- Copyright - Il Pensiero Scientifico Editore downloaded by IP 216.73.216.118 Fri, 04 Jul 2025, 01:42:51

Scale di misura e test statistici

Riassunto

La scelta del test statistico è discussa in relazione alla scala di misurazione (scale nominali, scale ordinali, scale di rapporti) con cui sono stati ottenuti i dati.

Parole chiave. *Test parametrici, scala di Likert, mediana, scale nominali, scale ordinali, scale di rapporti.*

Summary

Measurement scales and the choice of the statistical test

The choice of the statistical test is discussed in relationship with the type of measurement scale that generated the data (nominal scales, ordinal scales, ratio scales).

Key words. *Parametric tests, Likert's scale, median, nominal scales, ordinal scales, ratio scales.*

Molteplici sono i tipi di dati che si raccolgono nel corso di studi clinici. Ci sono dati quantitativi e dati qualitativi; tra i primi ci sono misure, ma anche conteggi, tra i secondi dati non ordinabili e dati ordinabili. Se è vero che talvolta non c'è un solo strumento statistico per analizzare i dati di un certo tipo, è altrettanto vero che taluni strumenti, utili per l'analisi di un certo tipo di dati, sono inappropriati per dati di altro tipo.

Cogliendo l'occasione fornita dal lavoro discusso nella rubrica "Casi clinici", in cui correttamente si osserva che dati ottenuti con l'impiego di scale psicometriche non possono essere considerati come provenienti da variabili normalmente distribuite, nella presente nota, si amplia il discorso alla scelta del test statistico in relazione alla natura dei dati. In tal modo si introduce l'argomento dei test non parametrici (che verranno trattati nel prossimo numero di CASCO), così importanti nell'analisi dei dati relativi agli studi clinici nel campo delle terapie di supporto.

Il test t di Student per il confronto tra due medie nel caso di campioni indipendenti – assai usato nella ricerca medica – poggia su due assunzioni:

a. nelle popolazioni target (cui vanno riferiti i risultati di uno studio condotto su campioni), il carattere oggetto di studio si distribuisca normalmente (cioè secondo la curva di Gauss);
b. la variabilità (misurata dalla varianza) delle due popolazioni sia la stessa.

La piena validità di tale test si raggiunge quando entrambi questi assunti sono rispettati. Tuttavia, si è dimostrato che anche moderate (ma non forti) violazioni di tali assunzioni non inficiano molto il risultato ottenuto, perché il test t di Student appartiene alla famiglia delle cosiddette statistiche "robuste".

Esempi. Dovendo confrontare i valori medi di una caratteristica antropometrica (ad es., la statura) o ematobiochimica (ad es., la glicemia) in soggetti normali, si può ammettere che le due assunzioni siano rispettate o al più moderatamente violate perché ciò è già noto da studi di biometria. Ma dovendo confrontare i redditi medi di due gruppi si può essere altrettanto certi che la violazione almeno del primo assunto sia forte perché il reddito si distribuisce in modo esponenziale (sono molti i soggetti poveri e pochi quelli ricchi).

Esistono test statistici volti a verificare se tali assunzioni siano rispettate o meno, ma, inspiegabilmente, sono raramente usati.

Un test che richieda una o più assunzioni sulla popolazione prende il nome di **test parametrico**.
Ad esempio, sono parametrici il test t di Student nelle sue varie applicazioni (test sul valore di una media; per il confronto tra due medie, nel caso di campioni indipendenti e nel caso di dati appaiati), il test F di Fisher-Snedecor per l'analisi della varianza, con tutti i test per il confronto tra gruppi che ne conseguono.

I dati oltre che quantitativi, possono essere anche qualitativi, sconnessi o ordinabili.

Tutti i caratteri qualitativi si manifestano nei singoli soggetti mediante un attributo (o aggettivo). Un carattere sconnesso ha modalità[1] non ordinabili. Il sesso (modalità: M e F) e il tipo di neoplasia sono esempi di caratteri sconnessi. Nella ricerca psicometrica (e in quella sociale), i caratteri sconnessi prendono il nome di scale nominali, dove "scala" esprime il desiderio di misurazione e "nominale" indica che si tratta di una scala che in realtà lo è solo di nome. Ovviamente, nel caso di un carattere sconnesso non è possibile calcolare alcuna media (tranne la moda[2]): l'analisi dei dati, dunque, è incentrata sulle frequenze.

I caratteri qualitativi ordinabili sono caratterizzati dall'avere modalità ordinabili, nel senso che date due modalità diverse, si può sempre individuare quella che viene prima e quella che viene dopo (cioè la maggiore e la minore). Come per i caratteri sconnessi, quelli ordinabili si manifestano nelle unità della

1. Si chiamano modalità i modi diversi di presentarsi del carattere nelle unità della popolazione.
2. Come nel linguaggio comune, la moda è la modalità che presenta la massima frequenza.

- Copyright - Il Pensiero Scientifico Editore downloaded by IP 216.73.216.118 Fri, 04 Jul 2025, 01:42:51

popolazione per mezzo di attributi che però, a differenza dei primi, sono ordinabili. Nella ricerca psicometrica i caratteri ordinabili prendono il nome di "scale ordinali". Oltre la moda (come nelle scale nominali), nelle scale ordinali è possibile calcolare anche la mediana[3]. Ovviamente non può essere calcolata la media aritmetica perché non possono essere sommati attributi. Lo stadio della neoplasia, il performance status e tutte le risposte valutate su una scala di Likert[4] sono esempi di scale ordinali. Le scale di Likert, assai usate per rilevare le risposte ai singoli item in un questionario sulla qualità di vita, apparentemente sono quantitative perché la scelta, ad es., è limitata a 9 punti (da 1 a 9), ma in realtà tali numeri sono modalità qualitative ordinabili, ma non modalità quantitative, perché la distanza tra due punti consecutivi non può mai essere considerata la stessa.
Ad es., in una scala a 5 punti (1 = malissimo, 2 = male, 3 = discretamente, 4 = bene, 5 = benissimo), la distanza che c'è tra 1 e 2 non è la stessa che c'è tra 4 e 5. Quindi, volendo sintetizzare le informazioni ottenute con una scala di Likert, sarebbe opportuno usare la mediana e non la media (come giustamente precisano gli autori del lavoro discusso in "Casi clinici"). Inoltre, una scala di Likert va, ad es., da 1 a 9 e ciò mal si concilia con l'assunzione di normalità, perché la variabile normale varia con continuità (cioè può assumere valori intermedi tra 1 e 2, tra 2 e 3, tra 3 e 4, e così via), ed è inoltre illimitata a sinistra e a destra (mentre la scala di Likert assume solo valori interi compresi tra 1 e 9).

L'ampia diffusione delle scale ordinali in Medicina è dovuta al fatto che in molti casi la variabile da misurare è quantitativa, ma non si riesce ad ottenere per essa misure precise. Allora ci si contenta di dividere il suo dominio in classi (poche) in base a cui è semplice classificare i pazienti.

Esempi.

1. Lo stadio di una neoplasia: è possibile pensare ad un *continuum* di valori di progressione della malattia (dalla mutazione di alcune cellule, nemmeno avvertita dal paziente, alla morte per progressione) che però non sono direttamente osservabili con precisione; allora l'intero intervallo viene diviso in alcuni sotto-intervalli in cui è semplice classificare il paziente in relazione alle sue condizioni cliniche osservabili direttamente o mediante analisi strumentali.
2. Volendo misurare la severità della nausea in un paziente, la variabile sottostante può immaginarsi continua (cioè che possa variare con continuità da "nessuna nausea" alla "massima nausea immaginabile"). A meno di non ricorrere ad analoghi visivi (non sempre consigliati per via dei problemi di riproducibilità da cui sono affetti), non c'è possibilità di misurare il livello esatto della nausea. Allora, l'intero intervallo in cui è definita la variabile continua "nausea", viene suddiviso in un certo numero di sotto-intervalli disgiunti in cui il paziente possa facilmente classificare se stesso. Nella pratica, si adotta una scala di Likert a 4 punti, ancorandola a precise condizioni abbastanza ben oggettivabili: 1 = no nausea, 2 = nausea lieve (compatibile con le attività quotidiane), 3 = nausea moderata (interferisce con le attività quotidiane), 4 = nausea grave (il paziente è costretto a letto per la nausea). In uno studio randomizzato di fase 3 sull'efficacia differenziale di due trattamenti antiemetici, il confronto tra i due bracci dell'intensità della nausea (misurata con la scala appena descritta) non può avvenire mediante un test t di Student perché non ha senso calcolare la severità media in quanto le distanze tra due modalità consecutive non sono note (ad esempio tra 1 e 2 c'è una distanza – cioè una differenza di condizioni – diversa da quella che c'è tra 3 e 4).

Schematizzazione operativa

1. ***Caratteri quantitativi***

Come sopra esposto, l'uso dei test parametrici è consentito solo quando il carattere sia quantitativo continuo (cioè una misura) e siano rispettate le assunzioni (sulla popolazione) che sono alla base di tali test. Ad esempio, se il carattere quantitativo è discreto (cioè non varia con continuità, come ad esempio un conteggio, dove la minima distanza tra due diverse modalità è 1), il test parametrico può essere usato, ma non sui dati originali, bensì su loro trasformazioni il cui risultato soddisfi, invece, le assunzioni del test parametrico.

Esempio. Siano 4; 9; 22; 36; 50 cinque conteggi (dati osservati). Si può procedere ad una loro trasformazione che consenta di considerare tali osservazioni come provenienti da una variabile che, invece, soddisfi l'assunzione di normalità per la successiva applicazione del test t di Student sul valore di una media (come, ad es., potrebbe accadere in uno studio di fase II). Nel caso di un conteggio, la trasformazione raccomandata è la radice quadrata dei valori osservati. Quindi, per applicare il test t si considerano i valori 2; 3; 4,7; 6; 7,1 che sono i dati trasformati, cioè quelli originali su cui si è eseguita l'operazione di radice quadrata. Malgrado tale risultato sia supportato da teoremi matematici, una volta

3. Si definisce mediana quel termine che divide la graduatoria (cioè la distribuzione ordinata) in due parti ugualmente numerose.
4. Le scale di Likert consentono di esprimere una misura di accordo con una certa proposizione contrassegnando il valore che maggiormente riflette la condizione o l'atteggiamento del rispondente. Ad esempio, volendo valutare l'impatto dell'emesi sulla qualità di vita, si può chiedere al paziente se l'emesi ha consentito lo svolgimento di tutte le attività quotidiane. Il rispondente sceglierà il punto che maggiormente rappresenta la sua condizione (1 = le ha consentite completamente, 2 = le ha ostacolate pochissimo, ..., 9 = le ha impedite del tutto). Di norma vengono fissati 5 o 7 o, al massimo, 9 modalità che vanno da "per nulla" a "moltissimo". Se sia preferibile un numero di punti pari o dispari è ancora oggetto di discussione.

- Copyright - Il Pensiero Scientifico Editore downloaded by IP 216.73.216.118 Fri, 04 Jul 2025, 01:42:51

eseguita la trasformazione va comunque verificato, con test *ad hoc*, che le assunzioni alla base del test t siano rispettate.

In alternativa alle trasformazioni dei dati (che furono particolarmente sviluppate nel periodo 1945-65), oggi si preferisce ricorrere a più sofisticate analisi statistiche (modelli lineari generalizzati).

2. *Caratteri sconnessi (scale nominali)*

In tal caso, i test parametrici non possono essere usati. Ad es., se la risposta è in termini di successo/insuccesso terapeutico (carattere dicotomico) a nessuno verrà mai in mente di usare il t-test per provare l'efficacia differenziale di due trattamenti. Le uniche analisi possibili sono quelle incentrate sulle frequenze.

3. *Caratteri ordinabili (scale ordinali)*

Sebbene in qualche lavoro alcuni autori tentino la strada dei test parametrici, questi sono, in linea di massima, da evitare non solo perché il loro campo di esistenza è finito (in una scala di Likert a 9 punti la variabile assume i soli valori interi da 1 a 9, senza variare con continuità), ma soprattutto perché le reali distanze tra due attributi consecutivi sono diverse nei vari settori della scala (il che rende la somma – e quindi la media – priva di significato).

Quando i test parametrici non possono essere usati – o ne fosse sconsigliato l'uso – non resta che ricorrere a **test non parametrici** che non hanno bisogno di alcuna assunzione sulla popolazione target. Per motivi di spazio, li esamineremo nel prossimo numero della rubrica, sottolineandone gli aspetti concettuali e trascurando completamente le formule, per la cui comprensione sarebbero necessarie conoscenze di calcolo combinatorio, formule d'altronde non essenziali per un primo approccio all'argomento in quanto oggi i test non parametrici sono calcolati da qualunque software statistico.

Enzo Ballatori