Análise de Dados

Distribuições de Amostragens

Prof. Dr. Daniel Caetano

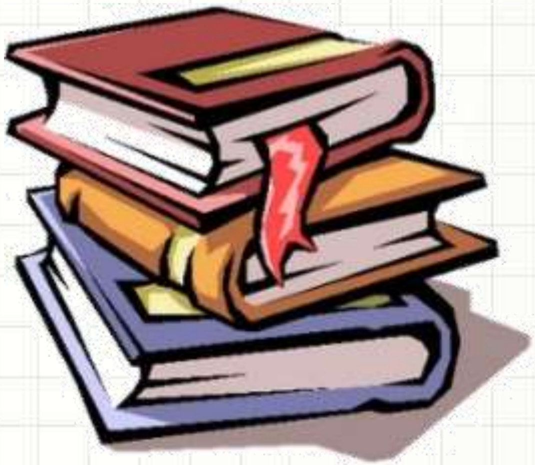
2020 - 1

Objetivos

- Compreender o conceito de Distribuição de Amostragens
- Trabalhar com as distribuições de amostragens da média aritmética e da proporção
- Compreender o Teorema do Limite Central

- **Atividade da Aula 10 no SAVA!**

Material de Estudo

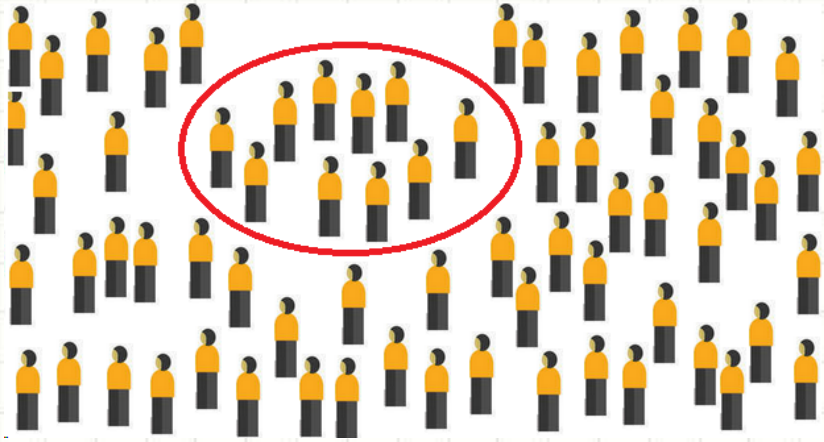


Material	Acesso ao Material
Apresentação	http://www.caetano.eng.br/ (Análise de Dados – Aula 10)
Minha Biblioteca	Estatística – Teoria e Aplicações usando MS Excel – Cap. 7 Estatística Aplicada e Probabilidade para Engenheiros – Cap. 7

Distribuições de Amostragens

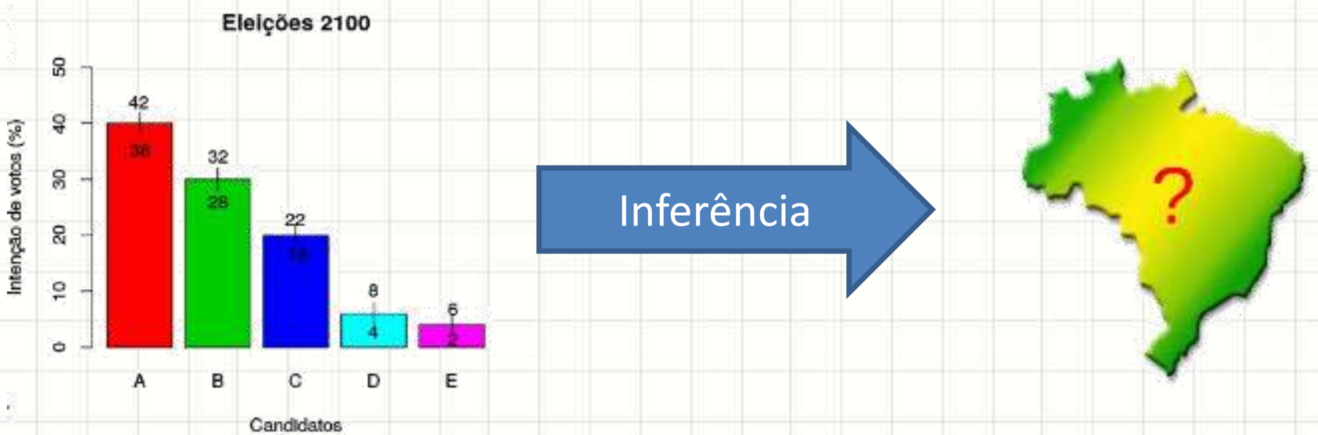
Da Amostra à População

- Coletamos e analisamos a amostra
  - Geramos estatísticas
  - Uma delas é a média aritmética amostral.
- Desejamos inferir sobre a população
  - Encontrar os parâmetros
  - Um deles é a média populacional.



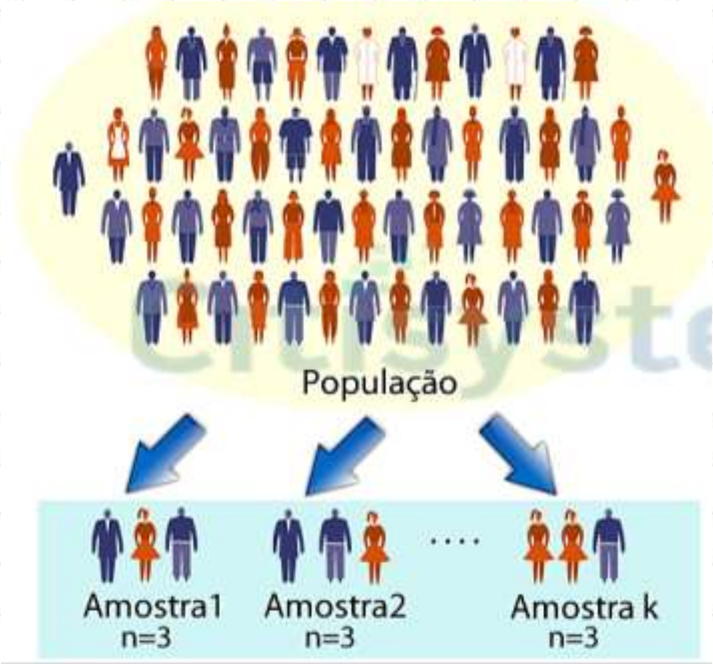
Da Amostra à População

- Podemos querer o mesmo para proporções
  - Que proporção da população tem a característica
  - Conhecendo a proporção da amostra?
- Exemplo:
  - Pesquisa eleitoral – proporção da amostra...
  - Leva a qual proporção da população?



Da Amostra à População

- Teoria: analisar todas as amostras possíveis
  - De um determinado tamanho
  - Distribuição de Amostragens.
- Na prática: só o resultado de uma amostra

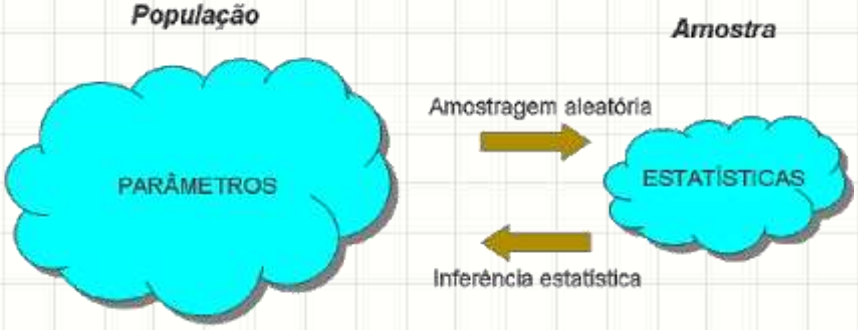


Distribuição de Amostragem da

Média Aritmética

D.A. da Média Aritmética

- Há diversas medidas de tendência central
  - Média aritmética é a mais amplamente usada.
- Média da amostra...
  - Usada para estimar a média da população.
- Distribuição de Amostragem da Média
  - Distribuição da média de todas as amostras
    - De um mesmo tamanho



Ausência de Viés

- Média das médias amostrais = média da pop.
  - Hein?
- Exemplo:
  - Imagine que temos 3 estagiários que foram indicados a realizar uma tarefa. A tabela abaixo indica o número de erros de cada um deles:

Estagiário	Número de Erros
Adão	3
João	2
Lila	4

Ausência de Viés - Exemplo

Estagiário	Erros
Adão	3
João	2
Lila	4

- Média?

$$\mu = \frac{3+2+4}{3} = \frac{9}{3} = 3\ erros$$

- Desvio Padrão?

$$\sigma = \sqrt{\frac{(0)^2+(-1)^2+(1)^2}{3}} = \sqrt{\frac{2}{3}} \cong 0{,}82\ erro$$

Ausência de Viés - Exemplo

Estagiário	Erros
Adão	3
João	2
Lila	4

- Vamos agora analisar por amostras $\mu = 3 \quad \sigma \cong 0{,}82$
  – Suponhamos todas as de 2 elementos

Amostra	Estagiários	Dados	Média
1	Adão, Adão	3; 3	3
2	Adão, João	3; 2	2,5
3	Adão, Lila	3; 4	3,5
4	João, Adão	2; 3	2,5
5	João, João	2; 2	2
6	João, Lila	2; 4	3
7	Lila, Adão	4; 3	3,5
8	Lila, João	4; 2	3
9	Lila, Lila	4; 4	4

E σ?

$$\mu = \frac{3 + 2{,}5 + 3{,}5 + 2{,}5 + 2 + 3 + 3{,}5 + 3 + 4}{9} = \frac{27}{9} = 3\ erros$$

!

Ausência de Viés - Exemplo

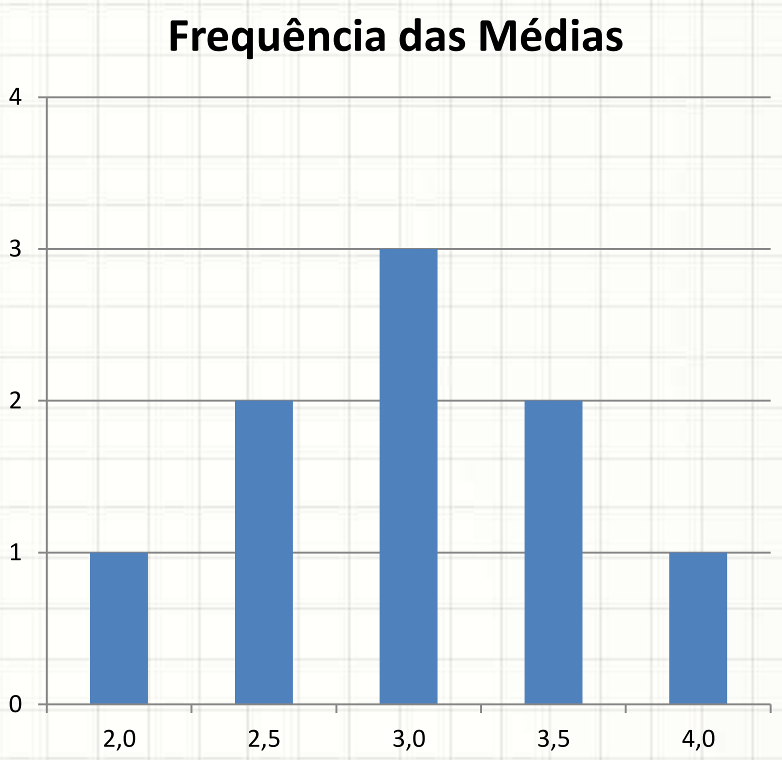
Estagiário	Erros
Adão	3
João	2
Lila	4

- E a distribuição das médias?

$\mu = 3 \quad \sigma \cong 0{,}82$

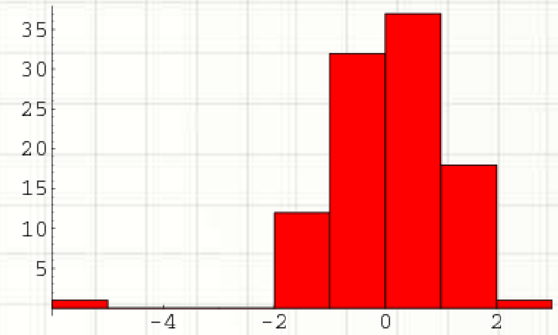
Amostra	Média
1	3
2	2,5
3	3,5
4	2,5
5	2
6	3
7	3,5
8	3
9	4

Valor	Freq.
2	1
2,5	2
3	3
3,5	2
4	1



Considerações sobre D.A. da Média

- Média de amostras...
  - Varia menos que o valor das amostras
- Se amostra pequena contém valor extremo
  - Isso impacta muito na média dela.
- Esse impacto diminui...
  - Quanto maior for a amostra!.
- Erro-padrão...
  - Desvio para todas as amostras possíveis



Erro-Padrão da Média Aritmética

- Erro-padrão...
  - Desvio para todas as amostras possíveis

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

    - σ é o desvio padrão na população
    - n é o tamanho da amostra.

- Equação vale para...
  - Amostra sem reposição
  - Amostra inferior a 5% da população.

Exemplo

- Considerando que o desvio padrão de um processo de abastecimento de cereais é de 15g, calcule o erro padrão para uma amostra de 25 caixas, em considerando uma produção de 10.000 caixas.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{25}} = \frac{15}{5} = 3$$

Amostragem: População “Normal”

- Se a população tem uma distribuição normal
  - A média aritmética das amostras também será
  - Com erro padrão

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

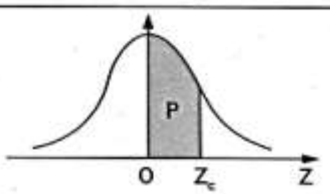
- Nesse caso, a equação vale...
  - Independente do tamanho da amostra!

Amostragem: População “Normal”

- Para encontrar a probabilidade de uma amostra ter uma média $\bar{x}$:

$$Z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Tabela III — Distribuição Normal Padrão
$Z \sim N(0, 1)$
Corpo da tabela dá a probabilidade p, tal que $p = P(0 < Z < Z_c)$



parte inteira e primeira decimal de Z_c	Segunda decimal de Z_c: 0	1	2	3	4	5	6	7	8	9	parte inteira e primeira decimal de Z_c
	p = 0										
1,0	34134	34375	34614	35769	35993	36214	1,0	5769	35993	36214	1,0
1,1	36433	36650	36864	37900	38100	38298	1,1	7900	38100	38298	1,1
1,2	38493	38686	38877	39796	39973	40147	1,2	9796	39973	40147	1,2
1,3	40320	40490	40658	41466	41621	41774	1,3	1466	41621	41774	1,3
1,4	41924	42073	42220	42922	43056	43189	1,4	2922	43056	43189	1,4

Exemplo

- Considerando que a média de conteúdo de uma caixa é 368g com distribuição normal e desvio padrão 15g, qual a probabilidade de uma amostra de 25 caixas possuir menos que 365g?

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{365 - 368}{\frac{15}{\sqrt{25}}} = \mathbf{-1,00}$$

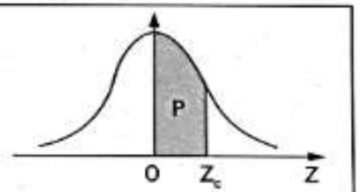
$$P(P < 365) = 0,5 - 0,34134$$

$$P(P < 365) = 0,15866$$

Tabela III — Distrib

$Z \sim$

Corpo da tabela dá a probabili



parte inteira e primeira decimal de Z_c	0	1	2	7	8	9	parte inteira e primeira decimal de Z_c
	p = 0						
1,0	34134	34375	34614	35769	35993	36214	1,0
1,1	36433	36650	36864	37900	38100	38298	1,1
1,2	38493	38686	38877	39796	39973	40147	1,2
1,3	40320	40490	40658	41466	41621	41774	1,3
1,4	41924	42073	42220	42922	43056	43189	1,4
1,5	43319	43448	43574	44179	44295	44408	1,5
1,6	44520	44630	44738	45254	45352	45449	1,6
1,7	45543	45637	45728	46164	46246	46327	1,7

Teorema do Limite Central

E se os valores não são “normais”

- População não tem distribuição normal?
  - Para amostras “grandes o suficiente”...
  - A distribuição das médias amostrais o será!.

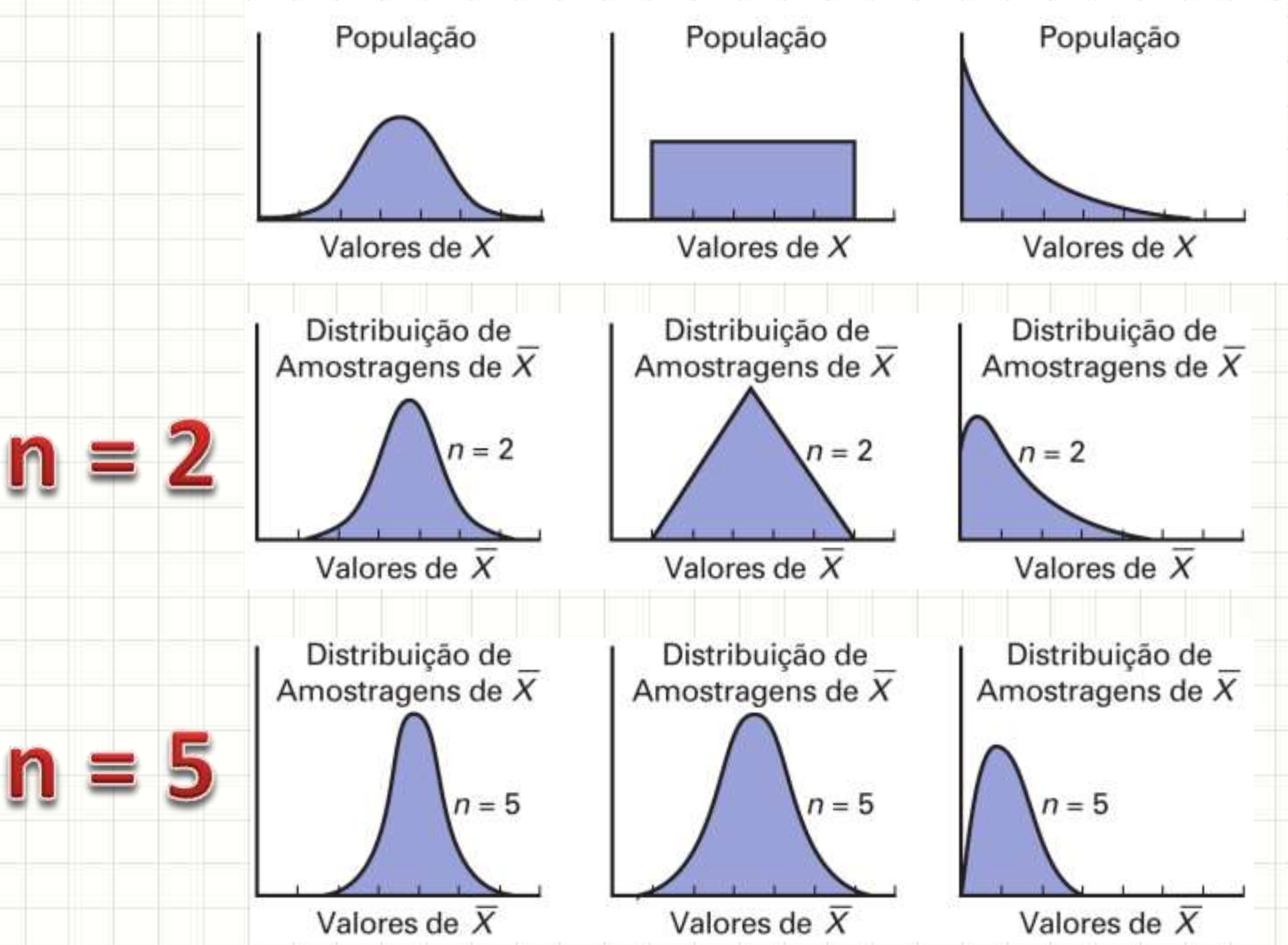
À medida que o tamanho da amostra vai se tornando *grande o suficiente*, a distribuição de amostragens da média aritmética passa a ser distribuída aproximadamente nos moldes da distribuição normal. Isso é verdadeiro, independentemente do formato da distribuição dos valores individuais dentro da população.

TEOREMA DO LIMITE CENTRAL

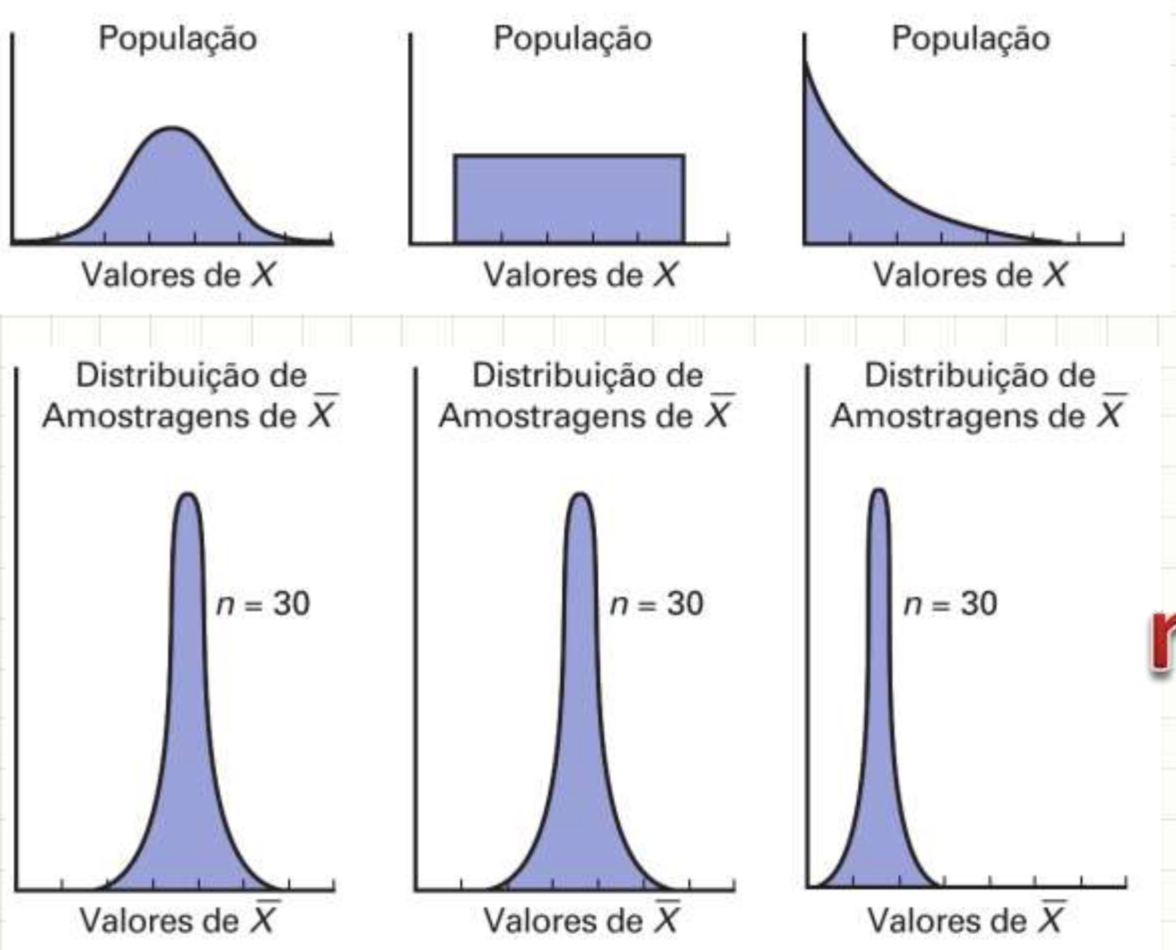
Teorema do Limite Central

- O que é uma amostra “grande o suficiente”?
  - Depende da distribuição
  - Em geral, maiores que 30 elementos
  - Se for próxima a um sino, maiores que 5...
  - Se for um sino, qualquer tamanho

Tamanho da Amostra: Observe!



Tamanho da Amostra: Observe!



Painel A
População Normal

Painel B
População Uniforme

Painel C
População Exponencial

n = 30

Tamanho da Amostra

- Pode-se derivar da equação Z

$$Z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

- A determinação do tamanho da amostra:

$$n = \left(\frac{\sigma . Z}{\bar{x} - \mu}\right)^2$$

Confiança	Z
95,0%	1,96
95,5%	2
99%	2,57

Distribuição de Amostragem da

Proporção

D.A. da Proporção

- Ex: deseja-se a preferência de uma marca
- A variável terá valores 1 ou 0
  - Ou é a preferida da pessoa, ou não é
- Em uma amostra, a média é a proporção

$$p = \frac{X}{n} = \frac{Número\ de\ Itens\ de\ Interesse}{Tamanho\ da\ Amostra}$$

- Pode-se estimar a proporção da população Π

Ausência de Viés e Erro

- Como é média, a proporção é isenta de viés
- O erro padrão será:

$$\sigma_p = \sqrt{\frac{\Pi.(1-\Pi)}{n}}$$

- A distribuição segue o padrão binomial

$$Z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{p - \Pi}{\sqrt{\frac{\Pi.(1-\Pi)}{n}}}$$

Exemplo

- Considerando que 77% dos adultos desejam acesso à internet quando estão em férias, determine a probabilidade de, em uma amostra de 200 pessoas em férias, mais de 80% desejar internet.

$$Z = \frac{p - \Pi}{\sqrt{\frac{\Pi.(1-\Pi)}{n}}} = \frac{0{,}8 - 0{,}77}{\sqrt{\frac{0{,}77.0{,}23}{200}}} = \mathbf{1{,}01}$$

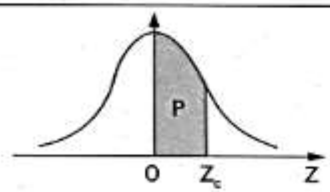
$$P(P > 80\%) = 0{,}5 - 0{,}34375$$

$$P(P > 80\%) = 0{,}15625$$

Tabela III — Distrib

$Z \sim$ I

Corpo da tabela dá a probabili



parte inteira e primeira decimal de Z_c	0	1	2	7	8	9	parte inteira e primeira decimal de Z_c
	p = 0						
1,0	34134	34375	34614	35769	35993	36214	1,0
1,1	36433	36650	36864	37900	38100	38298	1,1
1,2	38493	38686	38877	39796	39973	40147	1,2
1,3	40320	40490	40658	41466	41621	41774	1,3
1,4	41924	42073	42220	42922	43056	43189	1,4
1,5	43319	43448	43574	44179	44295	44408	1,5
1,6	44520	44630	44738	45254	45352	45449	1,6
1,7	45543	45637	45728	46164	46246	46327	1,7
1,8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Conclusões

Resumo

- Distribuição de Amostragens
- Estimativa da média populacional
  - Erro-padrão da média
- Teorema do Limite Central
  - Tamanhos mínimos de amostra
- Estimativa de proporção populacional
  - Erro padrão da proporção

- A avaliação está chegando!

PERGUNTAS?

Exercícios

Exercícios

1. Dada uma distribuição normal, com $\mu=100$ e $\sigma = 10$, caso selecione uma amostra com $n = 25$, qual a probabilidade de que a média esteja:

a) Abaixo de 95?

Exercícios

1. Dada uma distribuição normal, com μ=100 e σ = 10, caso selecione uma amostra com n = 25, qual a probabilidade de que a média esteja:

a) Abaixo de 95?

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{95 - 100}{\frac{10}{\sqrt{25}}} = \frac{-5}{2}$$

$$\boldsymbol{Z = -2,5}$$

$$P(\bar{x} < 95) = 0,5 - 0,49379$$

$$P(\bar{x} < 95) = 0,00621$$

Tabela III — Distribuição Normal Padrão

$Z \sim N(0, 1)$

Corpo da tabela dá a probabilidade p, tal que $p = P(0 < Z <$

parte inteira e primeira decimal de Z_c	Segunda decimal de Z_c						
	0	1	2	3	4	5	6
	p = 0						
2,0	47725	47778	47831	47882	47932	47982	480
2,1	48214	48257	48300	48341	48382	48422	484
2,2	48610	48645	48679	48713	48745	48778	488
2,3	48928	48956	48983	49010	49036	49061	490
2,4	49180	49202	49224	49245	49266	49286	493
2,5	49379	49396	49413	49430	49446	49461	494
2,6	49534	49547	49560	49573	49585	49598	496
2,7	49653	49664	49674	49683	49693	49702	497
2,8	49744	49752	49760	49767	49774	49781	497
2,9	49813	49819	49825	49831	49836	49841	498
3,0	49865	49869	49874	49878	49882	49886	498

Exercícios

1. Dada uma distribuição normal, com μ=100 e σ = 10, caso selecione uma amostra com n = 25:

b) Existe uma chance de 65% de estar acima de qual valor?

Exercícios

1. Dada uma distribuição normal, com μ=100 e σ = 10, caso selecione uma amostra com n = 25:

b) Existe uma chance de 65% de estar acima de qual valor?

Área superior = 65%, mas metade da gaussiana é 50%

Então, procura-se o Z para 15%, mas sinal negativo

$$\boldsymbol{Z = -0,39}$$

decimal de Z_c	0	1	2	3	4	5	6	7	8	9	decimal de Z_c
	p = 0										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,0
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,1

Exercícios

1. Dada uma distribuição normal, com μ=100 e σ = 10, caso selecione uma amostra com n = 25:

b) Existe uma chance de 65% de estar acima de qual valor? $\boldsymbol{Z = -0,39}$

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \Rightarrow \ -0{,}39 = \frac{\bar{x} - 100}{\frac{10}{\sqrt{25}}} \Rightarrow \quad -0{,}39.2 = \bar{x} - 100$$

$$\therefore \bar{x} = 100 - 0{,}78 \qquad \boldsymbol{\bar{x} = 99,22}$$

Exercícios

2. Um caixa de uma agência bancária gasta com cada cliente, em média populacional, $\mu=3{,}10$ minutos e desvio padrão $\sigma = 0{,}40$ minutos. Se você selecionar uma amostra de 16 clientes, qual a probabilidade de que a média por cliente seja de, pelo menos, 3 minutos?

Exercícios

2. Um caixa de uma agência bancária gasta com cada cliente, em média populacional, μ=3,10 minutos e desvio padrão σ = 0,40 minutos. Se você selecionar uma amostra de 16 clientes, qual a probabilidade de que a média por cliente seja de, pelo menos, 3 minutos?

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{3 - 3{,}1}{\frac{0{,}4}{\sqrt{16}}} = \frac{-0{,}1}{0{,}1}$$

$$\boldsymbol{Z = -1}$$

$$P(\bar{x} > 3) = 0{,}5 + 0{,}34134$$

$$P(\bar{x} > 3) = 0{,}84134$$

Tabela III — Distribuição Normal Padrão

$Z \sim N(0, 1)$

Corpo da tabela dá a probabilidade p, tal que $p = P(0 < Z <$

parte inteira e primeira decimal de Z_c	Segunda decimal de Z_c 0	1	2	3	4	5
	p = 0					
1,0	34134	34375	34614	34850	35083	35314
1,1	36433	36650	36864	37076	37286	37493
1,2	38493	38686	38877	39065	39251	39435

Exercícios

3. Para uma amostra de 50 domicílios, foi feita a pergunta: “Você tem um dispositivo Apple?”. Dos 50 respondentes, 20 afirmaram que sim. Determine a proporção da amostra e o erro-padrão da proporção.

Exercícios

3. Para uma amostra de 50 domicílios, foi feita a pergunta: “Você tem um dispositivo Apple?”. Dos 50 respondentes, 20 afirmaram que sim. Determine a proporção da amostra e o erro-padrão da proporção determinada.

$$p = \frac{X}{n} = \frac{20}{50} = 0{,}4$$

$$\sigma_p = \sqrt{\frac{0{,}4.\,(1 - 0{,}4)}{50}} = 0{,}11$$

Exercícios

4. Um instituto fez uma pesquisa eleitoral com duas candidatas. Supondo que na amostra uma delas receber pelo menos 55% dos votos, essa candidata terá o prognóstico de vencedora. Se selecionar uma amostra de 100 eleitores, qual a probabilidade de ela ter prognóstico de vencedora se o porcentual de votos pra ela na população for 50,1%?

$\Pi =$ 50,1%   n = 100 eleitores   p = 55%

Exercícios

4. Π = 50,1%; n = 100 eleitores; p = 55%

$$Z = \frac{p - \Pi}{\sqrt{\frac{\Pi.(1-\Pi)}{n}}} = \frac{0{,}55 - 0{,}501}{\sqrt{\frac{0{,}501.0{,}499}{100}}} = \mathbf{0,98}$$

$$P(x > 50{,}1\%) = 0{,}5 - 0{,}33646 \qquad P(x > 50{,}1\%) = 0{,}16354$$

decimal de Z_c	0	1	2	3	4	5	6	7	8	9	decimal de Z_c
	p = 0										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,0
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,1
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147	1,2
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774	1,3