



Universidade Federal do Ceará

Faculdade de Economia

Métodos Quantitativos

Vicente Lima Crisóstomo

Fortaleza, 2020

Sumário

- Introdução
- Estatística Descritiva
- Probabilidade
- Distribuições de Probabilidades
- Amostragem e Distribuições Amostrais
- Estimação
- Testes de Significância
- Análise de Variância
- Teste de Significância para Proporções
- Testes Não Paramétricos
- Correlação e Regressão

Estimação

- Estimação é um processo que consiste em utilizar dados amostrais para estimar parâmetros populacionais desconhecidos
- Características de uma população podem ser estimadas a partir de uma amostra aleatória
 - Exemplos
 - Estimação sobre produtos defeituosos
 - Estimação sobre clientes
 - Estimação sobre alunos
 - Estimação sobre empresas
 - Estimação sobre custos
 - Estimação sobre populações de animais

Estimação

- Estimativas **pontuais** e **intervalares**
 - Quanto ao valor do parâmetro estimado
 - Estimador
 - É uma estatística amostral usada para aproximar-se um parâmetro populacional
 - Média amostral por exemplo
 - Estimativa
 - Valor específico, ou intervalo de valores, usado para aproximar um parâmetro populacional

Estimação

■ Estimativas **Pontuais**

■ Estimativa pontual

- Valor único usado para aproximar parâmetro populacional

■ Estimação de um valor específico do parâmetro populacional

- Média amostral usada como estimação da média populacional
- Desvio padrão amostral usado como estimação do DP populacional
- Proporção de elementos numa amostra como estimação da proporção populacional

Estimação

■ Estimativas **Intervalares**

■ Estimativa intervalar ou **Intervalo de Confiança**

- *Amplitude, ou intervalo, de valores que têm probabilidade de conter o verdadeiro parâmetro populacional*

■ Estimação de um intervalo de valores possíveis, no qual o parâmetro populacional específico esteja contido

- Parâmetro amostral está próximo do populacional embora não seja exatamente igual
- Variabilidade amostral
- “Estimativa Intervalar”
 - Provê um intervalo de valores possíveis do parâmetro populacional

Exemplos de estimativas pontuais e intervalares

Estimativa pontual

- Consumo médio de 1litro de gasolina por 10Km
- Consumo médio de 1,5litros de água por dia
- O desvio padrão do desempenho estudantil é 10ptos
- O DP dos pesos das baleias é de 100Kg
- 10% da população é a favor da reforma eleitoral
- 5% dos eleitores vota em trânsito

Estimativa intervalar

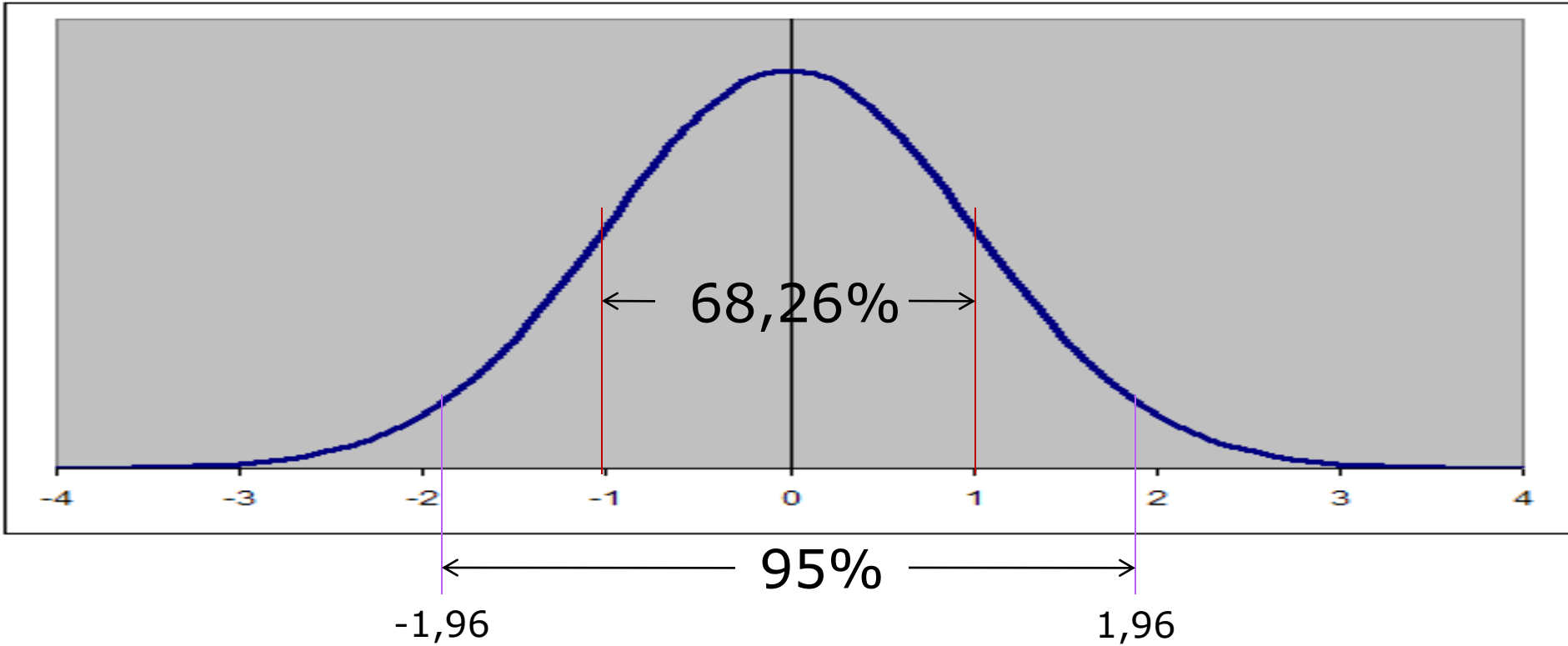
- Consumo médio de 1litro de gasolina entre 8 e 12Km
- Consumo médio entre 1 e 2litros de água por dia
- O desvio padrão do desempenho estudantil está entre 8 e 12ptos
- O DP dos pesos das baleias varia de 80 a 120Kg
- Entre 8 e 12% da população é a favor da reforma eleitoral
- Entre 3 e 7% dos eleitores vota em trânsito

Estimação

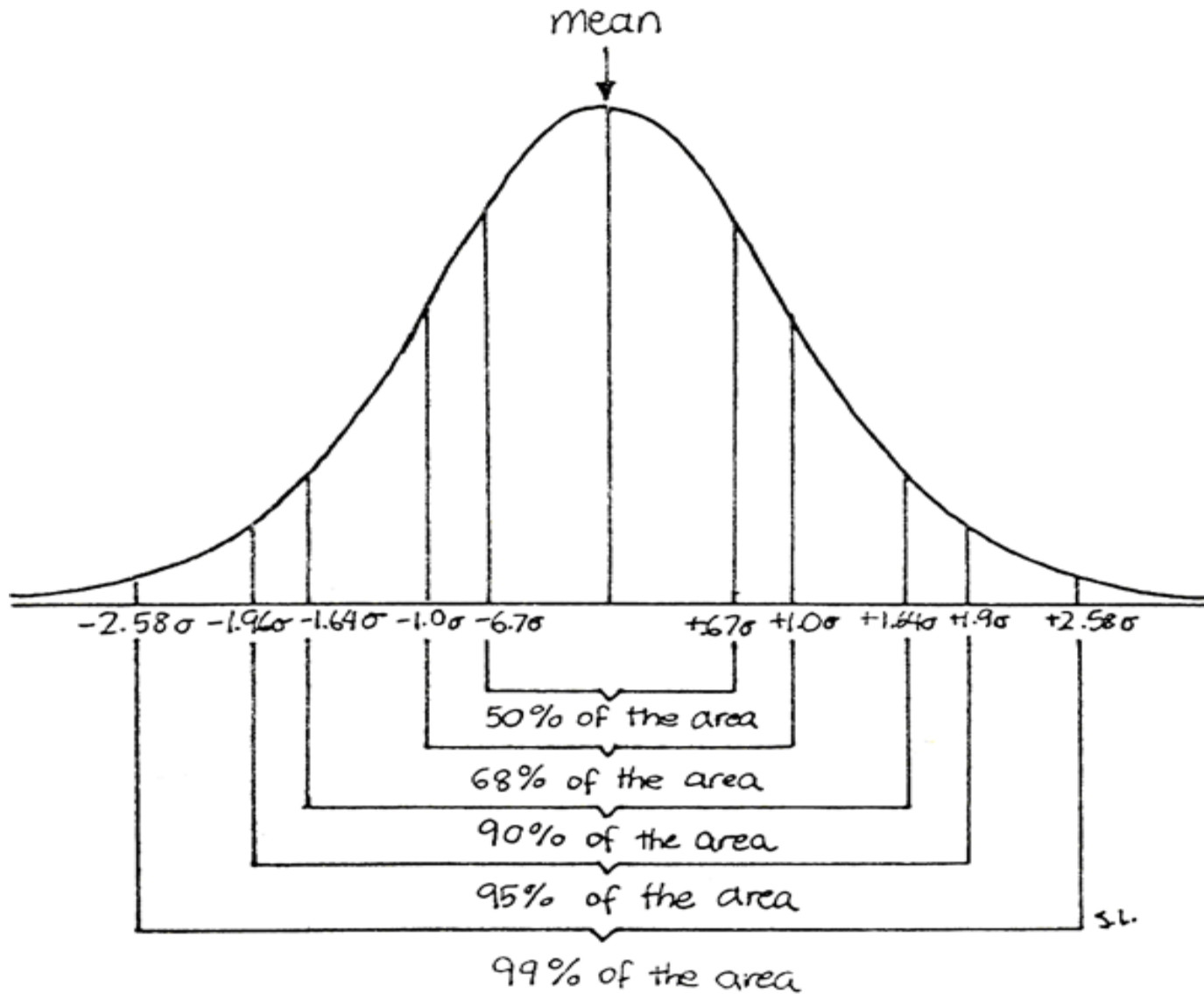
■ Lógica da Estimação

- Distribuição amostral usada como estimador
- Estatística amostral é uma observação da Distribuição Amostral
- Distribuição de médias amostrais é Normal ou aproximadamente Normal
 - 68,26% da estatística amostral está a menos de 1DP da verdadeira média
 - 31,74% das médias amostrais superam 1DP a partir da média amostral
 - 90% da estatística amostral está a menos de 1,65DP da verdadeira média
 - 10% supera 1,65DP em relação à média amostral
 - 95% da estatística amostral está a menos de 1,96DP da verdadeira média
 - 5% supera 1,96DP em relação à média amostral

Estimacão



Estimação



Estimação

- Assim, ao dizer-se que
 - A média amostral está a menos de 1DP da verdadeira média
 - Pode-se esperar estar certo em 68,26% das vezes (possibilidade de acerto) e errados em 31,74% das vezes
 - A média amostral está a menos de 1,96DP da verdadeira média
 - Pode-se esperar estar certo em 95% das vezes (possibilidade de acerto) e errados em 5% das vezes
- Como não se sabe o valor exato da verdadeira média populacional
 - Não se sabe exatamente sua distância da média amostral que se dispõe
 - Aceita-se a atribuição probabilística do intervalo no qual o verdadeiro valor **pode** estar contido
 - **INTERVALO DE CONFIANÇA**

Estimação

- **INTERVALO DE CONFIANÇA**

- **GRAU DE CONFIANÇA = Nível de Confiança = Coeficiente de Confiança**

- Indica a probabilidade que o verdadeiro parâmetro populacional tem de estar contido no Intervalo de Confiança
- Aceita-se com **GRAU DE CONFIANÇA** DE $(1 - P(\text{erro}))$ que o parâmetro populacional está contido no intervalo de confiança
- Intervalo de confiança de 90% (grau de confiança de 90%)
 - Tem associado um risco de 10% de erro
- Intervalo de confiança de 95% (grau de confiança de 95%)
 - Tem associado um risco de 5% de erro
- Intervalo de confiança de 99% (grau de confiança de 99%)
 - Tem associado um risco de 1% de erro

Estimação

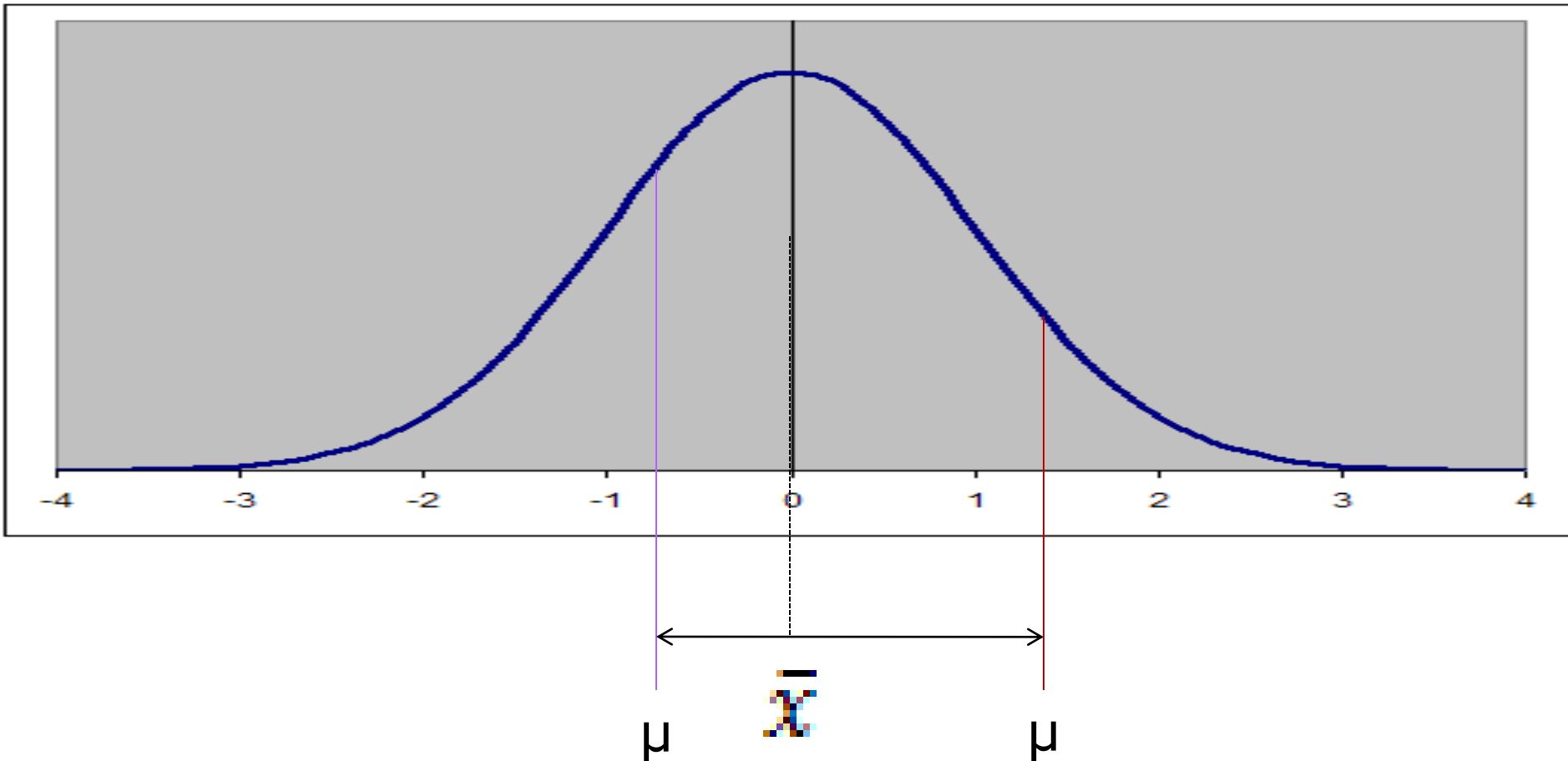
- *INTERVALO DE CONFIANÇA*
 - O IC é um intervalo de valores, **centrado na estatística amostral**, dentro do qual julga-se, com certo **grau de confiança**, e respectivo **risco de erro**, estar contido o verdadeiro parâmetro populacional.
- Intervalo de confiança de X%
 - Tem associado um risco de $(1 - X)\%$ de erro
 - Probabilidade do verdadeiro valor populacional não estar contido no intervalo de confiança
 - $(1 - X)\%$ dos intervalos não incluirão a verdadeira média populacional

Estimação

- **INTERVALO DE CONFIANÇA**
 - Aceita-se com GRAU DE CONFIANÇA DE $(1 - P(\text{erro}))$ que o parâmetro populacional está contido no intervalo de confiança
 - O risco de erro é reduzido à medida em que se aumenta o intervalo de confiança (valor de z)
- Independente do Intervalo de Confiança estabelecido
 - Média amostral é menor ou maior que o parâmetro populacional verdadeiro
 - Esta incerteza leva ao estabelecimento de um intervalo de valores possíveis

Estimação

- Média amostral é menor ou maior que o parâmetro populacional verdadeiro



Estimação

- *INTERVALO DE CONFIANÇA*
- Independente do Intervalo de Confiança estabelecido
 - Esta incerteza leva ao estabelecimento de um intervalo de valores possíveis delimitado por Valores Limítrofes
- Valor Limítrofe, ou Valor Crítico do Intervalo de Confiança
 - Valor limite do IC
 - Valores que excedem o VC são mais improváveis de ocorrer
 - ***Espera-se que o verdadeiro parâmetro populacional não exceda o VC com o grau de confiança estabelecido***

Estimação

- *INTERVALO DE CONFIANÇA*
- Ao usar-se dados amostrais para estimar parâmetros populacionais
 - Há uma **margem de erro** (E , ou ε)
 - Possibilidade de estimar-se erroneamente
 - É a diferença máxima provável entre o parâmetro amostral e o parâmetro populacional
 - com probabilidade $(1 - \alpha)$

Estimação

- *INTERVALO DE CONFIANÇA (Distribuição Normal)*
 - Esta incerteza leva ao estabelecimento de um intervalo de valores possíveis
 - $[\bar{X} - z \sigma_{\bar{X}} ; \bar{X} + z \sigma_{\bar{X}}]$
- *INTERVALO DE CONFIANÇA*
 - De 90%
 - $[\bar{X} - 1,64 \sigma_{\bar{X}} ; \bar{X} + 1,64 \sigma_{\bar{X}}]$
 - De 95%
 - $[\bar{X} - 1,96 \sigma_{\bar{X}} ; \bar{X} + 1,96 \sigma_{\bar{X}}]$
 - De 99%
 - $[\bar{X} - 2,58 \sigma_{\bar{X}} ; \bar{X} + 2,58 \sigma_{\bar{X}}]$

Estimação

- *INTERVALO DE CONFIANÇA*
 - É um intervalo de valores, centrado na Estatística Amostral, dentro do qual, aceita-se, ou julga-se, com uma margem de erro conhecida (risco), que o parâmetro populacional está contido
- Estimação da Média de uma População
 - Distância de determinada média da amostra para a média da distribuição amostral em unidade efetiva
 - Depende da variabilidade da distribuição amostral
 - Maior amostra => menor Desvio Padrão da distribuição amostral
 - Grandes amostras produzem médias amostrais mais próximas do parâmetro populacional

Estimação

- Estimação da Média de uma População
 - Conhecimento do desvio padrão populacional
 - **Desvio Padrão populacional conhecido:** Distribuição z
 - Estimativa pontual de μ_x : \bar{x}
 - Estimativa intervalar de μ_x : $\left[\bar{x} - z \sigma_{\bar{x}} ; \bar{x} + z \sigma_{\bar{x}} \right]$
 - Desvio Padrão populacional não conhecido
 - Situações mais comuns
 - Distribuição t de Student

Estimação

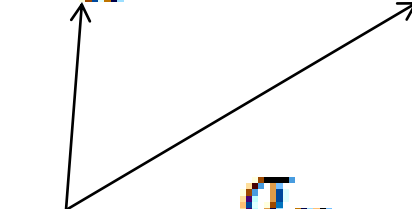
■ Estimação da Média de uma População

■ Conhecido DP Populacional

- Estimativa Intervalar

- Baseada na hipótese: Distribuição amostral das médias amostrais é Normal

- Estimativa intervalar de μ_x : $[\bar{x} - z \sigma_{\bar{x}} ; \bar{x} + z \sigma_{\bar{x}}]$

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


- Média amostral (\bar{x}) é o ponto médio do intervalo de confiança

Estimação

■ Estimativa Intervalar

■ Confiança desejada (tabela z)

- 90% : $z = \pm 1,65$
- 95% : $z = \pm 1,96$
- 99% : $z = \pm 2,58$
- Intervalo de Confiança

$$\left[\bar{x} - z \sigma_{\bar{x}}; \bar{x} + z \sigma_{\bar{x}} \right]$$

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

- Erro de Estimação

$$erro = z \frac{\sigma_x}{\sqrt{n}}$$

$$erro = z \frac{\sigma_x}{\sqrt{n}}$$

Confiança (%)	z	Tamanho amostra	MediaAmostr	DPPopul	LimInf	LimSup	erro	Amplitude
68,26	1,00	36	100	10	98,333	101,667	1,667	3,333
90	1,65	36	100	10	97,250	102,750	2,750	5,500
95	1,96	36	100	10	96,733	103,267	3,267	6,533
99	2,58	36	100	10	95,700	104,300	4,300	8,600

Confiança	z	Tamanho amostra	MediaAmostr	DPPopul	LimInf	LimSup	erro	Amplitude
68,26	1,00	512	100	10	99,558	100,442	0,442	0,884
90	1,65	512	100	10	99,271	100,729	0,729	1,458
95	1,96	512	100	10	99,134	100,866	0,866	1,732
99	2,58	512	100	10	98,860	101,140	1,140	2,280

Confiança	z	Tamanho amostra	MediaAmostr	DPPopul	LimInf	LimSup	erro	Amplitude
68,26	1,00	1024	100	10	99,688	100,313	0,313	0,625
90	1,65	1024	100	10	99,484	100,516	0,516	1,031
95	1,96	1024	100	10	99,388	100,613	0,613	1,225
99	2,58	1024	100	10	99,194	100,806	0,806	1,613

Estimação

- Fatores que interferem na Amplitude do Intervalo de Confiança
 - Coeficiente de confiança
 - Tamanho da amostra
 - Dispersão da população (desvio padrão)
- Determinação do tamanho da amostra

$$erro = z \frac{\sigma_x}{\sqrt{n}} \quad \therefore \quad \sqrt{n} = z \frac{\sigma_x}{erro} \quad \therefore \quad n = \left(z \frac{\sigma_x}{erro} \right)^2$$

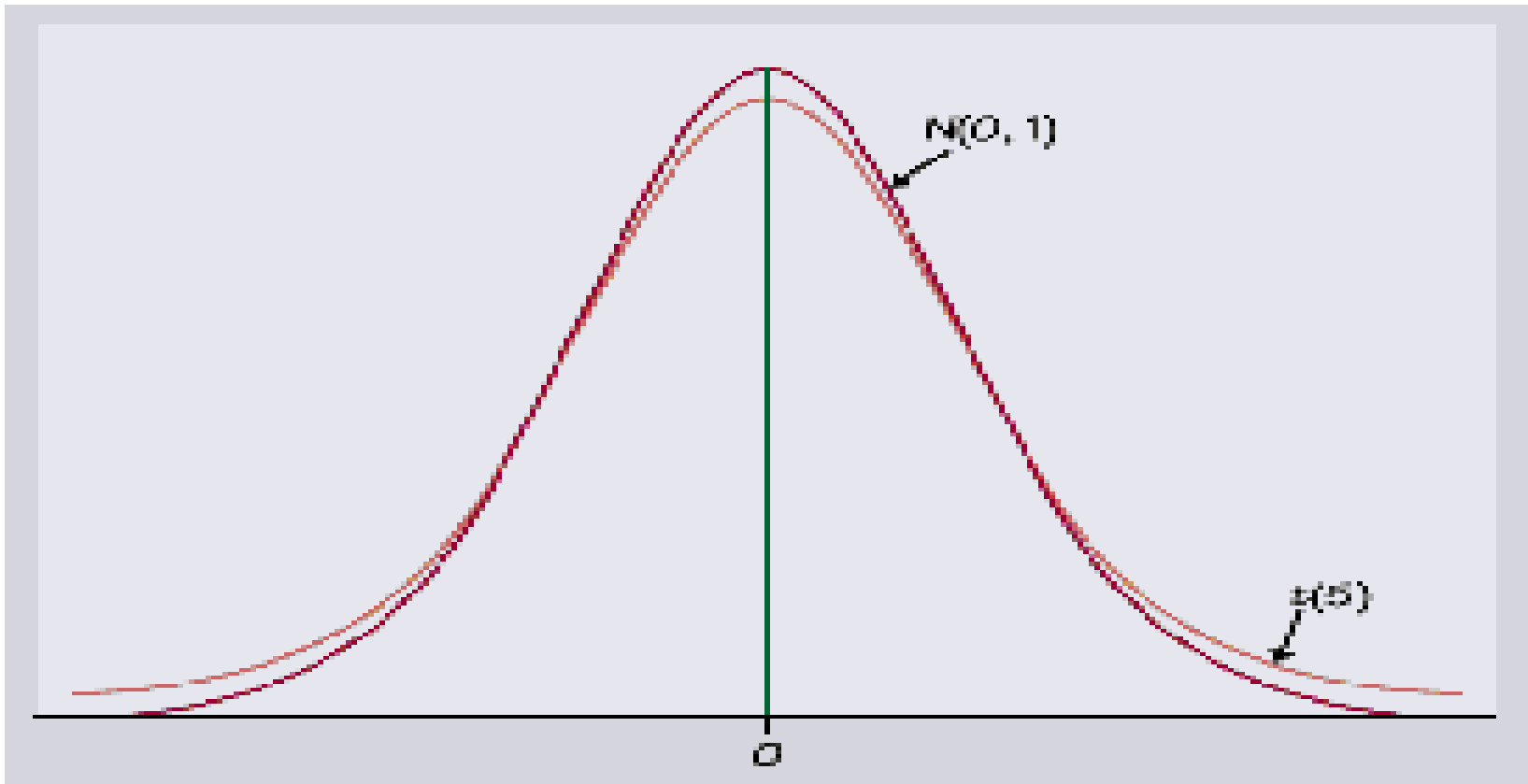
- E se não se conhece a dispersão?

Estimação

- Estimação da Média de uma População quando o *Desvio Padrão populacional não é conhecido*
 - Usa-se Desvio Padrão da Amostra (s_x) como estimativa
 - s_x é boa aproximação do verdadeiro valor
 - Teorema do Limite Central
 - Amostras com $N \geq 30$
 - Distribuição de médias aproximadamente Normal
 - $N < 30$
 - Deve-se usar a Distribuição ***t de Student***

Estimação

- Distribuição t de Student (W. S. Gosset)
 - Forma da Distribuição t parecida com a Normal
 - Distribuição t tem maior área nas caudas



Estimação

- Distribuição t de Student (W. S. Gosset)
 - Distribuição t tem maior área nas caudas
 - Para certo nível de confiança
 - t é um pouco superior ao correspondente z
 - Não é padronizada como a Normal
 - Distribuição t depende do tamanho da amostra
 - Há uma Distribuição t para cada amostra
 - Amostras com $n < 30$, distribuição t mais sensível ao tamanho amostral
 - Teoricamente, Distribuição t deve ser usada quando não se conhece DP populacional. Entretanto,
 - Grandes amostras pode-se usar z para aproximar t

Estimação

- ***Distribuição t de Student*** (W. S. Gosset)
 - Cálculo do Desvio Padrão da Amostral (s_x)

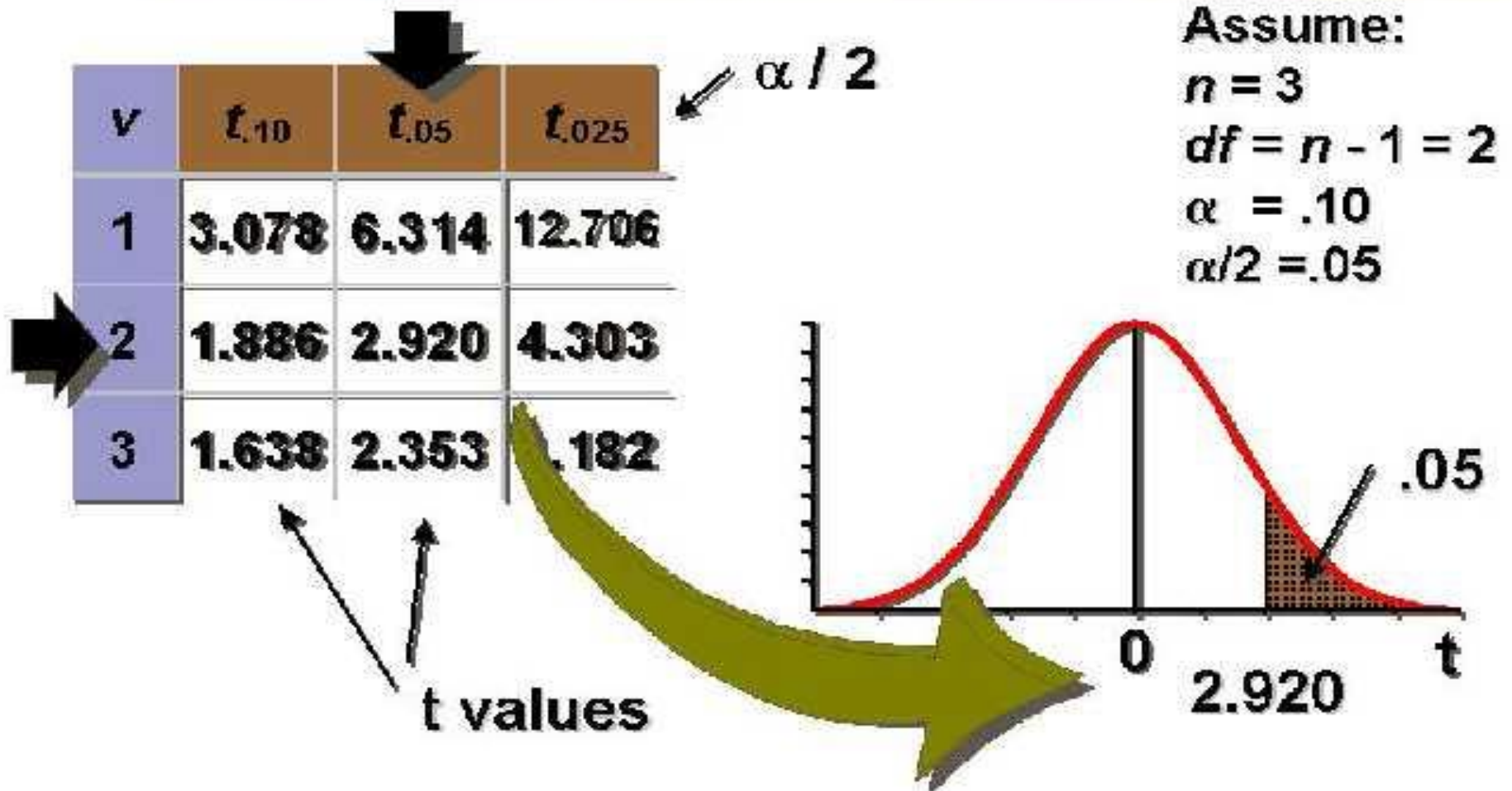
$$s_x = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

- $(n - 1)$ = **Graus de Liberdade** (tamanho da amostra menos um)

Estimação

- Distribuição t de Student (W. S. Gosset)
 - Tabela da Distribuição t
 - Tabela-se os principais valores de t
 - Para Graus de Liberdade ($n - 1$) e probabilidades (áreas) na(s) cauda(s) da curva
 - Probabilidades (áreas) no topo, ou abaixo
 - Referem-se a caudas da curva
 - Graus de Liberdade ao lado ($n - 1$)
 - **Células contêm valores de t**

Student's t Table



Probabilidades na cauda

Uma Cauda		0,100	0,050	0,025	0,010	0,005	GL
Duas Caudas		0,200	0,100	0,050	0,020	0,010	
D E G R E E S O F F R E E D O M	1	3,078	6,314	12,710	31,820	63,660	1
	2	1,886	2,920	4,303	6,965	9,925	2
	3	1,638	2,353	3,182	4,541	5,841	3
	4	1,533	2,132	2,776	3,747	4,604	4
	5	1,476	2,015	2,571	3,365	4,032	5
	6	1,440	1,943	2,447	3,143	3,707	6
	7	1,415	1,895	2,365	2,998	3,499	7
	8	1,397	1,860	2,306	2,896	3,355	8
	9	1,383	1,833	2,262	2,821	3,250	9
	10	1,372	1,812	2,228	2,764	3,169	10
	11	1,363	1,796	2,201	2,718	3,106	11
	12	1,356	1,782	2,179	2,681	3,055	12
	13	1,350	1,771	2,160	2,650	3,012	13
	14	1,345	1,761	2,145	2,624	2,977	14
	15	1,341	1,753	2,131	2,602	2,947	15
	16	1,337	1,746	2,120	2,583	2,921	16
	17	1,333	1,740	2,110	2,567	2,898	17
	18	1,330	1,734	2,101	2,552	2,878	18
	19	1,328	1,729	2,093	2,539	2,861	19
	20	1,325	1,725	2,086	2,528	2,845	20
	21	1,323	1,721	2,080	2,518	2,831	21
	22	1,321	1,717	2,074	2,508	2,819	22
	23	1,319	1,714	2,069	2,500	2,807	23
	24	1,318	1,711	2,064	2,492	2,797	24
	25	1,316	1,708	2,060	2,485	2,787	25
	26	1,315	1,706	2,056	2,479	2,779	26
	27	1,314	1,703	2,052	2,473	2,771	27
	28	1,313	1,701	2,048	2,467	2,763	28
	29	1,311	1,699	2,045	2,462	2,756	29
	30	1,310	1,697	2,042	2,457	2,750	30
	40	1,303	1,684	2,021	2,423	2,704	40
	60	1,296	1,671	2,000	2,390	2,660	60
	120	1,282	1,645	1,960	2,326	2,576	120

Valores de t para **90% de Confiança** (0,05 em cada cauda)

Tamanho Amostra	Graus de Liberdade (n-1)	Valor t
2	1	6,314
9	8	1,86
10	9	1,833
18	17	1,74
28	27	1,703
31	30	1,697

Valores de t para **95% de Confiança** (0,025 em cada cauda)

Tamanho Amostra	Graus de Liberdade (n-1)	Valor t
2	1	12,71
9	8	2,306
10	9	2,262
18	17	2,11
28	27	2,052
31	30	2,042

Valores de t para **99% de Confiança** (0,01 em cada cauda)

Tamanho Amostra	Graus de Liberdade (n-1)	Valor t
2	1	31,82
9	8	2,896
10	9	2,821
18	17	2,567
28	27	2,473
31	30	2,457

Estimação

- Distribuição *t* de Student (W. S. Gosset)
 - Intervalo de Confiança

$$\left[\bar{x} - t \frac{s_x}{\sqrt{n}} ; \bar{x} + t \frac{s_x}{\sqrt{n}} \right]$$

erro

$$s_x = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

$$erro = t \frac{s_x}{\sqrt{n}}$$

$$erro = t \frac{s_x}{\sqrt{n}}$$

Confiança	t	Tamanho amostra	GL	MediaAmostr	DPAmostr				
					(s_x)	LimInf	LimSup	erro	Amplitude
90%	1,711	25	24	20	1,500	19,487	20,513	0,513	1,027
95%	2,064	25	24	20	1,500	19,381	20,619	0,619	1,238
99%	2,797	25	24	20	1,500	19,161	20,839	0,839	1,678

Confiança	t	Tamanho amostra	GL	MediaAmostr	DPAmostr	LimInf	LimSup	erro	Amplitude
90%	1,083	10	9	20	1,500	19,486	20,514	0,514	1,027
95%	2,262	10	9	20	1,500	18,927	21,073	1,073	2,146
99%	3,250	10	9	20	1,500	18,458	21,542	1,542	3,083

Confiança	t	Tamanho amostra	GL	MediaAmostrai	DPAmostrai	LimInf	LimSup	erro	Amplitude
90%	1,671	61	60	20	1,500	19,679	20,321	0,32092	0,642
95%	2,000	61	60	20	1,500	19,616	20,384	0,384	0,768
99%	2,660	61	60	20	1,500	19,489	20,511	0,511	1,022

■ Intervalos de Confiança Unilaterais

■ Há situações em que

- Quer-se verificar se parâmetro populacional é menor que um referencial mínimo
- Quer-se verificar se parâmetro populacional é maior que um referencial máximo

■ Nestas situações

- O intervalo de confiança apresenta só um limite

	Desvio Padrão Populacional Conhecido (z)	Desvio Padrão Populacional Desconhecido (t)
Limite Inferior	$\bar{x} - z \frac{\sigma_x}{\sqrt{n}}$	$\bar{x} - t \frac{s_x}{\sqrt{n}}$
Limite Superior	$\bar{x} + z \frac{\sigma_x}{\sqrt{n}}$	$\bar{x} + t \frac{s_x}{\sqrt{n}}$

■ Intervalos de Confiança Unilaterais

- Uma amostra aleatória de 49 observações ($\Rightarrow z$) tem média 50 tem **desvio padrão populacional** igual a 6.
 - Estabeleça com 90% de segurança um valor superior para a média

$$\bar{x} + z \frac{\sigma_x}{\sqrt{n}} = 50 + 1,65 \frac{6}{\sqrt{49}} = 50 + 1,41 = 51,41$$

- Qual o risco (probabilidade) da média ser superior a 52

$$z = \frac{x - \bar{x}}{\sigma_{\bar{x}}} = \frac{x - \bar{x}}{\frac{\sigma_x}{\sqrt{n}}} = \frac{52 - 50}{\frac{6}{\sqrt{49}}} = 2,33$$

$$P(z > 2,33) = P(z > 0) - P(0 < z < 2,33) = 0,5 - 0,4901 = 0,0099 = 0,99\%$$

■ Intervalos de Confiança Unilaterais

- Uma amostra aleatória de 30 observações ($\Rightarrow t$) tem média 50. Sabendo que o **desvio padrão (amostral)** é igual a 7, pede-se:
 - Estabeleça com 90% de segurança um valor inferior para a média (unilateral)

$$\bar{x} - t \frac{\sigma_x}{\sqrt{n}} = 50 - 1,311 \frac{7}{\sqrt{30}} = 50 - 1,68 = 48,32$$

- Qual o risco (probabilidade) da média ser inferior 47?

$$t = \frac{x - \bar{x}}{s_{\bar{x}}} = \frac{x - \bar{x}}{\frac{s_x}{\sqrt{n}}} = \frac{47 - 50}{\frac{7}{\sqrt{30}}} = -2,35$$

$$P(t < -2,35) \approx P(z < -2,35) = 0,5 - 0,4906 = 0,0094 = 0,94\%$$