



UNIVERSITA' DEGLI STUDI DI PADOVA

FACOLTA DI SCIENZE STATISTICHE

CORSO DI LAUREA TRIENNALE
IN STATISTICA E TECNOLOGIE INFORMATICHE

TEST DI PERMUTAZIONE PER GLI EFFETTI FISSI DI UN MODELLO MISTO

RELATORE: ch.mo prof. Finos Livio

CORRELATORE: dott. Basso Dario

LAUREANDO: Novello Andrea, 600722

ANNO ACCADEMICO 2010-2011

Indice analitico

1 INTRODUZIONE AI MODELLI MISTI.....	5
1.1 Il modello misto.....	7
1.2 Stima dei parametri.....	9
<i>1.2.1 Algoritmi di massimizzazione....</i>	<i>11</i>
Newton-Rapson.....	13
Fisher scoring.....	14
Expectation-Maximization.	15
<i>1.2.2 Punti di partenza.....</i>	<i>16</i>
1.3 Proprietà statistiche del modello.....	17
2 TEST DI PERMUTAZIONE SUI MODELLI MISTI..	21
<i>2.1 Introduzione.....</i>	<i>21</i>
<i>2.2 Il modello.....</i>	<i>22</i>
<i>2.3 Il test.....</i>	<i>27</i>
3 I CASI IN ESAME.....	35
<i>3.1 Analisi Snijders Data.....</i>	<i>35</i>
<i>3.2 Analisi Bottles.....</i>	<i>37</i>
4 APPENDICE.....	43
<i>4.1 Funzione t2p.....</i>	<i>43</i>
<i>4.2 Funzione t.nptest.....</i>	<i>43</i>
<i>4.3 Funzioni random.coeff/2</i>	<i>45</i>
<i>4.4 Funzione make.Y.....</i>	<i>49</i>
5 BIBLIOGRAFIA.....	53

1 INTRODUZIONE AI MODELLI MISTI

I modelli lineari analizzati solitamente vengono detti Modelli lineari ad effetti fissi e sono costituiti nel seguente modo:

$$y = X\beta + \varepsilon$$

Dove y è il vettore risposta $n \times 1$ che costituisce la variabile risposta mentre X è una matrice $n \times m$ costituente le variabili esplicative (variabili indipendenti), β ($m \times 1$) è un vettore di parametri da stimare e ε è un vettore ($n \times 1$) detto dei “residui”: “errori” effettuati durante la stima cioè gli scarti della variabile y dal vettore $X\beta$.

In alcuni casi è possibile considerare la matrice X come una matrice stocastica di dati provenienti per esempio da una distribuzione normale (approccio stocastico) ma generalmente si considera la matrice X come deterministica e perciò non si assume nessuna distribuzione per essa (approccio deterministico), mentre il vettore ε al contrario è assunto come proveniente da una distribuzione che in questo caso assumeremo normale: $\varepsilon \sim N(0, \sigma^2)$ con elementi i.i.d..

Una delle caratteristiche dei modelli ad effetti fissi è quella che ogni osservazione è di per se indipendente dalle altre; tale assunzione però non può ovviamente essere sempre soddisfatta:

prendiamo per esempio un esperimento compiuto su dei pazienti ai quali è stata somministrato un medicinale e vogliamo osservare dei sintomi che crediamo siano influenzati dal medicinale. Prendiamo quindi più misure su ogni soggetto in momenti diversi corrispondenti ai giorni successivi l’assunzione del farmaco e proviamo a costruire un modello.

Ciò che si nota fin da subito è che ovviamente le misure riferite allo

stesso soggetto non possono essere tra loro indipendenti per ovvi motivi, possiamo però assumere che le realizzazioni tra (between) soggetto e soggetto siano tra loro indipendenti.

In tal caso quindi abbiamo che per ogni soggetto al suo interno (whithin) le osservazioni sono correlate ma tra un soggetto e l'altro la correlazione delle sue componenti è 0.

Supponiamo per esempio che tra un soggetto e l'altro i valori abbiano media completamente diversa ma una oscillazione molto simile all'interno dello stesso gruppo, in tal caso possiamo pensare che la variabile risposta y sia determinata dagli effetti fissi, ma anche da un vettore di coefficienti casuali b_i che descrivono le peculiarità individuali. Tale vettore cambia da soggetto a soggetto e sposta l'intercetta della retta formata dagli elementi di $X_i\beta$ ossia quelli corrispondenti al soggetto i .

In tal caso arriviamo ad un modello come questo:

$$y_{ij} = \alpha + b_i + \beta_1 \cdot x_{ij} + \varepsilon_{ij}$$

con alfa che corrisponde alla media delle osservazioni totali.

Casi come questi ovviamente possono essere svariati e non riguardano unicamente pazienti e medicinali.

I soggetti o gruppi (clusters) possono essere infatti di diverse tipologie, l'esempio considerato in precedenza riguarda misure ripetute sullo stesso soggetto ma più in generale gli individui possono semplicemente essere intesi come gruppi diversi che creano fonti di varianza multiple dovute alla diversificazione tra un gruppo e un altro oppure possono riguardare addirittura le serie temporali dove ogni elemento è legato ai precedenti dato che si tratta di misurazioni sullo stesso elemento in unità temporali distinte.

1.1 IL MODELLO MISTO

Ora procederemo a generalizzare i modelli sopra descritti in modo tale da includere molti dei casi possibili:

Siano Z_i matrici fissate $n_i \times k$ dove n_i corrisponde al numero di elementi all'interno del gruppo i .

Il modello lineare misto (Linear Mixed Model, LMM) generale è definito come segue:

$$y_i = X_i \beta + Z_i b_i + \varepsilon_i$$

dove:

y_i è un vettore $n_i \times 1$, X_i è una matrice $n_i \times m$, β è un vettore $m \times 1$ e ε_i è un vettore $n_i \times 1$ e b_i è un vettore $k \times 1$ dei random effects.

Come sempre assumiamo che $\varepsilon \sim N(0, \sigma^2)$ ma stavolta assumiamo anche che $b_i \sim N(0, D_*)$

con b_i indipendente da ε_i

Assumiamo d'ora in avanti: $D = \frac{D_*}{\sigma^2}$.

Diventa quindi evidente che:

$$E(y_i) = X_i \beta$$

$$VAR(y_i) = \sigma^2 (I + Z_i D Z_i')$$

Dove I è una matrice quadrata identità di dimensione n_i .

Osserviamo inoltre che dato y_{ij} con i che identifica il gruppo e j l'elemento al suo interno

$$COV(y_{ik}, y_{il}) = D_* \text{ con } 0 \leq k < l \leq n$$

Mentre:

$$COV(y_{kj}, y_{lj}) = 0 \text{ con } 0 \leq k < l \leq n$$

Ora che abbiamo definito i vari momenti e le distribuzioni dei vettori casuali possiamo arrivare alla seguente conclusione:

$$y_i \sim N(X_i \beta, \sigma^2(I + Z_i D Z_i'))$$

A questo punto è banale giungere alla funzione di verosimiglianza semplificata:

$$l(\theta) = -\frac{1}{2} \sum [\log(|V_i|) + (y_i - X_i \beta)' V_i^{-1} (y_i - X_i \beta)]$$

dove $V_i = I + Z_i D Z_i'$ e $\theta = (\beta, \sigma^2, \text{vech}'(D))'$: il vettore dei parametri da stimare (vech' è un operatore che trasforma la matrice in vettore formato da elementi unici dato che D è simmetrica).

Esempio:

Supponiamo per esempio di voler studiare l'altezza raggiunta dalle piante di una serra due mesi dopo aver applicato su di esse 2 diversi trattamenti (per esempio due diversi diserbanti), supponiamo inoltre che le piante siano di diverso tipo e perciò le dividiamo in 5 gruppi corrispondenti alle 5 specie di piante in esame, supponiamo inoltre avere 8 piante per ciascun gruppo e che i diversi gruppi differenzino le piante sia per la loro altezza base che per l'effetto che i trattamenti hanno su di loro (almeno per quanto riguarda l'incremento dell'altezza). Per ogni gruppo 4 piante vengono sottoposte al trattamento "A" e le 4 rimanenti al trattamento "B". Ci troviamo di fronte quindi ad un caso di modello a coefficienti casuali con X_i matrice 8×2 con la prima colonna di "1" e la seconda avente i primi 4 valori corrispondenti a "0" e i rimanenti a "1".

Notiamo che essendo un modello a coefficienti casuali Z_i corrisponde a X_i ($Z_i = X_i$) e inoltre X_i è la stessa per ogni cluster perciò $X_i = X \quad \forall i$.

Il modello derivante perciò sarà il seguente:

$$y_i = X(\beta + b_i) + \varepsilon_i$$

In questo caso quindi y_i si distribuisce:

$$y_i \sim N(X\beta, \sigma^2(I + XDX))'$$

Molto meno banale è invece la stima dei parametri.

1.2 Stima dei parametri

In questo paragrafo e nel successivo ci occuperemo della stima dei parametri del modello precedentemente descritto e lo faremo massimizzando la funzione di log-verosimiglianza (MLE) per via numerica utilizzando tre metodi differenti che però presentano caratteristiche in comune.

Innanzitutto è importante osservare che il cuore del nostro problema sta nella stima di D.

Se D fosse noto in precedenza infatti, la stima dei coefficienti β è ottenibile banalmente tramite GLS (Generalized least squares).

Partendo dal modello generale per i LMM considerato in precedenza è importante innanzitutto che la somma $\sum (X_i' X_i)$ sia non-singolare e che: $\sum (n_i) = N_i > m$ per rendere il modello identificabile per β mentre almeno una delle matrici $Z_i' Z_i$ deve essere definita positiva e deve essere $\sum (n_i - k) > 0$ perché sia invece identificabile per σ^2 e D.

Se $n_i = \text{cost}$ e $Z_i = Z \quad \forall i$ allora il modello è detto bilanciato e se oltre ad essi anche $X_i = X \quad \forall i$ allora il modello è detto bilanciato con coefficienti casuali.

Di vitale importanza però è lo spazio parametrico in quanto come già sappiamo $\beta \in \mathbb{R}^m$ e $\sigma^2 > 0$ ma in aggiunta D deve essere definita non-negativa (chiameremo D^+ lo spazio di non negatività di D).

Prima di procedere con i metodi di stima è importante stabilire un criterio generale per l'esistenza del MLE. Gli algoritmi di stima infatti potrebbero riuscire o fallire durante le iterazioni, ma la non riuscita non implica necessariamente che l'MLE non esiste, potrebbero fallire per esempio perché durante le iterazioni la matrice D non diventa definita non-negativa oppure perché la sequenza dei punti generati dall'algoritmo non convergono alla stima a causa di una cattiva scelta del punto iniziale; oppure molto più banalmente il pc in cui l'algoritmo è stato lanciato va in overflow obbligandolo a terminare prima della riuscita.

A tale scopo definiamo il seguente teorema per l'esistenza del MLE:

Sotto le usuali assunzioni definite in precedenza per il modello, l'MLE (Maximum likelihood estimates) e l'RMLE (Restricted Maximum likelihood estimates) per il modello ad effetti misti nello spazio parametrico definito in precedenza esiste se e solo se non esiste un

$\beta \in \mathbb{R}^m$ tale che per $r_1, \dots, r_N \in \mathbb{R}^k$, :

$$y_i = X_i \beta + Z_i r_i \quad i=1,2,\dots,N$$

Il quale implica praticamente che $\sum (n_i - k) - m > 0$.

In particolare per quanto riguarda la non-negatività della matrice D è

importante far notare che una delle possibilità da considerare è che il motivo per cui questa caratteristica non sia soddisfatta durante le iterazioni è che in realtà la matrice di covarianza dei random effects del nostro modello sia zero. Ciò ovviamente comporta non pochi problemi in fase di stima perciò è utile avere un criterio per capire in anticipo se una soluzione positiva per la matrice D esiste.

Consideriamo quindi il seguente teorema in cui OLS (ordinary least squares) denota la stima ai minimi quadrati ordinari del parametro da cui è preceduto:

Lasciamo $\hat{e}_i = y_i - X_i \hat{\beta}_{(OLS)}$ denotare il vettore $n_i \times 1$ dei minimi quadrati ordinari e $\hat{\sigma}_{(OLS)}^2 = \frac{\sum (\|\hat{e}_i\|^2)}{N_i}$ denotare la varianza ai minimi quadrati ordinari. Se la matrice $k \times k$

$$\sum (Z_i' \hat{e} \hat{e}' Z_i) - \sigma_{(OLS)}^2 \sum (Z_i' Z_i)$$

è definita non-negativa e non-zero allora la MLE di D è una matrice non-zero.

1.2.1 Algoritmi di massimizzazione

Passiamo ora agli algoritmi di massimizzazione della funzione di log-verosiglianza.

I tre tipi di algoritmi usati in statistica per massimizzare la funzione di log-verosiglianza sono:

Expectation-Maximization (EM), Fisher scoring (FS) e Newton-Raphson (NR).

Questi tre algoritmi hanno la forma seguente:

$$t_{(s+1)} = t_s + \lambda_s g_s$$

dove t_s è la stima di massima verosimiglianza alla s-esima iterazione, mentre $t_{(s+1)}$ è il vettore aggiornato all'iterazione successiva.

$0 < \lambda_s \leq 1$ è la lunghezza del passo e g_s è il vettore direzionale (o aggiustamento), calcolato come:

$$g_s = H_s^{-1} * g_s$$

Dove

$$g_s = \frac{\delta l(t)}{\delta t} \Big|_{t=t_s}$$

è il gradiente della funzione di log-verosimiglianza l e H_s è una matrice definita positiva.

La lunghezza del passo λ_s è necessaria per assicurare l'incremento della funzione l da iterazione a iterazione.

Le derivate delle funzione di verosimiglianza che compongono il gradiente sono le seguenti:

$$\frac{\partial l}{\partial \beta} = \sigma^{-2} \sum X_i' V_i^{-1} (y_i - X_i \beta)$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{1}{2} N_i \sigma^{-2} + \frac{1}{2} \sigma^{-4} \sum (y_i - X_i \beta)' V_i^{-1} (y_i - X_i \beta)$$

$$\frac{\partial l}{\partial D} = -\frac{1}{2} \sum [Z_i V_i^{-1} Z_i - \sigma^{-2} Z_i' V_i^{-1} (y_i - X_i \beta) (y_i - X_i \beta)' V_i^{-1} Z_i]$$

Che vanno ovviamente poste uguali a 0 per il calcolo delle stime di massima verosimiglianza.

Newton-Rapson:

L'algoritmo di Newton-Rapson utilizza come matrice H la matrice hessiana della funzione di log-verosimiglianza ed è definito come segue:

$$H = \begin{bmatrix} H_{11}^{m \times m} & H_{12}^{m \times 1} & H_{13}^{m \times k^2} \\ H_{12}' & H_{22}^{1 \times 1} & H_{23}^{1 \times k^2} \\ H_{13}' & H_{23}' & H_{33}^{k^2 \times k^2} \end{bmatrix}$$

con:

$$H_{11} = \sigma^{-2} \sum X_i' V_i^{-1} X_i ,$$

$$H_{12} = \sigma^{-4} \sum X_i' V_i^{-1} e_i ,$$

$$H_{13} = \sigma^{-2} \sum e_i' V_i^{-1} Z_i \times X_i' V_i^{-1} Z_i ,$$

$$H_{22} = \sigma^{-6} \sum e_i' V_i^{-1} e_i - 0.5 N_i \sigma^{-4} ,$$

$$H_{23} = 0.5 \sigma^{-4} \sum e_i' V_i^{-1} Z_i \times e_i' V_i^{-1} Z_i ,$$

$$H_{33} = 0.5 \sum \{ \sigma^{-2} (Z_i' V_i^{-1} e_i e_i' V_i^{-1} Z_i \times R_i + R_i \times Z_i' V_i^{-1} e_i e_i' V_i^{-1} Z_i) - R_i \times R_i \} .$$

Dove:

$$R_i = Z_i' V_i^{-1} Z_i$$

λ_s invece è scelto con la seguente regola: si comincia ponendolo uguale a 1; se $l_{(s+1)} > l_s$ lo si mantiene uguale a 1 mentre se $l_{(s+1)} \leq l_s$

allora prendiamo $\lambda_s = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots$ ecc finché $l_{(s+1)} > l_s$.

Il problema sostanziale di questo algoritmo tuttavia è che è possibile che la matrice hessiana H_s non diventi definita positiva durante le iterazioni, soprattutto nel caso in cui ci troviamo distanti dal punto di massima verosimiglianza.

Fisher scoring:

Nell'algoritmo di Fisher scoring la matrice H è l'opposto della matrice attesa derivata dall'hessiano che è chiamata anche matrice di informazione di Fisher.

$$I = \begin{bmatrix} \sigma^2 \sum X_i' V_i^{-1} X_i & 0 & 0 \\ 0 & 0.5 N_t \sigma^{-4} & 0.5 \sigma^{-2} \sum \text{vec}'(R_i) \\ 0 & 0.5 \sigma^{-2} \sum \text{vec}(R_i) & 0.5 \sum R_i \times R_i \end{bmatrix}$$

dove vec denota l'operatore di vettorizzazione delle matrici.

Si nota fin da subito che la matrice in questione è diagonale a blocchi, il che significa in questo caso che le stime di β e quelle di (σ^2, D) possono essere eseguite separatamente e perciò per semplificare il lavoro calcoleremo $\beta_{(s+1)}$ utilizzando la stima tramite GLS con la matrice D rimpiazzata con \hat{D} mentre σ^2 e D utilizzando il blocco in basso a destra della matrice I.

Ci sono tre principali motivi per preferire questo algoritmo rispetto a quello di Newton-Rapson:

1. la matrice I è sempre definita positiva mentre l'hessiano empirico no.
2. la matrice di informazione attesa all'ultima iterazione fornisce una migliore stima della varianza asintotica rispetto alla matrice empirica ed è più robusta in presenza di outliers.
3. l'uso della matrice di informazione semplifica la costruzione di differenti versioni di algoritmi di massimizzazione basati sulla log-verosimiglianza profilo e per gli RMLE (restricted maximum likelihood estimation).

L'algoritmo di fisher-score in sostanza è preferibile a quello di Newton-Rapson tranne forse per il fatto che è leggermente più lento rispetto ad esso a convergere.

Expectation-Maximization:

Nell'algoritmo di Expectation-Maximization aggiorniamo i parametri σ^2 e D con l'aggiornamento:

$$\Delta \sigma = 2 \frac{\sigma^2}{N_t} \frac{\partial l}{\partial \sigma^2} \quad \Delta D = \frac{2}{n} D \frac{\partial l}{\partial D} D$$

mentre β è ricalcolato con il solito metodo dei GLS.

Notiamo che questo algoritmo può essere visto come una forma speciale dell'algoritmo generico visto in precedenza con $\lambda_s = 1$.

Questo algoritmo presenta due vantaggi:

1. massimizza la funzione di log-verosimiglianza iterazione dopo iterazione se il gradiente non è zero.
2. la formula di aggiornamento della matrice $D_{(s+1)}$ genera sempre matrici definite positive purché la matrice D_0 di partenza sia definita positiva.

Tuttavia il maggior punto a sfavore di questo algoritmo è il fatto che mostra una maggiore lentezza di convergenza quando la matrice D è vicina a 0.

Dopo una attenta analisi dei tre algoritmi quindi giungiamo alla conclusione che il migliore sia effettivamente l'algoritmo di Fisher-scoring dato che dispone di molti aspetti positivi rispetto agli altri e perciò proseguiamo definendo un metodo di scelta per il punto di partenza di tale algoritmo.

1.2.2 Punti di partenza

E' importante chiaramente scegliere un punto di partenza intelligente per i nostri algoritmi in modo tale da evitare che essi convergano troppo lentamente o peggio non convergano affatto.

Consideriamo quindi la seguente equazione che corrisponde alla derivata della log-verosimiglianza eguagliata a 0 una volta posto per ipotesi $D=0$:

$$\sum (Z_i' Z_i) D (Z_i' Z_i) = \sum [\hat{\sigma}_{OLS}^2 Z_i' \hat{e}_i' \hat{e}_i Z_i - Z_i' Z_i]$$

Il valore di D che la soddisfa sarà quindi utilizzato come stima del

punto D_0 quindi \hat{D}_0 e applicando la vettorizzazione otteniamo:

$$\text{vec}(\hat{D}_0) = [\sum (Z_i' Z_i) \times (Z_i' Z_i)]^{-1} \times \text{vec}[\sum [\hat{\sigma}_{OLS}^{-2} Z_i' \hat{e}_i \hat{e}_i' Z_i - Z_i' Z_i]]$$

Per concludere inoltre osserviamo che per far sì che la matrice D resti definita non-negativa durante le iterazioni è possibile proiettarla nel sottospazio $D+$ dove tramite scomposizione spettrale moltiplicativa si controllano gli autovalori e si sostituiscono quelli minori di 0 con 0; la matrice rimoltiplicata dopo la modifica di quella diagonale sarà perciò proiettata su $D+$.

In alternativa è possibile proiettare direttamente la matrice \hat{D} finale su $D+$ consentendo quindi a D di non diventare non-definita negativa durante le iterazioni purché la matrice V_i sia definita positiva per $\forall i$ durante il processo di massimizzazione.

1.3 Proprietà statistiche del modello

Passiamo ora ad analizzare quelle che sono le proprietà statistiche dei modelli lineari misti.

Cominciamo con il calcolo della matrice di informazione di Fisher per i parametri del nostro modello.

Osserviamo fin da subito che i coefficienti β sono incorrelati con i parametri di varianza (σ^2, D) il che ci consente di poter stimare la matrice di informazione separatamente per i due casi.

Nel primo caso la matrice di covarianza si calcola nel seguente modo:

$$COV(\hat{\beta}_{ML}) = \hat{\sigma}^2 (\sum X_i' (I + Z_i \hat{D} Z_i')^{-1} X_i)^{-1}$$

Mentre per ciò che concerne il vettore dei parametri di varianza

$\psi = (\sigma^2, \text{vech}'(D))'$, la matrice di informazione è la seguente per il cluster i :

$$I = \frac{1}{2} \begin{bmatrix} n_i \sigma^{-4} & \sigma^{-2} \text{vec}'(R_i) D^{+'} \\ \sigma^{-2} D^{+} \text{vec}(R_i)' & D^{+} (R_i \times R_i) D^{+'} \end{bmatrix}$$

La matrice di varianza asintotica per l'intero campione invece è l'inversa della matrice di informazione totale:

$$COV(\hat{\psi}_{ML}) = 2 \begin{bmatrix} N_t \sigma^{-4} & \sigma^{-2} \text{vec}'(\sum R_i) D^{+'} \\ \sigma^{-2} D^{+} \text{vec}(\sum R_i)' & D^{+} (\sum R_i \times R_i) D^{+'} \end{bmatrix}^{-1}$$

Per quanto riguarda invece gli intervalli di confidenza sui coefficienti β osserviamo che potremmo pensare di applicare il classico test di Wald sulla componente β_j tenendo conto che come già sappiamo sotto H_0 definisce

$$\frac{\hat{\beta}_j}{s_j} \sim N(0,1) \quad \text{con } s_j \text{ scarto quadratico medio della componente } j\text{-esima.}$$

Ma se il numero di parametri stimati è relativamente grande allora il rapporto:

$$\frac{\hat{\beta}_j}{s_j} \sim t(N_t - m)$$

È però importante osservare che in realtà questo rapporto non si distribuisce realmente con una t di student come quella in quanto la matrice D del nostro modello è stimata anziché essere conosciuta a priori.

Molti autori infatti ritengono che l'intervallo di confidenza verosimiglianza-profilo (PL) sia migliore rispetto a quello di Wald usuale.

L'idea è la seguente: lasciamo j fissato e vogliamo costruire un intervallo di confidenza per la j -esima componente del vettore

$$y = (y_1, y_2, \dots, y_k)' \quad \text{basato sulla log-verosimiglianza } l(y) .$$

Sia \hat{y} l' MLE per $l(y)$ con $l_{max} = l(\hat{y})$. Fissiamo y_j e denotiamo il vettore $(k-1) \times 1$: $\tilde{y} = (y_1, \dots, y_{j-1}, y_{j+1}, \dots, y_k)'$ con la j -esima componente rimossa e troviamo il massimo di l rispetto a \tilde{y} , che chiameremo $\tilde{l}(y_j)$.

A questo punto le radici $x_1 < x_2$ (se esistono) della formula:

$$l(x) = l_{max} - \frac{1}{2} q^2_{(1 - \frac{\alpha}{2}, t_{(N-m)})} \quad (\text{con } q: \text{ quantile della } t \text{ di student})$$

costituiscono gli estremi dell'intervallo di confidenza verosimiglianza-profilo per y_j

Come sappiamo bene il test di ipotesi e il calcolo dell'intervallo di confidenza per un parametro sono equivalenti quindi per verificare il test di ipotesi $H_0: \beta_j = 0$ basta verificare l'appartenenza dello 0 all'intervallo di confidenza corrispondente.

Continuiamo col definire un buon test statistico per la presenza dei random effects.

Se i random effects non sono presenti nel nostro modello significa $D = 0$ quindi:

$$H_0: D = 0$$

Come è già stato precedentemente introdotto nei capitoli precedenti ma mentre in precedenza abbiamo presentato un metodo "computazionale" per la verifica di tale ipotesi adesso ne introdurremo uno di statistico.

Osserviamo fin da subito che non è semplice fare un test come questo in quanto ricordiamo che in 0 parlando del parametro D ci troviamo in corrispondenza del bordo dello spazio parametrico. Questo comporta

non pochi problemi in quanto il test rapporto di verosimiglianza non ha una distribuzione limitata χ^2 .

Il seguente test invece si distribuisce esattamente come una F di Snedecor:

$$\frac{(S_{OLS} - S_{min}) / (r - m)}{S_{min} / (N_t - r)} \sim F(r - m, N_t - r)$$

con S_{OLS} stima di σ^2 ai minimi quadrati ordinari, assumendo cioè che $D = 0$ mentre

$$S_{min} = \min_y \|y - W y\|^2$$

Dove W è la matrice congiunta X e Z ottenuta dall'unione delle colonne delle due rispettive matrici mentre $y = (\beta', b_1', \dots, b_n)'$ e $r = rk(W)$.

È inoltre possibile avere una stima dei random effect supponendo che σ^2 e D siano noti:

$$\tilde{b}_i = D Z_i' (I + Z_i D Z_i')^{-1} (y_i - X_i \hat{\beta})$$

dove $\hat{\beta}$ è ottenuto con la stima ai minimi quadrati generalizzati (GLS).

Per quanto riguarda i test di ipotesi generalizzati:

$$H_0: C \beta = 0 \quad \text{con } C \text{ è una matrice fissata } q \times m$$

sui parametri B dei modelli misti essi presentano la seguente forma:

$$\frac{(RSS_0 - RSS) / q}{RSS / (n - m)} \sim F(q, n - m)$$

Dove $RSS = (y - X \hat{\beta})' V^{-1} (y - X \hat{\beta})$ e $RSS_0 = (y - X \hat{\beta}_0)' V^{-1} (y - X \hat{\beta}_0)$

con $\hat{\beta}$: la stima GLS di β e $\hat{\beta}_0$ è la stima GLS sotto la restrizione $C \beta = 0$.

2 TEST DI PERMUTAZIONE SUI MODELLI MISTI

2.1 Introduzione

Dopo aver descritto generalmente cosa sono i modelli misti procediamo con l'analizzare un metodo per i test di ipotesi sugli effetti fissi in modello misto che poi utilizzeremo per l'analisi di due dataset.

Come abbiamo già detto in precedenza i test di ipotesi per i coefficienti degli effetti fissi per un modello misto sono un argomento delicato in quanto gli usuali test t o W non sono effettuabili come in modello a soli effetti fissi, in particolare il test t non si distribuisce realmente come una t di student e quindi il test non è esatto. Nell'introduzione abbiamo mostrato un test basato sulla verosimiglianza profilo, esso tuttavia non è adatto a tutti i casi riscontrabili nella realtà, soprattutto quando siamo in ambito multivariato e quando sono necessarie alternative unidirezionali anziché bilaterali.

Negli usuali modelli misti inoltre, come abbiamo già detto, la varianza totale è data dalla somma della varianza dei random effects con quella del termine d'errore che si suppone essere in comune tra i soggetti ma supponiamo per esempio di essere interessati alla stima della media relativa ad ogni soggetto e ci concentriamo sulla varianza di tale stimatore: è ovvio che essendo una media la varianza dello stimatore dipenderà dalla numerosità campionaria di ciascun soggetto quindi ci troveremo in un contesto di eteroschedasticità.

Quindi ci troviamo di fronte ad un modello multivariato dove la variabile risposta "y" è composta da più colonne: $Y(n_i \times q)$ dove q rappresenta il numero di variabili del multivariato, potremmo pensare per esempio di considerare le m variabili separatamente per poi conglobare le q inferenze in un p-value globale che le rappresenti. Un alternativa valida potrebbe essere quella di utilizzare il MCMC (Mixed-Effects modeling with crossed

random effects for subjects and item) tuttavia in questo caso la distribuzione a priori (multivariata) per tutti i termini di errore è assunta nota e i costi computazionali sono molto elevati.

Scomporre un'analisi multivariata in più analisi univariate inoltre è un'operazione pericolosa in quanto le m variabili non possono essere assunte come indipendenti e la strategia adottata comunemente è la correzione di Bonferroni che è nota essere adatta a i casi in cui la correlazione tra i p -value è elevata. Tuttavia una buona estensione al multivariato non sembra esserci né per l'approccio classico tramite MLE che per quello MCMC.

Quello che utilizzeremo in seguito invece è un metodo totalmente non parametrico che provvede ad un test multivariato esatto che è adattabile a modelli generali e che consente inoltre di condurre test con alternative unidirezionali.

Tale test è detto test di permutazione, (Permutation Test) e cominceremo a trattarlo discutendo riguardo al modello su cui si basa.

2.1 Il modello

Assumiamo che y_i sia un vettore $n_i \times 1$ di misurazioni tale che:

$$y_i | (\mu_i, R_i) \sim \phi(\mu_i, R_i) \quad i=1, \dots, N \quad (1)$$

Dove n_i è il numero di osservazioni (prove) nel soggetto i , $\phi(\circ)$ è una funzione di densità/probabilità,

$$\mu_i = h(X_i, \beta_i) \quad (2)$$

e R_i è la matrice di varianza/covarianza per le osservazioni di ciascun

soggetto, con β_i un vettore $p \times 1$ di coefficienti e X_i una matrice $n_i \times p$ di preditori (fissati).

Negli usuali modelli lineari misti univariati se le normali assunzioni sono soddisfatte (da ora in poi $\phi(\circ)$ sarà la densità normale multivariata),

$y_i = \mu_i + \varepsilon_i$ dove $\varepsilon \sim N(0, R_i)$, con $R_i = \sigma^2 I_{n_i}$ e $\mu_i = X_i(\beta + u_i)$, dove β è un vettore di effetti fissi, $u_i \sim N(0, \Sigma_u)$ e $COV(\varepsilon_i, u_i) = 0$. Qui u_i è un vettore $p \times 1$ di random effects con la matrice di varianza/covarianza: Σ_u che è comune per tutti i soggetti ed è assunta indipendente da ε_i .

Possiamo scrivere l'intero modello come:

$$y_i = X_i(\beta + u_i) + \varepsilon_i$$

quindi $E[y_i] = X_i\beta$ e $V[y_i] = X_i V[\beta + u_i] X_i' + V[\varepsilon_i] = X_i \Sigma_u X_i' + R_i$.

Questa specifica quindi estende i modelli lineari misti classici considerati nell'introduzione consentendo ad R_i di essere diverso per ogni soggetto, quando comunemente lo avremmo assunto costante ($R_i = R \forall i$).

Ovviamente la definizione (1) e (2) comprende modelli molto più generali e complessi.

Inoltre assumiamo che sia disponibile t_i : uno stimatore di β condizionatamente a ciascun soggetto, quindi:

$$t_i | u_i \sim (\beta + u_i, \Sigma_i)$$

Dove Σ_i è la matrice di varianza/covarianza $p \times p$ di t_i interna ai gruppi. Notare che Σ_i dipende solo da R_i e X_i ; per i modelli lineari definiti sopra diventa $\Sigma_i = \sigma_i^2 (X_i' X_i)^{-1}$. Questa parte del modello descrive la variabilità interna ai gruppi.

Per finire, assumiamo che:

$$\beta_i = \beta + u_i \sim (\beta, \Sigma_u)$$

che è il vero vettore dei parametri pescato da una distribuzione comune a tutti i soggetti. Questa parte del modello descrive la variabilità tra i soggetti a causa della presenza dei random effects. Mettendo tutto assieme otteniamo, incondizionatamente:

$$t_i \sim (\beta, \Sigma_u + \Sigma_i)$$

quindi: $E[t_i] = E[E(t|u_i)] = E[\beta + u_i] = \beta$ e

$V[t_i] = V[E(t_i|u_i)] + E[V(t_i|u_i)] = \Sigma_u + \Sigma_i$. Notare che qui gli stimatori t_i sono ottenuti condizionatamente a ciascun soggetto.

Se le assunzioni per un modello lineare tengono per ciascun soggetto e

$h(X_i, \beta_i) = X_i \beta_i$, condizionatamente alla realizzazione del random effect u_i gli stimatori di β_i possono essere ottenuti attraverso gli OLS, che sarebbero: $t_i = (X_i' X_i)^{-1} X_i' y_i$.

Questo ci porta, incondizionatamente, alla seguente formula:

$$V[t_i] = (X_i' X_i)^{-1} X_i' [X_i \Sigma_u X_i' + \Sigma_i] X_i (X_i' X_i)^{-1} = \Sigma_u + \Sigma_i$$

Poniamo adesso $t = [t_1', t_2', \dots, t_N']'$, ossia una matrice $N \times p$ dove ciascuna linea contiene gli stimatori degli effetti fissi relativi a ciascun soggetto/gruppo.

Le assunzioni del modello possono essere riassunte come segue:

$$(a1) \quad f(t_i - \beta_i | \beta_i) = f(-t_i - \beta_i | \beta_i)$$

$$(a2) \quad g(\beta_i - \beta) = g(-\beta_i - \beta)$$

$$(a3) \quad t_i | (\beta_i) \perp \beta_i$$

$$(a4) \quad t_i \text{ è indipendente da } t_l \text{ per } 1 \leq i < l \leq N$$

Notare che (a1) e (a2) implicano l'assunzione di simmetria dell'unione incondizionata delle distribuzioni dello stimatore, $f(t_i - \beta) = f(-t_i - \beta) \forall i$ (qui f è una funzione di densità non specificata). Questo è assicurato se le usuali assunzioni per i modelli misti sono soddisfatte, inclusa (a3). Questo in pratica corrisponde all'assunzione di simmetria e indipendenza per le distribuzioni condizionate degli stimatori e per la distribuzione di β_i ; l'ultima assunzione riguarda l'assunzione di indipendenza tra i soggetti.

Rimarchiamo inoltre che queste assunzioni riguardano solamente gli stimatori dei coefficienti t , la simmetria non è assunta per i dati originali y_i .

I modelli lineari generalizzati e i modelli lineari misti generalizzati sono perciò casi speciali di questo modello assumendo $\mu = g^{-1}(X_i[\beta + u_i])$ dove g è la funzione legame per la distribuzione di y_{ij} . In questi casi R_i dovrebbe essere scritta più propriamente come $R_i = R_i(\mu_i)$ dalla specifica del modello come per esempio in una regressione binomiale o di Poisson dato che il parametro di dispersione dipende dall'effetto della vera media.

Questo in realtà gioca un ruolo diretto su Σ_i tale che ci obbliga a scrivere $\Sigma_i = \Sigma_i(R_i(\mu_i))$.

Questo modello è inoltre adattabile a casi più generali, come per esempio alle statistiche basate sul rango. Se per esempio l'i-esimo soggetto è misurato sotto due diverse condizioni, la differenza delle medie dei ranghi può essere considerata: $t_i = \bar{r}_1 - \bar{r}_2$, dove $\bar{r}_k = n_k^{-1} r(y_i) X_k$, X_k è il vettore degli indicatori della k-esima condizione sperimentale, $k=1,2,\dots$ e $r(\circ)$ è la trasformazione rango del vettore delle osservazioni interno al soggetto i.

In questo esempio t_i è la stima della vera differenza tra i ranghi medi μ_i , che è una funzione linearizzata dei preditori in (2) mentre ϕ in (1) rimane una funzione non-specificata e sconosciuta.

A parte gli stimatori t_i basati sulle combinazioni lineari, ulteriori estensioni possono essere considerate. Come ad esempio nel caso delle "due condizioni per soggetto", noi potremmo considerare la differenza in mediana di tutte le osservazioni tra le due condizioni. In questo caso $\mu_i = h(X_{ij}, \theta_i)$ in (2) non può essere scritta come combinazione lineare dei preditori.

E' inoltre importante ribadire che gli effetti tra i soggetti possono essere presi in considerazione in questa struttura. La somma di tutti gli effetti è inclusa nella stima della media totale di ciascun soggetto i, perciò ogni modello lineare generalizzato con un intercetta compirà correttamente il lavoro.

Una ulteriore generalizzazione è rappresentata dal modello multivariato con q variabili dipendenti: in questo caso ogni t_i diventa una matrice $p \times q$ nella quale le righe contengono le stime del modello per ogni variabile (oppure come il vettore concatenato $t_i = [t_{i1}, \dots, t_{ip}, \dots, t_{i1q}, \dots, t_{ipq}]$), R_i diventa una matrice $nq \times nq$ e Σ_i diventa una matrice $pq \times pq$ dei preditori (notare che anche la dipendenza tra i preditori è considerata in questo caso).

I modelli possono essere stimati separatamente per ogni variabile (come faremo noi in seguito), oppure unitamente(come per esempio nei modelli normali multivariati).

Le assunzioni (a1)-(a3) devono essere pensate rispetto al vettore concatenato. I test parziali su ciascuna variabile del modello sono fatti simultaneamente e un p-value globale può essere ottenuto combinando queste informazioni, come discuteremo in seguito.

All'interno di questo modello siamo interessati a testare

$$H_0: \beta = \beta_0 \text{ contro } H_1: \beta \neq \beta_0$$

Più precisamente, siamo interessati a testare l'ipotesi nulla della k-esima componente:

$$H_{0k}: \beta_k = \beta_{0k} \text{ contro } H_{1k}: \beta_k \neq \beta_{0k}$$

tale che :

$$H_0 = \bigcap_{k=1}^p H_{0k}$$

2.2 Il test

Dalle assunzioni della sezione precedente, noi sappiamo che

$\{t_i; i=1, \dots, N\}$ ha elementi indipendenti. Nonostante questo però, essi non sono identicamente distribuiti nemmeno nell'ipotesi nulla, per il fatto che hanno la matrice di varianza $\Sigma_u + \Sigma_i$ che ovviamente cambia da soggetto a soggetto.

Il caso particolare in cui $\Sigma_i = \Sigma_i \forall (i, l)$ è raggiunto per esempio quando i soggetti hanno lo stesso numero di osservazioni ($n_i = n \forall i$), e R_i è assunta essere la stessa per ogni soggetto (i.e. $R_i = R$). In questo caso un approccio molto semplice ma efficace utilizza un t-test a un campione parametrico nella k-esima colonna di t per testare $H_{0k}, k=1, \dots, p$.

Quando lo stimatori del gruppo t_i' sono eteroschedastici noi abbiamo bisogno di adottare un test più complesso. Questo è dovuto al fatto che la stima della varianza residua diventa distorta e questo si riflette in una distribuzione di riferimento sbagliata del test statistico.

Al contrario invece, un test esatto può essere ottenuto seguendo le direttive di McNemar:

poniamo $T^{ob} = s'(t - 1\beta_0')$ come valore osservato del test statistico, dove 1 è il vettore $N \times 1$ di "1", e s è un vettore $N \times 1$ di valori uguali a $1/N$.

Per definire la distribuzione di permutazione (o orbita, l'insieme di punti con la stessa verosimiglianza sotto H_0), definiamo S come la collezione dei 2^N punti dello spazio $\{+1/N, -1/N\}^N$ riarrangiati in una matrice $2^N \times N$. Quindi lo spazio di permutazione è dato da $T = S(t - 1\beta_0')$.

Nonostante possa sembrare piuttosto banale, questo test è non-distorto, consistente ed esatto. L'elemento generico T^* dello spazio di permutazione T è dato dal punto le cui coordinate sono il vettore $p \times 1$:

$$\sum_{i=1}^N S_i \frac{(t_i - \beta_0)}{N}, \text{ dove } Pr\{S_i = -1\} = Pr\{S_i = 1\} = 1/2.$$

E' opportuno notare che la dipendenza interna tra gli stimatori dei parametri di ciascun soggetto (tipicamente descritta da Σ_i) è mantenuta da tutti i

componenti del vettore $\frac{(t_i - \beta_0)}{N}$ che sono moltiplicati dallo stesso segno.

Questa strategia di permutazione si applica anche nel caso del multivariato: quì la dipendenza tra le variabili è tradotta nella dipendenza tra gli stimatori

di ciascuna variabile e noi non abbiamo bisogno di esplicitare gli elementi rimasti dal fatto che le correlazioni sono gestite non-parametricamente dalla strategia di permutazione.

Una volta che lo spazio (multivariato) di permutazione è ottenuto, è semplice accedere all'ipotesi nulla globale nel k-esimo predittore applicando le combinazioni non-parametriche dei tests indipendenti (come è stato proposto).

Il valore osservato del test statistico per la k-esima componente di β è

dato $T_k = \frac{\sum_{i=1}^N (t_{ik} - \beta_k)}{N}$, il cui valore atteso e la varianza sono

rispettivamente uguali a 0 e $[\frac{\sum_u}{N} + \frac{\sum_{i=1}^N \Sigma_i}{N^2}]_{kk} = [\frac{\sum_{i=1}^n V(t_i)}{N^2}]_{kk}$, $k=1, \dots, p$ sotto

H_{0k} . E' da notare che questa varianza coinvolge solo gli elementi

diagonali della matrice $\frac{\sum_{i=1}^n V(t_i)}{N^2}$.

Le assunzioni di simmetria realizzano l'esattezza del test indipendentemente dalla corretta specifica del modello: ad esempio uno considera un modello lineare anziché un modello lineare misto.

Un problema può riguardare il potere del test per grandi differenze nelle componenti Σ_i che può compromettere la sensibilità del test. Infatti, se noi lasciamo R_i variare per ciascun soggetto, cioè se il modello è non bilanciato, allora Σ_i varierà anch'essa per ogni soggetto.

Una possibile soluzione è quella di pre-pesare gli estimatori di ciascun soggetto per l'inverso dei loro standard errors, similmente a quanto succede nei GLS, però qui noi pesiamo direttamente gli stimatori anziché le osservazioni originali.

Se noi sapessimo il vero valore $\Sigma_u + \Sigma_i$ per ciascun soggetto noi potremmo standardizzare i valori di T dividendoli per le loro deviazioni standard. Dato che queste matrici sono generalmente sconosciute, noi dovremmo stimare $diag(\Sigma_u + \Sigma_i)$. Come può facilmente essere immaginato, ci sono più strategie alternative per stimare le varianze residue. Essi possono condurci a differenti conclusioni inferenziali ovviamente. Comunque, l'unica caratteristica del test per un punto di vista procedurale è se la stima è fatta assumendo H_0 vera o non. Qui noi discuteremo le due scelte attraverso i due metodi più semplici.

Assumiamo di avere una stima di $diag(\Sigma_i)$ (per esempio la varianza degli stimatori dei parametri nel soggetto i, che chiameremo $\sigma_{ik}^2, (k=1, \dots, p)$, allora noi proveremo a stimare Σ_u . Faremo uso del metodo dei momenti sotto l'ipotesi nulla (la media di ogni stimatore è uguale a 0). In questo caso

$$V(T) = E[T' T] = E \frac{[\sum_{i=1}^N t_i t_i']}{N^2} ; \text{ quindi:}$$

$$diag(E[\sum_{i=1}^N t_i t_i']) = diag(N \Sigma_u + \sum_{i=1}^N \Sigma_i)$$

Poniamo σ_{uk}^2 come il k-esimo elemento della $diag(\Sigma_u)$; per ogni elemento nella diagonale noi otteniamo:

$$\sum_{i=1}^N t_{ik}^2 = diag(N \sigma_{uk}^2 + \sum_{i=1}^N \sigma_{ik}^2)$$

dove σ_{ik}^2 è la vera varianza di t_{ik} . Rimpiazzando σ_{ik}^2 con le loro stime otteniamo:

$$\hat{\sigma}_{uk}^2 = \frac{(\sum_{i=1}^N t_{ik}^2 - \sum_{i=1}^N \hat{\sigma}_{ik}^2)}{N}$$

dove la stima è forzata a 0 in caso sia negativa. Questo implica che non è presente un random effect nella k-esima componente del modello. Notiamo che la stima del σ_{uk}^2 è invariante rispetto alle permutazioni (perciò non è coinvolta dalle permutazioni dei segni delle stime), infatti

$[S_i(t_{ik} - \beta_k)]^2 = (t_{ik} - \beta_k)^2$ sia che S_i sia positivo o no. Noi chiamiamo questa la stima varianza sotto l'ipotesi nulla, che è ottenuta assumendo H_{0k} soddisfatta.

Una volta che gli stimatori della varianza sono disponibili, si può considerare il test statistico:

$$T' = s' \tilde{t}_i$$

Dove ciascun elemento di \tilde{t}_i è la versione standardizzata degli elementi di t_i :

$$\tilde{t}_{ik} = \frac{t_{ik}}{\sqrt{\hat{\sigma}_{uk}^2 + \hat{\sigma}_{ik}^2}} = \frac{\sqrt{N} t_{ik}}{\sqrt{\sum_{i=1}^N t_{ik}^2 - \sum_{l \neq i} \hat{\sigma}_{lk}^2 + (N-1) \hat{\sigma}_{ik}^2}}$$

Notare che questa trasformazione modifica solamente le varianze delle distribuzioni condizionate, ma non coinvolge le medie sotto H_0 (che sono sempre nulle), quindi il test rimane esatto anche nel caso in cui le varianze sono sottostimate o sovrastimate. Questo è il caso in cui il modello ad effetti misti è scelto anzichè un semplice modello lineare. In particolare , il

denominatore di \tilde{t}_{ik} è invariante rispetto alle permutazioni.

L'alternativa, possibile, strategia è quella di stimare la varianza degli stimatori sotto l'ipotesi alternativa, ossia quando essa è assunta vera. Questo è dovuto al fatto che, quando l'ipotesi alternativa è vera, le varianze sono sovrastimate. In questo caso le stime delle varianze dei random effect sono basate sulla varianza di t_i , piuttosto che sul loro secondo momento.

Usando ancora la stima dei momenti, gli elementi della diagonale k-esima diventano:

$$\sum_{i=1}^N t_{ik}^2 - N \bar{t}_k^2 = (N \sigma_{ik}^2 + \sum_{i=1}^N \hat{\sigma}_{ik}^2)$$

con:

$$\bar{t}_k = \frac{1}{N} \sum_{i=1}^N t_{ik}$$

Applicando questa scelta nelle stime, gli stimatori dei coefficienti modificati (e osservati) diventano uguali a:

$$\tilde{t}_{ik} = \frac{\sqrt{N} t_{ik}}{\sqrt{\sum_{i=1}^N (t_{ik} - \bar{t}_k)^2 - \sum_{l \neq i} \hat{\sigma}_{lk}^2 + (N-1) \hat{\sigma}_{ik}^2}}$$

Ora, tenendo conto che \bar{t}_k cambia ad ogni permutazione, il denominatore deve essere ristimato ad ogni permutazione.

Tirando le somme sulle possibilità che abbiamo illustrato, possiamo concludere che:

1. Il test originale è stato pensato per gli stimatori dei coefficienti fissi così come sono. La strategia di permutazione coinvolge cambiamenti indipendenti nei segni del vettore t_i ; questo test è esatto nonostante però possa essere un po' sensibile in alcuni casi.
2. Una prima modifica riguarda il ripesaggio degli stimatori de coefficienti con l'inverso dei loro standard error, quando H_0 è assunta vera. La strategia di permutazione è ottenuta modificando i segni dei vettori \tilde{t}_i ; quindi , il processo di ripesaggio è applicato solamente all'inizio.
3. Un ulteriore modifica riguarda il ripesaggio degli stimatori dei coefficienti ad ogni permutazione, tenendo conto che i loro standard error cambiano ad ogni permutazione, ci troviamo in questo caso sotto H_1 ;

Noi ribadiamo il fatto che, l'intera procedura può essere ripetuta forzando $\Sigma_u=0$, ossia considerando i modelli lineari generalizzati anziché quelli misti. Questo fa sì che la strategia 2 coincida con la strategia 3. Più interessanti, la scelta non coinvolge l'esattezza del test, questo chiarifica perchè l'asserto che questo test non è compromesso da una cattiva specifica del modello, mentre MLE invece ovviamente lo è.

Che tipo di test sia da utilizzare dipende dalle assunzioni del modello, in generale quindi noi suggeriamo la versione originale nel caso in cui siamo di fronte ad un modello bilanciato e la varianza interna è assunta uguale, la soluzione 2 è più adeguata nel caso in cui i random effect sono considerevolmente irrilevanti e la soluzione 3 invece copre i casi rimanenti.

Indipendentemente dalla scelta ribadiamo che questi test sono esatti, tuttavia possono differire riguardo alle prestazioni (la potenza del test).

Per concludere la trattazione teorica osserviamo che il test è facilmente estendibile all'inferenza multivariata tenendo conto che la distribuzione congiunta del test statistico è sempre disponibile senza ulteriori assunzioni nel modello.

3 I CASI IN ESAME

Dopo aver illustrato teoricamente le procedure per i test di permutazione sui modelli misti, ora procederemo ad applicarli su due dataset reali.

3.1 Analisi Snijders Data

Le ibridazioni genomiche comparative basate sui micro-array presentate da Snijders (2001) sono disponibili gratuitamente in internet.

Si tratta di un dataset di 15 cellule sottoposte all'azione di un cromosoma batterico artificiale (BACs) dal NIGMS deposito di cellule genetiche umane (Coriell Institute for Medical Research)

La media log₂-ratios degli spots di 2,463 geni (i.e. variabili) è disponibile per tutte le linee cellulari. Sono anche disponibili il numero di repliche valide e la deviazione standard osservata per ogni gene in ogni linea cellulare.

I geni che non hanno almeno 10 osservazioni sono esclusi, questo ci porta ad avere 2,293 geni rimanenti. Nel dataset risultante ancora un 8% di osservazioni sono mancanti.

Siamo interessati a trovare gli spots con log₂-ratio diverso da zero.

È uso comune in questi studi di affrontare il problema con un t-test per ogni gene sulla base del valore di mean log₂-ratios. In questa applicazione invece cercheremo di utilizzare anche le informazioni di “numero di repliche valide” e deviazione standard osservata. Per fare questo un modello misto adeguato allo scopo è il seguente:

- il modello ha $X_i = Z_i$ matrici $n_i \times 1$ di “1”
- l'unico coefficiente fisso β è la media degli elementi totali: $\beta = \bar{y}$ e $\beta_i = \bar{y}_i$ dove le medie \bar{y}_i sono i veri valori del parametro.
- assumiamo ε_i e b_i con varianza variabile da soggetto a soggetto

in definitiva il modello è definito come segue

$$y_i \sim \bar{y} + b_i + \varepsilon_i \quad \text{con } \varepsilon_i \sim N(0, \sigma_i^2 I_{n_i}) \text{ e } b_i \sim N(0, \sigma_{bi}^2 I_{n_i})$$

Quello a cui siamo interessati nella nostra analisi è il test di ipotesi di nullità:

$$H_0: \bar{y} = 0 \quad \text{contro} \quad \bar{y} \neq 0$$

per poterlo effettuare ovviamente dobbiamo essere a conoscenza di Σ_i che nel nostro caso è la varianza dello stimatore \bar{y}_i che sappiamo bene essere:

$$\frac{\sigma_i^2}{n_i} \quad \text{per ciascun genoma.}$$

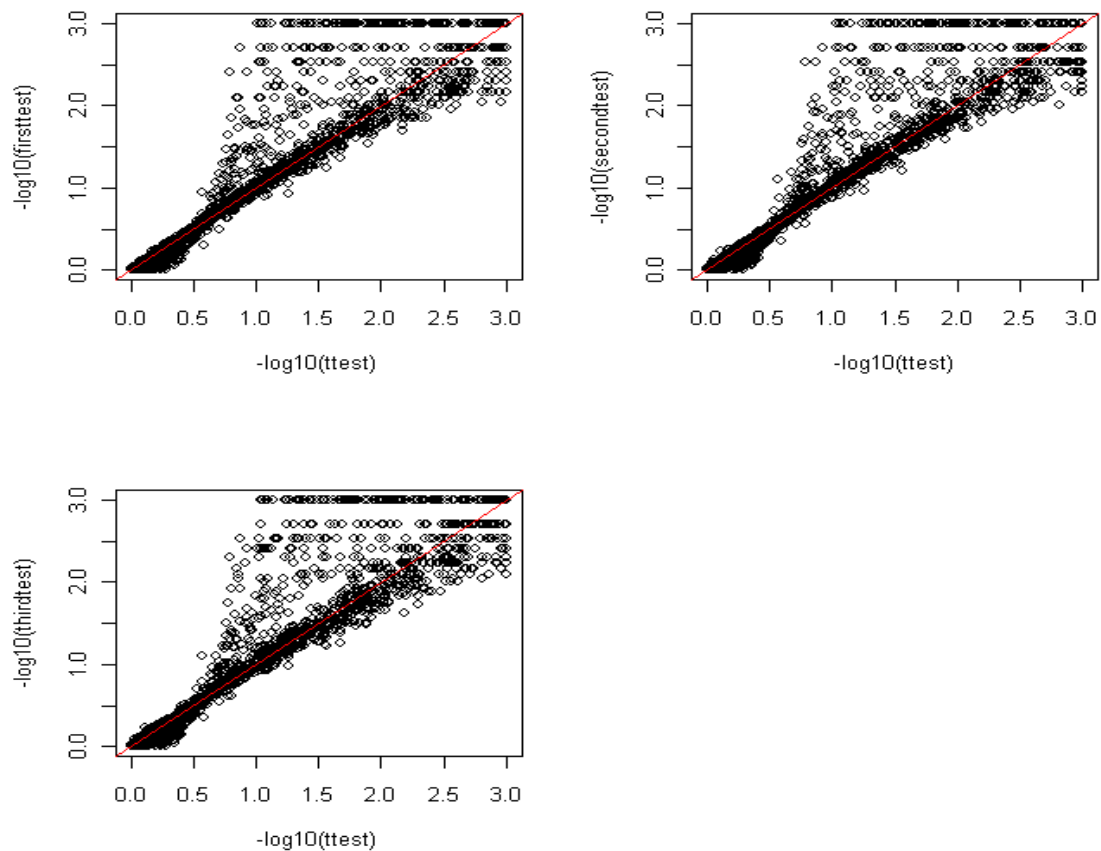
A tale scopo quindi sono forniti nel nostro dataset anche la matrice dei $\hat{\sigma}_i$ e le repliche tecniche (le numerosità campionare per ogni individuo) n_i che ovviamente hanno entrambe la stessa dimensione della matrice Y delle stime.

Per le 2293 variabili rimanenti da questo punto è possibile calcolare i p-value corrispondenti per i tre approcci di permutazione. I risultati sono descritti in figura 1.

Ciò che osserviamo è una nuvola di punti concentrata nella zona in alto a destra dei grafici, sopra la retta che divide il grafico, ciò significa che apparentemente i test di permutazione, esprimono meglio la significatività dei dati la cui media è evidentemente diversa da 0. Per conferma controlliamo anche la percentuale di punti sotto 5% e 1% per i 4 test e infatti otteniamo:

```
> c(mean(ttest<.01),mean(ttest<.05))  
[1] 0.3017881 0.4461404  
> c(mean(firsttest<.01),mean(firsttest<.05))  
[1] 0.3501962 0.4731792  
> c(mean(secondtest<.01),mean(secondtest<.05))  
[1] 0.3467074 0.4753598  
> c(mean(thirdtest<.01),mean(thirdtest<.05))  
[1] 0.3593546 0.4792848
```

Figura 1



3.2 Analisi Bottles

Il secondo dataset riguarda uno studio sperimentale sulla motricità. I risultati sono ancora in via di pubblicazione per cui in questa sede ci limiteremo ad analizzarne una parte e a darne una descrizione puramente in termini modellistici più che di contenuto. Il campione è costituito da 7 individui, a ciascun di essi è stato misurato più volte nell'atto di spostare una bottiglia. Gli indicatori rilevati riguardano il movimento compiuto: la latenza e l'ampiezza del movimento delle varie accelerazioni, decelerazioni, velocità compiute dalla bottiglia. Come anticipato, non daremo qui il dettaglio degli indicatori misurati, ma ci limiteremo a evidenziare che variabili totali sono 17 e quindi la risposta è di tipo multivariato.

Le diverse condizioni del test si basano sul fatto che in alcuni casi lo

spostamento è avvenuto quando l'individuo sapeva già il peso della bottiglia (sapeva quindi se la bottiglia era pesante o leggera) e nei casi rimanenti no oltre al fatto che alcune bottiglie erano pesanti e altre leggere.

Lo scopo del test è capire se i fattori Peso conosciuto/non conosciuto (variabile p_con del nostro modello), bottiglia pesante/leggera (variabile Peso del nostro modello) e interazione tra i due sono determinanti nel differenziare i 17 indicatori di risposta che menzionavamo in precedenza. Stiamo parlando di un modello di regressione fattoriale che però presenta ovviamente un random effect dovuto alla presenza degli individui (osserviamo che questo modello è un caso analogo a quello dell'esempio di pagina 4 con due fattori anziché uno).

Procederemo quindi come nel dataset precedente ma dovremo questa volta stimare un modello per ciascuna variabile e ciascun soggetto. Avremo quindi una matrice 7×68 di coefficienti e di standard errors.

Proviamo inizialmente a verificare quanto i due fattori influiscano sulle 17 variabili prese singolarmente. A tale scopo effettuiamo un test univariato analogo a quello effettuato per il dataset precedente rispetto ai 4 coefficienti in analisi: l'intercetta, il fattore perso, il fattore peso conosciuto, e l'interazione tra i due.

I p-value per i test di nullità di ciascun coefficiente riferiti ai tre approcci di permutazione si ottengono confrontando gli spazi di permutazione dei tre approcci con la statistica osservata e sono elencati in figura 1.

Ora però quello che ci interessa maggiormente è vedere quanto i due fattori e l'interazione siano significativi rispetto alle variabili risposta prese simultaneamente, questo risultato è ottenuto tramite una combinazione non parametrica di test di permutazione. La funzione di combinazione utilizzata in questo caso è quella di Fisher.

Quello che otteniamo per i nostri dati per ciascuno dei tre metodi è il seguente in Tabella 2:

Osserviamo che p_con non è significativo e può perciò essere tolto dalla nostra analisi mentre l'interazione è significativa o quantomeno border line.

Tabella 1:

	first	second	third
acc_LAT1 .PesoPesante	1	0.75	0.75
acc_LAT1 .p_conSI	0.8125	0.78125	0.78125
acc_LAT1 .PesoPesante:p_conSI	0.46875	0.5	0.5
acc_ACC1 .PesoPesante	0.1875	0.171875	0.171875
acc_ACC1 .p_conSI	0.5625	0.5625	0.5625
acc_ACC1 .PesoPesante:p_conSI	0.9375	0.890625	0.890625
acc_LAT2 .PesoPesante	0.328125	0.234375	0.234375
acc_LAT2 .p_conSI	0.0625	0.0625	0.0625
acc_LAT2 .PesoPesante:p_conSI	0.71875	0.78125	0.78125
acc_ACC2 .PesoPesante	0.203125	0.21875	0.21875
acc_ACC2 .p_conSI	0.84375	0.84375	0.84375
acc_ACC2 .PesoPesante:p_conSI	0.09375	0.09375	0.09375
acc_LAT3 .PesoPesante	0.03125	0.03125	0.03125
acc_LAT3 .p_conSI	0.65625	0.65625	0.65625
acc_LAT3 .PesoPesante:p_conSI	0.015625	0.015625	0.015625
acc_ACC3 .PesoPesante	0.03125	0.03125	0.03125
acc_ACC3 .p_conSI	0.09375	0.09375	0.09375
acc_ACC3 .PesoPesante:p_conSI	0.015625	0.015625	0.015625
dec_LAT1 .PesoPesante	0.71875	0.71875	0.71875
dec_LAT1 .p_conSI	0.953125	0.9375	0.9375
dec_LAT1 .PesoPesante:p_conSI	0.09375	0.046875	0.046875
dec_DEC1 .PesoPesante	0.546875	0.796875	0.796875
dec_DEC1 .p_conSI	0.1875	0.21875	0.21875
dec_DEC1 .PesoPesante:p_conSI	0.03125	0.03125	0.03125
dec_LAT2 .PesoPesante	0.265625	0.21875	0.21875
dec_LAT2 .p_conSI	0.34375	0.34375	0.34375
dec_LAT2 .PesoPesante:p_conSI	0.890625	0.84375	0.84375
dec_DEC2 .PesoPesante	0.9375	0.828125	0.828125
dec_DEC2 .p_conSI	0.625	0.578125	0.578125
dec_DEC2 .PesoPesante:p_conSI	0.09375	0.09375	0.09375
aper_LAT .PesoPesante	0.75	1	1
aper_LAT .p_conSI	0.65625	0.625	0.625
aper_LAT .PesoPesante:p_conSI	0.03125	0.03125	0.03125
aper_MaxAP .PesoPesante	0.4375	0.515625	0.515625
aper_MaxAP .p_conSI	0.75	0.78125	0.78125
aper_MaxAP .PesoPesante:p_conSI	0.265625	0.265625	0.265625
vel_LAT1 .PesoPesante	0.09375	0.078125	0.078125
vel_LAT1 .p_conSI	0.34375	0.34375	0.34375
vel_LAT1 .PesoPesante:p_conSI	0.71875	0.671875	0.671875
vel_PVEL1 .PesoPesante	0.671875	0.609375	0.609375
vel_PVEL1 .p_conSI	0.296875	0.28125	0.28125
vel_PVEL1 .PesoPesante:p_conSI	0.046875	0.0625	0.0625
vel_LAT2 .PesoPesante	0.015625	0.015625	0.015625
vel_LAT2 .p_conSI	0.609375	0.609375	0.609375
vel_LAT2 .PesoPesante:p_conSI	0.09375	0.09375	0.09375
vel_PVEL2 .PesoPesante	0.0625	0.0625	0.0625
vel_PVEL2 .p_conSI	0.53125	0.546875	0.546875
vel_PVEL2 .PesoPesante:p_conSI	0.3125	0.28125	0.28125
z_ALT .PesoPesante	0.015625	0.015625	0.015625
z_ALT .p_conSI	0.015625	0.015625	0.015625
z_ALT .PesoPesante:p_conSI	0.015625	0.015625	0.015625

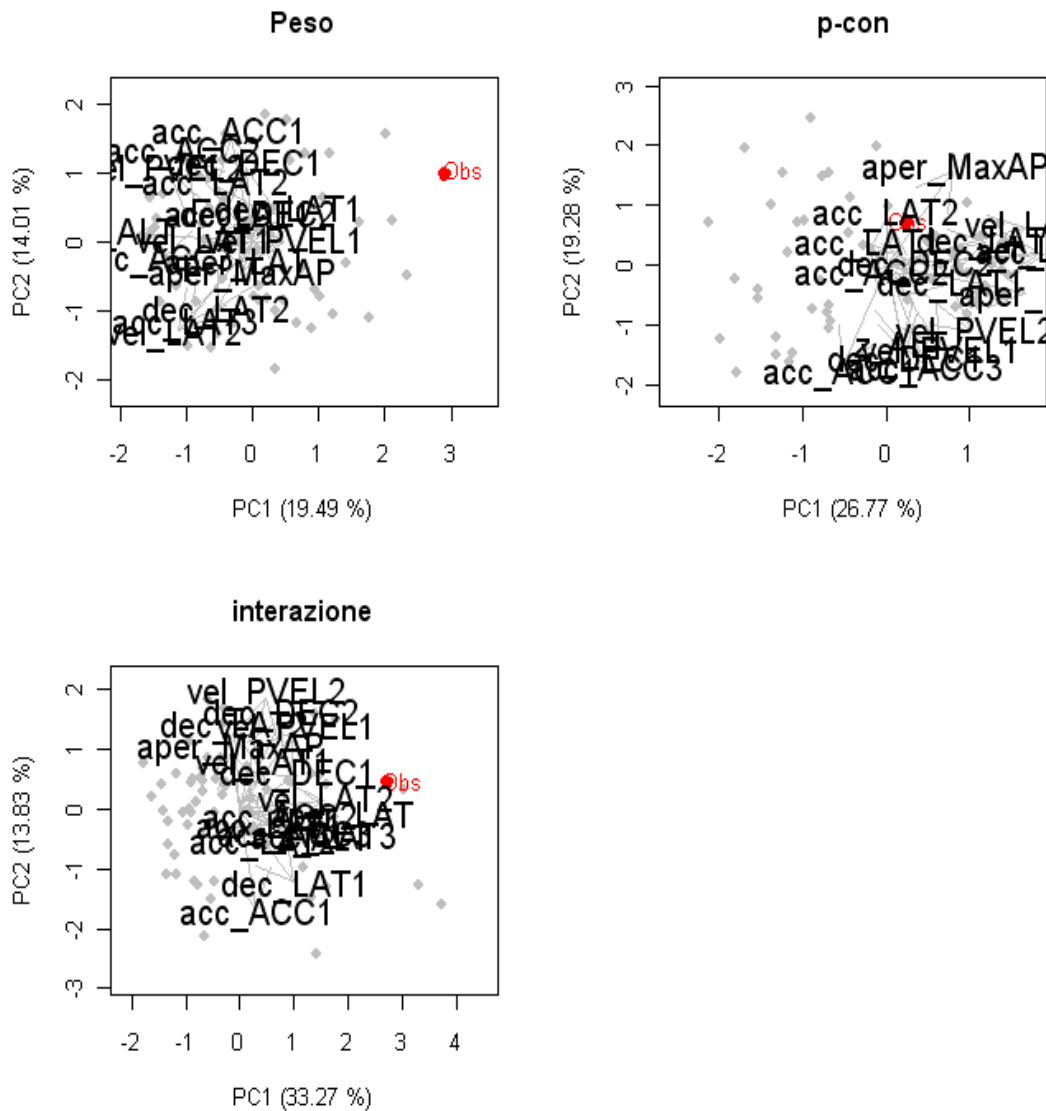
Tabella 2:

	Peso	p_con	interazione
primo test	0.03125	0.40625	0.046875
secondo test	0.03125	0.421875	0.046875
terzo test	0.03125	0.421875	0.046875

Proviamo a creare dei grafici per verificare la cosa osservando le componenti principali.

Ciò che otteniamo lo mostriamo in figura 2:

Figura 2:



Anche qui osserviamo che p_{con} è non significativo però anche in questo caso l'interazione è significativa. Ciò può essere dovuto al fatto che in realtà “ p_{con} ” influisce solamente su una delle due condizioni del primo fattore “Peso”. Quello che faremo quindi sarà capire quale delle due viene influenzata e quale no.

Creiamo quindi due diversi modelli condizionandoci al fattore peso: leggero o pesante.

Per quanto riguarda le bottiglie leggere osserviamo (riformulando il modello) un p-value non-significativo per la variabile p_{con} .

Ora proviamo a procedere analogamente per le bottiglie pesanti e osserviamo che il p-value come previsto è significativo. I risultati sono descritti in Tabella 3:

Tabella 3:

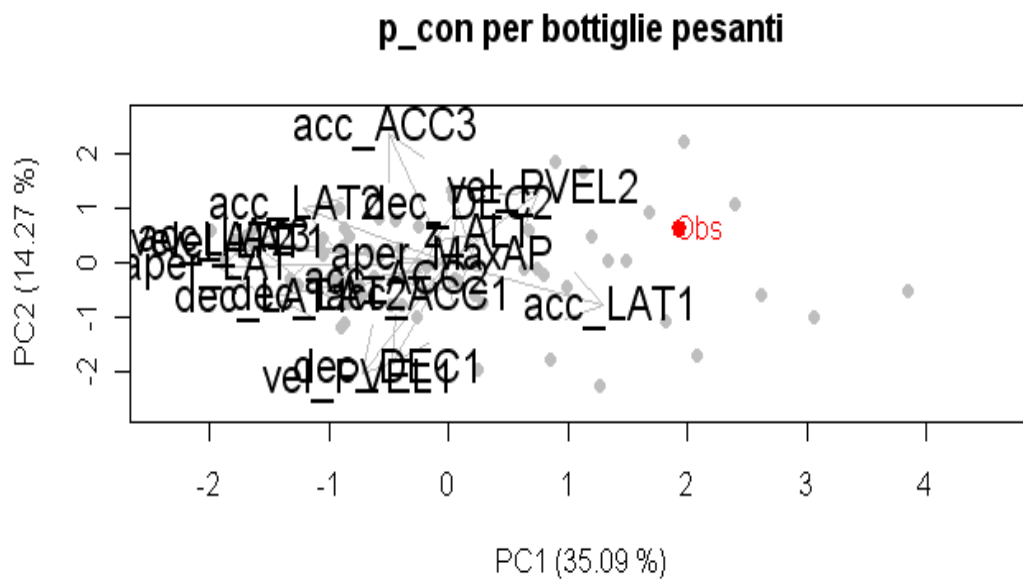
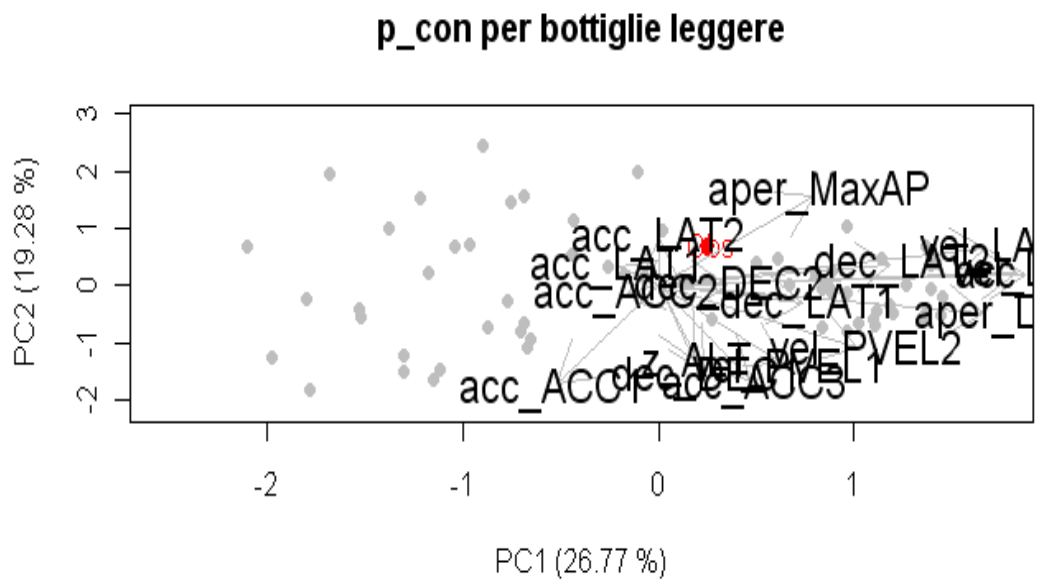
	global.pv.light	global.pv.heavy
primo test	0.40625	0.046875
secondo test	0.421875	0.046875
terzo test	0.421875	0.046875

Ciò in sostanza significa che la variabile p_{con} è significativa solamente se la bottiglia spostata è pesante.

Per concludere proviamo a creare due grafici analoghi a quello in figura 2 per i due modelli ridotti (Figura 3).

I grafici confermano le nostre ipotesi.

Figura 3:



4 APPENDICE

Di seguito elencheremo i codici in R delle funzioni interessate in questa analisi.

4.1 Funzione t2p

La funzione t2p calcola lo spazio di permutazione dei p-value quando obs.only=FALSE altrimenti calcola semplicemente il p-value per la statistica test osservata:

```
t2p<-function(T,obs.only=TRUE){  
  if(!is.matrix(T)) {T<-as.matrix(T)}  
  if(obs.only) {  
    return(apply(T,2,function(permy)mean(permy>=permy[1])))  
  }  
  else{  
    B<-dim(T)[1]  
    P=apply(-T,2,rank,ties.method="max")/B  
    P=as.matrix(P)  
    rownames(P)=c("p-obs",paste("p-*",1:(B-1),sep=""))  
  }  
  colnames(P)=colnames(T)  
  return(P)  
}
```

4.2 Funzione t.nptest

La funzione t.nptest effettua il primo dei 3 approcci per il calcolo della matrice di permutazione.

```
t.nptest <- function(Y, perms=1000, alternative =
```

```

NULL,permP=FALSE,permT=FALSE){
  if(!is.matrix(Y)) Y=as.matrix(Y)
  N=dim(Y)[1]
  if (is.list(perms)) {
if (is.numeric(perms$seed)) {
  set.seed(perms$seed)
  perms = perms$number
}
}

# if it is a matrix, just use it to switch signs
if(!is.matrix(perms)){
  if (2^N <= perms) {
    # all permutations if possible and if no stratas
    random <- FALSE
    library(e1071)
    perms<-bincombinations(N)
    perms[which(perms==0)] <- -1/N
    perms[which(perms==1)] <- 1/N
  } else {
    #otherwise random permutations
    random <- TRUE
    perms <- rbind(rep(1, N), matrix(1 - 2 * rbinom(N *
perms,1, 0.5), perms, N))/N
  }
}

Y[is.na(Y)]=0

permY <- perms %*% Y

```

```

#make hight values of the statistics to be significant
if (is.null(alternative)) {
  alternative = rep(0, dim(Y)[2])
} else if (length(alternative) != dim(Y)[2])
  alternative <- rep(alternative[1], dim(Y)[2])

permY[, alternative < 0] <- -permY[, alternative < 0]
permY[, alternative == 0] <- abs(permY[, alternative == 0])
  permY[is.na(permY)]=0

  p=t2p(permY)

  if(permP) permP=t2p(permY, obs.only=F)
  else permP=NULL

  if(permT) permT=permY
  else permT=NULL

  list(p=p,permP=permP,permT=permT)
}

```

4.3 Funzione random.coeff e random.coeff2

Le due funzioni calcolano rispettivamente le matrici di permutazione per il secondo e terzo approccio di permutazione.

```

random.coeff <- function(Yall, perms=1000, sigma2Th=NULL,
  alternative = NULL, permP=FALSE, permT=FALSE){
  Y=Yall$Y

```

```

se=Yall$se
N <- dim(Y)[1]
if(is.null(sigma2Th)){
  #stima di sigma2_Theta
  realN<-apply(Y,2,function(x){sum(!is.na(x))})
  Y[is.na(Y)]=0
  sigma2Thtemp <- (apply(Y^2,2,sum)-
apply(se^2,2,sum,na.rm=T))/realN
  sigma2Th <- pmax(0,sigma2Thtemp)
} else {
  sigma2Th=rep(sigma2Th,dim(Y)[2])
  sigma2Thtemp=sigma2Th
}

se <- sqrt(matrix(sigma2Th, byrow=TRUE,nrow=dim(se)
[1],ncol=length(sigma2Th)) + se^2)
Y <- Y/se
m <- apply(Y,2,mean)
Sigma2 <- sigma2Thtemp
# perform 1 sample test on Y
ps=t.nptest(Y=Y, perms=perms, alternative =
alternative,permP=permP,permT=permT)
list(permP=ps$permP, permT=ps$permT, res=cbind(Estimate=m,
Sigma2=Sigma2 , p.value=ps$p), Y=Yall)
}

```

```

random.coeff2 <- function(Yall, perms=1000, alternative =
NULL,testedCoef=NULL,permP=FALSE,permT=FALSE){

```

```

Y=Yall$Y
se=Yall$se
N <- dim(Y)[1]
if (is.list(perms)) {
  if (is.numeric(perms$seed)) {
    set.seed(perms$seed)
    perms = perms$number
  }
}

# if it is a matrix, just use it to switch signs
if(is.vector(perms)){
  if (2^N <= perms) {
    # all permutations if possible and if no stratas
    random <- FALSE
    library(e1071)
    SignsMatrix<-bincombinations(N)
    SignsMatrix[which(SignsMatrix==0)] <- -1
  } else {
    #otherwise random permutations
    random <- TRUE
    SignsMatrix <- rbind(rep(1, N), matrix(1 - 2 *
rbinom(N * perms,1, 0.5), perms, N))
  }
}

sumse=apply(Y^2,2,sum,na.rm=T)-apply(se^2,2,sum,na.rm=T)
permY=matrix(NA,dim(SignsMatrix)[1],dim(Y)[2])
realN<-apply(Y,2,function(x){sum(!is.na(x))})
Y[is.na(Y)]=0
for (i in dim(SignsMatrix)[1]:1) {

```

```

temp <- matrix(rep(SignsMatrix[i,],dim(Y)[2]),ncol=dim(Y)
[2]) * Y

#stima di sigma2_Theta
sigma2Th <- (sumse-apply(temp,2,mean)^2)/realN
sigma2Th <- pmax(0,sigma2Th)
tempse <- sqrt(matrix(sigma2Th,
byrow=TRUE,nrow=dim(se)[1],ncol=length(sigma2Th)) + se^2)
temp <- temp/tempse
permY[i,] <- apply(temp,2,sum,na.rm=T)
}

#make hight values of the statistics to be significative
if (is.null(alternative)) {
  alternative = rep(0, dim(Y)[2])
} else if (length(alternative) != dim(Y)[2])
  alternative <- rep(alternative[1], dim(Y)[2])
permY[, alternative != 0] <- permY[, alternative != 0] %*%
diag(alternative[alternative != 0])
permY[, alternative == 0] <- abs(permY[, alternative == 0])
permY[is.na(permY)]=0
colnames(permY)=colnames(Y)
if(permT) permT=permY
else permT=NULL
list(res=cbind(Estimate=NA, Sigma2=sigma2Th,
p.value=t2p(permY),sigma2sogg=NA),Y=Yall,
permP=t2p(permY,obs.only=F),permT=permT)
}

```


4.4 funzione make.Y

La funzione make.Y calcola la matrice dei coefficienti e degli standard error per il modello descritto dalla variabile model, dividendolo nei vari clusters:

```
make.Y <- function(model, clusters, se=NA, testedCoef=NULL){

  if (!is.list(model)) model=list(model)
  idClust=unique(clusters)
  names(idClust)=idClust
  #if se has not the same size of Y build it as matrix
  ncoef=sum(sapply(model,function(mod)
length(mod$coefficients)))
  if(!all(dim(as.matrix(se))==c(length(idClust),ncoef)))
se=matrix(se[1:min(ncoef,prod(dim(se))),ncol=ncoef,nrow=length(idCl
ust),byrow=TRUE)

  Yall <- lapply(idClust,function(idsel)
{res=list(coef=NULL,se=NULL,df.mod=NULL,df.res=NULL);
                                                    for(i in
1:length(model) ){

  newmodel = try(update(model[[i]],subset=(clusters==idsel)),silent
= TRUE)

  if(is(newmodel,"try-error")){

temp<-matrix(NA,nrow=length(model[[i]]$coef),ncol=2)
```

```

colnames(temp)<-c("Estimate","Std. Error")
}

else

temp <- summary(newmodel)$coef

if(dim(temp)[1]!=length(model[[i]]$coef))

for(j in 1:length(model[[i]]$coef)){

    if(is.na(names(model[[i]]$coef[j])!=rownames(temp)[j])||
names(model[[i]]$coef[j])!=rownames(temp)[j]){

        if(j==1){

            temp<-rbind(c(NA,NA),temp)

        }

        else

            if(j>dim(temp)[1])

temp<-rbind(temp,c(NA,NA))

else{

sub<-rbind(temp[1:(j-1)],c(NA,NA))

temp<-rbind(sub,temp[j:dim(temp)[1],])

```

```

}

rownames(temp)[j]<-names(model[[i]]$coef[j])
}

}

rownames(temp) <-
paste(names(model)[i],rownames(temp),sep=".")
res$coef=c(res$coef,temp[,"Estimate"])
res$se=c(res$se,temp[,"Std.
Error"]) res$df.mod =
c(res$df.mod, rep(dim(temp)[1],dim(temp)[1]))

res$df.res = c(res$df.res, rep(sum(clusters==idSel)-dim(temp)
[1],dim(temp)[1]))

res})

#stima di se
se[is.na(se)] <- t(sapply(Yall ,function(y) y$se))[is.na(se)]
df.res=t(sapply(Yall ,function(y) y$df.res))
df.mod=t(sapply(Yall ,function(y) y$df.mod))
Y <- t(sapply(Yall ,function(y) y$coef) )
rownames(se) <- rownames(df.res) <- rownames(df.mod) <-
rownames(Y)
colnames(se) <- colnames(df.res) <- colnames(df.mod) <-
colnames(Y)
if(!is.null(testedCoef)) { Y <- Y[,testedCoef,drop=FALSE]; se <-
se[,testedCoef,drop=FALSE]}
Yall=list(Y=Y, se=se, df.res=df.res,df.mod=df.mod)
}

```


5 BIBLIOGRAFIA

Eugene Demidenko, Wiley, “Mixed Models, Theory and Applications”.

Finos Livio, Basso Dario, “Exact Multivariate Permutation Tests for Fixed Effects in Mixed-Models” dispense.

Snijders, A. M., Nowak, N., Segraves, R., Blackwood, S., Brown, N., Conroy, J., Hamilton, G., Hindle, A. K., Huey, B., Kimura, K., Law, S., Myambo, K., Palmer, J., Ylstra, B., Yue, J. P., Gray, J. W., Jain, A. N., Pinkel, D., and Albertson, D. G. (2001). “Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat Genet*, 29(3):263–264.”

Pesarin, “ Multivariate Permutation Tests with applications in Biostatistics.” John Wiley and Sons, Chichester. 2001