# Analisi dati in Fisica

XIII Scuola Estiva di Fisica - UNISALENTO - 5 settembre 2023

# John Robert Taylor : introduzione all'analisi degli errori

ERRORI → SBAGLI ?

ERRORI → INCERTEZZE !

- INEVITABILI
- COSA POSSIAMO CERCARE DI FARE? → ridurli

  → STIMARLI

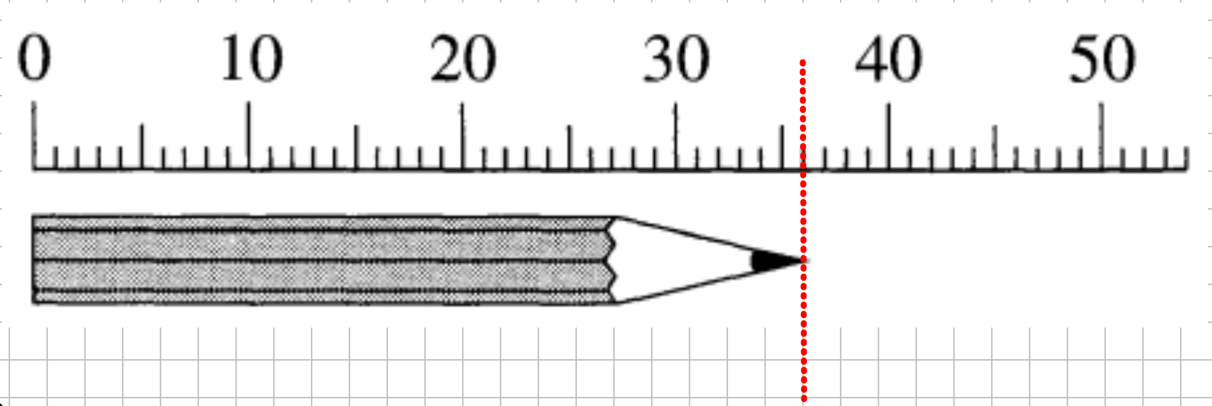
IMPORTANTE CONOSCERLI!

**Nelle scienze applicate**

**Nelle scienze di base**

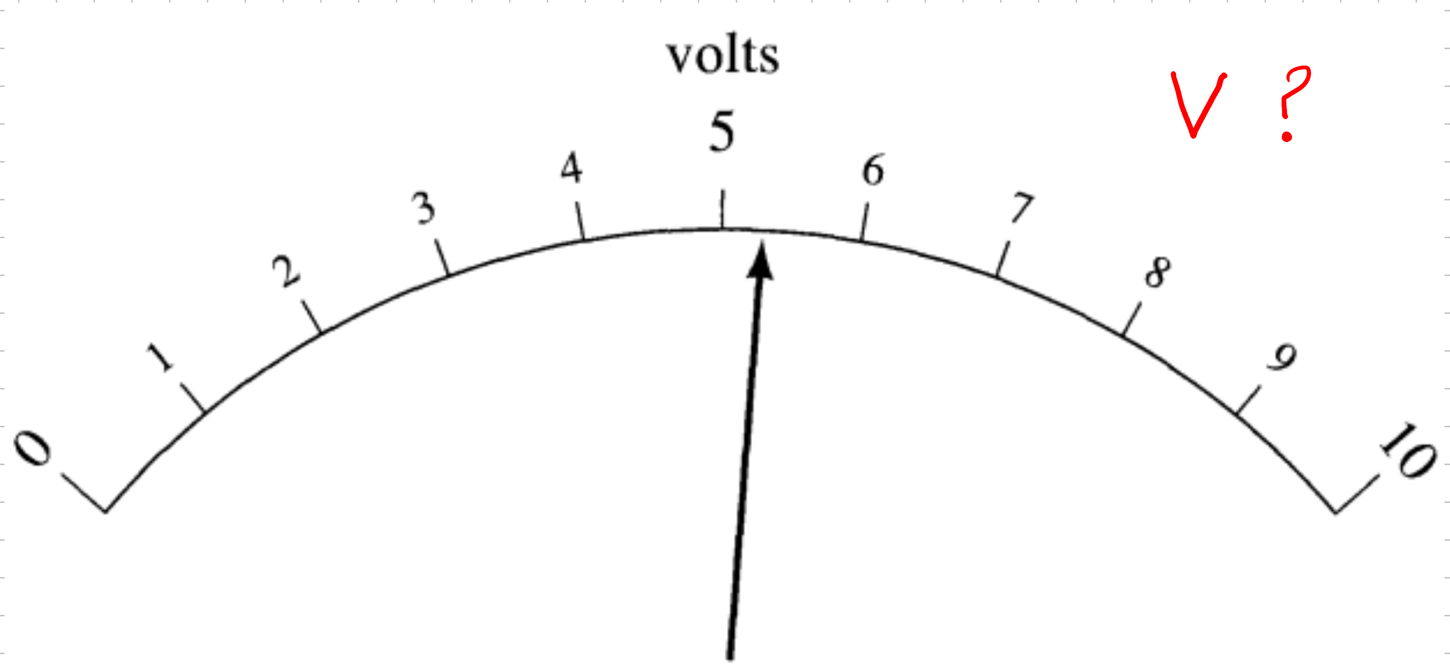
ad esempio quando si propone una nuova teoria che deve essere "validata" sperimentalmente. Le misure devono consentire di decidere tra 2 o più teorie.

# LA STIMA DEGLI ERRORI NELLE LETTURE DI SCALA




(mm)

$\ell$ ?




V ?

Anche se diversi osservatori possono non essere d'accordo con le valutazioni date, non si può negare che sono stime ragionevoli delle grandezze e delle incertezze (se legate solo alla lettura).

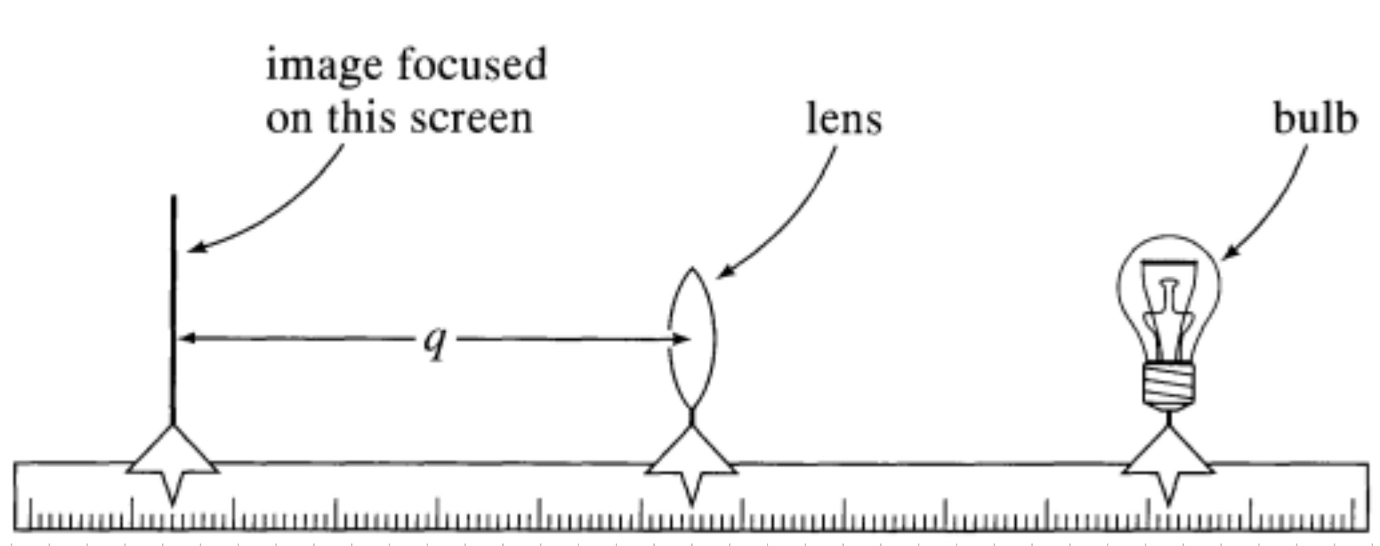

HT
HT63
AutoHiRes MIN
AC
240.36
60000
RANGE
HIRES
Hz%
REL Δ
AUTO
HOLD
MODE
MAX MIN
PEAK
OFF
TRMS
AC+DC
CAT IV 600V
CAT III 1000V
FUSED
COM

V = ?

## Problema di definizione

Di solito in una misura, oltre all'incertezza dovuta alla lettura su una scala, ci sono altre sorgenti di errore, da individuare e stimare per non sottostimare l'incertezza.

TROVARE TUTTE LE POSSIBILI CAUSE DI ERRORE E STIMARLI



# LA STIMA DEGLI ERRORI NELLE MISURE RIPETUTE

- INTANTO BISOGNA ASSICURARSI di poter ripetere la misura della stessa grandezza
- se si può, allora si può fare un approccio statistico (media, deviazione standard ...)
- ma ci possono essere errori che misure ripetute non evidenziano ---
  ?

RAPPRESENTARE

# Come rappresentare e utilizzare gli errori

→ stima migliore ± errore $\qquad T = (2{,}4 \pm 0{,}1)\ s$

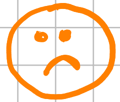
$$\text{misura di } x = x_{best} \pm \delta x$$

cioè siamo ragionevolmente certi che la quantità in questione si trova in quell'intervallo.

70%

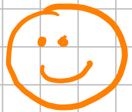
## → cifre significative

$$9{,}82 \pm 0{,}02385 \ m/s^2$$



**REGOLA D'ORO**: gli errori sperimentali dovrebbero essere arrotondati ad 1 cifra significativa

$$9{,}82 \pm 0{,}02 \ m/s^2$$



con piccole eccezioni: $\delta x = 0{,}14$ le lascio così

$\delta x = 0{,}24$ pure

$6051,78 \pm 30$ m/s

2 8

$6050 \pm 30$ m/s

DA CUI ALTRA REGOLA D'ORO: l'ultima cifra significativa in un risultato dovrebbe essere dello stesso ordine di grandezza dell'errore

$92,8 \pm 0,3$ $93 \pm 3$ $90 \pm 30$

# ALTRE Regole d'oro

* evita $1,61 \cdot 10^{-19}\, C \pm 5 \cdot 10^{-21}\, C$

meglio scrivere ---- ?

* Nei calcoli, i numeri usati dovrebbero essere tenuti con una cifra significativa in più e solo alla fine si esegue l'arrotondamento.

UTILIZZARE

La maggior parte delle esperienze sono <u>quantitative</u>. E se vogliamo che siano di qualche interesse si devono trarre conclusioni facendo ad esempio

- misura / misura
- misura / valore accettato

confronto

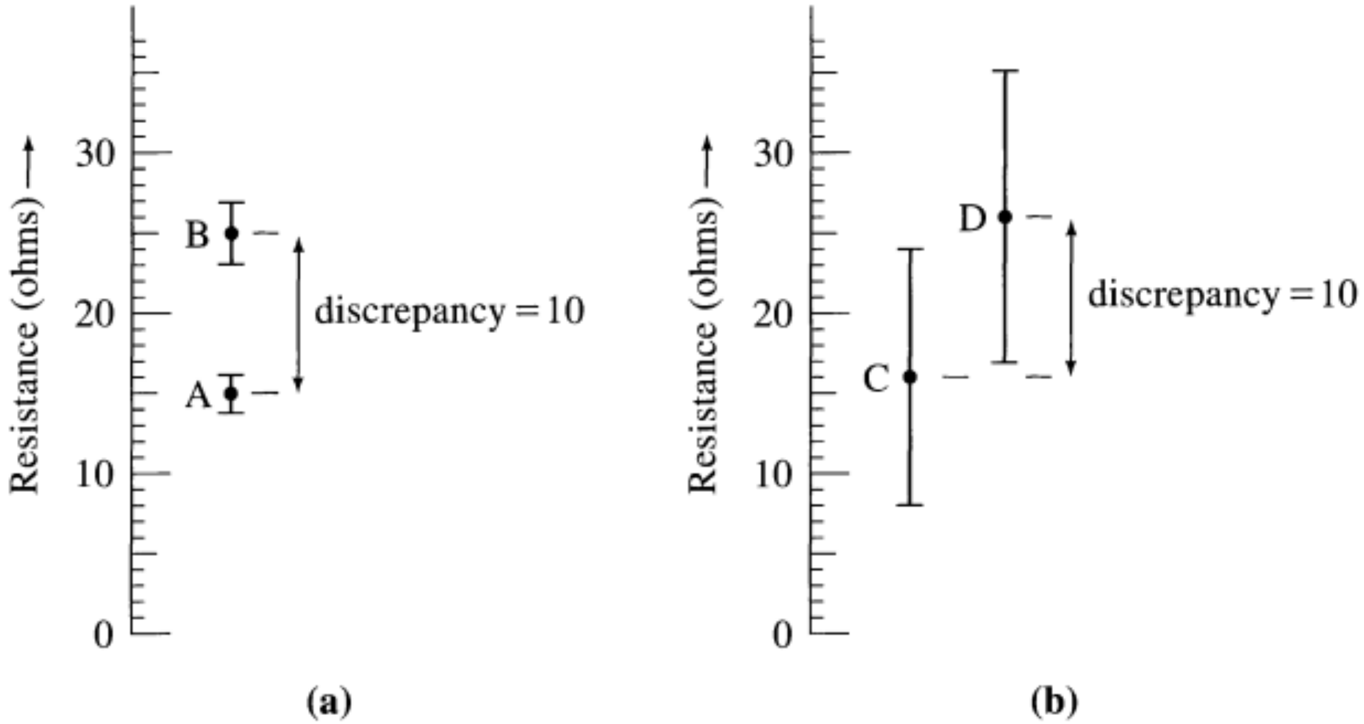
- verifica grafica della proporzionalità tra 2 grandezze

# CONFRONTO DI DUE MISURE (della stessa grandezza)

non compatibili                                    Compatibili

$25 \pm 2\ \Omega$

$15 \pm 2\ \Omega$




$26 \pm 9\ \Omega$

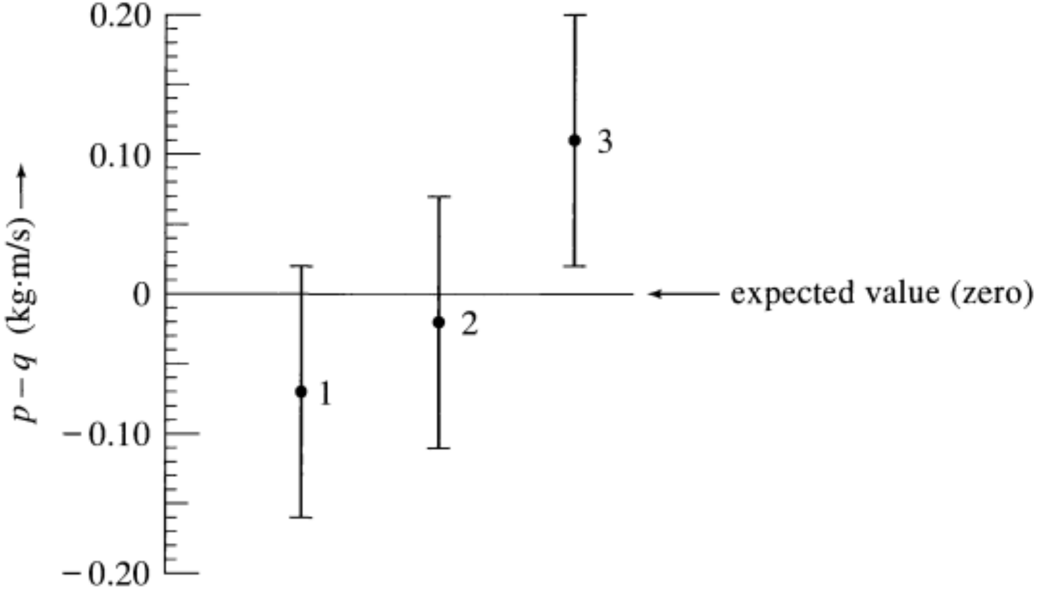
$21 \pm 8\ \Omega$

CONFRONTO DI DUE MISURE (una serie ...)

**Table 2.1.** Measured momenta (kg·m/s).

| Trial number | Initial momentum $p$ (all $\pm 0.03$) | Final momentum $q$ (all $\pm 0.06$) |
|---|---|---|
| 1 | 1.49 | 1.56 |
| 2 | 3.10 | 3.12 |
| 3 | 2.16 | 2.05 |
| etc. | | |

per facilitare la conclusione :




$\pm 0{,}09$ kg m/s

# CONFRONTO DI VALORI MISURATI E ACCETTATI

A volte si misura (laboratori didattici) una grandezza il cui valore accettato è noto.

↳ misure molto accurate, note a tutti.

$$c = 299792458 \text{ m/s}$$

certo noi non otterremo la stessa precisione

331 m/s

studente 1

329 ± 5 m/s

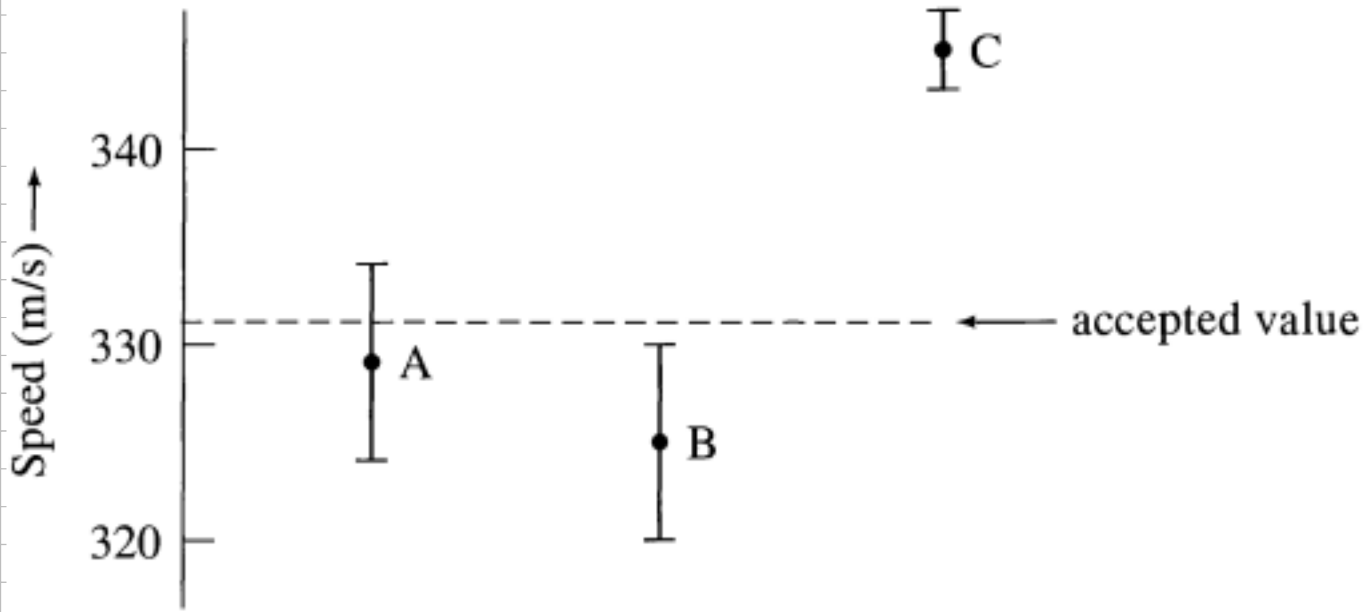
studente 2

345 ± 2 m/s

errore di calcolo? di lettura? di stima dell'errore? CHECK

sistematico?

# CONFRONTO DI VALORI MISURATI ED ACCETTATI



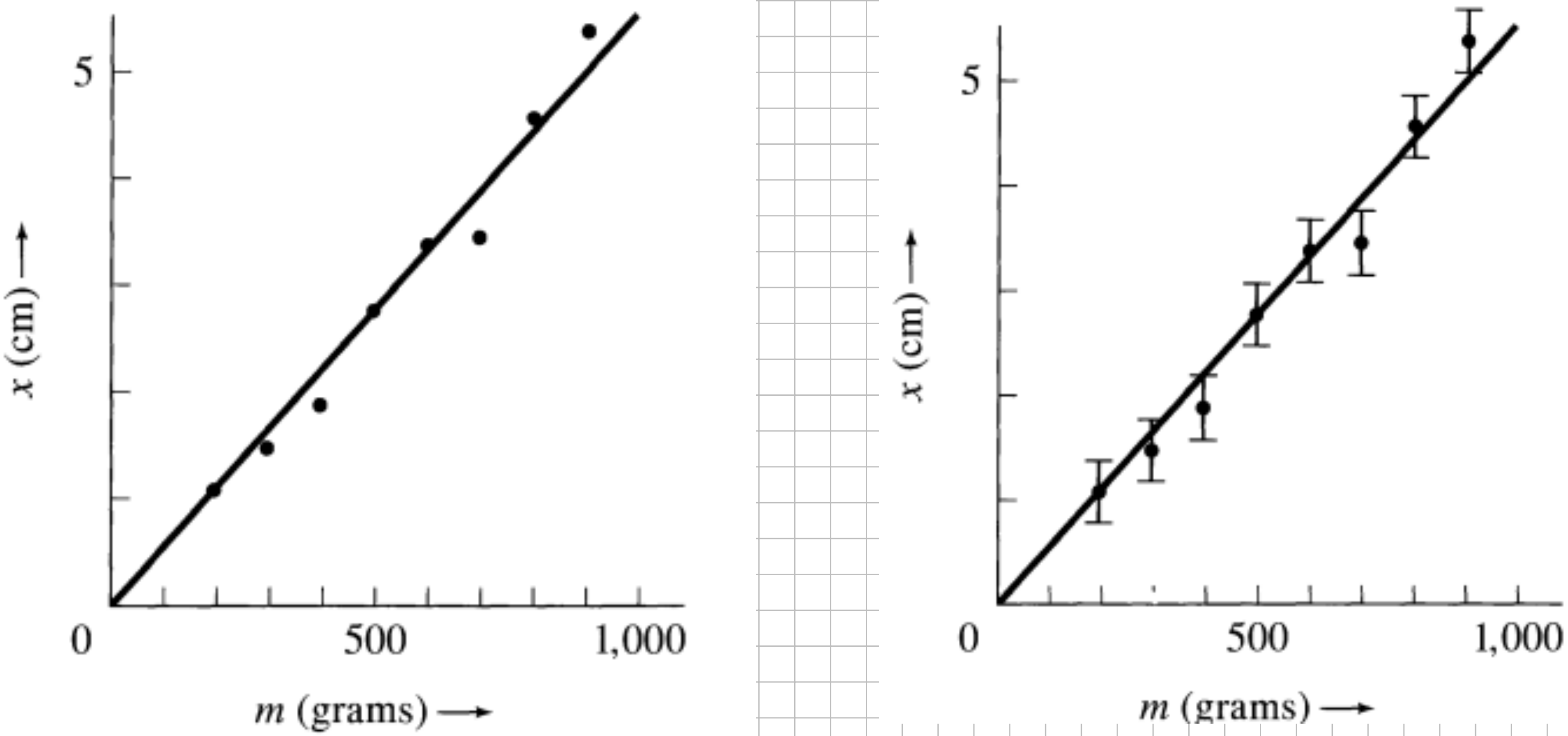
verifica grafica della proporzionalità

Tra 2 grandezze

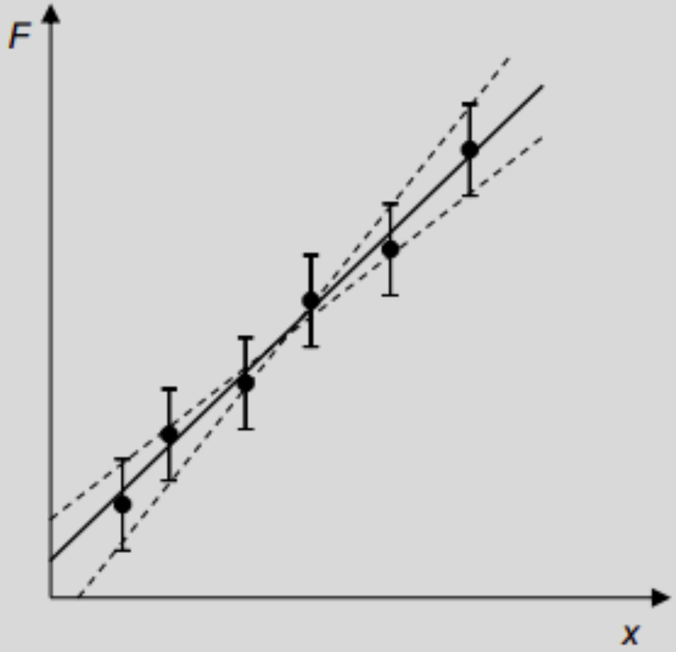
$$x = \left(\frac{g}{k}\right) m$$

legge di Hooke

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Load $m$ (grams) ($\delta m$ negligible) | 200 | 300 | 400 | 500 | 600 | 700 | 800 | 900 |
| Extension $x$ (cm) (all $\pm 0.3$) | 1.1 | 1.5 | 1.9 | 2.8 | 3.4 | 3.5 | 4.6 | 5.4 |



5
x (cm)
0
500
1,000
m (grams)
5
x (cm)
0
500
1,000
m (grams)

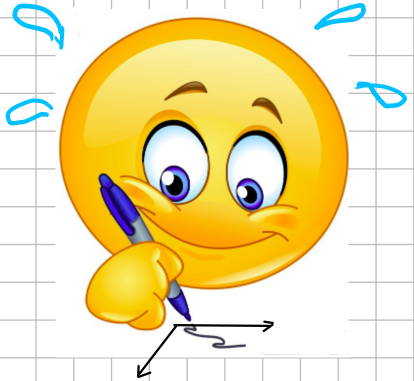



The gradients of each line were computed:

- Line of best fit: 4.7 N $m^{-1}$
- Line of worst fit (max): 4.9 N $m^{-1}$
- Line of worst fit (min): 4.4 N $m^{-1}$

Calculate the value for the spring constant *k* and its uncertainty.

# Accorgimenti per i grafici a mano



- usa la matita
- assi con unità di misura e simbolo
- di solito


- metti delle "tacchette" lungo gli assi con dei valori
- titolo
- sfrutta lo spazio: prendi le misure!
- non collegare i punti sperimentali con delle linee
- se visibili, inserisci le barre degli errori.

## Se la legge è lineare?

$$y = Ax + B$$

. Si procede come prima

$$A = A_{best} \pm \delta_A$$

pendenza della retta best fit

pendenza delle rette max e minima

$$\delta A = \frac{A_{max} - A_{min}}{2}$$

$$B = B_{best} \pm \delta_B$$

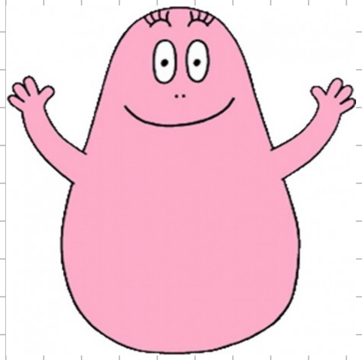
intersezione con asse y della retta best fit

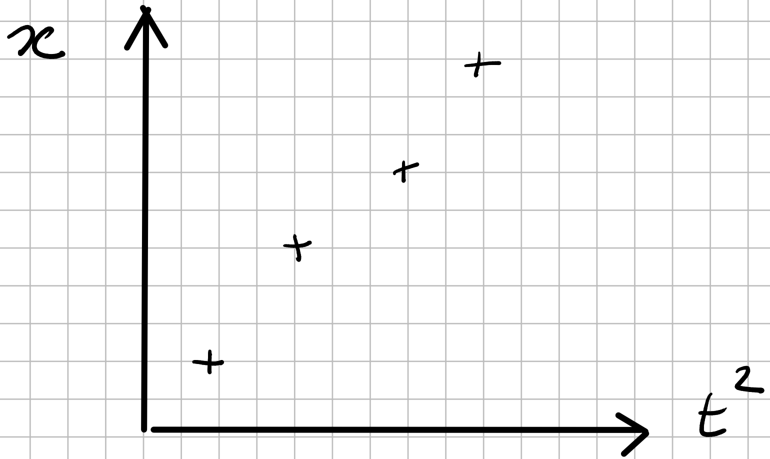
- - - -

# Non tutto è lineare!

Esempio $\qquad x = \frac{1}{2} g t^2$

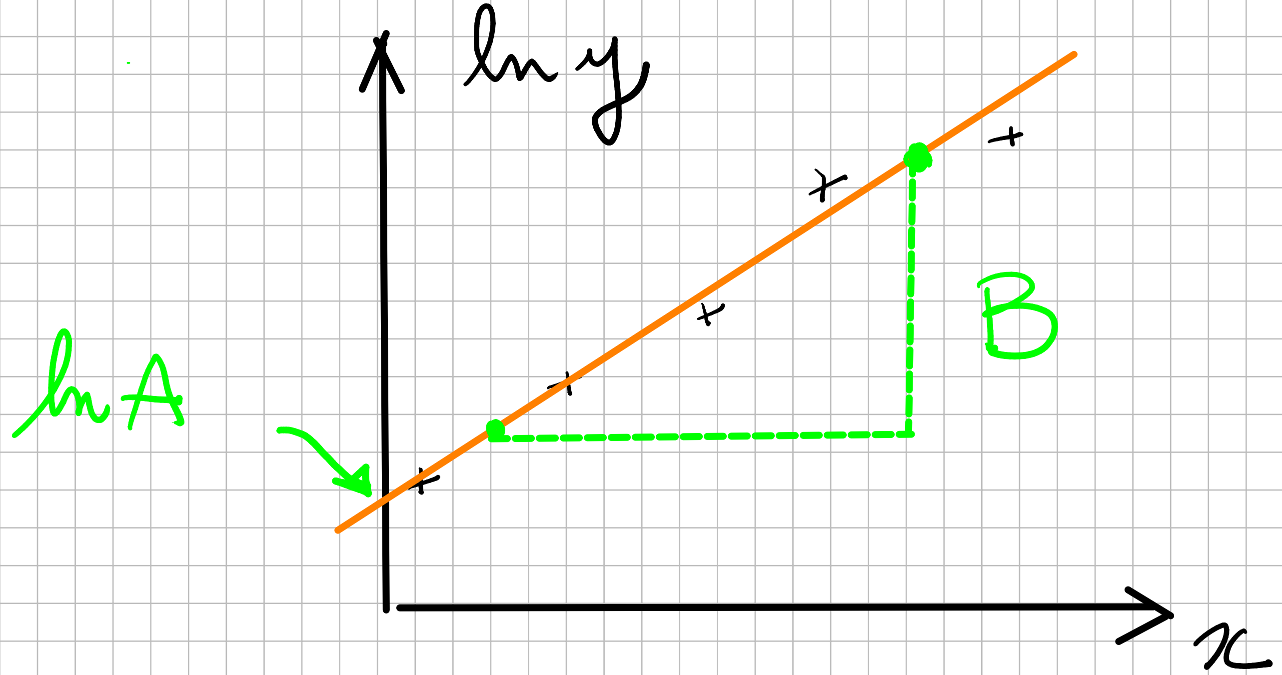
i punti sperimentali dovrebbero giacere su una parabola. Difficile da verificare.





$$y = Ae^{Bx} \rightarrow \ln y = \ln A + Bx$$

$A > 0$



# PROPAGAZIONE DEGLI ERRORI nelle misure indirette

$$\varepsilon_x = \frac{\delta x}{|x|} \qquad \varepsilon_{x\%} = \frac{\delta x}{|x|} \cdot 100$$

Somme e differenze :

$$q = x + \cdots + z - (u + \cdots + w) \longrightarrow \delta q \approx \delta x + \cdots + \delta z + \delta u + \cdots + \delta w$$

prodotti e quozienti :

$$q = \frac{x \cdots z}{u \cdots w} \longrightarrow \varepsilon_q = \varepsilon_x + \cdots + \varepsilon_z + \varepsilon_u + \cdots + \varepsilon_w$$

## Prodotto di una grandezza misurata per un numero esatto

$$q = Bx \longrightarrow \delta q = |B|\,\delta x$$

es: quanto vale lo spessore di un foglio se la misura di 100 di essi mi fornisce $S = 3,3 \pm 0,1$ cm ?

## Potenze

$$q = x^n \longrightarrow \varepsilon_q = n\,\varepsilon_x$$

es: $h = \frac{1}{2} g t^2$ $\quad h = 14.1 \pm 0.1$ m $\quad t = 1.6 \pm 0.1$ s

$g$ ?

$$g = \frac{2h}{t^2} \longrightarrow \varepsilon_g \approx \varepsilon_h + 2\,\varepsilon_t$$

$$g_{best} = 11.0\ \frac{m}{s^2}$$

$$\varepsilon_g \approx \frac{0.1}{1.6}\cdot 100 + 2\left(\frac{0,1}{14.1}\right)$$

$$\approx 0{,}7\% + 2\cdot(6.3\%) \approx 13{,}3\%$$

$$g = (11.0 \pm 1.4)\ \frac{m}{s^2}$$

sappiamo anche come migliorare.

# ERRORI INDIPENDENTI IN UNA SOMMA

$$q = x + y \longrightarrow \delta q \leq \delta x + \delta y$$

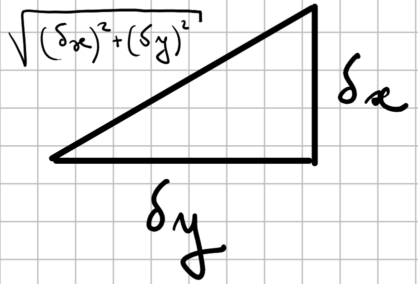
$q_{best} - (\delta x + \delta y)$ $\qquad$ $q_{best}$ $\qquad$ $q_{best} + (\delta x + \delta y)$

Come potrebbe accadere che il "valore reale" sia qui?

→ se ho sottostimato $x$ di $\delta x$
e $y$ di $\delta y$ → Poco probabile se $\delta x$ e $\delta y$ sono indipendenti e casuali.

In alcune condizioni quindi, una stima più realistica dell'incertezza è:

$$\delta q \simeq \sqrt{(\delta x)^2 + (\delta y)^2} \leq \delta x + \delta y$$




lo stesso vale per gli errori relativi

$$\varepsilon_q = \sqrt{\varepsilon_x^2 + \varepsilon_y^2}$$

somma in quadratura

## Casi in cui non si somma in quadratura

$x = (2.0 \pm 0.1)$ cm

$z = x^2 = x \cdot x$

quale?

$$\frac{\delta z}{z} = 2\,\frac{\delta x}{x}$$

$$\frac{\delta z}{z} = \sqrt{\left(\frac{\delta x}{x}\right)^2 + \left(\frac{\delta x}{x}\right)^2} = \sqrt{2}\,\frac{\delta x}{x}$$

oppure incertezze sistematiche.

Es

$$g = 4\pi^2 \frac{l}{T^2}$$

$l = 92.95 \pm 0.1$ cm

$T = 1.936 \pm 0.004$ s

$$g = g_{best} \pm \delta g$$

?

## Soluzione

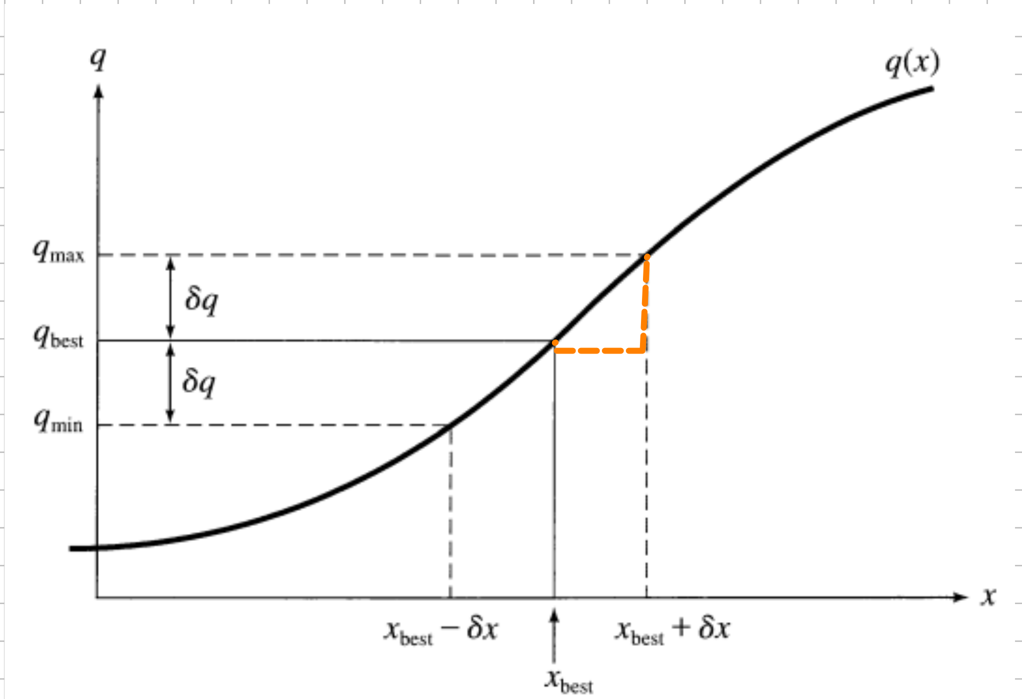
$$g_{best} = 4\pi^2 \frac{92.95\ \text{cm}}{(1.936\ \text{s})^2} = \underbrace{979.035}_{\text{abbondare}}\ \frac{\text{cm}}{\text{s}^2}$$

$$\frac{\delta g}{g_{best}} = \sqrt{\Big(\underbrace{\frac{\delta \ell}{\ell}}_{0.1\%}\Big)^2 + \Big(\underbrace{2\frac{\delta T}{T}}_{0.4\%}\Big)^2} = \sqrt{\left(\frac{0.1}{92.95}\right)^2 + \left(2 \cdot \frac{0.004}{1.936}\right)^2} \approx 0,43\,\%$$

$$\delta g \approx 4,18\ \frac{\text{cm}}{\text{s}^2}$$

$$g = (979 \pm 4)\ \text{cm/s}^2$$

# Funzioni arbitrarie di una variabile (approfondimento)




$q = q(x)$ se $\delta x$ è "piccolo"

$$\delta q = q(x_{best} + \delta x) - q(x_{best})$$

$$\delta q = \frac{dq}{dx} \delta x = q'(x)\, \delta x$$

$$\delta q = |q'(x)|\, \delta x$$

$$q = q(x, y) \longrightarrow \delta q \approx \left|\frac{\partial q}{\partial x}\right| \delta x + \left|\frac{\partial q}{\partial y}\right| \delta y$$

se le incertezze sono indipendenti e casuali

$$q = q(x, y \ldots z) \qquad \delta q = \sqrt{\left(\frac{\partial q}{\partial x} \cdot \delta x\right)^2 + \left(\frac{\partial q}{\partial y}\, \delta y\right)^2 + \cdots + \left(\frac{\partial q}{\partial z}\, \delta z\right)^2}$$

# Analisi statistica degli errori casuali

errori ~~sistematici~~ - casuali

supponiamo di poter ripetere + volte la misura delle stesse grandezze !

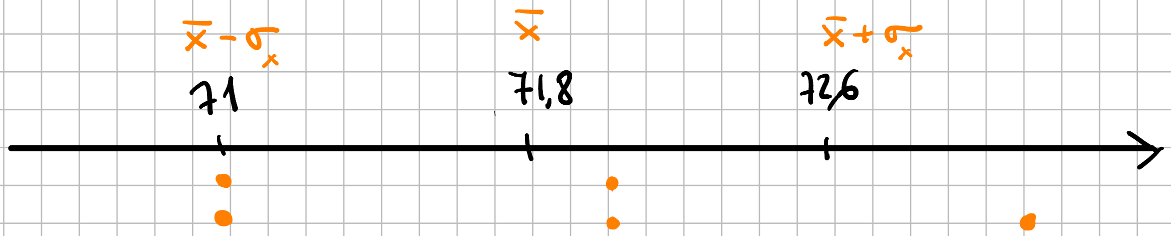
| Trial number $i$ | Measured value $x_i$ | Deviation $d_i = x_i - \bar{x}$ | Deviation squared $d_i^2$ |
|---|---|---|---|
| 1 | 71 | −0.8 | 0.64 |
| 2 | 72 | 0.2 | 0.04 |
| 3 | 72 | 0.2 | 0.04 |
| 4 | 73 | 1.2 | 1.44 |
| 5 | 71 | −0.8 | 0.64 |
| | $\Sigma x_i = 359$ | $\Sigma d_i = 0.0$ | $\Sigma d_i^2 = 2.80$ |

$$\bar{x} = 359/5 = 71.8 = x_{best}$$

$$\sigma_x = \sqrt{\frac{\Sigma d_i^2}{N}} \xrightarrow{\text{meglio}} \sigma_x = \sqrt{\frac{\Sigma d_i^2}{N-1}}$$

è una stima dell'incertezza in una singola misura

$\sigma_x = 0.8$




osservo

Circa il 30% dei dati giace fuori da $\bar{x} \pm \sigma_x$

circa il 70% delle nostre misure sono entro $\bar{x} \pm \sigma_x$

Se io dovessi fare una sesta misura c'è una probabilità del 70% che si trovi entro $\bar{x} \pm \sigma_x$

"Vi è un 70% di probabilità che una singola misura differisca meno di $\sigma_x$ dal valore vero, di cui $\bar{x}$ è una stima molto affidabile"

Quindi $\sigma_x = \delta_x$ hanno lo stesso significato

$k$ : $\frac{N}{m}$ : 86 85 84 89 86 88 88 85 83 85

quale incertezza attribuire alle singole misure ?

$\bar{x} = 85.9 \ \frac{N}{m}$ $\qquad$ $\sigma_x = 2 \ \frac{N}{m}$

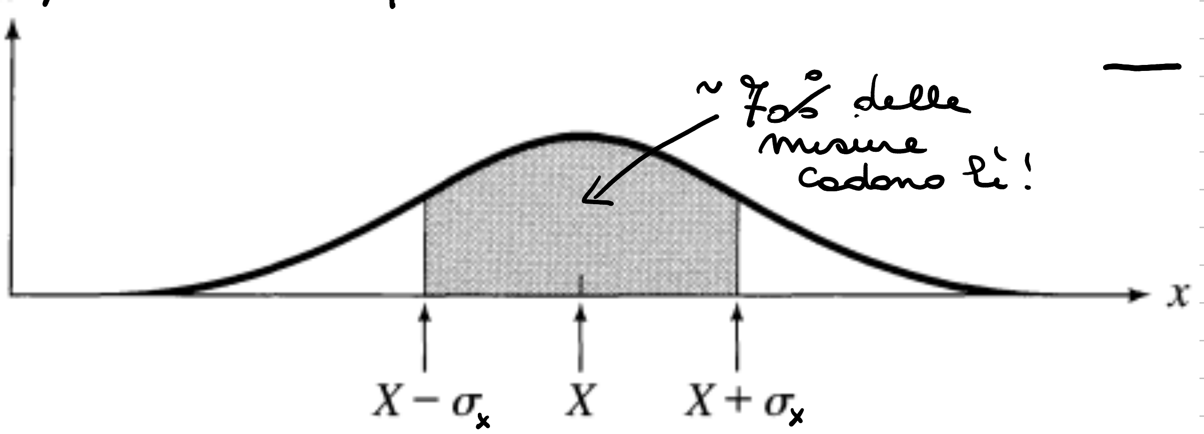

$\sigma_x = 2 \frac{N}{m}$ è la stima STATISTICA dell'incertezza da attribuire ed ogni SINGOLA MISURA cioè $84 \pm 2 \frac{N}{m}$ o

$89 \pm 2 \frac{N}{m}$

ma io ho fatto $N = 10$ misure della stessa grandezza, con media $85.9\ \mathrm{N/m} \longrightarrow \pm\ ?$

Vediamo come si distribuiscono le misure SINGOLE $\longrightarrow$ GAUSSIANA

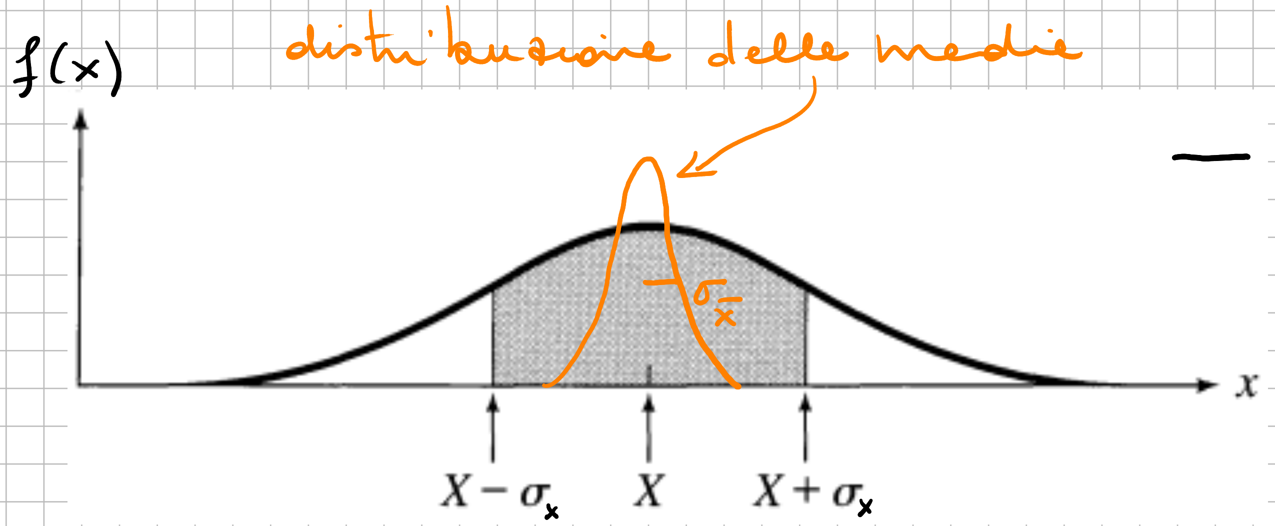
$f(x)$ = densità di probabilità




— singole misure

$\sigma_x$, per $N$ grande varia poco

e la distribuzione delle medie?




— singole misure

$\sigma_{\bar{x}}$ varia con N

ho lavorato tanto per fare N misure → merito un premio : $\boxed{\sigma_{\bar{x}}}$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N}}$$

diminuisce lentamente con N

$$x = 85.9 \pm \frac{2}{\sqrt{10}} \ \text{N/m}$$

$$x = 85.9 \pm 0.6 \ \text{N/m}$$

questo è il frutto del lavoro

Cose approfondire ?

- studio della distribuzione normale (errori distribuiti casualmente)

- Metodo dei minimi quadrati (o regressione lineare)

$x, y$ molte misure per diversi valori. Se prendiamo per garantito che $y = Ax + B$, come posso trovare analiticamente la migliore retta che interpola la mia serie di dati?

- $x$ e $y$ diverse misure (serie di dati)

i valori misurati $(x_1, x_2, \dots x_n)$, $(y_1, y_2, \dots, y_n)$ soddisfano realmente le nostre attese che $y$ sia lineare in $x$? (correlazione lineare)

fit con la calcolatrice programmabile o con Excel