

---

# **Levantamentos Florestais**

**Conceitos de Amostragem Aplicados  
ao Levantamento de Florestas**

---

**João L. F. Batista**

Departamento de Ciências Florestais

**UNIVERSIDADE DE SÃO PAULO**

Escola Superior de Agricultura “Luiz de Queiroz”

Piracicaba - 1998

---

**Levantamentos Florestais: conceitos de amostragem aplicados  
ao levantamento de florestas**

Copyright © 1998 João L. F. Batista

Registro no Escritório de Direitos Autorais  
Biblioteca Nacional - Ministério da Cultura  
No. Registro: 270.044 Livro: 485 Folha: 204

Departamento de Ciências Florestais  
Escola Superior de Agricultura “Luiz de Queiroz”  
Universidade de São Paulo  
Av. Pádua Dias, 11  
Caixa Postal 9  
13418-970, Piracicaba - SP

Email: [parsival@usp.br](mailto:parsival@usp.br)

---

**Sobre a utilização desta obra didática**

Esta obra didática pode ser obtida no endereço  
<http://lmq.esalq.usp.br/>  
na forma de arquivo PDF.

Ela é distribuída  
com finalidade exclusivamente didática  
para **uso individual e pessoal.**

É proibida a sua distribuição  
ou reprodução, parcial ou integral,  
sem a autorização escrita do autor.

*The shadow of the trees in the misty river  
is dissipated like smoke,  
while in the real branches above  
the turtle-doves lament.*

*How this pale landscape, o traveler  
reflected your pale self,  
and how sad was the lamenting in the lofty foliage  
of your drowned hopes!*

*The Shadow of the Trees,  
Paul Verlaine (1844-1896)*

---

## APRESENTAÇÃO

Estas notas foram inicialmente desenvolvidas para o *Curso III - Inventário Florestal* do II Programa de Reciclagem em Métodos Quantitativos do Departamento de Ciências Florestais, ESALQ, Universidade de São Paulo, lecionado em 1996. No decorrer do tempo elas foram utilizadas, integralmente ou em partes, em vários outros e disciplinas lecionados sofrendo um processo contínuo de aprimoramento.

Na presente forma, estas notas discutem os princípios básicos de amostragem e sua utilização em levantamentos e análise de vegetação nativas ou plantadas. Os delineamentos amostrais principais são apresentados junto com alguns métodos não convencionais de amostragem voltados, via de regra, para florestas. Ao longo das notas, uma série de exercícios são propostos assumindo a utilização de software estatístico moderno e flexível como o Splus.

---

## CAPÍTULO 1

# Conceitos Básicos de Amostragem

---

O estudo botânico ou ecológico de formações vegetais implica na caracterização da vegetação, seja através de variáveis qualitativas, seja por meio de variáveis quantitativas. Enquanto que na caracterização qualitativa pode-se justificar uma abordagem subjetiva para a escolha das áreas da floresta que serão estudadas mais detalhadamente, a caracterização quantitativa só se justifica se a representatividade das áreas estudadas puder ser demonstrada objetivamente. A representatividade de uma amostra da floresta deve se fundamentar nos princípios da amostragem estatística sendo necessário o entendimento dos conceitos básicos de probabilidade e estatística que fundamentam a amostragem quantitativa para se compreender os métodos e técnicas de levantamentos florestais.

### 1.1 Amostragem: Estatística × Não Estatística

O objetivo de todo levantamento de recursos naturais é a coleta de informações que possam ser utilizadas na tomada de decisão, sejam elas visando a caracterização da vegetação, o manejo florestal sustentável ou o diagnóstico dos impactos de atividades antrópicas sobre ecossistemas naturais e semi-naturais. Para que as decisões tomadas sejam apropriadas, as informações obtidas devem ser acuradas e confiáveis. Entretanto, a coleta da informação é sempre realizada sobre limitações de orçamento, tempo e disponibilidade de recursos humanos e existe uma relação direta entre a qualidade da informação (precisão, exatidão, confiabilidade) e a quantidade de recursos necessários para obtê-las. Qualidade da informação e disponibilidade de recursos são os dois elementos que definem a melhor técnica de coleta de informações.

As diferentes técnicas de coleta de informações podem ser agrupadas em categoriais bastante amplas: censo, mapeamento, amostragem não estatística e amostragem estatística.

#### Censo

Por censo devemos entender a enumeração completa da população. No estudo de uma floresta, censo implicaria a observação ou mensuração de todas as árvores. Se o custo de um censo está dentro do orçamento disponível, ele é a técnica mais desejável, não haverá nenhuma incerteza sobre as informações obtidas, uma vez que toda a população foi observada. Dada a magnitude dos recursos naturais e a disponibilidade orçamentária para levantamentos florestais o censo é raramente uma técnica viável.

Entretanto, a prática florestal faz uso do censo em alguns casos. Em inventários pré-corte, é comum a prática de se contar todas as árvores presentes num povoamento a ser cortado, sendo que algumas dessas árvores são medidas. Embora, apenas uma amostra das árvores foi medida, o número de árvores presentes no povoamento é obtido através do censo e, portanto, sem qualquer erro amostral ou incerteza. No estudo de recursos naturais em grandes regiões é comum o uso de imagens de satélites através das quais os tipos de vegetação podem ser identificados. Assumindo que não existe nenhum erro nos métodos de análise de imagens, as informações obtidas dessa forma consistem em um censo, pois, toda a região é estudada. Nesse caso, a subdivisão da região em áreas florestadas e áreas não florestadas, por exemplo, produz uma informação sem qualquer margem de incerteza amostral associada a ela.

Na maioria das circunstâncias, no entanto, o censo pode ser impraticável ou mesmo indesejável. Se o censo fosse utilizada para obter a produção de madeira de uma floresta plantada com 10000 *ha*, demoraria tanto tempo que quando a informação estivesse disponível ela já não seria atual sendo, portanto, inútil. Se por outro lado, o censo fosse executado por um grande número de pessoas para tornar o levantamento mais rápido, seriam cometidos tantos erros de procedimento e de mensuração no campo que a qualidade da informação final seria questionável.

## Mapeamento

Outra forma de obtenção de informações é a elaboração de mapas temáticos. A grande vantagem dos mapas é que além da informação *quanto*, adiciona-se a informação *onde*, permitindo que as dependências espaciais dos dados sejam exploradas. Da mesma forma que o censo, mapas bem feitos são bastante caros, pois se baseiam em informações detalhadas de campo, seja para a construção dos próprios mapas (e.g. levantamentos topográficos), seja para verificação das informações obtidas por sensoriamento remoto (e.g. fotografias aéreas, imagens de satélites e radares).

O sensoriamento remoto possibilita a redução dos custos de trabalhos de campo com um aumento dos custos na obtenção de imagens e processamento dos dados em laboratório. Essas relações nem sempre são vantajosas, informações detalhadas de uma área terão que se basear em fotografias aéreas de pequena escala, as quais tem um alto custo de obtenção. O aumento da escala das fotografias aéreas e o uso de imagens de satélite geram uma redução do custo por unidade de área, mas também resultam em redução do detalhamento da informação e, conseqüentemente, em redução da precisão espacial e temática dos mapas.

Um problema importante mas de difícil solução é a determinação da precisão e confiabilidade de mapas temáticos gerados por diferentes metodologias. Os mapas topográficos (altimétricos e plano-altimétricos) são os únicos que quantificam os erros garantindo, em geral, erros inferiores a 1% na determinação da área.

## Amostragem Não Estatística

A amostragem não estatística se baseia no julgamento subjetivo do técnico ou especialista para a seleção de amostras. O aspecto subjetivo desse tipo de amostragem resulta em dois aspectos negativos. O primeiro é que dois técnicos diferindo em experiência profissional e habilidade técnica podem produzir resultados bastante discrepantes. Em segundo lugar, a precisão e confiabilidade da informação não pode ser determinada. O grande apelo desse tipo de amostragem é o baixo custo.

Essa técnica foi utilizada em inventário florestal em países da Europa Central durante muito tempo, antes do surgimento das técnicas estatísticas de amostragem. Nos séculos 18 e 19, era parte do treinamento do engenheiro florestal a estimativa visual do estoque de crescimento de povoamentos florestais. A necessidade de manejo dos recursos florestais fez com que técnicas de inventário florestal surgissem muito antes do advento da teoria estatística de amostragem. Atualmente, o inventário florestal se baseia totalmente na amostragem estatística, pois, frequentemente, o conhecimento da confiabilidade de uma informação é tão importante quanto a informação em si.

## Amostragem Estatística

A amostragem estatística é uma metodologia mais objetiva de seleção de amostras e assegura a obtenção de uma *amostra representativa* da população estudada. Representatividade da amostra e *confiabilidade* das informações obtidas a partir dela são conceitos que andam juntos. Do ponto de vista estatístico, representatividade significa que cada unidade da população possui uma probabilidade não nula de ser selecionada na amostra e essa probabilidade é independente da pessoa que faz a amostragem. O conceito probabilístico de representatividade permite que a confiabilidade das informações obtidas por amostragem estatística possa ser quantificada. A confiabilidade é vista como o oposto da *incerteza*, medida em termos de probabilidade, que podemos associar com cada amostra dado que somente parte da população foi observada. Custo e confiabilidade da informação são aspectos antagônicos na amostragem estatística. Por exemplo, o custo de se saber qual a disponibilidade de madeira de eucalipto no Estado de São Paulo destinada a produção de celulose com precisão de  $\pm 1\%$  pode ser mais alto que o orçamento que qualquer empresa ou organização estaria individualmente disposta a pagar pela informação. Já uma precisão de  $\pm 10\%$  poderia ter um custo aceitável.

A representatividade da amostragem estatística também implica que algum mecanismo aleatório seja utilizado para selecionar as amostras. A única maneira se assegurar que um mecanismo aleatório seja de fato aleatório, isto é, independente da pessoa que o executa, é utilizar números aleatórios presentes em tabelas ou gerados por computador.

### 1.1.1 Parâmetro e Estimativa

Ao planejar a amostragem estatística é importante ter uma clara visão sobre a população sendo estudada. Do ponto de vista florestal, a população será definida como o objeto de interesse do levantamento, em geral, a floresta como um todo ou uma população vegetal ou animal em particular. Do ponto de vista estatístico, o termo *população* tem uma definição bem mais precisa. Uma *população estatística* é o conjunto de valores, numéricos ou categóricos, que uma determinada *variável pode assumir*. Por exemplo, um florestal deseja estimar a volume atual de madeira numa floresta de *Pinus caribea hondurensis* com 10000 *ha*. Para o florestal a população de interesse são as árvores da floresta, mas na perspectiva da amostragem estatística, a população são os valores de volume (em  $m^3/ha$ , por exemplo) que podemos obter ao medirmos diferentes porções (parcelas) dessa floresta. O conceito estatístico de população exige uma clara definição da variável sendo observada (volume de madeira em  $m^3/ha$  no exemplo acima).

As informações referentes à população são chamadas de *parâmetros*. Considerando diferentes subáreas de 1 *ha* da dita floresta de *Pinus caribea hondurensis*, a variável volume possuirá um valor *médio* e uma *variância*. Estas informações se referem a população como um todo e a única maneira de conhecê-las é medindo toda a floresta. Na prática, as informações referentes à população, são sempre desconhecidas, embora sejam o interesse principal do levantamento. Para representar tais informações utilizamos letras gregas:  $\mu$  para média e  $\sigma^2$  para a variância. É importante compreender que só existe um único valor para  $\mu$  e  $\sigma^2$  para cada população em particular e, portanto, os parâmetros ( $\mu$  e  $\sigma^2$ ) são *constantes*.

Por outro lado, ao fazermos um levantamento, obtemos de cada parcela medida um valor de volume na floresta. O conjunto desses valores constitui a *amostra*, para a qual também poderemos calcular um valor médio e uma variância. Digamos que na floresta de *Pinus caribea hondurensis* foram medidas 100 parcelas. Embora o valor da média ou da variância seja constante para essas 100 parcelas, se uma segunda amostra de 100 parcelas for tomada na mesma floresta esse valor de média ou variância mudará. Isso significa, que estatisticamente média e variância obtidas da amostra não são constantes, e sim uma função da amostra tomada na população. Esses valores de média e variância são, no entanto, “bons palpites” a respeito dos valores da média e variância da população e eles são chamados de *estimativas* dos parâmetros da população. As informações obtidas das amostras são representadas pelas letras gregas com um “chapéu” ( $\hat{\mu}$  e  $\hat{\sigma}^2$ ) ou por variações da letra latina utilizada para representar a variável de interesse: se  $v$  é a variável volume,  $\bar{v}$  é a média e  $s_v^2$  é a variância do volume.

A amostragem estatística nos permite associar uma medida de confiabilidade às estimativas obtidas a partir de uma *amostra*. A maneira de se calcular a estimativa dependerá de como a amostra foi tomada na população. A maneira de calcular é chamada de *estimador*, enquanto que a maneira como as amostras são selecionadas numa população é chamada de *delineamento amostral*.

### 1.1.2 Precisão, Viés e Exatidão

Independentemente da confiabilidade que desejamos numa *estimativa* qualquer, sempre esperamos que o *estimador* que a gerou tenha certas propriedades. Uma propriedade que qualquer pessoa deseja num estimador é a capacidade de produzir valores próximos ao valor do parâmetro da população, sem qualquer tendência a super-estimar ou sub-estimar. Qualquer tendência que distancie um estimador do valor do parâmetro é chamada de *viés* ou vício, e um bom estimador deve ser *não-enviesado* (sem viés) ou não-viciado. Outra propriedade de um bom estimador é a *precisão*. Se diferentes amostras forem tomadas numa mesma população, o estimador mais preciso é aquele cujos valores das estimativas variam menos entre as amostras, isto é, as diferentes amostras produzem estimativas próximas entre si. Portanto, o conceito de precisão é o oposto do conceito de variabilidade ou variância de um estimador, isto é, quanto menor a variância de um estimador, maior a sua precisão.

O estimador ideal é aquele que combina ausência de viés e alta precisão. A figura 1 mostra uma analogia entre tiro ao alvo e os conceitos de viés, precisão e exatidão.

## 1.2 Estatísticas Básicas

Na grande maioria dos levantamentos florestais estaremos trabalhando com a média e a variância como estatísticas básicas, sejam elas da população ou da amostra. Outras medidas obtidas, tais como totais (volume total, por exemplo), serão obtidos a partir da média.

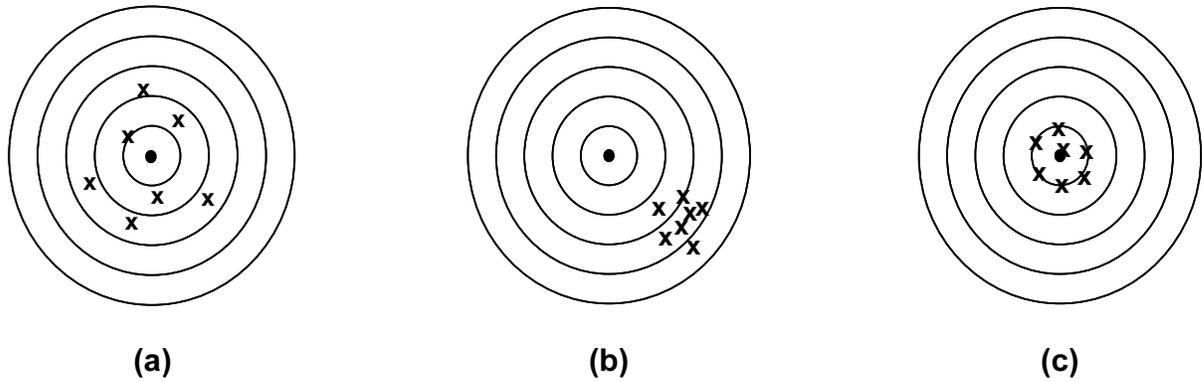


Figura 1: Analogia entre atiradores e os conceitos de viés, precisão e exatidão. O atirador (a) demonstra ausência de viés com baixa precisão, pois apesar da mosca estar no centro da região de dispersão dos tiros, essa região é muito ampla. O atirador (b) possui alta precisão pois a região de dispersão dos tiros é pequena, mas se encontra enviesada em relação à mosca. O estimador ideal é como o atirador (c): possui alta precisão sem qualquer viés.

### 1.2.1 Média e Variância

Digamos que a variável  $X$  é a variável de interesse, por exemplo volume de madeira, altura, DAP, área basal, densidade de árvores, etc.. Em termos de população, a média e a variância são obtidas pelas fórmulas:

$$\mu = \frac{\sum_{i=1}^N X_i}{N} \quad (1)$$

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N} \quad (2)$$

onde  $N$  é o tamanho da população (número total de árvores na floresta, número de parcelas que poderiam ser localizadas na floresta, etc.)

Se uma amostra  $\{x_1, x_2, \dots, x_n\}$  com  $n$  observações for obtida, os estimadores básicos para média e variância são respectivamente:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (3)$$

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \left(1 - \frac{n}{N}\right) \quad (4)$$

O estimador da variância apresentado acima difere ligeiramente da fórmula tradicional para a variância da amostra que a maioria das pessoas estão acostumadas. Na maioria dos casos, as populações estudadas são muito grandes (ou infinitas), de modo que mesmo tomando uma grande amostra a fração  $f = n/N$  permanece um valor muito pequeno, isto é, a parte da população observada é sempre desprezivelmente pequena em relação ao tamanho da população. Em alguns levantamentos florestais, entretanto, a população estudada é pequena e a amostra pode representar uma parte considerável dessa população. Nesses casos, o fator  $f = n/N$ , chamado de *fração amostrada*, não é pequeno ( $f > 0.05$ ) sendo necessário corrigir a variância. O fator de correção:

$$1 - f = 1 - \frac{n}{N} \quad (5)$$

é chamado de *correção para populações finitas* e resulta numa redução da variância tradicional que assume uma população infinita.

Esse conjunto de equações mostra que para uma dada população,  $\mu$  e  $\sigma^2$  (equações 1 e 2) são constantes, pois dependem de todos os valores da população. Já  $\bar{x}$  e  $s_x^2$  (equações 3 e 4) dependem dos valores observados na amostra. Como uma mesma população pode produzir diferentes amostras, a média e variância da amostra são sempre variáveis.

Como a variância é uma soma de desvios em relação à média *elevados ao quadrado*, elas possuem unidade diferente da média. Uma outra medida de variação frequentemente utilizada é o *desvio padrão*, definido

como a raiz quadrada da variância. O desvio padrão mede a variabilidade na mesma unidade da média, isto é, na unidade dos dados originais e por isso é uma medida de interpretação mais direta. Como no caso da variância, existe o desvio padrão da população ( $\sigma$ ), que é um parâmetro, e o desvio padrão da amostra:

$$\sigma = \sqrt{\sigma^2}$$

$$s_x = \sqrt{s_x^2}$$

Tanto a variância quanto o desvio padrão possuem unidades que dependem da variável sendo estudada. O coeficiente de variação ( $CV$ ) é uma medida de variação que independente da unidade de medida utilizada. Se a amostragem de uma variável  $X$ , resultou em média  $\bar{x}$  e desvio padrão  $s_x$ , o  $CV$  é calculado pela fórmula:

$$CV = \frac{s_x}{\bar{x}} 100 \quad (6)$$

Portanto, o  $CV$  mede a variabilidade em termos percentuais em relação ao valor da média.

### 1.2.2 Distribuição Normal

Das inúmeras distribuições estatísticas que existem, a distribuição Normal ou distribuição Gaussiana é central na estatística em geral e na amostragem estatística em particular. A distribuição Normal é definida pela *função de densidade probabilística*:

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/2\sigma^2} \quad \sigma > 0; -\infty < y < \infty;$$

que define a sua forma e, conseqüentemente, a probabilidade de ocorrência de eventos.

A distribuição Normal possui dois parâmetros  $\mu$  (média) e  $\sigma^2$  (variância) e o seu formato, definido pela expressão acima, é próximo ao formato de um sino com a variável  $y$  podendo assumir valores que vão de infinito negativo ( $-\infty$ ) até infinito positivo ( $+\infty$ ). A média é chamada de parâmetro de *localização*, pois controla onde o centro da distribuição fica. A distribuição Normal é perfeitamente *simétrica* em relação à média. A variância é o parâmetro de *dispersão*, pois controla o quanto os valores estão dispersos em relação à média. Quanto menor a variância os valores estarão mais “concentrados” nas proximidades da média e, portanto, menos dispersos. A figura 2 apresenta algumas propriedades da distribuição Normal em termos de probabilidade.

Na distribuição Normal, podemos definir intervalos ao redor da média cujas amplitudes são múltiplos do desvio padrão, isto é, intervalos do tipo

$$\mu \pm k \sigma$$

onde  $k$  assum valores inteiros (1, 2, 3, ...). A proporção das observações que caem dentro destes intervalos é constante e independe dos valores de  $\mu$  e  $\sigma$ . A figura 2 (b) mostra intervalos que incluem 91, 95 e 90% dos valores para uma distribuição com média zero e desvio padrão 1. Essa distribuição é chamada de *Distribuição Normal Padronizada* e a partir dela podemos obter probabilidades para qualquer distribuição através das expressões:

$$Z = \frac{X - \mu_x}{\sigma_x} \Leftrightarrow X = \mu_x + \sigma_x Z \quad (7)$$

onde  $X$  é uma variável Normal com média  $\mu_x$  e desvio padrão  $\sigma_x$  e  $Z$  é a variável Normal Padronizada. Com base nessas relações, é possível obter as probabilidades para qualquer variável Normal a partir da variável Normal Padronizada. A tabela 1 traz uma tabela de probabilidades para a variável Normal Padronizada.

#### EXERCÍCIO 1.1: A Distribuição Normal

Construa um gráfico mostrando a influência da variância sobre a distribuição Normal.

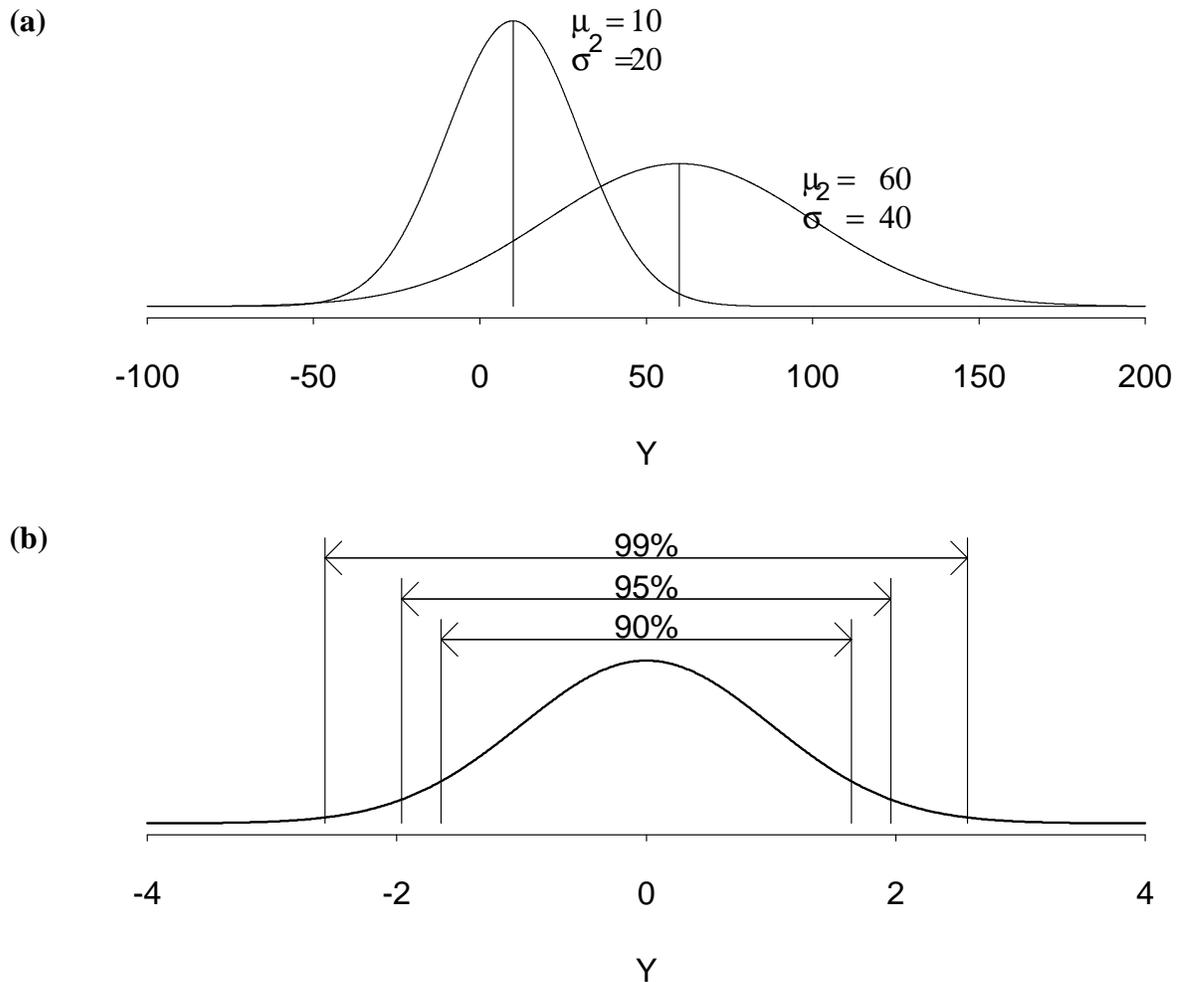


Figura 2: Distribuição Normal: (a) Efeito de média e variância sobre o formato da distribuição Normal. (b) Distribuição Normal Padronizada, isto é, uma distribuição Normal com  $\mu = 0$  e  $\sigma = 1$ , mostrando intervalos ao redor da média que cobrem diferentes proporção dos valores assumidos por uma variável Normal.



EXERCÍCIO 1.2: Estimador com Viés

- Construa um vetor ( $\mathbf{x}$ ) com 100000 valores amostrados de uma distribuição Normal com média zero e variância 100.
- Construa um vetor ( $\mathbf{A}$ ) que identifique 10000 amostras de tamanho 10 a partir dos valores de  $\mathbf{x}$ .
- Para cada amostra estime a variância utilizando o estimador de variância sem viés:

$$\widehat{\sigma^2} = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

- Construa um histograma com as estimativas  $\widehat{\sigma^2}$  obtidas identificando o valor médio das 10000 estimativas.
- Para cada amostra estime a variância utilizando o seguinte estimador *enviesado* da variância:

$$\tilde{\sigma^2} = \frac{\sum (x_i - \bar{x})^2}{n}$$

- Construa um histograma com as estimativas  $\tilde{\sigma^2}$  obtidas identificando o valor médio das 10000 estimativas.
- Compare os histogramas de  $\widehat{\sigma^2}$  e  $\tilde{\sigma^2}$ .

**1.2.3 Teorema Central do Limite**

A importância da distribuição Normal para a amostragem estatística se fundamenta num teorema chamado de Teorema Central do Limite. Em termos simples esse teorema mostra como a distribuição Normal surge quando se faz amostragem de uma variável aleatória  $X$ , a qual pode ter qualquer tipo de distribuição estatística (uniforme, exponencial, binomial, Poisson, etc.). Se  $X$  possui média  $\mu_x$  e variância  $\sigma_x^2$ , podemos obter amostras do tipo  $\{x_1, x_2, \dots, x_n\}$ , cada uma representada por uma série de observações  $x_i$  ( $i = 1, 2, \dots, n$ ) da variável  $X$  obtidas de modo **independente**. Então, à medida que o tamanho da amostra  $n$  cresce ( $n \rightarrow \infty$ ), a distribuição média amostral ( $\bar{x}$ ) se aproxima da distribuição Normal com média  $\mu_x$  e variância  $\sigma_x^2/n$ .

EXERCÍCIO 1.3: Gerando Variáveis com Diferentes Distribuições

Gere 1000 amostras com tamanhos 5, 10 e 30 das seguintes distribuições estatísticas:

- Uniforme entre 0 e 1;
- Exponencial com taxa = 0.1;
- Normal com média 10 e variância 25.

EXERCÍCIO 1.4: Verificação Empírica do Teorema Central do Limite

Utilizando as 1000 amostras geradas no exercício anterior,

- faça um histograma para das amostras geradas por distribuição e tamanho de amostra;
- calcule a média cada amostra gerada;
- faça um histograma das médias para cada distribuição e tamanho de amostra;
- teste se a distribuições são Normais.
- calcule a variância das médias amostrais para cada distribuição e tamanho de amostra.

### 1.2.4 Intervalo de Confiança

O Teorema Central do Limite assegura que para amostras suficientemente grandes a distribuição da média amostral se aproxima da distribuição Normal. Mas o que é uma amostra suficientemente grande? Em geral os estatísticos consideram amostras com  $n > 30$  como sendo grandes. Este número surgiu do trabalho de W.S. Gosset (conhecido pelo pseudônimo “Student”) que no começo do século deduziu uma distribuição estatística que se aproxima da Normal Padronizada à medida que o tamanho da amostra cresce. W.S. Gosset, ou “Student”, mostrou que para amostras aleatórias de tamanho  $n$  da variável  $X$  (média  $\mu_x$  e variância  $\sigma_x^2$ ) a variável

$$t_{n-1} = \frac{\bar{x} - \mu_x}{\sqrt{s_x^2/n}} \quad (8)$$

terá uma distribuição conhecida. Posteriormente, a distribuição desta variável foi chamada de distribuição  $t$  de Student em sua homenagem. Considera-se genericamente que uma amostra de tamanho igual ou maior que 30 é uma “amostra grande”, pois acima deste tamanho de amostra a distribuição  $t$  de Student é muito semelhante à distribuição Normal Padronizada.

O trabalho de Student mostrou que podemos utilizar a distribuição  $t$  para medir a incerteza associada com informações obtidas a partir de amostras estatísticas. Se a equação 8 for re-escrita na forma:

$$\bar{x} - \mu_x = (t_{n-1})s_x/\sqrt{n} \quad (9)$$

teremos uma expressão para a distância entre a média amostral ( $\bar{x}$ ) e a média populacional ( $\mu_x$ ). Esta distância é chamada de **erro amostral**. Logicamente o erro amostral é uma variável aleatória que depende do tamanho da amostra  $n$  e da variabilidade natural da população ( $s_x^2 \rightarrow \sigma_x^2$ , à medida que  $n$  cresce).

Para facilitar o uso dessa expressão definiu-se o *erro padrão da média* como

$$s_{\bar{x}} = \sqrt{\frac{s_x^2}{n}} = \frac{s_x}{\sqrt{n}} \quad (10)$$

Com base na distribuição  $t$ -Student podemos calcular um intervalo que contenha 95% das diferenças entre a média da amostra e a da população:

$$\begin{aligned} P((-t_{0.025;n-1})s_{\bar{x}} \leq \bar{x} - \mu_x \leq (t_{0.975;n-1})s_{\bar{x}}) &= 0.95 \\ P(\bar{x} - (t_{0.025;n-1})s_{\bar{x}} \leq \mu_x \leq \bar{x} + (t_{0.975;n-1})s_{\bar{x}}) &= 0.95 \end{aligned}$$

O intervalo assim construído é chamado de Intervalo de Confiança de 95%, pois se repetíssemos a amostragem inúmeras vezes, em 95% das vezes o intervalo construído incluiria a média da população. A fórmula geral para um intervalo de confiança de  $(1 - \alpha)100$  de confiança é

$$\bar{x} \pm (t_{1-\alpha/2;n-1})s_{\bar{x}} \quad (11)$$

A tabela 2 apresenta o valor de  $t$ -student para diferentes graus de liberdade ( $n - 1$ ) e níveis de confiança de 90, 95 e 91%. Note que o chamamos de **erro amostral** ( $(t_{1-\alpha/2;n-1})s_{\bar{x}}$ ) é a na verdade a amplitude do intervalo de confiança e, portanto, depende do **coeficiente de confiança**  $((1 - \alpha)100\%)$  utilizado.

Tabela 2: Tabela da distribuição de  $t$ -Student para vários graus de liberdade e níveis de confiança de 90, 95 e 99%.

Graus de Liberdade	Confiança			Graus de Liberdade	Confiança		
	90%	95%	99%		90%	95%	99%
1	6.314	12.706	63.657	21	1.721	2.080	2.831
2	2.920	4.303	9.925	22	1.717	2.074	2.819
3	2.353	3.182	5.841	23	1.714	2.069	2.807
4	2.132	2.776	4.604	24	1.711	2.064	2.797
5	2.015	2.571	4.032	25	1.708	2.060	2.787
6	1.943	2.447	3.707	26	1.706	2.056	2.779
7	1.895	2.365	3.499	27	1.703	2.052	2.771
8	1.860	2.306	3.355	28	1.701	2.048	2.763
9	1.833	2.262	3.250	29	1.699	2.045	2.756
10	1.812	2.228	3.169	30	1.697	2.042	2.750
11	1.796	2.201	3.106	40	1.684	2.021	2.704
12	1.782	2.179	3.055	50	1.676	2.009	2.678
13	1.771	2.160	3.012	60	1.671	2.000	2.660
14	1.761	2.145	2.977	120	1.658	1.980	2.617
15	1.753	2.131	2.947	$\infty$	1.645	1.960	2.576
16	1.746	2.120	2.921				
17	1.740	2.110	2.898				
18	1.734	2.101	2.878				
19	1.729	2.093	2.861				
20	1.725	2.086	2.845				

O intervalo de confiança ( e o erro amostral) é a maneira de associarmos uma medida de confiabilidade à informação obtida da amostra (média amostral). Assim podemos construir intervalos de 90%, 99% ou qualquer outro nível de confiança. Por outro, lado se fixarmos o nível de confiança, digamos 95%, nosso interesse será obter intervalos de confiança que pequenos, isto é, pequenos erros amostrais. Como o erro amostral (tamanho do intervalo de confiança) depende do tamanho da amostra ( $n$ ), temos um método objetivo para definir o tamanho de amostra necessário para se produzir um erro amostral arbitrariamente pequeno, dado um determinado nível de confiança.

**EXERCÍCIO 1.5:** Intervalo de Confiança

- Gere 10000 amostras de tamanho 30 da distribuição uniforme.
- Para cada amostra compute intervalos de confiança de 90% e 95%.
- Verifique a proporção de intervalos que contém a média verdadeira.
- A proporção encontrada difere da esperada?

---

## CAPÍTULO 2

# Amostragem Aleatória Simples (AAS)

---

A Amostragem Aleatória Simples é a forma mais simples e direta de se fazer uma amostragem estatística. Na AAS, a seleção das unidades amostrais é realizada de modo completamente aleatório. No caso de levantamentos florestais, isso implica que a localização espacial de cada parcela é completamente aleatória.

### 2.1 Seleção de Parcelas

Como o ser humano é incapaz de produzir padrões espaciais ou de caminhamento que sejam completamente aleatórios, utilizamos números aleatórios gerados em computadores para nos guiar na locação das parcelas no campo. Todo o processo de definição da localização das parcelas deve ser feito no escritório com base no mapa ou croquis da área a ser levantada.

**EXERCÍCIO 2.1:** Locação Aleatória de Parcelas I

Realize a locação de 20 parcelas aleatoriamente nas seguintes áreas:

- um povoamento com forma retangular de  $2000 \times 5000$  m;
- um povoamento com forma circular com 500 m de raio;

**EXERCÍCIO 2.2:** Locação Aleatória de Parcelas II

O arquivo de dados FOREHIPO.DAT contem dados de uma população com 400 unidades amostrais. Simule 100 amostras aleatórias simples de tamanho 30 desta população.

### 2.2 Estimandos os Parâmetros na AAS

Os estimadores utilizados na AAS são estimadores tradicionais de média e variância, isto é:

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad (12)$$

$$s_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} (1-f) \quad (13)$$

$$s_y = \sqrt{s_y^2} \quad (14)$$

$$s_{\bar{y}} = s_y / \sqrt{n} \quad (15)$$

onde  $f = n/N$  é a fração amostrada.

Se estivermos interessados no valor do total, como o volume total de madeira numa floresta, utilizamos a fórmula:

$$\hat{Y} = N \bar{y} \quad (16)$$

e a estimativa do total terá desvio padrão

$$s_{\hat{Y}} = N s_{\bar{y}} \quad (17)$$

O intervalo de confiança também é construído de maneira tradicional:

$$\begin{aligned} \text{Média Amostral:} & \quad \bar{y} \pm (t_{1-\alpha/2;n-1})s_{\bar{y}} \\ \text{Estimativa do Total:} & \quad \hat{Y} \pm (t_{1-\alpha/2;n-1})s_{\hat{Y}} \end{aligned}$$

**EXERCÍCIO 2.3:** Realizando uma AAS

Utilizando as amostras simuladas no exercício anterior:

- calcule as médias e variâncias amostrais de cada amostra simulada;
- analise a distribuição das médias e variâncias amostrais;
- compare as distribuições amostrais com os valores populacionais.

**EXERCÍCIO 2.4:** Intervalo de Confiança

Utilizando as amostras simuladas no exercício anterior, compute o intervalo de confiança de 95% para cada amostra.

**EXERCÍCIO 2.5:** Computando as Estimativas de uma AAS

Utilizando os dados no arquivo EUCALIPT.DAT, estime a média e intervalo de confiança de 95%, por espécie, para: (a) DAP, (b) porcentagem de falha e (c) volume ( $m^3/ha$ ). Ignore o fato de haver variação em termos de rotação e idade.

## 2.3 Tamanho da Amostra

Com base no cálculo do intervalo de confiança podemos obter uma expressão que permite estimar o tamanho da amostra (número de parcelas) necessário para se obter um intervalo de confiança com tamanho e confiança arbitrariamente estabelecidos. Essa expressão é:

$$n^* = \left( \frac{t \quad CV}{E} \right)^2 \quad (18)$$

onde:

$t = t_{1-\alpha/2;n-1}$  é o percentil  $(1 - \alpha)100$  da distribuição  $t$ -Student;

$(1 - \alpha)100\%$  é o coeficiente de confiança desejado;

$CV = (s/\bar{x})100$  é o coeficiente de variação;

$E$  é o erro amostral (percentual) aceitável, isto é, o tamanho aceitável do intervalo de confiança em termos de porcentagem da média.

O coeficiente de variação utilizado na fórmula pode ser proveniente de um levantamento piloto ou de conhecimento prévio sobre o comportamento da variável e floresta sendo levantada. Como o valor de  $t = t_{1-\alpha/2; n-1}$  depende do tamanho da amostra  $n$ , que está sendo estimado, essa fórmula deve ser resolvida por um processo iterativo (vide exemplo 2.1).

Em geral, o coeficiente de confiança de um intervalo é fixado em função do tipo de levantamento florestal, em levantamentos que visam estimar o volume de madeira de florestas, por exemplo, o coeficiente de confiança é de 95%. Já o erro amostral depende mais do tipo de floresta em que o levantamento é realizado, por exemplo, em florestas plantadas de rápido crescimento erros amostrais maiores que 5% não são aceitáveis, enquanto que em florestas nativas erros amostrais de 10%, e até de 20%, podem ser aceitáveis, dada a heterogeneidade da floresta.

**EXEMPLO 2.1:** Computando o Tamanho da Amostra

Suponhamos que um florestal deseja saber quantas unidades amostrais (parcelas) são necessárias para se obter, com 95% de confiança ( $\alpha = 0.05$ ), uma estimativa de produção da floresta ( $st/ha$ ) com um erro amostral de no máximo  $\pm 10\%$  ( $E = 10$ ). Ele acredita que a floresta tenha um  $CV = 25\%$  e estima que um bom número inicial seja 25.

$$n = 25 \Rightarrow t_{0.975;24} = 2.064 \Rightarrow n^* = \left( \frac{2.064 \cdot 25}{10} \right)^2 = 26.6 \approx 27$$

$$n = 27 \Rightarrow t_{0.975;26} = 2.056 \Rightarrow n^* = \left( \frac{2.056 \cdot 25}{10} \right)^2 = 26.4 \approx 27$$

e o processo convergiu para  $n^* = 27$ .

A convergência nesse processo é geralmente bem rápida mesmo que o “chute” inicial seja distante do valor ideal. Suponhamos que o florestal estimasse um número de 10 parcelas:

$$n = 10 \Rightarrow t_{0.975;9} = 2.262 \Rightarrow n^* = \left( \frac{2.262 \cdot 25}{10} \right)^2 = 42.9 \approx 43$$

$$n = 43 \Rightarrow t_{0.975;42} \approx 2.021 \Rightarrow n^* = \left( \frac{2.021 \cdot 25}{10} \right)^2 = 25.5 \approx 26$$

$$n = 26 \Rightarrow t_{0.975;25} = 2.060 \Rightarrow n^* = \left( \frac{2.060 \cdot 25}{10} \right)^2 = 26.5 \approx 27$$

$$n = 27 \Rightarrow t_{0.975;26} = 2.056 \Rightarrow n^* = \left( \frac{2.056 \cdot 25}{10} \right)^2 = 26.4 \approx 27$$

Em apenas quatro iterações se chegou ao resultado ideal, mostrando que a equação converge rapidamente.

Para tamanhos bem grandes de amostra ( $n > 120$ ) a distribuição  $t$ -Student é bem próxima à distribuição Normal e portanto o valor de  $t_{1-\alpha/2; n-1}$  converge para um valor único  $z_{1-\alpha/2}$  independente do tamanho da amostra. Nessas situações a fórmula acima tende a uma constante e, portanto, o tamanho ideal da amostra converge para o valor

$$n^* = \left( \frac{z_{1-\alpha/2} \cdot CV}{E} \right)^2.$$

Para um intervalo de confiança de 95% ( $\alpha = 0.05$ ) esse valor é

$$n^* = \left( \frac{1.960 \cdot CV}{E} \right)^2.$$

**EXEMPLO 2.2:** Computando o Tamanho da Grandes Amostra

Suponhamos que  $CV = 120$  e que o florestal estimasse um número ideal de 30 parcelas:

$$n = 30 \Rightarrow t_{0.975;29} = 2.042 \Rightarrow n^* = \left( \frac{2.041 \cdot 120}{10} \right)^2 = 600.4 \approx 601$$

$$n = 601 \Rightarrow t_{0.975;600} \approx z_{0.975} = 1.960 \Rightarrow n^* = \left( \frac{1.960 \cdot 120}{10} \right)^2 = 553.1 \approx 554$$

Note que o processo iterativo para automaticamente no momento que o tamanho da amostra se torna tão grande que os valores de  $t_{0.975;n-1}$  e  $z_{0.975}$  se tornam aproximadamente iguais.

**EXERCÍCIO 2.6:** Tamanho de Amostra para Eucalipto

Utilizando os dados no arquivo EUCALIPT.DAT, encontre o número ideal de parcela para as variáveis (por espécie) (a) DAP, (b) porcentagem de falha e (c) volume ( $m^3/ha$ ). Considere um intervalo de confiança de 95%, com erro amostral máximo de 5%, e ignore o fato de haver variação em termos de rotação e idade. O tamanho de amostra é o mesmo para as duas espécies e para as três variáveis?

## 2.4 A Estrutura Espacial da Floresta e a AAS

Um ponto de discussão entre pesquisadores é a influência da estrutura espacial da floresta, ou de populações arbóreas na floresta, sobre a “eficiência da amostragem”. A definição do significado de “eficiência da amostragem” é geralmente o aspecto mais polêmico desta questão. Um concepção errônea, mas frequente, é que a aleatorização das unidades amostrais “prejudicaria” a representatividade da amostra, uma vez que a estrutura espacial da floresta não é aleatória. Vejamos, através de exercícios qual o impacto da estrutura da floresta sobre os resultados da AAS.

**EXERCÍCIO 2.7: *Estrutura da Floresta e AAS***

Neste exercício utilizaremos os dados de quatro arquivos:

- REAL.DAT: dados de uma floresta hipotética, mas com estrutura realista. Ao todo são 400 unidades amostrais despostas numa floresta quadrada (20 por 20 unidades amostrais).
- ALEAT.DAT: os mesmos valores da floresta hipotética num padrao espacial totalmente aleatório.
- AGRUP.DAT: os mesmos valores da floresta hipotética, mas os valores foram agrupados para gerar agrupamentos de valores bem distintos sistematicamente organizados na floresta.
- PERIOD.DAT: novamente os mesmos valores da floresta hipotética, neste caso os valores foram agrupados para gerar um padrão espacial *periódico* de altos e baixos.

Para cada tipo de floresta (arquivo):

- a. Construa um gráfico mostrando a estrutura espacial.
- b. Simule 1000 AAS de tamanho 30 e compute a média e variância das amostras.
- c. Compare as distribuições da média e variância obtidas para os diferentes tipo de floresta.

---

## CAPÍTULO 3

# Amostragem Sistemática Simples (ASS)

---

Apesar da simplicidade da amostragem aleatória simples (AAS) e da teoria estatística demonstrar a representatividade das informações obtidas através dela, os práticos florestais sempre sentiram que a AAS é limitada quando se trata de captar a variabilidade espacial que ocorre nas florestas. Os detalhes dessa variabilidade espacial são, em geral, desconhecidos antes do levantamento sendo difícil incorporá-los no delineamento amostral. Outro aspecto incômodo na AAS é o elevado custo de localização e deslocamento de uma parcela a outra, dentro de grandes áreas florestais.

Desta forma, os levantamentos florestais sempre tiveram uma tendência a distribuir sistematicamente as parcelas sobre a área a ser estudada. O levantamento sistemático sempre teve um apelo de melhor cobertura espacial e, portanto, melhor “representatividade” da floresta. A “representatividade” entendida pelos florestais é, em geral, um conceito bem mais vago do que aquele definido pela teoria de amostragem estatística.

Entretanto, as vantagens da amostragem sistemática, em termos de custo e facilidade de execução, são inquestionáveis quando comparadas à AAS. A locação sistemática, no entanto, pode produzir forte viés se for definida unicamente por critérios subjetivos quando do levantamento de campo. Variações sistemáticas que existam na floresta produzirão tendenciosidades nas estimativas destituindo o processo de qualquer sentido de representatividade. Para evitar esse problema, a Amostragem Sistemática Simples (ASS) aleatoriza o ponto inicial a partir do qual todas as unidades amostrais (parcelas) são locadas sistematicamente. Apesar do início aleatório reduzir o risco de viés, ele não o elimina totalmente. A prática florestal, porém, tem mostrado que para os levantamentos e inventários florestais tradicionais a amostragem sistemática produz resultados confiáveis.

### 3.1 Amostragem por Conglomerados

Antes de discutirmos a ASS, é importante estudarmos a amostragem por conglomerados pois podemos definir a amostragem sistemática como um caso particular desta. Na amostragem por conglomerados, as unidades amostrais (parcelas) são na verdade conjuntos sub-unidades (*elementos*), chamados de *conglomerados*, sendo que as observações são realizadas e anotadas para cada elemento do conglomerado. A figura 3 mostra alguns exemplos de conglomerados utilizadas em levantamentos florestais.

Numa amostragem em conglomerados, são levantadas vários conglomerados (unidades amostrais), cada uma delas com um certo número de elementos. Além da variabilidade entre conglomerados, estima-se também a variabilidade interna dos conglomerados. Esse tipo de levantamento é útil quando o fenômeno estudado possui uma alta variabilidade numa escala espacial pequena (dentro do conglomerado) e uma baixa variabilidade numa escala espacial grande (entre conglomerados). Nesses casos, o levantamento em conglomerados evita que a alta variabilidade que ocorre em pequenas escalas influencie a estimativa média ao nível da floresta como um todo (escala espacial grande). O exemplo clássico desse tipo de fenômeno é a regeneração natural em florestas nativas. A ocorrência de uma variação de micro-habitat no subbosque em florestas nativas ou plantadas e a ocorrência de “manchas” tipos de solo são outras situações onde a amostragem por conglomerados também pode ser apropriada. A variação na estrutura da floresta de fragmentos florestais bastante degradados ocorre, em geral, numa escala espacial em que o uso de parcelas em faixa (transectos), subdivididas em subparcelas, torna o trabalho de campo mais prático e as estimativas por amostragem em conglomerados mais precisa.

#### 3.1.1 Estimativas para Conglomerados de Mesmo Tamanho

Numa amostragem em conglomerados, a variável observada deve ser representada por dois índices:

$y_{ij}$  onde:  $i = 1, 2, \dots, n$  representam os conglomerados e

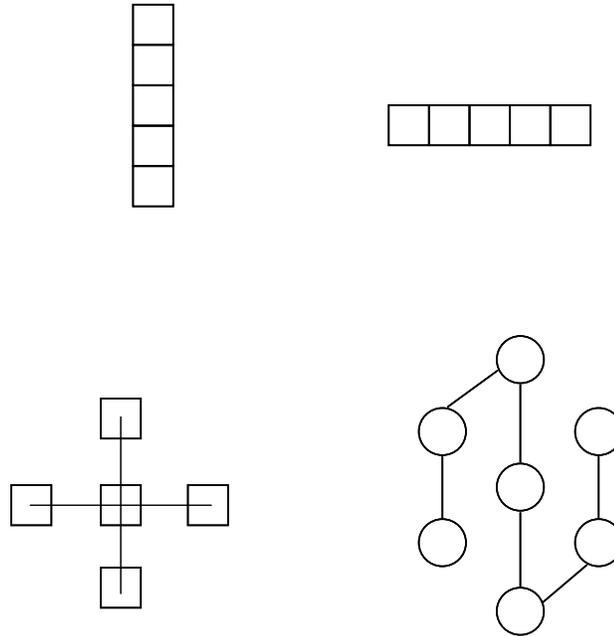


Figura 3: Exemplos de conglomerados utilizados em levantamentos florestais

$j = 1, 2, \dots, m$  representam os elementos dentro de cada conglomerado.

Note que neste caso o número de elementos por conglomerado ( $m$ ) é o mesmo.

A média da população será estimada por

$$\bar{y} = \left(\frac{1}{n}\right) \sum_{i=1}^n y_{i.} \tag{19}$$

onde  $y_{i.} = \sum_{j=1}^m y_{ij}$  é o total de cada conglomerado. (20)

A média ao nível de conglomerado é estimada por:

$$\bar{y}^* = \frac{\bar{y}}{m} \tag{21}$$

A variância entre conglomerados é estimada por

$$s_E^2 = (1 - f) \left(\frac{1}{n - 1}\right) \sum_{i=1}^n (y_{i.} - \bar{y})^2 \tag{22}$$

O erro padrão da média segue o estimador tradicional

$$s_{\bar{y}} = \sqrt{s_E^2/n},$$

enquanto que um erro padrão para média dos conglomerados também pode ser estimado:

$$s_{\bar{y}^*} = \sqrt{\frac{s_E^2}{m^2 n}} \tag{23}$$

A construção do Intervalo de Confiança se dá da maneira usual.

**EXERCÍCIO 3.1: Amostragem por Conglomerados**

Utilizando os dados do exercício 2.7, compare a amostragem por conglomerados com a AAS, sendo o número de parcelas da AAS igual o número total de elementos na amostragem por conglomerados. Para facilitar os cálculos, são funções fornecidas as seguintes funções do SPLUS:

- `Sample.cluster`: gera amostra por conglomerados de diferentes tamanhos de conglomerados a partir dos dados das florestas;
- `Sample.cluster2`: faz os cálculos para as amostras geradas por `Sample.cluster2`.

**EXERCÍCIO 3.2: Amostragem de Regeneração**

O arquivo IBICATU2.DAT apresenta os dados de regeneração de guarantã na Reserva Estadual de Ibicatu. Cada observação refere-se ao número de plantas numa área de  $5 \times 5m$  (subparcela), segundo os estágio

- "M2" plantas com altura entre 50cm e 2m;
- "E" plantas com altura superior a 2m e DAP inferior a 5cm.

As parcelas foram formadas locando os elementos de modo a formar uma faixa com 5m de largura e comprimento de 100m, totalizando 20 elementos por conglomerado.

**3.1.2 Estimativas para Conglomerados de Tamanhos Diferentes**

Quando os conglomerados possuem um número variável de elementos, o que é comum nos levantamentos florestais, as estimativas são obtidas por estimadores ligeiramente diferentes:

Tamanho Médio de Conglomerado:	$\bar{M} = M/N$ onde: $M$ é o número de elementos na população; $N$ é o número de conglomerados na população.
Tamanho Médio de Conglomerado Estimado:	$\bar{m} = \sum_{i=1}^n m_i / n$ utilizado quando $\bar{M}$ não é conhecido.
Média populacional:	$\bar{y} = (1/n) \sum_{i=1}^n y_i$ onde $y_i = \sum_{j=1}^{m_i} y_{ij}$ é total por conglomerado.
Média dos Conglomerados:	$\bar{y}^* = \sum_{i=1}^n y_i / \sum_{i=1}^n m_i$
Variância entre Conglomerados:	$s_E^2 = (\sum_{i=1}^n y_i^2 - \sum_{i=1}^n y_i \cdot m_i + \bar{y}^{*2} \sum_{i=1}^n m_i^2) / (n - 1)$
Variância da média dos Conglomerados:	$s_{\bar{y}^*}^2 = (s_E^2 / (n \bar{M}^2)) (1 - f)$

**EXERCÍCIO 3.3: Levantamento em Fragmentos**

Os dados do arquivo FRAGMEN.DAT se referem a transectos (parcelas em faixa com 10m de largura) localados num fragmento florestal na região de Piracicaba. A variável `picada` identifica o transecto, enquanto que a variável `dborda.m` registra a distância da borda do fragmento em múltiplos de 10m, identificando as subparcelas de  $10 \times 10m$ .

Analisando as variáveis

- área basal (`g.year1.m2`) e
- mortalidade (`n.mort`)

compare os resultados considerando que o delineamento amostral foi AAS e as parcelas foram localadas como:

- a. cada transecto como uma parcela independente;
- b. cada subparcela como uma parcela independente;
- c. cada subparcela como um elemento dentro de um conglomerado (transecto).

**EXERCÍCIO 3.4: Levantamento em Floresta Pluvial**

Os dados de um levantamento em floresta pluvial no Maranhão são apresentados no arquivo MARAN4.DAT. O levantamento foi realizado por parcelas em faixa com 10m de largura e comprimento variável, gerando parcelas de  $5000m^2$  e  $2500m^2$ . Analisando as variáveis

- `dap` (`dap.cm`),
- área basal (em  $m^2/ha$ ) e
- densidade (em  $arv/ha$ ),

Encontre as estimativas supondo uma AAS tomando

- a. cada parcela como uma parcela simples;
- b. cada parcela como um conglomerado.

### 3.2 Amostragem Sistemática Simples

A amostragem sistemática pode ser vista como uma amostragem em conglomerados onde *apenas um conglomerado* foi amostrado. Dessa forma a variância obtida é na verdade a variância interna do conglomerado. O que difere na amostragem sistemática é que os elementos do conglomerado não estão próximos umas dos outros, mas dispersos sistematicamente por toda floresta. A figura 4 mostra como essa analogia funciona.

Na prática as estimativas da amostragem sistemática são obtidas como se a amostragem fosse AAS (equações 12, 13 e 15). O intervalo de confiança e o tamanho ideal da amostra também seguem os procedimentos da amostragem aleatória simples. Surge, entretanto, um problema. Como toda a AAS é considerada um único conglomerado, o estimador da variância da AAS (fórmula 13) é na verdade um estimador da variância *dentro do conglomerado* sendo um estimador *viciado* da variância populacional. Para a maioria das variáveis analisadas em levantamentos florestais, no entanto, a prática mostrou que este estimador é conservador, isto é, ele *superestima* a variância da população.

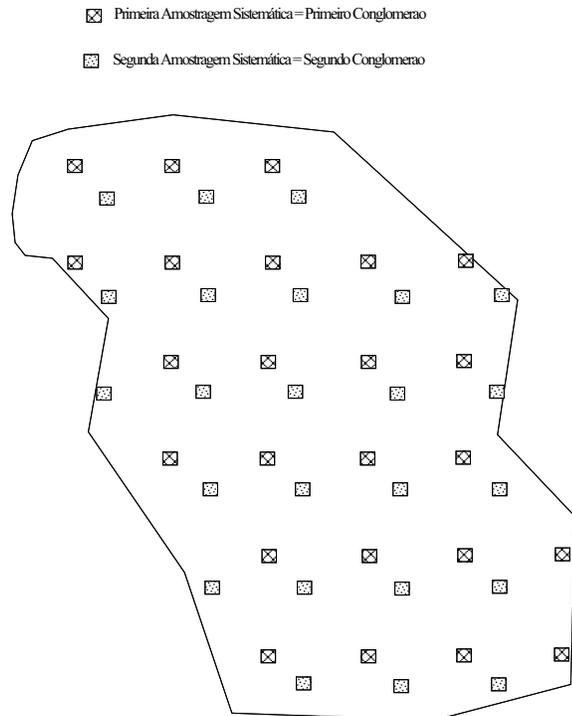


Figura 4: Relação entre amostragem por conglomerados e amostragem sistemática. Cada levantamento sistemático corresponde a um conglomerado cujas sub-unidades se distribuem sistematicamente por toda a floresta.

EXERCÍCIO 3.5: A Estrutura da Floresta e a ASS

Utilizando os dados das quatro florestas hipotéticas (arquivos REAL.DAT, ALEAT.DAT, AGRUP.DAT e PERIOD.DAT do exercício 2.7) compare as estimativas da média e variância obtidas pela ASS com as obtidas pela AAS com base em 1000 simulações. Utilize a função `Sample.system` do SPLUS para realizar a amostragem sistemática (esta função fixa o tamanho da amostra em 25).

EXERCÍCIO 3.6: A Estrutura da Floresta e a ASS - II

Utilizando os dados do exercício anterior, construa uma série de 1000 intervalos de confiança de 95% com base na ASS para cada floresta. O coeficiente de confiança é de fato alcançado em todas os tipos de floresta ?

---

## CAPÍTULO 4

# Tamanho de Parcela

---

Nos levantamentos florestais as unidades amostrais raramente são as árvores, pois isto implicaria na seleção aleatória destas o que é impraticável. Em geral, as unidades amostrais são constituídas de “parcelas”, ou pequenas áreas dentro da qual todas as árvores, arbustos, plântulas, etc., são enumeradas e medidas. Como vimos no capítulo anterior, uma unidade amostral pode ser um conglomerado composto de vários elementos. A parcela, concebida como um conglomerado, é formada de várias “subparcelas” e a questão de quantas subparcelas utilizar tem sido chamada como a definição do “tamanho ótimo” de parcela. Tamanho de parcela e número de parcelas (tamanho de amostra), não podem ser consideradas separadamente no planejamento de um levantamento florestal. Infelizmente, nenhuma solução definitiva existe para o problema, mas veremos algumas abordagens utilizadas para resolver essa questão.

### 4.1 Método do Coeficiente de Variação

O método do coeficiente de variação consiste em escolher primeiro o tamanho da parcela e depois o número ideal delas para se obter o erro amostral desejado. Em geral o  $CV$  é estimado medindo-se parcelas de tamanhos crescentes e calculando o  $CV$  entre elas para os diferentes tamanhos. A tendência é que o  $CV$  decresça até estabilizar, e o tamanho “ótimo” de parcela é o menor tamanho cujo o  $CV$  não difere “muito” do  $CV$  estabilizado. As aspas da frase anterior sinalizam que a decisão final é bastante subjetiva.

Uma vez obtido o valor do  $CV$ , obtém-se o número ideal de parcelas (tamanho da amostra) pela fórmula 18. Como o número ideal de parcelas é uma função do  $CV$ , podemos re-escrever essa fórmula da seguinte forma:

$$n_i = \left( \frac{CV_i \cdot t}{E} \right)^2 \quad (24)$$

Logo, o número de parcelas é função do  $CV$ , diminuindo-se o  $CV$  (aumentando o tamanho da parcela) reduz-se o número necessário de parcelas.

#### EXERCÍCIO 4.1: Tamanho de Parcela pelo CV

Utilizando a função `CV.calc` construa os seguintes gráficos:

- $CV$  em função do tamanho de parcela. Construa o gráfico com base em **várias simulações** e encontre uma curva de tendência.
- Tamanho de amostra ( $n_i$ ) (para cada valor de  $CV$  obtido) em função do tamanho de parcela. Considere aceitável o erro amostral de no máximo 10% com 95% de confiança.
- O **número total de subparcelas** (para cada valor  $n_i$  obtido) em função do tamanho de parcela. Número total de subparcelas =  $n_i \times$  (tamanho de parcela).

Qual a conclusão a que se pode chegar analisando os gráficos construídos?

EXERCÍCIO 4.2: Estrutura Espacial e Tamanho de Parcela

Utilizando a função `CV.calc2` analise o efeito do tamanho de parcela no levantamento das quatro floresta hipotética do exercício 2.7.

## 4.2 Relação CV e Tamanho de Parcela

A relação entre  $CV$  e tamanho de parcela tende a possuir um padrão básico. Vários pesquisadores tentaram captar essa relação por fórmulas aproximadas. A fórmula mais utilizada, é atribuída a Freese e estabelece que:

$$CV_i^2 = CV^2 \sqrt{\frac{P}{P_i}} \quad (25)$$

onde:

$CV$  é o o coeficiente de variação para parcelas de tamanho  $P$ , ambos conhecidos;

$CV_i$  é uma aproximação do coeficiente de variação para parcelas *hipotéticas* de tamanho  $P_i$ .

Através dessa fórmula é possível estimar o tamanho ideal de parcela com base em informações de um inventário piloto com parcelas de tamanho  $P$ . É importante ressaltar que essa fórmula é apenas uma aproximação e o seu comportamento em condições de florestas brasileiras ainda não é bem conhecido.

EXERCÍCIO 4.3: Tamanho de Parcela pela Fórmula de Freese

Utilizando os dados do levantamento florestal no Maranhão (arquivo `MARAN4.DAT`), compare as curvas de  $CV$  em função de tamanho de parcela utilizando a Fórmula de Freese quando esta curva é construída com base em:

- parcelas de  $5000m^2$ ;
- parcelas de  $2500m^2$ .

Compare as curvas de  $CV$  com base nas variáveis DAP médio ( $cm$ ) e área basal ( $m^2/ha$ ).

## 4.3 Método do Coeficiente Correlação Intraclasse

Os métodos baseados no  $CV$  permitem apenas definir o tamanho de uma parcela para um dado número ideal de parcelas, ou vice-versa. Eles não permitem estudar o tamanho de parcela e número de parcelas controlando essas variáveis simultaneamente.

O método do coeficiente de correlação intraclasse controla tamanho e número de parcelas com base no estudo da variação dentro das parcelas e entre as parcelas. Para utiliza-lo é necessário um levantamento onde cada parcela foi tratada como um conglomerado, isto é, os dados de campo foram anotados por **subparcela** (elemento) dentro da parcela (conglomerado). Assim, diferentes tamanhos de parcela (diferentes números de subparcelas) podem ser analisados.

Com base nesses dados se faz uma Análise de Variância com o seguinte modelo linear:

$$y_{ij} = \mu + P_i + \varepsilon_{ij}$$

onde

$y_{ij}$  é o dado da parcela  $i$ , subparcela  $j$ ;

$\mu$  é a média geral do levantamento;

$P_i$  é o efeito da parcela  $i$  sobre a média geral;

$\varepsilon_{ij}$  é o erro experimental, sendo a variação entre subparcelas dentro das parcelas.

O quadro de análise de variância para um experimento com  $r$  parcelas, cada uma com  $k$  subparcelas, fica:

Causa de Variação	Graus de Liberdade	Quadrado Médio	Esperança do QM
Varição entre Parcela	$r-1$	$s_E^2$	$\sigma^2(1 + (k-1)\rho)$
Varição dentro das Parcelas (Erro)	$r(k-1)$	$s_D^2$	$\sigma^2(1 - \rho)$

Enquanto  $\sigma^2$  é a **variância populacional** que desejamos estimar e que influencia diretamente o erro amostral, o coeficiente de correlação intraclasse ( $\rho$ ) indica o grau de autocorrelação espacial entre as subparcelas dentro de parcela.

A partir dessa tabela o coeficiente de correlação intraclasse pode ser estimado por

$$\hat{\rho} = \frac{s_E^2 - s_D^2}{s_E^2 + (k-1)s_D^2} \quad (26)$$

e a variância populacional por

$$\hat{\sigma}^2 = s^2 = \frac{s_D^2}{1 - \hat{\rho}}$$

Com base nessas estimativas o erro padrão da média fica:

$$s_{\bar{x}} = \sqrt{\frac{s^2}{kr} [1 + (k-1)\hat{\rho}]} \quad (27)$$

e podemos vislumbrar as seguintes situações em termos do valor do coeficiente de correlação intraclasse:

- Se  $\rho = 1$  então

$$s_{\bar{x}} = \sqrt{\frac{s^2}{r}} = \text{constante}$$

isto é, a variância entre subparcelas é nula e a variância entre parcelas é a melhor estimativa da variância populacional, **independentemente do tamanho da parcela**.

- Se  $\rho < 1$  então  $s_{\bar{x}}$  é uma **função decrescente** de  $k$ , isto é, quanto maior a parcela menor o erro amostral. Quanto mais distante de 1  $\rho$  estiver, maior a queda no erro padrão da média ( $s_{\bar{x}}$ ) em função do tamanho da parcela ( $k$ ).

A figura 5 mostra graficamente como o valor do coeficiente de correlação intraclasse influencia o erro padrão da média.

Para valores altos de  $\rho$ , o efeito do tamanho de parcela pode ser considerado desprezível quando comparado ao efeito do número de parcelas (figura 6). Já para valores pequenos de  $\rho$ , o tamanho de parcela pode ter influência sobre a precisão do levantamento comparável o número de parcelas.

Um aspecto importante é que a estimativa de  $\rho$  pode assumir valores negativos, mas sempre terá um limite inferior que depende o tamanho de parcela do levantamento utilizado. Assim a estimativa de  $\rho$  estará sempre no intervalo

$$-1/(k-1) \leq \hat{\rho} \leq 1.$$

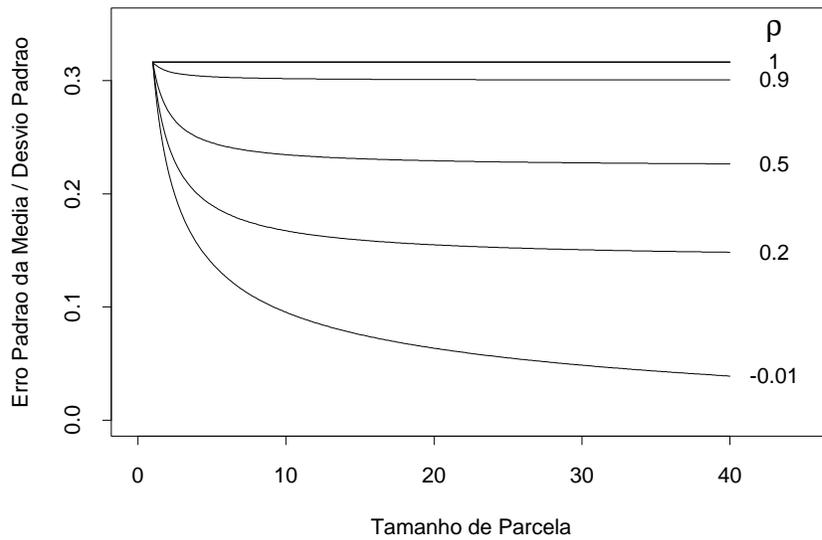


Figura 5: Efeito do coeficiente de correlação intraclasse ( $\rho$ ) e do tamanho de parcela sobre o erro padrão da média. Como o erro padrão da média também depende da variância populacional o gráfico mostra a razão do erro padrão da média pelo coeficiente de variação, mostrando uma tendência válida para qualquer população, independentemente da variância populacional.

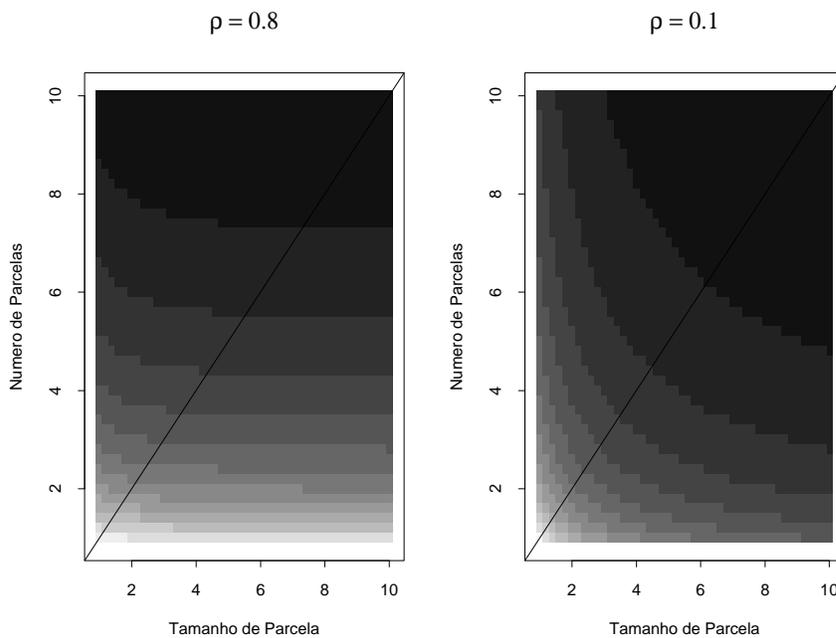


Figura 6: Influencia relativa do tamanho de parcela e número de parcelas sobre a precisão de um levantamento florestal para dois valores de  $\rho$ . A escala de cinza representa a razão do erro padrão da média pelo coeficiente de variação. Tons mais escuros implicam em valores mais elevados da razão (menor precisão no levantamento).

O limite superior é o limite correto para um coeficiente de correlação, mas o limite inferior depende diretamente do tamanho de parcela utilizado para estimar  $\rho$ . Quanto maior o número de subparcelas utilizados na estimativa, tanto mais próximo de zero estará o limite inferior, impedindo que valores negativos do coeficiente de correlação intraclasse possam ser adequadamente estimados por este método.

O tamanho de subparcela também influencia a estimativa do coeficiente de correlação intraclasse, pois o método se baseia na variação entre subparcelas. Para a maioria das variáveis medidas em levantamentos florestais, a variância entre unidades amostrais decresce com o tamanho destas. Escolha de subparcelas pequenas tenderão a gerar maior variância dentro da parcela do que subparcelas grandes, afetando a estimativa do coeficiente.

EXERCÍCIO 4.4: Estimando o Coeficiente de Correlação Intraclasse

Utilizando os dados do exercício 2.7, construa gráficos da distribuição da estimativa de  $\hat{\rho}$  para para os quatro tipos de floresta. Utilize a função `Estima.rho`.

EXERCÍCIO 4.5: Estimando o Coeficiente de Correlação Intraclasse - II

Utilizando os dados do arquivo `FRAGEMEN.DAT`, estime  $\rho$  para as variáveis quantitativas:

- `g.year1.m2` × 100 = área basal ( $m^2/ha$ ) total da floresta;
- `g.growth.m2` × 100 = crescimento em área basal ( $m^2/ha$ );
- `g.ingro.m2` × 100 = área basal ( $m^2/ha$ ) das árvores ingressantes;
- `n.live` × 100 = número de árvores ( $arv/ha$ ) sobreviventes;
- `n.mort` × 100 = número de árvores ( $arv/ha$ ) mortas;
- `n.ingro` × 100 = número de árvores ( $arv/ha$ ) ingressantes.

#### 4.4 Custo de Medição e Tamanho de Parcela

Para maioria das variáveis medidas em levantamentos florestais espera-se que os valores do coeficiente de correlação intraclasse sejam positivos ( $s_E^2 > s_D^2$ ), pois tamanhos crescentes de parcela sempre geram uma redução no erro amostral da média. A questão do tamanho de parcela mais apropriado passa a envolver os custos do levantamento, pois as parcelas maiores também implicam em maiores custos de medição.

Em geral, os custos variáveis num levantamento florestal estão diretamente ligados ao **tempo** necessário para realizar o levantamento. Podemos, portanto, classificar os custos em:

- $c_1$  - custo de deslocamento de uma parcela a outra,
- $c_2$  - custo de medição de uma subparcelas.

O custo total de um levantamento com  $r$  parcelas de tamanho  $k$  cada uma fica

$$T = r(c_1) + kr(c_2) \quad (28)$$

Para encontramos um tamanho apropriado de parcela devemos considerar duas funções:

$$\begin{aligned} \text{Erro Padrão: } s_{\bar{x}} &= \sqrt{\frac{s^2}{kr} [1 + (k-1)\hat{\rho}]} \\ \text{Custo Total: } T &= r(c_1) + kr(c_2) \end{aligned}$$

Com estas funções temos duas alternativas para encontrar o tamanho ideal de parcela:

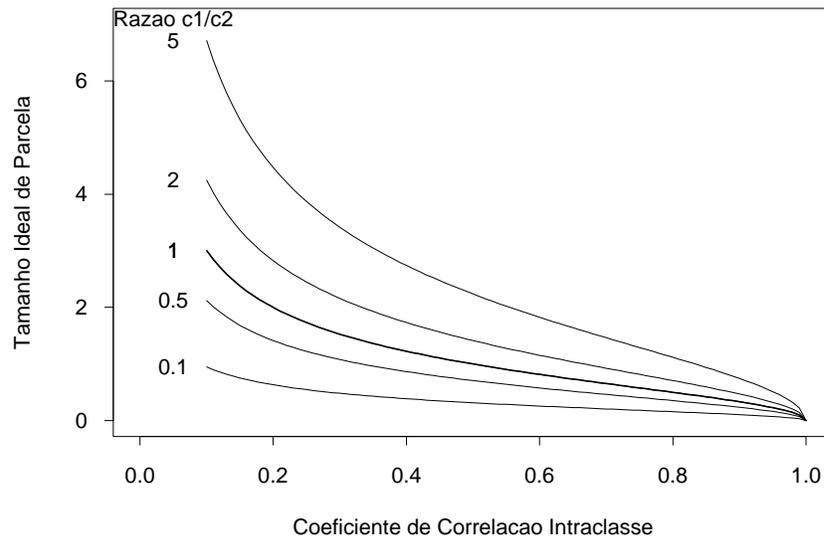


Figura 7: Influência do coeficiente de correlação intraclasse ( $\rho$ ) e da razão de custos de levantamentos florestais sobre o tamanho ideal de parcela. O tamanho ideal é apresentado em termos de número de subparcelas ( $k^*$ ).

1. minimizar o erro padrão da média ( $s_{\bar{x}}$ ) mantendo o custo total ( $T$ ) constante;
2. minimizar o custo total mantendo o erro padrão constante.

A solução das equações é a mesma para ambas alternativas, resultando no tamanho ideal:

$$k^* = \sqrt{k \frac{c_1}{c_2} \frac{s_D^2}{s_E^2 - s_D^2}} = \sqrt{\frac{c_1}{c_2} \frac{1 - \hat{\rho}}{\hat{\rho}}} \quad (29)$$

Note que o tamanho ideal depende da razão dos custos e do coeficiente de correlação intraclasse. A figura 7 mostra como estas duas grandezas influenciam o tamanho idela.

A razão de custos tem um forte efeito sobre o tamanho de parcela. Razões maiores que um indicam que o custo de deslocamento é maior que o custo de medição, o que acontece em levantamentos realizados em grandes áreas como levantamentos regionais. Já razões menores abaixo de um ocorrem em pequenas áreas ou quando um grande número medições ou medições complexas são tomadas em cada parcela. Neste caso, o método sugere o uso de pequenas parcelas. A influência da razão de custos, entretanto, torna-se muito pequena para valores do coeficiente de correlação intraclasse acima de 0.8.

EXERCÍCIO 4.6: Tamanho de Parcela e Custos de Medição

Faça gráficos do tamanho ideal de parcela para os dados de floresta hipotética (exercício 4.4) para razões de custo de 0.5, 1 e 5.

EXERCÍCIO 4.7: Tamanho de Parcela e Custos de Medição - II

Considere o levantamento em fragmentos florestais (arquivo FRAGMEN.DAT). Encontre o tamanho ideal de parcela para as variáveis quantitativas medidas quando a razão de custos for 0.5, 1 e 5.

Maiores detalhes sobre tamanho e número de parcela podem ser encontrados nos trabalhos de Zeide (1980), Stauffer (1982) e Gomes e Chaves (1988).

---

## CAPÍTULO 5

# Amostragem Aleatória Estratificada (AAE)

---

Em muitos levantamentos e inventários florestais é possível obter algum tipo de informação sobre a floresta a ser estudada anteriormente ao planejamento do levantamento. As informações preliminares permitem identificar dentro da floresta áreas que serão mais homogêneas quanto às variáveis estudadas e que constituirão subáreas com importância particular, devendo-se obter estimativas específicas para elas. Na teoria de amostragem estatística essas subáreas em que a área de estudo é subdividida são chamadas de *estratos*.

Alguns aspectos importantes que orientam a estratificação em levantamentos florestais são:

- dados topográficos;
- indicadores ecológicos: espécies indicadoras, variações ambientais, etc.;
- dados de tipologia florestal;
- dados de cadastro;
- sensoriamento remoto;
- dados de índice de sítio.

### 5.1 Estimativas dos Parâmetros

Na prática, a amostragem estratificada aleatória consiste num conjunto de AASs realizadas de modo independente em cada estrato. Portanto, para se obter estimativas a nível de estrato os estimadores apresentados para AAS podem ser utilizados. As estimativas dos parâmetros a nível da área de estudo como um todo (floresta) são diferentes pois devem combinar as informações dos diferentes estratos.

Se uma floresta é subdividida em  $L$  estratos cada um com tamanho  $N_h$  ( $N = \sum_{h=1}^L N_h$ ) as estimativas a nível de cada estrato são:

$$\text{Média: } \bar{y}_h = \left( \frac{1}{n_h} \right) \sum_{i=1}^{n_h} y_{ih} \quad (30)$$

$$\text{Variância: } s_h^2 = (1 - f_h) \left( \frac{1}{n_h - 1} \right) \sum_{i=1}^{n_h} (y_{ih} - \bar{y}_h)^2 \quad (31)$$

$$\text{Erro Padrão da Média: } s_{\bar{y}_h} = \sqrt{\frac{s_h^2}{n_h}} \quad (32)$$

onde

$n_h$  é o tamanho da amostra dentro de cada estrato  $h$ , número de parcelas locadas em cada estrato  $h$ ;

$f_h = n_h/N_h$  é a fração amostrada em cada estrato  $h$ .

As estimativas a nível da área estudada como um todo são obtidas ponderando-se as estimativas a nível dos estratos pelos tamanhos relativos dos estratos, isto é, a relação entre  $N_h$  e  $N$ :

$$\text{Média: } \bar{y}_{st} = \left( \frac{1}{N} \right) \sum_{h=1}^L N_h \bar{y}_h \quad (33)$$

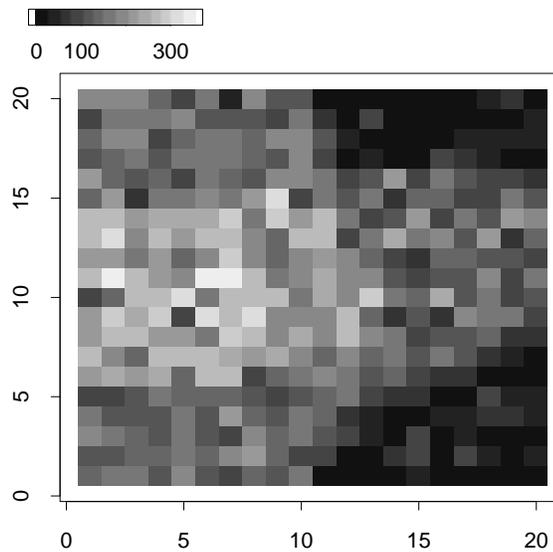


Figura 8: Padrão espacial da produção ( $st/ha$ ) em uma floresta hipotética.

$$\text{Erro Padrão da Média: } s_{\bar{y}_{st}} = \sqrt{\frac{1}{N^2} \sum_{h=1}^L \frac{N_h^2 s_h^2}{n_h} (1 - f_h)} \quad (34)$$

$$\text{Graus de Liberdade: } d = \frac{\left( \sum_{h=1}^L w_h s_h^2 \right)^2}{\sum_{h=1}^L (w_h s_h^2)^2 / (n_h - 1)} \quad (35)$$

(36)

onde  $w_h = N_h(N_h - n_h)/n_h$  é um peso relativo a cada estrato.

Com base nessas estimativas, o Intervalo de Confiança de 95% é construído

$$\bar{y}_{st} \pm (t_{0.975;d}) s_{\bar{y}_{st}}.$$

#### EXERCÍCIO 5.1: Comparação AAS vs. AAE

O arquivo FOREHIPO.DAT apresenta os dados da floresta hipotética com informação de estratificação. Assuma que cada medida da variável  $x$  se refere à produção de madeira ( $st/ha$ ) da floresta, sendo que o padrão espacial desta floresta segue a figura 8. Compare o desempenho da AAS contra a AAE quando se utiliza amostras de tamanho 9, 21 e 30.

#### EXERCÍCIO 5.2: Floresta do Maranhão

Utilizando os dados da floresta do maranhão (arquivo MARANHAO.DAT) obtenha a média e o respectivo intervalo de confiança de 95

EXERCÍCIO 5.3: *Floresta Plantada*

Num inventário em floresta de eucalipto (arquivo EUCALYPT.DAT), a floresta foi estratificada por espécie, rotação e idade de acordo com a seguinte tabela:

<i>E. grandis</i>			<i>E. saligna</i>		
Rotação	Idade (anos)	Área do Estrato (ha)	Rotação	Idade (anos)	Área do Estrato (ha)
1ª	3	175	1ª	3	201
	4	129		4	233
	5	215		5	239
	6	200		6	188
	7	240		7	168
	8	110		8	187
2ª	9	185	2ª	9	207
	10	280		10	145
	11	205		11	192
	12	150		12	223
	13	180		13	259
			14	271	199

Obtenha a média e o respectivo intervalo de confiança de 95

## 5.2 Alocação das Parcelas por Estrato

Ao alocar uma dada quantidade de unidades amostrais (parcelas) nos diferentes estratos podemos utilizar três critérios:

**Alocação Proporcional à Área:** cada estrato recebe uma proporção do número total de parcela ( $n$ ) de modo proporcional à sua área:

$$n_h = \left( \frac{N_h}{N} \right) n \quad (37)$$

**Alocação Ótima:** cada estrato recebe uma proporção de  $n$  em função da sua área combinada com a sua variabilidade:

$$n_h = \left( \frac{N_h s_h^2}{\sum_{k=1}^L N_k s_k^2} \right) n \quad (38)$$

**Alocação Ótima com Custo Variável:** é a alocação que considera que o custo de obtenção de uma amostra varia de estrato para estrato. Para definir essa alocação não é necessário saber com detalhes o custo em cada um dos estratos, basta saber a relação ou proporção dos custos de amostragem entre os estratos.

$$n_h = \left[ \frac{N_h s_h / \sqrt{c_h}}{\sum_{k=1}^L (N_k s_k / \sqrt{c_k})} \right] n \quad (39)$$

onde  $s_h = \sqrt{s_h^2}$  e  $c_h$  é o custo de amostragem de uma parcela em cada estrato  $h$ .

**EXERCÍCIO 5.4:** Alocação de Parcelas - Floresta no Maranhão

Com base no exemplo de inventário da floresta no Maranhão, verifique para as variáveis área basal (AB) e volume para serraria (V) como as parcelas seriam alocadas pelo três métodos de alocação. No método de custo variável, assumo que a relação de custos para os estratos A:B:C:D é 1:4:3:5.

**EXERCÍCIO 5.5:** Alocação de Parcelas - Floresta Plantada

Verifique como as parcelas seriam alocadas no exemplo de inventário de floresta plantada, segundo os métodos de alocação proporcional à área e segundo o método da alocação ótima.

### 5.3 Amostragem Sistemática com Pós-Estratificação

Frequentemente não possuímos informações detalhadas que nos permitem fazer uma estratificação previamente ao levantamento florestal. A amostragem sistemática, entretanto, fornece informações que podem ser utilizadas para fazermos a estratificação posteriormente ao levantamento (pós-estratificação). A amostragem sistemática tem uma cobertura regular sobre o terreno de modo que o tamanho de cada estrato pode ser estimado pelo número de parcelas classificadas como pertencente aquele estrato.

---

## CAPÍTULO 6

# Amostragem com Probabilidade Proporcional ao Tamanho

---

A amostragem com probabilidade proporcional ao tamanho (APPT) é um método típico das Ciências Florestais e desenvolvido especificamente para a amostragem de espécies arbóreas. O método de amostragem foi inicialmente proposto pelo Florestal austríaco Bitterlich na década de 40, posteriormente foi desenvolvido por uma série de Florestais norte-americanos, europeus e japoneses, destacando-se os trabalhos de Keen, Grosenbaugh e Hirata na década de 50. O nome “Amostragem com Probabilidade Proporcional ao Tamanho” não é o único nome utilizado, existindo uma grande variedade de nomes que são utilizados para designá-lo, como por exemplo:

- Amostragem por contagem angular;
- Amostragem por parcelas de raio variável;
- Amostragem por parcelas de área variável;
- Amostragem horizontal por pontos.

O método se baseia no chamado “Princípio de Bitterlich” enunciado por Bitterlich em 1947. Esse princípio estabelece que:

“ O número de árvores ( $N$ ) de um povoamento , cujos os DAPs vistos de um ponto fixo aparecem maiores a um dado valor ( $\alpha$ ) é proporcional à sua área basal por hectare ( $G$ ).”

Dessa forma é possível estimar a área basal da floresta ( $G$ ) diretamente através de um processo de contagem que enumere as árvores ( $N$ ) em função de um ângulo de visada ( $\alpha$ ).

### 6.1 Procedimento de Campo

O procedimento da APPT em campo incia-se com a seleção de um ângulo de visada ( $\alpha$ ) em campo. A partir de um ponto locado de modo aleatório na floresta, se faz um giro de 360 graus em que cada árvore visualizada é comparada com o ângulo  $\alpha$  selecionado. As árvores cujos os DAPs *aparecem* maiores que o ângulo  $\alpha$  são as árvores incluídas na amostrada, enquanto que as demais são ignoradas, isto é, não são amostradas. Para se obter a estimativa da área basal da floresta basta multiplicar o número de árvores amostradas no ponto ( $N$ ) por um fator  $F$  que é função do ângulo  $\alpha$  selecionado. Assim a estimativa da área basal para um ponto de específico de amostragem é dada pela fórmula:

$$\begin{aligned} \alpha &\implies F \\ G &= N \times F \end{aligned}$$

Árvores cujo DAP aparece exatamente igual ao ângulo de visada ( $\alpha$ ) não contadas como valendo 0.5. A figura 9 mostra como as árvores amostradas são definidas num ponto de amostragem.

A área basal da floresta ( $G$ ) não é obtida com apenas um único ponto. Na verdade, cada ponto funciona como uma unidade amostral ou parcela e a estimativa final da área basal é dada pela média dos valores

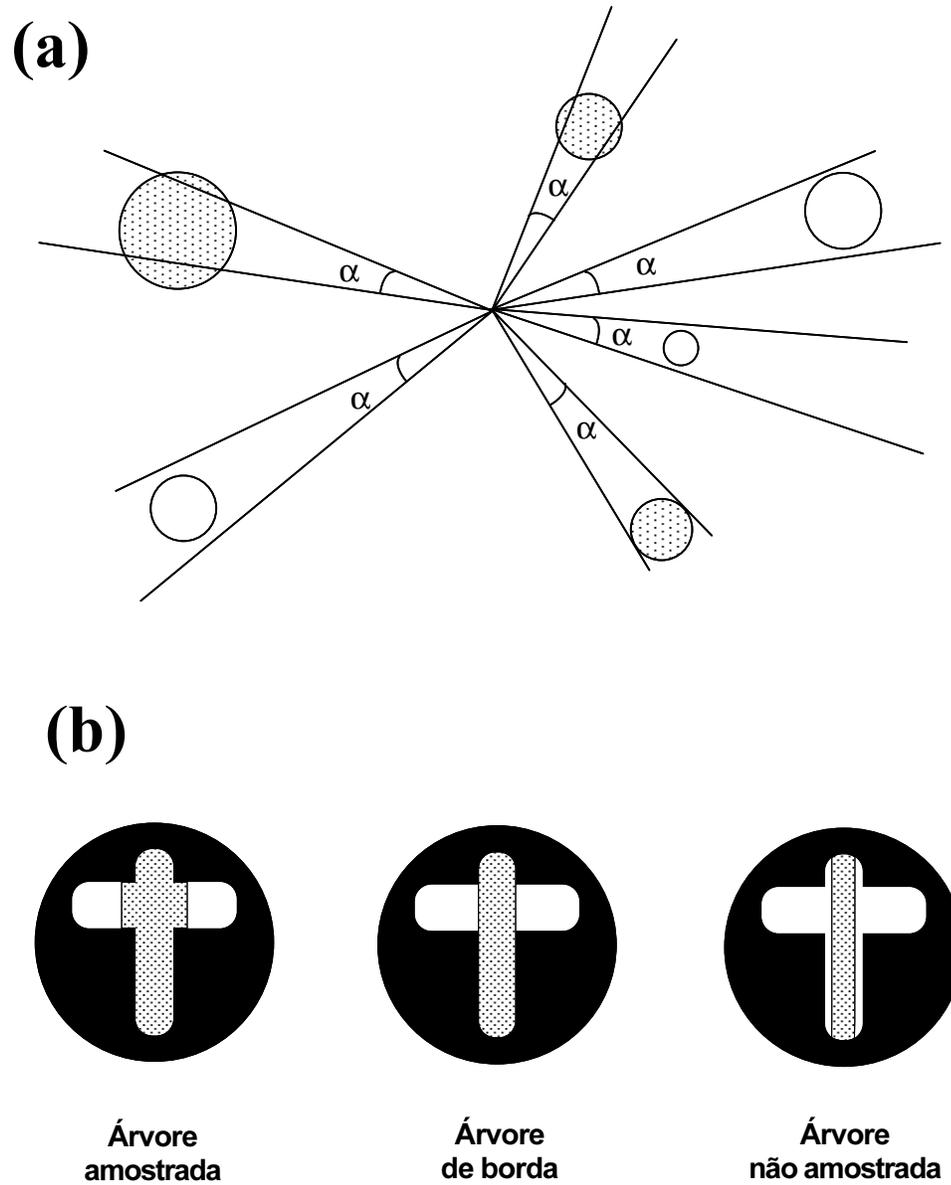


Figura 9: Modo de amostragem das árvores utilizando o princípio de Bitterlich. (a) As árvores cujos DAPs aparecem maiores que um dado ângulo  $\alpha$  são incluídas na amostra, enquanto as demais são ignoradas. (b) Um método prático de visualizar os DAPs das árvores é utilizar um tubo cujo comprimento e abertura resultam no ângulo  $\alpha$  desejado (Tubo de Panamá). Neste caso as árvores amostradas aparecerão maiores que a abertura vertical do tubo.

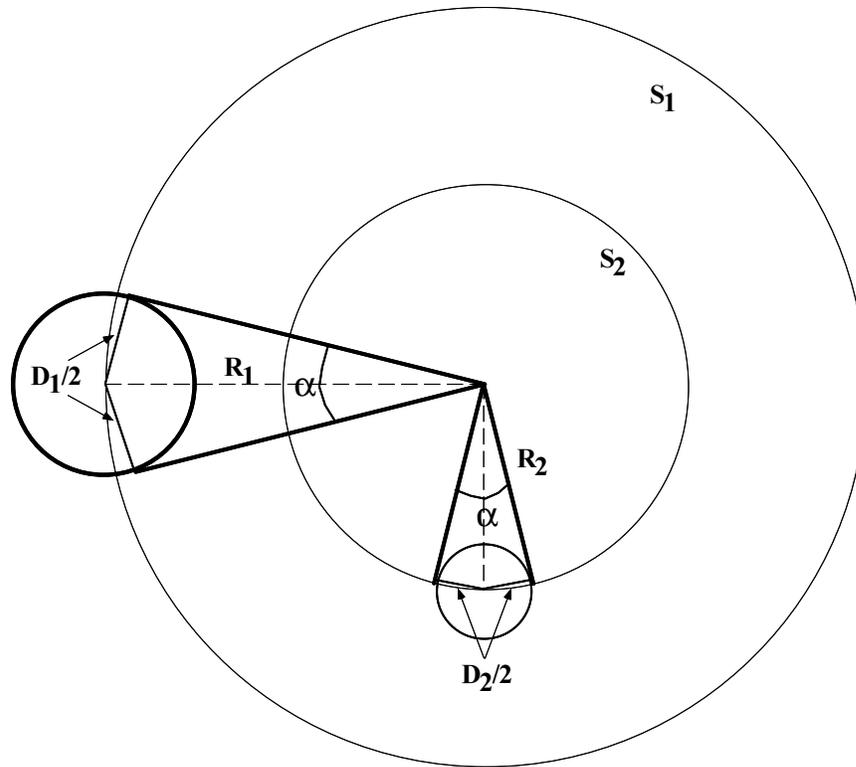


Figura 10: Formação hipotética de parcelas de amostragem num ponto de visualização para duas árvores com DAPs  $D_1$  e  $D_2$ .

obtidos em cada ponto. Como as unidades amostrais são os pontos de visada, a APPT pode ser aplicada em conjunto com os demais métodos de amostragem vistos (AAS, ASS e AEA). Dada a rapidez com que as medidas são tomadas em cada ponto, a APPT é mais frequentemente utilizada em conjunto com a amostragem sistemática estabelecendo um reticulado sobre a floresta onde cada intersecção do reticulado é tomado como um ponto de amostragem.

## 6.2 Princípio de Bitterlich

O princípio de Bitterlich funciona porque ele faz com que cada árvore amostrada contribua exatamente  $Fm^2/ha$  para estimativa da área basal da floresta ( $G$ ). Onde  $F$  é uma constante que depende unicamente do ângulo de visada ( $\alpha$ ). Vejamos como isso ocorre, acompanhando o que fazemos num ponto de amostragem:

1. As árvores são visualizadas através de um ângulo  $\alpha$  projetado na altura dos DAPs.
2. Uma série de parcelas circulares concêntricas são formadas (hipoteticamente), sendo que o raio ( $R$ ) de cada parcela está associado ao DAP ( $D$ ) de cada árvore (figura 10).
3. O raio  $R$  de cada parcela concêntrica é determinado pelo diâmetro de cada árvore e não é influenciado pela localização espacial da árvore. Seguindo a figura 10, toda árvore com diâmetro  $D_1$  que estiver a uma distância igual ou menor que  $R_1$  será amostrada. Igualmente, toda árvore com DAP  $D_2$  que estiver a uma distância igual ou menor que  $R_2$  será amostrada. Cada árvore forma assim uma parcela circular hipotética: a árvore com DAP  $D_1$  forma a parcela de raio  $R_1$  e área  $S_1$ ; a árvore com DAP  $D_2$  forma a parcela de raio  $R_2$  e área  $S_2$ .
4. A razão entre o DAP ( $D$ ) de cada árvore amostrada e o raio  $R$  que faz com que esta árvore seja amostrada

é constante:

$$k = \frac{D_1}{100R_1} = \frac{D_2}{100R_2}$$

$$k = \frac{D}{100R} = 2 \sin \frac{\alpha}{2}$$

O número 100 na equação usado para converter o diâmetro  $D$  de centímetros para metros, a mesma unidade de medida do raio  $R$ . É importante notar que a constante  $k$  depende única e exclusivamente do ângulo  $\alpha$  selecionado, sendo a mesma para todas as árvores amostradas.

5. O fato da razão entre o DAP ( $D$ ) de cada árvore amostrada e o raio ( $R$ ) da parcela circular hipotética que a contem ser a constante  $k$ , implica que a razão entre a área seccional  $((\pi/4)(D/100)^2)$  de cada árvore amostrada e a área  $(\pi R^2)$  da parcela hipotética que a contem também será constante.

$$k = \frac{D}{100R} \implies \frac{(\pi/4)(D/100)^2}{\pi R^2} = \text{constante}$$

6. O fator de área basal  $F$  é exatamente esta constante, tomando-se o cuidado de expressar a área seccional das árvores em  $m^2$  e a área da parcela em  $ha$  de modo que a área basal resultante seja em  $m^2/ha$ :

$$F = \frac{(\pi/4)(D/100)^2}{(\pi R^2)/10000} = \frac{D^2}{4R^2} = 2500k^2$$

7. Para uma interpretação mais clara, a equação do fator de área basal pode ser re-arranjada do seguinte modo:

$$F = \left[ \frac{\pi D^2}{4(100)^2} \right] \times \left[ \frac{10000}{\pi R^2} \right]$$

$$F = \left[ \begin{array}{c} \text{Área Seccional} \\ \text{de cada Árvore} \\ \text{Amostrada (m}^2\text{)} \end{array} \right] \times \left[ \begin{array}{c} \text{Fator} \\ \text{de Conversão} \\ \text{para m}^2/ha \end{array} \right]$$

Assim, cada árvore amostrada contribui exatamente  $F m^2/ha$  para a estimativa da área basal da floresta, independentemente do seu diâmetro e da sua distância do ponto de amostragem.

8. Como a constante  $k$  é função do ângulo de visada  $\alpha$  podemos escolher esse ângulo segundo o valor de  $F$  que desejarmos:

$$k = \sqrt{\frac{F}{2500}}$$

$$\alpha = 2 \arctan(k/2)$$

$F$	$k$	$\alpha$
1	0.0200	1.1460°
2	0.0283	1.6208°
3	0.0346	1.9852°
4	0.0400	2.2924°

Alguns aspectos práticos devem ser tomados em conta na escolha do melhor fator  $F$ .

### 6.3 Princípio de Bitterlich como APPT

Para a estimativa da área basal, o princípio de Bitterlich nos permite uma estimativa direta, baseada apenas na contagem das árvores amostradas. O mesmo princípio, no entanto, pode ser utilizado para estimar qualquer atributo da floresta como volume, número de árvores por hectare, índices de diversidade em florestas nativas, etc.

Este método será eficiente sempre que o atributo desejado possuir uma relação com a área basal da floresta ou com a área seccional das árvores individualmente.

Para que outros atributos possam ser estimados, o método utiliza um “Fator de Expansão” ( $F$ ) como variável de ponderação para os atributos medidos nas árvores individualmente. Na verdade, o  $FE$  é uma estimativa de quantas árvores por hectare cada árvore amostrada representa. Se tomarmos  $i$  como o índice de cada árvore amostrada teremos:

$$\begin{aligned}
 FE_i &= \frac{1 \text{ hectare}}{\text{Área da Parcela Circular com a árvore } i} \\
 &= \frac{10000}{\pi R_i^2} \\
 &= \frac{10000}{\pi \left(\frac{D_i}{100k}\right)^2} \\
 &= \frac{10000k^2}{\pi \frac{D_i^2}{10000}} \\
 &= \frac{2500k^2}{(\pi/4)D_i^2} \\
 &= \frac{F}{G_i} \left[ \frac{m^2/ha}{m^2} \Rightarrow \frac{1}{ha} \right]
 \end{aligned}$$

Para obtermos a estimativa do número de árvores num dado ponto de amostragem basta somarmos os  $FE_i$ :

$$\hat{N}(arv/ha) = \sum_{i=1}^n FE_i$$

onde  $n$  é o número de árvores amostradas num dado ponto. Se desejarmos o volume da floresta basta somarmos o produto do volume de cada árvore amostrada pelo seu  $FE$ :

$$\hat{V}(m^3/ha) = \sum_{i=1}^n FE_i V_i$$

onde  $V_i$  é o volume de cada da árvores amostrada em  $m^3$ .

#### EXERCÍCIO 6.1: Aplicação da APPT - Floresta Tropical

O arquivo ELDORADO.DAT traz os dados de um levantamento baseado em APPT realizado numa floresta nativa utilizando fator de área basal 2. Foram levantados 4 transectos de 1000m com 10 pontos em cada transecto, separados por distâncias de 100m.

- Encontre a média e o intervalo de confiança de 95 para a área basal e número de árvores por hectare.
- Encontre as estimativas de dominância (área basal) e densidade (número de árvores por hectare) para cada espécie.

---

## CAPÍTULO 7

# Monitoramento Florestal

---

A floresta é um sistema dinâmico que se altera de diversas maneiras no decorrer do tempo. O monitoramento florestal tem como meta acompanhar as mudanças temporais que ocorrem na estrutura e funcionamento da floresta. Em termos de amostragem, as mudanças que ocorrem na floresta possuem dois fatores:

1. Alteração do valor do *atributo* das unidades amostrais. As unidades amostrais podem ser árvores ou parcelas.
2. Alteração do valor *agregado* do atributo para a floresta.

Vejam os estes fatores em termos do crescimento da floresta, que é a mudança temporal mais estudada devido à sua importância econômica. A alteração do valor do atributo se refere ao crescimento das árvores ou parcelas individualmente. Já a alteração do valor agregado do atributo se aplica à variação do volume de madeira na floresta como um todo, isto é, no aumento volumétrico da floresta. O incremento volumétrico da floresta é resultado não só do crescimento individual das árvores, mas também de como este crescimento individual das árvores varia de local para local na floresta. Ocorrência de secas, geadas, incêndios e colheitas mudam o padrão de crescimento das árvores e atuam de maneira desuniforme na floresta como um todo.

O crescimento volumétrico da floresta também pode ser analisado em termos de seus componentes:

- Sobrevivência ( $S$ ) ou as árvores com tamanho acima de um dado limite que existem na primeira ocasião e sobrevivem até a segunda. O crescimento destas árvores é o componente principal para o crescimento da floresta entre as duas ocasiões de amostragem.
- Mortalidade ( $M$ ) ou as árvores que existem na primeira ocasião, mas morrem antes da segunda.
- Colheita ( $C$ ) ou as árvores que existem na primeira ocasião e são cortadas antes da segunda.
- Ingresso ( $I$ ) ou as árvore que cruzam o tamanho limite de amostragem entre a primeira e a segunda ocasião.

Assim podemos dizer que o volume da floresta na primeira ocasião de amostragem ( $V_1$ ) é dado pela fórmula:

$$V_1 = S_1 + M + C,$$

já na segunda ocasião o volume da floresta é:

$$V_2 = S_2 + I.$$

Estimar o crescimento “líquido” da floresta dado por:

$$\begin{aligned}\Delta V &= V_2 - V_1 \\ &= (S_2 - S_1) + I - M - C\end{aligned}$$

é diferente de estimar os componentes do crescimento, isto é,  $S_1$ ,  $S_2$ ,  $M$ ,  $C$  e  $I$ .

## 7.1 Objetivos do Monitoramento

Um sistema de monitoramento em florestas tem em geral os seguintes objetivos:

1. Estimar os parâmetros da população em momentos específicos do tempo. Parâmetros frequentemente de interesse são:
  - volume de madeira,
  - número de árvores,
  - taxa de ataque de insetos,
  - ocorrência de uma da espécie em floresta nativa,
  - diversidade de espécies.
2. Estimar a mudança dos parâmetros entre duas ocasiões específicas. O objetivo pode ser relacionar o crescimento da floresta com aspectos climáticos ou do solo, analisar o impacto de alterações antrópicas sobre o crescimento e estrutura da floresta, analisar o efeito de novas técnicas silviculturais ou de manejo.
3. Estimar os componentes da mudança temporal, como por exemplo os componentes de sobrevivência, mortalidade, colheita e ingresso no estudo do crescimento. Ao estimar os componentes podemos ter em mente diferentes questões:
  - Estimar as mudanças nas unidades, como por exemplo a mudança do tamanho ou vigor das árvores que sobreviveram entre duas ocasiões de amostragem, ou a mudança do grau de infestação de patógenos, ou o grau de desfolhamento nas árvores individuais.
  - Estimar a mudança “média” na unidades de amostragem; por exemplo o crescimento “típico” das árvores nos primeiros anos após o desbaste.
  - Acompanhar a variabilidade de uma dada característica durante um período de tempo para uma unidade específica; por exemplo a variação no incremento diamétrico anual de uma árvore durante um período de 5 anos.
4. Agregar informações sobre as unidades amostrais ao longo do tempo. As empresas florestais costumam manter um banco de dados da história silvicultural dos talhões. O biometrista florestal estaria interessado em saber como a história silvicultural se relaciona com as características de crescimento, mortalidade e vigor das árvores ao longo do tempo.
5. Acumular amostras ao longo do tempo, especialmente para subpopulações raras. Nos levantamentos de florestas nativas muitas de espécies são representadas por apenas uma ou algumas poucas árvores. Levantamentos sucessivos na mesma floresta resultam num maior número de árvores das espécies raras sendo amostradas ao longo do tempo.

## 7.2 Métodos de Amostragem

Existem três abordagens básicas de amostragem para o monitoramento florestal:

- Amostragem por Substituição Completa (ASCo);
- Amostragem por Remedição Completa (ARCo);
- Amostragem com Substituição Parcial (ASPa).

### 7.2.1 Amostragem por Substituição Completa (ASCo)

Esta abordagem também pode ser chamada de inventários sucessivos independentes, pois consiste em realizar levantamentos ou inventários totalmente independentes nas sucessivas ocasiões de amostragem. Nenhuma preocupação é tomada com a manutenção de parcelas permanentes ou com a consistência das medidas ao longo do tempo.

A ASCo tem como grandes vantagens a sua simplicidade e flexibilidade. Qualquer delineamento amostral pode ser utilizado e, se necessário, o delineamento pode ser modificado nas ocasiões sucessivas. O número de parcelas e o orçamento também podem ser alterados ao longo do tempo em função da disponibilidade de recursos ou precisão necessária em cada ocasião de amostragem. A ASCo é também um bom sistema para acumular amostras sobre populações raras e detectar a ocorrência de novos eventos na floresta. As desvantagens incluem a impossibilidade de se estimar os componentes da mudança na floresta ou de se agregar informações sobre as unidades ao longo do tempo, e as mudanças “líquidas” na floresta são estimadas ineficientemente (com alta variância).

### 7.2.2 Amostragem por Remedição Completa (ARCo)

A amostragem por remedição completa é o método mais utilizado no Brasil nas empresas que trabalham com florestas de rápido crescimento, sendo geralmente chamada de Inventário Florestal Contínuo. Na ARCo, todas as parcelas são permanentes, isto é, são instalada no início do monitoramento e re-visitadas em todas as ocasiões sucessivas de amostragem. É importante lembrar que somente as parcelas permanentes são acompanhadas. Assim, perda de parcelas e inclusão de novas parcelas se tornam um problema nesse tipo de amostragem.

Essa abordagem é eficiente para estimar os componentes de mudança na floresta, assim como para agregar informações na construção de um histórico de manejo. As informações geradas pela ARCo são as melhores informações para construir modelos de crescimento e produção, ou outros modelos que representem a estrutura e funcionamento da floresta ao longo do tempo.

A mais séria limitação desse método é que ele acompanha apenas as mudanças dos valores dos atributos para as unidades ao longo do tempo, não sendo capaz de detectar alterações na composição da floresta (valor agregado dos atributos). Mudanças que ocorrem na floresta como um todo alterando a sua estrutura e não ocorrem nas parcelas permanentes, ou mesmo que ocorrem de modo distinto na floresta e nas parcelas monitoradas, tornam este método problemático em termos da manutenção da representatividade ao longo do tempo.

### 7.2.3 Amostragem com Substituição Parcial (ASPa)

Na amostragem com substituição parcial, as unidades amostrais são re-visitadas um número variável de vezes. A cada ocasião de amostragem algumas unidades são substituídas por novas unidades, de modo que num monitoramento com 3 ocasiões, algumas unidades são visitadas 3 vezes, outras unidades são visitadas duas vezes, e ainda algumas unidades são visitadas apenas uma vez.

A ASPa é uma situação intermediária entre a ASCo e a ARCo que tenta contornar as limitações de ambas. A contínua introdução de novas unidades amostrais contorna o problema da ARCo de não detectar alterações na composição e estrutura da floresta que ocorrem de modo espacialmente diferenciado. A existência de parcelas permanentes permite agregar informações históricas em algumas unidades amostrais e estimar os componentes da mudança na floresta. A ASPa é a única abordagem capaz de alcançar todos os objetivos do monitoramento listados acima, embora não os alcance necessariamente com a mesma eficiência das outras abordagens.

A principal desvantagem da ASPa é a sua complexidade. Com um grande número de ocasiões de amostragem, os estimadores se tornam extremamente complexos, principalmente os estimadores de variância. Esta complexidade faz com que a combinação da ASPa com outros delineamentos amostrais, como a amostragem aleatória estratificada, se torne ainda mais complexa. Entretanto, a estratificação é, em geral, um procedimento quase essencial em levantamentos florestais.

### 7.2.4 ASPa em Duas Ocasões

Apresentamos aqui os estimadores para ASPa quando apenas duas ocasiões de amostragem foram realizadas. Definição das parcelas:

- $n$  unidades foram amostradas em cada uma das duas ocasiões;
- $m$  unidades são comuns às duas ocasiões (parcelas permanentes);
- $u$  unidades estão presentes somente na segunda ocasião (parcelas temporárias), assim  $u = n - m$ .

Utilizando as parcelas que foram medidas somente na segunda ocasião, temos um estimador para a média da floresta na segunda ocasião:

$$\bar{y}_{2T} = \sum_{i=1}^u y_{2i}/u$$

$$\text{Var}(\bar{y}_{2T}) = s_2^2/u$$

O estimador para a média da floresta com base nas unidades medidas em ambas ocasiões é:

$$\bar{y}_{2P} = \bar{y}_{2m} + \hat{\beta}(\bar{y}_{1n} - \bar{y}_{1m})$$

$$\text{Var}(\bar{y}_{2P}) = \frac{s_2^2}{m} \left(1 - \frac{u}{n} \hat{\rho}^2\right)$$

onde:

$\hat{\beta}$  é o coeficiente de regressão estimado a partir do modelo:

$$y_{2i} = \beta y_{1i} + \varepsilon_i$$

( $i = 1, 2, \dots, m$ ) envolvendo apenas as unidades medidas em ambas ocasiões.

$\hat{\rho}$  é o coeficiente de correlação entre  $y_{1i}$  e  $y_{2i}$ , isto é, os valores medidos nas unidades visitadas em ambas ocasiões.

O estimador da ASPa para a média da floresta na segunda ocasião combina os dois estimadores anteriores:

$$\bar{y}_2 = \frac{w_T \bar{y}_{2T} + w_P \bar{y}_{2P}}{w}$$

onde:

$$w_T = 1/\text{Var}(\bar{y}_{2T})$$

$$w_P = 1/\text{Var}(\bar{y}_{2P})$$

$$w = w_T + w_P$$

A variância deste estimador combinado é:

$$\text{Var}(\bar{y}_2) = w \left[ 1 + \frac{4}{w^2} \left( \frac{w_T(w - w_T)}{u - 1} + \frac{w_P(w - w_P)}{m - 1} \right) \right]$$

## Referências

- [1] Husch, B.; Miller, C.I.; Beers, T.W.; 1982 **Forest Mensuration** (3a. ed.). John Wiley & Sons, New York.
- [2] Avery, T.E.; Burkhart, H.E.; 1983 **Forest Measurements**. McGraw-Hill, New York.
- [3] Freese, F. 1962 Elementary Forest Sampling. U.S.D.A. Forest Service, *Agriculture Handbook 232*. Re-impressão O.S.U. Book Stores, 1989, Corvallis.
- [4] Schreuder, H.T.; Gregoire, T.G.; Wood, G.B; 1993 **Sampling Methods for Multiresource Forest Inventory**. John Wiley & Sons, New York.
- [5] de Vries, P.G. 1986 **Sampling Theory for Forest Inventory: a teach-yourself course**. Springer-Verlag, Berlin.
- [6] Scheaffer, R.L.; Mendenhall, W.; Ott, L.; 1986 **Elementary Survey Sampling** (3a.ed.) Duxbury Press, Boston.
- [7] Thompson, S.K. 1992 **Sampling**. John Wiley & Sons, New York.
- [8] Cochran, W.G. 1977 **Sampling Techniques** (3a. ed.) John Wiley & Sons, New York.
- [9] Zeide, B 1980 Plot size optimization. *Forest Science*, **26**(2):251-257.
- [10] Stauffer, H.B. 1982 A sample size table for fores sampling. *Forest Science*, **28**(4):777-784.
- [11] Gomes, F.P.; Chaves, R.; 1988 A amostragem ótima em inventário florestal. *IPEF*, (38):17-22.

## Funções do SPLUS

```

Sample.cluster <- function(forest = F.hipo, plot.sizes = c(1, 2, 3, 4, 5), samp.size = 60)
{
  N <- length(forest)
  l <- length(plot.sizes)
  test <- (samp.size/plot.sizes)
  if(sum(test == trunc(test)) != l)
    stop("Sample sizes is not compatible with plot sizes")
  out <- matrix(NA, ncol = 2 * l, nrow = samp.size)
  dimnames(out) <- list(NULL, sort(paste(rep(plot.sizes, 2), c("idx", "val"), sep = ".")))
  for(i in 1:l) {
    nm1 <- paste(i, "idx", sep = ".")
    nm2 <- paste(i, "val", sep = ".")
    out[, nm1] <- ceiling(1:samp.size/i)
    tmp <- sort(sample(1:N, samp.size/i))
    tmp2 <- sort(rep(tmp, samp.size/length(tmp))) + 0:(samp.size/length(tmp) - 1)
    out[, nm2] <- forest[tmp2]
  }
  out
}

Sample.cluster2 <- function( mtx = Sample.cluster(F.hipo) )
{
  varnames <- dimnames(mtx)[[2]]
  mtx.dim <- dim(mtx)
  k <- seq(1, mtx.dim[2], 2)
  out <- matrix(0, nrow = length(k), ncol = 6)
  dimnames(out) <- list(NULL, c("n", "m", "pop.med", "con.med", "var.entre", "ep.con.med"))
  #
  #
  #
  out[1,"n"] <- length(mtx[,2])
  out[1,"m"] <- NA
  out[1,"pop.med"] <- mean(mtx[,2])
  out[1,"con.med"] <- mean(mtx[,2])
  out[1,"var.entre"] <- var(mtx[,2])
  out[1,"ep.con.med"] <- sqrt( var(mtx[,2]) / out[1,"n"] )
  #
  #
  #
  for(i in 2:length(k)){
    totpar <- tapply( mtx[,k[i]+1], mtx[,k[i]], sum )
    totpar[ is.na(totpar) ] <- 0
    m <- tapply( mtx[,k[i]+1], mtx[,k[i]], length )[1]
    out[i,"n"] <- length(totpar)
    out[i,"m"] <- m
    out[i,"pop.med"] <- sum(totpar) / out[i,"n"]
    out[i,"con.med"] <- sum(totpar) / sum(m)
    out[i,"var.entre"] <- sum( (totpar - out[i,"pop.med"])^2)/(out[i,"n"] - 1)
    out[i,"ep.con.med"] <- sqrt( out[i,"var.entre"] / (m^2 * out[i,"n"]) )
  }
  out
}

CV.calc <- function(sizes = c(1, 2, 5, 10, 15, 20), n = 30, MED = 400, DP = 200)
{
  CV <- function(x)

```

```

    {
      100 * (sqrt(var(as.vector(x)))/mean(x))
    }
  l <- length(sizes)
  out <- numeric(l)
  for(i in 1:l) {
    idx <- rep(1:n, sizes[i])
    tmp <- rnorm(sizes[i] * n, MED, DP)
    tmp2 <- tapply(tmp, idx, mean)
    out[i] <- CV(tmp2)
  }
  out
}

CV.calc2 <- function(dat=F.hipo)
{
  CV <- function(x)
  {
    100 * (sqrt(var(as.vector(x)))/mean(x))
  }
  samp <- Sample.cluster(dat)
  idx <- paste(1:5, "idx", sep=".")
  val <- paste(1:5, "val", sep=".")
  out <- numeric(5)
  for( i in 1:5){
    idx.tmp <- match(samp[,idx[i]], 1:12)
    val.tmp <- samp[!is.na(idx.tmp), val[i]]
    idx.tmp <- idx.tmp[!is.na(idx.tmp)]
    tmp <- tapply(val.tmp, idx.tmp, mean)
    tmp <- tmp[!is.na(tmp)]
    out[i] <- CV(tmp)
  }
  out
}

Estima.rho <- function(dat=F.hipo)
{
  #
  # Retorna estimativa de Rho com base num levantamento onde
  #   k=4 e r=15
  #
  #
  tmp <- Sample.cluster(dat)
  idx <- tmp[, "4.idx"]
  x <- tmp[, "4.val"]
  idx <- idx[!is.na(x)]
  #
  #
  #
  if( length(table(table(idx))) != 1)
    return(rho=NA)
  #
  #
  #
  k <- as.numeric(names(table(table(idx))))
  r <- length(unique(idx))
  linmod <- lm( x ~ idx )
  ms <- summary.aov(linmod)[,"Mean Sq"]

```

```
var.entre <- ms[1]
var.dentro <- ms[2]
rho <- (var.entre - var.dentro) / ( var.entre + (k-1)*var.dentro)
return(rho)
}

Rho.calc <- function(x, index)
{
  k <- as.numeric(names(table(table(index))))
  r <- length(unique(index))
  linmod <- lm( x ~ index )
  ms <- summary.aov(linmod)[,"Mean Sq"]
  var.entre <- ms[1]
  var.dentro <- ms[2]
  rho <- (var.entre - var.dentro) / ( var.entre + (k-1)*var.dentro)
  return(rho)
}

Sample.system <- function(dat = F.period)
{
  x <- ceiling(5 * runif(1))
  y <- ceiling(5 * runif(1))
  x <- c(x, x + seq(5, 15, 3))
  y <- c(y, y + seq(5, 15, 3))
  out <- matrix(0, nrow = 5, ncol = 5)
  for(i in 1:5) {
    for(j in 1:5) {
      out[i, j] <- dat[x[i], y[j]]
    }
  }
  out
}
```