

La Statistica

Marco Martini

1 Origini della Statistica

La Statistica è una scienza relativamente giovane che in tre secoli ha avuto un enorme sviluppo. Il suo nome deriva da “Stato”: statistici erano coloro che discutevano delle cose relative al governo dello Stato sulla base di “numeri, pesi e misure” secondo l’impostazione suggerita da William Petty (1676) nella sua *Political Arithmetick*¹:

The Method I take to do this, is not yet very usual; for instead of using only comparative and superlative Words, and intellectual Arguments, I have taken the course (as a Specimen of the Political Arithmetick I have long aimed at) to express my self in Terms of Number, Weight, or Measure; to use only Arguments of Sense, and to consider only such Causes, as have visible Foundations in Nature; leaving those that depend upon the mutable Minds, Opinions, Appetites, and Passions of particular Men, to the Consideration of others².

Compito degli statistici del XVII e del XVIII secolo era dunque quello di raccogliere metodicamente i dati relativi a popolazioni numerose, alle loro ricchezze, ai loro comportamenti, per sintetizzarli in informazioni utili al fine di conoscere e, quindi, di assumere decisioni razionali.

La procedura più tradizionale per ottenere informazioni globali su insiemi di individui, utilizzata fin dall’antichità, era il loro conteggio che permetteva di rappresentare l’insieme stesso con un numero naturale. Uno dei più antichi libri della Bibbia, che si intitola, appunto, *Numeri*, contiene, tra l’altro la “registrazione di tutti i figli di Israele secondo le loro famiglie, dai vent’anni in su, cioè di tutti coloro che erano abili alla guerra”

¹William Petty, *Political Arithmetick*, Londra, stampato postumo nel 1690.

²Il metodo da me seguito non è quello usuale; invece di usare parole al comparativo o al superlativo e argomenti intellettuali ho preferito esprimermi mediante numeri, pesi e misure; usare argomentazioni deducibili dall’esperienza sensibile, considerare soltanto quelle cause aventi fundamenta visibili nella natura, lasciando ad altri quelle che dipendono dalle mutevoli idee, opinioni, ambizioni e passioni di singoli uomini.

(Numeri, 1, 19-47). A Roma, ogni cinque anni il *censor* faceva il *census* (censimento solenne) di tutti i cittadini, della loro attitudine a contribuire agli aggravii della *res publica* e dei loro costumi.

Gli statistici, partendo da queste tecniche antichissime e avvalendosi dei progressi della matematica affinarono via via le procedure volte a migliorare le informazioni numeriche sulle popolazioni e sui diversi fenomeni che le riguardavano, in vista delle decisioni, del controllo o del confronto con altri Stati.

Le sintesi da essi operate, anche se comportavano sempre una perdita di informazioni rispetto alla ricchezza dei dati individuali, si dimostrarono strumenti potenti per la conoscenza e il governo delle popolazioni.

Per questo motivo nei secoli XIX e XX l'impiego delle procedure statistiche si estese progressivamente allo studio di altri insiemi numerosi di elementi empirici dei quali poteva essere utile sintetizzare caratteristiche e comportamenti ai quali si dette il nome di *popolazioni* o di *collettivi* : da quello delle molecole che compongono un gas a quello delle particelle elementari della materia; da quello delle misure ripetute di grandezze fisiche o chimiche a quello dei risultati di numerosi esperimenti in campo medico o farmaceutico; da quello dei prodotti realizzati in serie a quello dei potenziali clienti delle imprese o elettori delle democrazie; da quello dei prezzi dei beni di consumo a quello dei lavoratori occupati e disoccupati e così via.

Poiché la nostra conoscenza del mondo procede per "universali", cioè per categorie che si predicano a pluralità di enti, la Statistica si dimostrò efficace in ogni area del sapere che riguardasse "collettivi" di oggetti empirici, trovando in ciascuna di esse stimoli e spunti per nuovi affinamenti delle procedure.

D'altra parte la pluralità dei campi applicativi ha sollecitato la precisazione dell'oggetto formale della disciplina, prescindendo dai diversi oggetti sostanziali: è nata così la fondamentale bipartizione della statistica odierna in due gruppi di discipline: le Statistiche metodologiche e le Statistiche applicate.

Le prime (che potrebbero anche chiamarsi discipline Statistiche formali) studiano le procedure statistiche nel loro aspetto formale con l'intento di precisarne e generalizzarne i concetti, il linguaggio e le tecniche, indipendentemente dallo specifico campo di applicazione.

Le seconde (che potrebbero anche chiamarsi Statistiche sostanziali) ricercano le procedure più adatte per produrre informazioni numeriche di sintesi atte a rispondere ai sempre nuovi problemi posti nelle specifiche area del sapere economico, sociale demografico, medico, fisico chimico, biologico ecc.

La dinamica fecondità della Statistica si fonda proprio sul confronto continuo tra i due approcci: il primo più orientato ad evidenziare aspetti generali del linguaggio statistico indipendentemente dai fini specifici; il secondo teso a trovare precise risposte a specifici problemi adattando e creando nuovi strumenti.

2 L'oggetto formale della Statistica

L'oggetto formale della Statistica è, come si è detto, ogni insieme di elementi empirici intorno ai quali si intendono raccogliere dati per sintetizzarli in informazioni numeriche atte ad assumere decisioni razionali in campo conoscitivo od operativo. Si consideri, ad esempio, il caso dell'insieme dei visitatori di un museo, che i responsabili del museo stesso desiderano studiare per meglio organizzare gli orari, il personale e l'offerta di materiali illustrativi.

L'insieme di elementi (i visitatori di un museo) per ragioni storiche viene denominato *popolazione* o *universo*; ogni singolo elemento (ogni visitatore) *unità statistica*.

Il *dato statistico* elementare consiste in una proposizione che associa alla singola unità statistica un predicato esprimente un attributo, una relazione o una azione: ad esempio il visitatore X è di nazionalità francese, ha l'età di vent'anni (attributi), è entrato nel museo domenica mattina (azione), con una comitiva di quindici persone (relazione).

Ogni predicato indica la modalità assunta da un *carattere* (ad esempio il carattere 'nazionalità' può assumere modalità: italiano, francese, tedesco ecc.; il carattere 'età' le modalità coincidenti con i numeri naturali, ecc.). Si osservi che un carattere è *statistico* non in sé stesso, ma in quanto è stato rilevato su diverse unità statistiche mediante una *rilevazione statistica*. Questa è l'operazione con la quale si è assegnata ad ogni unità statistica la modalità empirica su di essa assunta da ogni carattere considerato e può essere eseguita con strumenti diversi quali l'intervista verbale, il

questionario da compilare, l'uso di contatori meccanici, il telerilevamento, ecc.

Il *registro statistico* è il documento, prodotto su supporto cartaceo, magnetico, o di altro materiale, nel quale si sono riportate le sequenze dei segni predeterminati (codici numerici, alfabetici, alfa numerici, grafici ecc.), univocamente corrispondenti ad ogni unità rilevata e ad ogni modalità su di essa assunta da ogni carattere considerato.

3 Le statistiche

Mentre i problemi tecnici relativi al come effettuare la rilevazione, la trasmissione e la memorizzazione in un registro sono oggetto di studio delle scienze dell'organizzazione e dell'Informatica, la Statistica studia le procedure razionali mediante le quali trasformare i dati in informazioni sintetiche esprimibili con numeri: le *statistiche* utili per conoscere e per decidere.

Un insieme di dati si trasforma in informazioni statistiche per rispondere a precise domande. Sono le domande che rendono informativi i dati: se, infatti nessuno ponesse domande le operazioni di rilevazione e di registrazione si ridurrebbero (come spesso avviene) a noiosi quanto inutili adempimenti burocratici.

Sotto questo aspetto il compito dello statistico è duplice: da un lato, egli deve sapere ben comprendere le esigenze dell'operatore (nel caso ipotizzato quelle del responsabile del museo che si chiede, ad esempio, quanti fogli illustrativi predisporre e in quali lingue, che orari di apertura stabilire o come distribuire il personale nei diversi giorni e nelle diverse fasce orarie); dall'altro, egli deve tradurre le esigenze dell'operatore in domande espresse nel linguaggio della statistica, in modo che ad esse si possa rispondere con informazioni numeriche desumibili dai dati rilevati (o rilevabili). Nel nostro caso, le richieste del responsabile del museo possono dare luogo a domande semplici, quali ad esempio: quanti sono in media i visitatori delle diverse nazionalità? in quali giorni e in quali ore del giorno vi è il maggiore afflusso di visitatori? oppure più complesse, quali: vi sono modelli stabili di comportamento dei visitatori, quanto agli orari o alla durata delle visite? si possono individuare e quantificare gruppi omogenei di visitatori che presentano esigenze simili tra di loro ma diverse da quelle di altri gruppi?

In generale le domande e quindi le statistiche che ad esse rispondono possono riguardare sia i caratteri statistici, considerati singolarmente (quali sono le età più frequenti dei visitatori?) oppure congiuntamente (che relazione sussiste tra l'età e gli orari di visita?), sia le relazioni di somiglianza o dissomiglianza delle unità statistiche dal punto di vista dei caratteri considerati (i visitatori del mattino sono simili a quelli del pomeriggio?).

Per rispondere anche alle domande più semplici lo statistico deve elaborare le statistiche più opportune, cioè quelle informazioni sintetiche espresse in termini numerici che rendano minima la perdita, opportunamente misurata, rispetto alla conoscenza dei dati individuali. Quando tale perdita è nulla la statistica è detta *sufficiente*.

È opportuno richiamare che con gli stessi dati si possono costruire tante statistiche quante sono le domande suscettibili di ricevere una risposta numerica. Il pregiudizio secondo il quale la Statistica mistificherebbe la realtà, divulgato dalla famosa poesia di Trilussa e ancora oggi molto diffuso, deriva dalla errata convinzione che essa fornisca formule (la media aritmetica!) valide per ogni circostanza. Al contrario una statistica sufficiente per uno scopo non lo è per un altro: sapere che ogni abitante del quartiere consuma *in media* un pollo alla settimana è sufficiente per decidere l'approvvigionamento del pollivendolo (purché sia nota anche la statistica degli abitanti del quartiere), ma non per il medico che desidera conoscere la dieta dei suoi pazienti: per questa sono molto più interessanti altre statistiche: per esempio quelle che permettono di confrontare quanti mangiano due polli con quanti non ne mangiano nessuno!

Nel corso della rilevazione e della registrazione è inevitabile commettere errori (di misura, di omissione, di duplicazione, di interpretazione, di trascrizione ecc.) che, solitamente, sono tanto più frequenti quanto più elevato è il numero di unità statistiche e dei caratteri che entrano in gioco. Inoltre molto raramente le rilevazioni riguardanti popolazioni numerose sono continue nel tempo.

Per motivi di costo ed organizzativi esse vengono più spesso effettuate periodicamente. Di norma, tra una rilevazione e la successiva intercorre un lasso di tempo più esteso se le rilevazioni sono totali (censuarie), cioè compiute su tutte le unità statistiche, meno esteso se le rilevazioni sono parziali (campionarie), cioè effettuate su una parte (campione) delle unità stesse. Le rilevazioni su particolari fenomeni poi, possono essere solo

occasionali e parziali. Per questi motivi, in pratica quasi sempre le statistiche devono essere costruite a partire da dati che sono incompleti a causa sia degli errori commessi nel corso della rilevazione o della registrazione, sia degli errori dovuti alla discontinuità e/o alla parzialità delle rilevazioni stesse.

Mentre la disciplina riguardante le statistiche desumibili da dati completi viene solitamente chiamata *Statistica descrittiva*, quella che prende le mosse da dati incompleti viene denominata *Statistica induttiva o inferenziale*: questa deve infatti indurre o inferire informazioni riguardanti l'intera popolazione tenendo conto degli errori sopra ricordati.

In questo compito è stata di grande aiuto alla Statistica il *Calcolo delle probabilità*, cioè la disciplina che “realizza l'utilizzazione della matematica per raggiungere il massimo di certezza e razionalità in condizioni di informazione incompleta” (Carlo Felice Manara, 1993)³. Utilizzando i concetti matematici di *probabilità* e di *variabile casuale* la Statistica induttiva o inferenziale ha via via messo a punto criteri rigorosi che consentono di assumere decisioni razionali, in campo conoscitivo od operativo, nel trattamento degli errori nel controllo dei dati, nella stima dei dati mancanti e, soprattutto, nell'stime costruite a partire da dati campionari.

La tecnica delle indagini campionarie che studia i metodi più opportuni per scegliere i campioni consente oggi di ottenere, con notevoli risparmi nei costi, risultati informativi spesso qualitativamente più validi di quelli derivanti dalle indagini censuarie. Purtroppo, a causa dell'analfabetismo statistico ancora diffuso, non sempre le precise condizioni imposte dalla Statistica sono rigorosamente rispettate: molti cosiddetti sondaggi d'opinione presentati su giornali, riviste e trasmissioni televisive come indagini campionarie possono servire forse a “fare notizia”, cioè a suscitare emozioni e reazioni, ma non certo a “fare statistica”!

4 Le statistiche applicate

Come si è già avuto modo di osservare, le tecniche statistiche di costruzione di informazioni sintetiche atte a rispondere ai più svariati scopi conoscitivi od operativi hanno trovato applicazione in tutti i campi conoscitivi legati all'osservazione di *collettivi empirici*. Nel corso di laurea

³Purtroppo non si è stati in grado di identificare il volume cui l'autore si sta riferendo.

in Scienze Politiche, in particolare, interessa studiare la loro applicazione in campo economico e sociale: per questo motivo oltre ad insegnamenti di statistica formale variamente denominati, si impartiscono anche insegnamenti di statistica sostanziale quali: Demografia (statistica della popolazione), Statistica sociale, Statistica economica ed Econometria.

Se negli insegnamenti di statistica formale il linguaggio statistico è presentato in sé stesso con una impostazione generale suscettibile di qualsiasi impiego, nelle statistiche applicate invece esso è considerato, per così dire, in atto, cioè nella sua concreta applicazione a popolazioni ed a caratteri economici e sociali.

I collettivi considerati sono quelli dei soggetti (persone fisiche, imprese, o istituzioni) o degli oggetti (beni reali o finanziari) appartenenti a un dato sistema socioeconomico.

Le popolazioni possono essere guardate da un punto di vista statico, ossia come stock fotografati in un determinato istante temporale, di cui si considerano caratteri-attributi (come età, sesso, titolo di studio e residenza per i soggetti; quantità prezzo e valore per gli oggetti) e caratteri-relazioni (come la parentela tra soggetti o il possesso dei soggetti sugli oggetti).

Le stesse popolazioni possono essere osservate da un punto di vista dinamico, come flussi filmati in un certo intervallo temporale, considerando i caratteri-azioni che modificano nel tempo la consistenza e la struttura degli attributi e delle relazioni delle popolazioni medesime: azioni intransitive (come la nascita la morte o la mobilità) o transitive (come la produzione, il consumo o lo scambio di beni).

I bilanci tra stock e flussi stabiliscono relazioni contabili che devono essere comunque rispettate e che possono essere espressi in *conti* (contabilità macroeconomica, conti demografici, sociali, ambientali ecc.) che, a seconda del sistema considerato, possono essere internazionali, nazionali, regionali, aziendali e così via.

I dati statistici necessari per la costruzione e il continuo aggiornamento di tali conti derivano da diverse rilevazioni e registrazioni compiute da vari enti che fanno parte del Sistema Statistico nazionale e coordinate dall'Istituto Nazionale di Statistica (ISTAT):

- ▷ continue e totali, come quelle delle anagrafi delle persone, delle imprese o dei beni immobili;
- ▷ periodiche e totali, come i censimenti decennali della popolazio-

ne, delle abitazioni e delle attività produttive, i telerilevamenti del territorio, le registrazioni amministrative annuali dei redditi o delle retribuzioni e mensili del commercio con l'estero;

- ▷ periodiche e parziali, come l'indagine annuale multi scopo sulle famiglie, le indagini trimestrali sulle forze di lavoro o sul valore aggiunto, le indagini mensili sui prezzi al consumo o sulla produzione industriale;
- ▷ occasionali e parziali, come molte indagini di mercato o quelle su particolari fenomeni sociali o economici.

La necessità di integrare dati provenienti da più fonti e le esigenze dei confronti tra diversi paesi hanno posto due esigenze: la prima è quella di rendere uniformi le definizioni delle unità, dei caratteri, delle rispettive modalità (nomenclature e classificazioni) nonché dei criteri di registrazione (codificazioni); la seconda è quella di mettere a punto procedure razionali per misurare adeguatamente gli errori di ogni fonte, controllare la qualità dei dati, ed integrare dati provenienti da fonti diverse.

Partendo dalla conoscenza critica dei dati la Statistica applicata costruisce indicatori sintetici sulla struttura e sulla dinamica dei sistemi economici, demografici o sociali, adatti per conoscere, per prevedere e per prendere decisioni, come, ad esempio: l'indice generale dei prezzi al consumo, elaborato in vista delle decisioni di politica economica o dell'adeguamento dei contratti all'andamento del costo della vita; i tassi di natalità, di mortalità o di fecondità, costruiti in vista delle previsioni dello sviluppo della popolazione e delle decisioni di politica sociale, scolastica, sanitaria o previdenziale; il tasso di disoccupazione, calcolato in vista delle decisioni di politica dell'occupazione o del lavoro e così via.

Lo studio e l'interpretazione del funzionamento dei sistemi economici e sociali viene spesso realizzato con l'ausilio di modelli teorici nei quali l'andamento di alcune variabili viene messo in relazione con quello di altre. Si pensi, ad esempio ai modelli che spiegano i consumi in funzione del reddito e dei prezzi, o la produzione in funzione del lavoro e del capitale, oppure ai modelli econometrici formati da centinaia di equazioni come quelli elaborati dalla Banca Mondiale, dal Fondo Monetario Internazionale o dalla Banca d'Italia. Tali modelli, dai più semplici ai più complessi, derivano da teorie economiche, demografiche o sociologiche e, per potere

essere verificati alla luce dell'evidenza empirica, richiedono l'intervento della statistica applicata: in particolare, nella misura delle variabili, nella specificazione della forma delle equazioni matematiche che interpretano i legami tra le variabili, nella fissazione e applicazione di criteri razionali per indurre dai dati empirici informazioni sui parametri di tali equazioni e, infine, nella valutazione della corrispondenza dei modelli alla realtà.

5 La Statistica e le altre scienze

Da quanto detto appare chiaro il nesso tra la Statistica e le altre scienze formali (Matematica, Logica e Informatica) e sostanziali (Economia e Sociologia).

Nel suo aspetto formale la Statistica si avvale di concetti elaborati dalla Matematica, come quelli di insieme, di numero, di funzione, di limite, di probabilità, o dalla Logica, come quelli di proposizione non analizzata, di valore di verità o di connettivi e, in generale, condivide con le suddette due scienze il rigore e la precisione del linguaggio, dell'astrazione, della concettualizzazione e della dimostrazione.

Ciò che la differenzia da queste discipline è però il suo oggetto, costituito da popolazioni empiriche mentre quello della Matematica consiste in entità ideali e quello della Logica nelle regole del ben ragionare.

L'Informatica, che insegna a gestire l'informazione e a comandare gli strumenti materiali ed elettronici che permettono una elaborazione rapida, efficace ed economica dell'informazione stessa, rappresenta un supporto oggi indispensabile per le rilevazioni, le registrazioni e le elaborazioni di cui la Statistica si serve: in particolare, l'Informatica consente oggi alla Statistica di spingere le proprie procedure di sintesi e di decisione molto al di là dei limiti che fino a qualche tempo fa erano resi invalicabili dall'onerosità delle operazioni. È noto inoltre che i progressi della tecnica mettono oggi a nostra disposizione strumenti molto più potenti di quelli di puro calcolo numerico: in particolare *pacchetti statistici* e *sistemi esperti* sono in grado di sostituirsi all'uomo nella soluzione di alcuni problemi statistici predifiniti.

Proprio la presenza di tali strumenti rende ancor più necessaria la riflessione critica sul loro impiego, in modo da evitare la nociva pigrizia e la pericolosa dipendenza mentale che un loro impiego acritico può indurre: se ben intesa l'Informatica può favorire la presentazione da parte dell'in-

segnante alla assunzione da parte dello studente di una attitudine statistica critica e creativa.

Tale abito mentale può emergere in modo ancor più evidente nelle Statistiche applicate se le definizioni standardizzate, le fonti di dati, gli indicatori sintetici e le procedure econometriche, non sono presentate e assimilate come una dottrina cristallizzata, ma piuttosto come spunto per una formazione ad uno spirito critico e creativo nei confronti dei sempre nuovi problemi posti dal rapido evolversi della realtà economico sociale.

Le Statistiche applicate sono, da questo punto di vista, il luogo privilegiato del dialogo sulla realtà che cambia tra chi si esprime nel linguaggio delle scienze economico sociali e chi si esprime nel linguaggio delle procedure rigorose che conducono alle sintesi numeriche a partire dall'osservazione della realtà stessa.

Ma ogni dialogo, per essere produttivo, richiede una motivazione, un'attitudine all'ascolto e un mondo di riferimento comune agli interlocutori: ciò significa che la Statistica non può essere applicata alla realtà se lo statistico non possiede una solida cultura economica e sociale e, soprattutto, se non ha educato e sviluppato la sua attenzione e la sua responsabilità per i problemi delle persone che compongono i collettivi socioeconomici oggetto del suo studio.