

La statistica e il pensiero matematico

1. La lettura del recentissimo manuale « Statistica demografica ed economica » di P. LUZZATTO-FEGIZ ⁽¹⁾ mi ha dato continuamente l'impressione che la materia sia svolta in modo da interessare chi legga con mentalità matematica, da un lato agli argomenti trattati, e nello stesso tempo ai modi di ragionamento, di cui è tratto volta a volta a considerare le concordanze, le analogie e le peculiarità in rapporto a quelli che gli sono familiari dalla matematica e dalle scienze fisiche, imperniati sul rigore deduttivo e l'analisi critica delle premesse. Sono stato indotto così a cercar di chiarire tali considerazioni e reazioni, in primo luogo a me stesso, ed ora nel presente articolo, che non pretende di essere tuttavia che poco più di una recensione integrata da divagazioni personali.

L'aspetto primo e più concreto della statistica è quello di una descrizione dei fatti mediante dati quantitativi opportunamente raccolti, e tale aspetto presenta naturalmente un interesse tanto più grande quanto maggiore è l'interesse che destano i fatti investigati. E a sua volta l'interesse che desta in noi un argomento è tanto più vivo quanto meglio esso s'inquadra nella visione che abbiamo del più ampio campo di cui fa parte e quanto più numerosi e profondi ci si svelano i legami che ve lo inseriscono rendendo interdipendenti concezioni e problemi in un raggio sempre più vasto.

Da questo punto di vista, il manuale citato ha il pregio fondamentale di un ordinamento per materia che costituisce nel suo insieme un quadro sintetico della vita delle popola-

⁽¹⁾ U. T. E. T. Torino, 1940. pagg. XIX-518, con 93 grafici e 180 tavole numeriche, L. 90.

zioni studiata nel duplice aspetto demografico ed economico, tra loro strettamente collegati. Così la proverbiale « aridità » dei dati statistici viene prevenuta dall'interesse già apparso del problema cui risponde la loro analisi, resa inoltre agevole dall'appropriata ed accurata presentazione in tabelle, diagrammi, cartogrammi, e a sua volta ogni nuova nozione acquisita avvia a nuovi problemi che s'intravedono.

Per accostarsi allo studio o semplicemente alla conoscenza dei problemi sociali credo che la via seguita sia quella che meglio concilia l'ampiezza della visione con la concretezza delle informazioni, e quindi risulti, in particolare, assai rispondente alla preparazione di chi proviene da studi matematici o fisici, e troverebbe perciò troppo indeterminate le trattazioni e argomentazioni che non scendono ad analisi quantitativamente precisabili, e troverebbe, per contro, troppo aperte a dubbi, quelle che procedono in forma strettamente matematica partendo da assunti che spesso appaiono e sono insufficientemente fondati o artificiosamente costruiti.

Potrebbe infine giovare nell'insegnamento della matematica il ricorso ad esempi e applicazioni di carattere statistico, e soprattutto demografico-economico, che farebbero apparire non solo più varie, ma per molti più convincenti, le possibilità e le ragioni del metodo matematico; specie in nesso alle recenti direttive per l'educazione al pensiero matematico nella scuola, l'insegnante che cercasse di ricavare appropriati esempi svolgerebbe un'opera che potrebbe riuscire assai lodevole.

2. In ogni campo della scienza, si può dire che il pericolo da cui è necessario premunirsi nel ragionare è sempre uno solo: quello di dire delle cose prive di senso, e cioè o delle affermazioni monche per incompleta enunciazione delle circostanze atte a renderne univoco il significato, o delle affermazioni senza oggetto, riguardanti concetti illusori, non definiti, pseudoconcetti. Gli errori veri e propri, quando non sono conseguenza di semplici sviste, derivano infatti quasi sempre da simili equivoci. E i modi per cadere nel pericolo sono forse sempre riducibili a due: o il modo di pensare impreciso informato al linguaggio ordinario senza il sussidio dello spirito critico, che ha la sua più piena estrinsecazione nella matema-

tica, o, al contrario, l'uso di concetti definiti in forma matematica, ma senza il discernimento occorrente per accertarsi che essa traduca alcunchè di praticamente significativo.

Una delle forme più frequenti in cui si presenta, specie nella statistica, il pericolo di simili abbagli è data da certi concetti espressi da aggettivi di cui sembra spontaneo fare il comparativo, mentre perchè ciò abbia senso scientificamente occorre analizzare se e relativamente a quali concreti problemi una tale comparazione risponda ad utili scopi⁽¹⁾. Gustoso è — a tale riguardo — l'esempio di pag. 280 relativo al livello dei prezzi più o meno « caro » in due paesi con diverse abitudini in fatto di consumi: rispetto a ciascuno dei due paesi risulta che l'altro è più caro. E istruttiva è la conclusione, dopo riportata una formula proposta da alcuni autori per dare una risposta univoca ma convenzionale a tale confronto: « Non riteniamo però che questi compromessi risolvano le difficoltà teoriche. Non esiste una misura corretta della divergenza fra il costo della vita di due popolazioni differenti, per la semplice ragione che la stessa espressione « costo della vita » è un concetto vago e impreciso, che non può essere definito esattamente e quindi neppure misurato ».

Altra avvertenza importantissima, troppo spesso trascurata, e dall'A. lodevolmente e ripetutamente sottolineata, è quella contro l'uso inconsiderato di concetti portati con temeraria estrapolazione fuori del campo in cui legittimamente hanno significato (situazione con notissime analogie nella fisica!); ci riferiamo in particolare al concetto di prezzo, alle considerazioni sulla sua applicazione all'infuori del quantitativo di beni effettivamente contrattati, e in particolare all'estrapolazione estrema tendente a definire la misura della totale « ricchezza » di una nazione.

Così in ogni pagina, in ogni questione, è onestamente discusso e analizzato il valore delle conclusioni in rapporto ai

(1) Cfr. per es.: O. CHISINI: *Qualità, numeri indici e grandezze*, « Per. Mat. », vol. XIX, n. 5, 1939-XVIII, e, per il senso più speciale del termine « indice » nella statistica, la mia comunicazione alla I Riun. Scient. della Soc. It. di Statistica, (Pisa, 1939-XVII), *Indici statistici e « teoria delle strutture »*, « Supplem. Statistico ai Nuovi Problemi, ecc. », a. V nn. 2-3-4.

diversi aspetti dei problemi, conformemente al proposito esplicitamente annunciato nella prefazione di « abituare lo studioso di questioni demografiche e sociali al linguaggio cauto e coerente della scienza, avviandolo per una strada che non conduce sempre a scoperte e nozioni sicure, ma può preservare dalla superficialità e dalla tendenza (innata nelle persone intelligenti) di scoprire nessi di causa ed effetto dove esistono tutt'al più rapporti di dipendenza reciproca o correlazioni accidentali e ingannevoli ».

Oltre alle precauzioni necessarie in ogni scienza per evitare erronee impostazioni concettuali, ve n'è nella statistica di peculiari di cui è assai istruttivo e interessante seguire l'analisi, e che derivano dal fatto che gli stessi dati di partenza hanno un valore assai meno facilmente individuabile che nelle scienze fisiche. È ovvio ad es., che una diversità nel criterio di raggruppamento per professioni, o anche una diversa sfumatura nella distinzione tra professioni affini mal delimitate, rende impossibile ogni confronto, ma più delicati e ineliminabili sono i casi in cui tale incomparabilità deriva involontariamente per circostanze d'altra natura, come nello studio delle cause di morte, dove il progresso dei metodi diagnostici fa aumentare apparentemente la mortalità per certe cause precedentemente spesso non individuate (specie, sembra, pel cancro), oppure in tutti i casi ove la raccolta dei dati dipende più o meno dalla cura degli interessati (nel rispondere a questionari, ecc.) e si può presumere che per ciò stesso i dati riguardino un gruppo « selezionato ».

Naturalmente, tutte le osservazioni critiche devono prefiggersi non lo scopo negativo di evitare errori nel modo più rinunciatorio consistente nel non affermare nulla, ma quello positivo di cercare appropriatamente conclusioni attendibili, e sono svariati e spesso acuti e ingegnosi i criteri cui lo statistico deve ricorrere per la conferma indiretta di certe supposizioni. Per la citata supposizione sul cancro si è trovata, ad es., una conferma nel fatto che l'aumento deriva da casi situati in organi inaccessibili, e molti altri esempi istruttivi si troveranno in tutto il manuale. E si vedrà anche come in molti casi delle statistiche di cui appare scarso l'interesse diretto servano invece in realtà come opportuni surrogati di ricerche che per mancanza di elementi non possono eseguirsi;

Si vede quindi quanto ambigua e vuota possa risultare l'abusata formula « *ceteris paribus* »

Ma, più sostanzialmente ancora, la difficoltà di simili analisi proviene da quella che è la diversità essenziale delle relazioni che ha da mettere in luce lo statistico in confronto a quelle dei fisici: questi infatti hanno sempre a che fare in pratica con relazioni di dipendenza funzionale, chè, anche nei casi ove l'interpretazione teorica deve ricorrere a schemi probabilistici, le misure effettive (macroscopiche) derivano da masse innumerevoli, e conducono a relazioni funzionali, mentre la popolazione dello statistico, a volerla suddividere secondo tutti i caratteri che la differenziano, verrebbe ripartita in classi sempre più ristrette fino a ridursi ai singoli individui. E la natura probabilistica delle possibili conclusioni è pertanto questione che riguarda non solo l'enunciazione dei risultati, ma anche la raccolta ed elaborazione dei dati.

4. Quando si tratta di probabilità, gli statistici in genere (e non essi soli, purtroppo) hanno l'abitudine di vedere le cose in modo abbastanza superficiale; perciò non è il caso di fare un appunto particolare per ogni trattazione che non raggiunga in tale riguardo una difficile impeccabilità, e tanto meno per un manuale dove, per i suoi stessi intendimenti, della probabilità si parla raramente e quasi incidentalmente, salvo un cenno un po' più ampio a proposito delle probabilità di morte.

L'argomento si presterebbe a divagazioni senza limite, e cercheremo di limitarci a pochi cenni per mostrare come dall'introduzione del concetto di probabilità e dalle questioni apparentemente di mera indole critica che possono porsi al riguardo, discende direttamente l'atteggiamento di fronte a problemi del più concreto interesse positivo, come, ad es., quello dell'assicurazione, o di alta importanza metodologica, come, ad es., quello della perequazione, e al modo stesso di raffigurarsi il concetto e la funzione delle « leggi » nel campo statistico.

5. Mentre parecchi matematici, trovando non soddisfacente la definizione della probabilità mediante i famosi « casi favo-

attraverso congetture da discutere caso per caso, si può talvolta risalire alle conclusioni desiderate attraverso « sintomi » più accessibili.

3. Una delle questioni che più spesso si incontrano nella statistica consiste nello studio dell'influenza di determinati fattori sul fenomeno che interessa (per es., dell'età, sesso, professione, stato civile, statura, peso, ecc. sulla mortalità, complessiva o per singole cause, ecc.). Riducendosi alla schematizzazione matematica, si tratterebbe di studiare una grandezza f che dipende da più variabili x, y, z, \dots , e nulla vi sarebbe di capzioso per un matematico, anche se nello stesso linguaggio dell'analisi non è direttamente messo in luce (con effettivo pericolo per i principianti) che lo studio di f al variare « della sola x » è locuzione del tutto indeterminata, che acquista un senso preciso solo specificando quali sono le altre variabili (per es. y, z) che si considerano (e il significato è allora: « sulle linee $y = \text{cost.}, z = \text{cost.}$ »); se alle variabili x, y, z si sostituiscono x, u, v , il « variare della sola x » viene a significare tutt'altra cosa (cioè « sulle linee $u = \text{cost.}, v = \text{cost.}$ »). Ma è facile immaginare quale grado di indeterminatezza viene ad assumere il concetto nella statistica dove il numero delle variabili che si possono considerare è pressochè illimitato, e solo per necessità pratiche occorre limitarsi a quelle che appaiono più significative. Studiando ad es., l'eventuale influenza della statura sulla mortalità, si potrebbe trovare (esemplificazione immaginaria a titolo illustrativo) che la bassa statura dà luogo a mortalità maggiore, ma ciò per il fatto che la statura è in media più bassa in determinate regioni sanitariamente meno progredite, mentre in ogni singola regione (eliminata cioè l'influenza di tale fattore) la bassa statura dà mortalità minore; ma ciò a sua volta potrebbe derivare dal fatto che in ogni regione la statura è in media più bassa in certe categorie di professioni con vita più salubre, mentre a parità anche di tali fattori l'influenza torna a invertirsi, e così di seguito ⁽¹⁾.

(1) Per un più ampio esame critico di tale questione ed altre analoghe, si veda C. GINI: *I pericoli della Statistica*, (Discorso inaugurale della I Riun. Scient. della Soc. It. di Statistica, Pisa 1939-XVII) in « *Supplem. Statistico ai Nuovi Problemi*, ecc. », a. V, nn. 2-3-4.

revoli » e « possibili », hanno cercato di dare un significato più concreto alla probabilità ricollegandola alla nozione statistica di frequenza, è notevole che l'A. senta il bisogno di appoggiarsi alla definizione classica per introdurre il concetto di probabilità nella stessa statistica. Tale fatto di per sè conforta l'opinione che ho ripetutamente illustrata e spiegata, che, cioè, nè l'uno nè l'altro dei detti criteri conduca a *definire* il concetto di probabilità nè a chiarirne il significato, dato che il primo si limita a darne una particolare proprietà aritmetica, e il secondo a tradurre in una grossolana identificazione o confusione le delicate relazioni che possono stabilirsi fra le due nozioni completamente eterogenee di probabilità e di frequenza.

Su 43,7 milioni di abitanti essendone morti 615 mila, il quoziente di mortalità (ossia la *frequenza*) per l'Italia, 1938, fu di 14,1 per mille. Aggiunge ora l'A., ricordando la definizione mediante i casi favorevoli e possibili, che tale frequenza ha anche « il significato approssimativo di probabilità media di morte », purchè si ammetta che « tutti i casi siano ugualmente possibili », ossia che « tutti gli abitanti siano ugualmente esposti al rischio di morte », e ciò in quanto egli considera i morti come « casi favorevoli » e gli abitanti come « casi possibili ». Ma, per corrispondere effettivamente a tale computo di casi favorevoli e possibili, il problema dovrebbe essere un altro e formularsi così: qual'è la probabilità che scegliendo a caso fra i nomi dei 43,7 milioni di abitanti si estragga quello d'uno dei morti, che si sa essere 615 mila? E allora l'uguale possibilità si riferirebbe alle modalità della estrazione, non al rischio di morte. Considerando invece la probabilità di morte, non si ha affatto un esperimento unico con 43,7 milioni di diversi risultati possibili di cui 615 mila favorevoli, ma 43,7 milioni di esperimenti distinti ciascuno dei quali ammette due soli risultati possibili, sicchè il riferirsi alla definizione classica è fuori luogo. Con simile trasposizione di significato si verrebbe semplicemente a definire, nonostante la forma velata, che « probabilità = frequenza », e a ciò si aggiungerebbe la condizione di « uguale possibilità » dei diversi casi mentre essa prenderebbe un significato incompatibile con la stessa definizione. Infatti essa implica che si possano *confrontare* (e quindi, a fortiori, *definire*) le probabilità dei sin-

goli eventi, mentre la definizione basata sulla frequenza presuppone di considerare insieme intere masse di eventi. Il ricorso a quella restrizione risulta tuttavia, come intenzione, giustificato: giustificato quale correttivo della non corrispondenza alla nozione comune e intuitiva e genuina della probabilità delle definizioni che ne disconoscono il carattere soggettivo.

Che la *frequenza* sia risultata del 14,1 per mille è un dato di fatto e si riferisce ai casi *già osservati*; la *probabilità* esprime invece un'opinione, una previsione, e si riferisce ai casi *non ancora osservati*. Potrà darsi che nel valutarla terremo conto della frequenza osservata, così come il medico tiene nota dei dati osservati nel fare la prognosi, senza che questa sia per ciò un fatto obbiettivo, indipendente dalla scienza e coscienza del medico. Nel caso della probabilità, potremmo per altri motivi giudicarla diversa dal valore occasionalmente verificatosi per la frequenza, e, d'altro lato, potrà apparirci diversa per ogni singolo individuo o caso a seconda di circostanze di svariata natura.

6. Di circostanze che hanno influenza, per restare nell'esempio, sulle probabilità di morte, molte sono prese in esame dall'A., e molte più accennate. Di fronte a un individuo concreto, come fisserà il premio l'assicuratore? Benchè la pratica commerciale esiga una certa semplificazione, è ovvio che teoricamente il problema richiede la valutazione della probabilità per il singolo individuo, tenendo conto di tutte le sue caratteristiche che ne fanno un esemplare unico, e sarebbe assurda la risposta di certi autori, per cui l'assicurazione dello stesso individuo N. N. sarebbe equa a condizioni diverse a seconda del gruppo di caratteristiche considerate, cui corrispondono diverse frequenze (per es. considerando N. N. come « avvocato di 37 anni », o « 37-enne residente a Trieste », o « avvocato residente a Trieste », ecc.).

Uno dei casi più importanti in cui la stessa frequenza osservata non ci appare una base attendibile per la valutazione di una probabilità è quello che induce ad applicare una perequazione ai dati osservati. Se, ad es., i dati statistici mostrano che il quoziente di mortalità per l'età 37 è stato molto più elevato che per le età vicine, difficilmente troverò motivo di

ritenere che ciò sia dovuto a un particolare rischio proprio di quell'età piuttosto che costituire una irregolarità di cui non è prevedibile il ripetersi. Perciò le probabilità saranno comunque non le frequenze osservate ma dei valori *perequati*; si potrebbe dire in modo espressivo che tali valori perequati rappresentano *la prognosi*, relativa a un individuo *nelle stesse condizioni della collettività osservata, ma NON compreso nella statistica*, in base ai risultati da tale statistica desunti. Con tale distinzione, acquisterebbero chiarezza le considerazioni sulla perequazione, che non possono non lasciare incerti quando si accenna senz'altro ad eliminare « irregolarità dipendenti da cause accidentali », mentre la definizione adottata della probabilità non contemplava possibilità di eccezioni al suo rigore aritmetico giustificabili da distinzioni in cause accidentali o non (e in quale senso?).

7. Analoga è l'osservazione più generale che vorrei fare sul nesso tra osservazioni e conclusioni, ossia sull'interpretazione e il significato da attribuire alle « leggi » statistiche. In effetti le conclusioni sono sempre impregnate di un certo soffio vitale che trascende i dati osservati: sono cioè, come e nello stesso senso che la probabilità, soggettive. Cioè: tali da presupporre il vaglio da parte del pensiero di un soggetto, ciò che non significa escludere che egli si attenga nel modo più scrupoloso a quel che si suol chiamare « obbiettività », e cioè imparzialità o spregiudicatezza o serenità (temo che molte prevenzioni contro la teoria « soggettiva » nascano da tale possibile equivoco).

Lo scienziato, in ogni campo, elabora i dati, crea delle teorie per inquadrarli, e ciò è anzi vero in misura assai maggiore per le scienze fisiche, ma non è di questo che intendo dire: il fisico, qualunque teoria crei, sa sempre che dovrà risultarne che i fatti sperimentali vanno come l'esperimento ha dimostrato. Nella statistica non esistono fatti osservati che si possano ripetere identici, nè che abbiano, quand'anche fossero identici, necessariamente identico lo svolgimento, e in ciò la statistica è all'avanguardia sulla via dell'indeterminismo cui le scienze fisiche stanno indirizzandosi per motivi molto più riposti.

Invece la statistica si pone sempre sulla falsariga delle

scienze fisiche, e cerca di giustificare le proprie incertezze come difetti o manchevolezze particolari, assimilandole ad errori d'osservazione, a perturbazioni, ecc. Un franco capovolgimento di tale punto di vista mi sembra toglierebbe molte artificiosità e pericoli, e darebbe una visione più chiara dell'insieme del ragionamento. Ciò nel manuale è sempre analizzato e discusso con veramente rara premura di evitare illusioni e preservare da eccessive suggestioni: vorrei dire, in una parola, con onestà: ma, conservandosi fuori del piano schiettamente probabilista, ciò ha sempre piuttosto l'apparenza di scusare una cosa che dovrebbe andare così, ma non va proprio come dovrebbe, mentre preferirei e riterrei più conforme alla natura delle cose il dire direttamente che ciò va in quest'altro modo, perchè questo è il modo in cui ha da andare.

8. Analoghe osservazioni si potrebbero fare sul concetto di media: il mancato esplicito riconoscimento del suo valore relativo secondo la definizione di CHISINI (¹), troppo poco conosciuta e apprezzata ancor oggi dagli statistici, impedisce di sostituire con una sola parola veramente chiarificatrice tante precisazioni più o meno empiriche.

9. Per finire, scendiamo da considerazioni forse anche troppo generali per occuparci di qualche particolare formale relativo al manuale del LUZZATTO-FEGIZ.

Una cosa che ritengo dia ai nervi a tutti i matematici è il comparire di qualche « 100 », messo a ingombrare quasi sempre le formule di statistica per indicare l'uso di esprimere il risultato in « percento ». Mi sembra naturale una convenzione (che per mio conto ho sempre seguita) che elimina completamente l'inconveniente: considerare cioè il segno « ‰ » come simbolo dell'unità di misura « 0,01 », o, se si preferisce, come l'operatore a destra equivalente a « $\times 0,01$ », (e analogamente per « ‰‰ » = per mille), cosicchè si scriverà, ad es.:

$$1 = 100 ‰ = 1000 ‰‰, \quad 0,021 = 2,1 ‰ = 21 ‰‰, \quad \text{ecc.}$$

Tutti i calcoli si possono svolgere allora senza il superfluo

(¹) Sul concetto di media, « Per. Mat. », a. IX, n. 2, 1929.

moltiplicatore 100, e solo il risultato finale si tradurrà in « % »; anche nelle operazioni successive si procede agevolmente (per es. $25\% \times 80\% = 2000\% \text{‰} = 20\%$).

Una notazione poco indovinata sembra poi quella a/b per indicare l'intervallo da a a b ; infatti a/b indica la frazione a diviso per b , e comunque la barra / dà l'idea di una relazione di subordinazione, non di simmetria come tra i due estremi di un intervallo. Meglio l'espressiva notazione di PEANO $a^{1-}b$, o le usuali (a, b) , ecc.

Ben più numerosi e importanti, anche da un punto di vista formale, sarebbero i pregi, fra cui mi limiterò ad accennare l'accuratezza degli indici e delle tre appendici, contenenti rispettivamente una bibliografia opportunamente raggrupata per argomenti, un utilissimo « glossario dei termini tecnici », e una raccolta, tutta illustrata da esempi positivi e negativi, di regole sulle rappresentazioni grafiche.

Il manuale si rivolge specialmente ai pratici come guida per l'uso del discernimento logico e di una certa visione matematica nel campo dei dati statistici cui devono interessarsi; correlativamente, esso può per lo stesso motivo costituire un'ottima guida per il matematico che voglia accostarsi, col suo modo di raffigurarsi concetti e relazioni, ai grandi problemi della vita dei popoli, illustrati dalla statistica nel loro aspetto quantitativo.

BRUNO DE FINETTI
