

**Anais do  
VII Seminário Multidisciplinar ENIAC Pesquisa 2015  
VII Encontro Da Engenharia Do Conhecimento Eniac  
VII Encontro De Iniciação Científica Eniac  
VII Fábrica de Artigos**

# **A PROBABILIDADE DE INADIMPLÊNCIA EM CRÉDITO DE PESSOAS JURÍDICAS COM BASE EM DADOS DE BALANÇO**

---

**Marco Antonio Ferreira Lima**

Marco Antonio Ferreira Lima é Prof. Me. Faculdades Integradas Campos Salles,  
prof\_rmarcolima@yahoo.com.br.

---

## RESUMO

Este artigo apresenta um modelo para mensurar a probabilidade de inadimplência em crédito constituído por pessoas jurídicas. Em verdade trata-se de um exercício aplicativo para o cálculo da probabilidade de inadimplência (PD), que vem a ser uma das componentes da perda esperada, presente no Método Interno Próprio (IRB), discutido no segundo Acordo da Basileia, proposto pelo Comitê de Supervisão Bancária da Basileia. A condução da pesquisa contou com uma amostra de 500 empresas de diversos ramos de atividades, do segmento varejo de um banco nacional. Acompanhou-se o desempenho delas nos empréstimos contratados durante o ano de 2002 e definiu-se como inadimplência em crédito, aquela empresa que apresentasse atraso em suas obrigações, superior a 60 dias. O modelo desenvolvido é função de indicadores contábeis, obtidos dos demonstrativos financeiros disponíveis no cadastro do banco. O objetivo do trabalho foi verificar a capacidade preditiva das variáveis contábeis na explicação da inadimplência em crédito.

**Palavras chave:** Probabilidade de inadimplência, índices contábeis, regressão logística.

## ABSTRACT

This article presents the results of a model dedicated to measuring small companies' probability of default. It is an application of one of the components of the concept of expected loss, this method the Internal Self (IRB), discussed in the second Basel Accord, proposed by the Banking Supervision Committee of Basel. A sample comprising 500 small companies from several economic sectors was utilized in the research (clients of a Brazilian bank). All loans contracted during the year 2002 were monitored and those past due more than 60 days were considered defaulted. The model was developed to estimate the small companies' probability of default based on financial ratios, available from the bank's data. The goal of this report was to verify the prediction power of those ratios in estimating the probability of default.

**Keywords:** Probability of default, financial ratios, logistic regression.

## INTRODUÇÃO

Esta pesquisa apresenta um modelo para mensurar a inadimplência em crédito, e considera sua ocorrência quando o cliente atrasa mais de 60 dias. Trata-se assim de um exercício preliminar para o cálculo de inadimplência (PD), sendo, entretanto, mais conservador, quando se compara com o prazo estabelecido pelo BIS para o conceito de inadimplência.

O objetivo do estudo tem o objetivo de verificar se os dados contábeis quando tratados como índices são variáveis suficientemente robustas para serem empregadas em modelos quantitativos destinados a mensurar o risco de inadimplência no crédito. Trata-se de quantificar uma das componentes do risco de crédito, conforme o modelo IRB, em condições extremas, pois as variáveis explicativas limitam-se em dados provenientes de demonstrativos contábeis.

O problema de pesquisa - demonstrações contábeis de uma empresa são fontes de informações para as instituições financeiras analisarem a capacidade da empresa em tomar o crédito. O processo de avaliação utiliza-se da teoria de análise de balanços para justificar as expectativas do banco. Os indicadores econômicos financeiros obtidos na análise de balanço constituem o "*corpus*" no modelo quantitativo a ser empregado.

O estudo sugere que métodos quantitativos proporcionam uma padronização nos procedimentos de análise de risco de inadimplência, não havendo então, a possibilidade da situação descrita por Assaf Neto (2002, p.49): dois analistas ao empregarem as mesmas informações e técnica de análise podem obter resultados divergentes sobre uma mesma empresa. Não se trata de contestar o autor, mas considerar que a análise de risco de inadimplência passa a ser resultado de um modelo quantitativo, embasado em métodos estatísticos ou econométricos, portanto, não dependente do grau de experiência do analista.

Assim o problema de pesquisa

relaciona-se com o tratamento dos dados contábeis e pergunta-se: como construir um modelo para mensurar o risco de inadimplência com base em informação contábil, de tal sorte que o método proposto, possibilite meios para avaliar o risco de crédito?

A metodologia resposta aos objetivos traçados faz deste artigo, um exercício empírico. Desenvolve-se um modelo para mensurar o risco de inadimplência baseado em informações contábeis com base nos arquivos fornecidos por uma instituição bancária. Seleciona-se uma amostra de empresas, que apresentaram demonstrativos contábeis e eram solicitantes e usuárias de crédito no período determinado de tempo. A partir do estudo dos dados constrói-se o modelo destinado a avaliar o risco de inadimplência.

O referencial teórico metodológico deste trabalho fundamenta-se nas classificações conduzidas por Abramo (1979, p.34-44): Segundo a utilização de resultados: Pesquisa aplicada: tipo de estudo empregado para que os resultados tenham aplicação imediata na solução de problemas concretos. É a obtenção do conhecimento pela transformação da realidade. O resultado do estudo tem aplicação no mercado bancário e será uma tentativa de contribuir para o processo de mensuração do risco inadimplência bem como oferecer novas vertentes para outras análises. Segundo os processos de estudo: Estrutural: análise sistemática da forma, do funcionamento, dos elementos e das suas inter-relações. Estatístico: exame da frequência e da variabilidade da incidência de certos atributos. O processo de estudo leva em conta a análise sistemática das informações contábeis, seu

funcionamento e sua relação com a avaliação do crédito. A natureza estatística verifica-se no evento aleatório que é a capacidade de pagamento do crédito concedido. Segundo o grau de generalização: Por amostragem: subconjunto de um conjunto maior de dados.

Os resultados dos dados colhidos são generalizados para o todo, obedecendo a certas condições e procedimentos estatísticos. O grau de generalização do estudo emprega a Teoria de Amostragem. Seleciona-se uma amostra dos balanços das empresas solicitantes de crédito junto a uma instituição bancária. Segundo os métodos de análise: Construção de modelos: reconstruir a realidade e construir não uma representação ideal, mas uma situação, um evento, um fenômeno, entre outros. O método de análise emprega métodos quantitativos tanto para identificar um comportamento, quanto para estabelecer um modelo. No caso, um modelo de avaliação do risco de inadimplência de empresas.

## 1. A PROBABILIDADE DE INADIMPLÊNCIA EM CRÉDITO

A probabilidade de inadimplência em crédito de pessoas jurídicas com base em dados de balanço trabalha com o comitê da basileia sobre fiscalização bancária, no chamado basileia ii consiste de três partes:

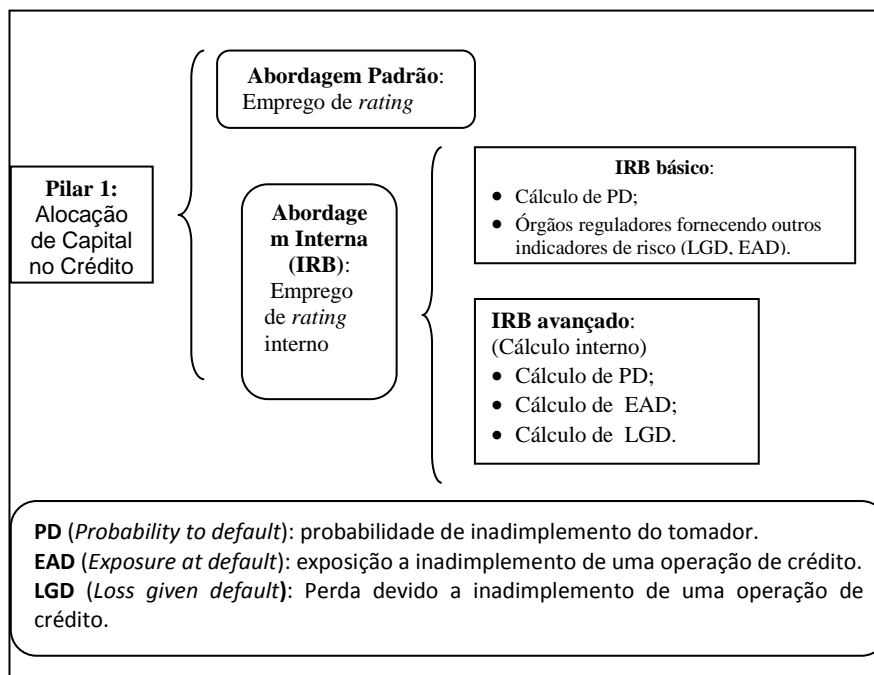
- i. Pilar 1: Necessidades mínimas de capital;
- ii. Pilar 2: Processo de exame da fiscalização;
- iii. Pilar 3: Disciplina de mercado.

Na realidade os documentos cobrem sete áreas, a saber:

- i. Método Padronizado do Risco de Crédito;
- ii. O Método Interno de Classificação do Risco de Crédito;
- iii. Securitização de Ativos;
- iv. Risco Operacional;
- v. Pilar 2: Processo de Exame da Fiscalização;
- vi. Administração e Fiscalização do Risco de Taxas de Juros nos Ativos Bancários;
- vii. Pilar 3: Disciplina no Mercado.

O Pilar 1, ao contemplar o risco de crédito, fornece a abordagem padrão e a baseada em classificação interna. A Figura 1 sintetiza a proposta:

Figura 1 – Alocação de capital no crédito, segundo o tipo de abordagem, no novo acordo da Basileia.



Fonte: adaptação própria baseada nos relatórios “ Visão geral do novo Acordo de Capital da Basileia “ e “ O método Baseado em Classificações Internas”. (COMITÊ DE SUPERVISÃO BANCÁRIA DA BASILÉIA, 2001).

de pelo menos um dos eventos:

Segundo Fortuna (2002, p.551), as instituições financeiras sendo avaliadas por agências especializadas em *rating*, ao aderirem ao Basileia II, estarão potencializando ganhos com a provável redução de alocação de capital. Assim, os bancos deverão desenvolver metodologias internas de classificação de *rating* para concessão de crédito ou na falta deste, solicitar *rating* externo das empresas clientes de crédito.

Observa-se na Figura 1, que a alocação de capital sob a ótica da abordagem interna, método avançado, exige o cálculo de três componentes, PD, EAD e LGD. O conceito de PD. Segundo o relatório “*Consultative document overview of the new Basel capital Accord: Basel committee on banking supervision*”; (Bank Of International Settlements, 2003, p.80) associa a ocorrência

- i.O banco considera que é improvável que o devedor pague suas obrigações de crédito ao grupo bancário como um todo, sem que o banco utilize recursos tais como execução de garantias (no caso de havê-las).
- ii.O devedor está inadimplente por mais que 90 dias em qualquer obrigação de crédito material junto ao grupo bancário. Saques a descoberto serão considerados como inadimplentes se o cliente ultrapassar limites estabelecidos ou tenha sofrido diminuição do limite para abaixo de sua posição corrente.

## 2. MODELO PARA CALCULAR A PROBABILIDADE DE INADIMPLÊNCIA

O método empregado para o cálculo da probabilidade de inadimplência é a Regressão Logística Múltipla, o que segundo Rosa (2000, p.18), trata-se de técnica usualmente empregada por profissionais da área financeira, pois:

- i. Tem forte poder discriminante e não exige suposições fortes como na Análise Discriminante (variáveis independentes com distribuição Normal Multivariada);
- ii. Os softwares estatísticos mais usuais apresentam o algoritmo;

Em um modelo de Regressão Logístico, a variável resposta ou variável dependente  $Y$ , assume valores dicotômicos:

$$Y = \begin{cases} 1, & \text{se contém o atributo;} \\ 0, & \text{em caso contrário.} \end{cases} \quad (4.1)$$

Nestas condições, a exemplo do modelo de regressão linear, a esperança condicional de  $Y$  dado um valor  $x$ , é expressa por  $E(Y | x)$ , (leia-se: esperança de  $Y$  dado o valor  $x$ ), onde  $x$  é a  $=0$  (não contém o atributo).

Para simplificar a notação, faça-se  $E(Y | x) = \pi(x)$ , o que representa a probabilidade de ocorrer o evento  $\{y=1\}$  dado uma relação com a variável  $x$ . Por exemplo, definindo que  $Y=1$  é o evento  $\{\text{o cliente é bom para crédito}\}$  e

este atributo é função da variável  $X = \{\text{Ativo Circulante sobre Passivo Circulante}\}$ , está se afirmando, no caso, que há uma relação entre ser bom cliente para crédito, com o indicador de liquidez corrente.

A versão mais simples do modelo de Regressão Logística, segundo Hosmer; Lemeshow (2000, p.1-9), com uma única variável independente, é expressa por:

$$\pi(X) = \frac{e^{\beta_0 + \beta_1 X_1}}{1 + e^{\beta_0 + \beta_1 X_1}} \quad (4.2)$$

Como na regressão linear simples, a função (4.2) apresenta um termo aleatório  $\varepsilon$  (resíduo, ou erro). Mais especificamente,  $E(Y | x) = \pi(x) + \varepsilon$ .

O problema é estimar os parâmetros desconhecidos  $\beta_0$  e  $\beta_1$ , tal que, em função da variável independente  $X$ , explique, ou represente bem a variável  $Y$ . Se não houver relação entre  $Y$  e  $X$ , então outras variáveis não presentes no modelo, mas representadas pela variável  $\varepsilon$  (resíduo, ou erro) produzem um efeito maior do que a variável  $X$ .

Para estimar os parâmetros  $(\beta_0, \beta_1)$  e sem perda de generalidades, havendo uma amostra de " $n$ " observações independentes dos pares  $(x_i, y_i)$ , sendo que cada  $y_i$ , é uma variável dicotômica como definido em (4.1), um conveniente método, é a função verossimilhança [1] para os pares  $(x_i, y_i)$ . O princípio de máxima verossimilhança objetiva estimar os valores de  $\beta$  tal que maximize a expressão (4.3). Para tanto, é mais fácil, matematicamente, empregar o logaritmo da função de verossimilhança:

$$L(\beta) = \ln[l(\beta)] = \sum_i^n \{y_i \ln[\pi(X)] + (1 - y_i) \ln[1 - \pi(X)]\} \quad (4.3)$$

Para achar o valor de  $\beta$  que maximiza  $L(\beta)$ , deriva-se parcialmente (4.3) em relação a  $\beta_0$  e a  $\beta_1$  e iguala-se a zero, criando assim, as equações verossimilhança:

As expressões (4.4) e (4.5) são não lineares e suas soluções exigem métodos especiais. Programas estatísticos próprios dão as soluções, por sua vez, segundo, McCullang; Nelder (1989) apud Hosmer; Lemeshow (2000, p.9), é possível encontrar suas soluções.

$$\sum_i^n [y_i - \pi(x_i)] \quad (4.4)$$

$$\sum_i^n x_i [y_i - \pi(x_i)] \quad (4.5)$$

A expressão do modelo a ser empregado de fato segue-se em (4.6). Trata-se de uma Regressão Logística Multinomial, por ter uma quantidade  $k$ , de variáveis independentes. A variável dicotômica  $Y$  conforme definido em (2.1), é função de um conjunto de variáveis independentes  $X_i$ ,  $i=1,2,\dots,k$ .

$$\pi(X) = \frac{e^{\beta_0 + \sum_{i=1}^n \beta_i X_i}}{1 + e^{\beta_0 + \sum_{i=1}^n \beta_i X_i}} \quad (4.6)$$

Deseja-se verificar se o atributo dado a variável  $Y$  (classificação a priori), é função de  $k$  variáveis independentes  $x_k$ . Para tanto é necessário estimar os coeficientes  $\beta_i$ ,  $i=1,2,\dots,k$ , da equação (2.6) através da maximização da função de verossimilhança em relação aos  $k+1$  elementos do vetor  $\beta$ . A estimativa dos coeficientes  $\beta$  e o cálculo do

predito  $Y$ , contou com o software estatístico *Statistical Package Social Science* (SPSS), versão 11.0.

### 3. LEVANTAMENTO DOS DADOS E RESULTADOS

Com acesso a base de dados de uma instituição bancária, daqui para frente, denominada Banco Gama, a coleta de dados, exigiu pesquisar dois sistemas e respectivos arquivos de dados:

- i. Arquivo 1: dados cadastrais, constituindo por contas constantes em balanços sintéticos, acessado no Banco Gama;
- ii. Arquivo 2: dados de risco, com informações quanto ao tipo de empréstimo que o cliente tomou, data de início e condições em que o crédito se encontra: em dia ou quando em atraso, por faixa de dias de atraso.

O plano de amostragem constituiu-se de três etapas:

- i. Levantamento aleatório no arquivo 1, de clientes que apresentassem o último balanço, com data de dezembro de 2001 e que possuísem um ou mais períodos de balanço anual, imediatamente anterior. O

objetivo foi criar um arquivo contendo empresas com no mínimo dois balanços anuais, para construção dos índices contábeis. Estes índices seriam as variáveis independentes no modelo da Regressão Logística;

ii. Selecionados os clientes, verificar quais se encontram no arquivo 2, durante o ano de 2002. A intenção nesta fase é determinar a variável dependente Y, ao classificar a *priori*, o cliente adimplente e o cliente inadimplente. Uma vez desenvolvido o modelo, é possível calcular a probabilidade de

adimplência para o período de um ano;

iii. Selecionar somente clientes tomadores de empréstimos de curto prazo, destinado ao giro de capital da empresa: capital de giro, conta garantida e cheque especial empresarial.

O resultado da seleção, extraindo duplicidades de clientes<sup>[2]</sup>, resultou em 500<sup>[3]</sup> empresas de diversos ramos de atividades. Os clientes classificados na amostragem como adimplentes, foram os que apresentaram durante o ano de 2002, no máximo 60 dias de atraso, em caso contrário eram considerados como inadimplentes; segue-se assim, a distribuição de frequência resultante:

Tabela 1- Classificação dos clientes pesquisados, segundo a situação no crédito

Tipo de cliente	Freqüência	%
Adimplentes	442	88,4
Inadimplentes	58	11,6
Total	500	100,0

Fonte: construção própria com base na associação entre os arquivos 1 e 2, extraindo duplicidades de clientes. São Paulo, 2016.

### 3.1. INDICADORES CONTÁBEIS

O estudo dos indicadores contábeis concentrou-se nas obras Estrutura e Análise de Balanço, Assaf Neto (2002) e Análise Financeira de Balanços, Matarazzo (1998), pois, naquilo que interessa: estudar e analisar os índices contábeis passíveis de fornecerem

informações quanto à capacidade de uma empresa em honrar seus compromissos de créditos, ambos os autores tratam o assunto com profundidade e, além disto, sendo conhecedores da legislação e prática contábil brasileira, enfocam o tema à nossa realidade. O Quadro 1 resume os índices avaliados:

Quadro 1 - Interpretação dos indicadores contábeis

Indicador	Fórmula	Interpretação
1- ENDIVID Participação de Capital de Terceiros	$\frac{\text{Capital de Terceiros}}{PL} \times 100$	Quanto à empresa tomou de Capitais de Terceiros para \$ 1 de Capital Próprio (Quanto menor, melhor)
2- PCCT Composição do Endividamento	$\frac{\text{Passivo Circulante}}{\text{Capital de Terceiros}} \times 100$	Qual a % de obrigações no curto prazo em relação às obrigações totais (Quanto menor, melhor)
3- IMOBCP Imobilização dos Recursos não Correntes	$\frac{\text{Ativo Permanente}}{PL + \text{Exigível a Longo Prazo}} \times 100$	Que % dos Recursos não Correntes (PL+ Exigível Longo Prazo) foi destinado ao Ativo Permanente (Quanto menor, melhor)
4- INDPFIN Independência Financeira	$\frac{PL}{\text{Ativo Total}}$	Qual o nível de independência da empresa em relação ao capital de terceiros (Quanto maior, melhor)
5- LIQCORR Liquidez corrente	$\frac{\text{Ativo Circulante}}{\text{Passivo Circulante}}$	Quanto a empresa possui de Ativo Circulante para cada \$ 1 de dívida total (Quanto maior, melhor)
6- LIQSEC Liquidez Seca	$\frac{\text{Ativo Circulante} - \text{Estoque} - \text{Despesas Antecipadas}}{\text{Passivo Circulante}}$	Quanto a empresa possui de Ativo Líquido para cada \$ 1 de Passivo Circulante (Quanto maior, melhor)
7- LIQGER Liquidez Geral	$\frac{\text{Ativo Circulante} + \text{Realizável a Longo Prazo}}{\text{Passivo Circulante} + \text{Exigível a Longo Prazo}}$	Quanto a empresa possui de Ativo Circulante + Realizável à Longo Prazo para a \$ 1 de dívida total (Quanto maior, melhor)
8- VAT Giro do ativo	$\frac{\text{Vendas Líquidas}}{\text{Ativo}}$	Quanto a empresa vendeu para cada \$ 1 de investimento total (Quanto maior, melhor)
9- LLV Margem líquida	$\frac{\text{Lucro Líquido}}{\text{Vendas Líquidas}} \times 100$	Quanto a empresa obtém de lucro para cada \$ 1 vendido (Quanto maior, melhor)
10- LLAT Rentabilidade do ativo	$\frac{\text{Lucro Líquido}}{\text{Ativo}} \times 100$	Quanto a empresa obtém de lucro para cada \$ 1 de investimento total (Quanto maior, melhor)

11-GAO Alavancagem Operacional	$\frac{\text{Variação no Lucro Operacional}}{\text{Variação no volume de Atividade}}$	Quantos % a empresa obtêm de lucro para cada 1 % de aumento nas vendas. É determinado pela estrutura custo da empresa, apresentando maior capacidade de alavancar o lucro, a empresa que apresentar maior custo fixo em relação ao custo total.
12- GAF Alavancagem Financeira	$\frac{\text{Variação Percentual no Lucro Líquido}}{\text{Variação Percentual no Lucro Operacional}}$	Determina a capacidade que recursos de terceiros apresentam ao elevar os resultados líquidos da empresa
13- GAT Alavancagem Total	GAO x GAF	Se o GAT= x%, então para cada 1 % de aumento no volume de venda, há uma elevação de x% no lucro líquido. Se GAT > 1,0 há indicação que qualquer variação no volume de atividade provoca mudanças mais que proporcional nos resultados líquidos da empresa.

Fonte: adaptação própria com base em Matarazzo (1998, P.158) e Assaf Neto (2002)

O cálculo dos indicadores contábeis segundo os critérios definidos no Quadro 1 tem como fonte os balanços sintéticos fornecidos pelo Banco Gama, relativo a 500

empresas aleatoriamente selecionadas. Como resultado da amostra, a distribuição dos períodos encontra-se na Tabela 2:

Tabela 2 - Número de empresas segundo o ano de publicação do balanço

Anos	Empresas	%
1999 e 2000 e 2001	307	61,4
2000 e 2001	193	38,6
Total	500	100,0

Fonte: amostra coletada no Banco Gama. São Paulo, 2015.

Neste ponto, ajustou-se a teoria apresentada no Quadro 1 com a realidade, ao

definir as variáveis independentes do modelo de Regressão Logística. O indicador ENDIVID: Participação de Capital de Terceiros foi redefinido. O denominador da equação, o Patrimônio Líquido pode assumir valores negativos. Em um processo de análise julgamental, o analista de crédito, julga saber avaliar seu impacto sobre valor da dívida do cliente, no entanto, em um modelo, há necessidade de desenvolver um algoritmo que

oriente a decisão a ser tomada, caso não se deseje que o resultado seja válido somente quando o Patrimônio Líquido assuma valores positivos. Como consequência, optou-se pelo indicador:

$$Divid = \left( \frac{PL - (Financiamentos + Instituição Financeira)}{Ativo Total} \right)$$

Sua interpretação é “quanto maior, melhor”. Pela mesma razão, (PL assumir valor negativo), não utilizamos como variável independente o indicador Imobilização do Patrimônio Líquido.

Ainda como parte do tratamento das variáveis independentes, os anos de publicação do balanço, 1999, 2000 e 2001 foram redefinidos como:

- i. Período 1: Balanços de 1999;
- ii. Período 2: Balanços de 2000;
- iii. Período 3: Balanços de

2001.

Desta forma, cada indicador foi calculado para cada período. Calculou-se também, a variação percentual entre os índices nos períodos 2 com 1 e 3 com 2. O Quadro 2 apresenta o resultado das relações:

Quadro 2 - Períodos avaliados pelos indicadores definidos

Indicadores	Períodos	Indicadores	Períodos
1- DIVID: Participação de Capital de Terceiros	1, 2, 3, 2/1 e 3/2	9- VAT: Giro do ativo	1, 2, 3, 2/1 e 3/2
2- PCCT: Composição do Endividamento	1, 2, 3, 2/1 e 3/2	10- LLV : Margem líquida	1, 2, 3, 2/1 e 3/2
4- IMOBPC: Imobilização dos Recursos não Correntes	1, 2, 3, 2/1 e 3/2	11- LLAT : Rentabilidade do ativo	1, 2, 3, 2/1 e 3/2

5- INDPFIN: Independência Financeira	1, 2, 3, 2/1 e 3/2	12-GAO : Alavancagem Operacional	2/1 e 3/2	
6- LIQCORR Liquidez corrente	1, 2, 3, 2/1 e 3/2	13- GAF Alavancagem Financeira	2/1 e 3/2	
7- LIQSEC :Liquidez Seca	1, 2, 3, 2/1 e 3/2	14- GAT Alavancagem Total	2/1 e 3/2	
8- LIQGER : Liquidez Geral	1, 2, 3, 2/1 e 3/2			

Observa-se ainda que, os valores dos índices são contínuos e embora a Regressão Logística permita o seu cálculo, ela trabalha melhor com valores discretos, ou categorizados<sup>[4]</sup>. Assim, como procedimento para categorização, calculou-se a estatística decíl das respectivas distribuições. O valor do índice contábil referente ao  $i$ -ésimo decíl,  $i=1,$

$2, \dots, 9$ ; passou a ser o limite superior da classe da distribuição de frequência. No caso dos indicadores de ordem decrescente, após o cálculo do decíl, reordenamos a posição do valor associado ao decil de ordem  $i$ ,  $i=1, 2, \dots, 9$ . A Tabela 3 apresenta os resultados do método:

Código		Decil										
Original	Recodificado	Total	Missing	10	20	30	40	50	60	70	80	90
LIQCOR1	liqcor1	481	19	0,574	0,837	1,067	1,265	1,530	1,802	2,214	3,362	6,151
LIQCOR2	liqcor2	499	1	0,509	0,784	1,045	1,235	1,421	1,731	2,150	3,154	5,763
LIQCOR3	liqcor3	499	1	0,543	0,857	1,045	1,250	1,441	1,654	2,034	2,819	4,720
LIQCOR21	liqcor21	481	19	-0,057	0,311	0,508	0,688	0,837	1,005	1,210	1,672	2,710
LIQCOR32	liqcor32	499	1	-0,056	0,332	0,548	0,718	0,882	1,012	1,186	1,499	2,946
LIQSEC1	liqsec1	481	19	0,182	0,383	0,567	0,699	0,881	1,119	1,508	2,278	4,195
LIQSEC2	liqsec2	499	1	0,157	0,350	0,509	0,678	0,844	1,114	1,377	2,078	3,783
LIQSEC3	liqsec3	499	1	0,244	0,403	0,534	0,676	0,848	1,125	1,373	1,729	3,800
LIQSEC21	liqsec21	481	19	-0,190	0,285	0,524	0,746	0,898	1,002	1,134	1,576	2,870
LIQSEC32	liqsec32	499	1	-0,416	0,244	0,556	0,745	0,876	0,998	1,116	1,469	2,567
LIQGER1	liqger1	481	19	0,464	0,697	0,873	1,009	1,163	1,438	1,821	2,657	5,133
LIQGER2	liqger2	499	1	0,384	0,613	0,837	0,992	1,157	1,421	1,700	2,233	3,882
LIQGER3	liqger3	499	1	0,434	0,632	0,860	1,017	1,160	1,358	1,650	2,126	3,465
LIQGER21	liqger21	481	19	-0,082	0,339	0,573	0,721	0,917	1,020	1,194	1,559	2,624
LIQGER32	liqger32	499	1	-0,009	0,381	0,591	0,822	0,925	1,030	1,145	1,321	2,049
IMOBBCP1	imbrnc1	481	19	0,028	0,118	0,231	0,363	0,527	0,674	0,805	0,955	1,228
IMOBBCP2	imbrnc2	500	0	0,029	0,134	0,236	0,380	0,533	0,659	0,832	1,002	1,275
IMOBBCP3	imbrnc3	500	0	0,041	0,130	0,253	0,392	0,543	0,675	0,826	0,996	1,249
ICP21	imbrnc21	472	28	-2,136	-0,460	-0,048	0,033	0,260	0,658	1,057	1,817	5,225
ICP32	imbrnc32	499	1	-2,558	-0,492	-0,034	0,000	0,164	0,395	0,752	1,317	2,612
INDEFIN1	indpfn1	481	19	0,056	0,179	0,280	0,376	0,467	0,575	0,670	0,755	0,876
INDEFIN2	indpfn2	500	0	0,039	0,164	0,267	0,365	0,449	0,537	0,623	0,737	0,845
INDEFIN3	indpfn3	500	0	0,053	0,162	0,269	0,366	0,436	0,525	0,614	0,714	0,818
INDPFI21	indpfn21	481	19	-3,486	-0,589	0,001	0,209	0,506	0,777	1,139	1,884	4,671
INDPFI32	indpfn32	500	0	-3,652	-0,639	0,011	0,352	0,715	1,056	1,363	2,423	6,520
divid1	divd1	481	19	-0,197	0,011	0,131	0,241	0,373	0,478	0,621	0,728	0,871
divid2	divd2	500	0	-0,192	-0,027	0,088	0,200	0,334	0,426	0,564	0,676	0,827
divid3	divd3	500	0	-0,243	-0,030	0,085	0,186	0,310	0,416	0,544	0,655	0,795
divi21	divd21	481	19	-1,003	-0,511	-0,317	-0,180	-0,090	-0,020	0,079	0,276	0,765
divi32	divd32	500	0	-1,135	-0,488	-0,274	-0,161	-0,057	0,005	0,086	0,259	0,874
V_AT1	giratv1	481	19	0,587	0,955	1,236	1,509	1,790	2,194	2,706	3,597	5,226
V_AT2	giratv2	500	0	0,532	0,882	1,287	1,573	1,841	2,192	2,689	3,722	5,473
V_AT3	giratv3	500	0	0,298	0,591	0,810	1,081	1,353	1,612	2,010	2,831	4,480
V_AT21	giratv21	483	17	-0,364	-0,229	-0,136	-0,079	-0,026	0,047	0,154	0,318	0,603
V_AT32	giratv32	498	2	-0,724	-0,575	-0,475	-0,332	-0,205	-0,095	0,006	0,136	0,346
LL_V1	mgll1	481	19	-0,059	-0,014	0,002	0,010	0,020	0,037	0,060	0,104	0,201
LL_V2	mgll2	498	2	-0,056	-0,010	0,004	0,011	0,020	0,033	0,056	0,090	0,188
LL_V3	mgll3	499	1	-0,081	-0,014	0,004	0,012	0,022	0,034	0,060	0,114	0,224
LL_V21	mgll21	483	17	-0,364	-0,229	-0,136	-0,079	-0,026	0,047	0,154	0,318	0,603
LL_V32	mgll32	483	17	-0,364	-0,229	-0,136	-0,079	-0,026	0,047	0,154	0,318	0,603
LL_AT1	renatv1	481	19	-0,104	-0,028	0,002	0,019	0,040	0,069	0,112	0,197	0,400
LL_AT2	renatv2	500	0	-0,077	-0,023	0,005	0,020	0,042	0,063	0,113	0,178	0,324
LL_AT3	renatv3	500	0	-0,068	-0,014	0,004	0,019	0,036	0,056	0,095	0,154	0,250
LL_AT21	renatv21	481	19	-1,886	-0,947	-0,722	-0,483	-0,264	-0,066	0,248	0,592	2,442
LL_AT32	renatv32	500	0	-2,045	-1,038	-0,778	-0,618	-0,431	-0,233	0,128	0,488	2,162
GAO21	rgao21	472	28	-23,589	-6,048	-2,418	-0,979	0,087	1,048	2,285	4,632	14,595
GAO32	rgao32	483	17	-14,370	-4,708	-1,354	0,207	0,771	1,159	1,832	3,872	12,332
GAF21	rgaf21	472	28	-0,218	0,468	0,820	0,965	1,000	1,005	1,094	1,291	1,952
GAF32	rgaf32	485	15	-0,354	0,498	0,855	0,964	1,000	1,000	1,054	1,191	1,569
GAT21	rgat21	472	28	-18,973	-5,057	-1,854	-0,458	0,485	1,380	2,822	5,832	15,325
GAT32	rgat32	483	17	-17,599	-4,979	-1,640	0,021	0,663	1,170	1,783	4,119	12,931

Tabela 3 – Índices contábeis distribuídos por decís

Nota-se que em determinados índices, houve ausência de valores, *missing values*,

fato verificado pela ausência do dado no cadastro original da empresa. A decisão foi

não retirá-la da amostragem e deixar para o final da análise, a representação do seu risco. Para melhor entender o resultado obtido na Tabela 3, exemplifica-se um caso: Seja a liquidez corrente do período 1 (licorr1).

A correspondente distribuição de decís (liqcor1) encontra-se na 1ª linha da Tabela 3. De posse desses dados, tem-se a distribuição de freqüência da variável em questão: liquidez corrente. O resultado encontra-se na Tabela 4:

TABELA 4 – CLASSES DE INTERVALOS DA LIQUIDEZ CORRENTE

Liquidez corrente	
1	Até 0,574
2	0,574 ---  0,837
3	0,837 ---  1,067
4	1,067 ---  1,265
5	1,265 ---  1,530
6	1,530 ---  1,802
7	1,802 ---  2,214
8	2,214 ---  3,362
9	3,662 ---  6,151
10	> 6,6151

Os demais índices passaram pelo mesmo processo exemplificado. Estas variáveis uma vez recodificadas são as variáveis independentes a serem testadas no modelo de Regressão Logística.

#### 4. A APLICAÇÃO DO MODELO DE REGRESSÃO LOGÍSTICA

#### MÚLTIPLA E RESULTADOS

Através do pacote estatístico SPSS, processou-se a Regressão Logística considerando como variáveis predictoras, as variáveis descritas na Tabela 3.

O modelo expresso na função (4.6) apresenta a seguinte relação:

$$\text{Cliente adimplente} = \text{função da distribuição de decís dos índices} \quad (5.2.1)$$

Destaca-se que o conceito adotado para cliente adimplente é aquele que tendo contraído um empréstimo durante o ano de 2002, quando em atraso, verificou-se no máximo por 60 dias.

Para a escolha das variáveis que melhor expliquem a relação entre ser cliente adimplente e os índices contábeis abordados, aplicou-se o método *Stepwise*, o que segundo Hosmer; Lemeshow (2000, p.116) trata-se de

um algoritmo estatístico de escolha de variáveis explicativas para o modelo segundo uma regra fixa. Optou-se então, pelo método *Forward: LR*, que se baseia no Teste da Razão de Verossimilhança<sup>[5]</sup>, com probabilidades de entrada e saída, respectivamente iguais a 0,05 e 0,10.

Segundo Makuch apud Rosa (2000, p.12), é comum no teste de desenvolvimento de modelos de Regressão Logística, empregar amostras onde o número de clientes adimplentes é igual ao de inadimplentes, pois, quando a amostra apresenta um número de clientes segundo o grupo, desproporcional, o modelo tende a discriminar melhor os que apresentam maior número de elementos.

A amostra de desenvolvimento utilizada neste estudo identifica 88% das 500 empresas selecionadas, em empresas adimplentes (ver Tabela 1), portanto, trata-se de uma amostra desproporcional. Uma solução para o problema descrito anteriormente é: se o tamanho da amostra permitir, é relevante selecionar um grupo de clientes adimplentes igual ao grupo de clientes inadimplentes e

empregá-la como base de desenvolvimento do modelo. Os que restarem, serão empregados como amostra de validação do modelo. Breiman et al. apud Rosa (2000, p.12), por sua vez, apresenta uma alternativa: gerar um número  $k$  de amostras particionadas de mesmo tamanho, e escolher  $k-1$  amostras como amostra para testar o modelo e o restante, para validar. Neste trabalho, apesar da amostragem ser estatisticamente de tamanho consistente, ou seja, 500 clientes; empregar a metodologia sugerida exigiria um *corpus* ainda maior. Desta forma, trabalhou-se com a real proporção de clientes considerados adimplentes e inadimplentes, como amostra de desenvolvimento do modelo. Para validar o resultado do modelo, procurou-se verificar se a taxa de previsão, quanto a recomendar ou não crédito, mantinha-se em períodos posteriores, no caso, nos meses de janeiro e fevereiro de 2003.

#### 4.1. Resultados obtidos

Sejam as medidas:

$$EB_t = \frac{(Total\ de\ bons\ recomendados)}{Total\ de\ bons} \times 100$$

$$EF_t = \frac{(Total\ de\ bons\ recomendados + Total\ de\ maus\ não\ recomendados)}{Total\ geral} \times 100$$

$$EM_t = \frac{(Total\ de\ maus\ não\ recomendados)}{Total\ de\ maus} \times 100$$

As probabilidades  $\pi(x)$  calculadas no modelo expresso em (4.6) no ponto de corte

0,88<sup>[6]</sup>, conforme definido em EF<sub>t</sub>, EB<sub>t</sub> e EM<sub>t</sub> estão dispostas na Tabela 5:

Tabela 5 - Percentagem de classificação correta

Passos do Stepwise	EF <sub>t</sub>	EB <sub>t</sub>	EM <sub>t</sub>
1	70,4%	72,8 %	52,7%
2	68,2%	67,3%	74,5 %
3	70,6%	70,0 %	74,5 %

A eficiência total do modelo, nas três simulações, no ponto de corte 0,88 é próxima a 70 %, indicando uma estabilidade no modelo. Neste estudo, optou-se pelos

resultados do passo três, porém qualquer outra decisão poderia ser tomada.

O modelo de Regressão Logística, no passo 3 é expresso na equação (5.2.2):

$$\pi(X) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3}} \quad (5.2.2)$$

Onde:  $\pi(x)$  é a probabilidade do i-ésimo cliente ser adimplente,  $i=1, 2, \dots, 500$ .

As variáveis selecionadas pelo modelo são:

i.X<sub>1</sub>: liquidez corrente no período 1 (anos 1999 e 2000);

ii.X<sub>2</sub>: Variação no giro do ativo entre os períodos 2 e 1 (variação entre os anos 2001 e 2000);

iii.X<sub>3</sub>: independência financeira no período 3 (anos 2000 e 2001).

Destaca-se ainda, que a probabilidade de inadimplência é verificada ao se calcular  $1 - \pi(x)$ .

A equação (5.2.2) é obtida dos resultados descritos no Quadro 3, onde apresenta as variáveis independentes escolhidas no passo 3, as respectivas estimativas dos parâmetros  $\beta$  (B), a Estatística de Wald, e Odds (exp(B)):

Quadro 3 - Variáveis selecionadas, correspondentes estimadores, estatística de wald e odds

Variáveis	B	Wald	Exp(B)
Liquidez corrente			
(LIQCOR1)		19,68	
Até 0,5736	1,942	4,691	6,974
0,5736 --  0,8369	1,302	2,631	3,676
0,8369 --  1,0667	1,028	1,764	2,797
1,0667 --  1,2648	1,753	3,983	5,771
1,2648 --  1,5300	-0,420	0,365	0,657
1,5300 --  1,8015	0,732	0,858	2,080
1,8015 --  2,2136	0,170	0,056	1,186
2,2136 --  3,3618	0,489	0,418	1,631
3,3618 --  6,1514	2,282	3,981	9,793
Maior que 6,1514	0,000		1,000
Variação no giro de			
ativos (V_AT21)		21,07	
Até -0,3641	-0,170	0,051	0,844
-0,3641 --  -0,2292	-0,876	1,477	0,417
-0,2292 --  -0,1355	1,101	1,334	3,007
-0,1355 --  -0,0793	0,232	0,084	1,261
-0,0793 --  -0,0257	-1,567	5,327	0,209
-0,0257 --  0,0472	-0,271	0,115	0,763
0,0472 --  0,1544	-1,145	2,563	0,318
0,1544 --  0,3180	-0,166	0,047	0,847
0,3618 --  0,6027	0,875	0,874	2,399
Maior que 0,6027	0,000		1,000
Independência			
financeira (NDEFIN3)		22,58	
Até 0,0525	-1,633	4,342	0,195
0,0525 --  0,1616	0,134	0,027	1,143
0,1616 --  0,2687	-0,264	0,113	0,768
0,2687 --  0,3658	-0,489	0,183	1,399
0,3658 --  0,4360	0,336	0,397	0,613
0,4360 --  0,5248	-0,433	0,342	0,649
0,5248 --  0,6143	0,815	0,853	2,259
0,6143 --  0,7138	1,434	2,251	4,196
0,7138 --  0,8175	2,013	3,024	7,482
Maior que 0,8175	0,000		1,000
Constante	1,716	4,826	5,564

## 5. TESTE DO MODELO PROPOSTO

Para validar o modelo (5.2.2), verificou-se o grau de acerto de sua previsão em recomendar ou não crédito, com o real comportamento da empresa durante os meses de janeiro e fevereiro de 2003, quanto ao nível de atraso. O parâmetro de decisão do modelo foi o valor 0,88<sup>[7]</sup>, lembrando que esta é a proporção de clientes adimplentes na amostra de desenvolvimento.

A metodologia consiste em Análise Comparativa entre os resultados previstos com os resultados reais obtidos. O valor 0,88 é comparado com a probabilidade  $\pi(x)$  calculada em (5.2.2): se inferior a 0,88, tem o crédito recusado, em caso contrário,

aprovado. Por outro lado, com referência nos meses de janeiro e fevereiro de 2003, verificam-se quais das empresas já estudadas, na base de desenvolvimento ainda estão utilizando os créditos estudados no desenvolvimento do modelo. Em caso afirmativo, qual o número de dias de atraso que a empresa apresenta; se superior a sessenta dias, é classificada como inadimplente, a exemplo da metodologia empregada na amostra de desenvolvimento. Lembrando as medidas,  $EF_t$ ,  $EM_t$  e  $EB_t$ , definidas anteriormente e calculadas no ponto de corte 0,88, temos para os dois instantes, na base de desenvolvimento e na de validação, os resultados do teste:

Tabela 6 – Percentagem de classificações corretas, segundo o tipo de arquivo

Medida de Classificação	Arquivo de desenvolvimento (2002)	Arquivo teste (jan/fev 2003)
$EF_t$	70,6%	69,1%
$EM_t$	74,5%	75,6%
$EB_t$	70,0%	68,3%

Da amostra inicial de 500 empresas analisadas na base de desenvolvimento, 375 (75%) apresentaram-se nos meses de janeiro e fevereiro de 2003. Dois motivos podem ser atribuídos à diferença: ou as empresas liquidaram a operação de crédito inicial e não a renovaram ou liquidaram e contrataram outro tipo de empréstimo que não faz parte deste estudo.

De imediato observa-se um acréscimo na taxa  $EM_t$  e uma conseqüente queda na taxa  $EB_t$ . É prematuro, no entanto, afirmar se o fato é devido à entrada de clientes com perfil diferente ao amostrado, o que significa que o

modelo já está “envelhecendo”, ou se a queda na taxa  $EB_t$  é resultado da tendenciosidade do modelo em discriminar melhor os clientes adimplentes do que os inadimplentes, por conta da desproporcionalidade na amostra de desenvolvimento. No conjunto, no entanto, e independentemente destas considerações, o modelo ainda mantém eficiência. Se de fato a queda registrada na medida  $EB_t$  é função da tendenciosidade, uma mudança no valor do ponto de corte fornece ao modelo as taxas de discriminação desejadas, e isto pode ser obtido através da estatística KS.

## 6. CONSIDERAÇÕES FINAIS

Este trabalho sugere um método estatístico para avaliar o risco de inadimplência. O modelo em questão afirma que o cliente é adimplente ou não em função de indicadores contábeis, obtidos nos demonstrativos das empresas selecionadas na amostragem.

O resultado do modelo é bastante satisfatório estatisticamente (ver Tabela 5). Os testes de validação conduzidos confirmaram a estabilidade e resultados próximos aos verificados na amostra de desenvolvimento (ver Tabela 6). A informação contábil por si só, forneceu condições de mensurar a probabilidade do cliente em cumprir ou não com o contrato de crédito.

Como consequência, e lembrando a abordagem sugerida pelo Novo Acordo da Basileia no modelo IRB avançado, onde define que o risco de crédito é função do risco de inadimplência do cliente (PD), da exposição do empréstimo tomado (EAD) e da taxa de perda em créditos inadimplentes (LGD), o modelo expresso em (5.2.2) presta ao atendimento do cálculo de PD.

Este artigo ao definir a inadimplência para atraso superior a 60 dias exige da instituição financeira uma posição mais conservadora do que a proposta pelo Comitê, uma vez que este somente considera inadimplência, quando há ocorrência de atraso superior a 90 dias.

### NOTAS

[1] Sobre Função de Verossimilhança ver, por exemplo, De Groot (1986).

[2] ) Na ocorrência de um cliente com mais de um empréstimo válido para o estudo, selecionou-se aquele que apresentasse empréstimo em situação de atraso pior.

[3] Supondo uma proporção máxima de clientes adimplentes e inadimplentes em 50 %, e um nível de significância de 5 %, a margem de erro para uma amostra de tamanho  $n= 500$  é de  $\pm 4,38\%$ .

[4] Segundo Rosa (2000, p.14), vantagens para categorizar as variáveis são: padronizar resultados, estabilizar o modelo e transformar as variáveis.

[5] Ver por exemplo, MENDENHALL; SINCICH (1996) ou consultar SPSS (1997).

[6] O ponto de corte separa os clientes recomendados pelo modelo, daqueles não recomendados. No caso aplicou-se a mesma proporção verificada na amostra selecionada.

[7] A manutenção deste valor tem como objetivo, verificar se as condições da amostra de desenvolvimento se mantêm. O mercado financeiro ajusta o ponto de corte à política de crédito ou então, utiliza a estatística de Kolmogorov-Smirnof (KS), para escolher o valor que mais discrimina clientes adimplentes dos inadimplentes.

### REFERÊNCIAS BIBLIOGRÁFICAS E ELETRÔNICAS

ABRAMO, Perseu. Pesquisa em ciências sociais. In Sedi Hirano (coord.). Pesquisa social: projeto e planejamento. São Paulo: T. A. Queiroz, Editor, 1979. p.21-88.

ASSAF NETO, Alexandre A. *Estrutura e análise de balanços: um enfoque econômico financeiro*. 7ed. São Paulo: Atlas, 2002. 320 p.

BANK OF INTERNATIONAL SETTLEMENTS. *Consultative document overview of the new Basel Capital Accord: Basel committee on banking supervision*. Basel: abril 2003. Versão disponível em: (<http://www.bis.org>).

COMITÊ DE SUPERVISÃO BANCÁRIA DA BASILÉIA. *O método baseado em classificações internas*. Tradução: Ernest & Young, 2001. versão 1.0. Texto disponível em: <<http://www.febraban.com.br>> Acesso em: 25 jul. 2001. Versão original em inglês disponível em: (<http://www.bis.org>).

COMITÊ DE SUPERVISÃO BANCÁRIA DA BASILÉIA. *Visão geral do novo acordo de capital da Basileia*. Tradução: Ernest & Young, 2001. versão 1.0. Texto disponível em: <<http://www.febraban.com.br>> Acesso em: 25 jul. 2001. Versão original em inglês disponível em: (<http://www.bis.org>).

DE GROOT, Morris H. *Probability and statistics*. 2ed. United States of America: Addison -Wesley Publishing Company, 1975. 723p.

FORTUNA, Eduardo. *Mercado financeiro: produtos e serviços*. 15. ed. Rio de Janeiro: Qualitmark, 2002. 624p.

HOSMER, David W.; LEMESHOW, Stanley. *Applied logistic regression*. 2ed. New York: A Wiley – Interscience Publication, 2000. 373p.

MATARAZZO, Dante. C. *Análise financeira de balanços*. 5ª. ed. São Paulo: Atlas, 1998. 471 p.

MENDENHALL, William, SINCICH, Terry. *A second course in statistics: regression analysis*. 5. ed. New Jersey: Prentice-Hall, Inc. 1996. 899p.

NORTH CAROLINA STATE UNIVERSITY. *Logistic regression*. Texto disponível em: <<http://www2.chass.ncsu.edu/garson/pa765/index.html>> Acesso em 15 jun. 2003.

ROSA, Paulo T. M. *Modelos de “credit score”: Regressão Logística, Chaid e Real*. 2000. 69f. Dissertação (Mestrado em Estatística). Instituto de Matemática e Estatística da Universidade de São Paulo. São Paulo,