Universidade Federal de Mato Grosso

AULA 06 – Estatística

Prof. Lucas Bianchi

Cuiabá, 05 de setembro de 2016

Na aula de hoje, estudaremos:

- ✓ Distribuição Amostral da média
- ✓ Distribuição Amostral da proporção
- ✓ Intervalos de Confiança para média e proporção.

É um processo de indução, na qual usamos dados extraídos de uma amostra para produzir inferência sobre a população. Esta inferência só será válida se a amostra for significativa.

Tipos de Estimações:

- Estimação Pontual;
- Estimação Intervalar.

Estimação Pontual

É usada quando a partir da amostra procura-se obter um único valor de certo parâmetro populacional, ou seja, obter estimativas a partir dos valores amostrais.

As estimativas são os valores amostrais obtidos para a média, variância, proporção, etc. Os valores de $\bar{X}$, S^2 , S estimam, respectivamente μ, σ^2 e σ.

Estimação Intervalar

Outra maneira de se calcular um estimativa de um parâmetro desconhecido, é construir um intervalo de confiança [a;b] para esse parâmetro com uma probabilidade de 1 - α (nível de confiança) de que o intervalo contenha o verdadeiro parâmetro, usando as distribuições de amostragem podemos obter expressões do tipo:

$$P(a \leq \mu \leq b) = 1 - \alpha$$

Dessa maneira α será o nível de significância, isto é, o erro que se estará cometendo ao afirmar que o parâmetro está entre o limite inferior e o superior calculado.

Intervalo de Confiança (IC)

É uma faixa de possíveis valores em torno da média amostral ou proporção amostral, e a probabilidade de que esta faixa seja contemplado com o valor real da média da população é determinado;

Nível de confiança (1-α): é a probabilidade de que o intervalo estimado contenha o parâmetro populacional;

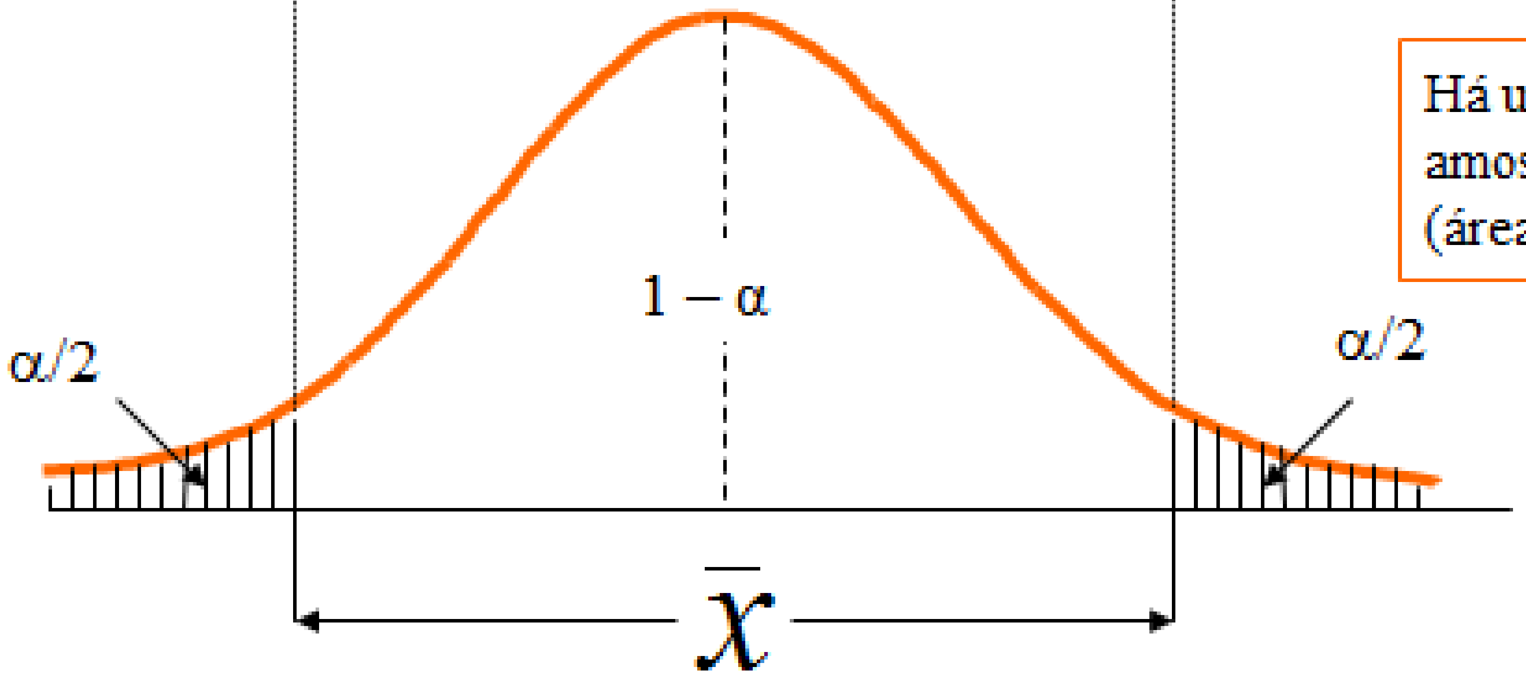
Nível de significância (α): é o erro que se comete ao afirmar que o parâmetro está entre o limite inferior e o superior calculado.

Há uma probabilidade de $1 - \alpha$ da média estar contida no intervalo definido

$1 - \alpha$ = nível de confiança

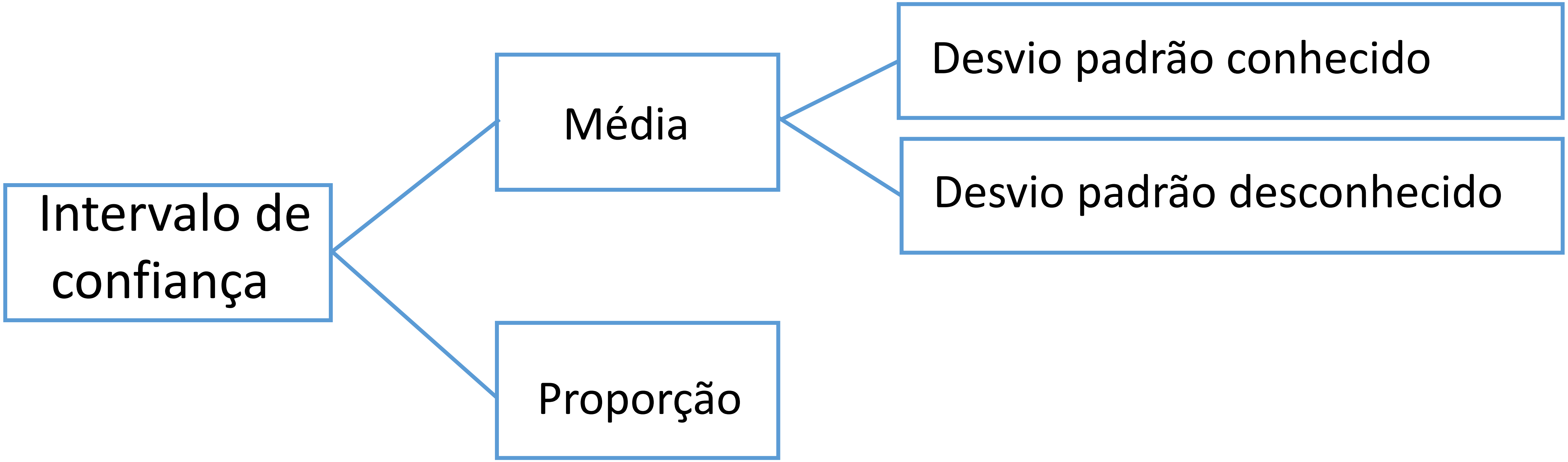
α = nível de significância (probabilidade de erro)

Há uma probabilidade α de a média amostral estar fora do intervalo definido (área hachurada)



Intervalo de confiança

Intervalo de Confiança (IC)



Intervalo de Confiança para a média populacional

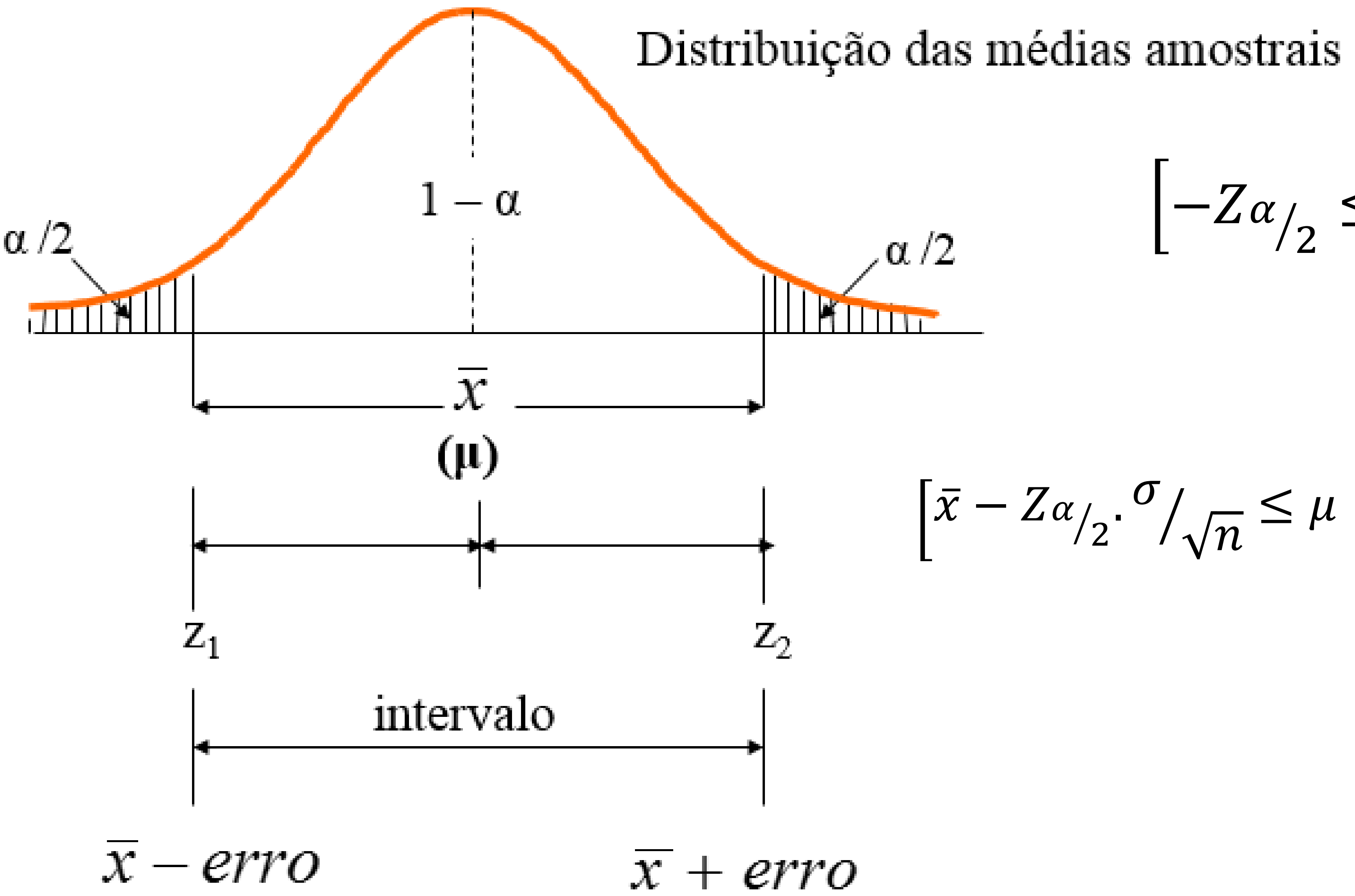
Quando o interesse é estimar a média de um parâmetro da população, por meio de uma amostra. Neste caso deve ser levado em consideração dois casos distintos:

- Desvio padrão da população conhecido;
- Desvio padrão da população desconhecido.

IC para média com Desvio Padrão conhecido

Considere uma v.a. simples $x_1, x_2, \dots, x_n$ obtida de uma população com distribuição normal, com média μ e variância σ^2 conhecida. Para construir o IC deve se considerar a distribuição da média amostral também é normal (μ) e variância σ^2/n.

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ assim temos que: } Z = \frac{\bar{x}-\mu}{\frac{\sigma}{n}} \sim N(0,1)$$



$$\left[-Z\alpha/_2 \leq Z \leq +Z\alpha/_2\right] = 1 - \alpha$$

$$\left[\bar{x} - Z\alpha/_2 . \sigma/_{\sqrt{n}} \leq \mu \leq \bar{x} + Z\alpha/_2 . \sigma/_{\sqrt{n}}\right] = 1 - \alpha$$

Exemplo 1: O projetista de uma indústria tomou uma amostra de 36 funcionários para verificar o tempo médio gasto para montar um brinquedo: $\bar{x}$=19,9 e σ = 5,73. Calcule o IC para o tempo médio gasto na montagem do brinquedo ao nível de 5% de significância.

$$\left[\bar{x} - Z\alpha_{/2} . \sigma/_{\sqrt{n}} \leq \mu \leq \bar{x} + Z\alpha_{/2} . \sigma/_{\sqrt{n}}\right] = 1 - \alpha$$

$$\left[19{,}9 - 1{,}96\left(\frac{5{,}73}{\sqrt{36}}\right) \leq \mu \leq 19{,}9 + 1{,}96\left(\frac{5{,}73}{\sqrt{36}}\right)\right] = 1 - 0{,}05$$

$$[18{,}02 \leq \mu \leq 21{,}77] = 0{,}95$$

IC para média com Desvio Padrão deconhecido

Vamos tratar de uma situação mais realista, quando a variância σ^2 da população é desconhecida. Considere uma v.a. simples $x_1, x_2, \ldots, x_n$ obtida de uma população com distribuição normal, com média μ e variância σ^2 desconhecida. Desta forma utilizaremos a variância amostral s^2 no lugar de σ^2, assim temos que:

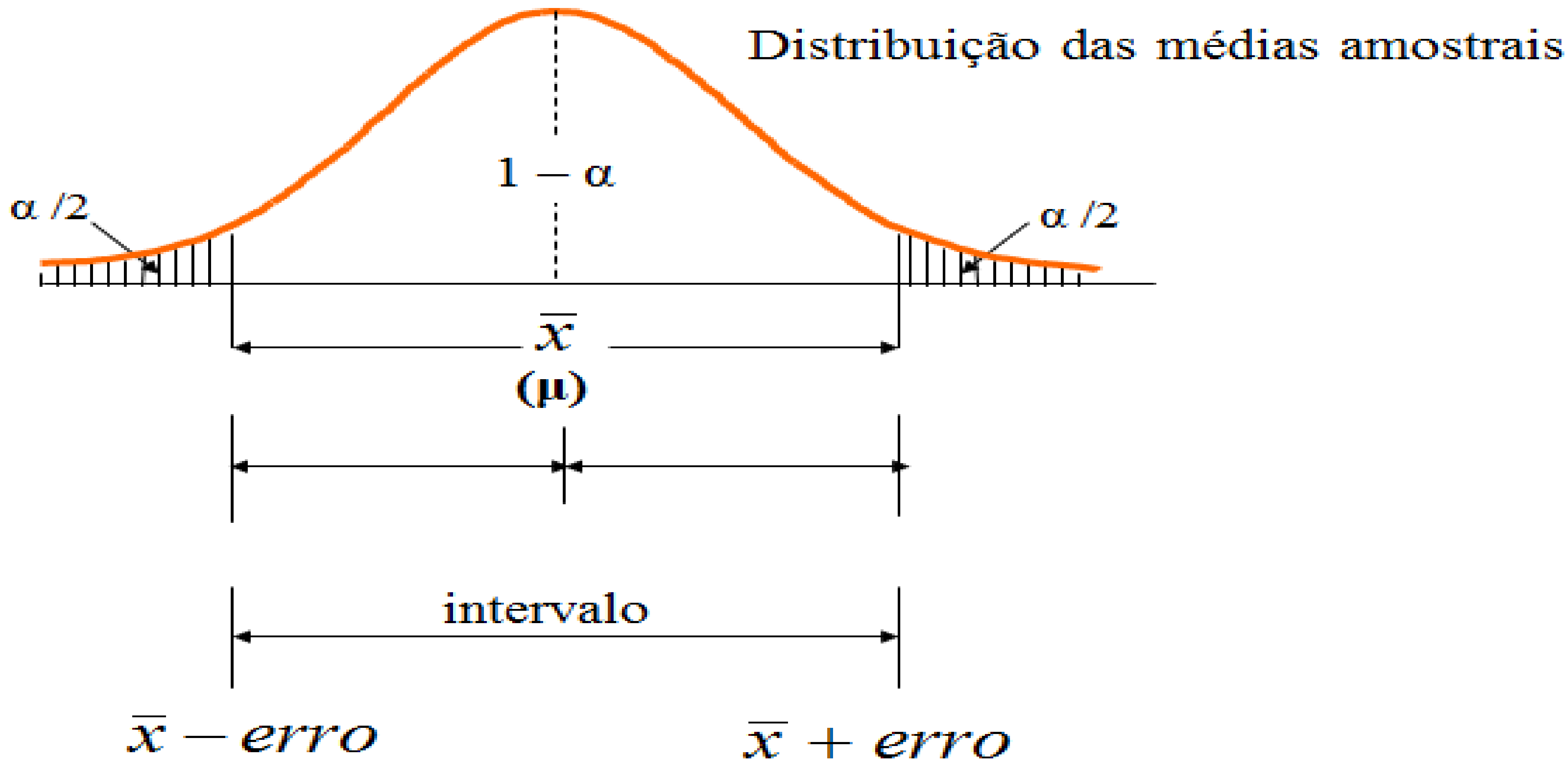
$$\mathrm{T} = \frac{\bar{x}-\mu}{\frac{S}{n}} \sim t_{n-1}$$

ou seja, T segue uma distribuição t de Student com n-1 graus de liberdade.

Ao fixarmos o nível de significância α, obtemos na tabela t-Student com n-1 graus de liberdade o valor $t_{(\alpha/_{2},\, n-1)}$ que satisfaz:

$$\left[-t_{(\alpha/_{2},\, n-1)} \le T \le t_{(\alpha/_{2},\, n-1)}\right] = 1 - \alpha$$

$$\left[\bar{x} - t_{(\alpha/_{2},\, n-1)}.{}^{s}/_{\sqrt{n}} \le T \le \bar{x} + t_{(\alpha/_{2},\, n-1)}.{}^{s}/_{\sqrt{n}}\right] = 1 - \alpha$$



Distribuição das médias amostrais
$1-\alpha$
$\alpha/2$
$\alpha/2$
$\bar{x}$
(μ)
intervalo
$\bar{x} - erro$
$\bar{x} + erro$

Exemplo 2: Foram realizados testes glicêmicos em 25 pacientes após um jejum de 8 horas. Os resultados estão apresentados na tabela abaixo. Encontrar um intervalo de 95% confiança para média μ.

Teste glicêmico (mg/dL)				
80	118	100	90	83
117	95	84	102	80
112	78	102	121	82
77	88	73	104	88
132	91	103	140	101

$$\bar{X} = 97{,}64 \quad S = 17{,}82$$

$$P\left[\bar{X} - t_{\left(\alpha/_{2},\, n-1\right)} \cdot {}^{s}/_{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\left(\alpha/_{2},\, n-1\right)}\right] = 1 - \alpha$$

$$P\left[97{,}64 - 2{,}064\left({}^{17{,}82}/_{\sqrt{25}}\right) \leq \mu \leq 97{,}64 + 2{,}064\left({}^{17{,}82}/_{\sqrt{25}}\right)\right] = 1 - 0{,}05$$

$$P[90{,}2839 \leq \mu \leq 104{,}9961] = 0{,}95$$

$$IC(90{,}28; 104{,}99) = 0{,}95$$

Interpretação: Podemos afirmar com 95% de confiança que o intervalo (90,28 ; 104,99) é um dos possíveis intervalos que contempla o índice glicêmico após jejum de 8 horas na população em estudo.

g.l.	0,200	0,150	0,100	Ps 0,050	0,025	0,010	0,005
1	1,376	1,963	3,078	6,314	12,71	31,82	63,66
2	1,061	1,386	1,886	2,920	4,303	6,965	9,925
3	0,978	1,250	1,638	2,353	3,182	4,541	5,841
4	0,941	1,190	1,533	2,132	2,776	3,747	4,604
5	0,920	1,156	1,476	2,015	2,571	3,365	4,032
6	0,906	1,134	1,440	1,943	2,447	3,143	3,707
7	0,896	1,119	1,415	1,895	2,365	2,998	3,499
8	0,889	1,108	1,397	1,860	2,306	2,896	3,355
9	0,883	1,100	1,383	1,833	2,262	2,821	3,250
10	0,879	1,093	1,372	1,812	2,228	2,764	3,169
11	0,876	1,088	1,363	1,796	2,201	2,718	3,106
12	0,873	1,083	1,356	1,782	2,179	2,681	3,055
13	0,870	1,079	1,350	1,771	2,160	2,650	3,012
14	0,868	1,076	1,345	1,761	2,145	2,624	2,977
15	0,866	1,074	1,341	1,753	2,131	2,602	2,947
16	0,865	1,071	1,337	1,746	2,120	2,583	2,921
17	0,863	1,069	1,333	1,740	2,110	2,567	2,898
18	0,862	1,067	1,330	1,734	2,101	2,552	2,878
19	0,861	1,066	1,328	1,729	2,093	2,539	2,861
20	0,860	1,064	1,325	1,725	2,086	2,528	2,845
21	0,859	1,063	1,323	1,721	2,080	2,518	2,831
22	0,858	1,061	1,321	1,717	2,074	2,508	2,819
23	0,858	1,060	1,319	1,714	2.069	2,500	2,807
24	0,857	1,059	1,318	1,711	2.064	2,492	2,797

Intervalo de Confiança para a proporção populacional

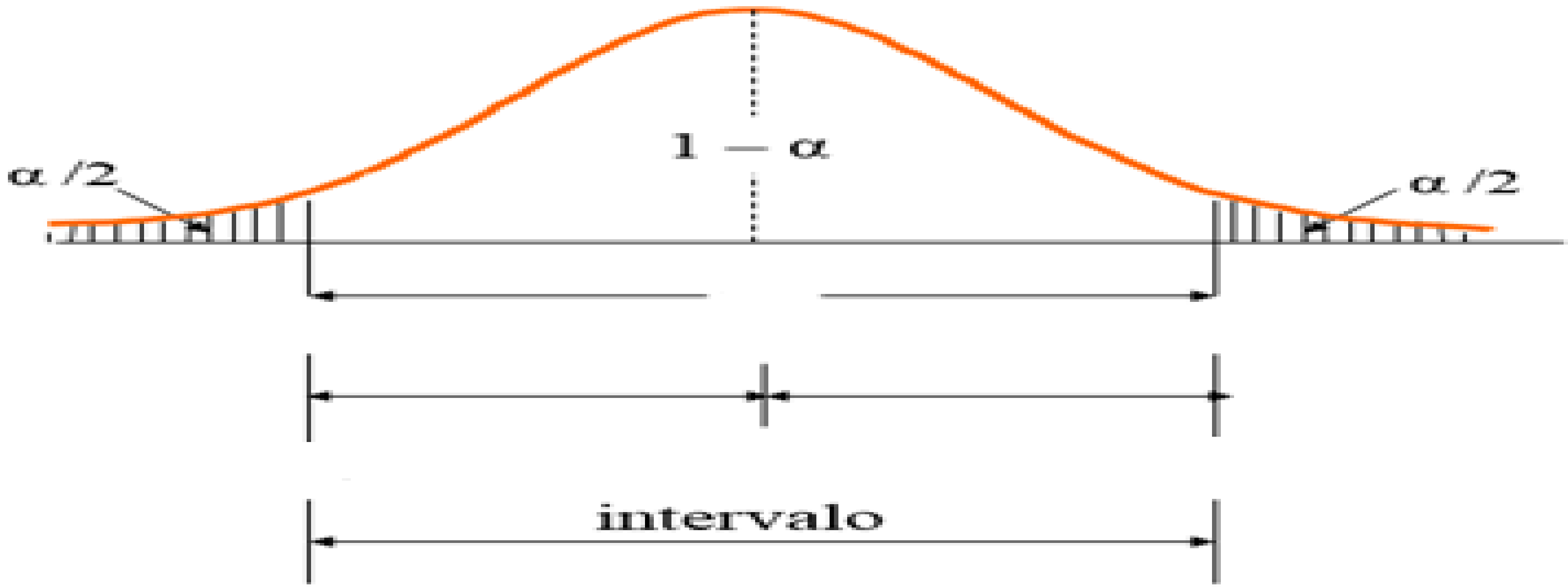
Consideremos uma população cujos elementos podem ser classificados em dois tipos: Sucesso e Insucesso. Pretende-se estimar a proporção p de sucessos na população.

Dada uma amostra de tamanho n, uma estimativa pontual de p da proporção de sucessos é dada por:

$$\hat{p} = \frac{x}{n}$$

Quando n for suficientemente grande $\hat{p}$ tem distribuição aproximadamente normal, com média $\mu_{\hat{p}} = p$ e variância $\sigma^2_{\hat{p}} = \frac{pq}{n}$ em que:

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$



$$\hat{p} - Z\alpha_{/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} \qquad \hat{p} + Z\alpha_{/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Fixando uma probabilidade de confiança (1- α), o intervalo de confiança para uma proporção pode ser obtido da seguinte forma:

$$P\left(\hat{p} - Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} \leq p \leq \hat{p} + Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) = 1 - \alpha$$

onde: $Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$ é a margem de erro da proporção e $Z_{\alpha/2}$ é o valor da curva normal padrão acima do qual encontramos uma área de $\alpha/2$.

Exemplo 3: Uma empresa de pesquisa de mercado faz contato com 30 pessoas para saber a satisfação a uma determinada marca de refrigerante, 12 delas respondem que gosta da referida marca. Obtenha o intervalo de confiança de 95% para proporção de pessoas que gostam da marca.

$$\hat{p} = \frac{x}{n} = \frac{12}{30} = 0{,}40$$

Na tabela Z tem se que para 95% de confiança que $Z\alpha/_2 = 1{,}96$

$$P\left(\hat{p} - Z\alpha/_2\sqrt{\frac{\hat{p}\hat{q}}{n}} \leq p \leq \hat{p} + Z\alpha/_2\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) = 0{,}95$$

$$P\left(0{,}40 - 1{,}96\sqrt{\frac{0{,}40\,X0{,}60}{30}} \leq p \leq 0{,}40 + 1{,}96\sqrt{\frac{0{,}40\,X0{,}60}{30}}\right) = 0{,}95$$

$$P(0{,}40 - 0{,}1753 \leq p \leq 0{,}40 + 0{,}1753) = 0{,}95$$

$$P(0{,}22 \leq p \leq 0{,}57) = 0{,}95$$

$IC_{95\%}(\mu)=[0{,}22 ; 0{,}57]$

Exercicio 1: Para estimar a proporção p de pessoa acometidas por uma certa gripe numa população, uma amostra aleatória simples de 1600 pessoas foi observada e constatou-se que, dessas pessoas, 160 estavam com gripe. Um intervalo aproximado de 95% de confiança para p será dado por:

a) (0,066;0,134) $\boxed{\text{b) } (0{,}085;0{,}114)}$

c) (0,058;0,142) d) (0,091;0,142)

e) (0,034;0,166)

Exercicio 2: Uma amostra aleatória simples de tamanho 400 de uma variável populacional normalmente distribuída com média desconhecida e variância igual a 25 foi observada e indicou uma média amostral igual a 12,52. O intervalo de 95% de confiança para é dado por:

a) (12,03;13,01) b) (11,65;13,39)

c) (10,99;15,05) d) (10,44;15,60)

e) (9,99;16,05)

Exercicio 3: A equipe de vendas imobiliária apresenta semanalmente relatórios que relacionam os contatos feitos com clientes durante a semana. Uma amostra de 65 relatórios semanais exibiu uma média amostral de 19,5 contatos com clientes por semana. O desvio padrão da amostra foi 5,2. Forneça os intervalos de confiança de 90% e 95% correspondentes ao número médio da população de contatos semanais com clientes feitos pela equipe de vendas.

$$IC\ 90\% = \bar{x} \pm Z_{\alpha/2} * \frac{\sigma}{\sqrt{n}} = IC(18{,}44;\ 20{,}55)$$

$$IC\ 95\% = \bar{x} \pm Z_{\alpha/2} * \frac{\sigma}{\sqrt{n}} = IC(18{,}23;\ 20{,}76)$$

Exercício 4: Uma pesquisa de 611 funcionários de escritório investigou seus hábitos de atendimento ao telefone, incluindo a frequência com que cada funcionário de escritório era capaz de atender às chamadas telefônicas e com qual frequência as chamadas telefônicas chegavam diretamente ao correio de voz (*USA Today,* 21 de abril de 2002). Ao todo, 281 funcionários de escritório indicaram que nunca precisavam do correio de voz e que eram capazes de responder a cada chamada telefônica.

a) Qual é a estimação por ponto da proporção da população de funcionários de escritório que são capazes de

atender a cada chamada telefônica?

b) Com 90% de confiança, qual é a margem de erro? **(Dica: Binomial)**

c) Qual é o intervalo de confiança de 90% da proporção da população de funcionários de escritório que são capazes de atender a cada chamada telefônica?

a)

$$p = \frac{281}{611} = 0,4599, logo\ q = 1\ - p = 1\ - 0,4599 = 0,5401$$

b)

$$Var(x) = npq = 611 * 0,46 * 0,54 = 151,768$$

$$Dp(x) = \sqrt{151,768} = 12,3194$$

$$E = \frac{12,3194}{\sqrt{611}} = 0,0202$$

c)

$$IC\ =\ (0,42; 0,49)$$