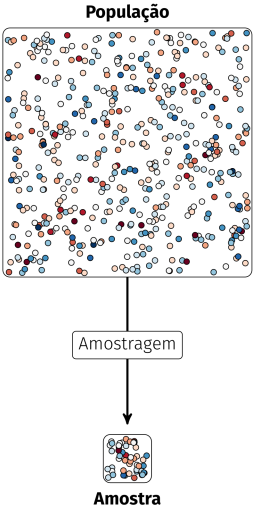
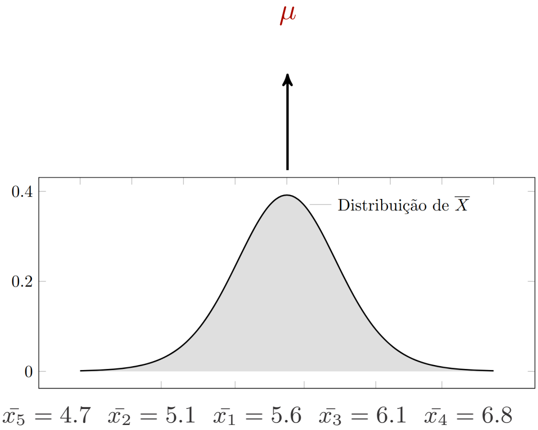
Distribuição amostral

Parte 2

Prof.: Eduardo Vargas Ferreira

Distribuição amostral da média





Parâmetros e Estatísticas

Denominação	População	Amostra
Média	μ	$\overline{X} = \sum \frac{X_i}{n}$
Variância	σ^2	$S^2 = \sum \frac{(X_i - \overline{X})^2}{(n-1)}$
Proporção	p	$\hat{p}$
Mediana	Q_2	q_2
Intervalo inter-quartil	$d_Q = Q_3 - Q_1$	$d_q = q_3 - q_1$
Função de densidade	$f(x)$	histograma
Função de distribuição	$F(x)$	$F_e(x)$

Distribuição amostral da média

Distribuição amostral da média

- Seja $X_1, X_2, \ldots, X_n$ uma sequência de variáveis aleatórias independentes e identicamente distribuídas, cada uma com $\mathbb{E}(X_i) = \mu$ e $\mathbb{V}ar(X_i) = \sigma^2$, então:

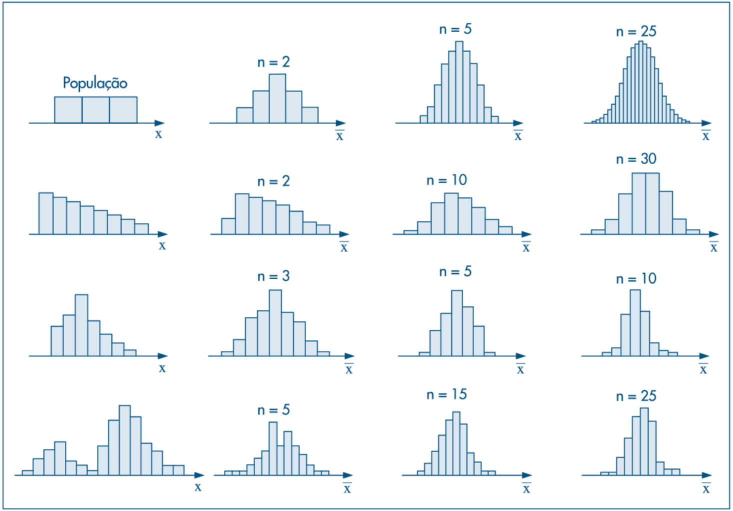
$$\mathbb{E}(\bar{X}) = \mu$$

$$\mathbb{V}ar(\bar{X}) = \frac{\sigma^2}{n}$$

- O **Teorema Central do Limite** nos diz que

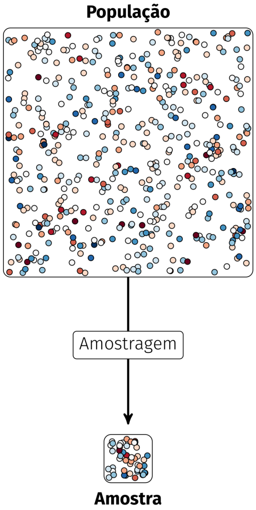
$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} N(0,1), \quad \text{quando} \quad n \to \infty$$

Caso geral

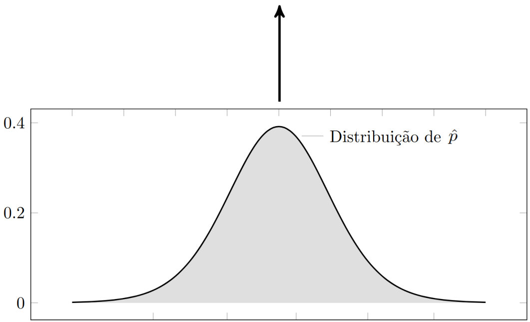


População
x
n = 2
$\overline{x}$
n = 5
$\overline{x}$
n = 25
$\overline{x}$
x
n = 2
$\overline{x}$
n = 10
$\overline{x}$
n = 30
$\overline{x}$
x
n = 3
$\overline{x}$
n = 5
$\overline{x}$
n = 10
$\overline{x}$
x
n = 5
$\overline{x}$
n = 15
$\overline{x}$
n = 25
$\overline{x}$

Distribuição amostral da média



p



$\hat{p_5} = 0.47 \quad \hat{p_2} = 0.54 \quad \hat{p_1} = 0.61 \quad \hat{p_3} = 0.70 \quad \hat{p_4} = 0.72$

Parâmetros e Estatísticas

Denominação	População	Amostra
Média	μ	$\overline{X} = \sum \frac{X_i}{n}$
Variância	σ^2	$S^2 = \sum \frac{\left(X_i - \overline{X}\right)^2}{(n-1)}$
Proporção	p	$\hat{p}$
Mediana	Q_2	q_2
Intervalo inter-quartil	$d_Q = Q_3 - Q_1$	$d_q = q_3 - q_1$
Função de densidade	$f(x)$	histograma
Função de distribuição	$F(x)$	$F_e(x)$

Parâmetros e Estatísticas

Denominação	População	Amostra
Média	μ	$\overline{X} = \sum \frac{X_i}{n}$
Variância	σ^2	$S^2 = \sum \frac{(X_i - \overline{X})^2}{(n-1)}$
$\boxed{\text{Proporção}}$	p	$\hat{p}$
Mediana	Q_2	q_2
Intervalo inter-quartil	$d_Q = Q_3 - Q_1$	$d_q = q_3 - q_1$
Função de densidade	$f(x)$	histograma
Função de distribuição	$F(x)$	$F_e(x)$

Distribuição amostral de uma proporção

Distribuição amostral de uma proporção

- Considere uma população em que a proporção de elementos portadores de certa característica seja p. Logo, podemos definir uma v.a. X como:

$$X = \begin{cases} 1 & \text{se o indivíduo for portador da característica} \\ 0 & \text{se o indivíduo não for portador da característica} \end{cases}$$

- Retirada uma AAS dessa população, seja $\displaystyle\sum_{i=1}^{n} X_i$, temos:

$$\sum_{i=1}^{n} X_i \sim b(n,p) \quad com \quad \begin{cases} \mathbb{E}\left(\displaystyle\sum_{i=1}^{n} X_i\right) = n \cdot p \\ \mathbb{V}ar\left(\displaystyle\sum_{i=1}^{n} X_i\right) = n \cdot p \cdot (1-p) \end{cases}$$

Distribuição amostral de uma proporção

- Então, pelo TCL, temos:

$$
\begin{aligned}
\frac{\sum_{i=1}^{n} X_i}{n} \quad &\sim \quad N\left[\,\mathbb{E}\left(\frac{\sum_{i=1}^{n} X_i}{n}\right)\,,\ \mathbb{V}ar\left(\frac{\sum_{i=1}^{n} X_i}{n}\right)\,\right] \\
&= \quad N\left[\,\frac{1}{n}\,\mathbb{E}\left(\sum_{i=1}^{n} X_i\right)\,,\ \frac{1}{n^2}\mathbb{V}ar\left(\sum_{i=1}^{n} X_i\right)\,\right] \\
&= \quad N\left(\,p\,,\ \frac{p\cdot(1-p)}{n}\,\right)
\end{aligned}
$$

- Assim, a distribuição amostral de uma proporção, $\hat{p}$ é dada por:

$$\hat{p} \quad \sim \quad N\left(\,p\,,\ \frac{p\cdot(1-p)}{n}\,\right)$$

Exemplo: vacina contra gripe

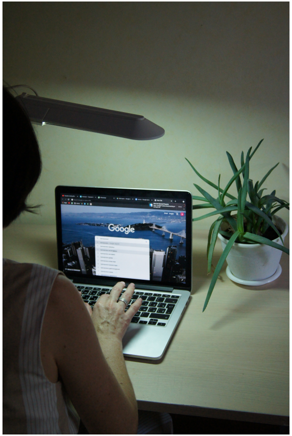
- Um fabricante afirma que sua vacina contra gripe imuniza em 80% dos casos. Uma amostra de 25 indivíduos foi sorteada para verificar sua imunização.

Se o fabricante estiver correto, qual a probabilidade da proporção de imunizados ser inferior a 0,75?

$$
\begin{aligned}
P\left(\hat{p} < 0.75 \mid p = 0.80\right) &= P\left(\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} < \frac{0.75 - p}{\sqrt{\frac{p(1-p)}{n}}}\right) \\
&= P\left(Z < \frac{0.75 - 0.8}{\sqrt{\frac{0.8(1-0.8)}{25}}}\right) \\
&= P\left(Z < -0.625\right) \approx 0.266.
\end{aligned}
$$

Exemplo: acesso à internet

- Uma pesquisa divulgou que 56% das famílias brasileiras têm acesso à internet. Para verificar a veracidade dessa informação, coletou-se uma amostra de 300 famílias.



Qual a probabilidade dessa proporção amostral estar próximo da proporção populacional em menos de 0,03?

$$\begin{aligned}
P(|\hat{p} - p| < 0.03) &= P\left(\frac{-0.03}{\sqrt{\frac{p\cdot(1-p)}{n}}} < \frac{\hat{p} - p}{\sqrt{\frac{p\cdot(1-p)}{n}}} < \frac{0.03}{\sqrt{\frac{p\cdot(1-p)}{n}}}\right) \\
&\approx P\left(\frac{-0.03}{\sqrt{\frac{0.56\cdot(1-0.56)}{300}}} < Z < \frac{0.03}{\sqrt{\frac{0.56\cdot(1-0.56)}{300}}}\right) \\
&= P(-1.046 < Z < 1.046) \approx 0.704
\end{aligned}$$

Exemplo: acesso à internet (cont.)

- Responda o item anterior considerando amostras de tamanho 600 e 1000.

n=600

$$
\begin{aligned}
P\left(-0.03 < \hat{p} - p < 0.03\right) &\approx P\left(\frac{-0.03}{\sqrt{0.56 \cdot (1-0.56)/600}} < Z < \frac{0.03}{\sqrt{0.56 \cdot (1-0.56)/600}}\right) \\
&= P\left(-1.480 < Z < 1.480\right) \approx 0,861
\end{aligned}
$$

n=1000

$$
\begin{aligned}
P\left(-0.03 < \hat{p} - p < 0.03\right) &\approx P\left(\frac{-0.03}{\sqrt{0.56 \cdot (1-0.56)/1000}} < Z < \frac{0.03}{\sqrt{0.56 \cdot (1-0.56)/1000}}\right) \\
&= P\left(-1.911 < Z < 1.911\right) \approx 0,944
\end{aligned}
$$

- Conforme n aumenta, é mais **improvável** de $\hat{p}$ estar distante de p.

Referências

- Bussab, WO; Morettin, PA. Estatística Básica. São Paulo: Editora Saraiva, 2006 (5ª Edição).
- Magalhães, MN; Lima, ACP. Noções de Probabilidade e Estatística. São Paulo: EDUSP, 2008.