

Estimativas e Tamanhos de Amostras

- 1 Aspectos Gerais
- 2 Estimativa de uma Média Populacional:
Grandes Amostras
- 3 Estimativa de uma Média Populacional:
Pequenas Amostras
- 4 Tamanho Amostral Necessário para Estimar μ
- 5 Estimativa de uma Proporção Populacional

Hipóteses

❖ **$n > 30$**

A amostra deve ter mais do que 30 valores.

❖ **Amostra Aleatória Simples**

Todas amostras de mesmo tamanho têm mesma chance de ser selecionada.

Dados coletados de forma imprecisa ou descuidada podem ser totalmente destituídos de valor, mesmo que a amostra seja suficientemente grande.

Definições

❖ Estimador

uma fórmula ou processo para utilizar dados amostrais para estimar um parâmetro populacional

❖ Estimativa

é um valor específico, ou um intervalo de valores, usado para aproximar um parâmetro populacional

❖ Estimativa Pontual

é um valor (ou ponto) usado para aproximar um parâmetro populacional

A média amostral \bar{x} é a melhor estimativa pontual da média populacional μ .

Definição

Intervalo de Confiança

(ou Estimativa Intervalar)

é uma amplitude (ou um intervalo) de valores usados para estimar o verdadeiro valor do parâmetro populacional

Menor # < parâmetro populacional < **Maior #**

Por exemplo:

Menor # < μ < **Maior #**

Definição

Grau de Confiança

(nível de confiança ou coeficiente de confiança)

é a probabilidade $1 - \alpha$ (geralmente expressa como o valor percentual equivalente) que é a frequência relativa das vezes que o intervalo de confiança realmente contém o parâmetro populacional, assumindo que o processo de estimação é repetido um grande número de vezes.

geralmente 90%, 95%, ou 99%

($\alpha = 10\%$), ($\alpha = 5\%$), ($\alpha = 1\%$)

Interpretando um Intervalo de Confiança

$$98,08^\circ < \mu < 98,32^\circ$$

Correto: Estamos 95% confiantes que o intervalo de 98,08 a 98,32 realmente contém o verdadeiro valor de μ . Isto significa que se construíssemos intervalos de confiança a partir de muitas amostras diferentes de mesmo tamanho, 95% deles conteriam efetivamente a média populacional μ .

Errado: Há uma chance de 95% de o verdadeiro valor μ estar entre 98,08 e 98,32.

Confidence Intervals from 20 Different Samples

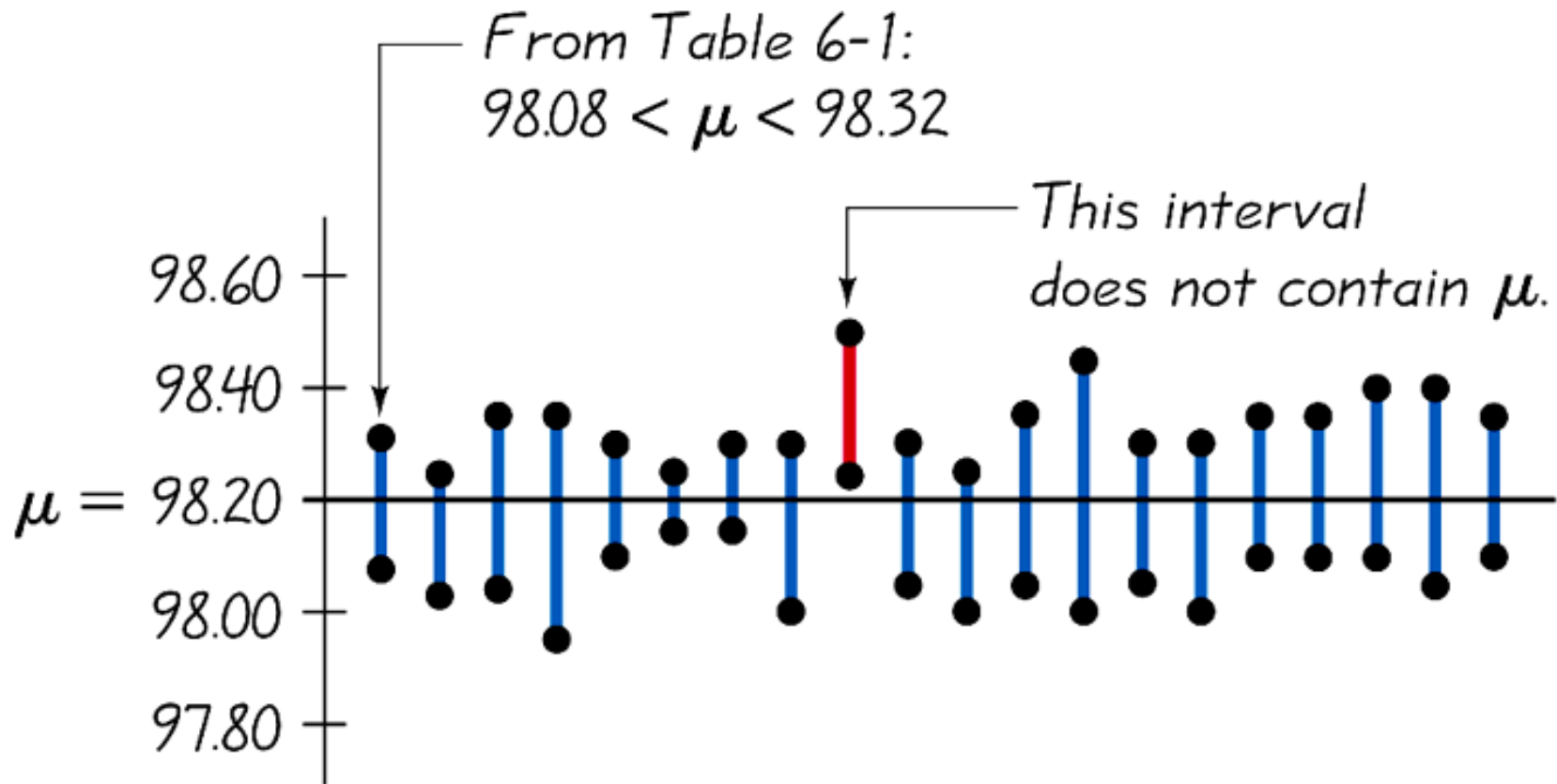


Figure 6-1

Definição

Valor Crítico

é o número na fronteira que separa os valores das estatísticas amostrais prováveis de ocorrerem, dos valores que têm pouco chance de ocorrer. O número $z_{\alpha/2}$ é um valor crítico que é um escore z com a propriedade de separar uma área de $\alpha/2$ na cauda direita da distribuição normal padronizada.

O Valor Crítico $z_{\alpha/2}$

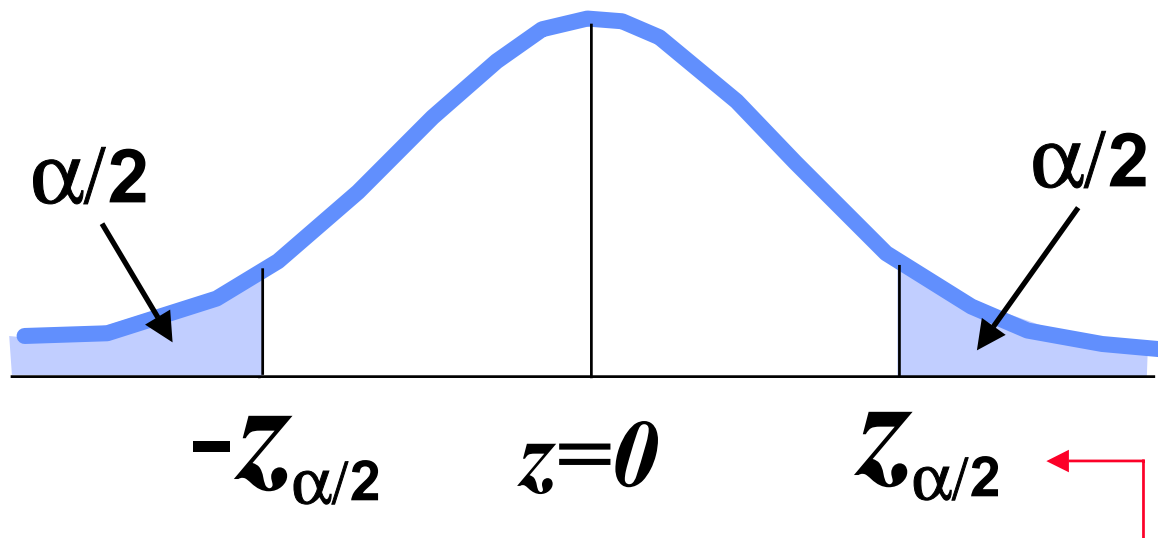
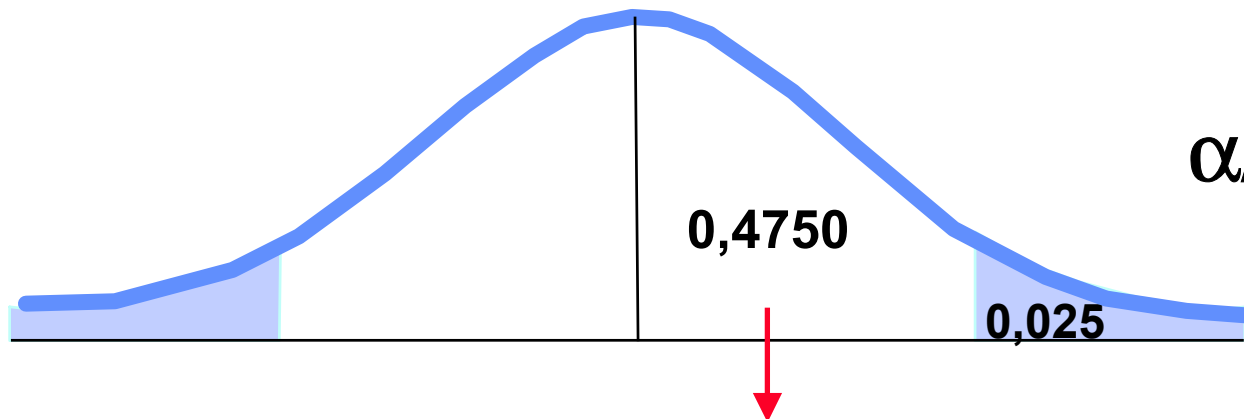


Figura 6-2

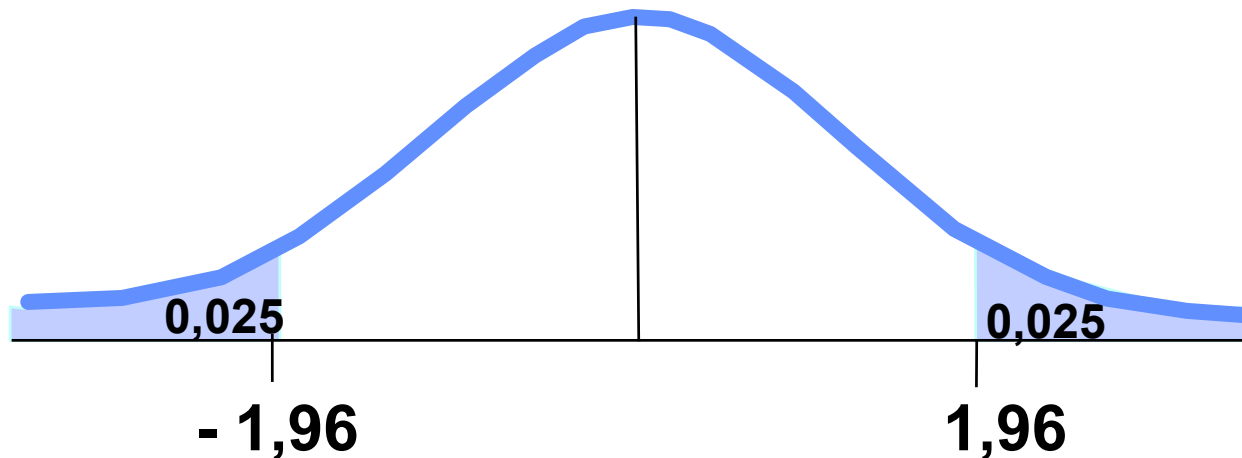
Lido na Tabela A-2
(corresponde à área de
 $0,5 - \alpha/2$)

Determinação de $z_{\alpha/2}$ para 95% de Grau de Confiança



$$\alpha = 0,05$$
$$\alpha/2 = 0,025$$

Usar a Tabela A-2 para
achar o valor de $z = 1,96$



$$z_{\alpha/2} = \pm 1,96$$

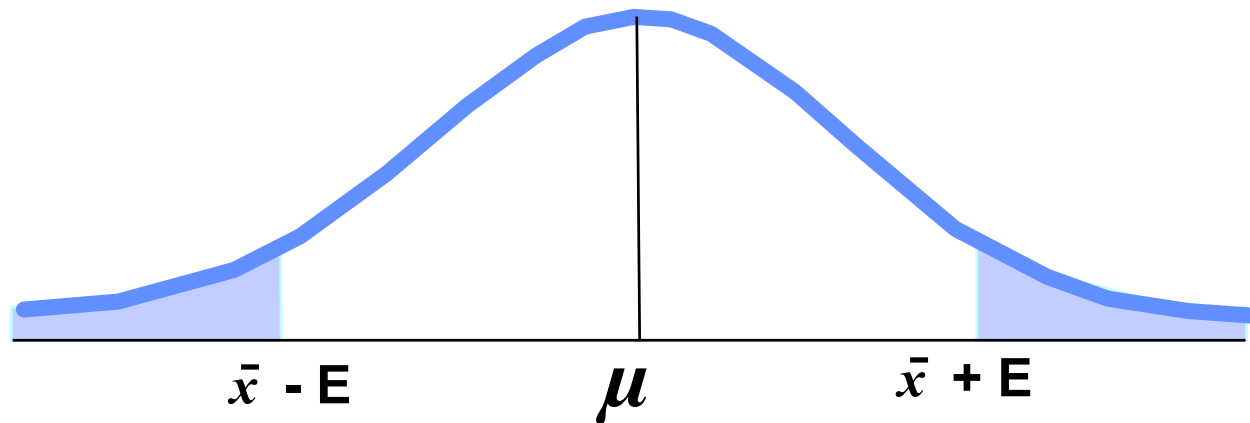
$z_{\alpha/2}$ para vários Grau de Confiança

Grau de Confiança	α	Valor Crítico $z_{\alpha/2}$
90%	0,10	1,645
95%	0,05	1,96
99%	0,01	2,575

Definição

Margem de Erro

é a máxima diferença provável entre a média amostral observada \bar{x} e a verdadeira média populacional μ . É denotada por E .



$$\bar{x} - E < \mu < \bar{x} + E$$

Limite inferior

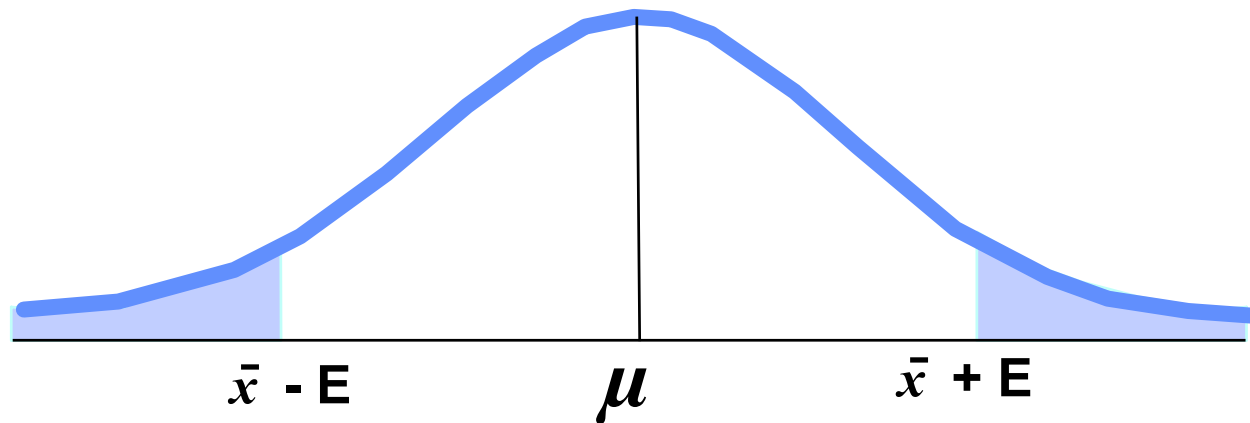
Limite superior

Definição

Margem de Erro

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Fórmula 6-1



É chamada também de **máximo erro da estimativa**

Cálculo de E Quando σ é Desconhecido

- ❖ Se $n > 30$, podemos substituir σ na Fórmula 6-1 pelo desvio-padrão amostral s .
- ❖ Se $n \leq 30$, a população deve ter distribuição normal e devemos conhecer σ para aplicar a Fórmula 6-1.

Intervalo de Confiança (ou Estimativa Intervalar) para a Média Populacional μ

(Com Base em Grandes Amostras: $n > 30$)

$$\bar{x} - E < \mu < \bar{x} + E$$

$$\mu = \bar{x} + E$$

$$(\bar{x} + E, \bar{x} - E)$$

Processo de Construção de um Intervalo de Confiança para μ

(Com Base em uma Amostra Grande: $n > 30$)

1. Determinar o valor crítico $z_{\alpha/2}$ correspondente ao grau de confiança desejado.
2. Calcular a margem de erro $E = z_{\alpha/2} \cdot \sigma / \sqrt{n}$.
Se o desvio-padrão populacional σ não é desconhecido, utilizar o desvio-padrão amostral, desde que $n > 30$.
3. Calcular os valores de $\bar{x} - E$ and $\bar{x} + E$. Levar esses valores na expressão geral do intervalo de confiança:

$$\bar{x} - E < \mu < \bar{x} + E$$

Exemplo: Uma pesquisa sobre o valor pago por um pacote de viagem para a Turquia envolvendo uma amostra aleatória de 64 pessoas mostrou que a despesa média era \$ 390, com um desvio-padrão de \$ 96. Estime o valor médio pago por “todos” os turistas com 95% de confiança.

$$n = 64$$

$$\bar{x} = 390$$

$$s = 96$$

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 1,96 \cdot \frac{96}{\sqrt{64}} = 23,52$$

$$\alpha = 0,05$$

$$\alpha/2 = 0,025$$

$$z_{\alpha/2} = 1,96$$

$$\bar{x} - E < \mu < \bar{x} + E$$

$$366,48 < \mu < 413,52$$

Baseado na amostra disponível, o intervalo de confiança para a média populacional é $98.08^\circ < \mu < 98.32^\circ$. Se fôssemos selecionar muitas amostras diferentes de mesmo tamanho, 95% dos intervalos de confiança conteriam a média populacional μ .

Tamanho da Amostra para Estimar a Média μ

Quando se conhece uma estimativa de p :

$$n = \frac{(z_{\alpha/2})^2 \sigma}{E^2}$$

Fórmula 6-2

Ao determinar o tamanho da amostra n , se a aplicação da amostra não conduz a um número inteiro, *umente* sempre o valor de n para o próximo inteiro

Exemplo: Um economista deseja estimar a renda média para o primeiro ano de trabalho de um bacharel. Quantos valores de renda devem ser tomados, se quer-se ter 95% de confiança em que a média amostral esteja a menos de \$500 da verdadeira média populacional? Suponha que saibamos por estudos prévios que $\sigma = \$6250$

$$n = \frac{[z_{\alpha/2} \sigma]^2}{E^2}$$

$$= \frac{(1,96 \cdot 6250)^2}{500^2}$$

$$= 600,25$$

$$= 601 \text{ bacharéis}$$

Com tal amostra teremos 95% de confiança e, que a média amostral difira em menos de \$500 da verdadeira média populacional.

Determinação da Estimativa Pontual e E a partir do Intervalo de Confiança

Estimativa pontual of μ :

$$\bar{x} = \frac{(\text{limite superior do intervalo}) + (\text{limite inferior do intervalo})}{2}$$

Margem de Erro:

$$E = \frac{(\text{limite superior do intervalo}) - (\text{limite inferior do intervalo})}{2}$$

3

Estimativa de Média Populacional: Pequenas Amostras

Pequenas Amostras

Hipóteses

Se:

1. $n \leq 30$
2. A amostra é uma amostra aleatória simples.
3. A população é normalmente distribuída.

Caso 1 (σ é conhecido): Pouco realístico. Utilizar mesmo método visto anteriormente (grandes amostras)

Case 2 (σ é desconhecido): Utilizar a distribuição t de Student.

Distribuição t de Student

Se a distribuição de uma população é essencialmente normal, então a distribuição de

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$

- ❖ é essencialmente uma Distribuição t de Student para todas amostras de tamanho n .
- ❖ é utilizada na determinação de valores críticos denotados por $t_{\alpha/2}$

Definição

Graus de Liberdade (*gl*)

corresponde ao número de valores amostrais que podem variar após terem sido impostas certas restrições a todos os valores.

$$gl = n - 1$$

nesta seção

Margem de Erro para Estimativa de μ

Baseada em σ Desconhecido e uma Amostra Aleatória Simples de uma População Normalmente Distribuída

Fórmula 6-2

$$E = t_{\alpha/2} \frac{s}{\sqrt{n}}$$

onde $t_{\alpha/2}$ tem $n - 1$ graus de liberdade

Intervalo de Confiança para a Estimativa de μ

Baseado em um σ Desconhecido e uma Amostra Aleatória Simples de uma População Normalmente Distribuída

$$\bar{x} - E < \mu < \bar{x} + E$$

onde $E = t_{\alpha/2} \frac{s}{\sqrt{n}}$

$t_{\alpha/2}$ lido na Tabela A-3

Tabela A-3 Distribuição *t*

Graus de liberdade	.05 (unilateral)	.01 (unilateral)	.025 (unilateral)	.05 (unilateral)	.10 (unilateral)	.25 (unilateral)
	.01 (bilateral)	.02 (bilateral)	.05 (bilateral)	.10 (bilateral)	.20 (bilateral)	.50 (bilateral)
1	63,657	31,821	12,706	6,314	3,078	1,000
2	9,925	6,965	4,303	2,920	1,886	,816
3	5,841	4,541	3,182	2,353	1,638	,765
4	4,604	3,747	2,776	2,132	1,533	,741
5	4,032	3,365	2,571	2,015	1,476	,727
6	3,707	3,143	2,447	1,943	1,440	,718
7	3,500	2,998	2,365	1,895	1,415	,711
8	3,355	2,896	2,306	1,860	1,397	,706
9	3,250	2,821	2,262	1,833	1,383	,703
10	3,169	2,764	2,228	1,812	1,372	,700
11	3,106	2,718	2,201	1,796	1,363	,697
12	3,054	2,681	2,179	1,782	1,356	,696
13	3,012	2,650	2,160	1,771	1,350	,694
14	2,977	2,625	2,145	1,761	1,345	,692
15	2,947	2,602	2,132	1,753	1,341	,691
16	2,921	2,584	2,120	1,746	1,337	,690
17	2,898	2,567	2,110	1,740	1,333	,689
18	2,878	2,552	2,101	1,734	1,330	,688
19	2,861	2,540	2,093	1,729	1,328	,688
20	2,845	2,528	2,086	1,725	1,325	,687
21	2,831	2,518	2,080	1,721	1,323	,686
22	2,819	2,508	2,074	1,717	1,321	,686
23	2,807	2,500	2,069	1,714	1,320	,685
24	2,797	2,492	2,064	1,711	1,318	,685
25	2,787	2,485	2,060	1,708	1,316	,684
26	2,779	2,479	2,056	1,706	1,315	,684
27	2,771	2,473	2,052	1,703	1,314	,684
28	2,763	2,467	2,048	1,701	1,313	,683
29	2,756	2,462	2,045	1,699	1,311	,683
Grande (z)	2,575	2,327	1,960	1,645	1,282	,675

Propriedades Importantes da Distribuição t de Student

1. A distribuição t de Student é diferente, conforme o tamanho da amostra.
2. A distribuição t de Student tem a mesma forma geral simétrica (forma de sino) que a distribuição normal, mas reflete a maior variabilidade (com distribuições mais amplas) que é esperada em pequenas amostras.
3. A distribuição t de Student tem média $t = 0$ (tal como a distribuição normal padronizada, com média $z = 0$).
4. O desvio-padrão da distribuição t de Student varia com o tamanho da amostra, mas é superior a 1 (ao contrário da distribuição normal padronizada, em que $\sigma = 1$).
5. Na medida em que aumenta o tamanho n da amostra, a distribuição t de Student se aproxima mais e mais da distribuição normal padronizada. Para valores $n > 30$, as diferenças são tão pequenas que podemos utilizar os valores críticos z em lugar de elaborar uma tabela muito maior de valores críticos t . (Os valores da última linha da Tabela A-3 são iguais aos valores críticos z correspondentes da distribuição normal padronizada.)

Distribuições t de Student para $n = 3$ e $n = 12$

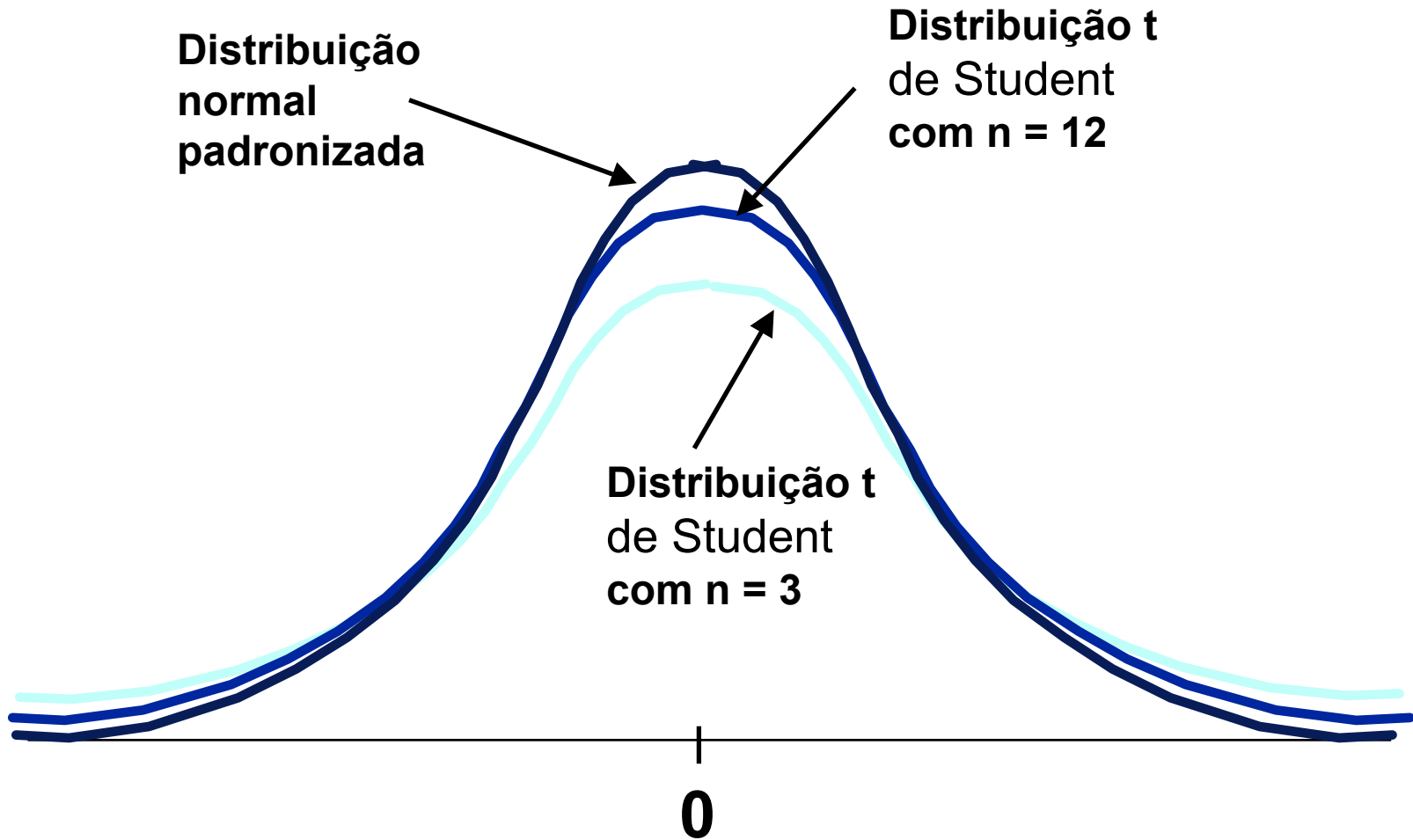


Figura 6-5

Exemplo: Para avaliar método de treinamento foi selecionada uma amostra de 15 empregados, verificando-se o tempo de treinamento necessário. Obteve-se uma média de 53,87 dias e um desvio-padrão de 6,82 dias. Determine o intervalo de 95% de confiança para μ , o tempo médio para treinar todos os empregados da empresa. (A distribuição dos 15 tempos parece ter a forma de sino.)

$$\bar{x} = 53,87$$

$$s = 6,82$$

$$E = \frac{t_{\alpha/2} s}{\sqrt{n}} = \frac{(2,145)(6,82)}{\sqrt{15}} = 3,78$$

$$\alpha = 0,05$$

$$\alpha/2 = 0,025$$

$$t_{\alpha/2} = 2,145$$

$$\bar{x} - E < \mu < \bar{x} + E$$

$$53,87 - 3,78 < \mu < 53,87 + 3,78$$

$$50,09 < \mu < 57,65$$

Estamos 95% confiantes que este intervalo contenha a média de tempo de treinamento de todos os empregados.

4

Estimativa de Proporção Populacional

Hipóteses

- 1. A amostra é uma amostra aleatória simples.**
- 2. Estão satisfeitas as condições para a distribuição binomial estão satisfeitas.**
- 3. A distribuição normal pode ser usada para aproximar a distribuição da proporção amostral, ou seja, as condições de que $np \geq 5$ e $np(1-p) \geq 5$ estão satisfeitas .**

Notação para Proporções

$p =$ proporção populacional

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

proporção amostral

de x **sucessos** em uma amostra de tamanho n

↑
(pronuncia-se
'p-chapéu')

Definição

Estimativa Pontual

A proporção amostral \hat{p} é a melhor estimativa pontual da proporção populacional p .

Margem de Erro da Estimativa de p

Fórmula 6-4

$$E = z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Intervalo de Confiança para a Proporção Populacional

$$\hat{p} - E < p < \hat{p} + E$$

onde

$$E = z_{\alpha/2} \sqrt{\frac{\hat{p} \hat{q}}{n}}$$

Tamanho da Amostra para Estimar a Proporção p

Quando se conhece uma estimativa de p :

$$n = \frac{(z_{\alpha/2})^2 \hat{p} (1 - \hat{p})}{E^2} \quad \text{Fórmula 6-5}$$

Quando **não** se conhece uma estimativa de p :

$$n = \frac{(z_{\alpha/2})^2 0.25}{E^2} \quad \text{Fórmula 6-6}$$

Duas Fórmulas para Determinação do Tamanho Amostral

$$n = \frac{(z_{\alpha/2})^2 \hat{p} (1 - \hat{p})}{E^2}$$

$$n = \frac{(z_{\alpha/2})^2 (0.25)}{E^2}$$

Exemplo: Desejamos determinar, com uma margem de erro de 4%, a proporção de donas de casa que usam e-mail. Assumindo que queremos uma confiança de 90% em nossos resultados, qual deveria ser o tamanho da amostra. Um estudo piloto indicou que 16,9 das donas de casa usam e-mail.

$$\begin{aligned}n &= \frac{[z_{\alpha/2}]^2 \hat{p} \hat{q}}{E^2} \\&= \frac{[1,645]^2 (0,169)(0,831)}{0,04^2} \\&= 237,51965 \\&= 238 \text{ donas de casa}\end{aligned}$$

Para estarmos 90% confiantes que nossa proporção amostral esteja no máximo 4% distante da verdadeira proporção de todas as donas de casa, deveríamos escolher aleatoriamente e pesquisar 238 donas de casa.

Exemplo: Desejamos determinar, com uma margem de erro de 4%, a proporção de donas de casa que usam e-mail. Considere uma confiança de 90% em nossos resultados. Qual deveria ser o tamanho da amostra. **Não há informação prévia sugerindo um possível valor para a proporção amostral.**

$$n = \frac{[z_{\alpha/2}]^2 (0,25)}{E^2}$$

$$= \frac{(1,645)^2 (0,25)}{0,04^2}$$

$$= 422,81641$$

$$= 423 \text{ donas de casa}$$

Como não há informação prévia, necessitamos uma amostra maior para alcançar os mesmos resultados com 90% de confiança e uma margem de erro de no máximo 4%.