71	Ensino Público	8,96	bom
72	Não	8,96	bom
73	Não	8,96	bom
74	Ensino Público	8,96	bom
75	Ensino Público	8,96	bom
76	Não	8,96	bom
77	Preto, pardo ou indígena	8,96	bom
78	Não	8,96	bom
79	Ensino Público	8,96	bom
80	Ensino Público	8,96	bom
81	Não	8,96	bom
82	Não	8,96	bom
83	Preto, pardo ou indígena	8,96	bom

84	Preto, pardo ou indígena	8,96	bom
85	Não	8,96	bom
86	Ensino Público	8,96	bom
87	Ensino Público	8,96	bom
88	Não	8,96	bom
89	Não	8,96	bom
90	Não	8,96	bom
91	Não	8,96	bom
92	Não	6,45	regular
93	Não	8,96	bom
94	Ensino Público	8,96	bom
95	Ensino Público	8,96	bom
96	Ensino Público	8,96	bom
97	Preto, pardo ou indígena	8,96	bom
98	Não	8,96	bom
99	Não	8,96	bom
100	Ensino Público	8,95	bom
101	Ensino Público	8,96	bom
102	Não	8,96	bom
103	Ensino Público	8,96	bom
104	Não	8,96	bom
105	Não	8,96	bom
106	Não	8,96	bom
107	Ensino Público	8,96	bom
108	Não	8,96	bom
109	Não	8,96	bom

110	Não	8,96	bom
111	Ensino Público	8,96	bom

Fonte: Autora

Posteriormente, para facilitar a análise comparativa, os resultados do desempenho individual foram agrupados e analisados separadamente para os grupos de alunos cotistas e não cotistas.

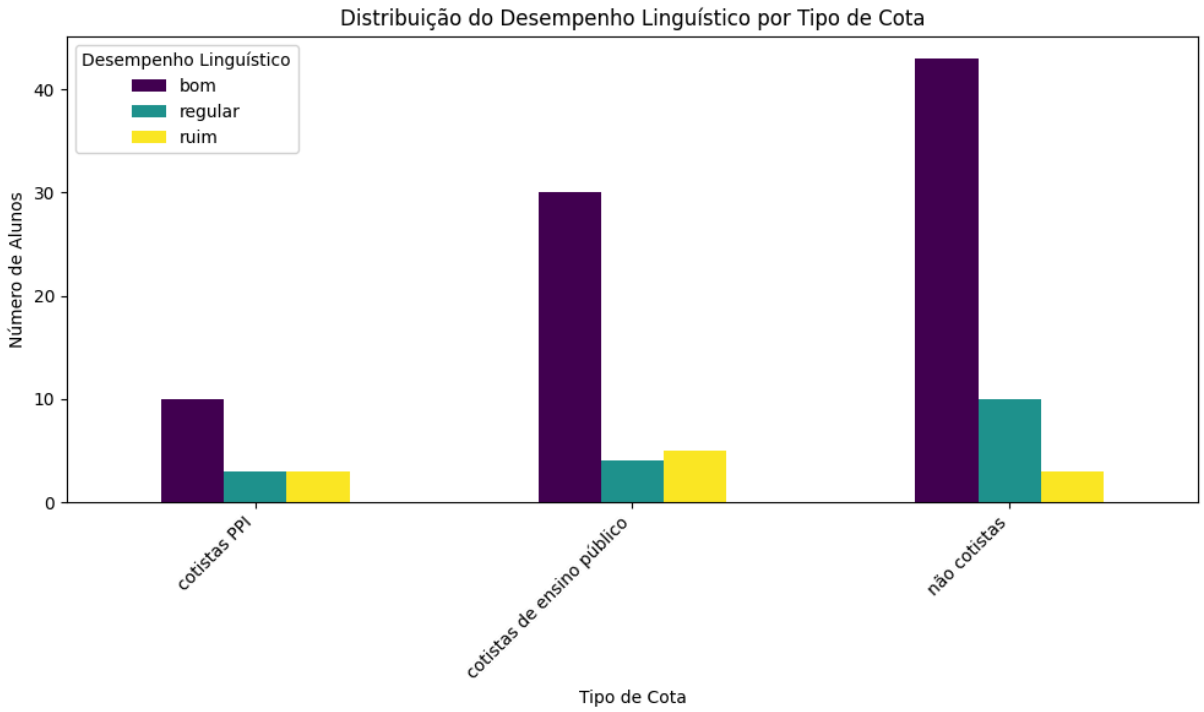
Tabela 32: Desempenho por cotas

Grupo/ Desempenho Geral Linguístico	Bom	Regular	Ruim	Total
PPI	10	3	3	16
Ensino Público	30	4	5	39
Não Cotistas	43	9	4	56

Fonte: Autora

A seguir, apresenta-se um gráfico para a visualização dos resultados obtidos.

Figura 27: Comparação Desempenho Geral Linguístico



Fonte: Autora

Para uma compreensão mais detalhada da distribuição do desempenho, foi conduzida uma análise percentual de cada grupo em relação às categorias "Bom", "Regular" e "Ruim". Essa abordagem permitiu identificar a proporção de alunos em cada nível de performance, revelando as tendências e as diferenças entre os grupos.

Tabela 33: Desempenho Percentual por Grupo

Grupo/ Desempenho Geral Linguístico (%)	Bom (%)	Regular (%)	Ruim (%)	Total de Alunos
PPI	62,5%	18,75%	18,75%	16
Ensino Público	76,92%	10,26%	12,82%	39
Não Cotistas	76,79%	16,07%	7,14%	56

Fonte: Autora

A análise da distribuição percentual do desempenho geral linguístico revela padrões de performance distintos entre os grupos. De forma geral, observa-se que a maioria dos alunos, independentemente de sua origem, foi classificada na categoria "Bom". O grupo de Ensino Público apresentou a maior proporção de alunos nessa categoria, com 76,92%, seguido de perto pelos Não Cotistas (76,79%) e, em seguida, pelo grupo Preto, Pardo ou Indígena (PPI), com 62,5%. Essa proximidade nas porcentagens sugere que, embora o número absoluto de alunos com bom desempenho seja diferente, a proporção de alunos com boa performance é elevada nos três grupos. Isso indica que a maioria dos estudantes, independentemente de seu perfil, foi classificada favoravelmente pelo sistema *fuzzy*.

Em relação às outras categorias, o grupo PPI apresenta uma maior concentração nas categorias “regular” e “ruim”, totalizando 37,5% dos alunos, enquanto esse percentual é menor para os alunos de ensino público (23,08%) e, sobretudo, entre os não cotistas (23,21%). Tal distribuição nos indica que, mesmo após o ingresso na universidade, persistem diferenças estruturais que, de alguma forma, impactam o desempenho acadêmico desses estudantes.

O grupo PPI apresenta a maior proporção de alunos classificados na categoria “ruim” quando comparado aos demais grupos. Enquanto nos grupos de ensino público e de não cotistas essa classificação aparece de forma mais limitada, entre os alunos PPI ela se mostra mais frequente, indicando a presença de um número maior de estudantes com maiores dificuldades acadêmicas. Essa diferença sugere que esses alunos enfrentam desafios

adicionais ao longo da trajetória universitária, possivelmente relacionados a desigualdades anteriores ao ingresso e a menor acesso a recursos de apoio acadêmico.

5.5 Validação

A validação do modelo proposto foi realizada a partir da análise de três especialistas. Cada especialista avaliou individualmente a tabela bruta dos dados, sem qualquer identificação dos alunos, considerando as seis variáveis de entrada do sistema: Classificação no Vestibular, Nota final no Vestibular, Coeficiente de Rendimento (CR), Média Ponderada, Aproveitamento de Disciplinas e Frequência.

Com base em sua análise individual e em seus próprios critérios de julgamento, cada especialista classificou aluno por aluno nas categorias linguísticas “ruim”, “regular” e “bom”, de acordo com sua percepção do desempenho acadêmico global. Em seguida, a classificação final atribuída a cada aluno foi definida por meio da regra de maioria, ou seja, considerando a categoria mais frequentemente indicada entre os três especialistas.

Essa média foi então utilizada como referência para a validação do modelo *fuzzy*, por meio da comparação com os resultados gerados pelo sistema, permitindo avaliar o grau de concordância entre o julgamento dos especialistas e o resultado produzido pelo modelo.

Na sequência, são apresentadas as análises realizadas pelos especialistas, reunidas e organizadas de acordo com os grupos: Cotistas PPI, Cotistas de Ensino Público e Não Cotistas, seguidas da média resultante.

Tabela 34: Resultados Validação - Especialista 1

Grupo/ Desempenho Geral Linguístico	Bom	Regular	Ruim	Total
PPI	10	1	5	16
Ensino Público	29	3	7	39
Não Cotistas	42	5	9	56

Fonte: Autora

Tabela 35: Resultados Validação - Especialista 2

Grupo/ Desempenho Geral Linguístico	Bom	Regular	Ruim	Total
PPI	5	10	1	16
Ensino Público	29	7	3	39
Não Cotistas	41	15	0	56

Fonte: Autora

Tabela 36: Resultados Validação - Especialista 3

Grupo/ Desempenho Geral Linguístico	Bom	Regular	Ruim	Total
PPI	10	3	3	16
Ensino Público	29	5	5	39
Não Cotistas	42	11	3	56

Fonte: Autora

Tabela 37: Resultados Validação - Média

Grupo/ Desempenho Geral Linguístico	Bom	Regular	Ruim	Total
PPI	10	3	3	16
Ensino Público	29	5	5	39
Não Cotistas	42	11	3	56

Fonte: Autora

A comparação entre os resultados gerados pelo modelo *fuzzy* e aqueles obtidos a partir da média da validação mostra um alto nível de concordância entre o sistema e a avaliação dos especialistas. De modo geral, as distribuições nas categorias “bom”, “regular” e “ruim” são muito próximas nos três grupos analisados, indicando que o modelo conseguiu reproduzir de forma consistente o julgamento humano.

No grupo PPI, a concordância é total, uma vez que tanto o modelo quanto a média da validação apresentam exatamente os mesmos resultados: 10 alunos classificados como “bom” (62,5%), 3 como “regular” (18,75%) e 3 como “ruim” (18,75%). Esse resultado nos mostra que, para esse grupo, o modelo captou muito bem os critérios utilizados pelos especialistas na avaliação do desempenho geral.

Para o grupo de ensino público, as diferenças observadas são pequenas. O modelo classificou 30 dos alunos como “bom” (76,92%) e 4 alunos como “regular” (10,26%),

enquanto a validação apontou 29 alunos como “bom” (74,36%) e 3 como “regular” (12,82%), mantendo-se idêntica a proporção de alunos classificados como “ruim” em ambas as abordagens (12,82%, correspondentes a 5 alunos). Essa variação percentual, de aproximadamente 2,56% entre as categorias “bom” e “regular”, indica que o modelo tende a classificar ligeiramente mais alunos como “bom”, sem alterar de forma significativa o perfil geral do grupo.

Situação semelhante é observada no grupo de não cotistas, no qual as diferenças entre o modelo e a validação também são pequenas. O modelo classificou 43 alunos como “bom” (76,79%), enquanto a validação indicou 42 alunos nessa categoria (75%), o que representa uma diferença de apenas um aluno, correspondente a 1,79%. Em contrapartida, a validação apresentou um número maior de alunos classificados como “regular”, passando de 9 alunos (16,07%) no modelo para 11 alunos (19,64%), o que representa uma diferença de 2 alunos, equivalente a 3,57%, enquanto a classificação “ruim” foi ligeiramente menor, reduzindo de 4 alunos (7,14%) no modelo para 3 alunos (5,36%) na validação, uma diferença de 1,78%.

De forma geral, as diferenças entre os resultados são pontuais e não comprometem a interpretação dos dados. A proximidade entre as classificações produzidas pelo modelo e aquelas obtidas na validação reforça que o sistema *fuzzy* proposto é adequado para representar o desempenho acadêmico dos alunos.

CONSIDERAÇÕES FINAIS

A matemática tradicional, com sua lógica binária de 0 ou 1, "verdadeiro ou falso", frequentemente se mostra limitada para modelar situações complexas do mundo real, onde a incerteza e a imprecisão estão presentes. É aqui que a lógica *fuzzy* se destaca. Diferente da lógica clássica, a lógica *fuzzy* introduz o conceito de graus de pertinência, permitindo que uma variável pertença parcialmente a um ou mais conjuntos.

Os conjuntos *fuzzy* servem como base para os Sistemas Baseados em Regras *Fuzzy* (SBRF). A aplicabilidade desses sistemas é ampla e diversa, sendo utilizados em diversos campos como controle automático de sistemas, análise financeira, medicina para diagnóstico de doenças e engenharia para otimização de processos.

Os SBRF oferecem uma abordagem de classificação mais flexível que a análise estatística tradicional, especialmente quando os critérios de classificação não são absolutos. Enquanto os métodos estatísticos convencionais estabelecem fronteiras rígidas entre categorias, os sistemas *fuzzy* permitem que um elemento pertença a diferentes categorias com graus variados de pertencimento, refletindo melhor a natureza gradual de muitos fenômenos reais.

Esta pesquisa buscou uma compreensão aprofundada do desempenho de estudantes de diferentes perfis do curso de matemática da UNESP, campus de Bauru, utilizando uma abordagem de duas frentes: a análise estatística e a modelagem por meio de um Sistema Baseado em Regras *Fuzzy*. Os resultados de ambas as metodologias, embora obtidos de maneiras distintas, convergem para uma série de conclusões importantes e complementares.

A análise estatística revelou que as diferenças mais significativas entre os grupos (Não Cotistas, Cotistas de Ensino Público e Cotistas PPI) se concentram no processo seletivo inicial. Em particular, o grupo PPI demonstrou uma classificação de vestibular significativamente inferior em comparação aos Não Cotistas. No entanto, a análise subsequente do desempenho acadêmico contínuo (CR, Média Ponderada, entre outros) não encontrou diferenças estatisticamente significativas entre os grupos. Isso sugere que, apesar das disparidades iniciais na entrada, o desempenho dos estudantes ao longo do curso tende a se nivelar.

A análise do Sistema Baseado em Regras *Fuzzy* forneceu uma perspectiva mais detalhada da distribuição de desempenho, além da comparação de médias ou distribuições e categorizando os alunos com base em múltiplos critérios simultaneamente. A maioria dos estudantes, independentemente do grupo, foi classificada na categoria de desempenho "Bom", o que valida a conclusão estatística de que há um nivelamento de desempenho no curso.

Em síntese, as análises se complementam. A estatística mostra que os grupos, ao longo do curso, se nivelam em termos de desempenho geral. Por outro lado, o SBRF nos permite enxergar que, dentro dessa performance geral similar, ainda existem variações na distribuição de desempenho. As conclusões combinadas de ambas as metodologias reforçam a complexidade do tema e a importância de ir além das métricas simples. A aplicação do SBRF mostrou-se uma ferramenta valiosa para classificar o desempenho dos alunos e identificar áreas de vulnerabilidade que podem não ser evidentes apenas com a análise estatística.

Os resultados obtidos na etapa de validação indicam que o modelo *fuzzy* proposto apresentou elevado grau de concordância com a avaliação realizada pelos especialistas. De modo geral, as classificações geradas pelo sistema mostraram-se bastante próximas àquelas definidas a partir da média entre os três avaliadores, evidenciando a capacidade do modelo em representar, de forma consistente, o julgamento humano sobre o desempenho acadêmico dos alunos.

Pesquisas que analisam o desempenho acadêmico a partir dos diferentes grupos de ingresso, especialmente no contexto das políticas de cotas, são fundamentais para uma maior compreensão dos efeitos dessas ações afirmativas, indo além de análises baseadas apenas em médias ou indicadores isolados. Nesse sentido, a utilização de sistemas *fuzzy* mostra-se particularmente adequada, pois permite lidar com a complexidade e a subjetividade inerentes ao desempenho acadêmico, incorporando múltiplas variáveis e termos linguísticos que se aproximam do raciocínio humano. Diferentemente de métodos estritamente estatísticos ou determinísticos, o modelo *fuzzy* possibilita capturar zonas de transição entre categorias, oferecendo uma análise mais sensível às nuances dos dados.

Dessa forma, os resultados obtidos assumem especial relevância no contexto do curso de Licenciatura em Matemática da UNESP de Bauru, ao permitir uma compreensão detalhada das características e similaridades no desempenho acadêmico dos diferentes grupos. Essa análise fornece subsídios importantes para orientar políticas internas do curso, especialmente

no que se refere a ações de acompanhamento, permanência e apoio pedagógico. Além disso, a metodologia adotada e os resultados apresentados podem servir como referência para a replicação do estudo em outros cursos e contextos institucionais.

Dando continuidade aos estudos desta dissertação, pretende-se aprofundar a pesquisa no doutorado, ampliando a análise do desempenho acadêmico para diferentes cursos da UNESP e incorporando novas variáveis relacionadas à permanência e à evasão. Serão aplicadas e comparadas abordagens baseadas em lógica *fuzzy*, K-Means e *Fuzzy* C-Means, com o objetivo de analisar a organização e a segmentação dos estudantes segundo seus perfis acadêmicos, desenvolver modelos de previsão do desempenho nos anos finais e investigar o impacto do sistema de cotas. Busca-se, dessa forma, produzir evidências que possam subsidiar estratégias institucionais de acompanhamento, permanência e apoio acadêmico.

REFERÊNCIAS BIBLIOGRÁFICAS

ANDERSON, B. D. O.; MOORE, J. B. **Optimal filtering**. Englewood Cliffs: Prentice-Hall, 1979.

BARRANTES, A. C. **Sistema de Inferência Fuzzy Aplicado na Avaliação Discente**. Trabalho de Conclusão de Curso (Licenciatura em Matemática) – Instituto Federal de São Paulo, São Paulo, 2011.

BARROS, L.; BASSANEZI, R. C. T**ópicos de lógica *fuzzy* e biomatemática**. Campinas: UNICAMP, 2006. Coleção IMECC. Textos Didáticos, v. 5.

BARROS, P. H. A.; LANZILLOTTI, R. S. Análise de desempenho dos estudantes mediante modelo intuicionista Fuzzy. In: **Congresso Brasileiro de Educação**. Anais. 2023.

BERTSEKAS, D. P. **Dynamic programming and stochastic control**. New York: Academic Press, 1976. Mathematics in Science and Engineering, v. 125.

BRASIL. Decreto nº 7.234, de 19 de julho de 2010. Dispõe sobre o Programa Nacional de Assistência Estudantil - PNAES. **Diário Oficial da União**, Brasília, DF, 20 jul. 2010. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2010/decreto/d7234.htm. Acesso em: 15 jan. 2025.

CASTILLO, O.; MELIN, P. **Type-2 *fuzzy* logic**: theory and applications. Berlin: Springer-Verlag, 2008.

CORCOLL-SPINA, C. O. **Lógica *fuzzy*:** reflexões que contribuem para a questão da subjetividade na construção do conhecimento matemático. 2010. Tese (Doutorado) - Universidade de São Paulo, São Paulo, 2010.

DORF, R. C.; BISHOP, R. H. **Sistemas de controle modernos**. 12. ed. Rio de Janeiro: LTC, 2011.

D'OTTAVIANO, I. M. L.; FEITOSA, H. A. **História da lógica e o surgimento das lógicas não clássicas**. Campinas: Centro de Lógica, Epistemologia e História da Ciência – CLE/UNICAMP, 2003.

DOYLE, J. C.; STEIN, G. Multivariable feedback design: concept for a classical/modern synthesis. **IEEE Transactions on Automatic Control**, v. 26, n. 1, p. 4-16, 1981.

FEITOSA, H. A.; PAULOVICH, L. **Um prelúdio à lógica**. São Paulo: UNESP, 2005.

GALHARDO, E.; VASCONCELOS, M. S.; FREI, F.; RODRIGUES, E. B. Desempenho acadêmico e frequência dos estudantes ingressantes pelo Programa de Inclusão da UNESP. **Avaliação: Revista da Avaliação da Educação Superior**, Campinas, v. 25, n. 3, p. 701-725, 2020.

GASPAR, J. S et al. **Introdução à análise de dados em saúde com Python**. Belo Horizonte, MG: Biblioteca J. Baeta Vianna: Universidade Federal de Minas Gerais, 2023. Disponível em: https://docs.bvsalud.org/biblioref/2023/06/1437637/introducao-a-analise-de-dados-em-saude-com-python-ciia-saude.pdf. Acesso em: 10. jul. 2025.

GOLZIO, A. C; PUERTA-DÍAZ, M. A Model for Analysis of Environmental Accidents Based on Fuzzy Logic: Case Study: Exxon Valdez Oil Spill. In: **Data and Information in Online Environments**. Springer, 2021. p. 313-327. DOI: 10.1007/978-3-030-77417-2_23.

GOMIDE, F. A. C.; GUDWIN, R. R. Modelagem, controle, sistemas e lógica *fuzzy*. **SBA Controle & Automação**, v. 4, n. 3, 1994.

HOLTZMAN, J. M. **Nonlinear system theory**. Englewood Cliffs: Prentice Hall, 1970.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Censo da educação superior 2023**: microdados. Brasília: INEP, 2024. Disponível em: https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/censo-da-educacao-superior. Acesso em: 20 jun. 2025.

KLIR, G. J.; YUAN, B. ***Fuzzy* sets and *fuzzy* logic**: theory and applications. Upper Saddle River: Prentice Hall, 1995.

KONAR, A. **Artificial intelligence and soft computing**: behavioral and cognitive modeling of the human brain. Boca Raton: CRC Press, 2000.

LIMA, H. P. **Uma abordagem para construção de sistemas *fuzzy* baseados em regras integrando conhecimento de especialistas e extraído de dados**. 2015. Dissertação (Mestrado) - Universidade Federal de São Carlos, São Carlos, 2015.

LUGER, G. F. **Artificial intelligence**: structures and strategies for complex problem solving. 5. ed. Boston: Pearson Education/Addison Wesley, 2005.

ŁUKASIEWICZ, J.; TARSKI, A. Untersuchungen über den Aussagenkalkül. **Comptes Rendus des Séances de la Société des Sciences et des Lettres de Varsovie, Classe III**, v. 23, p. 30-50, 1930.

MAMDANI, E.; ASSILIAN, S. An experiment in linguistic synthesis with a *fuzzy* logic controller. **International Journal of Man-Machine Studies**, v. 7, n. 1, p. 1-13, jan. 1975.

MARRO, A. A.; SOUZA, A. M. C.; CAVALCANTE, E. R. S.; BEZERRA, G. S.; NUNES, R. O. **Lógica *fuzzy***: conceitos e aplicações. Natal: Departamento de Informática e Matemática Aplicada (DIMAp), Universidade Federal do Rio Grande do Norte (UFRN), 2010.

MCNEILL, F. M.; THRO, E. ***Fuzzy* logic**: a practical approach. Boston: AP Professional/Academic Press, 1994.

MENDES JUNIOR, A. A. F. Uma análise da progressão dos alunos cotistas sob a primeira ação afirmativa brasileira no ensino superior: o caso da Universidade do Estado do Rio de Janeiro (UERJ). **Ensaio: Avaliação e Políticas Públicas em Educação**, Rio de Janeiro, v. 22, n. 82, p. 31-56, jan./mar. 2014. Disponível em: http://www.scielo.br/pdf/ensaio/v22n82/a03v22n82.pdf. Acesso em: 25 jan. 2024.

MENIN, M. S. S. et al. Representações de estudantes universitários sobre alunos cotistas: confronto de valores (UNESP/Marília). **Educação e Pesquisa**, São Paulo, v. 34, 2008.

NERI, M. (org.). **Você no mercado de trabalho**. Rio de Janeiro: FGV, 2008. 62 p.

NEVES, E. P.; DUARTE, M. A. Q.; ALVARADO, F. V. Sistema baseado em regras *fuzzy* para avaliação da qualidade da água. **C.Q.D. - Revista Eletrônica Paulista de Matemática**, Bauru, v. 14, 2019. Disponível em: https://sistemas.fc.unesp.br/ojs/index.php/revistacqd/article/view/177. Acesso em: 24 jan. 2024.

SAGE, A. P.; WHITE, C. C. **Optimum systems control**. 2. ed. Englewood Cliffs: Prentice-Hall, 1977.

SHAW, I. S.; SIMÕES, M. G. **Controle e modelagem fuzzy**. São Paulo: Edgard Blucher; FAPESP, 1999.

SILVA, G. C. T.; LUNETTA, C.; PEIXOTO, M. S. Uma abordagem fuzzy para o desempenho escolar no processo ensino-aprendizagem. In: SILVA, G. C. T. (Org.). **Tecnologias Educacionais e Inovações Pedagógicas**: perspectivas e desafios na Matemática. São Paulo: Editora Científica Digital. v. 1, p. 129-144. 2023

SOUZA, M. A. Desempenho dos candidatos no vestibular e o sistema de cotas na UERJ. **Ensaio: Avaliação e Políticas Públicas em Educação**, Rio de Janeiro, v. 20, n. 77, p. 701-724, out./dez. 2012.

TAKÁCS, M. **Approximate reasoning in *fuzzy* systems based on pseudo-analysis and uninorm residuum**. Edited by Bernard de Baets, János Fodor. Gent: Academia Press, 2004.

VASCONCELOS, M. S.; GALHARDO, E. O Programa de inclusão na UNESP: valores, contradições e ações afirmativas. **Revista Ibero-Americana de Estudos em Educação**, Araraquara, v. 11, n. esp. 1, p. 285-306, 2016.

WILGES, B.; MATEUS, G. P.; NASSAR, S. M.; BASTOS, R. C. Avaliação da aprendizagem por meio de lógica de fuzzy validado por uma Árvore de Decisão ID3. **CINTED-UFRGS Novas Tecnologias na Educação**, Porto Alegre, v. 8, n. 3, dez. 2010.

ZADEH, L. A. *Fuzzy* sets. **Information and Control**, v. 8, n. 3, p. 338-353, jun. 1965.

ZADEH, L. A. Outline of a new approach to the analysis of complex systems and decision processes. **IEEE Transactions on Systems, Man, and Cybernetics**, v. 3, n. 1, p. 28-44, 1973.

ZAGO, N. Do acesso à permanência no ensino superior: percursos de estudantes universitários de camadas populares. **Revista Brasileira de Educação**, Rio de Janeiro, v. 11, n. 32, p. 226-237, ago. 2006.

Apêndice A - Códigos Análises Estatísticas

- Estatísticas descritivas

```
import pandas as pd

colunas_analise = ['CR', 'Nota Final Vestibular', 'Classificação
Vestibular', 'Média Ponderada', 'Aproveitamento Disciplinas',
'Frequência Média']

print("Estatísticas para a tabela ncota:")
mean_ncota = ncota[colunas_analise].mean()
median_ncota = ncota[colunas_analise].median()
mode_ncota = ncota[colunas_analise].mode().iloc[0]
stats_ncota = pd.DataFrame({
    'mean': mean_ncota,
    'median': median_ncota,
    'mode': mode_ncota
})
display(stats_ncota)

print("\nEstatísticas para a tabela cota:")
mean_cota = cota[colunas_analise].mean()
median_cota = cota[colunas_analise].median()
mode_cota = cota[colunas_analise].mode().iloc[0]
stats_cota = pd.DataFrame({
    'mean': mean_cota,
    'median': median_cota,
    'mode': mode_cota
})
display(stats_cota)

colunas_analise_variancia = ['CR', 'Média Ponderada', 'Nota Final
Vestibular', 'Aproveitamento Disciplinas', 'Frequência Média']

print("\nVariância e Desvio Padrão para a tabela ncota:")
variancia_ncota = ncota[colunas_analise_variancia].var()
desvio_padrao_ncota = ncota[colunas_analise_variancia].std()
stats_variancia_std_ncota = pd.DataFrame({
    'Variância': variancia_ncota,
    'Desvio Padrão': desvio_padrao_ncota
})
display(stats_variancia_std_ncota)
```

```
print("\nVariância e Desvio Padrão para a tabela cota:")
variancia_cota = cota[colunas_analise_variancia].var()
desvio_padrao_cota = cota[colunas_analise_variancia].std()
stats_variancia_std_cota = pd.DataFrame({
    'Variância': variancia_cota,
    'Desvio Padrão': desvio_padrao_cota
})
display(stats_variancia_std_cota)

colunas_quartis = ['CR', 'Nota Final Vestibular', 'Média Ponderada',
'Aproveitamento Disciplinas', 'Frequência Média']

print("\nQuartis para a tabela ncota:")
display(ncota[colunas_quartis].quantile([0.25, 0.5, 0.75]))

print("\nQuartis para a tabela cota:")
display(cota[colunas_quartis].quantile([0.25, 0.5, 0.75]))
```

- Testes de Normalidade (Shapiro-Wilk)

```
from scipy.stats import shapiro
import pandas as pd # Garantir importação

colunas_teste_normalidade = ['CR', 'Nota Final Vestibular', 'Média
Ponderada', 'Aproveitamento Disciplinas', 'Frequência Média',
'Classificação Vestibular']

grupos = {
    'Não Cotistas': ncota,
    'Cotistas Ensino Público': cota_ensino_publico,
    'Cotistas Racial': cota_racial
}

for nome_grupo, df_grupo in grupos.items():
    print(f"Testes de Normalidade para: {nome_grupo}")
    for coluna in colunas_teste_normalidade:
        if coluna in df_grupo.columns and
len(df_grupo[coluna].dropna()) > 3:
            dados = df_grupo[coluna].dropna()
            try:
                statistic_shapiro, p_value_shapiro = shapiro(dados)
                print(f"  Coluna '{coluna}':")
```

```
                print(f"    Estatística de Shapiro-Wilk:
{statistic_shapiro:.4f}")
                print(f"    P-valor: {p_value_shapiro:.4f}")
                alpha = 0.05
                if p_value_shapiro < alpha:
                    print("    Resultado: Rejeitamos a hipótese nula
(Não Normal).")
                else:
                    print("    Resultado: Não rejeitamos a hipótese
nula (Pode ser Normal).")
            except Exception as e:
                print(f"  Coluna '{coluna}': Não foi possível realizar
o teste Shapiro-Wilk. Erro: {e}")
        else:
            print(f"  Coluna '{coluna}': Dados insuficientes ou coluna
não encontrada para realizar o teste.")
    print("-" * 30)
```

- Testes de Hipóteses Não Paramétricos

```
from scipy.stats import mannwhitneyu, kruskal
import pandas as pd # Garantir importação
import scikit_posthocs as sp # Para testes post-hoc

# Teste U de Mann-Whitney (Não Cotistas vs. Cotistas Combinados)
cr_nao_cotistas = ncota['CR'].dropna()
cr_cotistas = cota['CR'].dropna()
if len(cr_nao_cotistas) > 0 and len(cr_cotistas) > 0:
    statistic_cr, p_value_cr = mannwhitneyu(cr_nao_cotistas,
cr_cotistas, alternative='two-sided')
    print(f"Teste U de Mann-Whitney para a coluna 'CR':")
    print(f"Estatística do teste: {statistic_cr}")
    print(f"P-valor: {p_value_cr}")
else:
    print("Não há dados suficientes na coluna 'CR' em um ou ambos os
grupos para realizar o teste.")

nota_nao_cotistas = ncota['Nota Final Vestibular'].dropna()
nota_cotistas = cota['Nota Final Vestibular'].dropna()
if len(nota_nao_cotistas) > 0 and len(nota_cotistas) > 0:
    statistic_nota, p_value_nota = mannwhitneyu(nota_nao_cotistas,
nota_cotistas, alternative='two-sided')
```

```
    print(f"\nTeste U de Mann-Whitney para a coluna 'Nota Final
Vestibular':")
    print(f"Estatística do teste: {statistic_nota}")
    print(f"P-valor: {p_value_nota}")
else:
    print("Não há dados suficientes na coluna 'Nota Final Vestibular'
em um ou ambos os grupos para realizar o teste.")

media_ponderada_nao_cotistas = ncota['Média Ponderada'].dropna()
media_ponderada_cotistas = cota['Média Ponderada'].dropna()
if len(media_ponderada_nao_cotistas) > 0 and
len(media_ponderada_cotistas) > 0:
    statistic_mp, p_value_mp =
mannwhitneyu(media_ponderada_nao_cotistas, media_ponderada_cotistas,
alternative='two-sided')
    print(f"\nTeste U de Mann-Whitney para a coluna 'Média
Ponderada':")
    print(f"Estatística do teste: {statistic_mp}")
    print(f"P-valor: {p_value_mp}")
else:
    print("Não há dados suficientes na coluna 'Média Ponderada' em um
ou ambos os grupos para realizar o teste.")

aproveitamento_nao_cotistas = ncota['Aproveitamento
Disciplinas'].dropna()
aproveitamento_cotistas = cota['Aproveitamento Disciplinas'].dropna()
if len(aproveitamento_nao_cotistas) > 0 and
len(aproveitamento_cotistas) > 0:
    statistic_ad, p_value_ad =
mannwhitneyu(aproveitamento_nao_cotistas, aproveitamento_cotistas,
alternative='two-sided')
    print(f"\nTeste U de Mann-Whitney para a coluna 'Aproveitamento
Disciplinas':")
    print(f"Estatística do teste: {statistic_ad}")
    print(f"P-valor: {p_value_ad}")
else:
    print("Não há dados suficientes na coluna 'Aproveitamento
Disciplinas' em um ou ambos os grupos para realizar o teste.")

frequencia_nao_cotistas = ncota['Frequência Média'].dropna()
frequencia_cotistas = cota['Frequência Média'].dropna()
if len(frequencia_nao_cotistas) > 0 and len(frequencia_cotistas) > 0:
```

```
    statistic_fm, p_value_fm = mannwhitneyu(frequencia_nao_cotistas,
frequencia_cotistas, alternative='two-sided')
    print(f"\nTeste U de Mann-Whitney para a coluna 'Frequência
Média':")
    print(f"Estatística do teste: {statistic_fm}")
    print(f"P-valor: {p_value_fm}")
else:
    print("Não há dados suficientes na coluna 'Frequência Média' em um
ou ambos os grupos para realizar o teste.")

# Teste de Kruskal-Wallis (Três Grupos)
cr_nao_cotistas = ncota['CR'].dropna()
cr_cota_ep = cota_ensino_publico['CR'].dropna()
cr_cota_racial = cota_racial['CR'].dropna()
grupos_cr = [grupo for grupo in [cr_nao_cotistas, cr_cota_ep,
cr_cota_racial] if len(grupo) > 0]
if len(grupos_cr) >= 2:
    statistic_cr_kruskal, p_value_cr_kruskal = kruskal(*grupos_cr)
    print(f"\nTeste de Kruskal-Wallis para 'CR' entre os três grupos:")
    print(f"Estatística do teste: {statistic_cr_kruskal:.4f}")
    print(f"P-valor: {p_value_cr_kruskal:.4f}")
else:
    print("Dados insuficientes em pelo menos dois grupos para realizar
o teste de Kruskal-Wallis para 'CR'.")

nota_nao_cotistas = ncota['Nota Final Vestibular'].dropna()
nota_cota_ep = cota_ensino_publico['Nota Final Vestibular'].dropna()
nota_cota_racial = cota_racial['Nota Final Vestibular'].dropna()
grupos_nota = [grupo for grupo in [nota_nao_cotistas, nota_cota_ep,
nota_cota_racial] if len(grupo) > 0]
if len(grupos_nota) >= 2:
    statistic_nota_kruskal, p_value_nota_kruskal =
kruskal(*grupos_nota)
    print(f"\nTeste de Kruskal-Wallis para 'Nota Final Vestibular'
entre os três grupos:")
    print(f"Estatística do teste: {statistic_nota_kruskal:.4f}")
    print(f"P-valor: {p_value_nota_kruskal:.4f}")
else:
    print("Dados insuficientes em pelo menos dois grupos para realizar
o teste de Kruskal-Wallis para 'Nota Final Vestibular'.")

mp_nao_cotistas = ncota['Média Ponderada'].dropna()
mp_cota_ep = cota_ensino_publico['Média Ponderada'].dropna()
```

```
mp_cota_racial = cota_racial['Média Ponderada'].dropna()
grupos_mp = [grupo for grupo in [mp_nao_cotistas, mp_cota_ep,
mp_cota_racial] if len(grupo) > 0]
if len(grupos_mp) >= 2:
    statistic_mp_kruskal, p_value_mp_kruskal = kruskal(*grupos_mp)
    print(f"\nTeste de Kruskal-Wallis para 'Média Ponderada' entre os
três grupos:")
    print(f"Estatística do teste: {statistic_mp_kruskal:.4f}")
    print(f"P-valor: {p_value_mp_kruskal:.4f}")
else:
    print("Dados insuficientes em pelo menos dois grupos para realizar
o teste de Kruskal-Wallis para 'Média Ponderada'.")
```

- Correlação (Spearman)

```
from scipy.stats import spearmanr
import pandas as pd

# Correlação entre 'Nota Final Vestibular' e 'CR'
dados_correlacao_cr = pd.concat([
    ncota[['Nota Final Vestibular', 'CR']].assign(Grupo='Não
Cotistas'),
    cota_ensino_publico[['Nota Final Vestibular',
'CR']].assign(Grupo='Ensino Público'),
    cota_racial[['Nota Final Vestibular', 'CR']].assign(Grupo='PPI')
]).dropna()

print("Correlação de Spearman entre 'Nota Final Vestibular' e 'CR'
(Todos os grupos):")
if len(dados_correlacao_cr) > 1:
    coeficiente_spearman_total_cr, p_value_spearman_total_cr =
spearmanr(dados_correlacao_cr['Nota Final Vestibular'],
dados_correlacao_cr['CR'])
    print(f"  Coeficiente de Correlação:
{coeficiente_spearman_total_cr:.4f}")
    print(f"  P-valor: {p_value_spearman_total_cr:.4f}")
else:
    print("  Dados insuficientes para calcular a correlação total.")

dados_nao_cotistas_cr =
dados_correlacao_cr[dados_correlacao_cr['Grupo'] == 'Não Cotistas']
dados_cota_ep_cr = dados_correlacao_cr[dados_correlacao_cr['Grupo'] ==
'Ensino Público']
```

```
dados_cota_racial_cr = dados_correlacao_cr[dados_correlacao_cr['Grupo']
== 'PPI']

print("\nCorrelação de Spearman entre 'Nota Final Vestibular' e 'CR'
para Não Cotistas:")
if len(dados_nao_cotistas_cr) > 1:
    coeficiente_spearman_nc_cr, p_value_spearman_nc_cr =
spearmanr(dados_nao_cotistas_cr['Nota Final Vestibular'],
dados_nao_cotistas_cr['CR'])
    print(f"  Coeficiente de Correlação:
{coeficiente_spearman_nc_cr:.4f}")
    print(f"  P-valor: {p_value_spearman_nc_cr:.4f}")
else:
    print("  Dados insuficientes para calcular a correlação.")

print("\nCorrelação de Spearman entre 'Nota Final Vestibular' e 'CR'
para Ensino Público:")
if len(dados_cota_ep_cr) > 1:
    coeficiente_spearman_cota_ep_cr, p_value_spearman_cota_ep_cr =
spearmanr(dados_cota_ep_cr['Nota Final Vestibular'],
dados_cota_ep_cr['CR'])
    print(f"  Coeficiente de Correlação:
{coeficiente_spearman_cota_ep_cr:.4f}")
    print(f"  P-valor: {p_value_spearman_cota_ep_cr:.4f}")
else:
    print("  Dados insuficientes para calcular a correlação.")

print("\nCorrelação de Spearman entre 'Nota Final Vestibular' e 'CR'
para PPI:")
if len(dados_cota_racial_cr) > 1:
    coeficiente_spearman_cota_racial_cr,
p_value_spearman_cota_racial_cr = spearmanr(dados_cota_racial_cr['Nota
Final Vestibular'], dados_cota_racial_cr['CR'])
    print(f"  Coeficiente de Correlação:
{coeficiente_spearman_cota_racial_cr:.4f}")
    print(f"  P-valor: {p_value_spearman_cota_racial_cr:.4f}")
else:
    print("  Dados insuficientes para calcular a correlação.")

# Correlação entre 'Nota Final Vestibular' e 'Aproveitamento
Disciplinas'
dados_correlacao_ad = pd.concat([
```

```
    ncota[['Nota Final Vestibular', 'Aproveitamento
Disciplinas']].assign(Grupo='Não Cotistas'),
    cota_ensino_publico[['Nota Final Vestibular', 'Aproveitamento
Disciplinas']].assign(Grupo='Ensino Público'),
    cota_racial[['Nota Final Vestibular', 'Aproveitamento
Disciplinas']].assign(Grupo='PPI')
]).dropna()

print("\nCorrelação de Spearman entre 'Nota Final Vestibular' e
'Aproveitamento Disciplinas' (Todos os grupos):")
if len(dados_correlacao_ad) > 1:
    coeficiente_spearman_total_ad, p_value_spearman_total_ad =
spearmanr(dados_correlacao_ad['Nota Final Vestibular'],
dados_correlacao_ad['Aproveitamento Disciplinas'])
    print(f"  Coeficiente de Correlação:
{coeficiente_spearman_total_ad:.4f}")
    print(f"  P-valor: {p_value_spearman_total_ad:.4f}")
else:
    print("  Dados insuficientes para calcular a correlação total.")

dados_nao_cotistas_ad =
dados_correlacao_ad[dados_correlacao_ad['Grupo'] == 'Não Cotistas']
dados_cota_ep_ad = dados_correlacao_ad[dados_correlacao_ad['Grupo'] ==
'Ensino Público']
dados_cota_racial_ad = dados_correlacao_ad[dados_correlacao_ad['Grupo']
== 'PPI']

print("\nCorrelação de Spearman entre 'Nota Final Vestibular' e
'Aproveitamento Disciplinas' para Não Cotistas:")
if len(dados_nao_cotistas_ad) > 1:
    coeficiente_spearman_nc_ad, p_value_spearman_nc_ad =
spearmanr(dados_nao_cotistas_ad['Nota Final Vestibular'],
dados_nao_cotistas_ad['Aproveitamento Disciplinas'])
    print(f"  Coeficiente de Correlação:
{coeficiente_spearman_nc_ad:.4f}")
    print(f"  P-valor: {p_value_spearman_nc_ad:.4f}")
else:
    print("  Dados insuficientes para calcular a correlação.")

print("\nCorrelação de Spearman entre 'Nota Final Vestibular' e
'Aproveitamento Disciplinas' para Ensino Público:")
if len(dados_cota_ep_ad) > 1:
```

```
    coeficiente_spearman_cota_ep_ad, p_value_spearman_cota_ep_ad =
spearmanr(dados_cota_ep_ad['Nota Final Vestibular'],
dados_cota_ep_ad['Aproveitamento Disciplinas'])
    print(f"  Coeficiente de Correlação:
{coeficiente_spearman_cota_ep_ad:.4f}")
    print(f"  P-valor: {p_value_spearman_cota_ep_ad:.4f}")
else:
    print("  Dados insuficientes para calcular a correlação.")

print("\nCorrelação de Spearman entre 'Nota Final Vestibular' e
'Aproveitamento Disciplinas' para PPI:")
if len(dados_cota_racial_ad) > 1:
    coeficiente_spearman_cota_racial_ad,
p_value_spearman_cota_racial_ad = spearmanr(dados_cota_racial_ad['Nota
Final Vestibular'], dados_cota_racial_ad['Aproveitamento Disciplinas'])
    print(f"  Coeficiente de Correlação:
{coeficiente_spearman_cota_racial_ad:.4f}")
    print(f"  P-valor: {p_value_spearman_cota_racial_ad:.4f}")
else:
    print("  Dados insuficientes para calcular a correlação.")
```

Apêndice B - Códigos Sistema *Fuzzy*

```
import numpy as np
import skfuzzy as fuzz
from skfuzzy import control as ctrl
import matplotlib.pyplot as plt

# --- Definição das Variáveis Fuzzy (Antecedentes e Consequente) ---

# Antecedente: Classificação no Vestibular (Rank: 1 é melhor)
classification = ctrl.Antecedent(np.arange(1, 110.1, 0.1),
'Classificação no Vestibular')
classification['muito_bom'] = fuzz.trapmf(classification.universe, [1,
1, 15.71, 30.71])
classification['bom'] = fuzz.trimf(classification.universe, [18.33,
33.33, 48.33])
classification['regular'] = fuzz.trimf(classification.universe, [39,
54, 69])
classification['ruim'] = fuzz.trapmf(classification.universe, [68.48,
83.48, 110, 110])

# Antecedente: Nota Final no Vestibular (0 a 100)
vestibular_score = ctrl.Antecedent(np.arange(0, 100.1, 0.1), 'Nota
Final no Vestibular')
vestibular_score['ruim'] = fuzz.trapmf(vestibular_score.universe, [0,
0, 19, 34])
vestibular_score['regular'] = fuzz.trimf(vestibular_score.universe,
[28.07, 43.07, 58.07])
vestibular_score['bom'] = fuzz.trimf(vestibular_score.universe, [47.72,
62.72, 77.72])
vestibular_score['muito_bom'] = fuzz.trapmf(vestibular_score.universe,
[74.16, 89.16, 100, 100])

# Antecedente: Coeficiente de Rendimento (CR) (0 a 10)
cr = ctrl.Antecedent(np.arange(0, 10.1, 0.1), 'Coeficiente de
Rendimento (CR)')
cr['ruim'] = fuzz.trapmf(cr.universe, [0, 0, 2.4, 4.4])
cr['regular'] = fuzz.trimf(cr.universe, [3.2, 5.2, 7.2])
cr['bom'] = fuzz.trimf(cr.universe, [5.3, 7.3, 9.3])
cr['muito_bom'] = fuzz.trapmf(cr.universe, [7.3, 9.3, 10, 10])

# Antecedente: Média Ponderada (0 a 10)
```

```
weighted_average = ctrl.Antecedent(np.arange(0, 10.1, 0.1), 'Média
Ponderada')
weighted_average['ruim'] = fuzz.trapmf(weighted_average.universe, [0,
0, 2.4, 4.4])
weighted_average['regular'] = fuzz.trimf(weighted_average.universe,
[3.2, 5.2, 7.2])
weighted_average['bom'] = fuzz.trimf(weighted_average.universe, [5.3,
7.3, 9.3])
weighted_average['muito_bom'] = fuzz.trapmf(weighted_average.universe,
[7.3, 9.3, 10, 10])

# Antecedente: Aproveitamento de Disciplinas (%) (0 a 100)
discipline_success = ctrl.Antecedent(np.arange(0, 100.1, 0.1),
'Aproveitamento de Disciplinas (%)')
discipline_success['ruim'] = fuzz.trapmf(discipline_success.universe,
[0, 0, 21.54, 41.54])
discipline_success['regular'] = fuzz.trimf(discipline_success.universe,
[35.62, 55.62, 75.62])
discipline_success['bom'] = fuzz.trapmf(discipline_success.universe,
[69.16, 89.16, 100, 100])

# Antecedente: Frequência Média (%) (0 a 100)
attendance = ctrl.Antecedent(np.arange(0, 100.1, 0.1), 'Frequencia
Media (%)')
attendance['ruim'] = fuzz.trapmf(attendance.universe, [0, 0, 33.91,
53.91])
attendance['regular'] = fuzz.trimf(attendance.universe, [51.11, 71.11,
91.11])
attendance['bom'] = fuzz.trapmf(attendance.universe, [73.33, 93.33,
100, 100])

# Consequente: Desempenho geral do aluno (0 a 10)
overall_performance = ctrl.Consequent(np.arange(0, 10.1, 0.01),
'Desempenho geral do aluno')
overall_performance['ruim'] = fuzz.trapmf(overall_performance.universe,
[0, 0, 2.65, 4.65])
overall_performance['regular'] =
fuzz.trimf(overall_performance.universe, [4.08, 6.08, 8.08])
overall_performance['bom'] = fuzz.trapmf(overall_performance.universe,
[7.09, 9.09, 10, 10])

# --- Definição das Regras Fuzzy ---
```

```
rule1 = ctrl.Rule(
    classification['muito_bom'] & vestibular_score['muito_bom'] &
cr['muito_bom'] & discipline_success['bom'] & attendance['bom'] &
weighted_average['muito_bom'],
    overall_performance['bom'])

rule2 = ctrl.Rule(
    (classification['regular'] | classification['ruim']) &
(vestibular_score['regular'] | vestibular_score['ruim']) & (cr['bom'] |
cr['muito_bom']) & (weighted_average['bom'] |
weighted_average['muito_bom']) & (discipline_success['bom']),
    overall_performance['bom'])

rule3 = ctrl.Rule(
    ((classification['ruim'] & vestibular_score['ruim'] & cr['ruim']) |
(discipline_success['ruim'] & attendance['ruim'])) | (cr['regular'] &
weighted_average['regular']),
    overall_performance['ruim'])

rule4 = ctrl.Rule(
  (classification['muito_bom'] | vestibular_score['muito_bom']) &
(cr['muito_bom'] | weighted_average['muito_bom']),
    overall_performance['bom'])

rule5 = ctrl.Rule(
    (cr['ruim'] & weighted_average['ruim'])|(discipline_success['ruim']
& attendance['ruim']),
    overall_performance['ruim'])

rule6 = ctrl.Rule(
     cr['muito_bom'] & discipline_success['bom'] &
weighted_average['muito_bom'],
    overall_performance['bom'])

rule7 = ctrl.Rule(
    cr['ruim'] | weighted_average['ruim'], overall_performance['ruim'])

rule8 = ctrl.Rule(
    vestibular_score['bom'] & classification['bom'] & (cr['regular'] |
discipline_success['regular']),
    overall_performance['regular'])

rule9 = ctrl.Rule(
```

```
    attendance['bom'] & (cr['bom'] | weighted_average['bom']),
    overall_performance['bom'])

rule10 = ctrl.Rule(
     discipline_success['ruim'] & (attendance['regular'] |
attendance['ruim']) & cr['ruim'],
    overall_performance['ruim'])

rule11 = ctrl.Rule(
    classification['bom'] & vestibular_score['muito_bom'] &
cr['muito_bom'],
    overall_performance['bom'])

rule12 = ctrl.Rule(
    (classification['ruim'] & vestibular_score['ruim']) & (cr['bom'] |
cr['regular']) & (weighted_average['bom'] |
weighted_average['regular']),
    overall_performance['regular'])

rule13 = ctrl.Rule(
  (attendance['bom'] & cr['bom']) | discipline_success['bom'],
    overall_performance['bom'])

rule14 = ctrl.Rule(
    weighted_average['regular'] & attendance['regular'] & cr['regular']
& discipline_success['regular'],
    overall_performance['regular'])

rule15 = ctrl.Rule(
 classification['muito_bom'] & vestibular_score['bom'] &
attendance['bom'] & (cr['bom'] | weighted_average['bom']),
    overall_performance['bom'])

rule16 = ctrl.Rule(
    (classification['muito_bom'] | classification['bom']) &
(vestibular_score['muito_bom'] | vestibular_score['bom']) & (cr['ruim']
| weighted_average['ruim']),
    overall_performance['ruim'])

rule17 = ctrl.Rule(
     (classification['regular'] | classification['bom']) &
(vestibular_score['regular'] | vestibular_score['bom']) &
(cr['muito_bom'] | weighted_average['muito_bom']),
```

```
      overall_performance['bom'])

rule18 = ctrl.Rule(
    cr['bom'] & weighted_average['bom'] & (discipline_success['ruim'] |
attendance['ruim']),
    overall_performance['regular'])

rule19 = ctrl.Rule(
    (cr['ruim'] | cr['regular']) & (weighted_average['ruim'] |
weighted_average['regular']) &
(attendance['regular']|attendance['bom']) &
(discipline_success['bom']|discipline_success['regular']) ,
    overall_performance['regular'])

rule20 = ctrl.Rule(
     cr['regular'] & discipline_success['regular']
    overall_performance['regular'])

rule21 = ctrl.Rule(
    cr['muito_bom'] & (classification['regular'] |
classification['ruim']) & (vestibular_score['regular'] |
vestibular_score['ruim']),
    overall_performance['bom'])

rule22 = ctrl.Rule(
    (cr['ruim'] | weighted_average['ruim']) &
discipline_success['regular'],
    overall_performance['ruim'])

rule23 = ctrl.Rule(
    attendance['bom'] & discipline_success['bom'] & (cr['regular'] |
cr['bom']),
    overall_performance['bom'])

rule24 = ctrl.Rule(
    (classification['bom'] | classification['muito_bom']) &
(vestibular_score['bom'] | vestibular_score['muito_bom']) &
(cr['regular'] | weighted_average['regular']),
    overall_performance['regular'])

rule25 = ctrl.Rule(
    (cr['bom'] | cr['muito_bom']) & (weighted_average['bom'] |
weighted_average['muito_bom']),
```

```
    overall_performance['bom'])
all_rules = [rule1, rule2, rule3, rule4, rule5, rule6, rule7, rule8,
rule9, rule10, rule11, rule12, rule13, rule14, rule15, rule16, rule17,
rule18, rule19, rule20, rule21, rule22, rule23, rule24, rule25]

overall_performance_ctrl = ctrl.ControlSystem(all_rules)

overall_performance_sim =
ctrl.ControlSystemSimulation(overall_performance_ctrl)

try:
    df = pd.read_excel('Dados Matemática separado por cotas.xlsx')
    print("Arquivo Excel 'Dados Matemática separado por cotas.xlsx'
carregado com sucesso!")
except FileNotFoundError:
    print("Erro: Arquivo 'Dados Matemática separado por cotas.xlsx' não
encontrado. Por favor, verifique o nome do arquivo e se ele foi
carregado corretamente.")
    raise

relevant_columns = [
    'Classificação Vestibular',
    'Nota Final Vestibular',
    'CR',
    'Aproveitamento Disciplinas',
    'Frequência Média',
    'Média Ponderada'
]

if not all(col in df.columns for col in relevant_columns):
    missing_cols = [col for col in relevant_columns if col not in
df.columns]
    print(f"Erro: Colunas necessárias não encontradas no arquivo Excel:
{missing_cols}")
    raise ValueError("Missing required columns in DataFrame")

df_filtered = df[relevant_columns].copy()

overall_performance_results = []

for index, row in df_filtered.iterrows():
    try:
```

```
        overall_performance_sim.input['Classificação no Vestibular'] =
row['Classificação Vestibular']
        overall_performance_sim.input['Nota Final no Vestibular'] =
row['Nota Final Vestibular']
        overall_performance_sim.input['Coeficiente de Rendimento (CR)']
= row['CR']
        overall_performance_sim.input['Aproveitamento de Disciplinas
(%)'] = row['Aproveitamento Disciplinas'] * 100
        overall_performance_sim.input['Frequencia Media (%)'] =
row['Frequência Média'] * 100
        overall_performance_sim.input['Média Ponderada'] = row['Média
Ponderada']

        overall_performance_sim.compute()
overall_performance_results.append(overall_performance_sim.output['Dese
mpenho geral do aluno'])
    except Exception as e:
        print(f"Erro ao processar a linha {index}: {e}")
        overall_performance_results.append(np.nan)

df['Desempenho Geral Fuzzy'] = overall_performance_results

def get_linguistic_term(fuzzy_output):
    if pd.isna(fuzzy_output):
        return np.nan

    terms = list(overall_performance.terms.keys())
    memberships = {term:
fuzz.interp_membership(overall_performance.universe,
overall_performance[term].mf, fuzzy_output) for term in terms}

    max_membership = max(memberships.values())

    if max_membership == 0:
        return 'indefinido'

    linguistic_terms = [term for term, mem in memberships.items() if
mem == max_membership]

    if len(linguistic_terms) > 1:
        return '/'.join(linguistic_terms)

    return linguistic_terms[0]
```

```
df['Desempenho Geral Linguistico'] = df['Desempenho Geral
Fuzzy'].apply(get_linguistic_term)

display(df[['Classificação Vestibular', 'Nota Final Vestibular', 'CR',
'Aproveitamento Disciplinas', 'Frequência Média', 'Média Ponderada',
'Desempenho Geral Fuzzy', 'Desempenho Geral Linguistico']].head())

performance_by_cotas = df.groupby('Cotas')['Desempenho Geral
Linguistico'].value_counts().unstack(fill_value=0)

print("Comparação do Desempenho Geral Linguistico por Cotas (Sistema
Fuzzy Ajustado):")
display(performance_by_cotas)
```