

# LIMITI ED ABUSI NELL'UTILIZZO DELLE TECNICHE DI REGRESSIONE IN GEOLOGIA APPLICATA ED IN GEOTECNICA

C. CHERUBINI - CONCETTA IMMACOLATA GIASI (\*)

**G**ia in un articolo del 1985 Nova evidenziava alcuni inconvenienti cui si poteva andare incontro con un uso non accorto delle correlazioni scaturenti da elaborazioni statistiche di dati geotecnici, svolte in genere mediante metodologie ai minimi quadrati. Recentemente Li e White (1993), hanno puntualizzato oltre agli inconvenienti anche i presupposti teorici per l'uso dei modelli statistici di regressione.

Questo articolo in sostanza desidera mettere in evidenza quelle che sono le problematiche insite nell'uso dei modelli di regressione, proponendo altresì una scelta fra metodologie diverse, anche con l'ausilio di esempi esplicativi.

## GENERALITÀ SUI MODELLI DI REGRESSIONE

I classici modelli di regressione sono trattati su numerosi testi: si vedano ad es. Gennaro (1975), Afifi e Azen (1979), Chapra e Canale (1988); in particolare il testo di Draper e Smith (1981) è totalmente dedicato alle tecniche di regressione.

Implementazioni di codici atti a calcolare le rette (o, anche curve) di regressione ai minimi quadrati si trovano non solo su personal ma anche su pocket computer. Trattando per semplicità il solo modello di regressione lineare, si può scrivere:

$$Y = a + bX + \delta \quad (1)$$

dove X e Y sono rispettivamente la variabile indipen-

### ABSTRACT

Regression analysis and curve fitting are techniques commonly used in engineering.

In the field of soil mechanics particular problems occur in the application of these techniques, often due to lack of recognition of the assumptions implicit in the method.

To avoid the improper use of these techniques, the paper highlights some of the common mistakes in regression analysis and in particular:

- Problems related to the choice of variables involved in procedures;
- Heteroscedasticity (not constant variance);
- Autocorrelation;
- Extrapolation;
- Algebraic manipulation.

Alternative methods different from ordinary least squares (OLS) like Least Normal Squares, Reduced Major Axis Methods, etc. are also described and used to evidence the difference among them that are, in general, noticeable.

### RÉSUMÉ

*L'analyse de la regression et l'interpolation sont des techniques communément utilisées dans l'art de l'ingénieur.*

*Dans le domaine de la géotechnique l'application de ces techniques peut provoquer quelques inconvénients, connexes à l'absence du respect des hypothèses implicite dans la méthode.*

*Pour éviter l'impropre utilisation de ces techniques, cet article met en évidence quelques fautes de l'analyse de regression et en particulier:*

- *Problèmes connexes au choix des variables considérées dans les procédures de calcul;*
- *Problèmes connexes à la variabilité de la variance;*
- *Autocorrélation;*
- *Extrapolation;*
- *Manipulation algébrique;*
- *Sont pareillement décrites des méthodes différentes des minimales corréllés (LNS, RMA, etc...). Quelques exemples démontrent la différence des résultats obtenus.*

dente e dipendente, a e b sono costanti e  $\delta$  è la componente casuale.

Il modello di regressione è basato sulle seguenti assunzioni di base:

1. X e Y non hanno errori di misura;
2. X e Y non sono funzionalmente dipendenti;
3. esiste una relazione lineare fra X e Y;
4. la componente aleatoria ( $\delta$ ) ha varianza costante (omoscedasticità);
5. la variabile d non è autocorrelata.

Possiamo ottenere i coefficienti a e b della retta attraverso il metodo dei minimi quadrati (Ordinary Least Squares, OLS), minimizzando la somma dei quadrati degli scostamenti dei punti reali rispetto a quelli calcolati attraverso l'equazione della retta.

Altri tipi di minimizzazione sono possibili ma possono non fornire risultati soddisfacenti in determinati casi [Chapra e Canale, 1988]. La

stima ai minimi quadrati, se le precedenti ipotesi sono rispettate, è la "migliore stima lineare non distorta" (BLUE, Best Linear Unbiased Estimate).

Se, invece, tali ipotesi non sono totalmente, o anche in parte rispettate, il metodo dei minimi quadrati risulta non valido e inefficiente. Ciò significa che le conclusioni tratte su base statistica non sono correttamente dedotte e che le previsioni svolte sono meno affidabili di quanto ritenuto.

Altri metodi sono disponibili per stimare i coefficienti a e b della retta (e quindi d) [Troutman e Williams, 1987; Morris et al., 1987; Tukey, 1977; Velleman e Hoaglin, 1981].

Di essi parleremo dopo aver svolto alcune considerazioni e discusso alcuni problemi strettamente legati alle ipotesi di cui sopra.

(\*) Istituto di Geologia Applicata e Geotecnica - Facoltà di Ingegneria, Politecnico di Bari.

### CONDIZIONI DA RISPETTARE PER L'USO DEI MODELLI DI REGRESSIONE IN CAMPO GEOTECNICO

- Le proprietà del terreno da usare nella equazione di regressione devono essere delle proprietà fondamentali del terreno. Per esempio come riportato da Li e White (op.cit.) se desideriamo correlare  $C_c$  con il contenuto d'acqua e l'indice dei vuoti dobbiamo convenire che  $w$  e  $e_0$  non sono proprietà fondamentali [Li 1990, 1991] perché dipendono dalla tensione geostatica. Conseguentemente viene introdotta una variabilità aggiuntiva a causa delle differenti profondità di campionamento. Inoltre, la correlazione può risultare distorta (biased) se esistono grandi quantità di dati provenienti da un particolare intervallo di profondità.

- Nel caso in cui si ipotizzano due o più variabili indipendenti in un modello di regressione ( $X_1$ ,  $X_2$  e così via) queste variabili dovrebbero risultare funzionalmente indipendenti. Problemi di tal genere, che rendono la stima OLS inefficiente vanno sotto il nome di "multicollinearità". Lee (1992) ad esempio ha discusso una serie di equazioni per correlare la velocità delle onde trasversali col valore NSPT e con la profondità  $z$ . Poiché NSPT è legato alla profondità, le due variabili "indipendenti" NSPT e  $z$  sono implicitamente funzionalmente dipendenti. Date queste condizioni si pone il problema della multicollinearità.

- La variabile dipendente  $Y$  e quella indipendente  $X$  non devono essere funzionalmente dipendenti.

- Uno degli esempi più classici riguarda la presunta correlazione fra il limite di liquidità  $LL$  e l'indice di plasticità  $IP$ .

- Poiché  $IP = LL - LP$  di fatto esiste una relazione di dipendenza lineare fra  $LL$  ed  $IP$ .

Esempi di studi di relazioni siffatte si trovano in lavori di numerosi autori. Val la pena ricordare solo che Nagaraj e Jayadeva (1983) hanno riportato che esiste una unica relazione lineare  $PI = 0,74 \cdot (LL - 8)$  che lega il limite di liquidità e l'indice di plasticità, relazione caratterizzata da un elevato coefficiente di correlazione (0,957 su 520 rilevazioni relative a diverse argille affioranti in diverse parti del globo).

Li (1991) ha dimostrato che, generando  $LL$  casualmente fra i valori limite  $30 \div 120\%$  mentre il rapporto  $PI/LL$  era anch'esso generato casualmente nell'intervallo  $0.3 \div 0.3$  (valori tutti tipici per le argille inorganiche) si ottiene una retta molto simile a quella di Nagaraj e Jayadeva (op. cit.) con analogo coefficiente di correlazione (0,9627). Windisch e Yong (1990) riportano alcune correlazioni fra  $IP$  ed  $LL$  valide per argille canadesi caratterizzate da coefficienti di correlazione variabili fra 0,93 e 0,99 asserendo però che "Strong correlations between  $PI$  and  $LL$  are not surprising, since the dominant quantity in  $PI$  is precisely  $LL$  ( $PI = LL - LP$ ) with the plastic limit showing relatively small variations".

Un'altra correlazione comunemente usata in geotecnica e che coinvolge due variabili funzionalmente collegate è la correlazione

$$\frac{C_c}{1 + e_0}, w$$

(contenuto naturale d'acqua) [Abdrabbo e Mahmoud, 1990; Grace e Henry, 1957; Koutsoftas, 1987; Rendon e Herrero, 1980]. Ovviamente per un'argilla satura,  $e_0$  e  $w$  sono funzionalmente relazionate tramite la gravità specifica. Un problema analogo si ha quando le quantità  $X$  e  $Y$  sono entrambe dipendenti da una terza variabile  $Z$  anche se  $X$  e  $Y$  non sono fisicamente relazionate.

Questo caso si verifica molto comunemente in geotecnica perché molte proprietà dipendono dalla tensione geostatica.

- Per quanto riguarda l'omoscedasticità, condizione espressamente richiamata al paragrafo precedente, va detto che molte proprietà dei terreni mostrano eteroscedasticità con la profondità (ad esempio media crescente con la profondità con coefficiente di variazione approssimativamente costante). Un esempio di ciò è mostrato in **figura 1** che riporta la resistenza a taglio non drenata delle argille marine di Hong Kong [Lumb 1967]. Come detto, se non viene rispettata l'omoscedasticità, la stima con l'OLS non è efficiente. La

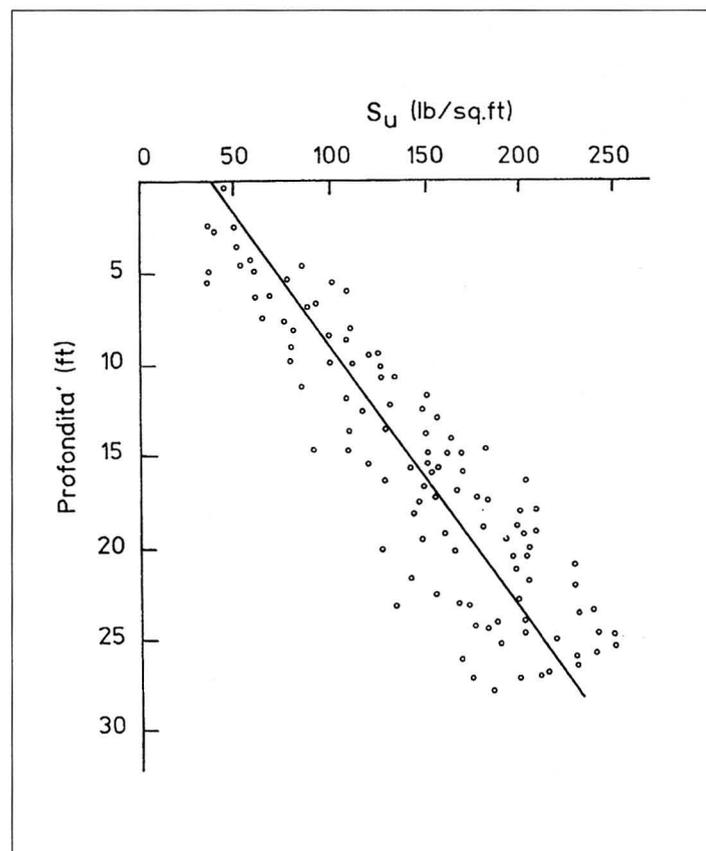


Figura 1 - Variazione con la profondità della resistenza a taglio non drenata delle Argille Marine di Hong Kong.

condizione suddetta è verificabile in prima approssimazione a occhio; all'uopo esistono comunque dei tests statistici specifici [Ramanathan, 1989; Pindyck e Rubinfeld, 1991].

- Alcuni studi indicano che le proprietà dei terreni tendono ad essere positivamente autocorrelate [Li e White, 1987; Jaska et al. 1992]. Cioè misure di caratteristiche geotecniche di campioni di terreno prelevati a distanza ravvicinata tendono ad essere maggiormente correlate rispetto a quelle di campioni prelevati a distanza maggiore. La presenza di autocorrelazione per la componente  $\delta$  viola una delle asserzioni fondamentali del modello di regressione lineare. Se non viene presa in considerazione l'autocorrelazione la procedura OLS non risulta BLUE. La varianza della componente casuale risulterà sottostimata ed il coefficiente di correlazione dell'equazione di regressione sarà sovrastimato, portando all'impressione errata che la proprietà del terreno è meno variabile e che la pendenza interpolata è più "perfetta" del reale. La presenza di autocorrelazione può essere rivelata dalla semplice osservazione del diagramma dei residui oppure da tests statistici come il test di Durbin-Watson [Ramanathan, 1989].

- Avendo stabilito una correlazione empirica su base statistica fra due proprietà dei terreni, si tende spesso erroneamente a svolgere previsioni al di là dell'intervallo dei dati di origine. Gli errori cui si può andare incontro possono essere connessi alla circostanza che una data legge può essere valida in un dato intervallo anche per ragioni di semplicità. In un intervallo più ampio il trend reale può essere diverso (si veda la **figura 2**). Inoltre le bande di confidenza tendono ad espandersi notevolmente quando si va ai limiti dell'intervallo dei dati originali, per cui le previsioni basate su estrapolazioni possono risultare inaffidabili.

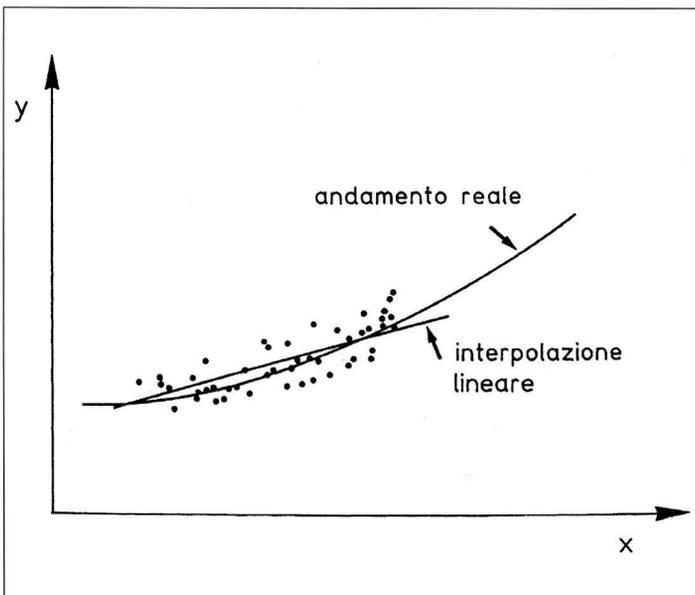


Figura 2 - Possibile errore cui si va incontro con l'estrapolazione del trend calcolato in un ben definito intervallo.

- Le equazioni di regressione vengono spesso manipolate algebricamente [Carrier e Beckman 1984; Williams e Troutman 1987]. La classificazione delle manipolazioni algebriche può essere così svolta:

- Riordinamento:

$$\text{L'equazione } Y = a + b \cdot X \text{ diviene } X = \frac{(Y - a)}{b}$$

- Eguaglianza di due equazioni:

$$\text{Se } Y = a + b \cdot X = c + d \cdot Z \text{ eliminando la variabile } Y \text{ si ha}$$

$$Z = \frac{[(a - c) + b \cdot X]}{d}$$

- Sostituzione:

Date due equazioni  $Y = a + b \cdot X$  e  $X = c + d \cdot Z$ , si crea una nuova equazione sostituendo  $X$  e cioè  $Y = (a + b \cdot c + b \cdot d \cdot Z)$

- Addizione - Sottrazione:

Se  $Y = a + b \cdot X$  e  $X = c + d \cdot Z$ , la nuova equazione che si ottiene è  $Y \pm Z = (a \pm c) + (b \pm d) \cdot X$ .

Molte delle manipolazioni algebriche descritte non producono stime corrette di alcune tecniche di regressione come indicato nella **tabella I** [Williams e Troutman 1987].

A titolo di esempio si discutono brevemente i risultati di Abdrabbo e Mahmoud (1990) che presentano una serie di correlazioni empiriche fra alcune proprietà di argille egiziane. Nella fattispecie l'indice di compressione viene correlato al contenuto naturale d'acqua  $w$  ed all'indice dei vuoti iniziale  $e_0$  secondo le seguenti espressioni:

$$C_c = 0,0066 \cdot w \quad (2)$$

$$C_c = 0,42 \cdot (e_0 - 0,5) \quad (3)$$

Eliminando dalle precedenti equazioni  $C_c$  si ottiene una relazione fra l'indice dei vuoti ed il contenuto naturale d'acqua.

$$e_0 = 0,01571 \cdot w + 0,5 \quad (4)$$

Tale legge, confrontata con l'insieme dei dati di partenza, non fornisce una valida interpretazione degli stessi, come si può ben vedere dall'esame della **figura 3**.

Lo stesso Nova (1985) evidenziava inconvenienti analoghi studiando le differenti relazioni che legano il coefficiente di spinta a riposo  $K_0$  (per terreni normal-

**Tabella I Manipolazioni algebriche consentite e non consentite per alcune tecniche di regressione.**

Tipo di manipolazione	OLS	LNS	RMA
Riordinamento	No	Si	Si (*)
Eguaglianza	No	No	Si (*)
Sostituzione	No	No	Si (*)
Addizione-Sottrazione	Si	No	No

(\*) Le pendenze calcolate possono avere segni errati. Ciò non è un problema se la correlazione è elevata.

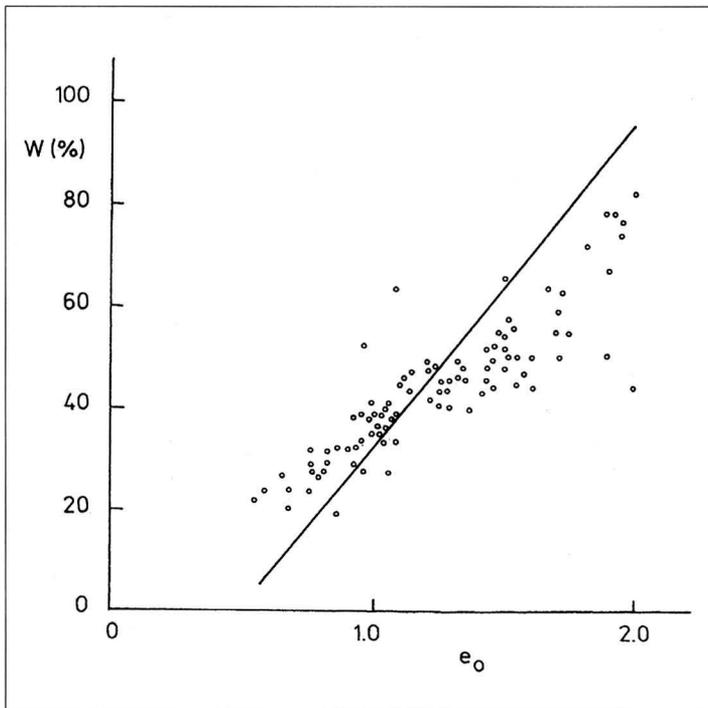


Figura 3 - Relazione lineare contenuto-naturale d'acqua, indice dei vuoti in situ ricavata attraverso correlazioni indirette con l'indice di compressione: tale relazione non è in grado di interpolare i dati sperimentali.

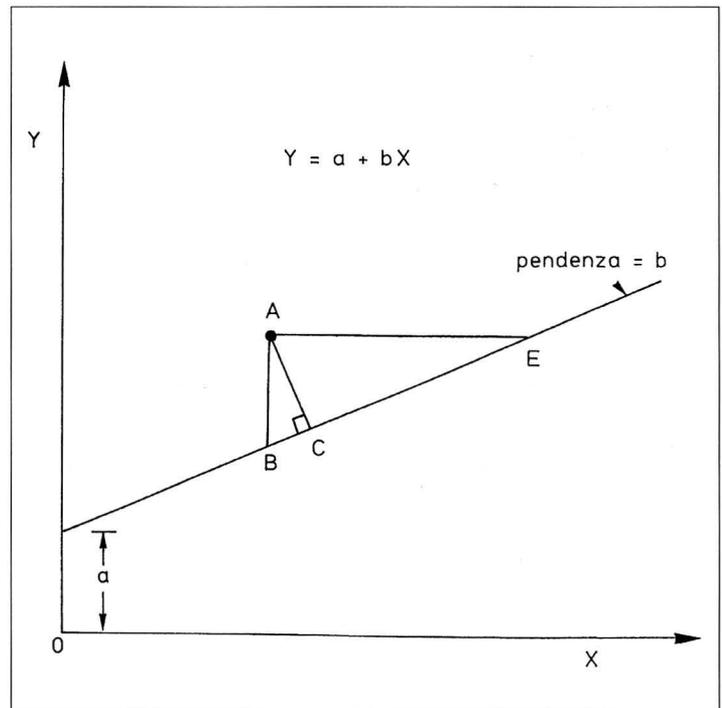


Figura 4 - Diversi metodi per minimizzare le distanze dei punti dalla retta interpolante.

consolidati) all'angolo di resistenza al taglio ed all'indice di plasticità.

#### MODELLI DI REGRESSIONE

Di seguito vengono esplicitate alcune metodologie per "adattare" una retta ad una serie di punti in un piano X-Y.

Va puntualizzato subito che quelle discusse di seguito sono solo alcune delle tecniche disponibili. Per una trattazione più ampia si rimanda ai testi riportati in bibliografia, nonché al lavoro di Rizkallah e Maschwitz (1983) per alcune tecniche di ottimizzazione che esulano però dagli scopi di questo articolo.

- Ols

Questo metodo, frequentemente usato, perviene alla determinazione della retta  $Y = a + b \cdot X$  per la quale la somma degli scostamenti al quadrato fra i punti considerati e quelli della retta, è minima. Se  $Y$  è la variabile da prevedere, le deviazioni sono assunte in direzione verticale, come la linea AB in **figura 4**. Se invece è  $X$  la variabile da prevedere e quindi la regressione è di  $X$  su  $Y$ , la deviazione appropriata dovrà essere rappresentata dalla distanza orizzontale AE.

C'è da notare che la minimizzazione della somma dei quadrati delle deviazioni orizzontali fornisce usualmente una retta diversa da quella ottenuta minimizzan-

do la somma dei quadrati delle deviazioni verticali. Ulteriori dettagli sul metodo si trovano in un qualsiasi testo di statistica per cui non è il caso di soffermarci a lungo.

Conviene richiamare, comunque, alcuni indici statistici di base utilizzati per calcolare i coefficienti  $a$  e  $b$  della retta. Essi sono:

$$S_x = \left( \frac{1}{N} \cdot \sum (X_i - \bar{X})^2 \right)^{\frac{1}{2}} \quad (5)$$

$$S_y = \left( \frac{1}{N} \cdot \sum (Y_i - \bar{Y})^2 \right)^{\frac{1}{2}} \quad (6)$$

$$S_{xy} = \frac{1}{N} \cdot \sum (X_i - \bar{X}) \cdot (Y_i - \bar{Y}) \quad (7)$$

$$r = \frac{S_{xy}}{S_x \cdot S_y} \quad (8)$$

dove:

$N$  è il numero dei punti considerati ( $X_i, Y_i$ );

$\sum$  indica ovviamente la sommatoria;

$\bar{X}$  è la media ed  $S_x$  è lo scarto quadratico medio dell'insieme dei valori  $X_i$ ;

$\bar{Y}$  è la media e  $S_y$  è lo scarto quadratico medio dell'insieme dei valori  $Y_i$ ;  $S_{xy}$  è la covarianza;

$r$  è il coefficiente di correlazione.

Le rette che usano le statistiche da (5) a (8) passano tutte attraverso il "baricentro" dei dati  $\bar{X}, \bar{Y}$ .

Quindi, in generale, si può valutare  $a$  imponendo il passaggio per il punto  $(\bar{X}, \bar{Y})$  e cioè:

$$a = \bar{Y} - b \cdot \bar{X} \quad (9)$$

dove  $Y$  è la variabile dipendente.

Per il metodo dei minimi quadrati assumendo  $Y$  come variabile dipendente, il valore di  $b$  risulta pari a:

$$b = r \cdot \left( \frac{S_y}{S_x} \right) \quad (10)$$

La scelta della variabile dipendente va effettuata in maniera opportuna in quanto ha, come già visto, influenza sui termini noti della equazione della retta che si ottiene.

- Averages (Medie)

Questo metodo consiste nel dividere i dati in due gruppi uguali o pressappoco uguali, dopo aver riportato gli stessi in un diagramma  $X$ - $Y$ . La scelta dei due gruppi viene svolta sulla base dei valori di  $X$ . Per il primo gruppo di dati caratterizzati da valori di  $X$  bassi si calcolano separatamente le relative medie di  $\bar{X}$  e  $\bar{Y}$ ; analogamente si opera sui dati caratterizzati da valori di  $X$  elevati.

Unendo i due punti medi precedentemente calcolati si ottiene ovviamente una linea retta; il calcolo di  $a$  e  $b$  è svolto secondo i metodi della geometria classica.

Una variante di questa tecnica prevede la suddivisione in più gruppi con l'individuazione di un numero corrispondente di valori medi ( $\bar{X}$ ,  $\bar{Y}$ ) per ogni gruppo e il tracciamento ad occhio di una linea retta che passi il più possibile nei pressi degli stessi.

- LNS (Least Normal Squares)

Minimizzando la sommatoria delle distanze al quadrato perpendicolari fra la linea e i punti (linea  $AC$  nella **figura 4**) si ottiene la linea dei minimi quadrati normali.

Per  $r + 0$  la pendenza  $b$  di questa retta vale:

$$b = \frac{(-A + (r^2 + A^2)^{\frac{1}{2}})}{r} \quad (11)$$

dove:

$$A = \frac{1}{2} \cdot \left( \frac{S_x}{S_y} - \frac{S_y}{S_x} \right) \quad (12)$$

e l'intercetta  $a$  si ha attraverso l'equazione 9.

- RMA (Reduced Major Axis)

Questo metodo minimizza la somma delle aree dei triangoli rettangoli formati dai punti e dalla linea retta che meglio approssima i dati, con i cateti paralleli agli assi  $X$  e  $Y$  (BAE in **figura 4**). In tal senso i metodi OLS e LNS possono essere interpretati come la minimizzazione della somma delle aree dei quadrati i cui lati rappresentano la deviazione del punto dalla retta. L'area trian-

golare tra ogni punto e la retta è:

$$\frac{1}{2} \cdot (base \cdot altezza) =$$

$$\frac{1}{2} \cdot (dev. verticale \cdot dev. orizzontale)$$

Nella **figura 4** si ha:

$$\frac{1}{2} \cdot (AB \cdot AE) \quad (13)$$

la pendenza  $b$  della linea  $BCE$  è semplicemente

$$b = \pm \left( \frac{S_y}{S_x} \right) \quad (14)$$

in cui  $b$  fornisce il segno del coefficiente di correlazione.

Questo metodo, analogamente all'LNS consente di ottenere una legge unica fra le due variabili non essendo necessario definire quale fra le due ( $X$  o  $Y$ ) sia da considerare variabile dipendente e quale variabile indipendente.

- Resistant line

Questo metodo fa parte dei metodi E.D.A. [Tukey, 1977; Velleman e Hoaglin, 1981].

Dopo aver riportato i dati su un diagramma  $X$ - $Y$  (**figura 5**), dividiamo gli stessi in tre porzioni, una con i valori più bassi di  $X$ , una con i valori medi di  $X$ , un'altra con i valori più elevati, approssimativamente attribuendo ad ogni regione un terzo del numero dei punti. Se  $(N/3)$  non è un numero intero, conviene far sì che il punto o i punti rimanenti siano attribuiti simmetricamente. In ogni porzione dei dati si trattano le  $X$  e  $Y$  separatamente.

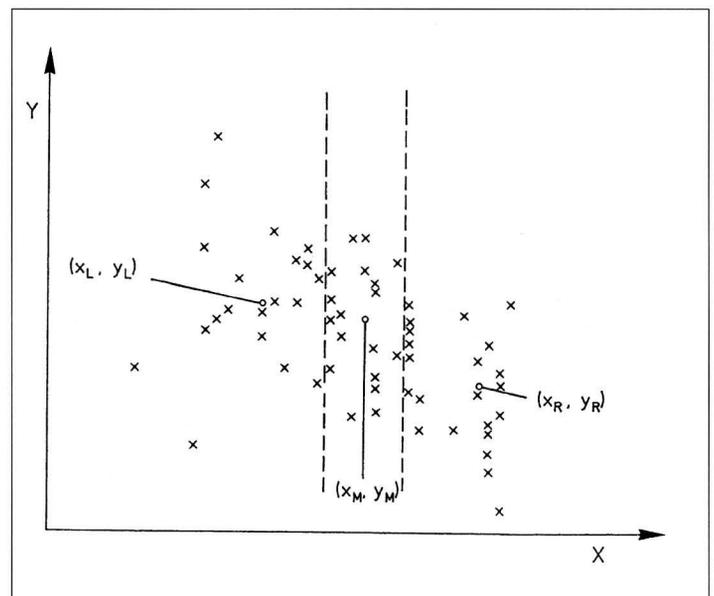


Figura 5 - Partizione dell'insieme dei punti al fine di individuare l'equazione della "resistant line".

Troviamo quindi le tre mediane di  $X$  e  $Y$  separatamente. I tre punti che individuano nel piano le tre mediane di  $X$  e le tre di  $Y$  non necessariamente quindi sono punti appartenenti all'insieme dei punti considerati.

Poiché tali punti sono scelti dalla metà di ogni terzo di punti, essi sintetizzano l'insieme dei punti in ogni regione. Conseguentemente, essi sono chiamati "summary points" (punti di compendio). Essi possono essere indicati col suffisso  $L$  (left)  $M$  (middle) e  $R$  (right) e cioè:

$$(X_L, Y_L), (X_M, Y_M), (X_R, Y_R)$$

Una volta trovati questi punti, possiamo facilmente calcolare i valori di  $a$  e  $b$ . Per la pendenza  $b$ , troviamo il valore attraverso la relazione

$$b = \frac{(Y_R - Y_L)}{(X_R - X_L)} \quad (15)$$

L'intercetta  $a$  deve essere "aggiustata" in modo che la retta passi il più possibile vicino al centro dei dati. È possibile farla passare per il punto medio (summary) in modo che:

$$a = Y_M - b \cdot X_M \quad (16)$$

Comunque, piuttosto che utilizzare solo il punto medio per determinare l'intercetta, è possibile utilizzare tutti e tre i punti "summary" e fare di essi la media e cioè calcolare:

$$a_L = Y_L - b \cdot X_L \quad (17)$$

$$a_M = Y_M - b \cdot X_M \quad (18)$$

$$a_R = Y_R - b \cdot X_R \quad (19)$$

e conseguentemente:

$$a = \frac{1}{3} \cdot (a_L + a_M + a_R) \quad (20)$$

Questa "resistant line" ha il pregio di non essere praticamente influenzata dai valori anomali (si ricordi che un valore è definibile anomalo quando si discosta inaspettatamente dagli altri). Però la insensibilità ai valori anomali ha un prezzo. I valori trovati in prima istanza per  $a$  e  $b$  non sono sempre i più appropriati. Un buon modo per controllare i valori che si sono trovati è quello di calcolare i residui, trattando i punti

$$(x, \text{residuo}) = [X_i, Y_i \cdot (a - b \cdot X_i)] \quad (21)$$

come dati X-Y e trovando i "summary points" come detto precedentemente. Se la pendenza  $b'$  per questi ulteriori punti è nulla o quasi nulla, la procedura precedentemente svolta può considerarsi corretta. Altrimenti è possibile modificare la pendenza originale aggiungendo la pendenza residua  $b'$  alla precedente. Ovviamente vanno in tal caso ricalcolati i residui per vedere se la nuova pendenza di essi è prossima allo zero. Può ac-

cadere di svolgere una correzione esagerata per cui i residui mostrano una pendenza in senso diverso da quello dei dati originali. Quando abbiamo due pendenze una troppo piccola (i residui hanno una pendenza positiva) e una troppo elevata (i residui hanno una pendenza negativa) possiamo dedurre che la pendenza corretta si trova all'interno di esse.

È spesso possibile migliorare la pendenza stimata usando la formula:

$$b_{\text{new}} = b_2 - b_2 \cdot \left[ \frac{(b_2 - b_1)}{(b_2 - b_1)} \right] \quad (22)$$

dove  $b_1$  e  $b_2$  sono le pendenze dei residui dei punti aventi  $b_1$  e  $b_2$  come pendenze.

#### ESEMPI DI CALCOLO

Utilizzando le diverse tecniche di regressione discusse nel paragrafo precedente, si vogliono valutare le differenze fra le stesse in termini di risultati. Intanto, si riporta nella **figura 6** un diagramma nel quale la correlazione fra percentuale di quarzo e distanza laterale è calcolata secondo diverse tecniche (Troutman e Williams, 1987) evidenziando differenze marcate. Altri dati utilizzati sono quelli tratti dal lavoro di Chen e Kulhawy (1993) che pongono in relazione, in varie combinazioni, le resistenze a taglio non drenate misurate con varie modalità ( $UU$ ,  $UC$ ,  $CIUC$ ).

Gli Autori calcolano tali relazioni con la tecnica OLS, minimizzando la somma dei quadrati delle differenze delle ordinate e cioè  $Y_i(\text{mis}) - Y_i(\text{calc})$ .

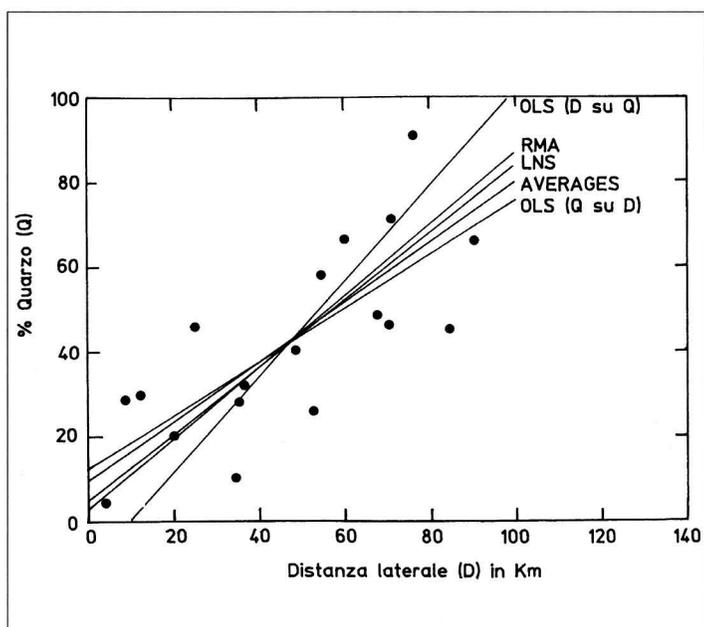


Figura 6 - Diagramma che lega la percentuale di quarzo con la distanza laterale, con le valutazioni statistiche svolte secondo differenti tecniche.

Nelle figure seguenti sono riportate le rette calcolate con i vari metodi proposti nel paragrafo precedente.

In particolare si sono esaminate le relazioni che legano:

$$\frac{S_u(CIUC)}{\sigma_{vo}} \text{ con } \frac{S(UU)}{\sigma_{vo}}$$

$$\frac{S_u(CIUC)}{\sigma_{vo}} \text{ con } \frac{S(UC)}{\sigma_{vo}}$$

dove  $\sigma_{vo}$  è, ovviamente, la tensione efficace verticale.

I risultati sono riportati nelle **figure 7 e 8** e nelle **tabelle II e III**.

Nel caso in cui la relazione è fra

$$\frac{S_u(CIUC)}{\sigma_{vo}} \text{ e } \frac{S_u(UU)}{\sigma_{vo}}$$

le differenze fra le varie tecniche appaiono apprezzabili dal punto di vista grafico. Dal punto di vista numerico

si può quantizzare lo scostamento della retta OLS dalle altre. In particolare il termine noto della retta trovata con la tecnica OLS è il più alto di tutti (0,233) mentre gli altri se ne discostano percentualmente in misura diversa. Analogamente la pendenza della retta OLS è la più bassa. Gli scostamenti percentuali di tale parametro sono anch'essi riportati nella tabella su indicata.

Nel caso della relazione fra

$$\frac{S_u(CIUC)}{\sigma_{vo}} \text{ e } \frac{S_u(UU)}{\sigma_{vo}}$$

le differenze fra le varie rette appaiono ancora più sensibili; ciò è tanto più evidente per quanto riguarda la "Resistant line".

La circostanza è evidentemente da mettere in relazione con la distribuzione alquanto "dispersa" dei punti tenendo altresì conto che le condizioni necessarie ad un buon funzionamento della OLS non sono sostanzial-

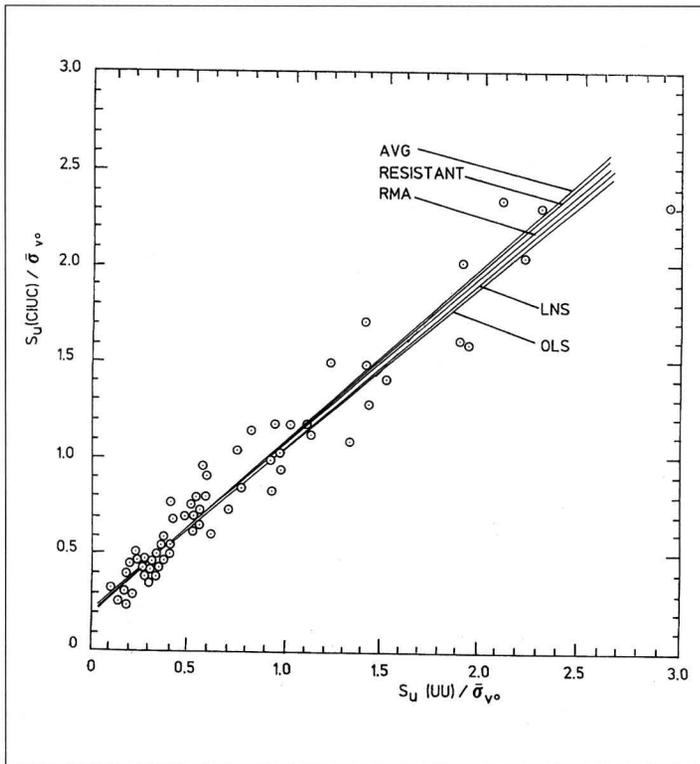


Figura 7 - Valutazione, secondo le diverse tecniche esaminate, della legge statistica che regola i rapporti indicati in figura.

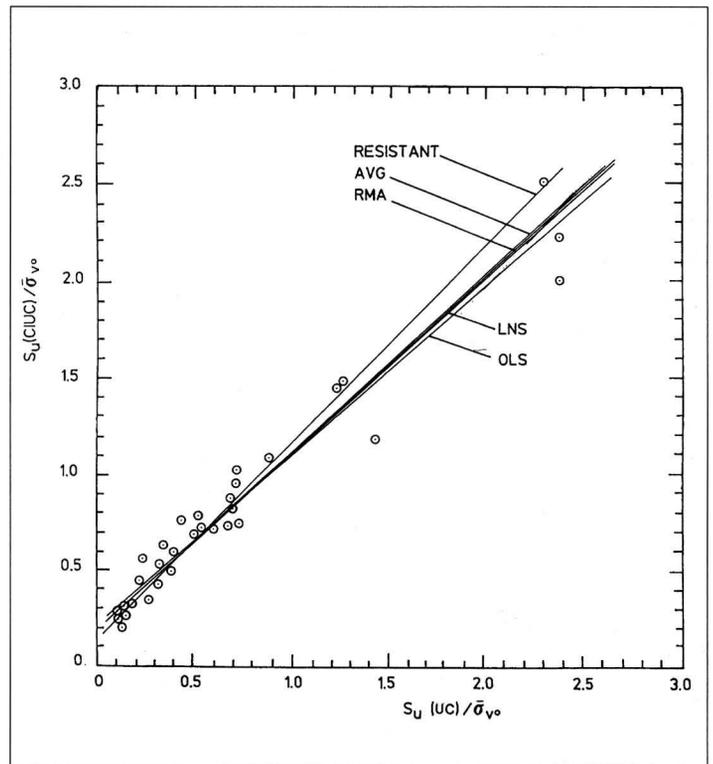


Figura 8 - Valutazione, secondo le diverse tecniche esaminate, della legge statistica che regola i rapporti indicati in figura.

Metodo	Scostamento percentuale Termine noto	Scostamento percentuale Pendenza
Ols	0.00 (%)	0.00 (%)
Lns	- 5.60 (%)	-2.00 (%)
Rma	- 6.50 (%)	-2.40 (%)
Avg	-15.90 (%)	+6.20 (%)
Resistant	-15.00 (%)	+5.60 (%)

Metodo	Scostamento percentuale Termine noto	Scostamento percentuale Pendenza
Ols	0.00 (%)	0.00 (%)
Lns	- 6.00 (%)	+ 2.30 (%)
Rma	- 6.50 (%)	+ 2.50 (%)
Avg	- 3.90 (%)	+ 3.10 (%)
Resistant	-31.50 (%)	+16.60 (%)

mente rispettate. Per quanto riguarda gli scostamenti percentuali essi risultano in tal caso più elevati rispetto alla figura precedente, particolarmente nel caso della Resistant line. Come considerazione generale si può dire che nel caso di dati strutturati come quelli precedentemente esaminati, il metodo LNS può essere considerato particolarmente adatto allo scopo di interpolare correttamente i dati in quanto minimizza i quadrati delle distanze dei punti della retta, in senso normale, non essendoci giustificazioni alcune per la minimizzazione in senso verticale o, addirittura in senso orizzontale.

## CONCLUSIONI

Sono stati presi in rassegna alcuni fra i più comuni errori che si compiono nell'applicazione "tout court" delle tecniche di regressione classiche alla geotecnica ed alla geologia applicata. L'uso sconsigliato di metodi non adatti può condurre a risultati anche incongruenti e, comunque a stime che non sono le "migliori" possibili. Sono state inoltre proposte e discusse, anche con un esempio su dati pubblicati molto recentemente, delle tecniche di regressione diverse dalla classica OLS. Sulla base delle considerazioni svolte e dei risultati ottenuti si è potuto constatare, pur nella limitatezza delle analisi svolte, che esistono tecniche svariate di "fitting" di dati che andrebbero utilizzate in maniera appropriata in funzione della struttura dei dati stessi. ●

## BIBLIOGRAFIA

Abdrabbo F.M., Mahmoud M.A. (1990) - Correlations between index tests and compressibility of Egyptian clays. *Soils and Foundations* vol. 20 n. 2 pp. 128-132.

Afifi A.A., Azen S.P. (1979) - Statistical analysis. A Computer Oriented Approach. Academic Press N.Y.

Carrier III W.D., Beckman J.F. (1984) - Correlations between index tests and the properties of remoulded clays. *Geotechnique*, vol. 34 n. 2 pp. 221-228.

Chapra S.C., Canale R.P. (1988) - Metodi numerici per l'ingegneria. McGraw Hill libri Italia s.r.l.

Chen Y.J., Kulhawy F.H. (1993) - Undrained strength interrelation among CIUC, UU and UC tests. *Journal of Geotechnical Engineering*, ASCE vol. 119 n. 11, pp. 1732-1749.

Draper N., Smith H. (1981) - Applied regression analysis. Second edition. John Wiley and Sons, N.Y.

Gennaro P. (1975) - Introduzione alla statistica. Etas Libri, Milano.

Grace H., Henry J.K.M. (1957) - The planning and design of the new Hong Kong airport. *Proc. ICE*, vol. 7 pp. 275-305.

Jaksa M.B., Kaggwa W.S., Brooker P.I. (1993) - Geostatistical modelling of the spatial variation of the shear strength of a stiff, overconsolidated clay. *Proc. Conference on Probabilistic Methods in Geotechnical Engi-*

*neering*, A.A. Balkema, pp. 185-194.

Koutsoftas D.C., Foot R., Handfelt L.D. (1987) - Geotechnical investigations offshore Hong Kong. *Journal of Geotechnical Engineering*, vol. 113 n. 2 pp. 87-105.

Lee S.H.-H. (1992) - Analysis of the multicollinearity of regression equations of shear wave velocities. *Soils and Foundations*, vol. 32 n. 1 pp. 205-214.

Li K.S. (1990) - The marine soils of Hong Kong. *Proc. 10th Southeast Asian Geotechnical Conference*, Taipei, vol. 1 pp. 431-436.

Li K.S. (1991) - Discussion of "Correlations between index tests and compressibility of Egyptian clays". *Soils and Foundations*, vol. 31 n. 3 pp. 146-148.

Li K.S., White W. (1987) - Probabilistic Characterization of Soil Profiles. *Research Report R19*, Department of Civil Engineering, Australian Defence Force Academy.

Li K.S., White W. (1993) - Use and misuse of regression analysis and curve fitting in geotechnical engineering. *Proceedings of the Conference on Probabilistic methods in geotechnical Engineering*. Canberra, pp. 145-152.

Lumb P. (1967) - Statistical methods in soil investigations. *Proc. 5th Australian-New Zealand Conference on Soil Mechanics and Foundation Engineering*, Auckland, pp. 26-33.

Morris P.H., Rynn J.M.W., Williams D.J. (1987) - Engineering implications of the linear regression of geotechnical data. *Civil Engineering Transactions*, IEAust, CE29(4), pp. 239-247.

Nagaraj T.S., Jayadeva M.S. (1983) - Critical reappraisal of plasticity index of soils. *Journal of Geotechnical Engineering*, vol. 109 n. 7, pp. 994-1000.

Nova R. (1981) - Su alcune correlazioni empiriche con l'indice di plasticità. *Rivista Italiana di Geotecnica*, Anno XV n. 3, pp. 170-175.

Pindyck R.S., Rubinfeld D.L. (1991) - *Econometric Models and Economic Forecasts*. McGraw-Hill Inc., New York.

Ramanathan R. (1989) - *Introductory Econometrics*. Harcourt Brace Jovanovick Publishers, San Diego.

Rendon-Herrero O. (1980) - Universal compression index equation. *Journal of Geotechnical Engineering*, ASCE, vol. 106 n. 11, pp. 1179-1200.

Rizkallah V., Maschwitz G. (1983) - Statistical evaluation of soil properties by using an extended regression analysis. *IV ICASP*, Firenze, pp. 919 - 930.

Troutman B.M., Williams G.P. (1987) - Fitting straight lines in earth sciences. In "use and abuse of Statistical Methods in the Earth Science" (Ed. WB Size) - Oxford University Press, New York, pp. 107-128.

Tukey J. (1977) - *Exploratory data analysis*. Addison Wesley Publishing Company.

Velleman P.F., Hoaglin D.C.C. (1981) - *ABC of EDA*. Duxbury Press, Boston.

Williams G.P., Troutman B.M. (1987) - Algebraic manipulation of equations of best-fit straight lines. In "use and abuse of statistical methods in the Earth Science" (Ed. WB Size), Oxford University Press, New York, pp. 129-141.

Windisch E.J., Yong R.N. (1990) - A statistical evaluation of some engineering properties of eastern Canadian clays. *Canadian Geotechnical Journal* Vol. 27 n. 3 pp. 373-386.