Metodologia Ecológica

Aula 4

Mais ainda sobre a hipótese nula: erro tipo I e II

Poder!

Prof. Marcus Vinícius Vieira - Instituto de Biologia UFRJ

Percentis e *box plots*

- Percentis mais usados: 90, 75, 50 (mediana)
- Coeficiente de variação
- Coeficiente de dispersão

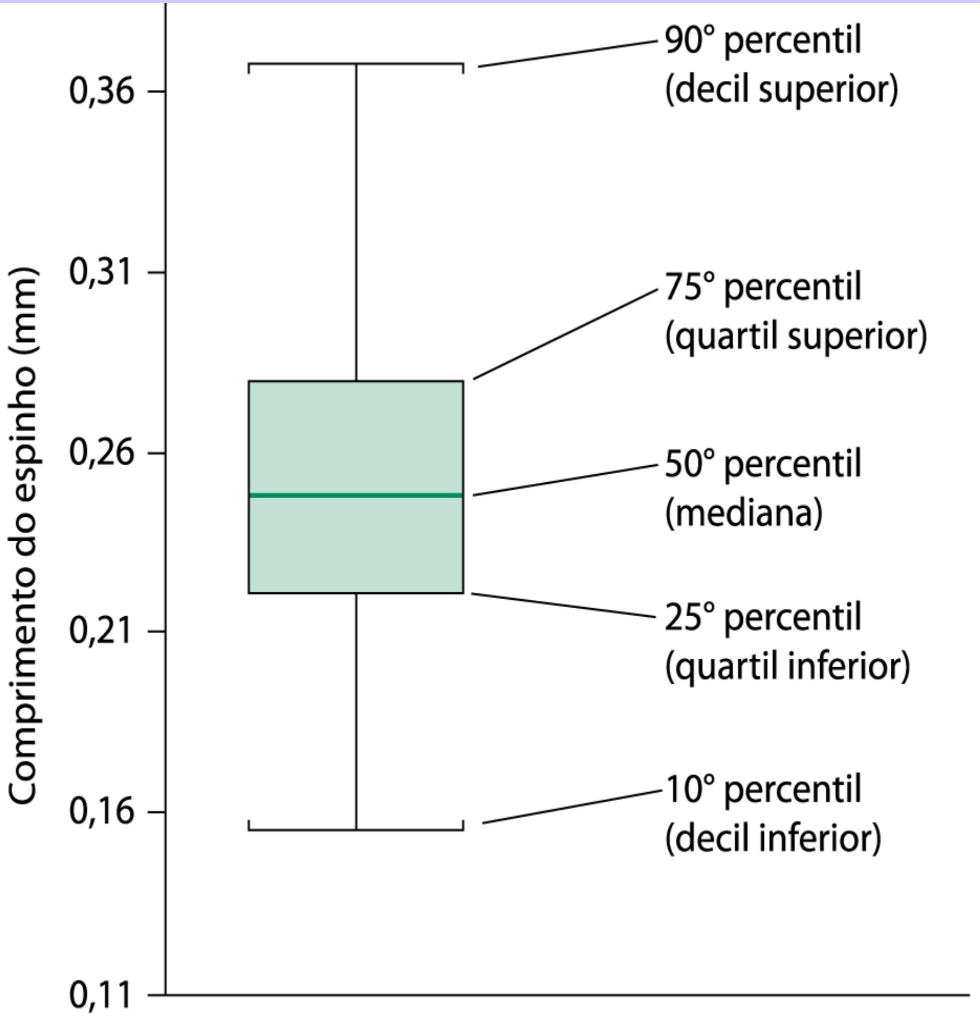

Figura 3.6 *Box plot* ilustrando os quantis dos dados da Tabela 3.1 ($n = 50$). A linha indica o 50° percentil (mediana) e a "caixa" engloba 50% dos dados, a partir do 25° até o 75° percentil. As linhas verticais se estendem do 10° ao 90° percentil.

Intervalos de confiança

P(média – 1,96ep ≤ μ ≤ média + 1,96ep) = 0,95

Intervalos de confiança generalizados

$$\bar{x} \pm t \bullet \frac{\sigma}{\sqrt{n}}$$

Distribuição *t* (de Student)

Intervalos de confiança

- Interpretação errônea: “Há uma chance de 95% que o valor real da população se encontre dentro do intervalo”
- Se o parâmetro estimado é fixo, isto não é possível: ou está ou não está no intervalo
- Correta: “Se o experimento for repetido 100 vezes, em 95 delas o intervalo conterá o valor real da população; em 5 delas não”.
- Mas e se o parâmetro varia na população?

Erro Tipo I : α

- Probabilidade de rejeitar Ho quando verdadeira
- Probabilidade de falso negativo
- Para calcular depende:

  basta especificar Ho

- Teste conservador

Erro Tipo II : β

- Probalidade de não rejeitar Ho quando falsa, ou
  probabilidade de falso positivo, β
- Para calcular depende:
  - Da hipótese alternativa
  - Do tamanho do efeito que se pretende detectar
  - Do tamanho da amostra
  - Do delineamento experimental
- Poder: probabilidade de rejeitar corretamente Ho quando falsa: 1- β
- Princípio de precaução em monitoramento ambiental

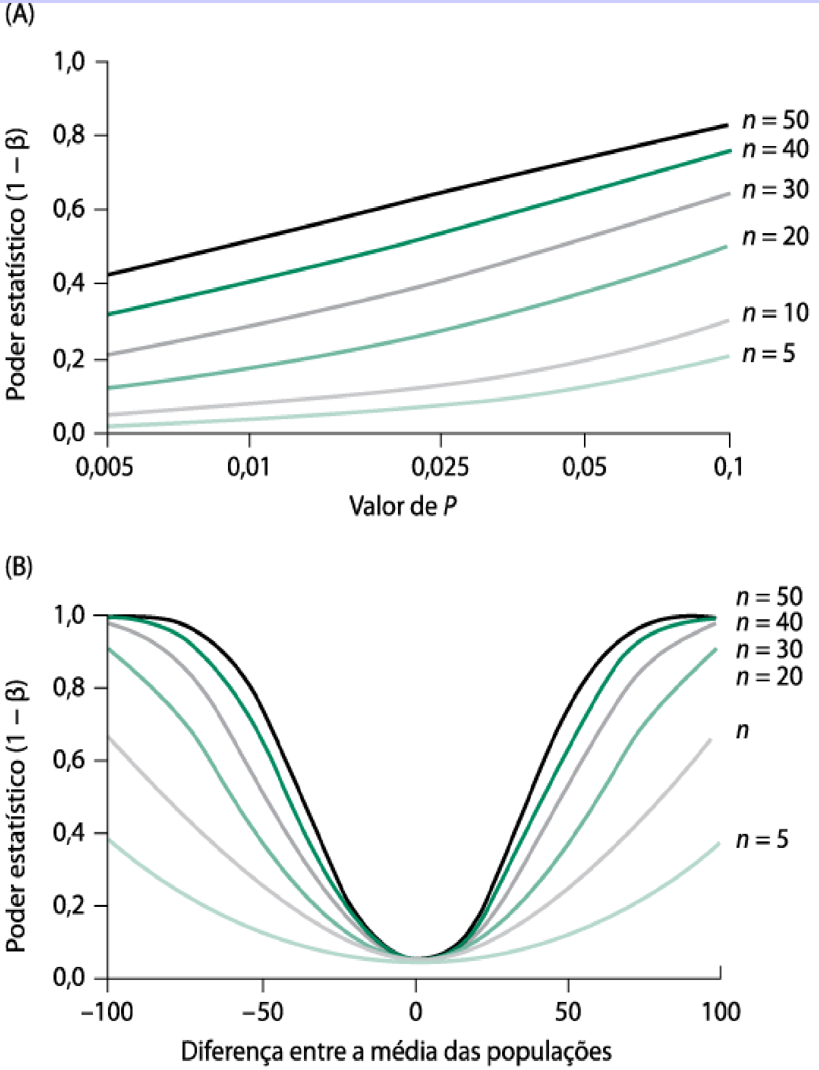

Figura 4.5 Relação entre poder estatístico, valor de *P* e tamanho do efeito observável em função do tamanho amostral. (A) O valor de *P* é a probabilidade de incorretamente rejeitar uma hipótese nula verdadeira, enquanto o poder estatístico é a probabilidade de corretamente rejeitar uma hipótese nula falsa. O resultado geral propõe que, quanto menor o valor de *P* usado para rejeitar a hipótese nula, menor o poder estatístico de corretamente detectar um efeito do tratamento. A um dado valor de *P*, o poder estatístico é maior quando o tamanho amostral é maior. (B) Quanto menor o tamanho do efeito observável do tratamento (i. e., quanto menor a diferença entre o grupo-tratamento e o grupo-controle), maior é o tamanho amostral necessário a um bom poder estatístico para detectar o efeito do tratamento.[21]

Momentos da distribuição normal

- Momento centrado (*Central Moment*)
- 1o. : desvio padrão
- 2o.: variância
- 3o.: assimetria (*skewness*)

- 4o.: curtose (*kurtosis*)

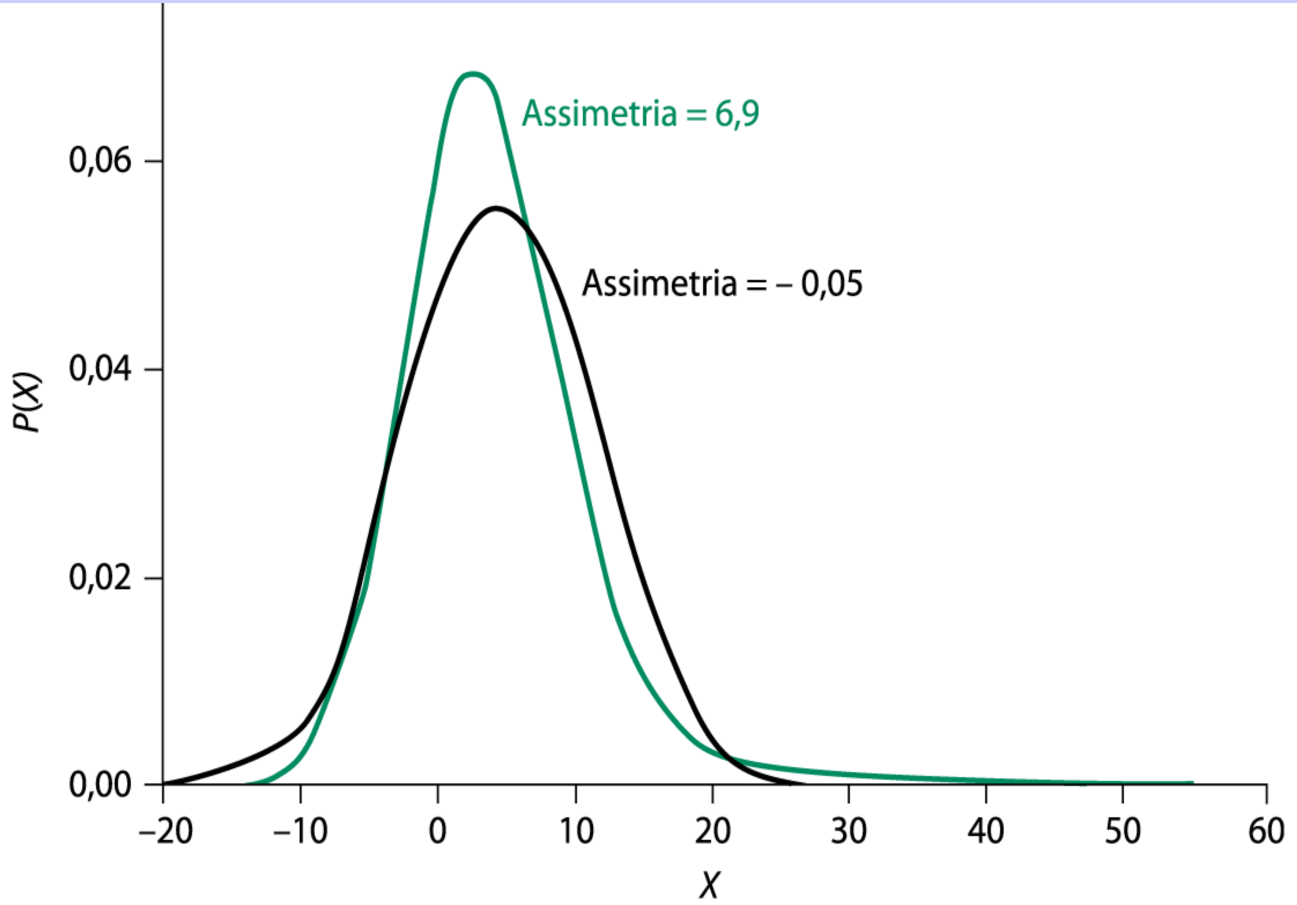

Figura 3.4 Distribuição contínua ilustrando a assimetria (g_1). A assimetria mede a extensão na qual a distribuição é assimétrica, com uma longa cauda de probabilidades à direita ou à esquerda. A curva verde é a distribuição log-normal ilustrada na Figura 2.8; ela tem assimetria positiva, com muito mais 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.

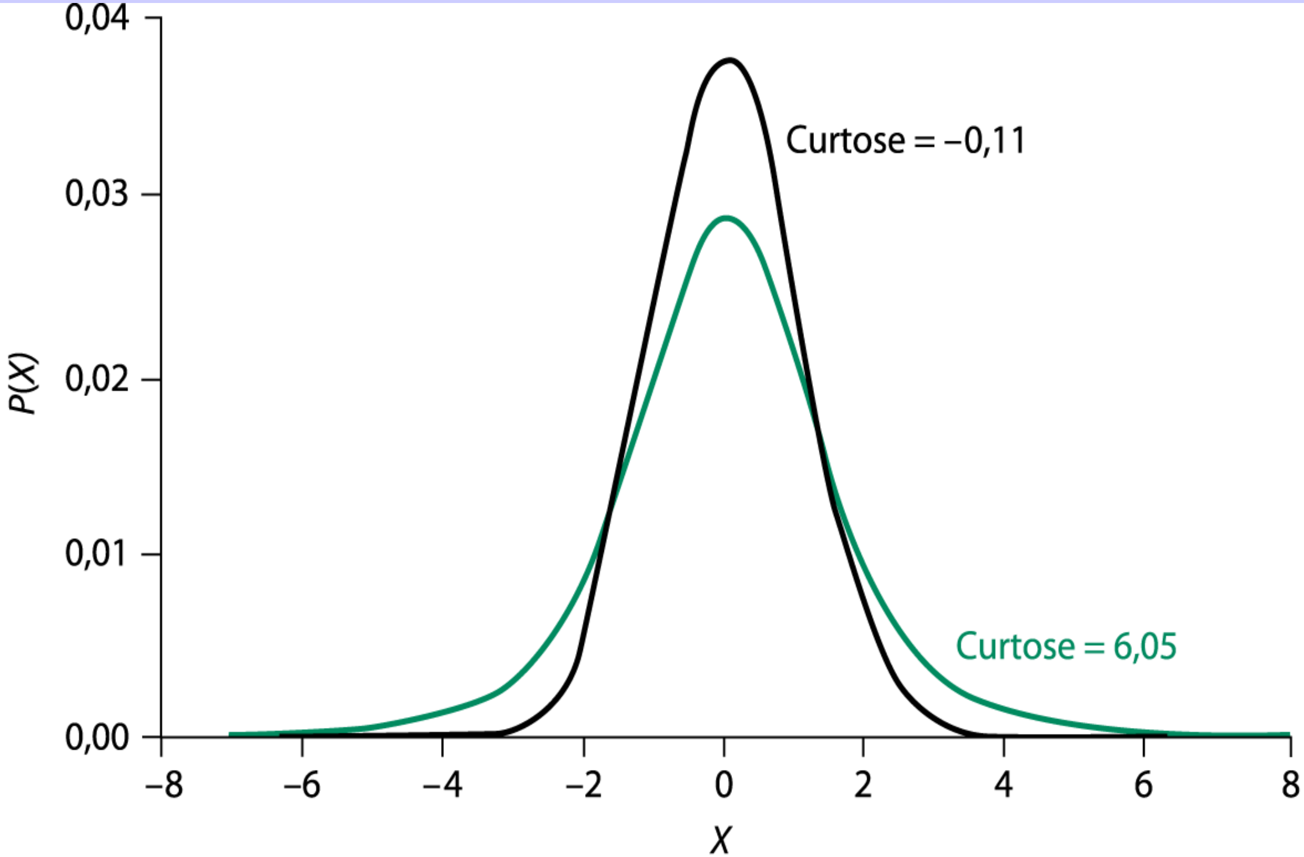

Figura 3.5 Distribuições ilustrando curtoses (g_2). A curtose mede a extensão na qual a distribuição é de cauda-pesada ou de cauda-leve, quando comparadas a uma distribuição normal padrão. Distribuições com caudas-pesadas são leptocúrticas e contêm relativamente mais área nas caudas da distribuição e 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).

Testes de hipótese nula comuns

- Diferença entre médias de duas amostras:
  - Teste *t*
  - Razão *F*
  - Premissas

Testes *t*

para testar Ho de diferença entre duas médias

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}}$$

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^{\,2}}{n_1} + \frac{s_2^{\,2}}{n_2}}$$

$$t = \frac{Diferença\_entre\_médias}{Erro\_padrão\_da\_diferença\_entre\_médias}$$

O significado de uma “variável” ser dita “aleatória”

- Definimos o espaço amostral Ω = {(captura), (fuga)} para os resultados possíveis dos evento visita de um inseto a uma planta carnívora.

- Para analisar estatisticamente esta informação, precisamos associar a cada elemento do espaço amostral um número.

- Precisamos de uma **função** que atribua um valor a cada elemento do espaço amostral.

- A esta função é dado o nome **variável aleatória**, normalmente representada por letras maiúsculas, como *X*.

Variáveis aleatórias

- Assim a variável aleatória na verdade é uma função cujos valores não são aleatórios (!?).

- É dita "aleatória" porque os valores de *X* dependem do resultado do experimento, que tem um grau de incerteza nos resultados, de "aleatoriedade".

- Uma variável associa a cada valor possível uma certa probabilidade.

Tipos de variáveis aleatórias

- Variáveis aleatórias podem ser discretas ou contínuas.
  - Variáveis aleatórias discretas têm um número finito de valores dentro de um intervalo.
    - Exemplos: $X \sim \mathrm{Bin}(n, p)$
  - Variáveis aleatórias contínuas podem ter um número infinito de valores dentro de um intervalo.
    - Exemplos

Tipos de variáveis aleatórias

- Variáveis aleatórias podem ser discretas ou contínuas.
  - Variáveis aleatórias discretas têm um número finito de valores dentro de um intervalo.
    - Exemplos: $X \sim \mathrm{Bin}(n, p)$
    - $X \sim \mathrm{Poisson}(\lambda)$
  - Variáveis aleatórias contínuas podem ter um número infinito de valores dentro de um intervalo.
    - Exemplos

Variáveis aleatórias discretas

- X ~ Bin(n, p) $$P(X) = \frac{n!}{X!(n-X)!} p^X (1-p)^{n-X}$$

- X ~ Poisson(λ) $$P(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

$$P(4\,plântulas) = \frac{0{,}75^4}{4!} e^{-0{,}75} = 0{,}0062$$

Variáveis aleatórias contínuas

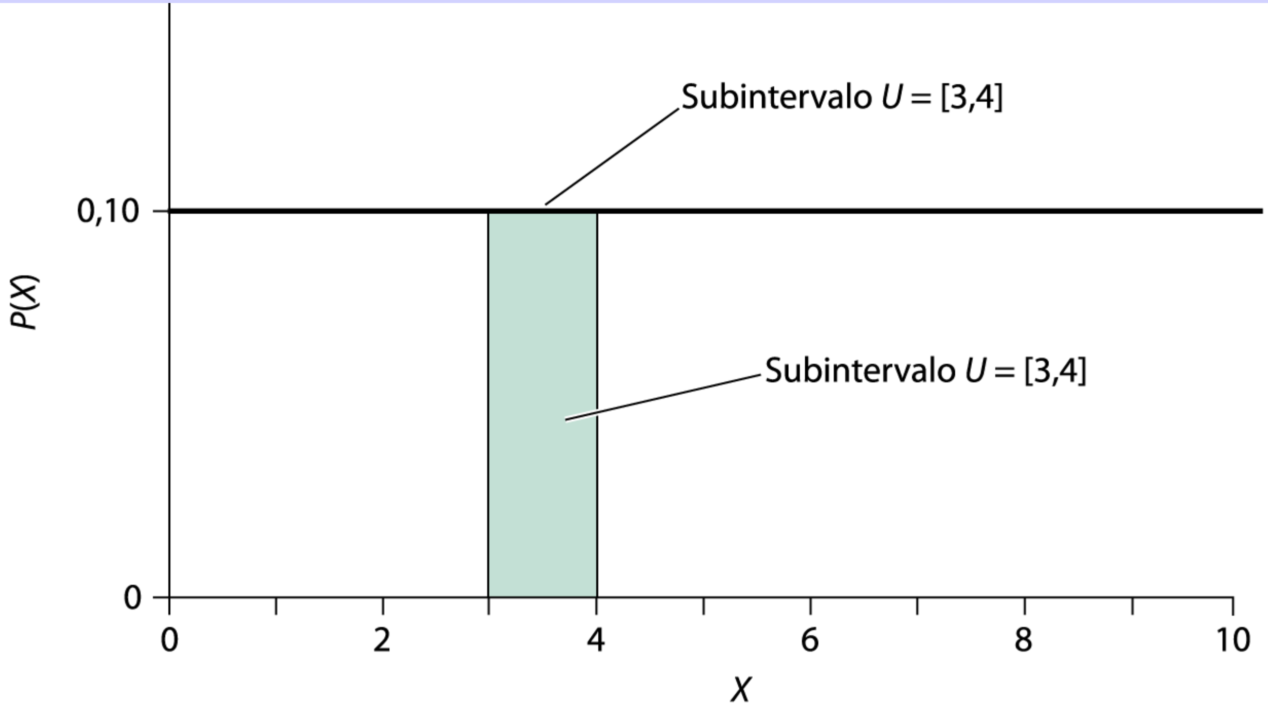

Figura 2.4 Distribuição uniforme com intervalo [0,10]. Em uma distribuição uniforme contínua, a probabilidade de um evento ocorrer em um subintervalo particular depende da área relativa do subintervalo; ela é a mesma independente de onde o subintervalo está inserido dentro dos limites da distribuição. Por exemplo, se a distribuição é delimitada por 0 e 10, a probabilidade de que um evento ocorra no subintervalo [3,4] é a área relativa delimitada por aquele subintervalo, que neste caso é 0,10. A probabilidade é a mesma para qualquer outro subintervalo com o mesmo tamanho, como [1,2] ou [4,5]. Se o subintervalo escolhido é maior, a probabilidade de um evento ocorrer naquele subintervalo será proporcionalmente maior. Por exemplo, a probabilidade de um evento ocorrer no subintervalo [3,5] é de 0,20 (desde que 2 dentre as 10 unidades do intervalo sejam transpassadas), e é de 0,6 para o subintervalo [2,8].

Variáveis aleatórias contínuas

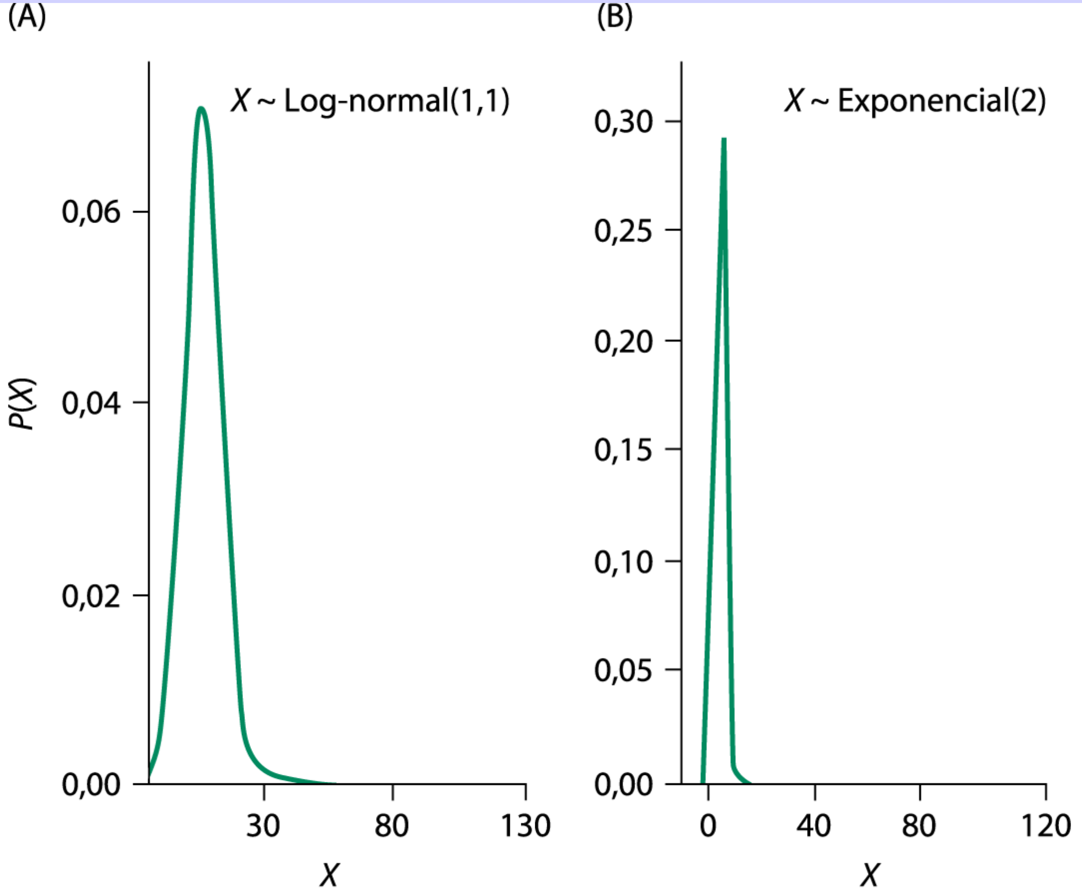

Figura 2.8 Distribuições log-normal e exponencial se ajustam a certos tipos de dados ecológicos, como a distribuição de abundâncias de espécies e distâncias de dispersão de sementes. (A) A distribuição log-normal é descrita por dois parâmetros, média e variância, ambos são 1 neste exemplo. (B) A distribuição exponencial é descrita por um único parâmetro *b*, que é 2 nesse exemplo. Ver a Tabela 2.4 para as equações usadas com as distribuições log-normal e exponencial. Ambas as distribuições, log-normal e exponencial, são assimétricas, com uma longa cauda a direita que desvia a distribuição à direita.

Variáveis aleatórias contínuas

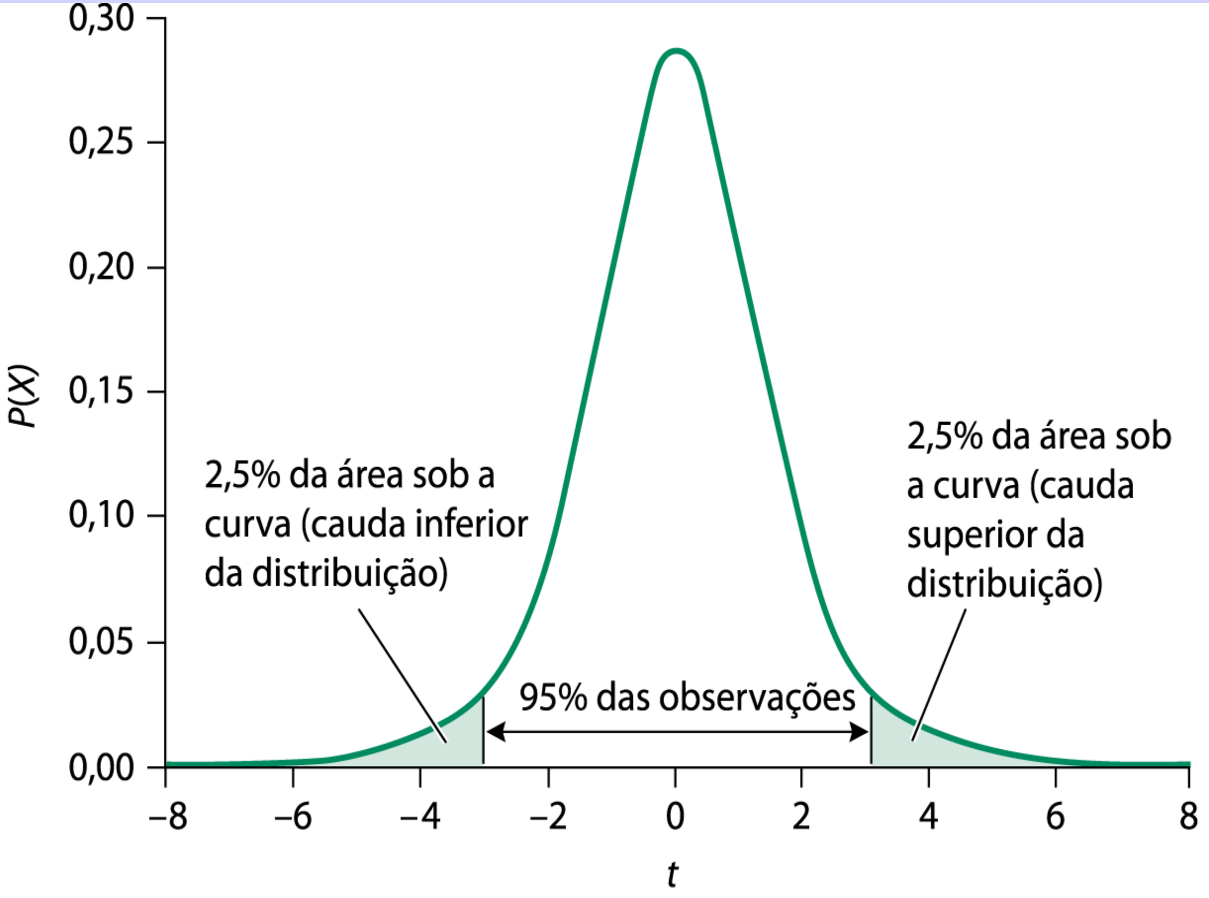

Figura 3.7 Distribuição-*t* ilustrando que 95% das observações, ou massa de probabilidade, caem dentro do percentil de ±1,96 desvio-padrão da média (média = 0). As duas caudas da distribuição contêm cada 2,5% das observações ou massa de probabilidade da distribuição. Elas somam 5% das observações, e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

Variáveis aleatórias contínuas

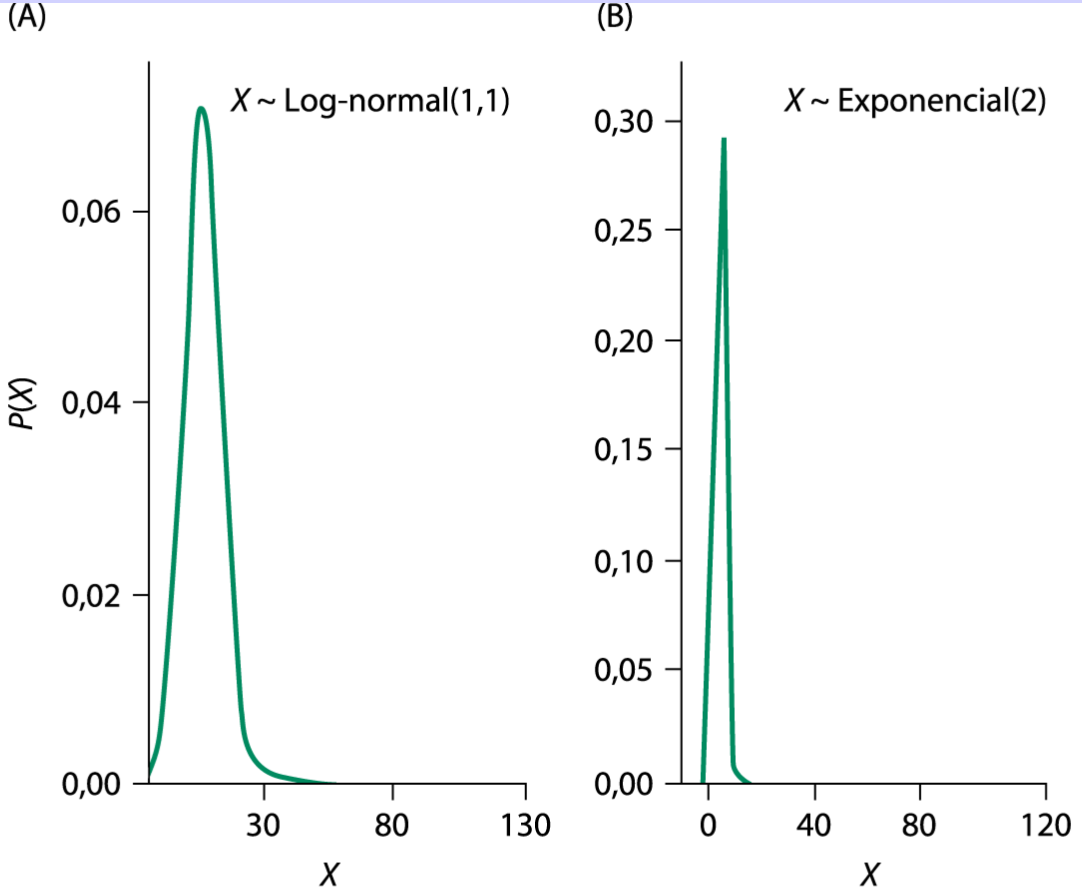

Figura 2.8 Distribuições log-normal e exponencial se ajustam a certos tipos de dados ecológicos, como a distribuição de abundâncias de espécies e distâncias de dispersão de sementes. (A) A distribuição log-normal é descrita por dois parâmetros, média e variância, ambos são 1 neste exemplo. (B) A distribuição exponencial é descrita por um único parâmetro *b*, que é 2 nesse exemplo. Ver a Tabela 2.4 para as equações usadas com as distribuições log-normal e exponencial. Ambas as distribuições, log-normal e exponencial, são assimétricas, com uma longa cauda a direita que desvia a distribuição à direita.

Estimativas

- Em estatística frequentista (ou assintótica)
  - Parâmetros na população são fixos
  - Amostrando-se a população repetidamente, infinitamente, a estimativa dos parâmetros irá convergir (na assíntota) nos valores reais dos parâmetros
- É uma premissa realista?