

PLSgui Tutorial

PLSgui Tutorial di Giorgio Arcara
è distribuito con Licenza [Creative Commons Attribuzione - Non commerciale 4.0 Internazionale](https://creativecommons.org/licenses/by-nc/4.0/).

Come utilizzare questo tutorial	2
SEZIONE 1	2
Introduzione teorica	2
Installare il software per la PLS	3
Preparare i files per utilizzare la PLS	3
Utilizzare il programma PLSgui	4
I tre passi dell'analisi tramite PLS	6
PASSO 1 - Session Profile for ERP data	7
PASSO 2 – Run PLS Analysis on ERP data	13
PASSO 3 – Show PLS Result for ERP DATA	15
Come interpretare i risultati della PLS	18
Mean-Centering PLS	18

Come utilizzare questo tutorial

Questo tutorial è organizzato in due sezioni . Nella sezione 1 sono spiegati i principali comandi della PLS in maniera abbastanza dettagliata. Nella sezione 2, “Un esempio pratico” è spiegato, passo per passo, come fare un’analisi PLS. Un buon modo per familiarizzare con le analisi PLS può essere guardare *prima* l’esempio pratico (per familiarizzare con il programma) e dopo leggersi la sezione uno di questo tutorial, che approfondisce più nel dettaglio i vari aspetti e le varie opzioni.

SEZIONE 1

Introduzione teorica

La PLS (Partial Least Squares) è un’analisi statistica multivariata particolarmente utile nelle situazioni in cui siano diverse variabili dipendenti, potenzialmente molto correlate tra loro, molte variabili indipendenti e non tante osservazioni a disposizione. È dunque un tipo di analisi molto indicata per le analisi degli ERP.

Fondamentalmente la PLS serve per rispondere alle due domande fondamentali delle analisi ERP: *Dove?* e *Quando?*.

Con *Dove* si intende, in che elettrodi sono presenti degli effetti?

Con *Quando* si intende, in quali finestre temporali sono presenti degli effetti?

Oltre a questo la PLS da due informazioni molto importanti:

1) Permette di identificare delle “**variabili latenti**” che spiegano il pattern di risultati. Queste variabili sono ortogonali (in un certo senso, “indipendenti” tra di loro). Questo significa che gli effetti trovati su due diverse variabili “latenti” indicano risultati che possono essere considerati come statisticamente dissociabili. Riguardo questa dissociabilità statistica riportano conclusioni del genere West (2007) e West e Kropinger (2005).

2) Permette di identificare associazioni tra effetti in diverse finestre temporali. Ad esempio in memoria prospettica in risposta ad un cue prospettico si osserva in genere una variabile latente che cattura un effetto in siti occipito-parietali intorno ai 200-400 ms (N300) e nella stessa variabile latente un effetto su siti frontopolari in una finestra temporale più tardiva (600-1000 ms). Il fatto che questi due effetti vengano “catturati” della stessa variabile latente suggerisce che essi siano degli eventi correlati tra loro.

Rispondendo a queste domande la PLS permette quindi di superare i limiti della scelta “classica” di finestre ed elettrodi (o cluster di elettrodi) su cui fare le analisi, che in genere si basa sull’ispezione visiva dei tracciati e sulle informazioni presenti in letteratura. Questi approcci tradizionali hanno due principali problemi: hanno insita avere una tendenza al bias di conferma e possono portare a trascurare altri risultati interessanti.

Per un’introduzione teorica alla PLS si rimanda all’articolo (scritto molto bene) di Lobaugh, West e McIntosh (2001).

<http://onlinelibrary.wiley.com/doi/10.1017/S0048577201991681/abstract>

Installare il software per la PLS

Il software per effettuare la PLS gira su matlab, quindi occorre avere installato matlab. Il software va scaricato dal sito del Rotman (<http://www.rotman-baycrest.on.ca/index.php?section=345>). Una volta scaricato occorre copiare la cartella all’interno della cartella toolbox di matlab e aggiungere il path ai percorsi automaticamente caricati da Matlab

Preparare i files per utilizzare la PLS

Per potere effettuare la PLS sui dati i dati devono avere una struttura un po’ particolare.

- 1) Ogni soggetto deve avere una cartella.
- 2) All’interno di ogni cartella ci deve essere un file .txt per ogni condizione.
- 3) Questo file .txt deve avere in riga gli elettrodi e in colonna i punti del file. Questo file .txt deve essere il file di average del soggetto rispetto alla condizione. Ad esempio. Il file del soggetto 1 per la condizione “ongoing” è un file .txt che rappresenta l’average (del soggetto) per la condizione “ongoing”. Se quindi ci sono 30 elettrodi, ogni sweep dura 2 secondi e il campionamento è a 500 Hz, il file avrà 30 righe e 1000 colonne (500*2).

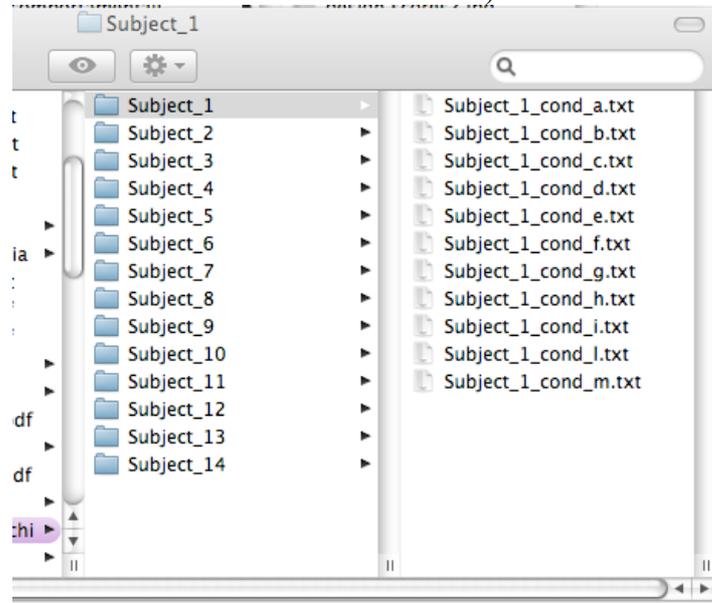
Per semplificare dei passaggi successivi è buona norma che i soggetti abbiano un nome progressivo (es. Subject_1, Subject_2, etc.) e che **anche i file con le condizioni abbiano un nome progressivo.**

Nella figura sotto mostro una organizzazione ideale del file. In questo esperimento c’erano 14 soggetti e 11 condizioni. Ci sono 14 cartelle (una per soggetto) e all’interno della cartella 11 file con le condizioni. Questi files hanno un nome alfabeticamente crescente.

In un passaggio successivo delle analisi sarà possibile associare ad ogni file un'etichetta per ciascuna condizione.

Se i file sono organizzati in questa maniera si risparmia un sacco di tempo nel programma per PLS.

E' accortezza di chi crea il file assicurarsi la coerenza dei file dei soggetti. (Es. rispetto all'esempio sotto, che tutti i file che hanno cond_a, si riferiscano effettivamente alla stessa condizione).



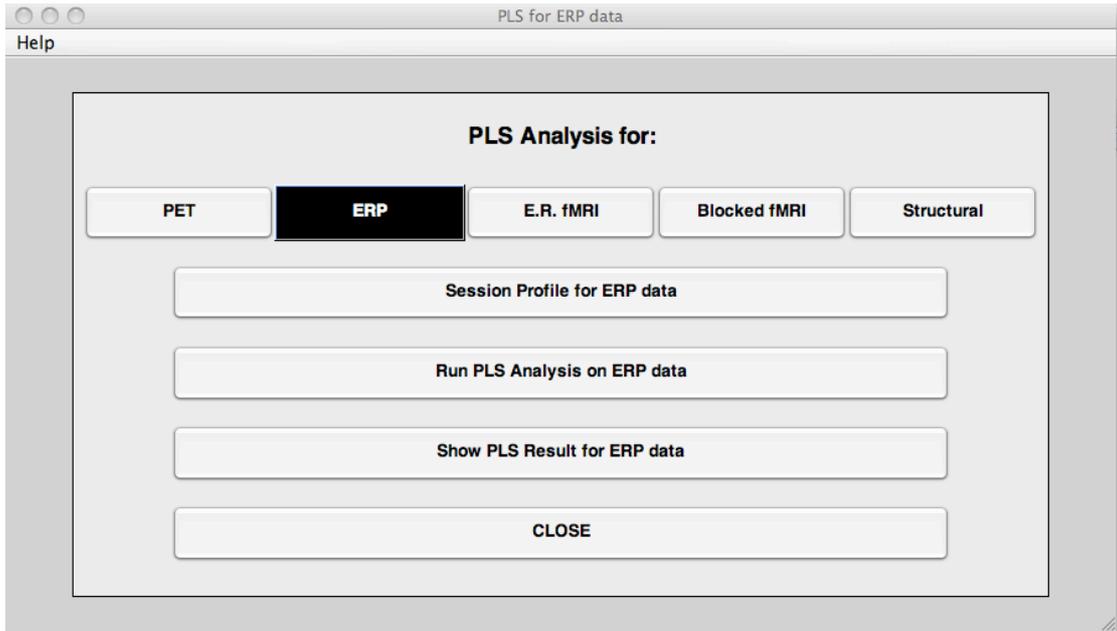
Utilizzare il programma PLSgui

Nel prompt di matlab scrivere

```
plsgui
```

Viene richiamata la gui per effettuare le analisi

Inizialmente è selezionata l'opzione per fare le analisi PET. Selezionare ERP.



1

I tre passi dell'analisi tramite PLS

2

L'analisi PLS si suddividerà sostanzialmente in 3 passi.

3

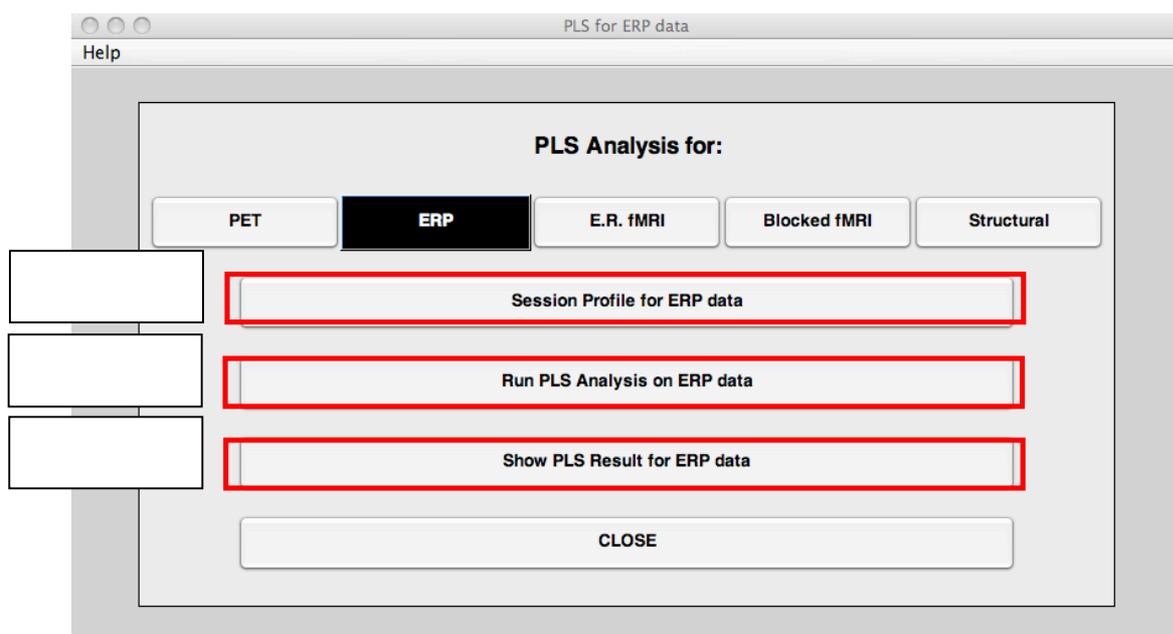
PASSO 1: Creare una **ERPsession**. Cioè importare tutti i files dei soggetti (vedi sopra) e specificare una serie di parametri (intervallo di campionamento, nomi degli elettrodi, nomi delle condizioni). Dalla sessione ERP si creano gli **ERPdatamat**, cioè dei dati pronti per essere analizzati dalla PLS. A partire dalla stessa ERPsession (che in genere include tutti i dati dell'esperimento), in genere si creano differenti ERPdatamat, che sono subset dalla stessa ERPsession. Ad esempio a partire dalla stessa ERPsession si possono creare ERPdatamat che includono solo alcuni soggetti, solo alcune condizioni, solo alcuni elettrodi, o solo certi intervalli temporali.

PASSO2: Effettuare le analisi. Le analisi statistiche associate alla PLS si basano su permutazioni e bootstrap. Per questa ragione ci mettono un po'.

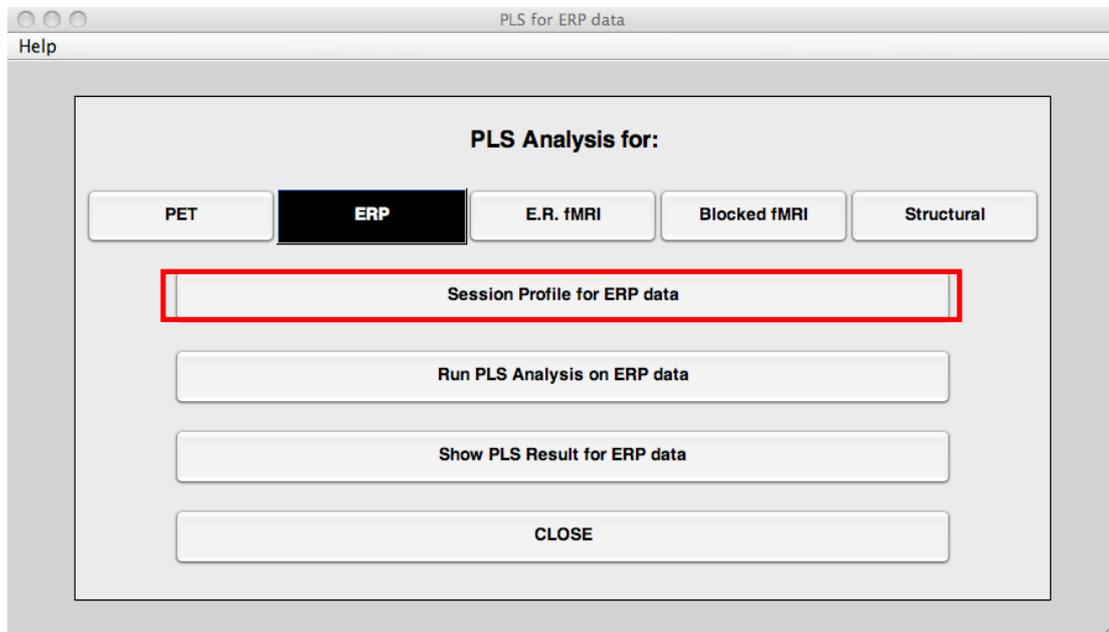
PASSO3: Mostrare i risultati. Nel terzo, e ultimo passo, si osservano i risultati.

NOTA: Durante i vari passi verranno creati numerosi file che termineranno con vari nomi **ERPsession**, **ERPdatamat**, **ERPresults**. È importante che modificando i nomi dei file, questi terminino sempre con questi "suffissi" (più l'estensione .mat). Questo perché il software per la PLS utilizza in automatico (per cercare i file nelle cartelle) un sistema di filtri. Se si modificano questi nomi c'è il rischio di avere difficoltà nel trovare i file corretti (oltre al rischio di fare confusione). Ad esempio va bene il nome "mio_esp_ERPsession.mat".

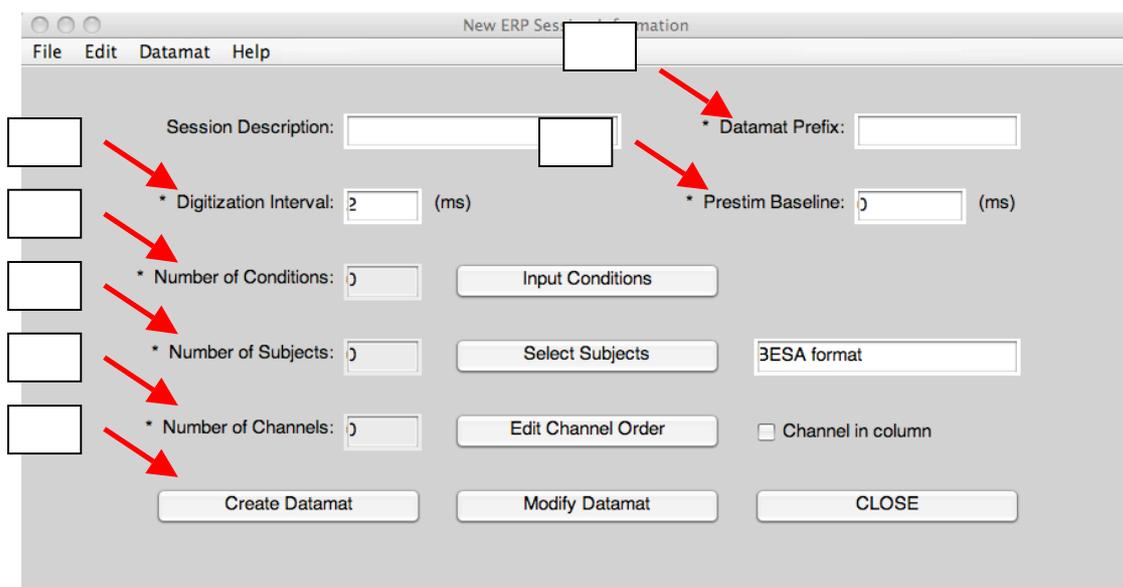
Non va bene il nome "mio_esp_ERPsession_prova.mat", perché in questo caso il nome del file (prima dell'estensione) non finisce con ERPsession).



PASSO 1 - Session Profile for ERP data.



Premendo la prima opzione per gli ERP (**Session Profile ERP data**) appare la finestra sotto. Tutte le opzioni segnate con un asterisco (*) sono da inserire obbligatoriamente. È una buona idea riempire questi campi nell'ordine in cui sono riportati i numeri nella figura.



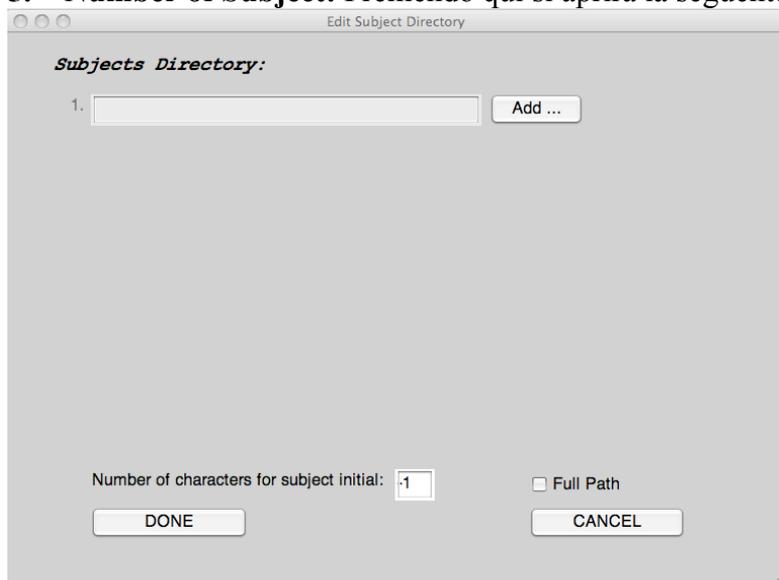
1. **Datamat Prefix:** qui va inserito un prefisso che verrà inserito automaticamente a tutti i file creati da qui in poi (in particolare a partire dal file ERPdatamat).

- Digitalization interval:** indica l'intervallo temporale che c'è tra ogni due punti dei file (i files .txt di ogni soggetto). Si ottiene a partire dalla frequenza di campionamento secondo la seguente formula: $(1/\text{freq.camp.}) * 1000$. Quindi per esempio se la frequenza di campionamento è 256 il digitalization interval è circa 3.9. **NOTA:** attenzione ad utilizzare il punto come decimale!
- Prestimulus Baseline.** La baseline in ms. Deve essere preceduta dal segno – (meno)
- Number of Condition:** Premendo qui si aprirà la seguente finestra.

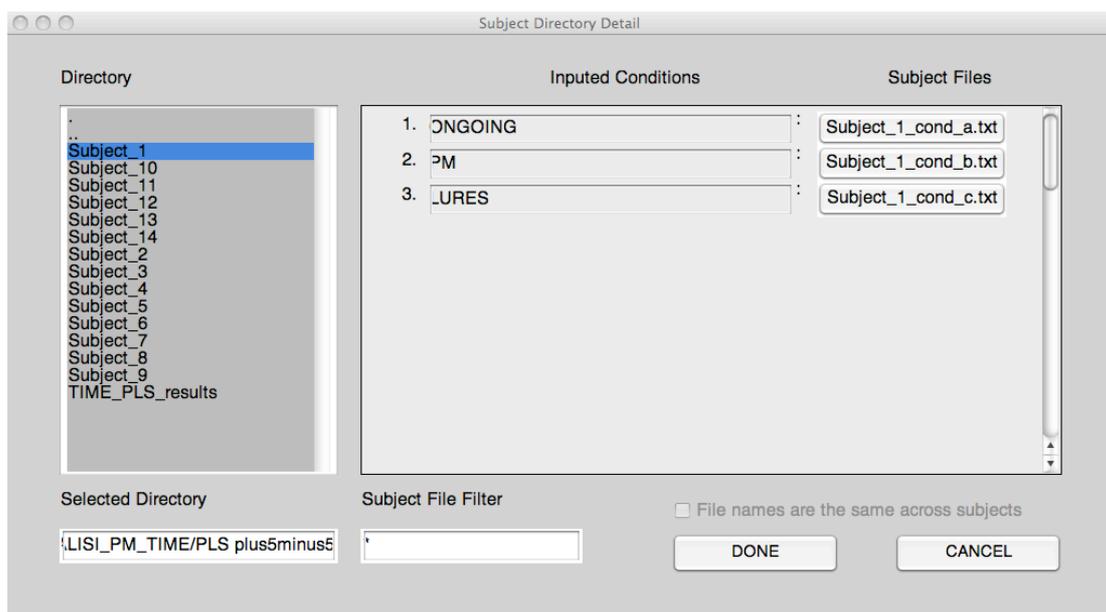


In questa finestra vanno inseriti manualmente i nomi delle condizioni del nostro esperimento. Dopo avere scritto il nome di una condizione, premere *Add* per aggiungerla. Ripetere finché non si sono inseriti i nomi di tutte le condizioni. Per semplificare le procedure è importante che i nomi delle condizioni vengano messi nello stesso ordine dei nomi dei file .txt dei soggetti. Ad esempio. In relazione all'esempio riportato nel paragrafo **Preparare il file per utilizzare la PLS**, Se inserisco le 10 condizioni che hanno nomi che finiscono *_a.txt*, *_b.txt*, *_c.txt*, etc. allora il nome della prima condizione inserita si riferirà al file che termina per *_a.txt*, il secondo al file che termina per *_b.txt* e così via. Premere *Done* quando si ha terminato

5. **Number of Subject:** Premendo qui si aprirà la seguente finestra.



NOTA: Questa finestra non si aprirà se prima non sono state specificate le condizioni sperimentali. Questa volta premendo *Add*, si aprirà quest'altra finestra



Questa finestra permette di scegliere la cartella con le condizioni del soggetto (è nella sezione directory, sulla destra). Se i files con le condizioni del soggetto sono stati specificati in accordo alle istruzioni fornite nel paragrafo sopra **Preparare il file per utilizzare la PLS**, allora in automatico le condizioni sono associate correttamente ai files.

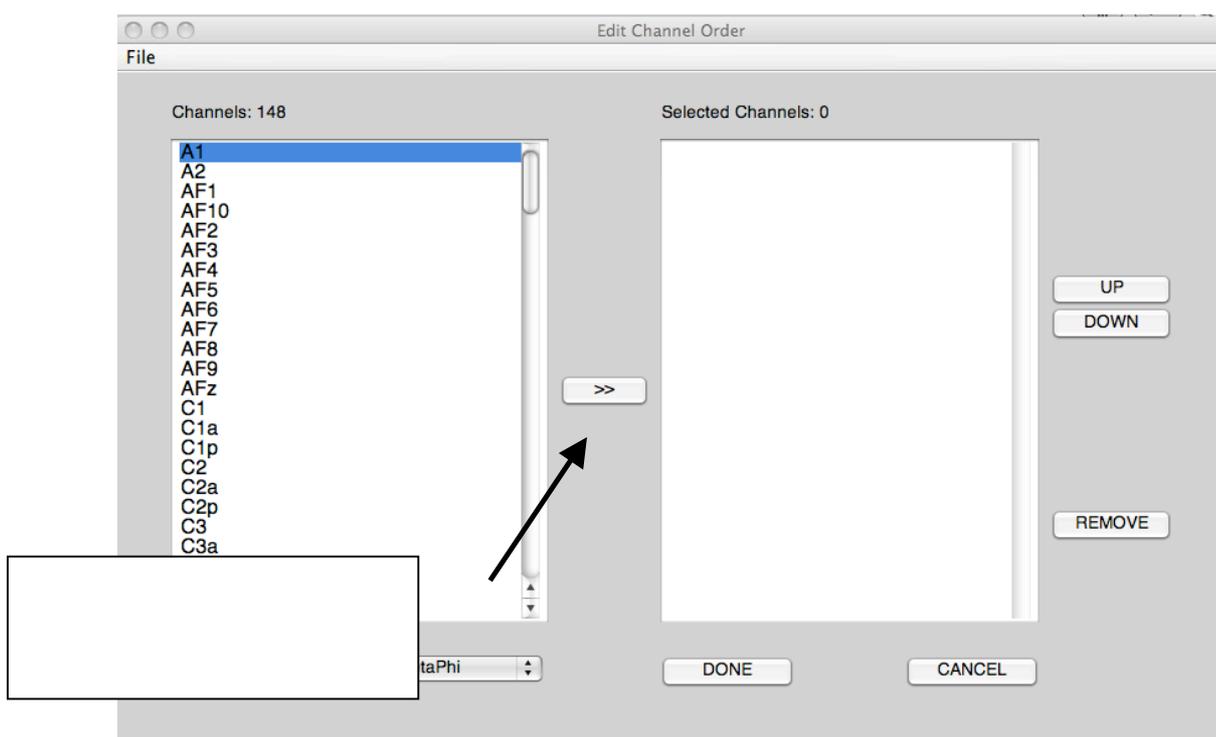
NOTA: In questo punto (sempre se i files .txt sono stati in maniera coerente e organizzati secondo quanto riportato prima). È possibile selezionare TUTTI i soggetti insieme (è sufficiente trascinare il cursore del mouse lungo tutte le cartelle dei soggetti quindi premere DONE). Premere nuovamente *Done* quando si ha terminato.

6. Edit Channel Order: In questa finestra si selezionano gli elettrodi utilizzati. È importante che vengano riportati nello stesso ordine con cui sono nei file .txt. Ad esempio supponiamo che nel file .txt di ogni soggetto le prime due righe siano Fp1 ed Fp2, allora bisogna mettere in ordine Fp1 ed Fp2 in questa schermata. Una volta selezionato il nome di un elettrodo sulla sinistra premere il pulsante con le due frecce al centro, tra i due riquadri, per indicare che quell'elettrodo è presente. Nella finestra a destra (quella intitolata "Selected Channels") il numero degli elettrodi deve ovviamente essere uguale al numero di righe dei file .txt.

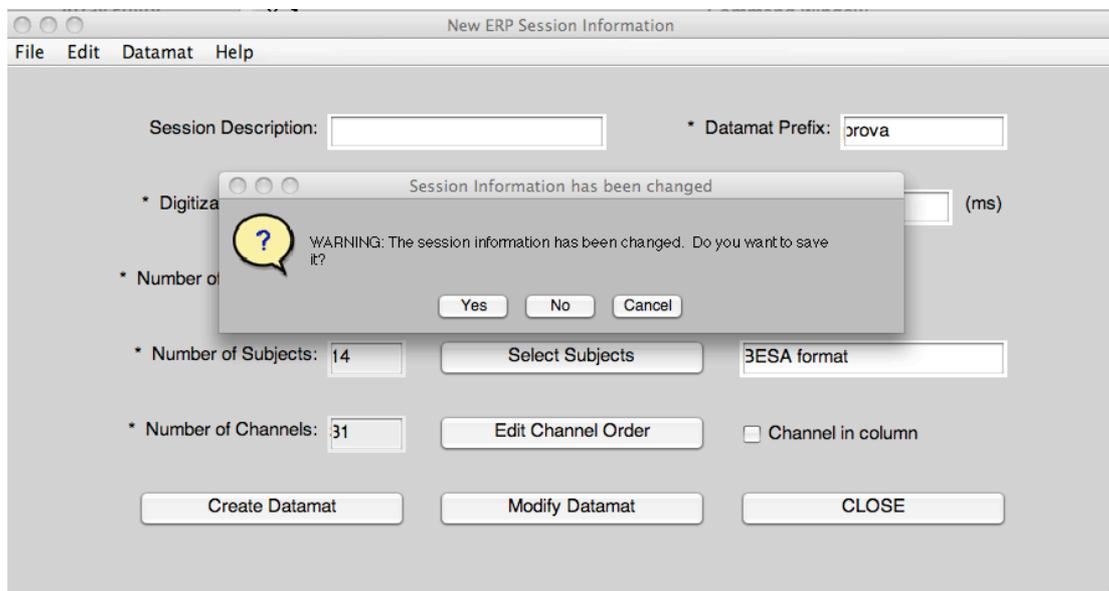
er includere un dopo averlo)
 Eventualmente si possono usare i pulsanti UP e DOWN (alla destra) per cambiare l'ordine degli elettrodi qualora siano stati inseriti in un ordine non corretto.

PRESTARE MOLTA ATTENZIONE A NON SBAGLIARE L'ORDINE DEGLI ELETTRODI!!!

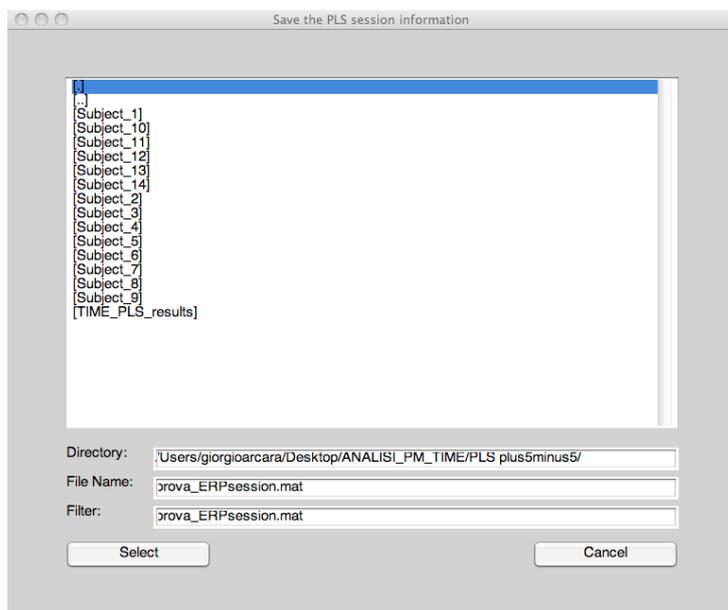
NOTA: questo procedimento può essere lungo e un po' noioso. Se in passato si sono create delle ERPsession con lo stesso settaggio di elettrodi una possibile strategia è possibile caricare queste ERPsession (Dalla schermata principale della Session Profile, File→Load e caricare la ERP session). Una volta caricate sostituire, soggetti, condizioni e gli altri parametri opportunamente, lasciando inalterati i canali.



7. Create Datamat: premendo questo tasto si crea appunto un file ERPdatamat (vedi sopra). Come prima cosa si aprirà questa finestra. In questo momento si salverà il file **ERPSession** (questo file “generale” che contiene tutte le informazioni).



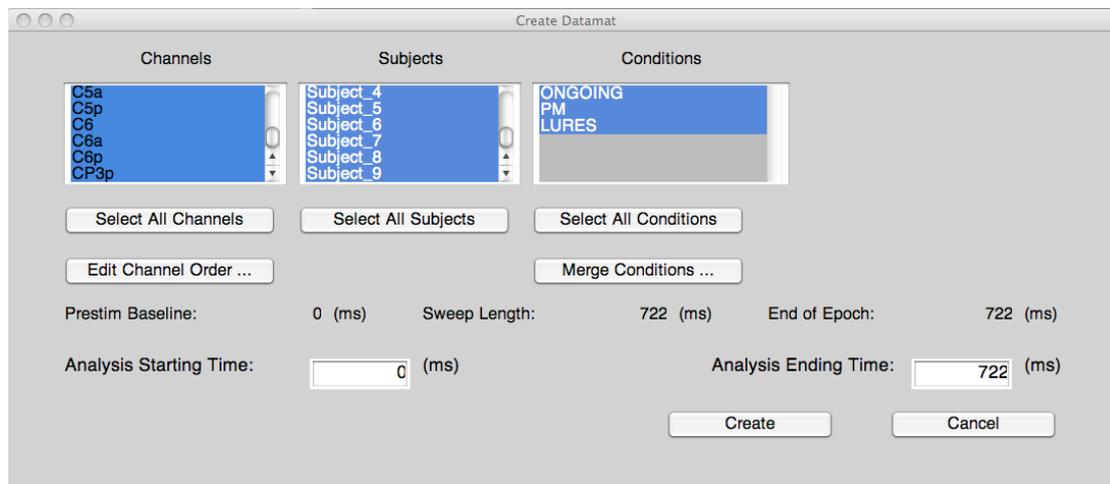
Chiederà di salvare un file, che avrà come prefisso il prefisso impostato (vedi punto 1.) e terminerà per ERPsession.mat



Eventualmente da questa schermata si può cambiare il nome del file che si sta creando (ma mi raccomando, non togliere il suffisso ERPsession.mat).

Dopo avere premuto “Select” si aprirà questa finestra. Da questa finestra è possibile selezionare quali canali, quali condizioni (e l’intervallo temporale) per creare il file **ERPdatamat** che è il file su cui effettivamente verranno effettuate le analisi. Cliccare “Create”, per creare il file datamat, che avrà lo stesso prefisso impostato nella ERPsession.

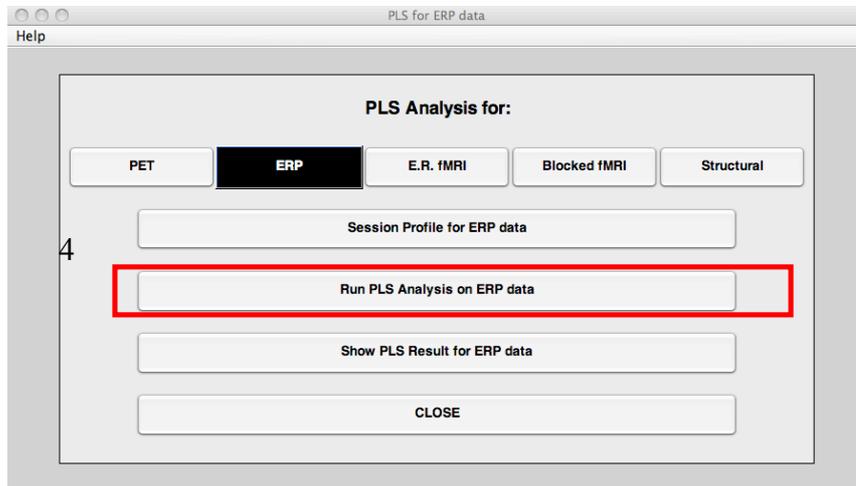
NOTA IMPORTANTE: In questa fase in genere si escludono gli elettrodi oculari (o mastoidi) e l’epoca selezionata non include l’intervallo di baseline. Altrimenti nelle analisi PLS vengono inseriti anche questi dati che determinano l’inserimento di rumore, e altri dati da elaborare, non necessari.



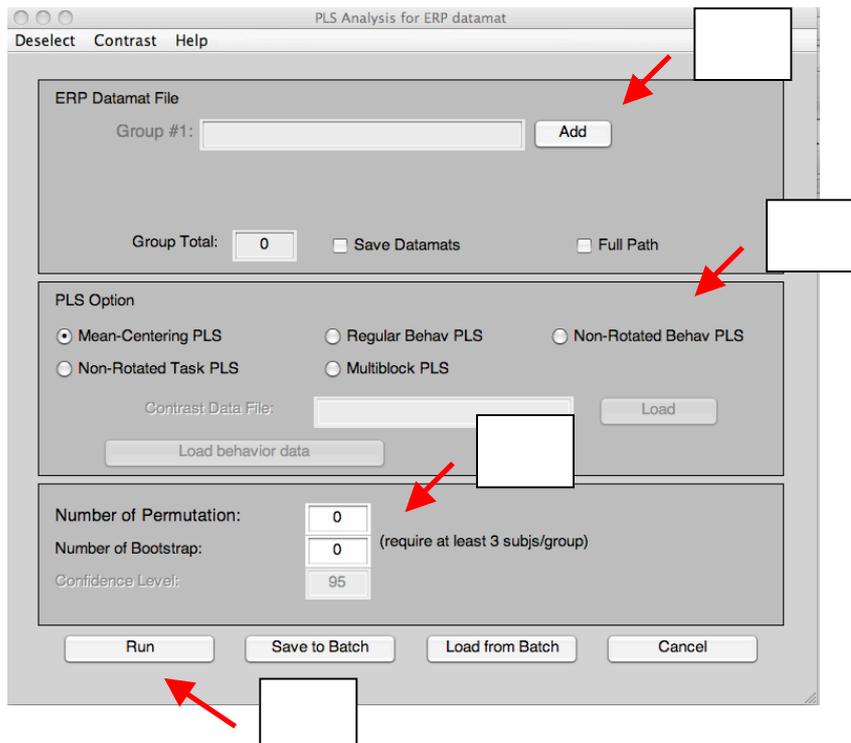
PASSO 2 – Run PLS Analysis on² ERP data.

Una volta creato il nostro file ERPdatamat possiamo effettuare la nostra analisi PLS. Per effettuarla andare nella schermata iniziale della gui e selezionare la seconda opzione: **Run PLS Analysis on ERP data**

3



La finestra che comparirà è quella riportata sotto



1. ERP datamat File. In questo riquadro si seleziona il file datamat su cui effettuare le analisi. Premere Add per aprire una finestra che permette di selezionare il file. Se si devono fare le analisi su più datamat (questo può accadere nel caso di un'analisi BETWEEN) è possibile continuare ad inserire altri datamat. Questi datamat, ovviamente devono essere consistenti in quanto a numero di elettrodi, numero di condizioni etc.

2. PLS option. In questa sezione è possibile selezionare che tipo di analisi PLS fare. A meno che non si conoscano le altre è meglio utilizzare la selezione di default cioè la Mean-Centering PLS.

Una breve spiegazione di cosa è questa Mean-Centering PLS (presa perlopiù dall'articolo di Bisiacchi, Schiff, Ciccola, Kliegel, 2009 e dall'articolo di Lobaugh, West, McIntosh, 2001).

In questo tipo di analisi inizialmente la matrice di dati ERP (che include tutti i soggetti) è centrata rispetto alla grand mean. La media di ogni task è dunque espressa come deviazione intorno a 0. Viene dunque effettuata una Singular Value decomposition (SVD¹) su queste matrici per ottenere le variabili Latenti (LV). Tramite questa SVD sono ottenuti tre output, che ci permettono di interpretare la relazione tra ampiezza ERP e disegno sperimentale. Il primo è un vettore di *singular values*, che rappresentano la magnitudine (non pesata) di ogni LV e viene utilizzata per calcolare la percentuale di *cross-block covariance* riconducibile a ciascuna LV².

Il secondo e il terzo output fornito da questa SVD sono i *Design Saliences* e le *electrode saliences*. Le *Design Saliences* servono per calcolare i *Design Scores*, che indicano in che modo una data variabile latente si esprime nelle diverse condizioni.

Le *electrode saliences* indicano invece il *dove* e il *quando* una determinata LV si esprime (i punti, in termini temporali e di elettrodi in cui sono presenti delle differenze).

Un'altra output cruciale è dato dagli *scalp scores*, che vengono calcolati a partire dalle *electrode saliences* di tutti gli elettrodi e indicano quanto ciascun soggetto esprime il pattern di una data LV.

Questa spiegazione può sembrare un casino (e in effetti lo è). Penso che non valga la pena stare a rileggerla per capirci qualcosa. Il modo migliore per capire cosa sono tutti questi valori introdotti e a cosa servono è leggere la sezione successiva, che spiega come sono mostrati i risultati della PLS.

¹ La Singular Value Decomposition è un particolare tipo di operazione matriciale.

² È importante notare che **la cross-block covariance NON È la stessa cosa della varianza spiegata**. Dal momento che nei risultati alla *cross-block covariance* viene associata una percentuale (vedi sezione risultati della PLS) si può confondere con la varianza spiegata (come per l' R^2 della regressione multipla, per capirci). In realtà qui non stiamo parlando di varianza dei dati ma di "covarianza" tra i blocchi. In soldoni, la percentuale di *cross-block covariance* indica quanto covarianza esistente tra i dati sulla base del disegno sperimentale (in un certo senso, quanta "correlazione" tra le varie condizioni), è riconducibile ad una specifica variabile latente.

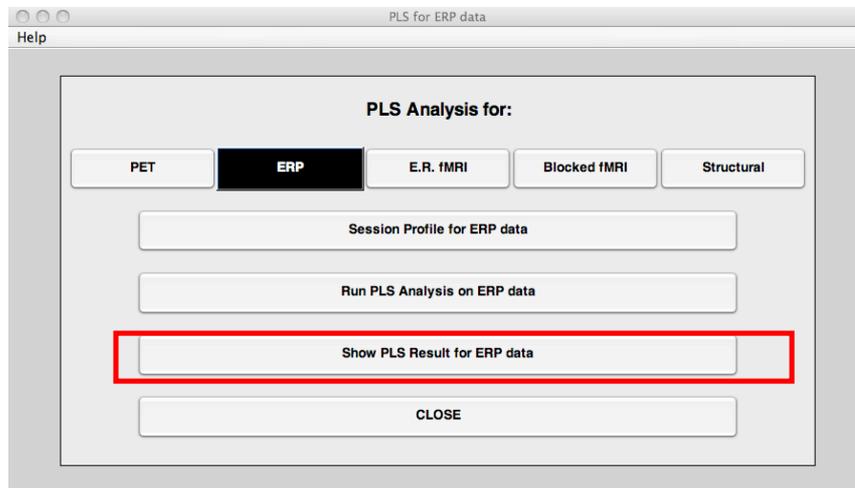
Si noti come invece la "varianza spiegata" (in una regressione multipla ad esempio), indichi un'altra cosa, cioè quanta variabilità dei dati è spiegata dal modello statistico

3. PLS option: In questa sezione è possibile specificare due parametri che servono per ottenere i valori di significatività statistica della PLS. Nell'articolo di Lobaugh, West e McIntosh (2001) sono riportati di inserire 1000 come Number of Permutation e 200 come number of bootstrap. Gli stessi valori si trovano in vari paper che utilizzano la PLS. Suggestisco pertanto di inserire sempre questi valori.

4. Run: Una volta impostati tutti i parametri precedenti. È possibile fare partire le analisi (che impiegheranno qualche minuto a completarsi) premendo appunto il pulsante "Run". Verrà innanzitutto chiesto con che nome salvare il file. I file saranno salvati in con il nome che termina per **ERPresult**. Quindi si chiederà con che percentuali di valori si vuole fare il bootstrap (default 50%), **lasciare il valore di default** (il programma stesso suggerirà di lasciare il valore di default).

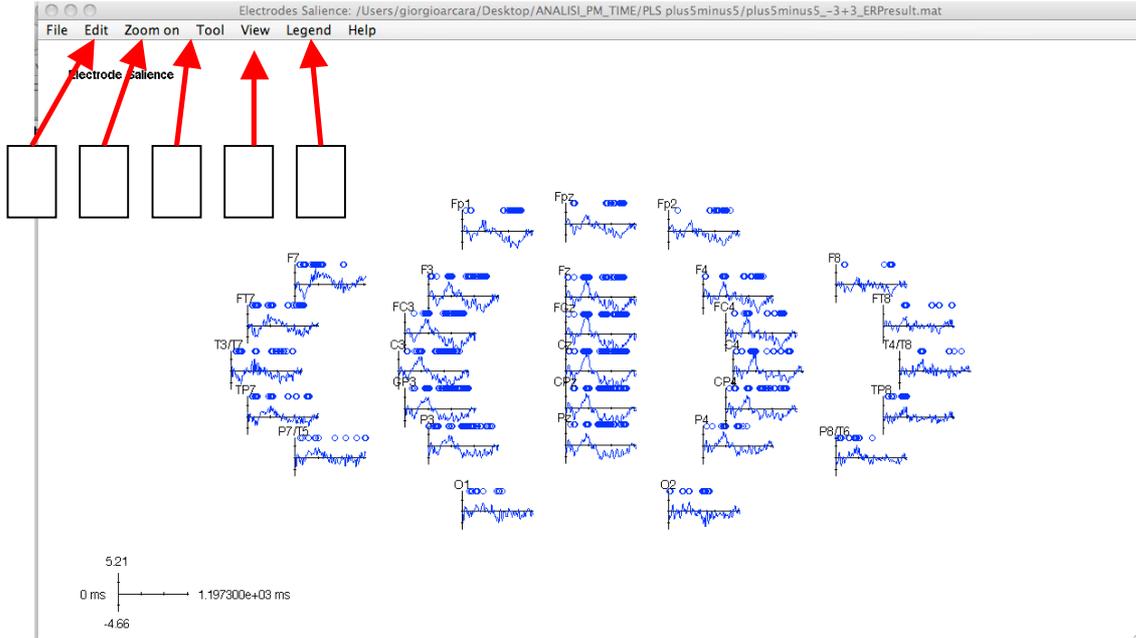
PASSO 3 – Show PLS Result for ERP DATA.

Una volta che si sono effettuate le analisi è possibile osservare i risultati. All'interno della finestra per mostrare i risultati sono presenti un'infinità di opzioni. È pertanto il caso di "esplorarle" personalmente. Di seguito riporto soltanto le opzioni più importanti e come interpretare i principali risultati di una PLS.



Premendo il pulsante **Show PLS Result for ERP data** verrà aperta una finestra per caricare il file (con filtro 'ERPresult.mat', vedi nota 1 per chiarimenti). La prima schermata che si aprirà mostrerà la electrode salience della prima variabile latente.

2 3 4 5



Di seguito elenco le principali funzioni associate a queste 5 opzioni.

NOTA: in seguito all'attivazione di alcune funzioni da questa barra, è possibile che le opzioni a disposizione cambino. Ad esempio se da View (punto 3) attivo Subject Amplitude, le opzioni disponibili in Tool cambieranno!

1. **Edit** tramite questo tasto è possibile selezionare gli elettrodi (sono possibili varie opzioni)
2. **Zoom on.** Tramite questo comando è possibile attivare/disattivare lo zoom. **ATTENZIONE:** se nella barra c'è scritto zoom on significa che lo zoom non è attivo (e occorre premere per attivare), se invece c'è scritto zoom off lo zoom è attivo (e bisogna premere nel pulsante zoom off per disattivarlo).
3. **Tool:** In questa opzione sono presenti tre voci.
 - a. **Display Option** si imposta di quale variabile latente si deve vedere la *saliency*.
 - b. **Bootstrap Option** si imposta di quale variabile latente si devono vedere i risultati del *bootstrap*. **NOTA IMPORTANTE:** In questa finestra si possono anche impostare positive e negative threshold del bootstrap. Questa threshold viene in maniera automatica calcolata dal programma ma non è chiaro se può essere lasciata o vada comunque reimpostata. Cambiare questo valore cambia i valori che vengono segnati come timepoints in cui la saliency è significativa. Nell'articolo di Lobaugh, West e McIntosh si riporta che dovrebbe essere 2, ma in alcuni articoli che usano la PLS si vede chiaramente che è impostata in maniera diversa (spesso più alta di 2). **Sembra che i valori di threshold del bootstrap vada riportato in un eventuale articolo.** Il maximum ratio e il minimum ratio indicano i massimi e i minimi di saliency osservati nei dati.
 - c. **Export data**, serve per esportare alcuni dei dati legati ai risultati, è un comando avanzato per cui non importante.

4. View: questa opzione è racchiude la maggior parte delle opzioni cruciali per vedere i risultati.

- a. Subject Amplitude:** selezionando questa opzione si aprirà una finestra con il plot dei dati ERP dei singoli soggetti (questa volta le ampiezze e non le salienze). Andando su Tool→Display option è possibile selezionare i soggetti da plottare. Se sono selezionati più soggetti verrà rappresentato un butterfly plot con quei soggetti.
- b. Average Amplitude:** selezionando questa opzione si aprirà una finestra con il plot dell'average dei dati ERP (questa volta le ampiezze e non le salienze). Andando su Tool→Display option è possibile selezionare quali condizioni plottare. Se selezionate più condizioni l'ERP risultante sarà l'average tra le condizioni selezionate.
- c. Plot Detail:** prima di selezionare questa opzione occorre selezionare uno o più elettrodi. Per selezionare gli elettrodi si può cliccare sul nome degli stessi, oppure selezionarli tramite le opzioni disponibili in **Edit**. Un elettrodo selezionato avrà il nome evidenziato in grassetto. Premendo **Plot Detail** dopo avere selezionato degli elettrodi farà aprire una finestra in cui verrà mostrato un ingrandimento dell'elettrodo.
- d. Plot singular Values:** selezionando questa opzione vengono rappresentati i Singular Value, uno dei risultati della PLS (poco interessante).
- e. Plot Design Scores:** questa opzione permette di mostrare i Design Scores, uno dei risultati principali della PLS. Dalla schermata che si aprirà si può selezionare di quale LV si vogliono osservare i Design Scores. Quando si apre questa schermata vengono anche mostrati gli Scalp Scores, utili per identificare dei soggetti outliers.
- f. Task PLS scalp scores with CI (INCERTO):** con questa opzione si mostrano gli stessi risultati dei Design Scores, ma questa volta con i CI (Confidence Intervals) associati a ciascuna condizione. Suppongo che questi CI permettano di concludere se due condizioni differiscano tra di loro. NOTA: anche se nel grafico nell'asse delle y si parla di brain scores, sono gli scalp scores (brain scores e scalp scores sono sinonimi).
- g. Plot temporal scores (INCERTO):** con questa opzioni dovrebbero essere plottati i temporal scores per soggetto. Cioè quanto ciascuna variabile latente si esprime nel tempo in ogni soggetto diviso per condizione. Ciascun grafico per soggetto dovrebbe risultare dall'attività mediata da tutti gli elettrodi.
- h. Subject Amplitude for Given Timepoints (INCERTO)**
- i. Channel Time:** opzione per togliere dalla schermata principale (quella che inizialmente mostra la salienza) il nome degli elettrodi
- j. Channel Axis:** opzione per togliere dalla schermata principale (quella che inizialmente mostra la salienza) con i grafici gli assi
- k. Channel Tickmark:** opzione per togliere dalla schermata principale (quella che inizialmente mostra la salienza) con i grafici i tickmarck (le tacche sugli assi).

5. Legend: Attiva/disattiva la visualizzazione della **Legenda**

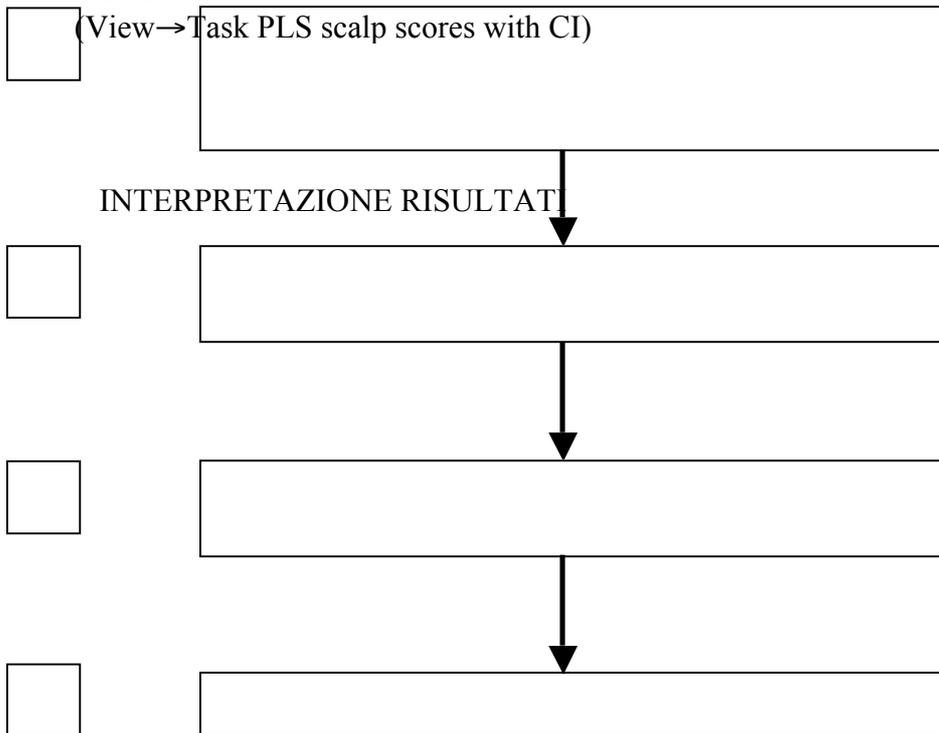
Come interpretare i risultati della PLS

(Schermata iniziale)

Mean-Centering PLS

Di seguito è riportato il significato dei vari risultati della PLS e l' algoritmo con cui procedo per l' interpretazione dei risultati della PLS. Sono anche riportate alcune considerazioni sulle interpretazioni

APPROFONDIMENTO VARIABILI LATENTI



STEP 1. Verifica significatività delle variabili latenti.

Come primo passo premere View→Plot Design Scores. Comparirà una figura come quella sotto. Sostanzialmente la PLS identifica un numero di variabili latenti pari al numero di condizioni inserite (quindi ad esempio nel nostro caso 6). Le variabili latenti sono riportate in ordine rispetto alla quantità di covarianza cross-block viene spiegata da ciascuna variabile.

Tramite i pulsanti a sinistra è possibile selezionare quale LV mostrare. La prima cosa da guardare è il riquadro in alto che oltre a riportare la quantità di covarianza cross-block è spiegata dalla variabile latente riporta un *p value*. Questo *p value* indica se la variabile latente è significativa e quindi se vale la pena discuterla. Tale significatività è ottenuta tramite il test di Permutazione (vedi sezione Run PLS Analysis on ERP data) e indica quanta probabilità c'è di avere ottenuto per puro caso questa variabile latente. Se questa probabilità è bassa (< .05) si considera il risultato come significativo.

Una volta assicurati che il risultato è significativo è possibile osservare come le condizioni si comportano.

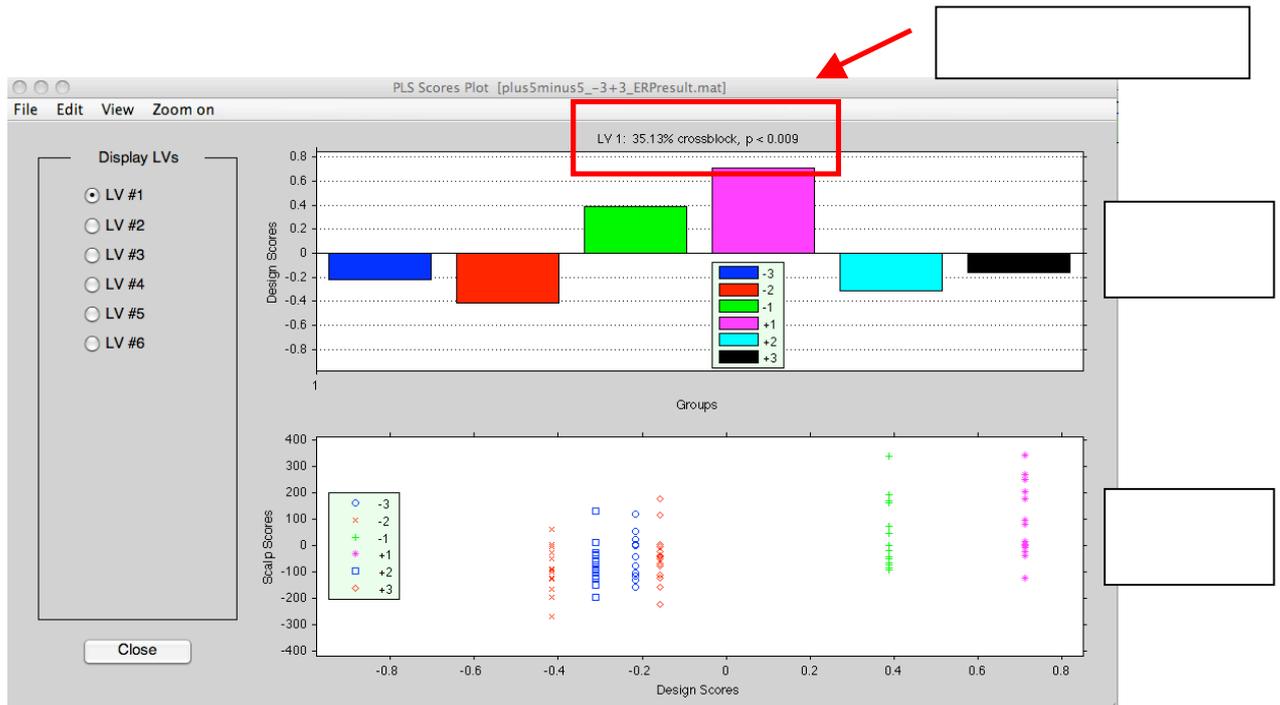
Il modo con cui procedo è innanzitutto osservare quali condizioni hanno Design scores positivi (nel mio esempio le condizioni -1 e +1) e quali negativi (nel mio

Design Scores

esempio le condizioni (-3, -2, +2, +3). Questa prima analisi qualitativa mi dice come si organizzano le condizioni.

Fatto questo cerco di vedere (sempre qualitativamente) quali condizioni sono associate maggiormente a questa variabile latente. In altre parole in che variabili c'è un Design Score più alto (o più basso). Ad esempio nell'esempio sotto si può vedere che la LV1 è catturata soprattutto dalla condizione +1, quindi dalla -1. Sembra che riesca soprattutto a distinguere queste condizioni, dalle condizioni -2 e +2. La LV1 si esprime invece poco nelle condizioni -3 e +3.

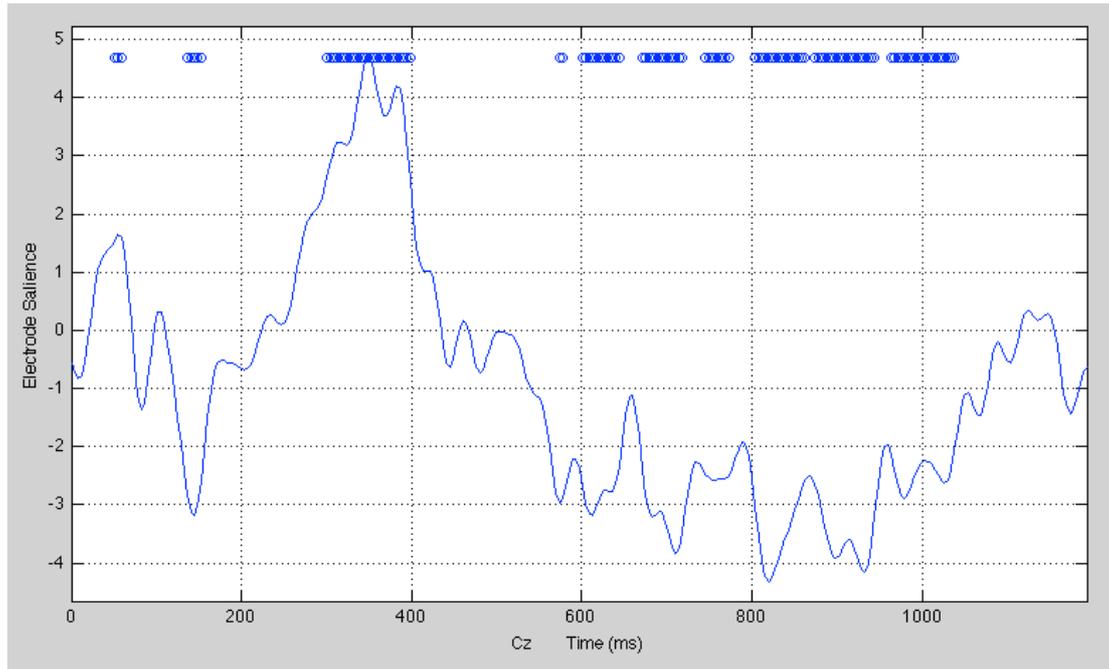
Da notare che la somma dei Design Scores positivi e la somma dei Design Scores negativi sono in valore assoluto uguali e pari (circa) ad 1.



Nella parte in basso del grafico è possibile invece vedere come si distribuiscono gli scalp scores (che ricordo sono dei punteggi riassuntivi per ciascun soggetto che indicano quanto un soggetto esprime il pattern di una data LV). Ogni simbolo rappresenta un soggetto. Ad esempio nell'esempio sopra ci sono 14 soggetti. Segue che per ogni condizione ci sono 14 simboli. Cliccando su un simbolo viene identificato a quale soggetto corrisponde e si evidenziano anche tutti i punti (per le altre condizioni) che corrispondono a quel dato soggetto.

Questo grafico è fondamentale perché ci permette l'identificazione di eventuali outliers. Anche se la procedura statistica alla base del calcolo della significatività dovrebbe proteggere dall'influenza di outlier, nella stessa guida della PLS (che si trova sul sito del Rotman), è suggerito di provare a togliere gli outliers e vedere se l'identificazione delle variabili latenti cambia rimuovendo questi outliers.

Il grafico sopra riporta una distribuzione degli scalp scores relativamente buona. Forse c'è un sospetto outlier nello scalp score di un soggetto per la condizione +1 (identificata dal simbolo + negli scalp scores) che è un po' troppo alto. Se vogliamo essere scrupolosi potremmo cliccare su questo simbolo per identificare il soggetto e



(INCERTO): Nella Mean-Centering PLS a ogni condizione viene in un certo senso associato un certo “peso” (vedi STEP 1). Ad esempio nell’esempio sotto la condizione +1 ha un Design Score maggiore della condizione -1 per cui ha un “peso” maggiore per questa variabile latente. Ciò dovrebbe riflettersi anche sulle salienze. Nell’esempio una salienza positiva indica che le due condizioni con Design Score positivo hanno una ampiezza maggiore di quelle con Design Score negativo, con un maggior peso della condizione +1 (perché ha appunto un Design Score maggiore) rispetto alla condizione -1.

NOTA IMPORTANTE: è importante assicurarsi che la salienza mostrata e il bootstrap mostrato siano effettivamente quelli della variabile latente che vogliamo studiare (per cambiarle Tool→Display Option e Tool→Bootstrap Options). La prima volta che si osservano i risultati PLS vengono mostrate come default salienza e risultati bootstrap della prima variabile latente, ma ho notato che se si osservano i grafici di altre variabili latente, poi si chiude il programma e poi si riapre vengono mostrate le ultime impostazioni utilizzate (quindi potenzialmente altre variabili latenti e non la prima.)