

LISTA 03 - AVALIAÇÃO DE DESEMPENHO DE SISTEMAS

Allyson Ryan

Julho, 2025

Inferência Estatística e Análise Exploratória de Dados

- Um usuário faz requisições a um serviço implantado em um servidor de aplicação. O usuário requisitou o serviço 60 vezes, e cada tempo de ida e volta foi registrado na Tabela 1. Calcule a estimativa pontual e o intervalo de confiança do tempo médio de ida e volta, levando em consideração um grau de significância de 1%. Use o Diagrama 1 para ajudá-lo a definir o método de estimativa a ser adotado.

Round Trip Time (ms)					
174.522	169.979	226.603	204.162	237.945	184.63
207.427	235.224	197.601	245.88	192.201	225.052
248.632	164.121	204.372	200.591	187.506	193.431
163.319	207.785	210.865	257.392	182.555	189.911
226.91	189.487	201.256	171.594	226.534	185.609
214.897	196.263	202.133	164.962	207.537	156.319
207.967	207.109	212.371	190.152	222.02	191.638
227.484	178.319	177.461	199.031	189.881	214.884
187.168	249.784	176.739	167.482	179.122	162.646
153.537	209.748	214.733	170.858	171.478	217.238

Tabela 1: Round Trip Time

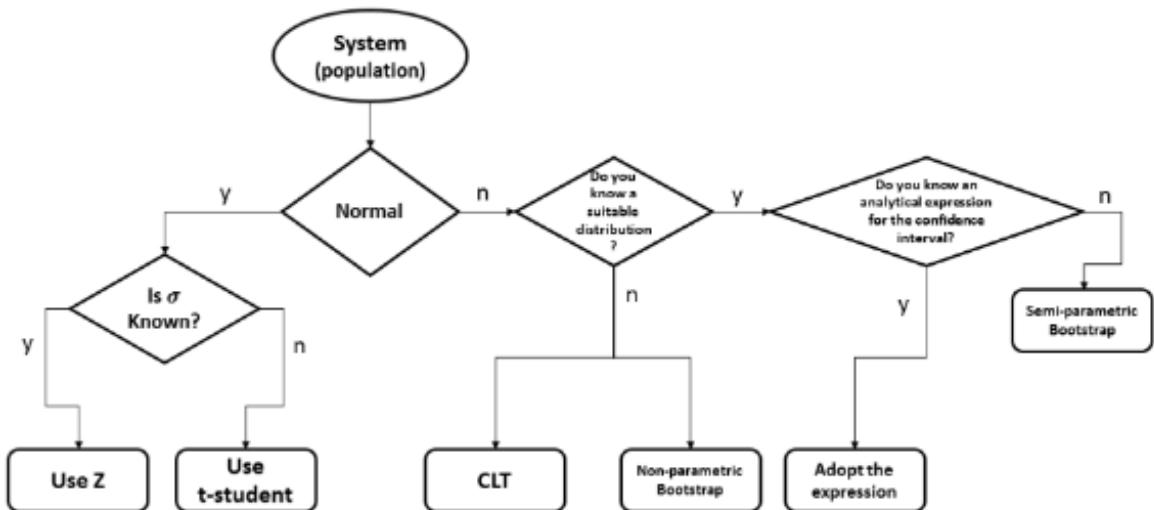


Diagrama 1: Orientação para Inferência Estatística de Médias

Estimativa pontual da média

A estimativa pontual da média (\bar{x}) representa o valor mais provável da média populacional com base na amostra coletada. Para obtê-la, divide-se a soma total dos tempos de resposta amostrados pelo tamanho da amostra ($n = 60$):

$$\bar{x} = \frac{11934,06}{60} = 198,901 \text{ ms}$$

Desvio padrão amostral

O desvio padrão amostral (s) mede a dispersão dos valores da amostra em relação à média. Para calculá-lo, utiliza-se a seguinte fórmula:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Nesse caso, o somatório dos quadrados das diferenças entre os valores da amostra e a média é igual a 36.537,26. Com $n = 60$, temos:

$$s^2 = \frac{36.537,26}{59} \approx 619,27559$$

$$s = \sqrt{619,27559} \approx 24,885 \text{ ms}$$

Erro padrão da média

O erro padrão da média (EP) estima a variabilidade da média amostral em relação à média populacional. É calculado como:

$$EP = \frac{s}{\sqrt{n}} = \frac{24,885}{\sqrt{60}} \approx \frac{24,885}{7,746} \approx 3,21267$$

Valor crítico da distribuição t

Como a variância populacional é desconhecida e $n = 60$ é relativamente pequeno, utiliza-se a distribuição t de Student. Para um nível de confiança de 99%, com 59 graus de liberdade ($df = n - 1 = 59$), o valor crítico é:

$$t_{0,005; 59} \approx 2,66$$

Margem de erro

A margem de erro é obtida multiplicando-se o valor crítico da distribuição t pelo erro padrão:

$$ME = t \cdot EP = 2,66 \cdot 3,21267 \approx 8,5457$$

Intervalo de confiança de 99%

Com a média amostral e a margem de erro, constrói-se o intervalo de confiança da média populacional:

$$IC = \bar{x} \pm ME = 198,901 \pm 8,5457$$

$$IC = [198,901 - 8,5457 ; 198,901 + 8,5457] = [190,355 ; 207,4467]$$

Resultados finais

- Estimativa pontual da média: 198,901 ms
 - Intervalo de confiança de 99%: [190,355 ms ; 207,4467 ms]
- 2 Considerando a amostra apresentada na Tabela 1, estime quantas medidas a mais são necessárias para estimar o intervalo de confiança da média se exigirmos um grau de confiança de 99% e um erro absoluto de 5 ms.

Cálculo do tamanho da amostra necessário para um erro máximo de 5 ms com 99% de confiança

Deseja-se determinar o tamanho mínimo da amostra (n) necessário para que a média estimada tenha um erro máximo absoluto (E) de 5 ms, com um nível de confiança de 99%. O desvio padrão da amostra já coletada é conhecido: $s = 24,885$ ms.

Fórmula utilizada

A fórmula para determinar o tamanho da amostra, com base no erro máximo permitido, é dada por:

$$n = \left(\frac{z \cdot s}{E} \right)^2$$

Onde:

- n = tamanho mínimo da amostra
- z = valor crítico da normal padrão para o nível de confiança desejado
- s = desvio padrão estimado da amostra
- E = erro máximo admissível (margem de erro)

Escolha do valor crítico z

Como se está estimando o tamanho amostral e espera-se um valor suficientemente grande de n , considera-se a aproximação pela **distribuição normal**. Para um nível de confiança de 99% em uma estimativa bilateral, o valor crítico é:

$$z_{0,005} = 2,576$$

Substituindo os valores na fórmula

$$n = \left(\frac{2,576 \cdot 24,885}{5} \right)^2 = \left(\frac{64,1037}{5} \right)^2 = (12,820752)^2 \approx 164,372$$

Arredondamento e conclusão

Como não se pode coletar uma fração de uma observação, o valor é arredondado para o inteiro imediatamente superior:

$$n = 165$$

Já há 60 amostras previamente coletadas. Portanto, para atingir o tamanho amostral necessário, são requeridas:

$$165 - 60 = 105 \text{ amostras adicionais}$$

Resultado final

- Tamanho mínimo necessário da amostra: 165
- Número de amostras adicionais a serem coletadas: 105

3 Usando a amostra retratada na Tabela 1, estime o intervalo de confiança do desvio padrão da população, levando em conta um grau de confiança de 95%. Explique por que você adotou esse método específico.

Intervalo de confiança para o desvio padrão populacional

Deseja-se calcular o intervalo de confiança para o **desvio padrão populacional** (σ), com base em uma amostra de tamanho $n = 60$ e desvio padrão amostral $s = 24,885$ ms. Considera-se que a variável segue uma distribuição aproximadamente normal.

Distribuição e parâmetros utilizados

Como a distribuição da variável é normal, e deseja-se estimar a **variância populacional** (σ^2), utiliza-se a **distribuição qui-quadrado** (χ^2), conforme a teoria apresentada nos slides.

A fórmula do intervalo de confiança para a variância populacional é:

$$\frac{(n - 1) \cdot s^2}{\chi_{1-\alpha/2}^2} < \sigma^2 < \frac{(n - 1) \cdot s^2}{\chi_{\alpha/2}^2}$$

Extraindo o intervalo para o desvio padrão, basta aplicar a raiz quadrada em cada extremidade:

$$\sqrt{\frac{(n - 1) \cdot s^2}{\chi_{1-\alpha/2}^2}} < \sigma < \sqrt{\frac{(n - 1) \cdot s^2}{\chi_{\alpha/2}^2}}$$

Para um nível de confiança de 95% ($\alpha = 0,05$), com $n = 60$ (logo, $df = 59$), os valores críticos da distribuição qui-quadrado são:

$$\chi_{0,975}^2 \approx 82,12 \quad \text{e} \quad \chi_{0,025}^2 \approx 39,66$$

Cálculo dos limites do intervalo

Com $s^2 = 619,275$ e $n - 1 = 59$, calcula-se o limite inferior:

$$\sigma_{\text{inferior}} = \sqrt{\frac{59 \cdot 619,275}{82,12}} \approx \sqrt{444,74} \approx 21,09 \text{ ms}$$

E o limite superior:

$$\sigma_{\text{superior}} = \sqrt{\frac{59 \cdot 619,275}{39,66}} \approx \sqrt{919,91} \approx 30,35 \text{ ms}$$

Justificativa do método

A utilização da distribuição qui-quadrado para construir esse intervalo de confiança é justificada porque:

- A variável (tempo de resposta) segue distribuição aproximadamente normal;
- O desvio padrão populacional (σ) é desconhecido;
- Deseja-se estimar a variabilidade da população com base na amostra.

Resultado final

- **Intervalo de confiança de 95% para σ :** [21,09 ms ; 30,35 ms]

4 Compare o desempenho de dois servidores (s1 e s2) em relação ao tempo de execução de um programa. O programa foi executado 30 vezes em s1 e 50 vezes em s2. Os tempos de execução são mostrados na Tabela 2. É necessário descobrir se os servidores têm desempenho equivalente. Adote 95% de grau de confiança.

Server s1 (ms)			Server s2 (ms)					
117.77	142.769	115.59	139.809	165.178	151.665	160.641	137.992	
133.565	128.848	146.908	152.971	146.88	168.352	149.04	164.09	
153.343	132.099	112.778	169.453	145.002	150.236	151.749	135.649	
112.393	135.215	133.737	135.328	143.022	172.957	154.346	153.114	
142.917	130.603	124.954	160.764	160.614	138.638	150.502	145.795	
137.151	131.024	128.206	155.959	153.015	135.985	150.853	148.505	
133.824	135.938	133.412	153.187	158.808	146.061	154.948	152.843	
143.192	119.181	119.593	160.994	145.952	149.612	140.984	141.328	
123.841	118.835	153.896	144.867	141.649	136.993	140.696	169.914	
107.698	137.072	134.679	131.415	138.591	138.343	155.893	153.899	

Tabela 2: Tempos de execução

Objetivo da análise

Deseja-se verificar se há diferença significativa entre o desempenho (tempo de execução) de dois servidores distintos, com base em duas amostras independentes de tempos coletados.

Informações fornecidas

- Servidor 1:
 - $n_1 = 30$
 - $\bar{x}_1 = 130,701$ ms
 - $s_1 = 11,7655$
- Servidor 2:
 - $n_2 = 50$
 - $\bar{x}_2 = 150,1016$ ms
 - $s_2 = 10,1416$
- Nível de confiança: 95%
- Teste: bilateral
- Hipóteses:

- H_0 : As médias dos servidores são iguais
- H_1 : As médias dos servidores são diferentes

Método de teste

Como temos duas amostras independentes com variâncias diferentes (os desvios padrão diferem), utilizamos o **teste t de Student para duas amostras independentes com variâncias desiguais** (também conhecido como Welch's t-test).

A estatística de teste é dada por:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Substituindo os valores:

$$\begin{aligned} t &= \frac{130,701 - 150,1016}{\sqrt{\frac{11,7655^2}{30} + \frac{10,1416^2}{50}}} \\ t &= \frac{-19,4006}{\sqrt{\frac{138,4269}{30} + \frac{102,852}{50}}} = \frac{-19,4006}{\sqrt{4,61423 + 2,05704}} = \frac{-19,4006}{\sqrt{6,67127}} \approx \frac{-19,4006}{2,5828} \\ t &\approx -7,511 \end{aligned}$$

Graus de liberdade (aproximados)

Como as variâncias são diferentes, utilizamos a fórmula de aproximação de Welch para os graus de liberdade:

$$\begin{aligned} df &\approx \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}} \\ df &\approx \frac{(4,61423 + 2,05704)^2}{\frac{(4,61423)^2}{29} + \frac{(2,05704)^2}{49}} = \frac{(6,67127)^2}{\frac{21,2911}{29} + \frac{4,2314}{49}} \approx \frac{44,505}{0,82045} \\ df &\approx 54,244 \approx 54 \end{aligned}$$

Decisão

Valor crítico de t para $\alpha = 0,05$ (teste bilateral) e $df = 54$:

$$t_{0,025,54} \approx \pm 2,001$$

Comparando:

$$|t_{\text{calculado}}| = 7,511 > 2,001$$

Portanto, rejeitamos a hipótese nula H_0 .

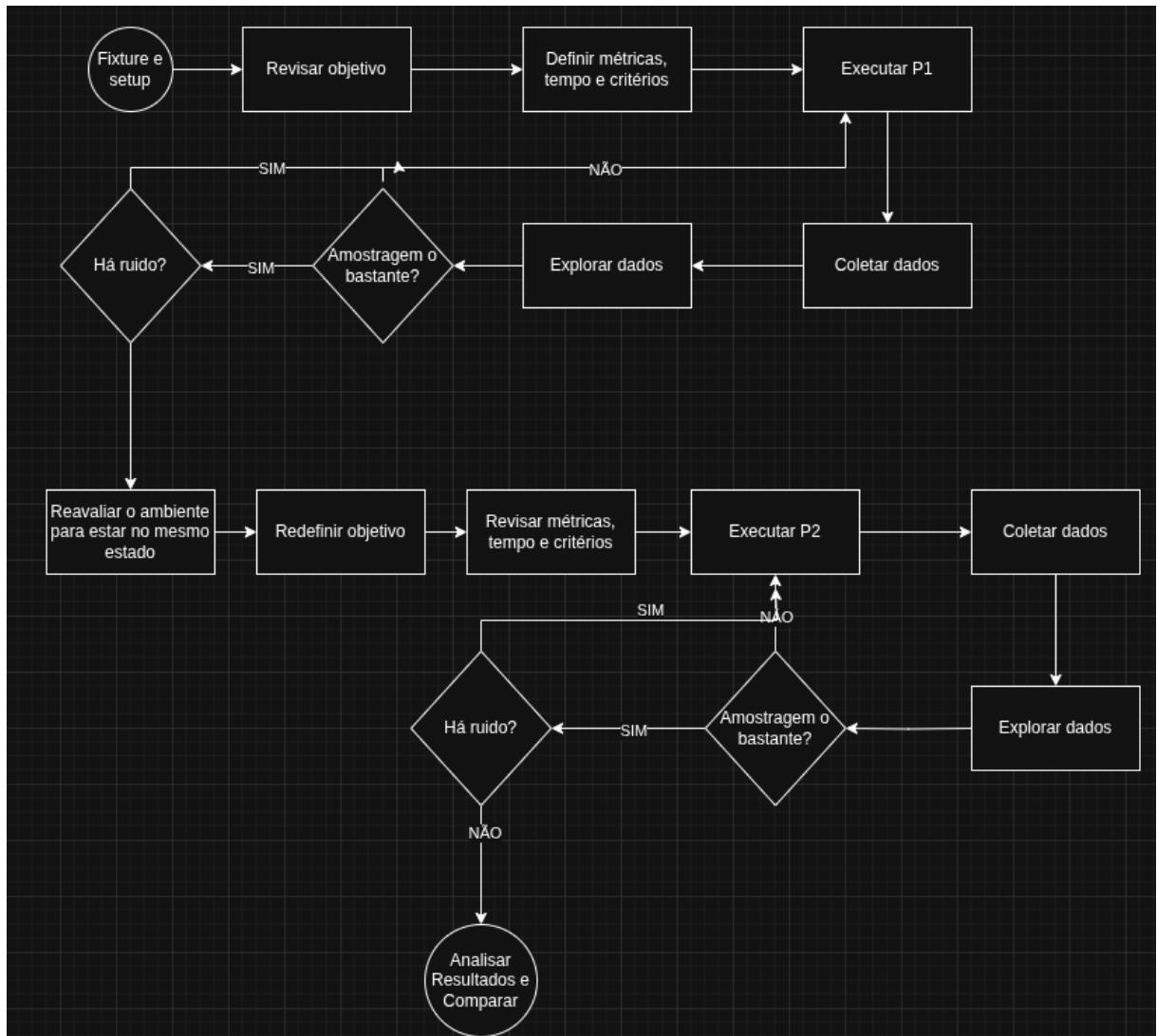
Conclusão

Com nível de confiança de 95%, há evidências estatísticas suficientes para concluir que os desempenhos dos servidores são significativamente diferentes. Considerando que a média do tempo de execução do servidor 1 (130,701 ms) é consideravelmente menor que a do servidor 2 (150,1016 ms), infere-se que o servidor 1 apresenta desempenho superior em termos de tempo de resposta.

5 Elabore uma metodologia para avaliar e comparar o desempenho de dois computadores pessoais (computadores à sua disposição), considerando como carga de trabalho quatro programas. A escolha de cada programa é sua responsabilidade. Defina as etapas da metodologia por meio de um diagrama de atividades e um documento que descreva as respectivas atividades, as ferramentas necessárias para executar cada atividade, as entradas, os produtos e as condições que sinalizam a conclusão de cada uma das atividades do processo. Adote o tempo de execução dos programas como a métrica primária para comparar o desempenho do computador. Cada um dos programas deve explorar um aspecto do sistema computacional:

- Programa 1: operações com inteiros
- Programa 2: operações com ponto flutuante
- Programa 3: operações com matrizes
- Programa 4: operações gráficas

Expandiria-se tal processo para os quatro distintos programas.



7. Uma empresa de software está examinando a eficácia de duas linguagens de programação diferentes para melhorar o desempenho da programação. Quinze de seus programadores especialistas, familiarizados com ambas as linguagens, foram solicitados a codificar uma função específica em ambas as linguagens. Os tempos necessários para implementar a função em cada linguagem de programação foram registrados (em minutos). A Tabela 3 mostra o tempo respectivo que cada programador gastou implementando a função. Há evidências para apoiar uma programação de melhor desempenho em qualquer uma das linguagens?

Objetivo

Verificar se há diferença significativa no tempo de execução entre duas linguagens de programação utilizando um teste estatístico apropriado para amostras emparelhadas.

Programme r	T PL1 (min)	T PL2 (min)
1	14	17
2	19	19
3	26	26
4	12	13
5	22	22
6	20	5
7	19	27
8	22	21
9	16	14
10	11	16
11	13	24
12	24	15
13	12	17
14	19	23
15	16	21

Tabela 3: Tempo para Programar

Características do problema

- As medições foram feitas sobre os mesmos programas implementados em duas linguagens distintas.
- As amostras são emparelhadas, pois comparam duas linguagens sobre o mesmo conjunto de tarefas.
- A análise será realizada com **teste t pareado** (também chamado de teste t para amostras dependentes).
- Nível de significância: $\alpha = 0,05$
- Tipo de teste: bilateral

Hipóteses

- H_0 : Não há diferença significativa entre os tempos de execução das linguagens (média das diferenças = 0)
- H_1 : Existe diferença significativa entre os tempos de execução das linguagens (média das diferenças $\neq 0$)

Dados

Diferenças entre os tempos de execução das linguagens:

$$-3, 0, 0, -1, 0, 15, -8, 1, 2, -5, -11, 9, -5, -4, -5$$

Quantidade de pares de observações: $n = 15$

Média das diferenças

Soma das diferenças: -15

$$\bar{d} = \frac{-15}{15} = -1$$

Cálculo do desvio padrão das diferenças

Subtrai-se cada diferença da média e eleva-se ao quadrado:

Diferenças em relação à média: $-2, 1, 1, 0, 1, 16, -7, 2, 3, -4, -10, 10, -4, -3, -4$

Quadrados: $4, 1, 1, 0, 1, 256, 49, 4, 9, 16, 100, 100, 16, 9, 16$

Soma dos quadrados: 582

$$s_d = \sqrt{\frac{\sum(d_i - \bar{d})^2}{n - 1}} = \sqrt{\frac{582}{14}} = \sqrt{41,5714} \approx 6,4475$$

Estatística do teste

A estatística t para dados pareados é:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} = \frac{-1}{6,4475/\sqrt{15}} \approx \frac{-1}{1,6647} \approx -0,6$$

Valor crítico

Para um teste bilateral com $\alpha = 0,05$ e $n - 1 = 14$ graus de liberdade, o valor crítico é:

$$t_{0,025,14} \approx 2,145$$

Decisão

$$|t| = 0,6 < 2,145 \Rightarrow \text{Não se rejeita } H_0$$

Intervalo de confiança para a média das diferenças

$$IC = \bar{d} \pm t \cdot \frac{s_d}{\sqrt{n}} = -1 \pm 2,145 \cdot \frac{6,4475}{\sqrt{15}} = -1 \pm 2,145 \cdot 1,6647$$

$$IC = -1 \pm 3,570 \Rightarrow \text{Intervalo: } [-4,57, 2,57]$$

Como o valor zero está contido no intervalo de confiança, não há evidência estatística para afirmar que há diferença entre as médias.

Conclusão

Com nível de confiança de 95%, os dados não fornecem evidências estatísticas suficientes para afirmar que existe diferença significativa no tempo de execução entre as duas linguagens. O intervalo de confiança inclui o valor zero, reforçando a conclusão de que a média das diferenças pode ser nula. Assim, não se pode afirmar que uma linguagem é mais rápida que a outra.