

## COS'È LA CHIMICA ANALITICA?

la chimica analitica è un approccio chimico-scientifico alla base di moltissime discipline. È quella branca della chimica che consiste nella determinazione **qualitativa** e **quantitativa** dei componenti di un determinato campione.

l'analisi che la chimica analitica opera è:

- 1) **qualitativa** = **identifica/riconosce l'entità** della sostanza chimica di interesse (analita) in un determinato campione
  - 2) **quantitativa** = **quantifica** la sostanza chimica di interesse identificata dall'analisi qualitativa
- L'analisi quantitativa è sempre accompagnata e preceduta dall'analisi qualitativa

### ANALISI QUANTITATIVA

Nell'analisi quantitativa si possono distinguere diverse tappe:

- 1) formulare una domanda su cosa voglio determinare, ossia **decidere l'ANALITA, ossia che sostanza voglio analizzare** di un determinato campione
- 2) **scegliere il METODO** di separazione della sostanza da analizzare: come?
  - esperienza
  - ricerca → PubMed (posso inserire una parola chiave che mi rimanderà a tutti gli articoli che la contengono per decidere che metodo utilizzare)
- 3) **scegliere il CAMPIONE** ("campionamento o sampling") e la sostanza da analizzare
- 4) **preparare il CAMPIONE**
  - la **matrice**, ossia tutto ciò che è al di fuori della sostanza da analizzare, potrebbe contenere sostanze che interferiscono con l'analisi dell'analita, gli **interferenti**: bisogna eliminarle o mascherarle (renderle non più interferenti).
- 5) **analizzare/misurare** le proprietà di interesse tramite il metodo scelto
- 6) valutare il dato sperimentale, facendo un **report** dei risultati ottenuti

### I METODI DELL'ANALISI QUANTITATIVA

I metodi dell'analisi quantitativa sono:

- ✓ **gravimetrico** = sfrutta la **precipitazione completa di un sale** per ottenere la massa dell'analita (es. filtrazione)
- ✓ **volumetrico** = sfrutta il **volume e la concentrazione di una soluzione** che reagisce completamente con l'analita per ottenere determinate informazioni dell'analita (es. titolazione)
- ✓ **elettroanalitico** = sfrutta le **proprietà elettriche** (potenziale, corrente, resistenza e quantità di carica elettrica) per ottenere determinate informazioni dell'analita (es. titolazione redox)
- ✓ **spettroscopico** = sfrutta l'**interazione** tra una **radiazione elettromagnetica** e gli **atomi** o le molecole di analita per ottenerne determinate informazioni (es. assorbimento nell'uv visibile)
- ✓ **spettrometria di massa** = sfrutta il rapporto **massa/carica** dell'analita per ottenerne determinate informazioni

### **esempio: ANALISI DI UNA BARRETTA DI CIOCCOLATO**

- 1) Voglio determinare la caffeina in una barretta di cioccolato
- 2) Faccio una ricerca su quali metodi analitici devo utilizzare per separare la caffeina dal cioccolato: il migliore è la cromatografia.
- 3) Seleziono un campione da analizzare, quindi una barretta di cioccolato. È preferibile prendere una barretta di cioccolato puro in quanto si rivela un campione abbastanza omogeneo.
- 4) Preparo il campione riducendo in piccole parti il cioccolato e poi rimuovendo o mascherando le specie che interferiscono con l'analisi chimica dell'analita, la caffeina: per una barretta di cioccolato si rimuovono i grassi tramite centrifugazione. Poi si preleva la soluzione filtrata contenente l'analita pronto per la cromatografia.
- 5) Si inietta la soluzione in una colonna cromatografica che separa l'analita e ne misura la quantità. Si misura la concentrazione dell'analita in molte identiche aliquote in modo da valutare l'incertezza

della misurazione e scongiurare grossi errori nell'analisi. Viene costruita una curva di calibrazione, ossia un grafico che illustra la concentrazione dell'analita rispetto a soluzioni standard contenenti concentrazioni note di caffeina pura. Così è possibile trovare la concentrazione della caffeina nel cioccolato

6) Viene fatto un attento report dei risultati e si traggono le conclusioni dell'analisi.

## GLI ERRORI NELLE ANALISI CHIMICHE

Tutte le misure sperimentali sono tutte affette da errori, grandi o piccoli a seconda del metodo o dal modo di operare dell'operatore. Bisogna tenere conto degli errori e, in base a questi, vedere se i dati sono validi per l'analisi.

### MISURE SPERIMENTALI

È importante ripetere più volte una misura per capire quale è il valore che si avvicina di più al valore reale. Le misure ripetute vengono chiamate replicati. I replicati possono essere:

- di campioni diversi → replicati analitici
- dello stesso campione → replicati tecnici

Il dato che è più probabile essere quello giusto/affidabile è il valore centrale. Qui possiamo distinguere due parametri statistici molto importanti:

- ✓ MEDIA = sommatoria dei dati diviso il numero di misurazioni effettuate.
- ✓ MEDIANA = valore centrale di un insieme di dati che sono stati ordinati in ordine crescente. Se abbiamo un numero pari di dati, allora la mediana corrisponde alla media dei due centrali. Se abbiamo un numero dispari di dati, allora la mediana corrisponde al valore centrale.

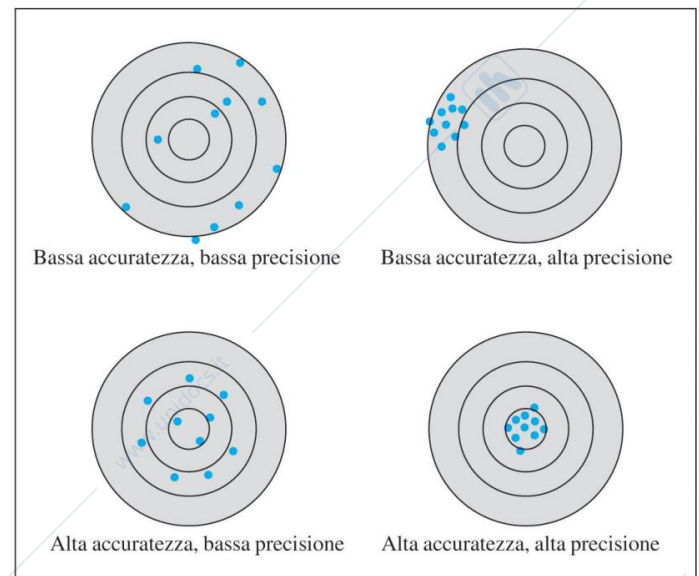
Questi parametri sono chiamati indici di posizione e sono i valori centrali intorno ai quali si distribuiscono tutti i dati. Ci fanno notare se sono presenti outliers, ossia dei valori distanti dalle altre osservazioni disponibili.

### PRECISIONE e ACCURATEZZA

Ogni misura è influenzata da incertezze, che si combinano producendo una dispersione dei risultati. In una serie di misurazioni si utilizzano i termini:

- PRECISIONE** = riproducibilità delle misure. Si ha una serie di misurazioni precisa se i valori non si disperdono tanto, ossia sono tutti simili tra loro. Può essere descritta dalla deviazione standard, dalla varianza e dal coefficiente di variazione
- ACCURATEZZA** = vicinanza delle misurazioni al valore reale o accettabile. Si ha una serie di misurazioni accurata se i valori sono vicini al valore reale ottenuto con un altro metodo già ben accreditato o a un valore di letteratura. Può essere descritta da

- errore assoluto: differenza tra il valore sperimentale  $X_i$  e il valore vero/accettabile  $X_t$   $E = X_i - X_t$
- errore relativo: rapporto tra l'errore assoluto  $X_i - X_t$  e il valore vero/accettabile  $X_t$  moltiplicato per 100  $Er = \frac{X_i - X_t}{X_t} \cdot 100$



### TIPI DI ERRORI NEI DATI SPERIMENTALI

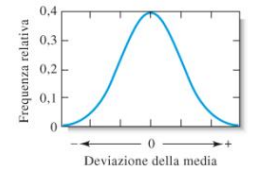
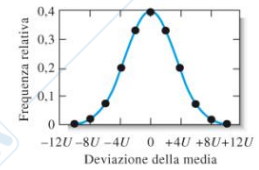
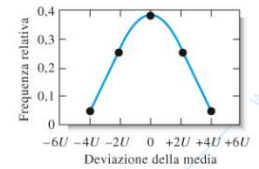
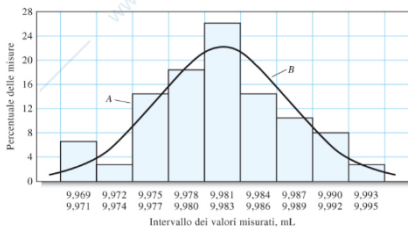
Esistono 3 tipi di errori assoluti nei dati sperimentali, la cui descrizione ci permette di analizzare la precisione e l'accuratezza delle misurazioni fatte:

- CASUALE** o indeterminato = è un errore di cui tutte le misure sperimentali sono affette e non può essere eliminato. È causato da molte variabili come incontrollate che accompagnano ogni misurazione. Determina una dispersione dei dati più o meno simmetrica intorno al valore medio. Influisce sulla precisione e non sulla accuratezza.

Come si rileva un errore casuale? È possibile rilevare un errore casuale calcolando la frequenza con cui si ha un determinato dato e quanto questo dista dal valore medio tramite la deviazione

standard. Si otterrà sempre una distribuzione dei dati "più fitta" intorno al valore medio e "meno fitta" lontana dal valore medio.

➔ Se costruisco un istogramma in cui sull'asse delle ascisse pongo la deviazione standard e sull'asse delle ordinate la frequenza con cui compaiono i relativi dati si ottiene una **DISPERSIONE NORMALE** con una particolare forma a campana o **CURVA GAUSSIANA** dell'errore. La curva è simmetrica rispetto a un punto centrale (punto medio). In questa curva i dati più attendibili sono quelli centrali e si verifica una decrescita esponenziale della frequenza all'aumentare dell'entità delle deviazioni. Maggiore è il numero di dati che si prende in considerazione, maggiore è l'attendibilità dei risultati.



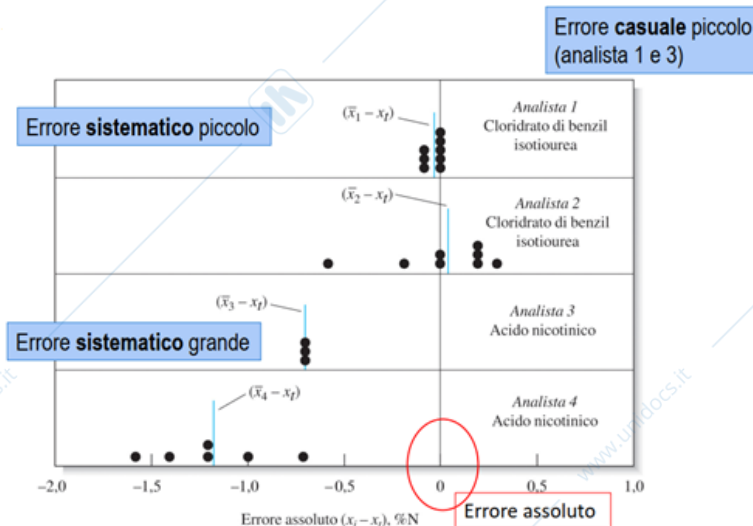
La curva gaussiana ci permette di comprovare se i dati sono precisi e accurati: ci permette di rilevare se in una serie di misurazioni è presente l'errore casuale.

In quasi tutti gli esperimenti analitici quantitativi la distribuzione di dati replicati è rappresentabile tramite una curva Gaussiana.

- 2) **GROSSOLANO** = è un errore dato dalla mancata attenzione da parte dell'operatore e può essere eliminato se ci si rende conto. Determina la comparsa di dati distanti dalla media.
- 3) **SISTEMATICO** o determinato = è un errore che si verifica continuamente e può essere eliminato. Determina la comparsa di dati vicini alla media ma la media della serie di dati differisce dal valore reale. Influisce sull'accuratezza e non sulla precisione. In questa categoria si possono distinguere 3 tipi di errori
  - ✓ Di metodo = dovuto ad un comportamento chimico o fisico non adeguato come lentezza o incompletezza di alcune reazioni, presenza reazioni secondarie che interferiscono con il processo di misura
  - ✓ Personali = errore del parallasse
  - ✓ Strumentali = dovuto ad uno strumento mal calibrato, malfunzionante ecc. Possono essere costanti (indipendenti dalla grandezza del campione da analizzare) o proporzionali (aumentano o diminuiscono in modo proporzionale alla quantità di campione)

Come rilevare l'errore sistematico?

- Analisi di campioni standard = confronto del valore con un campione standard, ossia una sostanza a titolo noto (di cui si conosce la concentrazione)
- Analisi indipendenti = confronto del valore con un valore ottenuto utilizzando un secondo metodo analitico in parallelo
- Analisi del bianco o della matrice del campione = analisi del solvente senza campione
- Variazione della quantità del campione = raddoppiare/triplicare ecc la quantità del campione e analizzare il risultato



## PROPAGAZIONE DELL'ERRORE

L'errore esprime la distanza tra il valore noto e il valore sperimentale. Per esprimere una misura con il suo relativo errore si scrive la misura **m (± errore assoluto)**. L'errore decide quante cifre decimali prendere della misura.

Es. 13,33 (± 0,02) → vuol dire che la misura è compresa tra 13,31 e 13,35 mL.

Quando ci sono più misure con la relativa incertezza (nella maggior parte degli esperimenti) allora l'errore viene propagato. Delle formule utili ai problemi sulla propagazione dell'errore sono

$$E_{rel} = \frac{E_{ass}}{\text{valore}} \quad E_{rel}\% = 100 \cdot E_{rel}$$

Per calcolare la propagazione dell'errore:

a) Nella somma e differenza si utilizza l'errore assoluto  $E_{ass} = \sqrt{e_1^2 + e_2^2}$

b) Nella divisione e moltiplicazione si utilizza l'errore relativo percentuale  $E\% = \sqrt{e_1\%^2 + e_2\%^2}$

c) Nel logaritmo si utilizza l'errore assoluto -  $E_{ass} = \frac{1}{\ln 10} \cdot \frac{ex}{x}$  dove  $y = \log_{10}x$

$$- E_{ass} = \frac{ex}{x} \quad \text{dove } y = \ln x$$

d) nell'esponenziale si utilizza - l'errore relativo percentuale  $E\% = a \cdot e_x$  dove  $y = x^a$

- l'errore relativo  $E_{rel} = e_x$  dove  $y = e^x$  /  $E_{rel} = (\ln 10)e_x$  dove  $y = 10^x$

Per calcolare la propagazione dell'errore sulla massa molecolare di un composto, per primo moltiplichiamo le masse atomiche di ciascun atomo e le loro incertezze per il numero di quegli atomi nel dato composto. Poi sommiamo le masse atomiche ottenute per trovare la massa molecolare. Per le incertezze utilizziamo la formula di propagazione dell'errore della somma/differenza (vedi slide per esercizio).

## CIFRE SIGNIFICATIVE

Le cifre significative di un numero sono tutte le cifre più la prima cifra incerta, ossia la cifra su cui ricade l'errore. Il numero di cifre significative è legato, quindi, all'incertezza della misura e l'arrotondamento di un risultato deve tener conto della probabile incertezza associata alla misura.

Le regole per determinare le cifre significative sono:

- 1) Gli zeri iniziali non sono significativi
- 2) Gli zeri finali non sono significativi purché non seguano una virgola decimale
- 3) Tutte le altre cifre sono significative inclusi gli zero compresi tra cifre diverse da zero
- 4) Quando la prima cifra del risultato è tra 1 e 2 prendo sempre una cifra significativa in più per ridurre l'errore

Nei calcoli si seguono le seguenti regole:

- Somma e differenza → risultato ha lo stesso numero di cifre decimali del numero con meno cifre decimali
- Prodotto e quoziente → risultato ha lo stesso numero di cifre significative del numero con meno cifre significative
- Logaritmi e esponenziali → risultato ha un numero di cifre decimali pari al numero di cifre significative del numero originale

## ARROTONDAMENTI

Bisogna sempre tenere più cifre del necessario durante i calcoli e poi arrotondare al numero appropriato alla fine, utilizzando il giusto numero di cifre significative e decimali. L'arrotondamento segue delle regole, ossia bisogna guardare la cifra dopo l'ultima cifra desiderata e:

- a) Se il numero è 5 o maggiore di 5 → si arrotonda per eccesso
- b) Se il numero è minore di 5 → si lascia invariato il valore della cifra

## I METODI STATISTICI NELLA CHIMICA ANALITICA

La statistica viene utilizzata nella chimica analitica perché consente il giudizio critico nei confronti dei dati sperimentali.

Inoltre, permette di verificare delle ipotesi per sviluppare modelli e descrivere i risultati sperimentali. I metodi statistici ci permettono di raggruppare i dati in categorie, di caratterizzarli in diversi modi e di prendere decisioni obiettive ed intelligenti circa la loro qualità ed interpretazione.

- ✓ POPOLAZIONE (concettuale) = insieme infinito di misure di interesse
- ✓ CAMPIONE (reale) = sottoinsieme che rappresenta l'idea popolazione: si possono dedurre le caratteristiche della popolazione dalle caratteristiche del campione (la media dei dati sperimentali del campione è uguale alla media dei dati sperimentali della popolazione se si aumentano il numero di misure che costituiscono il campione)

## LA CURVA GAUSSIANA

La curva gaussiana ci permette di trattare statisticamente l'errore casuale. Rappresenta i dati di ogni esperimento analitico quantitativo, che presentano necessariamente errori casuali. È una distribuzione simmetrica dei dati intorno alla media.

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

L'espressione della curva gaussiana è

in cui  $x$  = variabile: le misure sperimentali  
 $\mu$  = media popolazione  
 $\sigma$  = deviazione standard

La deviazione standard indica la distanza dei dati dal valore medio, è un indice della precisione di una popolazione di dati. È data dalla formula:

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

Questi permettono di descrivere la curva a seconda dei valori della deviazione standard, della media e del dato sperimentale.

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

Esiste anche  $z$ , ossia la deviazione relativa di un dato dalla media della popolazione: la **DEVIAZIONE RELATIVA** alla deviazione standard. Si ottiene con la formula a fianco. In questa equazione sono racchiusi i parametri più importanti della curva gaussiana.

- ✓ Se  $x - \mu = \sigma$ , allora  $z = \sigma$
- ✓ Se  $x - \mu = 2\sigma$ , allora  $z = 2\sigma$  ...

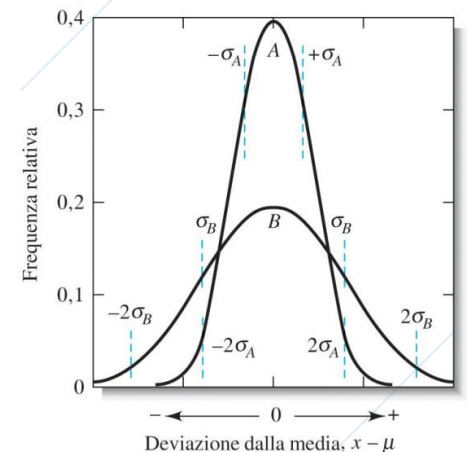
Se invece della deviazione standard sull'asse delle  $x$  viene posto  $z$ , in funzione di  $z$  si producono diverse curve alte-basse-strette-larghe che descrivono TUTTE le popolazioni di dati, indipendentemente dalla loro  $\sigma$ .  $z$  ci serve a capire in quale intervallo intorno alla media  $\mu$  si trovano i dati del nostro esperimento analitico

- Se la curva è stretta e alta  $\rightarrow$  accettabile: tutti i valori sono vicini al valore medio
- Se la curva è bassa e larga  $\rightarrow$  poco accettabile: i valori sono più distanti dal valore medio

Gli intervalli in cui si possono trovare i dati sono aree sottese dalla curva gaussiana e sono:

- 68,3%  $\rightarrow \pm\sigma$
- 95,4%  $\rightarrow \pm 2\sigma$
- 99,7  $\rightarrow \pm 3\sigma$

Tutto ciò era riferito alla popolazione...MA se ci riferiamo ad un campione?



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

Media del campione

Numero dei gradi di libertà

Per calcolare la deviazione standard di un **CAMPIONE** utilizzo la formula a fianco.

Il numero dei gradi di libertà è il numero di dati indipendenti (quelli non prevedibili dal resto delle analisi): è  $N-1$  ossia il numero dei dati meno uno perché l'ultimo dato può essere ricavato dai precedenti.  $\bar{x}$  è la media del campione, ossia la media aritmetica di un certo numero di dati presi dalla popolazione.

La VARIANZA del campione è il quadrato della deviazione standard  $s^2$ . È una stima della varianza della popolazione, che si ottiene con  $\sigma^2$ .

Altri modi con cui calcolare la deviazione standard sono:

- deviazione standard relativa (DSR):  $s_r = \frac{s}{\bar{x}}$
- deviazione standard relativa percentuale (CV):  $CV = \frac{s}{\bar{x}} \cdot 100$   
(varia in funzione della quantità di campione a parità di deviazione standard)

La DISPERSIONE o range è un altro indice della precisione di un insieme di risultati replicati ed è la differenza tra il valore più grande e quello più piccolo dell'insieme di dati:  $W = X_{max} - X_{min}$

La stima statistica è quella stima di un parametro che viene eseguita su un campione di dati. Quindi se  $\mu$  e  $\sigma$  vengono riferiti ad un campione sono stime statistiche.

### INTERVALLO DI FIDUCIA e LIMITI DI FIDUCIA

L'intervallo di fiducia è un intervallo intorno alla media  $\mu$  nel quale ci si aspetta di trovare la media  $m$  della popolazione. Le linee di confine di tale intervallo sono chiamati limiti di fiducia.

Il livello di fiducia ci dice quanta probabilità ci sia che la media  $\mu$  si trovi in un certo intervallo di fiducia.

È utilizzato per capire quanti dati possiamo accettare oltre quelli più vicini alla media nell'intervallo  $\pm\sigma$ .

Questo perché nella maggior parte delle situazioni incontrate in chimica analitica il valore vero della media  $\mu$  non può essere determinato poiché richiederebbe un enorme numero di misure.

L'ampiezza dell'intervallo di fiducia è legato alla deviazione standard  $s$  come la deviazione standard  $\sigma$  della popolazione:

- ✓ se  $s$  è una buona stima di  $\sigma$  o è nota  $\rightarrow$  l'intervallo di fiducia IF per la media  $\bar{x}$  di  $N$  misure replicate si può dire che è  $IF = \bar{x} \pm \frac{z\sigma}{\sqrt{N}}$ .
- ✓ se  $s$  non è una buona stima di  $\sigma$  o non è nota  $\rightarrow$  l'intervallo di fiducia IF per la media  $\bar{x}$  di  $N$  misure replicate si può dire che è  $IF = \bar{x} \pm \frac{ts}{\sqrt{N}}$ .

$t$  è chiamato  $t$  statistico e si ottiene con la formula

$$t = \frac{\bar{x} - \mu}{s}$$

e per la media di  $N$  misure

$$t = \frac{\bar{x} - \mu}{s/\sqrt{N}}$$

Notiamo dalla regola che se aumentiamo  $N$ , quindi il numero delle misurazioni, l'errore diminuisce.

Solitamente l'intervallo di fiducia preso è quello del  $\pm 2\sigma$  con un livello di fiducia del 95%.

### T TEST e Z TEST per media sperimentale/media nota

Per spiegare un'osservazione si IPOTIZZA un modello teorico che viene verificato dai TEST DI VERIFICA che sono:

- a) test  $t$  = test che confronta le medie di due campioni. In questo test NON si conosce  $\sigma$ .
- b) test  $z$  = test che confronta le medie di due popolazioni. In questo test si conosce  $\sigma$ .

Questi test si basano su ipotesi di uguaglianza (ipotesi nulla) e disuguaglianza (ipotesi alternativa) tra il risultato sperimentale del modello teorico e il risultato noto, verificando se quello noto conferma quello

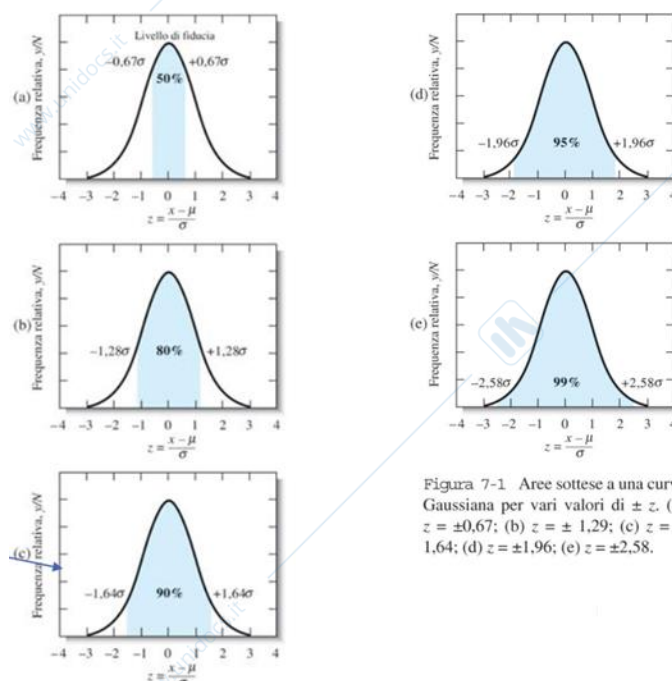


Figura 7-1 Aree sottese a una curva Gaussiana per vari valori di  $\pm z$ . (a)  $z = \pm 0,67$ ; (b)  $z = \pm 1,29$ ; (c)  $z = \pm 1,64$ ; (d)  $z = \pm 1,96$ ; (e)  $z = \pm 2,58$ .

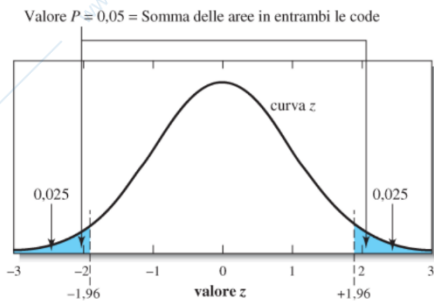
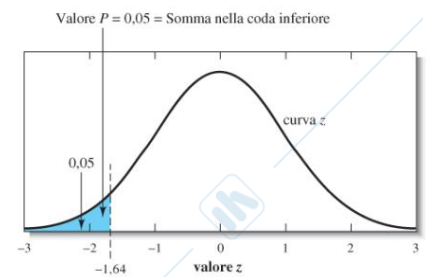
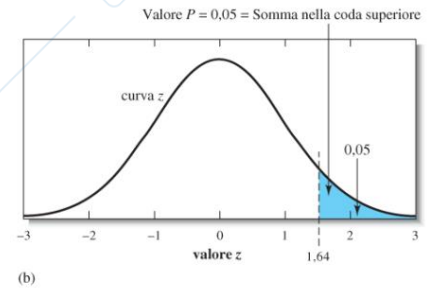
sperimentale: è una verifica statistica dell'ipotesi che  $\mu_0$  (media nota) è uguale a  $\mu$  (media sperimentale). Se il risultato sperimentale è uguale a quello noto allora il modello teorico contiene solo errori casuali ed è accettabile per ulteriori esperimenti MENTRE se il risultato sperimentale non è uguale a quello noto allora contiene errori sistematici e non è accettabile quindi deve essere elaborato un nuovo modello con nuove ipotesi oppure devono essere controllati i calcoli.

RIASSUMENTO, i test su popolazioni si basano sull'impostazione delle ipotesi:

- IPOTESI NULLA  $H_0: \mu_0 = \mu$
- IPOTESI ALTERNATIVA  $H_a$ : 1)  $\mu_0 \neq \mu \rightarrow$  si scarta l'ipotesi nulla se  $z \geq z_{critic}$  o se  $z \leq -z_{critic}$   
 2)  $\mu_0 > \mu \rightarrow$  si scarta l'ipotesi nulla se  $z \geq z_{critic}$   
 3)  $\mu_0 < \mu \rightarrow$  si scarta l'ipotesi nulla se  $z \leq -z_{critic}$

$z_{critic}$  è il **livello di significatività**  $\alpha$  che è una soglia che determina se un risultato di uno studio possa essere considerato statisticamente significativo. È legato al livello di fiducia: se il livello di fiducia è 95%, il livello di significatività è 5% in quanto è tutto ciò che sta al di fuori dell'intervallo di fiducia.

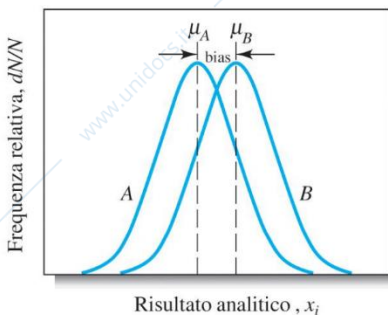
Quindi, il 95% sarà quel range in cui è valida l'ipotesi nulla MENTRE il 5% sarà quel range in cui è valida l'ipotesi alternativa. Se trasformiamo il 5% in 0,05 allora chiamiamo il livello di significatività "P value".



Il test può essere a una coda o a due code: è a una coda se i valori al di fuori del livello di fiducia sono solo a una estremità della curva di gauss (nel caso 2 e 3), è a due code se i valori al di fuori del livello di fiducia sono a entrambe le estremità della curva di Gauss (nel caso 1).

Oltre che su popolazioni, si può applicare il test t su campioni con  $\sigma$  non noto. In questo caso si usa il t statistico e quindi si avrà:

- IPOTESI NULLA  $H_0: \mu_0 = \mu$
- IPOTESI ALTERNATIVA  $H_a$ : 1)  $\mu_0 \neq \mu \rightarrow$  si scarta l'ipotesi nulla se  $t \geq t_{critic}$  o se  $t \leq -t_{critic}$   
 2)  $\mu_0 > \mu \rightarrow$  si scarta l'ipotesi nulla se  $t \geq t_{critic}$   
 3)  $\mu_0 < \mu \rightarrow$  si scarta l'ipotesi nulla se  $t \leq -t_{critic}$



Se nell'analisi di campioni le misurazioni fossero affette (oltre che da errori casuali) da ERRORI SISTEMATICI o BIAS allora avremo curve che si discostano dalla curva del risultato noto e quindi non affetto da errore sistematico. La curva non affetta da errore sistematico (A) differisce dalla curva con errore sistematico (B) di un valore pari al Bias =  $\mu_B - \mu_0$ . Per capire quale è la curva che effettivamente è affetta da errore sistematico faccio un confronto tra il Bias creato con  $\mu_B$  e con  $\mu_A$ : quello più lontano da  $\mu_0$  è quello con errore sistematico.

### T TEST e Z TEST per 2 medie sperimentali

Se il confronto non avvenisse tra una media nota e una sperimentale MA tra due medie sperimentali allora si dovrebbe:

- valutare la differenza tra le due medie e valutare se tale differenza sia risultato di errori casuali
- valutare se due metodi analitici diversi forniscono gli stessi valori
- valutare se usando gli stessi metodi si ottengono le stesse medie

Nei test t per il confronto tra due medie sperimentali si prendono in considerazione la media  $\mu_1$  dell'insieme di misure 1 e la media  $\mu_2$  dell'insieme di misure 1 ottenute da due analisti diversi:

- IPOTESI NULLA  $H_0: \mu_1 = \mu_2$
  - IPOTESI ALTERNATIVA  $H_a$ :
    - 1)  $\mu_1 \neq \mu_2$
    - 2)  $\mu_1 > \mu_2$
    - 3)  $\mu_1 < \mu_2$
- si scarta l'ipotesi nulla se  $t \geq t_{critic}$   
 perché grande differenza medie  
 - si accetta l'ipotesi nulla se  $t \leq t_{critic}$   
 perché piccola differenza tra medie

Assumiamo che le deviazioni standard  $s_1$  e  $s_2$  dei due insiemi di misure siano simili e che siano buone stime della deviazione standard  $\sigma$  della popolazione.

A questo punto il valore di t CAMBIA per il t test della differenza tra due medie sperimentali:

come si calcola t?

per calcolare t abbiamo bisogno delle due medie, già date, e della deviazione standard relativa alla differenza tra le medie.

$$\frac{s_d}{\sqrt{N}} = \sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}$$

Per calcolare la deviazione standard della differenza tra le medie si usa la formula

$$s_{m1}^2 = \frac{s_1^2}{N_1} \quad s_{m2}^2 = \frac{s_2^2}{N_2}$$

in quanto le varianze dell'operatore 1 e 2 erano

Per avere una migliore stima di  $\sigma$  rispetto a quella data da  $s_1$  o  $s_2$  da sole, usiamo la deviazione standard cumulata.

$$s_{cumulata} = \sqrt{\frac{\sum_{i=1}^{N_1} (x_i - \bar{x}_1)^2 + \sum_{j=1}^{N_2} (x_j - \bar{x}_2)^2 + \sum_{k=1}^{N_3} (x_k - \bar{x}_3)^2 + \dots}{N_1 + N_2 + N_3 + \dots - N_t}}$$

il quadrato delle deviazioni dalla media relative ad ogni sottoinsieme

Numero dei gradi di libertà

( $N_t$  è il numero totale di insieme di dati che devono essere cumulati, quindi nel nostro caso è 2)

Da ciò deriva che la deviazione standard che ci serve per il calcolo di t, inserendo la deviazione standard cumulata (perché migliore stima di  $\sigma$ ) nella formula della deviazione standard della differenza tra due

medie, è:

$$\frac{s_d}{\sqrt{N}} = \sqrt{\frac{s_{cumulata}^2}{N_1} + \frac{s_{cumulata}^2}{N_2}} = s_{cumulata} \sqrt{\frac{N_1 + N_2}{N_1 N_2}}$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{cumulata} \sqrt{\frac{N_1 + N_2}{N_1 N_2}}}$$

L'espressione di t diventa:

### esempio

In una investigazione forense è stato analizzato il contenuto di alcool di un bicchiere contenente vino rosso e di una bottiglia aperta al fine di determinare se il vino nel bicchiere provenisse da essa. Sulla base di 6 analisi, è stato stabilito che il contenuto medio in etanolo del vino nel bicchiere è 12,61%. Per il vino nella bottiglia, la media di 4 analisi è risultata pari al 12,53% di alcool. Le 10 analisi hanno prodotto una deviazione standard cumulata  $s_{cumulata}$  pari allo 0,070%. Sulla base di questi dati, è possibile affermare che c'è una differenza tra i due vini?

L'ipotesi nulla è  $H_0: \mu_1 = \mu_2$  e l'ipotesi alternativa è  $H_a: \mu_1 \neq \mu_2$ .

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{cumulata} \sqrt{\frac{N_1 + N_2}{N_1 N_2}}} = \frac{12,61 - 12,53}{0,07 \sqrt{\frac{6 + 4}{6 \times 4}}} = 1,771$$

Il valore critico di t al livello di fiducia del 95% per  $10 - 2 = 8$  gradi di libertà è 2,31. Poiché  $1,771 < 2,31$ , accettiamo l'ipotesi nulla al livello di fiducia del 95% e concludiamo che non c'è differenza nel contenuto alcolico dei vini.

### ERRORI NELLA VERIFICA DELLE IPOTESI

Anche nel caso di verifica delle ipotesi nulla può accadere che si commettano errori. Possiamo distinguere gli errori in

- Errori di TIPO 1:  $H_0$  viene scartata anche se vera a causa della scelta dell'intervallo di fiducia troppo stretto → FALSO NEGATIVO
- Errori di TIPO 2:  $H_0$  è accettata anche se falsa a causa della scelta dell'intervallo di fiducia troppo ampio → FALSO POSITIVO

## TEST F

È un test fatto prima del test T che serve per verificare che l'assunzione delle deviazioni standard come simili sia giusta.

È basato sull'IPOTESI NULLA  $H_0: \sigma^2_1 = \sigma^2_2$  e sull'IPOTESI ALTERNATIVA  $\sigma^2_1 > \sigma^2_2$  (test a una coda) oppure  $\sigma^2_1 \neq \sigma^2_2$  (test a due code).

F è definito come il rapporto delle due varianze  $s^2_1/s^2_2$  che sono una buona stima di  $\sigma_1$  e  $\sigma_2$  e deve essere confrontato con il valore critico  $F_{critic}$  al livello di significatività desiderato (di solito 5%) consultabile nelle tabelle.

## ANOVA

L'ANOVA o analisi della varianza è utilizzata per i confronti multipli ossia su più gruppi, quindi tra più di due medie di popolazioni.

Questi metodi utilizzano un singolo test per determinare se esiste, oppure no, una differenza tra le medie di popolazioni invece di utilizzare confronti a coppia come viene fatto nel test t. Dopo di ciò, identifica quali particolari medie di popolazioni differiscono dalle altre tramite procedure di confronto multiplo.

È utile se dobbiamo confrontare:

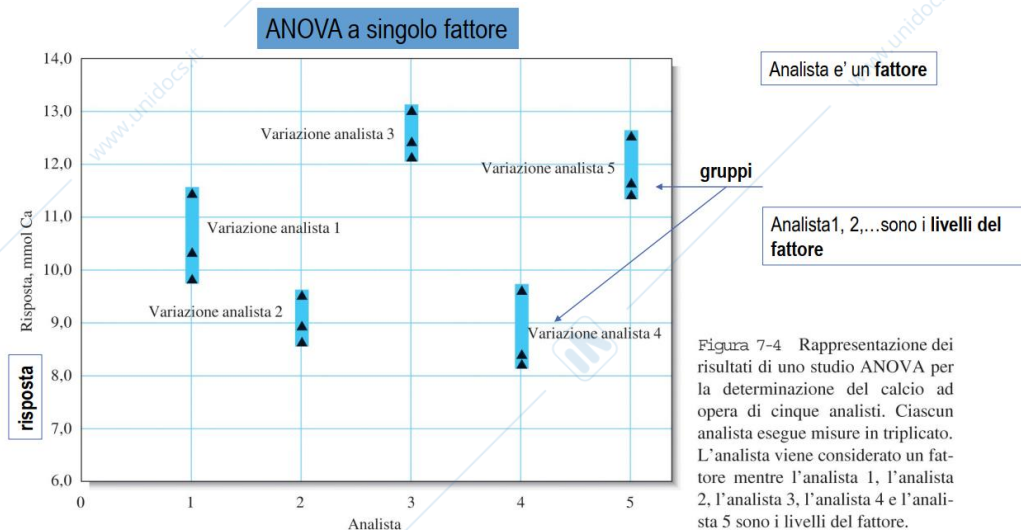
- un unico metodo sperimentale tra più laboratori.
- Più di due metodi sperimentali in un laboratorio.

Anche ANOVA si basa su: - IPOTESI NULLA  $H_0: \mu_1 = \mu_2 = \mu_3 \dots$

- IPOTESI ALTERNATIVA  $H_a$ : almeno due dei valori  $\mu_i$  sono differenti

3 sono le variabili su cui verte il metodo ANOVA:

- ✓ Fattore = caratteristica comune
- ✓ Livelli o Gruppi = valori diversi della caratteristica comune da confrontare
- ✓ Risposta = misura che quantifica il confronto dei diversi valori



Il fattore è la variabile indipendente e la risposta è la variabile dipendente. Per capire meglio questi 3 concetti possiamo porci una domanda da descrivere con ANOVA: c'è una differenza nei risultati di 5 analisti che determinano il calcio mediante un metodo volumetrico? in questo caso il fattore è

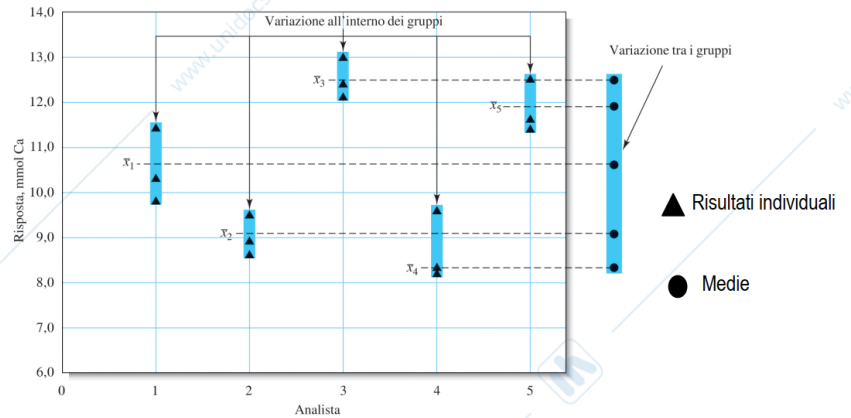
l'analista, i livelli sono analista 1/analista 2/analista 3/analista 4/analista 5. La risposta sono le moli di calcio determinate da ciascun analista.

Per applicare i metodi ANOVA abbiamo bisogno di fare una principale assunzione inerente le popolazioni oggetto di studio: le varianze delle popolazioni devono essere identiche. Questa assunzione si può

verificare confrontando le varianze massima e minima mediante un test F oppure controllando che il valore più grande di s non sia grande più di due volte il valore più piccolo di s al cubo (s)<sup>3</sup>.

**Il principio base di ANOVA è quello di confrontare la variazione tra i gruppi con quelle all'interno dei gruppi:**

- Quando H<sub>0</sub> è vera → la variazione tra le medie dei gruppi è simile alla variazione all'interno dei gruppi
- Quando H<sub>0</sub> è falsa → la variazione tra le medie dei gruppi è più grande della variazione all'interno dei gruppi.



Il test statistico di base usato per ANOVA è il test F, quindi è necessario calcolare un F e un F<sub>critico</sub>:

- un valore alto di F rispetto al valore critico di tabella può far scartare H<sub>0</sub> a favore dell'ipotesi alternativa H<sub>a</sub> (H<sub>0</sub> falsa)
- un valore basso di F rispetto al valore critico di tabella può confermare H<sub>0</sub> (H<sub>0</sub> vera)

Però, F non può essere calcolato con la formula del semplice test F: si richiedono altre quantità chiamate

$$F = \frac{MQF}{MQE}$$

somma dei quadrati infatti: dove MQF è la media dei quadrati dei livelli del fattore (stima delle varianze dei gruppi) e MQE è la media dei quadrati dell'errore (stima della varianza tra i gruppi e all'interno dei gruppi)

### ma come calcolare MQF e MQE?

0. Media complessiva di tutti i dati:

$$\bar{\bar{x}} = \left(\frac{N_1}{N}\right)\bar{x}_1 + \left(\frac{N_2}{N}\right)\bar{x}_2 + \left(\frac{N_3}{N}\right)\bar{x}_3 + \dots + \left(\frac{N_I}{N}\right)\bar{x}_I$$

(N<sub>1</sub> è il gruppo 1 e N è il numero totale delle misure)

1. Somma dei quadrati dovuti al fattore:

$$SQF = N_1(\bar{x}_1 - \bar{\bar{x}})^2 + N_2(\bar{x}_2 - \bar{\bar{x}})^2 + N_3(\bar{x}_3 - \bar{\bar{x}})^2 + \dots + N_I(\bar{x}_I - \bar{\bar{x}})^2$$

➤ media del singolo analista paragonata con la media complessiva

2. Somma dei quadrati dovuta agli errori:

$$SQE = \sum_{j=1}^{N_1} (x_{1j} - \bar{x}_1)^2 + \sum_{j=1}^{N_2} (x_{2j} - \bar{x}_2)^2 + \sum_{j=1}^{N_3} (x_{3j} - \bar{x}_3)^2 + \dots + \sum_{j=1}^{N_I} (x_{ij} - \bar{x}_I)^2$$

$$SQE = (N_1 - 1)s_1^2 + (N_2 - 1)s_2^2 + (N_3 - 1)s_3^2 + \dots + (N_I - 1)s_I^2$$

➤ somma dei diversi valori singoli paragonati con la media del singolo analista.

Variazione tra i gruppi e all'interno dei gruppi

3. Somma totale dei quadrati SQT = SQF + SQE

4. Numero di gradi di libertà: (I-1) per SQF e (N-1) per SQE  
(dove "I" sono i gruppi e "N" il numero totale delle misure)

5. Valori medi dei quadrati

Media dei quadrati dei livelli del fattore = MQF =  $\frac{SQF}{I - 1}$  → Stima della varianza tra i gruppi ( $\sigma^2_E + \sigma^2_F$ )

Media dei quadrati dell'errore = MQE =  $\frac{SQE}{N - I}$  → Stima della varianza all'interno dei gruppi ( $\sigma^2_E$ )

## Tabella ANOVA

Sorgente di variazione	Somma dei quadrati (SQ)	Gradi di libertà (gl)	Medie dei quadrati (MQ)	Stime delle medie	F
Tra gruppi (effetto del fattore)	SQF	I - 1	$MQF = \frac{SQF}{I - 1}$	$\sigma_E^2 + \sigma_F^2$	$\frac{MQF}{MQE}$
All'interno dei gruppi (errore)	SQE	N - I	$MQE = \frac{SQE}{N - I}$	$\sigma_E^2$	
Totale	SQT	N - 1			

### TEST Q

Il test Q serve per decidere se un risultato anomalo/isolato, un outlier, debba essere accettato o scartato. L'outlier Potrebbe essere il risultato di un errore grossolano non rivelato.

$$Q = \frac{|x_q - x_n|}{w}$$

Q si calcola con un rapporto

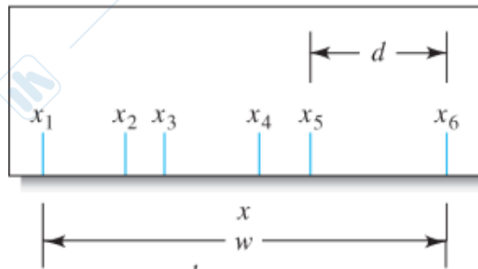
dove: -  $Xq$  è il valore anomalo

-  $Xn$  è il valore più vicino a  $Xq$  se i dati vengono disposti in ordine crescente

-  $w$  è la dispersione

$d$

Se  $Q > Q_{critico}$  allora rifiutiamo il dato: è un outlier dato dalla presenza di un errore grossolano.



$$d = x_6 - x_5$$

$$w = x_6 - x_1$$

$$Q = d/w$$

Se  $Q > Q_{critic}$  rigettare  $x_6$

# CAMPIONAMENTO, STANDARDIZZAZIONE, CALIBRAZIONE e VALUTAZIONE DEI METODI ANALITICI

Campionamento, standardizzazione e calibrazione sono gli step della procedura analitica che seguono la scelta del metodo.

## LA SCELTA DEL METODO: come si sceglie il metodo più adatto?

Tenendo conto di

- domanda che ci poniamo: se vogliamo definire la concentrazione dell'analita, allora utilizziamo il determinato metodo per trovare la concentrazione.
- livello di accuratezza richiesto, facendo un compromesso con il tempo e il denaro disponibili per l'analisi: se abbiamo bisogno di tanta accuratezza, il metodo scelto deve essere molto avanzato e deve utilizzare strumentazione sensibile.
- complessità di campioni e numero di componenti del singolo campione
- numero di campioni da analizzare per ottenere un risultato valido e statisticamente significativo
- quantità del campione: se parliamo di
  - ✓ > 0,1g → macroanalisi
  - ✓ Da 0,01 a 0,1g → semimicroanalisi
  - ✓ Da 0,0001 a 0,01g → microanalisi
  - ✓ < 10<sup>-4</sup>g → ultramicroanalisi
- quantità di analita: le quantità dell'analita determinano anche i diversi tipi di analiti. Se parliamo di
  - ✓ livello da 1% a 100% → principale
  - ✓ da 0,01 (100 ppm) a 1% → minore
  - ✓ da 1 ppb a 100 ppm → tracce
  - ✓ < 1 ppb → ultratracce

Maggiore è il livello e quindi la quantità dell'analita, maggiore è l'affidabilità dei risultati e minore l'errore: minore è la quantità, più la deviazione standard aumenta.

- effetti matrice nell'analisi di campioni reali: l'analisi di campioni reali è complicata dalla presenza della matrice del campione, ossia tutto ciò che non è analita. La matrice può contenere specie con proprietà chimiche simili analita e tali specie possono reagire con gli stessi reagenti allo stesso modo dell'analita oppure generare una risposta strumentale che può non essere facilmente distinta da quella proveniente dall'analita. Queste sostanze vengono chiamate INTERFERENTI perché interferiscono con l'analisi che si sta effettuando e quindi bisogna eliminarli o mascherarli. Inoltre, più complessa è la matrice, maggiore è la possibilità di avere interferenze.

*Ricorda: i campioni sono analizzati mentre i costituenti o le concentrazioni sono determinati (analisi del siero del sangue per il glucosio / determinazione del glucosio nel siero del sangue)*

## CAMPIONAMENTO

Il campionamento è il processo tramite il quale viene acquisita una frazione rappresentativa, chiamata CAMPIONE RAPPRESENTATIVO, della popolazione. I singoli prelievi per ottenere la frazione rappresentativa si chiamano incrementi di campionamento.

Infatti, l'analisi chimica non avviene mai su una popolazione, che è un insieme infinito di campioni, bensì su un campione, ossia un sottoinsieme finito/una piccola frazione dell'insieme infinito popolazione.

Affinché i risultati siano significativi, il campione deve essere rappresentativo della popolazione: deve avere le stesse caratteristiche chimico-fisiche della popolazione.

Le conclusioni derivanti dall'analisi del campione dovranno poi essere estese alla popolazione tramite l'analisi statistica: dall'analisi del campione possiamo estrapolare le conclusioni relative alla popolazione come la media e la deviazione standard tramite metodi statistici.

## COME OTTENERE UN CAMPIONE RAPPRESENTATIVO?

Inizialmente viene determinato un CAMPIONE GROSSOLANO, ossia un campione contenente tante piccole porzioni prelevate da diverse parti della massa del materiale. Quindi, per arrivare al campione grossolano

vengono realizzati tanti incrementi di campionamento rappresentativi della massa del materiale.

In seguito, il campione grossolano deve essere trasformato in CAMPIONE DA LABORATORIO per essere analizzato: il campione grossolano viene ridotto di dimensioni e reso omogeneo.

È da notare che entrambe le composizioni del campione grossolano e del campione da laboratorio devono riflettere quanto più possibile la composizione media della massa totale del materiale da analizzare.

## OBIETTIVI DEL CAMPIONAMENTO

Statisticamente, gli obiettivi del processo di campionamento sono:

1. Ottenere un **valore medio** della concentrazione dell'analita che sia una stima corretta della media della popolazione: deve essere valida l'ipotesi nulla.
2. Ottenere una **varianza** che sia una stima corretta della varianza della popolazione, in modo tale che si possano trovare intervalli di fiducia validi per la media e possano essere applicati i vari test statistici.

## INCERTEZZA DEL CAMPIONAMENTO

Anche il campionamento è soggetto a errore, che può essere grossolano, casuale e sistematico.

- Gli errori grossolani e sistematici possono essere controllati con attenzione ed eliminati tramite calibrazione mediante standard, bianchi e materiali di riferimento.
- Gli errori casuali possono essere mantenuti a livelli accettabili attraverso uno stretto controllo delle variabili che influenzano le misure,
- Gli errori dovuti al CATTIVO CAMPIONAMENTO non sono controllabili e quindi vengono trattati separatamente

$$\rightarrow s_{tot}^2 = s_c^2 + s_m^2$$

$s_c$  = deviazione standard campionamento

$s_m$  = deviazione standard metodo

Se l'incertezza di campionamento è elevata e non può essere migliorata, allora si utilizza un metodo di analisi meno preciso ma rapido per ottenere meno errori di campionamento: si aumenta il numero di campioni da analizzare. Questo perché la deviazione standard della media diminuisce

all'aumentare delle misurazioni  $s_{media} = \frac{s}{\sqrt{N}}$

## FASI DEL CAMPIONAMENTO

### IL CAMPIONE GROSSOLANO

Il campione grossolano è una replica in miniatura dell'intera massa del materiale che deve essere analizzato.

Per scegliere un campione grossolano rappresentativo è importante determinare il minor peso di materiale, quindi il minor numero N di particelle necessario a fornire l'informazione desiderata...ma come? stando attenti a due variabili che potrebbero influenzare la procedura di campionamento nello scegliere la quantità giusta di campione:

- 1) INCERTEZZA che può essere tollerata tra la composizione del campione e del campione grossolano
- 2) ETEROGENEITÀ del materiale, ossia se il materiale contiene particelle di dimensioni uguali (omogeneo) o particelle di dimensioni/granulometrie differenti (eterogeneo)

Solitamente, il numero di particelle richieste per un campione grossolano varia da poche a  $10^{12}$  particelle.

Distinguiamo, quindi due tipi di campionamenti del campione grossolano:

- a) CAMPIONAMENTO DI MATERIALE OMOGENEO = un esempio di campionamento di materiale omogeneo è quello di soluzioni omogenee di liquidi e gas. Il campione grossolano, in questo caso, può essere relativamente piccolo poiché:
  - sono omogenei fino al livello molecolare
  - piccoli volumi di campione possono contenere moltissime particelle.

Inizialmente il liquido o gas viene agitato. I liquidi vengono raccolti da sistemi in flusso continuo, mentre i gas possono essere campionati mediante diversi metodi come per esempio utilizzando un sacchetto di campionamento o intrappolandoli in un liquido/assorbendoli sulla superficie di un solido.

- b) **CAMPIONAMENTO DI MATERIALE ETEROGENEO** = un esempio di campionamento di materiale eterogeneo è quello di metalli e leghe. Il campione grossolano, in questo caso, deve essere di dimensioni più grandi in quanto sono completamente eterogenei. Il campionamento avviene mediante macchine segatrici, frantumatrici o trapanatrici che trapanano in vari punti in modo casuale per raccogliere più particelle di dimensioni differenti. Può anche avvenire utilizzando il metodo della fusione metallica e per rendere omogeneo il metallo/la lega.

### PREPARAZIONE DEL CAMPIONE DA LABORATORIO

Una volta creato il campione grossolano è necessario creare un campione da laboratorio analizzabile e misurabile.

Per fare questo bisogna capire se il campione grossolano:

- ✓ abbia le dimensioni adatte per il laboratorio, ossia al massimo poche centinaia di grammi.
- ✓ sia omogeneo

Se non rispetta questi requisiti si attua un processo di frazionamento per ridurre la grandezza delle particelle del campione grossolano e omogenizzarlo, **MANTENENDO INVARIATO** il numero delle particelle contenute nel campione

### NUMERO DI CAMPIONI DA LABORATORIO

Una volta preparati i campioni da laboratorio, l'interrogatorio che rimane è: quanto campione prelevare per l'analisi?

Il numero di campioni  $N$  da prelevare dipende da:

- a) intervallo di fiducia  $IF$  richiesto per il valore medio
- b) deviazione standard relativa del campionamento  $\sigma_c$

dato che  $IF = x \pm \frac{t \cdot sc}{\sqrt{N}}$  incertezza assoluta tollerata per un particolare livello di fiducia

L'incertezza relativa e quindi deviazione standard relativa  $\sigma_c$  sarà quindi  $\sigma_c = \frac{t \cdot sc}{x \cdot \sqrt{N}}$

E quindi il numero di campioni è  $N = \frac{t^2 \cdot sc^2}{x^2 \cdot \sigma_c^2}$

#### esempio

**La determinazione del rame in un campione di acqua di mare ha dato come valore medio 77,81  $\mu$ g/L e una deviazione standard  $sc$  di 1,74  $\mu$ g/L. Quanti campioni devono essere analizzati per ottenere nei risultati una deviazione standard relativa dell'1,7% ad un livello di fiducia del 95%?**

Soluzione

Iniziamo assumendo un numero infinito di campioni, che danno un valore di  $t$  pari a 1,96 ad un livello di fiducia del 95%. Poiché  $\sigma_r = 0,017$ ,  $sc = 1,74$  e  $x = 77,81$ ,

$$N = \frac{(1,96)^2 \times (1,74)^2}{(0,017)^2 \times (77,81)^2} = 6,65$$

$\times 7$  campioni,  $t$  per 6 gradi di libertà è 2,45. Usando questo valore di  $t$ , si può calcolare un secondo valore per  $N$  e si ottiene  $N = 10,38$ .

Se ora usiamo 9 gradi di libertà e  $t = 2,26$ , il valore successivo sarà  $N = 8,84$ .

Il processo di iterazione converge verso un valore di  $N$  approssimativamente pari a 9.

- ➔ Quello che si fa in questi esercizi è far convergere  $N$  con il valore dei gradi di libertà. Calcoliamo prima  $N$  con numero infinito di campioni e viene 6,65: prendo 7 come numero di campioni  $N$ , il numero più vicino a 6,65. Calcolo il numero di gradi di libertà e poi guardo sulle tabelle a quanto corrisponde  $t$ . Sostituisco  $t$  nell'espressione di  $N$  e controllo se il numero  $N$  e i gradi di libertà

convergono, sono valori vicini. Se così non è allora vado avanti prendendo un numero di campioni più grande e rifaccio il procedimento fino ad arrivare a  $N = N-1$ .

## STANDARDIZZAZIONE E CALIBRAZIONE

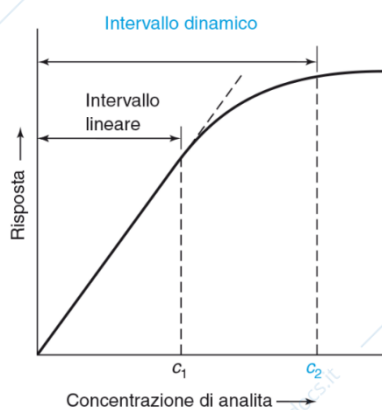
Dopo il campionamento si procede all'analisi chimica ma, prima di procedere, bisogna valutare che lo strumento stia funzionando correttamente mediante la calibrazione. La calibrazione determina, quindi, la relazione tra la concentrazione dell'analita e la risposta dello strumento. Per valutare questa relazione si utilizza una soluzione standard, di cui si conosce la concentrazione.

### CONFRONTO CON STANDARD

- ✓ **CONFRONTO DIRETTO:** Serve per verificare se le proprietà dell'analita da testare siano uguali o molto simili a quelle di uno standard. Si attua mediante un confronto tra le proprietà dell'analita e le proprietà della soluzione standard.  
Es. Si confronta il colore dell'analita con quello prodotto dalla stessa reazione degli standard. Se variando la concentrazione dello standard (per esempio per diluizione) si ottiene lo stesso colore che si ottiene dall'analita, allora si è stabilita la concentrazione dell'analita.
- ✓ **TITOLAZIONE:** l'analita, il titolato, reagisce con il titolante, un reagente standardizzato di cui si conosce la concentrazione, in una reazione a stechiometria nota. Si varia la quantità di titolante fino a che non si raggiunge il punto dell'equivalenza chimica ( $n^{\circ}\text{mol titolante} = n^{\circ}\text{mol titolato}$ ) e quindi la quantità di reagente standardizzato può essere messa in relazione alla quantità di analita presente e trovare la concentrazione dell'analita.

### CALIBRAZIONE CON STANDARD ESTERNO

Serve per definire una funzione di calibrazione dello strumento o curva di calibrazione, che mette in relazione la risposta dello strumento con le concentrazioni note degli standard dell'analita. Si attua preparando una serie di soluzioni standard dell'analita separatamente dal campione e misurandole con lo strumento voluto. Si riportano in un grafico le misurazioni sull'asse delle x e le risposte dello strumento sull'asse delle y. Otterremo una curva di calibrazione quadratica SPERIMENTALE relativa alla concentrazione dell'analita. Questa si può confrontare con una curva di calibrazione perfettamente lineare TEORICA dello standard dell'analita.

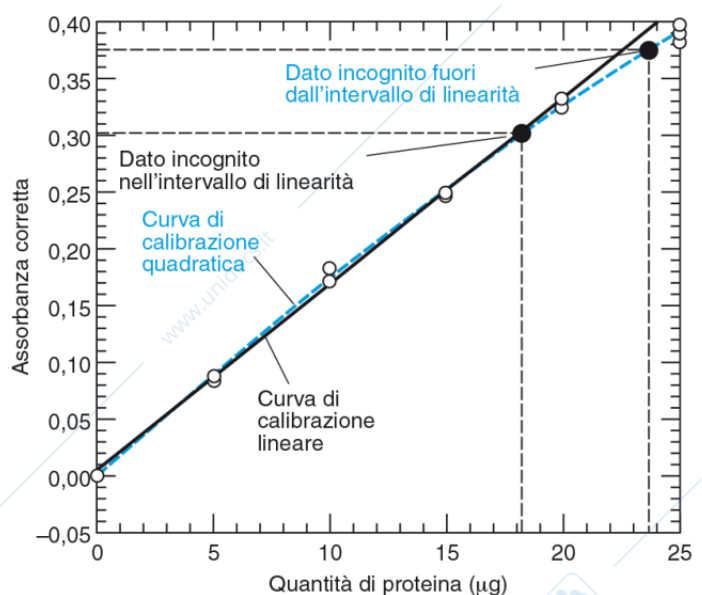


La curva di calibrazione è caratterizzata da:

- **INTERVALLO LINEARE** = intervallo di concentrazione di analita nel quale si ha una relazione lineare tra la risposta misurata dallo strumento e la concentrazione dell'analita
- **INTERVALLO DINAMICO** = intervallo di concentrazione di analita nel quale si ha una relazione NON lineare ma misurabile tra la risposta misurata dallo strumento e la concentrazione dell'analita

La curva di calibrazione teorica è una retta con equazione  $y=mx+b$  dove  $m$  è

data da  $\Delta y/\Delta x$  e  $b$  è l'intercetta.



La deviazione di ciascun punto sperimentale  $(x_i; y_i)$  dalla retta di calibrazione teorica è detto residuo:  $R = y_i - (mx_i + b)$ . Infatti, se non tutti i dati si trovano esattamente sulla curva di calibrazione lineare allora sono stati fatti degli errori determinati nel processo di misura: si deve cercare di disegnare la miglior retta tra i punti dei dati. Per fare ciò si utilizza il metodo dei MINIMI QUADRATI, che consente di tracciare la migliore retta che interpola una serie di dati sperimentali, aventi deviazioni casuali da un comportamento perfettamente lineare.

Per svolgere il metodo dei minimi quadrati si assume che

- 1) Esiste effettivamente una relazione lineare tra risposta misurata e concentrazione dello standard
- 2) Ogni deviazione dei singoli punti della retta deriva da un errore commesso nel corso della misurazione: non è associato errore ai valori di  $x$ .

Il metodo dei minimi quadrati calcola la **somma dei quadrati dei residui**  $SS_{resid}$  di tutti i punti della curva sperimentale e la minimizza: serve per rendere più piccola la deviazione in modo da costruire una retta che si avvicina molto a quella teorica.

$SS_{resid}$  è una misura della variazione dei valori di  $y$  che **non sono spiegati dalla relazione**, presunta lineare, tra  $x$  e  $y$ .

$$SS_{resid} = \sum_{i=1}^N [y_i - (b + mx_i)]^2$$

Si può definire anche una **somma totale dei quadrati**  $SS_{tot}$ , ossia una misura della variazione totale tra i valori di  $y$ .

$$SS_{tot} = S_{yy} = \sum (y_i - \bar{y})^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{N}$$

Per capire se la curva sperimentale coincide con la curva teoria allora ci serviamo del COEFFICIENTE DI DETERMINAZIONE  $R^2$ .

$$R^2 = 1 - \frac{SS_{resid}}{SS_{tot}}$$

Se  $R^2$  è vicina all'unità (-1 o 1), la curva di calibrazione teorica e sperimentale coincidono e quindi il 100% della curva sperimentale può essere spiegata con il modello lineare della teorica.

Si dice anche che le due variabili  $x$  e  $y$  hanno una **CORRELAZIONE LINEARE** che può essere positiva ( $R^2=1$ : all'aumentare di  $x$ , aumenta anche  $y$ ) o negativa ( $R^2=-1$ : all'aumentare di  $x$ , diminuisce  $y$ ).

Se  $R^2$  è 0 allora la curva di calibrazione sperimentale non coincide per niente con quella teorica.

In questo caso si dice che le due variabili  $x$  e  $y$  **NON** sono correlate linearmente.

L'equazione della curva di calibrazione che interpola i dati e che descrive come la  $y$  cambi in funzione della  $x$  viene ottenuta grazie a un'ANALISI di REGRESSIONE.

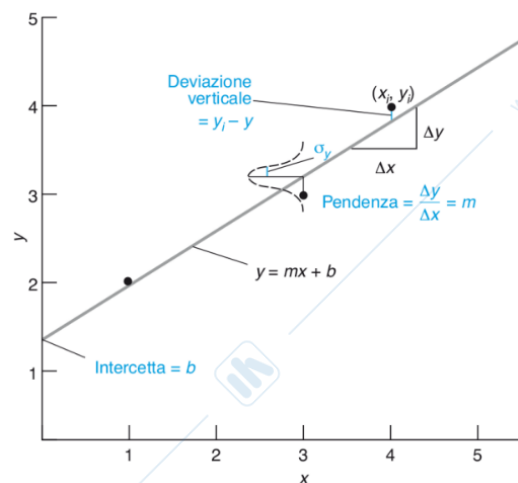
*La curva di calibrazione costituisce anche un metodo indispensabile nell'ambito della chimica analitica per la determinazione della concentrazione o della massa di un campione incognito per interpolazione, sostituendo la  $y$  e trovando la  $x$ .*

### ERRORI DI CALIBRAZIONE con standard esterno

Anche durante il processo di calibrazione con standard esterno possono comparire alcuni errori, che possono essere casuali o sistematici o legati al cattivo campionamento/calibrazione (strumentali).

Gli errori sistematici sono dovuti a uno standard non preparato correttamente; quindi, per eliminarli basta preparare accuratamente gli standard, che devono avere stato chimico uguale a quello dell'analita del campione.

Gli errori legati al cattivo campionamento/calibrazione vengono corretti mediante misurazione di un BIANCO. Il bianco ideale è identico al campione, ma senza analita. Molto spesso, però, è più conveniente

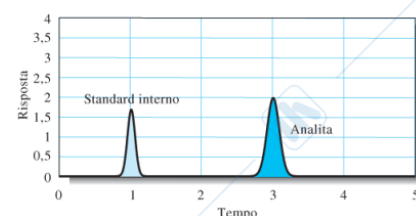
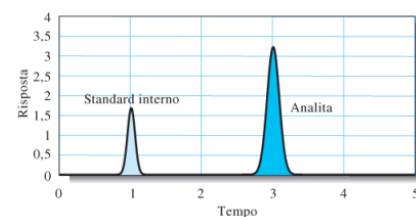
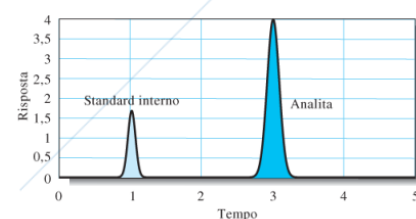


ricorrere ad un bianco reale che può essere un bianco del solvente, contenente lo stesso solvente nel quale è disciolto il campione, o un bianco del reagente che contiene il solvente più tutti i reagenti usati nella preparazione del campione.

### MINIMIZZAZIONE DEGLI ERRORI nelle procedure analitiche: preparazione del campione

L'accuratezza totale e la precisione di un'analisi potrebbero essere limitate da errori nella misurazione, nel campionamento e nella calibrazione: come minimizzarli?

- **SEPARAZIONI:** il campione è purificato mediante metodi di separazione (filtrazione, precipitazione, dialisi, estrazione con solvente, volatilizzazione, scambio ionico e cromatografia) per minimizzare gli errori dovuti ad eventuali interferenze nella matrice del campione.
- **SATURAZIONE:** al campione vengono aggiunte specie interferenti, gli standard e i bianchi, in modo che l'effetto delle interferenze diventi indipendente dalla concentrazione originale della specie interferente nel campione. Questo, però, può ridurre la sensibilità e la rivelabilità dell'analita
- **MODIFICATORE DI MATRICE:** al campione viene aggiunto un modificatore di matrice, ossia una specie, di per sé non interferente, che viene aggiunta ai campioni, agli standard e ai bianchi in quantità tale da rendere la risposta indipendente dalla concentrazione della specie interferente.
- **AGENTE MASCHRANTE:** al campione viene aggiunto un agente mascherante, ossia una sostanza che reagisce selettivamente con la specie interferente per formare un complesso che non interferisca.
- **DILUIZIONE:** viene diluito il campione per minimizzare l'effetto interferente di una specie se questa non produce un effetto significativo al di sotto di un determinato livello di concentrazione.
- **CORRISPONDENZA DELLA MATRICE:** viene duplicata la matrice del campione aggiungendo i principali componenti della matrice alle soluzioni degli standard e del bianco.
- **AGGIUNTE STANDARD:** viene aggiunta una quantità nota di una soluzione standard dell'analita al campione e l'altro campione viene lasciato inalterato. Le risposte dei due campioni vengono usate per ricavare la concentrazione dell'analita e evitano quel tipo di errori legati al campionamento (interferenze e contaminazioni).
- **STANDARD INTERNO:** viene aggiunto al campione, agli standard e ai bianchi uno standard dell'analita a concentrazione nota che rientra nell'intervallo lineare della curva di calibrazione e chimicamente-fisicamente simile all'analita. Si costruisce quindi un grafico che ha sull'asse delle y le risposte dell'analita e dello standard interno e sulle x la concentrazione dell'analita. Si ottengono così due curve di calibrazione che esprimono il rapporto tra risposta analita/concentrazione analita e risposta standard/concentrazione analita. Queste curve chiamate "picchi" sottendono un'area costituita dall'insieme dei punti ottenuti sperimentalmente che varia in funzione della concentrazione dell'analita. Il rapporto tra la risposta dell'analita e la risposta dello standard interno ci permetterà di ricavare la concentrazione dell'analita. Il metodo dello standard interno può compensare quegli errori che influenzano in egual modo l'analita e lo standard interno.



### VALUTAZIONE DEI METODI ANALITICI

I metodi analitici strumentali sono basati su un SEGNALE STRUMENTALE, ossia una grandezza fisica misurabile specifica dell'analita e proporzionale alla concentrazione dell'analita. È la risposta dello strumento alla concentrazione dell'analita.

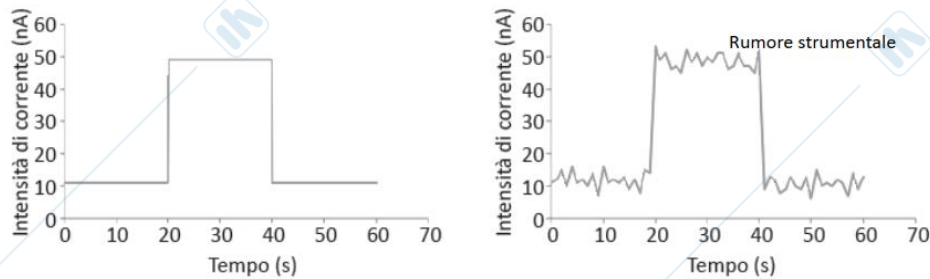
Il RUMORE è la componente del segnale priva di informazione analitica, legata alle fluttuazioni casuali del segnale. Il rumore va, quindi, a interferire sul segnale strumentale.

→ L'ampiezza delle fluttuazioni è data dalla deviazione standard strumentale RSD%. Questa è legata al segnale medio e alla deviazione standard del segnale.

$$RSD\% = \frac{\sigma_s}{S_m} \cdot 100$$

Il FONDO STRUMENTALE è il segnale analitico misurato in assenza di analita.

I rapporti segnale/rumore e segnale/fondo sono un indice di quanto riusciamo a distinguere il segnale analitico dal suo fondo o dal rumore, ossia quanto possiamo riconoscerlo. Se il rapporto è vicino a 1 allora segnale e fondo/rumore sono simili e non riusciamo a distinguerli.



## LE FIGURE DI MERITO

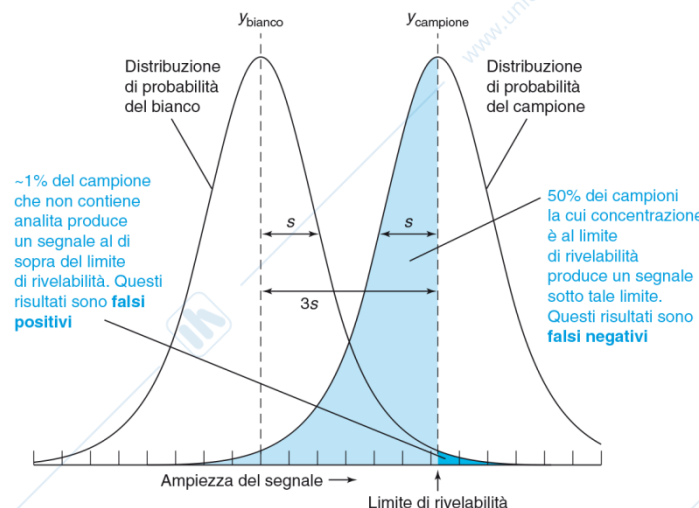
I metodi analitici vengono caratterizzati e valutati da diverse figure di metodo che sono

- 1) **LIMITE DI RIVELABILITÀ (LOD o DL):** è il limite al di sotto del quale non distinguiamo più l'analita dal suo fondo strumentale (segnale/fondo=1). La soglia al di sopra della quale il segnale è significativamente diverso dal fondo si dice **SEGNALE MINIMO RILEVABILE** ed è  $S_{min} = B + 3\sigma_B$  ( $B$  è il valore medio del fondo e  $\sigma_B$  è la deviazione standard del fondo). Il limite di rilevabilità è la quantità di analita che corrisponde al segnale minimo rilevabile.

$$DL = \frac{k\sigma_B}{m}$$

dove  $k$  è chiamato fattore di fiducia e  $m$  è la sensibilità della calibrazione. Di solito si sceglie  $k=2$  o  $k=3$ :  $k$  pari a 2 corrisponde ad un livello di fiducia del 92,1% e  $k$  pari a 3 ad un livello di fiducia del 98,3%

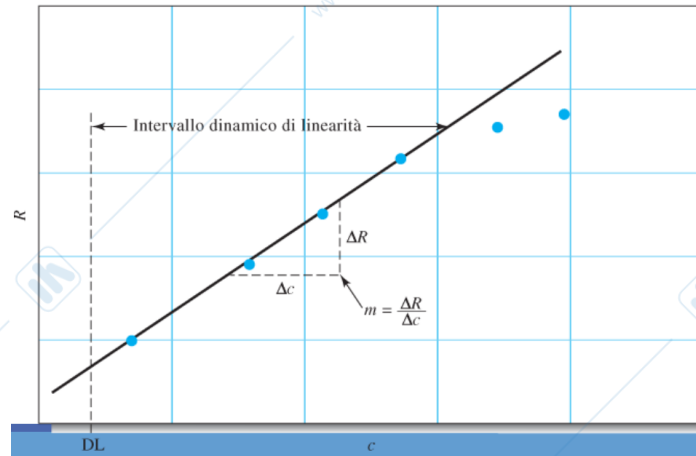
- 2) **LIMITE DI QUANTIFICAZIONE (LOQ):** è il limite al di sotto del quale non possiamo più effettuare un'analisi quantitativa perché la quantità di analita è troppo bassa. In questo caso il **SEGNALE MINIMO RILEVABILE** è  $S_{min} = B + 10\sigma_B$ . Il suo valore è sempre più alto di quello del limite di rivelabilità.



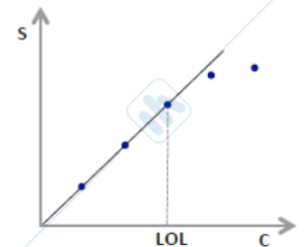
- 3) **SENSIBILITÀ ( $m$ ):** capacità dello strumento di rispondere in funzione di piccole variazioni di concentrazione di analita.

→ Sensibilità della calibrazione = variazione del segnale di risposta per unità di variazione della concentrazione dell'analita. Corrisponde alla pendenza della curva di calibrazione ( $m = \Delta R / \Delta c$ ): aumentando la pendenza della curva di calibrazione, aumenta la differenza di segnale a parità di differenza di concentrazione; quindi, aumentando la pendenza aumenta anche la sensibilità.

- Sensibilità analitica = rapporto tra la pendenza della curva di calibrazione e la deviazione standard del segnale analitico a una data concentrazione dell'analita



- 4) **SELETTIVITÀ**: capacità di un metodo analitico di riconoscere l'analita dal resto della matrice e quindi di non risentire della presenza di componenti diversi dall'analita (interferenti). Dipenderà dalla concentrazione dell'analita e dalla concentrazione dell'interferente:  $S = b + m_1 C_{\text{analita}} + m_2 C_{\text{interferente}}$ . Il coefficiente di selettività  $K = m_1/m_2$ .
- 5) **INTERVALLO DI LINEARITÀ (LOL)**: è la concentrazione massima entro la quale la relazione tra segnale analitico e concentrazione dell'analita ha andamento **LINERARE**. Il limite inferiore dell'intervallo è l'intervallo di rivelabilità. Il limite superiore dell'intervallo è il limite di linearità del 5%, ossia la concentrazione alla quale il segnale analitico o la pendenza della curva di calibrazione devia dalla linearità del 5% e quindi non ha più un andamento lineare. L'intervallo è, quindi, compreso tra il limite di rivelabilità e il limite di linearità.
- Se la concentrazione non rientra in questo intervallo allora si procede per diluizione. Ma il segnale che noi vediamo è dovuto alla soluzione diluita e la concentrazione reale non è quella diluita: dobbiamo tenerne conto. La prima soluzione avrà come concentrazione il fattore moltiplicativo della diluizione che abbiamo operato.
- 6) **PRECISIONE**= riproducibilità delle misure e quindi del risultato. Può essere strumentale, intra-analisi, inter-laboratorio, intermedia tra diversi operatori.
- 7) **INCERTEZZA DI MISURA**
- 8) **ACCURATEZZA** = vicinanza del dato sperimentale al dato reale. L'accuratezza si valuta analizzando un materiale di riferimento standard, utilizzando un metodo analitico differente, aggiungendo al bianco una quantità nota di analita, aggiungendo uno standard dell'analita all'incognito.
- 9) **RECUPERO** = indica la quantità di analita determinata da un metodo di analisi rispetto alla quantità totale.
- 10) **ROBUSTEZZA** = indica la capacità di un metodo analitico di non risentire degli effetti delle variazioni delle condizioni analitiche.
- 11) Velocità, facilità e convenienza, capacità dell'operatore, costi e disponibilità della strumentazione, costo per campione



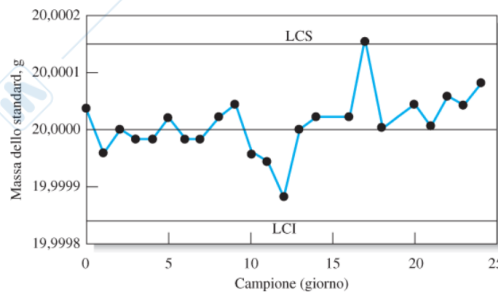
## GARANZIE DI QUALITÀ DEI RISULTATI ANALITICI

Quando si applicano metodi analitici bisogna valutare la qualità dei risultati e la qualità delle prestazioni dei mezzi e degli strumenti utilizzati. Questo è reso possibile dal controllo di qualità, dato dalle CARTE DI CONTROLLO, e dalla validazione dei risultati.

Una carta di controllo è un diagramma sequenziale di qualche caratteristica importante nella garanzia di qualità. La carta mostra i limiti statistici di variazione consentiti per il parametro che viene misurato:

$$\begin{aligned} - \text{ LCS} &= \text{limite di controllo superiore} \\ - \text{ LCI} &= \text{limite di controllo inferiore} \end{aligned} \quad \left. \begin{aligned} \text{LCS} &= \mu + \frac{3\sigma}{\sqrt{N}} \\ \text{LCI} &= \mu - \frac{3\sigma}{\sqrt{N}} \end{aligned} \right\}$$

dove  $\mu$  è la media della popolazione delle misure di massa,  $\sigma$  è la deviazione standard della popolazione delle misure e  $N$  è il numero delle repliche che devono essere effettuate per ogni campione.



La validazione è la conferma attraverso esami e apporto di prove oggettive che i requisiti particolari per uno specifico utilizzo previsto siano soddisfatti. In altre parole, è un parametro che determina l'adeguatezza di un'analisi a fornire l'informazione desiderata e può essere applicata ai campioni, alle metodologie e ai dati.

Ci sono diversi modi per validare i metodi analitici:

- ✓ analisi di materiali standard di riferimento, quando disponibili
- ✓ analisi con un metodo diverso e confronti tra metodi e inter-laboratorio
- ✓ analisi di campioni "arricchiti" (con aggiunta esterna di analita alla matrice)
- ✓ valutazione dei fattori che influenzano la misura e il risultato finale
- ✓ stima dell'incertezza associata al risultato
- ✓ analisi di campioni sintetici che simulano la composizione chimica dei campioni da analizzare