

## **ESTUDO SOBRE OS FATORES DETERMINANTES DA RENDA DE CLIENTES BANCÁRIOS**

**HENRIQUE DANTAS DA SILVA**

*Faculdade de Economia, Administração e Contabilidade da USP.*

### **Resumo**

A utilização de novas tecnologias revolucionou a forma como pessoas e instituições interagem. Desde os famosos *bureaus* de crédito até os modelos mais recentes de discriminação de bons e maus pagadores, o mercado evoluiu e a forma como se oferece crédito também. Contudo, 40% da população brasileira economicamente ativa, a qual movimenta anualmente 655 bilhões de reais, ainda não está bancarizada. Para contornar essa situação, existem propostas de se utilizar dispositivos móveis como forma de se reduzir os custos com bancarização e tornar possível que essas pessoas possam adentrar o mercado de crédito. Tendo em vista essa possibilidade, o objetivo deste estudo é identificar os fatores determinantes da renda de clientes bancários e assim fornecer insumo para o desenvolvimento de políticas de crédito a partir dos dispositivos móveis. A identificação dos principais fatores foi realizada através da técnica estatística de regressão linear e após a identificação das variáveis pelo Método *Stepwise*, foi possível a construção de um modelo estatístico para a mensuração da capacidade de pagamento (renda). O modelo poderá ser utilizado junto à tecnologia móvel e permitirá às instituições financeiras reduzirem os custos com a bancarização. Os resultados do estudo indicam que das 30 variáveis testadas, as que possuem maior poder de discriminação da renda de clientes bancários são: saldo devedor total, limite do cartão de crédito, tempo de conta corrente, profissão, investimento em CDB, investimentos mantidos na instituição, idade, financiamento de automóveis, estado civil, *rating*, risco total mantido na instituição, crédito imobiliário, saldo devedor no cartão de crédito, sexo, crédito consignado e crédito em atividades de agronegócio. Contudo o modelo possui como limitação a necessidade de que o indivíduo que esteja sendo avaliado, já tenha algum dia operado no mercado de crédito, pois das variáveis selecionadas, parte delas levam em consideração o histórico de contratação de crédito pelo indivíduo, além da própria quantidade de variáveis avaliadas.

**Palavras-Chave:** *CreditScoring*; Modelos de Estimação da Renda; Risco de Crédito; Bancarização; Método *Stepwise*.

### **1. Introdução**

O crédito fornecido às pessoas físicas representa uma das principais atividades realizadas pelas instituições financeiras. Os bancos comerciais (regulamentados pelo Banco Central do Brasil através da Resolução CMN 2.099, de 1994), espalhados através de suas 34,2 mil agências e pontos de atendimentos – PABs pelo Brasil, oferecem ao mercado pessoa física 1,2 trilhões de reais divididos nos mais diversos produtos como, por exemplo, Crédito Pessoal,

Cartão de Crédito, Crédito Consignado, Crédito Imobiliário, Financiamento de veículos, Financiamento rural, Cheque Especial, dentre outros. (Federação Brasileira de Bancos - FEBRABAN, 2012).

O crédito obtido pelos indivíduos tem como função estimular o consumo e/ou suprir uma necessidade desses consumidores no curto prazo (Caouette *et. al.*, 2009). Contudo, a população brasileira economicamente ativa, a qual é constituída por 86 milhões de brasileiros, apresenta 52% desses sem vinculação bancária (Cernev, 2008). Esses indivíduos correspondem a aproximadamente 45 milhões de brasileiros e movimentam anualmente 655 bilhões de reais. Esses brasileiros que não possuem relacionamento bancário podem ser divididos em três grupos: 1) aqueles que possuem dificuldade de acesso à rede bancária; 2) aqueles que não desejam ter relacionamento bancário por não acreditarem no papel das instituições financeiras; e 3) aqueles que já possuíram relacionamento com instituições financeiras, mas por estarem inadimplentes, não mais o possuem.

Os bancos comerciais sabem da existência desses indivíduos, como também o quanto eles movimentam anualmente em termos monetários, porém o motivo de não atuarem de forma massiva sobre essa população está relacionado ao custo *versus* benefício, tendo em vista o primeiro grupo citado. O custo ponderado pelos bancos está no gerenciamento de novas agências e pontos de atendimento – PABs em regiões com uma densidade populacional baixa, tendo em vista que dos 55 milhões de indivíduos não bancarizados, 40% são de baixa renda e 50% deles vivem nas regiões Norte e Nordeste do país.

Ou seja, para parte da população brasileira o acesso à rede bancária não é simples. Além do difícil acesso enfrentado, o Banco Central – BACEN através dos bancos, como também as próprias instituições financeiras, durante o processo de cadastro dos clientes, exigem uma grande quantidade de informações e de documentos, o que aumenta o custo do processo e inibe a bancarização (regulamentado pela Resolução CMN 2.025, de 1993). Tendo em vista essa situação, Breternitz (2009) analisou a importância da bancarização da classe de baixa renda e enxergou na utilização da tecnologia dos dispositivos móveis uma maneira de diminuição dos custos de bancarização, tornando assim a situação atrativa aos bancos comerciais e, conseqüentemente, à classe de baixa renda.

Breternitz (2009) recomendou ainda que modelos de negócios adotados em outros países fossem estudados a fim de fornecerem base para a estruturação de um novo sistema de bancarização através de dispositivos móveis no Brasil. Entretanto, o desenvolvimento de uma prática inovadora no setor bancário, como a criação de um aplicativo que realizasse a abertura de contas correntes e poupanças, exigiria que as instituições financeiras contornassem o problema enfrentado com a assimetria informacional.

Mesmo dentre os próprios clientes, as instituições por vezes desconhecem inúmeras informações, não possuindo confiabilidade sobre o real posicionamento do indivíduo no mercado financeiro. Esse problema pode se agravar, quando a informação desconhecida for à fonte e/ou o valor de renda do indivíduo, pois sem o cadastro ou atualização destas informações nenhum produto financeiro pode ser contratado.

Desta maneira, o desenvolvimento de um aplicativo que realize a abertura de contas correntes ou poupanças, o qual seria imprescindível no processo de bancarização da sociedade, como destaca Breternitz (2009), não será possível caso o aplicativo não permita às instituições financeiras identificarem a fonte de renda, o valor da renda, como também outras variáveis que ao se correlacionarem com a renda sustentem a veracidade da informação. Sendo assim, esta

pesquisa pretende responder a seguinte questão de pesquisa: Quais os fatores determinantes da renda de clientes bancários? Assim sendo, o trabalho tem como objetivo principal identificar os fatores determinantes da renda, a partir de uma amostra de 89.862 clientes bancários, com o retrato do posicionamento financeiro destes clientes em 31 de dezembro de 2014.

Nesse sentido, o trabalho contribuirá com o processo de análise de crédito ao identificar as principais variáveis capazes de determinar e prever a renda do cliente, tornando possível, futuramente, o desenvolvimento de um sistema para dispositivos móveis que exija as informações mais relevantes, de forma a auxiliar os processos das instituições financeiras e, conseqüentemente, a bancarização da população de baixa, média ou até mesmo alta renda.

Ressalta-se que durante o processo de análise da base de dados a pesquisa se fundamentou em alguns trabalhos, em especial no trabalho de Gouvêa *et. al.* (2012). Os autores, na ocasião, analisaram as técnicas de modelagem Regressão Logística *versus* Algoritmos Genéticos, para os procedimentos de solicitações de crédito, e identificaram que ambas as técnicas geraram modelos com resultados satisfatórios. Contudo, nos trabalhos que utilizam regressão logística, por exemplo, os autores buscam classificar um cliente em “bom” ou “mau” (variável de saída dicotômica), e assim optam pelas técnicas citadas anteriormente. Já o presente trabalho busca identificar quais variáveis possuem maior correlação com a renda dos clientes, ou seja, o trabalho assume como necessidade o relacionamento linear dos dados, sendo assim, será adotada a técnica estatística de regressão linear, visto que a variável dependente é métrica (Renda). Através da identificação dos principais fatores determinantes da renda, será possível a construção de um modelo de predição de renda, o qual poderá ser adotado por instituições e utilizado no aprimoramento de concessão de crédito.

Além desta introdução, o trabalho está organizado da seguinte forma: Na seção 2 será abordada a Fundamentação Teórica, a qual dará insumo ao desenvolvimento do trabalho. Na seção 3 é descrita a amostra, as principais variáveis e os procedimentos metodológicos utilizados. Na seção 4 são apresentados os resultados da pesquisa e, por fim, na seção 5, são apresentadas as considerações finais.

## **2. Fundamentação Teórica**

### **2.1. Modelos de Classificação de Risco de Crédito (*Credit Scoring*)**

Por melhor que seja a identificação do cliente e a diminuição da assimetria informacional, as instituições precisam continuamente utilizar modelos que lhe possibilitem mitigar o risco de inadimplência (ou risco de *default*). Para mitigar o risco de crédito, existem diversas metodologias de mensuração, exposição e mecanismos de gestão (Lima *et al.*, 2009). Dentre os métodos existentes, é possível citar: análise discriminante, regressão logística, regressão linear, árvores de decisão, programação linear e suas variações (Samejima *et al.*, 2003).

Porém, a quantidade de modelos existentes atualmente não revela o quão custoso à tarefa de análise de crédito era anteriormente. Apenas após o surgimento de computadores com capacidades operacionais mínimas fora possível a aplicação de modelos teóricos. Diante desta situação, somente na década de 90 os modelos de *credit scoring* começaram a ser usados no Brasil e, em decorrência, muitos estudos foram desenvolvidos.

Neste sentido, o trabalho irá inicialmente apresentar os modelos desenvolvidos nos estudos de Altman (1968), Elizabetsky (1976) e Kanitz (1976), pois são trabalhos seminais referentes a risco de crédito, e os quais contribuirão para o entendimento da evolução dos

modelos de classificação de risco de crédito. Vale ressaltar que estes trabalhos foram desenvolvidos tendo em vista a análise de instituições não financeiras, logo, será realizado um paralelo e análise da funcionalidade das técnicas na avaliação de clientes pessoa física para a utilização por instituições financeiras para prever o risco de crédito.

Em 1968, Altman desenvolveu um modelo de previsão de insolvência, porém tendo em vista o cenário macroeconômico, o modelo fora revisto e lançado em 1979, intitulado “Previsão de problemas financeiros em empresas”. O objetivo do trabalho era avaliar as falências de empresas nos Estados Unidos e a partir disso criar um modelo que permitisse a previsão desse acontecimento. Através da técnica de análise discriminante e regressão múltipla, o autor desenvolveu um modelo de previsão de insolvência. Na realização do estudo Altman (1968) utilizou uma amostra de empresas, a qual a dividiu em dois grupos, empresas com problemas e empresas sem problemas. O autor construiu dois modelos e realizou comparações para avaliar a sua efetividade. Após a escolha do primeiro modelo utilizou o teste de Lachenbruch, no qual ele retira uma empresa da amostra e aplica o modelo nas demais, buscando classificar a empresa que havia sido retirada e assim avaliar a assertividade do modelo.

Altman (1968) concluiu que o modelo possuía significância, porém, tendo em vista que os dados utilizados eram dados públicos e divulgados pelas empresas, havia o risco de confiabilidade informacional. O seu modelo fora desenvolvido no mercado americano, mas em 1999 o autor adaptou o modelo para o mercado brasileiro. Sua importância é tamanha ao passo que, ao lado de Kanitz e Elizabetsky, representa o início da avaliação do risco de crédito através de modelos estatísticos.

Kanitz foi pioneiro na utilização de modelos estatísticos para prever o risco de insolvência das empresas. Através das técnicas de regressão múltipla e análise discriminante, o autor desenvolveu em 1976 uma equação matemática com cinco índices de análise de balanço. O autor defende a possibilidade de se prever, com certo grau de segurança, a situação financeira e, assim, estimar o risco de insolvência. Essa previsão poderia ser realizada através da análise de índices apresentados no balanço das empresas. O objetivo era avaliar o comportamento dos índices ano a ano e identificar tendências de melhora ou piora.

Kanitz (1976) esclarece ao final do estudo os principais pontos que podem prever a insolvência das empresas: endividamento e fluxo de caixa. O estudo de Kanitz, apesar de ter como foco a pessoa jurídica, é referência para a criação de modelos estatísticos para prever o risco de insolvência da pessoa física, entretanto, as instituições financeiras não conseguem obter do cliente relatórios financeiros, como por exemplo, balanço ou fluxo de caixa. Desta forma, caso o indivíduo não seja cliente da instituição existirá assimetria informacional, o que colocara em dúvida a capacidade de pagamento do cliente.

Elizabetsky apresentou, em 1976, um modelo matemático para decisões de crédito em bancos comerciais. Através da técnica de Análise discriminante, Elizabetsky desenvolveu um indicador de falência, o qual busca analisar informações do cliente que revelem a sua relação com o mercado e a relação dele com instituições fornecedoras de crédito. Elizabetsky (1976) destaca em seu trabalho que variáveis como qualificação dos dirigentes, risco das operações *versus* rentabilidade, histórico de pagamentos, análise das demonstrações contábeis e utilização dos recursos pela empresa para avaliar o risco de insolvência da empresa. Elizabetsky (1976) defende que a utilização de modelos matemáticos para a avaliação do risco de crédito ou insolvência de pessoas jurídicas ou mesmo pessoas físicas é mais vantajoso para o negócio ao

passo que há redução do tempo de análise, diminuição da interferência humana, análise simultânea de diferentes casos, melhoria na qualidade da decisão, redução do tempo de avaliação, possibilidade de alterações na política, tendo em vista o desempenho da economia, ou em resumo a padronização do processo de oferecimento de crédito.

É perceptível que o mercado financeiro em na década de 70 estava sedento por essa padronização, a qual já havia sido difundida por Ford no início do século com a automação de veículos. Entretanto, essa automação só fora possível no mercado financeiro com a ampliação das tecnologias, e essa situação não é diferente de hoje, os modelos de *credit score* operam com as tecnologias vigentes, porém, com a modificação e expansão dessas tecnologias, os modelos ganham mais espaço de atuação, consequentemente, precisam ser adequados para esse novo cenário.

Neste sentido vale se atentar que Altman, Kanitz e Elizabetsky utilizaram a técnica de análise discriminante na construção de seus modelos. Altman e Kanitz, além da análise discriminante, utilizaram também a técnica de regressão linear múltipla. Sendo assim, serão avaliadas criticamente neste momento as seguintes técnicas estatísticas: análise discriminante, regressão linear múltipla, KMV e regressão logística.

A análise discriminante é uma ferramenta estatística utilizada na avaliação de pessoas jurídicas ou pessoas físicas, as quais são classificadas em bons e maus pagadores. Essa técnica assume como premissa a normalidade multivariada das variáveis independentes, igualdade na matriz de variância e covariância de grupos estipulados pelas variáveis dependentes, inexistência de multicolineariedade entre as variáveis independentes, relações lineares entre todas as variáveis e ausência de *outliers* (Virgillito *et al.*, 2008). A técnica tem como princípio pontuar os indivíduos analisados, tendo em vista a ponderação das variáveis analisadas e a partir da pontuação gerada, classificá-los em bons ou maus pagadores.

Segundo Mário (2002), para a construção da técnica de análise discriminante são necessários quatro passos: levantamento de amostras e segregação em grupos, definição dos coeficientes discriminantes para cada variável independente, apuração dos *scores* discriminantes de cada empresa, da média destes *scores* em cada grupo, e do *score* crítico ou ponto de corte, e por fim testar a função discriminante na amostra selecionada para verificação do grau de confiança ou acerto do modelo. Entretanto, a obtenção de amostra de indivíduos sabidamente bons pagadores e maus pagadores, para que seja possível a realização da amostra, é um empecilho, pois existe a dificuldade em se identificar essa característica. Em empresas, por exemplo, a confiabilidade das informações é necessária para que as informações das demonstrações financeiras possam ser utilizadas, e em se tratando de pessoa física, existe a necessidade de se confiar no método de comprovação de renda (capacidade de pagamento) para o desenvolvimento do modelo.

A técnica de regressão linear múltipla, a qual fora utilizada por Altman e Kanitz, busca identificar da mesma forma que a técnica de análise discriminante o grau de exposição ao risco de insolvência, porém, a regressão linear múltipla é uma coleção de técnicas estatísticas utilizadas na construção de modelos que descrevem de maneira razoável relações entre variáveis explicativas de um determinado processo. A diferença entre a regressão linear simples e a múltipla é que na múltipla são tratadas duas ou mais variáveis explicativas, logo, a regressão múltipla pode ser utilizada para as situações em que se analisam diversas variáveis, como nos processos de concessão de crédito. Assim, no presente trabalho, será utilizada a



regressão linear múltipla, visto que serão analisados 30 possíveis fatores determinantes da renda como variáveis explicativas.

Além desses recursos utilizados pelos autores, hoje o mercado possui modelos mais recentes que buscam se adequar as modificações do cenário macroeconômico e social. Neste sentido, o KMV ou *Credit Monitor*, representa mais uma das ferramentas utilizadas no gerenciamento de carteiras de crédito, sendo este utilizado na construção de modelos de *credit scoring* tanto para empresas, como para pessoas físicas. O modelo KMV foi desenvolvido através do aprimoramento do modelo de Merton, o qual defendia que o preço das ações negociadas no mercado refletia as expectativas do mercado sobre a empresa. Da mesma forma que o modelo de análise discriminante, o modelo KVM considera relevante as informações disponíveis no mercado sobre o preço dos ativos para a determinação da probabilidade de *default*.

Entretanto Santos *et. al.* (2009) expõem que: “A redução da perda [calculada através do modelo KMV] poderá afetar a expansão dos negócios e dos lucros. Há, nesse caso, uma relação custo–benefício que deverá ser analisada pelo analista e pela direção da empresa, de forma que esta possa maximizar os retornos aos acionistas dentro de um nível aceitável de risco.” Assim sendo, a técnica KMV, da mesma forma que a técnica discriminante, serve como balizador, mas não como única ferramenta aos gestores de carteiras de crédito, pois como é conhecido, as instituições financeiras executam um processo de clusterização, no qual busca não apenas aprovar ou não aprovar o crédito, mas sim definir quais produtos e quanto será aprovado a cada perfil de cliente. Neste sentido, a regressão logística se apresenta como uma técnica que se ajusta a essa classificação e, por isso, é frequentemente utilizada no mercado e nos estudos.

Contudo a regressão logística é uma metodologia estatística a qual prediz a probabilidade de um evento ocorrer e apresenta uma variável de saída dicotômica, podendo este estar entre 0 e 1. Como citado anteriormente, apesar de ser a técnica mais utilizada no mercado para o desenvolvimento de modelos de *credit scoring* (Rosa, 2000), a técnica não contemplaria o objetivo deste estudo, logo, será utilizada a regressão linear múltipla, considerando que a variável Y a ser estimada é métrica.

## 2.2. Estudos Empíricos

Neste tópico serão abordados alguns dos trabalhos nacionais, os quais utilizaram não somente técnicas estatísticas de regressão linear, como também outras técnicas na construção de modelos de *credit scoring* tanto para pessoa jurídica, como para pessoa física.

Assaf Neto (2006) apresentou um estudo no qual desenvolveu um modelo de classificação de empresas em solventes e insolventes, através da técnica estatística regressão logística. Ele utilizou uma amostra composta por empresas de capital aberto durante o período de 1994 e 2004 e a partir das demonstrações financeiras calculou índices financeiros, as quais utilizaram como variáveis independentes. Para a validação do modelo, o autor utilizou o método de Jackknife e a Curva ROC observando que, com bom nível de acurácia, era possível prever eventos de *default* com um ano de antecedência.

Gouvêa *et. al.* (2012) buscou realizar uma comparação entre a técnica de regressão logística e algoritmos genéticos. Os autores buscaram identificar qual das duas técnicas seria mais assertiva. Para tanto, os autores analisaram 14.000 dados fornecidos por uma grande instituição financeira. Os dados foram divididos em dois grupos, o primeiro com 8.000 dados

(sendo 4.000 bons e 4.000 maus) e, o segundo para a validação do modelo, com 6.000 dados (sendo 3.000 bons e 3.000 maus). Ao final do estudo, os autores identificaram que a regressão logística apresentou melhor ajuste aos dados, porém não houve uma diferença considerável, sendo que ambos os modelos apresentaram taxas de acerto próximas ou acima de 65%, as quais são consideradas boas por especialistas. (Picinini *et. al.*, 2003). Desta forma, os autores concluíram que a adoção de qualquer modelo traria bons resultados à instituição financeira, porém, existem poucos estudos com o modelo de algoritmos genéticos, o que motiva a utilização dos modelos de regressão logística.

Já no ano seguinte, Gouvêa *et. al.* (2013) desenvolveram um modelo de *credit scoring* através da técnica estatística de regressão logística. Na ocasião, os autores utilizaram uma amostra com 20.000 dados obtida junto a uma instituição financeira. A amostra foi dividida em três subamostras, uma para a construção do modelo e as outras duas para realizar a validação do mesmo. Durante a construção do modelo, os autores ressaltaram que, para o desenvolvimento de um modelo de *credit scoring*, é preciso definir o que a instituição considera como bom ou mau pagador, definição essa que está atrelada a política de crédito empregada. Por fim, os autores realizaram dois testes de significância, o teste de qui-quadrado e o teste de Hosmer e Lemeshow e concluíram que o modelo poderia ser utilizado por instituições financeiras para a tomada de decisão na concessão de crédito.

Diante deste cenário, este trabalho não tem como pretensão criar um modelo de *credit scoring*, mas sim identificar os fatores determinantes da renda e, assim, fornecer insumo para novos trabalhos, como também apresentar um modelo de previsão de renda. Vale ressaltar que a identificação do valor da renda é fundamental à partir do momento em que a renda de um cliente é a base para a concessão de crédito, tendo em vista que as políticas calculam a alavancagem a qual o cliente pode atingir através do valor de renda observado ou estimado.

### **3. Procedimentos Metodológicos**

#### **3.1. Amostra**

A amostra utilizada nesta pesquisa foi obtida junto a uma instituição financeira, porém sua identidade não pode ser revelada, da mesma forma que a identificação dos indivíduos da base fora excluída após a confecção da mesma, sendo criada outra chave de identificação.

A base inicial disponibilizada era composta por alguns milhões de registros, dos quais havia clientes ativos e inativos (12 meses sem movimentações de qualquer natureza). Desta forma, foram realizados os seguintes filtros para seleção dos clientes e composição da base utilizada como amostra: i) apenas clientes ativos, ii) apenas clientes que recebem salários na instituição, iii) apenas clientes que não transferem seus salários a outras instituições, iv) foram excluídos aqueles que foram desligados de seus respectivos empregos, v) como também foram excluídos os funcionários da instituição.

Após a realização destes filtros foram selecionados aleatoriamente 89.862 registros. A base de dados utilizada corresponde à posição dos clientes no mês de dezembro de 2014. Utilizou-se na construção dessa pesquisa a informação de clientes que recebem seus salários na instituição, pois esses são aqueles que possuem a variável renda (variável dependente) “real”, ou seja, a instituição reconhece, através de créditos na conta do cliente e através de convênio com as empresas, o salário bruto do indivíduo. Essa situação só não seria verdadeira quando o processo de demissão ocorrer e a informação não for repassada à instituição e, assim, esta só irá identificar o acontecimento após determinado período.

### 3.2. Definição das variáveis

#### 3.2.1. Variável Dependente

A variável dependente analisada nesta pesquisa é a Renda. Esse valor representa o poder de pagamento do indivíduo e é através dele que políticas de crédito e cálculos de *credit scoring* são realizados. A não utilização de clientes que não recebem seus salários através da instituição é devido aos diversos modelos de estimação de renda utilizados no processo de identificação da capacidade de pagamento do indivíduo. Nesse sentido, no intuito de reduzir interferências por outros modelos que não serão abordados nessa pesquisa, decidiu-se realizar o estudo com um público o qual a qualificação da renda fosse possível e precisa.

#### 3.2.2. Variáveis Independentes

Para a realização do estudo foram consideradas as seguintes variáveis independentes:

- *S\_devedor*: A variável “saldo devedor” é medida em R\$ (reais) e representa o saldo acima de 90 dias de atraso do cliente com a instituição.
- *L\_cartao*: A variável “limite do cartão” é medida em R\$ (reais) e representa o limite no cartão de crédito do indivíduo.
- *T\_conta*: A variável “tempo de conta corrente” é medida em dias e representa o tempo de existência de conta (relacionamento) do cliente na instituição.
- *Poupanca*: A variável “poupança” é medida em R\$ (reais) e representa o valor depositado em conta poupança pelo indivíduo na instituição.
- *CDB*: A variável “CDB” é medida em R\$ (reais) e representa o valor aplicado em CDB pelo indivíduo na instituição.
- *Prev*: A variável “previdência privada” é medida em R\$ (reais) e representa o valor aplicado em previdência privada pelo indivíduo na instituição.
- *Fundos*: A variável “fundos de investimentos” é medida em R\$ (reais) e representa o valor aplicado em fundos de investimentos pelo indivíduo na instituição.
- *Inv*: A variável “Investimentos” é medida em R\$ (reais) e representa a somatória dos valores aplicados em Poupança, CDB, Previdência Privada e Fundos de investimento pelo indivíduo na instituição.
- *Idade*: A variável “idade” é medida em dias e representa a idade do cliente.
- *Correntista*: A variável “correntista” é binária e representa se o cliente possui conta corrente (1) ou não (0) na instituição.
- *Risco\_hist*: A variável “risco histórico” é medida em R\$ (reais) e representa a soma de todos os contratos de crédito realizados pelo cliente com a instituição desde a abertura da conta corrente.
- *Risco\_merc*: A variável “risco no mercado” é medida em R\$ (reais) e representa o saldo atual de todos os contratos de crédito realizados pelo cliente no mercado de crédito.
- *Risco\_total*: A variável “risco total” é medida em R\$ (reais) e representa o saldo de todos os contratos de crédito realizados pelo cliente na instituição.
- *Reneg*: A variável “crédito com renegociação” é medida em R\$ (reais) e representa o saldo de contratos de renegociação realizados pelo cliente na instituição.
- *Imob*: A variável “crédito imobiliário” é medida em R\$ (reais) e representa o saldo de crédito imobiliário realizado pelo cliente na instituição.
- *Cartão*: A variável “cartão de crédito” é medida em R\$ (reais) e representa o saldo no cartão de crédito (saldo em atraso) realizado pelo cliente na instituição.



- *Cheque*: A variável “cheque especial” é medida em R\$ (reais) e representa o saldo no cheque especial (utilização do limite de cheque especial) realizado pelo cliente na instituição.
- *CP*: A variável “Crédito Pessoal” é medida em R\$ (reais) e representa o saldo de crédito pessoal realizado pelo cliente na instituição.
- *Consig*: A variável “Crédito Consignado” é medida em R\$ (reais) e representa o saldo de crédito consignado realizado pelo cliente na instituição.
- *Agro*: A variável “Crédito em Agronegócio” é medida em R\$ (reais) e representa o saldo de crédito concedido a atividades de agronegócio realizado pelo cliente na instituição.
- *Autos*: A variável “Financiamento de automóveis” é medida em R\$ (reais) e representa o saldo de crédito concedido ao financiamento de automóveis realizado pelo cliente na instituição.
- *Rating*: A variável “Rating” é medida na escala de 1 a 9 e representa o nível de confiança da instituição com o cliente, sendo 1 pior e 9 melhor.
- *Q\_depend*: A variável “quantidade de dependentes” é medida em valores inteiros a partir de 0 e representa a quantidade de dependentes que o cliente possui.
- *T\_emprego*: A variável “tempo de emprego” é medida em dias e representa há quanto tempo o cliente esta empregado em sua posição atual.
- *Prof*: A variável “profissão” esta codificada com valores de 1000 a 9999 e representa qual a profissão ou atividade principal do cliente. A escala foi construída tendo em vista as atividades mais rentáveis, sendo que quanto menor o número, maior a renda.
- *E\_civil*: A variável “estado civil” está codificada de 1 a 7 e representa a situação conjugal do indivíduo, sendo 1: Solteiro; 2: Casado sem comunhão de bens; 3: Casado com comunhão de bens; 4: Divorciado; 5: Separado; 6: União Estável; 7: Viúvo.
- *A\_atuacao*: A variável “área de atuação” esta codificada com valores de 1000 a 9999 e representa qual a área de atuação do indivíduo. A escala foi construída tendo em vista as atividades mais rentáveis, sendo que quanto menor o número, maior a renda.
- *Escol*: A variável “escolaridade” esta codificada com valores de 1 a 15 e representa qual o nível de instrução do indivíduo. Quanto menor o número menor o grau de instrução do indivíduo.
- *Sexo*: A variável “sexo” é binária é representa se o indivíduo é homem (1) ou mulher (0).

Vale ressaltar que para as variáveis mensuradas em moeda (R\$), fora utilizado como ln (logaritmo natural) visando reduzir problemas de heterocedasticidade em função da escala de mensuração das variáveis.

### 3.3. Tratamento dos Dados

Após a definição da base amostral foram calculadas estatísticas descritivas, com o intuito de identificar o comportamento da amostra. Na base amostral é possível identificar a existência de 30 variáveis explicativas e 89.862 observações. Tendo em vista que, em uma distribuição normal ou distribuição de Gauss, na qual 99,7% das observações concentram-se até 3 desvios padrão da média, fora construído um histograma a partir da variável renda e, assim, a hipótese de distribuição normal para renda fora validada.



São Paulo, 27 a 29 de Julho de 2016

## **Building Knowledge in Accounting**

A técnica estatística escolhida para a realização do estudo foi à regressão linear, a qual considera que a relação da resposta às variáveis é uma função linear dos parâmetros. Para se estimar o valor esperado, o qual determina a relação entre ambas as variáveis (Y e X), utiliza-se a seguinte equação:

$$y = \alpha + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i} + \text{erro}$$

Em que: y = variável dependente; x = variáveis independentes;  $\alpha$  = constante;  $\beta$  = coeficientes angulares.

Devido ao tamanho da amostra, assume-se a normalidade dos dados. Já a possível existência de multicolinearidade entre as variáveis explicativas será analisada por meio da correlação entre elas. Também serão consideradas as significâncias dos parâmetros de uma regressão robusta por meio da correção de White para os desvios-padrão dos coeficientes, buscando minimizar problemas de heterocedasticidade dos resíduos. Já a autocorrelação residual não deve ser analisada visto que a base não utiliza séries temporais de dados, mas constitui-se em um corte transversal (*cross-section*). As análises foram realizadas com *software* Stata 11.

#### **4. Resultados e discussões**

Fora construída inicialmente uma tabela com as estatísticas descritivas de todas as variáveis, evidenciada a seguir.

**Tabela 1. Estatísticas Descritivas**

| Variáveis   | Obs   | Média    | Desvio-Padrão | Mínimo | Máximo      |
|-------------|-------|----------|---------------|--------|-------------|
| renda       | 89862 | 5798,62  | 30813,72      | 0,00   | 8002468,00  |
| s_devedor   | 89862 | 36671,52 | 429820,60     | 0,00   | 82500000,00 |
| l_cartao    | 89862 | 10007,32 | 25882,61      | 0,00   | 900000,00   |
| t_conta     | 89862 | 3129,36  | 2928,51       | 0,00   | 9999,00     |
| poupanca    | 89862 | 4739,08  | 35029,95      | 0,00   | 4406614,00  |
| cdb         | 89862 | 32802,79 | 492985,40     | 0,00   | 69200000,00 |
| prev        | 89862 | 12055,84 | 239289,80     | 0,00   | 51800000,00 |
| fundos      | 89862 | 18112,76 | 341960,40     | 0,00   | 82500000,00 |
| inv         | 89862 | 58428,10 | 696267,30     | 0,00   | 82500000,00 |
| idade       | 89862 | 44,07    | 16,10         | 0,00   | 119,00      |
| correntista | 89862 | 0,88     | 0,32          | 0,00   | 1,00        |
| risco_hist  | 89862 | 41210,82 | 278503,70     | 0,00   | 44700000,00 |
| risco_merc  | 89862 | 45313,95 | 366727,90     | 0,00   | 51600000,00 |
| risco_total | 89862 | 15785,87 | 141242,00     | 0,00   | 24200000,00 |
| reneg       | 89862 | 921,02   | 7436,02       | 0,00   | 432564,70   |
| imob        | 89862 | 5974,91  | 96889,63      | 0,00   | 17500000,00 |
| cartao      | 89862 | 2726,22  | 7383,37       | 0,00   | 269932,00   |
| cheque      | 89862 | 598,57   | 3282,36       | 0,00   | 276759,00   |
| cp          | 89862 | 1630,07  | 26215,89      | 0,00   | 6710823,00  |
| consig      | 89862 | 1498,82  | 8072,71       | 0,00   | 480026,00   |
| agro        | 89862 | 956,63   | 86762,45      | 0,00   | 24200000,00 |
| autos       | 89862 | 578,07   | 4672,65       | 0,00   | 207406,10   |
| rating      | 89862 | 6,41     | 2,15          | 0,00   | 9,00        |
| q_depend    | 89862 | 0,51     | 1,22          | 0,00   | 99,00       |
| t_emprego   | 89862 | 296,11   | 137,90        | 0,00   | 981,00      |
| prof        | 89862 | 4988,22  | 3116,03       | 0,00   | 9990,00     |
| e_civil     | 89675 | 3,53     | 2,91          | 0,00   | 9,00        |
| a_atuacao   | 89862 | 8,47     | 5,88          | 0,00   | 19,00       |
| escol       | 89861 | 3,63     | 2,72          | 0,00   | 14,00       |
| residencia  | 89862 | 3,71     | 3,09          | 0,00   | 82,00       |
| sexo        | 89862 | 0,60     | 0,49          | 0,00   | 1,00        |

Notas: Obs: número de observações; s\_devedor: saldo devedor; l\_cartao: limite do cartão; t\_conta: tempo de conta corrente; inv: investimentos; risco\_hist: risco histórico; risco\_merc: risco no mercado; reneg: crédito com renegociação; imob: crédito imobiliário; cp: crédito pessoal; consig: crédito consignado; agro: crédito em agronegócio; autos: financiamento de automóveis; q\_depend: quantidade de dependentes; t\_emprego: tempo de emprego; prof: profissão; e\_civil: estado civil; a\_atuacao: área de atuação; e\_escol: escolaridade.

Na sequência foram analisados os pressupostos da regressão linear, dos quais dentre eles está à ausência de multicolinearidade, a qual foi analisada através da correlação de Pearson entre as variáveis explicativas métricas. A Tabela 2 a seguir apresenta os resultados obtidos.



São Paulo, 27 a 29 de Julho de 2016

## Building Knowledge in Accounting

**Tabela 2. Correlação de Pearson entre as Variáveis Dependente e Explicativas**

|             | renda | s_devedor | l_cartao | t_conta | poupanca | cdb         | prev | fundos | inv   | idade | correntista | risco_hist  | risco_merc | risco_total | reneg | imob | cartao | cheque | cp   | consig | agro | autos | rating | q_depend | t_emprego | prof | e_civil | a_atuacao | escol | residencia | sexo |      |      |  |
|-------------|-------|-----------|----------|---------|----------|-------------|------|--------|-------|-------|-------------|-------------|------------|-------------|-------|------|--------|--------|------|--------|------|-------|--------|----------|-----------|------|---------|-----------|-------|------------|------|------|------|--|
| renda       | 1,00  |           |          |         |          |             |      |        |       |       |             |             |            |             |       |      |        |        |      |        |      |       |        |          |           |      |         |           |       |            |      |      |      |  |
| s_devedor   | 0,10  | 1,00      |          |         |          |             |      |        |       |       |             |             |            |             |       |      |        |        |      |        |      |       |        |          |           |      |         |           |       |            |      |      |      |  |
| l_cartao    | 0,30  | 0,20      | 1,00     |         |          |             |      |        |       |       |             |             |            |             |       |      |        |        |      |        |      |       |        |          |           |      |         |           |       |            |      |      |      |  |
| t_conta     | 0,10  | 0,10      | 0,30     | 1,00    |          |             |      |        |       |       |             |             |            |             |       |      |        |        |      |        |      |       |        |          |           |      |         |           |       |            |      |      |      |  |
| poupanca    | 0,00  | 0,10      | 0,10     | 0,10    | 1,00     |             |      |        |       |       |             |             |            |             |       |      |        |        |      |        |      |       |        |          |           |      |         |           |       |            |      |      |      |  |
| cdb         | 0,40  | 0,20      | 0,30     | 0,10    | 0,00     | 1,00        |      |        |       |       |             |             |            |             |       |      |        |        |      |        |      |       |        |          |           |      |         |           |       |            |      |      |      |  |
| prev        | 0,00  | 0,30      | 0,20     | 0,10    | 0,00     | 0,20        | 1,00 |        |       |       |             |             |            |             |       |      |        |        |      |        |      |       |        |          |           |      |         |           |       |            |      |      |      |  |
| fundos      | 0,10  | 0,70      | 0,10     | 0,10    | 0,00     | 0,10        | 0,10 | 1,00   |       |       |             |             |            |             |       |      |        |        |      |        |      |       |        |          |           |      |         |           |       |            |      |      |      |  |
| inv         | 0,30  | 0,70      | 0,30     | 0,10    | 0,10     | <b>0,80</b> | 0,30 | 0,60   | 1,00  |       |             |             |            |             |       |      |        |        |      |        |      |       |        |          |           |      |         |           |       |            |      |      |      |  |
| idade       | 0,10  | 0,10      | 0,20     | 0,60    | 0,10     | 0,10        | 0,10 | 0,10   | 0,10  | 1,00  |             |             |            |             |       |      |        |        |      |        |      |       |        |          |           |      |         |           |       |            |      |      |      |  |
| correntista | 0,00  | 0,00      | 0,10     | 0,20    | 0,00     | 0,00        | 0,00 | 0,00   | 0,00  | 0,20  | 1,00        |             |            |             |       |      |        |        |      |        |      |       |        |          |           |      |         |           |       |            |      |      |      |  |
| risco_hist  | 0,20  | 0,30      | 0,20     | 0,10    | 0,00     | 0,10        | 0,00 | 0,10   | 0,10  | 0,10  | 0,00        | 1,00        |            |             |       |      |        |        |      |        |      |       |        |          |           |      |         |           |       |            |      |      |      |  |
| risco_merc  | 0,10  | 0,20      | 0,10     | 0,00    | 0,00     | 0,00        | 0,00 | 0,00   | 0,00  | 0,00  | 0,00        | 0,60        | 1,00       |             |       |      |        |        |      |        |      |       |        |          |           |      |         |           |       |            |      |      |      |  |
| risco_total | 0,10  | 0,30      | 0,20     | 0,00    | 0,00     | 0,10        | 0,00 | 0,00   | 0,10  | 0,00  | 0,00        | <b>1,00</b> | 0,60       | 1,00        |       |      |        |        |      |        |      |       |        |          |           |      |         |           |       |            |      |      |      |  |
| reneg       | 0,00  | 0,00      | 0,00     | 0,00    | 0,00     | 0,00        | 0,00 | 0,00   | 0,00  | 0,00  | 0,00        | 0,10        | 0,10       | 0,10        | 1,00  |      |        |        |      |        |      |       |        |          |           |      |         |           |       |            |      |      |      |  |
| imob        | 0,10  | 0,30      | 0,10     | 0,00    | 0,00     | 0,00        | 0,00 | 0,00   | 0,00  | 0,00  | 0,00        | <b>0,70</b> | 0,40       | <b>0,70</b> | 0,00  | 1,00 |        |        |      |        |      |       |        |          |           |      |         |           |       |            |      |      |      |  |
| cartao      | 0,20  | 0,10      | 0,70     | 0,20    | 0,00     | 0,20        | 0,10 | 0,10   | 0,20  | 0,10  | 0,10        | 0,20        | 0,10       | 0,20        | 0,00  | 0,10 | 1,00   |        |      |        |      |       |        |          |           |      |         |           |       |            |      |      |      |  |
| cheque      | 0,00  | 0,00      | 0,10     | 0,10    | 0,00     | 0,00        | 0,00 | 0,00   | 0,00  | 0,10  | 0,10        | 0,10        | 0,10       | 0,10        | 0,00  | 0,20 | 1,00   |        |      |        |      |       |        |          |           |      |         |           |       |            |      |      |      |  |
| cp          | 0,00  | 0,00      | 0,10     | 0,00    | 0,00     | 0,00        | 0,00 | 0,10   | 0,00  | 0,00  | 0,00        | 0,20        | 0,10       | 0,20        | 0,00  | 0,00 | 0,10   | 1,00   |      |        |      |       |        |          |           |      |         |           |       |            |      |      |      |  |
| consig      | 0,00  | 0,00      | 0,00     | 0,10    | 0,00     | 0,00        | 0,00 | 0,00   | 0,00  | 0,10  | 0,10        | 0,10        | 0,10       | 0,10        | 0,00  | 0,10 | 0,10   | 0,00   | 1,00 |        |      |       |        |          |           |      |         |           |       |            |      |      |      |  |
| agro        | 0,10  | 0,20      | 0,00     | 0,00    | 0,00     | 0,00        | 0,00 | 0,00   | 0,00  | 0,00  | 0,00        | 0,60        | 0,40       | 0,60        | 0,00  | 0,00 | 0,00   | 0,00   | 0,00 | 1,00   |      |       |        |          |           |      |         |           |       |            |      |      |      |  |
| autos       | 0,00  | 0,00      | 0,10     | 0,00    | 0,00     | 0,00        | 0,00 | 0,00   | 0,00  | 0,00  | 0,00        | 0,10        | 0,00       | 0,10        | 0,00  | 0,00 | 0,10   | 0,10   | 0,00 | 0,00   | 1,00 |       |        |          |           |      |         |           |       |            |      |      |      |  |
| rating      | 0,10  | 0,00      | 0,30     | 0,30    | 0,10     | 0,10        | 0,00 | 0,00   | 0,10  | 0,30  | 0,20        | 0,10        | 0,00       | 0,00        | -0,20 | 0,00 | 0,20   | -0,10  | 0,00 | 0,00   | 0,00 | 0,00  | 1,00   |          |           |      |         |           |       |            |      |      |      |  |
| q_depend    | 0,00  | 0,00      | 0,20     | 0,30    | 0,10     | 0,00        | 0,00 | 0,00   | 0,00  | 0,40  | 0,10        | 0,00        | 0,00       | 0,00        | 0,00  | 0,00 | 0,10   | 0,00   | 0,00 | 0,00   | 0,00 | 0,10  | 1,00   |          |           |      |         |           |       |            |      |      |      |  |
| t_emprego   | 0,00  | 0,00      | 0,00     | 0,30    | 0,10     | 0,00        | 0,00 | 0,00   | 0,00  | 0,40  | 0,00        | 0,00        | 0,00       | 0,00        | 0,00  | 0,00 | 0,00   | 0,00   | 0,00 | 0,00   | 0,00 | 0,10  | 0,20   | 1,00     |           |      |         |           |       |            |      |      |      |  |
| prof        | -0,10 | 0,00      | -0,10    | 0,00    | 0,00     | 0,00        | 0,00 | 0,00   | 0,20  | 0,00  | 0,00        | 0,00        | 0,00       | 0,00        | 0,00  | 0,00 | -0,10  | 0,00   | 0,00 | -0,10  | 0,00 | 0,00  | 0,00   | 0,10     | 1,00      |      |         |           |       |            |      |      |      |  |
| e_civil     | 0,10  | 0,10      | 0,20     | 0,40    | 0,10     | 0,10        | 0,00 | 0,00   | 0,10  | 0,50  | 0,20        | 0,10        | 0,10       | 0,10        | 0,00  | 0,00 | 0,20   | 0,10   | 0,00 | 0,10   | 0,00 | 0,10  | 0,20   | 0,30     | 0,30      | 0,00 | 1,00    |           |       |            |      |      |      |  |
| a_atuacao   | 0,00  | 0,00      | 0,10     | 0,10    | 0,00     | 0,00        | 0,00 | 0,00   | 0,00  | 0,20  | 0,10        | 0,00        | 0,00       | 0,00        | 0,00  | 0,00 | 0,10   | 0,00   | 0,00 | 0,00   | 0,00 | 0,10  | 0,10   | 0,10     | 0,10      | 0,20 | 0,10    | 1,00      |       |            |      |      |      |  |
| escol       | 0,00  | 0,00      | 0,00     | -0,10   | 0,00     | 0,00        | 0,00 | 0,00   | -0,10 | 0,00  | 0,00        | 0,00        | 0,00       | 0,00        | 0,00  | 0,00 | 0,00   | 0,00   | 0,00 | 0,00   | 0,00 | 0,00  | 0,00   | 0,00     | 0,00      | 0,00 | -0,10   | 0,00      | 1,00  |            |      |      |      |  |
| residencia  | 0,00  | 0,00      | -0,10    | -0,10   | 0,00     | 0,00        | 0,00 | 0,00   | -0,20 | -0,10 | 0,00        | 0,00        | 0,00       | 0,00        | 0,00  | 0,00 | -0,10  | 0,00   | 0,00 | 0,00   | 0,00 | 0,00  | -0,10  | -0,10    | -0,10     | 0,00 | -0,20   | -0,10     | 0,00  | 1,00       |      |      |      |  |
| sexo        | 0,00  | 0,00      | 0,10     | 0,10    | 0,00     | 0,00        | 0,00 | 0,00   | 0,10  | 0,00  | 0,00        | 0,00        | 0,00       | 0,00        | 0,00  | 0,00 | 0,10   | 0,00   | 0,00 | 0,00   | 0,00 | 0,00  | 0,10   | 0,10     | 0,10      | 0,00 | 0,10    | 0,00      | 0,10  | 0,00       | 0,00 | 0,00 | 1,00 |  |

Notas: s\_devedor: saldo devedor; l\_cartao: limite do cartão; t\_conta: tempo de conta corrente; inv: investimentos; risco\_hist: risco histórico; risco\_merc: risco no mercado; reneg: crédito com renegociação; imob: crédito imobiliário; cp: crédito pessoal; consig: crédito consignado; agro: crédito em agronegócio; autos: financiamento de automóveis; q\_depend: quantidade de dependentes; t\_emprego: tempo de emprego; prof: profissão; e\_civil: estado civil; a\_atuacao: área de atuação; e\_escol: escolaridade.



São Paulo, 27 a 29 de Julho de 2016

## **Building Knowledge in Accounting**

A ausência de multicolinearidade é uma das premissas para estabelecer um modelo de regressão linear múltipla. A multicolinearidade consiste em um problema comum em regressões, no qual as variáveis independentes possuem relações lineares exatas ou aproximadamente exatas. O indício mais claro da existência da multicolinearidade é quando o  $R^2$  é bastante alto, mas nenhum dos coeficientes da regressão é estatisticamente significativo. Como se pode observar nos resultados da Tabela 2, podemos identificar quatro cruzamentos que possuem correlação de Pearson acima de 0,70 (em negrito). No entanto, optou-se por estimar o modelo completo, com todas as variáveis, visto que posteriormente será estimado o modelo pelo método *Stepwise*, que já corrige problemas de multicolinearidade. Os resultados do modelo completo estão na Tabela 3, a seguir.



Tabela 3. Modelo de Regressão Completo

| Nº Observações          | =       | 89674,00 |         |                              |         |
|-------------------------|---------|----------|---------|------------------------------|---------|
| F (30, 89643)           | =       | 708,68   |         |                              |         |
| Prob > F                | =       | 0,00     |         |                              |         |
| R <sup>2</sup>          | =       | 19,17%   |         |                              |         |
| R <sup>2</sup> ajustado | =       | 19,14%   |         |                              |         |
| renda                   | Coef.   | Estat. t | valor p | Intervalo de Confiança (95%) |         |
| s_devedor               | -0,01   | -23,14   | 0,00    | -0,01                        | -0,01   |
| l_cartao                | 0,15    | 22,49    | 0,00    | 0,14                         | 0,17    |
| t_conta                 | -0,18   | -4,19    | 0,00    | -0,26                        | -0,09   |
| poupanca                | 0,00    | -0,20    | 0,84    | -0,01                        | 0,00    |
| cdb                     | 0,01    | 8,27     | 0,00    | 0,00                         | 0,01    |
| prev                    | -0,01   | -20,74   | 0,00    | -0,01                        | -0,01   |
| fundos                  | 0,00    | -8,29    | 0,00    | 0,00                         | 0,00    |
| inv                     | 0,01    | 20,96    | 0,00    | 0,01                         | 0,02    |
| idade                   | 27,53   | 3,31     | 0,00    | 11,21                        | 43,84   |
| correntista             | 597,78  | 1,99     | 0,05    | 8,50                         | 1187,06 |
| risco_hist              | 0,00    | -0,18    | 0,86    | -0,01                        | 0,01    |
| risco_merc              | 0,00    | 1,59     | 0,11    | 0,00                         | 0,00    |
| risco_total             | 0,01    | 1,06     | 0,29    | -0,01                        | 0,02    |
| reneg                   | 0,06    | 4,94     | 0,00    | 0,04                         | 0,09    |
| imob                    | 0,01    | 4,55     | 0,00    | 0,01                         | 0,02    |
| cartao                  | 0,43    | 23,44    | 0,00    | 0,40                         | 0,47    |
| cheque                  | 0,01    | 0,18     | 0,86    | -0,05                        | 0,06    |
| cp                      | 0,00    | -1,13    | 0,26    | -0,01                        | 0,00    |
| consig                  | 0,05    | 4,02     | 0,00    | 0,02                         | 0,07    |
| agro                    | 0,04    | 14,18    | 0,00    | 0,03                         | 0,04    |
| autos                   | 0,04    | 1,98     | 0,05    | 0,00                         | 0,08    |
| rating                  | 164,08  | 3,39     | 0,00    | 69,11                        | 259,05  |
| q_depend                | -113,75 | -1,36    | 0,17    | -277,27                      | 49,77   |
| t_emprego               | -5,91   | -7,84    | 0,00    | -7,38                        | -4,43   |
| prof                    | -0,18   | -5,73    | 0,00    | -0,24                        | -0,12   |
| e_civil                 | 198,48  | 5,12     | 0,00    | 122,45                       | 274,50  |
| a_atuacao               | -17,21  | -1,04    | 0,30    | -49,53                       | 15,11   |
| escol                   | 32,67   | 0,95     | 0,34    | -34,71                       | 100,05  |
| residencia              | 30,29   | 0,98     | 0,33    | -30,04                       | 90,63   |
| sexo                    | 830,77  | 4,29     | 0,00    | 451,63                       | 1209,92 |
| Constante               | 1538,58 | 2,98     | 0,00    | 527,01                       | 2550,15 |

Notas: s\_devedor: saldo devedor; l\_cartao: limite do cartão; t\_conta: tempo de conta corrente; inv: investimentos; risco\_hist: risco histórico; risco\_merc: risco no mercado; reneg: crédito com renegociação; imob: crédito imobiliário; cp: crédito pessoal; consig: crédito consignado; agro: crédito em agronegócio; autos: financiamento de automóveis;

q\_depend: quantidade de dependentes; t\_emprego: tempo de emprego; prof: profissão;  
e\_civil: estado civil; a\_atuacao: área de atuação; e\_escol: escolaridade.

Como se pode observar, o valor de Prob > F é igual à zero, o que demonstra que o modelo é estatisticamente significativo. Contudo, existem variáveis com p-valor superiores a 0,05. Neste sentido, antes de analisar os resultados, foi estimado o modelo final pelo método *stepwise*, a um nível de 0,05, com correção para heterocedasticidade pelo Método de White. O método *stepwise* identificação modelo ideal a partir de várias iterações das variáveis explicativas que são testadas, considerando variáveis sem problemas de multicolinearidade. O modelo obtido é apresentado na Tabela 4 a seguir.

**Tabela 4. Modelo de Regressão Final estimado pelo Método Stepwise**

|                               |              |                 |                |                                     |         |
|-------------------------------|--------------|-----------------|----------------|-------------------------------------|---------|
| <b>Nº Observações</b>         | =            | 89674,00        |                |                                     |         |
| <b>F (30, 89643)</b>          | =            | 1062,37         |                |                                     |         |
| <b>Prob &gt; F</b>            | =            | 0,00            |                |                                     |         |
| <b>R<sup>2</sup></b>          | =            | 19,16%          |                |                                     |         |
| <b>R<sup>2</sup> ajustado</b> | =            | 19,14%          |                |                                     |         |
| <b>renda</b>                  | <b>Coef.</b> | <b>Estat. t</b> | <b>valor p</b> | <b>Intervalo de Confiança (95%)</b> |         |
| s_devedor                     | 0,00         | -23,20          | 0,00           | 0,00                                | 0,00    |
| l_cartao                      | 0,20         | 28,60           | 0,00           | 0,10                                | 0,20    |
| t_conta                       | -0,20        | -4,30           | 0,00           | -0,30                               | -0,10   |
| prof                          | -0,20        | -6,00           | 0,00           | -0,20                               | -0,10   |
| cdb                           | 0,00         | 8,40            | 0,00           | 0,00                                | 0,00    |
| inv                           | 0,00         | 21,10           | 0,00           | 0,00                                | 0,00    |
| idade                         | 23,20        | 2,90            | 0,00           | 7,40                                | 39,00   |
| autos                         | 0,00         | 2,10            | 0,00           | 0,00                                | 0,10    |
| e_civil                       | 193,00       | 5,00            | 0,00           | 117,90                              | 268,20  |
| rating                        | 174,50       | 3,70            | 0,00           | 81,20                               | 267,70  |
| risco_total                   | 0,00         | 2,70            | 0,00           | 0,00                                | 0,00    |
| imob                          | 0,00         | 5,90            | 0,00           | 0,00                                | 0,00    |
| cartao                        | 0,40         | 25,20           | 0,00           | 0,40                                | 0,50    |
| sexo                          | 820,80       | 4,30            | 0,00           | 444,10                              | 1197,50 |
| consig                        | 0,10         | 4,30            | 0,00           | 0,00                                | 0,10    |
| agro                          | 0,00         | 17,50           | 0,00           | 0,00                                | 0,00    |
| Constante                     | 2275,60      | 5,50            | 0,00           | 1467,40                             | 3083,80 |

Notas: s\_devedor: saldo devedor; l\_cartao: limite do cartão; t\_conta: tempo de conta corrente; inv: investimentos; imob: crédito imobiliário; consig: crédito consignado; agro: crédito em agronegócio; autos: financiamento de automóveis; prof: profissão; e\_civil: estado civil.

Inicialmente, observa-se que diversas variáveis foram excluídas do modelo final, por não apresentarem significância estatística ou apresentarem problemas de multicolinearidade. O modelo final apresentou poder explicativo ( $R^2$ ) de 19,16% e possui as seguintes variáveis:

*S\_devedor(x1)*: a variável “saldo devedor” possui p-valor igual à zero, sendo estatisticamente significativa no modelo e apresentando valor de beta negativo (estatística t negativa, apesar de beta próximo de zero), ou seja, quanto menor o saldo devedor do cliente maior será a sua renda.

*L\_cartao(x2)*: a variável “limite do cartão de crédito” possui p-valor igual à zero, sendo estatisticamente significativa no modelo e apresentando valor de beta positivo, ou seja, quanto maior o limite no cartão de crédito do cliente maior será a sua renda.

*T\_conta(x3)*: a variável “tempo de conta corrente” possui p-valor igual à zero, sendo estatisticamente significativa no modelo e apresentando valor de beta negativo, ou seja, quanto menor o tempo de conta do cliente maior será a sua renda. Essa dinâmica acontece, pois após um ano de cadastro do valor de renda, o sistema da instituição considera a renda desatualizada e a subscreve por uma renda inferida, a qual, como se pode concluir, se apresenta de valor menor.

*Prof(x4)*: a variável “profissão” possui p-valor igual à zero, sendo estatisticamente significativa no modelo e apresentando valor de beta negativo, ou seja, quanto menor o código cadastrado (o que significa que melhor é a profissão em termos de remuneração) maior será a renda do cliente.

*CDB (x5)*: a variável “CDB” possui p-valor igual à zero, sendo estatisticamente significativa no modelo e apresentando valor de beta positivo, ou seja, quanto maior o valor aplicado em CDB, maior será a renda do cliente.

*Inv(x6)*: a variável “Investimentos” possui p-valor igual à zero, sendo estatisticamente significativa no modelo e apresentando valor de beta positivo, ou seja, quanto maior os valores investidos, maior será a renda do cliente.

*Idade(x7)*: a variável “idade” possui p-valor igual à zero, sendo estatisticamente significativa no modelo e apresentando valor de beta positivo, ou seja, quanto maior a idade, maior será a renda do cliente.

*Autos(x8)*: a variável “autos” possui p-valor igual à zero, sendo estatisticamente significativa no modelo e apresentando valor de beta positivo, ou seja, quanto maior o crédito para o financiamento de automóveis, maior será a renda do cliente.

*Estado\_civil(x9)*: a variável “estado civil” possui p-valor igual à zero, sendo estatisticamente significativa no modelo e apresentando valor de beta positivo, ou seja, quanto maior o código de estado civil, maior será a renda do cliente. Essa dinâmica ocorre, pois os menores códigos cadastrados representam indivíduos solteiros.

*Rating(x10)*: a variável “rating” possui p-valor igual à zero, sendo estatisticamente significativa no modelo e apresentando valor de beta positivo, ou seja, quanto maior o rating (melhor capacidade de crédito), maior será a renda do cliente.

*Risco\_total(x11)*: a variável “risco total” possui p-valor igual à zero, sendo estatisticamente significativa no modelo e apresentando valor de beta positivo, ou seja, quanto maior o “risco\_total”, maior será a renda do cliente. Isto ocorre, pois clientes com maior renda tem maior capacidade de crédito, aumentando a sua alavancagem e risco no banco.

*Imob(x12)*: a variável “crédito imobiliário” possui p-valor igual à zero, sendo estatisticamente significativa no modelo e apresentando valor de beta positivo, ou seja, quanto maior o crédito imobiliário, maior será a renda do cliente.

*Cartao(x13)*: a variável “saldo devedor no cartão de crédito” possui p-valor igual à zero, sendo estatisticamente significativa no modelo e apresentando valor de beta positivo, ou seja, quanto maior o saldo devido no cartão, maior será a renda do cliente. Contudo, essa suposição requer atenção na concessão de crédito, pois o saldo devido no cartão representa um indício de que a capacidade de pagamento do cliente foi afetada de alguma maneira, e talvez, sua renda registrada já não expresse a sua situação atual.

*Sexo(x14)*: a variável “sexo” possui p-valor igual à zero, sendo estatisticamente significativa no modelo e apresentando valor de beta positivo, ou seja, se a variável assumir valor igual a 1 (homem) a renda é maior em comparação a categoria 0 (mulheres). Isso revela a disparidade entre salários recebidos por homens e mulheres.

*Consig(x15)*: a variável “crédito consignado” possui p-valor igual à zero, sendo estatisticamente significativa no modelo e apresentando valor de beta positivo, ou seja, quanto maior o crédito consignado, maior será a renda do cliente.

*Agro(x16)*: a variável “crédito em agronegócio” possui p-valor igual à zero, sendo estatisticamente significativa no modelo e apresentando valor de beta positivo, ou seja, quanto maior o crédito em atividades agropecuárias, maior será a renda do cliente.

Sendo assim, o modelo de regressão linear elaborado possui a seguinte configuração:

$$y = 2275,6 - 0,00x_1 + 0,20x_2 - 0,20x_3 - 0,20x_4 + 0,00x_5 + 0,00x_6 + 23,20x_7 + 0,00x_8 + 193,00x_9 + 174,50x_{10} + 0,00x_{11} + 0,00x_{12} + 0,40x_{13} + 820,80x_{14} + 0,10x_{15} + 0,00x_{16}$$

Após a identificação do modelo, fora realizado o cálculo do índice de acertos (Ia), aplicando o modelo obtido em uma amostra segregada da base estudada. A utilização do modelo conferiu um índice de acerto igual a 0,72, ou seja, o modelo possui uma assertividade média de 72% do total de casos.

Considerando que, segundo Picinini *et. al.* (2003) e Selau *et al.* (2009), modelos de *credit scoring* com taxas de acerto acima de 65% são considerados bons, é possível assumir como aceitáveis os resultados encontrados na aplicação da regressão linear aos dados obtidos junto à instituição financeira. Logo, o modelo poderia ser aplicado e possivelmente traria resultados positivos na discriminação dos valores de renda dos indivíduos.

Vale ressaltar ainda que, dentre as 16 variáveis identificadas como significativas, *idade*, *estado civil*, *rating* e *gênero*, foram aquelas que se mostraram com maior destaque ao modelo, por possuírem os maiores betas.

## 5. Considerações Finais

O objetivo deste estudo foi identificar quais os principais fatores determinantes da renda de clientes bancários e, ainda, apresentar um modelo no qual seja possível a estimação da renda. Neste sentido, o trabalho utilizou a técnica de regressão linear e testou dentre 30 variáveis disponíveis quais possuem maior correlação com a renda do cliente e poderiam ser utilizadas na estimação da mesma. Dentre as 30 variáveis testadas, 16 variáveis apresentaram significância estatística após a técnica *stepwise*, sendo *idade*, *estado civil*, *rating* e *gênero* as variáveis mais significativas ao modelo.



São Paulo, 27 a 29 de Julho de 2016

## **Building Knowledge in Accounting**

O modelo desenvolvido possui em sua estrutura as seguintes variáveis: saldo devedor total, limite do cartão de crédito, tempo de conta corrente, profissão, CDB, investimentos mantidos na instituição, idade, financiamento de automóveis, estado civil, rating, risco total mantido na instituição, crédito imobiliário, saldo devedor no cartão de crédito, sexo, crédito consignado, crédito em atividades de agronegócio. Neste sentido é perceptível que, para a construção adequada do modelo, a instituição necessitaria obter junto à outra instituição informações básicas do cliente, como as apresentadas acima. (Adquirir a variável renda e assumi-la como atual, seria possível, porém desvantajoso, tendo em vista a desatualização que a informação pode sofrer, imaginando que o cliente comprove renda apenas uma vez e que as demais informações possam ser atualizadas mensalmente através das atividades do cliente).

Vale ressaltar, porém, que o estudo possui como limitação a necessidade de que o cliente tenha algum histórico no mercado, como também assume um posicionamento estático observado, ou seja, fora assumido na construção do modelo uma estrutura de posicionamento financeiro compatível com a situação econômica daquele período (Novembro de 2014). Segundo o Valor Econômico, em 07/05/2015, em abril de 2015 a caderneta de poupança registrou as maiores retiradas da série histórica acompanhada desde 1995, ou seja, diante do cenário macroeconômico, pode ocorrer uma migração da alocação de recursos, o que pode assim modificar o entendimento das variáveis que possuem maiores correlações com a renda de indivíduos.

Contudo, ainda assim, se tem que a identificação das variáveis poderá incentivar órgãos reguladores, como o Banco Central do Brasil (BCB), a exigirem que indivíduos realizem a apresentação de documentos que possibilitem a identificação do cliente, mas que, além disso, possa proporcionar às instituições a redução dos custos com bancarização, sem a perda do poder de discriminação dos clientes e suas capacidades de comprometimento, sendo assim possível oferecer à sociedade linhas de crédito adequadas. O modelo poderá auxiliar ainda instituições financeiras na identificação da capacidade de pagamento dos clientes e, assim, ampliar a bancarização no Brasil, auxiliando no crescimento e desenvolvimento de uma sociedade cada vez mais economicamente ativa.

### **REFERÊNCIAS**

- ALTMAN, E. *Financial ratios, discriminant analysis and the prediction of corporate bankruptcy*. The journal of finance. v.23, n.4, p.589-609, sept, 1968.
- ASSAF NETO, A.; BRITO, G.A.S. **Modelo de classificação de risco de crédito de empresas**. Revista Cont. Fin, USP São Paulo, v. 19, n. 46, p. 18 – 29, janeiro/abril 2008.
- BANCO CENTRAL DO BRASIL. Resolução CMN 2.025, de 1993, com a redação dada pela Resolução CMN 2.747, de 2000.
- BANCO CENTRAL DO BRASIL. Resolução CMN 2.099, de 1994.
- BRETERNITZ, V. J. **Contribuições ao processo de construção de estratégias para a bancarização da população de baixa renda com o uso de dispositivos móveis**. 2009. Tese (Doutorado em Administração) - Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2009.
- CAOQUETTE, J.B. *et al.* **Gestão de risco de crédito: o grande desafio dos mercados financeiros globais**. 2ª ed. Rio de Janeiro: Qualitymark, SERASA, 2009.
- CERNEV, Adrian Kemmer. **Mobile Banking: fenômeno de inclusão bancária no Brasil**. Revista Inter IT, v. 2, p. 54, 2008.





ELISABETSKY, R. **Um modelo matemático para decisões de crédito no banco comercial.** São Paulo, 1976. 190p. Escola Politécnica da Universidade de São Paulo.

CIAB FEBRABAN 2012. **A sociedade conectada. Setor Bancário em números, tendências tecnológicas e agenda atual.**

GOUVÊA, M. A.; GONÇALVES, E. B.; MANTOVANI, D. M. N. **Aplicação de regressão logística e algoritmos genéticos na análise de risco de crédito.** Revista Universo Contábil, [S.l.], v. 8, n. 2, p. 84-102, abril 2012.

GONÇALVES, E. B.; GOUVÊA, M. A.; MANTOVANI, D. M. N. **Análise de risco de crédito com o uso de regressão logística.** Revista Contemporânea de Contabilidade, Florianópolis, v. 10, n. 20, p. 139-160, ago. 2013.

KANITZ, S. C. **Indicadores contábeis financeiros de previsão de insolvência: a experiência na pequena e média empresa brasileira.** São Paulo, 1976. 187p. Tese (livre docência), Faculdade de Economia, Administração, Contabilidade e Atuária da Universidade de São Paulo.

LIMA, F. G.; PERERA, L. C. J.; KIMURA, H.; SILVA FILHO, A. C. **Aplicação de redes neurais na análise e na concessão de crédito ao consumidor.** Revista de Administração da USP, São Paulo, v. 44, n. 1, p. 34-45, 2009.

\_\_\_\_\_. **Como prever falências.** São Paulo: McGraw-Hill, 1978. 176p.

MÁRIO, P. C. **Contribuição ao estudo da solvência empresarial: uma análise de modelos de previsão - estudo exploratório aplicado em empresas mineiras.** 2002. Dissertação (Mestrado em Controladoria e Contabilidade: Contabilidade) - Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2002.

PICININI, R., OLIVEIRA, G. M. B.; MONTEIRO, L. H. A. **Mineração de critério de *credit scoring* utilizando algoritmos genéticos.** In: SIMPÓSIO BRASILEIRO DE AUTOMAÇÃO INTELIGENTE, 6., 2003, Bauru. Anais... SBAI, 2003. CD-ROM.

ROSA, P. T. M. **Modelos de *Credit Scoring*: Regressão Logística, CHAID e REAL.** 2000. 125 f. Dissertação (Mestrado em Estatística) – Curso de Pós-graduação em Estatística, Instituto de Matemática e Estatística da Universidade de São Paulo, São Paulo, 2000.

SAMEJIMA, K.; DOYA, K.; KAWATO, M. **Inter-module credit assignment in modular reinforcement learning.** Neural Networks, v. 16, n. 7, p. 985-994. 2003.

SANTOS, J. O.; SANTOS, J. A. R. **O modelo kmv e sua utilidade no processo de análise do risco de crédito.** REGE Revista de Gestão, São Paulo, v. 16, n. 2, p. 73-82, jun. 2009.

SELAU, L. P. R.; RIBEIRO, J. L. D. **Uma sistemática para construção e escolha de modelo de previsão de risco de crédito.** Revista Gestão e Produção, v. 16, n. 3, p. 398-413. 2009.

VALOR ECONÔMICO. Poupança tem maior retirada para meses de abril desde 1995. Disponível em <<http://www.valor.com.br/financas/4040088/poupanca-tem-maior-retirada-para-meses-de-abril-desde-1995>>. Acesso em 15/09/2015.

VIRGILLITO, S. B. **Estatística Multivariada na Construção de modelos para análise de risco de crédito e previsão de insolvência de empresas.** Integração (USJT), v. 1, p. 103-198, 2008.