

## Il ruolo della statistica nella lettura dei fenomeni socio-economico climatici

ESTà, giugno 2021

In una società complessa e sempre più caratterizzata dall'interazione tra uomo e macchina è inevitabile dover discutere di fenomeni economico-sociali e ambientali utilizzando dati empirici e informazioni raccolti sul campo. È altrettanto inevitabile l'utilizzo di appropriate metodologie di analisi di questi al fine di comprendere il reale contenuto informativo che essi possono fornirci. Nel corso degli ultimi anni, e la pandemia COVID ha dato una ulteriore spinta, sono entrati nel nostro vocabolario giornaliero diversi termini, quali ad esempio *Big Data*, *Intelligenza Artificiale*, *Data Drive*, o *Modelli empirici*. Tutti termini che in qualche misura hanno a che vedere con la disciplina della *Statistica* e dell'*Analisi dei dati*<sup>1</sup>.

Anche la società si è modellata di conseguenza: i telegiornali e i media offrono sempre più servizi su queste tematiche, invitando esperti dei settori o realtà connesse; sono nati molti corsi di laurea e master per formare esperti in analisi dei dati in grado di interpretare i fenomeni che ci circondano e tentare di trasformare i fenomeni in opportunità.

In effetti diventa sempre più difficile ottenere visibilità e credibilità se non ci si trova sulla frontiera di queste tematiche. Anche ESTà ha colto questa opportunità e ha cercato di espandere le proprie competenze interne sviluppando collaborazioni con il mondo accademico e della ricerca per sviluppare metodologie solide e utili per comprendere con occhi *oggettivi*, guidati dai dati, la sempre più intricata relazione tra lo sviluppo economico moderno, l'ambiente, il mondo dell'energia e la società in continua evoluzione.

Ma facciamo un passo indietro. Di cosa stiamo parlando davvero? Con quale realtà ci dobbiamo confrontare? Partiamo da una semplice definizione di Statistica: *la statistica è una disciplina che ha come fine lo studio quantitativo e qualitativo di un particolare fenomeno collettivo in condizioni di incertezza o non determinismo, cioè di non completa conoscenza di esso o di una sua parte*<sup>2</sup>.

Già nella definizione sono racchiusi molti concetti fondamentali che è necessario sviluppare. Innanzitutto il termine *disciplina* indica la Statistica come una materia, un insieme di metodologie e di teorie. Non si tratta di semplici formule o numeri. È un corpus di contenuti che ha una propria filosofia sottostante, delle regole e dei metodi che permettono di interpretare i *fenomeni*, ossia eventi, che ci circondano. I fenomeni sono collettivi, riguardano una *popolazione di interesse* (un gruppo di soggetti o oggetti che si vogliono analizzare), che vengono studiati in un contesto di *incertezza*. Incertezza è tra tutti il vero cuore della definizione, il punto focale. La parola *incertezza* esprime il concetto di ignoranza (in senso positivo chiaramente) su qualcosa, la mancanza di informazioni totali e solo una conoscenza parziale e limitata del fenomeno. Indica che non tutto è controllabile (appunto, non è deterministico) e che, pur facendo il massimo nei nostri sforzi di modellistica, qualcosa sfuggerà sempre al controllo e potrebbe determinare importanti variazioni tra ciò che si afferma e la realtà. *Incertezza* e *Caso*, sono i due concetti che distinguono la Statistica e i suoi adepti, gli *Statistici*, da altre figure professionali e di ricerca. In un certo senso è come dire che per quanto uno si sforzi nel comprendere un fenomeno, un errore (distanza tra realtà e il nostro risultato) dovrà essere sempre

---

1 Facciamo attenzione ad un aspetto davvero importante parlando di dati: due concetti connessi non per forza si sovrappongono. Statistica, Machine Learning, Ingegneria del dato, sono termini tra loro connessi e identificano campi e filosofie complementari, ma al tempo stesso contrastanti. Ognuno di questi campi ha visioni diverse e richiedono competenza diverse, ma devono comunicare per una corretta evoluzione della ricerca.

2 <https://it.wikipedia.org/wiki/Statistica>

tollerato. Il punto è come rendere questo errore minimo, poco influente, senza però avere la pretesa di eliminarlo o ignorarlo.

Già da qui è chiaro che si tratti di un mondo complesso, e a volte oscuro, che richiede una notevole cura e competenza nell'utilizzo degli strumenti che ci mette a disposizione. E a cosa serve davvero tutto questo? Una risposta semplice, ma non banale potrebbe essere quella di voler raccontare una storia. Prima di raccontarla, però, devo comprendere la storia e i dati ci danno una grande mano.

Torniamo al concetto di *Popolazione*, perché è da qui che nascono le parti più interessanti. Con la parola popolazione intendiamo qualunque gruppo di soggetti o oggetti accomunati da qualche caratteristica che li rappresenta e di cui siamo interessati ad ottenere delle informazioni. Con popolazione potremmo intendere ad esempio, l'insieme di tutte le imprese italiane operanti in un determinato settore oppure l'insieme di tutti gli studenti italiani in un dato momento. Oppure il numero di membri di una determinata specie di animali in una foresta o il numero totale di alberi di un'area.

Spesso non è possibile raggiungere tutta la popolazione complessiva. Perché troppo grande (quanti alberi ci sono esattamente in Italia? Difficile a dirsi), perché non si conoscono le 'generalità', i 'nomi' di tutti i membri della popolazione (quante sono esattamente le persone immigrate in Italia nell'ultimo anno? Tra migranti regolari e irregolari non è facile contare). Per questo motivo si utilizza un *campione* (un sottogruppo, sottoinsieme) della popolazione di interesse, lo si studia e analizza e infine si prova a trarre qualche conclusione sulla popolazione complessiva partendo dai risultati campionari. Questo processo di passaggio dai dati campionari a stime sulla popolazione complessiva prende il nome di *Inferenza Statistica*. Ovviamente, dal momento che utilizzo solo parte della popolazione e la inserisco nel campione, commetteremo un errore di approssimazione che è tanto più piccolo tanto più il campione è grande e tanto più esso *rappresenta* la vera natura della popolazione. Facciamo un esempio: se volessi stimare l'altezza media dei cittadini italiani (diciamo circa 60 milioni), dovrei costruire un campione quanto più ampio possibile (diciamo almeno qualche migliaio di persone) e questo campione deve rispettare la distribuzione della popolazione in termini di genere (stessa % di maschi nel campione e nella popolazione), di fascia di età (stessa % di persone tra i 40 e i 50 anni nel campione e nella popolazione) e di area geografica (stessa % di cittadini Lombardi nel campione e nella popolazione). Il campione così ottenuto è *rappresentativo* (rispetta le caratteristiche chiave, le proprietà) della popolazione. L'errore, come anticipato, esisterà sempre e non potrà essere eliminato. Al più possiamo cercare di ridurlo al minimo con degli appropriati metodi. L'incertezza è la chiave di comprensione dell'inferenza.

Partendo dall'inferenza posso poi fare delle *previsioni* nel tempo (previsione per serie storiche), nello spazio (previsione con modelli geo-statistici o spaziali) o su soggetti della popolazione non presenti nel campione (previsione cross-sezionale). Vale a dire, usando informazioni osservate posso ottenere delle stime di valori che sono ancora ignoti nel tempo (il meteo di domani), nello spazio (le concentrazioni di inquinanti in una certa città) o su un particolare soggetto (qual è la ricchezza di un soggetto che ha certe caratteristiche socio-economiche). Ancora una volta, anche la previsione dà luogo ad un *errore di previsione* che va considerato.

Ora, facciamo uno sforzo, e proviamo ad applicare questi concetti nei contesti dello sviluppo economico sostenibile, dell'ambiente, dell'inquinamento o della società. Ciò che EStà sta maturando in questi anni è la convinzione che alcuni fenomeni e trend ambientali ed economici possano essere interpretati e, sotto certe condizioni, previsti. Basti pensare ai cambiamenti climatici: i dati raccontano che le temperature in tutto il globo sono in evoluzione, abbiamo aree del mondo in cui la temperatura si alza e altre in cui si abbassa (lo scioglimento dei ghiacci polari ha questa assurda contraddizione

giudicata senza senso da tanti scettici, ad esempio l'ex presidente degli Stati Uniti, ma che ha solide basi scientifiche); il livello del mare si sta alzando un po' ovunque; le concentrazioni di inquinanti si stanno riducendo in varie aree d'Italia (la Lombardia ad esempio registra ossidi e particolati in forte diminuzione dal 2014 in avanti). Escludendo eventi esogeni, cioè che non dipendono dalla struttura economico-sociale di un sistema economico, anomali come la pandemia da COVID-19, anche lo stato di salute di una economia può essere analizzata e prevista. Ogni anno gli istituti nazionali e internazionali di statistica ci offrono stime sugli andamenti della ricchezza, sia nel tempo che a livello territoriale e le banche centrali tentano di interpretare l'evoluzione dei prezzi. Non bisogna stancarsi nel dire che, pur quando sono fornite da enti certificati e accreditati (come può essere ISTAT o Banca d'Italia), le stime possono essere errate proprio in virtù di quell'errore ci sempre accompagna gli statistici.

EStà sta impegnando molte risorse e tempo nello studio delle metodologie più appropriate per sviluppare modelli e analisi di qualità che permettano di indirizzare gli interlocutori verso scelte ragionate e basate quanto più possibile su criteri oggettivi. Fermo restando che non sempre sia possibile fornire una indicazione esatta, ma quanto meno sensata e solida nei fondamenti.