

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO
DEPARTAMENTO DE ECONOMIA**

MONOGRAFIA DE FINAL DE CURSO

MODELO DE ESCORAGEM DE CRÉDITO RENTABILIZADO

**Gabriel José Gama Ferreira
No. de matrícula 95015234-9-2**

Orientador: Paulo Levi

Dezembro de 1999

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO
DEPARTAMENTO DE ECONOMIA**

MONOGRAFIA DE FINAL DE CURSO

MODELO DE ESCORAGEM DE CRÉDITO RENTABILIZADO

**Gabriel José Gama Ferreira
No. de matrícula 95015234-9-2**

Orientador: Paulo Levi

Dezembro de 1999

“Declaro que o presente trabalho é de minha autoria e que não recorri para realizá-lo, a nenhuma forma de ajuda externa, exceto quando autorizado pelo professor tutor.”

“As opiniões expressas neste trabalho são de responsabilidade única e exclusiva do autor.”

ÍNDICE

INTRODUÇÃO

I - O MERCADO DE CONCESSÃO DE CRÉDITO BRASILEIRO

II - ANÁLISE E CONCESSÃO DE CRÉDITO

III - CREDIT SCORING

III.1 - BASE DE DADOS

III.2 - DEFINIÇÃO DE BONS E MAUS

III.3 - DEFINIÇÃO DAS VARIÁVEIS DO MODELO

III.4 - ANÁLISE BIVARIADA

III.5 - REGRESSÃO LINEAR

III.6 - TESTE DE PERFORMANCE - KOMOGOROV-SMIRNOV (KS)

IV - PONTO DE CORTE - INTRODUZINDO A RENTABILIDADE DO PRODUTO NO MODELO

V - ADENDO AO MODELO - CONSIDERAÇÕES MACRO-ECONÔMICAS

VI - CONCLUSÕES

ANEXO I - ANÁLISE BIVARIADA

ANEXO II - REGRESSÃO LINEAR E TESTES DE SIGNIFICÂNCIA E CORRELAÇÃO

ANEXO III - RENTABILIDADE DO PRODUTO

VII - BIBLIOGRAFIA

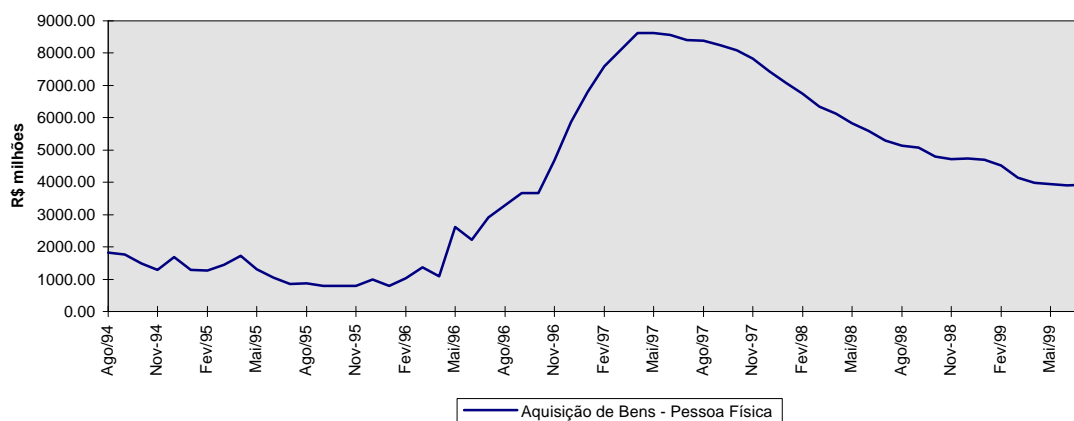
INTRODUÇÃO:

Esta monografia tem como objetivo explicitar os conceitos básicos envolvidos no que se convencionou traduzir como escoragem de crédito - “*credit scoring*” - além de propor o desenvolvimento de uma extensão do modelo convencional que leve em conta o cenário macroeconômico para o período de duração da operação de concessão de crédito e a rentabilidade da operação. O modelo desenvolvido será aplicado sobre o produto Crédito Direto ao Consumidor - Lojista pois, conforme o capítulo II irá demonstrar, tal produto é o mais representativo dentro do mercado de crédito brasileiro.

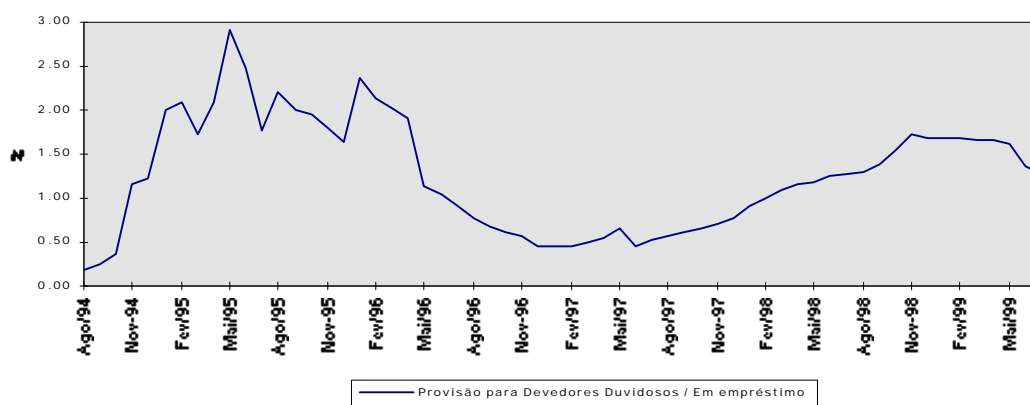
Tal modelagem sobre a análise do risco de concessão de crédito, amplamente difundida nos países desenvolvidos - notadamente nos EUA, ganha cada vez mais adeptos nas financeiras e bancos em atuação no mercado nacional devido a dois fatores conjuntos: a explosão da atividade creditícia após a implantação bem sucedida do programa de estabilização brasileiro de Julho de 1994 e o posterior estouro dos índices de inadimplência como reflexo direto da falta de cultura de obtenção de crédito da população brasileira - especialmente das classes C, D e E - em 1995 e da piora dos resultados macroeconômicos dos países emergentes em geral devido a eclosão da chamada Crise Asiática em 1997 que atingiu sensivelmente o público alvo de tal atividade através do crescimento dos índices de desemprego.

A análise gráfica abaixo visa exemplificar as mudanças observadas no mercado nacional de crédito direto ao consumidor em seguida a implantação do plano real em Julho de 1994.

Evolução do Mercado Brasileiro de Crédito Pós Plano Real



Evolução da Inadimplência do Mercado Brasileiro de Crédito Pós Plano Real



Conforme os gráficos acima explicitam o aumento dos índices de inadimplência se dá alguns períodos após o aumento da produção. Tal fato é compreensível devido ao fato de que o cliente que virá a se tornar inadimplente não é identificado como tal na data de abertura da operação. Somente ao longo do período de financiamento é que tal característica poderá ou não ser observada.

I) O MERCADO DE CONCESSÃO DE CRÉDITO BRASILEIRO:

- Crédito Imobiliário (37%)
 - Setor Público - Sistema Financeiro de Habitação¹
 - Setor Privado
- Crédito ao Consumidor (63%)
 - Empréstimos à Prazo (55%)
 - Crédito Direto ao Consumidor - Lojista (38%)
 - Empréstimo Pessoal (30%)
 - Arrendamento Mercantil - “*Leasing*” (30%)
 - Fomento Mercantil - “*Factoring*” (2%)
 - Crédito Rotativo
 - Cheque Especial (31,63%)
 - Cartão de Crédito (18,62%)
 - Cartão de Loja - “*Private Label*” (49,75%)

O resumo explicativo acima explicita a estrutura atual do mercado brasileiro de crédito ao consumidor para pessoa física² demonstrando que as operações ligadas ao comércio lojista (Crédito Direto ao Consumidor/Lojista + Cartão de Loja/“*Private Label*” e em alguma medida as operações com Cartão de Crédito) são as que dispõem de maior mercado consumidor. Tal fato é decorrente da carência existente no Brasil quanto a posse de utilitários domésticos básicos conforme a tabela abaixo² demonstra

¹ Não foram fornecidos os percentuais para as operações de Crédito Imobiliário.

² Fonte: Lloyds TSB Group - “Consumer Finance Report”/ Fevereiro de 1998.

através do percentual da população que possui o produto especificado sobre o total da população do país em questão:

Produto	Estados Unidos	União Européia	Brasil
Lava-Louças	51%	26%	4%
Secadora de Louças	52%	18%	5%
Microondas	86%	33%	6%
Freezer	40%	46%	12%
Geladeira	99%	97%	50%

Observação: As lojas de Departamento (eletrodomésticos) costumam ter um volume de vendas à prazo maior do que as de Artigos de Vestuário e do que os Supermercados logo, os produtos acima dispostos são bastante representativos das distorções entre o Brasil e o resto do mundo desenvolvido.

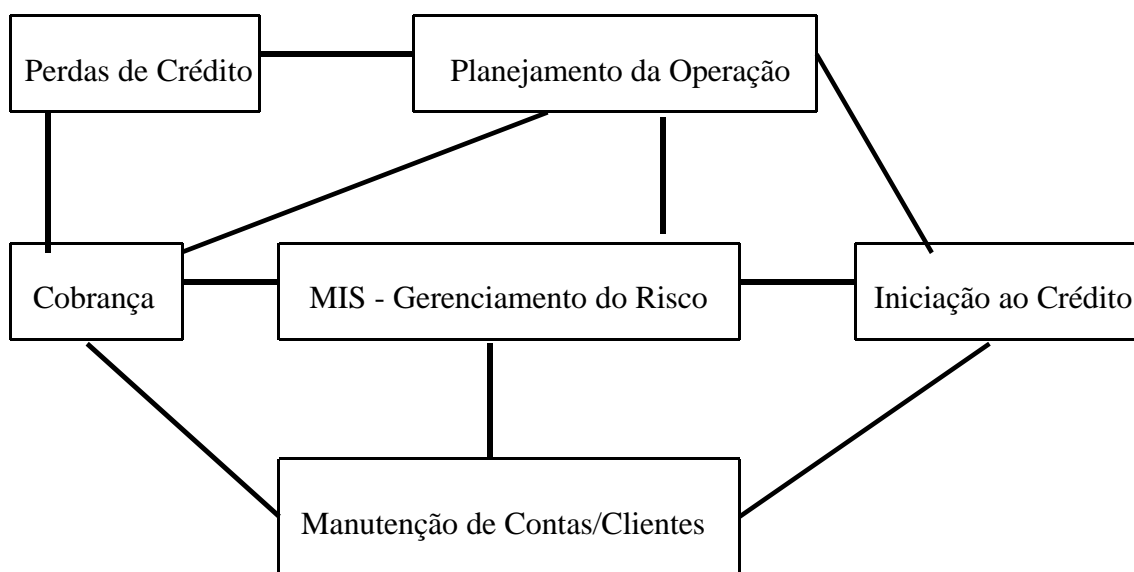
Conclui-se que a estrutura do mercado nacional de crédito aponta para taxas de crescimento positivas das vendas financiadas se assumirmos a hipótese de crescimento econômico sustentado sem altos índices de inflação. Paradoxalmente, é importante citar que a propensão ao consumo da população brasileira deve ser reduzida nos próximos meses com as mudanças na estrutura previdenciária que estão em tramitação no Congresso Nacional, de modo que, a questão que se impõe ao mercado de crédito quanto ao controle dos níveis da inadimplência independe de uma nova explosão creditícia, uma vez que a mesma não deve ocorrer até o final do atual governo federal.

II) ANÁLISE E CONCESSÃO DE CRÉDITO:

Qualquer instituição financeira que trabalhe com concessão de crédito tem que possuir uma estrutura tal que possibilite o desenvolvimento das seguintes ações:

- Primeiramente, planejar as operações a serem ofertadas e analisar as condições de pagamento do proponente ao crédito - *“Front-End”*
- Durante todo o processo, acompanhar os resultados das operações realizadas - *“MIS: Management Information System”*
- E finalmente, cobrar aqueles clientes que não efetuaram corretamente seus pagamentos - *“Back-End”*.

O estrutura abaixo³ visa demonstrar as interligações necessárias para que os diferentes departamentos de uma instituição financeira concedente de crédito possam alcançar bons resultados nas finanças e no tocante a conquista/manutenção de *“market-share”*.



³ Fonte: “Risco e Recompensa: O Negócio de Crédito ao Consumidor” - Credicard/Citibank - Página 24

Este trabalho se concentrará no desenvolvimento de um método econômico-estatístico que possibilite maximizar os resultados de “Front-End” a partir de uma análise precisa e ágil das condições de pagamento do proponente ao crédito num ambiente de concessão de crédito massificado, ou seja, a utilização de técnicas estatísticas sobre uma amostra representativa da população de proponentes ao crédito aliada aos cenários macroeconômicos do período da operação e as características financeiras de cada operação realizada possibilitará aos gerentes das linhas de crédito uma análise objetiva em substituição a análise subjetiva tradicionalmente realizada, aumentando assim, a capacidade de geração de novos negócios e a precisão das previsões quanto ao resultado.

O modelo tradicional para a análise de concessão de crédito foi desenvolvido nos anos 70's nos Estados Unidos em instituições tais como Citicorp, American Express, BankBoston, entre outros, e ficou conhecido como Escoragem de Crédito - “Credit Scoring”, onde a palavra escoragem se refere ao ranking que é montado com a pontuação obtida por cada cliente após a avaliação do modelo previamente desenvolvido.

A idéia central era de que a avaliação subjetiva feita pelos gerentes de linha baseada na teoria dos 3 C's - Caráter, Capacidade, Colateral - poderia ser aplicada a um método estatístico que escorasse uma base de clientes do passado próximo para que um ranking quanto aos índices de inadimplência destes mesmos clientes pudesse ser usado para identificar perfis de clientes mais ou menos arriscados, de modo que, os novos proponentes ao crédito teriam seus perfis analisados sobre a luz das conclusões tiradas na base de dados. Por exemplo, se a característica solteiro - encontrada em 90% dos

clientes da base de dados - se mostraram arriscados (maus pagadores⁴) é coerente que se olhe com mais atenção os clientes novos que sejam solteiros antes de conceder o crédito.

Tal processo seria feito para uma série de características - tal qual estado civil no exemplo, e para uma série de atributos - tal qual solteiro no exemplo, de modo que, o conjunto de informações fornecidas pelo proponente ao crédito em sua ficha cadastral forneceria uma base sólida para a análise de concessão de crédito. Tal técnica evoluiu até o que hoje é amplamente conhecido como Modelo de Escoragem de Crédito - “*Credit Scoring Model*”, que será o assunto do próximo capítulo.

⁴ A definição de maus pagadores será explorada na próxima seção em conjunto com a definição de bons

III) CREDIT SCORING

Abaixo serão listados os passos para o desenvolvimento de um modelo de escoragem de crédito original, sendo que as inovações propostas ao modelo serão apresentadas no sexto capítulo desta monografia.

III.1) BASE DE DADOS

Tal base de dados costuma ser extraída⁵ para o período de 1 ano corrido com o intuito de eliminar as distorções sazonais encontradas nos mercados de crédito - especialmente crédito lojista que concentra nos meses de Maio/Dia das Mães e Dezembro/Natal suas maiores produções. Além disso, tal amostra deve ser representativa da base de clientes atual para que o “*Credit Scoring*” a ser desenvolvido represente com fidelidade o universo de clientes da instituição financeira em questão - já que as características dos clientes das diferentes instituições são distintas na maioria dos casos.

Outra questão muito importante é a qualidade encontrada nos dados cadastrais. Talvez este seja o maior inimigo dos profissionais desenvolvedores de modelos de escoragem de crédito, uma vez que se os dados cadastrais não forem confiáveis o modelo pode sofrer dois tipos de viés: a diminuição do número de variáveis passíveis de análise e/ou a análise viesada de algumas variáveis que podem ter sido preenchidas corretamente para um universo muito pequeno de clientes - o que com toda certeza seria a pior das perdas.

pagadores.

Alguns modelos incluem na base de dados um número de observações referente aos clientes que tiveram suas propostas de crédito rejeitadas durante o período da base de dados. Tal providência se mostra coerente pois o modelo final deverá analisar o risco dos novos proponentes ao crédito de forma geral e não somente daqueles que seriam aprovados na análise subjetiva. Contudo, tal precaução é de difícil execução no Brasil pois as instituições financeiras não tem a cultura de armazenar os dados dos clientes rejeitados no passado.

Para o desenvolvimento deste modelo foi obtida uma amostra de 3.446 contratos junto a financeira Lecca, que atua principalmente no mercado de crédito direto ao consumidor lojista do Estado do Rio de Janeiro, para o período de Janeiro a Dezembro de 1998.

III.2) DEFINIÇÃO DE BONS E MAUS

Normalmente são considerados bons clientes aqueles que nunca atrasaram ou na pior das hipóteses foram moderadamente inadimplentes durante o financiamento, de modo que, o planejamento da empresa não foi afetado e o custo resultante de tal inadimplência foi amplamente coberto pelas receitas provenientes dos encargos diários e de mora. A maioria das empresas define como bom cliente aquele que nunca teve atraso superior a 30 ou 60 dias durante seu contrato.

⁵ Empresas que não possuem este tipo de banco de dados costumam comprar tal base de dados de algum concorrente que acredita-se possuir uma base de clientes e produtos semelhantes.

Semelhantemente, são considerados maus clientes aqueles que foram altamente inadimplentes durante o prazo do contrato ou que em algum momento pararam de pagar por completo as prestações, de modo que, geraram perdas/saldos devedores para a instituição. Usualmente estes clientes são aqueles que possuíram em algum momento do contrato atraso maior ou igual a 60 ou 90 dias⁶.

Para o desenvolvimento deste modelo foram considerados bons clientes aqueles que pagaram seu contrato integralmente e em nenhum momento incorreram em atraso maior ou igual do que 60 dias. De forma antagônica, foram considerados maus clientes aqueles que em algum momento do contrato, tenha sido ele integralmente pago ou não, incorreram em atraso maior ou igual do que 60 dias.

III.3) DEFINIÇÃO DAS VARIÁVEIS DO MODELO

Conforme explicitado anteriormente, o primeiro critério de seleção das variáveis a serem utilizadas no modelo deve ser a qualidade cadastral encontrada na base de dados, uma vez que variáveis com falhas no preenchimento devem ser descartadas. A seguir pode-se separar as variáveis nos seguintes grupos: cadastrais, comportamentais e do produto.

As variáveis cadastrais são aquelas que foram obtidas na ficha cadastral de requisição de crédito - por exemplo: posse ou não de telefone residencial, estado civil, idade, número de dependentes, bairro residencial, tipo de residência (alugada, própria, de

⁶ Clientes que possuem contratos ativos, ou seja, a última prestação ainda está por vencer, são deixados a margem da análise pois somente podemos ter uma fotografia correta do contrato após seu término.

terceiros, etc.), profissão, cargo, comprometimento de renda na data de abertura do contrato e nível de escolaridade entre outras.

Já as variáveis comportamentais se referem ao comportamento de pagamento do cliente nos contratos anteriores já liquidados - por exemplo: maior dia de atraso de todos os contratos do cliente, número de contratos liquidados, ticket médio dos contratos anteriores e intervalo médio entre a abertura e o término dos contratos anteriores, entre outras.

As chamadas variáveis do produto refletem as características do produto vendido - por exemplo: taxa de juros cobrada, número de prestações do financiamento, dias de carência, valor do crédito concedido, valor do IOC⁷ (Imposto sobre Operações de Crédito), forma de pagamento (cheque pré-datado⁸ ou carnê bancário) e valor da TAC (Tarifa de Abertura de Crédito) entre outras.

Para o desenvolvimento deste modelo as seguintes variáveis foram utilizadas: forma de pagamento (cheque pré-datado ou carnê), entrada (financiamento com ou sem entrada), idade do cliente, telefone residencial (possui ou não possui), profissão (autônomo, profissional liberal, aposentado, assalariado da iniciativa privada, funcionário público e empresário), comprometimento de renda (% do valor da prestação sobre a renda mensal do cliente) e SPC (passagem ou não pelo SPC nos últimos 6 meses).

⁷ Até o último dia 18 de Outubro de 1999 a tarifa do IOC era de 6% ao ano mais um adicional de 0,35% sobre o valor do crédito concedido. Após tal data foi fixada em 1,5% ao ano também sobre o valor do crédito concedido.

⁸ Estima-se que cerca de 60% de todas as operações de crédito no Brasil são feitas por intermédio de cheques pré-datados. Tal característica não é encontrada no resto do mundo.

É importante frisar que os resultados do modelo são tão melhores quanto melhores forem as variáveis incluídas no próprio. Infelizmente, devido ao pequeno número de informações obtidas junto a financeira citada anteriormente, a elaboração de algumas variáveis compostas se tornou inviável. Por outro lado, os resultados finais não devem ser extremamente prejudicados devido ao fato de que as variáveis citadas acima são tidas como principais na elaboração de qualquer modelo de escoragem de crédito.

III.4) ANÁLISE BIVARIADA

É chamado de Análise Bivariada o cálculo da frequência - número de contratos - que é feita sobre cada atributo de cada variável anteriormente definida. Deve-se entender como atributo os diferentes valores que cada variável pode assumir - por exemplo: a variável profissão pode ter como possíveis atributos autônomo e profissional liberal, entre os outros citados no item anterior. O nome Bivariada vem do fato de que a contagem será feita separadamente para cada tipo de contrato, ou seja, cada variável será contada em relação ao status do contrato “Bom ou Mau” de modo a gerar a frequência de contratos bons e maus por cada atributo de cada variável.

Após a obtenção de tais frequências é calculado o percentual de bons contratos sobre maus contratos para cada atributo. Tal percentual servirá de base para que diferentes atributos de uma mesma variável possam ser agrupados de modo a formar novas variáveis.

Exemplificando, temos que a variável Profissão possui no atributo Assalariado Privado o número de 1245 contratos bons e 70 contratos maus, assim como o atributo

Assalariado Público possui o número de 584 contratos bons em contrapartida a 31 contratos maus. O percentual de bons sobre maus do atributo Assalariado Privado seria $(1245/3228)/(70/218) = 1,20$ denotando que para cada contrato mau concedido a profissionais Assalariados Privados 1,20 contratos bons foram gerados, logo tal atributo deve ser considerado um bom indicador isolado do baixo risco de concessão de crédito de operações realizadas mediante o pagamento com cheque pré-datado. Observe ainda que tal cálculo quando feito para o atributo Assalariado Público tenha resultado em um percentual de bons contratos sobre maus contratos de 1,27. Pode-se concluir que tais atributos possuem isoladamente um semelhante indicador de risco, de modo que deverão ser agrupados de modo a gerar a variável Assalariado, por exemplo, que deverá ser submetida a uma nova Análise Bivariada. Tal processo prossegue até que os atributos resultantes possuam percentuais de bons sobre maus contratos significativamente distintos uns dos outros⁹. Abaixo estão listadas as classificações usuais para os diferentes resultados do percentual de bons contratos sobre maus contratos. Esta classificação define o critério usual para o agrupamento de variáveis.

→ Muito Mau	percentual até 0,50
→ Mau	percentual entre 0,60 e 0,88
→ Neutro	percentual entre 0,89 e 1,20
→ Bom	percentual entre 1,21 e 1,50
→ Muito Bom	percentual superior a 1,50

⁹ Também é levado em conta o número de observações encontradas para cada atributo, uma vez que não faz muito sentido manter isolado um determinado atributo que tenha uma frequência extremamente baixa (menos que 5% da amostra em geral). Para estes casos existe um relaxamento nos critérios de agrupamento. Além disso, para variáveis contínuas - tais como idade - espera-se que a distribuição de dos percentuais de bons sobre maus das variáveis resultantes da Análise Bivariada seja próxima a uma distribuição linear ou a uma curva.

O Anexo I traz os resultados obtidos para a análise bivariada deste modelo.

III.5) REGRESSÃO LINEAR

Uma vez agrupados os atributos das diferentes variáveis teremos uma série de variáveis dummy - por exemplo: o contrato é de um cliente Autônomo ou não, se sim a variável dummy autônomo assumirá valor 1, se não assumirá valor nulo. A partir de tais variáveis será rodada uma regressão linear multivariada com base no Método dos Mínimos Quadrados Ordinários.

É importante frisar que tal modelo pode ser desenvolvido através de regressões lineares, não havendo necessidade da utilização de modelos específicos para mensuração de probabilidade - tais como o modelo Probit -, já que o objetivo do modelo não é mensurar a probabilidade de um cliente se tornar bom ou mau pagador e sim, desenvolver um critério para a ordenação de todos os proponentes ao crédito, de modo que, os classificados como possíveis maus pagadores sejam o quão mais diferentes quanto possível daqueles possíveis bons pagadores. Sendo assim, o fato de que o modelo de regressão linear não resultará num intervalo de valores exatamente entre 1 e 0 não consiste num impecílio para que o objetivo do modelo seja alcançado, ou seja, a utilização de regressão linear - um método amplamente conhecido e de fácil compreensão atende perfeitamente o objetivo principal do modelo: separar a população de possíveis bons pagadores da população de possíveis maus pagadores.

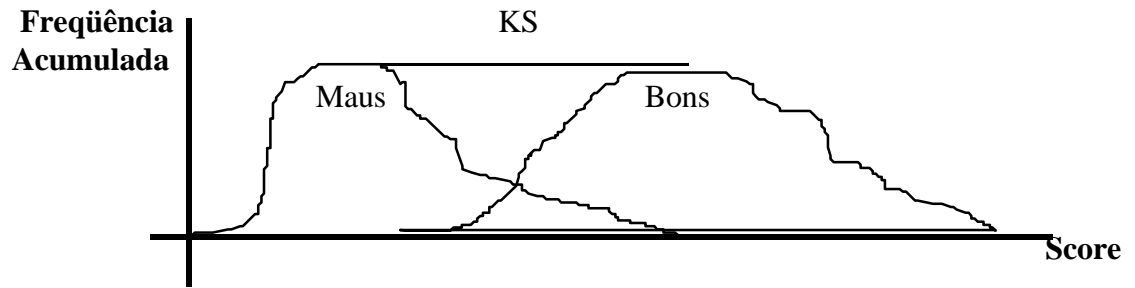
A partir dos resultados obtidos com este modelo original são feitos testes de validação - teste de significância de Student e de correlação de Spearman basicamente.

Tais testes servirão de base para a exclusão das variáveis dummy insignificantes estatisticamente ou altamente correlacionadas entre si. Após tais exclusões, se necessárias forem, roda-se novamente o modelo através de Regressão Linear com o Método dos Mínimos Quadrados Ordinários, onde a variável dependente Y_i representará o escore/pontuação de um determinado proponente ao crédito de acordo com as variáveis do mesmo (variáveis independentes X_i), de acordo com os coeficientes estimados pelo modelo (β_i 's) e de acordo com o erro estatístico aleatório (U_i).

O anexo II contém os resultados obtidos e os respectivos comentários para estas regressões lineares preliminares, assim como, para os testes de significância e correlação.

III.6) TESTE DE PERFORMANCE - KOMOGOROV-SMIRNOV (KS):

Dado que o objetivo de qualquer modelo de escoragem de crédito é analisar o proponente ao crédito de modo a tomar uma dentre duas decisões - aprovar ou não a concessão do crédito - e dado que segundo a modelagem até agora explicitada só existem dois tipos de clientes: bons, para os quais o crédito deve ser aprovado e maus, para os quais o crédito deve ser negado, tem-se que o modelo será tão bom quanto puder separar estes dois tipos de clientes, estas duas populações. Para tal fim, o teste usualmente aplicado é o Kolmogorov-Smirnov. O resultado deste teste explicita a maior diferença absoluta (em módulo) encontrada ao longo da distribuição acumulada de frequência dos contratos bons e maus por valor de score encontrado. O gráfico abaixo exemplifica com clareza tal teste.



Modelos de credit scoring são considerados eficientes se o índice de tal teste, o KS, estiver em torno de 35%.

I V) PONTO DE CORTE - INTRODUZINDO A RENTABILIDADE DO PRODUTO NO MODELO

Após a definição da equação que pontuará os proponentes definitivamente são geradas classes de score através da Análise Bivariada nos moldes do que foi exposto anteriormente, de modo a agrupar os score próximos. Tal Análise Bivariada também se encontra no Anexo I.

Como último passo do desenvolvimento do modelo se faz necessário o estabelecimento da regra de decisão para a concessão de crédito, ou seja, qual será a pontuação mínima exigida do proponente ao crédito para que sua proposta seja aprovada? Tal limite poderia ser estabelecido subjetivamente seguindo alguma política de conquista de “*market-share*”, ou seja, manutenção da taxa de aprovação em conjunto com a redução da taxa de perda ou manutenção da taxa de perda em conjunto com o aumento da taxa de aprovação.

O que se recomenda para que o modelo seja o mais coerente possível com a realidade da empresa e do produto analisado é que o ponto de corte seja tal que o número de contratos maus aceitos em conjunto com o número de contratos bons aceitos torne o produto em questão viável financeiramente. Em outras palavras, se faz necessário que a perda gerada pelos contratos maus seja “coberta” pelo lucro gerado pelos contratos bons. Consequentemente, é preciso calcular quanto é a perda gerada por um contrato mau e quanto é o ganho gerado por um contrato bom em cada classe de score resultante da análise bivariada anteriormente explicada para que se tenha a proporção de quantos contratos bons terão que ser vendidos para que um contrato mau possa também

ser vendido sem afetar a saúde financeira do produto. Tal cálculo é obtido com a chamada Análise de Rentabilidade do Produto, onde os contratos bons e maus são analisados separadamente no que diz respeito as receitas de juros, de refinanciamento e de atraso/antecipação e aos custos operacionais, de cobrança, de carregamento (float) e aos percentuais de perda do portfólio.

Devido ao caráter naturalmente sigiloso de tais informações referentes a rentabilidade do produto, e conseqüentemente da empresa, não foi possível obter os dados reais quanto a rentabilidade dos bons e dos maus clientes. Assim sendo, uma simulação foi feita, de modo a gerar resultados para o que se denominou Ganho dos Bons e Perda dos Maus para cada classe de score. Tal simulação foi feita utilizando alguns conceitos básicos na rentabilização de produtos de financiamento parcelado. Os resultados assim como os conceitos utilizados estão detalhados no Anexo III.

Finalmente, de posse da perda dos maus, do ganho dos bons e da freqüência ponderada de bons e maus contratos por classe de score é possível calcular a partir de que classe de score a operação se torna rentável. Tal corte costuma se dar em torno da classe de score que possui a maior diferença acumulada entre a distribuição de bons e maus contratos, de modo a formar a seguinte regra de decisão:

“Se o cálculo do score do proponente resultar num valor menor ou igual do que o limite inferior da primeira classe de score onde a operação começa a se tornar rentável, *Então* o crédito não será consentido. Caso contrário a proposta será aprovada!”

V) ADENDO AO MODELO: CONSIDERAÇÕES MACRO-ECONÔMICAS

Na medida em que o modelo até aqui apresentado tem como objetivo principal desenvolver um método estatístico-matemático para previsão do comportamento de pagamento dos clientes, classificados simplesmente como bons ou maus pagadores, uma questão que se impõe é a investigação quanto a origem de tal inadimplência.

Tal tarefa não é nada fácil uma vez que dezenas de fatores influenciam não só na capacidade de pagamento como também na pré-disposição para o pagamento dos indivíduos. Esta talvez seja a maior virtude e também o maior paradoxo dos modelos de escoragem de crédito até aqui apresentados. Na medida em que tais modelos não se preocupam com a origem do fenômeno da inadimplência e sim com a ordenação dos clientes de forma a separar o máximo possível os possíveis inadimplentes dos pouco possíveis inadimplentes, obtém resultados extremamente satisfatórios, reduzindo tal inadimplência, mas por outro lado não identificando os principais fatores que causam-na. É de se concluir que talvez por isso tais modelos não apresentem resultados expressivos para o coeficiente de determinação R^2 .

O objetivo deste capítulo é tentar agregar tal ferramenta aos modelos de escoragem de crédito, ou seja, incluir os cenários macroeconômicos no desenvolvimento do modelo.

Inicialmente deve-se esperar que qualquer variável econômica observada no período de abertura do contrato tenha um período de maturação até que produza efeitos na capacidade e disposição de pagamento dos indivíduos. Tal período foi estimado como

em dois anos¹⁰ para o produto Cartão de Crédito nos Estados Unidos. Como não foi possível obter dados ou estudos semelhantes para o Brasil este trabalho irá adotar tal medida também para o produto crédito direto ao consumidor lojista parcelado.

Com o objetivo de investigar as origens da inadimplência serão estimados os coeficientes lineares de indicadores macroeconômicos através do método dos mínimos quadrados ordinários. Os indicadores utilizados estão definidos abaixo¹¹:

Taxa de Juros: dada pela taxa CDI (Certificado de Depósito Interbancário) – Tal taxa foi escolhida devido a sua longa existência e por ser um conhecido indicador para o mercado financeiro em geral. Além disso, a intuição para inclusão desta variável é a de que quanto maiores as taxas mais se afastarão os bons pagadores, uma vez que os mesmos costumam ter maiores alternativas para compra do que somente o crediário que acaba sendo inundado de clientes com baixo poder aquisitivo. A citação encontrada no último parágrafo da página 162 do livro Credit Risk Modeling de Elysabeth Mays traduz bem este sentimento:

“ Globalization, deregulation, technological change, and more recently Medicaid and welfare reforms will continue to limit the incomes of households with relatively little education and training. These households will borrow as much as lenders are willing to provide in an effort to maintain their living standards in the face of shrinking incomes.”

¹⁰ Fonte: Mays, Elizabeth, Credit Risk Modeling, 1998, quinto parágrafo da página 156.

¹¹ Fonte: MCM Consultores

Vendas do comércio: Quanto maior a base de clientes mais difícil se torna o controle do risco da operação através de métodos subjetivos de análise de capacidade de pagamento, assim sendo, conforme já citado na introdução deste trabalho, o volume de vendas parece ser um forte indicador do nível de risco associado as operações de crédito. Tal variável é dada sob forma de variação em relação ao valor médio encontrado para o ano de 1998.

Salário Real deflacionado pelo INPC (Índice Nacional de Preços ao Consumidor): Tal variável tem como objetivo ser o indicador da renda dos consumidores deste produto, uma vez que o grande público deste tipo de financiamento se localiza nas classes econômico-sociais C, D e E.

Índice de Desemprego Aberto – IBGE: Tal indicador é de fácil intuição uma vez que uma pessoa desempregada, na grande maioria das vezes, perde toda sua fonte de renda, de modo que, não haverão mais recursos disponíveis para que as suas dívidas outrora admitidas continuem a ser saldadas. Esta variável foi utilizada sob forma de fator numérico.

INPC (Índice Nacional de Preços ao Consumidor): Além de estar altamente relacionado com o desenvolvimento das atividades creditícias, uma vez que ambientes com alta inflação historicamente afastam os proponentes ao crédito. Além disso, um processo inflacionário afeta diretamente o poder aquisitivo das classes econômico-sociais citadas anteriormente, que não costumam possuir qualquer tipo de proteção bancária frente ao chamado imposto inflacionário. Tal variável foi utilizada como fator numérico.

Como variável dependente foi utilizado o saldo das negativações e reabilitações do SPC (Sistema de Proteção ao Crédito) que representa o maior e mais usualmente utilizado indicador de inadimplência do país. O software SPSS foi utilizado para obtenção dos resultados da regressão linear multivariada através do método dos mínimos quadrados ordinários. Abaixo estão dispostos os resultados encontrados:

1ª Regressão Linear: incluindo todas as variáveis

Equation Number 1 Dependent Variable.. SPC

CDI DESEMPRE INPC SAL_RIO VENDAS__

Variable(s) Entered on Step Number

- 1.. VENDAS
- 2.. SALÁRIO
- 3.. DESEMPREGO
- 4.. CDI
- 5.. INPC

Multiple R	,81382
R Square	,66230
Adjusted R Square	,64709
Standard Error	36714,50365

Analysis of Variance

DF	Sum of Squares	Mean Square
----	----------------	-------------

Regression	5	293438547658,82760	58687709531,7655
Residual	111	149622980358,36920	1347954778,00333

F = 43,53834 Signif F = ,0000

----- Variables in the Equation -----

Variable	B	SE B	Beta	T	Sig T
CDI	-97157,23315	71087,53978	-,253856	-1,367	,1745
DESEMPRE	1891456,4733	253146,5719	,486260	7,472	,0000
INPC	31297,091949	69413,66810	,084754	,451	,6530
SALÁRIO	362,517354	201,244695	,106986	1,801	,0744
VENDAS	1268,086565	280,893908	,369696	4,514	,0000
(Constant)	-2004064,566	294730,8063		-6,800	,0000

Tais resultados sugerem a exclusão das variáveis CDI e INPC devido a baixa significância dos coeficientes obtidos. Após a exclusão de tais variáveis uma Segunda regressão foi desenvolvida:

2ª Regressão Linear: excluindo CDI e INPC

Equation Number 2 Dependent Variable.. SPC

DESEMPRE SAL_RIO VENDAS__

Variable(s) Entered on Step Number

1.. VENDAS

2.. SALÁRIO

3.. DESEMPREGO

Multiple R	,80558
R Square	,64895
Adjusted R Square	,63963
Standard Error	37100,20057

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	3	287525516281,36700	95841838760,4556
Residual	113	155536011735,82980	1376424882,61796

F = 69,63100 Signif F = ,0000

----- Variables in the Equation -----

Variable	B	SE B	Beta	T	Sig T
DESEMPRE	2111746,3808	224733,9052	,542892	9,397	0000
SALÁRIO	240,410941	193,048471	,070950	1,245	,2156
VENDAS	1642,869013	194,984197	,478960	8,426	,0000
(Constant)	-2337988,650	235246,4260		-9,938	,0000

Esta segunda regressão revela que a variável salário real deflacionado pelo INPC não é estatisticamente significativa, logo uma nova regressão foi rodada excluindo tal variável.

3ª Regressão Linear: excluindo SALÁRIO

Equation Number 3 Dependent Variable.. SPC

DESEMPRE VENDAS

Variable(s) Entered on Step Number

1.. VENDAS

2.. DESEMPREGO

Multiple R	,80258
R Square	,64413
Adjusted R Square	,63789
Standard Error	37189,73040

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	2	285390858608,11930	142695429304,060
Residual	114	157670669409,07750	1383076047,44805

F = 103,17251 Signif F = ,0000

----- Variables in the Equation -----

Variable	B	SE B	Beta	T	Sig T
DESEMPREGO	2164025,7935	221310,9590	,556332	9,778	,0000
VENDAS	1656,323616	195,154457	,482882	8,487	,0000
(Constant)	-2377277,081	233683,9546		-10,173	,0000

Esta última regressão apresentou resultados satisfatórios já que mesmo com o pequeno número de indicadores macroeconômicos o coeficiente de explicação do modelo R2 chegou a 64,413, indicando que ao redor de 60% do índice de inadimplência observado na década de 90 pode ser explicado pelas flutuações do índice nacional de desemprego aberto e pelas flutuações de produção/volume de créditos concedidos.

A partir de tais resultados existem basicamente dois métodos para a inclusão de tais indicadores macroeconômicos no modelo.

O primeiro seria rodar novamente a regressão final do modelo de escoragem de crédito acrescido das variáveis macroeconômicas. Tal procedimento deve aumentar o grau de explicação do modelo mas provavelmente irá gerar novos focos de correlação. É de se imaginar que variáveis tais como profissão, cargo e taxa de desemprego sejam correlacionadas por exemplo.

O segundo método seria a geração de um índice que funciona-se como um adicional ou redutor da pontuação originalmente obtida pelo cliente. Tal método poderia se utilizar de instrumentos tais como análise fatorial e modelos auto-regressivos. Um exemplo de

tal método é encontrado no livro *Credit Risk Modeling* da autora Elyzabeth Mays nas páginas 166 e 167, como segue:

$$\text{“ Pit} = a + b1 * \text{Sito} + b2 * \text{LEIit}_0$$

Where: Pit = Probability of Default by Customer i during period t,

Sit₀ = Credit Score for Customer i at time t₀

LEIit₀ = Leading Economic Indicator for Customer i at time t₀, and a, b1 and b2 are estimated parameters ”

Devido ao fato de que o banco de dados obtido junto a financeira Lecca não possui data de abertura, o que impossibilita a identificação do período referente a operação em questão, não foi possível aferir o resultado de algum destes dois métodos na prática. Assim sendo, fica a sugestão para que um futuro trabalho complemente tal teoria de modo a tornar cada vez mais eficientes estes tipos de modelo cada vez mais utilizados no mercado de crédito brasileiro.

VI) CONCLUSÕES

É importante salientar que a elaboração de métodos do tipo do apresentado neste trabalho representam mais do que simples ferramentas para aumento do lucro e da eficiência das instituições financeiras. Percebe-se, inclusive nos estudos oficiais¹² divulgados recentemente a preocupação de que quanto maiores os índices de inadimplência maiores serão as taxas de juros e conseqüentemente maiores serão as punições sobre os bons pagadores, que mesmo possuindo um histórico de pagamento

¹² Fonte: Juros e Spread Bancário no Brasil, Banco Central do Brasil, 1999

irretocável serão obrigados a pagar taxas de juros exorbitantes devido aos altos índices de inadimplência de qual não faz parte.

Além disso, pelo fato do país ainda possuir um estoque de injustiça social significativa, assim como, uma carência enorme por tecnologia e produtos básicos da vida moderna cotidiana, conceder crédito eficientemente é o caminho para a expansão cada vez maior do poder aquisitivo da parcela menos favorecida da população, uma vez que a experiência passada de nosso país não recomenda tal expansão pela via governamental paternalista, e sim pelo amadurecimento do relacionamento entre proponentes e concedentes, maior agilidade e regulamentação dos aspectos jurídicos envolvidos e constante modernização das técnicas empregadas.

VII) ANEXOS

ANEXO I - ANÁLISE BIVARIADA

• 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:

As variáveis contínuas com altos intervalos de valores tais como idade demandam o desenvolvimento de um agrupamento subjetivo inicial com intervalos de 05 anos (tabela a) para que uma melhor observação possa ser feita. Posteriormente, adota-se os critérios acima explicados (frequência e percentual de bons sobre maus) para agrupar definitivamente, gerando as classes explicitadas na tabela b.

a)

Classe	Bons			Maus			% Bom/% Mau
	Frequência Acumulada	Frequência	%	Frequência Acumulada	Frequência	%	
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)

Classe	Bons			Maus			% Bom/% Mau
	Frequência			Frequência			
	Acumulada	Frequência	%	Acumulada	Frequência	%	
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
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:

Para tal variável, foram necessárias duas análises.

a)

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
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:

Tal variável foi agrupada de acordo com os índices máximos exigidos nas diferentes situações de concessão de crédito. Na tabela a, os percentuais foram divididos com intervalos de 05 pontos percentuais. Na tabela b, utilizou-se os critérios usuais acima explicados (frequência e percentual de bons sobre maus) para agrupar definitivamente as classes dos atributos de tal variável.

a)

	Bons			Maus			
	Frequência			Frequência			
Classe	Acumulada	Frequência	%	Acumulada	Frequência	%	% Bom/% Mau
10	295	295	9,14%	10	10	4,59%	1,99
15	1136	841	26,05%	58	48	22,02%	1,18
20	2108	972	30,11%	113	55	25,23%	1,19
25	2893	785	24,32%	183	70	32,11%	0,76
30	3136	243	7,53%	209	26	11,93%	0,63
35	3200	64	1,98%	215	6	2,75%	0,72
40	3228	28	0,87%	218	3	1,38%	0,63
Total	-	3.228	100,00%	-	218	100,00%	

b)

	Bons			Maus			
	Frequência			Frequência			
Classe	Acumulada	Frequência	%	Acumulada	Frequência	%	% Bom/% Mau
10	295	295	9,14%	10	10	4,59%	1,99
20	2.108	1.813	56,16%	113	103	47,25%	1,19
40	3.228	1.120	34,70%	218	105	48,17%	0,72
		3.228	100,00%		218	100,00%	

- SPC (Sistema de Proteção ao Crédito):

Classe	Bons		Maus		% Bom/% Mau
	Frequência	Percentual	Frequência	Percentual	
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%	

- Score Final:

Score	# Bons	# Maus	% Bons	% Maus	% Bons/% Maus	Classe
<= 56	0	1	0,0000	0,0046	0,0000	1
57	0	0	0,0000	0,0000	-	2
58	1	1	0,0003	0,0046	0,0675	2
59	0	0	0,0000	0,0000	-	2
60	0	0	0,0000	0,0000	-	2
61	0	1	0,0000	0,0046	0,0000	3
62	0	0	0,0000	0,0000	-	3
63	0	1	0,0000	0,0046	0,0000	3
64	2	6	0,0006	0,0275	0,0225	3
65	3	1	0,0009	0,0046	0,2026	3
66	6	4	0,0019	0,0183	0,1013	4
67	1	2	0,0003	0,0092	0,0338	4
68	1	2	0,0003	0,0092	0,0338	4
69	3	2	0,0009	0,0092	0,1013	4
70	0	6	0,0000	0,0275	0,0000	4
71	6	4	0,0019	0,0183	0,1013	5
72	12	6	0,0037	0,0275	0,1351	5
73	8	5	0,0025	0,0229	0,1081	5
74	13	7	0,0040	0,0321	0,1254	5
75	20	2	0,0062	0,0092	0,6753	5
76	13	0	0,0040	0,0000	-	6
77	12	3	0,0037	0,0138	0,2701	6
78	24	4	0,0074	0,0183	0,4052	6
79	28	10	0,0087	0,0459	0,1891	6
80	13	8	0,0040	0,0367	0,1097	6
81	45	16	0,0139	0,0734	0,1899	7
82	27	7	0,0084	0,0321	0,2605	7
83	70	14	0,0217	0,0642	0,3377	7
84	28	2	0,0087	0,0092	0,9455	7
85	58	9	0,0180	0,0413	0,4352	7
86	18	6	0,0056	0,0275	0,2026	8
87	76	9	0,0235	0,0413	0,5703	8
88	93	9	0,0288	0,0413	0,6979	8
89	184	14	0,0570	0,0642	0,8876	8
90	41	0	0,0127	0,0000	-	8
91	184	14	0,0570	0,0642	0,8876	9
92	133	3	0,0412	0,0138	2,9940	9
93	71	3	0,0220	0,0138	1,5983	9
94	142	4	0,0440	0,0183	2,3975	9
95	281	7	0,0871	0,0321	2,7110	9
96	127	3	0,0393	0,0138	2,8589	10
97	240	3	0,0743	0,0138	5,4027	10
98	304	6	0,0942	0,0275	3,4217	10
99	76	2	0,0235	0,0092	2,5663	10
100	102	1	0,0316	0,0046	6,8885	10
> 100	762	10	0,2361	0,0459	5,1461	11
Total	3228	218				
Total Geral		3446				

ANEXO II - REGRESSÃO LINEAR E TESTES DE SIGNIFICÂNCIA E CORRELAÇÃO

Após a definição das classes das variáveis dada pela análise bivariada descrita acima, o próximo passo é rodar a regressão linear multivariada 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 |

A regressão inicial foi feita com todas as variáveis. O resultado abaixo foi obtido através do software Eviews.

LS // Dependent Variable is SCORE

Sample: 1 3446

Included observations: 3446

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

	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	

O resultado da regressão acima 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.

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

Teste de Hipótese)

H0: Coeficiente = 0

H1: Coeficiente \neq 0, onde \neq significa diferente de

Regra de decisão)

Se $|t\text{-value}| > t\text{-crítico}$ → Então, rejeito H_0 ao nível de significância de 5%.

Se $|t\text{-value}| < t\text{-crítico}$ → Então, não rejeito H_0 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), denotando 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 da mesma.

O segundo teste a ser feito é o de verificação da existência de correlação das variáveis. Através do software estatístico SPSS, foi possível 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. Abaixo estão expostos os resultados para os coeficientes de Correlação de Spearman.

Parte 1)

	APO	ASS	AT_10	AUTO	CA	CHE	
ASS	-,3230 ,000						
AT_10	-,0275 ,107	-,0120 ,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	-,6381	,0384	-,2107	-,0201	,0201	

	,000	,000	,024	,000	,237	,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
<i>Parte 2)</i>						
	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, em módulo, 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 deve-se excluir nenhuma variável do modelo devido a correlação, uma vez que é perfeitamente aceitável que as variáveis tais como comprometimento de renda entre 10% e 20% e comprometimento de renda entre 20% e 40%.

Conclui-se, a partir dos testes feitos anteriormente, que seria interessante excluirmos do modelo a variável comprometimento de renda entre 20% e 40% devido a baixa significância observada para o coeficiente C(12). O resultado final obtido após tal exclusão se encontra abaixo.

LS // Dependent Variable is SCORE

Sample: 1 3446

Included observations: 3446

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

	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

Após a definição do modelo de regressão final, testa-se a validade do 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/ou

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 freqüência 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 50,78%, ou seja, o K-S deste modelos atingiu o nível de 50,78% de separação

entre bons e maus contratos/clientes, o que pode ser considerado como um resultado muito bom, já que o z-value para tal coeficiente se mostrou significativo estatisticamente.

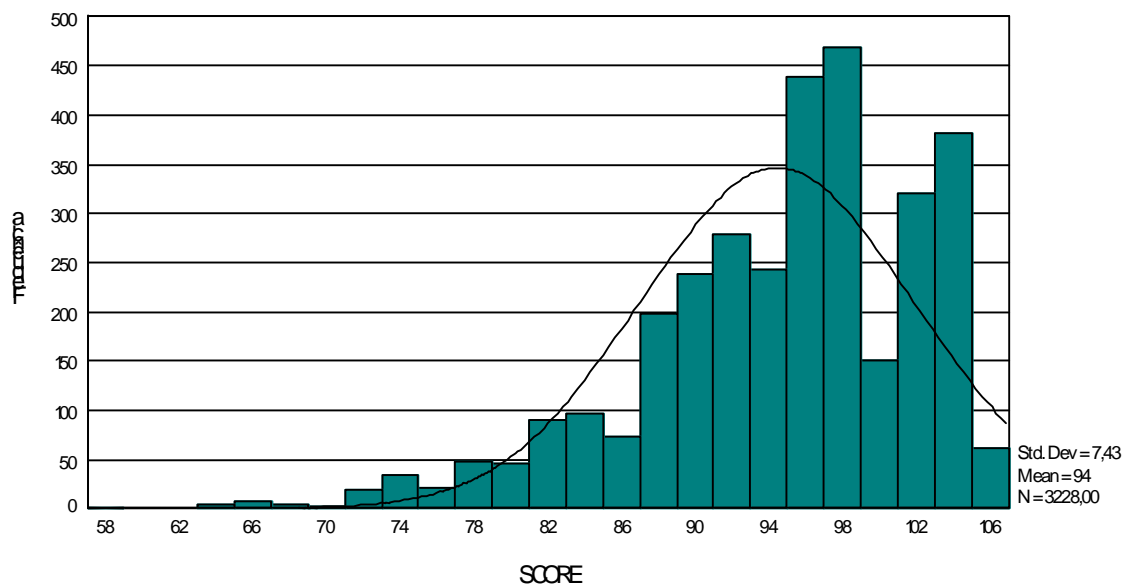
Logo, pode-se considerar satisfatório o resultado obtido pelo modelo, uma vez que, apesar do valor encontrado pelo R2 ter sido baixo – devido ao pequeno número de variáveis incluídas no modelo principalmente – a separação entre as populações de bons e maus clientes foi considerável. Deve-se esperar ainda, que a inclusão de novas variáveis seria interessante do ponto de vista explicativo do modelo – tornando-o entretanto pesado computacionalmente.

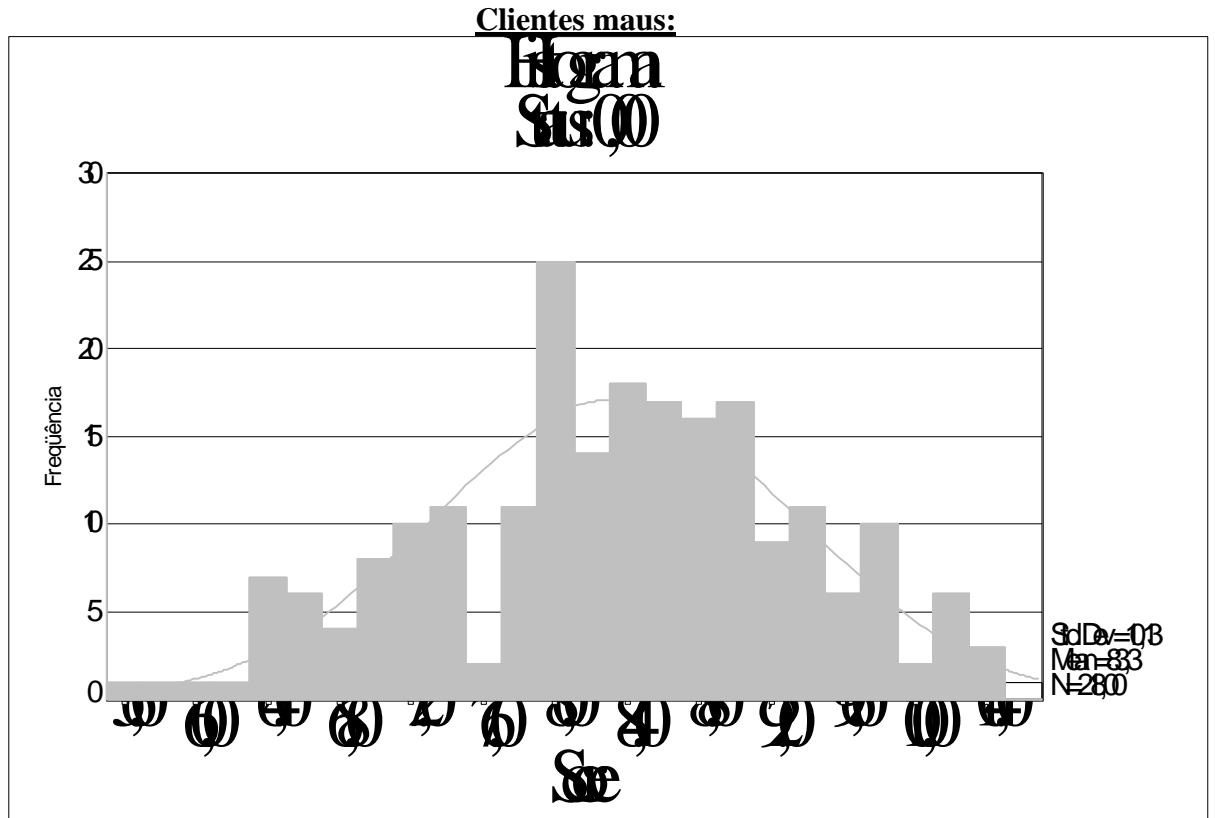
Abaixo apresentamos os gráficos de distribuição de frequência dos clientes bons e maus por faixa de score.

Cientes bons:

Histograma

Status: 1,00





ANEXO III - RENTABILIDADE DO PRODUTO

Abaixo estão descritas algumas informações constantes em qualquer operação de financiamento de crédito direto ao consumidor lojista. É importante salientar o que já foi explicado anteriormente, ou seja, tais valores foram simulados devido ao caráter sigiloso adotado pela grande maioria de bancos e financeiras em relação a tais informações:

Variáveis	Maus	Bons
Volume de Financiamento :	500,00	500,00
Plano Médio :	6	4,7
Contrato Médio :	500,00	420,00
TAC :	9,00	2,00
Taxa de Financiamento :	9,70%	7,43%

Taxa de Juros de Mora :	8,00%	8,00%
IOF :	0,0041%	0,0041%
IOF (fixo):	0,3800%	0,3800%
CPMF:	0,0000%	0,0000%
Confins:	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
Custo Fixo :	-	-
Índice de Perda :	50,00%	50,00%

Receitas

Líq. de Financto :	147	82
TAC :	9	2
Float :	1	1
Juro de Mora e Multa :	-	-
Custo do Atraso :	-	-
Total =>	156	86

Despesas

IOF :	2	1
IOF (fixo):	2	2
CPMF:	-	-
Confins:	6	3
PIS:	1	1
Perda :	328	-
Custo Variável :	25	30
Custo Fixo :	-	-
Total =>	363	37
Valor Bruto Resultado =>	(229)	53
Valor Líquido Resultado =>	(207)	49

Pela análise de rentabilidade simulada acima, a contribuição bruta dos clientes bons, ou melhor dizendo o ganho dos bons, seria igual a R\$ 53,00, enquanto que a perda bruta atribuída aos maus clientes seria R\$ 229,00. Desta maneira podemos observar que precisaríamos de 4,32 clientes bons para cobrir a perda gerada pelos maus.

VIII) 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., 1996.

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

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

Giambiagi, Fábio e Mesquita Moreira, Maurício, A Economia Brasileira nos Anos 90, BNDES, Rio de Janeiro, 1999.

Mansur Levy, Paulo e M. D. Hahn, Leda, A Economia Brasileira em Transição: O Período 1993/96, IPEA, Rio de Janeiro, 1997.

Consumer Finance Report, Lloyds TSB Group PCC, 1998.

Departamento de Estudos e Pesquisas - DEPEP, Juros e Spread Bancário no Brasil, Banco Central do Brasil, 1999.

Base de dados:

Financeira Lecca

MCM Consultores

Páginas na Internet:

www.bcb.gov.br

www.zonafinanceira.com

www.latinfocus.com

www.acsp.com