

# **Modelagem do Risco de Crédito: Um Estudo do Segmento de Pessoas Físicas em um Banco de Varejo**

Autoria: Antônio Alves Amorim Neto, Charles Ulises De Montreuil Carmona

## **Resumo:**

Nas últimas décadas diversos modelos estatísticos de probabilidade foram desenvolvidos pelas instituições financeiras. Estes modelos, no entanto, na maioria das vezes são específicos para análise do crédito de empresas. O segmento de pessoas físicas ao contrário do segmento de pessoas jurídicas é bastante homogêneo sob a ótica financeira, em outras palavras, este segmento possui poucos índices financeiros para serem analisados. As principais diferenças entre os clientes bancários do segmento de pessoas físicas estão relacionadas ao seu comportamento. A funcionalidade dos modelos estatísticos multivariados aplicados ao gerenciamento do crédito para pessoas físicas no Brasil ainda é uma incógnita. O propósito deste artigo, portanto, consiste em preencher uma lacuna existente no meio acadêmico brasileiro no que se refere aos modelos de gerenciamento e concessão do crédito para pessoas físicas. Os resultados encontrados nesta pesquisa trazem indícios que os modelos multivariados podem ser utilizados como ferramentas eficazes no gerenciamento do crédito bancário para o segmento de pessoas físicas no Brasil.

## **1. Introdução**

As mudanças testemunhadas na economia brasileira após a introdução do Plano Real em julho de 1994, alçaram a questão de administração de risco de crédito a uma posição de destaque nas instituições financeiras locais. Antes disso, a alta inflação havia inibido o crescimento do mercado de empréstimos no país. “Ao invés de emprestar a clientes do setor privado (a um risco de crédito maior), os grandes bancos comerciais preferiam financiar a dívida interna do país (a um risco de crédito menor), com maior liquidez, mas ainda assim cobrando taxas de juros substanciais” (PRADO, BASTOS e DUARTE JR., 2000).

A expansão das operações de crédito, no Brasil após o Plano Real foi abrangente em vários segmentos da economia, entretanto, as operações de crédito para pessoas físicas foram as que mais cresceram nos últimos anos. As informações consolidadas divulgadas pelo Banco Central do Brasil para cada modalidade de crédito do sistema financeiro nacional mostram que já no início do ano 2001 o saldo total das operações concedidas a pessoas físicas representava um volume superior a 50% do saldo total das operações concedidas a pessoas jurídicas.

O presente trabalho justifica-se por tratar de um tema de fundamental importância para as empresas que trabalham com crédito para pessoas físicas: A utilização de técnicas estatísticas multivariadas no desenvolvimento de modelos para previsão de inadimplência no segmento de pessoas físicas em um banco comercial brasileiro.

## **2. Referencial teórico**

### **2.1 Sistemas de pontuação de crédito (*credit scoring*)**

*Credit scoring* pode ser definido como o processo de atribuição de pontos às variáveis de decisão mediante técnicas estatísticas. Trata-se de processo que define a probabilidade de que um cliente com certas características, pertença ou não a um grupo possuidor de outras

determinadas características consideradas desejáveis, hipótese em que se aprova um limite de crédito, esta técnica, portanto, estabelece uma regra de discriminação de um determinado cliente solicitante de crédito (VICENTE, 2001, p.49).

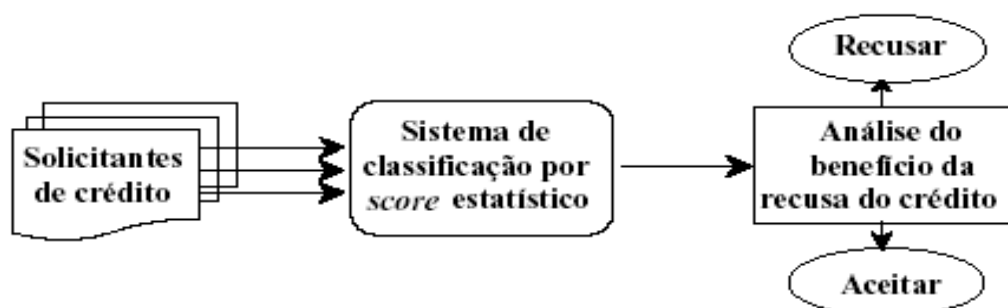
Os sistemas de pontuação de crédito definem a probabilidade de um cliente vir a ser “bom pagador” ou “mau pagador” com base em suas características: “Existem vários fatores que estão associados à possibilidade de inadimplência. Um modelo de escoragem de crédito combina os fatores mais importantes associados à possibilidade de inadimplência, determina o inter-relacionamento entre eles e atribui números para gerar o escore final. A prática tem por objetivo produzir um modelo de escoragem de crédito no qual quanto maior for o escore, menor será o risco de perda com devedores duvidosos” (GHERARDI e GHIEMMETTI, 2000).

Os sistemas de pontuação de crédito são encontrados em praticamente todos os tipos de análises de crédito, desde crédito ao consumidor até empréstimos comerciais. A idéia é essencialmente a mesma: A pré-identificação de certos fatores-chave que determinam a probabilidade de inadimplência, e sua combinação ou ponderação para produzir uma pontuação quantitativa (SAUNDERS, 2000, p. 13).

Os modelos tradicionais de *credit scoring* atribuem pesos estatisticamente predeterminados a alguns dos atributos dos solicitantes, para gerar um escore de crédito. Se esse escore é favorável, quando comparado a um valor de corte, então a solicitação é aprovada (CAOQUETTE, ALTMAN e NARAYANAN, 1998, p.182).

A figura abaixo, apresentada por Sousa e Chaia (2000) ilustra o processo de concessão de crédito através do uso de modelos de *credit scoring*:

**Figura 2.1 – Processo de concessão de crédito através do uso de modelos de *credit scoring***



**Fonte: SOUSA E CHAIA, 2000, p.21.**

O *credit scoring* representa um processo científico, porém, não inibe a possibilidade de se recusar um bom pagador ou se aceitar um “mau” pagador. Isto ocorre porque nenhum sistema de gestão de crédito consegue o total de informações relevantes na classificação do devedor, e, mesmo que conseguisse, o seu custo tornaria a análise economicamente inviável (SOUSA E CHAIA, 2000, p.21).

Os modelos de *credit scoring* se dividem em duas categorias: modelos de aprovação de crédito e modelos de escoragem comportamental, também conhecidos como *behavioural scoring* (CAOQUETTE, ALTMAN e NARAYANAN, 1998, p.180).

O *behavioural scoring* é uma ferramenta para previsão de eventos associados ao risco de crédito, como a inadimplência e pagamentos em dia, entre outras características. Para tanto leva em consideração aspectos comportamentais e de atividades de clientes da instituição (VICENTE, 2001, p.47).

A principal diferença entre as duas categorias é que nos modelos de escoragem comportamental, a instituição financeira analisa o comportamento do cliente em operações anteriores, enquanto que nos modelos de aprovação de crédito a instituição financeira não conhece o cliente: A informação adicional no *behavioural scoring* é o histórico de compras e pagamentos do cliente (THOMAS, 2000).

No *behavioural scoring* são utilizadas variáveis tais como número de atrasos no último ano, volume da transação, utilização média da linha. Desta maneira cada operação recebe uma classificação, que representa seu risco de crédito relacionado ao comportamento do tomador.

Enquanto os modelos de aprovação de crédito preocupam-se apenas com a concessão e o volume de crédito, os modelos de escoragem comportamental podem ser utilizados para gestão de limites de crédito rotativo, autorizações de compra acima do limite em caso de cartões de crédito, ofertas de crédito, cobrança preventiva, entre outras estratégias (VICENTE, 2001, p.48).

## 2.2 Histórico

David Durand em 1941 foi o primeiro a reconhecer que a técnica de análise discriminante, inventada por Fisher em 1936, poderia ser usada para separar bons e maus empréstimos (THOMAS, 2000, p. 6). Em “*Risk Elements in Consumer Installment Financing, 1941 (National Bureau of Economic Research, N.Y.)*”, Durand apresentou um modelo que atribuía pesos para cada uma das variáveis usando análise discriminante (KANG e SHIN, 2000, p.2198).

A popularização dos sistemas de *credit scoring*, no entanto, só ocorreu nos anos 60: “A chegada dos cartões de crédito na década de 60 fez com que os bancos intensificassem o uso de *credit scoring*. O número de pessoas que solicitavam cartão de crédito a cada dia tornou economicamente impossível ter mão-de-obra suficiente para decisões de empréstimos que não fossem automatizadas” (THOMAS, 2000).

No meio acadêmico os estudos começaram na década de 60. O modelo de Altman (1968) é considerado um marco teórico no estudo do risco de crédito:

Os estudos acadêmicos sobre o risco, no entanto, remontam ao final da década de 60, num dos trabalhos pioneiros sobre risco de crédito, Altman (1968) argumentava que o desenvolvimento de um novo modelo preditivo era necessário em função do crescimento das falências e mudanças financeiras nas organizações, agravado pelo drástico aumento do tamanho médio de empresas falidas. Apresentou então, o clássico modelo de análise discriminante de sete variáveis ZETA, continuação dos trabalhos já apresentados em 1968 com cinco variáveis”(BRUNI, FUENTES e FAMÁ, 1997, p.1)

O avanço na computação permitiu o uso de outras técnicas na construção de sistemas de *credit scoring*. Nos anos 80 foram introduzidas as técnicas de regressão logística e regressão linear, as duas principais técnicas utilizadas atualmente para a construção dos modelos. Mais recentemente técnicas de inteligência artificial como sistemas especialistas e redes neurais artificiais foram implantadas com sucesso (THOMAS, 2000, p.7).

### 2.3 Vantagens dos modelos de *credit scoring*

Sobre as vantagens relativas ao uso dos modelos de *credit scoring*, CAOUETTE, ALTMAN e NARAYANAN (1998) dizem:

“Os modelos de *credit scoring* oferecem muitas vantagens. São objetivos e consistentes, que são características desejáveis para qualquer instituição, e especialmente para aquelas que não possuem uma forte cultura de crédito. Se desenvolvidos apropriadamente, eles podem eliminar práticas discriminatórias nos empréstimos. Eles tendem a ser relativamente inexpressivos, bastante simples e de fácil interpretação. A instalação de tais modelos é relativamente fácil. As metodologias usadas para construir esses modelos são comuns e bem entendidas, assim como as abordagens usadas para avalia-los. Os regulamentadores aprovam modelos bem projetados e baseados em estatísticas. Uma instituição é capaz de proporcionar melhor serviço ao consumidor pela sua habilidade de aprovar ou negar um pedido de empréstimo rapidamente. Esse é um fator importante no mundo de mudanças rápidas como o atual.”

Por outro lado Silva (2000) conclui: “A agilidade que o banco ou a empresa que concede crédito ganha é altamente valiosa, pois ao invés do analista ficar examinando e concluindo sobre cada um dos índices, poderão dedicar seus tempos a outros assuntos relevantes e que não possam ser sistematizados”.

Parkinson e Ochs (1998, p.26-27) elaboraram um resumo com as principais vantagens do uso de sistemas de *credit scoring*:

Revisões de crédito consistente: Os dados históricos de outros devedores são um bom indicador de consistência para revisão de crédito.

Informações Organizadas: A sistematização e organização das informações contribuem para a melhoria do processo.

Eficiência no trato de dados fornecidos por terceiros: O processo de *Credit Scoring* torna as informações de banco de dados fornecido por terceiros, anteriormente classificadas como dados acessórios, parte integrante do sistema.

Diminuição da metodologia subjetiva: O uso de um sistema quantitativo parametrizado que minimiza o subjetivismo.

Compreensão do processo: O modelo construído sintetiza o processo de concessão de crédito da empresa, fornecendo maiores subsídios para entendê-lo.

Maior eficiência do processo: A análise de crédito é centrada em um número menor de fatores, reduzindo o tempo do processo e melhorando a eficiência.

### 2.4 Limitações e desvantagens dos modelos de *credit scoring*

Apesar das vantagens em relação a outros modelos citadas acima, os modelos de pontuação de crédito possuem também diversas desvantagens.

Entre as principais limitações dos modelos de *credit scoring*, Silva (2000) destaca o aspecto temporal da amostra:

“O tempo (a época) é uma das principais limitações apresentadas pelos modelos desenvolvidos a partir do uso de análise discriminante. Com o decorrer do tempo, tanto as variáveis quanto seus pesos relativos sofrem alterações. As variáveis que, segundo a análise discriminante, são as que melhor classificam sob determinada conjuntura econômica, podem não ser em outra situação.”

Caouette, Altman e Narayanan (1998) também destacam o aspecto temporal: “Um modelo de *credit scoring* pode degradar-se pelo tempo se a população que ele é aplicado diverge da população original que foi usada para construir o modelo”.

O aspecto geográfico é outra grande limitação para os sistemas de *credit scoring*: “Os aspectos da região geográfica limitam o uso de um modelo único, sendo que o desenvolvimento de diversos modelos poderá exigir que disponha de amostras muito grandes” (SILVA, 2000, p.308).

Parkinson e Ochs (1998, p.26-27) destacaram as seguintes desvantagens do uso de sistemas de *credit scoring*:

Custo de desenvolvimento: o desenvolvimento de um sistema de *credit scoring* traz não apenas os custos do sistema, mas altos custos de manutenção com suporte técnico e recursos humanos.

Modelos com “excesso de confiança”: Os modelos de *Credit Scoring* podem adquirir Status de “Perfeitos”, não permitindo crítica de seus resultados.

Problemas de “Valores não preenchidos”: A falta de dados não característicos de determinados clientes, ou mesmo a falta de informação pode causar problemas na sua utilização, assim como resultados não esperados.

Interpretação equivocada dos escores: O uso inadequado do sistema devido à falta de conhecimento pode ocasionar problemas para a instituição.

### **3. Métodos estatísticos aplicados em modelos de *credit scoring***

As três principais técnicas para elaboração de sistemas de *credit scoring* são a análise discriminante, a regressão logística e, mais recentemente modelos baseados em redes neurais.

A técnica de análise discriminante foi a primeira técnica estatística utilizada na criação de modelos de escoragem de crédito: “A aplicação de modelos matemáticos e estatísticos para a decisão de concessão de crédito ganhou reconhecimento quando David Duran, da *National Bureau of Economic Research* (Agência Nacional de Pesquisa Econômica), em Nova Iorque, publicou o seu estudo em 1941 intitulado “Risk Elements in Consumer Installment Financing” (Elementos de Risco no Financiamento a Prazo do Consumidor). Ele foi o primeiro a usar a análise discriminante para medir o risco de crédito. Subseqüentemente, foram feitas várias tentativas para investigar os sistemas de crédito estatísticos, mas a indústria de crédito não considerou seriamente os modelos de pontuação de crédito até meados de pontuação de crédito até meados da década de sessenta” (COLE e MISHLER, 1998).

A partir da década de 80 foi introduzida a técnica de regressão logística, e mais recentemente foram implantados métodos de análise de crédito baseados em redes neurais (THOMAS, 2000).

Os modelos mais utilizados atualmente continuam sendo os modelos de análise discriminante e os modelos de regressão logística: “A evolução das técnicas preditivas de insolvência possibilitou a utilização de modelos estatísticos mais sofisticados como a Análise Discriminante Múltipla (ADM) ou a Análise de Regressão Logística (Logit), presentes em boa parte dos modelos acadêmicos elaborados” (BRUNI, MURRAY e FAMÁ, 1998).

CAOUEITE, ALTMAN e NARAYANAN (1998, p.182) diz que em geral utiliza-se técnicas estatísticas multivariadas tais como análise discriminante e análise de regressão.

### 3.1 Análise discriminante

Os primeiros estudos acadêmicos sobre riscos de insolvência se baseavam em análises de quocientes, índices contábeis, e deduziam que empresas em processos de insolvência apresentavam quocientes bastante diferenciados das solventes. Nestes estudos, uma das principais questões levantadas consistia em se determinar quais índices seriam os mais importantes na predição e qual a ponderação ideal desses índices. Surgiu, então, a análise discriminante multivariada como capaz de solucionar este questionamento (BRUNI, MURRAY e FAMÁ, 1998).

KASSAI e KASSAI (1998, p.6) dizem: “A análise discriminante, também chamada de análise do fator discriminante ou análise discriminante canônica, é uma técnica estatística desenvolvida a partir dos cálculos de regressão linear e, ao contrário desta, permite resolver problemas que contenham não apenas variáveis numéricas, mas também variáveis de natureza “qualitativa”, como é o exemplo de empresas “solventes” e “insolventes”.

A análise discriminante permite descobrir as ligações que existem entre um caráter qualitativo a ser explicado e um conjunto de caracteres quantitativos explicativos. Também permite prever, por meio de um modelo, os valores da variável que derivam dos valores tomados pelas variáveis explicativas (GIMENES e URIBE-OPAZO, 2001, p.67).

O objetivo fundamental da análise discriminante é a alocação de todos os elementos da amostra em grupos bem definidos, evitando todas as formas de superposição. Os dados de cada elemento, de cada um dos grupos, são coletados e, em seguida procura-se derivar uma função, que nada mais é que uma combinação linear, para melhor discriminar os grupos entre si. O resultado almejado é a obtenção de um conjunto único de coeficientes para cada uma das variáveis independentes e que classifique, com a máxima precisão, cada elemento observado nos grupos previamente definidos (GUIMARÃES 2001, p.33).

A função discriminante linear tem a forma:

$$Y_x = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n$$

Onde:

$Y_x$  = Variável Dependente, reflete o valor do escore;

$\beta_0$  = Intercepto;

$\beta_1, \beta_2, \beta_3, \dots, \beta_n$  = Pesos atribuídos a cada variável;

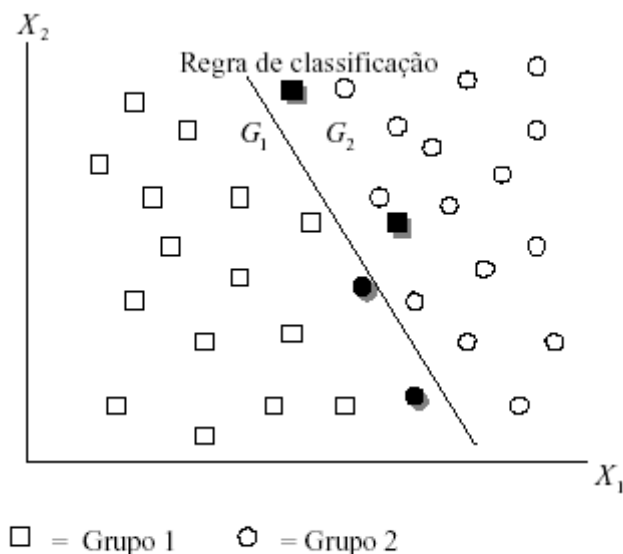
$X_1, X_2, X_3, \dots, X_n$  = Variáveis discriminantes;

A classificação de cada cliente é efetuada a partir da função discriminante estimada, e através do cálculo do valor dessa função para cada cliente (escore). De acordo com o critério

de classificação de Fisher para dois grupos, o cliente deverá ser classificado como bom pagador se estiver mais próximo do centróide deste grupo do que do centróide do grupo dos maus pagadores, isto é, se a distância entre o seu escore discriminante e o centróide do grupo 1 for menor que a distância entre o seu escore e o centróide do grupo 2, e no grupo dos maus pagadores no caso contrário (COSTA, 1992, p.62).

Sejam  $X_1$  e  $X_2$  as variáveis observadas para cada indivíduo de cada grupo. No gráfico abaixo estão representadas as observações efetuadas:

**Gráfico 3.1 – Espaço discriminante**



Fonte: GUIMARÃES, 2000, p.33.

### 3.2 Regressão logística

Os modelos de risco de crédito que utilizam a regressão logística começaram a ser usados na década de 80. “Desde OHLSON (1980) a análise LOGIT é usada freqüentemente para a avaliação de riscos de inadimplência, baseada em características financeiras (por exemplo, índices das empresas)”.

A análise de regressão logística é uma técnica de probabilidade condicional que é usada para estudar a relação entre uma série de características de um indivíduo e a probabilidade de que o indivíduo pertença a um entre dois grupos estabelecidos anteriormente (GIMENES e URIBE-OPAZO, 2001).

A técnica de regressão logística (LOGIT) é aplicada em situações nas quais se deseja prever a presença ou ausência de uma característica, ou resultado, baseado em valores de um conjunto de variáveis independentes. No caso da mensuração do risco de crédito, a regressão logística é utilizada para a avaliação da probabilidade de insolvência (ou inadimplência) de determinado grupo de clientes, relativos a situações de empréstimo ou financiamento (VICENTE, 2001, p.82).

Os modelos elaborados com base na análise LOGIT usam um conjunto de variáveis contáveis para prever a probabilidade de inadimplência do tomador de empréstimo, assumindo que a probabilidade de inadimplência é distribuída logisticamente, ou seja, estatisticamente assume uma forma funcional logística, e é, por definição, forçada a cair entre 0 e 1 (BRUNI, MURRAY e FAMA, 1998).

O modelo LOGIT cria para cada empresa um escore  $Z$ :

$$Z = \alpha + \beta X_i$$

Onde  $X_i$  é o valor da  $i$ -ésima variável.

Estudos comparativos entre modelos que utilizam técnicas de regressão logística e os modelos de análise discriminante mostrou que a capacidade de previsão desses modelos é similar, portanto, não existe uma técnica estatística predominante.

GIMENES e URIBE-OPAZO (2001, p.74) concluem a comparação entre as técnicas de análise discriminante e LOGIT com o seguinte comentário: “Na realidade, não há uma metodologia única para a construção de modelos de previsão de insolvência: muito menos existe um consenso teórico sobre qual das metodologias é a melhor”.

É importante destacar que na análise LOGIT não existem restrições à normalidade multivariada na distribuição das variáveis independentes, nem a igualdade de matrizes de covariância dos dois grupos, como ocorre na análise discriminante, onde se pressupõe a existência dessas condições (GIMENES e URIBE-OPAZO, 2001, p.70).

## **4. Metodologia**

### **4.1 Composição da amostra**

A amostra desta pesquisa está formada por dois grupos de indivíduos selecionados aleatoriamente entre os tomadores de empréstimo pessoa física de um banco comercial:

1. Grupo “A” = 172 clientes sem atraso no pagamento de suas operações, classificados como risco nível “A”;
2. Grupo “H” = 172 clientes inadimplentes.

Nessa pesquisa os clientes inadimplentes são caracterizados como indivíduos com atraso no pagamento de suas operações superior a 180 dias e inferior a 360 dias, classificados de acordo com a resolução 2.682/99 do Banco Central do Brasil como risco nível “H”.

### **4.2 Método de seleção**

Para selecionar os indivíduos da amostra utilizamos métodos de aleatoriedade baseados em números gerados por computador (aplicativo Microsoft Excel). Os dados utilizados referem-se a operações de crédito concedidas a pessoas físicas pelas agências sediadas na cidade de Recife (PE) de um banco comercial brasileiro. Estas informações estão disponíveis em seu sistema de informações interno através do aplicativo específico para esse fim. Os dados coletados foram impressos e arquivados para futuras eventuais consultas.

### **4.3 Parâmetros da amostra**

A classificação consiste em dados referentes à situação/classificação das operações no mês de julho do ano de 2001, e obedece aos seguintes parâmetros:

*A. Indivíduos com responsabilidade cujo valor seja superior a R\$ 200,00* – Foram relacionados apenas indivíduos cujo valor mínimo do somatório de suas operações seja de duzentos reais, pois operações cujo limite seja inferior a 200 reais muitas vezes são deferidas sem a necessidade da atualização de dados cadastrais, o que impossibilita a coleta de dados.



Será assumido o pressuposto de que as operações com valores inferiores a duzentos reais representam um percentual insignificante no montante da carteira de crédito;

*B. Clientes que possuam operações a pelo menos 12 meses* – Será assumido o pressuposto que clientes com operações com prazo menor que 12 meses em julho/2001 podem estar classificados indevidamente no grupo de “bons pagadores” devido ao curto período de tempo analisado;

*C. Clientes que permaneceram 12 meses no mesmo “status”* – Em julho/2002 foram eliminados da amostra clientes inadimplentes que pagaram suas dívidas, assim como, clientes adimplentes que se tornaram inadimplentes.

Para desenvolver os modelos de escoragem do crédito utilizamos duas técnicas estatísticas distintas, Análise Discriminante e Regressão Logística, sendo que para cada técnica foram desenvolvidos dois modelos: Um modelo classificatório para a concessão do crédito, e um modelo classificatório para o gerenciamento do crédito. Estes modelos foram comparados e avaliados.

#### 4.4 Avaliação dos modelos

“Se o modelo de previsão for exato, a exatidão da previsão será melhor que 80 ou 90% sobre o horizonte de previsão” (CAOQUETTE, ALTMAN e NARAYANAN, 1998, p.214). Assumindo que consideramos um modelo funcional quando este apresenta uma taxa de sucesso (número de acertos por casos totais) superior a 80%, podemos expressar a seguinte hipótese:

$$H_0: \mu_1 \geq 80\%$$

$$H_1: \mu_1 < 80\%$$

Onde:  $\mu_1$  = Média de acertos do modelo.

Logo, se  $H_0$  não for rejeitada, pode-se considerar o modelo de risco de crédito desenvolvido válido para o objetivo proposto.

#### 4.5 Principais limitações da metodologia

Abaixo enumeramos as principais limitações metodológicas desta pesquisa:

*Tamanho da amostra.* De uma forma geral quanto maior a amostra, maior será a credibilidade do modelo. No entanto, devido a limitações técnicas e temporais, a amostra limitou-se a 344 indivíduos;

*Ausência da amostra de validação.* Antes de desenvolver a fórmula de escoragem de crédito, o ideal é dividir a amostra total em: *amostra de desenvolvimento* e *amostra de validação*. A fórmula de escoragem de crédito é desenvolvida através da *amostra de desenvolvimento* e testada através da *amostra de validação*, ou seja, a comparação do comportamento previsto com o comportamento real, usando as amostras de desenvolvimento e validação. O teste de validação mostra se o sistema de escoragem de crédito está funcionando, independentemente da amostra na qual foi desenvolvido. No caso dessa pesquisa, e, em muitos outros trabalhos a limitação quantitativa da amostra total impede a divisão em dois grupos;

*Ausência de indivíduos com classificações de crédito intermediárias.* Neste trabalho não estamos analisando a deterioração do crédito, ou seja, incorporando as classes intermediárias entre a classe de risco “A” e “H”. A amostra presente nessa pesquisa limita-se a utilizar clientes sem atrasos, e, clientes com atraso superior a 180 dias;

*Período de análise.* Em geral, análise de período temporal maior traz uma maior compreensão dos modelos. A partir dessas informações seria possível arbitrar o período de análise ideal. A limitação técnica, no entanto, fez com que fosse utilizado o período referente à 1 ano;

*Ausência da taxa de recuperação de crédito.* Um percentual reduzido dos créditos inadimplentes é renegociado. Especialistas calculam que entre 1,5% e 3% das operações em prejuízo (atraso superior a 1 ano) são recuperadas. Nos modelos presentes nessa pesquisa a taxa de recuperação de crédito não é utilizada, pois, faltavam dados abertos sobre a recuperação de crédito, assim como o perfil dos clientes destas operações;

*Limitação institucional e geográfica.* A amostra está limitada a um único banco e uma única cidade. A segmentação do mercado bancário faz com que os bancos atuem em segmentos diferentes, o que pode ocasionar um viés da amostra. A cidade de escolhida, assim, como outras capitais, possui características econômicas peculiares que também pode trazer vieses a amostra;

*Número limitado de informações.* Um cadastro com maior número de informações de variáveis poderia aumentar a capacidade preditiva dos modelos. Informações tais como número de filhos, hábitos de consumo, propriedade de bens, endividamento com outras instituições, podem aumentar sensivelmente a qualidade dos modelos;

*Abstração de princípios éticos e legais.* Não foram analisados princípios éticos e legais nessa pesquisa. Em todo o mundo existem legislações que protegem o consumidor. Nos Estados Unidos a Lei *Equal Credit Opportunity* de 1975 é específica em relação à discriminação do consumidor. O estado civil e o sexo do consumidor, por exemplo, não podem ser motivo de recusa de crédito (MISHLER e COLE, 1998, p.191). No Brasil, o código do consumidor rege essas relações.

## 5. Resultados

### 5.1 Modelos de concessão de crédito

#### 5.1.1 Análise discriminante

A opção “STEPWISE” para análise discriminante do software SPSS 10.0 foi utilizada para gerar o seguinte modelo:

Tabela 5.1.1.a Modelo 1: Análise discriminante

VARIÁVEIS		Coefficientes
ESCOLAR – Grau de escolaridade	1= Nível Fundamental 2= Nível Médio 3 = Nível Superior	0,942
LRESID – Local onde reside	1 = Recife (capital) 2 = Interior de Pernambuco 3 = Fora do Estado	-0,532
NATOCUP – Natureza da ocupação principal	1 = Empresa Privada 2 = Economia mista 3 = Funcionário Público	0,282
IDADE – Idade (Constante)	Valor Absoluto em Anos	0,039 -3,332

O quadro 5.1.1.b mostra as médias dos escores e os parâmetros de classificação (Ponto de corte dos escores) dos dois grupos:

Quadro 5.1.1.b - Análise discriminante: médias e parâmetros de classificação do modelo 1

<b>Escore</b>	<b>Grupo “A” (Adimplentes)</b>	<b>Grupo “H” (Inadimplentes)</b>
Média	0,572.	-0,572.
Parâmetro de classificação	Escore > 0	Escore < 0

O modelo de concessão de crédito desenvolvido com a técnica de análise discriminante conseguiu alocar corretamente 73,3% dos indivíduos da amostra:

Quadro 5.1.1.c – Análise discriminante (modelo 1): percentual de acertos

<b>Classificação Original</b>	<b>Classificação do Modelo</b>	
	<b>Grupo “A” (Adimplentes)</b>	<b>Grupo “H” (Inadimplentes)</b>
<b>Grupo “A”</b> (Adimplentes)	71,5%	28,5%
<b>Grupo “H”</b> (Inadimplentes)	25%	75%

### 5.1.2 Regressão logística

O modelo obtido utilizando regressão logística foi bastante similar ao obtido com a utilização da análise discriminante (quadro 5.1.2.a):

Tabela 5.1.2.a Modelo 1: Regressão logística

<b>VARIÁVEIS</b>		<b>Coefficientes</b>
ESCOLAR – Grau de escolaridade	1= Nível Fundamental 2= Nível Médio 3 = Nível Superior	-1,014
LRESID – Local onde reside	1 = Recife (capital) 2 = Interior de Pernambuco 3 = Fora do Estado	0,623
NATOCUP – Natureza da ocupação principal	1 = Empresa Privada 2 = Economia mista 3 = Funcionário Público	-0,314
IDADE – Idade (Constante)	Valor Absoluto em Anos	-0,042 3,526

As regras de decisão para classificação diferem em relação à técnica estatística utilizada para gerar o modelo de *credit scoring*. O quadro 5.1.2.b mostra as médias dos escores e os parâmetros de classificação (ponto de corte) dos dois grupos:

Quadro 5.1.2b – Regressão logística: médias e parâmetros de classificação do modelo 1

<b>Escore</b>	<b>Grupo “A” (Adimplentes)</b>	<b>Grupo “H” (Inadimplentes)</b>
Média	0,375.	0,625.
Parâmetro de classificação	Escore < 0,50	Escore > 0,50

O percentual de acertos do modelo baseado em regressão logística, assim como aconteceu no modelo gerado através da aplicação de análise discriminante, mostra que o modelo é mais eficiente na classificação de clientes inadimplentes. O quadro 5.1.2.c apresenta o percentual de acerto do grupo “A” (clientes adimplentes) e grupo “H” (clientes inadimplentes):

Quadro 5.1.2.c – Regressão logística (modelo 1): percentual de acertos

Classificação Original	Classificação do Modelo	
	Grupo “A” (Adimplentes)	Grupo “H” (Inadimplentes)
Grupo “A” (Adimplentes)	70,9%	29,1%
Grupo “H” (Inadimplentes)	26,2%	73,8%

A taxa geral de acertos do modelo de concessão de crédito desenvolvido com a técnica de regressão logística (72,4%) foi bem próxima a taxa encontrada no modelo desenvolvido com a aplicação de análise discriminante (73,3%).

## 5.2 Modelos de escoragem comportamental

### 5.2.1 Análise discriminante

O software e os parâmetros utilizados para gerar o modelo de escoragem comportamental foram os mesmos utilizados para gerar o modelo de concessão de crédito. O modelo de escoragem comportamental, no entanto, agrega um número maior de variáveis. A tabela 5.2.1.a apresenta o modelo final de escoragem comportamental (*behavioural scoring*) gerado através da aplicação de análise discriminante:

Tabela 5.2.1.a Modelo 2: Análise discriminante

VARIÁVEIS	Formato/Valores	Coefficientes
ESCOLAR – Grau de escolaridade	1= Nível Fundamental 2= Nível Médio 3 = Nível Superior	0,665
LRESID – Local onde reside	1 = Recife (capital) 2 = Interior de Pernambuco 3 = Fora do Estado	-0,425
NATOCUP – Natureza da ocupação principal	1 = Empresa Privada 2 = Economia mista 3 = Funcionário Público	0,226
IDADE – Idade	Valor Absoluto em Anos	0,030
RETBX – Já possuiu restrição de crédito	0 = Não 1 = Sim	-0,816
PVEICULO – Possui financiamento de veículos	0 = Não 1 = Sim	2,008
PCARTAO – Possui cartão de crédito do banco.	0 = Não 1 = Sim	1,193
(Constante)		-2,213

Os parâmetros de classificação da análise discriminante aplicada a escoragem comportamental são os mesmos aplicados ao modelo de concessão de crédito. As médias dos escores dos dois grupos, no entanto, tornaram-se mais distantes do ponto de corte, o que indica uma melhoria de eficiência em relação ao modelo de concessão de crédito:

Quadro 5.2.1.b - Análise discriminante: médias e parâmetros de classificação do modelo 2.

Escore	Grupo “A” (Adimplentes)	Grupo “H” (Inadimplentes)
Média	0,771.	-0,771.
Parâmetro de classificação	Escore > 0	Escore < 0

O quadro 5.2.1.c apresenta o percentual de classificações corretas e incorretas efetuadas pelo modelo para cada um dos grupos (grupo “A” e grupo “H”):

Quadro 5.2.1.c – Análise discriminante (modelo 2): percentual de acertos

Classificação Original	Classificação do Modelo	
	Grupo “A” (Adimplentes)	Grupo “H” (Inadimplentes)
Grupo “A” (Adimplentes)	76,7%	23,3%
Grupo “H” (Inadimplentes)	14%	86%

Assim como no modelo 1 (concessão de crédito) verificou-se que o modelo 2 (escoragem comportamental) desenvolvido com a técnica de análise discriminante possui maior capacidade de predição para o grupo dos clientes inadimplentes (86%) do que no grupo dos adimplentes (76,7%).

O modelo 2 desenvolvido com a técnica de análise discriminante conseguiu alocar corretamente 81,4% dos indivíduos da amostra, o que demonstra um ganho considerável de eficiência em relação ao modelo 1 (73,3%).

## 5.2.2 Regressão Logística

O mesmo software (SPSS 10.0) e parâmetros foram usados para gerar o modelo 2 de regressão logística. O modelo de escoragem comportamental obtido com a utilização da regressão logística está ilustrado na tabela 5.2.2.a:

Tabela 5.2.2.a Modelo 2: Regressão logística

VARIÁVEIS	Formato/Valores	Coefficientes
<i>ESCOLAR</i> – Grau de escolaridade	1= Nível Fundamental 2= Nível Médio 3 = Nível Superior	-0,991
<i>LRESID</i> – Local onde reside	1 = Recife (capital) 2 = Interior de Pernambuco 3 = Fora do Estado	0,686
<i>NATOCUP</i> – Natureza da ocupação principal	1 = Empresa Privada 2 = Economia mista 3 = Funcionário Público	-0,332
<i>IDADE</i> – Idade	Valor Absoluto em Anos	-0,039
<i>RESTBX</i> – Já possuiu restrição de crédito	0 = Não 1 = Sim	1,405
<i>PVEICULO</i> – Possui financiamento de veículos pelo banco	0 = Não 1 = Sim	-8,049
<i>PCARTAO</i> – Possui cartão de crédito do banco.	0 = Não 1 = Sim	-1,667
(Constante)		3,213

O ponto de corte da regressão logística para o modelo 2 é o mesmo do modelo 1. Ainda em relação ao modelo 1, os escores se tornaram mais distantes do ponto de corte (0,50) indicando maior precisão do modelo 2:

Quadro 5.2.2.a – Regressão logística: médias e parâmetros de classificação do modelo 2

<b>Escore</b>	<b>Grupo “A” (Adimplentes)</b>	<b>Grupo “H” (Inadimplentes)</b>
Média	0,297.	0,703.
Parâmetro de classificação	Escore < 0,50	Escore > 0,50

O quadro 5.2.2.b apresenta o percentual de classificações corretas e incorretas para os dois grupos:

Quadro 5.2.2.b – Regressão logística (modelo 2): percentual de acertos

<b>Classificação Original</b>	<b>Classificação do Modelo</b>	
	<b>Grupo “A” (Adimplentes)</b>	<b>Grupo “H” (Inadimplentes)</b>
<b>Grupo “A” (Adimplentes)</b>	78,5%	21,5%
<b>Grupo “H” (Inadimplentes)</b>	15,1%	84,9%

O modelo 2 desenvolvido com a técnica de regressão logística, assim como os outros modelos desenvolvidos, também mostrou classificar melhor os clientes inadimplentes (84,9%) do que os clientes adimplentes (78,5%). A capacidade de predição desse modelo (81,7%) é ligeiramente superior ao modelo desenvolvido com a técnica de análise discriminante (81,4%).

## 6. Análises e conclusões

A análise dos modelos confirma a similaridade da eficiência das técnicas de análise discriminante e regressão logística. A diferença da taxa de sucesso nos dois modelos foi mínima: No modelo 1 a técnica de análise discriminante conseguiu uma taxa de sucesso superior (73,3%) a taxa obtida através da aplicação da regressão logística (72,4%), enquanto que no modelo 2, a técnica de regressão logística obteve taxa de sucesso (81,7%) superior ao da aplicação da técnica de análise discriminante.

Outro ponto que diz respeito à similaridade dos modelos refere-se as variáveis obtidas. Os dois modelos possuem as mesmas variáveis, e, respeitando-se as regras de decisão de cada modelo, estas variáveis possuem pesos (coeficientes) bem parecidos. O resultado é que o percentual de indivíduos que recebeu a mesma classificação dos modelos desenvolvidos com a aplicação da análise discriminante e regressão logística é de 98,5% no modelo 1, e 98% no modelo 2, ou seja, apenas 5 indivíduos receberam classificações diferentes no modelo de concessão de crédito, enquanto que no modelo de escoragem comportamental o quantitativo foi de 7 indivíduos.

As semelhanças acima citadas conjuntamente com a ausência de uma técnica estatística com taxa de sucesso predominante nos dois modelos corroboram com a idéia, presente em outros estudos comparativos, de que não existe uma técnica estatística predominante no desenvolvimento de modelos de *credit scoring*.

De acordo com a metodologia dessa pesquisa um modelo é considerado funcional quando apresenta uma taxa de sucesso superior a 80%, além disso, para ser aplicável na realidade um modelo tem que estar de acordo com a legislação sobre provisão de crédito.

Os estudos sobre finanças corporativas incorporam uma multiplicidade de índices, estratégias e variáveis. Os estudos sobre finanças pessoais, no entanto, limitam-se a poucos índices financeiros, assim, o comportamento humano é quem determina a heterogeneidade desse segmento. Mensurar o comportamento humano e suas variáveis, porém, é uma tarefa

árdua. É comum ver indivíduos de origem, formação e renda semelhante seguirem caminhos divergentes em suas finanças pessoais.

Os modelos desenvolvidos para fins de concessão de crédito desenvolvidos com as técnicas de análise discriminante e regressão logística tiveram taxas de sucesso de 73,3% e 72,4% respectivamente. Os modelos de concessão de crédito apresentam geralmente baixas de sucesso devido ao número limitado de variáveis. Nos modelos desenvolvidos nessa pesquisa, variáveis importantes como renda familiar, restrições do cônjuge, e quantidade de instituições que o cliente opera ficaram ausentes, apesar disso as taxas de sucesso média dos modelos de concessão de crédito ficaram próximas de 73% de acerto. A hipótese de funcionalidade dos modelos de concessão de crédito foi rejeitada, no entanto, esse modelo gerou importantes informações que podem melhor subsidiar as decisões de um analista de crédito.

Os modelos de escoragem comportamental envolvem um maior número de variáveis, e, conseqüentemente apresentam maiores taxas de sucesso. Os modelos de *behavioural scoring* desenvolvidos nessa pesquisa tiveram, além da limitação das variáveis, a limitação temporal, já que normalmente o período analisado é duas vezes maior que o aqui utilizado. Apesar disso o modelo gerado através da análise discriminante obteve uma taxa de sucesso de 81,4%, enquanto que o modelo gerado através de regressão logística obteve uma taxa de sucesso de 81,7%.

De acordo com esses valores a hipótese de funcionalidade dos modelos não foi rejeitada, ou seja, os modelos foram considerados funcionais para o objetivo proposto. Esses valores indicam, ainda, que com um cadastro mais completo, informações mais coerentes, e um período de análise maior, provavelmente seria possível obter uma taxa de sucesso maior.

A natureza “*ad hoc*” impede, no entanto, a generalização e inferência dos resultados obtidos neste trabalho para outras realidades. Apesar de ser improvável, é possível que em uma outra região geográfica, por exemplo, as amostras dos dois grupos (adimplentes e inadimplentes) sejam tão homogêneas que torne impossível desenvolver modelos de *credit scoring* funcionais.

## 7. Referências Bibliográficas

- ALTMAN, E. I., **Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy**. Journal of Finance, Sep 1968, p. 589-609.
- AMORIM NETO, Antônio A. **Modelagem do risco de crédito: um estudo do segmento de pessoas físicas em um banco de varejo**. Dissertação de Mestrado. Universidade Federal de Pernambuco. Recife: 2002.
- BRUNI, A. L., FUENTES J., FAMÁ, R. **Risco de crédito: Evolução teórica e mecanismo de proteção desenvolvido nos últimos 20 anos**. Anais do II SEMEAD da FEA/USP. São Paulo: 1997.
- BRUNI, A. L., MURRAY, A. D. , FAMÁ, R. **Modelos Brasileiros Preditivos de Risco de Crédito: Um Estudo Exploratório Atual sobre as suas Eficácias**. Periódico Tema, número 32, p. 148-167, janeiro/junho 1998.
- CAOQUETTE, J.B. E.J. ALTMAN, e P. NARAYANAN, **Gestão do Risco de Crédito: O próximo grande desafio financeiro**. – Rio de Janeiro: Qualitymark Ed. 1998.
- COLE, R., MISHLER L. **Gestão de Crédito**. 11ª Edição. São Paulo: Ed. McGraw-Hill, 1998.
- COSTA, Francisco C. **Avaliação de crédito pessoal**. Dissertação de Mestrado. Instituto Superior de Ciências do Trabalho e da Empresa – Portugal. Lisboa: Julho/1992.
- DUARTE JR., Antonio M. **Riscos: Definições, Tipos, Medição e Recomendações para o seu gerenciamento**. Working Paper, IBMEC. São Paulo: 2000.

GHERARDI, Carlo e GHIEMMETTI, Sílvia – Revista “Tecnologia do Crédito”, ANO 01, Nº 02, Artigo **Escoragem de Crédito: Metodologia que identifica Estatisticamente o Risco de Crédito** – São Paulo: Serasa, setembro/97.

GIMENES, Régio M. T., URIBE-OPAZO, Miguel A. **Modelos Multivariantes para a previsão de insolvência em cooperativas agropecuárias: Uma comparação entre a Análise Discriminante e a Análise de Probabilidade Condicional - Logit**. Caderno de Pesquisas em Administração, v.08, nº 03, pp 65-76. São Paulo: Julho/2001.

GUIMARÃES, Inácio A. **Construção e avaliação de uma regra de conhecimento e classificação de clientes de uma instituição financeira com base em análise multivariada**. Dissertação de Mestrado, Universidade Federal do Paraná. Curitiba: 2000.

HAND, David J. **Modelling consumer credit risk**. Journal of Management Mathematics, n 12, p. 139-155. London, U.K: 2001.

KANG, S., SHIN, K. **Customer credit scoring model using analytic hierarchy process**. Inform & Korms, Seoul, p.2197-2204. Korea:2000.

KASSAI, J. R., KASSAI S. **O termômetro de kanitz**. Trabalho apresentado no ENANPAD/98. Foz do Iguaçu/PR. Setembro/98.

OHLSON, J.A. **Financial Ratios and the Probabilistic Prediction of Bankruptcy**. Journal of Accounting Research, Spring, p.109-131, 1980.

PARKINSON, Kenneth L. & OCHS, Joyce R. **Using credit screening to manage credit risk**. Business Credit, p.23-27, mar., 1998.

PRADO, Renata G. A., BASTOS, Norton T., DUARTE JR, Antônio M. **Gerenciamento de Riscos de Crédito em Bancos de Varejo no Brasil**. Global Risk Management – Unibanco S.A. São Paulo, 2000.

SANVICENTE, Antônio Z., MINARDI, Andréa M. A. **Migração de risco de crédito de empresas brasileiras: Uma aplicação de análise de clusters na área de crédito**. Instituto Brasileiro de Mercado de Capitais, São Paulo: 1999.

SAUNDERS, Anthony. **Medindo o Risco de Crédito: Novas abordagens para value at risk e outros paradigmas**. Rio de Janeiro: Qualitymark Ed., 2000.

SECURATO, José R. **Um modelo para determinar o valor presente de uma carteira de crédito e de seus riscos**. Anais do III SEMEAD da FEA/USP. São Paulo:1998.

SICSU, Abraham L. Revista “Tecnologia do Crédito”, ANO 01, Nº 05, Artigo **Desenvolvimento de um Sistema de Credit Scoring**. São Paulo: Serasa, Março/98.

SILVA, José Pereira. **Gestão e análise do risco de crédito**. – 3 ed. São Paulo: Atlas, 2000.

SOUSA, Almir F., CHAIA, Alexandre J. **Política de crédito: uma análise qualitativa dos processos das empresas**. Caderno de Pesquisas em Administração, v.07, nº 03, pp 13-25. São Paulo: Julho/2000.

THOMAS, Lyn C. **A survey of credit and Behavioural Scoring; Forecasting financial risk of lending to consumers**. University of Edinburgh, Edinburgh, U.K, 2000.

VICENTE, Ernesto F. R. **A estimativa do risco na constituição da PDD**. Dissertação de Mestrado, USP. São Paulo: Maio/2001.