

# Aplicação de Modelagem *Probit* no Auxílio de Práticas Gerenciais de Marketing

Autora: *Luciana Bellini Rangel*

Resumo: Este artigo objetiva desenvolver, testar e diagnosticar um modelo *Probit*, o qual pode auxiliar no gerenciamento de marketing. O modelo proposto utiliza o programa STATA para analisar a relação causal da probabilidade da inadimplência pelo perfil do indivíduo. Com a finalidade de ajudar os processos gerenciais, esse modelo procura minimizar os riscos nas relações comerciais, auxiliando no desenvolvimento de estratégias de marketing e gestão.

Palavras-chave: Perfil do consumidor, marketing, ferramentas gerenciais.

## 1 Introdução

O cenário competitivo organizacional abrange fatores macro e microeconômicos em dimensões que ultrapassam as fronteiras das nações, de modo que, no atual ambiente de negócios, o fator sucesso determina a sobrevivência das empresas e as mantém em um mercado altamente competitivo e dinâmico. Nesse contexto, minimizar os riscos passou a ser tema de estudos no campo da gestão, com o desenvolvimento de trabalhos econométricos na busca de modelos que visam aumentar a precisão das informações, gerando melhores resultados financeiros aos negócios. Diante disso, mostrando uma modelagem capaz de minimizar os riscos operacionais, este artigo destaca a importância do desenvolvimento de estratégias de gestão e marketing, apresenta e aplica técnicas de microeconometria, bem como emprega microdados modelados por *Probit*.

O modelo apresentado neste artigo e criado com a aplicação do *Probit* envolve estimação, testagem e diagnóstico a partir da análise da relação causal da inadimplência como um fenômeno socioeconômico, com a utilização dos dados em *cross-section*. Como ferramenta básica para o desenvolvimento deste trabalho foi utilizado o programa STATA.

Tendo como fundamentação teórica para as estratégias corporativas e de gestão Mintzberg, Ahlstrand e Lampel (2000), Hitt, Ireland e Hoskisson (2008), Santos (2008), Certo, Samuel C, e Peter J. P. (2009), Barney e Hesterly (2011), além de Kotler P. e Keller (2013), o presente trabalho, baseando-se nos estudos de Glaser (1963), Boyd Jr., Westfall e Stasch (1977), Churchil (1999), Aaker, Kumar e Day (2004), Hewson (2006), Wicke e Silver (2009), Malhotra (2012), utiliza-se de dados secundários. Quanto ao modelo *Probit*, foram levadas em conta as pesquisas de Greenber (1980), Hahn E. D. e Soyer (2008), Wooldridge (2010-2011). No quarto tópico, foram apresentados os resultados obtidos, seguindo-se das considerações finais com possíveis recomendações para futuros trabalhos.

Sabe-se que a inadimplência é uma das causas de insolvência financeira que pode levar negócios à falência. Ao identificar a relação causal entre as características inerentes ao indivíduo e seu perfil adimplente ou não, o modelo econométrico desenvolvido objetiva auxiliar o desenvolvimento das estratégias de gestão e funcionais de marketing. Com isso, cria-se uma ferramenta que pode minimizar o risco de interações indesejáveis com clientes potencialmente inadimplentes, auxiliando o direcionamento das relações comerciais.

Ao promover a importância das estratégias mercadológicas, desde o nível da alta gestão, passando pelo funcional até chegar ao operacional, espera-se que este artigo sirva de base para que demais pesquisadores possam desenvolver trabalhos empíricos empregando microdados.

## 2 Fundamentação Teórica

O modelo preditivo desenvolvido neste trabalho busca auxiliar aos gestores quanto às informações básicas dos clientes, objetivando minimizar os riscos operacionais no desenvolvimento de estratégias de gestão e marketing.

Todo o processo de gestão parte da administração estratégica, cujos objetivos consistem em escolher e implementar estratégias que possam gerar vantagem competitiva aos negócios. Nesse caminho, uma ferramenta amplamente utilizada é o planejamento estratégico que, para Mintzberg, Ahlstrand e Lampel (2000), consiste em fixar objetivos, a partir da análise interna e externa do ambiente organizacional, realizando uma avaliação estratégica com o objetivo de desenvolver e operacionalizar as estratégias. Estas são definidas por Hitt, Ireland e Hoskisson (2008) como sendo “um conjunto integrado e coordenado de compromissos e ações definido para explorar competências essenciais e obter vantagem competitiva” (HITT, IRELAND E HOSKISSON, 2008, p. 4).

Quando definem uma estratégia, as empresas escolhem alternativas para competir. Assim, as estratégias direcionam o que a empresa pretende fazer, delineando a elaboração das ações. Desse modo, inovar criando formas de atendimento totalmente diferenciadas faz com que se obtenham maiores ganhos na eficiência do atendimento aos clientes (HITT, IRELAND E HOSKISSON, 2008).

As estratégias funcionais servem de base para a elaboração dos processos operacionais. Marketing, por exemplo, tem por objetivo facilitar as trocas entre organizações e seus compradores: “As organizações que visam lucros devem atrair e manter seus clientes para gerar vendas e lucros” (CERTO, SAMUEL C.; PETER, J. P., 2009, pág. 371).

Atuando de forma estratégica dentro de uma organização, o marketing funciona como uma grande central de informações mercadológicas, de forma que, por meio do conhecimento do mercado e dos clientes, auxilia nas tomadas de decisões e no processo estratégico (KOTLER, P; KELLER, 2013).

Para Santos (2008), as empresas devem buscar a fidelização de seus clientes rentáveis, o que significa medir, gerenciar e criar valor, transformando clientes satisfeitos em clientes fiéis.

Segundo Barney e Hesterly (2011), embora, não raro, seja difícil para a empresa identificar se está seguindo a melhor estratégia, é possível reduzir a probabilidade de erros no desenvolvimento da estratégia e garantir sucesso para qualquer negócio. O processo da administração estratégica pode aumentar ou não a probabilidade de uma empresa desenvolver bons negócios pelo desenvolvimento e pela implementação de táticas bem sucedidas (BARNEY E HESTERLY, 2011).

Vive-se hoje a sociedade do consumo, em que se sofrem bombardeios diários de ofertas sedutoras, o que gera desejos de consumo ilimitados diante de recursos financeiros limitados. Essa realidade gera grande disputa de mercado por parte das empresas, exigindo que as organizações minimizem o risco da inadimplência por meio do direcionamento do marketing para interações com maiores probabilidades de serem mais rentáveis.

### 3 Metodologia

Trata-se de um estudo com uma abordagem quantitativa cujo objetivo é desenvolver um modelo causal para identificar a probabilidade de um cliente ser inadimplente ou não diante de seu perfil. Para isso, trabalhos de pesquisa científica podem ser desenvolvidos por meio da utilização das fontes de dados primários ou das fontes de dados secundários.

Para o desenvolvimento do modelo econométrico proposto neste artigo, foram utilizados os dados secundários do banco de dados *Bankloan.sav* disponíveis na base do programa da IBM SPSS versão 20, que consiste num *software* de análise preditiva, tanto auxiliando na elaboração de informações para que as tomadas de decisões sejam mais

inteligentes, quanto resolvendo problemas e melhorando resultados, (<http://www-01.ibm.com/software/analytics/spss/>).

As fontes secundárias referem-se à utilização dos dados disponíveis, os quais já foram coletados anteriormente e podem ser reanalisados e transformados em informações por pesquisadores ávidos pela exploração do conhecimento (CHURCHILL, 1999; WICKE; SILVER, 2009; AAKER; KUMAR; DAY, 2004; MALHOTRA, 2012).

Com relação às definições sobre os dados secundários, cita-se também a descrição de Glaser (1963: 11): “o estudo dos problemas específicos através da análise dos dados existentes que foram originalmente coletadas para outra finalidade”. Hewson (2006) afirma que a análise de dados secundários é o aprofundamento do estudo de um conjunto existente de dados com o objetivo de abordar uma questão de pesquisa distinta da qual esses dados foram originalmente coletados, gerando novas interpretações e conclusões.

Os dados secundários são, basicamente, fontes especificadas, coletadas e registradas por pessoas que não são seus usuários atuais. Ou seja, pode ser um registro interno ou advir de fontes externas de consulta (AAKER; KUMAR; DAY, 2004). Estando disponível um volume de dados secundários, definidos pelos pesquisadores como aqueles que podem resolver o problema investigado, salva-se uma considerável quantidade de tempo e dinheiro (BOYD JR.; WESTFALL; STASCH, 1977).

Como dito, a maior vantagem da utilização dos dados secundários é a economia de tempo e de dinheiro. Pesquisadores, sozinhos ou com poucos assistentes, buscam informações de publicações compiladas e publicadas por outras pessoas. Enquanto um projeto de campo pode levar de 60 a 90 dias, ou até mais, dados secundários podem ser coletados no prazo de poucos dias (WENTZ, 1979; BOYD JR.; WESTFALL; STASCH, 1977; AAKER; KUMAR; DAY, 2004; MALHOTRA, 2012; BOYDWICKE; SILVER, 2009).

Analisando os dados disponíveis no banco de dados IBM SPSS, arquivo *bankloan.sav*, foi observado que o mesmo dispunha de um total de oitocentos e cinquenta eventos, dos quais setecentos estavam classificados como adimplentes ou inadimplentes, e cento e cinquenta demandavam classificação.

Ainda na descrição do banco utilizado, o mesmo possui oito variáveis, sendo uma categórica, e as demais, métricas:

1. Nível educacional (*ed*) → categórica
2. Idade (*age*) → métrica
3. Anos no atual emprego (*employ*) → métrica
4. Anos na atual residência (*adress*) → métrica
5. Renda (*income*) → métrica
6. Percentual da renda comprometida com dívida (*debtinc*) → métrica
7. Dívida no cartão de crédito (*creddeb*) → métrica
8. Outras dívidas (*othdebt*) → métrica

Considerando o banco de dados *bankloan.sav* satisfatório para o desenvolvimento e validação do modelo, foi utilizada a codificação da variável dependente na base de dados:

- 1 = Sim = Cliente inadimplente
- 0 = Não = Cliente adimplente

Devido ao fato de a variável dependente ser dicotômica, a princípio o modelo *Probit* foi escolhido para a definição da probabilidade de o indivíduo ser inadimplente. Após a definição das variáveis, o modelo foi testado com a utilização da técnica de mínimos quadrados ordinários, e os resultados comparados aos resultados do modelo desenvolvido pelo método *Probit*.

Para uma validação dos resultados, após o desenvolvimento do modelo *Probit* com os setecentos casos, o banco de dados inicial foi retomado, e uma amostra aleatória com 60% dos

casos originais foi criada para título de verificação. Assim, o modelo foi novamente rodado com quatrocentos e vinte casos, para verificação e comparação dos resultados alcançados.

As premissas respeitadas referem-se à independência dos casos e características das variáveis, uma vez que, segundo Wooldrige (2011), para esse tipo de modelagem, espera-se que haja pouca, ou nenhuma, multicolinearidade entre as variáveis independentes e que os casos não sejam correlacionados.

O método econométrico escolhido para este trabalho foi desenvolvido pela aplicação da modelagem *Probit*, que consiste num método de regressão não linear utilizado quando a variável resposta é qualitativa com dois resultados possíveis (WOOLDRIDGE, 2010).

Conforme Greenberg (1980), a função *Probit* teve sua primeira publicação realizada por Chester Ittner Bliss, em um artigo de 1934, no qual Bliss mostrou a transformação do percentual de mosquitos mortos por um pesticida em uma unidade de probabilidade, relacionando a dosagem do pesticida ao percentual de mortalidade de mosquitos. Com isso esperava obter uma linha mais ou menos reta, ou seja, uma distribuição normal. A partir desse trabalho, o modelo passou a ser amplamente utilizado e denominado *Probit*.

A abordagem é justificada, em particular, se a variação de resposta pode ser racionalizada como uma distribuição normal entre os indivíduos em teste, no qual a tolerância de um determinado assunto é a parcela apenas suficiente para a resposta de interesse (HAHN, E. D. SOYER, R., 2008).

Ao longo dos anos, o estudo apresentado por Bliss, em 1935, seguiu sua evolução pela dedicação de pesquisadores, sendo hoje uma técnica que pode ser amplamente aplicada em diversos campos do saber devido à evolução da computação e de sistemas com ampla capacidade para o processamento de dados complexos, como o STATA. Este é um *software* com capacidade robusta para a análise estatística de dados em um pacote integrado, dispondo ainda de grande amplitude e profundidade nas suas ferramentas analíticas, (<http://www.stata.com/>).

Wooldridge (2010; 2011) afirma que os economistas tendem a preferir a hipótese de normalidade da distribuição do erro, razão pela qual o modelo de resposta binária *Probit* é o mais popular em econometria. A hipótese de identificação relativa do modelo *Probit* é os betas em termos populacionais, necessitando-se, para isso, respeitar os seguintes pressupostos:  $E(u'x)=0$  exogeneidade estrita e  $E(x'x)=k$  rank cheio. O primeiro pressuposto se justifica para garantir a consistência do beta, de maneira que  $x$  e  $u$  sejam não correlacionados ( $x$  exógeno). O segundo pressuposto serve para garantir a invertibilidade de  $E(x'x)$ , de maneira que o beta possa ser estimado. Isso também remove a possibilidade de multicolinearidade perfeita.

Ao se comparar o modelo *Probit* com modelos de probabilidade linear, pode-se concluir que, embora o segundo seja fácil de ser estimado e usado, ele possui desvantagens com relação à probabilidade ajustada, que poderá ser menor que zero ou maior que um, e que o efeito parcial de qualquer variável explicativa é constante. Essas limitações podem ser compensadas pela utilização de modelos de respostas binárias (WOOLDRIDGE, 2011).

Para o desenvolvimento dos passos iniciais relativo à escolha das variáveis preditoras do modelo, o pesquisador fez a opção pela utilização do programa IBM SPSS, em função de o próprio pesquisador ter maior afinidade com os comandos desse software para a estimação.

Seguindo o passo a passo do desenvolvimento do trabalho, tem-se que: o modelo foi rodado com todos os casos a fim de identificar *outliers*:

- Usando a distância de Cook ( $dc$ ) → Nenhum *outlier* foi identificado (se  $dc > 1$  é *outlier*)
- Usando os resíduos padronizados ( $rp$ ) → 11 *outliers* foram identificados (se  $rp > |3|$  é *outlier*)

Como os resultados foram ligeiramente piores sem os *outliers*, foi decidido manter todos os casos. Considerando a estimação do modelo, para a seleção das variáveis, foi utilizado o método *Forward Stepwise (Likelihood Ratio)*. Outros métodos também foram

testados, como o *Backward Stepwise (Likelihood Ratio)*. De forma intuitiva foi considerada a hipótese de que salário não teria relação com o endividamento, o que foi comprovado pelo teste de hipótese. Todos os métodos de seleção resultaram na escolha das mesmas variáveis no modelo final, dando evidência de que, entre os modelos possíveis, o escolhido é o melhor.

Pela estimação dos coeficientes e verificação de multicolinearidade, foi observado que os coeficientes eram significativos e que não havia evidências de multicolinearidade no modelo, conforme tabela um.

Tabela 01 – Estimação de coeficientes – IBM SPSS.

		Variáveis da Equação						95% C.I. for EXP (B)	
		B	S.E	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 4	<i>employ</i>	-0,216	0,034	41,444	1	0,000	0,806	0,755	0,861
	<i>adress</i>	-0,071	0,024	8,605	1	0,003	0,931	0,888	0,977
	<i>debtinc</i>	0,076	0,023	10,788	1	0,001	1,079	1,031	1,130
	<i>creddebt</i>	0,535	0,109	23,916	1	0,000	1,708	1,378	2,117
	<i>_cons</i>	-0,818	0,323	6,425	1	0,011	0,442		

d. Variable(s) entered on step 4: address

Fonte: Próprio pesquisador; estimação dos coeficientes e verificação de multicolinearidade realizada pela utilização do programa IBM SPSS.

Dessa forma, o modelo testado e apresentado neste artigo é composto pelas seguintes variáveis:

1. Anos no atual emprego (*employ*) → métrica
2. Anos na atual residência (*adress*) → métrica
3. Percentual da renda comprometida com dívida (*debtinc*) → métrica
4. Dívida no cartão de crédito (*creddeb*) → métrica

Após a definição das variáveis, o modelo foi desenvolvido no programa STATA, e os resultados obtidos bem como suas respectivas interpretações estão descritos na próxima seção.

#### 4 Resultados

O desenvolvimento do modelo *Probit* foi realizado pela utilização do programa STATA, cujos resultados apresentados nesta seção iniciam-se com a análise descritiva das variáveis, conforme tabela dois:

Tabela 02 – Análise descritiva das variáveis – STATA.

Variable	Obs	Mean	Std. Dev.	Min	Max
<i>Employ</i>	700	8.388.571	6.658.039	0	31
<i>Adress</i>	700	8.278.571	6.824.877	0	34
<i>Debtinc</i>	700	1.026.057	6.827.234	.4	41.3
<i>Creddebt</i>	700	1.553.457	2.117.209	.01	20.56

Fonte: Próprio autor; desenvolvido pela utilização do programa STATA.

Conforme descrito anteriormente, o arquivo é composto por setecentos casos, com a utilização das variáveis *employ* (anos no mesmo emprego), *adress* (anos no mesmo endereço),

*debtinc* (percentual da renda comprometida com dívida), *creddebt* (dívida no cartão de crédito). Em média os indivíduos moram no mesmo endereço e trabalham no mesmo emprego por oito anos, sendo que a variável emprego possui um desvio padrão de seis e sessenta e cinco, de modo que o menor evento refere-se a zero, e o maior, a trinta e um anos. Já no quesito moradia, o desvio padrão é de seis e oitenta e dois, sendo que a menor observação também é zero, e a maior é trinta e quatro anos.

Com relação ao percentual da renda comprometida, em média pode-se afirmar que gira em torno de 10%, com um desvio padrão de seis e oitenta e dois, sendo que a menor observação é de zero quatro por cento, e a maior de quarenta por cento. Por fim, a dívida no cartão de crédito varia até cinquenta e seis reais; essa variável possui a média de um e cinquenta e cinco, e o desvio padrão de dois e onze.

O modelo desenvolvido por mínimos quadrados ordinários explicitou que os dados estão estatisticamente significantes a um por cento.

Tabela 03 – Regressão realizada por Mínimos Quadrados Ordinários – STATA.

Number of obs	=	700
F( 4, 695)	=	68.57
Prob > F	=	0.0000
R-squared	=	0.2830
Adj R-squared	=	0.2789
Root MSE	=	.37342

Source	SS	df	MS
Model	38.247.554	4	956.188.866
Residual	96.911.016	695	.139440312
Total	135.158.571	699	.193359902

Default	Coef.	Std. Err.	T	P>t	[95% Conf.	Interval]
<i>Employ</i>	-.023752	.0025039	-9.49	0.000	-.0286681	-.0188358
<i>Adress</i>	-.0072973	.0021964	-3.32	0.001	-.0116098	-.0029849
<i>Debtinc</i>	.0148191	.0025047	5.92	0.000	.0099014	.0197368
<i>Creddebt</i>	.0619107	.0088579	6.99	0.000	.0445194	.0793021
<i>_cons</i>	.2728572	.0355305	7.68	0.000	.2030973	.3426171

Fonte: Próprio autor, desenvolvido pelo programa STATA.

Foram utilizados setecentos casos, sendo as variáveis explicativas conjuntamente importantes para explicar a variável dependente. O resultado apresentado no teste considera que todos os coeficientes são estatisticamente iguais a zero. Em torno de 28% da variação das variáveis dependentes pode ser atribuído às variáveis explicativas.

Analisando o modelo pelos parâmetros estimados, todas as variáveis são estatisticamente significativas. Enquanto os valores negativos significam uma queda, os positivos significam o oposto.

Pôde ser observado que as variáveis “anos no mesmo emprego” e “anos no mesmo endereço de moradia” diminuem a probabilidade de o indivíduo ser inadimplente; já o percentual da renda comprometido com dívidas e as dívidas do cartão de crédito aumentam a probabilidade de o indivíduo ser mau pagador.

O método dos mínimos quadrados ordinários, ou (MQO) ou OLS (do inglês *Ordinary Least Squares*), consiste numa técnica que procura encontrar a melhor distância quadrática, ajustando um conjunto de dados com o objetivo de minimizar a soma dos quadrados das diferenças entre o valor estimado e os dados observados (os chamados resíduos), (WOOLDRIDGE, 2011).

O MQO, uma forma de estimação amplamente utilizada em econometria, é um estimador que minimiza a soma dos quadrados dos resíduos da regressão, de forma a maximizar o grau de ajuste do modelo aos dados observados. Os requisitos para a aplicação dessa técnica são: aleatoriedade da distribuição, normalidade e independência dos erros. Outro ponto que precisa ser observado é que o modelo precisa ser linear nos parâmetros, ou seja, as variáveis apresentam uma relação linear entre si. Caso contrário, deverá ser usado um modelo de regressão não linear (WOOLDRIDGE, 2011).

Desse modo, como técnica mais adequada ao alcance dos objetivos propostos, optou-se por aplicar o método *Probit*, cujos resultados são apresentados na tabela 4, com as respectivas observações.

Tabela 04 – *Probit* com todas as observações – STATA.

<i>Number of obs</i>	=	700
<i>LR chi2(4)</i>	=	246.41
<i>Prob &gt; chi2</i>	=	0.0000
<i>Pseudo R2</i>	=	0.3063

<i>Log likelihood</i> = -278.97522
------------------------------------

<i>default</i>	Coef.	Std. Err.	z	P>z	[95% Conf.	Interval]
<i>employ</i>	-.1351445	.0151654	-8.91	0.000	-.1648682	-.1054208
<i>adress</i>	-.0433977	.0108855	-3.99	0.000	-.0647329	-.0220626
<i>debtinc</i>	.0508484	.0106872	4.76	0.000	.0299019	.0717948
<i>creddebt</i>	.3219738	.049078	6.56	0.000	.2257828	.4181649
<i>_cons</i>	-.4942234	.1471011	-3.36	0.001	-.7825363	-.2059106

Fonte: Próprio autor, desenvolvido pelo programa STATA.

Foram analisados setecentos casos pelo teste LR  $\chi^2$ , cujas variáveis explicativas são conjuntamente importantes para explicar a variável dependente. O resultado apresentado no Prob  $\chi^2$  considera que todos os coeficientes são estatisticamente iguais a zero. O pseudo R<sup>2</sup> informa que trinta por cento da variação das variáveis dependentes podem ser atribuídas às variáveis explicativas.

Analisando o modelo pelos parâmetros estimados, todas as variáveis são estatisticamente significativas. É importante ressaltar que os valores dos parâmetros estimados na linguagem do modelo *Probit* significam a contribuição individual das variáveis no índice de utilidade que, embora não seja observável, está correlacionado com a probabilidade de que o evento ocorra. Enquanto os valores negativos significam uma queda na probabilidade enquanto que os valores positivos significam o oposto.

Foi possível observar que as variáveis “anos no mesmo emprego” e “anos no mesmo endereço de moradia” diminuem a probabilidade de o indivíduo ser inadimplente. Por sua vez, o percentual da renda comprometido com dívidas, bem como as dívidas de cartões de crédito aumenta a probabilidade de o indivíduo ser mau pagador. Para melhor aprofundar os estudos relativos aos microdados e expressar com mais clareza o modelo, é possível analisar o efeito marginal das

variáveis sob a probabilidade de o indivíduo ser inadimplente, digitando-se o comando `mfx` no STATA. Para análises individuais, o modelo *Probit* possibilita ainda o desenvolvimento do perfil do indivíduo e a respectiva de probabilidade de ser ou não mau pagador.

Foi possível ainda calcular a probabilidade de cada um dos indivíduos da amostra como adimplente ou inadimplente, diante das variáveis que compõem seu perfil. Para isso, o STATA dispõe do comando *Predict prob*, a partir do qual digita-se `list prob`. Com isso, o modelo identifica a probabilidade individual de o sujeito ser ou não um mau pagador. Por se tratar de setecentas observações, não cabe aqui demonstrar esse resultado.

É fato que o modelo *Probit* é estimado via máxima verossimilhança. Após essa primeira análise, pela utilização do comando `estat class`, foi possível identificar o percentual de acerto do modelo, que foi capaz de classificar corretamente oitenta e um por cento dos casos, conforme resultado apresentado na tabela 5.

Tabela 05 – Análise do modelo *Probit* com todas as observações – STATA.

<i>Classified</i>	D	~D	Total
+	90	37	127
-	93	480	573
Total	183	517	700

<i>Classified + if predicted Pr(D) &gt;= .5</i>
<i>True D defined as default != 0</i>

<i>Sensitivity</i>	Pr( + D)	49.18%
<i>Specificity</i>	Pr( --D)	92.84%
<i>Positive predictive value</i>	Pr( D +)	70.87%
<i>Negative predictive value</i>	Pr(~D -)	83.77%

<i>False + rate for true ~D</i>	Pr( +~D)	7.16%
<i>False - rate for true D</i>	Pr( - D)	50.82%
<i>False + rate for classified +</i>	Pr(~D +)	29.13%
<i>False - rate for classified -</i>	Pr( D -)	16.23%

Fonte: Próprio autor, desenvolvido pelo programa STATA.

Evitando-se dúvidas em relação à eficiência do modelo, além de o banco original ter sido acessado, foi feita, com quatrocentos e vinte casos, uma sub-amostra para servir como verificação, cujos resultados são apresentados na tabela 6:

Tabela 06 – Modelo para verificação realizado com quatrocentos e vinte casos – STATA.

<i>Number of obs</i>	420
<i>LR chi2(4) =</i>	118.42
<i>Prob &gt; chi2 =</i>	0.0000
<i>Pseudo R2 =</i>	0.2556

<i>Log likelihood = -172.47376</i>
------------------------------------

<i>default</i>	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]
----------------	-------	-----------	---	-----	----------------------

<i>employ</i>	-.1141725	.0184159	-6.20	0.000	-.1502669	-.078078
<i>adress</i>	-.0361802	.014154	-2.56	0.011	-.0639216	-.0084388
<i>debtinc</i>	.0430937	.0143433	3.00	0.003	.0149813	.0712062
<i>creddebt</i>	.3066015	.0653636	4.69	0.000	.1784913	.4347118
<i>_cons</i>	-.5954225	.1898246	-3.14	0.002	-.9674718	-.2233732

Fonte: Próprio autor, desenvolvido pelo programa STATA.

Novamente, os resultados foram estatisticamente significativos, sendo possível obter, pelo comando *estat class*, o percentual de casos classificados corretamente que passou a ser 82%, confirmando a efetividade do modelo criado.

Tabela 07 – Análise do modelo *Probit* com quatrocentos e vinte observações – STATA.

<i>Classified</i>	D	~D	Total
	41	15	56
	-60	304	364
Total	101	319	420

<i>Classified + if predicted Pr(D) &gt;= .5</i>
<i>True D defined as default != 0</i>

<i>Sensitivity</i>	Pr( + D)	40.59%
<i>Specificity</i>	Pr( ~D)	95.30%
<i>Positive predictive value</i>	Pr( D +)	73.21%
<i>Negative predictive value</i>	Pr( ~D -)	83.52%

<i>False + rate for true ~D</i>	Pr( + ~D)	4.70%
<i>False - rate for true D</i>	Pr( - D)	59.41%
<i>False + rate for classified +</i>	Pr( ~D +)	26.79%
<i>False - rate for classified -</i>	Pr( D -)	16.48%

<i>Correctly classified</i>	81.43%
-----------------------------	--------

Fonte: Próprio autor, desenvolvido pelo programa STATA.

Traçando um comparativo entre os três modelos, pode-se afirmar que as estimativas estão consistentes, os sinais dos coeficientes são os mesmos em todos os modelos, e as mesmas variáveis são estatisticamente significantes em cada um deles, conforme dados apresentados na tabela 8.

Tabela 08 – Comparativo entre os modelos desenvolvidos.

Variáveis Independentes	MQO	<i>Probit</i> 700	<i>Probit</i> 420
<i>Employ</i>	-0,237	-0,135	-0,114
<i>Adress</i>	-0,007	-0,043	-0,036
<i>Debtinc</i>	0,015	0,051	0,043
<i>Creddebt</i>	0,062	0,322	0,307
<i>Const</i>	0,273	-0,484	-0,595

Porcentagem correta prevista	81%	81,43%	82,14%
Valor log verossimilhança	_____	-279	-172
Pseudo R-quadrado	0,278	0,306	0,256

Fonte: Próprio pesquisador – base de dados do STATA.

O pseudo  $R^2$  do modelo *Probit* é baseado nas log-verossimilhanças. As magnitudes das estimativas de coeficiente entre modelos não são diretamente comparáveis, de modo que a maior diferença entre o MQO e o modelo *Probit* é que, enquanto o primeiro assume efeitos marginais constantes, o segundo implica magnitudes decrescentes dos efeitos parciais.

Apesar do desenvolvimento e da aplicação do modelo mais adequado ao alcance dos objetivos propostos, foi observada uma limitação relativa à endogeneidade, uma vez que o *Probit* não corrige esse erro do modelo.

## 5 Considerações finais

O objetivo deste trabalho foi realizar o desenvolvimento de um modelo *Probit* no programa STATA. Para isso, foi utilizada uma *cross-section*, proveniente do banco de dados *bankloan.sav* – IBM – SPSS, um modelo econométrico cuja finalidade é auxiliar o desenvolvimento das estratégias de gestão e funcionais de marketing, identificando-se a relação causal entre as características inerentes ao indivíduo e seu perfil adimplente ou não.

O estudo revelou que a mudança em uma variável independente provoca uma mudança na variável dependente. Por exemplo, ter mais um ano no mesmo endereço diminui a probabilidade de o indivíduo ser inadimplente. Como as variáveis econômicas são devidamente interpretadas como variáveis aleatórias, a ideia de probabilidade deverá ser utilizada para formalizar o sentido em que uma mudança gera outra. Trata-se da noção *ceteris paribus*, isto é, mantendo todos os outros fatores (relevantes), fixa-se no cerne para se estabelecer essa relação causal.

Com um percentual de acerto superior a 80%, esse modelo nasce como uma ferramenta que pode minimizar o risco de interações indesejáveis com clientes potencialmente inadimplentes, auxiliando no direcionamento das relações comerciais.

As características observadas para a classificação de bons ou maus pagadores foram, basicamente, as relações entre dívidas e rendas: as dívidas de cartão de crédito, os números de anos na mesma residência e o número de anos no mesmo emprego.

Pelo modelo *Probit*, os gestores poderão identificar a probabilidade de o indivíduo ser ou não inadimplente. Diante dessa informação, a empresa poderá optar entre aumentar o grau de relacionamento com seus clientes ou não.

## 6 Sugestões e Recomendações para Futuras Pesquisas

Este estudo aplicou o método *Probit* para o desenvolvimento de um modelo preditivo a partir dos dados secundários oriundos da base de dados *bankloan.sav*, disponível no programa IBM SPSS. O modelo desenvolvido busca minimizar os riscos operacionais, destacando a importância do desenvolvimento de estratégias de gestão e marketing.

O modelo estima, testa e diagnóstica a probabilidade de um consumidor ser inadimplente, bem como analisa os dados em uma *cross-section*. Como esses dados aplicam-se à economia americana, não representando a realidade do consumidor brasileiro, a autora propõe, como recomendações para trabalhos futuros, o desenvolvimento de um banco de dados real que parta da interação com os possíveis clientes, por meio de uma anamnese contendo as variáveis identificadas na modelagem desenvolvida por este artigo. Recomenda-se, além disso, um

levantamento quantitativo, a partir de um questionário, com o objetivo de obter informações básicas, inerentes aos consumidores, sendo, desse modo, fidedigno ao negócio e contribuindo de forma prática e objetiva para o desenvolvimento de estratégias lucrativas.

## 7 REFERÊNCIAS BIBLIOGRÁFICAS

AAKER, David A.; KUMAR, V.; DAY, George S. **Pesquisa de marketing**. Trad. Reynaldo Cavalheiro Marcondes. 2ª Ed., Ed. Atlas, São Paulo (SP), 2004.

BOYD JR., Happer W.; WESTFALL, Ralph; STASCH, Stanley F. **Marketing research: text and cases**. Ed. Richard D, INC, fourth edition. Homewood, Illinois, EUA, 1977.

CHURCHILL JR., Gilbert A. **Marketing Research: metodological foundations**. The Dryden Press, 7ª ed., Orlando, Flórida, EUA, 1999.

CRESWELL, J.W. **QUALITATIVE INQUIRY AND RESEARCH DESIGN: Choosing Among Five Approaches**, 3rd edition Thousand Oaks, CA: Sage. 2013.

GLASER, B. The use of secondary analysis by the independent researcher. In: **The American Behavioural Scientist**. Ed. 6;: 11–14, 1963.

Hahn, E. D. and R. Soyer. "Probit and Logit Models: Differences in a Multivariate Realm." Retrieved May 28, 2008, from <http://home.gwu.edu/~soyer/mv1h.pdf>.

GREENBERG, B. G. "Chester I. Bliss, 1899-1979." *International Statistical Review / Revue Internationale de Statistique* 8(1): 135-136. 1980.

HAHN, E. D. and R. Soyer. "Probit and Logit Models: Differences in a Multivariate Realm." Retrieved May 28, 2008, from <http://home.gwu.edu/~soyer/mv1h.pdf>.

HEWSON, C. Secondary Analysis, in Jupp, **The Sage Dictionary of Research Methods**, London, Sage, V ed., 2006.

HITT, Michael A.; IRELAND, R. Duane; E.HOSKISSON, Robert. **Administração Estratégica**. 2. ed. São Paulo: Cengage Learning, 2008. 415 p.

<http://www.ibm.com/software/analytics/spss/> acesso 28 de janeiro de 2015.

<http://www.stata.com/> acesso 28 de janeiro de 2015.

KLUYVER, Cornelis A. de; II, John A. Pearce. **Estratégia: Uma Visão Executiva**. 3. ed. São Paulo: Pearson, 2010. 226 p.

KOTLER, P. **Administração de Marketing: a Bíblia do Marketing**. 14ed. São Paulo, Prentice Hall, 2013.

MALHOTRA, Naresh. **Pesquisa de marketing: uma orientação aplicada**. Trad. Lene Belon Ribeiro e Monica Stefani. Bookman, 6ª ed., Porto Alegre (RS), 2012.

MINTZBERG, Henry; AHLSTRAND, Bruce; LAMPEL, Joseph. Safári de Estratégia. Porto Alegre: Bookman, 2000. 299 p.

SANTOS, D. Fidelização de clientes. 2008. Disponível em internet: <http://www.administração.com.br/artigos/marketing/fidelizacao-de-clientes/20882/>. Acesso em 29 de janeiro de 2015.

STONER, A.F James; FREEMAN, R.Edward. Administração. 5. ed. Rio de Janeiro: Editora Ltc, 2012. 533 p.

WHITTINGTON, Richard. **O que é Estratégia**. 1ed São Paulo: Thomson, 2002.154 p.

WENTZ, Walter B. **Marketing research: management and methods**. Harper & Row Publishers, New York, 1979.

WICKE, T.; SILVER, R. C. A community responds to collective trauma: An ecological analysis of the James Byrd murder in Jasper. In: **American Journal of Community Psychology**. Ed. 44, 233–248. Texas, 2009.

WOOLDRIDGE, J, M **Introdução à econometria:uma abordagem moderna** 2 edição. Cengage – São Paulo, 2010

WOOLDRIDGE, J, M “**Econometric Analysis of Cross-Section and Panel Data**”, MIT Press, 2011.