

# Indice

<b>Premessa</b>	<b>v</b>
<b>1 Introduzione allo studio</b>	<b>1</b>
1.1 Il metodo . . . . .	2
1.2 I limiti dello studio . . . . .	3
<b>2 I metodi di combinazione non parametrici</b>	<b>5</b>
2.1 La teoria . . . . .	5
2.2 Funzioni di combinazione non parametrica . . . . .	7
2.3 Tecniche di ricampionamento condizionato (C.M.C.) . . . . .	9
<b>3 Problemi con dati appaiati: soluzione di permutazione</b>	<b>13</b>
3.1 Il disegno sperimentale . . . . .	13
3.2 Caso univariato . . . . .	14
3.3 Soluzione parametrica: t di Student . . . . .	18
3.3.1 Dati appaiati normalmente distribuiti . . . . .	18
3.4 Soluzione non parametrica: test di permutazione . . . . .	20
3.5 Caso multivariato . . . . .	23
3.6 Cross-over design . . . . .	28
3.7 Digressione sui disegni cross-over . . . . .	30
3.8 Verifica d'indipendenza . . . . .	32
<b>4 Test di permutazione con missing data</b>	<b>41</b>
4.1 Introduzione al problema dei missing data . . . . .	41
4.2 Esempi dei processi che generano i <i>missing-data</i> . . . . .	42
4.3 Approccio di permutazione . . . . .	45
4.4 Modelli non-MAR . . . . .	47
4.5 Combinazione non parametrica . . . . .	48
4.6 Modelli MCAR . . . . .	49
4.7 Permutazioni di <i>missing data</i> : struttura e analisi . . . . .	49
4.8 Soluzione per problemi MCAR . . . . .	51

4.8.1	Estensione ai modelli non-MAR . . . . .	53
4.9	Dati appaiati e <i>missing value</i> . . . . .	54
4.10	Intention to treat analysis . . . . .	55
<b>5</b>	<b>Risultati dello studio</b>	<b>59</b>
5.1	Analisi con <i>missing values</i> . . . . .	59
5.2	Analisi con ITT ( <i>Intention To Treat</i> ) . . . . .	62
<b>A</b>	<b><i>ECDF</i> delle simulazioni</b>	<b>69</b>
<b>B</b>	<b>Risultati stratificazione</b>	<b>75</b>
	<b>Bibliografia</b>	<b>81</b>
	<b>Ringraziamenti</b>	<b>87</b>

UNIVERSITÀ DEGLI STUDI DI PADOVA  
FACOLTA' DI SCIENZE STATISTICHE  
CORSO DI LAUREA IN  
SCIENZE STATISTICHE ED ECONOMICHE

TESI DI LAUREA

ANALISI MULTIVARIATA PER  
OSSERVAZIONI APPAIATE CON DATI  
MANCANTI

Relatore: Chiar.mo Prof. Fortunato Pesarin

Co-relatore: Prof. Giovanni Fava

Laureanda: Chiara Brombin

ANNO ACCADEMICO 2004-2005



Alla mia famiglia



# Premessa

I metodi parametrici utilizzati per la soluzione di problemi di carattere univariato e multivariato hanno, come limitazione, la necessità di dover ricorrere all'introduzione di ipotesi molto restrittive, spesso ingiustificate se non impossibili da giustificare, irrealistiche, non sempre chiare, difficilmente interpretabili, formulate *ad hoc* per poter fare inferenza. A questo si deve aggiungere che le assunzioni che rendono valida l'applicazione di tali metodi (normalità, omoschedasticità, indipendenza e identica distribuzione della componente stocastica erratica) sono di norma raramente soddisfatte e, quand'anche soddisfatte, i risultati sono spesso ottenuti tramite approssimazione.

Ci sono casi in cui, anche assumendo la normalità della distribuzione degli errori, le soluzioni parametriche non forniscono risultati esatti: ci si riferisce per esempio ai test per campioni con osservazioni appaiate quando i parametri di scala dipendono dalle unità o ci sono dati mancanti, allo studio dell'analisi della varianza a blocchi (two-way ANOVA), a problemi derivanti dal fatto che il numero delle variabili rilevate è superiore alla dimensione del campione, oppure nei test a due campioni quando si vuole fare inferenza congiuntamente su media e varianze e le risposte sono positive.

Sempre più spesso, per problemi multivariati più complessi studiati in ambito biomedico, ingegneristico, psicologico, farmacologico, negli esperimenti clinici, nel controllo della qualità, quando non è noto il modello distributivo, non si può invocare la normalità o ancora quando l'inferenza riguarda variabili di tipo qualitativo, si passa da un approccio parametrico ad uno non parametrico, ovviando così, senza perdita sostanziale di efficienza, le limitazioni sopra accennate. E tra i metodi non parametrici, un ruolo di grande rilievo per le proprietà generali di cui godono, vi sono i metodi di permutazione.

Non bisogna però lasciarsi trarre in inganno, ovvero non concludere erroneamente che i metodi non parametrici di permutazione siano una *panacea*, per tutti i problemi inferenziali di interesse: se, sotto  $H_0$ , non ci si condiziona ad un insieme di statistiche sufficienti e non si assume l'ipotesi di scambiabilità dei dati le soluzioni ottenute saranno tutt'altro che esatte [1].

Nei problemi multivariati, la soluzione di permutazione si ottiene applicando un preciso metodo di combinazione non parametrica di test dipendenti [2]. Tale metodo si basa sulla decomposizione delle ipotesi da testare in  $k$ , con  $k > 1$ , sottoipotesi per ciascuna delle quali si calcola un appropriato test di permutazione parziale, utilizzando poi un processo di simulazione consistente in  $B$  ricampionamenti casuali senza reinserimento, tramite condizionamento all'insieme dei dati osservati, si ottiene una stima della distribuzione di permutazione della statistica test. Infine i risultati parziali della simulazione condizionata vengono combinati in una statistica test del secondo ordine tramite una funzione reale (misurabile) continua, monotona decrescente e non degenera [3].



# Capitolo 1

## Introduzione allo studio

Recenti studi in psicologia e neuropsichiatria infantile hanno messo in luce un aumento del numero di bambini che presentano disturbi comportamentali, disagi di tipo psicologico come ansia, depressione, stress, incapacità di concentrazione, disturbi alimentari (anoressia, bulimia, obesità) che si traducono in una difficoltà nella gestione dei rapporti interpersonali e si riflettono nei processi evolutivi, d'apprendimento e di sviluppo [4] [5].

Nell'intento di prevenire abusi e sofferenze nei bambini e negli adolescenti, associazioni ed enti no-profit per la tutela e protezione dei diritti d'infanzia si adoperano sempre più per promuovere interventi scolastici mirati a ristabilire l'equilibrio affettivo nel bambino. Accanto ad interventi della tradizione psicologica più consolidata, come la terapia cognitivo-comportamentale basata sulla correzione delle distorsioni cognitive e sull'insegnamento di una serie di abilità coping e problem solving per migliorare l'adattamento psicologico in situazioni di stress, si sta diffondendo un filone di ricerca d'impronta americana noto come *Positive Psychology*.

Il concetto cardine della *Positive Psychology* è la necessità di distogliere lo sguardo dai deficits psicologici e comportamentali che caratterizzano lo sviluppo del bambino e di concentrarsi piuttosto sulle competenze e i punti forza del bambino stesso, ma anche della sua famiglia, per promuovere una crescita e un rafforzamento da un punto di vista psicologico.

Lo scopo della *Positive Psychology* è di indurre il soggetto a trovare nell'evento stressante qualcosa di positivo, utilizzando strategie di coping, come rivalutazione positiva degli eventi, strategie di problem-solving, la ricerca di un supporto sociale e di una espressione costruttiva delle emozioni [6] [7]. Ne

dovrebbe conseguire un miglioramento a livello di carattere, di relazioni, di gestione di una vita sociale, di un incremento delle prospettive.

Su questa scia è stato sviluppato un protocollo d'intervento per l'applicazione della *Well-Being Therapy (WBT)*, un aspetto della *Positive Psychology*, in ambito scolastico, articolato in una serie d'incontri di gruppo, supportato da metodologie laboratoriali ed esperienziali, come giochi, simulazioni e discussioni di gruppo, e avente come obiettivo quello di educare i ragazzi a migliorare l'espressione di emozioni e sentimenti e a favorire i processi relazionali e comunicativi.

## 1.1 Il metodo

Nel corso dell'anno scolastico 2003/2004 è stato effettuato uno studio su una popolazione scolastica costituita da studenti delle scuole medie inferiori. Sono state esaminate 6 classi, per un totale di 120 partecipanti, di 4 scuole medie di Padova. Oggetto dell'intervento in aula era un breve corso di educazione all'affettività da parte di due psicologhe cliniche (4 incontri per classe di 2 ore ciascuno). I primi due incontri, uguali in tutte le 6 classi, si sono incentrati in maniera precipua sul riconoscimento delle emozioni e sul miglioramento del clima emotivo nelle classi.

Nei successivi due incontri l'insieme delle 6 classi è stato suddiviso, in maniera casuale, in due gruppi, gruppo *A* e gruppo *B*. Ogni gruppo comprendeva quindi 3 classi di 60 soggetti ciascuna. Nel gruppo *A* si è cercato di migliorare il benessere e favorire le esperienze positive applicando il modello di benessere psicologico (*Well Being Therapy, WBT*), nel gruppo *B*, seguendo esclusivamente le teorie cognitive (*Cognitive Behaviour Therapy, CBT*) l'attenzione è stata rivolta alla risoluzione dei problemi e dei conflitti nei ragazzi. Prima e dopo gli incontri, i ragazzi hanno compilato due scale di autovalutazione del proprio stato psicologico: il *Kellner's Symptom Questionnaire (SQ)* [8] [9] e la forma breve per adolescenti del *Psychological Well-Being (PWB)* [10] [11]. Il *Kellner's Symptom Questionnaire (SQ)* è una scala di autovalutazione con 92 item che comprende 8 sottoscale, 4 di disagio psicologico (ansia, depressione, sintomi somatici, ostilità) con un punteggio che va da 0 a 17, per cui maggiore è il punteggio maggiore è il disagio, e 4 di benessere psicologico (rilassamento, contentezza, benessere fisico

e buona disponibilità verso gli altri) con un punteggio che va da 0 a 6, per cui maggiore è il punteggio minore è il benessere.

L' SQ è composto da item che indicano sintomi e item che indicano stati di benessere (well-being). Ad esempio la scala della depressione consiste di item come *sentimenti di disperazione, che non c'è più nulla da fare* e item come *felice*. Se il paziente risponde SI per *sentimenti di disperazione, che non c'è più nulla da fare*, viene conteggiato un punto per la scala della depressione; come pure se segna NO per *felice* viene conteggiato un punto. Di conseguenza per ogni soggetto vengono calcolati i sintomi per i quali risponde affermativamente e le espressioni di benessere per le quali risponde negativamente [12].

La forma breve del PWB è un questionario con 18 item per la valutazione del benessere psicologico validato per l'Italia dall'équipe del prof. Fava, che ha messo a punto anche la versione breve per adolescenti utilizzata in questo protocollo. Il questionario fornisce una valutazione di 6 aree del benessere (autonomia, padronanza dell'ambiente, crescita personale, relazioni positive, scopo nella vita e accettazione di sé). I soggetti dovevano fornire risposte SI/NO agli item indicati.

Il punteggio assegnato alle risposte va da 0 a 3, per cui maggiore è il punteggio, maggiore è il benessere. L'obiettivo dello studio è di scoprire se ci sono differenze significative tra i due gruppi e l'ipotesi che si vuole testare è se l'intervento focalizzato sul benessere, caratterizzato da attività più divertenti e piacevoli, è stato più efficace.

## 1.2 I limiti dello studio

Le limitazioni sono legate alla natura preliminare dello studio. Innanzitutto, il numero esiguo di incontri (4 incontri di 2 ore ciascuno) e il campione di soggetti prevalentemente sani non hanno di certo agevolato la rilevazione di differenze significative tra tempo 0 e tempo 1 (pre e post trattamento). Il numero stesso dei soggetti era ridotto perché con 92 item il campione dovrebbe essere costituito da circa 900 unità.

I partecipanti, poi, non sono stati sottoposti a delle terapie standard (WBT, CBT), ma ad un protocollo d'intervento assolutamente preliminare, soprattutto nel caso del trattamento WBT, adattato in un contesto scolasti-

co, avente in comune con la terapia classica completa solo alcuni degli aspetti salienti.

Il protocollo finale della WBT infantile è infatti suscettibile a variazioni che richiedono una serie di prove di fattibilità dei singoli steps del protocollo anche a livello individuale e su una popolazione clinica, i risultati delle quali potranno essere utili per perfezionare il protocollo della WBT infantile nel trattamento dei disturbi più frequenti nell'infanzia/adolescenza. Da non sottovalutare l'autoselezione delle scuole che non sono state randomizzate per ricevere il trattamento, ma si sono spontaneamente proposte per essere incluse nello studio (hanno aderito spontaneamente allo studio). Il piano sperimentale non era ben progettato: i risultati finali infatti potrebbero essere condizionati da alcuni fattori quali la mancanza di omogeneità nella composizione delle classi esaminate, sia per quanto riguarda il sesso sia per quanto riguarda l'età degli studenti, che potrebbe mascherare un effetto *leader*, la tendenza negli adolescenti alla *desiderabilità sociale*, il fattore *dislocazione delle scuole*, se queste si trovassero in zone residenziali o più in periferia, e legato a questo il fattore *condizione sociale* dei partecipanti. Altre caratteristiche dello studio che hanno costituito una limitazione alla sua piena riuscita sono la mancanza di un follow-up in grado di rilevare gli effetti del protocollo WBT a lungo termine e il fatto che entrambi gli interventi condividono tecniche e aspetti in comune, la WBT è infatti una tecnica specifica all'interno dell'ampio insieme delle strategie cognitivo-comportamentali [13].

Tra tutti questi aspetti, a rendere un po' più complicata l'analisi dei dati è stata anche la presenza di *missing values* e questo lavoro di tesi affronterà il problema seguendo due diversi approcci, tramite l'applicazione del principio dell'*Intention To Treat (ITT) analysis* e utilizzando i metodi non parametrici di permutazione.

# Capitolo 2

## I metodi di combinazione non parametrici

### 2.1 La teoria

Caratteristica dei test di permutazione è il condizionamento all'insieme dei dati osservati che è un insieme di statistiche sufficienti qualunque sia il modello sottostante di riferimento. E' questo il motivo per cui i test di permutazioni vengono chiamati *distribution free*, ossia che le distribuzioni dei test prescindono completamente dalla legge che governa la variabile aleatoria su cui si vuol fare inferenza e che non è necessario fare assunzioni stringenti sulla distribuzione dei termini d'errore [14].

Le assunzioni principali riguardanti la struttura dei dati, l'insieme dei test parziali e le ipotesi d'interesse per i test, nel contesto della combinazione non parametrica, possono essere così schematizzate:

1. L'insieme  $q$ -dimensionale dei dati, o il vettore  $q$ -variato delle risposte è indicato con

$$\begin{aligned} X &= \{X_j, j = 1, \dots, C\} = \{X_{ji}, i = 1, \dots, n_j, j = 1, \dots, C\} \\ &= \{X_{hji}, i = 1, \dots, n_j, j = 1, \dots, C, h = 1, \dots, q\} \end{aligned}$$

Il vettore  $q$ -variato delle risposte  $X$  è definito dal modello statistico  $(X, \mathcal{X}, \mathcal{B}, P)$ , dove  $\mathcal{X}$  è lo spazio campionario da cui prende i suoi valori,  $\mathcal{B}$  è una  $\sigma$ -algebra e  $P$  è una distribuzione di probabilità di solito non specificata proveniente da una famiglia  $\mathcal{P}$  di distribuzioni non degeneri. L'insieme dei dati  $X$  è costituito da  $C \geq 2$  campioni o gruppi

di ampiezza  $n_j \geq 2$ , con  $n = \sum_j n_j$ ; i gruppi sono rappresentativi di  $C$  livelli di un trattamento e i dati  $X_j$  sono supposti *i.i.d.* con distribuzione  $P_j \in \mathcal{P}$ ,  $j = 1, \dots, C$ , secondo un modello MANOVA. Se sono a disposizione delle covariate, è semplice il passaggio ad un modello MANCOVA. Per coerenza con una rappresentazione unità per unità, l'insieme  $X$  si può riscrivere con  $X = \{X(i), i = 1, \dots, n; n_1, \dots, n_C\}$ , in cui si assume che i primi  $n_1$  vettori di dati provengano dal primo campione, i successivi  $n_2$  dal secondo e così via.

2. Sotto l'ipotesi nulla le distribuzioni multivariate delle risposte sono uguali nei  $C$  gruppi e i dati sono scambiabili tra i  $C$  campioni. Il sistema d'ipotesi è:

$$\begin{aligned} H_0 &= \{P_1 = \dots = P_c\} = \left\{X_1 \stackrel{d}{=} \dots \stackrel{d}{=} X_c\right\} = \bigcap_{1 \leq i \leq k} H_{0i} \quad (2.1) \\ H_1 &= \{\text{almeno una } H_{0i} \text{ è falsa}\} \end{aligned}$$

Supponiamo che si vogliano analizzare e verificare più aspetti di un problema, così che l'ipotesi nulla  $H_0$  può essere scomposta in un insieme finito di sottoipotesi  $H_{0i}$ ,  $i = 1, \dots, k$ , ciascuna adatta all'aspetto parziale di interesse.  $H_0$  è vera se tutte le  $H_{0i}$  sono congiuntamente vere. Di qui si può definire  $H_0$  come ipotesi nulla globale, tale che  $H_0 = \left\{ \bigcap_{i=1}^k H_{0i} \right\}$ . Sempre considerando la scomposizione in  $k$  sottoipotesi, l'ipotesi alternativa globale  $H_1$  è vera se almeno una delle sottoipotesi  $H_{0i}$  è falsa, quindi  $H_1 = \left\{ \bigcup_{i=1}^k H_{1i} \right\}$

3. con  $T = T(X)$  si indica il vettore  $k$ -dimensionale di statistiche test, le cui componenti rappresentano i test univariati non degeneri di primo ordine ciascuno idoneo alla verifica delle sub-ipotesi  $H_{0i}$  vs  $H_{1i}$ .

Le assunzioni riguardanti l'insieme di test parziali  $T = \{T_i, i = 1, \dots, k\}$  sufficienti per la combinazione non parametrica sono:

- (A.1) Tutti i test parziali  $T_i$  sono permutazionalmente esatti, ovvero le variazioni delle statistiche sotto  $H_0$  dipendono solo dalla casualità delle permutazioni, e sono marginalmente corretti, sono stocasticamente significativi per valori grandi, vale a dire che sotto  $H_1$  le loro distribuzioni sono stocasticamente più grandi rispetto ad  $H_0$

(A.2) I test di permutazione  $T_i$  sono consistenti, cioè:

$$\Pr \{T_i > T_{i\alpha} | H_{1i}\} \geq \alpha, \forall \alpha > 0, i = 1, \dots, k$$

e al tendere di  $n$  all'infinito si ha che

$$\lim_{n \rightarrow \infty} \Pr \{T_i > T_{i\alpha} | H_{1i}\} \rightarrow 1$$

dove  $n$  è la numerosità campionaria e  $T_{i\alpha}$  è il valore critico, assunto finito, di  $T_i$  al livello  $\alpha$ .

## 2.2 Funzioni di combinazione non parametrica

La combinazione non parametrica prende in considerazione i  $p$ -value  $\lambda_i$  di permutazione associati alle statistiche test  $T_i$ ,  $i = 1, \dots, k$ .

Il test combinato di secondo ordine  $T'' = \psi(\lambda_1, \dots, \lambda_k)$  si ottiene tramite una funzione  $\psi \div (0, 1)^k \rightarrow \mathcal{R}^1$  continua, non crescente, univariata, non degenera, reale (misurabile) che soddisfi alle seguenti proprietà:

- i)  $\psi$  deve essere non crescente in ogni suo argomento, ovvero  $\psi(\dots, \lambda_i, \dots) \geq \psi(\dots, \lambda'_i, \dots)$  se  $\lambda_i < \lambda'_i$ ,  $i \in (1, \dots, k)$
- ii)  $\psi$  assume il suo valore massimo  $\bar{\psi}$ , che potrebbe non essere finito, quando almeno uno degli argomenti raggiunge lo 0, cioè  $\psi(\dots, \lambda_i, \dots) \rightarrow \bar{\psi}$  se  $\lambda_i \rightarrow 0$ ,  $1 \leq i \leq k$
- iii) il suo valore critico  $T''_\alpha$  è finito e tale che  $T''_\alpha \leq \bar{\psi}$ ,  $\forall \alpha > 0$

Queste proprietà definiscono una classe  $\mathcal{C}$  di funzioni di combinazione. Le funzioni di combinazione  $\psi$  più utilizzate sono:

- a) La funzione di combinazione di *Fisher*

$$T''_F = -2 \cdot \sum_i \log(\lambda_i) \quad (2.2)$$

Se tutti i  $k$  test parziali sono indipendenti e continui, sotto l'ipotesi nulla  $T''_F$  si distribuisce come un  $\chi^2$  centrale con  $2k$  gradi di libertà

b) La funzione di combinazione di *Liptak*

$$T_L'' = \sum_i \Phi^{-1}(1 - \lambda_i) \quad (2.3)$$

dove  $\Phi$  è la funzione di ripartizione di una normale standard. Se tutti i  $k$  test parziali sono indipendenti e continui, sotto l'ipotesi nulla  $T_L''$  ha distribuzione normale, con media 0 e varianza  $k$  [15]. Una versione della funzione di combinazione di Liptak considera la trasformazione logistica dei *p-values*:

$$T_P'' = \sum_i \log[(1 - \lambda_i)/\lambda_i] \quad (2.4)$$

c) La funzione di combinazione di *Tippett*

$$T_T'' = \max_{1 \leq i \leq k} (1 - \lambda_i) \quad (2.5)$$

la cui distribuzione sotto l'ipotesi nulla, se tutti i  $k$  test parziali sono continui e indipendenti, si comporta come il più grande di  $k$  valori casuali estratti da una variabile uniforme nell'intervallo aperto  $(0, 1)$

d) La funzione di combinazione di *Lankaster*

$$T_G'' = \sum_i \Gamma_{r,a}^{-1}(1 - \lambda_i) \quad (2.6)$$

dove  $\Gamma_{r,a}^{-1}$  è l'inversa della funzione di ripartizione di una variabile casuale gamma centrale con parametro di scala noto  $a$  e  $r$  gradi di libertà. Se tutti i  $k$  test parziali sono indipendenti, sotto l'ipotesi nulla,  $T_G''$  ha distribuzione gamma centrale con parametro di scala  $a$  e  $rk$  gradi di libertà.

e) Un' interessante sottoclasse di  $\mathcal{C}$  è costituita dall'insieme di funzioni  $\mathcal{C}_D$  di combinazione non parametrica *diretta*, espressa nella forma

$$T_D'' = \sum_i T_i \quad (2.7)$$

che si può utilizzare quando tutte le statistiche test parziali sono omogenee, per cui condividono la stessa distribuzione di permutazione asintotica (per esempio, si distribuiscono come normali standard o sono del



tipo  $\chi^2$  con gli stessi gradi di libertà) e il loro supporto comune asintotico è almeno non limitato a destra. Le distribuzioni di permutazione di tutti i test parziali possono essere le stesse solo per numerosità campionarie abbastanza elevate, quindi per numerosità campionarie finite, questa condizione può essere soddisfatta solo approssimativamente poiché le distribuzioni di permutazione sono essenzialmente dipendenti dai dati osservati. La funzione di combinazione diretta consente di evitare i calcoli abbastanza intensivi dell'algoritmo ed inoltre, anche se la funzione diretta opera come nel caso univariato, è sostanzialmente una combinazione non parametrica perché il problema di verifica viene spezzato in  $k$  sotto problemi, una statistica test globale  $T \div \mathcal{R}^k \rightarrow \mathcal{R}^1$  non è disponibile e le relazioni di dipendenza tra i test parziali sono implicitamente catturate dalla procedura di combinazione.

### **2.3 Tecniche di ricampionamento condizionato (C.M.C.)**

Vi sono due criteri per permutare i dati, uno riguarda la permutazione sistematica di tutti i dati, l'altro prende in considerazione solo un campione estratto casualmente dallo spazio di permutazione. Pur assolvendo alla stessa funzione, l'utilizzo del secondo metodo si traduce in un vantaggio in termini di riduzione dei calcoli, vantaggio di notevole importanza soprattutto quando il numero di permutazioni è tanto elevato da richiedere tempi di elaborazione molto lunghi anche per calcolatori potenti. E' stato dimostrato che il fatto di ridurre il numero di permutazioni non porta ad una minor attendibilità del risultato o ad una perdita di potenza del test, anzi, per quanto riguarda quest'ultimo aspetto, ossia quello relativo alla potenza, questa è sempre vicina a quella del test non parametrico più potente. Una soluzione approssimata, legata al secondo, è basata sul metodo di simulazione di *Monte Carlo Condizionato (CMC)* dall'orbita di permutazione, tramite il quale si può condurre una tecnica di ricampionamento condizionato all'insieme dei dati osservati. Il campionamento Monte Carlo condizionato altro non è se non la replicazione dei campionamenti senza reinserimento (Without Replacement Resampling). La stima della distribuzione di permutazione sarà tanto più

precisa quanto maggiore è il numero di iterazioni (di solito una buona stima si ottiene con 10000 replicazioni). La procedura che consente di ottenere una stima, tramite CMC, della distribuzione di permutazione dei test combinati consta di due fasi: la prima relativa alla stima della distribuzione  $k$ -variata di  $T$ , la seconda riguarda la stima della distribuzione di permutazione del test combinato  $T''_{\psi}$  e utilizza i risultati ottenuti via CMC nella prima fase.

In ambito multivariato dunque, l'algoritmo di simulazione per la stima della distribuzione  $k$ -variata di  $T$  può essere descritto nelle seguenti fasi:

- a.1) si calcola il valore osservato di  $T : T_0 = T(X)$
- b.1) si considera un componente  $g^*$ , casualmente rilevato da un appropriato gruppo di trasformazioni  $G$ , e i valori del vettore  $T^* = T(X^*)$  dove  $X^* = g^*(X)$ . Una permutazione  $X^*$  del file di dati si può anche ottenere considerando una permutazione  $(u_1^*, u_2^*, \dots, u_n^*)$  delle etichette  $(1, 2, \dots, n)$ , ciascuna indicante una unità del campione di partenza, e assegnando poi il vettore di dati con le etichette permutate al gruppo appropriato, per esempio nel caso di due campioni di numerosità  $n_1$  e  $n_2$  tali che  $n_1 + n_2 = n$  si avrà che  $X(u_i^*)$  viene assegnato al primo gruppo se l'etichetta  $i$  soddisfa la condizione  $1 \leq i \leq n_1$ , altrimenti al secondo gruppo
- c.1) si ripete  $B$  volte la fase descritta in b.1); l'insieme dei risultati del CMC  $\{T_r^*, r = 1, \dots, B\}$  è un campione casuale dalla distribuzione nulla  $k$ -variata di  $T$
- d.1) la funzione di distribuzione empirica simulata  $k$ -variata (*EDF*)

$$\widehat{F}_B(z|X) = \left[ \frac{1}{2} + \sum_r I(T_r^* \leq z) \right] / (B + 1) \quad \forall z \in \mathcal{R}^k$$

con  $I(\cdot)$  che vale 1 se la relazione è soddisfatta 0 altrimenti, fornisce una stima della distribuzione di permutazione  $k$ -dimensionale  $F(z|X)$  di  $T$ . Si ha poi che

$$\widehat{L}_i(z|X) = \left[ \frac{1}{2} + \sum_r I(T_{ir}^* \geq z) \right] / (B + 1), \quad i = 1, \dots, k$$

dà una stima  $\forall z \in \mathcal{R}^1$  della funzione di permutazione marginale del livello di significatività  $L_i(z|X) = \Pr(T_i^* \geq z|X)$ ; così che  $\widehat{L}_i(T_{i0}|X) = \widehat{\lambda}_i$  fornisce una stima non distorta e consistente del vero  $p$ -value marginale  $\lambda_i = \Pr(T_i^* \geq T_{i0}|X)$  relativa al test  $T_i$ ,  $\forall i = 1, \dots, k$ . E' da notare che rispetto agli stimatori EDF standard,  $1/2$  e  $1$  sono rispettivamente aggiunti al numeratore e al denominatore allo scopo di ottenere un valore stimato nell'intervallo aperto  $(0, 1)$  in modo che le trasformazioni inverse di funzioni continue, come  $-\log(\lambda)$  o  $\Phi^{-1}(1 - \lambda)$ , siano continue e sempre definite. Dato che  $B$  è molto grande queste quantità aggiunte sono alterazioni praticamente irrilevanti che non influenzano il comportamento degli stimatori sia per campioni di numerosità finita che asintoticamente. Al posto di  $1/2$  e  $1$  si può dunque mettere una qualsiasi quantità positiva  $\varepsilon$  e  $2\varepsilon$ , assegnando ad  $\varepsilon$  un valore di gran lunga inferiore rispetto al numero di ricampionamenti  $B$ .

La combinazione non parametrica segue le seguenti fasi:

- a.2) i  $k$   $p$ -value  $\widehat{\lambda}_i = \widehat{L}_i(T_{i0}|X)$  osservati sono stimati sui dati  $X$  dove  $T_{i0} = T_i(X)$ ,  $i = 1, \dots, k$  rappresenta i valori osservati dei test parziali e  $\widehat{L}_i$  è la  $i$ -esima funzione marginale del livello di significatività ottenuta tramite ricampionamento CMC sull'insieme dei dati nell'ultimo step dell'algoritmo della prima fase (vedi d.1) )
- b.2) il valore osservato combinato dei test di secondo ordine utilizza i risultati ottenuti nella prima fase con CMC ed è dato da  $T_0'' = \psi(\widehat{\lambda}_1, \dots, \widehat{\lambda}_k)$
- c.2) l'  $r$ -esimo vettore statistico simulato combinato è calcolato partendo da  $T_r''^* = \psi(\lambda_{1r}^*, \dots, \lambda_{kr}^*)$ , dove  $\lambda_{ir}^* = \widehat{L}_i(T_{ir}^*|X)$ ,  $i = 1, \dots, k$ ,  $r = 1, \dots, B$
- d.2) il  $p$ -value simulato globale è dato da  $\widehat{\lambda}_\psi'' = \sum_r I(T_r''^* \geq T_0'')/B$
- e.2) si rigetta l'ipotesi nulla globale a livello di significatività fissato e pari ad  $\alpha$  se  $\widehat{\lambda}_\psi'' \leq \alpha$ .



## Capitolo 3

# Problemi con dati appaiati: soluzione di permutazione

### 3.1 Il disegno sperimentale

Il più semplice tra i disegni sperimentali è quello basato sul confronto prima-dopo in un unico gruppo di soggetti, quindi senza gruppo di controllo. Varie sono le motivazioni che rendono attraente tale approccio: il fatto che il gruppo sottoposto al trattamento di controllo e quello sottoposto al trattamento sperimentale siano identici o quasi, se si considera l'eventualità che un soggetto esca prematuramente dallo studio, porta a credere che l'effetto della variabilità biologica sia minimizzato e che la possibilità di evidenziare la risposta al trattamento sia massimizzata. Inoltre, a parità di altre condizioni, la dimensione del campione è più piccola rispetto a quella richiesta dai disegni sperimentali alternativi, vale a dire che saranno necessari meno pazienti per raggiungere lo stesso obiettivo. Ma lo svantaggio più rilevante è che un disegno sperimentale di questo tipo non previene in alcun modo la distorsione. La distorsione o *bias* nella ricerca clinica, diversamente da un errore casuale, è un errore che sistematicamente favorisce un trattamento rispetto ad un altro. Se, da un lato, la presenza di un errore casuale comporta un aumento della variabilità della risposta e rende più complesso il riconoscimento di una reale efficacia del trattamento, dall'altro è possibile ridurlo, almeno in parte, gli effetti indesiderati operando sulla numerosità campionaria, ovvero scegliendo un campione sufficientemente numeroso da consentire la separazione dell'effetto del trattamento dalla variabilità complessiva con un grado di certezza,

o di rischio, ritenuto accettabile. Controllare e contenere gli effetti di un errore sistematico è piuttosto difficile, anche se si aumenta la numerosità del campione o si utilizzano tecniche di analisi statistica più raffinate che, tramite stime aggiustate, depurino i risultati dall'effetto del bias. L'unico modo per ridurre il *bias* è quello di scegliere, in sede di pianificazione dell'esperimento, un adeguato disegno sperimentale. Numerose sono le fonti di distorsione in un confronto prima-dopo tra cui le variazioni temporali della malattia, le variazioni temporali di personale, equipaggiamento e contesto di uno studio, la regressione statistica verso la media, fenomeno per cui una variabile che assume un valore estremo (molto alto o molto basso) nella prima misurazione tenderà ad essere più vicina alla media della popolazione nelle misurazioni successive, l'effetto apprendimento e l'effetto psicologico. Nonostante queste limitazioni, anche per ragioni di carattere etico, i disegni prima e dopo senza gruppo di controllo sono molto utilizzati in biostatistica. Quasi sempre nel confronto prima-dopo in un solo gruppo di soggetti è il dopo ad essere favorito con il risultato che gran parte dei trattamenti sperimentali con disegni di questo tipo tendono ad apparire efficaci per poi spesso fallire quando messi alla prova in disegni sperimentali meno triviali capaci di offrire maggior protezione contro la distorsione. E' forse l'elevata probabilità di un evento positivo la ragione per cui il disegno prima-dopo continua ad essere così popolare tra i ricercatori [16].

### 3.2 Caso univariato

Nei problemi con dati appaiati (*paired observations*) su ciascun individuo partecipante allo studio la variabile d'interesse viene rilevata prima e dopo la somministrazione di un trattamento, una terapia o un farmaco. Si hanno dunque due situazioni sperimentali: la situazione *A*, che corrisponde alla rilevazione pre-trattamento, chiamata anche *baseline observation*, e la situazione *B*, ovvero la rilevazione post-trattamento che viene effettuata dopo un periodo di tempo stabilito sulle unità trattate al fine di testare se il trattamento in questione ha indotto un effetto statisticamente significativo, un aumento o una riduzione, un miglioramento o un peggioramento rispetto alle condizioni iniziali.

Le risposte bivariate sono dipendenti con rispetto alle unità, dato che le

misurazioni vengono fatte in tempi diversi negli stessi soggetti, mentre le  $n$  coppie di osservazioni sono indipendenti, in quanto relative ad unità diverse. Se si assume che gli individui siano omogenei in relazione alle condizioni sperimentali, l'insieme dei dati appaiati  $\{(Y_{Ai}, Y_{Bi}), i = 1, \dots, n\}$  può essere visto come un campione casuale di  $n$  coppie *i.i.d.* di osservazioni estratte da una variabile bivariata  $(Y_A, Y_B)$ . A livello formale la rappresentazione dei dati è di tipo matriciale, le osservazioni sono raccolte in una matrice di  $n$  coppie  $(Y_A, Y_B) \in \mathcal{X}$  dove  $\mathcal{X}$  è lo spazio campionario dell'esperimento. Se, per esempio ci si aspetta che il trattamento riduca stocasticamente una certa variabile, il sistema d'ipotesi può essere così formulato:

$$\begin{aligned} H_0 & : \left\{ Y_A \stackrel{d}{=} Y_B \right\} = \left\{ P_A(z) = P_B(z), \forall z \in \mathcal{R}^1 \right\} \\ H_1 & : \left\{ Y_A \stackrel{d}{>} Y_B \right\} \end{aligned} \quad (3.1)$$

in cui  $P_A$  e  $P_B$  sono le distribuzioni marginali di  $Y_A$  e  $Y_B$ ,  $H_0$  e  $H_1$  sono rispettivamente l'ipotesi nulla e l'ipotesi alternativa. Nel modello sopra descritto sotto  $H_1$  vi è *dominanza stocastica* di  $Y_A$  rispetto a  $Y_B$ , ma a seconda della specificazione del problema varia anche la direzione dell'ipotesi alternativa. Tra i modelli più utilizzati per descrivere la variabile risposta osservata sono da citare:

i) i modelli con *effetti additivi fissi*

$$Y_{Ai} = \mu + Z_{Ai}, \quad Y_{Bi} = \mu - \delta + Z_{Bi}, \quad i = 1, \dots, n \quad (3.2)$$

in cui  $\mu$  è la costante relativa alla popolazione;  $\delta$  è l'effetto del trattamento, assunto finito e strettamente positivo sotto l'ipotesi nulla  $H_0$ ,  $Z_{Ai}$  e  $Z_{Bi}$  sono deviazioni casuali identicamente distribuite e centrate, chiamate anche termini o componenti d'errore non necessariamente indipendenti entro le unità, ma indipendenti rispetto alle unità.

ii) i modelli con *effetti additivi fissi e unità non omogenee*

$$Y_{Ai} = \mu + \eta_i + Z_{Ai}, \quad Y_{Bi} = \mu + \eta_i - \delta + Z_{Bi}, \quad i = 1, \dots, n \quad (3.3)$$

in cui le  $\eta_i$  sono quantità ignote relative e specifiche dell' $i$ -esimo individuo che non dipendono dal livello del trattamento; per quanto riguarda tutte le altre componenti del modello queste assumono lo stesso significato di quello che hanno nel modello descritto in (3.2).

iii) i modelli con *effetti additivi che variano da individuo a individuo*

$$Y_{Ai} = \mu + \eta_i + Z_{Ai}, \quad Y_{Bi} = \mu + \eta_i - \delta_i + Z_{Bi}, \quad i = 1, \dots, n \quad (3.4)$$

in cui  $\delta_i$  sono effetti del trattamento specifici dell' $i$ -esima unità e, sotto  $H_1$ , sono quantità finite non negative, almeno una delle quali è positiva.

iv) i modelli con *effetti stocastici generalizzati*

$$Y_{Ai} = \mu + \eta_i + Z_{Ai}, \quad Y_{Bi} = \mu + \eta_i + Z_{Bi} - \Delta_{Bi}, \quad i = 1, \dots, n \quad (3.5)$$

dove, sotto  $H_1$ , gli effetti casuali  $\Delta_{Bi}$  che possono dipendere in qualche maniera da  $(\mu, \eta_i, Z_{Ai}, Z_{Bi})$ , e sotto l'ipotesi alternativa sono quantità stocastiche non negative, almeno una delle quali è strettamente positiva.

L'ultimo modello risulta essere consistente per le variabili quantitative che presentano una qualche forma di dominanza stocastica, in particolare con le forme di dominanza di tipo moltiplicativo fisso o stocastico. Prendendo come modello di riferimento il modello con effetti additivi fissi (3.2), l'ipotesi nulla (3.1) può essere anche formulata come

$$H_0 : \{ \Pr(Y_A - Y_B \leq -z) = \Pr(Y_A - Y_B \geq z), \quad \forall z \in \mathcal{R}^1 \} \quad (3.6)$$

così che  $H_0$  è vera soltanto se la variabile differenza  $X = Y_A - Y_B = \delta + Z_A - Z_B$  è distribuita in maniera simmetrica attorno allo 0 mentre  $H_1$  è tale che  $\Pr\{X > 0 | H_1\} > 1/2$  e  $X$  è distribuita simmetricamente intorno al parametro di posizione  $\delta > 0$ , indicatore dell'effetto del trattamento. Quando si usa come variabile di riferimento la variabile differenza  $X$  il modello (3.2) e il modello (3.3) coincidono, infatti si ha che  $X_i = Y_{Ai} - Y_{Bi} = \delta + Z_{Ai} - Z_{Bi}$ ,  $i = 1, \dots, n$  in entrambe le formulazioni. Dunque se non vi è un reale effetto del trattamento ed eventuali variazioni osservate sono apportate solo dalla componente  $\eta_i$ , ovvero quando le covariate agiscono solo sul parametro specifico dell'individuo,  $X$  diventa una variabile *covariate-free*, consentendo così di superare l'ostacolo legato all'omogeneità delle unità, rispetto alle condizioni sperimentali, che caratterizza il modello (3.3). La simmetria di cui si è accennato sopra è dovuta alla scambiabilità delle componenti d'errore  $Z_A$  e  $Z_B$  entro le unità.



Posto  $Z_A = V$ ,  $Z_B = W$  e  $Z = Z_A - Z_B = V - W$ , le variabili  $V$  e  $W$  sono scambiabili entro le unità e questo implica che

$$F_V(z) = F_W(z), \quad \forall z \in \mathcal{R}^1 \quad (3.7)$$

e

$$F_{V|t}(z|W = t) = F_{W|t}(z|V = t), \quad \forall (t, z) \in \mathcal{R}^2 \quad (3.8)$$

in cui  $F_V$ ,  $F_W$ ,  $F_{V|t}$  e  $F_{W|t}$  sono le funzioni di ripartizione (*c.d.f*) delle variabili  $V$ ,  $W$ ,  $(V|W = t)$ ,  $(W|V = t)$ , funzioni che sono legate a  $P$  e sono ben definite. Quindi si può scrivere

$$\begin{aligned} F_{V-W}(z) &= \int_{-\infty}^{+\infty} F_{V|t}(z+t|W=t) \cdot dF_W(t) \\ F_{W-V}(z) &= \int_{-\infty}^{+\infty} F_{W|t}(z+t|V=t) \cdot dF_V(t) \end{aligned} \quad (3.9)$$

così si ha che  $\Pr\{V - W > z\} = \Pr\{V - W < -z\}$ ,  $\forall z \in \mathcal{R}^1$ , che è la condizione di simmetria di  $Z$  intorno all'origine 0. Come conseguenza della proprietà di simmetria si ha, sotto  $H_0$ , che:

- \*  $\Pr\{X < 0\} = \Pr\{X > 0\}$
- \*  $X$  ha media nulla se il valore atteso  $\mathbb{E}(Z) = 0$
- \*  $X$  ha mediana nulla se è  $\text{Med}(Z) = 1$
- \* Il vettore dei segni  $\{X_i/|X_i|, i = 1, \dots, n\}$  è stocasticamente indipendente da quello delle differenze  $\{X_i = Y_{Ai} - Y_{Bi}, i = 1, \dots, n\}$  (da sottolineare che se le  $X_i = 0$  allora le  $X_i$  e i relativi segni  $X_i/|X_i|$  vengono esclusi dall'analisi),

mentre sotto  $H_1$  si ha

- \*  $\Pr\{X < 0\} > (<) \Pr\{X > 0\}$ , a seconda che  $Y_A < (>) Y_B$ .

La distribuzione  $P_X$  (o  $P$  soltanto) di  $X$  è assunta ignota (nei parametri, nella forma analitica o del tutto) ed è tale che  $P \in \mathcal{P}$  ovvero appartiene ad

una famiglia di distribuzioni non parametriche e non degeneri. Assumendo che  $\mathbb{E}(Z)$  sia finito così che  $\mathbb{E}(X)$  può ben rappresentare l'effetto del trattamento  $\delta$ , un altro modo per scrivere il sistema di ipotesi è:

$$H_0 : \{\delta = 0\} \text{ vs } H_1 : \{\delta > 0\}$$

Il vettore delle differenze  $\{X_i = Y_{Ai} - Y_{Bi}, i = 1, \dots, n\}$  può essere considerato come un campione casuale di  $n$  osservazioni scambiabili e se i dati sono *i.i.d.* tutte le unità sono caratterizzate dalla stessa distribuzione. Nel caso del modello (3.4), ciascuna unità ha distribuzione propria  $P_i, i = 1, \dots, n$  e in quest'ultimo caso la soluzione può essere ottenuta solo via non-parametrica perché aumenta di molto il numero di parametri di disturbo che devono essere eliminati. Nel caso parametrico è necessario che  $\mathbb{E}(X)$  esista finito, mentre nel caso non parametrico è necessaria l'esistenza della mediana o comunque di un parametro di posizione robusto. E' spesso molto utile considerare delle trasformazioni dei dati indotte da una funzione  $\varphi$ , come  $\varphi(X), \varphi(Y) - \varphi(W), \varphi(Y, W)$ . In ambito non parametrico, il problema relativo alla scelta di una funzione di trasformazione che consenta di incrementare la potenza del test e ottenere il test migliore è ancora una questione irrisolta.

### 3.3 Soluzione parametrica: t di Student

#### 3.3.1 Dati appaiati normalmente distribuiti

Una soluzione al problema dei dati appaiati può essere ottenuta in un contesto parametrico se si assume che le variabili siano normalmente distribuite e abbiano varianza ignota. Il modello con effetti additivi fissi può essere scritto come

$$\{Y_{Ai} = \mu + \sigma \cdot Z_{Ai}, Y_{Bi} = \mu - \sigma \cdot \delta + Z_{Bi}, i = 1, \dots, n\} \quad (3.10)$$

in cui  $\mu$  è la costante di popolazione,  $\delta$  è l'effetto del trattamento,  $\sigma$  la deviazione standard, ignota, indipendente dalle unità e dal livello del trattamento e tale che  $0 < \sigma < +\infty$ ,  $Z_{ij}$ , quantità stocastiche d'errore, hanno distribuzione normale standard, ovvero  $Z_{ij} \sim \mathcal{N}(0, 1)$  con  $i = 1, \dots, n, j = A, B$  indipendenti tra le unità ma non necessariamente entro le unità.

La statistica test più comunemente usata è la statistica test  $t$  di Student per osservazioni appaiate che ha la forma,  $T = \bar{X} \cdot \sqrt{n}/\hat{\sigma}$  in cui  $\hat{\sigma}^2 = \sum_i (X_i - \bar{X})^2 / (n - 1)$  e  $\bar{X} = \sum_i X_{ij} / n$  con le  $X_i \sim \mathcal{N}(\delta, \sigma_X^2)$ . Sotto  $H_0$  la statistica  $T$  ha distribuzione  $t$  di Student centrale con  $(n - 1)$  *g.d.l.*, mentre sotto l'alternativa  $H_1$  è distribuita come una  $t$  di Student non centrale con un parametro di non centralità positivo così che valori grandi diventano significativi. Il parametro ignoto  $\sigma_X$  è solo un parametro di disturbo e  $T$  è una statistica invariante rispetto al valore assunto da questa quantità,  $\hat{\sigma}$  è una statistica sufficiente minimale completa per  $\sigma_X$  sotto  $H_0$  e sotto  $H_1$ .

### L'analisi della covarianza (ANCOVA)

Sempre rimanendo nell'ambito parametrico, il problema per campioni con dati appaiati può essere risolto tramite l'analisi della covarianza (ANCOVA) a condizione che le  $(Y_A, Y_B)$  siano indipendenti e identicamente distribuite con distribuzione normale bivariata e se  $Y_A$  è la variabile covariata del problema.

Per inciso, data  $Y$  la variabile dipendente (nel nostro caso corrisponde alla variabile rilevata dopo il trattamento  $Y_B$ ) e  $X$  la variabile indipendente covariata (nel nostro caso è rappresentata dalla misurazione iniziale  $Y_A$ ) gli assunti necessari per la validità delle procedure standard per le verifiche d'ipotesi previste dal problema dell'ANCOVA parametrica sono:

- La regressione lineare di  $Y$  su  $X$
- La distribuzione di  $Y|X = x$  deve essere normale per ogni popolazione considerata e con matrice di varianze e covarianze  $\Sigma$  indipendente da  $X$
- Le risposte iniziali dipendono solo dalle covariate e non anche dai soggetti

Quando si affrontano problemi inferenziali, nel formulare ipotesi circa il modello statistico sottostante la popolazione da cui provengono i dati e nel verificare, tramite opportuni test statistici, tali ipotesi, è fondamentale la specificazione degli assunti e restrizioni che garantiscono la validità del test. La violazione degli assunti che consentono l'applicabilità del test comporta una perdita di robustezza più o meno accentuata a seconda degli assunti violati. Le assunzioni alla base della procedura ANCOVA sono

1. Indipendenza dei dati
2. Normalità
3. Omogeneità della varianza
4. Omogeneità della regressione
5. Assenza di multicollinearità
6. Affidabilità dei dati sulle covariate

### 3.4 Soluzione non parametrica: test di permutazione

Le soluzioni di permutazione sono condizionali all'intero insieme dei dati osservati che, sotto l'ipotesi nulla  $H_0$ , è statistica sufficiente per ogni distribuzione sottostante non degenere  $P$ . Assumiamo che  $P$  sia ignota e che la famiglia non parametrica di probabilità  $\mathcal{P}$ , da cui proviene  $P$ , sia costituito solo da distribuzioni non degeneri. Da notare che, a causa di questo condizionamento ai dati osservati e dell'indipendenza delle  $n$  unità, la distribuzione multivariata  $P$  può essere ottenuta come  $\prod_i P_i$ , in cui  $P_i$  è la distribuzione relativa all' $i$ -esima unità. Questo implica che si possono formulare dei modelli del tipo (3.4) in cui gli errori  $(Z_{Ai}, Z_{Bi})$  sono non-invarianti rispetto alle unità e il fatto di condizionarsi a statistiche sufficienti fa venir meno la necessità di supporre che le unità abbiano identica distribuzione. Mentre, come abbiamo visto, nel caso parametrico  $\mathbb{E}(X)$  deve esistere finito, nell'ambito delle soluzioni non parametriche questa assunzione non è più necessaria e viene soltanto richiesta l'esistenza di un parametro  $\delta$  che sia un buon indicatore degli effetti del trattamento. L'ipotesi nulla  $H_0 : \left\{ Y_A \stackrel{d}{=} Y_B \right\}$  implica che ci sia scambiabilità delle variabili  $Y_A$  e  $Y_B$  entro ciascuna unità rispetto ai due tempi di rilevazione  $A$  e  $B$ . Il segno di ciascuna differenza  $X_i$ ,  $i = 1, \dots, n$  si può pensare sia attribuito con probabilità  $1/2$ . Una statistica test può essere data da  $T = \sum_i X_i$ . Se è vera l'ipotesi nulla e se i dati sono scambiabili è possibile applicare i metodi di permutazione e la distribuzione condizionata  $F_T(t|\mathbf{X})$  di  $T$ , in cui  $\mathbf{X} = \{X_i, i = 1, \dots, n\}$  è l'insieme dei punti osservati, si può derivare dalla distribuzione di  $T^* = \sum_i X_i^*$ ,

in cui le  $X_i^*$  sono ottenute attribuendo casualmente il segno + o - alla differenza  $X_i$ ,  $i = 1, \dots, n$ , con probabilità  $1/2$ . La distribuzione di probabilità di  $\mathbf{X}^* = \{X_i^*, i = 1, \dots, n\}$ , condizionata all'insieme  $\mathbf{X}$ , è uniforme dentro lo spazio di permutazione  $\mathcal{X}_{/\mathbf{X}}$ , ovvero tutti i punti sono equiprobabili.  $F(z|\mathbf{X}) = \Pr\{T^* \leq z|\mathbf{X}\}$ ,  $z \in \mathcal{R}^1$  è la funzione di ripartizione condizionata (*c.d.f.*) ottenuta via permutazione indotta da  $\mathbf{T}$  dato  $\mathbf{X}$  e tale funzione esiste sempre perché si assume che  $\mathbf{X}$  sia una quantità misurabile rispetto allo spazio misurabile  $(\mathcal{X}, \mathcal{B})$ , a sua volta esistente e ben definito. Sotto l'ipotesi alternativa  $H_1$ , la *c.d.f.* di  $T$  è stocasticamente più grande di quella che  $T$  ha sotto l'ipotesi nulla  $H_0$ , così che valori grandi di  $T$  diventano significativi e il test è non distorto. Indicato  $T_o = T(\mathbf{X})$  il valore osservato di  $T$ , se il *p-value*  $\lambda = \Pr\{T^* \geq T_o|\mathbf{X}\}$  è superiore al livello di soglia fissato  $\alpha$ ,  $H_0$  viene accettata, secondo le usuali regole dei test per la verifica d'ipotesi. Nel caso in cui i campioni abbiano bassa numerosità, la cardinalità dello spazio di permutazione  $\mathcal{X}_{/\mathbf{X}}$  è data da  $M = \#\left[\mathbf{X}^* \in \mathcal{X}_{/\mathbf{X}}\right]$ , in cui  $\mathcal{X}_{/\mathbf{X}}$ , che è sempre finito se la numerosità  $n$  è finita, per  $n$  non troppo piccolo ha, in generale, cardinalità troppo grande perché possano essere esaminati tutti i suoi punti.

L'idea perseguita da molti autori da Dwass [18] in poi è stata quella di attuare, via simulazione, un campionamento di punti condizionale all'insieme dei dati osservati  $\mathbf{X}$  dallo spazio di permutazione: si tratta della procedura di ricampionamento senza reinserimento già descritta nel secondo capitolo applicata, in questo caso, a problemi sviluppati nell'ambito univariato.

L'algoritmo in questo caso è costituito dai seguenti steps:

- s.1) calcolare il valore osservato  $T_o$ , sull'insieme  $\mathbf{X}$  osservato, della statistica  $T : T_o = T(\mathbf{X})$
- s.2) per ciascuna delle  $n$  differenze in  $\mathbf{X}$ , considerare un'attribuzione casuale dei segni in modo tale da ottenere l'elemento  $\mathbf{X}^*$  dallo spazio di permutazione  $\mathcal{X}_{/\mathbf{X}}$
- s.3) calcolare di  $T^* = T(\mathbf{X}^*)$
- s.4) ripetere  $B$  volte gli steps descritti in s.2) e s.3)
- s.5) le  $B$  permutazioni altro non sono se non un campionamento casuale da  $\mathcal{X}_{/\mathbf{X}}$ , i valori di  $T^*$  ottenuti con l'algoritmo simulano la distribuzione

nulla di permutazione di  $T$  e consentono di stimare la *c.d.f.* di permutazione  $F(z|\mathbf{X})$  e la funzione del livello di significatività  $L(z|\mathbf{X}) = \Pr\{T^* \geq z|\mathbf{X}\}$  utilizzando la funzione di ripartizione empirica (*EDF*)  $\widehat{F}_B^*(z) = \#(T^* \leq z)/B$  e la funzione  $\widehat{L}_B^*(z) = \#(T^* \geq z)/B, \forall z \in \mathcal{R}^1$ .

L'algoritmo descritto in s.2) per l'attribuzione casuale dei segni alle differenze  $X$  può essere basato sulla funzione  $X_i^* = X_i \cdot S_i^*$ ,  $i = 1, \dots, n$ , dove le variabili casuali  $S_i^*$  sono *i.i.d.* e ciascuna assume valore  $+1$  o  $-1$  secondo la funzione  $S^* = -1 + 2 \cdot [2 \cdot Rnd]$ , in cui  $Rnd$  diventa un numero pseudo-casuale uniformemente distribuito sull'intervallo aperto  $(0, 1)$  e  $[\cdot]$  indica la parte intera di  $(\cdot)$ . Le stime ottenute in s.5) sono tali che maggiore è il numero  $B$  di iterazioni Monte Carlo (*CMC*), le funzioni  $F(\cdot|\mathbf{X})$  e  $L(\cdot|\mathbf{X})$  sono meglio stimate da  $\widehat{F}_B^*(\cdot)$  e  $\widehat{L}_B^*(\cdot)$ . Ad ogni modo si possono utilizzare  $\widehat{F}_B^*(\cdot)$  e  $\widehat{L}_B^*(\cdot)$  al posto delle rispettive  $F(\cdot|\mathbf{X})$  e  $L(\cdot|\mathbf{X})$  al fine di valutare se i dati osservati hanno lo stesso comportamento ipotizzato sotto  $H_0$ . Il *p-value* stimato a partire dai valori della statistica osservata  $T_o$  corrispondente al livello di significatività osservato da confrontare con quello fissato come soglia per l'accettazione dell'ipotesi nulla è dato da:  $\widehat{\lambda} = \widehat{L}_B^*(T_o) = \#(T^* \geq T_o)/B$ , quindi se  $\widehat{\lambda} \leq \alpha$  si conclude con l'accettazione dell'ipotesi alternativa secondo le regole standard della verifica d'ipotesi.

Se al posto di fare  $B$  campionamenti dallo spazio delle permutazioni, si considera tutto lo spazio di permutazione, le funzioni  $F(z|\mathbf{X})$  e  $L(z|\mathbf{X})$  sono calcolate in modo esatto. Se il numero  $B$  di permutazioni (ricampionamenti) è molto alto e tendente all'infinito, la stima del *p-value*  $\widehat{\lambda}$  è molto vicina al vero valore  $\lambda$ . Se la numerosità  $n$  del campione è grande si può applicare il teorema centrale del limite per approssimare la distribuzione di permutazione  $F(\cdot|\mathbf{X})$  di  $T$ .

E' stato provato che

- ★ se  $n < 25$ , è possibile calcolare esattamente  $T^*$  su tutti i punti dello spazio campionario di permutazione e poi calcolare  $F[t|\mathbf{X}]$ ,
- ★ se invece  $n > 200$ ,  $\widehat{\sigma}_X$  è assunta finita e il rapporto  $(\sum_i X_i^4) / (\sum_i X_i^2)^2$  è piccolo, allora  $F[t|\mathbf{X}]$  si può approssimare con il teorema centrale del limite (CLT). Si ha che il valore atteso e la varianza della variabile  $S^*$

sono rispettivamente  $\mathbb{E}(S^*) = 0$  e  $\mathbb{V}(S^*) = 1$ . Quindi

$$\mathbb{E} \left\{ \left( \sum_i X_i \cdot S_i^*/n \right) | \mathbf{X} \right\} = 0$$

e

$$\mathbb{V} \left\{ \left( \sum_i X_i \cdot S_i^*/n \right) | \mathbf{X} \right\} = \sum_i X_i^2/n^2$$

perché condizionatamente a  $\mathbf{X}$ , le quantità  $X_i$  in  $T^*$  sono quantità fisse. La versione standardizzata è

$$K^* = \left( \sum_i X_i \cdot S_i^* \right) / \left( \sum_i X_i^2 \right)^{1/2} \quad (3.11)$$

che, essendo la somma di  $n$  variabili indipendenti standardizzate, è distribuita approssimativamente come una normale standard. Il test  $K^*$  è approssimativamente distribuito come una normale standard indipendentemente dalla distribuzione sottostante  $P$ , mentre la distribuzione  $t$  di *Student* per essere approssimata alla normale standard presuppone che anche  $P$  abbia distribuzione normale

★ in tutti gli altri casi  $F[t|\mathbf{X}]$  viene calcolata utilizzando i valori medi ottenuti nelle  $B$  iterazioni del *CMC*.

### 3.5 Caso multivariato

Assumiamo, a titolo d'esempio, che una variabile  $q$ -dimensionale  $X$  reale non degenera venga osservata in  $k$  tempi diversi, per esempio  $(\tau_1, \tau_2, \dots, \tau_k)$ , sulle stesse  $n$  unità statistiche in due situazioni sperimentali differenti corrispondenti a due livelli del trattamento  $A_1$  e  $A_2$ . Solitamente le osservazioni al livello  $A_1$  corrispondono alle rilevazioni pre-trattamento e quelle a livello  $A_2$  corrispondono alle rilevazioni post-trattamento. L'insieme dei dati si può indicare con:

$$\begin{aligned} \mathbf{X} &= \{X_{hjit}, t = 1, \dots, k, i = 1, \dots, n, j = 1, 2, h = 1, \dots, q\} \quad (3.12) \\ &= \{\mathbf{X}_{jit}, t = 1, \dots, k, i = 1, \dots, n, j = 1, 2\} \\ &= \{\mathbf{X}_{ji}, i = 1, \dots, n, j = 1, 2\} \end{aligned}$$

dove il medesimo simbolo  $\mathbf{X}$  è usato per rappresentare la variabile risposta  $q$ -variata  $(X_1, \dots, X_q)$ , l'andamento temporale  $q$ -variato  $(\mathbf{X}_1, \dots, \mathbf{X}_k)$  e l'intero insieme dei dati. Nonostante l'apparente complessità della notazione, sarà il contesto a rendere chiaro il suo significato. Le variabili risposta di interesse possono essere ben rappresentate da un modello di tipo additivo come quello sotto riportato:

$$X_{hjit} = \mu_h + \mu_{hit} + \delta_{hjt} + \sigma_{ht} \cdot (\delta_{hjt}) \cdot Z_{hjit}, \quad (3.13)$$

ove  $t = 1, \dots, k$ ,  $i = 1, \dots, n$ ,  $j = 1, 2$ ,  $h = 1, \dots, q$ ;  $\mu_h$  è una costante relativa alla popolazione dell'  $h$ -esima variabile;  $\mu_{hit}$  indica l'effetto del tempo sull'  $h$ -esima variabile osservata al tempo  $t$  e specifico dell'  $i$ -esima unità campionaria;  $\delta_{hjt}$ , che rappresenta un effetto temporale del trattamento al livello  $j$ -esimo sull'  $h$ -esima variabile e generalmente si assume che  $\delta_{h1t} = 0$  e  $\delta_{h2t} \neq 0, \forall(h, t)$ ;  $\sigma_{ht} \cdot (\delta_{hjt}) > 0$  sono coefficienti di scala per la popolazione dell'  $h$ -esima variabile al tempo  $t$ , invarianti rispetto alle unità ed eventualmente dipendenti dal livello del trattamento, come si può vedere dalla presenza di  $\delta_{hjt}$  (quindi dall'indice  $j$ ) nella formula, e tali che siano soddisfatte le relazioni di dominanza  $X_{h1} \stackrel{d}{<} (o \stackrel{d}{>}) X_{h2}$ ,  $h = 1, \dots, q$  quando  $\delta_{h2t} \neq 0$ ;  $Z_{hjit}$  sono errori casuali  $q$ -variati, per ipotesi scambiabili rispetto al livello del trattamento, indipendenti dalle unità, tali che  $E[Z] = 0$  (ove  $Z$  sta per il vettore degli  $Z_{hjit}$ ) e senza una specificata distribuzione di probabilità.

Assumiamo ancora che le  $X_{hjit}$  siano trasformate delle variabili risposta originarie  $U_{hjit}$ , cioè  $X_{hjit} = \phi_h(U_{hjit})$ , con  $t = 1, \dots, k$ ,  $i = 1, \dots, n$ ,  $j = 1, 2$ ,  $h = 1, \dots, q$  con  $\phi_h$  una adeguata funzione di trasformazione dei dati. Per adeguata si intende una funzione che cerchi di massimizzare la potenza del test, ovvero renda massima la probabilità di rifiutare l'ipotesi nulla quando questa è falsa, o che consenta di migliorare l'interpretazione dei risultati: la ricerca di una funzione siffatta dà adito ad un dibattito ancora irrisolto, ma la funzione 'media campionaria' può senz'altro essere utilizzata come appropriato indicatore degli effetti che la somministrazione del trattamento ha sulle unità campionate.

Gli effetti del trattamento  $\delta_{hjt}$  possono essere quantità di natura deterministica o stocastica. Se hanno natura stocastica vengono indicati solitamente con  $\Delta_{hjit} = \delta_{hjt} + \sigma_{\Delta} \cdot W_{hjit}$ , dove  $\delta_{hjt}$  e  $W_{hjit}$  vengono interpretate come medie e deviazioni degli effetti stocastici del trattamento nell'indi-



duo, mentre  $\sigma_{\Delta}$  è un coefficiente di scala per gli effetti stocastici, e possono essere dipendenti o indipendenti dagli errori  $Z_{hjit}$ . Nella seguente trattazione del problema gli effetti del trattamento  $\delta_{hjt}$  hanno natura deterministica o stocastica indipendente dagli errori  $W_{hjit}$ . Inoltre, quando si fa riferimento al generico tempo  $t$  di rilevazione, la coppia multivariata  $(W_{hjit}, Z_{hjit})$  degli errori, in quanto dipendenti dal tempo (come si vede dall'indice  $t$ ), possono essere considerati come processi stocastici discretizzati.

Si vuole verificare l'efficacia del trattamento, senza curarci della variabile temporale, nè delle dipendenze latenti, nè della distribuzione ignota. L'ipotesi nulla che si vuole testare è  $H_0$  tale che:

$$H_0 : \left\{ \bigcap_{t=1}^k [X_{1t} \stackrel{d}{=} X_{2t}] \right\} = \left\{ \bigcap_{t=1}^k \bigcap_{h=1}^q [\delta_{h2t} - \delta_{h1t} = 0] \right\} = \left\{ \bigcap_{t=1}^k \bigcap_{h=1}^q H_{0ht} \right\}$$

verso l'alternativa  $H_1$  tale che

$$H_1 = \left\{ \bigcup_{t=1}^k [X_{1t} <\neq> X_{2t}] \right\} = \left\{ \bigcup_{t=1}^k \bigcup_{h=1}^q H_{1ht} \right\}$$

in cui almeno una delle ipotesi nulle non è verificata. Sotto  $H_0$ , le variabili risposta (profile response) sono scambiabili rispetto ai livelli del trattamento, così che i due profili  $q$ -dimensionali  $X_{1t}$  e  $X_{2t}$ , con  $t = 1, \dots, k$ , sono scambiabili entro le unità. Per quanto riguarda la definizione di  $H_1$ , nel caso generale alcune delle sottoipotesi possono essere ad una coda, o ristrette, e altre a due code. In questo paragrafo, ci si muoverà nel contesto delle alternative ristrette multivariate. Siamo, insomma, interessati a fare inferenza sulla variabile 'differenza' tra le osservazioni relative ai due trattamenti, ovvero la variabile d'interesse è:

$$Y_{hit} = X_{h2it} - X_{h1it} = \delta_{h.t} + \sigma_{ht} \cdot (\delta_{h2t}) \cdot Z_{h2it} - \sigma_{ht} \cdot (\delta_{h1t}) \cdot Z_{h1it}$$

in cui  $t = 1, \dots, k$ ,  $i = 1, \dots, n$ ,  $h = 1, \dots, q$ , con  $\delta_{h.t} = \delta_{h2t} - \delta_{h1t}$ .

Da evidenziare che in questa forma, la variabile 'differenza tra i due trattamenti'  $Y_{hit}$  dipende soltanto dagli effetti del trattamento, gli errori sono scambiabili e che, sotto l'ipotesi nulla  $H_0$ ,  $\sigma_{ht} \cdot (\delta_{h2t}) = \sigma_{ht} \cdot (\delta_{h1t})$

In ambito parametrico la soluzione al problema sopra esposto risulta piuttosto ostica ([19]; [20]; [21]; [22]; [23]; [24]; [25]; [26]; [27]; [28]) soprattutto quando si ha a che fare con variabili risposta di tipo categoriale, oppure

quando non si può assumere la normalità della distribuzione della variabile rilevata o ancora se è ignota la matrice di varianze e covarianze (Robertson *et al.*, 1988). Il problema invece è facilmente risolvibile con il metodo delle permutazioni e la combinazione non parametrica dei test parziali.

I test di permutazione parziali hanno la forma:

$$T_{ht} = \varphi_{ht}\left(\sum_i Y_{hit}\right), \quad t = 1, \dots, k, \quad h = 1, \dots, q$$

in cui  $Y_{hit}$  sono le differenze osservate unità per unità e variabile per variabile,  $\varphi_{ht}$  indica la funzione valore assoluto, se la direzione della sotto-ipotesi alternativa  $H_{1ht}$  di interesse non è specificata, ossia  $H_{1ht} = \left\{X_{h2t} \stackrel{d}{\neq} X_{h1t}\right\}$ , oppure corrisponde alla funzione segno, positivo (+) o negativo (-), a seconda che  $H_{1ht} = \{Y_{hit} > 0\}$  o  $H_{1ht} = \{Y_{hit} < 0\}$  rispettivamente.

Tutti i test parziali  $T_{ht}$ ,  $t = 1, \dots, k$ ,  $h = 1, \dots, q$  sono marginalmente non distorti, ciascuna sotto-ipotesi  $H_{0ht}$  vs  $H_{1ht}$  relativa all'  $h$ -esima variabile può essere considerata come se fosse univariata, le funzioni di combinazione si possono opportunamente applicare.

Assunta la scambiabilità sotto  $H_0$ , nel problema sopra esposto, la distribuzione di permutazione multivariata di  $T = \{T_{ht}, t = 1, \dots, k, h = 1, \dots, q\}$  è generata dall'assegnazione casuale dei due trattamenti  $A_1$  e  $A_2$  ai soggetti. Questo implica che le permutazioni vengono fatte entro gli individui e rispetto al livello del trattamento ( $A_1$  e  $A_2$ ). Così ci sono due permutazioni per ciascun individuo e la cardinalità dello spazio campionario  $\mathcal{X}_{\mathbf{Y}}$  dove  $\mathbf{Y} = \{Y_{hit}, t = 1, \dots, k, i = 1, \dots, n, h = 1, \dots, q\}$  è  $2^n$ . Per ciascuna unità, alla variabile  $Y_{hit}$ , ovvero al vettore delle differenze, viene assegnato, con uguale probabilità, il segno + o -. La statistica test che si può utilizzare è

$$Y^* = \{[Y_{hit}^* = Y_{hit} \cdot S_i^*, t = 1, \dots, k, h = 1, \dots, q], i = 1, \dots, n\}$$

in cui  $S_i^* = 2 \cdot [2 \cdot Rnd] - 1$ ,  $i = 1, \dots, n$ .

I segni generati casualmente sono invarianti entro le unità e rispetto ad  $h$  e  $t$  e sono indipendenti rispetto alle unità. Questo garantisce che le relazioni di dipendenza siano preservate e l'assegnazione casuale dei segni ha uguali effetti nel vettore  $q$ -variato delle differenze relative alla stessa unità.

Se  $q = k = 1$  e le risposte sono omoschedastiche e distribuite normalmente, la verifica d'ipotesi si ottiene via parametrica utilizzando il classico

test  $T$  di *Student* per dati appaiati. Il problema per dati appaiati in ambito multivariato può anche essere visto come un test multivariato per la verifica dell'ipotesi di simmetria perché sotto l'ipotesi nulla le differenze  $Y_{hit}$  sono tutte simmetricamente distribuite intorno allo 0 per ogni  $(t, i, h)$ . Sotto l'alternativa, se l'ipotesi di omoschedasticità è ancora valida, le variabili sono disposte simmetricamente intorno al parametro  $\{\delta_{h,t}\}$ .

Per quanto riguarda la combinazione non parametrica dei test parziali, un metodo può essere quello di combinare in prima istanza rispetto al tempo entro ciascuna variabile, ovvero  $T''_h = \psi_1(T_{h1}, T_{h2}, \dots, T_{hk})$ ,  $h = 1, \dots, q$ , e in un secondo momento rispetto alle  $q$  variabili, ovvero  $T'' = \psi_2(T''_1, \dots, T''_q)$ , dove con  $\psi_1$  e  $\psi_2$  si indicano due appropriate funzioni di combinazione non necessariamente coincidenti. Il procedimento può essere anche eseguito all'inverso, ovvero prima si fa una combinazione rispetto alle variabili per ciascun tempo,  $T'''_t = \psi_2(T_{1t}, T_{2t}, \dots, T_{ht})$ ,  $t = 1, \dots, k$ , e poi rispetto al tempo,  $T''' = \psi_1(T'''_1, \dots, T'''_k)$ . Data la non linearità delle funzioni di combinazione la conclusione cui si arriva seguendo l'uno o l'altro dei due metodi,  $T''$  o  $T'''$ , è pressochè uguale, ma non necessariamente coincidente.

Supponiamo che il modello ora sia:  $X_{hjit} = \mu_h + \mu_{ht} + \delta_{hjt} + \sigma_{ht} \cdot Z_{hjit}$ , con  $t = 1, \dots, k$ ,  $i = 1, \dots, n$ ,  $j = 1, 2$ ,  $h = 1, \dots, q$ , in cui la notazione è la stessa utilizzata nel modello (3.13), ma con i coefficienti di scala  $\sigma_{ht}$  tutti positivi e finiti e il termine d'errore  $kq$ -dimensionale  $Z_{hjit}$  con covarianza finita. La distribuzione marginale di permutazione di tutti i test parziali nella forma standardizzata (3.11), passando per il teorema centrale del limite, si può approssimare ad una normale standard, cioè

$$K_{ht}^* = \frac{\sum_{1 \leq i \leq n} Y_{hit} \cdot S_i^*}{\left\{ \sum_{1 \leq i \leq n} Y_{hit}^2 \right\}^{1/2}} \sim \mathcal{N}(0, 1), \quad t = 1, \dots, k, \quad h = 1, \dots, q$$

dove  $Y_{hit} = X_{h2it} - X_{h1it}$ , mentre la distribuzione di permutazione multivariata di  $\mathbf{K}^*$  è normale multivariata, ovvero  $\mathbf{K}^* \sim \mathcal{N}_{k,q}(\mathbf{0}, \mathbf{\Gamma})$ , dove la matrice di correlazione  $\mathbf{\Gamma}$  è data da

$$\mathbf{\Gamma} = \left\{ \rho_{htjr} = \frac{\sum_{i=1}^n Y_{hit} \cdot Y_{jir}}{\left[ \sum_{i=1}^n Y_{hit}^2 \cdot \sum_{i=1}^n Y_{jir}^2 \right]^{1/2}}, \quad (t, r) = 1, \dots, k, \quad (h, j) = 1, \dots, q \right\}$$

dove

$$\rho_{htjr} = \mathbb{E}(K_{ht}^* \cdot K_{jr}^*) = \frac{\mathbb{E} \left\{ \sum_{i=1}^n \sum_{s=1}^n (Y_{hit} \cdot S_i^*) \cdot (Y_{jir} \cdot S_s^*) \right\}}{\left[ \sum_{i=1}^n Y_{hit}^2 \cdot \sum_{i=1}^n Y_{jir}^2 \right]^{1/2}}$$

e i segni relativi alle diverse unità sono indipendenti

$$\mathbb{E}(S_i^* \cdot S_s^*) = \begin{cases} 1 & \text{se } i = s \\ 0 & \text{se } i \neq s \end{cases} \quad (3.14)$$

### 3.6 Cross-over design

Un disegno *cross-over* è un tipo di *trial* clinico randomizzato in cui  $n$  unità sperimentali vengono suddivise in due gruppi: quelle assegnate casualmente al primo gruppo ricevono prima il placebo e poi il trattamento, quelle randomizzate al secondo gruppo ricevono prima il trattamento e poi il placebo. L'obiettivo dell'esperimento è di verificare, da un lato, se c'è un effetto specifico del trattamento, e, dall'altro, se c'è un effetto determinato dalla modalità di somministrazione del trattamento stesso. Per descrivere le variabili risposta si può utilizzare un modello del tipo:

$$Y_{jti} = \mu + \alpha_{ij} + \beta_{jt} + Z_{jti}, \quad i = 1, \dots, n_j, \quad t = 1, 2, \quad j = 1, 2$$

dove  $\mu$  è una costante ignota relativa alla popolazione; gli  $\alpha_{ij}$  rappresentano gli effetti legati all'individuo, di solito sono considerati parametri di disturbo e, come tali, non sono di interesse per l'analisi, i  $\beta_{jt}$  sono gli effetti del  $t$ -esimo trattamento nel  $j$ -esimo gruppo e gli  $Z_{jti}$  sono degli errori casuali supposti indipendenti rispetto alle unità. Entro le unità, dato che i dati sono appaiati, le differenze possono essere espresse nella forma

$$X_{jti} = Y_{j1i} - Y_{j2i} = \beta_{j1} - \beta_{j2} + Z_{j1i} - Z_{j2i} = \delta_j + \sigma(\delta_j) \cdot Z_{ji}$$

in cui vengono eliminati gli effetti legati all'individuo, i  $\delta_j$  rappresentano le variazioni dovute al trattamento, gli errori  $Z_{ji}$  sono simmetricamente distribuiti intorno allo 0 e i  $\sigma(\delta_j)$  sono coefficienti di scala che possono dipendere dagli effetti principali, ma sono invarianti rispetto alle unità. Le ipotesi sono così formulate:

$$H'_0 : \{\delta_1 = \delta_2 = 0\} \quad \text{vs} \quad H'_1 : \{(\delta_1 \neq 0) \cup (\delta_2 \neq 0)\} \quad (3.15)$$

e

$$H_0'' : \{X_1 \stackrel{d}{=} X_2\} = \{\delta_1 = \delta_2\} \text{ vs } H_1'' : \{\delta_1 \neq \delta_2\} \quad (3.16)$$

Questo secondo sistema d'ipotesi risulta particolarmente interessante quando  $H_0'$  è falsa, ma va comunque sottoposto a test indipendentemente dall'accettazione o meno del primo sistema: può infatti accadere che i due trattamenti abbiano effetti uguali, ma questo non implica che i due gruppi abbiano uguale distribuzione, ovvero accettare  $H_0'$  non è informativo della modalità di somministrazione del trattamento e di una sua eventuale efficacia. L'insieme delle statistiche sufficienti è rappresentato da  $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2\}$ , ovvero è l'insieme costituito da tutte le possibili differenze all'interno dei due gruppi. Verificare il sistema (3.15) presuppone che i dati siano scambiabili entro ciascuna unità, e corrisponde a testare l'ipotesi di simmetria rispetto a 0 per le differenze, così che lo spazio campionario di permutazione è costituito da tutte le possibili  $2^n$  permutazioni. Per testare invece  $H_0''$  vs  $H_1''$  i dati devono essere scambiabili tra i due gruppi perché si deve testare l'uguaglianza delle differenze in un problema a due campioni e le possibili permutazioni sono  $n!$ . I due test sono

$$T_1^* = \sum_i X_{1i} \cdot S_{1i}^* - \sum_i X_{2i} \cdot S_{2i}^* \quad (3.17)$$

in cui  $S_{ji}^*$  sono i segni attribuiti casualmente e

$$T_2^* = \sum_i X_{1i}^*/n_1 - \sum_i X_{2i}^*/n_2. \quad (3.18)$$

Le due statistiche test possono essere così riscritte:

$$T_1^* = \sum_i (\delta_1 + \sigma(\delta_1) \cdot Z_{1i}) \cdot S_{1i}^* - \sum_i (\delta_2 + \sigma(\delta_2) \cdot Z_{2i}) \cdot S_{2i}^* \quad (3.19)$$

e

$$T_2^* = \sum_i (\delta_1^* + \sigma(\delta_1^*) \cdot Z_{1i}^*)/n_1 - \sum_i (\delta_2^* + \sigma(\delta_2^*) \cdot Z_{2i}^*)/n_2$$

Sotto l'ipotesi nulla i valori che assumono entrambi i test  $T_1^*$  e  $T_2^*$  dipendono dalla permutazione di errori scambiabili. Tutti i test considerati sono esatti, consistenti e non distorti.

### 3.7 Digressione sui disegni cross-over

La caratteristica peculiare del disegno *cross-over* è che ciascun paziente riceve in successione più di un trattamento in studio. Ciò che cambia, da paziente a paziente, è la successione dei trattamenti: ciascun paziente viene randomizzato ad una delle possibili *sequenze* di trattamenti [16]. Se si hanno solo due trattamenti,  $A$  e  $B$ , le possibili sequenze sono due, ovvero  $AB$  e  $BA$ . Ciascun intervallo temporale in cui si somministra uno dei due trattamenti in studio è detto *periodo*. Nel disegno *cross-over* con due trattamenti si hanno due periodi: il periodo 1, nel quale il gruppo di pazienti randomizzato alla sequenza  $AB$  riceve  $A$  e quello randomizzato alla sequenza  $BA$  riceve  $B$ ; il periodo 2, nel quale il primo gruppo riceve  $B$  e il secondo  $A$ . Tale disegno è detto anche *cross-over*  $2 \times 2$  (due trattamenti e due periodi). Per evitare che l'effetto di un trattamento si protragga nel periodo successivo, il cosiddetto effetto *carry-over*, è necessario introdurre un adeguato intervallo tra la fine di un trattamento e l'inizio del successivo, chiamato *wash-out*, la cui durata è determinata dalle caratteristiche dei trattamenti in studio. Tra gli effetti che un disegno *cross-over* può mettere in luce sono da citare: l'effetto principale del trattamento, l'effetto principale del periodo, l'effetto *carry-over*, l'effetto di interazione tra periodo e trattamento, l'effetto sequenza e l'effetto soggetto. Il vantaggio di un approccio *cross-over* sta nella minore dimensione campionaria in quanto ciascun paziente contribuisce alla numerosità campionaria per più di un trattamento e la variabilità delle misurazioni entro soggetto è minore rispetto a quella tra soggetti. Apparentemente semplice ed estremamente vantaggioso, il disegno *cross-over* risulta essere molto complesso sul piano pratico-logistico e problematico sul piano metodologico. Per quanto riguarda il piano logistico, al singolo paziente si richiede una notevole disponibilità in termini di tempo, dal momento che devono essere ripetute tutte le procedure e le misurazioni dettate dal protocollo, tante volte quanti sono i trattamenti. Sul piano metodologico, le complicazioni principali sono date dagli abbandoni prematuri, chiamati in gergo *drop-out*, dalla presenza di un effetto "periodo estremo" e dalla presenza dell'effetto *carry-over*. Per quanto riguarda l'aspetto relativo ai *drop-out*, il fenomeno degli abbandoni prematuri deve essere contemplato già in fase di pianificazione, e può essere risolto aumentando il numero di soggetti randomizzati sulla base del

numero previsto di abbandoni e stabilendo a priori come valutare i pazienti nell'analisi intention-to-treat. Un numero di abbandoni molto superiore a quello previsto rappresenta sempre un problema qualunque sia il disegno sperimentale. Per effetto periodo estremo si intende una situazione in cui, nei periodi successivi al primo, la patologia si modifica al tal punto da violare in modo sostanziale uno o più criteri di elegibilità dei pazienti. La situazione limite, tutt'altro che impossibile in alcune patologie, si ha quando la maggior parte dei paziente è guarita o deceduta: ne consegue che molti farmaci, anche quelli efficaci nelle forme tipiche di una patologia, hanno un'azione minore o addirittura trascurabile, nelle forme molto lievi o molto avanzate, indipendentemente dall'effetto dei trattamenti, e, in aggiunta, che la popolazione in studio nel secondo periodo non è più quella *target* prevista dal protocollo. La presenza di un effetto *carry-over* rende difficile l'interpretazione dei dati raccolti nei periodi successivi al primo. Quando alla fine di uno studio cross-over ci si ritrova con un elevato numero di abbandoni con un carry-over statisticamente significativo, si utilizza solo il primo periodo dello studio: ci si riconduce quindi ad un disegno a *gruppi paralleli*, ovvero un disegno in cui ciascun paziente riceve un solo trattamento, quello del gruppo in cui è stato randomizzato, per la durata prevista dal protocollo e tutti i gruppi sono trattati e valutati in contemporanea (è per questo che i gruppi sono definiti paralleli). Anche questa soluzione presenta dei problemi, il primo dei quali è che la dimensione del campione diventa insufficiente riducendo, a livelli inaccettabili, la potenza dello studio. Di fatto in tali condizioni lo studio può al più generare delle ipotesi che dovranno essere confermate in altri studi. Esistono numerose situazioni in cui tale disegno rappresenta un'ottima soluzione, ovvero quando sono verificate le seguenti condizioni: l'effetto dei trattamenti si manifesta presto, per cui la durata di somministrazione dei trattamenti può essere breve; i trattamenti in studio hanno effetti reversibili, cioè non modificano il quadro clinico in modo prolungato rispetto alla durata dello studio o permanente; la malattia o condizione d'interesse è relativamente stabile nell'arco di tempo coperto dallo studio, cioè non va incontro ad un rilevante miglioramento o peggioramento; i trattamenti in studio hanno effetti che non perdurano anche nel periodo successivo rispetto a quello in cui sono stati somministrati.

### 3.8 Verifica d'indipendenza

Intuitivamente si potrebbe pensare che, essendo le permutazioni dei segni sul test  $T_1^*$  indipendenti dai valori delle permutazioni su  $T_2^*$ , i due test sono indipendenti per testare  $H_0 = H'_0 \cap H''_0$ . Le simulazioni effettuate per provare quanto affermato sopra hanno fornito risultati interessanti, per quanto trattandosi di test applicati congiuntamente sui medesimi dato ci sia da attendersi, secondo Lehmann (1986) [17], una dipendenza positiva.

Sono stati considerati due gruppi con osservazioni appaiate (secondo un disegno sperimentale basato sul confronto prima-dopo), di numerosità diverse: il campione può essere descritto dal modello:

$$X_{tji} = \mu_j + Z_{tji} + \eta_t + \delta_j \quad (3.20)$$

con

$t = 1, 2$  e  $t = 1$  indica la prima rilevazione (pre trattamento),  $t = 2$  la seconda rilevazione (post trattamento);

$j = 1, 2$  indica il gruppo;

$i = 1, \dots, n_j$  indica la numerosità del gruppo;

$\eta_t$  indica l'effetto del tempo ed è tale che  $\eta_t = 0$  se  $t = 1$  e  $\eta_t \neq 0$  se  $t = 2$ ;

$\delta_j$  indica l'effetto del trattamento e  $\delta_j = 0$  se  $t = 1$  e  $\delta_j \neq 0$  se  $t = 2$ ;

$Z_{tji}$  sono degli errori non normali, costruiti come il quadrato della somma di variabili uniformi nell'intervallo  $(0, 1)$ .

Quindi si ha che

$$\begin{cases} X_{11i} = \mu_1 + Z_{11i} & \text{se } j = 1 \text{ e } t = 1 \\ X_{21i} = \mu_1 + Z_{21i} + \eta_2 + \delta_1 & \text{se } j = 1 \text{ e } t = 2 \end{cases}$$

$$\text{con } Z_{11i} = \left( \sum_{h=1}^3 \mathcal{U}_h \right)^2 \text{ e } Z_{21i} = \left( \sum_{h=2}^4 \mathcal{U}_h \right)^2$$

$$\begin{cases} X_{12i} = \mu_2 + Z_{12i} & \text{se } j = 2 \text{ e } t = 1 \\ X_{22i} = \mu_2 + Z_{22i} + \eta_2 + \delta_2 & \text{se } j = 2 \text{ e } t = 2 \end{cases}$$



con  $Z_{12i} = \left( \sum_{h=5}^7 \mathcal{U}_h \right)^2$  e  $Z_{22i} = \left( \sum_{h=6}^8 \mathcal{U}_h \right)^2$ . Sono stati considerati 4 test,  $T_1$  per verificare se c'è un effetto del trattamento nel primo gruppo,  $T_2$  per verificare se c'è un effetto del trattamento nel secondo gruppo,  $T_3$  era il test combinato di  $T_1$  e  $T_2$  per verificare se almeno in uno dei due gruppi c'è un effetto del trattamento, ed infine  $T_4$  per verificare l'uguaglianza in distribuzione dei due gruppi. I test e i relativi sistemi d'ipotesi sono:

$$\begin{aligned} T_1 &\longrightarrow \begin{cases} H_{01} : \eta_2 + \delta_1 = 0 \\ H_{11} : \eta_2 + \delta_1 \neq 0 \end{cases} \\ T_2 &\longrightarrow \begin{cases} H_{02} : \eta_2 + \delta_2 = 0 \\ H_{12} : \eta_2 + \delta_2 \neq 0 \end{cases} \\ T_3 &\longmapsto \begin{cases} H_{03} : \eta_2 + \delta_1 = 0 \cap \eta_2 + \delta_2 = 0 \\ H_{13} : \eta_2 + \delta_1 \neq 0 \cup \eta_2 + \delta_2 \neq 0 \end{cases} \\ T_4 &\longmapsto \begin{cases} H_{04} : \delta_1 = \delta_2 \\ H_{14} : \delta_1 \neq \delta_2 \end{cases} \end{aligned}$$

In maniera particolare, sotto  $H_0$ , si ha:

- $\delta_1 = \delta_2 = 0$
- per il primo campione, per costruzione,  $n_1 = 30$ ;  $Y_{11} = (\mathcal{U}_1 + \mathcal{U}_2 + \mathcal{U}_3)^2$  e  $Y_{21} = (\mathcal{U}_2 + \mathcal{U}_3 + \mathcal{U}_4)^2$ , con  $Y_{11}$  e  $Y_{21}$  indicanti rispettivamente la rilevazione pre e post trattamento (o parimenti le rilevazioni in  $t_1$  e  $t_2$ );
- per il secondo campione, per costruzione,  $n_2 = 20$ ;  $Y_{12} = (\mathcal{U}_5 + \mathcal{U}_6 + \mathcal{U}_7)^2$  e  $Y_{22} = (\mathcal{U}_6 + \mathcal{U}_7 + \mathcal{U}_8)^2$ , con  $Y_{11}$  e  $Y_{21}$  indicanti rispettivamente la rilevazione pre e post trattamento (o parimenti le rilevazioni in  $t_1$  e  $t_2$ ).

Si vede dunque che le variabili pre e post, avendo in comune due addendi, sono dipendenti. La formulazione del modello è più compatta, rispetto a quella vista in (3.20). Le variabili  $Y_{tj}$ ,  $t = 1, 2$ ,  $j = 1, 2$ , sono rappresentative del termine d'errore, dell'effetto del tempo e dell'effetto del gruppo. Sono state fatte anche tre simulazioni sotto  $H_1$ , scegliendo diversi livelli di  $\delta_1$  e  $\delta_2$ . In maniera particolare, nella prima simulazione si ha:

- $\delta_1 = 0.2$  e  $\delta_2 = 0.4$

- nel primo campione  $n_1 = 30$ ;  $Y_{11} = (\mathcal{U}_1 + \mathcal{U}_2 + \mathcal{U}_3)^2$  e  $Y_{21} = (\mathcal{U}_2 + \mathcal{U}_3 + \mathcal{U}_4)^2 + 0.2$ , sempre con  $Y_{11}$  e  $Y_{21}$  indicanti rispettivamente le rilevazioni in  $t_1$  e  $t_2$ ,
- nel secondo campione  $n_2 = 20$ ;  $Y_{12} = (\mathcal{U}_5 + \mathcal{U}_6 + \mathcal{U}_7)^2$  e  $Y_{22} = (\mathcal{U}_6 + \mathcal{U}_7 + \mathcal{U}_8)^2 + 0.4$ , con  $Y_{12}$  e  $Y_{22}$  indicanti rispettivamente le rilevazioni in  $t_1$  e  $t_2$ ,

nella seconda simulazione si ha:

- $\delta_1 = 0.4$  e  $\delta_2 = 0.4$ ,
- nel primo campione  $n_1 = 30$ ;  $Y_{11} = (\mathcal{U}_1 + \mathcal{U}_2 + \mathcal{U}_3)^2$  e  $Y_{21} = (\mathcal{U}_2 + \mathcal{U}_3 + \mathcal{U}_4)^2 + 0.4$ ,
- nel secondo campione  $n_2 = 20$ ;  $Y_{12} = (\mathcal{U}_5 + \mathcal{U}_6 + \mathcal{U}_7)^2$  e  $Y_{22} = (\mathcal{U}_6 + \mathcal{U}_7 + \mathcal{U}_8)^2 + 0.4$ ,

ed infine, nella terza simulazione si ha:

- $\delta_1 = 0.3$  e  $\delta_2 = 0.7$
- nel primo campione  $n_1 = 30$ ;  $Y_{11} = (\mathcal{U}_1 + \mathcal{U}_2 + \mathcal{U}_3)^2$  e  $Y_{21} = (\mathcal{U}_2 + \mathcal{U}_3 + \mathcal{U}_4)^2 + 0.3$ ,
- nel secondo campione  $n_2 = 20$ ;  $Y_{12} = (\mathcal{U}_5 + \mathcal{U}_6 + \mathcal{U}_7)^2$  e  $Y_{22} = (\mathcal{U}_6 + \mathcal{U}_7 + \mathcal{U}_8)^2 + 0.7$ .

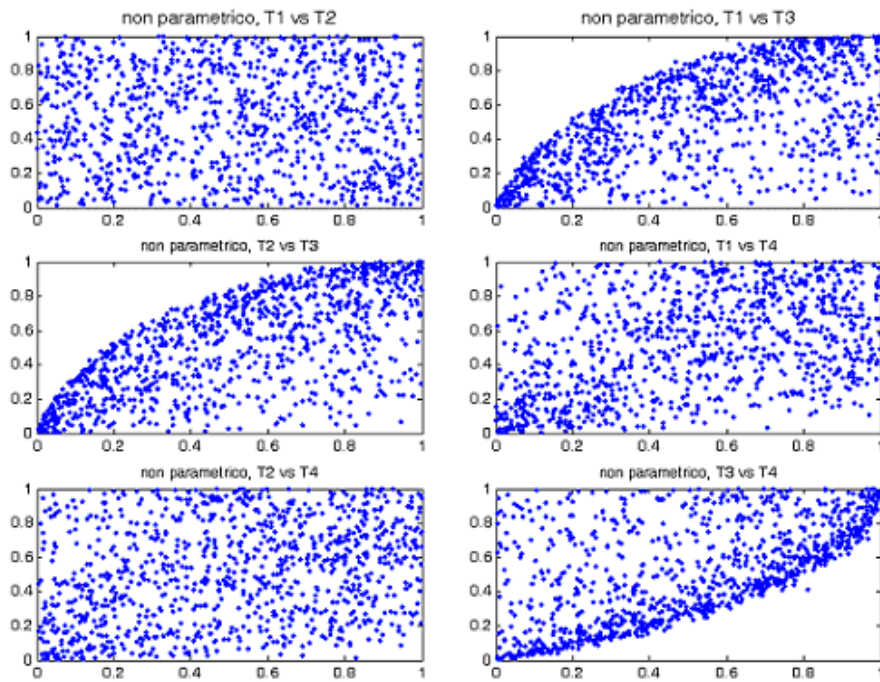
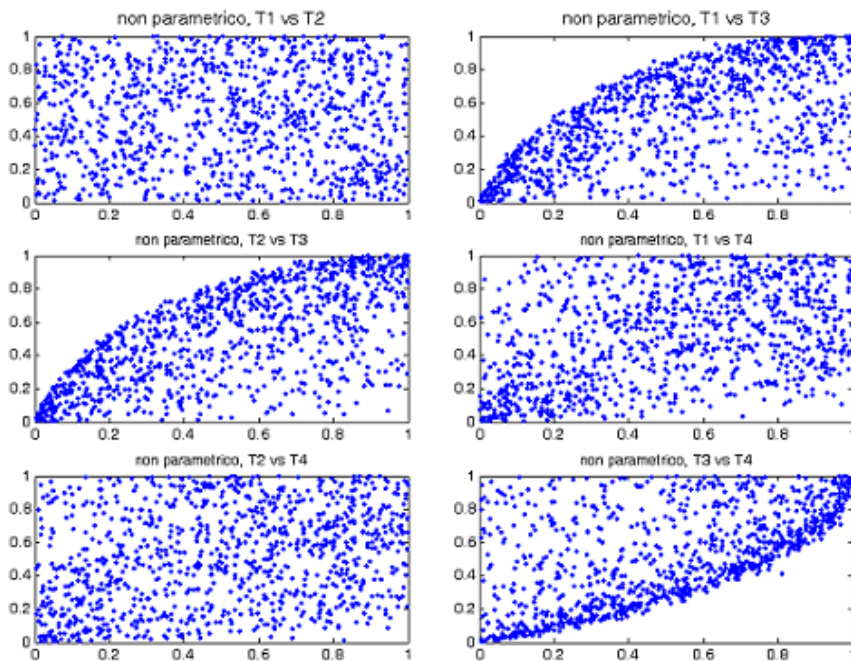
Ciascuno dei *p-value* dei test non parametrici, memorizzati nel vettore finale dei *p-value*, è stato ottenuto con  $B = 500$  e sono state fatti  $MC = 1000$  campionamenti Monte Carlo. Il procedimento è costituito dalle seguenti fasi:

1. Si generano i dati sotto  $H_0$  e sotto  $H_1$  con le modalità sopra descritte,
2. Si calcolano i test  $T_1$  e  $T_2$ , ovvero i test per campioni con dati appaiati nel primo e nel secondo campione rispettivamente, il test  $T_3$  combinazione di  $T_1$  e  $T_2$  (funzione di combinazione di Fisher) ed infine il test per campioni indipendenti  $T_4$ . Si fanno  $B = 500$  permutazioni, si ottengono i rispettivi p-value e si memorizzano  $\lambda_1^{OB}$ ,  $\lambda_2^{OB}$ ,  $\lambda_3^{OB}$  e  $\lambda_4^{OB}$ . Per

fare un confronto, questo secondo passaggio è stato eseguito anche per via parametrica, utilizzando i test parametrici consueti e in maniera particolare  $T_5$  e  $T_6$  sono i test *t-Student* per campioni con dati appaiati, per il primo e il secondo campione, il test  $T_7$  è la combinazione di *Fisher* di  $T_5$  e  $T_6$ , ed infine il test  $T_8$  che è sempre un test *t* questa volta per campioni indipendenti,

3. Si ripetono i passi 1.) e 2.) per 1000 volte. Si ottiene, alla fine, una matrice  $1000 \times 8$  che viene utilizzata per la verifica dell'ipotesi di indipendenza dei test  $T_3$  e  $T_4$  non parametrici e di  $T_7$  e  $T_8$ , ovvero dei corrispondenti test parametrici

Come si vede dai plot riportati nelle figure seguenti, sia sotto l'ipotesi nulla  $H_0$  sia sotto l'ipotesi alternativa,  $H_1$ , i test non parametrici  $T_3$  e  $T_4$  e i test parametrici  $T_7$  e  $T_8$ , risultano positivamente dipendenti. Riportato nell'ambito di un trial clinico, la dipendenza evidenziata dall'algoritmo di simulazione può essere interpretata come un effetto di interazione tra trattamento e gruppo.

Figura 3.1: test non parametrici: simulazioni sotto  $H_0$ Figura 3.2: test parametrici: simulazioni sotto  $H_0$

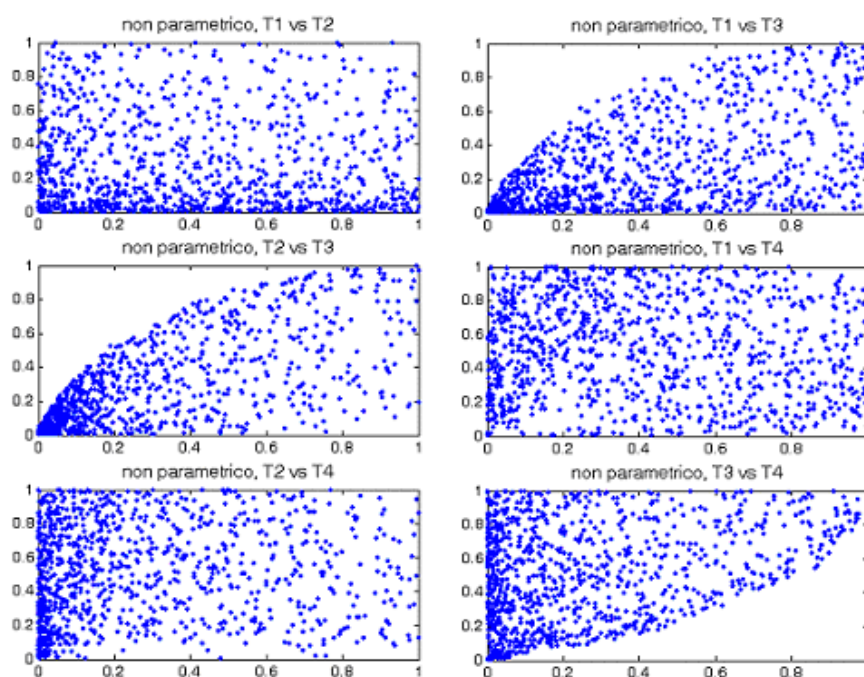


Figura 3.3: test non parametrici: simulazioni sotto  $H_1$  e  $\delta_1 = 0.2$ ,  $\delta_2 = 0.4$

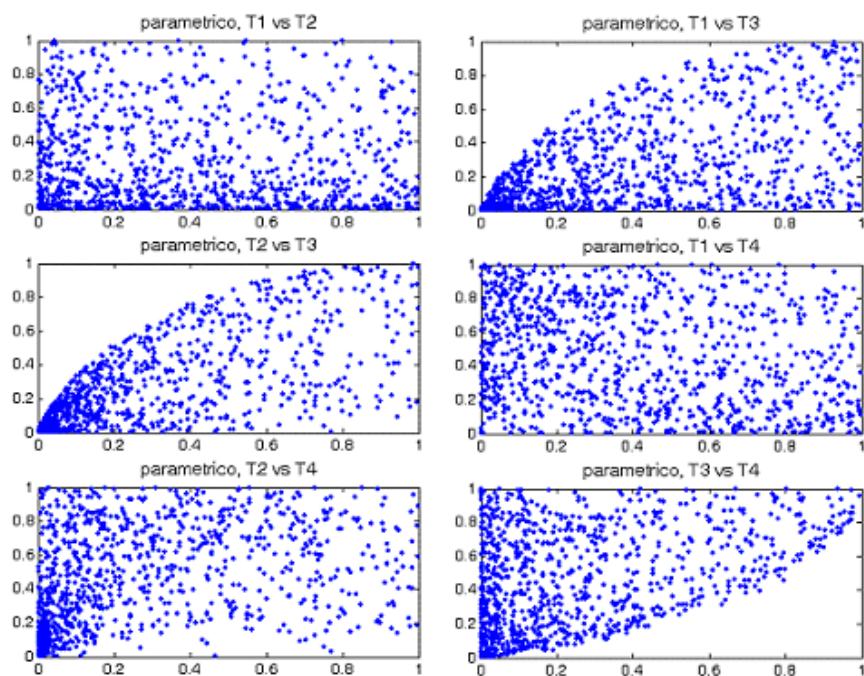
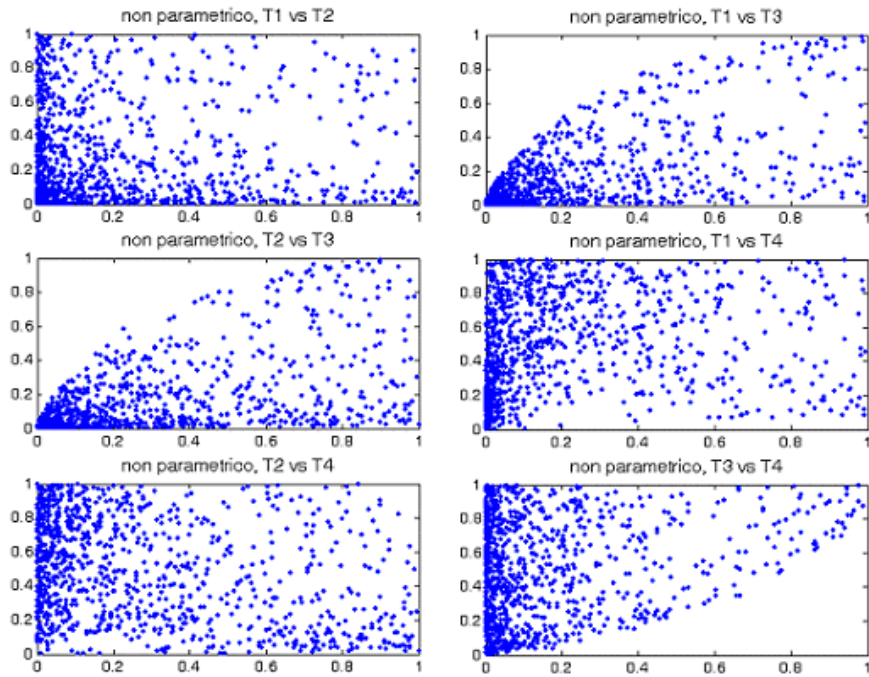
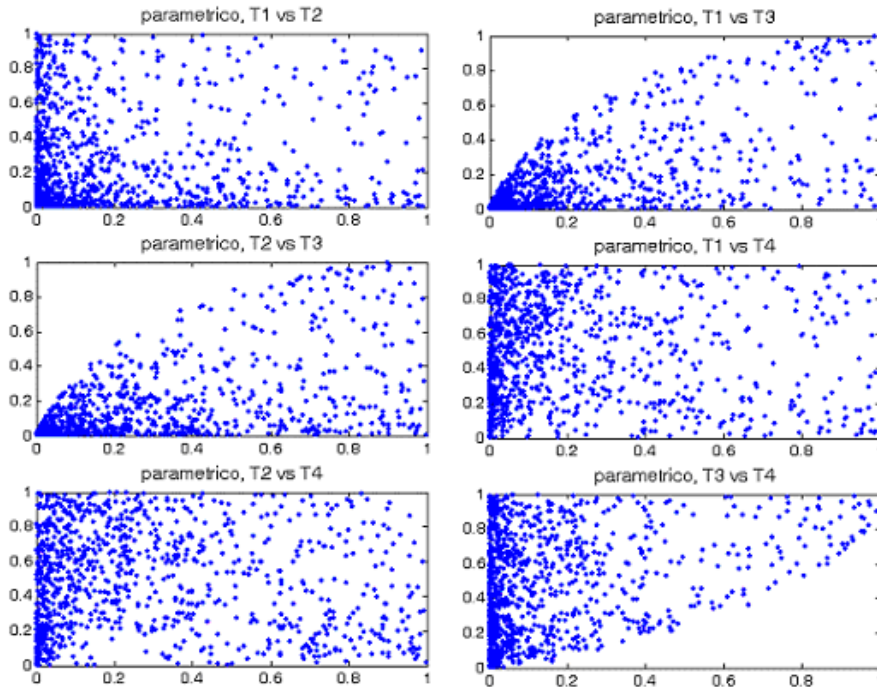


Figura 3.4: test parametrici: simulazioni sotto  $H_1$  e  $\delta_1 = 0.2$ ,  $\delta_2 = 0.4$

Figura 3.5: test non parametrici: simulazioni sotto  $H_1$  e  $\delta_1 = \delta_2 = 0.4$ Figura 3.6: test parametrici: simulazioni sotto  $H_1$  e  $\delta_1 = \delta_2 = 0.4$

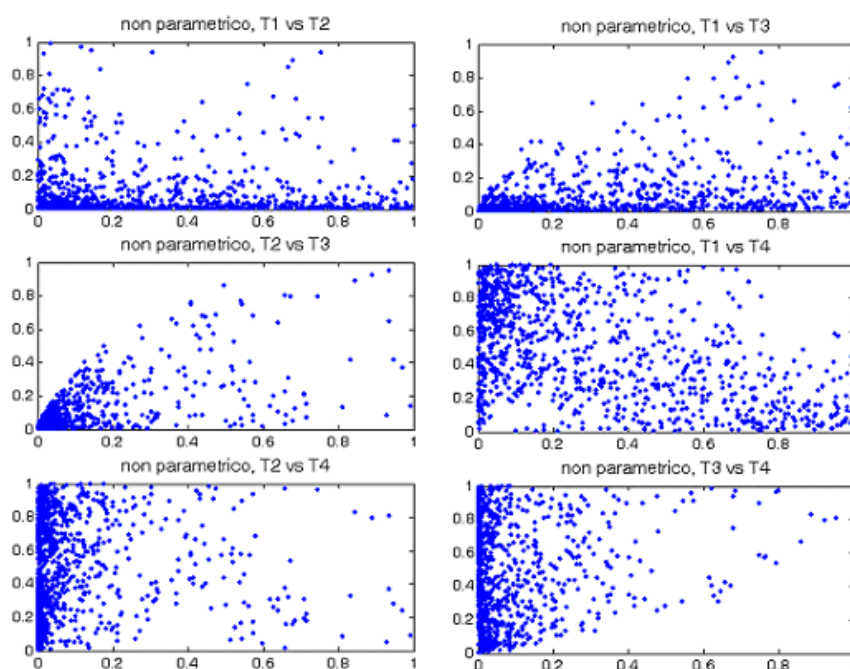


Figura 3.7: test non parametrici: simulazioni sotto  $H_1$  e  $\delta_1 = 0.3$ ,  $\delta_2 = 0.7$

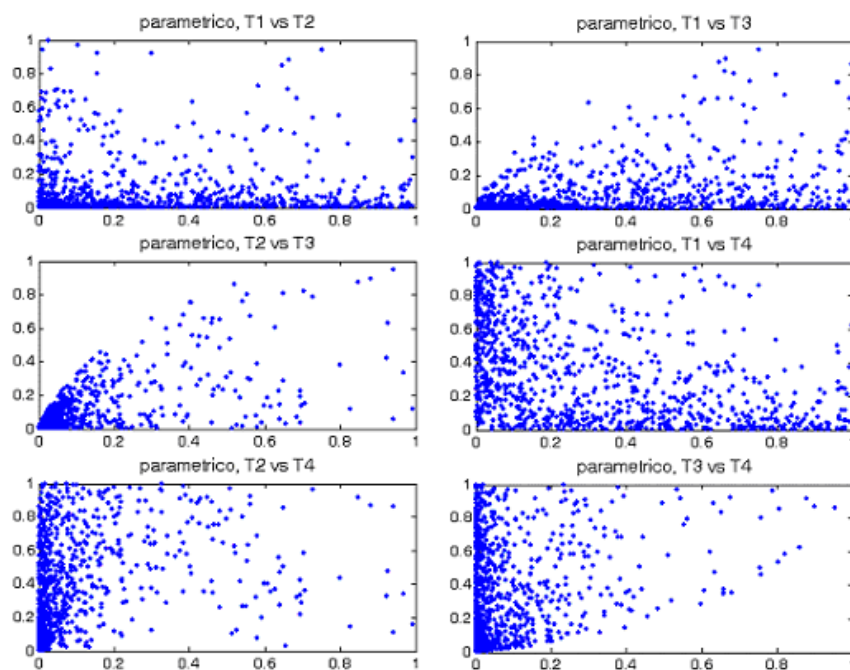


Figura 3.8: test parametrici: simulazioni sotto  $H_1$  e  $\delta_1 = 0.3$ ,  $\delta_2 = 0.7$





# Capitolo 4

## Test di permutazione con missing data

### 4.1 Introduzione al problema dei missing data

Il problema dei *missing data* è assai frequente ([29]; [30]; [34]; [35]; [31]; [32]; [33]) specie nelle indagini demografiche quando una popolazione sotto studio in un certo anno, al momento del *follow-up* effettuato dopo qualche anno si presenta in numero ridotto per ragioni tra le più svariate. Spesso le analisi di tali dati procedono con l'assunzione, esplicita o implicita, che il processo generatore di *missing-data* può essere ignorato. E' naturale chiedersi quando questa procedura è appropriata. Si tratta di un problema assai discusso in ambito statistico cui sono state proposte parecchie soluzioni. Un lavoro molto importante a riguardo è stato quello di Rubin (1976) [29]. L'autore fornisce questa definizione:

*“The missing data are missing at random if for each possible value of the parameter  $\phi$ , the conditional probability of the observed pattern of missing data, given the missing data and the value of the observed data, is the same for all possible values of the missing data. The observed data are observed at random if for each possible value of the missing data and the parameter  $\phi$ , the conditional probability of the observed pattern of missing data, given the missing data and the observed data, is the same for all possible values of the observed data. The parameter  $\phi$  is distinct from  $\theta$  if there are no a priori*

ties, via parameter space restrictions or prior distributions, between  $\phi$  and  $\theta$ .”

Sia  $U = (U_1, \dots, U_n)$  un vettore di variabili casuali con funzione di densità  $f_\theta$ . L'obiettivo è di fare inferenza su  $\theta$ , il parametro di questa densità. Sia  $M = (M_1, \dots, M_n)$  il vettore associato di variabili casuali indicatrici dei dati mancanti, in cui  $M_i$  assume il valore 0 o 1. La probabilità che il vettore  $M$  assuma il valore  $m = (m_1, \dots, m_n)$  dato che  $U$  assume il valore  $u = (u_1, \dots, u_n)$  è  $g_\phi(m|u)$ , dove  $\phi$  è il vettore di parametri di disturbo della distribuzione. La distribuzione condizionale  $g_\phi$  corrisponde al processo che causa i dati mancanti; se  $m_i = 1$ , il valore della variabile casuale  $U_i$  viene osservato, mentre se  $m_i = 0$ , il valore della variabile casuale  $U_i$  non viene osservato. Più precisamente, per compattare la notazione, si può definire il vettore di variabili casuali  $V = (V_1, \dots, V_n)$ , con range esteso per includere il valore \* indicante i *missing data*:  $v_i = u_i(m_i = 1)$  e  $v_i = *(m_i = 0)$ : oggetto dell'inferenza infatti non è la distribuzione di  $U$ , ma la distribuzione di  $V$ , ovvero la distribuzione dei valori osservati.

## 4.2 Esempi dei processi che generano i *missing-data*

*Esempio 1:* Supponiamo di avere  $n$  campioni di una lega metallica e su ciascuno vogliamo registrare alcune caratteristiche con uno strumento che ha una probabilità  $\phi$  costante di non riuscire a rilevare le caratteristiche in questione in tutti i campioni. Allora

$$g_\phi(m|u) = \prod_{i=1}^n \phi^{m_i} (1 - \phi)^{1-m_i} \quad (4.1)$$

*Esempio 2:* Sia  $u_i$  il valore della pressione sanguigna nell' $i$ -esimo soggetto ( $i = 1, \dots, n$ ) in un rilevamento ospedaliero. Supponiamo che  $v_i = *$  se  $u_i$  è inferiore a  $\phi$ , dove con  $\phi$  si indica il valor medio della pressione sanguigna nella popolazione; ovvero vengono registrati i valori della pressione sanguigna solo se sono superiori alla media. Allora

$$g_\phi(m|u) = \prod_{i=1}^n \delta \{ \gamma(u_i - \phi) - m_i \} \quad (4.2)$$

in cui  $\gamma(a) = 1$  se  $a \geq 0$  e 0 altrimenti;  $\delta(a) = 1$  se  $a = 0$  e 0 altrimenti.

*Esempio 3:* Le osservazioni vengono prese in sequenza finché una particolare funzione delle osservazioni sta in una regione critica  $C$ . Qui  $n$  è assunto infinito e, per un certo  $n_1$  che è una funzione delle osservazioni,  $v_i \neq * (i \leq n_1)$  e  $v_i = * (i > n_1)$ . Così:

$$g_\phi(m|u) = \prod_{i=1}^{n_1} \delta(1 - m_i) \prod_{i=n_1+1}^n \delta(m_i) \quad (4.3)$$

dove  $n_1$  è il minimo  $k$  tale per cui  $Q_k(u_1, \dots, u_n) \in C$ .

*Esempio 4* Sia  $n = 2$ . Se  $u_1 \geq 0$  : con probabilità  $\phi$ ,  $v_1 \neq *$  e  $v_2 = *$ , e con probabilità  $(1 - \phi)$ ,  $v_1 \neq *$  e  $v_2 \neq *$ . Se  $u_1 < 0$  : con probabilità  $\phi$ ,  $v_1 \neq *$  e  $v_2 = *$ , e con probabilità  $(1 - \phi)$ ,  $v_1 \neq *$  e  $v_2 = *$ , e con probabilità  $(1 - \phi)$ ,  $v_1 = *$  e  $v_2 \neq *$ . Così:

$$g_\phi(m|u) = \begin{cases} \phi & \text{se } m = (1, 0) \\ (1 - \phi)\gamma(u_1) & \text{se } m = (1, 1) \\ (1 - \phi)\{1 - \gamma(u_1)\} & \text{se } m = (0, 1) \\ 0 & \text{se } m = (0, 0) \end{cases} \quad (4.4)$$

Sia  $\tilde{v} = (\tilde{v}_1, \dots, \tilde{v}_n)$  una realizzazione nel campione della variabile  $V$ , cioè ciascuna  $\tilde{v}_i$  può essere un'osservazione, un numero noto, o un dato mancante,  $*$ . Se  $\tilde{v}_i$  è un numero, vuol dire che  $\tilde{m}_i$ , osservato in  $M_i$ , è 1 e che  $\tilde{u}_i$ , osservato in  $U_i$ , è uguale a  $\tilde{v}_i$ . Se invece  $\tilde{v}_i$  è un valore mancante  $*$ , vuol dire che  $\tilde{m}_i = 0$  e che il valore di  $U_i$  non è noto.

	Missing data,	Observed data,	
Example	missing at random	observed at random	$\phi$ distinct from $\theta$
1	Always MAR	Always OAR	Always distinct
2	MAR only if all $\tilde{m}_i=1$	OAR only if all $\tilde{m}_i=0$	Distinct only if mean blood pressure in the population is known a priori
3	Always MAR	Never OAR	Always distinct
4	MAR unless $\tilde{m}_i=(0,1)$	OAR unless $\tilde{m}_i=(1,1)$	Distinct if a priori $\phi$ is not restricted by $\theta$

Tabella 4.1: Classificazione degli esempi

Il valore osservato di  $M$ , chiamato  $\tilde{m}$ , effettua una partizione di ciascuno dei vettori delle variabili casuali e il vettore dei valori osservati in due vettori

corrispondenti a  $\tilde{m}_i = 0$  se il valore manca e  $\tilde{m}_i = 1$  se il valore è osservato. Per convenienza scriviamo:

$$U = (U_{(0)}, U_{(1)}), \quad V = (V_{(0)}, V_{(1)}), \quad u = (u_{(0)}, u_{(1)}), \quad v = (v_{(0)}, v_{(1)})$$

dove per definizione  $v_{(0)} = (*, \dots, *)$  e  $u_{(1)} = v_{(1)}$ . È importante ricordare che queste partizioni sono quelle corrispondenti a  $m = \tilde{m}$ , il *pattern* (modello, l'andamento) osservato dei valori mancanti. Per ulteriore convenienza di notazione, si può definire  $\tilde{u} = (u_{(0)}, \tilde{u}_{(1)})$ , in cui  $\tilde{u}$  è costituito da un vettore  $u_{(0)}$  i cui argomenti corrispondono alle variabili casuali non osservate e da un vettore di numeri noti,  $\tilde{u}_{(1)} = \tilde{v}_{(1)}$ , corrispondenti ai valori delle variabili casuali osservate. L'obiettivo è di usare  $\tilde{u}$ , oppure  $\tilde{m}$  e  $\tilde{u}_{(1)}$ , per fare inferenza sul parametro  $\vartheta$ . Nella pratica si ignora il processo che genera i valori mancanti quando si fa inferenza. La questione principale riguarda la necessità di trovare delle condizioni, le più deboli e semplici, su  $g_\phi$  che consentano, pur ignorando il processo generatore dei dati mancanti, di ottenere, tramite appropriata inferenza, informazioni sul parametro  $\vartheta$ . Ci sono tre condizioni rilevanti per rispondere a questa domanda.

**Definizione 1** I *missing data* sono *missing at random* (MAR) se, per ogni valore di  $\phi$ ,  $g_\phi(\tilde{m}|\tilde{u})$  assume lo stesso valore per tutti gli  $u_{(0)}$

**Definizione 2** Gli *observed data* sono *observed at random* (OAR) se, per ogni valore di  $\phi$  e  $u_{(0)}$ ,  $g_\phi(\tilde{m}|u)$  assume lo stesso valore per tutti gli  $u_{(1)}$ .

**Definizione 3** Il parametro  $\phi$  è distinto da  $\vartheta$  se il loro spazio parametrico congiunto fattorizza in un  $\phi$ -spazio e  $\vartheta$ -spazio, e se le loro distribuzioni a priori, quando sono specificate per  $\phi$  e  $\vartheta$ , sono indipendenti.

Se i *missing data* sono MAR e gli *observed data* sono OAR, i *missing data* sono *missing completely at random* (MCAR). In questo caso, la mancanza non dipende dai valori osservati o non osservati, e i valori osservati possono essere considerati come un sotto campione casuale dell'intero insieme dei dati. In questo caso è possibile ignorare il processo che genera i dati mancanti e far inferenza su  $\vartheta$ . Se i dati sono MAR e  $\phi$  e  $\vartheta$  sono distinti, allora l'inferenza di verosimiglianza su  $\vartheta$  può essere basata sulla funzione di verosimiglianza

senza includere un termine per il processo dei *missing data*, in questo caso il processo che genera i *missing data* è definito *ignorable* e l'analisi può essere condotta via condizionamento ai dati effettivamente osservati. Se i *missing data* sono *missing not at random* (non-MAR), per fare una corretta inferenza, in ambito parametrico, il processo generatore deve essere specificato. Dal momento che in molti casi i *missing data* provengono da indagini campionarie e le non risposte sono dovute ai più svariati motivi, in generale è molto complesso definire un modello che, tra tutti, meglio descriva i dati, siano essi osservati o mancanti. E' per questo che un'analisi condotta su dati non-MAR è di gran lunga più difficile rispetto a quella condotta su dati MCAR.

### 4.3 Approccio di permutazione

In situazioni di MCAR, le soluzioni parametriche sfruttano il *deletion principle* o un metodo di imputazione: il primo analizza i dati restanti dopo la rimozione di tutti i vettori incompleti, il secondo consiste nella sostituzione dei dati mancanti con appropriate funzioni, in genere la media, dei dati effettivamente osservati. Tutto questo perché, un approccio parametrico richiede che vi sia a disposizione un insieme completo di dati.

Un approccio di permutazione, invece, considera i dati come sono senza invocare il principio di eliminazione o imputazione: quando infatti si eliminano i vettori incompleti, si eliminano anche quelle informazioni che sono utili per l'analisi, quando si fanno delle imputazioni si possono introdurre informazioni distorte che possono influenzare negativamente l'analisi. Senza perdita di generalità, supponiamo di voler testare l'uguaglianza in distribuzione di  $C$  campioni, in cui alcuni dati sono mancanti .

Dati  $C$  gruppi di variabili risposta scambiabili  $q$ -dimensionali

$$\mathbf{X}_j = \{\mathbf{X}_{ji} = (X_{hji}, h = 1, \dots, q), i = 1, \dots, n_j, j = 1, \dots, C\}$$

con funzione di distribuzione  $P_j$ ,  $\mathbf{X}_{ji} \in \mathcal{R}^q$ , dove  $n = \sum_j n_j$  è l'ampiezza del campione, e il sistema d'ipotesi è dato da

$$\begin{aligned} H_0 &= \{P_1 = \dots = P_C = P\} = \left\{ \mathbf{X}_1 \stackrel{d}{=} \dots \stackrel{d}{=} \mathbf{X}_C \right\} \\ H_1 &= \{H_0 \text{ non è vera} \} \end{aligned} \quad (4.5)$$

Assumiamo che i dati, siano essi effettivamente osservati o mancanti, siano scambiabili rispetto ai  $C$  gruppi sotto l'ipotesi nulla. Consideriamo appropriate trasformazioni dei dati, che indichiamo con  $\varphi_h$ ,  $h = 1, \dots, q$ , che siano specifiche per ogni variabile. In modo particolare utilizziamo la trasformazione media campionaria come indicatore appropriato degli effetti del trattamento. L'insieme dei dati può essere così rappresentato:

$$\mathbf{Y} = \{Y_{hji} = \varphi_h(X_{hji}), i = 1, \dots, n_j, j = 1, \dots, C, h = 1, \dots, q\} \quad (4.6)$$

e i test di permutazione parziali possono essere costruiti utilizzando appropriate funzioni del totale del campione, ossia  $S_{hj}^* = \sum_i Y_{hji}^*$ , con  $j = 1, \dots, C$  e  $h = 1, \dots, q$ . Consideriamo anche un vettore di variabili indicatrici della presenza o meno dell'osservazione, ovvero

$$\mathbf{O} = \{O_{hji}, i = 1, \dots, n_j, j = 1, \dots, C, h = 1, \dots, q\} \quad (4.7)$$

in cui  $O_{hji} = 1$  se  $X_{hji}$  stato osservato e  $O_{hji} = 0$  altrimenti. Il vettore  $\mathbf{O}$  rappresenta la configurazione osservata del data set. L'insieme dei dati osservati può dunque essere riscritto come  $(\mathbf{Y}, \mathbf{O})$ . Definiamo dunque l'effettiva dimensione del campione con  $\nu_{hj} = \sum_i O_{hji}$ , dove  $j = 1, \dots, C$  e  $h = 1, \dots, q$ . Se si suppone dunque che, sotto  $H_0$ , i dati siano congiuntamente scambiabili rispetto ai gruppi su entrambe le variabili  $\mathbf{Y}$  e  $\mathbf{O}$ , il sistema d'ipotesi (4.5) può essere così riformulato:

$$\begin{aligned} H_0 &= \left\{ (\mathbf{Y}_1, \mathbf{O}_1) \stackrel{d}{=} \dots \stackrel{d}{=} (\mathbf{Y}_C, \mathbf{O}_C) \right\} \\ H_1 &= \{H_0 \text{ non è vera}\} \end{aligned} \quad (4.8)$$

In questo contesto, sempre sotto  $H_0$ ,  $P$  diventa la distribuzione congiunta multivariata  $(\mathbf{Y}_j, \mathbf{O}_j)$  con  $j = 1, \dots, C$  e si può anche scrivere che  $P = P_{\mathbf{O}} \cdot P_{\mathbf{Y}|\mathbf{O}}$  enfatizzando così la variabile indicatrice dell'inclusione nel campione e la variabile trasformazione dei dati osservati. La coppia  $(\mathbf{Y}, \mathbf{O})$  rappresentante l'intero insieme dei dati osservati, sotto l'ipotesi nulla, è congiuntamente un insieme di statistiche sufficienti per la distribuzione sottostante i dati osservati e il processo generatori dei dati mancanti. Se ci sono  $q$  differenti aspetti da porre sotto ipotesi, ci saranno anche  $q$  test parziali con la loro relativa combinazione non parametrica. Fatta questa considerazione,

il sistema d'ipotesi (4.5) può essere riscritto come

$$\begin{aligned} H_0 & : \left\{ \bigcap_{h=1}^q \left[ (Y_{h1}, O_{h1}) \stackrel{d}{=} \dots \stackrel{d}{=} (Y_{hC}, O_{hC}) \right] \right\} = \left\{ \bigcap_{h=1}^q H_{0h} \right\} \\ H_1 & = \left\{ \bigcup_h H_{1h} \right\} \end{aligned} \quad (4.9)$$

in cui  $H_0$  è vera se tutte le  $H_{0h}$  sono congiuntamente vere, mentre  $H_1$  è vera se almeno una  $H_{1h}$  è vera. Per testare  $H_0$  contro  $H_1$ , consideriamo un vettore  $q$ -dimensionale di statistiche test  $\mathbf{T} = \{T_1, \dots, T_C\}$ , la  $h$ -esima componente del quale è la statistica test per verificare l' $h$ -esima sottoipotesi  $H_{0h}$ . Assumiamo che i test parziali siano non degeneri, marginalmente non distorti, consistenti, significativi per valori grandi. Di qui il test combinato è funzione dei  $q$  test parziali dipendenti. Proprio per la presenza di relazioni di dipendenza sottostanti, la combinazione deve essere condotta via non parametrica.

## 4.4 Modelli non-MAR

Se i *missing data* sono non-MAR (*Missing Not at Random*), allora l'ipotesi nulla  $H_0$  nel verificare l'uguaglianza in distribuzione dei  $C$  gruppi deve prendere in considerazione il vettore dei dati osservati e raccolti  $\mathbf{Y}$  congiuntamente al vettore associato d'inclusione  $\mathbf{O}$ , in quanto sotto l'ipotesi alternativa il trattamento può influenzare anche il processo che causa i *missing data*. Di qui l'ipotesi nulla richiede l'uguaglianza in distribuzione del processo generatore dei dati mancanti nei  $C$  gruppi, espresso da  $\mathbf{O}$ , e delle variabili risposta  $\mathbf{Y}$  condizionatamente ad  $\mathbf{O}$ , cioè:

$$H_0 : \left\{ \left[ \mathbf{O}_1 \stackrel{d}{=} \dots \stackrel{d}{=} \mathbf{O}_C \right] \cap \left[ \left( \mathbf{Y}_1 \stackrel{d}{=} \dots \stackrel{d}{=} \mathbf{Y}_C \right) \mid \mathbf{O} \right] \right\} \quad (4.10)$$

L'assunzione di scambiabilità, sotto l'ipotesi nulla, degli  $n$  vettori  $(\mathbf{Y}, \mathbf{O})$ , con rispetto ai  $C$  gruppi, implica che gli effetti del trattamento sono nulli per tutte le variabili, osservate e non osservate. In altre parole, si assume che non vi sia differenza in distribuzione per la variabile d'inclusione multivariata  $\mathbf{O}_j$ ,  $j = 1, \dots, C$ , e condizionatamente a  $\mathbf{O}$ , per le variabili osservate presenti in  $\mathbf{Y}$ . Di conseguenza non è necessario specificare il processo generatore dei *missing data* e la distribuzione delle osservazioni, purchè siano a disposizione test parziali marginalmente non distorti. In particolare dato che la soluzione è non parametrica, non è necessario specificare le relazioni di dipendenza

sottostanti la distribuzione congiunta di  $(\mathbf{Y}, \mathbf{O})$ . L'ipotesi nulla globale può essere scomposta in  $2q$  sottoipotesi, ovvero:

$$\begin{aligned} H_0 & : \left\{ \left[ \cap_h \left( O_{h1} \stackrel{d}{=} \dots \stackrel{d}{=} O_{hC} \right) \right] \cap \left[ \cap_h \left( Y_{h1} \stackrel{d}{=} \dots \stackrel{d}{=} Y_{hC} \right) \mid \mathbf{O} \right] \right\} \quad (4.11) \\ & = \left\{ H_0^{\mathbf{O}} \cap H_0^{\mathbf{Y} \mid \mathbf{O}} \right\} = \left\{ (\cap_h H_{0h}^{\mathbf{O}}) \cap (\cap_h H_{0h}^{\mathbf{Y} \mid \mathbf{O}}) \right\} \end{aligned}$$

contro

$$H_1 = \left\{ (\cup_h H_{1h}^{\mathbf{O}}) \cup (\cup_h H_{1h}^{\mathbf{Y} \mid \mathbf{O}}) \right\}$$

dove  $H_{0h}^{\mathbf{O}}$  indica l'uguaglianza in distribuzione, rispetto ai  $C$  livelli, dell' $h$ -esima componente marginale del processo d'inclusione nel campione, ossia  $\mathbf{O}$ , e  $H_{0h}^{\mathbf{Y} \mid \mathbf{O}}$  indica l'uguaglianza in distribuzione dell' $h$ -esima componente di  $\mathbf{Y}$ , condizionatamente ad  $\mathbf{O}$ . Una statistica test di permutazione per ciascuna delle  $q$  sottoipotesi  $H_{0h}^{\mathbf{O}}$  può essere la statistica  $X^2$ . In alternativa si possono utilizzare altre statistiche appropriate per variabili categoriali, come quella di Anderson-Darling, oppure il test esatto di Fisher multivariato (*Fisher's exact multivariate probability test*), se le variabili categoriali sono ordinate.

## 4.5 Combinazione non parametrica

La combinazione non parametrica dei  $2q$  test parziali può essere fatta in almeno tre modi diversi:

1. considerare un'unica funzione di combinazione su tutti i  $2q$  test parziali, del tipo  $T_a'' = \psi(\lambda_1^{\mathbf{O}}, \dots, \lambda_q^{\mathbf{O}}; \lambda_1^{\mathbf{Y} \mid \mathbf{O}}, \dots, \lambda_q^{\mathbf{Y} \mid \mathbf{O}})$ ;
2. considerare  $q$  combinazioni, per ciascuna variabile componente, del secondo ordine  $T_{bh}'' = \psi_h(\lambda_h^{\mathbf{O}}; \lambda_h^{\mathbf{Y} \mid \mathbf{O}})$ ,  $h = 1, \dots, q$  e poi una combinazione del terzo ordine  $T_b''' = \psi(\lambda_{b1}'', \dots, \lambda_{bq}'')$ ;
3. considerare due combinazioni del secondo ordine,  $T_{c\mathbf{O}}'' = \psi_{\mathbf{O}}(\lambda_1^{\mathbf{O}}, \dots, \lambda_q^{\mathbf{O}})$  e  $T_{c\mathbf{Y} \mid \mathbf{O}}'' = \psi_{\mathbf{Y} \mid \mathbf{O}}(\lambda_1^{\mathbf{Y} \mid \mathbf{O}}, \dots, \lambda_q^{\mathbf{Y} \mid \mathbf{O}})$ , rispettivamente sul vettore delle variabili indicatrici dell'inclusione, e sul vettore di dati effettivamente osservati  $(\mathbf{Y}, \mathbf{O})$ , seguite da una combinazione del terzo ordine  $T_c''' = \psi(T_{c\mathbf{O}}'', T_{c\mathbf{Y} \mid \mathbf{O}}'')$



E' da notare che in tutte le fasi e in ciascuno dei tre metodi viene utilizzata la stessa funzione  $\psi$  e  $T''_a, T'''_b, T''''_c$  sono quasi permutazionalmente equivalenti, eccetto per le approssimazioni dovute a simulazioni finite e alla non linearità delle funzioni di combinazione.

## 4.6 Modelli MCAR

Se i *missing data* sono MCAR (*Missing Completely at Random*) si procede condizionandosi al vettore indicante l'inclusione osservato, cioè  $\mathbf{O}$ , e si ignora  $H_0^{\mathbf{O}}$  perché si assume che  $\mathbf{O}$  non fornisca alcuna informazione sugli effetti del trattamento. Essendo, per assunto, vere le sottoipotesi su  $\mathbf{O}$ , l'ipotesi  $H_0^{\mathbf{O}} : \{\mathbf{O}_1 \stackrel{d}{=} \dots \stackrel{d}{=} \mathbf{O}_C\}$  può essere ignorata. L'ipotesi nulla può essere riscritta in modo più semplice nella forma

$$H_0 = H_0^{\mathbf{Y}|\mathbf{O}} : \left\{ \bigcap_h \left[ (\mathbf{Y}_{h1} \stackrel{d}{=} \dots \stackrel{d}{=} \mathbf{Y}_{hC}) | \mathbf{O} \right] \right\} = \left\{ \bigcap_h H_{0h}^{\mathbf{Y}|\mathbf{O}} \right\} \quad (4.12)$$

contro

$$H_1^{\mathbf{Y}|\mathbf{O}} = \left\{ \bigcup_h H_{1h}^{\mathbf{Y}|\mathbf{O}} \right\}$$

in cui la notazione è la stessa di quella utilizzata nel paragrafo precedente. Questo problema può essere risolto da una combinazione non parametrica del tipo  $T''' = T'''_{c|\mathbf{Y}|\mathbf{O}} = \psi_{\mathbf{Y}}(\lambda_1^{\mathbf{Y}|\mathbf{O}}, \dots, \lambda_q^{\mathbf{Y}|\mathbf{O}})$ .

## 4.7 Permutazioni di *missing data*: struttura e analisi

Per capire meglio la soluzione non parametrica di permutazione in presenza di dati mancanti, è necessario considerare il ruolo degli indicatori d'inclusione permutati  $\mathbf{O}^* = \{O_{hji}^*, i = 1, \dots, n_j, j = 1, \dots, C, h = 1, \dots, q\}$  con particolare attenzione al numero di dati mancanti in tutti i punti dello spazio di permutazione  $(\mathcal{Y}, \mathcal{O})_{/(\mathbf{Y}, \mathbf{O})}$  associato alla coppia  $(\mathbf{Y}, \mathbf{O})$ . All'interno di ogni gruppo, con riferimento a ciascuna variabile componente, la dimensione effettiva del campione di permutazione, indicata con  $\nu_{hj}^* = \sum_i O_{hji}^*$ ,  $i = 1, \dots, n_j, j = 1, \dots, C, h = 1, \dots, q$ , varia a seconda dell'attribuzione casuale delle unità, e dei relativi *missing data*, nei  $C$  gruppi. Questo perché anche

le unità con *missing data* sono incluse nel meccanismo delle permutazioni ([36]; [37]; [38]; [39]; [40]; [41]).

Consideriamo un modello MCAR e un vettore di statistiche test parziali  $\mathbf{T}$  basate su funzioni dei totali campionari dei dati validi e indichiamo la sua distribuzione di permutazione multivariata con  $F[\mathbf{t} | (\mathbf{Y}, \mathbf{O})]$ ,  $\mathbf{t} \in \mathcal{R}^q$ . Osserviamo che l'insieme  $\mathbf{O}^*$  delle permutazioni di  $\mathbf{O}$ , cioè l'insieme delle possibili realizzazioni delle configurazioni dei dati attribuiti casualmente ai  $C$  gruppi, induce una partizione in sotto-orbite sull'intero spazio di permutazione  $(\mathcal{Y}, \mathcal{O})_{/(\mathbf{Y}, \mathbf{O})}$  associato a  $(\mathbf{Y}, \mathbf{O})$ . Queste sotto-orbite sono costituite dai punti (permutazioni) che hanno lo stesso numero  $\{\nu_{hj}^*, j = 1, \dots, C, h = 1, \dots, q\}$  di dati validi nel campione di permutazione. Per esempio due punti  $(\mathbf{Y}_1^*, \mathbf{O}_1^*)$  e  $(\mathbf{Y}_2^*, \mathbf{O}_2^*)$  stanno nello stessa sotto-orbita se le dimensioni dei rispettivi campioni di permutazione  $\nu_{1hj}^* = \sum_i O_{1hji}^*$  e  $\nu_{2hj}^* = \sum_i O_{2hji}^*$  sono uguali per ogni  $h$  e  $j$ , con  $h = 1, \dots, q$  e  $j = 1, \dots, C$ . Se le sotto distribuzioni di permutazione di  $\mathbf{T}$  fossero invarianti rispetto alle sotto-orbite indotte da  $\mathbf{O}^*$  potremmo stimare  $F[\mathbf{t} | (\mathbf{Y}, \mathbf{O})]$  con una procedura di simulazione *CMC*, ossia tramite il condizionamento ai dati del campione, ignorando tutte le sotto-orbite. L'invarianza rispetto alla configurazione di permutazione dei dati richiede che l'uguaglianza

$$F[\mathbf{t} | (\mathbf{Y}, \mathbf{O})] = F[\mathbf{t} | (\mathbf{Y}, \mathbf{O}^*)] \quad (4.13)$$

sia soddisfatta per ogni  $\mathbf{t} \in \mathcal{R}^q$ , per ogni specifica permutazione  $\mathbf{O}^*$  di  $\mathbf{O}$  e per tutti gli insiemi di dati  $\mathbf{Y}$ . Quando  $q > 1$  o la dimensione del campione è finita, è piuttosto difficile conseguire in maniera esatta questa invarianza in distribuzione. Quando  $q = 1$ , nei problemi ad una dimensione, questa invarianza in distribuzione può essere ottenuta in maniera esatta nei modelli MCAR perché, tramite condizionamento, è possibile ignorare la mancanza di dati eliminando dall'insieme dei dati tutte le unità non osservate. Come si vedrà in seguito, quando si hanno campioni con osservazioni multivariate appaiate, così che le differenze mancanti sono quantità permutazionalmente invarianti, i relativi test diventano esatti. In generale si utilizzano soluzioni approssimate.

## 4.8 Soluzione per problemi MCAR

Indichiamo con  $\nu = \{\nu_{hj}, j = 1, \dots, C, h = 1, \dots, q\}$  la matrice  $q \times C$  del numero effettivo di dati validi nel vettore indicatore dell'inclusione osservato  $\mathbf{O}$ . Assumiamo che le statistiche test siano basate sui totali di dati validi nel campione di permutazione  $\{W_{hj}^* = \sum_i Y_{hji}^* \cdot O_{hji}^*, j = 1, \dots, C, h = 1, \dots, q\}$ . Considerando che la distribuzione di  $W_{hj}^*$ , dei totali campionari, condizionali all'intero data set  $(\mathbf{Y}, \mathbf{O})$  considerato finito, dipende dal numero effettivo di addendi  $\nu_{hj}^*$ , l'uguaglianza (4.13) diventa quindi

$$F[\mathbf{t} | (\mathbf{Y}, \nu)] = F[\mathbf{t} | (\mathbf{Y}, \nu^*)] \quad (4.14)$$

dove  $\nu^* = \{\nu_{hj}^*, j = 1, \dots, C, h = 1, \dots, q\}$  rappresenta la matrice  $q \times C$  del numero effettivo, nel campione di permutazione, di dati validi associati a  $\mathbf{O}^*$ . Trovare statistiche test che abbiano sotto-distribuzioni nulle invarianti rispetto a  $\nu^*$  e a tutti i dati  $\mathbf{Y}$  è molto difficile e le condizioni (4.13) e (4.14) sono raramente soddisfatte in maniera esatta. Tuttavia una soluzione, basata sull'invarianza in media e varianza del vettore di statistiche test  $\mathbf{T}$  rispetto a  $\nu^*$ , può essere data in maniera approssimata. In altre parole si devono cercare statistiche, che abbiano medie e varianze invarianti rispetto alle sotto-orbite indotte da  $\mathbf{O}^*$  sullo spazio di permutazione  $(\mathcal{Y}, \mathcal{O})_{/(\mathbf{Y}, \mathbf{O})}$ . Senza perdita di generalità, consideriamo il caso di una variabile univariata  $Y$ , così da avere soltanto una statistica test  $T$ , e assumiamo che i test di permutazione considerati siano basati sui totali campionari univariati di dati validi,  $W_j^* = \sum_i Y_{ji}^* \cdot O_{ji}^*, j = 1, \dots, C$ . Il totale dei totali, ovvero  $W = \sum_j W_j$ , quantità non nulla, è permutazionalmente invariante perché in  $(\mathcal{Y}, \mathcal{O})_{/(\mathbf{Y}, \mathbf{O})}$  l'equazione

$$W = \sum_i Y_{ji} \cdot O_{ji} = \sum_j W_j^*$$

è sempre soddisfatta. Consideriamo il caso a due campioni  $C = 2$  e assumiamo che il test per verificare  $H_0^{\mathbf{Y}|\mathbf{O}}$  contro  $H_1^{\mathbf{Y}|\mathbf{O}}$  sia una combinazione lineare di  $W_1^*$  e  $W_2^*$ . Così il test è espresso nella forma

$$T^*(a^*, b^* | \nu^*) = a^* \cdot W_1^* - b^* \cdot W_2^*$$

dove  $a^*$  e  $b^*$  sono due coefficienti indipendenti dai dati effettivamente osservati  $\mathbf{Y}$ , ma non necessariamente permutazionalmente invarianti. Questi

coefficienti devono essere determinati assumendo che, sotto l'ipotesi nulla, la varianza  $\mathbb{V}[T^*(a^*, b^*)|\nu^*] = \zeta^2$  è costante, ovvero indipendente dalle effettive ampiezze campionarie, a permutazione avvenuta,  $\nu_j^*$ ,  $j = 1, 2$ , e che la media  $\mathbb{E}[T^*(a^*, b^*)|\nu^*] = 0$ . In accordo con le tecniche di campionamento casuale senza ripetizione da  $(\mathbf{Y}, \mathbf{O})$ , considerato finito, possiamo scrivere le seguenti equazioni:

$$\nu_1^* + \nu_2^* = \nu$$

$$W_1^* + W_2^* = W$$

$$\mathbb{E}(W_j^*) = W \cdot \nu_j^*/\nu, \quad j = 1, 2$$

$$\mathbb{V}(W_j^*) = \sigma^2 \cdot \nu_j^*(\nu - \nu_j^*)/(\nu - 1) = \mathbb{V}(\nu^*), \quad j = 1, 2$$

dove  $V$  è una funzione positiva e  $\nu = \nu_1 + \nu_2$ ,  $W = W_1 + W_2$  e  $\sigma^2 = \sum_{ji} (Y_{ji} - W/\nu)^2 \cdot O_{ji}/\nu$  sono quantità non nulle permutazionalmente invarianti. Per ciascuna coppia  $(\nu_1^*, \nu_2^*)$  indicante la numerosità, positiva, nel campione di permutazione, i due totali nel campione di permutazione  $W_1^*$  e  $W_2^*$  hanno la stessa varianza e il loro coefficiente di correlazione  $\rho(W_1^*, W_2^*) = -1$  perché la loro somma  $S$  è una quantità permutazionalmente invariante. Di qui, possiamo scrivere:

$$\mathbb{E}[T^*(a^*, b^*)] = a^* \cdot W \cdot \nu_1^* - b^* \cdot W \cdot \nu_2^* = 0$$

$$\mathbb{V}[T^*(a^*, b^*)] = a^{*2}V(\nu^*) + 2a^*b^*V(\nu^*) + b^{*2}V(\nu^*) = (a^* + b^*)^2V(\nu^*)$$

Le soluzioni di queste equazioni sono  $a^* = (\nu_2^*/\nu_1^*)^{1/2}$  e  $b^* = (\nu_1^*/\nu_2^*)^{1/2}$ , ignorando un irrilevante coefficiente positivo. Per  $C = 2$  e  $q = 1$  le statistiche test, invarianti in media e varianza, sono del tipo:

$$T^* = W_1^* \cdot (\nu_2^*/\nu_1^*)^{1/2} - W_2^* (\nu_1^*/\nu_2^*)^{1/2} \quad (4.15)$$

Quando non ci sono *missing value* così che  $\nu_j^* = n_j^*$ ,  $j = 1, 2$  la (4.15) è permutazionalmente equivalente al test per il confronto tra medie ovvero  $T^* \approx \sum_i Y_{1i}^*$ . Per dare una soluzione ben definita si assume che  $\nu_1^*$  e  $\nu_2^*$  siano congiuntamente positivi. Questo implica che si devono eliminare dallo spazio di permutazione  $(\mathcal{Y}, \mathcal{O})_{/(\mathbf{Y}, \mathbf{O})}$  tutti quei punti in cui anche un singolo componente della matrice di permutazione  $\nu^*$  è zero. Questa restrizione non ha effetti sulle conclusioni inferenziali.

Nel caso in cui si abbiano  $C > 1$  campioni e  $q = 1$ , una soluzione approssimata è data da

$$T_C^* = \sum_{j=1}^C \left\{ W_j^* \cdot [(\nu - \nu_j^*)/\nu_j^*]^{1/2} - (W - W_j^*) \cdot [\nu_j^*/(\nu - \nu_j^*)]^{1/2} \right\}^2 \quad (4.16)$$

Questa statistica test può essere vista come una combinazione diretta non parametrica (vedi 2.7) di  $C$  test parziali dipendenti, ciascuno ottenuto da un confronto, fatto via permutazione, tra il  $j$ -esimo campione e tutti i restanti  $C - 1$ , raggruppati insieme. Se  $q > 1$ , la soluzione sarà una combinazione non parametrica e per testare  $H_0 : \left\{ \cap_h H_{0h}^{\mathbf{Y}|\mathbf{O}} \right\}$  contro  $H_1 : \left\{ \cup_h H_{1h}^{\mathbf{Y}|\mathbf{O}} \right\}$  la soluzione diventa  $T'' = \psi(\lambda_1, \dots, \lambda_q)$  dove  $\psi \in \mathcal{C}$ ,  $\lambda_h$  è il p-value di

$$T_{Ch}^* = \sum_{j=1}^C \left\{ W_{hj}^* \cdot \left( \frac{\nu_h - \nu_{hj}^*}{\nu_{hj}^*} \right)^{1/2} - (W_h - W_{hj}^*) \cdot \left( \frac{\nu_{hj}^*}{\nu_h - \nu_{hj}^*} \right)^{1/2} \right\}^2$$

ciascuno relativo all' $h$ -esima variabile,  $h = 1, \dots, q$ .

### 4.8.1 Estensione ai modelli non-MAR

In un modello non-MAR, sempre per ottenere una soluzione non parametrica, si devono anche combinare le  $q$ -statistiche test fatte sulle componenti del vettore  $\mathbf{O}$  indicatore dell'inclusione nel campione, purché tutti i test siano marginalmente non distorti. Per testare il sistema d'ipotesi (4.11) si devono combinare i  $q$ -test  $T_h^{\mathbf{O}*}$  e i  $q$ -test  $T_h^{\mathbf{Y}|\mathbf{O}*}$ ,  $h = 1, \dots, q$ . Di qui il test combinato è del tipo

$$T'' = \psi(\lambda_1^{\mathbf{O}}, \dots, \lambda_q^{\mathbf{O}}; \lambda_1^{\mathbf{Y}|\mathbf{O}}, \dots, \lambda_q^{\mathbf{Y}|\mathbf{O}})$$

Per ciascuna delle  $q$  sotto-ipotesi  $H_{0h}^{\mathbf{O}}$  contro  $H_{1h}^{\mathbf{O}}$ , una statistica test appropriata può essere il  $\chi^2$  di Pearson o, comunque, si può usare una qualsiasi altra statistica test per dati categoriali. Per le  $q$ -sottoipotesi  $H_{0h}^{\mathbf{Y}|\mathbf{O}}$  contro  $H_{1h}^{\mathbf{Y}|\mathbf{O}}$  si rimanda al paragrafo precedente. I test parziali  $T_h^{\mathbf{O}*}$  sulle componenti di  $\mathbf{O}$  sono esatti, non distorti e consistenti, mentre i test  $T_h^{\mathbf{Y}|\mathbf{O}*}$  sulle componenti di  $\mathbf{Y}$  sono consistenti, ma sono esatti e non distorti solo approssimativamente, così che il test  $T''$  è consistente e approssimativamente esatto e non distorto per tutte le funzioni  $\psi \in \mathcal{C}$ .

## 4.9 Dati appaiati e *missing value*

Una variabile reale  $\mathbf{X}$ ,  $q$ -dimensionale e non degenere, viene osservata in  $k$  occasioni temporali diverse sulle stesse  $n$  unità in due situazioni sperimentali, corrispondenti a due livelli di un simbolico trattamento. I dati vengono così rappresentati:

$$\begin{aligned} \mathbf{X} &= \{X_{hjit}, t = 1, \dots, k, i = 1, \dots, n, j = 1, 2, h = 1, \dots, q\} \quad (4.17) \\ &= \{\mathbf{X}_{jit}, t = 1, \dots, k, i = 1, \dots, n, j = 1, 2\} \\ &= \{\mathbf{X}_{ji}, i = 1, \dots, n, j = 1, 2\} \end{aligned}$$

dove  $\mathbf{X}$  rappresenta la risposta  $q$ -variata  $(X_1, \dots, X_q)$ , l'andamento temporale  $q$ -variato  $(\mathbf{X}_1, \dots, \mathbf{X}_k)$  e l'intero insieme dei dati. Supponiamo che ci siano alcuni *missing data* e quindi che ci sia anche un vettore indicatore d'inclusione

$$\mathbf{O} = \{O_{hjit}, t = 1, \dots, k, i = 1, \dots, n, j = 1, 2, h = 1, \dots, q\}$$

in cui  $O_{hjit} = 1$  se  $X_{hjit}$  è osservato e  $O_{hjit} = 0$  se  $X_{hjit}$  è un *missing value*. Supponiamo di voler verificare se il trattamento ha effetto, senza curarci del tempo, delle dipendenze sottostanti e dell'ignota distribuzione. Se i dati sono *missing* secondo un modello MCAR, le ipotesi da testare sono formalmente espresse con

$$H_0 : \left\{ \bigcap_{t=1}^k (\mathbf{X}_{1t} \stackrel{d}{=} \mathbf{X}_{2t}) \mid \mathbf{O} \right\} = \left\{ \bigcap_{t=1}^k \bigcap_{q=1}^h H_{0ht} \right\}$$

verso

$$H_1 : \left\{ \bigcup_{t=1}^k (\mathbf{X}_{1t} \stackrel{d}{\neq} \mathbf{X}_{2t}) \mid \mathbf{O} \right\} = \left\{ \bigcup_{t=1}^k \bigcup_{q=1}^h H_{1ht} \right\}$$

La statistica test di permutazione assume la forma

$$T_{ht}^* = \varphi_{ht} \left[ \frac{\sum_i Y_{hit} \cdot O_{hit} \cdot S_i^*}{(\sum_i Y_{hit}^2 \cdot O_{hit}^2)^{1/2}} \right], \quad t = 1, \dots, k, \quad h = 1, \dots, q \quad (4.18)$$

dove  $Y_{hit} = X_{h1it} - X_{h2it}$  sono le differenze osservate unità per unità, tempo per tempo e variabile per variabile,  $O_{hit} = O_{h1it} \cdot O_{h2it}$  vale 1 se e soltanto se  $X_{h1it}$  e  $X_{h2it}$  sono entrambe osservate,  $S_i^*$  sono i segni di permutazione e  $\varphi_{ht}$  sono le funzioni valore assoluto o segno  $+/-$  a seconda che l' $h$ -esima alternativa  $H_{1ht}$  di interesse sia a due o ad una coda.

Tutti i test parziali  $T_{ht}^*$ ,  $t = 1, \dots, k$ ,  $h = 1, \dots, q$  sono permutazionalmente esatti, purchè le dimensioni effettive del campione  $\nu_{ht} = \sum_i O_{hit}$  siano positive, in quanto i  $\nu_{ht}$  sono quantità permutazionalmente invarianti e la distribuzione nulla dipende solo dalla scambiabilità degli errori. Questi hanno media di permutazione nulla e varianza di permutazione unitaria e sono marginalmente non distorti. Di qui la combinazione non parametrica è una soluzione adeguata per i test multipli. Tutti i test sono espressi in forma standardizzata così è possibile applicare una funzione di combinazione diretta e si ha  $T_h^{''*} = \sum_t T_{ht}^*$  per i test entro ciascuna variabile senza curarsi del tempo, e poi un test globale  $T''' = \sum_h T_h^{''*}$ . L'estensione ai modelli non-MAR è semplice, purchè i test parziali per le componenti del vettore indicatore d'inclusione  $\mathbf{O}$  siano marginalmente non distorti. Se i test parziali  $H_0^{\mathbf{O}} : \{\cap_h H_{0h}^{\mathbf{O}}\}$  contro  $H_1^{\mathbf{O}} : \{\cup_h H_{1h}^{\mathbf{O}}\}$  sono esatti, allora la combinazione non parametrica per l'ipotesi

$$H_0^{\mathbf{O}} : \left\{ \left[ \cap_h H_{0h}^{\mathbf{O}} \right] \cap \left[ \cap_h H_{0h}^{\mathbf{Y}|\mathbf{O}} \right] \right\}$$

contro

$$H_1^{\mathbf{O}} : \left\{ \left[ \cup_h H_{1h}^{\mathbf{O}} \right] \cup \left[ \cup_h H_{1h}^{\mathbf{Y}|\mathbf{O}} \right] \right\}$$

è pure esatta.

## 4.10 Intention to treat analysis

La valutazione dei risultati finali ottenuti in un trial clinico risulta problematica quando non vi è una completa adesione allo studio da parte dei pazienti. Molti autori si sono dedicati alla ricerca di un buon metodo che prevenga la non adesione e consenta di formulare stime corrette, dal punto di vista statistico, dell'efficacia di un trattamento. Uno di questi metodi è quello dell'ITT (*Intention-To-Treat*). Secondo il principio dell'ITT la popolazione di analisi (*analysis population*) è costituita da tutti i pazienti randomizzati, inclusi quelli che hanno interrotto lo studio prematuramente (e quindi hanno alcune valutazioni mancanti o, addirittura, sono usciti dallo studio senza aver mai ricevuto il trattamento) e quelli che presentano delle violazioni ai criteri di selezione (per cui non sarebbero dovuti entrare nello studio). Il termine, coniato da Hill (1961) [42], è stato più tardi definito come l'analisi che

*“Includes all randomized patients in the groups to which they were randomly assigned, regardless of their adherence with the entry criteria, regardless of the treatment they actually received, and regardless of subsequent withdrawal from treatment or deviation from the protocol.”* [43]

Altra soluzione al problema in questione è quella che si basa sul principio del TR (*Treatment Received*): l'idea di fondo è che i pazienti devono essere analizzati a seconda del trattamento effettivamente ricevuto, senza badare al trattamento cui erano stati, inizialmente, randomizzati.

Un'altra proposta di analisi è rappresentata dalla popolazione *per protocol* in cui vengono inseriti nello studio solo i pazienti che hanno rispettato i principali criteri di selezione e le principali procedure del protocollo, ovvero quei pazienti che hanno assunto il trattamento per la durata e nelle quantità previste, si sono presentati a tutte le visite e non hanno fatto uso, in contemporanea al trattamento, altri farmaci considerati proibiti. I criteri perché un paziente venga considerato sufficientemente aderente (*compliant*) al protocollo vanno definiti a priori nel protocollo stesso. Un'analisi di questo tipo ha come obiettivo quello di verificare qual è l'effetto del trattamento se assunto nelle migliori condizioni

Supponiamo, in un trial clinico, di avere dei pazienti randomizzati o al trattamento  $A$  o al trattamento  $B$  e supponiamo ancora che, durante lo studio, alcuni dei pazienti randomizzati ad  $A$  ricevano di fatto  $B$  (gruppo  $B_A$ ), e alcuni pazienti inizialmente randomizzati al gruppo  $B$  ricevano  $A$  (gruppo  $A_B$ ). Secondo il principio dell'ITT, l'analisi deve essere effettuata mettendo a confronto il gruppo  $A + B_A$  con il gruppo  $B + A_B$ , coerentemente con la randomizzazione iniziale, mentre la scelta di un approccio basato sull'TR implica che l'analisi sia condotta sui gruppi  $A + A_B$  e  $B + B_A$  coerentemente con il trattamento effettivamente ricevuto. E' ancora aperto il dibattito su quale delle due filosofie, sottostanti i due diversi approcci di analisi dei risultati ottenuti nei trial clinici, conduca a dei risultati corretti. In accordo con gli statistici, per lo meno nelle analisi preliminari, è da preferirsi un approccio di tipo ITT: infatti la circostanza in cui l'ITT produce risultati migliori è quella in cui l'obiettivo dello studio è di testare l'ipotesi nulla che non vi siano differenze tra i due trattamenti. Un'analisi di tipo TR ha senso invece se lo studio in questione è la componente base di un più ampio disegno sperimentale. Tuttavia schierarsi in modo rigido a favore di uno o dell'altro



approccio può portare a conclusioni errate.

La discussione sarà incentrata sul primo dei tre approcci. L'analisi basata sul principio dell'*intention-to-treat* non altera gli effetti della randomizzazione e garantisce pertanto confronti liberi da bias di selezione e di valutazione. Questo è il motivo per cui viene considerata primaria in tutti i contesti, a partire da quello regolatorio [16]. Oltre a questo l'analisi sulla popolazione *intention-to-treat* serve a valutare il trattamento sperimentale nelle condizioni più simili a quelle che si verificheranno una volta che il trattamento verrà commercializzato. Infatti il farmaco potrebbe finire, anche a causa di una diagnosi errata, nelle mani di pazienti di tipologia diversa rispetto a quelli sui cui era stata testata e provata la sua efficacia, potrebbe essere utilizzato per un periodo diverso, più breve o più lungo, rispetto a quello prescritto o essere somministrato in concomitanza a farmaci sconsigliati.

Si sono sviluppate e diffuse alcune varianti della definizione dell'*intention-to-treat* sopra riportata: una di queste stabilisce l'esclusione di quei pazienti che non hanno assunto neppure una dose del farmaco in studio e/o quelli senza alcuna valutazione post-basale. Queste varianti devono sempre tener conto del fatto che i pazienti analizzati devono rappresentare la quasi totalità dei pazienti randomizzati (i pazienti esclusi devono essere non più del 5%). Per quanto riguarda la valutazione dei pazienti che non hanno una rilevazione a fine studio o nelle visite di valutazione primaria, il metodo più diffuso a tal fine è quello di valutare questi soggetti con la loro ultima misurazione disponibile, ovvero con il metodo chiamato *last observation carried forward* (LOCF) oppure quello di valutare i pazienti secondo delle regole prestabilite e dipendenti dalla ragione di uscita dallo studio: per esempio, se un paziente esce dallo studio per insoddisfazione, inefficacia o evento avverso serio riceve il peggior punteggio possibile; ad un paziente uscito per una ragione indipendente dal trattamento viene assegnato un punteggio corrispondente alla media del gruppo cui questo paziente appartiene; se, infine, la causa dell'abbandono del trattamento è la scomparsa della sintomatologia al paziente viene assegnato il miglior punteggio possibile. Tra i fattori che alimentano il dibattito sulla validità o meno dell'impiego di una strategia ITT, degno di nota è la distorsione. Si consideri uno studio in cui i pazienti sono randomizzati per ricevere un trattamento, di cui si vuole testare l'efficacia nella riduzione del tasso di mortalità dovuta a cirrosi epatica. All'inizio, i pazienti sono rando-

mizzati per ricevere o la terapia del gruppo di controllo o il trattamento e, alla fine dello studio, si dovranno confrontare i tassi di mortalità dei pazienti in ciascun gruppo. L'ITT viene impiegato come espediente per circuire il problema legato ai *missing data* e procedere nell'analisi dei risultati. Un'analisi basata sul principio dell'ITT è senza dubbio conservativa, infatti non altera la randomizzazione iniziale, e prudentiale, perché prende in considerazione la possibilità che non tutti i pazienti seguano completamente le procedure previste dal protocollo specifico del gruppo cui sono stati assegnati. Ad esempio, può accadere che, durante lo studio, a pazienti randomizzati al gruppo di controllo, magari per una errata prescrizione, venga somministrato il farmaco. Avranno dunque le stesse esperienze dei pazienti sottoposti al trattamento ma, per mantenere la randomizzazione iniziale, verranno inclusi nel gruppo di controllo, aumentandone così il tasso di mortalità. Così pure può accadere che pazienti randomizzati al gruppo trattato (o attivo) non prendano il farmaco. Di conseguenza lo studio fornirà un'informazione spuria: i risultati ottenuti nel gruppo attivo saranno in parte dovuti alle risposte dei pazienti che in realtà non hanno ricevuto il trattamento e viceversa vale per i risultati ottenuti nel gruppo di controllo. Tutto questo maschera e altera l'effetto efficacia del trattamento, sopra o sotto stimandola e diminuendo così la potenza del trial clinico.

# Capitolo 5

## Risultati dello studio

### 5.1 Analisi con *missing values*

Inizialmente, è stato considerato il campione intero, includendo nell'analisi tutti i soggetti partecipanti allo studio, anche quelli per i quali le rilevazioni non erano complete. Come si è visto, i metodi non parametrici di permutazione consentono di ottenere test esatti anche in presenza di *missing values*. Il campione suddiviso nei due gruppi, per età e sesso, è descritto nelle tabelle 5.1 e 5.2. I due gruppi, quello trattato secondo il modello di benessere psicologico *Well Being Therapy*, (WBT) e quello trattato seguendo esclusivamente le teorie cognitive della *Cognitive Behaviour Therapy* (CBT) non hanno uguale composizione: per esempio si può notare che le classi in cui è stato applicato il CBT sono costituite in modo particolare da tredicenni, al contrario di quelle trattate con WBT, in cui le osservazioni sono ripartite nella fascia d'età compresa tra i 12 e i 14 anni.

WBT		
età	femmine	maschi
12	9	6
13	11	18
14	6	9
15	0	1
16	0	0

Tabella 5.1: composizione del gruppo trattato con WBT

CBT		
età	femmine	maschi
12	0	0
13	22	31
14	0	3
15	2	0
16	2	0

Tabella 5.2: composizione del gruppo trattato con CBT

Il primo obiettivo di uno studio basato sul confronto prima-dopo è quello di verificare se vi è un effetto del trattamento all'interno del gruppo: si tratta quindi di fare un'analisi per campioni con osservazioni appaiate e missing values (campioni dipendenti). La statistica test di riferimento è del tipo (4.18): le variabili di riferimento sono dunque le differenze tra pre e post, cui viene assegnato il valore 0 se viene a mancare o il punteggio della rilevazione pre-trattamento o il punteggio del retest o se mancano entrambe le rilevazioni. Nella tabella (5.3) sono riportati i *p-value* dei test parziali non parametrici, ottenuti con  $B = 10000$  permutazioni. L'efficacia del trattamento WBT si traduce in un aumento significativo dell'accettazione di sé ( $p= 0.034$ ) e in una riduzione significativa degli atteggiamenti di ostilità ( $p= 0.020$ ), mentre, per quanto riguarda il trattamento CBT, sono da evidenziare un aumento significativo nei punteggi delle variabili scopo vita ( $p= 0.004$ ) e accettazione di sé ( $p= 0.038$ ) e una riduzione dei punteggi nelle variabili ansia ( $p= 0.001$ ), depressione ( $p= 0.024$ ) e benessere fisico ( $p= 0.009$ ). Da ricordare che quest'ultima variabile, considerata dal *Kellner's Symptom Questionnaire* (SQ), ha una valenza negativa: pur appartenendo alla sottoscala di benessere psicologico, è tale che maggiore è il punteggio ottenuto minore è il benessere. Per quanto riguarda la combinazione non parametrica dei test parziali (vedi tabella 5.4), si può vedere che il trattamento WBT ha un effetto su tutte le variabili, congiuntamente, della scala PWB ( $p= 0.012$ ). Il trattamento CBT, invece, ha un effetto significativo sulla combinazione delle variabili ansia-relax ( $p= 0.001$ ), depressione-contentezza ( $p= 0.042$ ), sintomi somatici-benessere fisico ( $p= 0.014$ ), sulla combinazione di tutte le variabili della scala SQ ( $p= 0.006$ ), su quella delle variabili PWB ( $p= 0.002$ ) e infine un effetto globale su tutte le variabili considerate congiuntamente

	WBT	CBT
	p-value	p-value
autonomia	0.156	0.089
padronanza ambiente	0.100	0.193
crescita personale	0.109	0.090
relazioni	0.135	0.220
scopo vita	0.230	0.004
accettazione	0.034	0.038
ansia	0.106	0.001
depressione	0.518	0.024
sintomi somatici	0.122	0.115
ostilità	0.020	0.132
relax	0.691	0.163
contentezza	0.684	0.239
benessere fisico	0.904	0.009
buona disponibilità	0.612	0.420

Tabella 5.3: WBT vs. CBT (campioni dipendenti)

( $p=0.001$ ). Per le variabili relax e contentezza vale lo stesso discorso fatto precedentemente per la variabile benessere fisico: ad esse viene associato un significato negativo. Quanto alla combinazione non parametrica dei *p-value* è stata impiegata la funzione di *Fisher* e, come sopra, un numero di permutazioni pari a 10000. Il secondo obiettivo consiste nel verificare quali dei due

combinazioni	WBT	CBT
	p-value	p-value
ansia-relax	0.258	0.001
depressione-contentezza	0.612	0.042
sintomi somatici-benessere fisico	0.335	0.014
ostilità buona-disponibilità	0.088	0.215
variabili scala SQ	0.267	0.006
variabili scala PWB	0.012	0.002
globale	0.084	0.001

Tabella 5.4: Combinazione non parametrica dei *p-value* (campioni dipendenti)

trattamenti è migliore. Si tratta quindi di fare test per campioni indipendenti sempre sulle variabili differenza con *missing values*. Le statistiche test utiliz-

zate sono del tipo (3.17), (3.18) o espresse nella forma (3.19). Nella tabella 5.5 accanto alle variabili considerate e rilevate nello studio sono riportati, per ciascun gruppo, il numero di *missing values*. E' emerso che il trattamento WBT è più efficace del CBT nella riduzione degli stati d'ansia ( $p= 0.021$ ) e nell'aumento delle sensazioni di benessere fisico ( $p= 0.007$ ). Per quanto

	non parametrico	N.missing	
	p-value	WBT	CBT
autonomia	0.270	4	14
padronanza ambiente	0.498	4	14
crescita personale	0.427	4	14
relazioni	0.564	4	14
scopo vita	0.120	4	14
accettazione	0.390	4	14
ansia	0.021	3	13
depressione	0.083	3	13
sintomi somatici	0.511	3	13
ostilità	0.783	3	13
relax	0.138	3	13
contentezza	0.246	3	13
benessere fisico	0.007	3	13
buona disponibilità	0.342	3	13

Tabella 5.5: WBT vs. CBT (campioni indipendenti)

riguarda la combinazione non parametrica dei  $p$ -value, il WBT ha un'azione maggiore sulla combinazione delle variabili ansia-relax ( $p= 0.030$ ) e sintomi somatici-benessere fisico ( $p= 0.038$ ), ma non è stata riscontrata una maggiore efficacia a livello globale (vedi 5.6). Per ottenere i risultati riportati in tabella, sono state effettuate  $B = 10000$  permutazioni e, per quanto riguarda la combinazione non parametrica dei  $p$ -value, è stata utilizzata la funzione di combinazione di Fisher.

## 5.2 Analisi con ITT (*Intention To Treat*)

Le stesse analisi sono state effettuate sul campione costruito secondo il principio dell'*Intention To Treat*, ovvero un campione che include tutti i soggetti che hanno effettuato entrambe le misurazioni o che non hanno effettuato il

combinazioni	non parametrico
	p-value
ansia-relax	0.030
depressione-contentezza	0.126
sintomi somatici-benessere fisico	0.038
ostilità buona-disponibilità	0.566
variabili scala SQ	0.063
variabili scala PWB	0.355
globale	0.087

Tabella 5.6: Combinazione non parametrica dei *p-value* (campioni indipendenti)

retest, e che, invece, esclude quei soggetti per i quali non sono disponibili i punteggi relativi alla prima misurazione. Per quanto riguarda i soggetti che non hanno effettuato il retest, applicando il principio dell'ITT, nella seconda misurazione vengono ripetuti i punteggi ottenuti, in ciascuna variabile, nella prima rilevazione. In modo particolare escono dallo studio 9 soggetti, 3 appartenenti al gruppo WBT, che quindi è composto da 57 soggetti anziché 60, e 6 appartenenti al gruppo trattato con CBT, costituito ora da 54 soggetti anziché 60. 9 soggetti invece sono stati valutati secondo l'ITT, di questi 1 appartenente al gruppo WBT e 8 al gruppo CBT. La composizione del nuovo campione è rappresentata nelle tabelle 5.7 e 5.8.

WBT		
età	femmine	maschi
12	9	5
13	11	17
14	6	8
15	0	1
16	0	0

Tabella 5.7: Composizione del gruppo trattato con WBT e applicazione dell'ITT

E' stato quindi risolto il problema dei *missing values*. L'analisi per campioni con dati appaiati (campioni dipendenti) ha messo in evidenza che il trattamento WBT aumenta in maniera significativa il livello di accettazione di sé ( $p= 0.031$ ) e riduce l'ostilità ( $p= 0.012$ ), mentre il trattamento CBT

CBT		
età	femmine	maschi
12	0	0
13	21	27
14	0	2
15	2	0
16	2	0

Tabella 5.8: Composizione del gruppo trattato con CBT e applicazione dell'ITT

comporta un aumento significativo dei punteggi ottenuti nelle variabili scopo vita ( $p= 0.004$ ) e accettazione di sé ( $p= 0.035$ ) e una riduzione di questi nelle variabili ansia ( $p= 0.001$ ), depressione ( $p= 0.020$ ) e benessere fisico ( $p= 0.016$ ). Per quanto riguarda la combinazione non parametrica dei test parziali risulta che il trattamento WBT ha un effetto sulla combinazione delle variabili ostilità-buona disponibilità ( $p= 0.048$ ) e sulle variabili, considerate in maniera congiunta, della scala PWB ( $p= 0.013$ ). Il trattamento CBT, invece, ha un effetto significativo sulla combinazione delle variabili ansia-relax ( $p= 0.001$ ), depressione-contentezza ( $p= 0.046$ ), sintomi somatici-benessere fisico ( $p= 0.025$ ), sulla combinazione di tutte le variabili della scala SQ ( $p= 0.008$ ), su quella delle variabili PWB ( $p= 0.002$ ) e infine un effetto globale ( $p= 0.002$ ). I risultati si trovano nelle tabelle 5.9 e 5.10. Nel confronto diretto tra i due trattamenti (campioni indipendenti), è stata rilevata una maggior efficacia del trattamento WBT rispetto al trattamento CBT nell'aumento del benessere fisico ( $p= 0.009$ ). Per quanto riguarda la combinazione non parametrica dei  $p$ -value, il WBT ha un'azione maggiore solo sulla combinazione delle variabili sintomi somatici-benessere fisico ( $p= 0.038$ ), ma non è migliore a livello globale. I risultati si trovano nelle tabelle 5.11 e 5.12. I valori dei  $p$ -value non parametrici, nell'analisi per campioni con osservazioni appaiate e in quella per campioni indipendenti, sono stati ottenuti con  $B = 10000$  permutazioni, per la combinazione non parametrica, è stata impiegata la funzione di combinazione di Fisher. Miglioramenti significativi in alcune scale del PWB e dell'SQ sono stati apportati sia dal trattamento WBT, incentrato sul riconoscimento e la valorizzazione delle emozioni positive secondo i concetti proposti dal modello di benessere psicologico di Carol



	WBT	CBT
	p-value	p-value
autonomia	0.167	0.080
padronanza ambiente	0.124	0.191
crescita personale	0.100	0.094
relazioni	0.142	0.186
scopo vita	0.242	0.004
accettazione	0.031	0.035
ansia	0.064	0.001
depressione	0.476	0.020
sintomi somatici	0.107	0.143
ostilità	0.012	0.097
relax	0.539	0.135
contentezza	0.547	0.305
benessere fisico	0.881	0.016
buona disponibilità	0.532	0.425

Tabella 5.9: WBT *vs.* CBT e applicazione dell'ITT (campioni dipendenti)

Ryff, sia dall'intervento CBT, focalizzato sul riconoscimento delle emozioni negative secondo i concetti proposti dalla teoria cognitiva. L'efficacia degli interventi scolastici ad orientamento cognitivo-comportamentale è già stata ampiamente dimostrata in letteratura. Il fatto che anche un protocollo basato sulla promozione del benessere psicologico, anziché sulla rimozione della sintomatologia, sia risultato in qualche modo efficace sia nella riduzione della sintomatologia che nell'aumento del benessere psicologico dei soggetti, mette in evidenza l'applicabilità di tale protocollo in ambito scolastico e su una popolazione adolescenziale, sia in un'ottica terapeutica che preventiva.

Da un punto di vista prettamente statistico si possono fare le seguenti considerazioni: per quanto riguarda il confronto tra pre e post trattamento, il primo metodo di analisi, ovvero il metodo che utilizza l'intero campione inclusi i *missing data*, e il secondo metodo, che sfrutta il principio dell'ITT, forniscono gli stessi risultati. Fa eccezione la significatività della combinazione ostilità-buona disponibilità, rilevata dal secondo metodo, che, a tutti gli effetti, si può definire una significatività al limite dell'accettazione dato che il *p-value* è pari a 0.048 ( $\sim 0.05$ ). Quanto al confronto diretto tra WBT e CBT,

combinazioni	WBT	CBT
	p-value	p-value
ansia-relax	0.157	0.001
depressione-contentezza	0.525	0.046
sintomi somatici-benessere fisico	0.302	0.025
ostilità buona-disponibilità	0.048	0.173
variabili scala SQ	0.178	0.008
variabili scala PWB	0.013	0.002
globale	0.061	0.002

Tabella 5.10: Combinazione non parametrica dei *p-value* e applicazione dell'ITT (campioni dipendenti)

il primo metodo, a differenza del secondo, evidenzia la significatività della variabile ansia e della combinazione delle variabili ansia e relax. In conclusione, nonostante queste lievi differenze, la cosa fondamentale è che il principio dell'ITT comporta una riduzione della numerosità del campione, escono infatti dallo studio 9 soggetti, e falsa i punteggi ottenuti nella seconda rilevazione per 9 soggetti, mentre il metodo improntato sui metodi non parametrici di permutazione non elimina alcuna informazione e non solo rispecchia di più la realtà ma, come si è visto, fornisce risultati esatti.

	non parametrico
	p-value
autonomia	0.356
padronanza ambiente	0.581
crescita personale	0.517
relazioni	0.656
scopo vita	0.197
accettazione	0.523
ansia	0.070
depressione	0.124
sintomi somatici	0.636
ostilità	0.846
relax	0.191
contentezza	0.342
benessere fisico	0.009
buona disponibilità	0.411

Tabella 5.11: WBT *vs.* CBT e applicazione dell'ITT (campioni indipendenti)

combinazioni	non parametrico
	p-value
ansia-relax	0.078
depressione-contentezza	0.186
sintomi somatici-benessere fisico	0.048
ostilità buona-disponibilità	0.646
variabili scala SQ	0.117
variabili scala PWB	0.519
globale	0.175

Tabella 5.12: Combinazione non parametrica dei *p-value* e applicazione dell'ITT (campioni indipendenti)



# Appendice A

## *ECDF* delle simulazioni

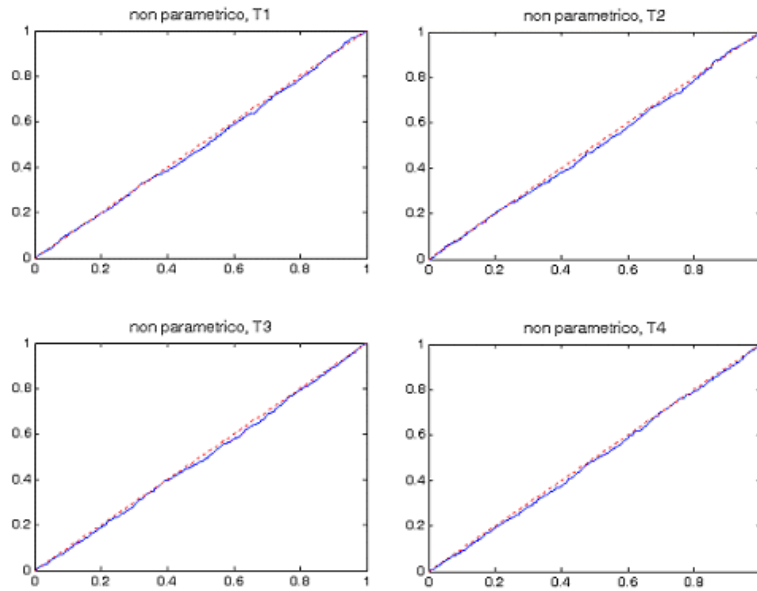


Figura A.1: *Empirical cumulative distribution function non parametrica sotto  $H_0$*

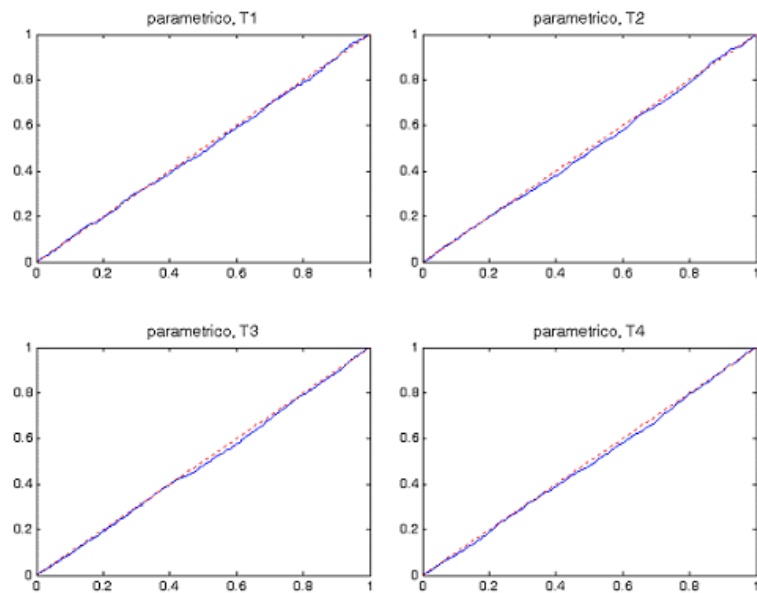


Figura A.2: *Empirical cumulative distribution function parametrica sotto  $H_0$*

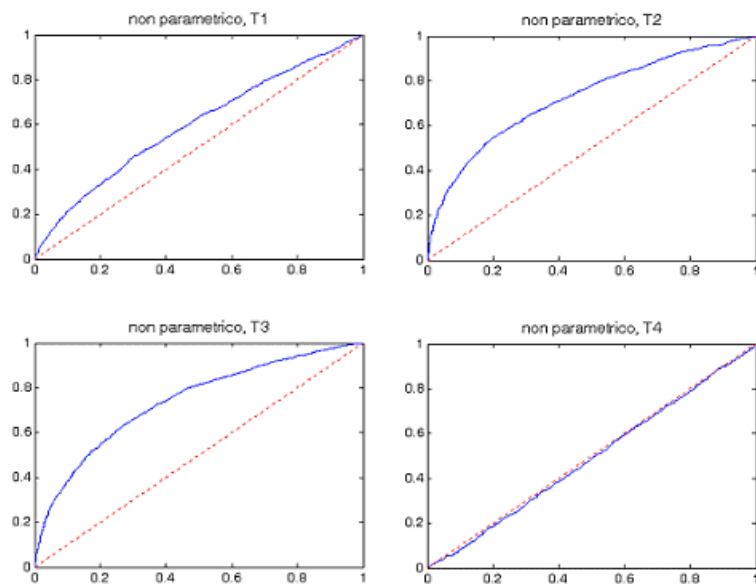


Figura A.3: *Empirical cumulative distribution function* non parametrica sotto  $H_1$  ( $\delta_1 = 0.2$ ,  $\delta_2 = 0.4$ )

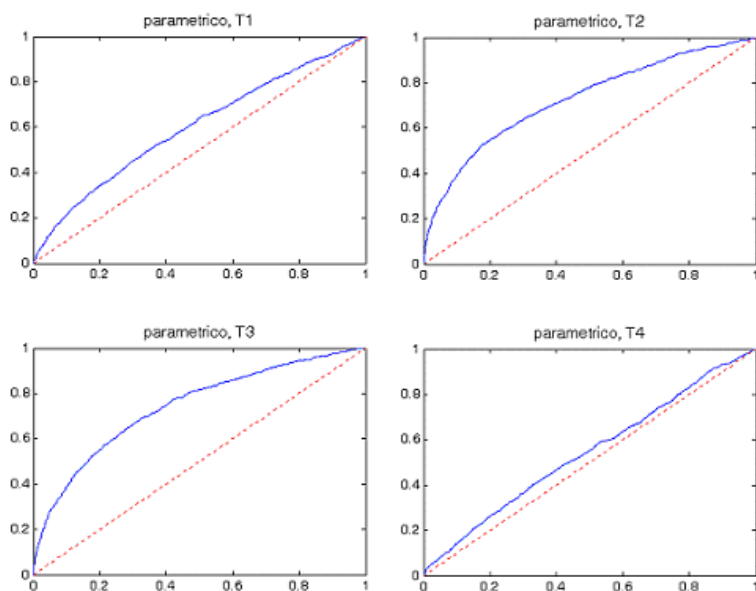


Figura A.4: *Empirical cumulative distribution function* parametrica sotto  $H_1$  ( $\delta_1 = 0.2$ ,  $\delta_2 = 0.4$ )

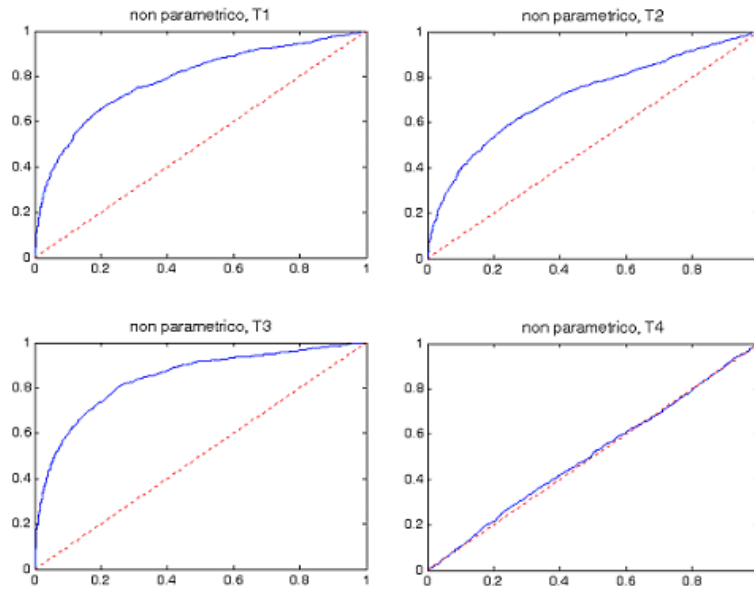


Figura A.5: *Empirical cumulative distribution function* non parametrica sotto  $H_1$  ( $\delta_1 = 0.4$ ,  $\delta_2 = 0.4$ )

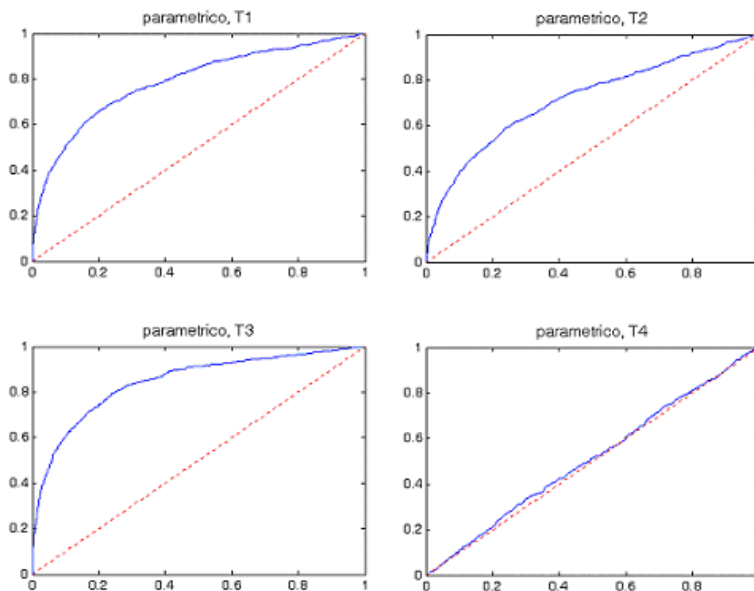


Figura A.6: *Empirical cumulative distribution function* parametrica sotto  $H_1$  ( $\delta_1 = 0.4$ ,  $\delta_2 = 0.4$ )



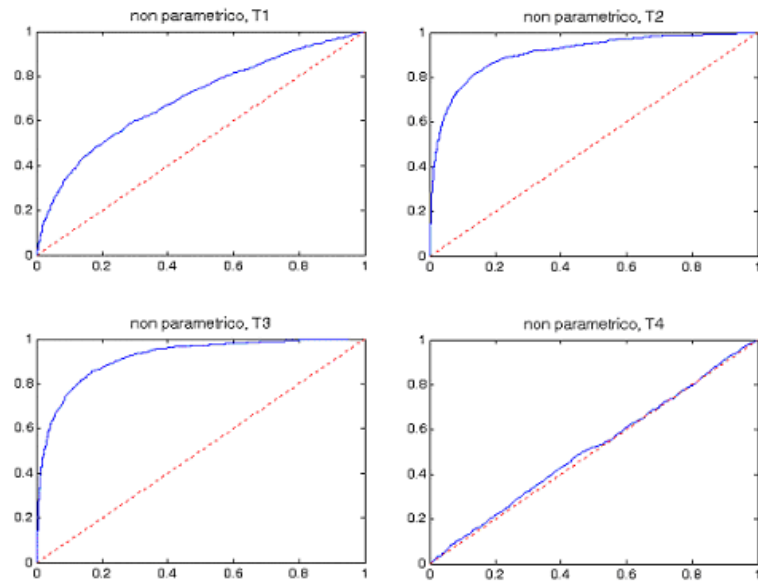


Figura A.7: *Empirical cumulative distribution function* non parametrica sotto  $H_1$  ( $\delta_1 = 0.3$ ,  $\delta_2 = 0.7$ )

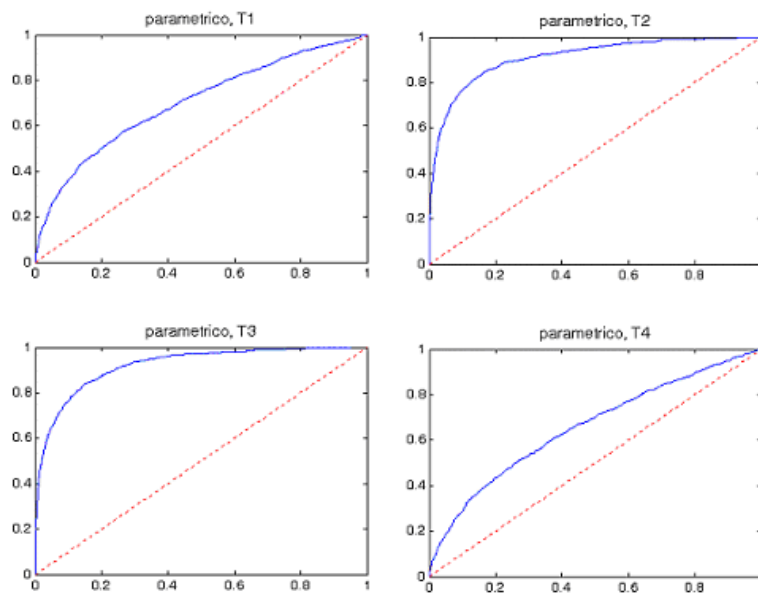


Figura A.8: *Empirical cumulative distribution function* parametrica sotto  $H_1$  ( $\delta_1 = 0.3$ ,  $\delta_2 = 0.7$ )



# Appendice B

## Risultati stratificazione

A titolo di curiosità, per completare l'analisi per campioni indipendenti con osservazioni appaiate, è stata introdotta la variabile sesso, prima ignorata, come variabile di stratificazione. L'obiettivo è quello di verificare se l'appartenenza ad uno specifico strato (femmine, maschi) incida sulla maggior efficacia del trattamento WBT rispetto al trattamento CBT. Questa "indagine" è stata effettuata, sempre via permutazione, prima sull'intero campione, lasciando quindi indicati i *missing values*, e poi sul campione costruito sulla base del principio dell'*intention to treat*. Osservando i valori assunti dai p-value relativi ai test parziali nel campione stratificato con dati mancanti, si può vedere questa maggior efficacia del WBT, rispetto al CBT, si traduca, nello strato costituito dalle femmine, in un miglioramento delle relazioni, una riduzione degli stati d'ansia e in un incremento del benessere fisico e, nello strato costituito dai maschi, in un miglioramento in termini rilassamento, contentezza e benessere fisico. Per quanto riguarda la combinazione di secondo livello, si può procedere in due diversi modi: il primo consiste nell'eseguire una combinazione non parametrica, all'interno di ciascun strato, di tutte le variabili poste sotto test e poi combinare rispetto agli strati (combinazione *Within Strata*); la seconda invece consiste nell'effettuare una combinazione non parametrica di tutti i test corrispondenti alla stessa variabile nei diversi strati e poi combinare rispetto alle variabili (combinazione *Within Variables*). Nel primo caso, facendo riferimento ai nostri dati, si ottiene un test per la combinazione globale delle variabili dello strato delle femmine, uno per la combinazione globale delle variabili dello strato dei maschi, e un test derivante dalla combinazione dei due precedenti, nel secondo caso si otten-

gono 14 test, tanti quanti le variabili processate, ottenuti dalla combinazione di due test parziali, uno proveniente dallo strato delle femmine e uno dallo strato dei maschi, relativi ad una stessa variabile. La combinazione *Within Strata* ha messo in evidenza una leggera significatività del test globale, derivante dalla combinazione rispetto agli strati, mentre nella combinazione *Within Variables* sono risultate significative le variabili ansia, contentezza, benessere fisico, e la combinazione globale. Tutti i risultati sono stati ottenuti applicando la funzione di combinazione non parametrica di Fisher. I risultati sono riportati nelle tabelle B.1, B.2 e B.3.

	femmine	maschi
	p-value	p-value
autonomia	0.605	0.140
padronanza ambiente	0.251	0.637
crescita personale	0.498	0.345
relazioni	0.014	0.971
scopo vita	0.252	0.171
accettazione	0.186	0.709
ansia	0.012	0.282
depressione	0.233	0.105
sintomi somatici	0.112	0.912
ostilità	0.444	0.907
relax	0.545	0.045
contentezza	0.837	0.010
benessere fisico	0.036	0.038
buona disponibilità	0.620	0.218

Tabella B.1: Test parziali

	p-value
strato femmine	0.096
strato maschi	0.091
globale	0.049

Tabella B.2: Within strata

Nel campione stratificato costruito con un approccio ITT, i p-value parziali significativi (vedi tabella B.4) nello strato costituito dalle femmine sono quelli relativi alle variabili relazioni, ansia e benessere fisico, mentre nello strato

	p-value
autonomia	0.280
padronanza ambiente	0.428
crescita personale	0.462
relazioni	0.064
scopo vita	0.168
accettazione	0.378
ansia	0.023
depressione	0.113
sintomi somatici	0.334
ostilità	0.765
relax	0.112
contentezza	0.045
benessere fisico	0.010
buona disponibilità	0.395
combinazione globale	0.043

Tabella B.3: Within variables

dei maschi l'unico p-value parziale significativo è quello relativo alla variabile contentezza. Per quanto riguarda la combinazione *Within Strata* (vedi tabella B.5) nessun test è risultato significativo, mentre per quanto riguarda invece la combinazione *Within Variables* (vedi tabella B.6) il trattamento WBT sembra più efficace del CBT, congiuntamente per maschi e femmine, nel migliorare il benessere fisico. Anche qui è stata scelta la funzione di combinazione non parametrica di Fisher.

	femmine	maschi
	p-value	p-value
autonomia	0.730	0.201
padronanza ambiente	0.347	0.744
crescita personale	0.609	0.469
relazioni	0.030	0.984
scopo vita	0.280	0.233
accettazione	0.296	0.788
ansia	0.022	0.476
depressione	0.256	0.157
sintomi somatici	0.135	0.964
ostilità	0.521	0.940
relax	0.623	0.055
contentezza	0.865	0.022
benessere fisico	0.046	0.068
buona disponibilità	0.681	0.241

Tabella B.4: Test parziali e applicazione ITT

	p-value
strato femmine	0.156
strato maschi	0.161
globale	0.117

Tabella B.5: Within strata e applicazione ITT

---

	p-value
autonomia	0.362
padronanza ambiente	0.523
crescita personale	0.573
relazioni	0.099
scopo vita	0.180
accettazione	0.491
ansia	0.063
depressione	0.160
sintomi somatici	0.375
ostilità	0.832
relax	0.125
contentezza	0.074
benessere fisico	0.017
buona disponibilità	0.417
combinazione globale	0.101

Tabella B.6: Within variables e applicazione ITT





# Bibliografia

- [1] Pesarin, F. (2001) *Multivariate Permutation tests: with application in Biostatistics*. John Wiley & Sons, Chichester-New York.
- [2] Celant, G., Pesarin, F., Salmaso L. (2000) Two sample permutation tests for repeated measures with missing values. *Journal of Applied Statistical Science*, 291-304.
- [3] Solari, A., Finos, L., Pesarin, F. (2003) Test di permutazione per la differenziazione genetica. Working Paper no. 2003.17, Dip. di Scienze Statistiche, Università di Padova.
- [4] World Health Organization (2004) Draft conclusions from the pre-conference The mental health of children and adolescents, organized by the European Commission, WHO European region and the Ministry of Health of Luxembourg, 12-20 September.
- [5] World Health Organization (2004) The world health report 2004: Changing history. Geneva, World Health Organization.
- [6] Nolen-Hoeksema, S. & Davis, C.G. (2002) Positive response to loss. Perceiving benefits and growth. In: Snyder, C.R., Lopez, S.J. (eds). *Handbook of positive psychology*. Oxford University Press, 598-607.
- [7] Nolen-Hoeksema, S. & Larson, J. (1999) *Coping with loss*. Erlbaum, Mahwah, NJ.
- [8] Fava, G.A., Kellner, R., Perini, G., Fava, M., Michelacci, L., Munari, F., Evangelisti, L.P., Grandi, S., Bernardi, M. & Mastrogiamomo, I. (1993) Italian validation of the Symptom Rating Test (SRT) and Symptom Questionnaire (SQ). *Can J Psych*, 28, 177-123.

- 
- [9] Kellner, R. (1987) A Symptom Questionnaire. *J Clin Psychiatry*, 48, 269-274.
- [10] Ryff, C.D. (1989) Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *J Pers Soc Psychology*, 57, 1069-1081.
- [11] Ryff, C.D. (1989) Beyond Ponce De Leon and life satisfaction: new directions in quest of successful aging. *Intern J Behav Development*, 12, 35-55.
- [12] Canestrari, R. (1982) *Nuovi metodi in psicomertia*. Organizzazioni speciali, Firenze.
- [13] Fava, G.A. (2000) Cognitive behavioral therapy. In M. Fink (Ed), *Encyclopedia of stress*. Academic Press, San Diego, pp.484-497.
- [14] Celant, G., Pesarin, F., Salmaso L. (2002) Some comparisons between a parametric and a non parametric solution for tests with repeated measures. *Metron*, LVIII, n.1-2, 64-79.
- [15] Liptak, I. (1958) On the combination of independent tests. *Magyar Tudományok Akademia Matematikai Kutato Intezenek Kozlomenyei*, 3, 127-141.
- [16] Bacchieri, A., Della cioppa, G. (2004) *Fondamenti di ricerca clinica*. Springer-Verlag, Milano
- [17] Lehmann, E.L. (1986) *Testing Statistical Hypotheses* (2nd edn). Wiley, New York.
- [18] Dwass, M. (1957) Modified randomization tests for nonparametric hypotheses. *Annals of Mathematical Statistics*, 28, 181-187.
- [19] Bartholomew, D.J. (1961) A test of homogeneity of means under restricted alternatives. *Journal of the Royal Statistical Society B*, 23, 239-281.
- [20] Kudo, A. (1963) A multivariate analogue of the one-sided test. *Biometrika*, 50, 403-418.

- [21] Nüesch, P.E. (1966) On the problem of testing location in multivariate populations for restricted alternatives. *Annals of Mathematical Statistics*, 37, 113-119.
- [22] Shorack, G.R. (1967) Testing against ordered alternative in model I analysis of variance: Normal theory and non-parametrics. *Annals of Mathematical Statistics*, 38, 1740-1753.
- [23] Perlman, M.D. (1969) One-sided testing problems in multivariate analysis. *Annals of Mathematical Statistics*, 40, 549-567.
- [24] Chatterjee, S.K. (1984) Restricted alternatives. In P.R. Krishnaiah e P.K. Sen (eds), *Handbook of Statistics*, 4, 327-345. North Holland, Amsterdam.
- [25] Robertson, T., Wrigth, F.T., Dykstra, R.L. (1988) *Order Restricted Statistical Inference*. Wiley, Chichester.
- [26] El Barmi, H., Dykstra, R. (1995) Testing for and against a set of linear inequality constraints in a multinomial setting, *Canadian Journal of Statistics*, 23, 131-143.
- [27] Dardanoni, V., Forcina, A. (1998) A unified approach to likelihood inference on stochastic orderings in a non-parametric context, *Journal of the American Statistical Association*, 93, 1112-1123.
- [28] Perlman, M.D., Wu, L. (1998) A class of conditional tests for a multivariate one-sided alternative. Technical report 328, Dip. di Scienze Statistiche, Università di Washington, Seattle..
- [29] Rubin, D.B. (1976) Inference and missing data. *Biometrika*, 63, 581-92.
- [30] Little, R.J.A. (1982) Models for nonresponse in sample survey. *Journal of the American Statistical Association*, 77, 237-250.
- [31] Barton, C.N., Cramer, E.C. (1989) Hypothesis testing in multivariate linear models with randomly missing data. *Communications in Statistics. Simulation and Computation*, 18, 875-895.

- 
- [32] Maritz, J.S. (1995) A permutation test allowing for missing values. *Australian Journal of Statistics*, 37, 153-159.
- [33] Little, R.J.A., Rubin, D.B. (1987) *Statistical Analysis with Missing Data*. Wiley, New York.
- [34] Wei, L.J., Lachin, J.M. (1984) Two-sample asymptotically distribution-free tests for incomplete multivariate observations. *Journal of the American Statistical Association*, 79, 653-661.
- [35] Wei, L.J., Johnson, W.E. (1985) Combining dependent tests withincomplete repeated measurements. *Biometrika*, 72, 359-364.
- [36] Giraldo, A., Pallini, A., Pesarin, F. (1994) A nonparametric testing procedure for missing data. In R. Dutter e W. Grossmann (eds), *COMPSTAT 1994- Proceedings in Computational Statistics*, 503-508. Physica-Verlag, Heidelberg.
- [37] Giraldo, A., Pesarin, F. (1992) Verifica d'ipotesi in presenza di dati mancanti e tecniche di ricampionamento. In *Atti della XXXVI Riunione Scientifica della Società Italiana di Statistica*, 2, 271-278.
- [38] Giraldo, A., Pesarin, F. (1993) A resampling procedure for missing data in testing problems. In G. Diana, L. Pace e A. Salvan (eds), *Atti della Conferenza 'Due Temi di Statistica Metodologica'*, 115-122. Curto, Napoli.
- [39] Pesarin, F. (1990) On a parametric combination method for dependent permutation tests with applications. *Psychotherapy and Psychosomatics*, 54, 172-179.
- [40] Pesarin, F. (1991) Tecniche di ricampionamento e verifica d'ipotesi multidimensionale. *Statistica*, 50, 483-501.
- [41] Bertacche, R., Pesarin, F. (1997) Treatment of missing data in multidimensional testing problems for categorical variables. *Metron*, 55, 135-149.
- [42] Bradford Hill, A. (1961) *Principles of medical Statistics*. 7th edition. The Lancet, London.

- [43] Fisher, L.D., Dixon, D.O., Herson, J., Frnkowski, R.K., Hearron, M.S., Peace, K.E. (1990) Intention to treat in clinical practice. In Peace, K.E., ed., *Statistical Issues in Drug Research and Development*, 331-350. Marcel Dekker, New York



# Ringraziamenti

Con un grazie particolare voglio esprimere tutta la mia stima e ammirazione al **Prof. Fortunato Pesarin**, mio relatore, perché si è sempre dimostrato disponibile, al **Prof. Giovanni Fava** che ha seguito nell'analisi dei dati, a **Carlotta Belaise** e **Chiara Ruini** per il materiale fornitomi.

Un enorme grazie al **Prof. Livio Finos**, per il suo preziosissimo aiuto e che ora può dimostrare a tutti che con i potenti mezzi di comunicazione ci si può laureare facilmente via Internet o per telefono.

Grazie ai miei **genitori**, perché hanno creduto in me, mi hanno incoraggiata e sostenuta nei momenti di sconforto, a mio fratello **Matteo**, che mi è stato indispensabile per la stesura della tesi e che, a quanto pare, dei ciccioni e ubriaconi irlandesi vorrebbero portarmi via, al nonno Bepi che è convinto diventerò deputato, alla nonna **Sofia** e alla nonna **Wanda** che mi voleva estetista.

A tutte le **Scopoline** e ai **Mazziani** con cui ho trascorso questi anni di università, alle mie compagne di facoltà **Carla DeFranceschi**, **Angelina Paggiaro** ed **Erica Passadore**, a **Guada**, che anche se distante rimarrà sempre la mia straniera preferita, a **Giovanni Fois**, che mi ha fatto muovere i primi passi nello spazio di lavoro scientifico (SWP).