

28.02.07

LA FORZA DELL'EFFETTO (effect size)

E' un indice che ci dice quanto è forte l'effetto che abbiamo misurato, (es. quanto è forte l'effetto dell'età sulla memoria)

È stato introdotto x alcuni punti deboli della logica falsificazionista, infatti il test di significatività ha un difetto: se riesco a raccogliere un campione abbastanza numeroso riesco a dimostrare qualsiasi idea, qualsiasi differenza, e quindi è stato introdotto questo parametro perché indipendente dalla numerosità del campione

Questo parametro mi misura la relazione fra **variabilità** dovuta al fattore considerato e la **variabilità** dovuta al **modello** (cioè quella riferita alla somma fra variabilità del fattore e variabilità errore); (la devianza è una misura di **variabilità** che a differenza della varianza corrisponde alla somma dei quadrati degli scarti; rispetto alla varianza manca la divisione per i gradi di libertà).

Il parametro più utilizzato per misurare la forza dell'effetto è il **Partial Eta squared**:

permette di confrontare tra di loro diversi studi che riguardano lo stesso argomento, e quindi di confrontare l'effetto riscontrato in diversi studi, proprio perché indipendente dalla numerosità.

Il difetto di questo parametro è che a differenza della significatività, non abbiamo una stima dell'errore.

[Eta squared corrisponde a un r quadro ovvero il coefficiente di determinazione dell'analisi della regressione lineare; r quadro, come eta squared ci da un indice della bontà del modello può assumere infiniti valori da 0 a 1;

0 significa che il modello non è in grado di rappresentare i miei dati, 1 che i miei dati sono rappresentati perfettamente dal modello, ed i dati stanno perfettamente sulla retta, perché il modello dell'analisi della regressione è una retta.

Cohen, autore riconosciuto a livello internazionale, per la ricerca in ambito psicologico, ha dato degli indici di riferimento x classificare un effetto come buono, medio o basso; dobbiamo sapere però che li ha definiti in modo arbitrario.

Inoltre se vogliamo spss ci da anche la potenza del test; da un punto di vista metodologico è corretto calcolarla prima attorno all'indice che abbiamo stimato, e non alla fine].

CRITICHE ALLA LOGICA FALSIFICAZIONISTA:

1. **ASIMMETRIA**: il fatto di essere tutta portata a falsificare delle ipotesi nulle, e quindi, le ipotesi nulle che non vengono falsificate non vengono prese in considerazione.
2. **RIGIDITA' DELLA SCELTA**: ovvero non si discute su quei risultati che superano anche di poco il limite alfa posto a 0,05 nel falsificare o non falsificare l'ipotesi nulla. (Il ricercatore onesto aumenterebbe il campione, il disonesto invece toglierebbe i decimali).
3. **ARBITRARIETA' DI ALFA** (collegata a quanto detto sopra).
4. **IL PROBLEMA DELLA NUMEROSITA'** (collegato a quanto detto sopra).
5. **NON PERMETTE DI CONFRONTARE DIRETTAMENTE DELLE IPOTESI ALTERNATIVE** (cioè devo sempre confrontare la mia ipotesi di partenza con l'ipotesi nulla, perché inevitabilmente viene sempre a costruirsi un'ipotesi nulla che è un'ipotesi di uguaglianza, e solo attraverso quella posso procedere nella mia inferenza).
6. **ERRORI POSSIBILI NELLA LOGICA FALSIFICAZIONISTA** (che bisogna evitare):

- Non valutare la potenza del test e la grandezza dell'effetto
- Confrontare i valori di probabilità ottenuti (perché le probabilità sono tutte calcolate in modo indipendente, quindi l'indice p non è confrontabile, ma indica semplicemente nel campione analizzato qual è la probabilità d'errore del 1° tipo).
- Utilizzare la logica falsificazionista per dimostrare delle uguaglianze (HO), perché è *UNO STRUMENTO UTILIZZATO X FALSIFICARE L'IPOTESI NULLA*.
- La significatività indica una differenza ma non è né un indice di senso né di forza.

LA LOGICA FALSIFICAZIONISTA E' STATO IL FRUTTO DI 2 DIVERSI APPROCCI:

- FISHER (1935) ha reso esplicito il p-value approach (PVA) cioè quello che oggi usiamo come logica falsificazionista.
- NEYMAN e PEARSON (1933) hanno elaborato il fixed alpha approach (FAA); anch'esso si basa sull'identificazione dell'intervallo alfa, che permette la falsificazione di un'ipotesi, però si è anche soffermato sui diversi errori di decisione (errore alfa, errore beta e potenza del test).

[Ancora oggi in ambito metodologico ci sono delle discussioni a sostegno dell'uno o dell'altro approccio].

Entrambi:

1. confrontano l'ipotesi sperimentale con l'ipotesi nulla.
2. utilizzano il p-value (ovvero un livello alfa arbitrario) per determinare la probabilità del verificarsi dell'evento, e confrontare questo con la probabilità dell'errore causale.

Il metodo FAA rappresenta un metodo alternativo alla logica falsificazionista, in grado di confrontare 2 ipotesi alternative (H1 e HO) e quindi calcolare un valore di p relativo all'ipotesi sperimentale e un valore di p relativo all'ipotesi nulla. (mentre nel metodo di Fisher si calcola solo il valore di p sull'ipotesi sperimentale). **Ciò è possibile perché attraverso la probabilità condizionata è possibile calcolare la p di HO e H1.** Dopodiché vengono confrontati con i valori di alfa e beta; quindi si sceglie tra le 2 quella che riduce al minimo l'errore alfa e l'errore beta. *[La p condizionata misura la probabilità di un evento (ad es. la p. di HO) dato un altro evento].* Questo metodo però non ha portato alla creazione di test statistici opportuni, e per ora ha un senso solo dal punto di vista teorico.

Dobbiamo attenerci ai **consigli dell'APA** (American Psychologist Association) organismo utilizzato come controllo in tutte le ricerche in ambito psicologico, per ovviare agli errori della logica :

- chiarezza nella presentazione dei dati [definire bene le distribuzioni dei nostri dati, presentazione dei dati mancanti (es. soggetti che hanno abbandonato la ricerca) e degli outliers (soggetti che danno risultati anomali e che spesso vengono esclusi dalla ricerca) e cosa comporta la loro mancanza, per non inficiare i risultati]. C'è un'analisi specifica che permette di analizzare anche l'effetto dei dati che sono andati persi: si confrontano i risultati con gli outliers con quelli senza gli outliers per vedere se sono diversi.
- Effettuare analisi il più semplice possibile (ogni volta che è possibile utilizzare delle analisi statistiche uniche (multivariate) per i diversi effetti che si vogliono misurare, rispetto ad analisi spezzettate.
- Rinunciare a presentare il risultato ottenuto come una scelta dicotomica fra l'accettazione o il rifiuto di H1; nella formulazione è preferibile dire: "sono emerse differenze significative". Inoltre occorre corredare il valore di p anche con una misura della forza dell'effetto.

DISEGNO DI RICERCA

E' MOLTO IMPORTANTE **DEFINIRE** IL DISEGNO DI RICERCA CHE STA ALLA BASE DELLA NOSTRA IPOTESI, OVVERO IL PROGETTO CHE CI SPECIFICA QUALI SONO **I LEGAMI FRA LE VARIABILI CHE VOGLIAMO ANDARE AD ANALIZZARE**, QUESTO PERCHE' ATTRAVERSO LA DEFINIZIONE DEL DISEGNO DI RICERCA, E' GIA' POSSIBILE ANDARE POI A VEDERE QUALI SONO I POSSIBILI RISULTATI CHE CI ASPETTIAMO DI TROVARE.

Ci possono essere 2 tipi di classificazione dei disegni di ricerca; non c'è concordanza fra gli autori:

- alcuni preferiscono una distinzione molto formale tra i diversi tipi di ricerca;
- altri hanno un'idea più pragmatica: non esistono dei disegni non sperimentali. (Benassi & Co. si ispirano a quest'ultima).

Classificazione formale:

Disegni sperimentali: disegni nei quali è possibile controllare tutte le variabili in gioco; il campione è raccolto con selezione idonea e casuale; le analisi dei dati sono fatte nel modo appropriato.

D. quasi sperimentali: sono caratterizzati dal parziale controllo di tutte le variabili in gioco.

Esempi:

- quelli caratterizzati da un unico campione, sul quale vengono effettuate misure ripetute che vengono poi trattate come misure indipendenti.
- quelli con campioni non equivalenti ovvero i gruppi sono composti da soggetti che non sono stati selezionati in modo casuale
- disegni ex-post-facto: sono studi che indagano a posteriori qualcosa che è già accaduto.

D. non sperimentali:

Esempi:

- ricerche di osservazione: sono ricerche nelle quali l'attenzione è spostata sulla struttura del problema che si sta indagando (descrittive).
- ricerche d'archivio: ove si fa riferimento a dati raccolti in archivi.
- ricerche sui casi singoli (es. descrizione di un caso grave singolo) secondo la Benassi anche se il campione è costituito da un caso singolo, non significa che non si possa parlare di disegno sperimentale).

01.03.07

Classificazione pragmatica

I disegni sperimentali sono distinti in base alle relazioni fra le variabili indagate nella nostra ricerca.

Disegni fattoriali:

- **sono presenti tutte le possibili combinazioni di fattori**
- **è possibile analizzare tutti gli effetti a cui i criteri di classificazione utilizzati fanno riferimento e tutte le possibili interazioni tra gli effetti principali**

[es. se per caso mi mancassero dei punteggi perché qualcuno non ha risposto ad un questionario, mi mancherebbero alcune combinazioni di variabili e quindi non lo potrei considerare come un disegno fattoriale, perché non sono in grado di analizzare tutte le combinazioni]

i disegni fattoriali possono essere suddivisi in altre sottoclassi:

Disegni entro i soggetti: chiamati anche per prove correlate o prove ripetute nei quali la misura è effettuata sullo stesso campione.

[es. gruppo di madri e gruppo di figli misurati alle scale di depressione vengono considerati come misure correlate; si utilizza il modello lineare generale per prove ripetute e non l'analisi della varianza che considera i soggetti come indipendenti fra loro].

Disegni fra i soggetti: nei quali è misurato l'effetto di una variabile in diversi gruppi: riguarda i confronti fra i soggetti; l'effetto studiato è determinato da un fattore di gruppo.

Disegni misti: nei quali le condizioni fra i soggetti e entro i soggetti sono presenti: si utilizza il modello lineare generale perché permette di considerare entrambe le condizioni in un'unica procedura; inoltre ci dà un'informazione (descrittiva) separata dovuta all'effetto entro e fra singolarmente.

esempi di quanto detto fin d'ora:

es. di disegno entro i soggetti:

in una ricerca sono stati indagati in diversi soggetti appartenenti ad un campione unico, la capacità nella percezione del movimento (puntini luminosi in movimento) in un compito di difficoltà crescente. Se un soggetto si ferma, viene perso, quindi spss li elimina e bisogna tenerne conto.

Prove ripetute (es. di disegno fattoriale):

1) Per indagare le differenze fra ogni livello e individuare quali sono i livelli che producono differenze. Ci sono due modi:

- attraverso dei contrasti ortogonali (preferibili dal punto di vista metodologico); i contrasti sono diversi e possiamo scegliere quello che preferiamo, (ad es. confronta il livello 1 con la media di tutti gli altri).
- attraverso i confronti post hoc, che confrontano tutte le combinazioni possibili, ma ciò è sbagliato da un punto di vista metodologico, perché aumenta l'errore alfa; per evitarlo si effettuano delle correzioni.

2) Per vedere, attraverso un grafico, l'andamento medio ai diversi livelli.

Attraverso le statistiche descrittive trovo il senso delle differenze che ho trovato attraverso il test statistico.

Se c'è un'oscillazione casuale molto elevata, (per cui non è in grado di darci delle misure precise), significa che il campione non è adeguato perché probabilmente abbiamo dei sottogruppi con delle prestazioni molto buone e altri sottogruppi con prestazioni pessime.

Es. di disegno fra soggetti:

Ricerca per vedere se c'è una relazione tra depressione della madre e punteggio ottenuto dai padri al questionario che misura la depressione.

Il punteggio va da 0 a 36 e quindi può essere considerato gaussiano, per cui può essere analizzato attraverso una semplice analisi della varianza, oppure visto che abbiamo 2 gruppi potremmo fare anche un t test.

Punteggio ottenuto dai padri: variabile dipendente.

Fattore di gruppo: avere o non avere la depressione nel gruppo delle madri

Facciamo le statistiche descrittive, che ci danno un senso di quello che andiamo ad indagare.

Vediamo che nell'analisi between abbiamo la possibilità di valutare la differenza fra i 2 gruppi che andiamo a considerare e in questo caso ci viene dato anche il valore dell'intercetta, ovvero si considera l'analisi della varianza sempre all'interno del modello lineare generale, perché si parte dal presupposto che esista un valore di base indipendente dal fatto di avere o non avere la depressione.

Su questo valore di base viene poi indagato l'effetto dovuto alla depressione.

[l'intercetta è il valore di base di partenza].

Se il risultato non è significativo, vado a vedere (attraverso le statistiche descrittive):

- se è perché la differenza è troppo piccola.

- se la deviazione standard (lo standard error) è troppo alto.
- la numerosità.

es. di disegno misti:

la ricerca precedente sulla capacità nella percezione del movimento (puntini luminosi in movimento) in un compito di difficoltà crescente, in questo caso misura 2 gruppi diversi: soggetti dislessici e soggetti di controllo: mi aspetto che la percezione del movimento in questi 2 gruppi sia diversa e che questa differenza vari a seconda del livello di difficoltà. Si utilizza ancora una volta il modello lineare generale. In questi disegni abbiamo la possibilità di indagare sia gli effetti principali, sia gli effetti dovuti all'interazione che mi indica se c'è un andamento diverso, nei diversi gruppi, nei diversi livelli di difficoltà. Quindi ad es. se i dislessici si differenziano dagli altri soggetti nei diversi livelli di difficoltà. Nel test multivariato abbiamo la significatività per la ripetizione e per l'effetto interazione. L'effetto between viene separato da questi due e viene analizzato da solo.

Risultano tutti significativi, ciò significa che:

c'è una differenza dovuta al gruppo.

c'è una differenza dovuta alla ripetizione (quindi alla difficoltà dei diversi livelli del compito); questo effetto viene calcolato aggregando i punteggi dei 2 gruppi (facendo le medie).

Ancora una volta le statistiche descrittive tornano utili per vedere il senso di quello che ho analizzato.

D. incompleti:

si ricorre ai disegni incompleti :

- quando non è possibile utilizzare tutte le combinazioni dei fattori (ma solo alcune)
- oppure quando tutte le possibili combinazioni dei fattori porterebbero all'utilizzo di un numero di sottogruppi molto elevato.

Perché sono da evitare tutte le situazioni in cui abbiamo un numero di sottogruppi molto elevato? Prima di tutto perché l'interpretazione dei risultati sarebbe molto difficile e, essendo troppo particolareggiate, ci farebbero perdere quel senso di generalità al quale invece aspira la ricerca. In secondo luogo perché dovremmo avere per ogni sottogruppo delle numerosità molto alte per ottenere dei livelli di significatività buoni, e quindi dovremmo raccogliere un campione enorme.

Le combinazioni di fattori prese in considerazione sono quelle che permettono lo studio degli effetti principali.

Esempi di disegni incompleti sono quelli basati sul:

QUADRATO LATINO: Utilizza un criterio di incompletezza ciclica (x cui toglie di volta in volta un elemento della combinazione).

A BLOCCHI RANDOMIZZATI: Vengono tolte in modo random alcune sottoclassi di un certo fattore, in modo tale però da poter indagare gli effetti principali. (per poter avere una interpretazione complessiva adeguata).

Es. se vogliamo indagare l'effetto dovuto dall'università di appartenenza e dal tipo di facoltà sulla media dei voti di laurea, dobbiamo raccogliere un numero sufficiente di università e a questo gruppo associare il sottogruppo delle facoltà e poi andare a misurare qual è la media dei voti di laurea. Non tutte le facoltà sono rappresentate in ogni città e quindi si farà in modo di analizzare quella possibile combinazione che comunque permette di studiare gli effetti principali. Togliendo di volta in volta una delle facoltà dalle città, ho la possibilità di indagare tutti gli effetti principali. Quindi ho la possibilità di studiare sia l'effetto dovuto alla città, sia quello dovuto alla facoltà, nonostante manchino alcune combinazioni.

D. personalizzati:

Importante: Sia nel caso in cui ci troviamo in un disegno fattoriale, sia in un disegno incompleto, possiamo effettuare la scelta di personalizzare il disegno stesso; perché il disegno sperimentale è quello che è condizionato dall'ipotesi di partenza e perché nell'ipotesi di partenza possono essere interessanti solo alcuni degli effetti e non altri.

In spss se non specifico niente il modello analizzato è un modello fattoriale completo; se invece lo personalizzo invece significa che tra tutte le combinazioni possibili sono interessata solo ad alcune (ad es. potrebbe non interessarmi l'effetto dato dal livello di difficoltà ma solo quello dovuto al gruppo tra dislessici e controllo) e devo specificarlo nell'analisi statistica utilizzata.

CAMPIONAMENTO

Il campione è lo strumento che ci permette di dimostrare l'idea sperimentale che si riferisce a quella popolazione generale perché ne è parte, ovvero dovrebbe essere rappresentativo dell'insieme universo della popolazione che andiamo ad analizzare attraverso la nostra idea.

E' determinato dall'ipotesi di ricerca e dalle variabili in gioco ed è molto importante quando presentiamo una nostra ricerca indicare quali sono i criteri di inclusione o di esclusione, oltre alla tecnica di campionamento che abbiamo utilizzato, perché questo giustifica l'analisi successiva, quindi i risultati che abbiamo ottenuto.

Popolazione di riferimento: insieme dei casi potenzialmente reperibili (universo) rispetto alla variabile che andiamo ad indagare.

Il campione deve essere:

Adeguato rispetto all'ipotesi di partenza e rispetto alla popolazione.

Rappresentativo rispetto alla popolazione di riferimento

Affidabile rispetto alle tecniche di campionamento utilizzate. (permette la replicabilità dei risultati in un altro campione raccolto con le stesse condizioni; questo perché è importante controllare la distorsione dovuta al campionamento).

La distorsione influenza i risultati della ricerca, non è possibile controllarla dal disegno sperimentale, ma è inevitabile ed è determinata dalla tecnica di campionamento utilizzata; vediamo come:

Paradosso centrale del campionamento (Alan Stuart)

C'è l'impossibilità di controllare e quindi di conoscere dall'esame del campione se questo è privo di distorsioni.

Si parla di **distorsione da selezione** perché si considerano campioni adeguati solo se selezionati in modo da non essere influenzati da delle variabili confondenti (che cioè non rientrano nel disegno sperimentale). L'unico modo per eliminare la distorsione da selezione è quello di selezionare il campione in modo casuale. La casualità mi permette di ottenere la replicabilità dei risultati, cioè risultati simili, uguali, a quelli ottenuti con il campione che abbiamo selezionato.

Riassumendo, il campione viene definito casuale perché privo di distorsione da selezione; ed è l'unico tipo attraverso il quale è possibile fare inferenza.

TECNICHE DI CAMPIONAMENTO

Ci sono diverse tecniche di campionamento, non esiste una tecnica migliore dell'altra, l'importante è specificare i criteri e le motivazioni per cui si è scelta quella tecnica piuttosto che un'altra.

- **CAMPIONAMENTO CASUALE SEMPLICE:**
(è la base di partenza per tutti gli altri tipi di campionamento)

permette di selezionare un campione dalla popolazione in modo da avere la stessa probabilità per ogni possibile campione studiato. Quindi ogni soggetto ha la stessa probabilità di appartenere al campione, e questo determina la casualità.

- **CAMPIONAMENTO STRATIFICATO**

Suddivide il campione in strati, ovvero dei sottogruppi della popolazione, questi sottogruppi sono creati in modo tale da rappresentare la popolazione di riferimento.

es. se voglio rappresentare la popolazione italiana, dovrò mantenere le proporzioni dei sottogruppi.

Affinché questi strati siano formati in modo casuale occorre procedere però con un campionamento casuale semplice.

Il campione è molto **preciso** rispetto alla popolazione perché cerca di delineare tutte le caratteristiche importanti rispetto alla popolazione. (Il criterio perseguito è la precisione).

Il punto debole è che la corretta stratificazione dipende molto dalle conoscenze del fenomeno che si sta studiando, da parte dello sperimentatore.

(il campionamento avviene a livello della popolazione)

- **CAMPIONAMENTO A GRAPPOLI**

La popolazione è suddivisa in gruppi, per aumentare la casualità di scelta, dopodiché di questi gruppi, viene selezionato casualmente uno.

I gruppi devono essere omogenei fra loro, ovvero uguali fra loro e che hanno la stessa probabilità di essere scelti.

(il campionamento, avviene a livello di unità di popolazione, quindi in un grappolo sono presenti le caratteristiche di quella unità di popolazione; questo comporta inevitabilmente una perdita di precisione, e un aumento all'interno dei grappoli della variabilità).

Un ulteriore passo del campionamento a grappoli è il:

- **CAMPIONAMENTO A STADI**

Dopo aver scelto due o più grappoli si seleziona all'interno di esso un individuo (sempre attraverso un campionamento casuale semplice); dopodiché si può andare avanti suddividendo ancora i grappoli in sottogruppi, a seconda delle esigenze dello sperimentatore, però si perde di precisione, perché l'individuo ha caratteristiche sue proprie che non sono più rappresentative della popolazione generale.

NUMEROSITA' CAMPIONARIA

Il campione oltre ad essere adeguato e rappresentativo deve essere anche di numerosità idonea. La numerosità va determinata a priori, attraverso il calcolo della numerosità adeguata per dimostrare quella minima differenza apprezzabile con quel grado di errore stimato arbitrariamente.

La numerosità dipende:

- dalla distanza tra H_0 e H_1 (+ è ALTA E + è NECESSARIO AVERE UN CAMPIONE NUMEROSO)
- dalla potenza del test (“ “ “ “)
- dalla minima differenza apprezzabile (+ piccola è la differenza che voglio andare a misurare e + il campione deve essere numeroso)
- dalla varianza casuale (+ alta è la varianza casuale e + alta è la numerosità che mi serve per dimostrare la mia idea)
- da alpha (+ basso è il valore di alfa e + alta è la numerosità che devo andare a raccogliere).

DIFFERENZE FRA LE TECNICHE DI CAMPIONAMENTO

- Quando la popolazione è omogenea rispetto alla variabile in studio è preferibile utilizzare un campionamento casuale semplice.
- Quando la popolazione non è omogenea, è preferibile un campionamento stratificato.

- Per avere più casualità nell'estrazione di un campione da una popolazione omogenea, si può utilizzare una tecnica di campionamento a grappoli o a stadi.