

## modulo 3

da dare i 3 moduli nello stesso anno  
registra le lezioni

$$X = \bar{x} \pm \Delta x \quad \text{con livello di confidenza } P \quad \text{intervallo di confidenza: risultato medio } \pm \text{ incertezza}$$

esempio risultato:  $(2,34 \pm 0,04) \mu\text{g/L}$ , 95%

l'errore richiede sempre una sola cifra significativa  
gli errori si sommano sempre

**intervallo di confidenza**: intervallo numerico nel quale il dato atteso cade con una certa probabilità, tale probabilità si chiama **livello di confidenza**

**significatività** ( $\alpha$  / alfa): complemento a 100 del livello di confidenza ( $\alpha = 1 - P$ )

$X \pm \Delta x$ , P    alfa più comune:  $0,05 = 5\%$

**media:**

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

**deviazione standard (s):**  
con  $n-1$ =gradi libertà

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

**varianza:** deviazione standard quadratica     $s^2$

**scarto quadratico:** differenza quadratica tra il dato e la media

deviazione standard: media degli scarti quadratici fratto il grado di libertà

**gradi di libertà (v)** : differenza tra il numero di misure effettuate e il numero di parametri da determinare  
(es: con l'assorbanza ho 2 parametri: A e C)

nel caso ideale con infinite misurazioni, la frequenza diventa una distribuzione di probabilità definibile dal grafico di gauss

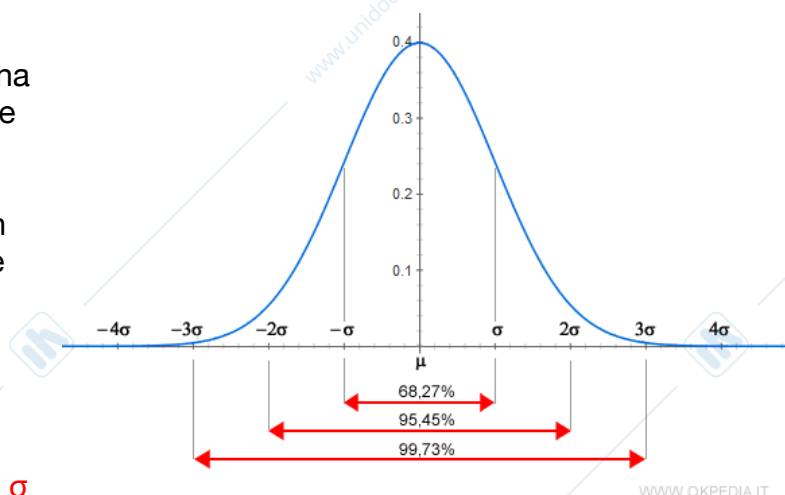
il massimo del grafico (indicato con  $\mu$ , mu) è la migliore stima del valore vero

la migliore stima della dev. std corrisponde a  $\sigma$  (sigma)

i punti di flesso sono definiti da  $\mu \pm \sigma$

intervallo di confidenza: integrale da  $\mu - \sigma$  a  $\mu + \sigma$     = 68%  
intervallo di confidenza: integrale da  $\mu - 2\sigma$  a  $\mu + 2\sigma$     = 95%  
intervallo di confidenza: integrale da  $\mu - 3\sigma$  a  $\mu + 3\sigma$     = 99,7%

$\mu \pm t\sigma \Rightarrow t$  di Student (è un parametro di  $\alpha$ )



all'aumentare del parametro t aumenta il livello di confidenza

la validazione di un metodo di analisi passa attraverso alcuni parametri di qualità, il primo è la precisione

**esattezza:** è la vicinanza tra il dato sperimentale ed un dato di riferimento in relazione con gli errori sistematici

**precisione:** è la capacità di un metodo di fornire misure replicate molto ravvicinate (è di conseguenza legata alla deviazione standard => più  $\sigma$  è piccolo e più la mis. è precisa) in relazione con gli errori casuali

**accuratezza:** esattezza con precisione (definizione IUPAC)

per avere un risultato preciso vogliamo avere sigma più piccolo possibile, ad un livello di confidenza sufficientemente alto

gli statisti hanno definito come un evento che capita meno di una volta su venti (5%) è definibile errore statistico casuale e può essere trascurato

=> fissando alfa al 5%, t è sempre uguale a 2

nel caso reale non faccio infinite misurazioni, ma un numero limitato (campione statistico) mentre n è limitato contano i gradi di libertà, poiché all'aumentare del grado di libertà v, t diminuisce. Cambiare sigma significa aumentare il livello di confidenza ma aumentare anche l'errore possibile

**teorema del limite centrale:** la distribuzione delle medie ha lo stesso valore vero  $\mu$  ma la sua dev. std è pari al rapporto tra la dev. std delle misure singole e la radice quadrata di n la varianza della media è invece uguale alla varianza delle misure fratto n

**deviazione standard delle medie:**

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

**intervallo di confidenza per la media:**

$$\bar{x} \pm t_{\alpha, n-1} \frac{s}{\sqrt{n}} \quad (s = \text{dev. std})$$

se abbiamo scelto un liv. di conf. del 95% => 95% risultati avranno un valore medio all'interno dell'intervallo  $S/\text{radice}(n)$

**cifre significative:** si definisce significativa una cifra di ordine superiore a quella dell'errore, l'errore si rappresenta sempre con una sola cifra significativa

formato di un errore:  $n * 10^m$ , con n = mantissa (num. naturale tra 1 e 9), mentre  $10^m$  è l'ordine di grandezza dell'errore

es:  $100,38794 \pm 0,004593 \text{ g}$  =>  $100,388 \pm 0,005$

il numero di cifre significative corrisponde al numero di cifre della mantissa nel formato esponenziale (es: 1349000 ha 4 cifre sig, 0.00017368 ne ha 5)

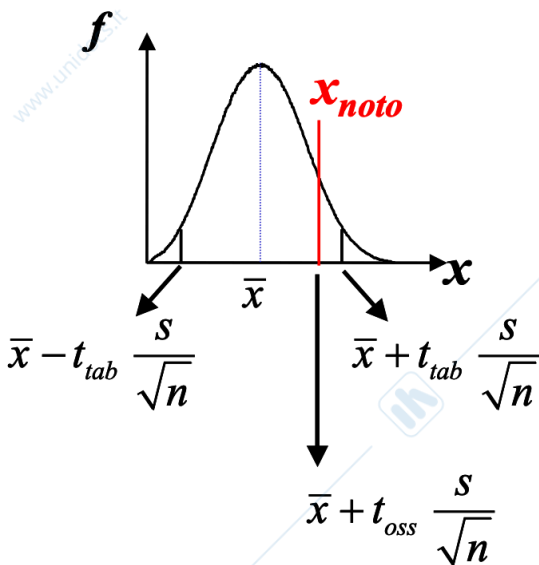
**test di significatività:** metodo statistico che ha l'obiettivo di confrontare se più risultati sono significativamente diversi (es: un risultato sperimentale con uno atteso o di riferimento, oppure uno sperimentale con un'altro sperimentale)

principali test di significatività: (ne studieremo 2)

- 1) t-test => controlla l'esattezza (vicinanza del dato centrale col dato atteso)
- 2) F-test => controlla la precisione
- 3) Q-test => controlla i dati anomali
- 4) chi quadro ( $\chi^2$ ) => controlla la normalità (ovvero l'appartenenza alla gaussiana)

il t-test è collegato agli errori sistematici, la precisione ha invece a che fare con gli errori casuali

**t-test per il confronto di una media con un valore noto:**



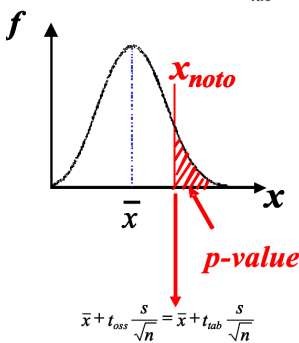
$$t_{oss} = \frac{|x_{noto} - \bar{x}|}{\frac{s}{\sqrt{n}}} \quad (3.2.1)$$

(l'ipotesi nulla è che non ci sia differenza significativa)

$t_{oss} < t_{tab} \Rightarrow$  l'ipotesi nulla è accettata, cioè non c'è differenza significativa tra  $\bar{x}$  e  $x_{noto}$  al livello di confidenza scelto.

$t_{oss} > t_{tab} \Rightarrow$  l'ipotesi nulla è rigettata, cioè c'è differenza significativa tra  $\bar{x}$  e  $x_{noto}$  al livello di confidenza scelto.

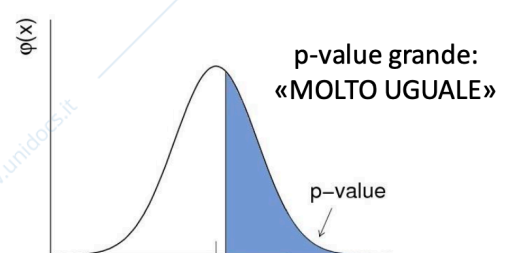
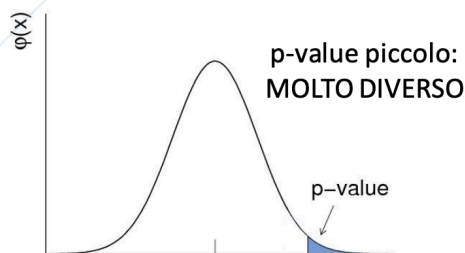
Il livello di confidenza  $P$  scelto, ovvero il livello di significatività  $\alpha$  scelto, e il valore di  $n$  determinano il valore numerico di  $t_{tab}$ .



**P-value:** rappresenta il grado di somiglianza tra un dato di riferimento e il punto massimo della gaussiana

è un valore di significatività che corrisponde alla situazione limite:

$$t_{oss} = t_{tab}$$



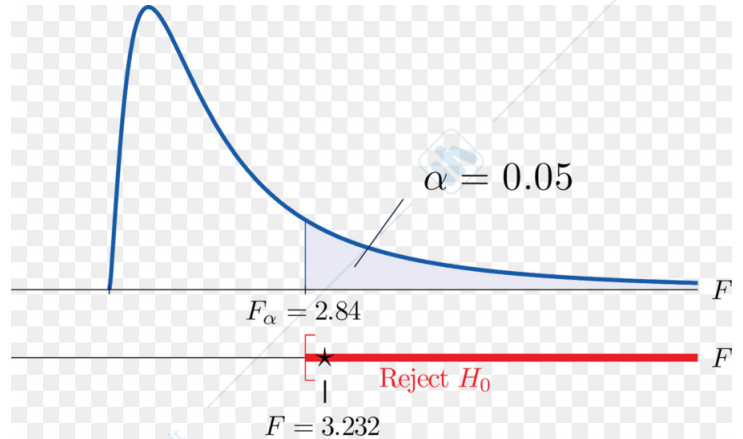
Nel caso di un problema a due code il p-value varia tra 0 e 1

Nel caso di un problema a una coda il p-value varia tra 0 e 0.5

mentre alfa è un riferimento soggettivo, il P-value è un dato oggettivo

### F-test per il confronto tra deviazioni standard

- è utilizzato per confrontare metodi diversi usati sullo stesso campione, al fine di determinare quale è più preciso
- è un test di accuratezza: mette a confronto due medie, due deviazioni standard e due gradi di libertà
- consideriamo il test ad una coda (solo quella di destra)



$$F_{oss} = \frac{s_1^2}{s_2^2}$$

l'ipotesi nulla è che non ci sia differenza significativa

F(oss) si calcola sempre con il rapporto tra la varianza maggiore al numeratore e la varianza minore al denominatore.

$$F_{oss} < F_{\alpha, v_1, v_2}^{1coda} \Rightarrow H_0 \text{ accettata, il metodo 2 non è più preciso del metodo 1.}$$

la funzione su excel è INV.F(alpha, v(max), v(min))

i gradi di libertà totali corrispondono alla somma dei gradi di libertà iniziali

**ANOVA (ANalysis Of VAriance)**: particolare tipo di F-test usato per confrontare molte serie di misure (ognuna definita da media, varianza e gradi di libertà)

si confrontano le **varianze intra-campioni** (ovvero dello stesso laboratorio) e le **varianze inter-campioni** (ovvero le varianze di diversi laboratori), se non presentano differenza significativa allora tutti i dati sono considerabili appartenenti alla stessa gaussiana

**varianza intra-campioni => delle singole misure**

**varianza inter-campioni => delle medie**

ci si aspetta che la varianza inter sia più grande della varianza intra, in quanto generalmente la variabilità aumenta cambiando campione, piuttosto che ripetendo la misura sullo stesso

con **n** serie di misure, e ciascuna con prove ripetute **m** volte (ed **N** numero totale di misure)

$$s_p^2 = \frac{\sum_{i=1}^n s_i^2}{n}$$

la **varianza intra-campioni** è la varianza media  $S(p)^2$ :

il **grado di libertà** associato alla varianza media è

$$v_p = \sum_{i=1}^n v_i = \sum_{i=1}^n (m-1) = nm - n = N - n$$

varianza inter-campioni:  $s_{\bar{x}}^2 = \frac{\sum_{i=1}^n (\bar{x}_i - \bar{\bar{x}})^2}{n-1}$ , dove la **grande media**  $\bar{\bar{x}}$  è data da:  $\bar{\bar{x}} = \frac{\sum_{i=1}^n \bar{x}_i}{n}$

la varianza inter-campioni è anche uguale al rapporto tra la varianza vera e il numero di misure ripetute

$$s_{\bar{x}}^2 = \frac{\sigma^2}{m}$$

Le  $n$  medie dello schema 3.7.1 sono non significativamente diverse se  $m s_{\bar{x}}^2$  non è significativamente più grande di  $s_p^2$  e cioè se il seguente  $F_{osservato}$ :

$$F_{osservato} = \frac{m s_{\bar{x}}^2}{s_p^2} = \frac{m \cdot (\text{varianza delle medie "inter"})}{(\text{media delle varianze "intra"})}$$

non è maggiore del seguente  $F_{calcolato}$  mediante la corrispondente funzione di Excel:

$$F_{calcolato} = \text{INV.F}(\alpha; n-1; N-n)$$

dove  $\alpha$  è la significatività scelta.

### (EXCEL)

- conviene sempre formattare la riga degli errori in notazione scientifica
- si calcola sempre l'intervallo di confidenza della media
- ogni volta che si usa uno stesso modello cambiando i valori iniziali bisogna controllare di nuovo il formato dell'errore e la misura
- con \$ blocca la casella (es: \$A\$3 blocca A3, se metto \$A3 blocca la riga A ma scorre)
- funzioni MEDIA(), SOMMA(), RADQ(), DEV.ST.C() (riferita al campione, da usare sempre questa, mai usare DEV.ST.P() )
- P-value a 2 code: DISTRIB.T.2T(t(oss); gradi lib.)
- t-tab: INV.T.2T(alfa, gradi lib)
- il t-oss si calcola con la formula intera su excel [=> |x(oss)-media| / (s/n^0.5) ]
- funzione F-test: INV.F(α;v(max);v(min))
- per F-test si usa DISTRIB.T.1T() per il P-value ad una coda
- funzione per varianza maggiore: SE(condizione; caso vero; caso falso)
- per fare la grande media basta selezionare tutti i dati su excel con la funzione media normale

### CALIBRAZIONE

**esempio con assorbanza:** la costante è determinabile statisticamente

$A = k \cdot c \Rightarrow$  relazione teorica

$A = a + b \cdot c \Rightarrow$  relazione nella realtà (con aggiunta polinomiale)

l'intercetta  $a$  vale 0 nel modello ideale, in chimica analitica è però sbagliato presupporre che  $a$  sia nulla, bisogna dimostrare che sia sufficientemente piccola da considerarla tale

nella realtà i colorimetri (o gli spettrofotometri) raggiungono un **asintoto orizzontale**, poiché il sensore luminoso alla fine del cammino ottico non ottiene mai un valore nullo, al massimo 1/1000 della luce in ingresso, per cui l'assorbanza massima è uguale a 3 (dato che  $A = \log(\text{luce incidente}/\text{luce passante})$ )

$\Rightarrow$  in laboratorio si lavora sempre sotto all'unità di assorbanza (<1.000)

**LOD (Limits Of Detection):** il limite di rilevabilità è il minimo segnale percepibile dallo strumento, è significativamente diverso dal bianco

$\Rightarrow A(\text{lod}) > A(\text{bianco})$  necessariamente (con  $A$ =assorbanza)

$$A = a + b \cdot c \quad (1)$$

$$A(\text{lod}) = A(b) + 3s(b) \text{ in cui } s(b) \text{ è la migliore stima di } s(y/x) \quad (2)$$

$$A(\text{lod}) \approx a + 3s(y/x)$$

$$\text{Con la formula (1): } A(\text{lod}) = a + b \cdot c(\text{lod}) \quad (3)$$

$$\text{ponendo (2) = (3): } \underline{c(\text{lod}) = 3s(y/x) / b}$$

## METODI QUANTITATIVI NELL'ANALISI STRUMENTALE

esistono 3 principali metodi per quantificare indirettamente un analita:

- 1) curva di taratura
- 2) aggiunte standard
- 3) standard interno

### 1) CURVA DI TARATURA

- si preparano vari campioni standard con diverse quantità note di analita ( $x(i)$ ) e si sottopone ognuno di essi alla misura sperimentale, registrando il rispettivo segnale analitico  $S(i)$ .
- si riporta un grafico ( $x, S$ ) e si determina un metodo statistico per determinare la curva che si adatta meglio ai punti sperimentali, questo procedimento prende il nome di **fitting**.
- si sottopone il campione incognito alla misura sperimentale e si rileva il segnale analitico  $S(0)$
- si esegue un'interpolazione sulla curva di taratura per determinare  $x(0)$

**interpolazione:** procedimento matematico finalizzato al calcolo di nuovi punti dato un insieme finito di dati, attraverso l'estrapolazione di una funzione che si adatta in maniera opportuna

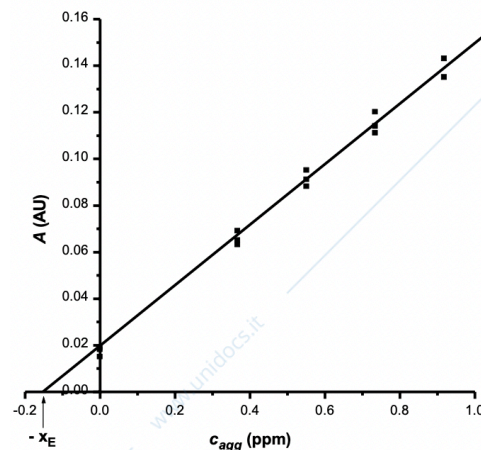
**intervallo lineare:** intervallo di valori i quali seguono una dipendenza lineare

esso è molto importante per via della semplice correlazione, dove si riesce ad impiegarlo la curva viene definita **retta di taratura**

**sensibilità:** è definita sensibilità la pendenza della curva di taratura. Nel caso di una retta di taratura la sensibilità coincide con il coefficiente angolare (in inglese è **sensitivity** non **sensibility**)

### 2) AGGIUNTE STANDARD

- si preparano  $n$  aliquote uguali dello stesso campione incognito
- si portano tutte allo stesso volume ottenendo  $n$  standard caratterizzati da una quantità aggiunta (con il primo senza aggiunta)
- si sottopone ciascuno standard alla misura strumentale e si rileva il segnale analitico  $S(i)$
- si estrapola la funzione che meglio si adatta alla curva, in questo caso sempre una retta
- si esegue una estrapolazione della retta in modo da determinare l'ascissa all'origine, così facendo si ottiene il valore incognito  $x(E)$



$$\text{Retta: } y = a + b x$$

$$x_E = \frac{a}{b} \quad (5.2.1)$$

### 3) STANDARD INTERNO

**caso 1:** uso dei **fattori di risposta**

- si individua una specie chimica simile all'analita per proprietà chimico-fisiche (standard interno)
- si prepara un campione contenente una quantità nota di analita e standard interno ( $x(a)$  e  $x(z)$ )
- si sottopone il campione alla misura e si rileva il rispettivo segnale analitico  $S(a)$  e  $S(z)$

$$\Rightarrow S(a) : S(z) = k(a)x(a) : k(z)x(z)$$

- si ricavano le  $k$  eseguendo misure ripetute

- partendo dal campione incognito si prepara un secondo campione in cui l'analita è presente in quantità incognita  $x$  e lo standard interno in quantità nota  $z$  (stesso ordine di grandezza di  $x$ )

- si sottopone il secondo campione alla misura strumentale  $\Rightarrow S(a)$  e  $S(z)$

$$\Rightarrow S(x) / S(z) = (k x) / z \quad \Rightarrow x = S(x)/S(z) * z/k$$

- si effettuano misure ripetute e si calcola  $x$

il caso 1 si usa quando il campione incognito contiene naturalmente uno standard interno con ordine di grandezza diverso da quello dell'analita (es: ricerca metanolo nel vino)

**caso 2:** uso di una **retta di regressione**

- si preparano vari campioni contenenti quantità note  $x(i)$  di analita e la stessa quantità  $z$  di standard interno. Questi campioni vengono chiamati standard

- si sottopone ciascuno standard alla misura strumentale. Per lo standard i-esimo si rileverà un segnale S(i) per l'analita ed un segnale S(z) per lo standard interno
- si riportano in grafico i punti sperimentali (x(i), S(i)/S(z))
- si applica un metodo statistico per determinare la retta che meglio si adatta (fitting)
- si sottopone il campione incognito alla misura strumentale, rilevando un segnale S(0)
- si esegue una interpolazione sulla retta di regressione per calcolare il valore x(0)/z, che corrisponde a S(0)/S(z), da cui si ricava x(0)

per applicare questo metodo è necessario che lo standard interno sia simile all'analita, che l'intervallo in cui si lavora sia lineare e che il recupero analitico relativo di analita e standard interno sia 100%

come vantaggi presenta il fatto che si può applicare dove una retta di taratura sarebbe poco precisa, e permette la determinazione quantitativa anche quando non si dispone di uno standard puro di analita

### REGRESSIONE LINEARE

un problema statistico di regressione si presenta quando si hanno n coppie di valori (x, y) corrispondenti a dati sperimentali, e si vuole determinare la curva y=f(x) che meglio si adatta ai dati sperimentali. Si parla di "regressione di y su x"

- condizione necessaria per poter eseguire una regressione è che l'errore sulla quantità sia trascurabile rispetto all'errore sul segnale => l'errore su y è indipendente da x (il quale è considerato senza errore)
- il procedimento matematico consiste nel minimizzare la sommatoria degli scarti quadratici medi delle ordinate
- una volta determinata la funzione analitica della curva di taratura, è necessario interpolare il valore della quantità nota

considerando la grande media degli x(i) e degli y(i), si definisce il centroide dei punti sperimentali. Il centroide non è in genere un punto sperimentale, ma ha la caratteristica di essere sempre presente sulla retta di regressione

**centroide:**  $(\bar{x}, \bar{y})$

errore su a e b:

$$\text{errore su } a = t_{\alpha, n-2} S_a$$

$$\text{errore su } b = t_{\alpha, n-2} S_b$$

equazioni per la regressione di una retta:

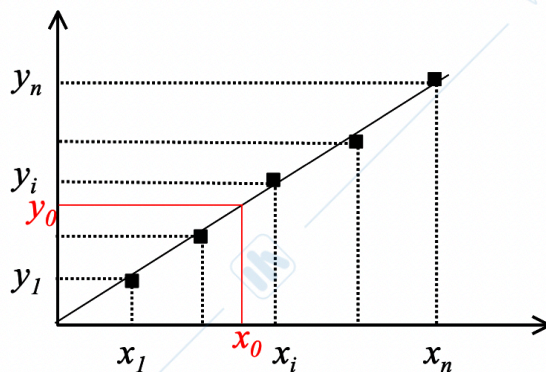
$$b = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}$$

$$a = \bar{y} - b \bar{x}$$

$$s_{y/x} = \sqrt{\frac{\sum_{i=1}^n (y_i - a - b x_i)^2}{n-2}}$$

$$s_a = s_{y/x} \sqrt{\frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}}$$

$$s_b = \frac{s_{y/x}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}$$



se non sono presenti misure replicate ma una regressione, la significatività non va indicata con la radice di n a denominatore

**interpolazione di una retta di taratura:**

$$x_0 = \frac{y_0 - a}{b}$$

con m = num. ripetizioni di y(0):

$$s_{x_0} = \frac{s_{y/x}}{b} \left[ \frac{1}{m} + \frac{1}{n} + \frac{(y_0 - \bar{y})^2}{b^2 \sum_{i=1}^n (x_i - \bar{x})^2} \right]^{1/2}$$

## PROPAGAZIONE DEGLI ERRORI

a) **errori assoluti statistici:**

- 1) misure dirette replicate
- 2) interpolazione

b) **errori tecnici:** dovuti dalla strumentazione o dall'uso di essa

es: pesata

$$m(\text{netta}) = m(\text{lorda}) - m(\text{tara})$$

$$dm(\text{netta}) = dm(\text{lorda}) - dm(\text{tara})$$

$$\Delta m(\text{netta}) = |\Delta m(\text{lorda})| - |\Delta m(\text{tara})|$$

$$\Delta m(\text{netta}) = \Delta m(\text{lorda}) + \Delta m(\text{tara}) \Rightarrow \text{gli errori si sommano sempre}$$

se gli errori sono solo casuali:  $\Delta m(\text{netta}) = (\Delta m(\text{lorda})^2 + \Delta m(\text{tara})^2)^{1/2}$

es: preparazione di una soluzione

$$c = m / (MM * V)$$

$$\ln(c) = \ln(m) - \ln(MM) - \ln(V)$$

$$dc / c = dm / m - dMM / MM - dV / V$$

$$\Delta c / c = \Delta m / m - \Delta MM / MM - \Delta V / V$$

$\Delta c / c \rightarrow$  errore relativo

se  $z = x_1 + x_2 + x_3 + \dots + x_n \Rightarrow \Delta z =$  sommatoria dei  $\Delta x$

si considera il p-value (nel t-test), per avere valori precisi  $a \rightarrow 1$ , mentre  $b \rightarrow 0$  (???)

## METODO DELLE AGGIUNTE STANDARD

(lezione prof)

Questo metodo si applica quando la misura di un campione viene significativamente influenzata dalla matrice in cui esso è immerso.

Questo effetto viene detto **effetto matrice**.

**Non si può dare per scontato che la matrice non influenzi il segnale analitico**

Ad esempio il calcolo dell'assorbanza per determinare la presenza di  $Pb^{2+}$  varia notevolmente se si considerano 2 campioni con stessa quantità di analita ma immersi in matrici diverse (come acqua e vino)

Il metodo delle aggiunte standard permette di definire in quali casi la matrice influenzi significativamente il segnale analitico

**svantaggi del metodo delle aggiunte standard:**

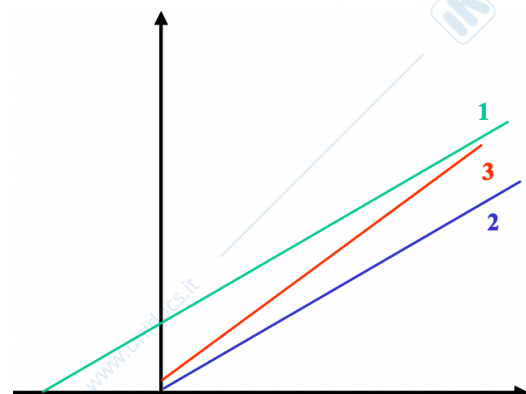
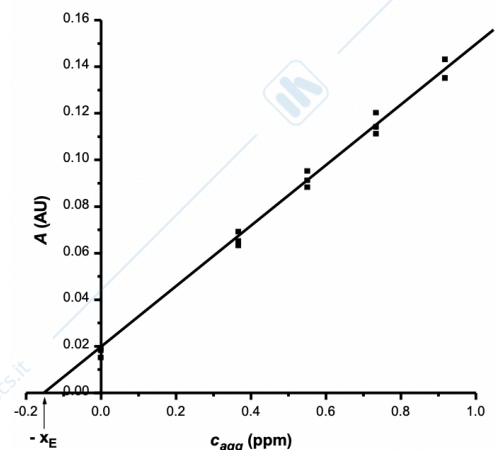
- 1) è poco preciso, legge lontano dal centroide (con standard interno è più vicino)
- 2) ogni campione incognito richiede l'impiego un processo lungo (preparazione di diverse soluzioni più estrapolazione con il grafico)

Nel grafico a lato sono mostrate le seguenti curve:

- 1) retta delle aggiunte standard
- 2) retta ottenuta col metodo della retta di taratura in caso di assenza di effetto matrice
- 3) retta ottenuta col metodo della retta di taratura da standard ideali in caso di presenza di effetto matrice

Il limite di rilevabilità (LOD) per il metodo delle aggiunte standard è lo stesso del metodo dello standard interno

Ciò accade poiché la pendenza della retta delle aggiunte standard è la stessa di una retta traslata sull'origine (dalla retta verde a quella blu)



## DOMANDE FREQUENTI ALL'ESAME

- 1) definizione di intervallo di confidenza
- 2) definizione di gradi di libertà di un problema statistico
- 3) scrivere la formula per l'intervallo di confidenza della media nel caso di n misure ripetute
- 4) definizione di t-test
- 5) definizione di F-test
- 6) definizione di esattezza. Quale test statistico serve a verificarla?
- 7) definizione di precisione. Quale test statistico serve a verificarla?
- 8) definizione di accuratezza. Quale test statistico serve a verificarla?
- 9) definizione di p-value
- 10) descrivere il procedimento della retta di taratura basata sulla legge di Lambert Beer e precisare come devono essere i p-value di intercetta e pendenza nel confronto con il valore nullo
- 11) discutere come si progettano gli standard per costruire una retta di taratura (viene dettata la formula per la deviazione standard sul dato interpolato)
- 12) definizione di sensibilità
- 13) scrivere la formula per l'intervallo di confidenza per un dato interpolato su una retta di taratura
- 14) descrivere il procedimento della retta delle aggiunte standard. Svantaggi del metodo delle aggiunte standard rispetto al metodo dello standard esterno
- 15) definizione di limite di rivelabilità e calcolo della minima concentrazione rilevabile da una retta di taratura o delle aggiunte standard
- 16) definizione dell'effetto matrice. Verifica statistica di tale effetto
- 17) definizione di test ANOVA. Un esempio di come lo abbiamo applicato a lezione:
  - 1) ANOVA per il confronto tra più di 2 intervalli di confidenza
  - 2) ANOVA per la verifica del modello "retta di regressione"
- 18) calcolo dell'errore tecnico nel preparare una soluzione a concentrazione assegnata
- 19) calcolo dell'errore tecnico in una titolazione
- 20) come si combinano errori statistici ed errori tecnici