

INTRODUZIONE ALLA STATISTICA

La *Statistica*, seppur basata su regole matematiche, quindi teoriche, trae le sue fonti e le sue principali spinte evolutive, non da ricerche dogmatiche, bensì da una stretta connessione con l'attività pratica: essendo, infatti, una scienza onnipresente nella ricerca scientifica e nelle analisi sperimentali, apporta un contributo significativo alle vicende quotidiane nel prendere decisioni razionali e coerenti.

Si tratta di una disciplina relativamente giovane, spesso confusa con le *statistiche* (dati, tabelle, grafici, indici, medie, ...): presente in tutte le scienze, infatti, non è altro che uno strumento essenziale per la scoperta di leggi e relazioni tra fenomeni. Occupandosi, in particolare, della raccolta e dell'analisi di dati, nonché dell'interpretazione dei risultati osservati, rappresenta, in tutte le situazioni nelle quali occorre assumere decisioni in condizioni di incertezza, un momento importante ed ineliminabile per la ricerca scientifica.

La parola *Statistica* deriva dal vocabolo italiano *Stato* e fa riferimento, nella quasi totalità dei linguaggi europei, alla constatazione per cui le prime informazioni su fenomeni reali sono state raccolte ed organizzate ad opera degli organismi statali che ne sono stati anche i principali utilizzatori. Sembrerebbe che la parola *Statistica*, intesa come raccolta di informazioni organizzate e gestite dallo Stato, sia stata utilizzata per la prima volta dall'italiano Girolamo Ghislini che, nel 1589, ha definito la Statistica come “descrizione delle qualità che caratterizzano e degli elementi che compongono uno Stato” e, tra il 1666 ed il 1668, ha dato alle stampe il *Ristretto della civile, politica, statistica e militare scienza*.

Tuttavia l'evoluzione storica della disciplina in esame nasconde ben due anime: la prima connaturata nell'*Homo Sapiens* che prende coscienza del mondo e dei suoi simili e che, nella lotta per la sopravvivenza, elabora comportamenti ottimali per nutrirsi, difendersi e riprodursi, strategie queste che, se da un lato, alternano sconfitte e successi, dall'altro, proprio grazie ad approssimazioni successive, consentono il diffondersi di acquisizioni e convinzioni, usi e costumi e, alla fine, di conoscenza, scienza e cultura. Quando poi tali convinzioni si sono codificate nella vita del clan, della tribù, di un popolo o di una nazione, ecco che nascono le regole di comportamento, ovvero la legge, cosicché l'analisi statistica diventa di interesse pubblico al punto tale che le nazioni, a mano a mano, iniziano a dotarsi di Istituti centrali di statistica, deputati per legge alla raccolta, organizzazione e diffusione di dati sulla popolazione, sulle abitazioni, sulle risorse economiche e su tutti gli aspetti rilevanti della vita di una nazione, di una comunità di stati (Unione Europea) o dell'intero pianeta (Nazioni Unite).

La seconda anima della statistica nasce da una constatazione differente: di fronte alla realtà che muta vi sono risultati che meritano più fiducia di altri perché si ripetono con maggiore regolarità, cosicché la mente umana registra regolarità senza certezze, convinzioni non sicurissime, ripetizioni di eventi non sempre garantiti da un esito univoco. Se, però, ciò genera, da un lato, paura ed impone cautele contro i rischi, dall'altro invita al gioco e alla scommessa (di qui il concetto di *probabilità* che, pur essendo ben presente nella storia e nella cultura sin dai primordi della civiltà, diventa importante e ben formalizzato solo a partire dal secolo XVIII, ovvero da quando sorge una nuova disciplina che sarà poi denominata *calcolo delle probabilità*; bisognerà attendere, però, l'inizio del '900 perché diventi e si diffonda lo studio dell'incertezza come fattore immanente nella scienza, nella tecnica e nelle attività umane, cosa questa che è avvenuta in primo luogo nelle scienze biologiche, fisiche e mediche, e poi si è trasferita nelle scienze umane, sociali, economiche e politiche).

La saldatura tra le due anime della statistica è avvenuta solo quando, di fronte alla natura sempre più sperimentale della conoscenza, ci si è posti il problema della validità delle ipotesi.

Oggi, gli organismi pubblici che solitamente raccolgono e diffondono informazioni statistiche sono innumerevoli ed agiscono secondo una gerarchia di competenze che individua nell'Ente locale la sede prioritaria di raccolta del dato elementare; la verifica, l'aggregazione e la pubblicazione, invece, sono di competenza dell'Ente centrale. In Italia tale Ente è l'ISTAT (Istituto Nazionale di Statistica) che dipende dalla Presidenza del Consiglio dei Ministri ed accentra istituzionalmente tutta la pubblicazione statistica nazionale, provvedendo a curare e diffondere il materiale raccolto in pubblicazioni a periodicità mensile, annuale, decennale o anche occasionale. Le rilevazioni contenute in tali pubblicazioni possono poi essere a carattere generale o relative a specifici campi di attività. Accanto all'Annuario Statistico Italiano e ai Bollettini mensili, a carattere generale, vi sono, sia gli annuari dedicati alle nascite e ai decessi, sia le pubblicazioni in campo sanitario e familiare, sia le statistiche della Pubblica Amministrazione, le statistiche giudiziarie, civili e penali. L'ISTAT, però, non è l'unica organizzazione che cura e gestisce l'ingente mole delle fonti statistiche nazionali: ogni anno, infatti, il Parlamento pubblica la *Relazione sulla situazione economica del Paese*, il Ministero del tesoro cura il *Bilancio di previsione dello Stato* e la Banca d'Italia pubblica il *Bollettino statistico*; vi sono poi tante altre pubblicazioni effettuate dalle Camere di Commercio, Confindustria, Enel, Rai e Ferrovie dello Stato. A livello internazionale, meritano un cenno le statistiche dell'ONU, del BIT, della OMS, dell'UNESCO e della FAO. Va citato, infine, l'Istituto Statistico delle comunità europee, l'EUROSTAT, che produce statistiche relative ai paesi aderenti all'unione.

La Statistica, pertanto, nell'applicare la sua metodologia ai problemi reali, utilizza alcuni criteri logico-concettuali di riferimento, i cosiddetti *paradigmi*.

Il paradigma statistico di base è quello della *sintesi delle informazioni*, in quanto di fronte all'analisi di fenomeni complessi si cerca un indicatore riassuntivo, utile per qualche scopo predefinito e si attribuisce a quest'unica misura il significato di "rappresentanza".

Un secondo paradigma statistico è la *scoperta del nuovo*, intendendo con ciò la messa in evidenza di fatti, congetture problematiche, connessioni, leggi che non erano note in precedenza o che andavano verificate.

Un terzo paradigma è il contenimento dell'errore, in quanto la connessione tra le informazioni possedute in ingresso, *input*, e quelle osservate in uscita, *output*, è alterata da una componente erratica, di natura aleatoria e quindi imprevedibile.

Un quarto paradigma essenziale della ricerca statistica, infine, è la *dialettica della conoscenza*, cioè l'accettazione dei risultati, la loro critica, la proposta di ipotesi alternative e lo studio di nuovi risultati per proporre, rifiutare e migliorare precedenti acquisizioni.

Ogni analisi empirica, quindi, si articola secondo il seguente schema:

obiettivi – informazioni – metodi statistici – risultati

dove per *obiettivo* si intende, sia la natura delle informazioni da raccogliere, sia gli strumenti mediante i quali esaminare tali informazioni.

La metodologia statistica, dunque, opera mediante stadi successivi:

- predefinire obiettivi delimitati ed agevoli da investigare sul piano empirico;
- individuare le informazioni necessarie per una puntuale conoscenza del fenomeno, oggetto di studio;
- applicare metodi adeguati per il trattamento delle informazioni;
- esaminare criticamente i risultati della metodologia statistica per saggiarne la coerenza, la validità e la generalizzabilità.

Risulta ora opportuno esaminare in dettaglio le *fasi di un'indagine statistica*, fondamentali per consentire alle informazioni raccolte di essere analizzate in modo appropriato con metodi statistici:

- 1) definizione degli obiettivi da raggiungere nella ricerca, obiettivi che devono essere molto particolareggiati, al fine di individuare con cura le informazioni da ricercare, evidenziando in particolare le variabili statistiche da osservare;
- 2) rilevazione dei dati: può essere *completa*, quando si esaminano tutti gli elementi oggetto di studio, o *parziale*, quando ci si limita a studiare un sottoinsieme, detto *campione*, dell'insieme di riferimento;

- 3) elaborazione metodologica dei dati: in questa fase si applicano schemi formali di tipo logico-deduttivo o inferenziale-induttivo, di natura matematica (tabelle o grafici), ovvero gli strumenti propri dell'analisi statistica;
- 4) presentazione ed interpretazione dei risultati: una presentazione poco efficace può vanificare sia l'utilità dei risultati raggiunti che le potenzialità del metodo utilizzato; un'accurata illustrazione dei risultati ed una disamina particolareggiata delle implicazioni operative dei medesimi, invece, sono elementi decisivi per il buon esito di un'indagine statistica;
- 5) utilizzazione dei risultati della ricerca: l'uso che può essere fatto dei risultati di una ricerca deriva anche dalle sue modalità di esecuzione e dai dettagli mediante i quali è stata condotta la misurazione o l'intervista; la statistica, del resto, non è una scienza fine a se stessa ma trova applicazione in diversi campi.

Lo sviluppo della Statistica, quindi, fa riferimento al processo conoscitivo dell'Uomo nel suo rapportarsi alla realtà esterna e si concretizza in tappe di apprendimento, funzionali all'obiettivo che intende raggiungere.

La prima percezione avviene mediante il senso della *vista* intesa come cattura della realtà, allo scopo di confrontare e sintetizzare fenomeni, eventi, problemi nella loro individualità e/o nella loro complessità. In tal senso la *Statistica* è la scienza che aiuta a vedere il mondo, a fotografarlo, a catalogarlo e classificarlo.

L'atto successivo al vedere è la riflessione e la comprensione, cioè *capire* la realtà all'interno di un'impostazione probabilistica nella quale l'esistente è esaminato in rapporto a ciò che poteva accadere o che verosimilmente accadrà.

Infine, senza l'azione l'Uomo non sopravvive né migliora; esistono, quindi, degli obiettivi da perseguire, progetti da realizzare, cose da fare: questa fase, detta dell'*agire*, serve a raggiungere scopi predefiniti e consentire la fruibilità del mondo, della società, delle risorse e delle opportunità che si offrono.

La descrizione (il *vedere*) e la comprensione (il *capire*) orientate verso l'azione concreta e l'obiettivo specifico (l'*agire*), dunque, producono sempre un modello statistico che svolge un ruolo strumentale, circoscritto, limitato ma utilissimo ed ineliminabile. Per questo, anche quando non specificato, la modellistica dei fenomeni reali è il momento finale della ricerca statistica proiettata verso le applicazioni.

La *Statistica*, dunque, è la *Scienza delle decisioni in condizioni di incertezza*, avente quali obiettivi principali il *sintetizzare*, ovvero predisporre i dati raccolti in una forma che consenta di comprendere meglio, e quindi di semplificare, i fenomeni rispetto ai quali è stata eseguita la

rilevazione, ed il *generalizzare*, ovvero estendere il risultato dell'analisi effettuata sui dati di un gruppo limitato di unità statistiche all'intera collettività di appartenenza.

All'interno della disciplina metodologica, inoltre, si possono distinguere due diverse correnti: la statistica *descrittiva* e la statistica *inferenziale*.

Con il termine di ***Statistica Descrittiva*** si intende un insieme di tecniche e strumenti finalizzati ad assolvere uno dei principali compiti assegnati alla Statistica, ovvero descrivere, rappresentare e sintetizzare in maniera opportuna un insieme o campione di dati relativamente ad un problema; tale branca, che ha come obiettivo quello di organizzare, riassumere e presentare i dati in modo ordinato attraverso strumenti di tipo sia grafico che numerico, si occupa di fotografare una data situazione e di sintetizzarne le caratteristiche salienti, ovvero di descrivere ciò che si osserva o ciò che i dati evidenziano nei loro tratti essenziali.

La ***Statistica Inferenziale*** o ***Inferenza Statistica***, invece, comprende le tecniche matematiche per quantificare il processo di apprendimento tramite l'esperienza; utilizza i dati statistici, anche opportunamente sintetizzati dalla statistica descrittiva, per fare previsioni di tipo probabilistico su situazioni future o comunque incerte; con la statistica inferenziale, quindi, si cerca di raggiungere conclusioni che si estendono oltre i dati raccolti nel loro immediato e che possono essere valide e riferibili ad un contesto più ampio rispetto a quello dei dati di quel singolo esperimento.

Dunque, mentre la statistica descrittiva si occupa di rappresentare l'informazione contenuta in un dato insieme o campione di dati, la statistica inferenziale utilizza tale informazione per fare delle affermazioni più generali riguardanti i parametri della popolazione da cui il campione è stato estratto. La statistica inferenziale, servendosi della teoria del calcolo della probabilità, pertanto, si occupa di risolvere il cosiddetto problema inverso, ovvero, sulla base delle osservazioni su un campione di unità selezionate con date procedure dalla popolazione, perviene a soluzioni valide entro dati livelli di probabilità e quindi di applicabilità anche per la popolazione stessa.