

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO

DEPARTAMENTO DE ECONOMIA

MONOGRAFIA DE FINAL DE CURSO

CREDIT SCORING

Pedro da Costa Carvalho

No. de matr 9515166

Orientador: Marcio Garcia

Junho de 1999

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO

DEPARTAMENTO DE ECONOMIA

MONOGRAFIA DE FINAL DE CURSO

Credit scoring

Pedro da Costa Carvalho

No. de matr 9515166

Orientador: Marcio Garcia

Junho de 1999

Índice	
Introdução	3
I - Tipos de análise	7
I.1- Método Subjetivo	8
I.2 – Método Objetivo	8
I.2.1- Credit and Behavioral score	8
I.2.2- Análise de carga e gastos	12
I.3 – Método Misto	12
II- Utilização da Scoring	13
II.1- Utilização no Brasil e no exterior	13
II.2- Utilização em Bureau de Crédito	15
III- Desenvolvimento de um modelo	17
III.1- Banco de Dados	17
III.2 – Definição das variáveis	19
III.2.1 – Variáveis do Produto	21
III.2.2 – Variáveis Cadastrais	21
III.2.3 – Variáveis da Capacidade de Pagamento	22
III.3 – Análise Bivariada	23
III.4- Regressão Original	28
III.5 – Analisando o Modelo	32
III.5.1 – Teste de significância	33
III.5.2 – Teste de correlação	34
III.6 – Regressão Final	38
IV – Testes de Performance	41

V – Ponto de Corte	44
VI- Utilização e características de um Credit Scoring	48
VI.1- Comparação das políticas de crédito	48
VI.2 – A necessidade do score	49
VII- Extensões do modelo	51
VIII- Conclusão	53
IX- Anexos	56
X- Bibliografia	59

Introdução

Uma das vantagens e consequência que a estabilização da nossa economia nos trouxe com certeza foi a expansão do crédito para pessoa física. Podemos dizer que isto se deu basicamente pelo aumento da capacidade das pessoas e firmas se planejarem, gerada pela manutenção do poder de compra. Que entre outros benefícios trouxe juros nominais mais ou menos constantes permitindo o uso de financiamento com taxas de juros pré-fixadas.

Sem dúvida nenhuma este crescimento foi responsável em grande parte pelo *boom* de consumo que houve nos primeiros meses após o real e depois disso se firmou como uma importante forma de pagamento para o consumidor. Porém as instituições financeiras brasileiras não estavam preparadas, em grande parte, para entrar mais forte neste mercado. Pois a inflação gerava um ganho para estas, não havendo assim a necessidade de se procurar fontes alternativas de receita.

Assim sendo os bancos passaram a se interessar mais por este nicho de mercado, alguns encarando como forma de especialização, outros como forma de ganhar escala e produtividade. Ainda temos instituições que buscaram outras formas de aumentar sua receitas como a cobrança de tarifas mais elevadas, ou que se especializaram em certos segmentos específicos do mercado de capitais. Outro fator relevante que aconteceu paralelamente foi a entrada de grandes bancos estrangeiros que fez a concorrência ficar mais acirrada e que normalmente já têm grande experiência neste segmento, o que lhes poupa bastante tempo e capital. Tudo isto fez com que várias instituições ao longo dos últimos cinco anos quebrassem ou simplesmente fechassem as portas, sendo que muitas devido a má concessão de crédito.

Dentro do crédito ao consumidor todos os segmentos tem características semelhantes, como: grandes volumes, pequeno / médio montante, processo de aprovação massificado, dados financeiros do cliente limitados e pobres e além de proporcionar um pequeno ganho por operação.

Diante destes fatos urge a necessidade de estudarmos e nos prepararmos melhor para a concessão de crédito massificado e um dos pontos mais importante para isto é a análise do crédito. O objetivo deste trabalho é analisar os diversos meios de elaborar uma política de crédito e de colocá-la em prática levando em consideração os custos e economias decorrentes de cada sistema.

Para tanto no, capítulo I, iremos descrever os diversos tipos ou métodos adotados para análise de crédito. No capítulo II, abordaremos brevemente a utilização no Brasil e no exterior do Credit Scoring e sua utilização em bureau de crédito. No capítulo III vamos, com uma massa de dados real, desenvolver um modelo de Credit Scoring relacionando alguns itens da nossa clientela com o conceito de bom cliente. No próximo capítulo iremos testar o modelo desenvolvido com objetivo de checar a sua performance. No capítulo V iremos definir o ponto de corte do modelo, ou seja, a partir de que score o cliente teria seu crédito aprovado. No capítulo VI, poderemos definir quando há realmente necessidade da utilização do scoring e como este deve ser utilizado. E finalmente no capítulo VII iremos apresentar algumas extensões do modelo desenvolvidas com a mesma tecnologia, porém com diferentes objetivos dentro do crédito ao consumidor.

I – Tipos de análise

Como qualquer outro tipo de decisão temos basicamente duas formas para avaliar e decidir o que fazer: analisar o problema objetivamente ou subjetivamente, ou até mesmo usando ambos os sistemas.

A parte que iremos estudar é simplesmente uma das formas de se explorar um banco de dados com uso de ferramentas estatísticas. Só relativo a área em questão poderíamos citar várias outras maneiras de se utilizar este método além de na concessão do crédito, como: marketing (*mail list*, *cross selling*, segmentação), gerencial (definir preço, fazer previsão de perda) e cobrança (quem cobrar primeiro, *self cure...*). Fora o uso de técnicas estatísticas em produtos não financeiros, como o seguro aonde estas já são utilizadas a muito mais tempo.

Apesar da *scoring* ter sido desenvolvida com o objetivo de melhorar e automatizar a concessão de crédito, vários outros usos foram dados aos modelos feitos. A análise de portfólios além de se basear somente no nível de perdas passadas, pode agora contar com esta tecnologia que tem um poder de previsão muito melhor. Isto se deve ao fato da análise revelar o nível dos créditos atuais e não do passado. Esta melhoria acabou por incentivar o mercado secundário de securitização pois há maior transparência na avaliação da qualidade dos recebíveis.

A própria administração do crédito também ganhou com uso desta ferramenta, pois por exemplo pode se dar mais atenção aos créditos com baixo scores. Possibilitando auditar uma maior quantidade destes ou dar prioridade no momento da cobrança. Outro processo facilitado é o de previsão de perdas, relacionando os scores com os relativas probabilidades de default. Ou seja a adoção deste método pode tornar mais fácil e consistente algumas decisões importantes.

É bom esclarecer que o que está em discussão é a análise do crédito e não as confirmações dos dados declarados pelo proponentes, que devem, se necessárias, feitas por profissionais experientes que conheçam todos os “truques” e fraudes relativos ao mercado específico.

I.1) Método Subjetivo

Este método consiste em definir uma política de crédito de forma subjetiva com base nos conhecimentos existentes do analista ou gerente de crédito. Por exemplo, no Brasil, temos que a categoria dos autônomos é considerada mau pagadora então para estas iremos exigir além dos pré-requisitos básicos uma experiência de crédito. Alguns conceitos teóricos também são utilizados, como por exemplo que o proponente só pode comprometer até 30% da renda com pagamento de financiamentos.

Podemos apontar como desvantagens do método subjetivo os pré-conceitos existentes que nem sempre estão de acordo com a realidade, a diferença de avaliação dentro de uma mesma instituição devido a diferentes interpretações das regras e até mesmo a interferência de questões de ordem emocional do analista de crédito na avaliação do mesmo. A única vantagem deste sistema, se bem implementado, em relação ao método objetivo é não estar correndo o risco de negar o crédito para bons clientes.

I.2) Método objetivo

Há alguns tipos diferentes de análise que usam este método, sendo que estes também podem ser usados em conjunto ou isoladamente.

I.2.1) Credit e Behavioral scoring

O Credit Scoring, ou escoragem de crédito, consiste basicamente em usar características reais do cliente e seus hábitos de pagamento e com métodos estatísticos desenvolver um score ou sistema de pontuação. O behavioral scoring, ou scoragem

comportamental, adota o mesmo procedimento só que usa o comportamento de pagamento do próprio cliente.

A teoria por trás destes sistemas está que o comportamento futuro do cliente irá ser igual a de clientes parecidos com ele que usaram o mesmo produto, ou no segundo caso que a experiência com o próprio poderá definir seu comportamento no futuro.

Deste modo o primeiro é utilizado quando o consumidor está solicitando o crédito pela primeira vez, e o segundo para a decidir se o limite do cliente deve ser aumentado, mantido, reduzido ou até mesmo zerado, ou também se devemos conceder um segundo empréstimo. No primeiro caso só podemos usar a scoragem de crédito já que o cliente ainda não possui experiência com a empresa no produto. Para o segundo poderíamos utilizar qualquer um dos dois sistemas, porém a comportamental tem se revelado um método mais preciso.

Podemos ainda diferenciar o score genérico do score específico, sendo o primeiro desenvolvido levando as informações cadastrais do cliente, seu atual comprometimento no mercado e seu comportamento passado. Já o outro, além de considerar estes dados leva em conta características do mercado ou da região para qual está sendo desenvolvido, ou ainda o tipo e/ou percentual de garantia dada pelo cliente. A vantagem do genérico é de ser mais barato e poder ser encontrado em pacotes já pré-definidos possibilitando assim a escolha entre determinados modelos e evitando a demora de um desenvolvimento. Porém, o específico como era de se esperar tem um grau de precisão maior, e torna o modelo mais simples para o usuário, na medida que fica mais fácil deste entender a decisão do modelo.

Podemos ver isto no setor imobiliário, onde garantia de uma agência governamental ou a proporção do imóvel a ser financiado é um fator relevante na

determinação da probabilidade de haver uma inadimplência. A tabela abaixo nos mostra a relação entre o percentual dado de entrada, a inadimplência e o total perdido .

Tabela I.1: Relação entre percentual de entrada, inadimplência e o total perdido.

	10 –70	71-80	81-90	91-95	96-100
Inadimplência	0,24	1,11	2,74	6,20	2,16
Total perdido	22,3	29,3	34,4	47,9	39,2

Fonte: Credit Risk Modeling, Design and Application

* Total perdido percentual não recuperado após a venda do imóvel, considerando os custos legais e de venda. Entretanto, sem levar o pagamento do seguro, normalmente obrigatório quando a entrada é menor que 20%.

Entre as desvantagens deste scores está o fato de ser imprescindível a existência de um banco de dados sobre o produto ou cliente em questão, além de que este deve ser confiável para tal utilização. Isto impossibilita o uso deste método para novos produtos. Outro ponto negativo é a grande quantidade de bons clientes rejeitados pelos modelos.

Para isto podemos utilizar algum tipo de repescagem para com isso minimizar a perda de bons clientes. Poderíamos também permitir aos gerentes a abertura de exceções tanto para aprovar quanto para negar o crédito indo contra o sistema. Contudo o nível de exceções não deve ultrapassar 5% do volume, pois isto indica o mau uso do sistema ou a sua desatualização. Além de que estas devem ser acompanhadas tanto a nível de fraude quanto de performance de cobrança.

Entre as suas vantagens podemos citar o custo de análise bem mais barato que usando analistas de crédito, a uniformidade das decisões de crédito, a previsibilidade e identificação das perdas, maior controle das operações, maior agilidade na liberação do

crédito . Entretanto para a primeira razão ser válida temos que ter um volume de operação que justifique a implantação e desenvolvimento de um sistema.

Devemos ter em mente que a qualidade dos dados são cruciais para o resultado da scoragem. Devido a este fato a ficha cadastral é muito importante para o score pois é a partir desta que será formada o banco de dados. Dependendo do tipo de crédito ao consumidor ela irá apresentar diferente tamanho e configuração, temos aqui que ressaltar que a mesma deverá ter a maior quantidade possível de informações, porém sem se tornar longas demais afastando assim alguns clientes. Além disso devemos pedir informação específica ao invés de categorias com idades de 21 a 29 anos. Outra importante questão além da segmentação da proposta é a sua regionalização para adequar as necessidades específicas de cada região.

Entre as dificuldades para implementação podemos citar:

- a) Tratamento do banco de dados difícil (*data warehousing, data mining*);
- b) Uso de complicadas técnicas estatísticas para seu desenvolvimento;
- c) Falta de firmas especializadas no assunto no Brasil e da falta de experiência das firmas estrangeiras no mercado brasileiro e
- d) resistência das pessoas acostumadas ao sistema antigo, principalmente devido a substituição de pessoas por máquinas e da sua complexidade.

Outra questão relevante é o sigilo em torno dos pesos e das variáveis para que possamos minimizar a possibilidade de fraude, isto é ainda mais relevante no caso do crédito indireto (através do revendedor ou lojista).

I.2.2) Análise de carga de gasto

Este tipo consiste em analisar a capacidade de pagamento do cliente no momento em que este está efetuando o financiamento com todos os dados relativos aos seus gastos e rendas disponíveis.

Assim sendo é fundamental para que se implemente este sistema de análise que se tenha conhecimento do maior número possível de compromissos do cliente e da sua renda. Esta última não é problema em geral, a não ser para os autônomos onde não há normalmente uma fonte segura para confirmar seus rendimentos. O grande desafio se encontra realmente em ter os gastos do cliente de forma rápida, completa e confiável. Esta questão é ainda maior no Brasil onde diferente dos Estados Unidos não há instituições especializadas neste tipo de serviço. Sendo isto não só problema para este tipo de análise mas para qualquer pois não temos qualquer forma de cadastro positivo do cliente.

I.3) Método misto

A mistura dos tipos pode se dar com a fixação dos pesos no sistema de scoring de modo subjetivo, ao invés de se utilizar métodos estatísticos. Esta pode ser a saída para instituições de menor porte que não tenham como arcar com o desenvolvimento de um sistema completo. Claro que poderá se estar incorrendo no mesmo erro de análise do que antes, mas já teremos o benefício da uniformização e da agilidade na concessão. Além de que os pesos podem ser ajustados através de relatórios de desempenho.

II- Utilização da Scoring

No exterior, sendo que o único mercado onde foi possível obter algum tipo de dado foi o norte americano, temos notícia da utilização dos métodos de scoring desde do início da década de 50 ou 60, dependendo do autor. Atualmente temos até a venda de produtos de score pré-definidos pela internet, além disso temos muitas firmas especializadas neste segmento ou só de scoring para cobrança.

II.1) Utilização no Brasil e no exterior

Este segmento se desenvolveu muito na carona do crescimento do setor de crédito para o consumidor (dados EUA), hoje em dia o uso da scoring lá já é rotina e está sendo amplamente utilizado em todos os mercados de crédito ao consumidor. Atualmente cerca de 95% das propostas de crédito na indústria de cartão de crédito não passa mais na mão de uma pessoa para avaliação. Apesar disto não é unanimidade a análise de crédito baseada somente no score, muito pelo contrário muitos consultores e agências de crédito discordam deste procedimento ou do grande grau de automação de alguns processos de concessão.

Temos também a utilização da scoring para créditos de maior valor, como o crédito imobiliário, só que sendo simplesmente como um dado complementar a análise do crédito. Para este segmento o impulso para utilização deste método só se deu na década de 1990. Principalmente a partir de 1995 quando as duas agências governamentais que compram créditos imobiliários, Fannie Mae e Freddie Mac, incentivaram os emprestadores iniciais a considerarem os scores dos bureau de crédito acima como um dado importante a ser analisado.

A GE Capital, maior firma do mundo em crédito ao consumidor, desenvolveu um modelo próprio para o crédito imobiliário a partir de sua base de dados de mais 6 milhões

de empréstimos nesta área. O OmniScore como ficou conhecido varia entre 0 e 1000 pontos, sendo quanto maior o score menor o risco. Colocando dados específicos do crédito imobiliário a Ge acredita conseguir uma precisão de duas a cinco vezes maior que se considerasse somente dados gerais. Apesar disso como eles próprio definem “OmniScore deve ser usado como uma ferramenta. Nós nos vimos como consultores de administração de risco”. Ou seja, a avaliação não é só baseada neste.

No Citibank, por exemplo, a utilização da scoragem na concessão do crédito se dá através do Credit Scoring, behavioral scoring e mail solicitation scoring, tanto na matriz quanto nas suas filiais ao redor do mundo. Porém não há obrigatoriedade para sua utilização até por que em certos casos não há como, por motivos já citados.

No Brasil, além dos já tradicionais mercado de Crédito Direto ao Consumidor, cheque especial, crédito pessoal, leasing, cartão de crédito e Private Label outras formas de crédito massificado estão surgindo como alternativa para pagamento financiado ou de empréstimo que poderão utilizar o sistema de scoragem. Podemos citar como exemplo o cartão de débito que só recentemente tomou força, o CDC – Eletrônico e o smart card.

Diferentemente no Brasil ainda consideramos o crédito imobiliário de alto valor para o uso destes tipos de sistemas, podemos relacionar isto a pouca experiência tanto neste segmento como também no uso dos referidos sistemas.

Em termos do uso por parte de empresas brasileiras há uma grande variação. A Credicard já usa este sistema pelo menos desde 1987, o que provavelmente se deve a participação do Citibank no capital desta. A Losango só recentemente criou uma área e contratou uma consultoria especializada (Fair Issac consultoria pioneira nos EUA) para desenvolver a sua utilização. A Fininvest só começou a pesquisar o assunto a partir de mais ou menos 1995.

Fora do mercado financeiro já temos outras empresas usando esta tecnologia para limitar o crédito de seus clientes, principalmente no setor de telecomunicações nas empresas recentemente privatizadas e nas suas concorrentes, como a ATL. Porém uma conclusão podemos tirar com certeza, o uso em firmas de menor porte, tanto dentro quanto fora do mercado financeiro, ou não existe ou é insignificante.

II.2) Utilização em credit bureau

Os credit bureau nada mais são do que empresas especializadas em colher, armazenar e vender informações de pessoas jurídicas e físicas. Nos Estados Unidos os relatórios de pessoas contém cinco partes:

1. Identificação do cliente: Endereço residencial, data de nascimento, telefone, endereço anterior, número da assistência social, empregador e seu endereço, renda, cargo e data de admissão.
2. Score desenvolvido pelo bureau
3. Informações judiciais e de cobrança
4. História de pagamento
5. Informação sobre outros pedidos de relatórios deste mesmo cliente

No Brasil já existem empresas especializadas como o SCI e a Serasa, porém somente para pessoas jurídicas. Para pessoas físicas temos no máximo as informações negativas e de outros pedidos do mesmo cliente através do SPC – Serviço de Proteção ao Crédito dos Clube Diretores Lojista (Associação de varejistas). E também uma série de empresas que informam as restrições relativas a cheques, como: sem fundos, extravio e roubo.

Há pouco tempo a Serasa lançou no mercado um produto completo como o norte-americano, porém ainda não houve adesão significativa. A princípio por dois motivos

básicos, primeiro pelo alto preço e segundo pelo “ciúme” das empresas em cederem seus bancos de dados para consulta de todos. Entretanto a questão do custo deve ser resolvida com a concorrência, o que primeiramente deve acontecer com o próprio SCI que a pouco tempo foi adquirido por um dos três maiores credit bureaus, o Equifax. E a parte sentimental deve ser resolvida com o crescimento do setor e a sua profissionalização, além da entrada de bancos estrangeiros no negócio que já usam este tipo de ferramenta.

A consolidação deste tipo de produto é muito importante para aumentar a segurança do emprestador, o que vai permitir um maior grau de investimento e uma diminuição no custo para o tomador final levando a consequente o crescimento do setor. Além das informações tradicionais positivas e negativas esses tipos de relatório trazem dentro de si credit scores pré concebidos com base nas suas informações para classificar o cliente, permitindo desta forma o acesso a pequenas e médias empresas este tipo de tecnologia. E mesmo para firmas maiores em algumas aplicações mais simples que envolvem pouco risco ou um pequeno montante de dinheiro podendo prevalecer a scoragem feita pelo bureau.

III – Desenvolvimento do Modelo

Antes de iniciar precisamos definir qual o objetivo da scoring: produto, região e público alvo. Como no exemplo a seguir iremos desenvolver um modelo específico é preciso que as características dos clientes e mercados sejam o mais homogêneo possível, pois isto irá influenciar diretamente na qualidade do modelo.

Esta importante questão que às vezes tem respostas simples como a diferenciação entre a concessão para cartão de crédito e financiamento imobiliário, pode se complicar quanto tratarmos, por exemplo, com crédito para veículos novos e usados ou cartão de crédito especial ou comum. A única maneira é rodar modelos com os produtos em separados e conjuntamente para observar a consistência e previsibilidade deles.

Temos ainda que escolher qual dos três métodos será utilizado para calcular o modelo. Análise discriminante o de mais fácil uso, regressão múltipla o mais utilizado ou rede Neural o mais recente. Iremos utilizar o método de regressão múltipla por se tratar de uma poderosa ferramenta estatística e a mais utilizada nestes casos.

III.1) Banco de dados:

Os scores são determinados por modelos estatísticos que atribuem pontos, ou pesos (dados pelos coeficientes obtidos na regressão final do modelo), a indicadores de comportamento de pagamento e a dados cadastrais do proponente (dados por variáveis dummy para cada classe de cada variável analisada).

O primeiro passo para o desenvolvimento de um credit scoring é a extração (ou aquisição de empresas semelhantes) de uma amostra de contratos passados que seja representativa do portfólio da empresa. Tal amostra deve contar com um número de observações ao redor de 1.500 contratos de clientes considerados bons pagadores e 1.500

contratos de clientes considerados maus pagadores. Além disso, nos modelos mais sofisticados recomenda-se a adição de 1.000 propostas rejeitadas, de modo que, a representatividade do modelo não seja influenciada pelas decisões de concessão tomadas, errada ou acertadamente, antes da implantação do credit scoring.

Tal amostra deve ser submetida a um ajuste que garanta a representatividade da carteira da empresa analisada para que a proporção de bons e maus clientes da amostra seja semelhante a proporção de bons e maus clientes da carteira real da empresa durante o período analisado.

É importante também selecionar contas ou financiamentos com inícios durante o ano todo para observarmos eventuais sazonalidades, sendo que alguns modelos a data da contratação é uma das variáveis explicativas. E como a tendência é que com o passar do tempo os modelos fiquem menos previsíveis devido as mudanças das características dos bons e maus clientes, é fundamental selecionarmos clientes recentes normalmente de até 24 meses atrás ou no caso de financiamento imobiliário de até 7 anos.

Para o desenvolvimento de nosso modelo foi extraída uma amostra já ajustada quanto a representativa dos clientes de Credito Direto ao Consumidor do Estado do Rio de Janeiro referente ao ano de 1998 da financeira Lecca. Tal amostra contém **218** contratos de maus clientes e **3.228** contratos de bons clientes, denotando que tal financeira possui um índice de **6,33%** maus clientes em sua carteira de CDC Lojista.

Precisamos então definir o conceito de bom e mau cliente. Normalmente é definido um número de dias que o cliente pode atrasar um pagamento com base num estudo feito na base de dados em questão. Sendo um parâmetro aceito geralmente quando mais de 50% dos atrasados passam para próxima faixa de atraso, o que geralmente ocorre na passagem da fase de até 60 para a de até 90 dias de atraso.

Algumas empresas tentaram incrementar esta definição com a implementação de conceitos mais complexos, como considerando o cliente mau se este atrasou mais de 30 dias por mais de cinco vezes ou 60 dias mais de 3 vezes ou ainda mais de 90 dias uma única vez. Não há neste caso como medir a melhor definição, mas até onde podemos avaliar usar uma única faixa de atraso além de manter a definição mais simples parece mais coerente. Pois decidir se o cliente é bom ou não, baseado na faixa onde a maioria dos créditos que atrasaram já foram recuperados tende a ser uma avaliação melhor do que pela quantidade de vezes que atrasou em várias faixas. Inclusive os clientes crônicos de trinta dias são em geral os mais lucrativos.

Uma forma de atenuar o problema da classificação podemos criar uma classe indeterminada, onde o cliente não é considerado nem bom e nem ruim. Desta forma, por exemplo, até 60 dias de atraso este seria considerado bom, entre 60 e 90 dias indeterminado e acima de 90 dias ruim. Neste caso o clientes classificado na categoria indeterminada não seria usada no desenvolvimento do modelo.

Neste caso vamos considerar maus clientes aqueles que atrasaram mais de 60 dias alguma prestação, e bons clientes aqueles que, ao contrário, não atrasaram mais de 60 dias nenhuma das prestações. Já as informações a respeito dos clientes rejeitados não serão usadas em nosso modelo, pois a manutenção dos dados referentes aos clientes negados não consiste ainda numa praxe das instituições financeiras brasileiras, de modo que, não foi possível a obtenção de tal amostra auxiliar junto a financeira citada anteriormente.

III.2) Definição das variáveis:

Cada instituição deverá usar seu próprio modelo de credit scoring já que a composição do portfolio depende do produto, da área de atuação e do público alvo, entre

outras características, de cada empresa. Assim sendo, a experiência individual de cada instituição será fundamental na definição tanto da representatividade da amostra e da definição de bom ou mau cliente quanto das informações que devem constar da ficha cadastral que o cliente deve preencher ao solicitar o crédito, já que será a partir de tais informações e de checagens junto as referências dadas pelo cliente e a Bureaus de crédito que o modelo será gerado.

Obviamente esta escolha esta limitada ao que temos no banco de dados, as mais utilizadas normalmente são idade, renda, faixa do CEP...Não podemos esquecer de levar em conta as diferenças regionais, como por exemplo a importância do cliente de baixa renda possuir um telefone residencial no Brasil e isto não ser relevante nos países desenvolvidos. Outro fator a ser considerado é a inclusão de aspectos como religião ou raça que poderiam ser considerados como discriminação, nos EUA já existe regulamentação sobre o assunto que define que se deve ter razões objetivas para se negar um crédito (Equal Credit Opportunity Act).

Outro ponto a ser levado em conta é se a variável é de fácil interpretação para os usuários finais. Pois pode tornar o modelo uma caixa preta, e conseqüentemente tornar de difícil aceitação o resultado final, pois estes irão achar as explicações pouco convincentes ou não irão entende-las.

Para o desenvolvimento de nosso trabalho, iremos utilizar um modelo simplificado onde somente as variáveis mais utilizadas neste tipo de modelagem serão incluídas, isto pois, a operacionalização do modelo se torna extremamente pesada computacionalmente na medida em que mais variáveis são inseridas na regressão. As variáveis escolhidas são listadas abaixo¹:

¹ Um demonstrativo do banco de dados se encontra no Anexo 2 deste trabalho.

III.2.1) Variáveis do produto:

- Tipo de produto: Cheque pré-datado ou carnê. Indica qual a forma de pagamento escolhida pelo cliente. Usualmente, operações com cheque pré-datado possuem menores índices de inadimplência, logo devemos esperar que o coeficiente da variável dummy da característica cheque pré-datado tenha um valor maior do que o coeficiente da variável dummy da característica carnê. Abaixo podemos ver a diferença de inadimplência entre os dois produtos.

Tabela 5.1: Índices de inadimplência por produto

	Até 30	30-60	60-90	90-120	120-150	150-180	> 180
Carnet	34,7%	17,8%	12,1%	11,3%	10,9%	10,7%	10,5%
Cheque	4,5%	3,3%	2,8%	2,5%	2,6%	2,6%	2,4%

Fonte: Lecca S/A

- Entrada: Com entrada ou sem entrada. Usualmente, clientes que optam por pagamentos com entrada possuem maior capacidade de pagamento. Neste caso, iremos ter uma variável dummy onde o coeficiente deverá ser multiplicado por 1 caso a operação possua entrada e multiplicado por 0 - assumindo pontuação zero para esta característica - caso a operação não tenha entrada.

III.2.2) Variáveis cadastrais:

- Idade: Idade do proponente ao crédito. Nesse ponto é importante frisar que, seguindo as políticas prévias de concessão de crédito adotadas pela maioria das instituições financeiras do mercado brasileiro, clientes com idade inferior a 18 anos e superior a 70 anos não são incluídos na amostra. Sendo que normalmente os clientes mais novos e os mais velhos carregam uma inadimplência maior.

- Telefone residencial: Indica se o cliente possui ou não telefone em sua residência. Tal variável é resultante da checagem de dados realizada após o preenchimento da proposta. Usualmente, clientes que possuem telefone residencial geram índices de inadimplência menores, logo devemos esperar que o coeficiente da variável dummy da característica telefone residencial tenha um valor positivo para que os possuidores de telefone tenham uma pontuação maior - dada pelo coeficiente multiplicado por 1 - neste atributo do que os que não tem - dado pelo coeficiente multiplicado por 0.
- Profissão: Foram analisadas as seguintes categorias: autônomo, profissional liberal, aposentado, assalariado da iniciativa privada, funcionário público, empresários e outros (como pastores, do lar...). Para cada atributo dessa característica será gerado um coeficiente de modo que a variável dummy aposentado de um cliente autônomo assumirá valor nulo em oposição ao coeficiente da variável dummy aposentado que será multiplicado por 1 para que a pontuação dos clientes aposentados seja diferente da pontuação concedida as demais profissões da amostra.

Neste caso esperamos que os aposentados e, principalmente, os autônomos tenham uma capacidade de pagamento menor que das outras classes. No primeiro caso devido ao baixo valor das aposentadorias e do grande índice de compra para terceiros² e no segundo devido a instabilidade no rendimentos auferidos por esta categoria profissional.

III.2.3) Variáveis da capacidade de pagamento:

- Comprometimento de renda: Denota o resultado das prestações do financiamento dividido pela renda mensal do cliente no momento da solicitação do crédito, indicando o percentual de endividamento do cliente. Mais uma vez, seguindo a tendência das

² Este é um dos principais motivos para atrasos, junto com o desemprego e doença na família.

principais financeiras, a financeira que disponibilizou a amostra a ser utilizada limita a concessão de crédito ao percentual máximo de 40% de comprometimento de renda.

Podemos esperar que quanto maior for o comprometimento da renda maior a tendência do cliente atrasar e conseqüentemente de ser um mau cliente.

- SPC (Sistema de Proteção ao Crédito): Indica se o cliente foi negativado ou não nos últimos 6 meses junto a este Bureau de Crédito. Esta variável terá um comportamento semelhante ao descrito para variável entrada somente alterando o sinal do coeficiente, ou seja, se o cliente tiver sido negativado terá seu coeficiente negativo multiplicado por 1, de modo que, sua pontuação será menor do que o cliente não negativado.

III.3) Análise Bivariada:

Após a definição das variáveis que serão incluídas no modelo, o passo seguinte é a realização da chamada Análise Bivariada. Tal procedimento visa agrupar observações tais como cliente com idade 18 anos e cliente com idade 22 anos, uma vez que tais clientes devem apresentar, na média, um comportamento de pagamento muito semelhante, de modo que, se torna mais coerente a utilização de classes de idade do tipo entre 18 e 25 anos.

Abaixo estão listadas as diferentes tabelas resultantes da classificação das variáveis acima descritas. É importante salientar que os dois critérios freqüentemente utilizados para o agrupamento de tais variáveis são a freqüência total de cada atributo de cada variável – já que não faz sentido considerarmos, por exemplo, o atributo aposentado se o mesmo significar apenas 0,001% da amostra – e o percentual de bons pagadores sobre maus pagadores calculado na última coluna da direita, que nos mostra o quão bom ou mau tal atributo é na média – quanto mais alto tal percentual melhor representado

estará este atributo, pois o número de bons clientes empresários, por exemplo, será muito maior do que o número de maus clientes empresários.

• **Produto:**

Classe	Bons		Maus		% Bom/ % Mau
	Frequência	%	Frequência	%	
Cheque	2.391	74,07%	113	51,83%	1,43
Carnê	837	25,93%	105	48,17%	0,54
Total	3.228	100,00%	218	100,00%	

• **Entrada:**

Classe	Bons		Maus		% Bom/ % Mau
	Frequência	%	Frequência	%	
C/ Entrada	2272	70,38%	87	39,91%	1,76
S/ Entrada	956	29,62%	131	60,09%	0,49
Total	3228	100,00%	218	100,00%	

• **Idade:**

Devido a enorme diversidade de observações do atributo idade, foi necessário o desenvolvimento de duas análises bivariadas. Inicialmente, agrupamos as diversas idades com intervalos de 05 anos (tabela a) para que uma melhor observação pudesse ser feita. Posteriormente, adotamos os critérios acima explicados (frequência e percentual de bons sobre maus) para agrupar definitivamente, gerando as classes explicitadas na tabela b.

a)

	Bons			Maus			
	Frequência			Frequência			
Classe	Acumulada	Frequência	%	Acumulada	Frequência	%	% Bom/% Mau
25	414	414	12,83%	44	44	20,18%	0,64
30	726	312	9,67%	70	26	11,93%	0,81
35	1175	449	13,91%	94	24	11,01%	1,26
40	1703	528	16,36%	120	26	11,93%	1,37
45	2035	332	10,29%	137	17	7,80%	1,32
50	2352	317	9,82%	153	16	7,34%	1,34
55	2626	274	8,49%	167	14	6,42%	1,32
60	2965	339	10,50%	184	17	7,80%	1,35
65	3100	135	4,18%	199	15	6,88%	0,61
70	3228	128	3,97%	218	19	8,72%	0,45
Total	-	3228	100%	-	218	100%	

b)

	Bons			Maus			
	Frequência			Frequência			
Classe	Acumulada	Frequência	%	Acumulada	Frequência	%	% Bom/% Mau
30	726	726	22,49%	70	70	32,11%	0,70
60	2965	2239	69,36%	184	114	52,29%	1,33
70	3228	263	8,15%	218	34	15,60%	0,52
Total	-	3228	100,00%	-	218	100,00%	

- **Telefone residencial:**

Classe	Bons		Maus		% Bom/% Mau
	Freqüência	Percentual	Freqüência	Percentual	
Tem telefone	2172	67,29%	90	41,28%	1,63
Não tem telefone	1056	32,71%	128	58,72%	0,56
Total	3228	100,00%	218	100,00%	

- **Profissão:**

Inicialmente, analisamos todas as profissões do banco de dados (tabela a) para depois, seguindo os critérios usuais, agruparmos os atributos semelhantes (tabela b)

a)

Classe	Bons		Maus		% Bom/% Mau
	Freqüência	Percentual	Freqüência	Percentual	
Autônomo	360	11,15%	60	27,52%	0,41
Aposentado	240	7,43%	21	9,63%	0,77
Ass. Privado	1.245	38,57%	70	32,11%	1,20
Ass. Público	584	18,09%	31	14,22%	1,27
Empresário	268	8,30%	12	5,50%	1,51
Prof. Liberal	404	12,52%	18	8,26%	1,52
Outros	127	3,93%	6	2,75%	1,43
Total	3.228	100,00%	218	100,00%	

b)

Classe	Bons		Maus		% Bom/% Mau
	Frequência	Percentual	Frequência	Percentual	
Autônomo	360	11.15%	60	27.52%	0.41
Aposentado	240	7.43%	21	9.63%	0.77
Assalariado	1,829	56.66%	101	46.33%	1.22
Outros	799	24.75%	36	16.51%	1.50
Total	3,228	100.00%	218	100.00%	

• **Comprometimento de renda:**

De acordo com o que foi explicado durante o item escolha das variáveis, tal índice será agrupado levando em conta parâmetros genéricos utilizados pelas instituições financeiras brasileiras. Na tabela a, agrupamos os percentuais com intervalos de 05 pontos percentuais. Na tabela b, adotamos os critérios acima explicados (frequência e percentual de bons sobre maus) para agrupar definitivamente as classes dos atributos de tal variável.

a)

Classe	Bons		Maus		% Bom/% Mau
	Frequência	%	Frequência	%	
10	295	9,14%	10	4,59%	1,99
15	841	26,05%	48	22,02%	1,18
20	972	30,11%	55	25,23%	1,19
25	785	24,32%	70	32,11%	0,76

30	243	7,53%	26	11,93%	0,63
35	64	1,98%	6	2,75%	0,72
40	28	0,87%	3	1,38%	0,63
Total	3.228	100,00%	218	100,00%	

b)

	Bons		Maus		
Classe	Frequência	%	Frequência	%	% Bom/% Mau
10	295	9,14%	10	4,59%	1,99
20	1.813	56,16%	103	47,25%	1,19
40	1.120	34,70%	105	48,17%	0,72
	3.228	100,00%	218	100,00%	

• SPC (Sistema de Proteção ao Crédito):

	Bons		Maus		
Classe	Frequência	Percentual	Frequência	Percentual	% Bom/% Mau
Negativado SPC	239	7.40%	71	32.57%	0.23
Não negativado SPC	2989	92.60%	147	67.43%	1.37
	3228	100.00%	218	100.00%	

III.4) Regressão Original:

Abaixo temos uma demonstração extremamente simplificada de qual seria o resultado final de um modelo de Credit Scoring. O cliente ao preencher o cadastro irá fornecer uma série de informações (no exemplo: X1, X2 e X3) que serão inseridas no

modelo previamente desenvolvido (que no exemplo irá definir os pesos B1, B2 e B3) de modo a gerar uma pontuação para tal cliente. B0 representa o coeficiente linear da equação - que costuma ser positivo para que nenhum cliente seja considerado mau pagador sem sofrer a devida análise anteriormente - e U o erro aleatório estatístico da equação. Tal pontuação será comparada com a pontuação mínima definida pelo modelo em conjunto com as diretrizes estratégicas traçadas pela empresa, de modo a definir se o cliente terá seu crédito aprovado ou não.

$$\text{SCORE DO CLIENTE} = B0 + B1*X1 + B2*X2 + B3*X3 + U$$

Após a definição das classes das variáveis dada pela análise bivariada descrita acima, o próximo passo é rodar a regressão linear pelo método dos mínimos quadrados ordinários em relação as variáveis dummy especificadas abaixo:

- Dummy cheque
se cheque (CH) = 1
se não (CA) = 0

- Dummy entrada
se tem entrada (ENT) = 1
se não (ENT) = 0

- Dummy idade 1
se idade até 30 (IDA30) = 1
se não (IDA30) = 0

- Dummy idade 2
se idade entre 30 e 60 (IDA60) = 1
se não (IDA60) = 0

	Dummy idade 3	se idade entre 60 e 70 (IDA70) = 1 se não (IDA70) = 0
→	Dummy telefone residencial	se tem telefone (TEL) = 1 se não (TEL) = 0
→	Dummy profissão 1	se autônomo (AUT) = 1 se não (AUT) = 0
	Dummy profissão 2	se aposentado (APO) = 1 se não (APO) = 0
	Dummy profissão 3	se assalariado (ASS) = 1 se não (ASS) = 0
	Dummy profissão 4	se outros (OUT) = 1 se não (OUT) = 0
→	Dummy % renda 1	se % até 10% (REN10) = 1 se não (REN10) = 0
	Dummy % renda 2	se % entre 10 e 20 (REN20) = 1 se não (REN20) = 0
	Dummy % renda 3	se % entre 20 e 40 (REN40) = 1 se não (REN40) = 0
→	Dummy SPC	se negativado (SPC) = 1

se não (SPC) = 0

O modelo usado para rodar a regressão foi o de Regressão Linear Multivariada, onde os coeficientes das variáveis dummy descritas acima foram estimados pelo Método dos Mínimos Quadrados Ordinários. A regressão inicial foi feita com todas as variáveis. O resultado abaixo foi obtido através do software Eviews.

LS // Dependent Variable is SCORE

Date: 06/15/99 Time: 22:37

Sample: 1 3446

Included observations: 3446

SCORE=C(1)+C(2)*APO+C(3)*ASS+C(4)*AUT+C(5)*CH+C(6)*CA+C(7)*ENT+C(8)*IDA30+C(9)*IDA60+C(10)*REN10+C(11)*REN20+C(12)*REN40+C(13)*SPC+C(14)*TEL+C(15)*OUT+C(16)*IDA70

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	1.992714	0.396729	5.022864	0.0000
C(2)	-1.054286	0.229309	-4.597670	0.0000
C(3)	-1.023566	0.228832	-4.473006	0.0000
C(4)	-1.082036	0.229200	-4.720929	0.0000
C(5)	0.714039	0.229580	3.110200	0.0019
C(6)	0.652620	0.229655	2.841748	0.0045

C(7)	0.066415	0.008696	7.637113	0.0000
C(8)	-0.819059	0.229596	-3.567398	0.0004
C(9)	-0.803977	0.229457	-3.503820	0.0005
C(10)	0.045222	0.017722	2.551781	0.0108
C(11)	0.029467	0.013025	2.262390	0.0237
C(12)	0.007631	0.014294	0.533853	0.5935
C(13)	-0.168028	0.013659	-12.30124	0.0000
C(14)	0.059997	0.008262	7.262005	0.0000
C(15)	-1.020008	0.228750	-4.459043	0.0000
C(16)	-0.861117	0.229191	-3.757202	0.0002

R-squared	0.123111	Mean dependent var	0.936738
Adjusted R-squared	0.119276	S.D. dependent var	0.243468
S.E. of regression	0.228488	Akaike info criterion	-2.947915
Sum squared resid	179.0685	Schwarz criterion	-2.919383
Log likelihood	205.5949	F-statistic	32.10374
Durbin-Watson stat	0.163914	Prob(F-statistic)	0.000000

III.5) Analisando o modelo:

Devemos observar que todas as variáveis se comportaram da maneira esperada, ou seja, que o resultado da regressão está de acordo com a teoria.

Porém o resultado acima especificado aponta para a necessidade de testes sobre a significância dos coeficientes obtidos e sobre a existência ou não de correlação entre as

variáveis para que as que estejam viesando o modelo possam ser identificadas e expurgadas do mesmo, de modo a gerar um modelo mais corretamente especificado.

III.5.1) Teste de significância:

O primeiro teste aplicado foi o de significância dos coeficientes estimados pelo modelo. Através do teste de hipótese abaixo descrito, e da sua regra de decisão correspondente, pudemos verificar quais coeficientes não se apresentaram estatisticamente significante.

Teste de Hipótese)

H0: Coeficiente = 0

H1: Coeficiente \neq 0

Regra de decisão)

Se $|t\text{-value}| > t\text{-crítico}$ → Então, rejeito HO ao nível de significância de 5%.

Se $|t\text{-value}| < t\text{-crítico}$ → Então, não rejeito HO ao nível de significância de 5%.

Temos que, usando $t\text{-crítico} = 1,96$ – dado pelo nível de significância de 5%, amostra = 3446 e grau de liberdade = 15. O único coeficiente que teve a hipótese nula não rejeitada foi C(12), dado que:

H0: C(12)=0

H1: C(12) \neq 0

$t_{\text{obs}} = 0,533$

$t_{2,5\%, 3.430} = 1,960$

$|t_{2,5\%, 3.430}| > |t_{\text{obs}}|$

Logo, não se rejeita a hipótese nula.

Isto denota que a variável dummy representativa da classe de 20% até 40% da variável comprometimento de renda pode ser excluída do modelo. Tal conclusão é reforçada pelo alto valor encontrado para o p-value,

III.5.2) Teste de correlação:

O teste de correlação das variáveis abaixo descrito foi obtido através do software estatístico SPSS, que possibilita a obtenção de testes de Spearman não paramétricos extremamente adequados para a análise de correlação entre variáveis dummy.

Tal teste nos informa, em sua primeira linha, o coeficiente de correlação entre a combinação de variáveis referente e em sua segunda linha uma medida da significância de tal correlação. Por exemplo, o primeiro registro indica os resultados das variáveis Aposentado e Assalariado, indicando uma correlação negativa de 0,3230 com uma significância correspondente de aproximadamente 0.

Coeficientes de Correlação de Spearman

Parte 1)

	APO	ASS	AT__10	AUTO	CA	CHE
ASS	-,3230					
	,000					
AT__10	-,0275	-,0120				
	,107	,482				

AUT.	-,1066 ,000	-,4203 ,000	-,0099 ,560			
CA	,0262 ,124	-,0178 ,296	,0192 ,260	,0322 ,058		
CHE	-,0262 ,124	,0178 ,296	,0192 ,260	-,0322 ,058	-1,0000 ,000	
ENT	,0971 ,000	,0077 ,650	,0197 ,248	-,2023 ,000	,0005 ,977	-,0005 ,978
18-30	-,1309 ,000	,0016 ,924	,0135 ,430	,1726 ,000	,0160 ,346	-,0161 ,345
30-60	,0372 ,029	,0216 ,206	-,0028 ,870	-,1369 ,000	-,0185 ,278	,0185 ,278
60-70	,1348 ,000	-,0382 ,025	-,0156 ,360	-,0323 ,058	,0065 ,702	-,0065 ,702
OUT	-,1619 ,000	-,6381 ,000	,0384 ,024	-,2107 ,000	-,0201 ,237	,0201 ,237

%10-20	,0394	-,0072	-,3487	-,0045	,0042	-,0043
	,021	,674	,000	,791	,803	,802
%20-40	-,0019	-,0066	-,1790	-,0046	,0185	-,0185
	,909	,699	,000	,789	,277	,277
SPC	-,0057	-,0299	-,0087	,0255	,0188	-,0188
	,740	,079	,610	,135	,270	,269
STATUS	-,0202	,0506	,0391	-,1218	-,1214	,1214
	,236	,003	,022	,000	,000	,000
TEL	,0247	,0112	-,0026	-,0779	-,0005	,0005
	,147	,511	,880	,000	,977	,978

Parte 2)

ENT	18-30	30-60	60-70	OUT	%10-20
18-30	-,1789				
	,000				
30-60	,1386	-,8041			
	,000	,000			

60-70	,0388	-,1683	-,4506			
	,023	,000	,000			
OUT	,0856	-,0528	,0565	-,0144		
	,000	,002	,001	,398		
%10-20	,0059	-,0105	-,0091	,0309	-,0126	
	,728	,538	,593	,070	,459	
\$20-40	-,0059	-,0183	,0436	-,0447	,0123	-,6428
	,730	,283	,011	,009	,471	,000
SPC	-,0316	,0130	-,0298	,0299	,0187	-,0191
	,063	,446	,080	,079	,273	,262
STATUS	,1525	-,0555	,0893	-,0646	,0468	,0437
	,000	,001	,000	,000	,006	,010
TEL	,0284	-,0225	,0085	,0197	,0312	,0717
	,096	,186	,618	,248	,067	,000

Parte 3)

%20-40	SPC	STATUS	
SPC	,0002		
	,990		
STATUS	-,0439	-,2140	
	,010	,000	
TEL	-,0838	-,0245	,1333
	,000	,150	,000

A partir dos coeficientes acima dispostos é possível listar os diferentes atributos das diferentes variáveis que possuem alta correlação entre si, onde é considerado como alta correlação valores acima de 0,50 (50%) - tais casos estão em negrito.

Conforme pode ser observado, somente atributos de um mesmo tipo de variável possuem alta correlação, de modo que, não devemos excluir nenhuma variável do modelo devido a correlação.

III.6) Regressão final:

Concluimos, a partir dos testes feitos anteriormente, que seria interessante excluirmos do modelo a variável comprometimento de renda entre 20% e 40%. O resultado final obtido após tal exclusão se encontra abaixo.

LS // Dependent Variable is PGTO

Date: 06/15/99 Time: 22:39

Sample: 1 3446

Included observations: 3446

$$\begin{aligned} \text{PGTO} = & C(1) + C(2)*\text{APO} + C(3)*\text{ASS} + C(4)*\text{AUT} + C(5)*\text{CH} + C(6)*\text{CA} + C(7)*\text{ENT} \\ & + C(8)*\text{IDA30} + C(9)*\text{IDA60} + C(10)*\text{REN10} + C(11)*\text{REN20} + C(13)*\text{SPC} \\ & + C(14)*\text{TEL} + C(15)*\text{OUT} + C(16)*\text{IDA70} \end{aligned}$$

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	1.991251	0.396678	5.019818	0.0000
C(2)	-1.053965	0.229284	-4.596765	0.0000
C(3)	-1.023507	0.228808	-4.473215	0.0000
C(4)	-1.082032	0.229176	-4.721404	0.0000
C(5)	0.718733	0.229388	3.133267	0.0017
C(6)	0.657414	0.229455	2.865109	0.0042
C(7)	0.066413	0.008695	7.637701	0.0000
C(8)	-0.817056	0.229541	-3.559521	0.0004
C(9)	-0.801804	0.229397	-3.495263	0.0005
C(10)	0.039896	0.014645	2.724178	0.0065
C(11)	0.024151	0.008395	2.876831	0.0040
C(13)	-0.168231	0.013653	-12.32219	0.0000
C(14)	0.059796	0.008252	7.245966	0.0000
C(15)	-1.019781	0.228726	-4.458524	0.0000
C(16)	-0.859309	0.229142	-3.750112	0.0002

R-squared	0.123038	Mean dependent var	0.936738
Adjusted R-squared	0.119460	S.D. dependent var	0.243468
S.E. of regression	0.228464	Akaike info criterion	-.948412
Sum squared resid	179.0834	Schwarz criterion	-.921664
Log likelihood	205.4517	F-statistic	34.38368
Durbin-Watson stat	0.163841	Prob(F-statistic)	0.000000

IV) Testes de performance:

Após a definição do modelo de regressão final, testaremos a validade de nosso modelo através do teste Kolmogorov-Smirnov. Tal teste é o mais indicado para a validação de modelos de Credit Scoring, uma vez que, o objetivo central do desenvolvimento de tais modelos é desenvolver um método capaz de separar a população de bons clientes da população de maus clientes para que a tomada de decisão quanto a aprovação ou não do crédito seja mais eficiente e coerente com os objetivos de market-share e rentabilidade traçados pela empresa.

Para a realização deste teste foi utilizado o software SPSS, que apresenta como resultado mais relevante a maior diferença absoluta encontrada entre a distribuição de frequência dos contratos bons e maus por valor de score encontrado.

Teste de Kolmogorov - Smirnov

Dados analisados:

Amostra 1) Status Mau = 0

Amostra 2) Status Bom = 1

Agrupado por Score

Ocorrências:

Amostra 1 → 218

Amostra 2 → 3228

Total → 3446

Resultados – Maiores diferenças extremas:

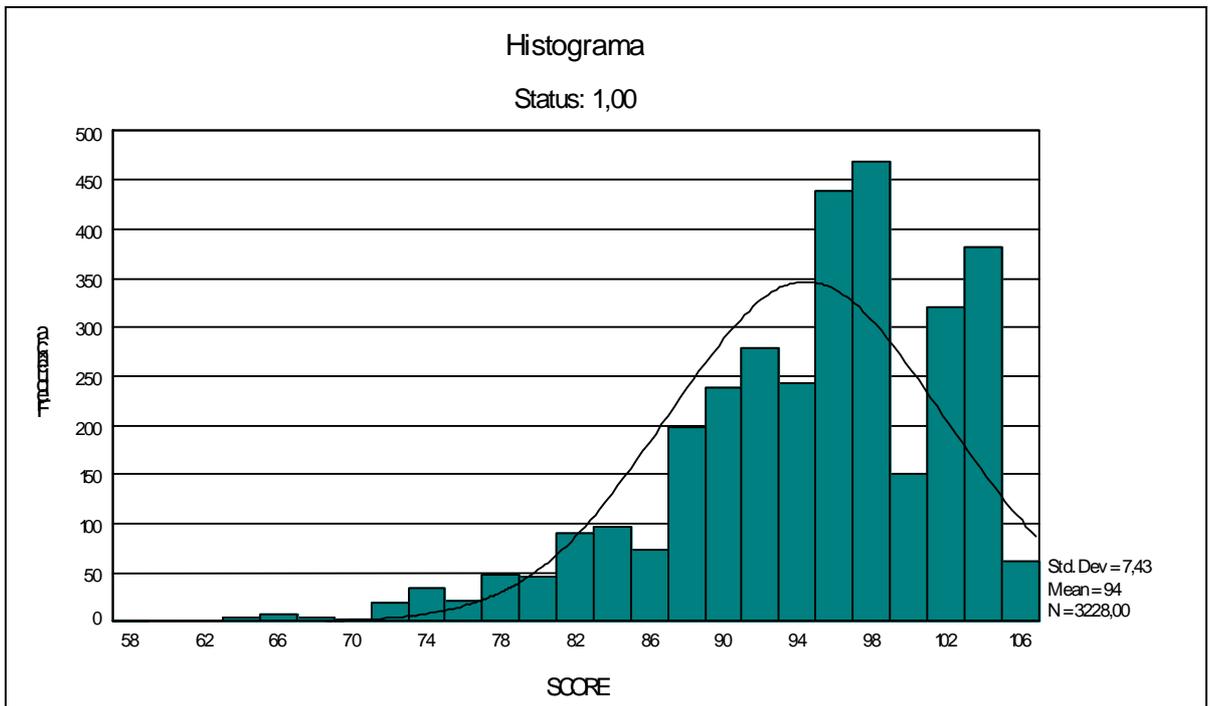
Absoluta	Positiva	Negativa	z-value	p-value
0,50782	0,00000	-0,50782	7,2569	0,0000

Assim sendo, o teste acima descrito apresenta a maior diferença absoluta entre a distribuição das ocorrências (número de contratos) de bons e maus clientes por faixa de score de 50,782%, ou seja, o K-S de nosso modelo atingiu o nível de 50,782%, o que pode ser considerado como um resultado muito bom, já que o z-value para tal coeficiente se mostrou significativo estatisticamente.

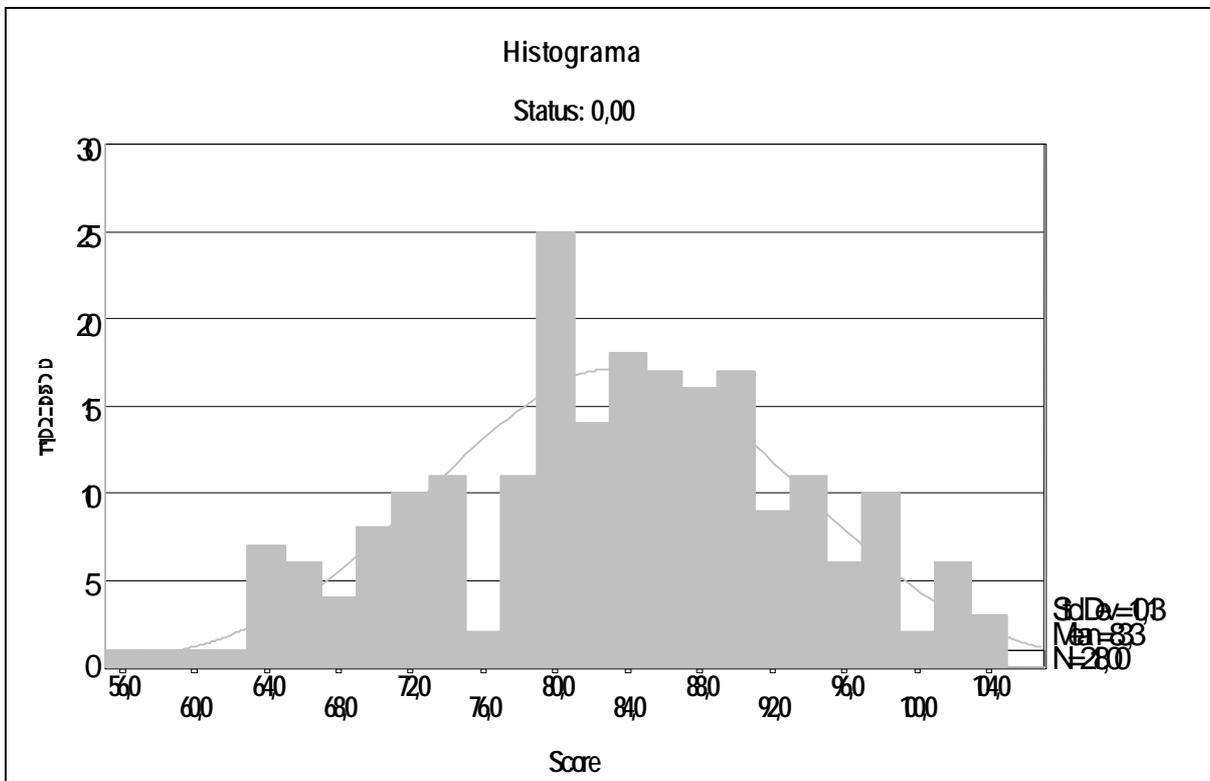
Logo, podemos considerar satisfatório o resultado obtido pelo modelo, uma vez que, apesar do valor encontrado pelo R2 ter sido baixo – devido ao pequeno número de variáveis incluídas no modelo – conseguimos separar consideravelmente as populações de bons e maus clientes. Devemos esperar ainda, que a inclusão de novas variáveis seria interessante do ponto de vista explicativo do modelo – tornando-o entretanto pesado computacionalmente – mas não fundamental do ponto de vista de resultados práticos do modelo.

Abaixo apresentamos os gráficos de distribuição de frequência dos clientes bons e maus por faixa de score.

Cientes bons:



Cientes maus:



V) Ponto de corte

O concedente de crédito é quem determina o nível de risco que quer assumir e não o credit scoring, este somente gera uma pontuação (score) para o cliente. Tal pontuação é confrontada com o ponto de corte estabelecido pelo concedente. Existem três critérios para esta definição:

- a) Manter a taxa de aprovação reduzindo a perda;
- b) Manter a taxa de perda aumentado a aprovação e
- c) Maximização dos lucros.

A escolha entre estes deverá se fazer conforme o planejamento estratégico da empresa, usaremos por ser o mais comum a otimização da rentabilidade.

Tal ponto de corte é estabelecido a partir da análise de rentabilidade do produto em questão, onde é medido quantas contas com maus comportamentos de pagamento são compensadas pela aquisição de contas com bons comportamentos de pagamento. Se o score estiver abaixo deste ponto de corte, o crédito não deverá, em princípio, ser concedido e se o score estiver acima deste ponto de corte, a concessão será recomendada.

A seguir apresentamos um modelo de análise de rentabilidade para o produto em questão, com os dados aproximados da financeira analisada.³

³ No anexo 1 iremos explicar os conceitos aqui usados

Condições da Operação :

Variáveis	Maus	Bons
Volume de Financiamento :	500,00	500,00
Plano Médio :	6	4,7
Contrato Médio :	500,00	500,00
TAC :	9,00	2,00
Taxa de Financiamento :	9,70%	7,15%
Taxa de Juros de Mora :	8,00%	8,00%
IOF :	0,0041%	0,0041%
IOF (fixo):	0,3800%	0,3800%
Cofins:	3,0000%	3,0000%
PIS:	0,6500%	0,6500%
Multa :	2,00%	2,00%
Taxa de Captação :	1,70%	1,70%
Pz de Pagto ao Lojista (dias) :	3	3
Custo Variável :	25,00	25,00
<u>Índices de Cobrança</u>		
Índice de Perda :	50,00%	0,00%

Resultado da Operação

	Maus	Bons
--	-------------	-------------

Financiamento :	500	500
------------------------	------------	------------

Receitas

Líq. De Financto :	147	79
TAC :	9	2
Float :	1	1
Juro de Mora e Multa :	-	-
Custo do Atraso :	-	-
Total =>	156	81

Despesas

IOF :	2	1
IOF (fixo):	2	2
Confins:	6	3
PIS:	1	1
Perda :	328	-
Custo Variável :	25	25
Total =>	363	32

Valor Bruto Resultado =>	(229)	53
Valor Líquido Resultado =>	(207)	49

Pela análise de rentabilidade, temos que a contribuição bruta dos clientes bons é igual a R\$ 53,00, enquanto que a perda bruta atribuída aos maus clientes é dada por R\$ 229,00. Desta maneira podemos observar que precisamos de **4,32** clientes bons para cobrir a perda gerada pelos maus.

A partir de tal proporção, podemos aplicar o conceito de taxa de sinistro de equilíbrio, usualmente utilizado na definição do ponto de corte de equilíbrio do modelo, dada pela relação abaixo, onde a unidade do numerador representa 1 cliente mau e o denominador representa a quantidade de bons para empatar um mau mais 1 cliente mau – representando o universo total.

$$\text{Taxa de Sinistro \%} = \frac{1}{(4,32 + 1)} \times 100 = 18,80 \%$$

Essa é a taxa de sinistro no equilíbrio do modelo corresponde a classe de score a partir de onde o produto começa a ser rentável. Conforme a tabela em anexo⁴ demonstra, a taxa de equilíbrio descrita acima corresponde ao ponto onde a determinação do score correspondente gera equilíbrio entre ganho e perda, ou seja, a partir deste valor para o score o produto é rentável e abaixo de tal valor o produto não é rentável.

Desta forma no nosso modelo o cliente com score acima de 81 pontos seria considerado bom cliente e conseqüentemente teria seu crédito aprovado. Podemos tirar outras conclusões da análise da tabela, como a taxa de rejeição de bons clientes que

⁴ Ver Anexo 3.

ficaria em 5,14%, a taxa de aprovação de maus clientes que ficaria em 65,13% e taxa de rejeição global que fica 7,02%.

Temos aqui o impressionante percentual de clientes maus aprovados, isto se deve basicamente a alta taxa de juros. Que permite uma alta margem e conseqüentemente um alto índice de perda sem que o produto de prejuízo. Poderíamos, ainda, tentar aumentar a rentabilidade deixando a classe com score entre 81 e 85 para ser analisada pelos autorizadores de crédito, para com isso reduzir o índice de 22,02% de maus clientes aprovados com este score.

VI – Utilização e características de um Credit Scoring

Usualmente é dito que a análise de crédito é uma arte e não uma ciência, o que com a crescente evolução dos scores está cada vez mais difícil de sustentar. Pois atualmente é impossível negar as vantagens de um método autorial sobre um subjetivo. Podemos dizer que as grandes questões atualmente são: qual o meio mais eficaz para desenvolver um modelo e como usá-lo no dia-a-dia.

VI.1) Comparação das políticas

O resultado da scoragem transforma completamente a política de crédito de uma instituição. Primeiro porque passam a ser considerados itens como faixa de idade ou até a época do ano em que o cliente está comprando, o que provavelmente nunca fizeram parte formalmente de uma política tradicional. Em segundo lugar podemos colocar a “frieza” e a objetividade apesar de ser um dos objetivos da scoragem muda radicalmente não só a política como a maneira de conceder o financiamento.

As decisões de concessão de crédito baseadas na experiência dos gerentes estão sujeitas a inconsistências causadas pelo estado de espírito desses profissionais. Por mais competentes que sejam, eles são seres humanos sujeitos a erros e a diferentes emoções ao longo das suas carreiras, de modo que, podemos esperar que um gerente em estado de extrema alegria (por exemplo: por estar próximo do nascimento de seu primeiro filho) tenderá a ser menos restritivo na concessão de crédito do que um outro gerente amargurado e infeliz (por exemplo: devido a descoberta de uma doença grave em sua família). Logo, as decisões num ambiente de credit scoring são imparciais, o que aumenta a eficiência.

Muitas empresas adequam seus sistemas de credit scoring de tal modo que, em casos marginais, ou seja, em casos onde os proponentes que não têm o score alto

suficiente para passar facilmente, nem é baixo o suficiente para serem reprovados, as propostas são enviadas para o gerente de crédito responsável que decide se a companhia irá estender ou não tal crédito. Isto possibilita que o volume de decisões subjetivas seja consideravelmente menor, de modo que, o gerente da linha terá maior eficiência na tomada desse tipo de decisão e maior tempo livre para viabilizar outros projetos relevantes.

Diferente da análise tradicional onde o proponente ao crédito têm que preencher todos os requisitos básicos da política de crédito. Como por exemplo: tempo mínimo de emprego e residência, quantidade máxima de informações solicitadas por outros concedentes, limite máximo de prestação em relação a renda entre outros. Com a utilização de um Credit Scoring não há regras rígidas para aprovação quanto a certas normas e itens.

Na escoragem o cliente pode ter a sua informação solicitada por diversas instituições e perder pontos por causa disto mas nunca terá seu crédito negado por este motivo. Pois poderá ganhar pontos suficientes em outros quesitos e desta forma ter alcançado o número de pontos suficientes para ter seu crédito aprovado.

VI.2) A Necessidade do Score

A questão da necessidade do score cabe somente em mercados ainda pequenos e pouco desenvolvidos onde temos o dilema de custo / benefício mau resolvido, pois em empresas de grande porte já está comprovado o uso do score como uma poderosa ferramenta na concessão do crédito.

Para firmas de pequeno porte a implementação de um modelo complexo pode ser “mortal”, ou devido ao alto custo de implementação e manutenção ou a conclusões tiradas erradamente da análise estatística feita sem o cuidado necessário.

O desenvolvimento de um sistema de scorage de crédito pode levar dois anos, necessitar de US\$ 25.000 a US\$ 250.000 em despesas, e envolver os esforços de várias centenas de pessoas, segundo informações obtidas de especialista na área.

A solução para este tipo de problema seria o uso de Credit Score genéricos ou ainda a scores desenvolvidos por credit bureaus. O problemas aqui é que diferentemente dos EUA não existem ainda tais tipos de solução no Brasil.

No crédito ao consumidor indireto, ou seja através de um revendedor ou lojista pode ser difícil a implementação de um score devido ao fato de termos neste produto uma terceira parte envolvida no financiamento que está interessada na venda, e que pode fazer pressão para que haja reconsiderações demais, inviabilizando o produto na prática. Neste caso uma solução seria mudar o método do ponto de corte para manter a taxa de aprovação do produto.

VII) Extensões do modelo

Como já colocamos, atualmente há várias utilizações dada para a técnica desenvolvida a partir da necessidade de maior segurança e velocidade da aprovação do crédito.

Behavioral credit scoring

Além de serem úteis na redução dos custos e na elevação da receita proveniente das decisões de concessão de crédito para novos clientes, os modelos de credit scoring também podem ser utilizados para monitorar contas já existentes. Por exemplo, o credit scoring de uma conta que tem limite de crédito baixo pode ajudar na decisão quanto o aumento ou não deste limite, e até no tamanho deste aumento.

Behavioural collect scoring

Tal aplicação pode melhorar o fluxo de caixa da empresa, ao passo que melhora o processo de cobrança. Por exemplo, uma conta em atraso com uma pontuação alta pode sugerir que o cliente geralmente em dia está passando por um período de dificuldades. A inadimplência nestes casos tem maiores chances de ser recuperada com um contato direto. Por outro lado, scores baixos podem sugerir problemas crônicos. Nestes casos, pode-se decidir suspender o crédito, ou enviar o contrato para as assessorias de cobrança mais cedo que nos casos usuais, de modo que, a probabilidade de recuperação será maior. Ou ainda, o modelo pode indicar aqueles clientes que sempre atrasam e que acabam pagando, e que desta forma não precisa receber atenção imediata.

Profitability credit scoring

Neste caso para decidirmos se o cliente é bom ou mau ao contrário de usarmos a quantidade de dias de atraso, usamos se o cliente deu lucro ou não para instituição. Desta forma na hora da aprovação não deixaríamos de aprovar clientes que têm como característica atrasar e pagar. É mais importante deixaríamos de ter um critério subjetivo para definição desta questão.

Cross-selling

Um outro uso do credit scoring está na utilização do banco de dados da própria empresa para avaliação da possibilidade de oferta de um produto a clientes de outros produto. Ou seja, analisar estatisticamente quais clientes têm tendência aceitar ou comprar outros produtos da empresa.

Pré-screening

Através da compra de Bancos de Dados de outras empresas podemos chegar a uma definição sobre a escolha dos clientes que se adequam ao perfil procurado pela empresa para a oferta de um determinado produto. A partir da análise do Credit Scoring buscamos a prospecção de tais clientes através de mala direta ou de outras formas de aquisição de market-share.

VIII- Conclusão

No negócio de crédito consumidor a relação risco / recompensa é o que determina as políticas a serem usadas. Se uma alta perda está incluída na taxa não haverá problemas, pois o produto será rentável. No Brasil, como o mercado ainda é pouco competitivo e desenvolvido, as altas taxas encobrem os altos índices de perda e também altos custos operacionais.

Este quadro tende a ser alterado com o aumento de concorrência, o primeiro passo neste sentido é a decisão de adquirir novas tecnologias, como o credit scoring. Porém o mais difícil é adequar as ferramentas às diferentes características do mercado brasileiro para conseguir extrair todo o potencial destas. Num exemplo clássico podemos citar o uso do cheque pré-datado como forma de pagamento a financiamentos, o que se não é pioneiro no mundo e não é comum.

Estes aprimoramentos são lentos e custosos, mas é o que irá ditar o futuro da indústria. E todos grandes *players* já perceberam isto e estão se mexendo nesta direção. As características da operação que exigem uma escala muito grande para diluir o custo e também o apetite dos bancos estrangeiros, irá acabar a curto prazo com os pequenos bancos e financeiras que atuam nesta área.

Isto tudo só é válido se continuarmos com o processo de estabilização bem sucedido, que permite o desenvolvimento do setor e encoraja os investimentos a longo prazo pelos investidores.

É preciso esclarecer que uma vez feito o modelo só completamos uma primeira etapa. Temos agora que monitorar os seus resultados constantemente para avaliar o desempenho, além de continuar a procura por novos métodos e tecnologias que possam aperfeiçoar o nosso score. Outro ponto é o uso da técnica desenvolvida para fazer testes

controlados, nos permitindo ter uma clara visão do resultado de diferentes possíveis estratégias a serem adotadas num determinado segmento ou para testar novos produtos.

Neste momento a competitividade da empresa passa a se basear no bom desenvolvimento do modelo, com a quantidade e escolha adequada de variáveis, a criatividade no uso, ter um bom sistema de repescagem....

Outra necessidade que fica latente ao estudarmos o assunto é a pouca ou nenhuma regulamentação governamental sobre os assuntos, o que por um lado possibilita as empresas de agirem da forma com que mais lhe convier, mas que por outro dificulta a obtenção de dados para avaliação do crédito.

Quanto ao desenvolvimento do modelo, observamos que as ferramentas estatísticas utilizadas não são muito complexas, porém isto não garante uma boa scoragem. Isto se deve a vários fatores, como:

- A amostra pode não ser representativa da população ou a população futura não ser parecida com a passada;
- O ponto de corte ser definido de forma errada;
- Escolha de variáveis pouco significativas;
- Banco de dados “pobre”, sem as informações importantes e
- Até a má implantação do modelo.

Os resultados do nosso modelo apesar da boa separação entre bom e mau cliente, nunca poderia ser utilizado devido as simplificações assumidas, como a não inclusão dos rejeitados e extração da amostra sem o uso de ferramentas estatísticas adequadas.

Teríamos também que estudar melhor alguns pontos como a distribuição do erro e de sua variância e as consequências disto na previsibilidade do modelo. Contudo o

resultado alcançado está de acordo com o objetivo do trabalho que era apresentar o Credit Scoring e mostrar como desenvolvemos um modelo.

IX) Anexos

Conceitos usados no modelo de rentabilidade:

Volume de financiamento => Valor total a ser emprestado (no caso igual ao valor do contrato médio)

Plano médio => Quantidade média de prestações

Tac => Taxa de Abertura de Crédito

Taxa de financiamento => Taxa de juros ao mês média, já com IOF incluso

Taxa de juros de mora => Taxa de juros ao mês paga quanto o cliente atrasa

IOF => Imposto sobre Operações Financeiras, alíquota ao dia vigente na época

IOF(fixo) => Percentual fixo do referido imposto

Cofins => Percentual de Cofins incidente na receita da operação

PIS => Percentual de PIS incidente na receita da operação

Multa => Percentual da multa contratual paga pelo cliente quanto atrasa

Taxa de captação => Taxa de juros ao mês de captação dos recursos utilizados

Pz de pagto ao lojista (dias) => Prazo entre o recebimento do contrato e o efetivo pagamento ao lojista

Custo variável => Custo operacional de cada contrato, com todos os custos incluídos

Índice de perda => Percentual do total a receber que vai ser jogado como perda

Líq. De financiamento => Receita de juros menos a despesa de captação

Tac => Receita total de Tac

Float => Receita gerada pela remuneração do capital até o pagamento efetivo do lojista

Custo do atraso => Custo de captação por carregar o atraso

IOF, IOF (fixo), CPMF, Cofins e Pis => Custo com os referidos impostos

Perda => Valor total jogado para perda

Custo variável => Total do custo variável

Valor bruto resultado => Valor futuro do resultado

Valor líquido resultado => Valor presente do resultado, descontado a taxa de captação

X) Bibliografia

Gujarat, Damodar N.: Basic Econometric, United States Military Academy, West Point, 1995.

Lewis, Edward M.: An Introduction to Credit Scoring, Fair, Isaac and Co., Inc., San Rafael, 1996.

Johnston, Jack e Dinardo Jack: Econometric Methods, McGraw-Hill, 1997.

Blatt, Adriano: Credit Scoring, Suma Econômica, Rio de Janeiro, 1998.

Mays, Elizabeth; Credit Risk Modeling, Design and Application, Amacom, 1998.

Boyes, William J., Hoffman, Dennis L. and Low, Stuart A.: Na Econometric Analysis of the Bank Credit Scoring Problem, Journal of Econometric 40, 1989.

Lawrence, David B.: O Negócio de Crédito ao Consumidor, Risco & Recompensa, Citicorp.

Material da Conferência de Cartões de Crédito Populares, São Paulo, 1997.

Páginas Internet:

www.crma.com

www.fico.com

Entrevista Osvaldo Antunes Maciel, ex-presidente da FININVEST