

# UM ALGORITMO ILS PARA MELHORIA DE EFICIÊNCIA DA ESTRATIFICAÇÃO ESTATÍSTICA

**José André de Moura Brito<sup>1</sup>**

IBGE – Instituto Brasileiro de Geografia e Estatística - Diretoria de Pesquisas – DPE/COMÉQ  
Av. Chile , número 500, 10º Andar, Centro – Rio de Janeiro – RJ.  
e-mail: jose.m.brito@ibge.gov.br

**Flávio Marcelo Tavares Montenegro**

IBGE – Instituto Brasileiro de Geografia e Estatística - Diretoria de Pesquisas – DPE/COMÉQ  
Av. Chile , número 500, 10º Andar, Centro – Rio de Janeiro – RJ.  
e-mail: flavio.montenegro@ibge.gov.br

**Luiz Satoru Ochi**

Instituto de Computação - Universidade Federal Fluminense  
Rua Passo da Pátria 156 - Bloco E - 3º andar, São Domingos Niterói - RJ  
e-mail:satoru@ic.uff.br

## RESUMO

Em diversas pesquisas amostrais, a questão de se produzir estimativas mais precisas está diretamente associada à resolução do problema de estratificação estatística, que consiste em determinar  $L$  estratos (agrupamentos) populacionais de forma que os elementos da população mais homogêneos entre si, de acordo com uma variável de estratificação, sejam alocados a um mesmo estrato. Em muitas aplicações práticas, a construção desses estratos é efetuada utilizando um algoritmo de agrupamento não hierárquico, como o  $k$ -médias. Esperando-se produzir estratos mais homogêneos e, em consequência, estimativas mais precisas, propõe-se no presente trabalho um novo algoritmo de estratificação que usa os conceitos básicos da metaheurística ILS e, diferentemente das abordagens tradicionais do problema, considera como função objetivo a própria expressão de variância correspondente ao plano amostral adotado. Resultados computacionais, obtidos a partir de simulações com dados reais do censo demográfico brasileiro, são apresentados e discutidos.

**PALAVRAS CHAVE.** Amostragem. Análise de Agrupamentos. Metaheurísticas. MH.

## ABSTRACT

In a number of sampling surveys, the production of more precise statistics is an issue that relates directly to solving the statistical stratification problem, which consists in determining a number  $L$  of population strata (clusters) in such a way that elements which are more homogeneous among themselves, according to a stratification variable, be allocated to the same stratum. In many practical applications, the construction of such strata is performed by non-hierarchical algorithms, like  $k$ -means. In order to produce even more homogeneous strata and, consequently, more precise estimates, this paper proposes a new stratification algorithm that uses the basic concepts of the ILS metaheuristics and, unlike the traditional approaches to this problem, consider as objective function the expression of variance itself adopted by the sampling design. Computational results obtained through simulations using real data from Brazilian demographic census are presented and discussed.

**KEYWORDS.** Sampling. Clustering Analysis. Metaheuristics. MH.

---

<sup>1</sup> “O IBGE está isento de qualquer responsabilidade pelas opiniões, informações, dados e conceitos emitidos neste artigo, que são de exclusiva responsabilidade dos autores”.

## 1. Introdução

A estratificação estatística desempenha um papel importante em muitas das pesquisas por amostragem, realizadas em diversas áreas e, especialmente, nos órgãos estatísticos, por permitir que, mesmo com uma amostra às vezes bastante limitada, se possam obter estatísticas com um bom grau de precisão, através da exploração adequada das homogeneidades freqüentemente observadas entre elementos da população de estudo. Assim, espera-se que a reunião dos elementos em grupos, ou *estratos*, mais homogêneos, possa resultar na produção de estimativas com um maior nível de precisão, ou seja, com um menor erro padrão associado.

A aplicação de uma estratificação estatística pode ser traduzida em se resolver o seguinte problema: dividir uma população com  $N$  elementos (unidade elementares) em  $L$  subpopulações, ou estratos, de forma que os elementos da população mais homogêneos entre si sejam alocados a um mesmo estrato.

A construção dos estratos é efetuada levando em conta uma variável de estratificação  $Y$ , cujos valores são disponíveis para todos os elementos da população. Em aplicações práticas, é muito comum o uso de um algoritmo de agrupamento não hierárquico, como o das  $k$ -médias (Mingoti, 2007), independentemente do plano amostral adotado (Bolfarine e Bussab, 2005).

Com o objetivo de produzir estratos ainda mais homogêneos, quando comparados aos estratos produzidos pelo algoritmo das  $k$ -médias, propõe-se no presente trabalho um novo algoritmo de estratificação que utiliza os conceitos básicos da metaheurística *ILS* (*Iterated Local Search*; ver Glover e Kochenberger, 2002), agregando, como função objetivo, a expressão de variância correspondente ao plano amostral adotado.

Este trabalho está dividido da seguinte forma: Na seção seguinte, são apresentados de forma resumida os principais conceitos de amostragem. Na seção 3, temos uma descrição detalhada do problema de estratificação abordado neste estudo. Na seção 4, apresenta-se o novo algoritmo de estratificação. Como conclusão do trabalho, tem-se um conjunto de resultados e análises obtidos a partir de simulações efetuadas com os dados do universo do *Censo Demográfico de 2000*.

## 2. Conceitos Básicos sobre Amostragem Probabilística

Atualmente, os levantamentos por amostragem vêm sendo amplamente utilizados em pesquisas para os mais variados fins, incluindo a tomada de decisão governamental, saúde pública, economia, pesquisas eleitorais, etc.

Muitas destas pesquisas seriam inviáveis caso se buscasse entrevistar todos os elementos da população de interesse (efetuar um censo), tendo em vista a limitação de tempo e os altos custos operacionais envolvidos para sua realização.

O levantamento por amostragem permite a obtenção de informações a respeito de valores populacionais desconhecidos, mediante a observação de apenas parte da população, denominada amostra. Ou seja, a partir da amostra, é possível produzir, por exemplo, estimativas de total e/ou da média para um conjunto de  $p$  variáveis de interesse. Tais estimativas constituem valores aproximados do que seriam os valores reais, caso todos os elementos da população (universo) fossem investigados.

Introduzindo a nomenclatura de amostragem, temos que os *elementos* de uma população são as unidades elementares de interesse (Bolfarine e Bussab, 2005). Matematicamente, a população consiste de todas as unidades elementares de interesse, sendo definida pelo conjunto  $U = \{1, 2, \dots, N\}$ , onde  $N$  é o tamanho da população. O termo *elemento populacional* é usado para denotar qualquer elemento  $i \in U$ , também conhecido por unidade elementar. Considerando uma pesquisa que envolva  $p$  variáveis de interesse, a  $j$ -ésima variável de interesse da pesquisa é, então, um vetor de informações associado aos elementos da população e definido por  $Y^j = (Y_1^j, Y_2^j, \dots, Y_i^j, \dots, Y_N^j)$ . A amostra é definida pelo conjunto  $S = \{1, 2, \dots, n\}$ , tal que  $S \subset U$ .

Exemplificando, consideremos uma pesquisa que vise estimar o salário médio de 5000 funcionários de uma companhia. Ao invés de se investigar toda a população, seleciona-se uma amostra de 500 funcionários, e anotam-se os seus salários. Neste caso, a variável  $Y$  (característica de interesse) a ser observada é o salário.

Ao estudar a distribuição dos salários dos funcionários na amostra e calcular o salário médio, espera-se que esta reflita a distribuição dos salários da população e o salário médio da população de 5000 funcionários, desde que a amostra tenha sido escolhida levando em conta os conceitos de amostragem probabilística (vide Bolfarine e Bussab, 2005).

A partir deste ponto, podemos definir os conceitos de estimador e de estimativa. O estimador corresponde a qualquer expressão (totais, médias, razões, etc) aplicada em cada unidade de uma amostra, considerando uma particular variável, e a estimativa é o valor assumido pelo estimador.

Para este exemplo, temos que  $\hat{Y} = \frac{1}{n} \sum_{i \in S} Y_i$  é um estimador da média populacional e o valor

observado para  $\hat{Y}$  corresponderá a uma estimativa da média populacional (utiliza-se símbolo  $\hat{\phantom{Y}}$  para representar uma estimativa).

Ao se definir um plano amostral, várias questões devem ser consideradas, dentre as quais destacamos: as restrições de orçamento, a melhoria dos estimadores, a precisão das estimativas (menor erro-padrão possível associado), a seleção dos instrumentos de coleta e as características que serão investigadas na população. A avaliação de tais questões é de vital importância para a determinação do esquema de amostragem que será adotado. A seguir, apresentamos os esquemas de amostragem que foram estudados no presente trabalho (Cochran, 1977):

- **Amostragem Estratificada.** Consiste na divisão de uma população em grupos, chamados de estratos, segundo alguma(s) característica(s) conhecida(s) na população sob estudo (Bolfarine e Bussab, 2005). Uma das principais razões para se realizar uma estratificação é a melhoria da precisão das estimativas.

Na amostragem estratificada, uma população  $U$  com  $N$  unidades é dividida em  $L$  subpopulações (estratos) com  $N_1, N_2, \dots, N_h, \dots, N_L$  unidades. Essas subpopulações não se superpõem e, juntas, abrangem a totalidade da população.

Depois de definidos os estratos, a partir do conhecimento de uma ou mais características da população, seleciona-se uma amostra em cada um deles, sendo as seleções feitas de maneira independente em cada um dos estratos, considerando algum esquema de amostragem probabilística (amostragem aleatória simples, amostragem com probabilidade proporcional ao tamanho, etc).

- **Amostragem de Conglomerados.** Neste esquema, os elementos da população são reunidos em grupos, os quais, por sua vez, são sorteados para compor a amostra. Um exemplo simples a ser considerado é a seleção de uma amostra composta por 500 estudantes. Ao invés de sortear os estudantes (elementos) diretamente de uma listagem de todos os alunos, pode-se sortear algumas escolas e considerar todos os alunos dessas escolas para compor a amostra.

A amostragem por conglomerados introduz uma economia na construção do cadastro, dispensando a necessidade de listar o total de elementos que compõe a população. Além disso, o custo de locomoção e acesso aos elementos da amostra para obtenção da informação desejada pode ser reduzido de forma significativa.

- **Amostragem com Probabilidade Proporcional ao Tamanho.** É um esquema de amostragem no qual a probabilidade de seleção não é a mesma para todas as unidades amostrais, e sim, proporcional ao tamanho de cada unidade. Para fins de exemplificação, suponha que tenhamos um cadastro (universo) com um conjunto de fazendas e suas respectivas áreas em hectares. As fazendas com maiores áreas devem ter maior probabilidade de serem selecionadas para amostra, o que implica em que a variável área seja utilizada como a variável de tamanho para o processo de seleção da amostra.

### 3. Descrição do Problema

O plano amostral estudado neste trabalho considera a aplicação da amostragem de conglomerados com estratificação e seleção com probabilidade proporcional ao tamanho. Inicialmente, são definidos  $L$  estratos populacionais, constituídos por um conjunto de  $M_h$  conglomerados ( $h = 1, \dots, L$ ) representados pelos setores censitários do universo (unidades de coleta do *Censo Demográfico Brasileiro*). Os setores censitários são definidos pelo IBGE obedecendo a critérios de operacionalização da coleta de dados, de tal maneira que abranjam uma área que possa ser percorrida por um único recenseador em um mês e que possua em torno de 250 a 350 domicílios (em áreas urbanas). Tais setores serão as unidades elementares da população.

A construção dos estratos populacionais, isto é, a definição dos grupos de conglomerados, é feita levando em conta os valores de uma variável  $Y$  (variável de estratificação), conhecida para todos os conglomerados do universo, sendo desejável que os conglomerados mais homogêneos entre si, segundo os valores de  $Y$ , sejam alocados a um mesmo estrato.

Em seguida, com a finalidade de produzir estimativas com um bom nível de precisão para a variável  $Y$  e para outras variáveis  $X$  que sejam correlacionadas com  $Y$ , seleciona-se, em cada um dos estratos, uma amostra de  $m_h$  ( $h = 1, \dots, L$ ) conglomerados, com probabilidade de seleção proporcional a uma variável de tamanho cuja informação esteja disponível para cada um dos conglomerados.

A figura 1 ilustra a aplicação deste plano amostral considerando uma população (retângulo) que foi dividida, considerando-se a variável  $Y$ , em 3 estratos homogêneos, compostos, respectivamente, por seis, três e quatro conglomerados (retângulos menores). Ou seja, foi realizada uma estratificação estatística. Os retângulos em cinza representam os conglomerados que foram, então, selecionados para compor a amostra.

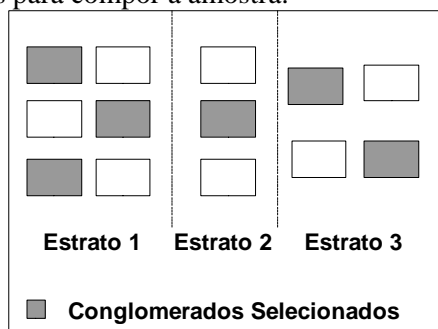


Figura 1 – Plano Amostral Combinando Estratificação e Conglomeração

É fato conhecido que a adoção deste tipo construção pode cooperar em muito para se obter estimativas precisas, ressaltando-se que o nível de precisão será tanto melhor quanto mais homogêneos forem os estratos produzidos e, caso sejam desejadas estimativas para uma variável  $X$ , quanto maior for a correlação entre as variáveis  $Y$ , de estratificação, e  $X$ .

Normalmente, os estratos são construídos mediante a aplicação de um algoritmo de agrupamento não hierárquico, tal como o algoritmo das  $k$ -médias (Mingoti, 2007 e Hartigan e Wong, 1979), que é provavelmente um dos mais conhecidos e mais utilizados em problemas reais. Na aplicação deste algoritmo, busca-se iterativamente alocar cada elemento àquele grupo cujo centróide (vetor de média(s) da(s) variável(eis)  $Y$ ) é o mais próximo do vetor de valores  $Y$  observados para o respectivo elemento. O processo completo é composto basicamente por quatro passos: (1) Escolha dos centróides, para inicializar o processo de agrupamento. (2) Cada elemento do conjunto de dados é comparado com cada centróide inicial, através de uma medida de distância, e o elemento é alocado ao grupo cujo centróide está a uma distância menor. (3) Recalculam-se os valores dos centróides para cada novo grupo formado e repete-se o passo (2), considerando os centróides destes novos grupos. (4) Os passos (2) e (3) são repetidos até que a diferença entre os centróides em duas iterações consecutivas seja não significativa.

Em geral, este algoritmo tende a convergir rapidamente para um mínimo local que pode estar distante do ótimo global do problema, que normalmente é muito difícil de ser encontrado, devido à alta complexidade computacional decorrente de seu aspecto combinatório. Caso seja aplicado um processo de busca exaustiva para garantir a obtenção da solução global, será necessário enumerar todas as soluções, ou seja, todas as possibilidades de combinação dos  $n$  objetos em  $k$  grupos. O número de possibilidades, neste caso, está associado ao número de Stirling de segundo tipo (ver Johnson e Wichern, 2002), dado por  $\frac{1}{k!} \sum_{j=0}^k (-1)^{k-j} \binom{k}{j} j^n$ .

Entretanto, é comumente possível encontrar soluções de melhor qualidade do que as obtidas pelo algoritmo das  $k$ -médias, utilizando-se de métodos de computação mais intensiva (embora não exaustiva), disponíveis na literatura, às custas de um maior tempo de processamento. Uma dessas possibilidades é de aplicação de uma metaheurística.

As metaheurísticas (Resende, 2004 e Glover e Kochenberger, 2002) vêm sendo desenvolvidas pelo menos desde os anos 80, sendo tal termo utilizado por Glover (1986). Enquanto as heurísticas tradicionais param, em geral, no primeiro ponto ótimo local encontrado, as heurísticas que utilizam os paradigmas das metaheurísticas possuem mecanismos que possibilitam escapar desses pontos, usualmente ótimos locais “pobres”, cujos valores associados estão normalmente muito distantes do valor da solução ótima.

Em face destas considerações, e na expectativa de se produzir extratos mais homogêneos do que os obtidos pelo método das  $k$ -médias, uma possível alternativa é construir os estratos através de um algoritmo de minimização que tenha, como função objetivo, a expressão de variância do plano amostral de conglomerados com estratificação.

Apresenta-se, a seguir, a notação e as fórmulas associadas ao plano amostral que foi objeto de estudo deste trabalho e que motivou a implementação do novo algoritmo de estratificação.

### 3.1 Notação Utilizada na Amostragem de Conglomerados com Estratificação

Inicialmente, suponhamos que os  $M$  conglomerados são agrupados em  $L$  estratos  $E_1, E_2, \dots, E_L$ , tendo-se associado a cada conglomerado o total da característica  $Y$ . Denotando por  $E_h$  um estrato genérico ( $h = 1, 2, \dots, L$ ), podemos definir:

- $Y_{hi}$  - Valor de  $Y$  no  $i$ -ésimo conglomerado do  $h$ -ésimo estrato do universo.
- $M_h$  - O número de conglomerados (setores) no estrato  $h$  do universo.
- $m_h$  - O número de conglomerados (setores) da amostra no estrato  $h$ .
- $N_h$  - Número de domicílios da população no  $h$ -ésimo estrato do universo.
- $N_{hi}$  - Número de domicílios da população no  $i$ -ésimo conglomerado do  $h$ -ésimo estrato.

$$Y = \sum_{h=1}^L Y_h \quad \text{Total da característica } Y \text{ no universo} \quad (1)$$

$$Y_h = \sum_{i=1}^{M_h} Y_{hi} \quad \text{Total da característica } Y \text{ no } h\text{-ésimo estrato do universo} \quad (2)$$

$$p_{hi} = \frac{N_{hi}}{N_h} \quad \begin{array}{l} \text{Tamanho relativo do } i\text{-ésimo conglomerado do} \\ \text{ } h\text{-ésimo estrato, que define a probabilidade seleção deste} \\ \text{conglomerado no sorteio para compor a amostra} \end{array} \quad (3)$$

Agora, selecionando-se em cada um dos  $L$  estratos uma amostra de conglomerados com probabilidade de seleção proporcional a uma variável tamanho, podemos estimar os valores  $Y_h$  a partir da aplicação das seguintes fórmulas:

$$Y'_{hi} \quad \begin{array}{l} \text{Valor de } Y \text{ no } i\text{-ésimo conglomerado selecionado no} \\ \text{ } h\text{-ésimo estrato para compor a amostra} \end{array} \quad (4)$$

$$p'_{hi} \quad \begin{array}{l} \text{Tamanho relativo do } i\text{-ésimo conglomerado do} \\ \text{ } h\text{-ésimo estrato, que define a probabilidade de seleção deste} \\ \text{conglomerado na amostra} \end{array} \quad (5)$$

$$\hat{Y}_h = \frac{1}{m_h} \sum_{i=1}^{m_h} \frac{Y_{hi}'}{p_{hi}} \quad \text{Estimativa do total da característica } Y \text{ no } h\text{-ésimo estrato} \quad (6)$$

$$\hat{Y} = \sum_{h=1}^L \hat{Y}_h \quad \text{Estimativa do total da característica } Y \quad (7)$$

Os estratos populacionais são construídos levando-se em conta a expressão de variância (equação 8) (combinação das variâncias dos estratos) ou o coeficiente de variação ( $cv$ ), que corresponde a uma medida de precisão relativa (equação 9). Observe-se que quanto menor o valor de (8) ou (9), mais homogêneos serão os estratos populacionais.

$$V(\hat{Y}) = \sum_{h=1}^L \frac{1}{m_h} \left( \sum_{i=1}^{M_h} \frac{Y_{ih}^2}{p_{ih}} - Y_h^2 \right) \quad (8)$$

$$CV(\hat{Y}) = 100 \cdot \sqrt{V(\hat{Y})} / Y \quad (9)$$

No caso de uma particular amostra, selecionada a partir dos estratos populacionais, segundo o plano amostral previamente descrito, pode-se avaliar a precisão da estimativa de  $Y$  ou de uma variável  $X$  a partir do cálculo da variância amostral ou de seu coeficiente de variação associado (equações 10 e 11).

$$v(\hat{Y}) = \sum_{h=1}^L \frac{1}{m_h \cdot (m_h - 1)} \left( \sum_{i=1}^{M_h} \frac{Y_{ih}'^2}{p_{ih}} - \hat{Y}_h^2 \right) \quad (10)$$

$$cv(\hat{Y}) = 100 \cdot \sqrt{v(\hat{Y})} / \hat{Y} \quad (11)$$

## 4. Algoritmo de Estratificação

### 4.1 Metaheurística ILS

A metaheurística ILS (*Iterated Local Search*), proposta por Lourenço, Martin e Stützle (Glover e Kochenberger, 2002), consiste, essencialmente, na aplicação iterativa de um procedimento de busca local em uma solução inicial  $s^0$ , que é previamente obtida a partir da utilização de um procedimento aleatório de construção ou considerando uma heurística de construção. O procedimento de busca local tem por finalidade melhorar a solução inicial e também aquelas produzidas após perturbações de soluções ótimas locais.

O êxito desta metaheurística está diretamente associado à escolha do **procedimento de busca local**, ao **procedimento de perturbação** aplicado sobre a solução corrente e ao critério de **aceitação das soluções**. Observamos que a implementação destes procedimentos, bem como do critério de aceitação, está intrinsecamente associada ao problema que será resolvido.

O pseudo-código abaixo mostra os passos básicos desta metaheurística: No passo (1) constrói-se uma solução inicial  $s^0$ , aplicando sobre solução uma busca local (passo 2), que produz uma solução  $s^*$ . Com a finalidade de obter soluções de melhor qualidade, aplica-se consecutivamente, a partir de  $s^*$ , um procedimento de perturbação e uma nova busca local, o que resulta em um nova solução  $s''$  a ser comparada com a solução  $s^*$ . Se a nova solução  $s''$  satisfaz um critério de aceitação baseado em  $s^*$ , então atribui-se  $s''$  a  $s^*$ , caso contrário, a solução  $s^*$  é mantida. Os passos (3), (4) e (5) são então aplicados durante  $m$  iterações, obtendo-se no final a melhor solução associada a  $s^*$ .

**Procedimento Iterated Local Search**

1.  $s^0 = \text{Gerar\_Solução\_Inicial};$
2.  $s^* = \text{Busca\_Local}(s^0);$

**Repita**

3.  $s' = \text{Perturbação}(s^*);$
4.  $s'' = \text{Busca\_Local}(s');$
5.  $s^* = \text{Critério\_Aceitação}(s^*, s'');$

**Até (Sejam efetuadas  $m$  iterações);**

**Fim-Procedimento**

Figura 2 - Pseudo-Código da Metaheurística ILS

#### 4.2 Algoritmo Proposto

O algoritmo ILS proposto neste trabalho consiste em um procedimento de geração, dois procedimentos de perturbação, dois procedimentos de busca local e um procedimento de aceitação, quais sejam:

**Geração da Solução Inicial:** São geradas 10 soluções, definindo-se como a solução inicial ( $s^o$ ) aquela com o menor valor da variância segundo a equação (8). Uma solução é representada por um vetor de  $N$  posições (número de setores na população) preenchidas com um valor aleatório entre 1 e  $L$  (número de estratos), indicando o estrato ao qual cada setor será alocado.

**Perturbação Tipo 1:** Consiste na troca de  $m$  setores entre dois estratos, com  $m$  assumindo um valor aleatório entre dois e cinco. Inicialmente, são sorteados dois valores entre 1 e  $L$ , que corresponderão aos índices dos estratos  $E_i$  e  $E_j$  entre os quais será efetuada a troca, observando que  $m = \text{mínimo}(m, |E_i|, |E_j|)$ . Em seguida, são sorteados  $m$  setores em cada um destes estratos, efetuando-se a sua troca.

**Perturbação Tipo 2:** Sorteia-se, respectivamente, um estrato  $E_i$  (observando que  $|E_i| > L$ ) dentre os  $L$  estratos e  $(L-1)$  setores deste estrato a serem alocados aos  $(L-1)$  estratos restantes.

**Busca Local Tipo 1:** Inicialmente, são efetuadas  $r$  trocas de setores entre dois estratos,  $E_i$  e  $E_j$ , selecionados aleatoriamente. Ou seja, em cada troca aloca-se um setor  $o_i \in E_i$  ao estrato  $E_j$  e um setor  $o_j \in E_j$  ao estrato  $E_i$ , sendo  $o_i$  e  $o_j$  também selecionados aleatoriamente.

Em seguida, ainda considerando os estratos  $E_i$  e  $E_j$ , são efetuadas  $r$  re-aloções de  $m$  setores do estrato  $E_i$  para o estrato  $E_j$  ou vice-versa. Para cada re-aloção, atribui-se a  $m$  um valor aleatório entre 1 e 4. Sorteia-se, também, um número real  $\rho \in (0,1)$ . Se  $\rho > 0.5$  o estrato  $E_i$  será o doador dos  $m$  setores e o estrato  $E_j$  será o receptor, caso contrário, o  $E_j$  será o doador e  $E_i$  será o receptor.

Após as  $r$  trocas e as  $r$  re-aloções, havendo uma redução no valor da variância em relação à melhor variância atual, atualiza-se a solução. Caso contrário, retorna-se à melhor solução obtida até o momento e selecionam-se dois novos estratos, aplicando-se novamente os dois tipos de movimentos.

**Busca Local Tipo 2:** Utiliza um procedimento de busca em vizinhanças, considerando a combinação de todos os estratos, tomados dois a dois. Para cada dois estratos  $E_i$  e  $E_j$  ( $i < j, i, j \in [1, 2, \dots, L]$ ) seleciona-se, aleatoriamente,  $q$  setores (com  $q$  variando entre 1 e 4), efetuando-se  $r$  trocas de  $q$  setores entre  $E_i$  e  $E_j$ . Após as  $r$  trocas, havendo redução da variância, continua-se na mesma vizinhança, selecionando-se  $q$  novos setores a serem trocados entre  $E_i$  e  $E_j$ . Em caso contrário, incrementa-se a vizinhança de um, buscando-se a troca de  $(q+1)$  setores entre os estratos. Quando a vizinhança máxima é atingida, novos estratos são definidos, aplicando-se, novamente, as  $r$  trocas de  $q$  setores entre  $E_i$  e  $E_j$ .

**Aceitação:** Por ordem de aplicação, consiste dos seguintes testes:

(1) Se  $f(s'') < f(s^*) \Rightarrow$  Substituir  $s^*$  (melhor solução até o momento) por  $s''$  (solução advinda da busca local).

(2)  $|f(s^*) - f(s'')| / f(s^*) \leq \epsilon \Rightarrow$  Substituir  $s^*$  (melhor solução até o momento) por  $s''$  (solução advinda da busca local).

Caso nenhum dos testes seja satisfeito, a solução  $s^*$  é mantida. O primeiro procedimento de perturbação foi aplicado nas iterações de número ímpar do algoritmo e o segundo nas iterações de número par. A busca local do tipo 1 foi aplicada no passo 2 do algoritmo (figura 2), logo após

o procedimento de construção e a busca local do tipo 2 foi aplicada em todas iterações do algoritmo (passo 4 do algoritmo).

## 5. Resultados Computacionais

A avaliação do algoritmo de estratificação (ILS), implementado em linguagem R ([www.r-project.org/](http://www.r-project.org/)), foi efetuada mediante a realização de um conjunto de experimentos computacionais com os dados do *Censo Demográfico* de 2000, produzido pelo IBGE. Todos os experimentos foram realizados em um computador Pentium com 2Gb de memória RAM e dotado de um processador de 3.0 Ghz (*Core 2 Duo*).

A base de dados utilizada no presente experimento constitui-se dos setores censitários do universo do Censo Demográfico de 2000 (dados de uso público), associados a algumas das unidades da federação. Cada setor censitário correspondeu a um conglomerado, sendo o total do rendimento dos responsáveis pelos domicílios e o total de anos de estudo dos responsáveis pelos domicílios, respectivamente, a variável de estratificação e a variável de estudo.

### 5.1 Aplicação do Algoritmo ILS e Simulações

A primeira fase do experimento consistiu da aplicação dos algoritmos ILS e *k*-médias, de forma a construir os estratos populacionais formados pelos conglomerados, considerando um conjunto de 15 unidades da federação.

A função objetivo utilizada correspondeu à variância expressa na equação (8) e o número de estratos foi fixado entre três e oito, o que implicou, por sua vez, em doze possíveis estratificações da população (seis do ILS e seis do algoritmo das *k*-médias) para cada uma das 15 unidades da federação.

Para a execução do algoritmo ILS, o número de iterações definido *a priori* foi de 50, o número de trocas e atribuições efetuadas pelos procedimentos de busca local foi igual a 800 e o valor do fator de tolerância  $\epsilon$  foi de 0,1.

A partir dos estratos populacionais construídos pelos dois algoritmos, foram selecionadas 400 amostras de conglomerados (setores) com a probabilidade de seleção de cada setor proporcional ao número de domicílios (variável de tamanho) no setor. Para compor cada uma das amostras, foram selecionados 20% dos setores em cada um dos estratos populacionais

Em seguida, calculou-se, a partir de cada uma das amostras, a estimativa de total (equação 7) da variável *Y* e da variável *X* e o coeficiente de variação destas estimativas (equação 11). Finalmente, considerando as 400 estimativas e os seus respectivos coeficientes de variação, calculou-se a média dos desvios relativos (equação 12) das estimativas de *Y* e *X*, em relação aos valores conhecidos do censo para as variáveis *Y* e *X*, e a média dos coeficientes de variação (equação 13).

$$\rho = \left( \sum_{i=1}^{400} 100x \left| \hat{Y}_i - Y \right| / \hat{Y}_i \right) / 400 \quad \hat{Y}_i = \text{Estimativa de } Y \text{ obtida na } i\text{-ésima amostra} \quad (12)$$

$$\mu = \sum_{i=1}^{400} cv(\hat{Y}_i) / 400 \quad cv(\hat{Y}_i) = \text{Coeficiente de variação associado à } i\text{-ésima estimativa} \quad (13)$$

Na tabela 1 e no gráfico 1, temos uma sumarização dos resultados relativos às estratificações produzidas pelos dois algoritmos.

A primeira coluna desta tabela contém, para cada uma das 15 unidades da federação, o nome da unidade, o número de setores do universo nesta unidade e o coeficiente de correlação entre *X* e *Y* (Bussab e Morettin, 2003). A coluna dois indica o número de estratos considerado. As colunas três e quatro trazem, respectivamente, os coeficientes de variação associados ao total populacional (equação 9) considerando os estratos contruídos pelo algoritmo das *k*-médias (*CV*<sub>1</sub>) e pelo algoritmo ILS (*CV*<sub>2</sub>). Em seguida, temos, nas colunas cinco e seis, a média dos desvios relativos (equação 12)  $\rho_1$  e  $\rho_2$  considerando as 400 estimativas produzidas a partir das amostras selecionadas dos estratos construídos pelos dois algoritmos. As colunas sete e oito trazem o coeficientes de variação médios  $\mu_1$  e  $\mu_2$  para as 400 amostras selecionadas. De maneira análoga, nas colunas de nove a dezesseis, temos os respectivos resultados para a variável *X*.



Para o gráfico 1, cada *box plot* em cinza claro contém as informações dos *cv's* da variável de estratificação renda, obtidos a partir da aplicação do algoritmo ILS em cada uma das quinze UFs, com número de estratos variando entre 3 e 8. Igualmente, cada *box plot* em cinza escuro contém os *CV's* do algoritmo das *k*-médias. Tais *CV's* foram calculados mediante aplicação da equação 9.

Inicialmente, analisando os valores dos coeficientes de variação apresentados na colunas 3 e 4 da tabela 1, é possível observar que o algoritmo ILS produziu estratos mais homogêneos que aqueles produzidos pelo algoritmo das *k*-médias. Para algumas UFs, a diferença entre os *CV's* do ILS e os *CV's* do *k*-médias chega a ser de quase 100%.

Pode-se observar, também, que as estimativas de total da variável de estratificação renda, produzidas a partir das 400 amostras selecionadas nos estratos definidos pelo ILS são mais acuradas (vide média dos desvios relativos – colunas 5 e 6) do que as estimativas produzidas a partir das amostras sorteadas nos estratos definidos pelo algoritmo *k*-médias. Ou seja, tais estimativas estão mais próximas do valor real de *Y*.

Por sua vez, o coeficiente de variação médio (colunas 7 e 8) calculado a partir da estimativa dos *cv's* de cada amostra apresenta valores bem próximos ao *CV* geral (colunas 1 e 2), o que indica a robustez dos estratos produzidos, ou seja, que a precisão foi mantida independentemente da amostra selecionada.

Destaca-se, também, que foram obtidos maiores ganhos de precisão para variável *anos de estudo* mediante a utilização dos estratos do algoritmo ILS. Tal observação está fundamentada tanto em relação aos valores dos *CV's* associados à função objetivo (colunas 9 e 10) quanto ao valor da média dos desvios relativos (colunas 11 e 12). Tal fato é importante, tendo em vista que nas pesquisas que utilizam planos amostrais são feitas estimativas para um conjunto de variáveis associadas à variável de estratificação. Provavelmente, tais ganhos de precisão foram favorecidos pela razoável correlação entre *X* e *Y*, pois, na maioria das UFs, o coeficiente de correlação linear (Bussab e Morettin, 2003) é da ordem de 0.8.

Finalmente, a análise do gráfico 1 volta a reforçar a observação de que o algoritmo ILS tende a produzir estratos mais homogêneos do que os produzidos pelo algoritmo das *k*-médias. Além disso, os maiores ganhos de precisão (maior redução da função objetivo) foram obtidos com número de estratos entre 3 e 6.

Ressalte-se que esta diferença de performance se deve principalmente ao uso de algoritmos que resultam em processos de computação mais intensiva, permitindo-se explorar uma quantidade bem maior de soluções. O custo desta melhoria pode ser expresso pela diferença nos tempos computacionais exigidos pelos algoritmos, os quais foram da ordem de minutos para o algoritmo das *k*-médias e da ordem de horas (com um máximo de cerca de dez horas) para os algoritmos propostos neste trabalho. Uma vez, entretanto, que a qualidade das soluções obtidas foi significativamente melhor, o que pode ocasionar uma redução nos custos, devido à necessidade de uma amostra menor, e/ou um aumento da precisão das estimativas, o uso de tais métodos intensivos de computação pode vir a ser bastante vantajoso.

Em trabalhos futuros, pretende-se avaliar a eficiência do algoritmo ILS considerando novas variáveis *X* e outros planos de amostragem, o que implica, por sua vez, em minimizar outras expressões de variância. Também serão implementados procedimentos de perturbação e busca local mais sofisticados, objetivando ganhos de precisão (redução da função objetivo) ainda maiores que os obtidos neste estudo.

Além disso, objetivando uma redução no tempo de processamento, em comparação com o algoritmo implementado em R (uma linguagem interpretada), espera-se implementar em futuros trabalhos novas versões dos algoritmos também em linguagem compilada, como C++ ou Java.

**Tabela 1 – Resultados da Estratificação**

UF	L	CVy <sub>1</sub>	CVy <sub>2</sub>	ρy <sub>1</sub>	ρy <sub>2</sub>	μcvy <sub>1</sub>	μcvy <sub>2</sub>	CVx <sub>1</sub>	CVx <sub>2</sub>	ρx <sub>1</sub>	ρx <sub>2</sub>	μcvx <sub>1</sub>	μcvx <sub>2</sub>	
<b>Rondônia</b> Setores 1839	3	3,03	2,01	2,31	1,58	2,99	2,01	1,92	1,67	1,59	1,35	1,92	1,66	
	4	2,82	1,67	2,40	1,32	2,77	1,65	1,92	1,57	1,56	1,27	1,93	1,57	
	5	2,73	1,56	2,41	1,29	2,67	1,55	1,87	1,56	1,49	1,24	1,86	1,57	
	6	2,70	1,41	2,15	1,14	2,66	1,40	1,87	1,53	1,52	1,21	1,88	1,53	
	<b>r=0,84</b>	7	2,54	1,25	2,04	0,99	2,52	1,24	1,87	1,52	1,55	1,19	1,86	1,52
	8	2,56	1,22	2,03	0,94	2,55	1,21	1,88	1,50	1,50	1,10	1,87	1,50	
	3	2,74	2,54	2,22	2,03	2,70	2,53	1,41	1,39	1,12	1,06	1,41	1,39	
	4	2,49	2,43	2,05	2,02	2,46	2,42	1,35	1,26	1,11	0,96	1,35	1,26	
<b>Amazonas</b> Setores 3151	5	2,27	1,98	1,78	1,53	2,24	1,97	1,20	1,13	0,94	0,90	1,19	1,14	
	6	2,22	1,98	1,80	1,68	2,20	1,97	1,19	1,11	0,98	0,88	1,18	1,11	
	<b>r=0,73</b>	7	2,15	1,78	1,60	1,35	2,10	1,76	1,14	1,08	0,91	0,88	1,14	1,08
	8	2,12	1,74	1,67	1,48	2,07	1,72	1,14	1,07	0,91	0,87	1,14	1,07	
	3	1,99	1,68	1,61	1,35	1,97	1,67	1,26	1,21	0,96	0,94	1,26	1,21	
	4	1,91	1,50	1,54	1,19	1,89	1,49	1,22	1,13	0,98	0,93	1,22	1,13	
	<b>Pará</b> Setores 5736	5	1,70	1,36	1,44	1,15	1,69	1,35	1,19	1,09	0,90	0,86	1,19	1,09
		6	1,62	1,27	1,46	1,00	1,63	1,26	1,17	1,07	0,95	0,82	1,17	1,07
<b>r=0,76</b>		7	1,63	1,29	1,33	1,04	1,60	1,29	1,18	1,07	0,98	0,84	1,18	1,07
8		1,58	1,30	1,26	1,06	1,53	1,29	1,17	1,05	0,92	0,83	1,17	1,05	
3		3,64	3,09	2,71	2,55	3,59	3,08	2,26	2,19	1,75	1,73	2,25	2,20	
4		3,01	2,73	2,27	2,07	3,00	2,73	2,00	1,98	1,61	1,55	1,99	1,96	
<b>Tocantins</b> Setores 1361		5	2,97	2,34	2,20	1,88	2,95	2,35	2,03	1,97	1,58	1,52	2,02	1,97
		6	2,73	2,24	2,24	1,80	2,64	2,23	1,97	1,92	1,59	1,57	1,96	1,91
	<b>r=0,84</b>	7	2,68	1,89	2,01	1,47	2,62	1,89	1,91	1,85	1,59	1,41	1,90	1,85
	8	2,64	1,69	2,10	1,34	2,60	1,67	1,88	1,78	1,51	1,45	1,87	1,77	
	3	2,39	2,27	1,91	1,90	2,35	2,28	1,13	1,08	0,90	0,87	1,13	1,07	
	4	2,21	1,92	1,78	1,45	2,20	1,91	1,06	1,01	0,86	0,77	1,05	1,01	
	<b>Espírito Santo</b> Setores 3163	5	1,99	1,86	1,59	1,52	1,99	1,86	1,02	0,99	0,82	0,78	1,02	0,99
		6	1,86	1,30	1,50	1,03	1,82	1,29	0,95	0,87	0,74	0,67	0,95	0,86
<b>r=0,83</b>		7	1,85	1,48	1,55	1,19	1,81	1,47	0,94	0,88	0,76	0,68	0,93	0,88
8		1,72	0,97	1,32	0,72	1,71	0,96	0,92	0,84	0,78	0,67	0,92	0,84	
3		1,28	1,14	1,03	0,91	1,28	1,14	0,56	0,55	0,45	0,42	0,56	0,55	
4		1,18	0,99	0,98	0,76	1,18	0,99	0,54	0,48	0,42	0,40	0,54	0,48	
<b>Paraná</b> Setores 12889		5	1,09	0,91	0,92	0,71	1,08	0,91	0,51	0,45	0,41	0,35	0,51	0,45
		6	1,03	0,87	0,83	0,66	1,03	0,87	0,49	0,44	0,42	0,33	0,49	0,44
	<b>r=0,81</b>	7	1,02	0,83	0,81	0,66	1,02	0,83	0,49	0,42	0,43	0,33	0,49	0,42
	8	1,01	0,77	0,82	0,62	1,00	0,77	0,48	0,40	0,39	0,31	0,48	0,40	
	3	1,53	1,22	1,22	0,99	1,52	1,21	0,69	0,66	0,58	0,51	0,69	0,66	
	4	1,39	1,08	1,15	0,88	1,38	1,07	0,65	0,61	0,50	0,49	0,65	0,61	
	<b>Santa Catarina</b> Setores 6748	5	1,34	0,95	1,04	0,79	1,32	0,95	0,65	0,57	0,51	0,45	0,65	0,57
		6	1,30	0,92	1,04	0,72	1,27	0,90	0,62	0,55	0,49	0,42	0,62	0,55
<b>r=0,83</b>		7	1,26	0,85	1,03	0,71	1,26	0,84	0,62	0,54	0,50	0,44	0,62	0,54
8		1,26	0,77	0,94	0,59	1,24	0,78	0,61	0,53	0,47	0,43	0,61	0,53	

(1) Algoritmo das *k*-médias

(2) Algoritmo ILS

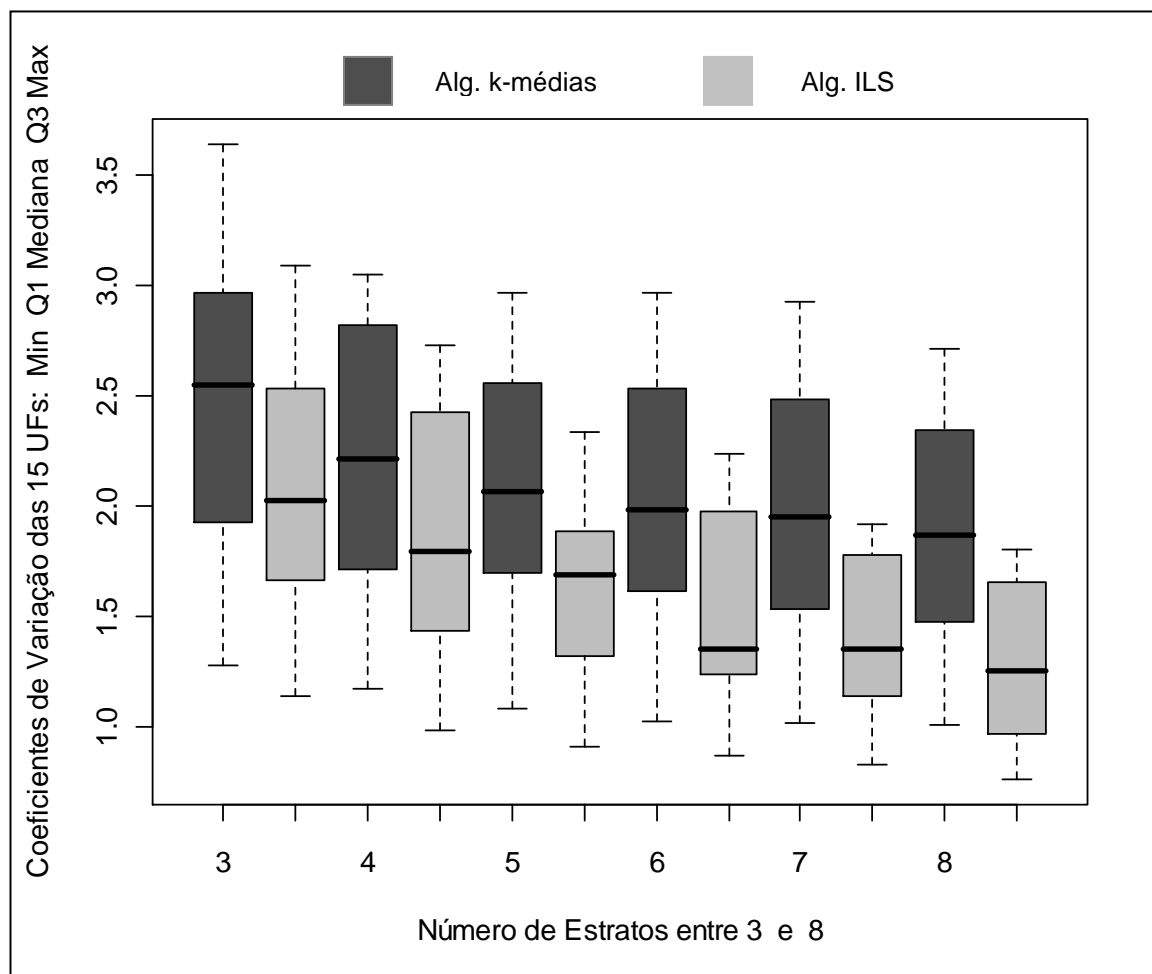
**Tabela 1 – Resultados da Estratificação (continuação)**

UF	L	CV <sub>y<sub>1</sub></sub>	CV <sub>y<sub>2</sub></sub>	$\rho_{y_1}$	$\rho_{y_2}$	$\mu_{cv_{y_1}}$	$\mu_{cv_{y_2}}$	CV <sub>x<sub>1</sub></sub>	CV <sub>x<sub>2</sub></sub>	$\rho_{x_1}$	$\rho_{x_2}$	$\mu_{cv_{x_1}}$	$\mu_{cv_{x_2}}$	
<b>Rio G. Sul</b> Setores 16634	3	1,08	0,86	0,89	0,64	1,08	0,85	0,47	0,43	0,35	0,32	0,47	0,43	
	4	0,98	0,75	0,77	0,57	0,98	0,75	0,45	0,39	0,35	0,31	0,45	0,39	
	5	0,92	0,68	0,75	0,56	0,92	0,68	0,43	0,37	0,35	0,31	0,43	0,37	
	6	0,89	0,62	0,69	0,52	0,89	0,62	0,42	0,35	0,32	0,27	0,42	0,35	
	<b>r=0,83</b>	7	0,87	0,58	0,68	0,46	0,87	0,58	0,42	0,34	0,32	0,26	0,42	0,34
	8	0,85	0,53	0,66	0,42	0,85	0,53	0,41	0,33	0,33	0,28	0,41	0,33	
	3	3,16	1,84	2,40	1,41	3,15	1,83	1,48	1,23	1,17	1,03	1,48	1,23	
	4	3,05	1,56	2,38	1,26	3,01	1,56	1,48	1,16	1,25	0,91	1,47	1,16	
<b>Mato Grosso</b> Setores 3175	5	2,90	1,40	2,34	1,15	2,87	1,41	1,45	1,13	1,16	0,93	1,44	1,13	
	6	2,97	1,24	2,56	0,99	2,91	1,24	1,46	1,11	1,18	0,85	1,45	1,10	
	<b>r=0,78</b>	7	2,93	1,16	2,32	0,94	2,91	1,16	1,45	1,08	1,23	0,85	1,45	1,09
	8	2,72	0,94	2,20	0,75	2,71	0,94	1,43	1,08	1,11	0,86	1,43	1,08	
	3	1,93	1,67	1,52	1,31	1,93	1,66	0,86	0,82	0,67	0,63	0,86	0,81	
	<b>Goiás</b> Setores 5944	4	1,72	1,44	1,37	1,12	1,71	1,43	0,81	0,77	0,63	0,58	0,81	0,77
		5	1,70	1,32	1,35	1,05	1,68	1,32	0,78	0,73	0,64	0,58	0,78	0,73
		6	1,65	1,24	1,37	1,02	1,66	1,24	0,78	0,71	0,63	0,56	0,78	0,71
<b>r=0,77</b>		7	1,54	1,14	1,20	0,91	1,52	1,13	0,75	0,68	0,63	0,53	0,75	0,68
8		1,48	1,05	1,17	0,82	1,48	1,05	0,75	0,65	0,59	0,51	0,75	0,65	
3		2,72	2,23	2,24	1,69	2,68	2,21	1,61	1,58	1,26	1,24	1,60	1,58	
<b>Maranhão</b> 6376		4	2,22	2,02	1,65	1,60	2,20	2,02	1,52	1,45	1,25	1,16	1,52	1,45
		5	2,15	1,89	1,73	1,46	2,13	1,88	1,51	1,41	1,25	1,07	1,51	1,41
	6	2,12	1,78	1,69	1,38	2,10	1,77	1,50	1,35	1,16	1,06	1,50	1,35	
	<b>r=0,77</b>	7	2,06	1,72	1,59	1,36	2,03	1,72	1,49	1,33	1,14	1,09	1,49	1,32
	8	2,02	1,66	1,65	1,31	2,00	1,66	1,46	1,30	1,17	1,07	1,46	1,31	
	3	2,91	2,80	2,34	2,14	2,87	2,78	1,83	1,78	1,47	1,35	1,82	1,76	
	<b>Piauí</b> Setores 3703	4	2,54	2,47	2,02	2,01	2,51	2,47	1,78	1,74	1,39	1,33	1,77	1,74
		5	2,45	1,88	1,91	1,49	2,41	1,87	1,61	1,55	1,19	1,14	1,61	1,55
6		2,35	2,07	1,82	1,65	2,30	2,06	1,56	1,51	1,23	1,16	1,56	1,51	
<b>r=0,79</b>		7	2,09	1,92	1,64	1,55	2,07	1,92	1,55	1,45	1,31	1,20	1,55	1,45
8		2,08	1,50	1,67	1,19	2,01	1,49	1,54	1,35	1,30	1,10	1,54	1,35	
3		2,18	2,04	1,69	1,60	2,16	2,02	1,33	1,18	1,06	0,98	1,33	1,18	
<b>Ceará</b> Setores 7931		4	2,03	1,97	1,61	1,57	2,03	1,97	1,14	1,00	0,94	0,82	1,14	1,00
		5	1,86	1,82	1,54	1,46	1,84	1,82	1,12	0,92	0,84	0,73	1,11	0,92
	6	1,74	1,52	1,56	1,21	1,73	1,52	1,05	0,82	0,90	0,67	1,05	0,82	
	<b>r=0,78</b>	7	1,69	1,43	1,34	1,16	1,68	1,43	1,03	0,79	0,80	0,61	1,03	0,79
	8	1,69	1,46	1,38	1,15	1,69	1,46	1,02	0,80	0,81	0,61	1,02	0,80	
	3	2,97	2,58	2,36	1,94	2,98	2,58	1,77	1,58	1,44	1,25	1,77	1,58	
	<b>Rio G. Norte</b> Setores 2613	4	2,84	2,47	2,35	1,94	2,82	2,45	1,56	1,43	1,22	1,14	1,55	1,43
		5	2,56	2,23	2,07	1,74	2,58	2,22	1,43	1,31	1,18	0,99	1,43	1,30
6		2,54	2,05	2,11	1,70	2,52	2,05	1,37	1,24	1,11	1,00	1,37	1,24	
<b>r=0,84</b>		7	2,49	1,83	1,99	1,40	2,44	1,83	1,38	1,13	1,08	0,89	1,38	1,13
8		2,35	1,81	1,85	1,35	2,28	1,78	1,36	1,10	1,09	0,89	1,35	1,09	
3		1,64	1,49	1,28	1,16	1,63	1,49	0,92	0,90	0,75	0,68	0,93	0,91	
<b>Bahia</b> Setores 15281		4	1,52	1,40	1,15	1,07	1,51	1,39	0,86	0,81	0,66	0,63	0,86	0,81
		5	1,45	1,05	1,14	0,81	1,45	1,05	0,83	0,68	0,67	0,56	0,83	0,68
	6	1,41	1,11	1,09	0,88	1,40	1,11	0,81	0,69	0,62	0,56	0,81	0,69	
	<b>r=0,78</b>	7	1,36	1,05	1,04	0,83	1,36	1,05	0,79	0,66	0,63	0,51	0,79	0,66
	8	1,36	1,05	1,02	0,82	1,35	1,04	0,79	0,65	0,60	0,52	0,79	0,65	

(1) Algoritmo das *k*-médias

(2) Algoritmo ILS

**Gráfico 1 – Box Plots relativos aos cv's dos algoritmos ILS e  $k$ -médias por número de estratos**



### Bibliografia

**Bolfarine, H. e Bussab, Wilton O.** (2005). Elementos de Amostragem. ABE – Projeto Fisher. Editora Edgard Blücher.

**Bussab, Wilton O. e Morettin P.A.** (2003). Estatística Básica. Editora Saraiva.

**Cochran, Willian G.** (1977). Sampling Techniques. Third Edition – New York, John Wiley.

**Glover, F.** (1986). “Future paths for integer programming and links to artificial intelligence”. *Comput. Operational 13*: 533-549.

**Glover, F. e Kochenberger, G. A.** (2002). “*Handbook of Metaheuristic*”, First Edition Norwell: Kluwer Academic Publishers.

**Hartigan, J.A. e Wong, M.A.** (1979). *A k-means clustering algorithm*, Applied Statistics 28, 100-108.

**Johnson A.R. e Wichern D.W.** (2002). *Applied Multivariate Statistical Analysis*. Prentice Hall. Fifth Edition.

**Mingoti, S.A** (2007). Análise de Dados Através de Métodos de Estatística Multivariada – Uma Abordagem Aplicada. Editora UFMG.

**Resende, M.G.C e Sousa, J.P** (2004). *Metaheuristics : Computer Decision-Making*. Kluwer Academic Publishers.

**Agradecimentos:** Aos pesquisadores André Wallace Nery da Costa, Denis Paulo dos Santos, Sonia Albieri e Antonio José Ribeiro Dias, da COMEQ/IBGE, pelas sugestões e comentários.