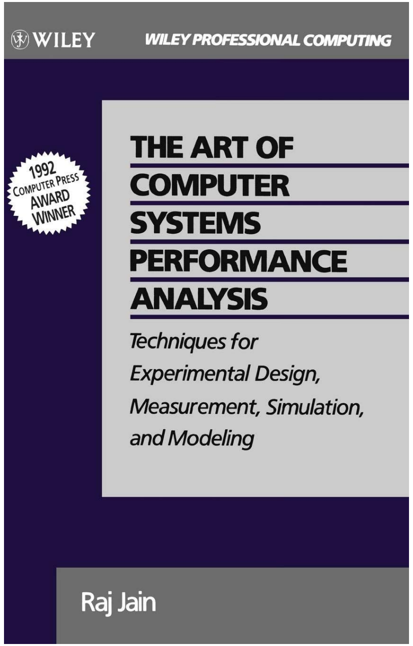
Avaliação de Desempenho de Sistemas Computacionais - ADS

Prof. Paulo Antonio Leal Rego
pauloalr@ufc.br

Universidade Federal do Ceará

Agenda

1. Resumindo os dados medidos
2. Amostra x População
3. Testes de hipóteses



Capítulos 12 e 13

Resumindo os dados medidos

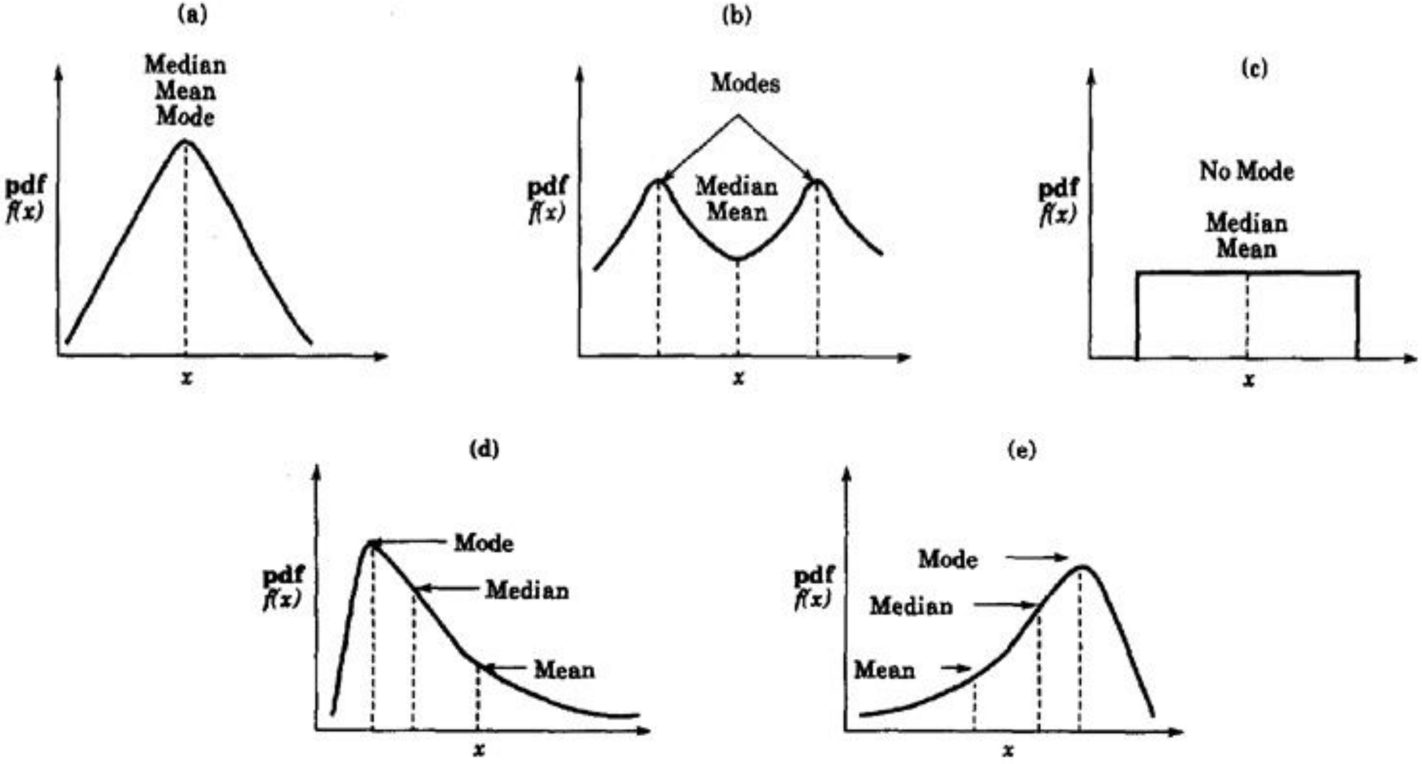
Medidas de centralidade

- Um único número pode ser utilizado para caracterizar um conjunto de dados (amostra).
- Geralmente a média é utilizada.
  - Para ser significativa, a média deve ser representativa da maior parte do conjunto de dados.
- Três alternativas populares para resumir uma amostra são especificar sua **média**, **mediana** ou **moda**.
  - Essas medidas são o que os estatísticos chamam medidas de tendência central/medidas de centralidade.

Medidas de centralidade

- A **média** da amostra é obtida tomando a soma de todas as observações e dividindo essa soma pelo número de observações na amostra.
- A **mediana** é obtida ordenando as observações em ordem crescente e tomando a observação que está no meio.
  - Se o número de observações for par, a média dos dois valores do meio é usada como mediana.
- A **moda** é obtida traçando um histograma e especificando o ponto médio do intervalo onde o histograma atinge o pico.
  - Em variáveis categóricas, a moda é fornecido pela categoria com mais frequência.

Medidas de centralidade



Medidas de centralidade

- O principal problema com a média é que ela é mais afetada por outliers do que a mediana e a moda.
  - Um único outlier pode causar uma mudança considerável na média. Isso é particularmente verdadeiro para pequenas amostras.
  - A mediana e a moda são resistentes a variações nas observações.

```
Amostra: [2, 5, 7, 10, 9, 11, 10, 10, 10, 200]
Média = 27
Mediana = Moda = 10
```

Medidas de centralidade

- A média atribui peso igual a cada observação e, nesse sentido, faz pleno uso da amostra. A mediana e a moda ignoram muitas informações.
- A média tem uma propriedade de aditividade ou linearidade em que a média de uma soma é uma soma das médias.
  - Isso não se aplica à moda ou mediana.

Medidas de dispersão/variabilidade

- A **amplitude** de uma amostra é a diferença entre o valor máximo e mínimo das observações.
- A **variância** da amostra é uma medida de dispersão que representa o quadrado da distância entre as observações e a média da amostra.

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

- O **desvio padrão** é a raiz quadrada da variância.

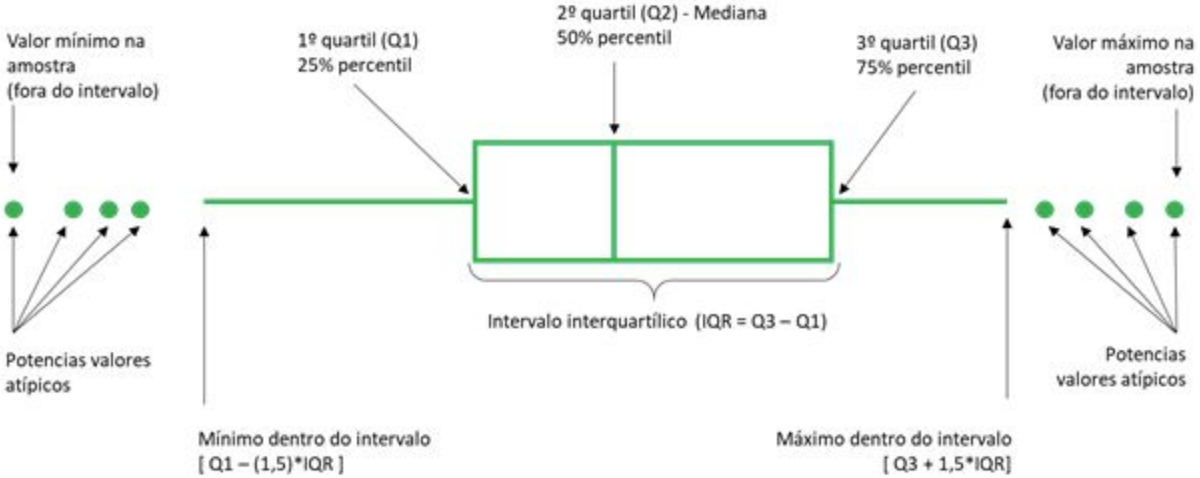
$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- O **coeficiente de variação** é definido como a razão do desvio padrão da amostra pela média amostral.

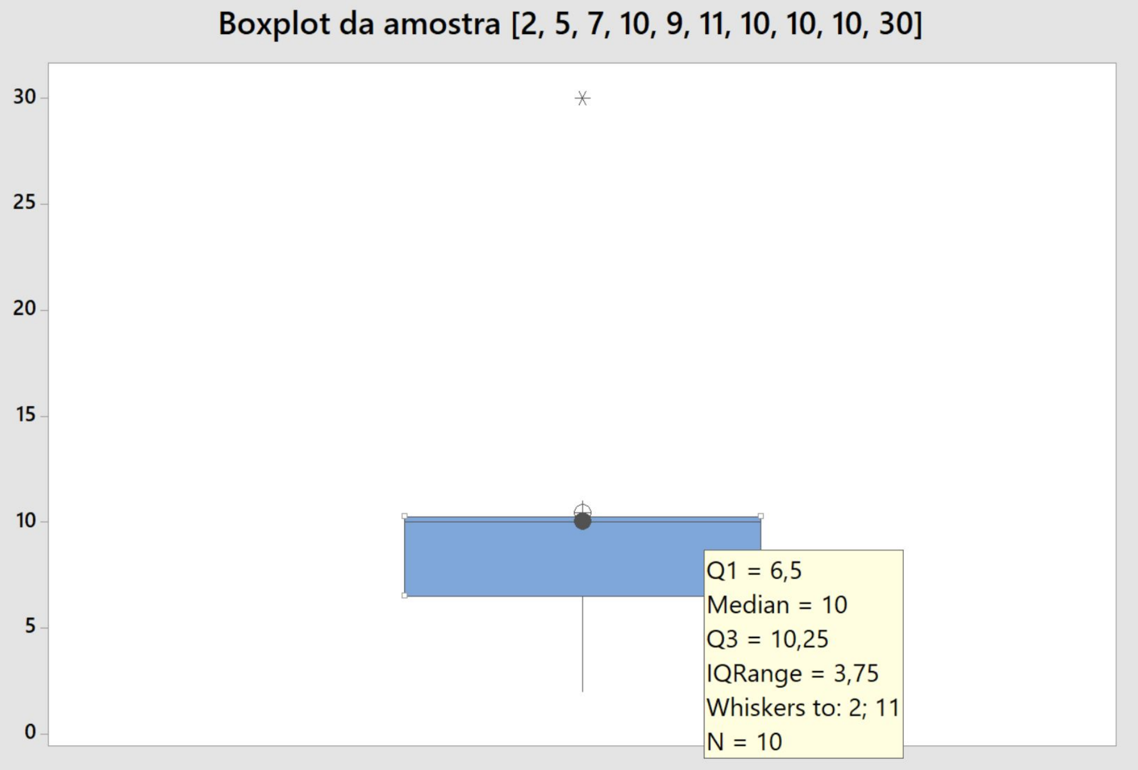
Medidas não centrais

- Quartis (Q1, Q2 e Q3) são valores dados a partir do conjunto de observações ordenado em ordem crescente, que dividem a amostra em quatro partes iguais.
- O **primeiro quartil (Q1)** é o número que deixa 25% das observações abaixo e 75% acima
- O **terceiro quartil (Q3)** deixa 75% das observações abaixo e 25% acima.
- O **segundo quartil (Q2)** é a mediana e deixa 50% das observações abaixo e 50% das observações acima.

Boxplot

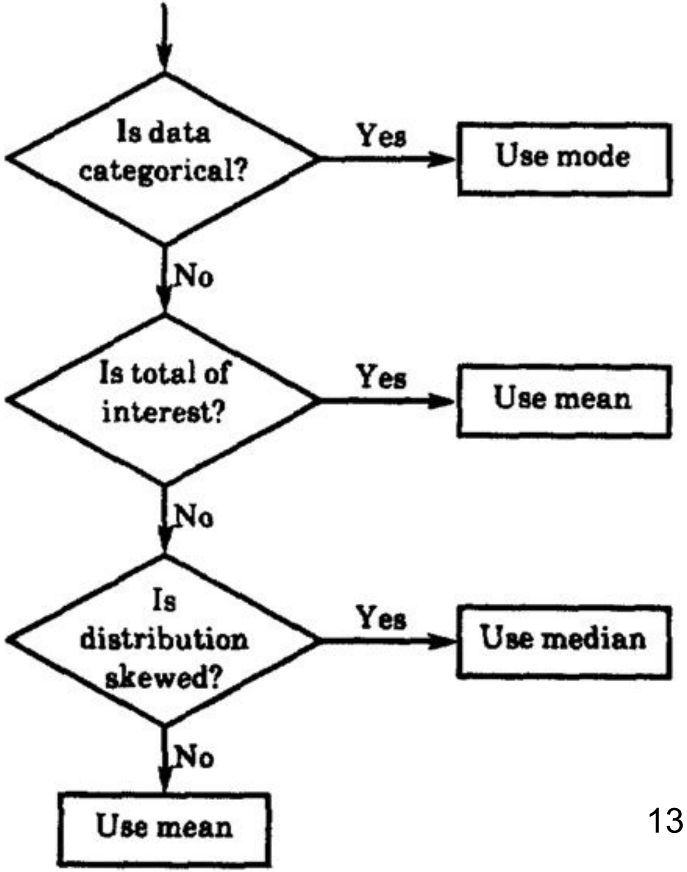

Valor mínimo na amostra (fora do intervalo)
1º quartil (Q1) 25% percentil
2º quartil (Q2) - Mediana 50% percentil
3º quartil (Q3) 75% percentil
Valor máximo na amostra (fora do intervalo)
Intervalo interquartílico (IQR = Q3 – Q1)
Potencias valores atípicos
Potencias valores atípicos
Mínimo dentro do intervalo [Q1 – (1,5)*IQR]
Máximo dentro do intervalo [Q3 + 1,5*IQR]

Boxplot


Boxplot da amostra [2, 5, 7, 10, 9, 11, 10, 10, 10, 30]
30
25
20
15
10
5
0
Q1 = 6,5
Median = 10
Q3 = 10,25
IQRange = 3,75
Whiskers to: 2; 11
N = 10

Escolha da medida de centralidade

- Um erro comum que analistas inexperientes cometem é escolher a medida de centralidade errada.
  - É comum especificar a média, mesmo sem saber se é a mais adequada em uma situação particular.
  - Exemplos:
    - Recurso mais usado: moda
    - Interarrival Time: mean
    - Configuração Média: mediana



Escolha da medida de centralidade

- Uso errado da média
  - Usar a média para valores significativamente diferentes
    - Nem sempre a média de qualquer amostra será útil.
    - A utilidade depende do número de valores e da variação, não apenas do tipo da variável.
    - Por exemplo: não é muito útil dizer que o tempo médio de CPU por consulta é de 505 ms, quando as duas observações são 10 e 1000 ms.

Escolha da medida de centralidade

- Uso errado da média
  - Usar a média sem considerar a distorção da distribuição

TABLE 12.1 System Response Times for 5 Days

	System A	System B
	10	5
	9	5
	11	5
	10	4
	10	31
Sum	50	50
Mean	10	10
Typical	10	5

Escolha da medida de centralidade

- Uso errado da média
  - Multiplicar as médias para obter a média de um produto
    - A média de um produto de duas variáveis aleatórias é igual ao produto das médias apenas se as variáveis aleatórias forem independentes.
      - Ex: sistema compartilhado com média de usuários 23 e média de subprocessos por usuário 2. 2*23=46 não é a média de subprocessos no sistema. Com muitos usuários usando o sistema, ele fica mais lento e tende-se a reduzir o número de subprocessos. Ou seja, existe uma correlação entre as variáveis.

Escolha da medida de centralidade

- Uso errado da média
  - Tirar a média de uma razão com diferentes bases.

$$\text{Average}\left(\frac{a_1}{b_1}, \frac{a_2}{b_2}, \ldots, \frac{a_n}{b_n}\right) = \frac{a_1 + a_2 + \cdots + a_n}{b_1 + b_2 + \cdots + b_n}$$
$$= \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} = \frac{(1/n)\sum_{i=1}^{n} a_i}{(1/n)\sum_{i=1}^{n} b_i} = \frac{\overline{a}}{\overline{b}}$$

Measurement Duration	CPU Duration Busy (%)
1	45
1	45
1	45
1	45
100	20
sum	200%
Mean	≠ 200/5 or 40%

$$\text{Mean CPU utilization} = \frac{\text{sum of CPU busy times}}{\text{sum of measurement durations}}$$
$$= \frac{0.45 + 0.45 + 0.45 + 0.45 + 20}{1 + 1 + 1 + 1 + 100} = 21\%$$

Escolha da medida de centralidade

- Existem outras médias
  - Média geométrica $\dot{x} = \left(\prod_{i=1}^{n} x_i \right)^{1/n}$
    - usada quando o produto das observações for uma quantidade de interesse.
  - Média harmônica

$$\ddot{x} = \frac{n}{1/x_1 + 1/x_2 + \cdots + 1/x_n}$$

Protocol Layer	Performance Improvement (%)
7	18
6	13
5	11
4	8
3	10
2	28
1	5

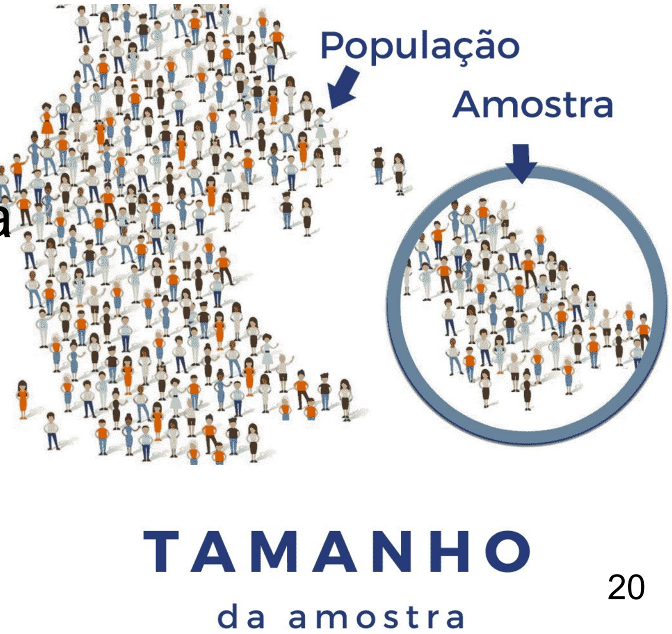
$\{(1.18)(1.13)(1.11)(1.08)(1.10)(1.28)(1.05)\}^{1/7} = 1.13$

13%

Amostra x População

Amostra x População

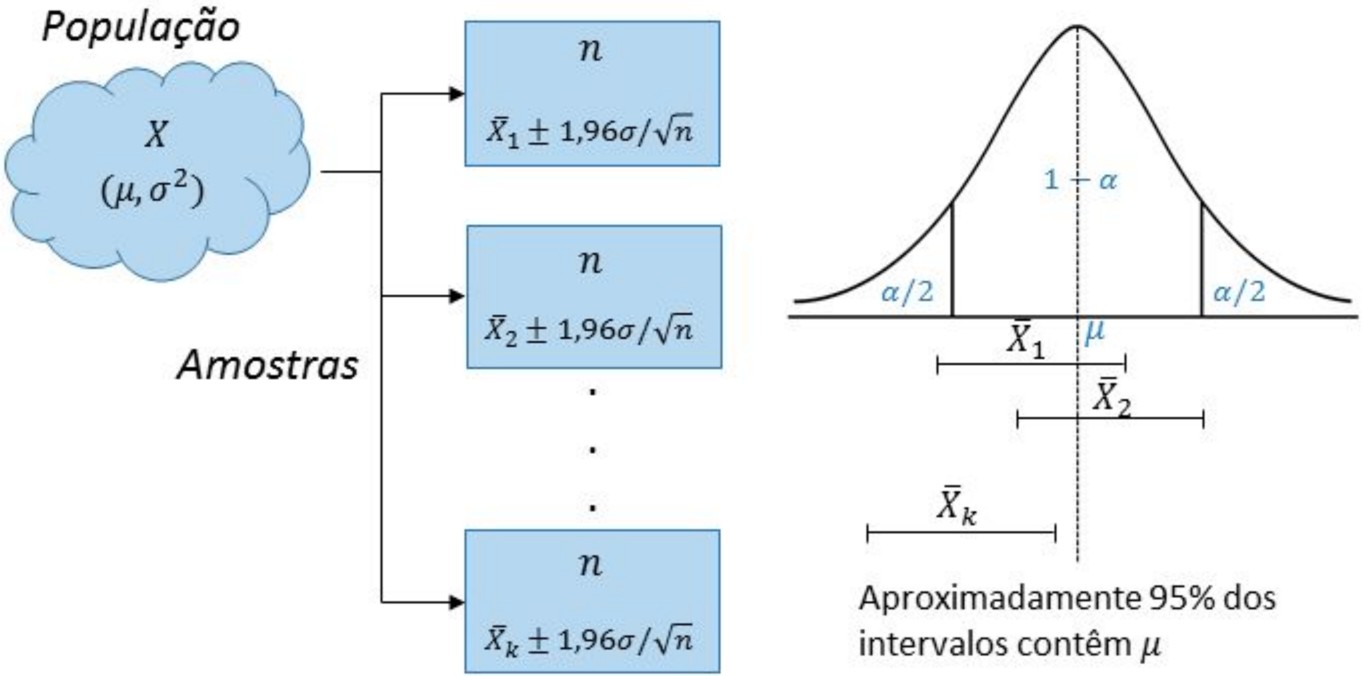
- Na maioria dos problemas do mundo real, as características da população são desconhecidas e o objetivo do analista é estimar essas características.
  - Ex: Ao medir o tempo para executar um programa. A média da amostra obtida a partir de uma única amostra de n observações é uma estimativa da média da população. É preciso repetir o experimento infinitas vezes para calcular com exatidão a média da população (o que é inviável).



Amostra x População

- Não é possível obter uma **estimativa perfeita** da média da população a partir de qualquer número finito de amostras de tamanho finito.
- O melhor que podemos fazer é obter limites probabilísticos. Assim, podemos ser capazes de obter dois limites, por exemplo, c_1 e c_2, de modo que haja uma alta probabilidade, $1 - \alpha$, de que a média da população esteja no intervalo (c_1, c_2):
    - Esse intervalo é chamado de **intervalo de confiança** para a média da população, α é o **nível de significância**, $100(1 - \alpha)$ é o **nível de confiança** e $1 - \alpha$ é **coeficiente de confiança**. Ex: Nível de confiança: 90 ou 95% ($\alpha = 0,1$ ou $0,05$).

Intervalo de confiança



Teorema Central do Limite

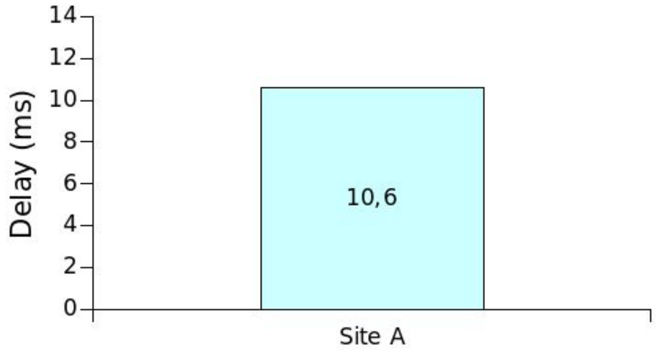
- Felizmente não é necessário coletar muitas amostras. Podemos determinar o intervalo de confiança com apenas uma amostra graças ao **teorema central do limite**, que permite determinar a distribuição da média da amostra.
  - O teorema afirma que se as observações em uma amostra {x1, x2,. . . , xn} são independentes e vêm da mesma população que tem uma média μ e um desvio padrão σ, então a média da amostra para muitas amostras se aproxima da **dist. normal** com média μ e desvio padrão $\sigma/\sqrt{n}$
    - À medida que n cresce, a distribuição da média amostral aproxima-se de uma distribuição normal com média μ.

Teorema Central do Limite

- A partir dessa expressão $\sigma/\sqrt{n}$, é fácil ver que, à medida que o tamanho da amostra n aumenta, o desvio padrão da média amostral (também chamado de erro padrão) diminui.
  - Atenção: o **erro padrão** é diferente do desvio padrão da população.
- O teorema é um importante resultado da estatística e a demonstração de outros teoremas estatísticos dependem dele.
  - Muitos testes estatísticos são realizados com a tabela de distribuição normal se o tamanho da amostra n é grande ou o desvio padrão da população é conhecido. Se o desvio σ é desconhecido e n < 30, a tabela T pode ser mais apropriada.

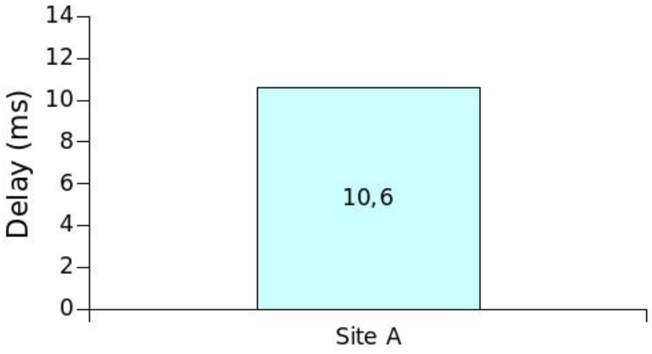
Intervalo de confiança

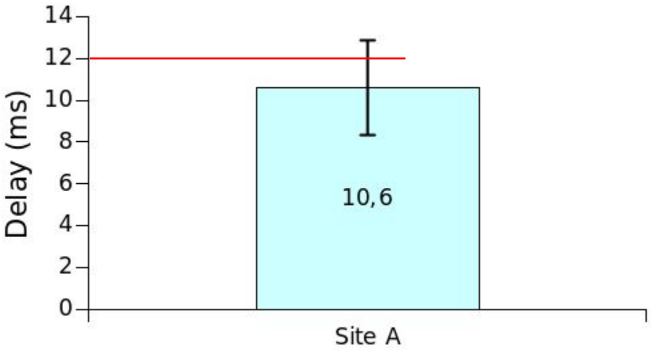
- Exemplos
  - Delay até o site A: [10, 7, 15, 2, 13, 16, 20, 7, 15, 1, 10, 7, 15, 2, 13, 16, 20, 7, 15, 1, 10, 7, 15, 2, 13, 16, 20, 7, 15, 1]
  - Delay para o site B: média 12ms
  - O delay para o site B é maior do que para o site A?



Intervalo de confiança

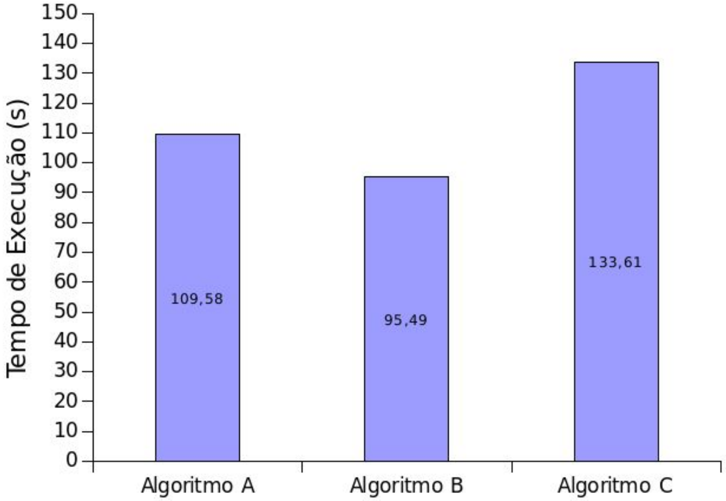
- Exemplos
  - Delay até o site A: [10, 7, 15, 2, 13, 16, 20, 7, 15, 1, 10, 7, 15, 2, 13, 16, 20, 7, 15, 1, 10, 7, 15, 2, 13, 16, 20, 7, 15, 1]
  - Delay para o site B: média 12ms
  - O delay para o site B é maior do que para o site A?





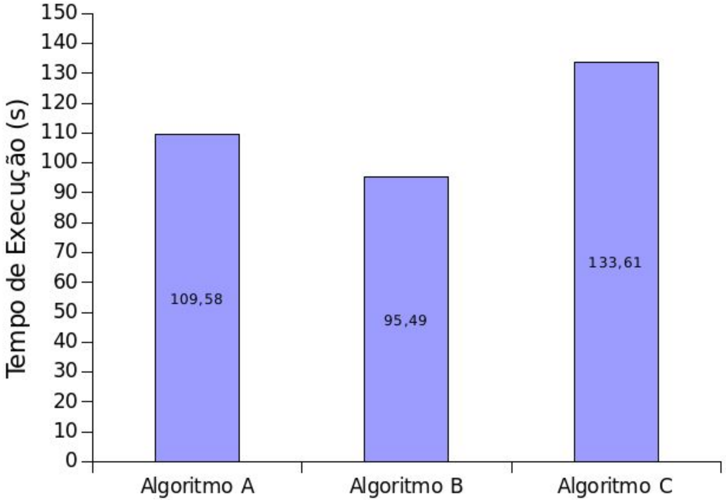
Intervalo de confiança

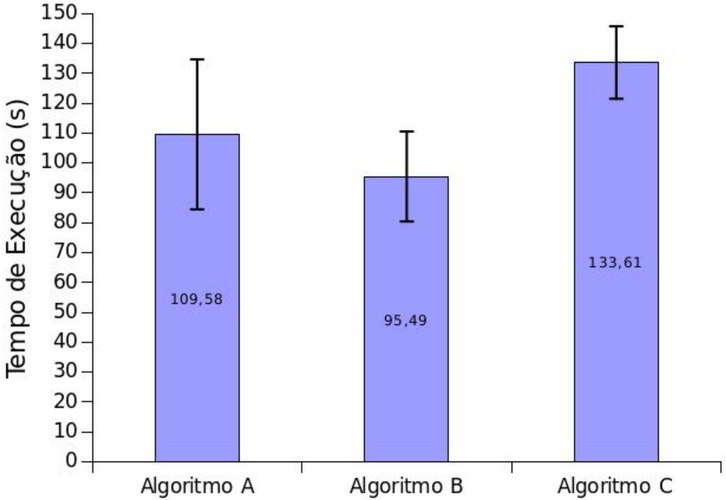
- Exemplos
  - Algoritmos A, B, C
  - Qual é melhor (executa mais rápido uma certa tarefa)?



Intervalo de confiança

- Exemplos
  - Algoritmos A, B, C
  - Qual é melhor (executa mais rápido uma certa tarefa)?





Testes de Hipóteses

Testes de Hipóteses

- É uma metodologia estatística que auxilia a tomar decisões sobre uma ou mais populações baseado nas informações obtidas nas amostras.
- Permite verificar se os dados amostrais trazem evidência que apoiem ou não uma hipótese estatística formulada.
  - Uma hipótese estatística é uma afirmação/conjectura sobre parâmetro(s) da distribuição de probabilidades de uma característica da população ou de uma variável aleatória.
  - Um teste de uma hipótese estatística é o procedimento ou regra de decisão que permite comprovar ou não uma hipótese, com base na informação contida na amostra.

Testes de Hipóteses

- H0 = hipótese nula
  - Afirmação que se quer testar. Presume-se que H0 é verdade, a não ser que existam evidências para provar o contrário.
- HA = hipótese alternativa
  - Aquilo que se pretende comprovar com a amostra (ao rejeitar a hipótese nula)
- Nível de significância: alfa (geralmente 0,05 ou 0,01).
- Exemplo:
  - H0: média = 500
  - HA: média != 500
- Resultado: p-value
  - Se p-value < alfa: rejeita H0
  - Senão: não há evidências suficientes para rejeitar H0

Testes de Hipóteses

- Teste de Normalidade
    - Tempo de Execução do Algoritmo A [139,9, 49,4, 167,7, 158,2, 30, 222,6, 164,1, 7,4, 166,4, 207,3, 193,8, 153,8, 46,6, 63,3, 29, 114,6, 39,7, 136,3, 96,1, 73,1, 157,5, 151,6, 200,7, 161,2, 83,3, 20,9, 51,2, 111,4, 128,2, 25,1]
        - H0: os dados estão normalmente distribuídos
        - HA: os dados não estão normalmente distribuídos
            - p-value = 0,066

Testes de Hipóteses

- Teste de Normalidade
    - Delay Site A [10, 7, 15, 2, 13, 16, 20, 7, 15, 1, 10, 7, 15, 2, 13, 16, 20, 7, 15, 1, 10, 7, 15, 2, 13, 16, 20, 7, 15, 1]
        - H0: os dados estão normalmente distribuídos
        - HA: os dados não estão normalmente distribuídos
            - p-value: 0,021

- Existem vários testes:
    - Anderson-Darling,
    - Kolmogorov-Smirnov,
    - Shapiro-Wilk, dentre outros...

Testes de Hipóteses

- Teste de hipóteses para a média (T test ou Z test)
  - Delay Site A [10, 7, 15, 2, 13, 16, 20, 7, 15, 1, 10, 7, 15, 2, 13, 16, 20, 7, 15, 1, 10, 7, 15, 2, 13, 16, 20, 7, 15, 1]
    - H0: média = 9
    - HA: média =! 9
      - p-value = 0,15
    - H0: média = 14
    - HA: média =! 14
      - p-value = 0,00454

Testes de Hipóteses

- Teste de hipóteses para as médias de duas amostras (T test ou Z test)
    - Amostras pareadas x não pareadas
    - Delay Site A [10, 7, 15, 2, 13, 16, 20, 7, 15, 1, 10, 7, 15, 2, 13, 16, 20, 7, 15, 1, 10, 7, 15, 2, 13, 16, 20, 7, 15, 1]
    - Delay Site B [9, 5, 10, 7, 5, 9, 5, 9, 10, 9, 6, 6, 12, 8, 9, 4, 6, 9, 6, 7, 9, 7, 13, 5, 6, 7, 12, 10, 7, 8]
        - H0: média Site A == média Site B
        - HA: média Site A != média Site B
            - p-value = 0,0247

Testes de Hipóteses

- Teste de hipóteses médias de mais de duas amostras (ANOVA)
  - Algoritmo A [139.9, 49.4, 167.7, 158.2, 30, 222.6, 164.1, 7.4, 166.4, 207.3, 193.8, 153.8, 46.6, 63.3, 29, 114.6, 39.7, 136.3, 96.1, 73.1, 157.5, 151.6, 200.7, 161.2, 83.3, 20.9, 51.2, 111.4, 128.2, 25.1]
  - Algoritmo B [70.3, 91.2, 19.5, 178.2, 129.2, 88.1, 116.2, 100.7, 93.8, 80.2, 94, 49.9, 181.6, 78.6, 100.2, 117.6, 127, 65.9, 91.7, 108.8, 80, 145.3, 130.1, 138, 62.4, 22.3, 75.1, 71, 133.2, 24.7]
  - Algoritmo C [136.3, 150.6, 150.6, 120.7, 144.8, 133.7, 129.8, 102, 157.4, 151.1, 74.9, 164, 117.9, 103.7, 55.7, 116.8, 124.1, 155.3, 137.1, 149.8, 162.6, 144.8, 135.9, 140.2, 62.4, 173.1, 161.6, 131.4, 108.6, 211.5]
    - H0: as médias de todas as amostras são iguais
    - HA: nem todas as médias são iguais
      - p-value = 0,009

Testes de Hipóteses

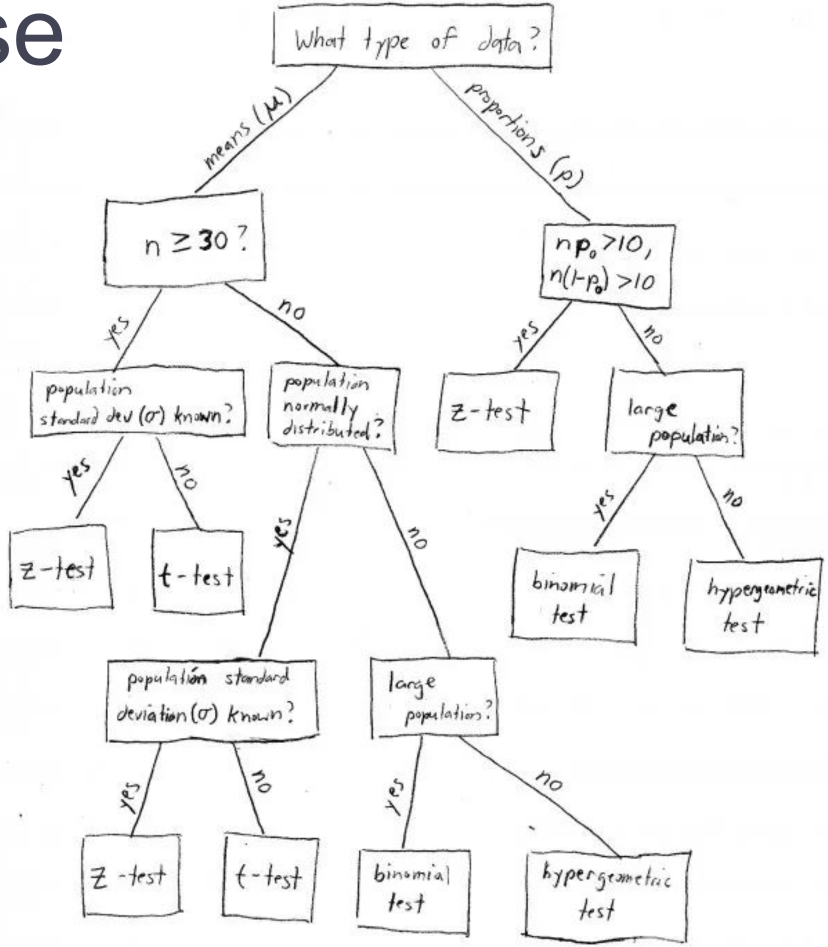
- Teste de hipóteses para uma proporção (Binomial test)
  - 300 usuários acessaram o sistema e conseguiram completar o cadastro sem erros
  - 25 usuários acessaram, mas não conseguiram completar
    - H0: 95% dos usuários conseguem completar. Ou seja, proporção = 95%
    - HA: proporção não é igual a 95%
      - p-value = 0,040

Testes de Hipóteses

- Teste de hipóteses para duas proporções (Chisquare ou Z test)
  - Computadores que usam software antivírus diferentes foram testados
  - Antivirus A: [1781 infectados, 135 não Infectados]
  - Antivirus B: [900 infectados, 89 não Infectados]
    - H0: as proporções são iguais
    - HA: as proporções são diferentes
      - p-value = 0,0724

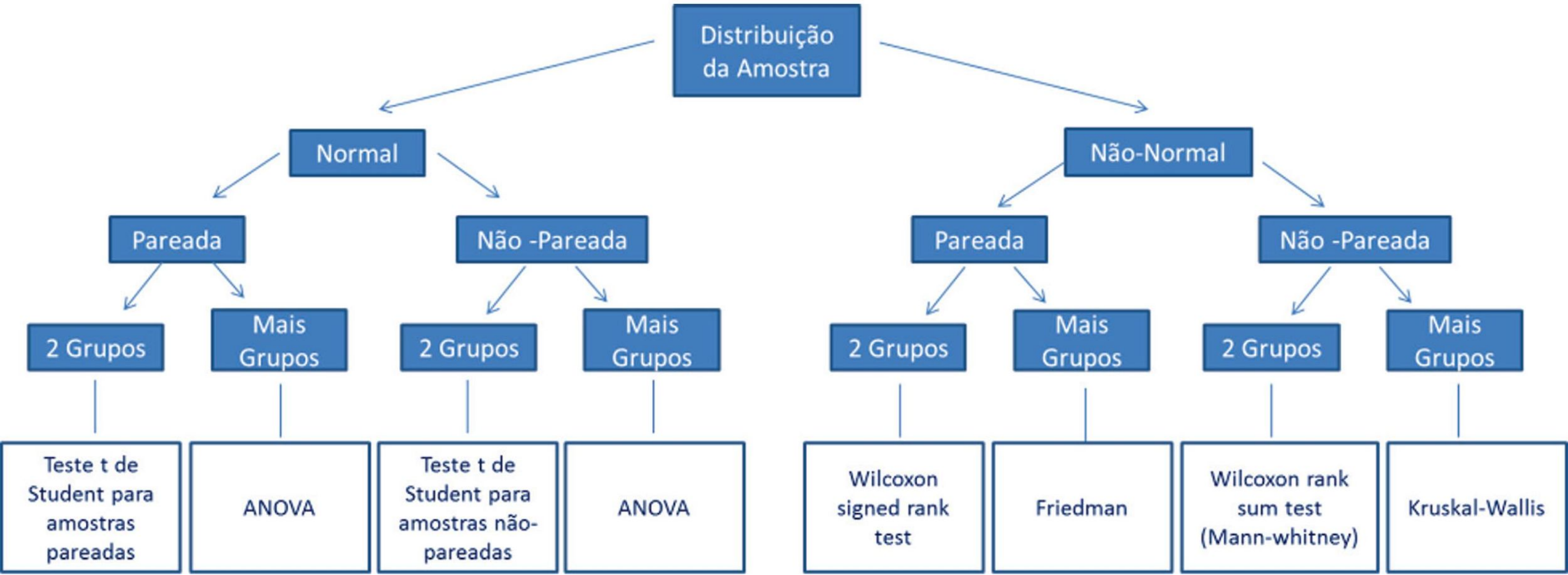
Testes de Hipótese

- Escolha dos testes



Testes de Hipóteses

- Catálogo de testes



Referências

- Bibliografia Básica:
  - JAIN, Raj. The art of computer systems perfomance analysis: techniques for experimental design, measurement, simulation, and modeling . New York, NY: John Wiley & Sons, 1991. xxvii, 685 p. ISBN 9780471503361.
  - KANT, K. Introduction to computer system performance evaluation. New York: McGraw-Hill, 1992.

Até a próxima aula!

Paulo Antonio Leal Rego
pauloalr@ufc.br

Universidade Federal do Ceará