**Nota generale**

Dobbiamo trovare più opzioni così lei ci può bocciare/approvare.

**Proposte di analisi su events recordings**

* Confronto tra medie (popolazioni normali) di tempi di risposta: vorremmo capire se lo switching ha impatto sulla media del tempo di risposta in schizofrenici.
  + Problemi di normalità: possiamo confrontare media di due popolazioni se una trasformata con boxcox? (Abbiamo provato con popolazione tempi di risposta sani e schizofrenici, entrambi non normali.) NO

No dobbiamo trasformarle entrambe in caso

* + Chiedere: ha senso distanza di Mahalanobis per togliere outliers? 7.5??
  + Test asintotici direttamente? Abbiamo 125 sani e 50 schizofrenici (>30). 🡺 fare test asintotici sulle medie dei tempi di reazione
  + Ha senso fare i test nonparametric? No

Soluzioni: se dobbiamo dimostrare che due vettori sono normali allora dimostro che la loro unione è normale. Sui tempi di riposta nelle varie combinazioni sì che possiamo far test tra le medie ?

* 1. Da fare in futuro: One-way ANOVA sui tempi di risposta(response times) con le 4 combinazioni ( schz-switch, schz-noswitch, control-switch, control-noswitch)

2. Da fare in futuro: Multi-way ANOVA su tempi di risposta (response times): spiegare le performance dei schz vs sani nelle varie condizioni sperimentali (switch/no switch, congruency/incongruency, correct response/incorrect response?).

* + Fattori: Schizofrenico, switch, congruency, correct (pochi dati però?)

DUBBI:

* + Problemi di normalità?
  + Chiedere: (v. risultati da ricerca Barratt e test affini): ha senso definire un nuovo indice di performance? Potremmo eventualmente usare lui come risposta al posto del solo tempo.
    - Proposte per indici di performance:

(possiamo fare riferimento alle percentuali di risposte esatte, ma bisogna tenere conto del fatto che sono valori percentuali che vanno mappati tra –∞ e +∞)

* Da fare in futuro: Random effects model per studiare impatto di durata del CSI/ITI (durata di tempo di Cue/Intervallo tra tasks: preparazione al successivo, ovvero tempo di preparazione al task) su variabilità di response time in schizofrenici. Sono stati osservati solo due cue time durations – long e short.
  + Ci aspettiamo che uno schizofrenico possa diventare incostante/imprevedibile se sottoposto a stimoli veloci; impaziente se tenuto troppo in attesa.
    - Proposte: collego con Barratt?
* Partendo dal fatto che: congruency/incongruency, switch/no switch sono treatments (condizioni che possiamo imporre durante l’esperimento) , mentre la correct response è un output dell’esperimento dunque va considerato come target

→ Facciamo una classificazione per vedere se i tempi di risposta influenzano la correttezza della risposta (potremmo dividere la popolazione fra sani e schizofrenici)

-> Ha senso invece classificare per capire se schizo o no? (anche in generale sul dataset)

Può essere usata come covariata anche la correct response in base al nostro scopo (non ai fini predittivi) ai fini del classificatore come modello esplorativo

**Proposte di analisi su zmaps**

* One-way ANOVA su zmap ridotta per regioni: g=2 groups, p=83 features=83 regioni:

Disponiamo già di valori di attivazioni medi per regione cerebrale (su ogni individuo): dividiamo individui in due gruppi (sani e schizofrenici) e cerchiamo features (regioni) che si attivano diversamente nei due gruppi.

* + Scopo desiderato: riduzione dimensionale, fissare attenzione per indagini successive solo su regioni rilevanti

DUBBI:

* + La zona 28 dobbiamo ignorarla, giusto? (Correlazione di segnale con sé stesso) NO, va tenuta.
  + Correlazione negativa nel singolo soggetto non significativa ai fini di interpretazione fenomeno? Significa che quei neuroni restano fermi a baseline?
  + Diffusione chimica di ossigeno in cervello farà sembrare attive zone che non lo sono?
  + Eccessiva dimensionalità? (82 regioni).
  + Presenza di dati mancanti (togliendo a prescindere i soggetti che hanno un NA nella linea scendiamo da 175 soggetti a 86 soggetti); comunque sulle regioni costruiamo tutti valori esistenti.
  + Come collegare con il resto delle analisi.
* 1. Segnale bold (coefficiente di Pearson)

Se c’è un valore negativo quelo che succede è che va al contrario (ovvero le aree si stanno spegnendo)

2. I dati mancati NA possiamo trascurarli perché non sono i zone di nostro interesse

* + - Proposte:
      1. Legame (tentativo di regressione?) tra regioni cerebrali rilevanti e tempo di risposta? (interpretare coefficienti di regressione per impatto su tempo di risposta)
      2. Collegare regioni specifiche a livelli alti di Barratt?
* Altri tipi di analisi su zmaps?
* PCA??? AIUTO: perché NON FUNZIONA? È colpa della variabilità troppo uguale di tutte le regioni? Dovrebbe funzionare per ridurre la dimensionalità
* Clustering su individui zmaps:

Esistono gradi/sfumature/tipi di schizofrenia?

Si, vedere se poi altre variabili che influenzano la schizofrenia (esempio: indice di massa corporea)

**Proposte di analisi su phenotype health**

* Molti dettagli sulla salute degli individui, più uno score che riassume il tutto.
  + Teniamo solo questo indice riassuntivo?

**Proposte di analisi su phenotype demographics**

* Abbiamo raccolto dai file “phenotype” i dati riguardanti il fumo (altre informazioni ci sembravano meno utilizzabili).

Abbiamo ricostruito il numero di sigarette fumate al giorno dal numero di pacchetti fumati al giorno (20 sigarette/pacchetto), il numero di anni di fumo dai mesi; e per i fumatori/ex-fumatori abbiamo plottato i dati essenziali ovvero: n° di sigarette fumate al giorno vs n° di anni per cui si è fumato.

Da una PCA e a occhio vediamo che la PC più importante è il n° di anni.

Consigliano di vedere se uno è fumatore o meno.

* + Ha senso utilizzare questa informazione congiuntamente con Barratt per indicazione su impulsività?
  + Dovremmo buttare informazione su numero di sigarette (ok?). Indice composto di entrambe grandezze poi standardizzato: indice “artificiale” di intensità di fumatore?

PCA

* + 18 dati mancanti su 51 osservazioni per ex fumatori (da quanto hanno smesso): inutilizzabile?

UPDATES POST 01/04 (incontro al MOX):

**Altre idee/possibilità di analisi (non ancora approfondite, generale):**

* Analisi normalità, medie etc su Barratt (eventualmente usando dati su popolazione generale da ricerca Cost)
* Lollo: ripartiamo da articolo 4 e 5 per idee test/interpretative
* Metodi non parametrici (permutazionali) per confronto medie su campioni non gaussiani (y=tempo di risposta): (approfondimento/l’anno scorso lo hanno spiegato ultimo giorno di lezione) -> Carlo
* Metodi non parametrici (permutazionali) per ANOVA/MANOVA etc su campioni non gaussiani (y=tempo di risposta): (approfondimento/l’anno scorso lo hanno spiegato ultimo giorno di lezione) -> Carlo
* Scott: coding of categorical variables-> why square root of something that looks like Hamming’s distance (but it is not?)
* Scott: Clustering of zmaps per analizzare dendrogramma e controllare se esistono “sfumature” di schizofrenia; ripetere con jittering dei dati. Da capire:
  + Incompatibilità con riduzione dimensionale (PCA)?
    - proiezione su seconda componente principale per catturare variabilità?
  + definizione di Distanza su zmaps di due individui
* vedere se nel tempo il tempo diminuisce e se le risposte corrette aumentano (analisi parallela). forse ci serve indice di performance. l'idea delle derivate temporali è abbastanza interessante, vedi sopra = ANALISI FUNZIONALE

**Idee di analisi confermate da incontro al MOX**

* opzione 1. One-way ANOVA sui tempi di risposta(response times) con le 4 combinazioni ( schz-switch, schz-noswitch, control-switch, control-noswitch)

opzione 2. Multi-way ANOVA su tempi di risposta (response times): spiegare le performance dei schz vs sani nelle varie condizioni sperimentali (switch/no switch, congruency/incongruency, correct response/incorrect response?).

* Problemi di normalità? Hanno detto provare e se abbiamo fortuna ok
* Indici di performance possono essere costruiti, controllare letteratura: pecentuale di risposte esatte?
* Facciamo una classificazione per vedere se i tempi di risposta influenzano la correttezza della risposta (potremmo dividere la popolazione fra sani e schizofrenici)
* Altre riduzioni dimensionali di zmaps? (v sotto però)
* Tenere solo alcuni indici per phenotype health e phenotype demographics (consigliati cigs cigs e BMI)

**Lollo, rush produttivo 03/04**

Sulla scorta di lettura articoli: test asintotici per provare che gli schizofrenici non hanno difficoltà aggiuntiva rispetto a sani per performare lo switch, ma sono generalmente più lenti nel task in entrambe le condizioni dell’esperimento.

Sviluppi possibili: NON LO SO ☹

Secondo rush produttivo poco concludente:

Ho cercato un pattern nei tempi di risposta medi per controllare se in media i soggetti migliorano la velocità facendo tanti trials. Risposta = NO

costanza pensa che non sia inutile

**Cos(t), 04/04**

Ho lavorato sullo script script\_zmap\_2 per fare PCA.

Versione aggiornata: script\_zmap\_2\_COST.

Su materiale -> plots ho caricato C\_plot\_1 che sono le varianze cumulate spiegate da ogni componente (viene decente, prima componente spiega quasi 45% di variabilità)

Su materiale-> plots -> PCA loadings ci sono jpeg che sono i pesi delle 82 regioni del cervello per le prime 10 componenti (ho scelto 10 arbitrariamente).

Lasciar perdere ANOVA a questo punto per zmaps perchè abbiamo già risultato di riduzione dimensionale accettabile e confermabile da loro risultati (v screenshot da locandina sotto) ??

il mio cervello sta producendo cose, spero non sceme:

perchè lasciar perdere la anova? ora sono mezzo convinto che se tu proietti gli 83-features data in un 20 features o forse 9/10 (tombari mostrava la 9th) possiamo applicarci la anova (one moment)

era per dire che mi sembra più papabile come dimensional reduction la pca che si interpreta piuttsto che tenere delle cose rilevanti solo per un possibile scelta di risposta che è quello che faremmo con anova

cioè possiamo includere in unica anova grossa già citata gli scores delle componenti principali che teniamo, così da avere un trattamento con tot livelli (anzichè 83 livelli)

più che trattamento è una covariata

ll treatment è “valore dello score nelle varie PCs”, il livello è “valore dello score nella specifica PC”

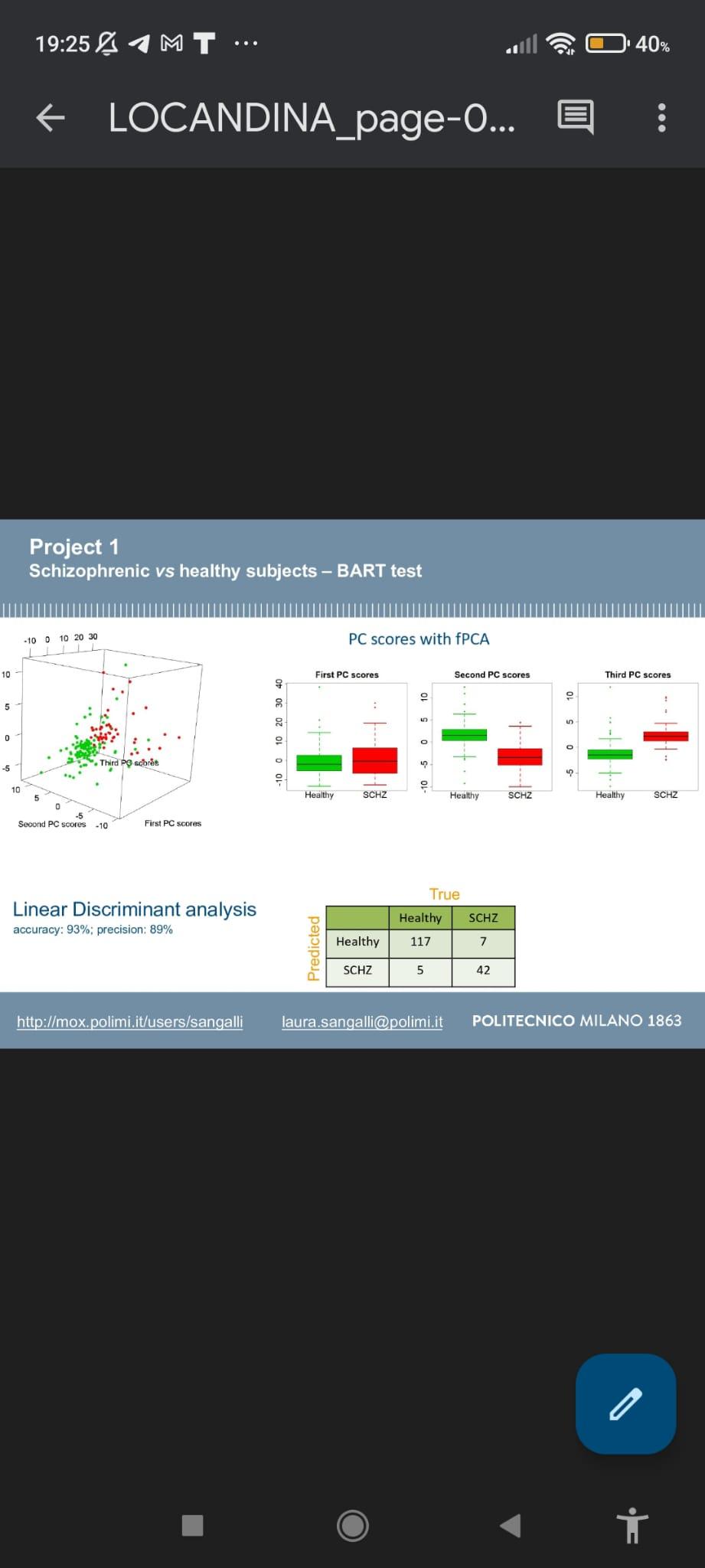
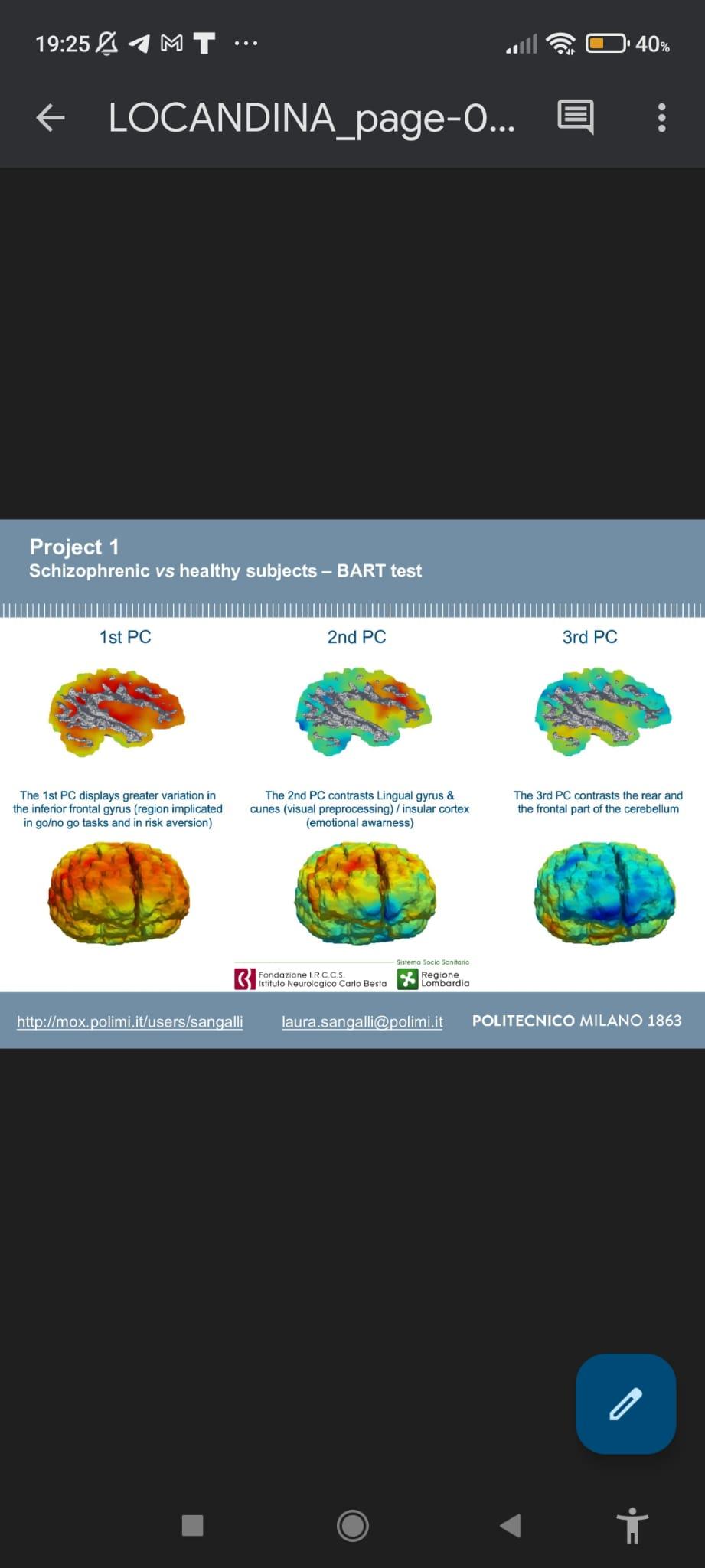
però sai che farlo per regioni non è malvagio

anova solo sulle zmaps

eh ma noi lo facciamo per PC e non per regioni APPOSTA per dare ad ogni PC un significato medico

Sviluppi possibili:

1. Interpretare loadings (eg prima componente principale:=”generale attivazione cerebrale”, etc) basandosi su funzionalità di zone cerebrali ed usando come ispirazione risultati su schizofrenici (ma bart test):



1. Provare a fare stessa PCA distinguendo tra sani e schz, interpretazione di risultati diversi?
2. Collegare scores “anomali” di individui lungo direzioni principali a Barratt “anomali”?