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

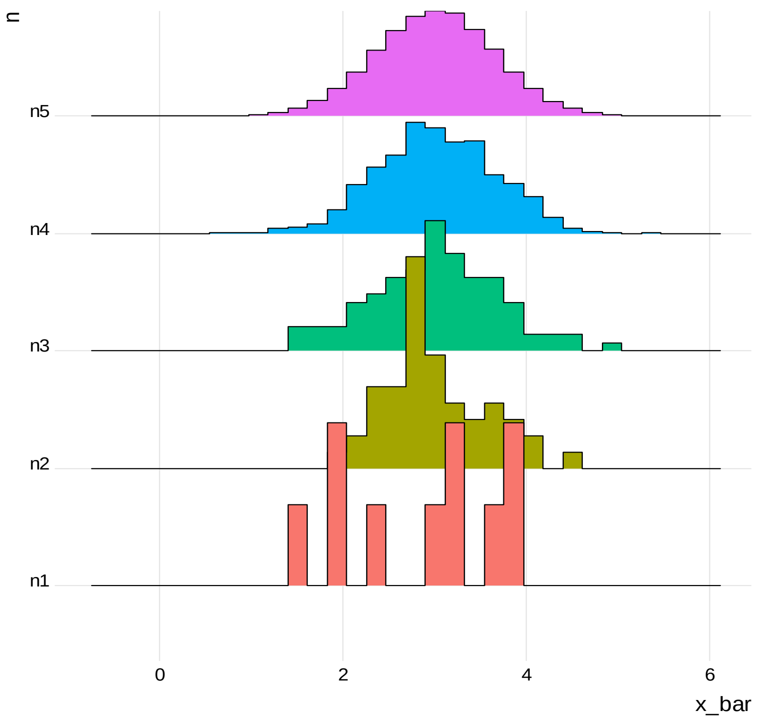
Estatística Geral

Professor
Fábio Prataviera

Distribuição Amostral e Estimação Intervalar

AULA DE HOJE

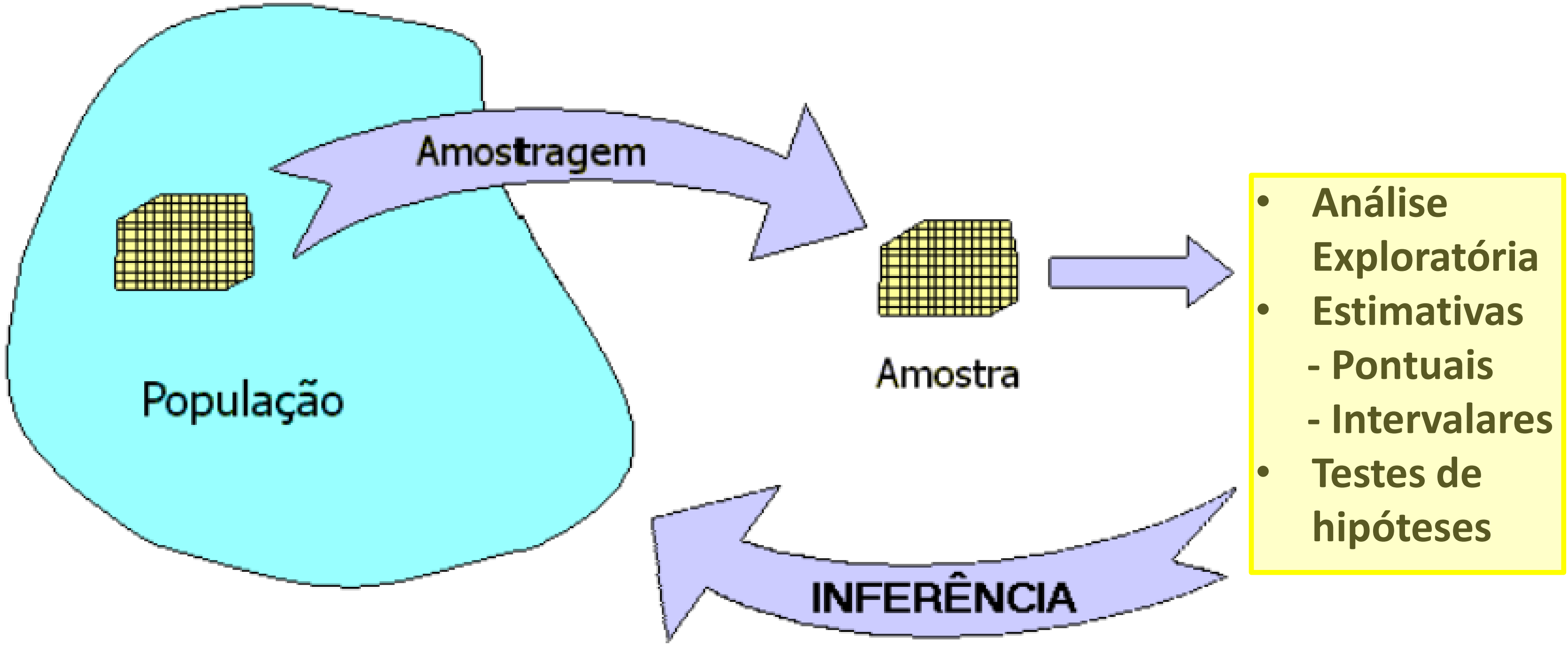
- Distribuição amostral da média;
- Distribuição amostral da proporção;
- Estimação intervalar da média;
- Estimação intervalar da proporção.



REVISÃO

- Conceitos de estatística descritiva e exploratória de dados;
- Natureza das variáveis;
- Processos de amostragem;
- Probabilidade e variáveis aleatórias;
- Alguns modelos probabilísticos;
- Estimação pontual e propriedades dos estimadores.

POPULAÇÃO E AMOSTRA



População e Amostra

Cada v.a. está associada a uma distribuição de probabilidade com vetor de parâmetros θ desconhecido.

Por exemplo:

$$\mathrm{X} \sim Exp(\lambda),\ \theta = \lambda$$

$$\mathrm{X} \sim Normal(\mu, \sigma^2),\ \theta = (\mu, \sigma^2)$$

$$\mathrm{X} \sim Normal(\mu, \sigma^2),\ \theta = \mu \text{ com } \sigma^2 \text{ conhecido}$$

$$\mathrm{X} \sim Bernoulli(p),\ \theta = p$$

População e Amostra

Parâmetros: São as quantidades da população, em geral desconhecidas, sobre as quais temos interesse.

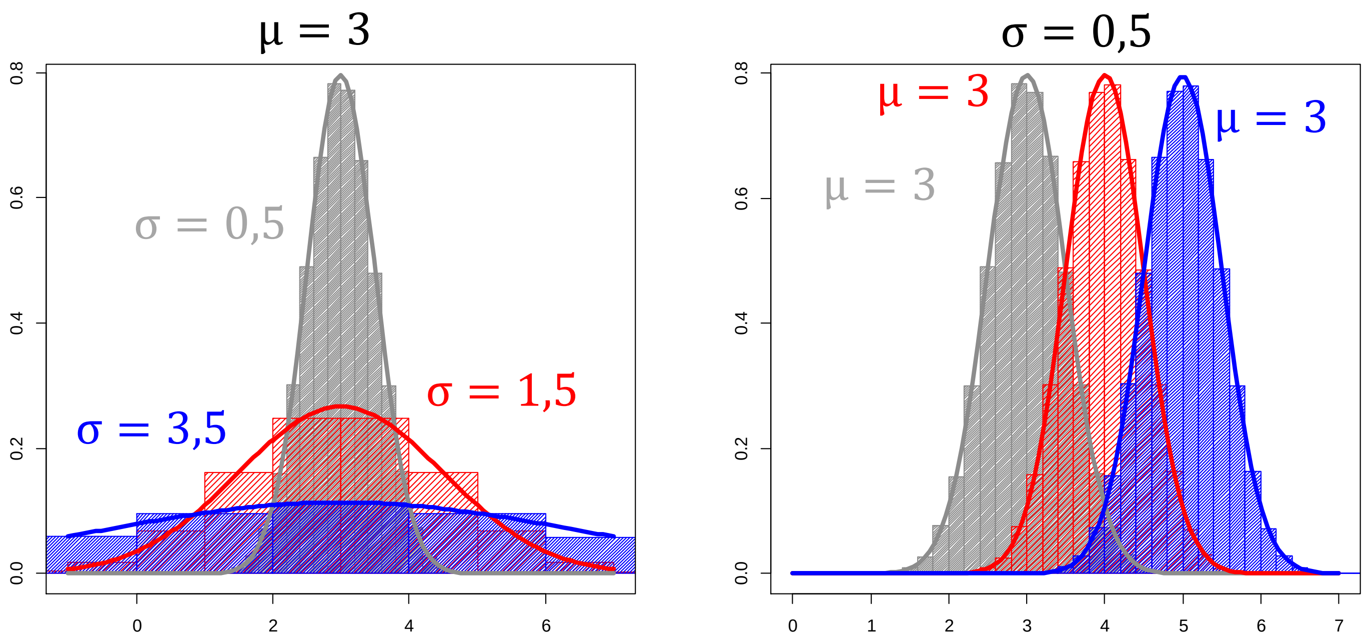
Estatísticas: É uma característica da amostra $X_1, \ldots, X_n$ de uma variável aleatória X. Ou seja, uma estatística T é uma função de $X_1, \ldots, X_n$, $T(X_1, \ldots, X_n)$ e portanto é uma variável aleatória.

	Parâmetro (População)	Estatísticas (Amostras)	Valores Observados
Média	$\boldsymbol{\mu}$	$\bar{\mathbf{X}}$	$\bar{\mathbf{x}}$
Variância	$\boldsymbol{\sigma}^2$	$\mathbf{S}^2$	$\mathbf{s}^2$
Proporção	$\mathbf{p}$	$\hat{\mathbf{P}}$	$\hat{\mathbf{p}}$

Distribuição Normal

Distribuição Normal: Seja $X_1, \ldots, X_n$ uma amostra de uma variável aleatória X com distribuição normal com média μ é variância σ^2. A função densidade de probabilidade de X é dada por (para $x, \mu \in \mathcal{R}$ e $\sigma^2 > 0$)

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}}$$



Distribuição Amostral

- Um estimador é uma função de variáveis aleatórias e portanto também é variável aleatória com uma distribuição de probabilidade denominada **Distribuição Amostral**.

- A distribuição das **médias** ou **proporções** amostrais representam a população de todas as **médias** ou **proporções** oriundas de uma amostra de tamanho n de uma variável aleatória.

- A convergência em forma de distribuição e dos parâmetros dessa distribuição são elucidadas pela **Lei dos Grandes Números** e pelo **Teorema Limite Central.**

Exemplo

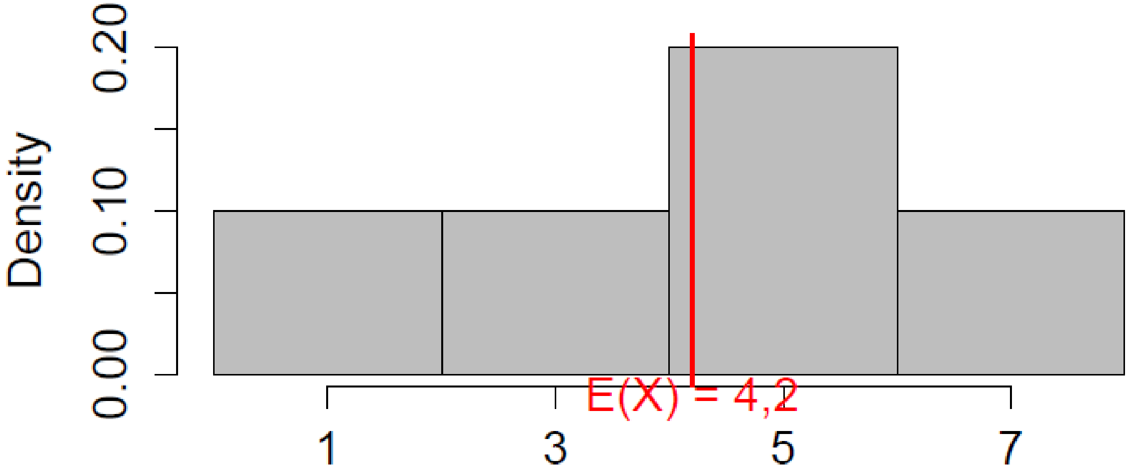
Considere uma população em que a variável X pode assumir um dos valores do conjunto {1, 3, 5, 5, 7}. A distribuição de probabilidade de X é

x	1	3	5	7
$\mathbb{P}(X = x)$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{5}$

Esperança e Variância:

$$\mathbb{E}(X) = \mu_X = 4,2$$
$$\mathbb{V}\text{ar}(X) = \sigma_X^2 = 4,16$$

EXEMPLO



$n = 1, Var[X] = 4{,}16$

Exemplo

Continuação do exemplo: Selecionar todas as amostras aleatórias simples de tamanho 2, $n = 2$**,** selecionadas ao acaso e com reposição da população X, e encontrar a distribuição do estimador pontual $\overline{\mathrm{X}}$, ou seja, encontrar a distribuição da média amostral

$$\overline{X} = \frac{X_1 + X_2}{2}$$

em que

- X_1 é o valor na primeira seleção.
- X_2 é o valor na segunda seleção.

Exemplo

Amostra (X_1, X_2)	Probabilidade	Média Amostral
(1,1)	1/25	1
(1,3)	1/25	2
(1,5)	2/25	3
(1,7)	1/25	4
(3,1)	1/25	2
(3,3)	1/25	3
(3,5)	2/25	4
(3,7)	1/25	5
(5,1)	1/25	3
(5,3)	1/25	4
(5,5)	2/25	5
(5,7)	1/25	6
(7,1)	1/25	4
(7,3)	1/25	5
(7,5)	2/25	6
(7,7)	1/25	7

EXEMPLO

Distribuição de $\overline{\mathrm{X}}$ para $n = 2$

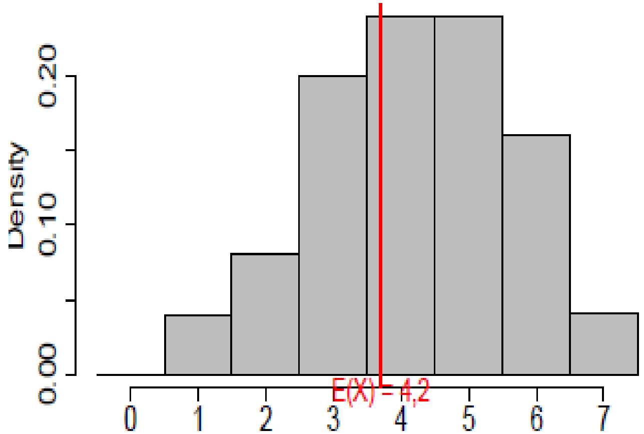
$\bar{x}$	1	2	3	4	5	6	7
$\mathbb{P}(\overline{X} = \bar{x})$	1/25	2/25	5/25	6/25	6/25	4/25	1/25

Esperança e Variância de $\overline{\mathrm{X}}$ para $n = 2$

$$\mathbb{E}(\overline{X}) = \mu_X = 4,2$$

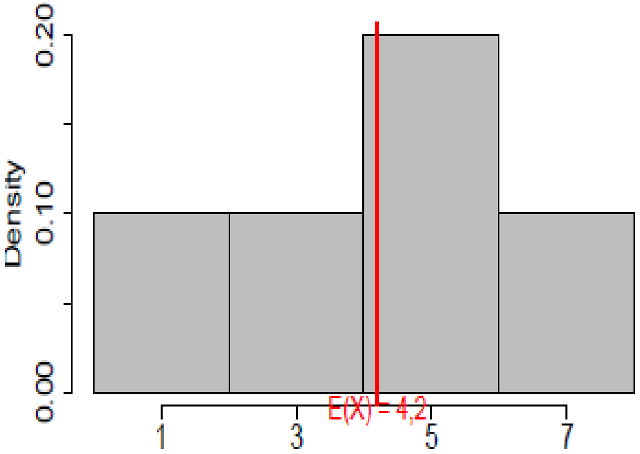
$$\mathbb{V}\mathrm{ar}(\overline{X}) = 2,08 = \frac{\sigma_X^2}{2}$$

EXEMPLO

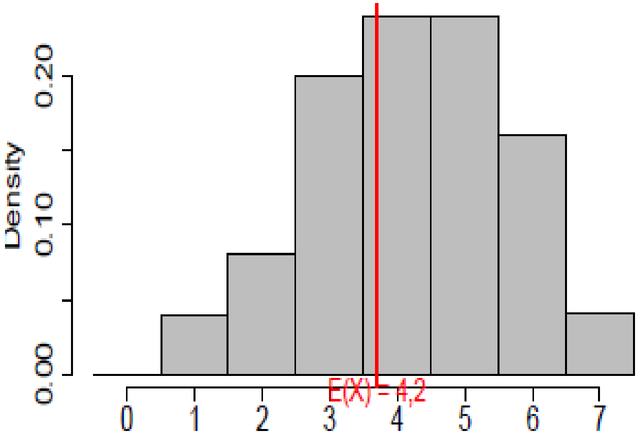


$n = 2, Var[\bar{X}] = 2{,}08$

EXEMPLO



$n = 1, Var[X] = 4{,}16$



$n = 2, Var[\bar{X}] = 2{,}08$

EXEMPLO

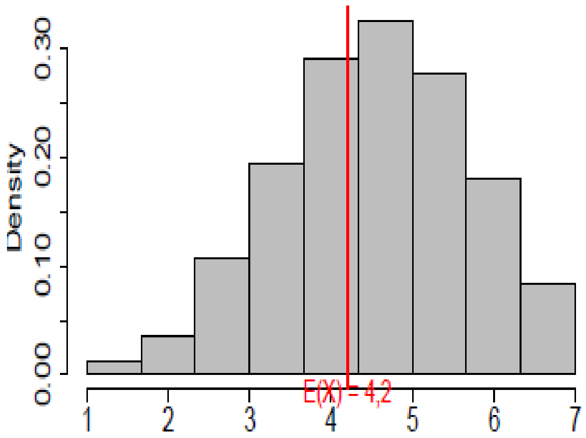
Distribuição de $\overline{X}$ para $n = 3$

$\bar{x}$	$\mathbb{P}(\overline{X} = \bar{x})$
1	1/125
5/3	3/125
7/3	9/125
3	16/125
11/3	24/125
13/3	27/125
5	23/125
17/3	15/125
19/3	6/125
1	1/125

Esperança e Variância de $\overline{X}$ para $n = 3$

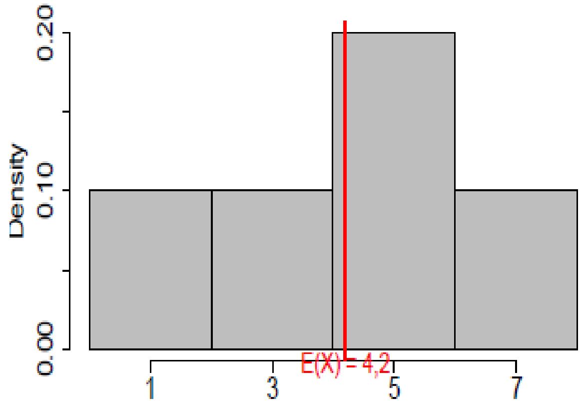
$$\mathbb{E}(\overline{X}) = \mu_X = 4,2 \text{ e } \mathbb{V}\mathrm{ar}(\overline{X}) = 1,39 = \frac{\sigma_X^2}{3}$$

EXEMPLO

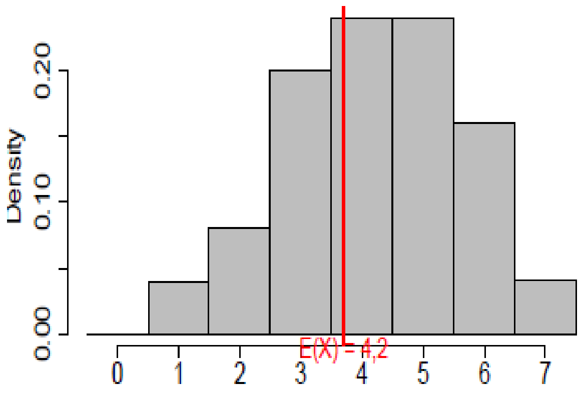


$n = 3, Var[\bar{X}] = 1{,}39$

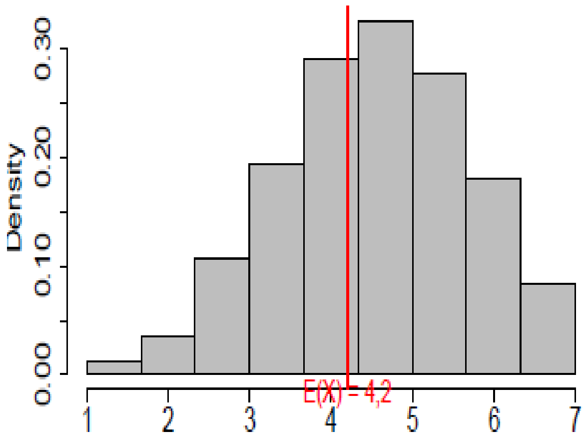
EXEMPLO



$n = 1, Var[X] = 4{,}16$

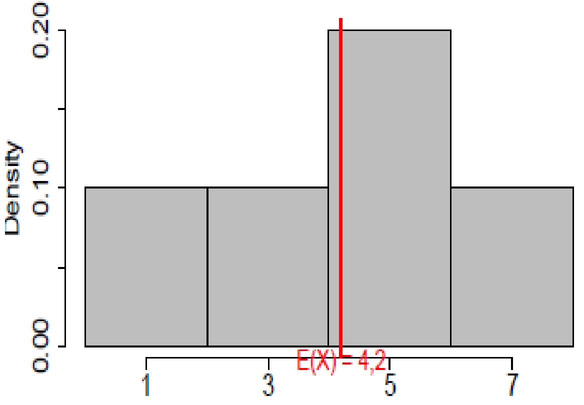


$n = 2, Var[\bar{X}] = 2{,}08$

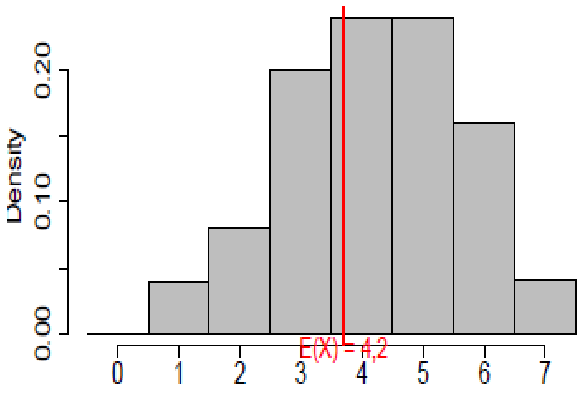


$n = 3, Var[\bar{X}] = 1{,}39$

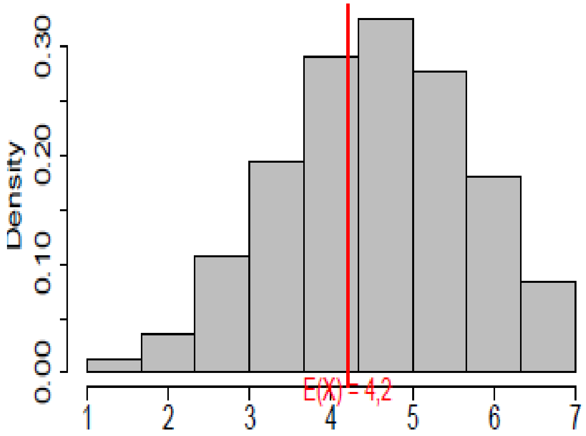
EXEMPLO



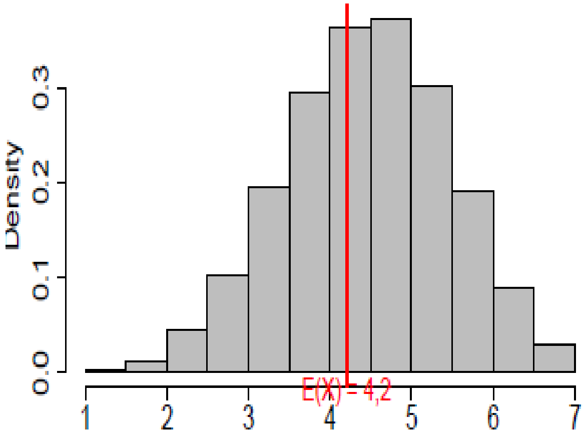
$n = 1, Var[X] = 4{,}16$



$n = 2, Var[\bar{X}] = 2{,}08$



$n = 3, Var[\bar{X}] = 1{,}39$



$n = 4, Var[\bar{X}] = 1{,}04$

Discussão do Exemplo

Análise dos Histogramas

- Conforme o tamanho da amostra aumenta, $\boldsymbol{n} \to \infty$, os valores de $\bar{X}$ tendem a concentrar-se cada vez mais em torno de $\mathbb{E}[\bar{\boldsymbol{X}}] = \mathbf{4,2}$.
- A **variância diminui** a medida que o tamanho da **amostra aumenta**.
- Para $\boldsymbol{n}$ suficientemente grande, a forma da distribuição (histograma) aproxima-se de uma **distribuição simétrica (normal)**.

DISTRIBUIÇÃO AMOSTRAL

Teorema

Sejam $\mathrm{X}_1, \ldots, \mathrm{X}_\mathrm{n}$ uma amostra aleatória, retiradas de uma população com média μ é variância σ^2 finita. Então, sendo $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ a média da amostra,

$$E[\bar{X}] = \mu \text{ e } \mathrm{Var}[\bar{X}] = \frac{\sigma^2}{n}$$

$$\boxed{E[\bar{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \text{ e } Var[\bar{X}] = Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]}$$

DISTRIBUIÇÃO AMOSTRAL

Teorema

Sejam $X_1, \dots, X_n$ uma amostra aleatória de uma **distribuição normal** com média μ é variância σ^2. Então,

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}).$$

DISTRIBUIÇÃO AMOSTRAL

Teorema

Sejam $X_1, \dots, X_n$ uma amostra aleatória de uma **distribuição normal** com média μ é variância σ^2. Então,

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}).$$

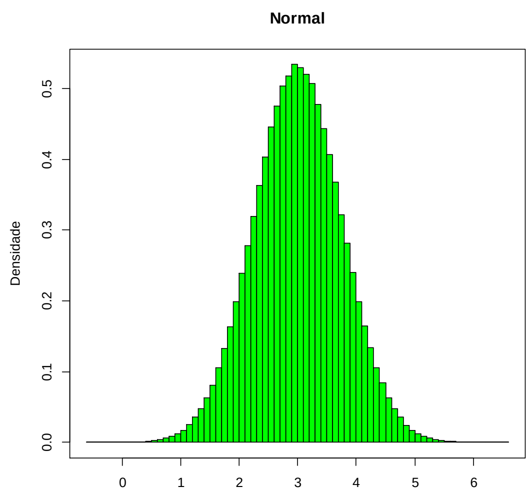
Usar esse resultado: $M_{\bar{X}}(t) = [M_X(t/n)]^n$

DISTRIBUIÇÃO AMOSTRAL

O que ocorre para qualquer distribuição?

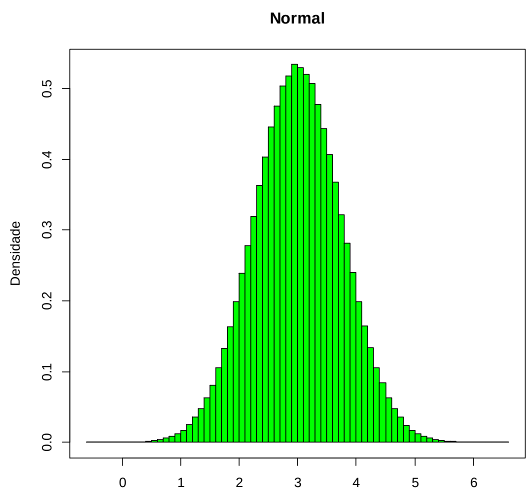
DISTRIBUIÇÃO AMOSTRAL

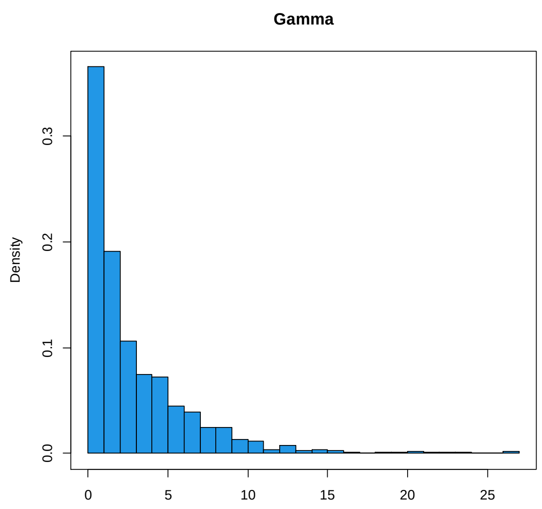
O que ocorre para qualquer distribuição?



DISTRIBUIÇÃO AMOSTRAL

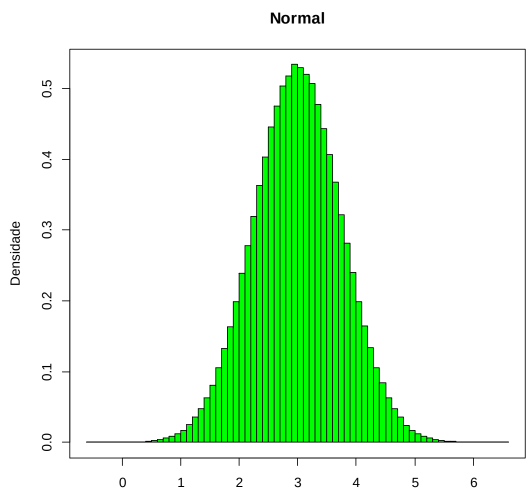
O que ocorre para qualquer distribuição?

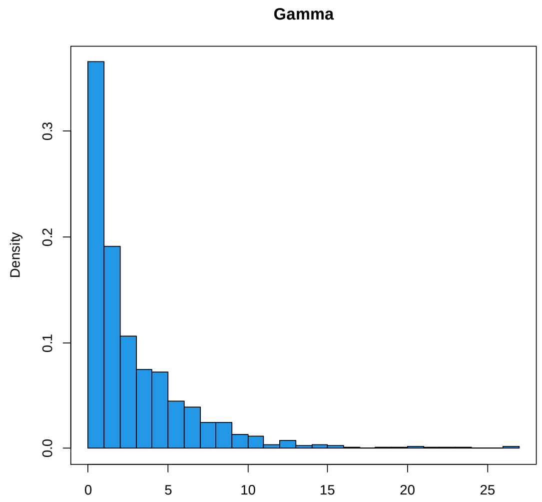


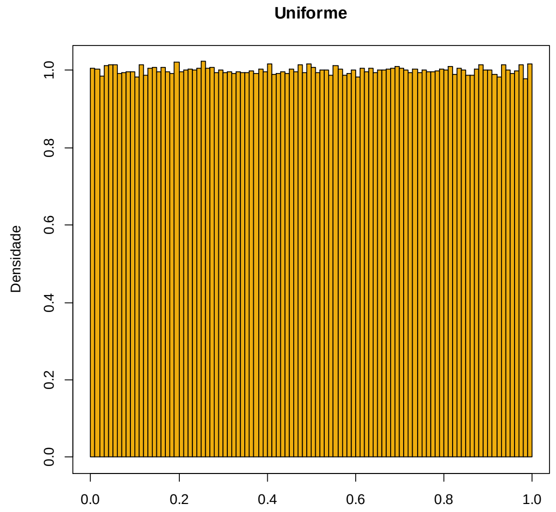


DISTRIBUIÇÃO AMOSTRAL

O que ocorre para qualquer distribuição?

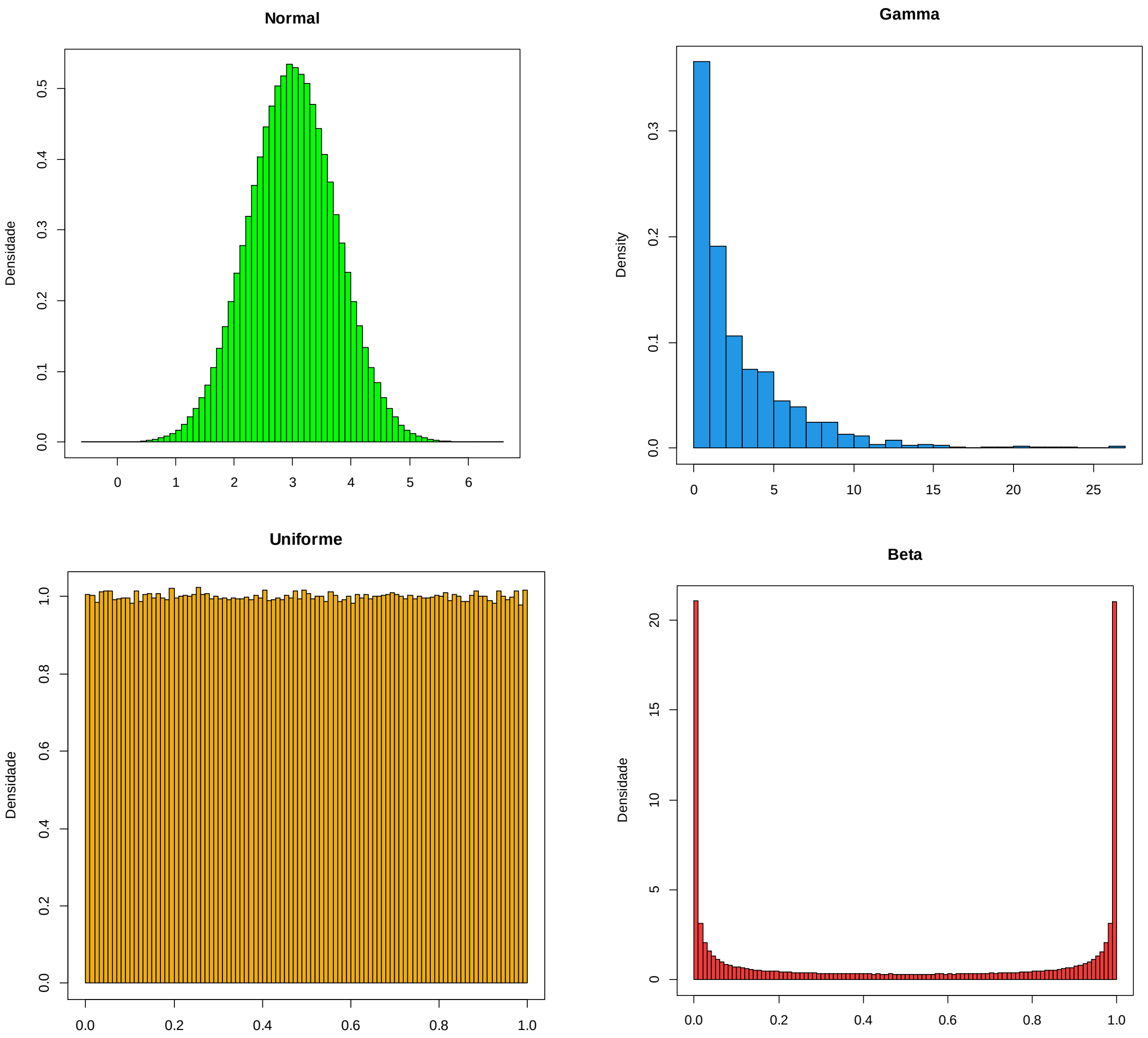




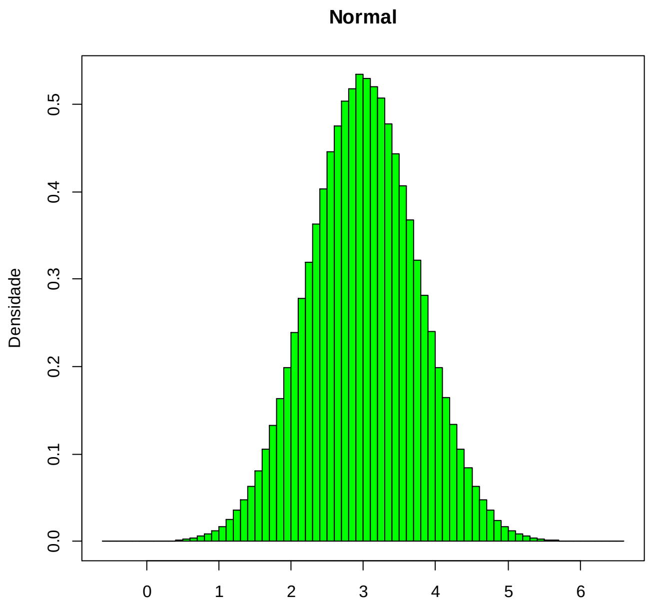


DISTRIBUIÇÃO AMOSTRAL

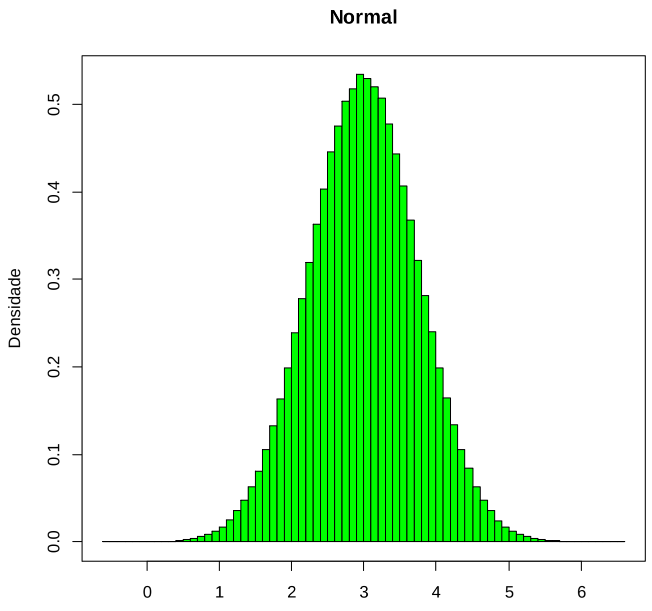
O que ocorre para qualquer distribuição?

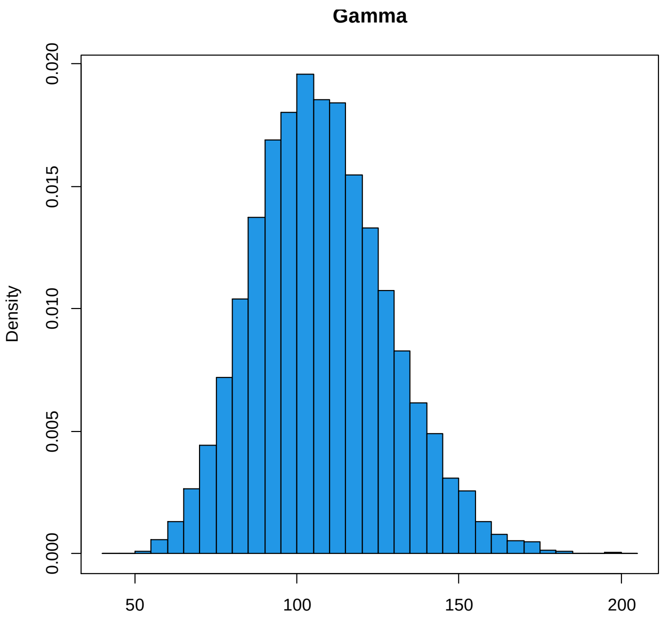


Distribuição Amostral

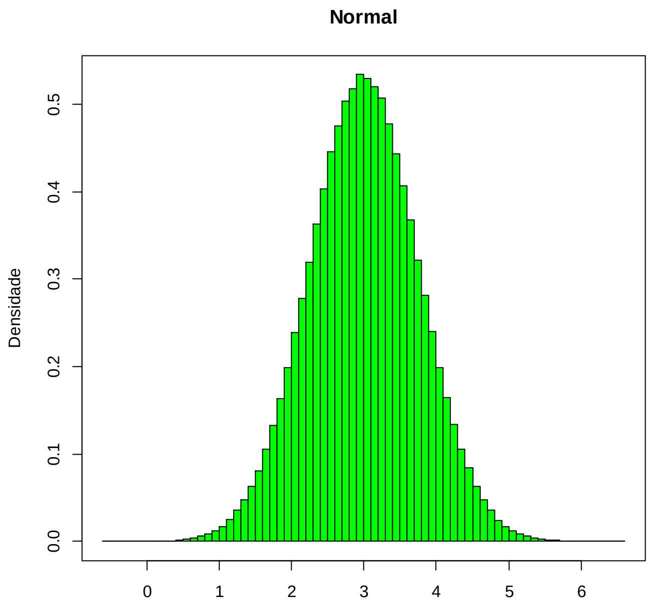


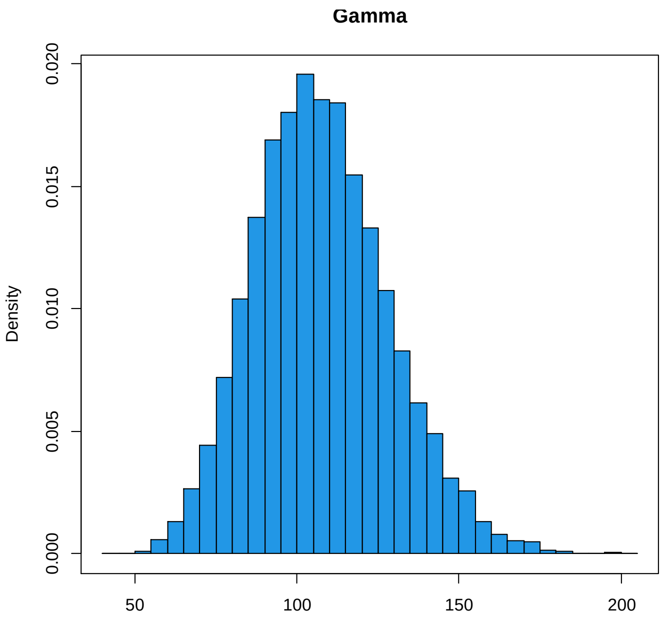
Distribuição Amostral

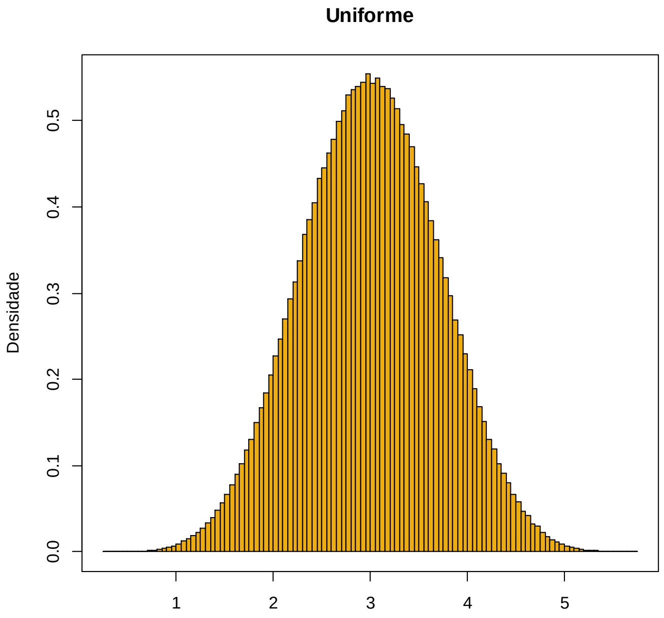




DISTRIBUIÇÃO AMOSTRAL

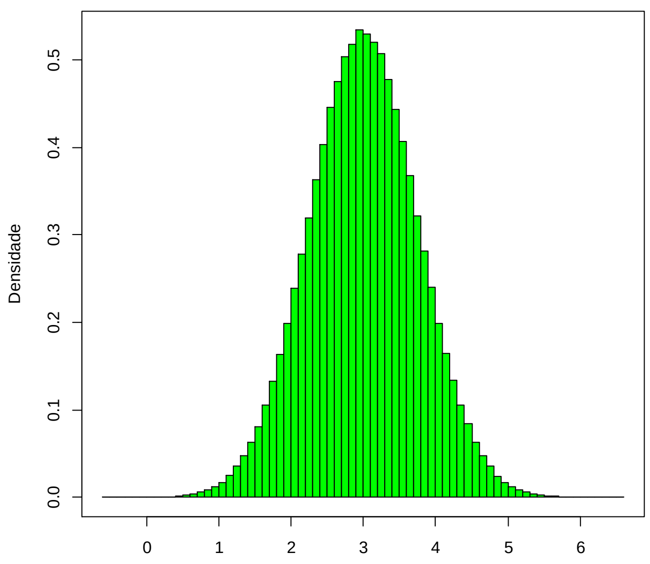




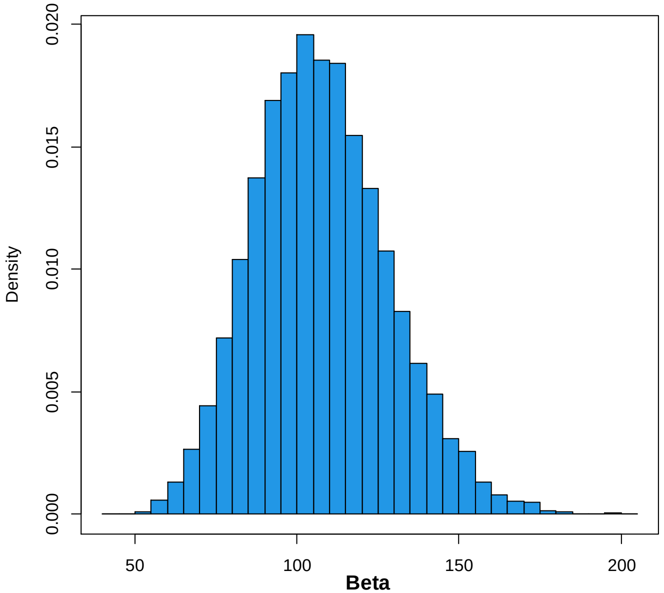


DISTRIBUIÇÃO AMOSTRAL

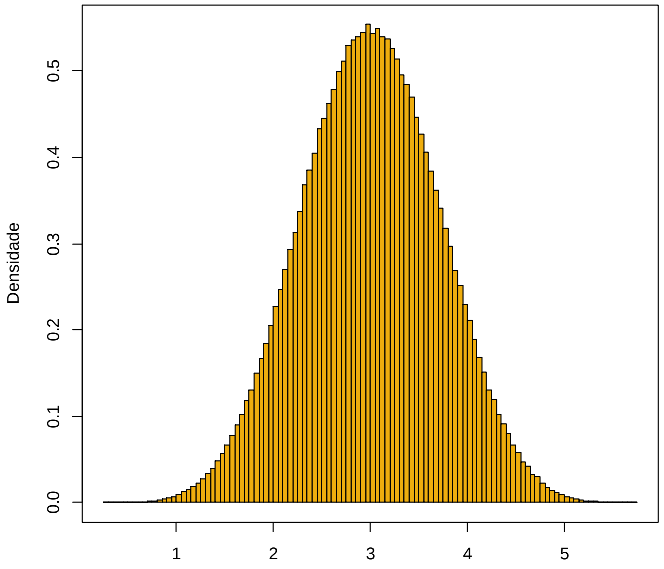
Normal



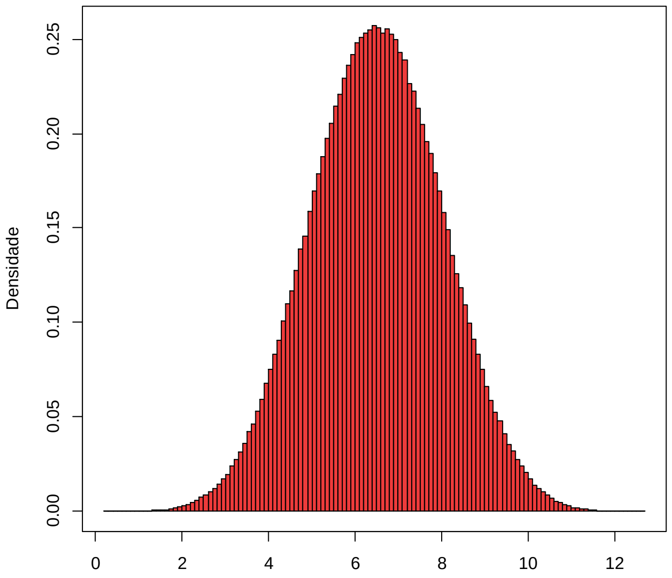
Gamma



Uniforme



Beta



TEOREMA LIMITE CENTRAL

Teorema Limite Central (TLC)

Para amostras aleatórias $X_1, \dots, X_n$ *iid*, retiradas de uma população com média μ é variância σ^2 finita, a distribuição amostral da média $\bar{X}$ aproxima-se, para n grande ($n \to \infty$), de uma distribuição normal com média μ é variância ${}^{\sigma^2}/_{n}$,

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \quad N(0,1)$$

- A variável aleatória $e = \bar{X} - \mu$ é denominado **erro amostral da média**.
- O desvio padrão ${}^{\sigma}/_{\sqrt{n}}$ é denominado **erro padrão da média**.

DISTRIBUIÇÃO AMOSTRAL DA MÉDIA

Exemplo: Sabe-se que o faturamento diário de um posto de gasolina segue uma distribuição de média 20 mil e desvio padrão 2 mil reais. Qual é a probabilidade de que num período de 60 dias, o faturamento total ultrapasse R$ 1.230.000,00?

Distribuição Amostral da Média

Exemplo: Sabe-se que o faturamento diário de um posto de gasolina segue uma distribuição de média 20 mil e desvio padrão 2 mil reais. Qual é a probabilidade de que num período de 60 dias, o faturamento total ultrapasse R$ 1.230.000,00?

- X é o faturamento diário do posto de gasolina.

DISTRIBUIÇÃO AMOSTRAL DA MÉDIA

Exemplo: Sabe-se que o faturamento diário de um posto de gasolina segue uma distribuição de média 20 mil e desvio padrão 2 mil reais. Qual é a probabilidade de que num período de 60 dias, o faturamento total ultrapasse R$ 1.230.000,00?

- $\boldsymbol{X}$ é o faturamento diário do posto de gasolina.
- $\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{X}] = \mathbf{20}$ e $\sigma = \sqrt{\mathbb{V}[\boldsymbol{X}]} = \mathbf{2}$.

Distribuição Amostral da Média

Exemplo: Sabe-se que o faturamento diário de um posto de gasolina segue uma distribuição de média 20 mil e desvio padrão 2 mil reais. Qual é a probabilidade de que num período de 60 dias, o faturamento total ultrapasse R$ 1.230.000,00?

- $\boldsymbol{X}$ é o faturamento diário do posto de gasolina.
- $\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{X}] = \mathbf{20}$ e $\sigma = \sqrt{\mathbb{V}[\boldsymbol{X}]} = \mathbf{2}$.
- $\boldsymbol{P}(\boldsymbol{X_1} + \cdots + \boldsymbol{X_{60}} > \mathbf{1230})$?.

Distribuição Amostral da Média

Resolução do Exemplo: Obtendo uma amostra de 60 valores de $\boldsymbol{X}$, representada por $\boldsymbol{X_1, \dots, X_{60}}$, com $\boldsymbol{X_i}$ representando o faturamento do posto no dia $\boldsymbol{i = 1, 2, 3, \dots, 60}$. Então

$$\mathbb{P}(\boldsymbol{X_1 + \cdots + X_{60} > 1230}) = \mathbb{P}\left(\frac{\boldsymbol{X_1 + \cdots + X_{60}}}{\boldsymbol{60}} > \frac{\boldsymbol{1230}}{\boldsymbol{60}}\right) \overset{\boldsymbol{1}}{=}$$

Distribuição Amostral da Média

Resolução do Exemplo: Obtendo uma amostra de 60 valores de $\boldsymbol{X}$, representada por $\boldsymbol{X_1, \dots, X_{60}}$, com $\boldsymbol{X_i}$ representando o faturamento do posto no dia $\boldsymbol{i = 1, 2, 3, \dots, 60}$. Então

$$\mathbb{P}(\boldsymbol{X_1 + \cdots + X_{60} > 1230}) = \mathbb{P}\left(\frac{\boldsymbol{X_1 + \cdots + X_{60}}}{\boldsymbol{60}} > \frac{\boldsymbol{1230}}{\boldsymbol{60}}\right) \overset{\mathbf{1}}{=}$$

$$\overset{\mathbf{1}}{=} \mathbb{P}(\boldsymbol{\bar{X} > 20,5}) \approx \mathbb{P}\left(\boldsymbol{Z} > \frac{\boldsymbol{\sqrt{60}(20,5-20)}}{\boldsymbol{2}}\right) = \boldsymbol{0,0262}$$

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Distribuição Amostral da Proporção

Distribuição amostral da proporção: Considere-se uma população em que a distribuição de elementos portadores de determinada caraterística é p e defina-se a v.a. $\mathrm{X}=1$, se o elemento possui a característica ou $\mathrm{X}=0$, se o elemento não possui a característica. Logo,

$\mathrm{X} \sim Ber(p)$, $\mu=\mathbb{E}[\mathrm{X}]=p$ e $\sigma^2=\mathbb{V}[\mathrm{X}]=\mathrm{p}(1-\mathrm{p})$.

Retirada uma amostra aleatória de tamanho n de X, temos

- Y_n- é o total de elementos portadores da característica de interesse;
- $\hat{P}=\frac{\mathrm{Y}_n}{n}$ - é a proporção amostral de elementos portadores dessa característica, com $\mathrm{Y}_n \sim Bin(n, p)$.

DISTRIBUIÇÃO AMOSTRAL DA PROPORÇÃO

Retirada uma amostra aleatória de tamanho n de X, temos

- Y_n- é o total de elementos portadores da característica de interesse;
- $\hat{P} = \frac{\mathrm{Y}_n}{n} = \overline{\mathrm{X}}$ - é a proporção amostral de elementos portadores dessa característica, com $\mathrm{Y}_n \sim Bin(n, p)$

Nas condições do TLC, para n grande, podemos considerar a distribuição amostral de $\hat{P}$ como aproximadamente normal,

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

Distribuição Amostral da Proporção

Exemplo: Visitantes estrangeiros nos Estados Unidos são cuidadosamente monitorados por organizações de segurança e pela indústria do turismo. Durante o mês de março de 2008 foram monitorados 4,7 milhões de visitantes internacionais nos EUA. Quarenta por cento de todos os visitantes da Europa Ocidental eram do Reino Unido (RU). Suponha que sejam selecionados aleatoriamente 120 visitantes do mês de março, vindos da Europa Ocidental, e que se determine o número dos visitantes que vieram do RU.

i) Qual a distribuição da proporção amostral de visitantes do RU, $\hat{P}$?

ii) Qual a probabilidade de que a proporção amostral seja maior do que 0,50?

iii) Qual a probabilidade de que a proporção amostral esteja entre 0,32 e 0,37?

Distribuição Amostral da Proporção

Resolução do Exemplo: A amostra selecionada foi de 120 turistas, n = 120. A proporção de visitantes da Europa Ocidental, p = 0,40.

i) $\widehat{P} \sim \mathcal{N}\left(p = 0,40, \frac{p(1-p)}{n} = \frac{0,40 \times 0,60}{120}\right) \Rightarrow \widehat{P} \sim \mathcal{N}(0,40; 0,002)$

DISTRIBUIÇÃO AMOSTRAL DA PROPORÇÃO

Resolução do Exemplo: A amostra selecionada foi de 120 turistas, n = 120. A proporção de visitantes da Europa Ocidental, p = 0,40.

i) $$\widehat{P} \sim \mathcal{N}\left(p = 0,40, \frac{p(1-p)}{n} = \frac{0,40 \times 0,60}{120}\right) \Rightarrow \widehat{P} \sim \mathcal{N}(0,40; 0,002)$$

ii) $$\mathbb{P}(\widehat{P} > 0,50) = \mathbb{P}\left(\frac{\widehat{P} - 0,40}{\sqrt{0,002}} > \frac{0,50 - 0,40}{\sqrt{0,002}}\right) \overset{1}{=}$$

$$\overset{1}{=} \mathbb{P}(Z > 2,24) = 1 - \mathbb{P}(Z \leq 2,24) \overset{2}{=}$$

$$\overset{2}{=} 1 - 0,9875 = 0,0125.$$

Distribuição Amostral da Proporção

Resolução do Exemplo: A amostra selecionada foi de 120 turistas, n = 120. A proporção de visitantes da Europa Ocidental, p = 0,40.

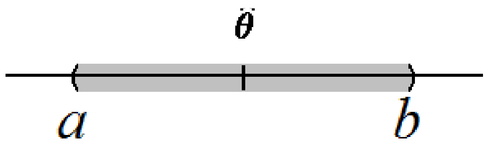
i) $\widehat{P} \sim \mathcal{N}\left(p = 0,40, \dfrac{p(1-p)}{n} = \dfrac{0,40 \times 0,60}{120}\right) \Rightarrow \widehat{P} \sim \mathcal{N}(0,40; 0,002)$

ii) $$\mathbb{P}(\widehat{P} > 0,50) = \mathbb{P}\left(\frac{\widehat{P} - 0,40}{\sqrt{0,002}} > \frac{0,50 - 0,40}{\sqrt{0,002}}\right) \overset{1}{=}$$
$$\overset{1}{=} \mathbb{P}(Z > 2,24) = 1 - \mathbb{P}(Z \leq 2,24) \overset{2}{=}$$
$$\overset{2}{=} 1 - 0,9875 = 0,0125.$$

iii) $$\mathbb{P}(0,32 \leq \widehat{P} \leq 0,37) = \mathbb{P}(-1,67 \leq Z \leq -0,67) \overset{1}{=}$$
$$\overset{1}{=} \mathbb{P}(Z \leq -0,67) - \mathbb{P}(Z \leq -1,67) \overset{2}{=}$$
$$\overset{2}{=} 0,2514 - 0,0367 = 0,2147.$$

Estimação por Intervalo

- No processo de "investigação" de um parâmetro θ, necessitamos ir além da estimativa pontual $\hat{\theta}$. Questionamentos quando não se conhece θ:
  - Quão próximo estamos do valor real de θ quando obtemos sua estimativa? Depende da precisão (ou variância) do estimador.
  - Uma maneira de contornar tal questionamento consiste em se encontrar um intervalo em torno de $\hat{\theta}$ que tenha alta probabilidade de englobar θ.



$$P(\text{ do intervalo } [a,b] \text{ englobar } \theta\ ) = \gamma$$

- O intervalo acima é construído com base na amostra dos dados, ou seja a partir dos dados e da distribuição amostral associada a $\hat{\theta}$.

INTERVALO DE CONFIANÇA PARA MÉDIA

Seja $X_1, \dots, X_n$ uma amostra de uma variável aleatória X com média μ é variância $\boldsymbol{\sigma^2}$**conhecida.** Para construir um intervalo de confiança para a média μ devemos considerar a distribuição da média amostral de $\bar{X}$.

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

O intervalo de confiança $\gamma = (1-\alpha)100\%$ para μ, temos que obter as constantes a e b, tal que $P(a \leq \mu \leq b) = (1-\alpha)$. A probabilidade $(1-\alpha)$ é chamada de **nível de confiança** do intervalo e α de **nível de significância**.

Assim, considerando ϵ um **erro amostral** ou **erro de estimação**, um intervalo de confiança para μ tem a forma $[\bar{X} - \epsilon; \bar{X} + \epsilon]$

$$\boxed{\frac{\epsilon\sqrt{n}}{\sigma} = z_{(1-\alpha/2)} \Rightarrow \epsilon = \frac{z_{(1-\alpha/2)}\sigma}{\sqrt{n}}}$$

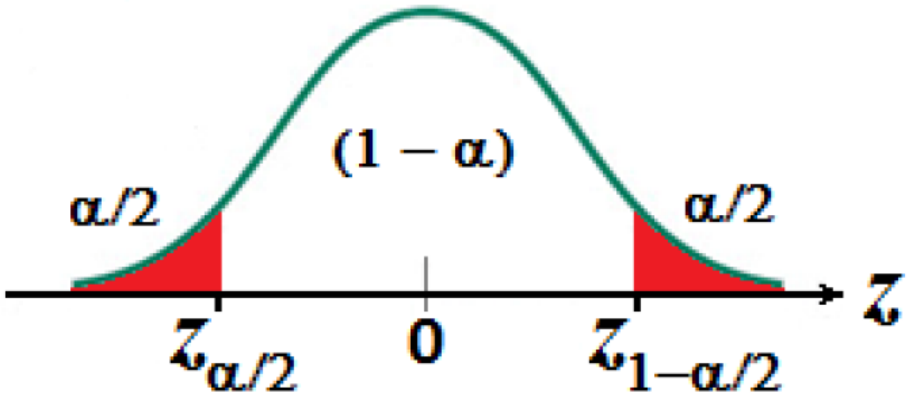
$$P(|\bar{X} - \mu| \leq \epsilon) = (1-\alpha)$$

$$\boxed{[\bar{X} - \frac{z_{(1-\alpha/2)}\sigma}{\sqrt{n}}; \bar{X} + \frac{z_{(1-\alpha/2)}\sigma}{\sqrt{n}}]}$$

INTERVALO DE CONFIANÇA PARA MÉDIA

Então da distribuição de $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$, temos

$$P\left(z_{\alpha/2} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z_{1-\alpha/2}\right) = (1-\alpha)$$

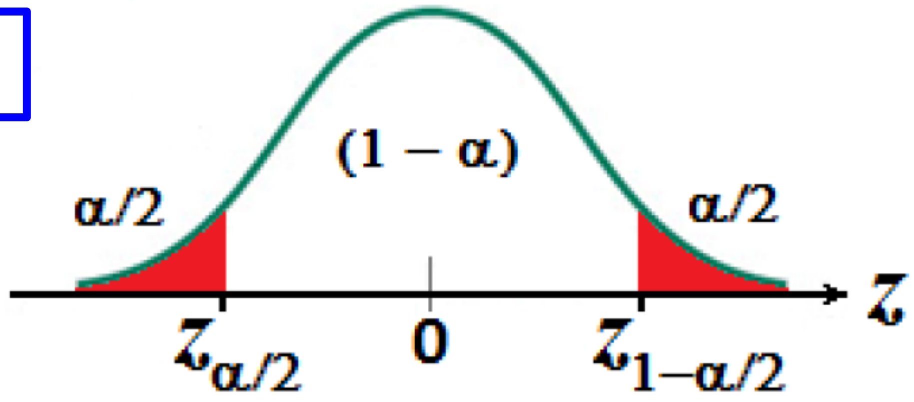


INTERVALO DE CONFIANÇA PARA MÉDIA

Então da distribuição de $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$, temos

$$P\left(z_{\alpha/2} \leq \boxed{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}} \leq z_{1-\alpha/2} \right) = (1-\alpha)$$

Quantidade Pivotal



INTERVALO DE CONFIANÇA PARA MÉDIA

Continuação:

$$\boxed{\epsilon}$$

$$P\left(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \ge \mu - \bar{X} \ge -z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = (1-\alpha)$$

$$P\left(\bar{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = (1-\alpha)$$

Como $z_{1-\alpha/2} = -z_{\alpha/2}$, teremos:

$$P\left(\underbrace{\bar{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}}_{a} \le \mu \le \underbrace{\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{b}\right) = (1-\alpha)$$

$$a = \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad \text{e} \quad b = \bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

INTERVALO DE CONFIANÇA PARA MÉDIA

Portanto, um intervalo de confiança $(1-\alpha)100\%$ para μ com σ^2 conhecido, é dado por

$$IC(\mu;\gamma) = \left[\bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \;;\; \bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right].$$

Se $\alpha = 0{,}05, {}^{\alpha}/_{2} = 0{,}025$ e $Z_{0,025} = -1{,}96$, logo um I.C. 95% para μ, com σ^2 conhecido, é dado por:

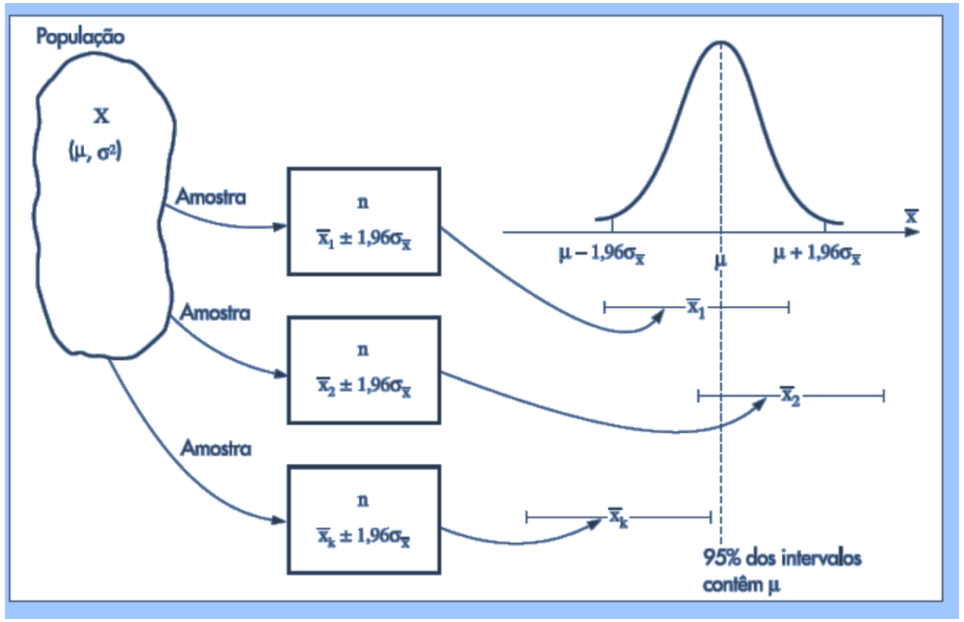
$$\left[\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \;;\; \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right].$$

Interpretação:

Se pudéssemos construir uma quantidade grande de intervalos (aleatórios!) da forma $\left[\bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \;;\; \bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$, *todos baseados em amostras de tamnho* n, *em média* $\gamma 100\%$ *deles conteriam o parâmetro* μ.

INTERVALO DE CONFIANÇA PARA MÉDIA

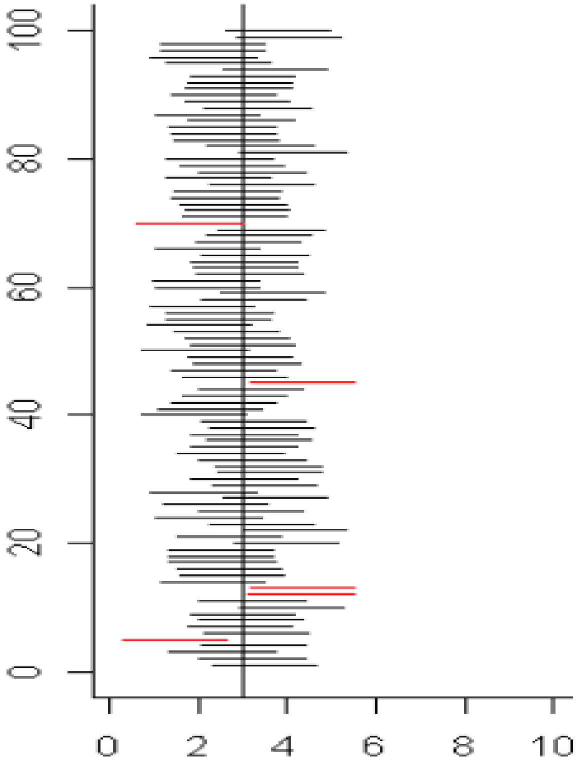
$N(3; 2{,}5^2)$



Normal

$\mu = 3 \quad \sigma = 2{,}5$

$n = 30, N_a = 100$



Intervalo de Confiança para Média

Exemplo: Testes de compressão foram aplicados em uma marca de cimento para avaliar sua resistência em concretos em MPa. Foram produzidos 13 corpos de prova e os testes foram aplicados no Laboratório de testes do Departamento de Engenharia Civil da UFRN.

- (O corpo de prova padrão brasileiro, normatizado pela ABNT, é o cilíndrico, com 15 cm de diâmetro, 30 cm de altura e o tempo de referência é 28 dias).

- Foi registrada a resistência à compressão simples (fc), para cada corpo de prova com o intuito de calcular a resistência característica do concreto à compressão (fck).

- Um concreto classe C30, por exemplo, corresponde a um concreto com fck = 30 Mpa (Mpa = 106Pa).

INTERVALO DE CONFIANÇA PARA MÉDIA

Continuação do exemplo: Dados dos 13 corpos de prova em Mpa

31.04	31.11	39.56	24.83	36.97	34.86	29.44
39.15	27.82	34.96	35.19	39.68	34.27	

A empresa afirma que o processo tem desvio padrão conhecido $\sigma = 5$MPa. Desta forma, deseja-se construir um intervalo de confiança 95% para a resistência à compressão média.

$$\bar{x} = 33.76$$

$$s = 4.665$$

Intervalo de Confiança para Média

Resolvendo o Exemplo:

Conhecemos a estatística: $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

$$P\left(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \geq \mu - \bar{X} \geq -z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = (1-\alpha)$$

$$\left[\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \; ; \; \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right].$$

Logo, [31,04; 36,48] é um I.C. 95% para a média μ.

Podemos dizer que aproximadamente 95% das médias μ
das amostras de tamanho 13 se encontram entre [31,04; 36,48].

INTERVALO DE CONFIANÇA PARA PROPORÇÃO

Como a proporção p é de fato a média amostral de uma a.a. cuja v.a. tem distribuição de Bernoulli(p), para se construir intervalos de confiança para p devemos seguir os mesmos procedimentos anteriores.

Considerando que o estimador da proporção $\hat{P}$ tem valor esperado p e variância $\frac{p(1-p)}{n}$ com distribuição

$$\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0\,;1),$$

Portanto, um intervalo de confiança $(1-\alpha)100\%$ para p, é dado por

$$\left[\hat{p}+z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\ ;\ \hat{p}-z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right]$$

INTERVALO DE CONFIANÇA PARA PROPORÇÃO

Exemplo: Testes de compressão em amostras de concreto, a empresa afirma que 90% da produção atende ao valor do fck = 30 Mpa. Desta forma, construir um I.C. de 95% para a proporção de corpos de prova com fc abaixa de fck.

31.04	31.11	39.56	$\boxed{24.83}$	36.97	34.86	$\boxed{29.44}$
39.15	$\boxed{27.82}$	34.96	35.19	39.68	34.27	

Dos 13 corpos de prova os valores 24.83, 29.44 e 27.82 são os menores do que o fck de 30Mpa. Então $\hat{p} = \frac{3}{13} = 0{,}231$.

INTERVALO DE CONFIANÇA PARA PROPORÇÃO

Resolução do Exemplo: considerando $p = 0.10$ são os 10% que não atendem aos valores segunda a empresa.

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} = 0.231 - 1.96\sqrt{\frac{0.10 \cdot 0.90}{13}} = 0.0679$$

$$\hat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} = 0.231 + 1.96\sqrt{\frac{0.10 \cdot 0.90}{13}} = 0.3941$$

Ou seja: $\mathbb{P}(0.0679 \leq p \leq 0.3941) = 0.95$

Portanto, [0.0679; 0,3941] é um I.C. 95% para a média p.

Podemos dizer que aproximadamente 95% das proporções p das amostras de tamanho 13 se encontram entre [0.0679; 0,3941].

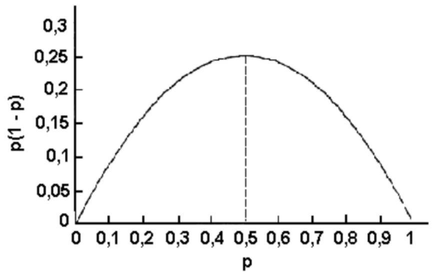
INTERVALO DE CONFIANÇA PARA PROPORÇÃO

Em geral não temos conhecimento sobre p, assim podemos construir intervalos de confiança para a proporção substituindo p e $(1-p)$ por $\hat{p}$ e $(1-\hat{p})$, respectivamente. Neste caso um intervalo de confiança $(1-\alpha)100\%$ para p, é dado por

$$\left[\hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \;\; ; \;\; \hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

Outra possibilidade é e construir um I.C. conservador para p assumindo $p = {}^1\!/_2$. Assim, $\frac{p(1-p)}{n} = \frac{1}{4n}$ com I.C. conservador para p

$$\left[\hat{p} + \frac{z_{\alpha/2}}{\sqrt{4n}} \;\; ; \;\; \hat{p} - \frac{z_{\alpha/2}}{\sqrt{4n}}\right]$$



Intervalo de Confiança para Proporção

Resolução do Exemplo:

- considerando $\hat{p} = 0.231$

$$\hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.231 - 1.96\sqrt{\frac{0.231 \cdot 0.769}{13}} = 0.0019$$

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.231 + 1.96\sqrt{\frac{0.231 \cdot 0.769}{13}} = 0.4601$$

Ou seja: $\mathbb{P}(0.0019 \leq p \leq 0.4601) = 0.95$

Portanto, [0.0019; 0,4601] é um I.C. 95% para a média p.

INTERVALO DE CONFIANÇA PARA PROPORÇÃO

Resolução do Exemplo:

- considerando $p = \frac{1}{2}$

$$\hat{p} + \frac{z_{\alpha/2}}{\sqrt{4n}} = 0.231 - \frac{1.96}{\sqrt{52}} = -0.0408 \qquad \textbf{(< 0 !!)}$$

$$\hat{p} - \frac{z_{\alpha/2}}{\sqrt{4n}} = 0.231 + \frac{1.96}{\sqrt{52}} = 0.5028$$

Portanto, [-0.0408; 0,5028] é um I.C. 95% para a média p.

Intervalo de Confiança para Proporção

Resolução do Exemplo:

- IC 95% considerando $p = 0{,}10$: [0.0679; 0,3941]
- IC 95% considerando $\hat{p}$: [0.0019; 0,4601]
- IC 95% considerando $p = 1/2$: [-0.0408; 0,5028]

$\Downarrow$

[0; 0,5028]

TAMANHO DE UMA AMOSTRA

Chamamos de **erro amostral** (da média ou da proporção) e designamos por *e* a v.a. que mede a diferença entre a estatística e o parâmetro. Assim,

$$e = \overline{X} - \mu \qquad \text{ou} \qquad e = \hat{p} - p.$$

Tendo-se

$$e \sim N\left(0, \frac{\sigma^2}{n}\right) \qquad \text{ou} \qquad e \sim N\left(0, \frac{p(1-p)}{n}\right).$$

Suponhamos que pretendemos determinar o valor de *n* que satisfaz

$$P(|\overline{X} - \mu|) \leq \varepsilon) \geq \gamma$$

com $0 < \gamma < 1$ e ε é o erro amostral (observado) máximo que podemos suportar. Definindo $\gamma = P(-z_\gamma < Z < z_\gamma)$, com $z_\gamma = \frac{\sqrt{n}\,\varepsilon}{\sigma}$, obtemos

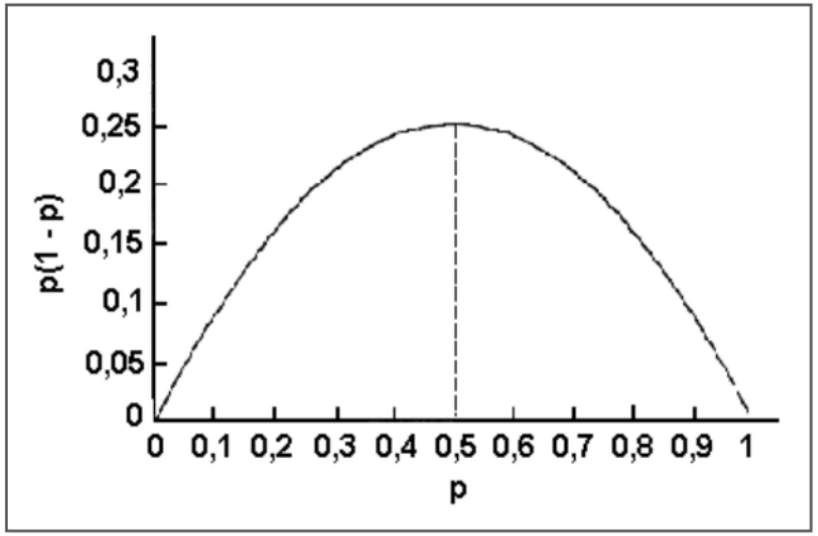
$$n = \frac{\sigma^2 \, z_\gamma^2}{\varepsilon^2}.$$

No caso de proporções,

$$n = \frac{p(1-p) \, z_\gamma^2}{\varepsilon^2}.$$

TAMANHO DE UMA AMOSTRA

Observando o gráfico da função $p(1-p)$, para $0 \leq p \leq 1$



vemos que:

- a função $p(1-p)$ é uma parábola simétrica em torno de $p = 0,5$;
- a função atinge o seu máximo (0,25) quando $p = 0,5$.

Deste modo, na prática, **se nada sabemos sobre o valor de** p, substituimos $p(1-p)$ por 0,25 obtendo-se

$$n = 0,25 \left(\frac{z_\gamma}{\varepsilon}\right)^2,$$

que pode fornecer um valor de n maior do que o necessário.

RESUMO DA AULA

- Estudo da distribuição amostral da média e da proporção;
- Construção de intervalo de confiança para média;
- Construção de intervalo de confiança para proporção;
- Resolução de exercicíos sobre o assunto abordado.

PRÓXIMAS AULAS

- Intervalo de confiança para a media com variância desconhecida;
- Intervalo de confiança para a diferença entre as medias de duas populações independentes;
- Intervalo de confiança para a diferênça entre proporções;
- Intervalo de confiança para a variância.