CDU 311.213.2:634.0.5
AGRIS K20

AMOSTRAGEM SISTEMÁTICA COM IGUAL PROBABILIDADE EM INVENTÁRIO FLORESTAL

JOSÉ ANTÔNIO ALEIXO DA SILVA
Prof. Adjunto do Departamento de Ciência Florestal/UFRPE.

ROBERT L. BAILEY
Prof. Titular da Escola de Florestas da Universidade da Georgia, USA.

A amostragem sistemática com igual probabilidade é um sistema amostral freqüentemente usado em inventários florestais devido as facilidades oferecidas por este sistema. A principal restrição ao uso do mesmo é com relação a um estimador sem tendência para a variância quando se usa só uma amostra aleatorizada. Um estimador sem tendência para tal variância pode ser usado através da fórmula da variância da amostragem em conglomerados, pois se usa várias amostras aleatorizadas. O presente trabalho é uma revisão bibliográfica dos métodos mais utilizados em amostragem sistemática com igual probabilidade, bem como apresenta uma modificação no sistema de amostragem sistemática repetida, o qual acreditamos ser o que proporciona maior número possível de amostras aleatórias distribuídas sistematicamente.

INTRODUÇÃO

Amostragem sistemática com aleatorização da primeira unidade de amostra tem sido freqüentemente usada em inventário florestal devido ao aspecto prático e conveniência. Administrativamente é simples, os custos são baixos e os resultados obtidos são semelhantes aos obtidos por uma amostragem inteiramente aleatória.

Com amostragem sistemática, a amostra tende a ser mais uniformemente distribuída na população florestal e isto, algumas vezes, proporciona resultados mais precisos que a amostragem inteiramente aleatória (COCHRAN, 1977).

Segundo RAJ (1968), amostragem sistemática é o mais conveniente método de seleção de unidade amostrais quando tais unidades estão seqüencialmente numeradas de 1 a N, porque com a seleção da primeira unidade de amostra, todas as outras unidades a serem amostradas estão automáticamente localizadas.

Uma tradicional definição de amostragem sistemática é que ela corresponde a seleção aleatória de uma primeira unidade de amostra na população, e todas outras unidades de amostras a serem mensuradas estão diretamente associadas a primeira unidade em função de uma distância fixa K chamada de "intervalo de amostragem", que é um número inteiro igual ou próximo a razão entre o tamanho da população e o tamanho da amostra (N/n). O valor de K é determinado no planejamento do inventário.

A aleatorização de uma única unidade de amostra é provavelmente a maior desvantagem da amostragem sistemática porque não permite o cálculo de uma estimativa sem tendência para a variância da população. Entretanto uma maneira de eliminar essa desvantagem é usar o sistema de amostragem com vários pontos aleatórios, sendo que cada um deles definirá um conjunto de unidades amostrais.

A teoria da amostragem sistemática é derivada do mais simples caso de amostragem em conglomerados, onde um único conglomerado é amostrado e não existe sub-amostragem dentro do conglomerado. A diferença entre amostragem sistemática e conglomerados é que existem k conglomerados, os elementos dentro de qualquer conglomerado estão separàdos entre si por uma distância K, e não existe em qualquer conglomerado dois elementos juntos a serem amostrados (MADOW, 1946).

Seleção da Primeira Amostra Aleatória

Para qualquer população florestal existem duas situações referentes a amostragem sistemática:

a) o número de unidades amostrais na população é exatamente um múltiplo de K, isto é , N=nK, onde n é o tamanho da amostra;

b) N não é múltiplo de K

No caso a, considerando que a primeira amostra aleatória define um conjunto de unidades de tamanho j, pode-se representar a distribuição dos conjuntos de amostras como:

CONJUNTO DE AMOSTRAS	COMPOSIÇÃO
1	1, K+1, 2K+1, ... , (j-1)K+1
2	2, K+2, 2K+2, ... , (j-1) K+2
.	.
.	.
.	.
K	K, 2K, 3K, ... , jK

Então o procedimento de seleção é a aleatorização de um número i, onde i=1, 2, ... ,N e em função deste número completar a distribuição das unidades de amostra no campo, sendo que cada uma delas deve ficar separada das outras vizinhas por uma distância K em sentido vertical e horizontal. Se todos os conjuntos de unidades amostrais forem considerados na amostra, o valor da média obtida pela amostragem sistemática é uma estimativa sem tendência da média da população.

Considere,

$$\bar{y} = (1/n)\sum_{i=1}^{n} y_{ij}$$

ser a média da amostra obtida na amostragem sistemática.

Então,

$$E(\bar{y}_{sist}) = (1/K)(\bar{y}_1 + \bar{y}_2 + \ldots + \bar{y}_k)$$

porque existem somente K conjuntos que podem ser amostrados, cada um com probabilidade 1/K. Então, E $(\bar{y}_{sist})$ pode ser escrita como:

$$E(\bar{y}_{sist}) = (1/K)\left[(Y_1/n) + (Y_2/n) + \ldots + (Y_k/n)\right]$$

onde $$Y_i = \sum_{i=1}^{n} y_i$$

$$E(\bar{y}_{sist}) = (1/Kn)(Y_1 + Y_2 + \ldots Y_k)$$

$$E(\bar{y}_{sist}) = (1/N)\sum_{i=1}^{k} Y_i = \bar{Y}$$

Mas, se $N \neq nK$, o uso deste procedimento proporciona uma estimativa com tendência para Y.

Para provar esta afirmação, consideremos uma população com tamanho N = 14 e K = 5, o que proporciona um valor não inteiro para n, conseqüentemente $N \neq nK$. Neste caso dois tamanhos de conjuntos podem ocorrer (três e duas unidades de amostra por conjunto).

CONJUNTOS

1	2	3	4	5
y_1	y_2	y_3	y_4	y_5
y_6	y_7	y_8	y_9	y_{10}
y_{11}	y_{12}	y_{13}	y_{14}	—

Então,

$$E(\bar{y}_{sist}) = (1/5)[(1/3(Y_1 + Y_2 + Y_3 + Y_4) + (1/2)(Y_5)]$$

$$E(\bar{y}_{sist}) = (1/15)(Y_1 + Y_2 + Y_3 + Y_4) + (1/10)(Y_5)$$

Mas, $\bar{Y} = (1/14)(y_1 + y_2 + \dots y_{14})$

$$E(\bar{y}_{sist}) \neq \bar{Y}$$

Considerando a mesma população, uma estimativa sem tendência para $\bar{Y}$ pode ser obtida usando o seguinte procedimento.

Uma unidade de amostra j ésima é aleatorizada na população (exemplo, j é a nona unidade). Então, j/K = 9/5 com um remanescente
r = 4. Observe que r < K, sendo que r pode assumir os seguintes valores: 1, 2, 3 e 4. Quadro r = 1 seleciona-se a amostra Y_1 , quando r = 2 seleciona-se a amostra Y_2, quando r = 3 seleciona-se a amostra

Y_3, quando r = 4 seleciona-se a amostra Y_4, quando r = 0 seleciona-se a amostra Y_5 como ponto de partida para distribuição de todas amostras, obedecendo o mesmo princípio de que uma unidade de amostra deve ficar a uma distância K em sentido horizontal e vertical das outras amostras vizinhas que também serão amostradas.

Uma vantagem deste procedimento é que a probabilidade de qualquer conjunto ser selecionado é n/N e não 1/K. Neste caso a probabilidade de seleção do conjunto 4, composto de y_4, y_9, y_{14} é 3/14 onde a probabilidade de cada elemento é 1/14.

Semelhantemente a probabilidade de seleção do conjunto 5 é 2/14.

Pr (selecionar o conjunto 4) = Pr $(y_4) + Pr(y_9) + Pr(y_{14})$
Pr (selecionar o conjunto 4) = Pr (1/14) + (1/14) + (1/14) = (3/4)
Então,

$$E(\bar{y}_{sist}) = (3/14)(y_1 + y_6 + y_{11}) + (3/14)(y_2 + y_7 + y_{12}) + (3/14)$$

$$(y_3 + y_8 + y_{13}) + (3/14)(y_4 + y_9 + y_{14}) + (2/14)(y_5 + y_{10})$$

$$E(\bar{y}_{sist}) = \left(\frac{3}{14}\right)\left[\frac{1}{3}(y_1 + y_6 + y_{11})\right] + \left(\frac{3}{14}\right)\left[\frac{1}{3}(y_2 + y_7 + y_{12})\right]$$

$$+ \ldots + \left(\frac{2}{14}\right)\left[\frac{1}{2}(y_5 + y_{10})\right]$$

$$E(\bar{y}_{sist}) = \left(\frac{1}{14}\right)(y_1 + y_2 + y_3 + \ldots + y_{14}) = \bar{Y}$$

Este procedimento produz estimativas sem tendência para $\bar{Y}$ em ambos casos, N = nK e N $\neq$ nK, mas não pode ser usado se o tamanho da população N não for conhecido, porque a probabilidade de seleção de qualquer conjunto de amostras é n/N (YAMANE, 1967).

A principal desvantagem da amostragem sistemática com aleatorização somente da primeira unidade de amostra, é que depois que esta é selecionada, todas as amostras que pertecem a seu conjunto tem probabilidade de serem amostradas igual a 1, enquanto que aquelas que não pertencem a tal conjunto tem probabilidade de serem selecionadas igual a 0. Isto significa que a maioria das unidades amostrais da população são excluidas da seleção por causa do sistema utilizado pela amostragem sistemática. Isto é contrário ao princípio básico da teoria de amostragem (FAO, 1982). Neste caso qualquer cálculo da variância possuí tendência.

Para melhor explanação deste fato, considere o caso geral de amostragem com igual probabilidade.

Assumindo,

n = tamanho da amostra
S_n = amostra S de tamanho n.

$A = \binom{N}{n}$ = número possível de amostras

$B = \binom{N-1}{n-1}$ = número possível de amostras contendo um elemento Ui.

$C = \binom{N-2}{n-2}$ = número possível de amostras contendo qualquer par de elementos Ui, Uj.

$$\Pr(U_i \in S_n) = (B / A) = [(N-1)!/(n-1)!(N-n)!][n!(N-n)!/N!]$$
$$\Pr(U_i \in S_n) = (n / N)$$

Usando a técnica de variáveis ponderadas, cujo teorema diz que se E (Wi) = constante para todos valores de i, $\hat{Y} = \sum_{i=1}^{n} W_i Y_i$ é uma estimativa sem tendência para Y se e somente se a constante for igual a 1, em outras palavras, E (Wi) = 1.

Wi = 0 se $U_i \notin S_n$ com probabilidade 1 - (n/N)
Wi = 1 se $U_i \notin S_n$ com probabilidade 1 - (n/N)

E, $V(\hat{Y}) = \sum_{i=1}^{N} y_i^2 V(W_i) + \sum_{i \neq j}^{N} y_i y_j COV(W_i, W_j)$

Como, E (W_i) = Ci [n/N + 0 (1 - (n/N)] = 1

$C_i (n/N) = 1 \rightarrow C_i = (N/n)$

A variância de Wi será:

$$V(W_i) = E(W_i^2) - [E(W_i)]^2$$

$$V(W_i) = (W_i^2) - 1.0$$

$$\text{porque } E(W_i) = 1.0$$

Mas,

$$E(W_i^2) = C_i^2 \text{ (n/N)} + 0^2 \text{ [1 - (n/N)]} = \text{(N/n)}^2 \text{ (n/N)} = \text{N/n}$$

Então,

$$V(W_i^2) = \text{(N/n) - 1} = \text{(N - N)/n}$$

$$COV(W_i, W_j) = COV(W_i, W_j) = E(W_i, W_j) - E(W_i)E(W_j)$$

$$E(W_i)E(W_j) = 1$$

$$E\text{ (Wi, Wj)} = \frac{\binom{N-2}{n-2}}{\binom{N}{n}} \text{ (N/n) (N/n)} = -\text{ [(N-n)/ n (n-1)]}$$

Então,

$$V\left(\hat{Y}\right) = \sum_{i=1}^{N} y_i^2 (N-n)/n - \sum_{i \neq j}^{N} y_i y_j [(N-n)/n(n-1)]$$

A fórmula de V $\left(\hat{Y}\right)$ é ajustável para amostragem sem reposição, e conseqüentemente poderia ser usada para amostragem sistemática, pois depois que uma amostra é mensurada esta não é reposta na população.

Segundo LOETSCH & HALLER (1964), qualquer unidade de amostra deve ser selecionada mutuamente independente e ao acaso, porquer esta é a única maneira que proporciona que todas unidades de amostra da população tenham igual chance de serem selecionadas para mensuração.

Como certos pares de unidades de amostra não têm chance de serem incluídas depois que a primeira unidade de amostra é selecionada na amostragem sistemática, a fórmula para V (Y) não pode ser usada para estimar a variância da população com os dados da amostra.

Entretanto, analisando a variância para amostragem sistemática quando todos os conjuntos de unidades amostrais são considerados, podemos observar que algumas vêzes o uso da amostragem sistemática pode aumentar a precisão da estimativa da variância. Considerando todos os conjuntos de uma amostragem sistemática (completa enumeração), a análise da variância da população pode ser calculada pelo seguinte procedimento.

Por definição a variância de $\bar{y}_{sist.}$ é:

$$V(\bar{y}_{sist}) = (1/K)\sum_{i=1}^{k} (\bar{y}_i - \bar{Y})^2$$

$$V(\bar{y}_{sist}) = (1/K)\sum_{i=1}^{k} \left[(1/n)\sum_{j=1}^{n} Y_{ij} - \bar{Y}\right]^2$$

$$V(\bar{y}_{sist}) = (1/K)(1/n^2)\sum_{j=1}^{k} \left[\sum_{j=1}^{n} (Y_{ij} - \bar{Y})\right]^2$$

$$V(\bar{y}_{sist}) = (1/Kn^2)\sum_{i=1}^{k} \sum_{j=1}^{n} (Y_{ij} - \bar{Y})^2 + 2\sum_{i=1}^{k} \sum_{j<j'}^{n} (Y_{ij} - \bar{Y})(Y_{ij'} - \bar{Y})$$

Para calcular o segundo termo da fórmula, consideremos:

$\rho = E(Y_{ij} - \bar{Y})(Y_{ij'} - \bar{Y}) / (Y_{ij} - \bar{Y})^2$, onde ρ é o coeficiente de correlação intraclasse de qualquer par de unidades pertencentes ao mesmo conjunto de amostras sistemáticas. Quando existe n unidades de amostra no conjunto, existe C_2^n diferentes pares de unidades a serem amostradas. Desde que existem K conjuntos, existirão Kn (n-1)/2 diferentes pares de unidades de amostra.

Então,

$$E\left(Y_{ij}-\overline{Y}\right)\left(Y_{ij'}-\overline{Y}\right)=2/Kn(n-1)\sum_{i=1}^{k}\sum_{j<j'}^{n}\left(Y_{ij}-\overline{Y}\right)\left(Y_{ij'}-\overline{Y}\right)$$

E,

$$E\left(Y_{ij}-\overline{Y}\right)^2=(1/N)\sum_{i=1}^{k}\sum_{j=1}^{n}\left(Y_{ij}-\overline{Y}\right)^2$$

$$E\left(Y_{ij}-\overline{Y}\right)^2=(N-1/N)(1/N-1)\sum_{i=1}^{k}\sum_{j=1}^{n}\left(Y_{ij}-\overline{Y}\right)^2$$

$$E\left(Y_{ij}-\overline{Y}\right)^2=(N-1/N)\,S^2$$

Então,

$$\rho=\left[2/Kn(n-1)\right]\sum_{i=1}^{k}\sum_{j<j'}^{n}\left(Y_{ij}-\overline{Y}\right)\left(Y_{ij'}-\overline{Y}\right)\left[N/(N-1)S^2\right]$$

$$\rho=\left[2/(n-1)\right]\sum_{i=j}^{k}\sum_{j<j'}^{n}\left(Y_{ij}-\overline{Y}\right)\left(Y_{ij'}-\overline{Y}\right)\left[1/(N-1)S^2\right]$$

Mas,

$$\sum_{i=1}^{k}\sum_{j=1}^{n}\left(Y_{ij}-\overline{Y}\right)\left(Y_{ij'}-\overline{Y}\right)=\left[(n-1)/2\right]\left\{\left[S^2(N-1)/1\right]\rho\right\}$$

Então,

$$V\left(\overline{y}_{sist}\right)=(1/K)(1/n^2)\left\{\sum_{i=1}^{k}\sum_{j=1}^{n}\left(Y_{ij}-\overline{Y}\right)^2+2\left[(n-1)/2\right](N-1)S^2\rho\right\}$$

$$V\left(\overline{y}_{sist}\right)=(1/Kn^2)\left[\sum_{i=1}^{k}\sum_{j=1}^{n}\left(Y_{ij}-\overline{Y}\right)^2+(N-1)S^2(n-1)\rho\right]$$

$$V\left(\overline{y}_{sist}\right)=(1/nN)\left[(N-1)S^2+(N-1)S^2(n-1)\rho\right]$$

$$V\left(\overline{y}_{sist}\right)=(1/nN)\left[(N-1)S^2\right]\left[1+(n-1)\rho\right]$$

$$V\left(\overline{y}_{sist}\right)=(S^2/n)(N-1/N)\left[1+(n-1)\rho\right]$$

Isto mostra que $V\left(\overline{y}_{sist.}\right)$ irá aumentar ou diminuir em função de ρ. Um alto e positivo valor de ρ é obtido quando as unidades de amostra são homogeneas na amostragem sistemática e um pequeno ou negativo valor de ρ será obtido quando as unidades de amostra forem heterogêneas na amostragem sistemática. Portanto para se obter o pequeno valor de $V(\bar{y}_{sist.})$, deve-se tentar tornar as unidades de amostra heterogeneas para manter ρ pequeno (YAMANE, 1967). Mas o problema é, que tipo de população irá proporcionar isto?

No geral se pode considerar quatro tipos de populações:

a) população aleatória;

b) população ordenada ou com tendência linear;

c) população autocorrelacionada

d) população periódica;

População aleatória é aquela em que os valores de $y_{i's}$ não são correlacionados e possuem mesmas esperanças matemáticas.

$E(y_i) = \mu$ para i=1, 2, ..., N

$$E(y_i - \mu)^2 = \sigma_i^2$$
$$E(y_i - \mu)(y_j - \mu) = 0$$

Então, neste caso ρ é igual a zero e o termo [1 + (n-1) ρ] será igual a 1. Com isto o valor esperado da variância da amostragem sistemática será igual ao da variância para amostragem inteiramente aleatória e amostragem estratificada.

Na população ordenada onde existe uma tendência linear entre os elementos da população, os valores de $y_{i's}$ não são correlacionados mas as esperanças matemáticas mudam linearmente com i.

$E(y_i) = \alpha + \beta_i$ i=1, 2, ..., n

onde, α e β = coeficientes da regressão.

$$V(y_i) = \sigma_i^2$$
$$COV(y_i, y_j) = 0 \qquad (i \neq j)$$

Segundo SCHREUDER et al. (1993), quando a população é ordenada o uso da amostragem sistemática em uma população produz resultados tão precisos como a amostragem inteiramente aleatória.

Os valores esperados das variâncias para a amostragem sistemáticas, amostragem inteiramente aleatória e estratificada são os seguintes (GAUTSCHI,1957).

$$E(V_{sist}) = (K - 1/nK)\sigma^2 + \beta^2(K^2 - 1/12)$$

$$E(V_{srt}) = (K - 1/nK)\sigma^2 + \beta^2(K^2 - 1/12n)$$

$$E(V_{ia}) = (K - 1/nK)\sigma^2 + \beta^2(K^2 - 1)(nK + 1)/12$$

Isto implica que $E(V_{srt}) < E(V_{sist}) < E(V_{ia})$ com estes valores iguais quando n=1.

Em populações autocorrelacionadas onde duas unidades de amostra são mais semelhantes se estão próximas do que se estão distantes, o valor de ρ irá aumentar ou diminuir em função da distância entre tais unidades de amostra (d). Os valores esperados para os parâmetros da população são:

$$E(y_i) = \mu \qquad \text{para } i = 1, 2, ..., N$$

$$E(y_i - \mu)^2 = \sigma^2$$

$$E(y_i - \mu)(y_{i+d} - \mu) = \rho_\alpha \sigma^2$$

onde $\rho_{d1} \geq \rho_{d2} \geq 0$ para $d_1 < d_2$

COOPER (1961) estudando o efeito do espaçamento no crescimento de árvores individuais em povoamentos com mesma idade, usando o teste de correlação de Spearman para diâmetro e distância de árvores com as árvores vizinhas, em 14 povoamentos de pinus ponderosa, com idades variando de 26 a 80 anos encontrou que somente 2 dos 14 coeficientes foram significativos a nível de 1% de probabilidade e 2 a nível de 5% de probabilidade, sendo que os demais não foram significantes. Ele concluiu que árvores em povoamentos jovens são distribuidas aleatóriamente e que essa distribuição possui uma ligeira tendência para uniforme a medida que o povoamento vai aumentando de idade. Não foi verificada a tendência de agrupamento de árvores com mesmas medidas e a média dos diâmetros em povoamentos homogêneos é altamente determinada pela densidade da população, isto é, número de árvores por unidade de área.

WATSON JR. (1967) estudando o efeito de diâmetro de árvores vizinhas e relação da distância entre essas árvores, em povoamentos de *Pinus elliottii*, encontrou que o diâmetro alcançado por uma árvore é independente dos diâmetros das árvores vizinhas. Uma possível explanação de tal fato é que tais árvores tiveram suficiente espaço para crescimento e a competição entre essas árvores já havia sido minimizada. Com relação a distribuição das distâncias entre classes diamétricas, ele encontrou que a redução da média de quadrados dos erros foi significante, e conseqüentemente a distribuição entre classes diamétricas não é um fenômeno aleatório. Como a distribuição de diâmetros individuais foi aleatória ele concluiu que possivelmente em povoamentos com igual espaçamento, diâmentros de árvores ocorrem aleatoriamente.

Além do problema de que a aleatorização de uma única unidade de amostra não é suficiente para estimativas de variâncias em amostragem sistemática, outro problema existe com respeito a populações periódicas, onde variações periódicas podem estar presentes e, às vêzes, são difícies de prever (PRODAN. 1968).

Para melhor entendimento do seja uma população periódica, considere a representação gráfica da curva senoidal e casualize um ponto sobre a mesma. Não importa onde o ponto fique, se o valor de K for igual a um ou mais períodos da curva, todo valor de y_i será o mesmo e a variância será zero.

Populações periódicas em florestas é um assunto controvertido. MADOW (1946) em um experimento constando de 420 observações em um viveiro florestal, encontrou evidência de periodicidade. O mais controvertido trabalho feito em florestas concernente a variações periódicas, foi feito por FINNEY (1949). Tal trabalho foi realizado na floresta de Dun Dehra no norte da Índia onde foram consideradas 292 linhas de amostragem, sendo que foi constatada periodicidade repetida a cada décima sétima linha. Ele afirma que tal periodicidade foi estudada por conhecedores da região e ninguém pode explicar o fenômeno. Esta periodicidade provocou um decréscimo na eficiência da amostragem sistemática quando comparada com amostragem estratificada, com mesma intensidade amostral. MILNE (1959) fez o seguinte comentário sobre o trabalho de FINNEY (1949) "o volume médio de madeira por unidade de área (acre) para cada linha foi calculado sem considerar as diferenças de comprimento de linhas". Então, depois de completo estudo sobre a região (Milne recebeu mapas e o procedimento de amostragem, enviados por Finney), demonstrou-se que tal periodicidade estava presente devido a erros de medição. Finalmente, ele concluiu que mesmo assumindo que variações periódicas estejam numa floresta, os seguintes fatos devem ocorrer para que a eficiência da amostragem sistemática seja comprometida:

a) O conjunto de unidades de amostra a ser considerado deve estar na mesma direção da variação periódica;

b) O intervalo de amostragem K deve ser o mesmo ou um múltiplo do intervalo de tal variação;

c) As amostras devem coincidir com os altos ou baixos valores da variação;

d) Os períodos de variação devem possuir os mesmos máximos ou mínimos valores, ou alternativamente, todo período deve ser amostrado;

e) Deve existir comparativas variações ao longo de qualquer linha contendo unidades de amostras.

Portanto, como se pode observar, populações periódicas em florestas não são fenômenos comuns, e se existem, certamente serão observadas e as devidas medidas corretivas serão tomadas.

Comparação entre uma e várias unidades aleatórias

Desde que na prática nunca se mensure todos os possíveis conjuntos de unidades amostrais, mas somente um ou alguns deles, os estimadores da variância devem ser calculados em função da amostra.
Como a aleatorização de unicamente uma unidade amostral não proporciona meios de se obter estimadores sem tendências para a variância, a melhor solução é aleatorizar várias unidades amostrais, sendo que cada uma define um conjunto de amostras (SHIUE, 1969).

Na aleatorização de várias unidades de amostras, a distribuição das demais unidades no conjunto é feita de maneira semelhante ao caso em que só uma unidade é aleatória, isto é, devem ficar sistematicamente separadas entre si por uma distância K. Com este procedimento quaisquer pares de unidades de amostra possui uma determinada probabilidade de estarem presente juntos. Conseqüentemente, estimadores sem tendências podem ser derivados. Mas, dependendo da variação da população, diferentes conclusões podem ser obtidas. Por exemplo, GAUTSCHI (1957) encontrou que em populações autocorrelacionadas, amostragem sistemática com uma única unidade foi mais precisa que a amostragem sistemática com várias unidades aleatórias. Esta conclusão foi explicada pelo fato de que a amostragem sistemática com uma unidade casualizada se distribuiu melhor na população que a amostragem sistemática com várias unidades aleatórias, isto com mesma intensidade amostral, além de que no segundo caso onde ocorre unidades mais próximas, a quantidade de informações sobre a população tende a ser menor, pois unidades próximas dão informações mais semelhantes. Como o trabalho de Shiue só foi publicado em 1960, provavelmente GAUTSCHI (1957) não usou fórmulas da amostragem em conglomerado para estimar a variância da amostragem com várias unidades aleatórias. Certamente se Gautschi tivesse simulado um maior número de conjuntos de amostras pelo método de Monte Carlo, também teria obtido diferente conclusão.

Alguns Métodos de Amostragem Sistemática

Existem muitos métodos de aplicação da amostragem sistemática. Neste trabalho todos eles não serão considerados pelo fato de que não é fácil quantificar exatamente quantos existem, e para fazer isto, este trabalho certamente seria extenso. Portanto serão considerados os comumente encontrados nos livros textos de amostragem, e aqueles mais usados no meio florestal.

A maneira mais simples de usar amostragem sistemática com uma unidade aleatória, mesmo considerado que não se pode derivar um estimador sem tendência para a variância, é o uso das fórmulas da amostragem inteiramente aleatória ou amostragem estratificada. OSBORNE (1942) estudando os erros de amostragem em levantamentos com fotografias aéreas, encontrou que o uso das fórmulas da amostragem inteiramente aleatória ou estratificada em amostragem sistemática, superestimaram o erro de amostragem. AVERY (1957) declarou "existem muitas defesas para a amostragem sistemática, infelizmente, o uso de fórmulas da amostragem inteiramente aleatória não é uma delas".

Para uma melhor visualização das unidades de amostra nos métodos , considere-se uma população de tamanho N = 256 e o tamanho da amostra n = 16.

Rede Quadrada ou Amostragem Alinhada

Neste caso a amostra aleatória é escolhida na população e todas as outras a serem amostradas são distribuidas na população obedecendo uma distância K entre si em sentido vertical e horizontal (figuara 1).

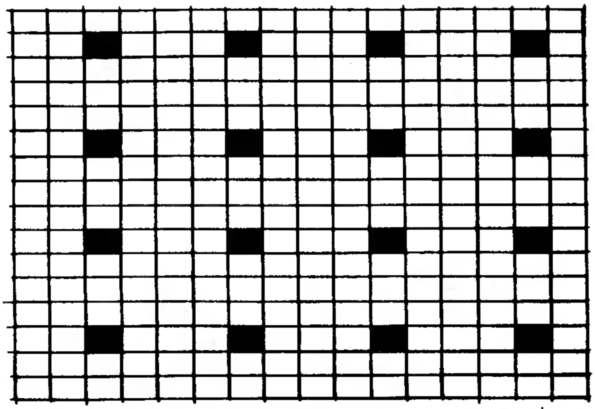

Figura 1 - Rede quadrada (1 dos possíveis 16 similares conjuntos de unidades amostrais)

Amostragem Desigual (não alinhada)

A população é dividida em estratos de tamanho igual a K, isto é, cada lado do estrato é de tamanho K. Dois pontos aleatórios definem as coordenadas da primeira unidade. Três pontos adicionais determinam as coordenadas horizontais das unidades restantes da primeira coluna de estratos (cada uma distante K entre si). Outros pontos aleatórios são necessários para determinar as coordenadas verticais das unidades restantes na primeira linha de estratos. A locação das outras unidades é feita em função da distância K utilizando exatamente o mesmo procedimento para a primeira linha e coluna de estratos, isto em sentido horizontal e vertical (figura 2).

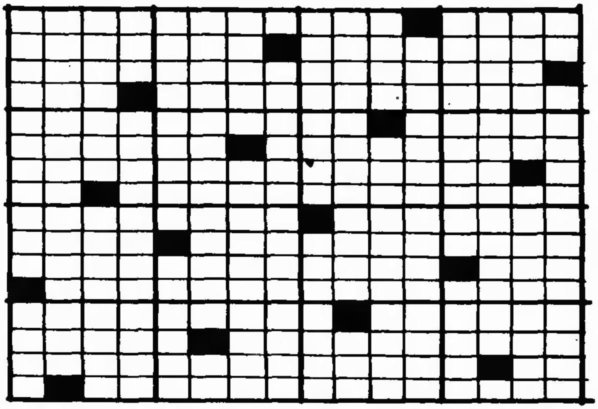

Figura 2 - Amostragem desigual (1 dos possíveis 16 similares conjuntos de unidades amostrais)

COCHRAN (1977) encontrou que este procedimento quando aplicado em populações autocorrelacionadas, é freqüentemente superior a rede quadrada ou amostragem estratificada.

Este método de amostragem sistemática é o que se aproxima mais da amostragem inteiramente aleatória, e isto pode ser uma desvantagem quando os custos de amostragem são diferentes por procedimento, pois o tempo de deslocamento entre duas unidades de amostra é proporcional a distância entre unidades e inversamente proporcional a velocidade de deslocamento entre essas duas unidades (ZEIDE, 1980). Neste caso a distância entre unidades tende a aumentar. MESAVAGE & GROSENBAUCH (1956) estudando a eficiência de vários métodos de amostragem, encontraram que o tempo permitido de deslocamento na amostragem inteiramente aleatória foi aumentado de 120% quando comparado com a rede quadrada. Como a amostragem desigual aumenta o tempo de amostragem, isto produz um decréscimo na eficiência relativa deste método quando os custos são considerados.

REDE RETANGULAR

Neste método as unidades amostrais a serem mensuradas são colocadas em algumas linhas separadas entre si por um múltiplo de K (figura 3).

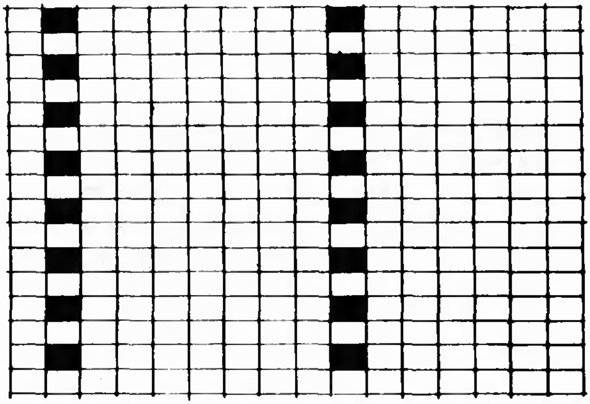

Figura 3 - Rede retangular (1 dos possíveis 16 similares conjuntos de unidades amostrais)

Este é o método que produz menores custos, mas concernente aos estimadores da média e da variância, é claro que a quantidade de informações sobre a população é menor, principalmente se esta for autocorrelacionada.

Os métodos 1,2 e 3, geralmente usam as fórmulas para média e variância proveniente de amostragem inteiramente aleatória ou estratificada.

MÉTODOS DAS PRIMEIRAS DIFERENÇAS

Este método consiste em se considerar as diferenças entre sucessivos pares de unidades amostrais, e proporciona aproximações para a variância da população. Se existe n unidades de amostra enumeradas na amostragem sistemática, existirá (n-1) diferenças. A variância é dada pela soma dos quadrados das diferenças dividida pelo dobro do número de diferenças , e a média é dada pela mesma fórmula aplicada na amostragem inteiramente aleatória.

$$\bar{y} = (1/n)\sum_{i=1}^{n} y_i$$

$$\hat{V}(\bar{y}) = \sum_{i=1}^{n-1} [d(y_i)]^2 / 2n(n-1)[1-(n/N)]$$

Onde, $d(y_i) = y_{i+1} - y_i$ $\qquad i = 1, 2, ..., (n-1)$

Este método é aplicado em amostragem sistemática com uma dimensão (uma linha de unidades de amostras).

Se existem várias linhas paralelas e equidistantes contendo as unidades de amostra (figura 1), o ideal é considerar cada linha como um estrato (FAO, 1982). Se o comprimento das linhas for diferente, a análise será feita proporcional ao tamanho do estrato. A média é dada por,

$$\bar{y} = \sum_{h=1}^{m} (Nh/N)\bar{y}_h$$

onde m = número de linhas.

Nh = número de unidades de amostra por estrato h.

N = número total de unidades de amostra na população.($N = \sum_{h=1}^{m} Nh$)

A variância de cada estrato pode ser estimada pela fórmula da variância por linha, e a variância da média da população será,

$$\hat{V}(\bar{y}_{sist.}) = \sum_{h=1}^{m} (Nh^2/N^2).V(\bar{y}_h)$$

HANSEN et al. (1966), declararam qua só pelo fato do agrupamento de pares de estratos, a estimativa da variância reflete tão bem como com estratificação. Se o agrupamento se estender para todos os possíveis pares de estratos contínuos, a precisão da estimativa será aumentada. A estimativa da variância relativa r = x/y é dada por,

$$\hat{V}_{(r)} = \left[\frac{1-(n-N)}{n}\right]\left[\frac{\sum_{i=1}^{n-1}(x_i - x_{i+1})^2}{2(n-1)\bar{x}^2} + \frac{\sum_{j=1}^{n-1}(y_i - y_{i+1})^2}{2(n-1)\bar{y}^2} - \frac{\sum_{i=1}^{n-1}(x_i - x_{i+1})(y_i - y_{i+1})}{(n-1)\overline{xy}}\right]$$

Para a mesma situação, YATES (1965) sugere o método das diferenças balanceadas do tipo,

$$d = (1/2)x_1 + x_2 + \ldots + (1/2)x_n$$

com estimativa da variância,

$$\hat{V}_{(r)} = \left[\frac{1-(n/N)}{n}\right]\left[\frac{\sum_{i=1}^{t} d_{x_i}^2}{t\sum_{i=1}^{nh} a_i^2} + \frac{\sum_{i=1}^{t} d_{y_i}^2}{t\sum_{i=1}^{nh} a_i^2} - \frac{2\sum_{i=1}^{t} d_{x_i} d_{y_i}}{t\sum_{i=1}^{nh} a_i^2}\right]$$

onde t = número de diferenças balanceadas

$\sum_{i=1}^{nh} a_i^2$ = soma de quadrados dos coeficientes de x_{ij} e $y_{i's}$

No caso de não se querer usar o método das diferenças balanceadas, mas simplesmente considerar estratos individuais, a fórmula para a variância será,

$$\hat{V}(\bar{y}_{sist}) = (N - n/Nn)\frac{\sum_{i=1}^{t} d_y^2}{t\sum_{i=1}^{nh} a_i^2}$$

Esta fórmula pode ser aplicada para uma ou mais limhas de tamanhos iguais ou diferentes. Em ambos casos depois de calculada

V(r), para estratos individuais ou partes de estratos, deve-se substituir os resultados na fórmula,

$$\hat{V}\left(\bar{y}_{sist}\right) = \sum_{i=1}^{t} \left(N_h^2 / N^2\right).v(r_h)$$

MÉTODOS DAS SEGUNDAS DIFERENÇAS

Este método é aplicado para estratos que não se superpõem, cada um com quatro unidades de amostra, sendo que a amostragem sistemática pode ser do tipo retangular ou quadrada. Um bom exemplo é considerar a figura 1, onde cada estrato possui 64 unidades de amostra sendo que em cada um, quatro unidades serão amostradas. Mas, neste método se deve escolher duas unidades em uma linha e outras duas paralelas e equidistantes na próxima linha a ser amostrada (FAO, 1982). Uma estimativa para a variância da média é dada por,

$$\hat{V}\left(\bar{y}_{SIST}\right) = \left[(1 - n/N)/3nn'\right]\left[\sum_{i=1}^{h} y_i^2 - 4\sum_{j=1}^{n} \bar{y}_j^2\right]$$

onde $\bar{y}_j$ = média do estrato j.

n' = número de estratos

Neste caso a variância é considerada nas duas direções.

MÉTODO DE VÁRIOS PONTOS ALEATÓRIOS (AMOSTRAGEM SISTEMÁTICA REPETIDA)

Neste caso a população é agrupada em conglomerados, sendo que cada um dos conglomerados possui um ponto (unidade de amostra) aleatório (SHIUE, 1960). Considerando 4 pontos aleatórios (a, b, c, d), a amostra será constituida dos seguintes números,

Ponto 1: a, a+K, a+2K, ... , a +K(n—1)

Ponto 2: b, b+K, b+2K, ... , b+K(n—1)

Ponto 3: c, c+K, c+2K, ... , c+K(n—1)

Ponto 4: d, d+K, d+2K, ... , d+K(n—1)

O procedimento de amostragem com vários pontos aleatórios pode ser descrito como: primeiro se deve decidir quantos pontos aleatórios serão considerados. Isto irá produzir n repetidos conglomerados com unidades de amostra distanciadas pelo intervalo de amostragem K de outra unidade no mesmo conglomerado (SHEAFFER et alii, 1979). No exemplo de N = 256, teremos n = 4 porque K = 4. A figura 4 mostra um dos 16 possíveis conglomerados independentes (sem superposição).

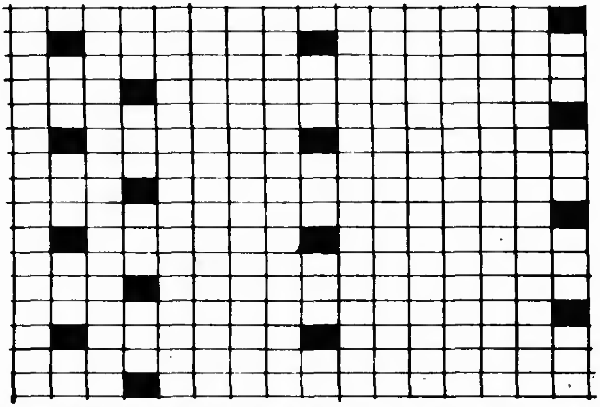

Figura 4 - Amostragem sistemática repetida (1 dos possíveis 16 similares conjuntos amostrais)

Então qualquer par de unidades de amostra possui uma probabilidade positiva de estar presente na amostra, e conseqüentemente estimadores sem tendência para a média e variância podem ser derivados.

Concernente ao tempo total de deslocamento entre unidades, este será igual ao da rede quadrada, porque as diferenças são entre as linhas que geralmente não são separadas por uma distância K.

Portanto os estimadores sem tendência para a média e a variância são os mesmos usados pela amostragem em conglomerados.

$$\bar{y}_{sist.} = \sum_{i=1}^{ns} \left(\bar{y}_i / n_s\right)$$

$$\hat{V}\left(\bar{y}_{sist}\right) = (N - n / N)\left[\sum_{i=1}^{ns} \left(\bar{y} - \bar{y}_{sist}\right)^2 / n_s(n_s - 1)\right]$$

A modificação proposta neste procedimento, e que também pode ser considerada como amostragem sistemática porque as amostras continuam em linhas, é casualizar 4 linhas de 1 a 16 e em cada linha casualizar 4 amostras de 1 a 16. A figura 5 mostra um dos possíveis independentes conglomerados.

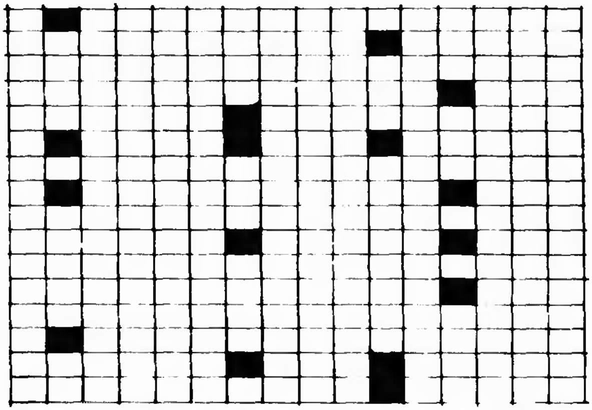

Figura 5 - Amostragem sistemática repetida (1 dos possíveis 16 similares conjuntos amostrais)

Os mesmos estimadores para a média e a variância usados na amostragem em conglomerados também podem ser usados neste caso. Pelo mesmo argumento anterior concernente ao tempo de deslocamento entre unidades, este procedimento proporciona uma redução ou igualdade no tempo total de amostragem.

MÉTODO DE LURY

Neste método a variação entre linhas de unidades de amostra é representada por uma função polinomial das posições das unidades de amostra nas linhas. A estimativa do valor médio por unidade de área é então obtida por uma função polinomial sobre toda a floresta. O grau da função polinomial dever ser suficiente para representar todas as tendências de variação. A variância por linha é estimada da soma de quadrados dos resíduos depois de removidos os componentes de cada polinomio (FINNEY, 1953).

No método de De Lury nenhum elemento aleatório é considerado, e a amostra selecionada em n linhas é unicamente função de regras polinomiais. Então, pode-se concluir que tal método é de difícil computação e de pouca aplicação no meio florestal, principalmente se o número de amostras for elevado.

CONCLUSÃO

O uso da amostragem sistemática em inventários florestais é frequente pela conveniência que este sistema de amostragem oferece. A principal desvantagem é com relação a obtenção de um estimador sem tendência para a variância, que não é possível usando uma única unidade de amostra aleatória. Com o uso de vários pontos aleatórios um estimador sem tendência pode ser obtido utilizando a fórmula da variância da amostragem em conglomerados.

Geralmente os trabalhos sobre amostragem sistemática alertam o "perigo"das populações com variações periódicas. Isto não é um problema em florestas, sendo que o mesmo argumento pode ser estendido para populações ordenadas e autocorrelacionadas. Referente a populações aleatórias, está claro que florestas não podem ser consideradas inteiramente como deste tipo, mas estão próximas, sendo que em plantações com mesmo espaçamento, existe tendência para que haja uma distribuição aleatória (WATSON JR., 1967).

Entre a declaração de AVERY (1975), citada anteriormente, e a de FREESE (1962), "apesar dos riscos, florestais não parecem querer desistir do uso da amostragem sistemática", se é preferível aceitar a segunda que é mais realista, sendo que a palavra riscos não é tão crítica no meio florestal.

Portanto, para segurança e precisão nos inventários florestais usando amostragem sistemática, deve-se utilizar vários pontos aleatórios com as fórmulas da amostragem em conglomerado.

ABSTRACT

Systematic sampling with equal probability is frequently used in timber inventory because of the convenience that this sampling desing offers. The main disadvantage concerns an unbiased estimador for the variance, which is not possible with only one randon starting point. With multiple randon starting points, an estimador for the variance can be obtained with cluster sampling formulae. This paper contains a literature review of the methods frequently used in systematic sampling with equal probality. It also presents a modification, in the case of repeated systematic sampling, which can produce a large number of randon starts for a systematic distribution of plots.

REFERÊNCIAS BIBLIOGRÁFICAS

1 AVERY, T. E. *Natural resources measurements*. New York, McGraw-Hill, 1975, 339 p.

2 CHACKO, V. J. A *manual on sampling for forest surveys*. Dehra Dun, Manager of Publications Survey of India Offices, 1965. 172 p.

3 COCHRAN, W. G. *Sampling techniques*. 3. ed New York, J. Wiley, 1977. 428 p.

4 COOPER, C. F. Pattern in ponderosa pine forests. *Ecology*, Durham, 42(3): 493-99, Jullho, 1961.

5 FAO. *Manual of forest inventory*. Roma, 1982. 200 p. (Fao Forest Paper, 27)

6 FREESE, F. *Elementary forest sampling*. Washington, Forest Service, 1962. 91 p. (Agriculture Handbook, 232).

7 FINNEY, D. J. An example of periodic variation in forest sampling. *Forestry*, London, 23: 96-111, 1949.

8 The estimation of the error in the systematic sampling in forest. *Journal of the indian Society of Agricultural Statistics*, New Delhi, 5:6-16, 1953.

9 GAUTSCHI, W. Some remarks on systematic sampling. *Annals of Mathematical Statistics*, Hayward, 28:385-94, 1957.

10 HANSEN, M. H.; HURWITZ, W. N.; MADOW W. E. *Sample surveys methods and theory*. New York, J. Wiley, 1966. v. 1.

11 LOETSCH, F. & HALLER; K. E. *Forest inventory*. Reinbey, BLV Verlagsgesellschaft Mnuchen Basel Wein, 1964. v. 1.

12 MADOW, L. H. Systematic sampling and its relation to others sampling designs. *Journal of the American Statiscal Association*, Washington, 41:207-14, 1946.

13 MADOW, W. G. & MADOW, L. H. On the theory of systematic sampling I. *Annals of Mathematical Statistics*, Hayward, 15:1-24, 1944.

14 MESAVAGE, C. & GROSENBAUCH, L. R. Efficiency of several cruising desing on small tract in North Arkansas. *Journal Forestry*, Washington, 56:596-76, 1956.

15 MILNE, A. The centric systematic are sample treated as a randon sample. *Biometrics*, Raleigh, 15:270-97, 1959.

16 OSBORNE,J. E. Sampling errors of systematic and randon surveys of covertype areas. *of the American Statistical Association*, Washington, 37:256-64, 1942.

17 PRODAN, M. *Forest biometrics*. Oxford, Pergamon Press, 1968. 447 p.

18 RAJ, Des. *Sampling theory*. New York, Mcgraw-Hill, 1968. 302 p.

19 SHEAFFER, R. L.; MENNHALL, W. ; OTT, L. *Elementary survey sampling*. North Scitvate, Duxbury Press, 1979. 278 p.

20 SCHREUDER, H. T.; GREGOIRE, T. G.; WOOD, G. B. *Sampling methods for mul-* tire sources forest inventory. New York, John Wiley & Sons, Inc. 1993, 445 p.

21 SHIUE, C. J. Systematic sampling with multiple randon starts. *Forest Science*, Washington, 6(1):42-50, 1960.

22 WATSON JR. W. F. *Effect of diameter of neighboring trees and spatial relations on tree diameter in selected slash pine plantations*. Athens, 1967. 29 p. Master of Science — University of Georgia.

23 YAMANE, T. *Elementary sampling theory*. New York, New York University, 1967. 305 p.

24 YATES, F. *Sampling methods for census and surveys*. London, C. Griffin, 1967. 330 p.

25 ZEIDE, B. Plot size optimization. *Forest Science*, Washinngton, 26(2):251-57, 1980.