Modelli di previsione per variabili dipendenti binarie.
Le variabili risposta di tipo binario presentano una codifica corrispondente a «1», *successo*, e «0», *insuccesso.* Un vettore n di risposte casuali e indipendenti dicotomiche potrebbe essere il seguente:

$$\mathbf{y} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 1 \end{bmatrix},$$

nel quale, ciascuna risposta segue una medesima distribuzione bernoulliana, $\Omega = \{0; 1\}$, con probabilità di «1» pari a p e probabilità di «0» pari a $1 - p$:

$$\begin{aligned} P(y_i = 1) &= \quad p; \\ P(y_i = 0) &= \quad 1 - p. \end{aligned}$$

Possiamo riscrivere la probabilità associata a ciascuna generica osservazione y_i nei termini di:

$$P(y_i) = p^{y_i}(1-p)^{1-y_i}, \text{con } y_i = 0, 1. \qquad \textbf{Eq. 35}$$

Avremo quindi per

$$\begin{aligned} y_i = 1, &\quad P(y_i) = p^1(1-p)^0 = p; \\ y_i = 0, &\quad P(y_i) = p^0(1-p)^1 = (1-p). \end{aligned}$$

Scriviamo ora la funzione di probabilità di massa per le nostre n osservazioni (…la nostra funzione di verosimiglianza), che sarà proporzionale al prodotto di n funzioni di Bernoulli (**Eq. 35**):

$$\begin{aligned} L(p; y_i) &= \prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i} \\ &= \prod_{i=1}^{n} p^{y_i} \frac{(1-p)^1}{(1-p)^{y_i}} \\ &= \prod_{i=1}^{n} \frac{p^{y_i}}{(1-p)^{y_i}} (1-p) \\ &= \left[\frac{p}{(1-p)}\right]^{\sum y_i} \prod_{i=1}^{n} (1-p) \end{aligned}$$

che si può riscrivere come

$$\begin{aligned} L(p; y_i) &= exp\left\{ln\left[\frac{p}{(1-p)}\right]^{\sum y_i}\right\} \prod_{i=1}^{n} (1-p) \\ &= exp\left\{\sum_{i=1}^{n} y_i \, ln\left[\frac{p}{(1-p)}\right]\right\} \prod_{i=1}^{n} (1-p) \end{aligned} \qquad \textbf{Eq. 36}$$

Prendendo il logaritmo naturale, (log-Verosimiglianza)

$$\begin{aligned} l(p; y_i) &= \sum_{i=1}^{n} y_i \, ln\left[\frac{p}{(1-p)}\right] + \sum_{i=1}^{n} ln(1-p) \\ &= \sum_{i=1}^{n} \left[y_i ln\left[\frac{p}{(1-p)}\right] + ln(1-p)\right] \end{aligned} \qquad \textbf{Eq. 37}$$

Assumiamo l'esistenza di un'associazione statistica tra la risposta binaria y_i ed una covariata x_i, che ci aiuta a ridefinire la previsione dell'esito di successo o insuccesso, dati i valori di x_i:

$$\mathbf{y} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}.$$

In questo caso, la probabilità associata a ciascuna osservazione y_i sarà **condizionata** alla conoscenza del valore in x_i, secondo i parametri β_j **di un adeguato modello statistico.** Formalmente, possiamo scrivere che:

$$P(y_i = 1) = \quad p(y_i = 1|x_i, \beta_j);$$
$$P(y_i = 0) = \quad 1 - p(y_i = 1|x_i, \beta_j).$$

La log-Verosimiglianza per n osservazioni casuali indipendenti, identicamente distribuite secondo una funzione di Bernoulli, può quindi essere riscritta come

$$l(\beta_j; y_i, x_i) =$$

Eq. 38

$$= \sum_{i=1}^{n} \left[y_i ln \left[\frac{p(y_i = 1|x_i, \beta_j)}{1 - p(y_i = 1|x_i, \beta_j)} \right] + ln(1 - p(y_i = 1|x_i, \beta_i)) \right]$$

Modello lineare di probabilità.
Potremmo usare una funzione di regressione lineare per specificare le probabilità degli esiti y_i.

- Si costruisce la variabile binaria y_i definita come:

$$y_i = \begin{cases} 1 & \text{se si realizza un evento "successo"} \\ 0 & \text{altrimenti (insuccesso)} \end{cases},$$

e si suppone che il suo valore atteso sia una funzione lineare di una variabile esplicativa:

$$E(y_i|x_i, \beta_j) = \alpha + \beta x_i. \qquad \textbf{Eq. 39}$$

- Poiché la variabile casuale y_i può assumere soltanto due valori [0,1], e ricordando la formula del valore atteso di una variabile aleatoria discreta

$$E(y_i) = \sum_{i=1}^{n} y_i p_i ,$$

si ha che

$$E(y_i|x_i, \beta_j) = 1 \times p(y_i = 1|x_i, \beta_j) + 0 \times p(y_i = 0|x_i, \beta_j)$$
$$= p(y_i = 1|x_i, \beta_j). \qquad \textbf{Eq. 40}$$

- Per l' **Eq. 39**, si ha quindi che la probabilità subordinata di successo è una funzione lineare della variabile esplicativa (*modello lineare di probabilità*), cioè

$$E(y_i|x_i, \beta_j) = p(y_i = 1|x_i, \beta_j) = \alpha + \beta x_i. \qquad \textbf{Eq. 41}$$

+

Modello lineare di probabilità.

- Facile da stimare applicando i minimi quadrati ordinari
- I suoi parametri β_j possono essere direttamente interpretati come gli effetti marginali delle variabili esplicative sulla variabile risposta dicotomica.
- Produce una buona approssimazione delle osservazioni quando le osservazioni sulle variabili esplicative sono vicine alla loro media.

–

- L'errore di previsione ε_i non ha una distribuzione normale attorno alla retta di regressione, bensì una distribuzione bernoulliana che può assumere soltanto due valori:

$$\varepsilon_i = y_i - E\left(y_i|x_i,\beta_j\right) =$$

$$= \begin{cases} 1-(\alpha+\beta x_i) & \text{quando } y_i = 1 \\ 0-(\alpha+\beta x_i) & \text{quando } y_i = 0\text{'} \end{cases}$$

con probabilità

$$P\left(\varepsilon_i = 1-(\alpha+\beta x_i)\right) = \quad \alpha+\beta x_i;$$
$$P\left(\varepsilon_i = -(\alpha+\beta x_i)\right) = \quad 1-(\alpha+\beta x_i).$$

Sulla base di queste probabilità si possono calcolare il valore atteso e la varianza della variabile aleatoria ε_i:

$$E(\varepsilon_i) = [1-(\alpha+\beta x_i)](\alpha+\beta x_i) +$$

$$-(\alpha+\beta x_i)[1-(\alpha+\beta x_i)] = 0 \qquad \textbf{Eq. 42}$$

Modello lineare di probabilità.

$$Var(\varepsilon_i) = [1-(\alpha+\beta x_i)]^2(\alpha+\beta x_i) +$$

$$-(\alpha+\beta x_i)^2[1-(\alpha+\beta x_i)]$$

Rinominiamo per semplicità la quantità $\alpha+\beta x_i$,

$$Var(\varepsilon_i) = \Big[1-\underbrace{(\alpha+\beta x_i)}_{a}\Big]^2\underbrace{(\alpha+\beta x_i)}_{a} +$$
$$-\underbrace{(\alpha+\beta x_i)}_{a}^2\Big[1-\underbrace{(\alpha+\beta x_i)}_{a}\Big]$$

riscrivendo la varianza dell'errore come:

$$Var(\varepsilon_i) = (1+a^2-2a)a + a^2(1-a)$$

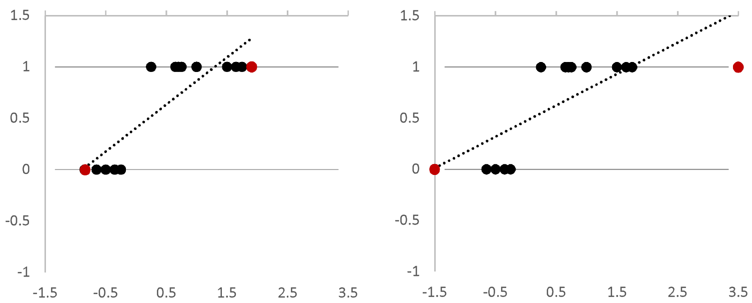
$$= a + a^3 - 2a^2 + a^2 - a^3 \qquad \textbf{Eq. 43}$$

$$= a - a^2 = (\alpha+\beta x_i)[1-(\alpha+\beta x_i)].$$

Quindi, se da un lato l'errore è indipendente in media dalle variabili esplicative (**Eq. 42**), la sua varianza non è unica e costante bensì dipende dallo specifico valore di x_i (**Eq. 43**, *eteroschedasticità*).

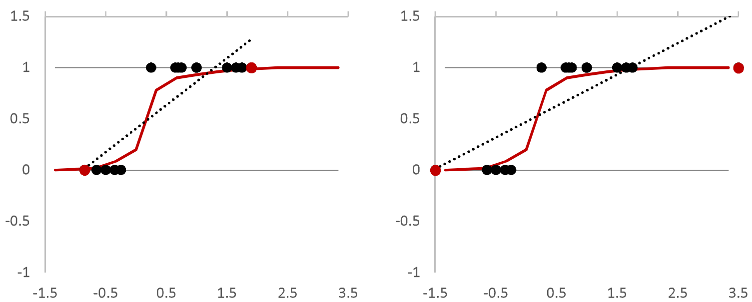
–

- Il modello lineare di probabilità non produce (automaticamente) un valore di probabilità vincolato all'intervallo [0, 1], soprattutto per osservazioni x_i lontane dalla media (Vedi figura successiva).



Modello lineare di probabilità.

La retta tratteggiata in figura rappresenta la funzione di regressione stimata; le rette parallele all'asse delle ascisse sono quelle in cui cadono le coppie di osservazioni (x_i, y_i), rappresentate dai punti in nero. Da notare come il modello lineare per la probabilità sia estremamente suscettibile alle osservazioni in x_i lontane dalla media (vedi punti in rosso).

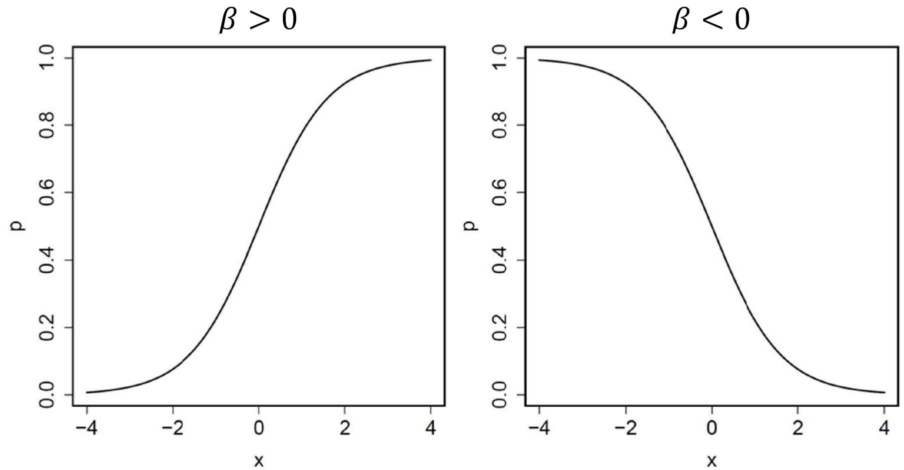


Modello NON lineare di probabilità.

Per superare gli inconvenienti connessi al modello di probabilità lineare è conveniente considerare una funzione che fornisca un valore ammissibile di $p(y_i = 1|x_i, \beta_j)$ per tutti i possibili valori delle variabili esplicative.

Modello Logit (di regressione logistica).
Nel modello *logit* la probabilità dell'evento «successo», condizionatamente a x_i, viene scritta come

$$p(y_i = 1|x_i, \beta_j) = \frac{exp\{\alpha + \beta x_i\}}{1 + exp\{\alpha + \beta x_i\}}, \qquad \textbf{Eq. 44}$$



detta *funzione di ripartizione logistica*, dalla quale otteniamo le due quantità fondamentali:

1. ***Complemento ad 1 della Eq. 44,***

$$1 - p(y_i = 1|x_i, \beta_j) =$$

$$= \frac{1 + exp\{\alpha + \beta x_i\} - exp\{\alpha + \beta x_i\}}{1 + exp\{\alpha + \beta x_i\}} \qquad \textbf{Eq. 45}$$

$$= \frac{1}{1 + exp\{\alpha + \beta x_i\}}$$

2. ***Logaritmo del rapporto delle probabilità, per dato x_i,***

$$\underbrace{ln\left[\frac{p(y_i = 1|x_i, \beta_j)}{1 - p(y_i = 1|x_i, \beta_j)}\right]}_{LOGIT} = ln\left[\frac{\frac{exp\{\alpha + \beta x_i\}}{1 + exp\{\alpha + \beta x_i\}}}{\frac{1}{1 + exp\{\alpha + \beta x_i\}}}\right] =$$

$$= ln[exp\{\alpha + \beta x_i\}] = \alpha + \beta x_i \qquad \textbf{Eq. 46}$$

da cui si individuano le peculiarità del modello *logit*: **è il *logit* che dipende linearmente dalle variabili esplicative e non la probabilità come nel caso del modello di probabilità lineare.** La quantità dentro alla parentesi (**rapporto di probabilità**, *odds*)

$$\frac{p(y_i = 1|x_i, \beta_j)}{1 - p(y_i = 1|x_i, \beta_j)} = \frac{p(y_i = 1|x_i, \beta_j)}{p(y_i = 0|x_i, \beta_j)}$$

definisce una sorta di pronostico in favore di una determinata «scelta». Se, ad esempio,

$$p(y_i = 1|x_i, \beta_j) = .25,$$

allora il pronostico è .333, cioè

$$\frac{p(y_i = 1|x_i, \beta_j)}{p(y_i = 0|x_i, \beta_j)} = \frac{.25}{.75} = \frac{1}{3}$$

3 a 1 in favore dell'evento $y_i = 0$.

Se, ad esempio, $p(y_i = 1|x_i, \beta_j) = .50$,
allora il pronostico è 1, cioè «parità» di scelta tra 1 è 0:

$$\frac{p(y_i = 1|x_i, \beta_j)}{p(y_i = 0|x_i, \beta_j)} = \frac{.50}{.50} = 1.$$

Se, ad esempio, $p(y_i = 1|x_i, \beta_j) = .75$,
allora il pronostico è 3, cioè

$$\frac{p(y_i = 1|x_i, \beta_j)}{p(y_i = 0|x_i, \beta_j)} = \frac{.75}{.25} = 3$$

3 a 1 in favore dell'evento $y_i = 1$.

Come si interpretano quindi α e β?

Interpretazione fondamentale di β

Il pronostico a favore di $y_i = 1$ si modifica moltiplicativamente di $exp\{\beta\}$ per ciascun incremento unitario in x_i.

Per $x_i = 0$ (o valore medio, se centrata), **l'Eq. 46** diventa:

$$\frac{p(y_i = 1|x_i = 0)}{1 - p(y_i = 1|x_i = 0)} = exp\{\alpha + \beta \cdot 0\} = exp\{\alpha\},$$

Per un incremento unitario in x_i abbiamo che

$$\frac{p(y_i = 1|x_i = 1)}{1 - p(y_i = 1|x_i = 1)} = exp\{\alpha + \beta\} = exp\{\alpha\} \cdot exp\{\beta\}.$$

Cosa succede al pronostico quando $\boldsymbol{x_i}$ passa a $\boldsymbol{x_i+1}$?
In corrispondenza di x_i

$$\frac{p(y_i=1|x_i,\beta_j)}{p(y_i=0|x_i,\beta_j)}=exp\{\alpha\}\cdot exp\{\beta\}^{x_i}.$$

In corrispondenza di x_i+1

$$\frac{p(y_i=1|x_i+1,\beta_j)}{p(y_i=0|x_i+1,\beta_j)}=exp\{\alpha\}\cdot exp\{\beta\}^{x_i+1}$$

$$=exp\{\alpha\}\cdot exp\{\beta\}^{x_i}\cdot exp\{\beta\}^{1}$$

$$=\frac{p(y_i=1|x_i,\beta_j)}{p(y_i=0|x_i,\beta_j)}exp\{\beta\}$$

$exp\{\beta\}$ è un odds ratio.

$$\frac{\frac{p(y_i=1|x_i+1,\beta_j)}{p(y_i=0|x_i+1,\beta_j)}}{\frac{p(y_i=1|x_i,\beta_j)}{p(y_i=0|x_i,\beta_j)}}=exp\{\beta\} \qquad \textbf{Eq. 47}$$

$$\beta=0.693147 \quad exp(\beta)=2=\frac{2}{1}\Big/\frac{1}{1}$$

*Con beta **positivo***, il pronostico a favore di $y_i=1$ aumenta (raddoppia) ad ogni incremento unitario in x_i;

$$\beta=-0.693147 \quad exp(\beta)=1/2=\frac{1}{2}\Big/\frac{1}{1}$$

*Con beta **negativo***, il pronostico a favore di $y_i=1$ diminuisce (dimezza) ad ogni incremento unitario in x_i;

Modello Logit (stima dei parametri).
Non ci rimane che stimare i parametri ignoti del modello mediante MLE, non prima di aver introdotto tre nuove regole di derivazione, i cui risultati ci consentono di ricavare velocemente le derivate prime e seconde della funzione di log-Verosimiglianza, nei parametri ignoti del modello logit. (vedi Appendice alla lezione).

Funzione di log-Verosimiglianza:
Riscriviamo l' **Eq. 38** inserendo le due quantità fondamentali derivanti dall'espressione della probabilità condizionate di successo nei termini della funzione di ripartizione logistica (**Eq. 45, 46**):

$$l(\beta_j;y_i,x_i)=$$

$$=\sum_{i=1}^{n}\left[y_i(\alpha+\beta x_i)+ln\left[\frac{1}{1+exp\{\alpha+\beta x_i\}}\right]\right]$$

$$=\sum_{i=1}^{n}[y_i(\alpha+\beta x_i)-ln(1+exp\{\alpha+\beta x_i\})]. \qquad \textbf{Eq. 48}$$

Derivate prime parziali [vedi **Eq. A.1-A.4**]:

$$\frac{\partial l(\beta_j;y_i,x_i)}{\partial\alpha}=\sum_{i=1}^{n}[y_i-p(y_i=1|x_i,\beta_i)] \qquad \textbf{Eq. 49}$$

$$\frac{\partial l(\beta_j;y_i,x_i)}{\partial\beta}=\sum_{i=1}^{n}[y_ix_i-p(y_i=1|x_i,\beta_i)x_i] \qquad \textbf{Eq. 50}$$

Derivate seconde parziali e cross- parziali[vedi **A.5** e **A.6**]:

$$\frac{\partial^2 l(\beta_j; y_i, x_i)}{\partial \alpha^2} = \frac{d}{d\alpha}\sum_{i=1}^{n}[y_i - p(y_i = 1|x_i, \beta_i)]$$

$$= -\sum_{i=1}^{n}\frac{d}{d\alpha}\left[\frac{exp\{\alpha + \beta x_i\}}{1 + exp\{\alpha + \beta x_i\}}\right]$$

$$= -\sum_{i=1}^{n} p(y_i = 1|x_i, \beta_i)[1 - p(y_i = 1|x_i, \beta_i)] \qquad \textbf{Eq. 51}$$

$$\frac{\partial^2 l(\beta_j; y_i, x_i)}{\partial \beta^2} = \frac{dy}{d\beta}\sum_{i=1}^{n}[y_i x_i - p(y_i = 1|x_i, \beta_i)x_i]$$

$$= -\sum_{i=1}^{n} x_i \frac{dy}{d\beta}\left[\frac{exp\{\alpha + \beta x_i\}}{1 + exp\{\alpha + \beta x_i\}}\right]$$

$$= -\sum_{i=1}^{n} {x_i}^2 p(y_i = 1|x_i, \beta_i)[1 - p(y_i = 1|x_i, \beta_i)] \qquad \textbf{Eq. 52}$$

$$\frac{\partial^2 l(\beta_j; y_i, x_i)}{\partial \alpha\beta} = \frac{dy}{d\beta}\sum_{i=1}^{n}[y_i - p(y_i = 1|x_i, \beta_i)]$$

$$= -\sum_{i=1}^{n}\frac{dy}{d\beta}\left[\frac{exp\{\alpha + \beta x_i\}}{1 + exp\{\alpha + \beta x_i\}}\right]$$

$$= -\sum_{i=1}^{n} x_i p(y_i = 1|x_i, \beta_i)[1 - p(y_i = 1|x_i, \beta_i)] \qquad \textbf{Eq. 53}$$

Derivate seconde parziali tutte di segno negativo (la funzione di Verosimiglianza va massimizzata).

Vettore derivate prime

$$\boldsymbol{u} = \begin{bmatrix} \dfrac{\partial l(\beta_j; y_i, x_i)}{\partial \alpha} \\ \dfrac{\partial l(\beta_j; y_i, x_i)}{\partial \beta} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n}[y_i - p(y_i = 1|x_i, \beta_i)] \\ \sum_{i=1}^{n} x_i[y_i - p(y_i = 1|x_i, \beta_i)] \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n}\left[y_i - \dfrac{exp\{\alpha + \beta x_i\}}{1 + exp\{\alpha + \beta x_i\}}\right] \\ \sum_{i=1}^{n} x_i\left[y_i - \dfrac{exp\{\alpha + \beta x_i\}}{1 + exp\{\alpha + \beta x_i\}}\right] \end{bmatrix}$$

Trattasi di equazioni esponenziali, in cui l'incognita compare come esponente, quindi, non riconducibile ad un polinomio uguagliato a zero. La soluzione di massima verosimiglianza ($\boldsymbol{u} = \boldsymbol{0}$) andrà cercata per approssimazione del risultato secondo i metodi iterativi che abbiamo visto.

In termini matriciali,

$$\mathbf{X}'(\mathbf{y} - \mathbf{p}) = \underbrace{\underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix}}_{\mathbf{2\times n}} \underbrace{\begin{bmatrix} y_1 - p(y_1 = 1|x_1, \beta_i) \\ y_2 - p(y_2 = 1|x_2, \beta_i) \\ \vdots \\ y_n - p(y_n = 1|x_n, \beta_i) \end{bmatrix}}_{\mathbf{n\times 2}}}_{\mathbf{2\times 1}} = \boldsymbol{u} \qquad \textbf{Eq. 54}$$

Dove $\mathbf{X}'$ è la matrice del modello lineare che abbiamo già visto, $\mathbf{y}$ è il vettore di risposte binarie osservate, e $\mathbf{p}$ è il vettore di valori previsti di probabilità di successo, secondo la funzione di ripartizione logistica.

Definiamo la quantità

$$a_i = p(y_i = 1|x_i,\beta_i)[1 - p(y_i = 1|x_i,\beta_i)]$$

$$= \frac{exp\{\alpha + \beta x_i\}}{1 + exp\{\alpha + \beta x_i\}}\left[1 - \frac{exp\{\alpha + \beta x_i\}}{1 + exp\{\alpha + \beta x_i\}}\right]$$

La matrice di derivate seconde (la matrice Hessiana)

$$\mathbf{H} = \begin{bmatrix} -\sum_{i=1}^{n} a_i & -\sum_{i=1}^{n} x_i a_i \\ -\sum_{i=1}^{n} x_i a_i & -\sum_{i=1}^{n} x_i^2 a_i \end{bmatrix},$$

In termini matriciali:

$$\mathbf{X}'\mathbf{W}\mathbf{X} = \underbrace{\underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix}}_{2\times n} \underbrace{\begin{bmatrix} -a_1 & 0 & 0 \\ 0 & -a_2 & 0 \\ 0 & 0 & -a_n \end{bmatrix}}_{n\times n} \underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}}_{n\times 2}}_{2\times 2}$$

$$= \underbrace{\underbrace{\begin{bmatrix} -a_1 & -a_2 & \cdots & -a_n \\ -x_1 a_1 & -x_2 a_2 & \cdots & -x_n a_n \end{bmatrix}}_{2\times n} \underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}}_{n\times 2}}_{2\times 2} = \mathbf{H} \qquad \textbf{Eq. 55}$$

Dove

- $\mathbf{W}$ denota una matrice diagonale $n \times n$ con elementi $p(y_i = 1|x_i,\beta_i)[1 - p(y_i = 1|x_i,\beta_i)]$ sulla diagonale principale,
- e dove $E[\mathbf{H}] = \mathbf{H}$ dal momento che la matrice non contiene nessuna funzione dei valori osservati y_i.

Vettore dei parametri ignoti $(j+1)$:

$$\boldsymbol{\varphi} = \begin{bmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_j \end{bmatrix},$$

nel nostro caso con una sola covariata $\boldsymbol{\varphi} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$

Il valore della log-Verosimiglianza sostituendo $\boldsymbol{\varphi}^{(t)}$**:**

$$l(\boldsymbol{\varphi}^{(t)}; y_i, x_i) = \sum_{i=1}^{n}\left[y_i(\alpha^{(t)} + \beta^{(t)}x_i) - ln(1 + exp\{\alpha^{(t)} + \beta^{(t)}x_i\})\right].$$

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

$$l(\boldsymbol{\varphi}^{(t+1)}; y_i, x_i) = \sum_{i=1}^{n}\left[y_i(\alpha^{(t+1)} + \beta^{(t+1)}x_i) - ln(1 + exp\{\alpha^{(t+1)} + \beta^{(t+1)}x_i\})\right]$$

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

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

$$= \boldsymbol{\varphi}^{(t)} - \left(\mathbf{X}'\mathbf{W}^{(t)}\mathbf{X}\right)^{-1}\mathbf{X}'(\mathbf{y} - \mathbf{p}^{(t)}) \qquad \textbf{Eq. 56}$$

Vediamo un esempio in R.

Appendice alla lezione

*Regola generale della derivazione di funzioni composte o **Chain Rule**:*

$$\frac{d}{dx}f[g(x)] = \frac{d}{dg}f[g] \cdot \frac{d}{dx}g(x),$$

così $\frac{d}{dx}(x^2+1)^3 = \frac{d}{d(x^2+1)}(x^2+1)^3 \cdot \frac{d}{dx}(x^2+1) = 3(x^2+1)^2 2x.$

$$\frac{d}{dx}exp\{f(x)\} = exp\{f(x)\} \cdot \frac{d}{dx}f(x),$$

Ad esempio,

$$\frac{d}{d\alpha}exp\{\alpha+\beta x_i\} = exp\{\alpha+\beta x_i\} \cdot \frac{d}{d\alpha}(\alpha+\beta x_i) = exp\{\alpha+\beta x_i\}(1\alpha^0) \qquad \textbf{Eq. A.1}$$

$$\frac{d}{d\beta}exp\{\alpha+\beta x_i\} = exp\{\alpha+\beta x_i\} \cdot \frac{d}{d\beta}(\alpha+\beta x_i) = exp\{\alpha+\beta x_i\}(x_i 1\beta^0) \qquad \textbf{Eq. A.2}$$

$$\frac{d}{dx}ln[f(x)] = \frac{1}{f(x)} \cdot \frac{d}{dx}f(x),$$

Ad esempio,

$$\frac{d}{d\alpha}ln[1+exp\{\alpha+\beta x_i\}] = \frac{1}{1+exp\{\alpha+\beta x_i\}} \cdot \frac{d}{d\alpha}\left(1+\underbrace{exp\{\alpha+\beta x_i\}}\right)$$

$$= \frac{exp\{\alpha+\beta x_i\}}{1+exp\{\alpha+\beta x_i\}} = p(y_i=1|x_i,\beta_i) \qquad \textbf{Eq. A.3}$$

$$\frac{d}{d\beta}ln[1+exp\{\alpha+\beta x_i\}] = \frac{1}{1+exp\{\alpha+\beta x_i\}} \cdot \frac{d}{d\beta}\left(1+\underbrace{exp\{\alpha+\beta x_i\}}\right)$$

$$= \frac{exp\{\alpha+\beta x_i\}}{1+exp\{\alpha+\beta x_i\}}x_i = p(y_i=1|x_i,\beta_i)x_i \qquad \textbf{Eq. A.4}$$

$$\frac{d}{dx}\left[\frac{g(x)}{h(x)}\right] = \frac{\left[\frac{d}{dx}g(x)\right]h(x) - \left[\frac{d}{dx}h(x)\right]g(x)}{h(x)^2},$$

Ad esempio,

$$\frac{d}{d\alpha}\left[\frac{exp\{\alpha+\beta x_i\}}{1+exp\{\alpha+\beta x_i\}}\right] = ? \qquad \frac{d}{d\beta}\left[\frac{exp\{\alpha+\beta x_i\}}{1+exp\{\alpha+\beta x_i\}}\right] = ?$$

Appendice alla lezione

$$\frac{d}{d\alpha}\left[\frac{exp\{\alpha+\beta x_i\}}{1+exp\{\alpha+\beta x_i\}}\right] =$$

$$= \frac{\left[\frac{d}{d\alpha}exp\{\alpha+\beta x_i\}\right](1+exp\{\alpha+\beta x_i\}) - \left[\frac{d}{d\alpha}(1+exp\{\alpha+\beta x_i\})\right]exp\{\alpha+\beta x_i\}}{(1+exp\{\alpha+\beta x_i\})^2}$$

Applicando la **Eq. A.1**

$$= \frac{exp\{\alpha+\beta x_i\}\cdot(1+exp\{\alpha+\beta x_i\}) - exp\{\alpha+\beta x_i\}\cdot exp\{\alpha+\beta x_i\}}{(1+exp\{\alpha+\beta x_i\})^2}$$

$$= \frac{exp\{\alpha+\beta x_i\} + exp\{\alpha+\beta x_i\}^2 - exp\{\alpha+\beta x_i\}^2}{(1+exp\{\alpha+\beta x_i\})^2}$$

$$= \frac{exp\{\alpha+\beta x_i\}}{1+exp\{\alpha+\beta x_i\}} \cdot \frac{1}{1+exp\{\alpha+\beta x_i\}}$$

$$= p(y_i=1|x_i,\beta_i)\cdot[1-p(y_i=1|x_i,\beta_i)] \qquad \textbf{Eq. A.5}$$

$$\frac{d}{d\beta}\left[\frac{exp\{\alpha+\beta x_i\}}{1+exp\{\alpha+\beta x_i\}}\right] =$$

$$= \frac{\left[\frac{d}{d\beta}exp\{\alpha+\beta x_i\}\right](1+exp\{\alpha+\beta x_i\}) - \left[\frac{d}{d\beta}(1+exp\{\alpha+\beta x_i\})\right]exp\{\alpha+\beta x_i\}}{(1+exp\{\alpha+\beta x_i\})^2}$$

Applicando la **Eq. A.2**

$$= \frac{exp\{\alpha+\beta x_i\}x_i\cdot(1+exp\{\alpha+\beta x_i\}) - exp\{\alpha+\beta x_i\}x_i\cdot exp\{\alpha+\beta x_i\}}{(1+exp\{\alpha+\beta x_i\})^2}$$

$$= \frac{exp\{\alpha+\beta x_i\}x_i + exp\{\alpha+\beta x_i\}^2 x_i - exp\{\alpha+\beta x_i\}^2 x_i}{(1+exp\{\alpha+\beta x_i\})^2}$$

$$= x_i\frac{exp\{\alpha+\beta x_i\}}{1+exp\{\alpha+\beta x_i\}} \cdot \frac{1}{1+exp\{\alpha+\beta x_i\}}$$

$$= x_i\cdot p(y_i=1|x_i,\beta_i)\cdot[1-p(y_i=1|x_i,\beta_i)] \qquad \textbf{Eq. A.6}$$