



**XX Seminário Nacional de Distribuição de Energia Elétrica  
SENDI 2012 - 22 a 26 de outubro  
Rio de Janeiro - RJ - Brasil**

<b>Gustavo Corteletti Venturini</b>	<b>Fabio Fabris</b>	<b>Flavio Miguel Varejao</b>
<b>Universidade Federal do Espírito Santo</b>	<b>Universidade Federal do Espírito Santo</b>	<b>Universidade Federal do Espírito Santo</b>
arodrigues.ufes@gmail.com	fabiofabris@gmail.com	fvarejao@inf.ufes.br

<b>Alexandre Loureiros</b>	<b>Rodrigo Marin Ferro</b>
<b>Universidade Federal do Espírito Santo</b>	<b>Espírito Santo Centrais Elétricas S.A.</b>
arodrigues.ufes@gmail.com	rodrigoferro@edpbr.com.br

**Otimização Amostral para Obtenção da Matriz de Perdas Não-Técnicas**

**Palavras-chave**

Amostragem  
Matriz de Perdas  
Perda não-técnica

**Resumo**

Neste trabalho desenvolveu-se uma metodologia para a geração da matriz de perdas não técnicas com a finalidade de direcionar mais acuradamente ações que objetivam diminuí-las. Para obter-se a situação real de certa população deveria-se inspecionar todo indivíduo pertencente a ela. No entanto limitações temporais e financeiras impedem esta quantidade de vistorias para tamanhos comuns de populações, sendo necessária uma restrição no número amostral. Assim, através de uma análise estatística do universo de consumo, nossa metodologia tenta otimizar a geração da amostragem para que obtenha-se, para um número pré-determinado de visitas, a máxima aproximação da realidade, ou simetricamente, de gerar amostras com certo padrão de confiabilidade que minimize a quantidade de clientes vistoriados.

**1. Introdução**

O consumo de energia elétrica do Brasil vem crescendo ao longo dos últimos anos. O aumento da população

aliado ao cenário econômico favorável resultou em um crescimento no consumo de 4% por ano [1][2], levando eventualmente a um aumento nas perdas de energia elétrica. Esta pode ser dividida em duas grandes fontes: perdas técnicas e não-técnicas. Perdas técnicas são inerentes ao sistema de distribuição e acontecem por fenômenos físicos relacionados ao transporte da energia elétrica que podem ser estimados com um bom grau de confiança por análise matemática.

Perdas não-técnicas ou comerciais são caracterizadas principalmente por consumidores clandestinos e irregulares, falhas nos medidores e falhas no cadastro de iluminação pública. Estas perdas podem ser organizadas em uma matriz que informa para cada fonte de perda, a energia não faturada. Uma Matriz de Perdas representativa informa aos especialistas quais são as fontes críticas de perda de energia elétrica no universo de consumo. Com essa informação é possível direcionar os investimentos de combate à perda comercial de forma mais eficaz. Uma vez que inspecionar todo o universo de consumo é notoriamente inviável, foi desenvolvida uma metodologia baseada em técnicas de amostragem para estimar a Matriz de Perdas.

Ligações clandestinas são ligações feitas por clientes não registrados à concessionária e são de difícil detecção. Perdas por Iluminação pública ocorrem devido a erros de cadastramento e falhas nos relés fotoelétricos. Irregularidades de consumidores cadastrados se dão quando este burla de alguma maneira o equipamento de medição, enquanto perdas por manipulação são registradas quando há fortes indícios de irregularidades mas estas não são encontradas.

## 2. Desenvolvimento

Baseando-nos nestas perdas, dividimos a metodologia de amostragem em duas frentes: uma para perdas oriundas de clientes cadastrados e outra para perdas baseadas em ligações a estações transformadoras. A primeira, que engloba irregularidades e manipulação de consumidores, assim como falhas nos medidores, pode ser feita visitando clientes em uma lista de identificadores, ponderadamente escolhidos a fim de se aumentar estatisticamente a acurácia do resultado. A segunda é feita colhendo resultados em uma listagem de estações transformadoras sorteadas, contando-se o número de ligações clandestinas encontradas em seus ramais e averiguando a conformidade do cadastro dos pontos de iluminação pública e seu funcionamento.

### 2.1 Clientes Cadastrados

Para criar o modelo estatístico, assumimos que os clientes cadastrados estão divididos em  $k$  grupos (chamados *estratos*) segundo algumas características derivadas da base de cadastro e consumo da população sob estudo. Em nosso estudo estamos considerando o grupo (alta, média ou baixa tensão), classe (comercial, residencial, industrial, etc) e faixa de consumo. Os extratos são formados por combinações dessas três características.

O problema de relevância prática e científica é determinar o número de clientes a serem amostrados em cada um dos  $k$  estratos de forma a obter globalmente uma estimativa de maior confiança. Para uma amostra de  $n_1, \dots, n_k$  clientes, em que  $n_j$  ( $j = 1, \dots, k$ ) representa o número de clientes a serem amostrados no estrato  $j$ , a estimativa de máxima verossimilhança [3] do total de energia não faturada devido a irregularidades no estrato  $j$ , digamos  $X_j$ , é estimada por:

$$X_j = N_j \mu_j p_j \quad (1)$$

onde:

$N_j$  - total de clientes cadastrados no estrato

$p_j$  - proporção de fraudadores entre os  $n_j$  clientes amostrados

$\mu_j$  - média da energia não faturada dos clientes detectados como fraudadores no estrato

Sendo portanto  $X$  a perda total estimada nos  $k$  estratos obtido por:

$$X = \sum_{j=1}^k X_j$$

No estimador acima estamos considerando que não existem erros de inspeção. Isto é, todos os clientes amostrados que são fraudadores tem suas irregularidades detectadas. De fato, estudos realizados pela EDP Escelsa/Bandeirante apontaram que a probabilidade de um técnico não detectar uma irregularidade não é desprezível. Sendo  $q_j$  a probabilidade de erro de inspeção no estrato  $j$ , então  $X_j$  é dado por:

$$X_j = N_j \mu_j \left( \frac{k_j}{1 - q_j} \right) \quad (2)$$

em que  $k_j$  é a proporção de irregularidades detectadas pelos técnicos. Note que uma estimativa da proporção geral de clientes com irregularidades é  $k_j/(1 - q_j)$ .

Uma vez definido o modelo para clientes fraudadores é necessário criar uma forma de indicar quais clientes devem ser amostrados para se obter um erro experimental menor que um determinado patamar. Dada uma margem de erro  $E$  e um nível de confiança  $1 - \alpha$ , o objetivo consiste em minimizar o total amostral  $N$  ( $N = n_1 + n_2 + \dots + n_k$ ) de forma que:

$$P(|X_{REAL} - X| < E) = 1 - \alpha \quad (3)$$

isto é, o objetivo é minimizar a soma dos  $n_1, \dots, n_k$  que assegure, com probabilidade  $1 - \alpha$ , que a diferença entre o valor real de energia não faturada e sua estimativa seja menor que a margem de erro. Pode-se expressar a incerteza da estimação em função da variância de  $X$ ,  $Var(X)$ . Com efeito, (3) será satisfeita se:

$$Var(X) = \frac{E^2}{z_{\alpha/2}^2} \quad (4)$$

em que  $z_{\alpha/2}$  é o quantil  $1 - (\alpha/2)$  da distribuição normal padrão. Por exemplo, para um nível de confiança de 95%,  $z_{\alpha/2} = 1,96$ . Assumindo independência entre os estratos, a variância de  $X$  pode ser decomposta como a soma das variâncias de  $X_j$ , isto é:

$$Var(X) = \sum_{j=1}^k Var(X_j)$$

O cálculo do tamanho amostral é portanto um problema de otimização inteira em que o objetivo é encontrar  $n_1, \dots, n_k$  que minimize  $N = n_1 + \dots + n_k$  restrito a:

$$Var(X) = \sum_{j=1}^k Var(X_j) = \sum_{j=1}^k N_j \mu_j \left( \frac{k_j}{1 - q_j} \right) \left( \frac{N_j - n_j}{N_j} \right) \leq \frac{E^2}{z_{\alpha/2}^2}$$

Este problema de otimização pode ser resolvido usando multiplicadores de Lagrange [4]. A solução tem

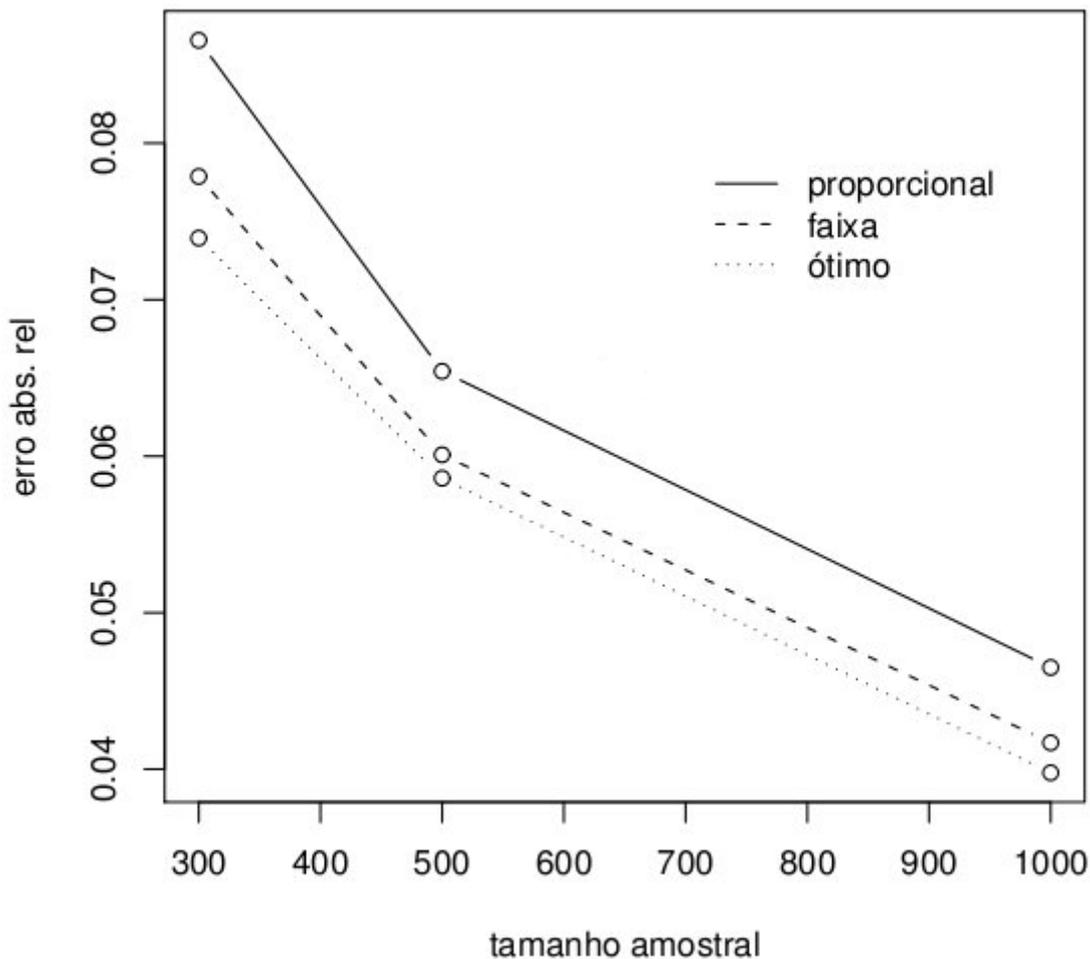
forma fechada e é dada por

$$n_j = \frac{\sqrt{a_j} \sum_{j=1}^k \sqrt{a_j}}{(E/z_{\alpha/2})^2 + \sum_{j=1}^k a_j/N_j} \quad (5)$$

em que  $a_j = N_j^2 \mu_j^2 p_j (1-p_j)$ .

Para determinar o tamanho amostral, os valores de  $\mu_j$  e  $p_j$ , a média de energia não faturada por fraudador e o percentual de clientes fraudadores no estudo  $j$ , devem ser conhecidos a priori. Em nossa aplicação usamos dados históricos da EDP Escelsa/Bandeirante para estimar estes parâmetros.

Para uma dada margem de erro  $E$  e nível de confiança  $1-\alpha$ , tomar  $n_j$  conforme (5) minimiza o tamanho amostral  $N$ . Analogamente, para um valor fixo de  $N$ , a distribuição das unidades da amostra pelos estratos de forma proporcional a (5) minimiza a margem de erro  $E$  (distribuição ótima). Dantas [5] sugere a distribuição das  $N$  amostras de forma proporcional ao consumo médio dos estratos. Outra proposta comumente usada é a distribuição de forma proporcional ao número de clientes por estrato. A Figura 1 mostra um estudo simulado comparando as distribuições ótimas determinadas pela equação (5), proporcionais ao número de clientes e proporcionais ao consumo médio dos estratos. Note que, de fato, a alocação gerada por (5) produz estimativas com erros menores.



**FIGURA 1 - Erro relativo absoluto entre o valor real de energia não faturada e suas estimativas provenientes de distribuições amostrais proporcionais ao número de clientes por estratos (linha cheia), a média de consumo por estratos (linha tracejada) e a distribuição ótima (linha pontilhada)**

Um modelo similar foi criado para representar o comportamento dos clientes em relação a manipulação. Finalmente, assumimos que o percentual de perda de energia devido a erro de medição de cada cliente segue uma distribuição Gaussiana com média  $m$  e variância  $z^2$ .

Foi atestado estatisticamente que a quantidade de clientes necessária para obter um erro aceitável para medidores é sempre inferior à quantidade de clientes necessária para obter o mesmo erro para clientes fraudadores. Portanto a equação (5) também é utilizada para determinar os clientes que terão os medidores aferidos.

## 2.2 Estações Transformadoras (ET)

Para modelar o número de ligações clandestinas usamos uma distribuição de probabilidades que é composta pela mistura de uma variável Bernulli e uma variável com distribuição Poisson. Este modelo é denominado modelo de Poisson com Zero-Inflado [6]. Desta forma, o número de ligações clandestinas em uma determinada ET é 0 (zero) com probabilidade  $p_{ET}$  e segue uma distribuição de Poisson com média  $m_{ET}$  com probabilidade  $1 - p_{ET}$ .

Para uma amostra de  $n$  ET's selecionadas aleatoriamente, uma estimativa do total de ligações clandestinas, digamos  $C$  é dada por:

$$C = N_{ET} \bar{X}_{ET} \quad (6)$$

Isto é, o número total de ET's vezes o número médio de ligações clandestinas nas  $N$  ET's inspecionadas. A perda total estimada pode ser calculada multiplicando-se esse número pelo consumo médio conhecido de um clandestino. Uma distribuição de Poisson com Zero-Inflada também foi usada para modelar o número de pontos de iluminação pública (IP) acesos (ou apagados) 24 horas e pontos de IP não cadastrados. A expansão é feita de forma similar ao estimador de clandestinos.

O desempenho dos estimadores apresentados nesta seção será avaliada através de um estudo simulado apresentado na Seção 2.3. A quantidade de trafos  $nE$  que devem ser inspecionados para um determinado erro  $E$  pode ser calculada de forma trivial a partir da seguinte equação

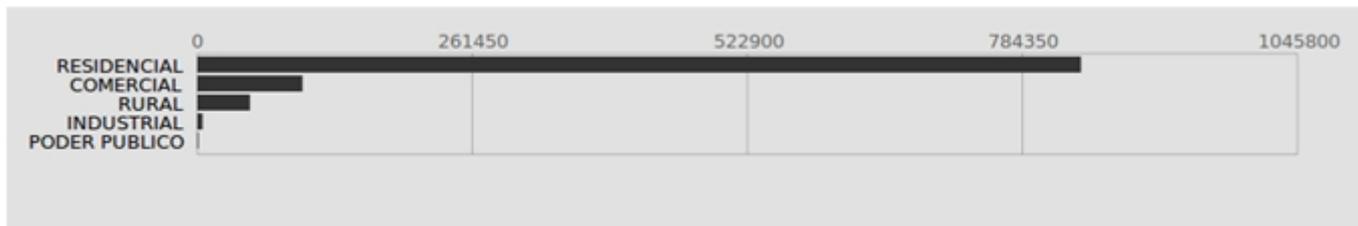
$$nE = \left( \frac{1.64}{\mu \times (1 - p) \times E} \right)^2 \times A \quad (7)$$

onde  $A = (m_{ET}^2 p_{ET} + m_{ET})(1 - p_{ET})$  é a variância da distribuição aleatória, As constantes foram determinadas a partir de inspeções piloto realizadas pela EDP Escelsa/Bandeirante no estado do Espírito Santo.

## 2.3 Experimentos

Simulamos um universo de consumo baseado nos modelos descritos pelas seções precedentes. Para avaliar a metodologia proposta o processo de amostragem de clientes cadastrados e estações transformadoras foi replicado cem vezes. Isto é, cada réplica representa uma amostra completa de clientes e ET's. Em cada réplica são selecionados aleatoriamente 2011 clientes e 400 ET's.

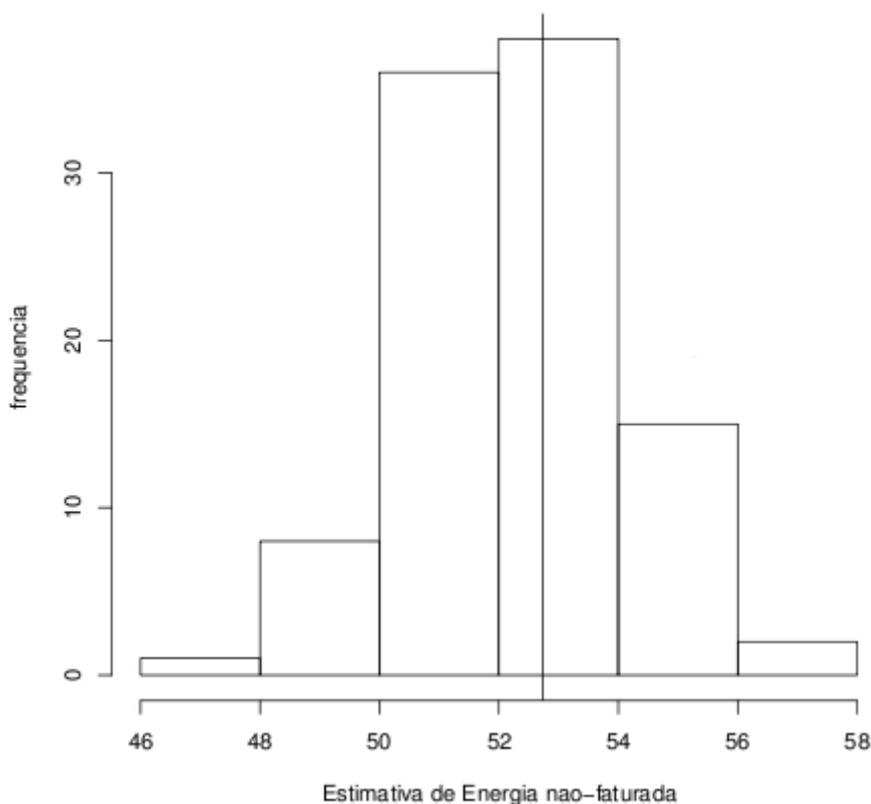
A base de dados construída possui cerca de 1 milhão de clientes e 20 mil estações transformadoras, com uma energia faturada de 192 GWh. A Figura 2 mostra a distribuição simulada dos dados nas classes, a grande maioria dos clientes se encontra na classe residencial, com baixa média de consumo.



**FIGURA 2 - Distribuição dos clientes na base criada**

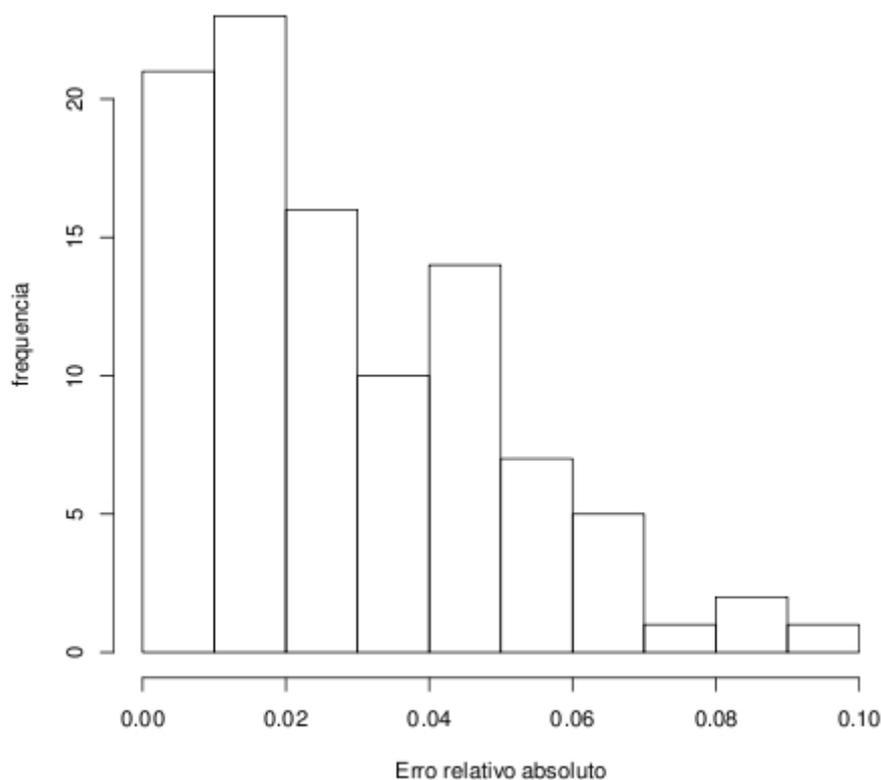
Com objetivo de avaliar a robustez da metodologia proposta, as estimativas resultantes destas replicações serão confrontadas com o valor total simulado de energia não faturada. A metodologia foi aplicada na base sintética criada a partir de parâmetros extraídos de uma base real de consumo de clientes. Todo o processo de amostragem e expansão foi realizado exatamente como seria em uma situação real. ET's e clientes foram selecionados e inspecionados artificialmente.

A Figura 3 mostra um histograma das estimativas de energia não faturada nas cem réplicas, exibindo como um traço vertical o valor exato de energia não faturado.



**FIGURA 3 - Frequências das Estimativas de Perda**

Percebe-se que as perdas estimadas obtidas concentram-se de fato em torno da perda real seguindo uma distribuição normal, com uma queda perceptível na frequência de resultados conforme distancia-se deste valor. A Figura 4 detalha o erro relativo absoluto, em percentual do valor alvo, das estimativas, acentuando a visualização da acurácia obtida pelos experimentos.



**FIGURA 4 - Frequências do Erros Relativos Absolutos**

### 3. Conclusões

Baseando-nos em condições reais obtidas por estudos pilotos, modelamos e simulamos amostragens para a estimação da matriz de perdas não-técnicas de forma consistente. A metodologia proposta resulta em estimadores não viciados, isto é, as estimativas formam uma distribuição simétrica em torno do verdadeiro valor total de energia não faturada. Esse fato é importante pois indica que a metodologia de geração de amostras está apropriada ao modelo estatístico criado. Se o universo de consumo real seguir as suposições estabelecidas na criação dos modelos então podemos afirmar que o resultado da expansão tem uma alta probabilidade de ser significativa.

Os resultados de nossas simulações aproximaram-se de um resultado conhecido com uma faixa de confiabilidade muito boa. Percebe-se que os erros percentuais estão concentrados em até 5%, ou seja, há uma alta probabilidade de que a diferença entre a perda real e a encontrada não ultrapasse 5%. Essa é uma característica muito desejada ao se realizar um experimento estatístico.

Confiantes na metodologia e nos resultados obtidos, está também em fase de implementação e teste um sistema que, baseando-se nas perdas estimadas, automatiza a otimização da alocação das ações a serem implantadas durante certo período de tempo. Estas ações são caracterizadas pelo seu custo e probabilidade de sucesso, e a meta a ser atingida definida pelo usuário é usada como parâmetro para um processo heurístico na busca por uma solução adequada.

#### 4. Referências bibliográficas

- [1] IBGE, I. B. de Geografia e E. *Tendências Demográficas: Uma análise dos Resultados da Sinopse Preliminar do Censo Demográfico 2000*. Departamento de População e Indicadores Sociais, 2001.
- [2] MME, M. das Minas e E. *Balanço Energético Nacional*. Brasília, 2002.
- [3] BOLFARINE, W. B. H. *Elementos de Amostragem*. [S.1.: s.n.], 2002. 290 p.
- [4] MINOUX, S. V. M. *Mathematical Programming: Theory and Algorithms*, [S.1.: s.n.], 1986, 518 p.
- [5] DANTAS, P. R. P. *Avaliação de Perdas de Energia Elétrica Não-Técnicas Metodologia aplicada no Município de Salvador/BA*. Dissertação de Mestrado, Salvador, BA, 2006.
- [6] LONG, J. S. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage Publications. 1997
-