Rivista di Statistica Applicata Vol. 19 n. 2 Giugno 1986

Osservazioni sull'interpolazione della retta senza intercetta e su un nuovo criterio di interpolazione

Roberto Prisco Istituto di Statistica e R.O. Università di Verona.

L'uso di R^2 come indice di aderenza di una funzione interpolante ai dati osservati è basato sulla proprietà di scomposizione della devianza della variabile dipendente. In questo lavoro viene avanzata la proposta di un indice alternativo, valido qualora la scomposizione suddetta non sia verificata (ad esempio per il modello $y_t = bx$). L'indice già presentato dal Mortara, nei casi "usuali" di interpolazione ai minimi quadrati di modelli lineari completi ($y_t = a+bx$) coincide con R^2. L'Autore esamina poi l'interpolazione di funzioni che rendono massimo l'indice di Mortara allo scopo di valutare se questo possa costituire un criterio di interpolazione accettabile. Mette quindi in luce alcune proprietà di queste interpolanti che risultano diverse, ma strettamente collegate con quelle che si ottengono usando il criterio dei minimi quadrati.

1. Introduzione

Nello studio delle relazioni tra grandezze le tecniche di interpolazione hanno lo scopo di individuare la funzione, di forma analitica fissata, che secondo il "criterio di accostamento" scelto meglio si adatta ai dati assegnati; ottenuta la funzione se ne evidenzia, con un indice, la "bontà dell'accostamento". La pratica comune ricorre come criterio di accostamento ai "minimi quadrati" e a R^2 come misura della bontà dell'accostamento. In questo lavoro si mostra come in certi casi R^2 non sia una valida misura della bontà dell'accostamento e si propone un indice alternativo, esaminando inoltre un criterio di accostamento alternativo ai minimi quadrati ad esso collegato.

Ci troviamo in presenza di n coppie di valori (x_i, y_i) che, rappresentate sul consueto piano cartesiano, individuano i punti corrispondenti; l'insieme dei punti costituisce il grafico di dispersione (non essendo possibili ambiguità i valori saranno indicati senza indice e le sommatorie andranno estese a tutti gli n elementi). Scopo dell'interpolazione è allora assegnare un valore ai parametri (indicati con le prime lettere minuscole dell'alfabeto) della funzione

$$y_t = y_t(x;\ a_0,\ a_1,\ ..\ a_k)$$

in modo che questa soddisfi il criterio scelto. I singoli valori assunti dalle variabili sono indicati con lettera minuscola, le variabili con lettera maiuscola ed i simboli di media $M(\ )$, varianza $\sigma^2(\ )$, covarianza $cov\ (\ ,\ )$ e coefficiente di correlazione $r\ (\ ,\ )$ hanno il significato consueto. Sono inoltre rilevanti nella trattazione le seguenti devianze:

devianza totale $= D_t = \Sigma(y-M(Y))^2$
devianza di interpolazione $= D_i = \Sigma(y_t-M(Y_t))^2$
devianza residua $= D_r = \Sigma(y_t-y)^2$
devianza della X $= D_x = \Sigma(x-M(X))^2$

Per cogliere più agevolmente le relazioni tra i dati e gli indici di bontà di accostamento è opportuno scindere D_t nella seguente somma di componenti, simile a quella già dimostrata da Leoni (1985):

$$D_t = D_i + D_r + n(M(Y_t)-M(Y))^2 + 2\,\Sigma(y-y_t)y_t - 2nM(Y)(M(Y)-M(Y_t)) \quad (1)$$

È immediato poi rilevare che se

a) $\Sigma(y-y_t)y_t = 0$
b) $M(Y) = M\ (Y_t)$

allora

$$D_t = D_i + D_r \quad (2)$$

e questo viene verificato nelle condizioni "usuali" in cui si interpola ai minimi quadrati [min(D_r)] una funzione lineare tale che sia vera anche la *b)*. La proprietà *a)* viene definita come proprietà di ortogonalità degli scarti e la *b)* come eguaglianza tra la media della Y osservata e quella della Y_t interpolata.

Conseguenza immediata della (2) è l'utilità di usare $R^2 = D_i/D_t$ come indice di bontà di accostamento. Alcuni Autori (Goldberger (1964; pag. 217), Montgomery Peck (1982; pag. 38 e segg.)) pur avendo puntualizzato quanto affermato finora usano poi questo indice anche per valutare interpolanti che non soddisfano entrambe le condizioni indicate. È per questo motivo che pare opportuno mettere in guardia gli utenti della statistica da un uso troppo disinvolto di questo indice.

2. L'indice di Mortara

È opportuno che un dice di bontà dell'interpolazione sia compreso tra 0 ed 1 ed assuma i valori estremi rispettivamente nelle due situazioni opposte di minima e massima attitudine delle y_t a sostituire le y (Kvaalseth (1985)). Un indice di questo genere, ricavato da quello con gli scarti assoluti proposto dal Mortara (Mortara (1922; pag. 343)), è

$$M^2 = D_i/(D_i + D_r) \tag{3}$$

il cui valore coincide con quello di R^2 nei casi "usuali", nei quali cioè valgono le proprietà a) e b), e ne differisce negli altri.

3. M² come indice di bontà di accostamento

Non è necessario cercare applicazioni complicate o funzioni scomode da interpolare per trovare il campo di impiego di M^2. È sufficiente infatti interpolare una retta senza intercetta (Montgomery Peck (1982; pag. 41), Leoni (1985)) perché non sia verificata l'applicabilità di R^2. In questa situazione infatti l'interpolazione a minimi quadrati viene attuata facendo assumere al coefficiente angolare b_1 il valore $\Sigma xy/\Sigma x^2$ ed è immediato verificare come non sia necessariamente $M(Y)=M(Y_i)$. Questa relazione è verificata infatti da

$$b_2 = \Sigma y/\Sigma x \tag{4}$$

e b_1 eguaglia b_2 (a parte i casi poco interessanti in cui le y sono uguali a 0 o le x sono uguali a 1) solo quando $y = b_1x$.

Le conseguenze di quanto enunciato in questo paragrafo emergono con chiarezza dall'esempio di Tab. I.

Tab. I - Popolazione e reddito di alcuni comuni della provincia di Verona relativi al 1981.

I valori dell'indicatore di reddito sono stati calcolati da A. Mazzali per uno studio su "Alcuni Indicatori dei Dislivelli Economici Territoriali" in corso di pubblicazione.

Comuni	Popolazione residente in migliaia	Indicatore di reddito prodotto in milioni
Dolcè	2,2	35
Cavaion	3,0	33
Zevio	9,9	87
Isola della Scala	10,0	84
S. Martino B.A.	13,0	110
Bussolengo	13,5	130
Cerea	15,0	120
S. Bonifacio	15,0	120
Villafranca	25,0	140

L'indicatore del reddito prodotto è stato calcolato con un algoritmo che tiene conto di diverse variabili (addetti nei vari rami produttivi) rilevate soltanto in occasione dei censimenti. È utile, quindi, porlo in relazione alla popolazione residente in quanto si ha disponibilità di questo dato anche per gli anni intermedi. È così possibile fare delle previsioni sul valore di questo indicatore per mezzo del modello interpolato. Il problema consiste perciò nel cercare di individuare un modello che fornisca una interpretazione della relazione tra le grandezze esaminate. Sia considerazioni teoriche, non vi è infatti reddito senza abitanti né vi sono abitanti senza reddito, sia l'esame del grafico di dispersione consigliano di ricorrere alla retta senza intercetta. È bene poi non restringere la determinazione del coefficiente angolare del modello $y_t = bx$ al solo b_1 (7,4246 sui questi dati), ma confrontarlo con b_2 allo scopo di scegliere il coefficiente che dà luogo alla retta più "buona". Si pone allora il problema di decidere come misurare il grado di "bontà" dell'interpolazione. Non si può infatti procedere alla sua valutazione ricorrendo all'indice

$$R^2 = D_i/D_t \tag{5}$$

poiché questi assume il valore 1,717. Conferma di questa impossibilità si ha eseguendo il calcolo con la formula "usualmente" equivalente

$$R^2 = 1 - D_r/D_t \tag{6}$$

che per questi dati vale 0,673.

Che R^2 sia un indice della bontà di accostamento compreso tra 0 ed 1 e che sia indifferente calcolarlo ricorrendo all'una o all'altra delle due formule ricordate discende infatti dalla (2). La violazione delle condizioni a) e b) conduce infatti a risultati discordanti e di difficile interpretazione. L'interpolante, allora, è appena accettabile ($R^2 = 0,673$) o più che ottima ($R^2 = 1,717$)? Come interpretare poi un valore maggiore di 1? Il grafico di dispersione fa intuire una interpolazione quasi buona, ma i valori eccessivamente discordi di R^2, uno troppo basso e l'altro troppo alto, non danno tranquillità di sorta.

È stato del resto già mostrato (Kvaalseth (1983), Hahn (1977)) che per modelli non usuali le formule alternative di R^2 cessano di dare valori coincidenti e talvolta possono dare valori esterni all'intervallo 0-1. La soluzione proposta da Kvaalseth, calcolare cioè R^2 con la (6), non elimina però tutte le difficoltà in quanto è possibile (come puntualizzato da Kvaalseth (1985)) che tale formula dia valori negativi.

L'indice M^2 (vedi anche Vianelli (1954; pag. 602)) calcolato con la formula alternativa

$$M^2 = 1 - D_r/(D_i + D_r)$$

presenta in ogni caso due valori coincidenti e sicuramente compresi tra 0 e 1, qualsiasi siano la curva interpolata, i parametri ed il criterio adottato per l'interpolazione. Nel nostro esempio $M^2 = 0{,}837$. L'interpolazione fatta ricorrendo alla (4) dà come risultati $b_2 = 8{,}0582$ e $M^2 = 0{,}838$. Risulta quindi difficile scegliere tra $b_1 = 7{,}4246$ e $b_2 = 8{,}0582$ (con valori di M^2 quasi uguali) e viene quasi immediato cercare una retta che soddisfi il criterio di rendere massimo il valore di M^2 [$Max(M^2)$].

4. L'interpolazione al massimo di M²

4.1 Generalità

Per completare la trattazione dell'esempio mostrato possiamo anticipare che questo criterio di accostamento dà come soluzione

$$b_3 = \Sigma y^2/\Sigma xy$$

e che sui dati della Tab. I si ottiene $b_3 = 7{,}7567$ e $M^2 = 0{,}842$. Ricerchiamo allora di quali proprietà godano le y_t ed i parametri delle funzioni interpolate ricorrendo al criterio proposto qui, per quanto ci consta, per la prima volta.

Questo criterio richiede che i valori dei parametri siano individuati dalla $(k+1)$-pla che rende massimo il valore di M^2. Se esistono le derivate e queste sono continue, condizione necessaria per l'individuazione della $(k+1)$-pla è che tutte le derivate prime siano eguali a 0.

$$\begin{cases} \dfrac{dM^2}{da_0} = \left(D_r \dfrac{dD_i}{da_0} - D_i \dfrac{dD_r}{da_0}\right) (D_i + D_r)^{-2} = 0 \\ \vdots \qquad\qquad\qquad \vdots \qquad\qquad\qquad \vdots \\ \dfrac{dM^2}{da_k} = \left(D_r \dfrac{dD_i}{da_k} - D_i \dfrac{dD_r}{da_k}\right) (D_i + D_r)^{-2} = 0 \end{cases} \tag{7}$$

In quel punto l'hessiano deve essere maggiore di 0 e la generica delle derivate seconde è

$$\begin{aligned} \frac{d}{da_j}\left(\frac{dM^2}{da_i}\right) = \Bigg[\Bigg(& \frac{dD_r}{da_j} \cdot \frac{dD_i}{da_i} + D_r \frac{d}{da_j}\left(\frac{dD_i}{da_i}\right) - \frac{dD_i}{da_j} \cdot \frac{dD_r}{da_i} - \\ & - D_i \frac{d}{da_j}\left(\frac{dD_r}{da_i}\right)\Bigg) (D_i + D_r)^2 + 2(D_i + D_r)\left(\frac{dD_i}{da_j} + \frac{dD_r}{da_j}\right) \cdot \\ & \cdot \left(D_r \frac{dD_i}{da_j} - D_i \frac{dD_r}{da_i}\right)\Bigg] \cdot (D_i + D_r)^{-4} \end{aligned} \tag{8}$$

Il criterio di $min(D_r)$ richiede invece che

$$\begin{cases} \dfrac{dD_r}{da_0} = 0 \\ \vdots \qquad \vdots \\ \dfrac{dD_r}{da_k} = 0 \end{cases} \tag{9}$$

è immediato rilevare allora che le due $(k+1)$-ple possono coincidere solo se $D_r = 0$ oppure se

$$\begin{cases} \dfrac{dD_i}{da_0} = 0 \\ \vdots \qquad \vdots \\ \dfrac{dD_i}{da_k} = 0 \end{cases} \tag{10}$$

4.2 Modelli lineari nei parametri

Verifichiamo ora la possibilità che i due criteri diano ai parametri gli stessi valori quando la funzione da interpolare sia lineare come:

$$y_t = a_0 + a_1x_1 + a_2x_2 + \ldots + a_kx_k$$

da cui

$$M(Y_t) = a_0 + a_1M(X_1) + a_2M(X_2) + \ldots + a_kM(X_k)$$

$$D_i = \Sigma(a_1(x_1 - M(X_1)) + a_2\,(x_2 - M(X_2)) + \ldots + a_k\,(x_k - M(X_k)))^2$$

è facile ricavare allora per $i = 0$

$$\frac{dD_i}{da_0} = 0$$

e per $i \neq 0$

$$\frac{dD_i}{da_i} = 2\Sigma(a_1(x_1 - M(X_1)) + \ldots + a_k\,(x_k - M(X_k)))x_i$$

e quindi la prima equazione della (7) diventa

$$\frac{dD_r}{da_0} = 0$$

che equivale alla prima equazione del sistema (9) dalla quale si ottiene $M(Y) = M(Y_t)$. Questa eguaglianza appare come una prima valida proprietà in quanto il criterio proposto, come quello di $min(D_r)$, garantisce il passaggio della funzione lineare completa per il punto $(M(X),M(Y))$. Sostituendo poi le derivate nel (10) si trova, per $a_i \neq 0\ \forall i$, $\Sigma\ (y_t - M(Y_t))^2 = 0$. Per modelli lineari con i coefficienti tutti non nulli il sistema (10) implica cioè che $D_i = 0$. Il sistema (7) quindi ammette soluzioni in comune con il (9) solo se $D_r = 0$ oppure se $D_i = 0$ con $a_i \neq 0\ \forall i$. Il criterio dei minimi quadrati però ammette che le y_t siano tutte eguali solo se $a_0 = M(Y)$ e $a_i = 0\ \forall i \neq 0$. La seconda alternativa non è quindi possibile e le funzioni interpolate possono coincidere solo se $D_r = 0$.

4.3 Interpolazione di $y_t = b_3x$

Verificato che per l'interpolazione di modelli lineari completi i due criteri danno risultati diversi, analizziamo l'interpolazione della retta priva di intercetta:

$$y_t = b_3x$$

Si ricava facilmente che

$$M(Y_t) = \mathrm{b}_3M(X)$$

$$D_i = b_3^2D_x$$

$$\frac{dD_i}{db_3} = 2b_3D_x$$

$$D_r = \Sigma(b_3x - y)^2$$

$$\frac{dD_r}{db_3} = 2\Sigma(b_3x - y)\ x$$

$$\frac{dM^2}{db_3} = 0 \Rightarrow \Sigma(b_3x - y)^2\ 2b_3D_x - b_3^2D_x - \\ - 2\Sigma(b_3x - y)x = 0$$

da cui per $D_x \neq 0$ e $b_3 \neq 0$ si ottiene la formula risolutiva anticipata nel capoverso 4.1

$$b_3 = \Sigma y^2/\Sigma xy$$

e, scrivendo la derivata di M^2 nella forma equivalente

$$\frac{dM^2}{db_3} = D_x \, 2\Sigma(y^2/b_3^3 - xy/b_3^2) \cdot$$
$$\cdot (D_x + \Sigma(x - y/b_3)^2)^{-2},$$

è facile verificare, dallo studio del suo segno, come la (7) individua un punto di massimo relativo. Altri massimi, anche assoluti, possono trovarsi per questa come per altre interpolanti agli estremi del campo di definizione.

4.4 Interpolazione di $y_t = a_1 + b_4x$

Consideriamo ora l'interpolazione della retta completa:

$$y_t = a_1 + b_4x.$$

Seguendo la procedura di calcolo del precedente paragrafo:

$$M(Y_t) = a_1 + b_4M(X)$$

$$D_i = b_4^2 D_x$$

$$\frac{dD_i}{da_1} = 0$$

$$\frac{dD_i}{db_4} = 2b_4D_x$$

$$D_r = \Sigma(a_1 + b_4x - y)^2$$

$$\frac{dD_r}{da_1} = 2\Sigma(a_1 + b_4x - y)$$

$$\frac{dD_r}{db_4} = 2\Sigma(a_1 + b_4x - y)\, x$$

e quindi per $D_x \neq 0$ e $b_4 \neq 0$ si ha con facili passaggi

$$\begin{cases} a_1 + b_4M(X) = M(Y) \\ b_4\Sigma(a_1 + b_4x - y)^2 - b_4^2\, \Sigma(a_1 + b_4x - y)x = 0; \end{cases}$$

si può osservare allora, come già rilevato nel capoverso 4.2 che:

i) i valori interpolati rispettano l'eguaglianza tra le medie
ii) il coefficiente angolare deve essere diverso da zero
iii) i parametri interpolati coincidono con quelli dei minimi quadrati solo quando la devianza residua è nulla.

Con ulteriori passaggi si trova

$$\begin{cases} a_1 = M(Y) - b_4 M(X) \\ b_4 = \sigma^2(Y)/cov(X,Y) \end{cases} \tag{11}$$

Come è possibile ricavare dalla (8), l'hessiano calcolato in questo punto vale $(D_i+D_r)^{-4}\ 4nD_xD_i(D_i—D_r)$; per la proprietà 4, di cui si darà dimostrazione più avanti, questa espressione è maggiore di 0. L'(11) individua quindi un punto di massimo relativo.

Di seguito vengono riportate alcune proprietà di cui godono le rette interpolate al $Max(M^2)$ iniziando dal confronto tra b_4 e il coefficiente angolare $b_5=cov(X,Y)\ /\ \sigma^2(X)$ ottenuto al $min(D_r)$.

PROPRIETÀ 1: La retta interpolata al $Max(M^2)$ ha coefficiente angolare maggiore in valore assoluto di quello della retta interpolata al $min(D_r)$.

PROPRIETÀ 2: La retta $y_t = y_t(x)$ interpolata al $Max(M^2)$ è l'inversa della retta $x_t = x_t(y)$ interpolata al $min(D_r)$.

Nella condizione di correlazione lineare nulla, essendo $cov(X,Y) = 0$ e in tal caso $b_5 = 0$, b_4 invece diverge e la retta interpolata diventa $x = M(X)$. Partendo da una situazione di correlazione nulla si spostino i punti facendoli addensare attorno ad una terza retta inclinata positivamente, si vedrà allora la retta interpolata ai minimi quadrati aumentare l'inclinazione e la retta interpolata al $Max(M^2)$ diminuirla sino a che le tre rette verranno a coincidere.

Altra proprietà rilevante riguarda la correlazione tra i valori y_t e l'errore ($e = y_t — y$); nella interpolazione "usuale" $cov(E,Y_t) = 0$, nella interpolazione esaminata in questo lavoro invece, come è facile dimostrare partendo dalla (1),

$$cov(E,Y_t) = b_4^2\sigma^2(X) - \sigma^2(Y) = \sigma^2(Y_t) - \sigma^2(Y)$$

PROPRIETÀ 3: L'errore $E = Y_t — Y$ e la variabile interpolata Y_t hanno covarianza pari alla differenza tra la varianza della variabile interpolata e quella della Y.

Esaminiamo ora con maggior dettaglio M^2 del quale si possono dare le seguenti espressioni equivalenti

$$M^2 = \frac{D_i}{D_i + D_r} = \frac{b_4^2\sigma^2(X)}{2b_4^2\sigma^2(X) - \sigma^2(Y)} = \frac{1}{2-R^2} \tag{12}$$

e ricordando l'espressione di b_4 si ricava inoltre

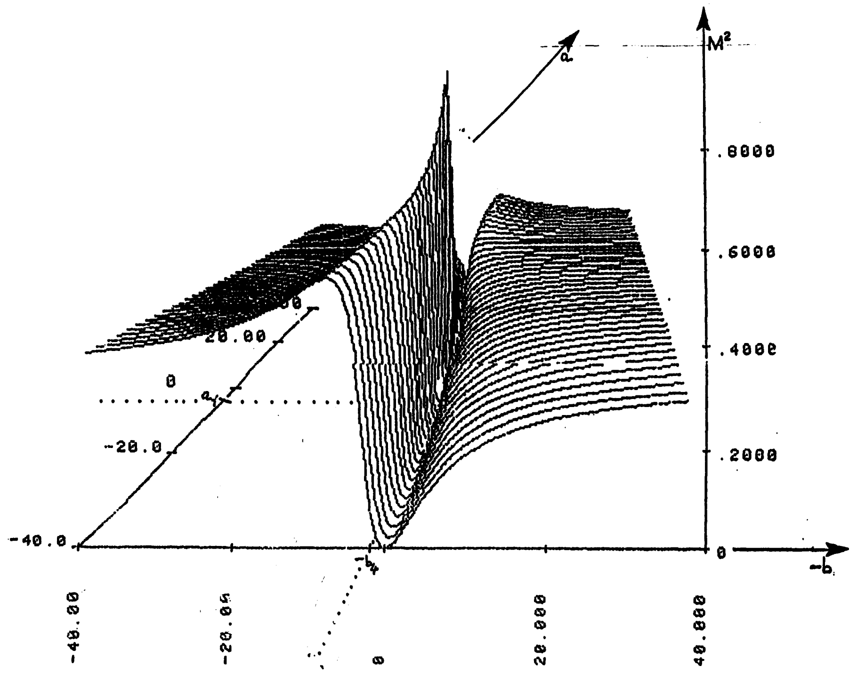

Fig. 1 - Grafico di $M^2 = M^2(a, b)$

$$\lim_{cov(X,Y) \to 0} M^2 = 0{,}5$$

$$cov(X,Y) = \sigma(X)\ \sigma(Y) \Rightarrow M^2 = 1$$

PROPRIETÀ 4: Il valore minimo di M^2 per le rette interpolate a $Max(M^2)$ è 0,5 ed inoltre M^2 è maggiore, o tutt'al più eguale, al valore di R^2 ottenuto con l'interpolazione al $min(D_r)$.

Nell'ambito dell'interpolazione al $min(D_r)$ vale la proprietà

$$r(Y_p, Y) = r(X,Y)$$

che è comune a tutte le rette che passano per il punto $(M(X), M(Y))$ e quindi vale la

PROPRIETÀ 5: Il coefficiente di correlazione calcolato sulle variabili interpolata ed osservata (Y_p, Y) eguaglia il coefficiente di correlazione calcolato sulle variabili indipendente e dipendente (X,Y).

L'andamento di M^2 come funzione di a e di b è piuttosto vario (si veda la Fig. 1); risulta infatti formato di una sorta di "altopiano" inciso da una "valle" in corrispondenza di $b = 0$ dal quale si erge una stretta "guglia" in corrispondenza di (a_1, b_4).

I valori di M^2 sono stati calcolati in riferimento all'interpolazione di una retta tra i punti (1,6), (2,9) e (15,18). Per motivi di migliore prospettiva del grafico l'asse delle b è stato rovesciato.

5. Conclusioni

In questo lavoro si è esaminata l'attitudine di M^2 a soddisfare due esigenze: essere un indice di bontà di accostamento e un criterio di accostamento. Si è visto che M^2 possiede buone qualità per essere un indice di bontà di accostamento, ma che le rette interpolate secondo questo criterio godono di proprietà di non immediata valutazione in confronto alle rette interpolate con il criterio dei minimi quadrati. In particolare la proprietà 2, ma anche le osservazioni successive sul modo di convergere delle due rette (al $Max(M^2)$ e al $min(D_r)$) al crescere della $cov(X,Y)$ e sulla relazione tra M^2 e R^2 sembrano suggerire che i due criteri siano piuttosto due diverse facce di una stessa medaglia. La differenza forse più notevole consiste comunque nel fatto che mentre un criterio impone l'assenza, l'altro impone la presenza della correlazione tra l'errore ed il valore interpolato.

Riferimenti bibliografici

Golberger A.S., *Econometric Theory*, John Wiley & Sons, New York, 1964.
Hahn G.J., Fitting Regression Models with No Intercept Term, *Journal of Quality Technology*, 9, 1977, 47-92.
Kvaalseth T.O., Note on the R^2 measure of goodness of fit for non linear models, *Bulletin of the Psychonomic Society*, 21, 1983, 79-80.
Kvaalseth T.O., Cautionary Note About R^2, *The American Statistician*, 39, 1985, 279-285.
Leoni R., Una osservazione sulla Scomponibilità della Devianza nel Modello di Regressione Lineare Multipla "Senza Intercetta", *Rivista di Statistica Applicata*, 18, 1985, 233-237.
Montgomery D.C. e Peck E.A., *Introduction to Linear Regression Analysis*, John Wiley & Sons, New York, 1982.
Mortara G., *Lezioni di Statistica Metodologica*, Edizione del Giornale degli Economisti e Rivista di Statistica, Città di Castello, 1922.
Vianelli S., *Metodologia Statistica delle Scienze Agrarie*, vol. I, Edizioni Agricole, Bologna, 1954.

Summary

Note on fitting straight lines with no intercept terms and on a new goodness of fit criterion

The partitioning of the sum of squared deviations of Y enables the use of R^2 as measure of goodness of fitting. Such a partitioning fails whenever models are not linear in the constants and also when linear models have no intercept terms. A statistic, already proposed by Mortara, is revised here. This statistic, called M^2, equals R^2 when linear models with intercept are fitted by means of least squares method. The Author argues about the properties of $Max(M^2)$ as a criterion suitable for evaluating goodness of fitting. Fitted models, although having different characteristics are related to the ones obtained by means of least squares.