

Benassi 14 marzo

Primo parziale: risultati all'inizio della prossima settimana sul sito della facoltà.

Se non è andato bene ci dobbiamo preparare meglio sulla prima parte, lei interroga quelli dei parziali, Bolzani e l'altra gli altri.

Secondo parziale: relazione su argomento a scelta da presentare il giorno dell'orale, è l'argomento a scelta su cui lei ci fa le domande, la cosa in più è il fatto di portarlo per iscritto; all'orale quindi dopo l'esposizione orale di questo resta solo l'ultima domanda. Due cartelle o un foglio protocollo. La strutturazione ce la dà lei.

Oggi iniziamo con la presentazione dei modelli che vengono utilizzati in psicologia, procedure statistiche avanzate che maggiormente vengono usate in psico. Panoramica generale.

Modelli utilizzati in psicologia

I modelli sono un'applicazione, una tendenza verso cui le scienze, non solo la psicologia, mirano ad arrivare, perchè attraverso un modello è possibile rappresentare la realtà che si sta studiando, il fenomeno; anche in statistica il modello ci permette di definire la relazione tra variabili in modo più generalizzabile possibile. E' possibile utilizzare i modelli in statistica in due diverse accezioni: si possono considerare come generalizzazione delle procedure, oppure si possono costruire per descrivere delle realtà sperimentali complesse. Nel primo caso il modello ha lo scopo principale di generalizzare i risultati e fa riferimento alla statistica inferenziale; nel secondo caso si costruisce il modello su quel campione, quei dati di riferimento, per avere un'idea, una rappresentazione, descrizione di quel campione, quindi si fa riferimento alla statistica descrittiva. Esempi del primo sono: il modello lineare generale (MLG) e la regressione logistica. Il modello lineare generale fa riferimento a tutti i test parametrici, e all'interno del quale vengono accorpate tutte le procedure di tipo parametrico (test), che si riferiscono allo studio di una o più variabili dipendenti, invece i test non parametrici non hanno un modello di riferimento che è in grado di accorpare tutte le caratteristiche diverse dei test parametrici; non c'è un modello vero e proprio in grado di assolvere alle funzioni che assolve il mod lin gen per i test parametrici. Quello che più si avvicina a questo è la regressione logistica, che però ha i suoi limiti. In ambito non parametrico sono stati creati molti test ad hoc per quelle situazioni specifiche e ancora non sono state inventate delle procedure in grado di accorparli tutti. Esempi di analisi statistiche che usano modelli per descrivere realtà complesse sono date dall'analisi fattoriale, MDS, analisi delle corrispondenze, modello di Rasch. L'analisi dei modelli viene effettuata attraverso:

- definizione del modello, definizioni delle variabili che mi interessano. Devo sapere qual è la funzione matematica che mi permette di studiare quelle relazioni. Il MLG presuppone per definizione una relazione lineare tra le variabili, e quindi somma ad una costante α i parametri β relativi alle diverse condizioni sperimentali che si stanno studiando. Quindi la definizione del modello viene data attraverso la funzione matematica che spiega la relazione tra le variabili in studio;
- stima dei parametri: i parametri devono essere misurati sul campione che si sta studiando, la loro stima viene effettuata su un campione apposito, per quantificare effettivamente la relazione tra queste diverse variabili, quindi l'apporto che ognuna di queste ha con la costruzione del modello
- valutazione della bontà del modello: quanto il modello è fedele e buono rispetto ai dati sperimentali raccolti
- in alcune procedure si calcola anche la sig dei parametri; essa dà anche la misura della bontà del modello, è in relazione con questa.

Si è soliti poi definire quei modelli che generalizzano le procedure '**modelli a struttura predeterminata**', perchè la struttura, quindi la definizione, del modello viene fatta a priori rispetto

al campione che poi si raccoglierà.

Esempio: io mi aspetto a priori che il peso di un individuo e la sua altezza siano in relazione lineare. Il mio modello di riferimento in questo caso è il modello della retta. Io so che le mie variabili hanno distribuzione gaussiana e voglio analizzare se la loro relazione è sig, allora valuto la relazione attraverso un modello che penso la rappresenti a priori, prima di raccogliere i dati. Dopo di che, faccio una stima dei valori dei parametri e calcolo i limiti di confidenza, perchè sono nell'ambito della statistica inferenziale, quindi è importante controllare la variabilità, l'apporto dato dall'errore casuale. Attraverso il calcolo dei limiti di confidenza dei parametri è possibile avere anche un'idea della stima dell'errore casuale; questo perchè io mi aspetto che i parametri che vado a stimare siano tutti significativi. Se io ad esempio studio la relazione tra sex di appartenenza, scolarità di ogni soggetto e il voto all'esame nel mio corso presuppongo che la relazione tra questi diversi fattori sia significativa, quindi che attraverso il MGL i valori siano significativi. Questo permette quindi di fare inferenza, di usare il mio modello per verificare le ipotesi sperimentali.

Modelli a struttura stimata: fanno riferimento alla statistica descrittiva. Si stima anche il modello stesso, che quindi non è più definito a priori, viene stimato sulla base dei dati sperimentali. Quindi io ho quei dati, valuto quale dei miei modelli meglio li rappresenta; solo in alcune situazioni calcolo la stima dei parametri attraverso i limiti di confidenza. Quindi la significatività viene perlopiù calcolata sui nuovi dati sperimentali. Mi interessa non la generalizzabilità, ma la descrizione del campione, quindi la sig potrebbe essere calcolata su un campione successivo, in teoria. Questo è il procedimento inverso rispetto a ciò che si fa nei modelli a struttura predeterminata. Mi aspetto di trovare non delle differenze, ma un modello attraverso il quale io riesco a strutturare una situazione analoga a quella dei dati sperimentali, cioè un'uguaglianza tra i dati e il modello, anche se non sarà mai identico, cerco quello che li rappresenta meglio. Esempio: voglio validare un questionario psicometrico. Lo somministro ad un certo numero di soggetto e voglio vedere in che relazione i diversi item stanno tra loro, perchè questo può dirmi come si strutturano i diversi item e come potrebbero corrispondere ai costrutti di riferimento. L'analisi fattoriale è in grado di dirmi questa cosa, riesce a costruire sui miei dati il miglior modello, ovvero riesce a stabilire matematicamente quali sono le relazioni che meglio rappresentano i miei dati; ovviamente una parte dei dati verrà persa, non vengono rappresentati tutti. Quindi l'analisi fattoriale mi permette di estrarre dalle variabili che studiamo un numero n di fattori che sono in grado di restituirci tutte le informazioni dal punto di vista matematico (tutte meno una certa parte che definiamo noi nella procedura del modello) e quindi abbiamo il quadro complessivo di tutte le relazioni possibili dei nostri dati, per esempio se la sottoscala che abbiamo creato per il questionario è adatta.

Ci sono due tipi di analisi fattoriale:

- di tipo esplorativo: usata come modello a struttura stimata, dà una descrizione della realtà e basta;
- confermativa: vuole andare oltre alla struttura stessa dell'analisi fattoriale e cerca di applicare su di essa anche un test di significatività; tuttavia questo tentativo viene fatto in modo sbagliato, perchè si usa il test di significatività per dimostrare H_0 , ovvero si cerca di dimostrare che il modello e i dati sono uguali. Questo però è un errore metodologico, non è MAI possibile dimostrare un'uguaglianza, e oltretutto utilizza un procedimento statistico fatto per cercare differenze! Ha un gran limite, ma viene usata molto in psicologia. Scopo: generalizzare i risultati, il modello ottenuto, ad altri dati, ma usa una procedura scorretta. Alcuni autori si rifiutano di applicarla. E' come allargare le potenzialità dei modelli a struttura stimata alle caratteristiche di quelli a struttura predeterminata. I modelli a struttura stimata non sono adeguati a generalizzare, perchè questi si riferiscono ad un campione particolare. Hanno i limiti e le potenzialità della statistica descrittiva. Servono a riassumere attraverso poche variabili che spiegano la maggior parte dei dati per avere un'idea della realtà complessiva troppo complicata da spiegare altrimenti. In un questionario strutturato è difficile compiere un'analisi, non ci potremmo soffermare su ciascun item per indagare le possibilità del soggetto, diventa più utile analizzare il questionario attraverso una struttura

che semplifica il questionario stesso. Per esempio raggruppando in sottogruppi. Inevitabilmente devono riferirsi, proprio perchè il modello è creato su quei dati, a quei dati sperimentali, e questo è il motivo per cui questi parametri vengono stimati su campioni molto numerosi, perchè questo mi autorizza poi a considerare l'analisi fatta come più affidabile.

Critiche all'uso dei modelli a struttura stimata per generalizzare le procedure: usi criticabili per due motivi principali:

1. non si può dimostrare H_0
2. comunque si fa riferimento ad uno strumento che viene creato per dimostrare delle differenze, e viene usato per dimostrare un'uguaglianza.

-> questi sono gli errori più importanti.

Un esempio di modello a struttura predeterminata è la regressione lineare; attraverso questa io ipotizzo a priori che i miei dati, che rappresentano la relazione tra le due variabili che studio, si distribuiscano su una retta. Quindi dico prima di raccogliere i dati che mi aspetto che la loro relazione sia data da una retta. Se invece voglio descrivere al meglio questi dati, vedo che la retta (di prima) non è esattamente sui miei dati, potrebbero essere descritti meglio da un altro tipo di modello, come una curva. Quindi nel caso di modelli a struttura predeterminata sono interessato non a trovare quel modello che identifica i miei dati ma non in modo perfetto ma a quello che perfettamente aderisce ai miei dati sperimentali. Mi permette di stimare i parametri, quindi di comprendere anche la relazione tra le variabili.

Esempio dell'applicazione di un modello a struttura predeterminata sullo stesso gruppo è dato dall'analisi di un questionario sulle modalità di utilizzo di internet. 5 sottoscale con punteggio Likert, sottoposto a studenti divisi in modo random in due gruppi. Se io applico lo stesso questionario ad entrambi i gruppi per vedere come funziona, cosa produce l'analisi fattoriale, vediamo che in realtà otteniamo due risultati diversi. Io mi aspetto che i due gruppi producano risultati diversi (che l'analisi fattoriale produca sempre risultati diversi), proprio perchè viene ogni volta elaborato un modello matematico su dati specifici, quindi i modelli costruiti sono sempre in modo diverso. Il tentativo è che questo modello sia perfettamente aderente ai miei dati, e quindi dev'essere costruito dopo la raccolta di essi.

Nel caso della regressione multipla sui due gruppi, la relazione tra le var che mi aspetto di trovare nel primo gruppo dovrebbe essere simile alla relazione che mi aspetto di trovare nel secondo. Quindi oltre alla stima dei param ho anche quella della loro oscillazione casuale. Mi aspetto che facendo un'ulteriore misurazione, ovvero un'ulteriore analisi della regressione su un campione analogo al primo, il parametro oscilli intorno alla regione di confidenza che ho misurato nel primo gruppo, ed effettivamente in questo esempio i parametri B, lo standard error (oscillazione di B) e l'intervallo di confidenza ci danno modo di fare inferenze, quindi di generalizzare i risultati ad altri campioni analoghi al primo. Infatti le significatività corrispondono, così come le regioni di confidenza dei parametri (in questo esempio). Questo perchè nei modelli a struttura predeterminata l'affidabilità dei parametri viene misurata tramite la stima della loro variabilità, ovvero si viene ad identificare quella regione di confidenza all'interno della quale è molto probabile trovare il valore vero del parametro, e quindi falsificare H_0 . Si parla dunque di probabilità, che permette di fare una previsione anche su dati successivi, di gruppi analoghi a quello raccolto, anche se non del tutto uguali.

Nei modelli a struttura stimata non c'è una definizione dell'oscillazione casuale dei parametri, quindi non ho una misura della variabilità del numero dei parametri, ma solo una misura della loro capacità di spiegare adeguatamente i dati sperimentali, ovvero della bontà del modello. Nel caso dell'analisi fattoriale, ad es. io so che i miei tre fattori che essa ha estratto spiegano il 70% della variabilità totale, quindi so che perdo il 40% delle variabili che studio attraverso quel modello. L'adeguatezza viene misurata attraverso diversi test che misurano la bontà del modello, ovvero quanto è aderente ai dati sperimentali.

Esistono modelli parametrici e non parametrici nell'ambito di questi modelli.

Caratteristiche di modelli parametrici e non parametrici:

-> i modelli non parametrici analizzano le var di tipo qualitativo, o variabili che non hanno distribuzione gaussiana. Il vantaggio di analizzare una var di questo tipo è che nella definizione della var stessa c'è una maggior certezza, si fa riferimento a presenza o assenza di qualcosa, ad esempio. Lo svantaggio è che avendo una maggior certezza la precisione è minore; più cerco di essere certo riguardo ad un fenomeno più perdo nella precisione della definizione di quel dato specifico. I modelli non parametrici hanno rispetto a quelli parametrici una maggior potenza nella falsificazione di H_0 , sono più sicura che h_0 sia stata falsificata nel modo corretto. Ci sono diversi tipi di test non parametrici, a seconda della situazione sperimentale. Difficilmente quindi è difficile identificare un modello che li raggruppi tutti, come succede nei test parametrici (vedi slide per lista di test a seconda delle serie di dati). Ogni singola situazione ha un test specifico, mentre nei test parametrici non è così. Abbiamo dunque molti tipi di modelli, come le tavole di contingenza e i modelli Log-Lineari (misurano di associazione tra le variabili), modelli regressivi invece stabiliscono una relazione tra le variabili che suppone che le variabili siano legate tra loro da una funzione logaritmica.

In questa lezione ha fatto una panoramica, poi vedremo tutto nei particolari. Solo modelli non parametrici.