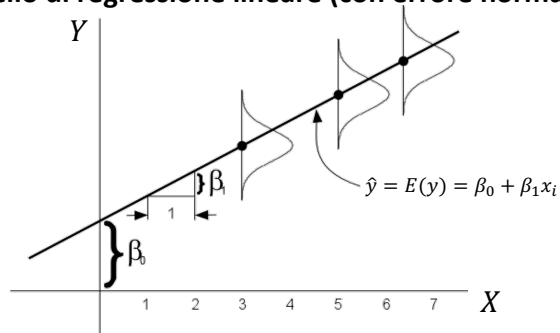


## Il modello di regressione lineare (con errore normale)



La figura fornisce una rappresentazione grafica del modello probabilistico di  $y = \beta_0 + \beta_1 x_i + \varepsilon_i$

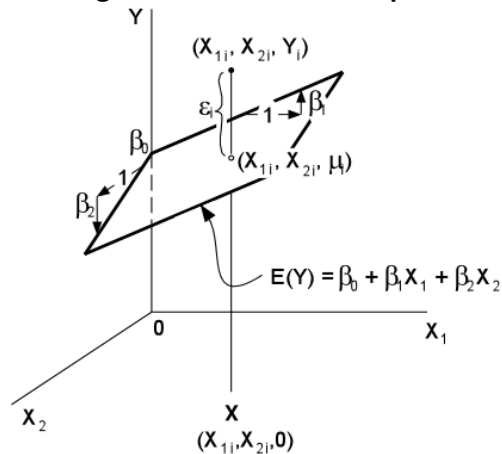
In corrispondenza di ogni valore di  $x$  c'è una popolazione di possibili valori  $y$ .

Questa popolazione ha media  $E(y)$  e varianza  $\sigma_y^2$ .

Per ciascuno dei possibili valori che  $X$  può assumere varierà solo  $E(y)$ , ma non la varianza  $\sigma_y^2$ .

Il modello probabilistico della regressione assume dunque che vi sia una diversa popolazione di risposte ( $y$ ) per ciascuno dei valori della variabile indipendente ( $x$ ), la cui media dipende dal valore della variabile indipendente ed è predetta dal modello della regressione.

## La regressione lineare multipla



- Nel modello della regressione multipla il valore atteso della variabile dipendente viene predetto da una combinazione lineare di  $j$  variabili indipendenti.
- Anziché trovare la retta che giunge il più vicino possibile a ciascuno di  $n$  punti posti in uno spazio bidimensionale, il problema diventa quello di trovare l'iperpiano  $j$ -dimensionale che giunge il più vicino possibile a  $n$  punti posti in uno spazio di dimensioni  $j+1$ .
- Nel caso di due variabili indipendenti, ad esempio, l'analisi della regressione multipla individua un piano che minimizza le distanze rispetto ad  $n$  punti posti in uno spazio tridimensionale.
- Per una singola osservazione  $y_i$ , il modello della regressione multipla è

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_j x_{ji} + \varepsilon_i$$

$$y_i = [1 \quad x_{1i} \quad x_{2i} \quad \dots \quad x_{ji}] \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_j \end{bmatrix} + \varepsilon_i$$

- Il coefficiente parziale della regressione multipla  $\beta_0$  denota il valore atteso della variabile dipendente quando tutte le variabili indipendenti sono uguali a zero.
- Il coefficiente  $\beta_j$  denota l'incremento atteso di  $Y$  in corrispondenza di un incremento unitario di  $X_j$ , quando tutte le altre variabili indipendenti sono tenute costanti,  $\varepsilon$  è una variabile aleatoria avente le stesse proprietà del termine d'errore del modello della regressione lineare bivariata.

- Per un campione di  $n$  osservazioni ci sono  $n$  equazioni come la precedente, che possono essere espresse in forma di matrice come:

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{\mathbf{y}}_{n \times 1} = \underbrace{\begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{j1} \\ 1 & x_{12} & x_{22} & \cdots & x_{j2} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{jn} \end{bmatrix}}_{\mathbf{X}}_{n \times (j+1)} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_j \end{bmatrix}}_{\boldsymbol{\beta}}_{(j+1) \times 1} + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}}_{\boldsymbol{\varepsilon}}_{n \times 1}$$

- Sviluppando il prodotto interno e la somma tra vettori

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 x_{11} + \beta_2 x_{21} + \cdots + \beta_j x_{j1} + \varepsilon_1 \\ \beta_0 + \beta_1 x_{12} + \beta_2 x_{22} + \cdots + \beta_j x_{j2} + \varepsilon_2 \\ \vdots \\ \beta_0 + \beta_1 x_{1n} + \beta_2 x_{2n} + \cdots + \beta_j x_{jn} + \varepsilon_n \end{bmatrix}$$

- In termini matriciali avremo che

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \text{Eq(18)}$$

in cui distinguiamo una componente *deterministica* (priva di errore), ricavata conoscendo il valore assunto dalle covariate:

$$E(\mathbf{y} | x_1, x_2, \dots, x_j) = E(\mathbf{y} | \mathbf{X}) = \mathbf{X}\boldsymbol{\beta} = \hat{\mathbf{y}} \quad \text{Eq(19)}$$

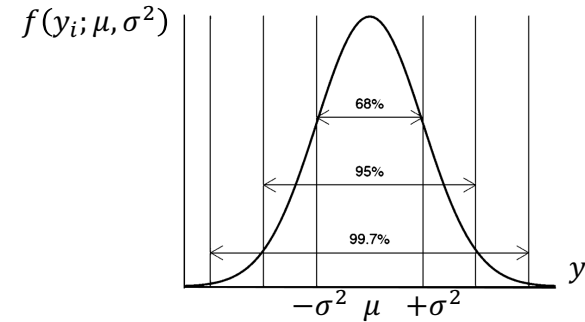
ed una *aleatoria*,

$$\boldsymbol{\varepsilon} = \mathbf{y} - \hat{\mathbf{y}},$$

con valore atteso nullo, (a) se sussiste associazione lineare tra le variabili (*linearità*) e (b) il modello è correttamente specificato:

$$E(\boldsymbol{\varepsilon}) = E(\mathbf{y} - \hat{\mathbf{y}}) = E(\mathbf{y} | x_1, x_2, \dots, x_j) - \hat{\mathbf{y}} = \mathbf{0}.$$

### Distribuzione normale, nessuna covariata.



dove  $E(\mathbf{y}) = \boldsymbol{\mu}$ , infatti

$$E(\mathbf{y}) = \hat{\mathbf{y}} = \begin{bmatrix} E(y_1) \\ E(y_2) \\ \vdots \\ E(y_n) \end{bmatrix} = \begin{bmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{bmatrix}$$

useremo quindi per ogni soggetto lo stesso valore previsto  $\mu$ :

$$f(y_i; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2} \times \frac{(y_i - \mu)^2}{\sigma^2}\right\}.$$

Rispetto al modello lineare espresso in forma matriciale, ciò equivale ad un effetto nullo di tutte le possibili covariate:

$$\hat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{j1} \\ 1 & x_{12} & x_{22} & \cdots & x_{j2} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{jn} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_0 \\ \vdots \\ \beta_0 \end{bmatrix}$$

dove  $\beta_0 = \mu$ , la stima di massima verosimiglianza di  $E(y_i)$ , la migliore stima possibile **in assenza di altre «informazioni»**.

**Distribuzione normale, una o più covariate (lineari).**

APPENDICE B

$$f(y_i; \mathbf{X}, \boldsymbol{\beta}, \sigma^2)$$

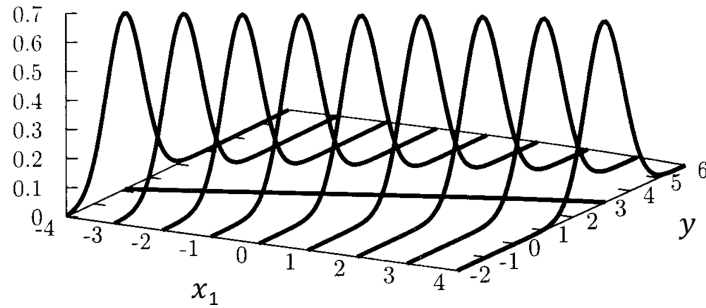


FIGURA B.3. Densità condizionali

$$E(\mathbf{y}|\mathbf{X}) = \hat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta},$$

In termini estesi,

$$\hat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{j1} \\ 1 & x_{12} & x_{22} & \cdots & x_{j2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{jn} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_j \end{bmatrix} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}.$$

Per l'osservazione  $y_i$  avremo che:

$$\begin{aligned} E(y_i) &= [\hat{\mathbf{y}}]_i = [\mathbf{X}\boldsymbol{\beta}]_i \\ &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_j x_{ji} \end{aligned}$$

Potremo quindi scrivere la funzione normale associata ad ogni osservazione  $y_i$ , come

$$f(y_i; \mathbf{X}, \boldsymbol{\beta}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2} \times \frac{(y_i - [\mathbf{X}\boldsymbol{\beta}]_i)^2}{\sigma^2}\right\} \quad \mathbf{Eq(20)}$$

- La funzione di probabilità congiunta delle  $n$  osservazioni  $y_i$  è:

$$P(\mathbf{y}) = \prod_{i=1}^n \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2} (y_i - [\mathbf{X}\boldsymbol{\beta}]_i)^2\right\}$$

- che possiamo riscrivere

$$\begin{aligned} P(\mathbf{y}) &= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - [\mathbf{X}\boldsymbol{\beta}]_i)^2\right\} \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n \varepsilon_i^2\right\}. \end{aligned}$$

- Inoltre, ponendo

$$\sum_{i=1}^n \varepsilon_i^2 = [\varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_n] \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \boldsymbol{\varepsilon}' \cdot \boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

- otteniamo la funzione di verosimiglianza

$$L(\boldsymbol{\beta}, \sigma^2; \mathbf{y}, \mathbf{X}) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\}.$$

Le stime di *massima verosimiglianza* di  $\boldsymbol{\beta}$  e  $\sigma^2$  renderanno *massimo* il valore della funzione di verosimiglianza.

- Per semplificare i calcoli consideriamo il logaritmo della funzione di verosimiglianza:

$$l(\boldsymbol{\beta}, \sigma^2; \mathbf{y}, \mathbf{X}) = \ln \left[ \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \right]$$

Riscriviamo la quantità in parentesi come

$$\left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} = (2\pi\sigma^2)^{-\frac{n}{2}} = (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}}$$

ottenendo

$$l(\boldsymbol{\beta}, \sigma^2; \mathbf{y}, \mathbf{X}) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

e troviamo (mediante software di calcolo simbolico) le derivate parziali per  $\boldsymbol{\beta}$  e  $\sigma^2$ :

- Matrix Calculus: <http://www.matrixcalculus.org/>

S. Laue, M. Mitterreiter, and J. Giesen, Computing Higher Order Derivatives of Matrix and Tensor Expressions, Advances in Neural Information Processing Systems. 31, 2750 (2018).

S. Laue, M. Mitterreiter, and J. Giesen, A simple and efficient tensor calculus, in AAAI, 2020.

derivative of  $-n/2 \cdot \log(2 \cdot \pi) - n/2 \cdot \log(\text{sigq}) - 1/(2 \cdot \text{sigq}) \cdot (\mathbf{Y} - \mathbf{X} \cdot \text{beta})' \cdot (\mathbf{Y} - \mathbf{X} \cdot \text{beta})$

w.r.t.  $\text{beta}$

$$\frac{\partial}{\partial \text{beta}} \left( (-n)/2 \cdot \log(2 \cdot \pi) - n/2 \cdot \log(\text{sigq}) - 1/(2 \cdot \text{sigq}) \cdot (\mathbf{Y} - \mathbf{X} \cdot \text{beta})' \cdot (\mathbf{Y} - \mathbf{X} \cdot \text{beta}) \right) = 1/\text{sigq} \cdot \mathbf{X}' \cdot (\mathbf{Y} - \mathbf{X} \cdot \text{beta})$$

where

X is a

Y is a

beta is a

n is a

pi is a

sigq is a

Export functions as

Common subexpressions

$$\frac{\delta}{\delta \boldsymbol{\beta}} l(\boldsymbol{\beta}, \sigma^2; \mathbf{y}, \mathbf{X}) = \frac{1}{\sigma^2} \mathbf{X}' (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) = \mathbf{0} \quad \text{Eq(21)}$$

Distribuendo il prodotto secondo la proprietà del prodotto di matrici,  $A(B + C) = AB + AC$ :

$$\frac{1}{\sigma^2} \mathbf{X}' \mathbf{y} - \frac{1}{\sigma^2} \mathbf{X}' \mathbf{X} \boldsymbol{\beta} = \mathbf{0}$$

$$\mathbf{X}' \mathbf{y} = \mathbf{X}' \mathbf{X} \boldsymbol{\beta}$$

$$(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{X} \boldsymbol{\beta}$$

$$(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y} = \mathbf{I} \boldsymbol{\beta} = \hat{\boldsymbol{\beta}} \quad \text{Eq(22)}$$

derivative of  $-n/2 \cdot \log(2 \cdot \pi) - n/2 \cdot \log(\text{sigq}) - 1/(2 \cdot \text{sigq}) \cdot (Y - X \cdot \text{beta})' \cdot (Y - X \cdot \text{beta})$   
w.r.t.  $\text{sigq}$

$$\frac{\partial}{\partial \text{sigq}} \left( (-n)/2 \cdot \log(2 \cdot \pi) - n/2 \cdot \log(\text{sigq}) - 1/(2 \cdot \text{sigq}) \cdot (Y - X \cdot \text{beta})' \cdot (Y - X \cdot \text{beta}) \right) = - (n/(2 \cdot \text{sigq}) - (t_0^T \cdot t_0)/(\text{sigq}^2 \cdot 2))$$

where  
 $t_0 = Y - X \cdot \text{beta}$

X is a    
Y is a    
beta is a    
n is a    
pi is a    
sigq is a

Export functions as     
Common subexpressions

$$= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = 0 \quad \text{Eq(23)}$$

$$-\frac{n}{2\sigma^2} = -\frac{1}{2(\sigma^2)^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Moltiplicando per  $-\frac{2(\sigma^2)^2}{n}$  si ottiene quindi la stima (distorta) di  $\sigma^2$ :

$$\hat{\sigma}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \frac{SQErr.}{n} \quad \text{Eq(24)}$$

Che andrà corretta per i gradi di libertà  $n - k - 1$ , dove  $n$  è la grandezza del campione,  $k$  sono le variabili indipendenti del modello di regressione, e  $-1$  si riferisce all'intercetta.

1. Ricavate con le formule di massima verosimiglianza **Eq(22)** e **Eq(24)** i coefficienti di regressione lineare e la stima della varianza costante  $\sigma^2$  (corretta per i gradi di libertà) per le seguenti variabili.

Y	X
3	2
2	3
2	1
7	4
6	5
7	8

2. Date le seguenti matrici,

$$\mathbf{y} = \begin{bmatrix} 3 \\ 2 \\ 2 \\ 7 \\ 6 \\ 7 \end{bmatrix}; \mathbf{X} = \begin{bmatrix} 1 & 2 & 2 & 3 \\ 1 & 3 & 2 & 7 \\ 1 & 1 & 1 & 2 \\ 1 & 4 & 5 & 6 \\ 1 & 5 & 9 & 8 \\ 1 & 8 & 2 & 9 \end{bmatrix}$$

- Stima con le soluzioni della massima verosimiglianza  $\boldsymbol{\beta}$  e  $\sigma^2$ , fornendo per quest'ultimo anche il valore corretto per i gradi di libertà.
- Per calcolare la varianza delle stime di massima verosimiglianza di cosa avreste bisogno? Provate a rispondere in modo esteso, rivedendo il materiale delle lezioni precedenti.

(Svolgete in R)

derivative of  w.r.t.

$$\frac{\partial}{\partial \beta} ((-1)/sigq \cdot X^T \cdot y - 1/sigq \cdot X^T \cdot X \cdot beta) = -1/sigq \cdot X^T \cdot X$$

where

X is a

beta is a

sigq is a

y is a

Export functions as

Common subexpressions  
 ON

Calcoliamo la derivata parziale di ordine secondo in  $\beta$  (derivata parziale della funzione derivata prima):

$$\frac{\delta^2 l(\beta, \sigma^2; \mathbf{y}, \mathbf{X})}{\delta \beta^2} = \frac{\delta \left( \frac{1}{\sigma^2} \mathbf{X}' \mathbf{y} - \frac{1}{\sigma^2} \mathbf{X}' \mathbf{X} \beta \right)}{\delta \beta} = -\frac{1}{\sigma^2} \mathbf{X}' \mathbf{X}$$

Sostituendo con la stima di massima verosimiglianza prima ottenuta,

$$\hat{\sigma}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X}\beta)' (\mathbf{y} - \mathbf{X}\beta) = \frac{SQErr}{n}$$

che rappresenta la stima della varianza costante (*omoschedasticità*) delle distribuzioni gaussiane di punteggi  $y_i$  attorno al valore previsto dal modello di regressione, ricaviamo la matrice  $(j+1) \times (j+1)$ :

$$\frac{\delta^2}{\delta \beta^2} l(\beta, \sigma^2; \mathbf{y}) = -\frac{n}{SQErr} \mathbf{X}' \mathbf{X} \quad Eq(25)$$

derivative of  w.r.t.

$$\frac{\partial}{\partial sigq} \left( \frac{-n}{2 \cdot sigq} + \frac{1}{2 \cdot sigq^2} \cdot (y - X \cdot beta)^T \cdot (y - X \cdot beta) \right) = \frac{n}{sigq^2 \cdot 2} - (t_0^T \cdot t_0) / sigq^3$$

where

$$t_0 = y - X \cdot beta$$

where

X is a

beta is a

n is a

sigq is a

y is a

Export functions as

Common subexpressions  
 ON

Calcoliamo la derivata parziale di ordine secondo in  $\sigma^2$  (derivata parziale della funzione derivata prima):

$$\begin{aligned} \frac{\delta^2 l(\beta, \sigma^2; \mathbf{y}, \mathbf{X})}{\delta (\sigma^2)^2} &= \frac{\delta \left( -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} (\mathbf{y} - \mathbf{X}\beta)' (\mathbf{y} - \mathbf{X}\beta) \right)}{\delta \sigma^2} \\ &= \frac{n}{2(\sigma^2)^2} - \frac{(\mathbf{y} - \mathbf{X}\beta)' (\mathbf{y} - \mathbf{X}\beta)}{(\sigma^2)^3} \quad Eq(26) \end{aligned}$$

Sostituendo con la stima di massima verosimiglianza  $n\hat{\sigma}^2 = (\mathbf{y} - \mathbf{X}\beta)' (\mathbf{y} - \mathbf{X}\beta)$ , ricaviamo lo scalare

$$\frac{\delta^2 l(\beta, \sigma^2; \mathbf{y}, \mathbf{X})}{\delta (\sigma^2)^2} = \frac{n}{2(\sigma^2)^2} - \frac{n\hat{\sigma}^2}{(\sigma^2)^3} = -\frac{n}{2(\sigma^2)^2} \quad Eq(27)$$

derivative of  $-n/(2 \cdot \text{sigq}) + 1/(2 \cdot \text{sigq}^2) \cdot (y - X \cdot \text{beta})' \cdot (y - X \cdot \text{beta})$  w.r.t.  $\text{beta}$

$$\frac{\partial}{\partial \text{beta}} \left( (-n)/(2 \cdot \text{sigq}) + 1/(2 \cdot \text{sigq}^2) \cdot (y - X \cdot \text{beta})' \cdot (y - X \cdot \text{beta}) \right) = -1/\text{sigq}^2 \cdot X^T \cdot (y - X \cdot \text{beta})$$

where

X is a

beta is a

n is a

sigq is a

y is a

Export functions as

Common subexpressions

Calcoliamo le due derivate cross-parziali di ordine secondo (derivata parziale in  $\beta$  della funzione derivata prima in  $\sigma^2$ ):

$$\frac{\delta^2 l(\beta, \sigma^2; \mathbf{y}, \mathbf{X})}{\delta \beta \sigma^2} = \frac{\delta \left( -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} (\mathbf{y} - \mathbf{X}\beta)' (\mathbf{y} - \mathbf{X}\beta) \right)}{\delta \beta}$$

$$= -\frac{1}{(\sigma^2)^2} \mathbf{X}' (\mathbf{y} - \mathbf{X}\beta) = -\frac{1}{(\sigma^2)^2} (\mathbf{X}' \mathbf{y} - \mathbf{X}' \mathbf{X} \beta) \quad \text{Eq(28)}$$

Sostituendo con la **stima di massima verosimiglianza**  $(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y} = \hat{\beta}$ , ricaviamo il vettore «nullo» (composto da valori zero, di ordine  $(j + 1) \times 1$ )

$$\frac{\delta^2 l(\beta, \sigma^2; \mathbf{y}, \mathbf{X})}{\delta \beta \sigma^2} = -\frac{1}{(\sigma^2)^2} (\mathbf{X}' \mathbf{y} - \mathbf{X}' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y})$$

$$= -\frac{1}{(\sigma^2)^2} (\mathbf{X}' \mathbf{y} - \mathbf{X}' \mathbf{y}) = \mathbf{0} \quad \text{Eq(29)}$$

derivative of  $1/\text{sigq} \cdot X^T \cdot (Y - X \cdot \text{beta})$  w.r.t.  $\text{sigq}$

$$\frac{\partial}{\partial \text{sigq}} (1/\text{sigq} \cdot X^T \cdot (Y - X \cdot \text{beta})) = -1/\text{sigq}^2 \cdot X^T \cdot (Y - X \cdot \text{beta})$$

where

X is a

Y is a

beta is a

sigq is a

Export functions as

Common subexpressions

(derivata parziale in  $\sigma^2$  della funzione derivata prima in  $\beta$ ):

$$\frac{\delta^2 l(\beta, \sigma^2; \mathbf{y}, \mathbf{X})}{\delta \sigma^2 \beta} = \frac{\delta}{\delta \sigma^2} \left( \frac{1}{\sigma^2} \mathbf{X}' (\mathbf{y} - \mathbf{X}\beta) \right)$$

$$= -\frac{1}{(\sigma^2)^2} \mathbf{X}' (\mathbf{y} - \mathbf{X}\beta) = -\frac{1}{(\sigma^2)^2} (\mathbf{X}' \mathbf{y} - \mathbf{X}' \mathbf{X} \beta) \quad \text{Eq(30)}$$

Sostituendo con la stima di massima verosimiglianza  $(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y} = \hat{\beta}$ , ricaviamo il vettore «nullo» (composto da valori zero, di ordine  $(j + 1) \times 1$ )

$$\frac{\delta^2 l(\beta, \sigma^2; \mathbf{y}, \mathbf{X})}{\delta \sigma^2 \beta} = -\frac{1}{(\sigma^2)^2} (\mathbf{X}' \mathbf{y} - \mathbf{X}' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y})$$

$$= -\frac{1}{(\sigma^2)^2} (\mathbf{X}' \mathbf{y} - \mathbf{X}' \mathbf{y}) = \mathbf{0} \quad \text{Eq(31)}$$

**Stima iterativa, metodo (più rapido) Fisher scoring.**

Vettore delle derivate prime, di dimensione  $(j + 2) \times 1$ , formato dalle Eq(21 e 23),

$$\begin{aligned} \mathbf{u} &= \begin{bmatrix} \frac{\delta l(\boldsymbol{\beta}, \sigma^2; \mathbf{y}, \mathbf{X})}{\delta \boldsymbol{\beta}} \\ \frac{\delta l(\boldsymbol{\beta}, \sigma^2; \mathbf{y}, \mathbf{X})}{\delta \sigma^2} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{\sigma^2} \mathbf{X}'\mathbf{y} - \frac{1}{\sigma^2} \mathbf{X}'\mathbf{X} \boldsymbol{\beta} \\ -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \end{bmatrix} \end{aligned} \quad \text{Eq(32)}$$

Matrice a blocchi (Hessiana) delle derivate seconde, nella quale inseriamo i risultati delle Eq(27, 29, 31), ottenuti sostituendo le stime di massima verosimiglianza nelle Eq(26, 28, 30),

$$E(\mathbf{H}) = \begin{bmatrix} \begin{matrix} \dots & \dots & \dots \\ \vdots & \ddots & \vdots \\ \dots & \dots & \dots \end{matrix} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ \begin{matrix} 0 & \dots & 0 \\ 1 \times (j+1) \end{matrix} & \begin{matrix} \dots \\ 1 \times 1 \end{matrix} \end{bmatrix} \quad \text{Eq(33)}$$

$$E(\mathbf{H}) = \begin{bmatrix} -\frac{1}{\sigma^2} \mathbf{X}'\mathbf{X} & \mathbf{0} \\ \mathbf{0} & -\frac{n}{2(\sigma^2)^2} \end{bmatrix}$$

Dove  $\mathbf{0}$  è un vettore (riga o colonna) «nullo» (composto da valori zero) di dimensione  $(j + 1) \times 1$  e  $1 \times (j + 1)$ .

Vettore dei parametri ignoti  $(j + 2)$ :

$$\boldsymbol{\varphi} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_j \\ \sigma^2 \end{bmatrix}$$

Il valore della Log-verosimiglianza sostituendo  $\boldsymbol{\varphi}^{(t)}$ :

$$l(\boldsymbol{\varphi}^{(t)}; y_i, \mathbf{X}) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^{2(t)}) - \frac{1}{2\sigma^{2(t)}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(t)})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(t)})$$

verrà confrontato con quello della stima successiva  $\boldsymbol{\varphi}^{(t+1)}$ ,

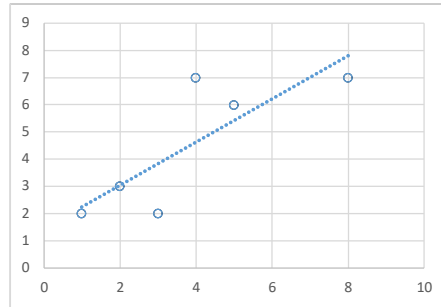
$$l(\boldsymbol{\varphi}^{(t+1)}; y_i, \mathbf{X}) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^{2(t+1)}) - \frac{1}{2\sigma^{2(t+1)}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(t+1)})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(t+1)})$$

ottenuto mediante la funzione generale («Newton-Raphson step»),

$$\underbrace{\boldsymbol{\varphi}^{(t+1)}}_{(j+2) \times 1} = \underbrace{\boldsymbol{\varphi}^{(t)}}_{(j+2) \times 1} - \underbrace{\left(E(\mathbf{H}^{(t)})\right)^{-1}}_{(j+2) \times 1} \mathbf{u}^{(t)}. \quad \text{Eq(34)}$$

Vediamo un esempio in Excel.

Y	X
3	2
2	3
2	1
7	4
6	5
7	8



OUTPUT RIEPILOGO

Statistica della regressione	
R multiplo	0,812353091
R al quadrato	0,659917545
R al quadrato corretto	0,574896931
Errore standard	1,583700763 (sigq*6/4)^(1/2)
Osservazioni	6

ANALISI VARIANZA

	gdl	SQ	MQ	F	Significatività F
Regressione	1	19,468	19,468	7,762	0,050
Residuo	4	10,032	2,508 (sigq*6/4) : correzione del denominatore		
Totale	5	29,5			per i gradi di libertà (n-k-1)

	Coefficienti	Errore standard	Stat t	Valore di significatività
Intercetta (beta0)	1,454054	1,270167	1,144774	0,316146
Variabile X (beta1)	0,794595	0,285209	2,786010	0,049513

```
> summary(glm(Y~X, family=gaussian(link=identity)))
```

```
Call:
glm(formula = Y ~ X, family = gaussian(link = identity))
```

```
Deviance Residuals:
    1      2      3      4      5      6
-0.04324 -1.83784 -0.24865  2.36757  0.57297 -0.81081
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.4541     1.2702   1.145  0.3161
X            0.7946     0.2852   2.786  0.0495 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 2.508108)
```

```
Null deviance: 29.500  on 5  degrees of freedom
Residual deviance: 10.032  on 4  degrees of freedom
AIC: 26.112
```

```
Number of Fisher Scoring iterations: 2
```

Y	X	Iterazione 0		Iterazione 1		Iterazione 2		Iterazione 3	
		phi0		phi(t+1)		phi(t+2)		phi(t+3)	
3	2	b0	1	b0	1,454	b0	1,454	b0	1,454
2	3	b	1	b	0,795	b	0,795	b	0,795
2	1	sigq	1	sigq	2,000	sigq	1,672	sigq	1,672
7	4								
6	5								
7	8								
		X		X		X		X	
		1	2	1	2	1	2	1	2
		1	3	1	3	1	3	1	3
		1	1	1	1	1	1	1	1
		1	4	1	4	1	4	1	4
		1	5	1	5	1	5	1	5
		1	8	1	8	1	8	1	8
		E(Y X)		Epsilon		E(Y X)		Epsilon	
		3	0	3,043	-0,043	3,043	-0,043	3,043	-0,043
		4	-2	3,838	-1,838	3,838	-1,838	3,838	-1,838
		2	0	2,249	-0,249	2,249	-0,249	2,249	-0,249
		5	2	4,632	2,368	4,632	2,368	4,632	2,368
		6	0	5,427	0,573	5,427	0,573	5,427	0,573
		9	-2	7,811	-0,811	7,811	-0,811	7,811	-0,811
		n	6	n	6	n	6	n	6
		logLik	-11,5136312	logLik	-10,1012	logLik	-10,0558	logLik	-10,0558
		logLik	-11,5136312	logLik	-10,10118	logLik	-10,05582	logLik	-10,05582
		Criterion	0,0000001		0,122676		0,004490		0,000000
					CAMBIA		CAMBIA		NO
		u		u		u		u	
		b0	-2	b0	0,000	b0	0,000	b0	0,000
		b	-14	b	0,000	b	0,000	b	0,000
		sigma	3	sigma	-0,246	sigma	0,000	sigma	0,000
		H		H		H		H	
		-6	-23	0	-3	-11,5	0	-3,6	-13,8
		-23	-119	0	-11,5	-59,5	0	-13,8	-71,2
		0	0	-3	0	0	-0,75	0,0	0,0
		phi(t+1)		phi(t+2)		phi(t+3)		phi(t+...)	
		b0	1,454054	b0	1,454054	b0	1,454054	b0	1,454054
		b	0,794595	b	0,794595	b	0,794595	b	0,794595
		sigma	2,000000	sigma	1,672072	sigma	1,672072	sigma	1,672072