

RECONHECIMENTO DE PADRÕES: METODOLOGIAS ESTATÍSTICAS EM CRÉDITO AO CONSUMIDOR

Por:

Inácio Andruski Guimarães
Anselmo Chaves Neto

RAE-eletrônica, Volume 1, Número 2, jul-dez/2002.

<http://www.rae.com.br/eletronica/index.cfm?FuseAction=Artigo&ID=1215&Secao=FINANÇAS2&Volume=1&Numero=2&Ano=2002>

©Copyright, 2002, RAE-eletrônica. Todos os direitos, inclusive de tradução, são reservados. É permitido citar parte de artigos sem autorização prévia desde que seja identificada a fonte. A reprodução total de artigos é proibida. Os artigos só devem ser usados para uso pessoal e não-comercial. Em caso de dúvidas, consulte a redação: redacao@rae.com.br.

A RAE-eletrônica é a revista on-line da FGV-EAESP, totalmente aberta e criada com o objetivo de agilizar a veiculação de trabalhos inéditos. Lançada em janeiro de 2002, com perfil acadêmico, é dedicada a professores, pesquisadores e estudantes. Para mais informações consulte o site www.rae.com.br/eletronica.

RAE-eletrônica
ISSN 1676-5648

©2002 Editora: Fundação Getulio Vargas – Escola de Administração de Empresas de São Paulo.



FUNDAÇÃO
GETULIO VARGAS



Escola de Administração
de Empresas de São Paulo

RECONHECIMENTO DE PADRÕES: METODOLOGIAS ESTATÍSTICAS EM CRÉDITO AO CONSUMIDOR

Inácio Andruski Guimarães

Professor do Departamento de Matemática do CEFET PR - Centro Federal de Educação Tecnológica do Paraná. Professor do Departamento de Administração de Empresas da FESP. Mestre em Métodos Numéricos em Engenharia pela UFPR.

E-mail: inacio@fesppr.br

Endereço: Rua sete de setembro, 3165 - Curitiba - PR, 80230-010

Interesses de pesquisa: Reconhecimento de Padrões, Análise de Crédito, Regressão Logística.

Anselmo Chaves Neto

Mestre em Estatística pela UNICAMP. Doutor em ENGENHARIA pela PUC-RIO.

E-mail: anselmo@est.ufpr.br

Endereço: C.P. 19081 Centro Politécnico, CEP 81531-990

Interesses de pesquisa: Métodos Estatísticos Multivariados, Análise de Séries Temporais, Controle Estatístico de Qualidade.

RESUMO

A inadimplência é um dos maiores problemas, senão o maior, enfrentado pelas administradoras de cartão de crédito. No estudo deste problema foi criado o conceito de risco, que é essencialmente a probabilidade de não recebimento dos créditos por parte das administradoras de cartões. Alguns autores, Caouette *et al.* (2000) e Silva (1988) referem-se às técnicas estatísticas multivariadas como ferramentas poderosas na administração do risco envolvido na concessão de crédito pessoal. Este trabalho apresenta a construção e avaliação de regras de reconhecimento de padrões baseadas em duas técnicas multivariadas: a Função Discriminante Linear de Fisher e a Regressão Logística para classificação de clientes de cartão de crédito em um de dois grupos. A eficiência dos procedimentos é avaliada por meio do Método de Lachenbruch, Lachenbruch (1975).

ABSTRACT

The non-payment (breach of contract) is one of the major, if not the major, problem faced by administrators (companies, agencies) of credit. In studies of such problems it was created the risk concept, that is essentially the probability of not receiving the credits from the administrators. Some authors, Caouette et al. (2000) and Silva (1988), refer the multivariate analysis as a very powerful tool in the risk administration of conceding the personal credit. This paper show the build and the evaluation of pattern recognition and classified rules based on the Discriminant Linear Function and the Logistic Regression, to classify the clients of credit card in one of two groups. The efficiency of the procedures was evaluated by the Lachenbruch Method, Lachenbruch (1975).

PALAVRAS-CHAVE

Reconhecimento de padrões, função discriminante linear, regressão logística, método de Lachenbruch.

KEY WORDS

Pattern recognition, credit analysis, discriminant linear function, logistic regression, Lachenbruch's method.

O SISTEMA DE SCORE NO CONTROLE DO RISCO DE INADIMPLÊNCIA

A palavra “crédito” pode ter mais de um significado, dependendo do contexto sob o qual esteja sendo tratada. Sob o ponto de vista meramente empresarial, a concessão de crédito significa a transferência da posse de um bem ou de uma quantia em dinheiro, mediante a promessa de pagamento futuro. De acordo com este conceito, pode-se entender o crédito à disposição de uma pessoa, física ou jurídica, como a capacidade da mesma em obter dinheiro, produtos ou serviços mediante compromisso de pagamento num determinado período de tempo.

Uma das maiores revoluções no crédito pessoal foi desencadeada pela criação do cartão de crédito. Também chamado “dinheiro de plástico”, é antes de mais nada um instrumento de crédito automático. Este sistema atingiu proporções que tornam obrigatória a permanente busca de técnicas que permitam o gerenciamento de um grande número de *portfólios* de empréstimos aos mais diversificados e descentralizados consumidores, de modo a obter simultaneamente o maior retorno possível.

No caso específico do crédito ao consumidor, as características observadas são:

- grandes volumes em pequenos montantes;
- processo de aprovação massificado;
- dados limitados e relativamente pobres;
- histórico de crédito do cliente disponível, mas geralmente incompleto, em grande parte dos casos limitando-se ao passado negativo ou positivo;
- utilização de bases estatísticas para avaliação do desempenho do gerenciamento do *portfólio*.

Inicialmente, é necessário distinguir, segundo enfoques estatísticos, “risco” de “incerteza”. O primeiro existe quando a tomada de decisões é baseada em probabilidades objetivas para a estimação de diferentes resultados. Desta forma a expectativa se fundamenta em dados históricos, permitindo que as decisões sejam tomadas a partir de estimativas consideradas aceitáveis. A incerteza é observada quando não se tem à disposição os dados históricos acima mencionados. Isto exige do tomador de decisões uma certa dose de sensibilidade, baseada em observações altamente subjetivas. Esta situação pode ser verificada em algumas empresas do setor varejista, onde a única informação disponível limita-se à busca de registros de negatividade junto aos serviços de proteção ao crédito.

O principal meio de controle do risco, ou pelo menos o mais utilizado, é o sistema de score. Este sistema consiste basicamente em avaliar características do novo cliente, atribuindo um determinado valor a cada característica. Em seguida os dados obtidos são usados na elaboração de um score. Com base no score obtido pelo cliente toma-se a decisão de conceder, ou não, o crédito. Para tomar tal decisão, o score é comparado com um valor previamente estabelecido, chamado *valor de corte*. É na obtenção deste último que reside a maior parte dos problemas enfrentados pelos profissionais envolvidos. A questão a ser resolvida neste ponto pode ser colocada da seguinte forma: como obter um valor de corte confiável a ponto de evitar perdas para a empresa, tanto pela aceitação, errada, de clientes que venham a se tornar inadimplentes quanto pela rejeição, igualmente errada, de clientes adimplentes.

PROPÓSITOS DO TRABALHO

Este trabalho tem os seguintes objetivos:

- utilizar as técnicas estatísticas multivariadas denominadas Função Discriminante Linear de Fisher (FDL) e a Regressão Logística (RL) na identificação de variáveis que permitam evidenciar, com certa antecedência, situações de inadimplência por parte de clientes de uma administradora de cartões de crédito a partir de informações cadastrais fornecidas pelos mesmos em propostas para adesão ao cartão de crédito administrado pela instituição..
- avaliar a eficiência dos modelos desenvolvidos, identificando aquele de melhor desempenho na identificação de variáveis acima mencionada. Pretende-se atingir este objetivo através do Método de Lachenbruch, uma técnica com características distintas da utilização de grupos de controle, prática normalmente adotada neste tipo de estudo.

JUSTIFICATIVA PARA OS MÉTODOS UTILIZADOS

A Análise Estatística Multivariada tem sido mencionada por alguns autores, como Silva (1988) e Caouette *et al.* (2000), como uma ferramenta bastante poderosa na administração do risco de inadimplência existente na concessão de crédito . Uma aplicação é a previsão do risco que corre um banco ou a administradora de crédito, e a conseqüente busca de uma forma de controle deste risco, via obtenção de um valor de corte calculado com base nas características apuradas junto à base de dados da companhia.

AS TÉCNICAS ESTATÍSTICAS MULTIVARIADAS DE RECONHECIMENTO DE PADRÕES E DE CLASSIFICAÇÃO DE OBSERVAÇÕES

As técnicas estatísticas multivariadas fazem parte do quadro de métodos quantitativos tidos como dos mais eficientes para a auxiliar a tomada de decisões. De acordo com Silva (1988, p. 93), “*O uso da estatística, da teoria das probabilidades, é um valioso instrumento para a tomada de decisão*”. Também, segundo o mesmo autor, “*Outros recursos, como o Teorema de Bayes, Análise Fatorial e Pesquisa Operacional, por exemplo, têm sua aplicabilidade nas áreas de crédito*”. Quando o analista de crédito tem à sua disposição uma regra de reconhecimento de padrões e classificação que indique previamente a chance de inadimplência de um futuro cliente, a decisão de concessão de crédito fica extraordinariamente facilitada; esse profissional pode então utilizar argumentos quantitativos em substituição a argumentos subjetivos e decidir com maior confiança

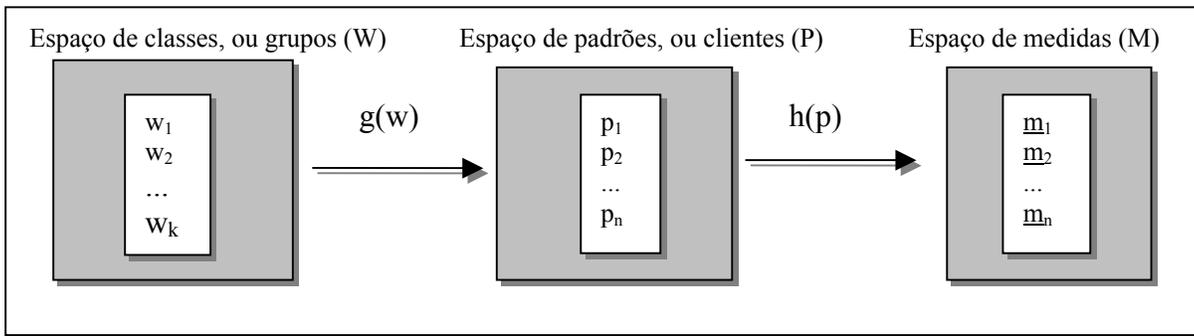
A ANÁLISE DISCRIMINANTE

A Análise Discriminante é uma técnica estatística multivariada usada na resolução de problemas que envolvem a **separação** de conjuntos distintos de observações e a **alocação** de novas observações em um conjunto específico. Integra o conjunto de técnicas usadas no Reconhecimento de Padrões, juntamente com técnicas de Programação Matemática e, mais recentemente, Redes Neurais e Algoritmos Genéticos. O reconhecimento de padrões, de um modo geral, está presente em áreas como:

- classificação de empresas, quanto à capacidade de solvência;
- processamento de sinais;
- reconhecimento de impressões digitais;
- Elaboração de perfis de consumidores;
- diagnóstico médico preliminar; entre outras.

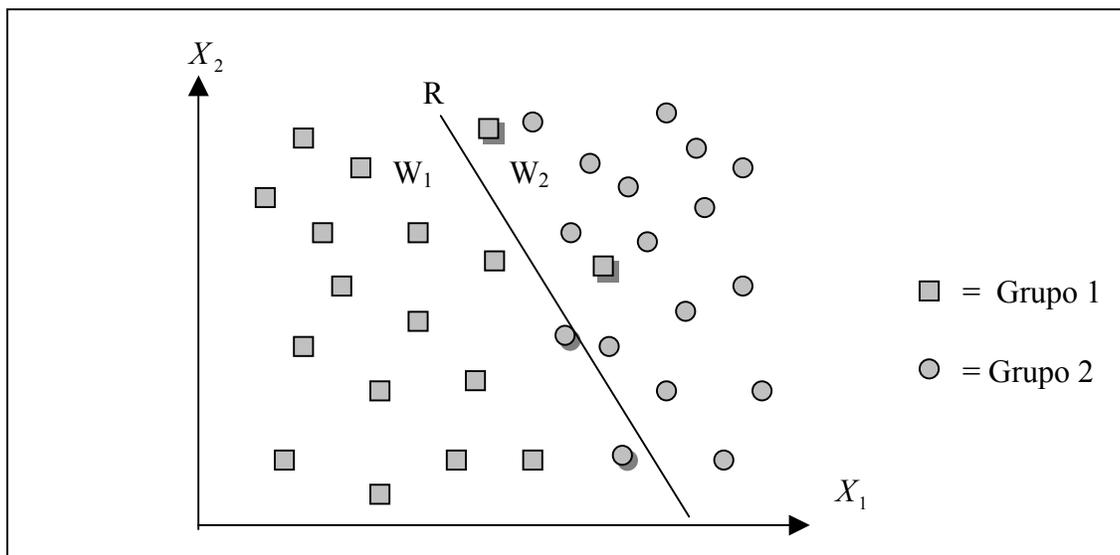
Um dos objetivos da Análise Discriminante é determinar a que grupo, dentre dois ou mais definidos *a priori*, pertence um novo elemento, com base em várias características observadas para o mesmo. Cada característica constitui uma variável independente, contribuindo para a classificação. A Análise Discriminante combina estas variáveis em uma ou mais funções, de modo a determinar para cada elemento escores de classificação com base em um banco de dados composto dos grupos de indivíduos pré-definidos. Neste trabalho, os grupos, ou classes, definidos *a priori* são dois: um de clientes adimplentes (padrões do primeiro grupo) e outro de clientes inadimplentes (padrões do segundo grupo), assim chamados os clientes com pagamento em dia de faturas e clientes com atraso superior a sessenta dias, respectivamente. O problema básico no reconhecimento de padrões pode ser apresentado da seguinte forma: dado um vetor de dimensão n de medidas das características, \underline{m}_i , obter um método de inversão do mapeamento nas relações g e h , de modo a identificar a classe geradora das medidas a partir de \underline{m}_i . Este raciocínio é ilustrado na figura 1, a seguir.

Figura 1: Mapeamento das Relações g e h .



Sejam, p.ex., dois grupos de observações resultantes da classificação segundo um determinado critério (neste trabalho, clientes adimplentes e inadimplentes) e, ainda, as variáveis X_1 e X_2 observadas para cada indivíduo pertencente a um dos grupos. Na figura 2, a seguir, tem-se o espaço discriminante e a regra de classificação, representada pela reta R , juntamente com observações dos grupos W_1 e W_2 .

Figura 2: Espaço Discriminante



Pode-se observar na figura acima que a regra de classificação, R , separa os pontos em dois grupos de forma razoável, havendo, porém uma “mistura” de observações. Esta “mistura” mostra a ocorrência de erros nas classificações. Estes erros no reconhecimento dos padrões são uma medida do poder discriminante das variáveis estudadas. Desta forma, o objetivo é a obtenção de uma regra de classificação que minimize a probabilidade de classificação errônea ou seja, otimize o reconhecimento dos padrões e a classificação dos mesmos. Também é desejável um método que permita determinar quais das variáveis estudadas contribuem de forma significativa para a discriminação dos padrões observados. Um método que possua esta propriedade permite que, na formação de um banco de dados, sejam desconsideradas as informações de pouca relevância, diminuindo desta forma não só o tempo necessário para a obtenção da referida regra de classificação, como também eventuais custos com o trabalho.

A FUNÇÃO DISCRIMINANTE LINEAR

Uma Função Discriminante Linear (FDL) tem a forma:

$$Z = \beta_0 + \beta_1 X_1 + \dots + \beta_i X_i + \dots + \beta_{p-1} X_{p-1} = \beta_0 + \sum_{i=1}^{p-1} \beta_i X_i \quad (1)$$

onde Z é o escore discriminante (variável dependente) e β_i , $i = 0, 1, \dots, p$ são os coeficientes da função nas variáveis X_i , $i = 0, 1, 2, \dots, p$, com $X_0 = 1$. A função retorna um valor Z , para um novo padrão \underline{m}_i , que é um vetor com as medidas das características observadas. O valor de Z funcionará como escore de classificação deste mesmo padrão.

A Função Discriminante Linear de Fisher, foi o primeiro método estatístico de discriminação e classificação. A idéia fundamental de *Fisher* foi transformar a observação multivariada \underline{X} , de dimensão p , na observação univariada y (escore), tal que os escores obtidos para as populações W_1 e W_2 sejam separados tanto quanto possível. Sendo $\underline{\mu}_1$, $\underline{\mu}_2$ e Σ respectivamente os parâmetros: vetor médio de W_1 , vetor médio de W_2 e matriz de covariância comum a ambas as populações, tem-se a função abaixo, que funciona como regra de classificação ao comparar y com o valor de corte m .

$$y = (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \underline{X} \quad (2)$$

Então pode-se expressar a regra de classificação como

- Alocar \underline{X}_0 em W_1 se $y_0 - m \geq 0$.
- Alocar \underline{X}_0 em W_2 se $y_0 - m < 0$.

Na realidade, os parâmetros $\underline{\mu}_1$, $\underline{\mu}_2$ e Σ não são conhecidos. Então trabalha-se com os seus estimadores: \bar{X}_1 , \bar{X}_2 e S_p , obtidos de amostras aleatórias dos grupos W_1 e W_2 com tamanhos n_1 e n_2 , respectivamente.

O estimador S_p , da matriz de covariância comum Σ , tem por expressão:

(3)

$$S_p = \frac{1}{n_1 + n_2 - 2} [(n_1 - 1)S_1 + (n_2 - 1)S_2]$$

onde S_1 e S_2 são as matrizes de covariâncias amostrais. Assim, a Função Discriminante Linear de Fisher Amostral é, segundo Johnson e Wichern (1988):

$$\hat{y} = (\bar{X}_1 - \bar{X}_2)' S_p^{-1} \underline{X} \quad (4)$$

O valor de corte m é estimado por

$$\hat{m} = \frac{1}{2} (\bar{y}_1 + \bar{y}_2) \quad (5)$$

onde \bar{y}_1 e \bar{y}_2 são as médias amostrais dos escores para W_1 e W_2 , respectivamente. Desta forma, a regra de classificação para uma observação \underline{X}_0 pode ser apresentada como

- Alocar \underline{X}_0 em W_1 se $y_0 = (\bar{X}_1 - \bar{X}_2)' S_p^{-1} \underline{X}_0 \geq \hat{m}$
- Alocar \underline{X}_0 em W_2 se $y_0 = (\bar{X}_1 - \bar{X}_2)' S_p^{-1} \underline{X}_0 < \hat{m}$

A REGRESSÃO LOGÍSTICA

A Regressão Logística consiste, fundamentalmente, na busca de um modelo que permita relacionar uma variável y , chamada “variável resposta”, aos “fatores” X_1, \dots, X_{p-1} , que, supõe-se, influenciam as ocorrências de um evento. A variável resposta deve ser do tipo dicotômica, assumindo apenas os valores 0 ou 1. Neste caso existe interesse apenas na ocorrência, ou não, do evento em questão. No presente trabalho, 0 designa “cliente inadimplente” e 1 designa “cliente adimplente”.

No modelo logístico a relação é dada por uma função tipo sigmóide, ou seja:

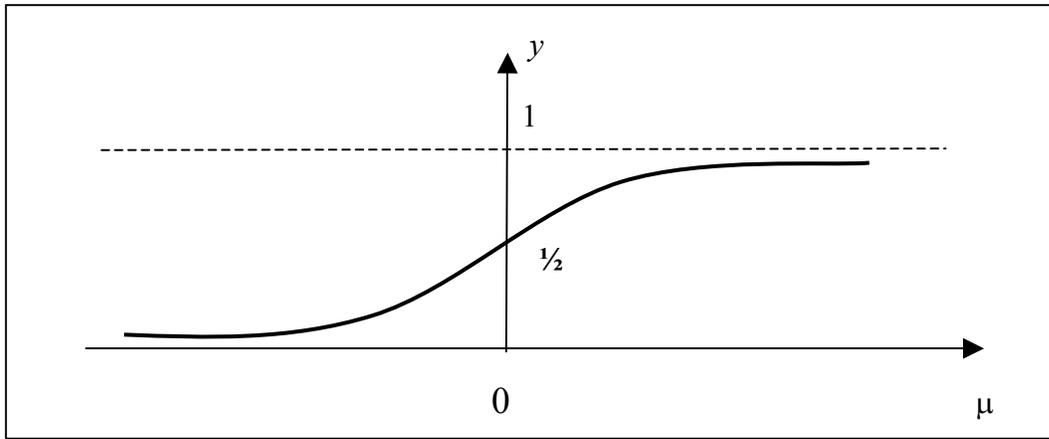
$$Y = \frac{e^\mu}{1 + e^\mu} \quad (6)$$

onde μ é uma expressão que relaciona os fatores acima e tem a forma

$$\mu = \underline{\beta}' \underline{X} + \varepsilon = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{p-1} X_{p-1} + \varepsilon = \beta_0 + \sum_{i=1}^{p-1} \beta_i X_i + \varepsilon \quad (7)$$

A função (6) assume valores entre 0 e 1 para $\mu \in (-\infty, \infty)$. A figura 3, a seguir, demonstra esse argumento.

Figura 3: Gráfico da Função Sigmóide



O escore y representa a probabilidade de ocorrência do evento associado a “1”, ou seja, $y = P(\underline{X} \in W_1) = P(\underline{X})$. Assim tem-se,

$$P(\underline{X}) = \frac{e^{\beta_0 + \sum_{i=1}^{p-1} \beta_i X_i}}{1 + e^{\beta_0 + \sum_{i=1}^{p-1} \beta_i X_i}} \quad (8)$$

que pode ser representada por

$$P(\underline{X}) = \frac{e^{\beta' \underline{X}}}{1 + e^{\beta' \underline{X}}} \quad (9)$$

A estimação dos coeficientes $\beta_0, \beta_1, \dots, \beta_{p-1}$ é feita através do Método da Máxima Verossimilhança, e o cálculo segue o Método de Newton, conforme descrito por Luenberger (1973).

O MÉTODO DE LACHENBRUCH

Tão importante quanto a obtenção de uma boa regra de reconhecimento e classificação é a determinação da eficiência da mesma. Uma regra que apresente uma taxa de erros elevada, pouca, ou nenhuma, utilidade terá. O Método de Lachenbruch, Lachenbruch (1975), é uma forma de avaliar a eficiência da regra de classificação. Esta técnica segue os passos apresentados a seguir:

1. Escolher um dos grupos (amostras).
2. Retirar uma observação do grupo.
3. Construir uma regra de classificação com as $n_1 - 1$ observações restantes do grupo escolhido e as n_2 observações do segundo grupo, isto é, para $n_1 - 1 + n_2$ observações.
4. Classificar a observação retirada, usando a função obtida anteriormente.
5. Realocar a observação descartada e repetir os passos 1 e 2 para todas as observações do primeiro grupo.
6. Repetir os passos 1 a 5 para o segundo grupo.
7. Finalmente, construir a regra de classificação com o total das $n = n_1 + n_2$ observações.

Assim obtém-se as probabilidades $P(2|1) = \frac{n_{1|2}}{n_1}$ e $P(1|2) = \frac{n_{2|1}}{n_2}$, onde $n_{i|j}$ é o número de observações do grupo i classificados como sendo do grupo j , com $i, j = 1, 2$. A proporção total esperada de erro é:

$$\hat{E}(AER) = \frac{n_{1|2} + n_{2|1}}{n_1 + n_2} \quad (10)$$

Desta forma obtém-se uma regra de reconhecimento e classificação construída com as n observações amostrais e testada com todas as referidas observações. Isto equivale a ter um grupo com n observações para o ajuste, ou treinamento, e outro grupo, também de tamanho n , para testar a eficiência do procedimento.

Uma abordagem alternativa, e utilizada com maior freqüência, é a formação de grupos de controle. Esta técnica, entretanto, carrega consigo o sério inconveniente de possibilitar que, na formação dos referidos grupos, sejam utilizadas observações responsáveis pela já mencionada “mistura”. Como resultado pode-se obter uma avaliação irreal para a regra de classificação, já que muitas das respostas estimadas para as mencionadas observações poderão não coincidir com as respostas efetivamente observadas.

SELEÇÃO DA AMOSTRA E VARIÁVEIS ESTUDADAS

Neste trabalho utilizou-se uma amostra com 707 observações, sendo 102 pertencentes ao grupo “0”, de clientes inadimplentes, e 605 pertencentes ao grupo “1”, de clientes adimplentes. Considera-se inadimplente o cliente que não honrar duas prestações mensais consecutivas, acumulando assim um atraso superior a sessenta dias para a primeira prestação. Foram estudadas as variáveis normalmente constantes em formulários de adesão utilizados pela instituição financeira fornecedora do dados, como p. ex. sexo, idade, renda e tempo de serviço no atual emprego. Foram excluídas todas as informações que de alguma forma possibilitassem a identificação dos clientes. Também foram definidas algumas variáveis “sintéticas”, com valores obtidos a partir de valores observados para outras variáveis, como p. ex. a razão entre o tempo de serviço e a idade. As variáveis são apresentadas nas tabelas 1 e 2, no final do texto, juntamente com os respectivos coeficientes. Nessas tabelas não constam as variáveis descartadas em função do baixo poder discriminante, avaliado com base no valor absoluto do respectivo coeficiente.

Juntamente com o banco de dados foi disponibilizado o resultado da avaliação sobre os dados fornecidos pelos clientes, efetuada através do método atualmente usado pela instituição fornecedora dos dados, para cada uma das observações. Estes resultados foram comparados com as classificações fornecidas pelas regras FDL e RL obtidas através dos métodos descritos neste trabalho. Os resultados são apresentados na Tabela 3, no final do texto. A comparação foi efetuada apenas para as taxas de acerto das regras envolvendo as variáveis correntes, já que o método em questão não é aplicado sobre as variáveis “sintéticas”.

OBTENÇÃO DA FUNÇÃO DISCRIMINANTE LINEAR E DO MODELO DE REGRESSÃO LOGÍSTICA

A matriz de dados de ordem $n \times p$, onde $n = 707$ observações do vetor \underline{X} , de dimensão $p = 27$ variáveis, foi utilizada para a obtenção de uma Função Discriminante Linear Amostral de Fisher, partindo-se da matriz de covariância estimada S_p (3), com 102 observações do grupo “0” (n_1), de clientes inadimplentes, e 605 observações do grupo “1” (n_2), de clientes adimplentes.

O cálculo dos estimadores de máxima verossimilhança em ambas as regras foi efetuado através de um programa computacional escrito em *Visual Basic 5.0*[®]. A Função Discriminante Linear de Fisher Amostral obtida tem seus coeficientes dados na Tabela 1. A eficiência foi avaliada pelo Método de Lachenbruch. Os resultados da avaliação constam na Tabela 3.

A mesma amostra utilizada para o ajuste da Função Discriminante Linear de Fisher foi também usada para modelar um ajuste logístico, através do mesmo programa. A matriz já mencionada foi usada para a obtenção de um Modelo de Regressão Logística, utilizando as mesmas observações e calculando os Estimadores de Máxima Verossimilhança através do Método de Newton. A eficiência foi igualmente avaliada pelo Método de Lachenbruch, com os resultados apresentados na Tabela 3.

AVALIAÇÃO DO DESEMPENHO DAS REGRAS USANDO O MÉTODO DE LACHENBRUCH

As taxas de acerto para a Função Discriminante Linear de Fisher e para o Modelo de Regressão Logística, obtidas com o acréscimo de algumas das variáveis sintéticas, são apresentadas na Tabela 3, possibilitando uma comparação das mesmas com o desempenho do método atualmente em uso. Na referida tabela percebe-se que o Modelo de Regressão Logística leva certa vantagem sobre os demais. Também é possível notar que a taxa de acertos do modelo atual para o grupo 1, de clientes adimplentes, é superior à taxa apresentada pela Função Discriminante Linear para o mesmo grupo. Entretanto, para o grupo 0, de clientes inadimplentes, esta última apresentou uma taxa consideravelmente superior. Na mesma tabela percebe-se que, entre os 136 clientes que se mostraram inadimplentes, 55 (40,44%) foram aprovados pelo método atual. Este índice é suficiente para colocar em dúvida a eficácia do atual método na previsão de inadimplência.

CONCLUSÕES

O emprego de métodos científicos, especialmente os quantitativos, na gestão de operações de crédito pessoal não é uma prática muito difundida no Brasil, exceto em empresas que operam em âmbito internacional e que costumam utilizar os referidos métodos em seus países de origem. As ferramentas computacionais disponíveis atualmente permitem, com relativa facilidade, a automatização de tais procedimentos. Desta forma não há razão que justifique a sua não utilização, principalmente se for levado em consideração que qualquer redução na inadimplência só tende a reduzir os custos da concessão de crédito.

A instituição de crédito tem nos métodos aqui estudados (Função Discriminante Linear de Fisher e Regressão Logística) uma ferramenta que, aliada à informática, possui inegável confiabilidade quanto aos resultados apresentados. Não se deve esquecer, contudo, que esta ferramenta necessita para o seu desenvolvimento e posterior aplicação, de um banco de dados da maior qualidade a fim de aumentar ainda mais o seu potencial de auxílio à tomada de decisões.

A diferença de eficiência, verificada pelo Método de Lachenbruch, entre os dois procedimentos apresentados não é significativa. Porém, ligeira vantagem favorece a Regressão Logística. O fundamental na atividade de análise de crédito é um banco de dados com informações que realmente possibilitem a discriminação entre os diversos padrões. Se o estudo desenvolvido for efetivamente implementado na instituição financeira, é possível também detectar a perda do “poder discriminante” de determinadas variáveis, traduzida pela diminuição do respectivo peso em valor absoluto. Este procedimento permite a obtenção de regras de classificação atualizadas, já que acompanha as alterações dos perfis dos clientes de ambos os grupos. Como exemplo de alteração de perfil pode-se citar a posse de aparelho de telefonia celular. Até cinco anos atrás a posse deste bem geralmente indicava que o cliente possuía uma renda sensivelmente superior à média. A abertura do setor de telecomunicações possibilitou que uma grande parcela da população tivesse acesso ao produto. Tal fenômeno acabou por tornar no mínimo temerária qualquer inferência sobre a renda de um cliente a partir desta informação. Desta forma fica enfatizada a necessidade de um banco de dados com informações discriminantes e permanentemente atualizadas.

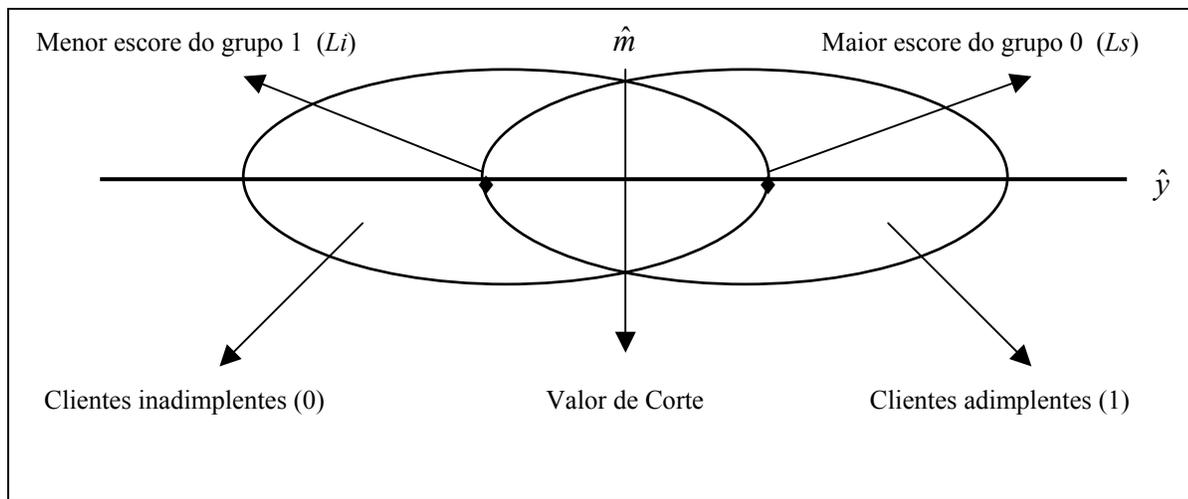
É necessário ressaltar que não foi dedicada atenção especial à “região de superposição de escores”, ilustrada na Figura 4, adiante. Esta região refere-se aos escores de clientes adimplentes, que de acordo com a regra teriam avaliação negativa, e também de clientes inadimplentes com escores que permitiriam uma classificação positiva. Situações como esta evidenciam a necessidade de normas gerenciais que orientem a tomada de decisão nestes casos. Vale lembrar que tais normas devem estar de acordo com o perfil da instituição, agressivo ou conservador, e levar em consideração o custo representado por uma decisão equivocada, representado tanto pela aceitação de um “mau” cliente, como pela rejeição de um “bom” cliente. A sugestão que se faz no trabalho é que se fixe limites para esta região de forma que um padrão situado neste intervalo (L_i , L_s) tenha uma análise com intervenção pessoal do analista de crédito, não ficando apenas na forma da rotina automatizada computacionalmente.

Diante das considerações anteriores pode-se concluir, ainda, que:

- Uma regra de reconhecimento e classificação obtida por qualquer dos métodos apresentados neste trabalho, só terá alguma eficácia como ferramenta de apoio à tomada de decisões se for obtida a partir de um banco de dados permanentemente atualizado.
- A busca por informações que apresentem a necessária consistência e alto poder discriminatório deve ser uma preocupação constante do tomador de decisões.

- Seja qual for o sistema escolhido, é de extrema importância a elaboração de normas que orientem a tomada de decisão em situações nas quais o escore obtido pelo cliente encontre-se na região de superposição. Para levar a efeito a elaboração destas normas é indispensável que se tenha em mente a cultura da empresa.

Figura 4: Região de Superposição de Escores



Finalmente, deve-se ressaltar que, qualquer que seja o método adotado, e não importando quão eficaz ele seja, é necessário ter a consciência de que uma regra de reconhecimento e classificação é uma ferramenta de apoio à tomada de decisão, integrando portanto o conjunto de técnicas à disposição dos profissionais que atuam na área. É importante, também, destacar a necessidade da mudança de cultura das instituições de crédito que preferem praticar juros extraordinariamente altos ao invés de, com base em uma metodologia científica, selecionarem adequadamente os seus clientes, de forma que se tenha um custo reduzido no mercado para o serviço oferecido.

TABELA 1 – COEFICIENTES DA FUNÇÃO DISCRIMINANTE LINEAR DE FISHER

| VARIÁVEIS | COEFICIENTES |
|-------------------------------|--------------|
| LIMITE | 0,0001 |
| SEXO | - 2,5097 |
| TEMPO DE RESIDÊNCIA | - 0,0806 |
| SEGURO AUTOMOTIVO | -1,0719 |
| SEGURO RESIDENCIAL | 0,1594 |
| CARTÃO SEGURADO | 2,2729 |
| SEGURO DE VIDA | 0,9609 |
| RENDA | - 0,0004 |
| IDADE | - 0,0166 |
| TEMPO NO ATUAL EMPREGO | - 0,1010 |
| IDADE DO CÔNJUGE | - 0,0051 |
| TELEFONE CELULAR | 7,1033 |
| ESTADO CIVIL | 0,2520 |
| TIPO DO DOCUMENTO APRESENTADO | - 0,2060 |
| ESCOLARIDADE | - 0,7977 |
| TIPO DE RESIDÊNCIA | - 1,1765 |
| SETOR DE ATIVIDADE | - 0,9055 |
| CEP | 0,0002 |

Tabela 2 – COEFICIENTES DO MODELO DE REGRESSÃO LOGÍSTICA

| VARIÁVEIS | COEFICIENTES ($\hat{\beta}$) |
|-------------------------------|--------------------------------|
| LIMITE | - 0,0006 |
| SEXO | 2,8811 |
| TEMPO DE RESIDÊNCIA | 0,1643 |
| SEGURO AUTOMOTIVO | 0,1260 |
| SEGURO RESIDENCIAL | - 0,3739 |
| CARTÃO SEGURADO | - 5,3658 |
| SEGURO DE VIDA | - 0,7594 |
| RENDA | 0,0012 |
| IDADE | 0,0298 |
| TEMPO NO ATUAL EMPREGO | 0,2443 |
| IDADE DO CÔNJUGE | 0,0092 |
| TELEFONE CELULAR | - 6,6774 |
| ESTADO CIVIL | - 0,0615 |
| TIPO DO DOCUMENTO APRESENTADO | 0,7448 |
| ESCOLARIDADE | 0,9614 |
| TIPO DE RESIDÊNCIA | 1,5737 |
| SETOR DE ATIVIDADE | - 0,8687 |
| CEP | - 0,0003 |
| CONSTANTE | 19,6195 |

TABELA 3 – RESULTADOS DA AVALIAÇÃO DOS MODELOS DE RECONHECIMENTO DE PADRÕES PELO MÉTODO DE LACHENBRUCH

| MODELO DE RECONHECIMENTO DE PADRÕES | VARIÁVEIS CORRENTES | | VARIÁVEIS CORRENTES E VARIÁVEIS SINTÉTICAS | |
|-------------------------------------|----------------------------|--------------------------|--|--------------------------|
| | Grupo 0 (Inadimplentes) | Grupo 1 (Adimplentes) | Grupo 0 (Inadimplentes) | Grupo 1 (Adimplentes) |
| Função Discriminante Linear | 92,16 % | 92,40 % | 69,61 % | 58,84 % |
| Modelo de Regressão Logística | 99,02 % | 99,83 % | 87,25 % | 98,84 % |
| Método Atualmente Utilizado | 59,55 % | 96,49 % | - x - | - x - |

Artigo recebido em 24.08.2001. Aprovado em 24.04.2002

REFERÊNCIAS BIBLIOGRÁFICAS

CAOUILLE, J. B. , ALTMAN, E. I. & NARAYANAN, P. Gestão do risco de crédito. O próximo grande desafio financeiro. São Paulo: Qualitymark , 2000.

JOHNSON, R. A. & WICHERN, D. W. – Applied multivariate statistical analysis. 2. ed. New Jersey: Prentice Hall International, Inc., 1988.

LACHENBRUCH, P. A. – Discriminant Analysis. New York: Hafner Press, 1975.

LUENBERGER , D. G. – Introduction to linear and nonlinear programming. Reading: Addison-Wesley Publishing Company , 1973.

SILVA, J. P. da – Análise e decisão de crédito. São Paulo: Atlas, 1988.