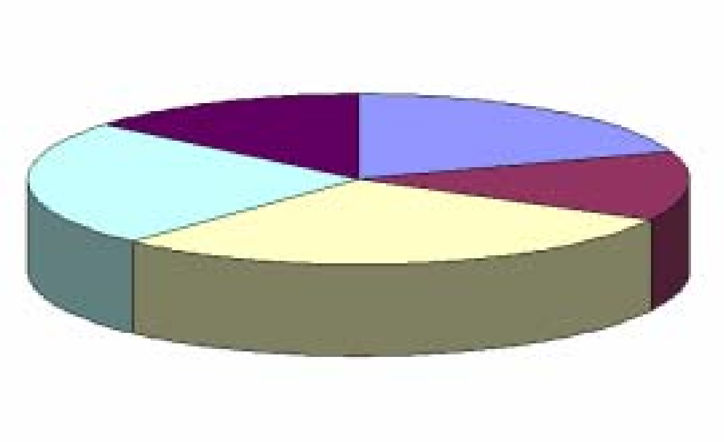
UNIVERSITÀ DEGLI STUDI DI SALERNO

DIPARTIMENTO DI FARMACIA

La statistica applicata alla chimica analitica



C. A. Mattia

Statistica

Forma di osservazione e induzione appropriata allo studio quantitativo di fenomeni che si presentano come pluralità o masse di casi, suscettive di variare senza regole assegnabili a tutto rigore. (R. Benini)

Metodi quantitativi di analisi di proprietà globali che interessano un numero elevato di entità facendo ricorso al calcolo della probabilità.

Metodi che permettono di organizzare e riassumere un insieme di dati e di evidenziarne determinate caratteristiche.

Classificazioni

Statistica descrittiva

Organizzazione e presentazione dei dati mediante metodi numerici e grafici.

Statistica induttiva
(statistica inferenziale)

Rilevazione di proprietà globali da analisi di una parte limitata.

Cenni storici

- La statistica, intesa come semplice raccolta di dati (censimenti,…), era usata in tutte le antiche civiltà.
- Le prime indagini statistiche si devono a John GRAUNT che fu incaricato dalla Società Reale di Londra di effettuare ricerche sull'andamento di caratteristiche della popolazione londinese (religione, malattie, provenienza, ceto economico,…).

Natural and Political Observations made upon the Bills of Mortality (1662)

Cenni storici

William PETTY, sempre nel XVII secolo, ne estese le applicazioni alle scienze economiche definendola

aritmetica politica.

Studio di fenomeni relativi alle strutture sociali, demografiche e in genere a tutte le discipline scientifiche.

Pascal, de Moivre, de Fermat, Bernouille, Poisson, Laplace, Fourier, Quetelet ("uomo medio"), Bayers,

Gauss,

Boole, Mendel (ereditarietà), Galton (regressione), Pearson (χ^2), Fischer, Gosset (Student), Savage.

Farmaco

Trattamento con farmaco

guariti: **947** non guariti: **319**

Trattamento con placebo

guariti: **614** non guariti: **330**

Il farmaco è efficace?

La domanda ha senso?

Che tipo di risposta ci aspettiamo?

Ereditarietà caratteri genetici

Mendel 1865

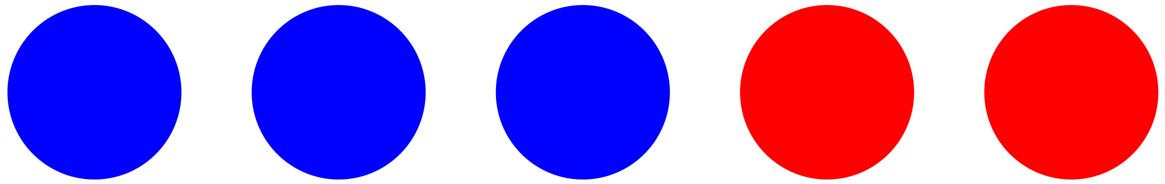
Teoria: ¾ deve avere la buccia liscia.

Teoria: ¾ deve essere di colore giallo.

	lisci	rugosi	
gialli	315	101	416
verdi	108	32	140
	423	133	556

I dati sperimentali sono in accordo con la teoria?

Estrazione



In un sacchetto vi sono 3 palline blu e 2 palline rosse.

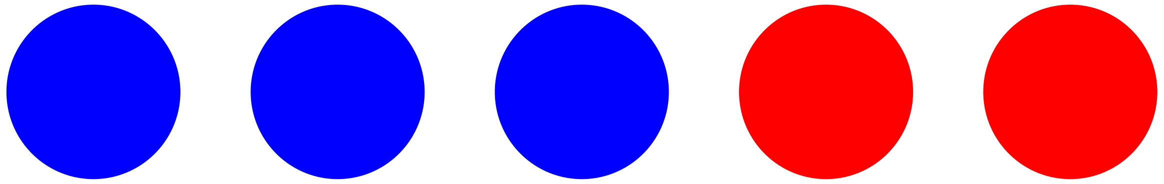
Scegliendo a caso (estrazione) di che colore sarà la pallina?

Ripetendo più volte l'estrazione cosa notiamo?

$$\frac{\text{numero estrazioni rosse}}{\text{numero estrazioni totali}} \qquad \frac{\text{numero palline rosse}}{\text{numero palline totali}}$$

Per un numero elevato di estrazioni i due rapporti tendono a coincidere.

Estrazione



Il rapporto tra le palline è 2/(2+3) = 0,40.

Il 40% delle palline è di colore rosso.

Il 60% delle palline è di colore blu.

Facendo 100 estrazioni ci aspettiamo un numero di palline estratte di colore rosso (n_R) molto vicino a 40.

Se il numero di estrazioni è elevato (1.000.000.000) n_R sarà vicinissimo al 40% (400.000.000).

Per una singola estrazione possiamo utilizzare i risultati delle osservazioni fatte e dire che il colore sarà rosso al 40%.

Ovvero diciamo che la probabilità sarà 0,40 o del 40%.

Assegnazione della probabilità

- **Punto di vista oggettivo**
  - **definizione frequenzistica: limite del rapporto tra il numero di eventi favorevoli (m) ed il numero di prove (n) per $n \to \infty$.**
  - **definizione classica: rapporto tra il numero di casi favorevoli ed il numero di casi possibili.**
  - **definizione assiomatica: assegna solo le regole, lascia aperto il problema dell'assegnazione dei valori.**
- **Punto di vista soggettivo**
  - **equa scommessa (misura del grado di fiducia che ciascuno di noi ripone nel verificarsi di un dato evento).**

L'equità in genere deriva da una stima di tipo "classica".

Statistica applicata

Cosa è la statistica?

La statistica è la scienza di:

- Raccogliere
- Descrivere
- Organizzare
- Interpretare

Dati

per transformarli in informazioni in modo da permettere di prendere decisioni efficienti.

Da chi viene usata?

- **Organismi ufficiali.**
- **Politici.**
- **Amministratori.**
- **Sportivi.**
- **Giornalisti.**
- **Commercianti.**
- **Medici.**
- **Addetti al controllo di qualità.**
- **etc.**

Tipi di statistica

Statistica descrittiva

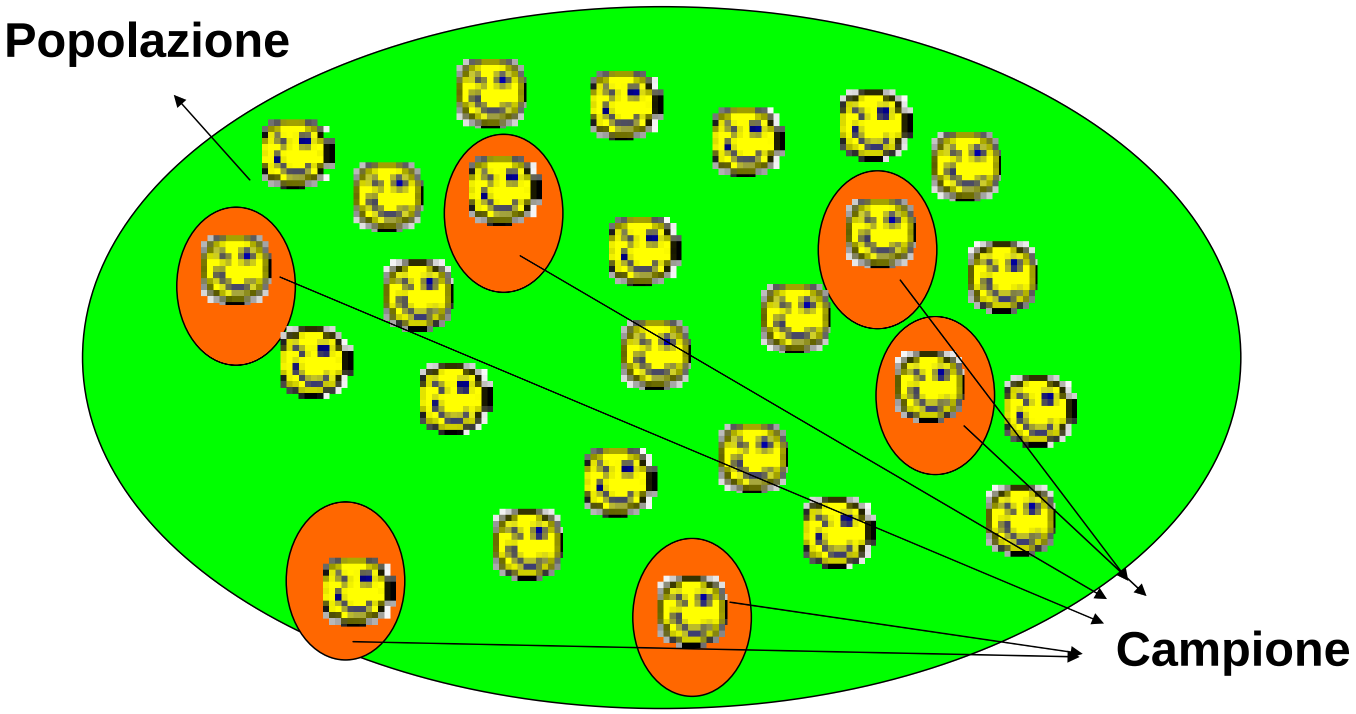
Metodo di raccogliere, organizzare, riassumere e presentare i dati sotto forma di informazioni.

- Esempio 1: I dati dell'ultimo censimento.
- Esempio 2: Il numero di ricoveri in ospedale durante l'anno.
- Esempio 3: I reati commessi nel mese scorso.
- Esempio 4: La richiesta di integratori alimentari.
- Esempio 5: I risultati delle ultime elezioni.

Tipi di statistica

- ***Statistica Inferenziale:*** metodi usati per determinare proprietà di ***popolazioni*** basandosi su analisi di ***campioni***.
- ***Popolazione*** è l'insieme di individui, oggetti o eventi di cui saranno analizzate le proprietà.
- ***Campione*** è, in generale, un sottoinsieme della popolazione in considerazione.

Popolazione e campione



Parametro e statistico

- ***Parametro***: *Valore numerico che riassume tutti i dati di una popolazione. Come simboli si utilizzano lettere dell'alfabeto greco:* μ σ.
- ***Statistico***: *Valore numerico che riassume i dati di un campione. Come simboli si utilizzano lettere dell'alfabeto inglese:* ***x s***.

Variabile

- *Variabile:* Caratteristica di interesse di ciascun elemento di una *popolazione* o di un *campione*.
- *Dato:* Valore della variabile associata ad una *popolazione* o ad un *campione*. É un numero, una parola, un simbolo, etc...
- *Dati:* Insieme dei valori collezionati per gli elementi relativi alla *popolazione* o al *campione*.

Tipi di variabili

- ***Qualitativa* o di *attibuto***

Classifica o descrive un elemento con valori che non costituiscono uno spazio metrico e pertanto le operazioni aritmetiche (somma, media, etc...) non sono significative.

Esempi: **Sesso, Nazionalità, Tipo di automobile, Grado di soddisfazione, etc...**

Tipi di variabili

- ***Quantitativa* o *Numerica***

Quantifica un elemento della popolazione. I valori costituiscono uno spazio metrico e pertanto le operazioni aritmetiche (somma, media, etc...) sono significative.

Esempi: **Numero di abitanti, Numero di figli, Chilometri percossi, Tempo di attesa, etc..**

Tipi di variabili

- Le variabili quantitative si possono classificare in **discrete** e **continue**.

Quantitative discrete: possono assumere solo alcuni valori caratterizzati da salti e interruzioni.

- Es. 1: Numero di ricoveri giornalieri (0,1,2,3,...).
- Es. 2: Punteggio F1 (25,18,15,12,10,8,6,4,2,1).
- Es. 3: Punteggio Tennis (15,30,40,...).
- Es. 4: Litri bottiglie champagne (1/2,3/4,1,2,...).

Tipi di variabili

- Le variabili quantitative si possono classificare in **discrete** e **continue**.

Quantitative continue: possono assumere qualsiasi valore entro uno specifico intervallo.

- Es. 1: Peso alla nascita.
- Es. 2: Stipendio di un impiegato.
- Es. 3: Tempo di percorrenza di un autobus.

Scala di misura

- Le variabili quantitative si misurano in scale ***nominali*** o ***ordinali***.
- ***Nominale:*** gli elementi si possono classificare in categorie ma senza un ordine o priorità.

- **Es. 1: Colore degli occhi.**
- **Es. 2: Stato civile.**
- **Es. 3: Luogo di residenza.**

Scala di misura

- Le variabili quantitative si misurano in scale ***nominali*** o ***ordinali***.
- ***Ordinale:*** gli elementi si possono classificare in categorie con un ordine o priorità.

- **Es. 1: Soddisfazione di un servizio.**
- **Es. 2: Stato di un malato.**
- **Es. 3: Stato socio-economico.**

Scala di misura

- Scale più sofisticate sono quelle ad ***intervalli*** o di ***rapporti***.
- ***Scala ad intervalli:*** non solo si possono ordinare le misure ma è anche nota la distanza tra due misure qualsiasi. É necessario definire una distanza unitaria e lo zero è arbitrario.

- Esempio: Temperatura in gradi Celsius.

Scala di misura

- Scale più sofisticate sono quelle ad ***intervalli*** o di ***rapporti***.
- ***Scala di rapporti:*** è il più alto livello di misurazione ed è caratterizzato dal fatto che può essere determinata sia l'uguaglianza di rapporti che di intervalli. Lo zero indica l'assenza totale della grandezza che si sta misurando.
- **Esempi:Tempo, statura, peso, concentrazione, etc..**

Verifica

- Dati
  - Campionamento
  - Assunzioni
- Distribuzione riferimento
  - Test
- Significatività
- Ipotesi
  - nulla (H_0): posta con lo scopo di essere screditata
  - alternativa (H_A)
- Decisione

Conclusioni

Bisogna sempre tener presente che le valutazioni sono di tipo probabilistico e che in generale si possono commettere errori di I[a] e II[a] specie.

Decisione	Ipotesi nulla	
	vera	falsa
Non rifiuto H_o	Scelta corretta	Errore tipo II (β)
Rifiuto H_o	Errore tipo I (α)	Scelta corretta

Farmaco

Trattamento con farmaco

guariti: **947** non guariti: **319**

Trattamento con placebo

guariti: **614** non guariti: **330**

Il farmaco è efficace?

La domanda ha senso?

Che tipo di risposta ci aspettiamo?

Farmaco

- Ipotesi nulla: equiprobabilità (non efficace).
- Distribuzione χ^2 per l'indipendenza.
- Gradi di libertà = 1 $\chi^2_{95\%} = 3,84$
- $\chi^2 = (947-894,22)^2/894,22+(614-666,78)^2/666,78+\ldots = 6,21$
- 6,21 è maggiore di 3,84:
  - l'ipotesi nulla può essere rigettata.
  - il farmaco ha, con una affidabilità del 95%, una certa efficacia.
  - se si varia l'affidabilità può non essere rigettata l'ipotesi nulla.
  - l'efficacia del farmaco non è provata se si considera una affidabilità del 99%.

Ereditarietà caratteri genetici

Mendel 1865

Teoria: ¾ deve avere la buccia liscia.

Teoria: ¾ deve essere di colore giallo.

	lisci	rugosi	
gialli	315	101	416
verdi	108	32	140
	423	133	556

I dati sperimentali sono in accordo con la teoria?

Mendel

- Ipotesi nulla: i dati sperimentali non sono in accordo con la teoria.
- Distribuzione delle differenze: χ^2
- Teoria:
  - GL = 3/4•3/4•556 = 9/16•556; GR = 3/4•1/4•556 = 3/16•556
  - VL = 1/4•3/4•556 = 3/16•556; VR = 1/4•1/4•556 = 1/16•556
- Sperimento: GL = 315; GR = 101; VL = 108; VR = 32
- gradi di libertà = 3 $\chi^2_{95\%} = 7,81$
- $\chi^2 = \Sigma[(\text{popol-camp})^2/\text{popol}] = \Sigma[(\text{teo-spe})^2/\text{teo}]$
- $\chi^2 = (315-312,75)^2/312,75 + \ldots = 0,47$
- **Essendo il valore 0,47 < 7,81 l'ipotesi nulla deve essere rigettata e si può affermare, con una affidabilità del 95%, che i dati sperimentali sono in accordo con la teoria.**

Come mentire con la statistica

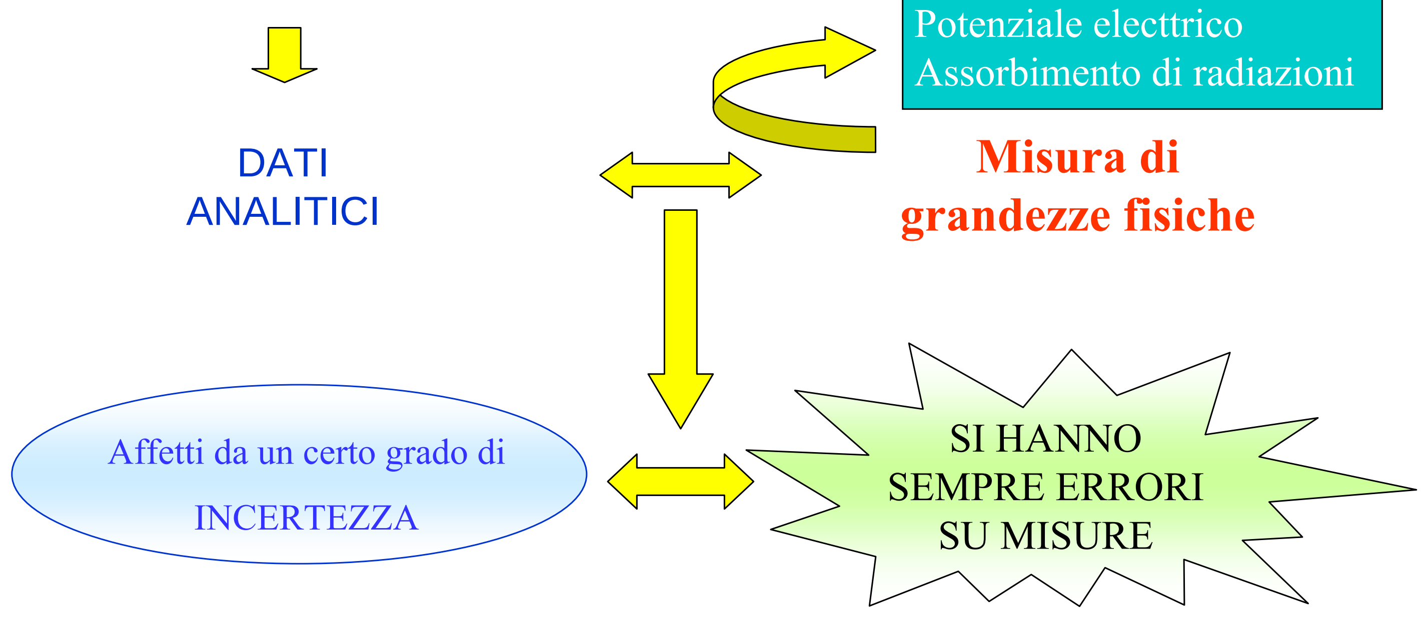
- **La probabilità di essere coinvolti in un incidente automobilistico aumenta con il tempo che si sta al volante.**

  Più si guida velocemente, minore è il tempo di percorrenza.
- **Nel 33% degli incidenti mortali sono coinvolti guidatori con elevato tasso alcolilco.**

  Pertanto, il rimanente 67% è stato causato da sobri.
- **Tenendo conto di ciò, è evidente che il modo più sicuro di guidare è da ubriachi ed ad alta velocità.**

Analisi

Analisi quantitativa



Qualunque misura fisica è soggetta a variabilità

Media e mediana

Sia dato un insieme di misure $x_1, x_2, \ldots x_N$.

Media: $$\bar{x} = \frac{\sum_i x_i}{N}$$

Mediana: avendo ordinato le misure in ordine crescente

N pari $$\tilde{x} = \frac{x_{N/2} + x_{N/2+1}}{2}$$ **N dispari** $$\tilde{x} = x_{N/2+1}$$

Nel caso delle misure: 10, 10, 12, 13, 13, 13, 15, 18, 25, 26, 26, 27, 28, 28, 35

la media è 19,93 e la mediana è 18.

Moda: è l'osservazione che compare con maggiore frequenza (13 tre volte)

Precisione: bontà dell'accordo tra i risultati di misurazioni successive.

Accuratezza o esattezza: bontà dell'accordo tra il risultato, x_i, o il valore medio dei risultati di un'analisi, ed il valore vero o supposto tale, x_t.

Media e deviazione standard

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

La media e la deviazione standard sopra definite, essendo valutate sulla base di un numero finito, e normalmente molto basso, di misurazioni, cioè di un **campione** delle infinite misurazioni che costituiscono l'intera **popolazione** delle misurazioni, sono solo stime della media e della deviazione della popolazione. Per un numero molto alto (tendente a infinito) di misurazioni si può scrivere:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} \qquad \sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}}$$

Normalmente, queste due ultime equazioni valgono per $N > 20$.

Errori sistematici

- Errori strumentali
  - malfunzionamento strumenti
  - errata calibrazione
  - condizioni non appropriate
- Errori di metodo
  - comportamento non ideale
  - reazioni indesiderate
- Errori personali
  - pregiudizio
  - arrotondamento
  - daltonismo, …
- Errori grossolani

Errori sistematici (δ)

Gli **errori sistematici** si manifestano nella tendenza deterministica di un dato metodo a **sovrastimare** (o **sottostimare**) il vero valore θ. Pertanto, l'universo delle misure che si possono virtualmente ottenere con tale metodo ha media **μ**, che differisce dal valore θ di una quantità **δ = μ - θ**.

Gli errori sistematici hanno cause ben determinate, inerenti o al **metodo** (***es.***: scarsa selettività del reagente usato per la titolazione di un certo soluto), o alle **condizioni di esecuzione** del procedimento analitico (***es.***: strumento non calibrato correttamente).

Una misura è tanto più accurata quanto minore è l'entità dell'errore sistematico (δ) da cui è affetta.

Errori sistematici

Gli errori sistematici possono essere identificati ed annullati mediante

- analisi di campioni standard, se disponibili;
- analisi del campione mediante un metodo indipendente, ovvero tale da prevedere l'utilizzo di strumentazione di provata affidabilità o di riferimento;
- analisi del ***bianco***, cioè di una soluzione contenente tutti i componenti presenti nel campione in esame eccetto l'analita di interesse; il bianco ideale è costituito dalla stessa matrice in cui è contenuto l'analita di interesse; l'analisi del bianco nelle titolazioni volumetriche consente, per esempio, di correggere l'errore connesso al volume di titolante necessario per far virare l'indicatore colorimetrico;
- analisi di campioni contenenti un diverso ammontare della variabile misurata (per es. si pensi alla perdita connessa alla solubilità durante il lavaggio con volumi diversi di acque di lavaggio).

Rigetto degli outliers

Spesso uno o più dati di una serie appaiono **irragionevolmente** diversi dagli altri.

L'eliminazione di un dato sospetto porta a variazioni notevoli del valore della media e a miglioramenti sostanziali di quello della deviazione standard, ma la decisione deve essere ***giustificata statisticamente*** e non dall'***intuito***, spesso falsato da bias personale.

L'eliminazione di un dato sospetto può essere decisa sulla base del test *Q* di Dixon (si usa anche il test T di Grubbs). L'equazione che permette di calcolare *Q* dipende dalla numerosità dei dati a disposizione. Avendo disposto i risultati in ordine crescente:

$$Q_{exp} = \frac{x_n - x_{n-1}}{x_n - x_1} \qquad Q_{exp} = \frac{x_2 - x_1}{x_n - x_1}$$

Q_{exp} deve essere confrontato con il valore critico tabulato in funzione del numero di osservazioni e del livello di fiducia (confidenza).

Rigetto degli outliers

Un metodo più semplice è dato dalla regola del 2,5 d.

il dato sospetto si elimina dalla media e si calcola, per i rimanenti n - 1 dati, la deviazione media dalla relazione:

$$\bar{d} = \frac{\sum_{i=1}^{n-1} |x_i - \bar{x}|}{n-1}$$

Il dato sospetto si rigetta se la sua deviazione

$$d_{sosp} = |x_{sosp} - \bar{x}|$$

è maggiore di 2,5 volte la deviazione media.

Esempi

❖ Verificare la presenza di outliers dal set di dati:

$X_1 = 23{,}23$; $X_2 = 21{,}29$;

$X_3 = 20{,}66$; $X_4 = 29{,}05$;

$X_5 = 23{,}33$;

$X_1 = 20{,}66$; $X_2 = 21{,}29$;

$X_3 = 23{,}23$; $X_4 = 23{,}33$;

$X_5 = 29{,}05$;

$$Q_{exp} = \frac{29{,}05 - 23{,}33}{29{,}05 - 20{,}66} = 0{,}682$$

Il valore critico è $Q_{\alpha=0{,}05;5}$ è 0,642. Dato che il valore sperimentale è maggiore di quello critico si può eliminare il dato (la probabilità di considerare accettabile il dato è inferiore al 5%).

La media è 22,1275, la deviazione media è 1,1525 la deviazione del quarto dato (29,05) è 6,9225 che è maggiore di 2,5•1,1525 = 2,88125.

Il valore per la regola del 2,5 d deve essere rigettato.

Rigetto degli outliers

Raccomandazioni per il trattamento degli outliers. Se un dato appare anomalo:

- accertarsi di non aver commesso un errore grossolano;
- ripetere l'analisi;
- eseguire un test di rigetto (Q, 2,5d,…);
- nel caso il dato sia confermato come outliers, eseguire una nuova replica;
- tenere presente che il rigetto è legato al livello di fiducia usato;
- **in casi limiti è consigliabile non eliminare il dato.**

Errori casuali

Gli ***errori casuali*** (detti anche ***indeterminati*** o "random"), causano una dispersione più o meno simmetrica dei dati intorno al valore medio.

Essi sono legati a fluttuazioni indefinite di una miriade di parametri sperimentali, quali temperatura, pH, pressione, umidità, punto d'arresto di una titolazione, forza ionica, ecc. oltre che alle tolleranze dei pesi delle bilance e della vetreria utilizzata per la misurazione di volumi e alle incertezze dei valori desunti dagli strumenti di misura.

Queste fluttuazioni avvengono anche cercando di lavorare con la massima cura.

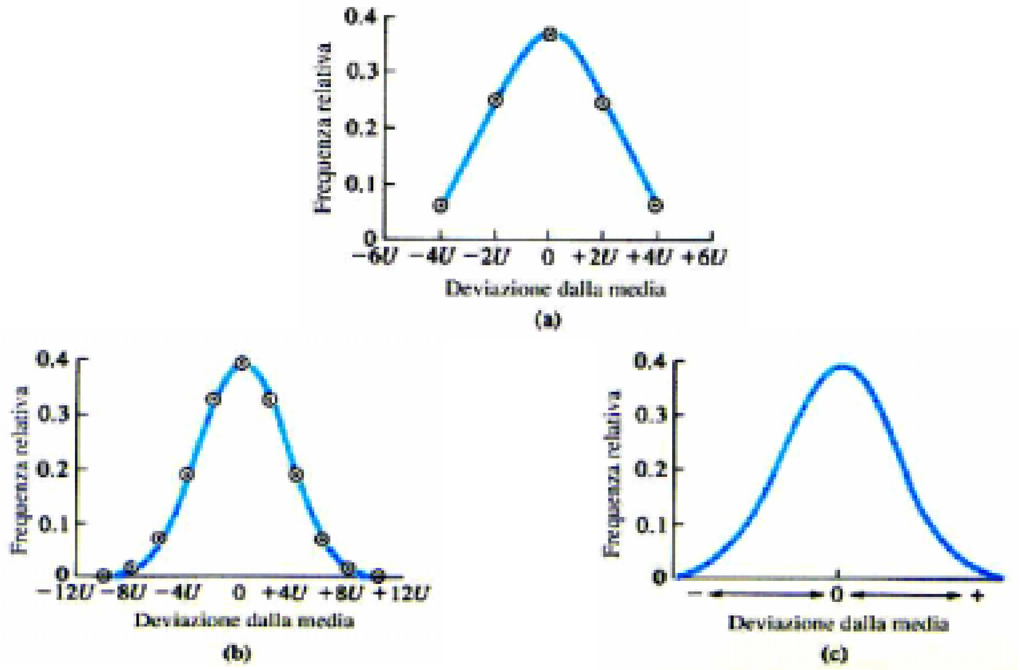
Gli errori casuali non possono essere eliminati, anche se possono essere ridotti operando con cura.

Errori casuali

Possibili combinazioni di quattro incertezze di uguale dimensione

Combinazione delle incertezze	Grandezza dell'errore indeterminato	Numero di combinazioni	Frequenza relativa
$+U_1+U_2+U_3+U_4$	$+4U$	1	1/16 = 0.0625
$-U_1+U_2+U_3+U_4$ $+U_1-U_2+U_3+U_4$ $+U_1+U_2-U_3+U_4$ $+U_1+U_2+U_3-U_4$	$+2U$	4	4/16 = 0.250
$-U_1-U_2+U_3+U_4$ $+U_1+U_2-U_3-U_4$ $+U_1-U_2+U_3-U_4$ $-U_1+U_2-U_3+U_4$ $-U_1+U_2+U_3-U_4$ $+U_1-U_2-U_3+U_4$	0	6	6/16 = 0.375
$+U_1-U_2-U_3-U_4$ $-U_1+U_2-U_3-U_4$ $-U_1-U_2+U_3-U_4$ $-U_1-U_2-U_3+U_4$	$-2U$	4	4/16 = 0.250
$-U_1-U_2-U_3-U_4$	$-4U$	1	1/16 = 0.0625

Distribuzione



Distribuzione di frequenza di misure affette da:
a) 4 incertezze indeterminate;
b) 10 incertezze indeterminate;
c) numerose incertezze indeterminate.

Errori casuali (ε)

Misurazioni dello stesso valore, ripetute in uno stesso procedimento analitico, e in condizioni il più possibile simili, portano spesso a misure differenti: non è possibile ripetere la misurazione in modo del tutto identico.

La somma di tutte le piccole e imprevedibili variazioni nell'esecuzione delle varie operazioni analitiche fa sì che le misure fluttuino attorno a un valore μ, che si scosta più o meno dal valore, a seconda dell'entità dell'errore sistematico.

Tali fluttuazioni attorno a μ ($\varepsilon = x - \mu$) sono dette errori casuali.

Una misura è tanto più precisa quanto minore è l'entità dell'errore casuale (ε) da cui è affetta.

Errore

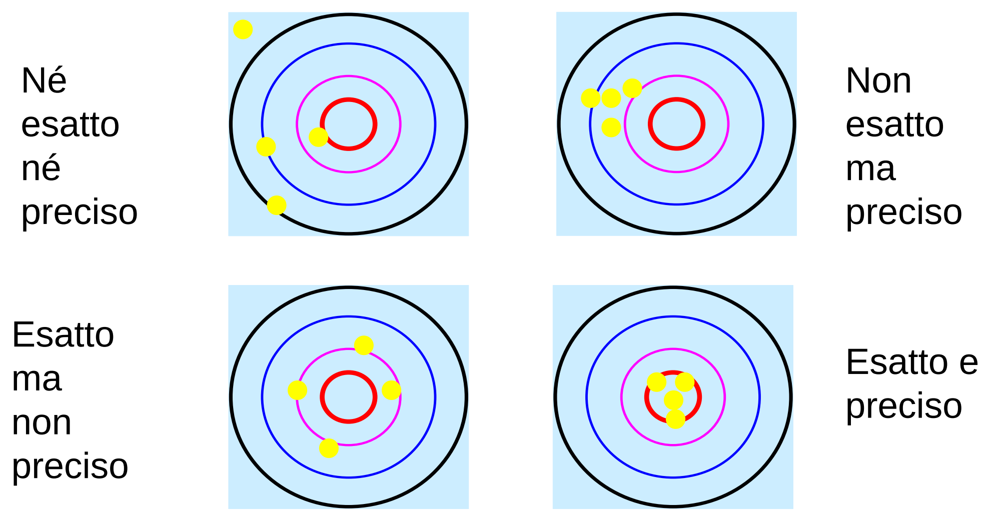
L'errore totale di una misura esente da errori grossolani può essere espresso come **somma** di una componente **sistematica** e di una componente **casuale**.

errore totale		errore sistematico		errore casuale
$(x-\vartheta)$	=	$(\mu-\vartheta)$	+	$(x-\mu)$
η	=	δ	+	ε
attendibilità		accuratezza		precisione

Precisione e accuratezza

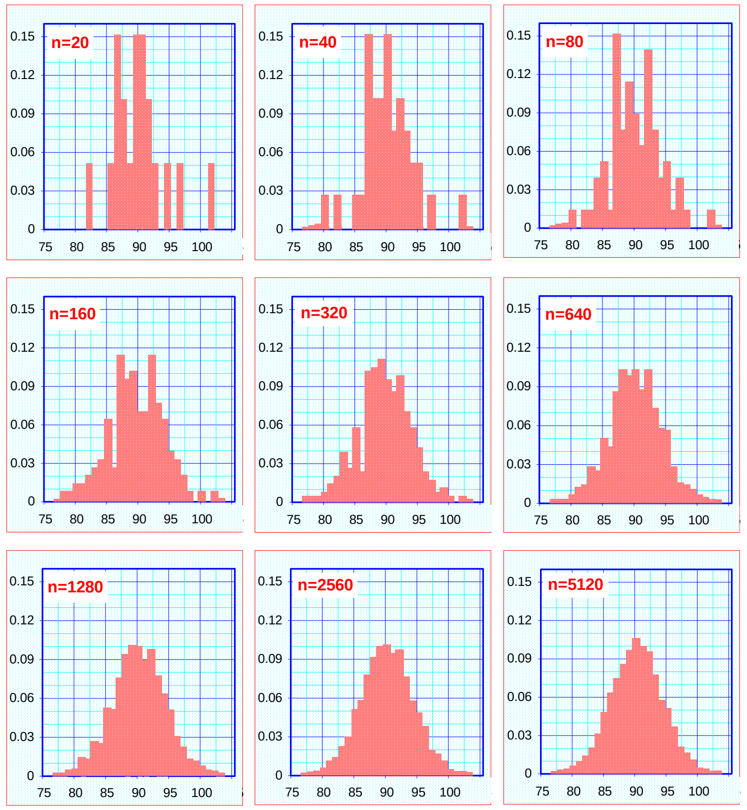
Gli errori possono essere **errori casuali** o **errori sistematici**.

Gli errori casuali influenzano la precisione, quelli sistematici l'esattezza.



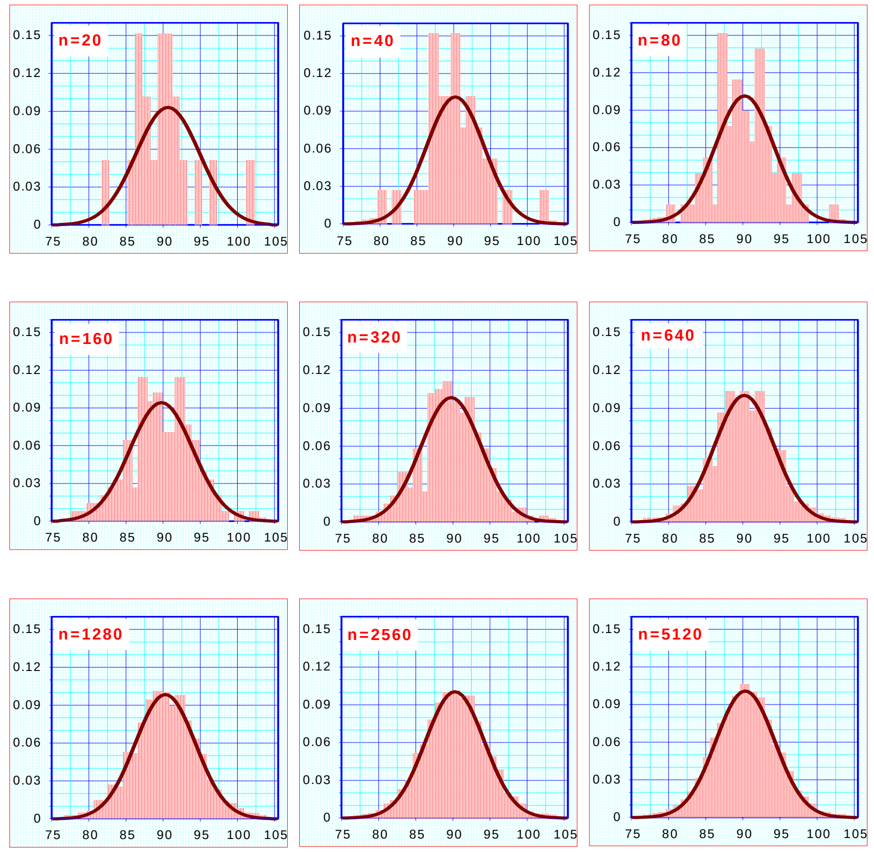
Distribuzione degli errori di misure

Si supponga di eseguire, in condizioni assai simili e con lo stesso metodo analitico, un gran numero di titolazioni di una soluzione di glucosio avente concentrazione 90 mg/dl, e di riportare in grafico le frequenze relative dei valori ottenuti (x) con le prime 20, 40, ... 5120 misure.



Distribuzione degli errori di misura

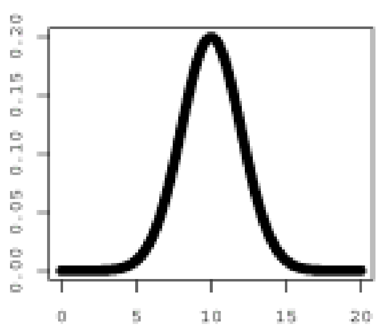
All'aumentare del numero di misure, i valori tendono ad accentrarsi attorno alla loro media e l'istogramma assume una forma *a campana* sempre più regolare, che può essere approssimata con una funzione reale nota come funzione di Gauss o funzione normale.



Gaussiana

La funzione gaussiana è definita da due parametri: μ (valore centrale) e σ (deviazione standard).

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



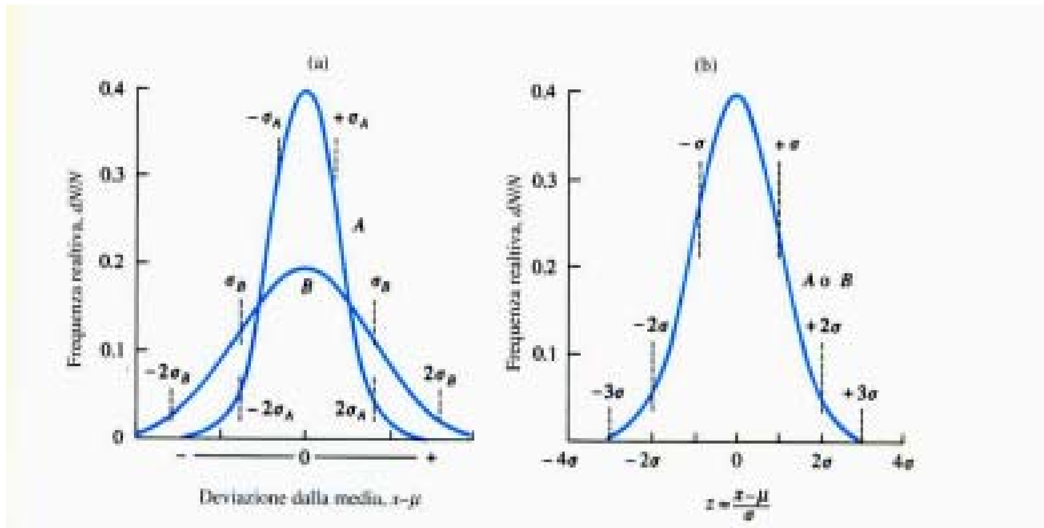
La distribuzione Gaussiana è ***simmetrica*** intorno al valore centrale (media e mediana coincidono) ed essendo una ***distribuzione di probabilità*** racchiude un'area unitaria.

Distribuzione normale

La distribuzione normale è quella che si ottiene scegliendo la media μ come origine degli assi cartesiani e la deviazione standard σ come unità di misura per l'asse x: la distribuzione normale rappresenta quindi ogni curva Gaussiana.

$$z = \frac{x-\mu}{\sigma}$$

$$y(z) = \frac{e^{-z^2/2}}{(2\pi)^{1/2}}$$



Gaussiana

ū = valore medio ascissa

P(|ū-μ| ≤ 0,67σ) = 50%

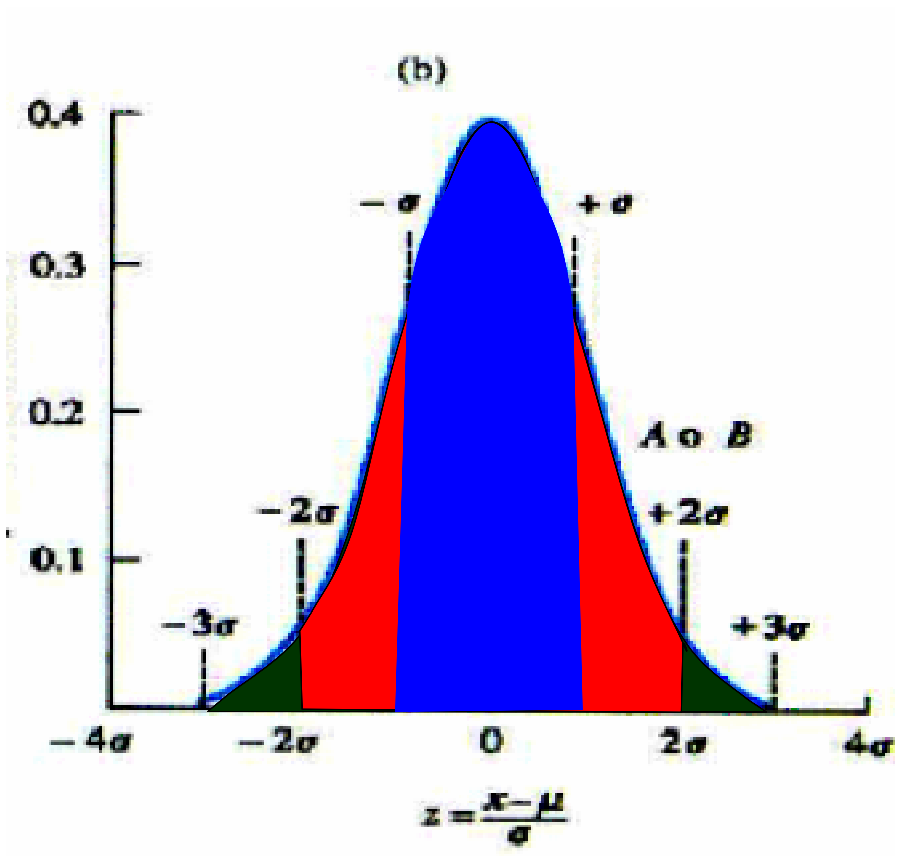
P(|ū-μ| ≤ σ) = 68,3%

P(|ū-μ| ≤ 2σ) = 95,4%

P(|ū-μ| ≤ 2,58σ) = 99%

P(|ū-μ| ≤ 3σ) = 99,7%

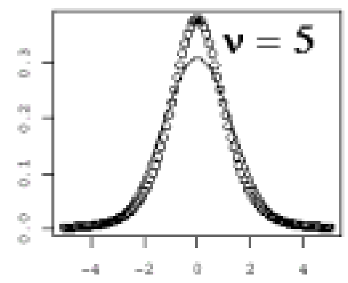
P(|ū-μ| ≤ 3,29σ) = 99,9%



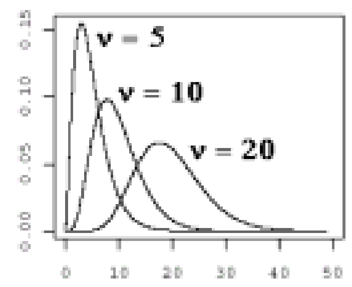
Altre distribuzioni

Si basano sul parametro ν (grado di libertà).

- t di Student
  - $\mu = 0$
  - $\sigma^2 = \nu/(\nu-2)$



- χ^2 di Pearson
  - $\mu = \nu$
  - $\sigma^2 = 2\nu$
  - non simmetrica



Intervallo di confidenza

- **Per una distribuzione normale si ha, per esempio, che circa il 95% dei valori possibili della media campionaria ricadrà entro due volte la deviazione standard.**
- **In generale, stimata la media campionaria, è possibile definire un intervallo in cui poter assumere ragionevolmente che in esso sia compreso il valore medio della popolazione.**

Intervallo di confidenza

- Tale intervallo si chiama **intervallo di fiducia o confidenza**, ed i suoi limiti estremi sono chiamati **limiti dell'intervallo di fiducia**. La probabilità che il valore atteso di un parametro stimato sia incluso in un intervallo stimato del parametro stesso si chiama livello di fiducia, e si indica con 1-α.
- Il livello di fiducia è espresso da un numero tra 0 e 1 (o in percento). La quantità complementare, α, si chiama **livello di significatività.**
- **Quindi la scelta di un determinato livello di fiducia non esclude totalmente la possibilità di fare previsioni sbagliate: se abbiamo scelto 1-α = 95% avremo comunque 5 possibilità su cento che il valore cercato (vero) cada al di fuori dell'intervallo di fiducia.**

Merendine

La percentuale di carboidrati in merendine viene dterminato mediante due diversi metodi (A e B) ottenedo i risultati seguenti:

METODO A	METODO B
37,51 %	36,44 %
38,01 %	37,01 %
36,99 %	35,89 %
37,78 %	36,22 %
37,20 %	36,01 %
38,25 %	

Metodi analisi

Si calcolano gli statistici media e varianza

$\bar{X}_A = 37{,}62\%$ $\bar{X}_B = 36{,}31\%$

$s_A = 0{,}48\%$ $s_B = 0{,}44\%$

e mediante il test ***F di Fischer*** ***(sulle varianze)*** e il test ***t di Sudent*** ***(sulle medie)*** si può affermare con un livello di confidenza del **95%** che il metodo **B** porta a risultati significativamente diversi rispetto al metodo **A** anche se i due metodi hanno varianze non significativamente diverse.

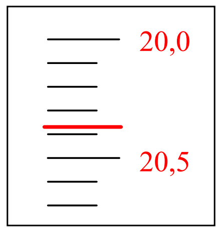
F_s=1,19; F_c=9,36 **t_s=4,68; t_c=2,26**

La convenzione delle cifre significative

Per rappresentare l'incertezza di una misura sperimentale fornita da uno strumento si può arrotondare il risultato in modo che esso contenga tutte le cifre certe più la prima incerta*.

Il volume prelevato da una buretta di 50 ml, tarata in decimi di millilitro, viene letto alla 2ª cifra decimale.

Il volume letto a fianco è 20,37 ml (o 20,38 ml).



In generale, per conservare n cifre significative bisogna eseguire i calcoli con (n+2) cifre.

$$\frac{(3{,}5^2 - 3{,}4^2)^2}{1{,}1} = 0{,}43_{282}$$

$$\frac{(3{,}509^2 - 3{,}419^2)^2}{1{,}143} = 0{,}34_{014}$$

* Più generalmente, il numero di cifre significative può essere deciso in base al valore della deviazione standard della media.
È importante arrotondare solo il risultato finale, mai quelli parziali.

Arrotondamento

Il risultato deve essere arrotondato alla prima (o seconda) cifra significativa dell'incertezza.

34,0967182736 ± 0,2703660271 = 34,10±0,27 ……

La prima (seconda) cifra significativa dell'incertezza corrisponde alla prima (seconda) cifra incerta della media.

Si preferisce l'uso di una sola cifra (dell'incertezza) se la prima cifra è **grande**, di due cifre se la prima è **piccola.**

34,0967±0,1703 ⇒ 34,10±0,17 **34,0967±0,7146 ⇒ 34,1±0,7**

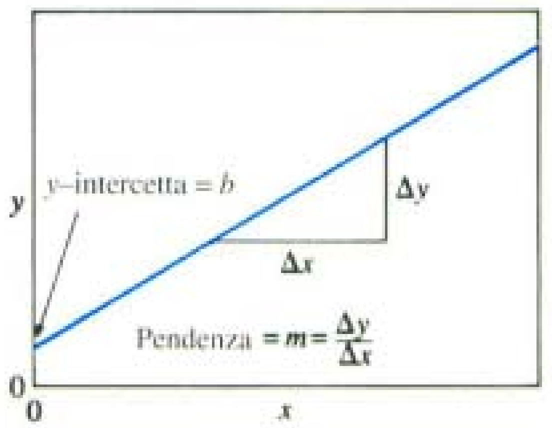
Il metodo dei minimi quadrati

La maggior parte delle determinazioni analitiche è effettuata utilizzando tecniche strumentali. Nel caso più frequente, la misurazione è di tipo indiretto. Si costruisce prima un diagramma di calibrazione analizzando campioni a concentrazione nota e riportando in grafico il segnale misurato (assorbanza, corrente, tensione, area di un picco, ecc.) in funzione della concentrazione. Si può quindi utilizzare il diagramma ottenuto per ricavare il valore della concentrazione del campione incognito da quello del segnale ad esso relativo.

L'equazione della retta di calibrazione è:

$y = m \cdot x + b$

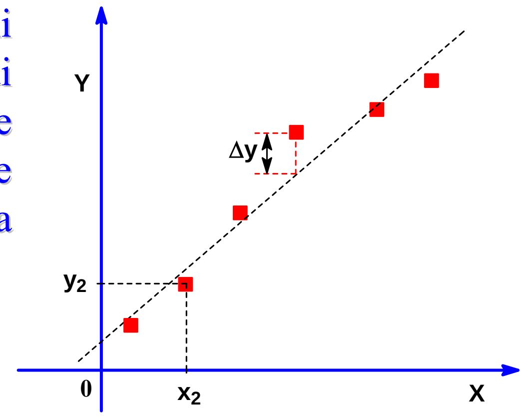
I punti sperimentali non sono mai allineati perfettamente: è necessario usare metodi obiettivi per tracciare la retta migliore che rappresenta i risultati sperimentali



C. A. Mattia

Il metodo dei minimi quadrati

Il metodo di regressione ai minimi quadrati è un metodo che permette di identificare la retta migliore mediante minimizzazione dei quadrati delle distanze tra i punti sperimentali e la retta supposta ideale (i residui).



Data una serie di risultati sperimentali
$(x_i;y_i)$ = concentrazione;segnale
si può dimostrare che pendenza ed intercetta sono calcolabili per mezzo delle equazioni:

$$m = \frac{\sum_i \left[(x_i - x_m)\cdot(y_i - y_m)\right]}{\sum_i (x_i - x_m)^2} \qquad b = y_m - m \cdot x_m$$

I calcoli sono normalmente eseguiti per mezzo di software (Excel, …).

C. A. Mattia

Esempi

❖ Trovare la retta di regressione ai minimi quadrati dei seguenti risultati, ottenuti analizzando il piombo in un campione di acqua potabile:

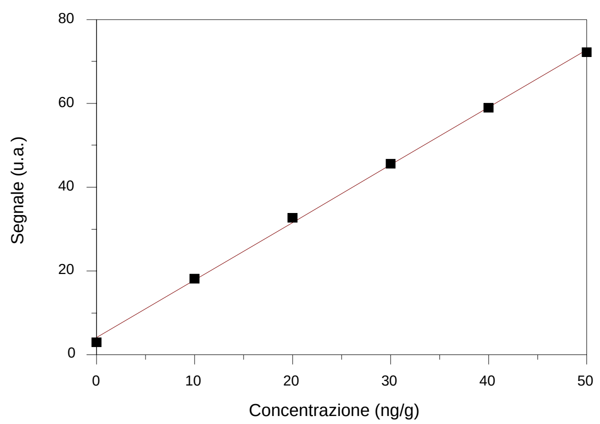
C(ng/g):	0	10	20	30	40	50
S(u.a.) :	3,0	18,2	32,7	45,6	59,0	72,2

m = 1,3751
b = 4,0714

S(C) = 1,3751•C + 4,0714

Infatti: S(30) = 45,324 u.a.

Residuo (30): -0,276

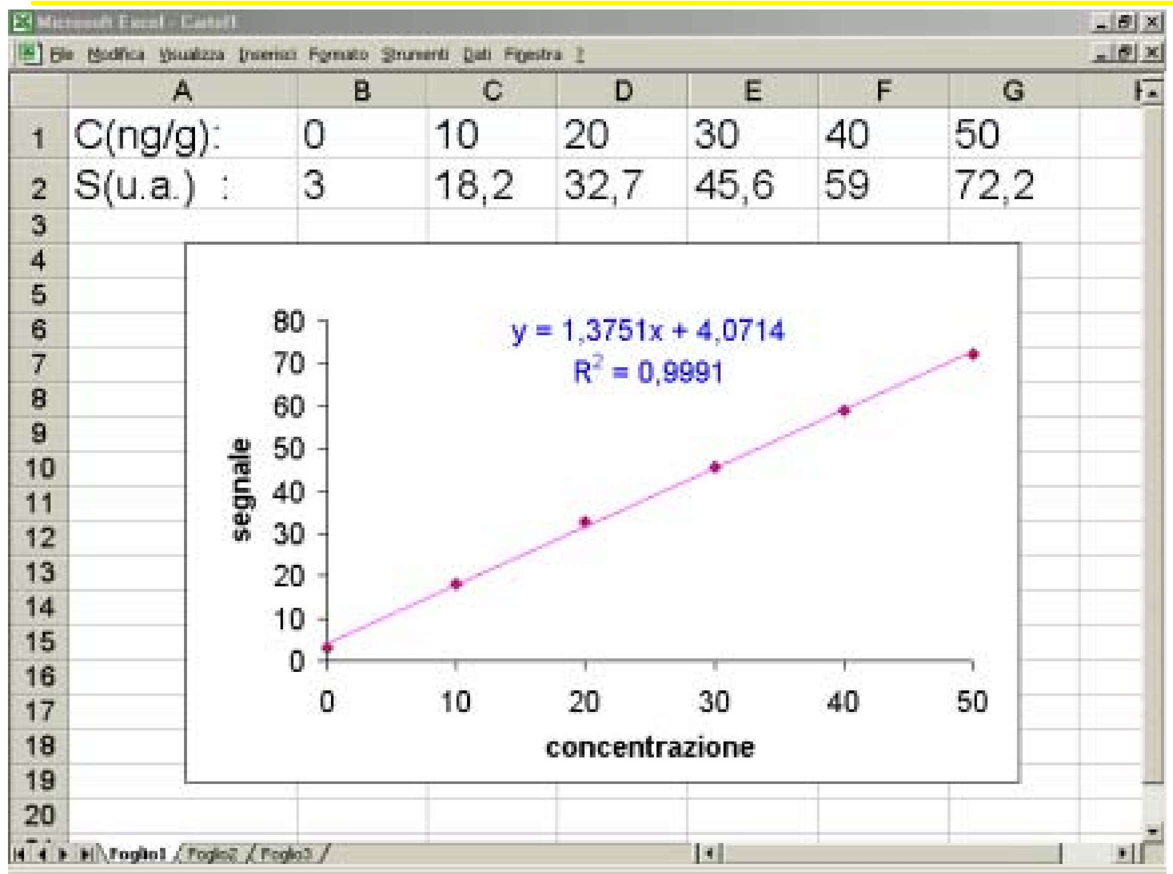


I valori di m e di b sono esatti o sono affetti da incertezza?

Il valore di C ottenuto sostituendo un valore di S è esatto o affetto da incertezza?

C. A. Mattia

Esempi



C. A. Mattia

Misurazioni

Nessuna misurazione fisica dà misure esatte, ma tutte comportano una

INCERTEZZA DI MISURA O ERRORE.

Quanto è grande questa incertezza?

1) **dipende dalla sensibilità dello strumento di misurazione (errore strumentale)**

(**sensibilità**: valore più piccolo della grandezza che uno strumento è in grado di apprezzare).

2) **dipende dall'operatore (errore soggettivo): è necessario ripetere più volte la misura e calcolarne il valore medio.**

Pesata

Es.: se pesassimo un anello d'oro troveremmo

	portata (g)	sensibilità bilancia	pesata anello (g)
1) bilancia di casa	5000	5g	5<…<10
2) bilancia droghiere	3000	1g	7<…<8
3) bilancia orefice	50	0,01g	7,35
4) bilancia analitica	100	0,0001g	7,3538

Errori

Ogni misurazione è affetta da errori sperimentali.

Gli errori sperimentali si combinano tra loro in modo da rendere ogni nuova misura più o meno diversa dalla precedente.

L'incertezza della misura sperimentale non può mai essere eliminata completamente perciò il valore vero di una quantità è sempre sconosciuto.

Tuttavia, spesso può essere valutata l'entità probabile dell'errore.

È possibile definire i limiti entro cui il valore vero di una quantità misurata cade con un dato livello di probabilità.

La stima dell'accuratezza dei dati sperimentali non è mai facile.

Errori

Dati con precisione e accuratezza ignoti sono privi di significato.

Errore assoluto: è la differenza, compreso il segno tra il valore misurato e il valore vero (o riconosciuto come tale).

Errore relativo: è l'errore assoluto diviso per il valore vero.

Errore relativo percentuale: è l'errore relativo per 100.

Dispersione dei dati

Errore (ε) in genere è uguale a f•σ, dove σ è la deviazione standard e f è un fattore che ne indica l'affidabilità.

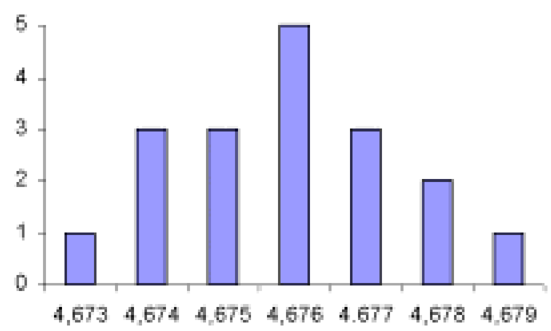
(Gaussiana: f = 1,96 => 95%; f = 2,56 => 99%).

Es. pH soluzione: 4,676; 4,679; 4,673; 4,674; 4,676; 4,675; 4,678; 4,677; 4,676; 4,675; 4,674; 4,677; 4,678; 4,674; 4,676; 4,677; 4,675; 4,676.

media (μ) = 4,67589

σ = 0,00160

int. disp. = 0,006



Rappresentazione

media (μ) = 4,67589 **σ = 0,00160**

valore min. = 4,673 **valore mass. = 4,679**

intervallo dispersione (4,679-4,673) = 0,006

affidabilità del 99,4% (f=3,0)

ε = 3,0•0,00160 = 0,00480

pH = 4,67589 ± 0,00480 **μ±3σ**

pH = 4,676 ± 0,003 **mediana ± ½int.disp.**

Un dato viene rappresentato da due valori:

stima del val. centrale (media, mediana, …)

stima della dispersione (dev. stand., errore, …).

Rappresentazioni

Il numero di **cifre significative** che identificano l'errore è limitato a **1** o **2**.

pH = 4,67589 ± 0,00480

pH = 4,6759 ± 0,0048

pH = 4,676 ± 0,005

L'errore si può riportare in parentesi

pH = 4,67589(480)

pH = 4,6759(48)

pH = 4,676(5)

La rappresentazione più usata (anche se meno soddisfacente) è mediante la convenzione delle cifre significative.

Cifre significative

Cifre significative: cifre che hanno significato in quanto sono quelle effettivamente registrate dallo strumento di misurazione, compatibilmente con la sua sensibilità.

Quanto maggiore è la sensibilità dello strumento, tanto più alto è il numero delle cifre significative le quali ci danno il grado di accuratezza di una misura.

<u>L'ultima cifra</u> indicata dalla misurazione viene definita **cifra incerta** in quanto rappresenta un valore approssimato (per eccesso o per difetto) e dipende dalla sensibilità dello strumento.

Cifra incerta

La **cifra incerta** è sempre l'ultima a essere scritta compreso lo **zero**.

Il numero di cifre significative in una misura è uguale a:

1. **gli interi non nulli**
2. **gli zeri compresi tra cifre non nulle**
3. **gli zeri finali solo se appartengono a numeri decimali**

($1{,}00$ / $1{,}00\bullet10^2$ / $8{,}050$ / $0{,}060$ / $12{,}30\bullet10^{-2}$)

Non sono da considerare gli **zeri finali dei numeri interi** (100 / 10 / 1000) se hanno valore di cifra significativa il numero va rappresentato con la notazione esponenziale

($1{,}00\bullet10^2$ / $1{,}0\bullet10$ / $1{,}000\bullet10^3$)

Esempio

Considerando l'esempio precedente (anello)

Sensib. bilancia (g)	Pesata anello (g)	Misura (g)	Cifre note	Cifra incerta	N. cifre Signif.
5	5<…<10	7		7	1
1	7<…<8	7,5	7	5	2
0,01	7,35	7,35	7,3	5	3
0,0001	7,3538	7,3538	7,353	8	5

Arrotondamento

1) **Se la prima cifra da eliminare è minore di 5, lasciare la cifra precedente così com'è (arrotondamento per difetto).**

Es. 5,9936 approssimato a un centesimo (tre cifre significative) 5,99

2) **Se la prima cifra da eliminare è maggiore di 5, incrementare di 1 la cifra precedente (arrotondamento per eccesso).**

Es. 34,581 approssimato a un decimo (tre cifre significative) 34,6

3) **Se la prima cifra da eliminare è 5 seguita da almeno una cifra diversa da 0 incrementare di 1 la cifra precedente (arrotondamento per eccesso).**

Es. 6319,4501 approssimata a un centesimo (sei cifre significative) 6319,46

4) **Se la prima cifra da eliminare è 5 seguita solo da zeri, si arrotonda per difetto, se la cifra precedente è pari, e per eccesso, se è dispari.**

Es. 47,3500 approssimata a un decimo (tre cifre significative) 47,4
68,850 approssimata a un decimo (tre cifre significative) 68,8

Cifre significative

Moltiplicazioni e divisioni: **il numero di cifre significative del risultato è uguale al numero di cifre significative del dato di partenza, usato nel calcolo, che ne ha meno.**

Es. $4{,}61 \times 1{,}4 = 6{,}5$ (riferito a 1,4)

$1{,}06\bullet10^{-3} : 6{,}185 = 1{,}71\bullet10^{-4}$ (riferito a $1{,}06\bullet10^{-3}$)

$2{,}86\bullet10^{4} \times 3{,}163\bullet10^{-2} : 1{,}8 = 5{,}0\bullet10^{2}$ (riferito a 1,8)

Cifre significative

Addizioni e sottrazioni: **il risultato ha lo stesso numero di cifre decimali del dato meno preciso usato nel calcolo.**

12,11 + 18,0 + 1,013 = 31,123 = 31,1 (riferito a 18,0)

21 – 13,8 = 7 (riferito a 21)

1,3 +	***1,30**** +
2,46=	2,46 =
3,76 = 3,8	3,76 = 3,76

*lo zero ci indica una misurazione effettuata con una sensibilità dello strumento dello **0,01** (centesimo dell'unità di misura); il centesimo calcolato, in questo caso è **zero**.

Anche lo zero è un numero!

Notazione esponenziale

n = C•10^e

n = **numero da convertire in notazione espon.**

C = **coefficiente**

10^e = **esponenziale**

e = **esponente = numero intero positivo o negativo**

Nella notazione esponenziale standard il coefficiente (C) è sempre maggiore o uguale a 1 e minore di 10.

Esempi

La massa dell'elio (He) è di:

0,00000000000000000000000664 g

$6,64 \cdot 10^{-24}$ g

In un litro di elio a 0 °C e 1 atm di pressione ci sono:

26.880.000.000.000.000.000.000 atomi

$2,6888 \cdot 10^{22}$ atomi

Il numero di Avogadro corrisponde a:

602.214.000.000.000.000.000.000

$6,02214 \cdot 10^{23}$

$6,022 \cdot 10^{23}$