

Lezione Benassi 6 marzo 2007

Campionamento: domande sulla piattaforma A3.

Esame: o analisi di ricerca scientifica, o alcune domande tipo quelle di A3.

Fallacia nella stima campionaria e nel test statistico

La prima caratteristica del ragionamento fallace è l'**ambiguità**, ovvero si pensa all'oggetto stesso del ragionamento come ad una cosa diversa, 'un x che sembra y ma è z', cioè si pensa all'oggetto del discorso e lo si classifica in modo ambiguo e quindi sbagliato. Non solo l'oggetto del discorso sembra qualcosa di diverso da quello che è, ma viene anche classificato come qualcosa di sbagliato. Le due caratteristiche principali del ragionamento fallace sono:

- ★ il fatto di essere ambiguo;
- ★ il fatto di ricadere in errore.

Il **campionamento** è l'unico modo che ci permette di fare inferenza, inoltre il campione pur riferendosi alla popolazione generale che vuole rappresentare la nostra idea sperimentale porta con sé un **errore campionario**, quindi qualcosa di sbagliato che lo differenzia rispetto alla nostra popolazione generale e a quella teorica di riferimento. L'unico modo per combatterlo è stimarlo, in diversi modi. La statistica e la metodologia si pongono come obiettivo quello di misurare questo errore.

Ci sono diversi tipi di errore. L'errore di campionamento ha particolari caratteristiche; si ovvia ad esso utilizzando una metodologia corretta. L'unico tipo di tecniche di campionamento che ci permette di stimare l'errore è il **campionamento probabilistico**. Il campione dev'essere rappresentativo rispetto alla popolazione, ovvero deve avere proprie caratteristiche di composizione che rispondono a quelle della popolazione rappresentata. Non solo per caratteristiche, ma anche per numerosità: il campione dev'essere sufficientemente numeroso per verificare la nostra idea. E' anche vero però che con campioni molto numerosi è possibile verificare idee che non sarebbero scientificamente interessanti. Il punto chiave per capire come va compresa la numerosità sta proprio nella metodologia utilizzata.

Es. Ricerche scientifiche su cose che non interessano, ma sono metodologicamente corrette.

La bontà di un'idea dunque non è data solo dal risultato significativo, ma dalla discussione etica e politica, dalle implicazioni che questi risultati hanno in sé.

Nelle **tecniche di campionamento** sono perseguibili due punti:

- ★ il campione dev'essere casuale e rappresentativo;
- ★ il campione dev'essere preciso.

La casualità completa, dal punto di vista teorico, non è possibile. Il campione non può mai essere privo di distorsioni: ha caratteristiche proprie rispetto alla popolazione generale, dovute al fatto che io ho scelto proprio *quel* campione.

Casualità e precisioni sono perseguibili dal punto di vista teorico, è ovvio che il ricercatore non va veramente a cercare queste caratteristiche, piuttosto va, ad esempio, in un ospedale e fa riferimento a quella struttura. Quindi il campione porta con sé caratteristiche che sono proprie della campionatura. L'importante è che il ricercatore nella ricerca definisca da dove proviene quel campione e quali sono state le metodologie della campionatura.

Piano di campionamento: modalità di selezione e parametri/stimatori da misurare. La misura effettuata sul campione si rapporta alla misura dell'errore che quel campione si porta dietro, attraverso la tecnica statistica applicata. Più io cerco di essere preciso rispetto alle caratteristiche del mio campione, più il mio campione perde di casualità. Il campionamento ottimale si fa rispetto all'ipotesi di partenza e al disegno sperimentale a cui si riferisce il ricercatore e gli errori (e quindi

l'efficienza della ricerca) vengono stimati attraverso la statistica inferenziale. La statistica descrittiva non permette di misurare l'errore, cioè l'effetto che stiamo studiando in rapporto all'errore.

Omogeneità e numerosità campionaria: per essere rappresentativo, un campione dev'essere il più possibile simile alle caratteristiche della popolazione generale; più la numerosità del campione si avvicina a quella della popolazione, più l'errore campionario diminuisce. Minore è la variabilità della stima campionaria e minore è l'affidabilità del campione che utilizziamo.

Esempi di fallacie collegate a distorsioni nel campionamento

Avvengono perché sono stati messi in discussione i principi di: casualità, idoneità e numerosità.

★ Se il campione non è **casuale**, io posso attribuire i risultati riferiti a quel campione, e quindi alla popolazione generale di riferimento, quando questo campione in realtà non la rappresenta. Se non è casuale non può essere considerato come rappresentativo, perché le variabili utilizzate per quel campione possono essere diverse rispetto alle variabili studiate, e quindi ci sfugge l'errore dovuto al campionamento, che invece noi crediamo di misurare attraverso la statistica. Esempio: applicazione di diversi test statistici sullo stesso campione, per dimostrare diverse cose rispetto alla stessa ipotesi, come applicare più t-test quando serve un'ANOVA (es. Differenze tra 3 gruppi indipendenti rispetto alla misura del QI), o quando applichiamo test univariati su esperimenti multivariati, come misurare due scale misuranti un unico costrutto, e che si pensa siano correlate tra loro. L'analisi opportuna consiste nell'inserire entrambe le scale come campioni indipendenti e applicare un test multivariato, avendo dunque un unico risultato con valore dimostrativo.

★ Quando le variabili misurate sono troppo **eterogenee** si inficia il principio di idoneità; si pensa di aver raccolto un campione rappresentativo, ma in realtà se ne è misurato uno con sottocampioni, ovvero caratteristiche peculiari proprie (es. Analisi della regressione su gruppi non omogenei, dovrebbe mostrare una retta, ma in realtà si identificano due gruppi ben separati. Il coefficiente R² è altissimo e la relazione è significativa dal punto di vista dell'analisi statistica, ma non tiene conto della diversità tra le due subpopolazioni. La retta viene trovata e il risultato è significativo, ma sbagliato; la retta NON è buona, i risultati non sono disposti su una retta!). Dunque il campione non è idoneo quando è troppo eterogeneo; si controlla verificando con le statistiche descrittive come sono disposti i nostri casi, qual è l'oscillazione casuale che abbiamo rispetto ai dati che misuriamo. Esempio: il paradosso di Simpson. Si vede come io posso arrivare a risultati fallaci non tenendo presente le cosiddette 'variabili confondenti'. Se vogliamo confrontare due terapie, A e B, rispetto ai miglioramenti su un certo caso patologico, possiamo avere questo paradosso: se consideriamo da prima i due gruppi rispetto alla percentuale di miglioramenti, possiamo trovarci nel caso in cui la terapia A è peggiore di B perché ha ottenuto il 10% in meno di miglioramenti. Se invece consideriamo due sottogruppi, di maschi e femmine, vediamo che il risultato di prima si ribalta a favore della terapia A. Infatti, considerando il fattore dovuto alla differenza di genere, la terapia A ha una percentuale di miglioramenti maggiore rispetto alla terapia B. Se avessimo considerato il primo caso, avremmo commesso una fallacia perché non avremmo considerato una variabile determinante nei miglioramenti, ovvero la variabile 'genere'.

★ Quando non trovo una differenza potrei concludere che essa non esiste senza ragionare sull'adeguatezza della **numerosità** del mio campione; è quindi sempre opportuno per non incappare in questo errore calcolare prima la numerosità sufficiente per dimostrare la nostra idea.

★ Spesso si commette un altro tipo di fallacia: dimostrare relazioni **causa-effetto** tra le variabili analizzate. Anche se il test statistico è costruito considerando una variabile come fattore, che si pensa modifichi l'altra variabile che si sta studiando, e quindi il test è costruito sul rapporto causa-effetto delle variabili, non si può concludere nei risultati che una variabile è causa di un'altra, anche se il risultato è significativo esso è di tipo relazionale, non causale.

Validità nella ricerca

Per essere valida, una ricerca sperimentale dev'essere falsificabile, generalizzabile e replicabile, e deve aver utilizzato metodi statistici adeguati. La validità di una ricerca si diversifica rispetto a quella di un test o di un questionario; il processo di validazione di questi ultimi permette di stabilire che quel questionario misura ciò che si vuole misurare.

Esistono diversi tipi di validità: la validità di una ricerca più importante è la **variabilità interna**. Si parla di essa rispetto alle variabili in studio, quando c'è una chiara definizione tra le variabili analizzate, e quando sono escluse le variabili confondenti. Ci sono diversi tipi di minaccia alla validità interna:

1. **Effetto regressione**: quando ripetiamo delle misure sullo stesso campione i risultati tendono a regredire verso la media, cioè col ripetersi della misura, lo stesso soggetto ha un punteggio sempre più vicino alla media, e scompaiono i punteggi estremi. Può minare la validità interna se non si tiene conto del fatto che le misure effettuate sono misure ripetute.
2. **Effetto apprendimento**: come sopra; se io somministro un compito in momenti successivi, chiaramente di quello stesso compito dovrò misurare anche l'effetto dovuto al fatto che il soggetto pian piano impara il compito. Anche in questo caso si può ovviare attraverso un test a prove ripetute, considerando non più le prove come indipendenti ma come correlate.
3. **Maturazione degli individui**: è possibile che io debba chiamare gli stessi soggetti in momenti diversi della loro maturità, e quindi che rispetto alle prime prove ci siano dei risultati dovuti alla maturazione degli individui.

In questi casi è opportuno inserire nel disegno sperimentale l'effetto dovuto alla ripetizione delle prove; ad esempio attraverso l'analisi per prove ripetute.

Esempio: disegno sperimentale di Solomon per il controllo dell'effetto trattamento, apprendimento e maturazione: suddivisione in gruppi diversi, ciascuno dei quali con caratteristiche peculiari.

Le **variabili confondenti** sono distinguibili in due gruppi:

★ **Sistematiche**: agiscono in modo incontrollabile;

★ **Asistematiche**: caratteristiche di situazioni di quel disegno sperimentale in particolare. Esempio: perdere dei soggetti durante la ricerca. Se la perdita è definitiva si parla di 'mortalità' dei soggetti; se invece è possibile recuperare e rimisurare i soggetti si parla di 'drop out'. In entrambi i casi è preferibile analizzare l'effetto dovuto a questa perdita. Ovvero: perché questi soggetti sono andati persi? Esiste un tipo di analisi che è sempre più richiesta, cioè 'intention-to-treatment-analysis' che permette di vedere l'effetto generato da questi soggetti e di compararlo a quello degli altri soggetti, confrontando i gruppi di chi ha terminato la ricerca e di chi non l'ha terminata. Questa analisi funziona 'alla rovescia' rispetto al solito, nel senso che io nella mia ricerca auspico che non ci siano differenze tra i due gruppi, perché altrimenti vorrebbe dire che non ho analizzato una variabile (la mortalità) che invece avrei dovuto analizzare.

E' importante che sia chiara la distinzione tra il ruolo delle diverse variabili nel disegno di ricerca, e che queste relazioni siano il meno possibile ambigue. Questa distinzione fatta tra relazioni genuine e spurie non è importante ai fini della validità della ricerca stessa, ma dal punto di vista teorico.

Validità esterna: oltre all'esplicitazione della relazione tra le variabili in studio, c'è la necessità che la mia ricerca produca dei risultati generalizzabili. La validità esterna si riferisce alla generalizzabilità dei risultati, dovuta soprattutto alla numerosità del campione (rappresentativo e idoneo).

Validità di costrutto: tipo di validità più labile; è più arbitrario il modo in cui viene definita. Si riferisce alla corrispondenza tra la ricerca e il costrutto che si vuole analizzare, quindi più difficilmente c'è la possibilità di trovare questa corrispondenza e ci si affida alla conoscenza degli esperti, che è fortemente arbitraria. Tra le validità descritte è la più difficile da discutere. Anche nei costrutti possono esserci costrutti confondenti; è lo sperimentatore che nel disegno della

ricerca, in riferimento alla validità interna, definisce in modo più o meno adeguato il costrutto di riferimento. Dunque tra le altre, la validità di costrutto è la meno importante.

Validità statistica o predittiva: valuta la corrispondenza tra la relazione delle variabili che abbiamo indagato e il calcolo della significatività vero e proprio, quindi si riferisce alla verifica statistica delle ipotesi sperimentali, attraverso l'utilizzo di test statistici appropriati. La validità statistica va ad indagare la relazione tra una variabile dipendente e il fattore ad esso associata, quindi è ridondante rispetto alla validità interna.

Validità ecologica: indaga la corrispondenza tra gli studi effettuati in laboratorio e quelli che presuppongono che il comportamento in laboratorio abbia un riscontro nella realtà.

Insieme alle variabili confondenti, vengono considerati altri fattori di disturbo della validità: **gli artefatti**.

Essi sono fattori di disturbo legati al soggetto o allo sperimentatore, e come le variabili confondenti agiscono e determinano dei cambiamenti nei risultati. Esempi di artefatti dovuti al ricercatore (**effetto Rosenthal**) sono quelli in cui il ricercatore trasmette in modo inconsapevole le proprie aspettative rispetto a ciò che va a misurare, e inevitabilmente favorisce una risposta adeguata nel soggetto 'suggerendogli' la risposta giusta. Questo effetto è controllato utilizzando studi 'a doppio cieco', in cui ai soggetti viene misurato l'effetto di una variabile che si sta studiando da un ricercatore che è all'oscuro dell'effetto o della soluzione del compito che si va ad indagare.

Un altro effetto è '**l'effetto Pigmalione**', ovvero un effetto Rosenthal applicato all'educazione. Si riferisce al fatto che il ricercatore più ha aspettative nei confronti del soggetto, più il compito stesso del soggetto sarà effettuato in modo positivo.

Altri fattori sono dovuti alla desiderabilità sociale, quindi è sempre opportuno che la ricerca venga condotta da più figure alternative rispetto al ricercatore che formula l'ipotesi. Ciò che spesso succede però è che i dati raccolti da chi è estraneo alla ricerca siano raccolti in modo sbagliato, producendo altri errori.

Fattori legati al soggetto: è possibile classificare i soggetti in base alle loro aspettative e ai comportamenti nei confronti del compito:

★ **Fiduciosi:** comportamento genuino e non distorto;

★ **Buoni:** tendono a dare una risposta che credono adeguata rispetto a ciò che pensano si voglia verificare;

★ **Negativi:** tendono a disconfermare le ipotesi che pensano il ricercatore formuli;

★ **Apprensivi:** pensano di essere valutati dallo sperimentatore e quindi rispondono in modo da dare un'immagine positiva di sé.

E' quindi necessario riferirsi sempre a queste quattro categorie di soggetti.

Fattori dovuti alla situazione sperimentale: riguardano tutti quei comportamenti che il soggetto mette in atto in modo incontrollato a causa delle aspettative e delle modifiche che genera la situazione sperimentale e che incidono sui risultati del compito.