

UNIVERSITÀ DEGLI STUDI DI PADOVA  
DIPARTIMENTO DI SCIENZE STATISTICHE  
CORSO DI LAUREA MAGISTRALE IN  
SCIENZE STATISTICHE



## **Analisi di tracciati EEG tramite Functional Data Analysis**

Relatore: Prof. Bruno Scarpa

Dipartimento di Scienze Statistiche

Correlatore: Dott. Enrico Bosco

Laureando: Francesco Zulian

Matricola N° 622481

Anno Accademico 2012/2013



UNIVERSITÀ DEGLI STUDI DI PADOVA  
DIPARTIMENTO DI SCIENZE STATISTICHE  
CORSO DI LAUREA MAGISTRALE IN  
SCIENZE STATISTICHE



## **Analisi di tracciati EEG tramite Functional Data Analysis**

Relatore: Prof. Bruno Scarpa

Dipartimento di Scienze Statistiche

Correlatore: Dott. Enrico Bosco

Laureando: Francesco Zulian

Matricola N° 622481

Anno Accademico 2012/2013



# Indice

<b>Introduzione</b>	<b>3</b>
Struttura della tesi . . . . .	4
<b>1 Elettroencefalografia del Coma</b>	<b>5</b>
1.1 Storia . . . . .	5
1.1.1 Il tracciato EEG . . . . .	7
1.1.2 I potenziali evocati . . . . .	8
1.2 Il coma . . . . .	9
1.2.1 Classificazioni conosciute . . . . .	10
1.2.2 Scale di valutazione del coma . . . . .	21
1.3 Lavori precedenti . . . . .	24
1.3.1 Il caso studiato . . . . .	24
1.3.2 Altri studi . . . . .	25
<b>2 Functional Data Analysis</b>	<b>27</b>
2.1 Cos'è l'analisi dei dati funzionali . . . . .	28
2.1.1 Rappresentazione dei dati funzionali . . . . .	28
2.1.2 Le basi . . . . .	28
2.1.3 Il modello . . . . .	33
2.2 Perché scegliere l'analisi dei dati funzionale . . . . .	36

<b>3</b>	<b>Il dataset studiato</b>	<b>39</b>
3.1	Variabili considerate . . . . .	42
3.1.1	Le variabili . . . . .	45
3.2	Studio dei dati funzionali . . . . .	48
3.2.1	<i>Smoothing</i> dei dati . . . . .	48
3.2.2	Intervalli di confidenza per il coefficiente . . . . .	50
3.2.3	Test di permutazione . . . . .	50
3.2.4	Un semplice modello . . . . .	53
3.2.5	Finestra temporale di 24 ore . . . . .	60
3.2.6	Finestra temporale di 48 ore . . . . .	63
	<b>Conclusioni</b>	<b>67</b>
	<b>A Codice Java</b>	<b>69</b>
	<b>B Codice R</b>	<b>83</b>

# Introduzione

Ci sono molte cose che differenziano l'essere umano dagli altri esseri viventi, di certo una su tutte è il cervello.

L'essere umano in milioni di anni di evoluzione ha sviluppato un cervello senza eguali in natura. Si tratta di un organo molto complesso, e proprio per questo motivo, tuttora ampiamente studiato.

L'ambito di studio di questo organo così straordinario trattato in questa tesi tuttavia non sarà quello delle sue splendide qualità, o di quanto un essere umano sia capace di fare grazie ad esso. Si concentrerà piuttosto su quanto succede quando questo organo non è in grado di funzionare correttamente in seguito ad un trauma. In particolare andrà ad analizzare quanto succede durante lo stato comatoso, momento in cui la maggior parte delle funzioni cerebrali sembrano interrompersi lasciando spazio ad uno stato di incoscienza di durata difficilmente prevedibile.

La tesi sarà suddivisa in capitoli che cercheranno di rispondere a queste domande:

- Quanto si conosce al momento rispetto alla classificazione dello stato comatoso.
- Esiste la possibilità di prevedere la capacità di un soggetto a compiere determinate azioni dopo il risveglio dal coma.

- Esistono degli intervalli temporali particolarmente utili e significativi a livello informativo durante lo stato comatoso, al fine di prevedere quanto detto sopra.

Questo testo perciò, non ha l'ambizione di essere un trattato esaustivo sul funzionamento del cervello durante lo stato comatoso, si tratta piuttosto di un'analisi dell'informazione che è possibile estrarre da un tracciato elettroencefalografico di un paziente in coma, al fine di prevedere l'esito della degenza.

## Struttura della tesi

**Capitolo 1 Elettroencefalografia del Coma:** breve introduzione sul coma e l'elettroencefalografia. Cenni sulla storia dell'elettroencefalografia, tipi di coma conosciuti e lavori precedenti sul tema.

**Capitolo 2 Functional Data Analysis:** cos'è, quali sono le sue applicazioni e come potrebbe rivelarsi utile per lo studio dei tracciati EEG.

**Capitolo 3 Il dataset studiato:** definizione delle variabili oggetto di studio e sintesi dei dati.

**Appendice A** Codice java per la creazione dei file di sintesi.

**Appendice B** Parte del codice R utilizzato per l'analisi.

# Capitolo 1

## Elettroencefalografia del Coma

### 1.1 Storia

L'elettroencefalografia consiste nella registrazione dell'attività elettrica dell'encefalo<sup>1</sup>. Questa tecnica venne inventata nel 1929 da Hans Berger, il quale scoprì che, inserendo degli aghi nello scalpo di un individuo, era possibile misurare una differenza di potenziale elettrico. In seguito tale tecnica fu perfezionata da Herbert Jasper.

La rilevazione del tracciato EEG ad opera di Berger consisteva nell'inserire due spilli d'argento sotto la cute dello scalpo, uno in posizione frontale ed uno in posizione posteriore, successivamente la tecnica si è via via affinata fino ad ottenere il sistema di posizionamento standard che va sotto il nome *Sistema internazionale 10-20*.

Tale sistema consiste nell'applicare degli elettrodi<sup>2</sup> sullo scalpo del paziente secondo un posizionamento standard che considera la distanza tra due

---

<sup>1</sup>Encefalo (dal greco *encephalon*, dentro la testa) è quella componente del sistema nervoso centrale completamente contenuta nella scatola cranica (Anastasi, 2007).

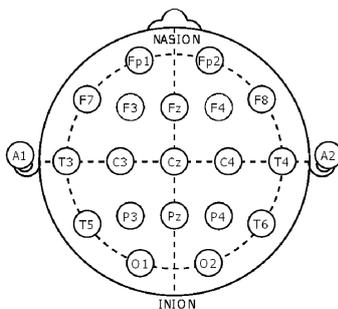
<sup>2</sup>Fortunatamente non si usano più spilli d'argento.

punti di repere cranici<sup>3</sup> chiamati *inion* (prominenza alla base dell'osso occipitale) e *nasion* (attaccatura superiore del naso), questa distanza va da 30 a 36 cm. Gli elettrodi vengono collocati lungo cinque linee:

- P1: longitudinale esterna
- P2: longitudinale interna destra
- C: Centrale
- P1: longitudinale esterna
- P2: longitudinale interna sinistra

La linea trasversale risultante dalle precedenti viene denominata montaggio P3.

Nel sistema internazionale 10-20 (Fig. 1.1), la dicitura 10-20 sta ad indicare la regola con cui gli elettrodi vengono disposti rispetto al 100% della distanza tra i punti *nasion* ed *inion*.



**Fig. 1.1:** Sistema internazionale 10-20.

---

<sup>3</sup>Si tratta di una serie di punti che sono individuabili utilizzando le parti del cranio apprezzabili tramite un'esplorazione superficiale dello stesso.

In particolare gli elettrodi fronto-polari son posizionati ad una distaza pari al 10%, sopra le sopracciglia; i frontali vengono disposti sulla stessa linea con una distanza del 20%; i centrali ed i parietali sempre con il 20% di distanza tra loro ed infine gli occipitali (sempre al 20%). In totale la distanza cumulata è del 90% quindi al 10% dall'inion. Una sigla è associata ad ogni posizione che gli elettrodi occupano sullo scalpo, questa è composta da due lettere a cui è associato un numero per indicare l'emisfero<sup>4</sup>:

- Fp: Frontopolare
- F: Frontale
- C: Centrale
- P: Parietale
- T: Temporale
- O: Occipitale

### 1.1.1 Il tracciato EEG

La rappresentazione grafica della registrazione elettroencefalografica è detta elettroencefalogramma. Questa rappresentazione viene prodotta da uno strumento (analogico o digitale) chiamato elettroencefalografo.

L'interpretazione del tracciato elettroencefalografico è un'operazione che richiede esperienza e può essere effettuata solamente da personale medico preparato. Tuttavia la logica di rappresentazione di quest'ultimo non è di difficile comprensione.

---

<sup>4</sup>Numeri dispari: emisfero sinistro. Numeri pari: emisfero destro.

Sull'asse delle ascisse sono riportate delle linee verticali in corrispondenza di ogni secondo di registrazione, per dare la possibilità di quantificare la durata dell'onda. Sull'asse delle ordinate, invece, sono riportate le varie zone cerebrali seguendo la dicitura standard del sistema internazionale 10-20 per permettere a chi legge il tracciato di identificare in quale zona cerebrale è localizzata l'onda. Poiché differenti onde hanno diversa ampiezza vi è anche la possibilità di modificare la scala di voltaggio, per poter visualizzare al meglio l'onda. Anche la frequenza di campionamento in Hz può essere modificata, per poter analizzare differenti onde. In figura 1.2 è rappresentato un tipico tracciato EEG con ampiezza impostata a  $200 \mu\text{V}$ .



**Fig. 1.2:** Tracciato EEG esemplificativo, in ascissa i secondi (linee rosse continue) ed in ordinata le zone cerebrali secondo il Sistema internazionale 10-20. Per ognuna sono riportate le misurazioni in  $\mu\text{V}$  del segnale ad ogni frazione di secondo.

### 1.1.2 I potenziali evocati

I potenziali evocati sono una tecnologia che sfrutta stimoli di differente tipo per verificare la presenza di risposte cerebrali. Servono a dare una misurazio-

ne affidabile ed obiettiva del funzionamento del sistema nervoso, soprattutto per identificare anomalie non identificabili tramite una visita esterna, tali anomalie vengono chiamate lesioni silenti (Nuwer, 1998). In presenza di un paziente sano, i potenziali evocati hanno delle curve stabili con ampiezza e latenza che si ripetono allo stesso modo se si misurano in momenti diversi in differenti anni.

Esistono diversi tipi di potenziali evocati, i più utilizzati clinicamente sono i potenziali evocati somatosensitivi (SEP), acustici (BAEP) e visivi (VEP). I potenziali evocati visivi studiano il nervo ottico e la corteccia visiva stimolando il paziente tramite la proiezione su di uno schermo di quadrati luminosi e non ad intervalli di tempo regolari, mentre per i pazienti in coma vengono utilizzati degli stimoli flash.

Per quanto riguarda i potenziali evocati acustici a breve latenza, questi vengono rilevati mediante degli elettrodi nello scalpo ripetendo prima in un orecchio e poi in un altro degli stimoli di breve durata (click) a frequenza costante per valutare il decorso del nervo acustico fino al tronco encefalico. Infine i potenziali evocati somatosensitivi si misurano stimolando il nervo mediano al polso e/o il nervo tibiale posteriore alla caviglia tramite un potenziale elettrico che viaggia lungo la via somatosensitiva fino alla corteccia somestesica, area deputata alla ricezione di stimoli sensitivi.

## **1.2 Il coma**

Il coma è un disturbo della coscienza. Esistono diversi gradi di compromissione della coscienza che possono passare dalla obnubilazione alla completa mancanza di risposte a qualsiasi tipo di stimolo (Husain, 2006). Il tracciato EEG può essere utilizzato per molteplici scopi, ad esempio viene utilizzato

nei casi di disturbi convulsivi come l'epilessia o per segnalare la presenza di alterazioni che possono indurre il neurologo a richiedere una TAC. Può anche essere utilizzato, come nel nostro caso, per valutare la profondità del coma.

### 1.2.1 Classificazioni conosciute

Esistono diversi tipi di conformazioni per le onde cerebrali registrate in un EEG, ed i pattern più frequentemente osservabili sono i seguenti:

- Attività di tipo *delta intermittente*
- *Onde trifasiche*
- Attività delta continua ad alto voltaggio
- Scariche periodiche di forma epilettica
- Scariche periodiche di forma epilettica generalizzate
- *Burst suppression*
- EEG non reattivo, lento, a basso voltaggio
- Inattività elettrocerebrale
- *Spindle coma*
- *Coma alfa*
- *Coma beta*

Questi pattern sono osservabili in diversi momenti durante lo stato comatoso e la loro identificazione aiuta i neurofisiologi nel processo di diagnosi e prognosi. A seconda di come si susseguono i suddetti è possibile definire (anche se con una notevole incertezza) se la situazione di coma sta migliorando o

peggiorando, ma è doveroso far notare che esistono molteplici classificazioni dei pattern anche precedenti a quella esposta. La più classica ed internazionalmente accettata ideata da Synek (1988), ad esempio, divide i pattern in cinque macro aree, tuttavia il fattore che accomuna tutti i tipi di classificazione, è quello di valutare principalmente la frequenza in Hz dell'onda e l'ampiezza della stessa misurata in  $\mu V$ , informazione utile per definire le variabili oggetto di studio, argomento che verrà approfondito nel paragrafo 3.1.

### **Le onde cerebrali**

Le differenti conformazioni elencate in precedenza sono caratterizzate dalla presenza di onde cerebrali di vario tipo. Le più conosciute ed anche le prime scoperte da Berger, sono le onde di tipo Alfa con frequenza 8-12Hz le quali possono essere registrate durante una fase rilassata di veglia ad occhi chiusi. Altro tipo di onde sono le onde Beta, vanno dai 14 ai 30 Hz e si registrano in un soggetto cosciente.

Altre onde conosciute sono le onde Delta che presentano frequenze tra 0,5 e 4 Hz e compaiono nelle fasi di sonno non REM (senza sogni). Infine le onde Theta hanno una frequenza che va dai 5 ai 7 Hz e si rilevano soprattutto durante la fase di sonno REM (con sogni) esse tuttavia possono anche essere segno precoce di patologia endocranica.

Il susseguirsi delle diverse onde cerebrali sopra elencate portano alla formazione dei pattern in seguito esposti.

### **Attività di tipo *delta intermittente***

L'attività di tipo *delta intermittente* (IRDA) è composta da onde sinusoidali di frequenza che va dai 2 ai 3 Hz, queste appaiono in maniera ritmica

ma intermittente. Negli adulti tale forma appare predominante nelle regioni frontali, mentre nei bambini è solitamente occipitale. Questo tipo di attività nei pazienti comatosi si presenta presto, subito dopo la perdita del ritmo alfa, quando le prime avvisaglie di stato di incoscienza si rendono presenti. La prognosi effettuabile di fronte ad onde con questo pattern dipendono fortemente dall'eziologia dello stato comatoso.



**Fig. 1.3:** Attività delta frontale intermittente le frecce indicano le curve che presentano tale conformazione (Synek, 1988).

### *Onde trifasiche*

Le *onde trifasiche* (TW) sono onde delta (sempre a frequenza 2-3 Hz) smussate e con un alto voltaggio<sup>5</sup> positivo, precedute e seguite da onde negative di bassa ampiezza. Questo tipo di onde si presenta in maniera predominante

<sup>5</sup>Il voltaggio viene misurato in  $\mu\text{V}$  quindi con alto voltaggio si intende un valore superiore a  $70 \mu\text{V}$ .

nelle regioni frontali e si manifesta durante lo stato iniziale del coma. Anche in questo caso fare una prognosi è difficile e dipende largamente dall'eziologia del coma.



**Fig. 1.4:** Onde trifasiche, le frecce indicano le curve che presentano la conformazione (Synek, 1988).

### Attività delta continua ad alto voltaggio

Questo tipo di attività è visibile più tardi rispetto alle onde trifasiche ed all'attività di tipo delta intermittente. Si tratta di onde che denotano un'attività delta di tipo polimorfico, con alta ampiezza e ad alto voltaggio, vanno da 1 a 2 Hz. Alla loro prima comparsa la frequenza può essere più rapida di 1 Hz ma via via che il coma continua la frequenza dominante si assesta e perde la sua reattività a stimolazioni esterne. La prognosi di fronte a questo tipo di onde è meno puntuale rispetto a quella delle due tipologie descritte precedentemente.



**Fig. 1.5:** Attività delta continua ad alto voltaggio (Synek, 1988).

### Scariche periodiche di forma epilettica

Le scariche periodiche di forma epilettica (PLED), sono una delle diverse forme periodiche che si riscontrano nell'EEG di un paziente in stato comatoso. Come dice il nome stesso, queste scariche si presentano asimmetricamente su un solo emisfero o un lobo preciso. Questo tipo di onde compaiono ad intervalli di tempo che vanno da 0,3 secondi a qualche secondo. La prognosi associata ad esse dipende come sempre dall'eziologia del coma ed è molto variabile.

### Scariche periodiche di forma epilettica generalizzate

A differenza delle *scariche periodiche di forma epilettica*, queste son generalizzate all'intero stato comatoso e con il protrarsi dello stesso l'ampiezza delle onde diminuisce e ne aumenta la durata. Le prognosi legate a questa forma d'onda dipendono dalla condizione del paziente, se conseguente ad una over-

dose da medicazione l'outcome risulta buono (recupero). Quando l'eziologia è una forma di anossia<sup>6</sup>, il recupero da parte del paziente è difficile.

### ***Burst suppression***

Uno schema di tipo *burst suppression*<sup>7</sup> è identificabile da picchi di alto voltaggio con frequenza variabile alternati a periodi di mancanza di attività dell'EEG. Questi possono avere durata variabile, generalmente all'approfondirsi dello stato comatoso la durata dei picchi diminuisce e i periodi di attività EEG aumentano. L'eziologia influisce pesantemente sulla prognosi di pazienti che presentano questa forma d'onda. Se presente in seguito ad un arresto cardiaco, l'esito è decisamente peggiore rispetto ad altre cause come ad esempio l'ipotermia.

### **EEG non reattivo, lento a basso voltaggio**

In questo tipo di pattern EEG l'attività predominante è a frequenza di tipo delta o theta e l'ampiezza tendenzialmente risulta inferiore ai 20  $\mu V$ , questo pattern è persistente e non risulta reattivo alle stimolazioni. Questa conformazione si presenta quando vi è un danneggiamento corticale e subcorticale ad esempio conseguente ad encefalopatie anossiche o a severi traumi cerebrali. La prognosi di fronte a questo tipo di onde è generalmente difficile.

---

<sup>6</sup>Con il termine anossia si definisce la diminuzione o totale mancanza di ossigeno a livello cellulare.

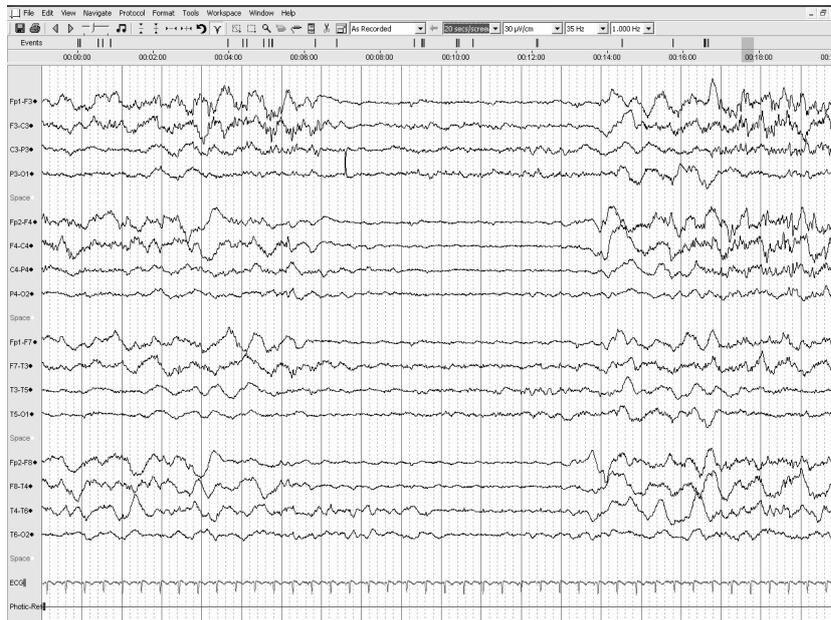
<sup>7</sup>Traduzione letterale: Esplosione e repressione.



**Fig. 1.6:** Scariche periodiche di forma epilettrica, le frecce indicano le curve che presentano la conformazione (Synek, 1988).



**Fig. 1.7:** Scariche periodiche di forma epilettrica generalizzate, le frecce indicano le curve che presentano tale conformazione (Synek, 1988).



**Fig. 1.8:** Burst suppression (Synek, 1988).



**Fig. 1.9:** EEG non reattivo, lento a basso voltaggio. La freccia indica il tracciato non reattivo (Synek, 1988).

## Inattività elettrocerebrale

Questo termine è utilizzato per identificare quei tracciati in cui non è presente alcun tipo di attività elettrica registrabile tramite elettrodi e che quindi sia almeno superiore ai  $2 \mu\text{V}$ ; questa soglia è data dal fatto che in presenza di voltaggio inferiore non è possibile distinguere il segnale dal cosiddetto rumore elettrico. L'inattività elettrocerebrale è il più grave tipo di anomalia riscontrabile in un EEG. Rappresenta un coma irreversibile ed i pazienti che presentano questa forma d'onda tendenzialmente arrivano al decesso o rimangono in uno stato vegetativo persistente.

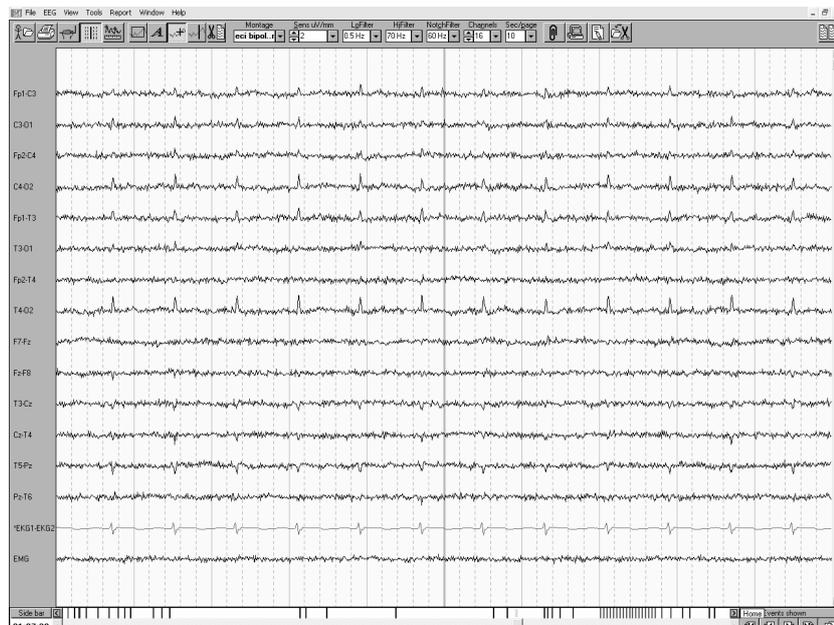


Fig. 1.10: Inattività elettrocerebrale (Synek, 1988).

## *Spindle coma*

Nel coma di tipo *spindle* l'attività predominante è sempre di tipo theta o delta, tuttavia sono frequenti scariche simmetriche (nei due emisferi) e sincrone di frequenza tra 11 e 14 Hz. Se queste onde si osservano per un lungo periodo

il paziente inizia a presentare un movimento degli occhi tipico della fase non REM, tuttavia lo stato di coscienza non muta. Questo tipo di onde spesso suggeriscono la preservazione degli emisferi cerebrali, dunque la prognosi di fronte a questo tipo di coma è spesso favorevole.

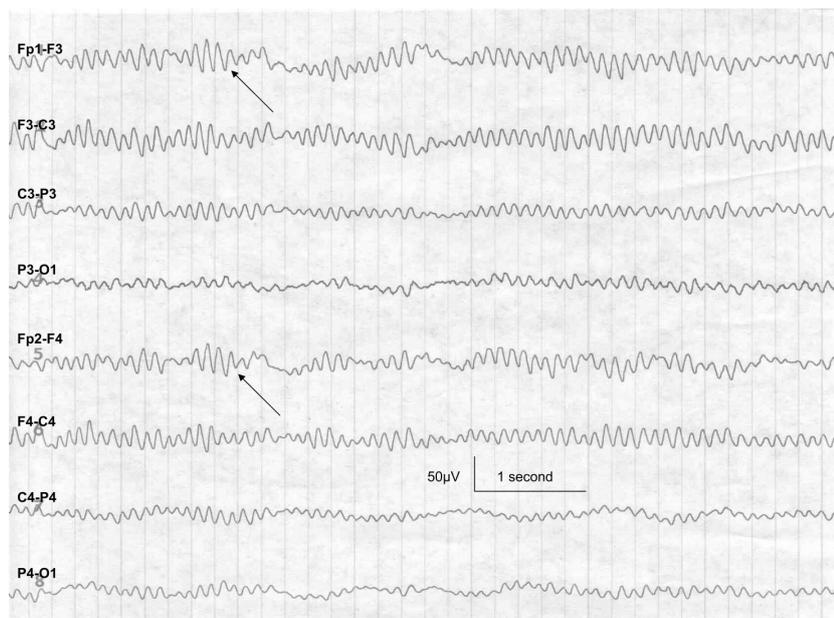


**Fig. 1.11:** *Spindle coma* le frecce indicano le curve che presentano la conformazione (Synek, 1988).

### *Coma alfa*

La dicitura coma alfa è utilizzata per descrivere un tipo di tracciato EEG nel quale vi è un'attività predominante di tipo alfa. In questi casi è necessario valutare la reattività del tracciato utilizzando una stimolazione algica o sonora e valutando se il tracciato si modifica. Generalmente nel paziente che recupera coscienza vi è reattività. Mentre nel paziente che rimane incosciente questa reattività viene a mancare. Eziologicamente questo tipo di coma può presentare tre differenti cause a seconda delle quali la prognosi è molto varia-

bile. In seguito ad un arresto cardiorespiratorio l'attività alfa è generalmente di 8-13 Hz, ad ampiezza da 10 a 50  $\mu\text{V}$ , il recupero da questo stato è difficile e raro. Un'altra causa può essere un'encefalopatia di tipo tossico (assunzione di droghe o grandi dosi di antidepressivi), il pattern è molto simile a quello dell'arresto cardiorespiratorio tuttavia son presenti frequenti onde beta. Con l'attenuarsi dello stato comatoso le onde si tramutano in un coma di tipo theta. Il terzo tipo di coma alfa si ha in pazienti in uno stato *locked-in*<sup>8</sup> in cui l'EEG sembra essere caratterizzato dallo stesso ritmo alfa presente negli individui normali. La prognosi solitamente risulta povera e la maggior parte dei pazienti in questo stato muore.



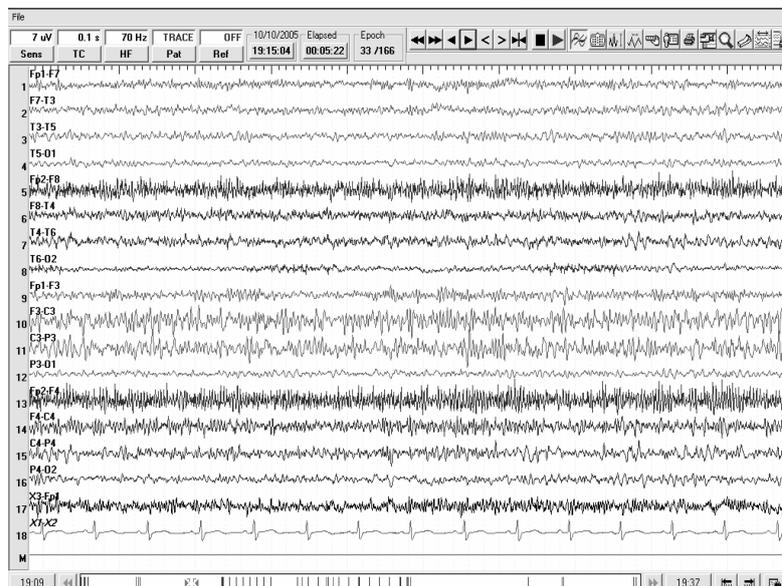
**Fig. 1.12:** Coma alfa, le frecce indicano le onde alfa presenti nel tracciato (Synek, 1988).

---

<sup>8</sup>La sindrome *locked-in* o sindrome del chiavistello è una condizione nella quale il paziente è cosciente e sveglio ma non può muoversi o comunicare a causa della completa paralisi di tutti i muscoli volontari del corpo (Bauer *et al.*, 1979).

## *Coma beta*

Il coma di tipo beta è caratterizzato da onde a frequenza che va dai 12 ai 16 Hz, generalmente con una predominanza nei lobi frontali. Occasionalmente possono essere osservate attività di frequenza di tipo alfa theta o delta. La reattività agli stimoli è mantenuta nel primo stadio del coma ma via via che questo si prolunga va perdendosi. Di solito questo stato è conseguente ad assunzione di sedativi o psicofarmaci. La prognosi è generalmente favorevole se il paziente riesce ad essere supportato durante la fase acuta di intossicazione.



**Fig. 1.13:** Coma beta (Synek, 1988).

### 1.2.2 Scale di valutazione del coma

I metodi di valutazione del coma più utilizzati sono la Glasgow Coma Scale (GCS) e la Glasgow Outcome Scale (GOS). Si tratta di due scale che hanno la funzione di valutare il coma durante la degenza del paziente (GCS) ed il recupero dello stesso una volta dimesso (GOS).

## Glasgow Outcome Scale

La scala di valutazione Glasgow valuta le capacità psicomotorie di soggetti che hanno passato un periodo in stato comatoso. È suddivisa in cinque livelli di classificazione che vanno dal decesso del paziente (livello 1), al *buon recupero* (livello 5) (Jennet e Bond, 1975). Questo processo di valutazione delle abilità alcuni mesi dopo la dimissione riesce a quantificare il grado di recupero del paziente e quindi risulta utile per studi come quello condotto, ma anche per valutare le terapie effettuate durante il ricovero, ecco i cinque livelli.

- Livello 1: Decesso
- Livello 2: Stato Vegetativo Persistente. Il paziente non presenta *ovvie funzioni corticali*<sup>9</sup>.
- Livello 3: Grave disabilità (Cosciente ma disabile). Il paziente dipende dagli altri per il supporto giornaliero, questo è dovuto ad una disabilità mentale, fisica o ad entrambe.
- Livello 4: Disabilità moderata (Disabile ma indipendente). Il paziente è indipendente per quanto riguarda la vita di ogni giorno. Le disabilità presenti possono includere diversi gradi di disfasia<sup>10</sup>, emiparesi<sup>11</sup>, o atassia<sup>12</sup>. Può presentare anche deficit di memoria e sbalzi di personalità.
- Livello 5: Buon recupero. Rispresa delle normali attività sebbene possano presentarsi deficit neurologici o fisici minori.

---

<sup>9</sup>Ritmo veglia - sonno.

<sup>10</sup>Disturbo del linguaggio. Vi sono 3 tipi di disfasie: della comprensione, della produzione e dell'articolazione.

<sup>11</sup>Difficoltà motoria che interessa una parte del corpo.

<sup>12</sup>Progressiva perdita delle capacità motorie dei muscoli volontari.

## Glasgow Coma Scale

Oltre alla scala di valutazione GOS, necessaria per valutare il grado di recupero di un paziente post coma, esiste un metodo di valutazione dell'evoluzione dello stato clinico del paziente durante il coma stesso, la scala di valutazione GCS. Tale metodo si basa su tre tipi di risposta agli stimoli (oculare, verbale e motoria) e si esprime attraverso un numero che è la somma delle valutazioni di ogni singola funzione. Il massimo punteggio è 15, il minimo 3 che denota un profondo stato di incoscienza (Teasdale e Jennet, 1974). In tabella 1.1 è riportata la scala di valutazione GCS in relazione alle differenti risposte agli stimoli.

GCS	Apertura occhi	Risposta verbale	Risposta motoria
1	Nessuna	Nessun suono emesso	Nessuna risposta
2	Allo stimolo doloroso	Suoni incomprensibili	Estensione allo stimolo doloroso
3	Allo stimolo verbale	Parla e pronuncia parole, ma incoerenti	Anormale flessione allo stimolo doloroso
4	Spontanea	Fraasi sconnesse, confusione	Flessione/ritrazione allo stimolo doloroso
5	-	Paziente orientato, conversazione appropriata	Localizzazione dello stimolo doloroso
6	-	-	Obbedisce ai comandi

**Tabella 1.1:** Rappresentazione dei differenti punteggi di Glasgow Coma Scale

L'indice GCS può essere riportato come somma dei tre punteggi oppure in forma analitica (EMV) con i tre punteggi separati. Generalmente le lesioni cerebrali sono classificate in tre modi:

- Grave,  $GCS \leq 8$
- Moderata,  $9 \leq GCS \leq 13$
- Minore,  $GCS \geq 14$

La Glasgow Coma Scale nasce per valutare il livello di coscienza in seguito ad un trauma cranico, successivamente il suo uso è stato esteso a tutti i pazienti in coma acuto, sia con patologie di natura medica che traumatica.

## 1.3 Lavori precedenti

Come si è visto, sono presenti differenti tipi di conformazioni in un tracciato elettroencefalografico. Tali pattern, come già detto più volte, sono particolarmente utili per effettuare previsioni sull'esito della degenza del paziente. L'utilità aumenta se ai pattern EEG si associa l'utilizzo dei potenziali evocati (paragrafo 1.1.2). Sfruttando queste informazioni sorge spontanea una domanda: è possibile utilizzare i tracciati EEG al fine di prevedere la capacità di un paziente, una volta dimesso, di compiere tutte le azioni che svolgeva prima dell'incidente e del conseguente periodo di coma?

### 1.3.1 Il caso studiato

Alla domanda espressa precedentemente Bosco *et al.* (2011) hanno cercato di rispondere in uno studio che prendeva in considerazione 68 pazienti in stato comatoso nel periodo 2007 - 2009, i cui familiari una volta dimessi, son stati sottoposti ad un'intervista telefonica nella quale si chiedevano quali fossero le capacità psicomotorie del familiare a tre mesi dalla dimissione. Le capacità andavano classificate secondo la scala *Glasgow Outcome Scale* (GOS paragrafo 1.2.2). Una volta associata la classificazione secondo la scala GOS a

ciascun paziente, ogni tracciato EEG è stato suddiviso in tre periodi temporali: il momento iniziale, il momento intermedio ed il momento finale. Ogni sezione temporale è stata identificata secondo la classificazione del tracciato EEG data in Synek (1988), ottenendo così tre variabili categoriali che riasumevano il tracciato EEG di ciascun paziente. Il tracciato poteva perciò indicare un processo di miglioramento o di peggioramento. Per valutare l'influenza sulla scala GOS dei vari tipi di pattern associati all'istante temporale in cui si presentavano, è stato realizzato un modello di regressione logistica con il quale è stato possibile osservare il grado di associazione tra EEG e classificazione GOS. Gli autori dell'articolo affermano che:

Utilizzando il modello di regressione logistica è stato osservato che il progressivo deterioramento del tracciato EEG era associato ad un incremento del rischio di decesso con un valore del 24% superiore rispetto ai pazienti il cui tracciato EEG non peggiorava.

Tale affermazione fa scaturire una successiva domanda: Il modello funziona bene utilizzando tre istanti temporali, ma è possibile realizzare un modello che sfrutti l'intero tracciato EEG?

L'intento dell'analisi condotta è stato quello di rispondere a questa domanda, ovviamente lo scenario che si è presentato era decisamente diverso, poichè lavorando sull'intero tracciato l'ambito di analisi diventa continuo e sostanzialmente non più discreto.

### **1.3.2 Altri studi**

In campo medico gli studi riguardanti pazienti affetti da disturbo della coscienza sono molteplici e le tecniche di analisi di tracciati EEG sono altrettanto numerose. Tali tecniche nascono tutte dalla necessità di sintetizzare

i differenti tracciati tramite un metodo di classificazione univoco ed il meno possibile soggetto ad errore, cosa che la sola analisi visiva non riesce ad effettuare. Tra i diversi metodi, quello più comunemente utilizzato per quantificare il contenuto in frequenza dell'attività cerebrale sfrutta l'analisi spettrale dei segnali e consiste nella scomposizione delle onde registrate in fase, ampiezza e frequenza tramite l'utilizzo della trasformata di Fourier. Questo metodo tuttavia non sarà presente in questo lavoro di tesi proprio perché già ampiamente utilizzato in ambiti simili.

Il metodo di analisi utilizzato si basa sullo studio dei dati funzionali, tecnica utilizzata già in altri studi e non ancora sfruttata del tutto nel campo dell'analisi dei tracciati EEG. In particolare tale tecnica è stata utilizzata in due studi del politecnico di Milano per analizzare due casi simili a quello in esame. Nel primo caso (Antonini, 2012) lo studio aveva l'obiettivo di trovare degli indici che permettessero di discriminare tra pazienti in stato vegetativo (VS) e di minima coscienza (MCS) sintetizzando l'informazione contenuta nei tracciati EEG. Nel secondo caso (Cardazzi, 2012) invece, i tracciati (SEEG) venivano utilizzati per identificare delle precise zone del cervello compromesse che davano luogo a delle forme di epilessia chiamate epilessie focali, l'identificazione di tali zone doveva essere il più accurata possibile per permettere la rimozione chirurgica della zona compromessa con il minimo effetto negativo per il sistema nervoso centrale.

Come già detto, in entrambi gli studi l'approccio all'analisi dei tracciati (EEG ed SEEG) avveniva tramite l'utilizzo dell'analisi funzionale, tecnica relativamente nuova e che ben si presta a questo tipo di studi. Per questo motivo si è scelto di utilizzare tale tecnica anche in questo elaborato.

## Capitolo 2

# Functional Data Analysis

Buona parte dei dati provenienti da diversi campi di applicazione si presenta a noi attraverso un processo naturale descrivibile come funzionale (Ramsay e Silverman, 1997).

L'approccio di tipo funzionale all'analisi dei dati è relativamente recente<sup>1</sup> e può essere applicato a tutti quei problemi nei quali l'analisi si concentra su dati caratterizzati da un dominio di tipo continuo.

I problemi analizzabili tramite l'approccio funzionale sono molteplici, alcuni esempi sono l'analisi della variazione di altezza di un campione di individui nel tempo, oppure la temperatura mensile rilevata in diverse stazioni meteo o la variazione di un determinato indice economico nel tempo. Perciò, come si può facilmente intuire, il fattore temporale è predominante nelle analisi di tipo funzionale, essendo questa la componente che si manifesta, in natura, in forma continua.

---

<sup>1</sup>Viene utilizzato per la prima volta nel 1996 da Jim Ramsey, il quale sfrutta tale approccio per analizzare il movimento delle labbra nell'articolo *Functional data analysis of lip motion*. Journal of the Acoustical Society of America, 99, 3718-3727.

L'assunzione di continuità del dominio, quindi, porta ad un approccio diverso da quelli standard per serie storiche che permette l'associazione ad ogni soggetto di una funzione nel tempo della variabile studiata che lo caratterizza.

## 2.1 Cos'è l'analisi dei dati funzionali

La filosofia che sta alla base dell'approccio funzionale, consiste nel considerare ogni funzione osservata come una singola entità e non come una serie di osservazioni individuali. Tuttavia, sebbene il dominio di osservazione del dato sia di natura continua, l'osservazione e la rilevazione di esso avviene in maniera discreta. Il termine funzionale perciò è relativo alla struttura intrinseca del dato e non tanto alla sua osservazione.

### 2.1.1 Rappresentazione dei dati funzionali

Un'osservazione funzionale  $x_i$  per l' $i$ -esimo individuo è composta di  $n$  coppie  $(t_i, y_i)$  dove  $y_i$  rappresenta l'osservazione di  $x(\cdot)$  al tempo  $t_i$ . Sui dati è possibile attuare un processo di *interpolazione* o un processo di *smoothing* se questi risultano affetti da errore.

### 2.1.2 Le basi

Come già detto, lo spirito dell'analisi funzionale è quello di specificare l'intera funzione rappresentante i dati e non soltanto i valori di un numero finito di istanti temporali. Per tale motivo è necessario un metodo che riesca a generalizzare, per ogni caso di studio, dove posizionare dei punti di valutazione della funzione, anche a fronte di misurazioni della stessa variabile ad istanti differenti per individui differenti. Inoltre, spesso, lavorando con funzioni è necessario aver la possibilità di effettuare delle derivate della funzione oggetto

di studio. Per questi motivi, per rappresentare un dato funzionale si rende necessaria la combinazione lineare di un insieme di elementi chiamati basi.

Una base è un insieme di funzioni  $\phi_k$  indipendenti tra loro che permettono di approssimare una funzione attraverso una combinazione lineare di  $K$  di esse, così espressa

$$x(t) = \sum_{k=1}^K c_k \phi_k(t). \quad (2.1)$$

Esistono diversi metodi per specificare delle basi di funzioni.

### **Serie di Fourier**

Le basi più conosciute e più utilizzate per le funzioni di tipo periodico sono ricavate dalla serie di Fourier:

$$\hat{x}(t) = c_0 + c_1 \sin(\omega t) + c_2 \cos(\omega t) + c_3 \sin(2\omega t) + c_4 \cos(2\omega t) + \dots$$

che, con riferimento alla 2.1, identifica le basi:

$$\phi_1(t) = 1$$

$$\phi_2(t) = \sin(\omega t)$$

$$\phi_3(t) = \cos(\omega t)$$

$$\phi_4(t) = \sin(2\omega t)$$

$$\phi_5(t) = \cos(2\omega t)$$

$\vdots$

con la costante  $\omega$  in relazione  $\omega = 2\pi/T$  con il periodo  $T^2$ . Tale sistema di basi è utilizzato soprattutto in presenza di funzioni regolari dove non sono

---

<sup>2</sup>Il periodo è l'intervallo di tempo necessario a compiere un'oscillazione completa della curva.

presenti forti sistematicità locali e dove le curve sono tendenzialmente dello stesso ordine in tutta la funzione. Tuttavia non di rado vengono utilizzate anche per approssimare funzioni non periodiche e non regolari proprio per la loro versatilità e semplicità di calcolo.

### **Basi polinomiali**

Basandosi sul classico monomio  $\phi_k(t) = (t - \omega)^k$ ,  $k = 0, \dots, K$ , se di fronte a istanti  $t_j$  equispaziati, è possibile ottenere una serie di basi polinomiali. Sfortunatamente, però, il prodotto cartesiano di due basi polinomiali  $\phi^T \phi$  può dare luogo ad una matrice singolare, per questo il parametro  $\omega$  va scelto con attenzione. Tuttavia questo tipo di basi, in presenza di tempi  $t_j$  equispaziati e di un valore  $\omega$  scelto opportunamente, implicano un onere computazionale di ordine  $O[(n + m)K]$ .

Come per le serie di Fourier, l'espansione polinomiale non riesce ad interpolare localmente in maniera precisa se non si usa un valore  $K$  molto grande, inoltre le basi polinomiali tendono ad approssimare bene al centro dell'intervallo e peggio agli estremi, per questo non sono indicate per l'estrapolazione o la previsione.

### ***Spline***

Uno dei sistemi di basi più versatile, perché capace di interpolare andamenti locali delle curve, sfrutta l'idea di costruire delle funzioni polinomiali a tratti rifacendosi alle asticcioline flessibili in legno, chiamate appunto *Spline*, utilizzate nei cantieri navali per la progettazione degli scafi. Queste asticcioline, venivano fissate in alcuni punti della sezione trasversale dello scafo mentre nello spazio tra un punto e l'altro erano lasciate libere di disporsi secondo la naturale tendenza; si determinavano così delle curve regolari con comporta-

mento preassegnato in alcune posizioni (Azzalini e Scarpa, 2009). Le *Spline* vengono costruite unendo diversi polinomi in punti  $\tau_k$  detti nodi. Il numero di nodi può essere scelto arbitrariamente o con delle procedure di convalida incrociata, mentre la relazione che regola la scelta del numero di basi è del tipo:

$$\text{numero di basi} = \text{grado del polinomio} + \text{numero di nodi interni}.$$

Dove con numero di nodi interni si intende tutti i nodi tranne quelli posizionati all'inizio ed alla fine del dominio della funzione (Ramsay *et al.*, 2009). L'interno della funzione sarà composta di polinomi di grado fissato  $d$ , e l'unico vincolo imposto per la scelta di questi sarà che polinomi adiacenti abbiano un opportuno grado di lisciamiento nei nodi di giunzione. Per questo verrà scelta una funzione che abbia derivate dal grado 0 al grado  $d - 1$  continue in ognuno dei punti  $\tau_k$ , in tal modo non sarà percepibile la discontinuità presente nei punti di giunzione. Già scegliere dei polinomi di grado  $d = 3$  (*Spline* cubiche) è sufficiente affinché la discontinuità non sia visibile all'occhio umano. Queste condizioni, perciò, si possono esprimere come

$$f(\tau_k) = y_k \quad \text{per } k = 1, \dots, K$$

$$f(\tau_k^-) = f(\tau_k^+), \quad f'(\tau_k^-) = f'(\tau_k^+), \quad f''(\tau_k^-) = f''(\tau_k^+) \quad \text{per } k = 2, \dots, K - 1$$

dove  $g(x^-)$  e  $g(x^+)$  indicano il limite da sinistra e da destra di una funzione  $g(\cdot)$  nel punto  $x$ .

Le *Spline* combinano la facilità computazionale dei polinomi con la capacità di cambiare comportamento localmente. Esistono molti modi di rappresentare le *Spline*, forse concettualmente il più semplice è una combinazione lineare di basi di funzioni

$$\phi_k(t) = (t - \tau_k)_+^d$$

dove  $u_+$  sta ad indicare la funzione che ritorna  $u$  se  $u \geq 0$  e 0 altrimenti, e dove solamente il numero  $K - 1$  di nodi interni viene utilizzato.

Sebbene questo sistema venga ancora utilizzato in alcune applicazioni, ha la tendenza a creare matrici singolari in presenza di un numero elevato di nodi. Nella pratica è meglio utilizzare delle *B-splines*, le quali presentano un supporto più compatto, cioè assumono il valore zero ovunque eccetto in un intervallo finito. Nel caso delle *Spline* cubiche ad esempio, una *B-spline* è una *Spline* cubica con supporto nell'intervallo  $[\tau_{k-2}, \tau_{k+2}]$ . Tale supporto permette sia di risolvere il problema di sigolarità della matrice prodotta, sia fornisce una computabilità di grado  $O(K)$  per ottenere i valori lisciati. La domanda che sorge è come scegliere dove posizionare i nodi  $\tau_k$ .

Sebbene alcune applicazioni suggeriscano dei punti precisi dove posizionare i nodi, la scelta del posizionamento è spesso arbitraria e dettata dall'esperienza. Esistono delle soluzioni che cercano di automatizzare tale scelta iniziando da una funzione con un'alta densità di nodi e via via eliminando i nodi non necessari tramite un algoritmo simile a quello della scelta delle variabili utilizzato nella regressione multipla (Ramsay e Silverman, 2002).

### ***B-splines***

Le *B-splines* sono state pensate originariamente da Curry e Schoenberg (1947) con la seguente definizione:

Consideriamo  $t = t_j$  una sequenza non decrescente (finita o infinita). La  $j$ -esima *B-spline* normalizzata di ordine  $k$  per la sequenza di nodi  $t$  si denoterà come

$$B_{j,k,t}(x) = (t_{j+k} - t_j)[t_j, \dots, t_{j+k}](\cdot - x)_+^{k-1} \quad \text{con } x \in \mathbb{R}.$$

Con  $(\cdot - x)_+^{k-1}$ , si indica che la differenza va calcolata mantenendo fisso il valore di  $x$  e cambiando di volta in volta il valore  $t$ . Il numero che risulterà

dipenderà ovviamente dallo specifico valore di  $x$  scelto. Mentre con  $(x)_+$  si indica la funzione troncamento  $\max(x, 0)$ . Le *B-splines* godono di una serie di proprietà. Per una più approfondita trattazione sulle *B-splines* e le loro proprietà è possibile consultare Boor (2001).

### 2.1.3 Il modello

Solitamente i modelli utilizzati per fare regressione lineare suppongono di avere  $y_1, \dots, y_N$  osservazioni di una variabile risposta ed  $x_1, \dots, x_N$  covariate. Applicando la regressione lineare si cerca di approssimare un modello della forma

$$y_i = \alpha + \sum_j \beta_j x_{ij} + \epsilon_i = \alpha + \langle \beta, x_i \rangle + \epsilon_i. \quad (2.2)$$

Dove  $\langle \beta, x_i \rangle$  indica il prodotto scalare tra il vettore  $\beta$  e quello delle covariate relative all'unità  $i$ .

Nell'ambito funzionale, l'estensione di tale modello è abbastanza semplice. Considerando sempre la variabile risposta  $y_i$  di tipo scalare e le covariate di tipo funzionale in un dominio temporale  $[0, T]$ , basterà sostituire il parametro  $\beta$  nell'equazione precedente con una funzione  $\beta$ , il modello sarà quindi così espresso

$$y_i = \alpha + \int_0^T \beta(s) x_i(s) ds + \epsilon_i = \alpha + \langle \beta, x_i \rangle + \epsilon_i. \quad (2.3)$$

Nella regressione lineare classica solitamente si assume che le variabili indipendenti siano osservate senza errore, tale condizione viene rispettata per quanto riguarda la regressione funzionale assumendo che, eventualmente, i dati siano stati preprocessati e lisciati. Per il calcolo del parametro  $\beta$  l'approccio classico di regressione lineare si basa sulla minimizzazione della somma degli scarti al quadrato, ed anche per quanto riguarda l'approccio funzionale si cerca di fare lo stesso non dimenticando le peculiarità dei dati

funzionali. Il calcolo perciò sarà

$$LMSSSE(\alpha, \beta) = \sum_{i=1}^N (y_i - \alpha - \langle \beta, x_i \rangle)^2 = \|y - \alpha - \langle \beta, x \rangle\|^2. \quad (2.4)$$

Dove  $x$  rappresenta il vettore delle funzioni covariate  $(x_1, \dots, x_N)^T$ , mentre l'operatore  $\|\cdot\|$  indica la norma euclidea del vettore calcolato.

Come si può facilmente notare, ci troviamo di fronte ad un sistema di equazioni diverso dal solito. Il coefficiente funzionale  $\beta$  è apparentemente indeterminabile, perché essenzialmente vi è un infinito numero di parametri  $\beta(s)$  ed un numero finito di condizioni  $y_i = \alpha + \langle \beta, x_i \rangle$  da soddisfare. Perciò, finché ci troviamo di fronte ad uno spazio delle funzioni che soddisfano l'equazione (2.4) infinito-dimensionali, indipendentemente dalla numerosità del campione, la minimizzazione della somma dei residui non produrrà uno stimatore consistente del parametro  $\beta$  del modello. Perciò per fornire un buon stimatore o almeno identificare unicamente la funzione  $\beta$  è necessario applicare dei metodi di *regolarizzazione*. L'approccio di *regolarizzazione* più semplice è quello che consiste nella discretizzazione della funzione in una serie finita e più 'grossolana' di punti sulla scala temporale, diminuendo così il numero di incognite da calcolare, ne esistono però diversi tipi e quelli più utilizzati sono principalmente due: *regolarizzazione tramite troncamento delle basi* e *regolarizzazione tramite roughness penalties*.

### **Regolarizzazione tramite troncamento delle basi**

Metodo abbastanza intuitivo, consiste nel diminuire il numero di basi scegliendone  $K < M$ , con  $M$  numero massimo di basi che si vuole proporre. Tale  $K$  dev'essere abbastanza grande da non comportare alcuna significativa perdita d'informazione.

Una volta scelto il valore  $K$ , supponiamo  $\phi = (\phi_1, \dots, \phi_K)^T$  sia un vettore di basi e che la funzione di regressione  $\beta$  si sviluppi come

$$\beta = \sum_{v=1}^K b_v \phi_v = b^T \phi$$

dove  $b^T$  rappresenta la trasposizione del vettore contenente i coefficienti della funzione di regressione  $\beta$ . L'espansione delle covariate sarà invece espressa come

$$x_i = \sum_{v=1}^K c_{iv} \phi_v = c_i^T \phi.$$

dove  $c_i$  sarà il vettore contenente i coefficienti relativi alle basi  $\phi$ .

Per poter generalizzare questo metodo anche a basi non necessariamente ortonormali (come ad esempio le *B-splines*), è necessario definire la matrice  $J$  composta da valori  $J_{ik}$  così calcolati

$$J_{ik} = \int \phi_j(s) \phi_k(s) ds = \langle \phi_j, \phi_k \rangle$$

utilizzando tale matrice, il prodotto scalare  $\langle \beta, x_i \rangle$  necessario nella 2.4 per la minimizzazione della somma degli scarti al quadrato sarà

$$\langle x_i, \beta \rangle = \sum_{j=1}^K \sum_{v=1}^K c_{ij} J_{jv} b_v.$$

Se consideriamo  $Z$ , la matrice contenente i coefficienti del modello, questa sarà composta dal prodotto vettoriale tra la matrice dei coefficienti  $C$  e la matrice  $J$ , al quale sarà giustapposto un vettore costante di valore 1 relativo ai coefficienti dell'intercetta, quindi espressa come  $Z = [1|CJ]$ .

Il modello (2.2) perciò, sarà riscritto semplicemente come

$$\hat{y} = Z\xi$$

con  $\xi = (\alpha, b_0, \dots, b_{M-1})^T$  vettore dei parametri i cui valori saranno la soluzione dell'equazione

$$Z^T Z \hat{\xi} = Z^T y.$$

## Regolarizzazione tramite *roughness penalties*

Questo approccio cerca di limitare l'eccessiva fluttuazione locale della funzione stimata, per farlo viene definita la somma dei quadrati degli scarti penalizzata

$$\begin{aligned} PENSSSE_\lambda(\alpha, \beta) &= \sum_{i=1}^N [y_i - \alpha - \langle x_i, \beta \rangle]^2 + \lambda \int_T [D^2\beta(s)]^2 ds \\ &= \|y - \alpha - \langle x_i, \beta \rangle\|^2 + \lambda \|D^2\beta\|^2. \end{aligned}$$

Dove l'espressione  $\int_T [D^2\beta(s)]^2 ds$ , quantifica la variazione della funzione  $\beta$  attribuendo un peso a quanto questa si discosta dalla retta.

Una funzione con variabilità alta avrà un valore grande, viceversa con bassa variabilità il valore sarà piccolo, questo comportamento è dovuto alla derivata seconda presente nell'espressione. Il parametro  $\lambda > 0$  invece gestisce il trade-off tra penalizzazione e variabilità. Questo può essere scelto sia arbitrariamente sia con tecniche di convalida incrociata. Ad esempio, fissati  $\alpha_\lambda^{(-j)}$  e  $\beta_\lambda^{(-j)}$  stime di  $\alpha$  e  $\beta$  ottenute minimizzando la somma dei quadrati degli scarti penalizzata basata su tutti i dati tranne  $(x_j, y_j)$ , è possibile definire la funzione di convalida incrociata (*cross validation*)

$$CV(\lambda) = \sum_{j=1}^N \left( y_j - \alpha_\lambda^{(-j)} - \langle x_j, \beta_\lambda^{(-j)} \rangle \right)^2$$

minimizzando  $CV(\lambda)$  rispetto al valore  $\lambda$  si ottiene il valore ottimale di tale parametro.

## 2.2 Perché scegliere l'analisi dei dati funzionale

Nel paragrafo precedente è stata esposta l'idea di analisi funzionale dei dati e di come questo tipo di analisi sia versatile e adatta all'utilizzo in diversi

ambiti. Si tratta di un approccio relativamente recente, e in letteratura non sono presenti casi di utilizzo dell'analisi funzionale per l'interpretazione di tracciati EEG relativi a periodi più lunghi di qualche minuto (paragrafo 1.3.2). Tuttavia la scelta è ricaduta sull'analisi funzionale dei dati poiché questa risultava essere una delle poche soluzioni non standard che potevano sfruttare a pieno l'informazione presente nei dati analizzandoli nel loro reale dominio e cioè nel continuo.

Questo tipo di analisi si presta molto all'interpretazione visiva delle informazioni ed all'identificazione di *pattern* che possono aiutare a valutare l'implicazione delle variabili scelte a determinati istanti temporali. Questa peculiarità dell'analisi funzionale si sposa con la tecnica classica di analisi dei tracciati operata dai neurofisiologi, i quali osservando il tracciato riescono ad identificare dei *pattern* conosciuti ed a fare delle previsioni sull'esito del coma.



# Capitolo 3

## Il dataset studiato

Dopo aver esposto gli strumenti utilizzati nell'analisi e l'idea che ha portato alla loro scelta, passiamo al dataset oggetto di studio.

I dati analizzati provengono dalla stessa popolazione di quelli considerati nell'articolo citato nel paragrafo 1.3. Di questi, per motivi computazionali spiegati in seguito (paragrafo 3.1) sono stati considerati soltanto 34 pazienti. Si tratta di pazienti in stato di coma al momento di ammissione al reparto di neurochirurgia intensiva dell'ospedale di Treviso tra il 2007 ed il 2009. L'inclusione nello studio era subordinata all'accettazione da parte dei familiari, ed alla presenza di emorragia subaracnoidea<sup>1</sup> e/o emorragia intracerebrale<sup>2</sup>. Inoltre dovevano presentare un grado di Glasgow Coma Scale<sup>3</sup>  $\leq 8$ . I familiari dei pazienti, tre mesi dopo la dimissione, dovevano prendere parte ad un'intervista telefonica per valutare lo stato di recupero tramite scala GOS (paragrafo 1.2.2). L'obiettivo del lavoro svolto è stato proprio prevedere il

---

<sup>1</sup>SAH: emorragia cerebrale situata tra le due membrane che circondano il cervello (aracnoide e pia madre), solitamente si manifesta in seguito ad un trauma cranico.

<sup>2</sup>ICH: consiste generalmente nella rottura, all'interno del tessuto cerebrale, di un vaso sanguigno, ed è più frequentemente localizzata nel cervelletto.

<sup>3</sup>GCS: vedi cap. 1.2.2

valore di tale scala sfruttando il tracciato EEG di ogni singolo paziente.

La popolazione considerata presenta una media di 7,79 giorni di ricovero con una deviazione standard di 5,20 giorni. Solo in quattro casi il ricovero è durato più di 13 giorni, mentre per la maggior parte dei ricoverati (28 pazienti) la durata si attesta tra 2 e 11 giorni (82%), come visibile in tabella 3.1.

<b>Giorni</b>	<b>Num. pazienti</b>
[0 – 5)	12
[5 – 10)	14
[10 – 15)	5
[15 – 20)	2
[20 – 25]	1

**Tabella 3.1:** Durata del ricovero in giorni.

Osservando quanto riportato in tabella 3.2, si può notare che per due classi è presente un solo paziente (classe 2 e classe 5). Per tale motivo è stata divisa la variabile in tre classi aggregando la classe 1 e la classe 2 nella nuova classe 1 e la 4 e la 5 nella nuova classe 4. Ottenendo così una nuova variabile con tre modalità corrispondenti alle vecchie classi 1-2, 3, 4-5.

<b>Valore GOS</b>	<b>Num. pazienti</b>
1	9
2	1
3	10
4	13
5	1

**Tabella 3.2:** Distribuzione pazienti secondo valore GOS.

Per quanto riguarda il genere, i pazienti risultano equidistribuiti. Mentre la suddivisione del campione per fasce d'età appare sbilanciata verso pazienti con una fascia d'età compresa tra i 40 ed i 70 anni, è presente solamente un paziente di 18 anni, e due tra i 30 ed i 40, come si può vedere in tabella 3.3.

<b>Età</b>	<b>Num. pazienti</b>
[18, 20)	1
[20, 30)	0
[30, 40)	2
[40, 50)	10
[50, 60)	8
[60, 70)	9
[70, 80]	4

**Tabella 3.3:** Distribuzione pazienti per fasce d'età.

I pazienti son stati monitorati per tutto il periodo di coma e la registrazione non è stata interrotta se non di fronte alla presenza del rischio di complicazioni e quindi alla necessità di cure specifiche. Va detto che la registrazione elettroencefalografica in alcuni casi è risultata molto sensibile alle sedazioni. In particolare conseguentemente alla somministrazione di particolari sedativi (Propofol e Remifentanil) si sono rilevate delle alterazioni dell'EEG (Burst Suppression). Identificare tali variazioni risulta abbastanza difficile, solo un occhio esperto sarebbe in grado di farlo, e qualora si identificassero non sarebbe possibile distinguere se le variazioni sono effetto dell'assunzione di sedativi o meno poichè ai tracciati non sono associate le somministrazioni di medicinali nei differenti istanti temporali. Tuttavia si potrebbe pensare un modello capace di individuare queste alterazioni per poterle eliminare dal tracciato. In questo caso si è preferito non implementare tale modello

utilizzando il tracciato come si presentava.

### 3.1 Variabili considerate

Le variabili considerate nell'analisi si riferiscono ad una precisa area dell'encefalo, quella frontale. I tracciati osservati, registrano l'attività dei lobi frontali sinistro e destro (F3 ed F4 vedi pragrafo 1.1). La scelta di tali zone è stata dettata dal fatto che in queste aree si concentrano la maggior parte delle configurazioni comatose conosciute, inoltre considerando altre zone si sarebbe appesantito notevolmente il già oneroso carico computazionale.

Sulla scelta delle variabili hanno influito molti fattori, in particolare la sempre maggiore capacità di immagazzinamento delle informazioni ha giocato un ruolo fondamentale.

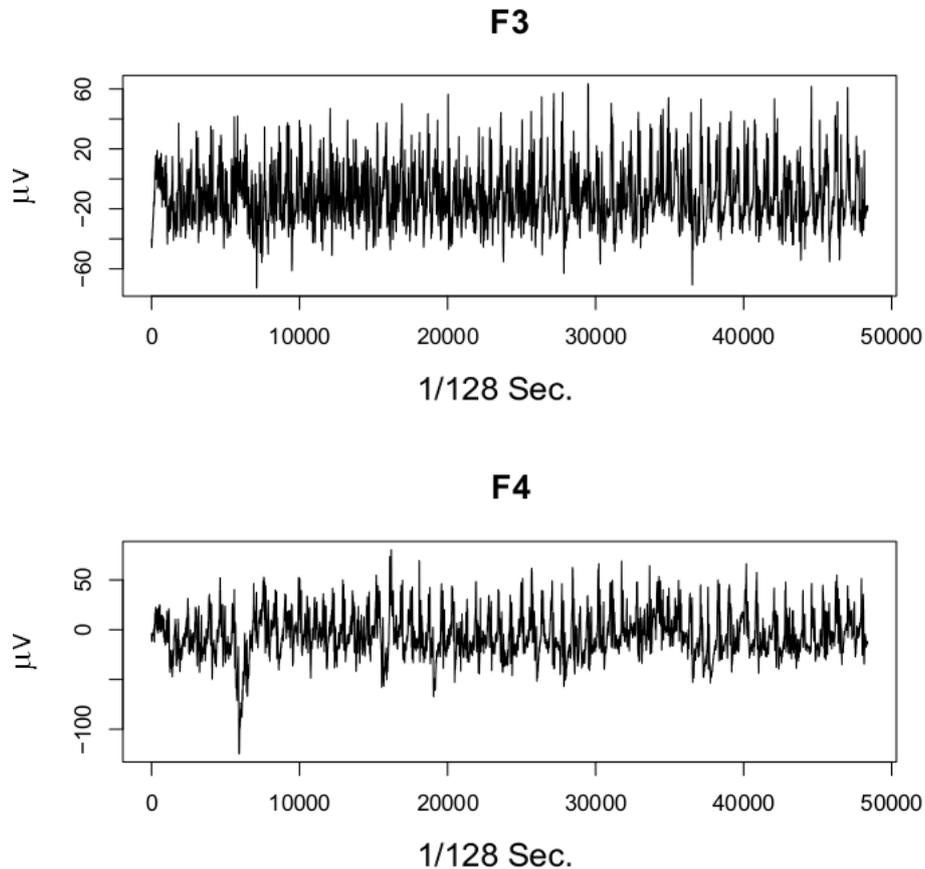
Uno strumento come quello utilizzato nell'ospedale di Treviso è in grado di effettuare 128 rilevazioni (in  $\mu V$ ) al secondo. Questo, corrisponde a 7.680 rilevazioni al minuto e 460.800 all'ora. Considerando, come detto in precedenza, che mediamente un paziente presentava un periodo di coma di otto giorni, si arriva ad avere una media di 88.473.600 rilevazioni per paziente, per singolo tracciato <sup>4</sup>. Questo porta presto ad identificare con quale tipo di variabili il lavoro è stato condotto e qual è stata la sfida principale per lo studio.

In figura 3.1 vi è un esempio dei primi sei minuti (46.080 rilevazioni) di tracciato EEG come si presentava prima delle sintesi operate sui dati. Una tale mole di dati è di certo difficile da gestire con un comune PC domestico, ed anche ricorrendo all'utilizzo di sistemi multiprocessore e con elevate quantità

---

<sup>4</sup>Va ricordato che per ogni paziente i tracciati sono due (Settori F3 ed F4).

di RAM, il carico computazionale richiesto risulta decisamente molto oneroso. Quindi come ridurre il carico computazionale senza perdere informazione?



**Fig. 3.1:** Primi 6 min di tracciato EEG grezzo di un paziente.

L'idea è quella di definire un numero ridotto (due) di variabili per ogni minuto, che riassumano le informazioni rilevanti presenti nei dati. L'approccio utilizzato, si basa su quanto osservato in merito ai diversi tipi di classificazioni esposte nel paragrafo 1.2.1. Studiandole si può notare che le variabili analizzate in fase di diagnosi sono sostanzialmente due: la frequenza dell'onda misurata in Hz e la differenza di potenziale elettrico tra i picchi presenti. Riuscendo ad estrapolare questi due indicatori per ogni secondo registrato, si

attua una notevole sintesi della mole di dati, senza perdere tuttavia l'informazione presente. Ciononostante questa soluzione risulta ancora insufficiente, poiché l'approccio di tipo funzionale esposto nel capitolo 2 e reso disponibile dalla libreria FdA di R, presenta notevoli problemi di calcolo anche di fronte all'utilizzo di un sistema con elevate capacità computazionali come *Hactar*<sup>5</sup>.

Un ulteriore passo verso la soluzione del problema computazionale, quindi, si basa sull'assunzione che ogni pattern identificato durante un periodo di coma si propaga nel tempo per qualche secondo fino a raggiungere anche la durata di un minuto. Assumendo tale ipotesi, sembra ragionevole pensare che ridurre l'osservazione di ogni minuto al primo secondo di tracciato potrebbe ben riassumere il periodo di coma senza perdere troppe informazioni. A questo va aggiunto che il periodo di coma da analizzare per i vari pazienti differiva di parecchi giorni, ed il problema computazionale si ripresentava con l'aumentare della finestra temporale analizzata. Come primo passo perciò ci si è concentrati su una finestra temporale pari a 15 ore, cioè il periodo di coma più corto riscontrato nel campione. Successivamente si è scelto di analizzare un periodo di 24 ed infine di 48 ore, valido per la totalità dei pazienti ed in grado di non sovraccaricare troppo il processo di calcolo. I tempi di calcolo dei differenti modelli in base alla finestra temporale utilizzata con associati i tempi di esecuzione del sistema, sono riportati in tabella 3.4.

Alla scelta delle variabili va ad aggiungersene un'altra, quella cioè di considerare maggiormente significative informazioni temporalmente più vicine al termine dello stato comatoso. Per tale motivo è stato necessario invertire i vettori contenenti i valori misurati, per poter attribuire un peso informativo maggiore alle informazioni recenti rispetto a quelle temporalmente più

---

<sup>5</sup>Hactar è una macchina composta da 1 server e 7 *blade* (lame) dotata di 64 CPU a 64 bit con frequenza 2.6 GHz ciascuno e 16 GB di RAM per *blade*.

Finestra temporale (ore)	Tempo di esecuzione (ore)
15	1
24	4
48	25
96	> 60*

**Tabella 3.4:** Tempi di esecuzione del sistema in base alla finestra temporale studiata.

\*Durante la stima del modello lo script termina l'esecuzione senza fornire risultati.

distanti.

### 3.1.1 Le variabili

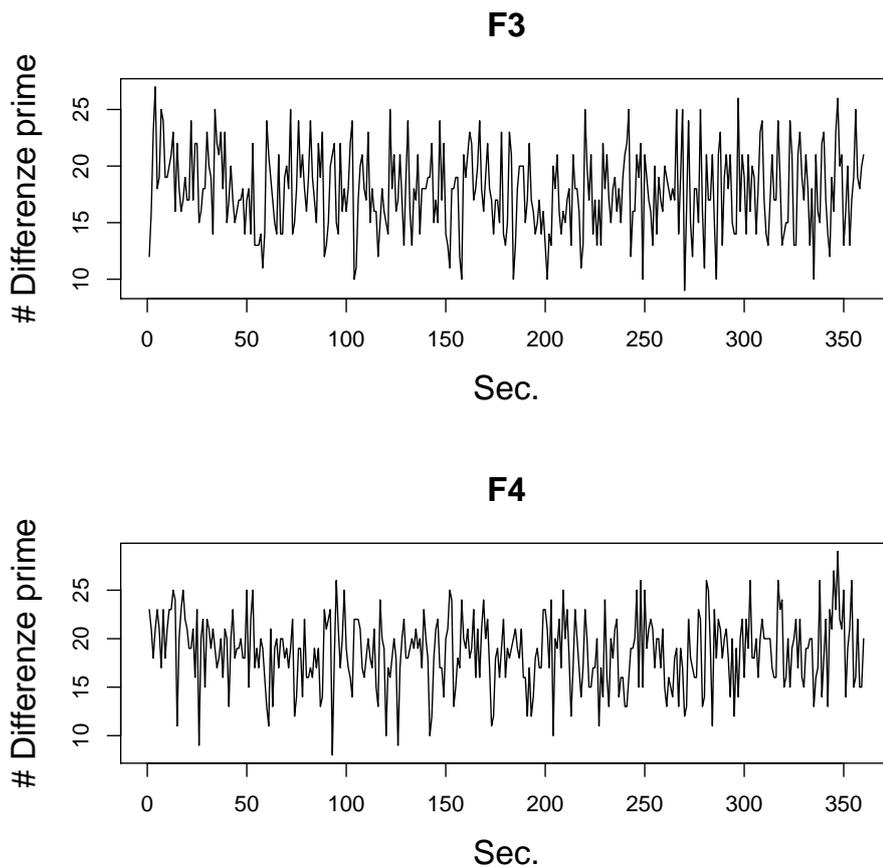
Come detto, la prima sintesi operata sul dataset è volta a creare due variabili che ben riassumano il dataset senza tuttavia perdere l'informazione che contiene. Per tale motivo sono state operate due sintesi:

- Sintesi tramite l'utilizzo delle differenze prime
- Sintesi tramite il calcolo della differenza di voltaggio

#### Sintesi del dataset tramite differenze prime

Per calcolare una variabile che sintetizzasse il numero di cicli ad ogni secondo di rilevazione, si è scelto di utilizzare il metodo delle differenze prime. Tale metodo si basa sul principio che se ci troviamo di fronte ad un cambio di segno nella differenza dei valori corrispondenti a due istanti di tempo successivi  $(x_i, x_{i+1})$ , allora siamo di fronte ad un cambio di tendenza della curva. Se la differenza è negativa passo da una curva crescente ad una decrescente viceversa per una differenza positiva (vedi funzione *diffPrime* appendice A). Contando il numero di cambi di pendenza della curva, si riesce ad ottenere

un numero intero che ne quantifica la frequenza al secondo come visibile in figura 3.2. Questo metodo è abbastanza semplice ed intuitivo e riesce a dare un buon indice della quantità di cicli al secondo pur essendo di fronte a curve come quelle elettroencefalografiche, che non presentano alcuna periodicità.

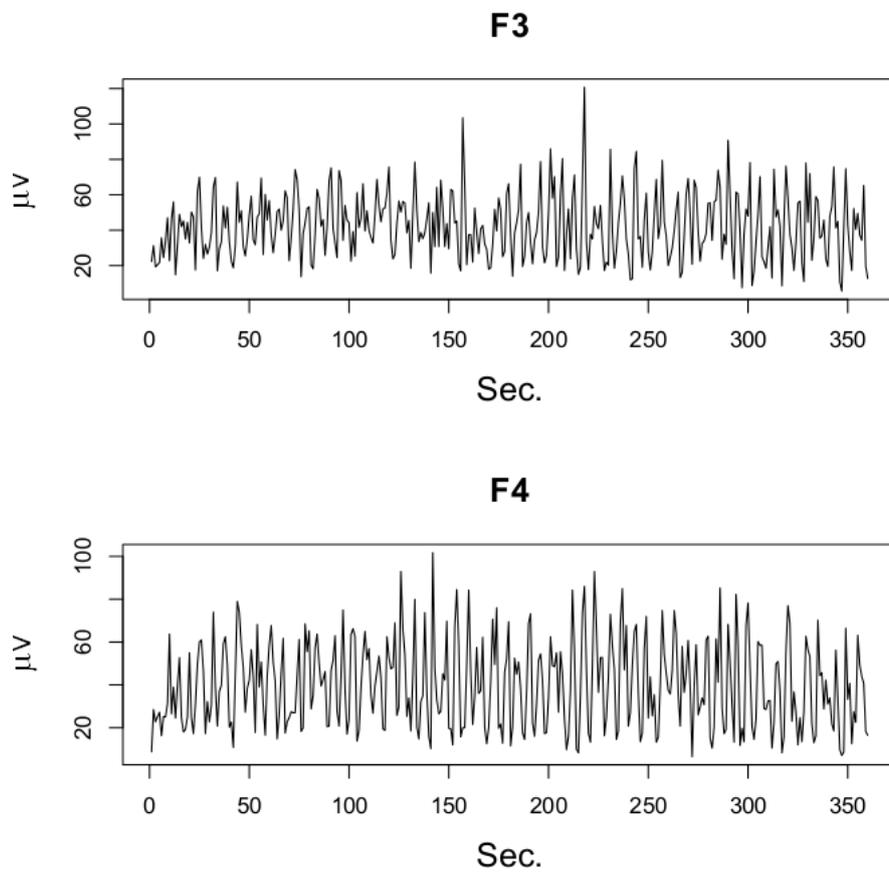


**Fig. 3.2:** Primi 6 min. tracciato EEG riassunti tramite differenze prime.

### Sintesi del dataset tramite differenza di voltaggio

Un'altra variabile estratta dal tracciato, è quella che valuta l'escursione che il voltaggio ha durante un secondo di registrazione. Per calcolarla si sono considerati per ogni secondo tutti i 128 voltaggi registrati, ed è stata calcolata

la differenza tra il voltaggio massimo ed il minimo ottenendo così il valore desiderato (vedi funzione *scarto* appendice A) un esempio del risultato è presente in figura 3.3.



**Fig. 3.3:** Primi 6 min. tracciato EEG riassunti tramite differenza di voltaggio.

Questa variabile, come quella ottenuta tramite differenze prime, è stata calcolata per ogni secondo di tracciato al fine di mantenere la natura funzionale dei dati pur riducendone la quantità.

## 3.2 Studio dei dati funzionali

Una volta definite le variabili da analizzare e le operazioni su di esse per rendere il carico computazionale meno oneroso, si è proceduto con l'analisi dei dati funzionali. Come primo passo necessario si è dovuto scegliere la quantità ed il tipo di basi da utilizzare per poter approssimare al meglio le funzioni relative ai tracciati di ogni singolo paziente. Questi non presentavano alcuna periodicità, quindi non era consigliabile utilizzare delle basi calcolate tramite serie di Fourier. Anche le basi di tipo polinomiale sono state escluse poiché non sono in grado di adattarsi al meglio a funzioni molto variabili se non tramite l'utilizzo di un polinomio di grado elevato. La scelta è perciò ricaduta su delle basi realizzate tramite *Spline* di regressione cubiche. Si è scelto un numero di nodi che permettesse di avere intervalli di lunghezza unitaria perciò in quantità pari al numero di istanti temporali registrati.

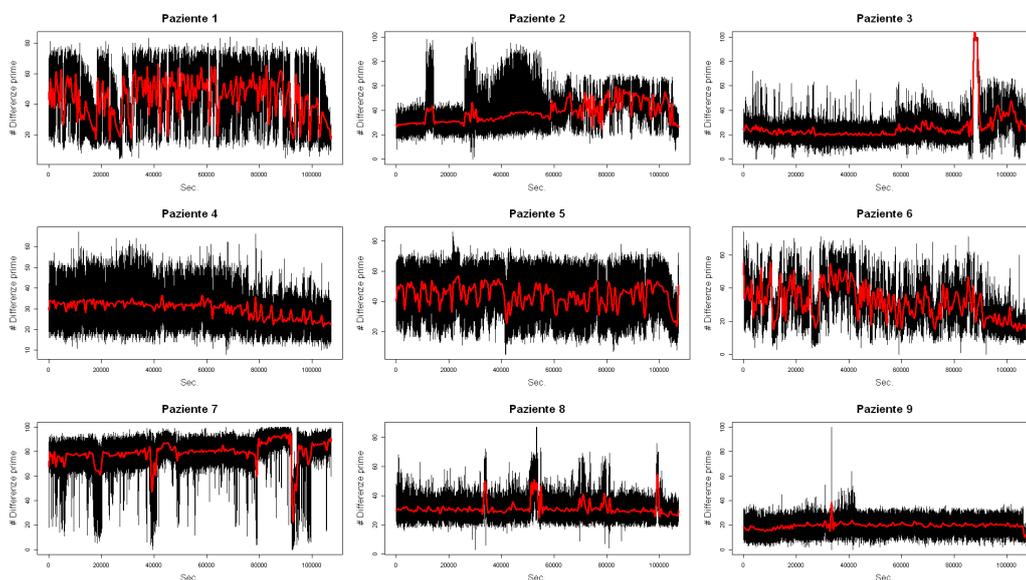
### 3.2.1 *Smoothing* dei dati

Il lisciamiento dei dati solitamente viene utilizzato per ripulire una funzione da errori di osservazione, come potrebbero essere quelli di uno strumento non sufficientemente accurato. Nel nostro caso i dati risultavano già privi di errori ma ancora molto variabili e, nonostante la variabilità fosse caratteristica stessa del tipo di dato, questa non faceva altro che appesantire l'analisi senza portarvi beneficio. Inoltre la riduzione degli errori di misurazione tramite una corretta taratura dello strumento non garantisce la totale assenza di 'rumore'. Per questo una scelta quasi obbligata è stata quella di effettuare dello *Smoothing* tramite *Spline*. A tal fine si è scelto di utilizzare delle *Spline* di lisciamiento (Azzalini e Scarpa, 2009).

L'interesse principale di un'analisi funzionale sta nell'interpretazione del-

la curva associata all'unità statistica e quindi una curva la cui variabilità presenta del 'rumore' o è molto influenzata dal caso o è poco interpretabile e non può essere utile ai fini analitici.

Si è operato uno *Smoothing* dei tracciati applicando la classica convalida incrociata leave one out resa disponibile dal software R ed il risultato è quello raffigurato in figura 3.4.



**Fig. 3.4:** *Smooth* dei dati tramite *Spline* di lisciamento (Azzalini e Scarpa, 2009), primi 9 pazienti.

Come visibile dalla figura i tracciati risultavano molto variabili, lo *Smoothing* degli stessi ha ricondotto le curve ad una dimensione più interpretabile pur mantenendone le caratteristiche che le differenziavano. Va notato inoltre che tutte le conformazioni caratteristiche dei differenti stati comatosi vengono mantenute anche dopo l'operazione di lisciamento seppur ricondotte ad un valore medio.

### 3.2.2 Intervalli di confidenza per il coefficiente

Una volta operato il lisciamento dei dati e condotta una regressione funzionale, per una corretta rappresentazione dei dati è necessario associarvi un intervallo di confidenza che misuri la precisione con cui sono stati stimati i vari  $\hat{\beta}_j(t)$ . In letteratura non sono riportati molti metodi per il calcolo di tale intervallo. Il più semplice ed utilizzato, si basa sulla classica assunzione che gli  $\epsilon_i$  siano indipendenti e identicamente distribuiti attorno allo 0 con varianza  $\sigma_\epsilon^2$  e distribuzione normale. Va notato che tale assunzione potrebbe risultare poco rispettata nel nostro caso, tuttavia, se rispettata, la covarianza di  $\epsilon$  si presenterà come

$$\Sigma = \sigma_\epsilon^2 I.$$

Utilizzando il metodo delta per il calcolo della varianza campionaria del parametro  $\hat{b}$  questa sarà

$$\text{Var}[\hat{b}] = (Z^T Z + R(\lambda))^{(-1)} Z^T \Sigma Z (Z^T Z + R(\lambda))^{(-1)}$$

dove  $R(\lambda)$  è la matrice di penalizzazione utilizzata per la stima del coefficiente  $\beta$ . Dai residui si potrà ottenere una stima di  $\sigma_\epsilon^2$ . Calcolati perciò tutti i valori necessari, l'intervallo sarà ottenuto sommando e sottraendo a  $\beta(t)$  due volte lo standard error.

Gli intervalli di confidenza così ottenuti, però, si basano su ipotesi troppo restrittive, per questo motivo è utile definire un test che si basi su ipotesi meno vincolanti come un test di permutazione.

### 3.2.3 Test di permutazione

Per verificare l'effettiva relazione tra la variabile esplicativa e le variabili risposta è sorta la necessità di realizzare un test ad hoc, poiché la natura

delle statistiche sui dati funzionali non dà modo di derivare una distribuzione teorica nulla sulla quale verificare le ipotesi. Per questo motivo si è deciso di ricorrere ad un test non parametrico.

Si è scelto un test di permutazione, il quale si basa sull'assunzione che se non vi fosse una relazione effettiva tra la variabile risposta e le covariate, non vi dovrebbe essere differenza se questo fosse ricalcolato associando in modo casuale le stesse. Basandosi su questa ipotesi, risulta abbastanza semplice creare una distribuzione nulla empirica, questa sarà ottenuta permutando le variabili risposta senza però modificare le esplicative, e di volta in volta ristimando il modello. Successivamente sarà quindi possibile ricalcolare il test ed avere una distribuzione nulla dello stesso. Andando a confrontare il valore ottenuto con la distribuzione sarà perciò visibile quanto il risultato ottenuto è distante dall'ipotesi nulla. Il calcolo del valore  $p$  si effettua semplicemente contando la proporzione di valori permutati che si trovano più distanti dal valore del test calcolato.

Il vantaggio di tale approccio è che il test può così prescindere da qualsiasi assunzione sulla distribuzione, lo svantaggio è che risulta molto difficile se non impossibile verificare la significatività di una sola covariata rispetto alle altre.

Nella libreria `fda` di R è disponibile un test di permutazione che si basa sulla statistica.

$$F = \frac{Var[\hat{y}]}{\frac{1}{n} \sum (y_i - \hat{y}_i)^2}$$

Questa differisce dalla classica statistica F poiché normalizza la somma dei quadrati degli scarti al numeratore ed al denominatore. Tale statistica è l'unica fornita e non risulta adatta per la variabile di tipo categoriale ordinale oggetto di studio quale è il GOS (paragrafo 1.2.2). Per tale motivo si è deciso di modificare la funzione presente nel pacchetto utilizzando una statistica specifica per questo tipo di variabile risposta, al fine di non incorrere in

errori di valutazione del modello dovuti alla scelta di un test non adatto.

Come già detto, la variabile risposta è una variabile di tipo categoriale ordinale, pensando ad un test adatto per tale tipo di variabili la scelta è ricaduta sul quello di Wilcoxon-Mann-Whitney sulla somma dei ranghi. Tale test viene solitamente utilizzato per verificare se il rango medio di due popolazioni differisce, in pratica può essere utilizzato come alternativa al test t di Student per dati appaiati. Per tale test si assume che i dati si presentino a coppie e provengano dalla stessa popolazione, che ogni coppia sia scelta casualmente ed indipendentemente e che i dati siano di tipo ordinale. Il sistema di ipotesi su cui si basa è il seguente:

- $H_0$ : La differenza media tra le coppie è pari a 0
- $H_1$ : La differenza media tra le coppie è diversa da 0

Pur essendo già presente in *R* una statistica test che utilizzasse il test di Wilcoxon, è stato necessario realizzarne una versione di permutazione ex novo adatta alle esigenze dello studio condotto. Il codice necessario alla realizzazione di tale versione è riportato in appendice B contenente i comandi *R* utilizzati.

### **La statistica Test di Wilcoxon**

Trattandosi di una statistica test di tipo non parametrico, la procedura di calcolo della stessa è composta dai passi di seguito riportati:

1. Per  $i = 1, \dots, N$  dove  $N$  è il numero di coppie, si calcola  $|x_{2,i} - x_{1,i}|$  e  $sgn(x_{2,i} - x_{1,i})$ .
2. Si escludono le coppie con  $|x_{2,i} - x_{1,i}| = 0$  e si considera il nuovo campione  $N_r$  così ottenuto.

3. Si ordinano le  $N_r$  coppie in ordine crescente e vi si assegnano dei ranghi ( $R_i$ ) a partire dal rango 1 assegnato al valore più piccolo.
4. Infine si calcola la statistica

$$W = \sum_{i=1}^{N_r} |[sgn(x_{2,i} - x_{1,i}) \cdot R_i]|$$

Questo procedimento verrà svolto per ogni permutazione delle variabili per confrontarle con quelle stimate attraverso il modello, al fine di verificare la distribuzione sotto  $H_0$  della statistica test considerata. Successivamente la distribuzione verrà confrontata con il valore della statistica ottenuta, per vedere quanto questa si discosta dall'ipotesi nulla e cioè di assenza di relazione tra le variabili esplicative e i valori stimati.

### 3.2.4 Un semplice modello

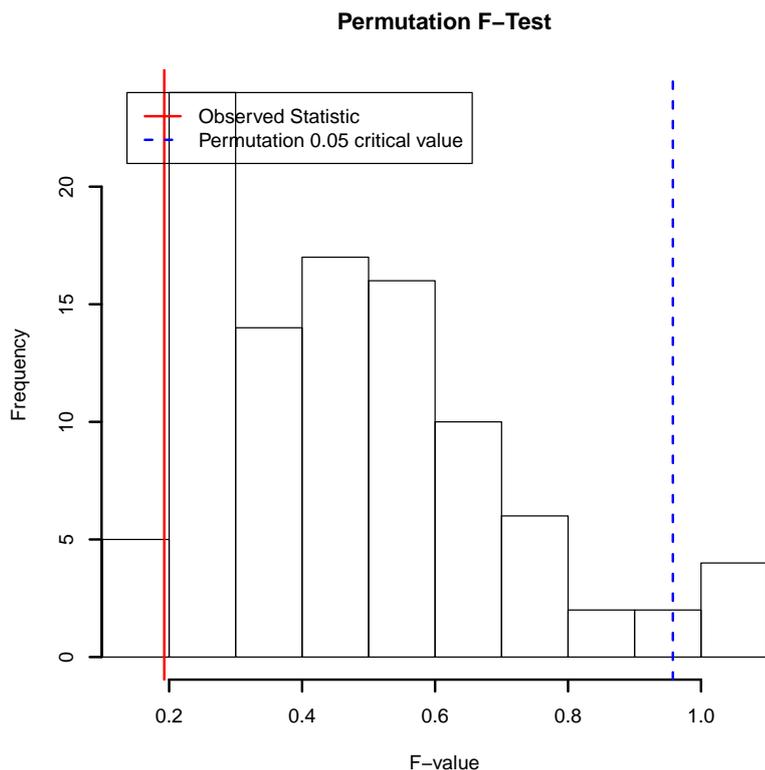
Per arrivare a scegliere un modello rappresentativo dei dati analizzati si sono sperimentate una serie di soluzioni, e grazie al test di permutazione esposto nella sezione precedente si è riusciti ad indirizzare l'analisi verso un modello che potesse ben sintetizzare i dati.

Il primo modello applicato si sviluppa considerando una finestra temporale di 15 ore sintetizzata come spiegato nel paragrafo 3.1. Le variabili considerate (differenze prime e differenze di voltaggio), sono state combinate semplicemente applicando la somma delle stesse in un semplice modello della forma

$$\hat{y}_i = \alpha_0 + \sum_{j=1}^q \beta_{1,j}[d_3(t_j) + d_4(t_j)] + \sum_{j=1}^q \beta_{2,j}[s_3(t_j) + s_4(t_j)] + \epsilon_i.$$

Dove  $d_3$  è la variabile che rappresenta la sintesi tramite differenze prime relativa al segnale  $F3$  e  $d_4$  quella relativa a  $F4$ . Mentre  $s_3$  ed  $s_4$  rappresentano

le differenze di voltaggio. Il risultato del test di permutazione ottenuto è visibile in figura 3.5, come si può vedere il test propende significativamente per non rifiutare  $H_0$  e quindi l'efficacia di tale modello non è avvalorata dal risultato del test.

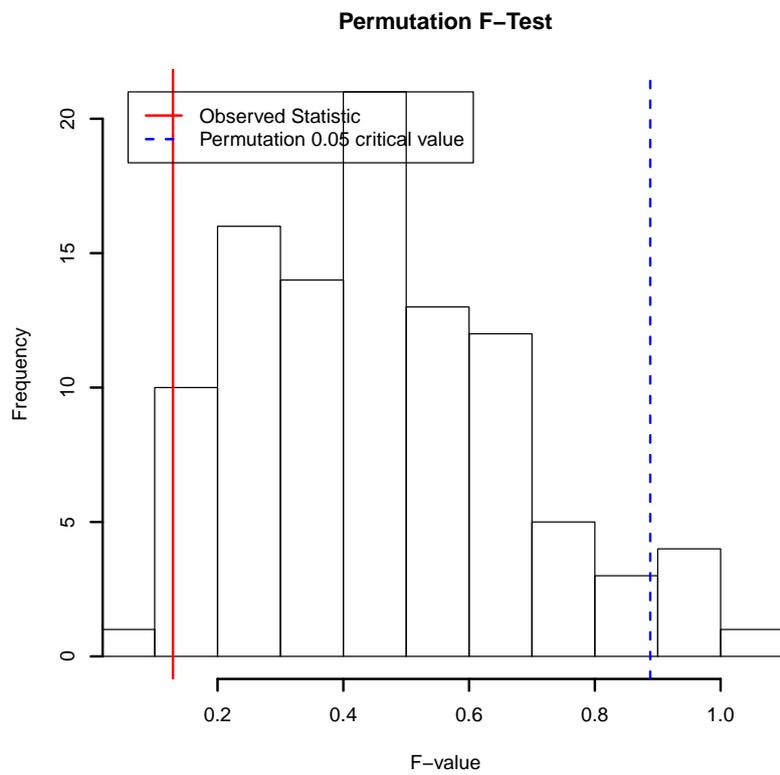


**Fig. 3.5:** Test di permutazione per il modello additivo.

Come passo successivo si è provato un modello che includesse una trasformazione dei dati relativi alle differenze di voltaggio, nello specifico si è scelto di calcolare la radice quadrata della somma di tali variabili, arrivando così ad un modello di questo tipo

$$\hat{y}_i = \alpha_0 + \sum_{j=1}^q \beta_{1,j} [d_3(t_j) + d_4(t_j)] + \sum_{j=1}^q \beta_{2,j} \sqrt{s_3(t_j) + s_4(t_j)} + \epsilon_i.$$

Anche in questo caso il test non rifiutava l'ipotesi nulla (vedi figura 3.6)

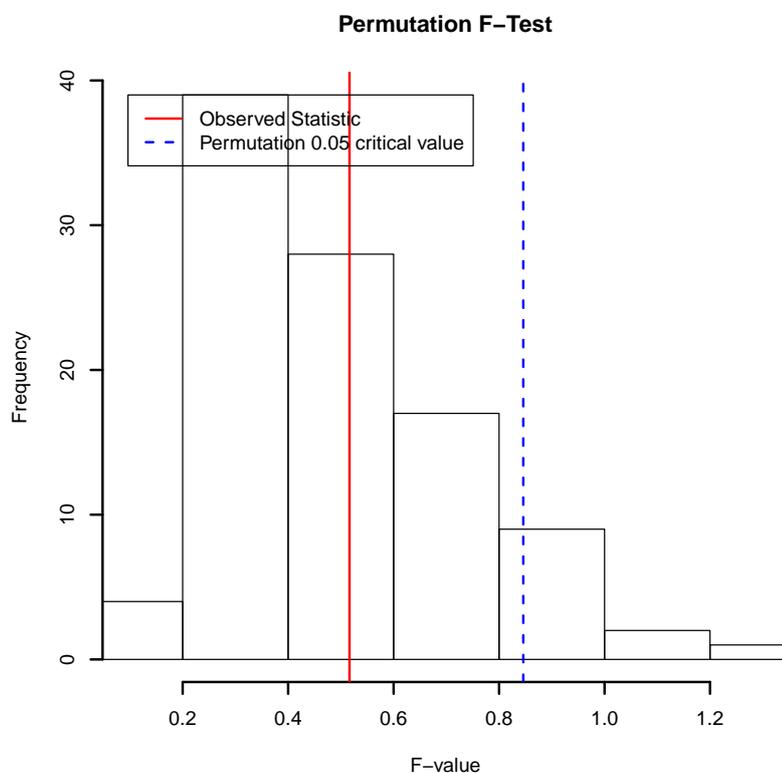


**Fig. 3.6:** Test di permutazione per il modello con radice quadrata delle differenze di voltaggio.

Di fronte all'assenza di risultati significativi con l'utilizzo di modelli additivi si è deciso di applicare un modello che considerasse la sottrazione delle differenze di voltaggio tra i due tracciati F3 ed F4, arrivando a questo modello

$$\hat{y}_i = \alpha_0 + \sum_{j=1}^q \beta_{1,j} [d_3(t_j) + d_4(t_j)] + \sum_{j=1}^q \beta_{2,j} [s_3(t_j) - s_4(t_j)] + \epsilon_i.$$

Il quale presentava risultati un po' più convincenti nell'ambito della significatività come dimostrato dal test in figura 3.7.



**Fig. 3.7:** Test di permutazione per il modello additivo con sottrazione delle differenze di voltaggio.

Poiché tale modello portava il test a non rifiutare  $H_0$  però si avvicinava sensibilmente alla zona borderline del test, si è pensato di applicare la sottrazione anche alle differenze prime. Questa scelta nasce anche dall'idea che

probabilmente se il lobo sinistro ed il lobo destro avessero manifestato lo stesso pattern, questo sarebbe stato in un certo senso annullato dalla differenza delle due curve, sia per le differenze prime sia per le differenze di voltaggio. In tal modo solamente le parti della curva effettivamente differenti tra le due zone craniche avrebbero avuto un coefficiente associato.

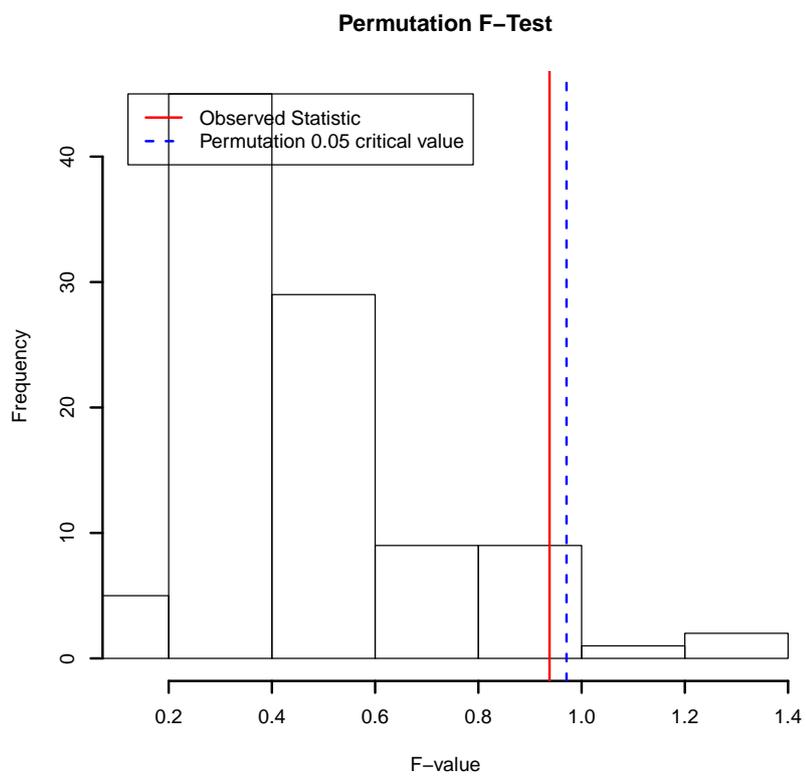
Il modello pensato sarà quindi

$$\sum_{j=1}^q \beta_{1,j}[d_3(t_j) - d_4(t_j)] + \sum_{j=1}^q \beta_{2,j}[s_3(t_j) - s_4(t_j)] + \epsilon_i.$$

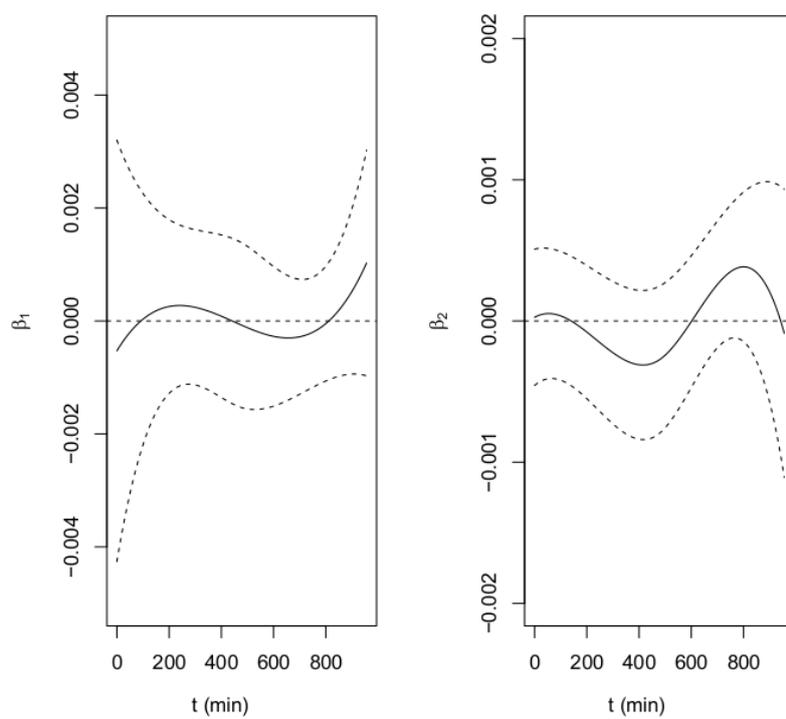
Effettuando il test di permutazione, quanto osservato non sembra ancora avere la significatività statistica sufficiente per dire che il modello sia effettivamente adattabile ai dati in questione: vedi figura 3.8. Tuttavia il risultato del test si trova in una zona borderline.

Di fronte a tale risultato si è deciso quindi di rappresentare i coefficienti del modello per vedere quale fosse la loro influenza nel tempo. Com'è possibile osservare, i coefficienti delle due curve si comportano in modo differente. Per quanto riguarda le differenze prime, la curva che rappresenta il coefficiente ha un andamento di fondo crescente che aumenta con la distanza dal momento della dimissione. Si può osservare che la curva inizia con un valore negativo per poi diventare positivo. Osservando invece la curva relativa alla differenza di voltaggio non è presente un andamento di fondo ben definito, la variabilità della curva è notevole e si denota una forte influenza negativa a circa metà periodo di coma. Gli intervalli di confidenza si presentano più ampi agli estremi mentre si restringono nella parte centrale dell'intervallo. Questo aspetto è più accentuato per il coefficiente  $\beta_1$  si veda figura 3.9.

Calcolando il test  $R^2$  per verificare l'adattabilità del modello il valore ottenuto è 0,222, la variabilità spiegata dal modello è decisamente ridotta.



**Fig. 3.8:** Test di permutazione su una finestra temporale di 15 ore.



**Fig. 3.9:** Coefficienti del modello con finestra temporale 15 ore (900 minuti considerando il primo secondo di ciascun minuto). Sull'asse delle ascisse gli istanti temporali sono 'invertiti', il tempo va dalla dimissione fino alle 15 ore precedenti.

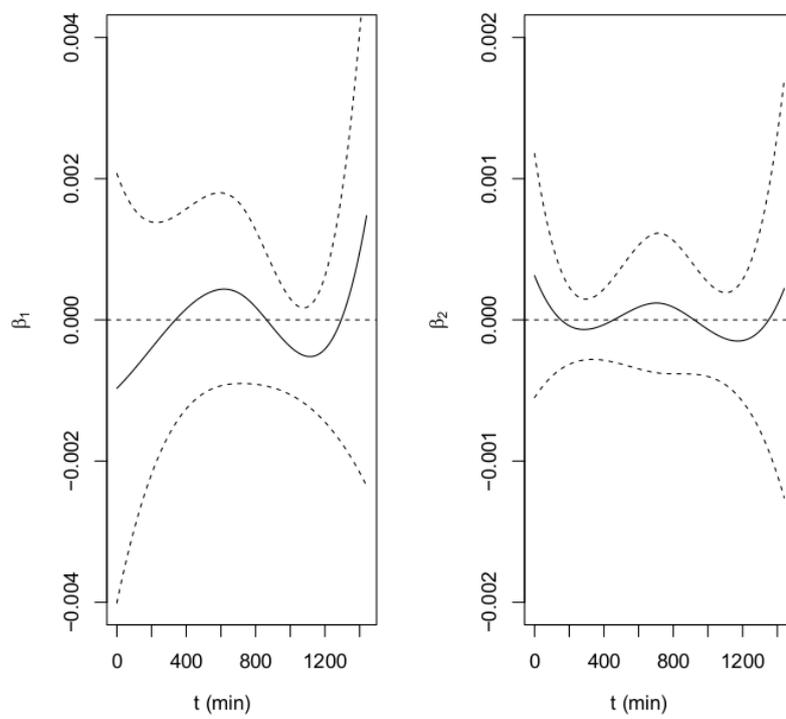
Visti i risultati ottenuti con il modello che includeva la sottrazione per entrambe le variabili si è provato ad applicare lo stesso su finestre temporali più ampie.

### 3.2.5 Finestra temporale di 24 ore

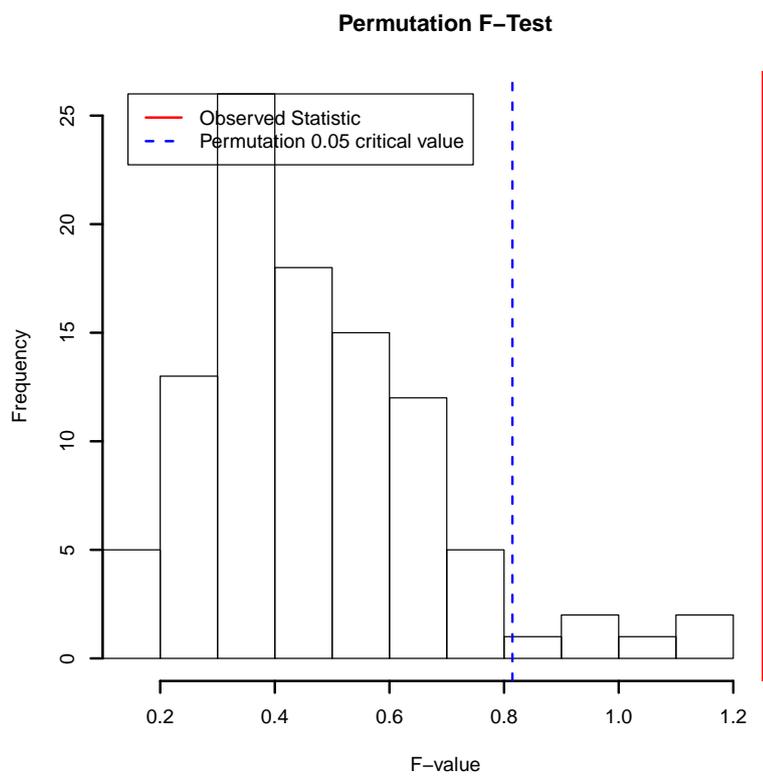
Sempre mantenendo lo stesso modello ma andando ad indagare una finestra temporale più ampia l'andamento di entrambi i coefficienti cambia decisamente, soprattutto per quanto riguarda il coefficiente relativo alla differenza di voltaggio. Come è possibile vedere dal grafico di destra nella figura 3.10 il coefficiente presenta valori positivi sia all'inizio sia al termine dello stato comatoso, mentre nella parte interna varia passando da valori negativi a valori positivi.

Per quanto riguarda invece il coefficiente relativo alle differenze prime, l'andamento crescente di fondo osservato nella finestra temporale di quindici ore, si manifesta anche in questo caso marcando di più il duplice cambiamento di pendenza presente all'interno della finestra temporale. Gli intervalli di confidenza calcolati si presentano molto più variabili rispetto a quelli visti per il modello precedente, tuttavia in questo modello la significatività dei valori ottenuti è comprovata da quanto ottenuto con il test di permutazione (vedi figura 3.11).

Il valore del test per la bontà di adattamento  $R^2$  ottenuto con questo modello è pari a 0,408, ciò implica un buon miglioramento rispetto alla finestra temporale di quindici ore, anche se rimane abbastanza distante da risultati ottimi.



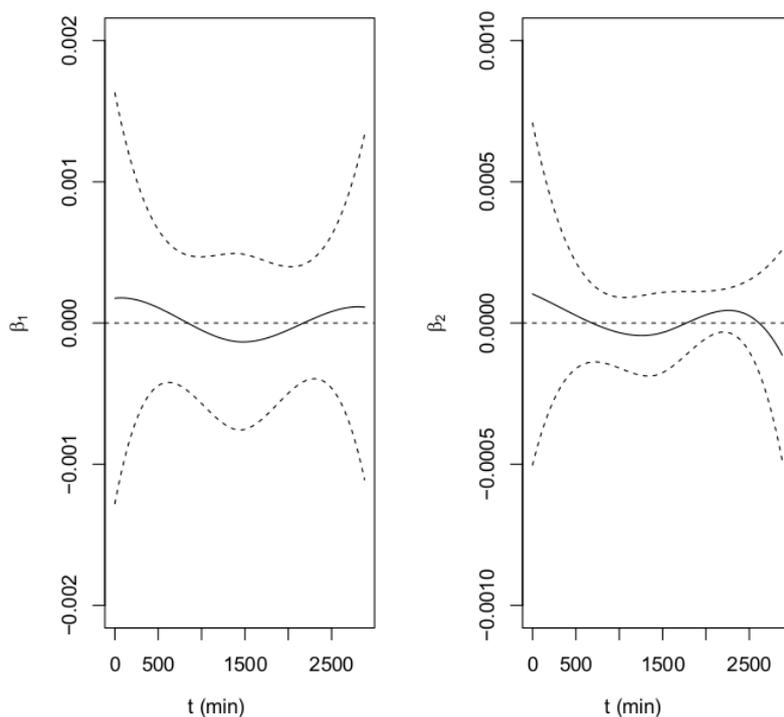
**Fig. 3.10:** Coefficienti del modello con finestra temporale 24 ore (1.440 minuti).



**Fig. 3.11:** Test di permutazione su una finestra temporale di 24 ore.

### 3.2.6 Finestra temporale di 48 ore

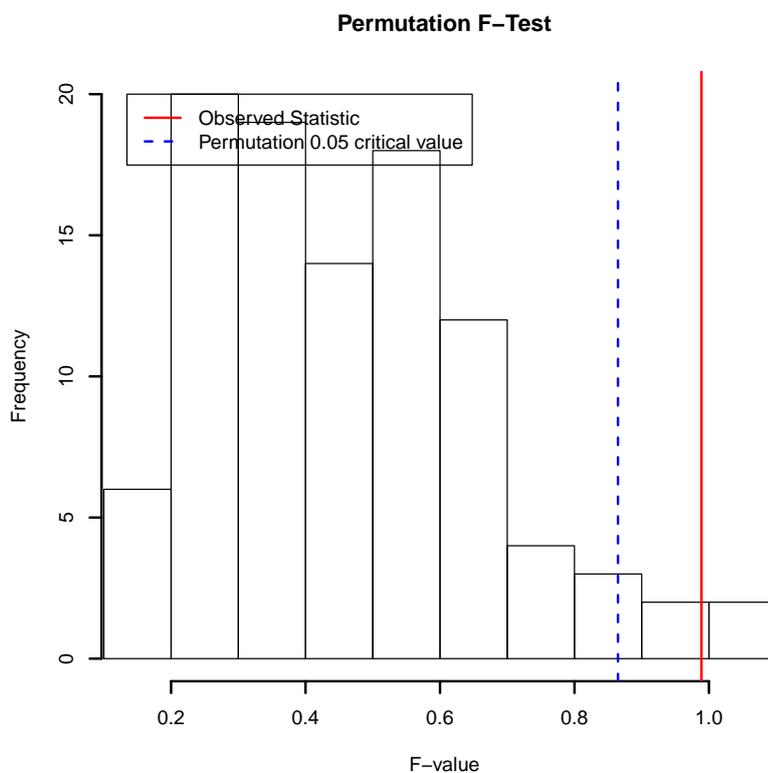
I coefficienti relativi alla finestra temporale di 48 (figura 3.12) ore delineano una situazione totalmente differente dalla precedente finestra temporale di lunghezza dimezzata.



**Fig. 3.12:** Coefficienti del modello con finestra temporale 48 ore (2.880 minuti).

Il coefficiente relativo alle differenze prime presenta un andamento quasi costante, parte da un valore positivo ad inizio ricovero per poi tornarci al termine, ed il cambio di pendenza presente a metà ricovero è poco accentuato. L'intervallo di confidenza di questa curva è molto ampio e variabile, motivo per cui è possibile immaginare la poca attendibilità della stessa. Invece per quanto riguarda il coefficiente relativo alla differenza di voltaggio l'intervallo è

abbastanza stretto attorno alla curva e la stessa presenta un lieve andamento decrescente di fondo. Anche in questo caso il test di permutazione avvalorava quanto osservato (vedi figura 3.13), inoltre gli intervalli di confidenza tendono a restringersi un po' nella parte centrale pur mantenendo un'ampiezza elevata agli estremi dell'intervallo.



**Fig. 3.13:** Test di permutazione su una finestra temporale di 48 ore.

Tuttavia calcolando il test  $R^2$  per questo modello si ottiene un valore pari a 0,115 decisamente un peggioramento nella quantità di variabilità spiegata dal modello, questo probabilmente spiega anche perché le curve relative ai coefficienti presentino un andamento così differente rispetto ai modelli precedenti.

Osservando quanto ottenuto con i diversi modelli studiati, è possibile supporre un impiego delle variabili scelte per un'analisi preliminare dello sviluppo di uno stato comatoso.

Preliminare poiché la misura dell'influenza di tali variabili risulta ancora difficile ed i test  $R^2$  effettuati dimostrano come la quantità di variabilità spiegata dal modello che include le due sole regioni frontali, non sia elevata.



# Conclusioni

Il compito di questo studio era quello di indagare la possibilità di sfruttare i tracciati EEG per ottenere un'indicazione sulla capacità di recupero di un paziente nei tre mesi successivi al risveglio dal coma. Tale possibilità era già stata dimostrata in un lavoro precedente considerando tre significativi istanti temporali durante il periodo di coma, la sfida dello studio era quella di poter considerare i tracciati nel continuo per una più accurata modellazione. Tale sfida è stata vinta in parte. Com'è stato visibile fin dall'inizio il problema si presentava particolarmente complesso sia dal punto di vista computazionale sia dal punto di vista della scarsa quantità di pazienti disponibili, tuttavia dopo un buon lavoro di pulizia ed interpretazione dei dati si è riusciti a compiere un'analisi funzionale che è possibile identificare come un primo passo verso uno studio più approfondito di questa tipologia di dati.

I risultati ottenuti, pur non presentando una facile interpretazione ed un'indicazione pratica, forniscono una prova concreta sulla possibilità di identificare un'influenza delle strutture dei diversi tracciati elettroencefalografici a fini previsivi ed analitici. Questi risultati quindi aprono le porte a possibili nuovi scenari di analisi, ad esempio:

- analisi più complete che includano anche altre zone cerebrali e non solo la regione frontale;

- definizione di altre sintesi dei dati diverse dalle differenze prime e differenze di voltaggio;
- utilizzo di metodi di calcolo in parallelo per il miglioramento dei processi di analisi;
- utilizzo dei Potenziali evocati<sup>6</sup> in concomitanza ai tracciati EEG.

Va precisato che le possibili analisi future saranno subordinate ai limiti computazionali imposti dagli strumenti utilizzati per le stesse. Senza dubbio le crescenti capacità di calcolo degli strumenti oggi utilizzati non possono far altro che allargare questi limiti, tuttavia un'ottimizzazione delle funzioni disponibili nella libreria `fda` potrebbe di certo portare dei benefici in questi termini. Inoltre non va dimenticato che per l'analisi svolta son state operate delle precise scelte riassuntive sul dataset per poter operare con gli strumenti a disposizione, non è da escludere perciò la possibilità di utilizzare differenti metodi per estrarre dai dati l'informazione necessaria senza perdere precisione nell'analisi. Alcune alternative potrebbero essere l'utilizzo della classica analisi spettrale dei segnali o di definizioni differenti della finestra temporale considerata.

---

<sup>6</sup>Esami che studiano la risposta del sistema nervoso centrale a stimoli esterni.

# Appendice A

## Codice Java

```
1
2 import java.io.*;
3 import java.util.*;
4 public class StringiFile
5 {
6     List<Double> f3Scarto = null;
7     List<Double> f4Scarto = null;
8     List<Integer> f3Min = null;
9     List<Integer> f4Min = null;
10    List<Integer> f3Magg = null;
11    List<Integer> f4Magg = null;
12    List<Integer> f3Diff1 = null;
13    List<Integer> f4Diff1 = null;
14    BufferedReader stdin;
15    String progressivo = null;
16
17    public StringiFile()
18    {
19        stdin = new BufferedReader(new InputStreamReader(System.in))
                ;
```

```

20     f3Scarto = new LinkedList<Double>();
21     f4Scarto = new LinkedList<Double>();
22     f3Diff1 = new LinkedList<Integer>();
23     f4Diff1 = new LinkedList<Integer>();
24     f3Min = new LinkedList<Integer>();
25     f4Min = new LinkedList<Integer>();
26     f3Magg = new LinkedList<Integer>();
27     f4Magg = new LinkedList<Integer>();
28 }
29
30 public static void main(String [] argv)
31 {
32     (new StringiFile()).exec(argv);
33 }
34
35 public void exec(String [] argv)
36 {
37     while(true)
38     {
39         String cmd = null;
40         System.out.print("Analizza/Esci_?_>_");
41         try
42         {
43             cmd = stdin.readLine();
44         }
45         catch(IOException ioe)
46         {
47             ioe.printStackTrace();
48             break;
49         }
50         if(cmd == null || cmd.equals("Esci"))
51         {
52             System.out.println("Ciao");

```

```

53         break;
54     }
55     else if(cmd.equals("Analizza"))
56     {
57         leggi();
58     }
59     else
60     {
61         System.out.println("Comando_<" + cmd + ">_non_
                ricnonsciuto");
62         continue;
63     }
64 }
65 }
66
67 public void leggi()
68 {
69     BufferedReader br = null;
70     String [] separazione = new String [3];
71     double [] analizzaF3 = new double [128];
72     double [] analizzaF4 = new double [128];
73     try
74     {
75         System.out.print("Percorso_>");
76         String nomeFile = stdin.readLine()+"/EEG.asc";
77         System.out.print("Numero_del_paziente_>");
78         progressivo = stdin.readLine();
79         br = new BufferedReader(new FileReader(new File(nomeFile))
                );
80         String val= null;
81         int index = 0;
82         int ciclo = 0;
83 // salto le prime 11 righe contenenti commenti

```

```

84     for (int i = 1; i <= 11; i++)
85     {
86         br.readLine();
87     }
88     while((val = br.readLine()) != null)
89     {
90         separazione = val.split("\t");
91         separazione[0] = separazione[0].replace(",",".");
92         separazione[1] = separazione[1].replace(",",".");
93         analizzaF3[index] = Double.parseDouble(separazione[0]);
94         analizzaF4[index] = Double.parseDouble(separazione[1]);
95         if(index == 127)
96         {
97             f3Min.add(diffPrimeMin(analizzaF3));
98             f4Min.add(diffPrimeMin(analizzaF4));
99             f3Magg.add(diffPrimeMagg(analizzaF3));
100            f4Magg.add(diffPrimeMagg(analizzaF4));
101            f3Diff1.add(diffPrime(analizzaF3));
102            f4Diff1.add(diffPrime(analizzaF4));
103            f3Scarto.add(scarto(analizzaF3));
104            f4Scarto.add(scarto(analizzaF4));
105            ciclo++;
106            index = 0;
107        }
108        else
109        {
110            index++;
111        }
112    }
113    br.close();
114    scriviFileRapido();
115 }
116 catch( IOException ioe1)

```

```

117     {
118         ioe1.printStackTrace();
119     }
120 }
121
122 public void scriviFileRapido()
123 {
124     String s1 = f3Scarto.toString();
125     String s2 = f3Diff1.toString();
126     String s3 = f3Min.toString();
127     String s4 = f3Magg.toString();
128     String s5 = f4Scarto.toString();
129     String s6 = f4Diff1.toString();
130     String s7 = f4Min.toString();
131     String s8 = f4Magg.toString();
132
133     s1 = s1.replace("[", "");
134     s1 = s1.replace(",_", "\n");
135     s1 = s1.replace("]", "");
136
137     s2 = s2.replace("[", "");
138     s2 = s2.replace(",_", "\n");
139     s2 = s2.replace("]", "");
140
141     s3 = s3.replace("[", "");
142     s3 = s3.replace(",_", "\n");
143     s3 = s3.replace("]", "");
144
145     s4 = s4.replace("[", "");
146     s4 = s4.replace(",_", "\n");
147     s4 = s4.replace("]", "");
148
149     s5 = s5.replace("[", "");

```

```

150     s5 = s5.replace(",_", "\n");
151     s5 = s5.replace("]", "");
152
153     s6 = s6.replace("[", "");
154     s6 = s6.replace(",_", "\n");
155     s6 = s6.replace("]", "");
156
157     s7 = s7.replace("[", "");
158     s7 = s7.replace(",_", "\n");
159     s7 = s7.replace("]", "");
160
161     s8 = s8.replace("[", "");
162     s8 = s8.replace(",_", "\n");
163     s8 = s8.replace("]", "");
164
165     String file1="/media/MioHD/Tesi/DatiRiassunti/F3/f3Scarto_"+
        progressivo + ".txt";
166     String file2="/media/MioHD/Tesi/DatiRiassunti/F3/f3Diff_"+
        progressivo + ".txt";
167     String file3="/media/MioHD/Tesi/DatiRiassunti/F3/f3Min_"+
        progressivo + ".txt";
168     String file4="/media/MioHD/Tesi/DatiRiassunti/F3/f3Magg_"+
        progressivo + ".txt";
169     String file5="/media/MioHD/Tesi/DatiRiassunti/F4/f4Scarto_"+
        progressivo + ".txt";
170     String file6="/media/MioHD/Tesi/DatiRiassunti/F4/f4Diff_"+
        progressivo + ".txt";
171     String file7="/media/MioHD/Tesi/DatiRiassunti/F4/f4Min_"+
        progressivo + ".txt";
172     String file8="/media/MioHD/Tesi/DatiRiassunti/F4/f4Magg_"+
        progressivo + ".txt";
173
174     OutputStreamWriter or1 = null;

```

```

175     OutputStreamWriter or2 = null;
176     OutputStreamWriter or3 = null;
177     OutputStreamWriter or4 = null;
178     OutputStreamWriter or5 = null;
179     OutputStreamWriter or6 = null;
180     OutputStreamWriter or7 = null;
181     OutputStreamWriter or8 = null;
182
183     try
184     {
185         File f1 = new File(file1);
186         if(f1.exists())
187         {
188             or1 = new FileWriter(f1);
189             or1.write(s1);
190             or1.flush();
191             or1.close();
192         }
193         else
194         {
195             or1 = new FileWriter(new File(file1));
196             or1.write(s1);
197             or1.flush();
198             or1.close();
199         }
200         File f2 = new File(file2);
201         if(f2.exists())
202         {
203             or2 = new FileWriter(f2);
204             or2.write(s2);
205             or2.flush();
206             or2.close();
207         }

```

```
208     else
209     {
210         or2 = new FileWriter(new File(file2));
211         or2.write(s2);
212         or2.flush();
213         or2.close();
214     }
215     File f3 = new File(file3);
216     if(f3.exists())
217     {
218         or3 = new FileWriter(f3);
219         or3.write(s3);
220         or3.flush();
221         or3.close();
222     }
223     else
224     {
225         or3 = new FileWriter(new File(file3));
226         or3.write(s3);
227         or3.flush();
228         or3.close();
229     }
230     File f4 = new File(file4);
231     if(f4.exists())
232     {
233         or4 = new FileWriter(f4);
234         or4.write(s4);
235         or4.flush();
236         or4.close();
237     }
238     else
239     {
240         or4 = new FileWriter(new File(file4));
```

```
241         or4.write(s4);
242         or4.flush();
243         or4.close();
244     }
245     File f5 = new File(file5);
246     if(f5.exists())
247     {
248         or5 = new FileWriter(f5);
249         or5.write(s5);
250         or5.flush();
251         or5.close();
252     }
253     else
254     {
255         or5 = new FileWriter(new File(file5));
256         or5.write(s5);
257         or5.flush();
258         or5.close();
259     }
260     File f6 = new File(file6);
261     if(f6.exists())
262     {
263         or6 = new FileWriter(f6);
264         or6.write(s6);
265         or6.flush();
266         or6.close();
267     }
268     else
269     {
270         or6 = new FileWriter(new File(file6));
271         or6.write(s6);
272         or6.flush();
273         or6.close();
```

```

274     }
275     File f7 = new File(file7);
276     if(f7.exists())
277     {
278         or7 = new FileWriter(f7);
279         or7.write(s7);
280         or7.flush();
281         or7.close();
282     }
283     else
284     {
285         or7 = new FileWriter(new File(file7));
286         or7.write(s7);
287         or7.flush();
288                 or7.close();
289     }
290     File f8 = new File(file8);
291     if(f8.exists())
292     {
293         or8 = new FileWriter(f8);
294         or8.write(s8);
295         or8.flush();
296         or8.close();
297     }
298     else
299     {
300         or8 = new FileWriter(new File(file8));
301         or8.write(s8);
302         or8.flush();
303         or8.close();
304     }
305     }
306     catch (IOException ioe2)

```

```

307     {
308         ioe2.printStackTrace();
309     }
310 }
311
312 // Calcolo le differenze prime finite all'indietro
313
314 public int diffPrime(double[] vettore)
315 {
316     double x1,x2,diff;
317     boolean crescente = true;
318     int contatore = 0;
319     for (int i = 1 ; i <= (vettore.length)-1; i++)
320     {
321         x2 = vettore[i];
322         x1 = vettore[i-1];
323         diff = (x2-x1);
324         if (crescente & (diff<0))
325         {
326             contatore++;
327             crescente = false;
328         }
329         if (!crescente & (diff>0))
330         {
331             contatore++;
332             crescente = true;
333         }
334     }
335     return contatore;
336 }
337
338 // Calcolo il numero di differenze prime finite all'indietro
    minore di 0

```

```

339
340 public int diffPrimeMin(double[] vettore)
341 {
342     double x1,x2;
343     int conta = 0;
344     for (int i = 1 ; i <= (vettore.length)-1; i++)
345     {
346         x2 = vettore[i];
347         x1 = vettore[i-1];
348         if((x2-x1)<0)
349         {
350             conta++;
351         }
352     }
353     return conta;
354 }
355
356 // Calcolo il numero di differenze prime finite all'indietro
      maggiore di 0
357
358 public int diffPrimeMagg(double[] vettore)
359 {
360     double x1,x2;
361     int conta = 0;
362     for (int i = 1 ; i <= (vettore.length)-1; i++)
363     {
364         x2 = vettore[i];
365         x1 = vettore[i-1];
366         if((x2-x1)>0)
367         {
368             conta++;
369         }
370     }

```

```
371     return conta;  
372 }  
373  
374 // Calcolo la differenza di potenziale elettrico tra il massimo  
    voltaggio ed il minimo  
375  
376 public double scarto(double[] vett)  
377 {  
378     Arrays.sort(vett);  
379     return vett[127] - vett[0];  
380 }  
381 }
```



# Appendice B

## Codice R

```
1
2 #####
3 # Script 1: Sintesi del dataset #
4 #####
5
6 # Carico i dati smooth
7 datif3Diff<-read.csv(file.choose(),header=TRUE,sep=" ")
8 datif4Diff<-read.csv(file.choose(),header=TRUE,sep=" ")
9 datif3Scarto<-read.csv(file.choose(),header=TRUE,sep=" ")
10 datif4Scarto<-read.csv(file.choose(),header=TRUE,sep=" ")
11
12 # Primi 5 secondi di ogni minuto
13 multipli<-seq(0,4)
14 indici<-seq(from = 1,to = nrow(datif4Scarto),by = 60)
15
16 f3DiffMin<-NULL
17 for(i in 1:length(indici)){
18   f3DiffMin<-rbind(f3DiffMin,datif3Diff[indici[i]+multipli,])
19 }
20 dim(f3DiffMin)
```

```

21 write.table(f3DiffMin , file="f3RiassPrimi5SecMinDiff.csv" ,row.
      names=FALSE)
22
23 f4DiffMin<-NULL
24 for(i in 1:length(indici)){
25   f4DiffMin<-rbind(f4DiffMin , datif4Diff[indici[i]+multipli ,])
26 }
27 dim(f4DiffMin)
28 write.table(f4DiffMin , file="f4RiassPrimi5SecMinDiff.csv" ,row.
      names=FALSE)
29
30 f3ScartoMin<-NULL
31 for(i in 1:length(indici)){
32   f3ScartoMin<-rbind(f3ScartoMin , datif3Scarto[indici[i]+multipli
      ,])
33 }
34 dim(f3ScartoMin)
35 write.table(f3ScartoMin , file="f3RiassPrimi5SecMinScarto.csv" ,row
      .names=FALSE)
36
37 f4ScartoMin<-NULL
38 for(i in 1:length(indici)){
39   f4ScartoMin<-rbind(f4ScartoMin , datif4Scarto[indici[i]+multipli
      ,])
40 }
41 dim(f4ScartoMin)
42 write.table(f4ScartoMin , file="f4RiassPrimi5SecMinScarto.csv" ,row
      .names=FALSE)
43
44 #####
45 # Script 2: Realizzazione modello e test di permutazione #
46 #####
47

```

```

48 datif3Diff<-read.csv( './DatiSmooth96/f3RiassPrim5SecMinDiff.csv '
    ,header=TRUE, sep=" ")
49 datif4Diff<-read.csv( './DatiSmooth96/f4RiassPrimi5SecMinDiff.csv
    ',header=TRUE, sep=" ")
50 datif3Scarto<-read.csv( './DatiSmooth96/f3RiassPrimi5SecMinScarto
    .csv ',header=TRUE, sep=" ")
51 datif4Scarto<-read.csv( './DatiSmooth96/f4RiassPrimi5SecMinScarto
    .csv ',header=TRUE, sep=" ")
52
53 dim(datif3Diff)
54 dim(datif3Scarto)
55
56 # Creo dati funzionali
57 library(fda)
58 secbasis<-create.bspline.basis(c(0,nrow(datif3Diff)),nbasis=nrow
    (datif3Diff))
59 f3Diff<-fd(as.matrix(datif3Diff),secbasis)
60 f4Diff<-fd(as.matrix(datif4Diff),secbasis)
61 secbasisScarto<-create.bspline.basis(c(0,nrow(datif4Scarto)),
    nbasis=nrow(datif4Scarto))
62 f3Scarto<-fd(as.matrix(datif3Scarto),secbasisScarto)
63 f4Scarto<-fd(as.matrix(datif4Scarto),secbasisScarto)
64
65 # Creo il vettore risposta
66 datiRisp<-read.csv( './tabella.txt ',sep=";",header=T)
67 res<-as.numeric(datiRisp$GOS)
68 res[which(res==2)]<-1
69 res[which(res==5)]<-4
70
71 comalist<-vector("list",3)
72 comalist[[1]]<-rep(1,length(res))
73 comalist[[2]]<-(f3Diff-f4Diff)
74 comalist[[3]]<-(f3Scarto-f4Scarto)

```

```

75
76 # Creo una base costante per l'intercetta e delle spline con
      cinque nodi per il coefficiente
77 constBasis<-create.constant.basis(c(0,nrow(datif3Diff)))
78 splinebasis1<-create.bspline.basis(c(0,nrow(datif3Diff)),5)
79 splinebasis2<-create.bspline.basis(c(0,nrow(datif3Scarto)),5)
80
81 betalis<-vector("list",3)
82 betalis[[1]]<-constBasis
83 betalis[[2]]<-splinebasis1
84 betalis[[3]]<-splinebasis2
85
86 # test di Permutazione riadattato
87 Fstat.fd<-function (y, yhat, argvals = NULL)
88 {
89     if (is.numeric(yhat)) {
90         yhat = as.vector(yhat)
91     }
92     if ((is.vector(y) & !is.vector(yhat)) | (is.fd(y) & !is.fd(
          yhat))) {
93         stop("y_and_yhat_must_both_be_either_scalars_or_
          functional_data_objects.")
94     }
95     if (is.fd(y)) {
96         rangeobs = y$basis$range
97         rangehat = yhat$basis$range
98         if (!prod(rangeobs == rangehat)) {
99             stop("y_and_yhat_do_not_have_the_same_range")
100        }
101        if (is.null(argvals)) {
102            argvals = seq(rangeobs[1], rangeobs[2], length.out =
              101)
103        }

```

```

104     yvec = eval.fd(argvals , y)
105     yhatvec = eval.fd(argvals , yhat)
106     F=as.numeric(wilcox.test(yhatvec ,yvec ,paired=T)$
        statistic)
107   }
108   else {
109     yvec = y
110     yhatvec = yhat
111     F=as.numeric(wilcox.test(yhatvec ,yvecc ,paired=T)$
        statistic)
112   }
113   return(list(F = F, argvals = argvals))
114 }
115
116 F.res<-Fperm.fd(res ,comalist ,betalis ,nperm=100)
117 save(file='./Test.rda',F.res)
118
119 #####
120 # Script 3: Analisi modello #
121 #####
122
123 load("./Test.rda")
124 m1<- F.res$fRegressList
125 library(fda)
126
127 # Creo il vettore risposta
128 datiRisp<-read.csv(file.choose(),sep=";",header=T)
129 res<-as.numeric(datiRisp$GOS)
130 res[which(res==2)]<-1
131 res[which(res==5)]<-4
132
133 # R^2
134 ess<-sum((res.hat-mean(res))^2)

```

```

135 tss<-sum((res-mean(res))^2)
136 r2<-ess/tss
137 r2
138
139 # INTERVALLI DI CONFIDENZA Per i BETA
140 resid<-res-res.hat
141 sigmaE<-sum(resid^2)/(34-m1$df)
142 sigmaE<-sigmaE*diag(rep(1,34))
143 idetit<-diag(rep(1,34))
144 stderrList<-fRegress.stderr(m1,idetit,sigmaE)
145
146 par(mfrow=c(1,2))
147 options(scipen=10)
148 betaestList<-m1$betaestlist
149 comabetafd<-betaestList[[2]]$fd
150 plot(comabetafd,xlab="t",ylab=expression(beta[1]),ylim=c
      (-0.004,0.004))
151
152 betaestList<-m1$betaestlist
153 comabetafd<-betaestList[[2]]$fd
154 betastderrList<-stderrList$betastderrlist
155 betastderrfd<-betastderrList[[2]]
156 lines(comabetafd+2*betastderrfd,lty=2,lwd=1)
157 lines(comabetafd-2*betastderrfd,lty=2,lwd=1)
158
159 betaestList<-m1$betaestlist
160 comabetafd<-betaestList[[3]]$fd
161 plot(comabetafd,xlab="t",ylab=expression(beta[2]),ylim=c
      (-0.002,0.002))
162
163 betaestList<-m1$betaestlist
164 comabetafd<-betaestList[[3]]$fd
165 betastderrList<-stderrList$betastderrlist

```

```
166 betastderrfd<-betastderrList [[3]]
167 lines (comabetafd+2*betastderrfd ,lty=2,lwd=1)
168 lines (comabetafd-2*betastderrfd ,lty=2,lwd=1)
```



# Bibliografia

Anastasi G. (2007). *Trattato di anatomia umana*. Ermes.

Antonini (2012). *Analisi statistica dei segnali elettrici cerebrali in pazienti in stato vegetativo e di minima coscienza*. Tesi di Laurea, Politecnico di Milano Facoltà di Ingegneria dei Sistemi.

Azzalini A., Scarpa B. (2009). *Analisi dei dati e data mining*. Springer.

Bauer G., Gerstenbrand F., Rimpl E. (1979). Varieties of the Locked-in syndrome. *Journal of Neurology*, **221**(2), 77–91.

Boor C. D. (2001). *A practical guide to Splines*. Springer.

Bosco E., Marton E., Feletti A., Scarpa B., Longatti P. (2011). Dynamic monitors of brain function: a new target in neurointensive care unit. *Critical Care*, **23**(3), 208–220.

Cardazzi (2012). *Analisi statistica multivariata e funzionale dei profili temporali di connettività di segnali EEG intracranici nelle epilessie focali*. Tesi di Laurea, Politecnico di Milano Facoltà di Ingegneria dei Sistemi.

Curry H., Schoenberg I. (1947). On Spline distributions and their limits: The polya distribution functions. *Bull. Amer. Math. Soc.*

- Husain A. M. (2006). Electroencephalographic assesment of coma. *Journal of Clinical Neurophysiology*, **23**(3), 208–220.
- Jennet B., Bond M. (1975). Assesment of outcome after severe brain damage. *Lancet*, **1**(1), 480–484.
- Nuwer M. R. (1998). Fundamentals of evoked potentials and common clinical applications today. *Electroencephalography and clinical Neurophysiology*, **106**(2), 142–148.
- Ramsay J., Silverman B. (1997). *Functional Data Analysis*. Springer.
- Ramsay J., Silverman B. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer.
- Ramsay J., Hooker G., Graves S. (2009). *Functional Data Analysis with R and MATLAB*. Springer.
- Synek V. M. (1988). Prognostically important EEG coma patterns in diffuse anoxic and traumatic encephalopathies in adults. *Journal of Clinical Neurophysiology*, **21**(5), 307–318.
- Teasdale G., Jennet B. (1974). Assesment of coma and impaired consciousness. a practical scale. *Lancet*, **13**(2), 4–81.