

IPEF - ESALQ
UNIVERSIDADE DE SÃO PAULO

SCIENTIA
FORESTALIS

ISSN 1413-9324
Nº 54, Dezembro, 1998

Comparando três métodos de amostragem: métodos de distâncias, contagem de quadrats e conglomerado adaptativo

Comparing three methods of sampling: distance methods, quadrats counts and adaptative cluster

Ilka Afonso Reis
Renato Martins Assunção

RESUMO: O problema de nosso interesse é a estimação do número de plantas ou árvores individuais que ocorrem numa região do espaço na situação onde esses indivíduos acham-se distribuídos de modo agregado na região. A estimação do tamanho de uma população distribuída espacialmente numa região é geralmente feita através da amostragem de sub-regiões ou ainda através do estudo da vizinhança de pontos selecionados ao acaso na região. Para resolver esse tipo de problema, existem várias técnicas desenvolvidas. O objetivo deste estudo é comparar três dessas técnicas: a estimação por contagem de quadrats, a estimação através dos métodos de distâncias e a estimação através do conglomerado adaptativo. Em simulações com uma configuração do tipo agregada, os estimadores que obtiveram o melhor desempenho foram os do conglomerado adaptativo (Hansen-Hurwitz e Horvitz-Thompson) e o da contagem de quadrats, sendo esse último o de menor variância dentre os dois. Entretanto, em regiões de difícil deslocamento, os estimadores do conglomerado adaptativo mostram-se os mais adequados, já que os quadrats amostrados estarão também agregados e, assim, os deslocamentos serão minimizados.

PALAVRAS-CHAVE: Conglomerado adaptativo, Contagem de quadrats, Métodos de distâncias, Amostragem

ABSTRACT: The size population of trees or plants spatially distributed in a planar region is estimated through sampling of sub-regions or through the study of the neighborhood of randomly selected points in the region. This paper involves the estimation of the population size when it is spatially clumped. This clumping makes the search for individuals in the study region difficult. Our objective is to compare three proposed methods to solve this estimation problem: quadrats counts, distance methods and adaptative cluster. In the simulation study, we worked with an aggregated population. The adaptative cluster estimators (Hansen-Hurwitz and Horvitz-Thompson) and the quadrat counts estimator had better performance compared to distance methods. The quadrat counts estimator had a smaller variance than the adaptative cluster estimators. However, when the movements across the region are difficult, the adaptative cluster estimators are more indicated than quadrat counts estimator because the sampled quadrats will be aggregated and hence the movements will be minimized.

KEYWORDS: Adaptative cluster, Quadrat counts, Distance methods, Sampling

INTRODUÇÃO

Em Ecologia e Ciências Florestais, um problema freqüente é a estimação do tamanho de populações de plantas que ocorrem numa determinada região. Ao conjunto dessas plantas, irregularmente distribuídas dentro de uma região do espaço, dá-se o nome de configuração espacial de pontos, pois os dados, nesse caso, estão na forma de um conjunto de coordenadas geográficas. A distribuição desses pontos na região de estudo pode acontecer de modo regular, agregado ou aleatório, caracterizando os tipos mais simples de configuração de pontos.

Em se tratando de florestas tropicais, vários trabalhos têm mostrado que a maioria das espécies arbóreas têm distribuição espacial agregada. Estudando uma floresta tropical na Costa Rica, Hubbell (1979) encontrou que todas as espécies têm distribuição espacial agregada ou aleatória. Usando os dados de Hubbell na Costa Rica e ainda florestas tropicais nas Ilhas Virgens e na Guiana Francesa, Armesto et al. (1986) encontraram que de 75 a 100 % das espécies estudadas têm distribuição espacial agregada. Silva e Lopes (1982), num estudo da distribuição espacial de onze espécies, ocorrendo numa região da Floresta Nacional do Tapajós, em Santarém - PA, encontraram que nove delas apresentam-se espacialmente agregadas.

Quando a configuração é do tipo agregada, o problema da estimação do tamanho da população de plantas torna-se mais complicado, devido à dificuldade de encontrar os aglomerados. Existem várias técnicas desenvolvidas para resolver esse tipo de problema e o objetivo deste estudo é comparar três dessas técnicas: a estimação por contagem de quadrats (Diggle, 1983), a estimação através dos métodos de distâncias (Diggle, 1983 e Byth, 1982) e a estimação através do conglomerado adaptativo (Thompson, 1990). Como o método do conglomerado adaptativo foi proposto para o caso desse tipo de configuração,

serão estudadas apenas situações onde os pontos se encontram agregados.

Geralmente, as comparações de estimadores do tamanho da população de plantas, denotado por τ , estão restritas a um mesmo tipo de amostragem. Byth (1982) e Engeman et al. (1994), por exemplo, compararam vários estimadores baseados em distâncias. Thompson (1990) comparou estimadores adaptativos com estimadores baseados em quadrats. Entretanto, pouca informação comparativa está disponível na literatura devido à dificuldade de se avaliar simultaneamente os três métodos considerados aqui. Essa dificuldade reside na equalização do esforço amostral dos três métodos. Neste artigo, os autores propõem comparar os métodos sob a restrição de que a área esperada de observação seja igual para os três métodos.

METODOLOGIA

Uma alternativa ao mapeamento intensivo dentro de uma região A seria registrar informações de um certo número de pequenas sub-regiões, por exemplo, o número de eventos em cada uma. Usualmente, a região A é particionada em uma grade regular de N sub-regiões, geralmente quadrados, dos quais se amostra m deles por amostragem aleatória simples sem reposição. Essas sub-regiões são chamadas quadrats e os dados referentes a elas são chamados contagens de quadrats. Supondo-se que os dados compreendem contagens n_1, n_2, \dots, n_m em m quadrats, cada um com área $|B|$. Um estimador simples e não viciado para λ , o número médio de eventos por unidade de área, é a contagem total dividida pela área total dos quadrats (1) (Thompson, 1992, pag 11). Outra situação comum é supor que os even-

$$\hat{\lambda} = \sum_{i=1}^m n_i / m | B | \quad (1)$$

tos são gerados por um processo de Poisson homogêneo com intensidade λ e que quadrats de lado r são escolhidos ao acaso, possivelmente com sobreposições, dentro de A com seu centro tendo uma distribuição uniforme em A , excluindo uma borda de tamanho r . Então, $\hat{\lambda}$ é sempre um estimador não viciado para λ , mas sua variância depende do tamanho e da forma dos quadrats individuais no esquema amostral (Diggle, 1983; Rosso, 1995). Neste caso, o que se quer é um estimador para o total de eventos na região A . Assim, um estimador razoável seria $|A| \hat{\lambda}$ onde $|A|$ é a área da região A .

Nos métodos de distância, a unidade básica de amostragem é um ponto, escolhido aleatoriamente na região de estudo, e a informação é registrada na forma de distâncias aos eventos vizinhos, por exemplo, as distâncias ao primeiro, segundo, terceiro, ..., k -ésimo vizinho mais próximo. Existem, na literatura, vários estimadores para λ baseados nos métodos de distância (Diggle, 1983). No entanto, como se trata de configurações onde os pontos se acham agregados, utilizar-se-á estimador proposto por Byth (1982), que se mostra o mais adequado a esse tipo de configuração.

O modelo probabilístico envolvido nessa estimação é o de um processo de Poisson homogêneo no plano observado apenas na região A . São selecionados n pontos com distribuição uniforme em A e, a cada ponto, é associado às distâncias ao primeiro, segundo, terceiro, ..., k -ésimo vizinho mais próximo. De posse dessa amostra de tamanho n , o estimador de momentos definido por Byth como em (2) onde x_{3i}^2 é a distância do i -ésimo ponto a seu terceiro vizinho mais próximo.

$$\hat{\lambda}_1 = 3n / \pi \sum_{i=1}^n x_{3i}^2 \quad (2)$$

Para estimar o total de pontos da região em estudo, trabalhar-se-á com o estimador (3) onde $|A|$ é a área da região de estudo.

$$\hat{\lambda}_1 = 3n |A| / \pi \sum_{i=1}^n x_{3i}^2 \quad (3)$$

A amostragem através do conglomerado adaptativo refere-se a desenhos nos quais um conjunto inicial de unidades (quadrats) é selecionado por algum procedimento probabilístico e, sempre que a variável de interesse de uma unidade selecionada satisfaz a um dado critério, unidades na vizinhança daquela unidade são adicionadas à amostra. Quando a configuração de pontos em estudo tende a ser agregada, muitos dos quadrats selecionados terão valores nulos para a variável de interesse. Assim, nessas situações de amostragem, os pesquisadores têm uma inclinação para investigar as unidades vizinhas a uma unidade amostral, onde a variável obteve valores altos ou interessantes. Nesse tipo de amostragem, a amostra final em geral será maior do que a amostra inicial, caso a variável de interesse de alguma das unidades inicialmente selecionadas satisfaça a dado critério e, então, as unidades da sua vizinhança também farão parte da amostra. Nessa situação, a população consiste de N quadrats com rótulos de $1, 2, \dots, N$ e variáveis de interesse associadas $y = \{y_1, y_2, \dots, y_N\}$. A amostra s é um conjunto de rótulos identificando as unidades selecionadas para observação. Os dados consistem nos valores y observados junto com os rótulos das unidades associadas. O parâmetro de interesse a ser estimado é a média populacional μ , definida como (4), ou o total $\tau = N\mu$. Neste caso, estima-se o total τ .

$$\mu = N^{-1} \sum_{i=1}^N y_i \quad (4)$$

Assume-se que, para toda unidade i na população, uma vizinhança A_i é definida como sendo uma coleção de unidades incluindo i . Essas vizinhanças não dependem dos valores y e populacionais e , no caso do conglomerado adaptativo, a vizinhança de cada unidade consiste dos vizinhos geograficamente mais próximos. A relação de vizinhança é simétrica: se a unidade j está na vizinhança da unidade i , a unidade i está na vizinhança da unidade j . Seleccionada a amostra inicial, se uma unidade seleccionada satisfaz a condição $C = \{y : y \geq c\}$, todas as unidades dentro da sua vizinhança são adicionadas à amostra e observadas. O valor de c é positivo, arbitrário e fixo e, nesse artigo, foi definido igual a 1. Para as unidades vizinhas que satisfizerem a condição, as unidades em suas vizinhanças também serão incluídas na amostra e assim por diante. Define-se um conglomerado como o conjunto de todas as unidades que são observadas devido à seleção inicial da unidade i . Tal conjunto pode consistir da união de várias vizinhanças. Dentro de cada conglomerado, está uma subcoleção de unidades, chamada rede, com a propriedade de que a seleção de uma unidade dentro da rede levará à inclusão na amostra de todas as unidades dessa rede. Uma unidade que não satisfaz a condição mas está na vizinhança de uma unidade que a satisfaz, é chamada de unidade de fronteira. Enquanto a seleção de uma unidade na rede resulta na inclusão de todas as unidades dessa rede e de todas as unidades de fronteira, a seleção de uma unidade de fronteira não resulta na inclusão de outras unidades. É conveniente considerar uma unidade que não satisfaz a condição como uma rede de tamanho um, tal que, dados os valores y , a população pode ser particionada unicamente em redes. Nesse estudo, trabalhar-se-á com amostragem sem reposição. Assim, quando a amostra inicial de n unidades for seleccionada, essas n unidades serão distintas. A unidade i será incluída na amostra, se uma unidade da rede à qual ela pertence for seleccionada na amostra inicial, ou se uma unidade de uma rede da qual i é uma unidade de fronteira for

seleccionada. Seja m_i o número de unidades na rede da qual i faz parte e a_i , o número total de unidades nas redes das quais i é uma unidade de fronteira. Note-se que, se a unidade i satisfaz a condição, tem-se $a_i = 0$, enquanto que, se a unidade i não satisfaz a condição, $m_i = 1$. A probabilidade de seleção da unidade i em uma das n retiradas é

$$p_i = \frac{(m_i + a_i)}{N} \quad \text{para } i = 1, 2, \dots, n \quad (5)$$

A probabilidade de que a unidade i seja incluída na amostra é:

$$\pi_i = 1 - \frac{\binom{N - m_i - a_i}{n}}{\binom{N}{n}} \quad \text{para } i = 1, 2, \dots, n \quad (6)$$

O tamanho da amostra final, v , é uma variável aleatória, que pode ser pensada como a soma de N variáveis aleatórias Bernoulli, cada uma com média π_i , $i = 1, 2, \dots, N$. Assim, no caso dos desenhos de conglomerado adaptativo, o valor esperado de v é dado por:

$$E(v) = \sum_{i=1}^N \pi_i$$

Nos desenhos amostrais de conglomerado adaptativo, as probabilidades de seleção não são conhecidas para cada unidade na amostra. Um estimador não viciado, sugerido por Thompson (1990), é uma modificação do estimador de Hansen-Hurwitz, usando as unidades que não satisfazem a condição somente quando elas forem seleccionadas na amostra inicial. O estimador modificado é baseado nas probabilidades de que uma unidade da rede seja interceptada pela amostra inicial. Seja y_i a rede que inclui a unidade i e m_i o número de unidades na rede. Seja w_i a média das observações na rede que inclui a i -ésima unidade da amostra inicial, tal que

$$w_i = \frac{1}{m_i} \sum_{j \in \psi_i} y_j \quad (7)$$

O estimador modificado para amostragem sem reposição é

$$\hat{\mu}_1 = \frac{1}{n} \sum_{i=1}^n w_i \quad (8)$$

A variância é dada por

$$\text{var}(\hat{\mu}_1) = \frac{(N-n)}{Nn(N-1)} \sum_{i=1}^N (w_i - \hat{\mu}_1)^2 \quad (9)$$

Um estimador não viciado para essa variância é

$$\hat{\text{var}}(\hat{\mu}_1) = \frac{(N-n)}{Nn(N-1)} \sum_{i=1}^N (w_i - \hat{\mu}_1)^2 \quad (10)$$

Thompson (1990) também mostrou que, para desenhos amostrais onde a probabilidade π_i de que a unidade i seja incluída na amostra é conhecida para todas as unidades, o estimador de Horvitz-Thompson, no qual cada valor y é dividido pela probabilidade de inclusão associada, é um estimador não viciado da média populacional. Como as probabilidades de inclusão não são conhecidas para todas as unidades incluídas na amostra nos desenhos de conglomerado adaptativo, um estimador não viciado para a média populacional pode ser obtido modificando o estimador de Horvitz-Thompson para utilizar as observações que não satisfazem à condição somente quando elas são incluídas na amostra inicial. Então, a probabilidade de que uma unidade seja utilizada no estimador pode ser calculada, mesmo que sua verdadeira probabilidade de inclusão na amostra seja desconhecida. O estimador modificado é baseado nas probabilidades de que a amostra inicial intercepte redes. Se a amostra inicial é selecionada sem reposição, define-se

$$\pi_i^* = 1 - \frac{\binom{N-m_i}{n}}{\binom{N}{n}} \quad (11)$$

onde m_i é definido como o número de unidades na rede que inclui a unidades i . Para uma unidade que não satisfaz a condição, $m_i = 1$. Define-se a variável indicadora J_i que valerá zero se a i -ésima unidade na amostra não satisfaz a condição e não foi selecionada na amostra inicial. Caso contrário, $J_i = 1$. O estimador modificado de Horvitz-Thompson, escrito em termos das unidades individuais da amostra, é

$$\hat{\mu}_2 = \frac{1}{N} \sum_{i=1}^N \frac{y_i J_i}{\pi_i^*} \quad (12)$$

Para obter a variância de $\hat{\mu}_2$ muda-se a notação a fim de se trabalhar com as redes, nas quais a população é particionada. Seja K o número de redes na população e seja ψ_k o conjunto de unidades da k -ésima rede. Seja x_k o número de unidades na rede k . O total dos valores y na rede k será denotado por $y_k^* = \sum_{i \in \psi_k} y_i$. A probabilidade π_i^* de que a unidade i seja utilizada no estimador é a mesma para todas as unidades dentro de uma dada rede k e essa probabilidade comum será denotada por α_k . Assim,

$$\alpha_k = 1 - \frac{\binom{N-x_k}{n}}{\binom{N}{n}} \quad (13)$$

para amostragem simples sem reposição.

Seja a variável indicadora z_k igual a um se alguma unidade da k -ésima rede está na amostra inicial e z_k igual a zero, caso contrário. Usando essa nova notação, o estimador pode ser escrito como

$$\hat{\mu}_2 = \frac{1}{N} \sum_{k=1}^K \frac{y_k^* z_k}{\alpha_k} \quad (14)$$

Para uma rede que não esteja na amostra, z_k será zero e também será zero para as redes de tamanho um como uma unidade de fronteira não selecionada inicialmente. A probabilidade α_{kh} de que a amostra inicial contenha ao menos uma unidade em cada uma das redes k e h é

$$\alpha_{kh} = 1 - \frac{\binom{N-x_k}{n} + \binom{N-x_h}{n} - \binom{N-x_k-x_h}{n}}{\binom{N}{n}} \quad (15)$$

quando a amostra é selecionada sem reposição. Com a convenção de que $\alpha_{kk} = \alpha_k$, a variância pode ser escrita como

$$\text{var}(\hat{\mu}_2) = \frac{1}{N^2} \sum_{k=1}^K \sum_{h=1}^K y_k^* y_h^* \frac{(\alpha_{kh} - \alpha_k \alpha_h)}{(\alpha_k \alpha_h)} \quad (16)$$

Um estimador não viciado para a variância em (16) é

$$\hat{\text{var}}(\hat{\mu}_2) = \frac{1}{N^2} \sum_{k=1}^K \sum_{h=1}^K y_k^* y_h^* z_k z_h \frac{(\alpha_{kh} - \alpha_k \alpha_h)}{(\alpha_k \alpha_h \alpha_{hk})} \quad (17)$$

EXPERIMENTOS SIMULADOS

Como o método do conglomerado adaptativo foi proposto para configurações de eventos do tipo agregada, utiliza-se o processo proposto por Neyman e Scott (1958) para gerar a população usada nas simulações. Essa população foi gerada dentro de um quadrado de lado 10. O número de “pais” foi gerado de uma distribuição de Poisson com média 10 e o número de “filhos”, de uma Poisson com média 20. As coordenadas dos eventos “pais” foram geradas de uma distribuição Uniforme (0;10) e as coordenadas dos eventos “filhos”, de uma distribuição Normal com média na coordenada do evento “pai” e desvio-padrão igual a 0.2. A configuração espacial da população escolhida é mostrada na Figura 1. Essa população é constituída de 233 eventos (total populacional a ser estimado) e a região é dividida em 100 quadrats de área igual a uma unidade.

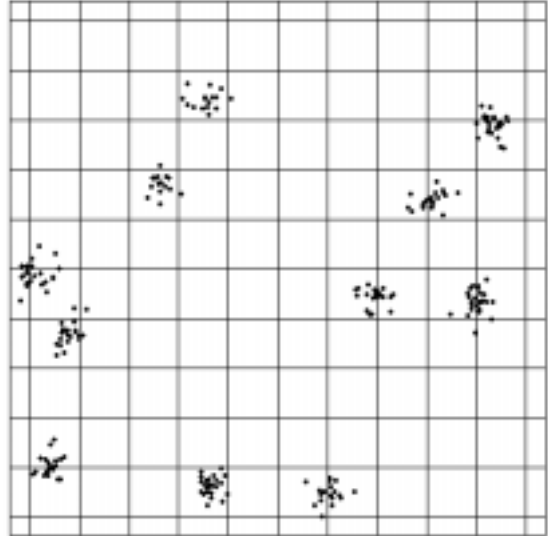


Figura 1

População de pontos gerada segundo um Processo Aglomerado de Poisson.

Point population generated according to a Poisson Clustered Process.

Para cada tamanho de amostra inicial ($n = 10, 15, 20, 25, 30$ e 35), foram realizadas 200 simulações, nas quais uma amostra foi retirada e as estimativas do total populacional segundo cada um dos quatro estimadores estudados foram calculadas. Para comparar os estimadores, utiliza-se o erro relativo médio (ERM), erro quadrático médio (EQM), vício ao quadrado (B^2) e variância (s^2). A medida ERM é definida como a média sobre as 200 simulações da diferença entre a estimativa e o valor populacional dividida pelo valor populacional. Note-se que o ERM pode ser negativo. A medida EQM é definida como a média sobre as 200 simulações do quadrado da diferença entre a estimativa e o valor populacional. A medida B^2 é o quadrado da diferença entre a média das 200 estimativas e o valor populacional do parâmetro. Finalmente, s^2 é a variância das 200 estimativas calculada como a soma dos quadrados das diferenças entre a estimativa e a média das estimativas dividida por 199. Observe que EQM não é exatamente igual à soma de B^2 e s^2 porque, na definição de s^2 , preferiu-se dividir por

199 para obter uma estimativa da variância não viciada.

A fim de tornar justa a comparação entre os estimadores, compararam-se os métodos de estimação sob a restrição de que a área de procura esperada seja igual para todos eles. Para isso, foi proposto igualar o tamanho esperado da área de procura para os três métodos, a partir do tamanho da área de procura esperado no método do conglomerado adaptativo, determinando assim, para cada método, qual deveria ser o tamanho de amostra utilizado.

No método da contagem de quadrats, a área total de procura é simplesmente o número de quadrats amostrados, pois os quadrats têm área igual a 1 unidade. No entanto, sabe-se que, no método do conglomerado adaptativo, o tamanho total da área de procura é aleatório, já que o tamanho da amostra final é aleatório. No método de distâncias, tem-se $A_D = \pi x_3^2$, onde A_D é a área de procura de um ponto amostral, é também uma variável aleatória, já que não se sabe *a priori* qual será a distância até o terceiro vizinho mais próximo utilizado no estimador. Portanto, A_T , a área total de procura no método de distâncias, também será uma variável aleatória.

Assim, a partir do tamanho da amostra inicial, o tamanho esperado da amostra final no método do conglomerado adaptativo foi calculado e usado como tamanho de amostra no caso da contagem de quadrats.

Para o caso do método de distâncias, sabe-se que

$$A_T = \sum_{i=1}^q A_{Di}$$

onde A_{Di} é a área de procura do i -ésimo ponto amostral e q é o tamanho da amostra a ser usada no método de distâncias. Tomando a esperança de A_T , tem-se

$$E[A_T] = \sum_{i=1}^q E[A_{Di}] = q \times E[A_D]$$

Através dessa expressão, conhece-se o valor da esperança de A_T for conhecido o valor de q , o

tamanho da amostra. Porém, é interessante justamente o valor de q . Assim, define-se o valor de $E[A_T]$ como sendo igual ao tamanho da área total de procura esperada no método do conglomerado adaptativo (tamanho esperado da amostra final) para se extrair qual deveria ser o tamanho da amostra no método de distâncias (q) que levaria a uma área total de procura esperada igual à área total de procura no método do conglomerado adaptativo. Os pontos amostrais eram selecionados aleatoriamente dentre os vértices dos quadrats, com exceção dos vértices pertencentes às bordas da região de estudo.

RESULTADOS

A Tabela 1 contém os resultados para os estimadores do conglomerado adaptativo (Hansen-Hurwitz e Horvitz-Thompson), métodos de distâncias (Byth) e contagem de quadrats, ao final de 200 simulações. Nas Figuras 2 e 3, observa-se que os estimadores de Hansen-Hurwitz, Horvitz-Thompson e o da contagem de quadrats mostram-se, em geral, não viciados para todos os tamanhos de amostra inicial. Nota-se ainda uma redução na variância desses estimadores à medida que aumenta o tamanho da amostra inicial, sendo o de contagem de quadrats, dentre os três, o de menor variância. O estimador proposto por Byth subestimou o verdadeiro valor do total de pontos na região em quase todas as simulações, com exceção de alguns pontos discrepantes. No entanto, é o que possui a menor variância dentre os quatro estimadores estudados.

A Figura 4 mostra os erros quadráticos para os estimadores estudados em todos os tamanhos de amostra inicial. Confirmam-se os resultados comentados anteriormente, onde o estimador de contagem de quadrats mostra-se o estimador de menor erro quadrático médio, dentre os quatro, seguido pelo estimadores do conglomerado adaptativo.

Tabela 1

Resultados para os estimadores Hansen-Hurwitz (HH), Horvitz-Thompson (HT), Byth e contagem de quadrats, em 200 simulações e diferentes tamanhos de amostra inicial (tamanhos de amostra final esperados estão entre parêntesis)

Results for Hansen-Hurwitz (HH), Horvitz-Thompson (HT), Byth e Quadrats counts estimators in 200 simulations and different sizes of initial samples (sizes of expected final sample are in parenthesis)

Amostra inicial		HH	HT	Byth	Quadrat
n = 10 (27)	ERM (%)	-4.243	-2.631	-57.485	-3.051
	EQM	15387.306	14778.840	17258.700	6038.214
	B ²	90.345	34.728	16580.800	46.694
	s ²	15373.830	14818.204	681.289	6021.627
n = 15 (37)	ERM (%)	-1.585	-3.925	-58.055	-1.635
	EQM	9173.032	7421.032	17558.008	4281.997
	B ²	12.610	77.316	16911.408	13.411
	s ²	9206.453	7380.622	649.849	4290.036
n = 20 (46)	ERM (%)	2.447	3.037	-59.743	-0.747
	EQM	5458.593	4784.609	18275.159	2575.756
	B ²	30.049	46.268	17908.690	2.802
	s ²	5455.824	4762.152	368.310	2585.884
n = 25 (53)	ERM (%)	1.099	1.593	-60.358	-0.720
	EQM	5174.911	4172.102	18487.342	1704.407
	B ²	6.065	12.741	18279.585	2.602
	s ²	5194.820	4180.263	208.802	1710.356
n = 30 (59)	ERM (%)	-4.459	-2.080	-60.896	0.159
	EQM	3470.581	2843.639	18846.119	1816.999
	B ²	99.745	21.718	18607.061	0.127
	s ²	3387.776	2836.102	240.260	1826.002
n = 35 (65)	ERM (%)	-0.213	1.028	-58.869	0.361
	EQM	3150.842	2188.968	17619.137	1387.160
	B ²	0.228	5.298	17388.778	0.652
	s ²	3166.446	2194.644	231.518	1393.475

ERM: erro relativo médio; EQM: erro quadrático médio; B² : vício ao quadrado; s²: variância amostral.

Na Figura 5, onde estão os vícios ao quadrado dos estimadores, para cada tamanho de amostra inicial, nota-se mais uma vez que o estimador proposto por Byth é o que possui o maior vício.

CONCLUSÃO

O problema de estimação do tamanho da população de plantas de uma determinada região,

torna-se particularmente mais difícil quando essas plantas acham-se distribuídas de modo agregado na região de estudo, formando os chamados aglomerados. Esse tipo de distribuição espacial é comumente encontrado entre as espécies de plantas de florestas tropicais (Hubbell, 1979; Armesto et al., 1986; Silva e Lopes, 1982). A dificuldade está em encontrar esses aglomerados. Roesch (1993) estudou uma adaptação dos estimadores propostos por Thompson, na qual a

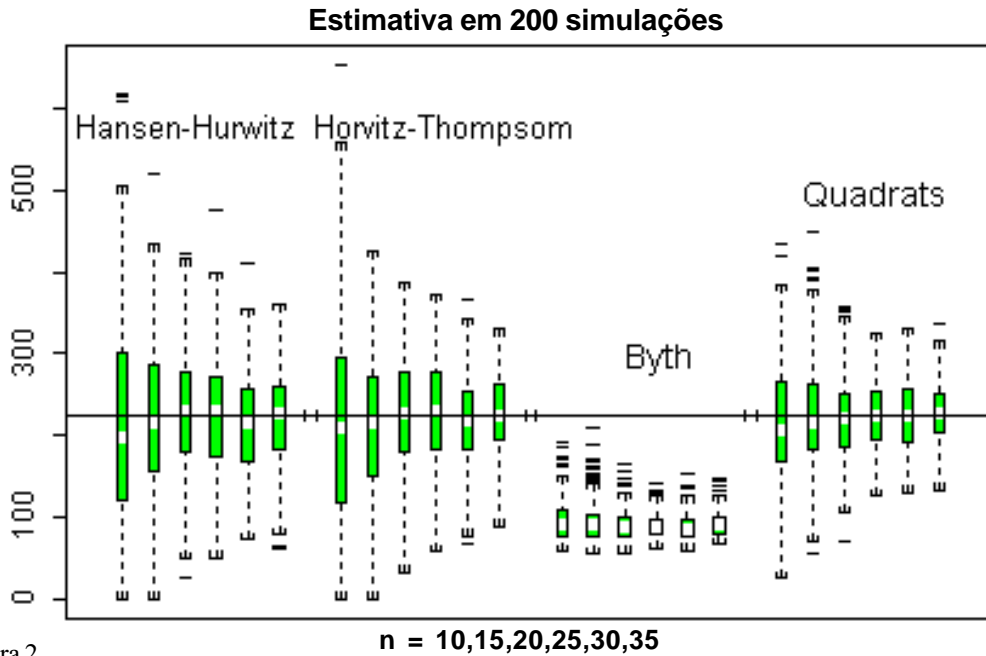


Figura 2

Estimativas para o total populacional segundo os estimadores HH, HT, Byth e Quadrats com diversos tamanhos de amostra inicial (n)

Total populational estimates according to HH, HT, Byth e Quadrats estimators for several sizes of initial sample (n).

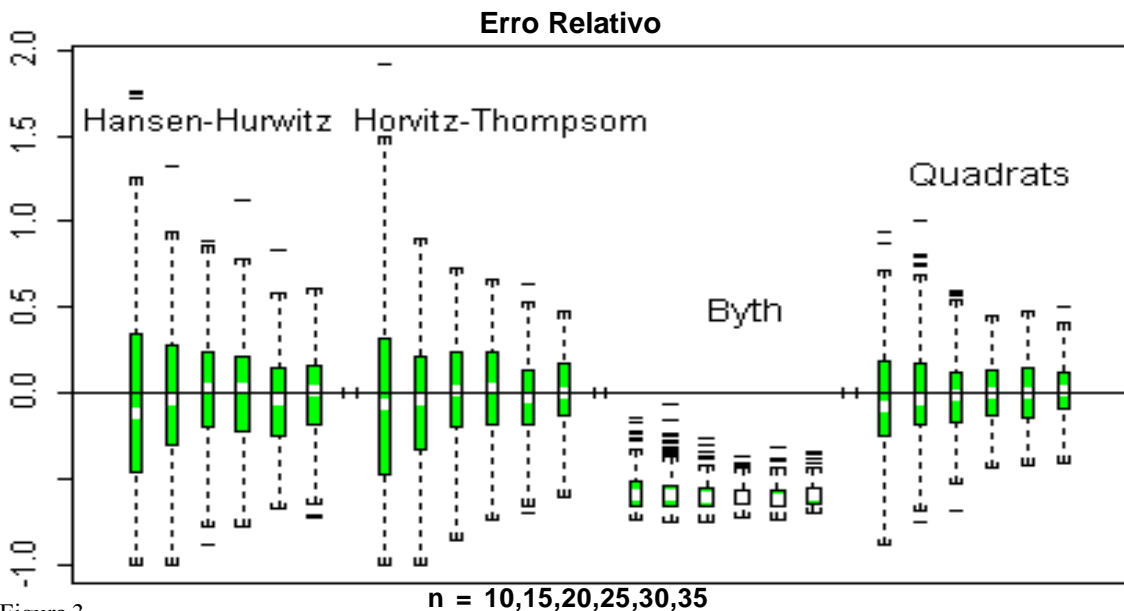


Figura 3

Erro relativo das estimativas para o total populacional segundo os estimadores HH, HT, Byth e Quadrats com diversos tamanhos de amostra inicial (n)

Relative error of total population estimates according to HH, HT, Byth e Quadrats estimators for several sizes of initial sample (n).

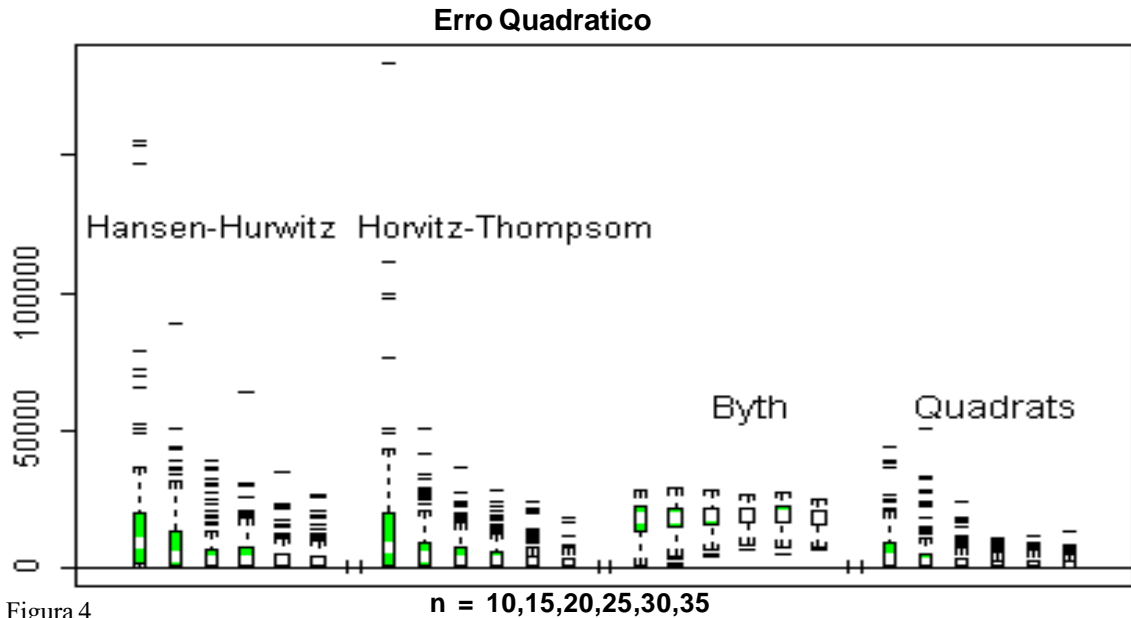


Figura 4

Erro quadrático das estimativas para o total populacional segundo os estimadores HH, HT, Byth e Quadrats com diversos tamanhos de amostra inicial (n)

Squared error of total populational estimates square error according to HH, HT, Byth e Quadrats estimators for several sizes of initial sample (n).

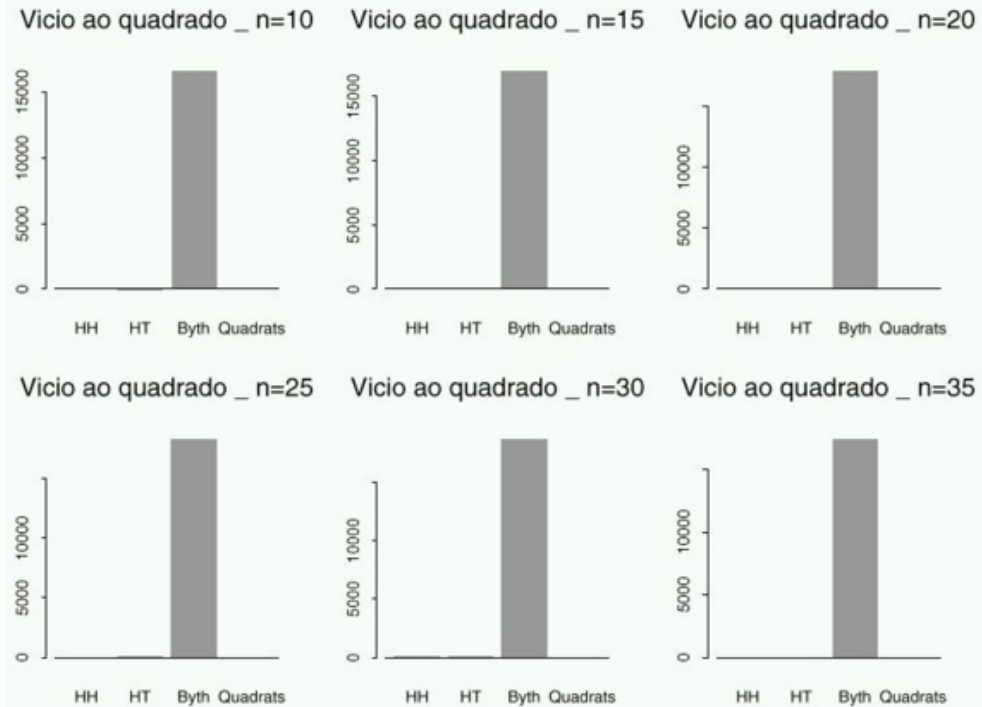


Figura 5

Vício (ao quadrado) das estimativas para o total populacional segundo os estimadores HH, HT, Byth e Quadrats com diversos tamanhos de amostra inicial (n)

Squared Bias of total populational estimates according to HH, HT, Byth e Quadrats estimators for several sizes of initial sample (n).

amostra inicial é selecionada com probabilidades desiguais. Contudo, não observou nenhuma vantagem desses estimadores sobre os propostos por Thompson no caso do conglomerado adaptativo.

Com uma configuração de pontos do tipo agregada, gerada em computador e mantida fixa, foram realizadas simulações das quais eram retiradas amostras, de vários tamanhos, nessa região de estudo e calculadas as estimativas para o total de pontos segundo cada um dos quatro estimadores estudados, Hansen-Hurwitz, Horvitz-Thompson, Byth e contagem de quadrats.

Ao final das simulações, o estimador proposto por Byth (método de distâncias) foi o que apresentou a menor variância, em todos os tamanhos de amostra, tendo, no entanto, o maior vício dentre todos os estimadores. Os estimadores de Hansen-Hurwitz e Horvitz-Thompson (conglomerado adaptativo) e o de contagem de quadrats, mostraram-se os estimadores de menor vício, sendo o estimador de contagem de quadrats o de menor variância dentre os três. A diferença entre

os estimadores do conglomerado adaptativo e de contagem de quadrats, em termos de qualidade de estimação medida pelo EQM, diminui à medida que aumenta o tamanho da amostra inicial.

Desse modo, em situações onde o deslocamento na região de estudo é difícil (pântanos, matas fechadas, acidentes geográficos etc.) e houver razões para se acreditar que a população a ser estudada distribui-se de modo agregado nessa região, os estimadores de Hansen-Hurwitz e Horvitz-Thompson são os mais adequados, já que, pela própria idéia do conglomerado adaptativo, os quadrats amostrados também estarão dispostos em aglomerados e os grandes deslocamentos serão menos frequentes. No caso da contagem de quadrats, estes estarão dispersos na região de estudo, implicando em maiores deslocamentos.

Nessa situação, o uso do método do conglomerado adaptativo para estimar o tamanho dessa população pode traduzir-se numa redução do trabalho de campo sem prejuízo da qualidade da estimação.

AUTORES E AGRADECIMENTOS

RENATO MARTINS ASSUNÇÃO é Professor Adjunto do Departamento de Estatística da UFMG - Caixa Postal 702 - Belo Horizonte, MG - 30161-970. É também pesquisador do CEDEPLAR-UFMG.

ILKA AFONSO REIS é Professora Assistente do Departamento de Estatística da UFMG - Caixa Postal 702 - Belo Horizonte, MG - 30161-970.

Os autores agradecem o apoio do CNPq e FAPEMIG.

REFERÊNCIAS BIBLIOGRÁFICAS

ARMESTO, J.J.; MITCHELL, J.D.; VILLAGRAN, C. A comparison of spatial patterns of trees in some tropical and temperate forests. *Biotropica*, v.18, n.1, p.1-11, 1986.

BYTH, K.B. On robust distance-based intensity estimators. *Biometrics*, v.38, p.127-135, 1982.

DIGGLE, P.J. *Statistical analysis of spatial point patterns*. New York: Academic Press, 1983. 148p.

ENGEMAN, R.M.; SUGIHARA, R.T.; PANK, L.F.; DUSENBENG, W.E. A comparison of plotless density estimators using Monte Carlo simulation. *Ecology*, v.75, n.6, p.1769-1779, 1994.

HUBBELL, S.P. Tree dispersion, abundance and diversity in a tropical dry forest. *Science*, v.203, n.4387, p.1299-1309, 1979.

- NEYMAN, J.; SCOTT, E.L. Statistical approach to problems in cosmology (with discussion). *Journal of Royal Statistical Society, Série B*, v.20, p.1-43, 1958.
- ROESCH JR., F. A. Adaptative conglomerado sampling for forest inventories. *Forest science*, v.4, p.65-669, 1993.
- ROSSO, S. Dimensionamento amostral em estudos descritivos de comunidades de organismos bênticos sésseis e semi-sésseis. *Oecologia brasiliensis*, v.1, p.193-223, 1995.
- SILVA, J.N.M.; LOPES, J.C.A. Distribuição espacial de árvores na Floresta Nacional do Tapajós. *Circular técnica. EMBRAPA - CPATU*, n.26, 1982.
- THOMPSON, S.K. Adaptative cluster sampling. *Journal of the American Statistical Association*, v.85, p.412, 1990.
- THOMPSON, S.K. *Sampling*. New York: John Wiley, 1992. 343p.

• **TRABALHOS DE PESQUISA** • EFEITO DA CAMADA DE RESÍDUOS FLORESTAIS NA COMPACTAÇÃO DO SOLO CAUSADA PELO TRANSPORTE PRIMÁRIO DA MADEIRA. • *FERNANDO SEIXAS; EZÉR DIAS DE OLIVEIRA JÚNIOR; CÍNTIA RODRIGUES DE SOUZA* • CARACTERÍSTICAS DO SISTEMA RADICULAR EM POVOAMENTOS DE EUCALIPTOS PROPAGADOS POR SEMENTES E ESTACAS. • *SÉRGIO LUÍS DE MIRANDA MELLO; JOSÉ LEONARDO DE MORAES GONÇALVES; LUIS EUGÊNIO GONÇALVES DE OLIVEIRA* • NÍVEIS CRÍTICOS DE FÓSFORO NO SOLO E NAS FOLHAS PARA A IMPLANTAÇÃO DE *EUCALYPTUS GRANDIS* HILL EX MAIDEN, EM QUATRO TIPOS DE SOLOS. • *JOÃO JOSÉ ISMAEL; SÉRGIO VALIENGO VALERI; LENINE CORRADINI; SILVIO FERNANDES ALVARENGA; CELINA FERRAZ DO VALLE; MANOEL EVARISTO FERREIRA; DAVID ARIIVALDO BANZATTO* • ADIÇÃO DE NUTRIENTES AO SOLO EM SISTEMA AGROFLORESTAL DO TIPO “CULTIVO EM ALÉIAS” E EM CERRADO NA REGIÃO DE BOTUCATU, SP. • *ÁLVARO LUIZ MAFRA; ANDREAS ATILA DE WOLINSK MIKLÓS; HUGO LUIZ VOCURCA; ALEXANDRE HUMBERTO HARKALY; EDUARDO MENDOZA* • VARIAÇÃO GENÉTICA DE INDICADORES DE TENSÃO DE CRESCIMENTO EM CLONES DE *EUCALYPTUS UROPHYLLA*. • *LOTHAR SCHACHT; JOSÉ NIVALDO GARCIA; ROLAND VENCOVSKY* • ESTABILIDADE DIMENSIONAL DO COMPENSADO UTILIZANDO RESINA DE ALTA REATIVIDADE. • *DIMAS AGOSTINHO DA SILVA; IVAN TOMASELLI; SETSUO IWAKIRI* • DURABILIDADE NATURAL DE 46 ESPÉCIES DE MADEIRA AMAZÔNICA EM CONTATO COM O SOLO EM AMBIENTE FLORESTAL. • *MARIA APARECIDA DE JESUS; JOSÉ WELLINGTON DE MORAIS; R. LIÉGE SOUZA DE ABREU; MARIA DE FÁTIMA C. CARDIAS* • ESTIMATIVAS E TESTES DA DISTRIBUIÇÃO DE FREQUÊNCIA DIAMÉTRICA PARA *EUCALYPTUS CAMALDULENSIS*, ATRAVÉS DA DISTRIBUIÇÃO S_B , POR DIFERENTES MÉTODOS DE AJUSTE. • *JOSÉ ROBERTO S. SCOLFORO; AGUINALDO THIERSCHI* • AVALIAÇÃO DE IMAGENS-ÍNDICE E IMAGENS-PROPORÇÃO NA IDENTIFICAÇÃO DE PLANTIOS FLORESTAIS DESFOLHADOS POR GEADAS E PELO ATAQUE DE INSETOS. • *FLÁVIO JORGE PONZONI* • COMPARANDO TRÊS MÉTODOS DE AMOSTRAGEM: MÉTODOS DE DISTÂNCIAS, CONTAGEM DE QUADRATS E CONGLOMERADO ADAPTATIVO. • *ILKA AFONSO REIS; RENATO MARTINS ASSUNÇÃO* • PROGRAMAS EDUCATIVOS COM FLORA E FAUNA (EXPRESSIONES DA BIODIVERSIDADE) E A EDUCAÇÃO AMBIENTAL. • *JOÃO LUIZ PEGORARO; MARCOS SORRENTINO* • **COMUNICAÇÕES** • ZONEAMENTO ECOLÓGICO DAS BACIAS DO PARANÁ E ALTO PARAGUAI (MS) PARA *EUTERPE EDULIS* MART. • *OMAR DANIEL; SILVIO NOLASCO OLIVEIRA NETO*