

**LA NASCITA DEL CONCETTO DI CORRELAZIONE
NELLA TEORIA STATISTICA INGLESE**

JUAN IGNACIO PIOVANI

DOTTORATO DI RICERCA IN METODOLOGIA DELLE SCIENZE SOCIALI.

**DIPARTIMENTO DI RICERCA SOCIALE E METODOLOGIA SOCIOLOGICA
"GIANNI STATERA" (RISMES), FACOLTÀ DI SOCIOLOGIA.
UNIVERSITÀ DEGLI STUDI DI ROMA "LA SAPIENZA"**

DIRETTORE: PROF. ENZO CAMPELLI

TUTORE SCIENTIFICO: PROF. ALBERTO MARRADI

**DOCENTI ESAMINATORI: PROF.SSA MARISA FERRARI OCCHIONERO
PROF. ANTONIO FASANELLA
PROF.SSA RITA PAVSIC**

RIASSUNTO

La tesi tratta della nascita del concetto di correlazione e dei relativi strumenti operativi nel periodo a cavallo fra '800 e '900 in Inghilterra, collegandolo ai precedenti tentativi di introdurre la quantificazione nelle scienze sociali in Gran Bretagna e sul continente, e in particolare alla scuola di Gottingen ed ai lavori dei francesi sulla teoria della probabilità. Particolare attenzione si rivolge ai mutamenti di significato del termine 'statistica' fra il '700 e la fine dell'800; ai fondamenti filosofici di questa nuova concezione di compiti e strumenti della statistica, con speciale riguardo per le articolate posizioni di Pearson; agli sviluppi tecnici dei vari strumenti, ricostruendo personalità e contributi di Galton, Edgeworth, Pearson, Yule e Spearman; e agli interessi di ricerca di questi protagonisti, sottolineando in particolare l'influenza dell'evoluzionismo e dell'eugenetica su tutto il percorso intellettuale di Galton e Pearson, l'eclettismo di Edgeworth, gli interessi esclusivi per la psicologia di Spearman e la genuina apertura ai problemi delle scienze sociali in generale di Yule.

INDICE

Introduzione	3
Capitolo 1: L'emergere della teoria statistica inglese: l'epoca, le idee, i protagonisti	6
1.1. Dalla <i>Political Arithmetic</i> alla <i>Statistics</i> : la statistica come indagine socio-politica	6
1.2. Teoria matematica della probabilità e ricerca empirica: verso una ridefinizione della statistica	12
1.3. Variabilità e correlazione all'origine della moderna teoria statistica: una vera svolta	16
1.4. Gli artefici della teoria della correlazione	20
Capitolo 2: Epistemologia, ontologia e metodologia nella teoria statistica inglese: l'elogio della correlazione e la critica dell'idea di causalità	28
2.1. L'ambiente intellettuale dell'ottocento: la scienza in discussione	28
2.2. Scienza e metodo nel pensiero statistico di Karl Pearson	32
2.3. Fenomenismo, sensismo e conoscenza: l'idea di realtà all'origine dell'epistemologia statistica	41
2.4. Correlazione anziché causalità: la nuova arma scientifica contro l'ultimo bastione della metafisica	46
Capitolo 3: Lo sviluppo della correlazione come strumento concettuale e operativo della ricerca empirica	50
3.1. I teorici degli errori e i fondamenti matematici della teoria della correlazione	50
3.2. La svolta di Galton: l'emergere dei concetti di regressione e correlazione	54
3.3. Formalizzazione e generalizzazione dei concetti di Galton: i moderni strumenti della correlazione	61
3.4. Dalla correlazione all'associazione	67
3.5. La controversia sulla rilevazione dell'associazione	71
Capitolo 4: Interessi sociali, problemi scientifici e ricerca empirica: il punto di partenza e il punto di arrivo dei nuovi concetti e strumenti statistici	77
4.1. Premessa	77
4.2. Eugenetica, biometria e antropometria: le origini sociali, politiche e scientifiche della correlazione	78
4.3. Correlazione e associazione nella ricerca economica e sociale	84
4.4. La psicologia correlazionale	88
Conclusioni	92
Riferimenti bibliografici	95

INTRODUZIONE

Per l'uomo contemporaneo la concezione statistico-probabilistica della realtà costituisce forse uno di quei sistemi di idee che si considerano naturali. Concepire i fatti della vita da una prospettiva statistica è così diffuso che spesso non ci si accorge che si tratta di una costruzione socio-storica, un prodotto del pensiero umano; non di una caratteristica naturale, intrinseca alla realtà stessa.

Il fatto che sia data per scontata mostra il successo storico che questa costruzione ha avuto. Hacking (1990) la descrive come un nuovo "stile di ragionamento"¹, che presenta come l'esito di un quadruplice successo: metafisico, epistemologico, logico ed etico.²

Hacking afferma anche (1990, 1) che "l'evento concettuale più decisivo del ventesimo secolo fu la scoperta che il mondo non è deterministico. La causalità, per molto tempo bastione della metafisica, fu rovesciata", e man mano emerse una nuova forma di "conoscenza oggettiva" — quella statistica. Ma questo fu solo lo sbocco di un processo di trasformazione graduale in cui si distinguono soprattutto tre tradizioni intellettuali con origini riconducibili alla metà del seicento: la *Political Arithmetic* inglese, la *Statistik* tedesca e la *Théorie mathématique des probabilités* francese.

Gli inglesi inaugurarono una tradizione di ricerca sociale empirica con ricorso ai dati numerici. Dall'incontro di questa con la statistica universitaria tedesca — cioè lo studio comparativo degli Stati — negli anni a cavallo tra il settecento e l'ottocento, si ridefinì la statistica come lo studio quantitativo della società e della politica. Poco dopo i cultori della nuova disciplina cominciarono a usare gli strumenti di analisi della teoria matematica della probabilità — già diffusi in astronomia — e così facendo aprirono la strada a una nuova visione, basata sulla registrazione delle frequenze osservate al fine di individuare regolarità stocastiche, che a loro volta fornivano ulteriori basi per l'applicazione delle probabilità (Porter 1986, 17).

Tuttavia, le frequenze si interpretavano alla luce del paradigma allora dominante: si privilegiavano le idee di tipicità e normalità, e si trascurava quella di variabilità, ritenuta una manifestazione di errore. Solo quando quest'ultima cominciò a essere riconosciuta come un problema scientifico legittimo — nell'Inghilterra di fine ottocento — si produsse la svolta che avrebbe dato origine alla moderna teoria statistica. In effetti, è in questo periodo — e come risultato della ricerca di nuovi strumenti matematici probabilistici adatti all'indagine sulle vicende umane — che la statistica acquisì il suo significato attuale più diffuso: la disciplina dedicata agli strumenti concettuali e operativi per l'analisi quantitativa dei dati.

Seguendo Mackenzie (1981), per teoria statistica — o statistica matematica, come la chiama Porter (1986) — intendo qualcosa di diverso "dall'attività di raccolta d'informazione

¹ Hacking prende in prestito quest'espressione da Crombie (1994). Altri preferiscono descriverla come un nuovo paradigma. Infatti, dall'introduzione del concetto di paradigma, negli anni sessanta, esso è divenuto uno strumento euristico fondamentale per capire come le idee scientifiche condivise in un periodo storico determinato — e date per scontate dalla maggior parte degli scienziati — siano delle complesse costruzioni sociali. Solo attraverso uno sforzo intellettuale si comprende come operano queste costruzioni, e quali sono i processi per i quali divengono idee così diffuse da trasformarsi in verità accettate come tali.

² Ovviamente, parlare di successo non implica una valutazione della bontà o dell'adeguatezza di questo modo di ragionamento; significa soltanto riconoscere la sua egemonia e la sua onnipresenza negli odierni modi di concepire la realtà e di costruire e giustificare un discorso su essa. Questa è una differenza importante, giacché la dominante ideologia del successo, sicuramente funzionale allo status quo, tende ad omologare l'ampia diffusione e accettazione di un artefatto — compresi quelli scientifici — con la sua bontà e qualità: "quest'uomo ha più successo (soprattutto economico) perché è più intelligente"; "questo prodotto ha successo perché è superiore"; "questo film ha vinto dieci *Oscar* perché è il meglio"; "questa concezione scientifica è più diffusa perché è quella giusta." I processi attraverso i quali una concezione scientifica raggiunge ampio consenso, al di là delle sue qualità, è un interesse centrale della moderna sociologia e storiografia della scienza.

quantitativa che tipicamente perseguivano gli enti ufficiali e gli scienziati sociali [...] e dalla teoria matematica della probabilità". Si tratta della "costruzione di un quadro teorico per l'analisi di dati numerici" e degli "strumenti che possono essere usati" a un tale proposito (MacKenzie 1981, 7).

Il primo esempio di ciò venne dalla correlazione, la cui costruzione storica è oggetto di questo lavoro. Intendo la correlazione in senso ampio, per indicare "l'interdipendenza fra dati quantitativi o qualitativi, che comprende anche l'associazione tra proprietà dicotomiche e la contingenza fra proprietà classificate in molteplici categorie" (Kendall e Buckland 1976/1980, 56). Ovviamente, nell'approfondire l'argomento farò opportune distinzioni tra il termine 'correlazione' inteso nel senso stretto — che si riferisce a variabili cardinali — e l'associazione fra variabili categoriali.

Si tratta evidentemente di un tema metodologico, ma la trattazione che ne propongo non è tecnica. Fa invece ricorso ad approcci e concetti di filosofia, sociologia e storia della scienza; e si articola attorno a una serie di tesi fondamentali nei confronti della storia della correlazione, che si possono riassumere nei seguenti punti:

- Il concetto di correlazione emerse dall'applicazione degli strumenti matematici della probabilità alla ricerca empirica, ma divenne concepibile solo quando la variazione fu riconosciuta come legittimo problema scientifico. Pertanto, la variabilità si trova all'origine di questa teoria e dei suoi strumenti;
- Il concetto di correlazione segna l'emergere della statistica moderna. Il suo sviluppo accompagnò — e giustificò socialmente e accademicamente — il consolidarsi e istituzionalizzarsi della statistica come disciplina autonoma. In questo senso essa costituì una vera svolta: la fine della statistica come studio dei fenomeni dello Stato e il suo inizio come complesso di tecniche per l'analisi dei dati;
- Il concetto di correlazione nacque legato a particolari interessi di ricerca in cui l'associazione fra proprietà era fondamentale; e nel quadro di una concezione ontologica ed epistemologica di stampo positivista, che respingeva energicamente l'idea dell'esistenza della cosa in sé e delle cause inerenti;
- La correlazione fu concepita come strumento capace di elevare allo *status* di vere scienze le discipline umane; giacché consentiva di realizzare gli assunti tradizionali della scienza — come allora concepita — anche in questi campi. Mi riferisco soprattutto al controllo impersonale di asserti su relazioni fra proprietà, espressi in termini matematici (Marradi 1997a);
- Contrariamente alla visione più diffusa circa lo sviluppo storico della scienza e dei suoi strumenti operativi, l'idea di correlazione fu concepita a partire da — e si applicò per prima a — problemi della ricerca sull'uomo, sul comportamento e sulla società, per poi diffondersi a tutto lo spettro delle discipline scientifiche;
- Questi processi si svolsero in Inghilterra tra il 1870 — quando Galton cominciò a lavorare alla costruzione di tecniche statistiche per le sue ricerche sull'eredità — e la Prima Guerra Mondiale — quando la maggior parte della teoria della correlazione e i suoi strumenti erano stati sviluppati. Nell'arco di un quarantennio, nel triangolo London-Oxford-Cambridge, un ristretto gruppo di intellettuali; lavorando a stretto contatto anche se non sempre in piena armonia, gettò le basi di quella che in sostanza è restata a tutt'oggi la strumentazione essenziale di cui si serve l'analisi dei dati.

Queste tesi sono trattate in modo approfondito in quattro capitoli di questo volume. Nel primo, di natura socio-storica, si ricostruiscono le tre tradizioni intellettuali sopra menzionate, al fine di illustrare le profonde radici della moderna teoria statistica nel pensiero scientifico europeo e di apprezzare il contesto in cui nacque il suo primo frutto: la correlazione. In altre parole, si tratta di mettere in evidenza il passaggio dal senso originario del termine 'statistica' (riferito ai dati sulle caratteristiche degli Stati) a quello di insieme di strumenti concettuali e

operativi per l'analisi dei dati. Tuttavia non è solo la storia di un cambiamento concettuale; è la descrizione di un complesso processo che coinvolse anche mutamenti di tipo culturale e sociale. In questa sede presento anche gli artefici di questa svolta, fornendo alcuni particolari di rilievo della loro vita e della loro carriera.

Nel secondo capitolo si intraprende un percorso di natura filosofica, considerando i fondamenti ontologici ed epistemologici della nascente teoria statistica inglese di fine ottocento. Si analizza in modo approfondito il pensiero di Karl Pearson, l'unico tra gli artefici della correlazione che riuscì ad articolare in un'opera scritta l'insieme di assunti in sostegno di questo nuovo artefatto scientifico. Egli si occupò appunto di giustificarlo come uno strumento in grado di garantire la conoscenza impersonale in un mondo di impressioni inerentemente soggettive, e di stabilire le relazioni tra i fenomeni osservati in un contesto che riteneva non deterministico.

Il terzo capitolo è interamente dedicato allo sviluppo storico degli strumenti della correlazione: dai suoi fondamenti matematici radicati nella teoria degli errori alla svolta di Galton, poi alla sua formalizzazione e generalizzazione. Questo è il capitolo più tecnico, e tuttavia il ricorso a formule matematiche è marginale; si tratta soprattutto di una ricostruzione di taglio storico.

Nel quarto e ultimo capitolo il discorso si orienta di nuovo verso la sociologia della scienza, mantenendo comunque la linea di continuità dell'approccio storico. Il suo obiettivo è sottolineare i particolari problemi di ricerca e i relativi interessi sottostanti che hanno influenzato lo sviluppo della correlazione e dell'associazione, e in cui trovarono le sue applicazioni pionieristiche.

Avverto che nel prosieguo di questo lavoro userò alcune convenzioni che qui richiamo:

- se cito un brano da una traduzione o da un'edizione successiva di un'opera, nella chiamata della *reference* pongo prima l'anno di pubblicazione originaria seguito da una sbarra e dall'anno dell'edizione dalla quale traggio la citazione;
- le citazioni fatte dall'autore di un brano che sto citando sono poste tra apici anziché virgolette;
- se non diversamente segnalato i corsivi e le virgolette nei brani citati sono degli autori.

Tra i molti che mi hanno aiutato desidero ringraziare:

- il direttore, prof. Giorgio Alberti, della sede staccata dell'università di Bologna a Buenos Aires e il personale della biblioteca (Yael Poggi e Gisela Pagliano) per avermi facilitato l'accesso a preziosi materiali bibliografici;
- il coordinatore, prof. Enzo Campelli, e la segretaria, sig.na Francesca Sperandio, del dottorato di metodologia delle scienze sociali di Roma, per aver facilitato in ogni modo possibile il mio cammino attraverso le norme accademiche;
- il prof. Giovanni Di Franco, il prof. Alberto Marradi e la dott.ssa Maria Mendiola per avermi ospitato nei soggiorni italiani;
- il prof. Campelli e il collegio docente del dottorato di Metodologia delle scienze sociali di Roma per aver ascoltato la mia illustrazione dello schema di questo lavoro;
- il prof. Marradi per averne discusso con me vari punti, e averne rivisto tutta la stesura.

CAPITOLO 1

L'emergere della teoria statistica inglese: l'epoca, le idee, i protagonisti.

The source of modern statistics is to be found [...] in the use of probability as a modeling tool to capture and analyze real variation in nature and society.

Theodor Porter (1986, 91).

Much of the material presented in modern courses on statistical methods for social sciences is superficially similar to texts available by 1830, and yet the adoption of these methods for the different purposes of the social scientists was so glacially slow that it amounted a reinvention.

Steven Stigler (1999, 3).

1.1. Dalla *Political Arithmetic* alla *Statistics*: la statistica come indagine socio-politica.

In senso ampio, la statistica britannica — in quanto attività di raccolta e organizzazione di dati relativi allo Stato — risale al diciassettesimo secolo. E largamente condivisa l'idea che il lavoro fondante nel campo dello studio statistico dei problemi sociali sia quello di John Graunt, *Natural and Political Observations on the Bills of Mortality* (1662).³ Pochi anni dopo, William Petty, in una lettera a Lord Anglesea del 17 dicembre 1672, introdusse l'espressione *political arithmethic* (Cullen, 1975) — verosimile antenato dell'oggi più diffusa *social statistics* — per designare questo tipo di studi:

Lo studio sistematico dei numeri sociali nello spirito della filosofia naturale ebbe origine durante il decennio del 1660, e si conobbe per circa un secolo e mezzo come *Political Arithmetic*. Il suo proposito [...] era la promozione di una solida e ben informata politica sociale (Porter 1986, 18).

I *London Bills of Mortality* erano allora l'unica fonte di dati demografici facilmente disponibile in Gran Bretagna: un materiale poco interessante, dal quale tuttavia Graunt trasse ampie conclusioni (Cullen 1975). Il fatto più importante non è pertanto la raccolta e organizzazione dei dati stessi: un lavoro incipiente, gestito principalmente da enti pubblici, secondo una tendenza crescente in vari stati europei, particolarmente la Francia e la Prussia. Seguendo Cullen, l'aspetto notevole, invece, sono l'analisi di questi dati, le conclusioni da essi tratte⁴, le riflessioni metodologiche⁵ e la giustificazione di una nuova disciplina — la *Political*

³ Si è discusso molto nell'ambiente accademico inglese quale sia in realtà il vero autore dell'opera citata sopra, se Graunt o Petty. Per un resoconto dettagliato di questo dibattito si veda Cullen (1975). Molti studiosi credono che Graunt sia l'autore delle *Observations*, mentre Petty sarebbe l'autore delle conclusioni e l'ideatore dell'inquadratura generale. Quest'affermazione si basa specialmente sul fatto della grande competenza aritmetica impiegata nello studio — non attribuibile a Petty. Ciononostante, Cullen sostiene che fu Petty l'artefice delle idee che generarono lo studio statistico della società.

⁴ Tra le conclusioni più importanti trasse da Graunt, Hull (1899) cita: la regolarità dei fenomeni sociali, la nascita di più maschi che non femmine, l'approssimativo equilibrio nella quantità di uomini e di donne nella società, l'alto tasso di mortalità nei primi anni di vita, il più alto tasso di morti in ambienti urbani rispetto a quei rurali.

⁵ Cullen (1975) ritiene che il lavoro di Graunt mostri grandi doti metodologiche. Sottolinea ad esempio che prima di analizzare statisticamente un dato, egli si preoccupa seriamente del problema della sua affidabilità.

Arithmetic — in un modo che continuerà a ispirare lo sviluppo della statistica per moltissimi anni, cioè quello del riformatore che intende la raccolta scrupolosa dei dati come requisito fondamentale di qualunque riforma sociale. Nelle parole del suo ispiratore — e persino inventore del suo nome — il compito della nuova disciplina era l'applicazione dei principi di Bacone all'arte di governare (Porter 1986). In effetti, adottando un parallelo tratto da Bacone tra il corpo naturale e quel politico, Petty scrisse:

Praticare sul corpo politico senza conoscere la sua simmetria, il suo tessuto e le sue proporzioni, è tanto superficiale ed informale come le operazioni [...] dei praticoni (*empiricks*)⁶ [sul corpo naturale] (1691, 129).

Secondo Lazarsfeld (1961), l'idea che le questioni sociali possono essere sottoposte all'analisi quantitativa si afferma nel clima intellettuale dell'era baconiana, per il desiderio d'imitare il successo delle scienze naturali, vale a dire la quantificazione⁷, e per le caratteristiche degli stati nazionali, che richiedevano una base per le decisioni pubbliche. Più specificamente, come si vedrà più avanti, hanno pesato lo sviluppo di sistemi di assicurazione ai quali serviva una solida base numerica, e la concezione mercantilista secondo la quale l'entità della popolazione era un fattore cruciale della ricchezza e del potere delle nazioni.

La *Political Arithmetic* si definì come una scienza della società le cui conclusioni dipendevano da numeri e misure. Pertanto, la qualità dei dati doveva migliorare, e non a caso Petty argomentò più volte a favore della creazione di un ufficio centrale di statistica. Tuttavia, le sue realizzazioni furono molto limitate, specialmente per le sue scarse conoscenze matematiche. Malgrado i suoi sforzi, nessuno continuò la linea proposta da Petty e l'espressione *political arithmetic* si svalutò. La stessa fortuna ebbe la sua idea di una scienza di tipo interdisciplinare, un ibrido di tutte le discipline sociali, in grado di usare tecniche quantitative. La statistica sociale rimase confinata alla demografia — un processo che Cullen (1975) chiama riduzione della *political arithmetic* alla demografia. Infatti, nel tardo seicento e nel settecento questo tipo di studi fiorì con l'obiettivo di descrivere la situazione degli Stati in termini demografici e costruire tabelle destinate a calcolare le assicurazioni e i tassi di rendita annuale, utili ai nascenti sistemi di assicurazione.

Ad ogni modo, una linea di continuità importante è costituita dal fatto che restò uguale l'impostazione e la giustificazione del lavoro: mettere queste conoscenze a disposizione del governo per la politica sociale. Questa giustificazione sarà importante nel successivo sviluppo della teoria statistica, dovuto alla necessità di risolvere problemi di ricerca legati al desiderio di "migliorare" la società.⁸

Nei cinquant'anni tra il 1780 e il 1830 alcuni segni di espansione e diversificazione delle statistiche sociali incominciarono ad intravedersi in Gran Bretagna: fu istituito un censimento nazionale, si rafforzò lo sviluppo delle statistiche mediche, si scoprirono nuove tecniche d'analisi e si applicarono tecniche già conosciute a problemi finora non quantificati (Cullen 1975). Si introdusse nella lingua inglese il termine *statistics* che subirà un graduale cambiamento di significato.

⁶ Petty utilizza la parola inglese arcaica *empiricks*, il cui significato è diverso da quello del moderno termine *empiricist* (empirista). Tratta dal suo contesto, *empirick* fa riferimento alla persona che opera sulla base d'una conoscenza derivata da esperienze personali, che però non è scientifica.

⁷ Lazarsfeld aggiunge opportunamente che queste sono le spiegazioni "convenzionali". Per quanto riguarda la quantificazione e la misurazione, per esempio, non tutti sono d'accordo. C'è ampio consenso nell'attribuire a Galilei — che riteneva che Dio avesse scritto il mondo nel linguaggio della matematica — l'impulso verso la quantificazione scientifica, giacché per capire il mondo si doveva misurare e calcolare. Tuttavia, la misurazione rimase per molto tempo confinata ad astronomia, geometria, meccanica e altre discipline classiche. La misurazione non ebbe un ruolo centrale in fisica e chimica almeno fino al 1840 (Kuhn 1961; Hacking 1990).

⁸ Si pensi per esempio all'importanza del progetto eugenetico nello sviluppo di quello che MacKenzie (1981) chiama "teoria statistica" e Porter (1986) chiama "statistica matematica". In effetti, alcuni dei suoi principali esponenti — Galton, Pearson, Fisher — condividevano l'obiettivo di attuare un progetto eugenetico (MacKenzie 1981).

Il termine fu impiegato per primo in inglese da Hooper (1770) come traduzione del tedesco *statistik*. Seguendo la tradizione tedesca, il neologismo inglese *statistics* fu definito come la scienza che "insegna l'ordinamento politico degli Stati moderni noti nel mondo".⁹ Tuttavia, il termine aveva in tedesco una storia più lunga, che risale agli studi del celebre Alfred Achenwall, professore a Göttingen.

Effettivamente, secondo Meitzen (1886), la statistica ha due radici: la *political arithmetic* britannica e l'*universitätsstatistik* (statistica universitaria tedesca). La seconda radice, da cui deriva il termine, è una tradizione che considerava gli aspetti rilevanti per uno Stato — aspetti fra i quali i dati numerici non erano necessariamente i più importanti; anzi, giocavano un ruolo piuttosto secondario (Lazarsfeld, 1961).

Secondo Hacking (1990) non è corretto attribuire la tendenza a raccogliere e organizzare dati ad un'unica nazione; praticamente tutti gli stati europei hanno seguito un proprio percorso nell'organizzazione delle informazioni per loro rilevanti: le città italiane del Rinascimento, culla della moderna idea di Stato, fecero indagini statistiche per prime in Europa, e gli spagnoli già nel tardo cinquecento organizzavano dei censimenti nelle colonie. Ma furono i tedeschi a proporre l'idea che lo Stato sia essenzialmente caratterizzato dalle sue statistiche, idea che si materializzò attraverso tre manifestazioni.

Una prima manifestazione, di tipo intellettuale, è proprio l'*Universitätsstatistik*, la cui origine è certamente anteriore ad Achenwall. Ad esempio Westergaard (1932) ritiene che la disciplina *Statistik* — indirizzata allo studio comparato e sistematico degli Stati — era già stata definita da Hermann Conring con il nome di *Staatenkunde*, nel secolo diciassettesimo.¹⁰ Già prima John (1884) aveva sostenuto che questa tradizione risaliva alle lezioni universitarie di Conring, intitolate *Notitia Rerum Publicarum* e *Notitia Statuum Germaniae*, e tenute in un periodo contemporaneo di Graunt e Petty. Pertanto, la radice inglese e quella tedesca si svilupparono contemporaneamente. Ma i lavori pionieristici di Conring, che scriveva fondamentalmente in latino, rimasero per molto tempo sconosciuti, e anche oggi molti non li riconoscono come l'origine della scuola tedesca di statistica universitaria. Achenwall sarebbe stato il primo a definire quest'oggetto di studio come una scienza autonoma chiamata *Statistik* (Cullen 1975) e a fare una presentazione ordinata, in tedesco, di questa tradizione (Lazarsfeld 1961).

Una seconda manifestazione è l'enumerazione sistematica della popolazione e di altre proprietà demografiche, a carico degli uffici pubblici, che cominciò in Prussia nella prima metà del 700, ed era già stata proposta ai governanti da Leibniz. Egli ebbe un vivo interesse nelle statistiche, particolarmente in quelle relative alle malattie, alla mortalità e alla popolazione in generale, e propose un sistema di 56 categorie per valutare lo Stato (inclusendo la distribuzione per sesso e per ceti, il tasso di mortalità infantile, il numero di donne in condizioni di sposarsi, etc.) (Leibniz 1700).

La terza manifestazione, di tipo privato e in uno stile più vicino a quello della *political arithmetic* britannica, è costituita dagli studi sulla popolazione. L'impresa più notevole in questa direzione fu compiuta nella Germania del settecento da Süssmilch: uno studio delle nascite e della mortalità (Hacking 1990).

Come già detto, la manifestazione che penetrò nell'ambiente intellettuale inglese, alla fine del settecento, è l'*Universitätsstatistik*. Apparentemente, il lavoro di Achenwall fu più noto non solo per il fatto di essere in lingua tedesca, ma anche per l'ambiente istituzionale in cui si svolse: l'Università di Göttingen, fondata nel 1737. Lì si costituì una scuola di fama europea, dove si raffinarono le idee di Conring e si considerarono nuovi strumenti metodologici. Uno degli interessi salienti della Scuola consisteva nella presentazione comparata di informazioni su differenti stati tedeschi¹¹ mediante uno schema a due dimensioni: in quella orizzontale si disponevano gli Stati da comparare, e in quella verticale le

⁹ Defizione di von Bielfeld, tradotta all'inglese da Hooper nel 1770.

¹⁰ Per Westergaard, le origini della disciplina si potrebbero far risalire ad Aristotele.

¹¹ Si ricordi che allora Germania era suddivisa in vari regni (Prussia, Sassonia, Baviera), principati, ducati, etc.

categorie rilevanti alla comparazione. Originariamente, le celle definite dall'incrocio delle dimensioni erano riempite con delle espressioni verbali. Gradualmente, si incominciò a far ricorso ad espressioni numeriche, un'inevitabile conseguenza di tale schematizzazione — quanto meno perché occupavano meno spazio (Lazarsfeld 1961). La tendenza alla presentazione numerica favorì a sua volta, in una sorta di circolo, i temi più affini alla quantificazione, presagendo l'odierna matrice dei dati, che non pochi (ad esempio Marradi 1980; 1996b) ritengono lo strumento intellettuale e operativo che caratterizza la ricerca sociale di tipo standard.

Dopo la traduzione del libro di von Bielfeld *The Elements of Universal Erudition, Containing an Analytical Abridgement of the Sciences, Polite Arts, and Belles Lettres*, eseguita da Hooper, il termine non fu usato in inglese fino al 1787, anno in cui fu tradotta un'altra opera tedesca. Si tratta in questo caso del saggio di von Zimmermann, *A Political Survey of Europe, in Sixteen Tables; Illustrated with Observations on the Wealth and Commerce, the Government, Finances, Military State, and Religions of the Several Countries*. Tuttavia, la versione anglicizzata del termine (con un significato un po' diverso da quello originario tedesco) non appare fino a 1791, con la pubblicazione del primo dei 21 volumi dello *Statistical Account of Scotland*, una compilazione a cura di Sir John Sinclair, centro di una rete di pastori presbiteriani il cui lavoro collettivo rese possibile quest'opera monumentale (Cullen 1975; Porter 1986). Per Sinclair la ricerca statistica fa "riferimento alla popolazione, le circostanze politiche, le produzioni ed altri affari dello Stato".¹² Secondo Porter (1986) Sinclair adottò deliberatamente il termine 'statistica', ma contemporaneamente fece notare che il suo progetto era diverso da quello tedesco; mentre le ricerche tedesche dette *statistik* s'interessavano agli affari dello Stato, quelle proposte da Sinclair s'indirizzavano ad accertare il *quantum* di felicità del popolo e i mezzi per il suo progresso. Peraltro l'edizione del 1797 dell'*Encyclopaedia Britannica* definiva il termine come "una parola introdotta recentemente per esprimere la visione o rilevamento di un regno, contea o parrocchia" (Mackenzie 1981, 7) — che è l'accezione tedesca.

L'incontro delle due tradizioni — la *political arithmetic* britannica e l'*universitätsstatistik* tedesca — verso la fine del settecento generò una controversia circa lo status scientifico e l'utilità della disciplina. La battaglia concettuale¹³ fu vinta, persino in Germania, dagli aritmetici politici. In effetti, Knies (1850) conclude che benché il termine 'statistica' si debba ad Achenwall, esso dovrebbe usarsi nel senso della *political arithmetic*. Dall'inizio dell'ottocento, questa fu l'accezione del termine. L'oggetto dell'antica tradizione tedesca fu considerato da allora in poi un interesse della scienza politica (Lazarsfeld 1961) che avrebbe dovuto chiamarsi *Staatskunde* (Hacking 1990), termine che secondo Westergaard si applicava agli studi di Conring prima che Achenwall li ribattezzasse come *Statistik*.

Non dovrebbe sorprendere, pertanto, che nei 30 o 40 anni dopo Sinclair si incominciò a legare progressivamente la statistica a qualche elemento di quantificazione.¹⁴ Tuttavia, un'accezione esclusivamente quantitativa doveva ancora emergere, perché malgrado il ricorso alla quantificazione, l'uso di tecniche di analisi matematica, salvo che a fini attuariali, era piuttosto poco diffuso (Cullen 1975).

Il grande sviluppo delle statistiche numeriche si ebbe nella prima metà dell'ottocento in Gran Bretagna con l'emergere del "movimento statistico" (*statistical movement*), tipico dei primi decenni del regno di Vittoria, e con l'istituzione degli uffici statistici. Nel 1832 fu fondato il Dipartimento di Statistica del Ministero del Commercio (*Board of Trade*), un ufficio centralizzato in grado di generare statistiche a sostegno delle riforme promosse da

¹² Citato da Cullen (1975, 10).

¹³ Sull'importanza delle battaglie sui concetti e sui termini nel mondo scientifico si veda Toulmin (1972); Marradi (1987).

¹⁴ Già nel 1801 Playfair suggerì che il termine 'statistica' doveva essere legato a qualche forma di quantificazione, ma egli non ebbe grande influenza sui contemporanei: non fu citato da un altro inglese fino a Jevons nel 1879 (Cullen 1975).

importanti attori politici e sociali. Inoltre, questo dipartimento doveva soddisfare il bisogno di informazione accurata e affidabile su quello che succedeva nelle varie periferie della Gran Bretagna — in particolare sulla produzione industriale e il commercio. Nel 1836 fu istituito un Ufficio Generale del Registro (*General Register Office*) incaricato della sistemazione dei dati relativi alle nascite, morti e matrimoni.

In quest'epoca si formarono anche le associazioni dedicate esclusivamente alla statistica: la prima a Manchester nel 1833. Poco prima era stata fondata la sezione di Statistica della *British Association for the Advancement of Science*, grazie alle influenti presenze di Malthus e di Quetelet nella riunione tenutasi a Cambridge; ciò anche se non tutti approvavano l'esistenza di una sezione di questo genere in seno all'*Association* (Porter 1986). I suoi fondatori furono anche membri attivi del nucleo della *Statistical Society of London*, creata nel marzo 1834, e poi divenuta *Royal Statistical Society* nel 1886:

"Anche se c'era qualche matematico tra i membri originari, c'erano ancora più economisti, politici, membri della camera dei Lords, ufficiali di governo e medici: il suo oggetto era l'informazione utile circa la società, non lo sviluppo dei metodi matematici" (Abrams 1968, 14).

Si formarono anche delle istituzioni analoghe nelle città importanti, particolarmente a Glasgow, che ebbe due associazioni contemporaneamente. Il desiderio di capire le trasformazioni sociali del tempo, e di stabilire una base scientifica per la politica sociale, era senz'altro congeniale al movimento statistico dei primi decenni del regno di Vittoria (Porter 1986).

Uno dei temi di maggior interesse del movimento statistico erano le "statistiche morali", relative all'educazione, la criminalità, e in minor misura alla religione. I suoi fautori erano generalmente contrari all'intervento dello Stato in economia, ma favorevoli al suo intervento in sede sanitaria e educativa.

Si cercavano le cause della degenerazione morale e dell'aumento del crimine e dell'agitazione sociale, problemi che preoccupavano quasi ossessivamente i membri delle società statistiche. La ricerca statistica avrebbe fornito sostegno empirico per le riforme da loro volute (*ibidem*). È chiaro che questi fini esplicitati dai membri delle Società erano accompagnati da "una latente funzione di controllo sociale [...] per la potenziale minaccia all'ordine pubblico costituita da ampi strati sociali che vivevano in condizioni miserande" (Pitrone 1996, 19).

L'urbanizzazione, più dell'industrializzazione, dominava le menti degli *statists*¹⁵, per la convinzione che il carattere e le abitudini della popolazione dipendessero dell'ambiente urbano. Questo spinse al rilevamento delle condizioni di vita della classe operaia urbana, in cui la Società di Manchester fu pioniere. Le Società volevano promuovere una classe operaia virtuosa; secondo le loro indagini, nella campagna le persone erano sobrie, frugali e laboriose; invece in città la situazione era allarmante. Le riforme nei settori sanitario e educativo erano ritenute due modi di affrontare la decadenza morale e la turbolenza politica della classe operaia (Cullen 1975).

Lo scontento sociale era molto diffuso negli anni 30 e 40 dell'ottocento. Attraverso il miglioramento della situazione sociale, le riforme cercavano in realtà — come già detto — di prevenire la minaccia delle forze anarchiche e socialiste. L'educazione significava una combinazione di istruzione fisica, morale e intellettuale; il suo proposito era, in sostanza, convertire una classe al sistema di valori dell'altra. La seconda riforma riguardava la sanità pubblica; si riteneva che le condizioni dell'ambiente fisico avessero effetti degradanti.

La prosperità raggiunta a metà del periodo vittoriano e la conseguente relativa tranquillità sociale scoraggiarono le teorie basate sull'urgenza delle riforme sociali, tipiche del movimento statistico. Le società periferiche, a eccezioni di quella di Manchester, crollarono, e

¹⁵ Dopo l'introduzione del termine *statistics*, gli uomini dedicati al lavoro con dati numerici erano noti nella Gran Bretagna con l'etichetta di *statists*. Il termine *statistician* risale invece al tardo ottocento.

il movimento, con le sue caratteristiche distintive, perse progressivamente l'impeto. A partire del 1850 circa, i temi trattati nelle istituzioni rimaste, particolarmente quella di Londra, si ampliarono, gli interessi si diversificarono, si cominciò a registrare una maggior varietà di opinioni, e un notevole aumento delle ricerche in economia (Cullen 1975).

La raccolta e compilazione di dati numerici era già ben affermata in Gran Bretagna verso il 1850. Ma la trattazione matematica non era sofisticata e il termine 'statistica' non aveva ancora la connotazione attuale. Come afferma Mackenzie (1981, 8) "nella Gran Bretagna della metà dell'ottocento non c'era una tradizione di teoria statistica". Pertanto, il movimento statistico dei primi decenni del regno di Vittoria non deve essere giudicato un precursore della moderna statistica. Esso non lasciò una tradizione di teoria statistica, e la raccolta d'informazioni quantitative rimase del tutto separata dagli sviluppi della teoria matematica della probabilità (*ibidem*).

Dalla nascita della *Statistical Society of London* fino al 1850, solo il 2% delle memorie presentate trattava del metodo statistico. Nell'introduzione al primo numero del suo *Journal*, uscito nel 1838, si definì la statistica sottolineando l'aspetto della raccolta di dati sociali e politici, e trascurando le questioni di metodo e di analisi dei dati:

La Scienza della Statistica [...] non discute cause [...] Solo cerca di raccogliere, organizzare e comparare, quella classe di fatti che da soli formano la base delle corrette conclusioni rispetto al governo sociale e politico.¹⁶

Secondo la concezione prevalente i dati grezzi, senza "opinioni", costituivano la caratteristica essenziale della scienza e il fondamento della statistica. La prima regola della *Statistical Society* era escludere i pareri. William Farr, in nome della Società, scrisse a Florence Nightingale, che era preoccupata per la mancanza di analisi in un contributo che al tempo stava preparando per la rivista: "non vogliamo impressioni, ma fatti."¹⁷ E questi "fatti" si presentavano in forma quantitativa perché i numeri rappresentavano i "veri" fatti.

Tuttavia, non tutti erano d'accordo con queste idee sulla statistica. Nel 1838 Robertson affermò: "Nessuna forma di registrazione ed organizzazione di dati può costituire da sé una scienza [...] La statistica deve intendersi come un metodo, un modo di organizzare i fatti che appartengono a varie scienze."¹⁸ La transizione dal movimento statistico originario implicò pertanto l'inizio della decadenza della sua classica concezione di statistica, e il timido emergere di un'altra più legata all'idea di metodo. Come afferma Cullen (1975), in questo periodo di transizione la *Statistical Society* non era in grado di decidere se l'obiettivo principale della disciplina fossero i fatti o il metodo; e già nel 1860 un suo *fellow*, Fox, rispecchiando i recenti sviluppi della disciplina nel continente, scrisse che essa si doveva considerare un settore della matematica.

In effetti, l'applicazione della matematica alle ricerche empiriche cominciava a manifestarsi nei paesi francofoni dell'Europa continentale. Questo non dovrebbe sorprendere, dato che la teoria della probabilità appartenne inizialmente alla tradizione intellettuale francese.

Quest'impostazione, in cui si combinano la teoria matematica e l'idea strumentale di metodo applicabile alle varie scienze empiriche, si consoliderà lentamente in Gran Bretagna nel periodo a cavallo tra l'ottocento ed il novecento, quando l'assimilazione della teoria della probabilità continentale, e la sua applicazione alla ricerca dei fenomeni umani, produrrà una vera svolta scientifica, e contribuirà all'emergere della moderna teoria statistica.

¹⁶ Citato da Porter (1986, 35-36).

¹⁷ Ivi, 36.

¹⁸ Ivi, 40.

1.2. Teoria matematica della probabilità e ricerca sociale empirica: verso una ridefinizione della statistica.

Si potrebbe cominciare un discorso sullo sviluppo e le applicazioni della teoria della probabilità alla ricerca empirica evidenziando un fatto storico piuttosto casuale: la nascita in Francia di questa tradizione intellettuale — lo studio della probabilità — contemporaneamente all'emergere del pensiero statistico britannico, nella forma della *political arithmetic*, e di quello tedesco, manifestato nell'*Universitätsstatistik*.

In effetti, l'emergere della teoria della probabilità si può ricondurre alla metà del seicento, con i lavori pionieristici di Fermat e Pascal. Alle sue origini, la teoria non intendeva sfidare la concezione della natura allora dominante, che Hacking (1990, cap. 2) chiama "dottrina della necessità"¹⁹, e si limitava a delle applicazioni astratte nel campo dei giochi (Stigler 1999), lontane da ogni considerazione circa caratteristiche della realtà.

Se la statistica britannica si presentava come un'applicazione dei principi di Bacone ai problemi politici e sociali, e pertanto come una manifestazione dell'orientamento empirico della filosofia e della cultura inglesi (Mitchell 1973, 147), la teoria della probabilità potrebbe inquadrarsi nella tradizione razionalista francese, come una manifestazione dell'orientamento astratto, matematico-deduttivo voluto da Descartes. Paradossalmente, anche se ancorata nella tradizione della razionalità, la dialettica della probabilità eroderà gradualmente alcuni dei principi su cui si basava la razionalità moderna; in particolare l'idea di certezza.

Come afferma Hacking (1990), nell'età della ragione, 'casualità' e 'irrazionalità' erano termini imparentati, che denotavano lo stesso atteggiamento sbagliato. L'uomo razionale copriva il caos apparente della realtà con il "velo delle inesorabili leggi naturali". Gradualmente, la teoria della probabilità iniziò a modellare le prospettive gnoseologiche e, alla fine, persino quelle ontologiche. Nel settecento, già si interpretava come il calcolo della razionalità del mondo in un contesto di conoscenza imperfetta (Porter 1986). Il mondo in sé era ancora deterministico, ma la sua conoscenza era diventata probabile. Nell'ottocento essa rese possibile concepire un mondo con regolarità e tuttavia non sottoposto a leggi universali. Alla fine del secolo, la casualità era già un'idea rispettabile, pronta a diventare la fedele ancella delle scienze naturali, biologiche e sociali (Hacking 1990). Dall'incontro tra la teoria della probabilità e i suoi strumenti euristici con la ricerca empirica nel campo dei fenomeni definiti come non predeterminati né soggetti a cause inesorabili, emerse la moderna teoria matematica della statistica. Nei paragrafi seguenti descriverò brevemente questo processo lungo e complesso.

Nonostante i suoi inizi legati alla speculazione astratta sui giochi aleatori, nel settecento si scoprì che la teoria della probabilità era in grado di fornire strumenti utili per la ricerca scientifica. Già Galilei aveva dichiarato che la misurazione e il calcolo erano cruciali nell'impresa scientifica, unico modo di capire una realtà che Dio aveva scritto nel linguaggio della matematica. Parallelamente Descartes aveva promosso la matematica come la chiave del pensiero scientifico organizzato. Tuttavia, la misurazione rimase confinata per molti anni all'astronomia e altre discipline classiche. Non dovrebbe sorprendere, pertanto, che la teoria matematica della probabilità trovasse le sue prime applicazioni in questa sede, in cui la misurazione era già affermata sin dai tempi di Galilei. Come afferma Stigler (1999), molti dei più rilevanti contributi alla statistica moderna, basata nella teoria della probabilità, emersero dalla considerazione di problemi scientifici concreti, e il caso della misurazione in astronomia è un esempio tipico.

¹⁹ Descartes è il suo primo difensore; anche Kant, negli scritti etici e metafisici, sosteneva un'idea analoga. Hume, pur essendo ritenuto un filosofo critico dell'idea di causalità, scrisse nell'*Enquiry Concerning Human Understanding* (1748) che niente esiste senza una causa per la sua esistenza. Il suo scetticismo si limitava al piano gnoseologico, alla possibilità di conoscere le cause, ma non nel piano ontologico, cioè, alla loro esistenza (Hacking 1990).

Infatti, il problema della misurazione favorì lo sviluppo degli strumenti concettuali ed euristici che si fondano sulla teoria degli errori, che "si svolse nel settecento e nell'ottocento come un'appendice [...] dell'astronomia" (MacKenzie 1981, 56). Gli astronomi "dovettero riconoscere che è impossibile misurare qualsiasi cosa con completa accuratezza ed esattezza". Tuttavia, il ricercatore poteva approfittare della possibilità di ripetere le sue misurazioni per accrescere la precisione. L'obiettivo della teoria degli errori era appunto questo: "ridurre al massimo possibile l'errore in una quantità data, e dare una stima affidabile del [suo] ammontare" (*ibidem*). Le applicazioni della teoria degli errori in astronomia si basavano largamente su due strumenti concettuali di eccezionale importanza, tuttora cruciali nella ricerca scientifica: la distribuzione normale e il metodo dei minimi quadrati.

Nel 1733, De Moivre²⁰ aveva introdotto nella teoria della probabilità la funzione della curva normale, allora conosciuta come legge degli errori.²¹ Per la prima volta era possibile applicare la teoria della probabilità a un numero indefinitamente grande di eventi indipendenti (Porter 1986).

L'innovazione di De Moivre, tuttavia, ebbe poco seguito fino agli anni 70 del settecento, quando Laplace cominciò a scrivere sulla teoria della probabilità. Laplace vede nella curva normale uno strumento eccellente, la cui applicabilità estende alla probabilità a posteriori: fino a una sua *Memoire* del 1774, infatti, l'idea di probabilità era piuttosto confinata al calcolo aprioristico dell'esito di giochi aleatori.

Laplace era convinto che le più importanti questioni della vita fossero dei problemi di probabilità, e pertanto suscettibili di soluzione con le sue teorie. Tuttavia, egli era altrettanto convinto che gli eventi del mondo fossero assolutamente determinati; il ruolo della casualità e della teoria matematica della probabilità era dovuto al divario esistente tra quello che è e quello che l'uomo può conoscere. La realtà è determinata da leggi non sempre aperte alla mente umana, e pertanto la probabilità era una proprietà gnoseologica, non ontologica (Richards 1997). Laplace ripropone pertanto il dogma della necessità: tutti gli eventi, persino quelli che per essere insignificanti non parrebbero seguire le leggi della natura, sono in realtà una loro necessaria conseguenza.

Un altro contributo importante all'applicazione della legge degli errori è legato al nome di Fourier. Egli era molto colpito dall'indefinita ripetizione di eventi che appaiono fortuiti, e che tuttavia, considerati complessivamente, fanno svanire ogni impressione d'irregolarità. Fourier era convinto dell'applicabilità della legge degli errori a una vastità di campi e oggetti, tutti caratterizzati da un'unità sottostante finora nascosta, svelata grazie all'universalità della curva degli errori (Porter 1986).

Il metodo dei minimi quadrati ha una storia più recente. Nel 1805, Legendre annunciò un metodo generale per ridurre le molteplici osservazioni di un oggetto, come una stella o un pianeta, anche se lo presentò senza alcuna giustificazione probabilistica. Questa è la prima comunicazione effettiva del metodo, giacché probabilmente Gauss — com'egli stesso più volte dichiarò — lo usava da quasi dieci anni, senza avere avuto occasione di renderlo pubblico.²²

Nel 1810, Laplace fornì una derivazione alternativa del metodo dei minimi quadrati, e stabilì che gli errori in astronomia, come negli studi della popolazione, dovevano distribuirsi secondo la legge degli errori (Porter 1986). Come afferma MacKenzie (1981, 57):

²⁰ Si noti che il libro di De Moivre, *The Doctrine of Chances* si pubblicò nel 1718. La funzione della curva normale fu introdotta nella seconda edizione del 1733. Peraltro, egli continuò la linea classica che non sfidava la dottrina della necessità. Analogamente a Hume, e poi a Laplace, egli affermava che tutto ciò che accade nella natura è il prodotto delle proprietà fisiche, anche se non le conosciamo.

²¹ Il termine 'normale' per riferirsi alla legge degli errori emerse nella seconda metà dell'ottocento, quando essa cominciava ad essere usata in contesti diversi dall'astronomia, particolarmente in statistica ed in altre scienze (Stigler 1999). La distribuzione normale è anche conosciuta come curva di Gauss, a sostegno della "legge della eponimia di Stigler", secondo la quale "nessuna scoperta scientifica prende il nome dell'originale scopritore" (Stigler 1999, 7).

²² Questa è la conclusione delle pazienti ricerche di Stigler (1981; 1999) sulla storia della statistica.

I teorici degli errori mostrarono che la miglior stima di una quantità [...] era di solito la media delle varie misurazioni, e che queste misurazioni tipicamente seguivano la distribuzione matematica che egli chiamavano 'legge della frequenza degli errori'

Gli astronomi e i matematici dell'ottocento produssero una gran mole di articoli sul metodo dei minimi quadrati.²³ Alla morte di Laplace nel 1827, la teoria della probabilità aveva raggiunto un alto livello. Verso la metà dell'ottocento, la conoscenza delle tecniche matematiche della probabilità era molto diffusa, accessibile a chiunque avesse conoscenze di analisi. L'applicazione di queste tecniche, particolarmente il metodo dei minimi quadrati, era *routine* in astronomia e geodesia, ma la sua diffusione in altri campi era lenta (Stigler 1999). Non dovrebbe sorprendere, pertanto, che l'estensione di questi metodi al campo dei fenomeni umani si debba a una figura con profonde conoscenze di astronomia — Aphonse Quetelet — e che sia legata alla cultura scientifica francofona²⁴; non solo per la sua posizione privilegiata come culla della teoria della probabilità, ma anche per lo status che la statistica aveva raggiunto in Francia all'inizio dell'ottocento.

Non pochi situano in Francia il centro del grande entusiasmo statistico dei primi decenni dell'ottocento²⁵, che pochi anni dopo avrà un seguito nel movimento statistico vittoriano. La passione per l'informazione quantitativa era già forte al tempo del Consolato e dell'Impero, particolarmente nell'ambito della sanità pubblica. Peraltro, come i loro colleghi britannici, gli statistici francesi erano borghesi riformisti che cercavano di sostituire "il lungo regno dell'opinione, gli interessi di partito e la confusione politica con un sicuro nucleo di ben stabiliti fatti sociali e verità rigorosamente dedotte" (Coleman 1982, 275).

Le leggi della probabilità, applicate inizialmente con successo all'astronomia, cominciarono lentamente ad essere esportate in altri campi. In un periodo in cui le scienze fisiche non avevano ancora bisogno della probabilità, la "scoperta" delle leggi statistiche negli affari umani caratterizzati da regolarità era una conseguenza prevedibile. La "legge dei grandi numeri", proposta da Poisson nel 1835, fornì fondamenti ancora migliori per applicare la teoria della probabilità agli affari sociali (Hacking 1990).

Quetelet sarà in larga misura il responsabile di questa svolta nella storia della statistica, rafforzando la tendenza alla diffusione della quantificazione a diverse aree oltre l'astronomia, e allo stesso tempo concedendo particolare attenzione ai metodi per il trattamento dei dati quantitativi (Lazarsfeld 1961). Profondamente influenzato dai lavori di Laplace e di Fourier sulla teoria della probabilità, Quetelet si convinse, verso il 1830, della possibilità di trattare le questioni umane applicando i metodi già consueti in astronomia, andando oltre la mera raccolta e classificazione dei dati. Secondo Lazarsfeld (1961, 295):

Questa combinazione della matematica astratta e la realtà sociale fornì la convergenza ideale delle due linee in cui si era sviluppata la mente di Quetelet

Inizialmente tentò di applicare le nozioni probabilistiche dei francesi alla misurazione del corpo umano — un interesse collegato al suo desiderio giovanile di diventare scultore — e in seguito pubblicò un lavoro che trattava dell'applicazione della probabilità alle "scienze politiche e morali".

Nel 1844 Quetelet annunciò che la legge degli errori si applicava anche alla distribuzione delle caratteristiche umane. Per lui le leggi statistiche si manifestavano come regolarità espresse nei termini di una distribuzione di frequenza. Il concetto centrale era quello di normalità, ed essa si rappresentava attraverso il valore medio della distribuzione. L'uomo medio (*homme moyen*) era l'ancora del gruppo, e prendeva il suo significato dal

²³ Nel 1877 Merriman elencò 408 libri e memorie sulla legge della frequenza degli errori e il metodo dei minimi quadrati (MacKenzie 1981).

²⁴ Quetelet era belga, ma studiò a Parigi ed era molto influenzato dalla cultura scientifica francese.

²⁵ Si veda ad esempio Hacking (1990).

gruppo. Quetelet era consapevole che la media non era sufficiente per comparare delle distribuzioni (Stigler 1999), ma tuttavia non s'interessò alla variabilità dei fenomeni umani. La sua assimilazione della normalità con quel che è giusto, corretto, buono — che ritroveremo in Durkheim — tenderà a concepire la variabilità come manifestazione patologica. Solo una reinterpretazione del termine, con la quale normalità diviene mediocrità, permetterà, nella Gran Bretagna di fine ottocento, di recuperare la variabilità come un problema scientifico rilevante, e così facendo aprirà la porta a una nuova svolta nella storia della statistica.

Verso il 1830-40, gli inglesi vennero in contatto con gli sviluppi scientifici continentali, e dovettero affrontare la sfida di assimilare la teoria della probabilità in una cultura in cui fra la religione e la scienza non c'erano ancora confini chiari (Richards 1997). La teologia naturale d'impostazione anglicana dominava gli ambienti universitari di Cambridge e Oxford, ed intellettuali influenti — ad esempio Whewell — si opposero all'impostazione dei "matematici continentali", particolarmente Laplace.

Fu Augustus de Morgan a presentare la teoria di Laplace al pubblico inglese in due lavori fondamentali: *Theory of Probabilities*, pubblicato nel 1838 nell'*Encyclopaedia Metropolitana* e un saggio sulla probabilità e la sua applicazione alle assicurazioni, uscito nella *Cabinet Cyclopaedia*. Attento a non sfidare le idee dominanti della teologia naturale, de Morgan si preoccupò, seguendo Laplace, di collocare le incertezze che la teoria della probabilità matematizzava nella mente umana, e non nel mondo esteriore (*ibidem*). Divenne comunque chiaro che la teoria della probabilità aveva importanti applicazioni pratiche, particolarmente in astronomia e nel campo delle assicurazioni, tema caro alla statistica britannica sin dalla fine del seicento.

Nel 1850, l'influente astronomo inglese Herschel aveva attirato l'attenzione degli scienziati inglesi sulle idee di Quetelet in tema di applicazione della probabilità ai dati sociali (Lazarsfeld 1961). Peraltro, Quetelet era già conosciuto negli ambienti intellettuali della Gran Bretagna. Negli anni 30 aveva avuto un ruolo decisivo nella creazione della sezione di statistica dell'*Association for the Advancement of Science*, e nel 1841 aveva presentato, in un incontro di statistici tenutosi a Plymouth, un lungo elenco di temi (la meteorologia, la fisica, la chimica, la botanica, l'agricoltura, la zoologia oltre alle questioni umane) che a suo avviso potevano essere studiati applicando i metodi statistici.

Tuttavia, gli strumenti concettuali della probabilità, e in particolare quelli della teoria degli errori, continuarono ad essere una specialità principalmente continentale. Dei quasi 500 saggi relativi alla legge degli errori e al metodo dei minimi quadrati, elencati da Merriman nel 1877, solo il 14% era stato pubblicato in Gran Bretagna (MacKenzie 1981). Un caso esemplare è Stanley Jevons: aveva studiato teoria della probabilità con de Morgan a Londra; sosteneva la sua utilità nelle scienze sociali e difendeva la matematizzazione dell'economia. Ciononostante, non fece alcun tentativo di profittare degli strumenti probabilistici di quantificazione dell'incertezza in queste discipline.

Il cambiamento di atteggiamento degli inglesi nei confronti della probabilità e del suo impiego pratico avvenne nell'ultimo quarto del secolo, in un ambiente in cui l'evoluzionismo darwiniano stava creando condizioni più favorevoli allo sviluppo di una scienza non deterministica. Fu proprio un cugino di Darwin, Francis Galton, il massimo responsabile della svolta che portò all'emergere della moderna teoria statistica.

L'impostazione evoluzionista condusse Galton a interessarsi all'eredità dei caratteri umani. Nella sua ricerca di strumenti statistici con cui affrontare il problema, egli si rivolse alla teoria degli errori, che Quetelet aveva già applicato con successo agli affari umani (*ibidem*). Tuttavia, Galton non si accontentava, come Quetelet, della mera applicabilità universale della legge degli errori. Inoltre, si oppose alla sua interpretazione della normalità: la deviazione dalla media non era necessariamente un difetto. Come riformatore sociale d'impostazione eugenetica, gli interessava il miglioramento della razza, e quindi la diffusione di quelle caratteristiche eccezionali che Quetelet riteneva devianti. *L'homme moyen* non era

per niente il modello ideale; era un mediocre le cui caratteristiche dovevano migliorarsi nei suoi discendenti.

La valorizzazione del concetto di variabilità era una conseguenza inevitabile. Infatti, Galton fu il primo a usare i metodi dell'analisi degli errori per trattare la variazione empirica; ma a questo proposito, le inadeguatezze tecniche della teoria degli errori cominciarono a diventare evidenti:

Per i teorici degli errori, la variabilità ("errore") era qualcosa da eliminare, o quanto meno da controllare e misurare. Gli obiettivi dei teorici degli errori erano pertanto contrari al trattamento della variabilità come un fenomeno di valore in se stesso (MacKenzie 1981, 58).

Studiare la variabilità richiedeva pertanto nuovi strumenti concettuali e operativi. La statistica moderna, in quanto disciplina autonoma orientata all'analisi dei dati, stava nascendo. Alla fine dell'ottocento, quando l'idea che ogni fatto dell'universo fosse determinato da una legge naturale necessaria stava perdendo forza (Hacking 1990), i concetti di regressione e correlazione e i relativi strumenti consentivano di analizzare la variabilità dei fenomeni e le loro relazioni. La scienza, come voluto da Galilei, continuò ad orientarsi verso la ricerca delle relazioni tra proprietà; ma esse non potranno più ritenersi necessarie e immutabili.

1.3. Variabilità e correlazione all'origine della moderna teoria statistica: una vera svolta.

Molto del materiale presentato oggi agli studenti di statistica sociale, dal punto di vista matematico, è simile a quello dei testi già disponibili verso il 1830; tuttavia, l'adozione di questi metodi nella ricerca sociale fu così lenta che richiese una reinvenzione. I problemi erano concettuali, non meramente matematici, e apparvero gradualmente man mano che gli scienziati si trovarono ad applicare al mondo sociale i costrutti matematici degli astronomi (Stigler 1999).

In astronomia, la probabilità si usava per rappresentare le imperfezioni della rilevazione; l'obiettivo principale era stimare o eliminare l'errore. Ma la moderna statistica emerse dall'uso della teoria della probabilità come uno strumento per catturare e analizzare la variabilità congeniale ai fenomeni naturali e sociali (Porter 1986).

Il passaggio di interesse dall'errore alla variabilità s'intravede in un certo senso nell'interpretazione della frequenza tipica del pensiero di Quetelet. In effetti, oggetto d'attenzione di Quetelet era la funzione probabile degli errori alla quale Galton si riferì poeticamente come "legge suprema dell'irrazionalità". Dalla sua introduzione nella prima metà del settecento, questa curva subì un processo di graduale mutamento concettuale nelle sue applicazioni alla ricerca empirica: dalla distribuzione degli errori di misurazione in astronomia alle deviazioni da un uomo medio idealizzato, poi alla genuina variabilità dei fenomeni naturali e umani. Secondo Porter (1986, 91), la reinterpretazione della curva normale "come la legge della genuina variazione, anziché del mero errore, fu il risultato centrale del pensiero statistico dell'ottocento". Esso permise lo sviluppo delle moderne tecniche di analisi statistica, a cominciare da regressione e correlazione, la cui invenzione Stigler (1999, 6) definisce come "uno dei più grandi eventi nella storia della scienza."

La quasi simultanea apparizione del libro *Natural Inheritance* e del metodo della correlazione segna l'inizio del periodo moderno della statistica. La correlazione come strumento prometteva di essere utile in ogni campo, specialmente in quelli dove stabilire chiare direttrici di causalità era problematico. Ma ai tempi di Galton, la teoria statistica come disciplina scientifica era effettivamente inesistente in Gran Bretagna: non s'insegnava come

corso universitario²⁶, non aveva una base istituzionale e non si pubblicava alcuna rivista specializzata dedicata alla disciplina.²⁷

Galton era consapevole di questo fatto, e presto si rese conto che regressione e correlazione avrebbero potuto essere il punto di partenza di una nuova disciplina scientifica; ma per ottenere questo, la conoscenza doveva essere organizzata e presentata in modo sistematico. Nella sua biografia, scritta da Karl Pearson (1914-1930), si riproduce una lettera di questo periodo: "Il momento è maturo affinché un matematico competente" riordini il materiale statistico. "Egli praticamente *fonderà* una scienza."

Deciso a progredire nella sua impresa, Galton tentò per molto tempo di attirare l'interesse di matematici esperti. Le sue conoscenze in questa disciplina erano limitate, e lo sviluppo della statistica applicata richiedeva il coinvolgimento di matematici più raffinati. Dalla fine del decennio del 1860 fino agli anni ottanta egli aveva lavorato sui problemi statistici con la collaborazione occasionale di qualche matematico, ma senza aiuti sostanziali. In fondo, egli rimase per vari decenni piuttosto isolato dal punto di vista intellettuale. A tutti gli anni 80, Edgeworth era l'unico intellettuale inglese, oltre a Galton, che lavorasse in modo serio e continuato sulla teoria statistica (MacKenzie 1981).

La diffusione dei nascenti concetti di regressione e correlazione, attorno al 1890, cominciò ad attirare l'attenzione di statistici noti, come Edgeworth, e di promettenti scienziati più giovani, come Weldon e Pearson. Il sogno di Galton di una nuova disciplina prendeva forma.

Edgeworth, che aveva compiuto un lavoro molto rilevante per gettare le fondamenta di una statistica matematica generale sin dagli anni 80 (Stigler 1999) cominciò a lavorare a un problema suggerito da Galton: la generalizzazione della distribuzione normale bivariata a un numero più alto di variabili. Contemporaneamente, nel 1892 Pearson cominciò una collaborazione con lo zoologo Weldon, applicando il concetto di correlazione alla ricerca biologica e affinando i metodi di Galton.

Verso il 1900 la natura della statistica usciva complessivamente trasformata dal lavoro di Galton, Edgeworth e Pearson. (Porter 1986). Da una disciplina limitata allo studio quantitativo dell'uomo e della società, essa stava diventando una disciplina in grado di fornire degli strumenti di analisi per una vasta gamma di scienze. Già nei suoi primi corsi di statistica avanzata, tenutisi all'Università di Londra a partire dalla metà degli anni 90, Karl Pearson insegnava che "statistica è [...] un termine utilizzato per misurazioni aggregate di qualsiasi fatto, sia sociale, fisico o biologico. La teoria pura della statistica è un ramo della matematica che si occupa della compilazione, rappresentazione e trattamento di [questi] aggregati numerici — indipendentemente dei fatti che i numeri rappresentino. La statistica applicata [infine] è l'impiego dei metodi della statistica pura a classi speciali di fatti — osservazioni biologiche, fisiche o politiche"²⁸ (Yule 1938, 200).

La moderna statistica matematica divenne possibile solo quando il pensiero statistico fu sufficientemente sofisticato e articolato da divenire un ramo autonomo della matematica (Porter 1986). Come dichiara il figlio di Karl Pearson²⁹: "lo sviluppo della statistica era

²⁶ Parlando degli anni 90, Yule (1938) afferma che un corso di teoria statistica semplice, organizzato, e logicamente articolato non poteva ancora esistere in quanto molti elementi fondamentali della disciplina avevano appena cominciato a svilupparsi.

²⁷ Si ricordi che la *Statistical Society of London*, già divenuta *Royal Statistical Society* al tempo della pubblicazione di *Natural Inheritance*, non era impegnata allo sviluppo della disciplina come intesa oggi. La sua rivista pubblicava allora solo molto occasionalmente degli articoli dedicati ai metodi statistici.

²⁸ Per Karl Pearson, quest'aspetto applicato della statistica era fondamentale: "Il principale proposito di tutto il [suo] lavoro era lo sviluppo dei metodi statistici applicati allo studio dei problemi dell'eredità e dell'evoluzione; sarebbe certamente sbagliato immaginare il Pearson di questo periodo [1890-1900] dedicato allo sviluppo della teoria statistica di per se stessa" (E. Pearson 1938/1948, 56). Egli era favorevole a "riavviare meglio la tradizione di applicare la scienza alla vita pratica", tema ampiamente discusso in una sua lezione al *Gresham College* nel 1890 (ivi, 48).

²⁹ Pearson ebbe due figlie e un figlio, Egon, che continuò la tradizione del padre. Egli fu uno dei più importanti ed influenti statistici del secolo xx.

[allora] strettamente collegato al problema della correlazione" (E. Pearson 1938/1948, 54), e questa tecnica trovava rapida applicazione in una gran varietà di problemi in biologia, economia, psicologia e sociologia.³⁰ Di conseguenza, "il nucleo di una nuova disciplina stava nascendo" (MacKenzie 1981, 102).

In generale, la scienza si stava professionalizzando velocemente negli ultimi decenni del periodo vittoriano, e gli scienziati cominciavano a perseguire ricerche sempre più specializzate (Lightman 1997). La frattura del contesto intellettuale comune assicurato dalla teologia naturale, conseguenza inevitabile del successo dell'evoluzionismo darwinista e dello sviluppo della teoria della probabilità, favorì la specializzazione e la professionalizzazione della scienza (Young 1985). Queste tendenze non costituiscono un fatto secondario, giacché hanno un ruolo centrale nel favorire l'istituzionalizzazione di ogni disciplina. Come afferma MacKenzie (1981) sono molteplici gli aspetti da considerare prima di poter parlare dell'emergere di una nuova disciplina. Occorrono, tra l'altro, una rete di scienziati interessati allo stesso tipo di problemi; un insieme di strumenti d'analisi e approcci più o meno condivisi; mezzi di comunicazione — formali ed informali — tra gli esperti; meccanismi di reclutamento e d'insegnamento, con una forma stabile e una base istituzionale; finanziamenti e altre risorse. La tendenza alla specializzazione e alla professionalizzazione della scienza è pertanto rilevante perché creò un ambiente favorevole all'istituzionalizzazione della statistica come disciplina autonoma.

In questo processo fu Pearson la figura cruciale: combinando le sue capacità matematiche con gli interessi di ricerca empirica promossi da Galton, e manifestando un entusiasmo per l'istituzionalizzazione che Porter (1986) giustamente chiama "imprenditorialità accademica", riuscì a stabilire — intellettualmente e socialmente — la statistica matematica come una nuova disciplina.

Dalla collaborazione tra Pearson e Weldon, iniziata nel 1892, ebbe origine la Scuola Biometrica. Dal 1894, quando Pearson cominciò a tenere il corso di statistica avanzata, fino agli anni 20 del novecento — quando Fisher stabilì un centro di ricerca alla stazione sperimentale di Rothamstead — la Scuola Biometrica fu l'unica istituzione britannica dedicata alla teoria statistica (MacKenzie 1981). Già nel decennio della sua creazione circa la metà di tutti gli articoli inglesi in teoria statistica si producevano in quest'ambiente istituzionale. *Biometrika*, la rivista della Scuola, fu per molti anni la pubblicazione più importante della Gran Bretagna per la teoria statistica, e continua ancora oggi ad essere una rivista di grande prestigio internazionale nella disciplina. Inoltre, la Scuola fu per molti anni la più importante fonte di insegnamento statistico, attraendo studenti che avrebbero raggiunto notorietà e posti di grande importanza istituzionale e accademica (Norton 1978).

Il gruppo di scienziati che contribuirono allo sviluppo della statistica inglese in questo periodo è molto ridotto. Usando la *Bibliography of Statistical Literature* di Kendall e Doig (1968), Mackenzie è riuscito a ricostruire la situazione della teoria statistica tra il 1900 e la Prima Guerra Mondiale: egli sostiene che solo ventisei individui avevano allora un interesse attivo nella disciplina, e fra loro, dodici possono essere considerati membri della Scuola Biometrica di Pearson. Tra i suoi docenti, si contavano "cinque delle quindici persone nominate dalla *International Encyclopedia of the Social Science* come gli autori dei più decisivi contributi allo sviluppo dell'odierno metodo statistico" (Ben David 1971, 151).

Pertanto, queste cifre mostrano che quest'ambito istituzionale, pur essendo il più importante, non era completamente egemonico.³¹ Man mano che si consolidava, la disciplina

³⁰ Valga come esempio l'affermazione di Lazarsfeld a proposito della sociologia. Alla fine dell'ottocento "la quantificazione in sociologia acquisisce la sua moderna funzione: tradurre le idee in operazioni empiriche e cercare relazioni regolari tra le variabili così create." La tecnica della correlazione sarà fondamentale in questa vicenda (Lazarsfeld 1961, 332-333).

³¹ Gli esperti di teoria statistica che allora non lavoravano alla Scuola Biometrica non erano né isolati, come nei tempi di Galton, né irrilevanti. Molti di loro partecipavano alle discussioni teoriche e tecniche del tempo, pubblicavano e insegnavano statistica in diversi ambiti. Persino Edgeworth, che probabilmente per ragioni di personalità non si sentiva parte di una Scuola ben definita, aveva una base istituzionale alla *Royal Statistical*

cominciava ad essere praticata in altri ambienti.³² Gli interessi intellettuali che avevano dato il primo impulso allo sviluppo delle tecniche di regressione e correlazione si diversificavano; e gli strumenti della Scuola Biometrica, concepiti per il trattamento dei problemi eugenetici e biologici, mostravano di essere applicabili in altre discipline (MacKenzie 1981).

La Scuola Biometrica continuò per molto tempo a occupare il centro della scena statistica britannica. La sociologia della scienza ci può aiutare a capire il perché. La Scuola era un gruppo sociale coeso sotto la chiara *leadership* di Pearson.³³ Questa situazione favorì un ambiente di forte coerenza intellettuale. Molte delle ricerche erano collettive, e si contava su una rivista propria e su varie collane di pubblicazioni legate ai laboratori di ricerca. Pearson controllava strettamente queste pubblicazioni, al punto che *Biometrika* è stata definita come "la rivista più personalmente editata mai pubblicata" (Yule 1936, 100).

La Scuola Biometrica non era solo un'organizzazione formale, ma un insieme di rapporti sociali relativamente stabili, che la dotavano di un'esistenza al di sopra dei singoli individui che ne facevano parte (MacKenzie 1981).

La posizione di potere istituzionale di Pearson si consolidò quando nel 1906, oltre alla direzione del Laboratorio di Biometria, assume la conduzione del Laboratorio di Eugenetica dell'Università di Londra. Su indicazione di Galton e con il suo contributo finanziario l'Università aveva stabilito un Ufficio di Registri Eugenetici (*Eugenics Records Office*). Nel 1906 Galton chiese a Pearson di assumerne la direzione, ed esso divenne il *Galton Laboratory of National Eugenics*. Nel 1911, Galton morì lasciando all'Università un fondo per creare una cattedra di eugenetica, con laboratorio e biblioteca. Inoltre, suggerì che il posto venisse offerto a Pearson. Aggiungendo donazioni e raccolte pubbliche, fu possibile costruire un edificio per ospitare il Laboratorio di Biometria ed il Laboratorio Galton di Eugenetica, fusi in un Dipartimento di Statistica Applicata. Questo fu il primo dipartimento inglese dedicato alla didattica avanzata e alla ricerca sulla teoria statistica.

La maggior parte delle tecniche più diffuse si svilupparono in quest'ambiente istituzionale, ed è abituale concepire la storia della disciplina come iniziata con Galton, se non con lo stesso Pearson. Levine (1996), in effetti, afferma che la statistica come disciplina scientifica fu modellata da Pearson, che le diede forma e direzione dominando la scena della teoria statistica dalla metà del 1890 fino alla Prima Guerra Mondiale.

Ma col tempo l'unificazione di statistica, biometria ed eugenetica, tipica dell'approccio di Galton e di Pearson, si allentò. La disciplina stava maturando, e il suo consolidamento erodeva il potere quasi monopolista di Pearson. Nuovi *leaders* intellettuali ed istituzionali, soprattutto Fisher, guadagnavano spazio nel mondo già meno ristretto della statistica matematica. Dopo il pensionamento di Pearson nel 1933, l'Università di Londra divise il dipartimento di Statistica Applicata, creando tre aree differenti: una cattedra specializzata di statistica, che occuperà per primo Egon Pearson; una cattedra di eugenetica, affidata a Fisher; e una cattedra di biometria affidata a Haldane.

Pearson criticò pesantemente questa decisione, convinto del fatto che essa indebolisse la posizione della disciplina che tanto aveva contribuito a far nascere. In realtà, la statistica stava ricevendo un nuovo impulso, rendendola indipendente da una particolare applicazione

Society e contribuiva con delle pubblicazioni e con la discussione del lavoro degli altri. Nella *Royal Statistical Society* stava emergendo un gruppo relativamente articolato di persone interessate alla moderna teoria statistica, guidato da Edgeworth stesso, da Yule e da Hooker. Il *Journal* cominciava a prestare più attenzione a questi temi. Alcuni degli articoli pubblicati nel *Journal* — particolarmente quelli di Yule — si possono giudicare decisivi contributi alla moderna teoria statistica.

³² Un caso notevole è quello di Gosset, noto nella teoria statistica per lo sviluppo della distribuzione *t*, detta 'student' per lo pseudonimo con il quale firmava gli articoli. Egli si unì temporaneamente alla Scuola Biometrica nel 1906, ma lavorava con problemi di statistica matematica in un ambiente del tutto diverso: era infatti un dipendente della birreria irlandese *Guinness*. Malgrado la sua frequentazione dell'accademia, egli decise di proseguire la carriera nell'area industriale, raggiungendo il posto di *manager* dell'impianto di Londra istituito nel 1935 (MacKenzie 1981).

³³ Più di quaranta persone lavorarono con Pearson nel periodo 1900-1914 (*ibidem*).

alla biometria e da un'impostazione eugenetica che diventava sempre meno sostenibile. Molti progressi teorici fondamentali erano stati già compiuti, e la teoria statistica come specializzazione scientifica maturava rapidamente. Negli anni trenta la situazione era radicalmente mutata da quella in cui Pearson aveva cominciato a lavorare: un gruppo di attivi ricercatori era esclusivamente dedicato alla disciplina; nell'*University College* di Londra c'era una nuova area orientata alla didattica e alla ricerca sulla teoria statistica, erede del suo vecchio dipartimento; e alla stazione di ricerca agricola di Rothamsted Fisher aveva creato un secondo centro specializzato. La rivista *Biometrika* continuava a pubblicare importanti contributi alla teoria statistica, e il *Journal* della *Royal Statistical Society* dava sempre più spazio ad articoli del genere.

Peraltro, il lavoro essenziale per la costituzione della teoria statistica come insieme di strumenti di analisi quantitativa applicabili alla ricerca empirica, legata allo sviluppo dei concetti di regressione e correlazione e dei relativi strumenti, era in gran parte già compiuto prima dello scoppio della Prima Guerra Mondiale.³⁴ Il coefficiente di correlazione fu definito matematicamente nel 1896, il tetracorico, Q e χ^2 nel 1900, ρ nel 1904. I personaggi notevoli da considerare sono Galton, Edgeworth, Pearson, Spearman e Yule.³⁵ Tornerò su questi protagonisti nella prossima sezione, per presentare alcuni aspetti rilevanti della loro vita e delle loro opere. Parafrasando quello che Stigler (1999) dice di Quetelet, si può affermare che non sono stati affatto *hommes moyens*.

1.4. Gli artefici della teoria della correlazione

In una serie di conferenze di Karl Pearson sulla storia della statistica, poi raccolte dal figlio Egon in un volume del 1978, egli sostenne:

È impossibile capire il lavoro di un uomo se non si capisce almeno qualcosa del suo ambiente. E il suo ambiente vuol dire lo stato degli affari sociali e politici del suo tempo.

Nei paragrafi precedenti ho inquadrato l'emergere del pensiero statistico in un dato ambiente culturale e sociale; in questi presento alcuni aspetti della vita dei protagonisti di questi avvenimenti, che ritengo rilevanti per apprezzare l'intero processo. Come afferma Stigler (1999, 64) non si può capire il lavoro di uno scienziato "senza giungere a conoscerlo, in qualche misura, come persona."

Francis Galton, il più piccolo di sette fratelli, nacque il 16 febbraio 1822 a Sparkbrook, vicino Birmingham. Suo padre era un ricco banchiere; la madre aveva legami di parentela con persone altolocate. Charles Darwin era suo cugino.

Francis fu educato in varie scuole private e nella famosa *King Edward VI Grammar School*. Sin dalla sua infanzia i suoi genitori prestarono molta attenzione alla sua educazione scientifica. Nelle *Memories of My Life* (1908) ricorda che il padre, ansioso che imparasse

³⁴ Come visto l'opinione più diffusa considera il periodo a cavallo tra l'ottocento e il novecento come il momento di nascita della moderna teoria statistica. Stigler (1999) invece afferma che la disciplina nacque formalmente nel 1933, quando dopo la pubblicazione di un celebre articolo di Egon Pearson e Neyman sui test di ipotesi più efficienti, si aprì la strada alla nuova scuola della statistica induttiva — probabilmente quella più influente e prolifica in seno alla statistica matematica.

³⁵ Stigler (1999) considera personaggi fondamentali Galton, Edgeworth e Pearson. Mackenzie (1981) pone Fisher al posto di Edgeworth. A parte le considerazioni personali di ogni autore, è chiaro che si possono individuare delle persone più influenti in un'impresa, ma risulta difficile, quando si tratta di complessi processi storici di costruzione scientifica, decidere chi sia il più importante. Altrettanto è vero per le figure centrali nello sviluppo della teoria della correlazione. Tuttavia, possiamo affermare che i cinque studiosi menzionati nel testo dettero dare i contributi decisivi.

qualcosa di scienze fisiche, si accordò con il proprietario di una gran farmacia affinché gli permettesse di trafficare per qualche giorno con gli alambicchi nel suo laboratorio.³⁶

Galton studiò inizialmente medicina al *Birmingham Hospital*, e completò gli studi al *King's College*. I suoi genitori volevano che facesse il medico, ma egli cambiò idea e tra il 1840 e il 1843 studiò matematica al *Trinity College* di Cambridge.

Il padre morì nel 1844, quando Francis aveva 22 anni, lasciando una considerevole fortuna ai figli. Non avendo bisogno di lavorare, Galton approfittò per viaggiare in luoghi esotici, e diventò un vero esploratore vittoriano, seguendo l'esempio del suo ammirato cugino Darwin. Nel periodo 1845-1850 visitò l'Egitto, il Medio Oriente e le zone sconosciute del Sudafrica; queste ultime esperienze furono raccolte nel libro *The Narrative of an Explorer in Tropical South Africa*, pubblicato nel 1853.

A partire degli anni sessanta i suoi interessi di ricerca s'indirizzarono sempre più verso il problema dell'eredità dei caratteri fisici e mentali. L'evoluzionismo di *On the Origin of Species*, scritto dal cugino e pubblicato nel 1859, e gli studi statistici di Quetelet in antropometria esercitarono enorme influenza su di lui. Le sue indagini sull'eredità hanno prodotto molte opere notevoli: *Hereditary Genius* (1869), *English Men of Science* (1874), *Inquiries into Human Faculty and Its Development* (1883) e *Natural Inheritance* (1889) — libro che attirò l'attenzione del giovane Karl Pearson.

Come si vedrà nel capitolo 4, queste ricerche sull'eredità sono strettamente collegate ai suoi contributi alla teoria statistica, particolarmente ai concetti di regressione e di correlazione. I temi delle sue indagini furono condizionati da forti interessi eugenetici, manifestati nella convinzione che l'umanità potesse migliorare il suo futuro attraverso il controllo della riproduzione.

La sua concezione eugenetica si può intendere come parte di un movimento più largo: il naturalismo scientifico. In effetti, come tipico esponente del naturalismo scientifico, Galton si confrontò senza esitazioni con le autorità religiose del tempo. Decisamente ateo, egli era convinto di poter dimostrare statisticamente la completa assenza di prove dell'esistenza di Dio (Galton 1869). Galton concepì l'eugenetica come una vera religione naturalistica alternativa a quella cristiana (MacKenzie 1981).

I suoi contributi scientifici, pubblicati in 220 articoli e quindici libri, coprono una vasta gamma di campi, dalla geografia alla psicologia, dall'antropometria alla teoria statistica. Non era un esperto matematico — anche se Pearson sostiene che Galton sottovalutava le sue capacità e conoscenze in questa disciplina — e molti dei suoi progressi hanno un'origine intuitiva, legata alla sua grande curiosità innata. A lui si devono un valore caratteristico di tendenza centrale (la mediana) e uno di variabilità (lo scarto interquartilico); il sistema dei percentili e i concetti di regressione e di correlazione. Fu membro della *Royal Statistical Society* dal 1860, quando ancora era la *Statistical Society of London*, ma la sua partecipazione a questa istituzione fu saltuaria; in tutta la sua vita presentò soltanto tre memorie alla Società. Fu invece più intensa la sua collaborazione alla *Royal Society* e all'*Anthropological Institute*.

Alla fine della sua vita Galton scrisse un romanzo, *Kantsaywhere*, che descrive la sua utopia eugenetica. Morì nel Surrey nel 1911, lasciando parte della sua fortuna — come abbiamo visto — all'*University College*.

Francis Ysidro Edgeworth nacque in Irlanda nel 1845, nel seno di una nobile famiglia trasferita durante il regno di Elisabetta I da Edgeware, nel Middlesex, alla contea di Longford in Irlanda. Alla fine dell'ottocento, il capo della famiglia era l'eccentrico Richard Lovell Edgeworth, padre di ventun figli e amico di Ricardo e di Jeremy Bentham (Kendall 1968). Dal suo sesto figlio e da una giovane catalana nacque Francis Ysidro.

³⁶ In un certo senso, questa preoccupazione del padre per socializzare scientificamente Galton sin dalla prima infanzia equivale alla socializzazione religiosa che altri genitori volevano per i loro figli; andare in laboratorio ha pertanto un significato simbolico simile ad andare in chiesa. Come si vedrà più avanti, non a caso Galton difese la scienza, nella sua crociata antimetafisica, come la nuova religione. E nella sua campagna eugenetica, propose persino che lo Stato inviasse dei veri “sagrestani eugenetici” in giro per il paese.

Studiò lingue straniere nel *Trinity College* di Dublino; poi si trasferì a Oxford con una borsa di studio, e vi si laureò in *Literae Humaniores* nel *Balliol College* nel 1869. Il suo interesse per la matematica emerse più tardi: Stigler (1999) afferma che verso la fine degli anni 70 egli perseguì un programma autonomo di apprendimento matematico equivalente a una laurea universitaria. Nei primi anni ottanta la sua preparazione copriva tutto lo spettro delle matematiche del tempo.

Il suo primo incarico accademico fu come insegnante di greco nel *Bedford College* dell'Università di Londra. Nel 1880 divenne insegnante di logica, e nel 1888 professore di economia politica al *King's College* di Londra. Nel 1891 fu designato *Drummond Professor* di economia politica a Oxford. Nel 1922 andò in pensione con il titolo di professore emerito. Fu presidente della *Royal Statistical Society*, vicepresidente della *Royal Economic Society* e *Fellow* della *British Academy*. Fu il primo editore dello *Economic Journal*, una delle riviste economiche più prestigiose a livello internazionale — posto che conservò dal 1891 fino alla morte nel 1926. Gli successe Keynes, già da qualche tempo suo co-direttore.

Il primo libro di Edgeworth, all'età di 32 anni, fu *New and Old Methods of Ethics* (1877), sicuramente poco comprensibile per i filosofi in generale, e per gli esperti di etica in particolare, dato l'estensivo uso di integrali ed altri simboli matematici. Nel 1881 uscì *Mathematical Physics*, che malgrado il titolo si occupa di propugnare l'applicazione della matematica alle scienze sociali. Vi sono chiaramente delineate le due direttrici lungo le quali si muoveranno i contributi futuri di Edgeworth: il concetto di utilità, attraverso il quale rende applicabili le matematiche ai problemi economici³⁷; e il concetto di *belief* (convinzione) col quale — controponendosi all'idea di conoscenza certa della Grecia classica — giustifica l'applicazione degli strumenti statistici ai fenomeni empirici, fisici o umani.

Edgeworth cominciò il lavoro alla teoria statistica indipendentemente da Galton, che poi gli suggerì di lavorare sui problemi della correlazione. È probabile che questo interesse gli venisse da Jevons, molto ammirato da Edgeworth e suo vicino di casa a Londra.

Nel 1887 pubblicò *Metretike*, un libro dedicato al metodo per misurare la probabilità e l'utilità, che intendeva la probabilità come la base del ragionamento induttivo (Kendall 1968).

I suoi contributi statistici più importanti sono nel campo dei numeri indici e della teoria delle distribuzioni di frequenza (allora nota come generalizzazione della legge degli errori). Stigler (1999) suggerisce una certa priorità di Edgeworth nello sviluppo della moderna analisi della varianza, compiuto molti anni dopo da Fisher. Per quanto i suoi contributi alla teoria della correlazione non generano consenso fra i contemporanei, almeno per i suoi interventi nei dibattiti e il rispetto che la sua figura ispirava, molti esperti affermano che "incoraggiando, criticando e applaudendo sin dall'inizio il lavoro degli statistici inglesi sulla correlazione" (Keynes 1926, 148) esercitò una grande influenza. La sua influenza fu sottile e profonda, e tuttavia è rimasto una delle meno conosciute tra le maggiori figure della storia della statistica (Stigler 1999).

La sua cultura classica era evidente; citava Orazio o Platone in una pagina e deduceva un'equazione differenziale nella seguente. Nel secolo XX era uno dei pochissimi che potevano ancora citare i classici in ogni occasione e in qualunque ambiente (Keynes 1926). Aveva predilezione per i polisillabi, ed era sempre pronto a lanciare un nuovo termine. La sua prosa non fu mai semplice, e la presentazione delle sue idee piuttosto disordinata, al punto che Pearson lo chiamò *Magister Obscurantissimus*. Keynes (1926, 146) afferma che in realtà Edgeworth aveva il segreto dello "strano ma attraente amalgama di poesia e pedanteria, scienza e arte, ingegno e erudizione"; ma nel tentativo di correggere il suo stile naturale riuscì solo ad "accrescere l'oscurità" delle sue presentazioni intellettuali. Per Stigler (1999) lo stile di

³⁷ Due erano allora i principali fautori dell'applicazione della matematica all'economia: Marshall e Edgeworth. Keynes (1926) sostiene che Marshall, che aveva studiato la disciplina a Cambridge, abbia meramente tradotto l'economia in linguaggio matematico, ma non abbia introdotto alcuna innovazione. A suo avviso, Edgeworth, pur essendo autodidatta, sapeva invece usare la matematica per fare progredire la scienza economica.

Edgeworth richiede un grande impegno al lettore, ma lo sforzo poi viene ricompensato e i suoi scritti appaiono incantevoli.

La sua vita non si limitava al lavoro intellettuale; era un grande appassionato di camminate, alpinismo, golf e canottaggio. D'estate nuotava tutte le mattine nelle gelide acque di Parson's Pleasure a Oxford, persino a ottanta anni d'età (O'Connor e Robertson 2003).

Mori di polmonite il 13 febbraio 1926, a Oxford. In quanto personaggio assolutamente singolare, non era discepolo di altri e non fondò alcuna scuola. Bowley fu il suo unico erede intellettuale. Ciononostante, è una figura di gran rilievo nella storia della statistica, cruciale nel consolidare il suo status come disciplina scientifica (Kendall 1968). Non pochi lo considerano il più grande statistico dell'ottocento (si veda ad esempio FitzPatrick 1960; Stigler 1999).

Carl Pearson nacque il 27 maggio del 1857. I suoi genitori erano originari dello Yorkshire. I Pearson erano discendenti di possidenti quaccheri, e gli antenati della madre erano marinai, capitani e proprietari di una flotta a Hull. Il padre era un avvocato, molto impegnato e deciso a salire nella scala sociale (MacKenzie 1981). Aveva un profondo interesse per gli studi storici, ereditato poi dal figlio Carl.

Carl era un ragazzo serio e studioso. Nel 1875 vinse una borsa di studi del *King's College* di Cambridge e nel 1879 si laureò tra i migliori nei *Cambridge Mathematics Tripos*.³⁸ Poco dopo andò in Germania per proseguire gli studi. A Heidelberg studiò fisica con Quincke, filosofia con Bruno Fischer e diritto romano con Mommsen. In questo periodo s'interessò alla storia tedesca — in particolare alla Riforma e alla figura di Lutero.³⁹

Nel 1880 pubblicò il suo primo libro, *The New Werther*, con il pseudonimo 'Loki'. Esso testimonia una crisi d'identità giovanile: "corro dalla scienza alla filosofia, e dalla filosofia ai nostri vecchi amici poeti; e allora, stanco dell'eccessivo idealismo, il mio senso pratico riemerge e ritorno alla scienza".⁴⁰ Questa frase premonitrice, pronunciata da *Arthur*, il personaggio del libro che i critici considerano autobiografico, fornisce un indizio del percorso filosofico seguito da Pearson, che sbocca in radicale scientismo anti-metafisico.⁴¹

Ma il frutto delle sue riflessioni e delle prime ricerche tedesche si manifesta in *The Ethic of Freethought*, pubblicato nel 1888. In quest'opera si intravede già il ricercatore perfezionista, che lavora con cura e minuzia. Il libro è una compilazione di articoli in cui sua la nascente posizione politica si presenta chiara. Tra l'altro, anticipando in qualche modo *Die protestantische Ethik und der Geist des Kapitalismus* di Max Weber, pubblicato nel 1904, il libro avanza l'ipotesi dell'influenza luterana sul benessere economico della Germania contemporanea.

La povertà dell'Inghilterra vittoriana e la compiaciuta superficialità di Cambridge avevano incominciato ad avvicinarlo al socialismo in una versione personale assolutamente eterodossa. Il capitalismo *laissez-faire* era per Pearson un sistema inefficiente di concorrenza anarchica, da sostituire con un sistema di pianificazione statale, con tutto il capitale concentrato nelle mani dello Stato. Il cambiamento non doveva risultare da una rivoluzione, ma da una transizione graduale, con dei risarcimenti ai capitalisti per le loro perdite (MacKenzie 1981). Il conflitto di classe era da evitare, cercando l'armonia sociale. Non deve

³⁸ Esame di laurea dell'Università di Cambridge.

³⁹ La cultura tedesca influenzò notevolmente la vita di Pearson. La sua impostazione filo germanica si manifestò persino nel cambio del nome: Karl al posto di Carl (Norton 1978). La lettera 'k' esercitò sempre un certo fascino, e non solo per la sua ammirazione della Germania; egli era consapevole, e più volte lo fece notare, che la lettera greca 'κ' non equivale alla moderna 'c', e per questo motivo decise di chiamare *Biometrika* la sua rivista.

⁴⁰ Citato da E. Pearson (1938 /1948, 31).

⁴¹ Arthur (o il giovane Carl) non aveva ancora sviluppato una concezione idealizzata di scienza. Egli denuncia la tirannia religiosa di tanti secoli, ma nutre pari diffidenza per la scienza e la sua prepotente autorità. Una posizione scettica nel confronto della scienza che il Pearson maturo non avrebbe sottoscritto. Peraltro, l'impostazione antireligiosa si era già manifestata quando da studente si era ribellato contro le lezioni obbligatorie di religione nel *King's College* dell'Università di Cambridge.

sorprendere, pertanto, la sua simpatia per il fabianesimo⁴² e la sua amicizia con alcuni dei suoi più noti esponenti come Sidney Webb e George Bernard Shaw.

Temeva il proletariato come una classe potenzialmente insurrezionale. Il cambiamento doveva venire dall'alto, e la società stratificata in base alla ricchezza doveva essere sostituita da una società stratificata sulla base dell'educazione e della cultura. MacKenzie (1981, 78) gli attribuisce un 'socialismo dei professori', sulla scorta dei *Katheder-Sozialisten* tedeschi che tanto ammirava.

Pearson fu un ardente darwinista, ma nell'epoca aurea dell'imperialismo, il darwinismo interno di Spencer divenne in Pearson un darwinismo esterno; una giustificazione della concorrenza economica e militare tra le nazioni avanzate e del loro sfruttamento dei popoli "inferiori".⁴³ Di solito si pensa al socialismo come opposto all'imperialismo, ma nell'Inghilterra della fine dell'ottocento le due tradizioni trovavano una confluenza nelle posizioni dei fabiani e di Pearson. La riforma sociale collettivista era vista come un imperativo per assicurare l'efficienza nazionale nella lotta inter-imperialista (MacKenzie 1981).

Nel 1884, all'età di ventisei anni, Pearson divenne professore di matematica applicata e meccanica all'*University College* di Londra, istituzione in cui avrebbe svolto tutta la sua carriera didattica e di ricerca. All'Università entrò in contatto con lo zoologo Weldon, che aveva incominciato a usare i nuovi strumenti statistici di Galton nella ricerca empirica, e cercava un partner competente in matematica per approfondire questa direttrice. Il rapporto con Weldon — che divenne presto amicale — e l'entusiasmo provocato dalla lettura di *Natural Inheritance* di Galton, convinsero Pearson della possibilità di applicare il ragionamento matematico al di là delle scienze fisiche. Weldon morì giovane, pochi anni dopo, ma questa collaborazione aveva segnato già il futuro della vita accademica di Pearson: il tentativo di dimostrare, attraverso l'applicazione dei nuovi strumenti della teoria statistica, le ipotesi della biologia evoluzionista. A questo fine fondò il Laboratorio di Biometria e poi, unendolo al Laboratorio Eugenetico di Galton, creò il Dipartimento di Statistica Applicata, primo nel suo genere. Insieme a Weldon fondò la rivista *Biometrika*, dedicata alla statistica applicata. Molti anni dopo fonderà *Annals of Eugenics*, ultimo tentativo di promuovere un'eugenetica "scientifica" sulla linea voluta da Galton. Un programma eugenetico⁴⁴ era la conseguenza logica dell'applicazione della teoria evoluzionista a un mondo caratterizzato dalla concorrenza internazionale (MacKenzie 1981).

Pochissimi suoi contemporanei potrebbero rivaleggiare con i molteplici interessi intellettuali di Pearson manifestati in pubblicazioni di poesia, storia dell'arte, germanistica, filosofia, epistemologia, politica, biografia, matematica, fisica, biologia, e soprattutto, teoria statistica. I suoi contributi in quest'ultima disciplina sono innumerevoli: la teoria della correlazione, il test χ^2 , il coefficiente η^2 , i fondamenti della moderna analisi in componenti principali.⁴⁵ Soprattutto, il contributo all'emergere, consolidare e istituzionalizzare la statistica come disciplina. Non pochi suoi contemporanei — basti Yule come esempio — e statistici attuali, lo considerano il padre fondatore della statistica moderna (Porter 1986; Norton 1978),

⁴² Hobsbawm (1968) definisce al fabianesimo come l'espressione politica della nascente classe media di colletti bianchi ed impiegati professionisti, che promuoveva un socialismo elitario, condotto da esperti ed amministratori, senza impegni politici nel confronto della classe operaia. La differenza tra il pensiero di Pearson e quello dei fabiani è che questo promuoveva il suffragio universale, al quale Pearson era contrario.

⁴³ Non deve sorprendere che questa posizione fosse criticata da Lenin, intransigente fustigatore, almeno sul piano retorico, dell'imperialismo europeo di allora.

⁴⁴ Il programma eugenetico di Pearson conteneva molte proposte fortemente contestate, per esempio l'intervento statale sulle gravidanze. I propagatori di "esseri umani non necessari" dovevano essere controllati (Pearson 1888, 433).

⁴⁵ Pearson è ritenuto il padre dell'analisi in componenti principali, anche se si dovrebbero considerare gli importanti apporti di Edgeworth e Macdonnell (si veda Di Franco e Marradi 2003, cap. 2).

il responsabile dello sviluppo della statistica come disciplina scientifica (Eisenhart 1974) o il creatore della statistica applicata.⁴⁶

Pearson fu una figura affascinante per le sue contraddizioni: socialista e allo stesso tempo imperialista, eugenista, difensore degli animali e amante della natura, libero pensatore e vicino al movimento femminista.⁴⁷ In effetti, mantenne sempre contatti amichevoli con il femminismo. Fu fondatore e membro attivo di un club⁴⁸ di uomini e donne progressisti, dedicato alla libera e "imparziale" discussione del rapporto tra i sessi. In questo club conobbe Maria Sharpe, la segretaria, che poi diventerà la sua prima moglie e la madre dei tre figli.

Le sue riserve sulla religione tradizionale, la sua aperta impostazione politica di tipo socialista, e la sua critica al rapporto tra i sessi nella società vittoriana, lo resero una figura marginale, ai "confini della società rispettabile" del tempo (MacKenzie 1981, 73).

Yule (1936), suo grande amico e collaboratore negli anni 90, poi allontanatosi, lo ricordava come un uomo di personalità singolare e forte, che sviluppava rapporti stretti con gli studenti. Il figlio, sforzandosi di fornire un giudizio imparziale, lo ritrae come un uomo eccezionale, instancabile lavoratore, ma con una tendenza all'intolleranza intellettuale e al dogmatismo (E. Pearson 1938), caratteristica che si evidenzia nei suoi molteplici confronti con colleghi sostenitori di posizioni diverse dalla sua, per esempio Yule e Fisher, o la Scuola Mendelista in Biologia. Morì nel 1936 nella sua casa di campagna. Già nel 1933 era andato in pensione, rammaricandosi di non essere in grado di seguire gli sviluppi della disciplina che aveva contribuito a istituzionalizzare, e con il timore che diventasse un ramo della matematica pura, abbandonando l'impostazione pratica che i *founding fathers* avevano sempre sostenuto.

Charles Spearman, nato a Londra nel 1863, in gioventù decise di arruolarsi nell'esercito britannico per poter studiare e fare ricerca, approfittando del tempo libero che in teoria esso gli avrebbe garantito. Da ufficiale partecipò alla guerra birmana e alla guerra dei Boeri. Dopo circa vent'anni lasciò la vita militare, e solo allora riuscì a riprendere un piano sostenuto di studi, finendo il dottorato di filosofia a 41 anni. I suoi studi dottorali si svolsero in Germania, sotto la supervisione di Wilhelm Wundt, che aveva fondato a Lipsia il primo laboratorio di psicologia sperimentale (William *et al.* 2003). Al tempo molti psicologi inglesi e americani facevano il loro apprendistato nella ricerca empirica con Wundt (Di Franco e Marradi 2003, cap. 2). Spearman fu influenzato anche da Galton, che aveva già avanzato l'idea di misurare le facoltà dell'intelletto, e dalla biometria di Pearson.

Dal 1907 fino al 1931 insegnò all'*University College* di Londra, nello stesso ambito istituzionale in cui lavorava Pearson. Occupò la cattedra di psicologia sperimentale e poi quella di filosofia della mente e logica. Fu *fellow* della *Society of British Psychology* e di altre prestigiose istituzioni europee e americane: la *Société Française de Psychologie*, la *Deutsche Psychologische Akademie* e la *Kentucky Academy of Science*.

I più notevoli contributi di Spearman si possono classificare in due grandi aree: la metodologia statistica e la psicologia sperimentale, e più particolarmente la teoria dell'intelligenza. Come si vedrà nel cap. 4, le aree sono strettamente collegate, dato che le sue innovazioni metodologiche derivarono dalla necessità di risolvere problemi della ricerca sull'intelligenza. Peraltro, i contributi più durevoli di Spearman nell'ambito della psicologia sono quelli relativi alla teoria dei test mentali, che hanno una forte valenza metodologica. Nel classico libro di Gulliksen, *Theory of Mental Tests* (1950), i primi lavori di Spearman sono considerati la base di tutte le formule ancora usate nella teoria dei test.

⁴⁶ Questo è, per esempio, il titolo conferito a Pearson dalla traduzione spagnola della biografia scritta dal figlio Egon.

⁴⁷ MacKenzie (1981) racconta che Pearson era amico di Olive Schreiner, la scrittrice sudafricana riscoperta dal movimento femminista negli ultimi 30 anni. In un ambiente scientifico dominato dagli uomini, Pearson era sempre disposto a promuovere le donne di talento. Sono moltissime le sue collaboratrici scientifiche; tra le più note, Alice Lee ed Ethel Elderton.

⁴⁸ Si tratta del *Men and Women's Club*, fondato da Pearson, Parker e dalle sorelle Maria, Laetitia ed Elizabeth Sharpe.

La correlazione di rango, direttamente rilevante alla storia del concetto e degli strumenti della correlazione, sarà approfondita nel cap. 3. Un contributo ancora più fondamentale di Spearman è l'analisi fattoriale.⁴⁹ In quanto basata su una matrice di correlazioni, e ispirata dalla necessità di dimostrare il carattere unitario dell'intelligenza (problema che informò tutta la riflessione di Spearman sul concetto di correlazione), essa sarà richiamata nel cap. 4. Inoltre, Spearman è considerato l'ideatore della tecnica *test-retest* per controllare l'affidabilità dei dati — corollario della sua permanente attenzione alla loro qualità.⁵⁰

Il suo contributo trova una sintesi in *The Abilities of Man* (1927), in cui presenta in modo sistematico la teoria fattoriale dell'intelligenza: tutte le capacità intellettuali hanno un fattore comune 'g'⁵¹ ed un fattore specifico 's'. Spearman continuò a fare ricerche fino alla morte, avvenuta nel 1945.

George Udny Yule nacque nel 1871 in Scozia e morì a Cambridge nel 1951. La sua famiglia, di antiche radici scozzesi, composta da ufficiali, funzionari pubblici e studiosi, godeva di una situazione piuttosto privilegiata. Il nonno William era un profondo conoscitore della cultura araba e persiana. Il padre George e lo zio Henry erano stati nominati cavalieri dalla regina Vittoria.

Yule studiò ingegneria civile all'*University College* di Londra dal 1887, e dopo cinque anni andò a Bonn per fare ricerca sulle onde elettriche sotto la supervisione di Hertz. Di questo periodo sono i suoi due primi articoli, entrambi dedicati alla fisica sperimentale. In seguito non si occupò più di questi argomenti. Kendall, suo collaboratore e amico, afferma (1952) che questo primo addestramento non sembrava aver lasciato tracce nello stile di pensiero di Yule, a parte la sua perizia nella — e preferenza per la — rappresentazione grafica.

Nell'estate del 1893 ritornò a Londra, e poco dopo accettò un'offerta di Pearson per collaborare alla sua cattedra di matematica applicata all'*University College*. Yule scoprì in Pearson "un maestro entusiasmante, e poco dopo stava dando dei veri contributi originali alla teoria statistica" (Kendall 1952, 156). Nel 1895 fu nominato *fellow* della *Royal Statistical Society*. La sua collaborazione con quest'istituzione, durata più di cinquant'anni, sarebbe stata molto fruttifera. Nel 1907 ne fu insignito della massima onorificenza, la *Guy Medal*, e dal 1924 al 1926 tenne la carica di presidente. Inoltre, fu nominato membro onorario di molte società statistiche straniere.

Malgrado il suo impegno all'Università, Yule dovette accettare altri lavori nella pubblica amministrazione al fine di arrotondare il suo magro stipendio. Ma il suo interesse per la ricerca e la didattica della statistica non diminuì: continuò a insegnare e a pubblicare. Nel 1911 uscì *Introduction to the Theory of Statistics*, il primo vero manuale generale di teoria statistica moderna uscito nel mondo.

Nell'anno successivo avvenne una svolta nella vita di Yule. L'Università di Cambridge gli offrì un posto come docente, e in seguito egli divenne il primo *reader*⁵² in statistica nella storia della prestigiosa università. Il rapporto con il *St. John's College* di Cambridge segnò la sua vita d'allora in poi; ne divenne *fellow* nel 1922.

I suoi contributi maggiori alla statistica sono nella teoria della correlazione e dell'associazione, che tratterò nel cap. 3, e che furono più rilevanti nel periodo 1900-1912. Ma Yule dette anche dei contributi alla ricerca sociale empirica in sociologia economica ed

⁴⁹ Spearman è universalmente riconosciuto l'ideatore dell'analisi fattoriale. Per una storia breve e precisa dello sviluppo di questa tecnica si veda Di Franco e Marradi (2003).

⁵⁰ Di Franco e Marradi (2003) ritengono giusto considerarlo il diffusore della tecnica, ma non l'ideatore, che più probabilmente fu Ebbinghaus (1885).

⁵¹ Il test d'intelligenza delle matrici progressive di Raven, ancora oggi utilizzato, è stato costruito lavorando fianco a fianco con Spearman, e viene considerato una delle operativizzazioni più accurate di g (Williams *et al.* 2003).

⁵² Docente universitario di status inferiore a quello di un professore ordinario.

epidemiologia, nei quali applicò in modo pionieristico i nuovi strumenti di analisi statistica che aveva contribuito a sviluppare.⁵³

A partire degli anni trenta, la teoria statistica stava seguendo un percorso sempre più tecnico e matematico, e Yule, che non era un grande esperto di matematica, cominciò a sentire che il suo tempo era passato e che non poteva più partecipare ai nuovi sviluppi della disciplina che tanto aveva contribuito a consolidare. Questo fatto, sommato a una malattia cardiaca intervenuta nel 1931 che lo lasciò parzialmente invalido, lo gettò in uno stato di depressione. La morte di Pearson nel 1936 approfondì questa sensazione, e con malinconia, malgrado gli scontri personali che avevano avuto con lui sin dai primi anni del secolo, scrisse a Kendall: "sento che l'era Karlovingia⁵⁴ è finita, e che [nella nuova] non posso giocare una parte" (Kendall 1952, 157).

Nel 1937 uscì una nuova edizione del suo classico manuale, ora rivista e ampliata da Kendall⁵⁵, che aveva conosciuto per caso nel 1935. Il libro ebbe un grandissimo successo di vendita e fu tradotto in varie lingue. Questo diede un nuovo impulso alla vita accademica di Yule, che ricuperando un interesse di famiglia per la filosofia e gli studi umanistici, cominciò a fare ricerche nell'area dell'analisi linguistica e letteraria. Applicando le sue care tecniche statistiche alla lunghezza delle frasi e alla frequenza di ogni parola, divenne un precursore della moderna analisi del contenuto nella sua impostazione quantitativa.⁵⁶ Il suo ultimo libro, *The Statistics of Literary Vocabulary* (1944) è il frutto di queste ricerche iniziate già negli anni trenta (Williams 2004).

Alla fine degli anni quaranta, le sue condizioni di salute peggiorarono e dovette passare gli ultimi anni di vita in una casa di riposo della città di Cambridge, fino alla morte, avvenuta nel 1951.

Yule apparteneva, come Edgeworth, a una famiglia altolocata in declino, la cui ricchezza era già svanita quando egli avrebbe potuto goderla. Politicamente è stato considerato un conservatore, e come tale contrario all'eugenetica, allora molto influente nel pensiero statistico (MacKenzie 1981). La sua personalità è stata descritta dai suoi amici come affabile; un uomo pratico, gradevole, di buon senso, non dogmatico, e con il tipico *humour* britannico. La sua versatilità si manifestò senz'altro nei suoi originali contributi alla teoria statistica e alla ricerca empirica.

⁵³ Alcuni di questi contributi si presenteranno nel cap. 4.

⁵⁴ Molti statistici di quel tempo — tra i quali Yule e Greenwood — usavano il termine 'Karlovingio', in chiaro riferimento a Karl (Pearson).

⁵⁵ Kendall aggiunse al libro i nuovi sviluppi della teoria del campionamento, che — come aveva presagito Pearson — era divenuta una delle aree più importanti e promettenti della teoria statistica.

⁵⁶ La paternità di questo contributo non gli è comunemente riconosciuta; ma il fatto è che non ci sono altri casi precedenti di applicazione della statistica all'analisi dei testi.

CAPITOLO 2

Epistemologia, ontologia e metodologia nella teoria statistica inglese: l'elogio della correlazione e la critica dell'idea di causalità.

The law of causation is a conceptual figment extracted from phenomena, it is not of their very essence. The actual problem before mankind is a far wider than that of "causation" [...] It is [the] conception of correlation between two occurrences embracing all relationship [...] the wider category by which we have to replace the old idea of causation.

Karl Pearson (1892/1957, 157)

2.1. L'ambiente intellettuale dell'ottocento: la scienza in discussione.

L'era vittoriana è stata frequentemente descritta come l'età della scienza, con un vero coro di uomini osannanti. Ciononostante, la scienza non aveva ancora avuto un completo successo nelle sue rivendicazioni; il suo ruolo sociale, culturale e intellettuale era oggetto di acceso dibattito (Kjærgaard 2002).

Nei primi decenni dell'ottocento molti autori inglesi tentavano di monopolizzare il modo "giusto" di definire la scienza. Non si era ancora consolidata una comunità scientifica che potesse arrogarsi legittimamente l'autorità sull'insieme di regole per la pratica scientifica (Winter 1997).

In una ricostruzione *ex post* risulta relativamente facile trovare nel pensiero e nelle scoperte scientifiche che segnarono l'inizio della modernità il seme che avrebbe finito per distruggere ogni traccia della visione teologico-metafisica della realtà. Ma nella sua infanzia, l'emergente scienza moderna dovette adattarsi ad una concezione che era ancora dominante. Un chiaro esempio ne è stato ricordato nel capitolo precedente: la teoria della probabilità sfidava la visione deterministica allora dominante; tuttavia, persino i suoi principali fautori si sforzarono di renderla compatibile con quella stessa concezione che era destinata ad erodere. Nell'ottocento, la tensione tra due cosmologie — la teologico-metafisica e la scientifico-naturalistica — stava raggiungendo il culmine.

Fayter (1997) afferma che questa crisi cominciò nel 1798 con la pubblicazione di *Essays on the Principles of Population* di Malthus, e arrivò al culmine con l'opera di Darwin, *On The Origin of Species* (1859), che situava l'uomo in un contesto cosmico che negava la storia biblica della creazione.⁵⁷

Gli intellettuali conservatori si sentivano sotto la minaccia posta dalla diffusione del pensiero naturalista e di quello materialista. D'altra parte, gli uomini di scienza temevano che il progetto scientifico e la diffusione della sua ideologia fallisse, e rivendicavano enfaticamente — proprio perché le loro pretese erano contestate — il diritto esclusivo di definire cosa fosse la scienza (Kjærgaard 2002).

Il conflitto tra naturalismo e visioni tradizionali del mondo — tipico del periodo medio e tardo vittoriano — non era pertanto un mero confronto d'idee astratte; in discussione era

⁵⁷ Il darwinismo comportava una difesa delle nascenti professioni scientificamente orientate contro l'ancora potente *establishment* della chiesa anglicana (MacKenzie 1981).

l'autorità di pronunciarsi sul cosmo, sulla società e sull'essere umano — e ovviamente, la capacità di appropriarsi dei vantaggi terreni che derivano di tale autorità.

La frattura del contesto intellettuale comune assicurato dalla teologia naturale condusse allo sviluppo della specializzazione e della professionalizzazione della scienza verso la fine dell'ottocento (Lightman 1997). Una comunità scientifica nel senso moderno cominciava a delinearsi; e di conseguenza, in quanto costituita da esperti e da scienziati professionisti, essa voleva assumersi l'incontestabile legittimità di pronunciarsi sul campo della scienza secolarizzata (Shapin 1990).

Non dovrebbe sorprendere, pertanto, che il pensiero della nascente comunità di statistici matematici, specialmente quello di Pearson — l'unico di loro che riuscì ad articolarlo in un'opera epistemologica — rispecchi in molti sensi questa corrente di laicismo e scetticismo religioso (Norton 1978).

The Grammar of Science (1892) è infatti il più importante sommario delle concezioni epistemologiche scritto nel periodo; un vero elogio della nuova visione della scienza e della società (Levine 1996). Ernst Nagel, scrivendo l'introduzione a un'edizione americana del 1957 di questa classica opera di Pearson, indica — in uno spirito paradossalmente più attribuibile a Fleck (1935), Kuhn (1962) o agli odierni sociologi della scienza — che frequentemente "perfino i più distinti uomini di scienza non possiedono idee chiare circa le nozioni da loro usate o circa la logica dell'indagine praticata" (Nagel 1957, v). Il commento di Nagel non poteva essere più opportuno: molti scienziati, persino quelli più prestigiosi, non avevano secondo Pearson idee chiare sulla scienza, e finivano molte volte per riprodurre atteggiamenti metafisici propri della cosmologia teologica. Pearson percepiva come nemici della scienza non soltanto i clerici tradizionalisti, ma anche non pochi scienziati — oggi si direbbe "normali" — che non si fermavano ad analizzarne i fondamenti ontologici, epistemologici e metodologici.

Nella prefazione alla prima edizione, scritta nel gennaio del 1892, Pearson afferma che la scienza doveva "portare avanti una difficile guerra contro la metafisica e il dogma". Ma per compiere quest'impresa occorreva "esaminare attentamente i suoi fondamenti", senza nascondere i difetti (Pearson 1892/1957, xvii). A suo avviso, il linguaggio della scienza dominante — la fisica — usato anche in altre discipline, compresa la sociologia, che egli allora riteneva legata alla biologia, richiedeva un riesame. I testi elementari di scienza non avevano un valore educativo; non incoraggiavano la chiarezza logica né l'esercizio del metodo scientifico. Anche molti fisici si trovavano intrappolati nelle maglie della teologia naturale e dello spiritualismo. Un chiaro esempio era il *Chapter of Science* di Stuart (1883), in cui si affermava che nessun risultato della scienza screditava una singola parola della Bibbia. Quest'opera scientifica era per Pearson ancora più retrograda della teologia di Cambridge, che almeno — ironizzava — già non dava credito alla storia dell'Arca di Noè (1892/1957, 135).

Pearson trovò nella scienza l'assoluto che la religione (abbandonata da giovane) e la metafisica (visitata nella prima gioventù tramite la filosofia e la letteratura del romanticismo), non erano state in grado di fornirgli in modo soddisfacente. Si ricordino a questo proposito le parole del giovane Arthur, protagonista del *New Werther* (1880), personaggio autobiografico della sua prima gioventù: "corro dalla scienza alla filosofia, e dalla filosofia ai nostri vecchi amici poeti; e allora, stanco dell'eccessivo idealismo, il mio senso pratico riemerge e ritorno alla scienza".

Pearson diviene così la voce principale della scienza positivista della fine dell'ottocento (Levine 2000). Il suo discorso epistemologico e metodologico, in cui la tradizione empirista britannica si amalgama armoniosamente con il postkantismo tedesco e con qualche elemento del razionalismo matematizzante di Descartes⁵⁸, tentava di eliminare la

⁵⁸ Secondo Norton (1978) *The Grammar of Science* evidenzia la familiarità di Pearson con la filosofia empirista inglese — particolarmente Berkeley — e con quella tedesca — in particolare Mach. Per MacKenzie (1981) la filosofia della scienza di Pearson emerge gradualmente dalle prime riflessioni su Kant al contatto con le idee di Clifford e di Mach. Tuttavia, la sua fonte epistemologica è la tradizione empirista. Levine (1996; 2000)

confusione che imputava tanto al pensiero metafisico quanto a quello scientifico del suo tempo.

Pearson presenta il suo saggio come un'opera intesa a stimolare il libero pensiero dei lettori; una fonte in cui i giovani studenti possano incontrare una scienza libera dalla metafisica, orientata ad esaltare il suo valore sociale e a favorire il suo apprezzamento. Peraltro, le sue idee suonavano allora piuttosto radicali, ed egli stesso lo riconosce ringraziando tutti quelli che, moderando il suo ardore, l'avevano aiutato a rendere il libro più digeribile per i suoi contemporanei. Ciononostante Lord Balfour, un influente personaggio della cultura vittoriana, reagì pesantemente contro le tesi del libro; sostenendo che la scienza non aveva più autorità in materia di conoscenza che le grandi tradizioni del pensiero etico e teologico (Levine 2000). Il suo saggio *The Foundations of Belief* (1894) è un'aperta critica della validità del metodo scientifico in tutte le aree dell'esperienza, e in particolare della crociata di Huxley a favore del darwinismo.⁵⁹ Pearson rispose irosamente, dando del reazionario e dell'ignorante al Lord.

Come sostiene Pearson nella prefazione alla seconda edizione (scritta alla fine del 1899) malgrado le critiche di Balfour e di altri, la visione proposta dal libro aveva avuto successo. Ma il conflitto tra scienza e dogma permaneva, perché "conoscere domanda sforzo, ed è intellettualmente più facile rinunciare, accettando di percorrere la strada che copre ciò che è sconosciuto con il velo dell'indefinibile".

La battaglia di Pearson, Huxley e altri per stabilire l'autorità della scienza in una cultura ancora predominantemente teologica si appoggiava su una fede illuminista nella razionalità come via per il miglioramento della società. Questo obiettivo, al quale Pearson ricorre per giustificare la seconda edizione di *Grammar*, apparsa nel 1900, prefigura la difesa di una ricerca scientifica libera delle intrusioni autoritarie, esposta da Popper in *The Open Society and its Enemies* (1945). L'ideale dell'illuminismo, l'impegno per la conoscenza e la razionalità come mezzi di progressivo miglioramento della condizione umana, trovò uno dei suoi ultimi sbocchi nel positivismo scientifico (Campelli 1999).

Per Pearson la capacità di promuovere il benessere della società umana, aumentare la felicità sociale e rafforzare la sua stabilità, era l'unica giustificazione accettabile di qualsiasi istituzione sociale o forma di attività umana; incoraggiando la ricerca e la diffusione della conoscenza scientifica si sarebbero formati cittadini più consapevoli e una società più stabile:

La scienza moderna, in quanto allenamento della mente all'esatta ed imparziale analisi dei fatti, è un tipo di educazione specialmente adatto alla promozione di una sana cittadinanza (Pearson 1892/1957, 9).

Il progresso sociale, a sua volta, serviva a prepararsi per un'inevitabile competizione internazionale per la sussistenza. Questa è la conclusione che Pearson trae dall'applicazione delle idee di Darwin all'ambito sociale e politico. Pearson è, come detto, un promotore del darwinismo sociale "esterno" (o internazionale)⁶⁰: la lotta per la sopravvivenza, in questa fase storica, già non si verifica al livello di uomo contro uomo, ma a quello di nazione contro nazione (Norton 1978). Contrariamente a Spencer, molti degli statistici britannici della fine dell'ottocento ritenevano che il *laissez-faire* in materia riproduttiva avrebbe condotto alla

sottolinea l'influenza del pensiero di Kant nell'anti-realismo di Pearson, e trova un elemento cartesiano nel tentativo di eliminare tutto ciò che sia dubbio e oscuro dai fondamenti della conoscenza. Riddle (1958) rileva tre componenti principali nell'epistemologia di Pearson: l'empirismo britannico, un'attenzione di matrice kantiana per il ruolo della mente nell'organizzazione e interpretazione delle sensazioni, e una fede cartesiana nella matematica come la chiave del pensiero scientifico organizzato.

⁵⁹ La sua critica sembra precorrere quella di Feyerabend (1970), solo che era articolata da un uomo ritenuto allora conservatore e tradizionalista. La stessa posizione che molti ora attribuiscono agli ortodossi difensori della visione standard della scienza nei confronti delle odierne critiche postmoderne.

⁶⁰ Si ricordi che c'erano allora due versioni del darwinismo sociale: Spencer difendeva una visione individualista e di libera concorrenza; Pearson e altri davano invece un'interpretazione collettivista (Norton 1978).

proliferazione degli incapaci. La scienza doveva intervenire per evitare la degenerazione della specie. Il darwinismo applicato ai problemi sociali fornì per molti anni la *raison d'être* del lavoro statistico di Pearson (MacKenzie 1981).

Una delle forme più importanti d'intervento era proprio la diffusione della conoscenza scientifica. Col passare del tempo, la concezione epistemologica positivista guadagnava consensi: "Al presente, il rischio che i dogmi metafisici ostacolino la ricerca scientifica non è molto grande. Sono passati i tempi in cui la filosofia hegeliana minacciava di soffocare la scienza nascente in Germania", scrive Pearson nell'ultima revisione del libro, compiuta nel 1911 (1892/1957, 17). Ciononostante, egli percepiva un rischio che non si poteva trascurare: "Il pericolo di ritardare la diffusione della conoscenza scientifica tra le persone non illuminate, adulando l'oscurantismo tramite il discredito del metodo scientifico" (ivi, 18).

La terza revisione di *Grammar of Science* era necessaria agli occhi di Pearson per la persistenza di un ultimo bastione metafisico nel cuore del pensiero scientifico: l'idea di causalità. Infatti, nella prefazione del 1911, Pearson afferma: "Rileggendo il libro dopo tanti anni, fu sorprendente scoprire che l'eterodossia degli anni ottanta è diventata la dottrina ufficiale di oggi. Nessuno crede ormai che la scienza *spieghi* qualcosa; tutti la vedono come una descrizione stenografica, come un'economia del pensiero." Ma il libro fu scritto in un'epoca in cui "il *leader* dei fisici matematici inglese affermava con fiducia che niente era così sicuro come l'esistenza oggettiva dell'etere" (ivi, xi). Potrebbe parere inutile, pertanto, rielaborare un libro il cui principale insegnamento sembrava già accettato. Tuttavia, rimaneva ancora un "feticcio metafisico" nella scienza moderna, vale a dire "la categoria di causa ed effetto":

È questa categoria altro che un limite concettuale all'esperienza, e carente di qualsiasi base nella percezione oltre un'approssimazione statistica? (ivi, xii).

Il contributo di Pearson alla statistica — in quanto modo di descrivere i fenomeni osservabili — è chiaramente collegato, pertanto, con le sue tesi sulla filosofia della scienza. Le innovazioni prodotte dal suo programma di ricerca devono essere giudicate il *pendant* matematico della sua filosofia della scienza (Norton 1978). Esse saranno proposte come l'unico modo per catturare tutta la molteplice e cangiante natura dell'esperienza (Levine 2000). La correlazione sarà l'arma concettuale e operativa da impiegare a questo effetto: uno strumento antimetafisico (Levine 1996).

In *The Grammar of Science*, la statistica è la base di una scienza che consente la formulazione di leggi che descrivono le sequenze di impressioni sensoriali nel modo più economico possibile. Alla metafisica della causalità si sostituisce la scienza della correlazione. Essa sarà promossa come lo strumento matematico capace di elevare allo status scientifico discipline come la psicologia, l'antropologia e la sociologia (Norton 1978).

The Grammar of Science è la quasi perfetta articolazione dell'epistemologia positivista nella sua veste anti-metafisica; il culmine di una ricerca dell'assoluto che ha ispirato la scienza sin dal suo moderno risveglio con Bacone e Descartes (Levine 2000). Essa costituì un importante contributo al pensiero positivista e fenomenista, l'arma nella battaglia ideologica — ancora non vinta — per stabilire la scienza come l'unico arbitro della credenza razionale (MacKenzie 1981, 81).

C'è stata una reazione contro il positivismo scientifico di Pearson, ma per molti scienziati a tutt'oggi il suo pensiero e la sua descrizione della scienza sono accettabili (*ibidem*). L'immagine della scienza che contribuì a costruire rimane ancora viva: essa aiuta a capire la *received view* del secolo XX (Kjærgaard 2002).

Tuttavia, come si vedrà nella sez. 2.3, le teorie epistemologiche di Pearson hanno qualche precorrimiento delle odierne tesi anti-positiviste. Il suo positivismo ha una forte radice scettica; è una prospettiva che pone l'accento sull'inevitabile elemento umano di qualunque fatto. Levine (1996) afferma che l'aspetto più sorprendente di questo positivismo radicale è che conduce logicamente al tipo di decostruzioni dell'individuo e dell'individualità che

caratterizzano le tesi attuali degli anti-positivisti. Il moderno costruttivismo che critica l'ingenua "oggettività" e il "realismo" della scienza ha paradossalmente un precursore nel positivismo anti-metafisico di Pearson: per lui lo scienziato inventa, non scopre (Levine 2000).

Pearson rasentò relativismo e soggettivismo per stabilire, tramite le idee dell'evoluzionismo e gli strumenti della correlazione, una conoscenza più solida e meno vulnerabile (*ibidem*). La statistica sarà la soluzione al solipsismo e al soggettivismo: essa organizza la conoscenza senza rimanere prigioniera delle particolarità dei singoli fenomeni e dell'idiosincrasia dell'essere conoscente individuale.

La scienza non ha materializzato il mondo, lo ha idealizzato; per la prima volta ha reso possibile vederlo come intelligibile, non come materiale (Pearson 1888). Questa è la conquista più notevole della mente e dell'esperienza umana: le leggi dell'universo fisico seguono i processi logici della mente; l'intelletto è la chiave per dare senso all'universo fisico. Per Levine (2000) l'empirismo scettico del positivismo di Pearson diventa in definitiva un idealismo: la conferma che il mondo non solo si adatta ai contorni della coscienza, ma che è coscienza stessa.

Il tentativo pearsoniano per fare della statistica un macchinario gnoseologico perfetto che consenta l'impersonalità in un sistema intrinsecamente soggettivo fu un'impresa notevole (Levine 1996), anche se oggi si considera criticamente segnalando le sue debolezze. *The Grammar of Science* fu la migliore espressione di questo tentativo. Nelle seguenti sezioni l'analizzerò in tutti i suoi molteplici aspetti.

2. 2. Scienza e metodo nel pensiero statistico di Karl Pearson.

L'epistemologia di stampo positivista sottesa al pensiero statistico inglese, di cui *The Grammar of Science* è il miglior esponente, anticipa per molti aspetti il pensiero del *Wiener Kreis* (Norton 1978), e persino la *standard view* (Mulkay 1979; Outhwaite 1987), una corrente che ha dominato la filosofia della scienza nella parte centrale del secolo scorso e nei paesi anglofoni, rappresentata dai classici lavori di Schlick (1918), Neurath (1931), Popper (1934; 1949; 1963), Reichenbach (1938), Carnap (1939), Hempel (1952; 1966) e Nagel (1961).

Intendo analizzare la concezione scientifica di Pearson organizzando le sue idee in nuclei tematici facilmente riconoscibili come appartenenti a questa lunga tradizione: la demarcazione del campo della scienza e la definizione dei fatti scientifici; la visione cumulativa della conoscenza; il rapporto tra le scienze sociali e il modello da seguire, le scienze naturali; l'unità del metodo scientifico e il suo ruolo come garante della conoscenza impersonale; la validità universale delle leggi scientifiche.

Il problema della demarcazione: qual è il campo legittimo della scienza?

Per Pearson, il campo della scienza non ha limiti: comprende ogni tipo di fenomeno naturale, ogni aspetto della vita sociale, ogni fase del passato e del presente:

Sostenere che esistono certi campi — per esempio la *metafisica* — dai quali la scienza è esclusa, in cui i suoi metodi non hanno applicabilità, significa che le regole dell'osservazione metodica e le leggi del ragionamento logico non sono applicabili ai fatti di quei campi. Questi campi, se esistono, si trovano fuori di ogni intelligibile definizione che si possa dare al termine *conoscenza* (Pearson 1892/1957, 15).

Pertanto, se non ci sono fenomeni osservabili, la possibilità di conoscenza scompare. Pearson si sforza di chiarire questo punto, che altrimenti rischierebbe di collocare i fatti non evidenti delle scienze sociali al di là della possibilità di conoscenza, confondendoli con delle speculazioni metafisiche. In effetti, la tesi dei metafisici era che la scienza ha dei problemi legittimi, ma ci sono campi in cui il suo metodo non è applicabile, e che quindi sono problemi filosofici. I fenomeni di questa natura, in effetti, erano spesso considerati oggetto della filosofia anziché della scienza. Ma per Pearson la filosofia raggiunge i suoi giudizi attraverso un oscuro processo di cogitazione interna, pericolosamente esposto all'influenza delle deformazioni personali.

Al fine di conquistare per la scienza l'universo dei problemi umani, e di tracciare il suo confine con la metafisica, Pearson fa ricorso ad un'idea ora classica. Egli afferma che nelle scienze mentali e sociali, i fatti sono più difficili da classificare, e i pregiudizi dovuti alle opinioni individuali sono più forti. Tuttavia, la scrupolosa analisi stava trasformando l'antropologia, la sociologia e la psicologia in delle vere scienze. I fatti mentali e sociali, pertanto, non si trovano al di là del trattamento scientifico: la limitazione attuale deriva del fatto che le analisi non sono state finora sufficientemente approfondite.

Il caso della metafisica è a suo avviso molto diverso. Essa è edificata "nell'aria o su sabbie mobili giacché parte da una mancata fondazione nei fatti, o crea una superstruttura prima di trovare una base nell'accurata classificazione dei fatti" (Pearson 1892/1957, 17). La metafisica è pericolosa, perché confonde le caratteristiche di due attività intellettuali diverse: la scienza e l'arte.

Il poeta è uno stimato membro della comunità, perché si sa che è un poeta [...]. Il metafisico è un poeta, sovente uno molto grande, però malauguratamente non è noto come poeta, perché si sforza di vestire la sua poesia con il linguaggio della ragione, e per questo può diventare un membro pericoloso della comunità (Pearson 1892/1957, 17).

Da queste considerazioni non si deve concludere che la scienza, come concepita da Pearson, neghi l'esistenza di alcuni problemi tipicamente considerati filosofici o metafisici. La scienza si trova adesso davanti ai problemi della vita e della mente, sottolinea Pearson (1892/1957, 20), come nel secolo diciassettesimo si trovava di fronte ai problemi cosmici. A sostegno di questa considerazione storica, che egli intende utilizzare come arma nella battaglia contro la metafisica e la teologia per la demarcazione del dominio legittimo della scienza, cita la dichiarazione che nel 1633 prelati e cardinali fecero a proposito della cosmologia di Galilei:

*Terram non esse centrum Mundi, nec immobilem, sed moveri motu etiam diurno, est item propositio absurda, et falsa in Philosophia, et Theologice considerata ad minus erronea fide.*⁶¹

Occorsero circa duecento anni per convincere il mondo teologico che i problemi cosmici erano oggetto esclusivo della scienza. Ciononostante, la pervadente cosmologia teologica era talmente radicata nelle istituzioni religiose che ancora nel 1819 i saggi di Galilei, Copernico e Keplero erano elencati nell'indice dei libri proibiti. Come nel tardo ottocento le posizioni tradizionali della Chiesa suonavano ridicole, prima o poi — riteneva Pearson — altrettanto ridicole sarebbero diventate le tesi metafisiche circa i fenomeni ancora non sufficientemente conosciuti della vita e della mente. Nel frattempo, l'atteggiamento corretto era riconoscere l'ignoranza umana:

Fino a quando lo studio scientifico della psicologia — per via dell'osservazione e della sperimentazione — non sarà portato oltre i suoi attuali limiti, la scienza può solo riconoscere la sua ignoranza davanti a questi problemi cosiddetti "metafisici" (Pearson 1892/1957, 19-20).

⁶¹ Citato da Pearson (1892/1957, 20).

Questo riconoscimento dell'ignoranza della scienza nei confronti di molti aspetti dei problemi sociali e mentali non comporta l'impossibilità di conoscere, ma una temporanea limitazione. Nell'epistemologia di Pearson questo è un aspetto fondamentale. La metafisica non era l'unico ostacolo al progresso delle scienze umane; altrettanto pericoloso era l'atteggiamento degli scienziati che, riconoscendo l'ignoranza della scienza circa alcuni fatti, si affrettavano a concludere che essa non si doveva all'immaturità di alcune discipline né alla temporanea difficoltà di applicare i metodi scientifici. Questa era, per esempio, l'impostazione di Du Bois-Reymond⁶² (1890), che tentando di dimostrare l'inerente impossibilità di conoscere alcuni problemi aveva concluso che su essi la scienza sarebbe restata per sempre muta.⁶³

Al contrario per Pearson l'attuale stato di sviluppo delle scienze sociali si deve non alla natura dei fenomeni mentali e spirituali che tentano di conoscere, ma alla loro insufficiente analisi. Pearson fa un'enfatica difesa delle scienze sociali⁶⁴, non solo dalle intromissioni metafisiche, ma anche dagli scienziati fisici che le disprezzano: "Occorreranno secoli prima di poter svelare gli enigmi psichici e biologici, ma chiunque dichiari che mai saranno svelati è come l'uomo del cinquecento, che avrebbe ritenuto completamente impossibile comunicare attraverso l'oceano Atlantico" (Pearson 1892/1957, 24) Dovunque ci sia almeno una piccola possibilità di conoscenza umana, c'è un legittimo problema scientifico. Pertanto, tutti i fenomeni, fisici e mentali costituiscono il campo della scienza, e non si può permettere che il dogma ed il mito erigano delle barriere intorno al territorio che essa non ha ancora occupato.

Nonostante, c'è una sottile questione da mettere in evidenza. I fenomeni fisici e mentali dei quali parla Pearson sono una condizione necessaria ma non sufficiente per definire l'oggetto della scienza. La scienza si occupa di questi fenomeni solo in quanto sono in grado di provocare delle impressioni sensoriali. La scienza è in realtà un'analisi dei contenuti della mente⁶⁵, non dei fenomeni stessi. Il metodo scientifico consiste nell'estrarre comparazioni ed inferenze appropriate dalle impressioni sensoriali memorizzate e dalle concezioni basate su esse. Fino a che l'impressione immediata non arriva al livello del concepire, o almeno a quello del percepire, non diviene oggetto della scienza (Pearson 1892).

Infatti, la scienza si basa largamente sulle inferenze tratte dalle nostre impressioni sensoriali immediate e memorizzate. In altre parole, essa inferisce l'esistenza di cose che giacciono al di là della diretta verifica sensoriale, o che al presente momento non possono essere controllate tramite le impressioni sensoriali. Le ipotesi scientifiche propongono asserti che nella maggior parte stanno al di là dell'immediatamente sensibile, e — come detto sopra — la scienza si occupa in prevalenza di concezioni derivate dalle impressioni sensoriali, non delle impressioni stesse.⁶⁶

⁶² Pearson aveva seguito le lezioni di Du Bois-Reymond in Germania.

⁶³ È curioso che Pearson, gran conoscitore della lingua e della cultura tedesca, non faccia alcun riferimento ad un'altra posizione, certamente non metafisica, ma tuttavia non negatrice della possibilità di conoscenza scientifica di certi fenomeni della mente. Mi riferisco alla tradizione ermeneutica, già nota al momento in cui si scrisse *The Grammar of Science*. Si ricordi che poco tempo prima, Pearson si era recato a Heidelberg e Berlino per motivi di studio. Allora già erano note molte notevoli opere propriamente riconducibili a questa tradizione: Schleiermacher (1838), Dilthey (1883), Rickert (1899), a parte i classici che l'avevano largamente ispirata, come ad esempio Vico.

⁶⁴ Sotto l'ombrello della classica concezione del rapporto tra "scienza pura" e "scienza applicata", Pearson traccia un parallelo fra discipline naturali e quelle sociali: così come la scienza fisica pura ha una diretta influenza nella vita quotidiana, la scienza della vita e della mente dovrà avere un grande impatto sulle abitudini e vita sociale. A suo avviso, le scoperte scientifiche non solo hanno rivoluzionato la nostra concezione dell'universo, ma stano rivoluzionando la nostra vita pratica, i nostri mezzi di trasporto, il trattamento delle malattie, etc.

⁶⁵ In linea con la tradizione empirista, per Pearson tutto il contenuto della mente deriva da impressioni sensoriali.

⁶⁶ Pearson sottolinea enfaticamente quest'idea, che lo scienziato medio della sua epoca non sempre condivideva. In effetto, era frequente pensare che il campo della scienza si limitava alle impressioni sensoriali, e questo comportava per lui il pericolo di lasciar molto spazio libero al dogma e la metafisica.

Anticipando Popper, afferma che la caratteristica che rende scientifica un'asserzione è la capacità di dimostrare la sua falsità. Al contrario, l'assunto sottostante alle inferenze popolari fallaci era che l'argomento più forte in favore della verità di un'asserzione fosse l'assenza o persino l'impossibilità di dimostrare la sua falsità.⁶⁷

L'inferenza legittima — quella scientifica — non si caratterizza per l'assoluta certezza; essa si dà entro i canoni della teoria della probabilità. Secondo Pearson, gradualmente avviene una transizione nel fondamento della convinzione, dalla fede acritica verso la valutazione della probabilità. È per questo che egli ritiene conveniente utilizzare il concetto di credenza (*belief*): gli asserti della scienza non sono caratterizzati dalla certezza assicurata per autorità esterne; essi acquisiscono credito come prodotto del bilanciamento delle probabilità (Pearson 1892/1957, 59) Come affermava Edgeworth, "la probabilità può essere descritta come [...] una forma di credenza incompleta" (1884, 223).

Scienze naturali e scienze umane: la prevalenza della fisica.

La visione pearsoniana della conoscenza scientifica è chiaramente cumulativa: "Ogni grande avanzamento della scienza apre i nostri occhi a fatti che finora non eravamo riusciti ad osservare... [e questo] è uno dei più straordinari aspetti del moderno progresso [...] La meta della scienza è [...] la completa interpretazione del mondo. Tuttavia questa meta è ideale — segna la *direzione* nella quale dobbiamo muoverci, mai una fase che realmente raggiungeremo" (Pearson 1892/1957, 13-14).

La visione cumulativa consente di porre un'analogia tra scienze naturali e umane che è fondamentale nell'epistemologia positivista: essa si materializza in una strategia che tenta di sottrarre i problemi umani dal dominio della metafisica e dallo scetticismo di molti scienziati, sostenendo una linea di sviluppo inesorabile secondo il modello della fisica.⁶⁸ Ovviamente, non si tratta di un contributo originale di Pearson; era un'idea condivisa e frequentemente promossa dagli statistici: "La sociologia potrebbe derivare istruzioni dall'esperienza della sua sorella più anziana, la scienza fisica"⁶⁹, affermava Edgeworth (1893, 670). E questa derivazione di istruzioni si riteneva possibile in quanto "la prevalenza della semplice legge degli errori [si presentava anche] nei fenomeni sociali" (ivi, 672; si veda anche Quetelet 1844; Galton 1886). Inseguire il modello delle scienze naturali comportava, pertanto, un

⁶⁷ Peralto, le inferenze fallaci della metafisica, la teosofia e lo spiritualismo si nascondono dietro un enorme fiume di parole, che non permette facilmente di esibire a nudo tutta la loro assurdità (Pearson 1892).

⁶⁸ L'idea di un'unica linea di sviluppo inesorabile ha caratterizzato il pensiero occidentale in molteplici sedi. In epistemologia e storia della scienza — come sopra menzionato — la linea di sviluppo sarebbe quella tracciata dalla fisica; e tutte le discipline, per diventare scientifiche, dovranno alla fine seguire le stesse tendenze. Anche nei primi passi dell'etnologia questa concezione lasciò la sua impronta. Urry (1984) spiega che verso la metà del ottocento, l'idea che le culture "primitive" sarebbero scomparse come conseguenza del contatto con la "superiore" cultura europea, il cui modello inevitabilmente dovrebbero seguire, spinse l'urgenza di ricercarle e "museificarle". Nell'economia e nella sociologia dello sviluppo, essa si manifestò nella teoria della modernizzazione (si veda ad esempio Germani 1965), che concepisce un'inevitabile transizione dalla società tradizionale alla società moderna, tramite i processi di secolarizzazione e d'istituzionalizzazione. L'America e l'Europa del Nord sarebbero i modelli da seguire, e lo sviluppo sociale, politico ed economico potrebbe solo percorrere questa strada. Questa concezione è stata fortemente contestata nei paesi dell'America Latina, dove si coniò la teoria della dipendenza (si veda ad esempio Cardoso e Faletto 1960; Sunkel 1971), che ritiene sviluppo e sottosviluppo come due aspetti congeniali al sistema capitalistico, e non come due fasi di una linea continua di svolgimento, come voleva la teoria della modernizzazione proposta da Germani (1965). In questa concezione c'è un elemento di dominio che i gruppi in situazione di debolezza — scienziati sociali, popoli poveri, donne — hanno criticato come scientismo, eurocentrismo, etnocentrismo, androcentrismo, etc.

⁶⁹ L'impatto di questa posizione è stato enorme persino nel pensiero strettamente sociologico. Comte, Durkheim e gli esponenti della sociologia accademica americana sono gli esempi più notevoli dell'accettazione di questa premessa. Peralto, essa è uno degli elementi fondamentali di quello che Giddens (1979) chiama "consenso ortodosso" della sociologia, costruito prevalentemente intorno al funzionalismo americano, la cui epistemologia non è altro che un adattamento alle scienze sociali della già riferita *standard view* (si veda Giddens 1979 anche per i commenti critici).

orientamento verso l'applicazione degli strumenti matematici e verso la logica della quantificazione:

Fino a che i fenomeni di qualunque ramo della conoscenza non siano soggetti alla misurazione e ai numeri, esso non assumerà lo status e la dignità di scienza (Galton 1879, 149).

L'assimilazione tra scienza e misurazione, inizialmente proposta da Galilei e per molto tempo confinata all'astronomia e alla geodesia, si propagava rapidamente in tutti i domini della scienza dell'ottocento. Nel 1876, in un'esibizione di strumenti scientifici a *South Kensington*, Londra, l'ingegner William Siemens affermò che tutte le grandi scoperte della scienza si basavano su un'accurata misurazione e sul paziente lavoro di setacciare i dati numerici (Schaffer 1997). Spearman (1904b, 252) era convinto che "la psicologia dovesse dare passo alla matematica", e Pearson vedeva nello sviluppo dei metodi della statistica il mezzo per rendere possibile il conseguimento di conoscenze valide sul comportamento umano.⁷⁰

Non dovrebbe sorprendere, pertanto, che il suo approccio alla scienza dia "un'enorme importanza alla produzione di *modi matematici di descrivere i fenomeni osservabili*" (Norton 1978, 13). Paradossalmente, quest'elogio della matematica e delle sue applicazioni alle scienze umane non era stata una caratteristica permanente del pensiero epistemologico di Pearson. Nella sua gioventù, quando il suo spirito ribelle si manifestava nella denuncia di ogni pretesa di certezza e di assoluto, egli denunciava (1880) il matematico che semplifica i misteri e la complessità della natura, fino al punto di renderla innaturale — si veda sopra, sez. 1.4.

Alla fine dell'ottocento, in pratica tutti erano d'accordo sull'utilità della matematica nelle scienze naturali, ma lo scetticismo nel confronto della sua applicazione nelle scienze sociali — persino nell'economia — era abbastanza diffuso. La critica non proveniva soltanto dai cultori delle *Geisteswissenschaften*, diffidenti nel confronto di metodi che giudicavano inadeguati per indagare i fenomeni umani. Essa era frequente anche tra i non esperti: Kendall (1968) cita un brano uscito nella *Saturday Review* del 1871, in cui s'ironizzava, in un modo che fa pensare alla *lay critique* della sociologia (si veda Giddens 1979), sul modo in cui la matematica nascondeva le più semplici affermazioni del senso comune sulla politica e la società dietro un insieme misterioso di lettere e di simboli.

Fino al periodo in cui cominciò la sua collaborazione con Weldon, sotto l'ispirazione di *Natural Inheritance* di Galton, Pearson riteneva che ci fosse "un considerevole pericolo nell'applicare i metodi delle scienze esatte ai problemi delle discipline descrittive, si tratti di questioni di eredità o di economia politica". C'era il rischio di fornire una descrizione assai semplificata della realtà sociale racchiudendola in una formula matematica: "L'eleganza e la parsimonia dei processi matematici esercitano fascino sullo scienziato descrittivo [spingendolo a cercare] ipotesi sociologiche che si aggiustino al suo ragionamento matematico".⁷¹

Ma la posizione di Pearson cambiò radicalmente negli anni Novanta⁷², ed egli finì per promuovere la statistica come lo strumento capace di elevare molte discipline allo status di vere scienze (Norton 1978). Come detto nella sezione precedente, la statistica sarà per lui l'insieme di strumenti concettuali e operativi in grado di realizzare gli assunti fondamentali della sua filosofia della scienza: una macchina epistemologica perfetta che consenta l'impersonalità — soluzione per il solipsismo e il soggettivismo — e contemporaneamente di quantificare i rapporti tra le sequenze di impressioni sensoriali. Peraltro, questi sono in sostanza i due assunti tipici della *received view*, che Marradi (1997a) ha individuato nella sua analisi dei vari approcci alla ricerca scientifica.

⁷⁰ Citato da E. Pearson (1938/1948, 46).

⁷¹ Citato da Norton (1978, 16).

⁷² Più volte Pearson ha dichiarato che il cambiamento fu prodotto del contatto con l'opera di Galton, che gli dimostrò l'applicabilità della matematica ai fenomeni umani.

In effetti, questa concezione scientifica comune, sottesa al metodo sperimentale e a quello dell'associazione, rende ragionevole la loro applicazione in quanto adotta "dei precisi assunti [...] sui compiti della scienza": (1) la produzione e il controllo "in modo impersonale di asserti (e nessi fra asserti) sulla realtà", (2) che "devono essenzialmente riguardare *relazioni fra proprietà* degli oggetti" (Marradi 1997a, 677-679).

L'accento sull'impersonalità e l'oggettività caratterizza il pensiero occidentale fin dalla filosofia greca (concetto di *ἐπιστήμη*) (Marradi 1997a), e l'epistemologia, almeno a partire da Descartes e Bacone, continuò questa impostazione associando la conoscenza autorevole con la capacità del soggetto conoscente di annientare se stesso (Levine 2000).

Pearson afferma che "una volta che si classificano e si capiscono i fatti, il giudizio basato su essi deve essere indipendente della mente che li esamina". Per lui, "la formazione di un giudizio assoluto in conformità a questa classificazione [...] riassume *gli obiettivi ed il metodo della scienza moderna*" (Pearson 1892/1957, 6). Il suo ideale di scienziato è quello del libero pensatore: un individuo che assimila i risultati della più elevata conoscenza scientifica e filosofica, e che si allena al giudizio impersonale come criterio intellettuale (Norton 1978, 17-18).

Il passo dalla conoscenza di una realtà individualmente costruita sulla base delle impressioni sensoriali alla conoscenza impersonale e condivisa, basata nella generalizzazione delle sensazioni, è uno degli aspetti ancora attuali dell'epistemologia di Pearson. La conoscenza è inesorabilmente relativa, perché si basa su un mondo che cambia secondo ogni costruzione mentale individuale, però gli individui che vivono in un periodo storico comune, con un dato livello di sviluppo delle facoltà intellettuali e degli strumenti d'osservazione, riescono a costruire delle realtà analoghe, condivisibili. La chiave è il concetto di normalità: i giudizi della scienza sono impersonali in quanto condivisibili da tutti gli uomini normali in un momento dato. L'impersonalità della conoscenza diviene intersoggettività.

Il secondo assunto al quale si fece riferimento propone che una scienza si deve occupare solo di relazioni fra proprietà. Hanno pertanto "importanza capitale le distribuzioni congiunte, in quanto permettono di stabilire relazioni matematiche (funzioni) che leghino diadicamente o n-adicamente gli stati su proprietà diverse" (Marradi 1997a, 679). Nelle parole di Pearson (1892/1957, 10) la scienza implica "la conoscenza approfondita di un piccolo gruppo di fatti, il riconoscimento delle loro relazioni, e le formule — leggi — che esprimono scientificamente le loro sequenze".

Il ruolo del metodo scientifico nella realizzazione di questi assunti è cruciale: esso consente contemporaneamente di "liberarsi del pregiudizio individuale nella formazione" degli asserti, e di stabilire i rapporti tra le sequenze delle impressioni sensoriali, grazie a strumenti quali la correlazione. La sua capacità di essere universalmente applicabile unifica tutti i domini della conoscenza:

L'unità di tutta scienza consiste solo nel suo metodo, non nel suo materiale. Non sono i fatti per sé che fanno la scienza, ma il metodo con il quale essi sono trattati (Pearson 1892/1957, 12).

Rivisitando la tradizionale idea che presenta le scienze naturali come modello, afferma che il miglior modo di acquisire competenza metodologica consiste nello studio approfondito di una disciplina naturale⁷³, perché "questo consente alla mente di raggiungere un potere di gran valore per trattare qualsiasi altro tipo di fatti" (Pearson 1892/1957, 7). L'esempio da seguire è Darwin: egli insegnò che il primo passo nel processo della scoperta scientifica

⁷³ L'importanza che Pearson attribuisce a queste discipline nell'apprendimento del metodo scientifico è così grande che propone l'intervento dello Stato per mettere l'istruzione nelle scienze naturali alla portata di tutti i cittadini. Il compito del docente sarebbe stato quello di insegnare i rudimenti del metodo scientifico anziché la conoscenza dei fatti (Pearson 1892/1957, 7).

consiste nel raccogliere una gran quantità di fatti seguendo i veri principi di Bacone⁷⁴ (1620), senza l'intervento di alcuna teoria. Infatti, Il metodo s'impara da quelli che hanno dedicato la loro vita alla ricerca empirica:

Il falegname che utilizza un attrezzo è miglior giudice della sua efficienza che il fabbro che lo forgia (Pearson 1892/1957, 32).

La scienza giustifica se stessa nei suoi metodi, al punto che solo la sua applicazione garantisce un prodotto con diritto al termine 'conoscenza'. Esso "consiste nell'attenta, e di solito laboriosa, classificazione dei fatti⁷⁵, nella comparazione delle relazioni e sequenze, e infine nella scoperta, con l'assistenza dell'immaginazione disciplinata, di un breve asserto o *formula* che riassume in poche parole un ampio raggio di fatti" (Pearson 1892/1957, 77).

Descartes (1637) fu il primo a dichiarare questo ruolo centrale del metodo nell'attività intellettuale nella tradizione filosofica occidentale. La sua concezione, sorprendentemente simile a quella di Bacone, associa l'idea di metodo a "delle regole certe e facili", che evitano di "consumare inutilmente alcuno sforzo della mente" (Descartes 1628, quarta *Regula ad directionem ingenii*). Bacone, a sua volta, intendeva il metodo come capace di mettere "quasi alla pari tutti gli ingegni, perché lascia poco spazio alle capacità individuali" (1620, aforisma 122).⁷⁶ Nel pensiero della nascente teoria statistica del tardo periodo vittoriano, questa concezione del metodo trova eco in Galton (1879): fare ricerca empirica in modo metodico richiede per lui automatismo, calma e neutralità mentale.

Pearson si riconosce esplicitamente erede della tradizione metodologica di Bacone e di Galton, e non pochi hanno segnalato persino degli elementi cartesiani. Tuttavia, la sua concezione di metodo riprende quella tradizionale quanto all'universalità e alla centralità, ma respinge l'idea di metodo come insieme di regole facili, automatiche e indipendenti delle capacità individuali.

Infatti, la scienza domanda un'altra dimensione della natura umana, non soddisfatta dai processi formali e razionali. Si tratta dell'aspetto estetico ed immaginativo, che è tipico dei poeti. La scienza — afferma Pearson (1892/1957, 30) — non può disprezzare queste dimensioni estetiche ed artistiche⁷⁷, perché l'immaginazione disciplinata si trova all'origine di tutte le grandi scoperte scientifiche. L'uomo senza immaginazione può raccogliere dati, ma non può fare grandi scoperte.

L'immaginazione e la creatività hanno un ruolo centrale nell'interpretazione dei dati. In effetti, la scoperta di un singolo asserto, di una breve formula dalla quale l'intero gruppo di fatti pare fluire, non è il lavoro del mero catalogatore, ma dell'uomo dotato di immaginazione creativa. Questo singolo asserto, la breve formula, le poche parole che sostituiscono nella mente un'ampia gamma di relazioni tra fenomeni isolati, è una legge scientifica.⁷⁸ Essa permette di raggruppare la vasta complessità dei fenomeni naturali o sociali; la sua

⁷⁴ Darwin considerava il metodo di Bacone la vera strada per raggiungere la conoscenza scientifica: "*I worked on true Baconian principles*". Pearson sottolinea anche che Laplace, altra notevole influenza nella sua concezione metodologica, condivideva questo apprezzamento: "[Bacone] a donné pour la recherche de la vérité, le précepte [...] et en indiquant la vraie méthode de s'élever aux causes générales des phénomènes, ce grand philosophe a contribué aux progrès immenses." Entrambi citati da Pearson (1892/1957, 32).

⁷⁵ Già Jevons (1874, 681) aveva affermato che "tutte le operazioni del metodo scientifico realmente coinvolgono la classificazione".

⁷⁶ Questo parallelismo tra Bacone e Descartes è sottolineato da Marradi (1996). A lui rinvio anche per un complessivo trattamento dell'origine del termine 'metodo' e le sue varie accezioni.

⁷⁷ Inoltre, ricorre a questa argomentazione per attaccare l'idea secondo la quale la scienza avrebbe distrutto inevitabilmente tutto ciò che c'era di bello e di poetico nella vita, affermando che la legge scientifica è l'unica interpretazione dei fenomeni che può soddisfare permanentemente il giudizio estetico.

⁷⁸ Secondo Pearson il desiderio di creare leggi scientifiche è un istinto naturale negli uomini, un desiderio insaziabile di riassumere in delle formule, o in brevi asserti, i fatti della loro esperienza. Questo istinto porta il selvaggio a deificare il vento, il fiume e l'albero; e porta all'uomo civilizzato ad esprimere la sua esperienza emozionale nell'opera d'arte, e quella fisica e mentale nelle leggi scientifiche.

"invenzione" è la peculiare funzione dell'immaginazione creativa. Tuttavia, ogni legge proposta da uno scienziato deve essere controllata e criticata in tutte le forme possibili; la critica è l'essenza dell'uso scientifico dell'immaginazione (Pearson 1892/1957, 31).

La legge scientifica e la sua validità universale

La concezione tradizionale della scienza ha riservato un luogo speciale per la "legge scientifica". Russell (1935/1987, 9) lo dimostra definendo la scienza come "un tentativo di scoprire [...] le leggi che collegano i fatti tra loro". Non dovrebbe sorprendere che Pearson abbia trattato il tema a lungo, impostandolo come parte della sua battaglia contro la metafisica.

La sua trattazione comincia con un chiarimento concettuale: riteneva importante distinguere tra il concetto denotato dal termine 'legge' nell'ambito scientifico e in quello giuridico. Egli desiderava evitare il rischio di confusione che emerge dall'attribuire al concetto scientifico caratteristiche totalmente estranee.

Austin (1869) aveva già sottolineato l'immensa differenza tra l'uso del termine 'legge' nella scienza e nella giurisprudenza; ma egli s'occupava prevalentemente delle conseguenze dannose di applicare all'ambito giuridico il termine nella sua accezione scientifica. Pearson, invece, voleva sottolineare i rischi speculari dell'applicazione nella scienza del termine giuridico.⁷⁹ La legge scientifica e quella giuridica hanno una caratteristica comune — entrambe sono il prodotto dell'intelligenza umana — ma la legge giuridica implica un comando e un dovere, una prescrizione. La legge scientifica è invece una descrizione, un'economia del pensiero.⁸⁰ La legge giuridica ha una validità limitata dal tempo e dallo spazio: è valida solo per una comunità particolare e in un momento storico determinato. La legge scientifica, pur essendo altrettanto relativa, ha un significato universale: è valida per tutti gli uomini normali, ma resta assoluta solo mentre la facoltà percettiva rimane allo stesso livello di sviluppo.⁸¹

Tuttavia, la legge scientifica non è inerente nel mondo esterno: essa non si può ritenere né incondizionata né indipendente dall'essere umano. La dottrina di Austin — allora molto diffusa — che attribuiva alla legge scientifica un'entità naturale, era criticata da Pearson come priva di senso: proporre che la legge scientifica esista prima dalla sua formulazione umana, come una "legge naturale", è un'inattendibile idea metafisica.⁸² Essa non solo assume che la

⁷⁹ Pearson avrebbe voluto sostituire il termine 'legge' con quello di 'formula', in quanto esso fa chiaro riferimento a "una breve asserzione [...] che riassume il rapporto tra un insieme di fatti" (1892/1957, 81). Ciononostante, egli era consapevole che sostituire un termine molto diffuso nel linguaggio sarebbe stata un'impresa difficile; e si accontentò di chiarire i due significati. Peraltro, si domandava: perché dovrebbe essere lo scienziato e non il giurista a rinunciare al termine?

⁸⁰ Ancora oggi si ritiene fondamentale questa caratteristica di economia o parsimonia della conoscenza scientifica. Anche molti scienziati sociali che criticano l'impostazione positivista, e respingono la possibilità di proporre leggi per i fatti sociali, condividono l'idea di parsimonia. Glasser e Strauss (1967), per esempio, nel tentativo di sviluppare una tecnica di analisi "qualitativa" dei dati, affermano che il corollario del metodo per la scoperta della *grounded theory* sarebbe una revisione complessiva che consenta di presentare un discorso articolato e *parsimonioso* circa i processi sociali studiati. Per Pearson, il concetto di parsimonia era dovuto a Newton, per cui il canone fondamentale del metodo scientifico consisteva nel divieto di cercare cause superflue: *Causas rerum naturalium non plures admitti debere, quam quae et verae sint et earum Phaenomenis explicandis sufficient.* *Natura enim simplex est et rerum causis superfluis non luxuriat* (Newton 1687, 402).

⁸¹ Pearson accetta l'idea che le facoltà percettive umane possano cambiare; anzi, riconosce la probabilità che ciò accada proprio nella sua epoca; ma ritiene che questi cambiamenti sarebbero stati minimi, quasi insensibili. A suo avviso la facoltà percettiva del genere umano (civilizzato) aveva raggiunto una ragionevole stabilità, se confrontata con precedenti periodi di rapidi cambiamenti accaduti nell'evoluzione umana dalle forme di vita inferiori. Pearson assume un atteggiamento evoluzionista; ma come tipico esponente della sua generazione ritiene che si stesse raggiungendo l'apice dell'evoluzione in tutte le sfere della vita umana.

⁸² Pearson propone una spiegazione del processo dal quale sarebbe emersa questa concezione. L'uomo proietta le sue sensazioni fuori di sé e si dimentica il fatto che sono condizionate dalla sua facoltà percettiva;

concatenazione di fenomeni descritta dalla legge scientifica appartenga al mondo esterno, ma che esista una sua ragione causale. Ma la ragione esiste solo se associata alla coscienza, e questa solo può essere attribuita agli esseri provvisti di un sistema nervoso. Pertanto, trovare una intrinseca ai fenomeni è metafisico, indipendentemente del fatto che la ragione si ritenga inerente alla natura o propria di un supposto creatore (Pearson 1892).

Pertanto, definire una legge scientifica come universalmente valida non equivale a considerarla come "naturale", o appartenente alla cosa in sé. Essa non esiste prima che l'uomo la proponga e l'esprima. La legge scientifica si caratterizza per la sua *relatività*, ed apprezzare quest'aspetto è per Pearson assai importante, perché persino la natura è relativa, cioè dipende dalla facoltà che la percepisce. È probabile che altre persone, in altre civiltà ed epoche storiche, abbiano percepito la natura in un modo diverso: "Il mondo del bambino e quello del selvaggio sono largamente differenti da quello dell'uomo civile normale odierno". Pertanto, la legge scientifica non giace in un universo esteriore incondizionato dagli uomini. Sostenere la "validità universale della legge naturale ha senso solo se applicata a un certo tipo di facoltà percettiva, vale a dire, quella degli uomini normali" (Pearson 1892/1957, 83-85).

Una sequenza di impressioni sensoriali non costituisce per sé una legge scientifica. Prima di poter diventare tale, una sequenza deve essere confrontata con altre, classificata e generalizzata: "concezioni ed idee, puri prodotti della mente, devono formarsi prima di poter fornire la descrizione di una gamma di sequenze che, per il fatto di essere concisa e comprensiva, sia degna del nome di legge scientifica" (ivi, 86). La legge della gravitazione, per esempio, non è tanto la scoperta che Newton fece di una regola che guida il movimento dei pianeti quanto "l'invenzione" di un metodo che descrive brevemente le sequenze delle impressioni sensoriali note come movimento planetario; un *résumé* mentale che sostituisce la lunga descrizione delle sequenze delle nostre impressioni. Si tratta essenzialmente di un prodotto della mente umana che non ha alcun significato al di là degli esseri umani:

C'è più significato nell'affermare che l'uomo dà la legge alla Natura, che non nel contrario, cioè, che la Natura dà la legge all'uomo (ivi, 87).

La legge scientifica non spiega *perché* la sequenza delle nostre percezioni abbia un certo ordine, né *perché* esso si ripete; la legge non introduce alcun elemento di necessità nella sequenza; semplicemente fornisce un conciso asserto circa *come* accadono i cambiamenti. La scienza non può dimostrare la necessità inerente di una sequenza, né provare con assoluta certezza che essa si debba ripetere:

La scienza è, per il passato, una descrizione, e per il futuro una credenza; non è, e mai è stata, una spiegazione, se con questa parola si intende [...] la *necessità* di qualsiasi sequenza di percezioni (ivi, 113).

La legge scientifica accerta una conoscenza probabile, ma anche provabile. I due termini hanno più di una somiglianza linguistica: "quando si parla della sfera delle percezioni si deve ricordare che, alla fine, il termine provabile equivale a probabile".⁸³ Solo in questo senso si deve giudicare la prova nel confronto dei fenomeni naturali: essa è sempre e unicamente probabile. In quanto certezza dimostrabile, la prova si applica esclusivamente alle scienze formali come la matematica. Provare, nella ricerca empirica, significa accertare una schiacciante probabilità in favore di una certa sequenza di percezioni (ivi, 140-141).

incoscientemente poi si stacca del prodotto della sua mente, già proiettato ai fenomeni, e si domanda quale sia la ragione dietro di loro (Pearson 1892/1957, 91).

⁸³ Questa idea si deve originalmente a Jevons (1874). Le sue considerazioni circa il problema della probabilità sono state molte influenti sul pensiero epistemologico degli statistici del tardo periodo vittoriano. Altrettanto influenti sono state le opere di De Morgan (1838), Galloway (1839) e Venn (1866).

Pearson pone l'accento sull'importanza della riproducibilità delle osservazioni e degli esperimenti scientifici al fine di accertare questa "schacciante probabilità": la possibilità di ripetere uno stesso esperimento da parte di ricercatori diversi costituisce la garanzia che una data inferenza sia valida per tutti gli essere umani normali, non il frutto di una facoltà percettiva anormale. Inoltre, anticipando le moderne idee di *peer review* e dell'oggettività come prodotto intersoggettivo, difende la prevalenza del dubbio e della critica come salvaguarda del progresso scientifico:

Uno futuro fatale — per niente impossibile — sarebbe per la scienza l'istituzione di una gerarchia che trattasse come eretico qualsiasi dubbio circa le sue conclusioni, o ogni critica dei suoi risultati" (ivi, 55).

2. 3. Fenomenismo, sensismo e conoscenza: l'idea di realtà all'origine dell'epistemologia statistica.

Giddens (1977) ha individuato tra le principali caratteristiche della filosofia positivista una preoccupazione per stabilire i limiti della conoscenza scientifica, in aperto conflitto con la metafisica; e una pretesa unità dei criteri logici e metodologici delle scienze naturali e sociali. Nella sezione precedente si è illustrato il modo in cui queste caratteristiche si sono manifestate nel pensiero epistemologico di Pearson. Si è suggerito, inoltre, che non si tratta di una concezione ottocentesca superficiale e ingenua. Al contrario, essa anticipa il modo in cui questi aspetti sarebbero stati trattati nelle più recenti versioni del positivismo logico e nella cosiddetta *received view*.

Queste anticipazioni nel pensiero di Pearson possono sorprendere lo studente di scienze sociali, che di solito viene in contatto con dei manuali che presentano l'epistemologia positivista come monolitica.⁸⁴ Ciononostante, come affermano anche alcuni suoi critici, non si tratta di una filosofia "carente di ogni sottilizza e varietà" (Hughes e Sharrock 1997/1999, 60); infatti, alcuni aspetti del pensiero di Pearson non si riconoscono tanto nei moderni eredi del positivismo quanto in alcune correnti antipositiviste.

Si tratta particolarmente della sua concezione ontologica, che parte della classica idea empirista che definisce la realtà come ciò che è percepito dai sensi — un'altra caratteristica del positivismo — ma che evolve verso una forma di costruttivismo relativista e anti-realista. Secondo Levine (2000) questa impostazione anticipa il tipo di decostruzione dell'individuo e dell'individualità che segna il moderno discorso antipositivista:

Non è esagerato dire che l'universo non era lo stesso per i nostri bisnonni [...], e in tutta probabilità sarà totalmente differente per i nostri bisnipoti [perché] dipende dall'acutezza e dalla struttura dei nostri organi sensoriali, e dalla finezza dei nostri strumenti di osservazione [...] L'universo è in gran parte la costruzione di ogni mente individuale (Pearson 1892/1957,15).

Come detto precedentemente, questa costruzione diviene possibile solo a partire delle impressioni sensoriali dei fatti. Per il pensiero epistemologico dei teorici della statistica, il campo legittimo della scienza è costituito dai *fatti* naturali e sociali, o — nel caso di Pearson

⁸⁴ Tanto nei manuali di tendenza positivista quanto in quelli dichiaratamente critici, spesso si illustra questa corrente epistemologica in modo superficiale e banale (Hughes e Sharrock 1997). Alcuni la difendono acriticamente come "la" filosofia della scienza, senza problematizzare i suoi assunti ontologici, epistemologici e metodologici. I critici la ritraggono come una filosofia dogmatica e ingenuamente realista (Levine 1996), senza sfumature. Paradossalmente, molti critici non si accorgono di quanto siano positivisti loro stessi (Phillips 1987, 44): è che malgrado la sua perdita di preminenza al meno al livello discorsivo, molti aspetti pratici della ricerca scientifica standard sono ancora determinati dal positivismo, anche nelle scienze sociali (si vedano Halfpenny 1982; Marradi 1984; Pawson 1989; Hughes e Sharrock 1997).

— dalle costruzioni mentali da essi derivate. Edgeworth per esempio, affermava che "i dati oggettivi sono un requisito" per "l'applicazione del calcolo della probabilità" nella ricerca scientifica (1884, 232). Pearson dichiarava che "se non ci sono dati, o sequenze tra loro da osservare, la possibilità di *ogni* conoscenza scompare" (1892/1957, 15). Il problema è come definire questi fatti e dati della scienza: qual è il criterio della loro realtà?

Questo è un tema trattato a lungo da Pearson. Una condizione *sine qua non* dell'esistenza di un oggetto reale — come appena detto — è una qualche impressione sensoriale.⁸⁵ Tuttavia, le impressioni sensoriali che determinano la realtà dell'oggetto esteriore sono forse pochissime; l'oggetto è in grande misura costruito da inferenze ed associazioni, ma qualche impressione sensoriale deve esistere per considerare l'oggetto reale, e non un mero prodotto dell'immaginazione. Nella costruzione dell'oggetto contano molto i vari tipi di esperienza passata, costituita a sua volta in base a precedenti impressioni che rimangono nella memoria. Un'impressione sensoriale, se abbastanza forte, lascia nel cervello una traccia più o meno permanente, che si manifesta sotto la forma di associazione allorquando capita un'impressione sensoriale simile. Gli effetti memorizzati di impressioni sensoriali passate formano in gran parte quello che siamo abituati a denominare "oggetto esteriore", che è pertanto prevalentemente costruito da noi (ivi, 40-41).

La concezione costruttivista di Pearson, tuttavia, non deve confondersi con la moderna idea che considera la realtà e la conoscenza come una costruzione sociale. La sua versione più diffusa, ispirata nell'analisi fenomenologico di Schutz (1932; 1962) è quella presentata da Berger e Luckman in *The Social Construction of Reality* (1966): ciò che l'uomo percepisce e conosce come reale varia da società a società, ed è prodotto, trasmesso e conservato attraverso processi sociali. La realtà della vita quotidiana si presenta come ordinata in un modo condiviso con gli altri (Berger e Luckman 1966/1969, 44), che si interiorizza soggettivamente nei processi di socializzazione primaria e secondaria.⁸⁶

La costruzione di cui parla Pearson è invece *individuale*. Solo perché gli uomini normali hanno organi sensoriali simili ed analoghe facoltà percettive si può inferire la costruzione — malgrado le individualità — di un mondo esteriore comune, di una realtà esterna *in sostanza* uguale.⁸⁷ In questo consiste la similitudine dell'universo per tutti gli esseri umani. L'empirismo soggettivista che risulta dalla costruzione individuale della realtà — e che finisce nel solipsismo — trova la sua via di uscita nell'idea di normalità. L'universalità della realtà — e della conoscenza scientifica — dipende della somiglianza delle facoltà percettive e razionali degli esseri umani normali (Levine 1996).

La definizione pearsoniana segue la linea di Morgan (1891): il termine *costrutto* fa riferimento a una combinazione di impressioni sensoriali immediate e passate. Il criterio di realtà di qualsiasi oggetto dipende, in definitiva, della sua possibilità di generare impressioni sensoriali; non dal fatto di essere percepito, ma dalla capacità di essere percepibile. C'è pertanto una sottile differenza tra 'reale', 'irreale' e 'ideale'. La scienza si interessa prevalentemente di oggetti reali, ma nelle sue ipotesi introduce spesso degli oggetti ideali, dei quali non abbiamo ancora un'esperienza sensoriale che comunque si potrebbe alla fine raggiungere. Un oggetto ideale è pertanto uno del quale non abbiamo ricevuto impressioni sensoriali; tuttavia, esse sono possibili, concepibili. Invece, un oggetto irreale è uno del quale non abbiamo esperienza diretta e che non potrebbe mai divenire un'impressione immediata. Ciononostante, la scienza si occupa di alcuni oggetti dei quali non abbiamo né potremmo avere un'impressione sensoriale diretta: le idee astratte delle scienze formali⁸⁸, come quelle

⁸⁵ Il sensismo di Pearson è fortemente ispirato dall'analisi anti-metafisica delle sensazioni proposta da Mach (1886), che riduce il mondo esteriore alla sua "superficie conoscibile", cioè alle sensazioni.

⁸⁶ La socializzazione secondaria è particolarmente importante nella costruzione della realtà e della conoscenza appartenente a sottomondi istituzionalizzati, per esempio quello della scienza.

⁸⁷ Per Pearson la costruzione del mondo esteriore di due uomini qualsiasi non può essere mai *esattamente* uguale, perché gli organi sensoriali e la facoltà percettiva variano dall'uno all'altro.

⁸⁸ Nella classificazione di Pearson sono "scienze astratte", che si occupano delle forme e dei concetti ideali. Si tratta di discipline strumentali, i cui concetti ci aiutano ad osservare e distinguere i fenomeni. Pearson considera

geometriche di curva e superficie, o i fatti storici. Apparentemente questo comporta una contraddizione, ma nella concezione ontologica di Pearson, questi non sono oggetti né ideali né irreali, semplicemente perché non sono oggetti. Si tratta dei fenomeni che Clifford (1879) chiamava *ejects*, appunto per distinguerli dagli oggetti (*objets*), e sono materia di studio legittima delle scienze.

Per chiarire il punto su oggetti ideali e irreali Pearson propone un esempio: il pianeta Nettuno era un oggetto ideale — non irreali — fino a quando — grazie allo sviluppo di strumenti di osservazione quali il telescopio — si raggiunse una sua impressione sensoriale immediata; a questo punto divenne un oggetto reale. La cosa in sé della quale scrive Kant⁸⁹, invece, è un oggetto irreali, perché non abbiamo né potremmo mai averne un'esperienza sensoriale diretta (Pearson 1892/1957, 41).

L'uomo è abituato a parlare del "mondo esteriore", della "realtà" che giace al di là di sé e che ritiene composta di oggetti individuali con un'esistenza indipendente. Per contrapposizione, definisce un "mondo interiore" costituito dal deposito delle impressioni sensoriali passate, dal pensiero e dalla memoria. Tuttavia, "la distinzione tra mondo interiore ed esteriore è arbitraria, chiaramente una di mera convenienza pratica nella vita quotidiana" (ivi, 65).

L'essere umano tende a proiettare verso l'esterno degli insiemi di impressioni sensoriali, definendole con l'etichetta di 'fenomeno'. Nella vita quotidiana, per motivi pratici, questi fenomeni si ritengono reali. Tuttavia, gli insiemi di impressioni che si proiettano contengono inevitabilmente aspetti estratti dalle impressioni memorizzate, che l'esperienza induce ad associare con ogni particolare impressione immediata. Pertanto, quello che chiamiamo "mondo reale" è una congiunzione di sensazioni immediate e di impressioni memorizzate.

Di fronte a un'argomentazione del genere, l'uomo della strada potrebbe rispondere: lì c'è un tavolo, io lo vedo, ha quattro gambe ed è fatto di lucida quercia nera; è una realtà. Ma se si fermasse a riflettere, sottolinea Pearson, "subito dopo si convincerebbe che il tavolo *reale* consiste nell'associazione permanente di un certo gruppo di impressioni sensoriali". Che cosa accadrebbe se qualcuno rompesse il tavolo e lo bruciasse? Esso si trasformerebbe in gas e cenere; la sua forma e misura, temperatura e colore, robustezza e solidità... tutto sarebbe svanito. È chiaro pertanto, prosegue Pearson, che il tavolo consisteva nel permanente apparire insieme di certe impressioni sensoriali, ma la sua forma e consistenza non esistevano fuori di chi le percepiva (ivi, 69-70).

Tradizionalmente i metafisici avevano cercato di trovare una spiegazione a quello che giace sotto le impressioni e che le dota di permanenza: la scuola materialista — allora influente — faceva ricorso all'idea di una materia sottostante; Berkeley a Dio; Kant e Schopenhauer alla volontà; e il naturalista Clifford alla *mind-stuff*. Pearson riteneva ugualmente inutili tutte queste considerazioni, e si riconosceva agnostico, nel senso originario voluto da Huxley⁹⁰, circa la possibilità di conoscere una pretesa cosa in sé.

A tutti gli effetti scientifici non occorre presumere una cosa in sé dietro di ogni insieme di impressioni sensoriali per dar conto della permanenza degli oggetti o della loro esistenza (ivi, 72). Proporre che l'uniformità e la permanenza della realtà dipenda da una cosa in sé è una fantasia, e pertanto non è oggetto di inferenza logica né di conoscenza umana:

la statistica come una di queste scienze astratte, definita come una disciplina interessata alle relazioni generali di discriminazione quantitativa.

⁸⁹ Il pensiero epistemologico dei teorici della statistica è indebitato alle idee di Kant. Pearson aveva già letto la *Kritik der reinen Vernunft* (1781) come studente universitario in Inghilterra, e approfittò della permanenza in Germania per approfondirne la lettura. A parte il rifiuto del concetto di cosa sé, egli fu molto attratto dalle idee della *Kritik*. Egli trovò in Kant una soluzione adeguata all'antagonismo tra empirismo e razionalismo. Si ricordi che era ammiratore di Bacone e di Descartes, e le loro tracce si possono trovare nel suo pensiero.

⁹⁰ Huxley introdusse il termine 'agnostico' per definire, secondo l'interpretazione di Pearson, quelli che limitano la possibilità di conoscenza a certi campi.

La ragione che incontriamo nei fenomeni naturali è sicuramente dovuta all'unica ragione della quale abbiamo alcuna esperienza, cioè, la ragione umana (ivi, 91).

La costruzione della realtà richiede pertanto la presenza di un'attività cosciente.⁹¹ È evidente, secondo Pearson, che senza impressioni sensoriali non ci sarebbe niente da registrare; ma è altrettanto evidente che senza la facoltà di ricevere impressioni permanenti e di memorizzarle non ci sarebbe la possibilità di pensare. Alla fine, senza il pensiero, che interviene nell'intervallo tra un'impressione sensoriale e la reazione conseguente, non ci sarebbe coscienza. Quando questo sforzo segue immediatamente un'impressione sensoriale, esso è involontario, un riflesso. Al contrario, quando è condizionato dalle impressioni memorizzate, si considera un'azione volontaria (ivi, 45).

L'impressione sensoriale immediata può esser considerata la scintilla che accende il pensiero, che chiama in causa le tracce di impressioni sensoriali passate. Ma il cervello umano è molto complesso, e le impressioni memorizzate si legano tra loro in diversi modi, rendendo spesso poco evidente il rapporto tra un'esperienza particolare e la risultante successione di pensiero. Analogamente, non sempre siamo in grado di rintracciare le impressioni immediate che diedero luogo ad una successione di pensiero specifica, perché il materiale del pensiero non si riduce a una combinazione di "oggetti esteriori" associati con le immediate impressioni sensoriali. Il pensiero, una volta stimolato, classifica, analizza o semplifica le caratteristiche delle impressioni e forma nozioni generali dei suoi modi e proprietà: "si passa dal *perceive* al *conceive*" (ivi, 47).

Questo processo di pensiero si può costatare, in senso stretto, soltanto in se stessi. Tuttavia, possiamo estenderlo ad altri esseri umani, perché essi hanno un organo fisico analogo — il cervello — che con ragionevole probabilità riceve impressioni uguali e forma "costrutti" simili:

Riconosciamo la coscienza nel nostro essere individuale, *presumiamo* che essa esista negli altri [...] La coscienza di un altro uomo [...] non può mai essere percepita come un'impressione sensoriale; la sua esistenza può solo essere *inferita* (ivi, 48-49).

La coscienza si associa al processo che *potrebbe* intervenire, nel cervello, tra la ricezione di un'impressione da parte di un nervo sensoriale e l'invio di uno stimolo all'azione tramite un nervo motorio. Pertanto, la coscienza è legata a un macchinario fisiologico, e si manifesta nel lasso di tempo tra un'impressione sensoriale e uno sforzo. Quando questo macchinario — o quest'intervallo — non si può osservare, non si ha il diritto d'inferire l'esistenza di coscienza (ivi, 57).

Ma quanto si può avvicinare questo macchinario al presunto mondo esteriore? Come opera l'accesso epistemico alla realtà? Esso funziona — metaforicamente — con delle limitazioni analoghe a quelle di un impiegato di una centrale telefonica. L'ego cosciente si trova nei terminali cerebrali dei nervi sensoriali, e non può andare oltre questi terminali per catturare il mondo esteriore. Il mondo esteriore ci giunge attraverso impressioni sensoriali, che sono dopo analizzate, classificate, memorizzate. Tuttavia, della natura delle cose in sé, di quello che potrebbe esistere oltre il nostro "sistema di fili telefonici" non sappiamo niente. La cosa in sé, la "realtà" dei metafisici, è sconosciuta e non conoscibile. La realtà del mondo esteriore consiste per la scienza nella combinazione di forma, colore e tatto: "siamo ingabbiati e confinati in un modo di impressioni sensoriali" (ivi, 63). Il mondo è condizionato dai nostri organi sensoriali; le loro peculiarità determinano la natura del mondo che costruiamo. Pearson

⁹¹ La coscienza si può inferire anche per le forme di vita inferiori, ma essa presenta una minore intensità. Nelle forme di vita animale non si può stabilire il punto in cui la coscienza svanisce. Pearson ironizzava: "quando non riconosciamo i vari gradi della coscienza ed intelligenza animale, assumiamo che un'anima sia impiantata in ogni essere umano al momento della nascita" (1892/1957, 128). Anche Galton era contrario a qualsiasi concetto soprannaturale di anima (MacKenzie 1981).

aggiunge il *sentio* alla massima cartesiana del *cogito*: "sperimento, pertanto una *routine* di percezioni esiste" (Levine 1996):

La ripetizione di una *routine* di impressioni sensoriali è una condizione essenziale del pensiero; l'effettivo ordine della sequenza non ha importanza, ma qualunque cosa sia, la sequenza deve ripetersi per rendere possibile la conoscenza (Pearson 1892/1957, 136).

Il potere di pensare, di associare insiemi di cose e di sequenze di impressioni sensoriali — immediate e memorizzate — sarebbe svanito se questi insiemi e sequenze non avessero qualche elemento permanente per classificarli e compararli. Tuttavia, non si può affermare l'esistenza di un'*assoluta* uniformità nella ripetizione dei fenomeni né si possono assimilare *assolutamente* tutti i comportamenti simili degli esseri razionali. La *routine* di percezioni è un'idea relativa che indica un certo grado di uniformità nella ripetizione, il cui limite — l'assoluta uniformità — è una pura nozione concettuale. Secondo la nostra esperienza, niente nell'universo si è ripetuto né si ripeterà esattamente (ivi, 153-154).

Per il pensiero, per l'azione nella vita quotidiana e per la ricerca scientifica è sufficiente l'indizio di una certa media o uniformità statistica, e non l'assoluta identità.⁹² Essa si può assumere solo in alcune aree specifiche della ricerca, quando gli oggetti o i comportamenti studiati sono tanto uniformi, che le insignificanti differenze nelle *routines* delle loro percezioni non ostacola la legittima inferenza: "Non va a discredito della grande struttura della moderna chimica fisica affermare che la presunta assoluta identità della molecola sia invece un'uniformità statistica" — con un errore probabile infimo — che consente allo scienziato, per motivi pratici, di assumere l'identità di tutti gli oggetti classificati insieme agli effetti delle asserzioni che controlla sperimentalmente (ivi, 156). Pearson era molto attento a questi assunti circa la natura degli oggetti e le loro proprietà. Non a caso, consapevole dell'inadeguatezza di assumere l'assoluta identità degli oggetti studiati, particolarmente nel campo delle scienze umane, indirizzò tutti i suoi sforzi allo sviluppo di tecniche che permettessero di realizzare gli scopi della scienza tradizionale⁹³ anche in questi casi.⁹⁴ Non a caso affermava che, una volta sviluppate le principali tecniche utili a questo proposito — correlazione e associazione — il successivo compito della statistica matematica sarebbe stato quello di raffinare la teoria e gli strumenti del campionamento casuale, indispensabili per la ricerca scientifica quando non si può assumere l'assoluta uniformità.

La visione pearsoniana dell'universo è analoga a quella di Galton: un intricato sistema di variabili correlate, approssimando ma mai raggiungendo la perfetta correlazione. Al mondo deterministico della causalità si sostituisce un mondo aleatorio — tuttavia regolare, uniforme — di fenomeni contingenti. L'universo della statistica è pertanto più complesso e sottile che quello deterministico della causalità; è un mondo in cui il dualismo e le dicotomie si considerano riduzionismi (Levine 2000):

L'universo è fatto di innumerevoli entità, ognuna probabilmente individuale, ognuna probabilmente non permanente; tutto ciò che l'uomo può raggiungere è classificare queste entità in delle classi di individui simili per via di misurazione o di osservazione delle loro caratteristiche. In queste classi la variazione può essere evidenziata, e il problema fondamentale della scienza è scoprire come la variazione in una classe sia correlata o contingente con la variazione nell'altra (Pearson 1892/1957, 165).

⁹² Per Pearson, inoltre, questa "convinzione nell'ordine uniforme dei fenomeni è essenziale per il benessere umano" (ivi, 138).

⁹³ Controllo impersonale di asseriti circa le relazioni tra le proprietà degli oggetti (di preferenza espresse in termini matematici). Si veda la sezione precedente.

⁹⁴ Si ricordi che uno dei suoi principali interessi era quello di trasformare la ricerca psicologica, antropologica, sociologica, etc. in "vera" ricerca "scientifica".

2. 4. Correlazione anziché causalità: la nuova arma scientifica contro l'ultimo bastione della metafisica.

Goldthorpe (2001, 1) afferma che "i fondatori della statistica moderna possono essere considerati rappresentati di un'era in cui il concetto di causalità era visto con scetticismo."⁹⁵ Almeno per Karl Pearson era un mero feticcio ereditato dal pensiero metafisico prescientifico, che doveva essere abbandonato e sostituito con quello di correlazione, più generale e preciso.

Infatti, nell'ultima revisione di *The Grammar of Science*, compiuta nel 1911, Pearson dichiarava che nella scienza rimaneva ancora un feticcio metafisico carente di ogni base percettiva: la categoria di causa ed effetto. La diffidenza nei confronti della causalità aveva già caratterizzato parte del pensiero statistico inglese del cuore del periodo vittoriano, che segnò la transizione della statistica dai fatti al metodo (dallo studio quantitativo della società allo sviluppo di strumenti di analisi quantitativa). In una lettera a Florence Nightingale — già citata nel primo capitolo — William Farr comunicava le sue riserve a mescolare causalità con statistica.⁹⁶ A suo avviso "lo statistico non ha niente a che vedere con la causalità", e questo comportava il rifiuto di ogni tentativo di provvedere una spiegazione dei fatti. La posizione di Pearson è molto più complessa ed elaborata; tuttavia, la presenza di un atteggiamento critico degli statistici tradizionali inglesi verso la causalità permette di stabilire una qualche linea di continuità fra loro e Pearson.

Abbagnano (1961/1963, 152) ha individuato due tendenze nell'evoluzione storica del concetto di causa. La più antica la concepiva come una relazione *razionale*, in cui la causa diveniva la ragione di un effetto, una *forza* che lo provoca e che è pertanto deducibile. L'altra, invece, definiva la causa come una relazione *empirica* costante e uniforme, non deducibile dall'effetto.

Di seguito propongo una breve e piuttosto arbitraria ricostruzione della storia del concetto, orientata a far capire l'impostazione positivista in generale, e quella pearsoniana in particolare.

La nozione metafisica emerse storicamente dal bisogno umano di trovare una spiegazione a ciò che accade nell'universo. Non dovrebbe sorprendere, di conseguenza, che essa sia stata fondamentale sin dall'inizio della riflessione filosofica (Ferrater Mora 1994). Nel pensiero greco, la 'causa' (αἰτία) implicava un'imputazione (idea legata all'origine giuridica del termine) e la produzione di qualcosa seguendo una certa norma o principio. Platone la definì appunto come il principio per il quale una cosa risulta ciò che è, e stabilì la classica distinzione tra causa prima (αἰτία) intelligibile, e causa seconda (αἰτία δευτέρα) sensibile, efficace (si veda Ferrater Mora 1994). Tuttavia, la prima trattazione sistematica del tema si trova in Aristotele, pioniere nell'affermare che la scienza e la conoscenza devono individuare le cause di ogni fenomeno, giacché tutto avviene a partire da qualcosa.

La dottrina aristotelica fu approfondita nel pensiero scolastico medioevale, particolarmente da San Tommaso. Egli attribuiva alla causa un elemento di necessità, tipico dell'impostazione metafisica. In generale, i filosofi razionalisti del sei-settecento — Descartes, Spinoza, Leibniz — continuarono ad assimilare il concetto di causa con quello di ragione, e coerentemente la definirono come un *principium*, cambiando solo l'impostazione ontologica — prevalente nell'antichità e nel medioevo — in una prospettiva gnoseologica (Ferrater Mora 1994).

La tradizione empirista britannica, antagonista del razionalismo continentale, inaugurò la seconda tendenza individuata da Abbagnano. Hume (1748) criticò pesantemente l'idea di

⁹⁵ In realtà la nozione di causa non era universalmente criticata in quell'epoca. Più giusto sarebbe rilevare che essa era considerata metafisica da molti pensatori positivisti (Hacking 1990), e che le loro idee guadagnavano rapida accettazione tra gli scienziati dell'epoca, compresi ovviamente i teorici della statistica.

⁹⁶ Citato da Porter (1986, 36).

rapporto necessario tra causa ed effetto: questo rapporto è solo un'inferenza fondata su una regolarità osservata. Rivisitando un'idea di ispirazione platonica, affermò che le relazioni necessarie si giustificano solo a livello delle idee, non in quello dei fatti. Nella sua critica al razionalismo e all'empirismo, Kant (1781) restrinse la causalità al mondo fenomenico, definendola come una categoria *a priori*, un "principio puro dell'intendimento". Tuttavia, concepì la natura come ordinata da relazioni causali, e pertanto, la sua "soluzione" al dibattito razionalista-empirista sul concetto di causalità è "solo verbale, un dogmatismo travestito" (Abbagnano 1961/1963, 157). Nel pensiero postkantiano, Schopenhauer fu uno dei più decisi difensori dell'idea metafisica di causalità, unica categoria originaria, legata per lui alla volontà.

Nel tardo periodo vittoriano, quando Pearson cominciò a sviluppare gli strumenti della correlazione, la nozione metafisica di causalità era ancora molto diffusa nella scienza. Le critiche di Pearson, non a caso, si indirizzarono prevalentemente alla concezione di Aristotele e dei suoi eredi scolastici, da un lato, e a quella di Schopenhauer dall'altro; entrambe ancora influenti nella scienza ottocentesca.

Bastava leggere un libro di fisica — considerata la disciplina più evoluta — per confermare quanto fosse diffusa la definizione della meccanica come la scienza della *forza*; causa inerente alla materia che produce o tende a produrre cambiamenti nel movimento. Pearson critica irosamente questo chiaro esempio di inaccettabile "intromissione della metafisica aristotelica" nella scienza. Seguendo Kirchhoff (1876), definisce la meccanica come "la scienza del movimento, il cui oggetto è la sua completa *descrizione* nel modo più *semplice* possibile" (Pearson 1892/1957, 115). Una "spiegazione" meccanica dei fenomeni implica solo la descrizione di una certa routine di percezioni nel conciso linguaggio della meccanica. Se si considera la forza come la causa del cambiamento, essa diviene uno dei tipici fantasmi con cui la metafisica riempie l'al di là delle impressioni sensoriali. La forza, pertanto, non è utile nella ricerca di una concezione scientifica di *causa* (ivi, 116-117).

La concezione "animista" e "volontaristica" di Schopenhauer aveva influenzato, per esempio, il pensiero scientifico articolato nell'opera astronomica di Herschel (1849). Per Pearson era ragionevole accettare che nelle prime fasi del loro sviluppo, gli esseri umani abbiano attribuito alla volontà la natura di agente causale. Solo nel lungo corso dell'evoluzione l'umanità cominciò a rendersi conto del fatto che la volontà è attribuibile solo ad esseri coscienti. A poco a poco la spiegazione spiritualista del movimento dovette lasciare il passo alla descrizione scientifica:

La necessità della legge naturale non ha la cogenza logica del teorema geometrico, né la cogenza categorica della legge giuridica voluto dal suo umano legislatore; è semplicemente la nostra esperienza di una *routine*, le cui fasi non hanno ordine logico né voluto (Pearson 1892/1957, 120).

La scienza non ha nulla da dire sulle prime cause: esse non possono essere inferite dalla ricerca scientifica, come supponeva Jevons.⁹⁷

La scienza moderna descrive come la volontà sia influenzata da desideri e passioni, e come questi derivino, a sua volta, dall'educazione, dall'esperienza, dall'eredità, e dalla malattia; tutte questioni associate al clima, alla classe, alla razza, e ad altri fattori dell'evoluzione. Pertanto, con il progresso della conoscenza positiva, l'uomo è riuscito a considerare gli atti individuali di volontà come cause secondarie in una lunga sequenza; come mere fasi in una *routine* descrivibile. Un atto di volontà indica il luogo dove la *routine* di sensazioni cambia dall'aspetto fisico a quello psichico:

L'idea di volontà come prima causa è meramente il limite che emerge dall'incapacità dei nostri poteri di seguire oltre il lato fisico di una *routine*, o di scoprire il suo lato psichico; è meramente un altro modo di dire: a questo punto comincia la nostra ignoranza [...] Analizzata

⁹⁷ Pearson è fortemente critico della concezione di causa di Jevons (1874).

nei suoi dettagli, la concezione che assegna alla volontà l'origine del movimento è priva di coerenza (ivi, 127).

A partire da questa critica comincia a emergere la nozione scientifica di causalità, più vicina alla tradizione intellettuale dell'empirismo inglese. Il termine 'causa' è utile solo se usato per individuare una particolare fase in una *routine* di percezioni. In questo modo, esso si applica al campo delle impressioni sensoriali, l'unica sfera su cui si può ragionare e raggiungere conoscenza. La causa è pertanto una fase della *routine* dell'esperienza, e non una necessità inerente. Seguendo Hume, Pearson afferma che non c'è niente in una causa scientifica che obblighi per inerente necessità a predire un dato effetto. Si tratta di una concezione certamente probabilistica della causalità: nessuna *routine* di esperienze implica necessità logica; ogni sequenza è caotica fino a che non diventa un'esperienza ripetuta:

Ogni volta che una sequenza di percezioni D, E, F, G è invariabilmente preceduta dalla percezione C, o le percezioni C, D, E, F, G si succedono in quest'ordine, cioè, formano una *routine* di esperienza, C si ritiene la *causa* di D, E, F, G, che si descrivono come gli effetti (ivi, 130).

Ciononostante, nessun fenomeno o fase di una sequenza ha un'unica causa: tutte le fasi precedenti sono cause successive, e come la scienza non ha ragione per inferire una prima causa, la successione di cause può essere rintracciata al limite della conoscenza esistente, ed oltre *ad infinitum* nel campo della conoscenza concepibile. La causa di un singolo evento si perde nell'incontrollabile storia dell'universo.⁹⁸ Dello stesso parere era Galton, per il quale i fenomeni erano determinati da cause molteplici e indipendenti.

Il termine 'causa', in questo senso, è piuttosto elastico: come proponeva John Stuart Mill (1843), esso comporta l'uniforme congiunzione nello spazio, l'antecedente uniforme nel tempo. La sua necessità logica, quando usata in una teoria empirica come quella del movimento planetario, o in una teoria formale come quella geometrica del circolo, si deve esclusivamente alle definizioni e agli assiomi della mente. La necessità appartiene al mondo delle concezioni; solo attraverso un processo inconscio — e ingiustificabile — viene trasferita al mondo delle percezioni (ivi, 134).

L'altissima probabilità dedotta dalle passate *routine* di esperienza consente di considerare l'universo come "invariabilmente ordinato"; tuttavia, la certezza ha senso nel campo delle idee. Pur non essendo né necessaria né certa, questa *routine* delle percezioni è il fondamento dell'idea di causalità, la cui assenza renderebbe in pratica impossibile l'esistenza degli esseri razionali capaci di agire. Fedele al *cogito* cartesiano, anche per Pearson il pensiero connota esistenza, ma l'azione della vita quotidiana richiede anche la percezione (*sentio*) di una *routine* di impressioni.

La nozione del rapporto causa-effetto non è altro che la cristallizzazione di questa necessità pratica. L'uomo l'ha trasformato nel fattore dominante dei fenomeni, al punto di convincersi non solo della sua assoluta verità, ma anche della sua corrispondenza con qualche realtà intrinseca all'esistenza stessa (ivi, 152):

La legge della causalità è una finzione concettuale estratta dai fenomeni, non è la loro essenza (ivi, 157).

Per Pearson, l'effettivo problema davanti all'umanità è più largo e complesso di quello della causalità: "se le cause hanno questo o quel grado di possibilità, quanto probabili saranno gli 'effetti'?" Dato che nel senso più ampio qualsiasi fenomeno *x* è una delle cause che precede

⁹⁸ Quest'affermazione mette in evidenza, secondo Pearson, quanto siano "connesse tra loro le varie discipline della conoscenza positiva; e quanto sia difficile per il grande edificio della scienza progredire con rapidità e sicurezza a meno che tutte le sue parti si tengano al passo una con l'altra" (ivi, 131). Un po' come Cristoforo Colombo, che affermava che nessuna flotta può navigare più rapida della sua nave più lenta.

o accompagna un altro fenomeno y , l'impostazione scientifica corretta è domandarsi in quale misura varia y quando varia x . Se la variazione di quest'ultimo non produce alcun effetto in y , i fenomeni sono assolutamente indipendenti; se si scopre invece che quella variazione di y è completamente legata a quella di x , c'è assoluta dipendenza. "Tra questi due limiti [...] tutti i gradi d'associazione possono capitare. Quando si varia una causa, il fenomeno cambia, ma c'è variazione nel suo cambiamento. Quanto minore sia la variazione in questo cambiamento, più strettamente la causa definisce il fenomeno, più strettamente si può accertare l'associazione o la correlazione. È questa concezione della correlazione tra due eventi, che abbraccia tutte le relazioni, dall'assoluta indipendenza fino alla completa dipendenza, la categoria più ampia con la quale si deve sostituire la vecchia idea di causalità" (*ibidem*).

I fenomeni sono pertanto contingenti. Per determinare la contingenza tra due fenomeni la scienza conta su uno strumento straordinario: la tabella di contingenza. Essa "è il sillogismo numerico che sostituisce per ogni proposito lo sterile sillogismo della vecchia logica aristotelica" (ivi, 160). Completa dipendenza e indipendenza sono i limiti estremi di una relazione. La preminenza delle nozioni fisiche ce le fa abitualmente ritenere come le due uniche categorie: indipendenza e causalità. In realtà sono limiti estremi della tabella di contingenza nella quale possiamo comprendere l'intera gamma di possibili associazioni di ogni coppia di proprietà (ivi, 163):

La vecchia visione della causa-effetto tentò di ridurre l'universo ai due limiti concettuali dell'esperienza — ma non poteva altro che fallire: nella nostra esperienza le cose non sono o indipendenti o causate (ivi, 166).

Il problema della contingenza è per Pearson alla base della scienza moderna. Da questo punto di vista, non trova distinzioni essenziali, ma solo di grado, tra i dati, il metodo, o le leggi risultanti dalla ricerca fisica, chimica, biologica o sociologica. Come affermò Blumer (1969), giustificando i percorsi seguiti nella ricerca interazionista simbolica, i metodi di ricerca e gli obiettivi della scienza devono essere coerenti con il modo in cui si concepisce la realtà. In quanto assunto ontologico, l'idea di un mondo contingente diventò la giustificazione di un modo particolare di intendere gli scopi della scienza:

Che l'universo sia una somma di fenomeni, alcuni dei quali sono più ed altri meno legati tra loro, è la concezione [...] che dobbiamo al presente estrarre dalla nostra esperienza [...] L'obiettivo della scienza [...] è cercare i fenomeni più fortemente correlati... Nessun fenomeno è causale; tutti i fenomeni sono contingenti, e il problema che dobbiamo affrontare è quello di misurare il grado di dipendenza, che si trova tra lo zero dell'indipendenza e l'uno della causalità. Questa è l'ampia prospettiva dell'universo che ci viene dettata dalla nostra esperienza (Pearson 1892/1957, 173-174).

Nell'ultima revisione del suo testo epistemologico, quando gli aspetti tecnici della correlazione e dell'associazione erano abbastanza sviluppati, Pearson sottolineava che i concetti di contingenza e di correlazione erano analoghi, e che entrambi i relativi coefficienti misuravano il grado in cui due fenomeni erano associati. Essi consentivano di determinare, per proprietà qualitative o quantitative, il legame tra qualsiasi coppia di proprietà nell'universo, e pertanto erano la base del nuovo approccio allo studio della natura (ivi, 177). La correlazione diviene "lo" strumento antimetafisico della conoscenza: la realtà rimane sempre sconosciuta, ma la correlazione rende possibile la conoscenza della sequenza d'impressioni (Levine 1996). Si tratta di uno strumento in grado di soddisfare i requisiti dell'epistemologia positivista: i numeri che stabiliscono l'intensità della correlazione sono impersonali, a condizione che si soddisfino certi requisiti in fatto di qualità dei dati e delle operazioni matematiche coinvolte.

Nel capitolo successivo ricostruirò il processo con cui questi strumenti della correlazione e dell'associazione sono stati costruiti.

CAPITOLO 3

Lo sviluppo della correlazione come strumento concettuale e operativo della ricerca empirica.

"Co-relation or correlation of structure" is a phrase much used in biology, and not least in that branch of it which refers to heredity, and the idea is even more frequently present than the phrase; but I am not aware of any previous attempt to define it clearly, to trace its mode of action in detail, or to show how to measure its degree.

Francis Galton (1888, 135)

3.1. I teorici degli errori e i fondamenti matematici della teoria della correlazione.

Il nome di Francis Galton è indissolubilmente legato alla teoria della correlazione. In effetti, egli è considerato il suo ideatore. Tuttavia, è difficile pensare a un'impresa intellettuale, pur rivoluzionaria che sia, che non abbia radici profonde, e molti storici della statistica, a cominciare da Karl Pearson, si sono dedicati a rintracciarle. Il fatto appena menzionato — come si vedrà più avanti — ha reso difficile individuare un primo antecedente intellettuale; infatti, anch'esso si può sempre collegare ad altri ancora più remoti.

Pearson (1896, 261) offrì una delle prime trattazioni delle origini della correlazione, individuando il saggio di Auguste Bravais, *Analyse Mathématique sur les Probabilités des Erreurs de Situation d'un Point* (1846) come "la prima e quasi esaustiva trattazione" dei teoremi della correlazione. Quest'idea fu poi approfondita da Yule (1897; 1909) e si diffuse con la pubblicazione del suo celebre manuale di statistica (1911). Nel 1920 Pearson presentò una storia alternativa, facendo risalire l'origine del calcolo correlazionale al saggio di Gauss, *Theoria combinationis observationum minimis obnoxiae* (1823), che in un certo senso "raggiunse la superficie normale di n variabili correlate" (Pearson 1920, 27). Denis (2000) ritiene che le radici matematiche della tecnica si trovano nel lavoro di Adrien Marie Legendre sui minimi quadrati, *Nouvelles méthodes pour la détermination des orbites des comètes* (1805).

Sembrerebbe che tutti siano d'accordo sul fatto che gli strumenti matematici — punti di partenza della correlazione — si fondino nella tradizione intellettuale della teoria degli errori, alla quale appartengono gli autori appena menzionati. Ciononostante, un completo consenso circa il suo principale precursore non si è finora raggiunto. Di seguito analizzerò brevemente le varie posizioni.

L'attribuzione di paternità a Bravais è forse la più diffusa. Egli era già stato accreditato della scoperta delle distribuzioni normali bivariata e trivariata quando Galton ancora lavorava sul tema:

I teoremi fondamentali della correlazione furono trattati per primo da Bravais [...] Egli sviluppò completamente la correlazione di due e tre variabili [...] La funzione di Galton (*Galton's function*) o coefficiente di correlazione [...] appare nel lavoro di Bravais, pur senza ricorso ad alcun simbolo (Pearson 1896, 261).

Tra le numerose memorie sulla teoria degli errori la più importante, in connessione [con la correlazione], è quella di Bravais, che già nel 1846 trattò la teoria degli errori dei punti nello

spazio. Tuttavia, egli non usa un singolo simbolo per il coefficiente di correlazione, ma la formula del prodotto-somma⁹⁹ può essere considerata un suo contributo (Yule 1909, 722).

Posizioni analoghe si trovano in Walker (1929), Westergaard (1932), Seal (1967) e Lancaster (1972). Tuttavia, essi cominciarono a sottolineare che Bravais non si era accorto delle conseguenze delle sue formule per il trattamento delle relazioni tra variabili: "gli elementi della teoria [della correlazione] si trovano nelle ricerche di Bravais del 1846, ma nella sua attuale forma essa si deve agli statistici inglesi [...] La teoria della correlazione non introdusse alcun nuovo principio; essa si basa su teoremi del calcolo delle probabilità interamente conosciuti. Ma fornì [...] delle formule facili, contenenti [...] un programma completo di ricerca, e una vasta applicazione [che le diede] un aspetto nuovo" (Westergaard 1932, 272).

Nel 1920 Pearson scrisse una memoria intitolata *Notes on the History of Correlation* in cui riconsiderava la sua originaria valutazione del lavoro di Bravais. Inoltre, fornì una spiegazione del processo per cui quest'ultimo era venuto ad essere diffusamente considerato il padre della correlazione. Pearson riporta di essere stato molto colpito dal lavoro di Galton sulla correlazione e sul problema dell'eredità, e di aver cominciato a sviluppare il concetto e i suoi strumenti: "il campo era vasto ed ero molto ansioso di fermarmi ad analizzare il lavoro degli altri [...] Volevo raggiungere nuovi risultati e applicarli [...] e quando diedi forma scritta agli appunti delle mie lezioni, chiesi l'aiuto di un collaboratore affinché riassumesse i lavori precedenti e me li riportasse" (Pearson 1920, 29). Molti anni dopo ebbe il tempo di rivisitare il saggio di Bravais, e si accorse di averne proposto delle interpretazioni fuorvianti, poi propagate attraverso gli scritti di alcuni suoi allievi, particolarmente Yule.

Rileggendo la memoria di Bravais, Pearson (1920, 29 ss.) arrivò alla conclusione che il suo obiettivo era solo misurare gli errori nella determinazione delle coordinate x, y, z di un punto nello spazio. Queste coordinate non erano misurate direttamente, ma solo funzioni degli elementi osservati $a, b, c, \dots$. Egli suppose che le quantità osservate fossero *non* correlate, nel senso moderno del termine. In sostanza, non c'è una sola riga nel saggio di Bravais che non si possa trovare nel lavoro precedente di Gauss (1823), il cui problema era esprimere la variabilità di X nei termini della variabilità delle quantità osservate $a, b, c, \dots$ e dei coefficienti differenziali A, B, C . Questo era lo stesso problema di Bravais. "L'unica parte di valore nella sua memoria è quella in cui si derivano e si trattano le proprietà di una superficie, i cui contorni sono le ellissi dell'ora familiare superficie normale", trovando inoltre la linea che corrisponde alla regressione di Galton. Tuttavia, questo non è stato il risultato dell'osservare X e Y , determinando la loro associazione, ma del fatto che X e Y erano funzioni di certe quantità indipendenti. Bravais si mantiene fedele alla tradizione gaussiana; le quantità osservate erano ritenute *assolutamente indipendenti* e lui non aveva alcuna idea circa la loro correlazione. I prodotti delle sue espressioni emersero dalle relazioni geometriche tra le quantità osservate e quelle da loro dedotte, non dalle relazioni *organiche* tra quantità direttamente osservate. Questa idea si deve completamente a Galton. Insomma, la conclusione di Pearson fu che Bravais non aveva dato alcun contributo di rilievo al problema della correlazione.

Secondo MacKenzie l'innovazione di Bravais consiste nel tentativo di stimare la posizione di un punto su un piano bidimensionale o tridimensionale.¹⁰⁰ La forma della distribuzione congiunta che raggiunse è identica a quella di Galton e Hamilton Dickson (si veda sez. 3.2), ma "determina soltanto la *forma* della legge" (MacKenzie 1981, 70). Per

⁹⁹ Yule usa l'espressione 'prodotto-somma' (*product-sum*) al posto della più frequente 'prodotto-momento' (*product-moment*).

¹⁰⁰ A sostegno della tesi secondo la quale tutte le imprese intellettuali trovano un loro antecedente — a volte persino sotto forma di un'analogia presentazione — Walker (1929) considera l'astronomo americano Adrian il primo a trattare la probabilità dell'occorrenza di due errori simultanei nella posizione di un punto. Inoltre, ella commenta i contributi di Laplace, Plana e Gauss in questa stessa direzione. I contributi di Gauss, come menzionato, erano già stati riconosciuti anche da Pearson (1920).

esporre completamente la dipendenza di X e Y egli avrebbe dovuto analizzare i coefficienti. Tuttavia, i suoi riferimenti alla correlazione di X e Y non sono stati seguiti da un tentativo di studiarla o misurarla; per lui questo non avrebbe avuto senso (*ibidem*).

Una considerazione parzialmente diversa merita, ad avviso di Seal (1967) e di MacKenzie (1981), un saggio di Schols pubblicato in Olanda nel 1875¹⁰¹, in cui tratta un problema analogo a quello di Bravais. Egli non presumeva l'indipendenza; anzi, credeva che gli errori nelle coordinate di un punto fossero il risultato di un gran numero di piccoli errori. Ciononostante, gli assi potevano essere determinati trattando gli errori *come se* fossero indipendenti. Avendo raggiunto questa conclusione che soddisfaceva i suoi interessi, Schols non tentò di formulare un'espressione dell'influenza dell'errore in una direzione o nell'altra.

Pertanto, i lavori di Bravais e Schols devono essere giudicati alla luce delle tendenze scientifiche del loro tempo; ciò significa considerarli esponenti di una data scienza normale — nel senso di Kuhn (1962) — in cui la questione della variazione e della correlazione non si poneva. Infatti, il paradigma astronomico dominante all'epoca poneva come problema fondamentale quello della valutazione degli errori di misurazione. A questo proposito contava su un orientamento teorico fondato sulla tradizione matematica del calcolo della probabilità, materializzata nella teoria degli errori e della loro distribuzione normale. Il suo atteggiamento era piuttosto formale, deduttivo, più vicino alla tendenza razionalista francese. Di conseguenza, "per nessuno di loro la dipendenza statistica costituiva un legittimo oggetto di attenzione" (MacKenzie 1981, 71).

Quelli che sottolineano il ruolo di Bravais come precursore della moderna correlazione abitualmente pongono l'accento sulla sua (apparente) considerazione di quello specifico problema. Al contrario, quelli che ricordano i nomi di Legendre e di Gauss puntano sul loro sviluppo degli strumenti matematici fondamentali del calcolo correlazionale — in particolare del metodo dei minimi quadrati — ma non della correlazione stessa. Del rapporto indissolubile tra teoria della correlazione e metodo dei minimi quadrati teneva il massimo conto Pearson (1920) nella sua ricostruzione dei precursori di Galton; il punto era già stato proposto da Yule:

Il metodo della correlazione è solo un'applicazione ai propositi della ricerca statistica del noto *metodo dei minimi quadrati*. È pertanto impossibile separare completamente la letteratura della teoria della correlazione [...] dal metodo dei minimi quadrati (Yule 1909, 722).

La controversia sulla priorità di Legendre o di Gauss nella storia della correlazione, pertanto, non è altro che un riflesso della disputa circa le origini del metodo dei minimi quadrati, brevemente riferita nel primo capitolo (si veda sez. 1.2). Molti autori che non tematizzano quest'aspetto presentano Legendre come il suo inventore, facendo commenti piuttosto marginali circa la controversia sopra menzionata (si veda ad esempio Denis 2000).¹⁰² Altri invece affermano che il metodo non nacque all'improvviso e mettono in relazione il lavoro di Legendre con alcuni dei problemi scientifici più rilevanti del suo tempo (Micheli e Manfredi 1995). Stigler (1986) valuta il metodo come il corollario di sviluppi iniziati a metà del settecento, e ritiene che se Legendre fu il primo a comunicarlo effettivamente, Gauss — come lui stesso aveva energicamente affermato — probabilmente lo usava dalla fine del settecento (Stigler 1981; 1999). Pearson (1920) non fa alcun riferimento né esplicito né implicito a Legendre, e ribadisce che le origini del calcolo correlazionale si trovano nelle memorie di Gauss sui minimi quadrati.

A parte la spinosa questione della sua invenzione, è chiaro che il metodo si svolse nella tradizione della teoria degli errori negli anni a cavallo tra il settecento e l'ottocento, legato al problema dell'errore nella misurazione in astronomia — già rilevante nella scienza

¹⁰¹ In bibliografia si riporta una traduzione francese del 1886.

¹⁰² Questo non deve ritenersi necessariamente un caso di mancata sensibilità storica. La scarsa attenzione per un tema si deve spesso al fatto che gli autori avevano un obiettivo diverso.

del settecento. I scienziati di questa disciplina riconoscevano l'impossibilità di misurare qualsiasi cosa con una completa accuratezza. Tuttavia una misurazione poteva essere ripetuta; e presto capirono l'utilità della ripetizione per ridurre l'errore nella determinazione di una quantità, così come per stimarne l'ammontare (MacKenzie 1981, 56). Allora era frequente usare la media aritmetica per determinare una quantità derivata da diverse misurazioni, che tipicamente seguivano la distribuzione normale. La stima del suo errore si chiamava errore probabile¹⁰³: "una devianza [...] dalla media delle osservazioni, di una grandezza tale che, se il numero delle osservazioni aumentasse indefinitamente, una metà degli errori sarebbe numericamente più grande e l'altra metà numericamente più piccola di esso"¹⁰⁴ (Walker 1929, 50).

Ciononostante, all'inizio dell'ottocento non c'era ancora un metodo sistematico in grado di "bilanciare" gli scarti di ogni misura rispetto alla media. La soluzione comunicata da Legendre in una sezione intitolata *Sur la Méthode des Moindres Quarrés*, del suo libro del 1805, consisteva nella somma dei quadrati degli scarti dalla media. Si trattava di un metodo matematicamente semplice, che assicurava un risultato coerente e che avrebbe sempre fornito la "miglior" soluzione in termini di riduzione dell'errore e di simmetria (Denis 2000). Il metodo fu subito accettato in astronomia e si propagò ad altre discipline, comprese le scienze umane (Stigler 1986).

Paradossalmente, malgrado questa rapida diffusione, Galton ignorava il metodo proposto da Legendre, così come ignorava il lavoro di Bravais (Pearson 1896; 1920).

Come si vedrà nella prossima sezione, l'influenza della teoria degli errori sullo sviluppo galtoniano della regressione e della correlazione si manifesta solo nell'uso della curva normale e del concetto di errore probabile per analizzare la variabilità. Quest'ultimo aspetto comportò un cambio di prospettiva: i teorici dell'errore lavoravano prevalentemente con le distribuzioni di una variabile, o al limite con variabili mutuamente indipendenti, con l'obiettivo di ridurre l'errore di misurazione; al contrario, Galton era genuinamente interessato all'errore e alla variabilità (si veda sezz. 1.2 e 1.3). I suoi concetti di regressione e di correlazione fornirono una nuova e rivoluzionaria inquadratura generale per il trattamento della dipendenza tra variabili. I precedenti strumenti tecnici sui minimi quadrati (Legendre e Gauss), e sulla superficie normale bi- o trivariata (Bravais) venivano così reinterpretati alla luce di questa nuova impostazione.

Pertanto, anche se formalmente identici ai fondamenti matematici della regressione e della correlazione moderne, è chiaro che non possono essere considerati una fonte di ispirazione per Galton, che non li conosceva affatto, né formulazioni preliminari di un problema che la scienza ancora non riconosceva. È evidente che il lavoro di Galton non emerse dal nulla; ma i suoi antecedenti intellettuali hanno una storia più complessa. La teoria degli errori permette dar conto solo di alcuni aspetti, e di solito la si privilegia per la concretezza dei suoi contributi strumentali. Altre influenze sono invece meno evidenti, ma non necessariamente meno importanti. Si pensi per esempio alla tendenza alla quantificazione nella ricerca empirica, al lavoro di Quetelet e alla sua applicazione di strumenti matematici nelle scienze umane; infine, alla tradizione naturalista britannica — in particolare le ricerche di Darwin. E tuttavia, anche se il discorso si allargasse con l'inclusione di queste e di altre tendenze intellettuali dell'epoca, sarebbe probabilmente difficile spiegare la nascita del moderno concetto di correlazione senza sottolineare il contributo originale di Galton: la scelta della variazione come legittimo problema di ricerca scientifica.

¹⁰³ Secondo Walker (1929) l'espressione 'errore probabile' ebbe origine tra gli astronomi matematici tedeschi all'inizio dell'ottocento. Per Denis (2000) essa risale alla metà del settecento.

¹⁰⁴ Si tratta, nei termini della moderna teoria statistica, della devianza interquartile.

3. 2. La svolta di Galton: l'emergere dei concetti di regressione e correlazione

Tra la fine degli anni sessanta e gli anni ottanta dell'ottocento, Galton lavorò su problemi statistici in modo relativamente isolato, compiendo una svolta teorica d'enorme importanza: i concetti di regressione e di correlazione (MacKenzie 1981, 10).

I suoi sviluppi sono strettamente legati alla ricerca empirica sul problema dell'eredità. Ci sono varie ipotesi circa le origini di questo interesse di ricerca che segnò uno spostamento verso la biologia evoluzionista e l'antropometria da precedenti interessi più vicini alla geografia, l'esplorazione naturalista e la meteorologia. Micheli e Manfredi (1995, 106) sottolineano un fatto personale: l'esaurimento nervoso legato alla constatazione che il matrimonio si era rilevato sterile avrebbe sancito quest'orientamento verso i processi genetici di trasmissione ereditaria. MacKenzie (1981, 58) trova invece una spiegazione eugenetica: l'interesse per i processi di eredità dei caratteri umani sarebbe stato condizionato da un progetto di miglioramento della specie. Altri riconducono quest'interesse, più genericamente, all'influenza della teoria evoluzionista del suo molto ammirato cugino Darwin, la cui opera principale *On The Origin of Species* (1859) fu letta da Galton negli anni sessanta. In effetti, commentando il libro Galton affermò che la nuova prospettiva "l'aveva incoraggiato a perseguire molte indagini [...] focalizzate sui temi centrali dell'eredità e del possibile miglioramento della razza umana" (Tankard 1984, 45).

Il primo frutto di questo nuovo orientamento fu il libro *Hereditary Genius* (1869), in cui Galton seguì l'impostazione di Quetelet nell'applicare la teoria degli errori alla ricerca sugli esseri umani. Egli non aveva una competenza matematica molto raffinata, ma "le tecniche basiche della teoria degli errori erano ampiamente conosciute in Inghilterra, e fu naturale che Galton si rivolgesse ad esse [...] particolarmente perché [...] Quetelet le aveva già applicato con successo ai dati umani" (MacKenzie 1981, 57). Alcuni anni dopo Galton affermava:

I metodi più elevati della statistica, che consistono nell'applicazione della legge della frequenza degli errori, si trovarono eminentemente adatti al processo dell'eredità (Galton 1889, 199-200).

E in una relazione come presidente dell'Istituto di Antropologia disse:

Non conosco nulla di così atto ad impressionare l'immaginazione che la straordinaria forma di ordine cosmico espressa dalla 'legge degli errori'. Il selvaggio, se potesse capirla, la venererebbe come un dio. Essa regna con riservata severità in mezzo alla più tempestosa confusione. Quanto più enorme è la massa e più grande l'anarchia, più perfetto risulta il suo impero. Si prenda un ampio campione d'elementi caotici e lì si disponga in ordine di grandezza, e allora, per quanto pazzamente irregolari appaiano, un'inattesa e bellissima forma di regolarità risulta essere stata sempre presente (Galton 1886, 494-95)

In *Hereditary Genius* Galton "non inventò nuovi strumenti statistici". Tuttavia, nell'ultimo capitolo "immaginò lo sviluppo d'una teoria quantitativa predittiva della discendenza" senza arrivare ancora allo sviluppo di una "legge matematica che collegasse la generazione dei genitori con quella della prole" (MacKenzie 1981, 57). Galton era perplesso a proposito di un fatto per il quale non trovava una soluzione ragionevole: "se il patrimonio ereditario che un individuo acquisisce è il prodotto di un larghissimo numero di fattori, tutti indipendenti l'uno degli altri [come si seguiva dalla teoria degli errori] [...] come avviene che un carattere singolo si ripeta identico da padre a figlio a nipote senza cambiare affatto?" (Micheli e Manfredi 1995, 106). Questo problema lo indusse a una svolta, che per il momento era semplicemente argomentale e piuttosto vaga. Ciononostante, cominciava a constatare l'inadeguatezza dell'impostazione della teoria degli errori per i suoi propositi. Nella sua autobiografia ricostruì la situazione in questo modo:

Gli obiettivi primari della legge gaussiana degli errori erano esattamente opposti, in certo senso, a quelli per cui li applicai. [I teorici degli errori] cercavano di liberarsi, o di provvedere una giusta riduzione, dell'errore. Ma questi errori e devianze costituivano proprio la cosa che m'interessava preservare e conoscere (Galton 1908, 305).

Secondo MacKenzie (1981) la rottura di Galton con l'approccio dei teorici degli errori divenne totalmente evidente in *Statistics by Intercomparisons* (1875). Infatti, in questa memoria egli si riferì al concetto di errore come "assurdo" (Galton 1875, 35): dal suo punto di vista definire un uomo estremamente abile come un enorme errore della natura era ridicolo. Ma gli aspetti interessanti di questo saggio per il futuro della correlazione vanno oltre la critica dell'idea di errore. Innanzitutto, si mette in evidenza un'altra questione che lo allontanò dalla tradizione precedente. I teorici degli errori seguivano un atteggiamento tipicamente razionalista e deduttivo. Galton dichiarava invece di prestare più attenzione alle osservazioni, e di ancorare le sue conclusioni all'analisi di distribuzioni "empiriche" anziché alle deduzioni aprioristiche:

La legge della frequenza degli errori afferma che le grandezze divergenti dalla media con questo o quel multiplo dell'errore probabile accadranno con questo o quel grado di frequenza. La mia proposta è invertire il processo [...]: questa o quella grandezza accadono con questo o quel grado di frequenza, pertanto le differenze tra esse e il valore medio sono tali e tali, espresse in unità di errore probabile (Galton 1875, 38).

Un altro aspetto da far notare è che, in questo periodo, Galton era già interessato a sviluppare "un metodo per ottenere semplici risultati statistici che [abbia] il merito d'essere applicabile ad una moltitudine di oggetti al di là dei correnti limiti della ricerca statistica [...] e che [sia] utile in vari rami della ricerca antropologica"¹⁰⁵ (ivi, 33).

Inoltre, spiega dettagliatamente le misure di tendenza centrale e di variabilità che pochi anni dopo avrebbe utilizzato nel calcolo della regressione e della correlazione: la media — che secondo la sua definizione sarebbe in realtà la moderna mediana¹⁰⁶ — e la devianza interquartilica.

Infatti, la sua media "rappresenta il valore medio di una serie almeno in uno dei molti sensi in cui il termine può essere usato". Per esempio, "la velocità media di un gruppo di cavalli [sarebbe stata] quella del cavallo che nel correre [occupava] il posto intermedio (*middlemost*)" (ivi, 34). Una volta determinata la media, "Il seguente passo di grande importanza consisteva nel determinare la divergenza"¹⁰⁷ della serie — cioè la tendenza dei singoli oggetti a divergere dal valore medio. La più conveniente misura di divergenza [risultava] da considerare, da un lato, l'oggetto che [aveva] il valore medio, e dall'altro, gli oggetti la cui divergenza, in ogni direzione, [era] tale che la metà di tutti quelli schierati su ognuno dei lati della media [aveva] una divergenza maggiore, e l'altra metà una divergenza minore. La differenza tra la media e ognuno di questi due oggetti [sarebbe stata] la misura in questione, tecnicamente [...] chiamata 'errore probabile'. Galton propose "semplicemente di

¹⁰⁵ Ovviamente non si riferiva alla correlazione, ma a un metodo utile per la comparazione di distribuzioni, che era già stato proposto e fino ha un certo punto utilizzato in *Hereditary Genius* (Galton 1869, 26). Questo orientamento verso la ricerca di metodi semplici e "universalmente" applicabili sarà sempre presente nel lavoro di Galton.

¹⁰⁶ Dato che Galton presuppone popolazioni distribuite normalmente, la media e la mediana, nel senso moderno, coincidono. Resta il fatto, però, che le osservazioni empiriche non sempre si attengono a questa distribuzione teorica, e non a caso i suoi successori notarono il problema di usare la mediana anziché la media nel calcolo della correlazione (si vedano ad esempio Weldon 1892 e Pearson 1896). Anzi, lo stesso Galton, nel parlare della devianza interquartilica, riconosce che le distribuzioni simmetriche, secondo la sua esperienza, sono molto rare (Galton 1875, 35).

¹⁰⁷ Galton usa la parola *divergency* (divergenza) invece del moderno termine statistico *deviation* (scarto).

prendere gli oggetti che [occupavano] rispettivamente il primo e il terzo quartile della serie"¹⁰⁸ (ivi, 35).

Due anni dopo la pubblicazione di *Statistics by Intercomparisons*, Galton tenne una relazione alla *Royal Institution* con il titolo di *Typical Laws of Heredity*.¹⁰⁹ Questo può considerarsi retrospettivamente come la prima fase della sua rivoluzione nella teoria statistica: l'originario sviluppo del concetto che più tardi sarebbe stato chiamato regressione lineare (MacKenzie 1981, 63). Galton avrebbe voluto usare dati antropometrici¹¹⁰; però la difficoltà di ottenerli lo spinse a proseguire la ricerca con delle misurazioni operate su due generazioni di semi di piselli.

Galton cominciò col prendere un gran campione di semi¹¹¹ che pesò per ottenere la media e l'errore probabile. Poi lo separò in sette gruppi, ognuno contenenti semi dello stesso peso: un pacchetto aveva i semi con i valori della media, e gli altri sei avevano semi una, due e tre volte in più e in meno l'errore probabile dalla media. Galton distribuì i semi tra i suoi amici, che dovettero farle crescere, e poi riportare la discendenza. Analizzando i dati, egli scoprì che i discendenti di ogni classe di semi originali avevano una variabilità analoga¹¹², e che si distribuivano normalmente. Inoltre, si rese conto che la media della discendenza dei semi più piccoli e di quelli più grandi non era così estrema come quella dei genitori, e che la relazione tra il peso dei semi della prima e della seconda generazione era lineare (si veda fig. 1). In questo modo arrivò al concetto di 'reversione'¹¹³, poco tempo dopo ribattezzato come 'regressione':

La 'reversione' è la tendenza della media del tipo filiale ideale a divergere dal tipo paterno, regredendo verso quello che, grossolanamente, potrebbe essere descritto come la media del tipo ancestrale (Galton 1877).¹¹⁴

¹⁰⁸ Galton riconosceva che il suo metodo non era in grado di fornire questo "errore probabile" numericamente; consentiva solo di "selezionare esemplari le cui differenze [erano] precisamente quelle cercate, e che [dovrebbero essere] apprezzate nel miglior modo possibile" (ivi, 35).

¹⁰⁹ Poi pubblicata nelle *Proceedings of the Royal Institution* (si veda Galton 1877).

¹¹⁰ Galton riferì questo particolare in una relazione tenutasi alla Sezione Antropologica della *British Association* (si veda Galton, 1885a).

¹¹¹ Al momento Galton era all'oscuro del lavoro di Mendel, che si diffuse in Gran Bretagna all'inizio del novecento.

¹¹² Questo non è altro che la moderna idea di *homoscedasticity* (Si veda Pearson 1920, 33).

¹¹³ Sembrerebbe che Galton arrivasse alla "legge di reversione" in un modo puramente empirico, semplicemente osservando i dati e deducendo la legge. Secondo MacKenzie questo è molto improbabile: "Galton aveva una nozione previa del tipo di legge che cercava: una semplice [...] asserzione matematica del rapporto tra la generazione dei genitori e quella della prole. Ci sono persino dei motivi per credere che i suoi dati non 'suggerivano' senza equivoco la legge di reversione" (MacKenzie, 1981: 63). È chiaro che, come aveva già fatto Bravais, ci si poteva arrivare aprioristicamente, per pura deduzione matematica. Tuttavia, per capire la regressione e la correlazione, era necessario un ancoraggio nelle risultanze empiriche, e una sensibilità alla variazione dei fenomeni e a quelle che Pearson (1920) chiamò "relazioni organiche".

¹¹⁴ Citato da Pearson (1920, 33).

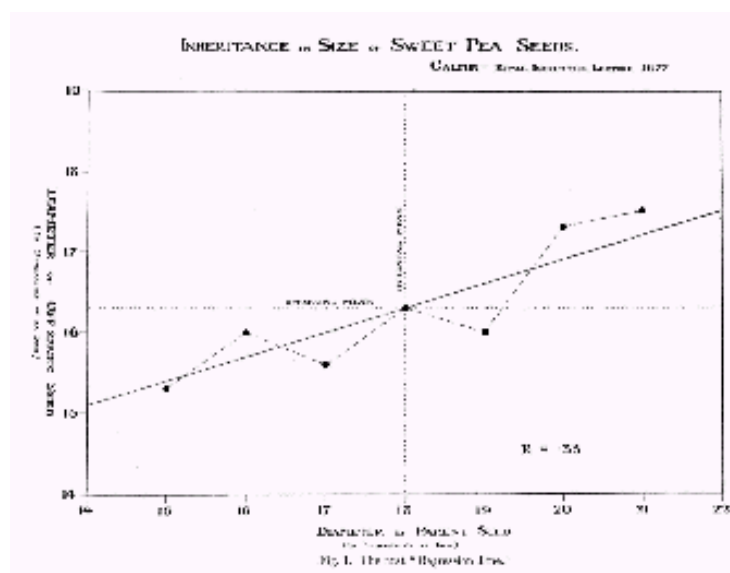


Fig. 1: Prima linea di regressione disegnata da Galton secondo la ricostruzione di Pearson.¹¹⁵

Incoraggiato dai risultati, Galton decise di replicare la sua ricerca sull'eredità dei caratteri umani. Negli anni a cavallo tra il 1870 e il 1880 si dedicò pazientemente a raccogliere questo genere di informazioni. Il Laboratorio Antropometrico stabilito a South Kensington era dedicato a quest'operazione. Verso il 1884, dopo aver offerto premi in cambio di registri familiari, riuscì ad ottenere il tanto ambito materiale empirico.

Galton decise d'investigare la statura perché era "facile da misurare, relativamente costante durante la vita adulta, e la sua distribuzione seguiva la legge della frequenza degli errori" (MacKenzie 1981, 64). Galton aveva dati sull'altezza di adulti e dei suoi discendenti. Per ogni discendente calcolò quello che si conosce come *mid-parent*, cioè la media della statura del padre e di quella della madre moltiplicata per 1,08. Poi si dedicò a cercare il rapporto tra la statura della prole e quella dei *mid-parent*. Si accorse che i dati umani manifestavano lo stesso modello dei semi, e in un modo ancora più chiaro. I primi risultati furono presentati nella sua relazione presidenziale alla Sezione Antropologica della *British Association* nel 1885. Si tratta di un lavoro di fondamentale importanza storica: secondo Pearson (1920, 35-36) esso contiene il primo diagramma mai pubblicato di due linee di regressione, la prima tabella di correlazione e la figura che indica il modo grafico della scoperta — basata su osservazioni — della superficie normale bivariata (fig. 2).

In effetti, Galton non si era limitato a riprodurre esattamente il suo lavoro precedente sui semi, ma era andato un passo oltre: "esaminando insieme le distribuzioni di frequenza [...] avvertì alcune strane caratteristiche". Usando il sistema di "plottaggio" già impiegato in altre ricerche precedenti, tracciò delle righe orizzontali e verticali. L'intersezione d'ogni paio di righe definiva quattro quadrati adiacenti. Galton scrisse in ognuno la frequenza corrispondente, e poi tracciò linee che univano tutte le frequenze di uguale valore. Con sorpresa scoprì che "le linee tracciate attraverso le frequenze d'uguale valore formavano una serie di simili ellissi concentriche", il cui centro era l'intersezione delle righe che rappresentavano la media della statura dei genitori e quella dei figli (si veda fig. 2). "Galton si rese conto che queste tendenze potrebbero essere state l'indizio di una comprensione più

¹¹⁵ Si veda Pearson (1920, 35; 1914-1930 vol. 3, cap. 14).

profonda della regressione" (MacKenzie 1981, 65). Nelle sue memorie Galton ricorda il modo quasi improvviso in cui la soluzione emerse:

Mentre aspettavo il treno in una stazione vicina *Ramsgate*, ed esaminando attentamente il diagramma nel mio taccuino, mi colpì che le linee di uguale frequenza formavano delle ellissi concentriche. I casi erano pochi per dare la certezza, però [...] mi avvicinavo a una soluzione. Un disegno più accurato confermò robustamente la mia impressione iniziale (1908, 302).

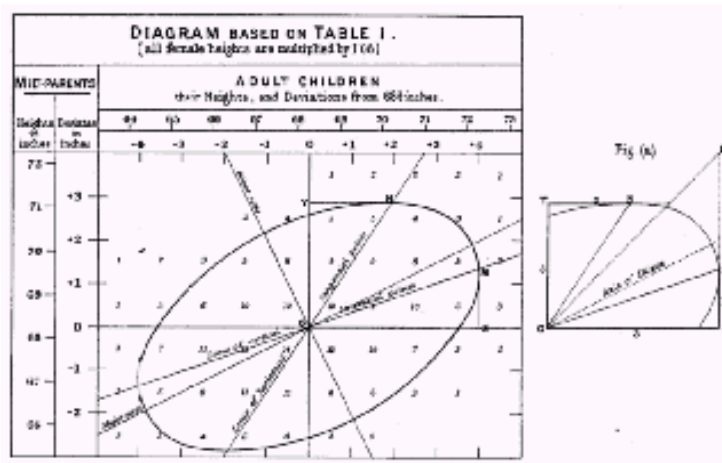


Fig. 2: Distribuzione congiunta della statura di padri e figli, con indicazione di una delle ellissi concentriche e della relazione geometrica tra le rette di regressione e le assi dell'ellisse.¹¹⁶

Galton tentò di costruire un'equazione che rappresentasse la superficie di frequenza congiunta descritta dalle ellissi. Insicuro delle proprie capacità matematiche, cercò la collaborazione di Hamilton Dickson, un noto esperto di Cambridge. La soluzione, analoga a quella proposta da Bravais nel 1846 (si veda sez. 3.1), emozionò profondamente Galton:

Mai ho sentito un bagliore di lealtà e di rispetto alla sovranità e al maestoso impero dell'analisi matematica come quando la sua risposta¹¹⁷ mi raggiunse, confermando, per via del puro ragionamento matematico, le mie varie e laboriose conclusioni statistiche molto più minutamente di quanto avessi osato sperare, dato che i dati originali erano piuttosto grezzi, e si doveva levigarli con tenera cautela (Galton 1885b, 255).

Le formule di Hamilton Dickson furono incluse come appendice nel saggio *Family Likeness in Stature* (1886). A parte la formalizzazione matematica, questo lavoro non presenta innovazioni nella teoria della correlazione. L'unico aspetto curioso è che Galton abbandonò temporaneamente la notazione r (che già dal 1877 simboleggiava la "reversione") e introdusse ω .

Il fondamentale passo successivo fu compiuto nel 1888, quando Galton cominciò a riconoscere che gli strumenti statistici derivati dalla ricerca sull'eredità avevano ampie implicazioni, avviando lo sviluppo del concetto di correlazione. "Agli occhi moderni, il passo dalla 'regressione' alla 'correlazione' sembra evidente" ma fu necessario un nuovo impulso, uno stimolo che venne dal sistema di identificazione personale ideato dal criminologo francese Bertillon, basato sulle misure antropometriche (MacKenzie 1981, 65). Nella celebre

¹¹⁶ Questo diagramma è riprodotto in Pearson (1920, 36; 1914-1930 vol. 3, cap. 14).

¹¹⁷ Fa riferimento alla risposta di Hamilton Dickson.

memoria *Co-relations and their Measurement, chiefly from Anthropometric data* (1888) Galton affrontò questo problema¹¹⁸ (si veda oltre, sez. 4.1) e usò per primo il termine 'correlazione'¹¹⁹ nel senso moderno.

Infatti, Galton non inventò il termine; esso era di uso frequente nella letteratura scientifica della seconda metà dell'ottocento. Grove (1846) e Carpenter (1851) lo utilizzarono per definire le relazioni mutue delle differenti forze della natura (Winter 1997). Anche Huxley (1864) lo usava. In biologia si impiegava per designare un principio d'interdipendenza degli organi. Per esempio Darwin scrisse sulla correlazione nel 1868, per descrivere le parti di un insieme connesse tra loro¹²⁰. Nel suo esemplare del libro di Darwin Galton aveva sottolineato le espressioni "sono fino a un certo punto" e "insignificante", entrambe riferite alle caratteristiche delle correlazioni. In *The Principles of Science* Jevons usò il termine in vari passaggi; per lui correlazione implicava che le cose "[erano] così legate che *quando una c'[era] l'altra c'[era], e quando una non c'[era] l'altra non c'[era]*" (1874, 354). Nella copia di Galton, conservato nei *Galton Archives*, questo passaggio è sottolineato, e si legge il commento "bella parola" (*nice wd.*). Secondo Stigler (1999), Galton prese il termine da Jevons, anche se quest'ultimo era contrario all'idea e alla sua utilità. Non a caso Galton cominciava il saggio richiamato sopra con questa affermazione:

'Co-relazione o correlazione della struttura' è una frase molto usata in biologia [...] e l'idea è ancora più frequentemente presente della frase; ma non sono al corrente di alcun precedente tentativo di definirla chiaramente, di tracciare in dettaglio il suo modo d'agire, o di mostrare come misurare il suo grado (Galton 1888, 135).

E poche righe dopo definiva il termine:

Due organi sono detti co-relati quando la variazione di uno è accompagnata di solito da una variazione nell'altro, e nella stessa direzione.¹²¹ Pertanto la lunghezza di un braccio è co-relata con quella della gamba, perché una persona con braccia lunghe solitamente ha gambe lunghe, e viceversa (*ibidem*).

Galton si rese conto che il problema della relazione tra le parti del corpo umano era simile ad altri già studiati, particolarmente a quello della discendenza familiare nella statura; e concluse che questi problemi analoghi esemplificavano una questione più generale¹²²: la correlazione.¹²³ La sua conclusione si materializzò in un articolo uscito poco dopo, *Kinship and Correlation* (1890). Galton la presentò così:

Il mio primo passo fu prendere un gran foglio di carta, tracciare linee traverse e segnare una scala appropriata alla statura verso l'alto e un'altra appropriata al cubito (cioè, la lunghezza dal gomito piegato fino alle dita estese) verso un lato. Dopodiché cominciai a 'plottare' ogni coppia di osservazioni della statura e del cubito della stessa persona [...] ponendo un segno di matita all'intersezione delle linee che corrispondevano a questi valori. Procedendo in questo modo, e

¹¹⁸ Galton aveva a sua disposizione una gran quantità di dati per cercare la correlazione delle varie parti del corpo umano. Per scrivere il suo articolo lavorò con un campione di 350 maschi di 21 anni e più, ottenuto nel suo Laboratorio Antropometrico (1888, 136).

¹¹⁹ In realtà Galton usò inizialmente il termine 'co-relazione', poi divenuto 'correlazione'.

¹²⁰ Citato da MacKenzie (1981, 66).

¹²¹ Non fu Galton, bensì Weldon, a concepire l'idea di una correlazione negativa (si veda oltre).

¹²² Due anni dopo Galton definiva la scoperta di questa generalizzazione: "pochi piaceri intellettuali sono più intensi di quelli goduti da una persona che, mentre è occupata in una ricerca specifica, improvvisamente percepisce che essa ammette un'ampia generalizzazione, e che i suoi risultati sono buoni anche in direzioni prima insospettite" (1890, 419).

¹²³ Nella sua autobiografia Galton racconta le circostanze in cui emerse l'idea della correlazione. Egli stava nei dintorni del Castello di Naworth, dove era stato invitato, quando una pioggia momentanea l'obbligò a cercare rifugio, e lì l'idea di correlazione "lampeggiò, e [dimenticò] tutto per un momento nella [sua] grande gioia" (Galton 1908, 300).

man mano che il numero di segni nel foglio aumentava, la forma della loro disposizione divenne più definita. All'improvviso mi colpì che la forma assomigliava a quella già familiare da quando mi occupavo di eredità [...] La riflessione presto mi fece capire che questi [...] problemi erano, in principi, identici a quello vecchio che avevo già risolto, e che tutti non erano altro che casi speciali di un problema molto più generale; vale a dire, quello della correlazione (Galton 1890, 420).

Se la variazione fosse dovuta "solo a cause comuni, la co-relazione sarebbe stata perfetta [...] Se non fosse dovuta per nessun rispetto a cause comuni, la co-relazione sarebbe stata nulla."¹²⁴ Tra questi due estremi c'era un'innumerabile quantità di casi intermedi, e [...] la forza della co-relazione in ogni caso particolare si poteva esprimere con un semplice numero" (Galton 1888, 135-136). Tuttavia, questo numero — la misura della correlazione — doveva essere lo stesso qualunque fosse la variabile considerata indipendente, e Galton sapeva ormai dalla ricerca sulla statura che i coefficienti di regressione non soddisfacevano questo requisito. Siccome ciò si doveva ai differenti errori probabili di ogni variabile, per trovare un coefficiente che soddisfacesse il requisito si doveva eliminare l'influenza dell'unità di scala di ciascuna variabile.

Dopo aver riscalato le variabili dividendole per il proprio errore probabile, Galton trovò che una semplice relazione teneva. Tutte e due le variabili mostravano una regressione lineare sull'altra, e i coefficienti di regressione erano equivalenti. Galton chiamò r il valore del coefficiente (MacKenzie 1981, 67), reintroducendo la notazione del 1877, che aveva abbandonato nel 1886.

Come si doveva operare questo ri-scalamento? Galton fornì un chiaro esempio che riguardava la lunghezza del dito medio sinistro e la statura: "le osservazioni hanno mostrato che uno scarto di un pollice nel dito è associata in media a una di 8 pollici e 19 centesimi nella statura; e che uno scarto di un pollice nella statura è associata in media con una di 6 centesimi nel dito. Non c'è reciprocità numerica in queste cifre, perché le [loro] scale di dispersione [...] divergono grandemente; nel raggio di 15 a 175". Tuttavia, i 6 centesimi moltiplicati per la frazione $175/15$, e le 819 centesimi moltiplicati per quella di $15/175$, concorrono nel dare "l'identico valore di 7 decimi, che è l'indice di correlazione (*index of correlation*)" (Galton 1890, 429-430).

Le caratteristiche fondamentali di questo "indice di correlazione" tra due variabili [Y e X], secondo Galton erano quattro: "(1) che $Y = rX$ per tutti i valori di Y ¹²⁵; (2) che r aveva un valore identico, indipendentemente da quale variabile fosse stata presa come indipendente; (3) che r era sempre inferiore al 1; (4) che r misurava la forza della co-relazione" (Galton 1888, 145).

Nel 1889 Galton pubblicò *Natural Inheritance*, probabilmente il suo libro più importante, e corollario di tutte le sue ricerche precedenti sui problemi di regressione e di correlazione nell'eredità. Non introdusse alcun nuovo sviluppo, ma presentò i concetti in

¹²⁴ Per Galton questo si capiva meglio con un esempio: "la relazione tra la forma e le caratteristiche di due fratelli è il risultato di tre gruppi d'influenze: (1) quelle che hanno influenzato entrambi i fratelli; (2) quelle che hanno influenzato il primo fratello ma non il secondo; (3) quelle che hanno influenzato il secondo fratello ma non il primo. Se non ci fossero le cause (2) e (3), i due fratelli sarebbero identici; se non ci fossero le cause (1), i fratelli non si assomiglierebbero per niente" (1890, 423).

¹²⁵ Pertanto definisce la linea di regressione (Pearson 1920).

modo articolato e sistematico, e il libro attirò l'attenzione di quelli che negli anni successivi avrebbero formalizzato e generalizzato gli strumenti della correlazione.

3. 3. Formalizzazione e generalizzazione dei concetti di Galton: i moderni strumenti della correlazione.

Pearson (1920, 41) afferma che una delle conseguenze più importanti della diffusione di *Natural Inheritance* fu l'attrazione che il concetto di correlazione esercitò su tre individui: Edgeworth, Weldon e lui stesso. Sin dall'inizio degli anni novanta, Edgeworth avrebbe curato alcuni degli sviluppi del concetto nel campo della statistica teorica; Weldon, invece, fu uno tra i primi ad usarlo estesamente nella ricerca empirica, e Pearson, sfruttando le sue capacità matematiche, dette alle formule la loro veste attuale.

Nel 1890 Weldon scrisse il suo primo articolo sulla correlazione. In esso comunicava i risultati di una ricerca sulle relazioni fra gli organi dei gamberetti, chiamando i coefficienti di correlazione 'funzioni di Galton' (*Galton's functions*), ma senza introdurre alcuna innovazione. Nel suo secondo saggio statistico sostituì la media aritmetica alla mediana nel calcolo delle correlazioni, aumentando l'accuratezza dei risultati, e riaffermò l'uso del simbolo r per il coefficiente (Weldon, 1892). Inoltre, trattò i primi coefficienti di correlazione negativi. Il terzo contributo di Weldon, pubblicato nel 1893, presentò i coefficienti di correlazione fra gli organi dei granchi di Plymouth e di Napoli. La necessità di calcolare l'errore probabile di r per valutare la sua significatività divenne patente in quest'articolo; l'argomento sarebbe stato poi approfondito da Pearson, che in questo periodo iniziò la sua collaborazione con Weldon.

Edgeworth scrisse la sua prima memoria sulla correlazione nel 1892, seguendo un suggerimento di Galton, che non era riuscito a risolvere matematicamente il problema della correlazione multipla. In *Correlated Averages* (1892), egli sviluppò la *Edgeworth's Expansion*, una generalizzazione al caso di tre variabili della distribuzione congiunta bivariata. L'articolo fu accolto favorevolmente¹²⁶, anche se conteneva errori di stampa e omissioni che rendevano molto difficile seguire la derivazione delle formule matematiche. Il suo contributo più duraturo fu la proposta di chiamare 'coefficiente di correlazione' la misura r che Galton aveva chiamato 'indice di correlazione' e Weldon 'funzione di Galton'.

Nel 1893 Edgeworth si occupò del problema della correlazione nei fenomeni sociali ma senza innovare teoricamente. Egli si limitò a mostrare come la "doppia legge degli errori"¹²⁷ fosse soddisfatta dai dati di Galton su esseri umani: "esiste una proporzione matematica, tanto quanto artistica, tra le parti della struttura umana" (1893, 671). Circa il dibattito sull'uso della media aritmetica o della mediana nel calcolo della correlazione — chiamato da Edgeworth *the battle of the means* — egli riteneva fosse futile nel caso di distribuzioni simmetriche, dato che entrambe avrebbero avuto lo stesso valore (cosa che ora ci risulterebbe ovvia). Tuttavia, particolarmente nell'applicare calcoli matematici ai fenomeni sociali, molta attenzione si doveva prestare alle distribuzioni asimmetriche. Contrariamente all'opinione di Weldon e di Pearson, concluse che "considerando aspetti teoretici quanto pratici, [sarebbe raccomandabile] in alcuni casi l'uso della mediana" (Edgeworth 1893, 673).

Senza dubbio il contributo più notevole alla teoria della correlazione dopo la pubblicazione degli articoli di Galton alla fine degli anni ottanta si trova nella memoria di Pearson *Mathematical Contributions to the Theory of Evolution: Regression, Heredity, and Panmixia* (1896). L'articolo è celebre per la presentazione della formula prodotto-momento e la sua giustificazione come la "miglior" formula del coefficiente di correlazione. Tuttavia, gli

¹²⁶ Nella sua storia della correlazione, Pearson (1920) valutò invece negativamente questo contributo di Edgeworth, che a suo avviso aveva lasciato il problema della correlazione multipla in uno stato incompleto.

¹²⁷ Distribuzione normale bivariata.

aspetti teorici rilevanti di questo saggio sono molteplici e richiedono una trattazione più approfondita.

Pearson incomincia con una frase che evoca per analogia quella usata da Galton nell'apertura del suo celebre *Co-relations and their Measurement, Chiefly from Anthropometric Data* (1888):

I problemi della regressione [...] sono stati trattati da Francis Galton nel suo straordinario lavoro *Natural Inheritance*; però anche se lui ha mostrato dei metodi esatti — tanto sperimentali quanto matematici — per trattare i problemi dell'eredità, non pare che i matematici abbiano sviluppato fino ad ora la sua trattazione [...]. La presente memoria è dedicata all'espansione e al completo sviluppo delle idee di Galton (Pearson 1896, 254-255).

Nel saggio, Pearson definisce la correlazione come la situazione in cui la media di una variabile — per esempio la dimensione di un organo — risulta una "funzione" dei valori di un'altra variabile (ivi, 257). La sua ricerca sulla correlazione intendeva "raggiungere [queste] fondamentali formule con una chiara deposizione *degli assunti*, e con speciale attenzione a quello che sembrava legittimo nel caso dell'eredità" (ivi, 261). Il problema fondamentale di Pearson era il metodo più pratico per determinare la formula del coefficiente di correlazione. Mentre le precedenti soluzioni di Galton e di Weldon non erano soddisfacenti, il modo più adatto era quello già proposto da Bravais nel 1846, che forniva il miglior risultato e che non presentava difficoltà di calcolo:

Pare che il risultato osservato sia il più probabile quando r riceve il valore $S(xy) / (n\sigma_1\sigma_2)$. Questo valore non presenta difficoltà pratiche nel calcolo, e pertanto dobbiamo accettarlo. È il valore dato da Bravais, salvo che lui non dimostrò che fosse il migliore (ivi, 265).

Pearson argomentò che $S(\xi\eta)$ — o $S(xy)$ come scrisse in altri passaggi della memoria — era un prodotto-momento¹²⁸, e che il suo svanire indicava l'assenza di correlazione.¹²⁹ Nella simbologia moderna si tratta della covarianza (Σxy) — la somma dei prodotti di ogni coppia di scarti di X e Y dalla rispettiva media — che risulta zero se le variabili non hanno relazioni e il cui limite superiore è l'unità. Da questo punto di partenza Pearson (ivi, 275) arrivò alla formula del coefficiente di correlazione, che propose di calcolare come segue¹³⁰:

$$r = \frac{S(xy)}{(n\sigma_1\sigma_2)}$$

La formula prodotto-momento di r fu comunicata in questo saggio, ma Yule afferma che Pearson la illustrava già almeno un anno prima nelle sue lezioni di statistica all'*University College*. Infatti, nei suoi appunti dalle lezioni del 1894-95 si legge: "il miglior valore di r , deduzione della formula prodotto-somma"¹³¹ (Yule 1938, 201). Tuttavia, come ricorda lo stesso Yule, Pearson dichiarava ai suoi studenti che la formula non era stata ancora applicata nella ricerca empirica.

Quando Pearson incominciò la collaborazione con Weldon ebbe l'opportunità di leggere i suoi articoli. Particolarmente quelli del 1892 e del 1893 convinsero Pearson della necessità di basarsi sull'errore probabile di r per dare un'interpretazione solida ai suoi risultati.

¹²⁸ Pearson aveva già usato il termine 'momento' in una lettera a *Nature* del 26 ottobre 1893.

¹²⁹ Come è noto il coefficiente r può assumere valore zero anche nel caso di due variabili correlate in modo non lineare (si veda Blalock 1960/1986, 395). Già Yule (1897, 821) aveva notato che se la regressione era molto lontana dal modello lineare, occorreva molta prudenza nell'uso di r .

¹³⁰ Usando la notazione contemporanea, i manuali di statistica presentano la formula seguente: $r = \Sigma xy / \sqrt{(\Sigma x^2)(\Sigma y^2)}$ che significa il rapporto tra la covarianza e la radice quadrata del prodotto delle devianze di X e di Y (Blalock 1960/1986, cap. xvii).

¹³¹ Si ricordi che Yule usava l'espressione prodotto-somma anziché prodotto-momento.

Non dovrebbe sorprendere che già nel 1896 egli fece il tentativo di introdurre una formula per l'errore probabile del coefficiente di correlazione. Tuttavia, essa era leggermente errata; il risultato corretto sarebbe stato raggiunto due anni dopo dallo stesso Pearson e dal suo collaboratore Filon (Pearson e Filon 1898).

Un altro aspetto fondamentale trattato da Pearson nel 1896 fu il problema della correlazione multipla. Galton aveva risolto in sostanza il problema della correlazione fra due variabili. Poi aveva affrontato la questione della sua generalizzazione ma senza riuscire a trovare una soluzione matematica. Nel 1892 Edgeworth considerò il problema della variazione congiunta di tre variabili, e suggerì che il metodo poteva essere esteso a modelli con oltre tre variabili. Pearson accolse questo suggerimento e si occupò di sviluppare gran parte la teoria della correlazione e regressione multipla, che "consiste nel definire un'equazione lineare tra una variabile X_1 e altre variabili X_2, X_3, X_4 , etc., in modo che la somma dei quadrati degli errori nello stimare X_1 dalle sue variabili correlate X_2, X_3, X_4 , etc. sia la minima possibile" (Yule 1897b, 817).

Ciononostante, la soluzione fu piuttosto teorica, giacché nel "caso generale di n variabili, la matematica diventava troppo complicata" per quell'epoca (Yule 1909, 722; si veda anche Yule 1897b, 837-838). Solo nel 1907 Yule introdusse una notazione speciale che rendeva più semplice il calcolo dei processi algebrici e aritmetici coinvolti nella correlazione multipla; e il primo trattamento complessivo dell'applicazione di una qualche forma di calcolo automatico alla correlazione (*Correlation and Machine Calculations*) fu pubblicato nel 1925 (Wallace e Snedecor).

Peraltro, Pearson si rese presto conto che includere moltissimi variabili in un'analisi di regressione e di correlazione di solito non aveva senso, e pertanto il problema pratico della difficoltà dei calcoli era meno grave. L'errore frequente era supporre che le diverse variabili incluse nell'analisi fossero "in se stesse *indipendenti*". Al contrario — riconosceva Pearson — tutte le proprietà considerate potevano essere associate o correlate tra loro in diversi gradi; ma l'esperienza mostrava che, introducendo ulteriori proprietà nel modello si ottenevano scarsi risultati:

La teoria della correlazione multipla mostra che la libertà di variare è compatibile con un numero indefinito di variabili determinanti, ma l'effettiva esperienza della correlazione mostra che sono solo poche variabili altamente correlate quelle che [effettivamente] contano (Pearson 1892/1957, 172-173).

Infine, nel testo del 1896 si trova "una delle prime trattazioni della correlazione spuria" (Melberg 2000, 4), che poi Pearson avrebbe ripreso in un saggio del 1897: *Mathematical Contributions to the Theory of Evolution: On a Form of Spurious Correlation which May Arise when Indices Are Used in the Measurement of Organs*. Secondo Aldrich (1995) Pearson usò l'espressione 'correlazione spuria' per distinguere fra quelle d'importanza scientifica e quelle che, invece, non indicavano relazioni organiche. Galton si interessò molto a questo nuovo concetto, e scrisse una breve nota in cui poneva l'accento sull'importanza di capire la sua genesi, dato che essa poteva avere *prima facie* un'apparenza paradossale. Egli riteneva fondamentale sensibilizzare gli statistici all'eventualità di correlazioni spurie per evitare conclusioni fallaci (si veda Galton 1897).

Yule riprese le considerazioni di Pearson sulla correlazione spuria, che a volte chiamava "illusoria" (Aldrich 1995) o "assurda" (Kendall e Buckland 1976/1980), per evidenziare il fatto che non era spuria una correlazione in se stessa, ma l'inferenza di una relazione tratta acriticamente dal valore facciale del coefficiente di correlazione (Melberg 2000). Come esempio di correlazione spuria Yule (1926) riportò il caso del rapporto tra il tasso di mortalità e la proporzione dei matrimoni celebrati nella Chiesa d'Inghilterra: il

coefficiente era .95; tuttavia era assolutamente privo di senso collegare razionalmente i due fenomeni, o stabilire tra loro un rapporto causale.¹³²

Nel 1897 si pubblicò un altro lavoro classico nella storia della correlazione. Si tratta del saggio di Yule: *On the Theory of Correlation*. Quest'articolo fu presentato dall'autore senza alcuna pretesa, sottolineando che non intendeva innovare; la sua intenzione era piuttosto didattica: raccogliere l'insieme degli sviluppi precedenti nel campo della teoria della correlazione e illustrarli diffusamente con esempi numerici.

Questa impostazione non era affatto priva di fondamento. Alla fine dell'ottocento gli statistici volevano assicurare lo status di "vere scienze" all'economia e alle altre discipline umanistiche. Tuttavia, "la ricerca delle relazioni [...] fra fenomeni economici presentava molti problemi di peculiare difficoltà, e si prestava a conclusioni fallaci. Lo statistico poteva sperimentare solo raramente, se non mai. Egli doveva accettare i dati dell'esperienza quotidiana e trattare al meglio possibile l'insieme di cambiamenti; non poteva, come il fisico, restringere il problema all'effetto di una variazione per volta". In queste circostanze, Yule riteneva necessario "offrire un metodo" di ricerca per le scienze sociali che, nonostante le sue difficoltà, "eccedesse in generalità qualsiasi altro che coprisse lo stesso terreno" (1897a, 812).

Nell'articolo di Yule le argomentazioni sono esposte in modo molto didattico e con grande chiarezza. Egli comincia rivisitando i lavori precedenti, e poi definisce i termini e gli assunti più elementari; pone l'accento sul fatto che la correlazione si applicava a "quantità numeriche".¹³³ Seguendo la tendenza tipica della statistica del suo tempo, propose di utilizzare il termine 'correlazione' e non l'espressione 'relazione causale' per riferirsi al rapporto tra variabili¹³⁴ (ivi, 812).

La sua trattazione della correlazione parte dalla regressione e dall'applicazione del metodo dei minimi quadrati. Si considerino due variabili correlate X e Y — per esempio "l'età delle donne e degli uomini al momento del matrimonio".¹³⁵ I valori si distribuiscono, con residui di lieve entità, attorno a una *curva di regressione*. In molte situazioni questa curva non diverge seriamente da una linea retta; infatti alla curva si può adattare una linea retta chiamata *linea di regressione* di Y su X , minimizzando le distanze alla linea. Per questo si usa il metodo dei minimi quadrati. Se la pendenza della linea è positiva si affermerà che, nel complesso, alti valori di Y si associano con alti valori di X . Se è negativa, si affermerà invece che alti valori di Y si associano con bassi valori di X .¹³⁶ La pendenza della linea è una misura pratica: indica il cambio nella media di Y dato un certo cambiamento di X . L'equazione della linea dà una risposta concisa a due importanti questioni statistiche: (1) se i valori alti di Y sono tendenzialmente associati a valori alti o bassi di X ; e (2) una misura del cambiamento medio del valore di Y corrispondente a un cambiamento unitario di X (Yule 1897a, 814).

Se invece di misurare X e Y in unità arbitrarie lo si fa prendendo lo scarto tipo (deviazione standard) come unità, si può calcolare r , il coefficiente di correlazione. Il miglior valore r si raggiunge con la formula prodotto-somma proposta da Bravais (1846) e ripresa da Pearson (1896).

¹³² Altri esempi interessanti sono forniti da Melberg (2000). Tra gli altri quello di Hendry (1980) che dimostrò una correlazione di .998 fra due serie storiche: l'inflazione e la pioggia caduta sul Regno Unito.

¹³³ Questo fatto mostra che già d'allora Yule era consapevole di come la natura delle proprietà analizzate limitasse il legittimo ricorso a certe tecniche (in questo caso la correlazione). Non a caso, sarebbe stato il primo a concepire tecniche alternative per determinare il rapporto tra proprietà non quantitative (si veda la sez. 3.4).

¹³⁴ Questo è un aspetto da sottolineare: Yule è stato forse il primo a riferirsi a una correlazione tra *variabili* e non tra *organi* o *caratteri*. Infatti, egli scrisse: "le quantità le cui relazioni si desidera investigare saranno chiamate *variabili*, poiché le loro grandezze variano" (Yule 1897, 812).

¹³⁵ Yule illustra questo con un diagramma relativo ai matrimoni italiani, apparentemente allora in esibizione nei salotti della *Royal Society*, elaborato dallo statistico italiano Perozzo.

¹³⁶ La pendenza, nell'equazione di Y su X ($Y = a + bX$), è designata da b . Yule (1897, 819) definisce b come la misura "del cambiamento di una variabile corrispondente in media a un cambiamento dato dell'altra; questi cambiamenti possono avere lo stesso segno o quello opposto [...] e si denominano convenientemente coefficienti di regressione."

Quando r è 1 le due variabili sono perfettamente correlate, ma quando è zero non si può dire che siano rigorosamente non correlate. La condizione $r = 0$ è necessaria ma non sufficiente (ivi, 821). Yule, anche se aveva dichiarato di non volere introdurre innovazioni, stava segnalando un punto fino ad allora trascurato: gli assunti di normalità delle distribuzioni e di linearità delle relazioni sottostanti a tutti gli sviluppi precedenti della teoria della regressione e della correlazione.¹³⁷

Il problema, per Yule, non risiedeva solo nell'ottenere la relazione ma anche nel modo di interpretarla. Nel caso di due variabili normalmente distribuite la questione era già stata risolta. Ma secondo Yule (1897, 842) la superficie normale bivariata implicava che:

- 1) la distribuzione di ogni variabile doveva essere normale;
- 2) le regressioni dovevano essere veramente lineari;
- 3) le deviazioni standard di tutte le variabili dovevano essere uguali;
- 4) tracciando delle linee di contorno attorno ai punti si doveva formare un sistema di ellissi simili e analogamente situate, il cui centro coincideva con il baricentro di tutta la superficie.

La "correlazione normale"¹³⁸ si applicava bene ai problemi di antropometria. Tuttavia Yule era più interessato all'economia e alle scienze sociali, in cui era raro trovare "correlazioni normali" (ivi, 851). Questo lo spinse a considerare il problema della correlazione fra variabili distribuite asimmetricamente (*skew correlation*), poco dopo ripreso anche da Pearson. Il *desideratum* era un "sistema appropriato di superfici che ci liberasse dalle limitazioni della superficie normale" (Pearson 1920, 44). Le intuizioni di Yule (1897a; 1897b) furono poi sviluppate da Edgeworth (1902; 1908) e dal suo allievo Bowley (1903).

Nel 1897, quando scrisse la memoria finora analizzata, Yule era un giovane assistente di Pearson. Ciononostante si era già affermato come autore di importanti contributi allo sviluppo della teoria della correlazione. La trattazione della correlazione tra variabili asimmetriche appena menzionata non era stato il suo primo contributo di rilievo; poco prima aveva introdotto il concetto di correlazione netta, poi ribattezzato 'correlazione parziale' su suggerimento di Pearson. La correlazione parziale proponeva dei coefficienti di correlazione tra due variabili qualsiasi, eliminando gli effetti della variazione di una terza (si veda Yule 1896; 1897a). Essa fu considerata poco dopo anche da Pearson e Alice Lee (1897).

Un altro contributo importante alla teoria della correlazione venne dallo psicologo Charles Spearman, pioniere nell'applicazione degli strumenti statistici alla ricerca in psicologia (si veda sez. 1.4). Nel 1904, tentando di "misurare oggettivamente" l'intelligenza, introdusse la tecnica della correlazione di ranghi e, come strumento operativo, il coefficiente ρ . Secondo Spearman (1904b, 222), la considerazione del metodo adatto per rilevare attributi psicologici lo costrinse "a entrare nella discussione generale dei metodi universalmente prevalenti per dimostrare l'associazione tra due eventi o attributi". Questo era stato appunto l'argomento principale di un altro classico saggio pubblicato lo stesso anno e nella stessa rivista, *The Proof and Measurement of Association Between Two Things* (1904a), in cui introdusse il coefficiente ρ . Egli cercava innanzitutto una misura che rispettasse il "fondamentale requisito della correlazione, vale a dire, *un'espressione quantitativa precisa*" (*ibidem*); "un valore numerico (variante convenientemente da 1 per la perfetta corrispondenza a 0 per la perfetta assenza di corrispondenza)" (ivi, 225):

¹³⁷ Peralto Pearson aveva avvertito: "Si assume che le dimensioni degli organi complessi siano determinate da molteplici e *indipendenti* cause contribuenti [...] Si assume che le variazioni nell'intensità delle cause contribuenti siano esigue se comparate con la loro assoluta intensità, e che queste variazioni seguano la legge della distribuzione normale" (Pearson 1896, 262).

¹³⁸ Questa è l'espressione usata da Yule per sottolineare il fatto che la teoria e la tecnica di Pearson si riferivano a situazioni estremamente particolari (distribuzioni normali multivariate, etc.) e che pertanto non erano automaticamente applicabili alle situazioni che non soddisfacevano questi assunti.

Il problema è molto chiaro [...]: si desidera accertare quanto tendono a corrispondere i ranghi osservati nelle diverse capacità [psicologiche] (ivi, 252).

Il coefficiente da lui proposto a questo effetto si calcola in sostanza attraverso la formula prodotto-momento, salvo che invece di utilizzare misure si impiegano i numeri che denotano le posizioni ordinali. Il coefficiente ρ , in effetti, "misura l'intensità della correlazione tra due insiemi di *rankings* o il grado di corrispondenza tra loro" (Kendall e Buckland 1976/1980, 59).

Pearson criticò pesantemente l'approccio di Spearman e respinse il concetto e il coefficiente di correlazione di ranghi. Alla base della sua impostazione giaceva la particolare concezione delle proprietà come continue e misurabili, e una conseguente ferrea opposizione a qualunque tentativo di sviluppare e applicare tecniche di correlazione che non rispettassero quest'assunto. Edgeworth e Yule sostenevano un'idea diversa e pertanto favorivano l'approccio di Spearman. Infatti, nel suo saggio del 1910, *Correlation Calculated from Faulty Data*, Spearman scrisse in nota che era compiaciuto di sapere che Yule, al contrario di Pearson, riteneva il suo metodo un passo avanti nella ricerca della correlazione fra proprietà non misurabili (si veda anche Williams 2004). Questa differente concezione della natura delle proprietà generò un'aspra polemica fra Pearson e Yule a proposito di come calcolare e analizzare i coefficienti adatti per le relazioni tra variabili categoriali, oggetto delle due sezioni seguenti.

Verso il 1910 lo sviluppo della teoria della correlazione — e secondo Yule (1909) la direzione principale che essa avrebbe preso negli anni successivi — si orientava alla considerazione delle relazioni per cui occorre equazioni non lineari (Yule 1909, 723). Pearson era stato il primo a trattare questo problema, che peraltro richiedeva ulteriori sforzi (si veda Pearson 1902; 1905).

A parte questo, la teoria e gli strumenti della correlazione erano già ben stabiliti; praticamente tutti i problemi fondamentali erano stati trattati, e molti già risolti. I diversi contributi di quasi tre decenni di lavoro ininterrotto, sparsi in tanti saggi e memorie di Galton, Edgeworth, Pearson, Yule, Spearman e altri potevano già esser presentati in modo complessivo e articolato — cosa del tutto impossibile quindici anni prima, quando i concetti e gli strumenti erano nelle prime fasi di sviluppo (Yule 1938).

Nel 1908 Hooker scrisse un articolo elementare e sintetico sulla teoria della correlazione, e nel 1911 Yule presentò la prima versione del suo classico manuale di statistica. Nelle sue quattordici edizioni questo sarebbe stato il testo di riferimento per docenti e studenti di statistica fino agli anni cinquanta.

Queste opere non presentavano innovazioni. La loro importanza risiede nella sistematizzazione della teoria e nel testimoniare due fenomeni interessanti. Da un lato l'enorme diffusione che i metodi avevano già raggiunto; essi erano applicati in campi diversi a da parte di ricercatori non esperti in statistica. Ormai l'uso degli strumenti non era limitato a quelli che avevano contribuito — o che erano in grado di contribuire — al loro sviluppo. Dall'altro lato indicava quanto fosse "matematizzata" e specializzata la statistica, al punto che i nuovi saggi sulla teoria della correlazione non potevano essere seguiti da lettori privi di una solida formazione matematica. Le opere di divulgazione erano diventate una necessità. Nella loro trattazione della correlazione si includevano non solo gli strumenti finora analizzati ma anche quelli dell'associazione fra proprietà non misurabili; che saranno presentati qui di seguito.

3. 4. Dalla correlazione all'associazione.

Nel pensiero statistico inglese dell'ottocento c'era una tradizione intellettuale — inaugurata da Jevons e poi proseguita da Edgeworth e Yule — tendente all'applicazione dei metodi statistici nella ricerca economica e sociale. I loro interessi si rivolgevano in particolare all'impossibilità di misurare molti aspetti rilevanti e alla natura asimmetrica di tante distribuzioni. Non a caso, come illustrato nella sezione precedente, fu Yule il primo a preoccuparsi seriamente dei limiti dell'assunto di normalità nella ricerca di correlazioni, e sarebbe stato anche il primo a proporre strumenti alternativi per determinare la forza di una relazione tra proprietà categoriali.

Già In *Mathematical Physics* (1881) Edgeworth aveva tentato di illustrare l'uso del ragionamento matematico senza dati quantitativi. Quando la tecnica della correlazione stava prendendo forma, Yule riprese questo suggerimento e nel 1900 introdusse il termine 'associazione' per definire le correlazioni in cui la variazione non poteva essere misurata. Questa era a suo avviso una questione di enorme rilevanza, che aveva ricevuto poca attenzione da parte degli statistici:

Nella tradizionale teoria della correlazione, normale o meno, sempre si assume di lavorare con dei materiali suscettibili di variazione continua, o quanto meno, variazioni con un *numero considerevole* di passi non continui (Yule 1900, 257).

In *On the Association of Attributes in Statistics* (1900) Yule affermò che solo i logici si erano occupati di questo problema: de Morgan (1847), Boole (1847; 1854) e Jevons (1870) erano stati in sostanza la sua fonte d'ispirazione. Jevons propose un metodo senza tuttavia mai tentare di applicarlo nelle sue ricerche empiriche.¹³⁹ Seguendo i logici, Yule usò i simboli a , b , c , etc. per denotare oggetti o individui che possedevano le qualità a , b , c , etc.; e le lettere greche α , β , γ , per denotare 'non a ', 'non b ', 'non c ', etc.¹⁴⁰

Il punto di partenza per la trattazione della relazione tra proprietà non misurabili — come affermò Yule in un saggio del 1903 — fu assumere "una semplice forma di classificazione: la 'divisione per dicotomia', assegnando gli oggetti o individui osservati ad una o altra di due classi¹⁴¹ mutuamente esclusive secondo il possesso o meno di qualche carattere o attributo" (Yule 1903, 121).

La regressione e la correlazione — come già illustrato — avevano trovato il loro punto di riferimento in una rappresentazione grafica — il diagramma a dispersione — che aiutò a Galton a "scoprire" la relazione tra variabili cardinali, fornendo inoltre un'idea della direzione e della forza della relazione. L'associazione avrebbe trovato un ancoraggio grafico analogo nella tabella di contingenza¹⁴².

¹³⁹ A proposito di Jevons Yule affermò: "mi sorprende che [lui] non abbia mai fatto un'applicazione pratica del suo metodo [...] fra la pubblicazione del suo articolo e la sua morte" (1900, 258). Anche Stigler (1999) deplora che Jevons non abbia sviluppato tecniche di analisi dei dati nel campo dell'economia.

¹⁴⁰ Sono infatti i simboli già usati da Jevons, solo che Yule mise le lettere greche al posto delle corsive, che Jevons impiegava per le categorie negative (non a , etc.). Le riteneva più convenienti per la lettura degli articoli nelle Società Scientifiche (pratica allora frequente). Egli era consapevole che le lettere greche non sempre coincidono con quelle romane; ma dato che le proprietà da considerare erano di solito poche, questo problema non si sarebbe presentato. Paradossalmente propose α , β , γ come corrispondenti di a , b , c : γ sarebbe stata l'equivalente greca di c !

¹⁴¹ Yule sottolineò che questo non comportava necessariamente l'esistenza di un limite naturale e preciso tra le due classi.

¹⁴² Esposto in Yule (1912) usando simboli già introdotti nel 1900 e nel 1903.

	Attributo b	Assenza di b (β)	
Attributo a	ab	a β	a
Assenza di a (α)	α b	$\alpha\beta$	α
	b	β	N

Tabella 1: Modello di tabella di contingenza ricostruito a partire da Yule (1900; 1903).

La notazione utilizzata aveva questo significato¹⁴³:

- N = numero dei casi.
- (a) = numero di oggetti che possedevano l'attributo a.
- (α) = numero di oggetti che non possedevano l'attributo a.
- (b) = numero di oggetti che possedevano l'attributo b.
- (β) = numero di oggetti che non possedevano l'attributo b.
- (ab) = numero di oggetti che possedevano entrambi gli attributi a e b.
- (a β) = numero di oggetti che possedevano l'attributo a e non b.
- (α b) = numero di oggetti che possedevano l'attributo b e non a.
- ($\alpha\beta$) = numero di oggetti che non possedevano né l'attributo a né b.

Secondo la sua definizione "due qualità o attributi si dicono indipendenti se la probabilità di trovarli insieme [è uguale al] prodotto delle probabilità di trovare ciascuno separatamente". Pertanto, "se due attributi (a) e (b) sono indipendenti, i prodotti delle frequenze contrarie di secondo ordine [devono essere] uguali" (Yule 1900, 270). A partire da questo principio, Yule tentò di costruire un coefficiente che fosse equivalente a quello di correlazione:

Mi sembra che il bisogno principale nel trattamento di statistiche di questo tipo sia qualche sorta di "coefficiente di associazione", che prenda il posto del 'coefficiente di correlazione' per variabili continue, e che sia una misura dell'associazione che oscilli tra la completa indipendenza da una parte e la completa associazione dall'altra (ivi, 271).

A suo avviso il coefficiente in questione doveva:

- 1) Assumere valore zero quando le variabili a, b erano indipendenti;
- 2) Assumere valore +1 quando, e solo quando, a e b erano completamente associati, cioè, quando tutti gli a erano b e tutti gli β erano α ;
- 3) Assumere valore -1 quando, e solo quando, a e β o b e α erano completamente associati, cioè tutti gli a erano β , e tutti i b erano α .

Sulla base delle considerazioni precedenti Yule propose il coefficiente di associazione Q, le cui caratteristiche si presentano di seguito alla formula:

$$Q = \frac{(ab)(\alpha\beta) - (a\beta)(\alpha b)}{(ab)(\alpha\beta) + (a\beta)(\alpha b)}$$

- 1) Quando a e b erano indipendenti il numeratore era zero e pertanto Q era zero;

¹⁴³ Yule chiamava "attributi positivi" quelli designati con lettere romane, e "attributi negativi" quelli designati con le lettere greche. Si noti che Yule di solito usava lettere romane maiuscole.

- 2) Quando $(a\beta) = 0$ o $(\alpha b) = 0$, o entrambi = 0, $Q = +1$;
 3) Quando $(ab) = 0$ o $(\alpha\beta) = 0$, o entrambi = 0, $Q = -1$.

Per illustrare le caratteristiche di Q Yule presentò degli esempi. Uno molto semplice — che riproduco — si riferiva al rapporto tra la mancata vaccinazione e un attacco di vaiolo. A questo proposito impiegò dati rilevati in varie città inglesi forniti da Humphrey (1897). Nella città di Sheffield, negli anni 1887-88, l'incidenza del vaiolo tra i bambini¹⁴⁴ vaccinati era del 7,9% e tra quelli non vaccinati del 67,6%. Se "attacco di vaiolo" si definiva come attributo a e "non vaccinato" come b , i dati si potevano rappresentare nella tabella di contingenza nella forma seguente¹⁴⁵ e si poteva calcolare il coefficiente Q per giustificare la loro associazione:

	Non vaccinato (b)	Vaccinato (β)
Con vaiolo (a)	67,6 %	7,9 %
Senza vaiolo (α)	32,4 %	92,1 %

Tabella 2: Associazione tra la vaccinazione e l'incidenza del vaiolo tra i bambini della città di Sheffield, 1887-88.

$$Q = \frac{67,6 \times 92,1 - 7,9 \times 32,4}{67,6 \times 92,1 + 7,9 \times 32,4} = .92$$

Le celle indicano ovviamente la percentuale di bambini vaccinati e non vaccinati, con e senza vaiolo: "il test d'associazione si basa necessariamente sulla *comparazione* di percentuali o di proporzioni [...] Il mero fatto che un certo numero di a siano b non fornisce informazioni; occorre infatti sapere quanti non- a siano b , o quale proporzione di a esiste in generale" (Yule 1903, 126).

È chiaro che Yule considerava le percentuali perché si era reso conto del significato della sua definizione di associazione: a e b sarebbero stati indipendenti "quando la proporzione degli a tra i b di un dato universo fosse la stessa di quella nell'intero universo". Pertanto, l'attacco di vaiolo e la vaccinazione sarebbero stati attributi indipendenti se "la percentuale di vaccinati fra gli attaccati fosse stata uguale a quella nella popolazione generale". Insomma, due attributi si dicevano positivamente o negativamente associati a seconda che (ab) fosse più o meno grande del valore che avrebbe avuto in caso d'indipendenza, o, in altre parole, a seconda che $(ab)/a$ fosse più o meno grande di $(b)/N$; o $(ab)/a$ fosse più o meno grande di $(a)/N$ " (ivi, 125-126).

Yule dava un'impostazione pratica al problema dell'associazione; era consapevole del fatto che Q avesse delle limitazioni importanti¹⁴⁶ e pertanto, senza molte pretese, dichiarò che sarebbe stato "perfettamente possibile [costruire] altre semplici funzioni con le stesse proprietà". Egli non desiderava attribuire "troppa importanza alla [sua] funzione"; il caso della

¹⁴⁴ Con meno di 10 anni di età.

¹⁴⁵ Si tratta di una ricostruzione fatta a partire dalle memorie di Yule (1900; 1903). Risulta un po' strano che abbia definito 'non vaccinato' invece di 'vaccinato' come attributo b . In ogni modo questa decisione non influenzava il risultato di Q né la sua interpretazione.

¹⁴⁶ Marradi ha dimostrato che il valore di Q è sensibilissimo a squilibri nei marginali. Esso sovrastima infatti pesantemente una relazione se i marginali sono squilibrati (1997b, par. 1.2).

relazione tra due attributi era per lui piuttosto semplice e la definizione di indipendenza pressoché intuitiva. Tuttavia, se si scegliesse Q, "le sue proprietà dovrebbero essere state investigate in modo approfondito" (Yule 1900, 272). Yule intendeva soltanto fornire uno schema generale della teoria adeguata ai casi che lo statistico doveva spesso trattare, e credeva che in mani più competenti la teoria e gli strumenti dell'associazione avrebbero potuto essere stati portati più lontano. Il suo era solo un primo passo nella direzione che considerava giusta:

Il metodo che ho suggerito ha il vantaggio di portare il caso dell'associazione in linea con quello della correlazione, assimilando il [suo] metodo e la [sua] concezione [...] a quelli dei campi meglio conosciuti (ivi, 283).

Nella memoria del 1903 Yule andò un passo oltre, occupandosi dell'associazione tra più di due proprietà. Secondo lui si doveva comunque cominciare dalla considerazione del rapporto fra ogni coppia di variabili, dopodiché si doveva rivedere quest'associazione nei sottoinsiemi definiti dagli altri attributi: "nel caso di tre attributi, ad esempio, non si deve trattare soltanto l'associazione tra a e b nell'universo generale, ma anche nell'universo dei c e nell'universo dei γ — associazioni che saranno testate confrontando $(abc)/(ac)$ con $(bc)/(c)$ o $(abc)/(bc)$ con $(ac)/(c)$, e $(ab\gamma)/(a\gamma)$ con $(b\gamma)/(\gamma)$ e così successivamente. Queste 'associazioni parziali' rivestivano grande importanza pratica come test della correttezza delle associazioni osservate. Quando a e b sono associati, si può spesso argomentare che la loro associazione non è 'diretta', ma è il risultato dell'associazione di a con c e di b con c. Quest'argomentazione poteva essere sottoposta a prova empirica cercando se a e b fossero ancora associati negli universi isolati di c e di γ " (Yule 1903, 126).

Per esemplificare la sua posizione Yule ricorse di nuovo al rapporto tra vaccinazione e insorgenza del vaiolo: esso poteva essere dovuto alla loro comune associazione con le condizioni sanitarie, assumendo che ci fosse una proporzione più elevata di persone vaccinate nella classe alta rispetto alla classe operaia, e che i primi svolgessero la loro vita in condizioni sanitarie migliori. Per controllare l'associazione "il materiale osservato doveva limitarsi a uno o l'altro [gruppo] di quelli che vivono in condizioni sanitarie adeguate o di quelli che vivono in condizioni sanitarie non adeguate" (ivi, 126).

Yule dimostrò che i test d'indipendenza diventavano più complessi quando il numero di proprietà considerate aumentava. "In quali condizioni si poteva concludere in modo legittimo che gli attributi abcd... fossero completamente indipendenti?". La risposta frequente era: "quando la probabilità di trovarli insieme era uguale al prodotto delle probabilità di trovarli separatamente"¹⁴⁷. Tuttavia questa era una risposta erranea: "dal punto di vista fisico la completa indipendenza per una serie d'attributi abcd... entro un universo dato si può sostenere solo quando ogni paio d'attributi esibisce indipendenza non solo entro l'universo in generale, ma anche in ogni sotto-universo specificato dai restanti attributi, o dai loro contrari" (ivi, 127).

In principio, *"la completa indipendenza in una serie di n attributi esiste se il criterio d'indipendenza si mantiene per tutte le frequenze di classe positiva fino a quella dell'ordine n^{esimo} "*. Tuttavia in pratica, nel caso della relazione tra tre proprietà, sarebbe stato sufficiente applicare il criterio d'indipendenza a quattro frequenze per dimostrare l'esistenza di un'associazione: $|ab|$, $|ac|$, $|bc|$, e $|ab|c|$.¹⁴⁸ Non sarebbe stato necessario controllare le nove differenti associazioni come suggerito dalla definizione (ivi, 130-131).

Il corollario delle sue riflessioni si riassunse nel seguente teorema: non si può inferire l'indipendenza di due proprietà entro un sotto-universo dal fatto che ci sia indipendenza nell'universo generale — conclusione che aveva già raggiunto nel 1900. E — specularmente

¹⁴⁷ Questa risposta emergeva, ovviamente, dall'applicazione acritica del criterio valido come test di indipendenza di due proprietà al caso di più di due proprietà.

¹⁴⁸ Nell'originale simbologia di Yule, $|ab|$ designa l'associazione tra gli attributi a e b; $|ab|c|$ designa l'associazione degli attributi a e b nell'universo dei c, e così via.

— due proprietà non esibivano necessariamente indipendenza nell'universo generale pur presentandola in ognuno degli sotto-universi considerati.

La trattazione del problema dell'associazione illusoria (*illusory association*), a volte chiamata anche *apparent* o *fictitious* dallo stesso Yule, era di conseguenza inevitabile. Essa non era altro che la controparte della correlazione spuria (*spurious correlation*) fra variabili cardinali: "ci sarà un'associazione apparente tra a e b nell'universo generale a meno che a o b siano indipendenti da c " (ivi, 134).

All'inizio del novecento Yule già aveva trattato complessivamente la teoria e gli strumenti dell'associazione. Egli aveva considerato il caso di due e più proprietà, aveva introdotto un coefficiente con caratteristiche simili a r (Q), e aveva riflettuto sulla questione dell'associazione parziale e illusoria. Pochi mesi dopo la pubblicazione del suo primo saggio del 1900, Pearson aveva proposto una trattazione fondata su assunti radicalmente diversi, dei quali derivò strumenti altrettanto diversi. In un principio sembrò che si trattasse di due approcci complementari, ma fra i due nacque un'aspra polemica. Per ben dieci o quindici anni, Yule da un lato e Pearson e i suoi collaboratori dall'altro continuarono a scrivere e a sviluppare i rispettivi approcci. Questa controversia, insieme alla teoria e gli strumenti di Pearson sono trattati nella prossima sezione.

3. 5. La controversia sulla rilevazione dell'associazione.

Mackenzie (1976; 1981) ha studiato in forma approfondita la controversia appena menzionata, che descrive giustamente come "la politica della tabella di contingenza": in apparenza, si trattò solo un "disaccordo puramente tecnico su una questione esoterica", ma gli approcci rispecchiavano in realtà differenti obiettivi nello sviluppo della teoria statistica, riconducibili a loro volta ai diversi interessi e atteggiamenti nei confronti dei problemi sociali e politici (MacKenzie 1981, 13).

Come visto nelle sezioni precedenti, verso la fine dell'ottocento gli statistici inglesi avevano raggiunto un apparente consenso su come misurare le correlazioni tra le variabili che — come la statura e il peso — consentivano una misurazione nel senso stretto del termine¹⁴⁹. Dal 1900 in poi l'attenzione cominciò a spostarsi verso l'associazione fra variabili categoriali — per le quali ovviamente non esisteva un'unità di misura.

Yule fece i primi tentativi in questa direzione, stabilendo alcuni principi teorici fondamentali dell'associazione e proponendo il coefficiente Q come una fra le possibilità alternative. Il problema generale fu ripreso poco dopo da Pearson, che seguì un percorso totalmente diverso.

Secondo MacKenzie (1981, 155 ss.; 1976, 39 ss.) l'approccio di Pearson si basò in un ragionamento più stringente dal punto di vista logico, ma meno sostenibile dal punto di vista pratico. Egli propose un coefficiente r (impiegando lo stesso simbolo già usato per la correlazione) che chiamerò invece ψ , seguendo la proposta di Yule (1912),¹⁵⁰ per non confonderli. L'assunto fondamentale nella derivazione di ψ era che, in una tabella di contingenza le categorie a_1 e a_2 (a e α nella notazione di Yule) e b_1 , b_2 (b e β per Yule) erano suddivisioni del *continuum* delle proprietà X e Y. Per esempio, la distribuzione dei valori della statura poteva essere ridotta alle categorie 'alto' e 'basso', scegliendo un punto della scala

¹⁴⁹ Si veda Marradi (1981).

¹⁵⁰ Yule (1912, 633 ss.), in un esteso lavoro dedicato al problema della misurazione dell'associazione, confronta in varie tabelle i risultati dei diversi coefficienti proposti fino ad allora applicandoli agli stessi dati. A questo proposito decise di utilizzare la lettera greca ψ per designare questo coefficiente. MacKenzie (1981) usa invece l'espressione r_T .

come criterio divisorio.¹⁵¹ Inoltre, per Pearson X e Y presentavano una distribuzione normale bivariata.

Pearson (1900) divise il piano formato da X e Y in quattro aree corrispondenti alle celle della tabella di doppia entrata. Attraverso complesse operazioni matematiche derivò il coefficiente ψ , chiamato 'tetracoric'¹⁵² (*tetrachoric coefficient*: dal greco *τετρα*, quattro, e *χορα*¹⁵³, regione). Pearson era consapevole degli assunti sottesi a questo coefficiente e del fatto che non fosse possibile controllarli.¹⁵⁴ Tuttavia, decise di considerare ψ l'unico coefficiente di correlazione adatto a una tabella di contingenza. Nel suo saggio considerò altri coefficienti di associazione, compreso Q, ma solo come approssimazioni a ψ , che pur essendo più semplici da calcolare, erano svantaggiosi in quanto si allontanavano dal valore di ψ .

Pearson capì che ψ era applicabile solo a tabelle 2x2, e mentre continuava a difenderlo, tentò di sviluppare altri coefficienti d'associazione. Il più importante di questi tentativi generò il coefficiente di contingenza C, derivato dal test χ^2 che lui stesso aveva proposto nel 1900. Com'è noto, χ^2 è una stima dell'impossibilità di una certa distribuzione dei dati nelle celle di una tabella relativa a un campione se le due variabili sono indipendenti nella popolazione. Peraltro, esso non serve come misura della forza dell'associazione: moltiplicando la frequenza d'ogni cella per una costante si moltiplica anche il valore di χ^2 . Pearson giunse all'ovvia conclusione che dividendo χ^2 per N (totale dei casi nella tabella) si evitava questo problema. Il coefficiente risultante è ϕ^2 , che Pearson chiamò *mean square contingency*:

Una misura basata su χ^2 era chiaramente attraente. Non si dovevano presupporre variabili sottostanti, e si poteva applicare a tabelle di ogni dimensione. Essa era perfino indipendente dell'ordine delle categorie delle variabili. Il problema era: quale misura basata su χ^2 si doveva usare?" (Mackenzie 1976, 42).

Pearson, come consueto nella sua impostazione, cercò un ancoraggio nel coefficiente di correlazione per proprietà continue distribuite normalmente, e scoprì un rapporto matematico tra r e ϕ^2 tale che $r = \pm \sqrt{(\phi^2/1 + \phi^2)}$.

$$\text{Quindi propose } C = \sqrt{\frac{\phi^2}{1 + \phi^2}}$$

Il coefficiente C, chiamato da Pearson 'primo coefficiente di contingenza', approssimava il valore di r nel caso che la tabella di contingenza derivasse da una sottostante distribuzione normale bivariata. Tuttavia, C non sostituì ψ . Pearson riteneva che C fosse adatto alle tabelle più grandi, perché in quelle piccole il rapporto tra C e r non teneva, e pertanto forniva una cattiva stima della correlazione.

Pearson tentò ulteriori sviluppi, sempre mantenendosi fedele all'idea di trovare un coefficiente di associazione direttamente comparabile con quello di correlazione. Il suo ultimo

¹⁵¹ È giusto notare che quando Pearson suggerisce questa riduzione di proprietà continue a categorie "apparentemente" discrete, sta pensando a variabili fondamentali per gli studi sull'eredità, argomento centrale delle sue ricerche empiriche. In questo senso, come si vedrà dopo, egli considera variabili quali 'colore della pelle', che si usava classificare in 'chiara' 'scura', etc., ma che avrebbe potuto misurarsi in termini di lunghezza di onda. Peraltro, se due proprietà potevano essere misurate, allora — nella gran maggioranza dei casi e per il calcolo della loro correlazione — non avrebbe avuto alcun senso ridurle a due classi.

¹⁵² Yule (1912) lo chiamava 'coefficiente normale'.

¹⁵³ Nei dizionari di greco non risulta un termine del genere.

¹⁵⁴ In questo senso Pearson si mantiene fedele alla tradizione inglese, inaugurata da Mill, e continuata da Edgeworth, secondo la quale in "tutte le applicazioni di concezioni matematiche all'esistenza reale [...] c'è sempre una combinazione di assunti e di ipotesi" (Edgeworth 1884, 227). Il problema è evidentemente determinare quando questi assunti ed ipotesi risultano plausibili.

tentativo consisté nella ricerca di un equivalente di ψ per tabelle maggiori di 2x2. La soluzione proposta insieme al figlio Egon fu un metodo iterativo che aggiustava qualsiasi tabella a doppia entrata a una distribuzione normale bivariata (Pearson e Pearson 1922). Il coefficiente risultante, detto 'policorico' (*polychoric coefficient*), richiedeva calcoli molto complessi che superavano le risorse computazionali del tempo (MacKenzie 1981).

Anche Yule tentò di sviluppare delle nuove misure di associazione: il coefficiente prodotto-somma r (*product-sum coefficient*)¹⁵⁵ e il coefficiente di collegamento ω (*coefficient of colligation*). Come Pearson, egli si mantenne fedele ai suoi assunti esposti nel 1900: i coefficienti dovevano soddisfare dei requisiti specifici. Dal punto di vista pratico, la differenza tra loro era che ω — come Q — assumeva il valore ± 1 in caso di associazione in un senso debole, mentre r lo assumeva solo in caso di perfetta associazione in un senso forte.¹⁵⁶

Il coefficiente prodotto-somma era il classico coefficiente di correlazione per proprietà misurabili applicato a tabelle di 2x2, ma senza assumere i presupposti eroici — in senso schumpeteriano — di Pearson. L'assunto fondamentale nella derivazione di Yule era che le due categorie di ognuna variabile corrispondevano ai valori discreti 0 e 1. I coefficienti r , Q e ω erano collegati matematicamente, ma in un senso meno stretto che i coefficienti di Pearson.

Per alcuni anni i lavori di Pearson e di Yule non crearono controversie. Anzi, le relazioni personali tra loro furono buone almeno fino alla fine del 1905. Pearson considerò Q nella sua memoria del 1900 come "estremamente elegante e semplice" (Pearson 1900, 17). Yule, nella memoria del 1903 asserì che i suoi sviluppi originali del 1900 dovrebbero esseri considerati "in connessione con un saggio successivo del Professor Pearson"¹⁵⁷ (Yule 1903, 127).

Tuttavia, nel 1906 egli criticò apertamente l'uso del coefficiente tetracorico ψ , particolarmente per il caso di tabelle più grandi di 2x2, e in generale l'assunto di normalità. Yule introdusse il concetto di isotropia: il prodotto incrociato di qualunque quartetto di celle adiacenti doveva essere uguale a quello di tutti gli altri quartetti possibili. Yule dimostrò che molte delle tabelle costruite da Pearson con dati empirici non erano isotrope, e pertanto l'uso di ψ non era giustificato.

Pearson si difese in un articolo del 1907. A suo avviso, il criterio di Yule era invalido perché non considerava l'errore probabile del risultato dei prodotti incrociati. Dato che le tabelle si costruivano su dati tratti da un campione di una popolazione più ampia, la mancata isotropia poteva essere dovuta alla fluttuazione casuale. Ad ogni modo, Pearson riconobbe che l'assunto di normalità non sempre teneva, e proprio per questa ragione sviluppò altri coefficienti come quello di contingenza.

La controversia, tuttavia, era ancora limitata e si esprimeva in termini piuttosto rispettosi. Essa si generalizzò dopo la pubblicazione del manuale di teoria statistica di Yule (1911) che dedicava ampio spazio al trattamento dell'associazione. Si trattavano a lungo i coefficienti Q e r (derivazione del coefficiente di correlazione per tabelle 2x2) e si faceva solo un riferimento superficiale a ψ . Nello stesso anno Heron, collaboratore e grande ammiratore di Pearson, scrisse su *Biometrika*¹⁵⁸ un articolo molto zelante in difesa dell'approccio di Pearson, in cui non risparmiava critiche nei confronti dei coefficienti di Yule e raccomandava di non usarli. Yule rispose con un lungo articolo nel 1912, che a sua volta suscitò una nuova

¹⁵⁵ Si tratta del coefficiente di correlazione di Pearson (r) applicato a tabelle 2x2 dopo essersi "liberato" di qualsiasi relazione "necessaria" con la teoria della "correlazione normale" (Yule 1912, 606). Su cosa intendesse Yule sul per 'correlazione normale' si veda sopra, nota 138.

¹⁵⁶ Il senso debole significava che il coefficiente era +1 quando $(a\beta)$ o $(\alpha\beta) = 0$. Il senso forte domandava invece che $(a) = (b)$ e $(a\beta) = (\alpha b) = 0$. E specularmente per la perfetta associazione negativa (si veda Yule 1912, 584).

¹⁵⁷ Yule si riferisce alla prima memoria di Pearson sul problema della "correlazione di caratteri non misurabili quantitativamente". Si veda Pearson (1900).

¹⁵⁸ Si ricordi che si tratta da una rivista fondata, diretta e strettamente controllata da Pearson.

replica — stavolta di Pearson e Heron — nel 1913. Secondo MacKenzie (1976; 1981) quest'ultimo saggio segnò la fine della fase aperta della controversia.

Yule cominciò con l'affermare che scegliere un coefficiente di associazione adatto non era una questione semplice: quelli fino ad allora sviluppati potevano dare risultati divergenti. La scelta si doveva fare, a suo avviso, tenendo conto di fattori quali la semplicità di calcolo e di interpretazione in ogni caso particolare. Così come ci sono diverse misure di tendenza centrale — diceva Yule — ci sono varie misure di associazione, e tutte possono essere utili a seconda di obiettivi e circostanze. Nessuna era perfetta, e non si giustificava il fatto di presentare il coefficiente tetracorico come l'unico giusto. Esso era il risultato di un "lavoro matematico eccezionale", però nel migliore dei casi era solo in grado di fornire "una correlazione ipotetica tra variabili presupposte". L'introduzione di "ipotesi né necessarie né controllabili" non gli sembrava "una procedura desiderabile nel lavoro scientifico" (Yule 1912, 610-612).

Pearson e Heron (1913) criticarono i coefficienti di Yule perché davano valori diversi per la stessa tabella. Inoltre, data una tabella formata partendo dalla dicotomizzazione di due variabili cardinali distribuite normalmente, nessuno di essi concordava con il valore del coefficiente di correlazione. Per queste tabelle, i tre coefficienti variavano a seconda di dove si facessero le divisioni (se vicine alla mediana o ai valori estremi). Argomentarono che Yule reificava le categorie, e che solo in rarissimi casi — quando le variabili erano veramente discrete — si potevano usare i suoi metodi. In questi casi r (coefficiente prodotto-somma derivato da Yule per tabelle 2x2) era il modo giusto di estendere la consueta teoria della correlazione. Insomma, la teoria di Yule era a loro avviso un formalismo vuoto (MacKenzie 1976; 1981).

All'origine della disputa c'era una diversa concezione della natura degli oggetti studiati e delle loro proprietà:

"Yule [...] non si ferma a discutere se gli attributi sono continui o discreti, o nasconde dietro una terminologia discreta delle variabili veramente continue" (Pearson e Heron 1913, 161).¹⁵⁹

Pearson era contrario alle dicotomie e all'idea di categorie discrete, che considerava come meri limiti concettuali. Tutte le proprietà erano per lui misurabili, solo che lo stato di sviluppo della scienza non sempre permetteva quest'operazione. Yule rispondeva ironicamente, traendo l'esempio dalle sue ricerche empiriche:

I non vaccinati sono ugualmente non vaccinati (assumendo la corretta osservazione): nessun individuo è più o meno vaccinato degli altri [...] Similmente, tutti quelli che sono morti per un attacco di vaiolo sono ugualmente morti: nessuno è più o meno morto degli altri (1912, 611-612).

Contrariamente a quanto affermavano Pearson e Heron, i saggi di Yule mostrano che lui era consapevole dei problemi della misurazione nelle scienze umane, e per questo criticava l'allargamento semantico del termine e la sua applicazione universale.¹⁶⁰ A suo avviso, la misurazione era divenuta un feticcio scienziata:

¹⁵⁹ Si noti che per Pearson e Heron dicotomie come "morto/sopravvissuto" erano solo la categorizzazione di una proprietà continua (1913, 161).

¹⁶⁰ L'uso acritico del termine 'misurazione' si trova particolarmente nell'opera di Stevens (1946), responsabile della diffusione di espressioni quali 'scala nominale'. Per una trattazione critica e una proposta alternativa si veda Marradi (1981).

La misurazione non comporta necessariamente progresso. Il fallito tentativo di misurare ciò che si desidera potrebbe far sì che la frenesia di misurare risulti nella misurazione di qualcosa di diverso — spesso dimenticando la differenza — o nell'ignoranza del fatto che alcune proprietà non possono essere misurate.¹⁶¹

Quando lavorava con proprietà discrete non misurabili né enumerabili lo statistico doveva accettare le limitazioni; la soluzione giusta non consisteva nella negazione di questi problemi o nello stabilire un'arbitraria analogia con le proprietà misurabili.

La controversia era ormai divenuta una battaglia: negli articoli si usava un linguaggio forte, non si risparmiavano ironie e a volte si scendeva ad attacchi personali. Pearson e Heron, per esempio, avvertirono del pericolo che i metodi di Yule comportavano per la moderna teoria statistica; erano preoccupati della loro diffusione e per l'invito a seguire la strada più facile anziché l'arduo cammino che la "buona pratica scientifica" domandava. Yule invitò a i suoi oppositori a trattare l'associazione su fondamenti meno ipotetici e più ragionevoli.

Il clima generale che la disputa aveva raggiunto si percepisce in un'ironica fantasia ideata da Greenwood¹⁶², in una lettera inviata a Yule l'8 novembre del 1913.¹⁶³ In essa si riproduce un passo fittizio del giornale *The Times*, datato il 1° aprile 1925:

G. Udney Yule, accusato di alto tradimento [...] fu giustiziato su una forca fuori della stazione Gower St. Una breve ma dolorosa scena avvenne sul patibolo. Mentre si aggiustava la corda, il criminale disse "il coefficiente normale è" e il seguito non fu udibile [...]. Snow fu il carnefice, e tra altri presenti c'erano lo sceriffo, Visconte Heron di Borkman e l'Onorevole W. Palin Elderton.¹⁶⁴

Al momento di andare in stampa, il mandato d'arresto di Greenwood non è stato ancora eseguito, ma la polizia ha quello che ritiene un importante indizio. Durante il consueto servizio mattutino alla Cattedrale di St. Paul, che è stato ben frequentato, il credo carlovingio¹⁶⁵ fu cantato dal coro come prescrive un decreto imperiale. Quando le solenni parole "Io credo nel sacro ed assoluto coefficiente tetracorico di correlazione" furono pronunciate, un uomo mal vestito, vicino alla porta nord, gridò "balle". In mezzo ad un'indescrivibile eccitazione, i sagrestani, armati di vari volumi di *Biometrika*, si precipitarono sul posto, ma [...] il criminale riuscì a fuggire.

Secondo Mackenzie (1976; 1981) i differenti obiettivi resero incommensurabili le posizioni di Pearson e di Yule. La logica e le dimostrazioni matematiche non erano sufficienti per risolvere il problema:

Entrambi sapevano che per qualsiasi tabella i tre coefficienti di Yule [...] normalmente non coinciderebbero [...]. Per Pearson questo era sufficiente per condannare l'intero sistema di Yule [...]. Per Yule, dall'altro lato, questo era assolutamente sperabile dato che [i tre coefficienti] erano semplicemente diversi modi di riassumere i dati osservati. Analogamente, entrambi accettavano che il valore del coefficiente di contingenza fosse influenzato dalla grandezza della

¹⁶¹ Passaggio di una lettera inviata a Kendall. Si veda Kendall (1952, 159).

¹⁶² Greenwood, con cui Yule mantenne uno stretto rapporto d'amicizia, aveva scritto pochi anni prima — essendo ancora collaboratore di Pearson nel Laboratorio Biometrico — un articolo che difendeva il coefficiente tetracorico, considerandolo il metodo "esatto" e "vero" per misurare l'associazione. Poi si avvicinò alla posizione di Yule, con il quale scrisse qualche articolo.

¹⁶³ Citata da Mackenzie (1976).

¹⁶⁴ Noti collaboratori di Pearson.

¹⁶⁵ Riferimento ironico a Karl Pearson.

tabella. Per Yule questo comportava una grave debolezza [...]. In certe circostanze il suo valore rispecchiava tanto il numero di celle della tabella quanto il grado di associazione fra le variabili. Per Pearson, dall'altro lato, questa caratteristica era scontata. Il coefficiente di contingenza era equivalente a quello di correlazione solo nel caso limite di un numero infinito di celle [...]. Nessuno contestava che, se applicato a dati tratti da due variabili cardinali distribuite normalmente, il valore di Q avrebbe variato considerabilmente a seconda di dove si facesse la dicotomizzazione (per esempio tra alto e basso). Per Pearson questo invalidava Q. Per Yule qualsiasi caratteristica di Q, che risultasse da una sua artificiale applicazione a variabili cardinali non inficiava la sua utilità per variabili categoriali (MacKenzie 1981, 168).

Con lo scoppio della Prima Guerra Mondiale il dibattito cominciò a svanire, ma la controversia non fu risolta. La dottrina statistica contemporanea tende a negare la possibilità di assegnare validità a uno solo di questi coefficienti. Goodman e Kruskal (1954, 763) ritengono che le misure d'associazione "devono essere prudentemente costruite in modo appropriato al problema sotto studio" per consentire delle interpretazioni operative. Per MacKenzie (1981, 180) "questo pluralismo teoretico è sufficientemente diffuso da concludere che l'approccio generale degli statistici moderni è più vicino all'impostazione di Yule." Il coefficiente Q è ancora trattato nei manuali di statistica e si può calcolare tramite il noto *software* SPSS. Esso continua ad essere molto popolare, in particolare tra i sociologi. Tuttavia, la costruzione di modelli per approssimare i dati — aspetto caratteristico dell'approccio di Pearson — ha guadagnato importanza negli ultimi decenni. Il coefficiente tetracorico non è totalmente scomparso. Il noto dizionario dei termini statistici di Kendall e Buckland include la voce 'coefficiente tetracorico', ma lo ritiene "uno stimatore particolare di forma piuttosto complicata" (Kendall e Buckland 1976/1980, 60).

La controversia su come misurare l'associazione mostra quanto siano stati importanti per ogni impostazione — a parte gli assunti ontologici e epistemologici — i particolari problemi di ricerca e i relativi interessi sottostanti. Nel prossimo capitolo esplorerò appunto interessi e problemi che hanno influenzato lo sviluppo degli approcci alla correlazione e all'associazione.

CAPITOLO 4

Interessi sociali, problemi scientifici e ricerca empirica: il punto di partenza e il punto di arrivo dei nuovi concetti e strumenti statistici.

I can only say that there is a vast field of topics that fall under the laws of correlation, which lies quite open to the research of any competent person who cares to investigate it.

Francis Galton (1890, 431).

It is to be hoped that statisticians working in other fields [...] will acquaint themselves with Galton's developments of new methods, and see how far can they be applied in their own fields.

John Dewey (1889, 334).

Galton provided, in the concepts of regression and correlation, the key tools for the treatment of two dependent variables [...] His work in this area arose directly from his eugenics concerns.

Donald MacKenzie (1981, 59).

3. 1. Premessa.

La formulazione originaria dei concetti di regressione e di correlazione — come già indicato — è legata alle ricerche di Galton sui problemi dell'eredità. Tuttavia, egli si rese presto conto che i nuovi metodi erano applicabili in altri campi. In *Kinship and Correlation* (1890), solo due anni dopo aver introdotto il termine 'correlazione' nel senso moderno, egli scrisse:

Mi auguro che presto si percepisca [...] che una gran varietà d'importanti questioni può essere avvicinata solo attraverso i metodi [della correlazione] [...] Sembra di esserci un ampio spazio per l'applicazione di questi metodi ai problemi sociali [...] C'è un vasto campo tematico che cade sotto le leggi della correlazione, che sono aperte alla ricerca di qualunque persona competente (ivi, 421 e 431).

Pearson imparò questa lezione dal suo ispiratore e la pose al centro della sua campagna per stabilire i metodi della statistica matematica come universalmente applicabili. A suo avviso essi sarebbero stati in grado di elevare varie discipline allo status di vere scienze:

Capii [da Galton] [...] che questa nuova concezione della correlazione portava la psicologia, l'antropologia, la medicina e la sociologia in grande misura dentro il campo del trattamento matematico [...] C'era qui per la prima volta una possibilità, non voglio dire la certezza, di

raggiungere conoscenze — valide come si ritenevano allora le conoscenze fisiche — nel campo delle forme viventi, e soprattutto, nel campo del comportamento umano.¹⁶⁶

Il problema della correlazione cominciò davvero a essere affrontato non solo in biometria e antropometria ma anche in altri ambiti che Galton aveva presagito e che Pearson riteneva bisognosi di metodi "affidabili": psicologia e sociologia economica.

In questo periodo di gestazione, come ricordava Egon Pearson (1938/1948), lo sviluppo degli strumenti della statistica era indissolubilmente legato ai problemi di ricerca; si trattava di una "statistica applicata". Con l'eccezione di Edgeworth, una considerazione teorica per sé era praticamente inesistente. La ricerca empirica nei campi menzionati, a cominciare da biometria e antropometria, fu l'ambito in cui si ebbero i contributi più duraturi alla teoria della correlazione in senso ampio. I concetti di Galton emersero dalle sue ricerche sull'eredità; la formula prodotto-momento fu proposta da Pearson in uno studio in questo stesso campo; molti aspetti rilevanti della teoria e degli strumenti dell'associazione si devono alle riflessioni di Yule su problemi di ricerca economica e sociale; infine, la correlazione di ranghi e i suoi strumenti emersero dalla ricerche psicologiche di Spearman.

Ciononostante, gli strumenti da loro costruiti si diffusero rapidamente, e poco dopo, l'analisi delle correlazioni era *routine* in molteplici discipline, in diversi ambiti istituzionali e nazionali. "La teoria della correlazione [...] fornì agli statistici delle formule facili, portanti [...] un completo programma per le loro ricerche", e questo aiuta a spiegarne l'enorme successo (Westergaard 1932, 272). Ovviamente sarebbe fuori luogo esplorare in questa sede tutte le applicazioni degli strumenti in questione; mi limito a considerare solo quelle che diedero origine ai primi sviluppi.

Come visto nel cap. 3, lo sviluppo di questi strumenti ebbe luogo interamente nell'Inghilterra degli anni a cavallo tra l'ottocento e il novecento. Ovviamente, dato che erano strettamente collegati alla ricerca empirica, anche le loro prime applicazioni ebbero luogo là. Tuttavia, è il caso di ricordare che le tecniche si diffusero rapidamente anche in America e nell'Europa continentale.

Già nel 1889, commentando *Natural Inheritance* di Galton, Dewey invitava ai suoi connazionali, in una pubblicazione dell'*American Statistical Association*, ad applicare i nuovi metodi statistici di Galton in diversi campi. All'inizio del novecento essi erano già impiegati con successo nelle ricerche sociali, economiche e finanziarie eseguite in America.

Nell'Europa continentale, ad esempio, regressione e correlazione erano già usate alla fine dell'ottocento nella ricerca psicologica, specialmente in Germania e Francia. In Italia, nel 1909 Bresciani-Turroni aveva pubblicato nel "Giornale degli Economisti" un articolo intitolato *Sui metodi per misurare la correlazione*, orientato alle sue applicazioni in economia. Verso il 1930 le varie università tedesche e svizzere contavano su laboratori che seguivano le orme di Galton e di Pearson nella ricerca statistica in biologia, medicina e genetica.

4. 2. Eugenetica, biometria e antropometria: le origini sociali, politiche e scientifiche della correlazione.

Gli interessi eugenetici di Galton sono noti. Negli ultimi dieci anni della sua vita si dedicò interamente alla campagna eugenetica. Nella sua monumentale biografia, scritta da Pearson (1914-1930), il periodo 1901-1911, fino alla sua morte, è dominato dal movimento

¹⁶⁶ Relazione di Karl Pearson. Si veda *Speeches delivered at a dinner held in University College, London, in Honour of Professor Karl Pearson*, 23 aprile del 1934. Stampati privatamente dalla Cambridge University Press, 1934. Citato da E. Pearson (1938/1948, 46).

eugenetico e dalle iniziative per istituzionalizzarlo. Nel Capitolo *Chief Epochs in the Life of Francis Galton* si trova questo schema:

Eugenics movement	...	...	...	...	...	1901—1911
Research Fellowship in Eugenics in University of London.						
“Eugenics Office”	...	...	...	...	...	1904
Transformation of “Eugenics Office” into the “Eugenics Laboratory” and its association with the Biometric Laboratory						1907
Death and by bequest Foundation of the Galton Professorship and Endowment of the Laboratory of Eugenics in the University of London	...	...	...	...	...	1911

Pearson (1914-1930, xxi).

Questo movimento era l'espressione politica e sociale di una particolare concezione della società. La premessa era che gli individui possedevano una quantità fissa — ereditata — di caratteristiche socialmente rilevanti che costituivano il valore civico: l'abilità naturale, il coefficiente intellettuale, etc.¹⁶⁷

Queste caratteristiche seguivano una distribuzione normale che era poi omologata a quella della società di classe inglese: allo strato più basso della popolazione era attribuita la minore quantità di valore civico. Il meglio della nazione stava nelle classi più alte: l'aristocrazia intellettuale, gli imprenditori di successo e i *leaders* delle professioni nascenti.¹⁶⁸

Il movimento eugenetico ispirato in Galton aveva pertanto due aspetti principali: 1) la costruzione di una visione della società le cui caratteristiche erano in sostanza il risultato degli elementi ereditati e misurabili dei suoi membri; 2) l'istituzionalizzazione di un gruppo di pressione che promuovesse le politiche sociali intese al miglioramento della costituzione ereditaria delle generazioni future.¹⁶⁹

Evidentemente, approfondire la conoscenza dei processi dell'eredità dei caratteri fisici e mentali era assolutamente prioritario.¹⁷⁰ Il legame tra la teoria statistica di Galton e l'eugenetica non è pertanto un'illusione: molti studiosi lo hanno riscontrato, a cominciare dallo stesso Pearson.

Cowan (1972) sottolinea che i sogni eugenetici di Galton fornirono la motivazione e la perseveranza mentale necessaria per districare i segreti della probabilità. Norton (1978, 9) afferma che Galton "era motivato da forti visioni eugenetiche, un uomo i cui tentativi di capire l'eredità umana furono spinti dal desiderio di mostrare il dominio della natura sull'allevamento (*nurture*); e questo, a sua volta, lo condusse a scoprire certe nozioni statistiche cruciali — in particolare quelle [...] della correlazione e della regressione". Analogamente MacKenzie (1981, 52 ss.) indica che "i bisogni dell'eugenetica determinarono in gran parte la sostanza della teoria statistica di Galton", giacché l'orientarono verso la

¹⁶⁷ Secondo MacKenzie (1981) si trattava di un insieme di assunti legati a un modello *a priori* la cui plausibilità derivava dall'analogia con l'eredità delle caratteristiche fisiche. Pearson dichiara (1914-1930, xxiv): "sono incline a concordare con Francis Galton sul fatto che l'educazione e l'ambiente producano un piccolo effetto nella mente delle persone, e che la maggior parte delle nostre qualità sia innata", e attribuisce questa tesi a Charles Darwin.

¹⁶⁸ Karl Pearson proponeva un'interpretazione analoga della struttura sociale: la sua concezione comprendeva quattro gruppi principali, definiti dalla nascita (aristocrazia tradizionale), dal capitale (imprenditori), dalla conoscenza (*élite* intellettuale della nascente classe media professionista), e dal lavoro manuale. Il polo positivo della distribuzione era costituito dalla *élite* intellettuale e il polo negativo dal proletariato.

¹⁶⁹ Secondo L. Darwin (1926, 138), presidente della Società Eugenetica fondata nel 1907, il fine della Società era "promuovere la fertilità dei migliori tipi che la nazione possiede, mentre si diminuisce il tasso di nascita tra quelli inferiori".

¹⁷⁰ Non a caso Galton propose di istituire laboratori antropometrici in cui le caratteristiche fisiche e quelle mentali di singoli individui e di intere famiglie potessero essere misurate (MacKenzie, 1981)

variabilità come un fenomeno d'importanza intrinseca. Per giustificare la sua concezione egli doveva sviluppare concetti radicalmente nuovi. In quanto eugenista, il suo principale interesse era l'impatto delle caratteristiche di una generazione sulla seguente. La dipendenza statistica di due variabili divenne cruciale: se non c'era dipendenza statistica tra le caratteristiche di due generazioni la politica di riforme eugenetiche diventava futile. Fu pertanto l'eugenetica che fece della dipendenza fra i fenomeni il *focus* della nascente teoria statistica.¹⁷¹

Queste idee non emersero dal nulla: esse sono strettamente legate alla sua interpretazione dell'evoluzionismo di Darwin e alle conclusioni tratte dall'osservazione giovanile — piuttosto dilettantesca — dei suoi contemporanei. Dalla nascita fino alla morte, il suo mondo sociale fu quello dell'aristocrazia intellettuale, e lui stesso più volte dichiarò che le sue convinzioni sull'eredità trovarono una fonte iniziale di ispirazione nell'osservazione dei suoi pari¹⁷²:

Ero stato molto impressionato da molti casi ovvi d'eredità tra i uomini di Cambridge che frequentavano l'Università nel mio tempo (Galton 1908, 288).

Le idee acquisite nel modo appena menzionato cominciarono a essere sistematicamente controllate nelle sue ricerche empiriche degli anni sessanta. *Hereditary Genius: An Inquiry into its Laws and Consequences* (1869) è una pionieristica applicazione degli strumenti statistici al problema dell'eredità.¹⁷³ Questo libro segna l'inizio di un percorso di ricerca che condusse allo sviluppo del concetto di regressione alla fine degli anni settanta e che culminò con quello di correlazione negli anni ottanta.

Gli interessi eugenetici sono stati più evidenti nella costruzione del concetto di regressione: le sue ricerche sul processo di trasmissione delle caratteristiche dei piselli (1877) e della statura all'interno delle famiglie umane (1885a; 1885b; 1886) sono riconducibili a

¹⁷¹MacKenzie punta giustamente sull'evidente esistenza di qualche rapporto tra lo sviluppo della scienza e l'ambiente sociale in cui esso avviene. Non sorprende che questa sia, pertanto, un'idea largamente condivisa dagli studiosi della scienza. Ciononostante, si verificano almeno due tendenze nel concepire questo rapporto. La prima è, secondo MacKenzie, più antica e influente, e considera molto limitati i modi in cui la società può indirizzare il percorso della scienza. Un esempio di questo atteggiamento nello studio sociologico della scienza è quello di Ben-David (1972). La seconda, invece, assegna un ruolo più deciso ai fattori sociali nello svolgersi della scienza. Una versione "debole" di questa seconda tendenza sostiene che la produzione di nuove idee è influenzata dalla società, ma poi sono giudicate con criteri oggettivi e generali propri dell'attività scientifica. Questa è, ad esempio, l'impostazione dell'epistemologo Klimovsky (1972) nel suo trattamento dei rapporti fra scienza e ideologia, in cui si evidenziano tracce popperiane e reichenbachiane — soprattutto per quanto riguarda la distinzione tra contesto della scoperta e contesto della giustificazione. La versione "forte", invece, sostiene che non solo la produzione di nuove idee, ma anche il processo attraverso il quale esse vengono accettate o respinte può essere condizionato dai fattori sociali. Quest'ultimo approccio, d'ispirazione kuhniana, è quello propriamente costruttivista, e lo studio sociologico di MacKenzie (1981) sullo sviluppo della statistica inglese nel periodo 1865-1930 è uno dei suoi importanti rappresentanti. Come visto sopra, per lui l'eugenetica non fu solo una fonte d'ispirazione; essa condizionò la direzione intrapresa nello sviluppo della teoria statistica. I tre artefici principali dell'emergere e del consolidarsi della moderna teoria statistica — a suo avviso Galton, Pearson e Fisher — condividevano forti interessi eugenetici e di classe. Tutti e tre erano impegnati nella promozione della nascente classe media professionale. Peraltro, anche il contributo più importante di Fisher — l'analisi della varianza — fu ideato inizialmente per una ricerca sull'eredità (Porter 1986).

¹⁷²A proposito dell'osservazione dei pari — e della già riferita prevalenza dell'eredità sulla socializzazione — non sarebbe giusto concludere che Galton non riconosceva per niente il ruolo della seconda. In *Psychometric Experiments* (1879, 158) esamina marginalmente la questione delle "idee prevalenti nello strato della società inglese" in cui nacque ed era cresciuto, e racconta un episodio circa l'effetto dell'educazione: "Mi trovavo in una casa di campagna con una compagnia molto piacevole di giovani e vecchi [...] la cui educazione e versatilità non era certamente sotto la media sociale. Una sera, giocammo un gioco che consisteva in che ognuno disegnasse uno scarabocchio tanto assurdo quanto fosse possibile rappresentando un fatto storico; i disegni furono dopo mescolati e passati successivamente da mano in mano, ognuno scrivendo indipendentemente la sua interpretazione del disegno per quanto riguardava l'evento storico che l'artista aveva inteso raffigurare. Fui stupito dall'identità delle nostre idee [...] e dal grande effetto della prima educazione nel fissarle".

¹⁷³Nella prefazione del libro Galton rivendica infatti il merito di essere stato il primo a indagare sull'eredità con tecniche statistiche.

questa concezione. La correlazione, invece, emerse piuttosto dall'attenzione dedicata al problema dell'identificazione personale, una questione che allora interessava sia gli statistici sociali sia i funzionari pubblici. L'idea era quella di poter individuare "le persone con propensioni criminali sulla base delle loro caratteristiche anatomiche" (Di Franco e Marradi 2003, 25). Seguendo questa impostazione generale Galton (1890) decise di indagare il modo di stimare la statura d'un uomo sconosciuto dalla lunghezza delle sue ossa e dalla relazione tra le varie dimensioni corporali. Tuttavia, anche se l'eugenetica non ebbe una presenza diretta nel suo lavoro sulla correlazione, i suoi interessi per l'identificazione personale furono parzialmente ispirati da essa (MacKenzie 1981).

Come visto nel capitolo precedente, le ricerche di Galton culminarono in *Natural Inheritance* (1889). Introducendo quest'opera egli ricorda:

Per molto tempo sono stato coinvolto nella ricerca dei problemi che giacciono alla base della scienza dell'eredità, e per vari anni ho pubblicato articoli tecnici su essi [...]. Questo volume contiene i risultati più importanti, ordinati e più completi di quello che finora era stato possibile [...]

L'indagine si riferisce all'eredità di qualità moderatamente eccezionali [...] ed è stata eseguita con metodi più raffinati di quelli abitualmente usati (ivi, 1).

L'impostazione evoluzionista di Galton in biologia¹⁷⁴ fu continuata da Weldon. Egli si occupò di problemi analoghi ma da un punto di vista morfologico. Le sue applicazioni della correlazione dimostrarono la linearità delle relazioni tra i caratteri fisici di granchi e di gamberetti, e suggerirono l'equivalenza dei coefficienti di correlazione tra gli organi di diverse specie locali. Le sue ricerche dell'inizio degli anni novanta, sebbene inquadrare nell'ambito della zoologia, fornirono i fondamenti di una nuova disciplina: la biometria. Essa comportava l'uso di metodi quantitativi per ricerche biologiche di impostazione evoluzionista, e comprendeva lo studio di piante, animali ed esseri umani. Gli studi su animali e piante — come aveva già proposto Galton nella ricerca sui piselli — spesso erano un mezzo per capire i processi di eredità in sé, e le loro conclusioni si generalizzavano agli esseri umani. Per Weldon "il problema dell'evoluzione animale era statistico" e la sua conoscenza era "l'unica base legittima per le speculazioni sulla storia e le predizioni del futuro" (1893, 329).

Ciononostante, le sue conoscenze matematiche non erano profonde e le ricerche si limitavano a un uso piuttosto elementare delle tecniche di Galton. L'affinamento degli strumenti fu la conseguenza della collaborazione con Pearson. Weldon e Pearson condividevano l'ambito di lavoro — l'Università di Londra — la profonda ammirazione per Galton e l'impostazione darwinista. I primi contributi di Pearson alla teoria statistica — dopo aver visitato altri campi di studio — emersero dal trattamento dei problemi proposti da Weldon. Essi sono strettamente collegati al tentativo di illustrare in modo matematico la biologia evolutiva. Questa matematizzazione del darwinismo si inquadra nella sua battaglia antimetafisica.

I più celebri contributi di Pearson allo sviluppo degli strumenti della correlazione si trovano infatti nella lunga serie di articoli intitolata *Contributions to the Mathematical Theory of Evolution*.¹⁷⁵ Soprattutto quello del 1896 — *Correlation, Heredity and Panmixia* — in cui presentò la formula prodotto-momento trattando il problema della selezione naturale, la selezione riproduttiva e l'endogamia.¹⁷⁶ Fino a 1901, quando i suoi interessi di ricerca

¹⁷⁴ Anche Dewey definiva biologiche le ricerche di Galton. Si veda la sua recensione di *Natural Inheritance* (1889)

¹⁷⁵ Egon Pearson (1938) afferma appunto che tutti i contributi statistici del padre devono essere giudicati come un insieme di strumenti applicabili alla soluzione dei problemi dell'eredità e dell'evoluzione.

¹⁷⁶ Pearson scrisse in quell'articolo: "La ricerca sulla correlazione che si presenterà [...] cerca di raggiungere le fondamentali formule necessarie [...] con speciale attenzione per quello che sembra legittimo nel caso dell'eredità" (1896, 261).

cominciarono a diversificarsi, egli lavorò quasi esclusivamente nel campo dell'eredità e dell'evoluzione.

Anche le ricerche in cui considerò la regressione e la correlazione multipla, la correlazione spuria e l'associazione tra proprietà categoriali appartenevano a questo ampio campo della biologia evoluzionista.

In effetti, al fine di costruire un modello che permettesse di determinare l'influenza degli antenati — oltre a quella più immediata dei genitori — ideò un'espressione per la distribuzione normale congiunta di n variabili, la superficie di correlazione multivariata normale:

Il passo molto rilevante nello sviluppo della teoria della correlazione multivariata normale emerse dalla preoccupazione per un problema biologico (Norton 1978, 12).

Analogamente, la necessità di tecniche per rilevare il grado di associazione emerse per lui dal desiderio "di misurare l'intensità delle correlazioni fraterne e ancestrali di caratteri come il colore degli occhi, dei capelli e della pelle" (E. Pearson 1938/1948, 59). Peraltro, come afferma MacKenzie (1981), questi bisogni specifici delle indagini sull'eredità condizionarono la natura stessa della sua teoria dell'associazione. Egli cercava di stabilire un nesso tra le caratteristiche fisiche e mentali, che avrebbe fornito un fondamento a interventi di politica eugenetica.¹⁷⁷

Le memorie di Pearson non si limitavano a riportare i risultati delle ricerche biometriche; ancora più importanti erano i già menzionati contributi alla metodologia statistica. Secondo E. Pearson (1938) essi si trovano — per il periodo 1893-1901 — in circa trentacinque saggi. La sua originale combinazione di ricerca empirica con soluzioni metodologiche di natura statistico-matematica disorientava non pochi contemporanei. I colleghi della *Royal Society* spesso non sapevano come catalogare i suoi articoli: si trattava di contributi alla biologia o alla matematica?

Tuttavia — a parte i notevoli contributi individuali di Pearson — l'aspetto più importante della sua collaborazione con Weldon fu probabilmente la fondazione e l'istituzionalizzazione della Scuola Biometrica. All'inizio del novecento essa già contava su un laboratorio e una rivista. Per molti anni fu la più importante fonte di didattica e ricerca statistica. I suoi membri lavoravano tanto nell'applicazione dei nuovi strumenti quanto al loro sviluppo e formalizzazione. Alcuni dei suoi studenti avrebbero poi avuto il compito di diffondere i metodi in altri ambiti istituzionali e di applicarli ad altri problemi scientifici, aggiungendo nuovi elementi alla teoria della correlazione. Le loro pubblicazioni hanno influenzato notevolmente il pensiero di biologi, psicologi e sociologi (Norton 1978).

Dopo la morte di Weldon nel 1906, Pearson si orientò sempre più verso la ricerca eugenetica. Il suo ruolo come direttore del Laboratorio Eugenetico e come professore della cattedra Galton di Eugenetica Nazionale si manifestò in una particolare combinazione di ricerca biometrica ed eugenetica, di taglio statistico, orientata alla riforma sociale: "il Laboratorio Biometrico sviluppò metodi statistici in un contesto biologico, ed il Laboratorio Eugenetico applicò questi lavori per dimostrare la prevalenza di *nature* su *nurture* negli affari umani" — in altre parole, la prevalenza dell'eredità sull'educazione e la socializzazione (Norton 1978, 4).

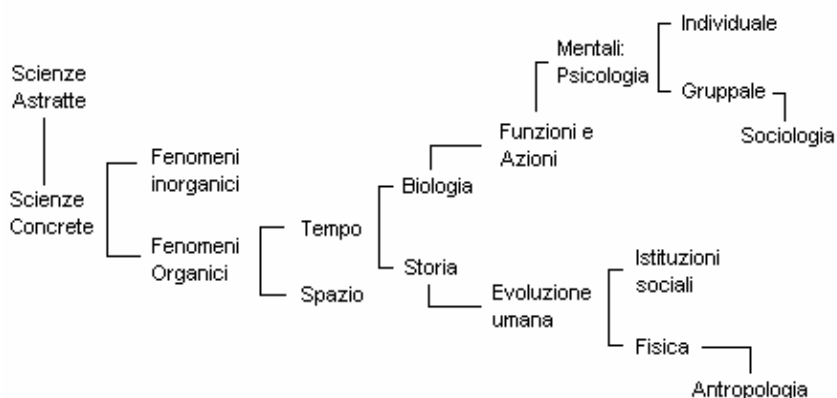
La creazione del Dipartimento di Statistica Applicata come frutto della fusione degli istituti di biometria e di eugenetica dell'*University College* segnò il destino della ricerca statistica di Pearson e dei suoi collaboratori nella direzione appena menzionata. In questo periodo le sue più importanti iniziative scientifiche furono motivate dalla passione per le

¹⁷⁷ Questo è il fondamento della conclusione di MacKenzie (1976) circa i differenti approcci di Pearson e di Yule al problema dell'associazione. Essi avrebbero risposto a diversi interessi sociali e ad atteggiamenti contrari nei confronti dell'eugenetica.

riforme sociali, basate su un programma eugenetico di miglioramento della razza inglese (Levine 1996) in preparazione per la concorrenza internazionale.

Si trattava di una particolare combinazione di biometria, epidemiologia, psicologia, antropologia fisica, sociologia e ricerca sulle istituzioni sociali. Gli studi sulla decadenza nazionale; sulle particolarità della razza inglese; sull'eredità degli attributi mentali e morali; sulle malattie infettive e le loro conseguenze; sull'alcoolismo e i suoi effetti fisici e psicologici nella prole; sui difetti mentali; sul rapporto tra fertilità e valore sociale; sul problema dell'immigrazione in Gran Bretagna; sul rapporto tra corpo e mente; sull'eredità delle malattie mentali sono esempi delle ricerche in cui si applicarono i nuovi strumenti della regressione e della correlazione nel dipartimento di Pearson.

Peraltro, quest'amalgama di temi così differenti in un unico programma di ricerca non era altro che la coerente realizzazione della classificazione pearsoniana delle scienze: si trattava in definitiva dello studio delle "forme viventi". In effetti, egli riteneva che la sociologia e psicologia si occupavano di azioni e funzioni con una base biologica; l'antropologia e le istituzioni sociali erano manifestazioni dell'evoluzione umana. Infine, la biologia e l'evoluzione umana erano due aspetti dei fenomeni organici studiati dalle scienze concrete, che si distinguevano da altri per la loro dimensione temporale. Ma la biologia faceva riferimento a fenomeni ricorrenti e l'evoluzione umana a fenomeni non ricorrenti, e pertanto era considerata un ramo della storia. Al fine di chiarire questa complessa e per niente ortodossa classificazione delle scienze presento qui una sua parziale ricostruzione schematica¹⁷⁸:



Senza dubbio i contributi statistici della scuola biometrica, soprattutto in tema di correlazione, hanno avuto un successo duraturo, pervadendo tutte le discipline scientifiche. Non si può affermare altrettanto nei confronti dei loro contributi alla biologia e alle scienze sociali. L'impostazione eugenetica che informava le loro ricerche era già contestata nel suo momento culminante, e dopo la Prima Guerra Mondiale — almeno in Inghilterra — divenne sempre più impopolare. Le conclusioni biologiche proposte dalla biometria pearsoniana erano considerate da molti colleghi — in particolare da quelli influenzati da Mendel — come matematicamente raffinate, ma biologicamente ingenuie. Il tentativo dei biometri di avanzare una concezione fenomenica e quantitativa dell'eredità, senza mediazioni teoriche, generò una vivace disputa con questa altra grande scuola biologica del tempo.¹⁷⁹

¹⁷⁸ Si veda E. Pearson (1938/1948, 209).

¹⁷⁹ Nel 1900 si era riscoperto il lavoro di Mendel sui meccanismi dell'eredità individuale. Bateson, uno dei più noti biologi degli anni a cavallo tra l'ottocento e il novecento, si interessò a questa prospettiva e fondò la scuola mendelista inglese.

Yule fece nel 1902 un tentativo di riconciliazione fra le due tradizioni antagoniste.¹⁸⁰ Egli era stato educato nel Laboratorio di Pearson, ma allo stesso tempo riteneva essenzialmente corretta la teoria di Mendel.¹⁸¹ Pertanto, come afferma Morrison (2002, 49), era in una posizione ideale per tentare di avvicinare gli approcci, o per "delimitare le [loro] sfere di applicazione [...] mostrando in che modo essi potevano coesistere". Il suo suggerimento indicava che il mendelismo poteva essere compatibile con l'idea darwinista di evoluzione continua.

Pearson (1904) riprese il suggerimento di Yule ma in generale rimase poco propenso ad accettare il mendelismo. Per lui, fondare la spiegazione dell'eredità sulla nozione del "fattore di Mendel"¹⁸² implicava imbarcarsi in speculazioni metafisiche circa non controllabili agenti causali.¹⁸³

Peraltro questa disputa fu solo una delle tante che segnarono la sua vita scientifica. In effetti, le conclusioni delle sue ricerche sociali non erano sempre ben accolte e spesso generavano aspre polemiche. L'esempio più noto è quello dell'influenza dell'alcoolismo paterno sulle capacità fisiche e mentali dei figli (si veda ad esempio Elderton 1910; Pearson e Elderton 1910). Alla tesi di Pearson e Elderton risposero polemicamente medici, scienziati sociali ed economisti — in particolare Marshall e Keynes (Stigler 1999).

Tuttavia, le tecniche sviluppate nella scuola biometrica guadagnavano gradualmente una certa autonomia relativa: esse non avevano limitazioni intrinseche tali da non poter essere applicate a indagini diverse da quelle volute da Pearson e dai suoi collaboratori. In un rapporto del 1930 alla Company of Drapers, *sponsor* di tante ricerche biometriche ed eugenetiche, Pearson stesso riconosceva che il suo obiettivo era stato costruire un laboratorio unico in cui si potesse applicare un nuovo calcolo ai problemi degli esseri viventi, e che questo proposito aveva comportato lo sviluppo di nuove forme di analisi matematica. Ma esse erano cresciute moltissimo attraverso il lavoro dei suoi discepoli sparsi nel mondo e attraverso coloro che studiavano i suoi scritti e che applicavano le sue idee nella ricerca economica, sociale e psicologica.¹⁸⁴

4. 3. Correlazione e associazione nella ricerca economica e sociale.

Come visto nel cap. 1, l'uso di metodi di analisi quantitativa nella ricerca sociale non comincia ovviamente con i moderni strumenti della correlazione. In effetti, come afferma Lazarsfeld (1961), l'aritmetica politica del seicento è stata una fase preparatoria in questo indirizzo, e il lavoro di Quetelet e LePlay inaugurò un secondo periodo in cui gli strumenti matematici della probabilità cominciarono ad essere applicati nella ricerca sociale empirica.

¹⁸⁰ Secondo E. Pearson (1938) le tesi biometrica e mendelista non erano in sostanza incompatibili. I loro percorsi di ricerca erano piuttosto complementari: oggi si riconosce grande valore alla combinazione di genetica, biologia e statistica. Tuttavia, all'inizio del novecento la situazione era diversa.

¹⁸¹ In una lettera scritta a Kendall Yule diceva di essere convinto che la biologia doveva molto più al mendelismo che alla biometria (si veda Kendall 1952).

¹⁸² Si tratta di un antenato del moderno concetto di gene.

¹⁸³ Secondo Morrison (2002) questa controversia sarebbe stata in sostanza risolta da Fisher attorno al 1920. Egli dimostrò infatti che l'incompatibilità tra il mendelismo e la biometria darwinista, non solo al livello metodologico — applicabilità delle tecniche statistiche sviluppate dalla scuola biometrica alla ricerca dell'eredità del mendelismo — ma anche nella sostanza, era solo apparente.

¹⁸⁴ Come già indicato, Pearson stesso applicò le tecniche della correlazione alla ricerca sociale e psicologica. Tuttavia, si trattava di ricerche con scarsa specificità disciplinare, molto legate alla sua particolare concezione dello status di queste scienze. Per lui esse dovevano ancorarsi su dati biometrici e antropometrici, e il più delle volte con un'ispirazione di intervento eugenetico.

A partire dalla seconda metà dell'ottocento questa impostazione trovò in Gran Bretagna dei ferventi sostenitori. Prima Jevons e poi Edgeworth si occuparono di giustificare l'applicabilità della matematica alle scienze sociali, e in particolare all'economia. Per Edgeworth la matematica era la chiave per far progredire la scienza economica.¹⁸⁵ Non dovrebbe sorprendere pertanto che lui abbia favorito l'uso della correlazione nella ricerca sociale — come del resto auspicava Galton sin dall'inizio.

Già nel 1893 — tre anni prima che Pearson proclamasse l'applicabilità dell'analisi delle correlazioni nel campo biologico, e contemporaneamente alle sue prime applicazioni effettive da parte di Weldon — Edgeworth scrisse sulla sua adeguatezza per la trattazione dei fenomeni sociali. A suo avviso la distribuzione congiunta degli errori permetteva studiare i problemi che dipendevano da una varietà di elementi o agenti indipendenti (1893, 672).

Edgeworth — coem appena segnalato — cominciò a promuovere l'applicazione della correlazione nel campo delle scienze sociali quando Weldon faceva i primi passi nella ricerca delle relazioni tra organi di animali, e persino prima che Pearson si occupasse del tema.

La differenza fondamentale, come suggerisce Kendall (1968) è che Edgeworth non riuscì a sollevare per l'applicazione della statistica ai problemi economici e sociali lo stesso interesse che Weldon e Pearson generarono in biologia. Questi ultimi fondarono la loro impresa su una solida base istituzionale: aprirono la strada a una scuola e a un nuovo ramo della biologia nato proprio dall'applicazione di regressione e correlazione. Edgeworth — come visto in sez. 1. 4 — non apparteneva ad alcuna scuola; non erano nel suo spirito il lavoro di gruppo e l'imprenditorialità accademica che sarebbero state utili a stabilire una scuola istituzionalizzata dedicata alla ricerca sociale con l'uso degli strumenti appena menzionati. Di conseguenza, la loro applicazione alle scienze sociali fu per molto tempo più dispersa e meno articolata.

La ricerca sociale empirica, piuttosto indebolita dopo il periodo d'oro del movimento statistico del primo periodo vittoriano, aveva trovato un nuovo impulso alla fine dell'ottocento. La relativa stabilità e benessere che in Gran Bretagna caratterizzò il periodo in cui avvenne il ristagno del movimento statistico — e la conseguente diminuzione dell'interesse per la ricerca e per le riforme sociali — stava lasciando il campo a un nuovo periodo di agitazione sociale. La condizione umana e sanitaria del proletariato urbano era molto preoccupante. Un nuovo periodo di fortuna della ricerca sociale empirica era prevedibile.

La *Royal Statistical Society*, erede dell'Associazione londinese che aveva istituzionalizzato il movimento statistico, sarebbe stata una delle tribune privilegiate per la discussione dei problemi sociali: condizioni di vita della classe operaia, problemi sanitari ed epidemiologici, mortalità infantile erano alcune delle questioni sul tappeto. Tuttavia, in questo periodo il ruolo dell'associazione non era del tutto chiaro. Come visto nel cap. 1, essa si manteneva ancora fedele a una tradizione statistica non impegnata nello sviluppo di nuove tecniche di analisi. La sua conversione in un'associazione statistica nel senso moderno del termine non avvenne fino ai primi decenni del novecento (MacKenzie 1981). Da altra parte, in quanto scienza empirica della società e dello Stato — come si intendeva quando l'associazione era stata fondata — la statistica stava perdendo senso di fronte al graduale consolidamento di discipline più specifiche quali la sociologia, l'economia politica e la scienza politica. È evidente che in questa fase di transizione un'associazione piuttosto conservatrice — e lenta ad adeguarsi al mutevole panorama intellettuale — non era in grado di fornire una base istituzionale solida per lo sviluppo di una ricerca sociale decisamente orientata all'uso delle nuove tecniche di analisi.¹⁸⁶

¹⁸⁵ Galton e Pearson difendevano una posizione analoga: anche per loro la matematizzazione avrebbe garantito la scientificità delle scienze sociali. Solo che i loro interessi di ricerca si trovavano sostanzialmente in altri campi.

¹⁸⁶ Intendo dire una base istituzionale equivalente a quella assicurata alla biometria dall'imponente edificio accademico costruito da Pearson: laboratori, biblioteche, posti di lavoro fissi, pubblicazioni, e soprattutto una direzione coerente e impegnata all'applicazione e allo sviluppo dei nuovi strumenti.

Ciononostante, all'associazione aderivano anche accademici interessati a far progredire i nuovi strumenti e le loro applicazioni alla ricerca economica e sociale. Si trattava in particolare di Edgeworth, Yule, Bowley e Hooker. Nel campo della correlazione il contributo di Edgeworth e di Bowley si può giudicare principalmente teorico, e quello di Hooker eminentemente applicativo. Yule invece ebbe un profilo in qualche senso paragonabile a quello di Pearson: i suoi contributi teorici furono accompagnati da molte ricerche empiriche (Kendall 1952).

Quando cominciò a lavorare alla teoria statistica Yule aveva avuto già qualche esperienza di ricerca empirica lavorando in Germania nel campo della fisica. Poi, sotto la supervisione di Pearson, eseguì indagini di tipo biologico dando dei contributi di rilievo — si ricordi ad esempio il suo tentativo di armonizzare il mendelismo e la biometria presentato nella sezione precedente. Tuttavia, come detto in altri capitoli, lui si sentiva un erede della tradizione di Jevons e di Edgeworth, e pertanto i suoi interessi si indirizzarono presto ai problemi economici e sociali. Molti dei suoi contributi teorici alla correlazione e all'associazione — come si vedrà più avanti — si fondarono su ricerche empiriche in quei campi.

Yule era ancora un collaboratore nella cattedra di Pearson quando applicò per primo la correlazione a un caso d'interesse economico e sociologico. In effetti, egli pubblicò nell'*Economic Journal* — allora editato da Edgeworth — due articoli successivi intitolati *On the Correlation of Total Pauperism with Proportion of Out-Relief* (1895; 1896). Si trattava di una ricerca basata sulle idee e i dati forniti da Charles Booth nella sua celebre inchiesta sulla condizione di vita del popolo londinese (1891). In questi lavori Yule tentò di stabilire la relazione tra povertà e *out-relief* (spese assistenziali). Come avrebbe dichiarato lui stesso quasi quindici anni dopo, queste ricerche fecero emergere un interesse per il problema della povertà, che fu trattato in modo più approfondito in un articolo del 1899 sulle cause dell'evoluzione della povertà nell'Inghilterra di fine Ottocento. La sua conclusione principale fu che i cambiamenti nel totale di poveri erano correlati con le variazioni nelle spese assistenziali, ma non con le variazioni della popolazione totale o della proporzione di anziani.

Nel 1909 uscì nel "Journal of the Royal Statistical Society" un importante saggio di Yule: *The Applications of the Method of Correlation to Social and Economic Statistics*. L'articolo comincia con una diagnosi della diffusione del metodo:

Il metodo della correlazione ha trovato molte applicazioni nello studio della biologia in tempi recenti, probabilmente più che in qualsiasi altra disciplina in cui i metodi della statistica sono utili, ma le applicazioni ai problemi che interessano specialmente agli studiosi di statistica economica e sociale sono state, a mio avviso, relativamente scarse (Yule 1909, 721).

Da questo punto di partenza Yule proseguì a perlustrare la letteratura in cui si riportavano ricerche su problemi sociali e economici che avevano applicato il concetto e gli strumenti della correlazione. Si tratta prevalentemente di studi compiuti in Inghilterra e di qualche lavoro straniero che aveva guadagnato notorietà negli ambienti accademici inglesi. A suo avviso occorre elencare ed organizzare queste ricerche disperse per comunicare più efficacemente la rilevanza della correlazione in questi campi, indirizzando l'attenzione degli scienziati sociali verso la tecnica appena menzionata e suggerendo sue ulteriori applicazioni.

In ordine cronologico, il secondo esempio dopo le ricerche di Yule sulla povertà fu uno studio di Bowley — pubblicato nel suo *Elements of Statistics* (1901) — in cui trattò la relazione tra il tasso di matrimoni ed il prezzo del grano in Inghilterra e Galles. Nel 1905 March ricercò varie relazioni tra le statistiche vitali, e nel 1906 Heron¹⁸⁷ studiò la relazione tra fertilità e status sociale in diversi distretti di Londra.

Secondo Yule (1909) era fondamentale applicare il metodo della correlazione ai dati sulla mortalità, con il proposito di illustrare l'influenza di certe proprietà misurabili sul

¹⁸⁷ Si ricordi che Heron era uno stretto collaboratore di Pearson nella Scuola Biometrica.

fenomeno. A suo avviso queste ricerche avrebbero potuto essere la base della politica sociale nel campo della sanità; esse tuttavia erano state scarse fino ai primi anni del novecento. Un'eccezione la costituivano le ricerche in cui Newsholme (1906; 1908) utilizzò il coefficiente di correlazione per accertare le cause della diminuzione del tasso di mortalità per tisi, con particolare riferimento ai pazienti ospedalizzati. Un'area d'interesse in qualche modo collegata alla precedente e non ancora sviluppata era il legame tra il tasso di lavoro femminile — tra donne sposate — e la mortalità infantile (Yule 1909).

Un altro campo di applicazione si trovava nella ricerca economica e finanziaria. Per esempio Hooker (1901a) aveva studiato la correlazione tra il tasso di matrimoni e il volume dei commerci. Poco dopo indagò sulle relazioni del prezzo del mais nei mercati di Berlino, Liverpool e Chicago (1901b). Il suo obiettivo era determinare quanto fosse internazionalizzato il mercato berlinese, ma i risultati furono piuttosto deludenti. Infine pubblicò una ricerca sull'influenza del clima sulla produzione agricola (1907).

L'applicazione della regressione e della correlazione al settore finanziario si sviluppò presto in America. Norton (1902) applicò il metodo nelle sue ricerche sul mercato del denaro di New York. Egli si interessò, per esempio, al rapporto tra il volume di depositi nelle banche e il tasso di sconto. A parte la sua utilità pratica, Yule (1909) valorizzò in particolare questa ricerca dato che introduceva regressioni non lineari vari anni prima che Pearson si occupasse del tema da un punto di vista teorico (si veda sez. 3. 3). Un altro importante uso della correlazione nelle finanze si trova nel lavoro di March citato sopra.

Alice Lee — una fedele collaboratrice di Pearson — studiò la relazione tra il tasso d'impiego e l'entità delle importazioni. Sfidando le tesi allora dominanti ella concluse che l'aumento delle importazioni non esercitava un'influenza costringente nel mercato di lavoro; il suo effetto era anzi opposto. Tuttavia, Yule (1909) manifestò qualche dubbio sui risultati.

Finora si sono presentati alcuni articoli che riferivano le prime applicazioni della correlazione in dei campi determinati. Ma le scienze sociali non sono state solo un'area di uso acritico degli strumenti menzionati.

Yule considerò ad esempio (1897) alcune proprietà rilevanti per le scienze economiche e la sociologia — le cui distribuzioni non seguivano un andamento simmetrico — per illustrare il problema delle "correlazioni non normali" e per evidenziare gli assunti sottesi alla teoria proposta fino a quel momento. Anche il concetto di correlazione parziale, originariamente denominato correlazione netta da Yule, fu giustificato in un modo analogo.

Tuttavia, il contributo teorico e tecnico più importante tra quelli compiuti a partire della trattazione di questioni sociali ed economiche fu l'associazione. MacKenzie (1981) afferma che Yule probabilmente capì la necessità di disporre di strumenti per determinare i rapporti tra proprietà non misurabili quando si dedicò a studiare un tema di rilevanza sociale: la vaccinazione dei minorenni. Allora molti personaggi influenti erano contrari alla vaccinazione. Per dimostrare la sua importanza da un punto di vista sociale, economico e sanitario, Yule sviluppò strumenti che permettessero di quantificare la relazione tra la vaccinazione e la guarigione (Yule 1900; 1903).

Anche il problema dell'associazione multipla, il concetto di associazione parziale e quello di associazione spuria emersero dalla considerazione di problemi empirici delle scienze sociali.

Tuttavia, come detto all'inizio di questa sezione, le applicazioni della correlazione nel campo socio-economico non godettero di un livello di articolazione o di una base istituzionale equivalente a quella della biometria. In generale, queste prime ricerche si limitarono a presentare i risultati di qualche correlazione empirica tra proprietà ritenute allora rilevanti. Un uso orientato alla costruzione di teorie sociologiche fondate empiricamente non si vedrà fino a qualche decennio dopo, nelle scuole più istituzionalizzate della sociologia empirica americana. Peraltro, Lazarsfeld affermò (1961), a partire delle sue indagini sulla storia della quantificazione in sociologia, che il primo sociologo ad applicare esplicitamente i coefficienti

di correlazione, e a dimostrare in modo competente la loro utilità nella ricerca strettamente sociologica, fu l'italiano Alfredo Niceforo¹⁸⁸ attorno al 1920.

4. 4. La psicologia correlazionale.

Stephen Stigler sostiene (1999) che la psicologia fu la prima disciplina umanistica in cui si verificò una rapida assimilazione delle tecniche statistiche delle scienze fisiche, grazie anche all'opera di Gustav Fechner, *Elemente der Psychophysik* (1860). Ma la differenza con altre scienze umane — per quanto riguarda la tendenza alla quantificazione e all'uso di questi strumenti — non si trova in realtà nella data di nascita. Come visto nel cap. 1, in questo periodo il ricorso agli strumenti della probabilità — in particolare alla legge degli errori — aveva qualche sostenitore anche in altre scienze umane. L'affermazione di Stigler si riferisce però al consolidamento e all'accettazione generalizzata.

In effetti, il successo di quest'approccio in psicologia era già assicurato verso gli anni ottanta, circa quarant'anni prima che in sociologia.¹⁸⁹ Ciò è dovuto al rapido riconoscimento della possibilità di condurre esperimenti e di manipolare le situazioni sotto studio. Gli esperimenti garantivano una base solida per la misurazione e il controllo dei fattori ritenuti importanti, e i loro risultati potevano essere impiegati come base della inferenza di leggi generali di ispirazione newtoniana. Verso il 1880 questo era una pratica frequente in Europa e America, per esempio nel lavoro di Ebbinghaus (1885)¹⁹⁰ e di Peirce e Jastrow (1885) rispettivamente (si veda Stigler 1999, cap. 10).

Inoltre, la psicologia empirica cominciava a trovare sbocchi istituzionali. Alla fine dell'ottocento c'era già un importante laboratorio in Germania fondato da Wilhelm Wundt. Molti psicologi americani e europei facevano il loro apprendistato con Wundt (Di Franco e Marradi 2003, cap. 2) e questo favoriva la comunicazione e gli scambi nella nascente comunità di psicologi sperimentali, che poteva anche profittare delle varie riviste specializzate pubblicate sulle due sponde dell'Atlantico.

Nella psicologia c'erano quindi tutte le condizioni per accogliere favorevolmente i nuovi strumenti della correlazione sin dalla loro nascita, e la loro applicazione assunse sin dall'inizio caratteristiche diverse che nelle altre scienze umane. Nelle sezioni precedenti si è riferito che in economia e ricerca sociale l'uso della correlazione fu per qualche tempo piuttosto disperso e occasionale: si mostravano relazioni tra fenomeni d'interesse in modo isolato, ma non si facevano dei tentativi di usarla complessivamente per fornire un ancoraggio empirico a teorie sulla società, o su specifici problemi sociali. In biologia, invece, essa era stata applicata in un modo più sistematico, come strumento privilegiato a sostegno di una concezione evoluzionistica. In psicologia la situazione delle prime applicazioni è stata più simile a quella della biologia. Soprattutto nel lavoro di Spearman, a partire dei primi anni del novecento, essa sarebbe stata la base strumentale per la costruzione di una teoria generale dell'intelligenza. Non a caso egli propose lo sviluppo di una "psicologia correlazionale" (*Correlational Psychology*) fondata nelle tendenze psichiche "che collegano insieme i cosiddetti 'test mentali' con le attività psichiche di una maggior generalità e interesse" (1904b, 205).

Tuttavia, l'applicazione della correlazione in psicologia non cominciò con Spearman. Galton stesso, mentre sviluppava i suoi concetti di regressione e correlazione, si era occupato di rilevare che una larga classe di fenomeni mentali poteva essere sottoposta a indagine

¹⁸⁸ Niceforo finì la sua carriera come professore emerito dell'università "La Sapienza" di Roma.

¹⁸⁹ Lo calcola Stigler (1999) in base alla storia della quantificazione in sociologia proposta da Lazarsfeld (1961).

¹⁹⁰ Questo lavoro è uno dei primi tentativi di applicare gli strumenti della probabilità in una ricerca puramente psicologica, non psico-fisica — quale invece si svolgeva nei laboratori di Wundt, Fechner e Weber.

statistica. In *Psychometric Experiments* (1879) sostenne che tutte le operazioni coscienti potevano essere registrate e trattate statisticamente. Il suo campo di applicazione preferito era la psicometria, "l'arte d'imporre la misurazione e i numeri alle operazioni della mente" (Galton 1879, 149). Nel 1883 — qualche anno prima della comunicazione formale del concetto di correlazione con la tecnica relativa — egli aveva suggerito la sua applicabilità nella ricerca psicologica quando riportò di avere trovato che i soggetti cui venivano attribuite migliori capacità in generale avevano tendenzialmente anche capacità più raffinate di discriminare minute differenze di peso. Tuttavia Galton non approfondì le ricerche di questo tipo e si orientò verso altri interessi. Il suo suggerimento di applicare la correlazione in psicometria tramite lo studio delle relazioni dei vari test mentali fu seguito da altri ricercatori, prima in Inghilterra e poi più estesamente in America. Alla fine dell'ottocento, si era diffuso anche nell'Europa Continentale, specialmente in Francia e Germania.

Non dovrebbe sorprendere pertanto che il celebre *Dictionary of Philosophy and Psychology* di Baldwin — uscito nel 1901 e considerato un classico — prevedesse la voce 'correlazione'. Uno dei significati a esso attribuiti faceva riferimento a una speciale legge che consente di asserire che una relazione riscontrata tra proprietà è sufficiente per presumerle associate anche nell'esperienza futura.

Nel 1904, in questo ambiente di rapida maturazione di una psicologia sperimentale impegnata all'uso dei metodi quantitativi, Spearman pubblicò la più famosa ricerca pionieristica basata sulla correlazione. Egli cominciò il suo lavoro con una perlustrazione delle ricerche psicometriche precedenti in cui si era utilizzata la correlazione — una sorta di *literature review* o stato dell'arte (1904b).

A suo avviso Oehrn, nel 1889, fu il primo a eseguire dei “veri” esperimenti di correlazione mentale; impiegò i test e poi calcolò la correlazione tra i suoi risultati. Boas nel 1891 cercò le correlazioni tra vista, udito e memoria di un campione di scolaretti. Due anni dopo Gilbert applicò test mentali a 1200 bambini e mise in relazione i risultati con la loro "abilità generale", stabilita secondo l'apprezzamento degli insegnanti. Contemporaneamente Dresslar investigò una nuova proprietà chiamata "illusione naturale": un bambino sarebbe stato considerato più intelligente di altri se — a parità di altre caratteristiche — mostrava una miglior capacità di associazione delle idee di dimensione e di peso. Nel 1894 Griffing utilizzò la correlazione per esaminare l'ampiezza (*amplitude*) dell'attenzione, anziché la sua intensità come era più frequente fare.

Nella Francia di fine ottocento gli psicologi Binet e Henri già incoraggiavano lo studio sistematico delle relazioni tra i differenti processi psichici. Tuttavia, il pioniere della ricerca psicometrica correlazionale francese fu Bourdon, che nel 1895 usò il concetto nella sua indagine sulle capacità di riconoscimento, discriminazione e associazione. Un anno dopo Binet cominciò effettivamente a percorrere la strada che fino ad allora aveva solo promosso verbalmente. Infatti, condusse una ricerca sulle facoltà intellettuali la cui conclusione principale fu l'esistenza di cinque tipi di carattere umano: il descrittivo, l'osservatore, l'erudito, l'emotivo, l'idealista. Poi proseguì le sue indagini con la collaborazione di Vaschide.

In Germania Wagner studiò da un punto di vista correlazionale il problema della fatica degli scolaretti, e pubblicò i suoi risultati nel 1896. Questo indirizzo di ricerca fu perseguito ed esteso negli anni seguenti da Ebbinghaus e Wiersma.

In America i lavori pionieristici furono di Sharp e Titchener. Nel periodo 1897-1899 Seashore — professore preso l'Università dello Iowa — studiò la relazione tra l'intelligenza generale (determinata soggettivamente dagli insegnanti) e le impressioni di forma, colore e peso in 200 bambini tra i sei e i quindici anni. Carman indagò nel 1899 le relazioni tra l'intelligenza generale dei bambini — seguendo la tradizionale impostazione che faceva ricorso a una valutazione da parte degli insegnanti — e la loro sensibilità al dolore.¹⁹¹ Un anno dopo Kirkpatrick studiò le attività motorie semplici in un campione di circa 500 bambini;

¹⁹¹ Secondo Spearman (1904b) questa fu la ricerca psicometrica più importante di fine ottocento.

e nel 1903 Peirce condusse esperimenti con 32 bambini per determinare la sensazione soggettiva del tatto. Altre ricerche importanti del periodo furono quelle di Thorndike e Woodworth, che tentarono per primi di ricercare correlazioni diacroniche.

Questa analisi dei lavori precedenti compiuta da Spearman (1904b) si concluse con un giudizio piuttosto negativo. La situazione non era a suo avviso per niente incoraggiante: le conclusioni circa la correlazione tra i risultati dei test mentali e le stime indipendenti di tipo pratico — per esempio la valutazione dell'intelligenza dei bambini secondo i maestri — si contraddicevano. Queste discordanze incoraggiavano una tendenza — in particolare tra i ricercatori più capaci — a negare completamente l'utilità delle correlazioni, tendenza che sarebbe stata fatale per la psicologia sperimentale come disciplina scientifica.

Si trattava — ad avviso di Spearman — di ricerche di scarso valore scientifico. Si noti ad esempio che solo il lavoro di Wissler a Columbia soddisfaceva "il primo e fondamentale requisito della correlazione, cioè, un'espressione quantitativa precisa". Molti avevano compilato tabelle numeriche molto elaborate senza focalizzarsi su un "singolo risultato esatto". Inoltre, ancora con l'eccezione di Wissler, nessuno aveva calcolato l'errore probabile, e pertanto non c'era alcuna possibilità di "giudicare in quale misura i suoi risultati si dovevano alla coincidenza accidentale" (Spearman 1904b, 222).

Un ultimo aspetto criticato da Spearman era che nessuno dei ricercatori sopra menzionati aveva considerato un'importante fonte di fallacie inevitabilmente presente in qualsiasi ricerca empirica, vale a dire gli "errori di osservazione": "nell'eseguire un esperimento e calcolare le correlazioni, si deve tener conto che quest'ultima non rappresenta la relazione matematica tra due gruppi di oggetti reali, ma solo tra due gruppi di misure ottenute con un processo più o meno fallibile. I risultati effettivamente ottenuti in qualsiasi esperimento di laboratorio sono necessariamente perturbati da varie contingenze che niente hanno a che vedere con le capacità reali del soggetto" (ivi, 223).

Di fronte alle radicali divergenze tra le varie teorie dell'intelligenza del suo tempo, Spearman si propose di sviluppare una concezione che potesse conciliare le diverse ipotesi, risolvendo armonicamente la disputa tra le teorie moniste — che difendevano l'idea di una sola facoltà generale — le teorie gerarchiche — che sostenevano l'esistenza di un castello di facoltà a vari livelli di generalità — e le teorie pluraliste — che avanzavano l'ipotesi di molteplici attitudini indipendenti e fra loro non subordinate. La sua strategia fondamentale fu quella di condurre esperimenti psicometrici e di applicare l'analisi di correlazione ai risultati, per controllare se le varie facoltà rilevate con i test erano correlate — formando un unico fattore d'intelligenza generale — o se erano indipendenti (Bernstein 1955). Pertanto era fondamentale contare su "una precisa espressione quantitativa derivata imparzialmente dal complesso dei dati disponibili [...]" (che variasse tra 1 per la perfetta corrispondenza e 0 per la perfetta assenza di corrispondenza) (Spearman 1904b, 252).

Nelle sue ricerche le correlazioni erano calcolate seguendo la formula prodotto-momento di Pearson; solo che invece di usare i valori esatti di ogni individuo si usava la sua posizione ordinale. In questo modo Spearman sviluppò il concetto di correlazione di ranghi e il coefficiente ρ .

Il suo metodo di analisi comprendeva quattro fasi: (1) calcolare le correlazioni; (2) determinare il loro errore probabile (per decidere se i risultati delle correlazioni era affidabile); (3) esaminare se qualche fattore irrilevante era stato considerato, e nel caso introdurre equazioni per eliminarlo; (4) rivedere criticamente l'intera argomentazione per arrivare a una stima finale della corrispondenza reale — probabile — e dell'affidabilità attribuita agli indicatori.

I risultati di questa ricerca — che riteneva totalmente riproducibili — mostrarono correlazioni molto forti tra le capacità "discriminazione generale" e "intelligenza generale". La conclusione da esse derivate fu che "tutti i rami della attività intellettuale hanno in comune una funzione fondamentale (o un insieme di funzioni)" (ivi, 284). Una conseguenza pratica di questa unità della funzione intellettuale era che le varie forme dell'attività mentale

costituivano una gerarchia stabile. Secondo Spearman la funzione si sviluppava completamente nei bambini attorno ai nove anni di età e determinava i loro risultati scolastici. Essa non risultava legata ad altri caratteri quali volontà o livello di adattamento. In età adulta non si trovavano differenze significative tra maschi e femmine. Tuttavia, la natura fisica di questi processi non era stata ancora investigata.

È evidente che gli interessi sostanziali di Spearman divergevano da quelli di Wundt, e "si allineavano piuttosto con gli interessi dei positivisti del tempo: in particolare la natura e le manifestazioni dell'intelligenza". La capacità di discriminazione sensoriale — oggetto privilegiato nelle ricerche scientifiche di Wundt e dei psicofisici — era per lui solo una manifestazione dell'intelligenza. "Spearman era convinto che l'intelligenza fosse una proprietà unitaria, di origine ereditaria" (Di Franco e Marradi 2003, 27).

Per confermare questa tesi impostò il problema in modo di dimostrare la dipendenza di tutte le variabili da un unico fattore, per la quale "i determinanti di second'ordine di una matrice di correlazione (cioè i prodotti incrociati dei coefficienti di correlazione fra tutti i possibili quartetti di variabili) dovevano avvicinarsi a zero" (*ibidem*). In questo modo gettò le basi dell'analisi fattoriale¹⁹². La sua tecnica — conosciuta come *vanishing tetrads* — continuò ad essere usata fino agli anni trenta, quando fu sostituita dal metodo dei centroidi di Thurstone.

Spearman dedicò tutta la sua vita al problema dell'intelligenza. Nel 1927 pubblicò *The Abilities of Man: Their Nature and Measurement*, in cui presentò in modo articolato la sua teoria sull'intelligenza. Essa presupponeva che se due abilità sono correlate, devono essere entrambe dipendenti da un fattore comune che determina la loro correlazione — chiamato *g* — e da un fattore specifico che determina la differenza tra le due abilità — chiamato *s* (Bernstein 1955). Secondo Spearman (1927/1955, 357) il primo e più importante risultato raggiunto in tutti gli anni dedicati alla ricerca dell'intelligenza fu "la scoperta di *g*, una vera rivoluzione copernicana nel modo della psicologia" e "un passo in avanti nella conquista di una autentica base scientifica" per questa disciplina (ivi, 360).¹⁹³

¹⁹² Spearman non usa il termine 'fattore' fino a 1927, nel libro *The Abilities of Man: Their Nature and Measurement*.

¹⁹³ Tuttavia, la sua tesi di un unico fattore generale dell'intelligenza subì attacchi quasi immediatamente. Uno dei più tenaci oppositori fu Burt. Egli criticava la mancanza di considerazione per le funzioni mentali superiori (astrazione, induzione, etc) che a suo avviso erano più rilevanti delle capacità sensoriali e di reazione motoria che usava Spearman (si veda Di Franco e Marradi 2003, cap 2). La teoria fattoriale fu criticata anche da Thorndike e Thompson, che proposero nuove versioni della teoria pluralista; e da Thurstone — per cui non esistevano *g* e *s* ma vari fattori di gruppo.

CONCLUSIONI

Ci sono diversi modi di impostare le conclusioni di uno studio. Uno dei più frequenti consiste nel riassumere, in modo articolato, le tesi trattate nei singoli capitoli. Un altro meno diffuso consiste nel proporre una riflessione più generale a proposito dei temi considerati, aggiungendo un elemento nuovo — e comunque non estemporaneo — alla presentazione. Mi auguro che le tesi proposte nel caso di questo lavoro siano state trattate in modo sufficientemente approfondito nei capitoli, e che le conclusioni principali da loro tratte si trovino — piuttosto esplicitamente — anche in quelle stesse sedi. Pertanto preferisco percorrere la seconda strada, meno convenzionale ma forse più suggestiva. Essa consente di mettere in relazione le argomentazioni dell'indagine con rilevanti problemi classici della disciplina, in questo caso della metodologia delle scienze umane. Le brevi riflessioni che propongo di seguito sono appunto a proposito del problema del metodo e le tecniche di ricerca.

Marradi (1996) chiarisce questo problema facendo ricorso a un'immagine:

Ogni ricerca ha un obiettivo cognitivo: vuole cioè migliorare, approfondire, articolare la conoscenza intorno a un certo argomento [...] Possiamo immaginare questo obiettivo come una radura in una foresta: si confida che sia raggiungibile, si spera di raggiungerla, ma non si sa esattamente dov'è; tanto meno si hanno le idee chiare su come arrivarci [...] Possiamo immaginare che il ricercatore parta da un punto qualsiasi ai bordi della foresta. La foresta da attraversare è percorsa per tratti più o meno lunghi, e in varie direzioni, da sentieri già tracciati, più o meno battuti: sono le tecniche che altri ricercatori hanno già ideato, modificato, sviluppato. Naturalmente è molto più comodo percorrere sentieri già battuti; ma non si sa se porteranno alla radura desiderata o da qualche altra parte [...] Compito del ricercatore-metodologo è scegliere via via il percorso [...] A ogni passo si deve scegliere se [...] seguire in proprio un sentiero battuto, oppure un altro appena tracciato, oppure più sentieri in combinazione; se ripercorrere fedelmente questi sentieri, oppure tentare brevi variazioni di percorso (modifiche a tecniche esistenti), oppure addentrarsi nella foresta, immaginando procedure del tutto nuove, magari per confrontarne gli esiti a quelli delle tecniche esistenti [...] L'essenziale del concetto di metodo sta in questo: nella scelta delle tecniche da applicare, nella capacità di modificare tecniche esistenti adattandole ai propri specifici problemi, e di immaginarne delle nuove (ivi, 82-83).

Galton voleva raggiungere una radura nella spessa foresta dell'evoluzione umana; cercava un rapporto ereditario tra le generazioni, e voleva esprimerlo matematicamente. Intraprese il suo percorso seguendo strade già conosciute, fornito degli strumenti — allora consueti — della teoria degli errori. Ma presto si rese conto che quella strada non portava alla radura tanto ambita; le tecniche a sua disposizione si erano rivelate non adatte e una loro reinterpretazione non bastava. Egli dovette scegliere un sentiero sconosciuto: passo dopo passo era costretto a decidere come proseguire; e di conseguenza il suo problema scientifico divenne duplice: doveva appunto costruire tanto nuova conoscenza scientifica quanto nuove tecniche strumentali adeguate. Il suo lavoro non era pertanto solo quello del ricercatore scientifico; Galton era diventato un ricercatore-metodologo.

Seguire strade conosciute e usare strumenti familiari in modo intelligente non è facile, ma almeno per la maggior parte degli scienziati è più rassicurante che avventurarsi nell'ignoto; e non solo per i rischi in termini di costruzione di nuova conoscenza ma anche per quanto riguarda l'accettazione da parte dei pari.

Ma Galton era stato da giovane un esploratore delle zone ignote del Sudafrica e pertanto seguire dei sentieri sconosciuti non lo intimidiva. Tuttavia non sempre era in grado di descrivere come fosse questo percorso, né di fornire una spiegazione razionale del modo in cui i suoi strumenti gli permettevano di percorrerlo. Meno che mai era allora in grado di

giustificare il loro uso. Parafrasando il poeta Machado, Galton si apriva strada camminando: già nel 1869, in *Hereditary Genius*, aveva presentato una prima idea di come immaginava la radura che stava cercando nella foresta; solo sei anni dopo, in *Statistics by Intercomparisons*, divenne evidente che gli strumenti statistici allora consueti non erano adatti. Occorsero un paio di anni in più per poter comunicare effettivamente il nuovo strumento della regressione, che tanto l'aveva avvicinato alla radura; e circa altri dieci anni per cominciare a formalizzare la correlazione, garanzia strumentale dell'arrivo al traguardo. Spesso raggiungeva una soluzione che lo soddisfaceva prima di poter spiegare come c'era arrivato; la sua esplicitazione gli richiedeva uno sforzo enorme, di anni. Egli dovette risolvere diversi problemi scientifici analoghi per rendendosi conto che la sua soluzione — dal punto di vista tecnico — non era specifica, ma ormai era una della quale avrebbero potuto profittare tutti gli scienziati in molteplici indirizzi. Le sue scelte erano più volte intuitive; solo attraverso la collaborazione di esperti si riusciva a formalizzare in un linguaggio matematico ciò che aveva immaginato, malgrado le sue scarse competenze che in teoria avrebbero dovuto precludergli anche l'immaginazione.

D'altra parte, i suoi contemporanei più competenti tecnicamente non sono stati in grado di ideare strumenti analoghi; forse avevano interessi diversi, ma comunque non riuscivano a farsi un'idea della radura immaginata da Galton o di quanto fosse promettente il nuovo sentiero che stava battendo. Hamilton Dickson, secondo quello che riporta Pearson (1914-1930), risolse matematicamente il problema della superficie bivariata normale e nemmeno si occupò di scrivere lui stesso la famosa appendice tecnica del saggio di Galton (si veda sez. 3.2). Tanta era la sua inconsapevolezza e insensibilità nei confronti della svolta epocale che aveva davanti ai suoi occhi! Forse gli mancava quell'immaginazione disciplinata tanto cara a Pearson, e che a suo avviso fa la differenza tra il lavoro meccanico di un tecnico della scienza e quello più artigianale di un genuino scienziato. In effetti, il lavoro di quest'ultimo dipende certamente della padronanza di non pochi strumenti tecnici, ma come quello dell'artista, richiede anche dimensioni tacite e conoscenza personale (Polanyi 1958; 1966), quella che consente Stigler affermare che "la scoperta di Galton della correlazione e regressione dovrebbe considerarsi uno dei più grandi eventi individuali nella storia della scienza" (1999, 6).

Tuttavia, le ricerche di Galton e gli strumenti da lui ideati non sono rimasti semplici "eventi individuali"; e in questo sta una delle differenze sostanziali con il lavoro dell'artista. Si tratta invece di un insieme di decisioni esemplari, di un nuovo paradigma, vale a dire della soluzione di un problema scientifico in un modo nuovo e tanto convincente che altri decisero di seguire lo stesso sentiero per trattare altri problemi.

Infatti, più volte Weldon e Pearson dichiararono che le loro ricerche correlazionali erano state ispirate dalla lettura di *Natural Inheritance*, che li convinse dell'adeguatezza degli strumenti di Galton per indagare su un'ampia gamma di argomenti. Da allora iniziò un lavoro di formalizzazione e generalizzazione — compiuto soprattutto da Pearson e Yule — che rese questi strumenti delle vere tecniche di ricerca: cioè un "complesso più o meno codificato di norme e modi di procedere, riconosciuto da una collettività, trasmesso o trasmissibile per apprendimento, elaborato allo scopo di svolgere una data attività manuale o intellettuale di carattere ricorrente" (Gallino 1978, 712-713).

Ogni volta che ai fini di una data ricerca era necessario determinare la relazione tra proprietà cardinali o categoriali — esprimendola attraverso un singolo numero — si poteva ricorrere agli strumenti della correlazione e dell'associazione. Infatti, "una volta che una procedura nuova, o una modifica a una procedura esistente, è stata ideata e viene codificata e diffusa, essa è già reificata e diviene una tecnica a disposizione della comunità dei ricercatori: non più una capacità privata ma un oggetto pubblico" (Marradi 1996, 83).

Tuttavia, questa reificazione comporta spesso un uso acritico delle tecniche; si dimentica il contesto e i fini per i quali sono state ideate ed esse diventano un fine in sé, un feticcio. Cercando rifugio nel mondo familiare delle tecniche si tenta di nascondere, o persino

negare, che la scelta di quelle giuste e il loro uso adeguato comporta anche la messa in gioco di aspetti non esplicitati della conoscenza personale. Infatti, l'immaginazione e la creatività non intervengono solo di fronte al problema che risulta del tutto nuovo o straordinario; esse sono fondamentali anche nell'affrontare problemi scientifici ricorrenti con tecniche standard. Le operazioni automatizzate che queste ultime richiedono non possono sostituire mai né la sensibilità nei confronti di un problema di ricerca, né il senso comune che guida il loro impiego. La rottura con le strutture del senso comune (auspicata da Bachelard 1938; Bourdieu *et al.* 1968) non deve annientare il ruolo del senso comune nella ricerca sociale.¹⁹⁴ "Le forme di attività intellettuale impegnate nel processo *continuo* che costituisce l'indagine sono molteplici. Ad un lavoro argomentativo esplicito si accompagnano fattori pragmatici e 'opachi', propri del ricercatore [...] alla intuizione 'creatrice' l'incerta inferenza induttiva" (Campelli 1991, 147).¹⁹⁵

Nei primi passi della sua formalizzazione si pensò che sostituendo l'idea metafisica di causalità, la correlazione sarebbe stata in grado di sradicare le conclusioni fallaci — specie nell'ambito delle scienze in cui non si poteva far ricorso alla sperimentazione per individuare l'effetto di una data variabile indipendente. Ma presto i suoi artefici si resero conto che, in quanto tecnica, essa non poteva garantire per sé una conclusione rilevante, significativa, appropriata. Occorreva qualcosa in più: si dovevano considerare i risultati prodotti dall'applicazione della tecnica introducendo degli elementi che essa, da sola, non poteva fornire. Il giudizio, la sensibilità, l'immaginazione erano essenziali come sempre. I concetti di correlazione spuria e di associazione illusoria sono solo due esempi piuttosto evidenti di questa consapevolezza.

In molti dei manuali correnti di statistica attuali si sostiene che gli articoli degli artefici della correlazione non hanno che un valore storico. Apparentemente non hanno niente da dire allo scienziato moderno. Le tecniche da loro ideate hanno acquisito una certa autonomia relativa nei confronti dei problemi e dei fini che le ispirarono, e la loro strumentazione, meno elaborata, si ritiene semplicemente estemporanea. Ritengo che quest'opinione sia profondamente sbagliata. I classici della teoria statistica non hanno solo rilevanza per lo storico; essi possono essere utili al ricercatore moderno e allo studioso di scienze umane. In *The Sociological Imagination*, Mills invita coloro che avevano prodotto lavori di qualità di raccontare in dettaglio come avevano operato: "solo in queste conversazioni con ricercatori esperti i giovani potranno acquisire un'utile sensibilità metodologica" (1959, 28). I saggi dei classici della teoria statistica sono suggestivi anche in questo senso: sono esempi di creatività nell'affrontare i problemi scientifici, presentano chiaramente i passi seguiti, e soprattutto, mostrano consapevolezza del fatto che l'essenziale di una tecnica di ricerca consiste nel suo essere utile a fini di conoscenza.

Ricordando la sua storia accademica con la tranquillità e la saggezza di un'anzianità lucida, Karl Pearson scrisse poco prima di ritirarsi dall'insegnamento attivo che la nuova forma di analisi matematica sviluppata nel suo tempo era stata ideata per risolvere concreti problemi di ricerca circa gli esseri viventi. Queste tecniche di analisi erano cresciute moltissimo, avrebbero continuato a svilupparsi, ma "solo con il dovuto senso della proporzione, in relazione con i bisogni pratici [...] Vale a dire, se [questo] calcolo non si trasforma in un campo di sfruttamento da parte dei matematici puri, esso deve rimanere legato a ricerche sui temi in cui il suo aiuto è più necessario; deve rimanere come scienza pratica, come statistica applicata" (E. Pearson 1938/1948, 179).

¹⁹⁴ Nel caso di Bourdieu *et al.* (1968), ciò significa che il ricercatore non deve prendere come naturale la situazione sociale che indaga o gli strumenti di cui si serve.

¹⁹⁵ Analogamente Bruschi (1996).

RIFERIMENTI BIBLIOGRAFICI

- ABBAGNANO, N. (1961) *Dizionario di Filosofia*. Torino: UTET [citazioni della trad. spagn.: *Diccionario de Filosofia*. México: FCE, 1963].
- ABRAMS, P. (1968) *The Origins of British Sociology*. Chicago: University Press.
- ALDRICH, J. (1995) *Correlations genuine and spurious in Perason and Yule*, in "Statistical Science" 10, 4.
- AUSTIN, J. (1869) *Lectures on Jurisprudence*. London: John Murray.
- BACHELARD, G. (1938) *La formation de l'esprit scientifique. Contribution à une psychanalyse de la conaissance objective*. Paris: Vrin.
- BACON, F. (1620) *Novum Organum*. London: Apud Joannem Billium Typographum Regiem.
- BALFOUR, A. J. (1894) *Foundations of Belief*. London: Longmans, Green.
- BEN-DAVID, J. (1971) *The Scientist's Role in Society: a Comparative Study*. Englewood Cliffs: Prentice Hall.
- BERGER, P. e LUCKMANN, Th. (1966) *The Social Construction of Reality*. New York: Doubleday [citazioni della trad. it.: *La realtà come costruzione sociale*. Bologna: Il Mulino, 1969].
- BERNSTEIN, J. (1955) *La Psicología Factorial de Spearman*, in Spearman, C. *Las Habilidades del Hombre: Su Naturaleza y Medición*. Buenos Aires: Paidós.
- BLALOCK, H. M. (1960) *Social Statistics*. New York: Mc Graw-Hill. [citazioni della trad. spagn., 2da ed.: *Estadística Social*. Mexico: FCE, 1986].
- BLUMER, H. (1969) *Symbolic Interactionism: Perspective and Methods*. Englewood Cliffs: Prentice-Hall.
- BOOLE, G. (1847) *The Mathematical Analysis of Logic, being an Essay Towards a Calculus of Deductive Reasoning*. London: Macmillan.
- BOOLE, G. (1854) *An Investigation of the Laws of Thought, on which are Founded the Mathematical Theories of Logic and Probabilities*. London: Walton and Maberley.

- BOOTH, C. (1891) *Labour and Life of the People*. London: Williams and Norgate.
- BOURDIEU, P. *et al.* (1968) *Le métier de sociologue. Problèmes épistémologiques*. Paris: Mouton.
- BOWLEY, A. L. (1901) *Elements of Statistics*. London: P.S. King and Son.
- BRAVAIS, A. (1846) *Analyse Mathématique sur les Probabilités des Erreurs de Situation d'un Point*, in "Mémoires présentés par divers Savants à l'Académie Royale des Sciences de l'Istitut de France" 9.
- BRESCIANI, C. (1909) *Sui Metodi per Misurare la Correlazione*, in "Giornale degli Economisti", marzo-aprile.
- CAMPELLI, E. (1991) *Il metodo e il suo contrario. Sul recupero della problematica del metodo in sociologia*. Milano, Franco Angeli.
- CAMPELLI, E. (1999) *Da un Luogo Comune. Elementi di Metodologia delle Scienze Sociali*. Roma: Carroci.
- CARDOSO, F. H. e FALETO, E. (1969) *Dependencia y Desarrollo en América Latina*. México: Siglo XXI.
- CARNAP, R. (1939) *Foundations of Logic and Mathematics*. Chicago: University Press.
- CARPENTER, W. (1851) *The Correlation of the Physical and Vital Forces*, in "British and Foreign Medico-Chirurgical Review", 8: 206-238
- CLIFFORD, W. K. (1879) *Lectures and Essays*. London: Macmillan.
- COLEMAN, W. (1982) *Death is a Social Disease: Public Health and Political Economy in Early Industrial France*. Madison.
- COWAN, R. S. (1972) *Francis Galton's Statistical ideas: the Influence of Eugenics*, in "Isis" 63.
- CROMBIE, A. C. (1994) *Styles of Scientific Thinking in the European Tradition*. London: Duckworth.
- CULLEN, M. J. (1975) *The Statistical Movement in Early Victorian Britain*. Brighton: Harvester Press.
- DARWIN, C. (1859) *On the Origin of Species*. London: John Murray.
- DARWIN, C. (1868) *The Variation of Animals and Plants under Domestication*. London: John Murray.

- DARWIN, L. (1926) *The Need for Eugenic Reform*. London: John Murray.
- DE MOIVRE, A. (1718) *The Doctrine of Chances: or, A Method of Calculating the Probability of Events in Play*. London: W. Pearson.
- DE MORGAN, A. (1838) *Theory of Probabilities*, in "Encyclopaedia Metropolitana" vol. 2. London.
- DE MORGAN, A. (1847) *Formal Logic; or The Calculus of Inference, Necessary and Probable*. London: Taylor and Walton.
- DENIS, D. J. (2000) *The origins of correlation and regression*. "61st Annual Convention of the Canadian Psychological Association", Ottawa, Canada.
- DEWEY, J. (1889) *Galton's Statistical Methods*, in "Publications of the American Statistical Association" 7.
- DI FRANCO, G. e MARRADI, A. (2003) *Analisi Fattoriale e Analisi in Componenti Principali*. Roma: Bonanno.
- DILTHEY, W. (1883) *Einleitung in die Geisteswissenschaften*, Leipzig: Dunker und Humblot.
- DU BOIS-REYMOND, P. (1890) *Über die Grundlagen der Erkenntnis in den Exacten Wissenschaften*. Tübingen.
- EBBINGHAUS, H. (1885) *Über das Gedächtnis*. Leipzig: Duncker & Humboldt.
- EDGEWORTH, F. Y. (1877) *New and Old Methods of Ethics*. London: James Parker.
- EDGEWORTH, F. Y. (1881) *Mathematical Physics: An Essay on the Application of Mathematics to the Moral Sciences*. London: Kegan Paul.
- EDGEWORTH, F. Y. (1884) *The Philosophy of Chance*, in "Mind" 9.
- EDGEWORTH, F. Y. (1887) *Metempsychosis*. London: Temple.
- EDGEWORTH, F. Y. (1892) *On Correlated Averages*, in "Philosophical Magazine", series 5, 34.
- EDGEWORTH, F. Y. (1893) *Statistical Correlation Between Social Phenomena*, in "Journal of the Royal Statistical Society" 56.
- EDGEWORTH, F. Y. (1902) *The Law of Error*, in "Encyclopædia Britannica", 28 Supplement to the 9th Edition.
- EISENHART, C. (1974) *Pearson*, in "Dictionary of Scientific Biography". New York: Charles Scribner's Sons.

- ELDERTON, E. (1910) *A first Study of the Influence of Parental Alcoholism on the Physique and Ability of the Offspring* (con la collaborazione di K. Pearson), in "Eugenics Laboratory Memoirs" X. London: Dulau.
- FAYTER, P. (1997) "Strange New Worlds of Space and Time: Late Victorian Science and Science Fiction", in Lightman, B (ed.) *Victorian Science in Context*. Chicago: University Press
- FECHNER, G. (1860) *Elemente der Psychophysik*. Leipzig: Breitkopf und Härtel.
- FERRATER MORA, J. (1994) *Diccionario de Filosofía*. Barcelona: Ariel.
- FEYERABEND, P. K. (1970) *Against Method. Outline of an Anarchistic Theory of Knowledge*, in M. Radner e S. Winokur (eds.) *Analyses of Theory and Methods of Physics and Psychology*. Minneapolis: University of Minnesota Press (Ristampato da New Left Books, London, 1975).
- FICHMAN, M. (1997) "Biology and Politics: Defining the Boundaries", in Lightman, B (ed.) *Victorian Science in Context*. Chicago: University Press
- FITZPATRICK, P. J. (1960) *Leading British Statisticians of the Nineteenth Century*, in "Journal of the American Statistical Association" 55, 289.
- FLECK, L. (1935) *Entstehung und Entwicklung einer wissenschaftlichen Tatsache. Einfuehrung in die Lehre vom Denkstil und Denkkollektiv*. Basel: Benno Schwabe
- GALLINO, L. (1978) *Dizionario di Sociologia*. Torino: UTET.
- GALLOWAY, T. (1839) *A treatise on Probability*. Edinburgh.
- GALTON, F. (1853) *The Narrative of an Explorer in Tropical South Africa*. London: John Murray.
- GALTON, F. (1869) *Hereditary Genius. An Inquiry into its Laws and Consequences*. London: MacMillan.
- GALTON, F. (1874) *English Men of Science: their Nature and Nurture*. London: Macmillan.
- GALTON, F. (1875) *Statistics by Intercomparison*, in "Philosophical Magazine" Series 4, 49.
- GALTON, F. (1877) *Typical Laws of Heredity*, in "Proceedings of the Royal Institution" 8.

- GALTON, F. (1879) *Psychometric Experiments*, in "Brain" 2.
- GALTON, F. (1883) *Inquiries into Human Faculty and its Development*. London: Macmillan.
- GALTON, F. (1885a) *Address to the Anthrpological Section of the British Association*, in "Nature" 32.
- GALTON, F. (1885b) *Regression Towards Mediocrity in Hereditary Stature*, in "Journal of the Anthropological Institue" 15.
- GALTON, F. (1886) *Family likeness in Stature*, in "Proceedings of Royal Society of London " 40.
- GALTON, F. (1888) *Co-relations and their Measurement, Chiefly from Anthropometric Data*, in "Proceedings of the Royal Society of London" 45.
- GALTON, F. (1889) *Natural Inheritance*. London: Macmillan.
- GALTON, F. (1890) *Kinship and Correlation*, in "North American Review" 150.
- GALTON, F. (1897) *Notes to the Memoir by Professor Karl Pearson, F.R.S., on Spurious Correlation*, in "Proceedings of the Royal Society of London" 60.
- GALTON, F. (1908) *Memories of My Life*. London: Methuen.
- GALTON, F. (1909) *Essays in Eugenics*. London: Eugenics Society.
- GAUSS, C. F. (1823) *Theoria combinationis observationum minimis obnoxiae*. Göttingen.
- GERMANI, G. (1965) *Política y Sociedad en una Época de Transición, de la Sociedad Tradicional a la Sociedad Moderna*. Buenos Aires: Paidós.
- GIDDENS, A. (1977) "Positivism and its critics", in Giddens, A. *Studies in Social and Political Theory*. London: Hutchison.
- GIDDENS, A. (1979) *Central Problems in Social Theory. Action, Structure and Contradiction in Sociological Analysis*. London & New York: Macmillan.
- GLASER, B. G. e
STRAUSS, A. L. (1967) *The Discovery of Grounded Theory. Strategies for Qualitative Research*. Chicago: Aldine.
- GOODMAN, L. A. e
KRUSKAL, W. H. (1954) *Measures of Association fo Cross Classifications*, in "Journal of the American Statistical Association" 49.

- GOLDTHORPE, J. H. (2001) *Causation, Statistics, and Sociology*, in "European Sociological Review" 17, 1.
- GRAUNT, J. (1662) *Natural and Political Observations on the Bills of Mortality*. London.
- GROVE, W. R. (1846) *The Correlation of Physical Forces, The Substance of a Course of Lectures*. London.
- GULLIKSEN, H. (1950) *Theory of Mental Tests*. New York: Willey.
- HACKING, I. (1990) *The Taming of Chance*. Cambridge: University Press.
- HALFPENNY, P. (1982) *Positivism and Sociology*. London: Allen and Unwin.
- HEMPEL, C. G. (1952) *Fundamentals of Concept Formation in Empirical Science*. Chicago: University Press.
- HEMPEL, C. G. (1966) *Philosophy of Natural Science*. Englewood Cliffs: Prentice Hall.
- HENDRY, D. F. (1980) *Econometrics. Alchemy or Science*, in "Economica" 47.
- HERON, D. (1906) *On the Relation of Fertility in Man to Social Status*, in "Drapers' Company Research Memoirs. Studies in National Deterioration", 1. London: Dulau.
- HERON, D. (1911) *The Danger of certain Formulae Suggested as Substitutes for the Correlation Coefficient*, in "Biometrika" 8.
- HERSCHEL, J. (1849) *Outlines of Astronomy*. London: Longman, Brown, Green and Longmans Publisher.
- HOBSBAWM, E. (1968) *Labouring Men: Studies in the History of Labour*. London: Weidenfeld and Nicolson.
- HOOKE, R. H. (1901a) *On the Correlation of the Marriage-rate with Trade*, in "Journal of the Royal Statistical Society" 64.
- HOOKE, R. H. (1901b) *The Suspension of the Berlin Produce Exchange and its Effect upon Corn Prices*, in "Journal of the Royal Statistical Society" 64.
- HOOKE, R. H. (1907) *The Correlation of the Weather and the Crops*, in "Journal of the Royal Statistical Society" 70.
- HUGHES, J. e
SHARROCK, W. (1997) *The Philosophy of Social Research*, Terza edizione. London: Longman. [citazioni della trad. spagn.: *La filosofía de la Investigación Social*. México: FCE, 1999].

- HULL, C. H., ed. (1899) *The Economic Writings of Sir William Petty together with the Observations upon the Bills of Mortality more probably by Captain John Graunt*. Cambridge.
- HUME, D. (1748) *An Enquiry Concerning Human Understanding*.
- HUMPHREY, N. A. (1897) *Vaccination and Small-Pox Statistics*, in “Journal of the Royal Statistical Society” 60.
- HUXLEY, T. H. (1864) *Lectures on the Elements of Comparative Anatomy, and on the Classification of Animals*. London: John Churchill and Sons.
- JEVONS, W. S. (1870) *On a General System of Numerically Definite Reasoning*, in “Memoirs of the Manchester Literary and Philosophical Society.”
- JEVONS, W. S. (1874) *The Principles of Science. A Treatise on Logic and Scientific Method*. London: Macmillan.
- JOHN, V. (1884) *Gesichte der Statistik*. Stuttgart: Ferdinand Encke.
- KANT, I. (1781) *Kritik der Reinen Vernunft*. Königsberg: Reiger.
- KENDALL, M. G. (1952) *George Udny Yule*, in “Journal of the Royal Statistical Society” 115, 1.
- KENDALL, M. G. (1968) *Studies in the History of Probability and Statistics XIX. Francis Ysidro Edgeworth*, in “Biometrika” 55, 2.
- KENDALL, M. G. e
BUCKLAND, W. R. (1976) *A Dictionary of Statistical Terms*. London: Longman [citazioni della trad. spagn.: *Diccionario de Estadística*, Madrid: Pirámide, 1980].
- KENDALL, M. G. e
DOIG, A. G. (1968) *Bibliography of Statistical Literature Pre-1940*. Edinburgh: Oliver and Boyd.
- KERN, S. (2004) *A Cultural History of Causality: Science, Murder Novels and Philosophy Since 1830*. Prentice: University Press.
- KEYNES, J. M. (1926) *Obituary. Francis Ysidro Edgeworth*, in “Economic Journal” 36, 141: 140-153.
- KIRCHHOFF, G. (1876) *Vorlesungen über Mathematische Physik*. Leipzig: B. G. Teubner
- KJAERGAARD, P. (2000) *Defending Science: ‘Genuine Scientific Men’ and the Limits of Natural Knowledge*. Aarhus Universitet.

- KLIMOVSKY, G. (1972) *Ciencia e Ideología*, in "Ciencia Nueva" 10.
- KNIES, K. G. (1850) *Die Statistik als Selbständige Wissenschaft*. Kassel.
- KUHN, Th. (1961) *The Function of Measurement in Modern Physical Science*, in "Isis" 52.
- KUHN, Th. (1962) *The Structure of Scientific Revolutions*. Chicago: University Press.
- LAPLACE, P. S. (1774) *Mémoire sur la probabilité des causes par les événements*, in "Memoires de mathématique et de physique, présentés à l'Académie Royale des Sciences, par divers savans" 6.
- LANCASTER, H. O. (1972) *Development of the Notion of Statistical Dependence*, in "Mathematics Chronicle" 2.
- LAZARSFELD, P. F. (1961) *Notes on the History of Quantification in Sociology. Trends, Sources and Problems*, in "Isis" 52, 2.
- LEGENDRE, A. M. (1805) *Nouvelles methodes pour la détermination des orbites des comètes*. Paris: Courcier.
- LEE, A. (1908) *On the Manner in which the Percentage of Employed Workmen in this Country is Related to the Import of Articles Wholly or Mainly Manufactured*, in "Economic Journal" 18.
- LEVINE, G. (1996) *Science and Citizenship: Karl Pearson and the Ethics of Epistemology*, in "Modernism/ Modernity" 3.3. <http://muse.jhu.edu/journals/modernism-modernity/v003/3.3levine.html>
- LEVINE, G. (2000) *Two Ways Not To Be a Solipsist: Art and Science, Pater and Pearson*, in "Victorian Studies" 43, 1.
- LIGHTMAN, B. (1997) "'The Voices of Nature': Popularizing Victorian Science", in Lightman, B. (ed.) *Victorian Science in Context*. Chicago: University Press.
- MACH, E. (1886) *The Analysis of the Sensations. Anti-metaphysical*, in "The Monist" 1.
- MACKENZIE, D. A. (1978) *Statistical Theory and Social Interests: A Case Study*, in "Social Studies of Science" 8, 1.
- MACKENZIE, D. A. (1981) *Statistics in Britain, 1865-1930. The Social Construction of Scientific Knowledge*. Edinburgh: University Press.

- MALTHUS, R. (1798) *Essays on the Principles of Population, as it affects the future improvement of society*. London: J. Johnson.
- MARCH, L. (1905) *Comparaison Numerique de Courbes Statistiques*, in "Journal de la Societ  de Statistique de Paris" 255.
- MARRADI, A. (1980) *Concetti e Metodi per la Ricerca Sociale*. Firenze: Giuntina.
- MARRADI, A. (1981) *Misurazione e Scale: Qualche Riflessione e una Proposta*, in "Quaderni di Sociologia" 29, 4.
- MARRADI, A. (1984) *Teoria: una tipologia dei significati*, in "Sociologia e Ricerca Sociale" 5, 13.
- MARRADI, A. (1987) *Linguaggio scientifico o torre di Babele?*, in "Rivista Italiana di Scienza Politica" 18, 1.
- MARRADI, A. (1996) *Metodo come arte*, in "Quaderni di Sociologia" 40, 10.
- MARRADI, A. (1997a) "Esperimento, associazione, insieme non standard?" in Bettin, G. (ed.) *Politica e Societ *. Padova: Cedam.
- MARRADI, A. (1997b) *Linee guida per l'analisi bivariata dei dati nelle scienze sociali*. Milano: Franco Angeli.
- MEITZEN, A. (1886) *Gesichte, Theorie und Technik der Statistik*. Berlin: Hertz.
- MELBERG, H. O. (2000) *From Spurious Correlation to Misleading Association: The Nature and Extent of Spurious Correlation and its Implications for the Philosophy of Science with Special Emphasis on Positivism*. Oslo: University of Oslo.
- MICHELI, G. A. e MANFREDI, P. (1995) *Correlazione e Regressione*. Milano: Franco Angeli.
- MILL, J. S. (1843) *A System of Logic, Ratiocinative and Inductive*. London.
- MILLS, C. W. (1959) *The Sociological Imagination*. New York: Oxford University Press.
- MITCHELL, D. (1973) *Storia della Sociologia Moderna*. Verona: Mondadori.
- MORGAN, C. L. (1891) *Animal Life and Intelligence*. London: Arnold.
- MORRISON, M. (2002) *Modelling Populations: Pearson and Fisher on Mendelism and Biometry*, in "British Journal of the Philosophy of Science" 53.

- MULKAY, M. (1979) *Science and the Sociology of Knowledge*. London: Allen & Unwin.
- NAGEL, E. (1957) "Introduction", in Pearson, K. *The Grammar of Science*. New York: Meridian.
- NAGEL, E. (1961) *The Structure of Science: Problems on the Logic of Scientific Explanation*. London: Routledge & Kegan Paul.
- NEURATH, O. (1931) "Physikalismus", in *Scientia* 50.
- NEWSHOLME, A. (1906) *An Inquiry into the Principle Causes of the Reduction of the Death-rate from Phthisis during the Last Forty Years, with Special Reference to the Segregation of Phthisical Patients in General Institutions*, in "Journal of Hygiene" 6.
- NEWSHOLME, A. (1908) *The Prevention of Tuberculosis*. London: Mathuen.
- NEWTON, I. (1687) *Philosophiæ Naturalis Principia Mathematica*. London: Jussu Societatis Regiæ ac Typis Josephi Streater.
- NORTON, B. J. (1978) *Karl Pearson and Statistics: The Social Origins of Scientific Innovation*, in "Social Studies of Science" 8, 1.
- NORTON, J. P. (1902) *Statistical Studies in the New York Money Market*. New York: Macmillan.
- OUTHWAITE, W. (1987) *New Philosophies of Social Science*. Basingstoke: Macmillan.
- PAWSON, R. (1889) *A Measure for Measures: A Manifesto for Empirical Sociology*. London: Routledge and Kegan Paul.
- PEARSON, E. S. (1938) *Karl Pearson: An Appreciation of Some Aspects of his Life and Work*. Cambridge: University Press. [citazioni della trad. spagn.: *Pearson, creador de la estadística aplicada*. Buenos Aires: Espasa-Calpe, 1948].
- PEARSON, K. ['Locki'] (1880) *The New Werther*. London: Kegan Paul.
- PEARSON, K. (1888) *The Ethic of Freethought*. London: T. F. Unwin.
- PEARSON, K. (1892) *The Grammar of Science*. London: Scott [citazioni della terza revisione, 1911. New York: Meridian, 1957].
- PEARSON, K. (1896) *Mathematical Contributions to the Theory of Evolution: Regression, Heredity, and Panmixia*, in "Philosophical Transactions of the Royal Society of London" 187.
- PEARSON, K. (1897) *Mathematical Contributions to the Theory of Evolution: On a Form of Spurious Correlation which May*

- Arise when Indices Are Used in the Measurement of Organs* ,
in "Proceedings of the Royal Society" 60.
- PEARSON, K. (1900) *Mathematical Contributions to the Theory of Evolution VII. On the Correlation of Characters not Quantitatively Measurable*, in "Philosophical Transactions of the Royal Society of London" 195.
- PEARSON, K. (1902) *On the Systematic Fitting of Curves to Observations and Measurements*, in "Biometrika" 1:
- PEARSON, K. (1904) *Mathematical Contributions to the Theory of Evolution, XII: On a Generalized Theory of Alternative Inheritance with Special reference to Mendel's Law*, in "Philosophical Transactions of the Royal Society", series A 203.
- PEARSON, K. (1905) *On the General Theory of Skew Correlation and Non-Linear Regression*, in "Draper's Company Research Memoirs". Biometric Series II. London: Dulau and Co.
- PEARSON, K. (1914-1930) *Life and letters of Francis Galton*. Cambridge: University Press.
- PEARSON, K. (1920) *Notes on the History of Correlation*, in "Biometrika" 13, 1.
- PEARSON, K. e
ELDERTON, E. (1910) *A Second Study of the Influence of Parental Alcoholism on the Physique and Ability of the Offspring: Being a Reply to Certain Medical Critics of the First Memoir and an Examination of the Rebutting Evidence Cited by Them*, in "Eugenics Laboratory Memoirs" XII. London: Dulau.
- PEARSON, K. e
FILON, L. N. G. (1898) *Mathematical Contributions to the Theory of Evolution, IV: On the Probable Errors of Frequency Constants and on the Influence of Random Selection on Variation and Correlation.*, in "Philosophical Transactions of the Royal Society", Series A 191.
- PEARSON, K. e
HERON, D. (1913) *On Theories of Association*, in "Biometrika" 9.
- PEARSON, K. e
LEE, A. (1897) *On the Distribution of Frequency (Variation and Correlation) of Barometric Height at Divers Stations*, in "Philosophical Transactions of the Royal Society", Series A 190.
- PEARSON, K. e

- PEARSON, E. S. (1922) *On Polychoric Coefficients of Correlation*, in "Biometrika" 14.
- PEIRCE, C. S. e JASTROW, J. (1885) *On Small Differences of Sensations*, in "Memoirs of the National Academy of Sciences for 1884" 3.
- PETTY, W. (1691) *The Political Anatomy of Ireland*, in *Economic Writings*, vol. 1.
- PHILLIPS, D. C. (1987) "The Demise of Positivism", in *Philosophy, Science and Social Enquiry*. Oxford: Pergamon.
- PITRONE, M. C. (1996) *Il sondaggio*. Milano: Franco Angeli.
- POLANYI, M. I. (1958) *Personal Knowledge. Towards Post-Critical Philosophy*. London: Routledge & Kegan Paul.
- POLANYI, M. I. (1966) *The Tacit Dimension*. New York: Doubleday.
- POPPER, K. R. (1934) *Logic der Forschung*. Wien: Springer.
- POPPER, K. R. (1945) *The Open Society and Its Enemies*. London: Routledge & Kegan Paul.
- POPPER, K. R. (1949) *Naturgetze und Theoretische Systeme*, trad ingl.: *Objective Knowledge; An Evolutionary Approach*. Oxford: Clarendon Press, 1972
- POPPER, K. R. (1963) *Conjectures and Refutations*. London: Routledge.
- PORTER, Th. (1986) *The Rise of Statistical Thinking, 1820-1900*. Princeton: University Press.
- QUETELET, A. (1844) *Recherches Statistiques*. Brussels: Hayez.
- REICHENBACH, H. (1938) *Experience and Prediction. An Analysis of the Foundations and the Structure of Knowledge*. Chicago: University Press.
- RICHARDS, J. (1997) "The Probable and the Possible in Early Victorian England", in Lightman, B (ed.) *Victorian Science in Context*. Chicago: University Press
- RICKERT, H. (1899) *Kulturwissenschaft und Naturwissenschaft* Tübingen: Mohr.
- RIDDLE, C. (1958) *Karl Pearson's Philosophy of Science*. (Tesi di dottorato non pubblicata. New York: Columbia University)

- RUSSELL, B. (1935) *Religion and Science*. London: Thornton and Butterworth [citazioni della trad. spag.: *Religión y ciencia*. Buenos Aires: FCE, 1987].
- SCHAFFER, S. (1997) "Metrology, Metrication, and Voctorian Values", in Lightman, B (ed.) *Victorian Science in Context*. Chicago: University Press
- SCHLEIERMACHER, F. D. (1838) *Hermeneutik und Kritik mit besonderer Beziehung auf das Neue Testament*. Berlin: Reiner.
- SCHLICK, M. (1918) *Allgemeine Erkenntnislehre*. Berlin: Springer.
- SCHOLS, C. M. (1886) *Théorie des Erreurs dans le Plan et dans l'Espace*, in "Annales de l'Ecole Polytechnique de Delft" 2.
- SCHUTZ, A. (1932) *Der Sinnhafte Aufbau der Sozialen Welt*. Wien: Springer.
- SCHUTZ, A. (1962) *The Problem of Social Reality* (Collected Papers, vol. I). Den Haag: Nijhoff.
- SEAL, H. L. (1967) *The Historical Development of the Gauss Linear Model*, in "Biometrika", 54.
- SHAPIN, S. (1990) "Science and the Public", in Olby, R. C.; Cantor, G. N.; Christie, J. R. R. e Hodge, M. J. S. (eds.) *Companion to the History of Modern Science*. London: Routledge.
- SINCLAIR, J. (1791-1797) *The Statistical Account of Scotland*. Edinburgh.
- SPEARMAN, C. E. (1904a) *The Proof and Measurement of Association Between Two Things*, in "American Journal of Psychology" 15.
- SPEARMAN, C. E. (1904b) *General Intelligence, Objectively Determined and Measured*, in "American Journal of Psychology" 15.
- SPEARMAN, C. E.. (1910) *Correlation Calculated from Faulty Data*, in "British Journal of Psychology" 3.
- SPEARMAN, C. E. (1927) *The Abilities of Man: Their Nature and Measurement*. London: Macmillan. [citazioni della trad. spag.: *Las Habilidades del Hombre: Su Naturaleza y Medición*. Buenos Aires: Paidós, 1955].
- STEVENS, S. S. (1946) *On the Theory of Scales of Measurement*, in "Science" 103.
- STIGLER, S. M. (1981) *Gauss and the Invention of Least Squares*, in "The Annals of Statistics" 9.

- STIGLER, S. M. (1986) *The History of Statistics: The Measurement of Uncertainty before 1900*, Cambridge, MA, and London: Belknap Press of Harvard University Press.
- STIGLER, S. M. (1999) *Statistics on the Table: The History of Statistical Concepts and Methods*, Cambridge, MA, and London: Harvard University Press.
- STUART, J. (1883) *A Chapter of Science*. London.
- SUNKEL, O. (1971) *El Subdesarrollo Latinoamericano*. México: Siglo XXI.
- TANKARD, J. W. (1984) *The Statistical Pioneers*. Cambridge: Schenkman.
- TOULMIN, S. E. (1972) *Human Understanding*. Princeton: University Press.
- URRY, J. (1984) "A History of Field Methods", In Ellen, R. (ed.) *Ethnographic Research: A guide to General Conduct*. London: Academic Press.
- VENN, J. (1866) *The Logic of Chance: an Essay on the Foundations and Province of the Theory of Probability, with Special Reference to its Logical Bearings and its application to Moral and Social Science*. London: Macmillan.
- VON BIELFELD, J. F. (1770) *Elements of Universal Erudition, Containing an Analytical Abridgement of the Sciences, Polite Arts, and Belles Lettres*. London.
- VON ZIMMERMANN, E. A. (1787) *A Political Survey of the Present State of Europe, in Sixteen Tables; Illustrated with Observations on the Wealth and Commerce, the Government, Finances, Military State, and Religion of the Several Countries*. London: C. Dilly.
- WALKER, H. M. (1929) *Studies in the History of Statistical Method*. Baltimore: Williams and Wilkins.
- WALLACE, H. A.
SNEDECOR, G. W. (1925) *Correlation and Machine Calculations*. Ames: Iowa State College Press.
- WEBER, M. (1904) *Die protestantische Ethik und der "Geist" des Kapitalismus*. Tübingen: Mohr.
- WELDON, W. F. R. (1890) *The Variations Occurring in Certain Decapod Crustacea*, in "Proceedings of the Royal Society" 47.
- WELDON, W. F. R. (1892) *Certain Correlated Variations in Crangon Vulgaris*, in "Proceedings of the Royal Society" 51.

- WELDON, W. F. R. (1893) *On Certain Correlated Variations in Carcinus Moenas*, in "Proceedings of the Royal Society" 54.
- WETERGAARD, H. (1932) *Contributions to the History of Statistics*. London: P. S. King.
- WILLIAMS, R. H. *et al.* (2003) *Charles Spearman: British Behavioral Scientist*, in "Human Nature Review" 3.
- WILLIAMS, R. H. (2004) *George Udny Yule: Statistical Scientist*, in "Human Nature Review" 4.
- WINTER, A. (1997) "The Construction of Orthodoxies and Heterodoxies in the Early Victorian Life Sciences", in Lightman, B (ed.) *Victorian Science in Context*. Chicago: University Press
- YOUNG, R. M. (1985) *Darwin's Metaphor: Nature's Place in Victorian Culture*. Cambridge: University Press.
- YULE, G. U. (1895-1896) *On the Correlation of Total Pauperism with Proportion of Out-Relief*, in "Economic Journal" 5.
- YULE, G. U. (1897a) *On the Significance of Bravais' Formulae for Regression, in the Case of Skew Correlation*, in "Proceedings of the Royal Society of London" 60.
- YULE, G. U. (1897b) *On the Theory of Correlation*, in "Journal of the Royal Statistical Society" 60.
- YULE, G. U. (1899) *An Investigation into the Causes of Changes in Pauperism in England during the Last Two Inter-censal Decades*, in "Journal of the Royal Statistical Society" 62.
- YULE, G. U. (1900) *On the Association of Attributes in Statistics*, in "Philosophical Transactions of the Royal Society of London" Series A 194.
- YULE, G. U. (1902) *Mendel's Laws and their Probable Relations to Intra-Racial Heredity*, in "New Phytologist" 1.
- YULE, G. U. (1903) *Notes on the Theory of Association of Attributes in Statistics*, in "Biometrika" 2, 2.
- YULE, G. U. (1906) *On a Property which hold good for all Groupings of a Normal Distribution of Frequency for two Variables*, in "Proceedings of the Royal Society of London" Series A, 77.
- YULE, G. U. (1909) *The application of the Method of Correlation to Social and Economic Statistics*, in "Journal of the Royal Statistical Society" 72.

- YULE, G. U. (1911) *An Introduction to the Theory of Statistics*. London: Griffin.
- YULE, G. U. (1912) *On the Methods of Masuring Asociation Btween Two Attributes*, in "Journal of the Royal Statistical Society" 75.
- YULE, G. U. (1926) *Why Do We Sometimes Get Nonsense Correlations between Time series? A Study in Sampling and the Nature of Time-series*, in "Journal of the Royal Statistical Society" 89.
- YULE, G. U. (1936) *Karl Pearson*, in "Royal Society" Obituary Notices 2.
- YULE, G. U. (1938) *Notes of Karl Pearson's Lectures on the Theory of Statistics 1894-96*, in "Biometrika" 30.
- YULE, G. U. (1944) *The Statistical Study of Literary Vocabulary*. Cambridge: University Press.
- YULE, G. U. e
KENDALL, M. G. (1937) *An Introduction to the Theory of Statistics*. London: Lippincott.