Victor Hugo Delvalle Souza

Estimação de Escores Binomiais Correlacionados: Uma aplicação em Credit Scoring

São Carlos

2008

Victor Hugo Delvalle Souza

Estimação de Escores Binomiais Correlacionados: Uma aplicação em Credit Scoring

Dissertação a ser apresentada ao Departamento de Estatística da Universidade Federal de São Carlos – DEs/UFSCar, como parte dos requisitos para obtenção do título de Mestre em Estatística.

Orientador:
Prof. Dr. Francisco Louzada-Neto

UNIVERSIDADE FEDERAL DE SÃO CARLOS
DEPARTAMENTO DE ESTATÍSTICA
DES/UFSCAR

São Carlos

2008

Ficha catalográfica elaborada pelo DePT da Biblioteca Comunitária da UFSCar

S729EE

Souza, Victor Hugo Delvalle.
Estimação de escores binomiais correlacionados: uma aplicação em Credit Scoring / Victor Hugo Delvalle Souza. -- São Carlos : UFSCar, 2008.
56 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2008.

1. Análise de regressão. 2. Regressão logística. 3. Modelos lineares generalizados. 4. Funções de estimação. 5. Equações de estimação generalizadas. 6. Customer Score. I. Título.

CDD: 519.536 (20ª)

Dedico este trabalho a minha noiva Juliana, e aos meus pais.

Agradecimentos

Agradeço este trabalho em primeiro lugar a Deus, que me deu força para concluir este trabalho.

A minha namorada e companheira de todos os momentos, Juliana Segato, que sempre esteve ao meu lado, me apoiando e ajudando.

Aos meus pais que me deram uma base sólida na educação apesar de todas as dificuldades enfrentadas.

Ao meu orientador Francisco Louzada Neto, pelos ensinamentos e pela dedicação para comigo.

A todos os professores que participaram da minha vida acadêmica e de alguma forma contribuiram na conclusão deste mestrado.

A todos os meus amigos que me apoiaram nesta jornada de infinito aprendizado.

A CAPES, que sem seu apoio financeiro, não teria conseguido concluir este sonho.

Aos meus superiores na Medial Saúde S/A, por toda a compreensão e apoio nas minhas faltas para a conclusão deste trabalho.

Resumo

Em grande parte das modelagens na área de risco de crédito, o modelo mais utilizado é o credit scoring, e como técnica estatística principal a regressão logistica binária, utilizada para decidir se um cliente é bom ou mau pagador. Neste trabalho propomos uma metodologia alternativa, onde a estimativa é feita diretamente nos escores dos clientes, com issa a resposta segue uma distribuição binomial. Nessa modelagem incluimos ainda a estimativa conjunta dos escores de vários produtos utilizados pelos clientes, levando em consideração a correlação existente entre estes escores.

Abstract

For the most part of modelings in the credit risk area, the most widely used model is the credit scoring, and as the main statistical technique, the binary logistic regression, used to determine whether a customer is a good or bad payer. In this academic work an alternative methodology is proposed, where the estimative is formed based on the scores obtained by customers; this means the response follows a binomial distribution. In this modeling the combined estimate of scores of various products used by customers is included, considering the correlation between these scores.

Sumário

Lista de Tabelas

1 Introdução

Nos últimos anos, com o advento do Plano Real e a estabilização da economia no Brasil, a concessão de crédito ao consumidor teve um aumento cada vez maior. Com esse crescimento, as instituições financeiras têm investido cada vez mais nessa fatia do mercado, por consequência, o risco assumido cresce junto. Seguindo essa linha, cada vez mais se faz necessário o uso de procedimentos mais sofisticados para decidir se haverá emprestimo ou não de capital a um proponente.

No início do século XX essa decisão era baseada exclusivamente no julgamento de um ou mais analistas (THOMAS et al., 2002), tornando essa decisão muito subjetiva. No entando com o desenvolvimento da técnica estatística de análise discriminante [1] por Fisher (1936), surgiram os primeiros modelos de credit scoring, modelos esses que permitem a ordenação dos clientes através de um escore, de acordo com a probabilidade dos mesmos pagarem os empréstimos concedidos. No início a substituição da experiência de especialistas por ferramentas estatísticas, ocorreu lentamente, provavelmente por se tratar de um método inovador, mas com o número de propostas cada vez maior, tornou-se inviável as análises individuais. Além disso, com menor custo, maior agilidade até mesmo melhor poder preditivo tornou o credit scoring muito popular sendo
atualmente muito utilizado na modelagem de risco de crédito (HAND; HENLEY, 1997).

Na década de 70 começaram a surgir no Brasil os primeiros modelos e após o

[1] Método em que a partir de características disponíveis de um indivíduo, cria-se uma regra de classificação que permite inferir a que população ele pertence.

Plano Real, com o crescimento na concessão de crédito, houve a difusão da técnica, sendo hoje utilizado pela maioria das instituições financeiras.

1.1 Modelos de concessão de crédito

Os modelos utilizados na concessão de crédito a clientes, por exemplo, a autorização de utilização do cheque especial em um banco, a entrada de uma pessoa num plano de saúde ou a concessão de um cartão de crédito a um cliente de um hipermercado, são chamados de *aplication scoring*, e tem como principal objetivo estimar a probabilidade de um indivíduo se tornar inadimplente após a obtenção do crédito em um determinado tempo, até hoje já foram desenvolvidas várias técnicas para a construção desse tipo de modelo. A mais utilizada é a regressão logística, no entanto, outras metodologias já foram estudadas, tais como: análise discriminante (HAND et al., 1998), regressão linear (ORGLER, 1970), modelos probito (GRABLOWSKY; TALLEY, 1981), árvores de decisão (ARMINGER et al., 1997), programação matemática (HAND, 1981), sistemas especialistas (SHOWERS; CHARKRIN, 1981), redes neurais (WEST, 2000), vizinho mais próximo (HENLEY; HAND, 1997), entres outras. Thomas (2000) concluiu que em relação a discriminação entre bons e maus clientes, não há diferenças significativas entre as técnicas utilizadas. Rosa (2000) e Thomas et al. (2002) descrevem as etapas para o desenvolvimento de um modelo de *application scoring*, onde se estima a probabilidade do cliente se tornar mau e, a partir do valor ajustado, pode-se classificá-lo como bom ou mau cliente. Neste método são utilizadas covariáveis inerentes ao indivíduo, como por exemplo, sexo, idade, renda, entre outras.

Modelos que estimam a probabilidade de um cliente que já possui um determinado produto da instituição ter problema de crédito em um certo tempo são conhecidos por *behavioural scoring*. Estes por sua vez tem uma grande vantagem sobre os modelos de *application scoring* pois, possuem um maior número de variáveis disponíveis para o ajuste, ou seja, além das variáveis disponíveis no momento da concessão, podemos utilizar variáveis comportamentais com relação ao produto analisado, por exemplo, o

tempo de utilização, a intensidade da utilização, etc.

Nos últimos anos, as instituições têm procurado mudar o foco dos produtos para os clientes. Com essa mudança podem ser adotadas estratégias integradas para grupos de clientes com os mesmos produtos. Isso pode evitar ainda que em um banco dois produtos de crédito sejam oferecidos a um mesmo cliente de modo concorrente, o que pode fazer com que o cliente tenha uma má impressão da instituição. O gerenciamento do risco de crédito baseado no foco do cliente também traz inúmeras vantagens, como por exemplo, a prevenção de concessão de um novo produto ou o aumento de limite já existente, para clientes com atraso ou *behavioural score* de alto risco em um outro produto. Com esse crescimento do foco no cliente, surgiu a preocupação em consolidar o risco de crédito do cliente em cada um dos produtos (dados pelos modelos de *behavioural scoring*) em uma única medida, essa modelagem é conhecida por *customer scoring*. Nesse modelo o objetivo é ordenar os clientes quanto a probabilidade de ter problema de crédito, em pelo menos um produto, dentro de um tempo pré-determinado. A grande vantagem dessa ferramenta é permitir uma visão geral do risco do cliente, facilitando assim a criação de políticas de crédito mais adequadas para a instituição. Por exemplo, um banco que possua três modelos de *behavioural scoring* de produto, pode ter grande dificuldade para criar estratégias de gerenciamento de risco de crédito para cada um dos possíveis resultados do vetor de escores do cliente, com o *customer scoring* essa tarefa é simplificada, pois substitui um vetor de três posições por uma única medida.

Entretanto, reportamos que o desenvolvimento de um *customer scoring* não significa que os modelos de *behavioural scoring* tornam-se obsoletos, pois estes podem indicar se o comportamento em algum produto em particular é o responsável pelo alto risco de crédito do cliente dado pelo modelo de *customer scoring*, além disso podem complementar o modelo do cliente na criação de política de crédito da instituição.

Apesar de sua importância, a falta de literatura é presente no desenvolvimento de modelos *customer scoring*. Uma possível explicação está no fato de que uma instituição que desenvolve um eficiente modelo de *customer scoring*, em geral, não tem

interesse em divulgá-lo para evitar que os concorrentes o utilizem e se beneficiem de seu bom desempenho.

Além dos modelos apresentados, existem outros que podem ser utilizados no segmento bancário como o *profit scoring* (OLIVER, 1993), onde o objetivo é ordenar os clientes quanto à probabilidade de dar lucro à instituição (ou ao valor desse lucro), esse modelo se desenvolvido corretamente, traz ganhos extraordinários no gerenciamento do risco de crédito, porém a sua construção é mais difícil do que se imagina em princípio e vários avanços ainda são necessários para que se consiga um modelo adequado. Os modelos de análise de sobrevivência também têm sido discutidos na área de crédito, eles estudam o tempo necessário para que o cliente se torne um problema para a instituição, estes modelos estão diretamente ligados ao *profit scoring* (THOMAS et al., 2001), já que o tempo até o cliente se tornar inadimplente está diretamente associado ao lucro que ele dará a instituição. Além desses modelos, vêm sendo construídos modelos para estimar a probabilidade do cliente pagar um empréstimo que já está em atraso (*collection scoring*), fraudar a instituição (*fraud scoring*)(HENLEY, 1995), comprar um produto após uma campanha de marketing (*propensity scoring*)(TSAI; YEH, 1999) e cancelar a conta ou um produto (*attrition scoring*). As metodologias em geral são as mesmas utilizadas tradicionalmente nos modelos de *application* e *behavioural scoring* típicos.

Neste trabalho nos focaremos na modelagem de *customer scoring*. Thomas et al. (2001) apenas apresentam o objetivo desses modelos, enquanto que McNab e Wynn (2000) discutem rapidamente o conceito, as componentes utilizadas no desenvolvimento, as vantagens e aplicações desses modelos. Já Groom e Gill (1998) discutem diversos aspectos importantes que devem ser observados no desenvolvimento de um modelo de *customer scoring*.

1.2 Motivação deste trabalho

Na maioria dos modelos do tipo *customer scoring* utilizados, os escores dos clientes são obtidos atráves dos escores dos produtos, isso pode ser feito de várias maneiras, por exemplo, escolhe-se como escore de cliente o valor do escore do produto de maior risco, ou o valor do escore do produto que for adquirido há mais tempo, a decisão de qual método utilizar na escolha é própria de cada instituição, neste trabalho optamos por utilizar o escore do produto de maior risco. Porém o objetivo nesse tipo de modelagem é o mesmo que os anteriormente citados, estimar a probabilidade do cliente tornar-se mau ou bom pagador após um tempo pré-fixado, e essa decisão de ser bom ou mau pagador é baseado no escore do cliente, sendo que os escores dos produtos dependem do comportamento da utilização dos mesmos durante um tempo pré-fixado, além de outras variáveis inerentes ao cliente.

Neste trabalho sugerimos uma proposta para modelagem alternativa em que obtemos conjuntamente os escores dos produtos e o escore do cliente. Para essa proposta utilizamos a metodologia de modelos lineares generalizados (MLG) (NELDER; WEDDERBURN, 1972) no caso especial do modelo logístico linear ou regressão logística, porém como nos MLG's não é levado em consideração a correlação entre as variáveis respostas utilizamos uma extensão para as técnicas de estimação dos parâmetros conhecidos como equações de estimação generalizadas (EEG)(LIANG; ZEGER, 1986). Essa técnica permite a estimação de modelos com dependência entre as respostas, quando as variáveis dependentes pertencem à família exponencial.

No Capítulo 2 apresentamos o embasamento teórico que envolve a metodologia proposta neste trabalho, como modelos lineares generalizados, regressão logística, funções de estimação e equações de estimação generalizadas.

No Capítulo 3 temos um estudo de simulação do modelo, com a finalidade de verificar a eficácia na predição dos escores.

Concluimos esse trabalho no capítulo 4 com dois exemplos de aplicação do modelo comparando os resultados obtidos com os modelos usuais de regressão logística.

2 *Metodologia*

Neste capítulo expomos as técnicas utilizadas. Na seção 2.1 são apresentados os Modelos Lineares Generalizados e a Regressão Logística Binomial. Na seção 2.2 temos as Funções de Estimação e as Equações de Estimação Generalizadas. O capítulo é finalizado com a seção 2.3 onde são apresentados alguns comentários finais.

2.1 Regressão Logística

A Regressão Logística (HOSMER; LEMESHOW, 1989) é uma técnica estatística utilizada para estudar a relação entre uma variável categorizada de interesse, e um conjunto de outras variáveis disponíveis no estudo, podendo essas variáveis serem discretas, contínuas, etc. Esse modelo é um caso particular dos modelos lineares generalizados (MCCULLAGH; NELDER, 1989).

2.1.1 Modelos Lineares Generalizados

Os Modelos Lineares Generalizados (MLGs), também denominados modelos exponenciais lineares, foram desenvolvidos por Nelder e Wedderburn (1972). Esta classe de modelos é baseada na família exponencial uniparamétrica, que possui propriedades interessantes para estimação, testes de hipótese e outros problemas de inferência. O MLG é definido por uma distribuição de probabilidade, membro da família exponencial de distribuições, para a variável resposta, um conjunto de variáveis independentes descrevendo a estrutura linear do modelo e uma função de ligação entre a média da

variável resposta e a estrutura linear. Várias distribuições de probabilidade importantes (discretas e contínuas) como normal, gama, Poisson, binomial, normal inversa (ou Gaussiana inversa) etc. são membros da família exponencial. Neste trabalho nos focaremos na distribuição binomial pois equivale à regressão logística. Uma importante característica dos MLGs é a suposição de independência, ou pelo menos de não-correlação, entre as observações. Porém neste trabalho apresetaremos uma técnica que elimina essa suposição extendendo assim os MLGs. De uma forma geral, a estrutura de um MLG é formada por três partes:

1. Uma *componente aleatória*, onde temos um vetor de observações $\mathbf{y} = (y_1, \ldots, y_n)^\top$ referente às realizações das variáveis aleatórias $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$, independentes, com médias $\mu = (\mu_1, \ldots, \mu_n)^\top$, essa parte de um MLG supõe que cada uma das componentes de $\mathbf{Y}$ seguem a mesma distribuição da família exponencial definida por

$$f_Y(y; \theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\} \tag{2.1}$$

onde $a(.), b(.)$ e $c(.)$ são funções conhecidas; $\phi > 0$ é denominado *parâmetro de dispersão* e θ é denominado *parâmetro canônico* que caracteriza a distribuição em (2.1), se ϕ é conhecido, a equação (2.1) representa a família exponencial uniparamétrica indexada por θ. Através de um trabalho algébrico utilizando a função de verossimilhança conseguimos encontrar alguns resultados importantes

$$\mathbb{E}(Y) = \mu = b'(\theta) \tag{2.2}$$

$$\mathrm{Var}(Y) = a(\phi)b''(\theta). \tag{2.3}$$

Da equação (2.2) podemos obter, univocamente, o parâmetro canônico θ.

2. Uma *componente sistemática* composta pela estrutura linear de um modelo de

regressão

$$\eta = \mathbf{X}\beta,$$

onde $\eta = (\eta_1, \ldots, \eta_n)^\top, \beta = (\beta_1, \ldots, \beta_n)$ e $\mathbf{X}$ é uma matriz $n \times p$ $(p < n)$ conhecida e com posto p. A função linear η dos parâmetros desconhecidos β é chamada de *preditor linear*.

3. Uma função monotônica diferenciável $g(.)$, que expressa a média μ do vetor y em função de η,

$$\mu_i = g^{-1}(\eta_i),\ i = 1, \ldots, n$$

denominada *função de ligação*. Algumas distribuições especiais têm uma função de ligação especial que está associada ao preditor linear η, estas ligações são chamadas canônicas e ocorrem quando $\theta = \eta$, onde θ é o parâmetro canônico definido em (2.1).

A matriz $\mathbf{X}$ é definida a partir de variáveis explicativas que podem ser contínuas, qualitativas (ou fatores) e combinações destes (interações) (MCCULLAGH; NELDER, 1989, cap. 3). Para estimar os parâmetros β's, existem diversas técnicas, tais como: estimação - M, Bayesiano, qui-quadrado mínimo e o método de máxima verossimilhança (MV), este último é o mais utilizado nos programas computacionais. O algoritmo de estimação por MV foi desenvolvido por Nelder e Wedderburn (1972) e baseia-se em um método semelhante ao de Newton-Raphson, conhecido como *Método Escore de Fisher*.

2.1.2 Modelo Logístico Linear

O modelo logístico linear, também conhecido como modelo de regressão logística, é um membro da classe dos MLGs servindo de alternativa para analisar respostas binárias através de um conjunto de variáveis explicativas. A relação entre a probabilidade de sucesso π e o conjunto de variáveis explicativas é dada através da função de

ligação logística,

$$g(\pi) = \text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$$

também chamada de logito. Tal relacionamento é sigmoidal, uma vez que a relação entre o $\text{logit}(\pi)$ e a matriz modelo é linear.

Suponha que temos n observações de Bernoulli sob a forma $y_i/m_i, i = 1,\ldots,n$, de modo que $\mathbb{E}(Y_i) = m_i\pi_i$, onde π_i é a probabilidade de sucesso correspondente à $i-$ésima observação. Assim, o *modelo de regressão logística* relaciona π_i com um conjunto de p variáveis explicativas $x_{i1}, x_{i2}, \ldots, x_{ip}$, associado a $i-$ésima observação, sendo expresso por

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}. \tag{2.4}$$

Podemos escrever (2.4) como

$$\pi_i = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip})}, \tag{2.5}$$

ou, denotando-se $\eta_i = \sum_j \beta_j x_{ij}$, de forma mais simples por

$$\pi_i = \frac{e^{\eta_i}}{1+e^{\eta_i}}.$$

Desde que y_i seja uma observação proveniente de uma distribuição binomial com média $m_i\pi_i$, o valor esperado de y_i é $\mathbb{E}(Y_i) = m_i\left(\frac{e^{\eta_i}}{1+e^{\eta_i}}\right)$. As equações (2.4) e (2.5) definem a componente sistemática do modelo de regressão logistica.

Ajuste do modelo

Para ajustarmos o modelo logístico linear é necessário, primeiramente, estimar os $p+1$ parâmetros $\beta_0, \beta_1, \ldots, \beta_p$. Estes parâmetros são estimados através do método de

máxima verossimilhança. Neste caso, a função de verossimilhança é dada por

$$L(\beta) = \prod_{i=1}^{n} \binom{m_i}{y_i} \pi_i^{y_i} (1-\pi_i)^{m_i - y_i}.$$

A função de verossimilhança pode ser considerada função dos parâmetros β's pois esta função depende das probabilidades de sucesso desconhecidas π_i, as quais dependem dos β's através da expressão (2.5). O problema agora é obter os valores $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p$ que maximizam $L(\beta)$ ou, equivalentemente, $\ell(\beta) = \log L(\beta)$, expresso por

$$\begin{aligned} \ell(\beta) &= \sum_{i=1}^{n} \left[\log \binom{m_i}{y_i} + y_i \log(\pi_i) + (m_i - y_i) \log(1-\pi_i) \right] \\ &= \sum_{i=1}^{n} \left[\log \binom{m_i}{y_i} + y_i \eta_i + (m_i - y_i) \log(1+e^{\eta_i}) \right], \end{aligned} \tag{2.6}$$

onde $\eta_i = \sum_{j=0}^{p} \beta_j x_{ij}$ e $x_{i0} = 1$ para todo $i = 1, \ldots, n$. Para tanto, é necessário calcularmos a derivada do logaritmo da função de verossimilhança em relação aos $p+1$ parâmetros desconhecidos β, dada por

$$\frac{\partial \ell(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} y_j x_{ij} - \sum_{i=1}^{n} m_i x_{ij} e^{\eta_i} (1+e^{\eta_i})^{-1}, \; j = 0, 1, \ldots, p.$$

Assim, igualando estas derivadas a zero obtemos um conjunto de $p+1$ equações não-lineares. As estimativas $\hat{\beta}_j$ correspondem à solução deste sistema e podem ser obtidas através do algoritmo do método escore de Fisher.

Bondade de ajuste

Existem diversas estatísticas que medem a discrepância entre as proporções observadas y_i/m_i e as proporções ajustadas $\hat{\pi}_i$. O *desvio* (D) é uma estatística de bondade de ajuste muito utilizada na literatura e baseia-se nas funções de log-verossimilhança maximizada sob o modelo de investigação $\hat{\ell}_p$ e sob o modelo saturado $\tilde{\ell}_n$, sendo ex-

presa por

$$D = 2(\tilde{\ell}_n - \hat{\ell}_p).$$

A partir desta expressão a log-verossimilhança maximizada para o modelo em investigação é dada por

$$\hat{\ell}_p = \sum_{i=1}^{n} \left[\log \binom{m_i}{y_i} + y_i \log(\hat{\pi}_i) + (m_i - y_i) \log(1 - \hat{\pi}_i) \right].$$

No modelo saturado as probabilidades ajustadas são idênticas às proporções observadas $\tilde{\pi}_i = y_i/m_i$. Assim a log-verossimilhança maximizada sob o modelo saturado é dada por

$$\tilde{\ell}_p = \sum_{i=1}^{n} \left[\log \binom{m_i}{y_i} + y_i \log(\tilde{\pi}_i) + (m_i - y_i) \log(1 - \tilde{\pi}_i) \right].$$

Logo, o desvio (D) reduz-se a

$$D = 2 \sum_{i=1}^{n} \left[y_i \log \left(\frac{\tilde{\pi}_i}{\hat{\pi}_i} \right) + (m_i - y_i) \log \left(\frac{1 - \tilde{\pi}_i}{1 - \hat{\pi}_i} \right) \right].$$

Neste trabalho não utilizaremos o desvio (D) para a analisar a bondade de ajuste, mas sim uma validação através de matrizes de confundimento, que será mais detalhada no capítulo posterior.

2.2 Equações de estimação generalizada

Apresentaremos agora as equações de estimação generalizadas (LIANG; ZEGER, 1986), que extendem os modelos lineares generalizados levando em consideração a correlação existente entre as observações, essas equações utilizam a teoria de funções de estimação (ARTES, 1997)(JØRGENSEN; LABOURIAU, 1994).

2.2.1 Funções de estimação

Uma função de estimação é uma função dos dados y e dos parâmetros de interesse θ. Um ponto importante no estudo dessas funções é o estabelecimento de condições que garantam que os estimadores dos parâmetros envolvidos possuam boas propriedades. Em geral, desejamos que os estimadores sejam consistentes e tenham distribuição assintótica conhecida.

Seja $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ um espaço de probabilidade, com $\mathcal{X} \in \Re$ e $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta \subseteq \Re^p\}$, para algum $p \in \mathbb{N}$ (p é um valor fixado referente à dimensão do espaço paramétrico Θ). Por definição, uma função $\psi : \mathcal{X} \times \Theta \to \Re^p$, é uma função de estimação se para cada $\theta \in \Theta$, $\psi(.;\theta) = (\psi_1, \ldots, \psi_p)^\top$ é uma variável aleatória.

Assumindo a existência de uma amostra de n vetores aleatórios independentes $\mathbf{y}_i = (y_{i1}, \ldots, y_{it})^\top$, $i = 1, \ldots, n$, e que a cada vetor esteja associada uma função de estimação ψ_i, estendemos o conceito de função de estimação para a amostra por

$$\Psi_n(\mathbf{y};\theta) = \sum_{i=1}^{n} \psi_i(\mathbf{y}_i;\theta),$$

com dimensão $(p \times 1)$, sendo $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$ um vetor $(N \times 1)$, $N = nt$.

Denotando $\Psi_n(\mathbf{y};\theta)$ por $\Psi_n(\theta)$, apresentamos algumas propriedades das funções de estimação:

1. As funções de estimação $\Psi_n(\theta)$ e $\Phi_n(\theta)$ são *funções de estimação equivalentes* se

   $$\Psi_n(\theta) = C(\theta)\Phi_n(\theta),$$

   sendo $C(\theta)$ uma matriz quadrada não singular e não estocástica, isto é, $\Psi_n(\theta)$ e $\Phi_n(\theta)$ tem as mesmas raízes;

2. A função de estimação $\Psi_n(\theta)$ é uma *função de estimação não viciada* se

   $$\mathbb{E}_\theta(\Psi_n(\theta)) = 0.$$

Se todas as funções de estimação ψ_i, $i = 1, \ldots, n$, são não viciadas, então a função de estimação Ψ_n baseada na amostra de tamanho n também será não viciada;

3. A *matriz de variabilidade* de uma função de estimação não viciada é dada por

$$\mathrm{V}_\Psi(\theta) = \mathbb{E}_\theta(\Psi_n(\theta)\Psi_n^\top(\theta)),$$

e a *matriz de sensibilidade* de uma função de estimação não viciada é dada por

$$\mathrm{S}_\Psi(\theta) = \mathbb{E}_\theta(\frac{\partial}{\partial\theta^\top}\Psi_n(\theta)),$$

e ambas têm dimensão $(p \times p)$.

4. Uma função de estimação $\psi = (\psi_1, \ldots, \psi_p)^\top : \mathcal{X} \times \Theta \to \Re^p$ é dita uma *função de estimação regular* quando as seguintes condições são satisfeitas para todo $\theta \in \Theta$ e para $i, j = 1, \ldots, p$:

i. $\mathbb{E}_\theta(\Psi_n(\theta)) = 0$, isto é, a função é não viciada;

ii. a derivada parcial de $\psi(y, \theta)$ com respeito a θ existe quase certamente para todo $y \in \mathcal{X}$;

iii. é possível permutar o sinal de integração e diferenciação da seguinte forma:

$$\frac{\partial}{\partial\theta_i}\int_{\mathcal{X}} \psi(y;\theta)\mathrm{d}\mathbb{P}_\theta = \int_{\mathcal{X}} \frac{\partial}{\partial\theta_i}\left[\psi(y;\theta)\right]\mathrm{d}\mathbb{P}_\theta;$$

iv. $\mathbb{E}_\theta(\psi_i(\theta)\psi_j(\theta)) \in \Re$ e $\mathrm{V}_\psi(\theta)$ é positiva definida;

v. $\mathbb{E}_\theta\left(\frac{\partial}{\partial\theta_l}\psi_i(\theta)\frac{\partial}{\partial\theta_k}\psi_j(\theta)\right) \in \Re$, com $l, k = 1, \ldots, p$ e $\mathrm{S}_\psi(\theta)$ é não singular.

Seja $\Psi_n(\theta)$ uma função de estimação regular, definimos a *matriz de informação de Godambe* de θ associada a Ψ_n por

$$\mathbf{J}_\Psi = \mathrm{S}_\Psi(\theta)^\top \mathrm{V}_\Psi^{-1}(\theta)\mathrm{S}_\Psi(\theta),$$

com dimensão $(p \times p)$.

A matriz de informação de Godambe $\mathbf{J}_\Psi(\theta)$ é igual a matriz de informação de Fisher se e somente se $\Psi_n(\theta)$ é equivalente à função escore. Por exemplo, se $\Psi_n(\theta)$ é a função escore e, portanto satisfaz as condições para ser uma função de estimação regular, então $\mathrm{S}_\Psi(\theta) = -\mathrm{V}_\Psi(\theta)$, fazendo com que sua matriz de informação de Godambe coincida com a matriz de informação de Fisher. De maneira geral, para todas as funções de estimação regulares que não sejam necessariamente funções escores, a informação de Godambe desempenha o papel da informação de Fisher.

Um conceito importante a ser destacado é o de otimalidade de uma função de estimação regular (GODAMBE, 1960). No caso de θ ser unidimensional, podemos definir uma função de estimação ótima como aquela cujas raízes possuem variância assintótica mínima. Esse conceito pode ser estendido para o caso multidimensional por meio da introdução de alguma ordenação das matrizes de covariância assintóticas (CHANDRASEKAR; KALE, 1984).

Sejam $Q_i(\theta)$ matrizes não estocásticas de postos completos e $u_i = u_i(y_i;\theta)$ vetores com média zero mutuamente independentes, $i = 1,\ldots,n$. A classe das *funções de estimação aditivas* geradas por u_i é definida por

$$\mathcal{L}(u) = \left\{ \Psi_n\theta \in \Re : \Psi_n\theta = \sum_{i=1}^{n} Q_i(\theta)u_i(y_i;\theta) \right\},$$

sendo que $\Re$ contém todas as funções regulares de θ e $u = (u_1^\top,\ldots,u_n^\top)^\top$.

Pode-se mostrar que nessa classe $\mathcal{L}(u)$, a *função de estimação ótima* é dada por (CROWDER, 1987)

$$\Psi_n^*(\theta) = \sum_{i=1}^{n} Q_i^*(\theta)u_i(y_i;\theta), \tag{2.7}$$

em que

$$Q_i^*(\theta) = \mathbb{E}_\theta\left(\frac{\partial u_i}{\partial\theta^\top}\right)^\top \mathrm{Cov}_\theta^{-1}(u_i).$$

É importante ressaltar que se considerarmos $C(\theta)\Psi_n^*(\theta)$, sendo $C(\theta)$ uma matriz

não singular, essa nova função de estimação também será ótima.

Para estabelecer condições para a normalidade assintótica de estimadores obtidos a partir de funções de estimação regulares, em (ARTES, 1997) é enunciado um teorema com sua prova e comentários.

2.2.2 Equações de estimação generalizadas

Na teoria, para o ajuste de modelos nas quais há dependência entre as observações, uma opção a ser utilizada é a análise de quasi-verossimilhança multivariada, no entanto, na prática, ela é uma teoria de difícil utilização por vários motivos. Esses pro-blemas práticos foram resolvidos com o desenvolvimento das equações de estimação generalizadas (EEG) por Liang e Zeger (1986). Elas permitem o ajuste de modelos para as situações nas quais mais de uma observação é tomada em uma mesma unidade amostral, gerando assim uma dependência entre elas. Observações de unidades amostrais diferentes são supostas independentes e a distribuição marginal da resposta pertence à família exponencial.

Seja t_i o número de observações obtidas para o indivíduo i. Defina $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{it_i})^\top$, $i = 1, 2, \ldots, n$ vetores independentes de variáveis aleatórias e assuma que y_{ij} pertence à família exponencial. Seja ainda $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \ldots, x_{ijp})^\top$ vetor de variáveis preditoras para a observação j da unidade amostral i e $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{it_i})^\top$. Admita também que $\mathbb{E}(y_{ij}) = \mu_{ij}$, $\mathrm{Var}(y_{ij}) = \phi^{-1} v(\mu_{ij})$ e $\mathrm{Corr}(\mathbf{y}_i) = \Gamma(u_i)$, defina $\mu_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{it_i})^\top$ e $v(.)$ é uma função positiva. Para facilitar a notação, será assumido, sem perda de generalidade, que $t_i = t$, $i = 1, 2, \ldots, n$.

Para a modelagem de μ_{ij} serão utilizadas as mesmas convenções usadas nos MLGs, isto é,

$$g(\mu_{ij}) = x_{ij}^\top \beta = \eta_{ij},$$

onde $g(.)$ é a função de ligação e β é o vetor de parâmetros.

A função de estimação ótima para β em $\mathcal{L}(y_i - \mu_i)$ é dada por $\Psi_n^*(\beta) = \sum_{i=1}^n \psi_i^*(\beta)$,

na qual

$$\psi_i^*(\beta) = \mathbf{D}_i^\top \mathbf{W}_i^{-1}\mathbf{u}_i = \mathbf{X}_i^\top \mathbf{H}_i \mathbf{W}_i^{-1}\mathbf{u}_i, \tag{2.8}$$

com
$\mathbf{W}_i = \text{Cov}(\mathbf{u}_i) = \phi^{-1}\mathbf{A}_i^{1/2}\Gamma(\mathbf{u}_i)\mathbf{A}_i^{1/2}$,
$\mathbf{A}_i = \text{diag}\{v(\mu_{i1},\ldots,v(\mu_{it}))\}$,
$\mathbf{u}_i = \mathbf{y}_i - \mu_i$ e
$\mathbf{H}_i = \text{diag}\left\{\frac{\partial h(\eta_i)}{\partial \eta_i}\right\}$, $h = g^{-1}$.

Embora a função (2.8) seja ótima entre as lineares geradas por $\mathbf{y}_i - \mu_i$, ela tem pouca utilidade prática. A matriz $\Gamma(\mathbf{u}_i)$ é a verdadeira matriz de correlação de $\mathbf{y}_i$ que, em geral é desconhecida. A solução, para esse problema, encontrada por Liang e Zeger (1986) foi substituir $\Gamma(\mathbf{u}_i)$ pela matriz $R(\alpha)$ denominada matriz de correlação de trabalho, sendo que α é um vetor de dimensão s que caracteriza completamente $R(\alpha)$. Admita, por exemplo, um caso em que $t = 3$ e no qual supõe-se que as correlações entre as variáveis sejam iguais. Tem-se então

$$R(\alpha) = \begin{pmatrix} 1 & \alpha_1 & \alpha_1 \\ \alpha_1 & 1 & \alpha_1 \\ \alpha_1 & \alpha_1 & 1 \end{pmatrix} \quad \text{e} \quad \alpha = [\alpha_1].$$

Caso seja admitido correlações diferentes entre as variaveis, tem-se

$$R(\alpha) = \begin{pmatrix} 1 & \alpha_1 & \alpha_2 \\ \alpha_1 & 1 & \alpha_3 \\ \alpha_2 & \alpha_3 & 1 \end{pmatrix} \quad \text{e} \quad \alpha = [\alpha_1, \alpha_2, \alpha_3]^\top.$$

O termo correlação de trabalho vem do fato de $R(\alpha)$ não precisar, necessariamente, ter a mesma estrutura de correlação de Γ, bastanto ter apenas as propriedades de uma matriz de correlação. O vetor α é tratado como um vetor de parâmetros de perturbação. Liang e Zeger (1986) sugeriram uma alteração em Ψ_n^*, obtendo assim a função de estimação

$$\Psi_n^G(\beta) = \sum_{i=1}^{n} \mathbf{D}_i^\top \hat{\Omega}_i^{-1} \mathbf{u}_i \tag{2.9}$$

em que $\hat{\Omega}_i = \hat{\phi}^{-1}\mathbf{A}_i^{1/2} R(\hat{\alpha})\mathbf{A}_i^{1/2}$ e na qual $\hat{\phi} = \hat{\phi}(\beta)$ e $\hat{\alpha} = \hat{\alpha}(\beta, \hat{\phi}(\beta))$ são estimadores de ϕ e α, respectivamente, que dependem apenas de β. Dessa forma, note que $\Psi_n^G(\beta)$ é função apenas de β. O Teorema a seguir traz as condições sob as quais a raiz de Ψ_n^G é um estimador consistente e assintóticamente normal de β.

Teorema 2.1. *Seja $\hat{\beta}_n$ a raiz de (2.9), sob condições gerais de regularidade, com $|\hat{\beta}_n - \beta| = O_p(1)$ e assumindo que*

a. *$\hat{\alpha}(\beta, \phi^{-1})$ é um estimador $\sqrt{n}-$consistente de α dados β e ϕ^{-1};*

b. *$\hat{\phi}^{-1}(\beta)$ é um estimador $\sqrt{n}-$consistente de ϕ^{-1} dado β;*

c. *$\left|\frac{\partial \hat{\alpha}(\beta,\phi^{-1})}{\partial \phi^{-1}}\right| \leq H(y,\beta)$, na qual $H(y,\beta)$ é uma função $O_p(1)$ de β e dos dados;*

então $\hat{\beta}_n$ é um estimador consistente de β e

$$n^{1/2}(\hat{\beta}_n - \beta) \xrightarrow{\mathcal{D}} \mathcal{N}_P(0, \bar{J}_G^{-1}),$$

quando $n \to \infty$, sendo que

$$\bar{\boldsymbol{J}}_G = \lim_{n\to\infty} \frac{\boldsymbol{J}_{nG}}{n},$$

sendo $\boldsymbol{J}_{nG}$ a matriz de informação de Godambe de β associada a $\Psi_n^G(\beta)$ e dada por

$$\boldsymbol{J}_{nG} = \left\{\sum_{i=1}^{n} \boldsymbol{S}_i\right\} \left\{\sum_{i=1}^{n} \boldsymbol{V}_i\right\}^{-1} \left\{\sum_{i=1}^{n} \boldsymbol{S}_i\right\},$$

sendo que $\boldsymbol{S}_i = -\boldsymbol{D}_i^\top \Omega_i^{-1} \boldsymbol{D}_i$ e $\boldsymbol{V}_i = \boldsymbol{D}_i^\top \Omega_i^{-1} Cov(\boldsymbol{u}_i) \Omega_i^{-1} \boldsymbol{D}_i$.

A prova do Teorema (2.1) está em Liang e Zeger (1986). Note que o teorema não exige que $R(\alpha)$ seja a verdadeira matriz de correlação de $\mathbf{y}_i$. Quando a estrutura de correlação definida pela matriz de correlação de trabalho coincide com a verdadeira estrutura, os estimadores de β terão um aumento de eficiência (LIANG et al., 1992).

Um estimador consistente para $\mathbf{J}_{nG}^{-1}$ é dado por

$$\hat{\mathbf{J}}_{nG}^{-1} = \left\{\sum_{i=1}^{n} \hat{\mathbf{S}}_i\right\}^{-1} \left\{\sum_{i=1}^{n} \hat{\mathbf{D}}_i^{\top} \hat{\Omega}_i^{-1} \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^{\top} \hat{\Omega}_i^{-1} \hat{\mathbf{D}}_i\right\} \left\{\sum_{i=1}^{n} \hat{\mathbf{S}}_i\right\}^{-1}, \tag{2.10}$$

sendo que todas as quantidades são avaliadas no ponto $\hat{\beta}$. Ele é conhecido como estimador sanduíche.

Estimação dos parâmetros

O processo iterativo para calcular $\hat{\beta}_G$ combina o método escore de Fisher para estimar β com o método dos momentos para estimar α e ϕ. Logo, expandindo a EEG dada na equação (2.9) em torno de um valor inicial $\beta_G^{(0)}$, o processo iterativo para estimar β é dado por

$$\begin{aligned} \hat{\beta}_G^{(m+1)} &= \hat{\beta}_G^{(m)} - \left\{\mathbb{E}_\beta\left[\frac{\partial}{\partial \beta^{\top}} \Psi_n^G(\hat{\beta}_G^{(m)})\right]\right\}^{-1} \Psi_n^G(\hat{\beta}_G^{(m)}) = \\ &= \hat{\beta}_G^{(m)} + \left\{\left[\sum_{i=1}^{n} \hat{\mathbf{D}}_i^{\top} \hat{\Omega}_i^{-1} \hat{\mathbf{D}}_i\right]^{-1} \left[\hat{\mathbf{D}}_i^{\top} \hat{\Omega}_i^{-1} (\mathbf{y}_i - \hat{\mu}_i)\right]\right\}^{(m)}, \end{aligned} \tag{2.11}$$

sendo $m = 0, 1, 2, \ldots$ o número de iterações. O índice m no lado direito das equações acima indica que as matrizes e os vetores são atualizados pelas estimativas de β, α e ϕ da $m-$ésima iteração.

Liang e Zeger (1986) utilizam o método dos momentos para estimar os parâmetros de correlação α e o parâmetro de escala ϕ, e os escrevem em função dos resíduos de Peason que é definido como

$$r_{ij} = \frac{y_{ij} - \mu_{ij}}{\sqrt{v(\mu_{ij})}}.$$

Note que $\mathbb{E}(r_{ij}) = 0$ e $\mathrm{Var}(r_{ij}) = \phi^{-1}$. O estimador do resíduo de Pearson para a

observação y_{ij} é dado por

$$\hat{r}_{ij} = \frac{y_{ij} - \hat{\mu}_{ij}}{\sqrt{v(\hat{\mu}_{ij})}}.$$

Assim, se o quarto momento de y_{ij} for finito, pode-se provar que

$$\hat{\phi}^{-1} = \sum_{i=1}^{n} \sum_{j=1}^{t} \frac{\hat{r}_{ij}^2}{nt - p}$$

é um estimado $\sqrt{n}-$consistente de ϕ^{-1} dado β. Observe que $\hat{\phi}^{-1}$ é um estimador da variância de r_{ij} que é igual a ϕ^{-1}.

O procedimento da EEG para estimar β permite que a estrutura de correlação entre as observações da mesma unidade experimental seja especificada de diferentes formas. Vemos algumas delas a seguir:

1. A matriz de correlação padrão uniforme assume que $\text{Corr}(y_{ij}, y_{il}) = \alpha, \forall j \neq l$ e $1 \leq j, l \leq t$. Então a estimativa de α é dada por

$$\hat{\alpha} = \hat{\phi} \sum_{i=1}^{n} \sum_{j>l}^{t} \hat{r}_{ij}\hat{r}_{il} / \left(\frac{1}{2} nt(t-1) - p \right).$$

2. A matriz de correlação $1-$dependente especifica que $\text{Corr}(y_{ij}, y_{i(j+1)}) = \alpha_j$, $j = 1, 2, \ldots, t-1$. Um estimador natural de α_j é dado por

$$\hat{\alpha}_j = \frac{\hat{\phi}}{n - p} \sum_{i=1}^{n} \hat{r}_{ij}\hat{r}_{i(j+1)}.$$

3. A matriz de correlação autoregressiva de primeira ordem, especifica que $\text{Corr}(y_{ij}, y_{il}) = \alpha^{|j-l|}$, $1 \leq j, l \leq t$. Para y_{ij} com distribuição Normal, $\mathbb{E}(\hat{r}_{ij}\hat{r}_{il}) \cong \alpha^{|j-l|}$. Então, α pode ser estimado pelo coeficiente angular da regressão em que a variável dependente é $\log(\hat{r}_{ij}\hat{r}_{il})$ e a independente é $\log|j-l|$.

4. Quando a matriz de correlação é a não estruturada, o elemento (i,j) da matriz $R(\alpha)$ é dado por $R_{ij} = 1$ se $i = j$ e $R_{ij} = \alpha_{ij}$ se $i \neq j$, $R_{ij} = R_{ji}$ e a mesma pode

ser estimada por

$$\hat{R}(\alpha) = \frac{\sum_{i=1}^{n} \hat{r}_i \hat{r}_i^\top}{n-p},$$

na qual $\hat{r}_i = (\hat{r}_{i1}, \ldots, \hat{r}_{it})^\top$.

Etapas para a estimação dos parâmetros β, α e ϕ

1. Estimar β assumindo independencia, isto é, utilizar como valor inicial a estimativa encontrada através de um MLG ordinário;
2. Estimar $r_{ij}, \forall i, j$, dado $\hat{\beta}$;
3. Estimar ϕ dado $\hat{r}_{ij}, \forall i, j$;
4. Definir uma estrutura de correlação a ser utilizada e estimar $R(\alpha)$ dadas estimativas encontradas nos passos anteriores;
5. Reestimar β utilizando as estimativas dos passos anteriores;
6. Repetir os passos 2–5 até que haja convergência.

Análises de diagnóstico

Análises de diagnóstico em EGG são feitas a partir de generalizações de técnicas usualmente utilizadas em modelos lineares generalizados. Venezuela (2003) e Hardin e Hilbe (2003) descrevem algumas dessas técnicas.

2.3 Comentários finais

Neste Capítulo apresentamos as técnicas estatísticas a serem utilizadas para a modelagem de dados financeiros, no Capítulo 3 apresentaremos o tipo de conjunto de dados utilizados para esse tipo de metodologia, além disso aplicaremos a metodologia proposta em dados artificiais, além de uma comparação com o método usual de regressão logística.

3 Estudo de simulação

Neste capítulo apresentaremos um estudo de simulação do modelo apresentado no Capítulo 2, de maneira que possa ser verificado a eficácia da metodologia proposta.

3.1 Geração de Binomiais correlacionadas

Para o estudo de simulação é necessário antes conhecer o procedimento para gerar distribuições binomiais correlacionadas, para isso utilizamos o algoritmo descrito por Park et al. (1996), esse artigo fornece um modo simples de gerar vetores aleatórios com distribuição bernoulli de dimensão arbitrária e não negativamente correlacionados. A partir das distribuições de bernoulli correlacionadas é fácil encontrar distribuições binomiais correlacionadas devido ao seguinte resultado.

Resultado 3.1. *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória com $I\!E(X_i) = \mu_X$ e $Var(X_i) = \sigma_{XX}$, e $Y_1, Y_2, \ldots, Y_n$ outra amostra aleatória com $I\!E(Y_i) = \mu_Y$ e $Var(Y_i) = \sigma_{YY}$, além disso*

$$Cov(X_i, Y_j) = \begin{cases} \sigma_{xy} & se\ i = j, \\ 0 & c.c. \end{cases}$$

e

$$Corr(X_i, Y_j) = \begin{cases} \rho_{xy} & se\ i = j, \\ 0 & c.c. \end{cases}$$

temos que $\rho_{XY} = \frac{\sigma_{xy}}{\sqrt{\sigma_{XX}}\sqrt{\sigma_{YY}}}$.

Agora seja $U = \sum_{i=1}^{n} X_i$ *e* $V = \sum_{j=1}^{n} Y_i$*, com isso temos*

$$\begin{aligned}
Corr(U,V) &= \frac{Cov(U,V)}{\sqrt{Var(U)}\sqrt{(Var(V))}} = \\
&= \frac{Cov(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} Y_j)}{\sqrt{Var(\sum_{i=1}^{n} X_i)}\sqrt{Var(\sum_{j=1}^{n} Y_j)}} = \\
&= \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, Y_j)}{\sqrt{\sum_{i=1}^{n} Var(X_i)}\sqrt{\sum_{j=1}^{n} Var(Y_j)}} = \\
&= \frac{\sum_{i=1}^{n} \sigma_{xy}}{\sqrt{\sum_{i=1}^{n} \sigma_{XX}}\sqrt{\sum_{j=1}^{n} \sigma_{YY}}} = \\
&= \frac{n\sigma_{xy}}{\sqrt{n\sigma_{XX}}\sqrt{n\sigma_{YY}}} = \\
&= \frac{n\sigma_{xy}}{n\sqrt{\sigma_{XX}}\sqrt{\sigma_{YY}}} = \\
&= \rho_{XY}
\end{aligned}$$

Uma distribuição binomial com probabilidade p e tamanho n pode ser escrita como uma soma de n bernoulli's independentes e identicamente distribuídas com probabilidade p, desse modo através do resultado 3.1 podemos mostrar que a correlação entre distribuições binomiais é igual a correlação das distribuições de bernoulli.

3.1.1 Geração de variáveis binárias correlacionadas

A idéia principal do algoritmo Park et al. (1996) consiste na propriedade de que qualquer variável aleatória com distribuição de Poisson pode ser expressa como uma convolução de outras variáveis aleatórias independentes, cada uma com distribuição de Poisson, assim as variáveis binárias são geradas com as correlações desejadas por compartilharem variáveis independentes com distribuição de Poisson.

Inicialmente vamos considerar o caso de 2 variáveis correlacionadas, ou seja, $t = 2$, depois generalizamos para mais variáveis.

Seja $X_i(\lambda)$ com $i = 1,2,3$ variáveis aleatórias mutuamente independentes com distribuição de Poisson com média $\lambda \geq 0$ e sejam U_1 e U_2 variáveis aleatórias definidas por

$$U_1 = X_1(\lambda_1 - \lambda_{12}) + X_3(\lambda_{12}) \tag{3.1}$$

e

$$U_2 = X_2(\lambda_2 - \lambda_{12}) + X_3(\lambda_{12}), \tag{3.2}$$

sendo λ_1, λ_2 e λ_{12} constantes não negativas. Como as variáveis X's são independentes, então U_1 e U_2 têm distribuição de Poisson com médias λ_1 e λ_2, respectivamente, e são não negativamente correlacionadas por causa do termo em comum $X_3(\lambda_{12})$. As variáveis binárias são definidas por

$$Z_1 = \mathbb{1}_{\{0\}}(U_1) \tag{3.3}$$

e

$$Z_2 = \mathbb{1}_{\{0\}}(U_2), \tag{3.4}$$

sendo $\mathbb{1}_{\{.\}}$ a função indicadora tal que

$$\mathbb{1}_A(y) = \begin{cases} 1 & \text{se} \quad y \in A \\ 0 & \text{c.c.} \end{cases}$$

As constantes λ_1, λ_2 e λ_{12} são escolhidas de modo que $\mathbb{E}(Z_1) = \pi_1$, $\mathbb{E}(Z_2) = \pi_2$ e $\text{Corr}(Z_1, Z_2) = \alpha_{12} > 0$. Logo, $\lambda_j = -\ln(\pi_j)$, para $j = 1,2$, pois

$$\mathbb{E}(Z_j) = \mathbb{P}(U_j = 0) = \mathbb{P}(X_j = 0, X_3 = 0) = \mathbb{P}(X_j = 0)\mathbb{P}(X_3 = 0) = e^{-\lambda_j}.$$

Por outro lado, temos que

$$\text{Var}(Z_j) = \pi_j(1 - \pi_j),$$

$$\mathbb{E}(Z_1Z_2) = \mathbb{P}(U_1 = 0, U_2 = 0) = \mathbb{P}(X_1 = 0, X_2 = 0, X_3 = 0) \overset{ind}{=} \pi_1\pi_2 e^{\lambda_{12}}$$

e

$$\text{Cov}(Z_1, Z_2) = \pi_1\pi_2 e^{\lambda_{12}} - \pi_1\pi_2 = \pi_1\pi_2(e^{\lambda_{12}} - 1).$$

Com isso, o coeficiente de correlação entre Z_1 e Z_2 é

$$\alpha_{12} = \frac{\pi_1\pi_2(e^{\lambda_{12}} - 1)}{\sqrt{\pi_1(1-\pi_1)}\sqrt{\pi_2(1-\pi_2)}}.$$

Portanto, devemos escolher

$$\lambda_{12} = \ln\left\{1 + \alpha_{12}\left[\frac{(1-\pi_1)(1-\pi_2)}{\pi_1\pi_2}\right]^{\frac{1}{2}}\right\}. \tag{3.5}$$

O Caso em que $\lambda_{12} = 0$ corresponde à independência entre Z_1 e Z_2.

O algoritmo é dado da seguinte maneira:

1. Dados π_1, π_2 e α_{12}, calcule $\lambda_1 = -\ln(\pi_1)$, $\lambda_2 = -\ln(\pi_2)$ e λ_{12} conforme a expressão (3.5).

2. Gere, independentemente, as variáveis de Poisson X_1 de média $(\lambda_1 - \lambda_{12})$, X_2 de média $(\lambda_2 - \lambda_{12})$ e X_3 de média λ_{12}.

3. Calcule U_1 e U_2 conforme as equações (3.1–3.2) e defina

$$Z_j = \begin{cases} 1 & \text{, se} \quad U_j = 0 \\ 0 & \text{, c.c.} \end{cases} \text{, para } j = 1, 2.$$

Caso geral

Para o caso geral em que temos mais de 2 variáveis correlacionadas ($t > 2$), temos as variáveis aleatórias $U_1, U_2, \ldots, U_t$ com distribuição de Poisson, como somas parciais das variáveis aleatórias independentes $X_1(\gamma_1), X_2(\gamma_2), \ldots, X_m(\gamma_m)$, com distribuição de

Poisson, para algum m inteiro não negativo e $\gamma_1, \gamma_2, \ldots, \gamma_m$ sendo números reais. Alguns X's podem aparecer simultaneamente em vários U's.

O valor esperado e a estrutura de correlação das variáveis binárias $Z_1 = \mathbb{1}_{\{0\}}(U_1)$, $Z_2 = \mathbb{1}_{\{0\}}(U_2), \ldots, Z_t = \mathbb{1}_{\{0\}}(U_t)$ são obtidos controlando o padrão de somas parciais dos X's em dada U e os valores de $\gamma_1, \gamma_2, \ldots, \gamma_m$.

Através do algoritmo a seguir descrevemos como obter m, $\gamma_1, \gamma_2, \ldots, \gamma_m$ e o padrão de somas parciais dados $\mathbb{E}(Z_j) = \pi_j$ e $\mathrm{Cov}(Z_i, Z_j) = \alpha_{ij}, \forall i \neq j, 1 \leq i, j \leq t$. Os passos do algoritmo para gerar mais de duas variáveis aleatórias binárias correlacionadas são:

1. Faça $k = 0$ e para $i \leq j$ e $1 \leq i, j \leq t$, calcule

$$\lambda_{ij} = \ln \left\{ 1 + \alpha_{ij} \left[\frac{(1-\pi_i)(1-\pi_j)}{\pi_i \pi_j} \right]^{\frac{1}{2}} \right\}. \tag{3.6}$$

   Observar que $\lambda_{ij} = \lambda_{ji}$.

2. Faça $k = k+1$. Determine $T_k = \{\lambda_{ij} : \lambda_{ij} > 0, i \leq j \text{ e } 1 \leq i, j \leq t\}$.
   Faça $\gamma_k = \lambda_{rs} = \min\{\lambda_{ij} : \lambda_{ij} \in T_k\}$, isto é, γ_k é menor elemento do conjunto T_k. Se $\lambda_{rr} = 0$ ou $\lambda_{ss} = 0$, então pare porque o algoritmo falhou, caso contrário, escolha um conjunto de índices S_k da seguinte maneira:
   Seja $S_k^0 = \{r, s\}$ e para $i = 1, \ldots, t$, faça

$$S_k^i = \begin{cases} S_k^{i-1} \cup i & \text{, se } \lambda_{ij} > 0 \forall\, j \in S_k^{j-1} \\ S_k^{i-1} & \text{, caso contrário.} \end{cases}$$

   Com isso, $S_k = S_k^t$.

3. Substitua $\lambda_{ij}, i, j \in S_k$, por $\lambda_{ij} - \gamma_k$. Se todo $\lambda_{ij} = 0$, então vá para o próximo passo, caso contrário, volte ao passo anterior.

4. Faça $m = k$ e para $i = 1, \ldots, t$, obtenha

$$\begin{aligned} U_j &= \sum_{k=1}^{m} X_k(\gamma_k) \mathbb{1}_{S_k}(i) \\ &\text{e} \\ Z_j &= \mathbb{1}_0(U_j). \end{aligned}$$

Exemplo de aplicação do algoritmo

Considere o caso que $t = 3$, $\hat{\pi}_1 = 0,9$, $\hat{\pi}_2 = 0,8$, $\hat{\pi}_3 = 0,7$ e as correlações entre as três variáveis são:

$$\mathbf{R}(\hat{\alpha}) = \begin{bmatrix} 1,0 & 0,1 & 0,5 \\ & 1,0 & 0,5 \\ & & 1,0 \end{bmatrix}.$$

A Tabela 3.1 nos mostra para cada k, $k = 1, \ldots, 6$, uma matriz contendo os valores de λ_{ij}, $\forall i \leq j$ e $1 \leq i, j \leq 3$, o par de índices (r,s) e o conjunto S_k. Nas matrizes que contém os valores de λ_{ij}, os números em negrito e sublinhados são os γ_k's e os números em negrito correspondem aos pares de índices (i, j), tais que $i, j \in S_k$.

Os valores de λ_{ij} em $k = 1$ foram calculados por (3.6). Como todos os λ_{ij}'s são maiores do que zero, então $T_1 = \{0,105;\ 0,017;\ 0,104;\ 0,223;\ 0,152;\ 0,357\}$. Em T_1, $\lambda_{12} = 0,017$ é o menor elemento, portanto, $\gamma_1 = 0,017$, $(r,s) = (1,2)$ e $S_1 = \{1,2,3\}$. Atualizando $\lambda_{ij} = \lambda_{ij} - \gamma_1$ para todo $i, j \in \{1,2,3\}$ e $i \leq j$. Os números resultantes são dados em $k = 2$. O conjunto T_2 de números positivos é dado por $T_2 = \{0,088;\ 0,087;\ 0,206;\ 0,135;\ 0,340\}$. O menor elemento de T_2 é $\gamma_2 = \lambda_{13} = 0,087$. Agora, $(r,s) = (1,3)$ e, portanto, $S_2 = \{1,3\}$. Atualizando $\lambda_{ij} = \lambda_{ij} - \gamma_2$ para todo $i, j \in S_2$ e $i \leq j$. Esses resultados são dados em $k = 3$.

O algoritmo continua até que todos os λ_{ij}'s sejam iguais a zero e, no nosso exemplo, ocorreu em $k = 6$ definindo $m = 6$. Notemos que $\lambda_{rr} > 0$ e $\lambda_{ss} > 0$ em cada passo k. Por fim, utilizando γ_k e S_k, $k = 1, \ldots, 6$, obtemos

$$
\begin{aligned}
U_1 &= X_1(0,017) \quad + X_2(0,087) \quad + X_3(0,002), \\
U_2 &= X_1(0,017) \quad\quad\quad\quad\quad\quad + X_4(0,135) \quad + X_5(0,071), \\
U_3 &= X_1(0,017) \quad + X_2(0,087) \quad\quad\quad + X_4(0,135) \quad\quad\quad + X_6(0,118).
\end{aligned}
$$

E, assim, definimos as variáveis binárias por

$$Z_j = \mathbb{1}_{\{0\}}(U_j) \text{ , j=1,2,3.}$$

Tabela 3.1: Valores referentes ao exemplo para gerar variáveis binárias

	Valores de λ_{ij}				
k	$i \leq j$ e $1 \leq i, j \leq 3$			(r,s)	S_k
	0,105	**0,017**	**0,104**		
1		**0,223**	**0,152**	(1,2)	{1,2,3}
			0,357		
	0,088	0,000	**0,087**		
2		0,206	0,135	(1,3)	{1,3}
			0,340		
	0,001	0,000	0,000		
3		0,206	0,135	(1,1)	{1}
			0,253		
	0,000	0,000	0,000		
4		**0,206**	**0,135**	(2,3)	{2,3}
			0,253		
	0,000	0,000	0,000		
5		**0,071**	0,000	(2,2)	{2}
			0,118		
	0,000	0,000	0,000		
6		0,000	0,000	(3,3)	{3}
			0,118		

No nosso caso para gerar um conjunto de 3 variáveis Binomiais correlacionadas Y_1, Y_2 e Y_3 de tamanho n, probabilidades π_1, π_2, π_3 e matriz de correlação $\mathbf{R}(\hat{\alpha})$, basta repetir o algoritmo acima n vezes gerando $Z_{i1}, Z_{i2}, \ldots, Z_{in}, i = 1,2,3$, com probabilidades π_1, π_2, π_3 e matriz de correlação $\mathbf{R}(\hat{\alpha})$ e por fim calcular $Y_i = \sum_{j=1}^{n} Z_{ij}$.

3.2 Apresentação dos dados

O conjunto de dados utilizados neste capítulo e no posterior é composto de uma amostra de dados artificiais com as mesmas características de uma base de dados reais que podem existir em vários setores da economia, tais como varejo, bancos, seguradoras, operadoras de planos de saúde, etc. A amostra foi adequadamente adaptada para aplicação da metodologia proposta.

A amostra de estudo engloba um total de 11000 clientes, os quais são identificados por uma variável de identidade (id), cada cliente possue dois produtos de créditos, produto 1 e produto 2. Um total de 13 variáveis são consideradas no estudo, como vemos na Tabela 3.2.

Tabela 3.2: Variáveis do banco de dados em estudo

Variável	**Descrição**	**Tipo**
id	Identificação do cliente	Discreta positiva
c	Relacionada ao cliente	Binária
t	Relacionada ao cliente	Discreta positiva
y_{12}	Relacionada ao produto 1	Contínua positiva
y_{22}	Relacionada ao produto 2	Discreta positiva
x_{11}	Relacionada ao produto 1	Binária
x_{12}	Relacionada ao produto 1	Discreta positiva
x_{21}	Relacionada ao produto 2	Discreta positiva
w_1	Relacionada ao produto 1	Discreta positiva
w_2	Relacionada ao produto 2	Discreta positiva
e_1	Escore do cliente no produto 1	Discreta variando de 0 a 1000
e_2	Escore do cliente no produto 2	Discreta variando de 0 a 1000
e_3	Escore final do cliente	Discreta variando de 0 a 1000

3.3 Construção do modelo

Para ilustrar a construção do modelo a ser utilizado neste trabalho vamos supor que um instituição possua dois produtos de crédito, por exemplo: no caso de um banco o cartão de crédito e o cheque especial, num comércio de varejo o parcelamento de

compras e o empréstimo pessoal em espécie, numa seguradora as apólices de seguro veicular e residencial, num operadora de planos de saúde os planos médico e odontológico. A cada um desses produtos temos associados 2 variáveis: para o produto 1 temos x_1 sendo por exemplo tempo de utilização do produto, e e_1 a variável resposta escore do cliente associado a esse produto, de maneira análoga para o produto 2 temos x_2 e e_2, além dessas variáveis temos 2 variáveis associadas diretamente ao cliente, x_c e e_c. Fazendo um paralelo as variáveis na tabela 3.2 para o produto 1 temos as variáveis y_{12}, x_{11}, x_{12}, w_1 e e_1, onde este último é o escore do cliente associado, e para o produto 2 temos y_{22}, x_{21}, w_2 e e_2, também este último sendo o escore do cliente e associado ao cliente temos as variáveis c, t e e_3, com este sendo o escore final associado ao cliente.

Voltando ao nosso exemplo, supomos agora que exista uma amostra de n clientes, então as variáveis observáveis para o cliente i, com $i = 1,2,\ldots,n$, são:
$x_{i1}, x_{i2}, x_{ic}, e_{i1}, e_{i2}$ e e_{ic}. Cada escore de produto, recebe um valor numérico entre 0 e 1000 associado ao desempenho do cliente no produto depois de um período de tempo pré-estabelecido, sendo 0 o pior caso e 1000 o melhor, e para o escore do cliente é escolhido o escore de pior desempenho entre os produtos, isto é, se um determinado cliente tem um escore 575 no produto 1 e 724 no produto 2, o escore final passa a ser 575. Assim temos definida a matriz das variáveis como:

Tabela 3.3: Matriz das covariáveis do exemplo

Cliente	Escores			Covariáveis		
i	e_{i1}	e_{i2}	e_{ic}	x_{i1}	x_{i2}	x_{ic}
1	e_{11}	e_{12}	e_{1c}	x_{11}	x_{12}	x_{1c}
2	e_{21}	e_{22}	e_{2c}	x_{21}	x_{22}	x_{2c}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	e_{n1}	e_{n2}	e_{nc}	x_{n1}	x_{n2}	x_{nc}

Como o objetivo é a estimação simultânea dos escores, levando em consideração a correlação existente, utilizamos a técnica de estimação através das EEGs descritas na Seção 2.2, porém para se adequar a metodologia das EEGs temos que mudar o formato da matriz apresentada na tabela 3.3, assim a nova matriz fica da seguinte maneira

Tabela 3.4: Matriz das covariáveis após à adequação

Cliente	Produto	Escore	Covariáveis			Indicadoras	
i	j	E	x_1	x_2	x_c	z_1	z_2
1	1	e_{11}	x_{11}	x_{12}	x_{1c}	1	0
1	2	e_{12}	x_{11}	x_{12}	x_{1c}	0	1
1	c	e_{1c}	x_{11}	x_{12}	x_{1c}	0	0
2	1	e_{21}	x_{21}	x_{22}	x_{2c}	1	0
2	2	e_{22}	x_{21}	x_{22}	x_{2c}	0	1
2	c	e_{2c}	x_{21}	x_{22}	x_{2c}	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	1	e_{n1}	x_{n1}	x_{n2}	x_{nc}	1	0
n	2	e_{n2}	x_{n1}	x_{n2}	x_{nc}	0	1
n	c	e_{nc}	x_{n1}	x_{n2}	x_{nc}	0	0

Podemos observar na Tabela 3.4 duas novas variáveis:

$z_1 = (z_{111}, z_{121}, z_{1c1}, z_{211}, z_{221}, z_{2c1}, \ldots, z_{n11}, z_{n21}, z_{nc1})^\top$

$z_2 = (z_{112}, z_{122}, z_{1c2}, z_{212}, z_{222}, z_{2c2}, \ldots, z_{n12}, z_{n22}, z_{nc2})^\top$

que são definidas como

$$z_{ij1} = \begin{cases} 1 \text{ se o produto } j \text{ do cliente } i \text{ for igual a 1,} \\ 0 \text{ caso contrário,} \end{cases}$$

e

$$z_{ij2} = \begin{cases} 1 \text{ se o produto } j \text{ do cliente } i \text{ for igual a 2,} \\ 0 \text{ caso contrário,} \end{cases}$$

Elas foram criadas para haver diferença entre os valores ajustados para as respostas de cada escore, para cada cliente i.

Definidos assim os novos dados e adicionando as devidas interações entre as variáveis explicativas e as variáveis indicadoras z_1 e z_2, utilizando a notação para GLM apresentada na Seção 2.1.1, temos a seguinte estrutura do modelo,

$$\begin{aligned} g(\mu_{ij}) &= \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + x_{ic}\beta_3 + z_{ij1}\beta_4 + z_{ij2}\beta_5 + \\ &+ x_{i1}z_{ij1}\gamma_1 + x_{i2}z_{ij1}\gamma_2 + x_{ic}z_{ij1}\gamma_3 + \\ &+ x_{i1}z_{ij2}\gamma_4 + x_{i2}z_{ij2}\gamma_5 + x_{ic}z_{ij2}\gamma_6 + \\ &+ x_{i1}x_{i2}z_{ij1}\gamma_7 + x_{i1}x_{ic}z_{ij1}\gamma_8 + x_{i2}x_{ic}z_{ij1}\gamma_9 + \\ &+ x_{i1}x_{i2}z_{ij2}\gamma_{10} + x_{i1}x_{ic}z_{ij2}\gamma_{11} + x_{i2}x_{ic}z_{ij2}\gamma_{12} + \\ &+ x_{i1}x_{i2}x_{ic}z_{ij1}\gamma_{13} + x_{i1}x_{i2}x_{ic}z_{ij2}\gamma_{14} \end{aligned} \tag{3.7}$$

onde

$g(\mu_{ij}) = \mathbb{E}(e_{ij})$;

β_0 é o intercepto do modelo;

$\gamma_k, k = 1, \ldots, 13$ são os parâmetros associados às interações do modelo;

$\beta_i, i = 1, \ldots, 5$ são os parâmetros associados às demais variáveis preditoras.

No contexto dos modelos de regressão logística, temos que $e_{ij} \sim \text{Bin}(\pi_{ij}, 1000)$, pois os escores podem assumir qualquer valor entre 0 e 1000, disso temos que $\mathbb{E}(e_{ij}) = \mu_{ij} = 1000\pi_{ij}$, então a expressão (3.7) torna-se

$$\begin{aligned} \text{logit}(\pi_{ij}) &= \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \\ &= \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + x_{ic}\beta_3 + z_{ij1}\beta_4 + z_{ij2}\beta_5 + \\ &+ x_{i1}z_{ij1}\gamma_1 + x_{i2}z_{ij1}\gamma_2 + x_{ic}z_{ij1}\gamma_3 + \\ &+ x_{i1}z_{ij2}\gamma_4 + x_{i2}z_{ij2}\gamma_5 + x_{ic}z_{ij2}\gamma_6 + \\ &+ x_{i1}x_{i2}z_{ij1}\gamma_7 + x_{i1}x_{ic}z_{ij1}\gamma_8 + x_{i2}x_{ic}z_{ij1}\gamma_9 + \\ &+ x_{i1}x_{i2}z_{ij2}\gamma_{10} + x_{i1}x_{ic}z_{ij2}\gamma_{11} + x_{i2}x_{ic}z_{ij2}\gamma_{12} + \\ &+ x_{i1}x_{i2}x_{ic}z_{ij1}\gamma_{13} + x_{i1}x_{i2}x_{ic}z_{ij2}\gamma_{14} \end{aligned} \tag{3.8}$$

3.4 Estudo de simulação do modelo

Para fazer um estudo de simulação do modelo utilizaremos algumas variáveis dos dados apresentados na seção 3.2, 2 covariáveis relacionadas ao produto 1, **x11** e **w1**, 2 relacionadas ao produto 2, **y22** e **w2**, e a covariável **c** relacionada diretamente ao cliente, além disso utilizamos também as covariáveis indicadoras dos produtos **z1** e **z2**, e uma série de interações entre essas variáveis.

Para a realização da modelagem foi utilizado o software estatístico SAS, versão 9.0, no procedimento PROC GENMOD está implementado a técnica de estimação através das EEGs.

Os coeficientes utilizados na simulação são dados pelos valores apresentados na tabela 3.5.

Tabela 3.5: Valores dos parâmetros para a simulação

Parâmetro	**Valor**	**Parâmetro**	**Valor**	**Parâmetro**	**Valor**
Intercepto	0,27435008	w2*z1	-0,00017057	c*x11*z1	-0,03555592
c	0,15981034	c*z2	-0,10447624	c*y22*z1	0,00042548
x11	-0,03685058	x11*z2	-0,02083361	c*w1*z1	0,00037340
y22	-0,00271120	y22*z2	-0,00078745	c*w2*z1	-0,00037638
w1	0,00157143	w1*z2	-0,00065712	x11*y22*z1	0,00042887
w2	0,00149954	c*x11	-0,10324706	x11*w1*z1	-0,00029867
z1	0,02278724	c*y22	-0,00079312	c*x11*z2	0,03163265
z2	0,17452864	c*w1	-0,00024142	c*y22*z2	0,00046839
c*z1	0,06033653	c*w2	0,00002921	c*w1*z2	0,0000422
x11*z1	-0,00300410	x11*y22	0,00074642	c*w2*z2	0,00026363
y22*z1	0,00067017	x11*w1	-0,00031166	x11*y22*z2	-0,00029652
w1*z1	0,00039123	x11*w2	-0,00040721	x11*w1*z2	0,00028653

Matriz de Correlação

1,0000	0,6108	0,9610
0,6108	1,0000	0,7704
0,9610	0,7704	1,0000

Em primeiro lugar encontramos os valores das probabilidades $\pi_{i1}, \pi_{i2}, \pi_{i3}$ para cada cliente *i* aplicando a expressão (3.8) com as covariáveis e coeficientes apresentados na

Tabela 3.5. Após isso foram geradas 50 amostras aleatórias, cada uma foi gerada da seguinte maneira:

1. Fazemos $j = 1$ e geramos para cada cliente i, 3 variáveis $U_{i1j}, U_{i2j}, U_{i3j}$, com distribuição Bernoulli de probabilidades $\pi_{i1}, \pi_{i2}, \pi_{i3}$ e matriz de correlação apresentada na tabela 3.5;

2. Repetimos o passo anterior para $j = 2, 3, \ldots, 1000$ e por fim calculamos as variáveis E_{i1}, E_{i2}, E_{i3} com distribuição binomial com probabilidades $\pi_{i1}, \pi_{i2}, \pi_{i3}$, tamanho 1000, fazendo $E_{i1} = \sum_{j=1}^{1000} U_{i1j}$, $E_{i2} = \sum_{j=1}^{1000} U_{i2j}$ e $E_{i2} = \sum_{j=1}^{1000} U_{i3j}$, essas variáveis representam os escores observados para cada produto da amostra;

3. Gerada toda a amostra, separamos os dados em uma amostra de aprendizagem com 70% dos dados escolhidos aleatoriamente, e uma amostra teste contendo os 30% restantes dos dados;

4. Fazemos a modelagem como apresentado na seção 3.3 utilizando a amostra de aprendizagem, estimamos assim, os coeficientes e a matriz de correlação de trabalho;

5. Com os parametros estimados calculamos os valores preditos dos escores com a amostra teste, aplicando a equação (2.5);

6. Com isso classificamos os escores observados (EO) em 10 grupos de modo crescente, ou seja, os escores que estiverem no intervalo

$$\min(EO) \vdash \left(\min(EO) + \frac{\max(EO) - \min(EO)}{10} \right)$$

pertencerá ao primeiro grupo, no segundo grupo estarão os escores pertencentes ao intervalo

$$\left(\min(EO) + \frac{\max(EO) - \min(EO)}{10} \right) \vdash \left(\min(EO) + \frac{2(\max(EO) - \min(EO))}{10} \right),$$

e assim por diante. De forma análoga aos escores observados, classificamos também os escores estimados em 10 grupos;

7. Após essa etapa construimos a matriz de confundimento, indicando para cada grupo, quantos clientes estão em cada combinação de grupo observado e estimado, por exemplo, dos escores observados pertencentes ao grupo 1, quantos estimados pertencem ao grupo 1, 2, 3, etc, e assim sucessivamente para os outros grupos de escores observados.

Após esses passos realizados, obtemos um total de 50 matrizes de confundimento para cada produto, com isso calculamos a média dos valores encontrados nessas matrizes e por fim calculamos as porcentagens de acertos dos valores estimados com relação aos valores observados, em cada um dos produtos, esse resultado vemos nas tabelas 3.6, 3.7 e 3.8.

Tabela 3.6: Matriz de confundimento para o escore no produto 1

Obs. / Esp.	Grupos dos escores									
	1	2	3	4	5	6	7	8	9	10
1	**58,10%**	**26,00%**	10,06%	3,40%	1,04%	0,36%	0,10%	0,00%	0,00%	0,00%
2	**26,87%**	**29,94%**	**21,53%**	11,10%	5,67%	2,67%	1,04%	0,40%	0,07%	0,01%
3	10,46%	**22,94%**	**24,49%**	**17,74%**	11,15%	7,11%	4,09%	1,92%	0,60%	0,05%
4	2,97%	11,50%	**17,94%**	**18,94%**	**14,71%**	11,88%	9,43%	7,76%	4,36%	0,76%
5	1,24%	5,98%	13,17%	**18,80%**	**18,94%**	**15,37%**	11,74%	8,76%	5,27%	1,18%
6	0,24%	2,45%	8,14%	15,69%	**20,23%**	**19,41%**	**15,28%**	10,61%	6,50%	1,68%
7	0,10%	0,93%	3,47%	9,73%	16,94%	**20,91%**	**20,98%**	**15,91%**	8,94%	2,02%
8	0,01%	0,21%	0,99%	3,57%	8,08%	14,88%	**22,11%**	**25,77%**	**19,50%**	5,00%
9	0,01%	0,05%	0,17%	0,87%	2,68%	6,31%	13,06%	**23,38%**	**34,59%**	**18,78%**
10	0,01%	0,01%	0,05%	0,17%	0,56%	1,13%	2,16%	5,48%	**20,15%**	**70,52%**
Total Intervalo	**84,97%**	**78,88%**	**63,96%**	**55,48%**	**53,88%**	**55,69%**	**58,37%**	**65,06%**	**74,24%**	**89,30%**
Média	**67,98%**									

Como podemos ver nessas tabelas, para os 3 escores temos um acerto de aproximadamente 70% levando em consideração o erro de apenas um grupo. Outro ponto interessante no resultado é que o modelo foi mais preciso nos grupos mais extremos, com um acerto de mais de 80%, além disso observamos também que a probalidade de um indivíduo que pertença a um grupo de alto risco, receba um escore do grupo de baixo risco é próxima de 0. O mesmo acontece com os indivíduos pertencentes ao grupo de baixo risco, que recebem escore de um grupo de alto risco com probabilidade quase nula.

Tabela 3.7: Matriz de confundimento para o escore no produto 2

Obs. / Esp.	Grupos dos escores									
	1	2	3	4	5	6	7	8	9	10
1	**61,33%**	**21,56%**	8,90%	4,38%	1,98%	0,87%	0,41%	0,34%	0,16%	0,03%
2	**24,90%**	**31,19%**	**21,39%**	12,46%	6,29%	2,41%	0,82%	0,23%	0,13%	0,02%
3	8,40%	**22,28%**	**24,30%**	**20,52%**	14,08%	6,95%	2,78%	0,57%	0,10%	0,15%
4	2,99%	11,59%	**19,35%**	**22,08%**	**20,75%**	14,08%	7,01%	1,80%	0,21%	0,03%
5	1,06%	5,71%	11,63%	**17,29%**	**21,24%**	**21,29%**	14,54%	5,96%	1,21%	0,02%
6	1,12%	5,78%	9,45%	13,35%	**17,51%**	**21,20%**	**17,78%**	10,56%	3,12%	0,36%
7	0,20%	1,65%	4,08%	7,67%	12,55%	**19,72%**	**24,98%**	**19,10%**	8,79%	1,24%
8	0,00%	0,22%	0,89%	2,17%	4,98%	11,16%	**21,87%**	**29,67%**	**22,66%**	6,62%
9	0,00%	0,01%	0,01%	0,07%	0,53%	2,07%	8,61%	**24,99%**	**40,16%**	**23,48%**
10	0,00%	0,00%	0,00%	0,01%	0,07%	0,25%	1,21%	6,77%	**23,47%**	**68,04%**
Total Intervalo	**86,23%**	**75,03%**	**65,04%**	**59,89%**	**59,50%**	**62,21%**	**64,63%**	**73,76%**	**86,29%**	**91,52%**
Média	**72,41%**									

Tabela 3.8: Matriz de confundimento para o escore no cliente

Obs. / Esp.	Grupos dos escores									
	1	2	3	4	5	6	7	8	9	10
1	**62,96%**	**26,68%**	8,12%	1,73%	0,29%	0,02%	0,00%	0,00%	0,00%	0,00%
2	**26,41%**	**34,98%**	**24,30%**	10,34%	3,03%	0,72%	0,15%	0,02%	0,00%	0,00%
3	8,52%	**24,07%**	**29,29%**	**22,40%**	10,56%	3,93%	1,12%	0,34%	0,03%	0,00%
4	1,84%	10,44%	**20,45%**	**23,26%**	**18,06%**	11,72%	7,69%	4,42%	1,83%	0,22%
5	0,21%	2,98%	11,82%	**20,12%**	**21,70%**	**17,16%**	12,31%	8,02%	4,62%	0,98%
6	0,05%	0,68%	4,36%	13,04%	**21,35%**	**22,26%**	**17,75%**	11,97%	6,72%	1,98%
7	0,01%	0,16%	1,24%	6,39%	15,46%	**22,24%**	**24,00%**	**18,49%**	9,68%	2,23%
8	0,01%	0,02%	0,34%	2,15%	7,27%	15,29%	**22,16%**	**26,02%**	**20,86%**	6,04%
9	0,00%	0,01%	0,08%	0,52%	1,97%	5,84%	12,63%	**23,69%**	**34,64%**	**20,50%**
10	0,00%	0,00%	0,01%	0,04%	0,32%	0,82%	2,19%	7,03%	**21,61%**	**68,05%**
Total Intervalo	**89,37%**	**85,73%**	**74,04%**	**65,78%**	**61,11%**	**61,66%**	**63,91%**	**68,20%**	**77,11%**	**88,55%**
Média	**73,55%**									

Esse resultado mostra que o modelo proposto, estima de maneira satisfatória os dados correlacionados, dando força para a aplicação da metodologia para variáveis respostas correlacionadas. É importante ressaltar que a verificação da eficácia do modelo foi feito através de uma validação dos valores estimados como foi apresentado na seção 2.1.2.

3.5 Considerações finais

Neste Capítulo acompanhamos a construção prática do modelo teórico estabelecido no capítulo anterior, vimos também como preparar o conjunto de dados para sua utilização. Além disso, verificamos um estudo de simulação do modelo proposto, nele observamos que a metodologia se aplica bem aos dados gerados. Devido ao tempo de processamento da simulação, o tamanho da amostra foi restrito.

No capítulo 4 veremos uma aplicação da metodologia através de exemplos, além de uma comparação com a metodoligia usual de regressão logística.

4 Aplicação

4.1 Exemplos aplicados

A seguir veremos 2 exemplos aplicados, onde comparamos os resultados obtidos através da modelagem proposta neste trabalho e a metodologia usual de regressão logística considerando os escores não correlacionados.

4.1.1 Exemplo 1

Os dados do exemplo 1 provêm de uma das 50 amostras geradas na simulação apresentada na seção anterior. Da mesma maneira que foi realizada na simulação a amostra foi a separada 2 subamostras: amostra treinamento, contendo 70% dos dados sorteados de modo aleatório e a amostra teste, contendo os 30% restantes.

Realizada a modelagem proposta na amostra de treinamento obtemos as estimativas expostas na tabela 4.1.

A tabela 4.1 mostra a existência de uma forte correlação entre as variáveis respostas, pois variam entre 60% a 95%, com essas estimativas validamos o modelo através das matrizes de confundimento, como mostra a tabela 4.2.

Na tabela 4.3 temos a matriz de confundimento do escore final do cliente quando os escores nos produtos 1 e 2 são estimados independentementes através de regressão logística simples e o escore final é o mínimo dos outros 2,

A partir da modelagem no exemplo 1, verificamos que com os altos valores das cor-

Tabela 4.1: Parâmetros estimados através das EEG's do exemplo 1

Parâmetro	Estimativa	Parâmetro	Estimativa	Parâmetro	Estimativa
Intercepto	0,5384	w2*z1	-0,0003	c*x11*z1	-0,0198
c	-0,0353	c*z2	0,3117	c*y22*z1	-0,0003
x11	-0,0497	x11*z2	0,0005	c*w1*z1	0,0003
y22	-0,0044	y22*z2	0,0028	c*w2*z1	-0,0002
w1	0,0008	w1*z2	0,0006	x11*y22*z1	0,0003
w2	0,0017	c*x11	-0,1035	x11*w1*z1	-0,0005
z1	0,2384	c*y22	-0,0016	c*x11*z2	0,0191
z2	-0,1423	c*w1	0,0001	c*y22*z2	-0,0040
c*z1	0,0451	c*w2	0,0002	c*w1*z2	-0,0006
x11*z1	0,0051	x11*y22	0,0006	c*w2*z2	0,0001
y22*z1	0,0009	x11*w1	0,0001	x11*y22*z2	0,0001
w1*z1	0,0005	x11*w2	-0,0006	x11*w1*z2	-0,0000

Matriz de Correlação de Trabalho		
1,0000	0,6090	0,9552
0,6090	1,0000	0,7733
0,9552	0,7733	1,0000

relações estimadas das variáveis respostas, a metolodogia proposta tem um acerto mais significativo que a metodologia usual de regressão logística, como vemos nas tabelas 4.2 e 4.3 onde a média de acertos levando em consideração um grupo de diferença é de 75% considerando a existência de correlação e de 44% considerando independência dos escores.

Tabela 4.2: Matriz de confundimento para o escore final do cliente do exemplo 1

Obs. / Esp.	Grupos dos escores									
	1	2	3	4	5	6	7	8	9	10
1	**60,34%**	**27,65%**	9,50%	2,23%	0,28%	0,00%	0,00%	0,00%	0,00%	0,00%
2	**25,98%**	**31,72%**	**25,68%**	13,29%	3,02%	0,00%	0,00%	0,30%	0,00%	0,00%
3	10,99%	**26,65%**	**29,12%**	**19,51%**	8,24%	3,02%	2.47%	0,00%	0,00%	0,00%
4	1,83%	9,15%	**18,29%**	**27,74%**	**25,30%**	11,59%	4,27%	1,52%	0,30%	0,00%
5	0,00%	3,93%	12,64%	**18,26%**	**26,97%**	**18,82%**	12,36%	3,93%	2,53%	0,56%
6	0,00%	0.87%	3,18%	10,98%	**17,63%**	**25,14%**	**21,39%**	13,58%	6,94%	0,29%
7	0,00%	0,00%	1,74%	4,94%	8,43%	**22,97%**	**22,97%**	**22,38%**	11,05%	5,52%
8	0,00%	0,00%	0,54%	2,16%	7,03%	14,32%	**21,35%**	**26,49%**	**20,00%**	8,11%
9	0,00%	0,00%	0,00%	1,46%	2,34%	3,51%	12,57%	**23,10%**	**37,13%**	**19,88%**
10	0,00%	0,00%	0,00%	0,29%	1,16%	0,58%	1,74%	8,14%	**21,80%**	**66,28%**
Intervalo de 1	**86,32%**	**86,02%**	**73,09%**	**65,51%**	**69,90%**	**66,93%**	**65,71%**	**71,97%**	**78,93%**	**86,16%**
Média	**75,05%**									

4.1.2 Exemplo 2

No exemplo 2, os dados referem-se aos dados apresentados na 3.2, e da mesma forma que foi realizado no exemplo 1, também foi separa em 2 subamostras, aprendizagem e teste, Após a seleção da amostra aprendizagem temos aproximadamente 8000 clientes selecionados, com um total de 13 variáveis, um resumo destes dados estão na tabela 4.4.

Após a adequação dos dados reorganizamos os dados de maneira semelhante a apresentada na seção 3.3, como podemos ver na tabela 4.5

Tabela 4.3: Matriz de confundimento para o escore final do cliente utilizando Regressão Logística no exemplo 1

Obs. / Esp.	Grupos dos escores									
	1	2	3	4	5	6	7	8	9	10
1	**55,87%**	**22,35%**	7,26%	3,07%	4,75%	2,79%	3,07%	0,84%	0,00%	0,00%
2	**24,17%**	**20,24%**	**16,31%**	11,18%	6,04%	6,04%	10,88%	4,53%	0,60%	0,00%
3	12,36%	**18,68%**	**10,44%**	**11,54%**	7,97%	10,44%	14,84%	9,34%	4,12%	0,27%
4	4,88%	14,02%	**11,28%**	**13,41%**	**8,54%**	8,23%	13,72%	13,41%	10,98%	1,52%
5	1,97%	14,33%	14,89%	**14,04%**	**11,80%**	**8,99%**	10,67%	10,39%	9,83%	3,09%
6	0,00%	4,91%	16,47%	13,29%	**10,69%**	**9,25%**	**9,54%**	12,72%	16,18%	6,94%
7	0,00%	3,49%	9,59%	15,99%	15,70%	**8,72%**	**7,56%**	**11,05%**	15,41%	12,50%
8	0,00%	1,89%	9,46%	10,54%	15,41%	14,32%	**8,92%**	**12,70%**	**14,86%**	11,89%
9	0,00%	0,00%	3,51%	6,14%	11,70%	18,71%	9,06%	**13,45%**	**16,37%**	**21,05%**
10	0,00%	0,00%	0,87%	1,16%	7,27%	12,21%	12,79%	11,05%	**11,63%**	**43,02%**
Intervalo de 1	**80,04%**	**61,27%**	**38,03%**	**38,99%**	**31,07%**	**26,96%**	**26,02%**	**37,20%**	**42,86%**	**64,07%**
Média	**44,65%**									

Tabela 4.4: Amostra de aprendizagem antes da adequação dos dados

id	x11	x12	x21	y12	y22	t	c	w1	w2	e1	e2	ec
1	1	1300	800	114,57	41	301	1	84	32	478	409	409
2	0	0	800	0	90	242	0	108	77	492	494	492
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
11609	1	440	300	67,03	50	14	0	16	14	555	518	518
11610	1	7000	3400	39,52	13	90	1	83	82	665	675	665

Com isso feito aplicamos o procedimento para a estimação dos parâmetros utilizando uma matriz de correlação de trabalho não estruturada, pois, após várias tentativas com os outros tipos de matrizes, não houveram resultados satisfatórios, essa foi a escolhida pois, teve resultados mais consistentes. Após excluir estimativas não significativas do modelo, obtivemos o seguinte resultado:

Tabela 4.5: Adequação dos dados da amostra de aprendizagem

id	x11	x12	x21	y12	y22	t	c	w1	w2	escore	z1	z2
1	1	1300	800	114,57	41	301	1	84	32	478	1	0
1	1	1300	800	114,57	41	301	1	84	32	409	0	1
1	1	1300	800	114,57	41	301	1	84	32	409	0	0
2	0	0	800	0	90	242	0	108	77	492	1	0
2	0	0	800	0	90	242	0	108	77	494	0	1
2	0	0	800	0	90	242	0	108	77	492	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
11609	1	440	300	67,03	50	14	0	16	14	555	1	0
11609	1	440	300	67,03	50	14	0	16	14	518	0	1
11609	1	440	300	67,03	50	14	0	16	14	518	0	0
11610	1	7000	3400	39,52	13	90	1	83	82	665	1	0
11610	1	7000	3400	39,52	13	90	1	83	82	675	0	1
11610	1	7000	3400	39,52	13	90	1	83	82	665	0	0

Como podemos observar na tabela 4.6 a correlação entre as variáveis respostas não é muito forte, com isso podemos verificar o impacto dessa informação na validação para o modelo com hipótese de correlação e o modelo considerando independência. Esses resultados podem ser vistos nas tabelas 4.7 e 4.8.

Tabela 4.6: Parâmetros estimados através das EEG's no exemplo 2

Parâmetro	Estimativa	Parâmetro	Estimativa	Parâmetro	Estimativa
Intercepto	0,27435008	w2*z1	-0,00017057	c*x11*z1	-0,03555592
c	0,15981034	c*z2	-0,10447624	c*y22*z1	0,00042548
x11	-0,03685058	x11*z2	-0,02083361	c*w1*z1	0,00037340
y22	-0,00271120	y22*z2	-0,00078745	c*w2*z1	-0,00037638
w1	0,00157143	w1*z2	-0,00065712	x11*y22*z1	0,00042887
w2	0,00149954	c*x11	-0,10324706	x11*w1*z1	-0,00029867
z1	0,02278724	c*y22	-0,00079312	c*x11*z2	0,03163265
z2	0,17452864	c*w1	-0,00024142	c*y22*z2	0,00046839
c*z1	0,06033653	c*w2	0,00002921	c*w1*z2	0,0000422
x11*z1	-0,00300410	x11*y22	0,00074642	c*w2*z2	0,00026363
y22*z1	0,00067017	x11*w1	-0,00031166	x11*y22*z2	-0,00029652
w1*z1	0,00039123	x11*w2	-0,00040721	x11*w1*z2	0,00028653

Matriz de Correlação de Trabalho		
1,0000	-0,3418	0,3005
-0,3418	1,0000	-0,3664
0,3005	-0,3664	1,0000

Observando a Tabela 4.7, vemos que no escore final do cliente temos um acerto variando entre 30% e 70% levando em consideração o erro de apenas um grupo, isto é, se por exemplo, o escore observado pertence ao grupo 5 e o estimado pertence ao grupo 4 ou 6. Na tabela 4.8 obtemos um resultado utilizando uma regressão logística simples e calculando o mínimo dos escores, e o que observamos é que os resultados são muito semelhantes, isso se deve ao fato da matriz de correlação estimada apresentar valores abaixo de 40% de correlação entre as variáveis respostas.

4.2 Considerações finais

Neste capítulo trabalhamos com duas aplicações, um contendo um forte correlação entre os escores e a outra com fraca correlação, nesses exemplos fizemos uma comparação do modelo proposto com a metodologia usual de regressão logística para estimação dos escores e calculando o mínimo, e verificamos que a metodologia proposta neste trabalho tem resultados equivalentes a metodologia usual quando as correlações

Tabela 4.7: Matriz de confundimento para o escore final do cliente no exemplo 2

Obs. / Esp.	Grupos dos escores									
	1	2	3	4	5	6	7	8	9	10
1	**33,33%**	**21,35%**	18,70%	12,94%	5,51%	4,57%	2,81%	0,91%	0,00%	0,00%
2	**18,10%**	**19,59%**	**15,86%**	15,88%	11,57%	8,86%	5,90%	3,33%	0,84%	0,00%
3	10,92%	**11,99%**	**15,58%**	**14,71%**	14,33%	13,43%	9,83%	5,76%	2,52%	0,87%
4	10,92%	10,53%	**9,92%**	**11,18%**	**11,85%**	14,86%	13,20%	8,48%	7,56%	1,16%
5	10,34%	11,99%	11,05%	**14,12%**	**11,85%**	**10,57%**	9,27%	10,30%	7,00%	3,49%
6	4,89%	8,48%	9,63%	8,82%	**12,95%**	**13,14%**	**14,04%**	12,73%	10,64%	4,65%
7	4,31%	7,31%	8,22%	7,65%	9,64%	**9,71%**	**13,20%**	**17,88%**	13,45%	8,72%
8	4,60%	4,68%	3,97%	10,00%	11,02%	12,86%	**11,80%**	**14,55%**	**14,57%**	12,21%
9	2,30%	2,92%	3,40%	2,35%	6,06%	6,57%	13,48%	**15,15%**	**22,69%**	**25,00%**
10	0,29%	1,17%	3,68%	2,35%	5,23%	5,43%	6,46%	10,91%	**20,73%**	**43,90%**
Intervalo de 1	**51,43%**	**52,93%**	**41,36%**	**40,01%**	**36,65%**	**33,42%**	**39,04%**	**47,58%**	**57,99%**	**68,90%**
Média	**46,93%**									

entre os escores têm valores baixos, enquanto quando temos alta correlação a nova metodologia se mostra mais eficaz que a atualmente utilizada.

Tabela 4.8: Matriz de confundimento para o escore final do cliente utilizando Regressão Logística no exemplo 2

Obs. / Esp.	Grupos dos escores									
	1	2	3	4	5	6	7	8	9	10
1	**34,57%**	**22,35%**	17,42%	11,05%	7,27%	2,79%	2,79%	0,92%	0,28%	0,00%
2	**18,29%**	**22,06%**	**14,61%**	14,45%	11,63%	6,69%	8,38%	3,67%	0,28%	0,29%
3	10,57%	**12,94%**	**12,36%**	**14,73%**	15,99%	14,76%	8,66%	7,95%	2,56%	0,29%
4	10,86%	13,82%	**12,92%**	**13,60%**	**12,79%**	12,81%	10,89%	6,12%	2,56%	1,74%
5	10,57%	7,35%	11,24%	**14,45%**	**15,12%**	**10,58%**	12,29%	8,87%	6,53%	2,62%
6	5,71%	6,18%	8,71%	7,65%	**9,30%**	**14,48%**	**12,85%**	16,51%	13,07%	5,52%
7	3,71%	6,18%	8,43%	9,07%	9,01%	**15,88%**	**11,45%**	**18,65%**	11,93%	6,98%
8	3,14%	5,59%	5,62%	7,93%	7,27%	8,36%	**12,85%**	**13,15%**	**19,32%**	16,57%
9	2,00%	2,35%	6,18%	5,10%	8,72%	7,80%	11,45%	**10,70%**	**22,44%**	**23,26%**
10	0,57%	1,18%	2,53%	1,98%	2,91%	5,85%	8,38%	13,46%	**21,02%**	**42,73%**
Intervalo de 1	**52,86%**	**57,35%**	**39,89%**	**42,78%**	**37,21%**	**40,94%**	**37,15%**	**42,50%**	**62,78%**	**65,99%**
Média	**47,95%**									

5 Conclusão

Ao verificar a aplicação do modelo proposto em dados artificiais, observamos um resultado satisfatório, equivalente ao resultado mostrado pela aplicação da metodologia usual quando a correlação entre os escores têm um valor baixo, porém quando a hipótese de correlação é válida o modelo apresentado se mostra superior aos utilizados normalmente.

Além disso através do estudo de simulação apresentado verificamos que a metodologia proposta, se adequa bem as hipóteses impostas aos dados.

Referências Bibliográficas

ARMINGER, G.; ENACHE, D.; BONNE, T. Analyzing credit risk data: a comparison of logistic discrimination, classification tree analysis, and feedforwad neural networks. *Computational Statistics*, v. 12, p. 293–310, 1997.

ARTES, R. *Extensões da teoria das equações de estimação generalizadas a dados circulares e modelos de dispersão*. Tese (Doutorado) — IME-USP, São Paulo, 1997.

CHANDRASEKAR, B.; KALE, B. K. Unbiased statistical estimation functions in presence of nuisance parameter. *Journal of Statistical Planning and Inference*, v. 9, p. 45–54, 1984.

CORDEIRO, G. M.; LIMA-NETO, E. de A. Modelos paramétricos. Curso apresentado no 16º SINAPE. 2004.

CROWDER, M. On linear and quadratic estimating function. *Biometrika*, v. 74, p. 591–7, 1987.

FISHER, R. A. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, v. 7, p. 179–188, 1936.

GODAMBE, V. P. An optimum property of regular maximum likelihood estimation. *Annals of Mathematical Statistics*, v. 31, p. 1208–11, 1960.

GRABLOWSKY, B. J.; TALLEY, W. K. Probit and discriminant functions for classifying credit applicants: a comparison. *Journal of Economics and Business*, v. 33, p. 254–261, 1981.

GROOM, G.; GILL, L. Customer scoring – practical issues for development success. In: *InterAct98 Conference*. Fair: Isaac and Company Inc., San Francisco, 1998.

HAND, D. J. *Discrimination and Classification*. Chichester: Wiley, 1981.

HAND, D. J.; HENLEY, D. J. Statistical classification methods in consumer credit sacoring: A review. *Journal of the Royal Statistical Society, Series A*, v. 160, p. 523–541, 1997.

HAND, D. J.; OLIVER, J. J.; LUNN, A. D. Discriminant analysis when the classes arise from a cotinuum. *Pattern Recognition*, v. 31, p. 641–650, 1998.

HARDIN, J. W.; HILBE, J. M. *Generalized Estimating Equations*. New York: Chapman and Hall, 2003.

HENLEY, W. E. *Statistical aspets of credit scoring*. Tese (Doutorado) — The Open University, Milton Keynes, 1995.

HENLEY, W. E.; HAND, D. J. Construction of a k-nearest neighbour credit scoring system. *IMA Journal of Mathematics Applied in Business and Industry*, v. 8, p. 143–151, 1997.

HOSMER, D. W.; LEMESHOW, S. *Applied logistic regression*. New York: John Wiley and Sons, 1989.

JOHNSTON, G. *Repeated measures analysis with discrete data using the SAS System*. Cary, NC, 1996.

JØRGENSEN, B.; LABOURIAU, R. S. *Exponential Families and Theoretical Inference*. 1994. Lecture Notes. Department of Statistical. University of British Columbia.

LIANG, K. Y.; ZEGER, S. L. Longitudinal data analysis using generalized linear models. *Biometrika*, v. 73, p. 13–22, 1986.

LIANG, K. Y.; ZEGER, S. L.; QAQISH, B. Multivariate regression analysis for categorical data. *Journal of the Royal Statistical Society*, v. 54, p. 3–40, 1992.

MCCULLAGH, P.; NELDER, J. A. *Generalized Linear Models*. 3. ed. London: Chapman and Hall, 1989.

MCNAB, H.; WYNN, A. *Principles and Practice of Consumer Credit Risk Management*. Kent: Financial World Publishing, 2000.

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. *Journal of the Royal Statistical Society, Series A*, v. 135, p. 370–384, 1972.

OLIVER, R. M. Effects of calibrations and discrination on proffitability scoring. In: UNIVERSITY OF EDINBURG. *Proceedings of Credit Scoring and Credit Control III*. Credit Research Centre, 1993.

ORGLER, Y. E. A credit scoring for comercial loans. *Journal of Money, Credit and Banking*, p. 31–37, november 1970.

PARK, C. G.; PARK, T.; SHIN, D. W. A simple method for generating correlated binary variates. *The American Statistician*, v. 50, p. 306, 1996.

PEREIRA, G. H. de A. *Modelos de risco de crédito de clientes: Uma aplicação a dados reais*. Dissertação (Mestrado) — Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2004.

ROSA, P. T. M. *Modelos de Credit Scoring Regressão Logística Chaid e Real*. Dissertação (Mestrado) — Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2000.

SHOWERS, J. L.; CHARKRIN, L. M. Reducing uncollectable revenue from residential telephone customers. *Interfaces*, v. 11, p. 21–31, 1981.

THOMAS, L. C. A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. *International Journal of Forecasting*, v. 16, p. 149–172, 2000.

THOMAS, L. C.; EDELMAN, D. B.; CROOK, J. N. *Credit Scoring and its Applications*. Philadelphia: Siam, 2002.

THOMAS, L. C.; HO, J.; SCHERER, W. T. Time will tell: behaviour scoring and the dynamics of consumer credit assessment. *IMA Journal of Management Mathematics*, v. 12, p. 89–103, 2001.

TSAI, H. T.; YEH, H. C. A two-stage screening procedure for mailing credit assessmen. *IMA Journal of Mathematics Applied in Bussines and Industry*, v. 10, p. 317–329, 1999.

VENEZUELA, M. K. *Modelos lineares generalizados para análise de dados com medidas repetidas*. Dissertação (Mestrado) — Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2003.

WEST, D. Neural network credit scoring problems. *Computers and Operational Research*, v. 27, p. 1131–1152, 2000.