Universidade de Brasília
IE- Departamento de Estatística

Lore Martins Bueno

ANÁLISE DE CRÉDITO: MEDIDAS DE AVALIAÇÃO DE MODELOS E APLICAÇÃO DA TEORIA *FUZZY* NA TOMADA DE DECISÃO

Brasília, DF
2011

LORE MARTINS BUENO

06/89378

ANÁLISE DE CRÉDITO: MEDIDAS DE AVALIAÇÃO DE MODELOS E APLICAÇÃO DA TEORIA *FUZZY* NA TOMADA DE DECISÃO

Relatório apresentado à disciplina Estágio Supervisionado II do curso e graduação em Estatística, Departamento de Estatística, Instituto de Exatas, Universidade de Brasília, como parte dos requisitos necessários para o grau de Bacharel em Estatística.

Orientação: Prof.° Luis Gustavo do Amaral Vinha

Brasília, DF

2011

B928a Bueno, Lore Martins.

Análise de Crédito: Medidas de avaliação de modelos e aplicação da teoria *fuzzy* na tomada de decisão / Lore Martins Bueno. – 2011

55 f. : il. ; 30 cm.

Inclui bibliografia.

Orientação: Luis Gustavo do Amaral Vinha

Monografia (graduação) – Universidade de Brasília, Instituto de Ciências Exatas, Departamento de Estatística, 2011

1. Algoritmo MLFE. 2. Lógica *Fuzzy*. 3. Métodos *Fuzzy* automáticos. 4. Modelo de Crédito. 5. Risco de Crédito.
I. Vinha, Luís Gustavo do Amaral (orient.). II. Título.

CDU 657.3/.4

LORE MARTINS BUENO

06/89378

ANÁLISE DE CRÉDITO: MEDIDAS DE AVALIAÇÃO DE MODELOS E APLICAÇÃO DA TEORIA *FUZZY* NA TOMADA DE DECISÃO

Monografia de graduação submetida à disciplina Estágio Supervisionado II do curso de graduação em Estatística do Departamento de Estatística, Instituto de Exatas, Universidade de Brasília, como parte dos requisitos necessários para o grau de Bacharel em Estatística.

Aprovada por:

_____

LUIS GUSTAVO DO AMARAL VINHA, MESTRE, EST/UNB.
ORIENTADOR

_____

AFRÂNIO MÁRCIO CORRÊA VIEIRA, DOUTOR, EST/UNB.
EXAMINADOR INTERNO

_____

JULIANA BETINI FACHINI, MESTRE, EST/UNB.
EXAMINADORA INTERNA

Brasília, DF
Julho de 2011

AGRADECIMENTOS

Em primeiro lugar, agradeço a Deus por todas as oportunidades que me foram oferecidas para que eu me tornasse um *outlier* nas estatísticas de um país que ainda luta contra o analfabetismo.

Em segundo, mas não menos importante, agradeço à imensa ajuda e paciência do meu orientador, Professor Luis Gustavo do Amaral Vinha, que inspirou a apuração do meu senso crítico e me deu dicas valiosas, sem as quais meu trabalho não seria concluído.

Agradeço ao Professor Doutor Alan Ricardo Silva, que me ensinou a "pensar como a máquina". Aos queridos Thaís e Guilherme Rodrigues, também agradeço pela ajuda essencial em linguagem de programação.

Agradeço a minha família pelo esforço de nunca deixar faltar recursos para que minha formação acadêmica se concretizasse. Agradeço à companhia alegre e incentivadora dos meus colegas de curso, daqueles que partilharam comigo a pressão do fim e o triunfo que a segue, às minhas amigas que esperaram pacientemente a conclusão do meu trabalho e também àquelas que estiveram presente durante a confecção do mesmo, sempre me apoiando com energias positivas.

Em especial, agradeço ao Gabriel Pereira Fortes por fazer o papel de meu colega da monografia e por me dar mais carinho e suporte do que eu imaginei que poderia ter, fazendo jus ao título de "companheiro para todas as horas", no sentido mais amplo da expressão.

"As far as the laws of mathematics
refer to reality, they are not certain,
as far as they are certain,
they do not refer to reality."

(Albert Einstein)

RESUMO

As instituições financeiras brasileiras têm voltado sua atenção para o gerenciamento de risco desde que o Banco Central aderiu ao segundo acordo de Basiléia, no qual as propostas foram elaboradas com o objetivo de tornar o sistema financeiro internacional mais homogêneo, sugerindo mudanças rigorosas na metodologia de gerência do risco e supervisão bancária. Para tal, é necessário que as instituições desenvolvam métodos eficazes na avaliação do risco e na decisão massiva de crédito. Esse trabalho tem por objetivo apresentar um sistema que auxilie na tomada de decisão do microcrédito baseado na Teoria *Fuzzy,* bem como comparar seu desempenho com um modelo de *Credit Scoring*, metodologia mais comum entre as instituições financeiras. As informações utilizadas no estudo – base de dados e modelo de *Credit Scoring* – foram fornecidas por uma instituição financeira atuante no mercado. Devido à quantidade de variáveis utilizadas na construção do sistema *fuzzy*, observou-se a necessidade de automatizar o processo de obtenção das regras e funções de pertinência. Para isso, foi desenvolvido um algoritmo em linguagem SAS/IML, adaptado do método automático MLFE para geração de sistemas *fuzzy*. O resultado da avaliação que comparou os dois modelos indicou que o sistema *fuzzy* se mostrou mais eficiente que o modelo de *Credit Scoring* na avaliação do crédito e concluiu-se que essa nova metodologia pode ser bem aceita no âmbito bancário de risco e ser aplicada em um sistema real de decisão de crédito.

Palavras-chave: algoritmo MLFE, lógica *fuzzy*, métodos *fuzzy* automáticos, modelo de crédito, risco de crédito.

ABSTRACT

Brazilian financial institutions have turned their attention to risk management since the Central Bank, BACEN, joined the second Basel agreement, where proposals were designed in order to make the international financial regulations more uniform, globalized and resilient, suggesting severe changes in the methodology of risk management and banking supervision. To enact these changes is necessary for institutions to develop effective methods in risk assessment and credit decisions. The method adopted to inform the large number of decisions should simplify the process and be reliable. This work aims to develop a credit decision system based on fuzzy theory as well as compare its performance with a credit scoring model, which is the most used methodology nowadays. The database and credit scoring model used on the comparison were provided by a Brazilian financial institution active in the market. Due to the amount of variables used in constructing the fuzzy system, there was a need to automate the process of obtaining the rules and membership functions. To make this possible, an algorithm in SAS/IML language was adapted from MLFE automatic method for generating fuzzy systems. The result of the evaluation that compared the two models indicated that the fuzzy system is more efficient than the credit scoring model in predicting the defaulters. The conclusion was that this new methodology can be well accepted in a real a system of credit decision and assessment.

Keywords: MLFE algorithm, automated *fuzzy* methods, credit model, credit risk, *fuzzy* logic.

SUMÁRIO

Página

1 INTRODUÇÃO

Após a implantação do Plano Real no Brasil em julho de 1994, houve grande crescimento na demanda por crédito. O novo plano foi capaz de segurar a hiperinflação presente a quase 15 anos no país, como também alcançar a estabilidade dos preços. Ao fim de 1994, o PIB cresceu 5,67% e o setor industrial apresentou expansão de 7%. A economia apresentava sinais de reaquecimento e a entrada das classes C e D no mercado consumidor gerou aumento da oferta de crédito. Com a inflação contida, isso significava prestações sem aumento todo mês.

Paralelamente ao aumento na concessão de crédito, ocorreu um aumento nas perdas bancárias, decorrente da maior concentração de negócios com clientes inadimplentes. Esse novo contexto econômico abriu um interessante campo para estudo ao relacionar a capacidade financeira de empresas e pessoas físicas com a ocorrência de fenômenos periódicos como redução inflacionária, volatilidade das taxas de juros, recessão econômica, desemprego etc.

Segundo Silva (2008), crédito, no conceito restrito e específico de que se trata esse trabalho, consiste na entrega de um valor presente mediante uma promessa de pagamento no futuro. Em um banco, por exemplo, essa transação pode ser traduzida como a compra da promessa de pagamento, onde a instituição coloca à disposição do cliente (tomador de recursos) um determinado valor para, no futuro, receber um valor maior.

O crédito é um dos principais meios de que as pessoas dispõem para adquirirem os bens e serviços de que necessitam e para usufruir de outros que a sociedade moderna oferece, desempenhando grande papel social:

- Estimula o consumo influenciando na demanda;
- Possibilita as empresas a aumentarem seu nível de atividade;
- Facilita as pessoas a obterem moradia, bens e até alimentos;
- Ajuda na execução de projetos para os quais as empresas não dispunham de recursos próprios.

O crédito, porém, não é concedido indiscriminadamente, já que a expectativa de recebimento de um montante de dinheiro numa data futura depende da capacidade do tomador de cumprir a promessa de pagamento. Sendo assim, existe risco de o

pagamento não acontecer, sendo que o lucro dos bancos credores depende diretamente da quantidade de clientes que quitam suas dívidas. Desse modo o risco de crédito é definido como a probabilidade de que o recebimento não ocorra. Nesse contexto surge o interesse em quantificar o risco de crédito.

O crédito pode ser concedido tanto à pessoa física, quanto à pessoa jurídica, seja esta uma pequena, média ou grande empresa. Para a agência financiadora, no entanto, investimentos solicitados por empresas de grande porte representam maior perda, caso a empresa venha a se tornar inadimplente. Assim, fez-se necessária a criação de uma regulamentação que protegesse o mercado financeiro de quebras decorrentes da má administração na concessão.

Um dos principais papéis do Sistema Financeiro Nacional é dar segurança ao próprio sistema e ao depositante. A preocupação com a solidez dos sistemas financeiros é universal e o Comitê de Basiléia tem prestado grande contribuição na busca de certa universalização de conceitos e procedimentos. Esse papel regulador tem forte interferência na estrutura organizacional das instituições financeiras. No Brasil, o risco de crédito teve significativa atenção e evolução com a Resolução n° 2.682/99 do Banco Central do Brasil (BACEN), obrigando os bancos a classificar suas operações de acordo com o risco e a efetuar o depósito da taxa de aprovisionamento adequada. A gestão de risco vem merecendo profunda atenção e requerendo elevados volumes de investimento em inteligência, sistemas e processos desde então.

A mensuração do risco, além de atender às exigências das autoridades monetárias do país, serve também como referencial para identificar a chance de perda de uma determinada operação e assim orientar sua precificação. Nos bancos, contribui no auxílio à redução de perdas decorrentes da responsabilidade de assumir riscos indevidos, bem como propiciando a busca da maximização do valor do banco a partir da tomada de decisão orientada pela avaliação da relação risco-retorno.

Sobehart (2001) afirma que muitas das grandes instituições financeiras mundiais desenvolveram modelos estatísticos que ajudam a medir, a monitorar e a gerenciar o risco das suas linhas de crédito. É prática comum adotar modelos logísticos, já que essa metodologia é bem aceita no mercado bancário, reconhecida pelo comprovado apoio prestado à gestão do risco.

À medida que as instituições se tornam especialistas nas técnicas de modelagem, o foco da pesquisa passa a ser a validação dos modelos. Entretanto, o Comitê de

Supervisão de Operações Bancárias de Basiléia (*Basel Committee of Banking Supervision*) recentemente identificou a validação desses modelos de risco como uma das tarefas mais desafiadoras para as instituições que apóiam suas decisões neles. Assegurar a adequação do modelo é uma tarefa crucial, pois a economia é dinâmica e uma realidade usada na construção de um modelo hoje pode não ser útil ou verdadeira no futuro (Sobehart, 2001).

No entanto, apesar de muito disseminada, a regressão logística possui algumas limitações, no que se refere à modelagem de relações complexas não-lineares e nem sempre é a melhor maneira de lidar com as questões que surgem no gerenciamento do risco. Com isso, buscou-se discutir outros métodos que pudessem dar assistência às decisões, como a teoria *fuzzy.* Essa é uma metodologia recente e foi pouco explorada no mercado de risco, apesar de estar bem difundida em outros campos que necessitam de apoio à tomada de decisão (Souza, 2003).

As instituições estão em busca de novas metodologias e técnicas capazes de atender às demandas de mercado e às recentes exigências regulatórias, portanto, a discussão e comparação de novos métodos com os métodos atuais são de grande importância para o mercado.

2 MOTIVAÇÃO

Um modelo de *Credit Scoring* é considerado bom quando consegue discriminar, entre os tomadores de empréstimos, os adimplentes dos inadimplentes. Entretanto os desafios relacionados ao uso contínuo destes modelos residem, muitas vezes, na falta de dados suficientes para o seu desenvolvimento. Nesse contexto, as instituições financeiras tem buscado cada vez mais medidas para a avaliação dos modelos de risco.

Devido a dificuldades em encontrar padrões nas avaliações de clientes inadimplentes, muitos testes estatísticos não são capazes de identificar o modelo mais eficaz ou mesmo indicar a necessidade de recalibrar ou reavaliar modelos em vigor. Logo, uma análise mais detalhada acerca dessas medidas se faz necessária para que as instituições financeiras tenham mais respaldo para a tomada de decisões (Sobehart, 2001).

No dia-a-dia, porém, nos deparamos com inúmeras dificuldades práticas, onde se torna difícil reunir dados suficientes e adequados para o ajuste de modelos. Uma sugestão é a adoção de um Sistema *Fuzzy* como ferramenta na decisão do crédito, que traz benefícios como flexibilidade e fácil adequação, mesmo a casos complexos como o discutido.

A vantagem do uso de um Sistema *Fuzzy* está na possibilidade de incorporar a experiência de um especialista no processo, para que se disponha da melhor estratégia no momento da decisão. A sistematização de conhecimento humano é possível na Teoria *Fuzzy*, pois ela emprega variáveis linguísticas ao invés de variáveis numéricas, tornando viável quantificar expressões numericamente vagas como “parcialmente correto” ou “mais ou menos arriscado”.

A teoria de conjuntos *fuzzy* se diferencia da teoria clássica no aspecto em que admite que um elemento pertença parcialmente a um conjunto. Assim, é possível que o sistema raciocine de forma semelhante ao homem, considerando a subjetividade no momento de classificar o cliente. Com essa estratégia é que se busca diminuir a confusão em identificar os perfis de indivíduos que venham a se tornar inadimplentes (Cesar, 2005).

3 OBJETIVOS

3.1 Objetivo Geral

Desenvolver um Sistema *Fuzzy*, apresentar de que forma ele pode auxiliar na tomada de decisão da análise de crédito e comparar sua eficiência com um modelo baseado em regressão logística.

3.2 Objetivos Específicos

- Apresentar detalhadamente a metodologia da Teoria *Fuzzy*;
- Desenvolver um sistema *fuzzy* capaz de lidar com grande quantidade de dados e variáveis de forma automática, simples e eficiente;
- Explorar a utilização dos métodos de avaliação da qualidade dos modelos de risco de crédito e apresentar as medidas de avaliação já conhecidas;
- Aplicar as medidas estudadas e o Sistema *Fuzzy* em dados reais;
- Comparar os resultados obtidos entre modelagens construídas através de regressão logística e Lógica *Fuzzy*.

4 REFERENCIAL TEÓRICO

4.1 Modelos de Crédito

O risco de uma solicitação de crédito pode ser avaliado de forma subjetiva ou de forma objetiva, com uso de metodologia quantitativa. Essa forma de decisão se tornou necessária com a popularização do microcrédito, um crédito pulverizado em quantias que, por não representarem a maior parte do lucro de uma agência financiadora, não se considera crucial ter uma equipe de gestores que analisem as propostas individualmente. Assim, as decisões podem ser automatizadas e ganha-se em agilidade.

Dentre os vários métodos quantitativos existentes, discutir-se-á o modelo de *Credit Scoring*, que tem o objetivo de estimar, na data da decisão do crédito, a probabilidade da concessão incorrer em perda para a instituição. O *credit score* é uma medida do risco e a forma como essa informação é utilizada na decisão do crédito cabe ao gestor do crédito. Aqui, o risco de crédito, ou seja, a probabilidade do tomador se tornar inadimplente, será chamada de probabilidade de *default*.

Como qualquer outra tentativa de previsão da realidade, o uso de modelos de *Credit Scoring* incorre em vantagens e limitações (Silva, 2008).

Vantagens:

- O uso de um modelo válido baseado em série histórica (experiência anterior da instituição) atribui certa segurança ao analista;
- A escolha adequada das variáveis e seus respectivos pesos elimina a subjetividade presente no julgamento de diferentes analistas;
- Ganho em agilidade pela padronização da avaliação;
- Confirmação de que algumas variáveis consideradas importantes não são necessariamente significativas na avaliação.

Limitações:

- As variáveis e seus respectivos pesos sofrem alterações com o decorrer do tempo;
- A instituição está sujeita a erros por má utilização do modelo;
- Características peculiares das diferentes classes de clientes limitam o uso de um modelo geral.

A tendência é que no crédito massificado, ou seja, quando a empresa trabalha com grande número de proponentes a realizar negócios de pequeno valor, sejam utilizados métodos estatísticos que possibilitem uma decisão rápida com a expectativa de adequado nível de segurança. Umas das metodologias usadas na construção de modelos de *Credit Scoring* é a regressão logística.

4.1.1 Regressão Logística

Hosmer e Lemeshow (2000) afirmam que o uso da regressão logística se estabeleceu efetivamente há menos de duas décadas. Inicialmente aplicada na pesquisa epidemiológica, acabou se consolidando como método de referência em campos diversos de pesquisa como biomedicina, finanças, ecologia, engenharia, entre outros. Tal qual em outros modelos de regressão, o objetivo da regressão logística é encontrar o modelo mais adequado e parcimonioso para descrever a relação entre a variável resposta (dependente) e um conjunto de variáveis preditoras (independentes).

O modelo de regressão logística é usado nos casos em que a variável resposta tem caráter não métrico, portanto, qualitativa. O modelo logístico binário é o mais utilizado e neste caso a variável resposta assume apenas dois níveis. Assim, procura-se ajustar uma função que permita estimar a probabilidade de uma observação pertencer a um dos dois grupos, ou que indique a presença ou não de certo atributo em razão do comportamento de um conjunto de variáveis independentes.

No presente estudo, o indivíduo a ser avaliado é o tomador de empréstimo, que poderá ser classificado como adimplente ou inadimplente. Logo, se cada observação Y_i assumir os valores 0 ou 1, Y_i terá distribuição Bernoulli com probabilidade de sucesso π_i, logo

$$P(Y_i = 1) = \pi_i \quad e \quad P(Y_i = 0) = 1 - \pi_i,$$

e variância dada por

$$Var(Y_i) = \pi_i(1 - \pi_i,).$$

Assim, no contexto de gerência de crédito, a regressão logística é utilizada para a avaliação da inadimplência de determinado grupo de clientes, assumindo que a probabilidade de *default* é logisticamente distribuída. O modelo de regressão logística é dado por

$$E(Y_i) = \pi_i(\boldsymbol{X}) = \frac{exp\ (\beta_0 + \beta_1 X_{1,i} + \cdots + \beta_p X_{p,i})}{1 - exp(\ \beta_0 + \beta_1 X_{1,i} + \cdots + \beta_p X_{p,i})}$$

onde $0 \leq \pi_i(\boldsymbol{X}) \leq 1$.

Pode-se, portanto, definir o escore como uma função das características do indivíduo

$$escore = \beta_0 + \beta_1 X_{1,i} + \cdots + \beta_p X_{p,i}.$$

De maneira geral, tem-se

$$P(inadimplente) = P(Y_i = 1) = \frac{e^{escore}}{1 + e^{escore}}.$$

Desse modo, no contexto de risco de crédito, tem-se que a variável resposta representa a situação do cliente. Ela assume o valor 1 para aqueles em situação de *default,* e valor 0 caso contrário. As variáveis independentes correspondem aos dados cadastrais coletados no momento da concessão e dizem respeito às características socioeconômicas que o analista acredita que influenciam no descumprimento do acordo entre cliente e instituição.

No ponto de vista matemático, segundo Hosmer e Lemeshow (2000), o modelo logístico apresenta vantagens como flexibilidade e fácil manuseio, possibilitando interpretação direta de seus parâmetros.

4.2 Teoria *Fuzzy*

O termo *fuzzy* vem da língua inglesa e pode ser traduzido como “nebuloso”, “vago”, “incerto” ou “impreciso”. Desde a Grécia antiga, tais expressões já perturbavam o ser humano, instigando os filósofos a buscarem uma forma de conceituá-las formalmente. Essa tentativa pode ser detectada num pensamento apresentado por Bertrand Russell (1872-1970): “Um homem tem a cabeça repleta de cabelos e, a partir de certo momento, começa-se a extração de fios, um a um, e a cada fio retirado pergunta-se se ele está calvo. Em que momento exatamente este homem ficará calvo?” Além disso, se o processo fosse repetido indefinidamente, o homem ficaria até careca, mas será que podemos definir a quantidade de fios de cabelo que estabelece a diferença entre calvo e careca?

No fim do século XIX, a comunidade científica também enfrentava um estado incômodo gerado pelo conceito de incerteza, quando os físicos notaram que a mecânica clássica já não era suficiente para resolver problemas de ordem molecular. O surgimento da mecânica quântica e de novos métodos aliados à estatística revolucionou a ciência no século passado, já que se tornou possível sintetizar o comportamento de vários agentes microscópicos em uma única medida e aplicá-la diretamente nos modelos macroscópicos adequados. Desde então, a influência da incerteza é considerada nos problemas, enquanto que a tentativa de construir modelos mais robustos nos permitiu alcançar soluções confiáveis e ao mesmo tempo, quantificá-la (Klir e Yuan, 1995).

No século XX, o filósofo quântico Max Black (1909-1989) publicou o artigo "*Vagueness: an exercise in logical analysis"*, texto precursor da idéia de conjunto *fuzzy*, que naquela época, porém, não chamou atenção dos filósofos. Só em 1965, o renomado engenheiro eletricista, Lotfi Asker Zadeh, lançou seu primeiro artigo sobre o assunto, intitulado "*Fuzzy Sets*". Devido a sua notável influência, o professor da Universidade de Berkeley na Califórnia, divulgou amplamente suas idéias e hoje é considerado precursor da Teoria de Conjuntos *Fuzzy*.

As teorias de Conjuntos *Fuzzy* e Lógica *Fuzzy* sustentam a base para geração de técnicas poderosas para a solução de problemas em áreas como sistemas de controle, tomada de decisão, reconhecimento de padrões e processamento de imagens digitais. O sucesso das aplicações motivou de tal forma o desenvolvimento da teoria *fuzzy* que hoje em dia máquinas de lavar roupas e outros eletrodomésticos são desenvolvidos usando essa lógica tendo em vista aprimorar seu funcionamento (Souza, 2003).

A Lógica *Fuzzy* (ou difusa) é inovadora devido a sua capacidade em tirar conclusões e gerar respostas baseadas em informações vagas, ambíguas, qualitativamente incompletas e imprecisas (Faria, 2006). Nesse aspecto, os sistemas *fuzzy* têm habilidade de raciocinar de forma semelhante à dos humanos. Seu comportamento é representado de maneira muito simples e natural, levando à construção de sistemas compreensíveis e de fácil manutenção.

4.2.1 Conjuntos *Fuzzy* e Lógica *Fuzzy*

No conceito de conjuntos *fuzzy* introduzido por Zadeh, esses conjuntos não tem limites precisos, ou seja, os elementos possuem um grau de pertinência aos conjuntos, que varia no intervalo [0,1], sendo a pertinência total denotada pelo valor 1. Esse conceito, na lógica proposicional, quer dizer que uma premissa pode ser parcialmente verdadeira, contrariando um dos axiomas da Lógica Clássica, o "Princípio do Terceiro Excluído". Ele estabelece que uma premissa poderá somente assumir os valores "verdadeiro" ou "falso", não existindo outra opção. Ao lidar com problemas do mundo real, no entanto, a informação não pode ser apenas "completamente verdadeira" ou "completamente falsa". Temos que lidar com situações incertas e desconhecidas, onde definições como parcialmente verdadeiro, verdadeiro sob tal ponto de vista e até verdadeiro com certa probabilidade são formas mais adequadas de classificar as ocorrências.

Muitas vezes, classificamos os fenômenos que observamos de forma subjetiva e as definições que temos disponíveis acerca das variáveis podem ser descritas de forma linguística. Essa informação é muito valiosa para o estudo de certos acontecimentos que são difíceis de modelar. De posse apenas desse tipo de informação, conclusões corretas são tomadas geralmente apenas por um *expert* do assunto. Médicos são capazes de classificar a taxa de gordura corporal de seus pacientes sem aferir uma medida sequer, baseando sua conclusão na observação do conjunto das proporções do indivíduo, como circunferência do abdômen e altura, usando, para isso, apenas variáveis linguísticas:

- "baixo", "médio", "alto" para a altura;
- "pequena", "média", "grande" para a circunferência do abdômen;
- "abaixo do normal", "ideal" e "acima do aceitável" para o volume do indivíduo.

Sua experiência lhe diz que indivíduos baixos com circunferência de abdômen grande apresentam volume acima do aceitável, portanto estão em situação de sobrepeso. Assim é possível avaliar a qualidade de vida do paciente e o risco de incidência de doenças relacionadas ao excesso de peso. O raciocínio usado para alcançar a conclusão é chamado de implicação, formado por uma premissa e um consequente (Tabela 1).

Tabela 1: Implicações.

SE	altura é “baixo” **E** circunferência é “grande”	**ENTÃO**	Volume é “acima do aceitável”
	altura é “alto” **E** circunferência é “pequena”		Volume é “abaixo do normal”
	altura é “médio” **E** circunferência é “pequena”		Volume é “ideal”

Apesar de muitas vezes não interferir na tomada de decisão do especialista, essa forma não quantificada do pensamento humano não é clara. Com o objetivo de agilizar e automatizar o processo de decisão pode-se quantificar para o computador, através da teoria *fuzzy*, o significado linguístico de cada regra estabelecida pelo profissional atribuindo graus de pertinência a elementos de um conjunto *fuzzy*.

Como exemplo, para um conjunto *fuzzy* das “altas temperaturas”, as temperaturas 35°C e 43°C são elementos desse conjunto, embora a temperatura 43°C possua um grau de pertinência maior. De maneira não muito compreendida, humanos tem a capacidade de associar um grau de pertinência a um objeto sem compreender conscientemente como se chega a ele. Por exemplo, em uma ficha de avaliação do professor, o aluno não tem dificuldade em conferir um grau de pertinência ao professor no conjunto “domínio do conteúdo”. Esse grau é alcançado imediatamente sem que se faça uma análise consciente sobre os fatores que influem nessa decisão.

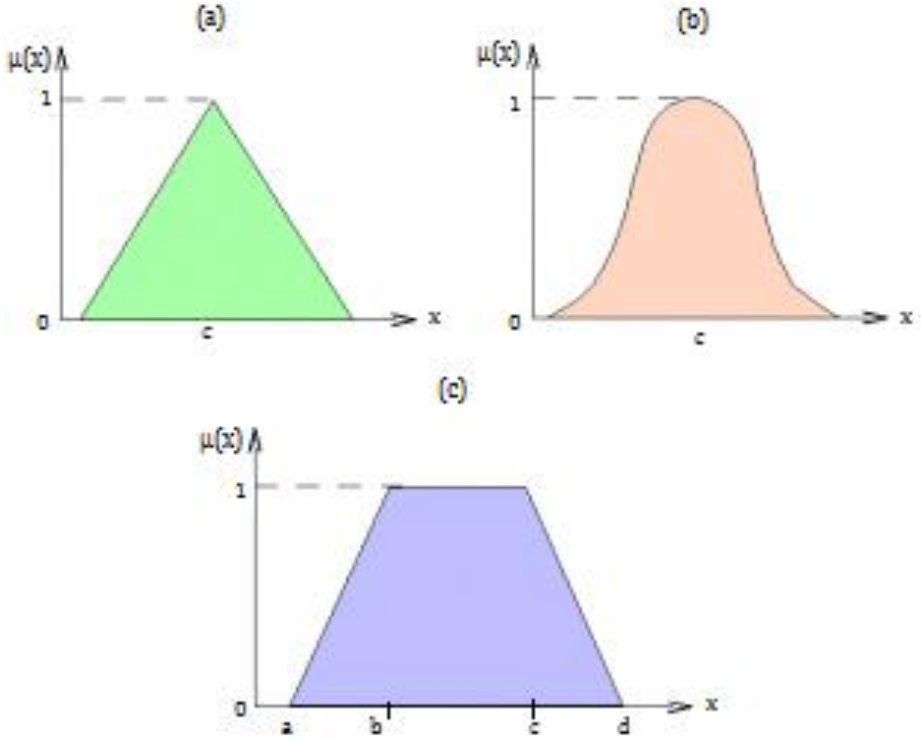
O grau de pertinência quantifica a compatibilidade que um termo linguístico tem com o significado da variável. A função de pertinência mostra todos os níveis de classificação das variáveis e com qual certeza cada significado pode ser atribuído a certo termo. Não se deve confundir a palavra certeza com probabilidade ou verossimilhança. A função de pertinência não é uma densidade de probabilidade, muito menos está contida num espaço de probabilidade. A certeza deve ser entendida por compatibilidade ou grau de verdade. A função de pertinência não quantifica comportamentos aleatórios, simplesmente diminui a imprecisão no significado linguístico conferido às variáveis.

Para definir uma função de pertinência é necessário determinar dois parâmetros:

- centro (c), valor ou intervalo de valores em que a função assume pertinência máxima;
- dispersão (s), determina a largura da curva.

Na Figura 1 são apresentadas algumas das funções mais utilizadas: triangular, trapezoidal e a gaussiana, seguidas da definição das funções gaussiana e triangular:

Figura 1 – Funções de pertinência: (a) triangular, (b) gaussiana, (c) trapezoidal.



Fonte – http://condicaoinicial.com/tag/logica-fuzzy.

$$\mu_{Gaussiana}(x) = exp\left\{-\frac{1}{2}\left(\frac{x-c}{s}\right)^2\right\};$$

$$\mu_{Triangular}(x) = \begin{cases} máx\left\{0; 1+\frac{x-c}{s}\right\}, & se\ x \leq c \\ máx\left\{0; 1+\frac{c-x}{s}\right\}, & se\ x > c \end{cases}.$$

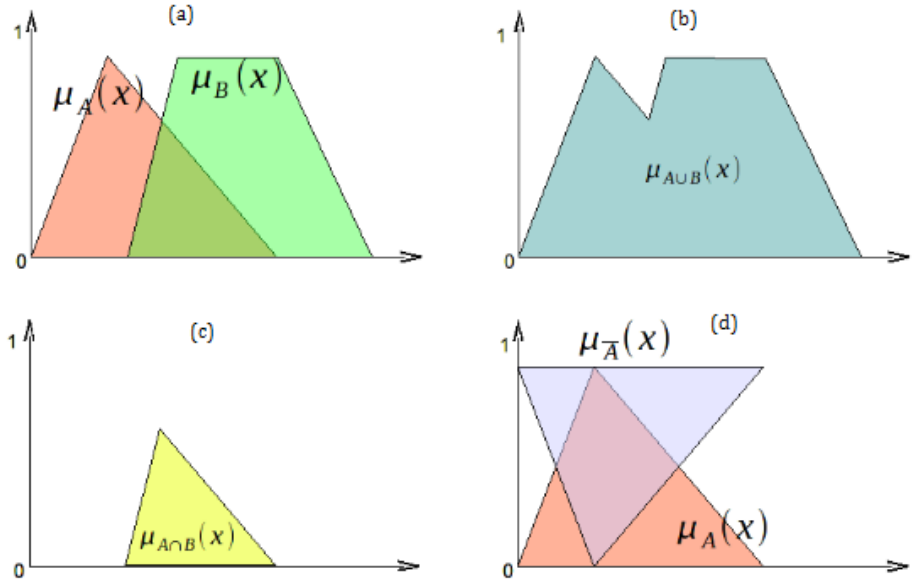
$$\mu_{Trapezoidal}(x) = \begin{cases} 0, & x < a,\ x > d \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b \leq x \leq c \\ \frac{d-x}{d-c}, & c \leq x \leq d \end{cases}$$

Além das apresentadas, existe uma vasta gama de funções de pertinência que podem ser usadas de acordo com a necessidade do problema. Cada curva representa um conjunto ou subconjunto *fuzzy.* O eixo das abscissas representa o elemento e o eixo das ordenadas, o grau de pertinência desse elemento ao conjunto. O grau de pertinência é denotado por $\mu_A(x)$, onde, por exemplo, $\mu_A(x) = 0{,}7$ significa que o grau de pertinência do elemento x ao conjunto A é de 0,7 ou ainda, em uma escala de zero a um, o elemento x é compatível com o conjunto A num grau de 0,7.

Os conjuntos *fuzzy* também podem ser manipulados algebricamente com operações de união, interseção, complemento, entre outras (Figura 2). Contudo, estas operações são definidas em termos do grau de pertinência. Supondo que o elemento x está contido em dois conjuntos, A e B, com graus de pertinência $\mu_A(x)$ e $\mu_B(x)$. A união, interseção e complemento são denotados por

$$\mu_{A\cup B}(x) = max[\mu_A(x), \mu_B(x)];$$
$$\mu_{A\cap B}(x) = min[\mu_A(x), \mu_B(x)];$$
$$\mu_{A^c}(x) = 1 - \mu_A(x).$$

Figura 2 – Operações com conjuntos *fuzzy*: (a) Conjuntos *fuzzy* A e B. (b) União de A e B. (c) Intersecção de A e B. (d) Complemento de A.



Fonte – http://condicaoinicial.com/tag/logica-fuzzy.

4.2.2 Sistema de Inferência *Fuzzy*

Segundo Klir e Yuan (1995), o conceito de Sistema de Lógica *Fuzzy* é primordial para desenvolver o raciocínio *fuzzy.* Ele faz um mapeamento não-linear de um vetor de entrada em uma saída escalar, usando um nível generalizado da regra afirmativa *Modus Ponens*. Essas regras são chamadas de implicação *fuzzy*, e são da forma: "Se x é A, então y é B". A parte da condição (SE) é chamada de antecedente ou premissa. O resultado (ENTÃO) é o consequente ou conclusão. A frase "se a pressão é alta, então o volume é pequeno" é um exemplo de regra *fuzzy*. Essencialmente, o sistema *fuzzy* é usado quando se tem um conjunto complexo de informações imprecisas, em maioria baseada no conhecimento de um especialista e necessita-se inferir acerca de uma variável resposta. Há basicamente dois tipos de sistema *fuzzy*:

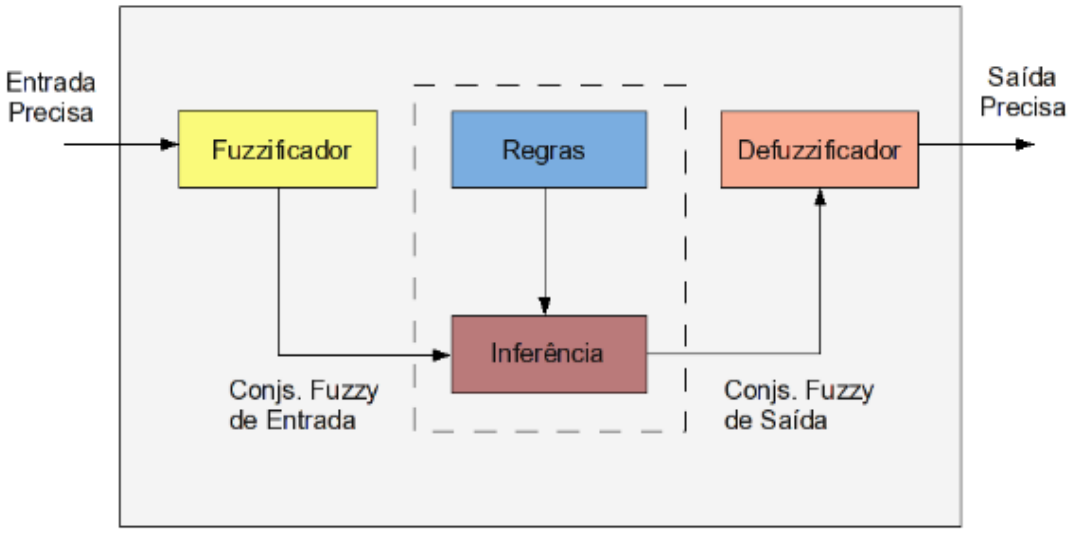
- *Standard Fuzzy System* (Sistema Padrão);
- *Functional Fuzzy System* (Sistema de Função).

Onde o primeiro é caracterizado por ter um consequente linguístico – variável de saída com níveis "pequeno", "médio" e "grande", por exemplo – e uma função de pertinência associada. No segundo, também chamado de Sistema Takagi-Sugeno, o formato do consequente é uma função arbitrária do tipo $y = f(x_i)$ – por exemplo, $y = x_1^2 + 5x_2 + 3$ – mas que também pode ser adaptado para funcionar como um sistema padrão.

Um sistema *fuzzy* pode ser construído a partir de uma ou mais variáveis de entrada e saída. Portanto podem ser classificados da seguinte forma:

- *Single Input and Single Output* - SISO;
- *Multiple Input and Single Output* - MISO;
- *Multiple Input and Multiple Output* - MIMO.

Figura 3 – Sistema de Inferência *Fuzzy*



Fonte – http://condicaoinicial.com/tag/logica-fuzzy.

Na Figura 3 são ilustradas as etapas de um sistema de inferência *fuzzy.* No *Fuzzificador* ocorre o mapeamento da entrada *crisp* (escalar) em uma função de pertinência. A *fuzzificação*, como o processo é chamado, pode ser *singleton* ou *non-singleton*. No primeiro caso, não há nenhum tipo de incerteza nas entradas e elas são escalares. Já no segundo caso, há incertezas nas entradas e elas são modeladas como números *fuzzy* (graus de pertinência). Quando a entrada consiste de mais de uma variável, ou seja, é formada por um vetor das variáveis de *input*, uma premissa é formada para representar o indivíduo ou cenário que possui todas aquelas características ao mesmo tempo. A função de pertinência da premissa é definida pela combinação das funções de pertinência de cada variável que a compõe, ou seja, ela é a interseção dos conjuntos *fuzzy* de entrada. Pode ser calculada usando o método de mínimo ou de produto:

$$\text{Método do Mínimo:} \quad \mu_{Premissa} = min\{\mu_{x_1}, \dots, \mu_{x_n}\};$$

$$\text{Método do Produto:} \quad \mu_{Premissa} = \prod \mu_{x_i}\,.$$

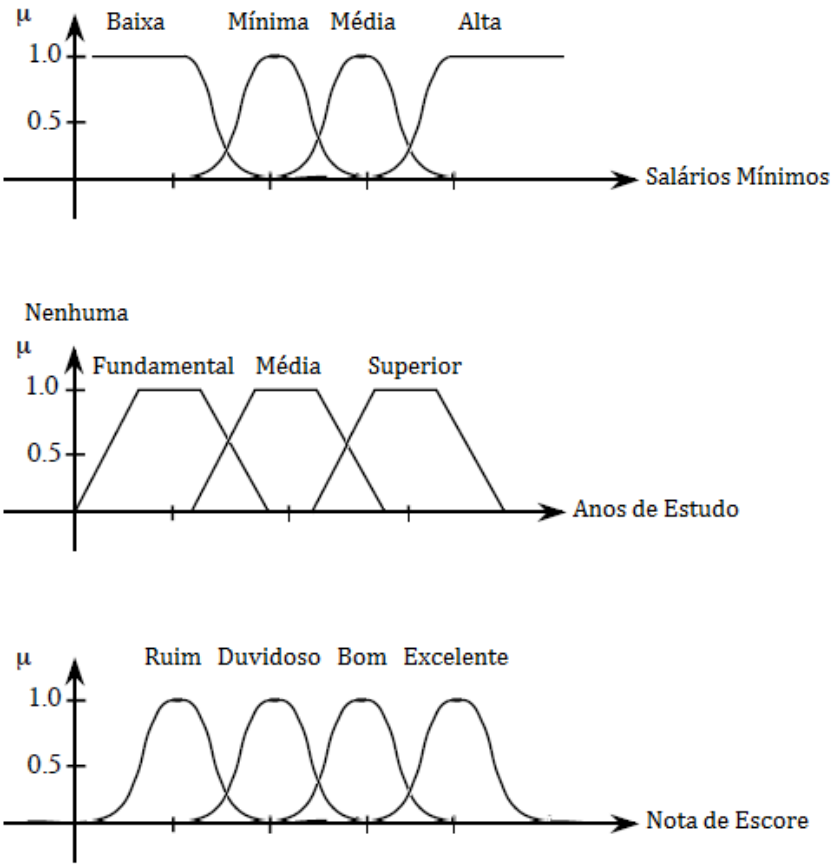
Essa abordagem é justificada pelo fato de que não é possível ter mais certeza de uma premissa do que se está certo de cada termo que a compõe, de forma individual.

No campo “Regras” está armazenada a base de conhecimento utilizada pelo Sistema de Inferência *Fuzzy.* Elas podem ser obtidas através de conhecimento especialista ou extraídas de dados numéricos, pesquisas ou resultados de modelos e são expressas na forma de estruturas SE-E-ENTÃO.

A descrição linguística das variáveis é importante para a montagem do sistema padrão, onde se atribui uma expressão que elucide as diferentes intensidades de cada categoria. Na Figura 4 são apresentadas exemplos de funções de pertinência para as variáveis:

- Renda: baixa, mínima, média, alta;
- Escolaridade: nenhuma, fundamental, média, superior;
- Avaliação do Crédito: excelente, bom, duvidoso, ruim.

Figura 4 – Funções de pertinência para Renda, Escolaridade e Avaliação do Crédito.



Fonte – a autora.

Nota-se que dependendo do tema abordado, uma vasta gama de funções de pertinência está disponível. É importante que não se confunda uma função de pertinência com uma densidade de probabilidade. Não há nenhuma definição estocástica na construção de um sistema *fuzzy* e as funções de pertinência não obedecem às leis da probabilidade.

A partir dessa definição, a construção das funções de pertinência se torna simples e a maneira como a combinação de regras influencia a variável resposta é intuitiva. Caso o nosso modelo considerasse apenas essas variáveis na decisão do crédito, pode-se ter uma base de regras como a mostrada na Tabela 2.

Tabela 2 – Base de regras para a Avaliação do Crédito dado Renda e Escolaridade.

Escolaridade	Renda			
	Baixa	Mínima	Média	Alta
Nenhuma	Ruim	Ruim	Duvidoso	Duvidoso
Fundamental	Ruim	Ruim	Bom	Bom
Média	Duvidoso	Duvidoso	Excelente	Excelente
Superior	Bom	Bom	Excelente	Excelente

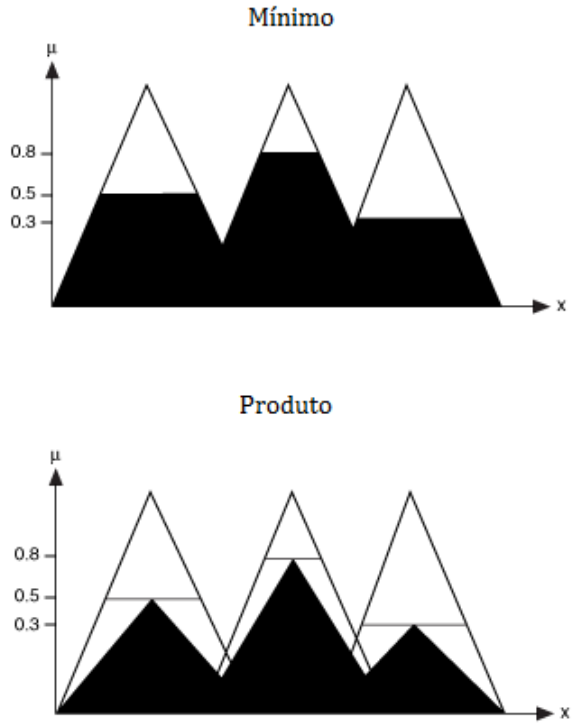
O conjunto de regras da Tabela 3 é característica de um sistema MISO, com duas variáveis de entrada e uma variável de saída, cada uma com quatro níveis, o que nos fornece $4^2 = 16$ regras distintas. Elas foram formadas respeitando os seguintes conceitos:

- Plenitude: existência de conclusões para qualquer combinação de entradas;
- Consistência: a conclusão de uma regra não entra em conflito com a conclusões obtidas por outras regras.

No entanto, num sistema real, onde se tem muita incerteza sobre o processo e às vezes mais do que duas variáveis de entrada, a base de regras disponível não conterá todas as combinações possíveis já esse número pode ser demasiadamente grande. Nesses casos, dentre as regras disponíveis, o sistema reúne e interpola aquelas que são relevantes na obtenção do consequente de uma premissa em específico. As regras que serão efetivamente usadas são aquelas que possuam pertinência não nula, ou seja, que agreguem informação no processo de inferência.

Somente nesse ponto, com a função de pertinência da premissa e as de regras relevantes para aquele cenário, que a inferência é feita, utilizando também o método do produto ou do mínimo para a implicação (Figura 5), o que nos gera um resultado *fuzzy*.

Figura 5 – Implicações *fuzzy*: métodos do mínimo e produto



Fonte – a autora.

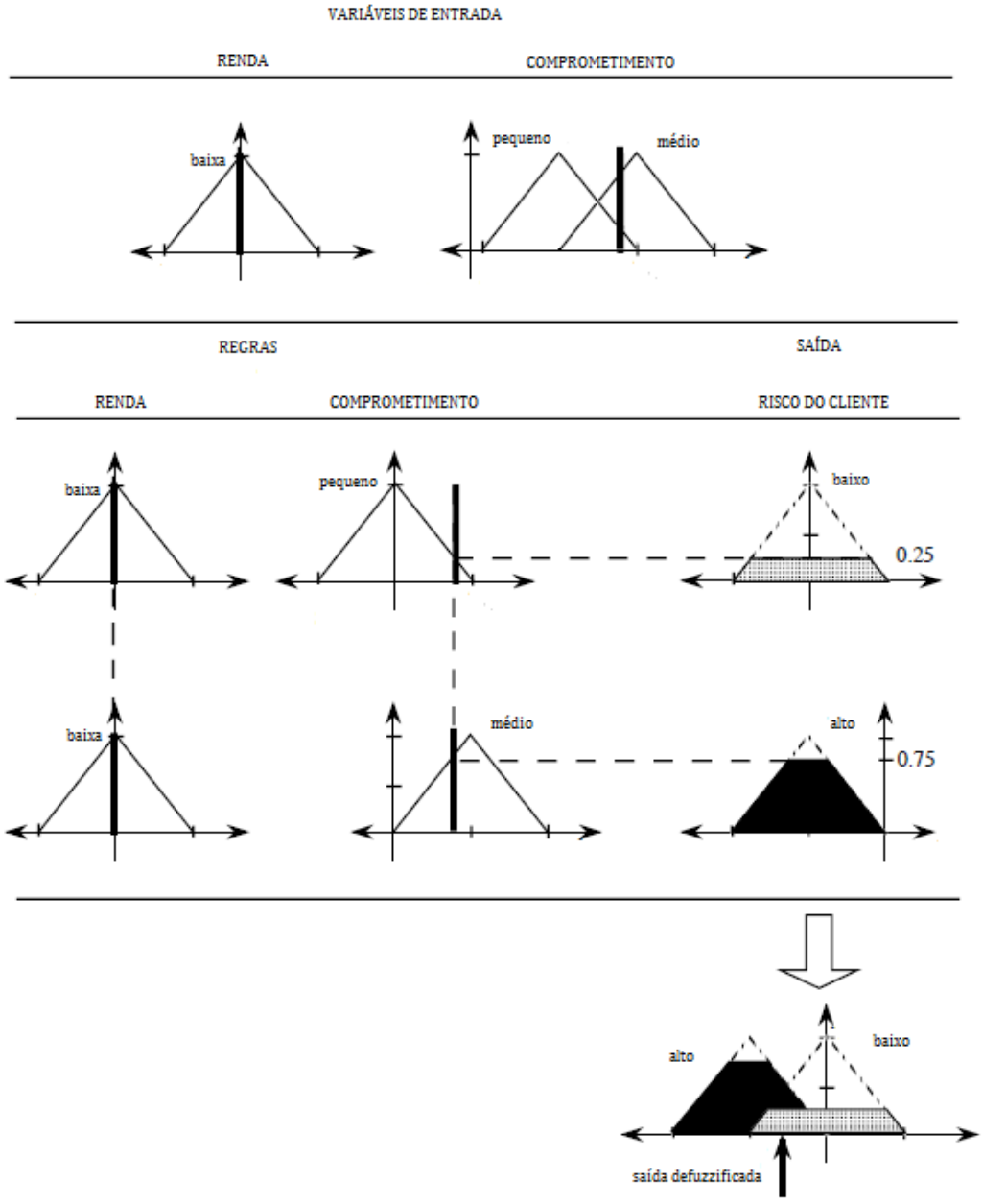
O resultado do sistema de inferência *fuzzy* deve ser transformado para a forma numérica, para que a interpretação seja direta. Existem muitos métodos de *defuzzificação*, porém apenas alguns são práticos e sua escolha é, de certa forma, subjetiva, já que não há estudos empíricos que comprovem a vantagem de um método sobre outro. Alguns exemplos são: média dos centros das premissas, centro de gravidade, *defuzzificação* por altura, centro da maior área, mais significativo dos máximos e centro máximo. (Passino e Yurkovich, 1998) Neste trabalho, foi utilizada a *defuzzificação* média dos centros, que é definida por

$$\hat{y} = \frac{\sum_{i=1}^{R} b_i \mu_{Premissa\ i}}{\sum_{i=1}^{R} \mu_{Premissa\ i}},$$

onde b_i denota o centro da função de pertinência da i-ésima premissa (ponto onde a função atinge o máximo da pertinência) e R é o número de regras ativas para a entrada que foi avaliada. A figura 6 ilustra esse processo para um sistema simples de decisão do crédito, com duas variáveis independentes, renda e comprometimento (percentual da renda que representa a parcela da dívida). Após a *defuzzificação*, a estimativa $\hat{y}$ estará na escala definida para a variável de saída e pronta para ser analisada.

A barra vertical indica o valor de entrada de cada variável. Nota-se no exemplo que a renda possui grau máximo de pertinência ao conjunto “baixa” e o comprometimento pertence a dois conjuntos *fuzzy*, com pertinências 0,25 para a categoria “pequeno” e 0,75 para a categoria “médio”. A seguir, tem-se as regras que determinam à qual categoria da variável de saída pertence a premissa. O valor da pertinência da saída é dado pelo método do mínimo.

Figura 6 – Processo de *defuzzificação*



Fonte – a autora.

Agora que todas as etapas do sistema foram definidas, é possível representá-lo através de uma função que sintetize as técnicas escolhidas para operarem em cada fase. Essa função, $f(\boldsymbol{x}|\boldsymbol{\theta})$, tem como argumentos o vetor $\boldsymbol{x}$ das entradas e o vetor de parâmetros $\boldsymbol{\theta} = \{\boldsymbol{b}, \boldsymbol{c}, \boldsymbol{s}\}$ relacionado com as regras, onde $\boldsymbol{b}$ denotam o vetor com os centros da variável de saída, $\boldsymbol{c}$ a matriz de pertinência das entradas, e $\boldsymbol{s}$ a matriz das dispersões das entradas, como demonstrado na seção 4.2.4.

4.2.3 Métodos automáticos para Sistemas *Fuzzy*

Muitas vezes a tarefa de modelar com precisão um processo natural através de um modelo matemático não-linear é muito difícil. Por vezes, quando a informação à priori sobre o processo é limitada, essa tarefa se torna impossível. No caso de modelos de crédito, reconhecer os perfis de clientes inadimplentes não é trivial. Além disso, construir uma base de regras do sistema se torna impraticável em tais situações. Felizmente, para essas situações, a modelagem *fuzzy* pode tornar a solução bem prática e pode tanto ser usada para a inferência quanto para a determinação dos parâmetros necessários (funções de pertinência e regras). Uma variedade de métodos automáticos pode ser encontrada em Passino e Yurkovich (1998): *Batch Least Squares* (BLS), *Recursive Least Squares* (RLS), *Gradient Methods* (GM's), *Clustering Methods* (CM's), *Learning from Example* (LFE) *e Modified Learning from Example* (MLFE).

Cada um deles diferencia-se basicamente pela quantidade e tipo de dado disponível para análise, sendo possível a combinação de dois ou mais métodos na construção de cada etapa do processo de inferência. Os métodos baseados em mínimos quadrados exigem um empenho computacional alto, e por isso tonam-se inviáveis para bancos de dados grandes, além de necessitarem que as funções de pertinência e base de regras estejam definidas. Já os métodos de gradiente são muito úteis na melhoria da *performance* do sistema quando combinados a outros métodos, abrangendo técnicas de aproximação como a de Newton e de Gauss-Newton. O método de agrupamento tem proposta similar à dos GMs, porém utilizando técnicas como a do Vizinho Mais Próximo e *c-means*. Com o LFE é possível extrair as regras de uma base de dados determinada para tal, chamada de base de treinamento, desde que as funções de pertinência sejam especificadas. O MLFE é ainda menos rígido nas exigências, pois permite a montagem da base de regras e a especificação das funções de pertinência a partir do banco de

treinamento, além de não exigir tanto esforço computacional, tonando-se a opção mais atrativa para desenvolvimento e aplicação nesse trabalho.

4.2.4 Algoritmo MLFE

Esse algoritmo possibilita tanto o cálculo dos parâmetros que definem as funções de pertinência quanto a formação da base de regras para o sistema *fuzzy,* a partir de dados a respeito do processo estudado, obtidos por meio de observação ou fornecidos por um especialista. A única pré-definição a ser feita é o tipo da função de pertinência (triangular, gaussiana, etc) e o algoritmo procura adaptar os parâmetros para que as regras representem os dados com mais eficiência.

Como o algoritmo é automático e iterativo, é importante que a escolha das variáveis seja feita com cautela, evitando problemas como perda de acurácia na estimativa ou superparametrização do modelo.

Primeiramente determinam-se as variáveis de entrada e saída que são de interesse para o estudo. Para a construção da base de regras, adota-se a primeira observação do banco de dados de treinamento como primeira regra. De posse de uma única regra, para uma nova observação, calcula-se a diferença entre a estimativa obtida e o resultado real:

$$|f(\boldsymbol{x}|\boldsymbol{\theta})_i - y_i|,$$

onde $f(\boldsymbol{x}|\boldsymbol{\theta})_i$ é a estimativa da i-ésima observação retornada pelo sistema e y_i é o i-ésimo resultado observado. Se a diferença entre elas superar a tolerância, a nova observação vira regra e o sistema "aprende" a partir dela, caso contrário, o perfil estudado possui equivalências no banco de regras e pode ser estimado a partir dele.

Os parâmetros centro e dispersão das funções de pertinência das entradas e centro da premissa são adicionados ao banco de regras de forma que uma regra não distorça o que outra já havia aprendido e que haja uma sobreposição suave entre as pertinências dos pontos de treinamento (Tabela 3). Uma vez que se obtém a base de regras, as estimativas podem ser calculadas para avaliação da eficiência do modelo *fuzzy*.

Tabela 3 – Formação da Base de Regras a partir do Banco de Treinamento

Entradas	Saída
$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$	$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$
Base de Regras	
Centro e dispersão das regras	Centro da premissa
$\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{R1} & c_{R2} & \cdots & c_{Rm} \end{bmatrix}$ $\begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ s_{R1} & s_{R2} & \cdots & s_{Rm} \end{bmatrix}$	$\begin{bmatrix} b_1 \\ \vdots \\ b_R \end{bmatrix}$

Para o evento estudado por este trabalho, o modelo adotado é MISO e a função utilizada na geração das regras e estimação da saída é dada por:

$$f(\boldsymbol{x}|\boldsymbol{\theta}) = \frac{\sum_{i=1}^{R} b_i \, min_j \left\{ exp\left(-\frac{1}{2}\left(\frac{x_j - c_{ij}}{s_{ij}} \right)^2 \right) \right\}}{\sum_{i=1}^{R} min_j \left\{ exp\left(-\frac{1}{2}\left(\frac{x_j - c_{ij}}{s_{ij}} \right)^2 \right) \right\}}, \qquad j = 1, \dots, m \qquad (1)$$

onde:

R é o número de regras disponíveis;

x_j é o valor da j-ésima variável do vetor linha de entrada da nova observação;

c_{ij} e s_{ij} são o centro e dispersão da função de pertinência da j-ésima variável, i-ésima regra;

b_i é o centro da função de pertinência da premissa da i-ésima regra.

A *fuzzificação* utilizada é do tipo *singleton*, as funções de pertinência para todas as variáveis são gaussianas, regra do mínimo para premissas e implicação e *defuzzificação* por média dos centros. Para mais detalhes na construção do MLFE e demais algoritmos, consultar Passino e Yurkovich(2001).

4.3 Medidas de Avaliação

A análise do poder de discriminação do modelo é fundamentada no uso de indicadores de desempenho. É importante ressaltar que a avaliação da *performance* do modelo segundo seu poder de discriminação é diferente de avaliar sua adequação. Para um mesmo conjunto de dados é possível ajustar vários modelos utilizando diferentes variáveis e técnicas. Só então as capacidades de discriminação de cada um são medidas

e posteriormente adota-se o melhor deles, baseando a decisão nas necessidades da empresa.

Segundo Sicsú (2010), não é indicado basear a análise em uma única medida, por isso, é prática comum as instituições adotarem pelo menos dois dos vários indicadores disponíveis. A seguir, são mostradas algumas dessas medidas (BCBS, 2005). Denota-se os tomadores adimplentes da amostra por C_a, os inadimplentes, C_i, suas respectivas estimativas (dadas pelo modelo) $\widehat{C}_a$ e $\widehat{C}_i$ e a nota de escore por k.

4.3.1 Teste de Kolmogorov-Smirnov (KS)

O objetivo dessa medida é obter a maior distância entre as funções de distribuição acumulada dos escores dos tomadores adimplentes e inadimplentes, $F_a(k)$ e $F_i(k)$, respectivamente,

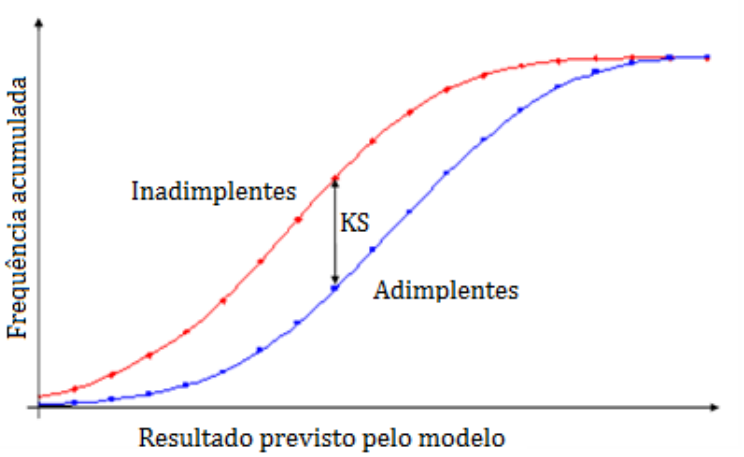
$$F_a(k) = \frac{\widehat{C}_a \ com\ k \leq k_0}{C_a} \qquad \text{e} \qquad F_i(k) = \frac{\widehat{C}_i \ com\ k \leq k_0}{C_i},$$

com k assumindo todos os valores do conjunto de possíveis escores. Calcula-se então a maior diferença entre as funções

$$KS = máx[F_a(k) - F_i(k)].$$

Nesse caso, quanto maior a distância entre as distribuições, melhor a discriminação do modelo, como pode ser mostrado na Figura 7.

Figura 7 – Teste KS



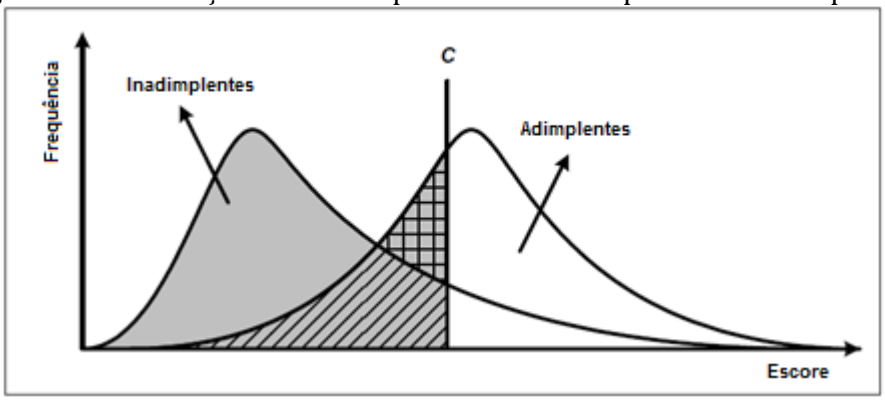
Fonte – adaptado de Selau, 2008.

4.3.2 Área abaixo da curva ROC (AUROC)

A construção da curva ROC (*Receiver Operating Characterisctic*) baseia-se nos conceitos de sensibilidade e especificidade e é ilustrada pela Figura 8, onde são mostradas as distribuições de escore para tomadores adimplentes e inadimplentes. A sensibilidade do modelo é definida pela proporção de clientes adimplentes que foram classificados corretamente. A especificidade é definida como a proporção de clientes inadimplentes que foram classificados corretamente. O desejado de um modelo é que se tenha alta sensibilidade e especificidade.

Caso o modelo fosse ideal, as distribuições de escores de adimplência e inadimplência seriam separadas, mas em uma situação real, a discriminação perfeita é impossível e as duas distribuições se sobrepõem em algum ponto.

Figura 8 – Distribuição de escores para clientes adimplentes e inadimplentes



Fonte – adaptado de BCBS, 2005.

Para que uma decisão seja tomada, um ponto de corte C é adotado, como mostrado na Figura 8, portanto os indivíduos com escore menor do que C são considerados inadimplentes, enquanto os que possuem escores maiores serão classificados como adimplentes. A sensibilidade e a especificidade dependem do valor do ponto de corte. Quando se aumenta o ponto de corte, a sensibilidade diminui e a especificidade aumenta. Dessa forma, a Tabela 4 ilustra as quatro situações possíveis, considerando que o evento seja a inadimplência.

Tabela 4: Classificação das decisões.

Decisão	Cliente	
	Inadimplente	Adimplente
Recusa crédito	Correta	Erro tipo I
Aprova crédito	Erro tipo II	Correta

A taxa de acerto, ou sensibilidade, T_{Ac} é definida por

$$T_{Ac} = \frac{\widehat{C}_i}{C_i},$$

onde $\widehat{C}_i$ é o número de adimplentes previstos corretamente pelo modelo e C_i é o número total de adimplentes no grupo, para o escore C.

A taxa de alarmes falsos T_{FA} é definida por

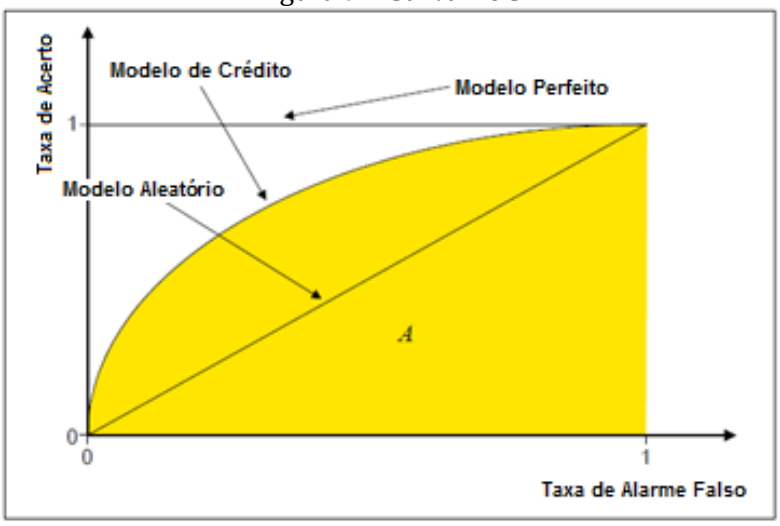
$$T_{FA} = \frac{\widehat{C}_{aerro}}{C_a},$$

onde $\widehat{C}_{aerro}$ é o número de clientes adimplentes que foram classificados incorretamente como inadimplentes dado determinado escore C. O número total de clientes adimplentes é dado por C_a.

Assim tem-se que $T_{Ac}(C)$, é a área sob a distribuição dos escores de clientes inadimplentes e à esquerda do valor de corte C. Logo, para cada ponto de corte C, são calculadas as taxas de acerto e de falso alarme. A curva ROC é um gráfico de T_{Ac} *versus* T_{FA} (Figura 9) ou Sensibilidade *versus* 1-Especificidade.

Teoricamente, é possível obter infinitos pontos de T_{Ac} e T_{FA}. Na prática, dispõe-se uma amostra de tamanho limitado, logo, a curva ROC é obtida através da interpolação linear desse conjunto de pontos. Considera-se o modelo com maior poder de discriminação quanto mais íngreme for a curva ROC. Portanto, quanto maior o valor AUROC (*Area Under* ROC), melhor a *performance* do modelo. Alguns *softwares* estatísticos já fornecem o valor AUROC.

Figura 9 – Curva ROC

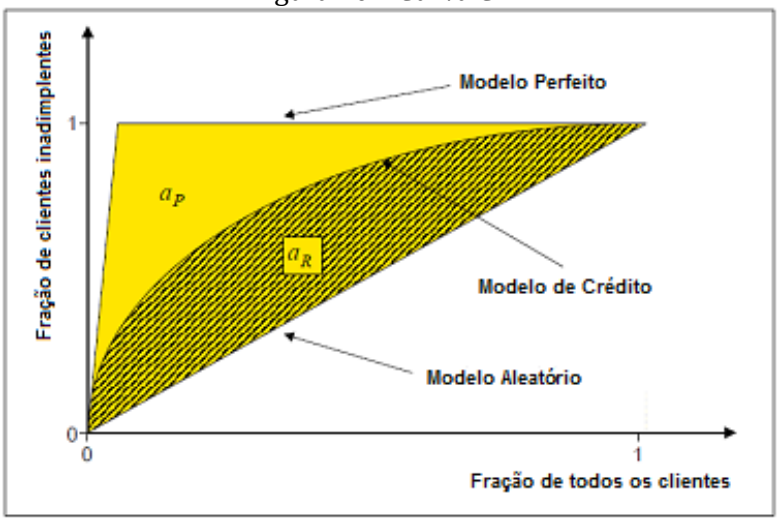


Fonte – adaptado de BCBS, 2005.

4.3.3 Razão de Acurácia (AR)

Para a obtenção da razão AR, primeiramente constrói-se a curva CAP (*Cumulative Accuracy Profile*). Os tomadores são classificados do menor para o maior escore, ou seja, do perfil mais arriscado para o mais confiável. Para a fração de clientes com escore até k, afere-se a proporção em situação de *default* (Figura 10).

Figura 10 – Curva CAP



Fonte – adaptado de BCBS, 2005.

Um modelo de discriminação perfeito seria capaz de atribuir os menores escores aos tomadores inadimplentes, enquanto que os clientes adimplentes receberiam os escores mais altos. Para um modelo sem qualquer poder discriminativo, a fração x de tomadores com os menores escores conterão $x\%$ do total de inadimplentes. Modelos

que avaliam situações reais estão entre esses dois extremos. A qualidade do modelo é medida pela razão AR, definida por

$$AR = \frac{a_R}{a_P},$$

onde a_P denota a área que um modelo perfeito ocuparia no gráfico e a_R é a área ocupada pelo modelo real. Logo, o modelo de crédito é considerado mais eficiente o quão próximo de 1 for a razão AR.

4.3.4 Escore de Brier (BRIER)

O escore de Brier é um método que avalia a qualidade de previsão de uma probabilidade e teve sua origem no campo de pesquisas meteorológicas. No entanto, pode ser aplicado diretamente na avaliação de modelos de risco.

Seja $p_0, p_1, \dots, p_k$ as probabilidades de *default* estimadas dos clientes inadimplentes nas k classes de escore. O Escore de Brier é definido por

$$B = \frac{1}{n}\sum_{j=1}^{n}(p_j - \theta_j)^2,$$

onde n denota a quantidade de clientes avaliados, p_j é a probabilidade estimada de *defaut* do j-ésimo cliente e θ_j é definido por

$$\theta_j = \begin{cases} 1, se\ há\ inadimplência \\ 0, caso\ contrário \end{cases}$$

Pela definição dada acima, segue que o escore de Brier está sempre entre zero e um. Quanto mais próximo de zero, melhor a estimativa das probabilidades. A desvantagem dessa medida é a queda da *perfomance* para probabilidades muito pequenas.

4.3.5 Distância de Mahalanobis (DM)

Conforme comentado em Santos (2002), o modelo com melhor desempenho é aquele que apresenta maior concentração de tomadores adimplentes com escores altos e inadimplentes com escores baixos. Pode-se então comparar os escores médios dos dois perfis de tomadores, levando em consideração a variabilidade dos dados. Nota-se que, dependendo da técnica utilizada para definir o escore, esse pode variar em um intervalo de valores muito diferentes. Distância de Mahalanobis é definida por

$$DM^2 = \frac{\left(\bar{k}_a - \bar{k}_i\right)^2}{S},$$

onde

$$S = \frac{n_a S_a^2 + n_i S_i^2}{n_a + n_i - 2},$$

e $\bar{k}_a$, $\bar{k}_i$ e S_a^2, S_i^2 denotam respectivamente a média e variância estimada dos escores médios dos indivíduos adimplentes e inadimplentes. O número de clientes adimplentes é n_a, o de inadimplentes é n_i.

A discriminação do modelo será melhor quanto maior for o valor da distância de Mahalanobis. Essa medida não possui intervalo de variação limitado, ou seja, varia entre zero e infinito.

5 METODOLOGIA

O banco de dados utilizado nesse trabalho consiste de propostas de concessão de crédito na modalidade parcelado e foi fornecido por uma instituição financeira atuante no mercado brasileiro com vasta experiência em concessão de crédito. Para proteger a idoneidade da instituição e facilitar seu manuseio, o banco de dados teve todas as variáveis categorizadas e não identificadas. Os indivíduos contidos no banco foram classificados de acordo com o modelo de *Credit Scoring* utilizado na instituição na época da concessão dos créditos, desenvolvido por meio da metodologia de Regressão Logística.

A comparação de desempenho entre esse modelo e o Sistema *Fuzzy,* é constituída basicamente por medidas diretas como as taxas de erro e acerto na predição e pelas medidas de discriminação propostas, que são as utilizadas na prática bancária para monitoramento de modelos de *Credit Scoring*, conforme Filho e Sleegers (2010).

Para a construção do Sistema *Fuzzy*, utilizou-se as mesmas variáveis do modelo de regressão logística, adotando-se inclusive, as mesmas categorias. Para o desenvolvimento do algoritmo MLFE, empregou-se apenas o *software* estatístico SAS versão 9.2, onde o sistema *fuzzy* proposto foi descrito em linguagem SAS/IML. As medidas também foram calculadas através de planilhas eletrônicas, SAS e R.

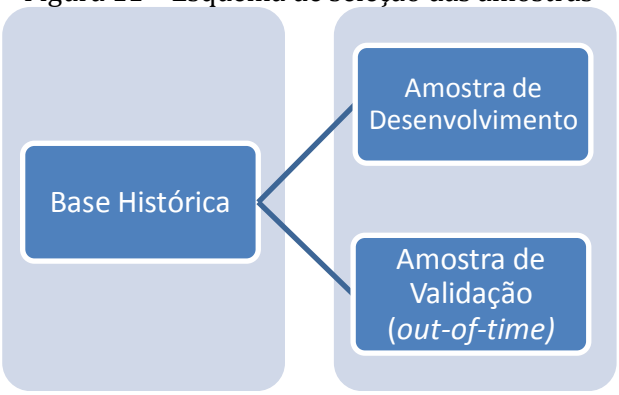
5.1 Seleção da amostra

A base de dados disponível possui informações reais de concessão de recursos, cujo produto oferecido aos tomadores corresponde a uma linha de crédito sem destinação específica, com limite pré aprovado e disponibilizado automaticamente na conta do cliente. O empréstimo tem taxa pré-fixada de prestações mensais e sucessivas calculadas pelo Sistema de Amortização Francês com vencimento escolhido pelo cliente quando da efetivação da transação. Os dados coletados referem-se à base histórica do período de setembro de 2007 a julho de 2009. As amostras foram retiradas considerando dois períodos de tempo distintos:

- Amostra de desenvolvimento: contratos firmados de setembro de 2007 a julho de 2008;
- Amostra de validação: contratos firmados de agosto de 2008 a julho de 2009.

De acordo com BCBS (2005), modelos de crédito são bastante sensíveis à escolha da amostra de validação. Para evitar dependência entre amostras, modelos quantitativos devem ser construídos e validados usando amostras transversais no tempo e universo. Uma amostra com proporção de *default* muito baixa diminui o poder do teste, aumentando assim a ocorrência do erro tipo I. Com a intenção de evitar esse problema e obter um conjunto de regras representativo tanto dos perfis adimplentes quanto dos inadimplentes, a amostra de desenvolvimento foi tomada de forma que houvesse proporção igual dos dois perfis. O esquema de amostragem é representado pela Figura 11.

Figura 11 – Esquema de seleção das amostras



Fonte – a autora.

Na fase de testes computacionais, foi verificado que as condições de Plenitude e Consistência da base de regras foram satisfatoriamente alcançadas quando estabelecida a partir de uma amostra de tamanho 7.000, da onde foram geradas 6.199 regras respeitando a tolerância adotada. A amostra de validação possui 30.000 observações.

Os escores obtidos em ambos modelos, CS e *Fuzzy*, é dado no intervalo [0,1]. No entanto, é prática comum das instituições trazer o escore para a casa das centenas e até dos milhares, com o objetivo de facilitar sua leitura. Aqui, a nota foi trazida para o intervalo [0,1000], seguindo o padrão utilizado pela instituição. Baseando-se no fato que a saída originalmente é um número entre zero e um, escolheu-se o limite de tolerância 0,2. O critério de escolha foi subjetivo, procurando apenas maximizar a quantidade de acertos no mínimo de tempo de processamento computacional.

5.2 Desenvolvimento do Sistema *Fuzzy*

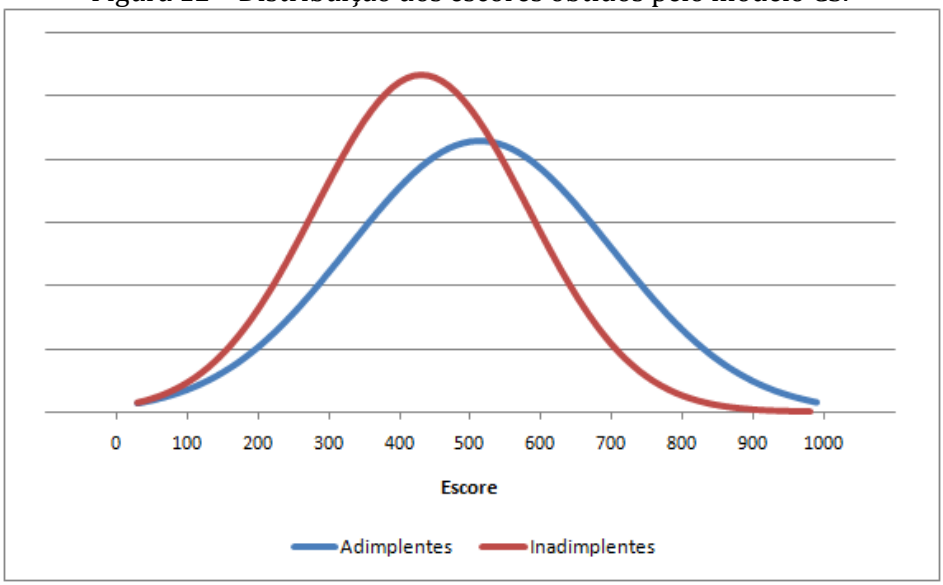
Utilizando o algoritmo MLFE, a amostra de desenvolvimento foi usada para a construção da base de regras. A adequação do modelo foi avaliada através da amostra de validação. Para ambos procedimentos, o sistema MISO determinado pela função (1) foi aplicado. As entradas do sistema consistem nas 18 variáveis categorizadas que são usadas pela instituição no modelo logístico atual. A saída é uma variável com domínio entre zero e um, que atribui um escore ao tomador. Os diagnósticos foram alcançados baseados no desempenho do modelo, segundo o ponto de corte adotado para o escore *fuzzy*.

A programação desenvolvida em SAS/IML se encontra no Apêndice A.

6 RESULTADOS

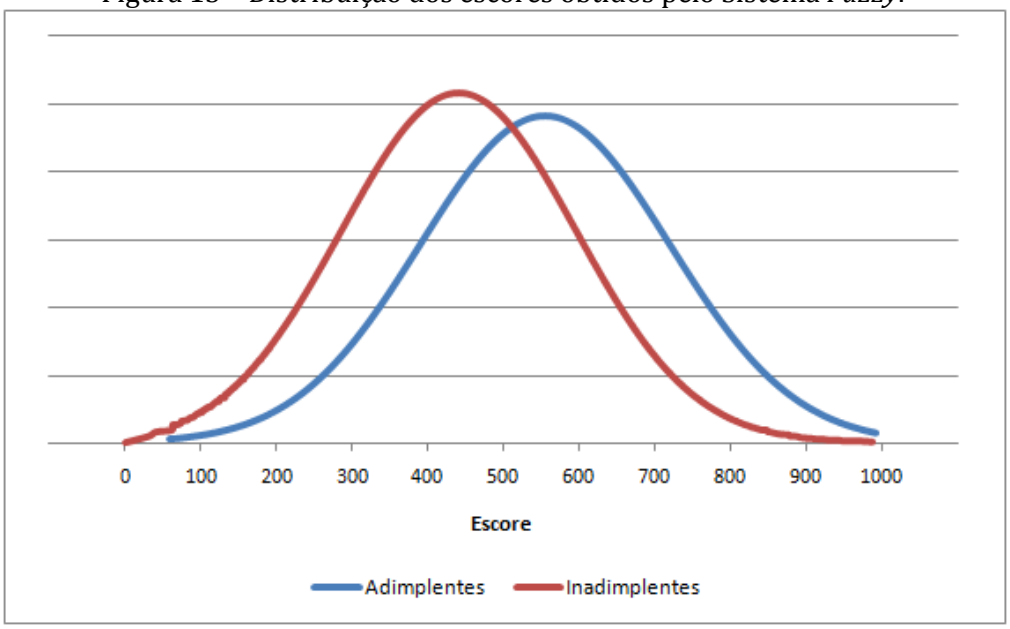
Inicialmente, os resultados das duas metodologias são apresentados em um formato de gráfico que é prática da instituição financeira. As distribuições dos escores para os tomadores adimplentes e inadimplentes são apresentadas na forma de distribuições normais com média e variância extraídas dos resultados. Porém, isso não quer dizer que a distribuição do escore seja necessariamente Normal. Da análise das Figuras 12 e 13, percebe-se que as distribuições de adimplentes e inadimplentes estão bastante sobrepostas em ambas abordagens e nos dois casos o escore médio dos adimplentes é mais alto.

Figura 12 – Distribuição dos escores obtidos pelo modelo CS.



Fonte – a autora.

Figura 13 – Distribuição dos escores obtidos pelo Sistema *Fuzzy*.



Fonte – a autora.

Nota-se também, que a diferença entre as médias das distribuições de adimplentes e inadimplentes é maior no Sistema *Fuzzy*, sugerindo maior poder de discriminação entre as duas categorias de cliente.

6.1 Avaliação com base no *Credit Score*

O modelo empregado no cálculo do escore de crédito foi construído a partir uma base mais antiga do que a disponibilizada para esse estudo, portanto há o risco de que tenha sofrido perda de adequação, o que poderia distorcer os diagnósticos citados nessa seção. Esse risco é minimizado pela utilização de rotinas de acompanhamento do desempenho do modelo, que buscam identificar alterações significativas na distribuição utilizada no desenvolvimento do modelo com a avaliada.

Os indicadores utilizados para essa finalidade demonstram que o modelo não sofreu alterações que justifiquem intervenções para ajustar perda de estabilidade e/ou capacidade de discriminação. Além disso, os parâmetros adotados pela instituição financeira para aceitação de clientes, tal como definição de ponto de corte e demais políticas, não estão retratados nesse trabalho.

O ponto de corte adotado para análise foi escolhido de forma que maximizasse a taxa de acertos na predição, como é visto na Tabela 5. Dessa forma, tomadores com nota

escore até 400 seriam ditos inadimplentes, enquanto que os restantes, inadimplentes. A decisão do ponto de corte é fundamental para avaliar os indicadores de modelos de *credit scoring*, porque é a referência para determinar a capacidade de acerto do modelo. Por isso nesse trabalho adotou-se o valor que maximizasse a taxa de acerto. Essa abordagem, no entanto, não leva em conta o custo dos erros, nem o critério de maximização de rentabilidade que em muitos casos são utilizados por instituições financeiras na decisão de concessão de crédito.

Tabela 5: Avalição do modelo de *Credit Scoring*

Estimado pelo modelo	Perfil real		Total
	Inadimplente	Adimplente	
Inadimplente	5.357	5.248	10.605
Adimplente	6.527	12.868	19.395
Total	11.884	18.116	30.000

Para o modelo de CS desenvolvido com base na Regressão Logística, a taxa de classificação correta de adimplência, ou sensibilidade, é de 71,03% enquanto que a especificidade é de 45,07%. A taxa global de acertos na estimativa é de 60,75%.

Na Tabela 6 são apresentados os valores das medidas de avaliação. Uma vez que não se teve acesso às referências adotadas pela instituição, optou-se por adotar os valores propostos por Filho e Sleegers (2010) em um trabalho publicado pela revista Tecnologia de Crédito, da Serasa *Experian* e são encontrados no Anexo A.

Tabela 6: Avalição do modelo de *Credit Scoring*

	KS	AR	AUROC	DM	BRIER
Medida	21,33	0,35	0,63	0,48	0,32
Classificação	Baixa	Aceitável	Baixa	Baixa	-

Nota: Não foram encontrados valores de referências para o escore de Brier.

Os resultados da avaliação apontam que o modelo por Regressão Logística apresentou baixa capacidade de discriminação, sendo que somente o indicador AR, obteve classificação aceitável.

6.2 Avaliação com base no Sistema *Fuzzy*

Para o sistema *fuzzy,* foi adotado o ponto de corte que maximizou os acertos nas estimativas, de acordo como mostrado na Tabela 7. Assim sendo, indivíduos com escores acima de 400 são classificados como adimplentes e o restante como inadimplentes. Nesse caso, o corte que maximizasse a taxa de predição correta, era um valor esperado, já que observou-se na amostra essa proporção de inadimplência.

Tabela 7: Avalição do modelo *Fuzzy*

Estimado pelo modelo	Perfil real		Total
	Inadimplente	Adimplente	
Inadimplente	5.104	3.570	8.674
Adimplente	6.780	14.546	21.326
Total	11.884	18.116	30.000

Nota-se que, para a configuração adotada para o sistema *fuzzy,* a sensibilidade caiu em relação ao modelo de CS, apresentando taxa de 42,95%. Em compensação, a especificidade aumentou para 80,29%. O modelo apresentou taxa global de acertos de 65,50%.Com base nos critérios usados (Tabela 8), o poder discriminativo do sistema *fuzzy* obteve avaliação satisfatória, com classificação "boa" para a medida AR e "aceitável" para as demais.

Não somente pela classificação, mas também considerando a diferença entre os valores obtidos para todos os indicadores, observa-se que o sistema *fuzzy* apresentou resultados bem superiores ao da Regressão Logística, conforme os parâmetros utilizados nesse estudo.

Tabela 8: Avalição do modelo *Fuzzy*

	KS	AR	AUROC	DM	BRIER
Medida	29,36	0,54	0.69	0,71	0,34
Classificação	Aceitável	Boa	Aceitável	Aceitável	-

Nota: Não foram encontrados valores de referências para o escore de Brier.

Como dito anteriormente, a variável de saída do sistema *fuzzy* não tem interpretação no universo das probabilidades, porém esse conceito foi extrapolado no cálculo do Escore de Brier, somente a título de curiosidade.

7 CONCLUSÃO

A compreensão humana da maioria dos processos é em grande parte baseada em conceitos imprecisos do nosso raciocínio. Essa imprecisão, quando comparada às quantidades exatas necessárias para que um computador atinja a mesma resposta, é sem dúvida uma informação que, se empregada da maneira correta, é de grande utilidade. A habilidade de incorporar tal raciocínio em problemas que até então são considerados intratáveis e complexos, é a característica que lógica *fuzzy* uma ferramenta eficaz.

Essa eficácia foi mostrada no diagnóstico apresentado na seção anterior, que mostra melhor desempenho geral do Sistema *Fuzzy,* de acordo com as medidas adotadas. Além de o Sistema *Fuzzy* ter apresentado melhor desempenho em acertos, as medidas, em geral, também apontam para o melhor poder de discriminação desse em relação ao modelo de *Credit Scoring.*

A construção do Sistema *Fuzzy* em linguagem SAS/IML permite a fácil conversão para outras linguagens de programação, além de admitir a incorporação de outras abordagens, adaptando-se totalmente ao tipo de dado disponível e aos objetivos que a instituição almeja atingir.

Para que os resultados sejam ainda mais satisfatórios, é necessário testar a influência de cada variável na estimação da resposta, tanto em separado quanto em grupos, e assim decidir qual é a melhor configuração para o sistema. Preferencialmente, esses testes devem ser conduzidos usando as variáveis quantitativas em sua forma original, o que permite agregar ao sistema o máximo de informação durante a inferência.

Por se tratar de um algoritmo que não necessita de um *software* específico para funcionar, o Sistema *Fuzzy* desenvolvido pode ser implementado em um sistema real de decisão de crédito para agilizar e amparar as decisões tomadas. O fato do ponto de corte adotado não levar em conta os custos que envolvem essa decisão é um fator que deve ser estudado com profundidade quando da implantação do sistema.

Como sugestão para trabalhos futuros, visando a melhora da *performance* do sistema, a inferência Bayesiana poderia ser utilizada quando da ausência de conhecimento especialista. Além disso, pode-se implementar algoritmos híbridos, ou seja, combinação de dois ou mais algoritmos automáticos na formação da base de regras

e até a combinação do sistema *fuzzy* com técnicas estatísticas. Outra saída seria testar a sensibilidade do sistema para diferentes funções de pertinência e empregar testes de exaustão para obter parâmetros mais adequados para a definição das funções de pertinência e tolerância.

REFERÊNCIAS

AI ACESSS. *Mahalanobis Distance.* Disponível em:
<http://www.aiaccess.net/English/Glossaries/GlosMod/e_gm_mahalanobis.htm>.
Acessado em: 3 jun. 2011.

AYEGÜL, I. *Credit Scoring Methods and Accuracy Ratio.* The Middle East Technical University, 2005.

BCBS - BASEL COMMITTEE ON BANKING SUPERVISION, *Working Paper n°14: Studies on the Validation of Internal Rating Systems.* Bank for International Settlements: 2005.

BRIER, G. W. Verification of Forecasts Expressed in Terms of Probability*. Monthly Weather Review,* abr. 1950. Disponível em: <http://docs.lib.noaa.gov/rescue/mwr/078/mwr-078-01-0001.pdf>. Acessado em: 2 mai. 2011.

CESAR, B. L.; MACHADO, M. A. S.; JUNIOR, H. A. O. Sistema de Inferência *Fuzzy* na Análise de Crédito Pessoal. *Revista Pesquisa Naval,* Brasília n. 18, p. 84-91, nov. 2005.

FARIA, M. P. C. *Análise de Crédito à Pequena Empresa: um Modelo de Escoragem Baseado nas Metodologias Estatísticas: Análise Fatorial e Lógica Fuzzy.* Rio de Janeiro: IBMEC, 2006.

FAWCETT, T. *An Introduction to ROC Analysis.* Elsevier, 2005.

FILHO, H. S.; SLEEGERS, L. C. Valores de Referência para os Principais Indicadores de Acurácia dos Modelos de Escoragem. *Serasa Experian Tecnologia de Crédito,* São Paulo, n. 73, p. 31-45, ago. 2010.

HOSMER, D. W., LEMESHOW, S. *Applied Logistic Regression.* Wiley, 2000.

KLIR, G. J.; YUAN, B. *Fuzzy Sets and Fuzzy Logic – Theory and applications*. Prentice Hall, 1995.

LOPES E FILHO ASSOCIADOS. O Novo Acordo de Capital da Basiléia (Basiléia II). *Boletim Risk Bank*, 2002. Disponível em: <http://www.riskbank.com.br/anexo/basileia2.pdf>. Acessado em: 15 mai. 2011.

MACHADO, A. R. *Modelos Estatísticos para Avaliação de Risco em Produtos de Crédito.* Brasília: UnB, 2010.

PASSINO, K. M.; YURKOVICH, S. *Fuzzy Control.* Addison-Wesley, 1998.

R Development Core Team. *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing, 2010.
URL http://www.R-project.org.

ROSS, T. J. *Fuzzy Logic with Engineering Applications.* Wiley, 2004.

SANTOS, S. K. Y. *Aplicação de Árvores de Decisão à Análise de Concessão de Crédito.* Curitiba: UFPR, 2002.

SAS Institute Inc. *SAS ® 9.2 Intelligence Platform: System Administration Guide*, ed. 2. Cary, NC, USA: SAS Institute Inc, 2011.

SELAU, L. P. R. *Construção de Modelos de Previsão de Risco de Crédito.* Porto Alegre: UFRGS, 2008.

SICSÚ, A. L. *Credit Scoring: Desenvolvimento, Implantação, Acompanhamento*. São Paulo: Blucher, 2010.

SILVA, J. P. *Gestão e Análise de Risco de Crédito.* São Paulo: Atlas, 2008.

SOBEHART, J. R.; KEENAN, S. C.; STEIN, R. M. Benchmarking Quantitative Default Risk Models: A Validation Methodology. *Algo Research Quarterly*, v. 4, p. 57-71, mar. 2001.

SOUZA, S. A. O. Alguns Comentários Sobre a Teoria *Fuzzy. Revista Online Exacta*, 2003. Disponível em: <http://portal.uninove.br/marketing/cope/pdfs_revistas/exacta/exacta_v1/exactav1_suzanaabreu.pdf>. Acessado em: 3 mai. 2011.

APÊNDICE A – Programação desenvolvida

```
/*********************************************************************/
/*      MLFE Modified Learning From Example Fuzzy System Method      */
/*                          Written by Lore Martins Bueno            */
/* Standard MISO Fuzzy System                                        */
/* Gaussian MF for inputs and output                                 */
/* Singleton Fuzzification                                           */
/* Minimum/Multiplication Method for premise and implication         */
/* Center Average Defuzzification                                    */
/*                                                                   */
/*********************************************************************/


/**********************SUB FUNCTIONS*************************/
title 'Gauss w=4 sigma=0.5 s=0.6 ef=0.2';
proc iml;
*GAUSS MEMBERSHIP FUNCTION;
start gauss(x,c,s);
      mu=exp(-(((x-c)/s)**2)/2);
return (mu);
finish gauss;

*TRIANGULAR MEMBERSHIP FUNCTION;
start triang(x,c,s);
            if x<=c then mu=max(0,(1+(x-c)/s));
            else mu=max(0,(1+(c-x)/s));
return (mu);
finish triang;

*PRODUCT FOR PREMISE AND IMPLICATION;
start prod(x);
xcol=j(nrow(x),1,1);
      do i=1 to nrow(x);
            do j=1 to ncol(x);
                  xcol[i]=xcol[i]#x[i,j];
            end;
      end;
return (xcol);
finish prod;

/*********************************************************************/
/*  SYSTEM INPUTS                                                    */
/*  MISO SYSTEM                                                      */
/*  n input variables (X1,...,Xn), 1 output variable (Y).            */
/*  Tolerance for fuzzy system output 'ef'.                          */
/*  Weight factor (overlap between MF of rules) 'W'.                 */
/*  Initial spread for MF of premises of the first rule 's'.         */
/*  Training Set 'G'.                                                */
/*                                                                   */
/*********************************************************************/

*PARAMETERS;
*Input Training Set ('in-time') MODEL CONSTRUCTION;
use training var{
var19 var20 var21 var22 var23 var24 var25 var26 var27 var28 var29 var30
var31 var32 var33 var34 var35 var36};
      read all into x;
use training var{cliente};
      read all into y;
```

```
ef=0.2;
w=4;
sigma=0.5;

/*********************************************************************/
/* MLFE-Develop both membership functions and rules using training set*/
/*********************************************************************/

*1st STEP - Define 1st data set as 1st rule;
B=Y[1,]; *1st Consequent = center for 1st output MF;
C=X[1,]; *MF Center designed by rule R;
S=j(1,ncol(x),1)*sigma; *Spread designed by rule R;
mu=j(1,ncol(x),1); *Membership values matrix;
*ADDING NEW RULES TO RULE BASE;

do i=2 to nrow(x);
      do j=1 to ncol(x);
            do k=1 to nrow(c);
                  mu[k,j]=gauss(x[i,j],c[k,j],s[k,j]); *Gauss or Triang;
            end;
      end;

            *PRODUCT FOR PREMISE AND IMPLICATION;
            muprod=prod(mu);
            do i=1 to nrow(mumin);
                  if muprod[i]=0 then muprod[i]=0.1;
            end;

            *MINIMUM FOR PREMISE AND IMPLICATION;
            mumin=mu[,><];

            *ESTIMATING G(x) BY F(x);
            f=mumin/mumin[+];

            f=muprod/muprod[+];

            f_est=b`*f;


            if abs(f est-y[i,])>ef then do;
                  c1=j(1,ncol(c),1);
                  c1=repeat(x[i,],nrow(c));
                  dif=abs(c-c1);
                  min=dif[><,];
                  min=min/w;
                  s=s//min;
                  b=b//y[i,];
                  c=c//x[i,];
                  end;
                  mu=j(nrow(c),ncol(c),1);
                  do n=1 to nrow(s);
                        do m=1 to ncol(s);
                              if s[n,m]=0 then s[n,m]=0.6;
                        end;
                  end;
end;
```

```
*CREATING DATA SETS FOR OUTPUTS;
show datasets;
create bib.rules var{b};
append; close bib.rules;
create bib.centers from c;
append from c; close bib.centers;
create bib.spreads from s;
append from s; close bib.spreads;

/*****************************************************************/
/* MLFE - TESTING FUZZY SYSTEM AND INFERING ABOUT INPUT DATA     */
/*****************************************************************/
proc iml;
use bib.centers5 var{
col1 col2 col3 col4 col5 col6 col7 col8 col9 col10 col11 col12 col13 col14
col15 col16 col17 col18};
      read all into c;

use bib.spreads5 var{
col1 col2 col3 col4 col5 col6 col7 col8 col9 col10 col11 col12 col13 col14
col15 col16 col17 col18};
      read all into s;

use bib.rules5 var{b};
      read all into b;

*Input Test Set ('out of time') MODEL VALIDATION;
use validation var{
var19 var20 var21 var22 var23 var24 var25 var26 var27 var28 var29 var30
var31 var32 var33 var34 var35 var36};
      read all into g;

mu=j(nrow(c),ncol(c),1);
f_est=j(nrow(g),1,1);

do i=1 to nrow(g);
      do j=1 to ncol(g);
            do r=1 to nrow(c);
                  mu[r,j]=gauss(g[i,j],c[r,j],s[r,j]);
            end;
      end;
muprem=mu[,><];
                        **** CHOOSE BETWEEN MINIMUM OR PRODUCT ;

muprem=prod(mu);

f=muprem/muprem[+];
f_est[i]=b`*f;
end;

create estimate from f_est;
append from f_est; close estimate;
quit;
```

APÊNDICE B – Gráficos das medidas de avaliação KS e AUROC

Figura 14: Gráfico KS para o modelo de *Credit Scoring*

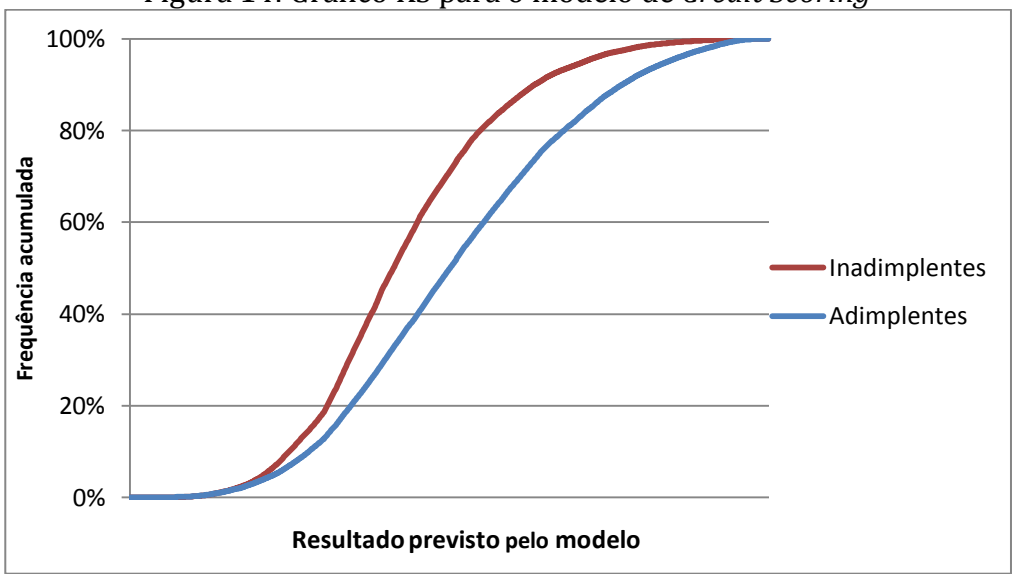


Figura 15: Gráfico KS para o Sistema *Fuzzy*

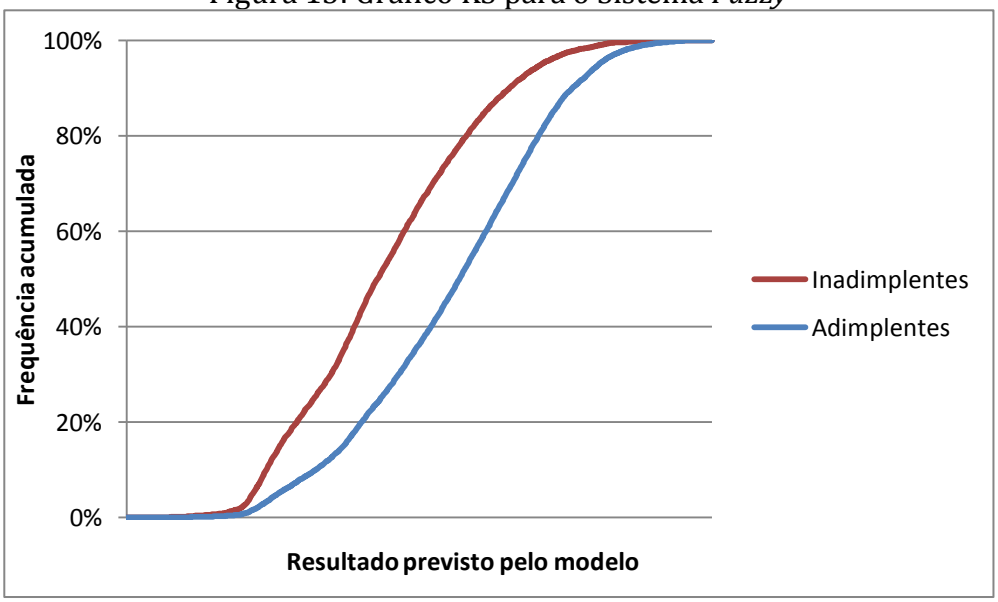


Figura 16: Curva ROC para modelo CS

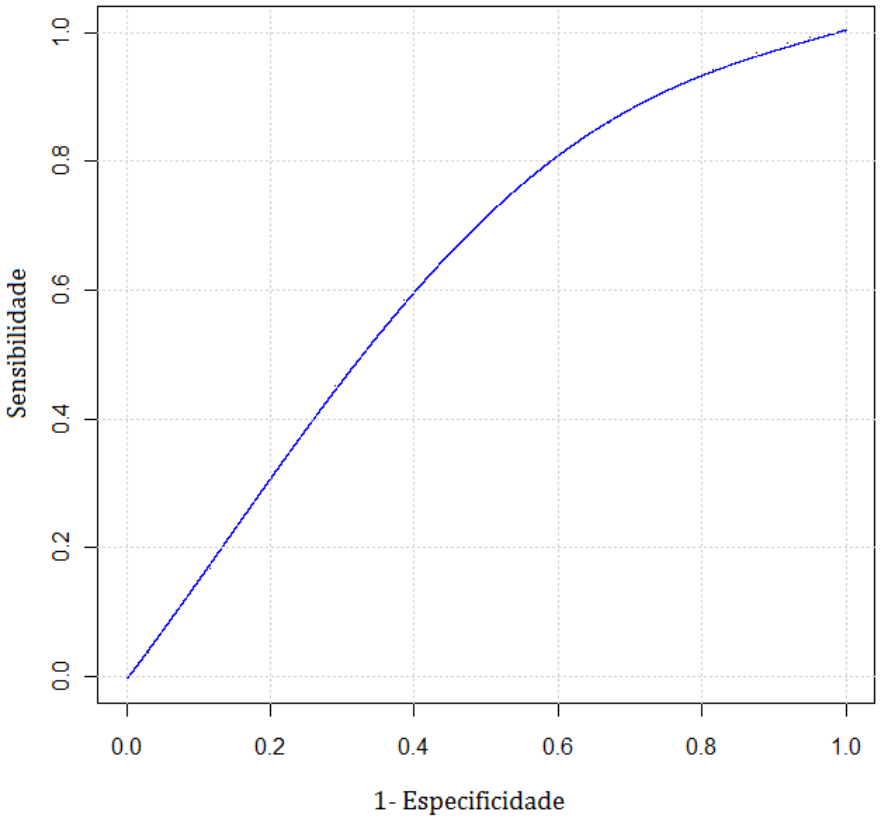
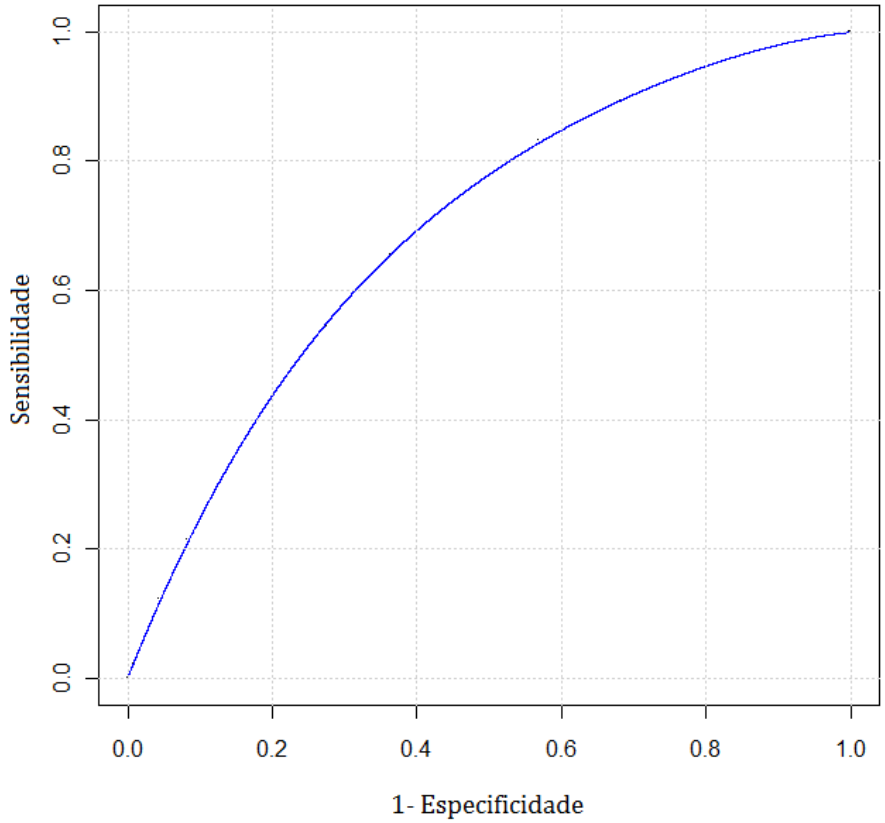


Figura 17: Curva ROC para Sistema *Fuzzy*



ANEXO A – Valores de referência para as medidas de avaliação adotadas

Tabela 9 – Referência para KS

Valor da medida	Discriminação
$KS < 15$	Muito baixa
$15 \leq KS < 25$	Baixa
$25 \leq KS < 35$	Aceitável
$35 \leq KS < 45$	Boa
$KS \geq 45$	Excelente

Fonte – Filho e Sleegers, 2010.

Tabela 10 – Referência para AUROC

Valor da medida	Discriminação
$AUROC < 0{,}60$	Muito baixa
$0{,}60 \leq AUROC < 0{,}68$	Baixa
$0{,}68 \leq AUROC < 0{,}74$	Aceitável
$0{,}74 \leq AUROC < 0{,}80$	Boa
$AUROC \geq 0{,}80$	Excelente

Fonte – Filho e Sleegers, 2010.

Tabela 11 – Referência para AR

Valor da medida	Discriminação
$AR < 0{,}20$	Muito baixa
$0{,}20 \leq AR < 0{,}35$	Baixa
$0{,}35 \leq AR < 0{,}48$	Aceitável
$0{,}48 \leq AR < 0{,}60$	Boa
$AR \geq 0{,}60$	Excelente

Fonte – Filho e Sleegers, 2010.

Tabela 12 – Referência para Distância de Mahalanobis

Valor da medida	Discriminação
$DM < 0{,}35$	Muito baixa
$0{,}35 \leq DM < 0{,}65$	Baixa
$0{,}65 \leq DM < 0{,}90$	Aceitável
$0{,}90 \leq DM < 1{,}21$	Boa
$DM \geq 1{,}21$	Excelente

Fonte – Filho e Sleegers, 2010.

Observação: Não foram encontrados valores de referência para o Escore de Brier.