# LCE0216
# Introdução à Bioestatística Florestal

## 8. Distribuições amostrais


Profa. Dra. Clarice Garcia Borges Demétrio

Monitor: Eduardo E. R. Junior

Escola Superior de Agricultura "Luiz de Queiroz"
Universidade de São Paulo


Piracicaba, 15 de maio de 2018

# Distribuições amostrais

- Estimadores são funções de variáveis aleatórias, portanto também são variáveis aleatórias;
- Sendo assim, **estimadores também são variáveis aleatórias** que seguem algum modelo de probabilidades;
- Nesse tópico, serão apresentadas as distribuições de alguns dos principais estimadores.

**Exemplo**: Seja uma população composta por quatro árvores dada na tabela a seguir:

| Árvore | Diâmetro (cm) |
|---|---|
| A | 8,0 |
| B | 20,0 |
| C | 24,0 |
| D | 27,0 |

A proporção de árvores com diâmetro inferior a 20cm,

$$\pi = \frac{1}{4} = 0,25.$$

O diâmetro médio ($\mu$):

$$\mu = \frac{\sum_{i=1}^{4} x_i}{4} = 19,75 \text{ cm}.$$

**Exemplo**: Seja uma população composta por quatro árvores dada na tabela a seguir:

| Árvore | Diâmetro (cm) |
|---|---|
| A | 8,0 |
| B | 20,0 |
| C | 24,0 |
| D | 27,0 |

A variância ($\sigma^2$):

$$\sigma^2 = \frac{\sum_{i=1}^{4}(x_i - \mu)^2}{4} = \frac{208,75}{4} = 52,1875 \text{ cm}^2.$$

O desvio padrão ($\sigma$):

$$\sigma = \sqrt{\sigma^2} = \sqrt{52,1875} = 7,2241 \text{ cm}.$$

# Distribuições amostrais

| Amostra | Elementos | $\hat{\pi}$ | $\hat{\mu}$ | $\hat{\sigma^2}$ |
|---|---|---|---|---|
| 1 | A,B | 0,50 | 14,0 | 72,0 |
| 2 | A,C | 0,50 | 16,0 | 128,0 |
| 3 | A,D | 0,50 | 17,5 | 180,5 |
| 4 | B,C | 0,00 | 22,0 | 8,0 |
| 5 | B,D | 0,00 | 23,5 | 24,5 |
| 6 | C,D | 0,00 | 25,5 | 4,5 |
| 7 | B,A | 0,50 | 14,0 | 72,0 |
| 8 | C,A | 0,50 | 16,0 | 128,5 |
| 9 | D,A | 0,50 | 17,5 | 180,0 |
| 10 | C,B | 0,00 | 22,0 | 8,0 |
| 11 | D,B | 0,00 | 23,5 | 24,5 |
| 12 | D,C | 0,00 | 25,5 | 4,5 |
| 13 | A,A | 1,00 | 8,0 | 0,0 |
| 14 | B,B | 0,00 | 20,0 | 0,0 |
| 15 | C,C | 0,00 | 24,0 | 0,0 |
| 16 | D,D | 0,00 | 27,0 | 0,0 |

# Distribuição amostral do estimador $P$

Vamos supor que uma árvore com menos de 20 cm de diâmetro não seja interessante para o mercado.

- Existe apenas uma árvore na população com determinada característica $\Rightarrow \pi = 1/4 = 0,25$.
- Estimar tal proporção observando árvores dessa população

$$\Downarrow$$

Observar uma amostra de tamanho dois, com reposição

$$\Downarrow$$

Estimar $\pi$ por meio da estatística

$$P = \frac{\text{número de casos favoráveis (sucessos)}}{\text{tamanho da amostra}}$$

# Distribuição amostral do estimador $P$

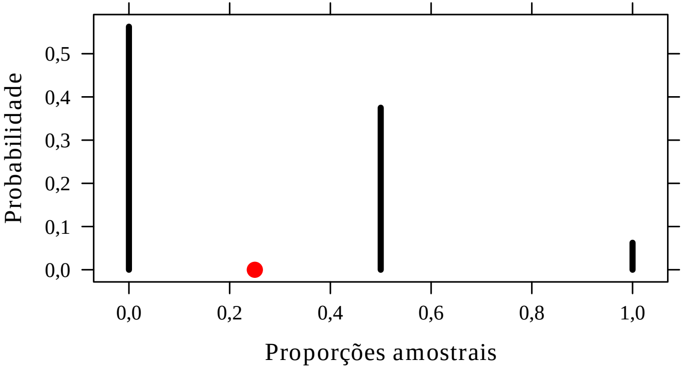
Perguntas:

1. Quais proporções amostrais podem ser obtidas?
2. Qual a probabilidade associada a cada uma?
3. Qual a forma da distribuição das proporções amostrais?
4. Qual a média da distribuição amostral dessas proporções?
5. Qual a variância da distribuição amostral dessas proporções?

# Distribuição amostral da estatística $P$

Distribuição amostral da proporção:

| $y_i$ | 0 | 1 | 2 |
|---|---|---|---|
| $\hat{p} = y_i/2$ | 0 | 0,5 | 1 |
| $\mathrm{P}(P = \hat{p})$ | 9/16=0,5625 | 6/16=0,3750 | 1/16=0,0625 |



# Distribuição amostral da estatística $P$

Distribuição amostral da proporção:

| $y_i$ | 0 | 1 | 2 |
|---|---|---|---|
| $\hat{p} = y_i/2$ | 0 | 0,5 | 1 |
| $\mathrm{P}(P = \hat{p})$ | 9/16=0,5625 | 6/16=0,3750 | 1/16=0,0625 |

- Média:

$$\mu_P = 0 \times 0,5625 + 0,50 \times 0,3750 + 1 \times 0,0625 = 0,25 = \pi$$

- Variância:

$$\begin{aligned}\sigma_P^2 &= (0 - 0,25)^2 \times 0,5625 + (0,50 - 0,25)^2 \times 0,3750 + \\ &\quad + (1 - 0,25)^2 \times 0,0625 \\ &= 0,09375 = \pi(1 - \pi)/n\end{aligned}$$

# Distribuição amostral da estatística $P$

$Y$: número de árvores com diâmetro inferior a 20 cm

$$\text{Se } Y \sim Bin(n, \pi).$$

Então,

$$\mu_Y = \mathrm{E}(Y) = n\pi \qquad \text{e} \qquad \sigma_Y^2 = \mathrm{Var}(Y) = n\pi(1-\pi).$$

Seja $P =$ proporção das árvores com diâmetro inferior a 20 cm. A distribuição amostral de $P$ poderá ser aproximada por uma distribuição normal com parâmetros:

$$\mu_P = \mathrm{E}(P) = \tfrac{\mu_Y}{n} = \pi \qquad \text{e} \qquad \sigma_P^2 = \mathrm{Var}(P) = \tfrac{\sigma_y}{n} = \tfrac{\pi(1-\pi)}{n}.$$

$$P \sim N(\mu_P, \sigma_P^2) \quad \Longrightarrow \quad Z = \frac{p - \mu_P}{\sigma_P} \sim N(0, 1)$$

**Observação**: Quando são utilizadas amostras sem reposição, deve-se fazer uma correção na variância.

**Exemplo**: Uma proporção de 37% dos visitantes de um parque favorecem a cobrança de taxas de entrada. Uma amostra aleatória de 200 visitantes foi tomada.

(a) Qual é a probabilidade que na amostra de 200 visitantes pelo menos 40% favoreçam a cobrança de taxas?

(b) Qual é a probabilidade que na amostra de 200 visitantes, a proporção dos que favorecem a cobrança de taxas fique entre 35% e 39%?

(c) Uma nova amostra de 10 visitantes foi tomada. Qual a probabilidade de que pelo menos 50% dos visitantes na amostra favoreçam a cobrança de taxas? É válido utilizar o mesmo método utilizado anteriormente? Qual método deveria ser utilizado nesse caso?

# Distribuição amostral da média

## Estimador $\bar{X}$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Sem dúvidas, o estimador mais utilizado na estatística aplicada.

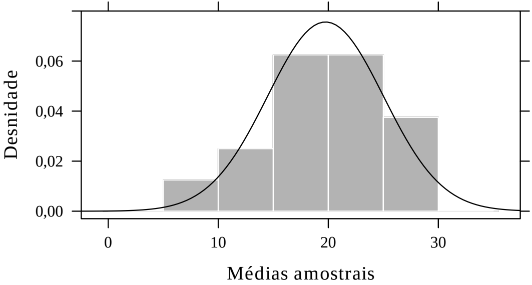
Considerando-se o exemplo de diâmetro das árvores. Agora o interesse é estimar o diâmetro médio ($\mu$).

| Amostra | Elementos | $\hat{\pi}$ | $\hat{\mu}$ | $\hat{\sigma^2}$ |
|---|---|---|---|---|
| 1 | A,B | 0,50 | 14,0 | 72,0 |
| 2 | A,C | 0,50 | 16,0 | 128,0 |
| 3 | A,D | 0,50 | 17,5 | 180,5 |
| 4 | B,C | 0,00 | 22,0 | 8,0 |
| 5 | B,D | 0,00 | 23,5 | 24,5 |
| 6 | C,D | 0,00 | 25,5 | 4,5 |
| 7 | B,A | 0,50 | 14,0 | 72,0 |
| 8 | C,A | 0,50 | 16,0 | 128,5 |
| 9 | D,A | 0,50 | 17,5 | 180,0 |
| 10 | C,B | 0,00 | 22,0 | 8,0 |
| 11 | D,B | 0,00 | 23,5 | 24,5 |
| 12 | D,C | 0,00 | 25,5 | 4,5 |
| 13 | A,A | 1,00 | 8,0 | 0,0 |
| 14 | B,B | 0,00 | 20,0 | 0,0 |
| 15 | C,C | 0,00 | 24, | 0,0 |
| 16 | D,D | 0,00 | 27,0 | 0,0 |

# Distribuição amostral da média

1. Qual a forma da distribuição das médias amostrais?
2. Qual a média da distribuição amostral dessas médias?
3. Qual a variância da distribuição amostral dessas médias?

# Distribuição amostral da média




- Forma: distribuição simétrica
- Média:

$$\frac{14,0 + 16,0 + \ldots + 27,0}{16} = 19,75\text{cm} = \mu$$

- Variância:

$$\frac{(14 - 19,75)^2 + (16 - 19,75)^2 + \ldots + (27 - 19,75)^2}{16} = 26,09 \text{ kg}^2 = \frac{\sigma^2}{n}.$$

# Distribuição amostral da média

$Y$: média do diâmetro das árvores (cm)

$$\text{Se } X \sim N(\mu, \sigma^2).$$

Seja $\bar{X}$ = a média amostral do diâmetro de $n$ árvores. A distribuição amostral de $\bar{X}$ terá distribuição normal com parâmetros:

$$\mu_{\bar{x}} = \mathrm{E}(\bar{X}) = \mu \qquad \text{e} \qquad \sigma^2_{\bar{x}} = \mathrm{Var}(\bar{X}) = \tfrac{\sigma^2}{n}.$$

$$\bar{X} \sim N(\mu_{\bar{X}}, \sigma^2_{\bar{X}}) \quad \Longrightarrow \quad Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} \sim N(0,1)$$

**Observação**: Quando são utilizadas amostras sem reposição, deve-se fazer uma correção na variância.

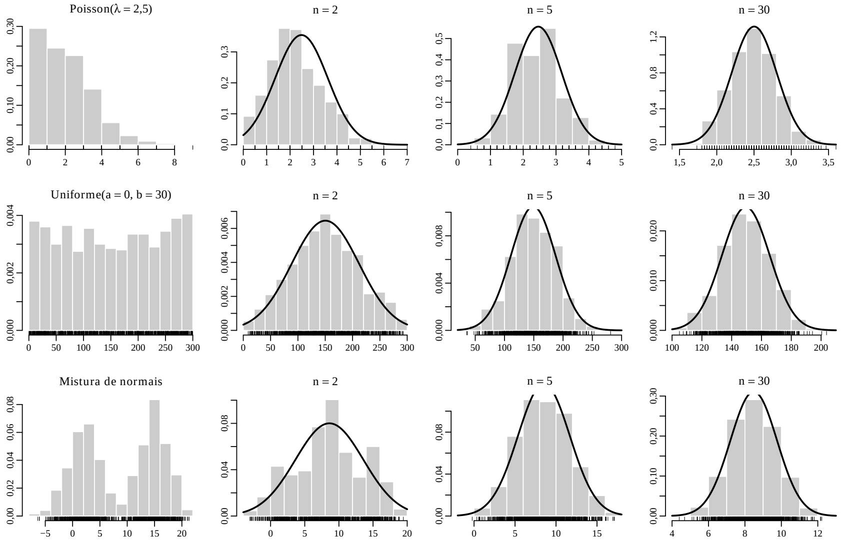
# Teorema Central do Limite

## Teorema Central do Limite

Se a população original tem uma distribuição qualquer com média $\mu$ e variância $\sigma^2$, para $n$ "suficientemente grande" (na prática, quando $n \geq 30$), $\bar{X}$ tem distribuição **aproximadamente** normal:

$$\left.\begin{array}{l} \mathrm{E}(X) = \mu \\ \mathrm{Var}(X) = \sigma^2 \end{array}\right\} \quad \Rightarrow \quad \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

# Teorema Central do Limite



# Distribuição amostral da média

**Exemplo**: Seja $X$ a produção anual de resina de árvores de *Pinus elliotti*. Suponha que $X$ segue uma distribuição normal com média 2,3 kg e desvio padrão 0,7 kg.

(a) Faça um esboço da distribuição de X.

(b) Foi tomada uma amostra aleatória de 16 árvores. Qual é a probabilidade de que a produção média das 16 árvores amostradas seja maior do que 2,8 kg?

(c) Uma amostra aleatória de 49 árvores foi tomada. Qual é a probabilidade de que a produção média das 49 árvores amostradas seja maior do que 2,8 kg?

(d) Uma amostra aleatória de 25 árvores foi tomada. Obter $\bar{x}$ tal que:

- $P(\bar{X} < \bar{x}) = 0,985$
- $P(\bar{X} < \bar{x}) = 0,975$