

Il prodotto degli scienziati: precauzioni d'uso.

Nei cartoni animati lo scienziato è un signore molto intelligente ed un po' bislacco, sicuramente infantile, con una grande abilità nel far uso di complicatissime formule matematiche..a volte cattivo, sempre fuori del mondo.

Per i filosofi lo scienziato è un bambino geniale che ha a disposizione le chiavi della verità ma non sa che farsene, per cui il filosofo, meno intelligente ma più saggio, gli sfilava dolcemente di mano quei disegni tanto carini e finalmente li usa per svelare verità al mondo basando su quei disegni, sicuramente veri ('diamine è scienza !') una definitiva visione del mondo.

Sia i cartoni animati che i filosofi comunque concordano sui due punti chiave: l'infantilismo degli scienziati e la centralità delle 'leggi di natura' chiare ed inesorabili.

I cartoni animati sono però epistemologicamente molto più smaliziati: sanno che della scienza non ci si può fidare completamente, soprattutto quando diventa troppo complicata. Will il Coyote ne sa qualcosa quando vede ridicolizzati uno dietro l'altro i suoi incredibili marchingegni da un intuitivo (ed un po' antipatico) Beep Beep. Molto meglio allora il 'semplice' Bugs Bunny che, grazie all'uso accorto (artigiano e quindi fatalmente locale) di due strumenti di base: la velocità nello scavare gallerie e l'abilità nei travestimenti ottiene un successo dietro l'altro.

Sia il quadro filosofico che quello fornito dalla Warner Bros sono delle caricature, quella della Warner Bros è ovviamente più vicina alla realtà, ma anch'essa mostra delle gravi manchevolezze. La caricatura filosofica, non essendo tamponata dalla saggezza dei disegnatori Warner Bros che ci indicano il carattere di finitezza ed imperfezione del fare scientifico ha chiaramente fatto più danni. Qui abbiamo poco tempo per una disamina di tutte le manchevolezze della caricatura filosofica dello scienziato con tutte le sue baggianate su falsificabilità, paradigmi ecc. La più grave è stata forse quella di accreditare la versione del primato della teoria nel fare scientifico: la moltiplicazione dell'icona di Einstein e della sua formuletta sono state così pervasive da far credere anche al grande pubblico che un lavoro molto simile a quello del falegname e del carpentiere fosse in realtà un esercizio di deduzione di teoremi.

Questo chiaramente non era un equivoco nato per caso, ma si iscriveva nella necessità di un pensiero idolatrico che rendesse 'necessario ed inevitabile' tutto ciò che gli scienziati producevano laddove la cifra primaria del fare scientifico è proprio la provvisorietà delle teorie che si usano fintanto che ci aiutano e poi si buttano (tanto sappiamo che sono per definizione incomplete), se mai, quello che ha vita più lunga sono le *metodologie*, noi ancora usiamo strumenti statistici sviluppati per provare teorie dementi. Per quanto questo possa sembrare strano e indurci a dubitare della solidità di questi metodi, basta darsi un'occhiata in giro e scoprire che il metodo di gran lunga più usato (e splendidamente funzionante in campi che vanno dalla fisica quantistica alla sociologia) per misurare le correlazioni è il coefficiente di Pearson sviluppato per dimostrare assurde follie eugenetiche. Pensiamo però che noi possiamo ancora godere dei versi di Dante fregandocene delle sue 'basi filosofiche', così come chiunque osservi un quadro di Caravaggio ne viene inesorabilmente colpito anche se ignora completamente i roveli del problema della Grazia in ambito controriformistico da cui quei quadri sono nati. *Ars longa vita brevis* dicevano gli antichi e più o meno si tratta di questo. Le metodologie (o meglio ancora l'uso personale che ogni scienziato fa delle metodologie che non sono sempre uguali a se stesse) sono il lato artistico della scienza, quello che gli scienziati fanno senza saperlo, le teorie il lato conscio (e quindi caduco), chi è credente come me potrebbe pensare che l'aspetto artigiano adombri il Senso, cioè il lato divino dell'agire umano, in qualsiasi sua attività.

Dopo i tremendi sfracelli generati dal modo di pensare 'teoriocentrico' (nazismo, colonialismo, sfruttamento, deumanizzazione...) insomma, forse è ora che la scienza riacquisti la sua nobiltà di

mestiere artigiano (nobile proprio perché umile), tanto più che di scienza ce n'è un gran bisogno se vogliamo mettere una pezza a problemi come il cambiamento climatico, le sperequazioni del mondo, la deforestazione... e lì di teorie decenti proprio non ne abbiamo.

Allora cerchiamo di trovare una chiave realistica per almeno abbozzare un quadro credibile del mestiere dello scienziato. Di modi ce ne sarebbero tanti, io ho scelto quello della 'retorica scientifica', i tipi di argomentazioni cioè che gli scienziati usano per difendere la plausibilità delle loro tesi. In questo campo incontreremo la statistica, lo strumento principe di argomentazione, qualcosa che tutti faremmo bene a conoscere se vogliamo capire di cosa si parla sulle pagine dei giornali o in tv e non solo quando si parla di scienza.

La statistica non è molto amata dalla maggioranza degli scienziati : è molto spesso considerata un male necessario e gli statistici sono al più tollerati. Da statistico questa cosa mi ha un po' amareggiato, poi ho capito i motivi: gli scienziati crescono con l'idea che se io osservo un certo stato di cose ci dovrà essere un motivo con relativa teoria che lo spiega, gli statistici partono dall'idea che quasi sicuramente quello che ho osservato di causa non ne ha proprio nessuna ed è frutto del caso.

Gli statistici hanno (sto arronzando molto, ma anche io sto nell'ambito della caricatura) come campione Pascal, gli scienziati Newton. Vediamo allora di provare a delineare questa caricatura della scienza basata sulla retorica.

Il lavoro scientifico inizia sempre con una domanda di cui in parte si immagina già di che tipo sia la risposta : la fase iniziale di scelta delle misure da acquisire, della strategia di raccolta dei dati, della definizione delle modalità sperimentali che precede la fase dell'analisi vera e propria è orientata a creare un contesto in cui la risposta sia il più possibile esauriente e con meno ambiguità possibili.

L'impressione che il lavoro dello scienziato sia molto simile ad un artificio retorico o a quello di un avvocato che prepari un'arringa è a questo punto non del tutto infondata¹. La retorica scientifica (quella buona beninteso, non le tronfie dichiarazioni di certi scienziati ai mass media...) ha però una caratteristica che la rende particolarmente affascinante e che costituisce buona parte del gusto del lavoro dello scienziato: tutti i suoi "raggiri" devono essere fatti "prima" che si inizi a giocare, a "bocce ferme". Una volta apparecchiato con la massima cura il quadro di riferimento, si dà il via all'imponderabile, al gioco vero e proprio, all'osservazione, all'esperimento, e la palla passa alla "Natura" o comunque ad un insieme di circostanze di cui noi non si ha controllo diretto e con cui non si può (e non bisogna assolutamente) interferire. A questo punto "si va a vedere" cosa è successo: se la configurazione è quella che ci dà ragione abbiamo avuto successo, altrimenti pazienza, avevamo torto (...questo è il risultato di gran lunga più frequente) oppure non avevamo organizzato le cose in maniera corretta, ci eravamo dimenticati qualcosa o chissà cos'altro.

In effetti il gioco del biliardo è forse la cosa che più si avvicina al mestiere dello scienziato: le combinazioni di regolarità e pura stocasticità sono più o meno nella stessa proporzione in cui si ritrovano nella ricerca scientifica e, soprattutto, i giocatori di biliardo professionisti sono obbligati a dichiarare in anticipo l'effetto del colpo: se per avventura si verifica un colpo ancora più sensazionale ed astruso, non ha lo stesso valore della realizzazione della predizione.

¹ La retorica (intesa nel senso buono di rendere convincenti le proprie posizioni) ha un ruolo importantissimo nella scienza come in qualsiasi altra attività pubblica (e notiamo per inciso che per definirsi pubblica un'attività basta che si rivolga anche ad un solo interlocutore, e molti di noi avranno ancora presenti gli immani sforzi di retorica fatti per convincere di qualcosa che a noi sembrava chiarissimo fidanzate, mariti, genitori, amici...).

Insomma, il bravo scienziato non si misura dal numero di volte che ci azzecca (cosa che dipende per larga parte dal caso) ma da come dispone chiaramente gli elementi del suo quadro così che le sue risposte appaiano chiare (sia in positivo che in negativo)².

Il cammino è molte volte più importante della meta e sicuramente più interessante, come qualsiasi decente escursionista sa bene, e come sapeva bene Martin Heidegger, che premetteva alle sue opere il detto “Wege, nicht Werke” (sentieri non opere compiute).

Il meccanismo retorico alla base di un qualsiasi decente pezzo di scienza può essere sintetizzato a grandi linee come segue: “ Se non ci fosse stato il fenomeno X (*che io dico di aver scoperto*) le cose sarebbero andate nella maniera A; invece gli avvenimenti sono andati nella maniera B, dimostrando l’esistenza di X”. Per mettere su un’arringa del genere abbiamo bisogno di tre elementi fondamentali:

- 1) Avere un modello condiviso (A) di come andrebbero le cose se lo stato del mondo fosse lo stato reale “diminuito” dell’ente X (questa è quella che i metodologi chiamano ipotesi “controfattuale”³)
- 2) Dimostrare in maniera plausibile che le cose sono andate in maniera B, *sensibilmente* diversa da A.
- 3) Rendere necessario il legame tra il non avverarsi di A e l’esistenza di X.

Per motivi di chiarezza espositiva inizieremo dal secondo punto, che è poi il meno difficile dei tre ed il più facilmente trattabile rimanendo sulle generali e senza entrare nel merito di un particolare problema. Per risolvere il punto 2) dobbiamo dimostrare che il particolare risultato osservato, se le cose stanno nella maniera A, è veramente improbabile e, a questo fine, ci serviamo di due concetti estremamente utili: il concetto di *popolazione* ed il concetto di *campione*. Per popolazione intendiamo l’intero insieme di misure teoricamente possibili di un certo fenomeno, per campione le misure effettivamente realizzate.

Il punto 2) allora si risolve stabilendo una misura di “plausibilità” dell’estrazione di un sottinsieme di misure B dalla popolazione di riferimento A.

Immaginiamo di voler eseguire un’indagine epidemiologica sulla relazione tra il livello di istruzione e il ricorso a qualche forma di psicoterapia. In questo caso, la nostra ipotesi di partenza potrebbe essere che la psicoterapia sia più diffusa tra le persone di istruzione medio-alta⁴. La prima cosa da fare sarebbe quella di situare la nostra indagine in un certa posizione spaziale e temporale, in modo da far emergere la nostra variabile di interesse (grado di istruzione) sul “rumore di fondo” provocato da tutti gli altri fattori potenzialmente predisponenti, e quindi decidere da chi è

² Alcuni epistemologi (Popper tra tutti) hanno individuato la cifra precipua del sapere scientifico nella “falsificabilità”, cioè nella possibilità delle teorie di essere smentite dai fatti. Noi avremmo molte cose da obiettare a questa posizione, ma non è questo il luogo per dilungarsi. Qui basta sottolineare ancora una volta come, indipendentemente dal risultato puntuale, sia la “procedura” per arrivare al risultato quello che più ci interessa e che secondo noi può fornire il maggior numero di insegnamenti e suggestioni a chi scienziato professionale non è.

³ In certi ambienti c’è una gran canizza sull’argomento, e si può ben immaginare perché se solo proviamo a pensare a che cosa significhi, ad esempio, “controllare l’effetto dell’appartenenza razziale su una certa malattia depurato dagli effetti accessori”. Saremmo costretti ad immaginare (e a costruire un campione acconcio) che so, negli Stati Uniti, di negri ricchi come i bianchi, che frequentano le stesse scuole, che abitano negli stessi posti, fanno lo stesso lavoro, etc.

⁴ In altre parole potremmo immaginare di avere un modello esplicativo per cui il riconoscimento di una qualche forma di disagio psichico e la conseguente individuazione della psicoterapia come mezzo di cura potrebbe dipendere dal raggiungimento di un certo livello culturale.

rappresentata la nostra popolazione di riferimento. Questo corrisponde a definire le condizioni al contorno della misura.

Nell'esempio proposto viene scelta la popolazione romana di sesso femminile e di età compresa fra i trenta ed i cinquanta anni. Questa popolazione, anche se non praticamente contattabile nella sua totalità, è perfettamente definita: in un qualunque momento (diciamo il 21 gennaio 2005) si può stabilire con assoluta certezza se una persona ha i requisiti per far parte della popolazione controllando i registri dell'anagrafe. La popolazione è costituita da donne con diverso livello di istruzione, che noi abbiamo bisogno di esprimere con una misura il più possibile appropriata. Scegliamo allora come indicatore il numero di anni di studio (STUD), anche se tutti sappiamo che non sempre l'istruzione di una persona è proporzionale al numero di anni passato sui banchi di scuola⁵. Tuttavia, decidiamo che "grossolanamente" la variabile STUD possa essere un'approssimazione di questo concetto, tanto più che ci basiamo su un campione piuttosto vasto in cui eventuali eccezioni si dovrebbero diluire. Una volta definita la popolazione di riferimento, dobbiamo stabilire se la "pesca casuale" di un sottoinsieme molto più limitato di questa popolazione fornisca dei valori della variabile STUD sensibilmente diversi da una "pesca mirata", cioè da una pesca in cui si accettano solo gli individui che hanno in più la caratteristica "aver partecipato a qualche forma di psicoterapia".

Lo scopo è verificare:

- Se il "campione mirato" è a tutti gli effetti considerabile un estratto qualsiasi della popolazione (e quindi ci fornisce dati coerenti con i dati generali della popolazione stessa), oppure è "distorto" dalla richiesta aggiuntiva relativa alla psicoterapia o da qualsiasi altra caratteristica possibilmente correlata al sottoporsi a una psicoterapia, per es. troppe donne divorziate nel campione;
- Se questa eventuale distorsione si riflette in un valore della variabile STUD non compatibile con una pura estrazione casuale.

Per fare questo occorre prima di tutto "condensare" i due nodi cruciali del problema e cioè l'entità delle differenze tra i due campioni e l'entità dell'effetto del caso sulle grandezze sintetiche utilizzate nel confronto. Per avere una misura relativa delle differenze tra i due campioni, la maniera più diretta di procedere è quella di stimare la media aritmetica della variabile STUD nei due campioni e confrontarne la differenza con la "variabilità attesa" nella popolazione.

Alla fine della raccolta dati il nostro materiale sarà sintetizzato dalla tabella seguente dove l'aver avuto esperienze di psicoterapia viene codificato nella classe PSIC, mentre non aver mai partecipato a terapie psicologiche si indica come NOPSIC. Essere PSIC o NOPSIC è a questo punto il discriminante per decidere l'appartenenza ad uno dei due campioni:

CAMPIONE 1

<i>Nome</i>	<i>Classe</i>	<i>STUD</i>
Anna	NOPSIC	8
Laura	NOPSIC	12
Anita	NOPSIC	18
Renata	NOPSIC	8
Maria	NOPSIC	13
Antonietta	NOPSIC	5
Giovanna	NOPSIC	13
Luisa	NOPSIC	8
Benedetta	NOPSIC	18
Marianna	NOPSIC	13
Felicita	NOPSIC	13
Annamaria	NOPSIC	8

⁵ Ammettendo di sapere di che si parla, parlando di 'cultura'.

Maria Luce	NOPSIC	8
Immacolata	NOPSIC	13
Ida	NOPSIC	5
Maria	NOPSIC	17
Olga	NOPSIC	12
Irene	NOPSIC	8
Flaminia	NOPSIC	13
Giuseppina	NOPSIC	13
Loredana	NOPSIC	8
Laura	NOPSIC	5
Veronica	NOPSIC	18
Ilaria	NOPSIC	8
Fiorenza	NOPSIC	8
Silvia	NOPSIC	8
Karen	NOPSIC	8
Teresa	NOPSIC	18
Marina	NOPSIC	5
Elena	NOPSIC	20
Rosa	NOPSIC	5

CAMPIONE 2

<i>Nome</i>	<i>Classe</i>	<i>STUD</i>
Barbara	PSIC	18
Rossana	PSIC	17
Ida Maria	PSIC	13
Letizia	PSIC	13
Ambra	PSIC	12
Elisabetta	PSIC	8
Emanuela	PSIC	8
Grazia	PSIC	17
Maddalena	PSIC	13
Cristina	PSIC	13
Maria	PSIC	5
Rosa Maria	PSIC	20
Rita	PSIC	8
Valentina	PSIC	3
Valeria	PSIC	5
Ornella	PSIC	10
Cristiana	PSIC	13
Silvana	PSIC	18
Bruna	PSIC	18
Lavinia	PSIC	18
Patrizia	PSIC	16
Paola	PSIC	13
Marina	PSIC	5
Tiziana	PSIC	13
Ilde	PSIC	17
Beatrice	PSIC	17
Manuela	PSIC	17
Dora	PSIC	17
Alberta	PSIC	14
Giulia	PSIC	18
Gina	PSIC	8
Federica	PSIC	13
Cleo	PSIC	13
Celeste	PSIC	13

Deborah	PSIC	13
Irma	PSIC	18

La media aritmetica⁶, cioè la somma di tutti i valori della variabile STUD diviso per il numero N degli individui, è semplicemente esprimibile dalla formula:

$$E(\text{STUD}) = \Sigma (\text{stud}(i))/N \quad [1]$$

Nella [1] la lettera E sta per “expectation” cioè “valore atteso” della variabile STUD⁷.

Nel caso proposto dall’esercizio, la media dei 31 individui del gruppo NOPSIC è pari a 10.87 anni di studio, e quella dei 36 individui del gruppo PSIC è pari a 13.19 anni di studio, con una differenza di 2.32 anni. Sono sufficienti circa 2 anni di differenza nelle medie a dire che il grado di cultura ha un effetto sensibile sul ricorso alla psicoterapia ?

Per confrontare questo differenziale di “circa due anni di studio” con la variabilità casuale, abbiamo bisogno di una stima di tale variabilità nella popolazione, ovvero di un indice che ci dica quanto, in media, ci aspettiamo che un’osservazione presa a caso differisca dalla media della popolazione. Si tratta ancora di un valore atteso, ma non più della grandezza tal quale, bensì degli scarti (differenze, distanze) dei singoli accadimenti rispetto al valor medio. Questo indice è la cosiddetta deviazione standard:

$$\text{Std. Dev. (STUD)} = \sqrt{\Sigma (\text{stud}(i) - E(\text{STUD}))^2 / N} \quad (2)$$

A ben vedere, non è altro che la media degli scarti delle singole osservazioni dal proprio centro con il solito trucco del quadrato seguito dalla radice per evitare la somma zero⁸. Bisogna dire che si sarebbe potuto prendere la somma dei valori assoluti, ma la deviazione standard in statistica viene basata sulla metrica Euclidea, che è più maneggevole dell’altra definizione. La deviazione standard insomma ci dà un’idea di quale sia l’entità di una differenza “normalmente attesa” tra un elemento della popolazione e il valor medio della popolazione stessa. È insomma una misura dell’incertezza della nostra media o, se si preferisce della sua scarsa rappresentatività.

La figura 1 ci fornisce un breve riassunto di questi concetti.

⁶ Possiamo immaginare la media aritmetica come il baricentro della distribuzione dei diversi valori della variabile considerata negli individui facenti parte del campione (o della popolazione).

⁷ Mai scelta fu più ambigua: lo slittamento di significato fra “atteso” e “vero” e, soprattutto, fra “atteso in termini di popolazione” (o come frequenza sul lungo periodo) e “atteso in una singola prova puntuale”, ha creato (ed ancora crea) notevoli fraintendimenti non solo fra i non addetti ai lavori ma anche fra scienziati di diverse discipline. Rimaniamo allora coi piedi per terra: la nostra media è semplicemente un’indicazione di dove localizzare il centro di una distribuzione (comunque ampia) di valori lungo un asse.

⁸ La (2) senza l’operazione di quadrato si chiama **varianza**, corrisponde a quello che i fisici chiamano momento di inerzia e, nei segnali, ha anche a che vedere con la “potenza” dei segnali... Noi però rimaniamo al significato intuitivo di “differenza media” tra un valore effettivo di una variabile e il centro della sua distribuzione.

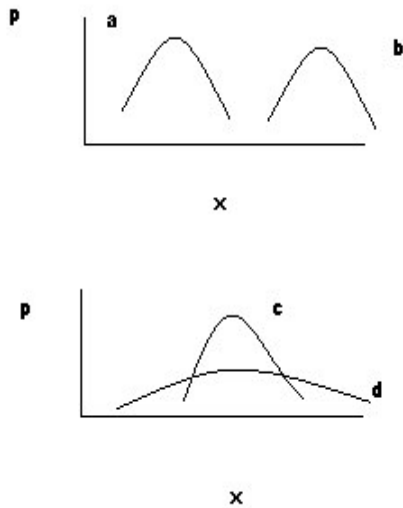


Figura 1. Il significato degli assi , in entrambi i pannelli della figura è identico: x sta per una misura osservata in diversi individui (come la nostra variabile *STUD*), p indica una frequenza, e misura quanto spesso si osserveranno valori della x corrispondenti dopo estrazione casuale da una certa popolazione. Le curve allora rappresenteranno differenti “distribuzioni di frequenza” e ci danno un’idea di quanto le popolazioni che descrivono sono omogenee (raccolte attorno alla loro media, individui molto simili fra di loro) o eterogenee

Nel pannello superiore vengono rappresentate due popolazioni **a** e **b** con la stessa varianza e diversa media, nel pannello inferiore due distribuzioni **c** e **d** con la stessa media e differente varianza. È allora chiaro che la conoscenza del “valore atteso” (media) di **c** ci darà un’informazione più rilevante per la conoscenza della popolazione **c** di quanto non ci fornisca la media di **a**. Ed ecco allora l’uso che potremmo fare dell’informazione sulla dispersione (varianza) per giudicare dell’entità delle differenze tra due popolazioni: semplicemente misurare quanto le due distribuzioni di frequenze sono sovrapposte tra di loro. Nella figura 2, la presenza di una notevole area di incertezza (colorata in nero) in cui un’ estrazione casuale potrebbe provenire da entrambe le distribuzioni rende la mia decisione se una certa osservazione venga dalla popolazione di destra o da quella di sinistra molto più incerta nel caso descritto sopra rispetto al caso descritto sotto.

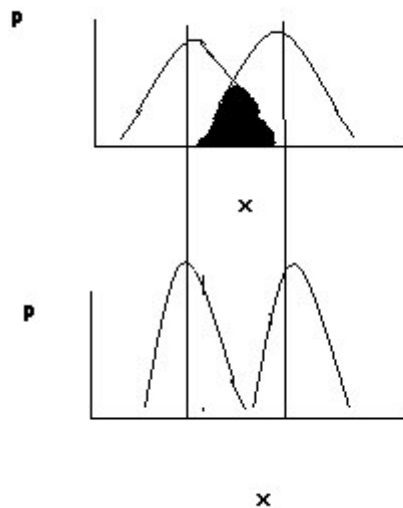


Figura 2
 Nel pannello superiore le due distribuzioni hanno esattamente la stessa differenza media fra di loro delle distribuzioni del pannello inferiore (la media è rappresentata dalle linee verticali). Ma la maggiore varianza delle distribuzioni del pannello superiore diminuisce la separabilità delle due distribuzioni rispetto alle distribuzioni del piano inferiore.

Torniamo alla nostra psicoterapia, la deviazione standard dei due campioni è praticamente coincidente, essendo pari a 4.55 per la classe NOPSIC e a 4.49 per la classe PSIC.

Questo ci dice che la differenza osservata di 2.32, dovuta o meno al caso, è comunque piuttosto piccola, essendo ampiamente all'interno della variabilità naturale. Possiamo farla finita qui? Potremmo, e sicuramente non ci saremmo persi un risultato eclatante, ma ci stiamo dimenticando di una sottigliezza legata ad un insidioso concetto che la scienza ha mutuato da quella potente quanto pericolosa alleata che è la matematica e cioè il concetto di “valore vero”.

Le cose stanno più o meno così: immaginiamo di poter effettivamente chiedere a TUTTE le donne di Roma di età compresa tra trenta e cinquanta anni se hanno mai frequentato sedute di psicoterapia e per quanti anni sono andate a scuola. Questa è un'operazione lunga e tediosa ma che non presenta impossibilità intrinseche. Immaginiamo che alla fine della rilevazione le due popolazioni⁹ PSIC e NOPSIC abbiano come valor medio e deviazione standard proprio gli stessi valori che noi abbiamo osservato nei nostri miseri campioni di 36 e 31 individui. A questo punto noi SAPREMMO CON CERTEZZA che le due popolazioni SONO DIVERSE in quanto PSIC ha media 13.19 e NOPSIC 10.87. Carino, no? Allora il nostro ragionamento di buon senso sull'entità delle differenze si dimostrerebbe un'imperdonabile leggerezza che ci celerebbe la verità. Ricordiamoci che alla base dell'inferenza statistica (come di qualsiasi tipo di inferenza) c'è l'ambizione di derivare “verità generali” da osservazioni parziali e cioè di generalizzare da un campione alla popolazione, e noi stiamo preparando un'arringa per dimostrare che la psicoterapia c'entra con lo studio IN GENERALE e non per le nostre 67 intervistate.

Il punto è quello di definire la rappresentatività del campione rispetto alla popolazione generale, in altri termini il problema diventa: con che verosimiglianza il valor medio della popolazione di riferimento è sovrapponibile a quello osservato nel campione?¹⁰ Questo ci permette di definire una misura di variabilità che, a differenza della deviazione standard, non misura più lo scarto medio di una osservazione dal suo insieme di riferimento, bensì lo scarto della media di un campione dalla media della popolazione da cui è tratto. Questa misura di variabilità si chiama Errore Standard (ES) e corrisponde alla Deviazione Standard diviso la radice della numerosità del campione:

$$ES = \text{Std.Dev.} / \sqrt{N} \quad (3)$$

Nel nostro caso, allora, le deviazioni standard dei due gruppi PSIC e NOPSIC che si situavano attorno a 4.5 corrispondono ad un errore standard attorno a 0.8, decisamente inferiore alla differenza di 2.32 osservata tra i valori medi della variabile STUD nei due gruppi. Infatti l'applicazione di un classico test inferenziale che associa alle distribuzioni osservate una densità di probabilità denominata Gaussiana¹¹ porta a stimare una probabilità del 4% che la differenza osservata fra i due gruppi sia dovuta al caso. Questo valore del 4% ci consente di dire che l'effetto

⁹ Attenzione a questo punto non sono più campioni, cioè sottinsiemi, ma sono proprio due popolazioni intere, rispettivamente: “donne romane fra i trenta e i cinquanta che sono state in psicoterapia” (PSIC) e “donne romane fra i trenta e i cinquanta che non sono mai state in psicoterapia” (NOPSIC)

¹⁰ Senza entrare nei dettagli matematici della dimostrazione, è comunque intuitivo capire come campioni più numerosi siano più rappresentativi di campioni più piccoli.

¹¹ L'adozione della Gaussiana implica una ben precisa assunzione sulla distribuzione degli allontanamenti dalla media. Non sempre questa assunzione è giustificata ed esistono dei casi in cui addirittura non è lecito fare alcuna assunzione su una specifica forma funzionale di distribuzione degli errori se non molto generale come l'ipotesi di continuità'. In questo caso si usano altri metodi inferenziali (es. test non parametrici). Qui però si presuppone che queste assunzioni siano sostenibili, in ogni caso il tipo di ragionamento che segue è assolutamente sovrapponibile anche per diverse assunzioni di forma degli ‘scarti dalla media’.

degli anni di studio sul ricorso alla psicoterapia è statisticamente significativo,¹² anche se solo marginalmente.

Quanto sopra è tutto vero, ma a me (e spero anche a voi) mi sembra di sentire puzzo di bruciato, un vago sentore di imbroglio, ancora più insidioso perché supportato dall'evidenza matematica. Insomma, siamo ammirati della sottigliezza del ragionamento, ma tutto sommato ci era più simpatico quel rozzo ma schietto confronto fra 2.3 e 4.5.

Torniamo allora all'equazione (3) e, come fanno i matematici, facciamo un discorso "al limite", e vediamo cosa succede stiracchiando oltre il ragionevole le condizioni al contorno. In particolare, se la numerosità del campione tende ad infinito (ovvero quando N è molto grande), è facile comprendere che, dato che la Deviazione Standard è un numero finito, aumentando il valore di N che è al denominatore della (3), ES tenderà a zero. Tutto giusto no ? Al crescere della numerosità, il campione diventerà sempre più rappresentativo della popolazione con cui tende a coincidere. Questo risultato però porta con sé una conseguenza perversa: qualsiasi differenza, comunque piccola, rispetto a qualsiasi cosa, con un campionamento sufficientemente grande, diventa statisticamente significativa, cioè 'vera'. E con questo la matematica seppellisce definitivamente la scienza sotto una montagna di ridicolo dimostrando come si possa rigorosamente affermare qualsiasi bestialità se si hanno abbastanza mezzi e tempo a disposizione per affrontare una sperimentazione sufficientemente vasta e quindi (secondo la vulgata corrente anche tra la maggioranza degli scienziati) rigorosa. Ovviamente non è che da questo si possa trarre la conclusione che i 'campioni piccoli sono meglio di quelli grandi' in quanto se i campioni troppo grandi sono affetti dal vizio della banalità, quelli troppo piccoli hanno il vizio complementare delle positività dovute al caso.

Cerchiamo allora di stare calmi e, mentre traiamo da questo risultato un'altra occasione di meditazione su quanto la sensibilità scientifica differisca da quella matematica, ma anche, cosa molto più importante, comprendere l'imbroglio potenzialmente celato dietro frasi come "...suffragato da una vastissima sperimentazione", vediamo di salvare la possibilità di un discorso razionale.

Per prima cosa separiamo il concetto di "statisticamente significativo" da quello di "rilevante". La significatività statistica si riferisce alla probabilità che due misure siano o no assolutamente identiche. Assolutamente identiche è molto diverso da ragionevolmente identiche, tant'è che mentre è praticamente impossibile che, diciamo, l'altezza di due persone misurata con la precisione del decimo di millimetro sia identica, il mondo è pieno di persone alte circa un metro e ottanta. Questo implica che la significatività statistica, come d'altronde qualsiasi altro strumento, è un concetto utilissimo ad un certo dettaglio di misura mentre, oltrepassato questo dettaglio, diventa addirittura controproducente.

Alla base del raggiungimento della significatività statistica della differenza di due campioni stanno tre ingredienti che intervengono alla pari nel raggiungimento di un certo valore di significatività:

- 1) Il valore effettivo della differenza osservata (Differenze grandi hanno più possibilità di essere significative).
- 2) L'entità della variabilità naturale della popolazione stimata dalla Deviazione Standard dei campioni (popolazioni più eterogenee saranno caratterizzate da un grado maggiore di incertezza rispetto alla loro media e quindi permetteranno con più difficoltà di osservare effetti significativi).

¹² Il limite di significatività statistica, cioè la probabilità di prendere una cantonata quando diciamo che due gruppi sono diversi quando invece la loro differenza dipende solo dal caso, convenzionalmente accettato come soglia per rifiutare o accettare un'ipotesi è del 5%.

- 3) La numerosità del campione (un grande numero di osservazione rende i campioni immagini più verosimili della popolazione di riferimento ed aumenta la potenza del test statistico).

Il bilancio di questi tre ingredienti deve essere accuratamente dosato dallo sperimentatore per evitare risultati assurdi, per cui una numerosità molto alta del campione ed una variabilità relativamente bassa della popolazione consente di osservare come significative anche differenze molto piccole tra le medie.¹³

Viceversa un campione poco numeroso in presenza di una notevole variabilità naturale non permetterà di dimostrare come significative anche differenze piuttosto elevate delle medie. Esistono delle formule facilmente accessibili per via Internet (ad esempio all' indirizzo http://department.obg.cuhk.edu.hk/ResearchSupport/Sample_size_EstMean.asp) che consentono di simulare i possibili esiti di un esperimento con domande del tipo “ In presenza di una Dev. Std. attorno a 10 e due campioni da 50 individui ciascuno posso individuare un' eventuale differenza di 20 punti dei miei due gruppi come significativa ?” o viceversa “quale è la differenza minima osservabile come significativa a partire da una Dev. Std. pari a 5 e due campioni da venti individui ?” e così' via....

Questa attività, da svolgersi prima dell'ottenimento dei risultati, consente di avere un'idea del tipo di esperimento da programmare sulla base delle risorse disponibili e, principalmente di cosa io ritengo una differenza degna di nota. Insomma uno sperimentatore che sia interessato ad un farmaco anti-ipertensivo inizia a ritenere un risultato rilevante se il farmaco abbassa la pressione del 20%, mentre non ha alcun interesse ad organizzare un test che dimostri in modo inequivocabile che VERAMENTE il suo farmaco abbassa la pressione in media dello 0.1% in quanto questo abbassamento di pressione ancorché verissimo, certificato da una impressionante e rigorosa sperimentazione, non cambia di nulla la condizione del paziente.

Insomma mentre in matematica esiste un bordo netto tra il VERO ed il FALSO, nelle scienze sperimentali il confine cruciale è tra il rilevante e l'irrelevante e quest'ultimo, anche se è vero, resta sconsolatamente irrilevante. Ed ecco allora che comprendiamo due punti chiave in parte contrari al nostro consueto modo lineare di vedere:

- 1) Che gli strumenti più sensibili non sono necessariamente i migliori.
- 2) Che il concetto di verità di un'informazione è in una posizione subordinata rispetto al concetto di rilevanza.

Il caso della psicoterapia era, però, rimasto in sospeso: del nostro risultato, possiamo dire che vale la pena approfondire o possiamo cavarcela facendo spallucce ?

Per rispondere a questa domanda dobbiamo prima rispondere ad un' altra che è solo apparentemente innocente e cioè ‘Ma perché lo vogliamo sapere?’

Insomma se io sono un assicuratore che vuole stabilire il premio di una polizza di assicurazione sanitaria che comprenda la psicoterapia sicuramente il dato è rilevante: lavorando sui grandi numeri posso pensare che un piccolo ritocco alle rate da pagare basato sul titolo di studio sia ragionevole e conveniente. Se io sono uno psicologo sperimentale che vuole costruire su questo risultato una teoria psicoanalitica sul ‘disagio dell'intellettuale’ è meglio che lascio perdere: il risultato così come è non mi dice alcunché sul singolo paziente che è poi il soggetto di ogni eventuale teoria. Ancora peggio, io non so se l' effetto da me osservato non dipenda da un ‘confondente’ , ad esempio semplicemente dal fatto che di solito chi ha un titolo di studio più alto ha anche maggiore disponibilità economica e quindi si può permettere la psicoterapia. La storia recente dell'epidemiologia è tutta un susseguirsi di proclami e smentite sulle supposte pericolosità di campi elettromagnetici, particolato diesel, coloranti, ecc. ecc. proprio per i

¹³ Si dice allora che il test ha una potenza elevata, è cioè molto sensibile.

problemi che abbiamo cercato di enucleare qui. Di fatto molti asseriscono che l'epidemiologia analitica basata su questo uso della statistica sia ormai arrivata al capolinea.

Qui non abbiamo tempo per parlare di rimedi (che comunque ci sono ma implicano un approccio statistico completamente nuovo, basato sulla correlazione individuale e non sul dato di popolazione) che comunque non cancellano la natura di base del problema e cioè la centralità dell'innocente domanda 'Ma perché lo vogliamo sapere?' che da sola inserisce la dimensione etica nel fare scienza annullando alla radice il mito del valore assoluto della conoscenza scientifica in quanto tale.

Ecco allora che possiamo tornare a Will Coyote e ad alcune 'precauzioni d'uso' che spero troverete utili:

- 1) Cercare sempre di scovare il punto in cui nell'articolo di giornale c'è scritta l'entità dell'effetto osservato. Non se l'effetto c'è o non c'è ma QUANTO è: insomma se X passa da 100 a 101 oppure da 100 a 1000, anche se non conosciamo di che si tratta, più o meno ci facciamo l'idea che nel primo caso l'informazione è perdibilissima (poi ci sarà un grillo parlante che ci dice che no, che pure una variazione dell' 1%... ma insomma lasciamo che sia lui ad inseguirci per strada), nel secondo caso vale la pena continuare a leggere.
- 2) Passare poi a vedere se si dice che variabilità ci si aspetta nel caso non si faccia nulla (variabilità spontanea dell'incidenza di tumori sul territorio ad esempio), se questo dato manca, buttare l'articolo nel cestino più vicino.
- 3) Cercare di capire le caratteristiche salienti della popolazione su cui si è fatto lo studio: non è che qui si parla di 'leggi di natura' comunque valide, qui le condizioni al contorno sono il 90% del gioco.
- 4) Vedere in che termini è stata posta la domanda, ricordo quindici-venti anni fa un articolo su *Science* che diceva di aver trovato 'il gene della follia' considerando 'affetti dalla stessa malattia' uno che si calava i calzoni nella metro, uno che aveva tentato il suicidio, uno che aveva bruciato una casa, ed uno molto ma molto melanconico (H.G. Brunner *et al.* (1993) "Abnormal Behavior Associated with a Point Mutation in the Structural Gene for Monoamine Oxidase A", *Science*, **262**, p.578).

Un po' poco forse per un 'ritratto alternativo' ma spero di essere per lo meno riuscito a dare il senso di un mestiere in cui buon senso (parola odiata dai veri scienziati della caricatura filosofica), utilizzo di strumenti matematici, e intuizione si fondono per costruire un'argomentazione ragionevole, mai definitiva e sempre provvisoria e discutibile ma potenzialmente onesta.

Alcuni autori suggeriscono l'uso della statistica Bayesiana come 'medicina' per risolvere le aporie del ragionamento induttivo. In poche parole, si definiscono Bayesiani tutti quegli approcci che inseriscono 'le credenze a priori' nella definizione delle probabilità' (cfr. F. Lad, su questo sito, sotto "Contributi a SD2").

Secondo me, mentre chiaramente tutti siamo implicitamente Bayesiani, in quanto nessuno fa delle misure a casaccio e nel momento stesso in cui decidiamo di misurare qualcosa 'crediamo' nella rilevanza di ciò che misuriamo, allo stesso modo deve intervenire un 'momento di sospensione' in cui le palline corrono sul biliardo senza che io gli dia delle bottarelle e ulteriori aggiustatine: fare insomma lavorare la natura senza troppe scocciature.

Non è un caso che il grande revival della statistica Bayesiana nelle scienze biomediche abbia coinciso con l'inizio, negli anni 80, del crollo (tuttora perdurante) dei farmaci che superavano i trial clinici come ultima speranza 'per far venire ancora qualcosa' ma questa è un'altra storia. Io credo che 'ricette magiche' che sostituiscano il buon senso, fortunatamente, non ce ne siano.

Inserito: 25 maggio 2005
Scienza e Democrazia/Science and Democracy
www.dipmat.unipg.it/~mamone/sci-dem