LCE0216
Introdução à Bioestatística Florestal

8. Distribuições amostrais

Profa. Dra. Clarice Garcia Borges Demétrio

Monitor: Silvio Gomes

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

Piracicaba, 5 de maio de 2020

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 (μ):

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

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

Árvore	Diâmetro (cm)	$x_i - \mu$	$(x_i - \mu)^2$
A	8,0	$8,0 - 19,75$	$(8,0 - 19,75)^2$
B	20,0	$20,0 - 19,75$	$(20,0 - 19,75)^2$
C	24,0	$24,0 - 19,75$	$(24,0 - 19,75)^2$
D	27,0	$27,0 - 19,75$	$(27,0 - 19,75)^2$
Total		0,0	208,75

A variância (σ^2):

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

O desvio padrão (σ):

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

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 π 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?

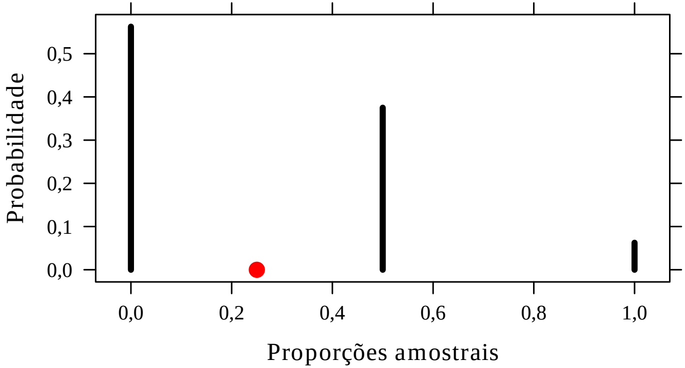
Distribuições amostrais

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

$$\hat{\mu} = \frac{\sum_{i=1}^{2} x_i}{2} \qquad s^2 = \hat{\sigma}^2 = \frac{\sum_{i=1}^{2}(x_i - \hat{\mu})^2}{n-1} = \frac{1}{n-1}\left[\sum_{i=1}^{2} x_i^2 - \frac{(\sum_{i=1}^{2} x_i)^2}{n}\right]$$

Distribuição amostral da proporção (amostras de tamanho 2):

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:

$$\begin{aligned}\sigma_P^2 = {} & (0 - 0,25)^2 \times 0,5625 + (0,50 - 0,25)^2 \times 0,3750 + \\ & + (1 - 0,25)^2 \times 0,0625 \\ & = 0,09375 = \frac{\pi(1-\pi)}{n} = \frac{0,25 \times 0,75}{3}\end{aligned}$$

Distribuição amostral da estatística P

Y: número de árvores com diâmetro inferior a 20 cm em amostras de tamanho n

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

Então,

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

Seja $P = \frac{Y}{n}$ proporção de árvores com diâmetro inferior a 20 cm. A distribuição amostral de P poderá ser aproximada por uma distribuição normal com parâmetros:

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

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

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?

Solução:

(a) $\pi = 0,37 \quad n = 200$

$P(P > 0,40) = 1 - P(P \leq 0,40) = 1 - P\Big(\frac{P - \mu_P}{\sigma_P} \leq \frac{0,40 - \mu_P}{\sigma_P}\Big) = 1 - P\Bigg(Z \leq \frac{0,40 - 0,37}{\sqrt{\frac{0,37 \times (1 - 0,37)}{200}}}\Bigg) = 1 - P(Z \leq 0,88) = 1 - 0,8106 = 0,1894$

(b) $\pi = 0,37 \quad n = 200$

$P(0,35 < P < 0,39) = ??$

(c) $\pi = 0,37 \quad n = 10$

Usando aproximação normal (não é adequada, pois n é pequeno)

$P(P > 0,50) = 1 - P(P \leq 0,50) = 1 - P\Bigg(Z \leq \frac{0,50 - 0,37}{\sqrt{\frac{0,37 \times (1 - 0,37)}{10}}}\Bigg)$

O ideal é usar a distribuição binomial – fazer a tabela de função de probabilidade e verificar qual valor satisfaz o que foi pedido.

Distribuição amostral da média

Estimador $\bar{X}$

Seja $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória de tamanho n de uma população e μ o parâmetro de interesse. Um estimador para μ é dado por

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

Sem dúvidas, o estimador mais utilizado na estatística aplicada.

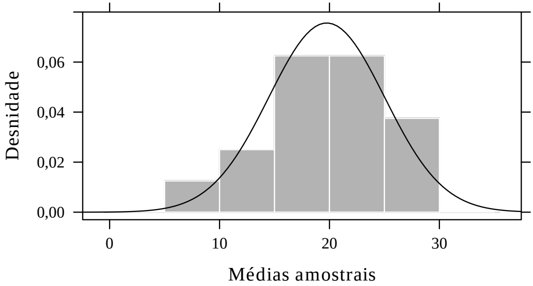
Considerando-se o exemplo de diâmetro das árvores. Agora o interesse é estimar o diâmetro médio (μ).

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

Seja $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória de tamanho n de uma população de árvores em que o DAP (cm) $X \sim N(\mu, \sigma^2)$. Seja μ o parâmetro de interesse e

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

a média amostral do diâmetro de n árvores, o estimador de μ. A distribuição amostral de $\bar{X}$ terá distribuição normal com parâmetros:

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

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

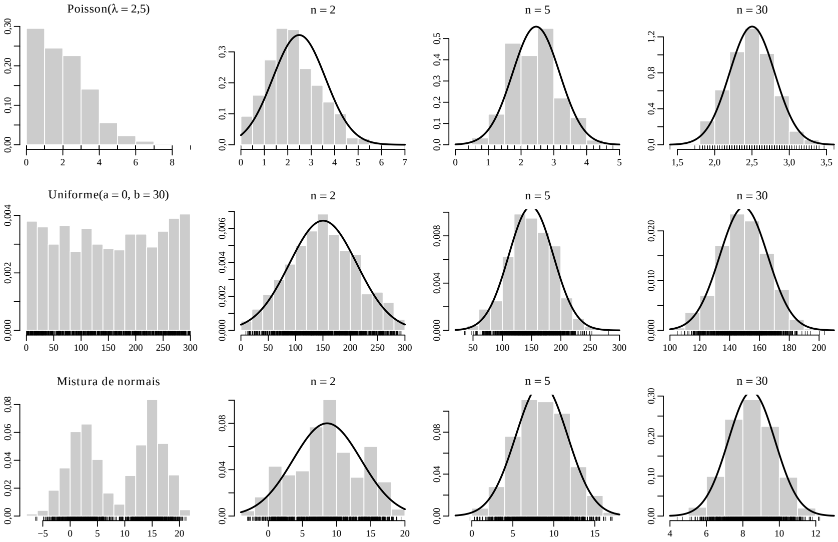
Observação: Quando são utilizadas amostras sem reposição, deve-se fazer uma correção na variância.

Teorema Central do Limite

Seja X uma variável aleatória de uma população que tem distribuição com média μ e variância σ^2. Seja $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória, de tamanho n, dessa população. Para n "suficientemente grande" (na prática, quando $n \geq 30$), o estimador $\bar{X}$ de μ tem distribuição **aproximadamente** normal com

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

Teorema Central do Limite



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

Distribuição amostral da média

Solução:

(a) $\mu_X = 2,3 \quad \sigma_X = 0,7kg \quad n = 16$

$$P(\bar{X} > 2,8) = 1 - P(\bar{X} \leq 2,8) = 1 - P\left(\frac{\bar{X} - \mu_X}{\sigma_{\bar{X}}} \leq \frac{2,3 - \mu_X}{\sigma_{\bar{X}}}\right) =$$

$$1 - P\left(Z \leq \frac{2,8 - 2,3}{\frac{0,7}{\sqrt{16}}}\right) = 1 - P(Z \leq 2,86) = 1 - 0,9979 = 0,0021$$

(b) $\mu_X = 2,3 \quad \sigma_X = 0,7kg \quad n = 49$

$$P(\bar{X} > 2,8) = 1 - P(\bar{X} \leq 2,8) = 1 - P\left(\frac{\bar{X} - \mu_X}{\sigma_{\bar{X}}} \leq \frac{2,3 - \mu_X}{\sigma_{\bar{X}}}\right) =$$

$$1 - P\left(Z \leq \frac{2,8 - 2,3}{\frac{0,7}{\sqrt{49}}}\right) = 1 - P(Z \leq 5,00) = 1 - 1 = 0$$

(c) $\mu_X = 2,3 \quad \sigma_X = 0,7kg \quad n = 25$

$$P(\bar{X} \leq \bar{x}) = 0,985 \Rightarrow P\left(\frac{\bar{X} - \mu_X}{\sigma_{\bar{X}}} \leq \frac{\bar{x} - \mu_X}{\sigma_{\bar{X}}}\right) = 0,985 \Rightarrow$$

$$\frac{\bar{x} - 2,3}{\frac{0,7}{\sqrt{25}}} = 2,17 \Rightarrow \bar{x} = 2,3 + 2,17 \times \frac{0,7}{5} = 2,60kg$$