

Università degli studi di Padova  
Dipartimento di Scienze Statistiche  
Corso di Laurea Triennale in  
Statistica e Tecnologie Informatiche



RELAZIONE FINALE

**VALIDITÀ DEI TEST BASATI SULLA VEROSIMIGLIANZA  
PENALIZZATA IN MODELLI LOG-LINEARI**

Relatore Prof. Nicola Sartori  
Dipartimento di Scienze Statistiche

Laureando Alessandro De Bettin  
Matricola N 1030482

Anno Accademico 2013/2014



# Indice

<b>Introduzione</b>	<b>9</b>
<b>1 Principi della teoria della Verosimiglianza</b>	<b>11</b>
1.1 Concetti di Base . . . . .	11
1.2 Stimatore di massima verosimiglianza e sue proprietà . . . . .	12
1.3 Test basati sulla verosimiglianza . . . . .	13
1.3.1 Il test log-rapporto di verosimiglianza . . . . .	13
1.3.2 Il test di Wald . . . . .	14
1.3.3 Il test score . . . . .	15
<b>2 Modelli Lineari Generalizzati: il modello Log-lineare</b>	<b>17</b>
2.1 Concetti generali . . . . .	17
2.2 Stima dei parametri: minimi quadrati ponderati iterati . . . . .	19
2.3 Il modello log-lineare . . . . .	20
<b>3 Riduzione della distorsione tramite funzione score penalizzata</b>	<b>23</b>
3.1 Metodo della funzione score penalizzata . . . . .	24
3.2 Modelli lineari generalizzati con legame canonico: verosimiglianza penalizzata . . . . .	25
3.2.1 La log-verosimiglianza penalizzata nel modello log-lineare	26
3.3 I test di verosimiglianza penalizzata . . . . .	27
3.4 Obiettivo . . . . .	28
<b>4 Metodi di simulazione</b>	<b>29</b>
4.1 Presentazione dell'algoritmo . . . . .	29

4.2	Generazione di tabelle di contingenza . . . . .	30
4.3	Analisi di ogni tabella . . . . .	31
4.4	Metodi di sintesi delle informazioni . . . . .	32
<b>5</b>	<b>Risultati</b>	<b>33</b>
5.1	Le tabelle reali . . . . .	33
5.2	Studi di simulazione . . . . .	35
5.2.1	Numerosità originale . . . . .	35
5.2.2	Numerosità ridotta . . . . .	39
5.3	Considerazioni finali . . . . .	43
	<b>Bibliografia</b>	<b>49</b>
<b>A</b>	<b>Codice R utilizzato</b>	<b>51</b>

# Elenco dei codici

A.1 Funzioni R e loro utilizzo . . . . .	51
--	----



## Elenco delle tabelle

5.1	Pazienti con malattie psichiatriche . . . . .	33
5.2	Coefficienti del primo modello ridotto . . . . .	34
5.3	Soddisfazione sessuale dei partner . . . . .	34
5.4	Coefficienti del secondo modello ridotto . . . . .	34
5.5	Quantili empirici ( $\times 100$ ) relativi ai valori tipici dell' $\alpha$ fissato .	35
5.6	Stime dei parametri e della distorsione . . . . .	45
5.7	Quantili empirici ( $\times 100$ ) relativi ai valori tipici dell' $\alpha$ fissato .	45
5.8	Stime dei parametri e della distorsione . . . . .	45
5.9	Quantili empirici ( $\times 100$ ) relativi ai valori tipici dell' $\alpha$ fissato .	46
5.10	Stime dei parametri e della distorsione . . . . .	46
5.11	Quantili empirici ( $\times 100$ ) relativi ai valori tipici dell' $\alpha$ fissato .	46
5.12	Stime dei parametri e della distorsione . . . . .	47





# Introduzione

L'obiettivo di questo studio è verificare la validità dei test basati sulla verosimiglianza nel caso in cui, al fine di ridurre la distorsione dello stimatore di massima verosimiglianza, venga usata una verosimiglianza penalizzata piuttosto che quella classica. L'ambito statistico in cui questo approccio verrà studiato è quello delle tabelle di contingenza, e in particolare del modello log-lineare. Infatti, in situazioni in cui la somma dei conteggi delle caselle di una tabella sia bassa, è probabile ottenere delle stime di massima verosimiglianza dei parametri del modello log-lineare ipotizzato molto distorte o addirittura infinite. Per studiare il comportamento dei test basati sulla verosimiglianza penalizzata verranno adottati metodi di simulazione; partendo da tabelle reali, verrà prodotto un gran numero di tabelle simili a queste, il cui modello generatore è completamente noto; per ognuna di esse sarà calcolato il  $p$ -value del test del log-rapporto di verosimiglianza, del test di Wald e del test score sotto l'ipotesi  $H_0$ , di cui si conosce a priori la fondatezza, utilizzando prima la funzione di verosimiglianza classica e poi quella penalizzata. Per ogni tipologia di  $p$ -value e di verosimiglianza verranno confrontati i quantili empirici dei  $p$ -value con quelli teorici; poiché i  $p$ -value sono calcolati sotto  $H_0$ , i quantili teorici di riferimento sono quelli di una variabile casuale uniforme in  $(0, 1)$ .

La trattazione è divisa in due sezioni principali: la sezione teorica, in cui verrà esposta la teoria che giustifica questo studio, e la sezione pratica, in cui saranno presentati i metodi di simulazione, i risultati ottenuti e una loro analisi. La prima seguirà questo ordine: verrà brevemente descritta la teoria inerente la verosimiglianza, in particolare riguardo gli stimatori di massima verosimiglianza e i test basati sulla verosimiglianza; di seguito verrà data una

presentazione dei modelli lineari generalizzati, con specifica attenzione per il modello log-lineare. Dopo una descrizione sommaria delle problematiche inerenti la distorsione, verrà presentato il metodo della funzione score penalizzata e la sua applicazione al modello log-lineare. La seconda parte comincerà con la spiegazione dei metodi di simulazione utilizzati; in seguito verranno riportati i risultati ottenuti da simulazioni di tabelle di contingenza a partire da tabelle reali; per concludere, verranno esposte delle considerazioni sulla validità del metodo proposto e sui suoi pro e i suoi contro.

# Capitolo 1

## Principi della teoria della Verosimiglianza

In questo capitolo viene presentato l'approccio all'inferenza statistica basato sulla verosimiglianza, con interesse particolare verso lo stimatore di massima verosimiglianza, le sue proprietà, e le proprietà dei test del log-rapporto di verosimiglianza, di Wald e score. Per approfondimenti vedere Pace e Salvan (2001).

### 1.1 Concetti di Base

Il metodo della massima verosimiglianza consente, partendo da un campione di osservazioni di una specifica variabile casuale, di stimarne i parametri tramite stimatori dalle caratteristiche desiderabili. Intuitivamente, il principio di questo metodo è: le stime più accurate dei parametri sono quelle che massimizzano la “probabilità congiunta” di aver osservato le unità statistiche facenti parte del campione. Più rigorosamente: sia  $\underline{y} = (y_1, y_2, \dots, y_n)$  un campione di osservazioni indipendenti e identicamente distribuite (*i.i.d.*) generato da  $Y \sim f(y, \theta)$ , con  $\theta$  appartenente allo spazio parametrico  $\Theta$ . La funzione di probabilità congiunta (densità congiunta per variabili continue) è  $p_{\underline{y}}(\underline{y}; \theta) = \prod_{i=1}^n f(y_i, \theta)$ . La funzione di verosimiglianza è la stessa funzione ma, se prima le variabili erano i valori assunti dalle variabili casuali, ora lo

sono i parametri. Essa è

$$L(\theta) = p_Y(y^{oss}; \theta),$$

con  $y^{oss}$  campione delle osservazioni effettive. Usualmente, per semplificare alcuni passaggi algebrici, si preferisce utilizzare il logaritmo della funzione di verosimiglianza, detta log-verosimiglianza:

$$l(\theta) = \log L(\theta).$$

La derivata prima di questa funzione è detta funzione score:

$$l_*(\theta) = \left[ \frac{\partial l(\theta)}{\partial \theta_1}, \dots, \frac{\partial l(\theta)}{\partial \theta_p} \right].$$

La funzione score è  $p$ -dimensionale, dove  $p$  è la dimensione del parametro  $\theta$ . La derivata seconda della log-verosimiglianza cambiata di segno è detta informazione osservata ed è una matrice  $p \times p$ :

$$j(\theta) = -l_{**}(\theta) = \left[ -\frac{\partial^2 l(\theta)}{\partial \theta_r \partial \theta_s} \right],$$

per  $r, s = 1, \dots, p$ . Il valore atteso dell'informazione osservata è l'informazione attesa:

$$i(\theta) = E[j(\theta)].$$

Quando  $\Theta$  è un sottoinsieme aperto di  $\mathbb{R}^p$  e  $l(\theta)$  è differenziabile almeno tre volte, con derivate parziali continue in  $\Theta$ , si dice che la verosimiglianza è regolare.

## 1.2 Stimatore di massima verosimiglianza e sue proprietà

Lo stimatore di massima verosimiglianza si ottiene massimizzando la funzione di verosimiglianza. Poiché la trasformazione logaritmica è monotona crescente, massimizzare la log-verosimiglianza è equivalente a massimizzare la verosimiglianza. Il punto di massimo della log-verosimiglianza è quello in cui la derivata prima di  $l(\theta)$  è zero,  $l_*(\theta) = 0$ , e la derivata seconda è negativa,  $j(\theta) > 0$ , nel caso  $\dim(\theta) = 1$ . Nel caso di verosimiglianza regolare, spesso la

soluzione è unica. Si definisce con  $\hat{\theta}$  lo stimatore di massima verosimiglianza di  $\theta$ .

Sotto condizioni di regolarità, lo stimatore di massima verosimiglianza possiede caratteristiche degne di nota:

- $\hat{\theta}$  è funzione della statistica sufficiente minimale per  $\theta$ ,
- Se  $\psi = \psi(\theta)$  è una funzione biunivoca di  $\theta$ , allora  $\hat{\psi} = \psi(\hat{\theta})$  è lo stimatore di massima verosimiglianza di  $\psi$  (proprietà di equivarianza),
- $\hat{\theta}$  è consistente,
- $\hat{\theta}$  è asintoticamente non distorto,
- $\hat{\theta}$  ha distribuzione asintotica normale multivariata, in particolare  $\hat{\theta} \sim N_p(\theta, i(\theta)^{-1})$ , dove il simbolo  $\sim$  indica “si distribuisce approssimativamente come”.

## 1.3 Test basati sulla verosimiglianza

I test basati sulla verosimiglianza sono tre: il test log-rapporto di verosimiglianza o test di Wilks, il test di Wald e il test score.

### 1.3.1 Il test log-rapporto di verosimiglianza

Dato il seguente test di ipotesi:

$$\begin{cases} H_0 : \theta = \theta_0, \\ H_1 : \theta \neq \theta_0, \end{cases}$$

la forma del test di Wilks è

$$W(\theta_0) = 2\{l(\hat{\theta}) - l(\theta_0)\},$$

dove  $l(\hat{\theta})$  è il valore massimo della log-verosimiglianza, mentre  $l(\theta_0)$  è il valore della medesima funzione sotto  $H_0$ . Sebbene in alcuni casi si possa ricavare la distribuzione esatta di questa statistica, spesso si utilizza la sua distribuzione asintotica. Essa è chi-quadrato con  $p$  gradi di libertà,  $W(\theta) \sim \chi_p^2$ . Nel caso

si faccia un'ipotesi su un sottoinsieme dei parametri, si utilizza il test log-rapporto di verosimiglianza profilo. Sia  $\theta = (\tau, \psi)$ , con  $\tau$  vettore dei primi  $p_0 < p$  parametri di  $\theta$  e  $\psi$  dei rimanenti  $p - p_0$ , e le seguenti ipotesi nulla e alternativa:

$$\begin{cases} H_0 : \psi = \psi_0, \\ H_1 : \psi \neq \psi_0. \end{cases}$$

Allora, si ha

$$W_p(\psi_0) = 2\{l(\hat{\tau}, \hat{\psi}) - l(\hat{\tau}_{\psi_0}, \psi_0)\},$$

con  $\hat{\tau}_{\psi_0}$  vettore delle stime di massima verosimiglianza di  $\tau$  sotto  $H_0$ , cioè con  $\psi = \psi_0$ . La sua distribuzione asintotica è chi-quadrato con  $p - p_0$  gradi di libertà,  $W_p(\psi_0) \sim \chi_{p-p_0}^2$ .

### 1.3.2 Il test di Wald

Siano le seguenti le ipotesi da verificare:

$$\begin{cases} H_0 : \theta = \theta_0, \\ H_1 : \theta \neq \theta_0. \end{cases}$$

Il test di Wald assume la seguente forma

$$W_e(\theta_0) = (\hat{\theta} - \theta_0)^T i(\theta_0) (\hat{\theta} - \theta_0),$$

dove  $i(\theta_0)$  è l'informazione attesa calcolata sotto  $H_0$ . Questo test ha la stessa distribuzione asintotica nulla del test  $W$ , e cioè  $W_e(\theta_0) \sim \chi_p^2$ . Nel caso si voglia testare solo una parte dei parametri,  $H_0 : \psi = \psi_0$ , il test di Wald assume la seguente forma

$$W_e(\psi_0) = (\hat{\psi} - \psi_0)^T \{[i(\theta_0)^{-1}]_{(\psi, \psi)}\}^{-1} (\hat{\psi} - \psi_0),$$

dove  $\theta_0 = (\hat{\tau}_{\psi_0}, \psi_0)$  e  $\{[i(\theta_0)^{-1}]_{(\psi, \psi)}\}^{-1}$  è la matrice inversa della sottomatrice relativa a  $\psi$  della matrice inversa dell'informazione attesa sotto  $H_0$ . La distribuzione asintotica, in questo caso, è la stessa di  $W_p(\psi_0)$ .

Il test di Wald, fra i test basati sulla verosimiglianza, è quello che dipende maggiormente dalla parametrizzazione del modello; è, quindi, anche il più sensibile alla distorsione degli stimatori, poiché quest'ultima dipende dal tipo

di parametrizzazione. La distribuzione asintotica del test non cambia se, come stimatori della varianza di  $\hat{\theta}$ , vengono usati  $i(\hat{\theta})^{-1}$  o  $j(\hat{\theta})^{-1}$ . Quest'ultima opzione è la più frequente.

### 1.3.3 Il test score

Il test score è l'unico fra i test basati sulla verosimiglianza a non dipendere dalla stima di  $\theta$ . Date le seguenti ipotesi:

$$\begin{cases} H_0 : \theta = \theta_0, \\ H_1 : \theta \neq \theta_0, \end{cases}$$

il test score è

$$W_u(\theta_0) = l_*(\theta_0)^T i(\theta_0)^{-1} l_*(\theta_0).$$

La sua distribuzione asintotica nulla è la stessa degli altri due test presentati, ovvero:  $W_u(\theta_0) \sim \chi_p^2$ . Sia  $\theta = (\tau, \psi)$ , la funzione score  $l_*(\theta) = \left[ \frac{\partial l(\theta)}{\partial \tau}, \frac{\partial l(\theta)}{\partial \psi} \right] = [l_\tau, l_\psi]$  e  $H_0 : \psi = \psi_0$  l'ipotesi da testare. Il test score è

$$W_u(\psi_0) = l_{\psi_0}^T [i(\theta_0)^{-1}]_{(\psi, \psi)} l_{\psi_0},$$

con  $\theta_0 = (\hat{\tau}_{\psi_0}, \psi_0)$ . Coerentemente con quella degli altri test, la sua distribuzione asintotica è  $W_u(\psi_0) \sim \chi_{p-p_0}^2$ .

Stimare la varianza di  $l_*(\theta)$ ,  $i(\theta)$ , con  $i(\hat{\theta})$  o  $j(\hat{\theta})$  non altera la distribuzione asintotica del test.





# Capitolo 2

## Modelli Lineari Generalizzati: il modello Log-lineare

Per la scrittura di questo capitolo, sono state di aiuto le dispense della Professoressa Menardi (2013-2014). Per approfondimenti sui modelli lineari generalizzati e sul modello log-lineare vedere Piccolo (2010).

### 2.1 Concetti generali

I modelli lineari generalizzati sono modelli statistici per dati provenienti da variabili casuali appartenenti alla famiglia di dispersione esponenziale. Essi constano di tre componenti fondamentali:

1. componente casuale,
2. componente sistematica,
3. funzione legame.

La componente casuale è la variabile risposta, con distribuzione appartenente alla famiglia esponenziale; nei modelli lineari generalizzati, la sua media viene messa in relazione con una combinazione lineare delle variabili esplicative. Una variabile aleatoria si dice avere distribuzione appartenente alla famiglia di dispersione esponenziale se e solo se la sua funzione di densità può essere

scritta nel seguente modo:

$$f(y; \gamma, \delta) = \exp \left\{ \frac{1}{\delta} (y\gamma - b(\gamma)) + c(y, \delta) \right\},$$

con  $\gamma$  parametro naturale e  $\delta$  parametro di dispersione o di scala. Alla famiglia di dispersione esponenziale appartengono molte variabili casuali ben note in statistica, ad esempio le variabili normale, Poisson e binomiale. Ciò rende questi modelli più versatili del diffuso modello lineare normale, in cui la variabile risposta può avere solo distribuzione normale.

La componente sistematica è la parte deterministica del modello, costituita da una combinazione lineare delle variabili esplicative, o loro trasformazioni. Essa è un vettore di  $n$  componenti, dove  $n$  è il numero di osservazioni. La sua forma generica è

$$\eta_i = \sum_{j=1}^p x_{ij} \beta_j = \tilde{x}_i^T \beta,$$

con  $i = 1, \dots, n$ ,  $\beta$  vettore dei parametri,  $p = \dim(\beta)$  e  $\tilde{x}_i$  vettore di variabili esplicative relative all' $i$ -esima osservazione.

La funzione legame, come suggerisce il nome, mette in relazione la componente casuale con quella sistematica. Essa è nota, monotona e derivabile. In particolare, il suo compito è legare la media della variabile casuale,  $\mu_i$ , con una combinazione lineare delle variabili esplicative,  $\eta_i$ . Solitamente viene identificata nel seguente modo:

$$g(\mu_i) = \eta_i.$$

Nel caso in cui la funzione legame sia uguale al parametro naturale della famiglia esponenziale della variabile risposta, la funzione si dice essere di legame canonico. La presenza di questa caratteristica comporta la semplificazione nel calcolo delle funzioni legate alla verosimiglianza.

Una forma generale per rappresentare un modello lineare generalizzato può quindi essere

$$g(\mu) = X\beta,$$

con  $\mu$  vettore  $n \times 1$  delle medie,  $X$  matrice delle variabili esplicative (la cui  $i$ -esima riga è  $\tilde{x}_i$ ) e  $\beta$  il vettore dei  $p$  regressori.

Al fine di valutare la bontà di un modello lineare generalizzato, uno degli elementi considerati è il confronto fra il modello in esame ed il modello saturo. Il modello saturo è quel modello che ha tanti parametri quante sono le osservazioni e che, conseguentemente, ha l'adattamento migliore ai dati. In particolare, i valori stimati per la variabile risposta coincidono con quelli osservati. Esso, pur non avendo una reale utilità come modello a se stante, ha valore come metro di paragone per modelli le cui esplicative sono un sottoinsieme delle sue, detti, proprio per questo, modelli ridotti. Per effettuare un confronto fra il modello saturo e un modello ridotto, il metodo standard consiste nell'effettuare un test per modelli annidati. Questa procedura viene di seguito descritta; sia  $M_1$  un modello di  $p \leq n$  parametri e  $M_0$  un modello ad esso annidato, con  $p_0 < p$  parametri. Le ipotesi da verificare sono le seguenti:

$$\begin{cases} H_0 : M_0 \equiv M_1, \\ H_1 : M_0 \not\equiv M_1. \end{cases}$$

Sia definito come  $\hat{\theta}_1$  il vettore delle stime dei parametri sotto  $H_1$  e  $\hat{\theta}_0$  quello sotto  $H_0$ . La forma del test è

$$W_p = 2(l(\hat{\theta}_1) - l(\hat{\theta}_0)),$$

nel caso venga usato il test del log-rapporto di verosimiglianza. Il test  $W_p$  ha distribuzione asintotica chi-quadrato con  $p - p_0$  gradi di libertà. Se l'ipotesi nulla viene accettata,  $M_0$  viene considerato tanto esplicativo quanto  $M_1$ . Nel caso venga testata l'equivalenza del modello saturo e di un modello ridotto di  $p_0 < n$  parametri, la distribuzione approssimata del test è  $W_p \sim \chi_{n-p_0}^2$ , poiché il modello saturo ha tanti parametri quante sono le osservazioni.

## 2.2 Stima dei parametri: minimi quadrati ponderati iterati

L'algoritmo dei minimi quadrati ponderati iterati consente di ricavare numericamente le stime di massima verosimiglianza dei parametri  $\beta$  di un modello lineare generalizzato. Nella maggior parte dei casi, infatti, le equazioni di verosimiglianza non hanno una soluzione analitica. Questo algoritmo

è un caso particolare dell'algoritmo dello scoring di Fisher, a sua volta ricavato dall'algoritmo di Newton-Raphson. L'obiettivo è trovare per quale valore dei parametri la funzione score assuma valore nullo. Il metodo è basato su approssimazioni della funzione score tramite polinomio di Taylor. Siano  $t$  il numero dell'iterazione corrente e  $i(\beta)$  l'informazione attesa. La  $t$ -esima iterazione è

$$\hat{\beta}^{(t)} = \hat{\beta}^{(t-1)} + i(\hat{\beta}^{(t-1)})^{-1} l_*(\hat{\beta}^{(t-1)}).$$

Il valore assunto da  $\hat{\beta}^0$ , ovvero il punto da cui parte l'algoritmo, è deciso tramite criteri specifici a seconda della distribuzione della variabile risposta. La procedura termina quando la distanza fra  $\beta^{(t)}$  e  $\beta^{(t+1)}$  è minore di un determinato valore, oppure quando il numero di iterazioni supera una soglia stabilita. Quest'ultimo caso si verifica quando le stime di massima verosimiglianza di  $\beta$  sono sulla frontiera dello spazio parametrico.

## 2.3 Il modello log-lineare

Il modello log-lineare è un modello lineare generalizzato adatto allo studio di tabelle di contingenza. Supponiamo di voler analizzare una tabella di conteggi raggruppati, per semplicità, secondo due variabili categoriali  $A$  e  $B$ , di  $R$  e  $C$  modalità rispettivamente; è di interesse capire in che modo ogni modalità assunta dalle due variabili categoriali influenzi le frequenze rilevate e, soprattutto, verificare se fra le due ci sia indipendenza o meno. Per fare ciò, il metodo migliore è l'utilizzo di un modello lineare generalizzato per dati di conteggio.

La componente sistematica di un modello del genere assume due modalità fondamentali: quella del modello saturo e quella del modello ridotto. La prima, per una generica casella di posizione  $rc$ , sarà del tipo

$$\eta_{rc} = \beta_1 + \beta_r^A + \beta_c^B + \beta_{rc}^{AB}.$$

Il numero di parametri è, quindi,  $R \times C$ , ovvero il numero di caselle della tabella. La seconda:

$$\eta_{rc} = \beta_1 + \beta_r^A + \beta_c^B.$$

con  $R + C - 1$  parametri. Nel seguito della trattazione, con modello saturo e con modello ridotto si intenderanno i modelli log-lineari con le componenti

sistematiche appena descritte. Come si può osservare, entrambe le possibilità contengono una parte dedicata alle due variabili categoriali e alle loro modalità:  $\beta_1 + \beta_r^A + \beta_c^B$  (effetto dovuto alla prima modalità di  $A$  e di  $B$ , effetto della categoria assunta da  $A$  rispetto alla sua prima modalità, effetto della categoria assunta da  $B$  rispetto alla sua prima modalità, rispettivamente). La differenza consiste nell'elemento  $\beta_{rc}^{AB}$ : l'interazione fra le due variabili. Quest'ultima componente rappresenta la dipendenza fra  $A$  e  $B$ : se è significativamente non nulla le variabili sono da considerarsi dipendenti.

A seconda del disegno di campionamento adottato per costruire la tabella, la componente casuale del modello assume una forma diversa. Nel caso sia stabilita a priori un'area di campionamento per lo studio, ad esempio un lasso di tempo o un'area vera e propria, il modello più adatto è un modello di Poisson; nel caso sia deciso a priori il numero di unità statistiche che verranno analizzate, è corretto usare un modello multinomiale. La variabile casuale di Poisson e la variabile multinomiale sono legate fra loro, infatti la distribuzione di  $n$  osservazioni provenienti da una Poisson( $\mu_i$ ),  $i = 1, \dots, n$ , condizionata al fatto che la loro somma è  $N$ , coincide con quella di una Multinomiale di parametri  $N$  e  $\pi = [\pi_1, \dots, \pi_n] = \left[ \frac{\mu_1}{\sum_{i=1}^n \mu_i}, \dots, \frac{\mu_n}{\sum_{i=1}^n \mu_i} \right]$ . Il valore atteso del conteggio nella casella di posizione  $rc$  è  $E(Y_{rc}) = \mu_{++}\pi_{rc}$  nel caso di variabile Poisson, o  $E(Y_{rc}|N) = N\pi_{rc}$  nel caso di Poisson condizionata alla somma dei valori di  $Y$  ( $Y$  vettore dei conteggi,  $\mu_{++} = \sum_{r=1}^R \sum_{c=1}^C \mu_{rc}$ ,  $\pi_{rc} = \frac{\mu_{rc}}{\mu_{++}}$ ).

La funzione legame, canonica per entrambi i modelli, è la funzione logaritmo. Il modello, in entrambe le situazioni può essere scritto nel medesimo modo

$$\log(\mu_{rc}) = X\beta.$$

Le funzioni di verosimiglianza dei due modelli differiscono di una quantità non dipendente dai parametri, conseguentemente, è possibile usare le stesse funzioni di verosimiglianza per entrambi i modelli. La funzione di log-verosimiglianza può essere scritta nel modo seguente:

$$l(\beta) = \sum_{i=1}^{R \times C} (y_i \tilde{x}_i \beta - e^{\tilde{x}_i \beta}).$$

La  $r$ -esima componente della funzione score:

$$\frac{\partial l(\beta)}{\partial \beta_r} = \sum_{i=1}^{R \times C} (y_i - e^{\tilde{x}_i \beta}) x_{ir},$$

$r = 1, \dots, p$ . Poiché la funzione legame coincide con il legame canonico, l'informazione osservata è uguale all'informazione attesa; la sua generica componente è

$$[j(\beta)]_{rs} = [i(\beta)]_{rs} = -\frac{\partial^2 l(\beta)}{\partial \beta_r \partial \beta_s} = \sum_{i=1}^{R \times C} e^{\tilde{x}_i \beta} x_{ir} x_{is}.$$

## Capitolo 3

# Riduzione della distorsione tramite funzione score penalizzata

La distorsione è una caratteristica legata al valore atteso di uno stimatore. In particolare, è la differenza tra il valore atteso dello stimatore e il parametro che si prefigge di stimare. La sua misura è data dalla funzione

$$b(\hat{\theta}; \theta) = E(\hat{\theta}) - \theta.$$

Sebbene non sia una caratteristica sufficiente a rendere uno stimatore un buon stimatore, la non distorsione è una proprietà auspicabile, poiché ci assicura che l'“obiettivo” dello stimatore sia almeno in media quello desiderato. Spesso non è possibile utilizzare stimatori non distorti; nel caso dello stimatore di massima verosimiglianza, ad esclusione di alcuni casi, la non distorsione è solamente una proprietà asintotica. Infatti, per lo stimatore di massima verosimiglianza, la distorsione può essere scritta in questo modo:

$$b(\hat{\theta}; \theta) = \frac{b_1(\theta)}{n} + \frac{b_2(\theta)}{n^2} + O(n^{-3}),$$

con  $n$  numero di osservazioni e  $b_1(\theta)$ ,  $b_2(\theta)$  funzioni di ordine  $O(1)$ , per  $n \rightarrow \infty$ . In situazioni in cui il numero di osservazioni sia basso, la distorsione può essere considerevole. Per questo sono stati ideati diversi metodi di correzione della distorsione per stimatori di massima verosimiglianza; essi si dividono in due tipologie: metodi espliciti e metodi impliciti. I primi consistono nel

calcolare la stima di massima verosimiglianza e nel correggerla con una stima della distorsione; metodi di questo tipo sono il jackknife, il bootstrap e il metodo di correzione asintotico; tutti questi metodi sono capaci di rimuovere il termine  $O(n^{-1})$  da  $b(\hat{\theta}; \theta)$ . Nel secondo caso, invece di partire dalle stime, si modifica la funzione score in modo che lo stimatore che ne deriva abbia una distorsione di ordine inferiore; in particolare, nel prossimo paragrafo verrà presentato il metodo implicito capace di rimuovere il termine  $O(n^{-1})$  da  $b(\hat{\theta}; \theta)$ . Per ulteriori informazioni sulla correzione della distorsione, si veda Kosmidis (2014).

### 3.1 Metodo della funzione score penalizzata

La funzione score di un modello regolare ha valore atteso uguale a zero, indipendentemente dalla parametrizzazione utilizzata. A seconda della parametrizzazione, gli stimatori di massima verosimiglianza possono o non possono essere funzione lineare delle osservazioni. Poiché il valore atteso è un operatore lineare, se lo stimatore non è funzione lineare dei dati, esso è distorto. Il ragionamento alla base di questo metodo è: inducendo una piccola distorsione nella funzione score, è possibile rimuovere il termine  $O(n^{-1})$  da  $b(\hat{\theta}; \theta)$ . La funzione score penalizzata sarà

$$\bar{l}_*(\theta) = l_*(\theta) - i(\theta)b(\hat{\theta}; \theta).$$

Per semplificare la spiegazione, verrà presentato di seguito il caso in cui  $\theta$  è scalare. Sia  $\underline{y}$  un campione di osservazioni *i.i.d.* di una variabile casuale con distribuzione  $f(y; \theta)$ . Una generica funzione score penalizzata per questo modello è

$$\bar{l}_*(\theta) = l_*(\theta) + A(\theta).$$

Sia  $\bar{\theta}$  lo stimatore di  $\theta$  ricavato dalla funzione score penalizzata. Riguardo la distorsione di  $\bar{\theta}$ , vale il seguente argomento:

$$E(\bar{\theta} - \theta) = \frac{1}{2}i^{-2}(\theta)(E[l_*^3(\theta)] + E[l_*(\theta)l_{**}(\theta)]) + i^{-1}(\theta)\alpha + O(n^{-\frac{3}{2}}),$$

dove  $\alpha = A(\theta)$  poiché, nel caso di modelli con un solo parametro,  $A(\theta)$  è scalare. Per il primo termine vale

$$\frac{1}{2}i^{-2}(\theta)(E[l_*^3(\theta)] + E[l_*(\theta)l_{**}(\theta)]) = \frac{b_1(\theta)}{n}.$$



Al fine di ridurre la distorsione, deve valere

$$i^{-1}(\theta)\alpha = -\frac{b_1(\theta)}{n} + O(n^{-\frac{3}{2}}).$$

Ad  $\alpha$  deve, quindi, essere imposto il seguente vincolo:

$$\alpha = -i(\theta)\frac{b_1(\theta)}{n} + O(n^{-\frac{3}{2}}).$$

Nel caso presentato, la funzione score penalizzata è

$$\bar{l}_*(\theta) = l_*(\theta) - i(\theta)\frac{b_1(\theta)}{n} + O(n^{-2}).$$

Per approfondimenti, si veda Firth (1993).

## 3.2 Modelli lineari generalizzati con legame canonico: verosimiglianza penalizzata

In modelli lineari generalizzati con legame canonico, la penalizzazione della funzione score assume la seguente forma ( $\theta$  è il vettore dei parametri,  $a_r$  è la  $r$ -esima componente di  $A(\theta)$ ):

$$a_r = \frac{1}{2}tr \left\{ i^{-1} \left( \frac{\partial i}{\partial \theta_r} \right) \right\} = \frac{\partial}{\partial \theta_r} \left\{ \frac{1}{2} \log |i(\theta)| \right\}.$$

Come è facile notare, una sua integrazione in  $\theta$  è immediata. Di conseguenza, è possibile ricavare una funzione di log-verosimiglianza penalizzata integrando la funzione score penalizzata. La sua espressione è

$$\bar{l}(\theta) = l(\theta) + \frac{1}{2} \log |i(\theta)|.$$

Allo stesso modo, è possibile ricavare una funzione di verosimiglianza penalizzata, tramite trasformazione esponenziale della log-verosimiglianza penalizzata:

$$\bar{L}(\theta) = L(\theta) |i(\theta)|^{\frac{1}{2}}.$$

La funzione di penalizzazione  $|i(\theta)|^{\frac{1}{2}}$  è la priori invariante di Jeffreys.

### 3.2.1 La log-verosimiglianza penalizzata nel modello log-lineare

Il modello log-lineare è un modello lineare generalizzato con legame canonico. Definiamo la matrice diagonale di dimensione  $(RC) \times (RC)$  con elementi sulla diagonale  $\mu_i = e^{x_i\beta}$ , con  $i = 1, \dots, R \times C$ :

$$W = \begin{bmatrix} \mu_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mu_{RC} \end{bmatrix}$$

La funzione di log-verosimiglianza penalizzata è, in questo caso:

$$\bar{l}(\beta) = \sum_{i=1}^{RC} (x_i\beta y_i - e^{x_i\beta}) + \frac{1}{2} \log |i(\beta)| = \sum_{i=1}^{RC} (x_i\beta y_i - e^{x_i\beta}) + \frac{1}{2} \log [\det (X^T W X)].$$

Siano  $H = W^{\frac{1}{2}} X (X^T W X)^{-1} X^T W^{\frac{1}{2}}$  e  $h$  il vettore degli elementi diagonali di  $H$ . La funzione score penalizzata è

$$\bar{l}_*(\beta) = X^T \left( y + \frac{1}{2} h - e^{X^T \beta} \right),$$

dove  $\mu = e^{X\beta}$  è il vettore delle medie e  $y$  è il vettore delle osservazioni. Nel caso il modello sia saturo, la matrice  $X$  è quadrata di dimensione  $(RC) \times (RC)$ ; la quantità  $\det (X^T W X)$  può essere scritta come  $|X| |W| |X|$ . La log-verosimiglianza penalizzata diventa, quindi,

$$\begin{aligned} \bar{l}(\beta) &= l(\beta) + \frac{1}{2} \log |W| = l(\beta) + \frac{1}{2} \log \prod_{i=1}^{RC} e^{x_i\beta} = \\ &= l(\beta) + \sum_{i=1}^{RC} \frac{1}{2} x_i\beta = \sum_{i=1}^{RC} \left( x_i\beta \left( y_i + \frac{1}{2} \right) - e^{x_i\beta} \right). \end{aligned}$$

La funzione score penalizzata è

$$\bar{l}_*(\beta) = X^T \left( y + \frac{1}{2} \mathbf{1} - e^{X^T \beta} \right).$$

Come è evidente, l'informazione attesa penalizzata per il modello saturo coincide con l'informazione attesa classica. Essa è

$$\bar{i}(\beta) = i(\beta) = X^T W X.$$

### 3.3 I test di verosimiglianza penalizzata

Per test basati sulla verosimiglianza penalizzata si intendono gli usuali test del log-rapporto di verosimiglianza, di Wald e score, in cui al posto delle funzioni di verosimiglianza vengono utilizzate le funzioni di verosimiglianza penalizzata. Sia  $\beta = (\tau, \psi)$  il vettore dei  $p$  parametri di un modello log-lineare saturo per tabelle a doppia entrata,  $M_{sat}$ . Sia  $\tau$  il vettore dei primi  $p_0 < p$  parametri di  $\beta$ , relativi all'effetto delle modalità delle variabili categoriali senza interazione, e  $\psi$  quello dei rimanenti  $p - p_0$  parametri, relativi all'interazione fra le due variabili categoriali. Il modello con i soli parametri  $\tau$  è il modello ridotto,  $M_{rid}$ . Il test di ipotesi

$$\begin{cases} H_0 : M_{rid} \equiv M_{sat}, \\ H_1 : M_{rid} \not\equiv M_{sat}, \end{cases}$$

può essere scritto nel seguente modo:

$$\begin{cases} H_0 : \psi = \psi_0 = 0, \\ H_1 : \psi \neq 0. \end{cases}$$

Il test del log-rapporto di verosimiglianza penalizzata è

$$\bar{W}_p = 2(\bar{l}(\bar{\tau}, \bar{\psi}) - \bar{l}(\bar{\tau}_{\psi_0}, \psi_0)).$$

con  $\bar{\tau}$  e  $\bar{\psi}$  stime di massima verosimiglianza penalizzata e  $\bar{\tau}_{\psi_0}$  stima di massima verosimiglianza penalizzata di  $\tau$ , posto  $\psi = \psi_0$ .

Il test di Wald penalizzato ha la seguente forma:

$$\bar{W}_e = (\bar{\psi} - \psi_0)^T \{[\bar{i}(\beta_0)^{-1}]_{(\psi, \psi)}\}^{-1} (\bar{\psi} - \psi_0),$$

con  $\beta_0 = (\bar{\tau}_{\psi_0}, \psi_0)$  e  $\bar{i}(\beta_0)$  informazione attesa penalizzata sotto  $H_0$ .

Sia  $\bar{l}_*(\beta) = \left[ \frac{\partial \bar{l}(\beta)}{\partial \tau}, \frac{\partial \bar{l}(\beta)}{\partial \psi} \right] = [\bar{l}_\tau, \bar{l}_\psi]$  la funzione score penalizzata. Il test score penalizzato è:

$$\bar{W}_u(\psi_0) = \bar{l}_{\psi_0}^T [\bar{i}(\beta_0)^{-1}]_{(\psi, \psi)} \bar{l}_{\psi_0}.$$

Le distribuzioni asintotiche dei test non cambiano se  $\bar{i}(\beta_0)$  viene stimato da  $\bar{i}(\bar{\beta})$ . Per questi test viene assunta la medesima distribuzione asintotica chi-quadrato con  $p - p_0$  gradi di libertà.

### 3.4 Obiettivo

Come visto, nei modelli lineari generalizzati con legame canonico, il metodo della funzione score modificata non solo permette di ottenere stimatori con distorsione di ordine inferiore, ma porta anche a una nuova serie di funzioni di verosimiglianza penalizzata. La domanda alla base di questa relazione è: quali sono le proprietà dei test basati sulla verosimiglianza penalizzata rispetto agli usuali test di verosimiglianza? In particolare, nel prossimo capitolo verrà esposto il metodo di simulazione adottato per studiare il comportamento dei test basati sulla verosimiglianza penalizzata in modelli log-lineari.

# Capitolo 4

## Metodi di simulazione

Al fine di confrontare il comportamento dei test basati sulla verosimiglianza classica e penalizzata in modelli log-lineari, si è implementato uno studio di simulazione. In questo capitolo verranno presentati i passi principali di questo studio. Il codice R utilizzato si trova nell'Appendice [A.1](#).

### 4.1 Presentazione dell'algoritmo

La procedura consiste nel calcolare, per un numero elevato  $Nsim$  di campioni generati casualmente sotto l'ipotesi nulla, i  $p$ -value di test di verosimiglianza e verosimiglianza penalizzata. I test che si confronteranno sono i tre di verosimiglianza e i corrispondenti tre di verosimiglianza penalizzata. I test vengono fatti per verificare l'equivalenza fra il modello saturo,  $M_{sat}$ , e il modello ridotto,  $M_{rid}$ . Le ipotesi nulla e alternativa sono

$$\begin{cases} H_0 : M_{rid} \equiv M_{sat}, \\ H_1 : M_{rid} \not\equiv M_{sat}. \end{cases}$$

Utilizzando la notazione del Paragrafo 3.3:

$$\begin{cases} H_0 : \psi = \psi_0 = 0, \\ H_1 : \psi \neq 0. \end{cases}$$

Alla fine della simulazione, per ogni test saranno disponibili  $Nsim$   $p$ -value. Poiché i test vengono calcolati sapendo che  $H_0$  è vera (il modello ridotto-

to è equivalente al modello saturo, poiché le tabelle sono generate da quel modello), la distribuzione dei  $p$ -value dovrebbe essere uniforme  $(0, 1)$ .

## 4.2 Generazione di tabelle di contingenza

Le tabelle utilizzate sono state generate da un modello log-lineare prestabilito. I parametri del modello di partenza vengono sempre ricavati da una tabella reale, che funge da sorgente per tutte le altre. Quindi, le frequenze vengono generate da una distribuzione multinomiale, in modo che  $N$ , la somma delle frequenze, sia uguale per tutte le tabelle. I passi sono:

- Partendo da una tabella a doppia entrata reale, si calcolano i parametri del modello log-lineare del tipo  $\log(\mu_{rc}) = \beta_1 + \beta_r^A + \beta_c^B$  (per riferimenti sulla notazione si veda il Capitolo 2).
- Utilizzando i parametri appena ricavati si ottiene il vettore di dimensione  $(RC) \times 1$   $\hat{\mu}_0 = e^{X\hat{\beta}_0}$ , con  $\hat{\beta}_0 = (\hat{\tau}_{\psi_0}, \psi_0)$ .
- Conoscendo il vettore  $\hat{\mu}_0$  è possibile calcolare le stime di massima verosimiglianza del parametro vettoriale  $\hat{\pi}_0$ , le cui componenti sono  $\hat{\pi}_{i0} = \frac{\hat{\mu}_{i0}}{\sum_{i=1}^{RC} \hat{\mu}_{i0}}$ . Esso è il vettore delle probabilità per il modello log-lineare multinomiale.
- A questo punto, una volta deciso quale sia la somma dei conteggi desiderata,  $N$ , è possibile generare valori pseudo-casuali da una multinomiale di parametri  $N$  e  $\pi = \hat{\pi}_0$ , in modo da ottenere il vettore delle frequenze che costituiranno la nuova tabella.
- Le frequenze vengono organizzate in un data-frame, di cui fanno parte anche le variabili categoriali, in modo che l' $i$ -esima riga contenga la frequenza e le modalità assunte da  $A$  e  $B$  per la  $i$ -esima casella, con  $i = 1, \dots, R \times C$ .

Poiché le log-verosimiglianze fra di esse sono equivalenti, generare da una Poisson o da una multinomiale non comporta differenze.

## 4.3 Analisi di ogni tabella

Una volta generata una tabella vengono svolte le operazioni di analisi necessarie ad ottenere i risultati cercati. Innanzitutto, vengono stimati i parametri di quattro diversi modelli log-lineari: modello saturo e modello ridotto, sia per verosimiglianza classica che per verosimiglianza penalizzata. Per fare ciò, la funzione di log-verosimiglianza (classica o penalizzata) viene massimizzata numericamente nei parametri. Ricordiamo che nel modello saturo il numero di parametri coincide col numero di caselle della tabella in esame ( $R \times C$  secondo la notazione adottata in precedenza); nel modello ridotto, invece, il numero di parametri è  $R + C - 1$ . Poiché l'obiettivo è testare l'equivalenza del modello ridotto e di quello saturo, il numero di parametri da testare sarà  $R \times C - R - C + 1 = (R - 1)(C - 1)$ .

Stimati i parametri, il passo successivo consiste nel calcolare i test di verosimiglianza. Le procedure utilizzate sono le stesse di quelle esposte in precedenza. Per il test del log-rapporto di verosimiglianza, le stime dei parametri vengono utilizzate per calcolare il valore della log-verosimiglianza (classica o penalizzata), al fine di confrontare le verosimiglianze del modello saturo e del modello ridotto. Nel test di Wald, le stime dei parametri di interesse vengono usate direttamente per un confronto con i loro valori sotto  $H_0$ , in questo caso 0. Per questo tipo di test di verosimiglianza classica, è necessario calcolare la quantità  $i^{-1}(\hat{\beta})$ . Questa operazione in alcuni casi non è possibile. La forma dell'informazione attesa per modelli log-lineari è  $i(\beta) = X^T W X$ , con  $W$  matrice diagonale sulla cui diagonale sono presenti i valori stimati delle medie,  $\hat{\mu}$ . Poiché nel modello saturo i valori stimati coincidono con quelli osservati, nel caso uno dei conteggi della tabella sia zero, anche la stima della media per quella casella è 0; in questo modo, la matrice  $W$  avrà una colonna e una riga di zeri. Poiché, in questo caso,  $W$  ha determinante zero, per le proprietà del determinante, anche  $i(\beta)$  ha determinante zero;  $i(\beta)$  è, conseguentemente, non invertibile. Il test, quindi, non può essere calcolato se la tabella ha valori pari a zero. Nello studio, si tiene traccia del numero di volte in cui questa situazione si verifica. In casi del genere il  $p$ -value del test viene considerato come mancante. Il problema non si verifica nel caso si utilizzi la verosimiglianza penalizzata. Nel test score,

occorre calcolare  $i(\beta_0)$ , dove  $\beta_0$  è il vettore dei parametri sotto  $H_0$ . Poiché  $H_0$  formula un'ipotesi solo per una parte dei parametri,  $\beta_0$  è formato dalla stima di massima verosimiglianza profilo (classica o penalizzata) dei parametri non considerati da  $H_0$  e dai valori ipotizzati da quest'ipotesi per i restanti.

## 4.4 Metodi di sintesi delle informazioni

La procedura presentata produce una grande quantità di dati. Per poter dare loro il giusto valore, occorre utilizzare dei metodi di sintesi validi. Essi consistono in grafici e tabelle.

I grafici che hanno la maggiore capacità di rappresentare quanto la distribuzione effettiva dei  $p$ -value si avvicini alla loro distribuzione teorica sono i grafici quantile-quantile. Essi rappresentano sulle ascisse i quantili teorici (in questo caso i quantili di una uniforme  $(0, 1)$ ) e sulle ordinate i  $p$ -value ordinati. I punti mostrano l'associazione fra il quantile teorico e il relativo quantile empirico. Poiché sono di particolare interesse i quantili solitamente usati come  $\alpha$  fissato nei test di ipotesi ( $0.01, 0.025, 0.05, 0.10$ ), è utile, con la stessa metodologia del grafico precedente, confrontare i quantili teorici con quelli empirici per questi valori. Se la distribuzione empirica coincide con quella teorica, i punti saranno allineati sulla bisettrice. Nel caso stiano sotto la bisettrice, il test in esame ha una probabilità di rifiutare  $H_0$  quando essa è vera (errore del primo tipo) più grande di quanto dovrebbe essere. Nel caso in cui stiano sopra la bisettrice, al contrario, il test rifiuta  $H_0$  con probabilità troppo bassa. Se i quantili sono eccessivamente distanti dalla bisettrice, il test è da considerarsi non buono.

Le tabelle di sintesi utilizzate sono due. La prima mette a confronto, per ogni  $p$ -value, i quantili empirici relativi ai valori solitamente assunti dall' $\alpha$  fissato. Essa non è altro che la trasposizione numerica del grafico di cui si è appena parlato. La seconda tabella confronta i veri valori dei parametri con la media di quelli ottenuti nei campioni simulati, utilizzando la verosimiglianza classica e quella penalizzata. In questo modo, è possibile apprezzare in quale misura il metodo della verosimiglianza penalizzata riduca in media la distorsione degli stimatori di massima verosimiglianza.



# Capitolo 5

## Risultati

In questo capitolo verranno riportate le sintesi dei dati ottenuti dagli studi di simulazione e le considerazioni che ne conseguono.

### 5.1 Le tabelle reali

Le tabelle che hanno avuto il ruolo di ispirare tutte le generazioni casuali sono due. La prima, Tabella 5.1, consiste nella catalogazione di un gruppo

	Affective disorders	Schizophrenics	Neurotics
Retarded	12	13	5
No retarded	18	17	25

**Tabella 5.1:** Pazienti con malattie psichiatriche

di 90 persone con disturbi psichiatrici. I pazienti vengono classificati secondo le modalità assunte da due variabili categoriali. La prima variabile, Activity (Ac), riguarda l'attività mentale dell'individuo; essa può essere ritardata o non ritardata. La seconda, Psychiatric (Ps), ha tre livelli che distinguono le unità statistiche secondo il loro disturbo psichiatrico; esso può essere: disturbo affettivo, schizofrenia o nevrosi. I parametri del modello log-lineare ridotto, calcolati per questa tabella, sono riportati nella Tabella 5.2. Le tabelle utilizzate nello studio di simulazione, quindi, avranno come modello generatore dei dati il modello log-lineare con i parametri appena presentati.

$\beta_1$	2.303
$\beta_2^{Ac}$	0.693
$\beta_2^{Ps}$	0
$\beta_3^{Ps}$	0

**Tabella 5.2:** Coefficienti del primo modello ridotto

I dati del secondo esempio sono riportati in Tabella 5.3; essa è composta da 4 righe e 4 colonne. Le unità statistiche sono 91 coppie monogame. Esse sono ordinate secondo le modalità assunte da due variabili casuali: la soddisfazione sessuale del partner maschile (M) e del partner femminile (F). Le variabili sono qualitative ordinali in modo crescente: la prima corrisponde al minor appagamento e l'ultima al maggiore. Le loro modalità sono A, B, C e D. I parametri del modello ridotto che verranno usati per la generazione

	A	B	C	D
A	7	7	2	3
B	2	8	3	7
C	1	5	4	9
D	2	8	9	14

**Tabella 5.3:** Soddisfazione sessuale dei partner

delle tabelle sono riportati nella Tabella 5.4.

$\beta_1$	0.918
$\beta_B^M$	0.051
$\beta_C^M$	0
$\beta_D^M$	0.552
$\beta_B^F$	0.847
$\beta_C^F$	0.405
$\beta_D^F$	1.012

**Tabella 5.4:** Coefficienti del secondo modello ridotto

## 5.2 Studi di simulazione

Le simulazioni sono state svolte in modo da evidenziare le differenze causate da un moderato o basso numero di unità statistiche. Per ogni tabella e numerosità iniziale si sono generati 10000 campioni. Prima, sia per la Tabella 5.1 che per la Tabella 5.3 sono state simulate tabelle con un numero di conteggi uguale a quello delle tabelle originali (90 e 91, rispettivamente). Poi, il numero di unità è stato ridotto a 30 per entrambi gli studi di simulazione.

### 5.2.1 Numerosità originale

I test associati alle *Nsim* tabelle generate dai parametri del modello ridotto associato alla Tabella 5.1 hanno prodotto i grafici in Figura 5.1. Come si può notare, tutti i *p*-value hanno un comportamento molto buono. Infatti, i punti dei grafici quantile-quantile sono allineati sulla bisettrice. Le assunzioni fatte sulla distribuzione dei test, quindi, paiono confermate. Più nello specifico, nella Tabella 5.5 sono riportati i quantili empirici corrispondenti ai valori di  $\alpha$  fissato usualmente utilizzati. Osservando i valori numerici assunti

$\alpha$	W	W penalizzato	Wald	Wald penalizzato	Score	Score penalizzato
1%	0.83	1.05	1.62	1.85	1.04	1.28
2.5%	2.14	2.55	3.40	3.73	2.47	2.91
5%	4.73	5.48	6.09	6.65	5.14	5.79
10%	9.39	10.48	10.99	11.86	9.93	10.94

**Tabella 5.5:** Quantili empirici ( $\times 100$ ) relativi ai valori tipici dell' $\alpha$  fissato

dai quantili di interesse, si può notare come quelli dei test penalizzati siano sempre leggermente più alti di quelli dei rispettivi test classici. Ad ogni modo, le differenze sono minime. In questo caso, l'utilizzo della verosimiglianza penalizzata piuttosto che quello della verosimiglianza classica è ingiustificato, poiché i due metodi portano a conclusioni quasi certamente analoghe. A conferma di ciò, nella Tabella 5.6 sono riportati i valori dei parametri stimati secondo la verosimiglianza classica e quella penalizzata, affiancati dai loro veri valori. Come si può notare, la distorsione per entrambi i casi è molto bassa.

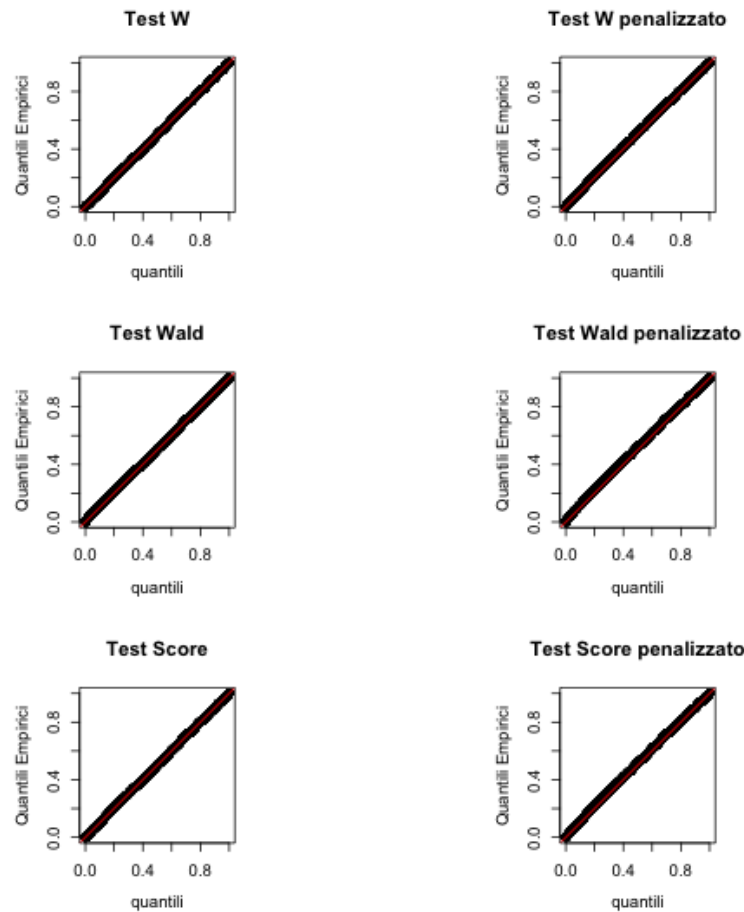
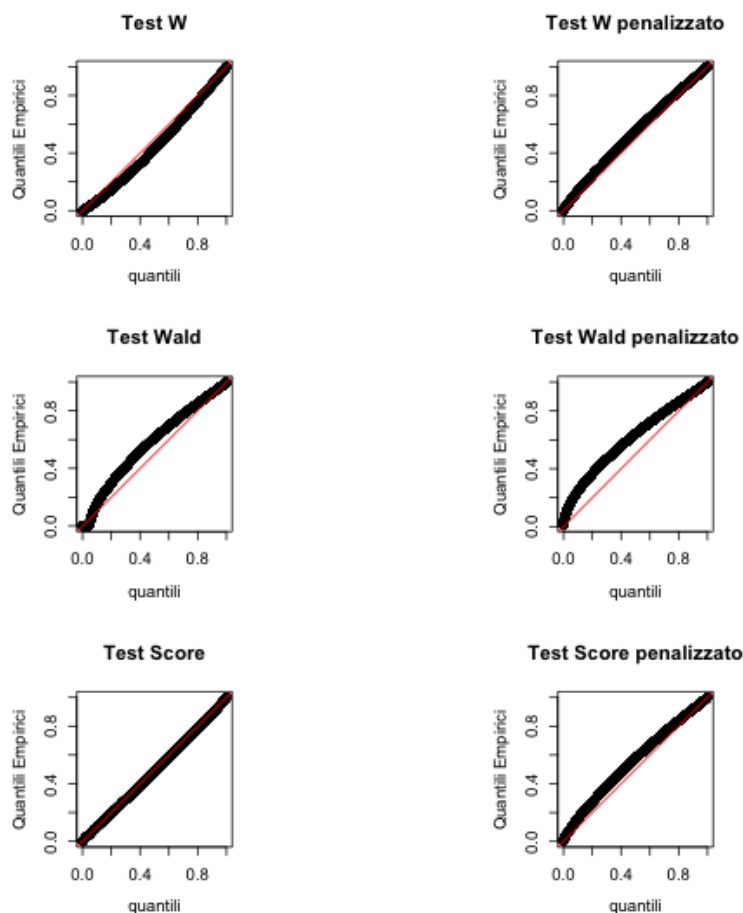


Figura 5.1: qqplot Tabella 5.1,  $N=90$

Per quanto riguarda gli esiti delle simulazioni partendo dalla Tabella 5.3, i grafici sono riportati nella Figura 5.2. I punti non sono sempre disposti perfettamente sulla bisettrice. Una tendenza che si può osservare, seppur non così evidente, è quella dei  $p$ -value dei test penalizzati di essere sempre “sopra” i  $p$ -value dei test calcolati con verosimiglianza classica. Concentrando la nostra attenzione sui quantili di maggiore importanza pratica, essi sono riportati nella Tabella 5.7, e una loro trasposizione grafica si trova alla Figura 5.3. I grafici in Figura 5.3 confermano che, a parità di tipologia di test e di quantile teorico, i test penalizzati hanno quantili empirici dai valori più alti. Il test che pare comportarsi in maniera migliore è il test score. Nel caso

Figura 5.2: qqplot Tabella 5.3,  $N=91$ 

della verosimiglianza classica il test è completamente in linea con ciò che ci si aspetta da un buon test, nel caso della verosimiglianza penalizzata la situazione peggiora leggermente. Il test  $W$  ha un comportamento discreto, anche se, per entrambi i tipi di verosimiglianza, non è pienamente soddisfacente; utilizzando la verosimiglianza classica, si ottiene un test che rifiuta troppo spesso (i punti giacciono sotto la bisettrice), mentre, con la verosimiglianza penalizzata, un test che rifiuta troppo poco spesso (i punti sono sopra la bisettrice). La scelta di un metodo migliore è quindi molto difficile. Nel caso del test di Wald, la scelta è ancora più complessa, poiché i limiti del test  $W$  sono presenti in maniera enfatizzata: le distanze dei punti dalla bisettrice

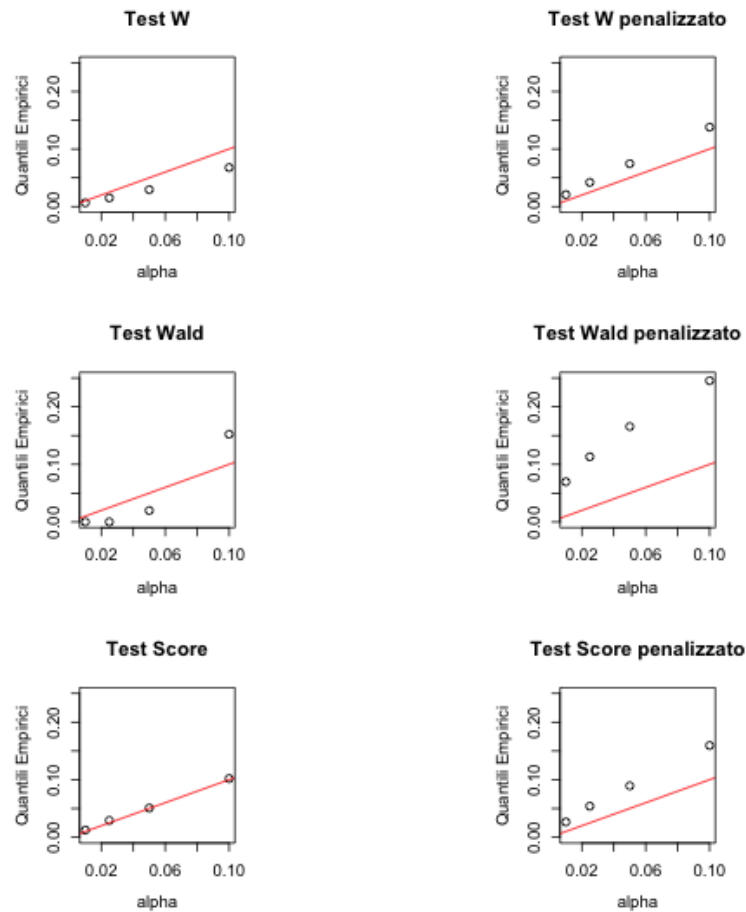


Figura 5.3: Quantili rilevanti Tabella 5.3, N=91

sono maggiori. Si ricorda che il test di Wald classico non può essere effettuato nel caso sia presente un conteggio pari a zero nella tabella analizzata (si veda il Paragrafo 4.3); per quanto riguarda questa simulazione, questo test non è stato eseguito per 39 volte. Sebbene, in questo frangente, la verosimiglianza penalizzata non abbia portato a miglioramenti oggettivi nei test, senza dubbio ha portato ad una notevole riduzione della distorsione. Infatti, come riportato nella Tabella 5.8, la distorsione è sostanzialmente inferiore nel caso si utilizzi la verosimiglianza penalizzata.

### 5.2.2 Numerosità ridotta

A partire dalla Tabella 5.1, la cui somma dei conteggi è 90, è stato condotta una simulazione in cui tutte le tabelle generate hanno numerosità totale 30. In Figura 5.4 sono riportati i grafici quantile-quantile. Le differenze fra i

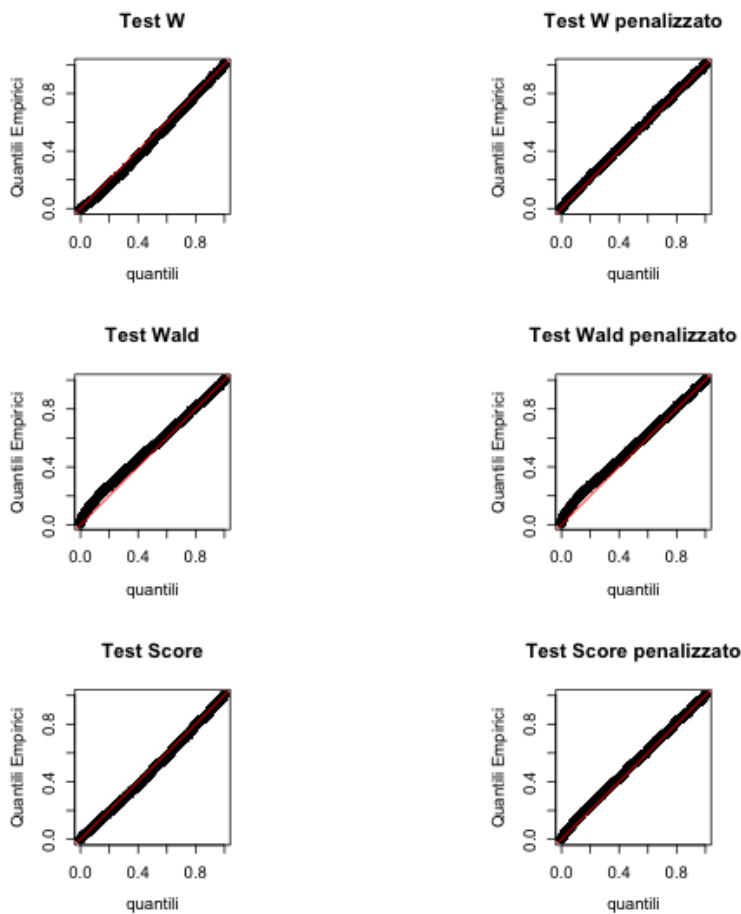


Figura 5.4: qqplot Tabella 5.1, N=30

test non sono evidenti a questo livello; l'unico che sembra avere dei quantili dei  $p$ -value non perfettamente allineati sulla bisettrice è il test di Wald, sia con verosimiglianza classica che penalizzata. Per quanto riguarda questo test nella sua versione classica, esso non è stato eseguito per 1 volta. Andando più nel dettaglio, le considerazioni che si possono trarre dai Grafici in Figu-

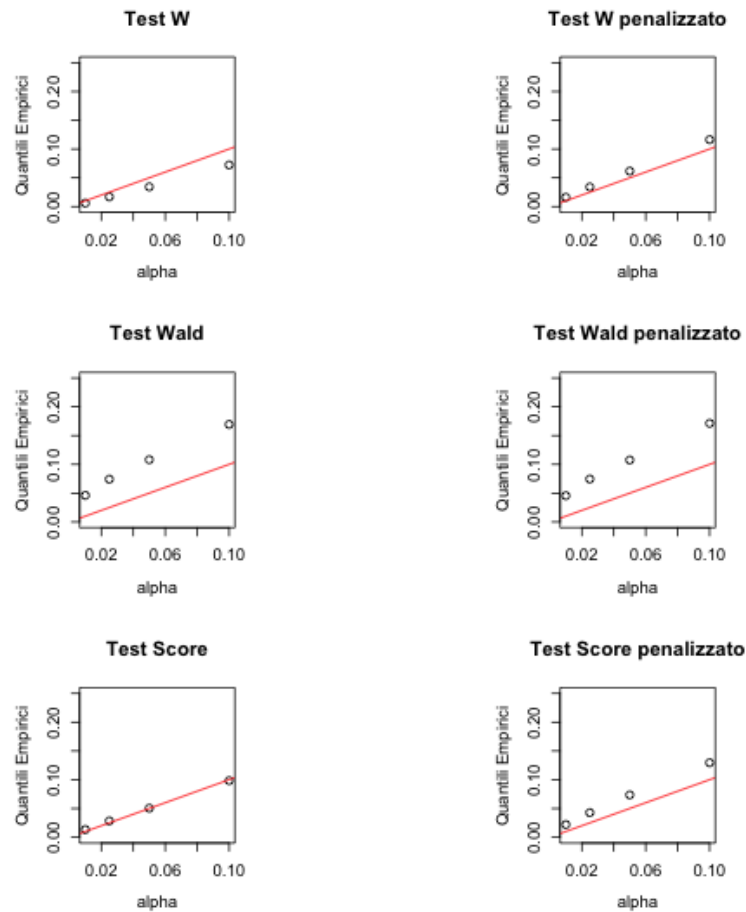


Figura 5.5: Quantili rilevanti Tabella 5.1,  $N=30$

ra 5.5 e dalla Tabella 5.9 sono analoghe a quelle fatte finora. Il test score classico è ancora il migliore; la sua versione penalizzata è lievemente troppo incline all'accettare l'ipotesi nulla. Il test  $W$ , seppur abbastanza preciso, rifiuta troppo nel caso classico e troppo poco in quello penalizzato. Il test di Wald risulta di nuovo il peggiore: in tutti e due i tipi di verosimiglianza risulta troppo conservativo. In una situazione del genere, anche se la numerosità del campione è bassa, l'utilizzo della verosimiglianza penalizzata non comporta dei vantaggi, dal punto di vista dei test. Per quanto riguarda la distorsione, riportata in Tabella 5.10, essa è lievemente inferiore nel caso della verosimiglianza penalizzata.



Le differenze fra i test con verosimiglianza classica e penalizzata si fanno più significative in quest'ultimo caso, in cui a partire dalla Tabella 5.3 sono state generate tabelle con numerosità complessiva pari a 30. In Figura 5.6,

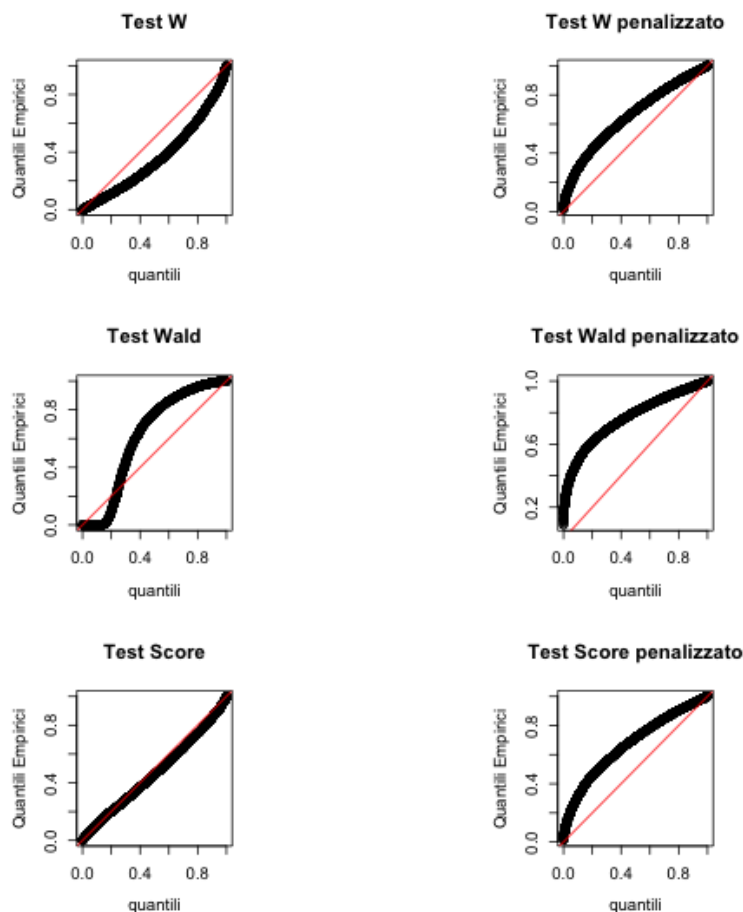


Figura 5.6: qqplot Tabella 5.3,  $N=30$

infatti, è chiaro come i quantili empirici dei  $p$ -value, nei test penalizzati, siano di valore maggiore di quelli nei corrispettivi test classici. Il test di Wald si conferma come il test peggiore: non solo i  $p$ -value non hanno palesemente distribuzione uniforme, ma il test stesso, nel caso della verosimiglianza classica, non è stato eseguito per 5778 volte, più della metà delle volte. Il test  $W$  non ha un buon adattamento in entrambe le sue forme: nel caso classico i quantili dei  $p$ -value sono sempre sotto la bisettrice, nel caso penalizzato

sempre sopra. Il test score classico, anche in una situazione in cui la numerosità è bassa, si comporta in modo adeguato. Al contrario, la sua versione penalizzata è eccessivamente permissiva, accettando troppo spesso l'ipotesi nulla. Queste considerazioni sono confermate dai grafici nella Figura 5.7 e dalla Tabella 5.11. La distorsione delle stime, riportata in Tabella 5.12, è,

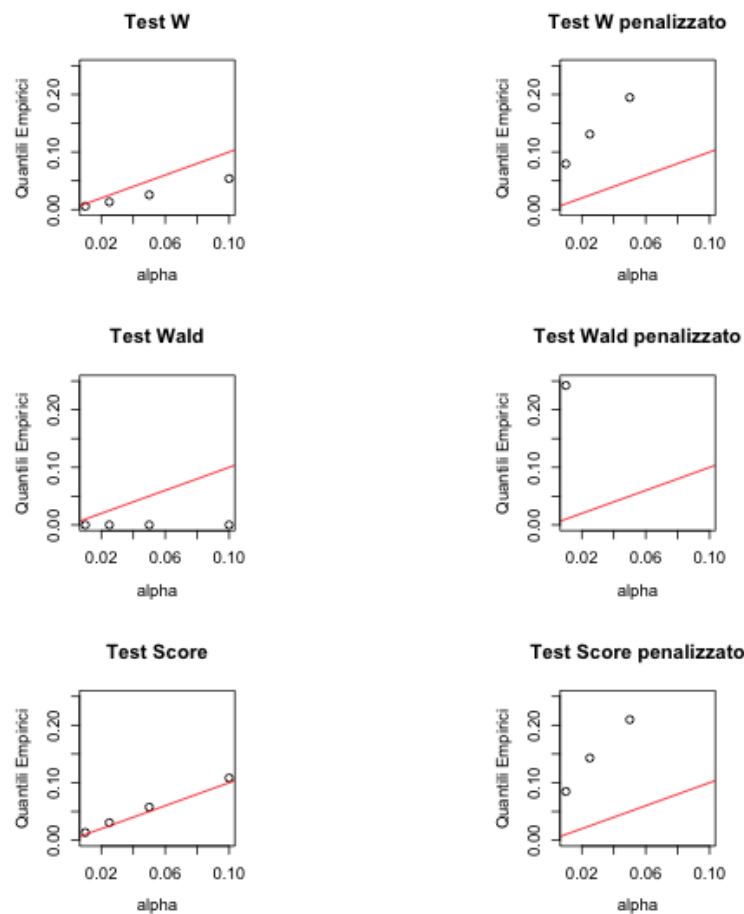


Figura 5.7: Quantili rilevanti Tabella 5.1,  $N=30$

in alcuni casi, molto grande per la verosimiglianza classica; per le stime di massima verosimiglianza penalizzata, la distorsione media non è mai maggiore di 1. Il metodo della verosimiglianza penalizzata si conferma un buon metodo per la diminuzione della distorsione.

## 5.3 Considerazioni finali

Lo studio presentato suggerisce alcune considerazioni riguardo gli effetti della distorsione, la natura dei test e la verosimiglianza penalizzata.

La tendenza che si può osservare è che, nei test  $W$  e di Wald classici, maggiore è la distorsione dei parametri, maggiore è la deviazione dagli assunti distributivi dei  $p$ -value. Ciò si verifica poiché questi test dipendono dalle stime dei parametri. Il test  $W$ , nella sua forma  $W_p(\psi_0) = 2\{l(\hat{\tau}, \hat{\psi}) - l(\hat{\tau}_{\psi_0}, \psi_0)\}$ , dipende sia dalle stime di massima verosimiglianza del vettore  $(\tau, \psi)$ , sia dalle stime di massima verosimiglianza vincolata del parametro di disturbo  $\tau_{\psi_0}$ . Il test di Wald, invece, dipende ancora maggiormente dalle stime, poiché oltre alle stime di massima verosimiglianza, utilizza anche l'informazione osservata, anch'essa dipendente dalle stime. Per questo motivo, il test di Wald è quello che ha il comportamento peggiore. Il test score classico, invece, è l'unico ad essere sempre affidabile. Infatti, esso non dipende in alcun modo dalle stime dei parametri, ma solo dal valore dei parametri sotto  $H_0$ .

I test basati sulla verosimiglianza penalizzata non si sono mai rivelati migliori dei corrispettivi test in versione classica. La ragione di ciò può essere una sbagliata assunzione di distribuzione della statistica test. Ad esempio, le ipotesi sulla distribuzione della statistica test score si basa sul fatto che  $l_*(\beta) \sim N(0, i(\beta))$ . Poiché la funzione score penalizzata è nella forma  $\bar{l}_*(\beta) = l_*(\beta) - i(\hat{\beta}) \frac{\hat{b}_1(\beta)}{n}$ , la distribuzione della funzione score penalizzata non ha media zero. Tuttavia, poiché per  $n \rightarrow \infty$  il termine  $\frac{\hat{b}_1(\beta)}{n}$  tende a zero, per  $n$  sufficientemente grande, la media può essere considerata zero. Alla luce di questa considerazione, è ragionevole supporre che la distribuzione effettiva dei test di verosimiglianza penalizzata dipenda dalla grandezza della distorsione; infatti, nello studio è emerso che più grande è la distorsione, maggiore è il distacco delle distribuzioni empiriche dei test di verosimiglianza penalizzata da quelle teoriche. Per poter utilizzare dei test basati sulla verosimiglianza penalizzata con successo, quindi, potrebbe essere utile studiare il rapporto fra la distribuzione dei test e la quantità  $\frac{\hat{b}_1(\beta)}{n}$ .

In conclusione, il metodo utilizzato in questo studio non comporta vantaggi se non nella correzione della distorsione. Dal punto di vista dei test, esso viene bocciato con riserva: modificando gli assunti distributivi, i test di

verosimiglianza penalizzata potrebbero essere migliori di quelli classici.

	Veri $\beta$	$\hat{\beta}$	$\bar{\beta}$	$B\hat{I}AS$	$B\bar{I}AS$
$\beta_1$	2.303	2.256	2.309	0.047	-0.007
$\beta_2^{Ac}$	0.693	0.721	0.693	-0.028	0.000
$\beta_2^{Ps}$	0	-0.003	-0.001	0.003	0.001
$\beta_3^{Ps}$	0	-0.001	-0.001	0.001	0.001
$\beta_2^{AcPs}$	0	0.000	-0.001	0.000	0.001
$\beta_3^{AcPs}$	0	-0.001	-0.001	0.001	0.001

Tabella 5.6: Stime dei parametri e della distorsione

$\alpha$	W	W penalizzato	Wald	Wald penalizzato	Score	Score penalizzato
1%	0.65	2.09	0.00	6.97	1.24	2.66
2.5%	1.50	4.21	0.00	11.32	2.93	5.40
5%	2.94	7.46	1.98	16.58	5.08	8.94
10%	6.79	13.80	15.23	24.55	10.21	15.95

Tabella 5.7: Quantili empirici ( $\times 100$ ) relativi ai valori tipici dell' $\alpha$  fissato

	Veri $\beta$	$\hat{\beta}$	$\bar{\beta}$	$B\hat{I}AS$	$B\bar{I}AS$
$\beta_1$	0.918	-0.901	0.944	1.820	-0.026
$\beta_B^M$	0.051	0.432	0.042	-0.381	0.009
$\beta_C^M$	0	0.148	-0.021	-0.148	0.021
$\beta_D^M$	0.552	2.026	0.532	-1.474	0.020
$\beta_B^F$	0.847	2.546	0.823	-1.699	0.024
$\beta_C^F$	0.405	1.677	0.387	-1.272	0.019
$\beta_D^F$	1.012	2.742	0.988	-1.731	0.023
$\beta_{BB}^{MF}$	0	-0.410	0.007	0.410	-0.007
$\beta_{CB}^{MF}$	0	-0.169	0.024	0.169	-0.024
$\beta_{DB}^{MF}$	0	-1.401	0.025	1.401	-0.025
$\beta_{BC}^{MF}$	0	-0.329	0.007	0.329	-0.007
$\beta_{CC}^{MF}$	0	-0.204	0.027	0.204	-0.027
$\beta_{DC}^{MF}$	0	-1.030	0.017	1.030	-0.017
$\beta_{BD}^{MF}$	0	-0.371	0.010	0.371	-0.010
$\beta_{CD}^{MF}$	0	-0.165	0.018	0.165	-0.018
$\beta_{DD}^{MF}$	0	-1.423	0.024	1.423	-0.024

Tabella 5.8: Stime dei parametri e della distorsione

$\alpha$	W	W penalizzato	Wald	Wald penalizzato	Score	Score penalizzato
1%	0.59	1.62	4.61	4.56	1.33	2.21
2.5%	1.70	3.41	7.42	7.45	2.84	4.27
5%	3.43	6.17	10.81	10.76	5.05	7.36
10%	7.23	11.63	16.95	17.13	9.84	12.95

**Tabella 5.9:** Quantili empirici ( $\times 100$ ) relativi ai valori tipici dell' $\alpha$  fissato

	Veri $\beta$	$\hat{\beta}$	$\bar{\beta}$	$B\hat{I}AS$	$B\bar{I}AS$
$\beta_1$	2.303	0.532	1.225	1.770	1.077
$\beta_2^{Ac}$	0.693	1.286	0.687	-0.593	0.006
$\beta_2^{Ps}$	0	-0.029	-0.011	0.029	0.011
$\beta_3^{Ps}$	0	-0.022	-0.008	0.022	0.008
$\beta_2^{AcPs}$	0	0.020	0.010	-0.020	-0.010
$\beta_3^{AcPs}$	0	0.028	0.009	-0.028	-0.009

**Tabella 5.10:** Stime dei parametri e della distorsione

$\alpha$	W	W penalizzato	Wald	Wald penalizzato	Score	Score penalizzato
1%	0.57	7.94	0	24.24	1.35	8.48
2.5%	1.32	13.10	0	31.53	3.05	14.29
5%	2.56	19.48	0	39.52	5.75	20.97
10%	5.39	29.26	0	49.46	10.82	31.32

**Tabella 5.11:** Quantili empirici ( $\times 100$ ) relativi ai valori tipici dell' $\alpha$  fissato

	Veri $\beta$	$\hat{\beta}$	$\bar{\beta}$	$B\hat{I}AS$	$B\bar{I}AS$
$\beta_1$	0.918	-9.546	0.051	10.465	0.867
$\beta_B^M$	0.051	-2.234	0.030	2.286	0.021
$\beta_C^M$	0	-2.871	-0.001	2.871	0.001
$\beta_D^M$	0.552	3.384	0.393	-2.832	0.159
$\beta_B^F$	0.847	6.192	0.654	-5.345	0.193
$\beta_C^F$	0.405	1.701	0.284	-1.296	0.121
$\beta_D^F$	1.012	7.257	0.782	-6.246	0.229
$\beta_{BB}^{MF}$	0	1.350	0.007	-1.350	-0.007
$\beta_{CB}^{MF}$	0	1.065	-0.020	-1.065	0.020
$\beta_{DB}^{MF}$	0	-0.213	0.122	0.213	-0.122
$\beta_{BC}^{MF}$	0	-0.040	0.011	0.040	-0.011
$\beta_{CC}^{MF}$	0	-0.436	-0.009	0.436	0.009
$\beta_{DC}^{MF}$	0	1.150	0.080	-1.150	-0.080
$\beta_{BD}^{MF}$	0	2.081	0.037	-2.081	-0.037
$\beta_{CD}^{MF}$	0	2.063	0.005	-2.063	-0.005
$\beta_{DD}^{MF}$	0	-0.391	0.160	0.391	-0.160

Tabella 5.12: Stime dei parametri e della distorsione





# Bibliografia

Firth, D. «Bias Reduction of Maximum Likelihood Estimates». In: *Biometrika* 80 (1993), pp. 27–38.

Kosmidis, I. «Bias in parametric estimation: reduction and useful side-effects». In: *WIRE, Computational Statistics* 6 (2014), pp. 185–196.

Menardi, G. *Il modello log-lineare, Dispense didattiche*. Università di Padova, 2013-2014.

Pace, L. e A. Salvan. *Introduzione alla statistica II*. Milano: Cedam, 2001.

Piccolo, D. *Statistica*. Il Mulino, 2010.



# Appendice A

## Codice R utilizzato

### Codice A.1: Funzioni R e loro utilizzo

---

```
psycho <- matrix(c(12,18,13,17,5,25),nrow=2)
```

```
psycho <- as.table(psycho)
```

```
psycho.df <- as.data.frame.table(psycho)
```

```
y1 <- c(7,2,1,2,7,8,5,8,2,3,4,9,3,7,9,14)
```

```
y1 <- as.table(matrix(y1,nrow=4))
```

```
y1.df <- as.data.frame.table(y1)
```

```
rtabBin<-function(beta,X,n)
```

```
{
```

```
  eta=as.vector(X%*%beta)
```

```
  mu=exp(eta)
```

```
  pi=mu/sum(mu)
```

```
  rmultinom(1,n,pi)
```

```
}
```

```
nlogL=function(beta,data,X)
```

```
{# -logver
```

```
  eta=as.vector(X%*%beta)
```

```

    -sum(eta*data-exp(eta))
  }

nlogL.pen=function(beta,data,X)
{# -logver penalizzata
  eta=as.vector(X**beta)
  pen=determinant(t(X)**diag(exp(eta))**X,log=TRUE)$modulus
  -sum(eta*data-exp(eta))-0.5*pen
}

wald.test=function(psi.hat,psi0,Jhat.inv.psi)
{
  1-pchisq( t(psi.hat-psi0)**solve(Jhat.inv.psi)**(psi.hat-psi0),length(
    psi0) )
}

score=function(beta,data,X)
{
  mu=as.vector(exp(X**beta))
  as.vector(t(X)**(data-mu))
}

score.bc=function(beta,data,X)
{# bias modified score function for Poisson model
  mu=as.vector(exp(X**beta))
  W=diag(mu)
  H=(W^0.5)**X**solve(t(X)**W**X)**t(X)**(W^0.5)
  h=diag(H)
  as.vector(t(X)**(data+0.5*h-mu))
}

library(numDeriv)

pgen<-function(y,A,B,n,Nsim,Jh0.wald=F,Jh0.score=T)
{
  data=data.frame(y,A,B)

```

```
modell1=glm(y~.,data=data,family=poisson)
modell2=glm(y~.^2,data=data,family=poisson)
b=modell1$coef
x=model.matrix(modell1)
X=model.matrix(modell2)

npar=length(modell2$coef)
nparRid=length(modell1$coef)
psi=rep(0,npar-nparRid)

p=rep(0,Nsim)
p.pen=p
p.wald=p
p.pen.wald=p
p.score=p
p.pen.score=p
beta.m=matrix(0,nrow=Nsim,ncol=npar)
beta.pen.m=beta.m

for(i in 1:Nsim)
{
  Y=rtabBin(b,x,n)
  min1=nlminb(rep(0,npar),nlogL,data=Y,X=X)
  min2=nlminb(rep(0,nparRid),function(x) nlogL(c(x,psi),data=Y,X)) #ver
  profilo
  L1=-min1$objective
  L0=-min2$objective

  min1.pen=nlminb(rep(0,npar),nlogL.pen,data=Y,X=X)
  min2.pen=nlminb(rep(0,nparRid),function(x) nlogL.pen(c(x,psi),data=Y,X
  ))
  L1.pen=-min1.pen$objective
  L0.pen=-min2.pen$objective

  p[i]=1-pchisq(2*(L1-L0),length(psi))
  p.pen[i]=1-pchisq(2*(L1.pen-L0.pen),length(psi))
}
```

```

beta=min1$par
beta.m[i,]=beta
beta.pen=min1.pen$par
beta.pen.m[i,]=beta.pen
if(any(Y==0))
{
  Jhat.inv=factor("errore")
}
else
{
  Jhat.inv=tryCatch(solve(hessian(nlogL,beta,data=Y,X=X)),error=
    function(e) factor("errore"))
}
Jhat.pen.inv=solve(hessian(nlogL.pen,beta.pen,data=Y,X=X))

if(Jh0.wald==F)
{
  if(Jhat.inv!="errore")
  {
    psi.hat=beta[seq(npar-length(psi)+1,npar,1)]
    Jhat.inv.psi=Jhat.inv[seq(npar-length(psi)+1,npar,by=1),seq(npar-
      length(psi)+1,npar,by=1)]
    p.wald[i]=wald.test(psi.hat,psi,Jhat.inv.psi)
  }
  else
  {
    p.wald[i]=NA
  }
  psi.pen.hat=beta.pen[seq(npar-length(psi)+1,npar,1)]
  Jhat.pen.inv.psi=Jhat.pen.inv[seq(npar-length(psi)+1,npar,by=1),seq(
    npar-length(psi)+1,npar,by=1)]
  p.pen.wald[i]=wald.test(psi.pen.hat,psi,Jhat.pen.inv.psi)
}

beta0=min2$par

```

```

Jhat0.inv=tryCatch(solve(hessian(nlogL,c(beta0,psi),data=Y,X=X)),error
  =function(e) factor("errore"))
beta0.pen=min2.pen$par
Jhat0.pen.inv=tryCatch(solve(hessian(nlogL.pen,c(beta0.pen,psi),data=Y
  ,X=X)),error=function(e) factor("errore"))

if(Jh0.wald==T)
{
  if(Jhat0.inv!="errore")
  {
    psi.hat=beta[seq(npar-length(psi)+1,npar,1)]
    Jhat.inv.psi=Jhat0.inv[seq(npar-length(psi)+1,npar,by=1),seq(npar-
      length(psi)+1,npar,by=1)]
    p.wald[i]=wald.test(psi.hat,psi,Jhat.inv.psi)
  }
  else
  {
    p.wald[i]=NA
  }
  psi.pen.hat=beta.pen[seq(npar-length(psi)+1,npar,1)]
  Jhat.pen.inv.psi=Jhat0.pen.inv[seq(npar-length(psi)+1,npar,by=1),seq(
    (npar-length(psi)+1,npar,by=1)]
  p.pen.wald[i]=wald.test(psi.pen.hat,psi,Jhat.pen.inv.psi)
}

if(Jh0.score==T)
{
  if(Jhat0.inv!="errore")
  {
    sc=score(c(beta0,psi),Y,X)[seq(nparRid+1,npar,1)]
    p.score[i]=1-pchisq(t(sc)%*%Jhat0.inv[seq(nparRid+1,npar,1),seq(
      nparRid+1,npar,1)]%*%sc,length(psi))
  }
  else
  {
    p.score[i]=NA
  }
}

```

```

    }
    if(Jhat0.pen.inv!="errore")
    {
      sc=score.bc(c(beta0.pen,psi),Y,X)[seq(nparRid+1,npar,1)]
      p.pen.score[i]=1-pchisq(t(sc)**Jhat0.pen.inv[seq(nparRid+1,npar,1),seq(nparRid+1,npar,1)]**sc,length(psi))
    }
    else
    {
      p.pen.score[i]=NA
    }
  }

  if(Jh0.score==F)
  {
    if(Jhat.inv!="errore")
    {
      sc=score(c(beta0,psi),Y,X)[seq(nparRid+1,npar,1)]
      p.score[i]=1-pchisq(t(sc)**Jhat.inv[seq(nparRid+1,npar,1),seq(nparRid+1,npar,1)]**sc,length(psi))
    }
    else
    {
      p.score[i]=NA
    }
    sc=score.bc(c(beta0.pen,psi),Y,X)[seq(nparRid+1,npar,1)]
    p.pen.score[i]=1-pchisq(t(sc)**Jhat.pen.inv[seq(nparRid+1,npar,1),seq(nparRid+1,npar,1)]**sc,length(psi))
  }
}

quantili=seq(0,1,1/Nsim)

par(mfrow=c(3,2), cex=0.6, mar=c(5, 5, 5, 5), mex=0.8,pty="s")
qqplot(quantili,p,main="Test W",ylab="Quantili Empirici")
abline(0,1,col=2)

```



```
qqplot(quantili,p.pen,main="Test W penalizzato",ylab="Quantili Empirici"
)
abline(0,1,col=2)
qqplot(quantili,p.wald,main="Test Wald",ylab="Quantili Empirici")
abline(0,1,col=2)
qqplot(quantili,p.pen.wald,main="Test Wald penalizzato",ylab="Quantili
Empirici")
abline(0,1,col=2)
qqplot(quantili,p.score,main="Test Score",ylab="Quantili Empirici")
abline(0,1,col=2)
qqplot(quantili,p.pen.score,main="Test Score penalizzato",ylab="Quantili
Empirici")
abline(0,1,col=2)

data.frame(p,p.pen,p.wald,p.pen.wald,p.score,p.pen.score,beta.m,beta.pen
.m)
}

#####

psycho.p=pgen(psycho.df$Freq,psycho.df$Activity,psycho.df$Psychiatric
,90,10000)

pvalues=psycho.p

beta=pvalues[,seq(7,7+(ncol(pvalues)-6)/2-1,1)]
beta.pen=pvalues[,seq(7+(ncol(pvalues)-6)/2,ncol(pvalues),1)]
p=pvalues$p
p.pen=pvalues$p.pen
p.wald=pvalues$p.wald
p.pen.wald=pvalues$p.pen.wald
p.score=pvalues$p.score
p.pen.score=pvalues$p.pen.score
numeroNAwald=length(p.wald)-sum(p.wald>=0,na.rm=T)
numeroNAwald
```

```

numeroNAscore=length(p.score) - sum(p.score>=0, na.rm=T)
numeroNAscore
numeroNAscore.pen=length(p.pen.score) - sum(p.pen.score>=0, na.rm=T)
numeroNAscore.pen

alpha=c(0.01,0.025,0.05,0.1)
round(100*data.frame(W=quantile(p,alpha),WPenalizzato=quantile(p.pen,alpha)
),Wald=quantile(p.wald,alpha,na.rm=T),WaldPenalizzato=quantile(p.pen.
wald,alpha),Score=quantile(p.score,alpha,na.rm=T),ScorePenalizzato=
quantile(p.pen.score,alpha,na.rm=T)),2)
b=glm(Freq~.,data=psycho.df,family=poisson)$coefficients
round(data.frame(VeriBeta=c(b,rep(0,ncol(beta) - length(b))),BetaHat=
colMeans(beta),BetaHatPen=colMeans(beta.pen),Bias=c(b,rep(0,ncol(beta)
-length(b))) - colMeans(beta),BiasPen=c(b,rep(0,ncol(beta) - length(b))) -
colMeans(beta.pen)),3)

par(mfrow=c(3,2), cex=0.6, mar=c(5, 5, 5, 5), mex=0.8,pty="s")
plot(alpha,quantile(p,alpha),ylim=c(0,0.25),main="Test W",ylab="Quantili
Empirici")
abline(0,1,col=2)
plot(alpha,quantile(p.pen,alpha),ylim=c(0,0.25),main="Test W penalizzato",
ylab="Quantili Empirici")
abline(0,1,col=2)
plot(alpha,quantile(p.wald,alpha,na.rm=T),ylim=c(0,0.25),main="Test Wald",
ylab="Quantili Empirici")
abline(0,1,col=2)
plot(alpha,quantile(p.pen.wald,alpha),ylim=c(0,0.25),main="Test Wald
penalizzato",ylab="Quantili Empirici")
abline(0,1,col=2)
plot(alpha,quantile(p.score,alpha,na.rm=T),ylim=c(0,0.25),main="Test Score
",ylab="Quantili Empirici")
abline(0,1,col=2)
plot(alpha,quantile(p.pen.score,alpha,na.rm=T),ylim=c(0,0.25),main="Test
Score penalizzato",ylab="Quantili Empirici")
abline(0,1,col=2)

```

```

##

sex.p=pgen(y1.df$Freq,y1.df$Var1,y1.df$Var2,91,10000)

pvalues=sex.p

beta=pvalues[,seq(7,7+(ncol(pvalues)-6)/2-1,1)]
beta.pen=pvalues[,seq(7+(ncol(pvalues)-6)/2,ncol(pvalues),1)]
p=pvalues$p
p.pen=pvalues$p.pen
p.wald=pvalues$p.wald
p.pen.wald=pvalues$p.pen.wald
p.score=pvalues$p.score
p.pen.score=pvalues$p.pen.score
numeroNAwald=length(p.wald)-sum(p.wald>=0,na.rm=T)
numeroNAwald
numeroNAScore=length(p.score)-sum(p.score>=0,na.rm=T)
numeroNAScore
numeroNAScore.pen=length(p.pen.score)-sum(p.pen.score>=0,na.rm=T)
numeroNAScore.pen

alpha=c(0.01,0.025,0.05,0.1)
round(100*data.frame(W=quantile(p,alpha),WPenalizzato=quantile(p.pen,alpha)
),Wald=quantile(p.wald,alpha,na.rm=T),WaldPenalizzato=quantile(p.pen.
wald,alpha),Score=quantile(p.score,alpha,na.rm=T),ScorePenalizzato=
quantile(p.pen.score,alpha,na.rm=T)),2)
b=glm(Freq~.,data=y1.df,family=poisson)$coefficients
round(data.frame(VeriBeta=c(b,rep(0,ncol(beta)-length(b))),BetaHat=
colMeans(beta),BetaHatPen=colMeans(beta.pen),Bias=c(b,rep(0,ncol(beta)
-length(b)))-colMeans(beta),BiasPen=c(b,rep(0,ncol(beta)-length(b)))-
colMeans(beta.pen)),3)

par(mfrow=c(3,2),cex=0.6,mar=c(5,5,5,5),mex=0.8,type="s")
plot(alpha,quantile(p,alpha),ylim=c(0,0.25),main="Test W",ylab="Quantili
Empirici")
abline(0,1,col=2)

```

```

plot(alpha,quantile(p.pen,alpha),ylim=c(0,0.25),main="Test W penalizzato",
      ylab="Quantili Empirici")
abline(0,1,col=2)
plot(alpha,quantile(p.wald,alpha,na.rm=T),ylim=c(0,0.25),main="Test Wald",
      ylab="Quantili Empirici")
abline(0,1,col=2)
plot(alpha,quantile(p.pen.wald,alpha),ylim=c(0,0.25),main="Test Wald
      penalizzato",ylab="Quantili Empirici")
abline(0,1,col=2)
plot(alpha,quantile(p.score,alpha,na.rm=T),ylim=c(0,0.25),main="Test Score
      ",ylab="Quantili Empirici")
abline(0,1,col=2)
plot(alpha,quantile(p.pen.score,alpha,na.rm=T),ylim=c(0,0.25),main="Test
      Score penalizzato",ylab="Quantili Empirici")
abline(0,1,col=2)

##

psycho.p30=pgen(psycho.df$Freq,psycho.df$Activity,psycho.df$Psychiatric
      ,30,10000)

pvalues=psycho.p30

beta=pvalues[,seq(7,7+(ncol(pvalues)-6)/2-1,1)]
beta.pen=pvalues[,seq(7+(ncol(pvalues)-6)/2,ncol(pvalues),1)]
p=pvalues$p
p.pen=pvalues$p.pen
p.wald=pvalues$p.wald
p.pen.wald=pvalues$p.pen.wald
p.score=pvalues$p.score
p.pen.score=pvalues$p.pen.score
numeroNAwald=length(p.wald)-sum(p.wald>=0,na.rm=T)
numeroNAwald
numeroNAScore=length(p.score)-sum(p.score>=0,na.rm=T)
numeroNAScore
numeroNAScore.pen=length(p.pen.score)-sum(p.pen.score>=0,na.rm=T)

```

```

numeroNAScore.pen

alpha=c(0.01,0.025,0.05,0.1)
round(100*data.frame(W=quantile(p,alpha),WPenalizzato=quantile(p.pen,alpha),
  Wald=quantile(p.wald,alpha,na.rm=T),WaldPenalizzato=quantile(p.pen.wald,alpha),
  Score=quantile(p.score,alpha,na.rm=T),ScorePenalizzato=quantile(p.pen.score,alpha,na.rm=T)),2)
b=glm(Freq~.,data=psycho.df,family=poisson)$coefficients
round(data.frame(VeriBeta=c(b,rep(0,ncol(beta)-length(b))),BetaHat=colMeans(beta),
  BetaHatPen=colMeans(beta.pen),Bias=c(b,rep(0,ncol(beta)-length(b)))-colMeans(beta),
  BiasPen=c(b,rep(0,ncol(beta)-length(b)))-colMeans(beta.pen)),3)

par(mfrow=c(3,2),cex=0.6,mar=c(5,5,5,5),mex=0.8,type="s")
plot(alpha,quantile(p,alpha),ylim=c(0,0.25),main="Test W",ylab="Quantili Empirici")
abline(0,1,col=2)
plot(alpha,quantile(p.pen,alpha),ylim=c(0,0.25),main="Test W penalizzato",ylab="Quantili Empirici")
abline(0,1,col=2)
plot(alpha,quantile(p.wald,alpha,na.rm=T),ylim=c(0,0.25),main="Test Wald",ylab="Quantili Empirici")
abline(0,1,col=2)
plot(alpha,quantile(p.pen.wald,alpha),ylim=c(0,0.25),main="Test Wald penalizzato",ylab="Quantili Empirici")
abline(0,1,col=2)
plot(alpha,quantile(p.score,alpha,na.rm=T),ylim=c(0,0.25),main="Test Score",ylab="Quantili Empirici")
abline(0,1,col=2)
plot(alpha,quantile(p.pen.score,alpha,na.rm=T),ylim=c(0,0.25),main="Test Score penalizzato",ylab="Quantili Empirici")
abline(0,1,col=2)

##

sex.p30=pgen(y1.df$Freq,y1.df$Var1,y1.df$Var2,30,10000)

```

```

pvalues=sex.p30

beta=pvalues[,seq(7,7+(ncol(pvalues)-6)/2-1,1)]
beta.pen=pvalues[,seq(7+(ncol(pvalues)-6)/2,ncol(pvalues),1)]
p=pvalues$p
p.pen=pvalues$p.pen
p.wald=pvalues$p.wald
p.pen.wald=pvalues$p.pen.wald
p.score=pvalues$p.score
p.pen.score=pvalues$p.pen.score
numeroNAwald=length(p.wald)-sum(p.wald>=0,na.rm=T)
numeroNAwald
numeroNAScore=length(p.score)-sum(p.score>=0,na.rm=T)
numeroNAScore
numeroNAScore.pen=length(p.pen.score)-sum(p.pen.score>=0,na.rm=T)
numeroNAScore.pen

alpha=c(0.01,0.025,0.05,0.1)
b=glm(Freq~.,data=y1.df,family=poisson)$coefficients
round(100*data.frame(W=quantile(p,alpha),WPenalizzato=quantile(p.pen,alpha),
Wald=quantile(p.wald,alpha,na.rm=T),WaldPenalizzato=quantile(p.pen.wald,alpha),
Score=quantile(p.score,alpha,na.rm=T),ScorePenalizzato=quantile(p.pen.score,alpha,na.rm=T)),2)
round(data.frame(VeriBeta=c(b,rep(0,ncol(beta)-length(b))),BetaHat=colMeans(beta),
BetaHatPen=colMeans(beta.pen),Bias=c(b,rep(0,ncol(beta)-length(b)))-colMeans(beta),
BiasPen=c(b,rep(0,ncol(beta)-length(b)))-colMeans(beta.pen)),3)

par(mfrow=c(3,2),cex=0.6,mar=c(5,5,5,5),mex=0.8,type="s")
plot(alpha,quantile(p,alpha),ylim=c(0,0.25),main="Test W",ylab="Quantili Empirici")
abline(0,1,col=2)
plot(alpha,quantile(p.pen,alpha),ylim=c(0,0.25),main="Test W penalizzato",ylab="Quantili Empirici")
abline(0,1,col=2)

```

---

```
plot(alpha,quantile(p.wald,alpha,na.rm=T),ylim=c(0,0.25),main="Test Wald",
      ylab="Quantili Empirici")
abline(0,1,col=2)
plot(alpha,quantile(p.pen.wald,alpha),ylim=c(0,0.25),main="Test Wald
      penalizzato",ylab="Quantili Empirici")
abline(0,1,col=2)
plot(alpha,quantile(p.score,alpha,na.rm=T),ylim=c(0,0.25),main="Test Score
      ",ylab="Quantili Empirici")
abline(0,1,col=2)
plot(alpha,quantile(p.pen.score,alpha,na.rm=T),ylim=c(0,0.25),main="Test
      Score penalizzato",ylab="Quantili Empirici")
abline(0,1,col=2)
```

---