

1. INTRODUZIONE

1.1. Problemi analitici quantitativi

I metodi chimico-analitici strumentali hanno lo scopo di quantificare o di determinare proprietà chimico-fisiche di uno o più *analiti* contenuti in una *matrice*.

Tali metodi consistono nel mettere in relazione un *segnale analitico* con la quantità (massa o concentrazione) o con la proprietà cercata per l'analita di interesse.

Anche le determinazioni “qualitative” (che rispondono alla domanda: l'analita c'è o non c'è?) sono in realtà quantitative, perché in ogni caso la risposta analitica va corredata di una informazione numerica che quantifica l'affidabilità della risposta stessa.

1.2. Errori nell'analisi quantitativa

Nessun risultato analitico ha senso se non corredata di:

- *errore*
- *livello di confidenza* = probabilità di dare una risposta vera
- *livello di significatività* = probabilità di dare una risposta falsa.

livello di confidenza + *livello di significatività* = 100%

1.3. Tipi di errore

Gli errori nell'analisi quantitativa si possono classificare in:

1.3.1. *errori grossolani.*

- Sono dovuti a sviste macroscopiche.
- Determinano la presenza di *outliers*

1.3.2. *errori casuali.*

- Fanno sì che le singole misure siano casualmente in eccesso o in difetto rispetto al *valore vero*.
- Sono dovuti a fluttuazioni incontrollabili delle condizioni sperimentali.
- Determinano la *precisione* della misura ed incidono sulla *ripetibilità e riproducibilità*.

1.3.3. *errori sistematici*

- Fanno sì che le singole misure siano tutte in eccesso o tutte in difetto rispetto al *valore vero*.
- Sono dovuti a non calibrazione degli strumenti o a pregiudizi dell'operatore (es: errore di parallasse)
- Determinano l'*esattezza* di una misura.

2. STATISTICA DELLE MISURE RIPETUTE

2.1 Media e deviazione standard

Supponiamo di avere effettuato n misure *ripetute* della grandezza X . Indichiamo i risultati di tali misure con x_i con $i=1, \dots, n$.

- Si definisce *media* delle misure ripetute la quantità:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (2.1.1)$$

- Si definisce *numero di gradi di libertà* la differenza tra il numero di misure ripetute e il numero di parametri da determinare.
- Si definisce *deviazione standard* delle misure ripetute la quantità:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}, \text{ dove } n-1 = n^\circ \text{ gradi di libertà} \quad (2.1.2)$$

Si dimostra che:

$$s = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n-1} - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n(n-1)}} \quad (2.1.3)$$

- Si definisce *varianza* delle misure ripetute la quantità: s^2 .
- Si definisce *coefficiente di variazione (CV)* o *deviazione standard relativa (RSD)* la quantità, espressa in %:

$$CV = RSD = \frac{s}{\bar{x}} 100 \quad (2.1.4)$$

2.2 Distribuzione dei risultati di una misura

Supponiamo di avere effettuato n misure *ripetute* della grandezza X .

- I risultati di tali misure (x_i , $i=1,\dots,n$) sono un *campione* appartenente ad una *popolazione*=l'insieme di tutti i possibili risultati che verrebbero da infinite misure di X .

- Supponiamo che il risultato j -esimo x_j si presenti m_j volte. Si definisce *frequenza* del risultato j -esimo la quantità m_j . Si definisce *frazione* di misure che hanno dato risultato x_j la quantità:

$$f_j = \frac{m_j}{n} \quad (2.2.1)$$

- Si dimostra che:

$$\sum_{j=1}^k f_j = 1 \quad (2.2.2)$$

- Si definisce *media pesata* delle misure ripetute la quantità:

$$\bar{x} = \frac{\sum_{j=1}^k m_j x_j}{n} \quad (2.2.3)$$

Ovviamente:

$$\frac{\sum_{i=1}^n x_i}{n} = \frac{\sum_{j=1}^k m_j x_j}{n} = \bar{x} \quad (2.2.4)$$

2.2.1. Distribuzioni

Nella misura di una grandezza X , si definisce *distribuzione* dei risultati la curva che descrive la *frequenza* dei possibili risultati in funzione del valore dei risultati stessi. La *distribuzione* dei risultati di una misura è una proprietà della *popolazione*.

Il modello matematico che descrive la *distribuzione* dei risultati di una misura è la *distribuzione Gaussiana* o *distribuzione normale*:

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2.2.5)$$

2.2.2. Proprietà della distribuzione gaussiana

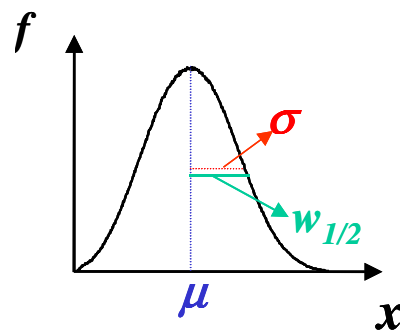


Fig. 2.2.1

- | | | |
|-----------------------------------|---|---------------------------------------|
| (1) Punto di massimo: | $x = \mu$ | $\mu = \text{valore vero}$ |
| (2) Massimo: | $f(x) = \frac{1}{\sigma \sqrt{2\pi}}$ | |
| (3) Punti di flesso: | $x = \mu \pm \sigma$ | $\sigma = \text{deviazione standard}$ |
| (4) Semi-larghezza a metà altezza | $w_{1/2} = \sigma \sqrt{2 \ln 2}$ | |
| (5) Larghezza a metà altezza | $W_{1/2} = \sigma \sqrt{8 \ln 2} \Rightarrow W_{1/2}^2 = 5.54 \sigma^2$ | |

Probabilità che una misura cada tra $x=a$ e $x=b$ $P = \int_a^b f(x) dx$

Dimostrazione delle proprietà (1)-(5)

$$(1) \quad f'(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \left[-\frac{1}{\sigma^2}(x-\mu) \right] \Rightarrow \text{Punto di massimo:}$$

$x = \mu$ Immediata, sostituendo μ al posto di x nell'equazione 2.2.5

$$(3) \quad f''(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \left[-\frac{1}{\sigma^2}(x-\mu) \right]^2 + \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \left[-\frac{1}{\sigma^2} \right]$$

$$f''(x) = \frac{1}{\sigma^3 \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \left[\frac{1}{\sigma^2}(x-\mu)^2 - 1 \right]$$

$$f''(x) = 0 \Rightarrow \frac{1}{\sigma^2}(x-\mu)^2 - 1 = 0 \Rightarrow (x-\mu)^2 = \sigma^2 \Rightarrow \text{punti di flesso: } x \pm \mu$$

(4) Metà altezza è pari alla metà del massimo (v. (2))

$$f(x)_{1/2} = \frac{1}{2} \frac{1}{\sigma \sqrt{2\pi}} \Rightarrow \frac{1}{2} = e^{-\frac{(x-\mu)^2}{2\sigma^2}} \Rightarrow -\ln 2 = -\frac{(x-\mu)^2}{2\sigma^2} \Rightarrow$$

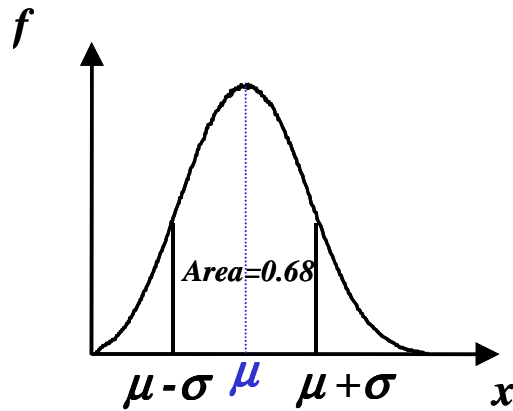
$$\Rightarrow 2\sigma^2 \ln 2 = (x-\mu)^2 \Rightarrow x = \mu \pm \sigma \sqrt{2 \ln 2} \Rightarrow$$

$$\Rightarrow \text{Semi-larghezza a metà altezza} = \sigma \sqrt{2 \ln 2}$$

(5) Immediata, ponendo $W_{1/2} = 2 w_{1/2}$

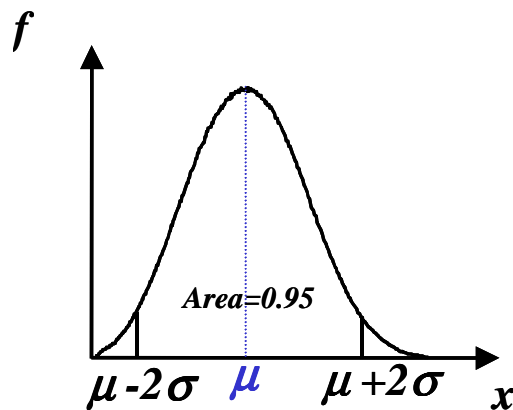
2.2.3. *Distribuzione gaussiana: normalizzazione e probabilità*

Condizione di normalizzazione $\int_{-\infty}^{+\infty} f(x) dx = 1$



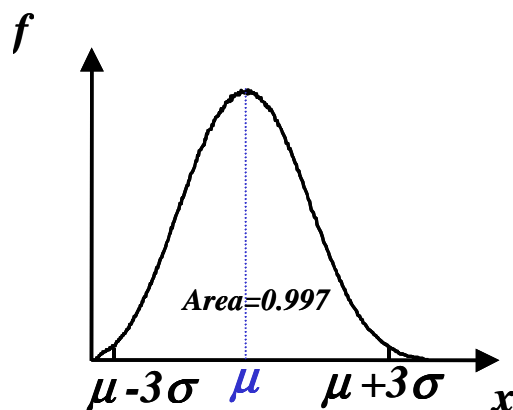
Probabilità che una misura cada tra $x = \mu - \sigma$ e $x = \mu + \sigma$:

68%



Probabilità che una misura cada tra $x = \mu - 2\sigma$ e $x = \mu + 2\sigma$:

95%



Probabilità che una misura cada tra $x = \mu - 3\sigma$ e $x = \mu + 3\sigma$:

99.7%

2.2.4. Intervallo di confidenza: caso ideale

Consideriamo il caso ideale in cui riusciamo ad eseguire infinite misure di X . Possiamo dire che:

il 95% dei risultati è compreso tra $x=\mu-2\sigma$ e $x=\mu+2\sigma$

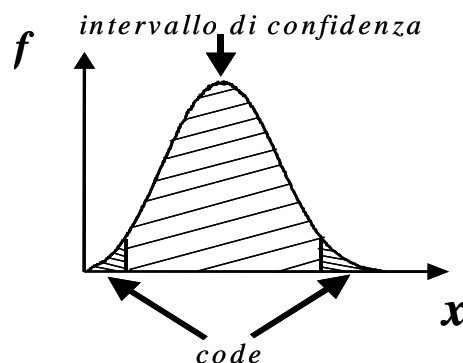
una ulteriore misura di X cadrà nell'intervallo $x=\mu\pm 2\sigma$
con una *probabilità* P del 95%.

all'*intervallo di confidenza* $x=\mu\pm 2\sigma$ associamo un
livello di confidenza del 95%.

Definizione: livello di significatività = 100% - livello di confidenza

livello di confidenza P = probabilità di cadere nell' intervallo di
confidenza

livello di significatività α = probabilità di cadere nelle *code* della
distribuzione



2.2.5. Intervallo di confidenza: caso reale

Consideriamo il caso reale in cui eseguiamo un numero finito n di misure di X , cioè abbiamo un *campione* appartenente ad una certa *popolazione*.

- La migliore stima di μ è la *media*:

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

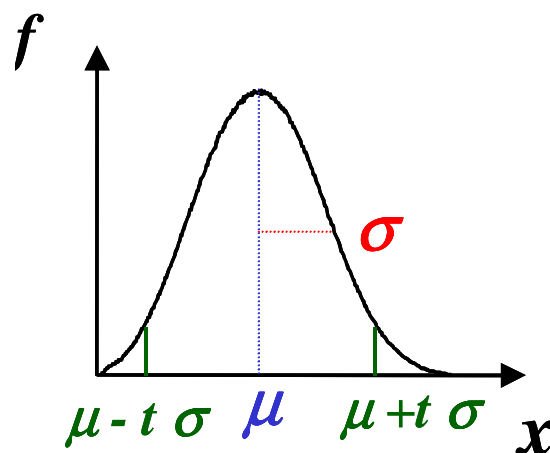
- La migliore stima di σ è la *deviazione standard* s :

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

- La migliore stima dell'*intervallo di confidenza* è:

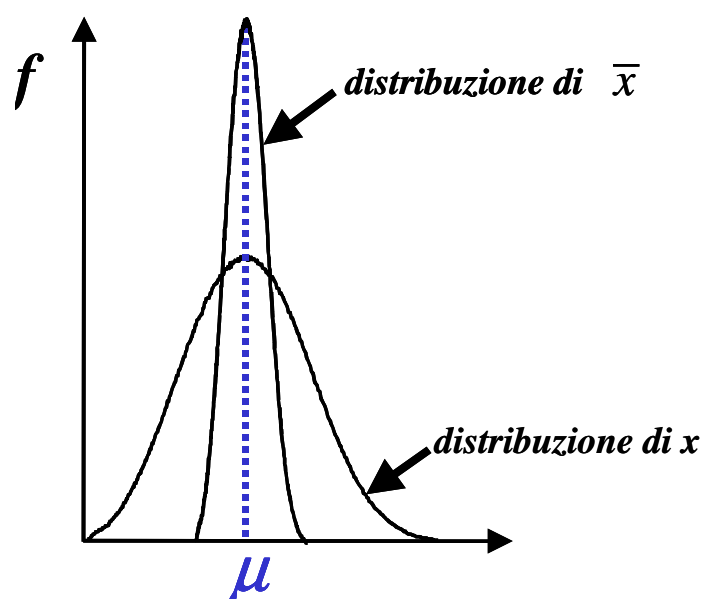
$$\bar{x} \pm t_{\alpha, n-1} s$$

dove t è un parametro statistico chiamato *t di Student*.



2.3 Distribuzione delle medie

Supponiamo di eseguire infiniti esperimenti in ciascuno dei quali misuriamo la *media* con n misure. Si dimostra che vale il *teorema del limite centrale*: la distribuzione delle medie ha lo stesso *valore vero* ma la sua *deviazione standard* è pari alla deviazione standard delle misure singole, divisa per la radice quadrata di n .



$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} \quad (2.2.6)$$

2.4 Intervallo di confidenza per la media

In base al teorema del limite centrale, l'intervallo di confidenza per una media è:

$$\bar{x} \pm t_{\alpha, n-1} \frac{s}{\sqrt{n}} \quad (2.2.7)$$

2.5 Presentazione dei risultati

- *Cifre significative* = tutte le cifre certe più la prima su cui cade l'errore. La prima cifra incerta è determinata dal numero che dà l'errore, espresso con una sola cifra significativa.

Es: $0.0137995 \pm 0.0003 = 0.0138 \pm 0.0003$

- Esprimendo il risultato numerico in forma esponenziale, le cifre significative sono quelle del fattore pre-esponenziale.

Esempi:

0.00017368 (errore su 8) = $1.7368 \cdot 10^{-4}$ ha 5 cifre significative

1349000 (errore sul 9) = $1.349 \cdot 10^6$ ha 4 cifre significative

3. TESTS DI SIGNIFICATIVITÀ

3.1. Definizione di test di significatività

- Un *test di significatività* è un metodo statistico che consente di stabilire se più risultati siano o meno significativamente diversi.

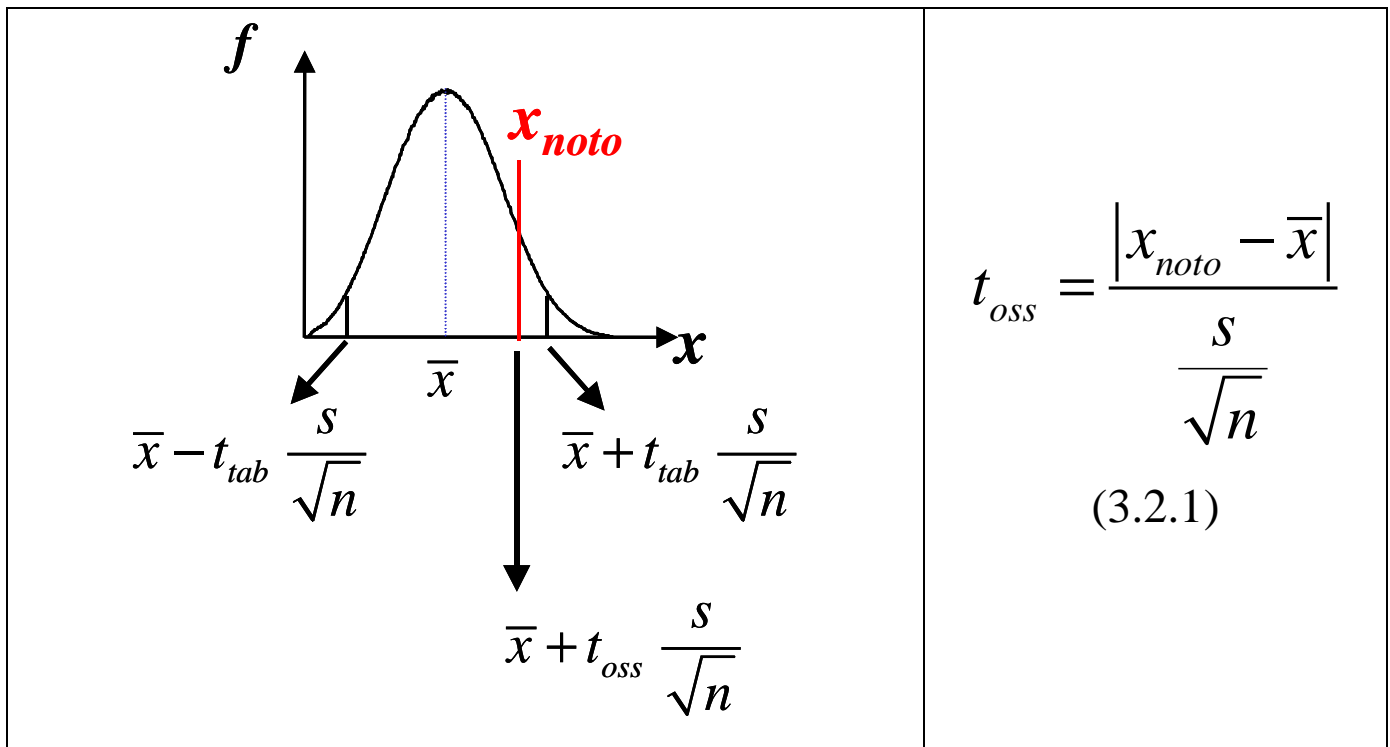
- Si parte sempre da una *ipotesi nulla* H_0 , che è l'ipotesi che *non* ci sia differenza significativa tra i risultati confrontati. Si stabilisce quindi se l'*ipotesi nulla* è vera o falsa al livello di confidenza scelto.

- Principali test di significatività:
 - ✓ *t-test* per la verifica dell'*esattezza*.
 - ✓ *F-test* per la verifica della *precisione*.
 - ✓ *Q-test* per la verifica di *dati anomali*.
 - ✓ *Test* χ^2 per la verifica della *normalità* di una distribuzione

- Definizione: *accuratezza* = *esattezza* & *precisione*

- Definizione: *validazione* = *verifica di accuratezza*

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



$t_{oss} < t_{tab} \Rightarrow$ 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 α scelto, e il valore di n determinano il valore numerico di t_{tab} .

Il *t*-test è un test di esattezza.

3.2.1. Esempio di confronto di una media con un valore noto

Si sottopone un campione a concentrazione nota ad un metodo analitico. Si vuole verificare se tale metodo dà il risultato atteso.

$$x_{\text{noto}} = 38.9 \text{ ppb}$$

Si eseguono 3 misure e si ottengono i risultati seguenti:

$$x_1 = 38.9 \text{ ppb}; x_2 = 37.4 \text{ ppb}; x_3 = 37.1 \text{ ppb}$$

$$\bar{x} = \frac{38.9 + 37.4 + 37.1}{3} = 37.8 \text{ ppb}$$

$$s = \sqrt{\frac{(38.9 - 37.8)^2 + (37.4 - 37.8)^2 + (37.1 - 37.8)^2}{3 - 1}} = 0.964 \text{ ppb}$$

$$t_{\text{tab}} = t_{0.05, 2} = 4.3$$

$$t_{\text{oss}} = \frac{|37.8 - 38.9|}{\frac{0.964}{\sqrt{3}}} = 1.98$$

Poiché $t_{\text{oss}} < t_{\text{tab}}$ l'ipotesi nulla è accettata, cioè non c'è differenza significativa tra il risultato ottenuto e il valore noto, al livello di confidenza del 95%.

Il *t-test* appena eseguito equivale a calcolare:

$$\bar{x} = 37.8 \pm 4.3 * \frac{0.964}{\sqrt{3}} = 37.8 \pm 2.4$$

$$35.4 < \bar{x} < 40.2 \text{ (intervallo di confidenza)}$$

Si osserva che x_{noto} cade dentro l'intervallo di confidenza.

3.3. *t*-test per il confronto tra due medie

Si vogliono confrontare due risultati ottenuti con due *tecniche diverse* sullo *stesso campione*:

risultato 1. \bar{x}_1, s_1, n_1, v_1

risultato 2. \bar{x}_2, s_2, n_2, v_2

Un simile problema si ha per esempio quando si vuole validare un metodo analitico mediante uno standard certificato. In questo caso il confronto tra le due medie è una *verifica di esattezza* del metodo sottoposto a validazione.

Verificare l'*ipotesi nulla* che non ci sia differenza significativa tra \bar{x}_1 e \bar{x}_2 equivale a verificare che non ci sia differenza significativa tra $|\bar{x}_1 - \bar{x}_2|$ e lo zero.

3.3.1. Caso 1: s_1 e s_2 non sono significativamente diverse

Si esegue un t -test in cui:

$$t_{oss} = \frac{|\bar{x}_1 - \bar{x}_2|}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (3.3.1)$$

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{\nu_1 s_1^2 + \nu_2 s_2^2}{\nu_1 + \nu_2} \quad (\text{varianza pooled}) \quad (3.3.2)$$

$$\nu = \nu_1 + \nu_2 = \text{numero di gradi di libertà del problema} \quad (3.3.3)$$

$t_{oss} < t_{P,\nu} \Rightarrow$ ipotesi nulla accettata

non c'è differenza significativa tra i due risultati

Esempio.

Due metodi hanno dato i seguenti risultati sullo stesso campione.

$\bar{x}_1 = 28.0$ ppm, $s_1 = 0.3$ ppm, $n_1 = 10$ misure

$\bar{x}_2 = 26.3$ ppm, $s_2 = 0.2$ ppm, $n_2 = 9$ misure

$$s = \sqrt{\frac{(10-1) \cdot 0.3^2 + (9-1) \cdot 0.2^2}{10+9-1}} = 0.251$$

$$t_{oss} = \frac{|28.0 - 26.3|}{0.251 \cdot \sqrt{\frac{1}{10} + \frac{1}{9}}} = 14.7, \nu = 10+9-1=18$$

$$t_{tab} = t_{0.05,18} = 2.1$$

$t_{oss} > t_{tab} \Rightarrow$ c'è differenza significativa tra i due risultati.

3.3.2. Caso 2: s_1 e s_2 sono significativamente diverse

Si esegue un t -test in cui:

$$t_{oss} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (3.3.4)$$

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 + 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 + 1}} \quad (3.3.5)$$

e si arrotonda ν all'intero più vicino.

3.4. Paired *t*-test

Il test delle differenze accoppiate si applica quando si abbiano h campioni diversi e su ciascuno si esegua una singola misura col metodo 1 e una singola misura col metodo 2.

Siano:

$x_{i,1}$ il risultato ottenuto sul campione i -esimo col metodo 1

$x_{i,2}$ il risultato ottenuto sul campione i -esimo col metodo 2

$$\begin{pmatrix} x_{1,1} & x_{1,2} \\ \dots & \dots \\ x_{i,1} & x_{i,2} \\ \dots & \dots \\ x_{h,1} & x_{h,2} \end{pmatrix} \rightarrow \begin{pmatrix} d_1 \\ \dots \\ d_i \\ \dots \\ d_n \end{pmatrix} \rightarrow \bar{d}, s_{\bar{d}}, h-1 \text{ gradi di libert\`a}$$

Per verificare se ci sia differenza significativa tra i vari risultati ottenuti con le due tecniche, si fa l'*ipotesi nulla* che $\bar{d}=0$ e si esegue un *t*-test con:

$$t_{oss} = \frac{|\bar{d}|}{s_{\bar{d}}} \quad (3.4.1)$$

Esempio

Campione	metodo1	metodo2	d
1	71	76	-5
2	61	68	-7
3	50	48	2
4	60	57	3

$$\bar{d} = \frac{-5 - 7 + 2 + 3}{4} = -1.75$$

$$s_{\bar{d}} = \frac{\sqrt{\frac{(-5+1.75)^2 + (-7+1.75)^2 + (2+1.75)^2 + (3+1.75)^2}{4-1}}}{\sqrt{4}} = 2.50$$

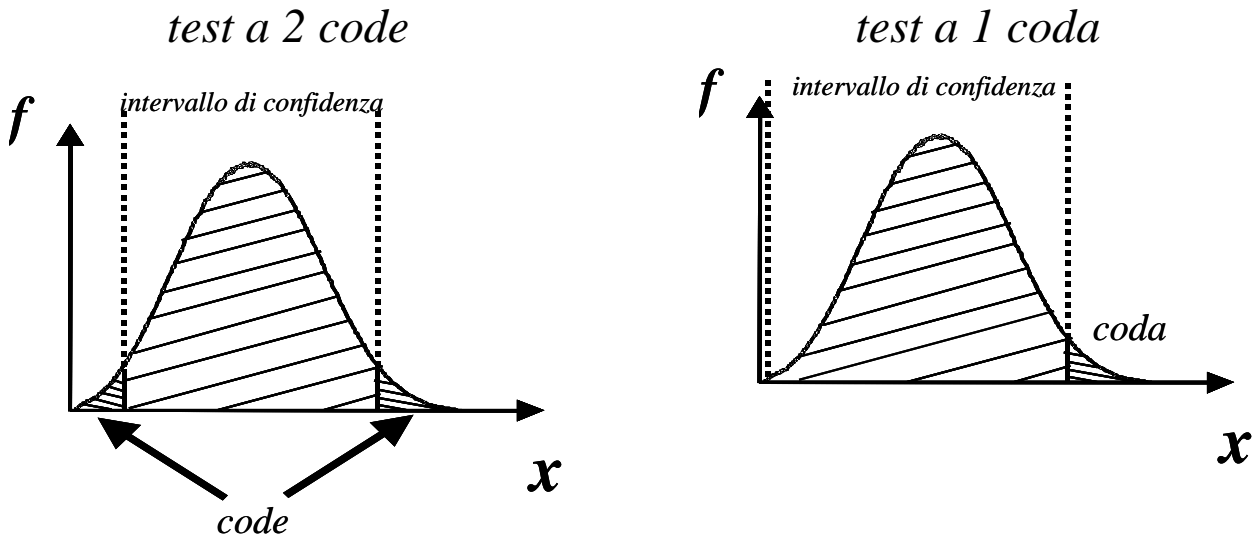
$$t_{oss} = \frac{1.75}{2.50} = 0.7$$

$$t_{0.05,3} = 3.18$$

$t_{oss} < t_{tab} \Rightarrow$ ipotesi nulla accettata:

i due metodi danno risultati *non significativamente differenti*.

3.5. Test a una coda, test a due code



I *t*-test finora descritti sono del tipo *a due code* perché si considera l'eventualità più generale che i dati di confronto rispetto ad una media possano cadere sia *al di sopra* (*coda* di destra) che *al di sotto* (*coda* di sinistra) rispetto all'*intervallo di confidenza*.

Quando c'è una motivazione sperimentale al fatto che il dato di confronto possa essere solo più grande o solo più piccolo rispetto ai dati trovati allora si applica un test *a 1 coda*.

Fissati un *livello di significatività* α e ν gradi di libertà, si ha:

$$t_{\alpha,\nu}^{1coda} = t_{2\alpha,\nu}^{2code} \quad (3.5.1)$$

$$t_{\alpha,\nu}^{1coda} < t_{\alpha,\nu}^{2code} \quad (3.5.2)$$

Dove non specificato, t è riferito al caso di *2 code*.

La funzione di Excel™ $\text{INV.T}(\alpha;\nu)$ è riferita al caso di *2 code*.

Esempio

Titolando 25 ml di un acido forte 0.1 M con una base forte 0.1 M, ci si è resi conto di avere usato come indicatore fenolftaleina troppo diluita e di avere apprezzato la comparsa del colore rosso in ritardo rispetto al punto equivalente, commettendo un errore in eccesso.

Si vuole verificare, con un *livello di confidenza* del 95%, se questo errore ha comportato una significativa mancanza di *esattezza*.

Risultati per il volume V di titolante aggiunto (ml).

25.06, 25.18, 24.87, 25.51, 25.34, 25.41

$$\bar{V} = 25.228 \text{ ml}$$

$$s = 0.238 \text{ ml}$$

$$t_{oss} = \frac{|25.228 - 25|}{\frac{0.238}{\sqrt{6}}} = 2.35$$

$$t_{0.05,5}^{1\text{coda}} = t_{0.10,5}^{2\text{coda}} = 2.01$$

$t_{oss} > t_{tab} \Rightarrow$ il dato trovato è *significativamente maggiore* rispetto al dato atteso.

N.B. se si fosse usato come t_{tab} il valore $t_{0.05,5}^{2\text{coda}} = 2.57$ si sarebbe concluso che il dato trovato *non è significativamente differente* rispetto al dato atteso.

3.6. *F*-test per il confronto tra deviazioni standard

Supponiamo di sottoporre a misura *uno stesso campione con due metodi diversi*.

Vogliamo confrontare le *precisioni* dei due metodi.

Dobbiamo confrontare le *varianze* s_1^2 e s_2^2 ottenute coi due metodi.

Definizione: $F_{\alpha, \nu_1, \nu_2} = \frac{s_1^2}{s_2^2}$, ν_1 e ν_2 = numeri di gradi di libertà. (3.6.1)

A numeratore si pone sempre la varianza più grande.

Ipotesi nulla H_0 : che *non* ci sia differenza di precisione tra i 2 metodi.

3.6.1. *F*-test a 1 coda

Si applica quando si vuole verificare se il metodo 2 è *più preciso* del metodo 1, al *livello di significatività α* .

$$F_{oss} < F_{\alpha, \nu_1, \nu_2}^{1coda} \Rightarrow H_0 \text{ accettata}$$

il metodo 1 *non* è più preciso del metodo 2.

3.6.2. *F*-test a 2 code

Si applica quando si vuole verificare se ci sia *differenza significativa* tra le *precisioni* dei due metodi, al *livello di significatività α* .

$$F_{oss} < F_{\alpha, \nu_1, \nu_2}^{2code} \Rightarrow H_0 \text{ accettata}$$

le *precisioni* dei due metodi *non* sono *significativamente* differenti

$$F_{\alpha, \nu_1, \nu_2}^{1coda} < F_{\alpha, \nu_1, \nu_2}^{2code}$$

$$F_{\alpha, \nu_1, \nu_2}^{1coda} = F_{2\alpha, \nu_1, \nu_2}^{2code}$$

La funzione di Excel™ $\text{INV.F}(\alpha; \nu_1, \nu_2)$ è riferita al caso di *1 coda*.

Esempio

Un analista propone un nuovo metodo (metodo 2) ed afferma che esso è *più preciso* di un metodo già validato (metodo 1). I risultati sono:

Metodo	Media (mg ml ⁻¹)	s (mg ml ⁻¹)	Numero misure
1	72	3.31	8
2	72	1.51	8

$$F_{oss} = \frac{3.31^2}{1.51^2} = 4.8$$

$$F_{0.05,7,7}^{1\text{coda}} = 3.787 = \text{INV.F}(0.05;7;7)$$

$$F_{oss} > F_{\alpha, v_1, v_2}^{1\text{coda}} \Rightarrow H_0 \text{ rigettata}$$

il metodo 2 è più preciso del metodo 1.

Esempio

Un analista propone un nuovo metodo (metodo 2) e si chiede se la sua precisione sia *diversa* rispetto ad un metodo già validato (metodo 1).

I risultati sono:

Metodo	Media (mg ml ⁻¹)	s (mg ml ⁻¹)	Numero misure
1	28.0	0.3	10
2	26.25	0.23	10

$$F_{oss} = \frac{0.3^2}{0.23^2} = 1.7$$

$$F_{0.05,9,9}^{2\text{code}} = 4.026 = \text{INV.F}(0.025;9;9)$$

$$F_{oss} < F_{\alpha, v_1, v_2}^{2\text{code}} \Rightarrow H_0 \text{ accettata}$$

le precisioni del metodo 2 *non sono significativamente diverse*.

3.7. *Q*-test per la verifica di dati anomali

Supponiamo di avere sottoposto uno stesso campione a più misure con lo stesso metodo. Uno dei dati risulta ad occhio molto diverso dagli altri. Si vuole verificare se esso sia *anomalo*.

Definizione:

$$Q = \frac{|\text{valore sospetto} - \text{valore più vicino}|}{(\text{valore massimo} - \text{valore minimo})} \quad (3.7.1)$$

$$Q_{oss} < Q_{tabulato} \Rightarrow \text{il dato sospetto non è rigettabile} \quad (3.7.2)$$

Valori critici di *Q* al livello di significatività del 5%

Numero misure	<i>Q</i>
4	0.831
5	0.717
6	0.621
7	0.570
8	0.524
9	0.492
10	0.464

Esempio

Dati raccolti (mg l⁻¹)

0.403, 0.410, 0.401, 0.380

$$Q = \frac{|0.380 - 0.401|}{0.410 - 0.380} = 0.7$$

Il dato sospetto risulta *non significativamente* anomalo al livello di *confidenza* del 95%.

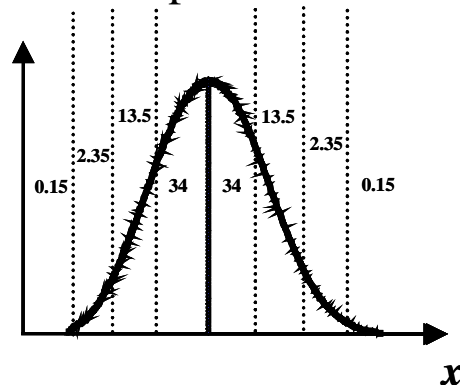
3.8. Test χ^2 per la verifica della normalità di una distribuzione

Supponiamo di avere effettuato n misure ($n > 50$) di una grandezza x . Vogliamo verificare se le misure ripetute sono conformi ad una distribuzione gaussiana.

Ipotesi nulla: non c'è differenza significativa tra la distribuzione delle misure osservate e una distribuzione gaussiana.

Dalle misure ripetute calcoliamo \bar{x} e s .

Suddividiamo l'asse x in k intervalli. È calcolabile la probabilità P_k che un dato cada nel k -esimo intervallo. Per es. nel caso $k=8$ intervalli ottenuti spostandosi di una σ a partire dalla media si ha:



Chiamiamo E_k il numero di misure che ci si attende che cada nell'intervallo k -esimo. È:

$$E_k = n P_k \quad (3.8.1)$$

Chiamiamo O_k il numero di misure che si osservano cadere nell'intervallo k -esimo.

Definiz.: $\chi^2 = \sum_k \frac{(O_k - E_k)^2}{E_k}$, $k =$ numero di gradi di libertà (3.8.2)

Fissato il *livello di significatività* α si ha che:

$$\chi_{\text{oss}}^2 < \chi_{\alpha, k}^2 \Rightarrow \text{i dati sono distribuiti normalmente} \quad (3.8.3)$$

I valori critici di χ^2 sono tabulati e calcolabili con la funzione di Excel™ $\text{INV.CHI}(\alpha; k)$

3.9. Test della frequenza cumulativa

Supponiamo di avere effettuato n misure ($n < 50$) di una grandezza x . Vogliamo verificare se le misure ripetute sono conformi ad una distribuzione gaussiana.

Ipotesi nulla: non c'è differenza significativa tra la distribuzione delle misure osservate e una distribuzione gaussiana.

Si ordinano i dati in modo crescente, assegnando a ciascuno un numero d'ordine k , chiamato anche *frequenza cumulativa*.

Si calcola:

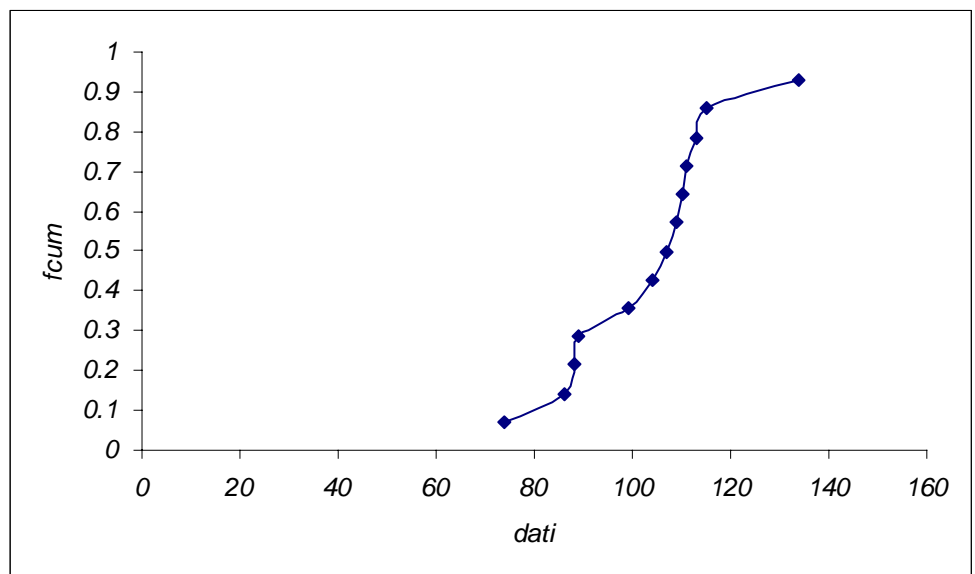
$$f_{cum} = \frac{k}{n+1} \quad (3.9.1)$$

E si riporta in grafico f_{cum} in funzione dei dati.

Se i dati appartengono ad una distribuzione gaussiana si deve ottenere una sigmoide.

Esempio.

dati	k	f_{cum}
74	1	0.071429
86	2	0.142857
88	3	0.214286
89	4	0.285714
99	5	0.357143
104	6	0.428571
107	7	0.5
109	8	0.571429
110	9	0.642857
111	10	0.714286
113	11	0.785714
115	12	0.857143
134	13	0.928571



3.10. Tipi di errore

➤ **ERRORI DI PRIMO TIPO:**

- l'ipotesi nulla è **VERA** ma viene **RIGETTATA**

cioè si cade erroneamente fuori dall'intervallo di confidenza

- Es: si stabilisce che valore noto e valore sperimentale

sono significativamente diversi quando

non sono significativamente diversi

➤ **SI MINIMIZZANO MASSIMIZZANDO IL LIVELLO DI CONFIDENZA OVVERO MINIMIZZANDO IL LIVELLO DI SIGNIFICATIVITÀ**

➤ **ERRORI DI SECONDO TIPO:**

- l'ipotesi nulla è **FALSA** ma viene **ACCETTATA**

cioè si cade erroneamente dentro l'intervallo di confidenza

- Es: si stabilisce che valore noto e valore sperimentale

non sono significativamente diversi quando

sono significativamente diversi

➤ **SI MINIMIZZANO MINIMIZZANDO IL LIVELLO DI CONFIDENZA OVVERO MASSIMIZZANDO IL LIVELLO DI SIGNIFICATIVITÀ**

4. CONTROLLO DI QUALITÀ

4.1 Parametri di qualità nelle misure ripetute

4.1.1. *Accuratezza: esattezza e precisione* Si definisce ACCURATEZZA la contemporanea sussistenza di ESATTEZZA e PRECISIONE.

Si definisce ESATTEZZA l'accordo tra il valor medio ottenuto da un ampio set di misure e un valore di riferimento accettato.

La mancanza di esattezza è in relazione con gli *errori sistematici*

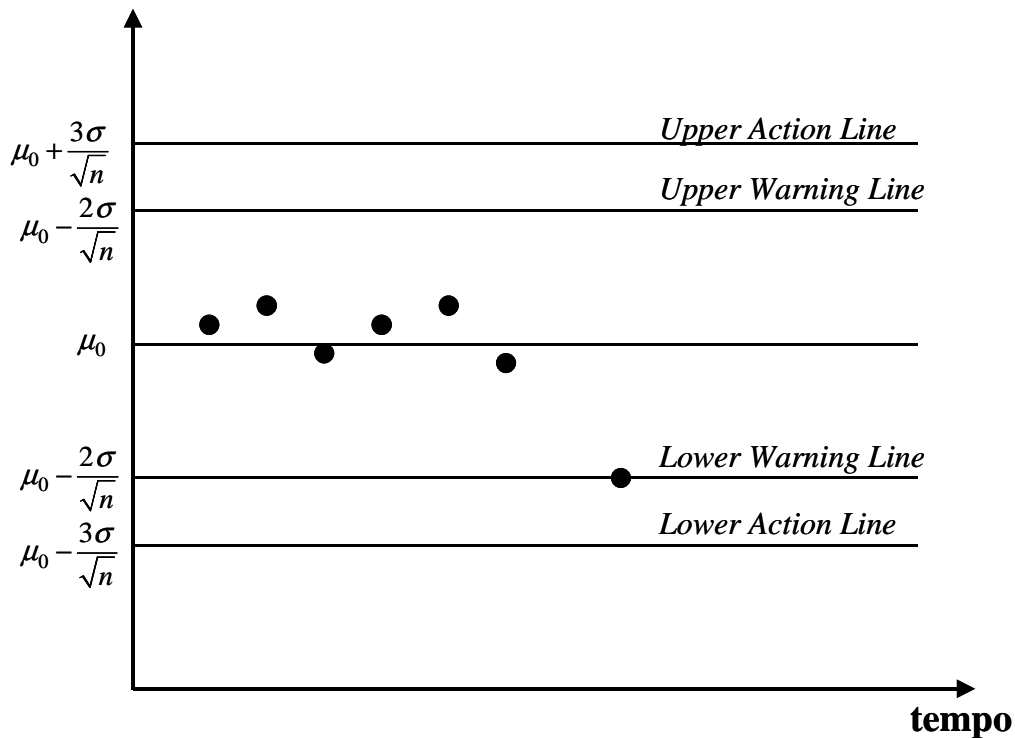
Si definisce PRECISIONE l'accordo, ovvero la vicinanza reciproca, tra i risultati di più misure replicate.

La mancanza di precisione è in relazione con gli *errori casuali*.

4.1.2. *Ripetibilità e riproducibilità*

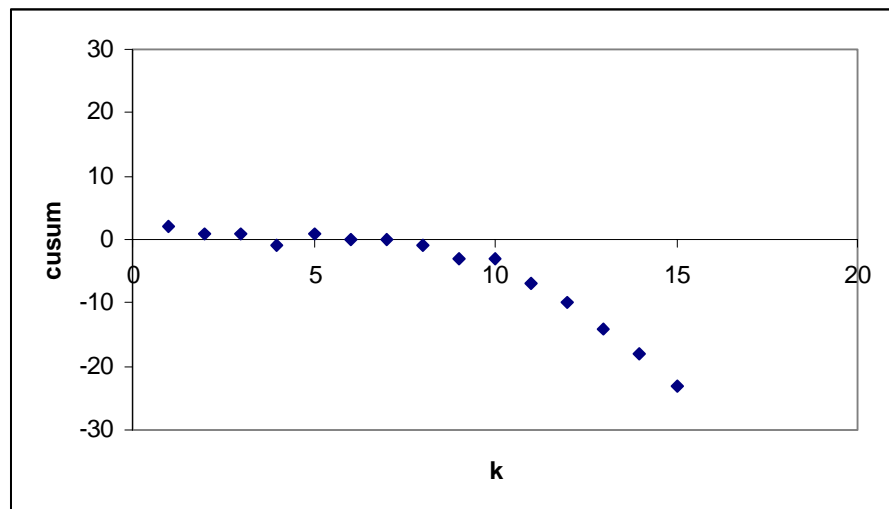
- *Ripetibilità ristretta*: precisione ottenuta con lo stesso operatore, la stessa strumentazione, in un breve lasso di tempo
- *Ripetibilità intermedia*: precisione ottenuta con diversi operatori, la stessa strumentazione, in un breve lasso di tempo
- *Riproducibilità*: precisione ottenuta con diversi operatori, strumentazioni diverse e/o in un lungo lasso di tempo

4.2 Carte di controllo



4.3 Carte *cusum*

<i>k</i>	<i>dato</i>	<i>dato-dato atteso</i>	<i>cusum</i>
1	82	2	2
2	79	-1	1
3	80	0	1
4	78	-2	-1
5	82	2	1
6	79	-1	0
7	80	0	0
8	79	-1	-1
9	78	-2	-3
10	80	0	-3
11	76	-4	-7
12	77	-3	-10
13	76	-4	-14
14	76	-4	-18
15	75	-5	-23



Dato atteso: 80.

Definizione: $cusum = \sum_i (\text{dato} - \text{dato atteso})_i$

Se il metodo analitico è sotto controllo, gli scarti dei valori misurati rispetto al valore atteso saranno casualmente positivi e negativi e la somma cumulativa sarà zero. Se il metodo va fuori controllo, *cusum* devierà rispetto al valore nullo.

5. METODI QUANTITATIVI NELL'ANALISI STRUMENTALE

Nella chimica analitica strumentale la quantificazione di un analita viene ottenuta *indirettamente* a partire da un *segnale analitico*.

Vi sono tre fondamentali metodi quantitativi:

- 1) Metodo della curva di taratura
- 2) Metodo delle aggiunte standard
- 3) Metodo dello standard interno

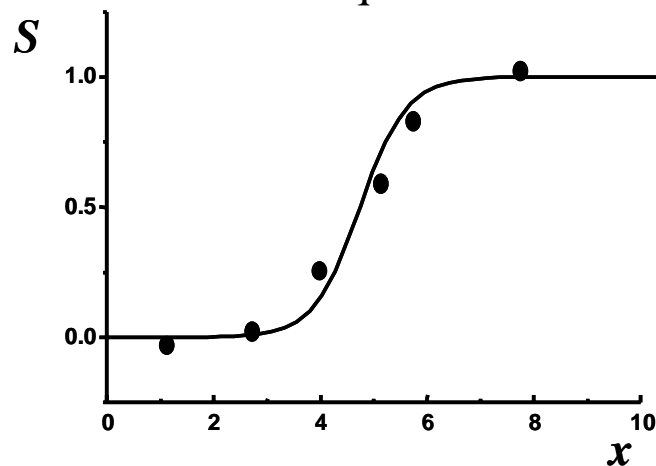
5.1 Metodo della curva di taratura

5.1.1 Procedimento:

- 1) Si preparano vari campioni contenenti *quantità note* (x_i) di analita nella stessa *matrice* del campione incognito o in una matrice equivalente. Questi campioni vengono chiamati *standard*.
- 2) Si sottopone ciascuno *standard* alla misura strumentale. Per lo standard i -esimo si rileverà un segnale S_i .
- 3) Si riportano in grafico i punti sperimentali (x_i, S_i).
- 4) Si applica un metodo statistico per determinare la curva che meglio si adatta ai punti sperimentali, ovvero si esegue un *fitting*.
- 5) Si sottopone il campione incognito alla misura strumentale, rilevando un segnale S_0 .
- 6) Si esegue una *interpolazione* sulla curva di taratura per calcolare il valore di x , chiamato x_0 , che corrisponde a S_0 .

5.1.2 La curva di taratura

La curva di taratura è la curva corrispondente all'andamento del segnale analitico in funzione della quantità di analita.



5.1.3 Correlazione

Condizione necessaria affinché si possa usare una curva di taratura è che esista una *correlazione* tra il *segnale analitico* S e la *quantità di analita* x , cioè che S dipenda da x . L'esistenza di correlazione tra y e x si può verificare mediante un parametro statistico chiamato *coefficiente di correlazione* R :

$$r = \frac{\sum_i [(x_i - \bar{x})(y_i - \bar{y})]}{\left\{ \left[\sum_i (x_i - \bar{x})^2 \right] \left[\sum_i (y_i - \bar{y})^2 \right] \right\}^{1/2}} \quad (\text{caso di una retta}) \quad (5.1.1)$$

Buona correlazione lineare \Rightarrow il valore assoluto di r è vicino a 1.

5.1.4 t-test per la correlazione

Ipotesi nulla: non c'è correlazione tra y e x

$$t_{oss} = \frac{|r| \sqrt{n-2}}{\sqrt{1-r^2}}, \quad t_{oss} > t_{tab} \Rightarrow \text{c'è correlazione} \quad (5.1.2)$$

5.1.5 Intervallo dinamico e intervallo lineare

Si definisce *intervallo dinamico* quell'intervallo di valori della *quantità* di analita (x) entro il quale il *segnale analitico* S varia al variare di tale quantità.

Si definisce *intervallo lineare* quell'intervallo di valori della *quantità* di analita (x) entro il quale l'intervallo dinamico è lineare.

È particolarmente conveniente lavorare nell'intervallo lineare, cioè è particolarmente conveniente il caso della *retta di taratura*, perché le elaborazioni statistiche per il trattamento dati sono molto semplici.

5.1.6 Sensibilità

Si definisce *sensibilità* la *pendenza* della curva di taratura. Nel caso di una retta di taratura $y=a+bx$ la sensibilità è uguale a b .

5.1.7 Applicabilità del metodo della curva di taratura

Per poter applicare il metodo della curva di taratura è necessario:

- 1) disporre della matrice che compone il campione incognito, oppure dimostrare che non esiste *effetto matrice*, cioè che il segnale dipende solo da x ed è indipendente dalla matrice.
- 2) Disporre di quantità adeguate di *analita puro*.
- 3) Scegliere i valori di x_i in un intervallo che comprenda x_0 .

5.1.8 Vantaggi del metodo della curva di taratura

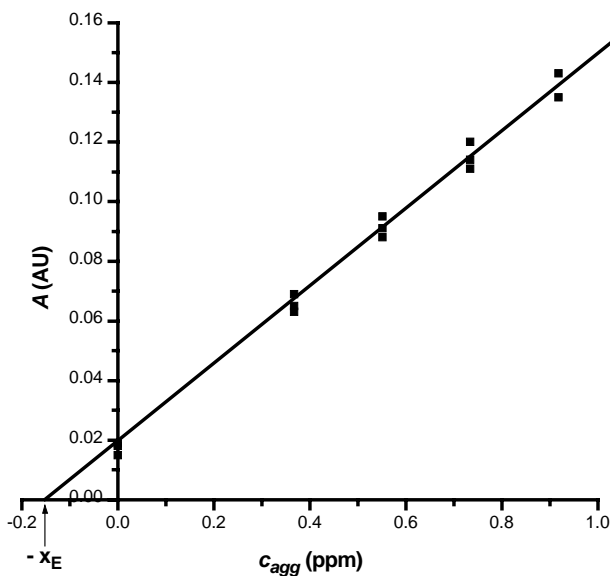
La curva di taratura può essere usata per molte analisi, salvo controllarne periodicamente la validità mediante uno *standard* di riferimento.

5.2 Metodo delle aggiunte standard

5.2.1. Procedimento:

- 1) Si preparano n aliquote uguali dello *stesso* campione incognito.
- 2) Alle aliquote dalla 2 alla n si aggiungono quantità note di analita puro, mentre non si fa alcuna aggiunta all'aliquota 1.
- 3) Si portano tutte le aliquote allo stesso volume, che può essere uguale a quello iniziale se le aggiunte sono state tali da non modificare significativamente il volume delle aliquote. Si ottengono così n standard caratterizzati da una certa *quantità aggiunta* $x_{agg,i}$.
- 4) Si sottopone ciascuno *standard* alla misura strumentale. Per lo standard i -esimo si rileverà un segnale S_i .
- 5) Si riportano in grafico i punti sperimentali ($x_{agg,i}$, S_i).
- 6) Si applica un metodo statistico per determinare la retta che meglio si adatta ai punti sperimentali.
- 7) Si esegue una *estrapolazione* della retta in modo da determinare l'*ascissa all'origine*. Si ottiene così il valore incognito x_E .

5.2.2. La retta delle aggiunte standard



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

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

5.2.3. Applicabilità del metodo delle aggiunte standard

Per poter applicare il metodo delle aggiunte standard è necessario:

- 1) Lavorare all'interno dell'*intervallo lineare*
- 2) Disporre di quantità adeguate di *analita puro*.

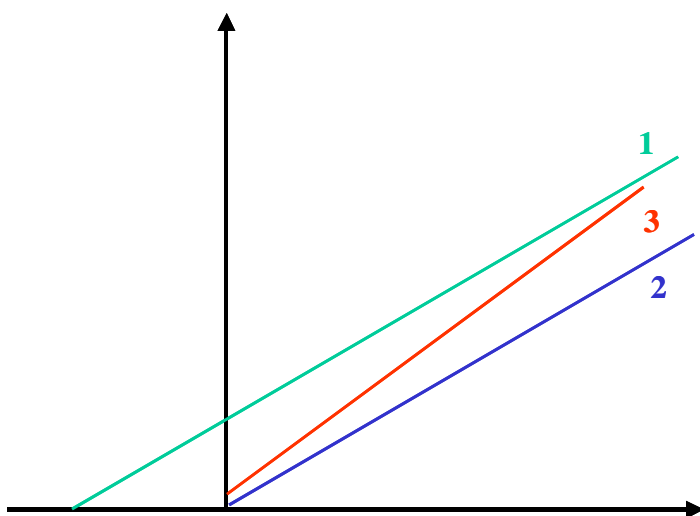
5.2.4. Vantaggi del metodo delle aggiunte standard

Permette una determinazione quantitativa senza conoscere la natura della *matrice*

5.2.5. Svantaggi del metodo delle aggiunte standard

- 1) È un metodo per *estrapolazione* il che comporta minore *precisione*
- 2) Ogni volta che si fa una determinazione occorre preparare gli standard e costruire la retta

5.2.6. Verifica dell'effetto matrice



- 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

5.3 Metodo dello standard interno

5.3.1 Caso 1: uso dei fattori di risposta

Procedimento

- 1) Si individua una specie chimica molto simile all'analita per proprietà chimico-fisiche (*standard interno*).
- 2) Si prepara un campione (*standard 1*) contenente *quantità note* di analita e di *standard interno*, indicate con x_A e x_Z rispettivamente.
- 3) Si sottopone lo *standard 1* alla misura strumentale. Si rileverà un segnale S_Z per lo *standard interno* e un segnale S_A per l'analita.
- 4) Sarà:

$$S_A : S_Z = k_A x_A : k_Z x_Z \quad (5.3.1)$$

$$\text{Definizione: fattore di risposta } k = \frac{k_A}{k_Z} \quad (5.3.2)$$

$$\frac{S_A}{S_Z} = \frac{k \ x_A}{x_Z} \quad (5.3.3)$$

$$k = \frac{S_A \ x_Z}{S_Z \ x_A} \quad (5.3.4)$$

- 5) Dai dati noti, si calcola k , eseguendo m misure ripetute.
- 6) Partendo dal campione incognito, si prepara un campione (*standard 2*) in cui l'analita è presente in quantità incognita x e lo standard interno è presente in quantità nota z dello stesso ordine di grandezza di x .
- 7) Si sottopone lo *standard 2* alla misura strumentale. Si rileverà un segnale S_Z per lo *standard interno* e un segnale S_X per l'analita.
- 8) Sarà:

$$\frac{S_X}{S_Z} = \frac{k \ x}{z} \quad (5.3.5)$$

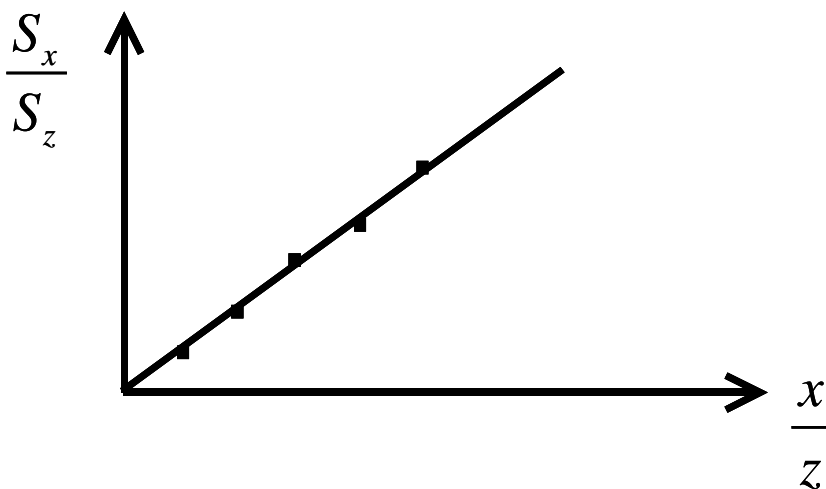
$$x = \frac{z \ S_X}{k \ S_Z} \quad (5.3.6)$$

- 9) Si eseguono misure ripetute e si calcola x .

5.3.2 Caso 2: uso di una retta di taratura

Procedimento

- 1) Si preparano vari campioni contenenti *quantità note* (x_i) di analita e *la stessa* quantità z di standard interno. Questi campioni vengono chiamati *standard*.
- 2) 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.
- 3) Si riportano in grafico i punti sperimentali ($x_i, S_i/S_Z$).
- 4) Si applica un metodo statistico per determinare la retta che meglio si adatta ai punti sperimentali, ovvero si esegue un *fitting*.
- 5) Si sottopone il campione incognito alla misura strumentale, rilevando un segnale S_0 .
- 6) Si esegue una *interpolazione* sulla curva di taratura per calcolare il valore x_0/z , che corrisponde a S_0/S_Z , e da esso si ricava x_0 .



$$\frac{S_x}{S_z} = a + b \frac{x}{z}$$

5.3.3 Applicabilità del metodo dello standard interno

Le condizioni necessarie per poter applicare il metodo sono:

- 1) Disporre di uno *standard interno* molto simile all'analita.
- 2) Lavorare all'interno dell'*intervallo lineare*.
- 3) Verificare che il *recupero analitico relativo* di *analita e standard interno* sia 100%.

5.3.4 Vantaggi del metodo dello standard interno

- 1) è applicabile anche laddove una *retta di taratura* darebbe scarsa precisione.
- 2) Rende possibile una determinazione quantitativa anche quando *non* si disponga di uno *standard puro di analita* ma si conoscano i fattori di risposta.

5.3.5 Quando usare i casi 1 e 2

Caso della retta

Si usa quando il campione incognito contiene naturalmente uno *standard interno* e quando lo *standard interno* è presente in quantità di diverso ordine di grandezza rispetto all'analita.

Esempio: ricerca del metanolo nel vino.

Caso del fattore di risposta

Si usa ove non sia opportuno l'altro caso.

5.4 Regressione lineare

- Un problema statistico di *regressione* si presenta quando si hanno n coppie di valori (x, y) che corrispondono a *dati sperimentali* e si vuole determinare la curva $y=f(x)$ che meglio si adatta ai punti sperimentali. Si parla di *regressione di y su x* .
- I metodi analitici quantitativi basati su una *curva di taratura* richiedono una *regressione*: occorre determinare la curva che meglio si adatta alle coppie di valori:
(*quantità nota, segnale analitico*).

In questo caso, condizione necessaria per poter eseguire una *regressione di y su x* è che l'errore sulla *quantità* sia trascurabile rispetto all'errore sul *segnale* e che l'errore su y sia indipendente da x .

- Il procedimento matematico che permette di determinare la *curva di regressione* consiste nel minimizzare la sommatoria degli scarti quadratici medi delle ordinate dei *dati sperimentali* rispetto alle corrispondenti ordinate dei *punti sulla curva*. Ciò equivale a minimizzare la quantità:

$$s_{y/x} = \sqrt{\frac{\sum_i^n (y_i - \hat{y}_i)^2}{\nu}} \quad (5.4.1)$$

y_i = ordinata *sperimentale* corrispondente al punto di ascissa x_i

\hat{y}_i = ordinata del *punto sulla curva di regressione* avente ascissa x_i

ν = numero di *gradi di libertà* del problema

- Una volta determinata la funzione analitica della *curva di taratura*, è necessario poter *interpolare* il valore di *quantità nota* che corrisponde al *segnale analitico* del campione incognito.

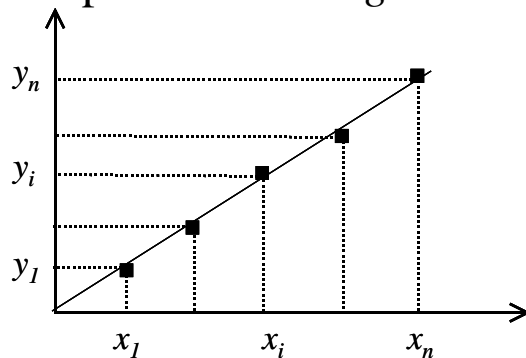
5.4.1 Regressione su una retta

- Un problema statistico di *regressione lineare* è un problema di regressione in cui i *dati sperimentali* hanno forma (x_i, y_i) e la *curva di regressione* ha la forma:

$$y = a_0 + a_1 f_1(x) + \dots + a_j f_j(x) + \dots + a_k f_k(x) \quad (5.4.2)$$

La regressione di y su x fornisce i valori dei parametri a_i .

- Il più comune problema di *regressione lineare* è *su una retta*:



$$y = a + b x$$

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (5.4.3)$$

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad (5.4.4)$$

$$b = \frac{\sum_{i=1}^n [(x_i - \bar{x})(y_i - \bar{y})]}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \quad (5.4.5)$$

$$a = \bar{y} - b \bar{x} \quad (5.4.6)$$

$$s_{y/x} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2}} = \sqrt{\frac{\sum_{i=1}^n [y_i - (a + b x_i)]^2}{n-2}} \quad (5.4.7)$$

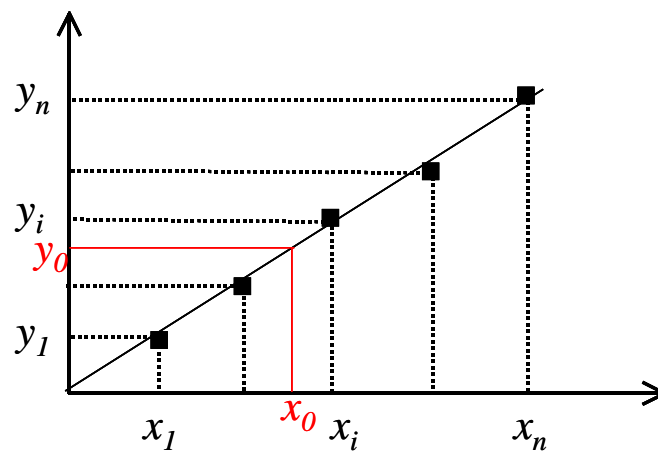
$$s_a = s_{y/x} \sqrt{\frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}} \quad (5.4.8)$$

$$\text{errore su } a = t_{\alpha, n-2} s_a \quad (5.4.9)$$

$$s_b = \frac{s_{y/x}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (5.4.10)$$

$$\text{errore su } b = t_{\alpha, n-2} s_b \quad (5.4.11)$$

5.4.2 Interpolazione su una retta di taratura



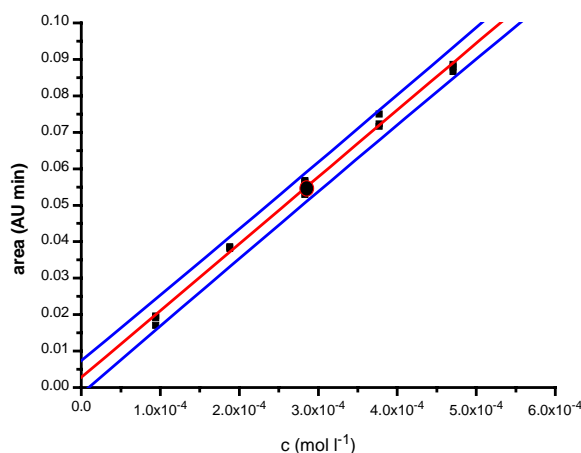
$$x_0 = \frac{y_0 - a}{b} \quad (5.4.12)$$

$$s_{x_0} = \frac{s_{y/x}}{b} \left[\frac{1}{m} + \frac{1}{n} + \frac{(y_0 - \bar{y})^2}{b^2 \sum_{i=1}^n (x_i - \bar{x})^2} \right]^{1/2} \quad (5.4.13)$$

m = numero di ripetizioni per y_0

$$\Delta x_0 = t_{\alpha, n-2} \cdot s_{x_0} \quad (5.4.14)$$

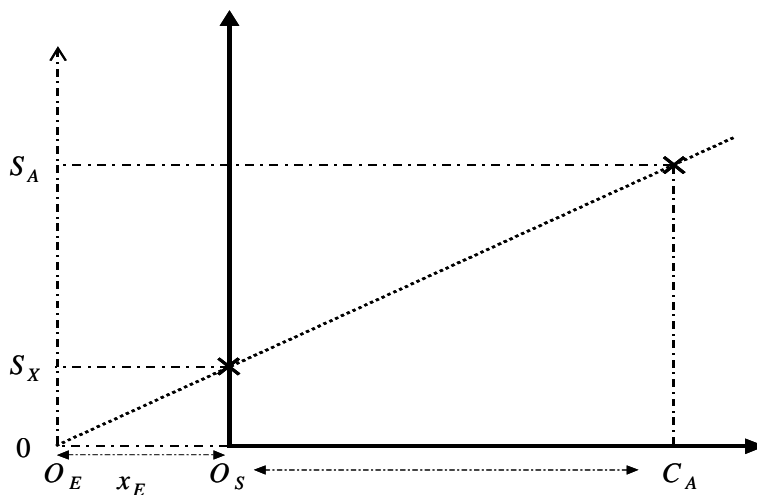
5.4.3 Iperboli di confidenza (regressione non pesata)



La retta di regressione di y su x passa sempre per il centroide dei punti (\bar{x}, \bar{y}) .

Dall'equazione 5.4.13 si possono calcolare le *iperboli di confidenza*.

5.4.4 Stima dell'errore nel metodo delle aggiunte standard



$$s_{x_E} = \frac{s_{y/x}}{b} \sqrt{\frac{1}{n} + \frac{\bar{y}^2}{b^2 \sum_i (x_i - \bar{x})^2}} \quad (5.4.15)$$

5.5 Regressione lineare pesata

Quando non è verificata la condizione che l'errore sulla y sia indipendente da x è possibile eseguire una regressione lineare, purché ai vari punti sperimentali venga attribuito un opportuno *peso statistico*.

$$w_i = n \frac{\frac{1}{s_i^2}}{\sum_{i=1}^n \frac{1}{s_i^2}} \quad (5.5.1)$$

$$\bar{x}_w = \frac{\sum_{i=1}^n w_i x_i}{n} \quad (5.5.2)$$

$$\bar{y}_w = \frac{\sum_{i=1}^n w_i y_i}{n} \quad (5.5.3)$$

$$b = \frac{\sum_{i=1}^n w_i x_i y_i - n \bar{x}_w \bar{y}_w}{\sum_{i=1}^n w_i x_i^2 - n \bar{x}_w^2} \quad (5.5.4)$$

$$a = \bar{y}_w - b \bar{x}_w \quad (5.5.5)$$

$$s_{(y/x)_w} = \left[\frac{\left(\sum_{i=1}^n w_i y_i^2 - n \bar{y}_w^2 \right) - b^2 \left(\sum_{i=1}^n w_i x_i^2 - n \bar{x}_w^2 \right)}{n-2} \right]^{1/2} \quad (5.5.6)$$

$$s_a = s_{y/x} \sqrt{\frac{\sum_i w_i x_i^2}{n \sum_i (w_i x_i - \bar{x})^2}} \quad (5.5.7)$$

$$\Delta a = t_{\alpha, n-2} s_a \quad (5.5.9)$$

$$s_b = \frac{s_{y/x}}{\sqrt{\sum_i (w_i x_i - \bar{x})^2}} \quad (5.5.10)$$

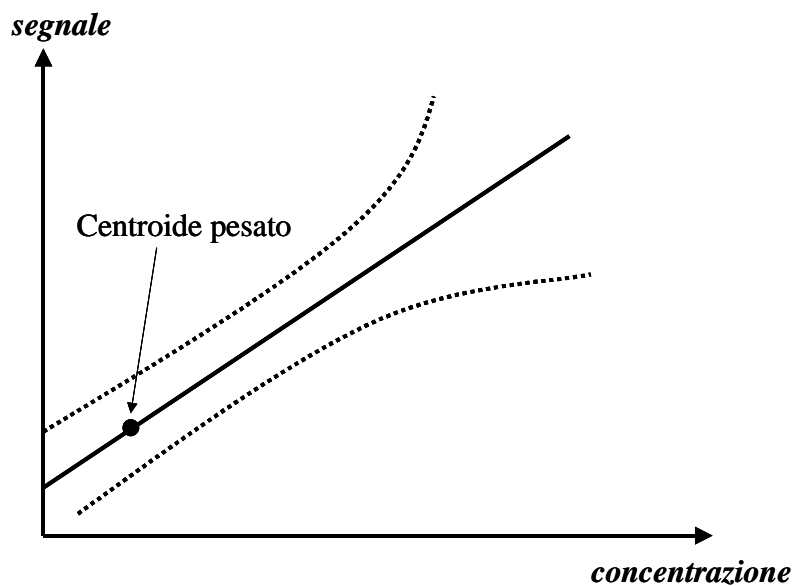
$$\Delta a = t_{\alpha, n-2} s_b \quad (5.5.11)$$

5.5.1 Interpolazione su una retta di regressione pesata

$$x_0 = \frac{y_0 - a}{b} \quad (5.5.12)$$

$$s_{x_0} = \frac{s_{(y/x)_w}}{b} \left[\frac{1}{w_0} + \frac{1}{n} + \frac{(y_0 - \bar{y}_w)^2}{b^2 \left(\sum_{i=1}^n w_i x_i^2 - n \bar{x}_w^2 \right)} \right]^{1/2} \quad (5.5.13)$$

$$\Delta x_0 = t_{\alpha, n-2} \cdot s_{x_0} \quad (5.5.14)$$



La retta di regressione di y su x passa sempre per il centroide pesato dei punti (\bar{x}_w, \bar{y}_w) .

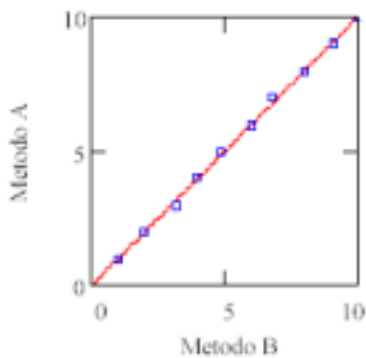
Dall'equazione 5.5.13 si possono calcolare le *iperboli di confidenza*.

5.6 Confronto tra metodi mediante regressione lineare

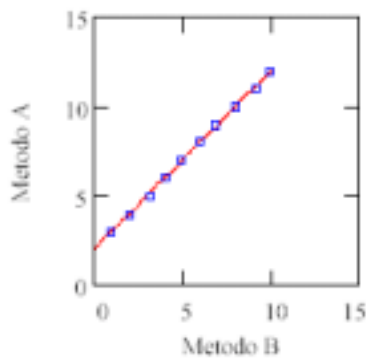
A e B sono due metodi analitici di cui il secondo molto più preciso del primo.

Si sottopongono a misura n campioni, ordinati per quantità crescente di analita, con entrambi i metodi e si riportano i risultati di A in funzione di quelli di B.

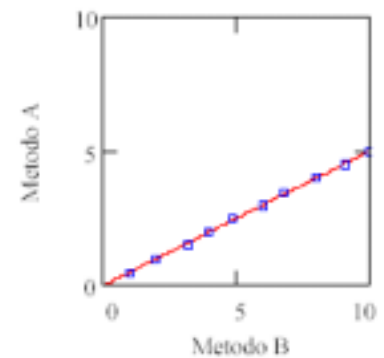
Si esegue una regressione lineare. Se i metodi danno risultati non significativamente diversi, allora l'intercetta deve essere non significativamente diversa da zero e la pendenza deve essere non significativamente diversa da uno.



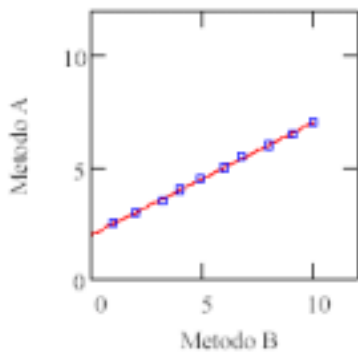
Teorico



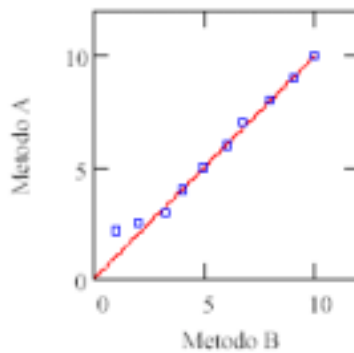
Errore sistematico costante positivo



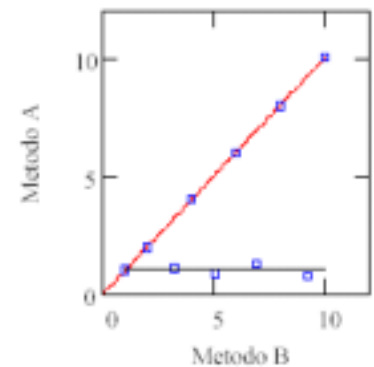
Errore sistematico proporzionale negativo



Errore sistematico costante positivo più errore sistematico proporzionale negativo



Deviazione dalla linearità



Problemi di speciazione

5.7 Rivelabilità e quantificabilità

I metodi analitici strumentali hanno un fondamentale vantaggio rispetto a quelli non strumentali: permettono di *rivelare e quantificare quantità più piccole* di analita.

Allo scopo di quantificare il “più piccole” si definiscono i seguenti *parametri di qualità*.

- LIMITE DI DECISIONE

Quantifica la capacità di affermare l'*assenza* dell'analita.

Si definisce x_D la minima quantità di analita che dà un segnale (y_D) *diverso* da quello del *bianco*.

- LIMITE DI RIVELABILITÀ

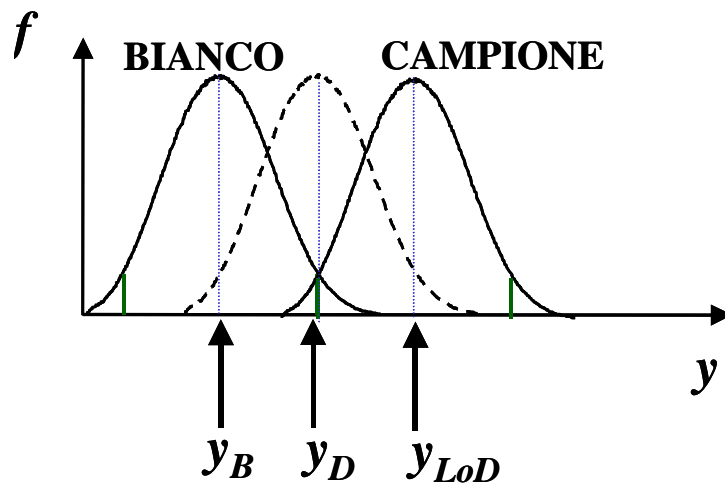
Quantifica la capacità di affermare la *presenza* dell'analita.

Si definisce x_{LoD} la minima quantità di analita che dà un segnale (y_{LoD}) *significativamente diverso* da quello del *bianco*.

- LIMITE DI QUANTIFICAZIONE

Misura la capacità di eseguire una *misurazione quantitativa*.

Si definisce x_{LoQ} la minima quantità di analita che può essere *quantificata* con un certo *livello di confidenza*.

5.7.1. Convenzione sulla stima di LoD e LoQ 

Una convenzione molto diffusa sul LoD stabilisce che, nel caso di segnale crescente con la quantità, si ponga:

$$y_{LoD} = y_B + 3 s_B \quad (5.7.1)$$

Il coefficiente 3 è in relazione ad un *livello di confidenza* del 99.7% nello stabilire che la quantità di analita è *significativa*: si ha una probabilità dello 0.3% di commettere un *errore del I tipo* affermando erroneamente che la quantità di analita è *significativa*.

Una convenzione molto diffusa sul LoQ stabilisce che, nel caso di segnale crescente con la quantità, si ponga:

$$y_{LoQ} = y_B + 6 s_B \quad (5.7.2)$$

Il coefficiente 6 di tale equazione è scelto in modo da rendere trascurabili *errori del I tipo*.

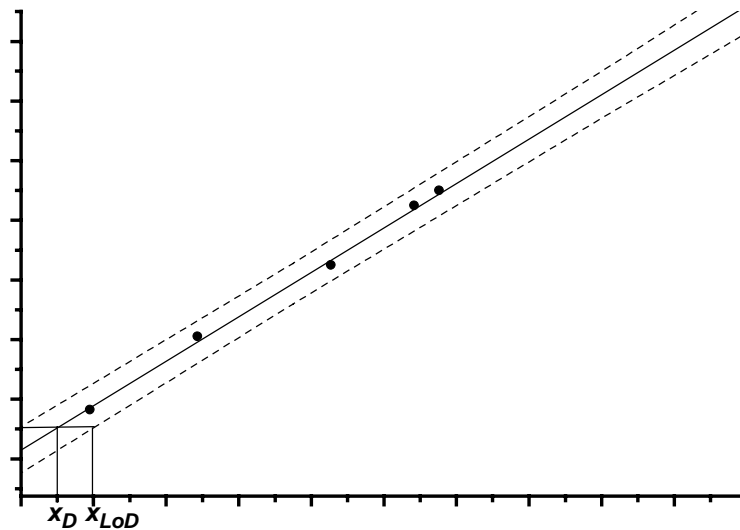
Esistono 3 modi di *stimare* x_{LoD} basandosi sulla convenzione qui proposta:

- 1) da misure ripetute sul bianco
- 2) da una retta di taratura o delle aggiunte standard
- 3) dal rapporto segnale/rumore

5.7.2. Stima del LoD da misure ripetute sul bianco

Si sottopone il *bianco* a misura strumentale e si applica una statistica delle misure ripetute per determinare y_{LoD} , da cui poi si ricava x_{LoD} .

5.7.3. Stima del LoD dalla retta di taratura



I modo: da $s_{y/x}$ e dalla sensibilità

$$y_{LoD} = y_B + 3 s_B = a + 3 s_{y/x} = a + b x_{LoD} \quad (5.7.3)$$

$$x_{LoD} = 3 \frac{s_{y/x}}{b}$$

II modo: dalle iperboli di confidenza

x_{LoD} è l'ascissa del punto dell'iperbole inferiore che ha come ordinata l'intercetta dell'iperbole superiore.

5.7.4. Stima del LoD dalla retta delle aggiunte standard

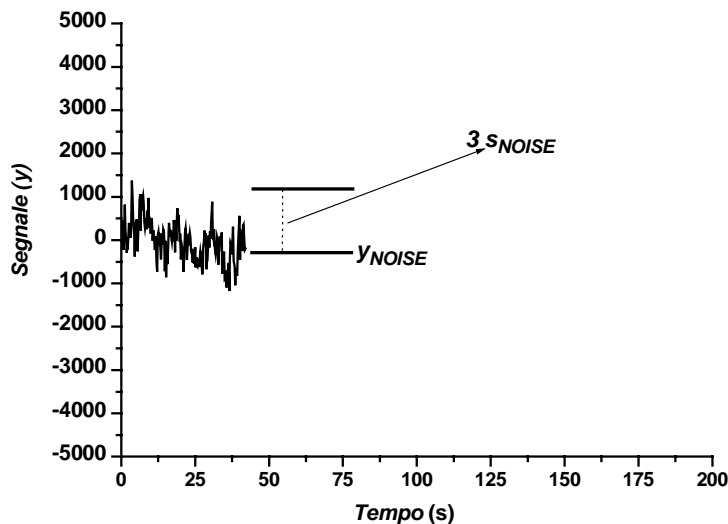
Poiché la retta delle aggiunte standard ha la stessa pendenza che avrebbe una retta di taratura ottenuta da standard preparati nella *stessa* matrice, vale ancora:

$$x_{LoD} = 3 \frac{s_{y/x}}{b} \quad (5.7.4)$$

Invece per ottenere y_{LoD} occorre eseguire una regressione lineare sui dati $(x_{agg,i} + |x_E|, y_i)$: si ottiene una intercetta $a_{traslata}$ che è la migliore stima del minimo segnale significativamente diverso dal quello del bianco.

5.7.5. Stima del LoD dal rapporto S/N

Il *rumore strumentale* è la variazione del segnale nel tempo.



y_{LoD} è calcolato come segue:

$$y_{LoD} = y_{NOISE} + 3 s_{NOISE} \quad (5.7.5)$$

5.8 Un test di linearità

Quando si sceglie il metodo quantitativo della retta di taratura occorre essere certi che tra il *segnale* e la *quantità di analita* vi sia una relazione lineare.

La regressione lineare su una retta $y=a+bx$ fornisce il *coefficiente di correlazione* r , ed fatto che $|r|$ sia vicino a 1 è una condizione solo necessaria, ma non sufficiente per affermare che scegliendo la retta come funzione fittante sia stata fatta la scelta migliore.

Un verifica di linearità tra segnale e quantità si può ottenere applicando il test *lack of fit* (LoF) descritto di seguito.

1) costruire di una retta di taratura eseguendo m misure ripetute su ciascuno degli n standard

2) Calcolare la VARIANZA ATTORNO ALLA RETTA (SS_{LoF})

$$SS_{LoF} = \sum_{i=1}^n m(\hat{y}_i - \bar{y})^2, \quad v_{LoF} = n - 2 \quad (5.8.1)$$

3) Calcolare la VARIANZA TOTALE (SS_{PE})

$$SS_{PE} = \sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \bar{y}_i)^2, \quad v_{PE} = n(m-1) \quad (5.8.2)$$

4) Calcolare il parametro F come segue

$$F = \frac{SS_{LoF} / v_{LoF}}{SS_{PE} / v_{PE}} \quad (5.8.3)$$

5) Test:

$$F_{OSS} < F_{\alpha, v_{LoF}, v_{PE}} \Rightarrow c' \text{ è linearità} \quad (5.8.4)$$

METODI NON PARAMETRICI

Si parla di *metodi statistici non parametrici* quando non si fa alcuna assunzione sul tipo di distribuzione che caratterizza i dati da trattare.

È utile applicare i metodi non parametrici nella statistica chemiometrica quando si hanno campioni ridotti ed i metodi parametrici non sono applicabili.

Descriviamo qui solo due dei parametri su cui si basano i suddetti metodi.

6.1 Mediana

La mediana M rappresenta per i metodi non parametrici ciò che la media rappresenta per una distribuzione gaussiana. Per calcolare M si dispongono i dati in ordine crescente, poi:

1) se n è dispari $M = \frac{n+1}{2}$ esimo valore

2) se n è pari, M è la media tra l' $\frac{n}{2}$ esimo valore e l' $\left(\frac{n}{2} + 1\right)$ esimo valore

Esempio

Una titolazione ha dato i seguenti risultati: 25.01, 25.21, 25.04, 25.06 ml

Dati ordinati: 25.01, 25.04, 25.06, 25.21

$$M = \frac{n_2 + n_3}{2} = \frac{25.04 + 25.06}{2} = 25.05$$

Esempio

Una titolazione ha dato i seguenti risultati: 25.01, 25.21, 25.11, 25.04, 25.06 ml

Dati ordinati: 25.01, 25.04, 25.06, 25.11, 25.21

$$M = \frac{n_3}{2} = 25.06$$

6.2 Differenza interquartile

La differenza interquartile è per un metodo non parametrico ciò che la deviazione standard è per una distribuzione gaussiana.

La differenza interquartile è la differenza tra quartile superiore e quartile inferiore.

Il quartile superiore (inferiore) è la mediana della metà superiore (inferiore) dei dati, disposti in ordine crescente.

La metà superiore (inferiore) dei dati è l'insieme dei dati più grandi (piccoli) della mediana di tutti i dati.

6.3 Percentili

Data una serie di dati, disposti in ordine crescente, l' X esimo percentile è quel dato tale che l' $X\%$ degli altri dati ordinati è più piccolo.

Esempio: distribuzione dimensionale

In cromatografia si usano polveri a granulometria nota, per le quali è importante il diametro medio e la dispersione dei diametri. Di tali campioni vengono forniti d_{10} , d_{50} , d_{90} . Il d_{90} (novantesimo percentile) è quel valore tale che il 90% delle particelle ha diametro più basso.

7. APPENDICE: propagazione degli errori

$$q = x + \dots + z - u - \dots - w \Rightarrow$$

$$\Delta q = \Delta x + \dots + \Delta z + \dots + \Delta u + \dots + \Delta w \text{ (limite superiore dell'errore)}$$

$$\Delta q = \sqrt{(\Delta x)^2 + \dots + (\Delta z)^2 + \dots + (\Delta u)^2 + \dots + (\Delta w)^2} \text{ (errori indipendenti e casuali)}$$

$$q = \frac{x \cdot \dots \cdot z}{u \cdot \dots \cdot w} \Rightarrow$$

$$\frac{\Delta q}{|q|} = \frac{\Delta x}{|x|} + \dots + \frac{\Delta z}{|z|} + \dots + \frac{\Delta u}{|u|} + \dots + \frac{\Delta w}{|w|} \text{ (limite superiore dell'errore)}$$

$$\frac{\Delta q}{|q|} = \sqrt{\left(\frac{\Delta x}{x}\right)^2 + \dots + \left(\frac{\Delta z}{z}\right)^2 + \dots + \left(\frac{\Delta u}{u}\right)^2 + \dots + \left(\frac{\Delta w}{w}\right)^2} \text{ (errori indipendenti e casuali)}$$

$$q = B \cdot x, B = \text{costante priva di errore} \Rightarrow \Delta q = B \cdot \Delta x$$

$$q = q(x) \Rightarrow \Delta q = \left| \frac{dq}{dx} \right| \Delta x$$

$$q = x^n, \Rightarrow \frac{\Delta q}{q} = |n| \frac{\Delta x}{x}$$

INDICE

1.	INTRODUZIONE	1
1.1.	Problemi analitici quantitativi	1
1.2.	Errori nell'analisi quantitativa	1
1.3.	Tipi di errore	2
1.3.1.	errori grossolani	2
1.3.2.	errori casuali	2
1.3.3.	errori sistematici	2
2.	STATISTICA DELLE MISURE RIPETUTE	3
2.1	Media e deviazione standard	3
2.2	Distribuzione dei risultati di una misura	4
2.2.1.	Distribuzioni	5
2.2.2.	Proprietà della distribuzione gaussiana	5
2.2.3.	Distribuzione gaussiana: normalizzazione e probabilità	7
2.2.4.	Intervallo di confidenza: caso ideale	8
2.2.5.	Intervallo di confidenza: caso reale	9
2.3	Distribuzione delle medie	10
2.4	Intervallo di confidenza per la media	10
2.5	Presentazione dei risultati	11
3.	TESTS DI SIGNIFICATIVITÀ	12
3.1.	Definizione di test di significatività	12
3.2.	<i>t</i>-test per il confronto di una media con un valore noto	13
3.2.1.	Esempio di confronto di una media con un valore noto	14
3.3.	<i>t</i>-test per il confronto tra due medie	15
3.3.1.	Caso 1: s_1 e s_2 non sono significativamente diverse	16
3.3.2.	Caso 2: s_1 e s_2 sono significativamente diverse	17
3.4.	Paired <i>t</i>-test	18
3.5.	Test a una coda, test a due code	20
3.6.	<i>F</i>-test per il confronto tra deviazioni standard	22
3.6.1.	<i>F</i> -test a 1 coda	22
3.6.2.	<i>F</i> -test a 2 code	22
3.7.	<i>Q</i>-test per la verifica di dati anomali	24
3.8.	Test χ^2 per la verifica della normalità di una distribuzione	25
3.9.	Test della frequenza cumulativa	26
3.10.	Tipi di errore	27
4.	CONTROLLO DI QUALITÀ	28
4.1	Parametri di qualità nelle misure ripetute	28
4.1.1.	Accuratezza: esattezza e precisione	28
4.1.2.	Ripetibilità e riproducibilità	28
4.2	Carte di controllo	29
4.3	Carte <i>cusum</i>	29

5.	METODI QUANTITATIVI NELL'ANALISI STRUMENTALE	30
5.1	Metodo della curva di taratura.....	30
5.1.1	Procedimento:.....	30
5.1.2	La curva di taratura.....	31
5.1.3	Correlazione.....	31
5.1.4	t-test per la correlazione.....	31
5.1.5	Intervallo dinamico e intervallo lineare	32
5.1.6	Sensibilità	32
5.1.7	Applicabilità del metodo della curva di taratura	32
5.1.8	Vantaggi del metodo della curva di taratura	32
5.2	Metodo delle aggiunte standard.....	33
5.2.1.	Procedimento:.....	33
5.2.2.	La retta delle aggiunte standard	33
5.2.3.	Applicabilità del metodo delle aggiunte standard.....	34
5.2.4.	Vantaggi del metodo delle aggiunte standard	34
5.2.5.	Svantaggi del metodo delle aggiunte standard.....	34
5.2.6.	Verifica dell'effetto matrice.....	34
5.3	Metodo dello standard interno.....	35
5.3.1	Caso 1: uso dei fattori di risposta.....	35
5.3.2	Caso 2: uso di una retta di taratura	36
5.3.3	Applicabilità del metodo dello standard interno	37
5.3.4	Vantaggi del metodo dello standard interno	37
5.3.5	Quando usare i casi 1 e 2	37
5.4	Regressione lineare	38
5.4.1	Regressione su una retta	39
5.4.2	Interpolazione su una retta di taratura.....	40
5.4.3	Iperboli di confidenza (regressione non pesata)	41
5.4.4	Stima dell'errore nel metodo delle aggiunte standard.....	41
5.5	Regressione lineare pesata.....	42
5.5.1	Interpolazione su una retta di regressione pesata.....	43
5.6	Confronto tra metodi mediante regressione lineare	44
5.7	Rivelabilità e quantificabilità	45
5.7.1.	Convenzione sulla stima di LoD e LoQ.....	46
5.7.2.	Stima del LoD da misure ripetute sul bianco.....	47
5.7.3.	Stima del LoD dalla retta di taratura.....	47
5.7.4.	Stima del LoD dalla retta delle aggiunte standard	48
5.7.5.	Stima del LoD dal rapporto S/N	48
5.8	Un test di linearità.....	49
6.	METODI NON PARAMETRICI.....	50
6.1	Mediana	50
6.2	Differenza interquartile.....	51
6.3	Percentili	51
7.	APPENDICE: PROPAGAZIONE DEGLI ERRORI	52