

Calculado intervalo de confiança e
comparando resultados com uma
pequena amostra

Distribuição t

- Fórmula quase a mesma:

$$\bar{x} \pm t_{[1-\alpha/2; n-1]} \left(\frac{s}{\sqrt{n}} \right)$$

- Usável para populações normalmente distribuídas!
- Mas funciona para pequenas amostras
- $n-1$ indica o grau de liberdade

Exemplo da Distribuição t

- 10 amostras de chegada de transações: 148 166 170 191 187 114 168 180 177 204
- Média da amostra $\bar{x} = 170.5$.
Desvio padrão $s = 25.1$, $n = 10$
- Calcule o intervalo de confiança de 90%:

$$\bar{x} \pm t_{[1-\alpha/2; n-1]} \left(\frac{s}{\sqrt{n}} \right)$$

- Quanto é $t_{[1-0.1/2; 10-1]}$

α						
Degrees of freedom	.005 (one tail) .01 (two tails)	.01 (one tail) .02 (two tails)	.025 (one tail) .05 (two tails)	.05 (one tail) .10 (two tails)	.10 (one tail) .20 (two tails)	.25 (one tail) .50 (two tails)
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
Large (z)	2.575	2.327	1.960	1.645	1.282	.675

Exemplo da Distribuição t

- 10 amostras de chegada de transações: 148 166 170 191 187 114 168 180 177 204
- Média da amostra $\bar{x} = 170.5$.
Desvio padrão $s = 25.1$, $n = 10$
- Calcule o intervalo de confiança de 90%:

$$\bar{x} \pm t_{[1-\alpha/2; n-1]} \left(\frac{s}{\sqrt{n}} \right)$$

$$170.5 \pm (1.833) \frac{25.1}{\sqrt{10}} = (156.0, 185.0)$$

- Calcule agora o intervalo de 99% de confiança

α						
Degrees of freedom	.005 (one tail) .01 (two tails)	.01 (one tail) .02 (two tails)	.025 (one tail) .05 (two tails)	.05 (one tail) .10 (two tails)	.10 (one tail) .20 (two tails)	.25 (one tail) .50 (two tails)
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	3.255	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
Large (z)	2.575	2.327	1.960	1.645	1.282	.675

Exemplo da Distribuição t

- 10 amostras de chegada de transações: 148 166 170 191 187 114 168 180 177 204
- Média da amostra $\bar{x} = 170.5$.
Desvio padrão $s = 25.1$, $n = 10$
- Calcule o intervalo de confiança de 90%:

$$\bar{x} \pm t_{[1-\alpha/2; n-1]} \left(\frac{s}{\sqrt{n}} \right)$$

$$170.5 \pm (3.250) \frac{25.1}{\sqrt{10}} = (144.7, 196.3)$$

Example 13.2

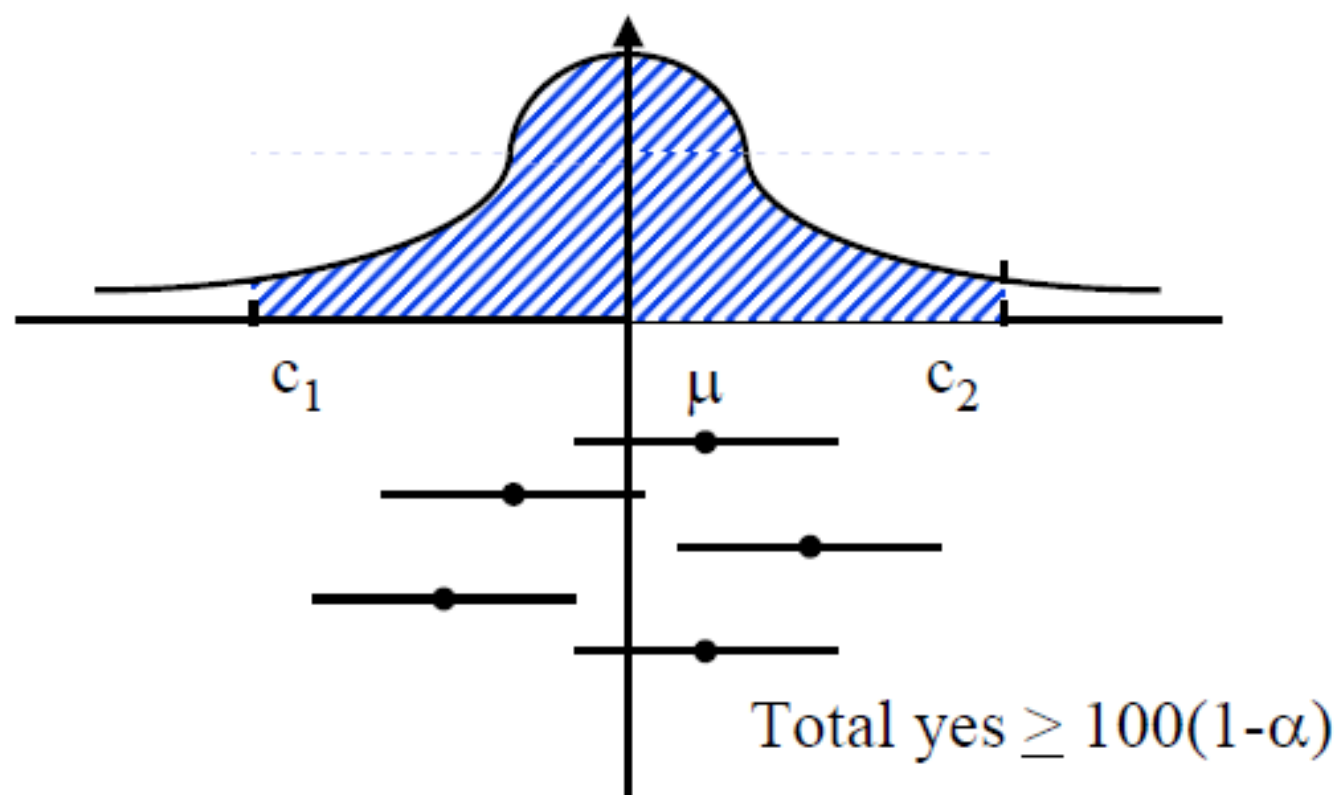
- ❑ Sample: -0.04, -0.19, 0.14, -0.09, -0.14, 0.19, 0.04, and 0.09.
 - ❑ Mean = 0, Sample standard deviation = 0.138.
 - ❑ For 90% interval: $t_{[0.95;7]} = 1.895$
 - ❑ Confidence interval for the mean
- $$0 \mp 1.895 \times 0.138 = 0 \mp 0.262 = (-0.262, 0.262)$$

Tomando decisões sobre os dados experimentais

- Por que usamos intervalos de confiança?
 - Sumarizar o erro na média da amostra
 - Prover elementos para saber se a amostra é significativa
 - Permitir comparações à luz dos erros

Confidence Interval: Meaning

- If we take 100 samples and construct confidence interval for each sample, the interval would include the population mean in 90 cases.



Comparando alternativas

- Em um projeto de pesquisa, normalmente, procura-se o melhor sistema, o melhor algoritmo, ...
- Métodos diferentes para observações pareadas (com par) e não pareadas (sem par).
 - Pareadas se o i -ésimo teste em cada sistema foi o mesmo
 - Não pareadas, caso contrário.

Pareadas

- Trata o problema como uma amostra de n pares
- Para cada teste, calcule a diferença dos resultados
- Calcule o intervalo de confiança para a diferença média
- Se o intervalo inclui 0 (zero), os objetos de comparação não são diferentes com a confiança especificada
- Se o intervalo não inclui zero, o sinal da diferença indica qual dos objetos é melhor com a confiança especificada e baseado nos dados experimentais

Pareados

- Considere dois métodos de busca A e B que são avaliados em função do número de documentos relevantes (em um total de 100) que cada um retorna
- Num teste com várias consultas, o algoritmo A retorna mais documentos relevantes que o B?
- Amostra de testes com 14 consultas:

■ Alg. A	4	5	0	11	6	6	3	12	9	5	6	3	1	6
■ Alg. B	2	7	7	6	0	7	10	6	2	2	4	2	2	0

Pareados

- Diferenças entre A e B: 2 -2 -7 5 6 -1 -7 6 7 3 2
1 -1 6
- Média 1.4, nível de confiança de 90% :
 - (-0.75, 3.6)
 - Não se pode rejeitar a hipótese de que a diferença é 0 e portanto os dois algoritmos tem desempenho similar.
- Com nível de confiança de 70% é (0.10 , 2.76)
 - A tem desempenho melhor que B

Example 13.3

- Difference in processor times: $\{1.5, 2.6, -1.8, 1.3, -0.5, 1.7, 2.4\}$.
- Question: Can we say with 99% confidence that one is superior to the other?

$$\text{Sample size} = n = 7$$

$$\text{Mean} = 7.20/7 = 1.03$$

$$\text{Sample variance} = (22.84 - 7.20*7.20/7)/6 = 2.57$$

$$\text{Sample standard deviation} = \sqrt{2.57} = 1.60$$

$$\text{Confidence interval} = 1.03 \mp t * 1.60/\sqrt{7} = 1.03 \mp 0.6t$$

$$100(1 - \alpha) = 99, \alpha = 0.01, 1 - \alpha/2 = 0.995$$

$$t_{[0.995; 6]} = 3.707$$

- 99% confidence interval = $(-1.21, 3.27)$

Example 13.5

- Performance: $\{(5.4, 19.1), (16.6, 3.5), (0.6, 3.4), (1.4, 2.5), (0.6, 3.6), (7.3, 1.7)\}$. Is one system better?
- Differences: $\{-13.7, 13.1, -2.8, -1.1, -3.0, 5.6\}$.

Sample mean = -0.32

Sample variance = 81.62

Sample standard deviation = 9.03

Confidence interval for the mean = $-0.32 \pm t\sqrt{(81.62/6)}$
 $= -0.32 \pm t(3.69)$

$t_{[0.95,5]} = 2.015$

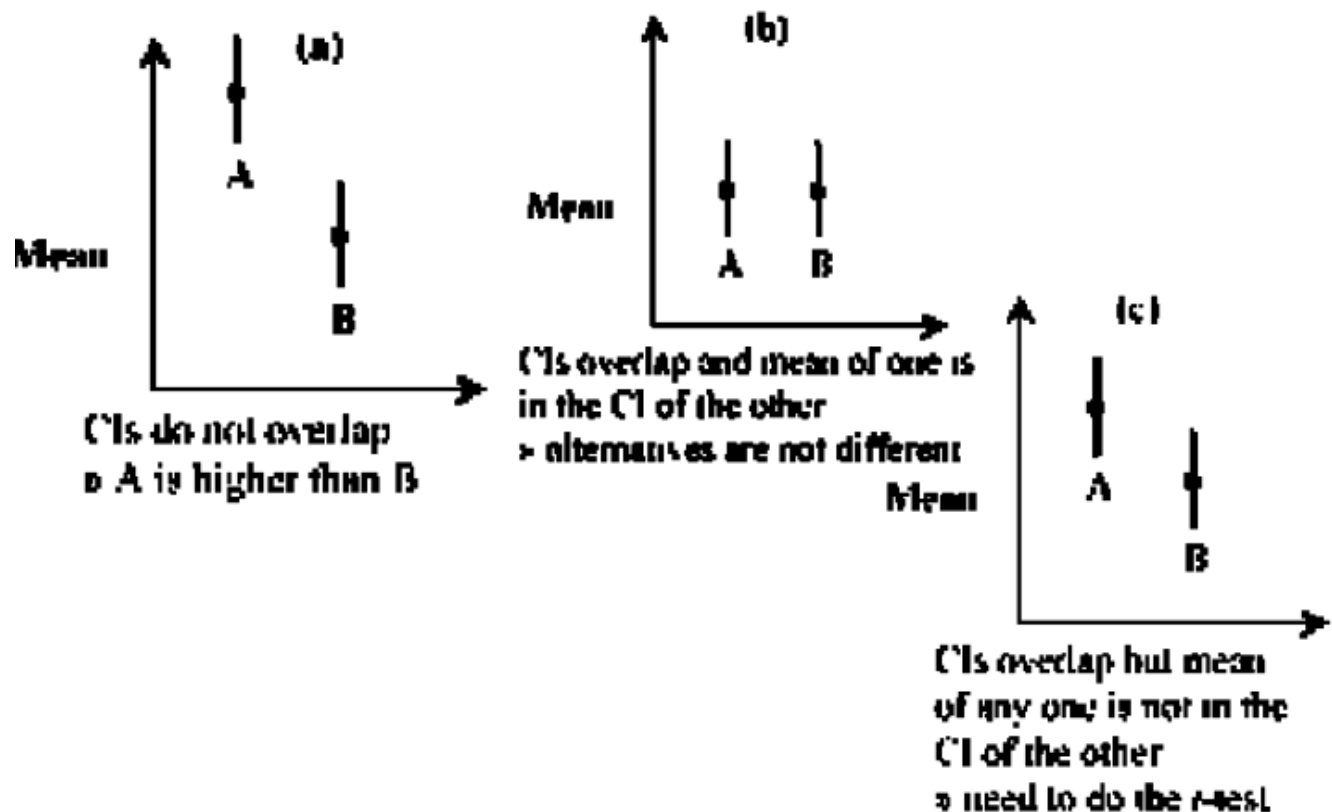
90% confidence interval = $-0.32 \pm (2.015)(3.69)$
 $= (-7.75, 7.11)$

- Answer: No. They are not different.

Não pareadas

- O número de experimentos comuns não precisa ser o mesmo

Approximate Visual Test



Teste t

1. Compute as médias das amostras \bar{x}_a e \bar{x}_b
2. Compute os desvio-padrões s_a e s_b
3. Compute a diferença das médias = $\bar{x}_a - \bar{x}_b$
4. Compute o desvio padrão das diferenças:

$$s = \sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}$$

Teste t

5. Compute os graus efetivos de liberdade:

$$\nu = \frac{(s_a^2 / n_a + s_b^2 / n_b)^2}{\frac{1}{n_a + 1} \left(\frac{s_a^2}{n_a} \right)^2 + \frac{1}{n_b + 1} \left(\frac{s_b^2}{n_b} \right)^2} - 2$$

6. Compute o intervalo de confiança:

$$(\bar{x}_a - \bar{x}_b) \mp t_{[1-\alpha/2; \nu]} S$$

7. Se o intervalo inclui zero, não há diferença

Teste t

O tempo de processamento necessario para executar uma tarefa foi medido em dois sistemas.

Os tempos no sistema A foram:

{5.36, 16.57, 0.62, 1.41, 0.64, 7.26}.

Os tempos no sistema B foram:

{19.12, 3.52, 3.38, 2.50, 3.60, 1.74}.

Os dois sistemas sao significativamente diferentes?

Teste t

Diferença das médias $\overline{x}_A - \overline{x}_B = -0.33$

Desvio Padrão da diferença média $s = 3.698$

Graus de liberdade $\nu = 11.921$

0.95 – *quantil* da *va* t com 12 graus de liberdade

$$t_{[0.95,12]} = 1.71$$

IC de 90% para a diferença das médias = (-6.92, 6.26)

IC inclui 0, logo os dois sistemas NAO são diferentes neste nível de confiança

Nível de confiança

- Os níveis de confiança entre 90% e 95% são os mais comuns em artigos científicos em computação
- Em geral, use o maior valor que lhe permita estabelecer conclusões sólidas num processo experimental.

Referências

- Raj Jain. The Art of Computer System Performance Analysis: Techniques for Experimental Design, Measurement, Simulation and Modeling, John Wiley & Sons, Inc., 1991.
- Jay L. Devore, PROBABILIDADE E ESTATÍSTICA PARA ENGENHARIA E CIÊNCIAS. Cengage Learning, 2006.
- Material didático da profa. Jussara Almeida – DCC/UFMG.