Capitolo 4

La regressione multipla

4.1 Introduzione

Nel caso della regressione lineare multipla si considerano p variabili indipendenti (o predittive) $X_1, \ldots, X_p$ ed una variabile dipendente Y. Il modello si scrive pertanto:

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon \,, \tag{4.1}$$

che viene chiamato *modello di regressione multipla* e descrive un iperpiano nello spazio $\mathbb{R}^{p+1}$; le variabili $X_1, \ldots, X_p$ vengono chiamati *regressori*. Da un punto di vista geometrico, l'equazione $\mathbb{E}(Y) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$ descrive un iperpiano. I coefficienti $\beta_0, \beta_1, \ldots, \beta_p$ sono i coefficienti regressione; il generico parametro β_j, ($j = 0, 1, \ldots, p$) fornisce la variazione media della variabile risposta Y in corrispondenza di una variazione unitaria del regressore X_j, supponendo costante il valore delle rimanenti variabili; per questo motivo i β_j ($j = 1, \ldots, p$) vengono anche chiamati *coefficienti di regressione parziale*.

Ponendo $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$ e $\mathbf{x} = (1, x_1, x_2, \ldots, x_p)'$, il modello (4.1) si può scrivere:

$$Y = \boldsymbol{\beta}' \mathbf{x} + \varepsilon \,.$$

Nota 4.1 Il modello (4.1) può essere anche utilizzato per studiare strutture più complesse. Consideriamo ad esempio il modello polinomiale:

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon$$

può essere ricondotto al modello (4.1) ponendo $x_1 = x$, $x_2 = x^2$ e $x_3 = x^3$, così che risulta:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

che è evidentemente un modello di regressione lineare. Anche modelli che includono effetti di interazione possono essere ricondotti al modello (4.1). Considerato:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$$

ponendo $x_3 = x_1 x_2$ possono essere scritti nella forma (4.1) come nel caso precedente. ♣

Assegnato l'insieme $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ e indicato con x_{ji} il valore della i-esima osservazione ($i = 1, \ldots, n$) nella j-esima variabile ($j = 1, \ldots, p$), possiamo scrivere:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + \varepsilon_i = \boldsymbol{\beta}' \mathbf{x}_i + \varepsilon_i \, .$$

Per le variabili di errore si assumono le stesse ipotesi (2.7), (2.8) e (2.9) viste in precedenza.

$$\mathcal{E}(\boldsymbol{\beta}) := \sum_{i=1}^{n} (y_i - f_i(\beta_0, \beta_1, \ldots, \beta_p))^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \cdots - \beta_p x_{pi})^2 \, . \tag{4.2}$$

La stima $\mathbf{b} = (b_0, b_0, \ldots, b_p)$ viene ottenuta in base al metodo dei minimi quadrati, cioè in modo tale che risulti minima la somma dei quadrati begli scarti:

$$\begin{aligned} \mathbf{b} &:= \underset{\beta_0, \beta_1, \ldots, \beta_p \in \mathbb{R}^{p+1}}{\arg} \min \mathcal{E}(\boldsymbol{\beta}) = \underset{\beta_0, \beta_1, \ldots, \beta_p \in \mathbb{R}^{p+1}}{\arg} \min \sum_{i=1}^{n} (y_i - f(\mathbf{x}_i))^2 \\ &= \underset{\beta_0, \beta_1, \ldots, \beta_p \in \mathbb{R}^{p+1}}{\arg} \min \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \cdots - \beta_p x_{pi})^2 \, . \end{aligned}$$

In questo caso il sistema di equazioni normali (2.13) si scrive:

$$\begin{aligned} \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{0i} - \cdots - b_p x_{pi}) &= 0 \\ \sum_{i=1}^{n} x_{1i} (y_i - b_0 - b_1 x_{1i} - \cdots - b_p x_{pi}) &= 0 \\ \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots &= 0 \\ \sum_{i=1}^{n} x_{pi} (y_i - b_0 - b_1 x_{1i} - \cdots - b_p x_{pi}) &= 0 \end{aligned}$$

da cui si ottiene

$$\begin{aligned} n b_0 + b_1 \sum_{i=1}^{n} x_{1i} + \cdots + b_p \sum_{i=1}^{n} x_{pi} &= \sum_{i=1}^{n} y_i \\ b_0 \sum_{i=1}^{n} x_{1i} + b_1 \sum_{i=1}^{n} x_{1i}^2 + b_2 \sum_{i=1}^{n} x_{2i} x_{2i} + \cdots + b_p \sum_{i=1}^{n} x_{1i} x_{pi} &= \sum_{i=1}^{n} y_i x_{1i} \\ \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots &= \sum_{i=1}^{n} y_i x_{ji} \\ b_0 \sum_{i=1}^{n} x_{pi} + b_1 \sum_{i=1}^{n} x_{1i} x_{pi} + b_2 \sum_{i=1}^{n} x_{2i} x_{pi} + \cdots + b_p \sum_{i=1}^{n} x_{pi}^2 &= \sum_{i=1}^{n} y_i x_{pi} \, . \end{aligned} \tag{4.3}$$

Dalla prima equazione delle (4.3) si ricava

$$b_0 = \bar{y} - b_1 \bar{x}_1 - \cdots - b_p \bar{x}_p$$

dopo aver ricavato $b_1, \ldots, b_p$ dalle restanti equazioni del sistema (4.3).

Infine indichiamo al solito con $e_i = y_i - f(x_i)$ $(i = 1, \ldots, n)$ i residui, e con σ_{je} la covarianza fra la j-esima variabile ed i residui. Con procedimenti analoghi al caso univariato, si dimostra che la media dei residui è nulla, cioè $\bar{e} = 0$, ed inoltre:

$$\sigma_{1e} = \sigma_{2e} = \cdots = \sigma_{pe} = 0 \ .$$

4.2 Un approccio matriciale

Assegnato un insieme di n osservazioni di $Y, X_1, \ldots, X_q$, in base al modello (4.1) possiamo scrivere:

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \ldots + \beta_p x_{1p} + \varepsilon_1 \\ Y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \ldots + \beta_p x_{2p} + \varepsilon_2 \\ \ldots &= \ldots \\ Y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \ldots + \beta_p x_{np} + \varepsilon_n \end{aligned} \tag{4.4}$$

Analogamente a quanto visto nel paragrafo 2.6.1, possiamo introdurre il vettore delle osservazioni $\mathbf{y}$, la matrice delle variabili indipendenti $\underset{\sim}{\mathbf{X}}$, il vettore dei parametri da stimare $\boldsymbol{\beta}$, il vettore degli errori $\boldsymbol{\varepsilon}$ ed il vettore n-dimensionale $\mathbf{1}$.

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \underset{\sim}{\mathbf{X}} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \tag{4.5}$$

$$\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}, \quad \mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \tag{4.6}$$

In generale, quindi, $\mathbf{Y}$, $\boldsymbol{\varepsilon}$ e $\mathbf{1}$ sono vettori a n dimensioni, $\underset{\sim}{\mathbf{X}}$ è una matrice di ordine $n \times (p+1)$ e $\boldsymbol{\beta}$ è un vettore a p dimensioni. In base a quanto scritto sopra, il sistema (4.4) in termini matriciali viene scritto in maniera formalmente analoga a quanto visto nella sezione precedente per la regressione semplice:

$$\mathbf{Y} = \underset{\sim}{\mathbf{X}}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

La stima $\mathbf{b}$ di $\boldsymbol{\beta}$ in base al metodo dei minimi quadrati si ottiene esattamente come in precedenza: si ottiene dapprima il sistema di equazioni normali:

$$\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}}\mathbf{b} = \underset{\sim}{\mathbf{X}}'\mathbf{y} \tag{4.7}$$

che conduce alla soluzione:

$$\mathbf{b} = (\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'\mathbf{y} \tag{4.8}$$

che risulta formalmente uguale alla (2.33), ove si assume che la matrice $\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}}$ sia invertibile e ciò accade se i regressori sono linearmente indipendenti.

Si noti che la (4.2) può essere scritta come:

$$\begin{aligned}\mathcal{E}(\boldsymbol{\beta}) &= (\mathbf{y} - \underset{\sim}{\mathbf{X}}\boldsymbol{\beta})'(\mathbf{y} - \underset{\sim}{\mathbf{X}}\boldsymbol{\beta}) = \mathbf{y}'\mathbf{y} - \boldsymbol{\beta}'\underset{\sim}{\mathbf{X}}'\mathbf{y} - \mathbf{y}'\underset{\sim}{\mathbf{X}}\boldsymbol{\beta} + \boldsymbol{\beta}'\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}}\boldsymbol{\beta} \\ &= \mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\underset{\sim}{\mathbf{X}}'\mathbf{y} + \boldsymbol{\beta}'\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}}\boldsymbol{\beta}\end{aligned}$$

da cui

$$\underset{\beta_0,\beta_1,\ldots,\beta_p\in\mathbb{R}^{p+1}}{\arg\min}\,\mathcal{E}(\boldsymbol{\beta}) = \underset{\beta_0,\beta_1,\ldots,\beta_p\in\mathbb{R}^{p+1}}{\arg\min}\{\mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\underset{\sim}{\mathbf{X}}'\mathbf{y} + \boldsymbol{\beta}'\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}}\boldsymbol{\beta}\}$$

cioè **b** è la soluzione di:

$$\frac{\partial\mathcal{E}(\mathbf{b})}{\partial\mathbf{b}} = \mathbf{0} \qquad \text{cioè} \quad \frac{\partial(\mathbf{y}'\mathbf{y} - 2\mathbf{b}'\underset{\sim}{\mathbf{X}}'\mathbf{y} + \mathbf{b}'\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}}\mathbf{b})}{\partial\mathbf{b}} = -2\underset{\sim}{\mathbf{X}}'\mathbf{y} + 2\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}}\mathbf{b} = \mathbf{0}$$

da cui

$$\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}}\mathbf{b} = \underset{\sim}{\mathbf{X}}'\mathbf{y} \qquad \text{e quindi} \qquad \mathbf{b} = (\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'\mathbf{y}$$

che ovviamente coincide con la (4.8).

Una volta ricavato **b** possiamo calcolare la generica stima della risposta:

$$\hat{y}_i = \mathbf{x}_i'\mathbf{b} \qquad i = 1,\ldots,n$$

dove $\mathbf{x}_i = (1, x_{i1}, x_{i2}, \ldots, x_{ip})'$ e quindi il vettore delle stime $\hat{\mathbf{y}} = (\hat{y}_1, \ldots, \hat{y}_n)'$ che, in base alla (4.8), è dato da:

$$\hat{\mathbf{y}} = \underset{\sim}{\mathbf{X}}\mathbf{b} = \underset{\sim}{\mathbf{X}}(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'\mathbf{y} = \mathbf{P}\mathbf{y} \tag{4.9}$$

dove $\mathbf{P} = \underset{\sim}{\mathbf{X}}(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'$ è la matrice di stima (*hat matrix*) così chiamata in quanto associa al vettore dei valori empirici $\mathbf{y} = (y_1, \ldots, y_n)'$ il vettore delle stime della risposta $\hat{\mathbf{y}} = (\hat{y}_1, \ldots, \hat{y}_n)'$. Possiamo scrivere i residui in notazione matriciale, posto $\mathbf{e} = (e_1, \ldots, e_n)'$:

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{P}\mathbf{y} = (\mathbf{I} - \mathbf{P})\mathbf{y}\ . \tag{4.10}$$

4.2.1 Un'interpretazione geometrica dei minimi quadrati

Un'interpretazione geometrica dei minimi quadrati è abbastanza utile dal punto di vista intuitivo. Possiamo pensare al vettore di osservazioni $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$ come un vettore dall'origine al punto A nello spazio campionario $\mathcal{C}$ che è n-dimensionale, cioè $\mathcal{C} \subseteq \mathbb{R}^n$, dove $(y_1, y_2, \ldots, y_n)$ sono quindi le coordinate del punto A.

La matrice $\mathbf{X}$ ha dimensioni $n \times (p+1)$, e quindi presenta $p+1$ vettori colonna n dimensionali: $\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p$ ciascuno dei quali definisce un vettore dall'origine nello spazio campionario. Questi $p+1$ vettori formano uno spazio $(p+1)$-dimensionale detto *spazio di stima*, e che verrà denotato con $\mathcal{S}$, dove $\mathcal{S} \subseteq \mathbb{R}^{p+1}$, con $p+1 < n$; pertanto $\mathcal{S} \subset \mathcal{C}$. Pertanto possiamo rappresentare ciascun punto in $\mathcal{S}$ come combinazione lineare di $\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p$. Pertanto qualunque punto nello spazio di stima ha la forma di $\mathbf{X}\boldsymbol{\beta}$. Sia B il punto in $\mathcal{S}$ individuato da $\mathbf{X}\boldsymbol{\beta}$. La distanza quadratica fra A e B è allora data da:

$$d(A,B) = (\mathbf{y} - \underset{\sim}{\mathbf{X}}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\ .$$

La distanza $d(A, B)$, per $B \in \mathcal{S}$ è minima in corrispondenza del punto $C \in \mathcal{S}$ definito da $\hat{\mathbf{y}} = \underset{\sim}{\mathbf{X}}\mathbf{b}$. Pertanto, poichè $\mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \underset{\sim}{\mathbf{X}}\mathbf{b}$ è perpendicolare allo spazio di stima $\mathcal{S}$, risulta allora:

$$\mathbf{X}'(\mathbf{y} - \underset{\sim}{\mathbf{X}}\mathbf{b}) = \mathbf{0} \qquad \text{cioè} \quad \underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}}\mathbf{b} = \underset{\sim}{\mathbf{X}}'\mathbf{y}$$

che costituiscono le equazioni normali (4.7).

4.3 Proprietà delle stime dei minimi quadrati

In analogia al caso univariato, indichiamo con $\mathbf{B}$ lo stimatore di $\boldsymbol{\beta}$. Le proprietà statistiche degli stimatori $\mathbf{B}$ possono essere dimostrate facilmente Consideriamo dapprima la distorsione. Dalla (4.8), scrivendo il vettore aleatorio $\mathbf{Y}$ anzichè le osservazioni $\mathbf{y}$ si ha:

$$\begin{aligned}
\mathbb{E}(\mathbf{B}) &= \mathbb{E}[(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'\mathbf{Y}] = \mathbb{E}[(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'(\underset{\sim}{\mathbf{X}}\boldsymbol{\beta} + \boldsymbol{\varepsilon})] \\
&= \mathbb{E}[(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}}\boldsymbol{\beta}] + \mathbb{E}[(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'\boldsymbol{\varepsilon}] \\
&= \mathbb{E}[\boldsymbol{\beta}] + (\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'\,\mathbb{E}[\boldsymbol{\varepsilon}] \\
&= \boldsymbol{\beta}
\end{aligned}$$

in quanto $\mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ e $(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}} = \mathbf{I}_{p+1}$. Pertanto $\mathbf{B}$ è uno stimatore non distorto di $\boldsymbol{\beta}$.

Le proprietà della variabilità di $\mathbf{B}$ sono espresse dalla matrice di varianze e covarianze:

$$\text{Cov}(\mathbf{B}) = \mathbb{E}\{(\mathbf{B} - \mathbb{E}(\mathbf{B}))(\mathbf{B} - \mathbb{E}(\mathbf{B}))'\} = \mathbb{E}\{(\mathbf{B} - \boldsymbol{\beta})(\mathbf{B} - \boldsymbol{\beta})'\} = \mathbb{E}(\mathbf{B}\mathbf{B}') - \boldsymbol{\beta}\boldsymbol{\beta}' \,.$$

Tenendo presente che la matrice $\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}}$ è simmetrica[1] si ha:

$$\begin{aligned}
\mathbb{E}(\mathbf{B}\mathbf{B}') &= \mathbb{E}\{(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'\mathbf{Y}[(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'\mathbf{Y}]'\} = \\
&= \mathbb{E}\{(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'\mathbf{Y}\mathbf{Y}'\underset{\sim}{\mathbf{X}}(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\} = \mathbb{E}\{(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'(\underset{\sim}{\mathbf{X}}\boldsymbol{\beta} + \boldsymbol{\varepsilon})(\underset{\sim}{\mathbf{X}}\boldsymbol{\beta} + \boldsymbol{\varepsilon})'\underset{\sim}{\mathbf{X}}(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\} \\
&= \mathbb{E}\{[(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}}\boldsymbol{\beta} + (\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'\boldsymbol{\varepsilon}][\boldsymbol{\beta}'\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}}(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1} + \boldsymbol{\varepsilon}'\underset{\sim}{\mathbf{X}}(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}]\} \\
&= \mathbb{E}\{[\boldsymbol{\beta} + (\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'\boldsymbol{\varepsilon}][\boldsymbol{\beta}' + \boldsymbol{\varepsilon}'\underset{\sim}{\mathbf{X}}(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}]\} \\
&= \mathbb{E}\{(\boldsymbol{\beta}\boldsymbol{\beta}') + (\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'\boldsymbol{\varepsilon}\boldsymbol{\beta}' + \boldsymbol{\beta}\boldsymbol{\varepsilon}'\underset{\sim}{\mathbf{X}}(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1} + (\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\underset{\sim}{\mathbf{X}}(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\} \\
&= \mathbb{E}(\boldsymbol{\beta}\boldsymbol{\beta}') + \mathbb{E}\{(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'\boldsymbol{\varepsilon}\boldsymbol{\beta}'\} + \mathbb{E}\{\boldsymbol{\beta}\boldsymbol{\varepsilon}'\underset{\sim}{\mathbf{X}}(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\} + \mathbb{E}\{(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\underset{\sim}{\mathbf{X}}(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\} \\
&= \mathbb{E}(\boldsymbol{\beta}\boldsymbol{\beta}') + (\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'\mathbb{E}(\boldsymbol{\varepsilon})\boldsymbol{\beta}' + \boldsymbol{\beta}\mathbb{E}(\boldsymbol{\varepsilon}')\underset{\sim}{\mathbf{X}}(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1} + (\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')\underset{\sim}{\mathbf{X}}(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1} \\
&= \boldsymbol{\beta}\boldsymbol{\beta}' + (\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')\underset{\sim}{\mathbf{X}}(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}
\end{aligned}$$

tenendo conto che $\mathbb{E}(\boldsymbol{\beta}\boldsymbol{\beta}') = \boldsymbol{\beta}\boldsymbol{\beta}'$ (esssendo $\boldsymbol{\beta}$ costante) e che $\mathbb{E}(\boldsymbol{\beta}) = \mathbf{0}$. Essendo $\boldsymbol{\varepsilon}$ un vettore n-dimensionale, la matrice $\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')$ è una matrice diagonale avente i valori $\mathbb{E}(\varepsilon_h^2)$ ($h = 1, \ldots, n$) sulla diagonale e pertanto essendo:

$$\mathbb{E}(\varepsilon_h^2) = \text{Var}(\varepsilon_h) = \sigma^2 \qquad h = 1, \ldots, n$$

[1] Infatti, se $\mathbf{A}$ è una matrice $n \times p$, allora $\mathbf{A}'\mathbf{A}$ è una matrice quadrata simmetrica di dimensioni $p \times p$.

segue

$$\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2 \mathbf{I}_n$$

dove $\mathbf{I}_n$ è la matrice identità n-dimensionale, e quindi, essendo σ^2 una quantit 'a scalare, segue:

$$\begin{aligned}\mathbb{E}(\mathbf{B}\mathbf{B}') &= \boldsymbol{\beta}\boldsymbol{\beta}' + (\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'\sigma^2\mathbf{I}_n\underset{\sim}{\mathbf{X}}(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1} \\ &= \boldsymbol{\beta}\boldsymbol{\beta}' + \sigma^2(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}}(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1} \\ &= \boldsymbol{\beta}\boldsymbol{\beta}' + \sigma^2(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1} .\end{aligned}$$

Si ottiene pertanto:

$$\mathrm{Cov}(\mathbf{B}) = \mathbb{E}(\mathbf{B}\mathbf{B}') - \boldsymbol{\beta}\boldsymbol{\beta}' = \boldsymbol{\beta}\boldsymbol{\beta}' + \sigma^2(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1} - \boldsymbol{\beta}\boldsymbol{\beta}' = \sigma^2(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}$$

che è una matrice simmetrica $p \times p$ il cui j-esimo elemento sulla diagonale è uguale alla varianza di B_j e l'elemento ij è uguale alla covarianza fra B_i e B_j. Posto $\mathbf{C} = (\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}$, allora la varianza di B_j è uguale a $\sigma^2 C_{jj}$ e la covarianza fra B_i e B_j è uguale a $\sigma^2 C_{ij}$. In particolare lo stimatore $\mathbf{B}$ è il migliore stimatore non distorto di $\boldsymbol{\beta}$ (teorema di Gauss Markov).

4.3.1 Stima di σ^2

Come nel caso della regressione semplice, possiamo costruire uno stimatore di σ^2:

$$\mathrm{SS}_e = \sum_{i=1}^{n}(y_i - f(x_i))^2 = \sum_{i=1}^{n} e_i^2 = \mathbf{e}'\mathbf{e} .$$

Posto $\mathbf{e} = \mathbf{y} - \underset{\sim}{\mathbf{X}}\mathbf{b}$, segue:

$$\begin{aligned}\mathrm{SS}_e &= (\mathbf{y} - \underset{\sim}{\mathbf{X}}\mathbf{b})'(\mathbf{y} - \underset{\sim}{\mathbf{X}}\mathbf{b}) = \mathbf{y}'\mathbf{y} - \mathbf{b}'\underset{\sim}{\mathbf{X}}'\mathbf{y} - \mathbf{y}\underset{\sim}{\mathbf{X}}\mathbf{b} + \mathbf{b}'\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}}\mathbf{b} \\ &= \mathbf{y}'\mathbf{y} - 2\mathbf{b}'\underset{\sim}{\mathbf{X}}'\mathbf{y} + \mathbf{b}'\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}}\mathbf{b} .\end{aligned}$$

Essendo $\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}}\mathbf{b} = \underset{\sim}{\mathbf{X}}'\mathbf{y}$, quest'ultima relazione si scrive:

$$\mathrm{SS}_e = \mathbf{y}'\mathbf{y} - \mathbf{b}'\underset{\sim}{\mathbf{X}}'\mathbf{y}.$$

Poichè $\mathrm{SS}_y = \mathbf{y}'\mathbf{y}$, dalla precedente relazione segue:

$$\mathrm{SS}_y = \mathrm{SS}_f + \mathrm{SS}_e$$

cioè la (2.22) nel caso multiplo, dove

$$\mathrm{SS}_f := \mathbf{b}'\underset{\sim}{\mathbf{X}}'\mathbf{y}$$

La somma dei quadrati dei residui presenta $n - (p+1)$ gradi di libertà in quanto vi sono p parametri che vengono stimati dal modello. Segue quindi:

$$s^2 = \frac{\mathrm{SS}_e}{n-(p+1)} = \frac{\mathrm{SS}_e}{n-p-1} . \tag{4.11}$$

Poichè si può dimostrare che il valore atteso della quantità ora ricavataè uguale a σ^2, ne segue che S^2 è uno stimatore non distorto di σ^2. Inoltre, come rilevato nel caso semplice, lo stimatore di σ^2 dipende dal modello.

i	Y	X_1	X_2
1	10	2	1
2	17	3	2
3	48	4	5
4	27	1	2
5	55	5	6
6	26	6	4
7	9	7	3
8	16	8	4

Tabella 4.1:

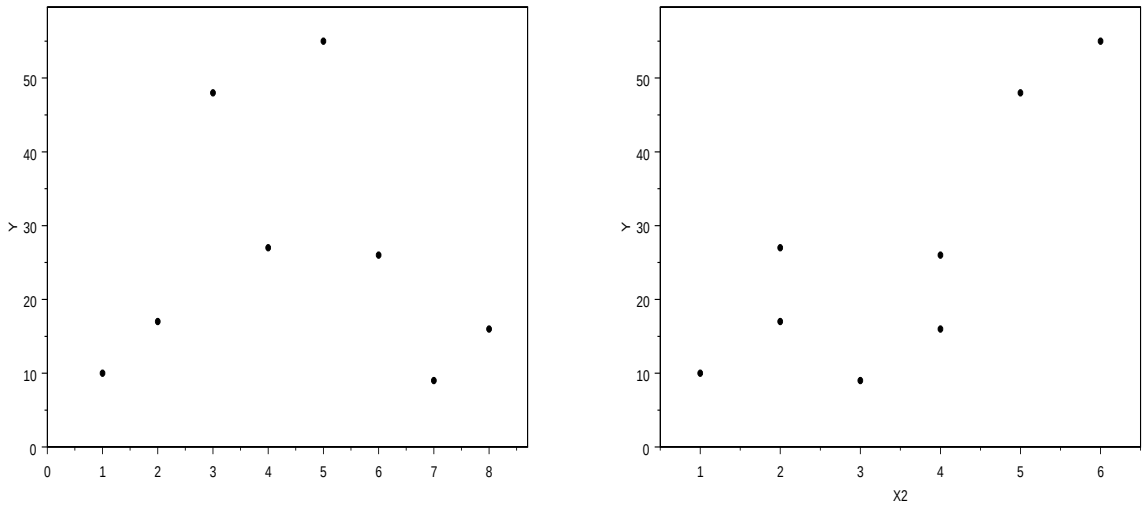

Figura 4.1: *Diagrammi a dispersione di Y rispetto a X_1 e di Y rispetto a X_2 per i dati in Tabella 4.1.*

4.3.2 Inadeguatezza dei diagrammi a dispersione nella regressione multipla

Abbiamo visto in precedenza che, nel caso della regressione semplice, il diagramma a nube di punti costituisce un importante strumento nell'analisi esplorativa della relazione fra una variabile dipendente Y ed una variabile esplicativa X. Nel caso della regressione multipla invece, l'esame dei diagrammi di Y rispetto a $X_1, \ldots, X_p$ non sempre, in generale, fornisce indicazioni sulla relazione fra la variabile risposta e le variabili esplicative.

Per illustrare l'inadeguatezza del diagramma a dispersione consideriamo un problema con due regressori sulla base dei dati in Tabella 4.1. In Figura 4.1 riportiamo i diagrammi a dispersione di Y rispetto a X_1 e di Y rispetto a X_2. I dati sono stati generati a partire dall'equazione

$$f(x) = 8 - 5x_1 + 12x_2 \ .$$

Il diagramma di Y rispetto a X_1 non evidenzia alcuna apparente relazione fra le due variabili; il

diagramma di Y rispetto a X_2 suggerisce una relazione lineare con pendenza uguale approssimativamente a 8. Si noti che entrambi i diagrammi a dispersione comunicano informazioni errate. Poichè i dati contengono due coppie di punti per le quali la variabile X_2 assume lo stesso valore – rispettivamente $X_2 = 2$ per $i = 2, 4$ e $X_2 = 4$ per $i = 6, 8$ – ed in questo caso possiamo misurare l'effetto di X_1 per X_2 fissato per entrambe le coppie. Si ottiene quindi:

$$b_{y1.2} = \frac{17 - 27}{3 - 1} = -5 \quad \text{per } X_2 = 2 \quad (i = 2, 4)$$

$$b_{y1.2} = \frac{26 - 16}{6 - 8} = -5 \quad \text{per } X_2 = 4 \quad (i = 6, 8)$$

che forniscono, in entrambi i casi, il valore corretto. Noto b_1 si potrebbe studiare l'effetto di X_2; in generale però questa procedura non è utile poichè non sempre vi sono insiemi di dati che contengono coppie di punti di questo tipo.

Questo esempio mostra come la costruzione di diagrammi a dispersione di Y rispetto a $X_1, \ldots, X_p$ può risultare fuorviante persino nel caso in cui si lavora con due regressori ed in assenza di quantità di errore. Ovviamente situazioni più realistiche con un numero di regressori superiore a 2 ed in presenza di quantità di errore possono risultare maggiormente fuorvianti.

4.4 Trasformazioni lineari

In alcuni casi, anzichè lavorare direttamente sulla matrice dei dati $\underset{\sim}{\mathbf{X}}$, è meglio costruire un modello di regressione operando preliminarmente un'opportuna trasformazione lineare dei dati. Le trasformazioni lineari costituiscono uno dei principali strumenti dell'analisi di dati multidimensionali. Al fine dello studio di un fenomeno, in molti casi risulta più utile prendere in considerazione ed analizzare solo alcune combinazioni lineari delle variabili piuttosto che tutte le variabili originali, ciò poichè spesso si riduce la dimensione dei dati. Le trasformazioni lineari inoltre possono semplificare la struttura della matrice di varianze e covarianze, rendendo più semplice l'interpretazione dei dati.

Sia $\mathbf{a} = (a_1, \ldots, a_p) \in \mathbb{R}^p$ e consideriamo una combinazione lineare di $\mathbf{x}_h = (x_{h1}, \ldots, x_{hp}) \in \mathbb{R}^p$:

$$u_h = a_1 x_{h1} + \cdots + a_p x_{hp} = \mathbf{a}'\mathbf{x}_h \qquad h = 1, \ldots, n$$

dove $a_1, \ldots, a_p$ sono assegnati. In base alle (1.7) e (1.9) la media e la varianza dei y_h sono date rispettivamente da:

$$\bar{u} = \frac{1}{n}\sum_{h=1}^{n} u_h = \frac{1}{n}\mathbf{a}'\sum_{h=1}^{n}\mathbf{x}_h = \mathbf{a}'\bar{\mathbf{x}}$$
$$s_u^2 = \frac{1}{n}\sum_{h=1}^{n}(u_h - \bar{u})^2 = \frac{1}{n}\sum_{h=1}^{n}\mathbf{a}'(\mathbf{x}_h - \bar{\mathbf{x}})(\mathbf{x}_h - \bar{\mathbf{x}})'\mathbf{a} = \mathbf{a}'\mathbf{S}_x\mathbf{a}\ ,$$

dove $\mathbf{S}_x = \frac{1}{n}\sum_{h=1}^{n}(\mathbf{x}_h - \bar{\mathbf{x}})(\mathbf{x}_h - \bar{\mathbf{x}})'$ è la matrice $p \times p$ di varianze e covarianze di $\mathbf{X}$. In generale possiamo considerare trasformazioni lineari $\mathbb{R}^p \to \mathbb{R}^q$ (con $q \leq p$) del tipo:

$$\mathbf{u}_h = \mathbf{A}\mathbf{x}_h + \mathbf{b} \qquad h = 1, \ldots, n$$

che può essere scritta

$$\underset{\sim}{\mathbf{U}} = \underset{\sim}{\mathbf{X}}\mathbf{A}' + \mathbf{1}\mathbf{b}' \tag{4.12}$$

dove $\underset{\sim}{\mathbf{U}}$ è una matrice $n \times q$, $\mathbf{A}$ è una matrice $(q \times p)$ e $\mathbf{b}$ è un vettore q-dimensionale.

Il vettore delle medie e la matrice di varianze e covarianze $\mathbf{S}_u$ (di dimensione $q \times q$) di $\mathbf{U}$ sono allora dati da:

$$\bar{\mathbf{u}} = \mathbf{A}\bar{\mathbf{x}} + \mathbf{b} \tag{4.13}$$

$$\begin{aligned}\mathbf{S}_u &= \frac{1}{n}\sum_{h=1}^{n}(\mathbf{u}_h - \bar{\mathbf{u}})(\mathbf{u}_h - \bar{\mathbf{u}})' \\ &= \frac{1}{n}\sum_{h=1}^{n}[\mathbf{A}\mathbf{x}_h + \mathbf{b} - (\mathbf{A}\bar{\mathbf{x}} + \mathbf{b})][\mathbf{A}\mathbf{x}_h + \mathbf{b} - (\mathbf{A}\bar{\mathbf{x}} + \mathbf{b})]' \\ &= \mathbf{A}\mathbf{S}_x\mathbf{A}' \ . \end{aligned} \tag{4.14}$$

In particolare se $q = p$ e $\mathbf{A}$ è non singolare[2] allora:

$$\mathbf{S}_x = \mathbf{A}^{-1}\mathbf{S}_u(\mathbf{A}')^{-1} \ .$$

Esistono numerose trasformazioni lineari di interesse in statistica, qui di seguito ne vedremo alcuni. Per semplicità tutte le trasformazioni lineari sono centrate intorno allo $\mathbf{0}$.

4.4.1 La trasformazione di scala

Consideriamo la trasformazione $\mathbb{R}^p \to \mathbb{R}^p$ data da:

$$z_{hi} := \frac{x_{hi} - \bar{x}_i}{s_i} \quad h = 1, \ldots, n \quad i = 1, \ldots, p \tag{4.15}$$

dove $s_i = \frac{1}{n-1}\sum_{h=1}^{n}(x_i - \bar{x}_i)^2$. La (4.15) opera un cambiamento di scala in modo tale che ciascuna variabile abbia varianza unitaria e pertanto si elimina l'arbitrarietà nella scelta della

[2] **Definizione.** Una matrice quadrata $\mathbf{A}$ si dice *non singolare* se $|\mathbf{A}| \neq 0$; altrimenti se $|\mathbf{A}| = 0$ la matrice $\mathbf{A}$ si dice singolare.

Alcune proprietà delle matrici non singolari. Sia $\mathbf{A}$ una matrice quadrata non singolare e $\alpha \in \mathbb{R}$. Allora si ha:

(i) $\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}(A_{ij})'$.

(ii) $(\alpha\mathbf{A})^{-1} = \frac{1}{\alpha}\mathbf{A}^{-1}$.

(iii) $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

(iv) L'unica soluzione di $\mathbf{Ax} = \mathbf{b}$ `e $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.

(v) $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$.

(vi) $|\mathbf{A}^{-1}| = |\mathbf{A}|^{-1}$.

scala. Le quantità z_{hi} vengono chiamate *scarti standardizzati*. Considerata la matrice diagonale:

$$\mathbf{D} = \begin{pmatrix} s_1 & 0 & 0 & \cdots & 0 \\ 0 & s_2 & 0 & \cdots & 0 \\ 0 & 0 & s_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & s_p \end{pmatrix} = \mathrm{diag}(s_i)\,,$$

cioè la matrice avente i valori s_i ($i = 1, \ldots, p$) nella diagonale principale e zero altrove, possiamo riscrivere la relazione precedente (4.15) in termini vettoriali come segue:

$$\mathbf{z}_h = \mathbf{D}^{-1}(\mathbf{x}_h - \bar{\mathbf{x}}) \quad h = 1, \ldots, n\,. \tag{4.16}$$

Possiamo scrivere le (4.16) in forma compatto utilizzando il calcolo matriciale. Si ha infatti dalla (4.16)

$$\mathbf{z}_h = \mathbf{D}^{-1}\mathbf{x}_h - \mathbf{D}^{-1}\bar{\mathbf{x}}$$

e posto $\underset{\sim}{\mathbf{Z}}' = (\mathbf{z}_1, \ldots, \mathbf{z}_n)$, in base alla (4.12) e tenendo conto che $\bar{\mathbf{x}} = \frac{1}{n}\mathbf{X}'\mathbf{1}$, si ha:

$$\begin{aligned} \underset{\sim}{\mathbf{Z}} &= \underset{\sim}{\mathbf{X}}\mathbf{D}^{-1} - \mathbf{1}(\mathbf{D}^{-1}\bar{\mathbf{x}})' = \underset{\sim}{\mathbf{X}}\mathbf{D}^{-1} - \mathbf{1}(\mathbf{D}^{-1}\frac{1}{n}\mathbf{X}'\mathbf{1})' \\ &= \underset{\sim}{\mathbf{X}}\mathbf{D}^{-1} - \frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{X}\mathbf{D}^{-1} = (\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}')\underset{\sim}{\mathbf{X}}\mathbf{D}^{-1} \\ &= \mathbf{H}\underset{\sim}{\mathbf{X}}\mathbf{D}^{-1}\,. \end{aligned} \tag{4.17}$$

dove $\mathbf{H} = \mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}'$ è la matrice centrante introdotta in (1.11). Da un'applicazione di tale risultato – poichè $\mathbf{H1} = \mathbf{0}$[3] – segue:

$$\begin{aligned} \bar{\mathbf{z}} &= \frac{1}{n}\underset{\sim}{\mathbf{Z}}'\mathbf{1} = \frac{1}{n}\mathbf{D}^{-1}\underset{\sim}{\mathbf{X}}'\mathbf{H}\mathbf{1} = \mathbf{0} \\ \mathbf{S}_z &= \frac{1}{n}\underset{\sim}{\mathbf{Z}}'\mathbf{H}\underset{\sim}{\mathbf{Z}} = \frac{1}{n}\mathbf{D}^{-1}\underset{\sim}{\mathbf{X}}'\mathbf{H}'\mathbf{H}\mathbf{H}\underset{\sim}{\mathbf{X}}\mathbf{D}^{-1} = \frac{1}{n}\mathbf{D}^{-1}\underset{\sim}{\mathbf{X}}'\mathbf{H}\underset{\sim}{\mathbf{X}}\mathbf{D}^{-1} = \mathbf{D}^{-1}\mathbf{S}_x\mathbf{D}^{-1} = \mathbf{R}_x\,, \end{aligned} \tag{4.18}$$

essendo $\mathbf{H}$ una matrice idempotente, vedi (1.13).

Si noti che, ovviamente, gli stessi risultati si ottengono considerando le varianze corrette e quindi dividendo le devianze per $n - 1$ anzichè per n.

4.4.2 La trasformazione di Mahalanobis

Se $\mathbf{S}_x$ è definita positiva, cioè $\mathbf{S}_x > 0$ allora $\mathbf{S}_x^{-1}$ ha un'unica radice quadrata che viene indicata $\mathbf{S}_x^{-1/2}$. La trasformazione di Mahalanobis è $\mathbb{R}^p \to \mathbb{R}^p$ è definita da:

$$\mathbf{u}_r \;=\; \mathbf{S}_x^{-1/2}(\mathbf{x}_r - \bar{\mathbf{x}}) \qquad\qquad h = 1, \ldots, n\,. \tag{4.19}$$

[3] Un'altra propriet`a della matrice centrante `e che risulta $\mathbf{H1} = \mathbf{0}$, infatti

$$\mathbf{H1} = (\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}')\mathbf{1} = \mathbf{1} - \frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{1} = \mathbf{1} - \frac{1}{n}\mathbf{1}\cdot n = \mathbf{1} - \mathbf{1} = \mathbf{0}\,,$$

essendo $\mathbf{1}'\mathbf{1} = n$.

Analogamente a quanto fatto in precedenza, in forma matriciale, la (4.19) si può scrivere:

$$\underset{\sim}{\mathbf{U}} \;=\; \mathbf{H}\underset{\sim}{\mathbf{X}}\mathbf{S}_x^{-1/2}\,, \tag{4.20}$$

dove $\underset{\sim}{\mathbf{U}}' = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$. Segue poi, con ragionamenti analoghi a quelli visti in precedenza:

$$\bar{\mathbf{u}} = \frac{1}{n}\underset{\sim}{\mathbf{U}}'\mathbf{1} = \frac{1}{n}\mathbf{S}^{-1/2}\underset{\sim}{\mathbf{X}}\mathbf{H}'\mathbf{1} = \mathbf{0}$$
$$\mathbf{S}_u = \frac{1}{n}\underset{\sim}{\mathbf{U}}'\mathbf{H}\underset{\sim}{\mathbf{U}} = \frac{1}{n}\mathbf{S}_x^{-1/2}\underset{\sim}{\mathbf{X}}'\mathbf{H}'\mathbf{H}\mathbf{H}\underset{\sim}{\mathbf{X}}\mathbf{S}_x^{-1/2} = \frac{1}{n}\mathbf{S}_x^{-1/2}\underset{\sim}{\mathbf{X}}'\mathbf{H}\underset{\sim}{\mathbf{X}}\mathbf{S}_x^{-1/2} = \mathbf{S}_x^{-1/2}\mathbf{S}_x\mathbf{S}_x^{-1/2} = \mathbf{I}_p\,.$$

Pertanto tale trasformazione elimina la correlazione fra le variabili standardizza la varianza di ciascuna variabile.

Si noti che, anche in questo, gli stessi risultati si ottengono considerando le varianze corrette e quindi dividendo le devianze per $n-1$ anzichè per n.

4.5 Coefficienti di regressione standardizzati

Il coefficienti di regressione b_j ($j = 1, \ldots, n$) fornisce la variazione media della variabile risposta Y in corrispondenza di una variazione unitaria del regressore X_j. Usualmente è difficile però confrontare i coefficienti di regressione poichè la grandezza di b_j dipende dall'unità di misura del regressone X_j. Consideriamo ad esempio il modello $Y = f(x_1, x_2) + \varepsilon$ con :

$$f(x_1, x_2) = b_0 + b_1 x + b_2 x_2 = 5 + x_1 + 1000 x_2$$

dove y è misurato in litri, X_1 in millilitri e X_2 in litri. Notiamo che, benchè il valore di b_2 (=1000) sia considerevolmente più grande di quello di b_1 (=1), l'effetto di entrambi sulla risposta è evidentemente identico poichè la variazione di un litro in X_1 o in X_2, quando l'altro regressore è mantenuto costante, produce la stessa variazione in Y.

Generalmente, le unità di misura dei coefficienti di regressione β_j sono date in termini di unità di Y/unità di X_j. Per questa ragione, risulta a volte utile lavorare con trasformazioni di scala dei regressori che conducono a coefficienti di regressione adimensionali.

Una trasformazione che usualmente viene considerata a riguardo è la trasformazione di scala (4.15):

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \qquad i = 1, \ldots, n, \quad j = 1, \ldots, p$$
$$y_i^* = \frac{y_i - \bar{y}}{s_y} \qquad i = 1, \ldots, n$$

dove in queston caso s_j^2 è la varianza corretta del regressore X_j e s_y^2 è la varianza corretta di Y:

$$s_j^2 = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1} \qquad s_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}\,.$$

Tali trasformazioni conducono a regressori e risposte aventi media uguale a zero e varianza unitaria, ed il modello di regressione diventa:

$$Y_i^* = \beta_1 z_{i1} + \beta_2 z_{i2} + \cdots + \beta_p z_{ip} + \varepsilon_i \qquad i = 1, \ldots, n\,,$$

in particolare il modello risulta centrato nell'origine ($b_0 = f(0) = 0$). La stima dei minimi quadrati **b** di $\boldsymbol{\beta}$ è data da:

$$\mathbf{b}^* = (\underset{\sim}{\mathbf{Z}}'\underset{\sim}{\mathbf{Z}})^{-1}\underset{\sim}{\mathbf{Z}}'\mathbf{y}^* \ .$$

Un'altra trasformazione di scala conduce a variabili aventi devianze unitarie:

$$w_{ij} = \frac{x_{ij} - \bar{x}_j}{\mathrm{SS}_j^{1/2}} \qquad \text{e} \qquad y_i^0 = \frac{y_i - \bar{y}}{\mathrm{SS}_y^{1/2}} \tag{4.21}$$

dove SS_j è la devianza del regressore X_j, cioè $\mathrm{SS}_j = \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 = \sqrt{n-1}s_j$. Ovviamente segue $w_{ij} = z_{ij}/\sqrt{n-1}$ e $y_i^0 = y_i^*/\sqrt{n-1}$ quindi in generale $\underset{\sim}{\mathbf{W}} = \underset{\sim}{\mathbf{Z}}/\sqrt{n-1}$ e $\mathbf{y}^0 = \mathbf{y}^*/\sqrt{n-1}$.

In questa trasformazione, ciascun nuovo regressore W_j ha media $\bar{w}_j = 0$ e devianza unitaria $\sum_{i=1}^n (w_{ij} - \bar{w}_j)^2 = 1$. In termini di questi variabili, il modello di regressione è:

$$Y_i^0 = \beta_1 w_{i1} + \beta_2 w_{i2} + \cdots + \beta_p w_{ip} + \varepsilon_i \qquad i = 1, \ldots, n \ ,$$

e la stima dei minimi quadrati è:

$$\mathbf{b}^0 = (\underset{\sim}{\mathbf{W}}'\underset{\sim}{\mathbf{W}})^{-1}\underset{\sim}{\mathbf{W}}'\mathbf{y}^0 \ . \tag{4.22}$$

Tenendo conto delle relazioni prima ricavate $\underset{\sim}{\mathbf{W}} = \underset{\sim}{\mathbf{Z}}/\sqrt{n-1}$ e $\mathbf{y}^0 = \mathbf{y}^*/\sqrt{n-1}$, ne segue che $\mathbf{b}^* = \mathbf{b}^0$, infatti

$$\begin{aligned}
\mathbf{b}^0 = (\underset{\sim}{\mathbf{W}}'\underset{\sim}{\mathbf{W}})^{-1}\underset{\sim}{\mathbf{W}}'\mathbf{y}^0 &= \left(\frac{\underset{\sim}{\mathbf{Z}}'}{\sqrt{n-1}}\frac{\underset{\sim}{\mathbf{Z}}}{\sqrt{n-1}}\right)^{-1}\frac{\underset{\sim}{\mathbf{Z}}'}{\sqrt{n-1}}\frac{\mathbf{y}^*}{\sqrt{n-1}} \\
&= (n-1)(\underset{\sim}{\mathbf{Z}}'\underset{\sim}{\mathbf{Z}})^{-1}\frac{\underset{\sim}{\mathbf{Z}}'\mathbf{y}^*}{n-1} = (\underset{\sim}{\mathbf{Z}}'\underset{\sim}{\mathbf{Z}})^{-1}\underset{\sim}{\mathbf{Z}}'\mathbf{y}^* = \mathbf{b}^*
\end{aligned}$$

Inoltre la matrice di covarianza di $\underset{\sim}{\mathbf{W}}'\underset{\sim}{\mathbf{W}}$ ha la forma di una matrice di correlazione. Infatti essendo $\underset{\sim}{\mathbf{W}} = \underset{\sim}{\mathbf{Z}}/\sqrt{n-1}$, in maniera simile alla (4.18) segue:

$$\underset{\sim}{\mathbf{W}}'\underset{\sim}{\mathbf{W}} = \frac{1}{n-1}\underset{\sim}{\mathbf{Z}}'\underset{\sim}{\mathbf{Z}} = \frac{1}{n-1}\mathbf{D}^{-1}\underset{\sim}{\mathbf{X}}'\mathbf{H}'\mathbf{H}\underset{\sim}{\mathbf{X}}\mathbf{D}^{-1} = \frac{1}{n-1}\mathbf{D}^{-1}\underset{\sim}{\mathbf{X}}'\mathbf{H}\underset{\sim}{\mathbf{X}}\mathbf{D}^{-1} = \mathbf{D}^{-1}\mathbf{S}_x\mathbf{D}^{-1} = \mathbf{R}_x \ ,$$

I coefficienti di regressione $\mathbf{b}^* = \mathbf{b}^0$ ora ricavati vengono usualmente denominati *coefficienti standardizzati di regressione*. Si può dimostrare che la relazione fra i coefficienti di regressione calcolati a partire dalle variabili originarie $\underset{\sim}{\mathbf{X}}$ e quelli basati sulle variebili standardizzati $\underset{\sim}{\mathbf{Z}}$ è:

$$\begin{aligned}
b_j &= b_j^* \left(\frac{\mathrm{SS}_y}{\mathrm{SS}_j}\right)^{1/2}, \qquad j = 1, \ldots, p \\
b_0 &= \bar{y} - \sum_{j=1}^p b_j \bar{x}_j \ .
\end{aligned}$$

Si noti infine che molti software statistici riportano in uscita sia i coefficienti originari di regressione che quelli standardizzati, che vengono spesso denominati come "coefficienti beta". Nell'interpretazione dei coefficienti standardizzati, bisogna ricordare che sono sempre coefficienti di regressione parziali (cioè b_j misura l'effetto di X_j in presenza degli altri regressori X_i, con $i \neq j$ nel modello); inoltre b_j^* dipendono dall'intervallo di variazione delle variabili di regressione. Ne segue che può essere molto fuorviante utilizzare l'ampiezza dei b_j^* come una misura dell'importanza relativa del regressore X_j.

4.6 Valutazione della bontà del modello

La valutazione della bontà del modello costituisce un importante aspetto della costruzione di un modello regressione multipla. Molte delle tecniche a tal scopo utilizzate, sono estensioni di quelle utilizzate nella regressione lineare semplice.

4.6.1 Il coefficiente di determinazione multiplo

Analogamente al caso della regressione lineare semplice, consideriamo un modello $y_i = f(\mathbf{x}_i) + \varepsilon_i$ dove $f(\mathbf{x}_1) = b_0 + b_1 x_{1i} + \cdots + b_p x_{pi}$, la devianza totale si scompone nella devianza di regressione e nella devianza dell'errore o residua:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - f(x_i))^2 + \sum_{i=1}^{n}(f(x_i) - \bar{y})^2 \tag{4.23}$$

o, equivalentemente:

$$\mathrm{SS}_y = \mathrm{SS}_e + \mathrm{SS}_f \ . \tag{4.24}$$

L'indice:

$$R^2 \ = \ \frac{\mathrm{SS}_f}{\mathrm{SS}_y} \ = \ 1 - \frac{\mathrm{SS}_e}{\mathrm{SS}_y} \tag{4.25}$$

prende il nome di *coefficiente di determinazione multiplo* e rappresenta la parte della devianza totale di Y che è spiegata (o determinata) dalla supposta relazione lineare con X_1 e X_2. La radice quadrata di R^2 viene chiamata *coefficiente di correlazione multipla*.

- $R^2 = 1$: quando gli scarti e_i sono tutti nulli, cioè quando $y_i = f(\mathbf{x}_i)$ per ogni $i = 1, \ldots, n$.
- $R^2 = 0$: quanto i punti sono massimamente dispersi rispetto al piano di regressione, cioè quando la relazione lineare attraverso cui ci si propone di interpretare il comportamento del fenomeno Y in funzione di $\underset{\sim}{\mathbf{X}}$, non fornisce alcun utile informazione al riguardo.

Anche in questo caso, la devianza di Y presenta $n - 1$ gradi di libertà in quanto le quantità $y_1 - \bar{y}, y_2 - \bar{y}, \ldots, y_n - \bar{y}$ soddisfano il vincolo $\sum_i (y_i - \bar{y}) = 0$. La devianza di errore presenta $n - p - 1$ gradi di libertà. I $p + 1$ gradi di libertà che vengono sottratti a n riflettono il fatto che i residui possono essere calcolati dal modello di regressione che richiede la stima di $p + 1$ parametri $\beta_0, \beta_1, \ldots, \beta_p$. Per sottrazione la devianza di regressione $\sum_{i=1}^{n}(f(x_i) - \bar{y})^2$ presenta p gradi di libertà.

La relazione (4.24) in termini di gradi di libertà viene scritta come:

$$n - 1 \ = \ p + (n - p - 1) \ . \tag{4.26}$$

Dalle (4.24) e (4.26) possiamo costruire la seguente tabella dell'*analisi della varianza*:

Sorgente di variazione	*Gradi di libertà*	*Devianza*
regressione	p	$\mathrm{SS}_f = \sum_{i=1}^{n}(f(x_i) - \bar{y})^2$
residui	$n - p - 1$	$\mathrm{SS}_e = \sum_{i=1}^{n}(y_i - f(x_i))^2$
totale	$n - 1$	$\mathrm{SS}_y = \sum_{i=1}^{n}(y_i - \bar{y})^2$

Il coefficiente di determinazione multiplo corretto Il valore di R^2 aumenta sempre (o comunque non decresce) quando si aggiunge un nuovo regressore al modello; ciò non significa che necessariamente il nuovo modello risulta superiore al precedente. Per questa ragione, nella regressione multipla si considera il coefficiente R^2 *corretto* (*adjusted* $\bar{R}^2$) che penalizza l'inclusione di variabili non necessarie nel nostro modello. Tale coefficiente viene definito sostituendo nella (4.25) i valori di SS_e e SS_y con i corrispondenti valori divisi per i rispettivi gradi di libertà:

$$\bar{R}^2 = 1 - \frac{\mathrm{SS}_e/(n-p-1)}{\mathrm{SS}_y/(n-1)} = 1 - \frac{n-1}{n-p-1}(1-R^2)\,. \tag{4.27}$$

Se i valori di R^2 e $\bar{R}^2$ sono molto differenti, allora verosimilmente il modello risulta sovradimensionato, cioè sono stati inclusi termini che non constribuiscono significativamente all'adattamento dei dati.

4.6.2 Analisi dei residui

Come nel caso della regressione semplice, anche per la regressione multipla l'analisi dei residui costituisce un elemento importante per la valutazione del modello. In questo caso è interessante analizzare i grafici dei residui rispetto a ciascun regressore X_i, ($i = 1, \ldots, p$). Nel caso multidimensionale si considerano altre tecniche di analisi dei residui che qui brevemente vengono accennate.

Residui parziali. I grafici basati sui *residui parziali* consentono di evidenziare maggiormente la relazione fra i residui ed il regressone X_i ($i = 1, \ldots, p$). Si definisce *residuo parziale* per il regressore X_i la quantità:

$$\begin{aligned} e^*_{ih} &:= y_h - b_1 x_{h1} - \cdots - b_{i-1} x_{h,i-1} - b_{i+1} x_{h,i+1} - \cdots - b_p x_{hp} \\ &= e_i + b_i x_{hi} \quad h = 1, \ldots, n\,. \end{aligned} \tag{4.28}$$

Il grafico dei e^*_{ih} rispetto a x_{ih} viene chiamato *grafico dei residui parziali*. Analogamente all'usuale grafico dei residui di e_h rispetto a x_{hi}, il grafico dei residui parziali risulta utile nelle rilevazione dei valori anomali e di condizioni di eteroschedasticità. In base alla (4.28) si evidenzia subito che il grafico dei residui parziali e^*_{ih} rispetto a x_{hi} avrà una pendenza pari a b_i anzichè pari a zero come nell'usuale grafico dei residui. Ciò consente di valutare semplicemente l'ammontare della divergenza in presenza di valori anomali e di eteroschedasticità.

In particolare se la relazione fra Y e X_i è non lineare, allora il grafico dei residui parziali fornisce utili indicazioni su opportune trasformazioni dei dati al fine di poter ottenere una relazione lineare fra la risposta ed i dati trasformati.

Grafici del regressore X_h rispetto al regressore X_j. Questo tipo di grafici può essere utile per studiare la relazione fra due variabili di regressione. In particolare, se due regressori sono fortemente correlati fra di loro, si dice che i dati presentano *multicollinearità*. Fenomeni di multicollinearità possono alterare fortemente le stime dei minimi quadrati e, in alcuni casi, possono rendere inutile il modello di regressione.

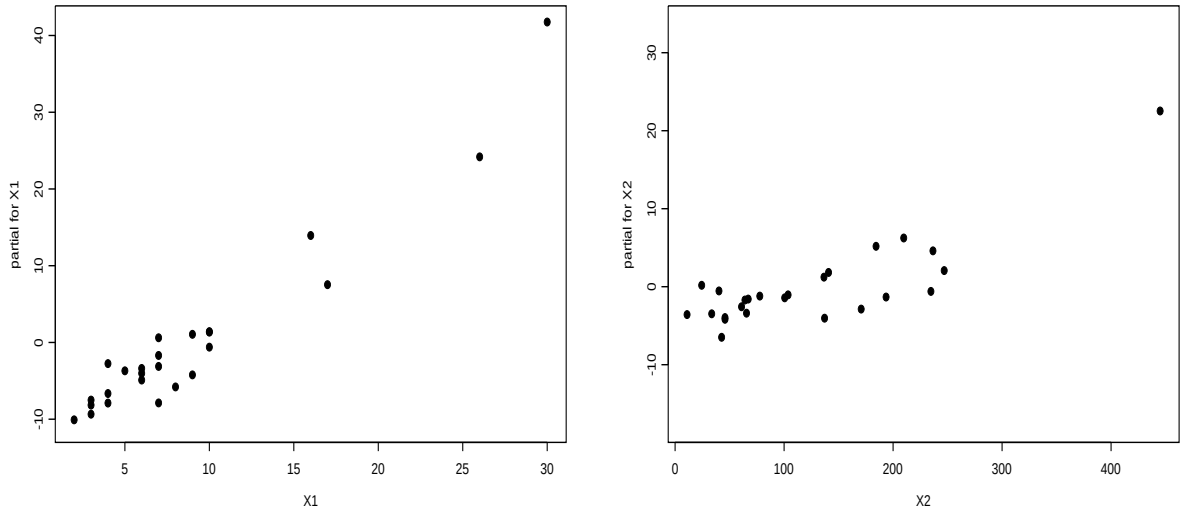

Figura 4.2: *Diagrammi a dei residui parziali per i dati in Tabella 3.4 in accordo al modello di regressione* $\hat{y} = 2.3412 + 1.6159x_1 + 0.0472x_2$.

4.6.3 Correlazione fra i residui

In generale, in un modello di regressione in cui $p+1$ parametri vengono stimati da n osservazioni, i residui presentano $n-p-1$ gradi di libertà. Pertanto i residui non possono essere indipendenti e necessariamente esiste una correlazione fra di essi. In analogia a quanto visto nella sezione 2.7.1, sia $\mathbf{E} = \mathbf{Y} - \hat{\mathbf{Y}}$ il vettore differenza fra $\mathbf{Y}$ e la risposta $\hat{\mathbf{Y}}$ e si ha:

$$\mathbf{E} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \underset{\sim}{\mathbf{X}}\mathbf{b} = \mathbf{Y} - \underset{\sim}{\mathbf{X}}(\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}})^{-1}\underset{\sim}{\mathbf{X}}'\mathbf{Y} = (\mathbf{I}_n - \mathbf{P})\mathbf{Y}$$

da cui

$$\mathbb{E}(\mathbf{E}) = (\mathbf{I}_n - \mathbf{P})\mathbb{E}(\mathbf{Y}) = (\mathbf{I}_n - \mathbf{P})\underset{\sim}{\mathbf{X}}\boldsymbol{\beta}\,.$$

Poichè

$$\mathbf{E} - \mathbb{E}(\mathbf{E}) = (\mathbf{I}_n - \mathbf{P})\mathbf{Y} - (\mathbf{I}_n - \mathbf{P})\underset{\sim}{\mathbf{X}}\boldsymbol{\beta} = (\mathbf{I}_n - \mathbf{P})(\mathbf{Y} - \underset{\sim}{\mathbf{X}}\boldsymbol{\beta}) = (\mathbf{I}_n - \mathbf{P})\boldsymbol{\varepsilon}$$

allora la matrice di varianze e covarianze dei residui $\mathbf{E}$ è data da:

$$\text{Cov}(\mathbf{E}) = \mathbb{E}[(\mathbf{E} - \mathbb{E}(\mathbf{E}))(\mathbf{E} - \mathbb{E}(\mathbf{E}))'] = (\mathbf{I}_n - \mathbf{P})\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')(\mathbf{I}_n - \mathbf{P})'\,.$$

Si noti che $\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2\mathbf{I}_n$ e che la matrice $\mathbf{I}_n - \mathbf{P}$ è simmetrica; essendo inoltre $\mathbf{P}$ idempotente – cioè $\mathbf{P}^2 = \mathbf{P}$ – segue pertanto:

$$\text{Cov}(\mathbf{E}) = (\mathbf{I}_n - \mathbf{P})(\mathbf{I}_n - \mathbf{P})\sigma^2 = (\mathbf{I}_n - 2\mathbf{P} + \mathbf{P}^2)\sigma^2 = (\mathbf{I}_n - 2\mathbf{P} + \mathbf{P}^2)\sigma^2(\mathbf{I}_n - \mathbf{P})\sigma^2.$$

In particolare la varianza dell'h-esimo residuo E_h è data da:

$$\text{Var}(E_h) = \sigma^2(1 - p_{hh})$$

dove p_{hh} è l'i-esimo elemento diagonale di $\mathbf{P}$. Poichè $0 \leq p_{hh} \leq 1$, allora la quantità s_e^2 della varianza dei residui in realtà fornisce una sovrastima della varianza dei residui $\text{Var}(E_h)$; inoltre $\text{Var}(E_h)$ dipende dalla posizione di $\mathbf{x}_h$.

4.7 Multicollinearità

I modelli di regressione sono largamente utilizzati in vari campi applicativi. Un problema che può avere un impatto drammatico sull'effettivo utilizzo di un modello di regressione concerne la *multicollinearità*, che consiste nel fatto che alcune variabili di regressione presentano una forte interdipendenza lineare.

I regressori costituiscono le colonne della matrice $\underset{\sim}{\mathbf{X}}$, per cui se una di esse può essere espressa esattamente come combinazione lineare di una o più delle altre colonne di $\underset{\sim}{\mathbf{X}}$, ne segue che la matrice $\underset{\sim}{\mathbf{X}}'\underset{\sim}{\mathbf{X}}$ risulta singolare. La presenza di una forte interdipendenza lineare fra i regressori può fortemente alterare la capacità predittiva di un modello di regressione lineare multipla. Consideriamo il seguente semplice esempio:

X_1	X_2
5	20
10	20
5	30
10	30
5	20
10	20
5	30
10	30

Considerata la trasformazione (4.21) $\underset{\sim}{\mathbf{X}} \rightarrow \mathbf{W}$, allora la matrice $\mathbf{W}'\mathbf{W}$ sarà uguale alla matrice di correlazione **R**:

$$\mathbf{W}'\mathbf{W} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

e

$$(\mathbf{W}'\mathbf{W})^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Consideriamo l'Esempio 3.5in cui si ha:

$$\mathbf{W}'\mathbf{W} = \begin{pmatrix} 1 & 0,824215 \\ 0,824215 & 1 \end{pmatrix}$$

da cui segue:

$$(\mathbf{W}'\mathbf{W})^{-1} = \begin{pmatrix} 3,11841 & -2,57023 \\ -2,57023 & 3,11841 \end{pmatrix}$$

Se consideriamo le varianze standardizzate dei coefficienti di regressione B_1^0 e B_2^0, otteniamo:

$$\frac{\text{Var}(B_1^0)}{\sigma^2} = \frac{\text{Var}(B_2^0)}{\sigma^2} = 1$$

mentre per l'insieme di dati prima considerato, si ottiene:

$$\frac{\text{Var}(B_1^0)}{\sigma^2} = \frac{\text{Var}(B_2^0)}{\sigma^2} = 3,11841\ .$$

Nell'insieme dei dati dell'Esempio 3.5, le varianze di regressione sono *inflazionate* a causa della multicollinearità. Questo fenomeno risulta evidente dal valore degli elementi non appartenenti alla diagonale della matrice $\mathbf{W}'\mathbf{W}$. I valori appartenenti alla diagonale di $(\mathbf{W}'\mathbf{W})^{-1}$ forniscono una misura della dipendenza lineare fra i regressori e vengono spesso chiamati *fattori di inflazione della varianza* (*VIF, variance inflation factor*), per cui risulta in quel caso:

$$\text{VIF}_1 = \text{VIF}_2 = 3,11841$$

mentre per il caso ipotetico precedente risulta

$$\text{VIF}_1 = \text{VIF}_2 = 1$$

che implica che i due regressori X_1 e X_2 sono ortogonali. Si può dimostrare che, in generale, il VIF relativo al j-esimo coefficiente di regressione è uguale a :

$$\text{VIF}_j = \frac{1}{1 - R_j^2}$$

dove R_j^2 è il coefficiente di determinazione ottenuto costruendo un modello di regressione di X_j rispetto alle restanti variabili. Chiaramente se X_j presenta una rilevante dipendenza lineare dagli altri regressori, allora R_j^2 sarà prossimo a uno e quindi VIF_j assume un valore grande. Valori dell'indice VIF maggiori di 10 indicano fenomeni rilevanti di multicollinearità. Modelli di regressione che vengono adattati con il metodo dei minimi quadrati, in presenza di forti multicollinearità, sono poco affidabili come strumento di previsione; inoltre i coefficienti di regressione sono spesso molto sensibili ai dati nel campione in esame.

Capitolo 5

Cenni sulla regressione logistica

Il modello di regressione logistica si distingue dal modello di regressione considerato in precedenza in quanto la variabile risposta Y in questo caso è binaria o dicocotomica, assume cioè solo due valori $0, 1$.

Il problema è quindi quello in cui si desidera studiare la relazione fra una certa variabile quantitativa X e la presenza/assenza di un certo carattere; più in generale si intende studiare la relazione che intercorre fra p variabili quantitative $X_1, \ldots, X_p$ e una variabile risposta $Y \in \{0, 1\}$ di tipo dicotomico.

Un possibile approccio consiste nel costruire modelli in cui la variabile risposta è costituita dalla funzione di distribuzione della normale (analisi *probit*). Un secondo metodo consiste nel modellare la risposta utilizzando la funzione logistica

$$f(z) := \frac{\exp(z)}{1 + \exp(z)}$$

il cui grafico è riportato in Figura 5.1. Pertanto si considera il seguente modello:

$$Y = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} + \varepsilon = \psi(x) + \varepsilon \tag{5.1}$$

chiamato modello di *regressione logistica*, dove:

$$\psi(x) := \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}. \tag{5.2}$$

Come nel caso della regressione lineare, il modello (5.1) è costituito dalla somma di una componente strutturale e di una quantità di errore ε.

Il modello di regressione logistica pu ‘o essere visto anche come problema della discriminazione fra due classi Ω_0 e Ω_1, corrispondenti ai due valori della variabile risposta $Y \in \{0, 1\}$. In particolare Ω_0 e Ω_1 costituiscono una partizione di una certa popolazione statistica Ω, cioè $\Omega_0 \cup \Omega_1 = \Omega$ e quindi $\Omega_1 = \Omega \setminus \Omega_0 = \Omega_0^c$. Ad esempio, se $Y = 0$ e $Y = 1$ corrispondono rispettivamente all'assenza ed alla presenza di un certo carattere, allora Ω_0 è l'insieme delle unità statistiche di Ω in cui il carattere Y risulta assente ($Y = 0$) e Ω_1 è l'insieme delle unità statistiche di Ω in cui il carattere Y risulta presente ($Y = 1$). Consideriamo la probabilità a posteriori $P(\Omega_1|x) = P(Y = 1|x)$ che l'unità statistica ω, sulla base della propria caratteristica numerica x, venga assegnata nella classe Ω_1. Il problema che ci poniamo è quello di valutare

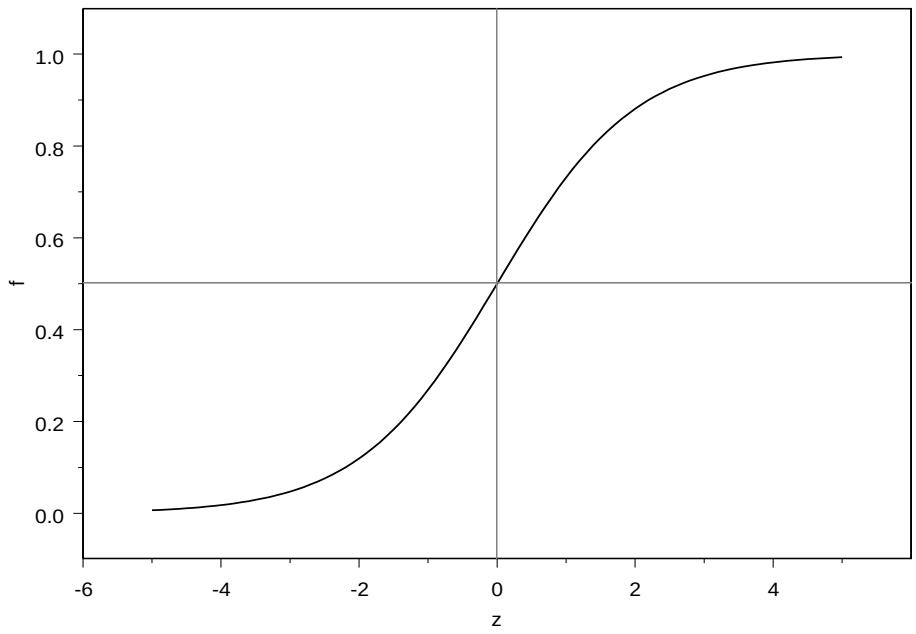

Figura 5.1: Diagramma della funzione logistica.

la probabilità $P(\Omega_1|x)$ mediante una funzione lineare di **x**; ovviamente si pone un problema in quanto una funzione lineare del tipo $P(\Omega_1|x) = \beta_0 + \beta_1 x$ non assume valori compresi fra 0 e 1.

Pertanto la funzione $\psi(x)$ introdotta in (5.2) viene vista come un modello per valutare la probabilità a posteriori $P(\Omega_1|x)$.

Nello studio della regressione logistica notevole importanza riveste la seguente trasformazione di $\psi(x)$:

$$g(x) := \ln\left(\frac{\psi(x)}{1-\psi(x)}\right) = \beta_0 + \beta_1 x \tag{5.3}$$

che viene chiamata *trasformazione logit*. L'importanza di questa trasformazione è che possiede molte delle proprietà del modello di regressione lineare visto in precedenza, in particolare la funzione logit $g(x)$ è lineare nei parametri, è continua ed a valori in tutto $\mathbb{R}$.

Nel modello di regressione (2.2) si assume usualmente che la componente di errore ε abbia distribuzione normale con media zero e varianza finita; nel caso qui in esame, ovviamente non è possibile fare una tale ipotesi.

Nel modello (5.1) la quantità può assumere solo due valori:

$$\begin{aligned} &\text{se } Y = 1 \text{ allora } \varepsilon = 1 - \psi(x) \text{ con probabilità } \psi(x) \\ &\text{se } Y = 0 \text{ allora } \varepsilon = -\psi(x) \text{ con probabilità } 1 - \psi(x) \end{aligned}$$

cioè $Y|x \sim B(1, \psi(x))$. Risulta inoltre:

$$\begin{aligned}
\mathbb{E}(\varepsilon) &= (1-\psi(x))\psi(x) + (-\psi(x))(1-\psi(x)) = 0 \\
\mathrm{Var}(\varepsilon) &= \mathbb{E}(\varepsilon^2) = (1-\psi(x))^2\psi(x) + (-\psi(x))^2(1-\psi(x)) \\
&= (1-2\psi(x)+\psi^2(x))\psi(x) + \psi^2(x)(1-\psi(x)) \\
&= \psi(x) - 2\psi^2(x) + \psi^3(x) + \psi^2(x) - \psi^3(x) \\
&= \psi(x) - \psi^2(x) = \psi(x)(1-\psi(x)).
\end{aligned}$$

Nel modello di regressione per variabili di risposta dicotomiche si ha quindi:

1. La speranza condizionale dell'equazione di regressione deve essere formulata in modo tale che assuma valori fra 0 e 1; in particolare il modello (5.2) soddisfa tale vincolo.

2. La distribuzione degli errori è binomiale anzichè normale come nel caso della regressione lineare.

Nel caso generale, si considera la dipendenza della variabile risposta da p regressori $X_1, X_2, \ldots, X_p$ e quindi la funzione:

$$\psi(\mathbf{x}) := \frac{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p)}. \tag{5.4}$$

e quindi il il logit del modello di regressione logistica è dato da:

$$g(\mathbf{x}) := \ln\left(\frac{\psi(\mathbf{x})}{1-\psi(\mathbf{x})}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p \tag{5.5}$$