

## Lezione Benassi 27 febbraio

(inizia la lezione con un questionario sulle nostre abitudini informatiche, poi fa un ripasso della lezione precedente)

### Ripasso

Statistica descrittiva vs. Inferenziale, è importante sapere cosa si può fare con l'una e con l'altra (ovvero: descrizioni del campione preso in esame per la prima, parte da un campione preciso e identifica delle idee generali da dimostrare; la dimostrazione di esse avviene per mezzo della seconda). Le due statistiche sono complementari, nessuna delle due, da sola, è totalmente informativa, è inutile dimostrare un'idea se non abbiamo la descrizione del campione. La statistica descrittiva viene utilizzata anche per validare dei questionari psicometrici: in psicologia la maggior parte delle misure si fa attraverso questionari creati ad hoc, perché l'oggetto di misura è particolare, quindi è necessario creare un questionario per misurare ciò che vogliamo conoscere. I questionari non sono fatti per misurare qualsiasi oggetto, devono essere validati. Questo processo di validazione si fa tramite procedure statistiche descrittive, che si attuano su un campione molto ampio; si fa riferimento a dei test statistici, come l'analisi fattoriale, appropriati per quella funzione. Gli stessi risultati si riferiscono quindi solo al campione che abbiamo considerato.

L'altra differenza fondamentale è che mentre la statistica inferenziale è basata sul calcolo probabilistico, l'altra non lo è: i parametri della statistica descrittiva non sono correlati dalla loro probabilità. Invece ogni volta che dimostriamo un'idea per mezzo dell'inferenza abbiamo la probabilità della nostra dimostrazione. Questo è il punto cruciale, è fondamentale per l'esame sapere quali sono le proprietà del concetto di probabilità (es. Limiti) e che abbiamo una buona conoscenza del calcolo probabilistico. SE NON VI RICORDATE QUALCOSA CERCATE SUGLI APPUNTI DEL PRIMO ANNO!!

Un'altra cosa importante è lo schema del procedimento della statistica inferenziale, della dimostrazione dell'ipotesi sperimentale. L'inferenza è basata su un procedimento rigido, che viene effettuato seguendo passi successivi che hanno una successione temporale ben definita.

Primo passo: formulazione idea sperimentale; buon ricercatore ha chiaro in testa cosa vuole dimostrare, qual è la relazione causale tra le variabili. Es. Cattiva idea sperimentale: valutare l'andamento dell'anno scolastico in certi soggetti. Buona idea sperimentale è trovare le differenze tra le carriere degli studenti della specialistica a seconda delle modalità di accesso alla laurea stessa (se il fatto di aver fatto o meno la prova di ingresso determini differenze). Dev'essere un'idea generale, relativa ad una popolazione e non solo ad un campione considerato. Esso quindi dev'essere casuale (scelto a caso all'interno della popolazione generale) ma anche idoneo (rappresentativo rispetto alla popolazione di riferimento). Per dimostrare l'idea generale è necessario individuare  $H_0$ ;  $H_1$  è di differenza,  $H_0$  è di uguaglianza. Attraverso il test statistico, dopo aver raccolto i dati, è il confronto tra la distribuzione di probabilità che abbiamo ottenuto rispetto ai dati raccolti (campione misurato) e la distribuzione nulla corrispondente alla distribuzione di probabilità che sarebbe vera se  $H_0$  fosse vera. Questo si vede anche dal punto di vista grafico. Si confrontano le due distribuzioni di probabilità attraverso il test statistico:  $H_1$  e la curva sottesa ad  $H_1$  corrisponde alla distribuzione di probabilità di ipotesi sperimentale; io mi aspetto che, data la mia ipotesi sperimentale, troverò differenze tra i due gruppi. Rispetto al parametro misurato troverò differenze rispetto al parametro teorico che ci sarebbe se  $H_0$  fosse vera, cioè se non ci fossero differenze tra i due gruppi. Confronto quindi le due distribuzioni. Ci sono diversi tipi di errore. L'errore di misura e l'errore di decisione. Gli errori di misura sono dovuti allo strumento di misura e sono ineliminabili, si possono ridurre al max attraverso un controllo dello strumento; anche gli altri sono ineliminabili, ma è possibile controllarli attraverso la loro stima. L'errore alfa si riferisce all'errore che si compie quando si considera  $H_0$  falsa, pur essendo vera. Corrisponde al limite del 5% relativo alla distribuzione nulla, e viene fissato in modo arbitrario prima della raccolta dei dati; è il limite al di sotto del quale si dimostra  $H_1$ . Nel caso di una curva

gaussiana, errore alfa viene diviso in una parte a dx e una a sx della media, cioè 0, per 2.5% da una parte e 2.5% dall'altra. Quando confrontiamo le due curve determiniamo anche l'errore beta; esso viene fissato solitamente in modo arbitrario perché corrisponde alla parte di curva relativa ad  $H_0$  che invece sta dalla parte opposta ad alfa. Più io cerco di ridurre l'errore alfa, più inevitabilmente beta viene a crescere perché spostato il suo limite. L'errore beta è l'errore relativo alla non falsificazione di  $H_0$  quando  $H_0$  è da falsificare. Assieme ad alfa e beta, si definisce la potenza del test, data dall'inverso di beta, cioè la probabilità di non respingere  $H_0$  quando questa è da respingere (1-beta). La potenza del test è determinata da  $H_1$ . I parametri di cui sopra sono legati l'uno all'altro, e dipendono da diversi fattori: numerosità, minima differenza che si vuole misurare, varianza stimata sul nostro campione e dagli stessi errore alfa e potenza del test. Quindi questi parametri sono legati fra loro perché nel calcolo del parametro questi parametri rientrano nella formula ( fa l'esempio del calcolo della numerosità con il programma Primit, cambiamento dei parametri in funzione della numerosità - se aumento il numero dei soggetti nel campione e aumenta la varianza, cosa succede nella mia curva sperimentale  $H_1$ ? Cosa cambia nella potenza del test? Aumenta beta e quindi diminuisce la potenza del test, innalza la larghezza della curva, e più la curva è larga mantenendo le stesse medie, più errore beta aumenta. Se invece aumento la numerosità aumenta la potenza, perché diminuisce la varianza e la probabilità che mi permette di dimostrare la nostra idea -?- ). Prima di fare la ricerca il ricercatore deve già sapere quale differenza si deve aspettare, ad esempio tra due gruppi che vuole misurare. Es. La WISC, ciò che discrimina tra i soggetti con QI nella norma o no si aggira intorno a 15, cioè chi ha QI sopra il 85 sono nella norma, al di sotto No. Quindi ad esempio mi aspetto che i due gruppi differiscano per 15, che è la deviazione standard (anche questa dobbiamo pensarla a priori prima di raccogliere i dati), che si calcola come radice quadrata della varianza, si parla quindi sempre di variabilità, è il corrispettivo (continua con l'esempio di prima). Si considera buona potenza quella del 0.8, cioè 80%. Quando facciamo una ricerca inferenziale diventa importante stabilire prima la maggior parte dei parametri da andare a misurare dopo.

Se durante l'inferenza non riesco a falsificare  $H_0$  e quindi dimostrare  $H_1$  non significa che  $H_0$  sia vera. Non è mai possibile dimostrare l'uguaglianza. Il fatto di non respingere  $H_0$  può essere dovuto a diversi fattori: il primo in ordine di importanza è che il campione ha una varianza molto grande, poi può essere che il campione abbia una numerosità ridotta, troppo piccola, oppure il campione che ho misurato non soddisfa le condizioni relative alla distribuzione, ad esempio non ho visto che il mio campione era distribuito diversamente rispetto al test statistico che ho applicato, oppure il campione non è rappresentativo, o non c'è sufficiente separazione tra  $H_0$  e  $H_1$ , le differenze che cercavo erano talmente piccole da non essere significative. Se non riusciamo a respingere  $H_0$  dobbiamo riflettere su tutte queste cose. In teoria su queste cose bisognerebbe riflettere PRIMA di raccogliere i dati, calcolando prima la potenza del test e la numerosità ideale del campione (quella minima che mi permette di falsificare  $H_0$  se  $H_0$  è falsa). Potenza del test e numerosità andrebbero stabilite prima di raccogliere i dati ed effettuare il test statistico; non sempre questo avviene, a volte proprio perché non ci sono le possibilità concrete. Quando non avviene, se io non riesco a falsificare  $H_0$  devo riflettere su queste cose. E' possibile far calcolare ai software statistici sia la potenza che la possibilità di falsificare  $H_0$ . Solo nel caso in cui la potenza del test è molto alta e il campione è molto numeroso e noi non abbiamo falsificato  $H_0$ , allora possiamo pensare che la differenza che cercavamo era troppo piccola, solo in quel caso dobbiamo pensarlo. La logica falsificazionista è tutta basata su falsificazione di  $H_0$ , tutte le ricerche che non riescono a falsificare  $H_0$  non vengono condivise nella comunità scientifica e vanno perse.

### **Vediamo un articolo da analizzare insieme.**

-> possibile esercizio proposto all'esame, in italiano!

Strutturazione articolo: introduzione con ipotesi sperimentale e risultati ottenuti fino a quel momento, poi presentazione dei metodi di ricerca utilizzati nel campione e la descrizione dei risultati. Questa struttura degli articoli è molto rigida, così come è rigido il processo di

dimostrazione.

In questo caso, si analizzano degli operatori sanitari; in particolare, vengono valutati tramite un indice di capacità di lavoro, misurato attraverso uno strumento opportuno, che si chiama ?.

**Analizziamo qual è l'ipotesi sperimentale, se è espressa bene o no, quali sono le variabili che prende in considerazione, se sono identificabili, se fanno riferimento ad un disegno sperimentale multivariato o univariato, quali sono le var dip e i fattori, in che caso è possibile rappresentare queste var e quali statistiche vengono applicate (test param o non param, se c'è corrispondenza tra il test usato e la var presa in considerazione e se è possibile identificare una distribuzione di probabilità).**

[legge velocemente l'articolo, vedi fotocopia se l'ha distribuita, o ascolta registrazione a 50 minuti]  
Utilizzo strumento: indice di capacità di lavoro;

Ipotesi da verificare: esiste una relazione tra età e indice di capacità di lavoro;

L'ipotesi è espressa in modo chiaro? No. Oltre all'età, vengono prese in considerazione altre variabili che nell'introduzione non vengono considerate; questo fattore cambia l'indice di capacità di lavoro, e sono coinvolti anche altri fattori. E' una ricerca fatta senza avere chiara l'idea sperimentale.

E' un'ipotesi generale? Che popolazione riguarda? Gli operatori sanitari, riguarda la popolazione generale che lavora in ambito sanitario, quindi in teoria è un'ipotesi generale, anche se non espressa in modo esplicito.

E' possibile considerare il campione utilizzato rappresentativo? Corrisponde all'ipotesi generale un campione adeguato, casuale e idoneo? Il campione non è adeguato, si riferisce solo ad una regione d'Italia, non a tutta la popolazione generale del comparto ospedaliero. L'idea non è chiara.

Materiali e metodi: informazioni su stratificazioni del campione, si delinea una particolare popolazione; la popolazione generale in realtà corrisponde ad una popolazione particolare, in cui infermieri professionali e medici sono una maggioranza dei soggetti, non si parla più di operatori sanitari in generale. In teoria si dovrebbe avere la stessa percentuale di ogni tipo di professione rappresentata. In questo caso il campione è stato raccolto in base alla disponibilità di soggetti, quindi la maggior parte dei risultati ottenuti riguardano infermieri professionali e medici, che insieme costituiscono il 73% della popolazione. Già dalla popolazione selezionata ci facciamo un'idea dei risultati ottenuti. Critica: no distribuzione equa nella popolazione rispetto alle professioni rappresentate.

Punteggio finale del questionario: da 7 a 49, indice delle possibili distribuzioni di probabilità, e quindi del test statistico da applicare. Punteggio è quantitativo, viene poi trasformato in qualitativo (fasce), si formano 4 gruppi di soggetti in base alla capacità di lavoro dei soggetti. Quindi che analisi si usa? Correlazione tra punteggio ed età è inferenziale, ma la statistica descrittiva si usa come complemento, quindi usiamo entrambe le statistiche. Lo scopo è vedere se tra età e l'altro indice c'è correlazione, quindi prima inferenziale. Abbiamo divisione per età e anche per punteggio, utilizzabile in modo quantitativo ma anche qualitativo. Fattori: età, sex, tipo di lavoro. Var dipendente: se è qualitativa usiamo test non parametrico, se è quantitativa va da 7 a 49, è possibile usare test parametrico con questo punteggio? Non basta che la variabile sia quantità per applicare test parametrico, occorre che la distribuzione sia gaussiana, in questo caso min 7 e max 49, mi aspetto che intorno alla metà si concentri la maggior parte dei casi e che questa variabile sia riconducibile ad una distribuzione gaussiana, quindi è accettabile applicare su di essa test parametrici. Se invece andasse da 7 a 15, più difficilmente avrebbe questa possibilità e sarebbe meno accettabile applicare ad un questionario che ha una scarsa oscillazione (attorno a 15-20 punti per ogni valore possibile) un test parametrico, a meno che assieme al punteggio scarso ci sia la distribuzione effettiva del campione. In questo caso il campione è molto ampio e la nostra curva molto più probabilmente si avvicinerà alla distribuzione gaussiana: all'aumentare dei casi tutte tendono alla distribuzione gaussiana.

Nei risultati si coglie che si diversificano i gruppi in base al sesso di appartenenza e vengono calcolate le differenze tra M e F rispetto agli indici qualitativi ottenuti. Ci sono differenze? Si

applica X<sup>2</sup>, applicato sulla differenza di percentuali. Trovano due differenze e applicano il t-test. Poi altra analisi per gruppo delle donne per identificare se all'interno dei vari gruppi di età ci fossero delle possibilità di ottenere degli indici diversi. Non viene riportata però nessuna significatività, ma solo descrizioni. Poi si analizza l'indice di età, e non viene presentata la probabilità di errore. E' stata analizzata in termini qualitativi o quantitativi? Per fasce d'età. Cioè, loro per ogni fascia d'età hanno effettuato un test statistico, divisi comunque per professione e sesso. L'età è qualitativa, il test applicato è ANOVA, ne hanno applicata una per ogni fascia d'età e quindi hanno ottenuto un valore di p per ognuna, dove aveva abbastanza persone. Errore grave! Hanno applicato più test sullo stesso campione, mentre era possibile applicare un'unica procedura statistica, anche se poco conosciuta.

Quindi in questo caso è possibile rilevare un test statistico appropriato perché è parametrico per quelle variabili, ma non è appropriato perché utilizza più test statistici quando era possibile utilizzarne uno solo, aumentando così l'errore alfa e rendendo più incomprensibili i risultati. Non c'è un risultato unico, ma solo alcune cose che riguardano popolazioni suddivise arbitrariamente.