La regressione mediana

Paolo Radaelli
Dipartimento di Metodi Quantitativi per l'Economia
Università degli Studi di Milano-Bicocca

Indice

Premessa

Questo lavoro é rivolto a studenti universitari che abbiano piena padronanza degli argomenti trattati in un corso base di statistica e, oltre a conoscere la regressione multipla, abbiano le nozioni di base dell'algebra lineare.
Si mostrerà come, nello studio della regressione, sia possibile utilizzare un criterio di accostamento alternativo al metodo dei minimi quadrati; in particolare si adotterà, come criterio di accostamento, la minimizzazione della somma dei valori assoluti dei residui. Tale metodologia, seppur introdotta negli anni '50, è stata sicuramente meno approfondita ed esplorata di altre, per la nota difficoltà di dover operare coi valori assoluti; di converso, la minimizzazione della somma dei quadrati dei residui risulta di più facile trattazione matematica e porta a notevoli applicazioni e vantaggi nell'ambito dell'inferenza statistica.
Verso la fine degli anni '70, con l'articolo di Koenker e Basset [5], la metodologia che verrà di seguito illustrata è stata generalizzata alla regressione dei quantili.

1 Introduzione

Si supponga di disporre di n osservazioni di $p+1$ variabili: $Y; X_1, \ldots, X_j, \ldots, X_p$. Tali osservazioni possono essere così rappresentate:

$$\mathbf{y} = [y_1 \ldots y_i \ldots y_n]';$$

$$\underset{(n\times p)}{X} = \begin{bmatrix} x_{11} & \ldots & x_{j1} & \ldots & x_{p1} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{1i} & \ldots & x_{ji} & \ldots & x_{pi} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{1n} & \ldots & x_{jn} & \ldots & x_{pn} \end{bmatrix}.$$

Si supponga inoltre che sussista una relazione di dipendenza di Y dalle p variabili X_j ($j = 1, \ldots, p$); l'obiettivo della regressione è di individuare un modello che sia in grado di descrivere "al meglio" la variabile dipendente Y in funzione delle variabili esplicative $X_1, \ldots, X_j, \ldots, X_p$:

$$\widehat{Y} = f(X_1, \ldots, X_j, \ldots, X_p). \tag{1.1}$$

Il modello (1.1) sarà tanto più adatto a rappresentare la relazione che lega Y alle p variabili esplicative, quanto più i valori forniti dallo stesso in corrispondenza di una osservazione:

$$\widehat{y}_i = f(x_{1i}, \ldots, x_{ji}, \ldots, x_{pi})$$

risulteranno prossimi ai valori osservati y_i. In altre parole, si vuole che gli scarti

$$r_i = y_i - \widehat{y}_i \qquad i = 1, \dots , n \tag{1.2}$$

siano il più possibile prossimi a zero.
La scelta del modello (1.1) da utilizzare non è ovviamente univoca; nel presente lavoro si considerano modelli lineari nei parametri:

$$\widehat{y}_i = c_1 x_{1i} + c_2 x_{2i} + \ldots + c_j x_{ji} + \ldots + c_p x_{pi} = \sum_{j=1}^{p} c_j x_{ji}. \tag{1.3}$$

Si osservi che è sempre possibile prevedere l'intercetta per il modello (1.3) aggiungendo una variabile esplicativa fittizia di valore costante e pari a 1.
L'obiettivo sarà quindi di determinare i valori degli ignoti parametri $c_1, \dots , c_j, \dots , c_p$ che consentano di minimizzare, secondo un qualche criterio, gli scarti:

$$r_i = y_i - \widehat{y}_i = y_i - \sum_{j=1}^{p} c_j x_{ji} \qquad i = 1, \dots , n. \tag{1.4}$$

Il criterio di minimizzazione degli scarti più noto e diffuso è il *metodo dei minimi quadrati* secondo il quale si individuano i valori dei parametri c_j $(j = 1, \dots , p)$ che rendono minima la somma del quadrato degli scarti tra valori osservati y_i e valori forniti dal modello $\widehat{y}_i$; il problema è quindi:

$$\min_{c_1,\dots,c_j,\dots,c_p} \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 . \tag{1.5}$$

Esempio 1.1 *Per due variabili Y ed X si sono osservati i valori riportati nella seguente tabella:*

X	1	3	4	6	9
Y	3	4	5	10	12

Si vuole determinare l'equazione della retta di Y rispetto ad X secondo il metodo dei minimi quadrati. A questo proposito si osservi che la retta è un caso particolare del modello (1.3) ottenuta per $p = 2$ e ponendo $x_{1i} = x_i$ e $x_{2i} = 1$ per $i = 1, \dots , n$. Si tratta quindi di determinare, utilizzando il criterio di accostamento dei minimi quadrati, i parametri c_1 e c_2 della retta:

$$\widehat{y} = c_1 x + c_2.$$

Nella figura 1 sono rappresentati i punti $(x_i; y_i)$ e la retta a minimi quadrati.

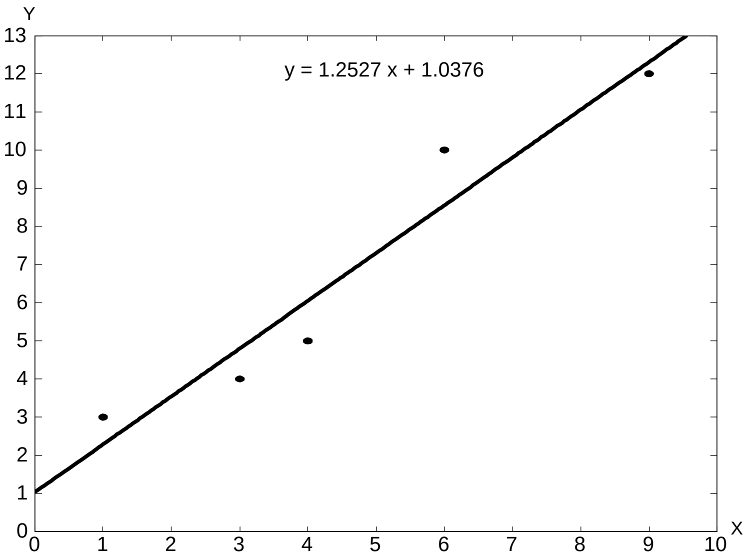


Figura 1: La retta a minimi quadrati

In questo lavoro si presenterà un criterio di accostamento alternativo a quello dei minimi quadrati; in particolare, si vedrà come determinare i parametri $c_1, \ldots, c_j, \ldots, c_p$ del modello (1.3) che rendono minima la somma dei valori assoluti degli scarti r_i. Il problema sarà quindi:

$$\min_{c_1,\ldots,c_j,\ldots,c_p} \sum_{i=1}^{n} |y_i - \widehat{y}_i| = \min_{c_1,\ldots,c_j,\ldots,c_p} \sum_{i=1}^{n} |r_i| \,. \tag{1.6}$$

Nel caso in cui vi sia una sola variabile esplicativa, si tratterà di determinare i parametri della retta di regressione:

$$\widehat{y} = c_1 x + c_2 \tag{1.7}$$

che rendono minima la somma dei valori assoluti degli scarti r_i. Il procedimento risolutivo che si illustrerà è stato proposto da Karst [4] e risulta abbastanza semplice.
Qualora vi siano due o più variabili esplicative, la risoluzione richiede l'utilizzo di tecniche di programmazione lineare; per tale motivo, nel seguito verranno brevemente richiamati i concetti fondamentali della teoria della programmazione lineare.

2 Richiami di programmazione lineare

Esempio 2.1 *Un'azienda di trasporti dispone di un autocarro che viene normalmente utilizzato per trasportare due tipi di prodotti, P_1 e P_2, confezionati in casse del peso di un quintale*

ciascuna, ma di dimensioni diverse; ogni cassa di merce P_1 occupa m^3 $0,5$ e ogni cassa di merce P_2 occupa m^3 $1,5$. Per ciascuna cassa di prodotto trasportata, l'azienda realizza un utile rispettivamente di 3 euro per il prodotto P_1 e di $4,5$ euro per il prodotto P_2.
Sapendo che la capacità dell'autocarro è di m^3 50, mentre la sua portata massima è di q 40 determinare la ripartizione del carico in modo che venga massimizzato l'utile.
Il problema consiste nel determinare il numero di casse x_1 del prodotto P_1 e x_2 del prodotto P_2 da trasportare in modo da massimizzare l'utile del servizio di trasporto tenendo conto però della capacità e della portata dell'autocarro.
Si indichi con $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ il vettore delle variabili decisionali. La funzione da massimizzare (funzione obiettivo) é l'utile complessivo $z = 3\ x_1 + 4,5\ x_2 = \mathbf{c}^T\mathbf{x}$, dove $\mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 4,5 \end{bmatrix}$ é il vettore dei coefficienti della funzione obiettivo.
Nella determinazione di x_1e x_2 è necessario rispettare i vincoli posti dal problema; in particolare il volume occupato dalle casse trasportate non deve essere superiore alla capacità dell'autocarro (m^3 50) mentre il peso complessivo delle casse trasportate non può essere superiore alla portata dell'autocarro (q 40). I due vincoli possono essere espressi mediante le seguenti disuguaglianze:

$$\begin{array}{ll} x_1 + x_2 \leqslant 40 & \textit{(capacità)} \\ 0,5\ x_1 + 1,5\ x_2 \leqslant 50 & \textit{(portata)} \end{array}.$$

Inoltre esiste un vincolo di non negatività sulle variabili decisionali:

$$x_1, x_2 \geqslant 0.$$

Indicando con $A = \begin{bmatrix} 1 & 1 \\ 0,5 & 1,5 \end{bmatrix}$ la matrice dei coefficienti dei vincoli e con $\mathbf{b} = \begin{bmatrix} 40 \\ 50 \end{bmatrix}$ il vettore dei termini noti, è possibile riscrivere i vincoli come segue:

$$\begin{cases} A\mathbf{x} \leqslant \mathbf{b} \\ \mathbf{x} \geqslant \mathbf{0} \end{cases}.$$

In definitiva il problema di programmazione lineare può essere così formalizzato:

$$\begin{cases} \max_{\mathbf{x}}\ z = \mathbf{c}^T\mathbf{x} \\ \textit{soggetto a} \quad A\mathbf{x} \leqslant \mathbf{b} \\ \qquad\qquad\quad \mathbf{x} \geqslant \mathbf{0} \end{cases}$$

o, equivalentemente:

$$\begin{cases} \max_{\mathbf{x}}\ z = 3\ x_1 + 4,5\ x_2 \\ s.a. \quad x_1 + x_2 \leqslant 40 \\ \qquad\ 0,5\ x_1 + 1,5\ x_2 \leqslant 50 \\ \qquad\ x_1, x_2 \geqslant 0 \end{cases}$$

L'esempio 2.1 è un classico problema di programmazione lineare. In generale si ha un problema di programmazione lineare quando si vogliono determinare i valori delle variabili $x_1, \ldots, x_p$ tali da rendere massima (minima) una funzione obiettivo lineare, dovendo soddisfare m vincoli lineari sulle variabili stesse; in generale si ha quindi:

$$\max_{\mathbf{x}} \quad \left(\text{oppure } \min_{\mathbf{x}}\right) \quad z = c_1x_1 + c_2x_2 + \ldots + c_px_p$$

soggetta ai vincoli

$$\left\{\begin{array}{cccccccc} a_{11}x_1+ & a_{12}x_2+ & \ldots & +a_{1p}x_p & \leqslant & (\geqslant) & (=) & b_1 \\ a_{21}x_1+ & a_{22}x_2+ & \ldots & +a_{2p}x_p & \leqslant & (\geqslant) & (=) & b_2 \\ \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ a_{m1}x_1+ & a_{m2}x_2+ & \ldots & +a_{mp}x_p & \leqslant & (\geqslant) & (=) & b_m \end{array}\right. . \tag{2.1}$$

Un problema di programmazione lineare è detto in **forma standard** quando è del tipo:

$$\max_{\mathbf{x}} \; z = c_1x_1 + c_2x_2 + \ldots + c_px_p \tag{2.2}$$

soggetta ai vincoli

$$\left\{\begin{array}{cccccc} a_{11}x_1+ & a_{12}x_2+ & \ldots & +a_{1p}x_p & \leqslant & b_1 \\ a_{21}x_1+ & a_{22}x_2+ & \ldots & +a_{2p}x_p & \leqslant & b_2 \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{m1}x_1+ & a_{m2}x_2+ & \ldots & +a_{mp}x_p & \leqslant & b_m \end{array}\right. \tag{2.3}$$

$$x_j \geqslant 0, \qquad j = 1, \ldots, p. \tag{2.4}$$

Un problema di programmazione lineare è detto in **forma canonica** quando è del tipo:

$$\max_{\mathbf{x}} \; z = c_1x_1 + c_2x_2 + \ldots + c_px_p \tag{2.5}$$

soggetta ai vincoli

$$\left\{\begin{array}{cccccc} a_{11}x_1+ & a_{12}x_2+ & \ldots & +a_{1p}x_p & = & b_1 \\ a_{21}x_1+ & a_{22}x_2+ & \ldots & +a_{2p}x_p & = & b_2 \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{m1}x_1+ & a_{m2}x_2+ & \ldots & +a_{mp}x_p & = & b_m \end{array}\right. . \tag{2.6}$$

$$x_j \geqslant 0, \qquad j = 1, \ldots, p. \tag{2.7}$$

E' possibile dimostrare che ogni problema di programmazione lineare, anche con variabili non necessariamente non negative, può essere risolto riconducendosi ad un problema di programmazione lineare in forma standard o in forma canonica.

2.1 Brevi considerazioni geometriche

Nella programmazione lineare risulta di particolare utilità la rappresentazione grafica del problema sul piano, nel caso in cui vi siano solo due variabili decisionali. In primo luogo è necessario individuare il dominio della funzione obiettivo così come definito dai vincoli posti dal problema. Si introduce pertanto il concetto di *regione ammissibile*. La regione ammissibile, per un problema di programmazione lineare in forma standard, (2.2), (2.3) e (2.4), può essere definita come:

$$R = \{\mathbf{x} \in I\!R^p : A\mathbf{x} \leqslant \mathbf{b}; \mathbf{x} \geqslant \mathbf{0}\}. \tag{2.8}$$

In altre parole R è l'insieme dei vettori $\mathbf{x} \in I\!R^p$ che soddisfano i vincoli del problema.
Si riprenda l'esempio 2.1 e si rappresenti graficamente la regione ammissibile (figura 2). I vincoli sono indicati rispettivamente con I e II e sono rappresentati dalle linee continue; le linee tratteggiate e parallele indicano le rette di livello della funzione obiettivo per diversi valori di z, infine la freccia indica il verso di aumento della funzione obiettivo. La regione ammissibile è data dal poligono convesso definito dai vertici A,B,C e D. Si osserva che, trattandosi di un problema di massimo, è possibile aumentare il valore della funzione obiettivo fintanto che la stessa ha almeno un punto in comune con la regione ammissibile. Il punto di ottimo corrisponde al valore massimo di z per il quale la funzione obiettivo risulta avere almeno un punto "in comune" con la regione ammissibile; nell'esempio tale punto è rappresentato dal vertice C.
Si può dimostrare che la regione ammissibile R è un sottoinsieme convesso di $I\!R^p$.
Particolare importanza rivestono, per quanto si vedrà in seguito, i punti estremi di R.

Definizione 2.1 *Punto estremo*
Un punto $\widetilde{\mathbf{x}} \in R$ è detto punto estremo di R se non può essere espresso come combinazione lineare convessa di altri due distinti punti di R.
$\widetilde{\mathbf{x}} \in R$ è quindi un punto estremo se e solo se $\forall\ \mathbf{x}^1$ e $\mathbf{x}^2 \in R$, $\forall \lambda : 0 < \lambda < 1$, si ha

$$\widetilde{\mathbf{x}} = \lambda \mathbf{x}^1 + (1-\lambda)\,\mathbf{x}^2 \quad \Rightarrow \quad \widetilde{\mathbf{x}} = \mathbf{x}^1 = \mathbf{x}^2.$$

Riprendendo l'esempio 2.1 si può agevolmente verificare che i punti estremi sono i vertici A, B, C e D.

L'individuazione grafica del punto di ottimo risulta ovviamente meno agevole nel caso di $p = 3$ variabili decisionali (si avrebbe una rappresentazione grafica in $I\!R^3$) e diventa impraticabile per $p > 3$.
Per una trattazione completa dell'argomento si rinvia alle indicazioni bibliografiche [2], [6] e [7].

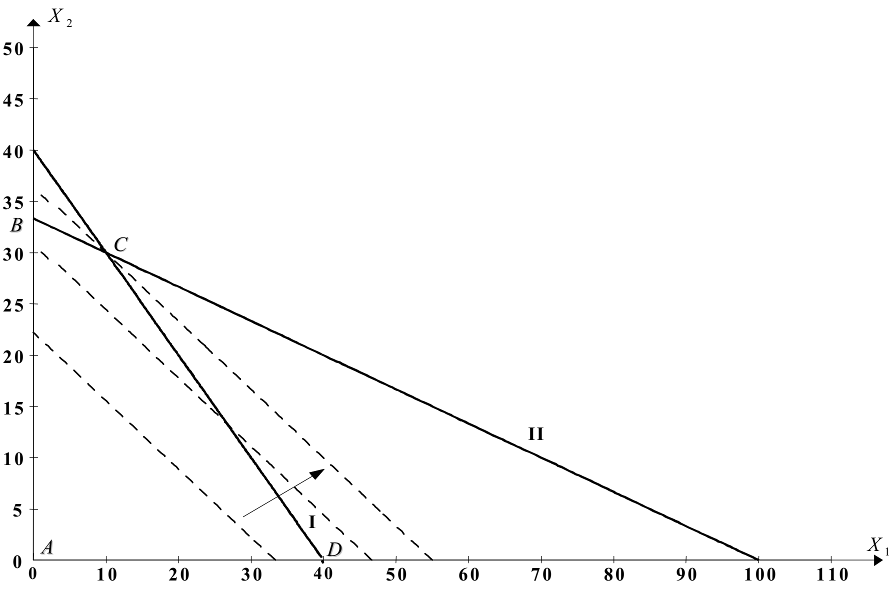


Figura 2: La regione ammissibile

2.2 Soluzione di base

Si consideri un problema di programmazione lineare in forma canonica (2.5), (2.6) e (2.7) che, riscritto in forma matriciale, è:

$$\max_{\mathbf{x}} \; z = \mathbf{c}^T\mathbf{x} \tag{2.9}$$

soggetta ai vincoli

$$A\mathbf{x} = \mathbf{b} \tag{2.10}$$

$$\mathbf{x} \geqslant \mathbf{0} \tag{2.11}$$

Con A matrice di ordine $m \times p$, $\mathbf{c} \in I\!R^p$, $\mathbf{x} \in I\!R^p$ e $\mathbf{b} \in I\!R^m$. Indicando con $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_p$ le p colonne della matrice A, riscriviamo la (2.10) come:

$$x_1\mathbf{A}_1 + x_2\mathbf{A}_2 + \ldots + x_p\mathbf{A}_p = \mathbf{b}. \tag{2.12}$$

In relazione alla matrice A dei coefficienti dei vincoli si assume che $m \leqslant p$ e che sia possibile individuare m colonne di A linearmente indipendenti[1]. Siano $i_1, i_2, \ldots, i_m$ gli indici delle

[1] In altre parole si assume che la matrice A abbia rango m. E' possibile dimostrare che è sempre possibile ricondursi ad un problema di programmazione lineare in forma canonica con queste proprietà partendo da un problema in forma standard.

colonne di A scelte in modo che siano linearmente indipendenti. E' possibile “riordinare” le colonne di A in modo che le prime colonne siano le m linearmente indipendenti; le restanti $p-m$ colonne sono ora le ultime nella “matrice A riordinata”. Si indichi con B la matrice ottenuta accostando le m colonne linearmente indipendenti di A :

$$B = [A_{i_1}\ A_{i_2} \dots\ A_{i_m}]; \tag{2.13}$$

le colonne di B formano una base per $I\!R^m$. Si indichi con D la matrice ottenuta accostando le restanti $p-m$ colonne; possiamo riscrivere la matrice A come $A = [B \vdots D]$.
Lo stesso “riordinamento” effettuato sulle colonne della matrice A deve essere operato sul vettore delle variabili decisionali $\mathbf{x}$ in modo che le variabili corrispondenti alle colonne della matrice B siano le prime nel vettore $\mathbf{x}$, mentre le restanti, quelle associate alla matrice D, siano le ultime. Si ottiene quindi il vettore delle variabili decisionali come accostamento di due vettori:

$$\mathbf{x} = [\mathbf{x}_B \vdots \mathbf{x}_D]^T$$

dove $\mathbf{x}_B = [x_{i_1}\ x_{i_2} \dots\ x_{i_m}]^T$ è detto vettore delle variabili di base mentre $\mathbf{x}_D \in I\!R^{p-m}$ è il vettore costituito dalle restanti $p-m$ variabili non di base.
A questo punto possiamo riscrivere la (2.10) come:

$$A\mathbf{x} = [B \vdots D] \begin{bmatrix} \mathbf{x}_B \\ \mathbf{x}_D \end{bmatrix} = B\mathbf{x}_B + D\mathbf{x}_D = \mathbf{b}. \tag{2.14}$$

Ricordando la metodologia di risoluzione dei sistemi di equazioni lineari, otteniamo la soluzione generale del sistema (2.14) come segue:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_B \\ \mathbf{x}_D \end{bmatrix} = \begin{bmatrix} B^{-1}\mathbf{b} - B^{-1}D\mathbf{x}_D \\ \mathbf{x}_D \end{bmatrix}. \tag{2.15}$$

Nella (2.15) le $p-m$ variabili decisionali del vettore $\mathbf{x}_D$ sono variabili libere.
Se nella soluzione generale (2.15) si pone $\mathbf{x}_D = \mathbf{0}$ si ottiene la soluzione particolare detta *soluzione di base:*

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_B \\ \mathbf{x}_D \end{bmatrix} = \begin{bmatrix} B^{-1}\mathbf{b} \\ \mathbf{0} \end{bmatrix}. \tag{2.16}$$

Se la soluzione di base (2.16) è a componenti non negative, cioè soddisfa il vincolo (2.11) allora è detta *soluzione ammissibile*[2] *di base.* Inoltre la soluzione di base (2.16), eventualmente ammissibile, è detta non degenere se ha solo $p-m$ componenti nulle, le variabili decisionali del vettore $\mathbf{x}_D$, mentre è degenere se le componenti nulle sono almeno $p-m+1$.

[2] Una soluzione $\mathbf{x}$ è detta ammissibile se appartiene alla regione ammissibile R.

Teorema 2.1 *Sia R la regione ammissibile di un problema di programmazione lineare;* $\mathbf{x} \in R$ *é un punto estremo di R se e solo se è soluzione ammissibile di base per il problema di programmazione lineare stesso.*

Teorema 2.2 *(Teorema fondamentale della programmazione lineare)*
Dato un problema di programmazione lineare in forma canonica:

$$\max_{\mathbf{x}} \ \left\{\mathbf{c}^T\mathbf{x} : \mathbf{x} \in R\right\}$$

dove $R = \{\mathbf{x} \in \mathbb{R}^p : A\mathbf{x} = \mathbf{b}, \mathbf{x} \geqslant \mathbf{0}\}$ *e A è una matrice $m \times p$ di rango m, allora:*

i) *Se esiste una soluzione ammissibile* $(\mathbf{x} \in R)$*, allora esiste una soluzione ammissibile di base* $\overline{\mathbf{x}} \in R$*;*

ii) *Se esiste una soluzione ammissibile* $(\mathbf{x} \in R)$ *ottimale (finita), allora esiste una soluzione ammissibile di base ottimale* $\mathbf{x}^* \in R$*.*

Alla luce del teorema 2.2 è quindi possibile risolvere un problema di programmazione lineare calcolando il valore della funzione obiettivo z sulle soluzioni ammissibili di base. Si osservi infatti che il numero di soluzioni ammissibili di base è finito e coincide col numero di matrici B non singolari e di ordine m che si possono "estrarre" a partire dalla matrice A selezionando, di volta in volta, m colonne linearmente indipendenti. Tali possibili scelte sono al più

$$\binom{p}{m} = \frac{p!}{m!\,(p-m)!} \tag{2.17}$$

valore ottenuto quando qualunque sottoinsieme di m colonne delle p della matrice A risulta essere una base per $\mathbb{R}^m$. Pertanto, posta la limitatezza della funzione obiettivo, si potrebbe individuare il punto di ottimo considerando ciascuna delle soluzioni ammissibili di base ottenute in corrispondenza di diverse scelte della matrice B e valutando, per ciascuna di esse, la funzione obiettivo. Questo modo di procedere risulta ovviamente impraticabile al crescere del numero $\binom{p}{m}$ ed è pertanto sconsigliato. Da tale considerazione nasce il metodo del simplesso che verrà brevemente descritto nel paragrafo successivo.

2.3 Il metodo del simplesso

Il metodo del simplesso è un algoritmo iterativo sviluppato da G.B. Dantzig nel 1974; tale algoritmo, partendo da un punto estremo della regione ammissibile R (che in forza del teorema 2.1 è una soluzione ammissibile di base), si "sposta" su un punto estremo adiacente in modo che il valore della funzione obiettivo incrementi o, al peggio, rimanga invariato. Il metodo si arresta in corrispondenza del punto di ottimo o nel caso in cui il problema dato non ammetta soluzione ottimale finita.

L'algoritmo del simplesso si compone quindi di due fasi:

1. un metodo per verificare se una data soluzione ammissibile di base è ottimale;
2. un metodo per "passare" ad una soluzione ammissibile di base adiacente alla precedente in modo da ottenere un miglioramento nella funzione obiettivo.

La terminazione dell'algoritmo è garantita dal fatto che il numero di punti estremi è finito e che nessun punto estremo viene esaminato più di una volta.
Esistono diverse versioni dell'algoritmo del simplesso; nel seguito ne verrà presentata una in particolare tratta da Kolman e Beck [6]; si rimanda alle indicazioni bibliografiche [6] e [7] per una panoramica esaustiva.
Si supponga di avere un problema di programmazione lineare in forma standard (2.2), (2.3) e (2.4); si supponga inoltre che il vettore dei termini noti $\mathbf{b}$ sia non negativo ($\mathbf{b} \geqslant \mathbf{0}$). Per ricondursi ad un problema in forma canonica è necessario ridurre i vincoli in forma di uguaglianza; a tale scopo si introducono delle opportune variabili non negative dette *variabili slack* $s_1, s_2, \dots, s_m$. L'insieme dei vincoli diventa così:

$$\begin{cases} a_{11}x_1+ & a_{12}x_2+ & \dots & +a_{1p}x_p & +s_1 & & & = & b_1 \\ a_{21}x_1+ & a_{22}x_2+ & \dots & +a_{2p}x_p & & +s_2 & & = & b_2 \\ \vdots & \vdots & & \vdots & & & \ddots & & \vdots \\ a_{m1}x_1+ & a_{m2}x_2+ & \dots & +a_{mp}x_p & & & +s_m & = & b_m \end{cases}$$

$$x_j \geqslant 0, \quad j = 1, \dots, p; \qquad s_t \geqslant 0, \quad t = 1, \dots, m$$

Si ha dunque un sistema di vincoli di m equazioni in $p+m$ incognite, oltre al vincolo di non negatività, per tutte le $p+m$ variabili. La matrice A dei coefficienti dei vincoli, il vettore dei coefficienti della funzione obiettivo e il vettore delle variabili decisionali sono rispettivamente:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} & 1 & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & a_{2p} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mp} & 0 & 0 & \dots & 1 \end{bmatrix} \tag{2.18}$$

$$\mathbf{c} = \begin{bmatrix} c_1 & c_2 & \dots & c_p & 0 & 0 & \dots & 0 \end{bmatrix}^T \tag{2.19}$$

$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_p & s_1 & s_2 & \dots & s_m \end{bmatrix}^T. \tag{2.20}$$

Nell'esempio 2.1, la forma canonica associata al problema è:

$$\max_{\mathbf{x}} z = 3\ x_1 + 4,5\ x_2 + 0\ s_1 + 0\ s_2$$

soggetta ai vincoli

$$\begin{cases} x_1 + \quad x_2 + \quad + s_1 \quad = 40 \\ 0,5\ x_1 + \quad 1,5\ x_2 + \quad + s_2 = 50 \end{cases}$$

$$x_1 \geqslant 0,\ x_2 \geqslant 0,\ s_1 \geqslant 0,\ s_2 \geqslant 0$$

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1/2 & 3/2 & 0 & 1 \end{bmatrix}$$

$$\mathbf{c} = \begin{bmatrix} 3 & 9/2 & 0 & 0 \end{bmatrix}^T$$

$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 & s_1 & s_2 \end{bmatrix}^T.$$

Il punto di partenza del metodo del simplesso è una soluzione ammissibile di base. Per come è stato formulato il problema, è possibile ottenere una soluzione ammissibile di base ponendo come variabili di base le m variabili slack, mentre le restanti sono non di base; in altre parole:

$$x_1 = x_2 = \ldots = x_p = 0 \qquad \text{e} \qquad s_1 = b_1,\ s_2 = b_2, \ldots, s_m = b_m.$$

Tale soluzione è ottenuta selezionando le ultime m colonne della matrice 2.18 che costituiscono, come si osserva facilmente, una matrice identità di ordine m. Pertanto la soluzione inziale è:

$$\mathbf{x}_B = B^{-1}\mathbf{b} = \mathbf{b}$$

ed è senz'altro ammissibile data l'ipotesi $\mathbf{b} \geqslant \mathbf{0}$.
Nell'esempio, la soluzione ammissibile di base di partenza è definita come:

$$x_1 = x_2 = 0 \qquad \text{e} \qquad s_1 = 40,\ s_2 = 50$$

dove

$$\mathbf{x}_B = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 40 \\ 50 \end{bmatrix} = \begin{bmatrix} 40 \\ 50 \end{bmatrix}.$$

	x_1	x_2	s_1	s_2	z	
s_1	1	1	1	0	0	40
s_2	1/2	3/2	0	1	0	50
	-3	$-9/2$	0	0	1	0

Tabella 1: La tabella inziale

Il valore della funzione obiettivo corrispondente a questa soluzione di partenza è:

$$z = \mathbf{c}^T\mathbf{x} = \begin{bmatrix} 3 & 9/2 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 40 \\ 50 \end{bmatrix} = 0$$

A questo punto risulta utile riportare il problema e la soluzione di partenza in forma tabellare. Per fare questo si riscrive la funzione obiettivo come:

$$z - 3x_1 - 4,5x_2 = 0; \tag{2.21}$$

z viene ora vista come una nuova variabile.
La tabella iniziale (tabella 1) viene costruita sulla base delle seguente indicazioni:

- nella riga iniziale, che riporta le intestazioni delle colonne, si inseriscono i nomi di tutte le variabili (x_1, x_2, s_1, s_2, z);
- nella riga marginale, detta riga obiettivo, si riportano i coefficienti delle variabili come risultano nella (2.21);
- le prime due righe riportano i coefficienti dei due vincoli;
- nella colonna che riporta le intestazioni delle righe si inseriscono i nomi delle variabili di base nelle corrispondenti equazioni; nel nostro esempio per la prima equazione (vincolo) è di base la variabile s_1 mentre per il secondo s_2;
- nella colonna marginale si riporta il valore delle variabili di base;
- nell'incrocio tra riga e colonna marginale si riporta il valore della funzione obiettivo per la soluzione in esame.

In generale la tabella iniziale sarà:

	x_1	x_2	$\dots$	x_p	s_1	s_2	$\dots$	s_m	z	
s_1	a_{11}	a_{12}	$\dots$	a_{1p}	1	0	$\dots$	0	0	b_1
s_2	a_{21}	a_{22}	$\dots$	a_{2p}	0	1	$\dots$	0	0	b_2
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
s_m	a_{m1}	a_{m2}	$\dots$	a_{mp}	0	0		1	0	b_m
	$-c_1$	$-c_2$	$\dots$	$-c_p$	0	0	$\dots$	0	1	0

.

Criterio di ottimalità

Data una soluzione ammissibile di base è necessario stabilire se essa è ottimale. Per valutare l'ottimalità è necessario verificare se la funzione obiettivo può essere aumentata incrementando il valore di una variabile non di base; tale miglioramento è possibile se nella riga obiettivo vi sono dei valori negativi in corrispondenza di variabili non di base e zero in corrispondenza delle variabili di base.

Se la riga obiettivo di una tabella ha valori nulli in corrispondenza della variabili di base e valori non negativi in corrispondenza delle variabili non di base allora la soluzione rappresentata dalla tabella stessa è ottimale e l'algoritmo termina in quanto è stato raggiunto l'ottimo della funzione obiettivo.

Nella tabella 1 si osserva che per entrambe le variabili non di base x_1 e x_2 si hanno, nella riga obiettivo, valori negativi, rispettivamente -3 e $-9/2$. Le celle della riga obiettivo in corrispondenza delle variabili di base s_1 e s_2 sono invece nulle. Pertanto è possibile aumentare il valore della funzione obiettivo incrementando una delle variabili non di base in modo tale che una variabile di base si annulli in quanto il numero delle variabili di base deve restare immutato. Si osservi che tale procedura consiste nel passare dal punto estremo attuale ad uno adiacente.

Avendo entrambe le variabili non di base valori negativi nella riga obiettivo, il problema è ora di stabilire un criterio per l'individuazione della variabile da incrementare e che quindi entra in base; il criterio che viene generalmente adottato (ne esistono in realtà diversi) è di selezionare la variabile con il valore nella riga obiettivo più elevato in valore assoluto (nel nostro caso x_2). La variabile così individuata viene definita *variabile entrante*. Nell'esempio si ha $4,5 > 3$ pertanto la variabile entrante risulta essere x_2 e viene segnalata con il simbolo $\downarrow$ (tabella 2). Si tratta ora di determinare il valore con cui far entrare in base tale variabile in modo tale che una variabile esca dalla base e che vengano rispettati i vincoli di non negatività delle variabili[3]. Per individuare la *variabile uscente* è necessario determinare il minimo dei rapporti tra i valori delle variabili di base, riportati nella colonna marginale, e i

[3] E' necessario che sia mantenuta l'uguaglianza $A\mathbf{x} = \mathbf{b}$

$$
\begin{array}{c|c|ccccc|c|}
\multicolumn{1}{c}{} & \multicolumn{1}{c}{} & & \downarrow & & & & \multicolumn{1}{c}{} \\
\cline{2-8}
 & & x_1 & x_2 & s_1 & s_2 & z & \\
\cline{2-8}
 & s_1 & 1 & 1 & 1 & 0 & 0 & 40 \\
\leftarrow & s_2 & 1/2 & \boxed{3/2} & 0 & 1 & 0 & 50 \\
\cline{2-8}
 & & -3 & -9/2 & 0 & 0 & 1 & 0 \\
\cline{2-8}
\end{array}.
$$

Tabella 2: L'individuazione della "variabile entrante ↓" e della "variabile uscente ←"

corrispondenti valori nella colonna della variabile entrante. Nella tabella 1:

$$\min \left\{ \frac{40}{1}; \frac{50}{3/2} \right\} = \frac{100}{3}$$

il rapporto minimo si ha in corrispondenza della riga intestata a s_2 che è pertanto la variabile uscente e viene indicata con il simbolo $\leftarrow$ nella tabella 2; inoltre 100/3 è il valore con cui la variabile x_2 entra in base e che satura il secondo vincolo annullando s_2 che esce dalla base. La nuova soluzione ammissibile sarà:

$$x_1 = 0, \;\; x_2 = 100/3 \qquad \text{e} \qquad s_1 = 40, \;\; s_2 = 0$$

e il valore della funzione obiettivo:

$$z = \mathbf{c}^T\mathbf{x} = \begin{bmatrix} 3 & 9/2 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 100/3 \\ 40 \\ 0 \end{bmatrix} = 150$$

Per ricavare la tabella 3, che riporta la nuova soluzione ammissibile, in primo luogo si sostituisce, nella colonna che riporta le intestazioni delle righe, la variabile uscente con la variabile entrante, nell'esempio s_2 con x_2. A questo punto è necessario costruire la matrice B (2.13) costituita dalle colonne di A relative alle nuove variabili di base; tali colonne devono seguire lo stesso ordine con cui compaiono le variabili di base nella tabella. La prima colonna di B sarà pertanto la colonna di A intestata ad s_1 mentre la seconda sarà la colonna intestata ad x_2:

$$B = \begin{bmatrix} 1 & 1 \\ 0 & 3/2 \end{bmatrix}.$$

Trascurando per il momento la riga obiettivo, per ricavare i valori da inserire nelle colonne della tabella 3 è sufficiente premoltiplicare l'inversa della matrice B per le colonne della matrice A.

$$B^{-1} = \begin{bmatrix} 1 & -2/3 \\ 0 & 3/2 \end{bmatrix}$$

Nella colonna intestata ad $x_1$si avrà quindi:

$$B^{-1}A_1 = \begin{bmatrix} 1 & -2/3 \\ 0 & 3/2 \end{bmatrix} \begin{bmatrix} 1 \\ 1/2 \end{bmatrix} = \begin{bmatrix} 2/3 \\ 1/3 \end{bmatrix}.$$

Si procede analogamente per le restanti colonne.
Il valore delle nuove variabili di base, colonna marginale della tabella 3, è:

$$\begin{bmatrix} s_1 \\ x_2 \end{bmatrix} = B^{-1}b = \begin{bmatrix} 1 & -2/3 \\ 0 & 3/2 \end{bmatrix} \begin{bmatrix} 40 \\ 50 \end{bmatrix} = \begin{bmatrix} 20/3 \\ 100/3 \end{bmatrix}$$

Non resta che calcolare i valori da inserire nella riga obiettivo della tabella 3; a questo scopo riportiamo nel vettore $\mathbf{c}_B$ i coefficienti delle variabili di base nella funzione obiettivo:

$$\mathbf{c}_B^T = \begin{bmatrix} 0 & 9/2 \end{bmatrix}$$

e calcoliamo, per ciascuna delle nuove colonne della tabella 3, la differenza:

$$z_j = \mathbf{c}_B^T B^{-1} A_j - c_j$$

dove j indica il numero della colonna mentre c_j è il coefficiente della $j - ma$ variabile nella funzione obiettivo. Nell'esempio si avrà, per la variabile x_1 :

$$z_1 = \mathbf{c}_B^T B^{-1} A_1 - c_1 = \begin{bmatrix} 0 & 9/2 \end{bmatrix} \begin{bmatrix} 2/3 \\ 1/3 \end{bmatrix} - 3 = -3/2$$

per la variabile x_2 :

$$z_2 = \mathbf{c}_B^T B^{-1} A_2 - c_2 = \begin{bmatrix} 0 & 9/2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} - 9/2 = 0$$

per s_1 :

$$z_3 = \mathbf{c}_B^T B^{-1} A_3 - c_3 = \begin{bmatrix} 0 & 9/2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} - 0 = 0$$

infine per s_2 :

$$z_4 = \mathbf{c}_B^T B^{-1} A_4 - c_4 = \begin{bmatrix} 0 & 9/2 \end{bmatrix} \begin{bmatrix} -2/3 \\ 2/3 \end{bmatrix} - 0 = 3.$$

		x_1	x_2	s_1	s_2	z	
$\leftarrow$	s_1	$2/3$	0	1	$-2/3$	0	$20/3$
	x_2	$1/3$	1	0	$2/3$	0	$100/3$
		$-3/2$	0	0	3	1	150

Tabella 3: La soluzione ammissibile di base successiva

	x_1	x_2	s_1	s_2	z	
x_1	1	0	$3/5$	-1	0	10
x_2	0	1	$-1/2$	1	0	30
	0	0	$9/4$	$3/2$	1	165

Tabella 4: La soluzione finale

La tabella per la nuova soluzione ammissibile di base è:
La soluzione rappresentata nella tabella 3 non è ottimale in quanto nella riga obiettivo vi è un valore negativo in corrispondenza di x_1 che sarà quindi la variabile entrante nella nuova soluzione. Procedendo in maniera analoga alla precedente iterazione, osserviamo che la variabile che esce dalla base è s_1. Per la costruzione della tabella 4, che rappresenta la nuova soluzione ammissibile con x_1 e x_2 variabili di base, si ripetono le medesime operazioni illustrate in precedenza, considerando come matrice B la matrice ottenuta affiancando le colonne di A delle variabili di base x_1 e x_2 :

$$B = \begin{bmatrix} 1 & 1 \\ 1/2 & 3/2 \end{bmatrix}.$$

Si giunge al termine alla seguente tabella:
La soluzione rappresentata soddisfa il criterio di ottimalità ed è quindi il punto di ottimo del problema considerato:

$$x_1 = 10,\ x_2 = 30,\ s_1 = 0,\ s_2 = 0.$$

Il valore della funzione obiettivo in corrispondenza di tale soluzione è:

$$z = \mathbf{c}^T\mathbf{x} = \begin{bmatrix} 3 & 9/2 & 0 & 0 \end{bmatrix} \begin{bmatrix} 10 \\ 30 \\ 0 \\ 0 \end{bmatrix} = 165.$$

Pertanto trasportando 10 casse del prodotto P_1 e 30 casse del prodotto P_2 l'azienda consegue il massimo utile (165 *euro*).

Nella trattazione appena esposta, non sono state affrontate alcune situazioni degeneri che si possono presentare nella soluzione di un problema di programmazione lineare con il metodo del simplesso. Si rimanda alle indicazioni bibliografiche [6] e [7] per ulteriori approfondimenti.

2.4 La dualità

Si consideri la coppia di problemi di programmazione lineare:

$$\left.\begin{array}{ll} \max_{\mathbf{x}} & z = \mathbf{c}^T\mathbf{x} \\ \text{s.a.} & A\mathbf{x} \leqslant \mathbf{b} \\ & \mathbf{x} \geqslant \mathbf{0} \end{array}\right\} \tag{2.22}$$

$$\left.\begin{array}{ll} \min_{\mathbf{w}} & z' = \mathbf{b}^T\mathbf{w} \\ \text{s.a.} & A^T\mathbf{w} \geqslant \mathbf{c} \\ & \mathbf{w} \geqslant \mathbf{0} \end{array}\right\} \tag{2.23}$$

con A matrice di ordine $m \times p$, $\mathbf{x}, \mathbf{c} \in I\!R^p$ e $\mathbf{b}, \mathbf{w} \in I\!R^m$. Il problema (2.22) è detto *problema primale* mentre il problema (2.23) è detto *problema duale*.
Si riprenda l'esempio 2.1, il problema primale è:

$$\left.\begin{array}{c} \max_{\mathbf{x}} \; z = \begin{bmatrix} 3 & 9/2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ \text{soggetto a} \\ \begin{bmatrix} 1 & 1 \\ 1/2 & 3/2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leqslant \begin{bmatrix} 40 \\ 50 \end{bmatrix} \\ x_1 \geqslant 0, \, x_2 \geqslant 0 \end{array}\right\}$$

mentre il duale è:

$$\left.\begin{array}{c} \min_{\mathbf{w}} \; z' = \begin{bmatrix} 40 & 50 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \\ \text{soggetto a:} \\ \begin{bmatrix} 1 & 1/2 \\ 1 & 3/2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \geqslant \begin{bmatrix} 3 \\ 9/2 \end{bmatrix} \\ w_1 \geqslant 0, \, w_2 \geqslant 0 \end{array}\right\}.$$

Primale	**Duale**
Max	Min
Coefficienti della funzione obiettivo	Termini noti dei vincoli
Coefficienti dell'$i-mo$ vincolo	Coefficienti dell'$i-ma$ variabile
$i-mo$ vincolo è una disuguaglianza del tipo $\leqslant$	$i-ma$ variabile è $\geqslant 0$
$i-mo$ vincolo è un'uguaglianza	$i-ma$ variabile è $\gtreqless 0$
$j-ma$ variabile è $\gtreqless 0$	$j-mo$ vincolo è un'uguaglianza
$j-ma$ variabile è $\geqslant 0$	$j-ma$ è una disuguaglianza del tipo $\geqslant$
Numero di variabili	numero di vincoli

Tabella 5: Relazioni tra il problema primale e il duale

Si riportano di seguito brevemente le regole da seguire nella costruzione del problema duale: Nella teoria della dualità sono di fondamentale importanza i seguenti teoremi.

Teorema 2.3 *Dato un problema primale, il duale del suo duale è il primale stesso.*

Il teorema 2.3 ci consente di leggere la tabella 5 sia da sinistra verso destra che viceversa.

Teorema 2.4 (Dualità debole) *Se $\overline{\mathbf{x}}$ è soluzione ammissibile del problema primale (2.22) e $\overline{\mathbf{w}}$ è soluzione ammissibile del problema duale (2.23) allora $\mathbf{c}^T\overline{\mathbf{x}} \leqslant \mathbf{b}^T\overline{\mathbf{w}}$.*

Teorema 2.5 *Se il problema primale ammette soluzioni ma la funzione obiettivo è illimitata allora il problema duale non ammette soluzioni e viceversa.*

Teorema 2.6 *Il problema primale ammette soluzione ottimale finita se e solo se il suo duale ammette soluzione ottimale finita.*

Teorema 2.7 (di dualità) a) *Se il problema primale (duale) ha una soluzione ammissibile con funzione obiettivo finita allora il problema duale (primale) possiede soluzione ammissibile con lo stesso valore della funzione obiettivo.*

b) *Se il problema primale (2.22) e il duale (2.23) ammettono soluzioni ammissibili, allora:*

- *Il problema primale ammette soluzione ottimale $\overline{\mathbf{x}}$;*
- *Il problema duale ammette soluzione ottimale $\overline{\mathbf{w}}$;*
- $\mathbf{c}^T\overline{\mathbf{x}} = \mathbf{b}^T\overline{\mathbf{w}}$.

E' possibile sintetizzare le relazioni tra problema primale e duale e i casi che si possono presentare nella seguente tabella dove le intestazioni di colonne e righe indicano rispettivamente: nessuna = nessuna soluzione ammissibile; finita = soluzione ottimale con funzione obiettivo finita; illimitato = soluzione ammissibile con funzione obiettivo illimitata.

	Primale		
Duale	*Nessuna*	*Finita*	*Illimitato*
Nessuna	Si	No	Si
Finita	No	Si	No
Illimitato	Si	No	No

3 La regressione mediana

Nel paragrafo 1 è stato introdotto, come criterio di accostamento alternativo ai minimi quadrati, la minimizzazione della somma dei valori assoluti degli scarti tra i valori osservati e i valori riprodotti dal modello (1.3):

$$\min_{c_1,\dots,c_j,\dots,c_p} \sum_{i=1}^{n} |y_i - \widehat{y}_i| = \min_{c_1,\dots,c_j,\dots,c_p} \sum_{i=1}^{n} \left| y_i - \sum_{j=1}^{p} c_j x_{ji} \right| = \min_{c_1,\dots,c_j,\dots,c_p} \sum_{i=1}^{n} |r_i| \,.$$

Nel seguito si presentaranno distintamente il caso in cui si disponga di una sola variabile esplicativa dal caso in cui le variabili esplicative siano almeno due.

3.1 La retta

Qualora vi sia una sola variabile esplicativa ($p = 1$) il problema si riduce alla determinazione dei parametri c_1 e c_2 della retta[4]:

$$\widehat{y} = c_1 x + c_2 \tag{3.1}$$

che, dati n punti (x_i, y_i), renda minima la quantità:

$$S = \sum_{i=1}^{n} |y_i - \widehat{y}_i| = \sum_{i=1}^{n} |y_i - c_1 x_i - c_2| \,. \tag{3.2}$$

[4]Avendo una sola variabile esplicativa, si pone per comodità $x_1 = x$. L'intercetta può sempre essere inclusa ponendo $x_{2i} = 1$, $i = 1, \dots, n$.

Il problema è stato affrontato, e risolto con un semplice procedimento, da Karst [4] il quale, nel suo lavoro, distingue:

a) il caso in cui si impone che la retta 3.1 passi per un punto specifico;

b) il caso in cui non si pone alcuna limitazione alla retta 3.1.

Nel seguito si illustreranno entrambi casi.

3.1.1 Il caso in cui si fissa un punto per il quale deve passare la retta

Si vuole che la retta 3.1 passi per il punto di coordinate (x^*, y^*), non necessariamente uno degli n punti dati (x_i, y_i).
In primo luogo trasliamo i valori originari (x_i, y_i) in modo che l'origine del nuovo sistema di riferimento sia (x^*, y^*); in altre parole consideriamo i nuovi valori:

$$\begin{aligned} x_i' &= x_i - x^* \\ y_i' &= y_i - y^* \end{aligned} \qquad i = 1, \dots, n. \tag{3.3}$$

E' chiaro che il problema si riduce ad individuare il coefficiente angolare della retta che rende minima la (3.2).
Con i valori traslati, il problema può essere riformulato come segue: determinare il valore di b nella retta:

$$\widehat{y}' = bx' \tag{3.4}$$

che minimizza:

$$S = \sum_{i=1}^{n} |y_i' - \widehat{y}_i'| = \sum_{i=1}^{n} |y_i' - bx'| \,. \tag{3.5}$$

Una volta determinato il valore di b, l'equazione della retta 3.1, riferita ai valori originari, si otterrà dalla trasformazione:

$$y - y^* = b(x - x^*). \tag{3.6}$$

Per spiegare il procedimento che si seguirà per la determinazione di b è necessario premettere alcune considerazioni geometriche sul minimo di S.
Si consideri il contributo di un singolo addendo nella sommatoria (3.5) al valore di S:

$$S_i = |y_i' - bx_i'| \,. \tag{3.7}$$

Rappresentando graficamente la (3.7) in un sistema di assi cartesiani riportando sulle ascisse i valori di b e sulle ordinate i valori di S_i, otteniamo la figura 3 nella quale osserviamo che dal punto di minimo di S_i, di coordinate $\left(\dfrac{y_i'}{x_i'}, 0\right)$, si dipartono due semirette le cui equazioni sono:

$$\begin{cases} S_i = y_i' - b\,|x_i'| & \text{per } \; b < \dfrac{y_i'}{|x_i'|} \\ \\ S_i = -y_i' + b\,|x_i'| & \text{per } \; b > \dfrac{y_i'}{|x_i'|} \end{cases} \tag{3.8}$$

con pendenza, rispettivamente, di $-|x_i'|$ e $|x_i'|$.

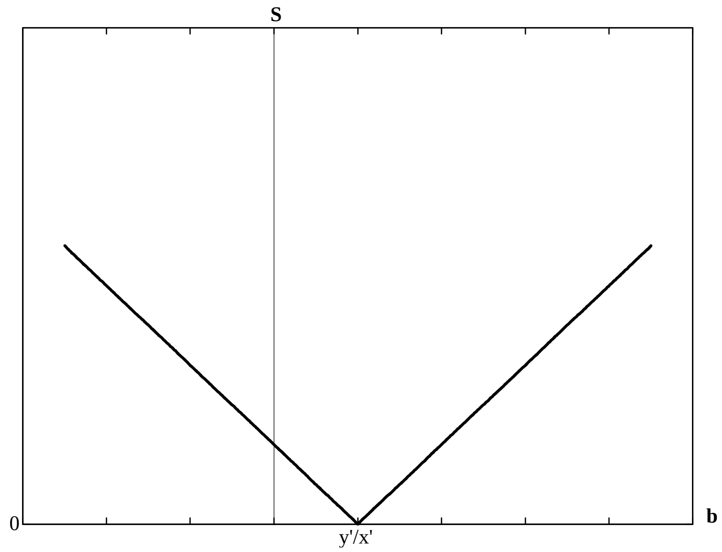


Figura 3: Il contributo di un singolo scarto $|y_i' - bx_i'|$ al valore di S.

Un'analoga rappresentazione grafica per ciascuno degli n punti (x_i', y_i'), consente di "costruire" il grafico di S come somma verticale delle semirette 3.8. Consideriamo ad esempio i punti di coordinate $(1;1)$, $(2;4)$ e $(1;3)$, il grafico di S è riportato nella figura 4.
Risulta chiaro, che il grafico di S è una spezzata poligonale convessa composta da una successione di segmenti adiacenti ciascuno dei quali ha una pendenza pari alla somma algebrica delle pendenze delle corrispondenti semirette. Così il primo segmento, quello più a sinistra, avrà pendenza $-\sum |x_i'|$ mentre l'ultimo avrà pendenza $\sum |x_i'|$. Si può dimostrare che il punto di minimo di S si ha in corrispondenza di uno dei suoi vertici.

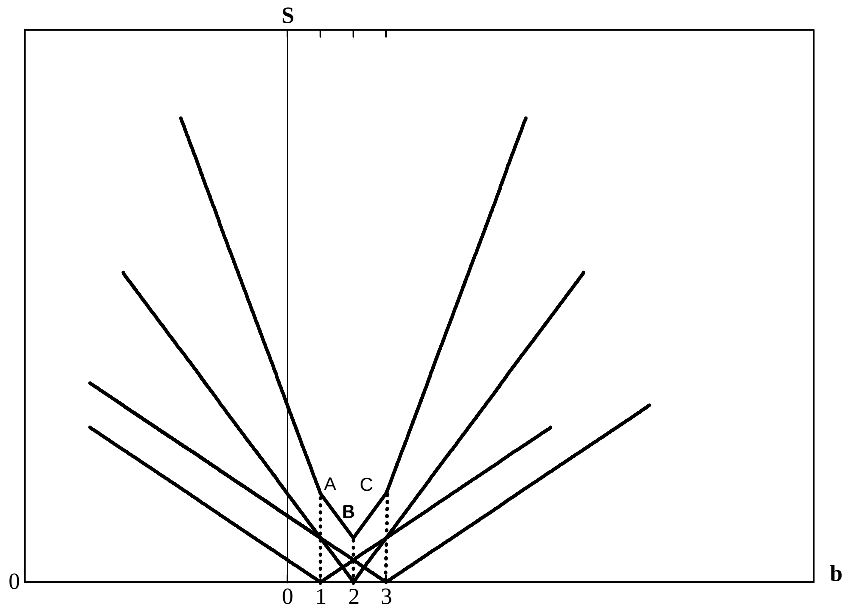


Figura 4: La rappresentazione grafica della somma degli scarti assoluti S.

La possibilità di individuare il valore di b per il quale S è minimo deriva dall'osservazione che i vertici della spezzata di S, cioè i punti in cui vi è una variazione di pendenza dei segmenti adiacenti, si hanno in corrispondenza delle ascisse $\frac{y_i'}{x_i'}$, $i = 1, \ldots, n$. Per determinare il punto di minimo è quindi necessario individuare per quale dei rapporti $\frac{y_i'}{x_i'}$ l'inclinazione dei segmenti che compongono la spezzata cambia di segno e in particolare diventa positiva.
Prima di illustrare, con un esempio, le fasi che porteranno all'individuazione del punto di minimo di S, è necessario aggiungere che, scorrendo la spezzata da sinistra verso destra, l'inclinazione dei segmenti adiacenti aumenta di $2\ |x_i'|$ in corrispondenza dei vertici; ad esempio il secondo segmento, partendo da sinistra, ha pendenza $-\sum |x_i'| + 2\ |x_i'|$.

(1) x_i	(2) y_i	(3) $x'_i = x_i - \overline{x}$	(4) $y'_i = y_i - \overline{y}$	(5) y'_i/x'_i	(6) Pos	(7) $\widetilde{x}_j$
1	3	$-6,3$	$-3,5$	$0,5556$	7	$-2,3$
3	2	$-4,3$	$-4,5$	$1,0465$	9	$1,7$
4	6	$-3,3$	$-0,5$	$0,1515$	4	$-1,3$
5	11	$-2,3$	$4,5$	$-1,9565$	1	$-3,3$
6	7	$-1,3$	$0,5$	$-0,3846$	3	$5,7$
7	5	$-0,3$	$-1,5$	$5,0000$	10	$6,7$
9	4	$1,7$	$-2,5$	$-1,4706$	2	$-6,3$
11	10	$3,7$	$3,5$	$0,9459$	8	$3,7$
13	8	$5,7$	$1,5$	$0,2632$	5	$-4,3$
14	9	$6,7$	$2,5$	$0,3731$	6	$-0,3$
73	65	0	0			

Tabella 6: Determinazione del coefficiente angolare della retta

Esempio 3.1 *Si determini la retta che minimizza la somma degli scarti assoluti dai punti le cui coordinate* (x_i, y_i) *sono riportate nelle prime due colonne della tabella 6 imponendo che la stessa passi per il punto che ha per coordinate le medie aritmetiche delle due serie di valori.*
Calcoliamo in primo luogo le medie aritmetiche di X *e di* Y :

$$
\begin{aligned}
\overline{x} &= \frac{73}{10} = 7,3 \\
\overline{y} &= \frac{65}{10} = 6,5
\end{aligned}
$$

I valori traslati rispetto al punto di coordinate $(7,3;\ 6,5)$*, calcolati secondo la (3.3), sono riportati rispettivamente nella terza e quarta colonna della tabella 6.*
Per ciascuno dei 10 *punti, si calcolano ora i rapporti* $\frac{y'_i}{x'_i}$ *che sono le ascisse dei vertici della spezzata; i valori sono riportati nella quinta colonna della tabella 6.*
E' necessario ora considerare l'ordinamento non decrescente dei rapporti $\frac{y'_i}{x'_i}$*; nella sesta colonna della tabella 6 è riportata, per ciascuno dei rapporti* $\frac{y'_i}{x'_i}$*, la posizione* j *che lo stesso occupa nell'ordinamento non decrescente. Si prenda ad esempio il rapporto*

$$\frac{y'_9}{x'_9} = \frac{1,5}{5,7} = 0,2632.$$

Tale rapporto occupa la quinta posizione nell'ordinamento non decrescente di tutti i 10 rapporti.
Si indichino con $\widetilde{x}_j$ *i valori della terza colonna della tabella 6 riordinati secondo la posizione* j*, riportata nella sesta colonna della tabella 6, dei corrispondenti rapporti* $\dfrac{y'_i}{x'_i}$*. Tali valori sono riportati nell'ultima colonna della tabella 6.*
Si possiedono ora tutti gli elementi necessari per calcolare il coefficiente angolare della retta, passante per il punto $(\overline{x}, \overline{y})$*, che minimizza la somma* S *degli scarti in valore assoluto.*
Dopo aver calcolato $-\sum |x'_i|$ *sarà sufficiente sommare progressivamente i valori* $2\,|\widetilde{x}_j|$ *che, si ricorda, rappresentano l'incremento di pendenza di segmenti adiacenti della spezzata, arrestandosi in corrispondenza del valore che comporta un cambiamento di segno nel risultato.*
Il rapporto $\dfrac{y'}{x'}$ *che si ha in corrispondenza di tale valore, è il coefficiente angolare della retta ricercata.*

$$\begin{array}{rr}
-\sum |x'_i| : & -35,6 \\
2\,|\widetilde{x}_1| : & \underline{\quad 4,6} \\
 & -31 \\
2\,|\widetilde{x}_2| : & \underline{\quad 3,4} \\
 & -27,6 \\
2\,|\widetilde{x}_3| : & \underline{\quad 2,6} \\
 & -25 \\
2\,|\widetilde{x}_4| : & \underline{\quad 6,6} \\
 & -18,4 \\
2\,|\widetilde{x}_5| : & \underline{\quad 11,4} \\
 & -7 \\
2\,|\widetilde{x}_6| : & \underline{\quad 13,4} \\
 & 6,4
\end{array}.$$

Il cambiamento di segno si ha in corrispondenza di $\widetilde{x}_6$*; pertanto il rapporto* $\dfrac{y'}{x'}$ *che occupa la posizione* 6 *(si veda la penultima colonna della tabella 6) nell'ordinamento non decrescente è il coefficiente angolare della retta ricercata. Tale rapporto è:*

$$\frac{y'_{10}}{x'_{10}} = \frac{2,5}{6,7} = 0,3731.$$

L'equazione della retta riferita ai valori originati x_i, y_i*, ottenuta secondo la (3.6), è:*

$$\begin{aligned}
\widehat{y} - \overline{y} &= 0,3731\ (x - \overline{x}) \\
\widehat{y} - 6,5 &= 0,3731\ (x - 7,3)
\end{aligned}$$

$$\widehat{y} = 0,3731\ x + 3,7764. \tag{3.9}$$

Il valore di S per tale retta è:

$$S = \sum_{i=1}^{n} |y_i - 0,3731\ x_i - 3,7764| = 18,3883$$

Si osservi che la retta 3.9 passa per il punto di coordinate $(10; 9)$ Tale proprietà è sempre soddisfatta, infatti la retta che minimizza la somma degli scarti in valore assoluto, con l'imposizione che passi per un prefissato punto, passa sempre per uno degli n punti dati ed in particolare per il punto in corrispondenza del quale si individua il coefficiente angolare della retta stessa.
Il grafico della retta 3.9 è riportato nella figura 5 (linea continua). Nello stesso grafico è rappresentata la retta a minimi quadrati (linea tratteggiata) la cui equazione è:

$$\widehat{y} = 0,3909\ x + 3,6461.$$

Si osservi che le due rette si intersecano nel punto le cui coordinate corrispondono alle medie aritmetiche di X e di Y. La retta a minimi quadrati passa sempre per tale punto, per la prima equazione del sistema normale, mentre la retta 3.9 vi passa per costruzione; ciò significa che se si fosse scelto un altro punto anziché $(7,3; 6,5)$, la retta di migliore adattamento, secondo il criterio degli scarti assoluti, non necessariamente sarebbe passata per il punto $(\overline{x}, \overline{y})$.

Prima di illustrare la determinazione dei parametri della retta nel caso in cui non vi siano restrizioni, è necessario sottolineare che può accadere che S non abbia un unico punto di minimo. Tale circostanza si verifica quando lo stesso valore di S, il minimo, è ottenuto in corrispondenza di più valori del coefficiente angolare, e in particolare per valori del coefficiente angolare compresi in un intervallo. Per ulteriori dettagli si rimanda al lavoro di Karst [4].

3.1.2 Il caso in cui non si impongono restrizioni alla retta

Se non si impone che la retta 3.1 passi per un prefissato punto (x^*, y^*), oltre al coefficiente angolare, è necessario determinare anche l'intercetta della retta. In altre parole si vogliono determinare i valori dei parametri a e b che minimizzano la somma:

$$S = \sum_{i=1}^{n} |y_i - \widehat{y}_i| = \sum_{i=1}^{n} |y_i - bx_i - a|\,. \tag{3.10}$$

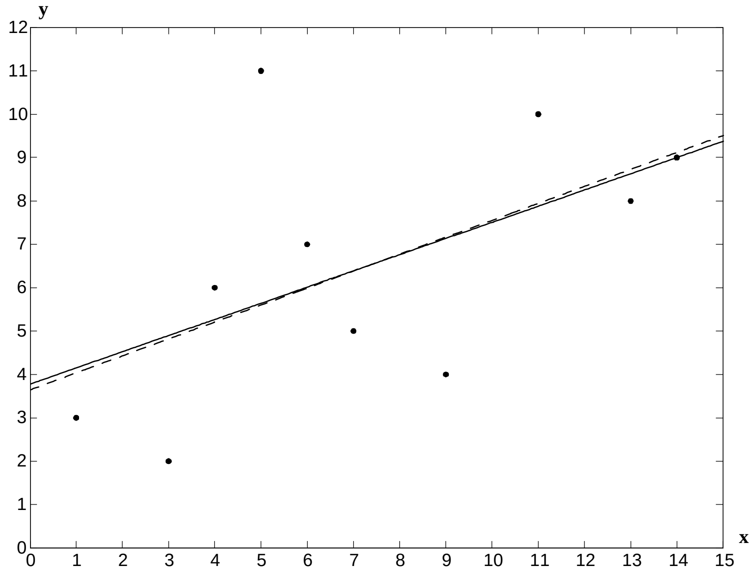


Figura 5: La retta che minimizza gli scarti in valore assoluto (linea continua) e la retta a minimi quadrati (linea tratteggiata).

Anche in questo caso la procedura da seguire nasce da considerazioni geometriche sul grafico di S, che nella fattispecie si dimostra essere un poliedro convesso nello spazio a tre dimensioni (S, a, b).
Tuttavia nel seguito ci si limiterà ad illustrare il procedimento per la determinazione dei parametri a e b invitando alla lettura dell'articolo di Karst [4] per ulteriori approfondimenti. Dati n punti (x_i, y_i), è possibile dimostrare che la retta che minimizza la somma degli scarti assoluti passa per almeno due degli n punti dati. Si potrebbe quindi calcolare il valore di S in corrispondenza di tutte le $\binom{n}{2} = \frac{n(n-1)}{2}$ rette che si possono costruire selezionando due degli n punti (x_i, y_i), scegliendo infine l'equazione della retta in corrispondenza del valore minimo di S. Tale procedura, seppur semplice, diventa ovviamente impraticabile per n elevato.
Karst [4] illustra un algoritmo che consente di esaminare solo un numero limitato delle $\binom{n}{2}$ rette, in particolare al massimo se ne esaminano n. Le fasi di tale procedura sono le seguenti:

a) si seleziona uno qualunque degli n punti (x_i, y_i); si indichi con (x_1, y_1) [5] tale punto;

b) si determina, secondo la procedura illustrata nel paragrafo precedente, il coefficiente angolare b_1 della retta di migliore adattamento (nel senso degli scarti assoluti) passante per (x_1, y_1) [6]. L'equazione della retta di migliore adattamento, rispetto ai dati originari,

[5] Il pedice di x e di y non fa riferimento all'indice i di (x_i, y_i).
[6] La procedura deve essere applicata agli $(n-1)$ punti ottenuti ignorando il punto (x_1, y_1).

è, secondo la 3.6:

$$\widehat{y} - y_1 = b_1(x - x_1). \tag{3.11}$$

Sia S_1 il valore si S calcolato per la retta 3.11.

Per quanto visto nel paragrafo precedente, la retta 3.11 passerà necessariamente per uno degli altri $(n-1)$ punti, indichiamo tale punto con (x_2, y_2);

c) si determina, secondo la procedura illustrata nel paragrafo precedente, il coefficiente angolare b_2 della retta di migliore adattamento (nel senso degli scarti assoluti) passante per (x_2, y_2) [7]. L'equazione della retta di migliore adattamento, rispetto ai dati originari, è:

$$\widehat{y} - y_2 = b_2(x - x_2). \tag{3.12}$$

Sia S_2 il valore di S calcolato per la retta 3.12. Si può dimostrare che $S_2 \leqslant S_1$.

La 3.12 passerà necessariamente per uno degli altri $(n-1)$ punti, indichiamo tale punto con (x_3, y_3). A questo punto si possono presentare due casi:

i) se il nuovo punto (x_3, y_3) coincide con il punto (x_1, y_1) della fase **b** allora significa che la retta 3.12, che coincide in questo caso con la 3.11, è quella che rende minimo il valore di S; l'algoritmo quindi termina;

ii) se il nuovo punto (x_3, y_3) *non* coincide con il punto (x_1, y_1) allora si ritorna alla fase **b** e si considera (x_3, y_3) come punto per cui si vuole passi la retta.

Illustriamo la procedura con un esempio.

Esempio 3.2 *Si riprendano i dati riportati in Zenga [11] (pag. 320) per i quali è già disponibile l'equazione della retta a minimi quadrati. I dati si riferiscono al numero di abbonati* (Y) *ad una rivista scientifica per gli anni* (X) *dal* 1979 *al* 1985.

Anni	$\mathbf{x}_i$	$\mathbf{y}_i$
1979	0	6284
1980	1	7603
1981	2	8591
1982	3	9278
1983	4	10206
1984	5	10614
1985	6	11569

[7] La procedura deve essere applicata agli $(n-1)$ punti ottenuti ignorando il punto (x_2, y_2).

Si determini l'equazione della retta:

$$\widehat{y} = bx + a \tag{3.13}$$

che minimizza:

$$S = \sum_{i=1}^{7} |y_i - \widehat{y}_i| = \sum_{i=1}^{7} |y_i - bx_i - a|\,. \tag{3.14}$$

Scegliamo come punto iniziale i valori dell'anno 1981 ($x_1 = 2$; $y_1 = 8591$)[8]*. Applicando sui restanti punti la procedura illustrata nel paragrafo 3.1.1 si ottiene l'equazione della retta passante per il punto* (2; 8591) *che minimizza la somma degli scarti in valore assoluto:*

$$\widehat{y} = 807,5\ x + 6976 \tag{3.15}$$

in corripondenza della quale si ha $S = S_1 = 1644,5$*. La retta 3.15 passa per il punto di coordinate* (4; 10206)[9] *che diventa il nuovo punto di riferimento* (x_2, y_2) *per la determinazione del nuovo coefficiente angolare. Applicando nuovamente la procedura del paragrafo 3.1.1 si ottiene la retta passante per il punto* (4; 10206) *che minimizza* S*:*

$$\widehat{y} = 867,\overline{6}\ x + 6735,\overline{3}. \tag{3.16}$$

Il valore di S *per la 3.16 è:*

$$S_2 = 1464 < 1644,5 = S_1.$$

La retta 3.16 passa per il punto di coordinate $(1; 7603)$ *che non coincide con il punto precedente* (2; 8591)*, pertanto è necessario ricercare la nuova retta, passante per* (1; 7603)*, che minimizzi* S*. Tale retta è:*

$$\widehat{y} = 793,2\ x + 6809,8 \tag{3.17}$$

in corrispondenza della quale si ha $S = S_3 = 1194,4 < 1464 = S_2$*. La retta 3.17 passa per il punto* $(6; 11569)$ *che è diverso dal punto precedente* $(4; 10206)$ *pertanto si prosegue l'algoritmo.*

---

[8] E' possibile scegliere qualunque degli n punti dati.

[9] E' proprio in corrispondenza di questo punto che si è determinato il coefficiente angolare della retta, infatti:

$$\frac{10206 - 8591}{4 - 2} = 807,5.$$

Si ottiene quindi la retta passante per $(x_4 = 6;\ y_4 = 11569)$ *che minimizza S:*

$$\widehat{y} = 793,2\ x + 6809,8. \tag{3.18}$$

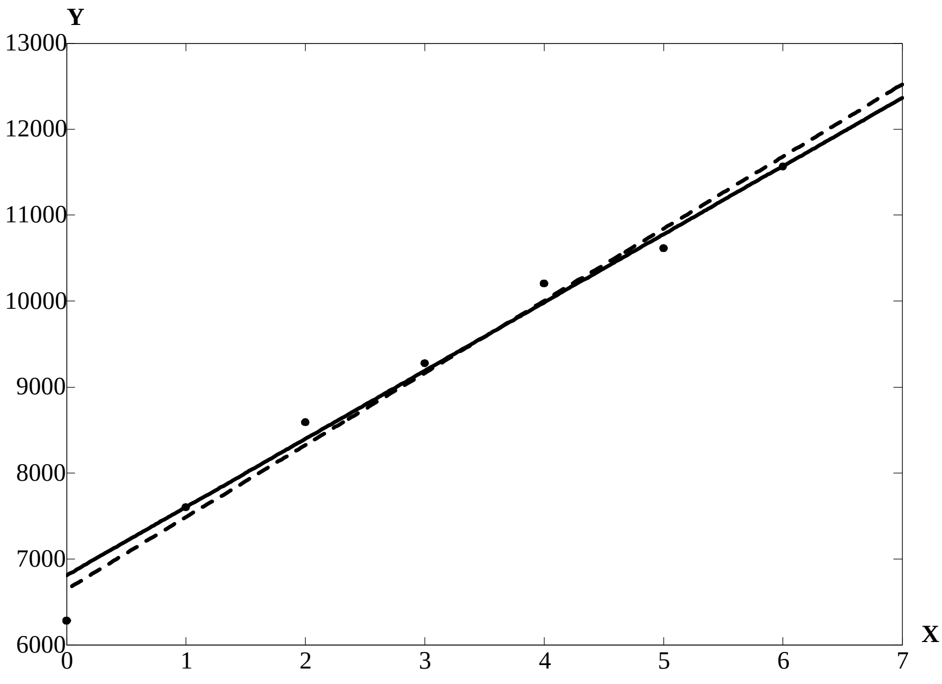
La retta 3.18 coincide con la 3.17, infatti la 3.18 passa per il punto di riferimento precedente $(1; 7603)$ *e l'algoritmo termina.*
La 3.18 è quindi l'equazione della retta che minimizza la somma degli scarti in valore assoluto rispetto agli n=7 punti dati. Il valore minimo di tale somma è:

$$S = S_4 = S_3 = 1194,4.$$

La retta 3.18 è rappresentata nella figura 6 (linea continua) insieme alla retta determinata col metodo dei minimi quadrati (linea tratteggiata) la cui equazione è:

$$\widehat{y} = 839\ x + 6646,57. \tag{3.19}$$

Figura 6: La retta che minimizza la somma degli scarti in valore assoluto (linea continua) e la retta a minimi quadrati (linea tratteggiata).



Un'ultima considerazione in relazione ai residui. Per entrambe le rette 3.18 e 3.19 si sono calcolate le somme dei quadrati e dei valori assoluti dei residui $(y_i - \widehat{y}_i)$. *I risultati sono:*

Somme dei residui	**Retta della regressione mediana**		**Retta a minimi quadrati**
$\sum(y_i - \widehat{y}_i)^2$	$3983449,44$	$>$	$334945,71$
$\sum\lvert y_i - \widehat{y}_i\rvert$	$1194,4$	$<$	$1403,43$

3.2 L'iperpiano di regressione

Si è visto come la procedura iterativa proposta da Karst [4], ed illustrata nel paragrafo (3.1), risulti di immediata comprensione; tuttavia la sua applicazione è limitata alla situazione in cui vi sia un'unica variabile esplicativa ($p = 1$). Quando si vuole "spiegare" il comportamento della variabile dipendente Y mediante due o più variabili esplicative ($p \geqslant 2$) il problema può essere risolto con l'ausilio di tecniche di programmazione lineare.
Nel seguito si illustrerà una procedura inzialmente proposta da Charnes, Cooper e Ferguson [3] nel 1955 e successivamente formalizzata da Wagner [10] nel 1959; faremo principalmente riferimento al secondo lavoro.
Si supponga di disporre di n osservazioni di una variabile dipendente Y e di p variabili esplicative: $X_1, \ldots, X_j, \ldots, X_p$. Si vogliono determinare i coefficienti $c_1, c_2, \ldots, c_p$ del modello:

$$\widehat{y} = c_1x_1 + c_2x_2 + \ldots + c_jx_j + \ldots + c_px_p = \sum_{j=1}^{p} c_jx_j \tag{3.20}$$

che minimizzano la somma degli scarti in valore assoluto:

$$\sum_{i=1}^{n} |r_i| = \sum_{i=1}^{n} |y_i - \widehat{y}_i| = \sum_{i=1}^{n} \left| y_i - \sum_{j=1}^{p} c_jx_{ji} \right| . \tag{3.21}$$

In altre parole il problema che si vuole risolvere è:

$$\min_{c_1,\ldots,c_j,\ldots,c_p} \sum_{i=1}^{n} |y_i - \widehat{y}_i| = \min_{c_1,\ldots,c_j,\ldots,c_p} \sum_{i=1}^{n} \left| y_i - \sum_{j=1}^{p} c_jx_{ji} \right| . \tag{3.22}$$

3.2.1 La rappresentazione dei residui come differenza di due quantità non negative

Il generico scarto $r_i = y_i - \widehat{y}_i$ può essere negativo, positivo o nullo. E' noto che ogni numero reale può essere espresso come differenza di due quantità non negative, così possiamo esprimere il generico scarto come:

$$r_i = \varepsilon_i - \delta_i \quad \text{con } \varepsilon_i, \delta_i \geqslant 0. \tag{3.23}$$

La rappresentazione 3.23 non è in generale unica ma diventa unica quando si impone il vincolo che la somma $\varepsilon_i + \delta_i$ sia minima; in questo caso si ha:

$\varepsilon_i = r_i$	$\delta_i = 0$	se $r_i > 0$
$\varepsilon_i = 0$	$\delta_i = r_i$	se $r_i < 0$
$\varepsilon_i = 0$	$\delta_i = 0$	se $r_i = 0$

.

Esempio 3.3 *Si supponga di voler rappresentare il valore $r = 2$ come differenza di due numeri non negativi ε e δ. Una scelta possibile è $\varepsilon = 3$ e $\delta = 1$ di modo che $r = 2 = \varepsilon - \delta = 3 - 1$. Si potrebbe anche scegliere $\varepsilon = 5$ e $\delta = 3$ di modo che $r = 5 - 3 = 2$; le possibili coppie (ε, δ) sono infinite. Tuttavia se si impone che la somma $\varepsilon + \delta$ sia minima l'unica rappresentazione possibile si ottiene scegliento $\varepsilon = 2$ e $\delta = 0$; in tutti gli altri casi $\varepsilon + \delta > 2$. Si supponga ora di voler rappresentare il valore $r = -4$. Una scelta possibile è $\varepsilon = 1$ e $\delta = 5$ di modo che $r = -4 = \varepsilon - \delta = 1 - 5$. Si potrebbe anche scegliere $\varepsilon = 3$ e $\delta = 7$ di modo che $r = 3 - 7 = -4$; ancora le possibili coppie (ε, δ) sono infinite. Tuttavia imponendo il vincolo che la somma $\varepsilon + \delta$ sia minima l'unica rappresentazione possibile si ottiene scegliento $\varepsilon = 0$ e $\delta = 4$; in tutti gli altri casi $\varepsilon + \delta > 4$.*

Alla luce di questa possibile rappresentazione, si consideri il singolo residuo della 3.21; ovviamente si ha:

$$r_i \begin{cases} > 0 & \text{se } y_i > \sum_{j=1}^{p} c_j x_{ji} \\ = 0 & \text{se } y_i = \sum_{j=1}^{p} c_j x_{ji} \\ < 0 & \text{se } y_i < \sum_{j=1}^{p} c_j x_{ji} \end{cases} ; \tag{3.24}$$

Dalla 3.23 si ottiene pertanto la seguente rappresentazione:

$$r_i = \varepsilon_i - \delta_i \begin{cases} \varepsilon_i = r_i & \delta_i = 0 & \text{se } y_i > \sum_{j=1}^{p} c_j x_{ji} \\ \varepsilon_i = 0 & \delta_i = r_i & \text{se } y_i = \sum_{j=1}^{p} c_j x_{ji} \\ \varepsilon_i = 0 & \delta_i = 0 & \text{se } y_i < \sum_{j=1}^{p} c_j x_{ji} \end{cases} . \tag{3.25}$$

Ovviamente la rappresentazione 3.25 può essere adottata per ciascuno degli n punti; inoltre si avrà:

$$\widehat{y}_i + \varepsilon_i - \delta_i = \sum_{j=1}^{p} c_j x_{ji} + \varepsilon_i - \delta_i = y_i \tag{3.26}$$

Le precedenti considerazioni possono essere interpretate anche in termini geometrici. Si supponga per comodità che il modello utilizzato sia la retta:

$$\widehat{y} = \sum_{j=1}^{p} c_j x_j = c_1 x + c_2. \tag{3.27}$$

Preso uno degli n punti (x_i, y_i), $i = 1, \dots, n$, si possono presentare tre casi (si veda la figura 3.27):

i) il punto giace sopra la retta come nel caso del punto (x_t, y_t). Per questo punto si ha $y_t > \widehat{y_t} = c_1 x_t + c_2$ pertanto, secondo la 3.25, sarà $r_t = \varepsilon_t$ e $\delta_t = 0$;

ii) il punto giace sulla retta come nel caso del punto (x_l, y_l). Per questo punto si ha $y_l = \widehat{y_l} = c_1 x_l + c_2$ pertanto, secondo la 3.25, sarà $r_l = \varepsilon_l = \delta_l = 0$;

iii) il punto giace al di sotto della retta come si ha per il punto (x_k, y_k). In questo caso $y_k < \widehat{y_k} = c_1 x_k + c_2$ pertanto, secondo la 3.25, sarà $r_k = \delta_k$ e $\varepsilon_k = 0$.

3.2.2 La determinazione dei coefficienti dell'iperpiano di regressione come problema di programmazione lineare

Sulla base della rappresentazione dei residui illustrata nel paragrafo precedente, è possibile riscrivere la somma degli scarti assoluti 3.21 come:

$$\sum_{i=1}^{n} |r_i| = \sum_{i=1}^{n} |\varepsilon_i - \delta_i| \tag{3.28}$$

Inoltre, ricordando che, per ciascuno degli n punti, almeno uno tra ε_i e δ_i é nullo, la 3.28 diventa:

$$\sum_{i=1}^{n} |\varepsilon_i - \delta_i| = \sum_{i=1}^{n} \varepsilon_i + \sum_{i=1}^{n} \delta_i \tag{3.29}$$

A questo punto risulta evidente che la ricerca dei parametri dell'iperpiano di regressione (3.20) che minimizzano la somma degli scarti assoluti equivale a ricercare $c_1, c_2, \dots, c_p$ che rendono minima la 3.29 tenedo conto che deve valere, per ciascuno degli n punti, la 3.26 e deve essere $\varepsilon_i, \delta_i \geqslant 0$ $i = 1, \dots, n$.

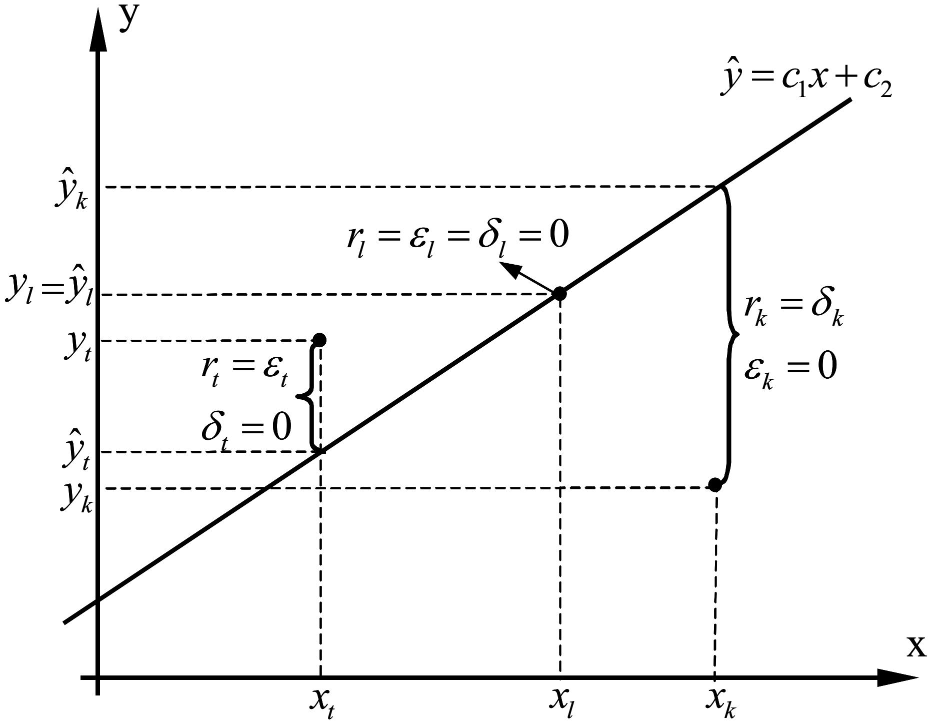


Figura 7: La rappresentazione dei residui come differenza di due quantità non negative.

Il problema di minimo 3.22 può essere pertanto riformulato come il seguente problema di programmazione lineare:

$$\min_{\substack{c_1,\dots,c_p;\\ \varepsilon_1,\dots,\varepsilon_n\\ \delta_1,\dots,\delta_n}} z = \sum_{i=1}^{n} \varepsilon_i + \sum_{i=1}^{n} \delta_i \tag{3.30}$$

soggetto ai vincoli

$$\sum_{j=1}^{p} c_j x_{ji} + \varepsilon_i - \delta_i = y_i \qquad i = 1, \dots, n \tag{3.31}$$

$$c_j \gtreqless 0 \qquad j = 1, \dots, p \tag{3.32}$$

$$\varepsilon_i, \delta_i \geqslant 0 \qquad i = 1, \dots, n \tag{3.33}$$

Rispetto a quanto visto nel paragrafo 2, dedicato alla programmazione lineare, nel problema 3.30 non vi è il vincolo di non negatività sulle p variabili decisionali c_j $(j = 1, \ldots, p)$. Tuttavia, secondo considerazioni del tutto analoghe a quelle viste nel paragrafo 3.2.1, è possibile rappresentare ciascuno dei coefficienti c_j come differenza di due quantità non negative[10]:

$$c_j = d_j - w_j.$$

Inoltre al fine di poter riscrivere il problema 3.30 secondo la forma canonica 2.5 ci si può ricondurre ad un problema di massimo semplicemente moltiplicando per -1 la funzione obiettivo.
In definitiva il problema di minimo 3.30 può essere riscritto come:

$$\max_{\substack{d_1,\ldots,d_p \\ w_1,\ldots,w_p \\ \varepsilon_1,\ldots,\varepsilon_n \\ \delta_1,\ldots,\delta_n}} z' = -\sum_{i=1}^{n} \varepsilon_i - \sum_{i=1}^{n} \delta_i \tag{3.34}$$

soggetta ai vincoli

$$\sum_{j=1}^{p} (d_j - w_j)x_{ji} + \varepsilon_i - \delta_i = y_i \qquad i = 1, \ldots, n \tag{3.35}$$

$$d_j, w_j \geqslant 0 \qquad j = 1, \ldots, p \tag{3.36}$$
$$\varepsilon_i, \delta_i \geqslant 0 \qquad i = 1, \ldots, n \tag{3.37}$$

Il problema consta di $2p + 2n$ variabili $(d_1, \ldots, d_p; w_1, \ldots, w_p; \varepsilon_1, \ldots, \varepsilon_n; \delta_1, \ldots, \delta_n)$ e di n vincoli di uguaglianza 3.35 oltre ai vincoli di non negatività sulle variabili.
E' possibile rappresentare il problema anche in termini matriciali; a questo proposito si osservi che:

i) i coefficienti delle variabili d_j, w_j $(j = 1, \ldots, p)$ nella funzione obiettivo sono nulli;

ii) i coefficienti delle variabili ε_i, δ_i $(i = 1, \ldots, n)$ nella funzione obiettivo sono -1;

iii) i coefficienti delle variabili d_j, w_j $(j = 1, \ldots, p)$ nel generico $i - mo$ vincolo sono rispettivamente x_{ji} e $-x_{ji}$;

[10] In realtà gli algoritmi implementati in alcuni pacchetti informatici rendono superfluo questo ulteriore passaggio in quanto consentono di specificare il limite inferiore e il limite superiore per ciascuna variabile decisionale.

iv) i coefficienti delle variabili ε_i, δ_i $(i = 1, \dots, n)$ nel generico $i - mo$ vincolo sono rispettivamente $+1$ e -1.

In definitiva il problema espresso in termini matriciali è:

$$\left.\begin{array}{ll} \max_{\mathbf{x}} & z' = \mathbf{c}^T\mathbf{x} \\ \text{s.a.} & A\mathbf{x} = \mathbf{y} \\ & \mathbf{x} \geqslant \mathbf{0} \end{array}\right\} \tag{3.38}$$

dove:

$$A = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{p1} & -x_{11} & -x_{21} & \dots & -x_{p1} & 1 & 0 & \dots & 0 & -1 & 0 & \dots & 0 \\ x_{12} & x_{22} & \dots & x_{p2} & -x_{12} & -x_{22} & \dots & -x_{p2} & 0 & 1 & \dots & 0 & 0 & -1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \dots & x_{pn} & -x_{1n} & -x_{2n} & \dots & -x_{pn} & 0 & 0 & \dots & 1 & 0 & 0 & \dots & -1 \end{bmatrix}$$

$$\mathbf{c} = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & -1 & -1 & \dots & -1 & -1 & -1 & \dots & -1 \end{bmatrix}^T$$

$$\mathbf{x} = \begin{bmatrix} d_1 & d_2 & \dots & d_p & w_1 & w_2 & \dots & w_p & \varepsilon_1 & \varepsilon_2 & \dots & \varepsilon_n & \delta_1 & \delta_2 & \dots & \delta_n \end{bmatrix}^T$$

$$\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix}^T.$$

Se si indica con $\underset{(n\times p)}{X}$ la matrice che riporta per riga, i valori osservati delle p variabili esplicative e con I_n la matrice identità di ordine n, possiamo utilmente riscrivere la matrice dei coefficienti dei vincoli A come:

$$A = [X \mid -X \mid I_n \mid -I_n].$$

Dopo aver risolto il problema 3.38, i coefficienti dell'iperpiano di regressione si ottengono per differenza:

$$\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_p \end{bmatrix} - \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix}.$$

Inoltre per ogni valore dell'indice i, almeno una delle variabili ε_i e δ_i deve essere nulla[11].

[11] Saranno entrambe nulle quando il punto giace sull'iperpiano di regressione.

Esempio 3.4 *Si riprenda l'esempio a pag. 9 del testo Zenga [12]. Per* $n = 20$ *aziende agricole sono state rilevate 4 variabili:* Y = *reddito annuo in milioni di lire;* X_1 = *superficie (in ettari);* X_2 = *numero di bovini. I dati sono riportati nella seguente tabella:*

$\mathbf{X}_1$	6	22	18	8	12	10	17	11	16	23	7	12	24	16	9	11	22	11	16	8
$\mathbf{X}_2$	18	0	14	6	1	9	6	12	7	2	17	15	7	0	12	16	2	6	12	15
Y	96	83	126	61	59	90	82	88	86	76	102	108	96	70	80	113	76	74	98	80

Si ricavi l'equazione del piano di regressione:

$$\widehat{Y} = c_1X_1 + c_2X_2 + c_3 \tag{3.39}$$

che minimizza:

$$S = \sum_{i=1}^{n} |y_i - c_1x_{1i} - c_2x_{2i} - c_3| \,. \tag{3.40}$$

Si tratta di determinare i coefficienti c_1, c_2 *e* c_3 *del piano 3.39 che rendono minima la somma 3.40. Si osservi che è sempre possibile includere nell'equazione del piano l'intercetta* c_3 *inserendo una variabile esplicativa fittizia* X_3 *che assume valore* 1 *per ciascuna delle* $n = 20$ *osservazioni.*
Il problema di programmazione lineare associato ha: $2p + 2n = 2 \cdot 3 + 2 \cdot 20 = 46$ *variabili sulle quali grava il vincolo di non negatività. Ci saranno inoltre* $n = 20$ *vincoli del tipo:*

$$\sum_{j=1}^{3}(d_j - w_j)x_{ji} + \varepsilon_i - \delta_i = y_i$$

$$d_1x_{1i} + d_2x_{2i} + d_3 - w_1x_{1i} - w_2x_{2i} - w_3 + \varepsilon_i - \delta_i = y_i$$

Le matrici e i vettori associati al problema:

$$\left.\begin{array}{ll} \max_{\mathbf{x}} & z' = \mathbf{c}^T\mathbf{x} \\ s.a. & A\mathbf{x} = \mathbf{y} \\ & \mathbf{x} \geqslant \mathbf{0} \end{array}\right\}$$

sono:

$$\underset{(20,46)}{A} = \begin{bmatrix} 6 & 18 & 1 & -6 & -18 & -1 & 1 & 0 & \dots & 0 & -1 & 0 & \dots & 0 \\ 22 & 0 & 1 & -22 & 0 & -1 & 0 & 1 & \dots & 0 & 0 & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 8 & 15 & 1 & -8 & -15 & -1 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & -1 \end{bmatrix}$$

$$\underset{(46,1)}{\mathbf{c}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & \dots & -1 & -1 & -1 & \dots & -1 \end{bmatrix}^T$$

$$\underset{(46,1)}{\mathbf{x}} = \begin{bmatrix} d_1 & d_2 & d_3 & w_1 & w_2 & w_3 & \varepsilon_1 & \varepsilon_2 & \dots & \varepsilon_{20} & \delta_1 & \delta_2 & \dots & \delta_{20} \end{bmatrix}^T$$

$$\underset{(20,1)}{\mathbf{y}} = \begin{bmatrix} 96 & 83 & \dots & 80 \end{bmatrix}^T.$$

Il problema è stato risolto sia usando la funzione "linprog" disponibile per il pacchetto applicativo Matlab 6.5 sia utilizzando la funzione "lin_prog" fornita dalla libreria IMSL-CNL. I risultati ottenuti, pressoché identici[12]*, sono:*

$$\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 889,930 \\ 1038,485 \\ 1039,218 \end{bmatrix} - \begin{bmatrix} 852,763 \\ 1036,918 \\ 1036,185 \end{bmatrix} = \begin{bmatrix} 1,567 \\ 3,033 \\ 37,167 \end{bmatrix}$$

pertanto l'equazione del piano di regressione che minimizza la somma degli scarti assoluti è:

$$\widehat{Y} = 1,567\ X_1 + 3,033\ X_2 + 37,167 \tag{3.41}$$

e il corrispondente valore di S è:

$$S = \sum_{i=1}^{n} |y_i - 1,567\ x_{1i} + 3,033\ x_{2i} + 37,167| = 113,342.$$

L'equazione del piano di regressione ottenuto con il metodo dei minimi quadrati è:

$$\widehat{Y} = 2,138\ X_1 + 3,257\ X_2 + 28,546.$$

Per entrambi i piani si sono calcolate le somme dei quadrati e dei valori assoluti dei residui $(y_i - \widehat{y}_i)$. I risultati sono:

Somme dei residui	**Piano della regressione mediana**		**Piano a minimi quadrati**
$\sum(y_i - \widehat{y}_i)^2$	1160,2	>	993,5033
$\sum \lvert y_i - \widehat{y}_i \rvert$	113,342	<	120,041

Prima di concludere è necessario aggiungere che Wagner [10] nel suo lavoro non risolve direttamente il problema di programmazione lineare (3.30), (3.31) e (3.32), ma propone di utilizzare il corrispondente problema duale con il vantaggio di ottenere una considerevole riduzione del numero di variabili decisionali da $2p+2n$ a n. Nel presente lavoro si è preferito non appesantire la trattazione dell'argomento; si invita il lettore interessato ad approfondire l'argomento.

[12] Differiscono nell'ordine della quarta cifra decimale.

Riferimenti bibliografici

[1] Abdelmalek N.N. (1974), "On the Discret Linear L_1 Approximation and L_1 Solutions of Overdetermined Linear Equations", *Journal of Approximation Theory*, 11, 38-53.

[2] Carcano G. (2002), *Elementi di Programmazione Lineare*, Edizioni Datanova, Milano.

[3] Charnes A., Cooper W.W., Ferguson R.O. (1955), "Optimal Estimation of Executive Compensation by Linear Programming", *Management Science*, 1, 138-151.

[4] Karst O.J. (1958), "Linear Curve Fitting Using Least Deviations", *Journal of the American Statistical Association*, 53, 118-132.

[5] Koenker R., Basset G. Jr (1978), "Regression Quantiles", *Econometrica*, 46, 33-50.

[6] Kolman B., Beck R.E. (1995), *Elementary Linear Programming with Applications*, second edition, Academic Press.

[7] Murty K.G. (1983), *Linear Programming*, John Wiley & Sons.

[8] Ostasiewicz W. (1994). *Line Fitting to Scattered Data*, materiale didattico per il Dottorato di Ricerca in Statistica, Facoltà di Economia - Università di Trento.

[9] Sposito V.A., Kennedy W.J., Gentle J.E. (1977), "Algorithm AS 110: L_p Norm Fit of a Straight Line", *Applied Statistics*, 26, 114-118.

[10] Wagner H.M. (1959), "Linear Programming Techniques for Regression Analysis", *Journal of the American Statistical Association*, 54, 206-212.

[11] Zenga M. (1988). *Introduzione alla Statistica Descrittiva*, Vita e Pensiero.

[12] Zenga M. (1994). *Metodi Statistici per l'Economia e l'Impresa*, Giappichelli Editore, Torino.