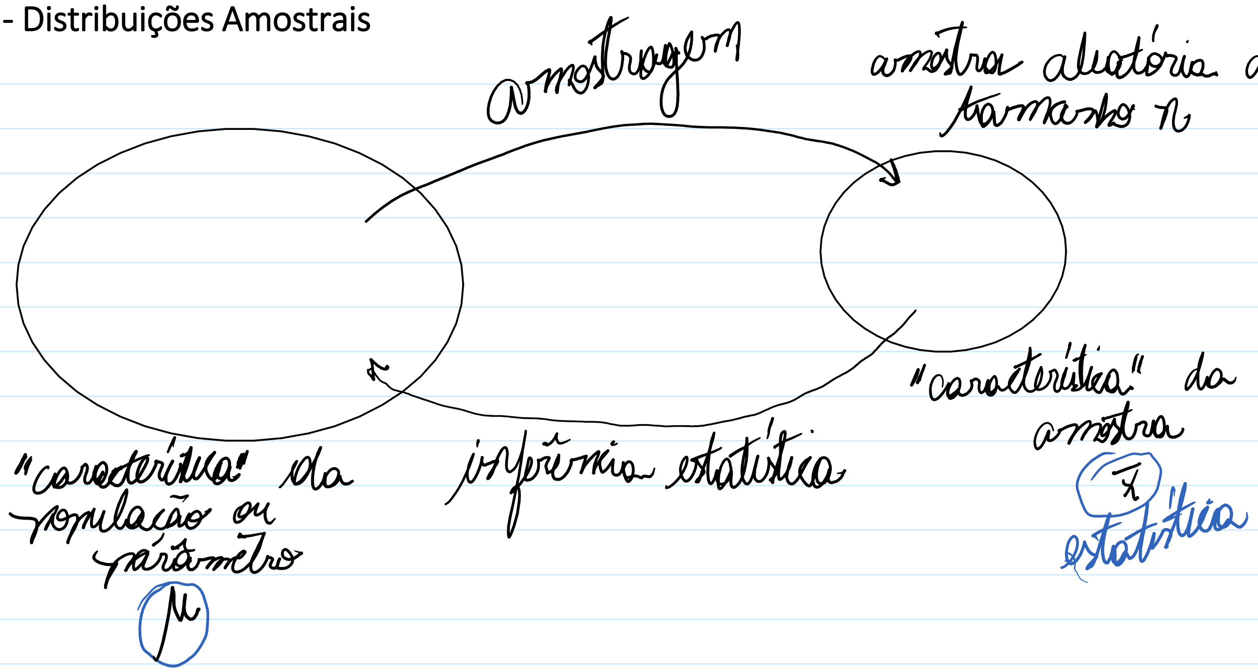
Aula 11 - Distribuições Amostrais



Exemplos de estatísticas:

- *Média da amostra:* $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
- *Proporção de elementos da amostra com determinada característica:* P
- *Variância da amostra:* $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$
- *Menor valor da amostra:* $X_{(1)} = \min\left(X_1, X_2, \cdots, X_n\right)$
- *Maior valor da amostra:* $X_{(n)} = \max\left(X_1, X_2, \cdots, X_n\right)$ *onde* $X_{(i)}$ *é a i-ésima maior observação da amostra*

*Um **parâmetro** é uma medida usada para descrever uma característica da população.*

Símbolos mais comuns para alguns parametros da população

	Parametro	Estatística
Média	μ	$\bar{X}$
Variancia	σ^2	S^2
Proporção	p	$\hat{P}$

Exemplo: Seja uma população composta por quatro árvores, com os correspondentes diâmetros apresentados 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 20 cm:

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

O diâmetro médio (μ):

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

A variância (σ^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 (σ):

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

Distribuição amostral da estatística 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

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

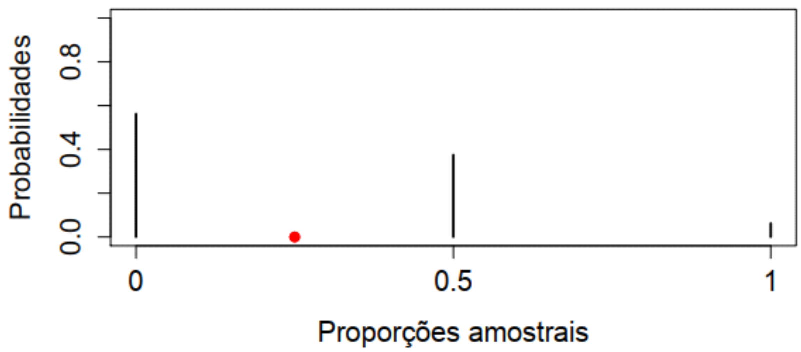
Perguntas:

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?

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 da proporção:

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



- Forma: distribuição assimétrica;

- Média:

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

- Variância:

$$(0-0,25)^2 \times 0,5625 + (0,50-0,25)^2 \times 0,3750 + (1-0,25)^2 \times 0,0625 =$$

$$= 0,09375 = \pi(1-\pi)/n$$

Vimos que...

Y: número de árvores com diâmetro inferior a 20 cm

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

Então,

$$\mu = n\pi \qquad \text{e} \qquad \sigma^2 = n\pi(1-\pi).$$

Logo, a distribuição amostral das proporções poderá ser aproximada por uma distribuição normal comparâmetros:

$$\mu = \pi \qquad \text{e} \qquad \sigma^2 = \frac{\pi(1-\pi)}{n}.$$

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.

$p = 0,37 \qquad \hat{p} = 0,37 \qquad n = 200$

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

(b) Qual é a probabilidade que na amostra de 200 visitantes, a porcentagem dos que são favoráveis 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 seja favorável à cobrança de taxas? É válido utilizar o mesmo método utilizado anteriormente? Qual método deveria ser utilizado nesse caso?

anteriormente? Qual método deveria ser utilizado nesse caso?

$\mu = \pi = 0{,}37 \qquad \sigma^2 = \frac{\pi(1-\pi)}{n} = \frac{0{,}37(0{,}63)}{200}$

$X \sim N\left(0{,}37 ; \frac{0{,}37(0{,}63)}{200}\right)$

(a) $P(p \geq 0{,}40) = P(Z \geq 0{,}88)$

$Z = \frac{0{,}40 - 0{,}37}{\sqrt{\frac{0{,}37(0{,}63)}{200}}} = 0{,}88$

$= 1 - P(Z < 0{,}88)$
$= 1 - 0{,}8106$
$= 0{,}1894$

(b) $P(0{,}35 < p < 0{,}39) = P(-0{,}59 < Z < 0{,}59) =$

$Z_1 = \frac{0{,}35 - 0{,}37}{\sqrt{\frac{0{,}37(0{,}63)}{200}}} = -0{,}59$

$= P(Z < 0{,}59) - P(Z < -0{,}59)$
$= 0{,}7224 - 0{,}2776$
$= 0{,}4448$

$Z_2 = \frac{0{,}39 - 0{,}37}{\sqrt{\frac{0{,}37(0{,}63)}{200}}} = 0{,}59$

(c) $n = 10 \qquad p = 0{,}37$

$\sigma^2 = \frac{0{,}37 \cdot (0{,}63)}{10}$

$P(p \geq 0{,}5) = P(Z \geq 0{,}85) = 1 - P(Z < 0{,}85)$
$= 1 - 0{,}8023$
$= 0{,}1977$

$Z = \frac{0{,}5 - 0{,}37}{\sqrt{\frac{0{,}37 \cdot (0{,}63)}{10}}} = 0{,}85$

Como a amostra é muito pequena estatística t

$\boxed{n \text{ pequeno}}$

Exemplo: Um processo de encher garrafas de vinho fornece 10% de garrafas com volume abaixo do especificado. Extraída uma amostra de 400 garrafas enchidas por esse processo, qual a probabilidade de a proporção amostral de garrafas com volume abaixo do especificado, P, estar entre $P_1 = 0,09$ (9%) e $P_2 = 0,11$ (11%)?

$p = 0,1$

$n = 400$

$P(0,09 < p < 0,11) = P(-0,67 < Z < 0,67)$

$$Z_1 = \frac{0,09 - 0,1}{\sqrt{\frac{0,1 \cdot 0,9}{400}}} = -0,67$$

$$Z_2 = \frac{0,11 - 0,1}{\sqrt{\frac{0,1 \cdot 0,9}{400}}} = 0,67$$

$= P(Z < 0,67) - P(Z < -0,67)$

$= 0,7486 - 0,2514$

$= 0,4972$

Distribuição amostral da média

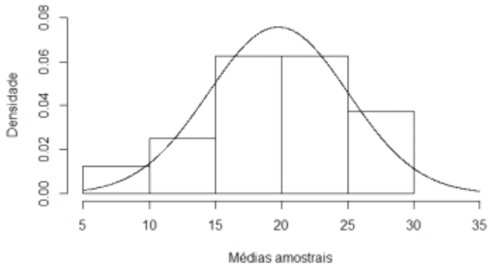
Distribuição amostral da estatística $\bar{X}$

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

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

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?



- Forma: distribuição simétrica
- Média:

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

- Variância:

$$\frac{(14,0 - 19,75)^2 + (16,0 - 19,75)^2 + \ldots + (27,0 - 19,75)^2}{16} =$$

$$= 26,09 \text{ kg}^2 = \frac{\sigma^2}{n}.$$

Assim...

Y: média do diâmetro das árvores (cm)

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

Então, a média amostral seguirá a distribuição normal com parâmetros:

$$\mu = \mu \qquad \text{e} \qquad \sigma^2 = \frac{\sigma^2}{n}.$$

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

Teorema Central do Limite

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

$$\left.\mathrm{E}(X) = \mu \right\} \Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma^2}{\ }\right)$$

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

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?

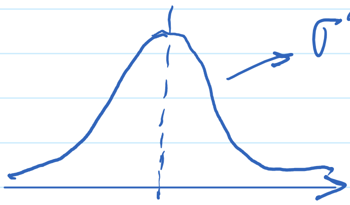
(c) Uma amostra aleatória de 49 árvores foi tomada. Qual é a probabilidade de que a produção média das 49 árvores amostradas seja maior do que 2,8 kg?

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

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

$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \to n \geq 30$

(a)



2,3

$\sigma^2 = (0,7)^2 = 0,49$

(b) $n = 16$

$P(\bar{X} > 2,8) = P(Z > 2,86) = 1 - P(Z < 2,86)$
$= 1 - 0,9979$
$= 0,0021$

$Z = \dfrac{2,8 - 2,3}{\sqrt{\frac{0,49}{16}}} = 2,86$

(c) $n = 49$

$P(\bar{X} > 2,8) = P(Z > 5) = 0$

$Z = 2,8 - 2,3 = 5$

$$Z = \frac{2,8 - 2,3}{\sqrt{\frac{0,49}{49}}} = 5$$

(d) $n = 25$

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

$Z = 2,17$

$$2,17 = \frac{\bar{x} - 2,3}{\sqrt{\frac{0,49}{25}}}$$

$\bar{x} = 2,6038 \cong 2,6$

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

$Z = 1,96$

$$1,96 = \frac{\bar{x} - 2,3}{\sqrt{\frac{0,49}{25}}}$$

$\bar{x} = 2,5744 \cong 2,6$