Ratio Math. n. 7 - 1994
13 - 24

CONSIDERAZIONI SULL'INFERENZA STATISTICA E LA VERIFICA DI IPOTESI

Enzo Ballone e Vittorio Colagrande***
**Facoltà di Medicina e Chirurgia, Università "G.d'Annunzio" - Chieti*
***Liceo Scientifico "G. Galilei" - Lanciano (CH)*

1. PREMESSA

Nella società odierna la quasi totalità delle indagini statistiche viene effettuata su campioni, diventa allora opportuno introdurre nell'insegnamento secondario superiore le principali tecniche di campionamento e le metodologie statistico-probabilistiche che permettono l'inferenza, ossia la generalizzazione dei risultati ottenuti attraverso una rilevazione per campione a dei collettivi più ampi detti "popolazione".

D'altra parte non si può trascurare la valenza formativa del processo induttivo. La Logica, in senso stetto, offre alle scienze sperimentali strumenti per respingere ("falsificare") una data ipotesi o una teoria quando queste sono in contrasto con un risultato sperimentale. Tuttavia la sola falsificazione è restrittiva per le esigenze delle diverse scienze: è necessario trovare altri strumenti che consentano di "imparare dall'esperienza" con maggiore ampiezza. Si tratta di risalire dalla verifica di un enunciato in uno o più casi particolari alla validità dell'enunciato stesso. È questa l'"induzione" ed è alla base della "logica dell'incerto" e, in particolare, dell' inferenza statistica.

I problemi dell'inferenza statistica sono sostanzialmente due: la stima statistica dei parametri di una popolazione e la verifica delle ipotesi statistiche concernenti la validità o meno di date assunzioni.

Nel presente lavoro, che trae origine dalla relazione tenuta dagli autori nell'ambito del Convegno Interregionale della MATHESIS su "Calcolo delle Probabilità Statistica e Crittografia nell'insegnamento", di Isernia 2-4 Aprile 1992, si esporranno alcune problematiche attinenti la verifica di un'ipotesi sul valore di una frequenza o sulla differenza di due frequenze.

La trattazione vuole essere una proposta didattica che serva di "stimolo" all'approfondimento e di invito alla presentazione "in classe" di alcuni procedimenti di inferenza statistica.

2. IL CAMPIONAMENTO E L'INFERENZA STATISTICA

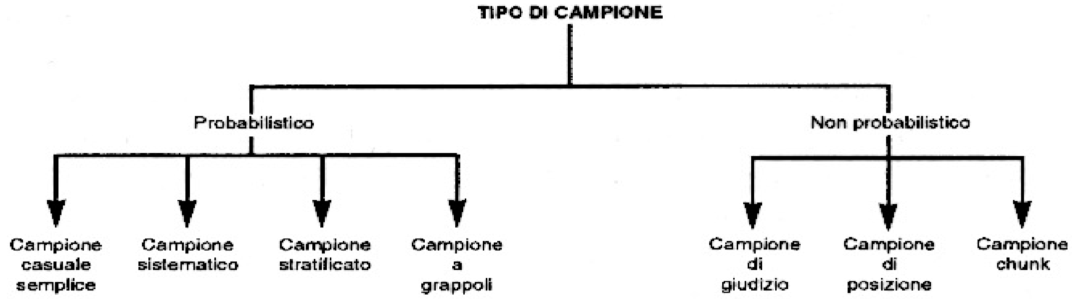
Ogni indagine statistica è mirata a studiare (prendere decisioni su) un carattere X di una ben definita popolazione attraverso alcuni parametri caratteristici (media, varianza, proporzione, ecc.), una parte dei quali o tutti frequentemente sono incogniti.

Supponendo che X sia una variabile casuale (v.c.) e che si distribuisca, nell'ambito della popolazione, secondo una funzione di densità (distribuzione) di probabilità $f(x|\Theta)$ dipendente da un vettore di parametri $\Theta \in R^n$, l'indagine ha come obiettivo quello di stimare o prendere decisioni sul vettore Θ. Nel caso che X sia una v.c. continua con distribuzione normale, $N_{m,s}(x) = 1 / [(2\pi)^{1/2}s] \ \exp[-(x - m)^2 / (2s^2)]$, il vettore dei parametri ha per componenti il valore medio m e lo scarto quadratico medio s.

L'analisi statistica spesso viene condotta su di un gruppo più o meno ristretto di persone od oggetti appartenenti alla popolazione stessa. Tale gruppo costituisce un campione ed è individuato in modo da consentire la generalizzazione dei risultati campionari all'intera popolazione.

Uno schema dei diversi tipi di campionamento è presentato nella fig. 1 seguente:

Fig. 1



La scelta degli elementi che formeranno un campione può essere effettuata in modo che ogni unità della popolazione abbia la stessa probabilità di entrare a far parte del campione. Questo tipo di estrazione, detto **campionamento casuale semplice,** fa parte delle tecniche randomizzate (probabilistiche) e si può provare che, utilizzando queste, il divario tra le informazioni fornite dal campione e quelle della popolazione è dovuto solo al caso.

Le operazioni di scelta degli elementi, poi, possono essere condotte con ripetizione (campionamento bernoulliano) o senza ripetizione (campionamento esaustivo).

Se ogni unità elementare della popolazione viene identificata con un numero intero progressivo a partire da 1, la scelta di ogni singola unità del campione porta alla definizione della variabile X_i = "valore di X per l'i-ma unità" e, quindi, la n-pla $(X_1, X_2, \dots, X_n)$ costituisce il campione. Dunque ogni campione di dimensione n di X corrisponde a n misure ripetute di X fatte nelle stesse condizioni. Nel caso di campionamento bernoulliano le variabili casuali X_i sono indipendenti.

Dopo aver estratto dalla popolazione un un campione che la rappresenti, sui dati campionari si calcolano le statistiche corrispondenti ai parametri della popolazione (media campionaria, deviazione standard, proporzione campionaria, ecc.). Attraverso "la teoria della stima" e "i test statistici" è possibile, con diversi livelli di probabilità di errore, "inferire" dai risultati campionari a quelli dell'intera popolazione, vero oggetto di ricerca.

I motivi che spingono a condurre analisi su campioni vanno ricondotti principalmente a problemi organizzativi, di costi e di tempi per l'acquisizione ed elaborazione dei dati. D'altra parte le indagini campionarie permettono un'analisi più particolareggiata di quanto consentirebbe uno studio di tutte le unità della popolazione oggetto di interesse.

Bisogna ovviamente analizzare se e quando è possibile estendere le conclusioni campionarie a tutto il collettivo. Il campione dovrebbe riflettere le caratteristiche della popolazione ("riprodurre in miniatura" l'intera popolazione) oggetto di studio, ma la mancata conoscenza di tali caratteristiche non consente di giudicare della rappresentatività del campione stesso. Si deve allora ripiegare su un campione che almeno non sia "distorto" e ciò è possibile attraverso la scelta di una tecnica di campionamento appropriata e della dimensione minima ottimale del campione. Stabilita la tecnica di campionamento più appropriata e fissata la dimensione, i risultati dell'indagine campionaria possono essere estesi poi all'intera popolazione, sia pure con un margine di errore prefissato (espresso in termini probabilistici).

Difficilmente, in realtà, si ottengono campioni che riproducano esattamente le caratteristiche della popolazione, tuttavia **nell'Universo dei campioni** (collezione di tutti i possibili campioni di data dimensione che si potrebbero estrarre da una popolazione) ci saranno certamente campioni maggiormente rappresentativi. In generale puo affermarsi che quanto più il campione è numeroso, tanto più sarà rappresentativo della popolazione di provenienza; d'altra parte, però, al crescere della dimensione del campione crescono i costi e le difficoltà tecniche di organizzazione dell'indagine.

3. CARATTERISTICHE DELLA VERIFICA DI IPOTESI STATISTICHE

La verifica di ipotesi in una ricerca statistica è compiuta al fine di confrontare un'ipotesi sulla "composizione" relativa ad un carattere X, in una popolazione oggetto di studio, con i risultati sperimentali che scaturiscono dall'osservazione di un campione estratto dalla popolazione stessa.

In generale un'indagine di tipo statistico parte con la formulazione di una ipotesi H_0 (ipotesi nulla). Essa postula frequentemente la completa accidentalità dell'esito sperimentale. Dall'analisi dei dati campionari, secondo regole di "comportamento induttivo", si arriva alla decisione di accettare o di rifiutare H_0. L'ipotesi nulla quindi è un'assunzione sui parametri della popolazione, ossia:

$$H_0 : \Theta = \Theta_0, \quad \text{con } \Theta_0 \text{ particolare valore di } \Theta$$

Tale ipotesi viene posta a confronto con una alternativa H_1 che viene formulata sul valore che lo stesso parametro può assumere in alternativa a quello considerato con H_0. Ad esempio:

$$H_1 : \Theta \neq \Theta_0$$

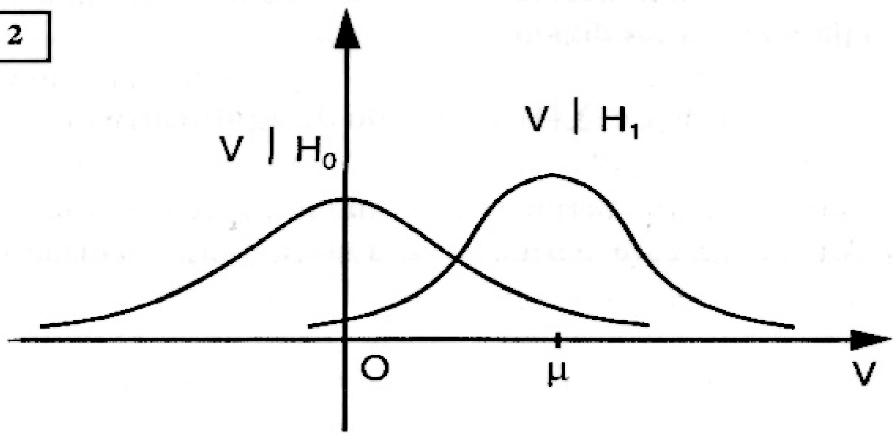
Lo strumento di decisione è un test statistico basato su una opportuna funzione definita sull'universo dei campioni **(funzione test)**. Poiché gli elementi costituenti un campione sono delle variabili casuali, è v. c. anche ogni funzione test. Esiste quindi la possibilità di derivare la distribuzione di probabilità di tale variabile in funzione della distribuzione degli elementi campionari.

Al variare delle n-ple campionarie, la funzione test è tale che:

1) quando H_0 è vera, è descritta da una variabile aleatoria V condizionata ad H_0 ($V|H_0$) avente un'opportuna distribuzione (densità) di probabilità;
2) quando H_1 è vera, è descritta da una variabile aleatoria $V|H_1$ avente una distribuzione (densità) della stessa classe della variabile $V|H_0$, ma con valori diversi dei parametri.

La fig. 2 illustra una possibile situazione concernente la funzione test sotto l'assunto che $V|H_0$ e $V|H_1$ siano distribuite normalmente rispettivamente con medie zero e μ e varianze $var(V|H_0)$ e $var(V|H_1)$.

Fig. 2



La funzione test spesso è costruita definendo una "distanza", espressa in un'opportuna unità di misura, tra la distribuzione (densità) di probabilità che l'ipotesi nulla permette di assegnare ai "punti" dell'universo campionario e un punto campionario scelto opportunamente (valor medio, mediana, moda, ecc).

Convenzionalmente si decide di rifiutare l'H_0 se la distanza tra il campione osservato x e il punto "privilegiato" supera un prefissato valore-soglia. Per la definizione di quest'ultimo, l'insieme Ω dei punti campionari viene suddiviso in due sottoinsiemi, Σ e $\Omega\backslash\Sigma$, in modo tale che si decide di rifiutare H_0 se $x \in \Sigma$ e di accettarla (non rifiutarla) se $x \in \Omega\backslash\Sigma$. Lo spazio Σ di rifiuto di un'ipotesi viene detto "regione critica" e l'appartenenza di un punto campionario a Σ viene stabilita in base al valore assunto su di esso dalla funzione test "opportunamente" scelta.

Ovviamente la decisione presa può essere corretta o no e non sarà dato sapere con "certezza" se la particolare ipotesi formulata sia vera o falsa. Questo perché la prova di verifica di ipotesi statistiche discende da un processo induttivo che utilizza quantità limitate di informazioni.

Vanno introdotti allora tutti i possibili errori che si possono commettere nell'estendere le conclusioni che possono emergere su base campionaria alla popolazione da cui proviene il campione x.

Un errore, errore di prima specie, consiste nel rifiutare l'ipotesi nulla quando, in realtà, questa è vera; per definirlo bisogna considerare la probabilità condizionata

$$\alpha = \mathrm{Prob}(x \in \Sigma \mid H_0) = \textbf{livello di significatività}$$

e la quantità $1 - \alpha$ è interpretabile come la capacità del test statistico scelto di indicare correttamente come vera l'ipotesi nulla sulla base del campione o anche come livello di "fiducia" che si dà al test.

Bisogna considerare anche un secondo errore, errore di seconda specie, che si può commettere se l'ipotesi nulla viene considerata vera, mentre in realtà non lo è. La probabilità di questo tipo errore è espressa da

$$\beta = \mathrm{Prob}(x \in \Omega \backslash \Sigma \mid H_1)$$

e la quantità 1 - β, detta **potenza del test,** esprime la capacità del test di far rilevare la falsità dell'ipotesi nulla.

Il tutto può essere schematizzato nella seguente fig. 3

Fig. 3

		SITUAZIONE REALE NELLA POPOLAZIONE	
		È vera H_0	Non è vera H_0
SITUAZIONE STIMATA IN BASE AL CAMPIONE	Si considera vera H_0	Decisione corretta Prob = 1 - α	Errore di II specie Prob = β
	Si considera non vera H_0	Errore di II specie Prob = α	Decisione corretta Prob = 1 - β

Se alle ipotesi H_i vengono assegnate delle probabilità iniziali $p_i = \mathrm{Prob}(H_i)$, la probabilità degli eventi legati a comportamenti corretti è data dalla

$$\mathrm{Prob}((\{x \in \Sigma\} \cap H_1) \cup (\{x \in \Omega\backslash \Sigma\} \cap H_0)) = p_1 (1 - \beta) + p_0 (1 - \alpha))$$

e quella di comportamenti erronei dalla

$$\mathrm{Prob}((\{x \in \Sigma\} \cap H_0) \cup (\{x \in \Omega\backslash\Sigma\} \cap H_1)) = p_0\, \beta + p_1\, \alpha.$$

Si cerca, ragionevolmente, di rendere massima la probabilità degli eventi legati a comportamenti corretti e minima quella relativa a comportamenti errati. Pertanto se, come usualmente si procede, il livello di significatività α viene fissato e si suppone l'equiprobabilità delle ipotesi ($p_i = 1/2$), il problema di ottimizzazione consiste nel far sì che β sia minimo. Operativamente, fissato α, si cerca di individuare un test che minimizzi β o, in altri temini, di individuare una regione critica che massimizzi la potenza del test.

Il livello di significatività, comunque, non può essere troppo piccolo in quanto una diminuzione dell'errore di prima specie provoca un aumento di quello di seconda specie. Si può arrivare ad una corretta valutazione solo conoscendo la natura del problema in questione e bilanciando i pro e i contro degli errori e delle loro conseguenze.

Ovviamente quanto appena visto è la problematica che si presenta prima dell'osservazione del campione. Dopo l'estrazione di quest'ultimo, la probabilità che sia vera H_0 avendo osservato $x \in \Sigma$ e quella che sia vera H_1 avendo osservato $x \in \Omega \setminus \Sigma$ sono date rispettivamente, per il teorema di Bayes, da

$$\mathrm{Prob}(H_0 \mid x \in \Sigma) = \frac{p_0\, \alpha}{p_0\, \alpha + p_1\,(1 - \beta)} = \frac{\alpha}{\alpha - \beta + 1},$$

$$\mathrm{Prob}(H_1 \mid x \in \Omega \setminus \Sigma) = \frac{p_1\, \beta}{p_1\, \beta + p_0\,(1 - \alpha)} = \frac{\beta}{\beta - \alpha + 1}.$$

La condizione di ottimo richiesta, allora, premette di rendere minima sia la probabilità di verità dell'ipotesi nulla quando l'esito campionario cade nella regione critica, sia quella dell'ipotesi alternativa quando l'esito si trova nella regione di accettazione.

Si osservi, infine, che trattandosi di analisi fatte su campioni, in tutto questo processo inferenziale deve entrare in gioco anche la numerosità campionaria. Infatti risulta immediato che per un campione di numerosità pari a quella della popolazione verrebbero ad annullarsi in due tipi di errori e per uno di piccole dimensioni α e β potrebbero risultare troppo grandi.

4. INFERENZA SU PERCENTUALI

Problemi che spesso si incontrano nella realtà quotidiana e che possono costituire interesse da un punto di vista didattico sono quelli legati a inferenze su "**popolazioni dicotomiche**".

Una popolazione dicotomica è un collettivo le cui unità sono classificate secondo due modalità di un carattere, ad esempio malato o sano, guarito non guarito, presente o assente, promosso o respinto, ecc. In tal caso interessa analizzare la frequenza relativa di una delle modalità del carattere.

La verifica d'ipotesi sul valore π_0 di una frequenza relativa o proporzione π di una data modalità di una variabile X dicotomica, consiste nel provare statisticamente se le osservazioni campionarie siano conformi o meno alla congettura fatta sul valore della frequenza. Il valore può essere "stimato" sulla base di conoscenze precedenti, di "intuizioni" sulla incidenza o prevalenza del fenomeno, attraverso un'indagine pilota, ecc.

Si supponga quindi di estrarre dalla popolazione dicotomica un campione casuale bernoulliano di dimensione n. Indicato con f il numero di unità che nel campione presentano la modalità oggetto d'interesse e con $p = f/n$ la corrispondente frequenza relativa, si prova che, nell'ambito dell'universo campionario, la v.c. frequenza assoluta F si distribuisce in modo binomiale e che per n grande (la regola empirica generalmente usata suggerisce di prendere n tale che $n\pi > 5$ e $n(1 - \pi) > 5$) la distribuzione della proporzione campionaria $P = F/n$ può essere approssimata da una gaussiana con media π_0 e varianza $\pi_0(1 - \pi_0)/n$.

La verifica dell'ipotesi nulla consiste nel confrontare il valore trovato p con il valore ipotizzato π_0: piccole divergenze potranno essere attribuite a fluttuazioni campionarie, divergenze notevoli, invece, denotano che le osservazioni campionarie non si conformano al valore ipotizzato per la frequenza della popolazione.

Fissata l'ipotesi nulla

$$H_0 : \pi = \pi_0. \qquad (\delta = \pi - \pi_0 = 0)$$

occorre considerare anche l'ipotesi alternativa H_1 che presume una popolazione caratterizzata da una frequenza relativa π_1 e può essere quindi:

bidirezionale: $H_1 : \pi_1 \neq \pi_0 \quad (\delta = \pi_1 - \pi_0 \neq 0)$;

unidirezionale: $H_1 : \pi_1 > \pi_0 \quad (\delta = \pi_1 - \pi_0 > 0)$ oppure

$H_1 : \pi_1 < \pi_0 \quad (\delta = \pi_0 - \pi_1 > 0)$.

La quantità δ rappresenta la differenza minima tra frequenze che si ritiene rilevante e di cui, quindi, l'indagine deve essere in grado di evidenziare la significatività statistica. In pratica la reale differenza tra frequenze non è nota prima dell'indagine e la sua valutazione rappresenta, spesso, proprio lo scopo dello studio statistico.

Per la verifica di potesi si può far ricorso alla funzione test

$$Z = \frac{p - \pi}{[p_0 (1 - \pi_0) / n]^{1/2}}$$

la cui distribuzione può essere approssimata con una normale standardizzata.

Parimenti, supponendo di voler confrontare le proporzioni π_1 e π_2 relative ad una stessa modalità di un carattere X in due popolazioni dicotomiche distinte, si estraggono due campioni bernoulliani di dimensioni n_1 e n_2 dalle rispettive popolazioni e si analizza la differenza tra le proporzioni di soggetti p_1 e p_2 che presentano la modalità nei campioni da confrontare. Sempre relativamente a campioni di grandi dimensioni la v.c. $P_1 - P_2$ si distribuisce normalmente con media $\pi_1 - \pi_2$ e varianza

$$[\pi_1(1 - \pi_1)/n_1] + [\pi_2(1 - \pi_2) / n_2].$$

In questo caso le ipotesi da prendere in esame sono:

$$H_0 : \delta = \pi_1 - \pi_2 = 0$$

contro una delle alternative:

$$H_1 : \delta = \pi_1 - \pi_2 \neq 0;$$
$$H_1 : \delta = \pi_1 - \pi_2 > 0 \quad \text{oppure} \quad H_1 : \delta = \pi_2 - \pi_1 > 0$$

e, assumendo:

$$p^* = (n_1 p_1 + n_2 p_2) / (n_1 + n_2)$$

come stima della proporzione incognita $\pi = \pi_1 = \pi_2$ sotto l'ipotesi H_0, la funzione test utilizzabile sarà:

$$Z = \frac{p_1 - p_2}{[p^* (1 - p^*)(1 / n_1 + 1 / n_2)]^{1/2}} . \tag{1}$$

Sia nel caso di confronto tra proporzione campionaria e quella della popolazione che in quello tra due proporzioni, il rifiuto o meno dell'ipotesi nulla è stabilito in base al valore z assunto dalla funzione test Z utilizzata sui dati campionari. Se z, in valore assoluto, risulta maggiore di un opportuno "valore critico", l'ipotesi nulla viene rifiutata; se risulta minore o uguale "in pratica" l'H_0 viene accettata. Il valore critico considerato, nel caso di "test unidirezionali", e dato dalla determinazione Z_α della variabile normale standardizzata tale che $\text{Prob}(|Z| > Z_\alpha) = \alpha$ (livello di significatività). Per "test bidirezionali" il valore critico viene individuato sostituendo a z_α il valore di $z_{\alpha/2}$.

Si consideri il seguente esempio. Da due popolazioni, una di non vaccinati (NV) e l'altra di vaccinati (V) con un vaccino antinfluenzale, si estraggono casualmente 800 e 700 soggetti rispettivamente. Nel primo campione si riscontrino 150 malati ($p_1 = 0.19$) e nel secondo 100 ($p_2 = 0.14$); ci si chiede la capacità immunitaria del vaccino.

L'ipotesi nulla postula nessuna efficacia del vaccino ($\pi_1 = \pi_2$) e l'ipotesi alternativa una significativa attività immunitaria ($\pi_1 \neq \pi_2$). Si tratta quindi di applicare un "test bidirezionale".

Data l'alta numerosità campionaria si può ricorrere alla funzione test (1).

Una stima di π può essere data da $p^* = 250/1500 = 0.17$, che esprime l'incidenza "cumulata" della malattia nell'ambito dei due campioni (NV e V) in esame. Il denominatore della (1) (detto, in termini tecnici, errore standard di una proporzione) vale $[0.17 \cdot 0.83 \cdot (1/800 + 1/700)]^{1/2} = 0.02$ e quindi $|z| = 2.85$. Pertanto, supposto un livello di significatività $\alpha = 5\%$, si ha $|z| > z_{\alpha/2} = 1.96$ e, quindi, si è indotti a rifiutare H_0 ed accettare H_1, ossia ad assumere l'efficacia del vaccino antinfluenzale per ridurre l'incidenza dei casi.

Qualora la stessa indagine fosse stata effettuata su due campioni rispettivamente di 80 e 70 soggetti con 15 e 10 malati rispettivamente, applicando la (1) si otterrebbe un valore di z uguale a 0.83, valore chiaramente non significativo (essendo minore di $z_{\alpha/2}$). Ciò mette in evidenza l'importanza della numerosità campionaria. Infatti la stessa incidenza percentuale di malati (19% nei non vaccinati e 14% nei vaccinati) risulta significativa (a favore del vaccino) quando i campioni hanno numerosità di 800 e 700 unità e di nessuna significatività statistica qualora l'indagine venga svolta su campioni di piccole dimensioni (80 e 70).

BIBLIOGRAFIA

[1] M.L. BERENSON - D.M. LEVINE (1989), *Statistica per le Scienze Economiche*, Zanichelli, Bologna.

[2] F. DI ORIO (1990), *Statistica Medica*, La Nuova Italia Scientifica, Roma.

[3] G. PRODI (1992), *Metodi Matematici e Statistici*, McGrawHill Italia, Milano.

[4] C. ROSSI - G. SERIO (1990), *La metodologia statistica nelle applicazioni biomediche*, Springer-Verlag, Berlin.

[5] M. R. SPIEGEL (1992), *Statistica*, Etas Libri, Milano.