

Analisi descrittiva delle distribuzioni doppie

Riassunto

Sono esposti gli strumenti statistici per eseguire l'analisi descrittiva di una tabella doppia in relazione alla natura del carattere "risposta" ("effetto"), cioè del carattere le cui variazioni siano spiegate da quelle del suo antecedente logico ("causa").

Parole chiave. Indipendenza, media, mediana, scarto quadratico medio, coefficiente di variazione.

Summary

Descriptive analysis of two-ways tables

The statistical instruments able to perform a descriptive analysis of two-ways tables were described.

Key words. Independence, mean value, median, root mean square, coefficient of variation.

I contenuti della presente rubrica non sono rintracciabili nei testi di Statistica perché sono a cavallo tra questa disciplina e la metodologia della ricerca. Con la loro esposizione, basata prevalentemente su esempi, si coglie anche l'occasione per ricordare alcuni elementi importanti della ricerca clinica.

In "Statistica per concetti" del precedente numero di CASCO, sono state definite le distribuzioni doppie ed è stato introdotto l'indice chi-quadro (o chi-quadrato) come misura della strettezza della relazione tra i due caratteri associati. Poiché quanto sarà esposto può essere considerato come una prosecuzione dell'argomento trattato nel numero precedente, si invita il Lettore a rileggerlo.

Se l'indice chi-quadro è uguale a zero, i due caratteri associati sono indipendenti, cioè, al variare dell'uno, l'altro non varia, denotando così l'assenza di ogni legame associativo; in tal caso ogni successiva analisi della relazione tra i due caratteri diventa inutile in quanto la relazione non esiste.

Viceversa, se è $\chi^2 > 0$ (non può mai essere negativo), allora al variare di un carattere anche l'altro varia, attestando così l'esistenza di una relazione tra i caratteri associati; pertanto l'analisi prosegue. Per comodità del Lettore, riportiamo di seguito la classificazione dei caratteri, già esposta nel numero precedente, perché importante ai fini di quanto sarà trattato. I caratteri possono essere:

a. **qualitativi sconnessi (scale nominali).** Si manifestano nelle unità del collettivo con attributi (aggettivi). Dati due attributi rilevati su due unità, si può dire solo se sono uguali o diversi (esempi: sesso, professione, tipo di neoplasia);

b. **qualitativi rettilinei (scale ordinali).** Si manifestano anch'essi come attributi; dati due attributi rilevati su due unità, non solo si può dire se sono uguali o diversi (come per i caratteri sconnessi) ma si può anche stabilire l'ordine, ossia quale dei due attributi precede l'altro (ad es., è maggiore o minore dell'altro). Esempi: stadio della malattia, ECOG performance status, punteggi su una scala di Likert;

c. **caratteri quantitativi (scale di rapporti).** Tali caratteri, nelle unità del collettivo si manifestano in forma di intensità (cioè con un numero) (esempi: peso, superficie corporea, dose di un farmaco, sopravvivenza).

Nella presente nota consideriamo il caso in cui almeno uno dei due caratteri sia qualitativo (sconnesso o ordinabile). Il caso di entrambi i caratteri quantitativi sarà esposto nel prossimo numero. Trattandosi di analisi descrittiva, le tematiche di inferenza non sono considerate.

Se è $\chi^2 > 0$, per procedere ad una corretta analisi descrittiva della distribuzione doppia rappresentata in forma di tabella doppia (o tabella a doppia entrata), è importante individuare quale dei due caratteri associati sia l'antecedente e quale il conseguente logico. Usando in modo inappropriato un linguaggio deterministico, potremmo dire quale sia la causa e quale l'effetto.

Siano (X,Y) i caratteri associati. X è antecedente logico di Y (che è il conseguente) se ha senso valutare solo come varia Y al variare di X.

Esempi (1)

- a. In un gruppo di pazienti affetti dalla stessa malattia allo stesso stadio, X sia il sesso (carattere sconnesso) e Y il punteggio complessivo degli score riportati in un questionario di qualità di vita del paziente. Ha senso chiedersi come varia Y al variare di X, cioè se i maschi presentano un punteggio complessivamente maggiore o minore di quello riscontrato sulle donne, ma non ha alcun senso chiedersi come vari il sesso al variare del punteggio: il sesso è l'antecedente logico (nel linguaggio deterministico diremmo che il sesso è la "causa" e lo score l'"effetto").
- b. Siano T1 e T2 due trattamenti farmacologici (X) e R la risposta in termini di successo o insuccesso terapeutico (Y). Ha senso chiedersi come varia Y al variare di X, cioè se i successi sono stati più frequenti nei pazienti trattati con T1 o in quelli trattati con T2: il trattamento

costituisce l'antecedente logico del risultato. Non ha senso valutare come varia il trattamento al variare della risposta.

Talvolta ciascun carattere può essere visto come antecedente logico dell'altro e, quindi, ha senso valutare sia come varia Y al variare di X, sia come varia X al variare di Y.

Esempi (2)

1. X = statura e Y = peso dei giovani alla visita di leva. In tal caso, la statura può essere considerata l'antecedente logico del peso, ma può anche avere interesse valutare come varia la statura al variare del peso.
2. X = Sopravvivenza e Y = score di qualità di vita. Può avere interesse investigare sia come varia la qualità di vita al variare della sopravvivenza, ma anche come varia la sopravvivenza al variare della qualità di vita.

Le analisi che possono essere condotte fanno riferimento al tipo di carattere "conseguente logico" ("effetto"), indipendentemente dalla natura dell'antecedente logico.

A. Conseguente logico ("effetto"): carattere sconnesso (scala nominale)

Esempio 3: due caratteri dicotomici

Sia X il trattamento: ad es., X1 = terapia, X2 = placebo.

Sia Y la risposta al trattamento: Y1 = successo, Y2 = insuccesso.

Al termine dello studio i risultati sono sintetizzati nella seguente tabella (1):

Tabella 1. *Pazienti secondo il trattamento e il risposta.*

Trattamento		
Risposta	terapia	placebo
Successo	84	50
Insuccesso	36	75
Totale	120	125

Il trattamento è l'antecedente logico. Si tratta di vedere come varia il conseguente (risposta) al variare del trattamento. Occorre, quindi, confrontare la percentuale di successi tra i due trattamenti: la percentuale di successi è stata del 70% (84/120) con la terapia e del 40% (50/125) con il placebo. Sarà il test per il confronto tra due frequenze a stabilire se la diversità osservata possa essere attribuita al caso o, invece, se deriva dal fatto che la terapia ha una sua efficacia farmacologica. Da un punto di vista descrittivo si può dire che, nei pazienti osservati, la terapia è stata più efficace del placebo.

Esempio 4: dose finding.

Sia Y la risposta definita come nell'esempio 3. Nel *dose finding* il carattere antecedente logico sono le diverse dosi del farmaco.

Tabella 2. *Pazienti secondo la risposta e la dose del farmaco.*

Dose del farmaco (mg)			
Risposta	0	75	150
Successo	33	72	75
Insuccesso	77	48	50
Totale	110	120	125

L'analisi descrittiva di tale tabella è del tutto analoga a quella vista sopra: si confrontano le percentuali di successi tra i tre trattamenti: 30% se il trattamento non viene somministrato, 60% se è somministrato alla dose di 75mg, 60% alla dose di 150mg. Il trattamento è attivo e le due dosi (75 e 150mg) hanno la stessa efficacia.

Osservazione (1). Se la risposta è di tipo soft in quanto dipende da una valutazione soggettiva (del paziente e/o del medico), lo studio clinico cui si riferiscono i risultati va tassativamente programmato in doppio cieco e ai pazienti randomizzati a ricevere una dose nulla del farmaco va somministrato un placebo.

B. Conseguente logico: carattere qualitativo rettilineo (scala ordinale)

Esempio 5. In uno studio trasversale sulla *fatigue* in pazienti neoplastici, è stato usato il BFI (Brief Fatigue Inventory) che consta di 10 item. Il ricercatore vuole analizzare la risposta all'item 2 in relazione all'età (meno di 50, 50-64, 65 anni ed oltre). Item 2: "Valuti il suo affaticamento (stanchezza) facendo una crocetta sul numero che esprime meglio il suo ABITUALE grado di affaticamento nelle ultime 24 ore"; segue una scala di Likert a 11 punti con gli estremi ancorati all'esperienza del paziente: 0 = nessun affaticamento, 10 = il peggior affaticamento che si possa immaginare.

Tabella 3. *Pazienti secondo l'affaticamento abituale nelle ultime 24 ore e l'età (anni).*

BFI \ Età →	< 50	50-64	65 e oltre
0	16	27	24
2	5	6	3
4	3	4	1
7	8	5	8
8	6	10	7
9	2	8	7
Totale	40	60	50

Vi sono vari modi di analizzare la tabella 3. Infatti, in relazione allo scopo dello studio il ricercatore

- a. potrebbe essere interessato solo alla fatigue severa, ponendo il cut-off da 8 in su;
- b. potrebbe invece interessarsi a chi non prova fatigue o ne prova in misura lieve (cut-off: da 0 a 2);

c. potrebbe decidere di eseguire un'analisi complessiva.

I casi (a) e (b) ricadono in quanto già esposto in precedenza.

Per il caso (a), le percentuali di chi prova fatigue severa sono 20% (8/40), 30% (18/60), 28% (14/50) rispettivamente per le tre classi di età, concludendo così che la fatigue severa è più frequente dai 50 anni in su.

Per il caso (b) le percentuali di chi prova fatigue lieve (BFI = 2) o non ne prova affatto (BFI = 0) sono 52,5% (21/40), 55% (33/60), 54% (27/50) rispettivamente per le tre classi di età: la presenza di una fatigue lieve (2) o nulla (0) è praticamente indipendente dall'età.

Il conseguente logico è una scala ordinale, per cui tra le tre classi di età si possono confrontare le mediane (v. CASCO 13, Statistica per concetti e CASCO 9, Statistica per concetti 2). Esse risultano sempre uguali a 2 per ciascuna classe di età e pertanto non variano al variare dell'età. In conclusione, l'influenza dell'età sulla risposta all'item 2 del BFI è complessivamente debole; solo una fatigue severa è più frequente sopra i 50 anni. Come si può constatare, dicotomizzando il carattere (BFI) si possono ottenere informazioni che altrimenti sarebbero nascoste per l'effetto media nella massa dei casi osservati.

Osservazione (2). *Non è consigliabile confrontare la media aritmetica del punteggio per le 3 classi di età perché le risposte all'item 2 del BFI sono solo apparentemente quantitative, in quanto non sono note le distanze tra due score diversi. Ad esempio, la distanza (differenza) tra 2 e 3 non è la stessa che c'è tra 7 e 8.*

C. Conseguente logico: carattere quantitativo (scala di rapporti)

Esempio 6. La severità di vomito.

Nei trial sui farmaci antiemetici, talvolta si confrontano i gruppi sperimentali (T1, T2) rispetto al numero di episodi di vomito (NVOM). È una prassi errata in quanto in tal modo si vengono a mescolare due indicatori diversi: la presenza di vomito e la sua severità. Infatti, se ci sono abbastanza pazienti che sono stati protetti dal vomito (NVOM = 0), solo per questo NVOM potrebbe risultare complessivamente inferiore in un gruppo rispetto all'altro, nascondendo la reale differenza in termini di severità. La prassi corretta confronta anzitutto la presenza di vomito e, successivamente, il numero di episodi assunto come indicatore di gravità dell'emesi, ma solo nei pazienti che hanno vomitato.

Riportiamo i risultati di uno studio di fase III, per il confronto della severità dell'emesi tra i pazienti che ricevono una profilassi antiemetica T1 e quelli trattati con T2.

Tabella 4. *Pazienti che hanno vomitato secondo il numero di episodi di vomito (NVOM) e la profilassi antiemetica*

NVOM \ Tratt. →	T1	T2
1	18	10
2	6	4
3	2	5
5	1	2
8	3	7
14	0	2
Totale	30	30

In tal caso, non solo possono essere eseguite tutte le analisi descritte al precedente punto (B), ma possono essere anche confrontati i valori di diversi indici statistici in grado di dare informazioni interessanti.

Infatti, si potrebbe fissare un cut-off per individuare un vomito particolarmente severo (ad esempio, 8 episodi ed oltre). In tal caso, la percentuale di pazienti con almeno 8 episodi è pari al 10% (3/30) nel gruppo T1 ed al 30% (9/30) in T2. Il che ci porta a concludere che nei pazienti osservati la profilassi antiemetica T1 protegge maggiormente da un vomito particolarmente severo.

Si potrebbe calcolare anche la mediana delle due distribuzioni: il numero mediano di episodi di vomito è pari a 1 nel gruppo T1, mentre è pari a 3 in T2.

In aggiunta a quanto esposto al punto (B), in questo caso possono essere calcolati vari indici statistici.

1. La media aritmetica: il numero medio di episodi di vomito è pari a 2,17 (65/30) nel gruppo T1 e a 4,23 (127/30) in T2: nei pazienti che hanno vomitato, la profilassi antiemetica T1 riduce la severità del vomito rispetto a T2. Da un'attenta lettura dei dati risulta che, se la randomizzazione è 1:1, è verosimile che la protezione dal vomito sia all'incirca la stessa per i due trattamenti (se, in entrambi i gruppi, 30 è il numero dei pazienti che hanno vomitato, la percentuale di quelli protetti dal vomito sarà all'incirca la stessa nei due gruppi); quindi T1 risulta preferibile a T2 perché, se è vero che T1 non aumenta la protezione dal vomito, almeno riduce la severità dell'emesi nei pazienti che vomitano.
2. Il confronto tra media e mediana in entrambi i gruppi depone per un'asimmetria positiva dovuta alla presenza di pazienti che hanno un vomito particolarmente severo (v. Statistica per concetti in CASCO 13).
3. Interessante potrebbe essere il confronto tra la variabilità delle due distribuzioni, da condursi per mezzo del coefficiente di variazione (perché le due medie sono diverse, v. Statistica per concetti in CASCO 14). Riportiamo alcuni calcoli per consentire al Lettore volenteroso di verificarli.

T1: Dev (X) = 136,18; $\sigma = \sqrt{(136,18/30)} = 2,13$; CV = $(2,13/2,17) \times 100 = 93,8\%$

T2: Dev (X) = 423,37; $\sigma = \sqrt{(423,37/30)} = 3,76$; CV =

$$(3,76/4,23) \times 100 = 88,9\%$$

Come si vede, lo scarto quadratico medio del numero di episodi di vomito nei pazienti trattati con T1 è uguale a 2,13: ogni paziente ha avuto un numero di episodi di vomito diverso da quello medio, in media, di 2,13 episodi. Nei pazienti trattati con T2, invece, lo scarto quadratico medio è risultato pari a 3,76: ogni paziente trattato con T2 ha avuto un numero di episodi di vomito diverso da quello medio, in media, di 3,76 episodi.

La variabilità del numero di episodi di vomito è leggermente superiore in T1 (CV = 93,8%) che in T2 (CV = 88,9%).

Conclusioni

Nell'analisi descrittiva di una tabella doppia anzitutto è importante individuare quale dei due caratteri, X e Y, sia l'antecedente logico dell'altro ("causa"), così da eseguire le analisi mostrando come le sue variazioni inducano variazioni nel conseguente logico ("effetto").

Può accadere che ciascuno dei due caratteri, X e Y, possa

essere considerato antecedente logico dell'altro; in tal caso si può valutare sia come varia X al variare di Y, sia come varia Y al variare di X (ad es., statura e peso).

Nell'analisi, è il tipo di carattere "conseguente logico" a suggerire con quali strumenti statistici essa possa essere condotta. Si è mostrato che se il conseguente logico è un carattere quantitativo, lo strumentario statistico per l'analisi è molto ampio. In qualche occasione capita di vedere che, pur di non rinunciare alla ricchezza esplicativa di tali strumenti, è stato analizzato come varia l'antecedente logico al variare del conseguente. Solo raramente, però, questo tipo di analisi può essere accettata.

Nella presente nota, per motivi di spazio, non è stato considerato in caso in cui entrambi i caratteri siano quantitativi. Infatti, quando X e Y sono entrambi quantitativi può essere condotta un'analisi assai più raffinata, quella della concordanza (ad es., se e quanto cresce Y al crescere di X), basata sulla regressione lineare. Tale situazione verrà trattata nel prossimo numero.

Enzo Ballatori