

Speciale N° 12: atti dal workshop  
Scienza e Metodo

## LA STATISTICA E IL METODO SCIENTIFICO

di Donata Marasini e Piero Quatto \*

Nei giorni 10 - 11 Ottobre 2013, presso i Laboratori Nazionali del Gran Sasso si è svolto il Workshop "Scienza e Metodo" su iniziativa dell'Istituto Nazionale di Fisica Nucleare (INFN).

L'evento che ha visto la partecipazione di più di cento insegnanti, era rivolto a docenti di discipline scientifiche e filosofiche per un aggiornamento a carattere interdisciplinare.



\* Università degli Studi di Milano-Bicocca, Dipartimento di Economia, Metodi quantitativi e Strategie d'impresa

La Statistica è un insieme di metodi e di modelli con i quali conoscere e interpretare fenomeni che fanno parte della realtà e che sono di interesse per diverse scienze o discipline come la Fisica, la Biologia, la Medicina, l'Economia, la Sociologia e così via; resta così chiarito l'attributo di ancillare che spesso viene attribuito alla Statistica. Quando i metodi sono impiegati per descrivere o esplorare la realtà, la Statistica prende il nome di «descrittiva». Per esempio, in presenza di una molteplicità di dati, può essere opportuna una sintesi che si può realizzare mediante la *cluster analysis* con la quale si formano gruppi omogenei rispetto a qualche caratteristica di interesse.

Prende il nome generico di Statistica «multivariata» quando identifica e impiega modelli per spiegare un fenomeno in funzione di altri. Per esempio, se la variabile dipendente è qualitativa e quelle indipendenti sono qualitative o quantitative, un modello sovente impiegato è quello della regressione logistica.

Quando, ipotizzata l'esistenza di un modello, la Statistica ricorre a un metodo che richiede la verifica sperimentale e a una procedura di induzione per convalidarlo o meno, prende il nome di Statistica «inferenziale» o «inferenza statistica».

Nel seguito l'attenzione verrà fissata su quest'ultima perché più vicina al metodo scientifico, almeno nell'accezione adottata in questo contesto.

Deve essere anticipato che la Statistica inferenziale può essere affrontata sotto diversi aspetti secondo l'interpretazione che viene attribuita alla probabilità che è strumento necessario nell'induzione. Se si adotta l'interpretazione frequentista, la probabilità è strettamente legata alle frequenze relative con cui un determinato evento si è verificato «nel lungo andare», ossia in un gran numero di prove; se si adotta l'interpretazione bayesiana, la probabilità è strettamente legata al grado di fiducia che un soggetto attribuisce a un evento.

Con riguardo al metodo scientifico, si tratta di fornire una definizione confrontandola poi con il metodo statistico impiegato nell'inferenza, quando quest'ultima viene intesa secondo l'accezione frequentista. Per definire il metodo scientifico, si è scelto di rappresentarlo con questa circolarità [Bellini, 2005]: esiste una «rappresentazione», cioè una realtà da cui partire offerta da dati, teorie, ipotesi; dalla rappresentazione si possono dedurre, anche tramite



l'intuizione, altre ipotesi; nasce la «predizione» la predizione deve essere verificata tramite esperimenti *ad hoc* e l'osservazione di quanto accade; si analizzano i dati ottenuti dall'esperimento; si confrontano con la predizione; si eseguono gli opportuni controlli; la fase finale prevede una procedura di «induzione» che riporta all'inizio del processo con il rafforzamento, o meno, della predizione.

Il metodo statistico nell'inferenza prevede invece i seguenti punti: esiste un modello di partenza che può assimilarsi alla rappresentazione; si ricorre a un campione o si realizza un esperimento che consente l'«osservazione»; si analizzano i dati; si eseguono gli opportuni controlli; con l'induzione si ritorna al modello di partenza che può o meno essere convalidato.

Come emerge dal confronto, il metodo statistico si articola in un numero inferiore di passi rispetto a quello che si è scelto per interpretare il metodo scientifico. Mancano la deduzione, la predizione e il conseguente confronto dei dati con la predizione; deduzione e predizione sono infatti peculiari di alcune scienze o discipline come la Fisica, la Biologia, la Medicina nella ricerca di base, ma non di altre come la Medicina nella ricerca clinica, l'Economia, la Sociologia.

I due esempi seguenti potranno essere di chiarimento.

Con riguardo al campo medico, la ricerca di base può non avere particolari obiettivi se non quello della curiosità del ricercatore che vuole conoscere, per esempio, il meccanismo molecolare che sta alla base di un particolare fenomeno biologico. È la ricerca realizzata nei laboratori che può fornire importanti informazioni e costituisce il punto di partenza per ogni altra scoperta, per esempio, quella di riscontrare proprietà di una molecola del tutto inaspettate. Da questo tipo di ricerca, attraverso la deduzione, si perviene alla predizione.

Sempre in campo medico, la ricerca clinica comprende invece gli studi che si occupano di identificare il trattamento più efficace su pazienti. Così, per esempio, si vuole sperimentare un nuovo farmaco da affiancare o sostituire alla chemioterapia tradizionale; la sperimentazione viene realizzata su un gruppo di pazienti di controllo e su uno i cui componenti vengono trattati con la nuova terapia. Sotto questo profilo rimane la circolarità di cui si è detto senza deduzione e predizione.

Si vuole sottolineare che l'assenza della deduzione e della predizione non rendono una disciplina «non scienza» come invece sembra trasparire in alcune argomentazioni [Bellini, 2005] o, addirittura, nella perentoria dichiarazione del famoso scienziato Ernest Rutherford secondo il quale «nella scienza esiste solo la Fisica; tutto il resto è collezione di francobolli».

## Il metodo statistico nell'inferenza

Premesso che il fenomeno oggetto di interesse sia interpretato da un modello  $X$  che viene chiamato variabile casuale, si supponga di essere interessati ad avere informazione sull'ignota media  $\mu$ . Con questo proposito vengono realizzate  $n$  prove che danno luogo a un campione o a un esperimento; per esempio, si osservano  $n$  individui componenti di una prefissata collettività, oppure si realizzano  $n$  prove indipendenti in un laboratorio biologico.

Indicati con  $(x_1, \dots, x_i, \dots, x_n)$  i risultati delle  $n$  prove, con il proposito di stimare  $\mu$ , gli  $n$  valori

campionari vengono sintetizzati in un numero, per esempio nella media  $\bar{x} = \frac{1}{n} \sum_i x_i$

calcolata sul campione.

Quando si adotta  $\bar{x}$  come stima di  $\mu$ , il circolo di cui si è detto si è compiuto.

Resta da considerare se l'induzione si è conclusa «bene», cioè se nel caso in esame la stima  $\bar{x}$  rappresenta  $\mu$  in modo soddisfacente.

Per potere giungere a qualche conclusione è opportuno associare alla stima uno stimatore, cioè una variabile casuale, studiandone qualche sua proprietà.

Poiché si fa riferimento all'interpretazione frequentista, questa comporta il possibile, anche se non effettivo, ripetersi delle  $n$  prove, così che in ciascuna delle ideali ripetizioni gli  $n$  valori osservabili possono cambiare dando luogo a diversi valori della media campionaria; l'insieme di tutti questi valori crea lo stimatore.

Una proprietà può essere la correttezza che si ha se il valore atteso dello stimatore è

proprio  $\mu$ , come accade per lo stimatore media campionaria  $\bar{X}$  si può poi misurare la sua dispersione perché tanto più è piccola tanto più è probabile ottenere una stima vicina a  $\mu$ . Conoscendo la legge di distribuzione di  $\bar{X}$  è possibile quantificare in termini probabilistici il divario  $\varepsilon$ , dove  $|\bar{x} - \mu| < \varepsilon$ ; è possibile determinare un intervallo entro il quale potrebbe essere compresa  $\mu$  con una prefissata garanzia e così via.

Il metodo statistico si avvicina di più al metodo scientifico, secondo il circolo di cui si è detto, quando si affronta l'argomento della verifica di ipotesi, dal momento che entrano in gioco «ipotesi» da verificare sperimentalmente.

L'intero circolo viene percorso con riguardo al seguente esempio.

In Fisica, l'esistenza del bosone di Higgs era stata intuiva negli anni Sessanta da Peter Higgs e da Francois Englert perché le diverse sperimentazioni portavano alla predizione dell'esistenza di questa particella denominata anche «la particella dannata», perché introvabile. Per questa ipotesi si dovevano portare nuove evidenze rispetto all'ipotesi «non esiste il bosone di Higgs»; solo nel luglio del 2012 è stato dato l'annuncio ufficiale, confermato un anno dopo dal CERN, il tutto reso possibile con l'impiego del metodo statistico.

Il circolo più ridotto invece si verifica, per esempio, nel controllo della qualità in ambito industriale, dove un impianto viene tarato per produrre cuscinetti a sfera di un particolare diametro, ipotesi da sottoporre a verifica. L'impianto deve essere tenuto sotto costante monitoraggio perché un calo di corrente, un cambio di temperatura, un errore nella procedura, può portare a produrre pezzi difettosi. Il monitoraggio si realizza osservando un campione di pezzi per giungere a una decisione sull'esattezza o meno della taratura.

Seguendo l'impostazione proposta da Sir Ronald Fisher negli anni Trenta, un test di significatività è un metodo volto a sondare l'incidenza del caso nei risultati di un esperimento realizzato per la verifica di un'ipotesi di interesse e la sua formalizzazione è la seguente.

Sia  $H_0 : \theta = \theta_0$  l'ipotesi da verificare su un ignoto parametro di interesse  $\theta$ ; costruita un'opportuna misura di discrepanza  $d$  tra quanto osservato e quanto ipotizzato, se detta misura è «grande» l'ipotesi viene rifiutata e accettata, in caso contrario, dove la soglia  $d^*$  che delimita l'accettazione o il rifiuto è la soluzione dell'equazione:

$$P(D \geq d^* | H_0) = \alpha \text{ e l'espressione } |H_0 \text{ deve intendersi «essendo vera l'ipotesi } H_0 \text{»}.$$

Resta, pertanto, identificata la zona di accettazione dell'ipotesi  $H_0$ , ossia  $d < d^*$ , che racchiude probabilità pari a  $(1-\alpha)$  e la zona di rifiuto che racchiude probabilità  $\alpha$ .

Conviene dedicare attenzione alla probabilità  $\alpha$ . Mentre in molti casi pratici, e in accordo anche con i suggerimenti di Fisher, la probabilità in questione viene fissata a livelli standard come 0,05, 0,01, 0,001, nel mondo della Fisica, per esempio, può essere fissata pari a 0,000000573 corrispondente a 1/3,5 milioni o anche a un livello inferiore; ne discende che  $(1-\alpha)$  è pari a 0,999999427 e di conseguenza che la zona di accettazione dell'ipotesi  $H_0$  è a larghissimo spettro. Il che ha una sua giustificazione logica molto evidente; prima di abbandonare l'ipotesi che non prevedeva l'esistenza del bosone di Higgs è necessario avere pressoché la certezza della sua falsificabilità.

Nella pratica dei test, l'attenzione viene spesso fissata su  $d'$  anziché su  $d^*$ , dove  $d'$  è il risultato sperimentale; valutata la probabilità  $P(D \geq d' | H_0) = p$  se  $p \leq \alpha$  si rifiuta

l'ipotesi ritenendo poco probabile ottenere un valore non superiore a quello osservato, essendo vera l'ipotesi stessa, e la si accetta in caso contrario. Il ricorso a  $p$ , che prende il nome di *p-value*, ha il vantaggio di potere verificare nel caso che sia  $p \leq \alpha$  qual è il livello  $\alpha$  che consentirebbe il rifiuto dell'ipotesi e comunque i software statistici forniscono  $p$  e non  $d^*$ .

Conviene ancora ricordare che, accanto o in sostituzione ai test di significatività, Jerzy Neyman e Egon Pearson, sempre negli anni Trenta, hanno introdotto altri tipi di test che all'ipotesi  $H_0$ , detta ipotesi nulla, hanno contrapposto una ipotesi alternativa  $H_1$ . In questo nuovo contesto non si tratta più di accettare o meno l'ipotesi  $H_0$  ma,

nel caso quest'ultima non sia confermata dai dati sperimentali, di accettare l'ipotesi alternativa. Possono così nascere due tipi di errore: quello di prima specie che consiste nel rifiutare  $H_0$  quando è vera e quello di seconda specie che consiste nell'accettare la stessa  $H_0$  quando è vera l'alternativa; nel primo caso la probabilità di commettere l'errore è  $\alpha$ , nel secondo è  $\beta$ , essendo  $\beta = P(D < d^* | H_1)$ .

Ritornando al *p-value*, conviene osservare che molte sono le errate interpretazioni; per esempio, con riguardo al campo fisico, nel caso della «scoperta» dell'esistenza del bosone di Higgs, il *p-value* è stato identificato con la probabilità che l'ipotesi fosse vera; è accaduto che anche il suo complemento sia stato identificato come la probabilità che l'ipotesi «esiste il bosone di Higgs» fosse vera.

Questa interpretazione deriva da una confusione tra probabilità di un'ipotesi  $H_0$  dato un risultato sperimentale  $E$  e probabilità del risultato sperimentale data un'ipotesi, ossia  $P(H_0|E) = P(E|H_0) \times [P(H_0)/P(E)]$ , espressione che si ricava dal teorema delle probabilità composte e che, con alcune semplici estensioni, porta al Teorema di Bayes e alla conseguente impostazione bayesiana dell'inferenza statistica.

Ricordando che nel presente contesto l'impostazione adottata è quella frequentista, le precedenti probabilità  $P(H_0|E)$  e  $P(H_0)$  non hanno spazio perché non possono essere verificate sperimentalmente. Tuttavia, sempre nel campo fisico, rigorosamente frequentista, le due ipotesi «esiste il bosone di Higgs» e «esistono particelle più veloci della luce», ipotesi quest'ultima che ha avuto risonanza con una dichiarazione della sua esistenza nel 2011 smentita poi nel febbraio 2012, hanno avuto associato da parte dei ricercatori un grado di fiducia, molto elevato la prima e molto basso la seconda, tanto che dopo molte sperimentazioni si è accettata la prima mentre per la seconda si è cercato (e trovato) l'errore che ha portato a un particolare risultato sperimentale che sembrava convalidarla.

## Conclusioni

Si è parlato di metodo statistico come supporto di induzione in diverse scienze realizzato tramite la stima e la verifica di ipotesi che hanno le loro origini molto lontane nel tempo ma la loro sistematizzazione nei primi decenni del secolo scorso. Ma che cosa accade in questo secolo anche se solo all'inizio?

È emblematica una frase di Bradley Efron, che ha introdotto nella letteratura un metodo di ricampionamento noto come *Bootstrap*, «*Something new is happening in the twenty-first century...*» Ed è così; l'impiego dei computer da supercalcolo, internet, anche tramite i *social network*, consentono di disporre di un numero inverosimile di dati relativi a una massa enorme di fenomeni. In questo contesto David Donoho, Professore di Statistica alla *Stanford University* ha vinto un premio di 1 milione di dollari per essere stato un pioniere dei *Big Data*, cioè della nuova realtà che, comunque, la Statistica non è ancora pronta ad affrontare.

Donata Marasini - Ordinario di Statistica

Piero Quatto - Professore di Statistica

(Università degli Studi di Milano-Bicocca, Dipartimento di Economia, Metodi quantitativi e Strategie d'impresa)

## Riferimenti bibliografici

Bellini G. (2005), *Divulgazione e metodo scientifico*, Dipartimento di Fisica Università di Milano, INFN - Sezione di Milano.

Birks J.B. (1962), *Rutherford at Manchester*, Heywood, London.

Costantini D., Geymonat L. (1982), *Filosofia della probabilità*, Feltrinelli, Milano.

D'Agostini G. (2012), *Scoperte scientifiche annunciate a colpi di sigma*, Dipartimento di Fisica, Roma La Sapienza.

D'Agostini G. (2011), *Probably a discovery: bad mathematics means rough scientific communications*, ArXiv:1112.3620.v2.

De Cristofaro R. (2003), *The analytical solution to the problem of statistical induction*, Statistica Vol. LXIII, n. 2.

- De Finetti B. (1970), *Teoria della probabilità*, Boringhieri, Torino.
- Efron B. (2008), *Microarrays, empirical Bayes and the two groups model*, *Statistical Science*, Vol. 23, n.1.
- Fisher R.A. (1956), *Statistical methods and scientific inference*, Oliver and Boyd, London.
- Landenna G., Marasini D. (1986), *Uno sguardo alle principali concezioni probabilistiche*, Giuffrè, Milano.
- Lehmann E.L. (1993), *The Fisher, Neyman-Pearson theories of testing hypotheses*, *Journal of American Statistical Association*, Vol. 88, n.424.
- Neyman J., Pearson S.E. (1932), *The testing of statistical hypotheses in relation to probabilities a priori*, *Proceedings of the Cambridge Philosophy Society XXIX*.
- Sandrini M.G.(2009), *Filosofia dei metodi induttivi*, Firenze University Press.