Metodologia Ecológica

Aula 2

Estimativas de tendência e variação
Como quantificar sua confiança em um número (probabilidades)
Intervalos de confiança de estimativas

Prof. Marcus Vinícius Vieira - Instituto de Biologia UFRJ

Medidas de tendência central e variação

- Variável aleatória: *Y*
- Cada observação da variável *Y*: Y_i
- Tamanho da amostra: *n* ou *N*
- Média aritmética: $\bar{Y}$
- Parâmetros desconhecidos, como valores e variância esperados: letras gregas

μ = ?

σ^2 = ?

Medidas de tendência central

- Média aritmética $\overline{Y} = \dfrac{\sum_{i=1}^{n} Y_i}{n}$

- Valor esperado $E(X) = \sum_{i=1}^{n} Y_i p_i$

- Relação entre valor esperado e média aritmética
- (fórmula com limite)

Condições para que a média aritmética seja um estimador sem viés de μ

- Observações são feitas em indivíduos escolhidos aleatoriamente
- Observações na amostra são independentes uma da outra
- Observações foram tiradas de uma população que pode ser descrita por uma variável normal aleatória

Lei de Grandes Números

- Teorema fundamental da Estatística
- Com o aumento do tamanho de amostra, *n*, a média de Y_n se aproxima do valor esperado de *Y*

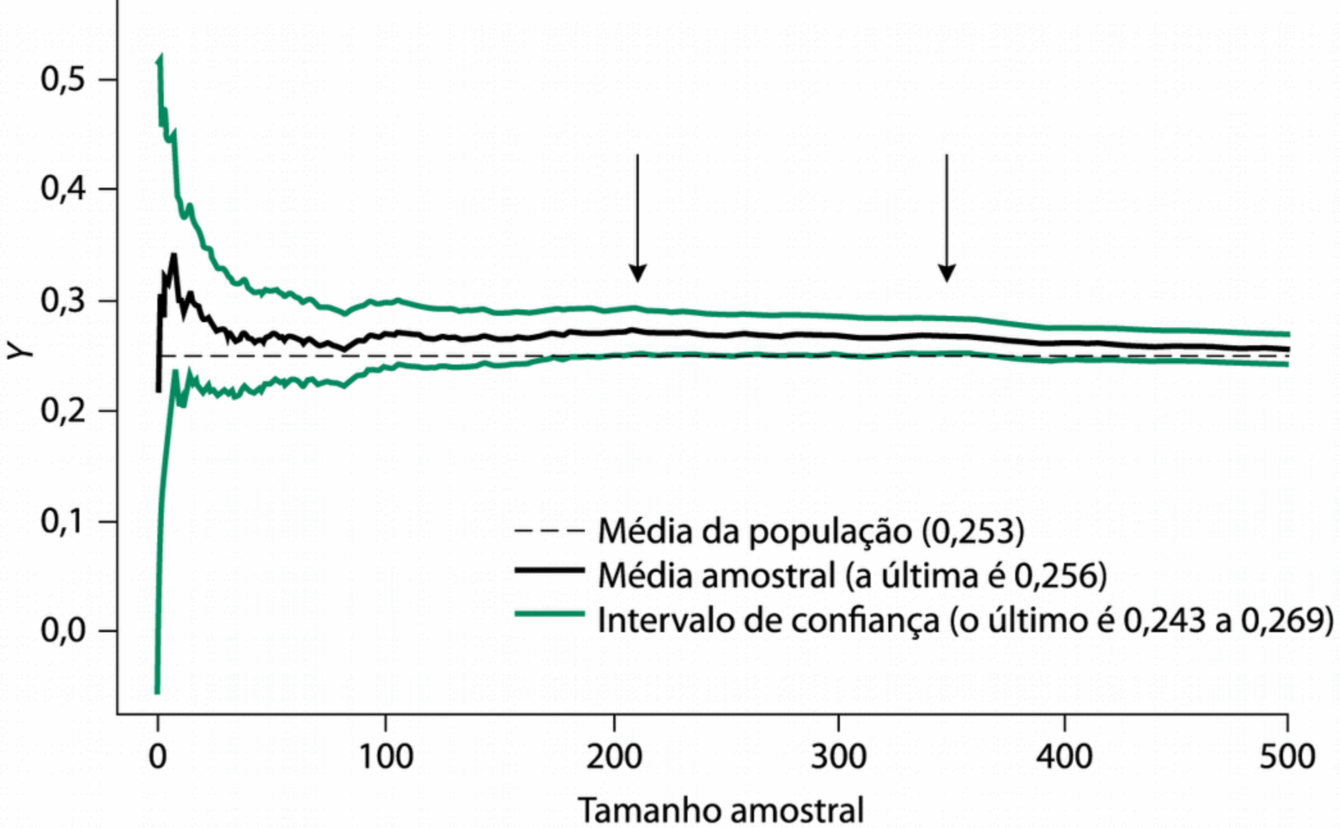

Figura 3.1 Ilustração da Lei dos Grandes Números e a construção de intervalos de confiança usando os dados dos espinhos tibiais de aranhas da Tabela 3.1. A média da população (0,253) é indicada pela linha pontilhada. A média amostral para amostras de tamanhos crescentes (*n*) é indicada pela linha sólida central e ilustra a Lei dos Grandes Números: conforme o tamanho amostral aumenta, a média amostral se aproxima da verdadeira média da população. As linhas sólidas superior e a inferior ilustram o intervalo de confiança de 95% ao redor da média. A largura do intervalo de confiança decresce conforme o tamanho amostral aumenta. Intervalos de confiança de 95% construídos dessa forma devem conter a verdadeira média da população. Note, contudo, que existem amostras (entre as setas) para as quais o intervalo de confiança não inclui a verdadeira média da população. As curvas foram construídas usando algoritmos e códigos do S-Plus publicados por Blume e Royal (2003).

Média geométrica

$$MG_Y = e^{\left[\overline{Z} = \frac{1}{n}\sum_{i=1}^{n}\ln(Y_i)\right]}$$

- Relação com a distribuição log-normal

$$MG_Y = \sqrt[n]{Y_1 Y_2 \ldots Y_n} \qquad MG_Y = \sqrt[n]{\prod_{i=1}^{n} Y_i}$$

Medidas de tendência central

- Média aritmética $\overline{Y} = \dfrac{\sum_{i=1}^{n} Y_i}{n}$

- Valor esperado $E(X) = \sum_{i=1}^{n} Y_i p_i$

- Relação entre valor esperado e média aritmética
- (fórmula com limite)

Exemplo: taxa de crescimento

- População inicial de 1000 indivíduos crescendo a uma taxa anual de 1,4

  (λ ou R em Ecologia de Populações)
- 1o. ano: 1000 x 1,4 = 1.400 indivíduos
- 2o. ano: 1.400 x 1,4 = 1.960
- 3o. ano: 1.960 x 1,4 = 2.744
- Média aritmética: 2.034,6
- Média geométrica: 1.960

Outro exemplo

- O território de um animal é um quadrado de 2 km de lado.
- A cada manhã o animal percorre os limites do território, mas ...
  - Começa a passos lentos, a 1 km/h
  - No segundo lado chega a 2 km/h,
  - no terceiro a 4 km/h,
  - e no último está cansado e reduz para 1 km/h.
- Qual a velocidade média?

Outro exemplo

- (1 + 2 + 4 + 1) / 4 = 8 / 4 = 2 km/h
- Mas, 2km a
  - 1 km/h = 2 h
  - 2 km/h = 1 h
  - 4 km/h = 0,5 h
  - 1 km/h = 2 h -> 5,5 h para percorrer 8 km
- 8 / 5,5 = 1,4545 km/h

Média harmônica

$$MH_Y = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{Y_i}}$$

- O inverso da média dos inversos (???)
- No nosso exemplo:
  (1 + 2 + 4 + 1) / 4 = 8 / 4 = 2 km/h

  1 / [(1/1 + ½ + 1/3 + 1/4)/4]
  = 1 / [2,75/4]
  = 1/0,6875 = 1,4545

- *MG* é ligeiramente menor que *MA*
- *MH* é ligeiramente menor que *MG*

Média harmônica

$$MH_Y = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{Y_i}}$$

- Exemplo Gotelli & Ellison: tamanho populacional efetivo (N_e)
- A média harmônica dá mais peso às observações de menor valor e menor peso às de maior valor
- É indicada então quando observações de menor valor influenciam mais o valor esperado que observações de maior valor
- É o caso do tamanho efetivo da população

Mediana e Moda

- Mediana: valor que divide uma distribuição ao meio
- Isto é, o valor observado com um mesmo número de observações acima e abaixo
- Para um número par de observações, é o ponto intermediário entre as observações $n/2$ e $(n/2)+1$
- Moda: valor mais freqüente

Por quê a média aritmética é mais usada?

- Teorema do Limite Central
- médias de grandes amostras de variáveis aleatórias seguem uma distribuição normal (= de Gauss)
- Mesmo que a distribuição de cada amostra não seja normal

Mediana ou Moda

- Distribuições assimétricas (uma das caudas mais comprida que a outra)
- Distribuições que não se encaixam em distribuições de probabilidade conhecidas
- Presença de valores extremos (influenciam muito qualquer uma das médias)

Medidas de tendência central e variação

- Variável aleatória: *Y*
- Cada observação da variável *Y*: Y_i
- Tamanho da amostra: *n*
- Média aritmética: $\bar{Y}$
- Parâmetros desconhecidos, como valores e variância esperados: letras gregas

$\forall$ μ = ?

$\forall$ σ^2 = ?

Medidas de variação

- Já vimos a variância esperada como $E[Y-E(Y)]^2$
- Para uma variável aleatória é representada por σ^2
- Valor desconhecido, tem que ser estimado da amostra, s^2

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

- Soma dos quadrados: $$SQ_Y = \sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

Viés (*bias*)

- É uma medida de acurácia, diferente de precisão
- A estimativa da variância

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \bar{Y} \right)^2$$

  resulta numa estimativa enviesada (*biased*) da variância em amostras "pequenas"
- A estimativa sem viés (*unbiased*) é

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(Y_i - \bar{Y} \right)^2$$

Medidas de variação

- Já vimos a variância esperada como $E[Y\text{-}E(Y)]^2$
- Para uma variável aleatória é representada por σ^2
- Valor desconhecido, tem que ser estimado da amostra, s^2

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

- Soma dos quadrados:

$$SQ_Y = \sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

Viés (*bias*)

- É uma medida de acurácia, diferente de precisão
- A estimativa da variância

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

  resulta numa estimativa enviesada (*biased*) da variância em amostras “pequenas”
- A estimativa sem viés (*unbiased*) é

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

Incerteza

- Qual deve ser a relação entre incerteza em uma estimativa e sua variância?
- E qual deve ser a relação com *n*?
- Incerteza α s^2 / n
- E as unidades? Como trazer esta medida para a mesma unidade da média?

Erro padrão $$SE_{\bar{y}} = \sqrt{\frac{s^2}{n}} = \frac{s}{\sqrt{n}}$$

- Desvio padrão de uma série de médias
- Quando apresentar *se* ou *sd*?
- Se a amostra é representativa de toda a população, *se*
- Ex.: observações naturais, em larga escala, com amostras grandes
- Se não é possível generalizar a partir da amostra, *sd*
- Ex. Experimentos em pequena escala, controlados e com pequeno *n*
- Desde que se apresente *n*, é possível converter de um para o outro

Intervalos de confiança

- Quando σ (desvio padrão da população) é conhecido, OU para tamanhos de amostra grandes (> 200):

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

- Quando σ é estimado da amostra por *s*, OU *N* < 200

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

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?

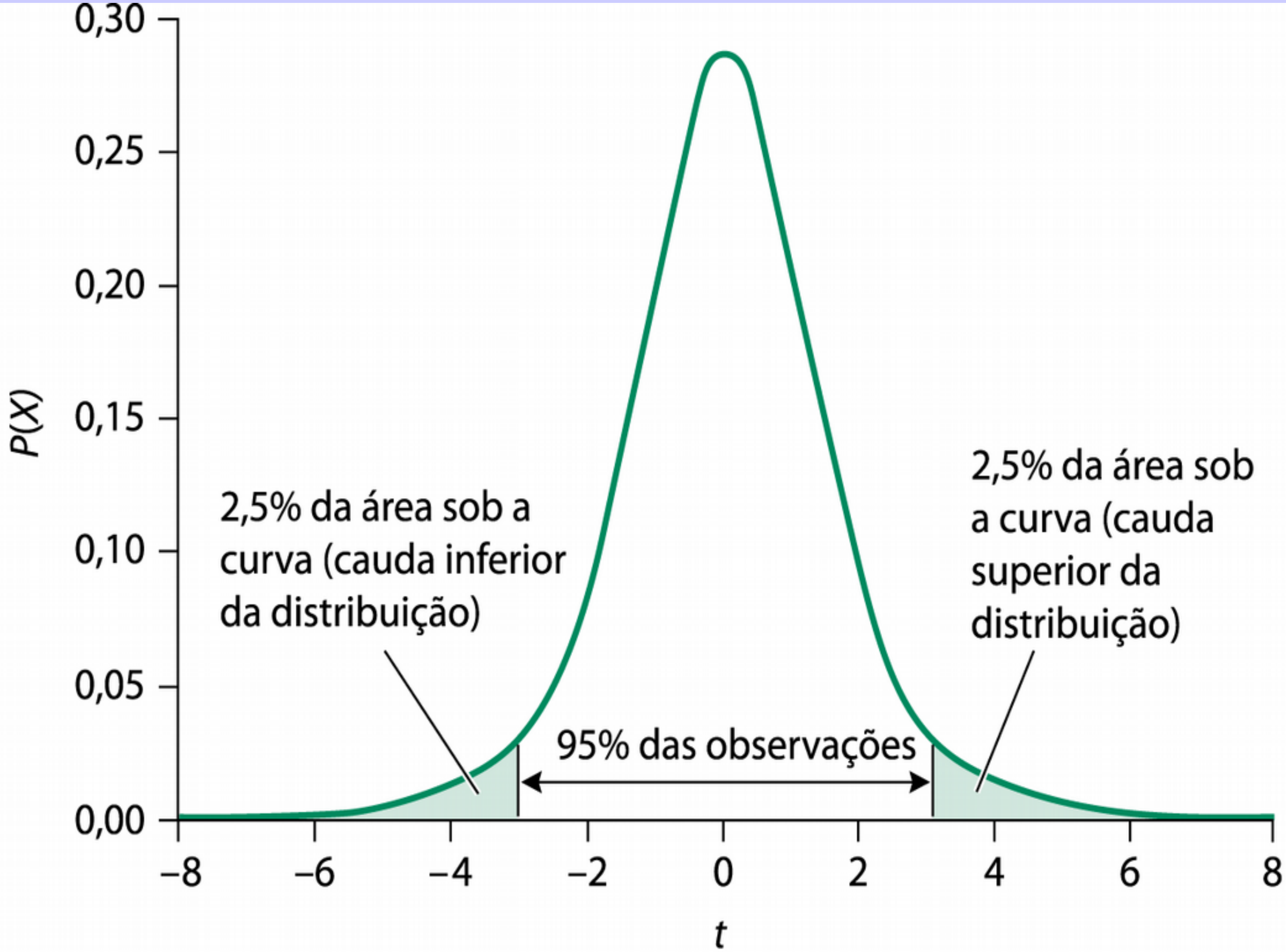

Figura 3.7 Distribuição-*t* ilustrando que 95% das observações, ou massa de probabilidade, caem dentro do percentil de ±1,96 desvio-padrão da média (média = 0). As duas caudas da distribuição contêm cada 2,5% das observações ou massa de probabilidade da distribuição. Elas somam 5% das observações, e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