Métodos Quantitativos para Ciência da Computação Experimental

Erros de medição - Comparando sistemas através de dados amostrais – Ch. 13 do Jain

Virgílio A. F. Almeida

22 de Abril de 2010

Departamento de Ciência da Computação
Universidade Federal de Minas Gerais

Erros em Medições Experimentais

- Fontes de erros
- Acurácia, precisão, resolução
- Um modelo matemático de erros
- Intervalos de confiança
  - Para média
  - Para proporções
- Quantas medições são necessárias para níveis desejados de erro?
- Seções do Jain: 13-1, 13-2, 13.8 e 13.9

Por que precisamos de estatística?

1. Ruidos, ruídos,….

OK????

Por que precisamos de estatística?

445 446 397 226
388 3445 188 1002
47762 432 54 12
98 345 2245 8839
77492 472 565 999
1 34 882 545 4022
827 572 597 364

2. Agregar dados em informação relevante....

$$\overline{x} = \ldots$$

O que é *estatística*?

- “ quantidades calculadas a partir de uma amostra [de dados].”

→ Sumarizações de grandes coleções de dados.

Relembrando os objetivos do curso

- Prover base conceitual intuitiva para o uso de ferramentas estatíticas em processos computacionais experimentais!!!!
  - Obter/tirar conclusões significativas na presença de medições com ruídos (normal!!!).
  - Permitir que sejam aplicadas técnicas estatísticas em novas situações e problemas.

→ Não use simplestemente as fórmulas... saber porque...

Fontes de Erros Experimentais

- Acurácia, precisão, resolução

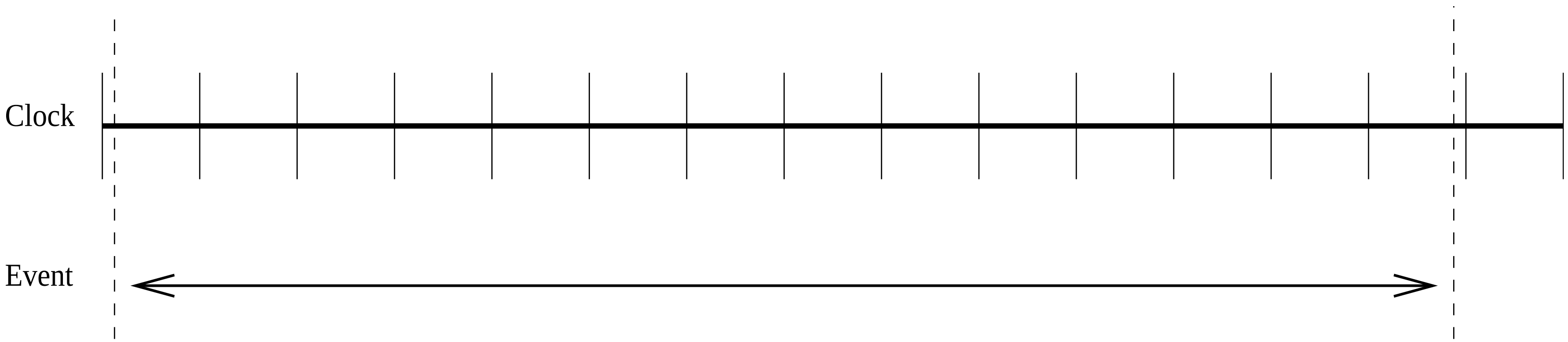
Erros experimentais

- Erros → *ruidos* nos valores medidos (ruidos na medição)
- *Erros Sistematicos*
  - Resultado de um probelma (“mistake”) experimental
  - Tipicamente produz variações tendenciosas lentas ou padrão tendecnioso constantes
- Controlado através da experiência/habilidade do experimentador
- Exemplos
  - Mudanças de temperatura podem afetar o ‘’clock”.
  - Esquecer de limpar o cache antes de medir tempos de execução

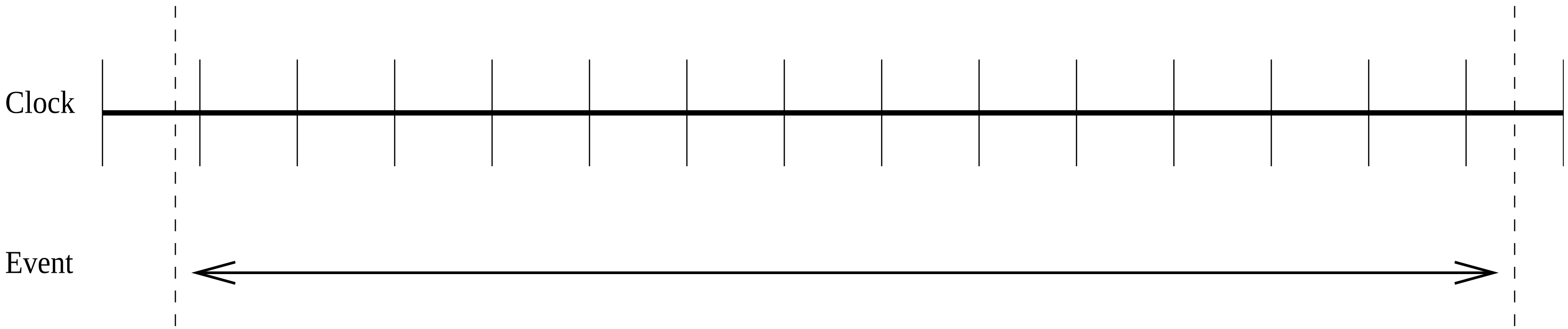
Erros experimentais

- Erros *Random*icos
  - Imprevisíveis e não-determinísticos
  - Não tendenciosos (unbiased) → probabidade igual de aumentar ou diminuir valores medidos
- Resultantes de
  - Limitações das ferramentas de medições
  - Obtenção/registro dos dados de medição pelo observador
  - Processo randomicodentro do sistema, sem causa conhecida
- Tipicamente dificil de ser controlado
  - Uso de ferramentas estatíticas para caracterizar e avaliar

Exemplo: Discretização (quantum)
→ Erro Randomico



(a) Interval timer reports event duration of n = 13 clock ticks.



(b) Interval timer reports event duration of n = 14 clock ticks.

Erro de discretização (Quantization)

- Resolução do Timer

  → erro de quantização
- Medições repetidas

  $X \pm \Delta$

  Completamente imprevisível

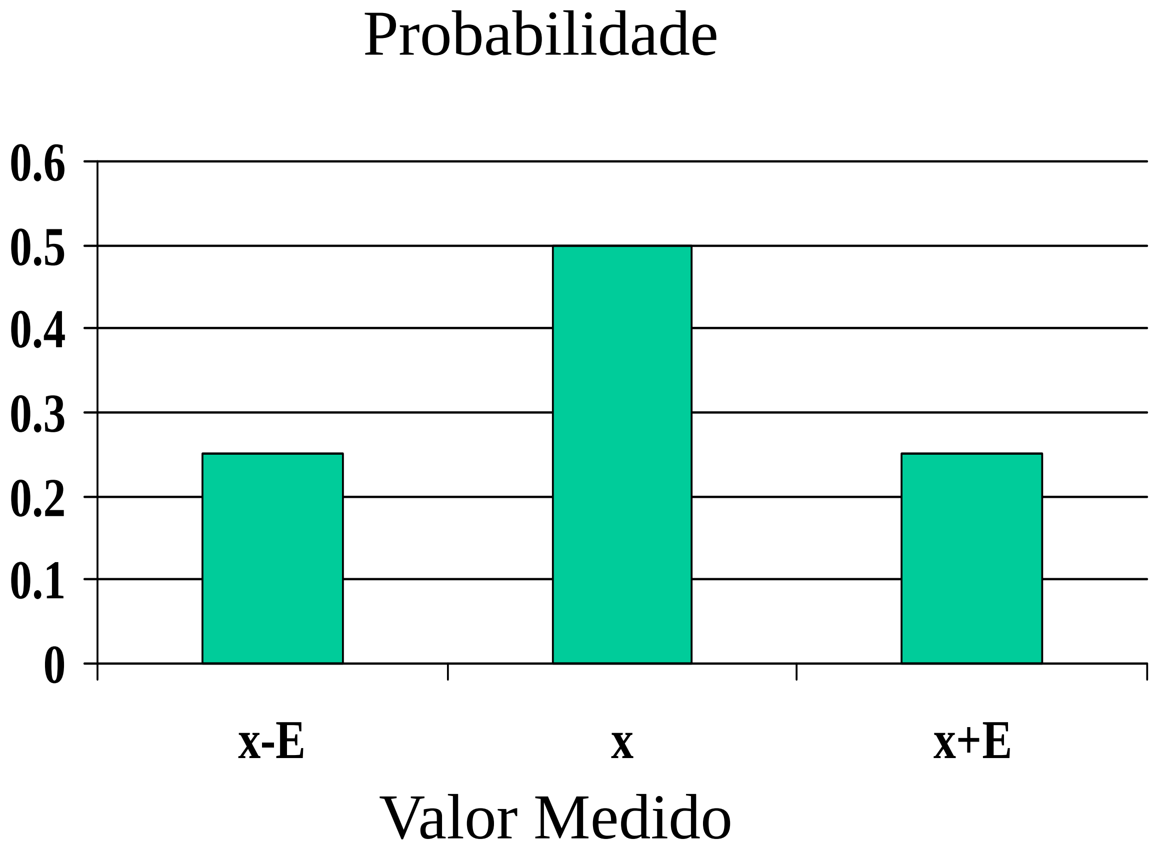
Um Modelo de Erros

Erros	*Valor Medido*	*Probabilidade*
-E	x – E	½
+E	x + E	½

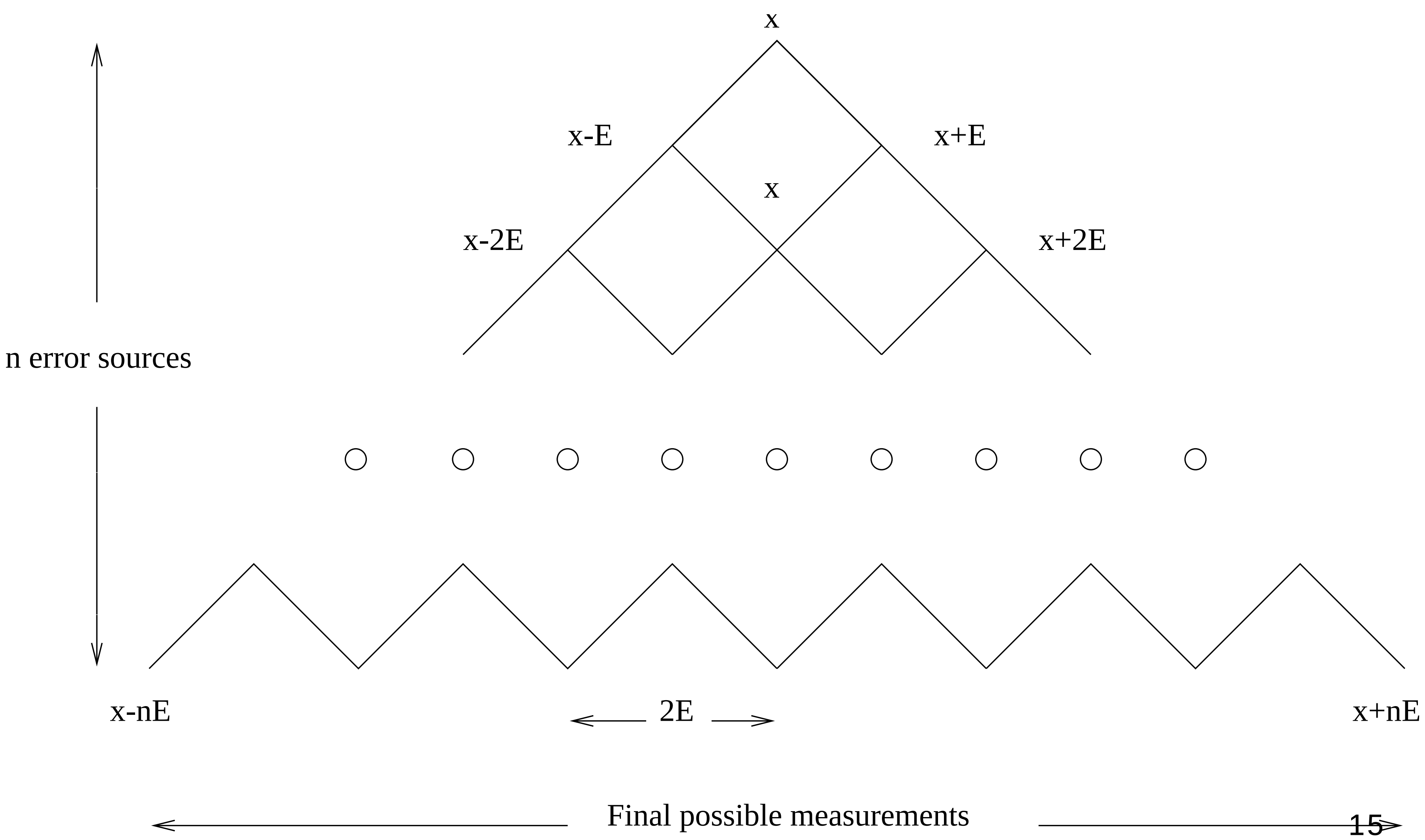
Um modelo de erros

Erro 1	*Erro 2*	*Valor Medido*	*Probabilidade*
-E	-E	$x – 2E$	¼
-E	+E	x	¼
+E	-E	x	¼
+E	+E	$x + 2E$	¼

Um Modelo de Erros



Probabilidade de Obter um Valor Medido Específico

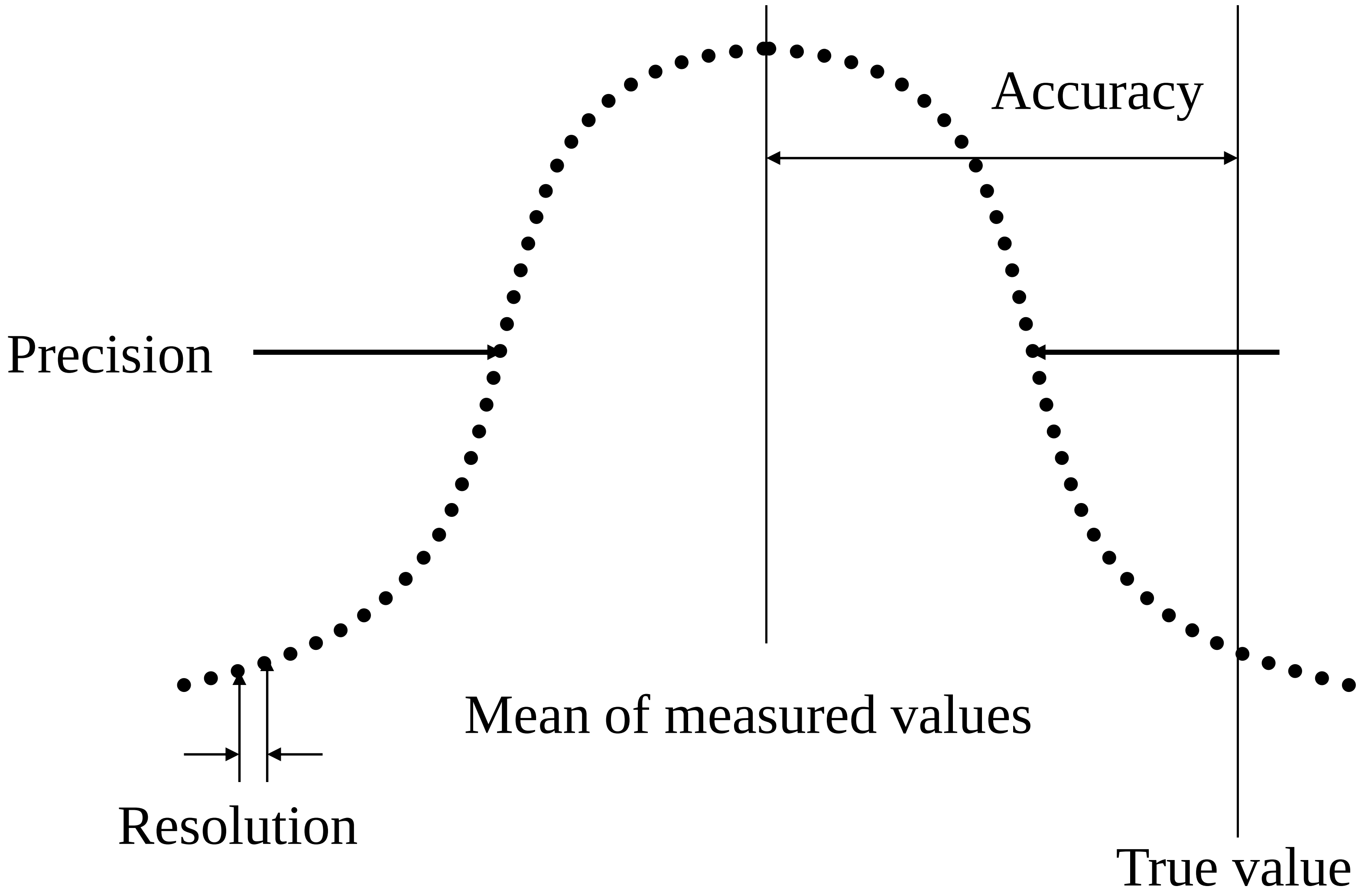


Um modelo de erros

- $Pr(X=x_i) = Pr($ medição $x_i)$

  = número de caminhos do valor real até x_i
- $Pr(X=x_i) \sim$ distribuição binomial
- A medida que o número de fontes de erro cresce
  - $n \rightarrow \infty$,
  - Binomial → Gaussian (Normal)
- Assim, a bell curve

Frequencia de Medição de Valores Específicos



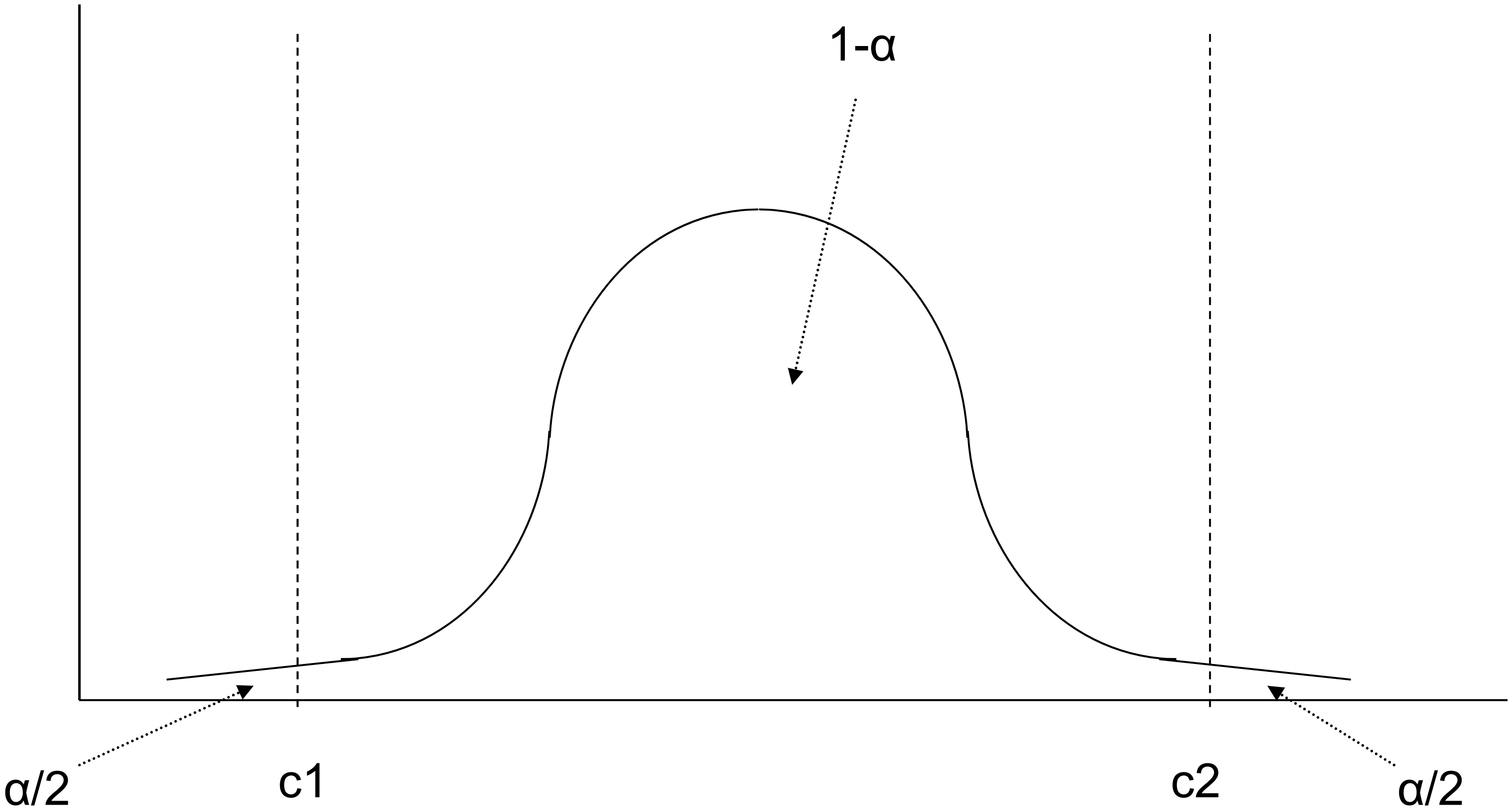
Acurácia, Precisão, Resolução

- Erros Sistematicos → acurácia
  - Quão próximo a média dos valores medidos está do valor verdadeiro
- Erros Randomicos → precisão
  - Repetibilidade das medições
- Caracteristicas das ferramentas → resolução
  - Pequenos incrementos entre medições

Quantificando Acurácia, Precisão, Resolução

- Acurácia
  - Dificil de determinar a verdadeira acurácia
  - Relativa a um padrão pre-definido
    - Ex.: definição de um “segundo”
- Resolução
  - Dependente das ferramentas (quais em computação?)
- Precisão
  - Quantificar quantidade de *imprecisão* usando ferramentas estatísticas

Intervalo de Confiança para Média



Normalize *x*

$$z = \frac{\bar{x} - x}{s / \sqrt{n}}$$

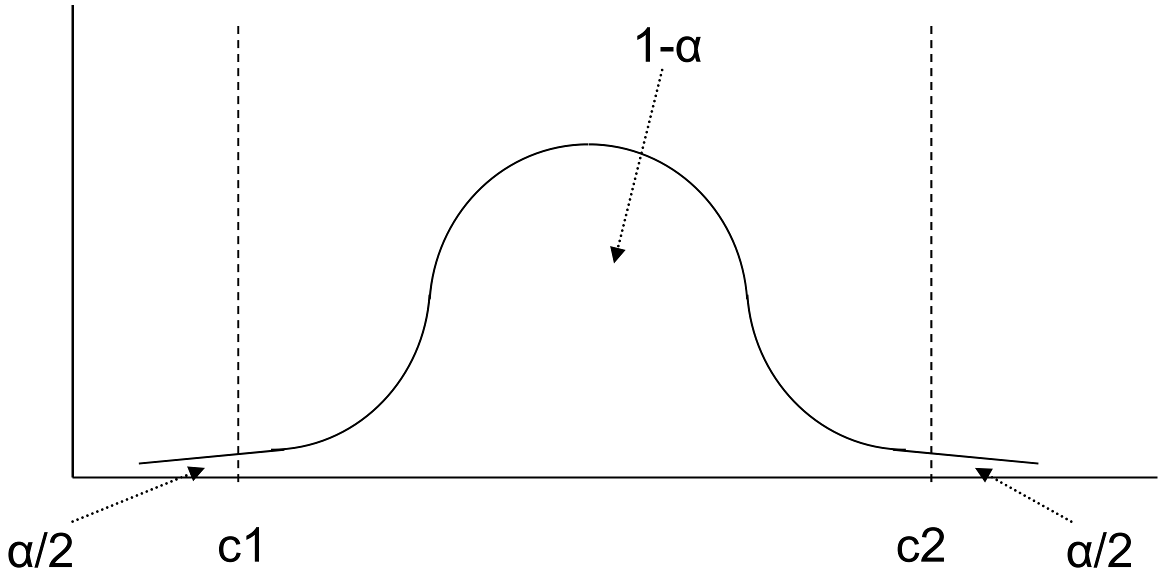
$$n = \text{number of measurements}$$

$$\bar{x} = \text{ mean} = \sum_{i=1}^{n} x_i$$

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

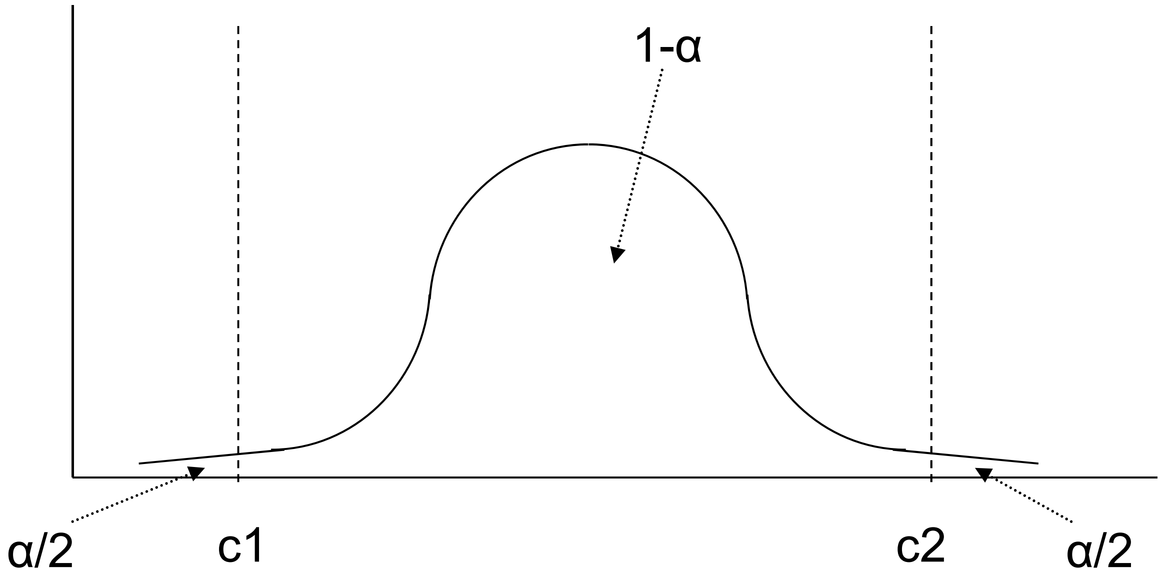
Intervalo de Confiança para Média

- *z* normalizado segue a distribuição *Student's t*
  - (n-1) graus de liberdade
  - Área à esquerda de $c_2 = 1 - \alpha/2$
  - Valores tabelados para *t*



Intervalo de Confiança para Média

- Quando $n \rightarrow \infty$, distribuição normalizada torna-se Gaussiana (normal)



Intervalo de Confiança para Média

$$c_1 = \bar{x} - t_{1-\alpha/2;n-1} \frac{s}{\sqrt{n}}$$

$$c_2 = \bar{x} + t_{1-\alpha/2;n-1} \frac{s}{\sqrt{n}}$$

Then,

$$\Pr(c_1 \leq x \leq c_2) = 1 - \alpha$$

Exemplo

Experimento	Valor Medido
1	8.0 s
2	7.0 s
3	5.0 s
4	9.0 s
5	9.5 s
6	11.3 s
7	5.2 s
8	8.5 s

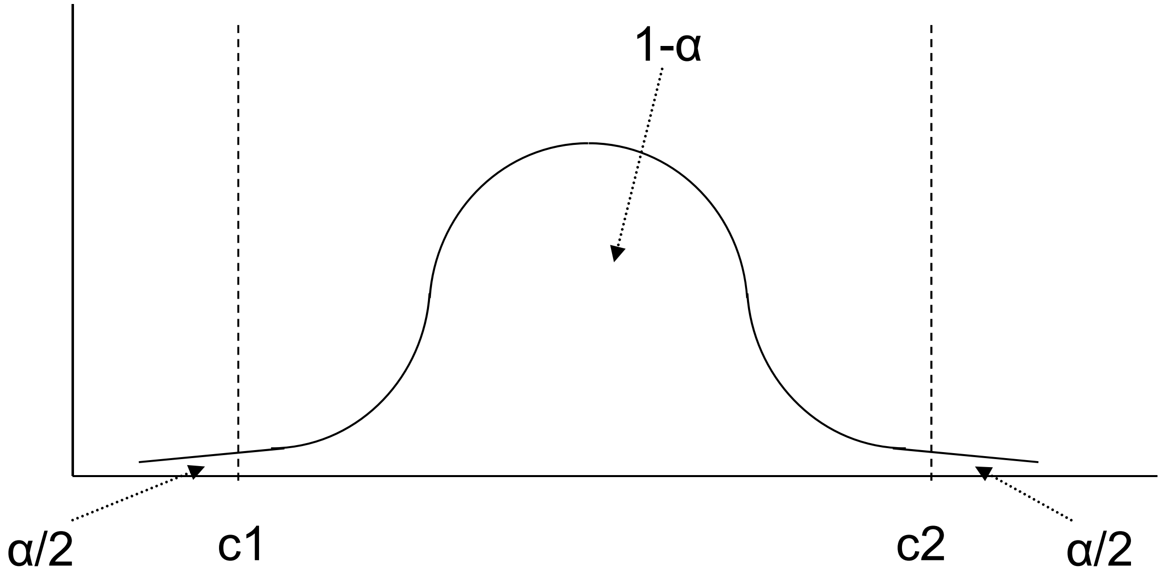
Exemplo (cont.)

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = 7.94$$

$$s = \text{sample standard deviation} = 2.14$$

Exemplo (cont.)

- 90% CI → 90% probabilidade do valor real no intervalo
- 90% CI → $\alpha = 0.10$
  - $1 - \alpha /2 = 0.95$
- $n = 8$ → 7 graus de liberdade



90% Intervalo de Confiança

$$a = 1 - \alpha / 2 = 1 - 0.10 / 2 = 0.95$$

$$t_{a;n-1} = t_{0.95;7} = 1.895$$

$$c_1 = 7.94 - \frac{1.895(2.14)}{\sqrt{8}} = 6.5$$

$$c_2 = 7.94 + \frac{1.895(2.14)}{\sqrt{8}} = 9.4$$

	a		
n	0.90	0.95	0.975
…	…	…	…
5	1.476	2.015	2.571
6	1.440	1.943	2.447
7	1.415	1.895	2.365
…	…	…	…
∞	1.282	1.645	1.960

95% Intervalo de Confiança

$$a = 1 - \alpha / 2 = 1 - 0.10 / 2 = 0.975$$

$$t_{a;n-1} = t_{0.975;7} = 2.365$$

$$c_1 = 7.94 - \frac{2.365(2.14)}{\sqrt{8}} = 6.1$$

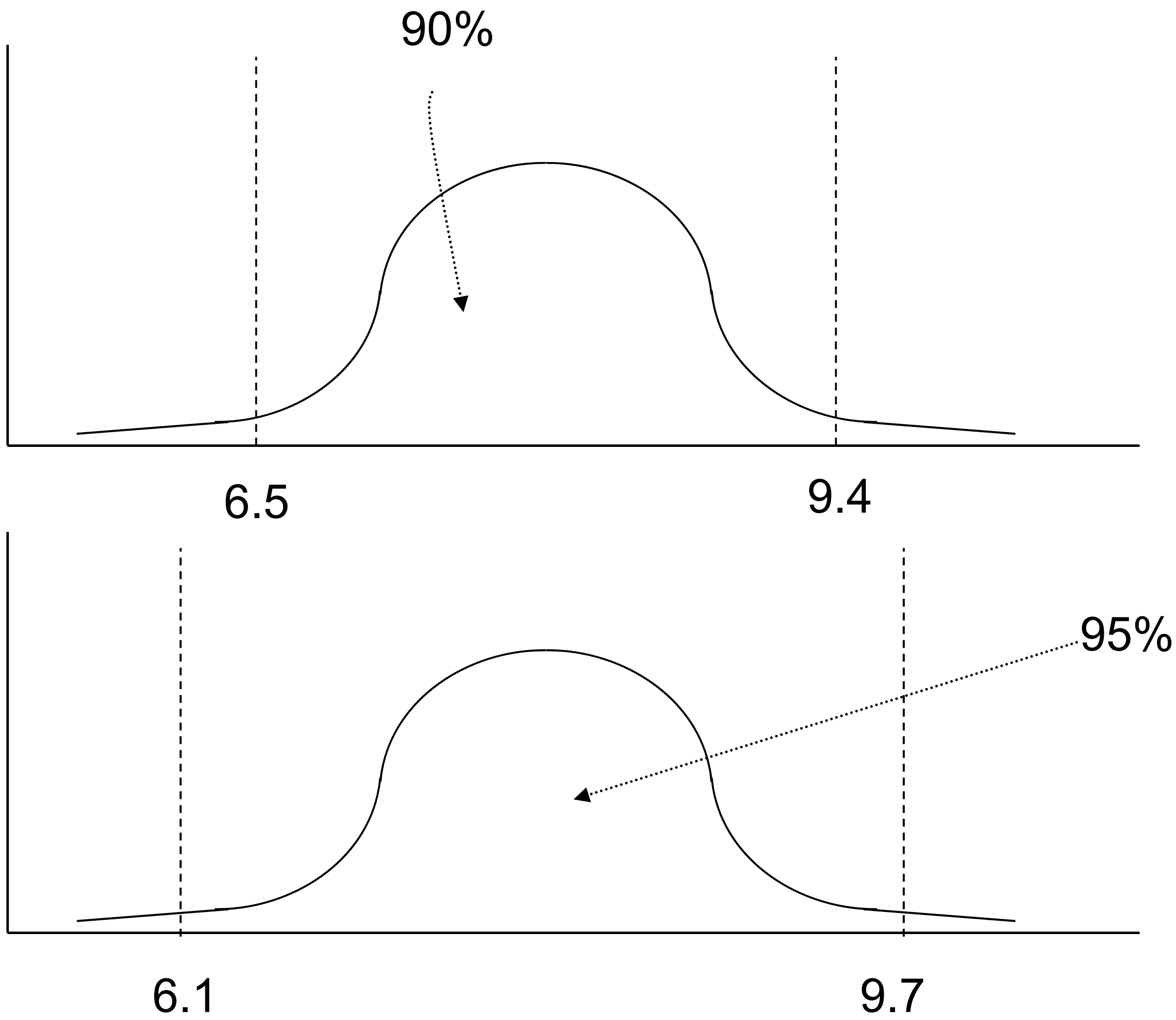
$$c_2 = 7.94 + \frac{2.365(2.14)}{\sqrt{8}} = 9.7$$

	a		
n	0.90	0.95	0.975
…	…	…	…
5	1.476	2.015	2.571
6	1.440	1.943	2.447
7	1.415	1.895	2.365
…	…	…	…
∞	1.282	1.645	1.960

Qual o significado?

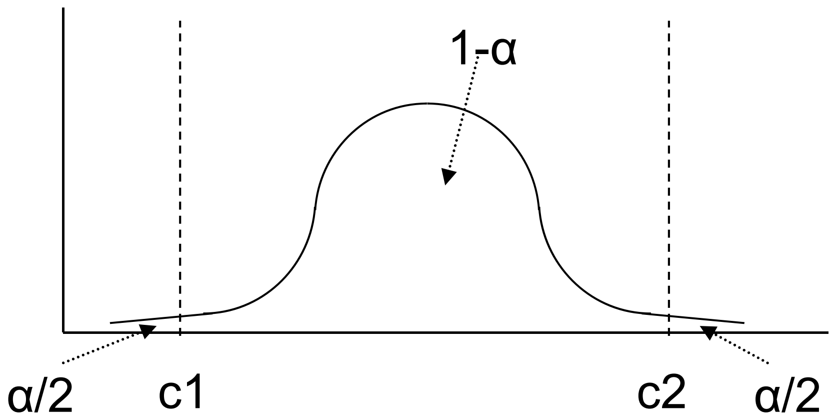
- 90% CI = [6.5, 9.4]
  - 90% probabilidade que o valor real está entre 6.5, 9.4
- 95% CI = [6.1, 9.7]
  - 95% probabilidade que o valor real está entre 6.1, 9.7
- Por que o intervalo é mais largo quando temos mais confiança?

Maior Confiança → Intervalo mais largo?



Hipótese chave

- Erros de medição são Normalmente distribuidos.
- Isso é verdade para maioria das medições em em sistemas reais de computação?



Hipótese chave

- Devido ao Teorema do Limite Central
- ***Sum of a "large number" of values from any distribution will be Normally (Gaussian) distributed.***

- O que é um "large number?"
  - Tipicamente assuma >≈ 30.

Quantas medidas?

- Largura do intervalo inversamente proporcional à $\sqrt{n}$
- Desejo de minimizar o número de medições [por que?]
- Determine um intervalo de confiança para média, tal que:
  - Pr(média real está no intervalo) = $(1 - \alpha)$

$$(c_1, c_2) = \left[(1-e)\bar{x}, (1+e)\bar{x}\right]$$

Quantas medidas?

$$(c_1, c_2) = (1 \mp e)\bar{x}$$

$$= \bar{x} \mp z_{1-\alpha/2} \frac{s}{\sqrt{n}}$$

$$z_{1-\alpha/2} \frac{s}{\sqrt{n}} = \bar{x}e$$

$$n = \left(\frac{z_{1-\alpha/2} s}{\bar{x}e} \right)^2$$

Quantas medidas?

- Mas *n* depende do conhecimento da média e desvio padrão!

- Estime *s* com pequenos número de medições

- Use esse *s* para determinar *n* necessário à obtenção da largura desejada de intervalo.

Quantas medidas?

- Média = 7.94 s
- Desvio padrão = 2.14 s
- Deseja-se intervalo de confiança de 90% dentro de 7% da média real.

Quantas medidas?

- Média = 7.94 s
- desvio padrão = 2.14 s
- Deseja-se intervalo de confiança de 90% dentro de 7% da média real.
- *$\alpha = 0.90$*
- *$(1-\alpha/2) = 0.95$*
- *Error = ± 3.5%*
- *$e = 0.035$*

Quantas medidas?

$$n = \left(\frac{z_{1-\alpha/2}s}{\overline{x}e} \right)^2 = \left(\frac{1.895(2.14)}{0.035(7.94)} \right) = 212.9$$

- 213 medições

→ 90% intervalo de confiança de 90% dentro do intervalo de +- 3.5% da média real.

Proporções

- p = Pr(success) em *n tentativas de um* experimentos binomial

- Estime a proporção: $p = m/n$
  - m = número de sucessos
  - n = número total de tentativas

Proporções

$$c_1 = \overline{p} - z_{1-\alpha/2}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

$$c_2 = \overline{p} + z_{1-\alpha/2}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

Proporções

- Quanto tempo o processador gasta executando o SO (*system time*)?

- Interrupção a cada 10 ms

- Contadores incrementais
  - n = número de interrupções
  - m = número de interrupções em PC dentro do SO

Proporções

- Quanto tempo o processador gasta executando o SO (*system time*)?
- Interrupção a cada 10 ms
- Contadores incrementais
  - n = número de interrupções
  - m = número de interrupções em PC dentro do SO
- Execute por 1 minuto
  - n = 6000
  - m = 658

Proporções

$$(c_1, c_2) = \overline{p} \mp z_{1-\alpha/2}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

$$= 0.1097 \mp 1.96\sqrt{\frac{0.1097(1-0.1097)}{6000}} = (0.1018, 0.1176)$$

- 95% de confiança para proporção
- Assim, estamos 95% de certeza (confiança) que o processador gasta 10.2-11.8% de seu tempo em SO (overhead?)

Número de Medições para Proporções

$$(1-e)\overline{p} = \overline{p} - z_{1-\alpha/2}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

$$e\overline{p} = z_{1-\alpha/2}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

$$n = \frac{z_{1-\alpha/2}^{2}\,\overline{p}(1-\overline{p})}{(e\overline{p})^{2}}$$

Número de Medições para Proporções

- Por quanto tempo deve-se executar o experimento com SO?

- Deseja-se 95% confiança
- ± 0.5%

Número de Medições para Proporções

- Por quanto tempo deve-se executar o experimento com SO?
- Deseja-se 95% confiança
- $\pm$ 0.5%
- *e = 0.005*
- *p = 0.1097*

Número de Medições para Proporções

$$n = \frac{z_{1-\alpha/2}^{2}\,\overline{p}(1-\overline{p})}{(e\overline{p})^{2}}$$

$$= \frac{(1.960)^{2}(0.1097)(1-0.1097)}{\left[0.005(0.1097)\right]^{2}}$$

$$= 1{,}247{,}102$$

- 10 ms interrupção

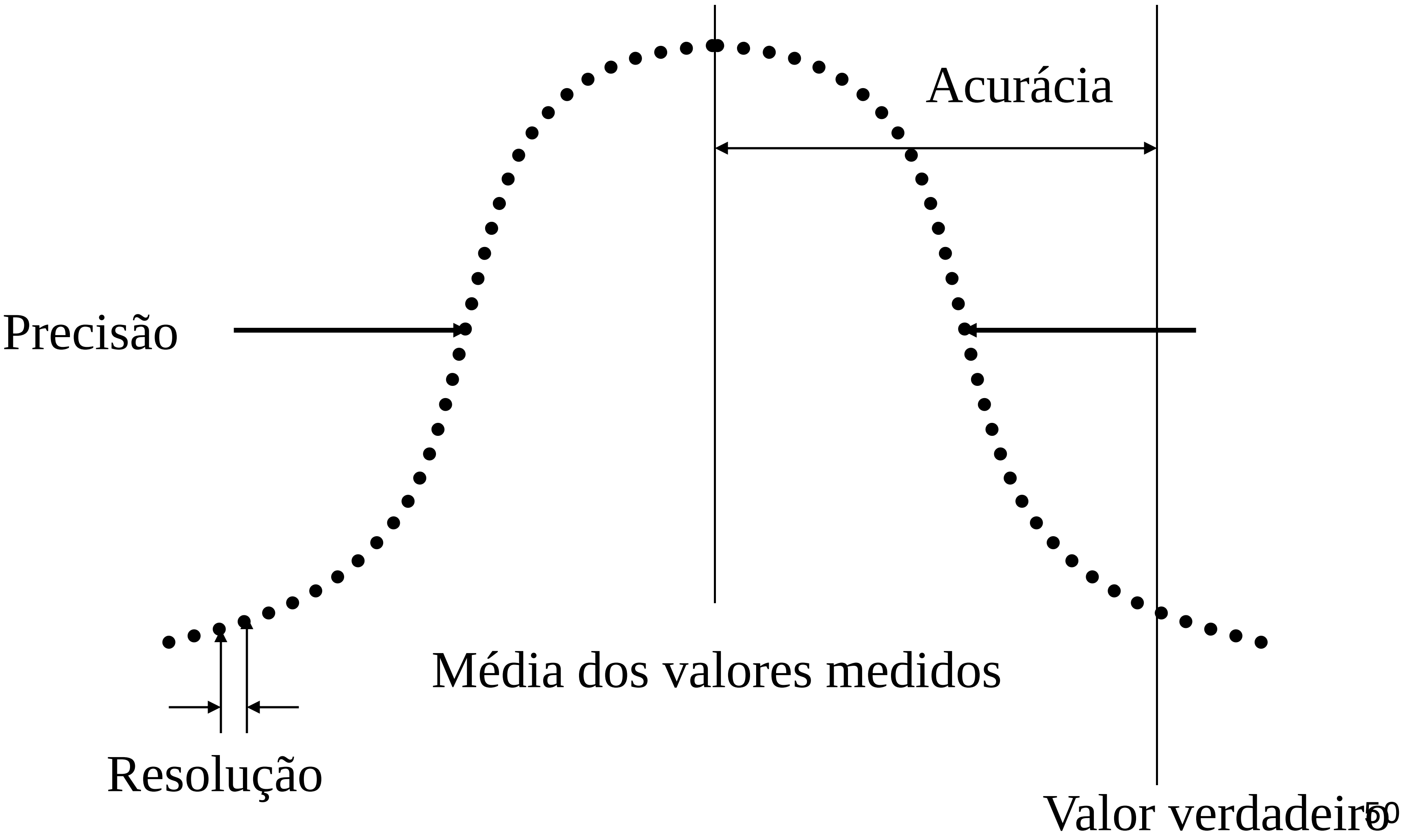
→ 3.46 horas

Pontos Conclusivos

- Use estatistica para
  - Tratar com medições com ruído;
  - Agregar grandes massas de dados
- Erros na medição são devidos a:
  - Acurácia, precisão, resolução das ferramentas
  - Outras fontes de erros

  → erros sistemáticos e randômicos

Pontos importantes: Modelo de erros com a curva da Normal



Pontos Conclusivos

- Uso de intervalos de confiança para quantificar precisão
- Intervalos de confiança para
  - Média de *n* amostras
  - Proporções
- Nível de Confiança
  - P [média verdadeira situar em um intervalo calculado]
- Computar o número de medidas necessárias para a desejada largura de um intervalo.

Exercício

- Considere um software de monitoramento que analisa o perfil de execução de programas em Linux. O experimento feito monitora o tempo de sistema operacional na execução do benchmark H. Em particular, o monitor interrompe a execução do programa H a cada 10 ms. Dois contadores são mantidos, um (c_a) para o número de interrupções e outro (c_b) para o número de vezes que a execução se encontrava em "system mode". A razão desses dois contadores representa a estimativa da fração do tempo gasto pelo sistema operacional. No primeiro teste, os contadores medidos foram: (c_{1a}) = 1300203 (c_{1b}) = 142892. Uma nova versão do Linux foi instalada e o mesmo teste foi executado, e obteve-se: (c_{2a}) = 999382 (c_{2b}) = 84876.

- Pode-se dizer com 90% de confiança que as duas versões do Linux tem performance equivalentes, no que se refere ao tempo gasto pelo sistema operacional para execução do benchmark H? Responda e mostre porque.