projecto Quark!

Escola de Física para Jovens

Análise de Dados Experimentais

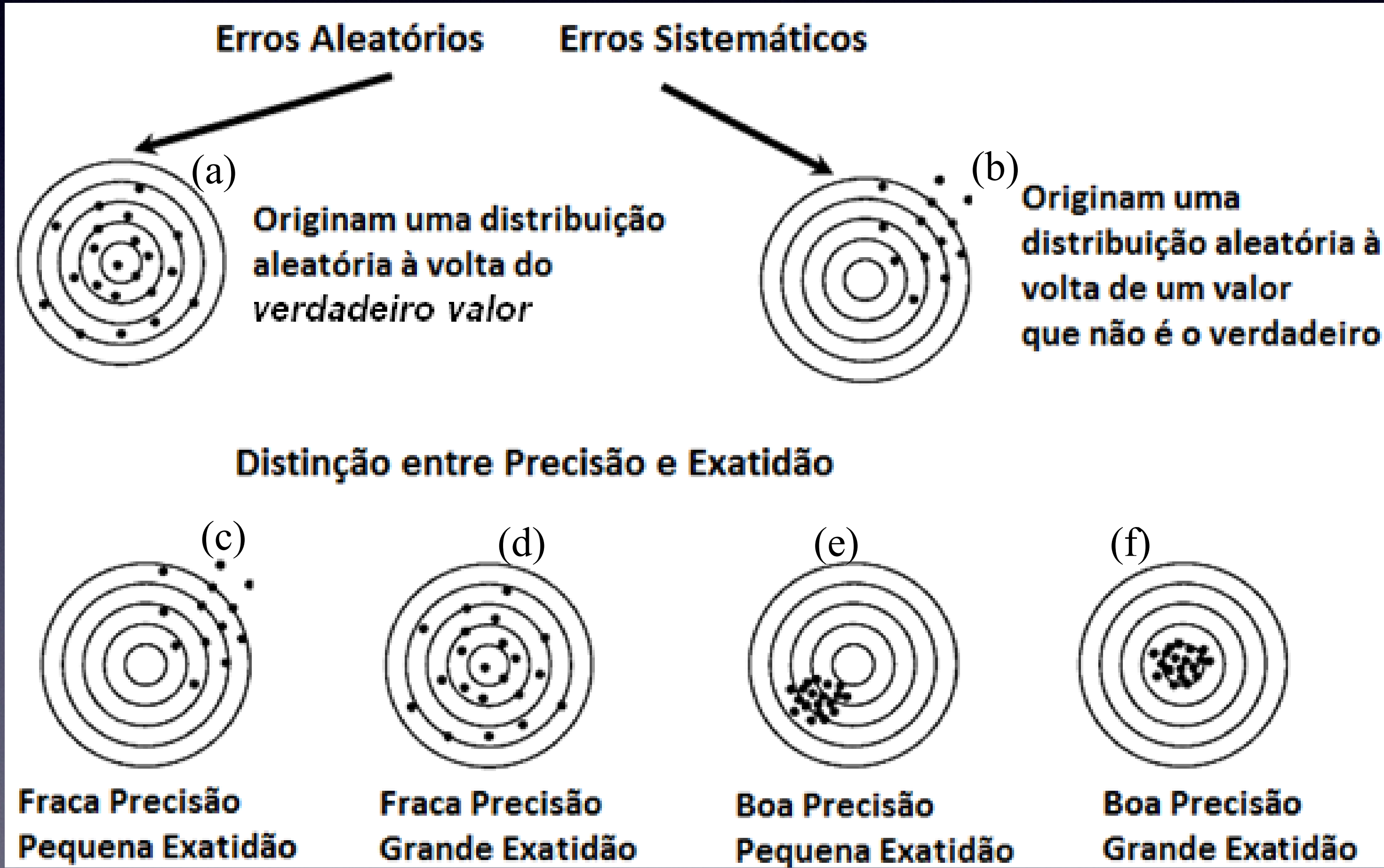
José António Paixão

· U C ·

Erros

Errare humanum est, perseverare diabolicum

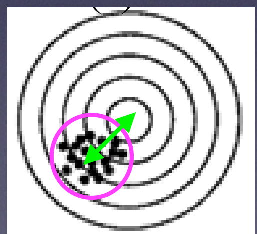
“Erros” aleatórios/sistemáticos



Exatidão/Precisão

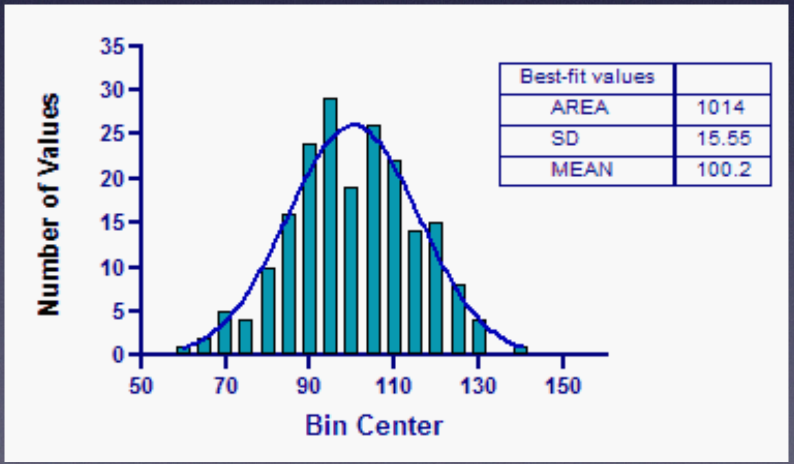
- Precisão: “largura” da distribuição dos resultados em medidas repetidas

- Exatidão: quão próximo está o “centro” da distribuição das medidas repetidas do “valor verdadeiro”



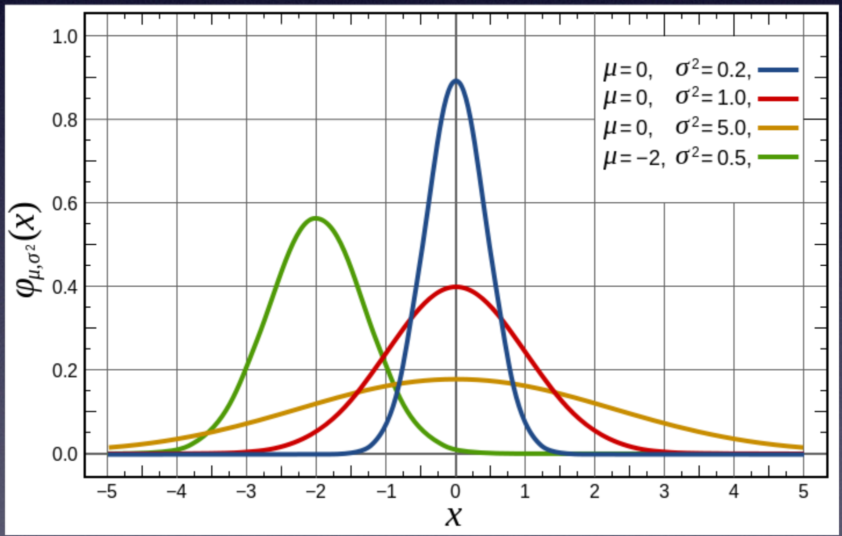
Distribuição Gaussiana

- Na sequência da análise de medidas de dados astronómicos, Gauss descobre que as medições repetidas seguem uma lei de distribuição que é conhecida por lei de Gauss:



Distribuição Gaussiana

$$\phi(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Centro da distribuição: μ

Largura da distribuição: σ

Distruibuição Gaussiana

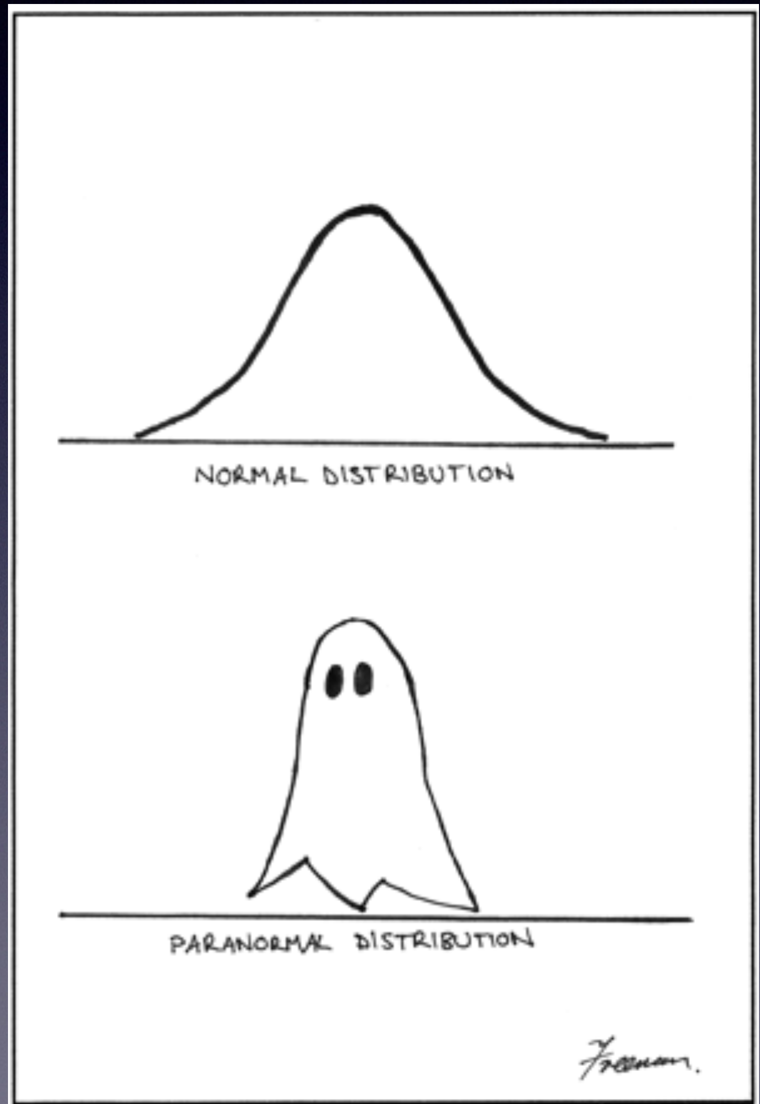
- Porque é que os dados experimentais, sujeitos a erros aleatórios seguem uma distribuição de Gauss?
- Teorema do limite central:

“toda soma de variáveis aleatórias independentes de média finita e variância limitada é aproximadamente gaussiana (normal), desde que o número de termos da soma seja suficientemente grande”

Uma recolha de dados numa experiência está sujeita a um elevado número de variáveis aleatórias (temperatura, flutuações da rede eléctrica, erros de paralaxe, etc.), sendo aplicável o teorema do limite central…

Distruibuição Gaussiana

…também conhecida por distribuição “normal”



Desvio padrão

- Dado um conjunto de *N* medidas como estimar a largura da distribuição?

$$\sigma = \lim_{N\to\infty} \sqrt{\sum_{i=1}^{N} \frac{(x_i - \bar{x})^2}{N-1}}$$

- Se o número de medidas não for pequeno, uma boa estimativa é

$$\sigma \sim \sqrt{\sum_{i=1}^{N} \frac{\left(x_i - \bar{x}\right)^2}{N}}$$

Variância

- Seja X uma variável aleatória

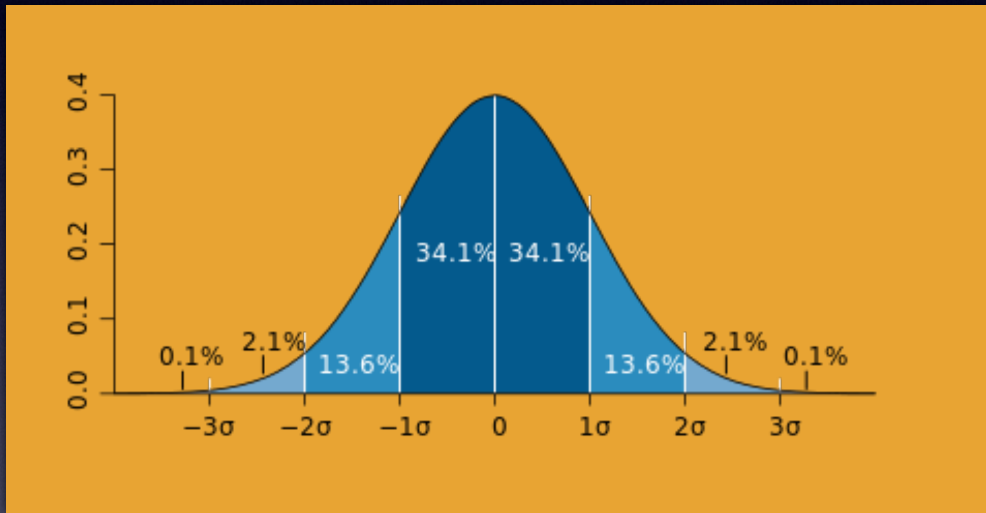
$$\mathrm{var}(X) = \sigma^2(X) = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i^2 + \bar{X}^2 - 2X_i\bar{X})$$

$$\mathrm{var}(X) = \sigma^2(X) = \bar{X}^2 - (\bar{X})^2$$

$$\mathrm{var}(X_1 + X_2) = \mathrm{var}(X_1) + \mathrm{var}(X_2)$$

$$\mathrm{var}(cX) = c^2\mathrm{var}(X)$$

Distribuição gaussiana



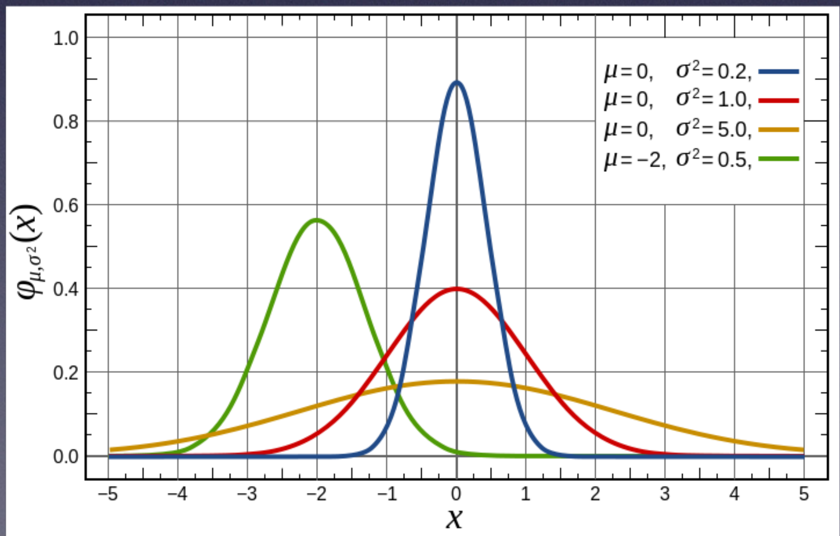
Cerca de 68% das medidas encontram-se no intervalo

$$[\mu - \sigma, \mu + \sigma]$$

O “verdadeiro” valor

- O objetivo de uma medição é estimar o “verdadeiro” valor da grandeza que estamos a medir e estimar a incerteza nesse valor.

- Se apenas estiverem presentes erros aleatórios, aceita-se que o verdadeiro valor é o máximo da distribuição de Gauss que é coincidente com a média dos valores medidos.



A incerteza na estimativa do “verdadeiro valor”

- E como estimar a incerteza no valor de μ ?

- O desvio padrão da distribuição das medidas não é uma boa estimativa da incerteza no valor de μ !

- A incerteza na estimativa do valor médio da distribuição de Gauss é inferior à largura da distribuição dos valores individuais!

Desvio padrão da média

- Fazendo *n* medidas repetidas (cada uma com *N* valores) para estimar a distribuição de μ :

$$\begin{aligned}\operatorname{var}(\mu) &= \operatorname{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\operatorname{var}\left(\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{var}(X_i) = \frac{n}{n^2}\operatorname{var}(X) = \frac{1}{n}\operatorname{var}(X).\end{aligned}$$

$$\sigma_\mu^2 = \frac{\sigma^2}{n}$$

$$\sigma_\mu = \frac{1}{\sqrt{n}}\sqrt{\sum_{i=1}^{N}\frac{(x_i-\bar{x})^2}{N}} = \frac{\sigma}{\sqrt{n}}$$

Médias pesadas

$$X = \{x_1, x_2, x_3, \ldots\}$$

$$\{\sigma_1, \sigma_2, \sigma_3, \ldots\}$$

$$\bar{X} = \frac{\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}}$$

$$\frac{1}{\sigma^2} = \frac{1}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}}$$

Erros de leitura

- Quando lidamos com instrumentos de medida há um erro mínimo associado à leitura do aparelho.
- Estes erros são especificados pelo fabricante do instrumento e, por vezes, referem-se à exatidão e não à precisão da medida…
- Quando efetuamos medidas com esses instrumentos temos de verificar se o desvio padrão das medidas não é superior ao erro de leitura do instrumento!


“Erros” aleatórios
Tipo A
Tipo B

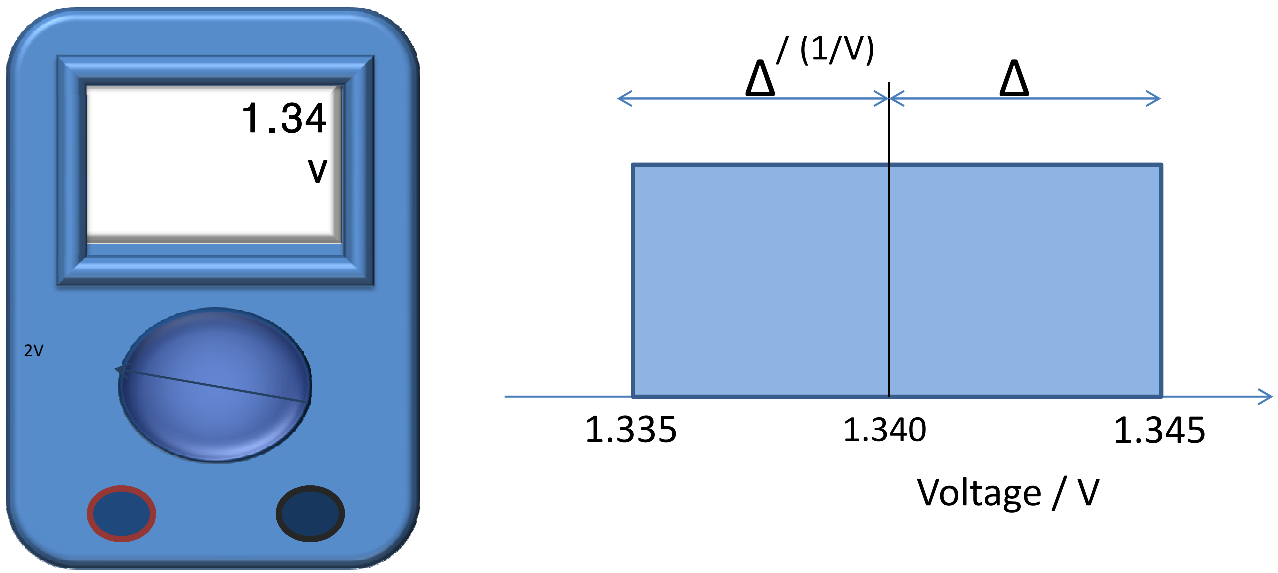
Tipo A

Avaliados estatisticamente através de medidas repetidas

Tipo B

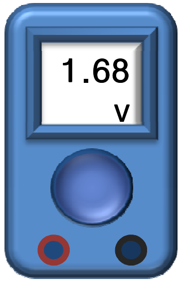
Estimados de outras formas (p.ex., através da “exatidão” dos instrumentos)

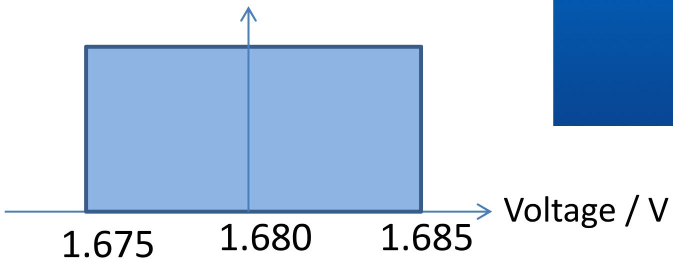
Leitura de um aparelho digital



$(1.340 \pm \varepsilon)$ V

Que incerteza (ε) devemos atribuir à leitura?



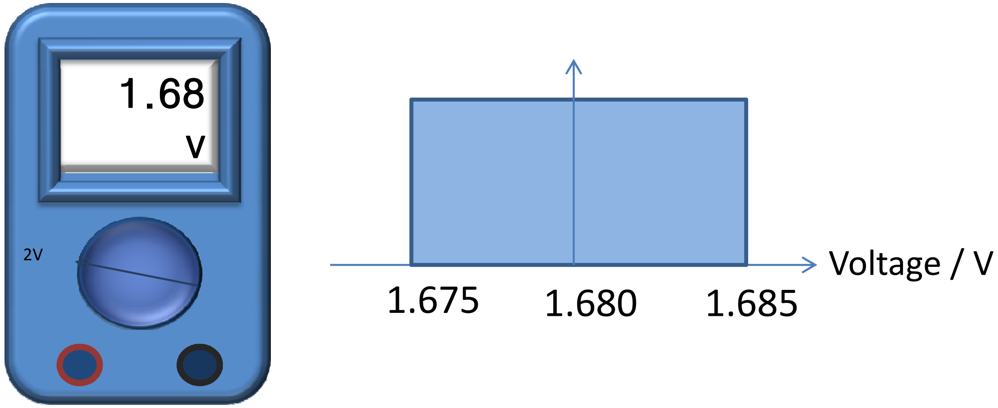


Uniform probability distribution

Length of the interval 2Δ = 1.685-1.675= 0.010 V

Half of the length Δ =0.010/2 =0.005 V

Standard uncertainty is $\pm\ \Delta/\sqrt{3} = 0.003\text{V}$

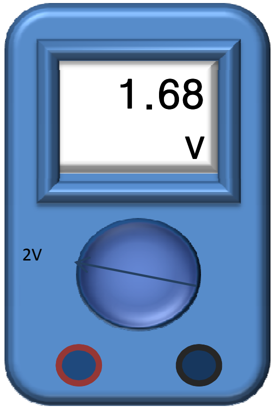


(1.680 ± 0.003) V

Uniform pdf

65% confidence

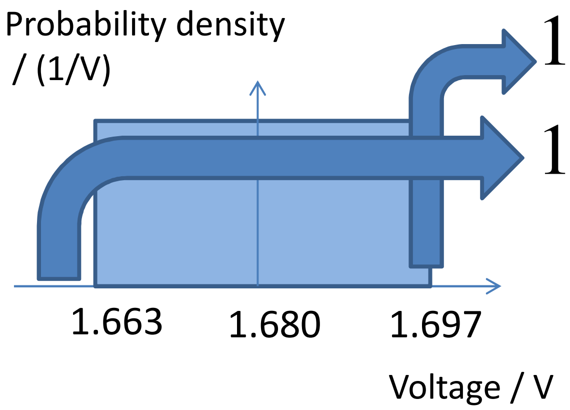
Sources of type B uncertainty



Digital resolution or scale or finite resolution

$$u_s = 0.0028\text{V}$$

Accuracy or rating ± 1%



$$1.680 + 0.01 \times 1.680 = 1.697$$

$$1.680 - 0.01 \times 1.680 = 1.663$$

$$2\Delta = 0.034$$

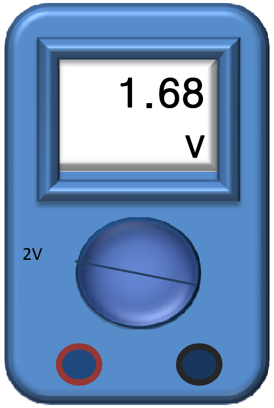
$$\Delta = 0.017$$

$$u_r = \Delta / \sqrt{3} = 0.0098\text{V}$$

Combined standard uncertainty

$$u_s = 0.0028\text{V} \qquad u_r = 0.0098\text{V}$$

$$u = \sqrt{u_s^{\,2} + u_r^{\,2}} = \sqrt{0.0028^2 + 0.0098^2} = 0.010\text{V}$$



(1.68± 0.01) V

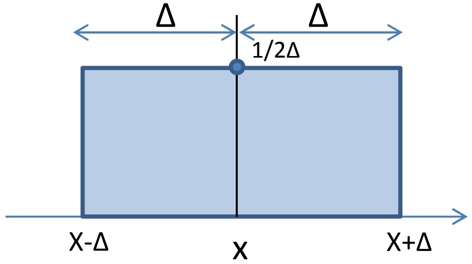
(1.680± 0.010) V

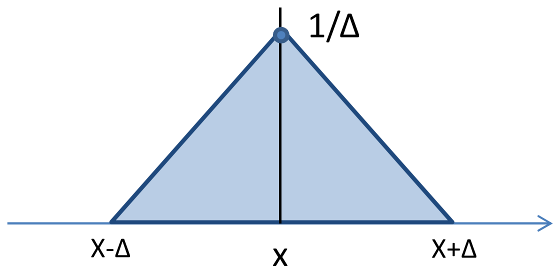
Coverage probability 68%

Common Type B Probability Distributions

Digital measuring instrument

$$u = \Delta / \sqrt{3}$$





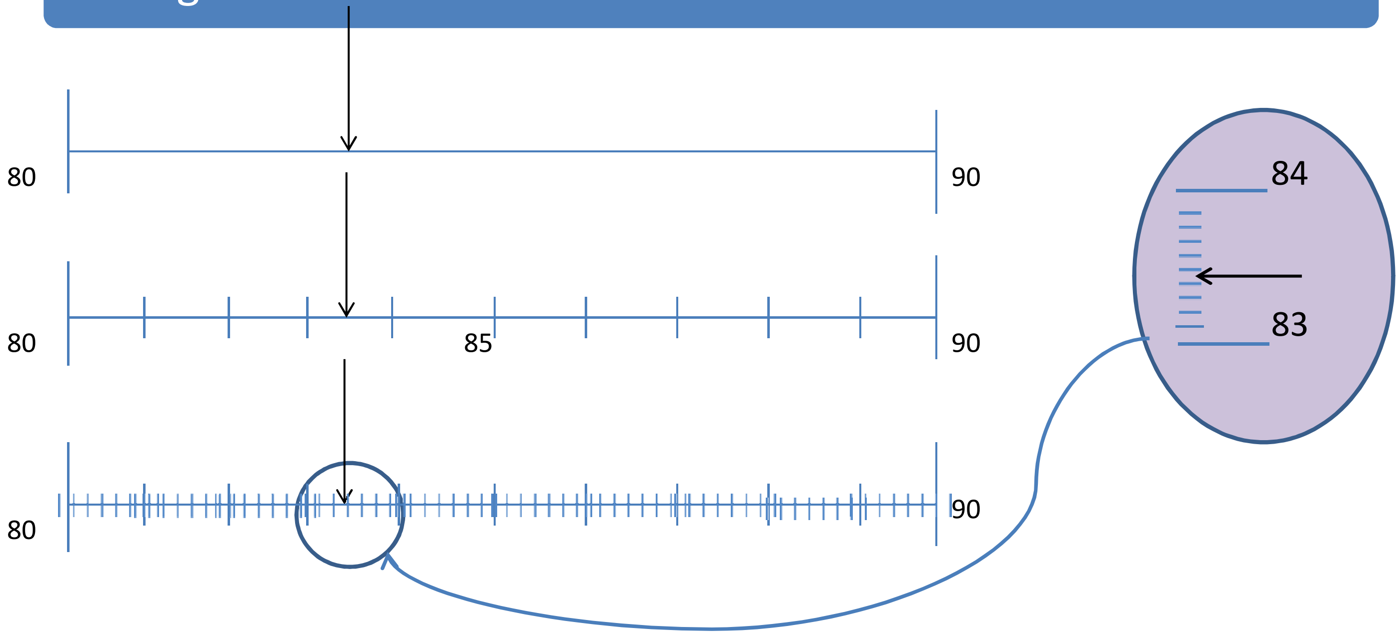
Analog instruments

$$u = \Delta / \sqrt{6}$$

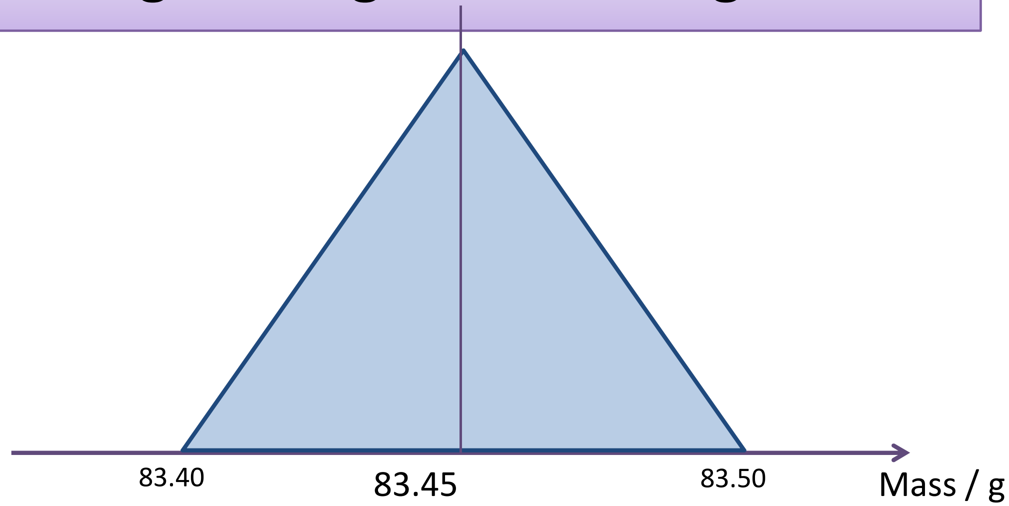
Digital scale

- 83.6 g
- 83.62 g
- 83.627 g

Analog scale



Reading 83.45 g on an analog scale



Δ = 0.05 g

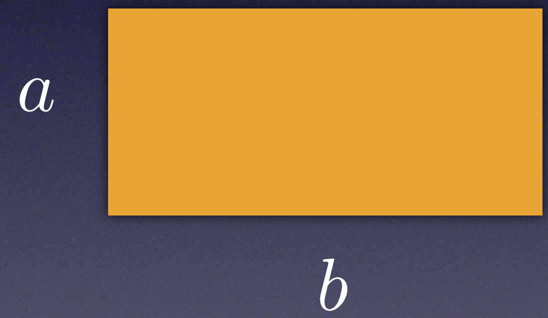
u = 0.02 g

Estimate of the measurand is (83.45 $\pm$ 0.02) g assuming a triangular PDF

Propagação de erros

Consideremos uma medida indireta, por exemplo da área *A* de um retângulo a partir das medidas (diretas) do comprimento dos lados *a* e *b*:

$$A = a \times b$$



Como é que as incertezas nos valores de *a* e *b* se propagam à estimativa do melhor valor de *A*?

Incerteza máxima

- Incerteza máxima no valor de A:

$$\sigma_A = \frac{1}{2}\left[(a+\sigma_a)\times(b+\sigma_b) - (a-\sigma_a)\times(b-\sigma_b)\right]$$

Desprezando nesta expressão os produtos $\sigma_a \sigma_b$

$$\sigma_A \sim a\,\sigma_b + b\,\sigma_a$$

ou ainda

$$\frac{\sigma_A}{A} \sim \frac{\sigma_a}{a} + \frac{\sigma_b}{b}$$

Incerteza máxima

- Será a incerteza máxima uma boa estimativa do erro associado à medida indireta da área?

- Sim, do ponto de vista **pessimista**!

- Mas o mais provável é que os erros não conspirem contra nós resultando no cenário mais desfavorável!

Incerteza mais provável

- Para o cálculo da incerteza mais provável numa medida indireta, as incertezas das medidas diretas combinam-se em “quadratura”!

$$f(x_1, x_2, ..., x_n)$$

$$\sigma_f^2 = \sum_{i=1}^{n} \left(\frac{\partial f}{x_i} \sigma_i \right)^2$$

- Exemplo:

$$A = a \times b$$

$$\sigma_A^2 = (b\sigma_a)^2 + (a\sigma_b)^2 \Leftrightarrow \left(\frac{\sigma_A}{A}\right)^2 = \left(\frac{\sigma_a}{a}\right)^2 + \left(\frac{\sigma_b}{b}\right)^2$$

"Regras" da propagação de "erros"

Caso	Função	Regra
1)	$z = ax \pm b$	$\Delta z = a\Delta x$
2)	$z = x \pm y$	$\Delta z = \left[(\Delta x)^2 + (\Delta y)^2\right]^{1/2}$
3)	$z = cxy$	$\frac{\Delta z}{z} = \left[\left(\frac{\Delta x}{x}\right)^2 + \left(\frac{\Delta y}{y}\right)^2\right]^{1/2}$
4)	$z = c(y/x)$	$\frac{\Delta z}{z} = \left[\left(\frac{\Delta x}{x}\right)^2 + \left(\frac{\Delta y}{y}\right)^2\right]^{1/2}$
5)	$z = cx^a$	$\frac{\Delta z}{z} = a\frac{\Delta x}{x}$
6)	$z = cx^a y^b$	$\frac{\Delta z}{z} = \left[\left(a\frac{\Delta x}{x}\right)^2 + \left(b\frac{\Delta y}{y}\right)^2\right]^{1/2}$
7)	$z = \sin x$	$\frac{\Delta z}{z} = \Delta x \cot x$
8)	$z = \cos x$	$\frac{\Delta z}{z} = \Delta x \tan x$
9)	$z = \tan x$	$\frac{\Delta z}{z} = \frac{\Delta x}{\sin x \cos x}$

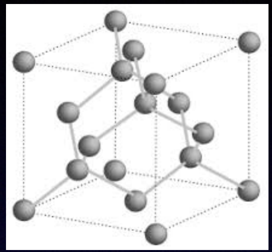
“Boas práticas”

I - Otimização da recolha de dados

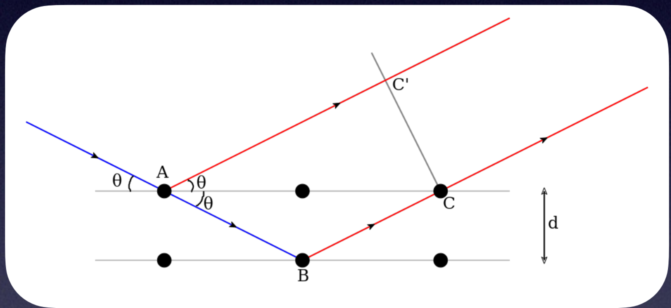
Um padrão“kg” de silício

Problema experimental: medir a distância entre os átomos do silício usando difração de RX

Lei de Bragg

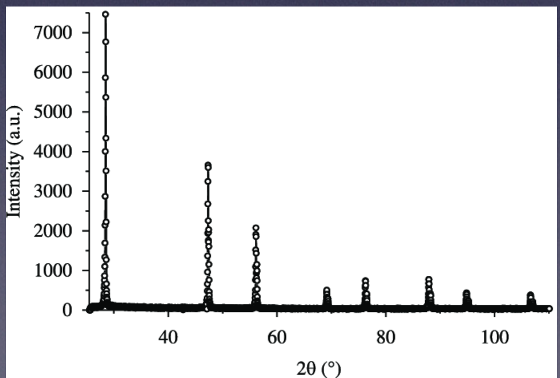


$$n\lambda = 2d\sin\theta$$



$$0 = 2(\delta d\sin\theta + d\cos\theta d\theta)$$

$$\frac{\delta d}{d} = -\cot\theta d\theta$$



$$\theta \sim \frac{\pi}{2} \rightarrow \frac{\delta d}{d} \sim 0$$

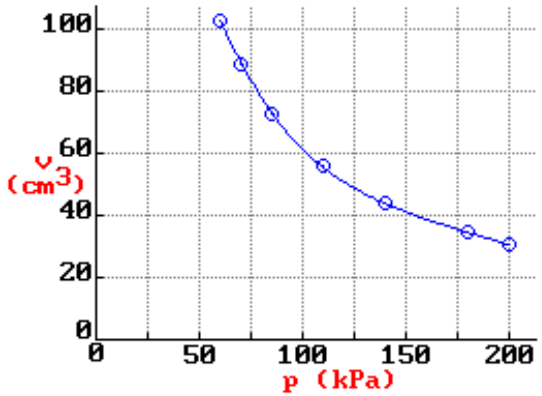
Para minimizar a incerteza devemos escolher um comprimento de onda tal que a reflexão medida tenha um ângulo de Bragg próximo de 90º.

Linearização

Lei de Boyle

$$PV = nRT$$

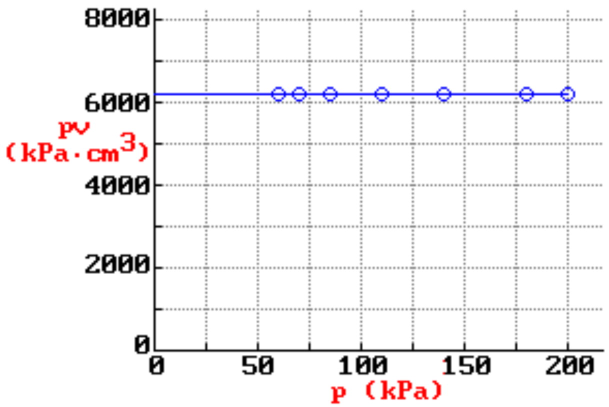
Pressure p(kPa)	Volume (cm^3)
200	31
180	34
140	44
100	62
85	73
70	88
60	103



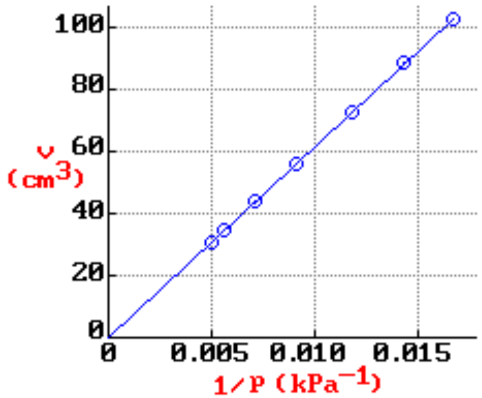
Será que a função representada no gráfico é mesmo uma hipérbole?

Lei de Boyle

Pressure p(kPa)	Volume (cm^3)	pv ($kPa \cdot cm^3$)
200	31	$6.2x10^3$
180	34	$6.2x10^3$
140	44	$6.2x10^3$
100	62	$6.2x10^3$
85	73	$6.2x10^3$
70	88	$6.2x10^3$
60	103	$6.2x10^3$

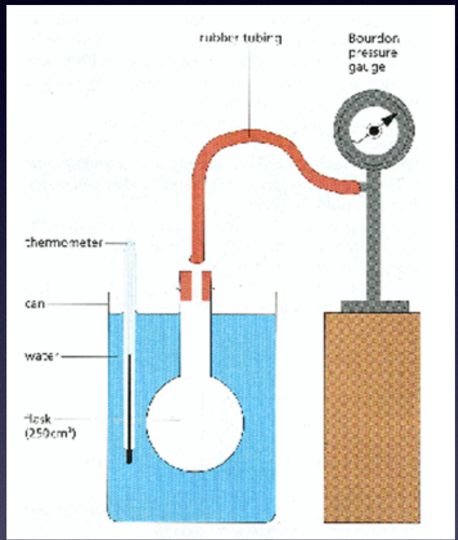


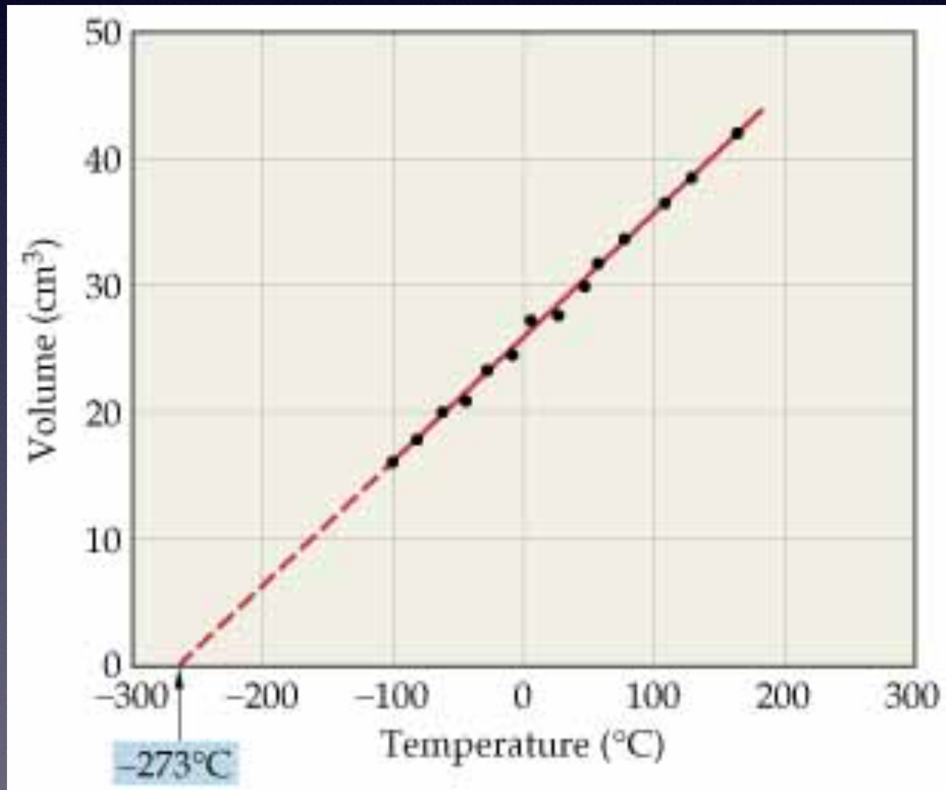
Pressure p(kPa)	Volume (cm^3)	1/p (kPa^{-1})
200	31	0.0050
180	34	0.0056
140	44	0.0071
100	62	0.0091
85	73	0.0118
70	88	0.0143
60	103	0.0167



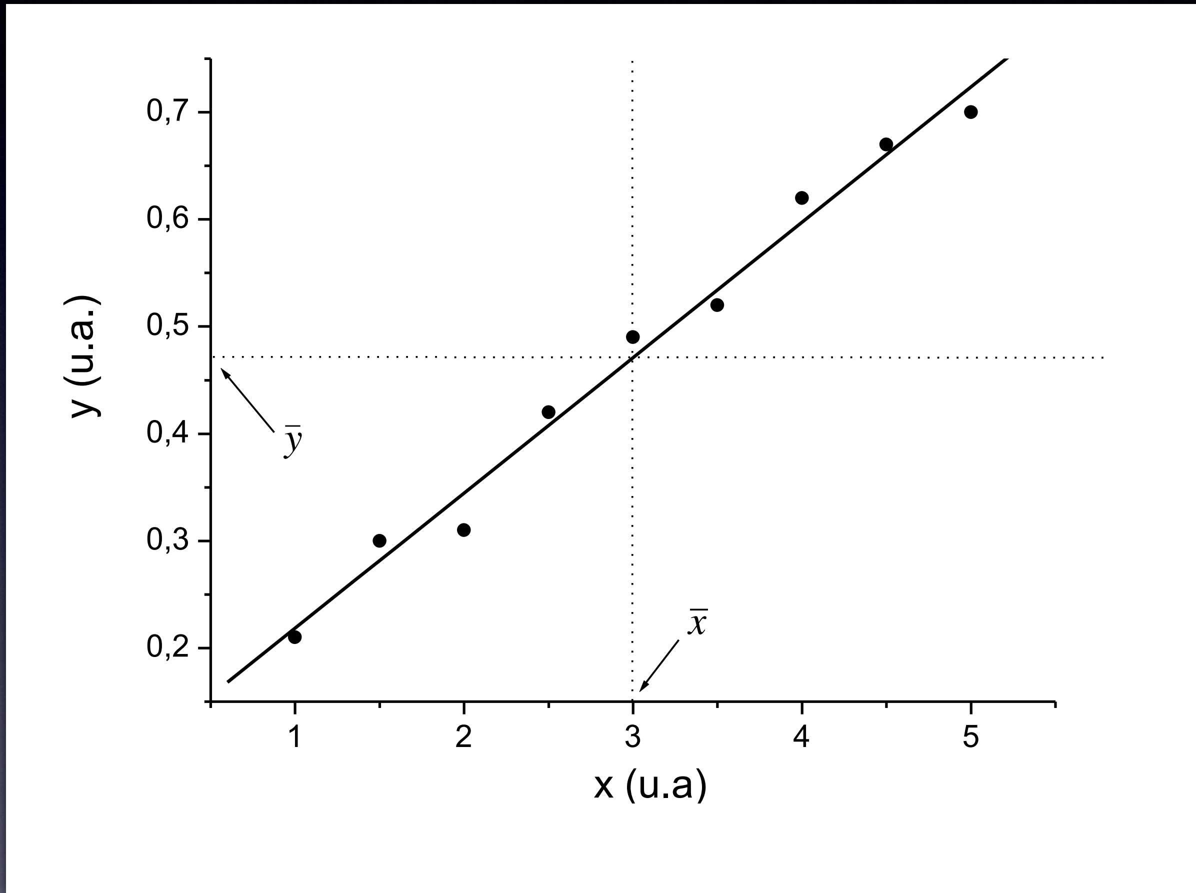
A representação de V em função de $1/P$ mostra claramente a (esperada) relação linear, demonstrando que o gás segue a lei de Boyle, de forma mais convincente que o gráfico de V em função de P!”

Lei de Gay-Lussac

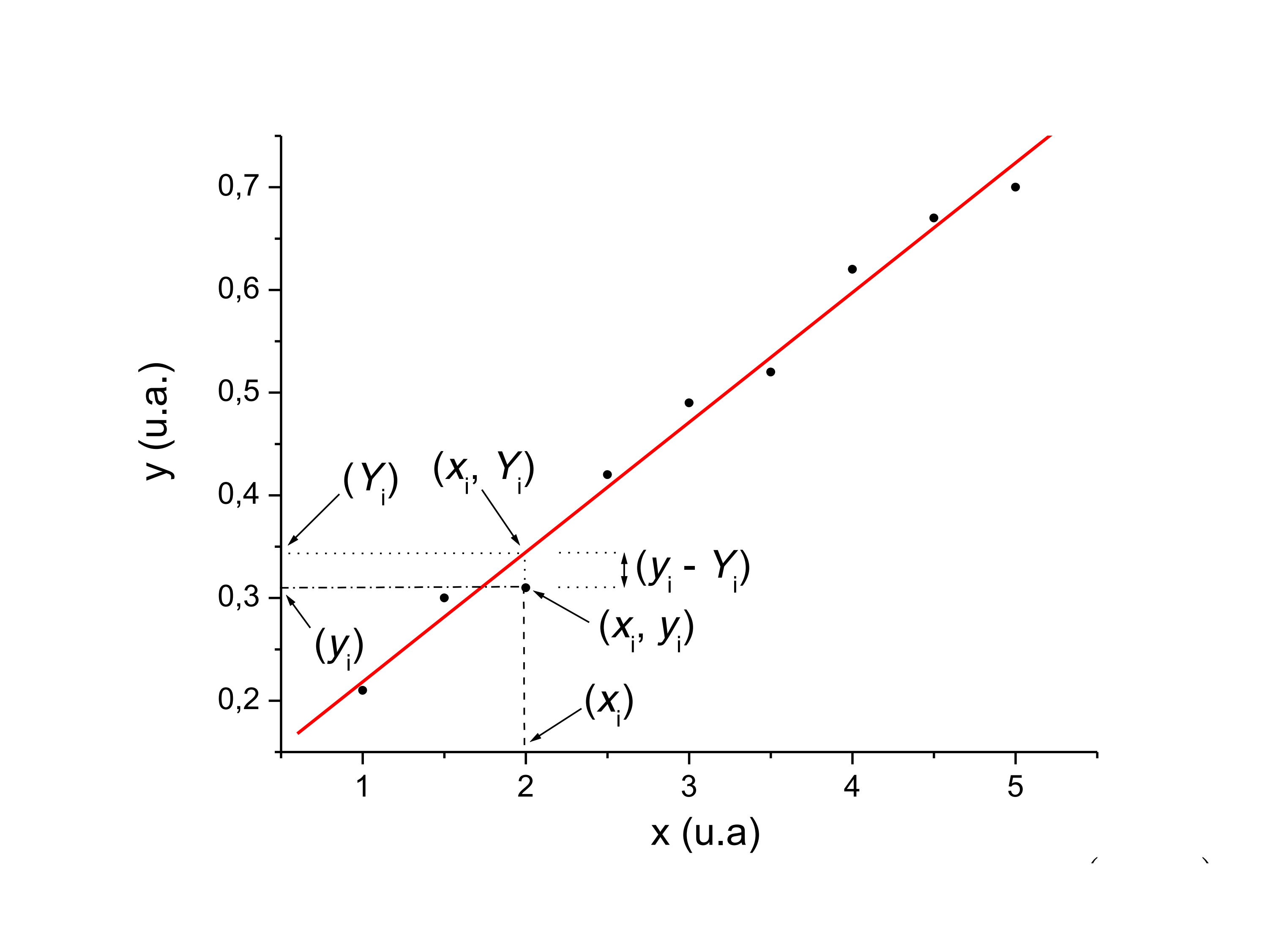




Ajustes lineares

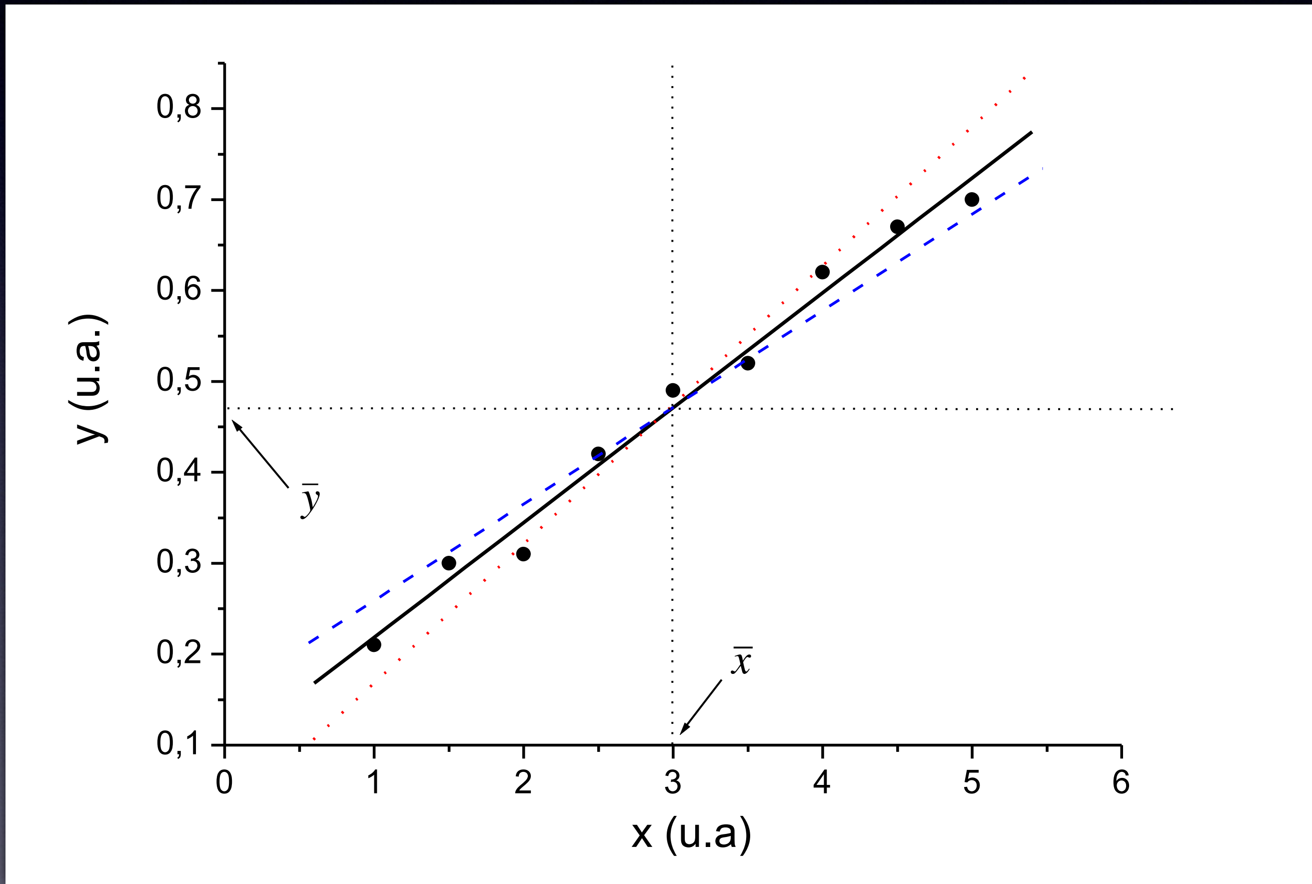


Ajustes lineares



Ajustes lineares

Método manual (gráfico):



$$\sigma_A = \frac{\frac{|A_{max} - A_{min}|}{2}}{\sqrt{N}} \text{ e } \sigma_B = \frac{\frac{|B_{max} - B_{min}|}{2}}{\sqrt{N}}$$

Regressão linear

A "melhor" reta que se ajusta aos dados experimentais também pode ser obtida por "regressão linear" ou "ajuste por mínimos quadrados" que corresponde a determinar a reta que minimiza a soma (pesada às incertezas) do quadrado dos desvios dos dados em relação ao modelo linear:

$$y = A + Bx$$

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - A - Bx_i)^2}{\sigma_y^2}$$

Função a minimizar

Condições de minimização:

$$\frac{\partial\chi^2}{\partial A} = \frac{-2}{\sigma_y^2}\sum_{i=1}^{N}(y_i - A - Bx_i) = 0, \qquad AN + B\sum x_i = \sum y_i,$$

$$\frac{\partial\chi^2}{\partial B} = \frac{-2}{\sigma_y^2}\sum_{i=1}^{N}x_i(y_i - A - Bx_i) = 0, \qquad A\sum x_i + B\sum x_i^2 = \sum x_i y_i$$

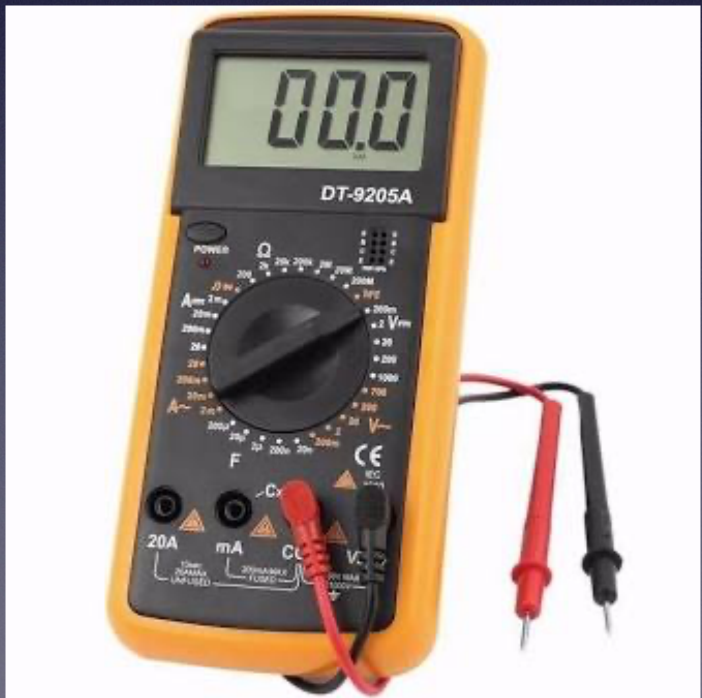
$$A = \frac{\sum_i x_i^2 \sum_i y_i - \sum_i x_i \sum_i x_i y_i}{\Delta} \qquad B = \frac{N\sum_i x_i y_i - \sum_i x_i \sum_i y_i}{\Delta} \qquad \Delta = N\sum_i x_i^2 - \left(\sum_i x_i\right)^2$$

$$\sigma_A = \sigma_y\sqrt{\frac{\sum_i x_i^2}{\Delta}} \qquad \sigma_B = \sigma_y\sqrt{\frac{N}{\Delta}}$$

O cálculo de *A* e *B* (e, eventualmente de σ_A e σ_B) estão pré-programados na maioria das calculadoras científicas (função regressão linear) e em folhas de cálculo (Excel)

IPhO

"Candidates must be aware that instruments affect measurements"



±0.9% + 1 dig.

10.00 V ± ?

10 V ± (0.09+0.01)

10 V ± (0.1)

Regra de ouro

- Usem o vosso senso comum!