

Benassi 27 marzo

Il 23 aprile chi non ha passato il parziale può ricominciare da zero...la valutazione negativa non incide per niente (dice); la relazione si può fare lo stesso.

La prossima settimana fa vedere gli esempi sbagliati e fa il ripasso generale.
[perdiamo un sacco di tempo discutendo dei parziali]

Tecniche descrittive

Due principali: cluster analysis e analisi discriminante. Permettono di semplificare le relazioni tra le var che si studiano, e quindi di comprendere meglio il fenomeno da un punto di vista particolare. Entrambe fanno riferimento a dei modelli, i quali sono esclusivamente da considerare come esplorativi, perché ci offrono una prospettiva particolare da cui comprendere il fenomeno, che non è unica né onnicomprensiva. In entrambi i casi si fa riferimento ad un modello, creato ad hoc sui dati sperimentali e quindi ha molte limitazioni nella generalizzabilità. Il modello può essere ulteriormente modificato, prodotto attraverso diverse interazioni, passi successivi che permettono il suo affinamento. Vengono proposti diversi modelli e lo sperimentatore in alcuni casi deve scegliere quali tra questi sia il migliore, e non sempre la scelta può essere sostenuta senza critiche, implica arbitrarietà nella decisione. Comunque nonostante l'arbitrarietà l'utilizzo non è limitato, perché queste tecniche esplorative permettono di formulare queste idee, di avere una prospettiva sul fenomeno che si studia, che permette di formulare ipotesi di ricerca; sono utilizzate per validare gli strumenti sia dal punto di vista della validità interna che da quella esterna.

Analisi discriminante

Come la cluster, viene usata per suddividere il campione in gruppi; si è di fronte ad una situazione in cui l'obiettivo è individuare dei sottogruppi per vedere se caratterizzano il fenomeno. Es. la dislessia evolutiva è caratterizzata da sintomi molto diversi tra loro, e al deficit fonologico si associa a volte un deficit visivo, a volte attentivo, altre volte un deficit dell'apprendimento implicito. Questi diversi tipi di var sono analizzati con diversi strumenti ed è utile unire le var per vedere se permettono di identificare dei sottogruppi caratterizzati da peculiarità specifiche. Quindi la cluster analysis e la discriminante possono essere utili. L'analisi discriminante si caratterizza perché è un metodo basato su due fasi principali: una di addestramento e una di analisi. Ha bisogno di un criterio di suddivisione del gruppo preesistente. Su questo criterio costruisce una funzione in grado di discriminare i soggetti analoghi di riferimento.

[fa vedere un esempio. Dobbiamo avere una var che permette di discriminare il campione; cioè per alcuni soggetti io conosco già l'appartenenza al gruppo. Dai dati io sono già in grado di distinguere i gruppi, cioè dislessici e controlli?]

Assunti:

- i fattori predittivi devono avere distribuzione gaussiana (quantitativi);
- i fattori devono essere scarsamente correlati tra loro;
- se esistono delle correlazioni, esse devono essere costanti all'interno dei gruppi;
- le medie e deviazioni standard dei fattori non devono essere correlate tra loro.

I fattori costituiscono i parametri necessari per discriminare i soggetti (nella fase di addestramento), vengono considerate le variabili che mi permettono di distinguere i soggetti. La **fase di addestramento** è quella in cui si crea il modello della funzione

discriminante, in cui la variabile discriminante diventa determinante nel calcolo dei parametri. La funzione discriminante considera l'apporto che ogni soggetto dà al gruppo; nella funzione discriminante è possibile valutare il valore dei b per ogni soggetto. Il valore di b è dato dalla fase di addestramento. La var di gruppo è sempre qualitativa (indica se si appartiene al gruppo 1 o 2 ecc). Io mi aspetto che le var che considero fattori siano determinanti nella divisione in gruppi della var stessa.

Nella fase di addestramento valuto quanto la distinzione in gruppi sia influenzata dalle variabili indipendenti che considero. Per ogni var valuto l'influenza e il senso dell'influenza. Si crea la funzione discriminante; si calcolano i parametri che danno l'indice di quanto i valori considerati influenti lo siano effettivamente (nella discriminazione tra i gruppi). Es. quanto l'attenzione è importante per distinguere dislessici e non? Quanto la capacità visiva? Ecc. I gruppi vengono dunque definiti a priori nella var che inserisco nel modello in questa fase. Es. da considerazioni cliniche si può sapere se i sogg sono dislessici o no.

In che modo si creano i sottogruppi, anche per soggetti non classificati? In modo che ci sia max differenza tra le medie di gruppo, quindi costruisce un modello attraverso la funzione discriminante in modo che i due gruppi siano al massimo diversi tra loro (media 0 e varianza 1). Dopo la fase di addestramento c'è la **fase di analisi**, in cui i parametri vengono utilizzati per classificare i soggetti per i quali prima non avevamo alcuna classificazione; si applica la funzione discriminante ai soggetti che non erano stati raggruppati. La classificazione si basa sul calcolo del valore soglia, al di sotto del quale i soggetti si classificano al gruppo 1 e al di sopra del quale nel gruppo 2.

La relazione lineare è di tipo additivo: l'effetto di ciascun fattore viene sommato all'effetto degli altri. Questo è il motivo per cui si spera che le var siano poco correlate tra loro, perché se la var relativa a questo parametro è molto correlata con quella relativa ad un altro, se le sommo perdo un'info dovuta alla loro relazione.

[seguono domande varie, lei risponde sventolando un esempio in SPSS sui bambini dislessici...]

La funzione discriminante mi calcola il valore b, dovuto nell'esempio al punteggio per le varie variabili (visiva, attenzione ecc). Se il sogg ha avuto valori che non possono distinguerlo tra i due gruppi, allora la sua probabilità di appartenere ad un gruppo o ad un altro sarà calcolata con un risultato vicino al 50%, i suoi punteggi non lo caratterizzano bene. I valori globali hanno media 0 e varianza unitaria (1), questo permette di distinguere i due gruppi al massimo. Per ogni soggetto, oltre a b si calcola la probabilità di appartenere a ciascun gruppo; si procede assegnando il soggetto al gruppo 1 o 2, e questa probabilità è associata al valore di soglia. La soglia sta nel punto di mezzo dei valori delle medie dei gruppi. Nel caso più semplice abbiamo 2 gruppi, se sono di più, occorre calcolare più funzioni discriminanti, tante quante il numero di gruppi -1. Questo perché per ciascuna combinazione di fattori occorre calcolare una soglia.

Al termine della fase di addestramento i soggetti vengono riassegnati ai gruppi; può succedere che ci sia concordanza tra le due classificazioni, oppure no. Se non c'è, la funzione discriminante dev'essere considerata non valida per discriminare i due gruppi. Mi aspetto che ci sia una sufficiente concordanza di giudizio tra i risultati della funzione e la var di gruppo che si è utilizzata nella fase di addestramento. Questo è comunque un processo descrittivo, quindi produce una visione particolare di ciò che si studia, che può essere considerata più o meno buona a seconda della corrispondenza tra i risultati e le ipotesi che ho verificato, oltre che a seconda della corrispondenza tra i due valori di giudizio. Anche le stesse var sono valutate come influenti o meno sulla funzione in base al peso che hanno nella possibilità di discriminare nei diversi gruppi. Se abbiamo tutte le var

con un peso (coefficiente) utile per la discriminazione molto basso, ci può far supporre che la loro influenza sia relativa rispetto alla classificazione nei gruppi in fase di partenza.

Quindi il problema è nelle var che abbiamo considerato.

1. ipotesi di partenza, legato alla consistenza effettiva del legame tra fattore di gruppo e fattori quantitativi, e quindi al loro legame;
2. forza dei parametri, legato a quanto ciascun parametro è in grado di discriminare la var di gruppo.

L'utilità della funzione discriminante non sta solo nel classificare die nuovi soggetti in un gruppo di appartenenza, ma anche nell'individuare i fattori che incidono maggiormente nella discriminazione tra più gruppi. Questa misura è data dai coefficienti che ci indicano il peso della relazione tra il fattore studiato e la funzione discriminante. Il peso del valore è utile quando è intorno ad 1. Se è molto al di sotto, es. 0.3, il peso è 30%. Sono valori arbitrari, ma solitamente i valori intorno al 50% vanno bene.