

SMAC – 1: Statistical Methods in Analytical Chemistry vol. 1

©2003 – gian piero luciani

Email : gianpiero@mac.com homepage: <http://homepage.mac.com/gianpiero>

Premessa:

Questo set di programmi consente di eseguire i test statistici ed i calcoli necessari a valutare la precisione (ripetibilità stretta) di un metodo di analisi chimica (in particolare) e di una serie di misure (in generale).

Questo programma nasce dalla necessità di eseguire verifiche secondo le procedure previste dal Manuale Unichim 179/01 (edizione 2001): *Linee guida per la validazione di metodi analitici nei laboratori chimici* – “Valutazione della precisione (ripetibilità stretta) di un metodo analitico eseguito in unico laboratorio da un solo operatore su di un unico strumento in un breve intervallo di tempo.”

Sono stati aggiunti alcuni test proposti dal Professor Giovanni Mori – Dipartimento di Chimica Generale ed Inorganica, Chimica Analitica, Chimica Fisica - Università di Parma, illustrati nel corso “Qualità del dato analitico e validazione dei metodi di analisi chimica” organizzato da CIMACQ (Consorzio Interuniversitario di Ricerca Metodologie Analitiche e Controllo di Qualità) presso l’Università di Ferrara 15-16 aprile 2003.

Fonte di ispirazione ed elemento fondamentale nell’impostazione del programma è stato il volume monografico Peter C. Meier – Richard E. Zittel “Statistical Methods in Analytical Chemistry” editore Wiley Interscience Publication, da cui deriva il titolo di questo software.

Licenza:

Questo software è libero.

Chiunque sia interessato a questo software può utilizzarlo e modificarlo liberamente senza che questo comporti alcun impegno nei confronti dell’autore.

La distribuzione di questo software è consentita a condizione che vengano distribuiti tutti i file che compongono questo software senza alcuna modifica. E altresì è consentita la distribuzione di forme modificate – adattate di questo software alle condizioni che l’autore originale venga sempre citato e che non venga richiesto alcun compenso per il prodotto modificato.

Questo software non è fornito di alcuna garanzia. L’autore ha posto la massima attenzione nella verifica delle tabelle statistiche e del codice, tuttavia non è in grado di garantirne l’affidabilità assoluta. Per questo l’utilizzatore che adotti questo software lo fa a proprio rischio e pericolo e sarà unico a rispondere di eventuali danni o anomalie derivanti dall’uso di questo software.

Installazione:

Questo programma è stato sviluppato su di un calcolatore (calcolatrice grafica programmabile) Hewlett Packard 48GX, il programma dovrebbe funzionare correttamente anche sui modelli HP48G, HP48G+ e HP 49; il programma non dovrebbe funzionare sui calcolatori della serie precedente HP 48S e HP 48SX. Il programma è distribuito in forma di directory e in forma di library installabile in qualunque porta.

La prima operazione da fare sarebbe quella di verificare se i dati sono distribuiti normalmente, infatti tutte le successive elaborazioni partono dal presupposto che la serie dei dati ubbidisca ad una distribuzione normale (o gaussiana).

Le misure in campo chimico soddisfano la distribuzione normale praticamente sempre, se questo test non viene superato. Pertanto necessario verificare l'intera procedura analitica per comprenderne il motivo poiché molto probabile che l'anomalia sia dovuta a gravi errori sistematici.

L'algoritmo per il test di Shapiro-Wilk è seguente:

si mettono in ordine crescente le misure:

$$x_1 \leq x_2 \leq x_3 \leq \dots \leq x_n$$

Si calcola la somma dei quadrati:

$$SQ = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2$$

si calcola il parametro b:

dove i coefficienti a_i sono tabulati in funzione di n numero dei dati

$$b = \sum_{i=1}^n a_{n-i+1} (x_{n-i+1} - x_i)$$

Si calcola il parametro W:

$$W = \frac{b^2}{SQ}$$

Si trasforma il parametro W in k_p :

Dove γ , η e ε sono tabulati in funzione di n numero dei dati.

$$k_p = \gamma + \eta \cdot \ln \frac{W - \varepsilon}{1 - W}$$

Il valore k_p si confronta con il valore tabulato per $\alpha=0.05$ e per $\alpha=0.01$, quando si verifica la condizione $k_p < k_{p,\alpha=0.05}$ allora si può ritenere che i dati ottenuti seguono una distribuzione normale o di Gauss.

Importante:

Questo test si applica per serie di dati compresi tra 3 e 50 misure.

$$k_{p,\alpha=0.05} = 1,960$$

$$k_{p,\alpha=0.01} = 2,576$$

Il programma richiede di inserire i dati sotto forma di matrice, ogni serie di dati dovrà comporre una colonna della matrice, il test viene quindi eseguito per ogni colonna della matrice.

Come output il programma fornisce una lista composta da “-” quando il test viene superato (distribuzione normale), da “*” quando il test non viene superato per $\alpha=0.01$ ma viene superato per $\alpha=0.05$ e da “**” quando il test non viene superato per $\alpha=0.05$ (e ovviamente $\alpha=0.01$).

```
1:SWtest: { "- " "- " "- " "*" "**" "**" }
```

Verificato che i dati soddisfino la condizione di “normalità” si prosegue cercando i dati “anomali”.

I dati “anomali” sono dati significativamente diversi dal valore medio, talmente diversi da poter essere considerati frutto di errori grossolani; è opportuno identificarli in modo da poter valutare criticamente (su base statistica e non arbitrariamente) la possibilità di scartarli per le successive elaborazioni (calcolo del valor medio, dell'intervallo di confidenza, etc...).

I test considerati sono 4, sarebbero da applicare tutti e 4 e ciascuno fornisce dati indipendenti, non tutti sono egualmente potenti.

Ogni test segnalerà nessuno, uno, due o più dati con uno o due asterischi.
I dati contrassegnati da 2 asterischi ** sono anomali con probabilità 99%
I dati contrassegnati da 1 asterisco * sono anomali con probabilità 95%
Gli altri dati non sono anomali.

DXtest

Test di anomalia di Dixon

TEST DI ANOMALIA DI DIXON

ΣDAT [[.7 1.5] [....

MATRICE STATISTICA DEI DATI

Il test di Dixon serve ad individuare 1 solo dato anomalo che sarà il più grande o il più piccolo della serie.

L'algoritmo per il test di Dixon \square seguente:

si mettono in ordine crescente le misure: $x_1 \leq x_2 \leq x_3 \leq \dots \leq x_n$

si calcolano in funzione di n numero dei dati i seguenti rapporti:

per $3 \leq n \leq 7$

$$R_{10} = \frac{x_2 - x_1}{x_n - x_1}$$

$$R_{10} = \frac{x_n - x_{n-1}}{x_n - x_1}$$

per $8 \leq n \leq 12$

$$R_{11} = \frac{x_2 - x_1}{x_{n-1} - x_1}$$

$$R_{11} = \frac{x_n - x_{n-1}}{x_n - x_2}$$

per $12 \leq n \leq 40$

$$R_{22} = \frac{x_3 - x_1}{x_{n-2} - x_1}$$

$$R_{22} = \frac{x_n - x_{n-2}}{x_n - x_3}$$

Il valore maggiore dei due R calcolati si confronta con un valore di R tabulato ai livelli di confidenza $\alpha=0.05$ e $\alpha=0.01$, si hanno le 3 possibilità

$$R_{\text{calcolato}} < R_{\alpha=0.05}$$

Il dato non **an**omalo

$$R_{\alpha=0.05} < R_{\text{calcolato}} < R_{\alpha=0.01}$$

Il dato **anomalo** 1 asterisco

*

$$R_{\text{calcolato}} > R_{\alpha=0.01}$$

Il dato **anomalo** 2 asterischi

Il programma richiede di inserire i dati sotto forma di matrice, ogni serie di dati dovrà comporre una colonna della matrice, il test viene quindi eseguito per ogni colonna della matrice.

Quindi viene prodotto il seguente output:

```
1:DXtest: { "-" "7,7**" }
```

G1test

Test di Grubbs per un dato anomalo.

```

TEST DI ANOMALIA GRUBBS 1 DATO
ΣDAT [[ .7 1.5 ] [ ....

```

MATRICE STATISTICA DEI DATI

L'algoritmo per il test di Grubbs \square seguente:

Si calcola il valore medio \bar{x} e la deviazione standard s con le seguenti formule:

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

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

$$G_1 = \frac{x_1 - \bar{x}}{s}$$

$$G_n = \frac{x_n - \bar{x}}{s}$$

$G_{\text{calcolato}} < G_{\alpha=0.05}$	Il dato non è anomalo	
$G_{\alpha=0.05} < G_{\text{calcolato}} < G_{\alpha=0.01}$	Il dato è anomalo 1 asterisco	*
$G_{\text{calcolato}} > G_{\alpha=0.01}$	Il dato è anomalo 2 asterischi	**

Quindi viene prodotto il seguente output:

G2test

Test di Grubbs per 2 dati anomali

```
TEST DI ANOMALIA GRUBBS 2 DATI
ΣDAT [[.7 1.5] [.000]]
```

MATRICE STATISTICA DEI DATI

Questo test è particolarmente utile ad individuare i dati anomali in quei casi in cui due valori troppo grandi o troppo piccoli esercitano un effetto di “mascheramento” reciproco impedendo ai test precedenti di essere individuati.

si mettono in ordine crescente le misure: $X_1 \leq X_2 \leq X_3 \leq \dots \leq X_n$

$$G = \frac{s_{n-1,n}^2}{s_0^2}$$
$$\begin{aligned} s_0^2 &= \sum_{i=1}^n (x_i - \bar{x})^2 \\ s_{n-1,n}^2 &= \sum_{i=1}^{n-2} (x_i - \bar{x}_{n-1,n})^2 \\ \bar{x}_{n-1,n} &= \frac{1}{n-2} \sum_{i=1}^{n-2} x_i \end{aligned}$$
$$G = \frac{s_{1,2}^2}{s_0^2}$$
$$\begin{aligned}s_0^2 &= \sum_{i=1}^n (x_i - \bar{x})^2 \\ s_{1,2}^2 &= \sum_{i=3}^n (x_i - \bar{x}_{1,2})^2 \\ \bar{x}_{1,2} &= \frac{1}{n-2} \sum_{i=3}^n x_i\end{aligned}$$

$G_{\text{calcolato}} > G_{\alpha=0.05}$	Il dato non è anomalo	
$G_{\alpha=0.05} > G_{\text{calcolato}} > G_{\alpha=0.01}$	Il dato è anomalo 1 asterisco	*
$G_{\text{calcolato}} < G_{\alpha=0.01}$	Il dato è anomalo 2 asterischi	**

Quindi viene prodotto il seguente output:

Test di anomalia di Huber

MATRICE STATISTICA DEI DATI

Il test di Huber, al contrario dei test precedenti, serve ad individuare un numero di dati anomali fino al 50% dei dati totali della serie di misure; il buon senso suggerisce di controllare il metodo quando i dati da scartare sono molti.

L'algoritmo per il test di Huber \square seguente:

si mettono in ordine crescente le misure: $x_1 \leq x_2 \leq x_3 \leq \dots \leq x_n$
 si individua la mediana x_{med}
 si calcolano i residui r_i $r_i = x_i - x_{med}$
 si mettono in ordine crescente i valori $|r_1| < |r_2| < \dots < |r_n|$
 assoluti di r_i
 si individua la mediana dei residui r_{med}
 Si confrontano i residui con il fatto $3,8r_{med}$ e con $4,5r_{med}$

Si esegue il test

$ x_i - x_{med} = r_i < 3,8 \cdot r_{med}$	Il dato non \square anomalo	
$3,8 \cdot r_{med} < r_i < 4,5 \cdot r_{med}$	Il dato \square anomalo 1 asterisco	*
$ x_i - x_{med} = r_i > 4,5 \cdot r_{med}$	Il dato \square anomalo 2 asterischi	**

Il programma richiede di inserire i dati sotto forma di matrice, ogni serie di dati dovr \square comporre una colonna della matrice, il test viene quindi eseguito per ogni colonna della matrice.

Quindi viene prodotto il seguente output:

1:Htest: { ",64* ,63** 7,7**" "1,5** 2** 2**" }

Eseguiti questi test e valutati i risultati, qualora si siano scartati dei dati considerati anomali sarebbe necessario procedere nuovamente con un test di normalità si Shapiro-Wilk.

Alla fine si possono eseguire i calcoli statistici ad "un livello".

CMONO — VMONO

Calcolo e visualizzazione dei risultati.

```

VALORI COLONNA N:1
: x: .733333333333
: med: .74
: range: .17
: Var: .00296969697
: s-dev: .5.449492609... ↓
< > [FNC] [L] [OK]

```

Il programma CMONO esegue i calcoli e li memorizza nella variabile statistica Σ SIGMA, il programma VMONO si occupa della loro visualizzazione. Si descrivono di seguito i calcoli eseguiti:

Simbolo	Descrizione	Formula
\bar{x}	Valore medio	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
med	Mediana	si pone: $x_1 < x_2 < x_3 < \dots < x_n$ per n dispari $med = x_{(n+1)/2}$

range	Dispersione dei dati	per n pari $med = (x_{n/2} + x_{n/2+1})/2$
s-d	Deviazione standard	$s = \sqrt{\frac{x_{\max} - x_{\min}}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$
CV%	Coefficiente di variazione %	$CV\% = \frac{s}{\bar{x}} \cdot 100$
$s - \bar{x}$	Deviazione standard della media	$s - \bar{x} = \frac{s}{\sqrt{n}}$
If	Intervallo di fiducia	$\square t_{p=1-\alpha/2} \cdot s$
If \bar{x}	Intervallo di fiducia della media	$\square t_{p=1-\alpha/2} \cdot \frac{s}{\sqrt{n}}$

Il programma richiede di inserire i dati sotto forma di matrice, ogni serie di dati dovrà comporre una colonna della matrice, il programma richiede di inserire anche il livello di probabilità desiderato (0.95 per $p = 95\%$ per $\alpha = 0.05$).

Quindi esegue i calcoli e per la loro visualizzazione richiama il programma VMONO. Quest'ultimo consente di scorrere i dati di ogni singola colonna con i tasti frecce \uparrow e \downarrow e di passare da una colonna e l'altra con i tasti programmati $>$ e $<$. Premendo "OK" o "ENTER" una copia del valore evidenziato viene riportata nello STACK dell'hp, premendo "CANCL" si esce dal programma.

Con questo si esaurisce lo studio del caso ad un unico livello; il tipico esempio di calcolo della esattezza e della precisione di un metodo di analisi eseguendo un certo numero di misure su di una miscela (standard) certificata. Seguendo i passaggi sino a questo punto si ha grado di verificare se il metodo è esatto e qual è la sua precisione. Questi test sono utili anche per verificare se la precisione del metodo al livello di concentrazione analizzato è conforme ai limiti (di legge, richiesti dagli standard interni, etc.) desiderati.

Esiste per anche il caso in cui è necessario verificare se la deviazione standard è meno costante all'interno dell'intero intervallo di misura (omoschedasticità) qualora la deviazione standard dipendesse dalla concentrazione sarebbe utile calcolare una espressione per esprimerla in funzione della concentrazione.

Per fare questo è necessario applicare i test che saranno argomento del programma "SMAC -2" in fase di realizzazione.

gian piero luciani
Comacchio 4 giugno 2003