

GIORGIO BERTOLAZZI

**LA FALSIFICAZIONE NELLA SCIENZA MODERNA**

1. Introduzione
2. Falsificazione e corroborazione
3. Approccio statistico alla verifica d'ipotesi
4. Statistica test e p-valore
5. Visione probabilistica e fallibilità della scienza
6. Onestà e oggettività nella verifica d'ipotesi
7. Logica bayesiana nella verifica d'ipotesi
8. Conclusioni

**ABSTRACT: THE FALSIFIABILITY CRITERION AND ITS IMPACT ON THE MODERN SCIENTIFIC RESEARCH**

Karl Popper's falsifiability criterion represents a pivotal point in modern epistemology. This criterion has revolutionized the methodology of scientific research by changing the approach to hypothesis testing. Popper argues that theoretical hypotheses must be empirically tested by assessing whether observations are in contrast with the theory of interest. Popper's ideas provide the theoretical basis for scientific hypothesis testing methodology. Throughout the twentieth century, scientific methods progressively embraced most of the Popperian concepts while rejecting some others. In this brief paper, we aim to highlight both the affinities and discrepancies between Popper's ideas and the statistical methodology adopted by the scientific community. Specifically, we will explore modern developments in the concept of falsifiability. Nowadays, the scientific community wholeheartedly adopts the statistical approach to hypothesis testing, considering it a fundamental cornerstone in research methodology. We will present the statistical methodology for hypothesis testing and the foundational concepts underlying these methods. Finally, an introduction to the Bayesian approach will be provided.



*Chiunque proponga una teoria scientifica, deve essere in grado di rispondere, come fece Einstein, alla domanda:*

*“Sotto quali condizioni dovrei ammettere che la mia teoria sia insostenibile?”*

*Karl R. Popper, La ricerca non ha fine*

**1. Introduzione**

Il criterio di falsificabilità proposto da Karl Popper rappresenta un punto di svolta nell'epistemologia moderna. Questo criterio ha

rivoluzionato la metodologia della ricerca scientifica mettendo in discussione l'approccio confermativo alla verifica d'ipotesi scientifiche<sup>1</sup>.

Popper sostiene che le ipotesi teoriche, formulate mediante deduzione, debbano essere sottoposte a un controllo empirico che consiste nel valutare se vi siano esperienze in grado di falsificare, ossia di smentire, la teoria in questione<sup>2</sup>. Secondo Popper, una teoria può essere considerata "scientifica" solo se esiste una procedura empirica che potrebbe falsificarla<sup>3</sup>. Inoltre, il processo di falsificazione permette di delineare i limiti e le fragilità della teoria in esame, consentendo al ricercatore di definire le condizioni sotto le quali la propria teoria possa essere ritenuta valida<sup>4</sup>.

Il pensiero di Popper rappresenta il fondamento teorico dei metodi per la verifica delle ipotesi scientifiche. Difatti, nel corso del Novecento, la metodologia scientifica si è avvicinata spontaneamente alle tesi popperiane, pur rifiutandone alcuni concetti.

Nonostante il pensiero di Popper abbia influenzato gli sviluppi del metodo scientifico, molti autori hanno criticato le posizioni popperiane relative all'approccio probabilistico<sup>5</sup>. In particolare, uno dei principali limiti dell'impostazione popperiana è legato al

---

<sup>1</sup> K. Popper, *Logica della scoperta scientifica* (1934), tr. it. Einaudi, Torino 1970.

<sup>2</sup> *Ibid.*

<sup>3</sup> Al fine di definire il carattere scientifico di una teoria, Popper scrive: «ammetterò certamente come empirico, o scientifico, soltanto un sistema che possa essere *controllato* dall'esperienza. Queste considerazioni suggeriscono che, come criterio di demarcazione, non si deve prendere la *verificabilità*, ma la *falsificabilità* di un sistema. In altre parole: da un sistema scientifico non esigerò che sia capace di esser scelto, in senso positivo, una volta per tutte; ma esigerò che la sua forma logica sia tale che possa essere messo in evidenza, per mezzo di controlli empirici, in senso negativo: *un sistema empirico deve poter essere confutato dall'esperienza*» (*ibid.*, p. 21).

<sup>4</sup> *Id.*, *Unended Quest: An Intellectual Autobiography*, Fontana, London and Glasgow 1976.

<sup>5</sup> N. Maxwell, *A Critique of Popper's Views on Scientific Method*, in «Philosophy of Science», 39, 2, June 1972, pp. 131-152. Si veda anche R.C. Jeffrey, *Probability and Falsification: Critique of the Popper Program*, in «Synthese», 30, 1/2, 1975, pp. 95-117.

rifiuto del metodo induttivo e della logica probabilistica su cui si fondano le procedure inferenziali della scienza moderna<sup>6</sup>.

In questo breve scritto verranno evidenziate le affinità e le discrepanze presenti tra il pensiero di Popper e la metodologia statistica adottata dalle discipline scientifiche. Particolare attenzione verrà rivolta alle criticità del pensiero popperiano in relazione al metodo induttivo, agli sviluppi moderni del concetto di falsificazione e ai metodi statistici impiegati per svolgere inferenza. L'approccio statistico alla verifica d'ipotesi è oggi unanimemente condiviso dalla comunità scientifica e rappresenta uno dei fondamenti metodologici della ricerca. Verranno presentati gli strumenti statistici per la verifica d'ipotesi e i concetti alla base di tali metodi. Infine, verrà fornita un'introduzione sulla logica bayesiana impiegata nel caso di procedure inferenziali che prevedono l'attuazione di molti esperimenti. Il presente articolo non ha la pretesa di trattare tutti gli aspetti del pensiero di

---

<sup>6</sup> In relazione al concetto di *probabilità di un'ipotesi*, Popper scrive: «Io non credo che sia possibile costruire un concetto di probabilità delle ipotesi che possa essere interpretato come quello che esprime un "grado di validità" dell'ipotesi, in analogia coi concetti "vero" e "falso" (e che oltre a ciò stia, col concetto di "probabilità oggettiva", cioè di frequenza relativa, in una relazione abbastanza stretta da giustificare l'uso della parola "probabilità")» (K. Popper, *Logica della scoperta scientifica*, cit., p. 289).

Popper rifiuta fermamente la logica delle probabilità come criterio di legittimazione del metodo induttivo, sostenendo che un'asserzione probabilistica dovrebbe essere associata essa stessa a un ulteriore grado di probabilità: «Per quanto riguarda la valutazione stessa [...], se la si considera "probabile", è necessaria una nuova valutazione: per così dire, la valutazione di una valutazione, e perciò una valutazione di un livello più alto. Ma ciò significa che siamo intrappolati in un regresso all'infinito. L'appello alla probabilità delle ipotesi non è in grado di migliorare la precaria situazione della logica induttiva. [...] "Una volta per tutte - scrive Heymans - la teoria delle probabilità non è capace di spiegare i ragionamenti induttivi: infatti, esattamente lo stesso problema che si cela nell'una si cela anche nell'altra (nell'applicazione empirica della teoria della probabilità). In entrambi i casi la conclusione va oltre ciò che è contenuto nelle premesse" (G. Heymans, *Die Gesetze und Elemente des wissenschaftlichen Denkens*, Barth, Leipzig 1915). Dunque, sostituendo alla parola "vero" la parola "probabile" e alla parola "falso" la parola "improbabile", non si guadagna nulla»; *ibid.*, p. 290. Sempre in relazione al metodo induttivo, Popper rifiuta fermamente ogni sua possibile applicazione: «L'induzione, cioè l'inferenza fondata su numerose osservazioni, è un mito. Non è né un fatto psicologico, né un fatto della vita quotidiana, e nemmeno una procedura scientifica»; K. Popper, *Congetture e confutazioni* (1963), tr. it. Il Mulino, Bologna 1972, p. 96.

Popper. La finalità perseguita consiste nella contestualizzazione del concetto di falsificazione in relazione alla metodologia statistica contemporanea.

## **2. Falsificazione e corroborazione**

Poniamo di essere interessati a valutare la validità dell'ipotesi  $H$ . L'approccio popperiano prevede l'attuazione di esperimenti empirici volti a falsificare, ovvero a smentire, l'ipotesi di interesse. Se questi esperimenti forniscono un esito negativo, l'ipotesi in esame ne risulta corroborata, ovvero la sua validità è supportata da delle prove empiriche. Inoltre, la procedura di falsificazione consente di individuare quei casi in cui  $H$  perde la sua validità; quindi, tramite la procedura di falsificazione siamo in grado di delineare i limiti dell'ipotesi  $H$ .

In Popper il problema di definire una misura che quantifichi il grado di corroborazione di un'ipotesi rimane insoluto. Inoltre, nonostante Popper metta in evidenza il fatto che una procedura di falsificazione possa anch'essa essere soggetta a degli errori, l'impostazione popperiana non propone misure che quantificano il grado di fiducia nei confronti di una falsificazione.

La metodologia moderna ha adottato un approccio statistico alla verifica d'ipotesi. Quest'approccio propone il  $p$ -valore come misura empirica che quantifica il nostro grado di fiducia nella falsificazione di un'ipotesi.

Prima di soffermarci sull'interpretazione del  $p$ -valore è necessario introdurre la logica alla base dell'approccio statistico alla verifica d'ipotesi, evidenziando le differenze rispetto alla falsificazione popperiana.

## **3. Approccio statistico alla verifica d'ipotesi**

L'approccio statistico alla verifica d'ipotesi si basa sulla definizione di un sistema costituito da due ipotesi; un'ipotesi nulla, detta  $H_0$ , sulla quale si focalizza la procedura di

falsificazione, e un'ipotesi alternativa, detta  $H_1$ , che rappresenta l'ipotesi in esame, ovvero la nuova teoria scientifica che vogliamo valutare<sup>7</sup>.

L'ipotesi alternativa  $H_1$  è definita in modo tale che se  $H_0$  è falsa,  $H_1$  è da considerarsi valida ( $H_1 = H_0^c$ ). Immaginiamo di voler dimostrare la presenza di correlazione tra due variabili. A tal fine, calcoliamo un coefficiente di correlazione  $\rho$  sui dati osservati. La nostra ipotesi di ricerca  $H_1$  è quindi la presenza di correlazione, mentre l'ipotesi  $H_0$  consiste nell'assenza di correlazione<sup>8</sup>:

$$\begin{cases} H_0 : \rho = 0 \\ H_1 : \rho \neq 0 \end{cases}$$

Questo sistema d'ipotesi, per come è stato costruito, prevede implicitamente l'adozione di  $H_1$  come ipotesi attendibile qualora  $H_0$  venisse falsificata. Inoltre, durante la procedura di verifica, l'ipotesi nulla è da ritenersi valida fino a quando non vi siano prove empiriche atte a falsificarla.

L'approccio statistico rappresenta una variante dell'approccio popperiano classico. Difatti la falsificazione popperiana si focalizza sull'ipotesi  $H_1$ , mentre l'approccio statistico è orientato sulla falsificazione di  $H_0$ , tale falsificazione rappresenta una prova a supporto di  $H_1$ .

L'approccio statistico, apparentemente macchinoso, trova la sua giustificazione nella necessità di definire una procedura che sia cautelativa nei confronti di nuove teorie scientifiche. Per introdurre una nuova teoria devo fornire delle prove<sup>9</sup> empiriche

---

<sup>7</sup> R.A. Fisher, *The Design of Experiments* (1935), Hafner press, New York 1971, pp. 15-17.

<sup>8</sup> K. Pearson, *Note on regression and inheritance in the case of two parents*, in «Proceedings of the Royal Society of London», 58, 347-352, 1895, pp. 240-242.

<sup>9</sup> L'espressione *evidenza empirica* si riferisce a una prova a favore di una teoria. Nonostante questa espressione sia ormai consolidata nel linguaggio scientifico, il termine *evidenza* è una cattiva traduzione del termine inglese *evidence*. Nessuna teoria scientifica può essere ritenuta evidente o sicuramente vera.

solide che falsificano l'ipotesi  $H_0$  generalmente ritenuta valida<sup>10</sup> dalla comunità scientifica. Per tale ragione, la procedura di verifica d'ipotesi tutela la validità di  $H_0$  rispetto a un suo rifiuto. Nella pratica, sono in grado di falsificare  $H_0$  a favore di  $H_1$  solo se posseggo prove empiriche solide che contrastano  $H_0$ .

Per saggiare le ipotesi che definiscono il sistema, l'approccio statistico prevede il calcolo di una misura probabilistica, detta  $p$ -valore, che quantifica la forza di falsificazione di  $H_0$ . Nel prossimo paragrafo approfondiremo il concetto di  $p$ -valore e presenteremo gli strumenti matematici impiegati per ottenere tale misura.

#### **4. Statistica test e $p$ -valore**

La statistica test è lo strumento utilizzato per calcolare il  $p$ -valore e valutare la falsificazione di un'ipotesi. Questo strumento è una funzione matematica la cui realizzazione dipende dai dati empiricamente osservati. La realizzazione empirica di una statistica test corrisponde a un valore numerico che può essere coerente o incoerente con l'ipotesi nulla  $H_0$ . Un valore della statistica test coerente con  $H_0$  corrobora tale ipotesi, mentre un valore incoerente rappresenta una prova a favore della falsificazione di  $H_0$ .

Immaginiamo di aver costruito una statistica test in modo tale che più è elevato il valore della statistica test maggiore è la forza di falsificazione nei confronti di  $H_0$ . In questo contesto il  $p$ -valore è definito come la probabilità di osservare un valore della statistica test maggiore del valore empiricamente osservato<sup>11</sup>. Questo equivale a dire che il  $p$ -valore corrisponde alla probabilità

---

<sup>10</sup> Una teoria che negli anni si è affermata dalla comunità scientifica è ritenuta *valida* fino a quando non viene falsificata. Non è corretto parlare di *verità* di una teoria o di un'ipotesi. Difatti, la scienza non conduce a un sapere definitivo e, certo, il rifiuto di un'ipotesi è parte integrante della pratica scientifica.

<sup>11</sup> Il  $p$ -valore è pari alla seguente probabilità;  $p = Prob\{T \geq t\}$ , dove  $T$  è la variabile casuale chiamata statistica test, mentre  $t$  è la realizzazione empirica della statistica test, ovvero la prova empiricamente osservata.

di ottenere una prova contro  $H_0$  forte almeno tanto quanto la prova empiricamente osservata.

Il  $p$ -valore permette di quantificare probabilisticamente il nostro grado di fiducia nella falsificazione di  $H_0$ ; più è piccolo il  $p$ -valore maggiore è la forza di falsificazione. Per esempio, un  $p$ -valore pari a 0.04 indica che la validità di  $H_0$  è poco probabile<sup>12</sup>. Quindi, anche se l'ipotesi  $H_0$  non può essere totalmente esclusa, sono portato a pensare che non sia valida in quanto poco probabile. Un  $p$ -valore che possiede un potere falsificante nei confronti di  $H_0$  viene detto significativo, viceversa un  $p$ -valore che non ha nessuna forza falsificante viene detto non-significativo.

È importante sottolineare che un  $p$ -valore non-significativo non rappresenta una prova a favore di  $H_0$  (l'ipotesi  $H_0$  è da ritenersi valida fino a prova contraria). Tale risultato, non essendo in grado di smentire la validità di  $H_0$ , consolida la fiducia in tale ipotesi, per questo motivo si parla di corroborazione di  $H_0$ .

Riassumendo: la mancata falsificazione dell'ipotesi  $H_0$  aumenta il grado di fiducia nei confronti di questa ipotesi, ovvero un  $p$ -valore non-significativo corrobora l'ipotesi nulla. Ciononostante, il  $p$ -valore non permette di quantificare il grado di corroborazione dell'ipotesi nulla. Difatti, due  $p$ -valori non-significativi hanno lo stesso contenuto informativo<sup>13</sup>.

---

<sup>12</sup> È importante sottolineare che il  $p$ -valore non coincide con la probabilità che l'ipotesi nulla sia falsa. Il  $p$ -valore corrisponde alla probabilità di osservare una prova empirica contro l'ipotesi nulla che sia forte almeno tanto quanto la prova empirica osservata. Tale probabilità è calcolata sotto l'assunzione che l'ipotesi nulla sia valida. In altre parole, ci chiediamo quanto è improbabile osservare ciò che abbiamo osservato assumendo che l'ipotesi nulla sia valida.

<sup>13</sup> È errato ritenere che due  $p$ -valori non-significativi distanti tra loro, come 0.8 e 0.3, abbiano un valore informativo diverso. Questa convinzione deriva dall'idea errata che, sotto  $H_0$ , un  $p$ -valore pari a 0.8 sia maggiormente probabile rispetto a valori più bassi. In realtà, da un punto di vista probabilistico, i  $p$ -valori 0.8 e 0.3 corrispondono a due valori di una statistica test che presentano delle probabilità di realizzazione perfettamente comparabili (sotto  $H_0$ , il  $p$ -valore ha una distribuzione uniforme, per questo motivo la differenza tra questi valori è trascurabile in termini probabilistici).

### 5. *Visione probabilistica e fallibilità della scienza*

L'idea che la pratica di ricerca scientifica sia costantemente esposta a degli errori è il principale punto di rottura tra la scienza moderna e il pensiero neopositivista. Popper sottolinea che anche una procedura di falsificazione può risultare errata; quindi, un'ipotesi che è stata falsificata può essere riabilitata in seguito. Ma Popper rifiuta un approccio probabilistico nella quantificazione degli errori legati alla verifica d'ipotesi.

Inoltre, Popper ritiene che non sia possibile formulare teorie valide universalmente e rifiuta il metodo induttivo, fino ad affermare che l'adozione di una logica delle probabilità non può risolvere il problema dell'induttivismo<sup>14</sup>. Questa è la maggiore divergenza tra il pensiero di Popper e la metodologia moderna, i cui metodi inferenziali sono basati su un approccio probabilistico. Il rifiuto di una logica inferenziale basata su concetti probabilistici è da considerarsi il maggiore limite della filosofia di Popper. Nella scienza moderna le prove a favore di un'ipotesi devono essere sempre accompagnate dalla quantificazione della probabilità che queste prove siano errate.

Se affermiamo che la percentuale di disoccupati in Italia è compresa tra il  $p_1\%$  e il  $p_2\%$  devo fornire la probabilità che la percentuale di interesse non si trovi all'interno dell'intervallo proposto, altrimenti l'informazione trasmessa non ha una validità scientifica. In altre parole, una procedura inferenziale è legittimata dalla quantificazione dell'errore. Quindi, in risposta a Popper, dopo aver osservato un miliardo di corvi neri posso ragionevolmente affermare che tutti i corvi sono neri, ma devo quantificare l'errore che potrei commettere sostenendo questa tesi. Gli errori che si possono commettere durante una verifica d'ipotesi sono:

- Errore di *I* specie: falsificazione di  $H_0$  nel caso in cui questa ipotesi sia valida

---

<sup>14</sup> *Ibid.*

- Errore di *II* specie: mancata falsificazione di  $H_0$  nel caso in cui questa ipotesi sia falsa

Durante una procedura inferenziale è sempre necessario quantificare le probabilità di commettere questi errori. Ovvero è necessario calcolare la probabilità  $\alpha$  di commettere un errore di *I* specie e la probabilità  $\beta$  di commettere un errore di *II* specie.

Commettere un errore di *I* specie è generalmente considerato più grave rispetto al commettere un errore di *II* specie. Questo è dovuto al fatto che, per ragioni cautelative, una nuova ipotesi deve essere supportata da prove empiriche solide per essere accettata dalla comunità scientifica. Per questa ragione, la probabilità di errore di *I* specie  $\alpha$  non deve superare una soglia, generalmente pari a 0.05, fissata a priori dal ricercatore.

Nella pratica, ritenere accettabile una probabilità del 5% di errore di *I* specie implica che un  $p$ -valore minore di 0.05 è da ritenersi statisticamente significativo. Questo implica che, qualora l'ipotesi nulla  $H_0$  fosse valida, vi è una probabilità del 5% di commettere un errore di *I* specie.

Il motivo per il quale non posso fissare una soglia troppo bassa sulla significatività del  $p$ -valore è dovuto al fatto che le probabilità degli errori di *I* e di *II* specie sono tra loro inversamente proporzionali. Quindi una probabilità di errore di *I* specie molto bassa determina una probabilità di errore di *II* specie molto elevata, e questo invaliderebbe la procedura di verifica d'ipotesi.

### **6. Onestà e oggettività nella verifica d'ipotesi**

L'approccio statistico alla verifica d'ipotesi è considerato *bona fide*. Ovvero per saggiare l'ipotesi  $H_1$  adotto un approccio cautelativo che tutela la validità di  $H_0$ . Questo ci costringe a produrre delle prove empiriche solide per falsificare  $H_0$  a favore di  $H_1$ .

Immaginiamo di voler mostrare che un nuovo farmaco determini una percentuale di guarigioni  $p_1$  maggiore rispetto alla percentuale di guarigioni  $p_2$  prodotta dal farmaco attualmente in commercio. Per saggiare tale ipotesi posso usare il seguente sistema d'ipotesi:

$$\begin{cases} H_0 : p_1 \leq p_2 \\ H_1 : p_1 > p_2 \end{cases}$$

Questo tipo di studio prende il nome di studio di superiorità, ed è perfettamente *bona fide*.

Immaginiamo adesso di volere mostrare che il nuovo farmaco sia equivalente al farmaco in commercio (l'uso del nuovo farmaco può avere dei vantaggi economici, quindi la sua commercializzazione è ammessa anche se non c'è un miglioramento della salute del paziente; in questo caso devo dimostrare che non può esserci un peggioramento dell'efficacia). Per saggiare l'ipotesi di equivalenza definisco il seguente sistema d'ipotesi:

$$\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 < p_2 \end{cases}$$

Questo sistema d'ipotesi non può essere ritenuto *bona fide* perché sto dando un vantaggio considerevole all'ipotesi che voglio sostenere. Per tale ragione, nel caso di uno studio di equivalenza<sup>15</sup> il sistema d'ipotesi appropriato è:

$$\begin{cases} H_0 : p_1 < p_2 \\ H_1 : p_1 = p_2 \end{cases}$$

Nonostante sia sconsigliato l'uso di procedure che non siano *bona fide*, nella pratica scientifica sono presenti alcune eccezioni. Per esempio, molti sistemi d'ipotesi per saggiare la presenza di distribuzione normale dei dati, come il test di Shapiro-Wilk<sup>16</sup>, assumono a priori che i dati abbiano una distribuzione normale.

<sup>15</sup> A. Juneja, A. R. Aggarwal, T. Adhikari, A. Pandey, *Testing of Hypothesis*, in «J Clin Diagn Res.», 10, 4, 2016, pp. LG01-LG03.

<sup>16</sup> S. S. Shapiro, M.B. Wilk, *An analysis of variance test for normality (complete samples)*, in «Biometrika», 52(3/4), 1965, pp. 591-611.

Un'altra importante caratteristica propria dei test statistici è l'oggettività. Una procedura di verifica d'ipotesi si dice oggettiva se, fissata a priori una soglia sull'errore di I specie, vi è un criterio univoco per saggiare la validità dell'ipotesi nulla. Per esempio, se stabiliamo una soglia del 5% sull'errore di prima specie, un  $p$ -valore minore di 0.05 è da ritenersi statisticamente significativo.

Generalmente le procedure di verifica d'ipotesi posseggono la proprietà di oggettività. Vi sono però delle rare eccezioni, come il caso degli studi di equivalenza, i quali necessitano della definizione di un intervallo di significatività<sup>17</sup>.

### **7. Logica bayesiana nella verifica d'ipotesi**

La qualità dello strumento scelto per svolgere inferenza dipende direttamente dalla probabilità  $\alpha$  di errore di I specie e dalla probabilità  $\beta$  di errore di II specie. Prima di svolgere una procedura di verifica d'ipotesi è necessario chiedersi:

- Se l'ipotesi nulla fosse valida, qual è la probabilità  $\alpha$  di falsificarla erroneamente?

- Se l'ipotesi nulla fosse falsa, qual è la probabilità  $\beta$  che la procedura di falsificazione fallisca?

Immaginiamo, adesso, di aver svolto un esperimento che ci ha condotto alla falsificazione di un'ipotesi. Ciò che ha senso chiedersi è quale sia la probabilità di avere falsificato erroneamente l'ipotesi nulla. Questa probabilità non coincide con la probabilità  $\alpha$  di errore di I specie. Inoltre, le prove empiriche raccolte influenzano direttamente la probabilità di validità di un'ipotesi.

Tale ragionamento è analogo alle considerazioni alla base delle procedure diagnostiche proprie della pratica clinica; la scelta di uno strumento diagnostico dipende dalla probabilità di

---

<sup>17</sup> European Medicines Agency, *Guideline on the choice of the noninferiority margin*, Doc. Ref. EMEA/CPMP/EWP/2158/99, London 2005.

diagnosticare la malattia nel caso in cui il paziente sia malato (i.e., sensibilità del test) e dalla probabilità di diagnosticare l'assenza della malattia nel caso in cui il paziente sia sano (i.e., specificità del test). Dopo che il medico avrà effettuato la diagnosi, le probabilità appena menzionate non avranno nessuna utilità clinica. Se il medico diagnosticasse la malattia ci interesserebbe la probabilità che il paziente sia realmente malato (i.e., valore predittivo positivo), mentre se il medico diagnosticasse l'assenza di malattia ci chiederemmo quale sia la probabilità che il paziente sia realmente sano (i.e., valore predittivo negativo).

Questo modo di ragionare può essere definito bayesiano-empirico; le probabilità a posteriori di corretta diagnosi vengono calcolate sulla base dei risultati empirici prodotti dall'esperimento.

Nel caso di una procedura volta a saggiare più ipotesi statistiche, il *false discovery rate* (FDR)<sup>18</sup> corrisponde alla proporzione attesa d'ipotesi nulle falsificate erroneamente sul totale delle ipotesi falsificate. Per esempio, se il FDR avesse un valore intorno al 5%, ci aspettiamo che, su 200 ipotesi falsificate, circa 10 ipotesi siano state falsificate erroneamente<sup>19</sup>.

Nel contesto dell'inferenza di larga scala, ovvero in presenza di molte ipotesi da saggiare, per fare in modo che il valore del FDR sia intorno al 5%, la probabilità  $\alpha$  di errore di I specie dei singoli test deve essere inferiore al 5%. Da un punto di vista pratico, la soglia di significatività dei  $p$ -valori potrebbe

---

<sup>18</sup> J. Storey, *A Direct Approach to False Discovery Rates*, in «Journal of the Royal Statistical Society Series B: Statistical Methodology», Volume 64, Issue 3, August 2002, pp. 479-498.

<sup>19</sup> Dal punto di vista del calcolo delle probabilità, l'errore di I specie è pari alla probabilità di falsificare l'ipotesi nulla dato che questa sia valida:  $\alpha = \text{Prob}\{\text{Falsificazione} \mid H_0 \text{ valida}\}$ . Mentre, la definizione bayesiana dell'FDR si basa sulla probabilità che l'ipotesi sia valida dato che siano state osservate delle prove empiriche che hanno condotto alla sua falsificazione:  $\text{Prob}\{H_0 \text{ valida} \mid \text{Falsificazione}\}$ ; cfr. B. Efron, *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*, Cambridge University Press, Cambridge 2010, Chapter 5, "Local False Discovery Rates". Da notare come, nelle due probabilità appena descritte, il condizionamento risulti invertito.

risultare molto più bassa di  $0.05$  e alcuni  $p$ -valori molto piccoli potrebbero essere non-significativi. Procedendo in questo modo stiamo rendendo più stringente il criterio decisionale relativo a ciascun test al fine di operare un controllo sull'errore globale legato all'intera procedura inferenziale<sup>20</sup>.

Tali considerazioni sono di fondamentale importanza nel contesto dell'inferenza di larga scala. Difatti, le procedure statistiche dipendono fortemente dal numero d'ipotesi sotto esame (più esperimenti svolgo, maggiore è la probabilità di ottenere un risultato statisticamente significativo). Per tale ragione, ignorare il numero di esperimenti potrebbe determinare una frequenza troppo alta d'ipotesi falsificate erroneamente.

## 8. Conclusioni

Nel presente articolo sono stati presentati i principali metodi statistici impiegati nella verifica d'ipotesi, evidenziando le ripercussioni pratiche del concetto di falsificazione sulla metodologia statistica e i limiti dell'impostazione popperiana. In particolare, l'approccio statistico propone il  $p$ -valore come misura per quantificare il nostro grado di fiducia nella falsificazione di un'ipotesi. La costruzione di un sistema d'ipotesi permette di stimare le probabilità associate ai possibili errori che possono manifestarsi durante una procedura di falsificazione. Inoltre, il punto di vista bayesiano alla verifica d'ipotesi permette di valutare a posteriori le probabilità di errore a fronte dei risultati osservati empiricamente.

L'impostazione probabilistica della metodologia contemporanea è in netto contrasto con il rifiuto del metodo induttivo presente nel pensiero di Popper. Difatti, l'impostazione probabilistica

---

<sup>20</sup> In questo esempio, il FDR rappresenta l'errore globale legato a una procedura inferenziale che comprende  $N$  test. Un'altra misura dell'errore globale consiste nel *false non-discovery rate* (FNR), questo corrisponde alla proporzione attesa d'ipotesi nulle erroneamente non-falsificate sul totale delle ipotesi non-falsificate. Come nel caso delle probabilità di errore di I e di II specie, il FDR e il FNR presentano tra loro un rapporto inversamente proporzionale.

rappresenta la maggiore divergenza tra il concetto popperiano di falsificazione e la visione scientifica contemporanea. Tramite la metodologia statistica, la scienza ha legittimato le procedure inferenziali, riabilitando così la prospettiva induttivista criticata da Popper.

Nonostante i limiti del pensiero popperiano, il criterio di falsificazione rimane un concetto cardine nella ricerca scientifica. Le discrepanze tra il pensiero di Popper e la verifica d'ipotesi statistica sono spesso più pratiche che teoriche, difatti il concetto di fallibilità della scienza e l'approccio *bona fide* alla falsificazione sono degli aspetti unanimemente condivisi.

GIORGIO BERTOLAZZI è Ricercatore presso il Dipartimento di Scienze economiche, aziendali e Statistiche dell'Università di Palermo

[giorgio.bertolazzi@unipa.it](mailto:giorgio.bertolazzi@unipa.it)