

SSC0951

Desenvolvimento de código otimizado

Análise Estatística dos Resultados

Profa. Sarita Mazzini Bruschi

sarita@icmc.usp.br

Material baseado no conteúdo da disciplina “Avaliação de Desempenho de Sistemas Computacionais”

de Marcos José Santana e Regina H. C. Santana

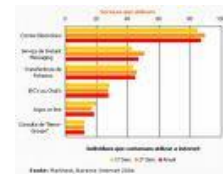
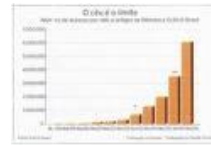
Conteúdo

- **Análise de resultado**
 - **Introdução**
 - Medidas de Desempenho Frequentemente Utilizadas
 - Análise Estatística dos Resultados
 - Comparação de Resultados
 - Procedimento para análise de resultados

Introdução

- Estatística Descritiva

X	Y	Al ₂ O ₃	SiO ₂	Fe ₂ O ₃	CaO	MgO
-70700	-129700	9.69	3.78	2.06	43.43	2.93
-70700	-129690	8.08	2.97	1.91	45.31	2.61
-70700	-129680	10.59	4.00	1.57	40.79	4.86
-70700	-129670	10.39	3.87	1.90	39.97	5.56
-70700	-129660	10.23	3.86	1.81	40.14	5.53
-70700	-129650	5.58	2.08	1.05	44.21	5.62
-70700	-129640	6.94	2.58	1.45	43.91	4.75
-70700	-129630	8.17	3.13	1.50	42.37	5.17



Exemplo 1 - Massa de 10 dados de diferentes formatos e cores. Cada dado representa um valor diferente de uma população de dados. A população de dados é formada por 10 dados de diferentes formatos e cores.

Formato	Cores	Valor
1	Amarelo	1
2	Verde	2
3	Vermelho	3
4	Azul	4
5	Roxo	5
6	Laranja	6
7	Verde-claro	7
8	Vermelho-claro	8
9	Azul-claro	9
10	Roxo-claro	10



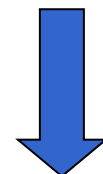
H
 μ \bar{x}
 σ^2 σ

Coleta de
Dados



Massa de
Dados

Organização
dos Dados



Dados
agrupados

Caracterização
dos Dados

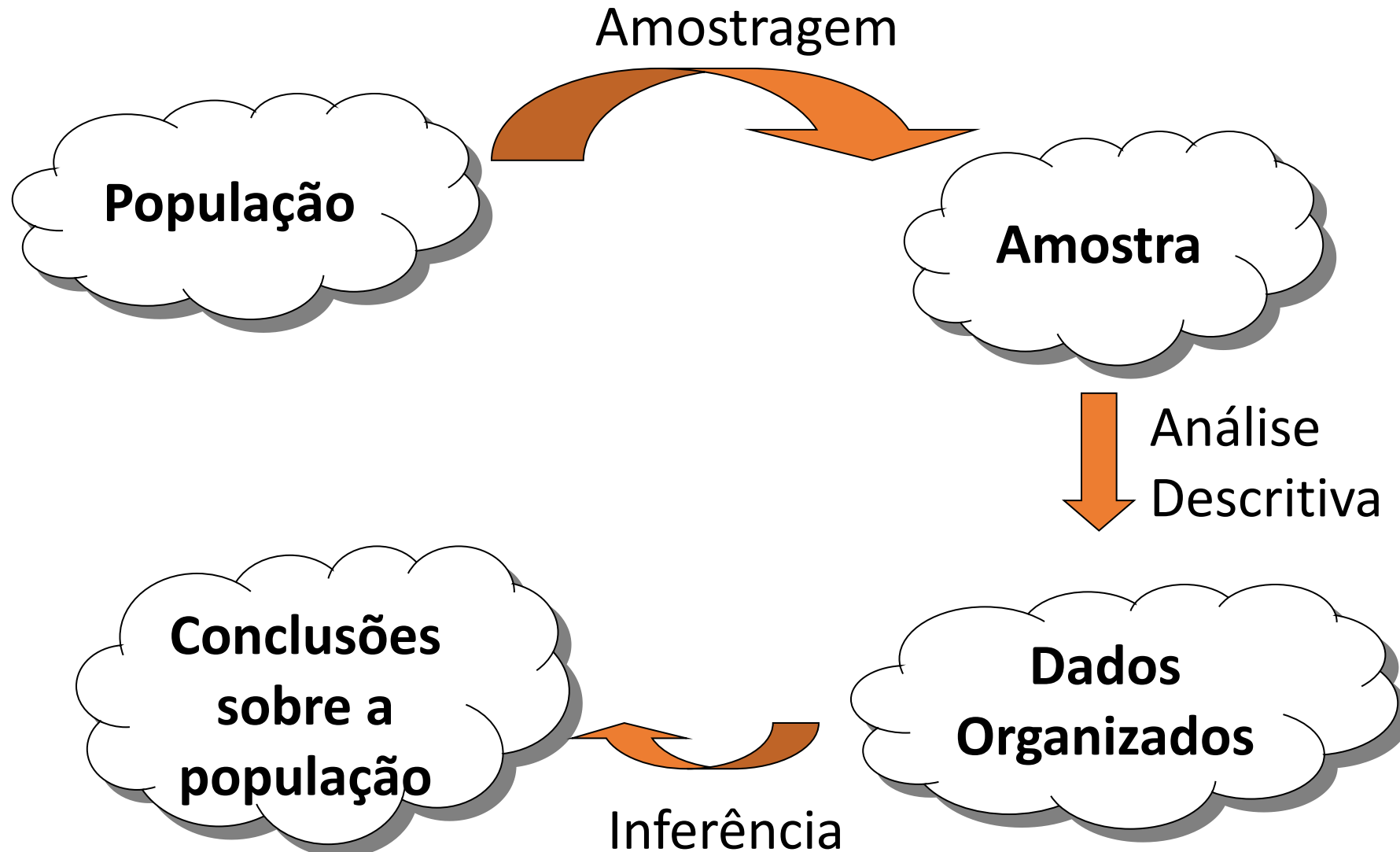


Parâmetros
Típicos

Introdução

- Inferência Estatística
 - A partir de amostras é possível tirar conclusões sobre a população
 - Erro amostral: elementos atípicos que não representam a amostra
 - Diferença entre amostra e população
 - Amostra viciada: tendência maior em selecionar algum tipo de elemento

Análise de Resultados - Introdução

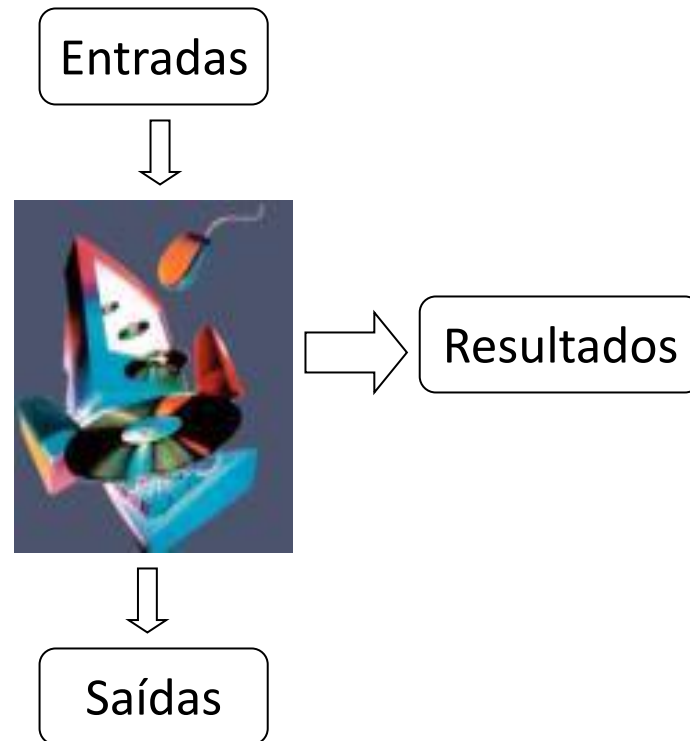


Análise de Resultados

- Procedimento **errado** normalmente utilizado para uma avaliação
 1. Desenvolvimento de um procedimento para avaliação
 - no nosso caso, seria a coleta das informações utilizando as ferramentas de monitoração
 2. Validação e verificação do sistema de avaliação
 3. Obtenção dos resultados através de **uma** execução da forma de avaliação escolhida
 4. Conclusões sobre o sistema em estudo
- **Equivalente a se considerar uma amostra unitária**

Análise de Resultados

- Por que esse Procedimento está **errado**?
- Aferição



Sistema a ser Avaliado

Análise de Resultados

- Por que esse Procedimento está **errado**?
 - Tem-se controle de todo o sistema?
 - Como são controladas as entradas do sistema?
 - Quais as condições iniciais do sistema?
 - O que mais o sistema está processando no momento da avaliação?
 - Como controlar as interrupções?
 - Diferentes características a serem consideradas, p.ex. onde estão localizadas as informações no disco?

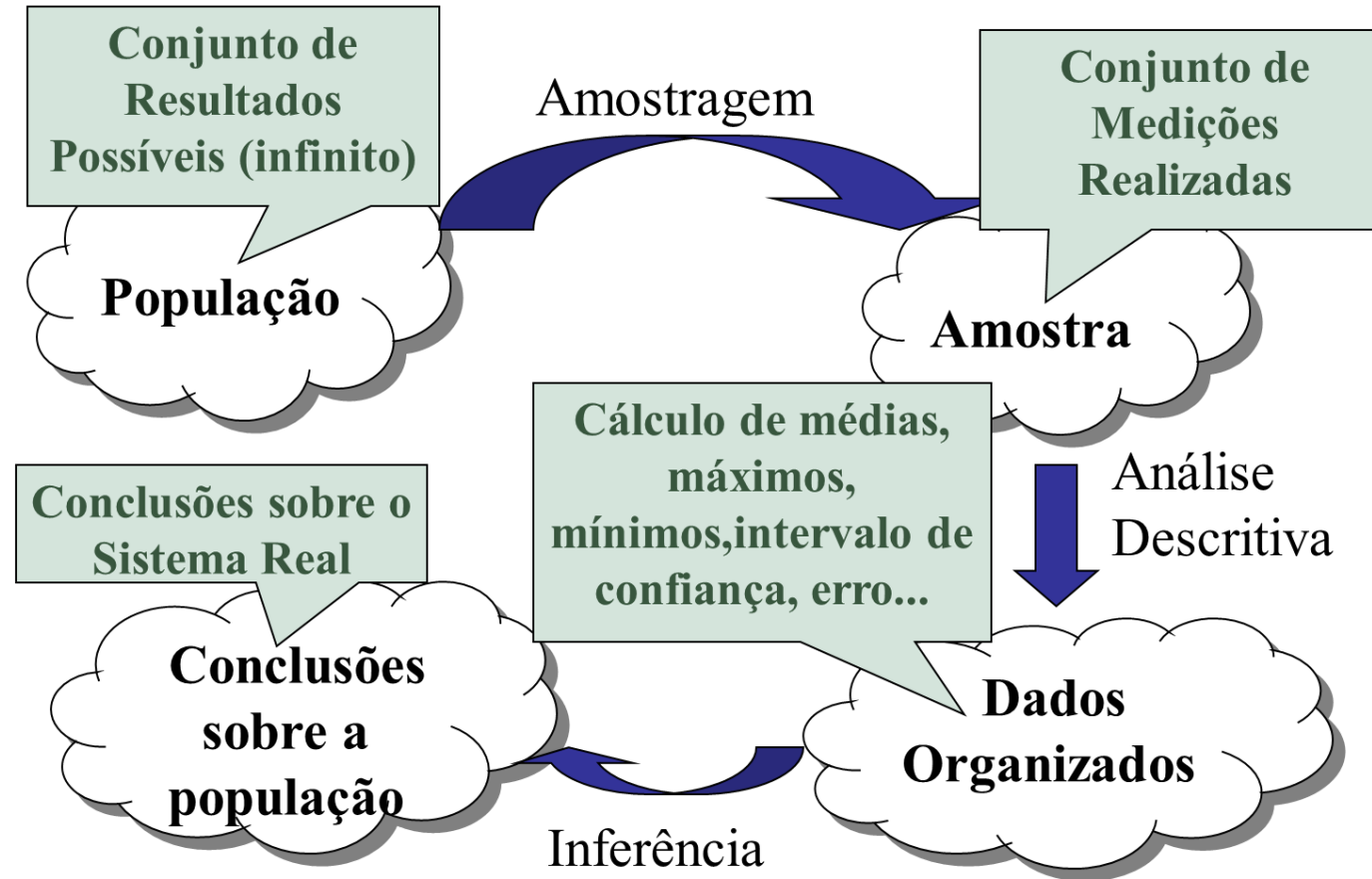
Análise de Resultados

- Por que esse Procedimento está **errado**?
 - Tempo para Execução de um processo em um Sistema Operacional
 - Uma tarefa programada pode ser ativada durante a execução
 - Interrupção do clock
 - Variável necessária está no cache? Na primeira vez que o processo executa pode não estar.
 - Quais as condições iniciais do sistema?
 - Que outros processos estão executando?

Análise de Resultados

- Portanto,
 - Em uma aferição, a medida obtida é somente **uma** dentre um conjunto de possibilidades
 - Deve-se utilizar as ferramentas que a estatística nos oferece para analisar os resultados dos experimentos.

Análise de Resultados



Análise de Resultados

- Os números são lindos!!!!
- Mas são também traiçoeiros....
- Primeiro ponto a ser considerado
 - Quais métricas estão sendo utilizadas?
 - O que representam os valores obtidos?

Conteúdo

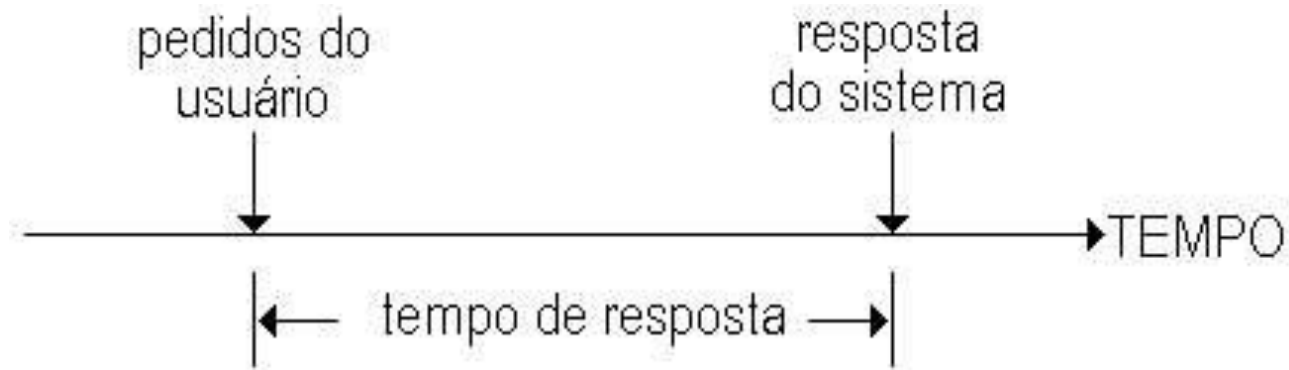
- **Análise de resultado**
 - Introdução
 - **Medidas de Desempenho Frequentemente Utilizadas**
 - Análise Estatística dos Resultados
 - Comparação de Resultados
 - Procedimento para análise de resultados

Medidas de Desempenho Frequentemente Utilizadas

- Tempo de resposta
 - Intervalo entre o pedido do usuário e a resposta do sistema
 - Pedido e resposta instantâneos
 - Pedido e resposta realistas
 - Duas definições:
 1. fim da requisição e início da resposta
 2. fim da requisição e fim da resposta
 - Geralmente o tempo de resposta cai na medida que a carga de trabalho diminui

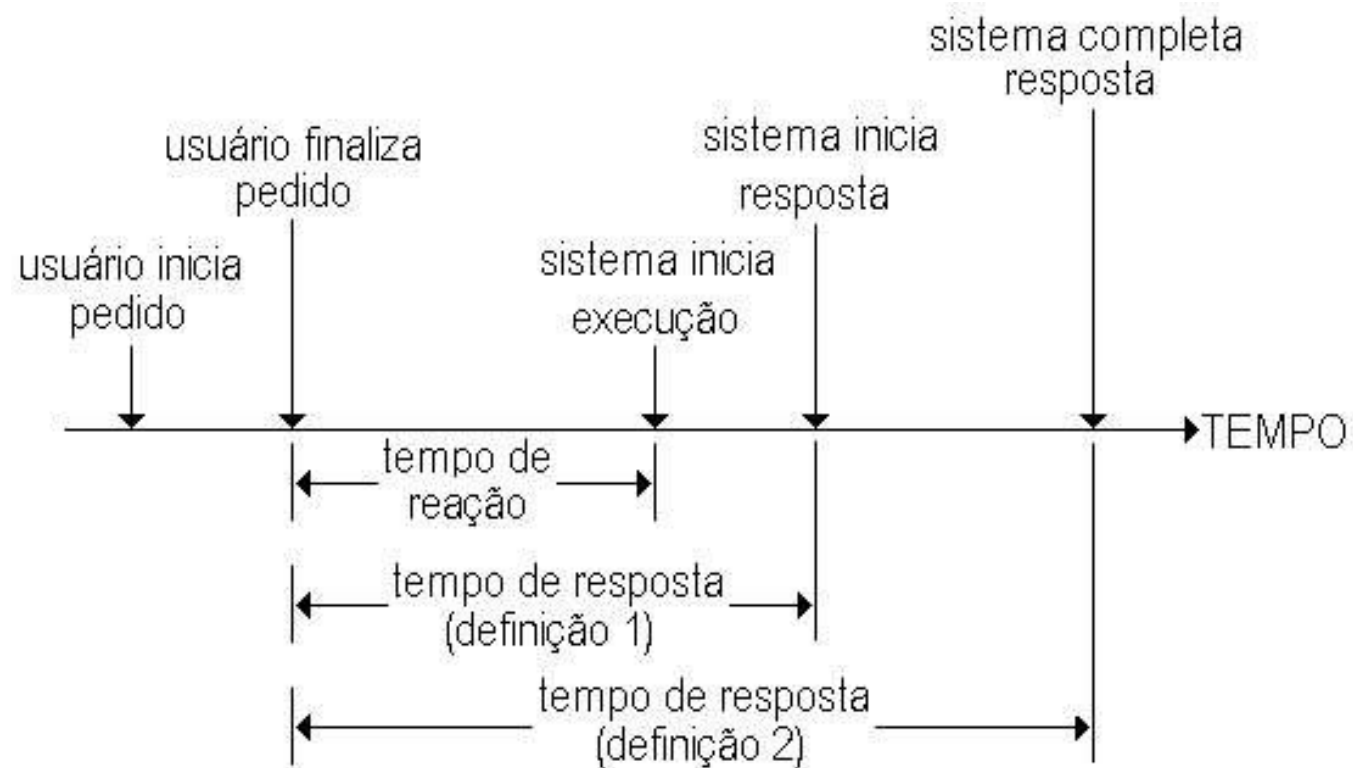
Medidas de Desempenho Frequentemente Utilizadas

- Tempo de resposta com pedido e resposta instantâneos



Medidas de Desempenho Frequentemente Utilizadas

- Tempo de resposta com pedido e resposta realistas

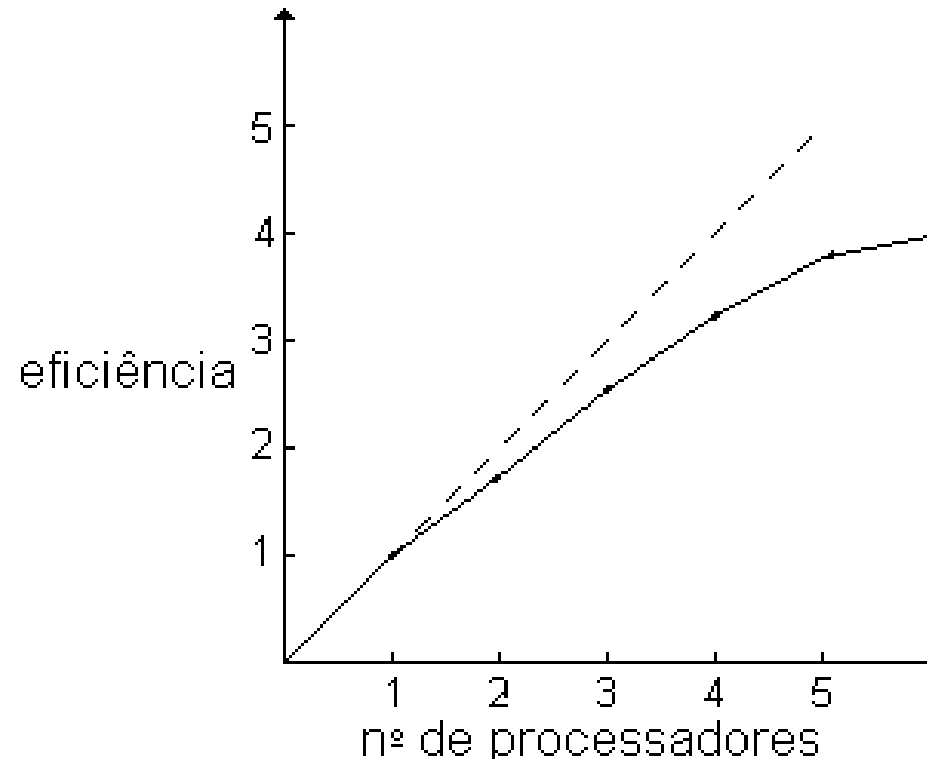


Medidas de Desempenho Frequentemente Utilizadas

- *Troughput*
 - Definido como uma taxa de pedidos que podem ser servidos pelo sistema (pedidos por unidade de tempo)
 - Capacidade nominal
 - O *troughput* geralmente aumenta até certo ponto, depois começa a cair
 - Caracteriza a eficiência

Medidas de Desempenho Frequentemente Utilizadas

- *Troughput*
 - Eficiência (exemplo)
 - Gráfico de eficiência em vários processadores



Medidas de Desempenho Frequentemente Utilizadas

- Utilização
 - Fração de tempo em que o recurso está ocupado
 - Sistema ocioso (*idle time*)
 - deve haver balanceamento de carga para que nenhum recurso seja mais utilizado que outro
- Confiabilidade
 - tempo provável em que o sistema fica livre de erros

Medidas de Desempenho Frequentemente Utilizadas

- Disponibilidade
 - Fração de tempo em que o sistema está disponível
- Custo/Desempenho
 - Melhor relação

Medidas de Desempenho Frequentemente Utilizadas

- *Speed Up* = tempo serial/tempo paralelo
- Problemas com *speed up*
 - Considere dois algoritmos, A e B

ALGORITMO	A	B
tempo em serial (s)	10	7
tempo em paralelo (s)	5	4
SPEEDUP	2	1.75

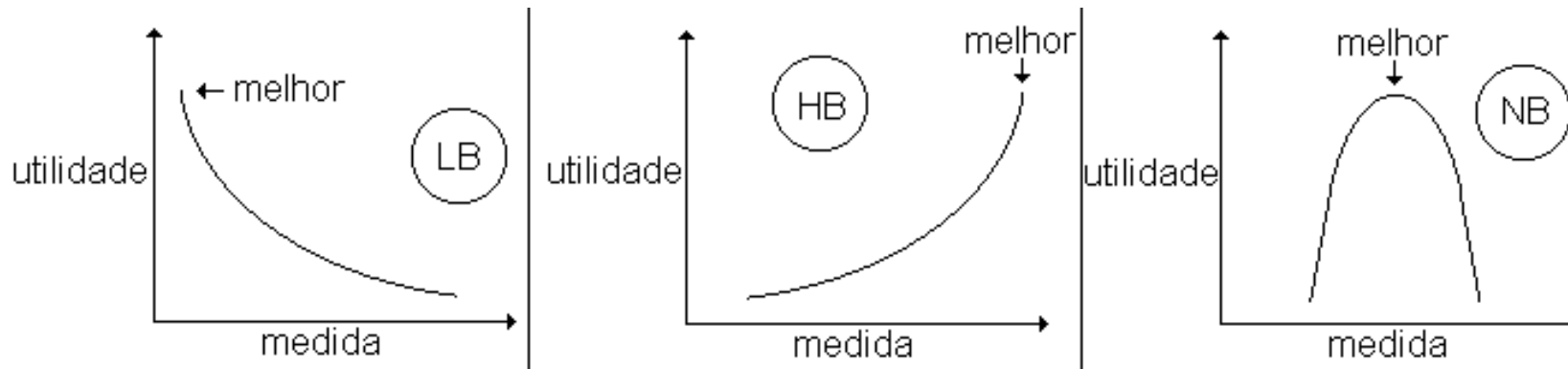
Deve-se evitar esta métrica?

Medidas de Desempenho Frequentemente Utilizadas

- MHZ - problemas
- Métricas populares:
 - MIPS, GIPS, TIPS (*M/G/T instructions per second*);
 - MFLOPS, GFLOPS, TFLOPS (*M/G/T floating operations per second*);
 - PPS (*packets per second*), BPS (*bits per second*)
 - etc.

Medidas de Desempenho Frequentemente Utilizadas

- Categorias das medidas de desempenho
 - Menor é melhor (*Low is Better - LB*);
 - Maior é melhor (*High is Better - HB*);
 - Nominal é melhor (*Nominal is Better - NB*).



Medidas de Desempenho Frequentemente Utilizadas

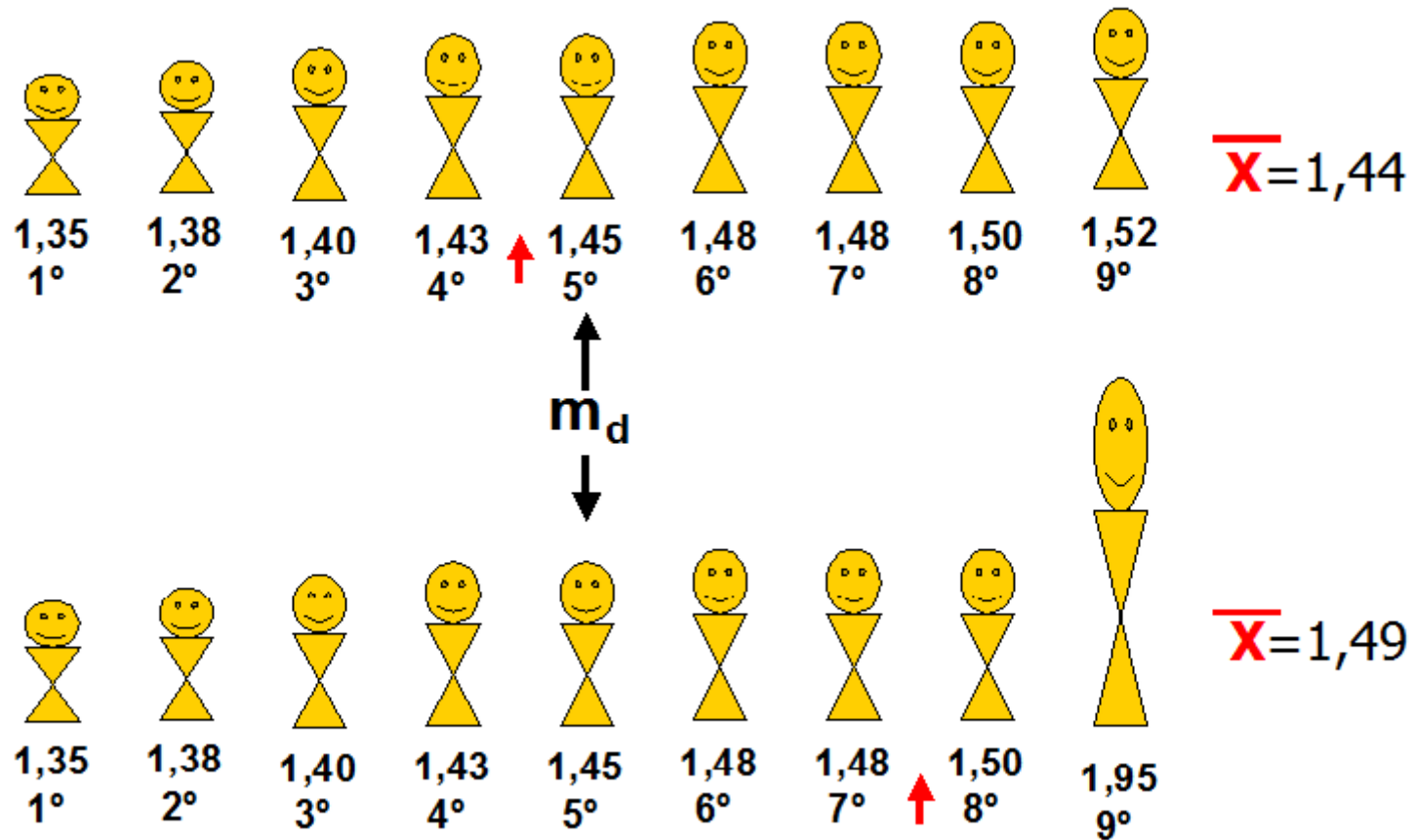
- Medidas de Posição
 - Média
 - Moda
 - Mediana
- Medidas de Dispersão
 - Desvio Padrão
 - Variância
 - Percentis
 - Box-Plot

Medidas de Posição

- Média
 - média aritmética dos valores
 - valores extremos “puxam” a média
- Moda
 - valor que ocorre mais vezes
- Mediana
 - divide o conjunto em duas partes iguais
 - em um conjunto ordenado a mediana está no centro
 - Não é influenciada por valores extremos

Medidas de Posição

- Média x Mediana



Medidas de Dispersão

- Variância

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (v_i - m)^2$$

- Desvio Padrão

$$\sigma = \sqrt{\text{variância}}$$

- Análise de percentis
- Box-Plot

Medidas de Dispersão

- Percentis – dividem os dados em cem partes de tamanhos iguais
- Quartis: dividem em quatro grupos, delimitados pelos percentis 25, 50 e 75

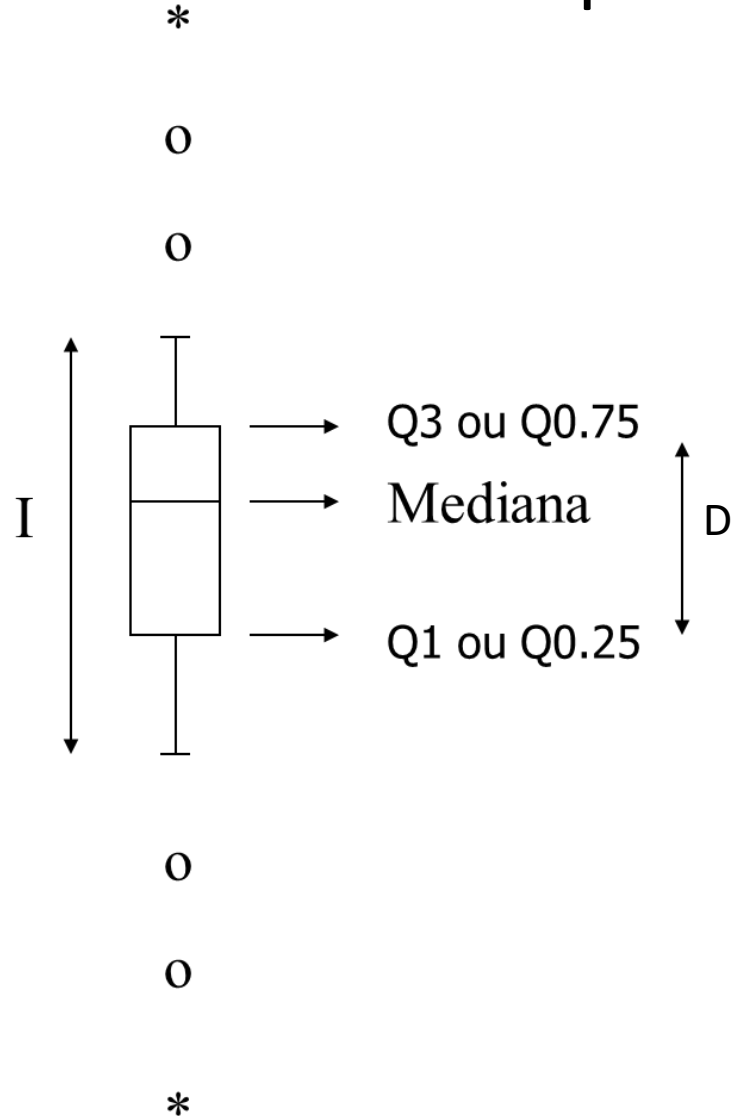
Medidas de Dispersão

- Quartis (Q1 e Q3)
 - 1º Quartil Q1 ou Q0.25 - valor a que corresponde a percentagem cumulativa de 25%
 - 3º Quartil Q3 ou Q0.75 - valor a que corresponde a percentagem cumulativa de 75%
- Mínimo (min) - Valor a que corresponde a percentagem cumulativa de 0%
- Máximo (max) - Valor a que corresponde a percentagem cumulativa de 100%

Medidas de Dispersão

- Box Plot
 - Mostra a mediana, primeiro e terceiro quartis de uma distribuição (pontos 50%, 25% e 75% na distribuição acumulada)
 - Noção de outlier: observação que se encontra a mais de um dado múltiplo (1.5 ou 3.0) do intervalo inter-quartílico, acima ou abaixo dos percentis 75% e 25%, respectivamente

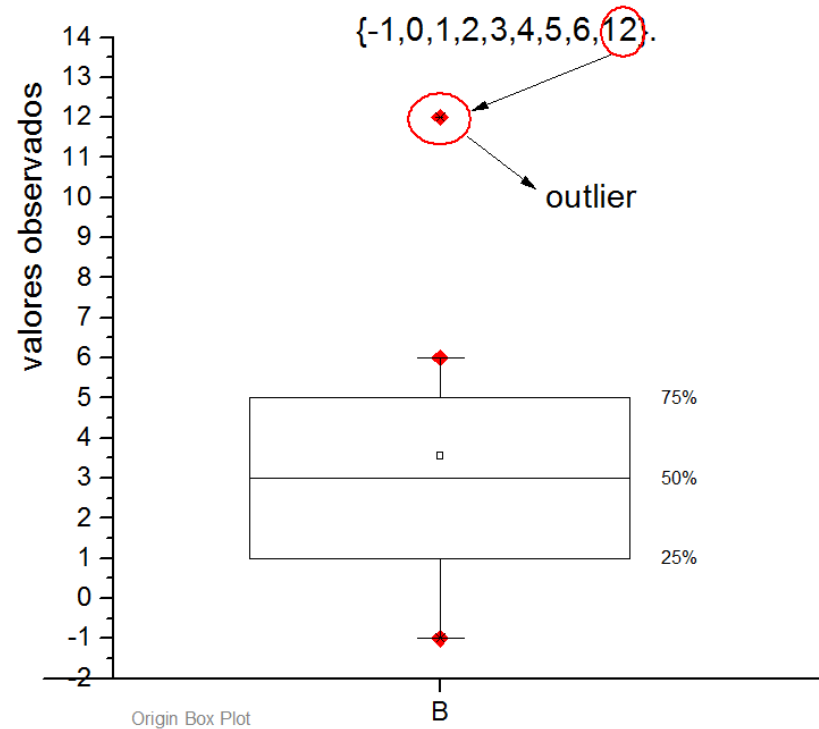
Medidas de Dispersão



- *Outliers* – valores extremos da população.
 - Moderados (1ª espécie) (o)
 - $[Q3+1.5(Q3-Q1) ; Q3+3(Q3-Q1)]$ e $[Q1-1.5(Q3-Q1) ; Q1-3(Q3-Q1)]$
 - Severos (2ª espécie) (*)
 - $[Q3+3(Q3-Q1) ; \text{máximo}]$ e $[\text{mínimo} ; Q1-3(Q3-Q1)]$
- D – Distância entre Quartis
- I – Intervalo das amostras

Medidas de Dispersão

- Mediana: 3
- Q1: 1
- Q3: 5
- D: $Q3 - Q1 = 4$



Medidas de Posição

- Valores extremos *Outliers*
 - Dados díspares, muito grandes ou muito pequenos, em relação aos demais
 - Influenciam muito as médias
 - Podem distorcer conclusões
 - É fundamental sua detecção e tratamento
- Possíveis causas de valores espúrios:
 - Erro na fase de mensuração (tomada da medida)
 - Erro na transcrição ou anotação do registro
 - Mudanças (reais) não-controláveis nas condições experimentais.
 - Característica da variável (ex.: instabilidade)

Medidas de Dispersão

- Detectando Observações Atípicas (*Outliers*)
 - As técnicas de estatística descritiva são de extrema utilidade, pois permitem a inspeção visual da distribuição dos dados.
 - Assim, para este fim, utilizam-se os seguintes recursos:
 - Histogramas e box-plots (valores maiores que $Q3 + 3*(Q3-Q1)$ e menores do que $Q1 - 3*(Q3-Q1)$);

Medidas de Dispersão

- Tratamento de *Outliers*
 - Muita controvérsia...
 - Não existe um critério ou metodologia para rejeição de dados
 - Problema maior quando tem-se pequena quantidade de dados ou uma distribuição normal não pode ser garantida.

Medidas de Dispersão

- Analisando Observações Atípicas (Exemplos)
 - Exemplo 1 – amostra sendo obtida em um sistema durante o mês de abril de 2a. a 6a. das 9 as 17;
 - 21/04 – 6a.feira – feriado - deve ser desprezado
 - Exemplo 2- Requisições de um servidor Web
 - Exemplo 3 –Acessos a um servidor Web ao longo de um dia

Medidas de Dispersão

- Podem-se desprezar dados atípicos?
 - Apenas o especialista na área pode responder.....

Conteúdo

- **Análise de resultado**
 - Introdução
 - Medidas de Desempenho Frequentemente Utilizadas
 - **Análise Estatística dos Resultados**
 - Comparação de Resultados
 - Procedimento para análise de resultados

Análise de Resultados

- Considera-se que alguma técnica para avaliação de desempenho e obtenção dos resultados tenha sido utilizada
 - No nosso caso, foi utilizada uma técnica chamada monitoração, ou aferição, onde os dados são coletados diretamente no sistema que está sendo avaliado

Análise de Resultados

- Em qualquer experimentação, três problemas devem ser considerados:
 1. Condições iniciais da experimentação
 2. Quando parar uma experimentação
 3. Resultado de uma execução oferece um resultado dentre muitos outros possíveis

Análise de Resultados

- **Terceiro problema:** durante a obtenção de dados sobre sistemas computacionais, tem-se que os resultados de uma medição oferecem um resultado dentre muitos outros possíveis
 - Deve ser utilizado em qualquer experimento que gere um conjunto de resultados possíveis
 - Após a obtenção dos resultados, estes devem ser analisados independente da técnica utilizada
 - Qual resultado deve ser considerado?
 - Como comparar dois conjuntos de resultados?

Análise de Resultados

- Como analisar os diferentes resultados de uma avaliação?
 - Primeiro objetivo da análise estatística dos resultados é estimar o erro ou intervalo de confiança
 - Deve-se considerar diversos resultados provenientes de diferentes execuções
 - Para simulação: utilizar conjuntos de números aleatórios sem correlação – diferentes sementes
 - Para **aferição**: considerar diversas medidas

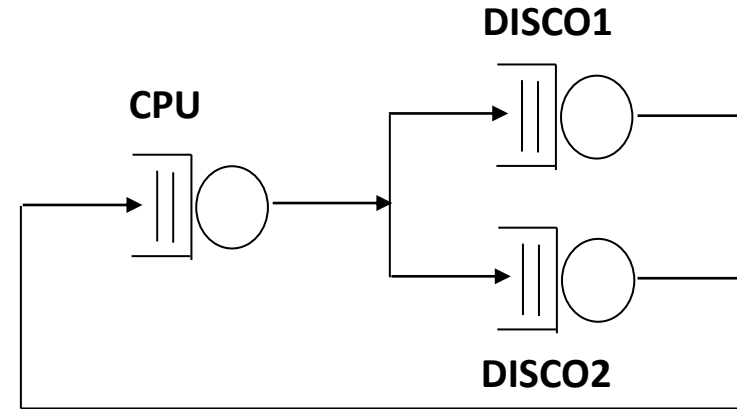
Análise de Resultados - Exemplo

- Utilização da CPU

- Execução 1: 0,36123
- Execução 2: 0,32571
- Execução 3: 0,32510
- Execução 4: 0,29999
- Execução 5: 0,35813

- Como analisar estes resultados?

- Perigo utilizar resultados de uma única simulação

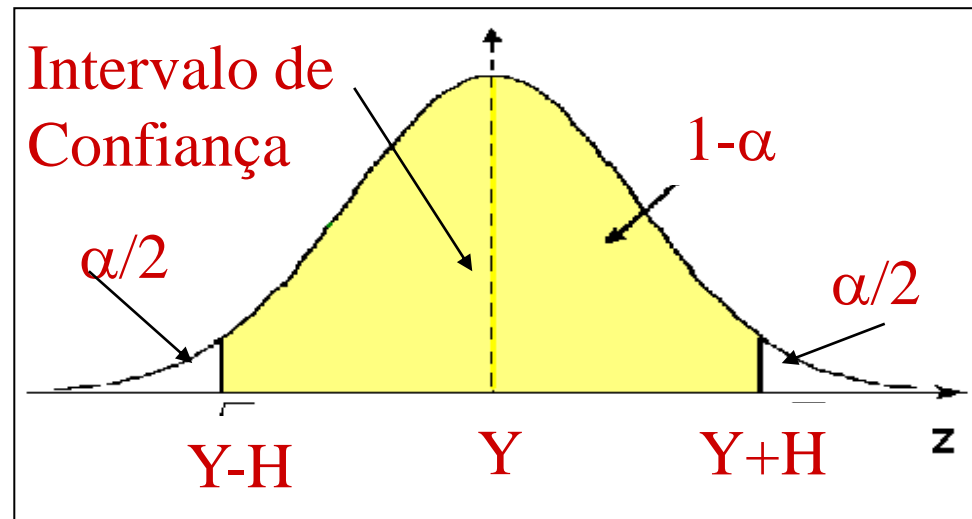


Análise de Resultados

- Utilização de Intervalos de Confiança
 - A partir de um conjunto de resultados possíveis queremos estimar o comportamento de um sistema
 - Podemos utilizar:
 - Valores fixos – Estimativa pontual
 - Intervalos – Estimativa intervalar
 - Valores fixos não permitem estimar o erro cometido
 - Intervalos de confiança permitem avaliar a confiança no resultado

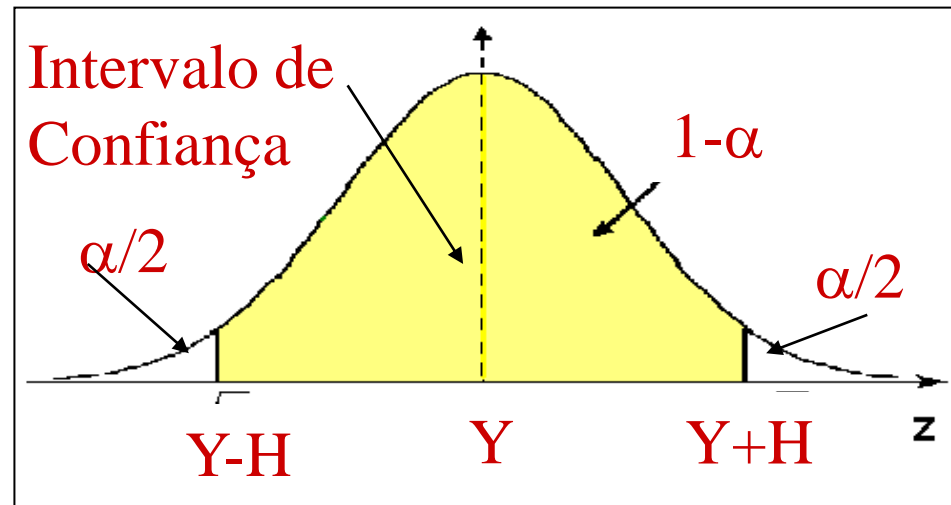
Análise de Resultados

- O que é o Intervalo de Confiança?
 - Intervalo que com uma determinada probabilidade (confiança) contém o valor do parâmetro estudado
 - Nível de confiança: $100 * (1-\alpha)\%$
 - α = probabilidade de erro
 - Y = média da amostra
 - H = largura do Intervalo de Confiança



Análise de Resultados

- O que significa Intervalo de Confiança?
 - Se Confiança = 95%, tem-se 95% de chances de que o parâmetro estará dentro do intervalo
 - Nada garante que o resultado de uma única execução (Y_i) cairá no intervalo
 - O resultado de uma única execução poderá estar na área definida por $\alpha/2$



Análise de Resultados

- Como determinar o Intervalo de confiança?
 1. Ordenar os valores obtidos
 2. Eliminar os $\alpha/2$ maiores valores
 3. Eliminar os $\alpha/2$ menores valores
 4. Obtém-se o intervalo procurado

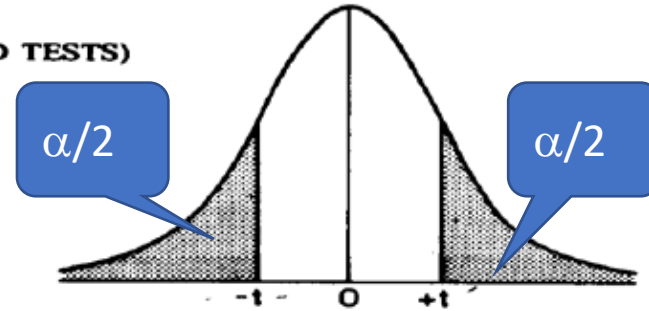
Análise de Resultados

- Como determinar o Intervalo de confiança?
 - Para amostras menores que 30 e com desvio padrão da população desconhecido
 - Utiliza-se o Teorema do Limite Central e a Tabela t-student
 - Para amostras maiores que 30 ou com desvio padrão da população conhecido
 - Utiliza-se o Teorema do Limite Central e a Tabela Z

Análise de Resultados

- Como determinar o Intervalo de confiança?
 - Para amostras menores que 30 e com desvio padrão da população desconhecido
 - Utiliza-se o Teorema do Limite Central e a Tabela t-student
 - Conheço o desvio padrão da amostra - DP

TABLE A 4
DISTRIBUTION OF t (TWO-TAILED TESTS)



Degrees of Freedom	Probability of a Larger Value, Sign Ignored								
	0.500	0.400	0.200	0.100	0.050	0.025	0.010	0.005	0.001
1	1.000	1.376	3.078	6.314	12.706	25.452	63.657		
2	0.816	1.061	1.886	2.920	4.303	6.205	9.925	14.089	31.598
3	.765	0.978	1.638	2.353	3.182	4.176	5.841	7.453	12.941
4	.741	.941	1.533	2.132	2.776	3.495	4.604	5.598	8.610
5	.727	.920	1.476	2.015	2.571	3.163	4.032	4.773	6.859
6	.718	.906	1.440	1.943	2.447	2.969	3.707	4.317	5.959
7	.711	.896	1.415	1.895	2.365	2.841	3.499	4.029	5.405
8	.706	.889	1.397	1.860	2.306	2.752	3.355	3.832	5.041
9	.703	.883	1.383	1.833	2.262	2.685	3.250	3.690	4.781
10	.700	.879	1.372	1.812	2.228	2.634	3.169	3.581	4.587
11	.697	.876	1.363	1.796	2.201	2.593	3.106	3.497	4.437
12	.695	.873	1.356	1.782	2.179	2.560	3.055	3.428	4.318
13	.694	.870	1.350	1.771	2.160	2.533	3.012	3.372	4.221
14	.692	.868	1.345	1.761	2.145	2.510	2.977	3.326	4.140
15	.691	.866	1.341	1.753	2.131	2.490	2.947	3.286	4.073
16	.690	.865	1.337	1.746	2.120	2.473	2.921	3.252	4.015
17	.689	.863	1.333	1.740	2.110	2.458	2.898	3.222	3.965
18	.688	.862	1.330	1.734	2.101	2.445	2.878	3.197	3.922
19	.688	.861	1.328	1.729	2.093	2.433	2.861	3.174	3.883
20	.687	.860	1.325	1.725	2.086	2.423	2.845	3.153	3.850

Análise de Resultados

- Como determinar o Intervalo de Confiança?
 - Média Amostral $\rightarrow Y_i$ = Média das observações de uma execução
 - Estimativa Global $\rightarrow Y$ = Média das médias amostrais
- Variância Amostral \rightarrow
$$s^2 = \sum_{i=1}^N (y_i - y)^2 / (n - 1)$$
- Desvio Padrão (DP) \rightarrow
$$\sqrt{s^2}$$
- $t_{\alpha/2, N-1}$ \rightarrow distribuição t-Student com N-1 graus de liberdade e nível de confiança igual a $1 - \alpha$

Análise de Resultados

- Como determinar o Intervalo de Confiança?

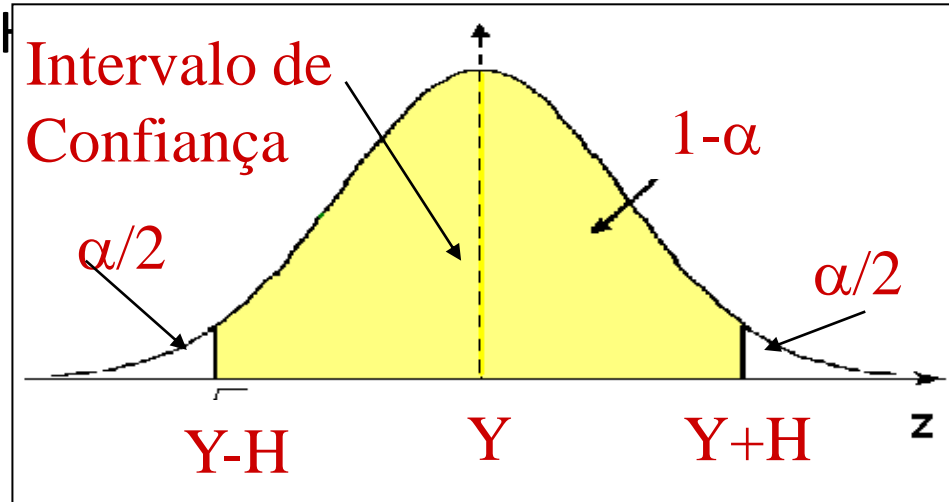
- Largura do Intervalo de Confiança de $100 * (1-\alpha)\%$

- $H = t * \text{desvio} / \sqrt{N}$

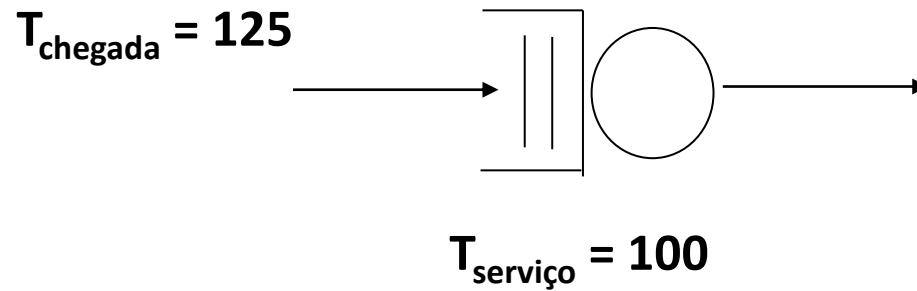

$$\sqrt{s^2}$$

- $H = t_{\alpha/2, N-1} * \sqrt{\frac{s^2}{N}}$

- Intervalo de Confiança: $Y \pm H$



Análise de Resultados



- X_i : tempo na fila para o cliente i
- X : tempo médio na fila para 5.000 clientes
- μ : média real para o tempo na fila ... não conhecido

Análise de Resultados

- Simulação executada 10 vezes para diferentes conjunto de números aleatórios, obtendo-se
 - $Y = 406,554$
- Esse valor está suficientemente próximo de μ ?

331,993	447,532
366,052	420,858
403,524	355,959
464,856	492,144
393,393	389,200

Análise de Resultados

- Simulação executada 10 vezes

$$1-\alpha = 0,95 \rightarrow \alpha = 0,05$$

$$\text{Média} = Y = 406,551$$

$$\text{Variância} = S^2 = \sum_{i=1}^{10} \frac{(331,993 - 406,551)^2 + \dots + \dots}{9} = 2539,082$$

$$\sqrt{s^2} = \sqrt{2539,082} = 50,39$$

DP =

331,993	447,532
---------	---------

366,052	420,858
---------	---------

403,524	355,959
---------	---------

464,856	492,144
---------	---------

393,393	389,200
---------	---------

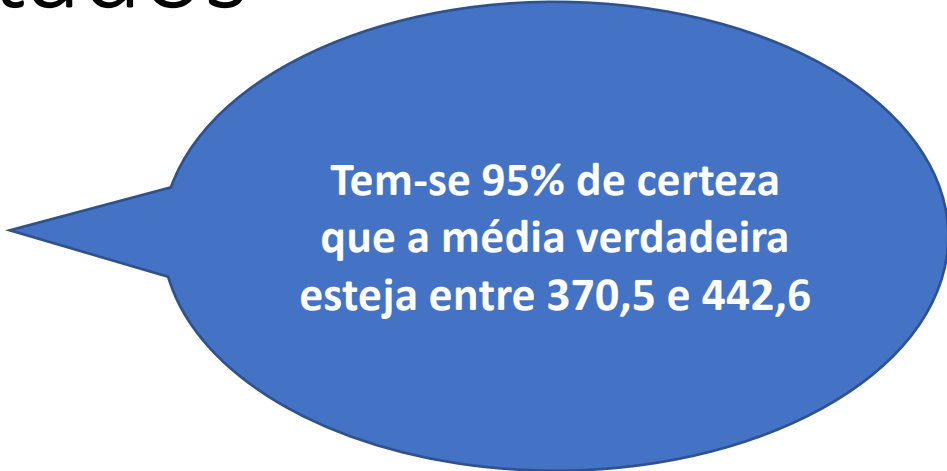
$$t_{0,05/2;9} = 2,26$$

$$H = t_{0,05/2;9} * DP / \sqrt{N} = 36,04$$

IC: 370,5 – 442,6

Análise de Resultados

- $H = 36,04$
- IC: 370,5 – 442,6



Tem-se 95% de certeza
que a média verdadeira
esteja entre 370,5 e 442,6

- Se a simulação for repetida várias vezes e em cada vez for determinado o intervalo de confiança, 95% destes intervalos irão conter a média verdadeira
- Valores individuais podem estar fora do intervalo de confiança

Análise de Resultados

- Média: $Y = 406,554$
- $H = 36,04$
- IC: $370,5 - 442,6$
- Amplitude do intervalo de confiança = $72,08$
 - $17,7\%$ do valor médio
- Não é um valor muito grande?
- Como diminuir?
 - **Aumentando o número de replicações**

Análise de Resultados

- Controle do Erro
- Utilizando-se a técnica de replicações:
 - Não pode-se determinar a precisão desejada
 - Pode-se determinar, aproximadamente, quantas replicações adicionais deve-se considerar para controlar o erro

Análise de Resultados -Exemplo

- Para o exemplo anterior
- Suponha que se queira $2 \cdot H \leq 15\%$ da Média $\Rightarrow H \leq 30,49$
- Seja r = as próximas replicações

r	$t_{r-1;0,95}$	$\sqrt{\frac{22851,74}{r-1}}$	$t_{r-1;0,95} * \sqrt{\frac{22851,74}{r-1}}$
11	2,23	15,934	35,53393
12	2,20	15,193	33,42448
13	2,18	14,546	31,71058
14	2,16	13,975	30,18703

- **São necessárias mais 4 replicações para atingir o erro máximo desejado**

Conteúdo

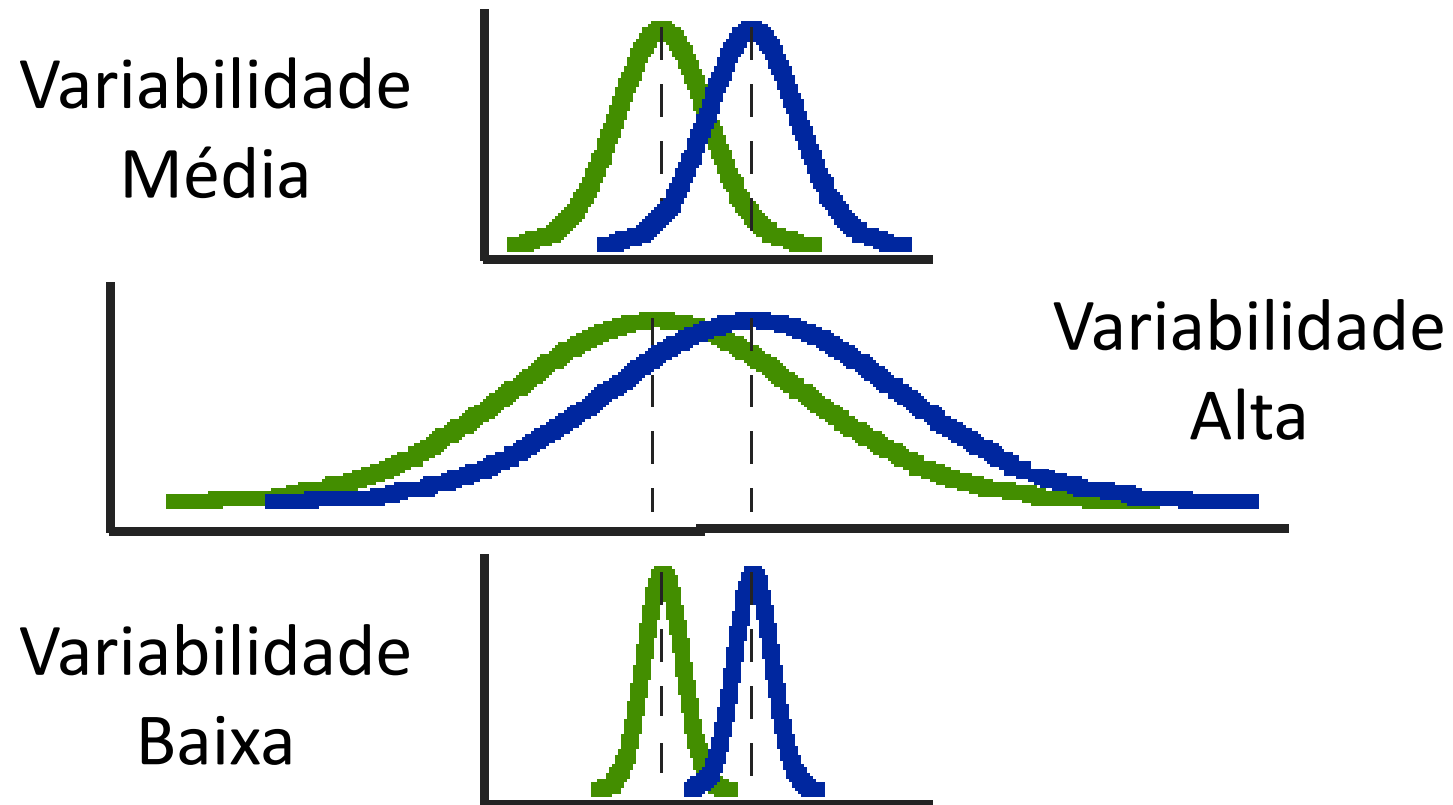
- **Análise de resultado**
 - Introdução
 - Medidas de Desempenho Frequentemente Utilizadas
 - Análise Estatística dos Resultados
 - **Comparação de Resultados**
 - Procedimento para análise de resultados

Análise de Resultados

- Comparação entre dois experimentos
 - Testes estatísticos podem ser utilizados para definir se os resultados provenientes de dois experimentos são conclusivos
 - Definem se os resultados são estatisticamente diferentes

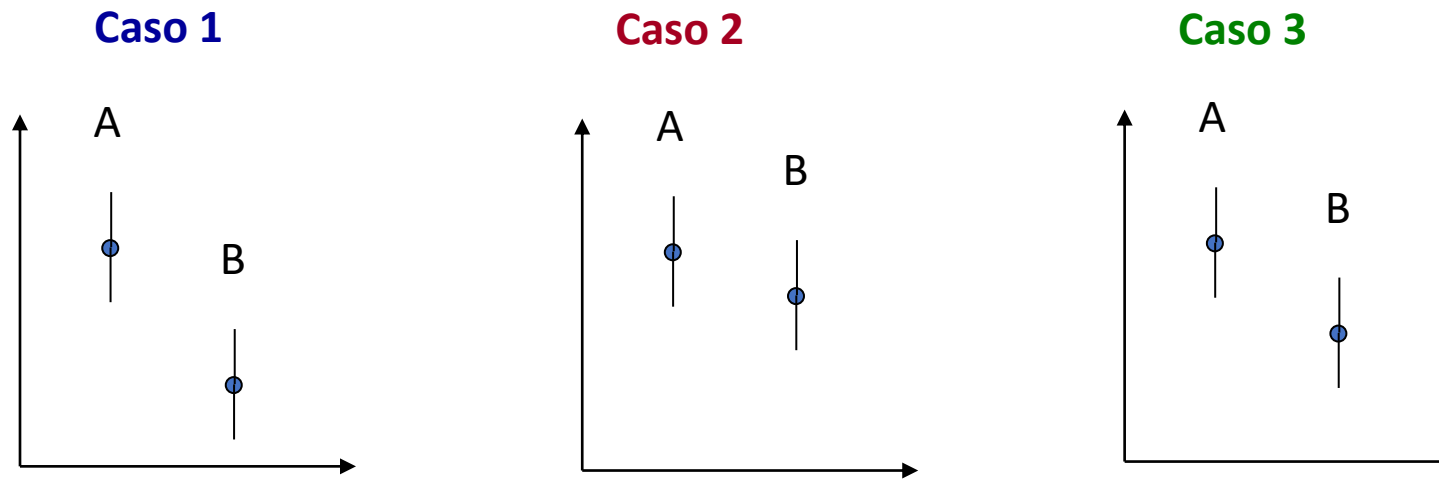
Análise de Resultados

- Comparação entre dois experimentos: **teste visual**



Análise de Resultados

- Comparação entre dois experimentos: **teste visual**



- **Caso 1** – ICs não sobrepostos $\Rightarrow A > B$
- **Caso 2** – Média de um está inserida no IC do outro $\Rightarrow A = B$
- **Caso 3** – ICs sobrepostos mas média está fora \Rightarrow necessário outro teste

Comparação entre dois experimentos

- Área de Estatística oferece grande número de testes para comparação entre experimentos:
 - Teste de hipótese:
 - Teste-t: para comparar a média de duas amostras
 - Análise de variância (ANOVA): para comparar média de três ou mais amostras
 - Teste para amostras pareadas
 - Teste para amostras não pareadas
 - Chi-Quadrado e Poisson - para valores não contínuos

Teste de hipótese

- Assume-se que uma proposição inicial é verdadeira e então testa essa proposição com os dados amostrais
 - Hipótese nula (H_0): proposição inicial, normalmente definida por pesquisas prévias ou pelo senso comum
 - Hipótese alternativa (H_1): a que se acredita que pode ser verdadeira
 - Valor de probabilidade (valor-p):
 - Se o valor-p é menor ou igual a um determinado nível de significância (nível α), a hipótese nula é rejeitada e a hipótese alternativa é apoiada.
 - Se o valor-p é maior que o nível α , a hipótese nula não pode ser rejeitada e a hipótese alternativa não tem apoio.

Teste de hipótese

Teste-t

- Cálculo do valor de t para a amostra:

$$t = \frac{\bar{X}_T - \bar{X}_C}{\sqrt{\frac{var_T}{n_T} + \frac{var_C}{n_C}}}$$

- Cálculo do número de graus de liberdade -> $n = n_T + n_C - 2$
- Entrar na tabela t-student com n e confiança desejada: t_{tab}
 - Se $t > t_{tab}$: médias são diferentes
 - Se $t < t_{tab}$: não existe diferença significativa entre as médias

Análise de Resultados

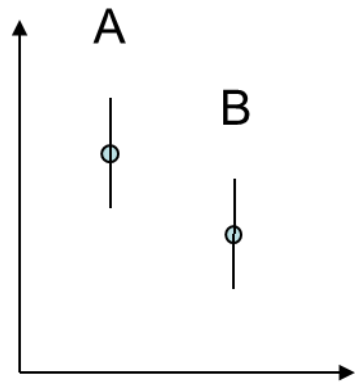
- Após 10 replicações de um programa de simulação, avaliando-se o tempo médio na fila de um recurso, obtiveram-se as médias das amostras e intervalo de confiança para 95% e para 90% representados na tabela a seguir:

	A	B
Média	12	11
H₁ (0,05)	0,8	0,9
H₂ (0,1)	0,4	0,5

- O que se pode concluir?

Análise de Resultados

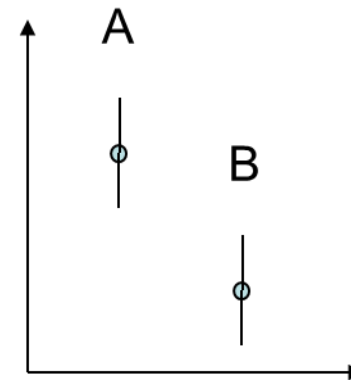
	A	B
Média	12	11
H₁ (0,05)	0,8	0,9
H₂ (0,1)	0,4	0,5



H₁ (0,05)

A - 11,2 – 12,8

B – 10,1 – 11,9



H₂ (0,1)

A – 11,6 – 12,4

B – 10,5 – 11,5

Análise de Resultados

$$H_0: \mu_A = \mu_B$$

$$H_1: \mu_A \neq \mu_B$$

$$\text{Estatística } t_{obs} = \frac{X_A - X_B}{\sqrt{\frac{var_A}{n_A} + \frac{var_B}{n_B}}}$$

$$var_A = 1,13$$

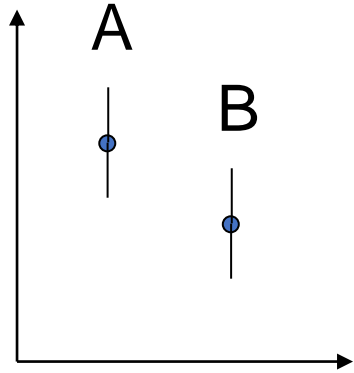
$$var_B = 1,27$$

$$n = n_A + n_B - 2 = 18$$

$$t_{obs} = \frac{12 - 11}{\sqrt{\frac{1,13}{10} + \frac{1,27}{10}}} = 2,040$$

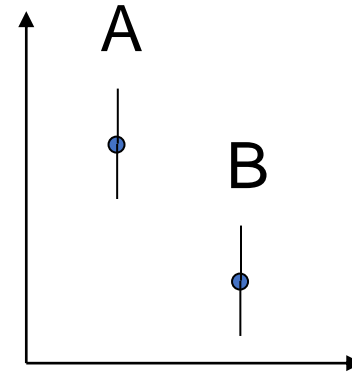
	A	B
Média	12	11
H₁ (0,05)	0,8	0,9
H₂ (0,1)	0,4	0,5

Análise de Resultados



$$H_1(0,05) \rightarrow t_{\text{obs}} = 2,040 < 2,101 = t_{\text{tab}}$$

Não existe diferença significativa



$$H_1(0,1) \rightarrow t_{\text{obs}} = 2,040 > 1,734 = t_{\text{tab}}$$

Médias diferentes

Região Crítica (RC): se o t_{obs} não pertencer a reunião crítica, aceita-se H_0

$$\{t \in \mathbb{R}; t < t_1 \text{ ou } t > t_2\}$$

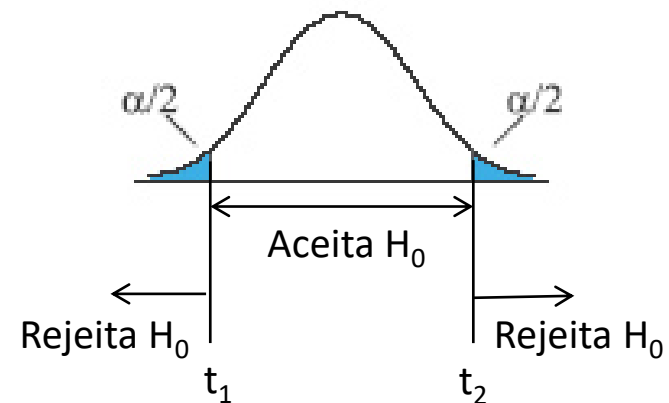
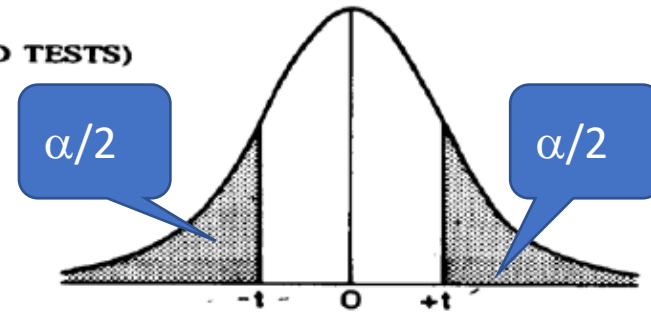


TABLE A 4
DISTRIBUTION OF t (TWO-TAILED TESTS)



Degrees of Freedom	Probability of a Larger Value, Sign Ignored								
	0.500	0.400	0.200	0.100	0.050	0.025	0.010	0.005	0.001
1	1.000	1.376	3.078	6.314	12.706	25.452	63.657		
2	0.816	1.061	1.886	2.920	4.303	6.205	9.925	14.089	31.598
3	.765	0.978	1.638	2.353	3.182	4.176	5.841	7.453	12.941
4	.741	.941	1.533	2.132	2.776	3.495	4.604	5.598	8.610
5	.727	.920	1.476	2.015	2.571	3.163	4.032	4.773	6.859
6	.718	.906	1.440	1.943	2.447	2.969	3.707	4.317	5.959
7	.711	.896	1.415	1.895	2.365	2.841	3.499	4.029	5.405
8	.706	.889	1.397	1.860	2.306	2.752	3.355	3.832	5.041
9	.703	.883	1.383	1.833	2.262	2.685	3.250	3.690	4.781
10	.700	.879	1.372	1.812	2.228	2.634	3.169	3.581	4.587
11	.697	.876	1.363	1.796	2.201	2.593	3.106	3.497	4.437
12	.695	.873	1.356	1.782	2.179	2.560	3.055	3.428	4.318
13	.694	.870	1.350	1.771	2.160	2.533	3.012	3.372	4.221
14	.692	.868	1.345	1.761	2.145	2.510	2.977	3.326	4.140
15	.691	.866	1.341	1.753	2.131	2.490	2.947	3.286	4.073
16	.690	.865	1.337	1.746	2.120	2.473	2.921	3.252	4.015
17	.689	.863	1.333	1.740	2.110	2.458	2.898	3.222	3.965
18	.688	.862	1.330	1.734	2.101	2.445	2.878	3.197	3.922
19	.688	.861	1.328	1.729	2.093	2.433	2.861	3.174	3.883
20	.687	.860	1.325	1.725	2.086	2.423	2.845	3.153	3.850

Teste de hipótese X Intervalo de Confiança

- Teste de hipótese
 - Resposta: aceita ou rejeita a hipótese
 - Conclusivo: não deixa dúvida
 - Não oferece maiores informações
 - Difícil de interpretar
 - O que significa $t_{obs} = 2,101 > 2,040$?

Teste de hipótese X Intervalo de Confiança

- Intervalo de confiança
 - Informações adicionais
 - Intervalo pequenos -> parâmetro bem estimado
 - Valores com o mesmo significado que as medidas originais -> mais fácil de entender e analisar
 - Significado de Média=12 e H(95%)=0,6

Conteúdo

1. Introdução
2. Planejamento de Experimentos
3. Técnicas para Avaliação de Desempenho
4. **Análise de resultado**
 - Introdução
 - Medidas de Desempenho Frequentemente Utilizadas
 - Análise Estatística dos Resultados
 - Comparação de Resultados
 - **Procedimento para análise de resultados**

Procedimento para análise de resultados

1. Escolher as variáveis de resposta adequadas e suficientes para atingir o objetivo
2. Escolher os fatores e níveis adequadamente
3. Realizar o experimento quantas vezes forem necessárias
4. Pensar na melhor forma de apresentar os dados
5. Fazer o tratamento estatístico adequado para os resultados
6. Observar os resultados e correlaciona-los com o que se conhece do sistema sendo avaliado

Procedimento para análise de resultados

1. Escolher as variáveis de resposta adequadas e suficientes para atingir o objetivo
 - Não considerar variáveis essenciais pode levar a erros na análise
 - Considerar que variáveis desnecessárias contribuem para aumentar a complexidade da análise
 - Sempre tentar analisar conjuntos não muito grandes de variáveis e, se necessário, realizar a análise em diversas fases

Procedimento para análise de resultados

2. Escolher os fatores e níveis adequadamente

- Escolher poucos fatores e, se possível, apenas dois níveis por fator
- Para os fatores com grande influência nas variáveis de resposta, detalhar processo separadamente.
- Considerar um grande número de fatores e de níveis em um primeiro momento da avaliação, torna a análise suscetível a erros.

Procedimento para análise de resultados

3. Realizar o experimento quantas vezes forem necessárias
 - Utilizar um dos métodos apresentados para determinar o ponto de parada de coleta de dados
 - Não tirar conclusões baseando-se em um único resultado

Procedimento para análise de resultados

4. Pensar na melhor forma de apresentar os dados
 - Tabelas são ótimas para observar detalhes e valores precisos
 - Gráficos são adequados para melhor visualizar os resultados
 - Nos gráficos, cuidado com escalas e origem dos eixos
 - Muitos valores em uma tabela ou em um gráfico tornam a análise mais complexa, e possíveis resultados mais difíceis de serem identificados

Procedimento para análise de resultados

5. Fazer o tratamento estatístico adequado para os resultados
 - Não tirar conclusões considerando-se apenas médias
 - Valores médios só fazem sentido quando acompanhados de desvio padrão, variância, intervalo de confiança, etc.
 - Valores médios com máximo e mínimo podem ajudar na análise mas não levam a resultados conclusivos, sem a presença de uma métrica que indique a dispersão dos dados

Procedimento para análise de resultados

6. Observar os resultados e correlaciona-los com o que se conhece do sistema sendo avaliado
 - Desconfie de resultados não esperados
 - Tente relacionar os diferentes resultados obtidos
 - Tente explicar os resultados obtidos