Intervalos Estatísticos para uma única Amostra - parte I

Intervalo de confiança para média

Marcos Oliveira Prates

2012/02

1. Introdução
2. Intervalo para média com variância conhecida
3. Intervalo de confiança para μ com amostra grande
4. Intervalo para média com variância desconhecida

Objetivos

Ao final deste capítulo você deve ser capaz de:

- Construir intervalos de confiança para média de uma população.
- Esses intervalos serão construídos usando a distribuição normal e a distribuição t-student.

Introdução

- Um parâmetro é estimado a partir de dados amostrais.
- Exemplo: queremos saber a viscosidade média de um produto.
- Uma estimativa seria

$$\hat{\mu} = \bar{x} = 1000\,.$$

- Essa estimativa não diz o quão próximo $\hat{\mu}$ está do verdadeiro valor μ.
- É provável que a média esteja entre 900 e 1100?
- E entre 990 e 1010?
- Precisamos de limites que representam valores plausíveis de μ.
- Esses limites são um **intervalo de confiança**.

- Uma estimativa de intervalo para um parâmetro:
  - **intervalo de confiança**.
- Não podemos ter certeza de que o intervalo contém o valor verdadeiro do parâmetro.
- Pois utilizamos apenas a informação de uma amostra.
- Porém podemos ter alta confiança de que o intervalo contém o valor verdadeiro.

- Outros tipos de intervalo:
  - intervalo de tolerância;
  - intervalo de previsão.

Intervalo de tolerância

- Exemplo: considere que os dados de viscosidade de um produto têm distribuição normal.
- Podemos encontrar limites que delimitam 95% dos valores de viscosidade.
- Encontrar limites tais que 95% das observações estão dentro deles.
- Para o caso da normal esse intervalo é

$$\mu - 1,96\sigma, \mu + 1,96\sigma$$

Intervalo de previsão

- Fornece limites para observações futuras.
- Exemplo: queremos encontrar os limites da viscosidade para uma nova medida.
- Podemos encontrar limites para o preço de um produto no próximo mês.

Intervalo para média com variância conhecida

- Vamos considerar temos uma população normal.
- Supomos que a variância σ^2 é conhecida.
- Essa não é uma suposição razoável.
- Geralmente não sabemos o verdadeiro valor da variância.
- Vamos considerar primeiro esse caso mais simples.
- Em seguida trataremos casos mais gerais.

- Suponha que
$$X_1, X_2, \ldots, X_n$$
é uma amostra de uma distribuição normal.
- A média μ é desconhecida e a variância σ^2 é conhecida.
- Sabemos que
$$\bar{X} \sim N(\mu, \sigma^2/n)\,.$$
- Padronizamos o $\bar{X}$
$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$
e
$$Z \sim N(0, 1)\,.$$

- Uma estimativa do intervalo de confiança para μ é

$$l \leq \mu \leq u \, .$$

- Onde os extremos l e u são calculados a partir da amostra.
- Diferentes amostras podem produzir valores distintos de l e u.
- Eles são valores observados de variáveis aleatórias L e U.
- Queremos determinar os valores de L e U tal que

$$P(L \leq \mu \leq U) = 1 - \alpha \quad \text{para} \quad 0 \leq \alpha \leq 1 \, .$$

Observação:

- o parâmetro μ está fixo;
- as variáveis aleatórias são L e U.

- Existe uma probabilidade $1 - \alpha$ de selecionarmos uma amostra tal que:
  - o IC conterá o verdadeiro valor do parâmetro.
- Depois de observado o valor da amostra

$$X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$$

calculamos l e u e o intervalo resultante é

$$l \leq \mu \leq u\,.$$

- Os valores l e u são denominados:
  - l é o **limite inferior** do intervalo;
  - u é o **limite superior** do intervalo.
- O valor $1 - \alpha$ é o **coeficiente de confiança** do intervalo.

- Como Z tem distribuição normal padrão

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

onde $z_{1-\frac{\alpha}{2}}$ é obtido a partir da tabela do normal padrão de modo que

$$P(-z_{1-\frac{\alpha}{2}} \leq Z \leq z_{1-\frac{\alpha}{2}}) = 1-\alpha\,.$$

- Podemo isolar o μ e ficamos com

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

- Então

$$L = \bar{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \quad U = \bar{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\,.$$

Intervalo para média com variância conhecida

- Seja $\bar{x}$ a média observada de uma amostra aleatória.
- Suponha que a amostra tem tamanho n.
- Ela é proveniente de uma população normal com variância conhecida σ^2.
- Um intervalo de $100(1-\alpha)\%$ de confiança para μ é

$$\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

onde $z_{1-\frac{\alpha}{2}}$ é tal que

$$P(Z \leq z_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}\,.$$

Exemplo:

- São feitas medidas de energia de impacto em 10 corpos de prova.
- Os valores observados são

$$64,1;\quad 64,7;\quad 64,5;\quad 64,6;\quad 64,5;\quad 64,3;\quad 64,6;$$

$$64,8;\quad 64,2;\quad 64,3\,.$$

- Suponha que a energia de impacto é normalmente distribuída com variância 1J.
- Querermos encontrar o IC de 95% de confiança para μ.

Exemplo: (solução)

- Temos que

$$z_{1-\frac{\alpha}{2}} = z_{0,975} = 1,96 \quad n = 10 \quad \sigma = 1 \quad \bar{x} = 64,46\,.$$

- O intervalo com 95% de confiança é

$$\bar{x} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

$$64,46 - 1,96\frac{1}{\sqrt{10}} \leq \mu \leq 64,46 + 1,96\frac{1}{\sqrt{10}}$$

$$63,84 \leq \mu \leq 65,08\,.$$

- Uma faixa de valores altamente plausíveis para μ é $[63,84;\ 65,08]$J.

Interpretação do intervalo de confiança

- Considere um intervalo com confiança $100(1-\alpha)\%$.
- Se repetissem o experimento um número infinito de vezes.
- Se para cada uma desses experimentos calculássemos o intervalo dessa forma.
- $100(1-\alpha)\%$ deles iriam conter o verdadeiro valor de μ.

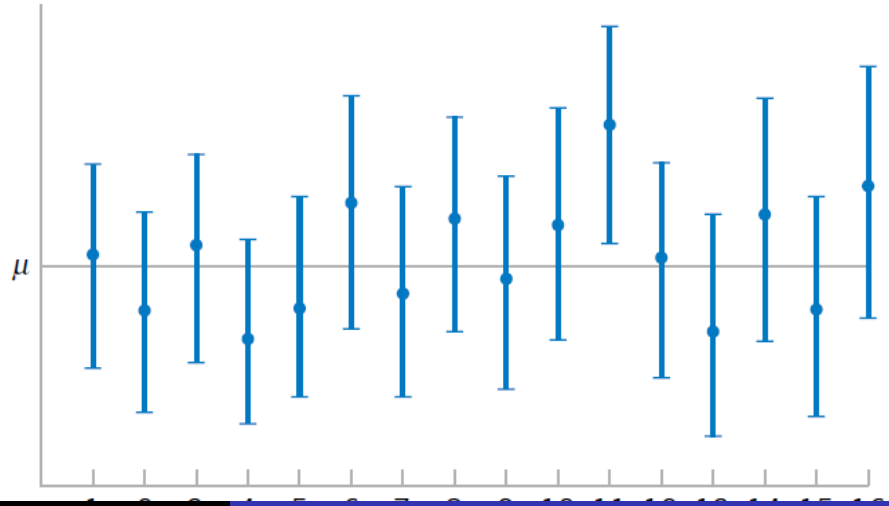


Figure 8-1 Repeated construction of a con

Confiança vs Precisão

- Se quisermos uma confiança de 99% ao invés de 95%.
- Devemos aumentar o comprimento do intervalo.
- O comprimento com 95% é

$$2(1,96\sigma/\sqrt{n})$$

enquanto com 99% é

$$2(2,58\sigma/\sqrt{n})\ .$$

- O **comprimento** do intervalo mede a **precisão** da estimativa.
- Se quisermos ter muita confiança teremos um intervalo menos preciso.
- A precisão é **inversamente** relacionada com o nível de confiança.

Escolha do tamanho de amostra:

- Quando estimamos μ usando $\bar{x}$ comentemos um erro

$$|\bar{x} - \mu| \leq z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

  (veja a figura)
- Fixado um nível de confiança $100(1-\alpha)\%$.
- Podemos escolher um tamanho de amostra *n* dependendo do erro máximo (margem de erro) que queremos cometer

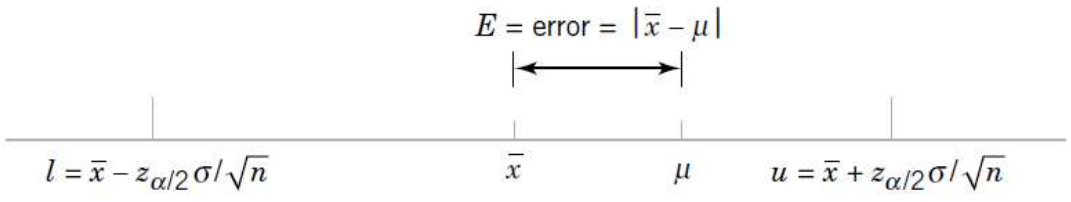


Figure 8-2 Error in estimating μ with $\bar{x}$.

- Escolhemos n tal que o erro máximo cometido é

$$E = z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

com $100(1-\alpha)\%$ de confiança.

- Isolando o n temos que

$$n = \left(\frac{z_{1-\frac{\alpha}{2}} \sigma}{E} \right)^2 .$$

Exemplo:

- Considere o exemplo das energias medidas em corpos de prova.
- Queremos construir um intervalo com 95% de confiança para energia média μ.
- O intervalo deve ter comprimento máximo de 1J.
- Então o erro de estimação máximo é 1/2J.
- Temos então que

$$E = 0,5 \quad \sigma = 1 \quad z_{1-\frac{\alpha}{2}} = 1,96\,.$$

- O tamanho de amostra requerido é

$$n = \left(\frac{z_{1-\frac{\alpha}{2}}\sigma}{E}\right)^2 \Rightarrow n = \left(\frac{(1,96)(1)}{0,5}\right)^2 = 15,37\,.$$

- O tamanho de amostra mínimo é 16.

Limites unilaterias de confiança

- Podemos fazer

$$l = -\infty \quad \text{ou} \quad u = \infty\,.$$

- Seja $\bar{x}$ uma estimativa para μ.
- O **limite superior** com $100(1-\alpha)\%$ de confiança é

$$\mu \leq \bar{x} + z_{1-\alpha}\sigma/\sqrt{n}.$$

- O **limite inferior** com $100(1-\alpha)\%$ de confiança é

$$\bar{x} - z_{1-\alpha}\sigma/\sqrt{n} \leq \mu.$$

Intervalo de confiança para μ com amostra grande

- Suponha que temos uma amostra de tamanho grande (pelo menos 40).
- Não vamos mais supor que as variáveis tem distribuição normal e nem que a variância é conhecida.
- Temos uma amostra

$$X_1, \ldots, X_n$$

  com uma média μ e variância σ^2 desconhecidas.
- Mesmo não sabendo a distribuição da população
  - o Teorema Central do Limite garante a distribuição de $\bar{X}$ se aproxima de uma normal padrão.
- O valor de σ é estimado por

$$S = \sqrt{\frac{\sum_i (X_i - \bar{X})^2}{n-1}}$$

Intervalo para média com amostra grande

- Queremos estimar μ.
- A estimativa pontual é $\bar{x}$.
- Não sabemos o valor de σ^2.
- Se n é grande

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

  tem uma distribuição que se aproxima da normal padrão.
- O intervalo com $100(1-\alpha)\%$ de confiança é dado por

$$\bar{x} - z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \, .$$

Exemplo:

- Uma amostra de 53 peixes é selecionada de um lago da Flórida.
- Mediu-se a concentração de mercúrio no tecido muscular.
- As figuras abaixo mostram o histograma e o gráfico de probabilidade para essa amostra.
- Ambos mostram que a distribuição não é normal.

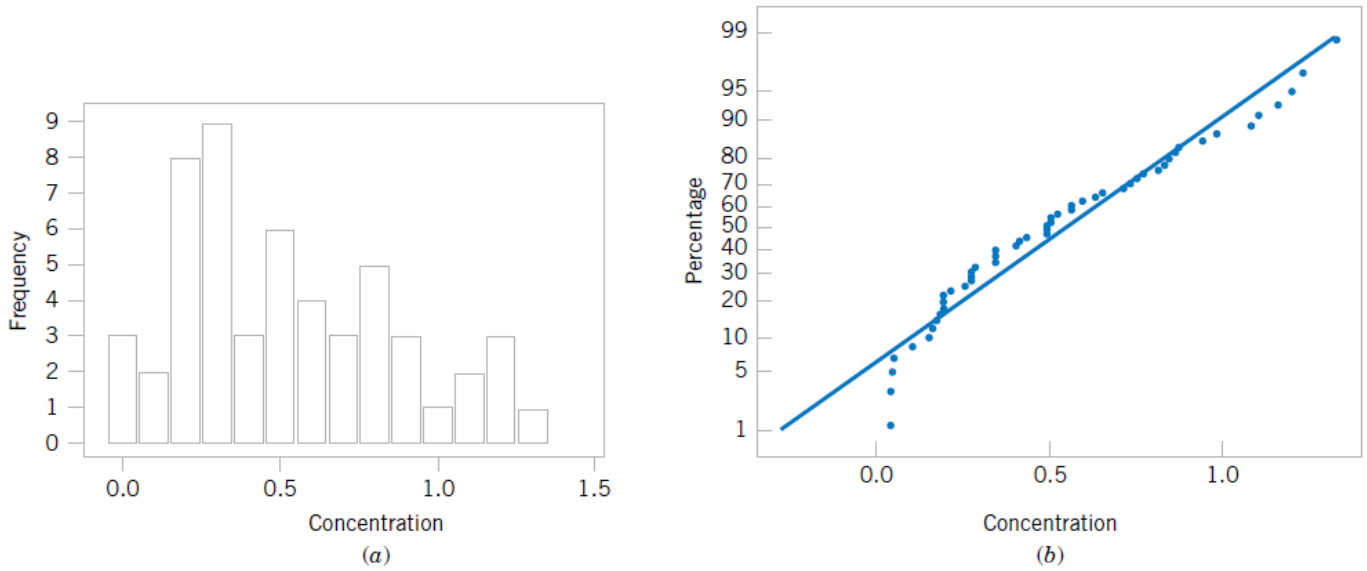


Figure 8.3 Mercury concentration in largemouth bass (a) Histogram, (b) Normal probability plot

Exemplo: (continuação)

- Queremos um intervalo de confiança para μ com 95% de confiança.
- $n > 40$ então não precisamos supor normalidade.
- Os dados são

$$n = 53 \quad \bar{x} = 0,5250 \quad s = 0,3486 \quad z_{0,975} = 1,96\,.$$

- O intervalo para μ é

$$\bar{x} - z_{0,975}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{0,975}\frac{s}{\sqrt{n}}$$

$$\bar{x} - 1,96\frac{0,3486}{\sqrt{53}} \leq \mu \leq \bar{x} + 1,96\frac{0,3486}{\sqrt{53}}$$

$$0,4311 \leq \mu \leq 0,6189\,.$$

Intervalo para média com variância desconhecida

- Vamos agora considerar casos em que:
  - a amostra é pequena;
  - a população é normalmente distribuída;
  - a variância é **desconhecida**.
- Muitas populações encontradas na prática são bem aproximadas pela normal.

Distribuição t

- Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de X.
- X tem distribuição normal com média μ e variância σ^2.
- μ e σ^2 são desconhecidos.
- Seja S o desvio padrão amostral

$$S = \sqrt{\frac{\sum_i (X_i - \bar{X})^2}{n-1}}\,.$$

- A variável aleatória

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

tem distribuição t-student com $n-1$ graus de liberdade.

- O gráfico abaixo mostra alguns exemplos de distribuição t.
- São parecidas com a normal.
- A distribuição t tem mais probabilidade na calda.

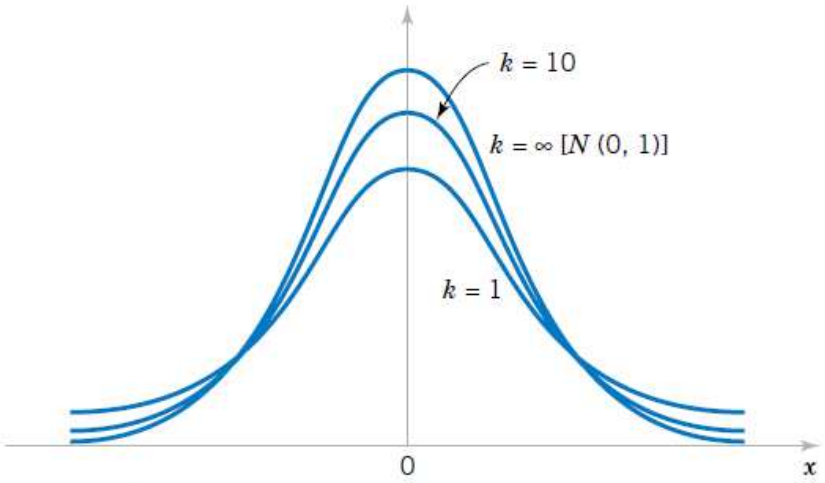


Figure 8-4 Probability density functions of several t distributions.

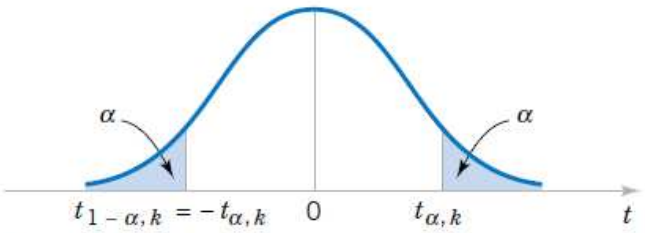


Figure 8-5 Percentage points of the t distribution.

- A Tabela V do apêndice mostra pontos percentuais da distribuição t.
- O valor $t_{\alpha;k}$ é o ponto da distribuição t com k graus de liberdade que deixa uma área α acima dele.
- Exemplo:

$$P(T_{10} > t_{0,05;10}) = P(T_{10} > 1,812) = 0,05\,.$$

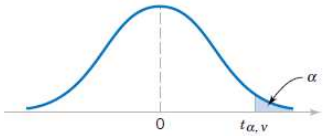


Table IV Percentage Points $t_{\alpha,\nu}$ of the t-Distribution

ν \ α	.40	.25	.10	.05	.025	.01	.005	.0025	.001	.0005
1	.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	.289	.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	.277	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	.271	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	.267	.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	.265	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	.263	.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	.262	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	.261	.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	.260	.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587

- A distribuição t é simétrica em torno do zero.
- Portanto

$$t_{1-\alpha,n} = -t_{\alpha,n}\,.$$

- Exemplo:

$$t_{0,95;10} = -t_{0,05;10} = -1,812\,.$$

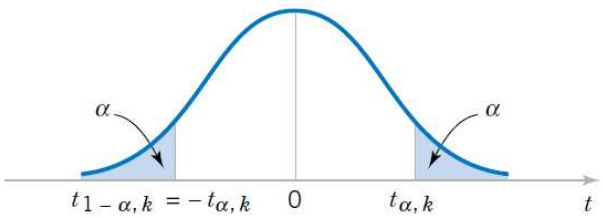


Figure 8-5 Percentage points of the t distribution.

Observação:

quando os graus de liberdade crescem a distribuição t se aproxima da Normal.

- Para encontrarmos o IC com $100(1-\alpha)\%$ vamos proceder como antes.
- Sabemos que

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

tem uma distribuição t com $n-1$ graus de liberdade.

- Encontramos $t_{\alpha/2;n-1}$ tal que

$$P(-t_{\alpha/2;n-1} \leq T \leq t_{\alpha/2;n-1}) = 1 - \alpha$$

ou seja

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

- Isolando o μ temos que

$$P\left(\bar{X} - t_{\alpha/2;n-1}S/\sqrt{n} \leq \mu \leq \bar{X} + t_{\alpha/2;n-1}S/\sqrt{n}\right) = 1 - \alpha\,.$$

- O intervalo de confiança para μ fica

$$\bar{x} - t_{\alpha/2;n-1}s/\sqrt{n} \leq \mu \leq \bar{x} + t_{\alpha/2;n-1}s/\sqrt{n}\,.$$

Intervalo para média com variância desconhecida

- Considere uma amostra

$$X_1, \ldots, X_n$$

vinda de uma população normal com média μ e variância σ^2.

- Os valores μ e σ^2 são desconhecidos.
- Sejam $\bar{x}$ e s a média e desvio padrão observados para essa amostra.
- Um intervalo com $100(1-\alpha)\%$ de confiança para μ é

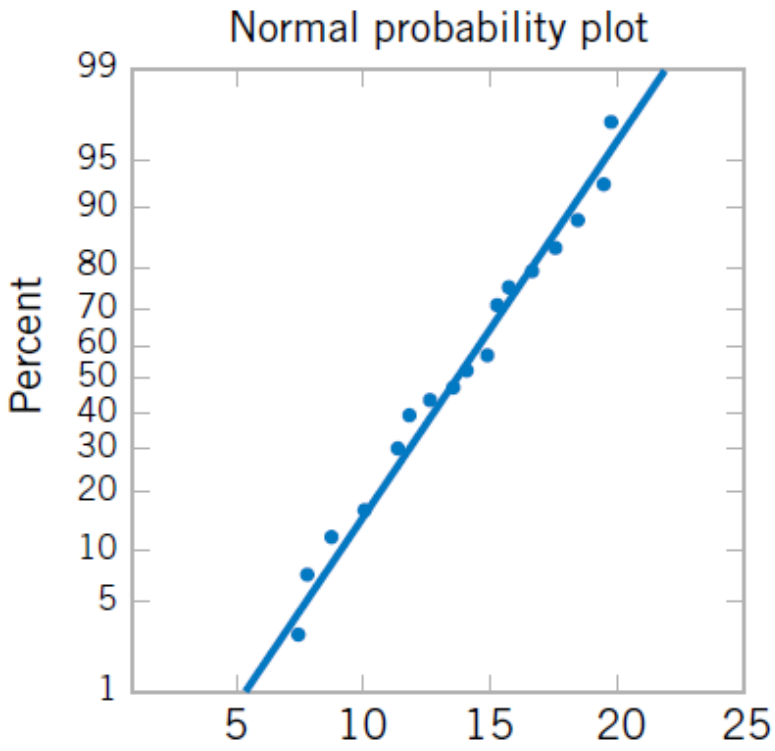
$$\bar{x} - t_{\alpha/2;n-1} s/\sqrt{n} \leq \mu \leq \bar{x} + t_{\alpha/2;n-1} s/\sqrt{n}$$

onde $t_{\alpha/2;n-1}$ é o ponto da distribuição t com $n-1$ graus de liberdade tal que

$$P(T_{n-1} \leq t_{\alpha/2}) = 1 - \alpha/2\,.$$

Exemplo:

- 22 corpos de prova são analisados.
- São registradas as cargas no ponto de falha.
- O gráfico de probabilidade abaixo mostra que a distribuição é próxima da normal.



Exemplo: (continuação)

- Os dados são

$$\bar{x} = 13,71 \quad s = 3,55 \quad n = 22\,.$$

- Queremos um intervalo com 95% de confiança.
- Os graus de liberdade são $n - 1 = 21$.
- Pela tabela

$$t_{0,025;21} = 2,080\,.$$

ν \ α	.40	.25	.10	.05	.025	.01	.005	.0025	.001	.0005
1	.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	.289	.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	.277	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	.271	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	.267	.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	.265	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	.263	.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	.262	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	.261	.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	.260	.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	.260	.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	.259	.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	.259	.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	.258	.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	.258	.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	.258	.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	.257	.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	.257	.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	.257	.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	.257	.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	.257	.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819

Exemplo: (continuação)

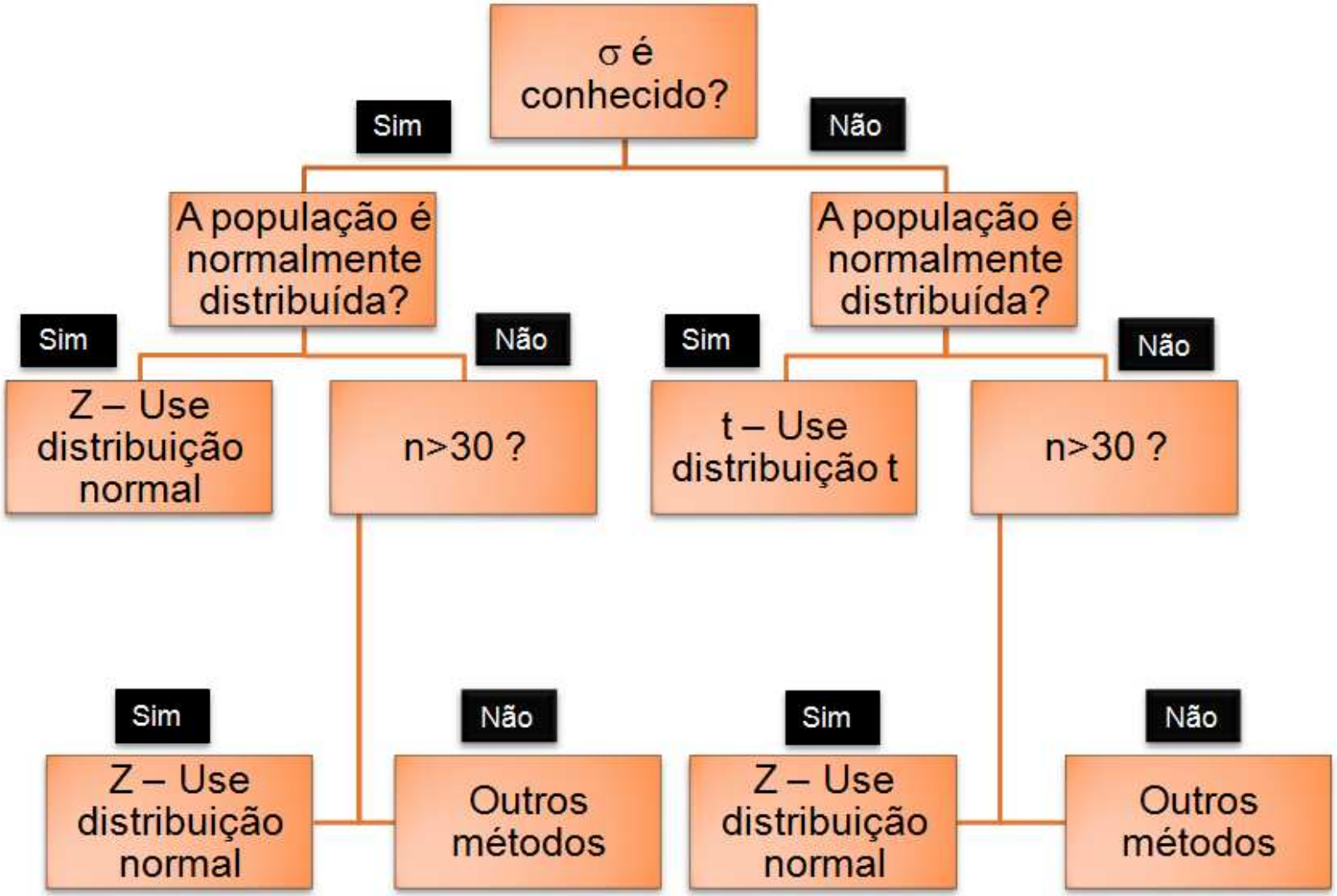
- O intervalo resultante é

$$\bar{x} - t_{\alpha/2;n-1} s/\sqrt{n} \leq \mu \leq \bar{x} + t_{\alpha/2;n-1} s/\sqrt{n}$$

$$13,71 - 2,080(3,55)/\sqrt{22} \leq \mu \leq 13,71 + 2,080(3,55)/\sqrt{22}$$

$$12,14 \leq \mu \leq 15,28\,.$$

- O intervalo é razoavelmente amplo por causa da variabilidade dos dados.


σ é conhecido?
Sim
Não
A população é normalmente distribuída?
A população é normalmente distribuída?
Sim
Não
Sim
Não
Z – Use distribuição normal
n>30 ?
t – Use distribuição t
n>30 ?
Sim
Não
Sim
Não
Z – Use distribuição normal
Outros métodos
Z – Use distribuição normal
Outros métodos