

## **Riflessioni**

Le numerose e contrastanti informazioni che sommergono la comunicazione sul Covid-19, e le accese discussioni che ne conseguono, chiamano in causa così pesantemente l'uso di dati e modelli che una riflessione sui compiti della Statistica si rende più che mai necessaria.

Quando ci viene chiesto di intervenire sui problemi reali in ragione della nostra competenza statistica, possiamo appellarci a quella strategia che fa del nostro mestiere una scienza, come è ormai ben consolidato da circa un secolo: stimolati da ciò che accade, ci chiediamo quali siano i meccanismi generatori degli eventi di interesse per inferire su di essi e, quindi, proporre soluzioni in relazione agli obiettivi che si volevano raggiungere. Tale strategia non è banale perché richiede competenze e strumenti analitici in ciascuna sua fase e questa correttezza condiziona ogni tappa, producendo proposte che, in funzione della qualità da cui provengono, possono risultare utili e ragionevoli o, al contrario, superficiali e inverosimili.

Nello specifico, una determinazione seria del dato di cui si discute è prioritaria: ogni informazione è utile in funzione dell'obiettivo. Ad esempio: il numero degli asintomatici può non essere rilevante per la programmazione delle terapie intensive; o ancora, effettuare tamponi a tappeto senza accompagnare gli esiti con una stima di falsi positivi e falsi negativi non aiuta nessun decisore, etc. La diffusione di un dato aggregato ha il vantaggio di una maggiore stabilità e serve ad aggiornare e rasserenare l'opinione pubblica; d'altra parte coloro che operano direttamente con i malati esigono interventi "locali". Allora, obiettivi differenti richiedono informazioni differenti. Sicuramente, andrebbe imposta una maggiore omogeneità nelle rilevazioni, perché la semplice diffusione di informazioni riassuntive (numerosità di morti, contagiati, dimessi, etc.) rappresenta una indicazione poco utilizzabile se queste nomenclature nascondono definizioni differenti (per contenuto, territorio, tempi di raccolta, etc.).

L'informazione che più colpisce deriva dai dati su cui ci si sofferma più spesso nel dibattito, come se fossero la soluzione e non un riflesso di un fenomeno latente. Eppure, il nocciolo del problema statistico si evidenzia quando ci cerca di esaminare, alla luce della nostra disciplina, il fatto "Coronavirus" nel suo manifestarsi e diffondersi: qui, occorre ribadire che non possiamo analizzare alcunché senza interscambio culturale e scientifico con chi ha competenze consolidate in questo ambito, anche se spesso studio e ricerca, in una prima fase, riposano solo su delle analogie. In questo momento storico durante il quale dinamica, persistenza e modalità di diffusione di un contagio sono oggetto costante di studio e di discussione, è poco scientifico dedurre solo dai dati delle valutazioni "come se" conoscessimo i meccanismi di propagazione sia deterministici sia probabilistici.

A tale proposito, la Statistica deve raccontare con molta chiarezza che vi sono due momenti, tuttora avvolti da molta nebbia, da cui deriva un'incertezza conoscitiva che si aggiunge all'ineliminabile incertezza probabilistica.

- i) Caratteristiche legate all'arrivo del virus nell'organismo umano (modalità, tempi, diagnostica, sviluppi, letalità, latenze, ritorni, etc.);
- ii) Aspetti legati alla diffusione del virus nello spazio e nel tempo.

La costruzione, quindi, di “modelli di contagio” è cosa niente affatto univoca, non solo perché –per propria natura- la Statistica non predice con certezza né prevede valori medi come se fossero certezze, ma anche perché qualsiasi affermazione che oggi si deduce da uno “schema astratto” (il modello) è inficiato da una debolezza intrinseca. Questa si aggiunge all’ineliminabile incertezza di chi fa ricerca ritenendo che solo la probabilità possa accompagnare un’informazione seria, l’unica meritevole di essere veicolata nella comunità scientifica e tra i cittadini.

Questa constatazione è complicata dal fatto che è difficile far digerire all’opinione pubblica (fatta da persone comuni e sensibili) che, nella comunicazione dei dati, è un atto di serietà ragionevole ed ineliminabile l’assunzione dell’incertezza. In un Paese dove una fetta della popolazione regala un fiume immenso di denaro, ogni anno, a maghi e ciarlatani di ogni risma e dove, con pari coerenza, per decenni si sono raccontate favole pseudo-scientifiche, sono state proposte ricette economiche e sociali miracolose, e si sono discusse per troppi mesi problematiche di cortissimo raggio (per curare la seconda cifra decimale dei sondaggi di turno invece del bene pubblico), è complicato raccontare oggi l’onesta ignoranza della scienza autentica di fronte ad un fenomeno mai visto e di complessità sconosciuta, aggravata notevolmente dalla circostanza che viviamo su un pianeta interconnesso e mobile, come mai prima d’ora.

Si è circondati da giornalisti ossessionati da “domande a risposta impossibile”, perché sono sbagliate le domande: quando arriva il punto di svolta? quando finisce la pandemia? quando è corretto riaprire? quando si torna come prima? Per noi statistici non esiste soluzione –eticamente valida e razionalmente qualificata- che ribadire che la scienza è un continuo conoscere senza acquisizioni definitive, un cammino faticoso e costoso che conduce l’uomo ad essere capace di fare cose impensabili ma che bisogna sfatare la diffusa consapevolezza dell’onnipotenza tecnologica, perché è una colossale bufala.

Circondati da politici la cui cultura statistica, con rare eccezioni, viene messa in ombra dal desiderio di fornire ricette e soluzioni ad un pubblico “affamato” di certezze e consolazioni a breve termine, occorre in tutte le occasioni (pubbliche e private) raccontare che solo progressivamente –rendendo accurate le informazioni di base e grazie alla certezza quotidiana degli scienziati di laboratorio (che scopriranno sempre più cose su questo virus)- si cominceranno a costruire rappresentazioni un po’ meno rozze di quelle basate su modelli e previsioni deterministiche (cioè matematiche), oggi costruite su informazioni deboli e destinati a fallimenti più o meno clamorosi.

Vi è uno spazio immenso ove la scienza statistica deve esercitare le sue competenze contribuendo al campionamento dei casi da trattare, al tracciamento del contagio, alla mappatura della sua diffusione, all’uso dei dati che si raccolgono in tempo reale, alla presentazione succinta ed efficace dei risultati, all’uso pubblico e generalizzato delle informazioni nel rispetto della privacy, alla logistica, trasporti e produzione, allo scopo di limitare il contagio, e così via.

Ma accanto a ciò (ed oggi ancor prima di ciò), occorre affermare subito –senza girarci attorno, senza metafore, con poche parole semplici ed efficaci- che se le persone vogliono previsioni puntuali e personalizzate l’onestà intellettuale ci impone di raccontare che sono, invece, quelle basate sui valori medi ad essere più affidabili. Perché noi statistici

scommettiamo sugli eventi più probabili e su questi possiamo essere utili al bene comune: difficilmente rispondiamo a soluzioni parziali (questo si chiama azzardo).

Il fascino esercitato dagli “algoritmi” e dall’intelligenza artificiale -sia presso il grande pubblico che nei mass-media e negli apparati decisionali- può condurre ad una rinuncia pericolosa del paradigma della ricerca statistica a favore di una credenza nella macchina tecnologica che sfiora la magia: nell’acquisizione inconsapevole di informazioni massicce (ma alla rinfusa), nella mancata esplicitazione dei processi assunti come generatori dei dati, nelle (sempre implicite e mai dichiarate) assunzioni inferenziali e nei meccanismi finali di classificazione e previsione, discriminazione e profilazione i cui obiettivi sono ben raramente esplicitati.

La lezione che impareremo da questa epocale e troppo lunga vicenda, che drammaticamente coinvolge tutti, riguarderà aspetti ben più profondi e duraturi che la riflessione sulla nostra professione e sulla nostra competenza disciplinare. Può essere l’occasione per qualificarci come quelli che sanno imparare dai dati, riconoscere la propria ignoranza, contribuire con tutti gli altri scienziati al bene comune nella costruzione progressiva e sempre perfezionabile di una comprensione della realtà che non è mai definitiva ma che si consolida anche attraverso gli insuccessi.

Come statistici, possiamo e dobbiamo rafforzare e diffondere la nostra consapevolezza che questa “scienza delle decisioni in condizioni di incertezza” è un formidabile valore aggiunto della conoscenza umana e che le sue potenzialità sono immense se e solo se viene utilizzata in modo proprio: secondo procedure corrette e metodologie ineccepibili perseguite da donne e uomini che hanno imparato, grazie all’evidenza empirica, ad essere sapienti ma, ancor più, molto umili.

*Domenico Piccolo*