9 La regressione

Nell'analisi dei fenomeni economici, sociali, naturali, spesso, è interessante studiare le relazioni, di **dipendenza** o di **interdipendenza**, che si possono venire a creare tra due o più variabili, rilevate sugli stessi soggetti o oggetti.

Quando fra due variabili, o fra una e più variabili, è possibile individuare una **relazione unidirezionale**, si suole parlare di "dipendenza".

Soffermiamoci sull'analisi della dipendenza fra **variabili quantitative**.

La **dipendenza statistica** non è una dipendenza di tipo logico, cioè di tipo *causa-effetto*. Si pensi, ad esempio, alla relazione esistente tra i redditi di una coppia di coniugi, dovuta probabilmente al fatto che i due coniugi appartengono in genere alla stessa classe sociale.

Distinguiamo due tipologie di variabili:

- le variabili **indipendenti o esplicative**, che si suppone assumano valori determinati;
- le variabili **dipendenti o di risposta**, affette da errori accidentali.

La scelta dell'una o dell'altra variabile come indipendente o come dipendente non è arbitraria, ma è legata alla natura del fenomeno; si pensi, ad esempio, alla relazione reddito-consumi, in cui sono i consumi a dipendere dal reddito e non viceversa.

Se non è possibile stabilire quale variabile possa essere considerata come logicamente "antecedente" e quale come "conseguente", ci si può interessare alla misura dell'**interdipendenza** (coefficiente di correlazione lineare di Bravais-Pearson). Si pensi al tipo di relazione esistente fra **statura** e **peso**, in cui le due variabili si influenzano reciprocamente.

9.1 La regressione lineare semplice

Supponiamo di aver rilevato due sole variabili, una indipendente X e l'altra dipendente Y; disponiamo dunque di n coppie di osservazioni (x_i, y_i):

x_i	y_i
x_1	y_1
x_2	y_2
x_3	y_3
.	.
.	.
.	.
x_n	y_n

Ci chiediamo: qual è la "vera" relazione funzionale esistente fra le due variabili X e Y? Essa potrebbe essere determinata con esattezza:

$$Y=f(x),$$

se la variabile dipendente Y non fosse affetta da errori accidentali.

In realtà, cioè, noi non rileviamo, per ogni soggetto/oggetto, la vera grandezza Y_i, ma rileviamo un dato y_i affetto da errore:

$$y_i=Y_i+\varepsilon_i.$$

Si pensi, ad esempio, alla relazione *reddito-consumi*, per cui è irrealistico pensare che percettori dello stesso reddito abbiano la stessa spesa per consumi. Gli errori ε_i tengono conto di tutti quei fattori che influiscono sui consumi e che sono diversi dal reddito (altre rendite, propensione alla spesa, ecc…).

Se, rappresentate su un sistema di assi cartesiani le n coppie di punti (x_i,y_i), è possibile ipotizzare che la relazione teorica che lega le due variabili X e Y sia di tipo lineare:

$$Y=f(x)=\alpha+\beta x,$$

il problema è quello di individuare, fra le infinite rette che si ottengono al variare di α e β, quella più probabile, cioè quella che presumibilmente ha generato la nostra serie di dati.

Assumendo che gli errori abbiano distribuzione normale, con media 0 e varianza costante σ^2:

$$\varepsilon_i \sim N(0, \sigma^2),$$

il miglior metodo per “stimare” i parametri α e β è il metodo dei minimi quadrati.

Il **metodo dei minimi quadrati** consiste nel minimizzare la somma dei quadrati degli scarti tra valori osservati y_i e valori teorici Y_i:

$$R = \sum_{i=1}^{n} \left(y_i - Y_i\right)^2 = \sum_{i=1}^{n} \left(y_i - \alpha - \beta x_i\right)^2 = \text{minimo.}$$

Calcoliamo dunque le derivate parziali di R rispetto ad α e β e uguagliamole a zero; si ottiene un sistema di due equazioni in due incognite, detto **sistema di equazioni normali**:

$$\frac{\partial R}{\partial \alpha} = -2\sum_{i=1}^{n} \left(y_i - \alpha - \beta x_i\right) = 0 \qquad (1)$$

$$\frac{\partial R}{\partial \beta} = -2\sum_{i=1}^{n} \left(y_i - \alpha - \beta x_i\right) x_i = 0 \qquad (2)$$

Dalla equazione (1) si ha:

$$\sum_{i=1}^{n} y_i - n\alpha - \beta \sum_{i=1}^{n} x_i = 0 \qquad \Rightarrow \qquad \hat{\alpha} = \frac{\sum_{i=1}^{n} y_i}{n} - \beta \frac{\sum_{i=1}^{n} x_i}{n} = M_y - \beta M_x$$

in cui β si ricava dalla (2).

Dalla equazione (2) si ha:

$$\sum_{i=1}^{n} x_i y_i - \alpha \sum_{i=1}^{n} x_i - \beta \sum_{i=1}^{n} x_i^2 = 0$$

e sostituendo ad α l’espressione trovata:

$$\sum_{i=1}^{n} x_i y_i - \left(\frac{\sum_{i=1}^{n} y_i}{n} - \beta \frac{\sum_{i=1}^{n} x_i}{n} \right) \sum_{i=1}^{n} x_i - \beta \sum_{i=1}^{n} x_i^2 = 0$$

$$\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n} + \beta \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} - \beta \sum_{i=1}^{n} x_i^2 = 0$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} = \frac{M_{XY} - M_X M_Y}{{}_q M_X^2 - M_X^2} = \frac{\sigma_{XY}}{\sigma_X^2}$$

Le stime di α e β sono dunque:

$$\hat{\alpha} = M_Y - \hat{\beta} M_X$$

$$\hat{\beta} = \frac{\sigma_{XY}}{\sigma_X^2}$$

dove:

$\hat{\alpha}$ rappresenta l’intercetta con l’asse delle ordinate;

$\hat{\beta}$ rappresenta il “coefficiente angolare” della retta, dunque la sua inclinazione, pertanto -∞< $\hat{\beta}$ <+∞. Il segno di $\hat{\beta}$ dipende ovviamente dal segno della covarianza. $\hat{\beta}$ esprime di quanto varia, in media, *Y* al variare di un’unità di *X*; se $\hat{\beta}$ =*0*, *Y* è indipendente da *X*.

Le stime di α e β si possono determinare con maggiore facilità considerando gli scarti $\overline{x}_i = x_i - M_X$ in luogo di x_i. Ciò equivale a considerare una traslazione dell’origine nel punto di coordinate (M_x, *0*), e quindi una traslazione dell’asse *Y*, di cui bisogna tener conto quando si stima α. La funzione di perdita è in tal caso:

$$R = \sum_{i=1}^{n} \left(y_i - Y_i\right)^2 = \sum_{i=1}^{n} \left(y_i - \alpha - \beta \overline{x}_i\right)^2 = \text{minimo}$$

e il sistema di equazioni normali diviene:

$$\frac{\partial R}{\partial \alpha} = -2 \sum_{i=1}^{n} \left(y_i - \alpha - \beta \overline{x}_i\right) = 0 \qquad (1)$$

$$\frac{\partial R}{\partial \beta} = -2\sum_{i=1}^{n}\left(y_i - \alpha - \beta\overline{x}_i\right)\overline{x}_i = 0 \qquad (2)$$

Dalla equazione (1) si ha:

$$\sum_{i=1}^{n} y_i - n\alpha - \beta\sum_{i=1}^{n}\overline{x}_i = 0 \qquad \Rightarrow \qquad \hat{\alpha} = \frac{\sum_{i=1}^{n} y_i}{n} = M_y$$

cui bisogna sottrarre βM_x , se si vuol tornare al vecchio sistema di riferimento.

Dalla equazione (2) si ha:

$$\sum_{i=1}^{n}\overline{x}_i y_i - \alpha\sum_{i=1}^{n}\overline{x}_i - \beta\sum_{i=1}^{n}\overline{x}_i^2 = 0 \;\Rightarrow\; \sum_{i=1}^{n}\overline{x}_i y_i - \beta\sum_{i=1}^{n}\overline{x}_i^2 = 0 \;\Rightarrow\; \hat{\beta} = \frac{\sum_{i=1}^{n}\overline{x}_i y_i}{\sum_{i=1}^{n}\overline{x}_i^2},$$

dove $\sum_{i=1}^{n}\overline{x}_i y_i = \sum_{i=1}^{n}\overline{x}_i \overline{y}_i$. Infatti:

$$\sum_{i=1}^{n}\left(x_i - M_X\right)\left(y_i - M_Y\right) = \sum_{i=1}^{n}\left(x_i - M_X\right)y_i - M_Y\sum_{i=1}^{n}\left(x_i - M_X\right) = \sum_{i=1}^{n}\left(x_i - M_X\right)y_i \;.$$

SCOMPOSIZIONE DELLA DEVIANZA TOTALE

La devianza di Y (*DEVT*) si può scomporre nella somma di due componenti, la devianza residua (*DEVE*) e la devianza di regressione (*DEVR*):

$$\sum_{i=1}^{n}\left(y_i - M_Y\right)^2 = \sum_{i=1}^{n}\left(y_i - \hat{Y}_i\right)^2 + \sum_{i=1}^{n}\left(\hat{Y}_i - M_Y\right)^2 .$$

Infatti è:

$$\sum_{i=1}^{n}\left(y_i - M_Y\right)^2 = \sum_{i=1}^{n}\left(y_i - \hat{Y}_i + \hat{Y}_i - M_Y\right)^2 =$$

$$= \sum_{i=1}^{n}\left(y_i - \hat{Y}_i\right)^2 + \sum_{i=1}^{n}\left(\hat{Y}_i - M_Y\right)^2 + 2\sum_{i=1}^{n}\left(y_i - \hat{Y}_i\right)\left(\hat{Y}_i - M_Y\right),$$

dove, se consideriamo gli scarti $\overline{x}_i = x_i - M_X$ in luogo di x_i, si ha:

$$\sum_{i=1}^{n}\left(y_i - \hat{Y}_i\right)\left(\hat{Y}_i - M_Y\right) = \sum_{i=1}^{n}\left(y_i - \hat{Y}_i\right)\left(\hat{\alpha} + \hat{\beta}\overline{x}_i - M_Y\right) =$$

$$= \sum_{i=1}^{n} \left(y_i - \hat{Y}_i\right)\left(M_Y + \hat{\beta}\bar{x}_i - M_Y\right) = \hat{\beta} \sum_{i=1}^{n} \left(y_i - \hat{Y}_i\right)\bar{x}_i = 0$$

essendo $\sum_{i=1}^{n} \left(y_i - \hat{Y}_i\right)\bar{x}_i = 0$ l'equazione (2) del sistema normale.

Una misura della bontà di adattamento della retta di regressione ai dati è data dal "coefficiente di determinazione"

$$R^2 = \frac{DEVR}{DEVT} = 1 - \frac{DEVE}{DEVT}, \quad 0 \le R^2 \le 1,$$

che assume valore 0 se *DEVR=0* e valore 1 se *DEVE=0*. *DEVR=0* se la retta di regressione coincide con la retta passante per M_Y; in tal caso, non c'è dipendenza di *Y* da *X*, essendo tale retta parallela all'asse X. *DEVE=0* se tutti i dati osservati giacciono sulla retta di regressione, ovvero la retta passa esattamente per i punti e l'adattamento può ritenersi ottimo.

Dimostriamo che $\frac{DEVR}{n} = \frac{\sigma_{XY}^2}{\sigma_X^2}$.

Se consideriamo $\bar{x}_i = x_i - M_X$ in luogo di x_i, si ha:

$$\frac{DEVR}{n} = \frac{\sum_{i=1}^{n} (\hat{y}_i - M_Y)^2}{n} = \frac{\sum_{i=1}^{n} \left(\hat{\alpha} + \hat{\beta}\bar{x}_i - M_Y\right)^2}{n} = \frac{\sum_{i=1}^{n} \left(M_Y + \hat{\beta}\bar{x}_i - M_Y\right)^2}{n} =$$

$$= \hat{\beta}^2 \frac{\sum_{i=1}^{n} \bar{x}_i^2}{n} = \frac{\sigma_{XY}^2}{\sigma_X^4} \sigma_X^2 = \frac{\sigma_{XY}^2}{\sigma_X^2}$$

R^2 è dunque pari al quadrato del coefficiente di correlazione lineare. Infatti:

$$R^2 = \frac{DEVR}{DEVT} = \frac{DEVR}{n\sigma_Y^2} = \frac{\sigma_{XY}^2}{\sigma_X^2 \sigma_Y^2} = \rho^2.$$

Facciamo un esempio.

Nella seguente tabella sono riportate le "quantità di precipitazioni *Y*" (in mm) e le "temperature medie *X*" (in gradi centigradi) registrate in 10 stazioni meteorologiche:

y_i	x_i
29	18
35	16
87	14
32	19
112	11
14	20
26	17
120	12
190	9
85	13
730	**149**

Il valore del coefficiente di correlazione lineare indica una forte interdipendenza lineare fra le due variabili di tipo inverso:

x_iy_i	x_i^2	y_i^2
522	324	841
560	256	1225
1218	196	7569
608	361	1024
1232	121	12544
280	400	196
442	289	676
1440	144	14400
1710	81	36100
1105	169	7225
9117	**2341**	**81800**

$$M_X = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{149}{10} = 14,9 \qquad M_Y = \frac{\sum_{i=1}^{10} y_i}{10} = \frac{730}{10} = 73$$

$$M_{XY} = \frac{\sum_{i=1}^{10} x_i y_i}{10} = \frac{9117}{10} = 911,7$$

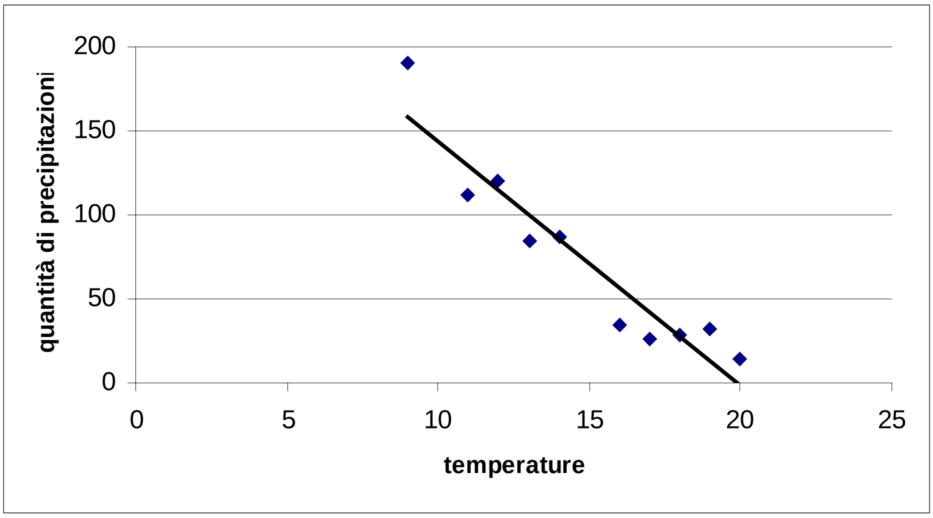
$$\sigma_{XY} = M_{XY} - M_X M_Y = -176$$

$$\sigma_X^2 = \frac{\sum_{i=1}^{10} x_i^2}{10} - M_X^2 = \frac{2341}{10} - (14{,}9)^2 = 12{,}11$$

$$\sigma_Y^2 = \frac{\sum_{i=1}^{10} y_i^2}{10} - M_Y^2 = \frac{81800}{10} - (73)^2 = 2850{,}49$$

$$\rho = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2 \sigma_Y^2}} = -0{,}95\,.$$

La relazione lineare fra le due variabili si evince anche dallo scatterplot di *Y* su *X*:



Determiniamo la retta di regressione di *Y* su *X*:

$$\hat{\beta} = \frac{\sigma_{XY}}{\sigma_X^2} = \frac{-176}{12{,}09} = -14{,}56$$

$$\hat{\alpha} = M_Y - \hat{\beta} M_X = 73 + 14{,}56 \cdot 14{,}9 = 289{,}91$$

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i = 289{,}91 - 14{,}56 x_i\,.$$

All'aumentare della temperatura di 1°, dunque, le quantità di precipitazioni diminuiscono in media di circa 15 *mm*.

La retta si adatta bene ai dati osservati essendo R^2 molto vicino ad 1:

$R^2 = \rho^2 = (-0{,}95)^2 = 0{,}9$.

Volendo stimare le quantità di precipitazioni in corrispondenza di un valore di *X* non osservato, ad esempio $x_i=10$, si può utilizzare la retta di regressione:

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i = 289{,}91 - 14{,}56 \cdot 10 = 144{,}33\,.$$

Nell'esempio considerato ha senso calcolare anche la retta di regressione di *X* su *Y*. In tal caso, le stime dei due parametri saranno:

$$\hat{\beta}' = \frac{\sigma_{XY}}{\sigma_Y^2} = \frac{-176}{2851} = -0{,}06$$

$$\hat{\alpha}' = M_X - \hat{\beta}' M_Y = 14{,}9 + 0{,}06 \cdot 73 = 19{,}41$$

e la retta di *X* su *Y* sarà:

$$\hat{x}_i = \hat{\alpha}' + \hat{\beta}' y_i = 19{,}41 - 0{,}06 y_i\,.$$

Le due rette di regressione si incontrano sempre nel punto di coordinate (M_X, M_Y).

Infatti, la retta $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$ passa per il punto (M_X, M_Y):

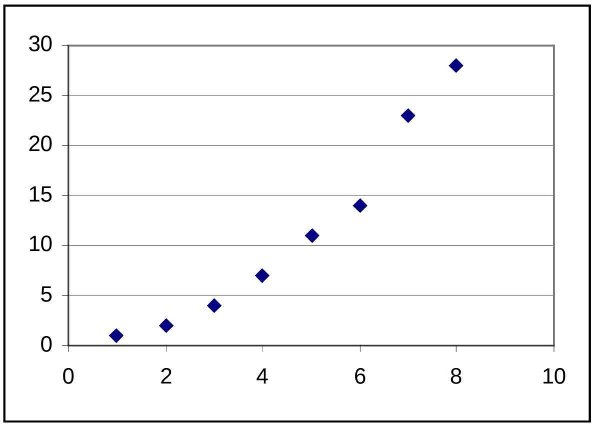
$M_Y = M_Y - \hat{\beta} M_x + \hat{\beta} M_x$..

Analogamente si dimostra che la retta $\hat{x}_i = \hat{\alpha}' + \hat{\beta}' y_i$ passa per lo stesso punto.

Se $\rho=\pm 1$, le due rette sono coincidenti, se $\rho=0$ le due rette sono perpendicolari e quindi le due variabili sono indipendenti linearmente; non è detto però che fra *X* e *Y* non ci sia una dipendenza di altro tipo, ad esempio parabolica.

9.2 La regressione non lineare

Non sempre le *n* coppie (x_i, y_i) dei dati rilevati si dispongono intorno ad una retta, per esempio:



Non sempre, dunque, possiamo ipotizzare che la relazione teorica che lega le due variabili *X* e *Y* sia di tipo lineare.

Nel caso in esame, possiamo pensare che la relazione "vera" fra le due variabili sia di tipo parabolico:

$$Y=f(x)=a+bx+cx^2.$$

Per determinare i parametri *a*, *b* e *c* ricorriamo al "metodo dei minimi quadrati":

$$R=\sum_{i=1}^{n}\left(y_i-Y_i\right)^2=\sum_{i=1}^{n}\left(y_i-a-bx_i-cx_i^2\right)^2=\text{minimo}.$$

Se, per semplicità di calcolo, si considerano gli scarti dalla media aritmetica $\overline{x}_i=x_i-M_x$, la funzione da minimizzare sarà:

$$R=\sum_{i=1}^{n}\left(y_i-a-b\overline{x}_i-c\overline{x}_i^2\right)^2.$$

Derivando parzialmente rispetto ai parametri e uguagliando a zero le derivate ottenute, si ha:

$$\frac{\partial R}{\partial a}=2\sum_{i=1}^{n}\left(y_i-a-b\overline{x}_i-c\overline{x}_i^2\right)(-1)=0$$

$$\frac{\partial R}{\partial b}=2\sum_{i=1}^{n}\left(y_i-a-b\overline{x}_i-c\overline{x}_i^2\right)\left(-\overline{x}_i\right)=0$$

$$\frac{\partial R}{\partial c}=2\sum_{i=1}^{n}\left(y_i-a-b\overline{x}_i-c\overline{x}_i^2\right)\left(-\overline{x}_i^2\right)=0$$

Si risolve dunque il sistema, ad esempio mediante il metodo di sostituzione o di Cramer:

$$\begin{cases} \sum_{i=1}^{n} y_i = na + b\sum_{i=1}^{n} \overline{x}_i + c\sum_{i=1}^{n} \overline{x}_i^2 \\ \sum_{i=1}^{n} \overline{x}_i y_i = a\sum_{i=1}^{n} \overline{x}_i + b\sum_{i=1}^{n} \overline{x}_i^2 + c\sum_{i=1}^{n} \overline{x}_i^3 \\ \sum_{i=1}^{n} \overline{x}_i^2 y_i = a\sum_{i=1}^{n} \overline{x}_i^2 + b\sum_{i=1}^{n} \overline{x}_i^3 + c\sum_{i=1}^{n} \overline{x}_i^4 \end{cases}$$

Per una delle proprietà della media aritmetica (cfr. par. 3.4) è $\sum_i \overline{x}_i = 0$. Inoltre, se i valori x_i costituiscono una progressione aritmetica, gli scarti dalla media con esponente dispari sono tutti nulli.

Supponiamo di aver osservato i seguenti valori:

y	*x*
1	0
2	1
4	2
7	3
14	**6**

Se non consideriamo gli scarti dalla media, occorre determinare la seguente tabella :

xy	x^2	x^2y	x^3	x^4
0	0	0	0	0
2	1	2	1	1
8	4	16	8	16
21	9	63	27	81
31	**14**	**81**	**36**	**98**

e risolvere il sistema:

$$\begin{cases} 14 = 4a + 6b + 14c \\ 31 = 6a + 14b + 36c \\ 81 = 14a + 36b + 98c \end{cases}$$

Utilizzando il metodo di Cramer si ha:

$$\Delta_a = \begin{vmatrix} 14 & 6 & 14 \\ 31 & 14 & 36 \\ 81 & 36 & 98 \end{vmatrix} = 80 \qquad \Delta_b = \begin{vmatrix} 4 & 14 & 14 \\ 6 & 31 & 36 \\ 14 & 81 & 98 \end{vmatrix} = 40$$

$$\Delta_c = \begin{vmatrix} 4 & 6 & 14 \\ 6 & 14 & 31 \\ 14 & 36 & 81 \end{vmatrix} = 40 \qquad \Delta = \begin{vmatrix} 4 & 6 & 14 \\ 6 & 14 & 36 \\ 14 & 36 & 98 \end{vmatrix} = 80$$

$$\hat{a} = \frac{\Delta_a}{\Delta} = \frac{80}{80} = 1 \qquad \hat{b} = \frac{\Delta_b}{\Delta} = \frac{40}{80} = 0{,}5 \qquad \hat{c} = \frac{\Delta_c}{\Delta} = \frac{40}{80} = 0{,}5$$

Se, invece, consideriamo gli scarti dalla media, occorre determinare la seguente tabella:

$\bar{x}y$	$\bar{x}^2$	$\bar{x}^2 y$	$\bar{x}^4$
-1,5	2,25	2,25	5,0625
-1	0,25	0,5	0,0625
2	0,25	1	0,0625
10,5	2,25	15,75	5,0625
10	**5**	**19,5**	**10,25**

e risolvere il sistema:

$$\begin{cases} 14 = 4a + 5c \\ 10 = 5b \\ 19{,}5 = 5a + 10{,}25c \end{cases}$$

9.3 La regressione multipla

Supponiamo di aver rilevato, su ciascuna unità statistica, k variabili indipendenti $X_1, X_2, \ldots, X_k$, ad esempio altezza, peso, circonferenza torace, ecc…

Si parla, in questo caso, di "regressione multipla".

Quando rileviamo $k+1$ variabili su n soggetti/oggetti, non disponiamo più di una serie doppia di valori (x_i, y_i), ma di un vettore di osservazioni per la variabile dipendente, l'età ad esempio, e di una matrice $n \times k$ di osservazioni, relative alle variabili indipendenti:

y_1	x_{11}	x_{12}	…	…	…	x_{1k}
y_2	x_{21}	x_{22}	…	.	…	x_{2k}
.	.	.		.		.
.	.	.		.		.
y_i	.	.	…	x_{ij}	…	.
.	.	.		.		.
.	.	.		.		.
y_n	x_{n1}	x_{n2}	…	…	…	x_{nk}

L'elemento generico x_{ij} rappresenta il valore della *j-ma* variabile indipendente rilevata sull'*i-mo* soggetto/oggetto.

Se ipotizziamo che la relazione teorica che lega la variabile Y alle altre sia di tipo lineare

$$Y=f(x_1, x_2, \ldots, x_k)=a_0+a_1x_1+a_2x_2+\ldots+a_kx_k,$$

otteniamo il "**modello di regressione lineare multipla**", che da un punto di vista grafico è rappresentato da un iperpiano nello spazio a ($k+1$) dimensioni.

Quando $k=2$, $f(x_1,x_2)=a_0+a_1x_1+a_2x_2$ rappresenta un piano nello spazio tridimensionale.

Tra gli infiniti piani che si ottengono al variare dei parametri a_0, a_1, a_2, l'obiettivo è individuare quello da cui, con maggiore probabilità, hanno avuto origine i dati osservati.

METODO DEI MINIMI QUADRATI

$$R=\sum_{i=1}^{n}\left(y_i-Y_i\right)^2=\sum_{i=1}^{n}\left(y_i-a_0-a_1x_{i1}-a_2x_{i2}\right)^2=\text{minimo.}$$

Derivando R rispetto ad a_0, a_1, a_2 e uguagliando a zero le derivate parziali, si dovrà risolvere il seguente sistema:

$$\begin{cases} \sum_{i=1}^{n} y_i = na_0 + a_1\sum_{i=1}^{n} x_{i1} + a_2\sum_{i=1}^{n} x_{i2} \\ \sum_{i=1}^{n} x_{i1}y_i = a_0\sum_{i=1}^{n} x_{i1} + a_1\sum_{i=1}^{n} x_{i1}^2 + a_2\sum_{i=1}^{n} x_{i1}x_{i2} \\ \sum_{i=1}^{n} x_{i2}y_i = a_0\sum_{i=1}^{n} x_{i2} + a_1\sum_{i=1}^{n} x_{i1}x_{i2} + a_2\sum_{i=1}^{n} x_{i2}^2 \end{cases}$$

Considerando gli scarti dalla media aritmetica $\bar{x}_i = x_i - M_x$, sarà:

$$R = \sum_{i=1}^{n} \left(y_i - Y_i\right)^2 = \sum_{i=1}^{n} \left(y_i - a_0 - a_1 \bar{x}_{i1} - a_2 \bar{x}_{i2}\right)^2 = \text{minimo.}$$

Derivando R rispetto ad a_0, a_1, a_2, uguagliando a zero le derivate parziali e considerando che

$$\sum_{i=1}^{n} \bar{x}_{ij} y_i = \sum_{i=1}^{n} \bar{x}_{ij} \bar{y}_i \qquad \text{per } j=1,\ 2,$$

si ricava $a_0 = M_y$ e si perviene al sistema:

$$\begin{cases} \sum_{i=1}^{n} \bar{x}_{i1} \bar{y}_i = a_1 \sum_{i=1}^{n} \bar{x}_{i1}^2 + a_2 \sum_{i=1}^{n} \bar{x}_{i1} \bar{x}_{i2} \\ \sum_{i=1}^{n} \bar{x}_{i2} \bar{y}_i = a_1 \sum_{i=1}^{n} \bar{x}_{i1} \bar{x}_{i2} + a_2 \sum_{i=1}^{n} \bar{x}_{i2}^2 \end{cases}$$

dove i termini noti rappresentano le codevianze fra la variabile dipendente e le variabili indipendenti, mentre i coefficienti dei parametri incogniti costituiscono la matrice di devianze e codevianze delle X_j ($j=1,\ 2$).

Come esempio supponiamo di aver rilevato la seguente matrice dei dati:

y	x_1	x_2
12	2	2,5
18	4	3
27	6	4,5
35	8	6
92	**20**	**16**

Per stimare i parametri del modello è conveniente costruire la seguente tabella:

x_1y	x_2y	x_1^2	x_2^2	x_1x_2
24	30	4	6,25	5
72	54	16	9	12
162	121,5	36	20,25	27
280	210	64	36	48
538	**415,5**	**120**	**71,5**	**92**

Se non si considerano scarti dalla media, occorre risolvere il sistema:

$$\begin{cases} 92 = 4a_0 + 20a_1 + 16a_2 \\ 538 = 20a_0 + 120a_1 + 92a_2 \\ 415{,}5 = 16a_0 + 92a_1 + 71{,}5a_2 \end{cases}$$

Ricorrendo al metodo di Cramer si ha:

$$\Delta_{a_0} = \begin{vmatrix} 92 & 20 & 16 \\ 538 & 120 & 92 \\ 415{,}5 & 92 & 71{,}5 \end{vmatrix} = 28 \qquad \Delta_{a_1} = \begin{vmatrix} 4 & 92 & 16 \\ 20 & 538 & 92 \\ 16 & 415{,}5 & 71{,}5 \end{vmatrix} = 60$$

$$\Delta_{a_2} = \begin{vmatrix} 4 & 20 & 92 \\ 20 & 120 & 538 \\ 16 & 92 & 415{,}5 \end{vmatrix} = 56 \qquad \Delta = \begin{vmatrix} 4 & 20 & 16 \\ 20 & 120 & 92 \\ 16 & 92 & 71{,}5 \end{vmatrix} = 24$$

da cui:

$$\hat{a}_0 = \frac{\Delta_{a_0}}{\Delta} = \frac{28}{24} = 1{,}17 \qquad \hat{a}_1 = \frac{\Delta_{a_1}}{\Delta} = \frac{60}{24} = 2{,}5 \qquad \hat{a}_2 = \frac{\Delta_{a_2}}{\Delta} = \frac{56}{24} = 2{,}3\,.$$