

Definindo uma estratégia de amostragem

Censo X Amostra



O que amostrar?

Definição do objeto a ser amostrado - O que é um indivíduo?



O que amostrar?

Definição do objeto a ser amostrado - O que é um indivíduo?



O que amostrar?

Definição do objeto a ser amostrado - O que é um indivíduo?

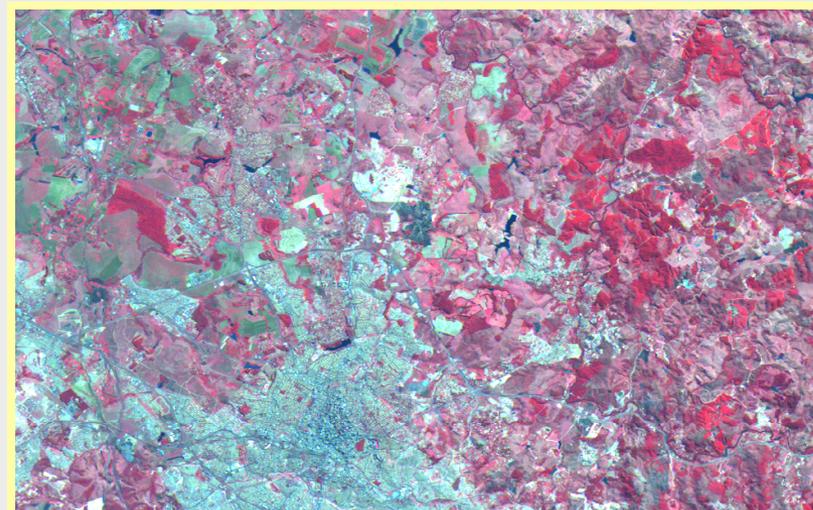
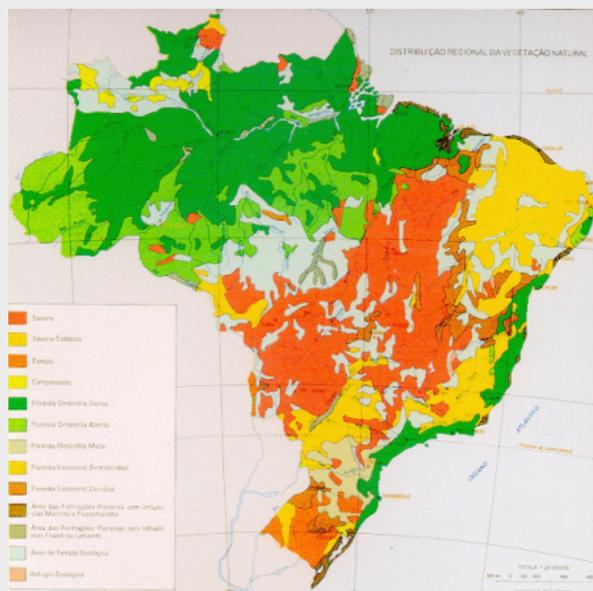


O que amostrar?

Definição das variáveis a serem amostradas



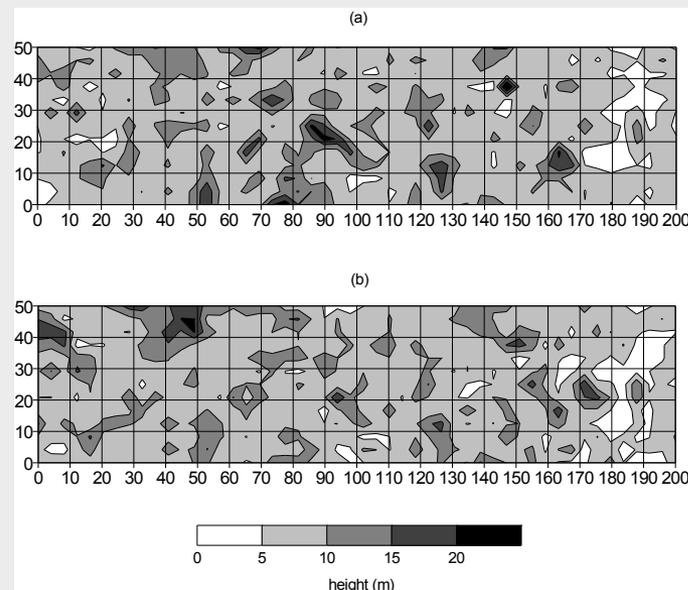
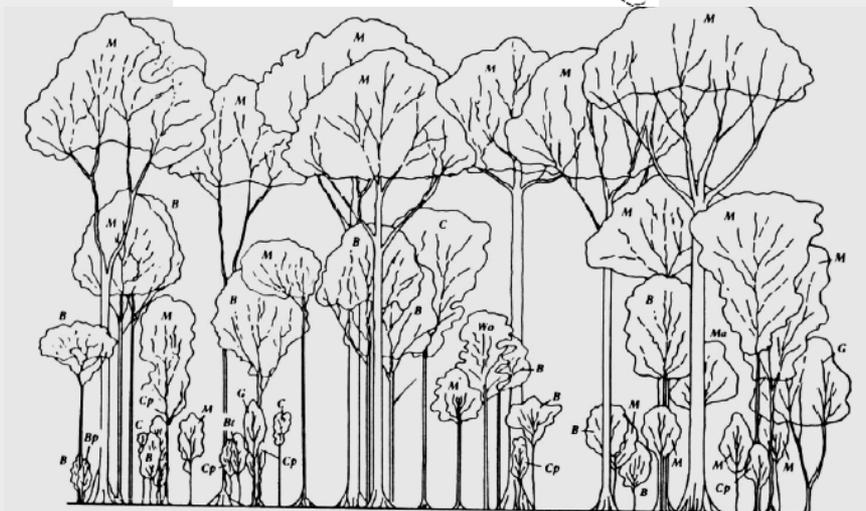
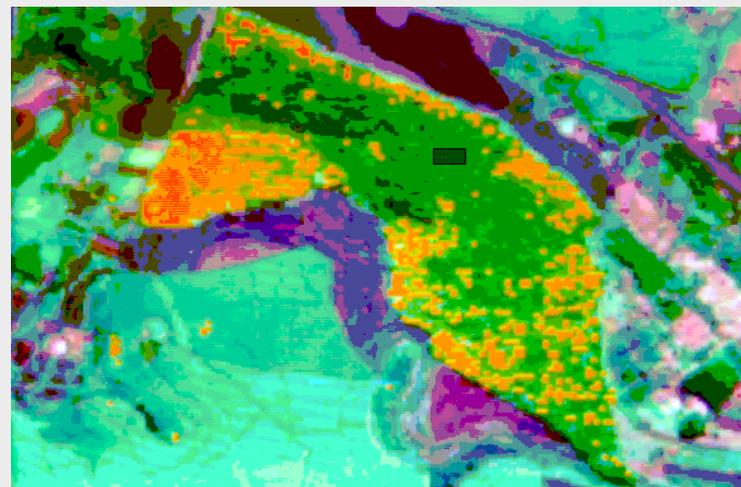
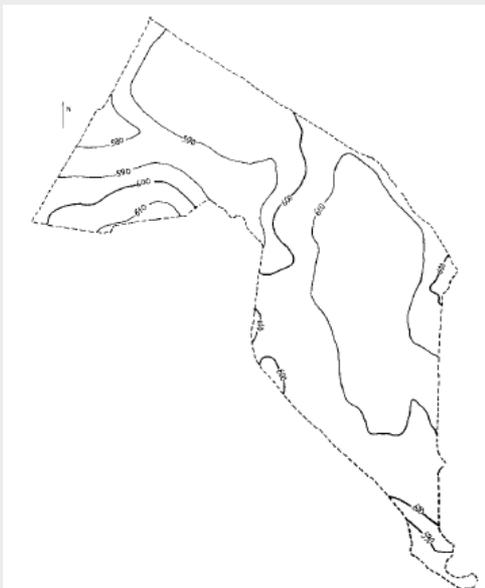
O que amostrar?
Definição do universo amostral
Escala, Heterogeneidade
Granulação do ambiente



O que amostrar?

Definição do universo amostral

Escala, Heterogeneidade, Granulação do ambiente



Como amostrar?

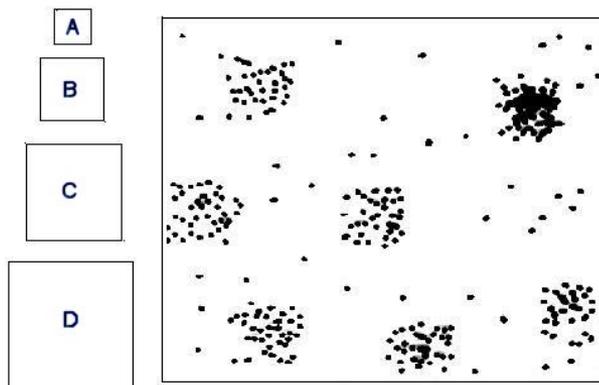
A definição do método de amostragem

Amostra X Unidade amostral

Amostragem **por área** ou **sem área**

Amostragem **por área**:

Uso de parcelas – Área fixa, número de indivíduos variável.



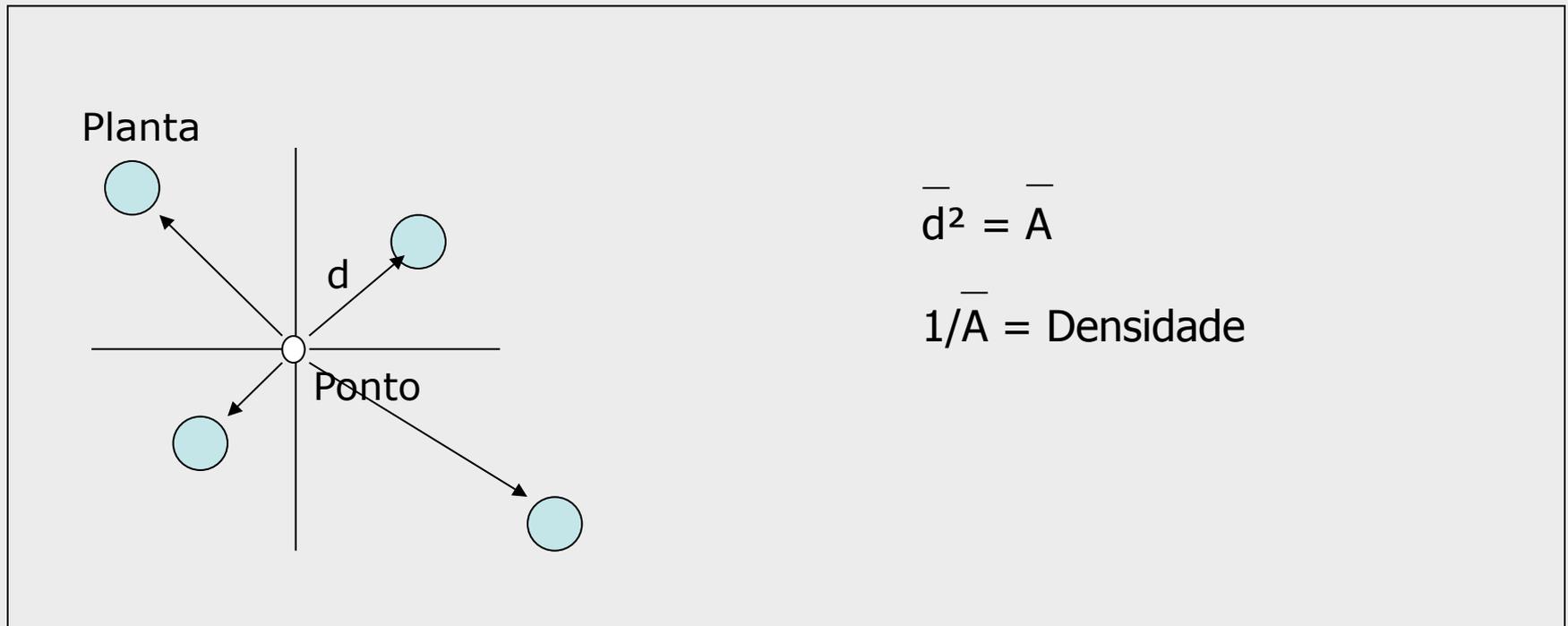
Número, tamanho e forma das unidades amostrais

Densidade mínima amostrada

Heterogeneidade dentro de cada unidade amostral (reduzir variância entre unidades) ou heterogeneidade entre unidades amostrais (aumentar variância entre unidades)?

Amostragem **sem área**:

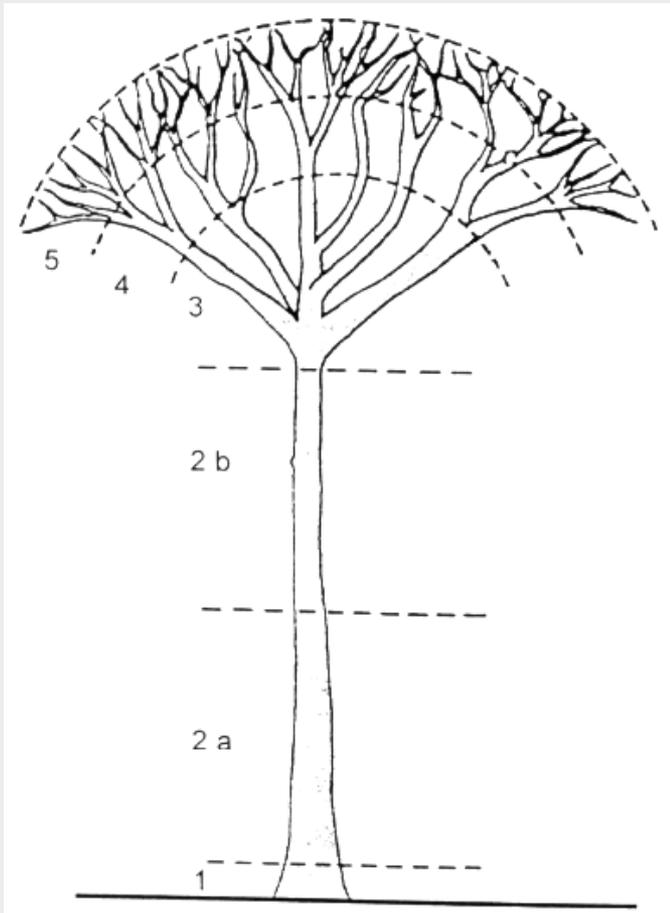
Uso de métodos de distância – Ponto-quadrante – Número de indivíduos fixo, área variável.



Amostragem **sem área**:

Uso de métodos baseados em indivíduos – localização nos indivíduos – Amostragem de epífitas.

Indivíduo



AMOSTRAGEM

Amostragem **sem área**:

Uso de técnicas de marcação e recaptura

$$N = \frac{M(n+1)}{R+1}$$

N = número de indivíduos

M = indivíduos capturados e marcados em t

R = indivíduos recapturados

n = indivíduos capturados em t+1

Alexander et al. (1997). Ecology 78: 1230-1237.

Planta perene com ciclo de vida > 25 anos

Floresce no início do outono

Acompanhamento por 4 anos



Asclepia meadii

Como amostrar?

A definição do sistema de amostragem

Amostragem **ao acaso**

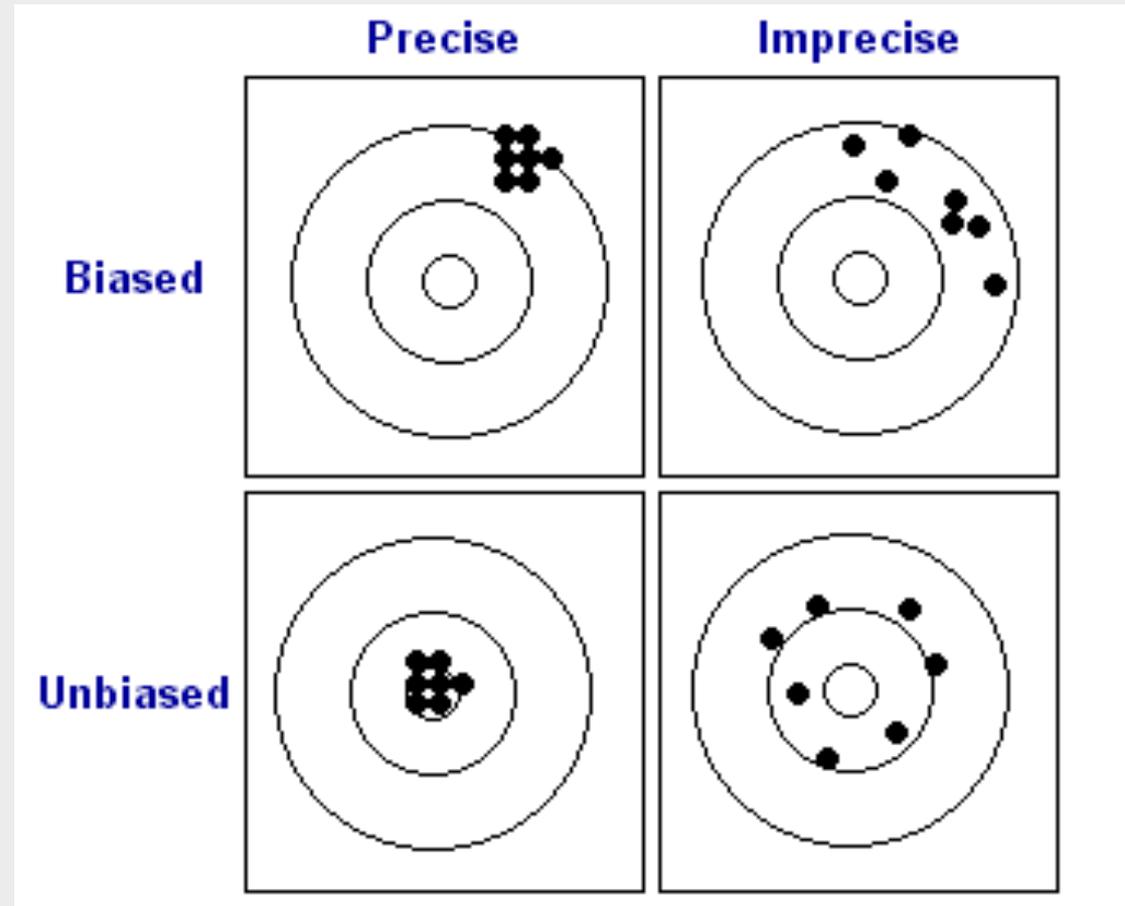
Amostragem **sistemática**

Amostragem **estratificada**

Usada quando a população é heterogênea, possuindo sub-populações homogêneas, que podem ser isoladas (estrato). Amostragem não estratificada é mais apropriada quando a população é homogênea.

Como amostrar?
A definição do método

Precisão X Acurácia



<http://www.stats.gla.ac.uk/steps/glossary/sampling.html>

Qual o tamanho da amostra?

Definição do nível de precisão ou margem de erro aceitável:

- quanto menor o erro aceitável, maior terá que ser a amostra.
- geralmente definido como 5%

Definição dos limites de confiança

- representa o grau de incerteza que temos.
- quanto maior o limite de confiança, maior terá que ser a amostra.
- geralmente definido como 95% ($t_{0,05} = 1,96$), 90% ($t_{0,10} = 1,645$) ou 99% ($t_{0,01} = 2,576$).

- assim, os limites de confiança da estimativa são determinados pela fórmula: $x \pm t_{\infty,\alpha} \frac{s}{\sqrt{n}}$ onde x = média, t = valor de t da distribuição de Student, s = desvio padrão, n = tamanho da amostra.

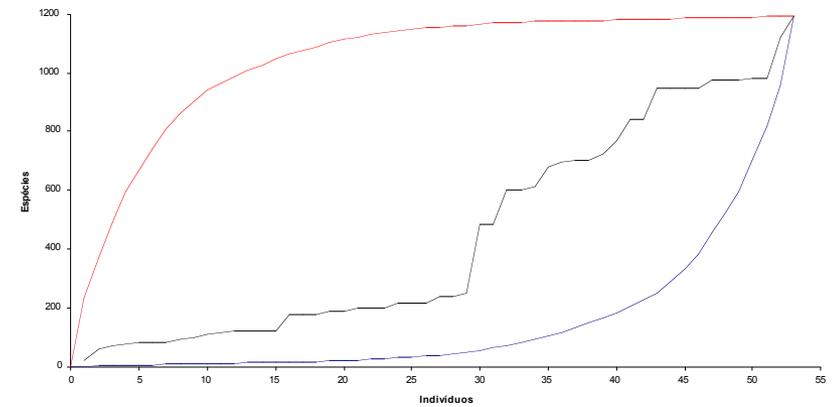
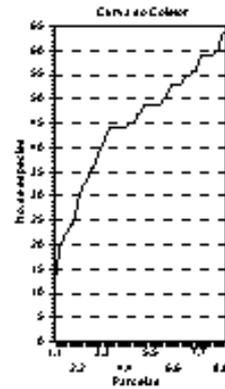
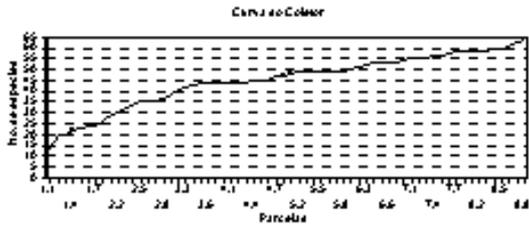
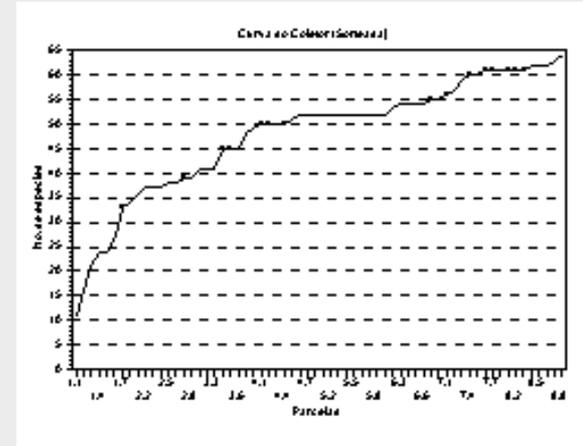
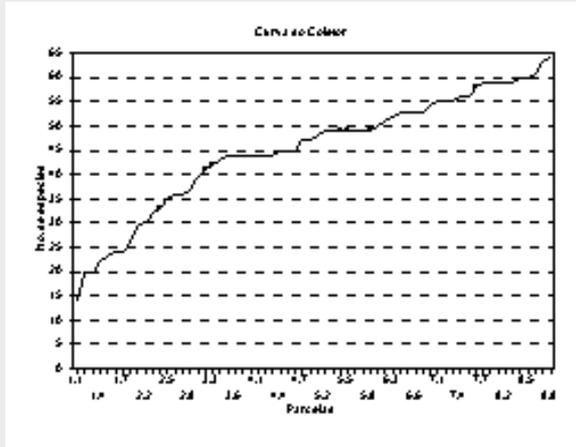
Número de amostras:

- existem várias formas de obter uma estimativa do número de amostras.
- exemplo: $n_0 = \frac{t_{\infty,\alpha}^2 \cdot s^2}{e^2}$ onde n_0 é o tamanho estimado da amostra e “e” é o erro aceitável.

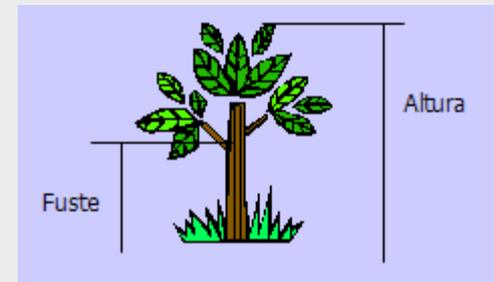
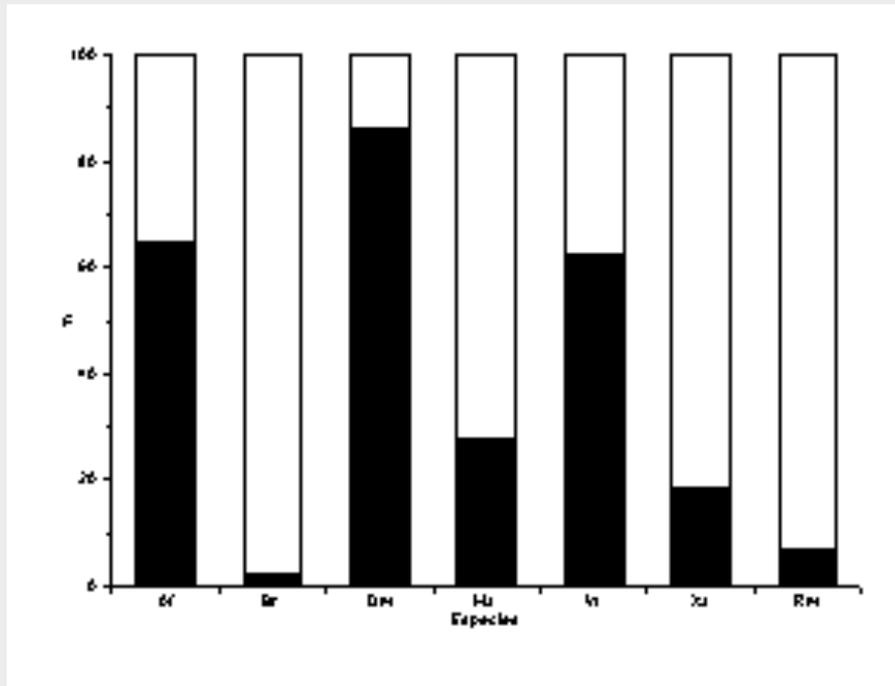
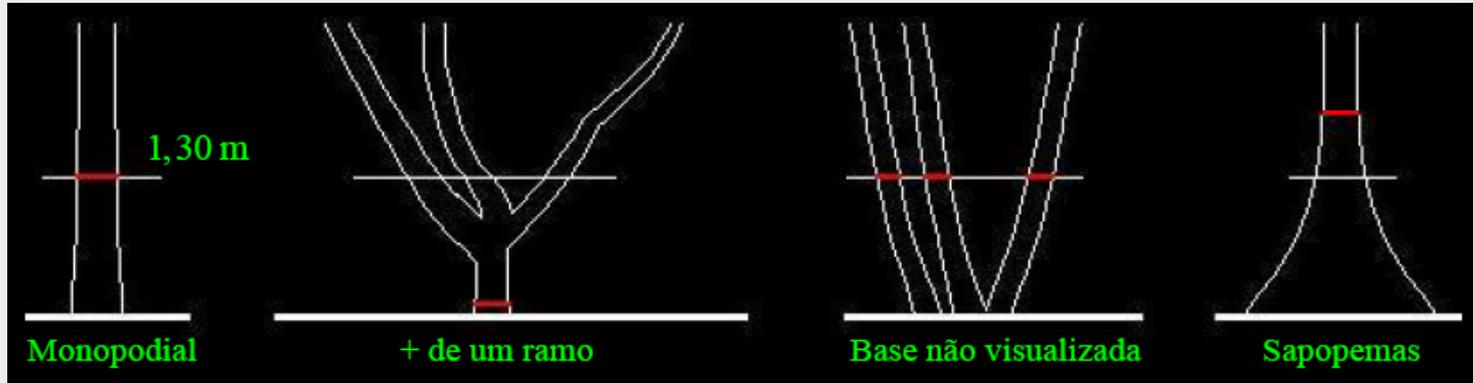
Problema: precisa ter uma boa estimativa de s da população (σ)

Suficiência amostral?

$$\text{Número de amostras} = x \pm t_{1-(1/2)\alpha} \frac{s}{\sqrt{n}}$$

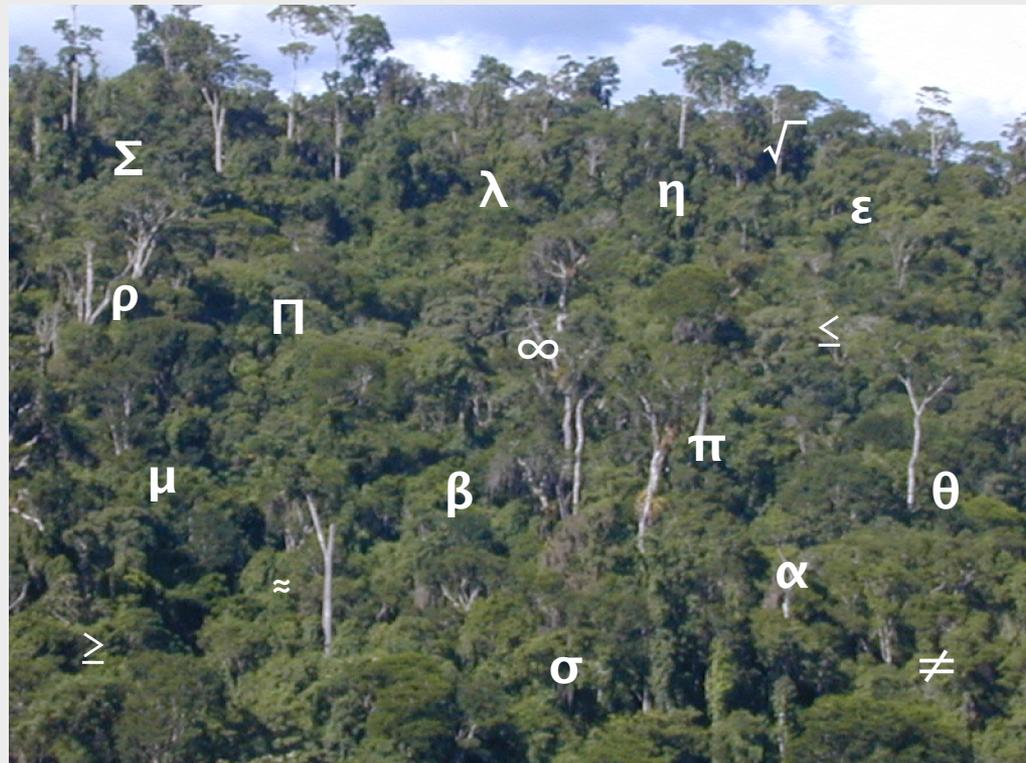


Critério de inclusão (tamanho mínimo do indivíduo)



Como analisar os dados obtidos?

Para que usar estatística?
Para que serve a estatística?



Como descrever resumida e objetivamente os resultados obtidos?
Como avaliar diferenças entre amostras?

Procedimentos de Análises

Teste adequado de hipótese – a importância da definição da hipótese a ser testada e da hipótese nula.

Hipótese estatística \neq Hipótese Biológica – Testes estatísticos são ferramentas a serem usadas pelo pesquisador.

O que significa a afirmativa de que duas médias de duas amostras diferem significativamente a $p = 0,01$?

O que significa erro tipo I e erro tipo II em estatística?

		<i>Statistical decision</i>	
		Null hypothesis	
<i>Actual situation</i>		Accepted	Rejected
Null hypothesis	True	Correct decision	Type I error
	False	Type II error	Correct decision

Estatística Paramétrica X Não-Paramétrica

Sobre médias, variâncias e desvios / medianas, amplitudes e quartis
X
Ausência de parâmetros

Médias

Média Aritmética $\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$

$\bar{Y} \rightarrow \mu$ Se:

1. Observações são feitas sobre indivíduos selecionados ao acaso
2. Observações são independentes entre si
3. Observações são feitas sobre uma população grande que pode ser descrita por uma variável aleatória normal

Média geométrica

$$GM_Y = \sqrt[n]{\prod_{i=1}^n Y_i}$$

$$\bar{\lambda} \neq (\lambda_1 + \lambda_2) / 2$$

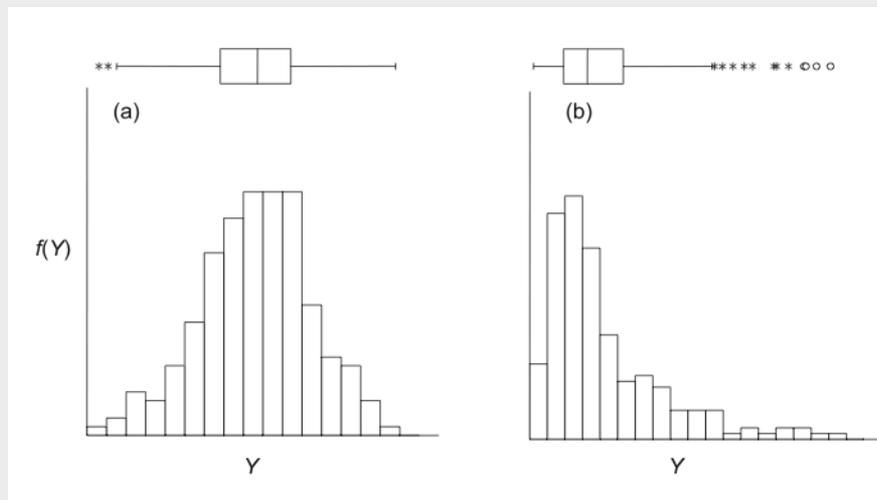
$$\bar{\lambda} = \sqrt{\lambda_1 \times \lambda_2}$$

Nem sempre a média aritmética é um bom descritor de tendência central.

Um exemplo é quando queremos estimar a taxa média de crescimento populacional ou o rendimento médio de uma aplicação.

Moda

Mediana

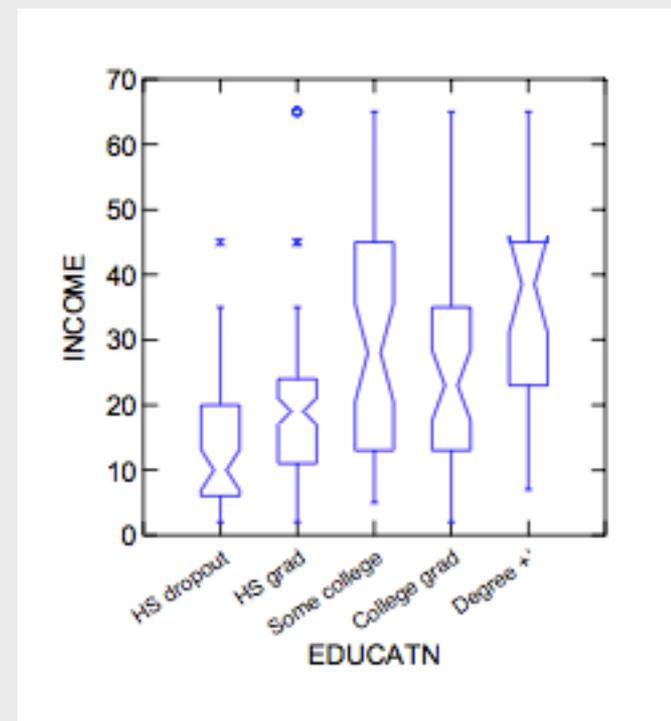
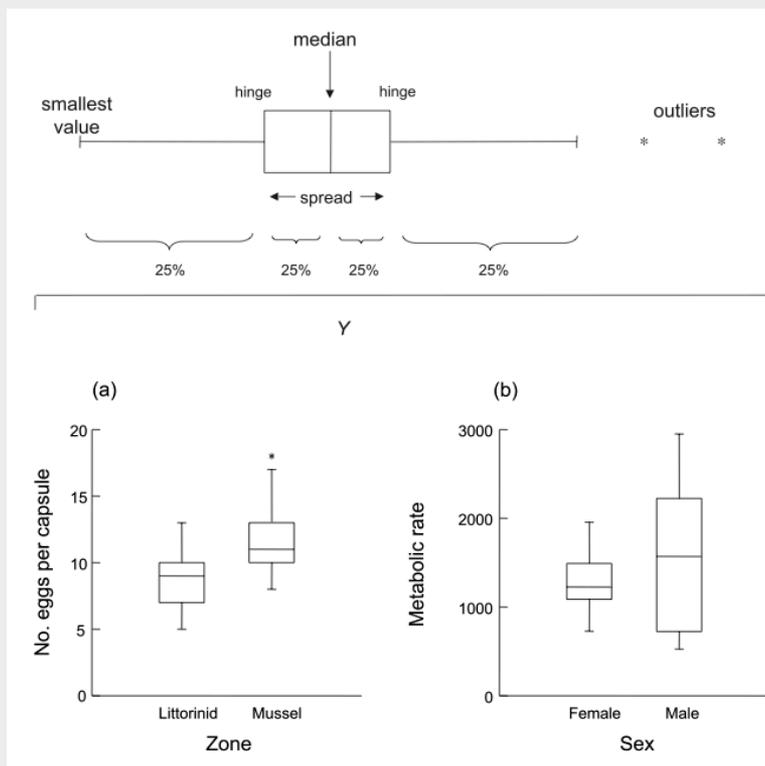


Variância

Desvio padrão

Erro padrão

Quartis



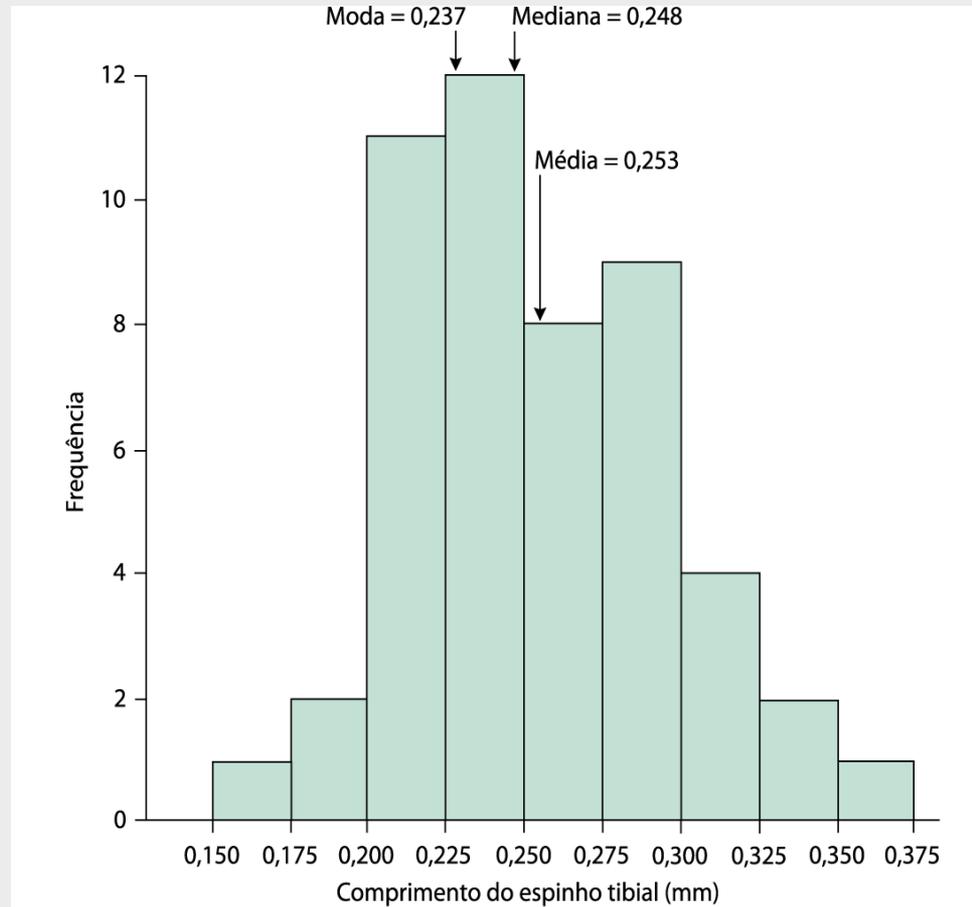


Figura 3.2 Histograma dos dados dos espinhos tibiais da Tabela 3.1 ($n = 50$) ilustrando a média aritmética, a mediana e a moda. A média é a expectativa dos dados, calculada como a média aritmética das medidas contínuas. A mediana é o ponto central do conjunto de observações ordenado. Metade de todas as observações é maior que a mediana e a outra metade é menor. A moda é a observação mais frequente.

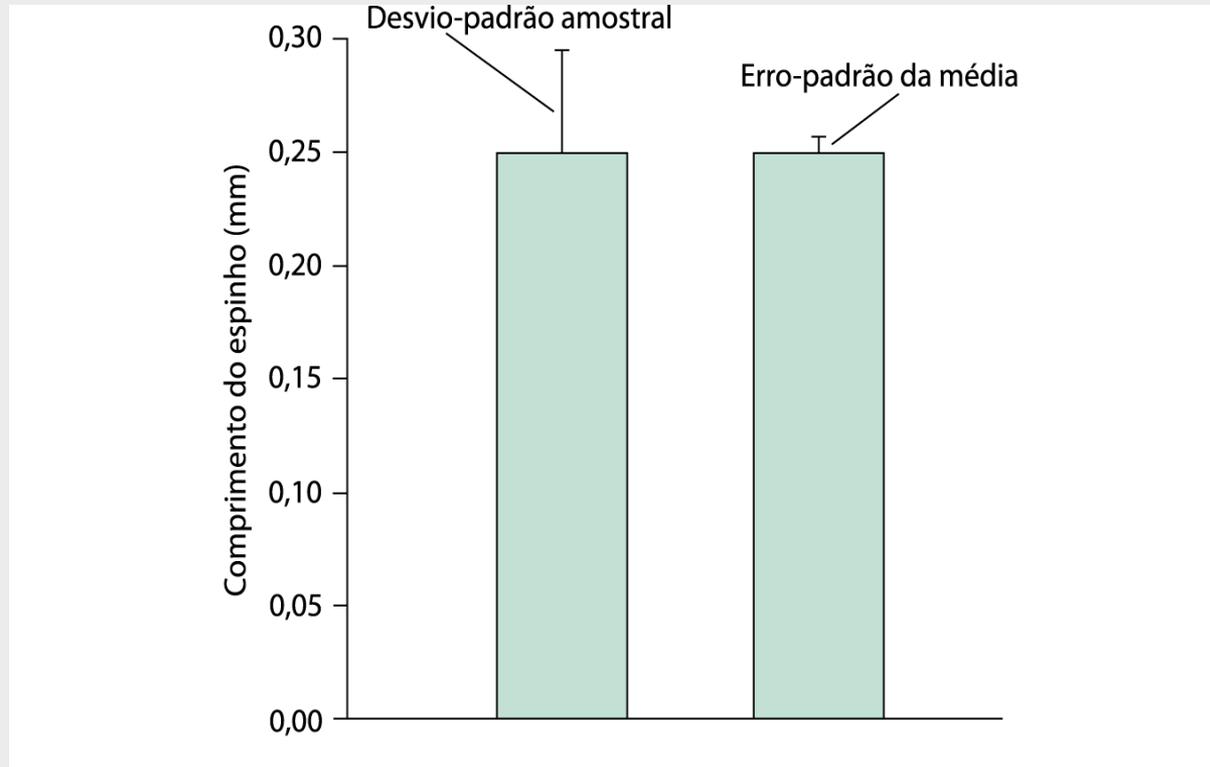
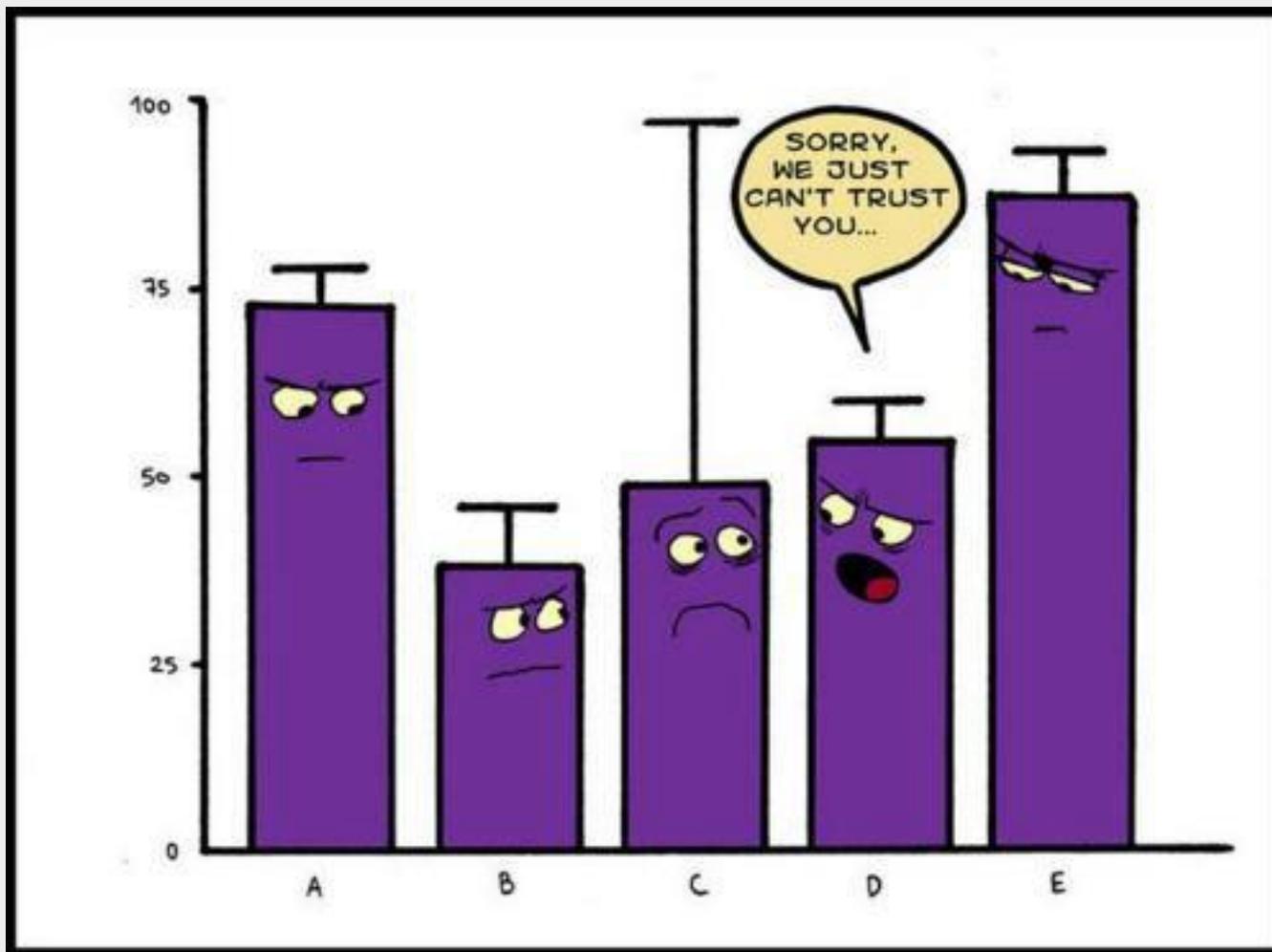


Figura 3.3 Diagrama de barras mostrando a média aritmética para os dados dos espinhos da Tabela 3.1 ($n = 50$), com barra de erro indicando o desvio-padrão amostral (barra esquerda) e o erro-padrão da média (barra direita). Enquanto o desvio-padrão mede a variabilidade de medidas individuais acerca da média, o erro-padrão mede a variabilidade da estimativa da média por si só. O erro-padrão se iguala ao desvio-padrão dividido por \sqrt{n} , então ele sempre será menor que o desvio-padrão, muitas vezes consideravelmente. As legendas e os textos de figuras devem sempre fornecer o tamanho amostral e indicar claramente quando o desvio-padrão ou o erro-padrão foi usado para construir as barras de erros.



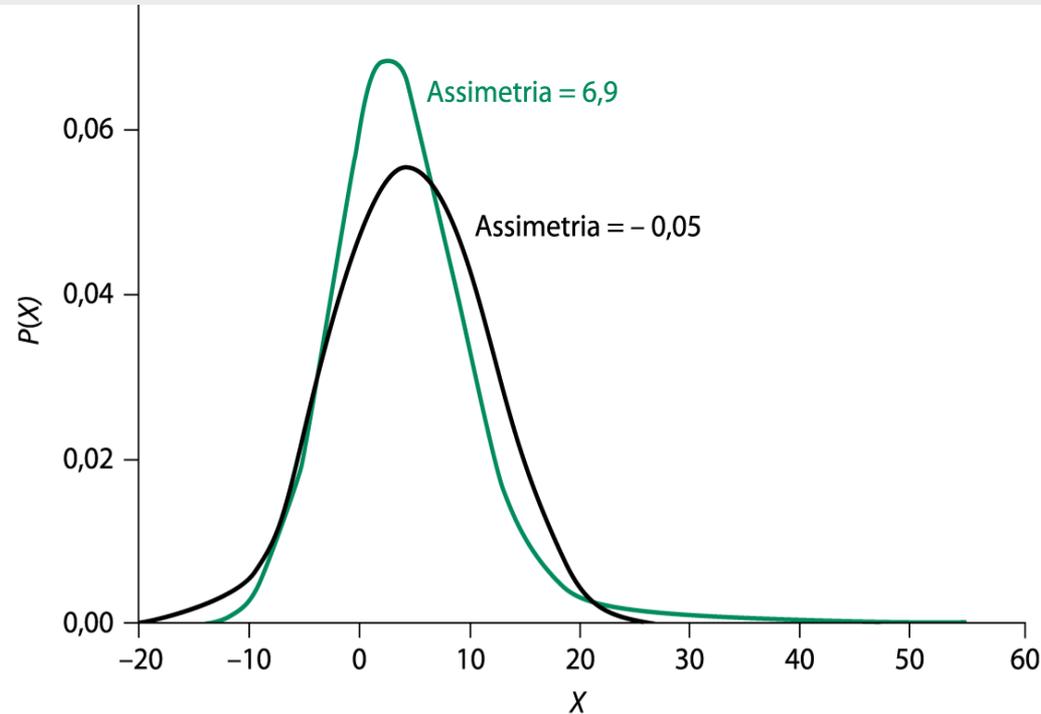


Figura 3.4 Distribuição contínua ilustrando a assimetria (g_1). A assimetria mede a extensão na qual a distribuição é assimétrica, com uma longa cauda de probabilidades à direita ou à esquerda. A curva verde é a distribuição log-normal ilustrada na Figura 2.8; ela tem assimetria positiva, com muito mais observações à direita da média do que à esquerda (uma longa cauda à direita), e uma medida de assimetria de 6,9. A curva preta representa uma amostra de 1.000 observações de uma variável aleatória normal com média e desvio-padrão idênticos aos da distribuição log-normal. Como esses dados foram tirados de distribuições normais simétricas, eles têm aproximadamente o mesmo número de observações em cada lado da média e a assimetria medida é aproximadamente 0.

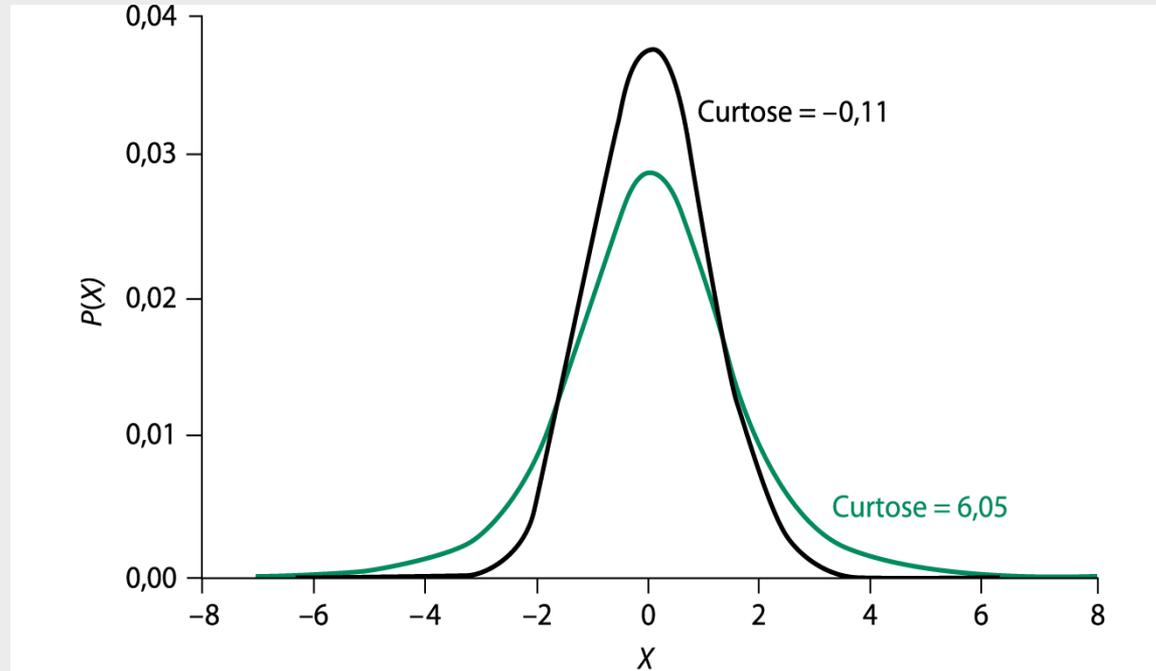


Figura 3.5 Distribuições ilustrando curtoses (g_2). A curtose mede a extensão na qual a distribuição é de cauda-pesada ou de cauda-leve, quando comparadas a uma distribuição normal padrão. Distribuições com caudas-pesadas são leptocúrticas e contêm relativamente mais área nas caudas da distribuição e menos no centro. Distribuições leptocúrticas possuem valores positivos para g_2 . Distribuições com caudas-leves são platicúrticas e contêm relativamente menos área nas caudas da distribuição e mais no centro. Distribuições platicúrticas têm valores negativos para g_2 . A curva preta representa uma amostra com 1.000 observações de uma variável aleatória normal com média 0 e desvio-padrão 1 ($X \sim N(0,1)$); sua curtose é próxima de 0. A curva verde é uma amostra de 1.000 observações de uma distribuição t com 3 graus de liberdade. A distribuição t é leptocúrtica e tem curtose positiva ($g_2 = 6,05$ neste exemplo).

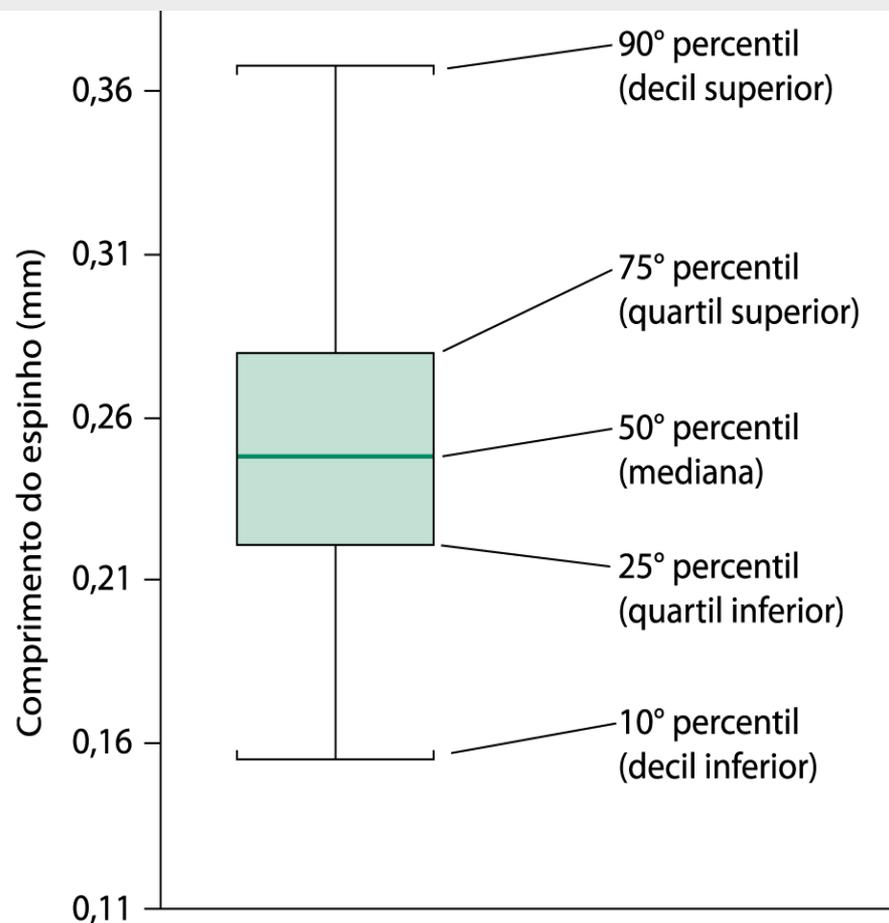


Figura 3.6 *Box plot* ilustrando os quantis dos dados da Tabela 3.1 ($n = 50$). A linha indica o 50° percentil (mediana) e a “caixa” engloba 50% dos dados, a partir do 25° até o 75° percentil. As linhas verticais se estendem do 10° ao 90° percentil.

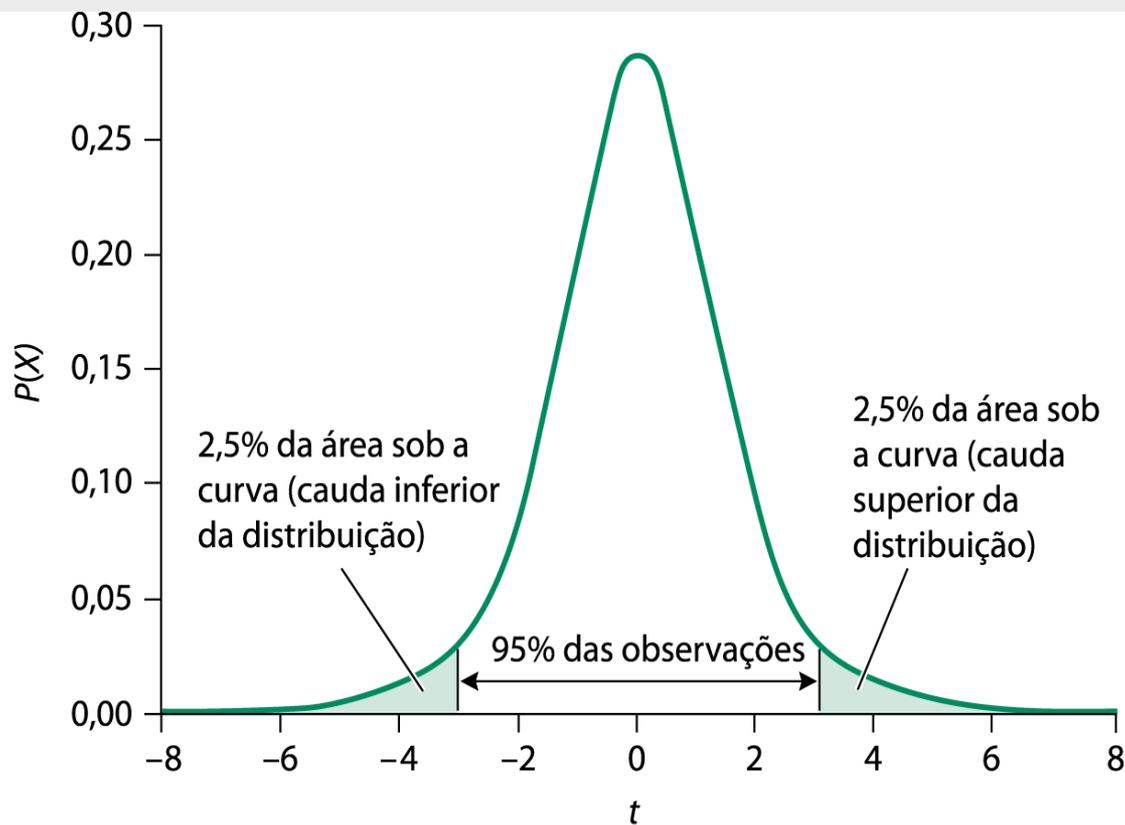
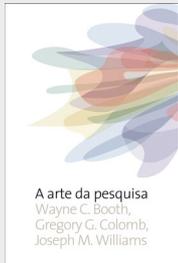


Figura 3.7 Distribuição-t ilustrando que 95% das observações, ou massa de probabilidade, caem dentro do percentil de $\pm 1,96$ desvio-padrão da média (média = 0). As duas caudas da distribuição contêm cada 2,5% das observações ou massa de probabilidade da distribuição. Elas somam 5% das observações, e a probabilidade $P = 0,05$ de que uma observação caia nessas caudas. Esta distribuição é idêntica à distribuição-t ilustrada na Figura 3.5.

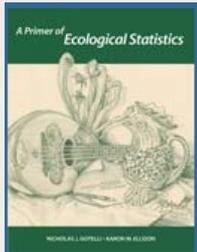
Referências Bibliográficas:



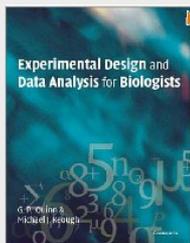
Barros Neto, B., Scarminio, I.S. & Bruns, R.E. 2002. Como fazer experimentos. Editora da UNICAMP



Booth, W.C., Colomb, G.C. & Williams, J.M. 2000. A arte da pesquisa. Martins Fontes



Gotelli, N.J. & Ellison, A.M. 2004. A Primer of Ecological Statistics. Sinauer

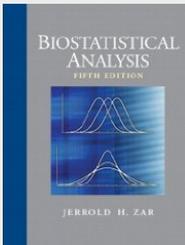


Quinn, G.P. & Keough, M.J. 2002. Experimental Design and Data Analysis for Biologists. Cambridge University Press

Referências Bibliográficas:



Sokal, R.R. & Rohlf, F.J. 1995. Biometry: The Principles and Practices of Statistics in Biological Research. 3rd ed. W.H. Freeman.



Zar, J.H. 2009. Biostatistical Analysis. 5th ed. Prentice Hall.

Bradshaw, A.D. 1987. Comparison - its scope and limits. New Phytologist 106 (supl.): 3-21

Schilling, A.C. & Batista, J.L.F. 2008. Curva de acumulação de espécies e suficiência amostral em florestas tropicais. Revista Brasileira de Botânica 31: 179-187

Sites:

<http://www.socialresearchmethods.net/kb/sampling.php>

<http://www2.chass.ncsu.edu/garson/PA765/sampling.htm>