FELICE VINCI

PROF. ORD. DI STATISTICA NELLA R. UNIVERSITÀ DI BOLOGNA

MANUALE DI STATISTICA

INTRODUZIONE ALLO STUDIO QUANTITATIVO DEI FATTI SOCIALI

VOLUME PRIMO

NICOLA ZANICHELLI EDITORE

BOLOGNA 1934-XII

FELICE VINCI
PROF. ORD. DI STATISTICA NELLA R. UNIVERSITÀ DI BOLOGNA

MANUALE DI STATISTICA

INTRODUZIONE ALLO STUDIO QUANTITATIVO
DEI FATTI SOCIALI

VOL. I

LABORAVI FIDENTER

NICOLA ZANICHELLI EDITORE
BOLOGNA 1934-XII

Stabilimenti Poligrafici Riuniti - II - 1934-XII - Bologna

PREFAZIONE

La parola statistica ha assunto con l'andar del tempo diversi significati. Le descrizioni delle condizioni politiche, geografiche, economiche, morali degli Stati — che al principio dell'evo moderno diedero luogo in alcune città italiane ad una fioritura di opere di vario pregio, tra le quali meritano di essere segnalate quelle di Francesco Sansovino (1521-1586) e di Giovanni Botero (1540-1617) — sboccarono ben presto in una nuova disciplina, alla quale in Germania Gottfried Achenwall (1719-1772) diede il nome di statistica, termine derivato dalla parola italiana: statista, uomo di Stato.

Coll'estendersi delle rilevazioni quantitative da parte degli Stati e col sorgere e l'affermarsi in Inghilterra — soprattutto per opera di Graunt e di Petty — dell'indirizzo dei cosidetti aritmetici politici, intenti a ricercare leggi empiriche dei fatti sociali, quelle descrizioni andarono raccogliendo sempre più largamente dati numerici, i quali infine prevalsero tanto sulle notizie e gli elementi qualitativi, che furon chiamati a loro volta statistiche. E pertanto si cominciò a parlare di statistica delle nascite, di statistica dei prezzi.....

Intanto, con lo sviluppo del calcolo delle probabilità e con l'impulso di speciali sodalizi e dei congressi, i metodi di analisi di cotesti dati si sviluppavano e si affinavano finchè formarono un corpo organico di metodi, a cui si diede pure il nome di statistica. E furon chiamati statistici i cultori di essa.

Poichè tali metodi si rivelarono della stessa natura di quelli usati in meteorologia, biologia, ecc., non si mancò di trarre giovamento dai rapidi progressi ch'essi avevano compiuto in queste discipline; anzi si attuò una così stretta cooperazione di lavoro (uno « schema teorico » va sotto il nome del demografo Lexis e un « metodo di traslazione » è dovuto all'economista Edgeworth), che anche quei naturalisti incominciarono a parlare di statistica meteorologica, biologica,.... e pure di statistica intesa come corpo di metodi.

Oggi tali metodi sono molto progrediti e formano l'oggetto di trattazioni formali, indipendenti dai problemi concreti, onde hanno tratto origine: limitandoci alla più recente letteratura, ricordiamo le opere per molti aspetti pregevoli di Whittaker-Robinson e di Jordan.

Ma si è sentito pur sempre il bisogno di trattazioni metodologiche rispondenti alle particolari esigenze dei varî rami di applicazione del metodo statistico, e son venuti fuori trattati di statistica ad uso degli attuarî, dei fisici, dei biologi, dei psicologi, dei cultori di scienze sociali e così via.

Questo secondo genere di trattazioni non è meno importante del primo, perchè mantiene il metodo a stretto contatto con le esigenze concrete, ch'esso è chiamato a soddisfare; gli offre continuo alimento e spesso la soluzione intuitiva di problemi, che i mezzi analitici non siano ancora in grado di dare a tutto rigore; infine mette in guardia contro le oziose raffinatezze e le facili aberrazioni.

Esso è specialmente opportuno per le scienze sociali, data la modesta preparazione matematica di molti, anche valorosi, cultori delle svariate discipline, che negli ultimi cento anni sono germogliate dal ceppo della statistica e dell'aritmetica politica, e data la necessità di diffondere tra di essi almeno la conoscenza dei principî fondamentali e delle molteplici applicazioni sociali del metodo.

Questo manuale si rivolge appunto ai cultori di scienze sociali e, dopo quel che abbiamo detto, crediamo superfluo giustificarne il particolare contenuto e i limiti, il carattere — che ogni pagina rivela — di introduzione allo studio quantitativo dei fatti sociali.

Compio il dovere di ringraziare pubblicamente i professori Raffaele D'Addario, Giovanni Lasorsa, Giuseppe Medici e Dionisio Tenderini della cordiale sollecitudine, con cui si sono compiaciuti di correggere le bozze di stampa di questo manuale, e dei loro utili suggerimenti.

Bologna, R. Università, Settembre 1933-XI.

F. VINCI

ERRATA CORRIGE

VOLUME PRIMO

Pag. 57, riga 7: Invece di « corrispondente dei morti », leggasi « corrispondente dei morti, a partire — si badi — dall'origine 800 ».

Pag. 64, riga 3: Invece di « che dà luogo », leggasi « che nella figura 22 dà luogo ».

Pag. 68, riga 4: Invece di « diagrammi alle », leggasi « diagrammi relativi alle ».

Pag. 83, riga 8: Invece di « lunghezza delle aste », leggasi « lunghezza rimanente delle aste ».

Pag. 92, riga 23: Invece di « corrisponderebbe al rapporto », leggasi « sarebbe il valore ».

Pag. 92, riga 24: Invece di « del gruppo e, pertanto, tale rapporto più frequente sarebbe dato dalla », leggasi « del gruppo logaritmico, valore che pertanto corrisponderebbe alla ».

Pag. 104, riga 34: Invece di « calcolati », leggasi « costruiti ».

Pag. 104, riga 41: Invece di « simili », leggasi « quasi simili ».

Pag. 109, riga penultima: Invece di « parti », leggasi « parti uguali ».

Pag. 111, righe 20 e 21: Invece di « l'elemento », leggasi « come primo elemento ».

Pag. 118, righe 6 e 10: Invece di « 68,08 », leggasi « 68,08 per cento ».

Pag. 121, formula 39: Invece di « r » leggasi « s ».

Pag. 137, nota (1): Aggiungere « Ove manchi a_0, mancherà la prima equazione (62), e così dicasi per gli altri coefficienti ».

Pag. 189, riga 21: Invece di « ai coefficienti. », leggasi « ai coefficienti, con qualche variante, ove manchi la costante additiva ».

Pag. 196, riga 7: Invece di « con », leggasi « per ».

Pag. 199, righe 18 e 19: Invece di « criminalità, analfabetismo, morbilità, mortalità », leggasi « criminalità, morbilità ».

[illegible]

INTRODUZIONE

Capitolo I.

NOZIONI PRELIMINARI

1. Le diverse scuole statistiche. – Nella metodologia statistica non esistono altre scuole che quelle di coloro che si fermano per istrada e quella di coloro che percorrono l'intero cammino.

È naturale che le prime a loro volta si distinguano secondo il punto di arresto e pertanto, ad es. nel campo degli studî sociali, abbiamo la cosidetta scuola amministrativa, che limita il compito dello statistico all'apprestamento di questionarî per la rilevazione di fatti interessanti la pubblica amministrazione, allo spoglio di essi e alla redazione di bollettini e annuarî; la scuola degli indici o dei valori segnalatori, che aggiunge al precedente bagaglio formule di valori medî, ecc.; la scuola, che impiega altresì schemi teorici e metodi interpolatorî, sostenendo peraltro che la giustificazione di essi non è compito dello statistico; e così via sino ad arrivare alla scuola di chi abbraccia tutto il campo della metodologia statistica.

Si vorrebbe da taluni opporre una statistica matematica ad una statistica non matematica; ma anche questa distinzione rientra nel quadro sopra tracciato, perchè il metodo statistico, fondandosi in generale su dati quantitativi o che si possano ricondurre a forma quantitativa, è per sua natura matematico, e la sola distinzione lecita sarebbe quella tra procedimenti che presuppongono nozioni di matematica elementare, e procedimenti che si fondano invece su nozioni di matematiche superiori.

Poichè questo manuale si rivolge ai cultori di scienze sociali e non può presumer note che le nozioni di algebra e di geometria analitica, che formano il patrimonio matematico delle persone di cultura generale, non abbiamo mancato di rispettare quest'ultima distinzione.

2. L' indagine scientifica e la metodologia statistica. – Dicendo che il metodo statistico è per sua natura matematico, non abbiamo inteso punto definirlo; per spiegare con qualche precisione che cosa veramente sia, non basta segnalare il mezzo più proprio di cui esso si avvale, ma occorre metterne in luce le caratteristiche fondamentali.

Anche chi non abbia familiarità coi problemi scientifici conosce che l'analisi di alcuni fenomeni — pure nel caso che sia possibile e lecito l'esperimento — presenta speciali difficoltà, derivanti dalla molteplicità di altri fenomeni (accadimenti, condizioni, circostanze, ecc.), con cui quelli sono in relazione.

Ciò, infatti, rende particolarmente ardua la scoperta di tali relazioni e, quand'anche esse siano note, la determinazione della forma in cui si presentano.

In tali casi l'analisi scientifica suole essere condotta o in modo astratto o su una grande massa di osservazioni — sperimentali o non — relative ai fenomeni considerati.

Il vantaggio del primo modo di procedere sta nel fatto che, conosciuta o ammessa per ipotesi l'esistenza di una relazione, esso permette di isolarla mentalmente in guisa da poterne accertare il comportamento; d'altra parte il secondo modo offre il vantaggio che in grandi masse di osservazioni il numero di quelle relazioni si presenta spesso fortemente ridotto: volgarmente si dice che nella massa molte influenze si elidono.

Ognuno di tali procedimenti ha i suoi pregi e i suoi difetti; ond'è che in molti campi di ricerche ormai s'impiega l'uno e l'altro simultaneamente: tra le scienze sociali è questo il caso, tipico, dell'economia politica, dove peraltro i due procedimenti hanno potuto avere così largo impiego da far persino riconoscere l'esistenza di un'economia politica deduttiva e di un'economia politica induttiva.

E mentre da una parte, sulle tracce di Walras e di Pareto (¹), si studiano ad es. le condizioni di equilibrio di un mercato in regime di perfetta concorrenza e nell'ipotesi di gusti costanti, di un numero assegnato di servizî produttori, di un dato stato della tecnica, ecc.; d'altra parte si rilevano ad es. i legami intercedenti tra l'andamento nel tempo del livello generale dei prezzi e dell'attività produttiva.

È pure il caso della demografia che, sorta come studio di masse di dati riguardanti lo stato e il movimento della popolazione, come ricerca empirica dei fattori di equilibrio dei sessi, dell'influenza delle condizioni economiche sui matrimonî, e così via, ha dato luogo ad una teoria che si avvale largamente dell'analisi matematica, partendo da ipotesi relativamente semplici: una schiera di studiosi, da Bortkewitsch a Lotka (²), si è proposta ad es. di determinare la composizione per età, che una popolazione tenderebbe a raggiungere stabilmente, se non fosse soggetta a

(¹) Vilfredo Pareto, *Manuale di economia politica* con una introduzione alla scienza sociale. Milano, Società Editrice Libraria, 1909.

(²) Louis I. Dublin and Alfred J. Lotka, *On the true rate of natural increase* as exemplified by the population of the United States, 1920, in « Journal of the American Statistical Association », 1925. Cfr. anche: Robert R. Kuczynski, *Fertility and reproduction*. New York, Falcon Press, 1932. – Sugli argomenti trattati nel testo il lettore troverà alcuni chiarimenti nell'Appendice *A*) II a questa Parte I.

movimenti migratorî e se per ogni classe di età si mantenesse costante nel tempo una data frequenza delle morti e delle nascite.

Quel secondo modo di procedere (per grandi masse), sebbene presenti speciali esigenze secondo la natura dei fenomeni considerati e non possa in ultimo non sboccare nel primo, ha fatto sorgere un insieme di criterî di generale e frequente applicazione. La giustificazione e la coordinazione logica di tali criterî e dei metodi, che ne discendono, formano l'oggetto della metodologia statistica.

Lo scopo di questo volume, però, non consente una trattazione così generale e ci consiglia di esporre quella parte della metodologia statistica, che trova impiego nelle scienze sociali e in quegli argomenti (in gran parte di antropometria), che hanno con esse stretta attinenza o particolari legami.

3. La prima categoria di problemi, di cui consta la metodologia statistica. – In due grandi categorie ci sembra utile raccogliere quei criterî e quei metodi. La prima, presupponendo note solo masse di osservazioni di *un dato fenomeno*, mira a determinare la natura generica dei fenomeni influenti e talvolta il senso e l'intensità d'azione di qualcuno di essi. Classificando le osservazioni del fenomeno in ordine di grandezza, rilevando la forma a cui tende la distribuzione di tali osservazioni al crescere del numero di esse ed esaminando per via di confronti se tale forma sia eventualmente quella a cui tende univocamente qualche noto concerto schematico di altri fenomeni, si riesce ad assimilare il fenomeno studiato a fenomeni della stessa natura di quelli predisposti nello schema e persino a scomporre la distribuzione in relazione all'influenza di qualche speciale gruppo di fenomeni.

Ad es., disponendo delle stature di un gruppo di maschi coetanei, noi possiamo classificarle in ordine di grandezza, e così costruire la cosidetta *distribuzione statistica*, la quale ci mostrerà che, al crescere delle stature, il numero degl'individui che le presentano in generale dapprima cresce sino a raggiungere un massimo e poi decresce sino ad annullarsi: rivelerà, cioè, un addensamento di individui attorno ad una statura tipica ed un diradamento di essi al disopra e al disotto di tale statura.

Ecco un esempio (tav. I), tratto da una statistica inglese [1], in cui le stature sono state raccolte per brevità in classi di due pollici. Il numero di individui presenta in alcuni gruppi una cifra decimale uguale a 5 perchè è stato diviso in due parti uguali e attribuito a due classi consecutive ogni dato corrispondente al limite comune delle due classi.

Adesso prendiamo un'urna contenente metà palle bianche e metà palle nere, tali che praticamente differiscano fra loro solo nel colore; se

(1) Karl Pearson and Alice Lee, nella rivista « Biometrica », II, 1903, pag. 415.

da essa eseguiamo ad es. una serie di 100 estrazioni, rimettendo dopo ogni estrazione la palla nell'urna per mantenerne inalterata la composizione, otterremo un certo numero di palle bianche su 100. Torniamo a fare un'altra serie di 100 estrazioni, e poi un'altra sino a raggiungere un totale di 1.078 serie di 100 estrazioni ciascuna.

TAVOLA I.

Stature in pollici	Numero di individui
59,5 - 61,5	3,5
61,5 - 63,5	24
63,5 - 65,5	100
65,5 - 67,5	237,5
67,5 - 69,5	323
69,5 - 71,5	236
71,5 - 73,5	105
73,5 - 75,5	37,5
75,5 - 77,5	8
77,5 - 79,5	3,5
Totale	1.078

TAVOLA II.

Palle bianche estratte	Numero delle serie di estrazioni
26 - 30	1
31 - 35	4
36 - 40	63
41 - 45	101
46 - 50	352
51 - 55	324
56 - 60	112
61 - 65	84
66 - 70	34
71 - 75	3
Totale	1.078

Tra queste serie determiniamo il numero di quelle che presentano la minore proporzione di palle bianche e poi il numero di quelle che presentano proporzioni via via più grandi. Tale distribuzione mostrerà un addensamento delle 1.078 serie attorno ad una proporzione tipica di palle bianche ed un diradamento di esse al disopra e al disotto di tale proporzione.

Con opportuni criterî, tendenti a rendere omogenea cotesta variabile (percentuale di palle bianche) con la precedente (statura in pollici), supponiamo di aver potuto raccogliere i risultati delle nostre estrazioni in classi confrontabili con le precedenti e di avere ottenuto la tav. II. Se si può dimostrare che, malgrado la diversa natura ed unità di misura delle variabili, le forme di tali distribuzioni tenderebbero a coincidere in un gran numero di osservazioni, e che d'altra parte nessun altro insieme di fenomeni condurrebbe alla forma, a cui hanno dato luogo i fenomeni influenti sulle nostre estrazioni, noi potremo concludere che sulle stature dei nostri maschi coetanei hanno influito fenomeni della stessa natura di quelli predisposti nello schema dell'urna, e saremo sulla buona via per l'identificazione di tali influenze e di quel grado di purezza etnica del gruppo considerato, che non meno dell'antropologo interessa il demografo.

Altri schemi teorici potrebbero escogitarsi per altri fenomeni ed in generale può essere utile uno studio sistematico di siffatti schemi. Esso è stato già in gran parte compiuto ad opera di Bernoulli, Gauss, Quetelet, Poisson, Lexis, Pearson, Boltzmann, Edgeworth, ecc., e forma la prima di quelle due categorie di criterî e di metodi, in cui abbiamo distinto la metodologia statistica.

4. La seconda categoria di problemi. – La seconda categoria, presupponendo note invece masse di osservazioni relative a *più fenomeni*, si propone di determinare se, in qual forma e, in alcuni casi particolari, con quale intensità tali fenomeni siano in relazione; e ciò in base all'analisi delle variazioni che presentano i valori corrispondenti delle osservazioni.

Riprendendo l'esempio precedente, poichè per ognuno di quei 1.078 maschi coetanei si conosce, non solo la sua statura, ma anche quella del padre, possiamo esaminare come variano, al crescere delle stature dei nostri coetanei, le stature dei loro padri.

Ecco i dati (tav. III), ricavati dalla menzionata statistica inglese, ed in cui i decimali delle caselle risultano dal fatto che ogni dato (coppia di stature) corrispondente ai limiti comuni delle classi è stato attribuito in parti eguali alle due o alle quattro caselle adiacenti.

Tavola III.

Stature dei figli in pollici	Stature dei padri, in pollici									
	da 57,5 a 59,5	da 59,5 a 61,5	da 61,5 a 63,5	da 63,5 a 65,5	da 65,5 a 67,5	da 67,5 a 69,5	da 69,5 a 71,5	da 71,5 a 73,5	da 73,5 a 75,5	Totali
59,5 - 61,5	—	—	1	1,5	1	—	—	—	—	3,5
61,5 - 63,5	—	1	5	7,25	8,5	1,75	0,5	—	—	24
63,5 - 65,5	3	3	12	35,25	30,5	11,75	4,5	—	—	100
65,5 - 67,5	—	5	15,75	53	86,5	53,25	21,5	2,5	—	237,5
67,5 - 69,5	—	2,5	14,75	38	90	106	53,25	17,5	1	323
69,5 - 71,5	—	—	1	16,5	50	75,75	65,5	22,75	4,5	236
71,5 - 73,5	—	—	—	4	11,5	32,5	34,75	20,75	1,5	105
73,5 - 75,5	—	—	1	1,5	1,5	10,5	11,5	9	2,5	37,5
75,5 - 77,5	—	—	—	—	—	3	1,5	3,5	—	8
77,5 - 79,5	—	—	—	—	—	1	1	1,5	—	3,5
Totali	3	11,5	50,5	157	279,5	295,5	194	77,5	9,5	1.078

Costruiremo le cosidette *serie statistiche* a due variabili o, più sinteticamente, come nell'esempio, quelle *tavole a doppia entrata* che, con opportune elaborazioni e in questo caso quasi a primo sguardo, permettono di giudicare dell'esistenza e della forma della relazione intercedente tra stature dei padri e stature dei figli, e di contribuire a quello studio della ereditarietà della statura umana e, in generale, dei caratteri antropologici, che ha un'importanza fondamentale per il cultore di scienze sociali.

La tavola III già rivela che nel gruppo esaminato i figli di bassa statura provenivano da padri in cui prevalevano le basse stature, i figli di alta statura da padri in cui prevalevano le alte stature.

Potremmo ancora procedere alla costruzione di una tavola a tripla, quadrupla.... entrata, se, ad es., di ognuno dei nostri coetanei si conoscesse anche il reddito familiare, ecc., e si volesse pure esaminare l'influenza che sulla statura esercitano le condizioni economiche, e così via.

La tavola IV contiene il numero regionale degli addetti alle nostre industrie al 15 ottobre 1927 (data dell'ultimo censimento industriale e commerciale) e il numero dei cavalli-vapore impiegati, in corrispondenza al gettito regionale, disposto in ordine crescente, delle tre principali imposte dirette nel 1927. Essa mostra ben chiara la solidarietà d'andamento delle tre successioni e — anche senza alcuna elaborazione statistica — fa pensare ad un alto grado di « correlazione »: molti operai, molte macchine e molta ricchezza si rivelano già tre fatti conciliabili, anzi strettamente collegati fra loro.

L'influenza di alcuni fenomeni potrà essere preliminarmente eliminata a mezzo di operazioni elementari (generalmente col calcolo di quozienti) o di

TAVOLA IV.

Gettito delle tre imposte dirette nel 1927 (milioni di lire)	Numero degli addetti alle industrie nel 1927 (migliaia)	Forza motrice delle industrie nel 1927 (migliaia di H. P.)
21,8	19,4	14,3
41,3	60,1	99,3
53,2	48,6	304,4
59,2	54,2	273,5
60,1	84,7	112,5
68,0	66,5	105,3
116,8	80,4	93,4
134,2	113,3	169,8
175,6	132,2	97,4
244,8	212,9	246,1
282,1	221,7	695,0
298,7	225,7	528,2
334,9	328,2	635,3
355,2	179,1	349,9
398,1	331,2	495,7
471,0	249,0	442,3
635,7	549,0	1.841,7
1.190,5	1.046,6	2.074,0

procedimenti meno semplici: possiamo ad es. togliere le diverse quote regionali di evasione dal gettito delle imposte, riportato nella tavola, od anche dividere i dati delle tre successioni per le popolazioni regionali rispettive.

Sorgerà il problema di determinare la forma analitica e in certi casi una misura sintetica delle relazioni accertate. Talvolta uno dei fenomeni osservati è il tempo od un attributo qualitativo suscettibile di due o più modalità, e si avranno le cosidette serie, o tavole a più entrate, storiche o qualitative.

Ci si accorge subito che questa seconda categoria di criterî e di metodi offre un campo di applicazione ben più grande della prima.

5. Come nelle applicazioni una precisa distinzione tra quelle due categorie non sia sempre possibile. – Occorre, però, notare che nelle applicazioni una precisa distinzione tra quelle due categorie non è sempre possibile. Infatti, quando si applica la prima, la scelta e l'interpretazione degli schemi suole essere agevolata da qualche nozione, che già si possegga sulle relazioni, in cui il fenomeno esaminato sta con altri fenomeni. Ed accade persino di vedere applicato uno schema teorico a gruppi di osservazioni, dai quali sia stata già eliminata l'influenza di qualche fenomeno ben conosciuto.

È noto, ad es., che le nascite variano da un anno all'altro sotto l'influenza di molte circostanze: movimenti migratorî, buoni o cattivi raccolti, epidemie, guerre, ecc. Ma, se vogliamo esaminare la cosidetta mascolinità delle nascite in tempi o luoghi diversi, converrà calcolare per ogni gruppo di nati vivi la percentuale in cui il numero dei nati di sesso maschile sta a quello corrispondente dei nati di sesso femminile e costruire una distribuzione statistica di tali percentuali, sulle quali influirà solo quel minor numero di circostanze, che affetta la proporzione dei sessi alla nascita. Si osserverà, tra l'altro, quel singolare addensamento di casi intorno alla proporzione di circa 106 maschi per 100 femmine, che fu una delle prime e più grandi «meraviglie» della statistica e che nel 1741 Süssmilch — uno dei precursori della moderna demografia — additò persino come opera della Divina Provvidenza, evidentemente contraria alla poligamia (1)!

D'altra parte, quando si applica la seconda categoria di metodi statistici, suole avvenire che per la molteplicità stessa dei fenomeni influenti,

(1) J. P. Süssmilch, *Die göttliche ordnung in den veränderungen des menschlichen geschlechts aus der geburt, dem tode und der fortpflanzung desselben erwiesen*. Berlin, 1775-76 (4a edizione).

Sulla storia della statistica può sempre consultarsi con vantaggio: Antonio Gabaglio, *Teoria generale della statistica*, Milano. Hoepli, 1888 (Vol. I: Parte storica). I più recenti contributi sono quelli di Harald Westergaard, *Contributions to the history of statistics*. London, P. S. King & Son, 1932.

oltre a quelli che eventualmente si riesca ad accertare, ne esistono altri ignoti, dei quali conviene almeno conoscere la natura generica. Ad es. nello studio della dinamica dei prezzi molti ritengono di avere isolato le fluttuazioni stagionali dalle tendenze generali e dalle fluttuazioni di lungo ciclo, se possono infine dimostrare che le oscillazioni residue obbediscono allo schema dei fenomeni casuali.

6. La rappresentazione dei dati. – La trattazione dei problemi esposti presuppone la conoscenza di modi acconci per la rappresentazione dei dati e precisamente per la rappresentazione numerica e grafica di essi (distribuzioni e serie statistiche, tavole a più entrate, diagrammi, cartogrammi, stereogrammi), per la rappresentazione mediante valori caratteristici, medie di valori, di scarti, di differenze, ecc.

In base, poi, al concetto di funzione, sorge il problema di rappresentare distribuzioni, serie statistiche e tavole a più entrate mediante funzioni analitiche opportunamente scelte, di determinare i valori più convenienti dei coefficienti di coteste funzioni, di giudicare della bontà dell'adattamento di esse ai dati osservati.

Tali rappresentazioni si dimostreranno in gran parte arbitrarie, ma offriranno un così grande ausilio nella trattazione di quei problemi scientifici, che forse non si riuscirà nemmeno a comprendere per qual altra via essa potrebbe fruttuosamente condursi.

È questo un concetto, che bisogna avere ben chiaro nello studio della metodologia statistica per non cadere in erronei apprezzamenti: i problemi, di cui essa si occupa, sono essenzialmente scientifici, in quanto mirano a fornire dei mezzi per la ricerca delle relazioni tra i fenomeni; peraltro essa si avvale di procedimenti arbitrarî, giustificati dalla loro fecondità nella risoluzione di quei problemi e suscettibili di essere abbandonati, se altri procedimenti si rivelassero ancor più fecondi a quei fini scientifici.

7. Partizione della materia. – Da quanto abbiamo esposto riteniamo giustificata la seguente partizione della materia:

1°) I modi di rappresentare i gruppi di osservazioni.

2°) L'analisi di un gruppo di osservazioni.

3°) Le relazioni tra gruppi di osservazioni.

A queste tre parti premettiamo un capitolo sul metodo statistico dall'aspetto logico, che, offrendo una veduta d'insieme del campo dei nostri studî, delle seduzioni che offre e delle insidie che tende, renderà meno disagevole il cammino che dovremo percorrere e farà ancor più apprezzare i risultati con esso raggiunti.

CAPITOLO II.

IL METODO STATISTICO DALL'ASPETTO LOGICO

1. IL PROCESSO LOGICO DELL'ANALISI STATISTICA. – Un'analisi statistica compiuta, sia essa condotta con la prima o la seconda categoria di metodi esposti nel precedente capitolo, dal punto di vista logico consta ugualmente di un processo di rilevazione, di classificazione, di confronto, di semplificazione e di generalizzazione.

Per offrire una nozione ancora più chiara della nostra disciplina, crediamo opportuno dare in forma elementare ed accessibile anche ai profani un'idea generale di cotesti stadî e degli errori più comuni e grossolani, che in essi si sogliono commettere (1).

2. FORME E LIMITI DELLE RILEVAZIONI. ERRORI DI RILEVAZIONE. – La rilevazione dei dati può assumere diverse forme: può essere continua e saltuaria, diretta e indiretta o congetturale, privata e pubblica, individuale e collettiva, ecc., e può eseguirsi (con e senza strumenti) a mezzo di registri, ruoli, questionarî, moduli. Ad es., i censimenti demografici, industriali, del bestiame sono rilevazioni saltuarie, dirette, pubbliche, collettive, fatte a mezzo di questionarî, che nei censimenti demografici sono le cosidette schede di famiglia; i valori del commercio internazionale di uno Stato, i dati dei nati, dei morti e dei matrimonî, la maggior parte delle quotazioni dei prezzi delle merci, si ottengono mediante rilevazioni continue, dirette, pubbliche, eseguite con registri e moduli; la ricchezza d'Italia dà luogo ad una rilevazione saltuaria, congetturale, privata, individuale, eseguita in parte a mezzo di questionari; la maggior parte dei dati biologici, antropologici, psicologici sono rilevati in modo saltuario, diretto, privato, individuale, mediante speciali strumenti e l'uso frequente di questionarî. E così via (2).

(1) Su questo argomento vedansi le classiche pagine di ADOLPHE QUETELET, *Lettres sur la théorie des probabilités.* Bruxelles, 1846; di JOHN STUART MILL, *A system of logic,* London, 1872; di ANGELO MESSEDAGLIA, *Prelezioni,* in « Biblioteca dell'economista ». Torino, Utet, Serie V, Vol. XIX. Ce ne siamo largamente occupati in uno studio: *Le illusioni statistiche,* pubblicato nel « Giornale degli economisti » del 1925. — Sulla « storia ideale eterna dell'errore », cfr. BENEDETTO CROCE, *Logica.* Bari, Laterza.

(2) Nell'Appendice *A*) I, alla Parte I di questo manuale esponiamo, a titolo di esempio, i principali criterî di rilevazione adottati nel nostro settimo censimento demografico.

In ogni caso una rilevazione presuppone la determinazione esatta del fenomeno o dei fenomeni, che si vogliono rilevare, e dei limiti di tempo e di spazio, entro i quali si ritiene opportuno eseguire la rilevazione; l'assegnazione ad essa di un dato grado di precisione e la conoscenza degli errori sistematici degli strumenti eventualmente usati e dell'organo o degli organi di rilevazione; una sufficiente capacità, preparazione, obiettività, diligenza di questi ultimi.

È chiaro che, se non si determina con precisione il *significato*, che s'intende attribuire alle parole « nato-morto » o « aborto » o « disoccupato », ecc. (poco importa che esso sia o non sia il significato giuridico), e il *tempo* e il *luogo* in cui la rilevazione ha da compiersi, i risultati si presteranno ad interpretazioni equivoche e non condurranno ad analisi feconde.

A seconda, poi, degli scopi, ai quali è rivolta l'indagine, le rilevazioni possono assumere tutti i gradi di *precisione*, che vanno da una stima grossolana alla determinazione più esatta consentita dagli eventuali strumenti di misura e dalle caratteristiche degli osservatori. A questo riguardo è da rilevare che — assegnato un dato grado di precisione alla ricerca — l'*imperfezione degli strumenti di misura* dà luogo generalmente ad errori sistematici, per l'eliminazione dei quali si sono escogitati speciali procedimenti, naturalmente variabili da caso a caso; e così, ad es., in antropometria si parla di errori del quadro a massima, ancora largamente usato nella misura dei diametri del capo invece del costoso compasso di spessore, e si cerca di eliminarli considerando che quel quadro tende a dare al capo una forma un po' più rotondeggiante. È, inoltre, da rilevare che l'*imperfezione dei nostri sensi* fa variare dall'uno all'altro osservatore l'intensità o le modalità di una sensazione; donde la necessità — se si vuol raggiungere un notevole grado di precisione — di determinare per ognuno di essi l'errore sistematico in cui suole cadere, ossia la cosidetta « equazione personale ».

Quanto agli errori prodotti da *incapacità intellettuale* o da *ignoranza*, i più insidiosi sono certo quelli, in cui s'incorre quando la rilevazione non viene eseguita direttamente da coloro che ne debbono usare, ma a mezzo di altre persone. È questo, ad es., il caso dei censimenti e delle grandi inchieste, in cui occorre fidarsi della capacità dei denuncianti e degli organi raccoglitori. Ed è notorio che — quantunque si cerchi sempre di contenere quelle rilevazioni nei limiti, e di dare ad esse la forma, consentiti dal livello intellettuale e culturale dell'ambiente in cui si compiono — pure nei censimenti demografici ad es., quando si chiede l'età, si dichiarano di preferenza età rotonde; quando si chiede l'anno di nascita, si dichiarano spesso erroneamente anni memorabili, come ad es. per l'Italia il 1848 o il 1859 o il 1866; e, quando si chiede l'una e l'altra, le dichiarazioni frequentemente discordano.

All'esatta rilevazione dei dati sogliono pure essere di ostacolo i *preconcetti specifici*, che su di essi si abbiano; preconcetti, che spesso sono così tenaci, da far addirittura chiudere gli occhi — come si dice — davanti alla realtà, la quale, o non si vede, o si vede in modo conforme a quei preconcetti.

È vero: « Is tamen humano intellectui error est proprius et perpetuus, ut magis moveatur et excitetur affermativis quam negativis »; (¹) ma spesso si è così imbevuti dell'opinione preconcetta, che si assumono solo i fatti che depongono a favore di essa, trascurando sommariamente il resto (²).

Ci racconta Quetelet che i fisiologi dei primi anni del secolo scorso ritenevano che il polso dei vecchi fosse meno frequente di quello dei giovani, ed avevano creduto di trovare nei dati statistici una precisa conferma di questa opinione. Successivamente, però, più corrette rilevazioni mostrarono il contrario e precisamente che, mentre le dimensioni del corpo umano crescono sino alla maturità e diminuiscono nella vecchiaia, il numero delle pulsazioni per unità di tempo diminuisce nella maturità e cresce nella vecchiaia.

Bene spesso, però, il preconcetto è alimentato dall'*interesse*, il quale di per sè è un'altra causa frequentissima di errate rilevazioni, specie quando queste si eseguano per via di denuncie o, comunque, su individui interessati ad alterare la verità. È l'interesse a far ritardare di un anno il servizio militare, che in molti paesi spinge i genitori a denunciare nel gennaio successivo la nascita dei loro figli maschi, avvenuta nel dicembre. Sono interessi quelli che per lo più consigliano alcune famiglie a fare apparire pei loro cari — malgrado i più rigorosi controlli — cause di morte diverse dalle reali, quando queste siano la tubercolosi, la sifilide, l'alcoolismo, ecc.; o a non dichiarare nei censimenti l'idiozia di qualche loro membro o a dichiarare legittima un'unione illegittima. È l'interesse a non dare la prestazione dovuta, che spinge i contribuenti ad occultare i loro beni al fisco o qualche giovane ventenne ad alterare i movimenti respiratori o ad incurvare la schiena nelle visite militari.

Non ci dilungheremo, poi, sugli errori, che si possono commettere per *disattenzione* o *negligenza*, perchè ognuno di noi ne fa esperienza quotidiana: sono equivoci, errori di interpretazione, di lettura, di trascrizione, di computo, rilevazioni di cause o di effetti inesistenti, ecc.

È singolarissimo, per la sua enormità, il caso di Saraw — riferito

---

(¹) FRANCISCI DE VERULAMIO (Francesco Bacone), *Instauratio magna*. Londini, 1620. Pars secunda, quae dicitur « Novum Organum » Aphorismus XLVI. Cfr. *The works of Francis Bacon*, London, Longman, 1864, Vol. I.

(²) Sulla distinzione tra preconcetti generici e specifici e su alcune opinioni manifestate su questo argomento da H. POINCARÉ in *La science et l'hypothèse* (Paris, Flammarion, 1923, Capo IX), rimandiamo al nostro studio citato: *Le illusioni statistiche*.

da Quetelet — , che pretendeva di avere rilevato nell'isola di Santa Croce (Antille Danesi) un bassissimo ed inverosimile rapporto tra il numero annuale di negri schiavi morti e la media annuale dei viventi. Tale risultato, ottenuto in piena buona fede, derivava semplicemente dal fatto che i bambini negri, morti in età inferiore ad un anno e che — come in tutti i paesi del mondo — davano un alto contributo alla morte, erano esenti dalla capitazione e non venivano inscritti nei registri demografici

Analogamente sino al 1842 la mortalità nel Belgio si riteneva singolarmente elevata, senza riflettere che nelle statistiche dei morti ogni Comune comprendeva indistintamente tanto i nati fuori del Comune, che in esso morivano, quanto i nati nel Comune, che morivano fuori di esso.

Infine, quanto agli errori di rilevazione cosidetti *accidentali* o *casuali*, la natura, le forme ed il comportamento di essi sono stati sviscerati a tal grado, che ormai se ne possiede una teoria, che per compiutezza e perfezione può considerarsi come una delle più alte conquiste del pensiero. Sono ormai nozioni comuni che in ogni misura si commette sempre un grande numero di piccoli errori, indipendenti fra loro e varianti senza una legge rigorosamente assegnabile; che, per l'influenza esclusiva di siffatti errori, le misure, ripetute in gran numero, tendono a distribuirsi simmetricamente seguendo una certa curva, detta di Gauss in omaggio a chi, primo, ne verificò la validità; e che la media aritmetica di quelle misure tende, sotto certe condizioni, alla vera misura del fenomeno osservato.

Avremo occasione, nella Parte II, di ritornare su questo argomento.

3. Le classificazioni dei dati. Esempi di classificazioni errate. – Abbiamo notato nel precedente capitolo l'utilità di una classificazione dei dati statistici ed abbiamo riportato, tra l'altro, l'esempio della classificazione di un gruppo di stature (tavola I).

Una classificazione di dati statistici — come una qualsiasi altra classificazione — può certamente eseguirsi nel modo più arbitrario; ma ciò non significa che la classificazione scelta non debba avere un significato, che non si debba ben conoscere la natura e la portata di essa e che sia lecito finir col seguire una classificazione diversa da quella prescelta.

D'altra parte, assegnato un dato scopo alla rilevazione, la scelta del criterio di classificazione non è più arbitraria, ma viene limitata dall'idoneità di questa a raggiungere lo scopo. Nè una classificazione, utile per un dato scopo, è sempre idonea a far raggiungere uno scopo diverso.

Invece spesso accade di assistere alle classificazioni più inverosimili di dati statistici, suggerite per lo più da mera ignoranza o negligenza.

È rimasta famosa, per la gravità degli equivoci a cui diede luogo, la classificazione che per lungo tempo fu seguita dal Ministero della Guerra in Francia, relativamente alla statura dei giovani chiamati alle armi (1). Quetelet aveva notato che le stature di un gruppo di coetanei di ugual sesso tendono, in certe condizioni, a distribuirsi secondo un certo schema teorico, cosicchè — come abbiamo visto nel precedente capitolo — vi sarà, fra l'altro, una statura od una classe di stature che sarà presentata da un numero prevalente di individui, mentre gl'individui con statura più alta o più bassa si disporranno simmetricamente attorno a quella statura o a quel gruppo di stature, e tanto più diminuiranno di numero, quanto più ne saranno discosti.

Ora Bertillon padre, studiando le stature dei giovani ventenni nati nel dipartimento del Doubs, notò in esse, invece di un solo gruppo centrale più numeroso, due gruppi separati da una ben marcata depressione. E senz'altro attribuì quella forma singolare ad una sovrapposizione di due distribuzioni, di cui quei massimi laterali fossero i vertici; spiegando tale sovrapposizione col fatto che, nel dipartimento considerato, alla razza celtica — la più antica abitatrice della regione e con statura tipica più bassa — si sarebbe sovrapposta la razza cimbrica, immigratavi in epoche relativamente più recenti e con statura tipica più alta.

Tale spiegazione fu condivisa da Lagneau, che credette di poterla convalidare con abbondanti prove storiche, e sembrò acquisita alla scienza. Invece si trattava di un errore, in cui si era caduti nella classificazione di quelle stature. Per dirla in breve, si erano raggruppati gl'individui in classi aventi un intervallo costante di un pollice; ma, poichè la rilevazione era stata fatta originariamente in centimetri, nell'eseguire la riduzione dei centimetri in pollici — che era l'unità di misura ancora seguita per certi scopi pratici — era accaduto che nella classe centrale, corrispondente all'intervallo da 5 piedi e 5 pollici a 5 piedi e 6 pollici si erano compresi coloro che presentavano solo le due misure di 166 e 167 cm.; mentre nelle classi precedenti e seguenti si erano compresi individui che presentavano in centimetri non due, ma tre misure diverse. Era, quindi, naturale che la classe centrale, di un terzo meno estesa delle due contigue, presentasse quella marcata depressione. Si era, in sostanza, seguita una classificazione in pollici ad intervalli uguali, nell'illusione che essa corrispondesse ad un'analoga classificazione in centimetri.

Quando il nostro antropologo Ridolfo Livi mise in luce, dapprima l'inverosimiglianza teorica della spiegazione di Bertillon, e poscia lo svarione in cui si era caduti, la depressione sparì nè si parlò più di celti e cimbri.

(1) Ridolfo Livi, *Antropometria*. Milano, Hoepli, 1900; e *Antropometria nei suoi rapporti con la medicina sociale*. Milano, Vallardi, 1907.

L'errore esposto aveva almeno l'attenuante dell'impossibilità di fare esattamente corrispondere, nella formazione delle classi, i centimetri ai pollici; ma accade spesso di vedere pure assunte classificazioni ad intervalli esplicitamente disuguali come aventi intervalli uguali e di assistere alle più strane costruzioni geometriche ed alle più inaspettate descrizioni di « gobbe » e di « piramidi interrotte ».

Altre volte le classificazioni si eseguono in guisa che alcune classi comprendono ambedue i valori limiti — come nella tavola I —, altre solo i limiti inferiori o solo quelli superiori, altre ancora escludono ambedue i limiti, altre infine non presentano un criterio univoco per tutti i dati che vi sono compresi.

Quest'ultimo errore dà luogo ad un'altra categoria di classificazioni errate: quelle equivoche, le quali si presentano frequentemente nel caso di attributi qualitativi.

Non intendiamo alludere alla maggiore o minore nè alla variabile o costante ampiezza delle classi adottate, perchè la scelta dell'ampiezza di esse è in relazione alla precisione ed allo scopo che si vuole raggiungere con l'indagine. Ond'è che non sapremmo a rigore muover rimprovero a chi, fondandosi ad es. su un censimento industriale o professionale, raccogliesse in poche classi alcune industrie o professioni caratteristiche, collocando le rimanenti in un'unica classe genericamente denominata « altre industrie » o « professioni non specificate ». Potremmo solo rilevare l'insufficienza di essa a far raggiungere alcuni scopi, come quello di fornire una nozione soddisfacente della composizione qualitativa delle industrie o professioni rilevate. Ma quando quell'ultima classe risulta piuttosto numerosa ed in essa sono stati compresi i casi ignoti e le industrie o le professioni, che solo in parte presentavano le caratteristiche delle classi precedenti, abbiamo ragione di affermare che quella classificazione è errata, perchè equivoca.

Equivoche sogliono pure risultare le classificazioni fondate sulle cause del fenomeno esaminato. Se è vero che la causa di un fenomeno è l'insieme degli altri fenomeni influenti sull'avveramento di esso, si comprende facilmente quanto sia difficile che sull'avveramento di due o più fenomeni influisca esattamente, in qualità e misura, lo stesso insieme di altri fenomeni; e quanti arbitri si possano commettere nel considerare la diversità di tali fenomeni influenti come elemento differenziale di una classificazione. Concepisce il lettore la possibilità di una corretta classificazione dei fallimenti o delle separazioni coniugali, secondo le cause determinanti? È questo motivo intrinseco, che vizia le statistiche delle cause di morte, ormai fornite da tutti gli Stati progrediti, e rende molti studiosi così scettici sui risultati di esse, malgrado il progresso delle scienze mediche ed i perfezionamenti che a quella classificazione si sono apportate col cosidetto elenco nosologico internazionale e continuamente si apportano ad opera dell'Istituto Internazionale di Statistica (¹).

Occorre, infine, ben guardarsi dalle interpretazioni arbitrarie delle classificazioni: si è, ad es., erroneamente ritenuto che le distribuzioni per paesi di destinazione delle merci esportate da un dato paese, fornite dalla statistica doganale in base alle dichiarazioni dei commercianti, corrispondano alle distribuzioni effettive; e che le ripartizioni regionali delle sentenze pronunciate dai giudici di un dato grado in cause penali o civili corrispondano alla ripartizione regionale dei reati o della materia del contendere.

4. I CONFRONTI E L'OMOGENEITÀ DEI DATI. – Le due categorie di criterî e di metodi, di cui ci siamo occupati nel precedente capitolo, si fondano essenzialmente su confronti.

Suol dirsi comunemente che i confronti debbono essere istituiti tra elementi omogenei; ma dobbiamo bene intenderci sul significato di questo precetto.

Ad es., è nozione ormai comune che la variazione dal tempo 0 al tempo 1 del livello generale dei prezzi delle merci in un dato luogo viene per lo più misurata, calcolando una media dei rapporti (numeri indici) tra i prezzi del tempo 1 e i prezzi del tempo 0.

Per l'impossibilità di considerare i prezzi di tutte le merci, si scelgono i prezzi di quelle merci, che intervengono in sì grande proporzione nel valore totale degli scambî o sono comunque così rappresentative, da rendere plausibile l'ipotesi che l'importanza delle merci rimanenti sia trascurabile rispetto a quella delle prime. Ne consegue che tale scelta è subordinata alle condizioni economiche della collettività, di cui si tratta, e che sarebbe erroneo, per due collettività a diversa struttura economica, confrontare medie di rapporti, che considerassero le medesime merci (²).

Tali confronti assumono infatti quegl'indici, non come meri calcoli aritmetici, ma come misure delle variazioni del livello dei prezzi. E pertanto l'omogeneità non dev'essere ricercata nell'identità delle voci considerate, ma nel fatto che gl'indici rispecchino bene quelle variazioni e quelle soltanto: dev'essere proprio ricercata nella diversità delle voci considerate e precisamente nella rispondenza di esse alle speciali caratteristiche delle collettività poste a confronto.

(¹) Sulla storia della nomenclatura nosologica delle cause di morte e sui criterî di applicazione, approvati dalla IV Conferenza internazionale (Parigi, 1929), cfr. ISTITUTO CENTRALE DI STATISTICA, *Nomenclature nosologiche per la statistica delle cause di morte e dizionario delle malattie.* Roma, 1933-XI (IV edizione).

(²) Cfr. il nostro studio: *Sulle variazioni dei prezzi*, in « Rivista delle Società Commerciali ». Roma, 1919; e MAURICE OLIVIER, *Les nombres indices de la variation des prix.* Paris, Giard, 1927.

Non meno erroneo sarebbe l'impiego, che di quegl'indici si facesse allo scopo di mettere in luce gli effetti di un regime doganale o le variazioni nel valore della ricchezza o nelle condizioni economiche di un dato paese: questi scopi sono completamente diversi da quelli, che guidano nella scelta delle merci e dei criterî di costruzione di quei numeri indici.

Invero l'omogeneità non è una proprietà assoluta di due o più fenomeni, ma una proprietà relativa allo scopo, che col confronto si vuole raggiungere. È innegabilmente utile il confronto, per una data popolazione, tra il numero dei nati e quello dei morti di un dato anno, perchè ci mette in grado di conoscere la misura e il senso della variazione naturale di quella popolazione; ma, quando il confronto si esegua allo scopo di conoscere la rapidità, con la quale i nati si estinguono, occorrerà tener conto del tempo in cui l'estinzione si avvera e confrontare, non più i nati ai morti dello stesso anno, ma i nati dell'anno ai morti che da essi provengono ed abbracciare il periodo, ben più lungo dell'anno, in cui quei morti si distribuiscono.

Analogamente, ammettendo di disporre di una statistica attendibile delle cause di morte, il confronto tra il numero dei morti in un anno per una data causa ed il numero medio dei viventi nell'anno fornirà certo un utile dato di fatto, ma non potrà dare un'idea dell'intensità d'azione della causa specifica considerata. Quei morti comprenderanno, tra l'altro, individui, che avrebbero alterato i confronti relativi ad altre cause, se quella considerata non li avesse prima colpiti; mentre non comprenderanno quegli altri individui, che sarebbero stati colpiti da tale causa, se non fossero stati prima eliminati da altre cause. Ond'è che quel confronto è perturbato dall'influenza di tutte le rimanenti cause di morte, e non risponde allo scopo speciale di misurare l'intensità, con cui quella causa avrebbe agito, indipendentemente dalle altre.

Si cadrebbe in un errore grossolano se, in un paese singolarmente progredito nella prevenzione degli infortunî sul lavoro, una più alta proporzione di morti per alcoolismo, ottenuta con un confronto di quella specie rispetto a un altro paese, si attribuisse interamente ad un'alta intensità d'azione di questa causa di morte, anzichè, almeno in parte, alla piccola influenza degli infortunî sul lavoro.

Quante applicazioni errate non si sono fatte, nello studio della distribuzione dei redditi, del cosidetto metodo dell'accrescimento percentuale delle classi (1)! Esso consisteva nel calcolare per una data località le percentuali d'incremento delle classi corrispondenti di due distribuzioni di redditieri, rilevate in tempi diversi, e nel dedurre dal confronto di tali

(1) Cfr. Costantino Bresciani-Turroni, *Osservazioni critiche sul metodo del Wolf per lo studio della distribuzione dei redditi*, in « Giornale degli economisti », 1914, e la bibliografia ivi citata.

percentuali il grado d'incremento delle diverse categorie di redditi. Se, ad es., si notava che il numero d'individui con redditi compresi tra 5.000 e 10.000 lire era cresciuto dall'uno all'altro tempo in una proporzione minore del numero d'individui con redditi compresi tra 10.000 e 15.000 lire, ed ancora minore del numero d'individui compresi nella classe successiva di redditi, e così via, si deduceva che i piccoli redditi erano cresciuti meno dei grandi.

Non si pensava che la conclusione sarebbe stata corretta, se non ci fosse stato alcuno spostamento di redditieri dall'una all'altra classe e se quel diverso aumento relativo di essi nelle varie classi fosse derivato solo dal fatto che nel secondo tempo nuovi redditieri si erano aggiunti a quelli esistenti nel primo tempo. Chè se, invece, fosse derivato ad es. dal fatto che i piccoli redditieri erano saliti in classi più elevate in seguito ad un aumento dei loro redditi, quella deduzione sarebbe stata manifestamente infondata.

La negligenza o l'ignoranza dei sani criterî metodologici sono spesso causa di errati confronti, ma non meno spesso tali errori si commettono per l'influenza di preconcetti specifici.

Un vecchio procedimento è quello di confrontare i valori più bassi di un fenomeno, relativi ad es. ad un periodo di depressione, con quelli successivi più alti, relativi ad un periodo di slancio, al fine di dimostrare non il fatto dell' aumento dei valori dall' uno all' altro dei periodi considerati, ma la tendenza generale all'aumento dei valori del fenomeno; oppure valori più alti con altri successivi più bassi, oppure ancora valori successivi approssimativamente uguali, al fine di inferirne la tendenza generale alla diminuzione, o, rispettivamente, alla stasi.

Analogamente, scegliendo alternativamente valori successivi alti e bassi, si può gabellare come ciclico un andamento del tutto irregolare. Il procedimento riesce un po' meno grossolano, qualora si eseguano i confronti, invece che tra successivi dati singoli, tra medie relative a successivi periodi opportunamente scelti. Furono questi, purtroppo, i procedimenti, con cui si cominciarono a studiare le crisi economiche; e non è da escludere che qualche studioso, per dimostrare nel modo più semplice e persuasivo la periodicità ed il sincronismo di esse nei principali paesi, abbia talvolta ceduto inconsciamente, nella scelta dei massimi e dei minimi o dei successivi periodi, a quell'idea preconcetta.

D' altra parte, solo un forzamento mentale dei dati avrebbe potuto dare credito a quella periodicità rigorosa, che pur si ritenne caratteristica di quelle crisi, ed alle spiegazioni fisiche, che ne furon date (1).

---

(1) Sui cicli economici possono leggersi recenti e apprezzati lavori nel Vol. VI della «Nuova collana di economisti stranieri e italiani». Torino, Utet, 1932-X.

Sono pure preconcetti di tal genere quelli, che generalmente guidano nei soliti confronti tra le entrate e le spese o le pressioni tributarie dei diversi paesi o delle diverse regioni di uno stesso paese. L'impossibilità di ottenere in esse, sia pure una parvenza di omogeneità per quella stima relativa dei fattori del benessere economico, che per lo più si ha in vista, rende raggiungibili con un po' di buona volontà tutte le conclusioni desiderabili.

In tutti cotesti confronti per lo meno si conosce lo scopo, al quale si mira; ma si assiste pure a confronti eseguiti senza una precisa determinazione di esso. Dobbiamo, però, soggiungere che bene spesso l'incertezza è puramente formale, perchè in sostanza si tende ad uno scopo determinato, che però, esplicitamente dichiarato, infirmerebbe i confronti e le conseguenti deduzioni, che si ha interesse di trarre.

5. Le semplificazioni statistiche. - Scopo ultimo di ogni rilevazione, classificazione, confronto è conoscere la natura dei fattori influenti su un dato fenomeno.

In questa indagine, essenzialmente scientifica, il metodo statistico presenta vantaggi, allorchè sul fenomeno influisce una grande quantità di altri fenomeni. Infatti in tal caso l'analisi può in generale semplificarsi, considerando masse di osservazioni; ciò che appunto — come già si è detto — è la caratteristica fondamentale di quel metodo.

Con un grande numero di osservazioni possiamo, anzitutto, indagare la natura generica dei fenomeni influenti ed in certi casi esaminare l'azione di alcuni gruppi di essi, ossia applicare uno di quegli schemi teorici, di cui consta la prima categoria di metodi statistici.

Il primo schema di tal genere fu costruito da Bernoulli ed è relativamente semplice e facilmente maneggevole: è quello dell'urna, a cui abbiamo accennato nel precedente capitolo; ond'è accaduto che, sebbene si siano poi costruiti altri schemi atti a rappresentare gli svariati intrecci di fenomeni influenti, che si presentano nella realtà, taluni studiosi si sono illusi d'incontrarlo in una quantità di casi, che da esso ripugnano. Quetelet fu appunto il più autorevole di essi e forse anche il più spregiudicato nell'applicazione di quello schema, in quanto ritenne che le distribuzioni delle stature di coetanei seguano sempre il detto schema di Bernoulli, e che a tale schema obbediscano pure le distribuzioni dei pesi del corpo umano e di tutti gli altri caratteri fisici e morali dell'uomo (1).

---

(1) Adolphe Quetelet, *Sur l'homme et le développement de ses facultés, ou essai de physique sociale*. Bruxelles, 1835. (La seconda edizione, dal titolo leggermente diverso, fu pubblicata con notevoli aggiunte nel 1869); e *Anthropométrie ou mésure des différentes qualités de l'homme*. Bruxelles, 1871. Ambedue queste opere sono state tradotte in italiano e pubblicate nella « Biblioteca dell'economista », Serie III, Vol. II.

Invero egli fu dapprima alquanto cauto nel sostenere questa tesi semplicista; ma nell'*Anthropométrie* si decise, infine, ad abbracciarla e a dichiarare persino, parlando del peso umano, di annettere una grandissima importanza al suo metodo uniforme: « Mi fu dato osservare, dopo studî numerosi e verifiche penose, a quali semplificazioni si possa giungere studiando la natura nei suoi fenomeni più importanti. Trovo in ciò una conferma del fatto che i procedimenti più semplici e comodi non sono sempre i primi seguiti, quando si vuol giungere alla verità ».

Purtroppo, però, si trattava di un mero preconcetto del valente astronomo belga: in base allo sviluppo di nuovi schemi teorici e ad indagini ben più precise, anche limitatamente al materiale di cui egli allora disponeva, noi possiamo adesso affermare che ben si apponeva il Quetelet della prima maniera, e che i procedimenti più semplici e comodi non sono sempre quelli che conducono alla verità.

D'altra parte, non sempre nell'applicazione degli schemi si considera un numero di osservazioni sufficiente a far riconoscere la forma, a cui tende la distribuzione di esse.

Cotesta insufficienza si presenta anche fuori dell'applicazione degli schemi teorici e molto frequentemente in quella che abbiamo chiamato la seconda categoria di problemi di cui consta la metodologia statistica, ossia nell'indagine dei legami tra le variazioni di due o più fenomeni, dove appunto un gran numero di osservazioni permette di semplificare l'analisi e precisamente di isolare più o meno nettamente alcuni fenomeni e di esaminarne la speciale influenza.

L'insufficienza del numero delle osservazioni particolarmente si nota in quei rami di ricerche, in cui lo studioso sia costretto a raccogliere da sè il materiale statistico e non disponga di sufficiente larghezza di mezzi.

Broca, generalmente ritenuto come uno dei fondatori dell'antropometria, di fronte alla difficoltà di disporre di larghi materiali di studio, arrivò persino a sostenere che una media di 20 osservazioni omogenee possa servire di base alle analisi antropometriche. E, studiando il peso del cervello umano in rapporto all'età, credette lecito fondare le sue deduzioni su classi di età comprendenti da un minimo di 11 ad un massimo di 51 osservazioni pei maschi e da un minimo di 13 ad un massimo di 34 per le femmine (1).

I risultati ottenuti da una parte dimostrarono che il peso di quei cervelli cresceva sino ai quarant'anni e leggermente diminuiva nella vecchiaia; ma, d'altra parte, nei primi gruppi di età corrispondenti al minor numero di osservazioni, si presentarono così inverosimili che egli

(1) Paul Broca, *Instructions craniologiques et craniométriques*, in « Mémoires de la Société d'Anthropologie de Paris », Serie II, Vol. II, 1875.

fu costretto, per non ammettere ipotesi assurde, a trascurare due osservazioni, mostrandone l'« anomalia ».

Orbene tutto questo — è stato giustamente osservato — dimostrava precisamente un'altra cosa, cioè che il numero delle osservazioni era insufficiente.

Altri errori si commettono nel ritenere trascurabile l'azione di un fenomeno o di un gruppo di fenomeni su quello in esame. In alcuni casi è piccolo e tollerabile l'arbitrio insito in tale procedimento, ma vi sono casi in cui è manifestamente eccessivo. È vero che la scienza procede per approssimazioni successive, ma è anche vero che è troppo semplice una proposizione che raccolga solo un centesimo di verità.

Se si nota che, durante la grande guerra, i cambî dell'Italia variarono fortemente e, tenendo presente che in quel periodo si ebbero vincoli economici di varia natura, si riesce a connettere l'azione di questi vincoli al livello dei cambî, la relazione sarà vera, ma abbraccerà una troppo piccola parte di vero: trascurerà fenomeni d'importanza notevolissima. Se, d'altra parte, si potessero considerare tutti i fatti di natura economica, demografica, politica, militare, che hanno influito su quelle variazioni dei cambî, l'indagine riuscirebbe ad afferrare completamente la verità, ma sarebbe troppo complicata e difficile.

Affermando, invece, che quelle variazioni dei cambî risultarono dall'azione combinata di vincoli, sbilanci economici con l'estero e circolazioni monetarie, e cercando di stabilire i legami tra i cambî ed ognuna di coteste categorie di fenomeni, l'indagine sarà molto meno complessa e riuscirà ad abbracciare una gran parte di vero.

Sull'argomento della criminalità si esercitò con singolare larghezza la tendenza smoderatamente semplificatrice di alcuni studiosi. Ci fu chi sostenne l'influenza preponderante delle circostanze climatiche sulla disposizione al delitto; chi credette nell'influenza prevalente dell'ignoranza, specie riguardo ad alcune categorie di delitti; chi giurò sull'influenza decisiva della razza.

Non meno pregiudizievole è, in generale, il ritenere come semplici relazioni di causa ad effetto le relazioni di interdipendenza.

Troppe volte è stato sostenuto, anche da studiosi di professione, che la speculazione dà luogo ad un aumento dei prezzi; e, d'altra parte, obbiettato che è invece l'aumento dei prezzi che alimenta la speculazione: la verità è che tra codesti fenomeni ha sempre avuto luogo un seguito di azioni e reazioni, di cui non è facile impresa descrivere la natura, il senso, l'intensità: l'idea di causalità è molto semplice e comoda, ma in questo caso allontana enormemente dalla realtà.

Purtroppo, quell'idea di causalità ha dominato a lungo nelle menti degli studiosi ed è stato uno degli ostacoli più tenaci allo sviluppo di alcuni rami del sapere ed in ispecie delle scienze sociali: si può dire che non esistono fenomeni economici, demografici, ecc., sui quali non siano state escogitate teorie ispirate a concetti di causalità e ricercate in tal senso conferme induttive.

Altro errore frequentissimo di semplificazione è quello di considerare come relazioni di causalità o di interdipendenza quelle relazioni di concomitanza, che hanno luogo tra due o più fenomeni, allorquando essi risentono insieme l'influenza di altri fenomeni.

Trascurando questo speciale legame, ognun vede quanto si possa essere sviati nella ricerca della verità: invece di andare in cerca dei legami intermedî, che effettivamente avvincono i fenomeni in esame all'origine comune della concomitanza, si cercherebbe di sorprendere anelli di causalità o di interdipendenza direttamente tra quei fenomeni, con risultati naturalmente cervellotici.

In alcuni casi quella relazione di concomitanza è evidente: sarebbe stolto ad es. parlare di relazione di causalità o di interdipendenza tra le variazioni mensili, pure a un dipresso inversamente solidali in alcuni paesi, del numero dei nati e dei suicidi; è chiaro che ambedue le variazioni dipendono — in maggior o minore misura e più o meno direttamente — dalla successione di quell'insieme di circostanze atmosferiche e di altra natura, che sono connesse alle stagioni. Ma vi sono casi, in cui la relazione di concomitanza non è così evidente e non si lascia facilmente scoprire, anche perchè può coesistere con una relazione di causalità o di interdipendenza. Ad es. i prezzi di due o più beni possono essere legati fra loro in un certo senso (ad es. da rapporti di strumentalità: filati e tessuti di cotone, o da rapporti di complementarità: lame e manichi di coltelli) ed inoltre dipendere insieme dall'ammontare della circolazione monetaria.

Diremo adesso brevemente di alcuni errori di semplificazione, che si commettono nella enunciazione o formulazione dei legami, già accertati per un dato fenomeno. La grande complessità di tali legami rende in generale necessario di riassumerli in un'espressione verbale o di altra natura, ma comunque relativamente semplice ed intelligibile alla nostra mente.

Pearson (1) molto efficacemente definisce la legge scientifica come un « riassunto in stenografia mentale ». Noi, però, dobbiamo ben guardarci dai riassunti di quegli stenografi, che non sappiano afferrare bene il senso dei discorsi!

Vide ben chiaro Malthus — sulle tracce dei suoi predecessori — il freno che viene alla lunga esercitato dalle condizioni economiche sullo sviluppo demografico; ma non si possono prendere alla lettera — com'è stato pur fatto — le formulazioni matematiche dello sviluppo della popo-

---

(1) Karl Pearson, *A grammar of science*. London, 1892.

lazione e delle sussistenze, da lui illustrate in un libro famoso (1), senza precludersi la via all'indagine, sia pure largamente approssimativa, delle intricatissime condizioni dell'equilibrio demografico.

Trattandosi della teoria dei giuochi, fu notato giustamente che i guadagni o le perdite pecuniarie non debbono essere considerate nel loro ammontare assoluto, ma in relazione alla ricchezza, che già si possegga.

Il fatto, che il piacere o il dolore dell'acquisto, o, rispettivamente, della perdita di una data somma di denaro è generalmente minore in corrispondenza ai più alti patrimonî, non era certo rilevato numericamente, ma era il risultato di apprezzamenti o stime quantitative, fatte in base ad osservazioni individuali. Si poteva, d'altra parte, pensare che, essendo la moneta un bene strumentale, quel piacere o dolore dipendesse da quel gruppo di beni, assortito in conformità ai proprî gusti, che la perdita o il guadagno monetario venisse rispettivamente a togliere o ad aggiungere ai beni, che già si possedessero; e fosse, quindi, variabilissimo dall'uno all'altro individuo.

Ciò nondimeno, Cramer non esitò a ritenere che, almeno nei limiti dell'esperienza passata, il piacere procurato dalla ricchezza monetaria varî come la radice quadrata del suo ammontare; e Daniele Bernoulli che il piacere dell'acquisto di una somma di moneta sia proporzionale al logaritmo del rapporto tra la somma acquistata e il patrimonio originario (2).

Restando ancora nel campo dell'economia, chi può oggi più parlare seriamente di relazioni rigorose tra prezzi e consumi, tra redditi e patrimonî o affitti, e così via?

Anche in antropometria sono state formulate ed hanno lungamente dominato alcune leggi, addirittura inattendibili per la loro eccessiva semplicità; ed oggidì si parla con la maggiore circospezione e solo in qualche caso singolare di quelle «leggi auree» (proporzioni quadratiche, cubiche, ecc.), che deliziarono i vecchi antropologi.

6. Le generalizzazioni e le cosidette leggi statistiche. – Su un punto ancora dobbiamo soffermarci, che permetterà di apprezzare l'importanza, ai fini delle applicazioni, dei risultati a cui il metodo statistico conduce. Abbiamo già osservato che tali risultati consistono o nell'accertamento delle relazioni tra un fenomeno ed un complesso generico di altri fenomeni (1ª categoria di metodi) o nell'accertamento delle

(1) T. R. Malthus, *An essay on the principles of population.* London, 1798; tradotto in italiano e pubblicato nella «Biblioteca dell'economista», Serie I, Vol. IX.

(2) Cfr. Alberto de' Stefani, *L'ofelimità del denaro*, in «Atti dell'Istituto Veneto di Scienze, Lettere ed Arti», 1913; Ragnar Frisch, *New methods of measuring marginal utility.* Tübingen, Mohr, 1932; e il nostro lavoro: *L'utilità della moneta e l'imposta progressiva*, in «Rivista italiana di statistica, economia e finanza», 1933.

relazioni tra un fenomeno ed uno, due, tre..., ma non tutti gli altri fenomeni, con i quali il primo è in relazione (2ª categoria di metodi).

Ciò fa sorgere la seguente domanda: dal fatto che un intreccio di fenomeni si è rivelato in passato in una certa relazione con un dato fenomeno, siamo noi autorizzati ad estendere la validità di tale relazione oltre i limiti delle osservazioni? O meglio, siamo autorizzati a generalizzarla in base a quelle stesse condizioni, che renderebbero lecita l'estensione delle relazioni elementari? Non è dubbia la risposta negativa: la generalizzazione *sic et simpliciter* sarebbe evidentemente infondata.

Ad es., i morti nel primo anno di vita sono moltissimi, diminuiscono rapidamente di numero sino a toccare un minimo a 10-15 anni, crescono poi più o meno velocemente sino a circa 70 anni e diminuiscono infine gradatamente sino a ridursi a qualche caso erratico intorno ai 100 anni.

È questa, in forma grossolana, la relazione che lega il numero dei morti alla loro età ed a quel complesso di fattori, quali l'ambiente di vita, le occupazioni, ecc., che variano sistematicamente con l'età.

Sotto certe ipotesi, che trovano in generale riscontro nella realtà, quella relazione si osserva, tanto se si considera la distribuzione per età dei morti provenienti da una stessa generazione, quanto se si considera la distribuzione per età dei morti di un dato anno, qualunque sia la generazione da cui essi provengono.

Essendosi riconosciuto che coteste distribuzioni di morti seguono un certo schema teorico, è stato proposto di scomporle in alcune distribuzioni parziali, corrispondenti a diversi stadi della vita: si tratta naturalmente di operazioni alquanto arbitrarie, che comunque riescono ad isolare gruppi molto comprensivi di influenze e non permettono di riconoscere e determinare l'azione singola degli svariatissimi fattori della mortalità umana.

In tali condizioni, siamo noi autorizzati ad estendere la validità di quella relazione oltre i limiti delle osservazioni? E chi può assicurarci che i progressi dell'igiene, il miglioramento del tenore di vita, la diminuita pericolosità dei mestieri, la più rigorosa osservanza della età minima nel lavoro dei fanciulli, ecc., non riusciranno a far perdurare il basso numero di morti, che attualmente si osserva nell'adolescenza, sino ai 20 anni ed oltre?

E chi, d'altra parte, può escludere che si avveri il timore di alcuni studiosi, che una forte e continua diminuzione della mortalità nei primissimi anni d'età, dovuta ai miglioramenti economici, ai progressi sanitarî, ecc., accresca sensibilmente la mortalità negli anni successivi sino al punto da mutare radicalmente la forma di quella distribuzione?

Quella relazione, invero, nasconde un insieme di legami tra l'età dei morti ed altri fatti, che non si è ancora riusciti a conoscere in modo soddisfacente. E pertanto noi possiamo solo affermare che, se quell'insieme non sarà in futuro molto diverso da quello passato, la distribuzione dei morti per età non sarà molto diversa da quella descritta. Poichè in particolare quel tanto, che di quell'insieme conosciamo, ci fa ritenere sommamente improbabile che esso possa sensibilmente variare in brevissimo tempo, noi possiamo ammettere la validità approssimativa di quella relazione nel vicinissimo futuro, e non solo nei termini molto grossolani, in cui l'abbiamo enunciata, ma anche in qualche particolare, che ci è stato dato di cogliere.

E possiamo persino estendere in quel modo, per una data popolazione, i particolari che — ad esclusione di quelli dovuti a fatti varianti accidentalmente con l'età — si osservano nelle distribuzioni dei morti di quella popolazione.

È questo il solo modo lecito di generalizzare.

Ad analoghe conclusioni saremmo pervenuti, se, anzichè il numero dei morti, avessimo considerato quei quozienti che si chiamano comunemente saggi di mortalità; ossia se avessimo eliminato per via di rapporti l'influenza sistematica che alle diverse età esercita il numero degli esposti a morire. Applicando appunto a siffatte tavole il tipo di generalizzazione sopra illustrato, è stato possibile dar solida base alle operazioni di assicurazione sulla vita umana.

Nel campo economico, molta risonanza ha avuto in questi ultimi anni un « barometro » mensile degli affari, calcolato per gli Stati Uniti dal Comitato Economico Harvard, al fine di formulare previsioni, non in base all'andamento ciclico dell'attività economica — non essendo affatto costante la durata dei successivi cicli economici — ma in base ad una legge di successione nell'andamento ciclico di tre categorie di fatti economici.

Per un periodo di tempo abbastanza lungo, chiusosi nel 1914 con lo scoppio della guerra mondiale, fu notato che, fatte le debite eliminazioni, il ciclo della speculazione presentava il massimo ed il minimo alcuni mesi prima del cosidetto ciclo degli affari; e questo a sua volta presentava quei massimi e minimi alcuni mesi prima del ciclo dei saggi dell'interesse. Pertanto quel Comitato pensò di costruire tre curve mensili, di cui la prima rappresentava l'andamento medio del debito bancario a New York e dei corsi delle azioni (indice della speculazione); la seconda l'andamento medio del debito bancario in 140 città, esclusa New York, e del livello generale dei prezzi delle merci (indice degli affari); la terza l'andamento medio dei saggi dell'interesse (indice monetario). Sino al 1914 ottenne la conferma di quella legge di successione: con sorprendente regolarità, nei periodi di prosperità o di depressione economica la prima ad ascendere o, rispettivamente, a discendere era la curva della speculazione; ad essa seguiva quella degli affari e veniva ultima quella dell'interesse (¹).

Se ne inferì che, se nel periodo considerato si fosse seguito l'andamento delle tre curve, le crisi economiche, pur non presentando una periodicità rigorosa, si sarebbero potute prevedere, perchè l'andamento sistematicamente discendente della prima curva avrebbe fatto ritenere prossimo quello delle altre e quindi la depressione economica; e viceversa.

Tale legge di successione, naturalmente, non si è sempre verificata negli anni successivi, essendo profondamente variato l'insieme di circostanze, che aveva dato luogo a quella regolarità; ma frattanto, in armonia a quel felice ordine di idee, le aziende nord-americane, avvalendosi largamente di statistici professionisti, hanno lavorato e lavorano fervidamente per « business forecasting » nei riguardi delle loro particolari industrie, riuscendo spesso a risultati praticamente preziosi, sebbene in generale gelosamente nascosti, sulle relazioni di concomitanza e di successione di alcuni rami dell'attività economica (²).

Gli esempî potrebbero moltiplicarsi, anzi si può dire che sono di questa natura i risultati del metodo statistico: nei fenomeni, che si trovano in relazione con un grandissimo numero di altri fenomeni, e a cui si vuole applicare il metodo statistico, per quanto si esaminino una, due,... n relazioni, ne resteranno sempre altre, spesso d'importanza punto trascurabile, che non consentiranno una generalizzazione diversa da quella, più o meno limitata, di cui ci siamo occupati.

Vi saranno fenomeni, per i quali le variazioni sono lentissime, e fenomeni, in cui esse avvengono meno lentamente, con carattere evolutorio od ondulatorio, ed anche attraverso qualche brusca variazione. Ond'è naturale che i cosidetti casi adiacenti, ai quali quelle relazioni si estendano, dovranno essere tanto più vicini, quanto meno lente siano state le variazioni; e tanto più probabilmente confermeranno la relazione, quanto meno suscettibile siasi rivelato il fenomeno di variazioni brusche od ondulatorie. L'importanza di coteste generalizzazioni è però sempre grandissima, perchè, malgrado la nostra ignoranza, esse ci consentono con le debite cautele, ed entro certi limiti e condizioni, di far previsioni, che diversamente non sarebbero affatto possibili; e ci permettono, quindi,

---

(¹) Cfr. « The review of economic statistics », edita dal detto Comitato e dove sono stati riportati periodicamente i risultati di tali elaborazioni e quelli ottenuti negli altri Stati, che ne hanno seguito l'esempio. Sulle modificazioni, successivamente apportate a quel barometro, e su altri tentativi del genere, cfr. il volume VI della « Nuova collana », citato in fine al n. 4.

(²) Su questo nuovo indirizzo può consultarsi il nostro studio: *Sui metodi di previsione economica*, pubblicato in « Rivista bancaria » del 15 luglio 1930 e il nostro editoriale: *Il barometro economico*, pubblicato nella cit. « Riv. it. di statistica, economia e finanza » del marzo e del giugno 1932. Molti dettagli sono esposti da W. WALLACE, *Business forecasting and its practical application*. London, Pitman, 1928 (2ª edizione).

di procedere un po' meno ciecamente nella soluzione dei problemi concreti della vita.

A tali generalizzazioni si è voluto dare il nome di *leggi statistiche*, ma si tratta di approssimazioni sommarie a complessi, non pienamente scomponibili, di legami elementari; e si va sempre più accertando in ogni ramo del sapere che in fondo è questo il carattere, più o meno pronunciato, di ogni legge scientifica.

PARTE PRIMA

PARTE PRIMA

I MODI RAPPRESENTARE I GRUPPI DI OSSERVAZIONI

CAPITOLO I.

LE TAVOLE NUMERICHE

1. LE DISTRIBUZIONI STATISTICHE. – I dati statistici sono raccolti a mezzo di spogli in tavole primitive o provvisorie, naturalmente variabili da caso a caso e sulle quali sarebbe vano e poco serio tentar di costruire una dottrina. I demografi, gli economisti, i biologi, i psicologi, i meteorologi, ecc., adottano a questo riguardo i più svariati criterî, che la pratica loro suggerisce in relazione alle caratteristiche dei fenomeni che si propongono di rilevare, agli scopi speciali cui mirano ed ai progressi incessanti della tecnica, che fornisce sempre nuovi modelli di macchine selezionatrici di schede, calcolatrici, ecc.

Però, qualunque siano i particolari criterî seguiti per lo spoglio dei dati, conviene che questi risultino infine disposti in due modi diversi, secondo che si voglia accertare la relazione intercedente tra i valori osservati di un fenomeno e il numero di volte che essi si presentano, o tra i valori osservati di un fenomeno e quelli corrispondenti di altri fenomeni.

Nel primo caso giova graduare in ordine crescente i valori diversi del fenomeno e scrivere accanto ad ogni valore il numero corrispondente delle osservazioni.

Questa rappresentazione numerica dà luogo ad una distribuzione statistica (1). Ne abbiamo già dato un esempio con la tavola I del primo capitolo, e le tavole che qui riportiamo ne esemplificano i principali tipi (2).

a) È da notare, anzitutto, che i valori osservati possono essere o non essere suscettibili di variare in modo continuo.

La tavola V riguarda un gruppo di valori, per loro natura variabili con discontinuità; e precisamente rivela quante vedove con 0, 1, ...5

---

(1) Gli statistici inglesi la chiamano spesso *seriation*, espressione che alcuni statistici nostrani traducono letteralmente in seriazione.

Per maggiori sviluppi sugli argomenti trattati in questo capitolo, cfr. ARTHUR L. BOWLEY, *Elements of statistics*. London, King, 1926 (quinta edizione, tradotta in francese, Paris, Giard, 1929).

(2) I dati, di cui non citiamo la fonte, sono ricavati da pubblicazioni ufficiali.

figli minorenni passarono a seconde nozze a Milano nel 1916: è chiaro che il numero dei figli non può essere che un intero.

Lo stesso è a dire per la tavola VI, che dà la distribuzione delle famiglie naturali censite in Italia al 1° dicembre 1921, secondo il numero dei componenti di esse.

Riguardo alla distribuzione, secondo l'ammontare dei redditi, dei redditieri prussiani accertati nel 1913 ai fini dell'imposta generale sui redditi (tavola X), notiamo che sarebbero possibili variazioni continue, se quei redditi monetarî, legalmente accertati, potessero variare per quantità inferiori alle più piccola frazione legale dell'unità monetaria.

Invece la distribuzione per statura di un gruppo d'italiani a venti anni d'età nel 1913 (tavola VII); la distribuzione dei 4.748 giorni del tredicennio 1878-1890 secondo l'altezza barometrica osservata in ognuno di essi a Southampton (tavola VIII); la distribuzione, secondo l'età delle madri, dei nati (vivi e morti) a Parigi nel 1919 (tavola IX), riguardano misure per loro natura suscettibili di variare con continuità.

b) D'altra parte, se i valori diversi osservati siano alquanto numerosi e se la natura del fenomeno considerato renda paghi di deduzioni approssimative, nella prima colonna della distribuzione si porranno classi consecutive di valori, procedenti ad intervalli costanti ed aventi un'ampiezza rispondente al grado di approssimazione che si voglia raggiungere, e nella seconda colonna si scriverà il numero di osservazioni corrispondente alle classi medesime.

La tavola VIII riguarda appunto cotesto caso ed anche le tavole VI, VII, IX e X, se si prescinde dal fatto che nella VI è ignoto il limite superiore dell'ultima classe, nelle IX e X è ignoto il limite inferiore della prima e nella VII sono ignoti ambedue i limiti estremi.

Per qualche scopo speciale, o se la graduatoria dei valori sia molto lunga, sogliono anche adottarsi classi di ampiezza diversa, il che però può essere di ostacolo all'applicazione di alcuni procedimenti statistici d'importanza fondamentale. Nella tavola X si osserva, appunto, che il numero dei redditieri prussiani è distribuito in classi consecutive di redditi, procedenti ad intervalli sempre più estesi.

Si sogliono spesso riportare per brevità nella prima colonna della distribuzione i soli valori centrali delle classi, od anche i soli limiti inferiori di esse, seguiti da una lineetta. Pertanto nella tavola VIII alle altezze barometriche di 2.855, 2.875,... centesimi di pollice corrispondono le classi: 2.845-2.865, 2.865-2.885,..., e nella tavola IX alle età di α-, 15-,... anni corrispondono le classi: α-15, 15-20,...

c) Inoltre le classi possono essere distinte in guisa che in ognuna di esse siano compresi anche i valori estremi (tavola VI), oppure solo il limite inferiore (tavola IX) o quello superiore (tavola X) di esse. Può, altresì, dividersi in due parti uguali ed attribuirsi a due classi consecutive ogni osservazione avente un valore uguale al limite comune delle due classi. In questo caso (tavv. VIII e I) il numero delle osservazioni di ogni classe può presentare una cifra decimale uguale a 5.

In generale, però, a parità d'intervalli delle classi, l'uno o l'altro di cotesti procedimenti non influisce sensibilmente sui risultati dell'aggrup-

TAVOLA V.

Numero dei figli minorenni	Vedove passate a nuove nozze
0	109
1	60
2	35
3	17
4	5
5	1
Totale	227

TAVOLA VI.

Numero dei componenti	Famiglie naturali
1 - 2	2.237.146
3 - 4	2.812.157
5 - 6	1.992.314
7 - 8	984.242
9 - 10	362.759
11 - 12	116.069
13 - 14	45.117
15 - ω	44.419
Totale	8.594.223

TAVOLA VII.

Stature in mm.	Individui
a - 1.255	1
1.255 - 1.305	2
1.305 - 1.355	4
1.355 - 1.405	10
1.405 - 1.455	30
1.455 - 1.505	108
1.505 - 1.555	494
1.555 - 1.605	1.840
1.605 - 1.655	2.957
1.655 - 1.705	2.688
1.705 - 1.755	1.312
1.755 - 1.805	441
1.805 - 1.855	94
1.855 - 1.905	16
1.905 - ω	3
Totale	10.000

TAVOLA VIII.

Altezze barometriche in centesimi di pollice	Giorni
2.855	3
2.875	6
2.895	22
2.915	58,5
2.935	187
2.955	436
2.975	812
2.995	1.151
3.015	1.119,5
3.035	619,5
3.055	278
3.075	50,5
3.095	5
Totale	4.748

TAVOLA IX.

Età delle madri in anni	Nascite
a –	5
15 –	2.324
20 –	11.472
25 –	12.485
30 –	9.044
35 –	4.897
40 –	1.474
45 –	126
50 –	3
Totale: età note	41.830
» » ignote	806
Totale	42.636

TAVOLA X.

Redditi annui in marchi	Redditieri
a – 900	8.698.666
900 – 3.000	6.489.373
3.000 – 6.000	579.773
6.000 – 10.500	135.189
10.500 – 30.500	87.061
30.500 – 100.000	22.239
100.000 – 500.000	4.416
500.000 – 1.000.000	242
1.000.000 – 24.000.000	89
Totale	16.017.048

pamento e delle elaborazioni successive; onde talvolta persino s'ignora quale criterio sia stato seguito. È questo il caso della tavola VII, per la quale abbiamo tenuto presente che ad ogni statura è stata attribuita nella rilevazione la misura del più vicino centimetro e che quindi le stature rilevate come inferiori o uguali a 125 cm. erano invece inferiori a circa 1.255 mm., quelle rilevate tra 126 e 130 cm. erano comprese tra 1.255 e 1.305 mm. circa, e così via; senza peraltro poter determinare quale limite fosse compreso in ogni classe.

Talvolta, come accade in molte distribuzioni per età, per rendere palese l'esclusione del limite superiore, le classi si scrivono ad es. nella forma: 0-4, 5-9, 10-14,.... e quindi comprendono individui dalla nascita sino a 4 anni compiuti, ossia sino alla vigilia del quinto compleanno; da 5 anni precisi sino a 9 compiuti, ossia sino alla vigilia del decimo compleanno, e così via. È questo il sistema usato di regola dal nostro Istituto Centrale di Statistica.

d) Osserviamo ancora che una distribuzione — anzichè dei numeri delle osservazioni corrispondenti a valori singoli od a classi di valori — può constare dei rapporti, in cui quei numeri stanno al totale delle osservazioni. Ad es. la tavola VII riguarda gl'iscritti di leva nati nel 1893, il numero dei quali non è stato punto 10.000, ma 305.172: la seconda colonna di essa consta di rapporti, ottenuti dividendo i numeri originarî per il totale delle osservazioni e moltiplicando i risultati per 10.000. Sono stati raggruppati nelle classi estreme i cosidetti valori erratici, ossia quei valori estremi, che per il loro piccolo numero generalmente presentano grandi irregolarità d'andamento e i cui rapporti, singolarmente considerati, avrebbero dato luogo a decimali trascurabili.

Siffatti rapporti sono comunemente chiamati *frequenze*, e i numeri originarî *frequenze assolute*; pertanto le distribuzioni statistiche spesso si chiamano distribuzioni di frequenze.

e) Talvolta può essere opportuno costruire una distribuzione di osservazioni accumulate, ossia corrispondenti a classi di valori minori o uguali, oppure maggiori o uguali, a dati limiti.

Riportiamo qui di seguito le distribuzioni, di cui alle tavole VII e X, disposte rispettivamente per classi di valori minori o uguali ai limiti superiori delle classi di stature e maggiori dei limiti inferiori delle classi di redditi.

Nel primo caso la nuova distribuzione si è ottenuta, lasciando inalterato nella seconda colonna il numero di osservazioni della prima classe e sostituendo a quello della seconda classe la somma dei primi due, a quello della terza la somma dei primi tre, e così via; ed è chiaro che nel secondo caso è bastato cominciare quest'operazione dall'ultima classe, anzichè dalla prima.

TAVOLA VII *bis*.

Stature in mm.	Individui
$\leqq$ 1.255	1
» 1.305	3
» 1.355	7
» 1.405	17
» 1.455	47
» 1.505	155
» 1.555	649
» 1.605	2.489
» 1.655	5.446
» 1.705	8.134
» 1.755	9.446
» 1.805	9.887
» 1.855	9.981
» 1.905	9.997

TAVOLA X *bis*.

Redditi annui in marchi	Redditieri
> *a*	16.017.048
» 900	7.318.382
» 3.000	829.009
» 6.000	249.236
» 10.500	114.047
» 30.500	26.986
» 100.000	4.747
» 500.000	331
» 1.000.000	89

f) Sinora abbiamo parlato di valori di un fenomeno, non perchè non esistano distribuzioni di osservazioni qualitative (la tavola XI è una distribuzione dei censiti in Italia nel 1911 secondo la religione dichiarata), ma perchè gli attributi non offrono largo campo di applicazione alle elaborazioni statistiche. Tranne, come ben si intende, che gli attributi possano ricondursi a forma quantitativa, come è ad es. il caso della distribuzione degli esiti dei conflitti di lavoro nelle industrie italiane durante il 1913 (tavola XII), in cui le tre classi: completamente sfavorevoli, in parte favorevoli e completamente favorevoli agli operai, corrispondono in sostanza ad esiti riusciti favorevoli rispettivamente per una proporzione di 0 per cento, superiore a 0 e inferiore a 100 per cento, e di 100 per cento delle richieste; o come è pure il caso della distribuzione del colore degli occhi di un gruppo di militari osservati in una inchiesta antropometrica italiana (tavola XIII), colore che è graduabile in base all'intensità della pigmentazione.

TAVOLA XI.

Religioni	Censiti nel 1911
Cattolica	32.983.664
Evangelica e protestante . . .	123.253
Greco-scismatica	1.378
Israelitica . . .	34.324
Altre	822
Nessuna	874.532
Non dichiarata .	653.404
Totale	34.671.377

TAVOLA XII.

Esiti dei conflitti di lavoro	Conflitti
Completamente sfavorevoli agli operai	31,4
In parte favorevoli agli operai	44,6
Completamente favorevoli agli operai	17,4
Incerti	6,6
Totale	100

TAVOLA XIII.

Colore degli occhi	Militari
Celesti. . .	30.882
Grigi . . .	61.605
Castani . .	180.255
Neri. . . .	26.118
Totale	298.860

g) Una distribuzione può essere statica o dinamica, secondo che consti di dati riguardanti lo stesso tempo o tempi diversi. La tav. VIII è una distribuzione dinamica; e noi possiamo, ad es., rilevare quante famiglie con 1-2 componenti in un dato tempo diventano di 3-4.... componenti in tempi successivi e dare una distribuzione formalmente simile a quella della tavola VI, ma in sostanza diversa perchè dinamica; o quanti bambini, in un gruppo di stature non superiori a 1.255 mm., raggiungono una statura compresa tra 1.255-1.305, poi una statura compresa tra 1.305-1.355, ecc. (distribuzione dinamica della tavola VII); o le nascite successive in un gruppo di donne di età inferiore a 15 anni, che attraversano la scala delle età feconde (distribuzione dinamica della tavola IX).

2. LE SERIE STATISTICHE. – Ove si voglia accertare la relazione intercedente tra i valori osservati di un fenomeno e quelli corrispondenti di altri fenomeni (circostanze, ecc.), conviene disporre i valori del primo fenomeno in corrispondenza alla successione crescente di quelli relativi all'uno o all'altro fenomeno noto.

Nella tavola XIV si tien conto del tempo, in cui le diverse rilevazioni delle tonnellate di stazza netta delle navi a vela italiane sono state fatte, e si dispongono tali rilevazioni secondo la successione degli anni: 1890-1913.

Siffatta rappresentazione numerica dà luogo ad una serie statistica.

D'altra parte, conoscendosi ad es. le tonnellate di stazza netta delle navi a vapore nazionali negli stessi anni, quelle 24 osservazioni possono anche disporsi secondo l'intensità crescente di cotesto tonnellaggio. Ne risulta la tavola XV.

TAVOLA XIV.

Anni	Tonnellate di stazza netta delle navi a vela a fine d'anno (in migliaia)
1890	634,1
1891	625,8
1892	609,8
1893	588,3
1894	571,6
1895	555,6
1896	527,6
1897	526,8
1898	537,6
1899	558,2
1900	568,2
1901	575,2
1902	570,4
1903	584,2
1904	570,4
1905	541,2
1906	503,3
1907	468,7
1908	453,3
1909	439,9
1910	432,7
1911	411,0
1912	374,8
1913	356,0

TAVOLA XV.

Ton. di stazza netta delle navi a vapore (in migliaia)	Ton. di stazza netta delle navi a vela (in migliaia)
186,6	634,1
199,9	625,8
201,4	609,8
207,5	571,6
208,2	588,3
220,5	555,6
237,7	527,6
259,8	526,8
277,5	537,6
314,8	558,2
376,8	568,2
424,7	575,2
448,4	570,4
460,5	584,2
462,3	570,4
484,4	541,2
497,5	503,3
526,6	468,7
566,7	453,3
631,3	439,9
674,5	432,7
697,0	411,0
762,3	374,8
876,9	356,0

Se, come nel primo caso, il fenomeno o la circostanza graduata è il tempo, la serie statistica è comunemente chiamata serie storica.

Serie statistica sarebbe ancora quella della tavola XVI, che dà il numero regionale dei cavalli-vapore impiegati nelle industrie italiane al

15 ottobre 1927, corrispondente al gettito regionale delle tre principal imposte dirette nel 1927.

Disponendo quel numero di cavalli-vapore secondo il numero degli addetti a tutte le industrie, pure rilevato in quelle regioni nel 1927, si ottiene la serie statistica della tavola XVII.

Tavola XVI.

Gettito delle tre imposte dirette nel 1927 (milioni di lire)	Forza motrice delle industrie nel 1927 (migliaia di H. P.)
21,8	14,3
41,3	99,3
53,2	304,4
59,2	273,5
60,1	112,5
68,0	105,3
116,8	93,4
134,2	169,8
175,6	97,4
244,8	246,1
282,1	695,0
298,7	528,2
334,9	635,3
355,2	349,9
398,1	495,7
471,0	442,3
635,7	1.841,7
1.190,5	2.074,0

Tavola XVII.

Numero degli addetti alle industrie nel 1927 (migliaia)	Forza motrice delle industrie nel 1927 (migliaia di H. P.)
19,4	14,3
48,6	304,4
54,2	273,5
60,1	99,3
66,5	105,3
80,4	93,4
84,7	112,5
113,3	169,8
132,2	97,4
179,1	349,9
212,9	246,1
221,7	695,0
225,7	528,2
249,0	442,3
328,2	635,2
331,2	495,7
549,0	1.841,7
1.046,6	2.074,0

Estendendo il procedimento descritto al caso di più fenomeni, possono darsi serie statistiche colleganti più di due variabili. La tavola IV dell'Introduzione è appunto una serie statistica a tre variabili, in quanto riporta il numero regionale dei cavalli-vapore e degli addetti, di cui ci siamo or ora occupati, corrispondente al gettito, disposto in ordine crescente, delle tre imposte dirette.

Analogamente a ciò che abbiamo osservato alla lettera *f*) del numero precedente, notiamo che, se ci fossimo limitati a riportare ad es. i dati del numero dei cavalli-vapore secondo le regioni, avremmo avuto una successione non ordinata di valori, corrispondente ad un gruppo di attributi (denominazioni delle regioni), e che una successione della stessa natura avremmo ottenuto, se ad es., invece della tavola X, avessimo riportato l'elenco delle occupazioni dei redditieri prussiani coi redditi corrispondenti accertati.

Come abbiamo detto per le analoghe distribuzioni di caratteri qualitativi, in siffatti tipi di serie occorre nettamente distinguere il caso di veri e proprî attributi qualitativi — che non offrono largo campo d'impiego alle elaborazioni statistiche — dal caso di attributi, che possano ricondursi a forma quantitativa.

Questo secondo caso sarebbe offerto, ad es., da una serie statistica ordinata per attributi graduabili, come l'ammontare degli stipendî percepiti in un dato tempo dalle varie classi gerarchiche del personale civile e militare di uno Stato; classi gerarchiche, che sono per loro natura graduabili, anzi in generale legalmente graduate.

Una serie, in cui tutti i fenomeni considerati fossero attributi, si avrebbe se ad es. di un gruppo di morti rilevassimo le cause di morte, le occupazioni e i luoghi di nascita (serie a tre variabili) o di un gruppo di coscritti il colore degli occhi e quello dei capelli (serie a due variabili). In questo secondo caso le variabili sarebbero ambedue graduabili, mentre nel primo caso non lo sarebbe nessuna e la serie statistica assumerebbe ad es. la forma seguente:

TAVOLA XVIII.

Cause di morte	Occupazioni	Luoghi di nascita
Polmonite	Contadino	Bologna
Tubercolosi	Operaio	Milano
Apoplessia	Ecclesiastico	Roma
. . . .	. . . .	. . . .

3. DIVERSI ASPETTI DI UNA RAPPRESENTAZIONE NUMERICA. DATO E UNITÀ STATISTICA. – Alcune rappresentazioni numeriche possono considerarsi come distribuzioni o come serie statistiche, secondo il punto di vista dal quale si considerano.

Ad es., se si considera ogni nato vivo come un'osservazione, dobbiamo ritenere la seguente tavola XIX come una distribuzione, dinamica per classi annuali, delle date di nascita dei nati vivi: ma, se consideriamo come una sola osservazione un gruppo annuale di nati vivi, dobbiamo ritenere la detta tavola come una serie statistica di cinque osservazioni, disposte secondo la successione dei tempi. Se disponessino quelle cinque osservazioni secondo la successione crescente di un altro fenomeno, ad es. secondo il numero corrispondente dei matrimonî, otterremmo un'altra serie statistica; e così via.

Analogamente, se nell'ammontare del debito ipotecario italiano, accertato al 31 dicembre 1910, consideriamo un'ipoteca come un'osservazione, dobbiamo ritenere che la tavola XX risulti da una serie statistica

TAVOLA XIX.

Anni	Numero dei nati vivi
1924	1.124.470
1925	1.109.761
1926	1.094.587
1927	1.093.772
1928	1.072.316

TAVOLA XX.

Saggi percentuali d'interesse	Debito ipotecario (in lire)
0 – 3,5	470.942.203
3,5 – 5	3.084.591.030
5 – 8	621.537.258
8 – 10	65.063.625
10 – 15	11.770.014

dei valori delle singole ipoteche, disposti secondo la successione crescente del saggio percentuale d'interesse, stipulato per ognuna di esse; e che tale serie sia stata poscia disposta per classi di saggi d'interesse.

Se, invece, consideriamo una lira di debito ipotecario come una osservazione, quest'esempio dovrà considerarsi come la distribuzione delle unità monetarie concesse in ipoteca, secondo il saggio percentuale d'interesse, che esse fruttano.

Mentre nel primo caso saremmo di fronte ad un gruppo di dati relativi al fenomeno: valore dell'ipoteca, ed al gruppo corrispondente: saggio percentuale d'interesse stipulato per ogni ipoteca; nel secondo caso avremmo un gruppo di dati relativi al saggio percentuale d'interesse delle singole lire prestate con ipoteca.

Occorre, peraltro, avvertire, che quella rappresentazione per classi non si presta all'analisi del primo punto di vista e consiglia di attenersi al secondo.

Generalmente si chiama unità statistica ogni elemento rilevato e dato statistico ogni aggruppamento di unità; ma ognun vede quanto siano relativi questi concetti, di cui si fa tanto discorrere nei trattati di statistica.

4. LE TAVOLE A PARECCHIE ENTRATE E I GRUPPI SCELTI. – Quando alcune combinazioni di valori si ripetono, può convenire di rendere più espressivo il secondo tipo di rappresentazioni numeriche, ponendo in una colonna i valori diversi di uno dei due fenomeni, graduati in ordine crescente, e in una riga i valori diversi dell'altro fenomeno, pure graduati in ordine crescente. Disponendo opportunamente riga e colonna, risulteranno tante caselle, quante sono le combinazioni di valori dei due fenomeni, e si potrà riportare in tali caselle il numero di volte, che ognuna di quelle combinazioni si ripete.

Cotesto tipo di rappresentazione numerica dà luogo ad una tavola a doppia entrata, la quale può constare anche di classi più o meno ampie di valori.

Ne abbiamo già dato un primo esempio con la tavola III dell'Introduzione. Se si volesse indagare l'influenza che sull'età dei morti a Milano esercitano i diversi mesi dell'anno, considerando ogni morte come un'osservazione, conviene pure adottare questo tipo di rappresentazione numerica.

Per il 1916 e nei riguardi dei soli maschi di età nota, si avrebbe la tavola XXI, che qui riportiamo.

TAVOLA XXI.

Giorni del 1916	Età dei morti										
	0-1	1-5	5-10	10-15	15-20	20-30	30-40	40-60	60-80	80-ω	Totali
1 - 31	84	49	10	4	13	32	26	100	200	40	558
32 - 60	65	53	14	8	13	27	32	87	154	26	479
61 - 91	77	57	11	3	11	25	28	88	143	19	462
92 - 121	46	47	5	1	10	25	21	81	95	10	341
122 - 152	44	38	9	4	7	30	16	61	98	16	323
153 - 182	47	40	14	7	10	19	20	63	88	12	320
183 - 213	85	32	14	4	12	23	32	69	97	11	379
214 - 244	56	19	9	8	11	16	32	70	76	20	317
245 - 274	44	26	9	5	12	18	21	69	82	11	297
275 - 305	35	24	1	7	13	16	25	72	123	16	332
306 - 335	33	23	9	5	11	16	22	71	108	11	309
336 - 366	41	34	5	8	15	39	25	89	138	30	424
Totali	657	442	110	64	138	286	300	920	1.402	222	4.541

Ugualmente potrebbe rappresentarsi la distribuzione per età di una popolazione rilevata in successivi censimenti, la distribuzione per età degli sposi secondo i mesi dell'anno o secondo l'età delle spose, la distribuzione delle nascite secondo l'età delle madri e la durata del matrimonio, ecc.

Riportiamo ancora due interessanti tavole a doppia entrata: la prima (tavola XXII), ricavata da un'inchiesta compiuta a Firenze su un gruppo di 6.673 abitazioni, permette di analizzare la relazione intercedente tra il numero degli abitanti ed il numero delle stanze, di cui esse constano; la seconda (tavola XXIII), ricavata da un censimento australiano del 1915, permette di analizzare la relazione, in cui i redditi non inferiori a 200 sterline stanno ai patrimonî non inferiori a 1.000 sterline.

Tavole a doppia entrata, in cui ad ogni valore di un fenomeno corrisponda un solo valore dell'altro fenomeno, sono molto rare: in tal caso il numero di osservazioni corrispondente ad ogni linea della

TAVOLA XXII.

Numero delle stanze	Numero degli abitanti																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Totali
1	30	73	58	47	36	16	11	5	2	2	—	—	—	—	—	—	—	280
2	25	166	173	161	121	86	44	17	8	5	2	1	—	—	—	—	—	809
3	12	190	323	343	303	203	161	88	47	24	16	3	3	3	1	1	—	1.721
4	2	125	267	351	364	307	235	161	89	65	35	25	11	5	4	3	3	2.052
5	—	33	100	129	193	177	155	132	98	67	37	32	16	14	7	2	3	1.195
6	1	6	28	43	43	66	57	65	31	32	16	10	12	5	4	2	7	428
7	—	3	2	10	18	20	15	16	10	14	11	6	3	1	4	2	3	138
8	—	—	—	4	7	2	5	7	5	4	—	2	—	—	—	—	—	36
9	—	—	—	—	—	—	3	1	—	2	2	1	1	—	—	—	—	10
10	—	—	—	—	—	—	1	2	—	—	1	—	—	—	—	—	—	4
Totali	70	596	951	1088	1085	877	687	494	290	215	120	80	46	28	20	10	16	6.673

TAVOLA XXIII.

Ammontare dei redditi in sterline	Ammontare dei patrimonî in sterline										
	1.000 2.500	2.500-5.000	5.000-10.000	10 000 15.000	15.000-20 000	20 000-25.000	25.000 50.000	50 000 75 000	75.000-100.000	100.000 – ω	Totali
200 – 300	18.710	9.035	4.037	461	116	49	39	8	2	2	32.459
300 – 500	13.440	8.537	5.653	1.244	296	106	95	21	3	2	29.397
500 – 750	4.124	3.819	3.338	1.437	521	173	160	21	5	4	13.602
750 – 1.000	1.229	1.535	1.652	829	452	242	194	12	6	3	6.154
1.000 – 1.500	739	977	1.340	770	506	328	450	51	8	8	5.177
1.500 – 2.000	211	267	512	372	264	175	388	65	16	8	2.278
2.000 – 3.000	92	142	308	295	225	157	479	149	30	30	1.907
3.000 – 4.000	25	29	66	61	66	69	237	111	32	27	723
4.000 – 5.000	14	20	30	26	33	46	115	60	38	38	420
5.000 – ω	5	14	18	15	29	45	150	143	100	299	818
Totali	38.589	24.375	16.954	5.510	2.508	1.390	2.307	641	240	421	92.935

tavola, disposta per valori singoli, sarebbe evidentemente compreso in una sola casella.

E appena necessario avvertire che nelle tavole a doppia entrata — a differenza delle serie statistiche — i due fenomeni rappresentati sono invertibili, perchè i valori di ambedue i fenomeni vi sono graduati dal minimo al massimo e non può risultarne modificato l'aggruppamento delle osservazioni.

In sostanza, coteste tavole non sono che un intreccio di due distribuzioni relative a due fenomeni. La tavola XXIII, ad es., consta della distribuzione di 92.935 individui secondo l'ammontare dei loro patrimonî (righe dell'ammontare dei patrimoni e delle osservazioni totali) e della distribuzione dei medesimi individui secondo l'ammontare dei loro redditi (colonne dell'ammontare dei redditi e delle osservazioni totali). Le dieci righe e le dieci colonne parziali, formate dal numero di volte che si presentano le diverse combinazioni di valori delle due distribuzioni, danno luogo a distribuzioni parziali per righe e per colonne coi rispettivi totali parziali.

Come le distribuzioni statistiche, così pure le tavole a doppia entrata possono riguardare *frequenze*, ossia numeri di osservazioni ragguagliati al totale generale di esse; e possono presentare osservazioni accumulate, corrispondenti a classi di valori minori o uguali, oppure maggiori o uguali, a dati limiti. In quest'ultimo caso, la costruzione non presenta alcuna difficoltà: si trasformano, anzitutto, le distribuzioni parziali contenute nelle singole colonne della tavola in distribuzioni per classi di valori minori o uguali, oppure maggiori o uguali a dati limiti, ottenendosi in tal guisa una nuova tavola; e poscia si trasformano le distribuzioni parziali contenute nelle singole righe di questa nuova tavola in distribuzioni per classi di valori minori o uguali, oppure rispettivamente maggiori o uguali a dati limiti.

Disponendo la tavola III del primo capitolo per classi di valori minori o uguali ai limiti superiori delle classi corrispondenti delle stature

TAVOLA XXIV.

Stature dei figli, minori o uguali a pollici	Stature dei padri, minori o uguali a pollici								
	59,5	61,5	63,5	65,5	67,5	69,5	71,5	73,5	75,5
61,5	—	—	1	2,5	3,5	3,5	3,5	3,5	3,5
63,5	—	1	7	15,75	25,25	27	27,5	27,5	27,5
65,5	3	7	25	69	109	122,5	127,5	127,5	127,5
67,5	3	12	45,75	142,75	269,25	336	362,5	365	365
69,5	3	14,5	63	198	414,5	587,25	667	687	688
71,5	3	14,5	64	215,5	482	730,5	875,75	918,5	924
73,5	3	14,5	64	219,5	497,5	778,5	958,5	1.022	1.029
75,5	3	14,5	65	222	501,5	793	984,5	1.057	1.066,5
77,5	3	14,5	65	222	501,5	796	989	1.065	1.074,5
79,5	3	14,5	65	222	501,5	797	991	1.068,5	1.078

dei padri e dei figli, ossia sommando le colonne parziali dall'alto in basso e poi le righe risultanti da sinistra a destra, si è ottenuta la tav. XXIV.

Inoltre — come abbiamo già osservato per gli altri tipi di rappresentazioni numeriche — notiamo che, se le tavole a doppia entrata presentano uno o ambedue i fenomeni sotto forma di attributi, occorre distinguere il caso di attributi suscettibili, dal caso di attributi non suscettibili, di espressione quantitativa. Ad es. la tavola XXI riguarda un gruppo di morti distribuito per età e per mesi; i quali ultimi, pur essendo stati originariamente espressi secondo le loro denominazioni, si sono potuti trasformare in classi, diversamente comprensive, di giorni. Così pure la seguente tavola XXV riguarda il gruppo di militari, di cui alla tavola XIII, distribuito secondo il colore degli occhi e dei capelli, che sono attributi ambedue graduabili in base all'intensità della pigmentazione; la tavola XXVI dà la ripartizione dei viaggiatori nelle nostre Ferrovie dello Stato durante l'esercizio 1927-28, secondo le zone di percorrenza (carattere quantitativo) e la classe (carattere qualitativo graduato).

TAVOLA XXV.

Colore degli occhi	Colore dei capelli				
	Rossi	Biondi	Castani	Neri	Totali
Celesti	343	9.083	17.829	3.627	30.882
Grigi	518	8.187	39.467	13.433	61.605
Castani	819	7.031	117.522	54.883	180.255
Neri	37	217	4.945	20.919	26.118
Totali	1.717	24.518	179.763	92.862	298.860

TAVOLA XXVI.

Zone di percorrenza in chilometri	III Classe	II Classe	I Classe	Totali
– 20	37.850.350	2.233.887	185.179	40.269.416
21 – 50	28.146.495	3.726.275	374.623	32.247.333
51 – 100	12.303.146	2.520.596	326.243	15.149.985
101 – 400	7.921.692	2.847.495	743.909	11.513.096
401 –	1.635.273	906.171	388.736	2.930.180
Totali	87.856.956	12.234.324	2.018.690	102.110.070

Invece un gruppo di condannati, ripartiti secondo il reato commesso e la durata della pena, riguarderebbe un attributo (il reato), che non può ricondursi ad espressione quantitativa.

In generale su questo argomento, non essendo le tavole a doppia entrata che un modo più espressivo di rappresentare le serie statistiche, valgono per le prime tutte le considerazioni, che abbiamo già esposto per le seconde.

Infine osserviamo che, come esistono serie a tre e più variabili, così può darsi il caso di tavole a tre e più entrate. Tavole siffatte possono costruirsi facilmente: suddividendo nella tavola XXIII ogni riga relativa all'ammontare dei redditi in parti corrispondenti ad es. alle età dei redditieri, si potrebbe rappresentare il numero degli individui, che presentano diverse combinazioni di età, di redditi e di patrimonî.

Così procedendo, potrebbero ottenersi tavole con un numero qualsiasi di entrate e si riuscirebbe alla determinazione di *gruppi scelti* di osservazioni, corrispondenti a date combinazioni di un numero qualsiasi di fenomeni.

TAVOLA XXVII.

			Numero dei morti nei Centri	Numero dei morti negli altri Comuni	Numero dei morti Totali
0 – 1 mese	Legittimi	M.	5.697	15.412	21.109
		F.	4.702	12.112	16.814
	Illegittimi	M.	1.002	978	1.980
		F.	893	804	1.697
1 – 12 mesi	Legittimi	M.	17.578	33.529	49.107
		F.	13.564	28.773	42.337
	Illegittimi	M.	1.268	1.515	2.783
		F.	1.367	1.324	2.691
1 – 2 anni	Legittimi	M.	8.826	16.302	25.128
		F.	8.262	15.432	23.694
	Illegittimi	M.	369	543	912
		F.	376	461	837
2 – 3 anni	Legittimi	M.	3.525	6.157	9.682
		F.	3.262	5.967	9.229
	Illegittimi	M.	106	159	265
		F.	99	168	267
3 – 4 anni	Legittimi	M.	1.743	3.222	4.965
		F.	1.704	3.131	4.835
	Illegittimi	M.	45	101	146
		F.	43	69	112
4 – 5 anni	Legittimi	M.	1.163	2.222	3.385
		F.	1.079	2.157	3.236
	Illegittimi	M.	32	62	94
		F.	25	52	77
		Totali	74.730	150.652	225.382

La tavola XXVII riporta il numero dei morti in Italia nel 1926, di età inferiore a 5 anni, distribuiti secondo l'età, la legittimità dei natali (tra gl'illegittimi sono compresi gli esposti), il sesso ed i luoghi di morte, alternativamente classificati in «Centri» e «altri Comuni».

Il fatto, che cotesti gruppi scelti non vengono sempre presentati sotto forma di tavole a più entrate, non ne modifica naturalmente l'essenza nè le caratteristiche formali.

5. Dati originari e derivati. – Spesso le rappresentazioni numeriche non riguardano dati originarî, ma dati ricavati da essi con opportuni procedimenti, in vista di scopi speciali.

Ad es., dividendo i nati vivi della tavola XIX per la popolazione media dell'anno, a cui si riferiscono, e moltiplicando i risultati per 1000, si ottengono i numeri dell'ultima colonna della tavola seguente, ossia i cosidetti quozienti (saggi o tassi) di natalità, che — come facilmente si osserva — sin dal 1925 sono diminuiti velocemente, a cagione della persistente caduta del numero assoluto dei nati vivi, malgrado l'aumento della popolazione.

Tavola XIX *bis*.

Anni	Popolazione media	Nati vivi	Quoziente di natalità per 1000 abitanti
1924	39.550.000	1.124.470	28,4
1925	39.878.000	1.109.761	27,8
1926	40.228.000	1.094.587	27,2
1927	40.581.000	1.093.772	27,0
1928	40.970.000	1.072.316	26,2

La prima e l'ultima colonna della tavola formano una rappresentazione numerica di dati derivati dalla popolazione media e dai nati vivi negli anni corrispondenti.

Si noti che anche i dati della popolazione media sono derivati, in quanto risultano da aggiornamenti annuali del censimento del 1° dicembre 1921, eseguiti aggiungendo alla popolazione censita a questa data i nati e gli immigrati e togliendo i morti e gli emigrati nel dicembre 1921 allo scopo di ricondurre la popolazione censita alla fine dell'anno 1921, e poi continuando l'aggiornamento alla fine degli anni successivi sino al 1928. La semisomma delle due valutazioni relative al principio ed alla fine di ogni anno si è assunta come popolazione media dell'anno, nell'ipotesi (come vedremo) che la popolazione fosse cresciuta in progressione aritmetica.

Analogamente, dividendo le cinque classi di debito ipotecario della seconda colonna della tavola XX per il numero delle ipoteche di ogni classe, si otterrebbe una distribuzione, secondo il saggio percentuale d'interesse dell'ammontare medio di ogni ipoteca; dividendo il numero delle vedove rimaritatesi con 0, 1,.... figli nel 1916, di cui alla tavola V, per il numero medio delle vedove viventi con 0, 1,.... figli nel medesimo anno, si avrebbe il quoziente di nuzialità dei varî gruppi; dividendo i gruppi di nascite, di cui alla seconda colonna della tavola IX, per il numero medio delle donne viventi nei gruppi di età corrispondenti, si avrebbero quozienti di fecondità femminile alle varie età; e così via. Si tenga presente che, eseguendo tale ragguaglio anche sul totale delle distribuzioni, non si ottiene un risultato uguale alla somma dei risultati delle singole classi.

D'altra parte, rilevata la lunghezza delle gambe di 300 maschi adulti emiliani e il volume del loro tronco (riportati rispettivamente nelle tavole XXVIII e XXIX), se dividiamo ogni lunghezza di gambe per il corrispondente volume del tronco, otterremo i dati riassunti nella tavola XXX, la quale rivela quel grado di dispersione dei « megalosplancnici » e dei « microsplancnici » attorno agl'individui di proporzioni normali, che secondo la scuola di Achille De Giovanni e di Giacinto Viola avrebbe notevole influenza sulle predisposizioni morbose, la fecondità, i temperamenti dei gruppi demografici (1).

TAVOLA XXVIII.

Lunghezza arti inferiori, in cm.	Individui
65 —	1
67 —	6
69 —	11
71 —	16
73 —	32
75 —	53
77 —	51
79 —	52
81 —	42
83 —	16
85 —	10
87 —	7
89 —	2
91 —	1
Totale	300

TAVOLA XXIX.

Vol. del tronco in 1000 cm.^{3}	Individui
18 —	1
20 —	3
22 —	6
24 —	38
26 —	57
28 —	52
30 —	49
32 —	28
34 —	29
36 —	19
38 —	11
40 —	3
42 —	4
Totale	300

TAVOLA XXX.

Rapp. tra arti inf. e vol. tronco	Individui
1,60 —	5
1,80 —	7
2,00 —	23
2,20 —	50
2,40 —	59
2,60 —	57
2,80 —	51
3,00 —	28
3,20 —	14
3,40 —	2
3,60 —	3
3,80 —	1
Totale	300

(1) Cfr. il nostro studio: *La statistica e l'antropometria clinica*, in « Endocrinologia e patologia costituzionale », giugno 1932.

Analoghe elaborazioni possiamo eseguire sulle serie a tre e più variabili e sulle tavole a più entrate. Ragguagliando tutti i dati della tavola IV alla popolazione media regionale del 1927, si avrebbe la serie per abitante del gettito delle tre imposte dirette, della forza motrice nelle industrie e del numero degli addetti alle industrie, che potrebb' essere sensibilmente diversa da quella originaria. Ragguagliando i gruppi parziali di morti della tavola XXI alle medie dei viventi alle età ed ai mesi corrispondenti del 1916, avremmo una tavola a doppia entrata di quozienti di mortalità, dove naturalmente i totali non sarebbero uguali alle somme dei quozienti singoli.

Capitolo II.

LE RAPPRESENTAZIONI GRAFICHE

1. Rappresentazioni cartesiane nel piano. Diagrammi lineari. – L'accertamento della relazione intercedente tra i valori di un fenomeno ed il numero di volte che essi si presentano, o tra i valori di due o più fenomeni espressi in forma di serie statistiche o di tavole a più entrate, è pure agevolato da rappresentazioni geometriche nel piano e nello spazio.

Limitandoci, anzitutto, alle rappresentazioni cartesiane ortogonali nel piano, ricordiamo che esse si fondano su una coppia di rette (assi) ortogonali, su ognuna delle quali è stabilito il verso e l'unità di misura.

Chiamati con X (asse orizzontale) e con Y (asse verticale) i due assi ortogonali e con O il punto d'intersezione di essi, si considera sul primo asse come positivo il semiasse posto a destra di O e negativo quello posto a sinistra; e sul secondo asse come positivo il semiasse superiore e negativo quello inferiore. Così il piano resta diviso in quattro quadranti, nei quali si possono rappresentare tutte le possibili coppie di numeri reali.

Infatti, scelto un segmento rettilineo di data lunghezza come unità di misura (modulo) di X ed un segmento rettilineo di lunghezza uguale o diversa come unità di misura (modulo) di Y; fissati poscia sui due assi, con l'origine in O, i segmenti corrispondenti (in lunghezza e verso) ad una coppia di numeri reali, le perpendicolari dagli estremi dei due segmenti agli assi rispettivi, intersecandosi, individuano uno ed un solo punto nel piano. Ripetendo l'operazione per altre coppie di numeri reali, si ottiene una successione di punti, idonea a rappresentare univocamente la data successione di coppie di numeri.

Coteste coppie sono le coordinate dei punti corrispondenti, rispetto ad O (origine delle coordinate); e precisamente si chiamano ascisse i numeri rappresentati sull'asse X ed ordinate quelli rappresentati sull'asse Y.

Ora, assumendo come coordinate le coppie di numeri di una distribuzione statistica e per maggior chiarezza congiungendo fra loro con seg-

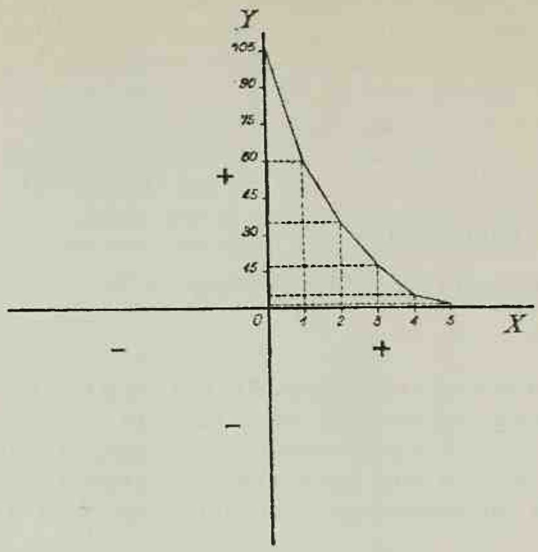


Fig. 1

menti rettilinei i punti corrispondenti, si ottiene una spezzata, il cui esame suole essere di molto ausilio nella determinazione della relazione intercedente tra i valori osservati del fenomeno ed il numero di volte, che essi si presentano.

Con questo metodo abbiamo rappresentato su un foglio di carta millimetrata e riprodotto nella figura 1, la distribuzione della tavola V, ponendo 1 figlio = 5 mm. e 1 vedova = 1/3 mm.

La somma dei segmenti, condotti perpendicolarmente dagli estremi e dai vertici della spezzata all'asse delle ascisse, è misurata dal numero totale delle osservazioni. Malgrado che questo numero sia sparuto, la spezzata mostra una netta concavità verso la direzione positiva dell'asse delle ordinate ed una tendenza asintotica, facilmente spiegabile, verso l'asse delle ascisse. Lo stesso è a dire, in generale, per i divorzî, come si rileva dalla figura 2, relativa ai divorzî, in migliaia, avvenuti in Prussia nel 1925.

Questo procedimento grafico può seguirsi per la rappresentazione delle serie statistiche formate da coppie di valori e per la rappresentazione delle distribuzioni accumulate: basta sostituire ai numeri sopra considerati rispettivamente le coppie di numeri della serie statistica o della distribuzione accumulata.

La figura 3 rappresenta in scala conveniente la serie mensile della « bilancia mercantile » del

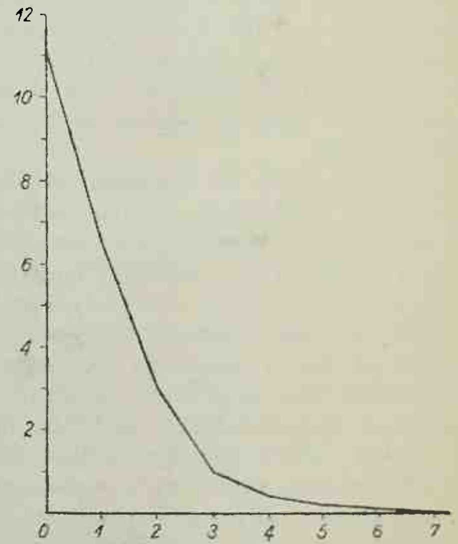


Fig. 2

nostro paese nel 1931 e nel 1932, e dimostra come nei mesi autunnali di cotesti anni si sia avuto un eccezionale, sebbene leggero eccesso di esportazioni sulle importazioni.

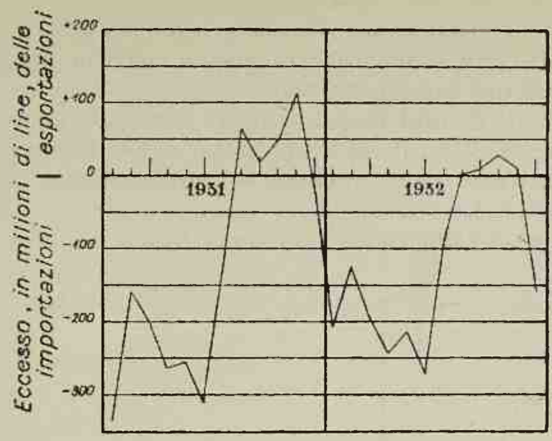


Fig. 3

La figura 4 rappresenta, invece, in scala conveniente, la distribuzione accumulata delle stature di cui alla tavola VII *bis*, ogni ordinata essendo data dal numero delle osservazioni aventi valori minori o uguali al valore dell'ascissa corrispondente. È facile connettere l'asimmetria della spezzata risultante alla leggera prevalenza delle alte stature nel complesso della popolazione del nostro paese.

Per brevità, in queste ultime figure abbiamo posto l'origine delle ascisse in corrispondenza al valore più piccolo di esse; in altri esempî avremmo potuto operare ugualmente per le ordinate.

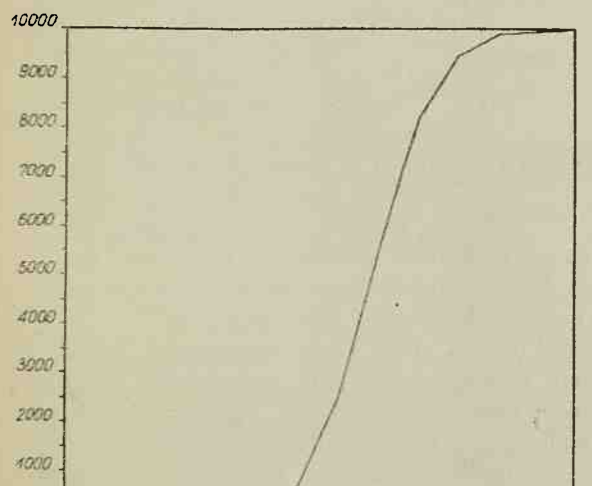


Fig. 4

Si noti che la fig. 4 rappresenta la proporzione, su diecimila, degl'individui con statura inferiore o uguale a dati limiti, e quindi permette di determinare, almeno approssimativamente, la statura che non è superata da una data frazione degl'individui esaminati: ad es. da $^1/_4$ (cioè 2.500) e così via.

La curva risultante da questo tipo di rappresentazione grafica è stata chiamata da Galton « ogiva » [1].

(1) Francis Galton, *Natural Inheritance*. London, Macmillan, 1889.

2. Diagrammi areali o istogrammi. – Se una distribuzione consta di classi consecutive di valori, conviene fissare sull'asse delle ascisse, a partire dall'origine, segmenti adiacenti che corrispondano ai valori compresi in ogni classe. Costruendo su ogni segmento un rettangolo, avente come base il segmento stesso e come altezza il segmento — misurato sull'asse delle ordinate — corrispondente al quoziente tra il numero delle osservazioni della classe e il numero delle unità di misura comprese nella classe medesima, l'area dei successivi rettangoli sarà misurata dal numero delle osservazioni di ogni classe.

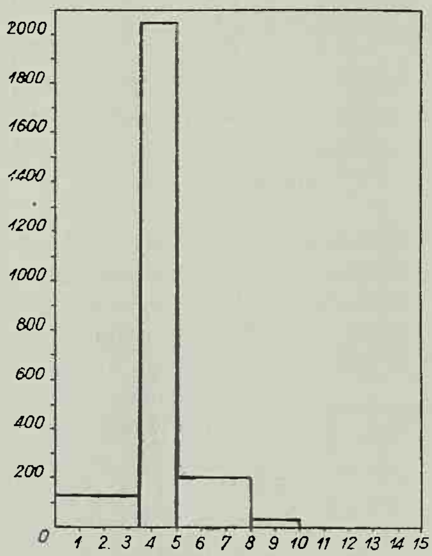


Fig. 5

Prescindendo dai segmenti comuni, la successione di rettangoli così costruita determinerà una figura, avente un'area misurata dal numero totale delle osservazioni, e darà luogo ad una spezzata, la cui forma fornirà la rappresentazione geometrica cercata.

Per la corretta interpretazione di cotesto tipo di rappresentazione grafica, da Pearson chiamato istogramma (1), dobbiamo avvertire che esso è tanto meno veritiero, quanto più ampio è l'intervallo delle classi.

Con esso abbiamo rappresentato (figura 5) la tavola XX, dopo aver diviso la prima classe dell'ammontare del debito ipotecario per 3,5, la seconda per 1,5, la terza per 3, la quarta per 2 e la quinta per 5. Risulta chiarissima la enorme prevalenza del debito ipotecario fruttante un interesse intorno al 4 per cento, saggio prevalente e tipico nel periodo in cui è stato stipulato.

S'intende facilmente che, ove le classi siano ugualmente estese, è superfluo dividere il numero corrispondente delle osservazioni per l'intervallo costante delle classi, perchè l'altezza dei successivi rettangoli si abbasserebbe proporzionalmente e la forma della spezzata resterebbe inalterata e non farebbe modificare il giudizio sulla forma della relazione.

Con questo criterio è stata rappresentata (figura 6) la distribuzione di stature della tavola VII. Lo scarsissimo numero di osservazioni comprese

(1) Karl Pearson, *Skew variations in homogeneous material*. In « Phil. Trans. of the Royal Society », 1895.

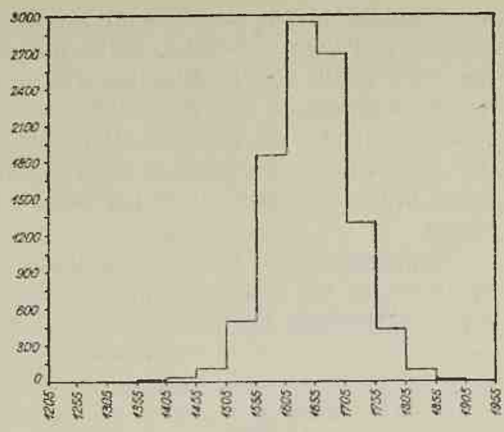


Fig. 6

nelle classi estreme ha consentito di considerare come 1.205 e 1.955 mm. le stature estreme della distribuzione.

Si noti, però, che mediante osservazioni accumulate si è potuta rappresentare graficamente cotesta distribuzione, e si sarebbero potute inoltre rappresentare le tavole IX e X, senza prescindere dalla prima classe nè ammettere per essa un limite inferiore arbitrario; e che avremmo potuto ancora rappresentare la tavola VI senza prescindere dall'ultima classe nè ammettere per essa un limite superiore arbitrario.

La rappresentazione grafica della tavola X a mezzo di osservazioni accumulate eviterebbe pure il forte arbitrio inerente all'ampiezza eccessivamente grande delle ultime classi.

Quanto alla tavola IX, osserviamo che essa potrebbe comunque rappresentarsi trascurando le età ignote, perchè, nell'ipotesi che esse si ripartissero a un dipresso proporzionalmente tra le varie classi della distribuzione — ipotesi *a priori* tanto più plausibile, quanto minore è la frequenza del gruppo dei valori ignoti —, la forma del diagramma non ne sarebbe sensibilmente alterata.

Se la seconda colonna della distribuzione riguardi quozienti tra due gruppi di osservazioni, si attribuirà il quoziente di ogni classe al valore centrale di essa e si costruirà un diagramma lineare; ma occorrerà andar cauti nell'interpretazione dei risultati.

3. Scale logaritmiche e numeri indici. – È proprietà dei logaritmi che a differenze uguali di essi corrispondono rapporti uguali tra i numeri; ond'è che, se in un diagramma cartesiano rappresentiamo le ordinate sostituendo ai numeri i loro logaritmi, ricavati da una comune tavola a pochi decimali, le differenze tra due ordinate logaritmiche permettono di valutare i rapporti tra due gruppi o valori osservati.

È stato notato che un diagramma a scala naturale può fare attribuire uguale importanza a variazioni notevolmente diverse e viceversa: ad es., nella tavola XIV si rileva che nel settennio 1890-1896 la stazza netta delle nostre navi a vela è diminuita quasi nella stessa misura che nel settennio 1907-1913, e precisamente in ambedue i periodi si rileva una diminuzione di poco più di 100 mila tonnellate (106,5 nel primo e 112,7 nell'altro settennio). In un diagramma a scala naturale le differenze fra quelle coppie di ordinate sarebbero rappresentate quasi dal medesimo segmento e non riuscirebbe evidente il fatto che la diminuzione relativa è stata nei due periodi notevolmente diversa (rispettivamente del 17 e del 24 per cento), ossia che il rapporto tra la settima e la prima ordinata è stato di 83 per cento nel primo periodo e di 76 nel secondo.

Rappresentando, invece, le ordinate su scala logaritmica (figura 7), la differenza tra le dette coppie di ordinate risulta sensibilmente maggiore nel secondo periodo che nel primo, e, trasportando tali differenze

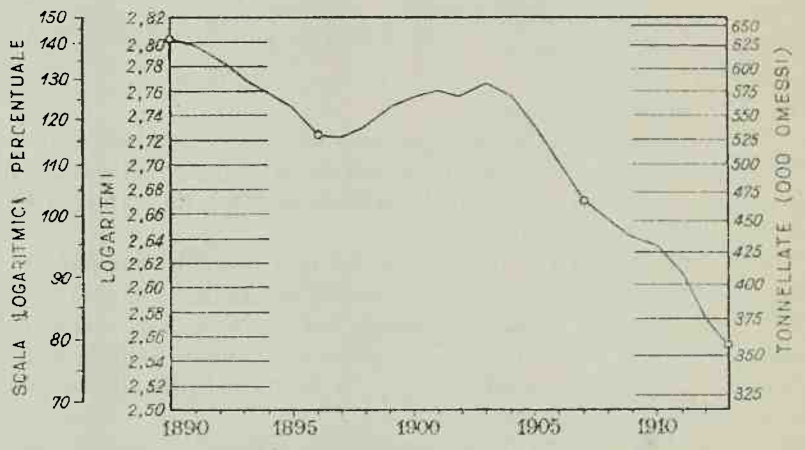


Fig. 7

sul ramo inferiore della scala esterna della figura, a partire dall'origine 100, è facile valutarle rispettivamente a poco meno di 85 ed a circa il 75 per cento: relativamente al livello raggiunto, la diminuzione delle navi a vela nel settennio precedente la guerra mondiale è stata in Italia sensibilmente maggiore della diminuzione osservatasi dal 1890 al 1896.

Analogamente potrebbe valutarsi a poco più del 10 per cento il transitorio aumento del periodo 1896-1903.

L'uso della scala logaritmica va diffondendosi nella pratica statistica, sebbene sia necessaria la scala dei rapporti, se si vogliono valutare numericamente le distanze verticali logaritmiche. Tale scala si ottiene facilmente, bastando trovare i logaritmi dei rapporti percentuali.... 70, 80,.... 100, 110,..., ossia 1,84510; 1,90309;... 2; 2,04139... e *costruirne la scala con lo stesso modulo usato per i logaritmi dei valori* *osservati*. Talvolta essa è costruita nella forma di scala angolare.

È forse inutile soggiungere che, usando scale logaritmiche, i confronti possono anche istituirsi tra spezzate relative a fenomeni diversi; e che, spostando la scala dei rapporti in modo che il rapporto 100 corrisponda ad un dato valore osservato, possono leggersi rapidamente su di essa gli altri valori in percentuali di quello scelto, ossia i cosidetti *numeri indici*, rispetto ad una data *base*.

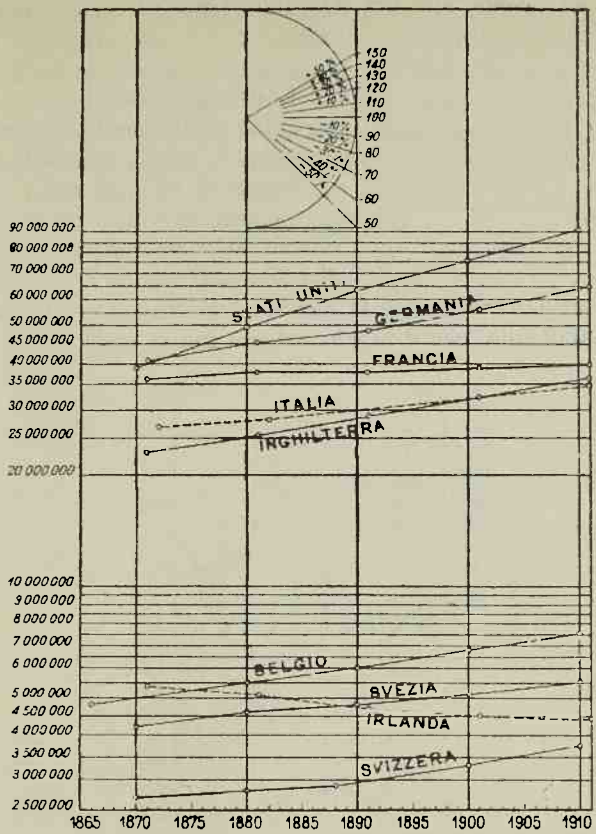


Fig. 8

Notiamo che la nostra figura 7 riporta a sinistra la scala dei logaritmi e a destra la scala corrispondente dei numeri con modulo necessariamente descrescente al crescere di essi; in generale, però, basta riportare solo la seconda scala — che, mentre mette in luce il valore numerico dei dati osservati, con l'andamento del modulo si rivela logaritmica — e quella logaritmica percentuale.

Altro esempio è fornito dalle serie storiche dei censimenti: rappresentando su una medesima scala logaritmica (fig. 8) i risultati dei censimenti di un gruppo di Stati dal 1865 al 1911 (a parità di territorio), è facile rilevare i maggiori aumenti relativi della popolazione degli Stati Uniti, specie nei primi due decennî; la stasi persistente della popolazione francese; la diminuzione di quella irlandese. La scala logaritmica percentuale, che abbiamo riportata nella doppia forma di scala lineare ed angolare e costruita allo scopo di confrontare i dati di due decennî consecutivi, fa valutare facilmente le variazioni percentuali: ad es., nella figura si rileva ad occhio che nel decennio 1881 - 91 la popolazione irlandese diminuì di circa il dieci per cento (1).

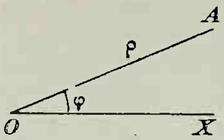
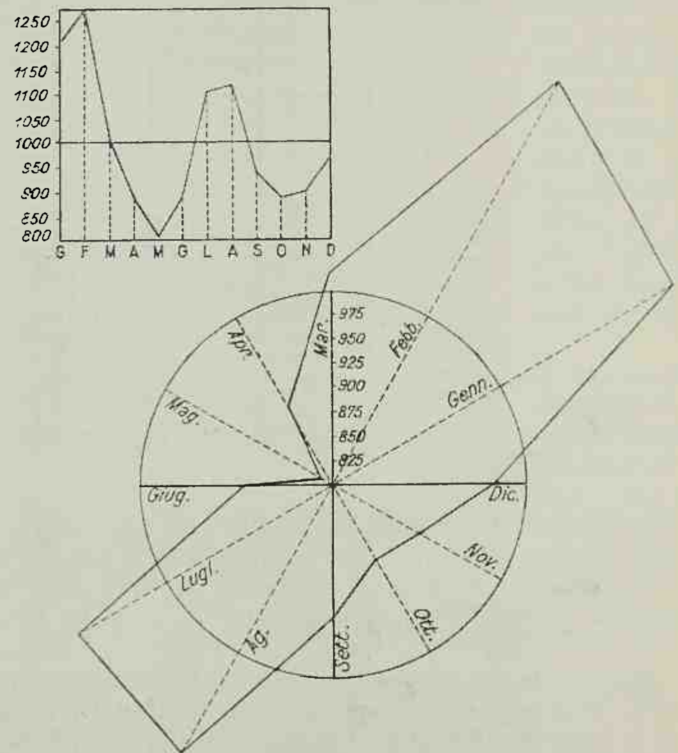


Fig. 9

4. Coordinate polari. – Oltre alle coordinate cartesiane ortogonali, sogliono usarsi altri tipi di coordinate, ad es. le cosidette coordinate polari.

Fissando su una retta X un punto O (polo), ricordiamo che un punto qualsiasi A del piano riesce pure univocamente determinato dalla sua distanza dal punto O (raggio vettore) e — fissato un dato verso per la misura degli angoli — dall'angolo (angolo polare o anomalia) che il raggio vettore forma colla semiretta OX. Il raggio vet-



Figg. 10 e 11

(1) G. H. Knibbs, *The mathematical theory of population, of its character and fluctuations, and of the factors wich influence them.* Melbourne, 1917.

Col semplice procedimento grafico esposto nel testo, a pag. 29 e segg. del volume citato si eseguono interessanti confronti tra un gran numero di Stati, a cominciare dal 1800.

tore ρ e l'anomalia φ sono le due coordinate polari del punto A (fig. 9). La semiretta OX dicesi asse polare.

Le figure 10 e 11 rappresentano, in coordinate rispettivamente cartesiane ortogonali e polari, le frequenze (relative) mensili — moltiplicate per 12.000 — dei morti in Italia nel 1923. Nella seconda figura ogni angolo rappresenta un mese, ogni raggio vettore la frequenza corrispon-

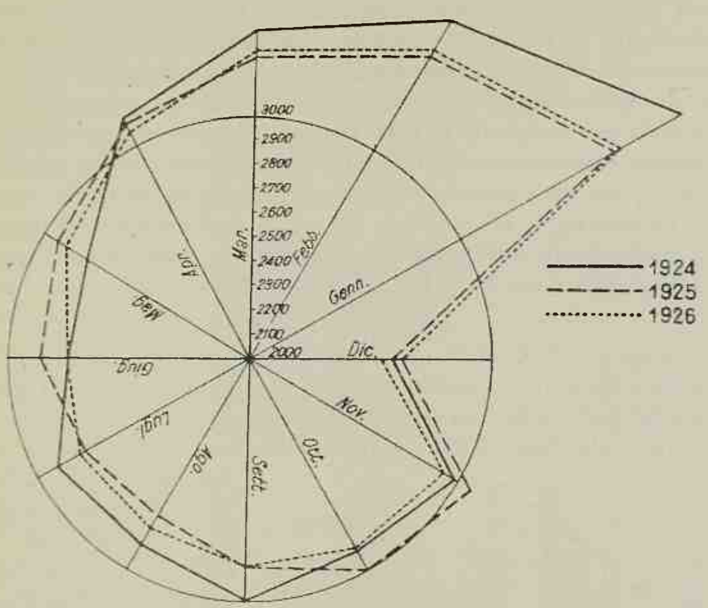


Fig. 12

dente dei morti; e, poichè i dati mensili dei morti sono stati ridotti a 30 giorni, — essendo stati divisi per il numero di giorni del mese corrispondente e poi moltiplicati per 30 — gli angoli sono stati costruiti tutti di 30 gradi.

Rispettando il valore 800 come origine della scala dei raggi, abbiamo sovrapposto al diagramma un cerchio di raggio 1000, corrispondente all'ipotesi di una equidistribuzione mensile dei morti; cerchio, che nella figura cartesiana corrisponde all'ordinata 1000 e ci permette di valutare l'intensità di quei massimi invernali e, in minore misura, estivi, che sono la caratteristica dell'andamento stagionale delle morti in Italia ed in altri paesi con clima non molto diverso.

La figura 12 rappresenta le medie giornaliere dei nati vivi nei successivi mesi del triennio 1924-1926 (moltiplicando tali medie per 30 si avrebbero le medie mensili, ridotte ad ugual numero di giorni) e mette in luce il vantaggio che per le serie periodiche presenta la rappresentazione polare, allorquando interessi di esaminare l'andamento generale di esse e quello ciclico.

Infatti le coordinate polari, mentre permettono di seguire il primo andamento attraverso lo svolgimento di una spezzata continua, danno luogo ad una sovrapposizione di cicli, che permette un confronto immediato di questi ultimi.

Dalla figura si rileva una tendenza all'involuzione della spezzata dei nati, rivelatrice di quella sinistra tendenza alla diminuzione delle nascite, che sin da quel triennio cominciava a notarsi nel nostro paese; e si rileva altresì — malgrado il mal costume, ormai in parte represso, di fare apparire come nati nel gennaio un gran numero di nati nel dicembre precedente e la conseguente perturbazione della spezzata in cotesti due mesi — l'esistenza di un massimo invernale-primaverile e di altro massimo, più attenuato, autunnale, che formano la caratteristica dell'andamento stagionale delle nascite nel nostro paese.

5. Cartogrammi e altri tipi di rappresentazione nel piano. — Per rendere più sinteticamente espressive le rappresentazioni geometriche, sogliono seguirsi diversi espedienti, di cui la figura 13 dà un esempio. Essa rappresenta dieci distribuzioni statistiche, riguardanti la composizione per età della popolazione italiana negli anni dal 1921 al 1925, distintamente pei due sessi (1): a mezzo di opportuni raggruppamenti dei rettangolini relativi ai dieci diagrammi areali, e di una conveniente distribuzione di bianco e nero in ognuno di essi, si può facilmente fissare lo sguardo sulle forme di ogni distribuzione, sulle varianti che tali forme presentano nei due sessi, sull'andamento dinamico di esse. In particolare notiamo che la guerra, avendo ar-

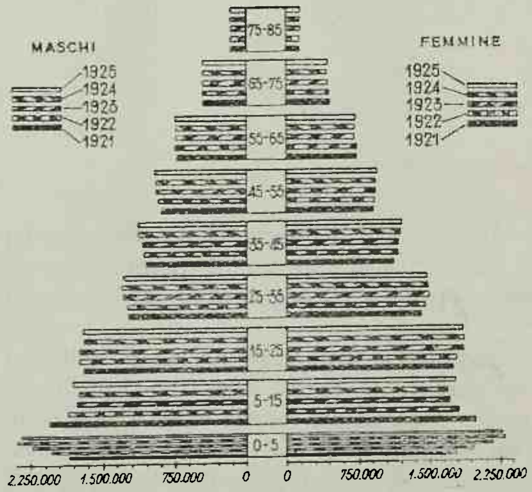


Fig. 13

(1) Cfr. il nostro studio: *L'età dei viventi e i saggi di mortalità*, in «Rivista it. di statistica», 1929, pag. 207.

restato quasi completamente il flusso migratorio, non produsse nel nostro paese, come altrove, uno squilibrio dei sessi; che, inoltre, la classe di età 0-5 anni presentò dal 1921 al 1924 per ambedue i sessi un forte incremento, per il fatto che andava riversando nella classe successiva le poco numerose schiere dei nati di guerra; e che nelle classi di 35-55 anni è facile scorgere gli effetti della già profonda contrazione del flusso emigratorio post-bellico.

Invece del bianco e nero e di semplici variazioni di disegno, avremmo potuto impiegare diversi colori.

L'esempio esposto fornisce una variante del metodo cartesiano, che rende immediato il passaggio ai tipi di rappresentazione grafica, fondati esclusivamente sull'impiego di figure, scale cromatiche, ecc. Cotesti tipi, comunemente chiamati *cartogrammi*, possono costruirsi in ogni caso, ma debbono necessariamente adottarsi, quando si vogliano eseguire rappresentazioni grafiche di serie a due variabili, una delle quali sia un attributo non riducibile a forma quantitativa.

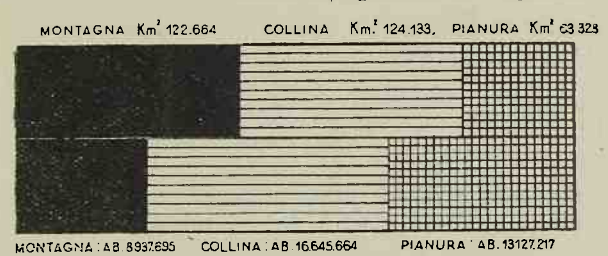


Fig. 14

Con la figura 14 ne diamo un esempio semplice: essa rappresenta la ripartizione, nelle zone di montagna, collina e pianura, della superficie territoriale del nostro paese (in alto) nel 1921, entro i nuovi confini, e della popolazione di esso (in basso) secondo il censimento del 1921. Il semplice avvedimento di scegliere le due scale in guisa da uguagliare i segmenti rappresentativi del totale della superficie e della popolazione permette di valutare ad occhio la nota preferenza della collina e principalmente della pianura da parte di quest'ultima.

Usando di una sola scala, abbiamo rappresentato con la figura 15 la superficie e il numero di abitanti delle cinque parti della Terra intorno al 1928 (¹).

Con l'aiuto di un *goniometro centesimale* abbiamo ancora costruito la figura 16, in base alla ripartizione percentuale della superficie territoriale del nostro paese secondo la destinazione e la qualità delle culture nel 1931. Vi si nota che l'incolto improduttivo (fabbricati, strade, acque, ferrovie, tranvie e sterili) forma una piccola frazione della superficie territoriale e la rimanente « superficie agraria e forestale » è prevalente-

(¹) I dati sono stati calcolati dall'Institut International de Statistique e pubblicati nel suo *Aperçu de la démographie des divers pays du monde*. La Haye, 1929.

mente destinata a seminativi (cereali, leguminose, piante industriali, da tubero, ortaggi e culture foraggere). La cultura del frumento vi occupa il 15,51 per cento.

Nella figura 17, i segmenti verticali AB e AE rappresentano rispettivamente la popolazione italiana al 30 giugno 1911 ed il numero dei nati vivi (moltiplicato per 100) in Italia nel 1911. Scegliendo arbitra-

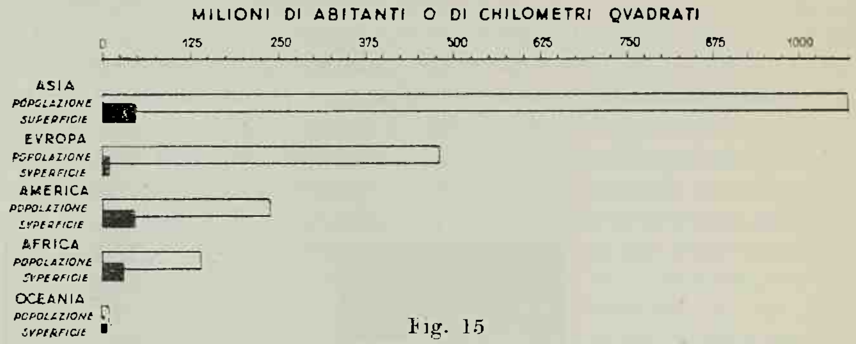


Fig. 15

riamente sull'asse orizzontale un segmento unitario AF, congiungendo l'estremo F di esso con B, e conducendo per il punto E la parallela a BF, si determina il punto H di ascissa 3,15, uguale al quoziente percentuale di natalità nel tempo e luogo considerati: è questo il procedimento di divisione grafica, che trovasi esposto in tutti i trattati di « calcolo grafico », come immediata applicazione del teorema di Talete; e da cui, invertendo, si ricava il procedimento di moltiplicazione grafica.

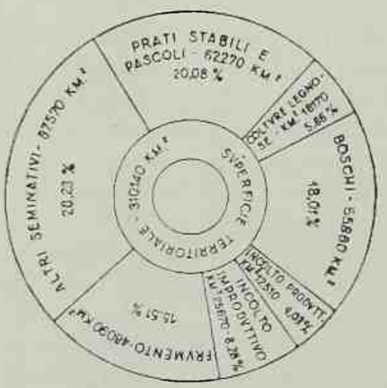


Fig. 16

Il rettangolo risultante $A\,B\,M\,H$, di base uguale al quoziente di natalità, di altezza uguale alla popolazione e di area uguale al numero dei nati, dà di questi tre elementi una veduta sintetica, che può porsi efficacemente a confronto con altra analoga, ad es. con quella che si ricava dal rettangolo $A\,B\,M'\,H'$, ottenuto con lo stesso procedimento in base al numero di morti (moltiplicato per 100), rappresentato dal segmento $A\,E'$. Sovrapponendo, ancora, i rettangoli $A\,C\,L\,G$ ed $A\,C\,L'\,G'$ relativi al 1928 (nuovi confini) risultano chiari da cotesti rettangoli — se non la diminuzione assoluta delle nascite nel tempo, data la tenuità di essa nei due anni posti a confronto — quel mancato incremento di nascite, malgrado l'aumento notevole della popolazione, quella sensibile diminuzione di morti, quell'aumento nell'eccedenza assoluta

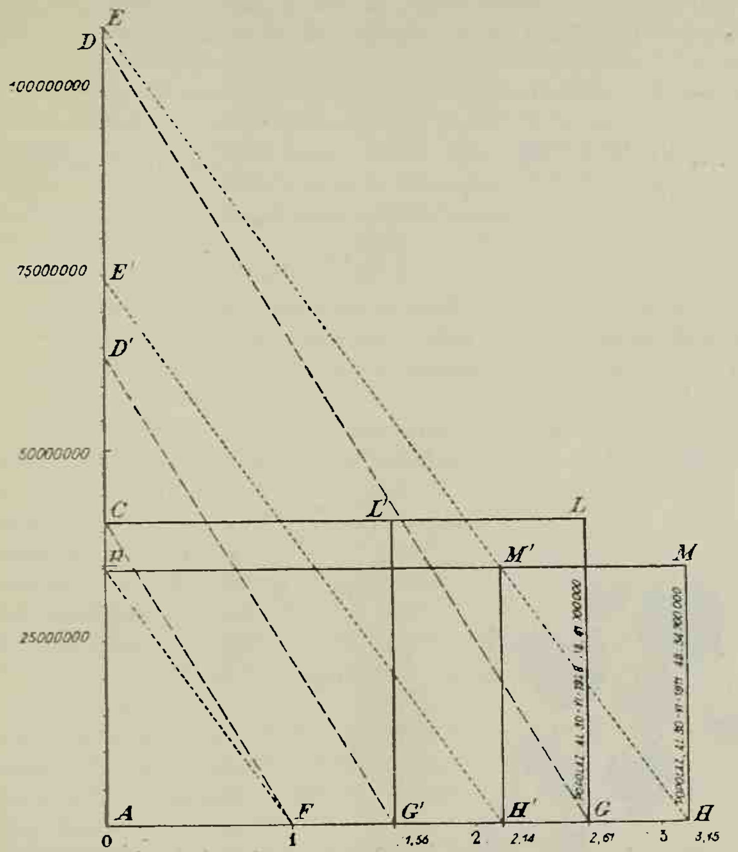


Fig. 17

delle prime sulle seconde, quella diminuzione generale dei quozienti, anche di natalità, che sono stati i dati più salienti della crisi demografica del nostro paese. Negli ultimi anni la situazione è peggiorata.

Naturalmente i confronti riescono più efficaci a mezzo di figure geometriche simili, ma allora bisogna rinunciare alle vedute d'insieme e limitare i confronti ad un solo fenomeno.

Per la costruzione geometrica di quadrati e cerchi, le cui superfici siano misurate da numeri assegnati, potrà seguirsi il procedimento esemplificato nella figura 18. In essa sono riportati su un asse orizzontale, a partire dall'origine O, segmenti proporzionali alla popolazione italiana censita nei centri non superiori a 2000 ab., nei centri in totale e nel Regno al 1° dicembre 1921.

Conducendo la circonferenza, di cui il primo segmento (OA') è il diametro, la distanza OA tra l'origine dell'asse e uno dei due punti di

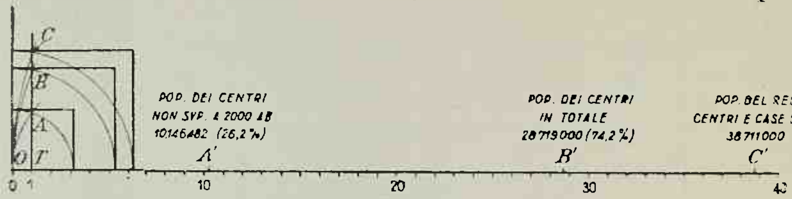


Fig. 18

incontro della circonferenza con la perpendicolare all'asse condotta per il punto di ascissa 1, è la radice quadrata del numero che misura quel primo segmento. Per convincersene, basta tener presente che: $OT/OA = OA/OA'$ e che $OT = 1$.

Pertanto, il quadrato di lato OA ha una superficie misurata (nella unità scelta) dal numero di abitanti nei centri non superiori a 2000 ab.; e similmente gli altri due quadrati della figura, costruiti in base alla popolazione dei centri ed a quella complessiva del Regno e pienamente comparabili.

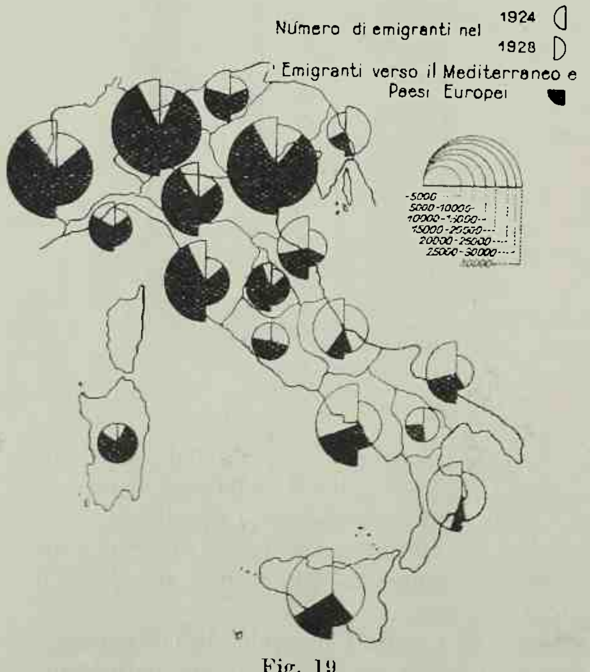


Fig. 19

Tenendo presente che l'area di un cerchio è proporzionale al quadrato del raggio, è ovvio che, anzichè i quadrati, avremmo potuto considerare i cerchi di raggio proporzionale ai lati dei quadrati medesimi o i quarti di tali cerchi, che trovansi disegnati nella figura.

Un ultimo esempio: la figura 19 è una cartina geografica dell'Italia, che in corrispondenza ad ogni regione riporta dei semicerchi accostati, rappresentanti con la loro superficie il numero di emigranti nel 1924 (semicerchio di sinistra) e nel 1928 (semicerchio di destra). La scala è stata ottenuta, facendo corrispondere alle sette classi, in cui sono stati raccolti per brevità i numeri regionali degli emigranti, le radici quadrate dei numeri da 1 a 7.

Sebbene il limite superiore dell'ultima classe sia ignoto, si è ritenuto trascurabile l'errore di ritenere che cotesta classe abbia un'ampiezza uguale a quella delle precedenti. Ogni semicerchio, inoltre, è diviso in due parti, le cui superfici stanno nel rapporto, in cui gli emigranti verso paesi transoceanici stanno a quelli diretti verso paesi del bacino del Mediterraneo e altri paesi d'Europa. Si nota la diminuzione del flusso emigratorio in tutte le regioni e specie in quelle che davano il maggior contributo all'emigrazione transoceanica (fa eccezione solo la Venezia Giulia),.

Negli atlanti geografici s'incontrano svariati esempi di quest'ultimo tipo di cartogramma, detto appunto geografico. In esso per lo più si colorano più o meno intensamente o diversamente le superfici corrispondenti alle varie parti di territorio, in base ad un largo e spesso elegante impiego di scale a tinte graduali o cromatiche; spesso troviamo, pure, rotte di piroscafi o linee ferroviarie, segnate più o meno intensamente o colorate a seconda dell'intensità del traffico mercantile, e talvolta rappresentate addirittura da nastri composti di strisce rappresentative della composizione del traffico (cartogrammi a nastro).

Altre forme di rappresentazione grafica, per quanto interessanti e spesso suggestive, non hanno importanza scientifica: sono per lo più figure umane o animali o simboliche, oggetti varî che, con le loro dimensioni od a mezzo di diversi colori, vengono assunte a rappresentare un gruppo di quantità o di attributi.

6. Rappresentazioni cartesiane nello spazio. Stereogrammi. — Per rappresentare geometricamente una serie a più di due variabili od una tavola a parecchie entrate, occorre fare rappresentazioni nello spazio.

A tal uopo — considerando, nel nostro spazio a tre dimensioni, serie a tre variabili o tavole a doppia entrata — ricordiamo che, fissati su due assi cartesiani ortogonali, con l'origine in O, i segmenti corrispondenti (in lunghezza e verso) ad una coppia di numeri reali, le perpendicolari dagli estremi dei due segmenti agli assi rispettivi, intersecandosi, individueranno uno ed un sol punto nel piano; e che, conducendo da cotesto punto un segmento corrispondente (in lunghezza e verso) a un altro numero reale, rappresentato su un terzo asse passante per il punto d'intersezione dei primi due e perpendicolare al piano, su cui essi giacciono, l'estremo di questo terzo segmento determinerà univocamente un punto nello spazio.

Cotesto punto avrà, pertanto, tre coordinate e sarà idoneo a rappresentare una terna di numeri reali.

Con questo procedimento, che dà luogo ai cosidetti stereogrammi, abbiamo rappresentato (figura 20) la tavola IV, prescindendo dalle ultime due terne di valori, che avrebbero ingrandito troppo la figura. Ab-

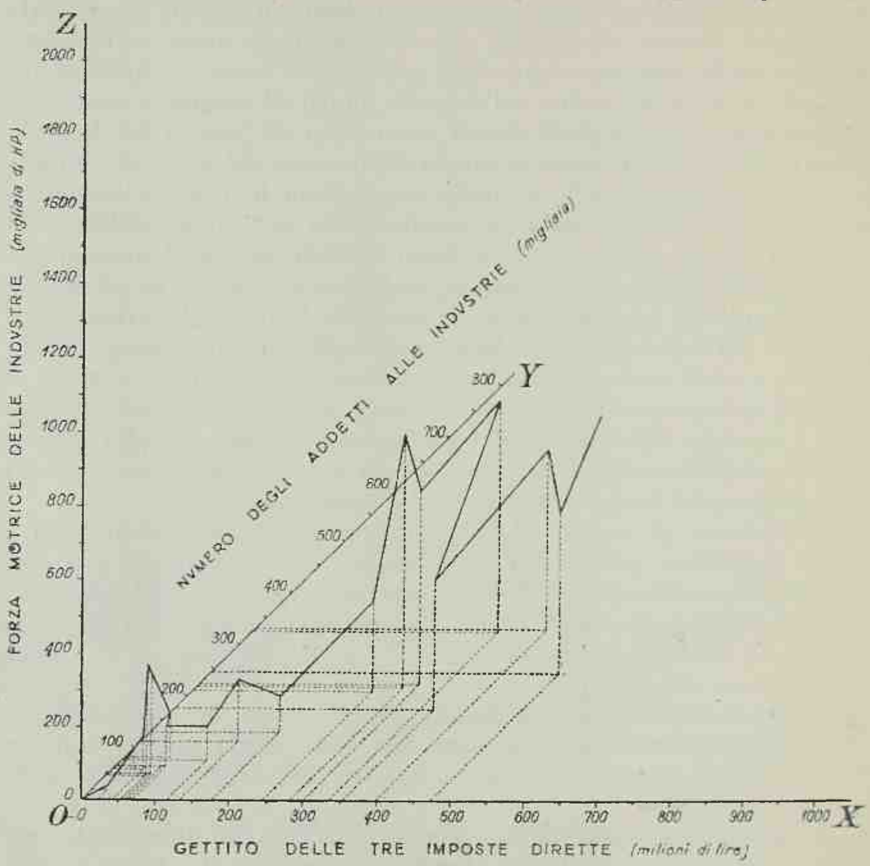


Fig. 20

biamo scelto lo stesso modulo per il gettito delle imposte (in milioni di lire) ed il numero degli addetti (in migliaia), un modulo dimezzato per la forza motrice (in migliaia di cavalli-vapore); e si vede chiarissima la tendenza alla solidarietà d'andamento dei tre fenomeni.

Congiungendo convenientemente con segmenti rettilinei i punti individuati nella figura, si ottiene un grafico molto complicato e che ancor più si sarebbe complicato, se tutti o parte dei valori considerati si fossero presentati più di una volta.

Meno complesse riescono per lo più le rappresentazioni geometriche delle tavole a doppia entrata, perchè, facendo corrispondere i numeri di osservazioni delle caselle ai segmenti perpendicolari al piano di base della figura, e congiungendo fra loro i punti estremi di quei segmenti, che corrispondono alle distribuzioni parziali per righe e per colonne, si ottiene in generale una superficie poliedrica più o meno regolare (1).

La somma dei segmenti perpendicolari al piano di base sarà misurata dal totale delle osservazioni.

Un esempio è offerto dalla figura 21, che rappresenta su scale convenienti la tavola XXII e mette in luce un regolare digradare del numero delle abitazioni intorno a quella combinazione di quattro stanze e cinque abitanti, che era a un dipresso il *tipo* delle abitazioni nel tempo e luogo considerati.

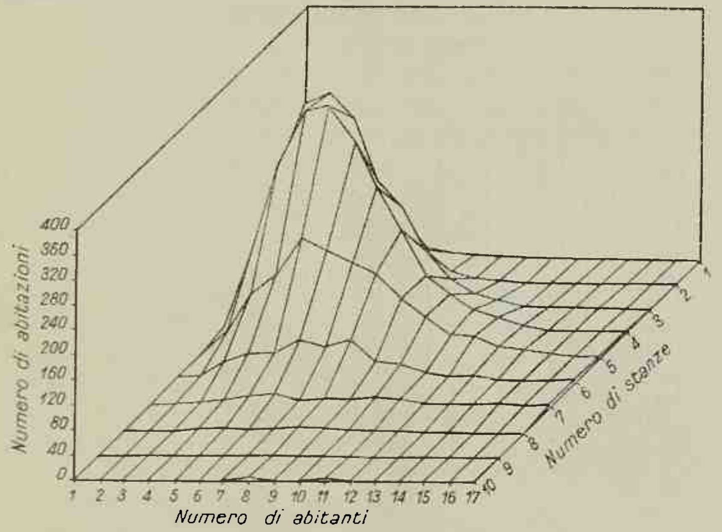


Fig. 21

Se la tavola a doppia entrata riguarda classi di valori, con procedimento analogo a quello esposto per le distribuzioni, si costruiranno prismi contigui, aventi come base rettangoli formati dai segmenti corrispondenti alle classi di valori rappresentati sui primi due assi, e come altezza i segmenti corrispondenti ai quozienti tra il numero di osservazioni di ogni coppia di classi e il prodotto delle unità delle due classi

(1) Luigi Perozzo, *Della rappresentazione grafica ed in particolare dei diagrammi a tre coordinate*, in « Annali di statistica ». 1880; e *Stereogrammi demografici*, in « Annali di Statistica », 1881. Cfr. anche G. Udny Yule, *An introduction to the theory of statistics*. London, Griffin, 1916.

rispettive. Ove si prescinda dalle facce comuni, il volume del solido formato da cotesti prismi sarà misurato dal totale delle osservazioni, e la superficie poliedrica di esso darà la forma geometrica della relazione.

La figura 22 è stata costruita in tal modo e rappresenta la tavola XXI, dove appunto i valori della circostanza « tempo » procedono per mesi, che sono formati da un diverso numero di giorni, e quelli dell'età dei morti procedono per classi sempre più estese. Il numero di osserva-

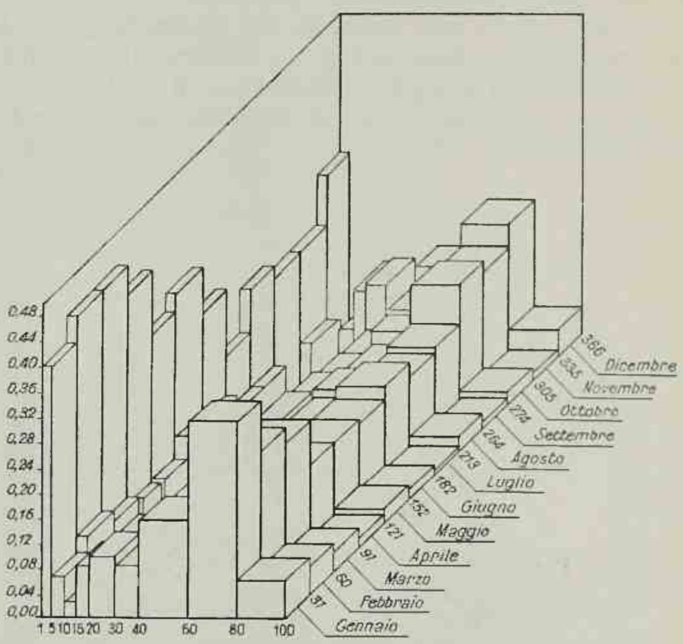


Fig. 22

zioni della prima linea e della prima colonna è stato diviso per 31×1, quello della prima linea e della seconda colonna per $31 \times 4 = 124$; quello della seconda linea e della seconda colonna per $29 \times 4 = 116$, ecc. Per economia di spazio, abbiamo omesso di rappresentare i morti di 0-1 anno d'età.

Malgrado il piccolo numero di casi considerati, la figura fornisce la forma caratteristica della distribuzione dei morti per età, col noto massimo iniziale e con quello intorno al 70° anno di età; e permette di abbracciare con uno sguardo le perturbazioni che tale forma presenta nei successivi mesi dell'anno, passando attraverso le alte quote dei mesi invernali e il piccolo e breve rialzo estivo.

Però, quanto più ampio è l'intervallo delle classi, tanto meno affidante ne sarà la rappresentazione geometrica; e, ove i valori dei fenomeni procedano ad intervalli costanti, sarà inutile dividere il numero delle osservazione per i prodotti delle unità delle classi corrispondenti, perchè tali prodotti saranno costanti e la forma della superficie non ne sarà modificata.

In base a quest'ultima considerazione, potrebbe rappresentarsi, senza alcuna elaborazione, la tavola III. Notiamo che in essa, non solo le classi di stature dei padri — come pure quelle dei figli — procedono ad intervalli costanti, ma le prime presentano intervalli uguali a quelli delle seconde, il che non sarebbe necessario per seguire il più semplice procedimento esposto.

La figura risultante presenterebbe una forma simile a quella della figura 21, ma, in base all'orientamento degli assi di cotesta figura, molto più appiattita lateralmente in direzione *NO - SE*, ossia quella forma a schiena d'asino, che è caratteristica delle tavole a doppia entrata riguardanti quei due fenomeni e li rivela strettamente solidali.

Se la tavola a doppia entrata riguardi osservazioni accumulate, si potrà ben seguire il procedimento esposto per le tavole riguardanti valori singoli, tenendo presente peraltro che in questo caso ogni segmento perpendicolare al piano di base della figura corrisponde ai valori minori o uguali, oppure maggiori o uguali, alle coppie di valori rappresentati dalle ascisse corrispondenti.

La figura 23 rappresenta appunto in tal guisa la tavola XXIV: è facile connetterne la forma a quella che — come abbiamo notato — sarebbe risultata dalla rappresentazione della tavola III.

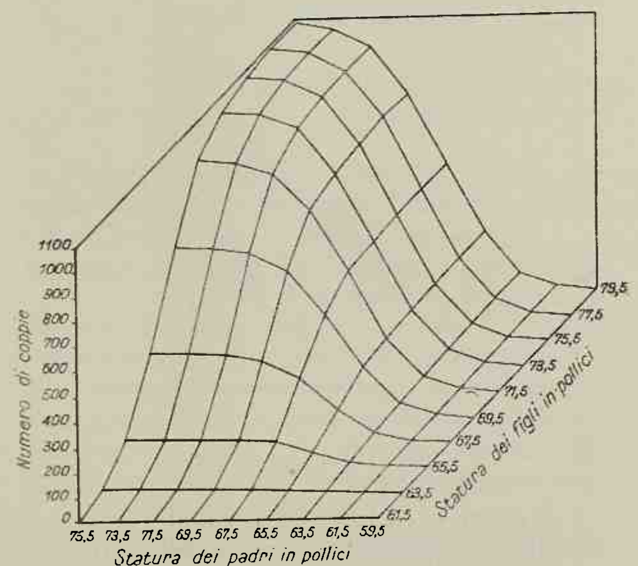


Fig. 23

Quanto alle tavole riguardanti quozienti fra due gruppi diversi di osservazioni (n. 5 del precedente capitolo), rimandiamo a ciò che abbiamo detto in fine al n. 2 per le rappresentazioni piane.

Infine, in base ai soliti tre assi ortogonali — dove rappresentiamo anni di nascita (X), anni di età (Y) e numero di individui (Z) — riportiamo sul piano XZ della figura 24 i gruppi decennali di nati in Italia dal 1872-81 al 1912-21; e sulle diagonali del piano XY, aventi origine nei punti dell'asse X corrispondenti alla fine degli anni 1881 e 1921, costruiamo i due diagrammi alle distribuzioni, per gruppi decennali di età, dei censiti nei vecchi confini del Regno.

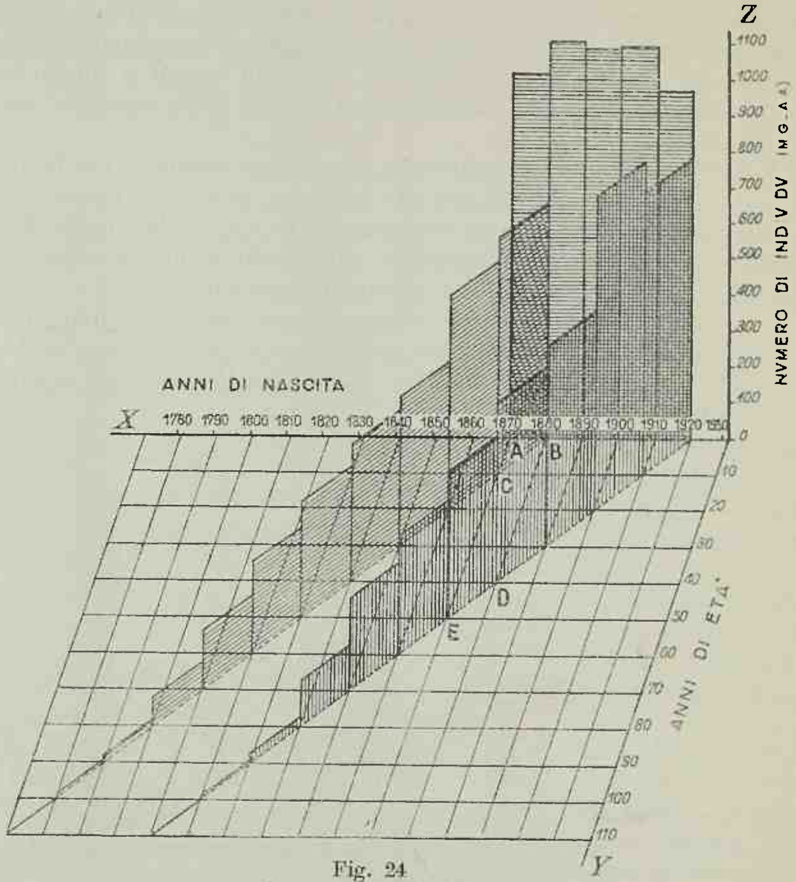


Fig. 24

È chiaro che i censiti alla fine del 1921 in età di 40-50 anni, rappresentati dal rettangolo con base DE posto sul secondo piano diagonale, provenivano dai censiti alla fine del 1881 in età di 0-10 anni, rappresentati dal rettangolo con base BC posto sul primo piano diagonale, e che questi a loro volta provenivano dai nati nel decennio 1872-81, rappresentati dal rettangolo sul piano XZ avente base AB.

Ora, se noi immaginiamo che in quella figura siano rappresentate tutte le distribuzioni di viventi — procedenti per intervalli piccolissimi di età — che possono pensarsi in corrispondenza ad ogni punto dell'asse X e che sulla superficie XZ siano rappresentati i gruppi di nati in ogni punto medesimo, riusciamo a concepire un solido, di cui — tra l'altro — le sezioni parallele al piano YZ attraversando i piani diagonali daranno le distribuzioni per età degli individui *provenienti da una stessa generazione di nati.*

Cotesto elegante tipo di rappresentazione grafica — chiamato stereogramma di Zeuner (1) — presenta un alto valore metodologico: su di esso è fondata la cosidetta teoria formale della popolazione. Nei limiti, in cui è possibile costruirlo, ha anche importanza descrittiva evidente: i pochi dati, che con esso abbiamo rappresentati, già mostrano l'incremento della nostra popolazione in tutti i gruppi d'età dal 1881 al 1921 e il basso livello del gruppo dei censiti a 0-10 anni d'età nell'ultimo censimento, in dipendenza della forte contrazione delle nascite nel decennio bellico 1912-'21.

7. Linee di livello. - Nella superficie faccettata della figura 21 possiamo considerare sezioni, i cui punti abbiano la medesima distanza dal piano di base; e possiamo pure tracciare le proiezioni di esse sul detto piano di base.

Sono, queste proiezioni, le linee di livello della tavola a doppia entrata, così chiamate per l'analogia che presentano con quelle largamente usate nelle carte topografiche.

La costruzione di esse implica in generale un sensibile arbitrio, dipendente dalla determinazione della quota nei punti intermedî a quelli corrispondenti ai numeri delle caselle.

Ad es., considerando la tavola XXII, le proiezioni sul piano di base delle sezioni, i cui punti abbiano un'altezza corrispondente a 100, 200, 300, non si possono esattamente disegnare, per il fatto che il numero di osservazioni delle caselle necessariamente varia con discontinuità.

Il modo più semplice per ovviare a questa difficoltà consiste nel costruire le linee di livello come nella figura 25, ossia nel far sì che, ad es., la linea di quota 100 divida il segmento AB — i cui punti estremi corrispondono nella tavola XXII ai numeri 33 e 125 — in parti proporzionali a $100-33=67$ ed a $125-100=25$; e così via.

Le linee così tracciate, e quelle altre che potrebbero facilmente tracciarsi assumendo quote intermedie od anche inferiori a 100 sino al limite consentito dalla dispersione dei dati, presentano nel nostro esempio forme

(1) Gustav Zeuner, *Abhandlungen aus der mathematischen statistik.* Leipzig, 1869. Cfr. su questo argomento i lavori citati di Perozzo e Wilhelm Lexis, *Abhandlungen zur theorie der bevölkerungs und moralstatistik.* Jena, Fischer, 1903.

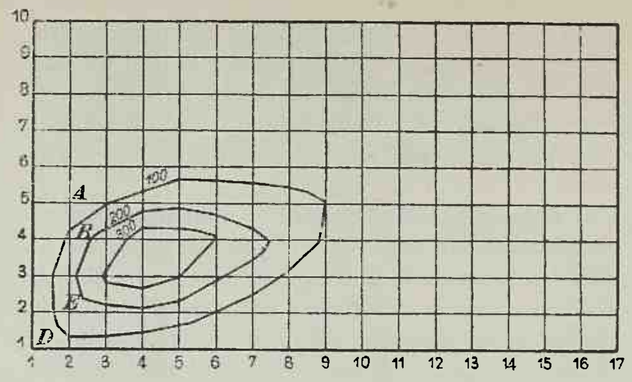


Fig. 25

ellissoidali, a causa di quello schiacciamento delle falde dello stereogramma, che rivela una sensibile solidarietà tra i fenomeni considerati, nel nostro esempio tra l'aumento delle stanze e quello degli abitanti nelle abitazioni rilevate. Tale schiacciamento non è molto evidente nella figura 21, a cagione del punto di vista, dal quale essa è stata disegnata; il che mette in luce uno degli aspetti vantaggiosi della rappresentazione con linee di livello.

Queste possono risultare aperte o chiuse e presentare le più diverse forme, in relazione ai legami esistenti tra i due fenomeni. Assumendo quote in progressione aritmetica, la maggior distanza fra le linee rivelerà una più lenta variazione di quota.

Dopo ciò che abbiamo detto nel numero precedente, si comprenderà facilmente che la costruzione di esse non dà luogo a difficoltà, quando la tavola a doppia entrata proceda per classi di valori ad intervalli costanti; e che, quando le classi siano di diversa ampiezza, converrà anzitutto formare una rete di rettangoli aventi lati proporzionali all'ampiezza delle classi, far corrispondere i numeri delle osservazioni ai punti centrali di tali rettangoli e attraverso i lati di essi far passare le linee nel modo dianzi illustrato, avendo però cura di fondare i calcoli sui numeri, divisi per l'area dei rettangoli, a cui essi si riferiscono.

8. Altri tipi di rappresentazioni nello spazio. – Se, invece delle coordinate cartesiane ortogonali, si adoperassero altri tipi di coordinate, ad es. le cosidette *coordinate geografiche* (raggio vettore, longitudine, colatitudine), si avrebbero altre forme di rappresentazioni nello spazio.

Esse, però, sono usate molto raramente, almeno nella pratica statistica, perchè non presentano la semplicità delle precedenti e si possono in generale evitare.

Le altre forme di rappresentazioni solide non sono che generalizzazioni di quelle forme piane, su cui ci siamo intrattenuti al n. 5.

Ad es. la figura 26 rappresenta su tre assi perpendicolari la superficie dell'Italia in migliaia di chilometri quadrati, la densità della popolazione per chilometro quadrato, i quozienti annui di natalità e di mortalità per abitante nel 1928: è chiaro che il piano di base della figura rappresenterà l'ammontare della popolazione e il volume totale di essa darà il numero annuo di nati, di cui è facile scorgere la parte corrispondente all'eccedenza dei nati sui morti.

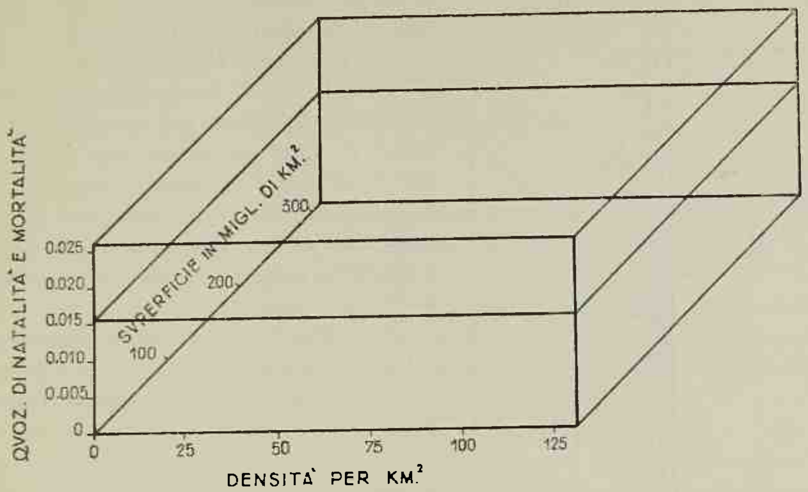


Fig. 26

Nei due solidi della figura 27 si riportano per il 1926 — rispettivamente riguardo ai centri aventi oltre 10.000 ab. e agli altri Comuni d'Italia — la popolazione, i quozienti di natalità per abitante e quelli di natimortalità (rapporto dei nati morti ai nati vivi e nati morti): si desume a primo sguardo che nei centri considerati la minor popolazione dà luogo ad un minor numero di nati e di nati morti, il quoziente di natalità è più basso, il quoziente di natimortalità più alto.

Infine, sul cerchio di base della figura 28 rappresentiamo la ripartizione in montagna, collina, pianura della popolazione italiana negli antichi confini del Regno al 1° dicembre 1921; sull'asse, ad esso perpendicolare

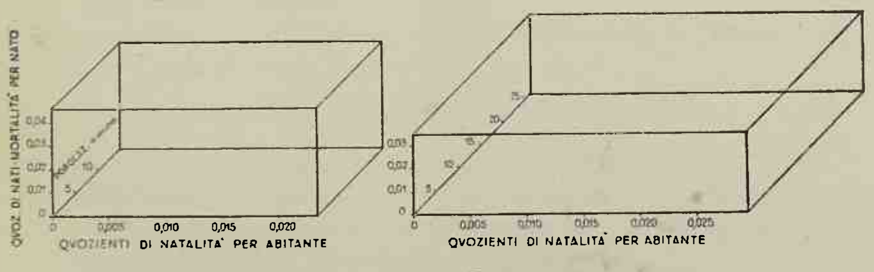


Fig. 27

e passante per il centro, il numero dei censiti di oltre 65 anni di età su 1000 abitanti. Le tre sezioni cilindriche, che ne risultano, dànno un'immagine eloquente dell'alta proporzione dei vecchi nelle montagne e della decrescente proporzione di essi in collina e in pianura: il volume di tali sezioni ne misura l'entità assoluta.

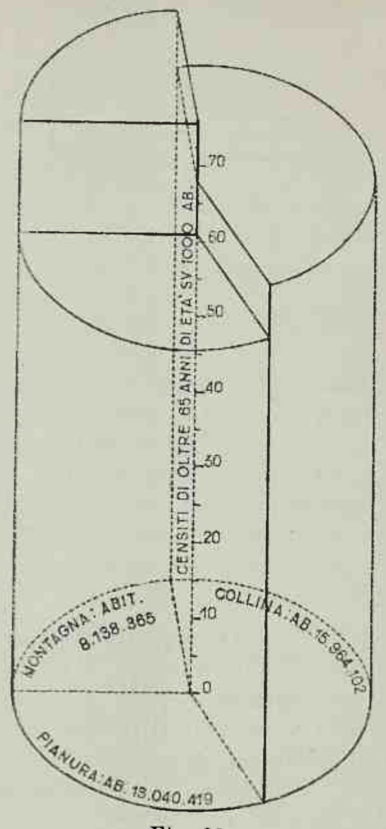


Fig. 28

Cotesti tipi di figure sono particolarmente consigliabili per la rappresentazione di quei quozienti, che risultano dal concorso di parecchi elementi, ad es. nella statistica dei trasporti per la rappresentazione del numero dei treni-chilometri, o degli assi-chilometri per km. di linea, e così via.

Su altre forme di rappresentazioni solide (allegoriche, ecc.) rimandiamo a ciò che, in fine al n. 5, abbiamo detto per le corrispondenti forme piane.

9. Figure a più dimensioni. – Per rappresentare una serie statistica a più di tre variabili od una tavola a più di due entrate, occorrerebbe costruire una figura a più di tre dimensioni. Senonchè figure geometriche siffatte non sono possibili nel nostro spazio tridimensionale. Ridotte ad esso, riescono poco chiare e non sono generalmente usate nella metodologia statistica.

Capitolo III.

LE COSTANTI CARATTERISTICHE

1. Il numero delle osservazioni. – I criterî ed i metodi, a cui abbiamo accennato nel primo capitolo, si fondano anche su alcune costanti caratteristiche dei gruppi osservati.

La più semplice di esse è ovviamente il numero delle osservazioni. Considerando, anzitutto, una distribuzione statistica, se indichiamo con x_i $(i = 1, 2, ..., n - 1, n)$ i valori diversi osservati e con y_i il numero di volte che essi si presentano,

$$N = y_1 + y_2 + \ldots\ldots + y_{n-1} + y_n \qquad (1)$$

è la detta caratteristica.

Il caso delle serie statistiche può formalmente ricondursi al precedente, ponendo, per il fenomeno che in esse si considera, $y_i = 1$.

2. La media aritmetica. – Una seconda caratteristica si ottiene rendendo sensibile la prima ai valori osservati, in guisa che un gruppo di valori più grandi dia luogo ad un risultato maggiore di quello ottenuto su un gruppo di valori più piccoli.

A tal uopo, attribuendo nella (1) ad ogni y_i un coefficiente d'importanza uguale al valore corrispondente, si ottiene:

$$S = x_1 y_1 + x_2 y_2 + \ldots\ldots + x_{n-1} y_{n-1} + x_n y_n \,. \qquad (2)$$

Cotesta formula, che dà la somma dei valori osservati, fornisce la seconda caratteristica nei riguardi delle distribuzioni. Per $y_i = 1$ si ha il caso delle serie statistiche.

Al fine di eliminare l'influenza, che sulla (2) esercita la (1), conviene ragguagliare l'una all'altra; onde:

$$M = \frac{S}{N} = \frac{x_1 y_1 + x_2 y_2 + \ldots\ldots + x_{n-1} y_{n-1} + x_n y_n}{y_1 + y_2 + \ldots\ldots + y_{n-1} + y_n} \qquad (3)$$

da la seconda caratteristica indipendentemente dalla prima.

Essa è la media aritmetica degli x_i, idonea a fornire un indice di posizione di essi corrispondente al valore comune che assumerebbero, se si ripartissero ugualmente la somma dei valori osservati.

Infatti la (3) può scriversi nella forma: $M N = S$.

La fig. 29 rappresenta un regolo poggiante su un fulcro e dà un'interpretazione meccanica della media aritmetica: la somma dei valori osservati, rappresentata nel braccio destro del regolo dai pesi y_1, y_2, ... posti alle distanze rispettive x_1, x_2, ... dal fulcro, è bilanciata nel braccio sinistro del regolo dalla somma di detti pesi y_1, y_2, ... posta alla distanza M dal fulcro.

Supponiamo, ad es., che si rilevino le età dei morti provenienti dal gruppo dei 992 mila nati in Italia nel 1932. Dopo la completa estinzione di quei nati, ossia fra poco più di un secolo, si sarebbe in grado di cal-

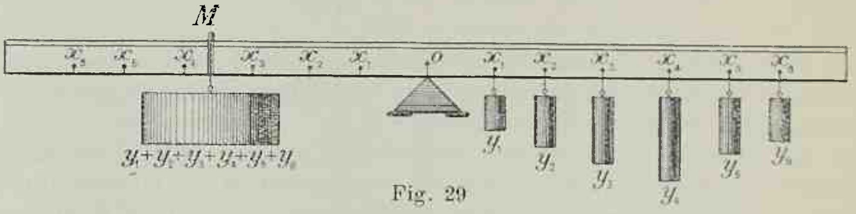


Fig. 29

colare la vita media del gruppo: moltiplicando ogni età (ad es. in anni e giorni) per il numero dei morti a quella età e sommando i risultati, si avrebbe la somma complessiva delle età dei morti, ossia il tempo complessivamente vissuto dai componenti quella generazione; dividendo cotesta somma per il numero totale dei morti (992.000), si avrebbe quella distanza media aritmetica delle date di morte da quelle di nascita, che si chiama brevemente « vita media ».

Estendendo la rilevazione ad altre generazioni annuali, ad es. a quelle del 1922, 1912, ..., si otterrebbero altri risultati, che compendierebbero efficacemente gli effetti dei miglioramenti igienici, economici, morali, ecc., sulle generazioni considerate.

Poichè questi miglioramenti nei paesi civili sono stati notevolissimi negli ultimi secoli, le vite medie ottenute sulle generazioni già estinte, che si son potute rilevare statisticamente, sono state in generale crescenti nel tempo. Ed è da ritenere che, se essi dureranno nei prossimi cento anni, la vita media dei nostri nati del 1932 segnerà un nuovo aumento, avvicinandosi a quei 65 anni, che per ragioni fisiologiche ne è un limite difficilmente superabile [1].

A scanso di equivoci, avvertiamo che la vita media di una generazione è una misura distinta dallo « intervallo medio tra due successive generazioni ». Per i nati del 1932 quest'ultimo si può calcolare (riguardo al sesso femminile) seguendo durante le età feconde le 484 mila femmine nate in quell'anno nel nostro paese e precisamente rilevando il numero

[1] Louis I. Dublin, *The possibility of extending human life*. New York, Metropolitan Life Insurance Co., 1922; riprodotto nella rivista « Metron », 1923.

delle figlie da esse procreate dall'età di 12 anni circa sino a circa 55 anni. Moltiplicando ogni età feconda per il numero corrispondente delle figlie, sommando i risultati e dividendo il totale per il numero complessivo delle figlie procreate, avremmo la distanza media aritmetica delle date di nascita delle figlie da quelle delle madri, ossia l'intervallo medio cercato.

Sembra che esso varii poco dall'uno all'altro Stato e oscilli intorno ai 30 anni (1).

Ciò è da attribuire anche al fatto che gli Stati, in cui — principalmente a causa della restrizione volontaria della prole — la proporzione tra le nascite e le donne atte a procreare diminuisce più rapidamente dopo i primi anni di età feconda, sono principalmente gli Stati più progrediti, nei quali è più alta la proporzione delle donne che, avendo raggiunto i 12 anni, sopravvivono ai 55 anni.

Cotesti calcoli potrebbero estendersi al sesso maschile — ad es. ai figli maschi, che nasceranno dai 508 mila maschi nati nel 1932 — ed anche all'insieme dei due sessi; ma è facile intendere le speciali difficoltà, a cui la rilevazione darebbe luogo riguardo ai figli naturali e adulterini.

Un altro esempio: in un dato anno accertiamo il valore dei redditi degli abitanti di uno Stato e calcoliamone la media aritmetica per abitante: moltiplicando ogni reddito monetario per il numero di abitanti che lo percepisce e dividendo la somma dai prodotti per il totale degli abitanti, otteniamo il reddito che ognuno avrebbe goduto se la somma dei redditi fosse stata distribuita in parti uguali tra gli abitanti medesimi.

Tenendo conto del potere d'acquisto della moneta rispetto a un gruppo assegnato di beni, quel reddito medio permetterà di apprezzare sinteticamente l'efficienza economica della popolazione, considerata come risultante dalla sua composizione per sesso, età, stato sanitario, occupazioni; e di rilevare quanto sia ancora piccola, anche nei paesi economicamente più progrediti, la quantità di beni che il reddito medio permette di acquistare e quanto sarebbe tenue per gl'indigenti il vantaggio derivante da una uguale ripartizione del reddito totale.

Tutto ciò naturalmente va detto nel caso, assai problematico, che l'uguale ripartizione del reddito totale non diminuisse l'ammontare di esso.

3. Medie aritmetiche semplici e ponderate. – Il calcolo della media aritmetica non dà luogo a difficoltà, quando si tratti ad es. di una distribuzione procedente per valori singoli (2).

---

(1) Robert R. Kukzynski, *op. cit.*, pag. 34, e *The balance of births and deaths*, Vol. I e II. Washington, The Brookings Institution, 1928-1931.

(2) I calcoli statistici sono molto facilitati da una macchina calcolatrice. Sono inoltre molto comode le *Barlow's Tables* (London, Spon), che danno senza errori i quadrati, i cubi, le radici quadratiche, le radici cubiche e i reciproci di tutti i numeri interi sino a 10.000. Per gl'interi sino a 100 saranno utili i prontuari, che riportiamo nelle appendici *B*) a questa Parte I.

Considerando la tavola V, si ha:

TAVOLA XXXI.

Numero dei figli minorenni x_i	Vedove passate a nuove nozze y_i	$x_i y_i$
0	109	0
1	60	60
2	35	70
3	17	51
4	5	20
5	1	5
Totali	227	206

$$M = \frac{206}{227} = 0,9075$$

Ogni gruppo di 100 vedove rimaritatesi aveva, quindi, in media 91 figli circa.

Questa media corrisponderà approssimativamente, nella figura 1, al punto di ascissa 0,91, molto vicino all'origine delle coordinate.

Se la distribuzione consta, invece, di classi di valori, il calcolo della media aritmetica non può certo presentare un grado di approssimazione maggiore di quello consentito dall'ampiezza delle classi. Il calcolo potrebbe eseguirsi rigorosamente, se per ogni classe di valori si conoscesse l'ammontare di essi; ma, poichè in generale non sono neppur note queste somme parziali, esso si esegue sostituendo ai valori di ogni classe la semisomma dei valori estremi di essa; e pertanto la media aritmetica della tavola VIII si calcola sui valori: 2.855, ecc., che appunto rappresentano le classi: 2.845 - 2.865, ecc. Con ciò si ammette l'ipotesi — tanto più plausibile *a priori* quanto, minore è l'intervallo adottato — che entro ogni classe i valori variino in progressione aritmetica; nel qual caso è facile constatare che la semisomma dei valori estremi corrisponde alla media aritmetica del gruppo.

Infatti nella progressione aritmetica:

$$x_1, \quad x_1 + d, \quad x_1 + 2d, \quad \ldots\ldots, \quad x_1 + (n-1)\,d,$$

tenendo presente che la somma dei primi n numeri della serie naturale è $\frac{n\,(n+1)}{2}$, si ha:

$$M = x_1 + \frac{n-1}{2}\,d = \frac{1}{2}\,[2x_1 + (n-1)\,d]$$

Lo stesso procedimento può seguirsi per le tavole VI, VII e IX, ove almeno si ritenesse lecito attribuire alle classi estreme aperte di esse un intervallo uguale a quello delle altre classi.

A volte, però, la scelta delle classi si subordina a scopi diversi da quello del calcolo della media aritmetica o di altre costanti. Ciò, in generale, risulta palese quando le classi siano molto ampie e per giunta di ampiezza diversa (tavola X); ed in tal caso — se non si possano seguire i procedimenti, di cui ci occuperemo nel seguente capitolo — converrà astenersi dal calcolo di quelle costanti.

Se, come nella tavola IX, esistono valori ignoti, è facile verificare che — nell'ipotesi che sia lecito ripartirli proporzionalmente tra le varie classi della distribuzione — essi non eserciteranno alcuna influenza sul calcolo della media aritmetica e potranno trascurarsi. L'ipotesi è naturalmente tanto più plausibile *a priori*, quanto minore è la proporzione dei detti valori ignoti al totale di essi.

Si noti che i confini dell'errore, che si può commettere calcolando la media aritmetica — come, del resto, ogni altro valore medio — su distribuzioni procedenti per classi di valori, invece che per valori singoli, si ottengono facilmente calcolando i due valori che il valor medio assumerebbe nelle due ipotesi che ad ogni classe possa attribuirsi il limite inferiore o quello superiore di essa. Si verifica, inoltre, facilmente che la semisomma delle due medie aritmetiche, così calcolate, è uguale alla media aritmetica calcolata sulle semisomme dei limiti di ogni classe; e che inoltre, se queste siano tutte di uguale ampiezza, le due medie differiranno fra loro di una quantità uguale all'intervallo comune delle classi; mentre, se le classi abbiano diversa ampiezza, le due medie differiranno di una quantità uguale alla media ponderata degl'intervalli delle classi.

Quanto alle serie statistiche, è generalmente agevole il calcolo della loro media aritmetica. Ponendo nella (3) $y_i = 1$, dalla tavola XIV si ricava che la media aritmetica delle tonnellate di stazza netta delle navi a vela nazionali, accertate alla fine degli anni 1890-1913, è:

$$M = \frac{634,1 + 625,8 + \ldots\ldots + 356}{24} = \frac{12.584,7}{24} = 524,4 \text{ migliaia di tonn.}$$

Si dice che in questo caso si è calcolata una *media aritmetica semplice*; mentre, se almeno un peso è diverso dall'unità — come nel caso precedente delle distribuzioni — si parla di *media aritmetica ponderata*. È facile, però, verificare che, se gli y_i sono comunque uguali fra loro, le due medie, semplice e ponderata, coincidono.

Si avverta che talora i pesi non danno il numero di volte, che i valori corrispondenti si sono ripetuti, ma un numero fittizio, proporzionale al grado di attendibilità che si attribuisce ai valori osservati. Ad es. nelle medie dei prezzi di un bene in un dato tempo e luogo si suol dare un peso maggiore ai prezzi ricavati da fonti più sicure. Si tratta, però, di manipolazioni soggettive dei dati, di cui bisogna fare uso prudente e pubblicamente dichiarato, e dalle quali è bene astenersi se non siano strettamente necessarie.

Notiamo ancora che la media aritmetica, non essendo sensibile all'ordine di successione dei valori, è indipendente dal fenomeno, in base al quale le osservazioni in esame sono state ordinate: ad es., la media aritmetica della tavola XIV coincide con quella, che potremmo ricavare dalla tavola XV e da qualunque altra serie a due o più variabili, in cui quelle stazze nette di navi a vela fossero riportate.

Il caso, poi, delle tavole a più entrate può ricondursi a quello delle distribuzioni statistiche: ad es. la media aritmetica delle stature dei padri, di cui alla tavola III, si ottiene facilmente considerando la distribuzione formata dalla riga delle stature dei padri e da quella delle osservazioni totali.

Riguardo agli attributi, è chiaro che essi — qualunque sia il tipo di rappresentazione numerica, di cui facciano parte — non ammettono medie aritmetiche nè altre costanti quantitative, tranne che corrispondano sostanzialmente a quantità, come ad es. gli esiti dei conflitti di lavoro (n. 1, *f* del primo capitolo) o che siano comunque graduabili, come ad es. il colore degli occhi e dei capelli.

In quest'ultimo caso resta, però, l'arbitrio della scelta della graduatoria, arbitrio che può dar luogo ad inconvenienti, specie se non si possa ammettere che gli attributi siano equidistanti tra loro o se ognuno di essi formi una classe.

Facendo ad es. corrispondere ai colori degli occhi della distribuzione, di cui alla tavola XIII, i numeri interi da 0 a 3, si ottiene:

TAVOLA XXXII.

Colore degli occhi	Graduatoria x_i	Militari y_i	$x_i y_i$
Celesti	0	30.882	0
Grigi	1	61.605	61.605
Castani	2	180.255	360.510
Neri	3	26.118	78.354
Totali		298.860	500.469

$$M = \frac{500.469}{298.860} = 1,6746.$$

Questo risultato dimostrerebbe che il colore medio degli occhi di quel gruppo di militari, rilevati nella nostra prima inchiesta antropometrica e che rispecchiano le caratteristiche della nostra popolazione maschile, è più vicino al castano che al grigio; ma, se quei colori corrispondano a classi di diversa ed imprecisabile ampiezza, la graduatoria adottata e la media calcolata su di essa saranno evidentemente errate e non ne sarà facile la correzione (1).

Questo esempio dimostri la grande cautela, che occorre impiegare nel calcolo delle costanti caratteristiche degli attributi graduabili.

4. Medie aritmetiche di quozienti. – Dai quozienti parziali si può ricavare il quoziente globale, calcolando una media aritmetica ponderata dei primi.

Riprendendo l'esempio della tavola XIX *bis* (n. 5 del primo capitolo) — ma ciò che diciamo può estendersi a qualunque altro tipo di rappresentazioni numeriche, comprese le tavole a più entrate —, il quoziente globale di natalità nel quinquiennio considerato è:

$$\frac{1.124.470 + 1.109.761 + \ldots\ldots + 1.072.316}{39.550.000 + 39.878.000 + \ldots\ldots + 40.970.000} = 0.0273$$

ossia il 27,3 per mille, e tale risultato è uguale a quello che si sarebbe ottenuto calcolando la media aritmetica dei cinque quozienti della tavola, assunto ciascuno di essi con peso proporzionale alla popolazione media dell'anno, a cui si riferisce:

$$\frac{\frac{1.124.740}{39.550.000} \times 39.550.000 + \frac{1.109.761}{39.878.000} \times 39.878.000 + \ldots + \frac{1.072.316}{40.970.000} \times 40.970.000}{39.550.000 + 39.878.000 + \ldots\ldots + 40.970.000} =$$

$$= 0,0273$$

In generale, chiamando con x_i e con y_i i due termini di una successione di rapporti, si ha:

$$\text{Quoziente globale:} = \frac{x_1 + x_2 + \ldots\ldots + x_n}{y_1 + y_2 + \ldots\ldots + y_n} = \frac{y_1 \frac{x_1}{y_1} + y_2 \frac{x_2}{y_2} + \ldots\ldots + y_n \frac{x_n}{y_n}}{y_1 + y_2 + \ldots\ldots + y_n}.$$

Si fermi bene in mente che la media aritmetica di un gruppo di rapporti, in generale, non è uguale al rapporto tra le somme o le medie aritmetiche dei termini di essi:

$$M\left(\frac{x_i}{y_i}\right) \neq \frac{M(x_i)}{M(y_i)} \tag{4}$$

(1) Per il colore degli occhi, dei capelli, della pelle, ecc., oggi s'impiegano scale cromatiche molto precise e dettagliate. Cfr. R. Martin, *Lehrbuch der anthropologie*. Jena, Fischer, 1928 (2ª edizione).

5. Le principali proprietà della media aritmetica. Gli scarti e i momenti. – Chiamando scarti (scostamenti, deviazioni) di x_i da m le quantità $x_i - m$, e chiamando *momento di grado s e di origine m* la espressione:

$$(x_1 - m)^s y_1 + (x_2 - m)^s y_2 + \ldots\ldots + (x_n - m)^s y_n, \tag{5}$$

possiamo mettere in luce alcune interessanti proprietà della media aritmetica.

Poniamo nella precedente espressione $m = M$, $s = 1$. Tenendo presente la (3), abbiamo:

$$(x_1 - M) y_1 + \ldots + (x_n - M) y_n = x_1 y_1 + \ldots + x_n y_n - M (y_1 + \ldots + y_n) = 0 \tag{6}$$

ossia: *è nulla la somma algebrica degli scarti dalla media aritmetica.*

Inoltre, sia z la differenza tra gli scarti rispetto ad un'origine qualsiasi m e gli scarti dalla media aritmetica:

$$x_i - m = x_i - M + z. \tag{7}$$

Elevando al quadrato ambo i membri di cotesta uguaglianza e sommando, otteniamo:

$$(x_1 - m)^2 y_1 + \ldots\ldots + (x_n - m)^2 y_n = (x_1 - M)^2 y_1 + \ldots\ldots + (x_n - M)^2 y_n + \\ + 2z [(x_1 - M) y_1 + \ldots\ldots + (x_n - M) y_n] + z^2 (y_1 + \ldots\ldots + y_n).$$

E poichè, per la (6), è nullo il doppio prodotto che compare nel secondo membro della precedente uguaglianza e l'ultimo termine è essenzialmente positivo, si ha sempre, per z non nullo:

$$(x_1 - M)^2 y_1 + \ldots\ldots + (x_n - M)^2 y_n < (x_1 - m)^2 y_1 + \ldots\ldots + (x_n - m)^2 y_n,$$

ossia: *la somma dei quadrati degli scarti dalla media aritmetica è un minimo, rispetto alla somma dei quadrati degli scarti da un'altra origine qualsiasi.*

Dalla (7) discende, tenendo conto della (6):

$$(x_1 - m) y_1 + \ldots\ldots + (x_n - m) y_n = z (y_1 + \ldots\ldots + y_n)$$

$$z = \frac{(x_1 - m) y_1 + \ldots\ldots + (x_n - m) y_n}{y_1 + \ldots\ldots + y_n} \tag{8}$$

$$M = m + z = m + \frac{(x_1 - m) y_1 + \ldots\ldots + (x_n - m) y_n}{y_1 + \ldots\ldots + y_n}. \tag{9}$$

Quest'ultima formula, come vedremo fra poco, permette in alcuni casi un calcolo rapido della media aritmetica.

Chiamando *momento diviso* l'espressione:

$$\mu_{m,s} = \frac{(x_1 - m)^s y_1 + \ldots\ldots + (x_n - m)^s y_n}{y_1 + \ldots\ldots + y_n} \tag{10}$$

si ha: $M = \mu_{0,1}$, $z = \mu_{m,1}$ e possiamo ricavare:

$$(x_1 - M)^s y_1 + \ldots\ldots + (x_n - M)^s y_n = (x_1 - m - \mu_{m,1})^s y_1 + \ldots\ldots + (x_n - m - \mu_{m,1})^s y_n$$

Considerando in questa uguaglianza le espressioni fra parentesi del secondo membro come formate dai due termini $x_i - m$ e $\mu_{m,1}$, in base alla formula dello sviluppo del binomio si ottiene:

$$\mu_{M,2} = \mu_{m,2} - \mu^2_{m,1} \tag{11}$$

$$\mu_{M,3} = \mu_{m,3} - 3\,\mu_{m,1}\,\mu_{M,2} - \mu^3_{m,1} \tag{12}$$

$$\mu_{M,4} = \mu_{m,4} - 4\,\mu_{m,1}\,\mu_{M,3} - 6\,\mu^2_{m,1}\,\mu_{M,2} - \mu^4_{m,1} \tag{13}$$

. . . .

Sono queste le relazioni, che legano i momenti divisi di grado 2, 3, 4, ..., rispetto alla media aritmetica, ai momenti divisi di uguale grado rispetto ad un'origine qualsiasi m (¹).

Avremo in seguito frequenti occasioni d'impiegarle.

6. Semplificazioni nel calcolo della media aritmetica. – Abbiamo affermato che la (9) permette in alcuni casi un calcolo rapido della media aritmetica: essa infatti mostra che, calcolando la media aritmetica degli scarti da un valore arbitrario convenientemente scelto (che sia ad es. un numero intero o con pochi decimali, corrispondente ad uno dei gruppi più numerosi di osservazioni, ecc.) ed aggiungendo il risultato al valore arbitrario, si ottiene la media aritmetica cercata.

Applicando ad es. la (9) alla distribuzione della tavola IX e ponendo arbitrariamente $m = 32{,}5$, si ha:

Tavola XXXIII.

Valori centrali delle classi di età delle madri x_i	Nascite y_i	$x_i - 32{,}5$	$(x_i - 32{,}5)\,y_i$
12,5	5	— 20	— 100
17,5	2.324	— 15	— 34.860
22,5	11.472	— 10	— 114.720
27,5	12.485	— 5	— 62.425
32,5	9.044	0	0
37,5	4.897	+ 5	+ 24.485
42,5	1.474	+ 10	+ 14.740
47,5	126	+ 15	+ 1.890
52,5	3	+ 20	+ 60
Totali	41.830		— 212.105
			+ 41.175
			— 170.930

$$M = 32{,}5 - \frac{170.930}{41.830} = 28{,}4137 = 28 \text{ anni e } 5 \text{ mesi circa.}$$

(¹) Sulle relazioni tra i momenti e sul calcolo di essi può consultarsi: W. Palin Elderton, *Frequency curves and correlation.* London, Layton, 1906.

Questo risultato non dà punto l'intervallo tra due generazioni successive di femmine, di cui abbiamo parlato al n. 2: infatti le nascite, da noi considerate, comprendono ambo i sessi e non derivano da una medesima generazione di madri, ma dalle diverse generazioni di madri contemporaneamente viventi alle età feconde nel 1916. Si tratta di una distribuzione statica (n. 1, *g* del primo capitolo).

Se l'intervallo delle classi è costante — come nel precedente esempio — una maggiore semplificazione di calcoli può ottenersi, dividendo ogni scarto arbitrario per l'intervallo costante e moltiplicando poscia per cotesto intervallo la somma dei prodotti dei nuovi scarti per il numero corrispondente delle osservazioni.

La semplificazione deriva dal fatto che gli scarti risultano uguali, in valore numerico e a partire dall'origine, alla successione dei numeri interi.

Nell'esempio precedente l'intervallo costante è uguale a 5, e risulta:

Tavola XXXIV.

Valori centrali delle classi di età delle madri x_i	Nascite y_i	$\frac{x_i - 32,5}{5}$	$\left(\frac{x_i - 32,5}{5}\right) y_i$
12,5	5	— 4	— 20
17,5	2.324	— 3	— 6.972
22,5	11.472	— 2	— 22.944
27,5	12.485	— 1	— 12.485
32,5	9.044	0	0
37,5	4.897	+ 1	+ 4.897
42,5	1.474	+ 2	+ 2.948
47,5	126	+ 3	+ 378
52,5	3	+ 4	+ 12
Totali	41.830		— 42.421
			+ 8.235
			— 34.186

$$M = 32,5 - \frac{5 \times 34.186}{41.830} = 28,4137 = 28 \text{ anni e } 5 \text{ mesi circa.}$$

7. Calcolo grafico della media aritmetica. – La media aritmetica può anche calcolarsi graficamente a mezzo di un semplicissimo procedimento, che adesso esemplifichiamo nel caso generale della media ponderata di n quantità qualsiansi e nell'ipotesi — che generalmente

può rispettarsi — che il punto d'incontro dei due assi cartesiani coincida con l'origine della scala delle y (1).

Premettiamo che spesso conviene ridurre dei segmenti in date proporzioni e che tale riduzione può eseguirsi meccanicamente a mezzo di un *compasso di riduzione*. Esso è costituito da due aste AB e $A'B'$ rigide, graduate e collegate con un perno spostabile (fig. 30). Collocando il perno P in modo che BP e $B'P$ risultino uguali ad $1/h$ (nella figura: 1/2) della lunghezza delle aste, i due triangoli $B'PB$ e APA' saranno simili e la distanza $B'B$ sarà uguale ad $1/h$ di AA'.

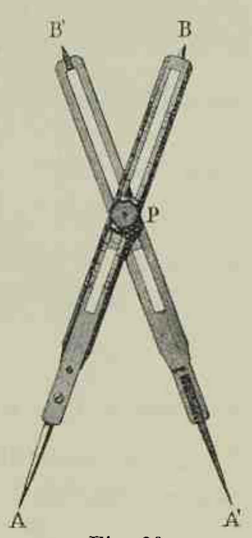


Fig. 30

Pertanto, spostando convenientemente il perno e facendo corrispondere ad AA' la lunghezza di un dato segmento, la distanza $B'B$ ne darà la frazione che si desidera.

La figura 31 rappresenta la distribuzione della tavola XXXIV e risponde ai fini del calcolo, in quanto il numero delle osservazioni di ogni classe si è fatto corrispondere al valore centrale di essa.

Il valore medio aritmetico è stato determinato nel modo seguente:

a) Ridotti per semplicità, col compasso di riduzione, ad $1/h$ (nell'esempio, ad 1/5) i segmenti $X_i\ F_i$, si è costruito, a partire da X_1 e perpendicolarmente all'asse orizzontale, il segmento $X_1\ C_1$ risultante dalla somma di quei segmenti ridotti.

b) Dal punto C_1 si è condotta una parallela all'asse orizzontale e si sono congiunti col punto X_1 i punti d'incontro C_2, C_3 ... di essa con i segmenti $X_i\ F_i$ o coi loro prolungamenti.

c) Tracciando dai punti F_i le parallele ad $X_1\ C_i$, si sono determinati i punti d'incontro D_i di esse con l'asse orizzontale.

d) Portando successivamente sul verso positivo di quest'asse, a partire dal punto X_1, i segmenti $D_i\ X_i$ ($D_1\ X_1$ è necessariamente nullo) si è ottenuto un segmento somma, che ridotto ad $1/h$ (nell'esempio, ad 1/5) è risultato uguale ad $X_1\ N$.

L'ascissa del punto N, che sulla scala si legge in 28,5 anni — il calcolo numerico ha dato 28,41 — è la media aritmetica della distribuzione.

Infatti, i due triangoli $X_1\ X_2\ C_2$ e $D_2\ X_2\ F_2$ sono simili e le misure dei loro lati sono legate dalla relazione:

$$(D_2\ X_2) = \frac{(X_1\ X_2)\ (X_2\ F_2)}{(X_2\ C_2)} = \frac{(X_1\ X_2)\ (X_2\ F_2)}{(X_1\ C_1)}$$

(1) DIONISIO TENDERINI, *Calcoli grafici*, in «Rivista italiana di statistica», 1930.

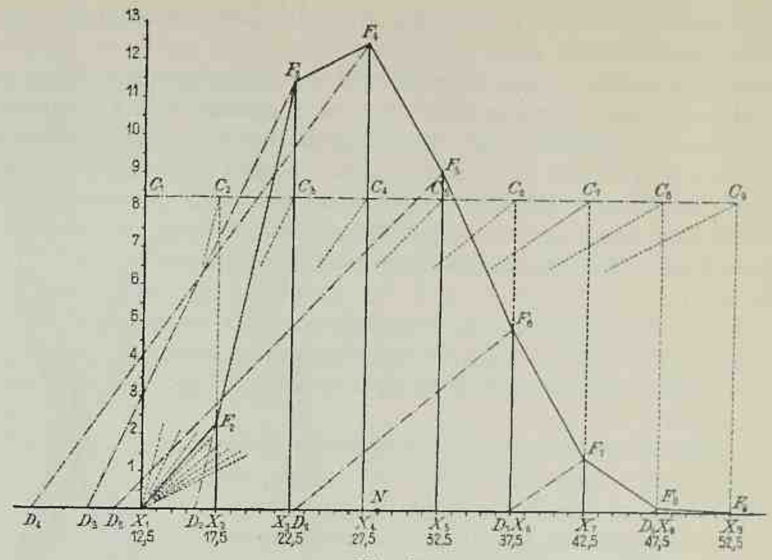


Fig. 31

E poichè, in base alle notazioni dei numeri precedenti, $X_1 X_2$ è misurato da $x_2 - x_1$, $X_2 F_2$ da y_2 e $X_1 C_1$ da $\frac{1}{h}$ $(y_1 + y_2 + \dots)$, il segmento $D_2 X_2$ sarà misurato da:

$$h \frac{(x_2 - x_1)\, y_2}{y_1 + y_2 + \dots}$$

e $X_1 N = \frac{1}{h}$ $(D_2 X_2 + D_3 X_3 + \dots)$ sarà misurato da $\mu_{x_1, 1}$ ossia, per la (9) e la (10), da $M - x_1$.

Adunque M è l'ascissa del punto N.

8. Indici speciali di posizione. La media armonica e la media geometrica. – Abbiamo considerato, come misura di posizione di un gruppo di osservazioni, la media aritmetica.

Però in alcuni casi, in vista di esigenze speciali, s'impiegano altre costanti caratteristiche.

Ci occuperemo, anzitutto, delle cosidette medie armoniche e geometriche, che talvolta sostituiscono la media aritmetica (1).

A. – Se da un gruppo di osservazioni si volesse risalire alla media aritmetica di un altro gruppo, legato al precedente da una nota relazione,

(1) Angelo Messedaglia, *Il calcolo dei valori medî e le sue applicazioni statistiche*, in « Biblioteca dell'economista », Serie V, Vol. XIX. Vi si espone il pensiero dei greci sulle principali medie, la cui origine peraltro si perde nella notte dei tempi.

occorrerebbe impiegare per il primo gruppo uua formula diversa da quella della media aritmetica, che tenesse conto di cotesta relazione.

Supponiamo, ad es., che da un gruppo di misure delle variazioni del livello dei prezzi si volesse risalire alla media aritmetica delle variazioni del corrispondente potere d'acquisto della moneta: è chiaro che in tal caso si dovrebbe calcolare la media aritmetica dei reciproci dei valori dati (purchè nessuno di essi sia nullo), perchè le variazioni del livello dei prezzi sono inverse a quelle del potere d'acquisto della moneta: quando, ad es., il livello medio dei prezzi raddoppia, il potere medio d'acquisto della unità monetaria dimezza e viceversa.

Chiamando media armonica il reciproco della media aritmetica dei reciproci:

$$M_a = \left(\frac{x_1^{-1} y_1 + \dots + x_n^{-1} y_n}{y_1 + \dots + y_n} \right)^{-1} = \frac{y_1 + \dots + y_n}{\frac{1}{x_1} y_1 + \dots + \frac{1}{x_n} y_n} \qquad (14)$$

notiamo che alla media aritmetica di una successione corrisponde la media armonica della successione dei reciproci, e viceversa.

Infatti nelle due successioni:

$$x_1, x_2, \dots x_n \qquad\qquad \frac{1}{x_1}, \frac{1}{x_2}, \dots \frac{1}{x_n}$$

si ha:

$$M(x) = \frac{x_1 + x_2 + \dots + x_n}{n} \qquad\qquad M_a\left(\frac{1}{x}\right) = \frac{n}{x_1 + x_2 + \dots + x_n}$$

od anche:

$$M_a(x) = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} \qquad\qquad M\left(\frac{1}{x}\right) = \frac{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}{n}$$

Possiamo pure scrivere:

$$\frac{1}{M(x)} = M_a\left(\frac{1}{x}\right) \qquad\qquad \frac{1}{M_a(x)} = M\left(\frac{1}{x}\right)$$

Nel triennio 1930, 1931, 1932 il livello medio dei prezzi all'ingrosso è disceso nel nostro paese da 411,04 a 341,57 e a 309,91, posto uguale a 100 il livello medio del 1913; e pertanto il potere medio d'acquisto della lira è salito da 24,33 a 29,28 e a 32,27, sempre posto uguale a 100 il potere medio d'acquisto del 1913.

Ora, alla media aritmetica delle tre variazioni del livello dei prezzi (354,17) non corrisponde la media aritmetica (28,63) dei tre poteri d'acquisto della lira, ma la media armonica (28,24) di essi, che ne è appunto il reciproco.

Similmente, se si rilevano i cambî giornalieri del dollaro a Milano (che sono espressi in lire) e i cambî giornalieri della lira a New York (che sono espressi in dollari) si constata che, per l'imperfetta azione degli « arbitraggi » e per l'influenza di svariate circostanze, le quotazioni corrispondenti non sono rigorosamente reciproche, cioè inverse. Ma, se noi vogliamo ottenere una misura media degli scarti in una data settimana, mese o anno, non dobbiamo confrontare le medie aritmetiche delle due successioni corrispondenti, ma la media aritmetica dell'una con la media armonica dell'altra (1).

Parecchi esempî s'incontrano anche in altri campi di applicazione della metodologia statistica. Seguiamo un gruppo annuale di nati (ad es. un milione) sino alla sua completa estinzione, sommiamo gli anni vissuti da ogni nato (ad es. 53,2 milioni di anni) e dividiamo cotesta somma per il milione di nati: avremo la «vita media» di ogni nato, che nell'esempio sarà di 53,2 anni.

Chiamiamo con l_x ($x = 0, 1, 2, \ldots n$) il numero dei nati (età 0) e dei sopravviventi al 1°, 2°, ... n° anno di età· con d_x ($=l_x - l_{x+1}$) il numero dei morti all'età x, $x+1$· ed ammettiamo che i morti nel primo, nel secondo ... anno di età si siano distribuiti uniformemente nel corso di ogni anno e siano quindi vissuti in media $\frac{1}{2}\,d_0$, $\frac{3}{2}\,d_1 \ldots$.

È facile constatare che il numero totale degli anni vissuti (53,2 milioni di anni) risulterà uguale alla media dei sopravviventi nel primo anno di età: $\frac{1}{2}\,(l_0 + l_1)$, più la media dei sopravviventi nel secondo anno: $\frac{1}{2}\,(l_1 + l_2)$ e così via:

$$\frac{1}{2}\,d_0 + \frac{3}{2}\,d_1 + \frac{5}{2}\,d_2 + \ldots = \frac{1}{2}\,(l_0 - l_1) + \frac{3}{2}\,(l_1 - l_2) + \frac{5}{2}\,(l_2 - l_3) + \ldots =$$

$$= \frac{1}{2}\,(l_0 + l_1) + \frac{1}{2}\,(l_1 + l_2) + \frac{1}{2}\,(l_2 + l_3) + \ldots$$

Ne viene che, se il rapporto di questa somma ad l_0 dà la vita media di ogni nato, il reciproco di esso è il rapporto tra il numero totale dei morti ($l_0 = d_0 + d_1 + \ldots d_n$) e la somma dei viventi nella media del primo, del secondo ... dell'n° anno di vita, ossia è un « quoziente medio annuo di mortalità ».

Poichè nell'esempio la vita media è risultata di 53,2 anni, il reciproco

(1) In periodi eccezionali gli scarti tra le medie aritmetiche delle quotazioni di due piazze possono essere molto grandi, com'è accaduto per il marco tedesco nel 1922-23. Cfr. C. Bresciani Turroni, *Le vicende del marco tedesco*, in « Annali di economia », Vol. VII. Milano, Università Bocconi. 1931.

di essa: 0,0188 dirà che in ogni anno di vita quel gruppo di nati ha avuto in media una mortalità del 18,8 per mille viventi.

Ciò posto, se noi calcoliamo le vite medie di diversi gruppi annuali di nati e ne facciamo la media aritmetica, il corrispondente quoziente medio annuo di mortalità non sarà dato dalla media aritmetica dei reciproci di ogni gruppo, ma dalla loro media armonica.

Infine, ponendo nella (5) $m = 1/M_a$ e sostituendo a x_i i loro reciproci, può verificarsi che *è nulla la somma algebrica degli scarti tra i valori reproci e il reciproco della media armonica, ed inoltre è minima la somma dei quadrati di questi scarti.*

Il calcolo grafico di essa, nel caso generale della (14), è facile quando nella rappresentazione grafica cartesiana il punto d'incontro dei due assi si faccia corrispondere all'origine di ambe le scale.

La fig. 32 rappresenta la distribuzione del numero (in milioni) dei biglietti giacenti nelle casse della Banca d'Italia al 30 giugno 1929, se-

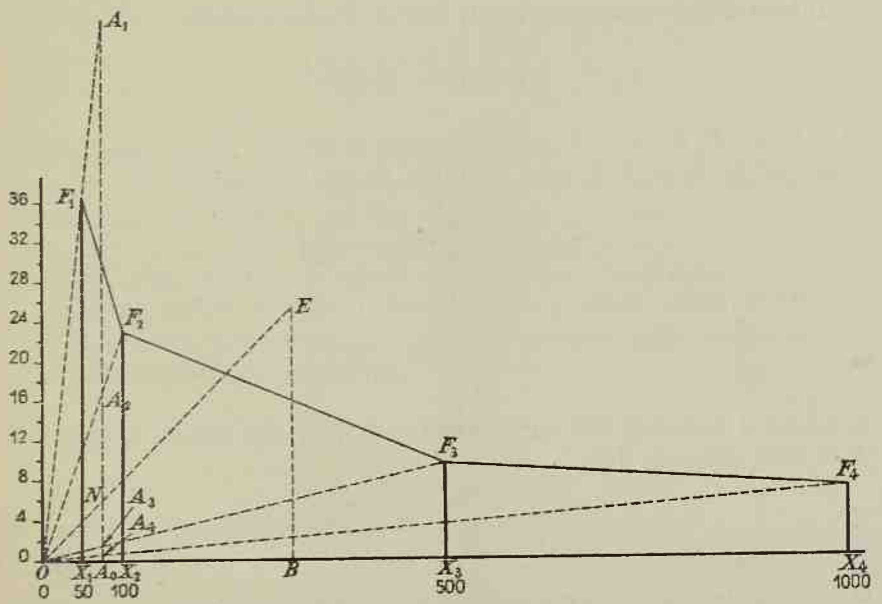


Fig. 32

condo il taglio di essi. Vi si legge che i biglietti da L. 50 ammontavano a 37,5 milioni, da L. 100 a 23,2, da L. 500 a 9,9 e da L. 1.000 a 7,1 milioni.

Si procede nel modo seguente [1]:

[1] DIONISIO TENDERINI, op. cit.

a) Da un punto arbitrario A_0, posto sul verso positivo dell'asse delle ascisse, si tira una parallela all'asse delle ordinate.

b) Si congiungono con segmenti rettilinei i punti F_i col punto O d'incontro dei due assi; tali segmenti, o i loro prolungamenti, incontrano la parallela arbitraria nei punti A_1, A_2, A_3, A_4.

c) Riducendo, col compasso riduttore, ad $1/h$ (nel nostro esempio ad $1/3$) i segmenti $A_0\,A_i$ e riportando il segmento corrispondente alla somma di essi sull'asse delle ascisse a partire da O, si ottiene il segmento $O\,B$.

d) Perpendicolarmente all'asse delle ascisse ed a partire da B, si traccia il segmento $B\,E$, somma dei segmenti $X_i\,F_i$ ridotti ad $1/h$ (nell'esempio ad $1/3$).

e) Si fa passare una retta per i punti O ed E: essa incontrerà la parallela arbitraria nel punto N; il segmento $A_0\,N$ sarà misurato, sulla scala delle ascisse, dalla media armonica della distribuzione, media che, nell'esempio scelto, risulta praticamente uguale alle L. 77 fornite dal calcolo numerico.

Infatti i due triangoli $O\,A_0\,A_1$ e $O\,X_1\,F_1$ sono simili:

$$(A_0\,A_1) = \frac{(OA_0)\,(X_1F_1)}{(OX_1)}$$

ed essendo $X_1\,F_1$ ed $O\,X_1$ misurati rispettivamente da y_1 e da x_1 e ponendo $(O\,A_0) = s$, $A_0\,A_1$ sarà misurato da sy_1/x_1, e quindi $O\,B$ da:

$$\frac{s}{h}\left(\frac{y_1}{x_1} + \frac{y_2}{x_2} + \frac{y_3}{x_3} + \frac{y_4}{x_4}\right).$$

D'altra parte, anche i due triangoli $O\,A_0\,N$ e $O\,B\,E$ sono simili:

$$(A_0\,N) = \frac{(OA_0)\,(BE)}{(OB)}$$

e, poichè il segmento BE corrisponde a $(y_1 + y_2 + y_3 + y_4)/h$, il segmento A_0N sarà misurato da:

$$\frac{y_1 + y_2 + y_3 + y_4}{\dfrac{y_1}{x_1} + \dfrac{y_2}{x_2} + \dfrac{y_3}{x_3} + \dfrac{y_4}{x_4}}$$

ossia dalla media armonica dei valori, la quale risulta espressa nell'unità di misura di essi.

È quasi superfluo aggiungere che, nel caso di valori negativi, ossia se i punti X_i cadessero anche nel verso negativo dell'asse delle ascisse, si potrebbe assumerli come positivi, salvo considerare come negativo il verso dei segmenti A_0A_i ad essi corrispondenti, ossia ottenere $O\,B$ mercè la somma algebrica degli A_0A_i.

Ma in tal caso la media armonica può dare risultati aberranti.

B. – Considerando quel numero che ha per logaritmo la media aritmetica dei logaritmi dei valori osservati:

$$\log M_g = \frac{y_1 \log x_1 + \dots + y_n \log x_n}{y_1 + \dots + y_n} \tag{15}$$

e passando dai logaritmi ai numeri, chiamiamo media geometrica delle quantità (positive) x_i l'espressione:

$$M_g = \sqrt[y_1+\dots y_n]{x_1^{y_1} \times \dots \times x_n^{y_n}}\,. \tag{16}$$

Ponendo nella (5) $m = \log M_g$ e sostituendo a x_i i loro logaritmi, può verificarsi che *è nulla la somma algebrica degli scarti:* $\log x_i - \log M_g = \log \frac{x_i}{M_g}$ ed inoltre *è minima la somma dei quadrati di questi scarti.* Dalla prima di queste due proprietà discende:

$$\left(\frac{x_1}{M_g}\right)^{y_1} \times \left(\frac{x_2}{M_g}\right)^{y_2} \times \dots \times \left(\frac{x_n}{M_g}\right)^{y_n} = 1\,.$$

L'impiego di M_g — che si assume sempre col segno positivo — è, tra l'altro, vantaggioso se, data una serie statistica formata da due successioni corrispondenti di cui una formi una progressione aritmetica e l'altra una progressione geometrica di quantità positive, si voglia determinare sulla seconda il valore corrispondente alla media aritmetica della prima.

È chiaro che tale valore corrispondente è dato dalla media geometrica. Infatti alle due successioni:

$$\begin{array}{llll} x_1, & x_1 + d, & x_1 + 2d, \dots & x_1 + (n-1)\,d \\ x'_1, & x'_1 q, & x'_1 q^2, \dots & x'_1 q^{n-1} \end{array}$$

corrispondono rispettivamente:

$$M = x_1 + \frac{n-1}{2}\,d \qquad M_g = x'_1 q^{\frac{n-1}{2}}\,.$$

Ad es., se le successive valutazioni annuali, per l'undicennio 1901-11, della popolazione di un dato paese si rivelassero in progressione geometrica, la popolazione per il 1906 corrisponderebbe alla media geometrica delle undici valutazioni annuali.

La media geometrica offre notevoli vantaggi, se si vuol risalire dalla media di un gruppo di rapporti alle medie dei loro termini e viceversa.

Infatti, la media geometrica dei rapporti:

$$\frac{x'_1}{x_1}, \frac{x'_2}{x_2}, \ldots\ldots \frac{x'_n}{x_n}$$

è:

$$M_g\left(\frac{x'}{x}\right) = \sqrt[n]{\frac{x'_1}{x_1} \times \ldots\ldots \times \frac{x'_n}{x_n}} = \frac{\sqrt[n]{x'_1 \times \ldots\ldots \times x'_n}}{\sqrt[n]{x_1 \times \ldots\ldots \times x_n}} = \frac{M_g(x')}{M_g(x)} \qquad (17)$$

ciò che dimostra come *la media geometrica di un gruppo di rapporti sia uguale al rapporto tra le medie geometriche dei termini di essi.*

Possiamo anche scrivere:

$$\frac{1}{M_g\left(\frac{x'}{x}\right)} = \frac{M_g(x)}{M_g(x')} = M_g\left(\frac{x}{x'}\right)$$

ed in particolare:

$$\frac{1}{M_g\left(\frac{1}{x}\right)} = M_g(x).$$

Coteste proprietà rendono, tra l'altro, consigliabile l'impiego della media geometrica nella costruzione delle medie dei rapporti, riferiti ad una data base, in quanto permettono di mutare rapidamente quest'ultima.

Infatti, nella seguente successione delle medie geometriche di n rapporti, aventi gli stessi denominatori:

$$\sqrt[n]{\frac{x'_1}{x_1} \times \frac{x'_2}{x_2} \times \ldots\ldots \times \frac{x'_n}{x_n}}$$

$$\sqrt[n]{\frac{x''_1}{x_1} \times \frac{x''_2}{x_2} \times \ldots\ldots \times \frac{x''_n}{x_n}}$$

. . . .

$$\sqrt[n]{\frac{x_1^{(m)}}{x_1} \times \frac{x_2^{(m)}}{x_2} \times \ldots\ldots \times \frac{x_n^{(m)}}{x_n}}$$

si può rapidamente portare la base al k^o termine, dividendo ogni rapporto medio per:

$$\sqrt[n]{\frac{x_1^{(k)}}{x_1} \times \frac{x_2^{(k)}}{x_2} \times \ldots\ldots \times \frac{x_n^{(k)}}{x_n}}$$

ossia per il k^o rapporto medio.

È facile verificare che questo procedimento non sarebbe corretto, se la successione media fosse calcolata con la media aritmetica.

In tal caso occorrerebbe ridurre alla nuova base i singoli rapporti e poi costruire su di essi la nuova successione media aritmetica.

Si abbiano, ad es., le seguenti tre successioni di rapporti percentuali:

Successione (1)	Successione (2)	Successione (3)	Successione media geometrica	Successione media aritmetica
100	100	100	100	100
96	102	104	100,61	100,67
92	98	86	91,87	92
91	88	91	89,99	90

Calcolandone la successione media geometrica (quarta colonna), il primo termine ne sarà la base (100); il secondo, calcolato mediante la (15), sarà:

$$\text{num}\,\frac{\log 96 + \log 102 + \log 104}{3} = 100,61$$

e così via.

Volendo trasportare la base di questa successione media ad es. nel secondo termine, non sarà necessario ricalcolare con la nuova base le singole successioni: basterà dividere ogni termine della successione media per: 1,0061.

Invece, dividendo ogni termine della successione media aritmetica (ultima colonna) per: 1,0067, si otterrà formalmente la base 100 nel secondo termine, ma si avranno risultati diversi da quelli, che correttamente si sarebbero ottenuti, dividendo anzitutto i termini di quelle tre successioni rispettivamente per 0,96, 1,02, 1,04 e calcolandone le medie aritmetiche: quei risultati ci direbbero soltanto in quale rapporto, mantenendo ferma la base nel primo termine, ogni termine della successione media sta al secondo; e nulla più.

Ciò spiega perchè l'uso della media geometrica incontri tanto favore nella costruzione di quelle speciali successioni di rapporti medi, che per lo più si assumono come numeri indici (index numbers) dei prezzi, ossia come variazioni medie percentuali del livello dei prezzi.

Il Consiglio Provinciale dell'Economia Corporativa di Milano rileva ogni settimana i prezzi medî all'ingrosso di 125 merci, che danno luogo a più frequenti transazioni o che comunque possono considerarsi come abbastanza rappresentative dell'andamento generale dei prezzi all'ingrosso nel nostro paese. Poi calcola i 125 rapporti di tali prezzi ai corrispondenti prezzi medi accertati nel 1913 e infine ricava da tali rapporti la media geometrica settimanale.

Coi dati settimanali calcola pure le medie geometriche mensili e annuali.

Poichè, anche entro il gruppo ristretto delle 125 merci, alcune hanno un'importanza maggiore di altre, si assumono per alcune merci (ad es. per il grano) i prezzi di diverse qualità, seguendo in tal modo il cosidetto metodo di ponderazione occulta.

Ora, tenendo presenti le proprietà della media geometrica, si potrà rapidamente trasportare la base di cotesti indici dal 1913 a qualsiasi altro tempo: ad es., conoscendo che dal marzo 1929 alla prima settimana di aprile 1933 quegl'indici sono discesi da 498,86 a 282,21, posti uguali a 100 i prezzi del 1913, possiamo affermare che, se il detto Consiglio Provinciale avesse eseguiti i suoi calcoli in base ai prezzi medî del marzo 1929, nell'aprile 1933 avrebbe ottenuto un indice di 282.21/4,9886 = 56.57 per cento del marzo 1929, che avrebbe dato la misura della gravità eccezionale della crisi economica, in cui tutto il mondo si è dibattuto in quel quadriennio.

Alcuni studiosi han cercato di dare un fondamento razionale all'uso della media geometrica nel calcolo dei numeri indici dei prezzi: è stato notato che per un gran numero di beni la distribuzione dei rapporti tra i prezzi di un tempo assegnato a quelli di un tempo base, ossia la distribuzione dei cosidetti «prezzi relativi» dà luogo ad una forma asimmetrica, a un dipresso come quella della tav. IX; e che, rappresentando graficamente tale distribuzione mediante i logaritmi dei prezzi relativi, si avrebbe una curva tendente alla simmetria. La media aritmetica dei logaritmi di quei rapporti corrisponderebbe, al rapporto più frequente o tipico del gruppo e, pertanto, tale rapporto più frequente sarebbe dato dalla media geometrica dei rapporti.

Ma ancora non si hanno verifiche abbastanza dimostrative del fatto, che la distribuzione dei logaritmi dei prezzi relativi tenda alla simmetria (1).

Il calcolo grafico della media geometrica, in base alla (16), riesce molto laborioso; ma, rappresentando i valori su scala logaritmica, può eseguirsi rapidamente sulla (15) col procedimento esposto al n. 7 per il calcolo grafico della media aritmetica.

9. Il valore prevalente (o normale) e il valore centrale (o mediano). – Alcuni valori osservati possono talvolta assumere l'importanza di caratteristiche di un gruppo. Tra essi meritano speciale men-

---

(1) Maurice Olivier, *Les nombres indices de la variation des prix*. Paris, Giard. 1927. Sui problemi relativi alla costruzione dei numeri indici dei prezzi, il volume citato contiene una scelta bibliografia. Cfr. pure Irving Fisher, *The making of index numbers*. London, Macmillan, 1922.

Nella «Biblioteca dell'economista», Serie V, Vol. XX, sono state tradotte e pubblicate tre memorie di F. Y. Edgeworth, presentate nel 1887-89 alla Commissione della Reale Società Britannica pel progresso delle scienze: sono fondamentali nello studio di questo argomento.

zione il valore prevalente del gruppo, che, se esiste, è quello al quale corrisponde il maggior numero di osservazioni; ed il valore centrale o i due valori centrali del gruppo (a seconda che il numero delle osservazioni sia dispari o pari), ossia quel valore o quei due valori consecutivi che dimezzano il gruppo, disposto in ordine di grandezza dei valori.

L'uso di queste nuove costanti è spesso consigliato dal fatto che esse sono indipendenti dai valori estremi, spesso mal noti, delle osservazioni e che meno delle medie di valori esse risentono l'eventuale grossolanità dei dati osservati.

Risulta chiaro, ad es., dalla tavola VI, che sin dal 1921 prevalevano nel nostro paese le famiglie con 3-4 componenti; che degli otto milioni e mezzo circa di famiglie, disposte in ordine crescente di componenti, la famiglia centrale (precisamente la 4.297.112^{ma} famiglia) aveva pure un numero di figli compreso fra 3 e 4; e che questi risultati non sarebbero stati punto perturbati da una diversa ripartizione delle famiglie nelle classi estreme.

A proposito della media aritmetica abbiamo parlato della vita media di un gruppo annuale di nati, come della somma degli anni vissuti da ognuno di essi, divisa per il numero dei nati medesimi. Ma, se i morti in età avanzata fossero di difficile accertamento (o perchè di alcuni vecchi si perdessero le tracce o perchè si ignorasse l'età della loro morte) e non fossero possibili buone correzioni, si potrebbe accertare la vita centrale o mediana del gruppo (ad es. di 100.000 nati), ossia l'età al decesso del 50.000^{mo} e del 50.001^{mo} morto, e trascurare quei morti di età avanzata, la cui distribuzione nel tempo non influirebbe punto sul nuovo risultato. Per questo alcuni demografi preferiscono la vita mediana (impropriamente chiamata vita probabile) alla vita media.

Inoltre il valore prevalente può ben determinarsi anche per le distribuzioni degli attributi qualitativi (nel qual caso si parla di modalità: ad es. è castano il colore prevalente nella tavola XIII), ed il valore o i valori centrali per le distribuzioni di attributi almeno graduabili, senza ricorrere all'arbitrio di una data graduatoria.

Osserviamo ancora che il valore o l'attributo prevalente in alcune distribuzioni può corrispondere alla caratteristica tipica del gruppo in esame, ossia — come sarà chiarito nella Parte II — al valore od all'attributo che le osservazioni assumerebbero per l'influenza di una circostanza o di un gruppo di circostanze costanti; e pertanto in alcune discipline, come in quelle biologiche, esso è spesso preferito sistematicamente alla media aritmetica (1).

Anche nella statistica criminale il valore prevalente ha speciale im-

---

(1) Cfr. tra gli altri: GIACINTO VIOLA, *La costituzione individuale*. Bologna, Cappelli, 1932.

portanza: « L'uomo incomincia di preferenza ad esercitare sui beni la sua tendenza al delitto. Dai 20 ai 30 anni, allorchè le sue forze sono complete, egli attenta anche alle persone: in questa età medesima la tendenza a tal genere di delitti raggiunge il suo maximum ». Sono parole scritte da Quetelet in quel libro IV della sua opera « Sur l'homme », che può considerarsi come il punto di partenza della moderna antropologia criminale e dei grandi contributi di Cesare Lombroso.

Si dice comunemente che la media aritmetica, quando non coincide col valore prevalente, dà un risultato puramente aritmetico, ed in questo senso si parla spesso di media aritmetica soggettiva o atipica o fittizia.

D'altra parte, il valore centrale o i due valori centrali consecutivi presentano la proprietà notevole di rendere minima la somma dei valori assoluti degli scarti calcolati rispetto ad essi (1).

La dimostrazione più elementare di questa proprietà può darsi nel modo seguente: Chiamando con $x_1, x_2, \dots x_N$ una successione di valori ordinati in guisa che $x_i \leq x_{i+1}$, la somma dei valori assoluti degli scarti tra x_1 ed i rimanenti valori sarà:

$$(x_2 - x_1) + (x_3 - x_1) + (x_4 - x_1) + \dots + (x_N - x_1) = \sum_{i=2}^{N} x_i - (N-1)\,x_1,$$

dove la lettera greca Σ (che si legge « sommatoria ») è il noto simbolo di addizione; tra x_2 ed i rimanenti valori sarà:

$$(x_2 - x_1) + (x_3 - x_2) + (x_4 - x_2) + \dots + (x_N - x_2) = \sum_{i=3}^{N} x_i - (N-3)\,x_2 - x_1 .$$

La differenza tra la prima e la seconda somma sarà:

$$(N-2)\,(x_2 - x_1),$$

e così via.

Calcolando tali somme e differenze rispetto ad ogni x_i, si ottiene la tavola XXXV.

Dall'ultima colonna di essa si rileva che le differenze tra una somma di scarti assoluti e la successiva dapprima sono positive e poscia negative; ed è pure facile constatare che per N dispari le somme corrispondenti a $x_{\frac{N-1}{2}}$, $x_{\frac{N+1}{2}}$, $x_{\frac{N+3}{2}}$ sono rispettivamente:

$$\sum_{\frac{N+1}{2}}^{N} x_i - x_{\frac{N-1}{2}} - \sum_{1}^{\frac{N-1}{2}} x_i , \qquad \sum_{\frac{N+3}{2}}^{N} x_i - \sum_{1}^{\frac{N-1}{2}} x_i , \qquad x_{\frac{N+3}{2}} + \sum_{\frac{N+3}{2}}^{N} x_i - \sum_{1}^{\frac{N+1}{2}} x_i$$

(1) LAPLACE (Le marquis de) *Théorie analytique des probabilités*. Deuxième supplément, février 1818, pag. 42 e seguenti. Cfr. la 3a edizione della *Théorie*, Paris, Courcier, 1820.

TAVOLA XXXV.

x_i	Somme degli scarti assoluti da x_i	Differenze tra le somme consecutive
x_1	$\sum_2^N x_i - (N-1)\,x_1$	$(N-2)\,(x_2 - x_1)$
x_2	$\sum_3^N x_i - (N-3)\,x_2 - x_1$	$(N-4)\,(x_3 - x_2)$
x_3	$\sum_4^N x_i - (N-5)\,x_3 - \sum_1^2 x_i$	
.	.	.
.	.	.
.	.	.
x_{N-2}	$\sum_{N-1}^N x_i + (N-5)\,x_{N-2} - \sum_1^{N-3} x_i$	$-(N-4)\,(x_{N-1} - x_{N-2})$
x_{N-1}	$x_N + (N-3)\,x_{N-1} - \sum_1^{N-2} x_i$	$-(N-2)\,(x_N - x_{N-1})$
x_N	$(N-1)\,x_N - \sum_1^{N-1} x_i$	

di cui la centrale è minima; e che per N pari le somme centrali corrispondenti a $x_{\frac{N}{2}}$ e $x_{\frac{N+2}{2}}$ sono minime ed uguali a

$$\sum_{\frac{N+2}{2}}^{N} x_i - \sum_{1}^{\frac{N}{2}} x_i .$$

Poichè coteste due somme minime e uguali fra loro possono scomporsi nella somma delle differenze assolute tra ogni coppia distinta di valori x_{N-i+1} e x_i, nella figura 33 si è rappresentata su un asse orizzontale una successione arbitraria di punti di ascisse x_1, x_2, ... e si sono fatte corrispondere quelle due somme minime alle somme dei segmenti paralleli tracciati al disopra e al disotto dell'asse.

Vi si rileva che, assumendo per N pari come valore centrale uno qualsiasi degli infiniti valori compresi tra i due centrali, resta costante la somma degli scarti assoluti e quindi valida anche per esso la proprietà dimostrata; e che, pertanto, per semplicità si può in ogni caso, sia N pari o dispari, ammettere l'esistenza di un valore mediano. È questo, appunto, l'uso ormai invalso nella pratica statistica, e generalmente, se N è pari e i due valori centrali non coincidono, si assume la semisomma di essi.

Infine riesce evidente quella scarsa sensibilità alle variazioni dei dati, che in taluni casi rende consigliabile — come abbiamo già detto — l'uso del valore centrale: basta considerare che il valore o la classe centrale di valori non si sposta, qualunque sia la posizione dei valori posti a sinistra o a destra di esso.

Fig. 33

Per questa via grafica la dimostrazione precedente sarebbe stata ben più rapida e intuitiva, ma la tav. XXXV ci sarà utile al n. 13 e successivi.

La proprietà di render minima la somma degli scarti numerici potrebbe far pensare che il valore centrale assicuri un adattamento ai dati migliore della media aritmetica e sia suscettibile di generale impiego; ma la scarsa sensibilità sopra illustrata rende illusorio quel migliore adattamento. Invece la media aritmetica è una funzione continua dei valori e, rendendo inoltre minima la somma dei quadrati degli scarti e nulla la somma algebrica di essi, raccoglie le condizioni ideall per un indice generale di posizione dei dati osservati.

Il calcolo del valore centrale non presenta difficoltà per un gruppo di osservazioni, di cui si conoscano i valori singoli, o gli attributi graduabili. Ad es. nella tavola V la vedova, che occupa il posto centrale nella graduatoria delle 227 vedove considerate, è la 114ma, a cui corrisponde un sol figlio minorenne: sarà dunque 1 il valore centrale o mediano della distribuzione. Nella tavola XIII sono due i militari che occupano il centro nella graduatoria dei 298.860 militari considerati: sono il 149.430^{mo} e il 149.431^{mo}, ai quali corrisponde il colore castano degli occhi (ammesso che questo colore non formi una classe). Esso, come si è detto, è anche l'attributo prevalente della distribuzione.

Ma non è facile la determinazione del valore centrale delle distribuzioni procedenti per classi di valori.

Per la tavola IX possiamo dire che, prescindendo dai valori ignoti, si avrà un valore centrale uguale alla semisomma dei due valori corrispondenti alla 41.830/2 = 20.915^{ma} ed alla (41.830 + 2)/2 = 20.916^{ma} osservazione e quindi evidentemente maggiore di 25 anni, età sino alla quale si hanno solo 13.801 osservazioni, e minore di 30 anni, età alla quale sono già comprese 26.286 osservazioni:

α - 15	5	5
15 - 20	2.324	2.329
20 - 25	11.472	13.801
25 - 30	12.485	26.286
.	.	.
.	.	.
.	.	.

ma non è possibile — almeno in base alle nozioni che sinora possediamo — determinarlo con precisione.

Quanto al valore prevalente, possiamo pure dire che, per la detta tavola, esso è verosimilmente compreso tra 25 e 30 anni; ma dobbiamo soggiungere che, a differenza del valore centrale, la determinazione di esso presenterebbe difficoltà anche nel caso di distribuzioni procedenti per valori singoli, e ciò a cagione sia dell'andamento irregolare dei gruppi di osservazione, sia dell'esistenza eventuale di parecchi massimi.

Però in alcuni casi semplici e, ove si rinunci ad una grande approssimazione, a mezzo di procedimenti grafici od in base a criterî intuitivi si possono ottenere risultati apprezzabili.

Ad es. rappresentando graficamente la distribuzione citata (figura 34, dove non abbiamo diviso preliminarmente il numero di osservazioni di ogni classe per l'intervallo delle classi, essendo questo costante) e ricavando graficamente dall'istogramma (*a*) le ordinate accumulate, osserviamo che — specie in corrispondenza alle età 20-30 anni — si può tracciare con piccolo arbitrio una curva (*b*) rappresentatrice della forma approssimativa, che si sarebbe ottenuta, se le osservazioni fossero state raccolte in classi infinitamente piccole.

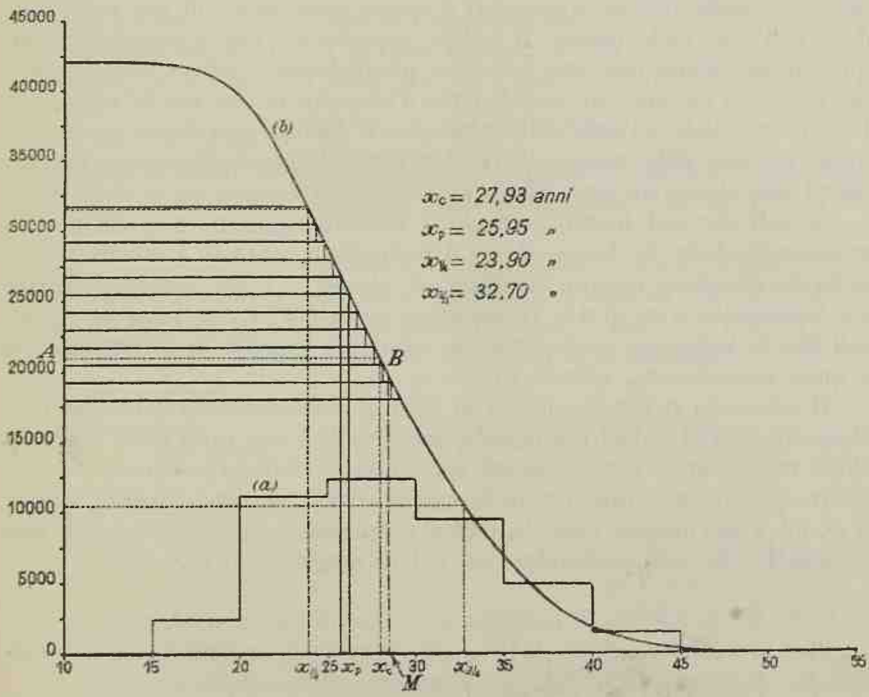


Fig. 34

Infatti il tipo della rappresentazione grafica, la relativa piccolezza e la costanza degl'intervalli della distribuzione, la natura del fenomeno considerato giustificano appieno la costruzione di una curva obbediente ai tre criterî: 1°) che essa passi per i punti assegnati; 2°) che negl'intorni dei punti estremi abbia pendenza nulla e negli intorni dei punti assegnati abbia pendenza uguale a quella di una retta passante per i due punti laterali; 3°) che presenti la massima semplicità.

Nè in base a tali criterî è possibile tracciare una curva sensibilmente diversa.

La seconda condizione è valida nell'ipotesi, generalmente plausibile, che alla curva empirica passante per tre punti consecutivi equidistanti si possa sostituire una parabola ordinaria [1].

Ora è chiaro che, dimezzando la prima ordinata (massima) di cotesta curva (b) e conducendo per il punto centrale A la parallela all'asse delle ascisse, questa incontrerà la curva (b) in B e la perpendicolare condotta da B all' asse delle ascisse segnerà su questo asse il valore centrale.

Sulla carta millimetrata risulta approssimativamente $x_c = 27{,}93$ anni.

È chiaro altresì, che al punto d'inflessione della (b), determinato dal passaggio dalle ordinate crescenti a quelle decrescenti di (a), corrisponderà sull'asse delle ascisse il valore prevalente. Per determinarlo con qualche precisione, conviene tracciare, parallelamente all'asse delle ascisse, dei segmenti equidistanti e dai punti d'incontro di essi con la curva abbassare parallele all'asse delle ordinate: il valore prevalente sarà compreso sull'asse delle ascisse tra le due parallele consecutive, che risulteranno più vicine tra loro. Si ha approssimativamente: $x_p = 25{,}95$ anni.

Si noti che nel nostro esempio è risultato $x_p < x_c$, ciò che era da attendersi, data la forma della distribuzione. Infatti l'addensamento della distribuzione intorno ai valori più piccoli fa sì che la perpendicolare corrispondente a x_p divide l'area della curva (a) in due parti disuguali, tali che la maggiore corrisponde ai valori più grandi: in corrispondenza a quest'area doveva, quindi, cadere x_c.

Il contrario si sarebbe avuto in caso di addensamento della distribuzione intorno ai valori più grandi; mentre, se i due rami della distribuzione fossero stati perfettamente simmetrici, avremmo avuto: $x_c = x_p$.

In quest'ultimo caso anche la media aritmetica sarebbe stata uguale a x_c ed a x_p; mentre, data la forma della nostra distribuzione, era ben necessario che essa presentasse un valore maggiore di x_p.

---

(1) Infatti nella parabola: $f(x) = a + b\,x + c\,x^2$, passante pei punti di coordinate $\left(-\frac{1}{2}, y_{-\frac{1}{2}}\right)$, $(0, y_0)$, $\left(\frac{1}{2}, y_{\frac{1}{2}}\right)$, si ha: $f'(0) = b = y_{\frac{1}{2}} - y_{-\frac{1}{2}}$.

Osserviamo ancora che, avendo già ottenuto per la nostra distribuzione: $M = 28{,}41$, possiamo scrivere: $x_p < x_c < M$ e che, se la distribuzione avesse presentato un addensamento verso i valori più grandi, avremmo invece avuto: $x_p > x_c > M$.

Coteste disuguaglianze sono caratteristiche dei tipi di curve ad un solo massimo e non fortemente asimmetriche, come quella considerata: il valore centrale è pure centrale tra il valore prevalente e la media aritmetica.

In esse, inoltre, si ha approssimativamente:

$$x_p = M - 3\,(M - x_c)\ , \tag{18}$$

formula che è spesso usata per la determinazione rapida di x_p nei detti tipi di curve, ma che dà spesso risultati alquanto lontani dal vero.

Nel nostro esempio si sarebbe avuto:

$$x_p = 28{,}41 - 3\,(28{,}41 - 27{,}93) = 26{,}97\ .$$

Abbiamo riportato nella figura anche i valori di $x_{1/4}$ e di $x_{3/4}$ (quartili), di cui ci occuperemo in seguito.

10. Lo scarto quadratico medio e la dispersione. – L'analisi del n. 5 ha messo in luce una caratteristica della media aritmetica, che è altra caratteristica di un gruppo di osservazioni.

Infatti, la somma dei quadrati degli scarti dalla media aritmetica:

$$(x_1 - M)^2\, y_1 + \ldots + (x_n - M)^2\, y_n$$

misurando il grado di approssimazione della media aritmetica ai valori osservati, risulta tanto maggiore, quanto maggiore è la divergenza che essi presentano fra loro.

Per rendere cotesta caratteristica indipendente dalla prima, occorre dividerla per il numero delle osservazioni; e per dare ad essa la stessa dimensione degli scarti originarî si potrà estrarne la radice quadrata.

Tenendo presente la definizione di momento diviso, data con la formula (10), risulta:

$$\sigma = \sqrt{\mu_{M,2}} = \sqrt{\frac{(x_1 - M)^2\, y_1 + \ldots\ldots + (x_n - M)^2\, y_n}{y_1 + \ldots\ldots + y_n}} \tag{19}$$

od anche, in forza della (11):

$$\sigma = \sqrt{\mu_{m,2} - \mu^2_{m,1}} = \sqrt{\frac{(x_1 - m)^2 x_1 + \cdots + (x_n - m)^2 y_n}{y_1 + \cdots + y_n} - \left(\frac{(x_1 - m)\, y_1 + \cdots + (x_n - m)\, y_n}{y_1 + \cdots + y_n}\right)^2} \tag{19bis}$$

dove m è un'origine arbitraria. Conoscendosi il valore di M, in base alla (9) si calcolerà rapidamente: $\mu^2_{m,1} = (M - m)^2$.

È questo lo scarto quadratico medio degli x_i, che, misurando la disuguaglianza media dei valori rispetto al loro livello medio, fornisce una misura della cosidetta dispersione (o variabilità) di essi, ossia della proprietà, che il fenomeno considerato presenta, di assumere differenti valori.

Ripigliando l'analogia col modello meccanico, di cui alla fig. 29, si sarebbe tentati di giustificare la somma dei quadrati degli scarti con le ragioni che in meccanica fanno considerare, oltre al cosidetto « momento statico », il « momento d'inerzia ».

Ma l'analogia avrebbe anche qui un valore puramente formale, perchè i dati statistici non sono punti materiali ma fatti sociali e naturali, che per mera convenzione si rappresentano graficamente.

Quando ci occupiamo, ad es., di distribuzione di redditi e di distribuzione per età di una popolazione; quando osserviamo i cambî di una data moneta, i prezzi di un dato bene o di un gruppo di beni, il numero dei matrimonî, delle nascite, dei decessi; noi constatiamo che i redditi e le età variano da individuo ad individuo, che i cambi, i prezzi, il numero dei matrimoni ecc., variano nel tempo, da luogo a luogo, e così via.

Una misura media dell'entità di tali variazioni può essere ottenuta in varî modi e solo criterî di convenienza possono guidare nella scelta: in generale s'impiega lo scarto quadratico medio, che dà una misura del grado di approssimazione dei valori osservati alla media aritmetica, ossia al valore comune che essi assumerebbero in caso di equidistribuzione.

Questa misura di dispersione è nulla, se tutti i valori sono uguali tra loro, è una funzione crescente (in valore numerico) di tutti gli scarti numerici ed ha il grande pregio di prestarsi facilmente agli sviluppi analitici e di essersi rivelata particolarmente feconda nell'analisi degli schemi teorici e delle relazioni tra gruppi di osservazioni, di cui ci occuperemo nelle parti successive di questo manuale.

Se per scopi particolari tali motivi si ritengano di scarso peso, nulla vieta che si ricorra ad altri indici di dispersione, che meglio rispondano a quegli scopi: al n. 13 faremo anzi alcune esemplificazioni a questo riguardo. Ma si tratterà di esigenze speciali.

Si noti, infine, che, trattandosi di attributi graduabili, lo scarto quadratico medio potrà calcolarsi sulle graduatorie di tali attributi, con le avvertenze che a questo proposito abbiamo già fatte riguardo alla media aritmetica.

Applichiamo adesso la (19) per $y_i = 1$, alla serie statistica delle stazze nette delle navi a vela, di cui alla tavola XIV. Le operazioni sono state riportate nella tavola XXXVI e dal totale dell'ultima colonna si ricava:

$$\sigma = \sqrt{\frac{138.566{,}07}{24}} = \sqrt{5.773{,}59} = \pm\, 75{,}98 \text{ migliaia di tonn.}$$

In generale si attribuisce a σ il segno positivo.

Anche σ, come M, è insensibile all'ordine di successione dei valori ed è quindi indipendente dal fenomeno, in base al quale la serie è stata

TAVOLA XXXVI.

Tonnellate di stazza netta delle navi a vela (in migliaia) x_i	x_i — 524,4	$(x_i - 524{,}4)^2$	Tonnellate di stazza netta delle navi a vela (in migliaia) x_i	x_i — 524,4	$(x_i - 524{,}4)^2$
634,1	+ 109,7	12.034,09	570,4	+ 46,0	2.116,00
625,8	+ 101,4	10.281,96	584,2	+ 59,8	3.576,04
609,8	+ 85,4	7.293,16	570,4	+ 46,0	2.116,00
588,3	+ 63,9	4.083,21	541,2	+ 16,8	282,24
571,6	+ 47,2	2.227,84	503,3	— 21,1	445,21
555,6	+ 31,2	973,44	468,7	— 55,7	3.102,49
527,6	+ 3,2	10,24	453,8	— 71,1	5.055,21
526,8	+ 2,4	5,76	439,9	— 84,5	7.140,25
537,6	+ 13,2	174,24	432,7	— 91,7	8.408,89
558,2	+ 33,8	1.142,44	411,0	— 113,4	12.859,56
568,2	+ 43,8	1.918,44	374,8	— 149,6	22.380,16
575,2	+ 50,8	2.580,64	356,0	— 168,4	28.358,56
Totale					138.566,07

ordinata: il risultato precedente coinciderebbe con quello che avremmo ottenuto sulle 24 osservazioni disposte in base a qualsiasi altra circostanza, ad es. in base a quella della tavola XV.

Altro esempio è dato dalla fig. 35, dove si rappresentano gl'indici mensili delle variazioni del volume degli scambî nel nostro paese (ai prezzi

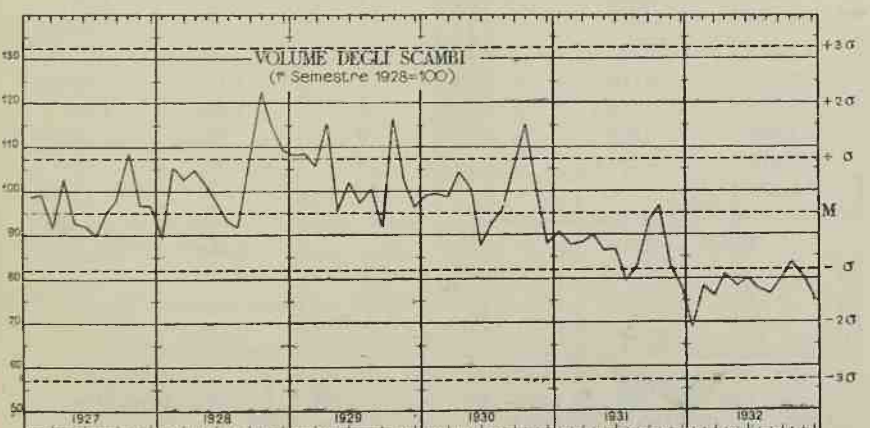


Fig. 35

del 1º semestre 1928 e posto uguale a 100 il detto semestre, immediatamente successivo alla rivalutazione della lira) durante il sessennio 1927-32. I valori della media aritmetica e dello scarto quadratico medio sono rappresentati da rette parallele all'asse delle ascisse. Abbiamo, inoltre, rappresentato i valori di alcuni multipli interi dello scarto quadratico medio per ragioni che saranno chiarite in seguito.

Applicheremo adesso la (19 *bis*) alla distribuzione della tavola VIII. Ammettendo l'ipotesi che ai valori di ogni classe si possa sostituire la semisomma degli estremi di essa (ipotesi tanto più plausibile, quanto minore è l'intervallo delle classi), e considerando che la distribuzione procede ad intervalli costanti, possiamo — come nella tavola XXXIV — ridurre gli scarti numerici, a partire da un'origine arbitraria, alla successione dei numeri interi.

TAVOLA XXXVII.

Altezze barometriche in centesimi di pollice x_i	Giorni y_i	$\frac{x_i - 2.995}{20}$	$\left(\frac{x_i - 2.995}{20}\right)^2$	$\frac{x_i - 2.995}{20} y_i$	$\left(\frac{x_i - 2.995}{20}\right)^2 y_i$
2.855	3	— 7	49	— 21	147
2.875	6	— 6	36	— 36	216
2,895	22	— 5	25	— 110	550
2.915	58,5	— 4	16	— 234	936
2.935	187	— 3	9	— 561	1.683
2.955	436	— 2	4	— 872	1.744
2.975	812	— 1	1	— 812	812
2.995	1.151	0	0	0	0
3.015	1.119,5	+ 1	1	+ 1.119,5	1.119,5
3.035	619,5	+ 2	4	+ 1.239	2.478
3.055	278	+ 3	9	+ 834	2.502
3.075	50,5	+ 4	16	+ 202	808
3.095	5	+ 5	25	+ 25	125
Totali	4.748			— 2.646 + 3.419,5 773,5	13.120,5

$$\mu_{m,1} = \frac{20 \times 773{,}5}{4.748} = 3{,}2582 \qquad \mu_{m,2} = \frac{20^2 \times 13.120{,}5}{4.748} = 1.105{,}3496$$

$$\sigma = \sqrt{1.105{,}3496 - 10{,}6159} = \sqrt{1.094{,}7337} = \pm 33 \text{ centesimi di pollice}$$

Essendo $M = 2.995 + 3{,}2582 = 2.998$ centesimi di pollice, nella figura 36 — dove abbiamo ritenuto inutile dividere preliminarmente gli y_i per l'intervallo costante — lo scarto quadratico medio corrisponde alle due ascisse $M \pm \sigma = 2.998 \pm 33$ ed è rappresentato da due rette parallele all'asse delle ordinate. Nella detta figura abbiamo inoltre rappresentato i valori di alcuni multipli interi dello scarto quadratico medio.

Sull'asse delle ascisse abbiamo anche riportato i valori limiti delle classi della distribuzione sotto forma di scarti $x_i - M$ (seconda scala), ed abbiamo scritto in una terza scala questi ultimi sotto forma di multipli dello scarto quadratico medio (scarti ridotti).

Inoltre, sull'asse delle ordinate abbiamo espresso in altra scala il numero delle osservazioni sotto forma di *frequenze* (n. 1 del primo capitolo), ossia mediante i rapporti:

$$f_i = \frac{y}{y_1 + \cdots\cdots + y_n}$$

che nell'esempio sono stati ottenuti, dividendo gli y_i per 4.748; ed abbiamo, infine, congiunto convenientemente fra loro, con segmenti rettili-

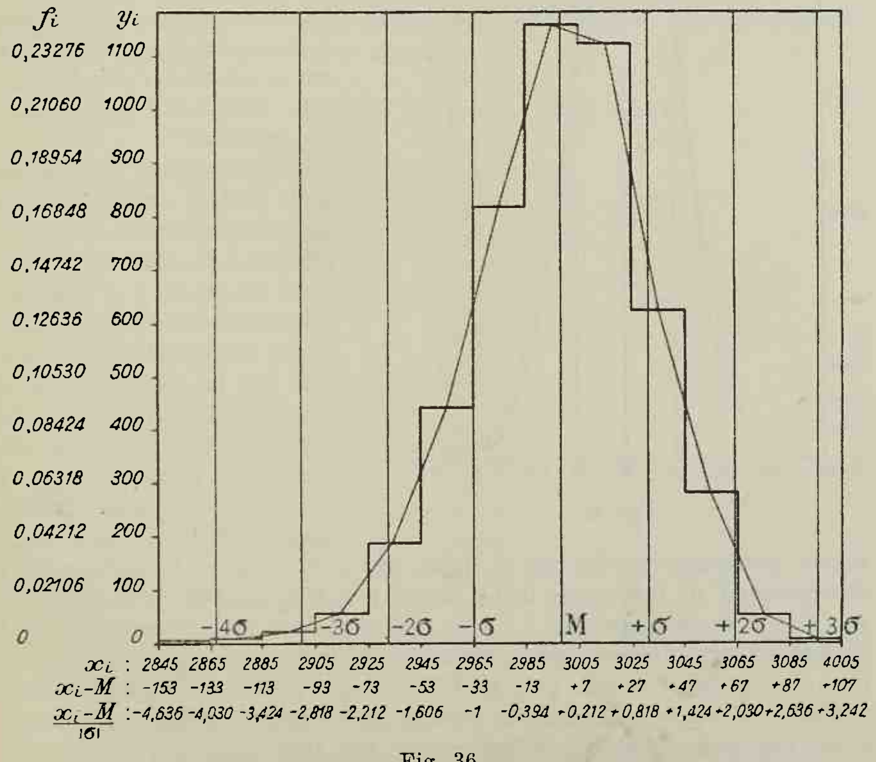


Fig. 36

nei, le altezze dei rettangoli adiacenti, in guisa da ottenere un andamento ancor più approssimato a quello, che avrebbe assunto la curva di distribuzione, se le osservazioni fossero state raccolte in classi infinitamente piccole.

Quest'ultimo metodo, detto dei « poligoni di frequenze », è spesso usato nella costruzione dei diagrammi areali e ne abbiamo dato esempio con la fig. 31. Ma, se esso è comodo pei calcoli grafici — condotti sull'ipotesi che ai valori delle classi si possa sostituire la semisomma dei limiti di esse —, è grossolano al fine di rappresentare la forma della distribuzione.

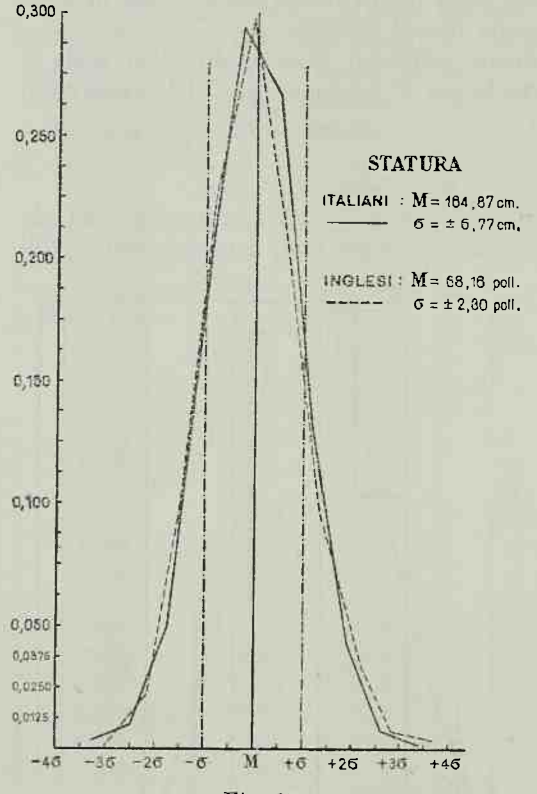


Fig. 37

Ammettendo di avere diviso gli y_i per gl' intervalli delle classi, un metodo migliore è quello di far passare una curva regolare tra le maglie dell'istogramma, in modo da rispettare l' uguaglianza tra l'area di ogni rettangolo e l'area corrispondente della curva adottata; ma i risultati variano dall'uno all'altro disegnatore.

In base alla terza scala delle ascisse (multipli dello scarto quadratico medio) e alla seconda scala delle ordinate (frequenze), si possono eseguire interessanti confronti tra distribuzioni, nell'ipotesi che la dispersione e il numero delle osservazioni fossero uguali: basterà sovrapporre i grafici calcolati con le stesse scale di frequenze e di multipli dello scarto quadratico medio. Ad es. nella figura 37 sono stati sovrapposti i due poligoni di frequenze delle tavole I e VII, usando le stesse scale per σ e le stesse scale per f_i; e si può affermare che, prescindendo dal diverso numero di osservazioni delle due distribuzioni e dalla loro rispettiva dispersione, le forme sono simili e che solo per le stature degl'italiani si nota una leggera asimmetria rispetto all'ordinata centrale con una sensibile prevalenza delle stature comprese tra M e $+\sigma$.

Tale prevalenza potrebbe giustificarsi colle proporzioni dei diversi tipi etnici, che compongono la nostra massa di coscritti.

Uguali sovrapposizioni possono eseguirsi con le serie statistiche: ad es. sulla figura 35 noi potremo disegnare con la stessa scala di σ, e a partire da M, l'indice delle variazioni del volume del commercio estero, al fine di esaminare se — *indipendentemente dalla diversa dispersione di tali fenomeni* — l'uno sia diminuito più fortemente dell'altro. Questi ultimi confronti sono frequentissimi nelle analisi economiche [1], dove s'incontrano fenomeni poco variabili, come ad es. i cambî della lira dopo la rivalutazione monetaria, e fenomeni profondamente variabili, eppure spesso strettamente connessi coi primi, come ad es. l'ammontare della circolazione monetaria: se non si elimina l'influenza della diversa dispersione, non è possibile mettere in luce le relazioni che legano cotesti fenomeni.

11. Il calcolo elementare dei momenti a mezzo di osservazioni accumulate. – Nel caso di distribuzioni ad intervalli costanti h, i calcoli esposti nei numeri precedenti possono essere ancora semplificati, scegliendo l'origine arbitraria m in modo che soddisfi all'uguaglianza: $m = x_1 - h$, ossia tale che si abbia:

$$\frac{x_1 - m}{h} = 1, \quad \frac{x_2 - m}{h} = 2, \; \ldots\ldots \; \frac{x_n - m}{h} = n$$

Infatti, in base alle espressioni:

$$\sum_1^n y_i = y_n + y_{n-1} + \cdots\cdots + y_2 + y_1$$

$$\sum_2^n y_i = y_n + y_{n-1} + \cdots\cdots + y_2$$

$$\cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot$$

$$\sum_{n-1}^n y_i = y_n + y_{n-1}$$

$$\sum_n^n y_i = y_n$$

possiamo scrivere:

$$\sum_1^n y_i + \cdots + \sum_n^n y_i = \sum_1^n i y_i = \sum_1^n \left(\frac{x_i - m}{h}\right) y_i = \frac{\sum_1^n y_i}{h} \mu_{m,1} \tag{20}$$

[1] Su tali applicazioni, cfr. il volume VI della « Nuova collana, ecc. », op. cit.

od anche, ponendo per brevità: $\sum\limits_1^n{}' y_i = \sum\limits_1^n y_i + ... + \sum\limits_n^n y_i$,

$$\mu_{m,1} = \frac{h}{\sum\limits_1^n y_i} \sum_1^n{}' y_i \,. \tag{21}$$

Dalle altre espressioni:

$$\begin{aligned}
\sum_1^n{}' y_i &= \sum_n^n y_n + \sum_{n-1}^n y_{n-1} + ... + \sum_2^n y_2 + \sum_1^n y_1 &&= n\, y_n + (n-1)\, y_{n-1} + ... + 2\, y_2 + y_1 \\
\sum_2^n{}' y_i &= \sum_n^n y_n + \sum_{n-1}^n y_{n-1} + ... + \sum_2^n y_2 &&= (n-1)\, y_n + (n-2)\, y_{n-1} + ... + y_2 \\
&\ldots\ldots\ldots\ldots &&\ldots\ldots\ldots\ldots \\
\sum_{n-1}^n{}' y_i &= \sum_n^n y_n + \sum_{n-1}^n y_{n-1} &&= 2\, y_n + y_{n-1} \\
\sum_n^n{}' y_i &= \sum_n^n y_n &&= y_n
\end{aligned}$$

si ricava:

$$\sum_1^n{}' y_i + ... + \sum_n^n{}' y_i = \sum_1^n \frac{(i+1)\, i}{2} y_i = \frac{1}{2}\left[\sum_1^n \left(\frac{x_i - m}{h}\right)^2 y_i + \sum_1^n \left(\frac{x_i - m}{h}\right) y_i \right] =$$

$$= \sum_1^n y_i \cdot \frac{1}{2} \left(\frac{1}{h^2}\, \mu_{m,2} + \frac{1}{h}\, \mu_{m,1} \right)$$

da cui, ponendo: $\sum\limits_1^n{}'' y_i = \sum\limits_1^n{}' y_i + ... + \sum\limits_n^n{}' y_i$

si ha:

$$\mu_{m,2} = \frac{h^2}{\sum\limits_1^n y_i} \left(2 \sum_1^n{}'' y_i - \sum_1^n{}' y_i \right) \tag{22}$$

e così via.

Ognuno vede quanto risulta in tal guisa semplificato il calcolo dei momenti divisi e, in particolare, della media aritmetica e dello scarto quadratico medio (¹).

Ad es., sommando successivamente le frequenze (su 10.000), con cui si sono presentati nelle classi quinquennali di età (limiti superiori esclusi) i maschi coniugatisi nel nostro paese durante il 1926, si ricava rapidamente:

(¹) Cfr. W. Palin Elderton, op. cit. Nella distribuzione di frequenze (relative) della tavola seguente, per non modificare i simboli, usiamo y_i invece di f_i.

TAVOLA XXXVIII.

Età degli sposi, in anni	Frequenze degli sposi	$\sum_{k}^{n} y_i$	$\sum_{k}^{n}{}' y_i$
15 – 20	430	10.000	32.219
20 – 25	3.249	9.570	22.219
25 – 30	3.610	6.321	12.649
30 – 35	1.312	2.711	6.328
35 – 40	575	1.399	3.617
40 – 45	295	824	2.218
45 – 50	174	529	1.394
50 – 55	118	355	865
55 – 60	82	237	510
60 – 65	73	155	273
65 – 70	46	82	118
70 – 75	36	36	36
Totali	$\sum_{1}^{n} y_i = 10.000$	$\sum_{1}^{n}{}' y_i = 32.219$	$\sum_{1}^{n}{}'' y_i = 82.446$

$$\mu_{m,1} = \frac{5}{10.000} \times 32.219 = 16{,}11$$

$$\mu_{m,2} = \frac{25}{10.000} (2 \times 82.446 - 32.219) = 331{,}68$$

Media aritmetica: $= m + \mu_{m,1} = 12{,}5 + 16{,}11 = 28{,}61$

Scarto quadratico medio: $\sqrt{\mu_{m,2} - \mu^2_{m,1}} = \sqrt{72{,}15} = \pm 8{,}49$

Potrebbe, adunque, affermarsi che l'età media dei maschi coniugatisi nel nostro paese durante il 1926 è stata poco maggiore di 28 anni e mezzo; e che in media tale età è variata di 8 anni e mezzo al disopra o al disotto della media, ossia del 29 per cento.

Cotesto procedimento elementare di calcolo può tradursi immediatamente in forma grafica, la quale presenta speciali vantaggi, se possa ammettersi la continuità dei valori osservati e si possa con minimo arbitrio sostituire alla spezzata empirica la linea, che presumibilmente rappresenterebbe l'andamento della distribuzione, se quei valori fossero raccolti in classi infinitamente piccole. Infatti in tal caso è possibile la cosidetta integrazione grafica successiva della funzione rappresentatrice di tale andamento e un calcolo più corretto dei momenti.

Questo argomento trovasi ampiamente svolto nei trattati di calcolo grafico (1).

(1) Cfr. GINO CASSINIS, *Calcoli numerici, grafici e meccanici*. Pisa, Mariotti-Pacini, 1928-VI.

12. Frequenze degli scarti eccedenti dati limiti. – La figura 36. nella forma in cui è stata costruita, permette di fare alcune interessanti considerazioni sulle frequenze degli scarti numericamente maggiori o uguali a dati multipli dello scarto quadratico medio.

Notiamo, anzitutto, che l'ammontare di tali frequenze è in relazione alla forma della distribuzione e che, se in base alla scala adottata in detto diagramma per i valori e le frequenze, la figura avesse presentato una forma più spianata, quelle frequenze sarebbero state maggiori.

In generale, però, qualunque sia la forma della distribuzione, può calcolarsi un limite superiore delle frequenze corrispondenti a scarti numericamente maggiori o uguali a $k\ |\sigma|$, dove è $k > 0$.

Infatti, considerando l'espressione:

$$\mu_{m,2} = a_1^2 f_1 + \ldots\ldots + a_n^2 f_n$$

dove a_i sono scarti da un'origine qualsiasi m, osserviamo che in essa alcuni scarti: a_1, ... a_{t-1} saranno in valore assoluto minori di una costante positiva l ed altri scarti: a_t, ... a_n, saranno in valore assoluto uguali o maggiori di l.

Ora, sostituendo alla precedente somma di prodotti quella dei prodotti di l^2 per le frequenze f_t .... f_n, corrispondenti agli scarti in valore assoluto uguali o maggiori di l, sarà:

$$\mu_{m,2} \geq l^2 \left(f_t + \ldots\ldots + f_n\right) \tag{23}$$

od anche:

$$\frac{\mu_{m,2}}{l^2} \geq f_t + \ldots\ldots + f_n$$

Scrivendo l sotto forma di multiplo di $|\sqrt{\mu_{m,2}}|$ ossia: $l = k\ |\sqrt{\mu_{m,2}}|$, si avrà infine:

$$\frac{1}{k^2} \geq F_k \tag{24}$$

che dimostra, appunto, che $1/k^2$ è limite superiore delle frequenze F_k corrispondenti a scarti in valore assoluto maggiori o uguali a $k\ |\sqrt{\mu_{m,2}}|$ e quindi, in particolare, maggiori o uguali a $k\ |\sigma|$.

In base a tale risultato, possiamo ad es. affermare che, qualunque sia la forma della distribuzione, gli scarti in valore assoluto maggiori o uguali a 2, 3,... volte lo scarto quadratico medio non possono abbracciare rispettivamente più di 1/4, 1/9,.... delle osservazioni [1].

Questo risultato — come si vedrà nella Parte II — è interessante dal punto di vista teorico, ma fornisce limiti troppo larghi per le applicazioni numeriche.

---

[1] Cfr. Fréchet et Halbwachs, *Le calcul des probabilités à la portée de tous.* Paris, Dunod, 1924.

Si noti che la (23) può estendersi a tutti i momenti divisi di grado pari, ossia a $\mu_{m,2h}$, dove h è un intero positivo:

$$\mu_{m,2h} \geq l^{2h} (f_t + \ldots\ldots + f_n)$$

donde:

$$\frac{\mu_{m,2h}}{k^{2h} \mu^h{}_{m,2}} \geq F_h \tag{25}$$

La (23) potrebbe pure estendersi ai momenti divisi di grado dispari, solo nel caso che in essi gli scarti si considerassero nel loro valore assoluto (1).

Partendo dalla disuguaglianza:

$$a_t^{2h} f_t + \ldots\ldots + a_n{}^{2h} f_n \geq l^{2h} (f_t + \ldots\ldots + f_n)$$

— che potrebbe pure estendersi ai momenti divisi di grado dispari, considerando gli scarti nel loro valore assoluto — si ottiene:

$$\frac{a_t^{2h} f_t + \ldots\ldots + a_n{}^{2h} f_n}{k^{2h} \mu^h{}_{m,2}} \geq F_h \tag{26}$$

Per $h = 1$, quest'ultimo risultato può scriversi nella forma:

$$\frac{a^2{}_t f_t + \ldots\ldots + a^2{}_n f_n}{a_1{}^2 f_t + \ldots\ldots + a^2{}_n f_n} \geq k^2 \frac{y_t + \ldots\ldots + y_n}{y_1 + \ldots\ldots + y_n}$$

la quale tra l'altro mostra che, assegnato un dato rapporto tra la somma dei quadrati degli scarti che in valore assoluto eccedono un dato limite e la somma dei quadrati di tutti gli scarti, non è più in nostro arbitrio la determinazione del rapporto tra il numero delle osservazioni eccedenti quel limite e il numero totale delle osservazioni; e viceversa.

13. Indici speciali di dispersione: lo scarto medio interquartile, lo scarto numerico medio, la differenza media. – È possibile calcolare altre misure di dispersione, atte a soddisfare speciali esigenze.

A. – Determinando i valori, che dividono un gruppo in quattro o in dieci parti, aventi un ugual numero di osservazioni, si sono costruiti indici di dispersione, dotati di quella minore sensibilità, per cui si raccomanda in alcuni casi l'uso del valore centrale (o mediano).

Tali valori si calcolano con procedimento analogo a quello esposto per quest'ultimo, e, quando i risultati forniscano due valori diversi, si suol prenderne pure la semisomma. Graficamente basta dividere in quattro o in dieci parti l'ordinata massima della curva (*b*) nella figura 34 e seguire per ogni punto di divisione il metodo già esposto per il calcolo di x_c.

(1) Cfr. P. Medolaghi, *La teoria del rischio e le sue applicazioni*, in « Atti del IV congresso degli attuarî ». Vienna, 1909; e F. P. Cantelli, *Intorno ad un teorema fondamentale della teoria del rischio*, in « Boll. dell' Assoc. degli attuarî italiani ». Milano, 1910.

Dividendo, ad es., la curva in quattro parti, designando con $x_{1/4}$ e $x_{3/4}$ e chiamando primo e terzo quartile i valori preceduti rispettivamente da 1/4 e da 3/4 delle osservazioni, è stato proposto come indice di dispersione il cosidetto « scarto medio interquartile », dato dalla metà della differenza interquartile (1):

$$q = \frac{1}{2}\left(x_{3/4} - x_{1/4}\right) \tag{27}$$

Nella detta figura 34 sono stati calcolati graficamente i valori dei detti quartili, dai quali si ricava:

$$q = \frac{1}{2}\,(32{,}70 - 23{,}90) = 4{,}4 \text{ anni} \ ,$$

risultato da preferire allo scarto quadratico medio, se tra l'altro non si ritenesse lecito ammettere, che alle classi estreme aperte si possa attribuire lo stesso intervallo quinquennale delle altre classi.

B. – Invece di q, s'impiega anche la media aritmetica degli scarti assoluti dal valore centrale — la quale, come si è dimostrato, è un minimo — :

$$s = \frac{\sum_{1}^{n} \left| x_i - x_c \right| y_i}{\sum_{1}^{n} y_i} \ . \tag{28}$$

Talvolta nella formula precedente si sostituisce x_c con M, e l'indice risultante — chiamato scarto numerico medio — è usato invece di σ, in vista della maggior semplicità e speditezza dei calcoli ed anche quando si richiedano risultati pienamente sensibili alle divergenze dei valori osservati.

Ma un uso così generale è pericoloso — non solo perchè la media degli scarti assoluti mal si presta agli sviluppi algebrici, di cui parleremo nelle parti successive di questo manuale, e non presenta rispetto a σ quegli altri vantaggi, di cui ci siamo occupati al n. 10 —, ma anche perchè, sostituendo nella (28) M a x_c non si ottengono indici dotati della piena sensibilità di σ. Basti considerare che, se a due scarti uguali nel valore numerico e nel segno (e quindi corrispondenti a uguali valori osservati) si sostituissero due scarti diversi ma tali che la loro somma fosse uguale a quella dei precedenti, si avrebbe:

$$2\,a = a + d + a - d \ ,$$

(1) Francis Galton, *Natural inheritance*. London, Macmillan, 1889; e G. Udny Yule, op. cit.

e, se i nuovi scarti avessero lo stesso segno dei primi, lo scarto numerico medio sarebbe il medesimo.

Invece, considerando i quadrati degli scarti, si avrebbe sempre, per d diverso da zero:

$$2\,a^2 < (a + d)^2 + (a - d)^2 = 2\,a^2 + 2\,d^2\ ,$$

e quindi lo scarto quadratico medio risulterebbe tanto maggiore, quanto maggiore fosse la divergenza tra gli scarti.

C. – Un altro indice di dispersione è stato escogitato in base alle differenze tra ogni valore e tutti gli altri.

È una vecchia idea, ripetutamente riesumata (1), quella che la media aritmetica delle differenze prese in valore assoluto $|x_i - x_k|$, corrispondenti a tutte le possibili disposizioni binarie dei valori, può essere assunta come indice di variabilità in quei casi, in cui convenga una misura, che non implichi alcun indice di posizione dei valori.

Ricordando che le disposizioni binarie di N elementi $x_1, x_2, \ldots x_N$ sono i gruppi che si possono formare riunendo questi elementi a due a due, alla condizione di considerare diversi l'uno dall'altro i gruppi formati con elementi identici ma diversamente disposti, sarà facile persuadersi che per quegli N elementi il numero delle disposizioni binarie che contengono l'elemento x_1 sarà $N - 1$, il numero delle disposizioni binarie che contengono l'elemento x_2 sarà pure $N - 1$, ecc.; onde si avranno in totale $N(N - 1)$ disposizioni binarie e la media delle differenze assolute corrispondenti avrà la forma:

$$\frac{\sum\limits_{i=1}^{N}\sum\limits_{k=1}^{N-1} \left| x_i - x_k \right|}{N(N-1)} \tag{29}$$

Si noti che la somma di tali differenze assolute non è che la somma delle espressioni contenute nella seconda colonna della tavola XXXV e che, nel caso generale che agli elementi $x_1, x_2, \ldots\ldots x_n$, corrispondano rispettivamente $y_1, y_2, \ldots\ldots y_n$ osservazioni, la differenza media sarà:

$$\Delta = \frac{\sum\limits_{i=1}^{n} y_i \sum\limits_{k=1}^{n-1} y_k \left| x_i - x_k \right|}{\sum\limits_{i=1}^{n} y_i \left(\sum\limits_{i=1}^{n} y_i - 1 \right)}\,. \tag{29 bis}$$

(1) E. Czuber, *Wahrscheinlichkeitsrechnung*. Leipzig, Teubner, 1908-10 (Vol. I, pag. 274). – G. Furlan, *Neue literatur zur einkommensverteilung in Italien*, in « Jahrbücher für nationaloekonomie und statistik », agosto 1911, pag. 255. – C. Gini, *Variabilità e mutabilità*. Bologna, Cuppini, 1912.

Poichè il calcolo diretto di questa formula è in generale molto laborioso, è vantaggioso ricorrere a formule semplificatrici. Una delle più usate (¹) è la seguente:

$$\Delta = \frac{2 \sum_{i=1}^{n} (A_i - B_i)\, y_i}{\sum_{i=1}^{n} y_i \left(\sum_{i=1}^{n} y_i - 1 \right)} \qquad (29\ ter)$$

dove A_i e B_i sono le somme dei prodotti $x_i\, y_i$, che occupano i posti non inferiori e, rispettivamente, non superiori ad i.

Tali somme, per $y_i = 1$, sono quelle che si ottengono sommando successivamente dal basso in alto e, rispettivamente, dall'alto in basso i valori osservati graduati in ordine crescente.

Tavola XXXIX.

x_i	y_i	$x_i\, y_i$	A_i	B_i	$A_i - B_i$	$(A_i - B_i)\, y_i$
12,5	5	62,5	1.188.475	62,5	+ 1.188.412,5	+ 5.942.062,5
17,5	2.324	40.670	1.188.412,5	40.732,5	+ 1.147.680	+ 2.667.208.320
22,5	11.472	258.120	1.147.742,5	298.852,5	+ 848.890	+ 9.738.466.080
27,5	12.485	343.337,5	889.622,5	642.190	+ 247.432,5	+ 3.089.194.762,5
32,5	9.044	293.930	546.285	936.120	— 389.835	— 3.525.667.740
37,5	4.897	183.567,5	252.355	1.119.687,5	— 867.332,5	— 4.247.327.252,5
42,5	1.474	62.645	68.787,5	1.182.332,5	— 1.113.545	— 1.641.365.330
47,5	126	5.985	6.142,5	1.188.317,5	— 1.182.175	— 148.954.050
52,5	3	157,5	157,5	1.188.475	— 1.188.317,5	— 3.564.952,5
Totali	41.830					+ 15.500.811.225
						— 9.566.879.325
						5.933.931.900

Ad es. sulla tavola IX i calcoli si eseguono nel modo esemplificato nella tavola XXXIX, dalla quale si ricava:

$$\Delta = \frac{2 \times 5.933.931.900}{41.830 \times 41.829} = 6{,}783 \text{ anni.}$$

Questo risultato ci dice che ogni età differisce dalle altre in media di 6,783 anni, e dovrà preferirsi quando — come dicevamo — convenga

(¹) F. Vinci, *Statistica metodologica* (Lezioni tenute nel R. Istituto Superiore Commerciale di Bari). Padova, Cedam, 1924 (pag. 54 e segg.) – E. Czuber, *Beitrag zur theorie statistischer reihen* (Versicherungswissenschaftliche mitteilungen, Vol. IX, Fasc. II), Wien, 1914. – A. De Gleria, in « Rivista it. di statistica », 1929 (ottobre) e 1930 (giugno).

una misura di dispersione, che non implichi alcun indice di posizione dei valori e — a differenza dallo scarto medio interquartile — sia funzione dei valori, anzichè della graduatoria di essi.

Ma, di fatto, siffatta convenienza si è rivelata molto rara nell'analisi statistica.

14. La dispersione relativa e la concentrazione. - Dividendo lo scarto quadratico medio per la media aritmetica, si ottiene una misura relativa della dispersione, da alcuni chiamata « coefficiente di variazione ».

Possiamo dire, ad es., per la tavola XXXVII che l'altezza barometrica giornaliera è variata in media di 33 centesimi di pollice e che, essendo:

$$100 \frac{\sigma}{M} = \frac{\pm 3.300}{2.998} = \pm 1.1 ,$$

tale variazione media corrisponde a circa l'uno per cento del livello medio dell'altezza barometrica.

Per la tavola XXXIV otteniamo $\sigma = \pm 6{,}1563$, che diviso per 28,4137 dà una variazione media dell'età delle madri di $\pm$ 21,7 per cento.

L'importanza di questi risultati sta nel fatto, che essi sono indipendenti dall'ordine di grandezza e dall'unità di misura dei fenomeni considerati. Il peso della formica è più o meno variabile di quello dell'elefante? L'uomo è più variabile per statura o per peso? È vero che i prezzi dei beni voluttuari sono nel tempo più variabili di quelli dei cosidetti generi di prima necessità? I redditi individuali sono più o meno variabili delle capacità intellettuali? Ecco domande alle quali si potrebbe con fondamento rispondere solo confrontando scarti medî relativi, non avendo senso un confronto tra un risultato ottenuto su misure piccolissime e quello ottenuto su misure grandissime, o tra risultati espressi in centimetri e in grammi o espressi in lire riguardanti rispettivamente chilogrammi di seta e chilogrammi di grano, o espressi in lire e in gradi di capacità intellettuale.

Si avverta, che essendo:

$$\frac{\sigma}{M} = \sqrt{\frac{\left(\frac{x_1 - M}{M}\right)^2 y_1 + \ldots\ldots + \left(\frac{x_n - M}{M}\right)^2 y_n}{y_1 + \cdots + y_n}} , \tag{30}$$

quel ragguaglio equivale ad assumere sotto segno di radice, in luogo della media quadratica degli scarti semplici, la media quadratica degli *scarti relativi*, che si ottengono mettendo in rapporto ogni scarto alla media aritmetica.

Per lo scarto medio interquartile (27) e lo scarto numerico medio (28), naturalmente converrà dividere i risultati per il valore centrale; mentre per la differenza media (29 *bis*) — in coerenza al criterio di non implicare alcun indice di posizione dei valori — si dovrà calcolare la media aritmetica delle differenze relative:

$$\frac{|x_i - x_k|}{x_i}.$$

Per renderne spedito il calcolo, abbiamo ricavato la seguente formula semplificatrice:

$$\frac{\Delta}{x} = \frac{\sum_{i=1}^{n-1} A_i \frac{y_{i-1}}{x_{i-1}} - \sum_{i=1}^{n-1} B_i \frac{y_{i+1}}{x_{i+1}}}{\sum_{i=1}^{n} y_i \left(\sum_{i=1}^{n} y_i - 1 \right)} \tag{31}$$

dove i simboli hanno lo stesso significato che nella (29 *ter*) [1].

Altri risultati si ottengono dividendo gl'indici di dispersione per il valore massimo che essi assumerebbero, se un solo elemento raccogliesse la somma dei valori osservati. In tal caso si parla di « rapporti di concentrazione », i quali ovviamente avranno significato solo quando si tratti di caratteri trasferibili. Ad es. per una distribuzione di redditieri quel rapporto direbbe in qual proporzione la dispersione di essa stia a quella che si sarebbe ottenuta, se un redditiere avesse accentrato il reddito totale della collettività; per una distribuzione di centri secondo il numero degli abitanti direbbe in qual proporzione la dispersione della popolosità di essi stia a quella che si sarebbe avuta, se un centro avesse raccolto la popolazione totale; per una distribuzione di famiglie secondo il numero dei figli minorenni quella proporzione sarebbe calcolata rispetto al caso che una famiglia avesse potuto adottare i figli di tutte le altre, e così via.

I valori massimi degli indici di dispersione si ottengono facilmente, in base alla formula ricavata dall'ipotesi che ad una osservazione corrisponda il valore totale $M \sum_{i=1}^{n} y_i$ ed alle altre $\sum_{i=1}^{n} y_1 - 1$ il valor 0.

Per lo scarto quadratico medio, ad es., si ottiene:

$$\sqrt{\frac{(M - 0)^2 \left(\sum_{i=1}^{n} y_i - 1 \right) + \left(M \sum_{i=1}^{n} y_i - M \right)^2}{\sum_{i=1}^{n} y_i}} = \sqrt{\sum_{i=1}^{n} y_i - 1}\; M \;; \tag{32}$$

---

[1] F. Vinci, *Sui coefficienti di variabilità*, in « Metron », 1920. L'inopportunità di dividere Δ per la media aritmetica dei valori, al fine di ottenere un indice di dispersione relativa, non sta solo in ragioni di coerenza, ma anche nel fatto che — come vedremo fra poco — con tal ragguaglio si cadrebbe in un « indice di concentrazione ».

per lo scarto numerico medio dalla media aritmetica:

$$\frac{(M-0)\left(\sum_{i=1}^{n} y_i - 1\right) + M \sum_{i=1}^{n} y_i - M}{\sum_{i=1}^{n} y_i} = 2\,\frac{\sum_{i=1}^{n} y_i - 1}{\sum_{i=1}^{n} y_i}\, M\,; \tag{33}$$

e per la differenza media — tenendo dapprima presente la (29), data la sua maggiore semplicità —:

$$\frac{2\left(\sum_{i=1}^{n} y_i - 1\right) M \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} y_i \left(\sum_{i=1}^{n} y_i - 1\right)} = 2\,M \tag{34}$$

Il più usato rapporto di concentrazione è quello desunto dalla differenza media, ossia dal rapporto tra la (29 *bis*) o la (29 *ter*) e la (34), perchè si presta ad un'interessante interpretazione.

Scriviamolo, per semplicità, nella forma:

$$R = \frac{\sum_{i=1}^{N} (A_i - B_i)}{M\,N\,(N-1)} \tag{35}$$

dove $N = \sum_{i=1}^{n} y_i$ è il numero, anche ripetuto, delle osservazioni ed A_i e B_i sono adesso le somme degli x_i, che occupano i posti non inferiori e, rispettivamente, non superiori ad i, ossia:

$$\begin{array}{ll}
A_1 = x_N + x_{N-1} + \ldots\ldots + x_2 + x_1 & B_1 = x_1 \\
A_2 = x_N + x_{N-1} + \ldots\ldots + x_2 & B_2 = x_1 + x_2 \\
\cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots \\
A_N = x_N & B_N = x_1 + x_2 + \ldots\ldots + x_{N-1} + x_N
\end{array}$$

È facile constatare che:

$$\sum_{i=1}^{N} A_i + \sum_{i=1}^{N} B_i = (N+1) \sum_{i=1}^{N} x_i$$

$$\sum_{i=1}^{N} (A_i - B_i) = (N+1) \sum_{i=1}^{N} x_i - 2 \sum_{i=1}^{N} B_i = (N-1) \sum_{i=1}^{N} x_i - 2 \sum_{i=1}^{N-1} B_i$$

e che, ponendo $q_i = B_i/B_N$, $p_i = i/N$, e ricordando la formula:

$$\sum_{i=1}^{N-1} i = \frac{N(N-1)}{2}$$

possiamo scrivere la (35) nella forma:

$$R = \frac{\sum_{i=1}^{N-1} (p_i - q_i)}{\sum_{i=1}^{N-1} p_i} \qquad (35\ bis)$$

Nel caso dei redditi di una collettività, il numeratore della (35 *bis*) darebbe la somma delle differenze tra frazioni successive p_i del numero dei redditieri con reddito non superiore a dati limiti e le frazioni corrispondenti q_i dell'ammontare dei loro redditi: se coteste frazioni corrispondenti fossero sempre uguali fra loro, è chiaro che ogni redditiere avrebbe il medesimo reddito e sarebbe $R = 0$ (concentrazione nulla); mentre, se coteste frazioni di reddito fossero nulle in $N - 1$ redditieri, perchè uno solo di essi raccogliesse i redditi di tutti, i q_i della (35 *bis*) sarebbero tutti nulli e risulterebbe $R = 1$ (concentrazione massima).

È quasi superfluo soggiungere che, se si disponesse di una distribuzione incompleta — ad es. di redditieri con reddito superiore al minimo esente da tassazione — i risultati varrebbero solo entro i limiti delle osservazioni e risentirebbero l'influenza di tali limiti.

Della (35 *bis*) può darsi un'elegante rappresentazione grafica (1), che esemplifichiamo con la distribuzione della tavola XL, che dà la popolazione presente nei nostri *centri* al 1° dicembre 1921, data del penultimo censimento.

Avvertiamo che il numero degli abitanti, riportato nella detta tavola, non è stato ottenuto moltiplicando il valore centrale delle classi per il numero dei centri, ma è stato fornito esattamente dal nostro Istituto Centrale di Statistica; e che inoltre i p_i ed i q_i sono stati calcolati per $i = 50$, 100, 250, ecc. (2).

Ciò posto, rappresentiamo su due assi cartesiani le coppie p_i, q_i usando le medesime scale: la spezzata risultante (tratteggiata nella fig. 38) risulta compresa in un quadrato di lato 1 e, congiungendo convenientemente i vertici di essa, si ottiene con approssimazione la « curva di concentrazione » corrispondente a intervalli piccolissimi: la diagonale tracciata nella figura è la cosidetta «retta di equidistribuzione», che si avrebbe nell'ipotesi che fosse $p_i = q_i$, ossia che la concentrazione fosse nulla; l'area compresa tra la retta di equidistribuzione e la curva di concen-

(1) M. O. Lorenz, *Methods of measuring the concentration of wealth*, in « Quarterly pubblications of the American Statistical Association », giugno 1905. – W. M. Persons, *The variability in the distribution of wealth and income*, in « The quarterly journal of economics », 1909 (pag. 417). – C. Gini, *Sulla misura della concentrazione e della variabilità dei caratteri*, in « Atti del R. Istituto Veneto », 1913-14.

(2) Cfr. R. D'Addario, *L'agglomeramento della popolazione nei compartimenti italiani*, in « Annali di Statistica », Serie VI, Vol. XVI.

trazione sarebbe misurata dal numeratore della (35 *bis*), ove gli i variassero con continuità, e nella stessa ipotesi l'area del triangolo ABC sarebbe misurata dal denominatore della detta formula.

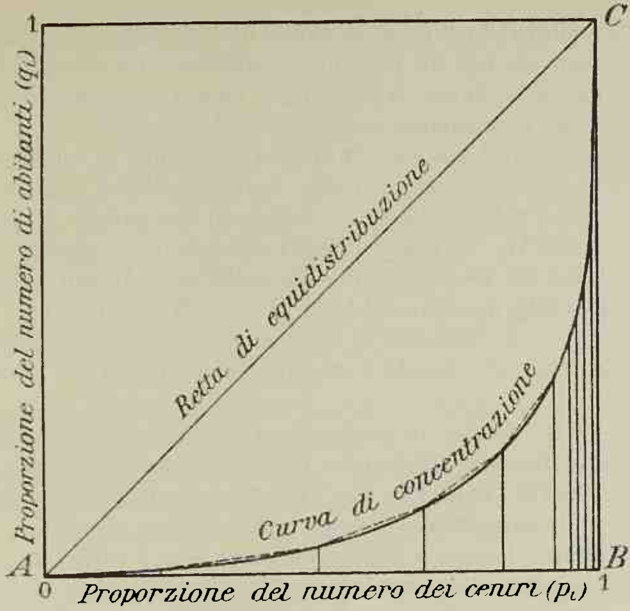


Fig. 38

Questo metodo grafico — oggi generalmente preferito — permette un calcolo di R più corretto di quello numerico, perchè si fonda su una curva di concentrazione più vicina a quella reale. È facile infatti osservare che, sostituendo la spezzata alla curva di concentrazione, R risulta approssimato per difetto.

Col metodo grafico, la valutazione dell'area compresa tra la retta di

TAVOLA XL.

Centri con abitanti	Centri		Abitanti	
	Numero assoluto	p_i	Numero assoluto	q_i
– 50	1.939	0,0684	65.340	0,0023
51 - 100	4.018	0,2102	302.623	0,0128
101 - 250	8.060	0,4946	1.317.625	0,0587
251 - 500	5.338	0,6830	1.914.777	0,1254
501 - 1.000	4.147	0,8293	2.930.900	0,2274
1.001 - 2.000	2.581	0,9204	3.615.217	0,3533
2.001 - 3.000	867	0,9510	2.107.422	0,4267
3.001 - 4.000	424	0,9660	1.470.393	0,4779
4.001 - 5.000	276	0,9757	1.235.724	0,5209
5.001 - 10.000	377	0,9890	2.678.078	0,6141
10.001 - 15.000	121	0,9933	1.441.599	0,6643
15.001 - 20.000	62	0,9955	1.092.120	0 7023
20.001 - 50.000	93	0,9980	2.677.312	0,7956
50.001 - 100.000	21	0,9995	1.256.609	0,8394
100.001 -	14	1,0000	4.613.156	1,0000
Totali	28.338		28.718.895	

equidistribuzione e la curva di concentrazione può eseguirsi per via elementare con un comune planimetro (1) o ritagliando su un cartone omogeneo la figura triangolare e quella risultante dalla curva di concentrazione e pesandole accuratamente.

Per tal via è stato appunto ottenuto su quella distribuzione di centri: $R = 68{,}08$, risultato che dimostra l'alta concentrazione della popolazione nei nostri centri (abbiamo trascurato la popolazione sparsa), soprattutto a cagione dell'inurbamento. Cotesto fenomeno è caratteristico dei paesi a civiltà capitalistica e da noi, peraltro, non raggiunge la più alta misura: dal 1871 al 1921 R è variato solo da 61,47 a 68,08, sebbene naturalmente in misura maggiore nei compartimenti del Nord, dove si sono sviluppati i grandi centri industriali e commerciali del Regno.

15. Indici di asimmetria. – Abbiamo avuto occasione di esaminare distribuzioni statistiche di diversa forma, alcune delle quali (tavole I, VI, VII, VIII, IX, XXXII) formate da due rami discendenti più o meno simmetricamente rispetto all'ordinata massima; e in fine al n. 9 abbiamo illustrato alcune relazioni, che i valori prevalente, centrale e medio aritmetico presentano in cotesti tipi di distribuzione. Ora diciamo che per essi si sogliono pure calcolare indici di asimmetria. Considerando, ad es., che le distanze tra i due quartili e il valore centrale sono uguali in caso di simmetria di quelle distribuzioni e che, a parità di grado di asimmetria, tali distanze sono tanto più divergenti, quanto maggiore è la dispersione dei dati, tenendo presente la (27) è stata costruita la formula:

$$S = \frac{\left(x_{3/4} - x_c\right) - \left(x_c - x_{1/4}\right)}{\frac{1}{2}\left(x_{3/4} - x_{1/4}\right)} = \frac{x_{3/4} + x_{1/4} - 2\,x_c}{\frac{1}{2}\left(x_{3/4} - x_{1/4}\right)}, \qquad (36)$$

la quale assume valore positivo, nullo o negativo. secondo che

$$x_{3/4} - x_c \gtreqless x_c - x_{1/4}\,.$$

Nell'esempio della figura 34 abbiamo:

$$S = \frac{32{,}70 + 23{,}90 - 2 \times 27{,}93}{\frac{1}{2}\left(32{,}70 - 23{,}90\right)} = +\,0{,}1682\,.$$

ossia la distribuzione di quelle nascite secondo l'età delle madri avrebbe un'asimmetria positiva del 16,82 per cento della sua dispersione.

(1) Sull'uso di esso, cfr. Gino Cassinis, op. cit.

Nella differenza, messa in rapporto allo scarto quadratico medio, tra la media aritmetica e il valore prevalente:

$$S' = \frac{M - x_p}{|\sigma|} \tag{37}$$

si è pure riconosciuto un buon indice di asimmetria, specie per quelle distribuzioni, dove il valore prevalente si possa ritenere tipico (cfr. n. 9).

Se questo fosse il caso per la tavola VIII, si avrebbe:

$$\frac{2.998 - 3.003}{33} = -0,15$$

cioè un'asimmetria negativa (prevalenza delle maggiori altezze barometriche) uguale a circa il 15 per cento della dispersione.

Questa seconda formula, però, non è di facile calcolo, in quanto presuppone la determinazione del valor prevalente, che — come si è visto — può calcolarsi per via elementare solo con larga approssimazione e, ricavato dalla formula (18) data in fine al n. 9, conduce spesso a notevoli errori.

Avvertiamo, infine, che il cosidetto scarto cubico medio, desunto dalla formula:

$$\sqrt[3]{\mu_{M,3}} = \sqrt[3]{\frac{\sum\limits_{i=1}^{n} (x_i - M)^3 y_i}{\sum\limits_{i=1}^{n} y_i}} \tag{38}$$

essendo calcolato rispetto alla media aritmetica, dà risultati equivoci, tranne casi particolari relativi a date famiglie di curve.

Ad es. nella distribuzione:

$$x_i = 1, \quad 2, \quad 3, \quad 4, \quad 5$$

$$y_i = 9, \quad 34, \quad 34, \quad 34, \quad 17$$

la media aritmetica è $3 + \frac{1}{8}$ e la (38) risulta nulla, nonostante l'asimmetria della distribuzione.

Poichè anche la (36) e la (37) possono condurre a simili aberrazioni, è stata proposta, come indice generale di asimmetria, la formula:

$$S'' = \frac{\Sigma \, | (x_i - M) + (x_{N-i+1} - M) |}{2 \, \Sigma \, | x_i - M |}$$

da calcolare sui valori, anche ripetuti, di x_i disposti in ordine di grandezza e dove i varia da 1 ad N; onde la sommatoria si estende da

$$(x_1 - M) + (x_N - M) \quad \text{sino a} \quad (x_N - M) + (x_1 - M),$$

per $N = \Sigma y_i$.

La formula precedente si annulla in caso di perfetta simmetria ed è sempre minore di 1.

Prendendo i quadrati, invece dei valori assoluti, si ottiene l'altra formula:

$$S''' = \frac{\Sigma [(x_i - M) + (x_{N-i+1} - M)]^2}{2\,\Sigma (x_i - M)^2} = 1 + \frac{\Sigma (x_i - M)(x_{N-i+1} - M)}{\Sigma (x_i - M)^2}$$

la quale gode delle medesime proprietà (1).

Sarà da preferirsi l'una o l'altra, secondo che si assuma per la misura della dispersione un indice semplice o un indice quadratico.

Ad es. nella distribuzione:

$$x_i = 0,\ 1,\ 2,\ 3$$

$$y_i = 8,\ 10,\ 6,\ 1$$

la (36) e la (37) dànno risultati nulli, malgrado l'evidente asimmetria, perchè il primo e il terzo quartile risultano uguali rispettivamente a 0 e a 2, ossia equidistanti dal valor centrale 1, e la media aritmetica risulta uguale a 1, che è pure il valore prevalente.

Invece, considerando gli scarti:

$$x_i - M = -1,\ 0,\ 1,\ 2$$

e i rispettivi y_i, si ha:

$$S'' = \frac{1+1}{16} = \frac{1}{8} = 0{,}125$$

$$S''' = 1 - \frac{2(2+6)}{18} = \frac{1}{9} = 0{,}111\,.$$

16. Alcune relazioni tra costanti caratteristiche. – Si è già osservato che alcune costanti discendono come casi particolari dall'espressione (10), che abbiamo chiamato momento diviso di grado s e di origine m.

---

(1) C. E. Bonferroni, *Elementi di statistica generale*. Torino, Litografia Gili, 1933-XI.

Ponendo adesso, in generale,

$$\sqrt[s]{\mu_{m,s}} = \sqrt[s]{\frac{\sum_{i=1}^{n} (x_i - m)^s \, y_i}{\sum_{i=1}^{n} y_i}} \tag{39}$$

può verificarsi che, per $s = -1$, la precedente espressione dà la formula generale della media armonica, e si dimostra altresì che il limite di essa per $s = 0$, dà la formula generale della media geometrica [1].

D'altra parte, per $s = -2, -3, \ldots$, da essa discendono altri infiniti tipi di medie; mentre per $s = 1, 2, 3, \ldots$ discendono immediatamente le medie aritmetica, quadratica, cubica, ... degli $x_i - m$.

Si dimostra che, se gli $x_i - m$ sono tutti positivi, al tendere di s a $-\infty$ ed a $+\infty$, la (39) tende rispettivamente a $x_1 - m$ e ad $x_n - m$ e cresce continuamente al crescere di s. Ciò significa che, ad es., *la media armonica di quantità positive* (M_a) *è sempre minore della geometrica* (M_g); *questa è minore dell'aritmetica* (M) *e così via.*

Queste ultime relazioni possono dimostrarsi rapidamente in modo elementare, in base ad un teorema di algebra secondo il quale, se la somma di più quantità positive è costante, il prodotto di esse è massimo quando siano tutte uguali fra loro.

Da esso, infatti, si ricava che la frazione:

$$\frac{\frac{1}{N}(x_1 + x_2 + \ldots + x_N)}{\sqrt[N]{x_1 \times x_2 \times \ldots \times x_N}}$$

diventa minima quando $x_1 = x_2 = \ldots = x_N$.

Poichè in questo caso il valore della frazione è 1, segue che, se gli x_i sono disuguali, quella frazione è sempre maggiore di 1 e $M > M_g$.

In base allo stesso teorema, se gli x_i sono disuguali, è:

$$\frac{\frac{1}{N}\left(\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_N}\right)}{\sqrt[N]{\frac{1}{x_1 \times x_2 \times \ldots \times x_N}}} = \frac{\frac{1}{M_a}}{\frac{1}{M_g}} > 1\,,$$

e quindi $M_g > M_a$.

---

(1) O. Dunkel, *Generalized geometrical means and algebraic equations*, in « Annals of mathematics », ottobre 1909. – U. Ricci, *Confronti tra medie*, in « Giornale degli economisti », luglio 1915. — C. E. Bonferroni, *Elementi di statistica generale*, op. cit., dove per via elementare si pongono altre formule generali e si discute anche della cosidetta media esponenziale, che s'impiega in matematica finanziaria per il calcolo della scadenza media, dell'età media nelle assicurazioni, ecc. Su questo argomento vedansi due note dello stesso autore in « Annuario del R. Istituto Superiore Commerciale di Bari », 1923-24 e 1924-25.

Sono state pure messe in luce alcune relazioni approssimate tra medie [1], dove interviene, tra l'altro, la cosidetta media antiarmonica:

$$M_{ant.} = \frac{x_1^2 + x_2^2 + \dots + x^2_N}{x_1 + x_2 + \dots + x_N}, \tag{40}$$

la quale invero ha un impiego rarissimo nelle analisi statistiche; e si son date interessanti relazioni tra indici di dispersione.

Ad es., se — come da qualche studioso si è fatto — si assume come indice di dispersione, anzichè la (29), la differenza media quadratica:

$$^2\Delta = \sqrt{\frac{\sum_{i=1}^{N} \sum_{k=1}^{N-1} \left(x_i - x_k\right)^2}{N(N-1)}} \tag{41}$$

o la differenza media quadratica con ripetizione, dove cioè si considerano anche le coppie formate da quantità uguali:

$$^2\Delta_R = \sqrt{\frac{\sum_{i=1}^{N} \sum_{k=1}^{N-1} \left(x_i - x_k\right)^2}{N^2}}, \tag{42}$$

è facile mostrare la relazione in cui queste espressioni stanno allo scarto quadratico medio.

Dai n. 5 e 10 sappiamo anzitutto che:

$$\mu_{m,2} = \mu_{M,2} + \mu^2_{m,1} = \mu_{M,2} + (M - m)^2 .$$

Se diamo ad m tutti gli N valori osservati, la media degli N valori di $\mu_{m,2}$, così ottenuti, sarà uguale alla media dei quadrati delle N^2 differenze tra ogni quantità e tutte le N quantità della distribuzione:

$$^2\Delta^2_R = \frac{1}{N} \sum_{i=1}^{N} \mu_{m_i,2} = \mu_{M,2} + \frac{1}{N} \sum_{i=1}^{N} (M - m_i)^2 = 2\,\mu_{M,2},$$

da cui

$$^2\Delta_R = \sqrt{2\,\mu_{M,2}} = \sqrt{2}\,\sigma \tag{43}$$

$$^2\Delta = \sqrt{\frac{2N}{N-1}}\,\sigma . \tag{44}$$

17. Conclusioni. – Le costanti caratteristiche, di cui ci siamo occupati in questo capitolo, sono una piccola parte delle innumerevoli costanti, che sono state o si potrebbero ricavare da un gruppo di quantità.

---

[1] Cfr. U. Ricci, opera ultima cit.

Cauchy ha addirittura affermato: « on appelle moyenne entre plusieurs quantités données une nouvelle quantité comprise entre la plus petite et la plus grande de celles que l'on considère.... D'après cette définition, il est clair qu'il existe une infinité de moyennes entre plusieurs quantités inégales et que la moyenne entre plusieurs quantités égales se confond avec chacune d'elles » (1).

Ma occorre fermare in mente che, se il numero delle medie tra più quantità e, in generale, delle costanti caratteristiche può considerarsi infinito, ben poche costanti si sono rivelate feconde nella metodologia statistica; e che tra queste primeggiano la media aritmetica e lo scarto quadratico medio, sia per il carattere generale delle esigenze a cui soddisfano, sia perchè meglio di altre si prestano ad analisi matematiche e alla costruzione dei metodi statistici.

Malgrado che — al pari di ogni altra costante — discendano da concetti arbitrarî, l'impiego di esse è stato sinora il mezzo più fecondo (n. 6 del primo capitolo dell'Introduzione) per quell'accertamento approssimato delle relazioni tra i fenomeni, che è l'oggetto altamente scientifico della metodologia statistica.

---

(1) A. L. Cauchy, *Cours d' analyse de l'École royale polytechnique*, Paris, Imprimerie Royale, 1821 (1.ère partie) pag. 14.

Capitolo IV.

LA SCELTA E L'ADATTAMENTO DELLE FUNZIONI

1. Formulazione del problema. – Chiamando funzione della variabile x una variabile y, che assuma uno o più valori determinati in corrispondenza ad ogni valore di x compreso in un dato intervallo, quel che abbiamo esposto nei precedenti capitoli ci autorizza ad estendere il concetto di funzione alle distribuzioni ed alle serie statistiche a due variabili. E precisamente possiamo in esse considerare i numeri della seconda colonna come valori di una funzione di una variabile, i cui valori corrispondenti siano quelli della prima colonna.

È quasi superfluo soggiungere che, se nella prima colonna figurano attributi graduabili, sostituendo ad essi una graduatoria possiamo considerare i gruppi di osservazioni come valori di una funzione del numero d'ordine della graduatoria, con riserve analoghe a quelle del n. 3 del precedente capitolo.

Fondandoci sulla definizione della funzione di più variabili, analoghe considerazioni possiamo fare tanto per le serie statistiche a più di due variabili, quanto per le tavole a parecchie entrate.

Tale estensione del concetto di funzione ha fatto compiere grandi progressi alle scienze naturali ed anche a quelle sociali, perchè ha reso in esse possibile l'impiego di metodi, che un tempo si ritenevano privilegio di pochi rami del sapere: soprattutto l'impiego delle due categorie di metodi, di cui — come si è detto nell'Introduzione — consta la metodologia statistica.

In questa, infatti, s'incontrano i problemi di rappresentare, con una funzione analiticamente determinata, gruppi di dati osservati; di adattare ad essi la funzione analitica scelta; infine di verificare la bontà dell'adattamento.

Sono tre problemi connessi, per la risoluzione dei quali è indifferente che si tratti di distribuzioni o serie statistiche o di tavole a parecchie entrate, e che i fenomeni considerati siano, in tutto o in parte, misure o attributi graduati: basta considerare successioni corrispondenti di numeri.

Il primo di tali problemi si risolve in base a criterî desunti dalla rappresentazione grafica dei dati, alle nozioni che si posseggano sulle proprietà delle funzioni e sui fenomeni in esame, ad analogie con altri fenomeni dei quali già si conoscano le funzioni analitiche, ed a quell'intuito scientifico che non si può ridurre a precetti; il secondo problema è noto comunemente col nome di problema dell'*interpolazione* o della *perequazione* — che sono invero due scopi particolari, che l'adattamento delle funzioni permette di raggiungere — e consiste nel calcolare opportunamente i valori dei coefficienti della funzione scelta; il terzo problema si propone di giudicare se i dati desunti dalla funzione si possano — per gli scopi che si perseguono — sostituire a quelli osservati, ossia se la funzione scelta sia idonea a rappresentare questi ultimi: a tal fine basta spesso la semplice ispezione dei risultati, ma talvolta occorre un esame sottile e approfondito, fondato sullo studio degli schemi teorici (Parte II).

Si noti, inoltre, che tali problemi non presenterebbero speciali difficoltà, se ci proponessimo di rappresentare rigorosamente i dati osservati; anzi, dovendosi senz'altro respingere quegli adattamenti che non facessero coincidere i dati osservati con quelli corrispondenti della funzione, il terzo problema non esisterebbe. In tal caso la funzione dovrebbe contenere un numero di coefficienti (parametri) uguale a quello delle condizioni (indipendenti e compatibili) poste dai dati osservati e i coefficienti di essa si potrebbero in generale calcolare coi metodi o cogli artifici suggeriti dall'algebra.

Quei problemi, invece, presentano speciali difficoltà, quando — com'è generalmente il caso per i fenomeni di cui ci occupiamo — i dati osservati e precisamente quelli assunti come valori della variabile dipendente siano affetti da perturbazioni, che si ritenga opportuno di trascurare: la convenienza di eliminare gli errori casuali e sistematici, e gli effetti di altre circostanze che non interessano al fine che si persegue, sono motivi che generalmente consigliano a non rispettare fedelmente i dati osservati.

Ben più difficile è in questo caso la scelta della funzione analitica; il problema dell'adattamento — affinchè non riesca indeterminato — viene formulato imponendo una data condizione di approssimazione dei dati osservati a quelli desunti dalla funzione, anche in relazione alle particolari esigenze del materiale disponibile; e sorge appunto il terzo problema di verificare la bontà dell'adattamento.

2. La scelta della funzione. – Nella scelta della funzione analitica talvolta ci si propone solo di descrivere i dati o una parte di essi più efficacemente che coi metodi esposti nei precedenti capitoli: si dice allora che si va in cerca di una *funzione empirica*, valevole per tutti

i dati o per una parte di essi. Spesso però si cercano dei principi, dai quali possa razionalmente dedursi un tipo di funzione, che risponda ai dati osservati: si dice, allora, che si cerca la *funzione teorica (o legge analitica)* del fenomeno o dei fenomeni considerati.

Generalmente, però, si risale alla funzione teorica dopo la scoperta di una buona funzione empirica e spesso si finisce col dedurre razionalmente e fare assurgere a teorica una funzione dapprima ritenuta empirica.

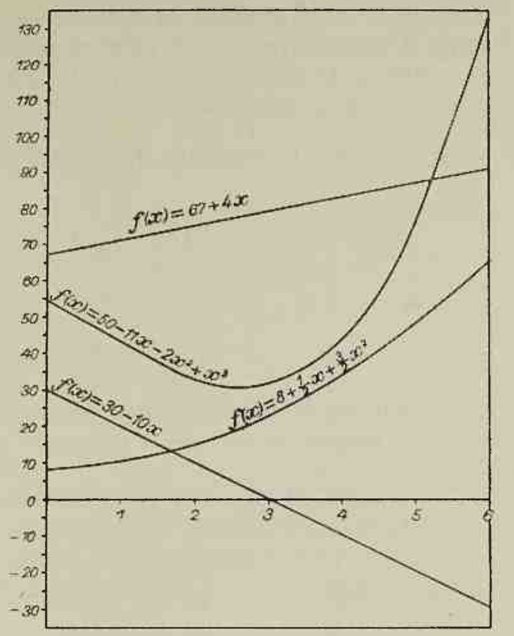


Fig. 39

Ma anche limitatamente alle funzioni empiriche e al fine meramente descrittivo che esse si propongono, non è facile una buona scelta della forma di tali funzioni. Per le distribuzioni e le serie statistiche a due variabili si fa largo impiego della funzione razionale intera:

$$z = f(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_r x^r \qquad (45)$$

la quale, considerando solo i primi due termini, dà luogo ad una funzione di primo grado, a cui corrisponde geometricamente una retta e che pertanto si chiama funzione lineare; considerando i primi tre, quattro,... termini, dà luogo rispettivamente a funzioni di secondo, terzo... grado, alle quali corrispondono parabole di secondo, terzo,... ordine [1].

Nella figura 39 sono dati alcuni esempi di coteste curve, costruite per punti e per valori di x compresi tra 0 e 6.

Chiamando *differenza prima, seconda, ...* di $f(x)$ le espressioni:

$$\left.\begin{array}{l} \Delta f(x) = f(x+h) - f(x) \\ \Delta^2 f(x) = \Delta f(x+h) - \Delta f(x) \\ \dots\dots\dots\dots\dots \\ \Delta^r f(x) = \Delta^{r-1} f(x+h) - \Delta^{r-1} f(x) \\ \dots\dots\dots\dots\dots \end{array}\right\} \qquad (46)$$

[1] Il largo impiego della (45) a scopi descrittivi deriva dal fatto che le funzioni teoriche ignote sono per lo più continue e che le funzioni continue — per un noto teorema — possono essere rappresentate dalla (45) con un' approssimazione grande quanto si vuole.

dove h è una costante arbitraria, si verifica facilmente che le differenze 1^a, 2^a, ... $(r-1)^{ma}$ sono rispettivamente polinomi di $(r-1)^{mo}$, $(r-2)^{mo}$,... 1^o grado in x, che la differenza r^{ma} è costante e le differenze $(r+1)^{ma}$, $(r+2)^{ma}$ ... sono nulle.

Ad es., per la funzione: $f(x) = 8 + \frac{1}{2}x + \frac{3}{2}x^2$ si ha:

$$\Delta f(x) = \frac{1}{2}h + \frac{3}{2}h^2 + 3hx$$

$$\Delta^2 f(x) = 3h^2$$

$$\Delta^3 f(x) = 0$$

e, facendo ad es. assumere successivamente ad x i valori della serie naturale dei numeri, si ottiene il prospetto seguente:

x_i	$f(x_i)$	$\Delta f(x_i)$	$\Delta^2 f(x_i)$	$\Delta^3 f(x_i)$
0	8	2	3	0
1	10	5	3	0
2	15	8	3	0
3	23	11	3	0
4	34	14	3	.
5	48	17	.	.
6	65	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

Sembra facile dedurre da tali proprietà un criterio di scelta delle funzioni razionali intere, che risponda ai nostri scopi: calcolando sulla variabile dipendente di una distribuzione o serie statistica a due variabili, in corrispondenza a valori equidistanti di x, le differenze dei successivi ordini, se s'incontra una colonna di differenze che (prescindendo dalle perturbazioni dei dati) possa ritenersi costante, l'ordine di tali differenze darà il grado della funzione razionale intera, idonea a rappresentare la distribuzione o la serie statistica. Ma sorgono difficoltà nel valutare l'influenza delle circostanze perturbatrici sulle successive differenze, per il fatto che le perturbazioni possono essere di svariata natura.

Ecco, nella tavola XLI, il prospetto delle differenze ricavate dai quozienti di natalità del nostro paese negli ultimi anni. Tali quozienti divergono leggermente da quelli della tav. XIX *bis*, essendo stati calcolati sulla popolazione del Regno corretta in base al censimento del 1931.

TAVOLA XLI.

Anni x_i	Nati vivi per 1000 ab. n_i	$\Delta(n_i)$	$\Delta^2(n_i)$	$\Delta^3(n_i)$	$\Delta^4(n_i)$	$\Delta^5(n_i)$	$\Delta^6(n_i)$	$\Delta^7(n_i)$	$\Delta^8(n_i)$
1924	29,0	— 0,7	+ 0,1	+ 0,3	— 1,3	+ 2,6	— 1,7	— 9,0	+45,8
1925	28,3	— 0,6	+ 0,4	— 1,0	+ 1,3	+ 0,9	—10,7	+36,8	
1926	27,7	— 0,2	— 0,6	+ 0,3	+ 2,2	— 9,8	+26,1		
1927	27,5	— 0,8	— 0,3	+ 2,5	— 7,6	+16,3			
1928	26,7	— 1,1	+ 2,2	— 5,1	+ 8,7				
1929	25,6	+ 1,1	— 2,9	+ 3,6					
1930	26,7	— 1,8	+ 0,7						
1931	24,9	— 1,1							
1932	23,8								

Parrebbe che una funzione razionale intera di grado inferiore all'8° non potesse ben rappresentare empiricamente cotesta serie storica e che non fosse affatto idonea la funzione di primo grado, dato che le $\Delta(n_i)$ variano da un minimo di — 1,8 a un massimo di + 1,1. Eppure, dando

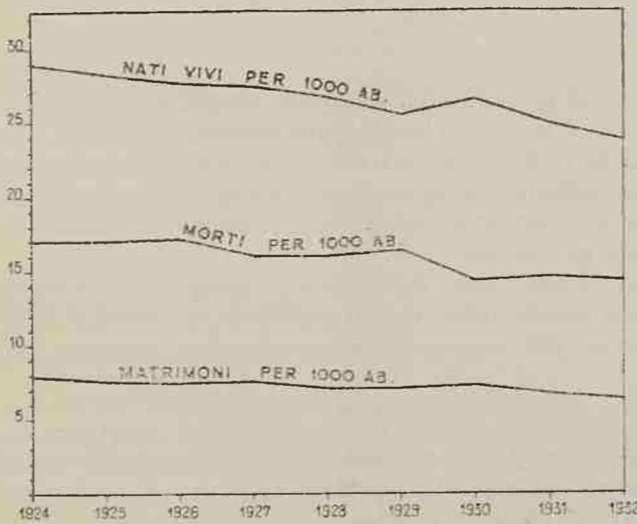


Fig. 40

uno sguardo alla figura 40 e tenendo presenti le perturbazioni di varia natura che la natalità del nostro paese ha risentito nell'ultimo decennio e specie nel 1929-30 per l'influenza di fattori meteorologici, se prescindiamo da tali perturbazioni e vogliamo rappresentare la caduta generale della natalità, non è *a priori* da respingere la funzione lineare.

Con la stessa funzione potremmo rappresentare la mortalità e la nuzialità, pure riportate nella detta figura, le quali con le loro diverse perturbazioni attraverso una linea pure discendente potrebbero confermare la bontà della nostra scelta, rivelando l'azione di un insieme di circostanze di tal natura, da costringere i dati del movimento naturale della nostra popolazione a una discesa marcatamente rettilinea.

Se si considerano altri tipi di funzioni, la scelta non è meno ardua, perchè non si riesce mai a valutare obiettivamente l'influenza delle perturbazioni e si è costretti a usare le parole « approssimativamente », « marcatamente », ecc., di significato elastico e incerto.

Limitandoci alle funzioni di uso più frequente nelle scienze sociali, rileviamo che la funzione:

$$z = a_0 \, x^{a_1} \tag{47}$$

per $a_1 = 2$ dà luogo ad una parabola, il cui asse coincide con quello dei valori di z, e per $a_1 < 0$ dà luogo ad un'iperbole; e che, passando ai logaritmi, essa diventa:

$$\log z = \log a_0 + a_1 \log x \, ,$$

ossia, ponendo $X = \log x$, $Z = \log z$, $A = \log a_0$, si trasforma nella funzione lineare:

$$Z = A + a_1 X \, .$$

Pertanto, se in una rappresentazione cartesiana i logaritmi delle coppie osservate x_i, y_i dessero luogo *approssimativamente* a una retta, sarebbe giustificata la scelta della (47) come funzione rappresentatrice dei dati. La verifica potrebbe pure eseguirsi, tenendo presente che nella (47) a valori di x in progressione geometrica corrispondono valori di z pure in progressione geometrica.

Se *sulla doppia scala logaritmica* la spezzata s'incurvasse in corrispondenza a piccoli valori di x e assumesse, al crescere di x, un andamento sempre più marcatamente rettilineo, si potrebbe ricorrere alla funzione:

$$z = a_0 \, x^{a_1} + a_2 \tag{48}$$

per $a_1 > 0$, nel qual caso il termine $a_0 \, x^{a_1}$ va assumendo al crescere di x un'importanza prevalente sul termine a_2 (¹).

Così pure la funzione esponenziale:

$$z = a_0 \, a_1^{x} \tag{49}$$

(¹) Su questo argomento un utile libro elementare è quello di M. Fréchet et R. Romann, *Représentation des lois empiriques par des formules approchées*. Paris, Eyrolles, 1930. Cfr. pure T. R. Running, *Empirical formulas*. London, Chapman and Hall, 1917.

passando ai logaritmi, diventa:

$$\log z = \log a_0 + x \log a_1 ,$$

ossia assume la forma lineare:

$$Z = A + B x .$$

Se, quindi, rappresentando su un piano cartesiano le coppie osservate x_i, $\log y_i$ si ottenesse una spezzata sensibilmente rettilinea, sarebbe giustificata la scelta della (49). Ci si potrebbe anche fondare sulla proprietà della (49), che a valori di x in progressione aritmetica corrispondono valori di z in progressione geometrica.

Se i risultati non fossero soddisfacenti, si potrebbe verificare la funzione:

$$z = a_0 \, a_1^x + a_2 \qquad (50)$$

tenendo presente che, per valori equidistanti di x, le differenze prime di questa funzione hanno la forma:

$$\Delta z = a'_0 \, a_1^x$$

e che quindi le coppie osservate x_i, $\log \Delta y_i$ dovrebbero corrispondere *a un dipresso* ad una retta; ossia che a valori di x in progressione aritmetica dovrebbero corrispondere valori di Δy_i in progressione *approssimativamente* geometrica.

Si potrebbe anche aggiungere alla (49) un fattore correttivo $a_2^{x^2}$ e verificare la funzione:

$$z = a_0 \, a_1^x \, a_2^{x^2} . \qquad (51)$$

Ponendo $a_0 = A^{b_0}$; $a_1 = A^{b_1}$; $a_2 = A^{b_2}$, la (51) può scriversi nella forma generale:

$$z = A^{\,b_0 + b_1 x + b_2 x^2} \qquad (52)$$

che, passando ai logaritmi, diventa

$$\log z = (b_0 + b_1 x + b_2 x^2) \log A .$$

Ond'è che la (51) potrebbe bene adottarsi — per ciò che abbiamo detto a proposito delle funzioni razionali intere — se, per valori equidistanti di x, le $\Delta^2 \log y_i$ fossero approssimativamente costanti.

Se $b_2 \log \text{nat} A < 0$, alla (52) può darsi la forma:

$$z = K \, e^{-h^2 (x-M)^2} \qquad (53)$$

dove $e = 2{,}71828 .....$ è la base dei logaritmi naturali e si ponga:

$$K = A^{b_0 - \frac{b_1^2}{4 b_2}} \; ; \quad h^2 = - b_2 \log \text{nat} A \; ; \quad M = - \frac{b_1}{2 b_2} .$$

Si osserva subito che, per x uguale a $\pm \infty$, $z = 0$ e che z raggiunge un massimo per $x = M$ e decresce simmetricamente rispetto a tale massimo.

Si dimostra che, rappresentata graficamente in coordinate ortogonali, essa dà luogo ad una particolare curva a campana, che tra l'altro — come vedremo nella Parte II — tra le ascisse corrispondenti ai due punti laterali di flesso comprende circa il 68,27 per cento dell'area totale.

Se le coppie osservate x_i, y_i danno luogo a una curva in cui, attraverso le irregolarità d'andamento, possa riconoscersi cotesta forma — come accade appunto per molte distribuzioni statistiche — la scelta della (53) come formula empirica è legittima.

Supponiamo, infine, che le coppie osservate consistano in una serie storica, ad es. mensile, di n termini, che presenti fluttuazioni stagionali lungo un periodo X.

Ponendo:

$$\varphi = 2\pi \frac{x}{X},$$

dove $\pi = 3{,}14159$ ..... è il noto rapporto della circonferenza al diametro di un cerchio, ossia trasformando la variabile x in un multiplo di 2π (lunghezza della circonferenza di raggio 1), le n coppie osservate si potranno rappresentare mediante la somma trigonometrica di ordine r:

$$\begin{aligned} z = a_0 + a_1 \cos \varphi + a_2 \cos 2\varphi + \ldots\ldots + a_r \cos r\varphi \\ + b_1 \operatorname{sen} \varphi + b_2 \operatorname{sen} 2\varphi + \ldots\ldots + b_r \operatorname{sen} r\varphi \end{aligned} \qquad (54)$$

dove $2r + 1 < n$.

Nella (54) — di cui si fa uso larghissimo pei fenomeni periodici [1] e che è chiamata serie di Fourier — ci si arresterà a quell'ordine, che si reputerà conveniente ad una *soddisfacente approssimazione.*

Oltre che al grafico dei dati ed alla conoscenza delle proprietà di alcuni tipi fondamentali di funzioni. qualche sussidio potrà chiedersi alla natura dei fenomeni in esame ed a criterî di analogia. Ma su questo punto non saranno inopportuni alcuni esempî concreti.

Si è supposto dai demografi che una popolazione numerosa in un intervallo piccolissimo di tempo sia suscettibile di variazione piccolissima e, poichè tale variazione nell'intervallo successivo entra a sua volta nel processo di riproduzione e di eliminazione, si è pensato che non si potessero commettere sensibili errori, ritenendo altresì ch'essa varii con legge analoga a quella della capitalizzazione ad interesse composto continuo (positivo in caso di aumento, negativo in caso di diminuzione della popolazione).

---

(1) Per un noto teorema, ogni funzione periodica continua può essere rappresentata dalla (54) con un'approssimazione grande quanto si vuole.

Pertanto, almeno entro brevi periodi, nei quali sia lecito ammettere che il saggio di variazione si mantenga a un dipresso costante, una popolazione numerosa suole rappresentarsi con la funzione esponenziale:

$$P_t = P_0\, e^{it} \qquad (55)$$

I brevi periodi non sogliono superare il decennio, essendosi osservato — ad es. sulla rappresentazione grafica dei $\log P_t$, data dalla figura 8 per un gruppo di Stati nell'ultimo quarantennio prebellico — che da un decennio al successivo le inclinazioni dei successivi segmenti delle spezzate cominciano a divergere sensibilmente, anche in periodi normali.

Essendo:

$$\log P_t = \log P_0 + (i \log e)\, t\,,$$

tali inclinazioni sono appunto proporzionali ad i.

È così grande la fiducia, che i demografi ripongono nella (55) — la quale in matematica finanziaria si dimostra appunto essere la legge di capitalizzazione, ad interesse composto continuo i, di una somma iniziale P_0 — che, attribuendo a P_0 e a P_t l'ammontare (approssimativo) della popolazione di un territorio popoloso in due successivi istanti alquanto vicini tra loro, il valore risultante di i è da essi ritenuto come saggio medio di variazione della popolazione nell'intervallo compreso tra quei due istanti.

Ad es., ammesso che la popolazione italiana presente nei confini del Regno sia ammontata a 37.988.000 al 1° dicembre 1921 ed a 41.177.000 al 21 aprile 1931 (ponendo $\log e = 0{,}434294$ e $t = 9{,}388333$), il nostro Istituto Centrale di Statistica (¹) ha ottenuto:

$$i = \frac{\log 41.177.000 - \log 37.988.000}{9{,}388333 \times 0{,}434294} = 0.0085868$$

ossia quel saggio annuo di variazione del 0,86 per cento, che si legge negli atti ufficiali.

In molti fenomeni economici è stato da tempo notato un andamento dinamico non dissimile da quello dei fenomeni meteorologici. Jevons — che fu un precursore nel campo della dinamica economica — affermò nel 1862, che « sarebbe necessario che tutte le fluttuazioni commerciali fossero investigate con gli stessi metodi scientifici, che ci sono familiari in altre scienze complesse, come specialmente la meteorologia e il magnetismo terrestre » (²). Ed ecco — tra gli altri — Amoroso mettersi su

---

Cfr. « Bollettino mensile di statistica » dell'Istituto Centrale di Statistica del Regno d'Italia. Suppl. alla « Gazzetta Ufficiale », 1933. Tav. I.

(²) W. Stanley Jevons, *Investigations in currency and finance*. London, Macmillan, 1909 (2ª edizione, pag. 4).

questa via e rappresentare i valori del commercio estero francese (importazioni + esportazioni) dal 1828 al 1900 con la funzione:

$$z = A + B \operatorname{sen} \frac{2\pi}{12} x + C a^x \, , \qquad (56)$$

dove x è espressa in anni e A, B, C, a sono i coefficienti (¹); ossia con un'onda ciclica (sinusoidale, perchè considera per semplicità solo la funzione seno) oscillante intorno ad una curva esponenziale: quest'ultima rappresenterebbe la tendenza generale (crescente) dei valori del commercio estero francese lungo quel settantennio.

Pareto (²), fondandosi su distribuzioni ricavate da fonti prevalentemente fiscali, notò che il numero dei redditieri con reddito superiore ad x decresceva rapidamente al crescere di x; ossia che la forma di tali distribuzioni era ad un dipresso quella della tavola X *bis*, dove pure i limiti dei redditi presentano intervalli sempre maggiori.

Sebbene i redditi degl'indigenti fossero di difficile accertamento ed in generale esenti da accertamenti fiscali, egli ritenne — in base ad ovvî criterî economici — che non potessero esistere individui con reddito nullo e che quelli con reddito vicinissimo a 0 fossero in numero sparuto: ammise, cioè, l'esistenza di un primo ramo ascendente delle distribuzioni, ove queste fossero poste nella forma della tavola X. Ma l'andamento rapidamente ascendente di questo primo ramo lo persuase che, adattando una funzione conveniente solo al secondo ramo discendente, avrebbe trascurato un piccolo numero di redditieri.

L'andamento di questo secondo ramo lo condusse a sospettare l'esistenza di una forma tipicamente iperbolica:

$$N_x = H x^{-\alpha} \qquad (57)$$

e pertanto egli rappresentò graficamente a doppia scala logaritmica le distribuzioni di cui disponeva. L'andamento riuscì marcatamente rettilineo, sebbene per alcune distribuzioni — attraverso le accidentalità di andamento — si notasse una lieve inflessione. Essa, ad es., può pure rilevarsi nella seguente figura 41, dove abbiamo costruito il grafico logaritmico della detta tavola X *bis*. Ond'egli, mentre affermò la validità della (57), propose in seconda approssimazione la funzione:

$$N_x = H (x + a)^{-\alpha} \qquad (58)$$

(¹) LUIGI AMOROSO, *Lezioni di matematica finanziaria*. Napoli, Maio, 1923 (Vol. II, pag. 236). Cfr. anche RODOLFO BENINI, *Principii di statistica metodologica*. Torino, Unione Tip. Ed. Torinese, 1906 (pag. 174).

(²) VILFREDO PARETO, *Cours d'economie politique*. Lausanne, Rouge, 1896-97 (Vol. II); e *Manuale di economia politica*, op. cit.

dove l'influenza di a diminuisce al crescere di x; ed ancora in terza approssimazione aggiunse un fattore che, per $\beta > 0$, è rapidamente decrescente al crescere di x:

$$N_x = H\,(x + a)^{-\alpha}\,e^{-\beta x} \tag{59}$$

Sono queste le formule di Pareto sulla distribuzione dei redditi, universalmente note ed applicate anche a scopi pratici; perchè, per valori di x non forniti dalle statistiche, danno con buona approssimazione i valori corrispondenti di N_x.

Nella (57) la disuguaglianza dei redditi diminuisce al crescere di α: se il valore di α risultasse ad es. uguale a 2, a 3, ..., il numero dei reddi-

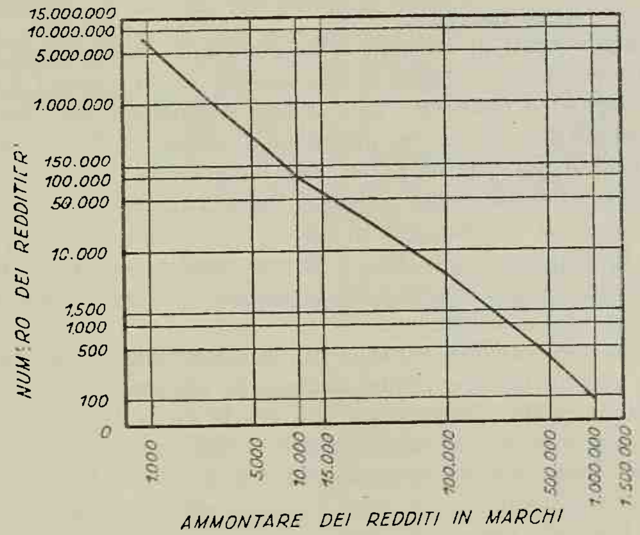


Fig. 41

tieri varierebbe in proporzione inversa al quadrato, al cubo .... dei redditi-limite. Pertanto α, per le distribuzioni obbedienti alla (57), suole spesso assumersi come indice di disuguaglianza dei redditi, invece del rapporto di concentrazione R, di cui ci siamo occupati al n. 14 del precedente capitolo, o di un qualsiasi indice di variabilità: il valore di α oscilla in generale tra 1 e 2 e varia da Stato a Stato, da regione a regione, da centro a centro in relazione all'assetto prevalentemente agricolo o industriale ed alle caratteristiche economiche generali di essi (1).

Potremmo riportare ancora altri esempî; ma, poichè nel corso di questo manuale avremo frequenti occasioni di esporne e su taluno di essi,

(1) COSTANTINO BRESCIANI-TURRONI, *La ricchezza delle città*, in « Annali del Seminario giuridico della R. Università di Palermo », 1912.

riguardante la (53) dovremo soffermarci a lungo nella Parte II, passiamo ad occuparci degli altri due problemi enunciati al n. 1, ossia del calcolo numerico dei coefficienti che compaiono nelle funzioni analitiche scelte e, in prima approssimazione, della verifica della bontà dell'adattamento.

3. Il metodo dei minimi quadrati. – Poichè, per le ragioni esposte al n. 1, le n coppie osservate x_i, y_i sono affette da perturbazioni, dalle quali si vuole prescindere, il numero dei coefficienti della funzione scelta è sempre minore di n. Pertanto il problema del calcolo dei coefficienti, affinchè non sia indeterminato, viene formulato in via generale imponendo una data condizione di approssimazione dei dati osservati y_i a quelli desunti dalla funzione.

Si sarebbe tentati di ritenere come condizione ideale di approssimazione quella che rende minima la somma dei valori assoluti degli scarti tra dati osservati e calcolati; però questa condizione, che sarebbe una generalizzazione di quella a cui obbedisce il valore centrale di un gruppo (n. 9 del precedente capitolo), può dar luogo all'inconveniente — analogo a quello del detto valore centrale, quando il numero dei dati è pari — di dar luogo ad una famiglia di curve, anzichè ad una curva sola: ed inoltre, se la funzione scelta non è lineare, quella condizione fa sorgere difficoltà analitiche gravi e talvolta insuperabili, a cagione dei valori assoluti, che in essa si assumono.

Abbiamo, però, veduto come la più generale misura di posizione di un gruppo sia quella che rende minima la somma dei quadrati degli scarti, ossia la media aritmetica, e che per le comuni serie statistiche il calcolo di essa equivale all'adattamento di una retta parallela all'asse delle ascisse; sorge, quindi, naturale l'idea di estendere tale condizione al caso in cui si sostituiscano ai dati osservati altri dati, desunti da una qualunque funzione assegnata.

Infatti la condizione di minimo della somma degli scarti quadratici tra dati osservati e calcolati non può non assicurare, in generale, un buon adattamento.

Limitandoci ancora alla considerazione di funzioni di una variabile, osserviamo che, se la funzione da adattare è ad es. la funzione razionale intera:

$$z = f(x) = a_0 + a_1 x + a_2 x^2 + \ldots\ldots + a_r x^r \qquad (60)$$

e le coppie x_i, y_i, formano un numero n di condizioni maggiore degli $r + 1$ coefficienti di questa funzione, la detta condizione può scriversi nella forma:

$$\Sigma (y_i - a_0 - a_1 x_i - \ldots\ldots - a_r x_i^r)^2 = \text{minimo} \qquad (61)$$

dove a_0, a_1, ... a_r sono i coefficienti incogniti e la somma s'intende estesa a tutti i valori osservati.

Semplici considerazioni di calcolo differenziale, su cui però dobbiamo sorvolare ([1]), fanno discendere dalla precedente condizione il sistema lineare di $r+1$ equazioni ad $r+1$ incognite:

$$\left.\begin{array}{l} \Sigma y_i = a_0 n + a_1 \Sigma x_i + \ldots + a_r \Sigma x_i^r \\ \Sigma x_i y_i = a_0 \Sigma x_i + a_1 \Sigma x_i^2 + \ldots + a_r \Sigma x_i^{r+1} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \Sigma x_i^r y_i = a_0 \Sigma x_i^r + a_1 \Sigma x_i^{r+1} + \ldots + a_r \Sigma x_i^{2r} \end{array}\right\} \quad (62)$$

risolvendo il quale si ottengono i valori dei coefficienti, per cui la condizione di minimo (61) riesce soddisfatta.

Le equazioni del sistema precedente si chiamano comunemente *equazioni normali*. La prima di esse assicura altresì — almeno per le funzioni razionali intere e in generale per quelle contenenti una costante additiva — l'uguaglianza tra la somma degli y_i e quella dei corrispondenti dati calcolati.

Questo metodo, al quale sono legati i nomi di Legendre (1806) e di Gauss (1809) — i quali peraltro lo dedussero da ipotesi molto più limitate di quelle, in base alle quali viene oggi largamente applicato — suol chiamarsi metodo dei minimi quadrati ([2]).

4. Formule semplificatrici. – Per calcolare la somma dei quadrati degli scarti tra dati osservati e calcolati, si consideri anzitutto che, ponendo:

$$\varepsilon_i = y_i - a_0 - a_1 x_i - \ldots - a_r x_i^r$$

quella somma può scriversi nella forma:

$$\begin{aligned} \Sigma \varepsilon_i^2 &= \Sigma \varepsilon_i (y_i - a_0 - a_1 x_i - \ldots - a_r x_i^r) \\ &= \Sigma \varepsilon_i y_i - a_0 \Sigma \varepsilon_i - a_1 \Sigma \varepsilon_i x_i - \ldots - a_r \Sigma \varepsilon_i x_i^r \end{aligned} \quad (63)$$

dove, in base alle (62) è:

$$\begin{array}{l} \Sigma \varepsilon_i = \Sigma (y_i - a_0 - a_1 x_i - \ldots - a_r x_i^r) = 0 \\ \Sigma \varepsilon_i x_i = \Sigma (y_i - a_0 - a_1 x_i - \ldots - a_r x_i^r) x_i = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \Sigma \varepsilon_i x_i^r = \Sigma (y_i - a_0 - a_1 x_i - \ldots - a_r x_i^r) x_i^r = 0 \end{array}$$

([1]) Uguagliando a zero le derivate parziali della (61) rispetto ad $a_0, a_1, \ldots a_r$, si ottiene:

$$-2 \Sigma (y_i - a_0 - a_1 x_i - \ldots - a_r x_i^r) x_i^s = 0 \qquad (s = 0, 1, \ldots r)$$

donde le (62), in cui il determinante dei coefficienti è in generale diverso da zero.

Talvolta la condizione di minimo (61) si esprime con un integrale definito, da cui si ricavano le equazioni normali in forma di integrali.

([2]) K. F. Gauss, *Theoria combinationis observationum erroribus minimis obnoxiae*, Comment. Soc. Göttingen, Vol. V, 1819-20 (e *Theoria motus corporum coelestium*, 1809, che venne alla luce tre anni dopo le *Nouvelles méthodes pour la détermination des orbites des comètes* di A. M. Legendre, dove già si parlava di un metodo dei minimi quadrati). Cfr. Guido Castelnuovo, *Calcolo delle probabilità*. Milano, «Dante Alighieri», 1919 (2ª edizione, Bologna, Zanichelli, 1925).

Pertanto la (63) diventa:

$$
\begin{aligned}
\Sigma \varepsilon_i^2 = \Sigma \varepsilon_i y_i &= \Sigma (y_i - a_0 - a_1 x_i - \dots - a_r x_i^r)\, y_i \\
&= \Sigma y_i^2 - a_0 \Sigma y_i - a_1 \Sigma x_i y_i - \dots - a_r \Sigma x_i^r y_i
\end{aligned}
\qquad (64)
$$

Quest'ultima formula è di comodo impiego, perchè le somme di essa, ad eccezione della prima, compaiono nel sistema (62) e quindi servono pure per il calcolo dei coefficienti.

Inoltre, se gli x_i sono in progressione aritmetica di ragione h, le somme delle potenze successive di essi, che compaiono nel sistema (62), possono ricavarsi assumendo rispettivamente i primi tre, quattro, ... termini della formula ricorrente:

$$
x_{n+1}^{s+1} = x_1^{s+1} + n h^{s+1} + (s+1)\, h^s\, \Sigma x_i + \binom{s+1}{2} h^{s-1}\, \Sigma x_i^2 + \dots + (s+1)\, \Sigma x_i^r, \qquad (65)
$$

e attribuendo rispettivamente ad s i valori: 1, 2, ... r; od anche dalla seguente:

$$
\Sigma x_i^s = n\, x_1^s + s\, x_1^{s-1}\, h\, \Sigma i + \binom{s}{2} x_1^{s-2}\, h^2\, \Sigma i^2 + \dots + h^s\, \Sigma i^s, \qquad (66)
$$

dove i varia da 1 ad $n-1$, le formule per il calcolo di Σi^s sono notissime e possono comunque ricavarsi, una volta per tutte, dalla (65).

In particolare si ha:

$$
\left.
\begin{aligned}
\Sigma x_i &= n x_1 + \frac{n(n-1)}{2} h \\
\Sigma x_i^2 &= n x_1^2 + 2 \frac{n(n-1)}{2} x_1 h + \frac{n(n-1)(2n-1)}{6} h^2 \\
\Sigma x_i^3 &= n x_1^3 + 3 \frac{n(n-1)}{2} x_1^2 h + 3 \frac{n(n-1)(2n-1)}{6} x_1 h^2 + \frac{n^2(n-1)^2}{4} h^3 \\
\Sigma x_i^4 &= n x_1^4 + 4 \frac{n(n-1)}{2} x_1^3 h + 6 \frac{n(n-1)(2n-1)}{6} x_1 h^2 + 4 \frac{n^2(n-1)^2}{4} x_1 h^3 \\
&\quad + \frac{6n^5 - 15n^4 + 10n^3 - n}{30} h^4 \\
&\dots\dots\dots\dots\dots\dots
\end{aligned}
\right| \qquad (67)
$$

Il calcolo di esse è rapidissimo, specie semplificando e mettendo in ognuna in evidenza i fattori comuni a gruppi di termini convenientemente scelti.

Infine, è utile osservare che, se si porta l'origine degli x_i alla loro media aritmetica semplice, ossia se si assumono, in luogo di essi, gli scarti dalla loro media aritmetica, il calcolo — comunque eseguito — delle successive potenze di x_i e delle altre espressioni che compaiono nelle (62) e (64) riesce generalmente più semplice.

Se poi gli x_i sono equidistanti, non solo la somma algebrica dei loro scarti, ma anche quella delle potenze dispari è nulla.

Non è, poi, difficile risalire dalla funzione degli scarti a quella dei valori. Ad es., assumendo la funzione:

$$f(x - M_x) = a'_0 + a'_1 (x - M_x) + a'_2 (x - M_x)^2 + a'_3 (x - M_x)^3 ,$$

questa può trasformarsi nella funzione:

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 ,$$

ponendo:

$$a'_0 - a'_1 M_x + a'_2 M^2_x - a'_3 M^3_x = a_0 \tag{68}$$

$$a'_1 - 2 a'_2 M_x + 3 a'_3 M^2_x = a_1 \tag{69}$$

$$a'_2 - 3 a'_3 M_x = a_2 \tag{70}$$

$$a'_3 = a_3 \tag{71}$$

Il calcolo è pure generalmente abbreviato, sostituendo anche agli y_i gli scarti dalla loro media aritmetica. In tal caso il passaggio dagli scarti ai dati originarî è immediato.

5. Applicazioni. – Se vogliamo ad es. adattare la funzione:

$$f(x) = a_0 + a_1 x$$

alla serie statistica del credito semestrale (valori provvisorî in milioni di lire) dei depositanti nelle nostre Casse ordinarie di risparmio durante gli anni 1920-25 (tavola XLII), dobbiamo risolvere il sistema (62) limitatamente alle prime due equazioni normali, i cui secondi membri constino dei primi due termini.

Ma, se consideriamo, invece degli x_i, i loro scarti dall'istante medio: fine dicembre 1922, sarà $\Sigma (x_i - 1922) = 0$ e si avrà:

$$a'_0 = \frac{\Sigma y_i}{n} , \qquad a'_1 = \frac{\Sigma (x_i - 1922)\, y_i}{\Sigma (x_i - 1922)^2} .$$

Per il calcolo numerico delle precedenti espressioni, anzitutto dalla seconda delle (67) ricaviamo immediatamente:

$$\Sigma (x_i - 1922)^2 = 13 \times 9 - \frac{1}{2} (13 \times 12 \times 3) + \frac{1}{24} (13 \times 12 \times 25) = 45{,}5$$

e poi si ha:

TAVOLA XLII.

x_i	$x_i - 1922$	y_i	$(x_i - 1922)\, y_i$	y_i^2
fine Dicembre 1919	— 3	5.454	— 16.362	29.746.116
» Giugno 1920	— 2,5	5.637	— 14.092,5	31.775.769
» Dicembre 1920	— 2	6.249	— 12.498	39.050.001
» Giugno 1921	— 1,5	7.003	— 10.504,5	49.042.009
» Dicembre 1921	— 1	7.448	— 7.448	55.472.704
» Giugno 1922	— 0,5	8.003	— 4.001,5	64.048.009
» Dicembre 1922	0	8.545	0	73.017.025
» Giugno 1923	+ 0,5	9.268	+ 4.634	85.895.824
» Dicembre 1923	+ 1	9.690	+ 9.690	93.896.100
» Giugno 1924	+ 1,5	10.331	+ 15.496,5	106.729.561
» Dicembre 1924	+ 2	10.982	+ 21.964	120.604.324
» Giugno 1925	+ 2,5	11.397	+ 28.492,5	129.891.609
» Dicembre 1925	+ 3	11.425	+ 34.275	130.530.625
Totali		111.432	— 64.906,5 + 114.552 49.645,5	1.009.699.676

$$a'_0 = \frac{111.432}{13} = 8.571{,}6923 \qquad a'_1 = \frac{49.645{,}5}{45{,}5} = 1.091{,}10989$$

Dalle (68) e (69) si ricava infine:

$$a_0 = 8.571{,}6923 - 1.091{,}10989 \times 1922 = -2.088.541{,}51628$$
$$a_1 = a'_1 = 1.091{,}10989$$
$$f(x) = -2.088.541{,}51628 + 1.091{,}10989\, x.$$

Calcolando i valori della funzione precedente per $x = 1919$; $1919{,}5$; 1920; ..... e ponendoli a confronto con quelli osservati (figura 42), la somma degli scarti risulterà nulla e quella dei loro quadrati sarà un minimo rispetto alla somma dei quadrati degli scarti risultanti da qualsiasi altra coppia di coefficienti.

Si noti che se, per semplicità, consideriamo come variabile: $x - 1922$, la funzione precedente diventa:

$$f(x - 1922) = 8571{,}6923 + 1.091{,}10989\,(x - 1922).$$

Avendo calcolato nella tavola precedente anche il valore di Σy_i^2, possiamo dalla (64) ottenere facilmente la somma dei quadrati degli scarti. Fondandoci per risparmio di calcoli su $f(x - 1922)$, si ha:

$$\Sigma \varepsilon_i^2 = 1.009.699.676 - 8.571{,}6923 \times 111.432 - 1.091{,}10989 \times 49.645{,}5 = 370.163{,}5825$$

Calcolando la media aritmetica ed estraendo la radice quadrata, si ottiene:

$$\sqrt{\frac{\Sigma \varepsilon_i^2}{n}} = \sqrt{\frac{370.163,5825}{13}} = \sqrt{28.474.12173} = \pm\ 168,74\ .$$

La formula precedente è analoga a quella dello scarto quadratico medio, di cui ci siamo occupati nel precedente capitolo; e, insieme ai

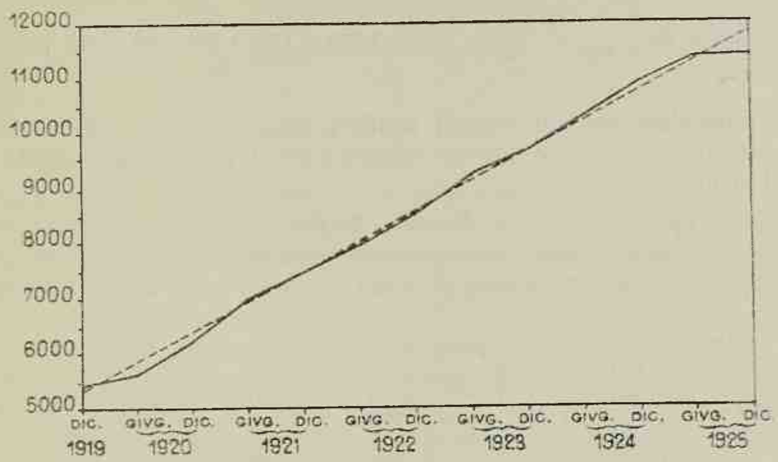


Fig. 42

valori singoli degli scarti, permetterà di esaminare se i dati desunti dalla funzione si possano sostituire a quelli osservati, ossia se la funzione scelta sia idonea a rappresentare questi ultimi.

Intanto l'indice precedente ci dice che, assumendo come credito semestrale dei depositanti i dati risultanti dalla funzione, invece di quelli effettivi, si commette un errore medio di $\pm$ 168,74 milioni, che, rispetto alla media aritmetica dei dati: $M_y = 8571,6923$, è l'1,97 per cento; e la figura, d'altra parte, rivela — almeno trascurando i dati estremi — un andamento dei segni degli scarti ed una piccolezza dello scarto massimo soddisfacenti. Sono questi i criterî generalmente adottati per giudicare della bontà dell'adattamento e noi li seguiremo in questo capitolo, riserbandoci di approfondirli nella Parte II.

Adatteremo adesso la funzione:

$$f(x) = a_0 + a_1 x + a_2 x^2$$

alla distribuzione, secondo il numero dei figli, delle vedove rimaritatesi a Milano nel 1916 (tavola V). A tal fine occorrerà risolvere il sistema (62) limitatamente alle prime tre equazioni normali, i cui secondi membri constino dei primi tre termini.

Essendo gli x_i poco numerosi e di piccola entità, conviene condurre i calcoli su di essi: la considerazione dei loro scarti dalla media aritmetica non abbrevierebbe molto i calcoli.

Poichè essi sono, nel nostro esempio, i primi sei termini della serie naturale dei numeri a cominciare da 0, si ha:

$$n = 6 \qquad \Sigma x_i = \frac{6 \times 5}{2} = 15 \qquad \Sigma x_i^2 = \frac{6 \times 5 \times 11}{6} = 55$$

$$\Sigma x_i^3 = \frac{36 \times 25}{4} = 225, \qquad \Sigma x_i^4 = \frac{6 \times 7.776 - 15 \times 1.296 + 10 \times 216 - 6}{30} = 979$$

Trattandosi però di numeri molto piccoli, anche in questo caso il calcolo diretto avrebbe fornito rapidamente i risultati precedenti. Infatti si ha:

TAVOLA XLIII.

x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$
0	109	0	0	0	0	0
1	60	1	1	1	60	60
2	35	4	8	16	70	140
3	17	9	27	81	51	153
4	5	16	64	256	20	80
5	1	25	125	625	5	25
15	227	55	225	979	206	458

Nella tavola precedente abbiamo, inoltre, calcolato le espressioni: $\Sigma x_i y_i$ e $\Sigma x_i^2 y_i$.

Possiamo, adunque, porre il sistema:

$$\begin{aligned} 227 &= a_0 \;\; 6 + a_1 \;\; 15 + a_2 \;\; 55 \\ 206 &= a_0 \;\; 15 + a_1 \;\; 55 + a_2 \;\; 225 \\ 458 &= a_0 \;\; 55 + a_1 \;\; 225 + a_2 \;\; 979 \end{aligned}$$

Risolvendolo con la nota regola di Cramer, si ha:

$$a_0 = 105{,}96 \qquad a_1 = -\,45{,}39 \qquad a_2 = 4{,}94$$

e quindi:

$$f(x) = 105{,}96 - 45{,}39\,x + 4{,}94\,x^2.$$

Calcolando i valori della funzione precedente per $x = 0, 1, \ldots\ 5$, e ponendoli a confronto con quelli osservati si ottiene:

TAVOLA XLIV.

x_i	y_i	$f(x_i)$	$\varepsilon_i = y_i - f(x_i)$	ε_i^2
0	109	105,96	+ 3,04	9,2416
1	60	65,51	— 5,51	30,3601
2	35	34,94	+ 0,06	0,0036
3	17	14,25	+ 2,75	7,5625
4	5	3,44	+ 1,56	2,4336
5	1	2,51	— 1,51	2,2801
Totali	227	226,61		51,8815

Dalla quarta colonna o dal totale della terza si può controllare che la somma degli ε_i positivi è uguale a quella dei negativi (con l'approssimazione consentita dai decimali trascurati nei calcoli). Questo ed altri eventuali controlli sono necessari, data la facilità di commettere errori in calcoli così laboriosi.

La somma dei quadrati degli ε_i, la quale dev'essere un minimo rispetto alla somma dei quadrati degli scarti risultanti da qualsiasi altro sistema di coefficienti, potrebbe calcolarsi a mezzo della (64); ma lo scarso numero e la piccola entità degli scarti rende in questo caso preferibile il calcolo diretto, che appunto trovasi riportato nell'ultima colonna della tavola precedente.

Ricavandone la media aritmetica ed estraendo da essa la radice quadrata, si ottiene:

$$\sqrt{\frac{\Sigma \varepsilon_i^2}{n}} = \sqrt{\frac{51,8815}{6}} = \sqrt{8,6469} = \pm 2,94 .$$

Il risultato precedente, insieme ai valori singoli degli scarti (posti meglio in evidenza nella figura 43), permetterà di esaminare se i dati desunti dalla funzione si possano sostituire a quelli osservati, ossia se la funzione scelta sia idonea a rappresentare la distribuzione in esame. Ci limitiamo ad osservare che, assumendo come numero di vedove rimaritatesi i dati risultanti dalla funzione, invece di quelli effettivi, si commette per ogni gruppo un errore medio di $\pm$ 2,94 vedove, che rispetto alla media aritmetica dei gruppi: $M_y = 37,83$, è il 7,77 per cento.

6. APPROSSIMAZIONI SUCCESSIVE SECONDO TCHEBYCHEF. METODO DI CAUCHY E APPLICAZIONI. – Se, per la risoluzione del sistema delle equazioni normali, si segue il noto metodo di sostituzione, i coefficienti a_0,

$a_1, \ldots a_r$ possono essere calcolati con molta facilità ed in modo che si abbia rapidamente una misura sommaria dell'approssimazione man mano raggiunta tra i dati osservati e quelli desunti della funzione.

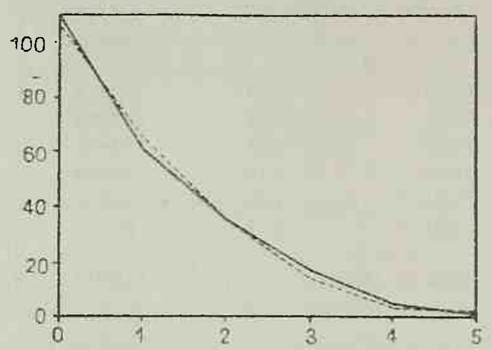


Fig. 43

Questo procedimento dev'essere preferito quando, nei riguardi delle coppie osservate, i criterî adottati abbiano fatto scegliere una funzione della forma (60), ma non abbiano permesso di determinarne il grado, ossia il numero dei termini di cui convenga valersi. Ciò significa che nel procedimento non si assume una data funzione, ma una funzione generica, che si ammette sviluppabile in serie di potenze o almeno rappresentabile con buona approssimazione mediante un polinomio, e della quale si vogliono determinare i coefficienti dei termini non trascurabili (1).

Sappiamo, anzitutto, che, assumendo:

$$f(x) = a_0 \qquad (72)$$

la condizione dei minimi quadrati è soddisfatta col porre:

$$\Sigma y_i = a_0 n \qquad \text{ossia:} \quad a_0 = \frac{\Sigma y_i}{n}. \qquad (A)$$

Dalla (64) si trae:

$$\Sigma \varepsilon_i^2 = \Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n} = \Sigma\left(y_i - \frac{\Sigma y_i}{n}\right)^2.$$

Non essendo l'approssimazione soddisfacente, si assumerà:

$$f(x) = a_0 + a_1 x \qquad (73)$$

dove sappiamo che a_0 e a_1 sono determinati, con la condizione dei minimi quadrati, dalle due equazioni normali:

$$\Sigma y_i = a_0 n + a_1 \Sigma x_i$$
$$\Sigma x_i y_i = a_0 \Sigma x_i + a_1 \Sigma x_i^2.$$

---

(1) Il procedimento esposto è dovuto a TCHEBYCHEF e trovasi ampiamente esposto in CHARLES JORDAN, *Statistique mathématique*. Paris, Gauthier-Villars, 1927 (Cap. II, n. 11 e Cap. X, n. 97). Su altri procedimenti di approssimazione, dovuti a Gram, allo stesso Jordan, a Lorenz, ecc., cfr. una nota di M. JACOB, *Un nuovo metodo di perequazione*, in « Rivista it. di statistica », 1931, Vol. III, pag. 346. – Avvertiamo che negli stadî successivi l'approssimazione non è necessariamente sempre crescente.

Ricavando dalla prima la relazione:

$$a_0 = \frac{\Sigma y_i}{n} - a_1 \frac{\Sigma x_i}{n} \tag{74}$$

e sostituendo nella seconda, si ottiene:

$$\Sigma x_i y_i = \left(\frac{\Sigma y_i}{n} - a_1 \frac{\Sigma x_i}{n}\right) \Sigma x_i + a_1 \Sigma x_i^2$$

da cui, ordinando,

$$\Sigma\left(y_i - \frac{\Sigma y_i}{n}\right) x_i = a_1 \Sigma\left(x_i - \frac{\Sigma x_i}{n}\right) x_i$$

od anche:

$$\Sigma\left(y_i - \frac{\Sigma y_i}{n}\right)\left(x_i - \frac{\Sigma x_i}{n}\right) = a_1 \Sigma\left(x_i - \frac{\Sigma x_i}{n}\right)^2 .$$

Semplificando i simboli, dalla precedente equazione si trae:

$$a_1 = \frac{\Sigma D x_i D y_i}{\Sigma (D x_i)^2} \tag{B}$$

Questa formula potrà calcolarsi direttamente, utilizzando il calcolo già eseguito per ottenere il valore di (A).

Sostituendo nella (64) ad x_i ed a y_i le espressioni Dx_i e Dy_i e tenendo presente che $\Sigma Dy_i = 0$, si ottiene

$$\Sigma \varepsilon_i^2 = \Sigma (D y_i)^2 - \frac{(\Sigma D x_i D y_i)^2}{\Sigma (D x_i)^2} \tag{75}$$

Essendo:

$$\Sigma (D y_i)^2 - \frac{(\Sigma D x_i D y_i)^2}{\Sigma (D x_i)^2} = \Sigma (D y_i - a_1 D x_i)^2 = \Sigma (y_i - a_0 - a_1 x_i)^2$$

la (75) è la somma dei quadrati degli scarti tra le Dy_i osservate e quelle desunte dalla funzione:

$$f(D x_i) = a_1 D x_i ,$$

od anche la somma dei quadrati degli scarti tra i dati osservati e quelli desunti dalla (73).

Se, in base alla (75) ed all'esame dei singoli scarti, l'approssimazione si reputerà soddisfacente, sarà facile passare dal calcolo di a_1 a quello di a_0, dato dalla (74), anche perchè il primo termine di quest'ultima formula è stato già calcolato in prima approssimazione, ed il secondo è stato pure calcolato per la determinazione delle Dx_i.

Se, invece, l'approssimazione non si reputerà soddisfacente, si assumerà:

$$f(x) = a_0 + a_1 x + a_2 x^2, \tag{76}$$

dove sappiamo che i coefficienti sono determinati, con la condizione dei minimi quadrati, dal sistema:

$$
\begin{aligned}
\Sigma y_i &= a_0\, n + a_1 \Sigma x_i + a_2 \Sigma x_i^2 \\
\Sigma x_i y_i &= a_0 \Sigma x_i + a_1 \Sigma x_i^2 + a_2 \Sigma x_i^3 \\
\Sigma x_i^2 y_i &= a_0 \Sigma x_i^2 + a_1 \Sigma x_i^3 + a_2 \Sigma x_i^4
\end{aligned}
$$

Dalla prima equazione si trae:

$$a_0 = \frac{\Sigma y_i}{n} - a_1 \frac{\Sigma x_i}{n} - a_2 \frac{\Sigma x_i^2}{n} \tag{77}$$

e, sostituendo nella seconda e terza ed ordinando:

$$\Sigma\left(y_i - \frac{\Sigma y_i}{n}\right) x_i = a_1 \Sigma\left(x_i - \frac{\Sigma x_i}{n}\right) x_i + a_2 \Sigma\left(x_i^2 - \frac{\Sigma x_i^2}{n}\right) x_i$$

$$\Sigma\left(y_i - \frac{\Sigma y_i}{n}\right) x_i^2 = a_1 \Sigma\left(x_i - \frac{\Sigma x_i}{n}\right) x_i^2 + a_2 \Sigma\left(x_i^2 - \frac{\Sigma x_i^2}{n}\right) x_i^2;$$

equazioni che possiamo scrivere nella forma:

$$\Sigma\left(y_i - \frac{\Sigma y_i}{n}\right)\left(x_i - \frac{\Sigma x_i}{n}\right) = a_1 \Sigma\left(x_i - \frac{\Sigma x_i}{n}\right)^2 + a_2 \Sigma\left(x_i^2 - \frac{\Sigma x_i^2}{n}\right)\left(x_i - \frac{\Sigma x_i}{n}\right)$$

$$\Sigma\left(y_i - \frac{\Sigma y_i}{n}\right)\left(x_i^2 - \frac{\Sigma x_i^2}{n}\right) = a_1 \Sigma\left(x_i - \frac{\Sigma x_i}{n}\right)\left(x_i^2 - \frac{\Sigma x_i^2}{n}\right) + a_2 \Sigma\left(x_i^2 - \frac{\Sigma x_i^2}{n}\right)^2.$$

Adottando le notazioni precedenti e ponendo ancora:

$$x_i^2 - \frac{\Sigma x_i^2}{n} = D x_i^2,$$

dalla prima di queste due equazioni si ricava:

$$a_1 = \frac{\Sigma D x_i D y_i}{\Sigma (D x_i)^2} - a_2 \frac{\Sigma D x_i^2 D x_i}{\Sigma (D x_i)^2} \tag{78}$$

e, sostituendo nella seconda e ordinando:

$$\Sigma\left(D y_i - \frac{\Sigma D x_i D y_i}{\Sigma (D x_i)^2} D x_i\right) D x_i^2 = a_2 \Sigma\left(D x_i^2 - \frac{\Sigma D x_i^2 D x_i}{\Sigma (D x_i)^2} D x_i\right) D x_i^2$$

od anche:

$$\Sigma\left(D y_i - \frac{\Sigma D x_i D y_i}{\Sigma (D x_i)^2} D x_i\right)\left(D x_i^2 - \frac{\Sigma D x_i^2 D x_i}{\Sigma (D x_i)^2} D x_i\right) = a_2 \Sigma\left(D x_i^2 - \frac{\Sigma D x_i^2 D x_i}{\Sigma (D x_i)^2} D x_i\right)^2.$$

Semplificando i simboli, dalla precedente equazione si trae:

$$a_2 = \frac{\Sigma D^2 x_i^2 D^2 y_i}{\Sigma (D^2 x_i^2)^2} \tag{C}$$

formula, che potrà calcolarsi direttamente, utilizzando i calcoli già eseguiti per ottenere (A) e (B).

D'altra parte, si ha:

$$\Sigma D^2 y_i = \Sigma D^2 x_i = 0$$

e pertanto la (64) porge:

$$\Sigma \varepsilon_i^2 = \Sigma (D^2 y_i)^2 - \frac{(\Sigma D^2 x_i^2 D^2 y_i)^2}{\Sigma (D^2 x_i^2)^2} \tag{79}$$

Essendo:

$$\Sigma (D^2 y_i)^2 - \frac{(\Sigma D^2 x_i^2 D^2 y_i)^2}{\Sigma (D^2 x_i^2)^2} = \Sigma (D^2 y_i - a_2 D^2 x_i^2)^2 = \Sigma (y_i - a_0 - a_1 x - a_2 a^2)^2 ,$$

la (79) è la somma dei quadrati degll scarti tra le $D^2 y_i$ osservate e quelle desunte dalla funzione:

$$f (D^2 x_i^2) = a_2 D^2 x_i^2 ,$$

od anche la somma dei quadrati degli scarti tra i dati osservati e quelli desunti dalla (76).

Se l'approssimazione si riterrà soddisfacente, si potrà facilmente passare dal calcolo di a_2 a quello di a_1 e di a_0 della (76) in base alle formule (78) e (77), anche perchè esse constano di elementi già calcolati.

In caso contrario, con procedimento in tutto analogo, potranno ancora considerarsi i termini: $a_3 x^3$, ... $a_r x^r$.

Si avverta che per le espressioni: $D x_i$, $D^2 x^2{}_i$, ..., che compaiono nelle formule (B), (C), ... potranno costruirsi prontuarî, che ne forniscano i valori in base a leggi assegnate di variazione degli x_i, e per dati valori di n.

Prontuarî siffatti sono stati costruiti nel caso che gli x_i variino in progressione aritmetica di ragione 1, come accade ad es. nelle serie storiche; e si comprende facilmente quanto ne risulti agevolato il calcolo di quelle formule [1].

Ma, indipendentemente dall'uso di prontuarî, il metodo esposto si presta ad un'interessante modificazione, dovuta a Cauchy (1835), la quale — pur eludendo la condizione dei minimi quadrati — spesso conduce a risultati poco diversi ed assicura una grande rapidità di calcoli.

[1] Cfr. VILFREDO PARETO, *Tables pour faciliter l'application de la méthode des moindres carrés.* Communication présentée à l'Assemblée annuelle des statisticiens officiels et de la Société Suisse de Statistique, Lausanne, 1898. – Altre tavole sono riportate da JORDAN, op. cit., pag. 27 e segg; ma il prontuario, che abbiamo riportato nell'Appendice *B*)II a questa Parte I può abbreviare molto i calcoli.

È questa rapidità di calcoli, che ne rende comodo l'impiego, specie in quei casi, in cui il metodo dei minimi quadrati imponga calcoli straordinariamente laboriosi senza condurre ad uno scarto quadratico medio sensibilmente minore.

Si tratta di una semplice modificazione delle equazioni normali (62), che noi esporremo nella forma delle approssimazioni successive, nella quale ha più generale impiego [1].

Nei successivi stadî di approssimazione si è pervenuti a relazioni e poscia a formule (A), (B), (C), ..., dove compaiono moltiplicatori della forma:

$$+1,\ D\,x_i\,,\ D^2\,x_i^2\,,\ .....$$

Ora, se come moltiplicatori si assumono invece i valori $+1$, 0, -1, secondo che le precedenti espressioni siano positive, nulle o negative, la prima di esse resterà $+1$, la seconda diventerà $sg\ Dx_i$ (espressione abbreviata di « segno di Dx_i »), e così di seguito.

Ne viene che la formula (A):

$$a_0 = \frac{\Sigma y_i}{n} \tag{A}$$

non subirà modificazioni; la formula (B) diventerà:

$$a_1 = \frac{\Sigma D\,y_i\,sg\,D\,x_i}{\Sigma D\,x_i\,sg\,D\,x_i} \tag{B'}$$

che potrà pure calcolarsi, utilizzando il calcolo eseguito per ottenere (A), e permetterà di risalire al valore di a_0 dato dalla (74) (si noti che $Dx_i\ sg\ Dx_i = |Dx_i|$); poscia, in luogo di (C), si avrà:

$$a_2 = \frac{\Sigma D^2\,y_i\ sg\,D^2\,x_i^2}{\Sigma D^2\,x_i^2\,sg\,D^2 x_i^2} \tag{C'}$$

che potrà calcolarsi in base alle espressioni:

$$D^2\,y_i = D\,y_i - \frac{\Sigma D\,y_i\,sg\,D\,x_i}{\Sigma D\,x_i\,sg\,D\,x_i}\,D\,x_i\ , \qquad D^2\,x_i^2 = D\,x^2 - \frac{\Sigma D\,x_i^2\,sg\,D\,x_i}{\Sigma D\,x_i\ sg\,D\,x_i}\,D\,x_i$$

$$D\,y_i = y_i - \frac{\Sigma y_i}{n} \qquad\qquad D\,x_i^2 = x^2 - \frac{\Sigma x_i^2}{n}$$

---

(1) Cfr. R. Radau, *Etudes sur les formules d'interpolation*. Paris, Gauthier-Villars, 1891. — E. Carvallo, *Le calcul des probabilités et ses applications*. Paris, Gauthier-Villars, 1912 (Cap. III, § 3).

cioè utilizzando i calcoli eseguiti per ottenere (A) e (B′), e permetterà di risalire al valore — analogo a quello fornito dalla (78) — che a_1 assume in questo terzo stadio:

$$a_1 = \frac{\Sigma D\, y_i\, sg\, D\, x_i}{\Sigma D\, x_i\, sg\, D\, x_i} - a_2 \frac{\Sigma D\, x_i^2\, sg\, D\, x_i}{\Sigma D\, x_i\, sg\, D\, x_i} \qquad (78\,bis)$$

e poi al valore di a_0, dato dalla (77). E così via.

È facile verificare che con le precedenti formule si annulla la somma algebrica degli scarti tra i dati osservati e quelli calcolati. Si ha infatti:

$$\Sigma\,(y_i - a_0) = \Sigma\left(y_i - \frac{\Sigma\, y_i}{n}\right) = 0$$

$$\Sigma\,(y_i - a_0 - a_1 x) = \Sigma\left(D\, y_i - \frac{\Sigma D\, y_i\, sg\, D\, x_i}{\Sigma D\, x_i\, sg\, D\, x_i}\, D\, x_i\right) = 0$$

$$\Sigma\,(y_i - a_0 - a_1 x - a_2 x^2) = \Sigma\left(D^2 y_i - \frac{\Sigma D^2 y_i\; sg\, D^2 x_i^2}{\Sigma D^2 x_i^2\, sg\, D^2 x_i^2}\, D^2 x_i^2\right) = 0$$

. . . .

Mentre, adunque, col metodo dei minimi quadrati — almeno nei riguardi delle funzioni, di cui ci occupiamo — si annulla la somma algebrica degli scarti ed inoltre si rende minima la somma dei quadrati di essi, col metodo ora esposto si raggiunge il primo scopo, ma si elude il secondo.

Si noti, infine, che con questo metodo non saranno più valide le formule per il calcolo rapido delle somme dei quadrati degli scarti tra i dati osservati e quelli calcolati successivamente; ma, poichè potrà sempre disporsi degli scarti singoli, sarà molto facile calcolare gli scarti medi numerici, interpretandoli con le debite cautele (vedasi il n. 13 del precedente capitolo).

Un esempio mostrerà la grande semplicità di calcoli, che le modificazioni di Cauchy assicurano, malgrado l'apparente complicazione dei simboli. Si voglia, ad es., verificare l'adattamento alla serie seguente — che dà il numero (in migliaia) degli emigranti italiani nei singoli anni del decennio 1904-13 —

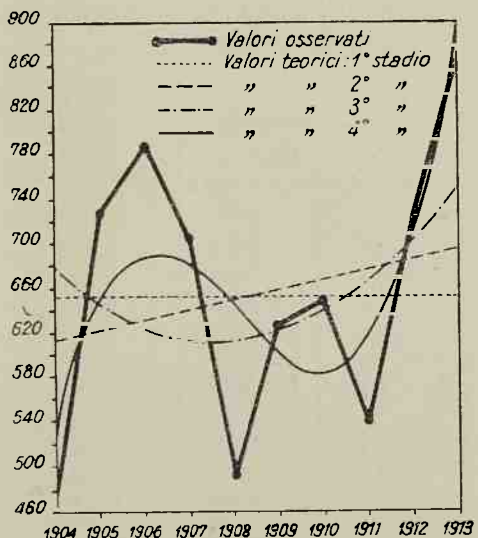


Fig. 44

di una funzione ignota per la sua complessità, di cui non si possa sicuramente affermare la forma trigonometrica, ma che si abbia ragione di ritenere ben rappresentabile con un polinomio.

TAVOLA XLV.

Anni	Numero degli emigranti, in migliaia	Anni	Numero degli emigranti, in migliaia
1904	471	1909	626
1905	726	1910	651
1906	788	1911	534
1907	705	1912	711
1908	487	1913	873

Sostituendo, per risparmio di calcoli, agli anni del decennio considerato i numeri interi da 0 a 9, si comincia con l'assumere la (72). In base alla formula (A), è:

$$f(x) = a_0 = \frac{\Sigma y_i}{n} = 657,2$$

Prescindendo da ogni altra considerazione, uno sguardo al diagramma (figura 44) persuade facilmente che l'approssimazione non è soddisfacente.

Assumendo, allora, la (73) e calcolando a_1 in base alla (B′), si ha:

TAVOLA XLVI.

$Dx_i = x_i - 4,5$	$Dy_i = y_i - 657,2$	$Dy_i \, sg \, Dx_i$	$Dx_i \, sg \, Dx_i$	$f(Dx_i) = 8,72 \, Dx_i$	$Dy_i - f(Dx_i) = D^2 y_i$
— 4,5	— 186,2	+ 186,2	+ 4,5	— 39,24	— 146,96
— 3,5	+ 68,8	— 68,8	+ 3,5	— 30,52	+ 99,32
— 2,5	+ 130,8	— 130,8	+ 2,5	— 21,80	+ 152,60
— 1,5	+ 47,8	— 47,8	+ 1,5	— 13,08	+ 60,88
— 0,5	— 170,2	+ 170,2	+ 0,5	— 4,36	— 165,84
+ 0,5	— 31,2	— 31,2	+ 0,5	+ 4,36	— 35,56
+ 1,5	— 6,2	— 6,2	+ 1,5	+ 13,08	— 19,28
+ 2,5	— 123,2	— 123,2	+ 2,5	+ 21,80	— 145,00
+ 3,5	+ 53,8	+ 53,8	+ 3,5	+ 30,52	+ 23,28
+ 4,5	+ 215,8	+ 215,8	+ 4,5	+ 39,24	+ 176,56
	| 1.034 |	+ 626 — 408	25		| 1.025,28 |
		218			

$$a_1 = \frac{\Sigma D y_i \, sg \, D x_i}{\Sigma D x_i \, sg \, D x_i} = \frac{218}{25} = 8,72 \qquad f(D x_i) = 8,72 \, D x_i \, .$$

La quinta colonna della tavola precedente riporta i dati desunti dalla precedente funzione e l'ultima colonna contiene gli scarti di essi dalle Dy_i; scarti che coincidono con quelli tra i dati osservati e i dati ricavati dalla (73). Essi sono ancora molto grandi e dimostrano che il guadagno di approssimazione è stato minimo: basta considerare che lo scarto numerico medio è adesso: 102,528, mentre prima era (seconda colonna): 103,4.

Non conviene, adunque, arrestarsi nel procedimento.

Per calcolare a_2 in base alla (C′), notiamo che le $D^2 y_i$ sono fornite dall'ultima colonna della tavola precedente e che occorrerà calcolare l'espressione:

$$D^2 x_i^2 = D x_i^2 - \frac{\Sigma D x_i^2 \, sg \, D x_i}{\Sigma D x_i \, sg \, D x_i} D x_i$$

dove sappiamo che è:

$$D x_i^2 = x_i^2 - \frac{\Sigma x_i^2}{n} .$$

Si avrà anzitutto:

TAVOLA XLVII.

x_i^2	$Dx_i^2 = x_i^2 - 28{,}5$	$Dx_i^2 \, sg \, Dx_i$	$Dx_i \, sg \, Dx_i$	$\frac{225}{25} Dx_i$	$D^2x_i^2 = Dx_i^2 - \frac{225}{25} Dx_i$
0	— 28,5	+ 28,5	+ 4,5	— 40,5	+ 12
1	— 27,5	+ 27,5	+ 3,5	— 31,5	+ 4
4	— 24,5	+ 24,5	+ 2,5	— 22,5	— 2
9	— 19,5	+ 19,5	+ 1,5	— 13,5	— 6
16	— 12,5	+ 12,5	+ 0,5	— 4,5	— 8
25	— 3,5	— 3,5	+ 0,5	+ 4,5	— 8
36	+ 7,5	+ 7,5	+ 1,5	+ 13,5	— 6
49	+ 20,5	+ 20,5	+ 2,5	+ 22,5	— 2
64	+ 35,5	+ 35,5	+ 3,5	+ 31,5	+ 4
81	+ 52,5	+ 52,5	+ 4,5	+ 40,5	+ 12
285		+228,5	25		0
		— 3,5			
		225			

Il totale dell'ultima colonna, che dev'essere sempre nullo, è stato calcolato per controllare l'esattezza dei calcoli.

Calcolando poscia, in base ai dati dell'ultima colonna della tav. XLVI ed ai segni dell'ultima colonna della tavola precedente, le espressioni

$D^2 y_i \, sg \, D^2 x_i^2$; e poi, in base ai valori assoluti di quest'ultima colonna, le espressioni $D^2 x_i^2 \, sg \, D^2 x_i^2$, si ottiene:

$$a_2 = \frac{\Sigma D^2 y_i \, sg \, D^2 x_i^2}{\Sigma D^2 x_i^2 \, sg \, D^2 x_i^2} = \frac{304,4}{64} = 4.756$$

$$f(D^2 x_i^2) = 4,756 \; D^2 x_i^2.$$

Nella tavola seguente si trovano riportati gli scarti tra le $D y_i$ e i dati corrispondenti desunti dalla funzione precedente:

TAVOLA XLVIII.

$D^2 y_i$	$f(D^2 x_i^2) = 4,756 \; D^2 x_i^2$	$D^2 y_i - f(D^2 x_i^2) = D^3 y_i$
— 146,96	+ 57,072	— 204,032
+ 99,32	+ 19,024	+ 80,296
+ 152,60	— 9,512	+ 162,112
+ 60,88	— 28,536	+ 89,416
— 165,84	— 38,048	— 127,792
— 35,56	— 38,048	+ 2,488
— 19,28	— 28,536	+ 9,256
— 145,00	— 9,512	— 135,488
+ 23,28	+ 19,024	+ 4,256
+ 176,56	+ 57,072	+ 119,488
		934,624

Lo scarto numerico medio è adesso minore del precedente (93,4624, ossia il 14,2 per cento di 657,2, valor medio di y_i), ma ancora alquanto elevato: la figura 44 mette in luce i tre stadî del procedimento, dimostrando come — se lo scopo, a cui si mira, è quello di riprodurre i diversi cicli della serie osservata — convenga introdurre ancor nuovi termini nella nostra funzione, affinchè con successive curvature si riesca a rappresentare quei diversi cicli.

A tal fine notiamo che si dovrà calcolare la formula:

$$a_3 - \frac{\Sigma D^3 y_i \, sg \, D^3 x_i^3}{\Sigma D^3 x_i^3 \, sg \, D^3 x_i^3}$$

Le $D^3 y_i$ si trovano già riportate nell'ultima colonna della tavola precedente e le $D^3 x_i^3$ sono date dall'espressione:

$$D^3 x_i^3 = D^2 x_i^3 - \frac{\Sigma D^2 x_i^3 \, sg \, D^2 x_i^2}{\Sigma D^2 x_i^2 \, sg \, D^2 x_i^2} D^2 x^2,$$

dove è:

$$D^2\, x_i^3 = D\, x_i^3 - \frac{\Sigma\, D\, x_i^3\, sg\, D\, x_i}{\Sigma\, D\, x_i\, sg\, D\, x_i}\, D\, x_i$$

$$D\, x_i^3 = x_i^3 - \frac{\Sigma\, x_i^3}{n}.$$

Eseguendo i calcoli si ottiene:

$$\frac{\Sigma\, x_i^3}{n} = 202,5, \quad \frac{\Sigma\, D\, x_i^3\, sg\, D\, x_i}{\Sigma\, D\, x_i\, sg\, D\, x_i} = 73, \quad \frac{\Sigma\, D^2\, x_i^3\, sg\, D^2\, x_i^2}{\Sigma\, D^2\, x_i^2\, sg\, D^2\, x_i^2} = 13,5$$

e poi:

$$a_3 = \frac{571}{144} = 3,965$$

$$f\,(D^3\, x_i^3) = 3,965\ D^3\, x_i^3.$$

Risulta:

TAVOLA XLIX.

$D^3\, y_i$	$f\,(D^3\, x_i^3) = 3,965\ D^3\, x_i^3$	$D^3\, y_i - f\,(D^3\, x_i^3) = D^4\, y_i$
— 204,032	— 142,740	— 61,292
+ 80,296	0	+ 80,296
+ 162,112	+ 59,475	+ 102,637
+ 89,416	+ 59,475	+ 29,941
— 127,792	+ 23,790	— 151,582
+ 2,488	— 23,790	+ 26,278
+ 9,256	— 59,475	+ 68,731
— 135,488	— 59,475	— 76,013
+ 4,256	0	+ 4,256
+ 119,488	+ 142,740	— 23,252
		| 624,278 |

Lo scarto numerico medio adesso è, adunque, notevolmente minore del precedente (62,4278, ossia il 9,5 per cento del valor medio di y_i)

Se, anche in base all'esame analitico degli scarti, ci potessimo fermare a questo quarto stadio, potremmo scrivere:

$$a_2 = \frac{\Sigma\, D^2\, y_i\, sg\, D^2\, x_i^2}{\Sigma\, D^2\, x_i^2\, sg\, D^2\, x_i^2} - a_3 \frac{\Sigma\, D^2\, x_i^3\, sg\, D^2\, x_i^2}{\Sigma\, D^2\, x_i^2\, sg\, D^2\, x_i^2}$$

$$= 4,756 - 3,965 \times 13,5 = -48,7715,$$

$$a_1 = \frac{\Sigma\, D\, y_i\, sg\, D\, x_i}{\Sigma\, D\, x_i\, sg\, D\, x_i} - a_2 \frac{\Sigma\, D\, x_i^2\, sg\, D\, x_i}{\Sigma\, D\, x_i\, sg\, D\, x_i} - a_3 \frac{\Sigma\, D\, x_i^3\, sg\, D\, x_i}{\Sigma\, D\, x_i\, sg\, D\, x_i}$$

$$= 8,72 + 48,7715 \times 9 - 3,965 \times 73 = 158,2185,$$

$$a_0 = \frac{\Sigma y_i}{n} - a_1 \frac{\Sigma x_i}{n} - a_2 \frac{\Sigma x_i^2}{n} - a_3 \frac{\Sigma x_i^3}{n}$$

$$= 657{,}2 - 158{,}2185 \times 4{,}5 + 48{,}7715 \times 28{,}5 - 3{,}965 \times 202{,}5 = 532{,}292,$$

e quindi:

$$f(x) = 532{,}292 + 158{,}2185\, x - 48{,}7715\, x^2 + 3{,}965\, x^3.$$

Ripetiamo che questo procedimento di Cauchy si segue, anche quando sia assegnato il grado della (60). Applicandolo alla distribuzione delle vedove, di cui alla tavola V, si avrebbe:

$$f(x) = 105{,}25 - 45{,}82\, x + 5{,}142\, x^2.$$

Calcolando per $x = 0, 1, \ldots. 5$ i dati teorici risultanti dalla funzione precedente e ponendoli a confronto con quelli osservati, non si ottengono risultati sensibilmente diversi da quelli già ottenuti al n. 5 col metodo dei minimi quadrati, nonostante la maggior somma di scarti quadratici.

7. Metodo delle somme. – Se delle coppie osservate x_i, y_i formiamo tanti gruppi consecutivi, quanti sono i coefficienti della funzione scelta, possiamo costruire un sistema di altrettante equazioni in base alle condizioni che in ogni gruppo la somma dei dati osservati risulti uguale a quella dei dati calcolati.

Se la funzione scelta è la (60), si ha il sistema, lineare rispetto ai coefficienti:

$$\left.\begin{aligned} \sum_{1}^{\alpha} y_i &= a_0\, \alpha + a_1 \sum_{1}^{\alpha} x_i + \ldots\ldots + a_r \sum_{1}^{\alpha} x_i^r \\ \sum_{\alpha+1}^{\beta} y_i &= a_0 (\beta - \alpha) + a_1 \sum_{\alpha+1}^{\beta} x_i + \ldots\ldots + a_r \sum_{\alpha+1}^{\beta} x_i^r \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ \sum_{\lambda+1}^{n} y_i &= a_0 (n - \lambda) + a_1 \sum_{\lambda+1}^{n} x_i + \ldots\ldots + a_r \sum_{\lambda+1}^{n} x_i^r \end{aligned}\right\} \qquad (80)$$

È questo il cosidetto metodo delle somme, che, assicurando l'uguaglianza tra la somma dei dati osservati e quella dei dati calcolati in ognuno degli intervalli scelti, assicura altresì l'uguaglianza tra la somma di tutti i dati osservati e quella di tutti i dati calcolati, e quindi rende pure nulla la somma algebrica degli scarti tra i primi e i secondi (1).

---

(1) Sembra che tale metodo sia stato proposto per la prima volta da Tobia Mayer nel 1748. Cfr. E. T. Whittaker and G. Robinson, *The calculus of observations*. London, Blackie, 2ª ediz., 1926, (pag. 259). Nelle (80) il determinante dei coefficienti risulta in generale diverso da zero.

Si noti che i risultati di questo metodo non sono sensibili alle variazioni degli y_i entro ogni classe, in cui siano stati raggruppati. È questa proprietà, che ne rende particolarmente utile l'impiego in quei casi, in cui s'ignori la ripartizione degli y_i compresi in ogni classe o tale ripartizione sia inattendibile o si ritenga trascurabile ai fini che si perseguono.

Ad es., è stato affermato che le distribuzioni per età dei censiti in un dato istante e dei morti in un dato intervallo di tempo si possono rappresentare per brevi tratti (non superiori a un quindicennio e pei morti solo a cominciare dall'età di 5 anni) con una parabola quadratica (1).

Ora, avendo motivi per ritenere che i dati dei morti in Italia nel 1926 dai 68 agli 82 anni compiuti — a causa degli arrotondamenti, da cui sono affetti, specie in queste età avanzate — sono erroneamente addensati attorno alle età terminanti per 0 e per 5, possiamo considerare più attendibile la distribuzione in gruppi quinquennali, data nella terza colonna della tavola L.

TAVOLA L.

Età	Numero dei morti nel 1926 y_i	Σy_i
68	9.900	
69	9.679	
70	12.716	53.833
71	10.030	
72	11.508	
73	11.736	
74	12.534	
75	13.770	61.839
76	12.785	
77	11.014	
78	11.396	
79	9.758	
80	11.941	50.628
81	8.753	
82	8.780	

(1) Cfr. il nostro lavoro: *Sulla costruzione delle tavole di mortalità*, in « Giornale di matematica finanziaria », gennaio - giugno 1925; e gli studî di KING e di BAGNI, ivi discussi.

Sostituendo, per semplicità, alle età la successione dei numeri interi da — 7 a + 7, avremo:

$$\sum_{1}^{5} x_i = -25 \qquad \sum_{1}^{5} x_i^2 = 135$$

$$\sum_{6}^{10} x_i = 0 \qquad \sum_{6}^{10} x_i^2 = 10$$

$$\sum_{11}^{15} x_i = 25 \qquad \sum_{11}^{15} x_i^2 = 135$$

Risolvendo il sistema:

$$53.833 = a_0\ 5 - a_1\ 25 + a_2\ 135$$
$$61.839 = a_0\ 5 \qquad + a_2\ 10$$
$$50.628 = a_0\ 5 + a_1\ 25 + a_2\ 135$$

si ha:

$$a_0 = 12.521{,}664 \qquad a_1 = -64{,}1 \qquad a_2 = -76{,}868$$

e quindi:

$$f(x) = 12.521{,}664 - 64{,}1\ x - 76{,}868\ x^2 \quad .$$

Calcolando, per $x = -7, -6, ... +6, +7$, i dati teorici risultanti dalla funzione precedente, e ponendoli a confronto con quelli osservati, si ottengono i numeri della terza colonna della tavola seguente:

TAVOLA LI.

Età	y_i	$f(x_i)$	$\varepsilon_i = y_i - f(x_i)$
68	9.900	9.204	+ 696
69	9.679	10.139	— 460
70	12.716	10.920	+ 1.796
71	10.030	11.548	— 1.518
72	11.508	12.022	— 514
73	11.736	12.342	— 606
74	12.534	12.509	+ 25
75	13.770	12.522	+ 1.248
76	12.785	12.381	+ 404
77	11.014	12.086	— 1.072
78	11.396	11.638	— 242
79	9.758	11.035	— 1.277
80	11.941	10.279	+ 1.662
81	8.753	9.370	— 617
82	8.780	8.306	+ 474

Sull'ultima colonna può controllarsi che la somma degli scarti algebrici è nulla, sia in totale, sia in ognuna delle classi quinquennali, in cui la distribuzione è stata raccolta.

L'andamento degli scarti (vedasi anche la figura 45) mette in luce l'addensamento dei morti attorno alle età rotonde a spese dei gruppi laterali, e conferma a primo aspetto l'ipotesi che la distribuzione dei morti alle età considerate, se non fosse stata affetta dai detti errori, avrebbe presentato una forma non molto diversa da quella descritta dalla nostra parabola quadratica.

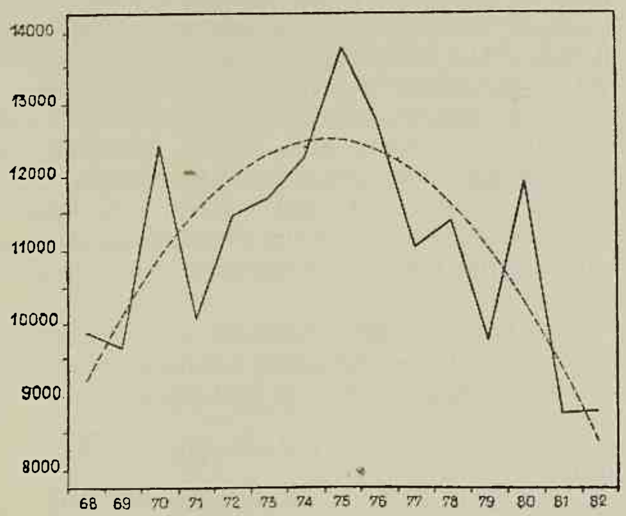


Fig. 45

Si è detto che il metodo esposto — quando le classi non siano imposte dal modo di rilevazione dei dati in relazione al numero dei coefficienti della funzione scelta — presenta l'inconveniente dell'arbitrarietà della scelta degl'intervalli. Ciò in generale è vero, anzi appunto per questo a tale vecchio metodo di adattamento si finì col preferire sistematicamente quello dei minimi quadrati; ma in alcuni casi — come nel nostro esempio — la libertà di scelta è un vantaggio, in quanto rende possibile l'assunzione degl'intervalli, che assicurano la compensazione degli errori sistematici degli y_i.

Le (80) sogliono porsi sotto diverse forme, anche in relazione al tipo di rappresentazione numerica dei dati da elaborare. Ad es., nel caso di distribuzioni statistiche, i primi membri delle (80), poste nella forma:

$$\left.\begin{array}{l} \sum_1^{\alpha} y_i = a_0\,\alpha + a_1 \sum_1^{\alpha} x_i + \ldots\ldots + a_r \sum_1^{\alpha} x_i^r \\ \sum_1^{\beta} y_i = a_0\,\beta + a_1 \sum_1^{\beta} x_i + \ldots\ldots + a_r \sum_1^{\beta} x_i^r \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \sum_1^{n} y_i = a_0\,n + a_1 \sum x_i + \ldots\ldots + a_r \sum_1^{n} x_i^r \end{array}\right\} \qquad (80\,bis)$$

non sono che osservazioni accumulate, nel senso spiegato al n. 1, *e* del primo capitolo.

S'intende, poi, facilmente che, se interessa che la curva da adattare — compatibilmente col numero di condizioni occorrenti — passi inoltre per punti assegnati, nel sistema precedente possono anche esser comprese equazioni tra valori singoli, ossia alcune classi possono restringersi in guisa da comprendere solo una osservazione.

Si avverte, infine, che, poichè nell'esempio scelto la variabile considerata (età) è suscettibile di variare con continuità, nei secondi membri del sistema precedente avremmo potuto assumere somme di valori varianti per quantità infinitesime e precisamente i cosidetti integrali definiti. Non lo abbiamo fatto, per non presupporre sin d'ora nozioni poco familiari alla massa dei lettori: diciamo solo che in tal guisa il metodo esposto presenta particolari vantaggi e va inteso col nome di metodo delle aree, in quanto assicura l'uguaglianza tra l'area corrispondente alla curva teorica e quella corrispondente alla curva empirica (1).

8. Metodo delle differenze. - Assumendo sempre la funzione (60) e ricordando le proprietà delle differenze successive di essa, vien naturale di porre, come condizione di adattamento:

$$\Sigma \Delta^k [f(x_i) - y_i] = 0 \qquad (k = 0, \ 1, \ \ldots\ldots \ r)$$

ossia:

$$\Sigma [\Delta^k f(x_i) - \Delta^k y_i] = 0$$

od anche, essendo: $\Delta^0 f(x) = f(x)$,

$$\left.\begin{array}{l} \Sigma f(x_i) = \Sigma y_i \\ \Sigma \Delta \, f(x_i) = \Sigma \Delta \, y_i \\ \Sigma \Delta^2 f(x_i) = \Sigma \Delta^2 y_i \\ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \\ \Sigma \Delta^r f(x_i) = \Sigma \Delta^r y_i \end{array}\right\} \qquad (81)$$

Infatti, è questo un sistema lineare di $r + 1$ equazioni ad $r + 1$ incognite, che sono appunto i coefficienti della funzione di grado r (2).

Il significato di tale sistema è evidente: mentre la condizione di ugua-

---

(1) Francesco Paolo Cantelli, *Sull'adattamento di curve ad una serie di misure o di osservazioni*. Roma, Bolognesi, 1905. Ne ha fatte applicazioni Giorgio Mortara, *Lezioni di statistica metodologica*. Città di Castello, 1922 (pag. 225). Per mantenerci nel campo dei metodi elementari, nelle analisi precedenti abbiamo supposto che in un istogramma sia sempre lecito attribuire alle osservazioni di ogni classe un solo valore.

(2) Felice Vinci, *Un nuovo metodo d'interpolazione: il metodo delle differenze*. Bari, Cressati, 1924. Nelle (81) il determinante dei coefficienti risulta in generale diverso da zero.

glianza della somma degli y_i alla somma dei valori corrispondenti della funzione, col metodo delle somme, si estende a gruppi parziali dei valori medesimi e, coi metodi dei minimi quadrati e di Cauchy, si estende alle altre somme che si ottengono, applicando alle prime dei moltiplicatori; col metodo ora esposto la condizione di uguaglianza si estende alle somme delle differenze non nulle della funzione, venendo in certa guisa a rispettare, per così dire, la fisionomia parabolica dei dati osservati. In esso è eliminata la necessità della scelta di un moltiplicatore o del raggruppamento in classi dei dati osservati, e può ottenersi un adattamento non molto diverso che coi metodi precedenti, specie quando l'andamento dei dati estremi non si allontani molto da quello della funzione scelta.

Se gli x_i formano una progressione aritmetica, si ha sempre:

$$\Delta^{k+1} x_i^k = 0$$

ed il sistema precedente si riduce a:

$$\left.\begin{array}{r} a_0 n + a_1 \Sigma x_i + a_2 \Sigma x_i^2 + \ldots + a_r \Sigma x_i^r = \Sigma y_i \\ a_1 \Sigma \Delta x_i + a_2 \Sigma \Delta x_i^2 + \ldots + a_r \Sigma \Delta x_i^r = \Sigma \Delta y_i \\ a_2 \Sigma \Delta^2 x_i^2 + \ldots + a_r \Sigma \Delta^2 x_i^r = \Sigma \Delta^2 y_i \\ \ldots\ldots\ldots\ldots\ldots \\ a_r \Sigma \Delta^r x_i^r = \Sigma \Delta^r y_i \end{array}\right\} \quad (81bis)$$

Pertanto, calcolate le somme che compaiono nel sistema, si potrà procedere rapidamente al calcolo dei coefficienti, a cominciare da a_r.

Le somme della prima equazione potranno ottenersi rapidamente, a mezzo delle (67), e quelle delle successive equazioni dalla formula:

$$\Sigma \Delta^k \alpha_i = (\alpha_n \pm \alpha_1) - (k-1)(\alpha_{n-1} \pm \alpha_2) + \binom{k-1}{2}(\alpha_{n-2} \pm \alpha_3) - \ldots \mp (\alpha_{n-k+1} \pm \alpha_k)$$

in cui dovrà adottarsi il segno posto in alto o in basso, secondochè k sia rispettivamente pari o dispari.

Da essa si trae in particolare:

$$\left.\begin{array}{l} \Sigma \Delta \alpha_i = \alpha_n - \alpha_1 \\ \Sigma \Delta^2 \alpha_i = (\alpha_n + \alpha_1) - (\alpha_{n-1} + \alpha_2) \\ \Sigma \Delta^3 \alpha_i = (\alpha_n - \alpha_1) - 2(\alpha_{n-1} - \alpha_2) + (\alpha_{n-2} - \alpha_3) \\ \Sigma \Delta^4 \alpha_i = (\alpha_n + \alpha_1) - 3(\alpha_{n-1} + \alpha_2) + 3(\alpha_{n-2} + \alpha_3) - (\alpha_{n-3} + \alpha_4) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{array}\right\} \quad (82)$$

Queste formule valgono, qualunque sia la legge di successione degli elementi considerati, e quindi possono fornire le somme delle differenze di y_i e di x_i^s. Ma, riguardo a questi ultimi, se gli x_i sono in progressione aritmetica, possiamo scrivere:

$$\begin{aligned}\Sigma \Delta^k x_i^s = & \left\{ \left[x_1 + (n-1)\,h\right]^s \pm x_1^s \right\} \\ & - (k-1) \left\{ \left[x_1 + (n-2)\,h\right]^s \pm (x_1 + h)^s \right\} \\ & + \binom{k-1}{2} \left\{ \left[x_1 + (n-3)\,h\right]^s \pm (x_1 + 2\,h)^s \right\} \\ & \cdots\cdots\cdots\cdots\cdots\cdots \\ & \mp \left\{ \left[x_1 + (n-k)\,h\right]^s \pm \left[x_1 + (k-1)\,h\right]^s \right\}\end{aligned}$$

donde:

$$\left.\begin{aligned} & \Sigma \Delta\, x_i = (n-1)\,h \\ & \Sigma \Delta\, x_i^2 = (n-1)\,h\,[2\,x_1 + (n-1)\,h] \\ & \Sigma \Delta^2 x_i^2 = 2\,(n-2)\,h^2 \\ & \Sigma \Delta\, x_i^3 = (n-1)\,h\,[3\,x_1^2 + 3\,x_1\,(n-1)\,h + (n-1)^2\,h^2] \\ & \Sigma \Delta^2 x_i^3 = 3\,(n-2)\,h^2\,[2\,x_1 + (n-1)\,h] \\ & \Sigma \Delta^3 x_i^3 = 6\,(n-3)\,h^3 \\ & \cdots\cdots\cdots\cdots\cdots\cdots \end{aligned}\right| \qquad (83)$$

Se, poi, gli x_i sono uguali ai primi n numeri della serie naturale a cominciare da 0, chiamando m l'ultimo termine della serie, le formule precedenti porgono:

$$\left.\begin{aligned} & n = m + 1 & & \Sigma x_i^3 = \frac{m^2\,(m+1)^2}{4} \\ & \Sigma x_i = \frac{m\,(m+1)}{2} & & \Sigma \Delta\, x_i^3 = m^3 \\ & \Sigma \Delta\, x_i = m & & \Sigma \Delta^2 x_i^3 = 3\,m\,(m-1) \\ & \Sigma x_i^2 = \frac{m\,(m+1)\,(2\,m+1)}{6} & & \Sigma \Delta^3 x_i^3 = 6\,(m-2) \\ & \Sigma \Delta\, x_i^2 - m^2 & & \cdots\cdots\cdots\cdots \\ & \Sigma \Delta^2 x_i^2 = 2\,(m-1) & & \cdots\cdots\cdots\cdots \end{aligned}\right| \qquad (83\ bis)$$

ed il sistema (81 *bis*), per $f(x) = a_0$, fornirà:

$$a_0 = \frac{1}{m+1} \Sigma y_i ;$$

per $f(x) = a_0 + a_1 x$:

$$\left.\begin{aligned} a_1 &= \frac{1}{m} \Sigma \Delta y_i \\ a_0 &= \frac{1}{m+1} \Sigma y_i - \frac{1}{2} \Sigma \Delta y_i ; \end{aligned}\right\} \qquad (84)$$

per $f(x) = a_0 + a_1 x + a_2 x^2$:

$$\left.\begin{aligned} a_2 &= \frac{1}{2(m-1)} \Sigma \Delta^2 y_i \\ a_1 &= \frac{1}{m} \Sigma \Delta y_i - \frac{m}{2(m-1)} \Sigma \Delta^2 y_i \\ a_0 &= \frac{1}{m+1} \Sigma y_i - \frac{1}{2} \Sigma \Delta y_i + \frac{m}{12} \Sigma \Delta^2 y_i \end{aligned}\right\} \qquad (85)$$

e così via.

Procedendo in tal guisa, le formule per il calcolo dei coefficienti contengono sempre tutti i termini delle corrispondenti formule precedenti, e quindi, ove non sia stabilito il grado della funzione e si voglia procedere per approssimazioni successive, non occorrerà in ogni stadio rifare i calcoli da capo, ma si potranno utilizzare completamente quelli già eseguiti negli stadî precedenti.

In un dato stadio si potrebbero anche utilizzare gli scarti, che si fossero già calcolati tra gli y_i ed i dati corrispondenti desunti dalla funzione.

Si noti che, se le differenze tra i successivi y_i non presentano variazioni di segno, od almeno se i valori estremi di y_i sono anche estremi nella graduatoria crescente di essi, l'uguaglianza tra le somme empiriche e quelle teoriche delle differenze prime assicura altresì l'uguaglianza tra il campo di variabilità dei dati calcolati e di quelli osservati.

Se gli x_i non formassero una progressione aritmetica, le differenze singole non presenterebbero le proprietà sommariamente ricordate in principio di questo numero; ma il metodo non perderebbe valore, non essendo legato a coteste proprietà.

Si voglia, ad es., adattare una retta tra i punti di coordinate (2 ; 8), (4 ; 7), (6 ; 15). Escludendo dai primi membri del sistema (81 *bis*) i termini successivi a quelli in x_i ed applicando le (82), avremo immediatamente:

$n = 3 \qquad \Sigma x_i = 2 + 4 + 6 = 12 \qquad \Sigma y_i = 8 + 7 + 15 = 30$

$\qquad \Sigma \Delta x_i = 6 - 2 = 4 \qquad \Sigma \Delta y_i = 15 - 8 = 7$

e quindi:

$$a_1 = \frac{7}{4} = 1{,}75$$

$$a_0 = \frac{30}{3} - 1{,}75 \cdot \frac{12}{3} = 3.$$

La funzione: $f(x) = 3 + 1{,}75\,x$ assicurerà l'uguaglianza delle somme dei valori e delle differenze prime degli y_i con quelle corrispondenti teoriche.

Per adattare una funzione parabolica di secondo grado ai dati annui dei matrimonî contratti in Italia nel ventennio dal 1908 al 1927, sostituendo agli anni gl'interi da 0 a 19, trovano applicazione le (85).

TAVOLA LII.

Anni	Migliaia di matrimonî	Anni	Migliaia di matrimonî
1908	283	1918	106
1909	266	1919	333
1910	269	1920	509
1911	260	1921	426
1912	265	1922	351
1913	264	1923	321
1914	252	1924	307
1915	186	1925	296
1916	106	1926	296
1917	97	1927	303

Si ha: $\Sigma y_i = 5.496$; e, in forza delle (82):

$$\Sigma \Delta\, y_i = 303 - 283 = 20$$

$$\Sigma \Delta^2 y_i = (303 + 283) - (296 + 266) = 24$$

donde:

$$a_2 = \frac{1}{36} \times 24 = \frac{2}{3}$$

$$a_1 = \frac{1}{19} \times 20 - \frac{19}{36} \times 24 = -11{,}614$$

$$a_0 = \frac{1}{20} \times 5.496 - \frac{1}{2}\,20 + \frac{19}{12} \times 24 = 302{,}8$$

$$f(x) = 302{,}8 - 11{,}614\,x + 0{,}667\,x^2.$$

Facendo assumere ad x i valori interi da 0 a 19, la funzione precedente fornisce i dati rappresentati a tratteggio nella figura 46, obbedienti alle condizioni poste.

Sebbene gli scarti raggiungano un'alta misura, pure — tenendo presente che il numero dei matrimonî contratti in Italia dal 1915 al 1918 subì una forte diminuzione a causa della guerra, e dopo il 1918 si ebbe un gran numero di matrimonî ritardati, che a un dipresso colmò il deficit precedente — la parabola descritta potrebbe assumersi *grosso modo* come rappresentativa della forma, che avrebbe presentato la curva dei matrimoni in quel ventennio, se non fosse intervenuta la guerra, per l'influenza combinata dell'aumento della popolazione e di tutti quegli altri fattori, che influiscono sul numero assoluto dei matrimonî.

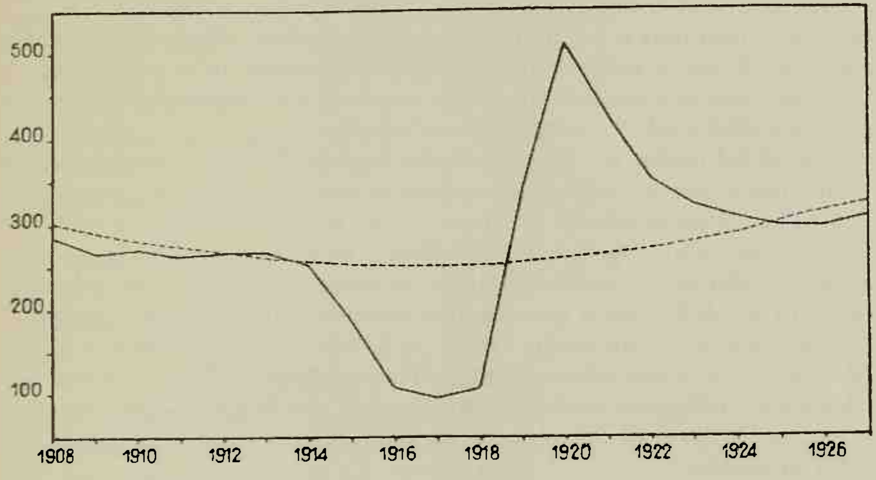


Fig. 46

Risultati ancora migliori avrebbero potuto forse ottenersi da una funzione parabolica di grado superiore, od anche combinando opportunamente questo metodo con quello delle somme: ad es. adattando una parabola quadratica obbediente alle tre condizioni che nel primo settennio e nel successivo tredicennio la somma dei dati osservati risulti uguale a quella dei dati teorici, e che inoltre la somma delle differenze prime empiriche sia uguale a quella delle differenze prime teoriche.

9. La perequazione e l'interpolazione. La cosidetta extrapolazione. – La rappresentazione analitica di una distribuzione o serie statistica a due variabili, oltre a fornire una nozione sintetica delle relazioni tra le variabili considerate, permette:

a) la perequazione (graduazione o aggiustamento) degli y_i, quando essi presentino variazioni inattendibili o trascurabili ai fini che si perseguono;

b) l'interpolazione di nuovi dati tra gli y_i.

Le figure 45 e 46 danno esempî evidenti di perequazione; dalla funzione lineare ottenuta al n. 5 possiamo inoltre ricavare, per $x = 1919.25$, ecc., l'ammontare dei risparmî alla fine del primo trimestre del 1920, ecc. mentre la funzione del n. 7, per $x = -6,5$ ecc. fornisce i morti in età da 68,5 a 69,5 anni, e così via.

Per la perequazione e l'interpolazione dei dati spesso non è necessario l'adattamento di una funzione idonea a rappresentare tutta la distribuzione o la serie statistica — i nostri esempî, ad eccezione di quello della figura 43, considerano appunto piccoli tratti di esse —, bastando una funzione, che possa rappresentarne una parte o parti successive.

Inoltre, se si tratta di una curva di osservazioni accumulate (n. 1, *e* del primo capitolo) e per un tratto di essa si vuole adattare ad *s* punti assegnati la curva rappresentatrice di una funzione ad *s* coefficienti, il problema dell'adattamento di essa si riduce al noto problema di far passare una curva ad *s* parametri per *s* punti.

Ma in tal modo, se ad es. abbiamo ragione di ritenere che i quattro punti della curva delle osservazioni accumulate (fig. 34), corrispondenti al numero di nascite da donne in età maggiore o uguale a 20, 25, 30, 35 anni siano pienamente attendibili e quel tratto possa essere bene rappresentato da un parabola cubica, adattando tale parabola possiamo ottenere per differenza il numero di osservazioni di ogni classe quinquennale ed anche — assumendo valido quell'andamento parabolico — i dati ben perequati corrispondenti a tutti gli intervalli annuali di x compresi tra 20 e 35 anni. Inoltre possiamo calcolare con precisione il valore centrale e prevalente (n. 9 del precedente capitolo), che sappiamo cadono in codesto intervallo.

I valori della funzione potrebbero ricavarsi direttamente dalla formula seguente (Lagrange, 1795):

$$\begin{aligned} f(x) = {} & y_0 \frac{(x - x_1)(x - x_2)\ldots\ldots(x - x_n)}{(x_0 - x_1)(x_0 - x_2)\ldots\ldots(x_0 - x_n)} \\ & + y_1 \frac{(x - x_0)(x - x_2)\ldots\ldots(x - x_n)}{(x_1 - x_0)(x_1 - x_2)\ldots\ldots(x_1 - x_n)} \\ & + y_2 \frac{(x - x_0)(x - x_1)\ldots\ldots(x - x_n)}{(x_2 - x_0)(x_2 - x_1)\ldots\ldots(x_2 - x_n)} \qquad (86) \\ & \ldots\ldots\ldots\ldots\ldots\ldots \\ & + y_n \frac{(x - x_0)(x - x_1)\ldots\ldots(x - x_{n-1})}{(x_n - x_0)(x_n - x_1)\ldots\ldots(x_n - x_{n-1})} \end{aligned}$$

che, per una funzione parabolica di r. esimo grado, risulta limitata all'$(r+1)^o$ termine $(r = n)$ e dove per $x = x_0$ si annullano tutti i termini ad eccezione del primo che risulta uguale ad y_0, per $x = x_1$ si annullano tutti i termini ad eccezione del secondo che risulta uguale ad y_1, e così via. Per qualsiasi valore intermedio di x, essa fornisce quello corrispondente di y obbediente all'andamento parabolico assegnato.

Però, essendo nel nostro caso costanti gl'intervalli di x, è preferibile usare la formula seguente (Newton-Gregory, 1670):

$$\begin{aligned} f(x) = y_0 &+ \frac{x - x_0}{h} \Delta y_0 + \frac{x - x_0}{h} \frac{x - x_0 - h}{2h} \Delta^2 y_0 \\ &+ \frac{x - x_0}{h} \frac{x - x_0 - h}{2h} \frac{x - x_0 - 2h}{3h} \Delta^3 y_0 \\ &\dots\dots\dots\dots\dots\dots \\ &+ \frac{x - x_0}{h} \frac{x - x_0 - h}{2h} \dots\dots \frac{x - x_0 - (n-1)h}{nh} \Delta^n y_0 \,, \end{aligned} \tag{87}$$

che, in base alle nozioni sulle differenze, date al n. 2, per una funzione parabolica di r. esimo grado risulta pure limitata all' $(r+1)^o$ termine, ossia al termine in $\Delta^r y_0$ ed inoltre:

per	$x = x_0$	dà	$f(x_0) = y_0$
»	$x = x_0 + h = x_1$	»	$f(x_1) = y_0 + \Delta y_0 = y_1$
»	$x = x_0 + 2h = x_2$	»	$f(x_2) = y_0 + 2\Delta y_0 + \Delta^2 y_0 = y_2$
. . . . .	. . . . .	. . . . .	. . . . .

e per qualsiasi valore intermedio di x fornisce quello corrispondente di y obbediente all'andamento parabolico assegnato (¹).

Nel nostro esempio si ha (tavola IX):

$x_0 \geq 20$	$y_0 = 39.501$	$\Delta y_0 = -11.472$	$\Delta^2 y_0 = -1.013$	$\Delta^3 y_0 = 4.454$
$x_1 \geq 25$	$y_1 = 28.029$	$\Delta y_1 = -12.485$	$\Delta^2 y_1 = 3.441$	
$x_2 \geq 30$	$y_2 = 15.544$	$\Delta y_2 = -9.044$		
$x_3 \geq 35$	$y_3 = 6.500$			

e quindi:

$$f(x) = 39.501 - \frac{x - 20}{5} 11.472 - \frac{x - 20}{5} \frac{x - 25}{10} 1.013 + \frac{x - 20}{5} \frac{x - 25}{10} \frac{x - 30}{15} 4.454.$$

Attribuendo ad x qualsiasi valore intermedio tra 20 e 35, si avrebbe rapidamente il numero degl'individui di età maggiore o uguale ad x e, per differenza, il numero degl'individui di età compresa tra x e $x + i$.

(¹) Su altre formule simili e sulle cosidette differenze centrali e differenze divise, cfr. E. T. Whittaker and G. Robinson, op. cit., Cap. I-IV; e G. Cassinis, op. cit., Cap. V.

Inoltre, chiamando $2\,y_c - 1$ il numero totale delle osservazioni (41.830), ad y_c corrisponderà l'età centrale x_c, che potrà dedursi dall'equazione precedente, ponendo $f(x) = 20.915{,}5$ (n. 9 del capitolo precedente).

Assumendo, però, più semplicemente — come suol farsi — l'ipotesi di un andamento rettilineo tra 25 e 30 anni e sostituendo nella (87) x_1 a x_0 e y_1 a y_0 in modo da poter mantenere le notazioni dell'esempio, avremo:

$$y_c = y_1 + \frac{x_c - x_1}{h}(y_2 - y_1)$$

donde:

$$x_c = x_1 + \frac{y_c - y_1}{y_2 - y_1}\,h,$$

che è la formula d'interpolazione per parti proporzionali, di cui si fa largo uso nelle tavole numeriche.

Nel nostro esempio si ottiene:

$$x_c = 25 + \frac{20.915{,}5 - 28.029}{15.544 - 28.029} \times 5 = 27{,}84$$

risultato che possiamo ritenere leggermente errato per eccesso, data la leggera concavità verso l'alto che dovrebbe presentare la curva nel tratto considerato.

Il valore prevalente (n. 9 del precedente capitolo) sarà quel valore, a cui corrisponde il punto d'inflessione della parabola cubica, valore che si dimostra [1] uguale a:

$$x_p = x_0 + h - h\,\frac{\Delta^2\,y_0}{\Delta^3\,y_0}.$$

Risulta:

$$x_p = 25 + \frac{5 \times 1.013}{4.454} = 26{,}137.$$

Ricordiamo che, con procedimento grafico, avevamo già ottenuto sulla fig. 34:

$$x_c = 27{,}93 \qquad x_p = 25{,}95.$$

Se l'adattamento di archi parabolici si esegue ripetutamente su tratti consecutivi della curva osservata, con opportuni avvedimenti si possono ottenere buone perequazioni e interpolazioni di lunghi tratti di essa.

A tal fine possono seguirsi parecchie vie, alle quali sono legati i nomi

---

[1] Uguagliando a zero la derivata seconda della formula di Newton-Gregory, limitata ai primi quattro termini, si ha:

$$\frac{d^2 f(x)}{d\,x^2} = \frac{1}{h^2}\Delta^2\,y_0 + \frac{x - x_0 - h}{h^3}\,\Delta^3\,y_0 = 0,$$

donde la formula del testo.

di Sprague, King, Karup (1): ci limiteremo ad esporre quella elementare e più aderente ai dati, da noi tracciata nel 1925 (2).

Abbiamo affermato al n. 7 che, distribuendo secondo classi annuali di età i viventi in un dato istante od i morti in un dato intervallo di tempo, quindici classi consecutive possono essere bene rappresentate da una parabola quadratica (almeno a prescindere dalle classi di morti relative ai primissimi anni di età). Possiamo anche dire che, sostituendo al profilo dei tre rettangoli corrispondenti a tre classi quinquennali consecutive di viventi o di morti un profilo parabolico ordinario, questo rispecchia bene l'andamento generale dei quindici dati osservati (ignoti o mal noti) compresi in quei tre quinquennî.

Fondandoci su cotesta nozione, chiamiamo con s_x la prima classe quinquennale riferita per semplicità alle cinque età intere da x a $x+4$. Se in base alla terna s_x, s_{x+5}, s_{x+10} tracciamo una parabola quadratica obbediente alle condizioni che in ognuno dei tre intervalli considerati le somme delle cinque ordinate teoriche siano uguali a quelle osservate, veniamo a determinare, col metodo delle somme, la classe annuale centrale u_{x+7}.

Ponendo uguale a zero cotesta età centrale ed all'unità l'intervallo quinquennale, dalla funzione parabolica:

$$f(\zeta) = a_0 + a_1 \zeta + a_2 \zeta^2$$

si ricava per il primo quinquennio:

$$f\left(-\frac{7}{5}\right) = a_0 - a_1 \frac{7}{5} + a_2 \frac{49}{25}$$

$$f\left(-\frac{6}{5}\right) = a_0 - a_1 \frac{6}{5} + a_2 \frac{36}{25}$$

$$f\left(-\frac{5}{5}\right) = a_0 - a_1 \frac{5}{5} + a_2 \frac{25}{25}$$

$$f\left(-\frac{4}{5}\right) = a_0 - a_1 \frac{4}{5} + a_2 \frac{16}{25}$$

$$f\left(-\frac{3}{5}\right) = a_0 - a_1 \frac{3}{5} + a_2 \frac{9}{25}$$

$$\sum_{-\frac{7}{5}}^{-\frac{3}{5}} f(\zeta) = 5\,a_0 - 5\,a_1 + \frac{27}{5}\,a_2$$

(1) Sono illustrate nei trattati di matematica attuariale, col nome di perequazioni o interpolazioni osculatrici. Cfr. pure E. Blaschke, *Vorlesungen über mathematische Statistik*. Leipzig, Teubner, 1906, pag. 231 e segg.

(2) Felice Vinci, *Sulla costruzione delle tavole di mortalità*, op. cit.; e *Per l'uniformità nei metodi di costruzione delle tavole di mortalità*, in « Bulletin de l'Institut Intern. de Statistique », Rome, 1925.

e per gli altri due quinquenni:

$$\sum_{-\frac{2}{5}}^{\frac{2}{5}} f(\zeta) = 5\,a_0 + \frac{2}{5}\,a_2\,,$$

$$\sum_{\frac{3}{5}}^{\frac{7}{5}} f(\zeta) = 5\,a_0 + 5\,a_1 + \frac{27}{5}\,a_2$$

Pertanto dal sistema:

$$\begin{aligned} 5\,a_0 - 5\,a_1 + \frac{27}{5}\,a_2 &= s_x \\ 5\,a_0 \qquad\quad + \frac{2}{5}\,a_2 &= s_{x+5} \\ 5\,a_0 + 5\,a_1 + \frac{27}{5}\,a_2 &= s_{x+10} \end{aligned}$$

si ottengono i valori di a_0, a_1, a_2, della funzione assunta e da essa le quindici classi annuali perequate, soddisfacenti alle condizioni poste.

Poichè, però, a noi interessa conoscere solo la classe annuale centrale, che è data da:

$$f(0) = a_0$$

possiamo risparmiare buona parte dei calcoli. Ricavando dal sistema precedente la formula di a_0, si ottiene:

$$\begin{aligned} u_{x+7} = a_0 &= \frac{1}{125}\left(-s_x + 27\,s_{x+5} - s_{x+10}\right) \\ &= 0{,}224\,s_{x+5} - 0{,}008\,(s_x + s_{x+5} + s_{x+10}) \end{aligned} \tag{88}$$

Applicando questa formula a terne consecutive di classi quinquennali, si ottengono le classi annuali perequate dei viventi o dei morti alle età centrali di ogni classe quinquennale, ad esclusione della prima e dell'ultima classe.

Ciò posto, notiamo ancora che, ricavando dal precedente sistema la formula:

$$a_1 = \frac{1}{10}\,(s_{x+10} - s_x)$$

essa fornisce la pendenza u'_{x+7} della parabola in corrispondenza alla ordinata centrale (1).

(1) È la derivata di $f(\zeta)$ nel punto di ascissa 0.

Se adesso, in base a due classi quinquennali s_{x+5}, s_{x+10}, tracciamo una parabola del quinto ordine, obbediente alle sei condizioni che in ognuno dei due quinquennî considerati la somma delle cinque ordinate teoriche sia uguale a quella delle cinque ordinate empiriche, e che negli anni centrali di tali quinquennî si abbiano ordinate e pendenze uguali rispettivamente a quelle già calcolate, riusciremo a determinare le ordinate annuali intermedie a quelle centrali con tutti gli elementi, di cui possiamo disporre.

Ponendo uguali a zero ed all'unità l'anno centrale del primo e rispettivamente del secondo quinquennio, dalla funzione:

$$\varphi(\zeta) = A_0 + A_1 \zeta + A_2 \zeta^2 + A_3 \zeta^3 + A_4 \zeta^4 + A_5 \zeta^5$$

si ricava il sistema:

$$\begin{aligned}
\varphi\,(0) &= A_0 = u_{x+7} \\
\varphi'\,(0) &= A_1 = u'_{x+7} \\
\varphi\,(1) &= A_0 + A_1 + A_2 + A_3 + A_4 + A_5 = u_{x+12} \\
\varphi'\,(1) &= A_1 + 2\,A_2 + 3\,A_3 + 4\,A_4 + 5\,A_5 = u'_{x+12}
\end{aligned}$$

$$\sum_{-\frac{2}{5}}^{\frac{2}{5}} \varphi(\zeta) = 5\,A_0 + \frac{2}{5}\,A_2 + \frac{34}{625}\,A_4 = s_{x+5}$$

$$\sum_{\frac{3}{5}}^{\frac{7}{5}} \varphi(\zeta) = 5\,A_0 + 5\,A_1 + \frac{27}{5}\,A_2 + \frac{31}{5}\,A_3 + \frac{4.659}{625}\,A_4 + \frac{1.159}{125}\,A_5 = s_{x+10}$$

donde, in definitiva:

$$\left.\begin{aligned}
u_{x+7} &= \frac{1}{125}\left(-\,s_x + 27\,s_{x+5} - s_{x+10}\right) \\
u_{x+8} &= \frac{1}{7.765.625}\left(-\,163.868\,s_x + 1.547.729\,s_{x+5} + 191.771\,s_{x+10} - 22.507\,s_{x+15}\right) \\
u_{x+9} &= \frac{1}{7.765.625}\left(-\,158.685\,s_x + 1.159.430\,s_{x+5} + 642.195\,s_{x+10} - 89.815\,s_{x+15}\right) \\
u_{x+10} &= \frac{1}{7.765.625}\left(-\,89.815\,s_x + 642.195\,s_{x+5} + 1.159.430\,s_{x+10} - 158.685\,s_{x+15}\right) \\
u_{x+11} &= \frac{1}{7.765.625}\left(-\,22.507\,s_x + 191.771\,s_{x+5} + 1.547.729\,s_{x+10} - 163.868\,s_{x+15}\right)
\end{aligned}\right\} \quad (89)$$

In base alle formule precedenti considerando quaterne successive di classi quinquennali, si riusciranno a determinare le classi annuali, ad esclusione delle prime sette e delle ultime sette.

Un vantaggio di tali formule sta nel fatto che, pur essendo strettamente vincolate ai dati osservati, esse non costringono i dati teorici ad uguagliarsi a questi ultimi proprio nei quinquenni, in cui questi sono stati raccolti e che risentono sempre le irregolarità d'andamento dei dati annuali (1).

Inoltre, stante la simmetria dei coefficienti, le operazioni si potranno disporre in guisa da ridurre a metà il numero delle moltiplicazioni da eseguire.

Ad es., in base alle seguenti quattro classi fornite dalla distribuzione per età della popolazione italiana nel 1921:

Età	25 – 29	30 – 34	35 – 39	40 – 44
Numero di viventi	2.860.806	2.629.418	2.352.108	2.142.963

le formule precedenti porgono:

Età	32	33	34	35	36
Numero di viventi	526.251	515.561	503.849	491.743	480.209

In base alle altre quattro classi:

Età	30 – 34	35 – 39	40 – 44	45 – 49
Numero di viventi	2.629.418	2.352.108	2.142.963	1.949.849

si ottiene:

Età	37	38	39	40	41
Numero di viventi	469.876	460.861	452.111	444.208	436.423

Questo metodo può evidentemente applicarsi al caso di gruppi diversamente comprensivi ed a parabole di diverso ordine; ciò che condurrebbe a formule diverse.

Prescindendo dalle condizioni delle pendenze, si ottiene forse una maggiore rapidità di calcoli; ma non si assicura il collegamento soddisfacente dei dati ricavati dagli archi successivi: non si è più nel campo delle *perequazioni o interpolazioni osculatrici.*

---

(1) Questo vantaggio è stato inopportunamente trascurato da GINI e GALVANI nella costruzione delle nostre tavole di mortalità per il 1921, dove appunto occorreva preliminarmente scomporre classi quinquennali di censiti e di morti in classi annuali di età. Sebbene fosse ben noto, per esperienza nazionale ed internazionale, che entro le classi quinquennali 0-4, 5-9, ... anni di età gli errori di rilevazione e in ispecie gli arrotondamenti delle età non si compensano mai rigorosamente, i detti autori hanno escogitato e applicato un metodo di perequazione, che — oltre ad essere eccessivamente arbitrario — rispetta rigorosamente le classi osservate. Cfr. C. GINI e L. GALVANI, *Tavole di mortalità della popolazione italiana*, in « Annali di statistica », Serie VI, Vol. VIII, Roma, 1931-X. – L. GALVANI, *Confronto tra le recenti tavole di mortalità italiane e quelle di altri paesi*, in « Atti dell'Istituto Nazionale delle Assicurazioni », Vol. III, Roma, 1931-IX. – Di cotesti studi abbiamo fatto un'analisi critica in « Rivista it. di statistica, », Vol. III, pag. 248.

TAVOLA LIII.

Età (a)	Morti in Italia nel 1926 – Dati osservati (b)	Dati perequati – con la formula (90) (c)	Dati perequati – con la formula (91) (d)	Scarti (b) — (c)	Scarti (b) — (d)
5	4.562				
6	3.569				
7	1.780	2.507	2.069	— 727	— 289
8	1.392	1.920	1.263	— 528	+ 129
9	1.231	1.568	1.326	— 337	— 95
10	1.628	1.584	1.554	+ 44	+ 74
11	1.808	1.712	1.794	+ 96	+ 14
12	1.860	1.921	1.886	— 61	— 26
13	2.034	2.100	2.035	— 66	— 1
14	2.276	2.329	2.256	— 53	+ 20
15	2.524	2.606	2.565	— 82	— 41
16	2.949	2.936	2.899	+ 13	+ 50
17	3.245	3.196	3.329	+ 49	— 84
18	3.687	3.468	3.543	+ 219	+ 144
19	3.575	3.628	3.733	— 53	— 158
20	3.884	3.733	3.758	+ 151	+ 126
21	3.747	3.757	3.812	— 10	— 65
22	3.771	3.754	3.785	+ 17	— 14
23	3.810	3.685	3.740	+ 125	+ 70
24	3.560	.	.	.	.
25	3.535	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

Se ammettiamo, per fare l'ipotesi più semplice, che cinque classi consecutive annuali di morti in Italia nel 1926 possano essere ben rappresentate da una retta, ponendo uguale a zero l'età centrale del quinquennio che si considera ed all'unità l'intervallo quinquennale, col metodo dei minimi quadrati si ha il sistema:

$$\Sigma y_i = \Sigma f(\zeta_i) = a_0 n + a_1 \Sigma\zeta_i = a_0 5$$

$$\Sigma \zeta_i y_i = \Sigma \zeta_i f(\zeta_i) = a_0 \Sigma \zeta_i + a_1 \Sigma \zeta_i^2 = a_1 \frac{2}{5}$$

donde:

$$f(0) = \frac{\Sigma y_i}{5} . \qquad (90)$$

Ma, applicando successivamente questa formula, ossia attribuendo ad ogni classe osservata la media aritmetica delle cinque classi consecutive, di cui essa è la centrale, non si è sicuri di ottenere una buona graduazione, anzi nel nostro esempio (colonna (c) della tavola LIII e fig. 47, dove i gruppi sono riferiti al centro di ogni classe annuale di età) essa

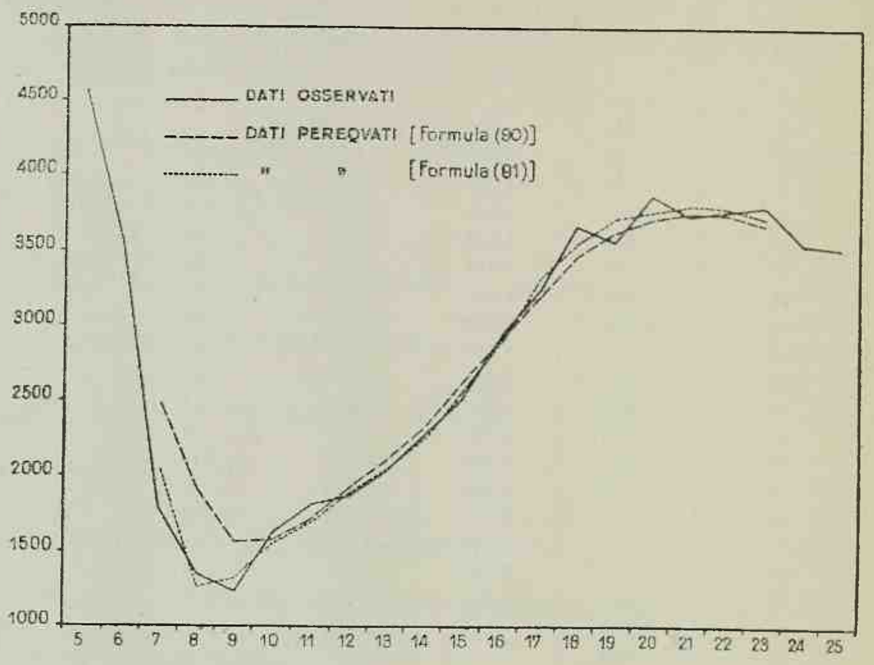


Fig. 47

è molto imperfetta; nè necessariamente migliorerebbe con l'assumere, anzichè cinque, sette o nove... dati consecutivi; ed è chiaro che in tal guisa aumenterebbe il numero dei dati estremi non suscettibili di perequazione.

E questo il cosidetto metodo di Wittstein di *perequazione per medie mobili* (1).

Se sostituiamo all'ipotesi dell'andamento rettilineo quella di un andamento secondo una parabola quadratica (2), otteniamo:

$$\Sigma y_i = \Sigma f(\zeta_i) = 5 a_0 + \frac{2}{5} a_2$$

(1) Su questo ed altri metodi simili (WOOLHOUSE, ecc.) si consulti E. BLASCHKE, op. cit.

(2) FELICE VINCI, *Sulla costruzione delle tavole di mortalità*, op. cit.

$$\Sigma \zeta_i \, y_i = \Sigma \zeta_i \, f(\zeta_i) = \frac{2}{5} a_1$$

$$\Sigma \zeta_i^2 \, y_i = \Sigma \zeta_i^2 \, f(\zeta_i) = \frac{2}{5} a_0 + \frac{34}{625} a_2$$

donde:

$$f(0) = \frac{-3\,(y_{-2/5} + y_{2/5}) + 12\,(y_{-1/5} + y_{1/5}) + 17\,y_0}{35} \qquad (91)$$

Con tale ipotesi la graduazione è migliore (colonna (d) della tavola LIII e fig. 47), e forse ancora migliorerebbe assumendo un numero di dati diverso da 5 o una parabola di grado superiore; ma siamo sempre lontani nè potremmo mai aspirare a quella perfetta graduazione, che il metodo precedente degli archi raccordati assicura.

Lo stesso è a dire per le perequazioni successive di dati perequati, le quali peraltro, oltre a sacrificare un gran numero di dati estremi, non hanno sempre fondamento ragionevole e spesso si ispirano a criterî, forse troppo nobilmente chiamati meccanici.

Tanto, allora, converrebbe eseguire semplici perequazioni o interpolazioni grafiche, ossia far passare ad occhio una curva semplice tra le maglie della spezzata empirica o per punti assegnati, come pure suol farsi e abbiamo già fatto nelle figure 34 e 36, o come il lettore con bel garbo può fare ad es. sulle figure 4, 5, 6 di questo volume.

Il ricorso a siffatti e così svariati procedimenti dimostra l'importanza delle perequazioni e interpolazioni, indipendentemente dall'esistenza di una funzione analitica generale. Basti pensare che con esse si riesce a calcolare più correttamente non solo i valori centrali e prevalenti, ma anche le altre costanti caratteristiche, di cui ci siamo occupati nel precedente capitolo: esse ci mettono in grado di ottenere dati più attendibili e più dettagliati di quelli di cui si disponga, che sono spesso molto errati o perturbati o raccolti in gruppi troppo estesi.

Strettamente connessa a questi procedimenti è la cosidetta extrapolazione, la quale consiste nell'estendere la validità di una funzione oltre i limiti delle osservazioni, e nel calcolare in base ad essa i dati teorici eccedenti quei limiti.

Si comprende facilmente quanto sia pericoloso ed infido tale procedimento, se — come per lo più accade — la scelta della funzione non possa giustificarsi che entro dati confini o abbia addirittura lo scopo di rappresentare comodamente una parte della curva empirica, a prescindere dalle altre parti, pur note, di essa.

Chi oserebbe estendere l'andamento lineare, col quale nella figura 42 si è rappresentato il credito semestrale dei depositanti nelle nostre Casse di Risparmio dal dicembre 1919 al dicembre 1925, a periodi di tempo diversi da quello esaminato? E non sarebbe forse assurdo ammettere valida, per l'intera scala delle età, quella forma generale parabolica, che nella figura 45 è stata assunta a rappresentare l'andamento delle classi annuali di morti da 68 a 82 anni?

Spesso, però, l'extrapolazione s'impiega utilmente per scopi speciali di analisi, com'è dimostrato dalla figura 48, ricavata dalla Relazione della nostra Amministrazione ferroviaria per l'esercizio 1928-29. In essa è rappresentato, con linea continua, il carico (in migliaia di tonnellate) delle merci sulle Ferrovie dello Stato negli esercizî finanziarî dal 1908-09 al 1928-29, e, con tratteggio, la retta dell'incremento medio del carico nel periodo prebellico 1908-14.

Si nota che, dopo la depressione bellica e la reazione postbellica, la spezzata del carico ha avuto tendenza negli ultimi due anni ad assumere l'andamento che si sarebbe potuto prevedere, extrapolando in base all'andamento lineare dell'ultimo sessennio prebellico.

Si potrebbe inferire che — dopo la perturbazione della guerra e del dopoguerra — l'insieme di circostanze, che influivano sull'incremento del traffico alla vigilia della guerra, avesse ripreso nel 1927-28 il suo ritmo normale.

Tenendo presente che le variazioni del carico ferroviario delle merci sono strettamente legate a quelle del volume complessivo degli scambi nel Regno, s'intenderà facilmente l'importanza *semiologica* di quel rilievo.

Talvolta l'extrapolazione consente il controllo di ipotesi, la valutazione degli effetti dell'avveramento di esse. Ad es., in base alle stime dell'Istituto internazionale di statistica (1), la popolazione mondiale intorno al 1920 ed al 1930 sarebbe ammontata rispettivamente a 1.788 e a 1.988 milioni.

Per ciò che si è detto al n. 2 a proposito della formula (55), il saggio medio annuo di variazione sarebbe stato circa:

$$\frac{\log 1.988.000.000 - \log 1.788.000.000}{10 \times 0{,}434294} = 0{,}010603$$

ossia l'1,06 per cento.

È facile rilevare che, se la popolazione mondiale avesse sempre presentato questo saggio d'incremento, sarebbero bastati 1.944 anni, perchè una coppia coniugale formasse per successive riproduzioni i 1.788 milioni di abitanti del 1920. Infatti dalla (55) si ricava:

$$t = \frac{\log 1.788.000.000 - \log 2}{0{,}010603 \times 0{,}434294} = 1.943{,}9.$$

(1) *Aperçu de la démographie des divers pays du monde*, publié par l'Office permanent de l'Institut international de statistique. La Haye, W. P. Van Stockum, 1932.

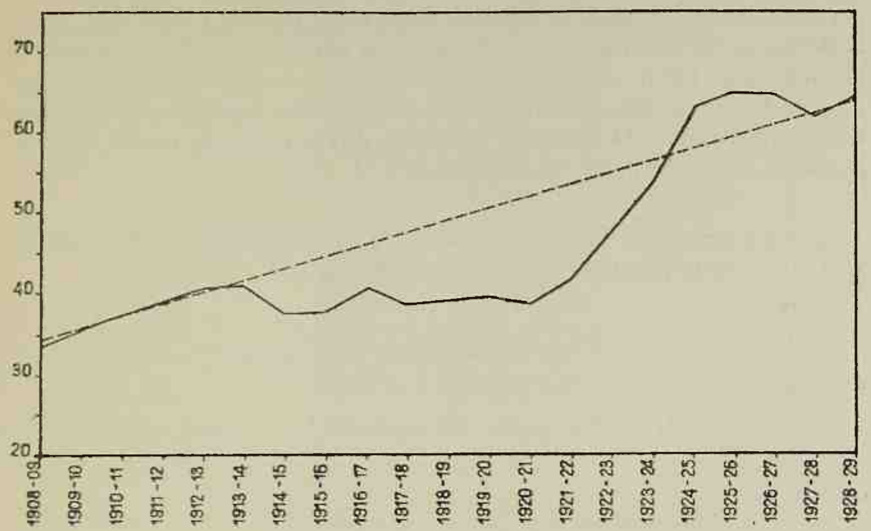


Fig. 48

Ma poichè è lecito ritenere che il solo Impero Romano, alla morte di Augusto, aveva una popolazione di circa 54 milioni di abitanti [1], dobbiamo concludere che il saggio medio annuo d'incremento della popolazione mondiale dev'essere stato in passato minore di quello accertato nell'ultimo decennio 1920-30.

D'altra parte, se la popolazione mondiale continuasse a crescere, in futuro, secondo il detto saggio annuo dell'1,0603 per cento, ponendo nella (55):

$$\frac{P_t}{P_0} = 2$$

si avrebbe:

$$t = \frac{\log 2}{0,010603 \times 0,434294} = 65,4$$

ossia in poco più di 65 anni la popolazione mondiale sarebbe il doppio di quella del 1930.

Fra un millennio, ponendo:

$$\log P_t = \log 1.988.000.000 + 0,010603 \times 1.000 \times 0,434294 = 13,90324$$

sarebbe:

$$P_t = 80.030.000.000.000$$

---

[1] GIULIO BELOCH, *La popolazione del mondo greco-romano*, in « Biblioteca di storia economica ». Milano, Società editrice libraria, 1909 (Vol. IV, pag. 447).

e valutando la superficie geografica della Terra (esclusi i mari e le regioni artiche e antartiche) a 132.160.000 km.², si avrebbe la bellezza di 605.544 abitanti per km.².

Poichè ciò sembra assurdo, è da ritenere che il saggio di variazione della popolazione nel prossimo millennio dovrà essere in media inferiore a quello accertato nello scorso decennio (¹).

10. Le funzioni lineari rispetto ai coefficienti. L'analisi armonica. – Se la funzione scelta abbia la forma:

$$z = f(x) = a_0 + a_1 g_1(x) + a_2 g_2(x) + \ldots\ldots$$

dove sia

$$g_1(x) = \frac{1}{x}, \qquad g_2(x) = \frac{1}{x^2}, \ldots\ldots$$

oppure

$$g_1(x) = \log x, \qquad g_2(x) = \log^2(x), \ldots\ldots$$

od anche

$$g_1(x) = \operatorname{sen} x, \qquad g_2(x) = \cos x, \ldots\ldots$$

o, in generale, siano $g_i(x)$ funzioni numericamente determinate di x, cosicchè la funzione $f(x)$ sia lineare rispetto ai coefficienti che in essa compaiono, anche in tali casi i coefficienti si potranno facilmente calcolare con la condizione dei minimi quadrati o con uno degli altri metodi esposti.

Nel caso dei minimi quadrati, dalla condizione:

$$\Sigma\left[y_i - a_0 - a_1 g_1(x_i) - a_2 g_2(x_i) - \ldots\ldots - a_r g_r(x_i)\right]^2 = \text{minimo}$$

potrà ricavarsi appunto un sistema lineare di equazioni, che differirà da (62) solo per il fatto che, al posto delle potenze successive di x, comparirà un'altra funzione numericamente determinata di x.

Dovendosi, ad es., adattare alle n coppie x_i, y_i la somma trigonometrica (54):

$$\begin{aligned} z = a_0 + a_1 \cos\varphi + a_2 \cos 2\varphi + \ldots\ldots \\ + b_1 \operatorname{sen}\varphi + b_2 \operatorname{sen} 2\varphi + \ldots\ldots \end{aligned} \qquad (54)$$

(¹) Cfr. il nostro studio: *Previsioni demografiche*, in «Studî del Laboratorio di Statistica del R. Istituto Superiore di Scienze Economiche di Venezia». Venezia, Libreria Emiliana, 1927. – Cfr. anche G. H. Knibbs, in «Scientia», 1925-1926.

si riuscirà al seguente sistema, lineare rispetto ai coefficienti:

$$\left.\begin{array}{l}
\Sigma y_i = a_0 n + a_1 \Sigma \cos \varphi_i + a_2 \Sigma \cos 2 \varphi_i + \ldots.. \\
\qquad + b_1 \Sigma \operatorname{sen} \varphi_i + b_2 \Sigma \operatorname{sen} 2 \varphi_i + \ldots.. \\
\Sigma y_i \cos \varphi_i = a_0 \Sigma \cos \varphi_i + a_1 \Sigma \cos^2 \varphi_i + a_2 \Sigma \cos 2 \varphi_i \cos \varphi_i + \ldots.. \\
\qquad + b_1 \Sigma \operatorname{sen} \varphi_i \cos \varphi_i + b_2 \Sigma \operatorname{sen} 2 \varphi_i \cos \varphi_i + \ldots.. \\
\Sigma y_i \cos 2 \varphi_i = a_0 \Sigma \cos 2 \varphi_i + a_1 \Sigma \cos \varphi_i \cos 2 \varphi_i + a_2 \Sigma \cos^2 2 \varphi_i + \ldots.. \\
\qquad + b_1 \Sigma \operatorname{sen} \varphi_i \cos 2 \varphi_i + b_2 \Sigma \operatorname{sen} 2 \varphi_i \cos 2 \varphi_i + \ldots.. \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
\Sigma y_i \operatorname{sen} \varphi_i = a_0 \Sigma \operatorname{sen} \varphi_i + a_1 \Sigma \cos \varphi_i \operatorname{sen} \varphi_i + a_2 \Sigma \cos 2 \varphi_i \operatorname{sen} \varphi_i + \ldots.. \\
\qquad + b_1 \Sigma \operatorname{sen}^2 \varphi_i + b_2 \Sigma \operatorname{sen} 2 \varphi_i \operatorname{sen} \varphi_i + \ldots.. \\
\Sigma y_i \operatorname{sen} 2 \varphi_i = a_0 \Sigma \operatorname{sen} 2 \varphi_i + a_1 \Sigma \cos \varphi_i \operatorname{sen} 2 \varphi_i + a_2 \Sigma \cos 2 \varphi_i \operatorname{sen} 2 \varphi_i + \ldots.. \\
\qquad + b_1 \Sigma \operatorname{sen} \varphi_i \operatorname{sen} 2 \varphi_i + b_2 \Sigma \operatorname{sen}^2 2 \varphi_i + \ldots.. \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots
\end{array}\right\} \quad (92)$$

Ponendo — come si è detto al n. 2 —:

$$\varphi_i = 2\pi \frac{x_i}{X},$$

i valori dei coefficienti si potranno ricavare senza difficoltà.

Il calcolo delle sommatorie del precedente sistema si può semplificare mediante le classiche formule, che permettono di trasformare in somme i prodotti delle funzioni seno e coseno:

$$2 \cos \alpha \varphi \cos \beta \varphi = \cos (\alpha + \beta) \varphi + \cos (\alpha - \beta) \varphi$$

$$2 \operatorname{sen} \alpha \varphi \cos \beta \varphi = \operatorname{sen} (\alpha + \beta) \varphi + \operatorname{sen} (\alpha - \beta) \varphi$$

$$2 \cos \alpha \varphi \operatorname{sen} \beta \varphi = \operatorname{sen} (\alpha + \beta) \varphi - \operatorname{sen} (\alpha - \beta) \varphi$$

$$2 \operatorname{sen} \alpha \varphi \operatorname{sen} \beta \varphi = \cos (\alpha - \beta) \varphi - \cos (\alpha + \beta) \varphi.$$

Infatti in tal modo tutte le sommatorie di prodotti nei secondi membri delle (92) si riducono alle due forme:

$$\Sigma \cos (\alpha \pm \beta) \varphi \quad \text{e} \quad \Sigma \operatorname{sen} (\alpha \pm \beta) \varphi.$$

Inoltre, *se i φ_i sono fra loro equidistanti e tutti compresi in un solo periodo* 2π, si dimostra facilmente, in base a note proprietà delle due funzioni circolari considerate, che le sommatorie Σ sen $m\,\varphi$ riescono tutte nulle e le sommatorie $\Sigma \cos m\,\varphi$ pure nulle, ad eccezione di quelle in cui è $m = 0$, perchè allora:

$$\Sigma \cos 0 = n.$$

Pertanto il sistema (92) si riduce a:

$$
\begin{aligned}
\Sigma y_i &= a_0 n \\
\Sigma y_i \cos\varphi_i &= \frac{1}{2} n a_1 \\
\Sigma y_i \cos 2\varphi_i &= \frac{1}{2} n a_2 \\
&\cdots\cdots \\
\Sigma y_i \operatorname{sen}\varphi_i &= \frac{1}{2} n b_1 \\
\Sigma y_i \operatorname{sen} 2\varphi_i &= \frac{1}{2} n b_2 \\
&\cdots\cdots
\end{aligned}
$$

donde:

$$
\left.
\begin{aligned}
a_0 &= \frac{1}{n} \Sigma y_i \\
a_1 &= \frac{2}{n} \Sigma y_i \cos\varphi_i \\
a_2 &= \frac{2}{n} \Sigma y_i \cos 2\varphi_i \\
&\cdots\cdots \\
b_1 &= \frac{2}{n} \Sigma y_i \operatorname{sen}\varphi_i \\
b_2 &= \frac{2}{n} \Sigma y_i \operatorname{sen} 2\varphi_i \\
&\cdots\cdots
\end{aligned}
\right\} \tag{93}
$$

Queste formule — chiamate formule di Bessel [1] — offrono il grande vantaggio che, ove non sia stabilito l'ordine della somma trigonometrica da adattare e si voglia procedere per approssimazioni successive, sarà possibile calcolare i coefficienti dei termini della somma, che man mano si considera, senza esser costretti a rifar sempre i calcoli da capo.

Per ottenere, appunto, un simile vantaggio con le funzioni razionali intere (60), è stato escogitato il metodo delle approssimazioni successive, che abbiamo già esposto al n. 6.

Si noti ancora che le (93) possono calcolarsi rapidamente, sostituendo ai dati osservati i loro scarti da un'origine arbitraria. In particolare, po-

---

[1] E. T. WHITTAKER and G. ROBINSON, *The calculus of observations*, op. cit., Cap. X; e G. CASSINIS, *Calcoli numerici, grafici e meccanici*, op. cit., Cap. VI. - Per un'esposizione elementare, M. FRÉCHET et R. ROMANN, *Représentation des lois empiriques* etc., op. cit., Cap. IV.

nendo $\varphi_1 = 0$, per $n = 12$ si avranno i risultati della tavola LIV, dove 0, $\frac{1}{6}\pi$, $\frac{1}{3}\pi$, ... (misure ciclometriche) corrispondono rispettivamente ad angoli di 0°, 30°, 60° .....; e le (93) diventano:

$$\left.\begin{array}{l}
a_0 = \frac{1}{12}(y_1 + y_2 + y_3 + y_4 + y_5 + y_6 + y_7 + y_8 + y_9 + y_{10} + y_{11} + y_{12}) \\
a_1 = \frac{1}{12}\left[2(y_1 - y_7) + \sqrt{3}\,(y_2 - y_6 - y_8 + y_{12}) + y_3 - y_5 - y_9 + y_{11}\right] \\
a_2 = \frac{1}{12}\left[2(y_1 - y_4 + y_7 - y_{10}) + y_2 - y_3 - y_5 + y_6 + y_8 - y_9 - y_{11} + y_{12}\right] \\
\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
b_1 = \frac{1}{12}\left[2(y_4 - y_{10}) + \sqrt{3}\,(y_3 + y_5 - y_9 - y_{11}) + y_2 + y_6 - y_8 - y_{12}\right] \\
b_2 = \frac{1}{12}\sqrt{3}\,(y_2 + y_3 - y_5 - y_6 + y_8 + y_9 - y_{11} - y_{12}) \\
\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots
\end{array}\right\} \quad (93bis)$$

TAVOLA LIV.

φ_i	$\cos\varphi_i$	$\operatorname{sen}\varphi_i$	$\cos 2\varphi_i$	$\operatorname{sen} 2\varphi_i$
0	$+1$	0	$+1$	0
$\frac{1}{6}\pi$	$+\frac{\sqrt{3}}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{\sqrt{3}}{2}$
$\frac{1}{3}\pi$	$+\frac{1}{2}$	$+\frac{\sqrt{3}}{2}$	$-\frac{1}{2}$	$+\frac{\sqrt{3}}{2}$
$\frac{1}{2}\pi$	0	$+1$	-1	0
$\frac{2}{3}\pi$	$-\frac{1}{2}$	$+\frac{\sqrt{3}}{2}$	$-\frac{1}{2}$	$-\frac{\sqrt{3}}{2}$
$\frac{5}{6}\pi$	$-\frac{\sqrt{3}}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$-\frac{\sqrt{3}}{2}$
π	-1	0	$+1$	0
$\left(1+\frac{1}{6}\right)\pi$	$-\frac{\sqrt{3}}{2}$	$-\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{\sqrt{3}}{2}$
$\left(1+\frac{1}{3}\right)\pi$	$-\frac{1}{2}$	$-\frac{\sqrt{3}}{2}$	$-\frac{1}{2}$	$+\frac{\sqrt{3}}{2}$
$\left(1+\frac{1}{2}\right)\pi$	0	-1	-1	0
$\left(1+\frac{2}{3}\right)\pi$	$+\frac{1}{2}$	$-\frac{\sqrt{3}}{2}$	$-\frac{1}{2}$	$-\frac{\sqrt{3}}{2}$
$\left(1+\frac{5}{6}\right)\pi$	$+\frac{\sqrt{3}}{2}$	$-\frac{1}{2}$	$+\frac{1}{2}$	$-\frac{\sqrt{3}}{2}$

Similmente, dividendo 2π in 6, 18, 24, ... parti e assumendo una qualsiasi somma trigonometrica di ordine inferiore ai noti limiti, si potranno calcolare le relative formule semplificatrici.

Un esempio: sommando le due ultime colonne della tavola XXI, abbiamo ricavato la seguente distribuzione mensile dei morti a Milano nel 1916 in età di 60 anni e più (tav. LV). Questi dati risentono l'influenza degli esodi estivi — come sui dati per il complesso del Regno influiscono le emigrazioni nette a scopi di lavoro e il movimento turistico —; ma dobbiamo accontentarcene, perchè non è possibile ricavarne i quozienti di mortalità dividendoli per il numero mensile corrispondente degli esposti a morire.

Le medie giornaliere dei morti, moltiplicate per 30, sono riportate nella terza colonna [1].

Scegliendo la somma trigonometrica di ordine 2 e calcolando su questi ultimi dati le (93 *bis*), si ottiene

$$\begin{aligned} a_0 &= 133{,}2500 \\ a_1 &= 49{,}8500 \\ a_2 &= 21{,}3333 \\ b_1 &= 7{,}6036 \\ b_2 &= 5{,}4848 \end{aligned}$$

e quindi, tenendo presente che nel nostro esempio è: $\varphi = 2\pi \dfrac{t}{12}$ (per t variabile tra 0 e 11),

$$\begin{aligned} z = 133{,}25 &+ 49{,}85 \cos \frac{\pi}{6} t + 21{,}3333 \cos \frac{\pi}{3} t \\ &+ 7{,}6036 \operatorname{sen} \frac{\pi}{6} t + 5{,}4848 \operatorname{sen} \frac{\pi}{3} t \,. \end{aligned}$$

Da questa formula si traggono i dati riportati nella quarta colonna della tavola LV, dove per il controllo dei calcoli abbiamo eseguito la somma dei dati osservati e di quelli calcolati. In base agli scarti che questi presentano dai corrispondenti dati osservati (ultima colonna), si ottiene:

$$\sqrt{\frac{\Sigma \varepsilon_i^2}{n}} = \sqrt{\frac{2.775{,}3929}{12}} = \sqrt{231{,}28274} = \pm\, 15{,}204\,,$$

il quale risultato, diviso per $M = 133{,}25$, rivela un'approssimazione dell'11,41 per cento, della quale forse potremmo accontentarci, data la esiguità e la conseguente irregolarità dei dati.

---

[1] La differenza tra gli $\bar{y}_i$ e gli y_i deriva dal fatto che i primi riguardano 366 giorni ed i secondi sono stati tutti ridotti a mesi di 30 giorni, ossia ad un anno di 360 giorni.

TAVOLA LV.

Mesi dell'anno x_i	Numero dei morti in età di 60 anni e più: greggio y_i	ridotto a 30 giorni $\bar{y}_i$	z_i	$\varepsilon_i = \bar{y}_i - z_i$
1	240	232	204,43	+ 27,57
2	180	186	195,64	— 9,64
3	162	158	158,84	— 0,84
4	105	105	119,52	— 14,52
5	114	110	99,49	+ 10,51
6	100	100	99,80	+ 0,20
7	108	105	104,73	+ 0,27
8	96	93	101,70	— 8,70
9	93	93	95,82	— 2,82
10	139	135	104,31	+ 31,31
11	119	119	136,17	— 17,17
12	168	163	178,54	— 15,54
Totali	1.624	1.599	1.598,99	

Se volessimo considerare anche i termini di ordine 3, basterebbe aggiungere nella tav. LIV le colonne cos 3φ e sen 3φ, determinare le formule abbreviate di a_3 e b_3 e procedere al calcolo di esse.

Ponendo:

$$a_s = \varrho_s \operatorname{sen} \alpha_s \ , \qquad b_s = \varrho_s \cos \alpha_s \ ,$$

in base ad una classica formula trigonometrica, la (54) assume la forma:

$$z = a_0 + \varrho_1 \operatorname{sen} (\varphi + \alpha_1) + \varrho_2 \operatorname{sen} (2\varphi + \alpha_2) + \ldots.. \qquad (54\ bis)$$

dove:

$$\varrho_s = + \sqrt{a_s^2 + b_s^2}$$

$$\operatorname{tg} \alpha_s = \frac{a_s}{b_s}$$

Nel nostro esempio si ha:

$$\varrho_1 = 50{,}43 \ , \qquad \varrho_2 = 22{,}02$$
$$\alpha_1 = 0{,}4518\,\pi \ , \qquad \alpha_2 = 0{,}4199\,\pi$$

e possiamo anche dire di avere adattato alla nostra serie mensile la somma di due funzioni seno.

Per ragioni d'ordine fisico, il secondo, terzo, ... termine della (54 *bis*) si chiamano *armoniche* d'ordine 1, 2, ...

Poichè i seni variano tra -1 e $+1$, è facile persuadersi che l'armonica di ordine s può variare tra $-\rho_s$ e $+\rho_s$: tale coefficiente dà dunque l'*ampiezza* dell'armonica.

Inoltre la variabile indipendente risulta spostata di α, e pertanto questo coefficiente dà il grado di sfasatura o senz'altro la *fase* dell'armonica.

Infine, quando $\varphi + \alpha_1$, $2\varphi + \alpha_2$, $3\varphi + \alpha_3$, ... variano di 2π, anche φ, 2φ, 3φ, ... variano di 2π, mentre φ varia rispettivamente di 2π, $\frac{2\pi}{2}$, $\frac{2\pi}{3}$, ...: sono questi i *periodi* delle successive armoniche, allorchè si assume come periodo fondamentale 2π.

Ond'è che la nostra funzione normale può esser considerata come la somma di una funzione fondamentale di periodo 2π e di altre funzioni armoniche di periodi uguali a sottomultipli di 2π.

Adattando la somma trigonometrica, si esegue adunque una scomposizione di un andamento periodico in oscillazioni od onde elementari; e per questo si dice che con l'adattamento della (54) si fa l'*analisi armonica* dei dati.

Allora, il fatto che nel nostro esempio ci siamo arrestati all'armonica di second'ordine significa che, dopo aver scomposto la nostra serie storica in due onde sinusoidali di ampiezza rispettivamente uguale a 50,43 e 22,02, abbiamo ritenuto trascurabili le onde residue, ossia i valori di ρ_3, ...

Quando si tenga presente l'influenza che sull'andamento stagionale della mortalità esercitano i rigori invernali e i forti calori estivi, e il fatto che i meteorologi rappresentano appunto con la funzione adottata le escursioni mensili della temperatura, potrà intendersi l'importanza del precedente risultato.

Se avessimo considerato, invece dei soli morti in età di 60 anni e più, la distribuzione mensile dei morti a qualsiasi età, due sole armoniche non sarebbero più state sufficienti a rappresentare i dati neppure con larga approssimazione e l'analisi si sarebbe notevolmente complicata, a cagione delle particolari oscillazioni dei morti alle altre età, specie infantili.

11. Le funzioni non lineari rispetto ai coefficienti. Il metodo dei momenti. – Supponiamo che la funzione scelta sia iperbolica, o sia la funzione esponenziale (49), o abbia comunque una forma non lineare rispetto ai coefficienti: in tali casi i metodi esposti daranno luogo a sistemi non lineari alquanto complessi.

A volte ne è possibile la risoluzione con semplici artifici. Ad es. la seguente funzione di Gompertz-Makeham:

$$z = k\, s^x\, g^{c^x} \qquad (94)$$

si è ritenuta idonea a rappresentare il numero degl'individui di una data generazione, che sopravvivono all'età x (almeno trascurando le età infantili e senili), e — per ragioni che non è questo il luogo di esporre — si è rivelata comoda nella pratica delle assicurazioni sulla vita.

Per determinare i quattro coefficienti: k, s, g, c e così adattare quella funzione ad una distribuzione per età di sopravviventi l_x provenienti da una stessa generazione, si pone anzitutto:

$$\log z = \log k + x \log s + c^x \log g \ .$$

Dividendo, poi, in quattro parti uguali i dati osservati compresi ad es. tra 17 e 88 anni e uguagliandoli col metodo delle somme ai dati teorici desunti dalla precedente equazione, si ottiene il sistema:

$$\left.\begin{aligned}
\sum_{x=17}^{34} \log l_x &= 18 \log k + 459 \log s + \frac{c^{17}(c^{18}-1)}{c-1} \log g \\
\sum_{x=35}^{52} \log l_x &= 18 \log k + 783 \log s + \frac{c^{35}(c^{18}-1)}{c-1} \log g \\
\sum_{x=53}^{70} \log l_x &= 18 \log k + 1.107 \log s + \frac{c^{53}(c^{18}-1)}{c-1} \log g \\
\sum_{x=71}^{88} \log l_x &= 18 \log k + 1.431 \log s + \frac{c^{71}(c^{18}-1)}{c-1} \log g
\end{aligned}\right\} \qquad (95)$$

che è lineare solo rispetto ai primi due coefficienti.

Facendo le differenze prime dei quattro numeri posti a primo membro, si ha:

$$\begin{aligned}
\Delta \sum_{x=17}^{34} \log l_x &= 324 \log s + \frac{c^{17}(c^{18}-1)^2}{c-1} \log g \\
\Delta \sum_{x=35}^{52} \log l_x &= 324 \log s + \frac{c^{35}(c^{18}-1)^2}{c-1} \log g \\
\Delta \sum_{x=53}^{70} \log l_x &= 324 \log s + \frac{c^{53}(c^{18}-1)^2}{c-1} \log g
\end{aligned}$$

e facendo ancora le differenze prime dei tre numeri posti a primo membro:

$$\begin{aligned}
\Delta^2 \sum_{x=17}^{34} l_x &= \frac{c^{17}(c^{18}-1)^3}{c-1} \log g \\
\Delta^2 \sum_{x=35}^{52} l_x &= \frac{c^{35}(c^{18}-1)^3}{c-1} \log g
\end{aligned}$$

donde, dividendo:

$$\frac{\Delta^2 \sum\limits_{x=35}^{52} \log l_x}{\Delta^2 \sum\limits_{x=17}^{34} \log l_x} = c^{18}.$$

Quest'ultima formula fornisce il valore di c, una delle due precedenti quello di g, una delle tre ancora precedenti quello di s, e risalendo ancora ad una delle equazioni del sistema (95) si otterrà il valore di k (1).

Altre volte, però, il sistema risultante è così complesso, da esserne impossibile o molto laboriosa la risoluzione per artifici.

In alcuni casi si può — sotto certe condizioni — trasformare la funzione non lineare rispetto ai coefficienti in una funzione lineare; ma di questo argomento non è possibile trattare in via sistematica ed elementare, perchè i procedimenti variano caso per caso e presuppongono nozioni di matematiche superiori.

Ci limitiamo a notare che la trasformazione può talora eseguirsi, assumendo solo i logaritmi delle variabili. Per fare un esempio dei più semplici, adatteremo la funzione iperbolica:

$$N_x = H x^{-\alpha} \tag{96}$$

alla distribuzione dei contribuenti con reddito superiore a L. 1.000, inscritti nei ruoli della tassa di famiglia in 23 città italiane (2).

Ponendo la (96) nella forma:

$$\log N_x = \log H - \alpha \log x$$

ed applicando il metodo delle differenze (n. 8), si ha:

$n = 7$ $\quad \Sigma \log x_i = 26{,}32222$ $\quad \Sigma \log y_i = 26{,}13131$

$\Sigma \Delta \log x_i = 4{,}39794 - 3 = 1{,}39794$ $\quad \Sigma \Delta \log y_i = 2{,}80956 - 4{,}77442 = -1{,}96486$

Tavola LVI.

Redditi in lire x_i	Numero di contribuenti con reddito maggiore di x_i y_i	$\log x_i$	$\log y_i$
1.000	59.486	3	4,77442
2.000	26.968	3,30103	4,43085
4.000	9.766	3,60206	3,98972
7.000	4.264	3,84510	3,62982
10.000	2.397	4	3,37967
15.000	1.310	4,17609	3,11727
25.000	645	4,39794	2,80956
Totali	104.836	26,32222	26,13131

(1) G. King, *A text-book of the principles of interest, life annuities and assurances.* London, Layton, 1902.

(2) R. Benini, op. cit., pag. 141.

e quindi:

$$\alpha = \frac{-1,96486}{1,39794} = -1,4055$$

$$\log H = \frac{26,13131}{7} + 1,4055 \times \frac{26,32222}{7} = 9,0182$$

$$\log N_x = 9,0182 - 1,4055 \log x.$$

Ne segue:

$$N_x = 1.042.800.000\ x^{-1,4055}$$

Le prime due colonne della tavola seguente riportano i valori di $\log N_x$ e gli scarti di essi dai logaritmi dei valori osservati. Le altre due colonne contengono i risultati ottenuti col metodo di Cauchy.

Metodo delle differenze		Metodo di Cauchy	
$\log N_x$	$\log y_i - \log N_x$	$\log N_x$	$\log y_i - \log N_x$
4,8017	— 0,0273	4,8354	— 0,0610
4,3786	+ 0,0522	4,3990	+ 0,0318
3,9555	+ 0,0342	3,9624	+ 0,0273
3,6139	+ 0,0159	3,6101	+ 0,0197
3,3962	— 0,0165	3,3854	— 0,0057
3,1487	— 0,0314	3,1301	— 0,0128
2,8369	— 0,0273	2,8085	+ 0,0011
26,1315	— 0,1025	26,1309	— 0,0795
	+ 0,1023		+ 0,0799

Passando dai logaritmi ai numeri, si ha:

Metodo delle differenze		Metodo di Cauchy	
N_x	$y_i - N_x$	N_x	$y_i - N_x$
63.120	— 3.634	68.450	— 8.964
23.910	+ 3.058	25.060	+ 1.908
9.026	+ 740	9.170	+ 596
4.111	+ 153	4.075	+ 189
2.490	— 93	2.428	— 31
1.408	— 98	1.350	— 40
687	— 42	643	+ 2

donde si rileva il fatto, già noto, che la funzione scelta si adatta male ai dati empirici in corrispondenza ai redditi più piccoli, e si constata altresì che i risultati del metodo delle differenze potrebbero preferirsi a quelli del metodo di Cauchy.

Gli è che l'uguaglianza tra le somme empiriche e quelle teoriche delle differenze prime viene ad attenuare la divergenza fra i valori estremi di log N_x, uguagliandola a quella empirica.

Risultati migliori si sarebbero ottenuti sui dati dei redditieri prussiani, di cui al diagramma logaritmico della figura 41. Questo, infatti, mostrando un andamento marcatamente rettilineo almeno in un primo tratto, fa prevedere in complesso un migliore adattamento della funzione iperbolica e una più netta conferma di una di quelle che, nel campo della distribuzione dei redditi, sono state chiamate « le leggi di Pareto ».

Ma — come si è detto al n. 2 — si potrebbe anche aggiungere alla (96) un fattore $e^{-\beta x}$, che passando ai logaritmi assumerebbe forma lineare. La costante a della (58) potrebbe, invece, calcolarsi solo per approssimazioni.

Occorre però tenere ben presente che le condizioni del metodo di adattamento impiegato sarebbero in tal guisa (come pure per la (94)) rispettate dai logaritmi, ma non dai numeri; e che pertanto quest'artificio può usarsi utilmente, se i risultati numerici si avvicinano alla condizione posta per i logaritmi.

Per rendere una funzione lineare rispetto ai coefficienti, analogo impiego dei logaritmi può farsi per le funzioni esponenziali (49).

Ci si potrebbe anche fondare sulle proprietà delle differenze dei varî ordini, od anche rinunziare all'utilizzazione di tutti i dati e far passare la curva per dati punti empirici, convenientemente scelti o perequati anche con uno dei procedimenti esposti al n. 9; si potrebbe — come *ultima ratio* — avvalersi di metodi grafici o numerici di approssimazione (1).

Indipendentemente dagli artifici e dai casi particolari esposti, v'ha un metodo, che si presta all'adattamento di molti tipi di funzioni non lineari.

Generalizzando una nozione data ai n. 5 e 16 del precedente capitolo e chiamando momento di grado s e di origine m la espressione: $\Sigma\,(x_i - m)^s\, y_i$ dove y_i siano adesso valori qualsiansi (e non solo numeri di osservazioni), possiamo dire anzitutto che, per adattare col metodo dei minimi quadrati la funzione razionale intera (60), le equazioni normali (62) si ottengono uguagliando i primi $r+1$ momenti effettivi con i corrispondenti momenti teorici della funzione scelta (rispetto all'origine 0 ed a cominciare dal momento di grado 0), ossia ponendo le $r+1$ condizioni:

$$\begin{aligned} \Sigma y_i &= \Sigma f(x_i) \\ \Sigma x_i\, y_i &= \Sigma x_i\, f(x_i) \\ &\cdots\cdots \\ \Sigma x_i^r\, y_i &= \Sigma x_i^r\, f(x_i) \end{aligned} \qquad (97)$$

(1) Molti esempî elementari si trovano nell'op. cit. di Fréchet e Romann. Cfr. anche Vilfredo Pareto, *L'interpolazione per la ricerca delle leggi economiche*, in « Giornale degli economisti », 1908.

Se, invece, la funzione scelta abbia una forma diversa dalla (60), le equazioni normali generalmente non si presenteranno più nella forma di eguaglianze tra momenti.

Pertanto, in tali casi, mantenendo un sistema di condizioni del tipo (97), si seguirà un metodo di adattamento diverso da quello dei minimi quadrati: è stato chiamato « metodo dei momenti » (1).

Notiamo subito, però, che cotesto metodo, se spesso permette di evitare le difficoltà, a cui i metodi precedenti conducono nel caso di funzioni non lineari rispetto ai coefficienti, in altri casi conduce a sistemi non lineari ancor più complicati. E ciò, nonostante che l'analisi venga facilitata, sostituendo integrali alle somme e quindi postulando la continuità dei dati empirici.

Per alcuni tipi molto frequenti nelle applicazioni statistiche, quell'inconveniente però non si avvera; anche perchè, riguardo ad alcune speciali funzioni che compaiono nell'adattamento di essi, esistono appositi prontuarî, in gran parte dovuti a Pearson e ad alcuni suoi seguaci (2).

Non è facile dare esempî elementari di applicazione di tal metodo. Ci limitiamo ad osservare che, assumendo che la funzione da adattare sia:

$$z = K\, e^{-h^2 (x-M)^2} \tag{53}$$

e prescindendo dall'esaminare se il metodo dei momenti conduca in questo caso a risultati uguali a quelli di altri metodi di adattamento, le eguaglianze dei momenti di grado 0, 1, 2 — desunti dai dati teorici nell'ipotesi di variazione continua della variabile — ai momenti desunti dagli n dati osservati x_i, y_i, conducono alle formule:

$$K = \frac{\Sigma y_i}{\sqrt{2\pi\sigma^2}}\,, \qquad h^2 = \frac{1}{2\sigma^2}\,, \qquad M = \frac{\Sigma x_i\, y_i}{\Sigma y_i}\,,$$

dove K si assume sempre col segno positivo.

Ciò nell'ipotesi che i momenti empirici non siano molto discosti da quelli che si ricaverebbero pure nell'ipotesi di variazione continua della variabile. E σ il noto simbolo dello scarto quadratico medio.

Per adattare la (53) ad un esempio concreto, dalla tavola I (o meglio, dalla distribuzione procedente per intervalli di un pollice, che per brevità abbiamo dovuto riassumere nella tavola I) ricaviamo:

$$M = 68{,}657$$

$$\sigma = \sqrt{7{,}5625} = \pm\, 2{,}75$$

---

(1) Karl Pearson, *Skew variation in homogeneous material*, op. cit.

(2) Karl Pearson, *Tables for statisticians and biometricians*, Part I, II. Cambridge, University Press (della Parte I c'è una seconda edizione, 1924).

donde:

$$h^2 = \frac{1}{2 \times 7{,}5625} = 0{,}066116$$

$$K = \frac{1.078}{\sqrt{2 \times 3{,}141593 \times 7{,}5625}} = \frac{1.078}{6{,}893} = 156{,}39054$$

Sarà, quindi:

$$z = 156{,}39\, e^{-0{,}066\,(x - 68{,}657)^2}$$

Facendo uso dei logaritmi o mediante un prontuario — di cui nella Parte II daremo notizia — sarà facile calcolare i valori della funzione precedente per valori di x uguali a 60,657; 62,657; ecc., ossia per scarti uguali a — 8, — 6, ecc., e porli a confronto coi dati osservati.

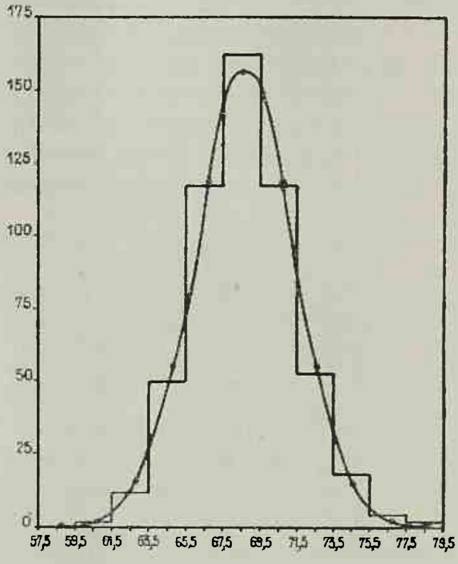


Fig. 49

I risultati sono rappresentati nella figura 49, dove la curva teorica è stata costruita ad occhio attraverso l'istogramma teorico (non rappresentato nella figura) e presenta un andamento non molto diverso da quello, che avremmo ottenuto con intervalli di scarti infinitamente piccoli.

Questo tipo di funzione è fondamentale nell'analisi degli schemi teorici, ossia nella prima categoria di metodi statistici, di cui ci occuperemo nella Parte II. La curva rappresentatrice di essa — come si è osservato al n. 2 — è asintotica rispetto all'asse delle ascisse e simmetrica rispetto all'ordinata massima corrispondente alla media aritmetica: proprietà, di cui abbiamo appunto tratto profitto nel tracciarla.

12. Le funzioni di più variabili e le superfici di frequenze. - Sinora ci siamo limitati a considerare il problema dell'adattamento di una funzione di una variabile ad una successione di coppie di numeri; ma ciò che abbiamo detto può estendersi immediatamente al problema più generale dell'adattamento di una funzione di più variabili a più successioni corrispondenti di numeri, sia che si tratti di serie statistiche a più variabili, sia che si tratti di tavole a parecchie entrate.

Infatti, se è data una funzione di più variabili, lineare rispetto ai coefficienti, si dimostra che la condizione:

$$\Sigma [y_i - a_0 - a_1 \varphi(t_i) - a_2 \psi(u_i) - a_3 \chi(v_i) - \ldots]^2 = \text{minimo}$$

dove $\varphi(t)$, $\psi(u)$, $\chi(v) \ldots$ siano funzioni numericamente determinate delle variabili t, u, v, ..., dà luogo ad un sistema lineare rispetto ai coefficienti, che permette il calcolo di questi ultimi con la condizione dei minimi quadrati.

Partendo dalla prima equazione normale, potremo ugualmente seguire il metodo di Cauchy, e nessuna difficoltà si opporrà all'applicazione del metodo delle somme e delle differenze.

Ammettendo che le terne di dati y_i, t_i, u_i riportati nella tavola IV si possano *grosso modo* rappresentare con la funzione

$$z = a_0 + a_1 t + a_2 u \ , \tag{98}$$

col metodo dei minimi quadrati si avrebbe il sistema:

$$\begin{aligned} \Sigma y_i &= a_0 n + a_1 \Sigma t_i + a_2 \Sigma u_i \\ \Sigma y_i t_i &= a_0 \Sigma t_i + a_1 \Sigma t_i^2 + a_2 \Sigma u_i t_i \\ \Sigma y_i u_i &= a_0 \Sigma u_i + a_1 \Sigma t_i u_i + a_2 \Sigma u_i^2 \end{aligned}$$

risolvendo il quale si avrebbero i valori di a_0, a_1, a_2.

Lo stesso può dirsi, ad es., per la funzione:

$$z = a_0 + a_1 \frac{1}{t^2} + a_2 \sqrt{u} \ \ldots..$$

e in generale per le funzioni lineari rispetto ai coefficienti.

Nel caso di funzioni non lineari rispetto ai coefficienti potrebbero altresì seguirsi artifici analoghi a quelli, a cui abbiamo accennato al numero precedente.

Ad es., alle successioni dei consumi y e dei prezzi medî x corrispondenti in anni successivi t, Schultz (1) ha adattato la funzione:

$$z = A\, x^{-\alpha}\, e^{\beta t} \ , \tag{99}$$

partendo dal presupposto che — almeno per brevi tratti — il consumo sia legato al prezzo da una relazione iperbolica (2) e al tempo da una relazione esponenziale.

---

(1) H. Schultz, *The shifting demand for selected agricultural commodities* 1875-1929, in « Journal of farm economics », 1932.

(2) Secondo le vedute di R. Benini, *Sull'uso delle formule empiriche nell'economia applicata*, in « Giornale degli economisti », 1907; e *Una possibile creazione del metodo statistico: l'economia politica induttiva*, in detto Giornale, 1908. - Cfr. anche F. Vinci, *L'elasticità dei consumi*, in « Rivista it. di statistica, », 1931; e *L'utilità della moneta e l'imposta progressiva*, in detta Rivista, 1933.

Passando dai numeri ai logaritmi, egli ha ottenuto la funzione lineare rispetto ai coefficienti:

$$\log z = \log A - a \log x + (\beta \log e) t$$

e quindi ha potuto applicare facilmente il metodo dei minimi quadrati.

Per rappresentare i gruppi di osservazioni di una tavola a doppia entrata, quand'essa abbia una forma non molto diversa da quella della tavola III, si usa spesso la funzione:

$$z = A\, e^{-a u^2 - b v^2 + 2 c u v} \tag{100}$$

Ponendo $z =$ costante, essa si trasforma in una ellisse, con coordinate passanti per il centro ma non coincidenti con gli assi principali; e attribuendo a z diversi valori, si ottengono altrettante ellissi concentriche, corrispondenti alle linee di livello, di cui ci siamo occupati al n. 7 del secondo capitolo.

La (100) è stata il punto di partenza per la trattazione dei problemi, di cui ci occuperemo nella Parte III di questo manuale; ossia per la costruzione di metodi idonei a mettere in luce le relazioni intercedenti tra gruppi di dati osservati relativi a due o più fenomeni.

Adattandola ad una tavola a doppia entrata col metodo dei momenti — in base ad ipotesi analoghe a quelle formulate nel numero precedente per la (53) — si ottengono le formule:

$$A = \frac{N}{2\pi \sqrt{\sigma_u^2 \sigma_v^2 - (M(u_i v_i))^2}}$$

$$a = \frac{\sigma_v^2}{2[\sigma_u^2 \sigma_v^2 - (M(u_i v_i))^2]}$$

$$b = \frac{\sigma_u^2}{2[\sigma_u^2 \sigma_v^2 - (M(u_i v_i))^2]}$$

$$c = \frac{M(u_i v_i)}{2[\sigma_u^2 \sigma_v^2 - (M(u_i v_i))^2]}$$

dove N è il numero delle coppie (osservate) degli scarti u_i, v_i; σ_u e σ_v sono i noti simboli degli scarti quadratici medî; $M(u_i v_i)$ è la media aritmetica dei prodotti degli u_i, v_i corrispondenti.

Ponendo:

$$r = \frac{M(u_i v_i)}{\sigma_u \sigma_v} \tag{101}$$

le formule precedenti possono scriversi:

$$A = \frac{N}{2\pi \sigma_u \sigma_v \sqrt{1 - r^2}}$$

$$a = \frac{1}{2\,\sigma_u^2\,(1 - r^2)}$$

$$b = \frac{1}{2\,\sigma_v^2\,(1 - r^2)}$$

$$c = \frac{r}{2\,\sigma_u\,\sigma_v\,(1 - r^2)}$$

Per adattare la (100) alla tavola III, chiameremo U e V le stature dei padri e, rispettivamente, quelle dei figli e calcoleremo anzitutto [1]:

$$M(U) = 67{,}70 \qquad M(V) = 68{,}66$$

$$\sigma_u = \pm 2{,}72 \qquad \sigma_v = \pm 2{,}75$$

$$r = 0{,}51$$

Attribuendo ai valori di σ e di A il segno positivo ed essendo $N = 1.078$, avremo:

$$A = \frac{1.078}{2 \times 3{,}141593 \times 2{,}72 \times 2{,}75 \sqrt{1 - 0{,}2601}} = 26{,}67099$$

$$a = \frac{1}{2 \times 2{,}72^2\,(1 - 0{,}2601)} = 0{,}09134$$

$$b = \frac{1}{2 \times 2{,}75^2\,(1 - 0{,}2601)} = 0{,}08935$$

$$c = \frac{0{,}51}{2 \times 2{,}72 \times 2{,}75\,(1 - 0{,}2601)} = 0{,}04607$$

e quindi:

$$z = 26{,}67\, e^{-0{,}09134\, u^2 - 0{,}08935\, v^2 + 0{,}09214\, u\, v} \tag{102}$$

Calcolando coi logaritmi i valori di z corrispondenti alle combinazioni osservate u_i, v_i (come valori di coteste variabili si dovranno sempre assumere gli scarti dei valori centrali delle classi dalle medie aritmetiche rispettive), si avranno i valori teorici corrispondenti ai numeri di osservazioni delle caselle.

Rappresentandoli graficamente su tre assi ortogonali, avremo lo stereogramma teorico con falde regolarmente discendenti al suo piano di base; mentre, rappresentando sul piano cartesiano le coppie u_i, v_i corrispondenti a valori assegnati di z, otterremo le ellissi teoriche corrispondenti alle linee di livello empiriche (da costruire, per ciò che si è detto al n. 7 del secondo capitolo, dividendo preliminarmente i gruppi di osser-

---

(1) I calcoli sono stati eseguiti sui dati pubblicati da Pearson, che sono molto più dettagliati di quelli che per brevità abbiamo raccolto nella tavola III. Cfr. anche G. U. YULE, op. cit., pag. 326 della terza edizione.

vazioni di ogni casella per il prodotto degl'intervalli corrispondenti, o — dato che gl'intervalli sono costanti e il prodotto di essi è 4 — facendo corrispondere alle ellissi teoriche le ellissi empiriche di quote quattro volte maggiori).

Ad es., per la costruzione dell'ellisse di quota 12, basterà ricavare dalla (102) la funzione:

$$\log z = \log 26{,}67 - 0{,}039668\, u^2 - 0{,}038804\, v^2 + 0{,}040015\, u\, v$$

e porre in essa $z = 12$. Risolvendola successivamente per $u_i = 58{,}5 - M(U)$, $60{,}5 - M(U)$, ..., le due radici reali saranno i corrispondenti valori teorici di v e permetteranno di costruire per punti l'ellisse teorica cercata. In base alla tavola III, l'ellisse empirica corrispondente sarà quella di quota 48.

Di questo argomento dovremo occuparci a lungo nella Parte III.

APPENDICI ALLA PARTE PRIMA

A) I. Il settimo censimento generale della popolazione italiana.
II. Appunti storici sull'applicazione del metodo statistico ai fatti economici e demografici.

B) I. Potenze dei numeri interi sino a 100.
II. Somme di potenze dei numeri interi sino a 100.
III. Radici, reciproci e potenze dei reciproci dei numeri interi sino a 100.
IV. Logaritmi, somme di logaritmi, ecc. dei numeri interi sino a 100.
V. Lunghezze degli archi da 1° a 180° e per frazioni di grado.

A) I. - Il VII Censimento generale della popolazione italiana [1].

1. – Il 21 aprile 1931 (Natale di Roma) in tutte le provincie d'Italia, nelle Colonie, nel Possedimento delle Isole dell'Egeo, si è proceduto al censimento della popolazione.

Il censimento generale della popolazione, che si compie a periodi regolari di tempo in tutti i paesi civili, è una delle operazioni più grandiose della statistica sociale ed interessa in egual modo gli uomini di governo, i pubblici amministratori, gli studiosi. Esso è uno dei mezzi tecnici più importanti per la rilevazione della popolazione di uno Stato, considerata nella sua condizione statica. È, per così dire, la riproduzione fotografica della popolazione, considerata nella sua composizione per sesso, per età, per stato civile, per condizioni biologiche ed etniche e per classi economiche e sociali, in un determinato momento. Il censimento rende possibile lo studio di quel fattore essenziale della vita dello Stato che è la popolazione, nei suoi svariati ed interessanti aspetti, e facilita a mezzo della comparazione le indagini riguardanti lo sviluppo demografico, politico, economico e sociale dei vari Paesi. Questo spiega le ragioni per le quali non solo i governi, ma anche gli uomini di scienza, attribuirono in ogni tempo, ma più specialmente nel secolo XIX, la più grande importanza ai censimenti e perchè queste grandiose rilevazioni, grazie ai suggerimenti degli studiosi, si vennero sempre più perfezionando e completando.

I caratteri che distinguono i censimenti, moderni dalle consimili rilevazioni sullo stato delle popolazioni che si fecero in passato, sono essen-

---

[1] Da una pubblicazione fatta dal nostro Istituto Centrale di Statistica in occasione del VII Censimento generale della popolazione del Regno (21 aprile 1931-IX) ricaviamo le notizie del testo, al fine di illustrare le difficoltà che s'incontrano nelle rilevazioni dei dati, la necessità di avere preliminarmente un'idea ben chiara dei criterî di elaborazione di essi e dei fini generali della ricerca, i tangibili progressi che il nostro ordinamento statistico va compiendo sotto lo stimolo veramente necessario di S. E. Mussolini.

La più importante e grandiosa operazione di statistica sociale, qual'è appunto un censimento demografico, non può non darne l'illustrazione migliore. Per maggiori dettagli il lettore può consultare la « Relazione preliminare » al detto censimento (Roma, Istituto Poligrafico dello Stato, 1933-XI).

zialmente tre: la periodicità, la simultaneità, l'universalità. Il censimento che, nella data storica del 21 aprile 1931, si è fatto in Italia, risponde a questi tre caratteri, pure apportando qualche notevole modificazione ed aggiunta a quanto è stato fatto con gli ultimi censimenti.

I censimenti moderni hanno anzitutto il carattere della *periodicità*. Si fanno, cioè, a periodi regolari di tempo, che per alcuni paesi comprendono dieci anni, con altri invece cinque anni.

Secondo la legge del 20 giugno 1871, n. 297, i censimenti generali della popolazione in Italia avrebbero dovuto succedersi a periodi decennali. Tale criterio venne sempre osservato dal 1861 al 1921, con la sola interruzione determinata, per male intese ragioni di economia, nel 1891.

Il censimento che si è fatto nel 1931, ha seguito a 10 anni di distanza il precedente ed è stato quindi il settimo della serie. Ma da questa data si inizierà in Italia, come già si fa in altri Stati, la serie dei censimenti quinquennali, poichè si è ritenuto che dieci anni costituiscano un periodo troppo lungo nei tempi dinamici nei quali viviamo.

L'importanza del censimento è stata accresciuta dalla rilevazione statistica delle abitazioni dei centri urbani, che permetterà di meglio conoscere lo sviluppo edilizio del nostro Paese e di studiare con maggiore ricchezza di dati il fenomeno, così interessante dal punto di vista sociale, dell'addensamento della popolazione.

Il secondo carattere del censimento moderno è quello della *simultaneità*, consistente nel compiere la rilevazione degli abitanti nello stesso giorno in tutte le parti dello Stato, cioè da noi nel 21 aprile.

I censimenti del 1861, 1871, 1881, furono eseguiti con riferimento alla mezzanotte del 31 dicembre, mentre quelli del 1901, 1911, 1921, furono rispettivamente riferiti al 10 febbraio, al 10 giugno, al 1° dicembre.

Si è ritenuto nel 1931 più opportuno di far cadere la scelta su un giorno dell'anno che non fosse esposto ai rigori dell'inverno o ai calori eccessivi dell'estate. Nell'aprile le giornate più lunghe si prestano ad un più fecondo rendimento da parte dei commessi incaricati di distribuire e di ritirare i modelli, e la temperatura più mite rende possibile l'accesso a tutte le abitazioni, sia nelle città che nelle campagne, nella pianura come nella montagna.

E poichè soltanto tre dei censimenti fino allora fatti si erano mantenuti fedeli alla stessa data, si è ritenuto opportuno di sceglierne una per quel censimento della popolazione, che non avesse più da essere abbandonata. E nessuna altra data che quella del 21 aprile, Natale di Roma, poteva, non solo per la stagione primaverile in cui è compresa, ma per il suo alto significato politico essere meglio scelta come punto immutabile per tutti i censimenti dell'avvenire.

Questa data del resto di poco si scosta da quella che è stata scelta per il censimento generale dell'agricoltura, ed essa coincide con la ripresa generale stagionale di quasi tutte le forme di attività, in esse comprese anche quella edilizia e del movimento dei forestieri.

Anche per l'ultimo censimento, come per i precedenti, lo stato della popolazione presente si è riferito ad un momento ben precisato, vale a dire al punto astronomico del passaggio da un giorno al successivo, dal 20 al 21 aprile, tenendo però conto di quei temperamenti intesi a far sì, che potessero essere compresi, nel foglio della rispettiva famiglia, anche coloro che alla mezzanotte ne fossero assenti, perchè si trovavano a teatro, nei ritrovi, in treno, ecc., ma che durante la notte stessa e al più tardi al mattino ritornarono alla propria abitazione senza essere stati censiti altrove.

Particolari norme sono state poi stabilite dall'Istituto Centrale di Statistica d'accordo coi Ministeri competenti per la rilevazione della popolazione delle Colonie e del Possedimento delle Isole dell'Egeo.

L'universalità, che è il terzo carattere del censimento moderno, consiste nel sottoporre alla rilevazione tutti i cittadini dello Stato e delle Colonie e Possedimenti. Tutti coloro che alla data del censimento vi dimorano e anche quelli che, risiedendovi, si trovano temporaneamente fuori, sono oggetto di rilevazione. Se, col censimento del 1871, furon compresi, fra gli abitanti della Italia unificata e redenta, per la prima volta i cittadini del Veneto e quelli di Roma, e col censimento del 1921 l'indagine venne estesa ai più giusti e larghi confini che la guerra vittoriosa aveva dato all'Italia, con quello del 1931 la determinazione riuscì ancora più precisa, sia per i territori riguardanti la madrepatria, che per quelli riferentisi alle Colonie, ecc.. Il censimento del 1921, eseguito a breve distanza dalla fine della guerra, non poteva rilevare tutta la popolazione ristabilitasi nelle sue sedi naturali, fervendo ancora i lavori di ricostruzione delle regioni devastate. Il grande numero dei morti, gloriosamente caduti nella grande guerra, aveva sebbene non profondamente, turbato la distribuzione delle classi per sesso, per età, ecc. esistente prima della guerra, incidendo le classi di età dei maschi fra i 18 e i 40 anni. La riduzione delle nascite negli anni decorrenti dal 1915 al 1919, e la riduzione cospicua dei viventi dei primi anni di età, resa più sensibile dalla maggiore mortalità infantile derivante dalle sofferenze patite dalle madri nel periodo di gestazione ed allattamento, contribuivano per altra via a turbare la situazione normale. I gruppi dei coniugati delle età più giovanili risentivano a loro volta delle riduzioni nel numero dei matrimonî verificatesi durante gli anni di guerra, mentre sensibili si manifestavano ancora i vuoti lasciati dalla epidemia della febbre influenzale. Aggiungendosi a queste cause di perturbamento quelle derivanti dalla ripresa, dopo la conclusione della pace, di cospicue correnti di emigrazione, specialmente verso quelle devastate regioni della Francia, che maggiormente avevano bisogno di braccia per le opere di ricostruzione e per colmare il vuoto lasciato dai morti, si comprende come nel 1921 la popolazione rilevata col censimento fosse ben lungi dal presentare una situazione normale. Il censimento del 1931, che è avvenuto in un periodo abbastanza lontano dalla data che segna la fine della guerra mondiale, ci ha dato la rappresentazione di una Italia veramente diversa.

E ha potuto farci conoscere, se non nel loro pieno sviluppo, almeno nei primi tangibili effetti, i risultati della politica demografica lungimirante, instaurata per volere di Mussolini dal Governo Nazionale Fascista e diretta ad evitare la triste emigrazione per l'estero, a far crescere la natalità, a rendere possibile con la sempre più vasta applicazione della bonifica integrale il maggiore rendimento del suolo patrio, l'intensificazione delle culture, un'alta proporzione di abitanti nelle campagne meglio coltivate, più salutari. E col necessario complemento della rilevazione relativa alle Colonie e ai Possedimenti, sempre meglio determinati nei loro confini e più solidamente assicurati nell'interno, ci ha mostrato come anche in quelle terre lontane il nostro paese abbia potuto conseguire benefici e progressi, che non era possibile prevedere alla data dell'ultimo censimento.

La particolare importanza e l'eccezionale significato del nostro ultimo censimento, sono il risultato di due singolari aspetti della politica del Regime Fascista: la valorizzazione della statistica e la grande portata attribuita al fattore demografico. Sta in fatto che il R. D. Legge 6 novembre 1930, n. 1503, ordinando il VII Censimento generale della popolazione, ha portato, rispetto ai precedenti censimenti, interessanti novità. Anzitutto — come si è detto — è stato disposto che d'ora innanzi i censimenti, da decennali, diventino quinquennali. In secondo luogo è stato stabilito che i censimenti siano eseguiti sempre alla stessa data, che è poi particolarmente significativa: il 21 aprile. In terzo luogo, il censimento è stato felicemente esteso alla popolazione delle Colonie di diretto dominio ed ai possedimenti, ed inoltre può ben dirsi che con le nuove disposizioni siano stati ordinati maggiori e più severi controlli e predisposta una più razionale organizzazione. Basti accennare, ad esempio, alle istruzioni diramate dall'Istituto Centrale di Statistica per la ripartizione del territorio del Comune in frazioni e sezioni, per la rilevazione dei piani topografici, per la revisione delle denominazioni delle vie e della numerazione dei fabbricati, per la costituzione ed il funzionamento delle Commissioni provinciali e comunali di censimento, e via dicendo.

Altra caratteristica del nostro ultimo censimento è costituita dalla speciale indagine sulle abitazioni, la quale è stata predisposta in 422 Comuni, i principali di tutte le regioni del Regno.

2. – Molti e importanti sono gli scopi scientifici e pratici, cui mira il censimento. Preme anzitutto stabilire la quota di incremento, assoluta e

relativa, della popolazione dall'uno all'altro censimento come elemento importantissimo per giudicare dell'avvenire demografico e politico della Nazione. Sempre in materia di numero assoluto degli abitanti, è ovvio che la distribuzione geografica della popolazione presenta particolare importanza per la politica agraria e sanitaria e che la distribuzione amministrativa in regioni, provincie e comuni è non meno importante nei riguardi di molti aspetti della politica tributaria, scolastica, dei lavori pubblici, ecc. È noto che non meno di trenta leggi in Italia si basano sulla conoscenza del numero assoluto degli abitanti di ciascuna unità amministrativa: l'istituzione degli uffici finanziarî, dei posti notarili, delle scuole, ecc., è basata appunto su tale conoscenza.

Scientificamente noi possiamo apprezzare la portata di molti tra i più cospicui fenomeni demografici solo riferendoli alla popolazione dalla quale scaturiscono: così si dica della quota di incremento della popolazione stessa, dell'andamento della nuzialità, della natalità, della mortalità, ecc.; così si dica anche di molti altri fenomeni che, dal punto di vista della loro frequenza, prendono ugualmente significazione quando vengano ragguagliati alla popolazione: criminalità, analfabetismo, morbilità, mortalità, carico tributario, ecc. Ne è a dire che, partendo dai risultati di un dato censimento, sia possibile aggiornare, di anno in anno, i dati della popolazione servendosi dei registri di anagrafe dei singoli Comuni. In proposito è da tenersi presente come il servizio anagrafico, anche se bene organizzato, non possa essere assolutamente preciso, troppo complesso essendo il movimento non solo intrinseco (nascite e morti), ma anche e sopratutto estrinseco della popolazione (emigrazione, immigrazione). Il censimento diventa pertanto di volta in volta strumento per il riordinamento delle anagrafi comunali.

A parte ciò, il censimento non si limita a rilevare il numero assoluto degli abitanti, ma rileva, di ogni cittadino, parecchi elementi la cui conoscenza non soddisfa solo un sentimento di curiosità. Nessuno può contestare ad esempio la grande importanza del carattere *sesso*, delle condizioni di equilibrio tra i due sessi. Secondo i censimenti dell'immediato dopo-guerra, Inghilterra, Germania e Francia contavano rispettivamente, per ogni 1000 femmine, 912, 915, 916 maschi; ma il censimento italiano del 1921 presentò una mascolinità dei censiti di 973.

C'è poi l'*età*. Si parla di *popolazioni giovani* e di *popolazioni vecchie* e siffatte distinzioni hanno una grande importanza dal punto di vista biologico e politico. Si discute di popolazioni che invecchiano e di altre che ringiovaniscono. Orbene tutto ciò si riferisce alla maniera in cui, in percentuali, una popolazione censita si distribuisce nelle singole età e alle variazioni nel tempo di tali percentuali. E non dico della utilizzazione dei dati della popolazione per età al fine della compilazione delle tavole di mortalità. Si può anche aggiungere che solo la conoscenza di tali dati distribuiti per età può, ad esempio, permettere di calcolare il presumibile gettito di una classe di leva.

Non vi è poi bisogno di spendere molte parole per illustrare l'importanza della rilevazione dello *stato civile.* Lo studio della nuzialità presuppone la conoscenza di tale elemento, mentre lo stato civile in combinazione con l'età costituisce il presupposto per l'interessante studio della fecondità matrimoniale. Analogo presupposto hanno le ricerche sulla durata media dei matrimonî e sulla durata media della convivenza feconda dei coniugi.

Poichè sarebbe troppo lungo discorso quello che volesse riferirsi a tutti gli elementi che si rilevano dal censunento, fermiamoci con l'analfabetismo. Quanti sono gli analfabeti in Italia? La risposta più esauriente è data dal censimento. Quanti erano nel 1921 e quanti nel 1931? Ecco un problema la cui risoluzione consente di fissare in cifre i progressi realizzati in siffatto campo dal nostro paese negli ultimi anni.

3. – Una apposita Commissione fu istituita per fare opera soprattutto di volgarizzazione, di propaganda. Essa cercò di far penetrare l'importanza e l'essenza di quella formidabile e complessa opera di rilevazione statistica nelle classi incolte, nelle quali dominano, sebbene ora in misura più attenuata del passato, i sospetti e i pregiudizi: si fece gran conto della collaborazione dei maestri, dei parroci, dei direttori di cattedre ambulanti, ecc.

È noto come serpeggi sempre, in occasione delle grandi rilevazioni statistiche, la preoccupazione di presunti scopi fiscali. A parte ciò, l'ignoranza o il capriccio possono talora suggerire risposte imprecise o false. Bisognò sgombrare ll terreno da questo ciarpame e mettere il censimento nella sua giusta luce.

C'era poi il vasto campo dell'alterazione fraudolenta della verità. Comuni che si ingrossano o si rimpiccioliscono per fini particolari, sempre condannevoli; impiegati che per il miglioramento di classe di un ufficio postale o per l'aumento di un posto di notaio spingono le popolazioni comunali, quando le differenze siano piccole, sino ai limiti richiesti dallo sperato vantaggio, e via dicendo.

Tutto questo ha avuto il più efficace e severo controllo: le infedeltà e le omissioni nelle risposte costituiscono una precisa violazione di legge e sono state perciò colpite da sanzioni.

L'organizzazione tecnica del nostro ultimo censimento ha avuto per base il foglio di famiglia e, per le convivenze diverse dalla familiare (alberghi, collegi, conventi, caserme, ospedali, carceri, navi, ecc.) il foglio di convivenza.

Uno speciale questionario è stato diramato in quei Comuni, nei quali si è svolta l'indagine sulle abitazioni.

A differenza di quel che si fece negli altri censimenti, i quali in unico foglio, o in unica busta, raccoglievano le notizie riguardanti tanto la famiglia quanto la convivenza, col nostro ultimo censimento sono stati distribuiti tre fogli: un foglio per la *famiglia*, un foglio per la *convivenza*, un questionario per la indagine sulle *abitazioni*. Anche le notizie concernenti le abitazioni, in tal modo, che per i censimenti precedenti trovavano luogo nel foglio o nella busta di famiglia, sono state raccolte a parte. La creazione dei tre distinti fogli — su cui poi sono state eseguite le operazioni di spoglio e di aggruppamento a macchina — ha facilitato e semplificato il compito degli Uffici centrali.

A questo proposito si è tornata a chiarire — come sempre si fa in occasione dei censimenti — il senso che i censimenti debbono dare alla parola *famiglia* e alla parola *convivenza*.

Non è da far meraviglia se, ai fini di questa o quella determinata ricerca scientifica o pratica, le parole abbiano volta per volta ad essere definite, potendo talvolta assumere esse significato alquanto diverso da quello con cui vengono adoperate nel comune parlare. Ogni scienza, o arte, ha sovente necessità di creare una sua propria nomenclatura, ai fini della precisione e dell'esattezza. E così ha fatto e fa la demografia, che in quella parte del suo compito che è lo studio dei criterî per la raccolta dei dati a mezzo del censimento, definisce, talvolta con senso diverso da quello attribuito dal comune linguaggio, le cose che esso censimento deve contare.

Famiglia, dunque, secondo i criterî del nostro censimento è un insieme di persone che convivono sotto il medesimo tetto; si può dire anche, e fu detto: che vivono intorno al medesimo focolare, o nella medesima abitazione intorno al medesimo desco. Come si vede, nella « famiglia » (di censimento) così concepita rientrano anche coloro che abitualmente convivono con i componenti la famiglia propriamente detta senza avere con costoro vincoli di sangue nè vincoli di affinità: i domestici, ad esempio, gli istitutori conviventi, e simili. Anche i « dozzinanti » conviventi entrano a far parte di tale « famiglia » di censimento quando partecipano al pasto della famiglia. Tuttavia, il foglio su cui sono state redatte tali notizie di « famiglia » ha permesso poi, agli Uffici centrali di spoglio e aggruppamento dei dati, di isolare e contemplare le famiglie « naturali » fuori dalla massa delle altre (di censimento). È da aggiungere che le persone viventi da sole, o perchè non hanno famiglia propria, o perchè pur avendola vivono separate da essa, han costituito pure di per sè una « famiglia » (di censimento) ai fini del censimento stesso. Vivere soli presso una data famiglia senza partecipare alla vita in comune della famiglia stessa, e cioè senza prender parte abitualmente ai pasti quotidiani di essa, è anche formare una « famiglia » (di censimento).

Così concepita, la « famiglia » dà luogo, per il censimento, ad un *foglio di famiglia*. Tante « famiglie » (di censimento), tanti fogli di famiglia.

firmati ciascuno dal *capo-famiglia*, che è la persona la quale ha su di sè il carico della famiglia o che come tale è considerato per vincoli di sangue o per altre ragioni.

La *convivenza*, per contro, è l'insieme di persone riunite in alberghi, locande, dormitorî, collegi, convitti, conventi, caserme, ospedali, carceri, ospizî, navi, barche, baracche, tende e simili. Della *convivenza* così concepita e definita non fanno parte il proprietario o conduttore o direttore o altre persone di amministrazione, servizio, assistenza, custodia, ecc., quando costoro abitino *con la propria famiglia* nei locali della convivenza. Essi costituiscono allora una di quelle « famiglie » di cui sopra, e hanno quindi compilato un *foglio di famiglia* che è altra cosa del foglio di convivenza, e ben distinto da esso.

4. – Novità di rilievo, nel nostro censimento del 1931, è stata l'inclusione di domande sulla natalità della donna italiana.

Indagine di grande interesse scientifico e pratico; poichè soltanto ponendo mano a materiale statistico in tal modo raccolto, e non poggiando ragionamenti e ipotesi su valutazioni dovute a impressioni subiettive e il più delle volte fallaci, si potrà, da un canto, giungere a mettere in evidenza le leggi che governano il mutar della natalità col mutar delle condizioni in cui essa si manifesta; e soltanto da esatte notizie in tal modo ricavate si potrà d'altro canto svolgere un'efficace polltica demografica.

I passati censimenti non permettevano nè permettono tali analisi, poichè dalle indicazioni numeriche da essi raccolte circa il numero dei figli è impossibile conoscere il numero *totale* di figli avuti dalla donna maritata censita. Il nostro censimento del 1931, per contro, pur limitando l'indagine alla donna maritata, o vedova (lasciando quindi da parte le nubili, per evidenti ragioni di opportunità), si è fermato per l'appunto a far tale conteggio. L'esclusione delle nubili, necessaria, non ha portato a gravi inconvenienti nelle risultanze che dall'indagine potranno essere ricavate.

La necessità, pertanto, di conoscere il numero *totale* di figli avuti da una donna, ha fatto richiedere, nel foglio di famiglia, alla donna maritata o vedova (e anche divorziata, se trovavasi in tale condizione):

a) il numero *complessivo* dei figli avuti; e per *complessivo* s'intendono tutti — dal primo cioè all'ultimo — contando per tali anche i nati morti. Si sono esclusi soltanto gli aborti, intendendosi per aborti gli espulsi durante i primi 6 mesi di gestazione;

b) l'età della donna al momento del matrimonio (e s'intende l'età al momento del *primo* matrimonio, se la donna si coniugò più volte); e ciò per aver notizia della durata della vita matrimoniale;

c) il numero dei figli — su quelli precedentemente indicati — viventi; e ciò per aver agio di accertare in qual modo il rapporto tra sopravviventi e nati può variare col mutar delle condizioni in cui esso si verifica.

È evidente che i dati tutti, così come li abbiamo indicati, già di per se stessi forniscono indicazioni del più vivo interesse: quando verranno combinati con gli altri del foglio stesso di famiglia, e cioè età, professione e dimora della donna (e professione del marito), sarà possibile ricercare — oltre al modo con cui varia la natalità a seconda dell'età, della durata del matrimonio, e simili — come essa varia a seconda della classe sociale e come da Regione a Regione d'Italia; come varia il numero di figli sopravviventi (sul totale dei figli nati) col variare di tutte le sopra dette condizioni, o parte di esse, ecc.

Questa serie di domande darà luogo, per la prima volta in Italia, a un esame largo e soddisfacente della natalità, quale è già stato fatto in altre fra le più progredite nazioni.

5. - In quanto alla domanda concernente la religione, si noti che tutti i censimenti italiani — meno due: quelli del 1881 e del 1921 — ebbero a porre il quesito. Ma il mutare della forma della domanda da censimento a censimento ben mostra quali e quante difficoltà, d'ordine vario, si presentano quando si fa tal domanda. In diversi modi, di volta in volta, si fecero — dai censimenti nostri e da altri all'estero — tentativi per attenuare, se non per eliminare, tali difficoltà. Il punto centrale sta qui: un censimento ha forse da chiedere ai censiti quale è la religione che essi effettivamente professano, vale a dire ha da chiedere una *dichiarazione di fede*, o di non fede, religiosa; oppure ha soltanto da chiedere quale è la religione cui il censito fu aggregato per mezzo del rito o del fatto concreto che alla nascita o in momento successivo lo dichiarò, per l'appunto, aggregato a questa o a quella religione?

Non è chi non veda la differenza tra i due concetti. Nel primo caso, si chiede una dichiarazione del modo di pensare. Nel secondo, non si indaga affatto su ciò, ma si domanda soltanto secondo quale rito o fatto concreto, alla nascita o posteriormente, fu consacrata o affermata l'« appartenenza » del soggetto a una data religione. Se con nessuno, il censito dovrà rispondere: *nessuno*; e in caso di mutamento, nell'appartenenza religiosa, sopravvenuto dopo la primitiva appartenenza, il censito si riferirà — nella risposta — alla appartenenza attuale.

Tra i due accennati sistemi, il censimento del 1931 ha scelto il secondo, stimando non doversi chiedere una dichiarazione di fede, ma doversi contare soltanto i componenti dei vari gruppi, secondo un segno obiettivo di carattere, diremo così, esterno e facilmente comprensibile e visibile. Non si è chiesto, cioè, quale religione concreta e positiva effettivamente *si pratica* o *si professa*, o in quale credo speciale si ha fede: le eventuali risposte a domande di tal genere potrebbero, del resto (come nelle discussioni preparatorie ai vari censimenti in Italia e fuori fu mostrato), essere inquinate da inesattezze, potendo qualche censito credere di ricavar profitto o di aver a subire qualche danno col dare questa o quella risposta. Si chiede invece uno stato di fatto che si manifesta o si è manifestato con segni esterni, a proposito dei quali la risposta si dà quasi automaticamente, in modo semplice, netto e incapace di generare equivoci.

Senza dubbio si potevano porre e l'una domanda e l'altra (appartenenza e fede effettivamente professata) sì che dal raffronto tra le due risposte si sarebbero potute ricavare utili notizie. Ma gli inconvenienti, or ora enunciati a proposito dell'indagare sulla *credenza* professata, lo hanno sconsigliato.

Quasi tutti i censimenti dei paesi esteri, o in un modo o nell'altro, pongono il quesito della religione e sono costretti, per le difficoltà sopra indicate e per i possibili equivoci che la maggior parte dei modi di porre la domanda può portare seco, a dare chiarimenti su ciò che la domanda effettivamente significa, e quindi sul modo di rispondere. L'ultimo censimento italiano, ponendo la domanda nel senso sopra spiegato, non poteva che limitarsi, nelle istruzioni, al breve avvertimento: « Chi ha appartenuto successivamente a più religioni, dovrà indicare l'ultima ».

Certamente, anche qui si sarebbero potute fare due domande: sull'ultima e sulla precedente religione. Ma si sarebbe andati incontro a un'indagine — riguardante le conversioni — che non è necessario (data la delicata natura dell'indagine stessa) sia compresa tra gli scopi di un censimento. Tanto più quando il censimento, circa il quesito della religione, si prefigge, come è stato detto, non l'indagine sulla fede o credenza, ma sulla « appartenenza » nel senso sopra detto.

A) II. - Appunti storici sull'applicazione del metodo statistico ai fatti economici e demografici.

1. – I primi cultori sistematici dell'analisi quantitativa (per grandi masse) dei fatti economici furono i cosidetti aritmetici politici, cioè i seguaci degli studî di un mercante londinese: John Graunt (1620-74).

Già W. Petty (1623-87), amico e ammiratore di Graunt, dopo essersi occupato largamente di problemi demografici e finanziarî inglesi, si era proposto di conoscere « con numeri, pesi e misure » il valore dei terreni, la spesa media degli abitanti, ecc.; scrivendo infine quella *Political Arithmetic* (pubblicata nel 1691), che diede appunto il nome al nuovo indirizzo di studî [1].

[1] A. Gabaglio, *Teoria generale della statistica*. Vol. I, parte storica. Milano, Hoepli, 1888. – C. H. Hull, *The economic writings of Sir William Petty*. Cambridge, Univ. Press, 1899.

Alle sue ricerche seguirono quelle della triade: Vauban (il famoso maresciallo francese: 1633-1707), Gregory King (1648-1712) e Davenant (1656-1714).

Del primo — perseguitato autore del *Projet d'une dixme royale* e che invero aveva avuto nelle sue analisi un grande predecessore, Sully, e il vantaggio deli' opera degli *intendenti*, già felicemente avviata da Colbert — si ricordano ancora il progetto di un'accurata stima catastale e le raccomandazioni di creare organi ufficiali preposti alla rilevazione statistica dei fatti demografici ed economici.

Il secondo — autore di *Natural and political observations and conclusions upon the state and conditions of England* (1696) — è ancora ricordato per una famosa tabella, con la quale cercava di stabilire una relazione tra l'entità del raccolto del frumento e il suo prezzo di mercato.

Il terzo, seguace dello stesso indirizzo ma critico dei risultati sin'allora raggiunti, scrisse un *Essay upon the probable methods of making a people gainers in the balance of trade* (1699) e alcuni *Discours* (1698), dove si definisce l'aritmetica politica come « l'arte di ragionare coi dati sulle cose attinenti allo stato ».

Per più di un secolo le indagini di economia induttiva si aggirarono intorno al pensiero di quei pionieri — emersero il nostro Carli col suo *Breve ragionamento sopra i bilanci economici delle nazioni* (1769) e, prima ancora, P. Verri con la costruzione statistica dei *Bilanci del commercio dello Stato di Milano* (1752 e '62) testè editi da « La Riforma Sociale » (¹) — finchè in Inghilterra alcuni studiosi, spinti allo studio delle variazioni dei prezzi, che da tempo preoccupavano popoli e governi, e su cui già si era formata un'abbondante letteratura al fine pratico di porvi rimedio (²), intesero il bisogno di ragionare su basi più reali della logica astratta, il cui impiego si andava diffondendo con lo sviluppo dell'economia classica anche in tema di circolazione monetaria, e più solide di testimonianze o dati di fatto incompostamente e sporadicamente raccolti, su cui in materia economica si era lungamente fermato lo studio degli aritmetici politici.

Ma, data la poca estensione nel tempo e la rudimentalità delle statistiche ufficiali, era naturale che si cominciasse con la raccolta e l'elaborazione dei dati, e che le prime analisi di essi presentassero carattere frammentario. Ciò spiega perchè, accanto alle opere descrittive colossali

---

(¹) In « Piccola collezione di scritti inediti o rari di economisti », diretta da Luigi Einaudi. Torino, 1932.

(²) Cfr.: L. Cossa, *Introduzione allo studio dell'economia politica*, III ed., Milano, 1892, p. 196 e *passim*. Intorno alle proposte di G. Shuckburg Evelin (1798), di J. Lowe (1822), di G. Poullet Scrope (1833), di G. Porter (1836), cfr.: W. S. Jevons, *Money and the mechanism of exchange*, London 1875, Chap. XXV, 22° Ed. 1909; e id., *Investigations in currency and finance*, London, 1884, 2ª Ed., London, 1909, p. 115.

di Tooke sulla storia dei prezzi dal 1784 [1], e poscia di Rogers sulla storia dell'agricoltura e dei prezzi dal 1259 [2], siano apparse e per un pezzo permangano solo analisi minute, vaghe e senza sistema, di statistiche economiche.

Nondimeno il primo carattere, che in essi chiaramente si vede rilevato, è l'esistenza di fluttuazioni *periodiche* della circolazione monetaria, dei prezzi e dei fatti economici in genere.

Ma non si trattava di un'improvvisa rivelazione: erano troppo sensibili i danni che di tempo in tempo si risentivano, e troppo vivaci le dispute teoriche e pratiche che insorgevano, per non essere già nota quella periodicità!

Porter [3], ad esempio, arriva persino a sorprendersi nel constatare come, dopo tanto discutere di economisti teorici da un lato, pur essi scissi in due scuole ben note, e di pratici dall'altro, dopo tanti provvedimenti dello Stato, a cominciare dalla costituzione del famoso *Bullion Committee* del 1810, non si fosse ancora riusciti ad alcuna conclusione sui mezzi per evitare « le ruinose alternative, adesso costantemente ricorrenti ». E soggiunge: « In ogni occasione in cui il mercato monetario è stato soggetto ad uno di questi parossismi, gli uomini esperti si sono proposti di spiegarne le cause e di poterne in futuro evitar subito i danni; e a chi avrà la pazienza di esaminare gli argomenti e le asserzioni usate in ambo i lati della controversia (pratici e teorici) deve far meraviglia osservare come ci sia un'identità di opinione, e quasi di espressione, tra gli scrittori che si sono schierati dallo stesso lato della questione in differenti periodi; cosicchè negli opuscoli pubblicati nel 1811 o nel 1826 egli ritroverà tutti gli argomenti principali e riferentisi alle medesime serie di circostanze che hanno formato la sostanza degli opuscoli scritti nel 1837 ».

Ciò che, dunque, rimaneva avvolto nel mistero era la causa di tali fluttuazioni; ed alla ricerca di essa conversero le prime analisi induttive.

J. Wilson, fondatore del grande giornale *The Economist*, in un volumetto edito nel 1840, citato da Jevons [4], già attribuisce « il frequente ricorso di periodi di eccitamento e di depressione negl'interessi monetarî e commerciali del paese alle enormi fluttuazioni dell'ammontare dei

---

(1) T. Tooke, *History of prices, and of the state of circulation from* 1792-1856, completata da G. Newmarch, London, 1838-57, e testè ristampata.

(2) T. Rogers, *A history of agriculture and prices in England, from the year after the Oxford Parliament* (1259) *to the commencement of the Continental War* (1793), Oxford, 1866, 1875, 1887.

(3) G. R. Porter, *The progress of the nation, in its various social and economical relations from the beginning of the nineteenth century to the present time*, London 1836, 1838, Sect. III, Chap. XII: *Currency*.

(4) J. Wilson, *Fluctuations in currency, commerce and manifactures, referable to the corn laws*, London, 1840.

mezzi, che da tempo a tempo sono occorsi al paese per la necessaria sussistenza della vita; o, in altre parole, alle fluttuazioni dei prezzi dei generi alimentari (grano, ecc.) ».

Poco dopo CLARKE (1) vede nel mondo fisico, senza peraltro essere riuscito a rinvenirla in fenomeni meteorologici o astronomici, la causa di questi « periodi di eccitazione e di giuoco, troncati in un istante e risolti in periodi di sofferenze e di perdite », dei quali egli naturalmente non si sorprende, come di cosa « così risaputa che sarebbe notevole se essa destasse meraviglie », e che, del resto, dice di avere già rilevato in uno studio pubblicato sin dal 1838, dove, prima della crisi del 1847, avrebbe notato periodi circa decennali facenti capo alle crisi del 1837, 1826, 1815, 1804, 1793.

È tale l'importanza e l'estensione ch'egli attribuisce a questo campo di ricerche, che non esita a farne oggetto di una nuova scienza: l'*economia fisica*.

Ma GILBART già aveva anche parlato di un ciclo annuale della circolazione. In seguito alla crisi del 1837, egli, riferendo in seno ad un *Select Committee of the House of Commons*, formato nel 1840, preliminarmente all'atto di PEEL, allo scopo di accertare gli effetti prodotti sulla circolazione monetaria del Regno Unito dall'opera delle varie banche di emissione, aveva compilato delle tabelle statistiche, poi riprodotte in uno studio speciale (2), che, tra l'altro, gli avevano fatto rilevare una quasi generale fluttuazione mensile, facente capo all'autunno, della circolazione della Banca d'Inghilterra, delle *Country Banks*, e delle Banche d'Irlanda e di Scozia.

Furono queste fluttuazioni mensili, attribuite poscia da LANGTON all'influenza delle stagioni sul commercio (3), che, insieme ad una quantità di scritti minori dei quali s'è perduta la traccia, spianarono la via agli studî di JEVONS (4).

I quali, infatti cominciano con una breve analisi (1862) delle fluttuazioni medie mensili del saggio dello sconto della Banca d'Inghilterra, nei periodi 1845-61 e 1824-61, e dei fallimenti, del prezzo del consolidato e del grano, poi estesa (1866) in uno studio sistematico sulle crisi autun-

(1) H. CLARKE, *Physical economy. A preliminary inquiry into the physical laws governing the periods of famines and panics*, in « Railway Register », 1847, ampiamente analizzato da JEVONS, *op. cit.*,

(2) J. W. GILBART, *The laws of the currency*, in « Foreign and Colonial Review », 1844; rifuso del suo trattato: *The history, principles and practice of banking*, ediz. riveduta da E. Sykes, London, 1907, Vol. II, pp. 114, 157, 181, 241. La prima edizione di questo trattato, riguardante solo la pratica bancaria, risale al 1827.

(3) W. LANGTON, *Observations on a Table showing the balance of account between the mercantile public and the Bank of England*, in « Transactions of the Manchester Statistical Society », 1857-58.

(4) Raccolti nel cit. volume postumo: *Investigations in currency and finance*, con un'Introduzione di H. S. FOXWELL.

nali, fatto principalmente in base ai bilanci della Banca d'Inghilterra; e, come di cosa risaputa, vengono ad occuparsi delle variazioni decennali del credito, dei prezzi, del valore dei mattoni fatti nel Regno Unito, dei carichi di legname importatovi ecc., e precisamente a proposito della compilazione dei numeri indici dei prezzi.

In questo prezioso contributo alla statistica metodologica (1863, 1865, 1869) Jevons, con metodo più corretto di quello già usato da Porter (1) nei suoi numeri indici mensili dei prezzi di cinquanta articoli in Inghilterra pel periodo 1833-37, ed anche in base ai dati di Tooke, analizza lungamente l'influenza delle «notissime grandi fluttuazioni decennali» per eliminarla dallo studio dell'andamento generale del valore dell'oro.

I suoi studi posteriori su tali grandi fluttuazioni (1875, 1878), fuorchè per una ricerca storica da lui fatta sui dati di Rogers, non hanno più, dunque, carattere espositivo, ma lo scopo di ripudiare le vedute di MILLS (2), che pure nel 1867 aveva attribuito a cagioni mentali e psicologiche i cicli di credito, e precisamente a stati successivi di scoraggiamento, fiducia, eccitamento, delusione e panico nella vita economica del paese; e di cogliere, invece, una speciale relazione, già sostenuta da HERSCHEL e da CARRINGTON, tra le variazioni delle macchie solari e le fluttuazioni dei prezzi agricoli: di dimostrare finalmente che la causa prima dei cicli decennali fosse quella già intraveduta da Clarke e precisamente una periodicità decennale dei raccolti, specie in India, come conseguenza di un'analoga periodicità delle stagioni e delle macchie solari.

A parte l'arditezza di tale spiegazione fisica — forse derivante da una seducente analogia con alcune fluttuazioni stagionali —, è di capitale importanza rilevare che quelle analisi di Jevons erano ormai il prodotto di un'intima e profonda convinzione della necessità, ch'egli già sentiva e andava predicando in tutti i suoi scritti (3), di un uso esteso dell'induzione quantitativa per lo sviluppo ulteriore della scienza economica; e che, secondo afferma Foxwell, buona parte della sua vita egli avrebbe impiegato intorno ad un gran trattato di economia politica, rimasto incompiuto, dove avrebbe lavorato su un immenso materiale statistico, raccolto nello spazio di più di venti anni, e del quale le sue analisi induttive non sarebbero che minuti frammenti.

Ma nessuna importanza egli attribuì alle formule empiriche, nonostante che nella sua cit. *Theory* (IV, 16) egli avesse pure rappresentato con un'espressione analitica le variazioni del prezzo del frumento in fun-

---

(1) G. R. PORTER, *The progress of the nation, op.* e *loc. cit.*

(2) J. MILLS, *On credit cycles and the origin of commercial panics*, in «Transactions of the Manchester Statistical Society», (1867-68).

(3) Cfr. l'Introduzione alla sua *Theory of political economy* (1871), ed i suoi *Principles of science. A treatise on logic and scientific method*, London, 1874, Vol. 2°, pp. 457 e segg.

zione dell'entità del raccolto in base ai dati della tabella di King (da lui però attribuita a Davenant); e che inoltre fosse riuscito ad ottenere dei prezzi teorici strettamente concordanti con quelli osservati.

Le fluttuazioni decennali venivano nel contempo studiate con maggiore ampiezza da Juglar (1), sulla base del portafoglio, degl'incassi ed in genere dei bilanci delle grandi banche, ed anche su altri ordini di fenomeni, come prezzi, commercio estero, matrimonî, nascite, morti, entrate pubbliche, quotazioni dei fondi pubblici, in un volume rimasto famoso per il sincronismo, che per la prima volta mette in luce, tra la periodicità di quelle crisi nei paesi più progrediti, quali la Francia, l'Inghilterra e gli Stati Uniti d'America. È però da lamentare la pesantezza e l'oscurità, talvolta notevoli, dell'esposizione, e la rudimentalità dei procedimenti adottati, che spesso ad esempio confondono le oscillazioni cicliche con quelle accidentali e con l'andamento generale delle serie, e si dimostrano assolutamente inferiori alla limpida metodologia jevoniana.

Così venne a concretarsi un complesso di nozioni scientifiche, che in seguito assunse il nome di teoria delle crisi e segnò il primo passo nello studio della dinamica economica.

Invero non si trattava ancora di una teoria, ma di un insieme di osservazioni e di prime conclusioni, che abbisognavano di una maggiore estensione ed elaborazione per poter essere composte a sistema. Senonchè l'importanza di queste nuove ricerche non fu percepita dagli economisti, nonostante la considerevole influenza sul pensiero economico che Foxwell si riprometteva dalla pubblicazione degli studî di Jevons, le crescenti rilevazioni di statistica economica che nel contempo gli Stati venivano operando ed i progressi della metodologia statistica.

Infatti quei metodi languirono lungamente negli studî prevalentemente descrittivi e nei calcoli, pur notevolissimi anzi fondamentali per il risveglio seguente, di Giffen, Palgrave, Sauerbeck, Le Play e seguaci, Newmarck, De Foville, Soetbeer, Engel, Lexis, Conrad, e di altri studiosi, in parte sparsi su riviste ed annali; nelle raccolte storiche di Rogers, Cunningham, Roscher, Soetbeer, Schmoller, Levasseur, D'Avenel, e di altri, anche per l'evo antico — ricerche più o meno feconde di dati quantitativi, che però in massima parte ancora attendono impiego —; in taluni studî sporadici, come ad es. quelli di Bela Weitz (2) sulle relazioni tra le variazioni della nuzialità, ed anche della mortalità, e le variazioni dei prezzi del grano in Prussia, Austria,

(1) C. Juglar, *Des crises commerciales et de leur rétour périodique en France, en Angleterre et aux Etats-Unis*, Paris, 1862, 2ª ediz. 1889; e in « Bulletin de l'Institut International de Statistique », tome XIII, 4e Livr., Rome, 1903.

(2) In « Statist. Monatschrift », *V Jahrg XI Heft*, riportata in « Annali di Statistica », Roma, 1880.

Francia, Inghilterra, Belgio, Svezia, Finlandia; di KREMP [1], che, nonostante le deficienze delle prime statistiche agrarie, tenta di analizzare l'influenza dei raccolti e dei depositi, del commercio internazionale ecc., sui prezzi del grano in Russia, Sassonia, Würtemberg, Austria, Ungheria, Francia; di POYNTING [2] tra le variazioni dei prezzi del grano e le importazioni di cotone e seta in Gran Bretagna; di RAWSON [3] tra i matrimonî ed i raccolti di grano nella Svezia; ed infine nelle ricerche di alcuni scrittori tedeschi sulle relazioni tra affitti e redditi, tra redditi e patrimonî, in base ai dati statistici offerti dal regime fiscale di taluni Stati germanici [4].

Ciò sinchè quel nuovo metodo di ricerca non sboccò nella ricerca di indici dello stato di benessere o di malessere dei popoli, nella così detta *semiologia economica*. Le analisi di ENGEL [5], di NEUMANN SPALLART [6], il grafico di DE FOVILLE [7], il tentativo del nostro BENINI [8], ed anche i calcoli di DES ESSARS [9], che però hanno riguardo solo alle crisi, si propongono appunto di trovare l'indice migliore, semplice o complesso, delle condizioni economiche di un paese; ad esse seguì un'analisi di PANTALEONI [10] sul carattere, i metodi ed i limiti della semiologia economica.

Si può dire che a PARETO si debba la prima applicazione dell'idea jevoniana di un'induzione quantitativa negli studî economici, sistematicamente condotta, e l'impostazione teorica del fenomeno delle crisi.

Infatti tutta la seconda parte, che è la più estesa, del suo primo trattato [11] ha appunto lo scopo di avvicinare la teoria economica pura alla realtà, mediante ulteriori deduzioni e induzioni quantitative per la prima volta accertate o già note o rielaborate su dati statistici raccolti da fonti originali e derivate ed anche storiche, come ad es. gli studî cit. di ROGERS, LEVASSEUR ecc.

È ancora vivo il ricordo dell'impressione profonda prodotta da quell'originale trattato, e dalle suggestive analisi dell'andamento dei capitali personali (popolazione), dei prezzi, del commercio internazionale, delle

---

(1) I. H. KREMP, *Ueber den einfluss des ernteausfalls auf die getreidepreise während der jahre* 1846-1876 *in den hauptsächlichsten ländern Europas*, Jena 1879, riportata come sopra.

(2) In « Journal of the Statistical Society », London, 1884.

(3) In Journal cit., 1885.

(4) Per la bibliografia, cfr. C. GINI, *Indici di concentrazione e di dipendenza*, cap. X e XI, in « Biblioteca dell'economista », serie V, Vol. XX.

(5) E. ENGEL, *La consommation comme mesure du bien-être des individus, des familles et des nations*, in « Bulletin de l'Institut Intern. de Statistique », 1887.

(6) F. NEUMANN-SPALLART, *Mesure des variations de l'état economique et social des peuples*, in « Bulletin », cit., id.

(7) In « Journal de la Société de Statistique de Paris », 1888.

(8) In « Giornale degli economisti », 1892.

(9) In Giornale cit., 1895.

(10) M. PANTALEONI, *Osservazioni sulla semiologia economica*, in « Revue d'economie politique », 1892; ristampate in *Scritti vari di economia*, Vol. I, Palermo, 1904.

(11) V. PARETO, *Cours d'economie politique*, Lausanne 1896-1897, *op. cit.*, 2 Voll.

crisi e dei redditi, condotte con l'impiego di una corretta metodologia, di cui in seguito Pareto doveva dare dei saggi fondamentali.

Soprattutto grande, anche per l'abbondante e pregevole letteratura statistica derivatane sui rapporti di distribuzione, concentrazione ecc., fu l'interesse destato dalla scoperta paretiana di un'approssimazione soddisfacente, che in molti casi si osservava tra i dati dei contribuenti alle imposte generali sui redditi, opportunamente classificati in ordine di grandezza dei redditi, e i dati teorici rispettivi desunti dall'adattamento di una funzione iperbolica.

Tale risultato — di cui ci siamo occupati nella Parte I, Cap. IV di questo manuale — veniva a rischiarare alcune sommarie indagini che quasi contemporaneamente faceva AMMON (1) sulle distribuzioni dei contribuenti all'imposta generale sui redditi in Sassonia allo scopo di dimostrarne semplicemente l'analogia con le distribuzioni degli uomini secondo i caratteri intellettuali, che già GALTON (ma anche QUETELET) aveva sostenuto seguissero la curva degli errori accidentali.

Seguendo queste nuove direttive, gli studî di BENINI sulla curva paretiana dei redditi e sulle sue relazioni con quella dei patrimonî ereditarî, sulle rappresentazioni analitiche di taluni consumi in funzione dei prezzi rispettivi — pure, questa, evidente derivazione di alcune considerazioni che Pareto aveva fatto sulle ripercussioni esercitate dalle variazioni di prezzo del sale e dello zucchero in taluni paesi sui consumi rispettivi, al fine di indurre l'entità dei redditi minimi (*Cours*, II, § 937) — finirono col proclamare l'esistenza di un nuovo campo di ricerche puramente *empiriche*, unicamente fondate sul metodo statistico: *l'economia politica induttiva o empirica* (intendi: *quantitativa*) (2).

Sia gli studî semiologici, che questa proclamata economia induttiva, hanno goduto largo favore in Italia e fuori dai cultori di statistica, e possiamo annoverare, tra le principali e prescindendo da quelle di carattere prevalentemente metodologico o aventi per obietto la semplice raccolta dei dati, le elaborazioni di JULIN (3), SORER (4), MORTARA (5),

---

(1) OTTO AMMON: *Die gesellschaftsordnung und ihre natürlichen grundlagen*, Jena 1895; trad. francese, Paris, 1900.

(2) R. BENINI, *Di alcune curve descritte da fenomeni economici aventi relazioni con la curva del reddito e con quella dei patrimoni*, in « Giornale degli economisti » 1897: id, *Sull'uso delle formule empiriche nell'economia applicata*, in detto, 1907; id. *Una possibile creazione del metodo statistico*: « *L'economia politica induttiva* », in detto, 1908.

(3) A. JULIN, *The economic progress of Belgium from* 1880 to 1908, in « Journal of the Stat. Society », London, 1911.

(4) R. SORER, *Einige indexzahlen zur wirtschaftlichen entwicklung Oesterreichs*, in « Bulletin de l'Inst. Intern. de Stat. », 1913, cfr. pure « Giornale degli economisti », 1914, dove l'autore dà la misura delle intense correlazioni che intercedono tra alcuni indici dello sviluppo economico dell'Austria.

(5) G. MORTARA, *Sintomi statistici delle condizioni economiche d'Italia, e... delle condizioni economiche della Francia*, in « Giornale degli economisti », 1914.

ed inoltre di BABSON, FLUX ecc., da una parte, e dall'altra i tentativi di BRESCIANI (1), GINI (2), DEL VECCHIO (3), LENOIR (4).

È però da notare che la molteplicità dei dati semiologici, via via assunti nel primo genere di ricerche, e la necessità di analizzare anzitutto la natura e l'andamento di tutte le rilevazioni economiche di un paese, di determinarne le reciproche dipendenze, ossia le relazioni funzionali, hanno in fatto trasformato la ricerca del dato semiologico definitivo in una ricerca di economia induttiva nel senso sopra riferito.

Questa compenetrazione dei due generi di studio non ha mancato di condurre a notevoli risultati, poichè ha dimostrato il poco valore di talune equazioni empiriche riducenti ad una o due o tre variabili le relazioni intercedenti tra tutti i fatti economici; poichè, insomma, ha confermato le vedute della cosidetta teoria dell'equilibrio economico, che pone a base di fatto delle sue indagini l'interdipendenza dei fatti economici, e la necessità di tener conto dei risultati dell'analisi astratta.

Ad esempio, gli studî citati di Mortara hanno posto in rilievo una correlazione, tra i consumi delle merci più frequentemente considerate nei precedenti calcoli induttivi ed i prezzi rispettivi, molto minore di quella strettissima, che invece lega quei consumi ad alcuni indici sintetici delle condizioni economiche, che appunto riflettono, comunque, l'azione di una quantità di altri fattori influenti.

Tal genere d'indagini già s'incontra nelle opere di TUGAN-BARANOWSKY (5), AFTALION (6), MITCHELL (7) e, più recentemente, di H. L. MOORE (8) che si propone di « utilizzare i nuovi metodi statistici e le più recenti teorie economiche per ricavare dai dati riferentisi ai salari, o nuove verità o altre verità sotto una nuova forma che renda possibile di metterle in relazione feconda con le uniformità della scienza econo-

---

(1) C. BRESCIANI, *Il rapporto fra pigione e reddito secondo alcune recenti statistiche*, in « Giornale degli econ. », 1906.

(2) C. GINI, *Prezzi e consumi*, in « Giorn. degli econ. », 1910; e *Indici di concentrazione e di dipendenza*, cap. X e XI, in « Bibl. dell'econ. », op. cit.

(3) G. DEL VECCHIO, *Relazioni fra entrata e consumi*, in « Giorn. degli econ., 1912, in cui si tenta la determinazione analitica di una relazione posta da Engel.

(4) M. LENOIR, *Etudes sur la formation et le mouvement des prix*. Paris, 1913.

(5) M. TUGAN-BARANOWSKY, *Studien zur theorie und geschichte der handelskrisen in England*. Jena, 1901.

(6) A. AFTALION, *Les crises périodiques de surproduction*, Paris, 1913.

(7) W. C. MITCHELL, *Business cycles* in « Memoirs of the University of California », 1913. Cfr. la recente edizione del 1927.

(8) H. L. MOORE, *Laws of wages. An essay in statistical economics*, New York. Macmillan, 1911. Cfr. inoltre: *The statistical complements of pure economics*, in « The Quart. Journ. of Economics », 1908, dove MOORE espone il pensiero di COURNOT e di EDGEWORTH a questo riguardo, descrive sommariamente i principali procedimenti statistici usati nell'induzione economica, e fa voti per una conciliazione della teoria economica col metodo statistico. La sua ultima opera: *Synthetic economics*, New York, Macmillan, 1929, s'inspira a tali vedute.

mica »; e specialmente nelle opere di I. FISHER che, con gli studî sul saggio dell'interesse e sul potere acquisitivo della moneta (1), si propone appunto di utilizzare vasti materiali statistici con l'uso della metodologia statistica, ma sulla base e col sussidio incessante di premesse teoriche note.

Qualunque possano essere i pareri sul valore di tali indagini, non crediamo si possa negare ch'esse abbiano rischiarato molte idee, ed abbiano spianato la via agli studî di dinamica economica, che dopo la guerra mondiale hanno riscosso sì grande favore.

2. – Riguardo all'analisi quantitativa (per grandi masse) dei fatti demografici, ai quali si rivolse in ispecial modo l'attenzione dei primi cultori di calcolo delle probabilità, ci limitiamo a riportare cronologicamente nella tavola seguente alcuni eventi particolarmente importanti e di per sè eloquenti (2).

Anni	Eventi	Autori o fonti bibliografiche
1532	Prima compilazione di statistiche settimanali dei morti a Londra.	Secondo C. H. HULL, *Econ. Writ. of Sir W. Petty*, pagina LXXXI.
1539	Inizio della rilevazione ufficiale dei battesimi, matrimonî e morti in Francia.	Secondo F. FAURE, *Hist. Stat.*, pag. 242.
1608	Inizio del registro parrocchiale in Svezia.	Secondo E. AROSONIUS, *Hist. Stat.*, pag. 537.
1662	Pubblicazione delle « *Natural and political observations mentioned in a following index and made upon the bills of mortality* ».	JOHN GRAUNT (1620-1674).

(1) I. FISHER, *The rate of interest*, New York, 1907; id, *The purchasing power of money*, New York, 1911; ed inoltre i noti studi di KEMMERER, KINLEY, ENGLAND, ecc.

È notevole vedere, in questo periodo, tratto dal dimenticatoio, in cui si trovava sin dai tempi di JEVONS, lo studio delle variazioni stagionali dei fatti economici, principalmente ad opera di KEMMERER, *Seasonal variations in the relative demand for money and capital in the United States*, New York, 1910, pubb. della « National Monetary Commission ». Cfr. pure le ricerche di HOOKER, *The meat supply in the United Kingdom*, in « Journal of the Stat. Society », 1909; e di CONRAD, *Die Monatspreise des getreides*, in « Jahrbücher für Nationalökonomie und Statistik », III Folge, Bd. IX. Gli studi di Fisher e di Kemmerer provocarono una verifica critica di BRESCIANI, *Le variazioni cicliche dei prezzi*, in « Annali del Seminario Giuridico » della R. Università di Palermo, 1913; e *Ueber die jahreszeitlichen schwankungen des gesamtwarenpreisniveaus*, in « Weltwirtschaftliches Archiv », Jena, 1914.

(2) Abbiamo tratto profitto da un elenco — alquanto incompleto nella bibliografia e che abbiamo avuto cura d'integrare, per quanto ci è stato possibile — pubblicato da R. PEARL, *Introduction to medical biometry and statistics*, Philadelphia and London, Saunders, 1927. Esso fu ricavato principalmente da uno spoglio del « Journal of the Roy. Stat. Society ». Le citazioni di *Hist. Stat.* si riferiscono al volume: *The History of Statistics, their development and progress in many countries*. New York, Macmillan, 1918.

Anni	Eventi	Autori o fonti bibliografiche
1666	Primo censimento del Canada (il primo censimento demografico moderno).	Secondo E. H. GODFREY, *Hist. Stat.*, pag. 179.
1693	Pubblicazione di « *Estimate of the degrees of mortality of mankind* », in « Philosophical Transactions of the Royal Society ».	EDMUND HALLEY, (1656-1742).
1713	Pubblicazione della « *Ars Conjectandi* ».	JACQUES BERNOULLI (1654-1705).
1718	Pubblicazione della « *Doctrine of chances* ».	A. DE MOIVRE (1667-1754).
1735	Inizio delle rilevazioni demografiche in Norvegia.	Secondo A. N. KIAER, *Hist. Stat.*, pag. 447.
1741	Pubblicazione di « *Die göttliche ordnung in der veränderungen des menschlichen geschlechts aus der geburt, dem tode und der fortpflanzung desselben erwiesen* ».	JOHANN PETER SÜSSMILCH (1707-1767).
1746	Pubblicazione della prima tavola francese di mortalità sotto il titolo « *Essai sur les probabilités de la durée de la vie humaine.*	A. DEPARCIEUX (1703-1768).
1748	Inizio della statistica demografica ufficiale in Svezia.	Secondo E. AROSONIUS, *Hist. Stat.*, pag. 540.
1749	Primo censimento demografico completo in Svezia.	Secondo W. S. ROSSITER, *A Century popul. growth*, pagina 2.
1753	Primo censimento demografico in Austria.	Secondo R. MEYER, *Hist. Stat.*, pag. 85.
1766	Pubblicazione delle « *Tables of mortality based upon the Swedish population* ».	WILHELM WARGENTIN (1717-1783).
1769	Primo censimento demografico di Danimarca e Norvegia.	Secondo A. JENSEN, *Hist. Stat.*, pag. 201.
1790	Pubblicazione delle « *Riflessioni sulla popolazione delle nazioni per rapporto all'economia nazionale* ».	GIAMMARIA ORTES (1713-1790)
1790	Primo censimento demografico federale negli Stati Uniti.	
1801	Primo censimento demografico completo in Gran Bretagna.	Secondo W. S. ROSSITER, *A Century pop. growth.*, l. c.
1801	Primo censimento demografico completo in Francia.	Secondo W. S. ROSSITER, op. cit.
1812	Pubblicazione della « *Theorie analytique des probabilités* ».	PIERRE SIMON LAPLACE (1749-1827).
1825	Pubblicazione del « *Memoire sur les lois des naissances et de la mortalité à Bruxelles* ».	ADOLPHE QUETELET (1796-1874).

Anni	Eventi	Autori o fonti bibliografiche
1829	Primo censimento demografico del Belgio.	Secondo A. JULIN, *Hist. Stat.*, pag. 128.
1832	Pubblicazione delle « *Recherches sur la reproduction et sur la mortalité de l'homme au different ages et sur la population de la Belgique* ».	ADOLPHE QUETELET e E. SMITS.
1834	Si fonda la « Royal Statistical Society » (London).	
1835	Pubblicazione di « *Sur l'homme et le developpement de ses facultés, ou essai de physique sociale* ».	ADOLPHE QUETELET (1796-1874).
1837	Pubblicazione delle « *Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précedées des règles générales du calcul des probabilités* ».	S. D. POISSON (1781-1840).
1837	Rilevazione civile della statistica demografica in Inghilterra. Istituzione dell'ufficio del Registrar General.	Secondo A. BAINES, *Hist. Stat.*, pag. 370.
1839	Organizzazione della « American Statistical Association » (New York).	Secondo *Hist. Stat.*, pag. 3.
1861	Primo censimento demografico completo in Italia.	
1869	Pubblicazione di « *Hereditary genius* ».	FRANCIS GALTON (1822-1907).
1869	Si fonda la « Societé de Statistique de Paris ».	
1877	Pubblicazione di « *Zur theorie der massenerscheinungen der menschlichen gesellschaft* »	WILHELM LEXIS.
1885	Si fonda all'Aia lo « Institut International de Statistique ».	
1890	Introduzione degli spogli meccanici nei censimenti (John S. Billings e Herman Hollerith).	
1901	Pubblicazione del primo fascicolo di « Biometrika ».	FRANCIS GALTON, KARL PEARSON, W. F. R. WELDON, C. B. DAVEMPORT.
1910	Pubblicazione di « *An Introduction to the theory of statistics* ».	G. UDNY YULE.
1911	Pubblicazione di « *A problem in age-distribution* », in « The London, Edinburg and Dublin Philosophical Magazine ».	A. J. LOTKA (e F. R. SHARPE).
1917	Pubblicazione di « *The mathematical theory of population* ».	G. H. KNIBBS.

B) I. – Potenze dei numeri interi sino a 100.

n	n^2	n^3	n^4	n^5	n^6
1	1	1	1	1	1
2	4	8	16	32	64
3	9	27	81	243	729
4	16	64	256	1 024	4 096
5	25	125	625	3 125	15 625
6	36	216	1 296	7 776	46 656
7	49	343	2 401	16 807	117 649
8	64	512	4 096	32 768	262 144
9	81	729	6 561	59 049	531 441
10	100	1 000	10 000	100 000	1 000 000
11	121	1 331	14 641	161 051	1 771 561
12	144	1 728	20 736	248 832	2 985 984
13	169	2 197	28 561	371 293	4 826 809
14	196	2 744	38 416	537 824	7 529 536
15	225	3 375	50 625	759 375	11 390 625
16	256	4 096	65 536	1 048 576	16 777 216
17	289	4 913	83 521	1 419 857	24 137 569
18	324	5 832	104 976	1 889 568	34 012 224
19	361	6 859	130 321	2 476 099	47 045 881
20	400	8 000	160 000	3 200 000	64 000 000
21	441	9 261	194 481	4 084 101	85 766 121
22	484	10 648	234 256	5 153 632	113 379 904
23	529	12 167	279 841	6 436 343	148 035 889
24	576	13 824	331 776	7 962 624	191 102 976
25	625	15 625	390 625	9 765 625	244 140 625
26	676	17 576	456 976	11 881 376	308 915 776
27	729	19 683	531 441	14 348 907	387 420 489
28	784	21 952	614 656	17 210 368	481 890 304
29	841	24 389	707 281	20 511 149	594 823 321
30	900	27 000	810 000	24 300 000	729 000 000
31	961	29 791	923 521	28 629 151	887 503 681
32	1 024	32 768	1 048 576	33 554 432	1 073 741 824
33	1 089	35 937	1 185 921	39 135 393	1 291 467 969
34	1 156	39 304	1 336 336	45 435 424	1 544 804 416
35	1 225	42 875	1 500 625	52 521 875	1 838 265 625
36	1 296	46 656	1 679 616	60 466 176	2 176 782 336
37	1 369	50 653	1 874 161	69 343 957	2 565 726 409
38	1 444	54 872	2 085 136	79 235 168	3 010 936 384
39	1 521	59 319	2 313 441	90 224 199	3 518 743 761
40	1 600	64 000	2 560 000	102 400 000	4 096 000 000
41	1 681	68 921	2 825 761	115 856 201	4 750 104 241
42	1 764	74 088	3 111 696	130 691 232	5 489 031 744
43	1 849	79 507	3 418 801	147 008 443	6 321 363 049
44	1 936	85 184	3 748 096	164 916 224	7 256 313 856
45	2 025	91 125	4 100 625	184 528 125	8 303 765 625
46	2 116	97 336	4 477 456	205 962 976	9 474 296 896
47	2 209	103 823	4 879 681	229 345 007	10 779 215 329
48	2 304	110 592	5 308 416	254 803 968	12 230 590 464
49	2 401	117 649	5 764 801	282 475 249	13 841 287 201
50	2 500	125 000	6 250 000	312 500 000	15 625 000 000

Segue: **Potenze dei numeri interi sino a 100.**

n	n^2	n^3	n^4	n^5	n^6
51	2 601	132 651	6 765 201	345 025 251	17 596 287 801
52	2 704	140 608	7 311 616	380 204 032	19 770 609 664
53	2 809	148 877	7 890 481	418 195 493	22 164 361 129
54	2 916	157 464	8 503 056	459 165 024	24 794 911 296
55	3 025	166 375	9 150 625	503 284 375	27 680 640 625
56	3 136	175 616	9 834 496	550 731 776	30 840 979 456
57	3 249	185 193	10 556 001	601 692 057	34 296 447 249
58	3 364	195 112	11 316 496	656 356 768	38 068 692 544
59	3 481	205 379	12 117 361	714 924 299	42 180 533 641
60	3 600	216 000	12 960 000	777 600 000	46 656 000 000
61	3 721	226 981	13 845 841	844 596 301	51 520 374 361
62	3 844	238 328	14 776 336	916 132 832	56 800 235 584
63	3 969	250 047	15 752 961	992 436 543	62 523 502 209
64	4 096	262 144	16 777 216	1 073 741 824	68 719 476 736
65	4 225	274 625	17 850 625	1 160 290 625	75 418 890 625
66	4 356	287 496	18 974 736	1 252 332 576	82 653 950 016
67	4 489	300 763	20 151 121	1 350 125 107	90 458 382 169
68	4 624	314 432	21 381 376	1 453 933 568	98 867 482 624
69	4 761	328 509	22 667 121	1 564 031 349	107 918 163 081
70	4 900	343 000	24 010 000	1 680 700 000	117 649 000 000
71	5 041	357 911	25 411 681	1 804 229 351	128 100 283 921
72	5 184	373 248	26 873 856	1 934 917 632	139 314 069 504
73	5 329	389 017	28 398 241	2 073 071 593	151 334 226 289
74	5 476	405 224	29 986 576	2 219 006 624	164 206 490 176
75	5 625	421 875	31 640 625	2 373 046 875	177 978 515 625
76	5 776	438 976	33 362 176	2 535 525 376	192 699 928 576
77	5 929	456 533	35 153 041	2 706 784 157	208 422 380 089
78	6 084	474 552	37 015 056	2 887 174 368	225 199 600 704
79	6 241	493 039	38 950 081	3 077 056 399	243 087 455 521
80	6 400	512 000	40 960 000	3 276 800 000	262 144 000 000
81	6 561	531 441	43 046 721	3 486 784 401	282 429 536 481
82	6 724	551 368	45 212 176	3 707 398 432	304 006 671 424
83	6 889	571 787	47 458 321	3 939 040 643	326 940 373 369
84	7 056	592 704	49 787 136	4 182 119 424	351 298 031 616
85	7 225	614 125	52 200 625	4 437 053 125	377 149 515 625
86	7 396	636 056	54 700 816	4 704 270 176	404 567 235 136
87	7 569	658 503	57 289 761	4 984 209 207	433 626 201 009
88	7 744	681 472	59 969 536	5 277 319 168	464 404 086 784
89	7 921	704 969	62 742 241	5 584 059 449	496 981 290 961
90	8 100	729 000	65 610 000	5 904 900 000	531 441 000 000
91	8 281	753 571	68 574 961	6 240 321 451	567 869 252 041
92	8 464	778 688	71 639 296	6 590 815 232	606 355 001 344
93	8 649	804 357	74 805 201	6 956 883 693	646 990 183 449
94	8 836	830 584	78 074 896	7 339 040 224	689 869 781 056
95	9 025	857 375	81 450 625	7 737 809 375	735 091 890 625
96	9 216	884 736	84 934 656	8 153 726 976	782 757 789 696
97	9 409	912 673	88 529 281	8 587 340 257	832 972 004 929
98	9 604	941 192	92 236 816	9 039 207 968	885 842 380 864
99	9 801	970 299	96 059 601	9 509 900 499	941 480 149 401
100	10 000	1 000 000	100 000 000	10 000 000 000	1 000 000 000 000

B) II. – Somme di potenze dei numeri interi sino a 100.

n	Σn	Σn^2	Σn^3	Σn^4	Σn^5	Σn^6
1	1	1	1	1	1	1
2	3	5	9	17	33	65
3	6	14	36	98	276	794
4	10	30	100	354	1 300	4 890
5	15	55	225	979	4 425	20 515
6	21	91	441	2 275	12 201	67 171
7	28	140	784	4 676	29 008	184 820
8	36	204	1 296	8 772	61 776	446 964
9	45	285	2 025	15 333	120 825	978 405
10	55	385	3 025	25 333	220 825	1 978 405
11	66	506	4 356	39 974	381 876	3 749 966
12	78	650	6 084	60 710	630 708	6 735 950
13	91	819	8 281	89 271	1 002 001	11 562 759
14	105	1 015	11 025	127 687	1 539 825	19 092 295
15	120	1 240	14 400	178 312	2 299 200	30 482 920
16	136	1 496	18 496	243 848	3 347 776	47 260 136
17	153	1 785	23 409	327 369	4 767 633	71 397 705
18	171	2 109	29 241	432 345	6 657 201	105 409 929
19	190	2 470	36 100	562 666	9 133 300	152 455 810
20	210	2 870	44 100	722 666	12 333 300	216 455 810
21	231	3 311	53 361	917 147	16 417 401	302 221 931
22	253	3 795	64 009	1 151 403	21 571 033	415 601 835
23	276	4 324	76 176	1 431 244	28 007 376	563 637 724
24	300	4 900	90 000	1 763 020	35 970 000	754 740 700
25	325	5 525	105 625	2 153 645	45 735 625	998 881 325
26	351	6 201	123 201	2 610 621	57 617 001	1 307 797 101
27	378	6 930	142 884	3 142 062	71 965 908	1 695 217 590
28	406	7 714	164 836	3 756 718	89 176 276	2 177 107 894
29	435	8 555	189 225	4 463 999	109 687 425	2 771 931 215
30	465	9 455	216 225	5 273 999	133 987 425	3 500 931 215
31	496	10 416	246 016	6 197 520	162 616 576	4 388 434 896
32	528	11 440	278 784	7 246 096	196 171 008	5 462 176 720
33	561	12 529	314 721	8 432 017	135 306 401	6 753 644 089
34	595	13 685	354 025	9 768 353	180 741 825	8 298 449 105
35	630	14 910	396 900	11 268 978	333 263 700	10 136 714 730
36	666	16 206	443 556	12 948 594	393 729 876	12 313 497 066
37	703	17 575	494 209	14 822 755	463 073 833	14 879 223 475
38	741	19 019	549 081	16 907 891	542 309 001	17 890 159 859
39	780	20 540	608 400	19 221 332	632 533 200	21 408 903 620
40	820	22 140	672 400	21 781 332	734 933 200	25 504 903 620
41	861	23 821	741 321	24 607 093	850 789 401	30 255 007 861
42	903	25 585	815 409	27 718 789	981 480 633	35 744 039 605
43	946	27 434	894 916	31 137 590	1 128 489 076	42 065 402 654
44	990	29 370	980 100	34 885 686	1 293 405 300	49 321 716 510
45	1 035	31 395	1 071 225	38 986 311	1 477 933 425	57 625 482 135
46	1 081	33 511	1 168 561	43 463 767	1 683 896 401	67 099 779 031
47	1 128	35 720	1 272 384	48 343 448	1 913 241 408	77 878 994 360
48	1 176	38 024	1 382 976	53 651 864	2 168 045 376	90 109 584 824
49	1 225	40 425	1 500 625	59 416 665	2 450 520 625	103 950 872 025
50	1 275	42 925	1 625 625	65 666 665	2 763 020 625	119 575 872 025

Segue: **Somme di potenze dei numeri interi sino a 100.**

n	Σn	Σn^2	Σn^3	Σn^4	Σn^5	Σn^6
51	1 326	45 526	1 758 276	72 431 866	3 108 045 876	137 172 159 826
52	1 378	48 230	1 898 884	79 743 482	3 488 249 908	156 942 769 490
53	1 431	51 039	2 047 761	87 633 963	3 906 445 401	179 107 130 619
54	1 485	53 955	2 205 225	96 137 019	4 365 610 425	203 902 041 915
55	1 540	56 980	2 371 600	105 287 644	4 868 894 800	231 582 682 540
56	1 596	60 116	2 547 216	115 122 140	5 419 626 576	262 423 661 996
57	1 653	63 365	2 732 409	125 678 141	6 021 318 633	296 720 109 245
58	1 711	66 729	2 927 521	136 994 637	6 677 675 401	334 788 801 789
59	1 770	70 210	3 132 900	149 111 998	7 392 599 700	376 969 335 430
60	1 830	73 810	3 348 900	162 071 998	8 170 199 700	423 625 335 430
61	1 891	77 531	3 575 881	175 917 839	9 014 796 001	475 145 709 791
62	1 953	81 375	3 814 209	190 694 175	9 930 928 833	531 945 945 375
63	2 016	85 344	4 064 256	206 447 136	10 923 365 376	594 469 447 584
64	2 080	89 440	4 326 400	223 224 352	11 997 107 200	663 188 924 320
65	2 145	93 665	4 601 025	241 074 977	13 157 397 825	738 607 814 945
66	2 211	98 021	4 888 521	260 049 713	14 409 730 401	821 261 764 961
67	2 278	102 510	5 189 284	280 200 834	15 759 855 508	911 720 147 130
68	2 346	107 134	5 503 716	301 582 210	17 213 789 076	1 010 587 629 754
69	2 415	111 895	5 832 225	324 249 331	18 777 820 425	1 118 505 792 835
70	2 485	116 795	6 175 225	348 259 331	20 458 520 425	1 236 154 792 835
71	2 556	121 836	6 533 136	373 671 012	22 262 749 776	1 364 255 076 756
72	2 628	127 020	6 906 384	400 544 868	24 197 667 408	1 503 569 146 260
73	2 701	132 349	7 295 401	428 943 109	26 270 739 001	1 654 903 372 549
74	2 775	137 825	7 700 625	458 929 685	28 489 745 625	1 819 109 862 725
75	2 850	143 450	8 122 500	490 570 310	30 862 792 500	1 997 088 378 350
76	2 926	149 226	8 561 476	523 932 486	33 398 317 876	2 189 788 306 926
77	3 003	155 155	9 018 009	559 085 527	36 105 102 033	2 398 210 687 015
78	3 081	161 239	9 492 561	596 100 583	38 992 276 401	2 623 410 287 719
79	3 160	167 480	9 985 600	635 050 664	42 069 332 800	2 866 497 743 240
80	3 240	173 880	10 497 600	676 010 664	45 346 132 800	3 128 641 743 240
81	3 321	180 441	11 029 041	719 057 385	48 832 917 201	3 411 071 279 721
82	3 403	187 165	11 580 409	764 269 561	52 540 315 633	3 715 077 951 145
83	3 486	194 054	12 152 196	811 727 882	56 479 356 276	4 042 018 324 514
84	3 570	201 110	12 744 900	861 515 018	60 661 475 700	4 393 316 356 130
85	3 655	208 335	13 359 025	913 715 643	65 098 528 825	4 770 465 871 755
86	3 741	215 731	13 995 081	968 416 459	69 802 799 001	5 175 033 106 891
87	3 828	223 300	14 653 584	1 025 706 220	74 787 008 208	5 608 659 307 900
88	3 916	231 044	15 335 056	1 085 675 756	80 064 327 376	6 073 063 394 684
89	4 005	238 965	16 040 025	1 148 417 997	85 648 386 825	6 570 044 685 645
90	4 095	247 065	16 769 025	1 214 027 997	91 553 286 825	7 101 485 685 645
91	4 186	255 346	17 522 596	1 282 602 958	97 793 608 276	7 669 354 937 686
92	4 278	263 810	18 301 284	1 354 242 254	104 384 423 508	8 275 709 939 030
93	4 371	272 459	19 105 641	1 429 047 455	111 341 307 201	8 922 700 122 479
94	4 465	281 295	19 936 225	1 507 122 351	118 680 347 425	9 612 569 903 535
95	4 560	290 320	20 793 600	1 588 572 976	126 418 156 800	10 347 661 794 160
96	4 656	299 536	21 678 336	1 673 507 632	134 571 883 776	11 130 419 583 856
97	4 753	308 945	22 591 009	1 762 036 913	143 159 224 033	11 963 391 588 785
98	4 851	318 549	23 532 201	1 854 273 729	152 198 432 001	12 849 233 969 649
99	4 950	328 350	24 502 500	1 950 333 330	161 708 332 500	13 790 714 119 050
100	5 050	338 350	25 502 500	2 050 333 330	171 708 332 500	14 790 714 119 050

B) III. – **Radici, reciproci e potenze di reciproci dei numeri interi sino a 100.**

n	$\sqrt{n}$	$\sqrt[3]{n}$	$\frac{1}{n}$	$\frac{1}{n^2}$	$\frac{1}{n^3}$
1	1,0000000	1,0000000	1,000000000	1,000000000	1,000000000
2	1,4142136	1,2599210	0,500000000	0,250000000	0,125000000
3	1,7320508	1,4422496	333333333	111111111	037037037
4	2,0000000	1,5874011	250000000	062500000	015625000
5	2,2360680	1,7099759	200000000	040000000	008000000
6	2,4494897	1,8171206	166666667	027777778	004629630
7	2,6457513	1,9129312	142857143	020408163	002915452
8	2,8284271	2,0000000	125000000	015625000	001953125
9	3,0000000	2,0800837	111111111	012345679	001371742
10	3,1622777	2,1544347	100000000	010000000	001000000
11	3,3166248	2,2239801	090909091	008264463	0007513148
12	3,4641016	2,2894286	083333333	006944444	0005787037
13	3,6055513	2,3513347	076923077	005917160	0004551661
14	3,7416574	2,4101422	071428571	005102041	0003644315
15	3,8729833	2,4662121	066666667	004444444	0002962963
16	4,0000000	2,5198421	062500000	003906250	0002441406
17	4,1231056	2,5712816	058823529	003460208	0002035416
18	4,2426407	2,6207414	055555556	003086420	0001714678
19	4,3588989	2,6684016	052631579	002770083	0001457938
20	4,4721360	2,7144177	050000000	002500000	0001250000
21	4,5825757	2,7589243	047619048	002267574	0001079797
22	4,6904158	2,8020393	045454545	002066116	0000939144
23	4,7958315	2,8438670	043478261	001890359	0000821895
24	4,8989795	2,8844991	041666667	001736111	0000723380
25	5,0000000	2,9240177	040000000	001600000	0000640000
26	5,0990195	2,9624960	038461538	001479290	0000568958
27	5,1961524	3,0000000	037037037	001371742	0000508053
28	5,2915026	3,0365889	035714286	001275510	0000455539
29	5,3851648	3,0723168	034482759	001189061	0000410021
30	5,4772256	3,1072325	033333333	001111111	0000370370
31	5,5677644	3,1413806	032258065	001040583	0000335672
32	5,6568542	3,1748021	031250000	0009765625	0000305176
33	5,7445626	3,2075343	030303030	0009182736	0000278265
34	5,8309519	3,2396118	029411765	0008650519	0000254427
35	5,9160798	3,2710663	028571429	0008163265	0000233236
36	6,0000000	3,3019272	027777778	0007716049	0000214335
37	6,0827625	3,3322218	027027027	0007304602	0000197422
38	6,1644140	3,3619754	026315789	0006925208	0000182242
39	6,2449980	3,3912114	025641026	0006574622	0000168580
40	6,3245553	3,4199519	025000000	0006250000	0000156250
41	6,4031242	3,4482172	024390244	0005948840	0000145094
42	6,4807407	3,4760266	023809524	0005668934	0000134975
43	6,5574385	3,5033981	023255814	0005408329	0000125775
44	6,6332496	3,5303483	022727273	0005165289	0000117393
45	6,7082039	3,5568933	022222222	0004938272	0000109739
46	6,7823300	3,5830479	021739130	0004725898	0000102737
47	6,8556546	3,6088261	021276596	0004526935	00000963178
48	6,9282032	3,6342411	020833333	0004340278	00000904225
49	7,0000000	3,6593057	020408163	0004164931	00000849986
50	7,0710678	3,6840314	020000000	0004000000	00000800000

Segue: **Radici, reciproci e potenze di reciproci dei numeri interi sino a 100.**

n	$\sqrt{n}$	$\sqrt[3]{n}$	$\frac{1}{n}$	$\frac{1}{n^2}$	$\frac{1}{n^3}$
51	7,1414284	3,7084298	0,019607843	0,0003844675	0,000007538579
52	7,2111026	3,7325111	019230769	0003698225	000007111971
53	7,2801099	3,7562858	018867925	0003559986	000006716954
54	7,3484692	3,7797631	018518519	0003429355	000006350658
55	7,4161985	3,8029525	018181818	0003305785	000006010518
56	7,4833148	3,8258624	017857143	0003188776	000005694242
57	7,5498344	3,8485011	017543860	0003077870	000005399772
58	7,6157731	3,8708766	017241379	0002972652	000005125261
59	7,6811457	3,8929965	016949153	0002872738	000004869047
60	7,7459667	3,9148676	016666667	0002777778	000004629630
61	7,8102497	3,9364972	016393443	0002687450	000004405655
62	7,8740079	3,9578915	016129032	0002601457	000004195898
63	7,9372539	3,9790571	015873016	0002519526	000003999248
64	8.0000000	4,0000000	015625000	0002441406	000003814697
65	8,0622577	4,0207256	015384615	0002366864	000003641329
66	8,1240384	4,0412401	015151515	0002295684	000003478309
67	8,1853528	4,0615480	014925373	0002227668	000003324877
68	8,2462113	4,0816551	014705882	0002162630	000003180338
69	8,3066239	4,1015661	014492754	0002100399	000003044057
70	8,3666003	4,1212853	014285714	0002040816	000002915452
71	8,4261498	4,1408178	014084507	0001983733	000002793991
72	8,4852814	4,1601676	013888889	0001929012	000002679184
73	8,5440037	4.1793392	013698630	0001876525	000002570582
74	8.6023253	4,1983364	013513514	0001826150	000002467771
75	8,6602540	4,2171633	013333333	0001777778	000002370370
76	8,7177979	4,2358236	013157895	0001731302	000002278029
77	8,7749644	4,2543210	012987013	0001686625	000002190422
78	8,8317609	4,2726586	012820513	0001643655	000002107251
79	8,8881944	4,2908404	012658228	0001602307	000002028237
80	8,9442719	4,3088695	012500000	0001562500	000001953125
81	9,0000000	4,3267487	012345679	0001524158	000001881676
82	9,0553851	4,3444815	012195122	0001487210	000001813671
83	9,1104336	4,3620707	012048193	0001451589	000001748903
84	9,1651514	4,3795191	011904762	0001417234	000001687183
85	9,2195445	4,3968296	011764706	0001384083	000001628333
86	9,2736185	4,4140049	011627907	0001352082	000001572189
87	9,3273791	4,4310476	011494253	0001321178	000001518596
88	9,3808315	4,4479602	011363636	0001291322	000001467412
89	9,4339811	4,4647451	011235955	0001262467	000001418502
90	9,4868330	4,4814047	011111111	0001234568	000001371742
91	9,5393920	4,4979414	010989011	0001207584	000001327015
92	9,5916630	4,5143574	010869565	0001181474	000001284211
93	9,6436508	4,5306549	010752688	0001156203	000001243229
94	9,6953597	4,5468359	010638298	0001131734	000001203972
95	9,7467943	4,5629026	010526316	0001108033	000001166351
96	9,7979590	4,5788570	010416667	0001085069	000001130281
97	9,8488578	4,5947009	010309278	0001062812	000001095683
98	9,8904949	4,6104363	010204082	0001041233	000001062482
99	9,9498744	4,6260650	010101010	0001020304	000001030610
100	10,0000000	4.6415888	010000000	0001000000	000001000000

B) IV. – **Logaritmi, somme di logaritmi ecc,, dei numeri interi sino a 100.**

n	$\log n$	$\Sigma \log n$	$\Sigma\, n \log n$	$\Sigma (\log n)^2$
1	0,0000000	0.0000000	0,0000000	0,0000000
2	0,3010300	0.3010300	0,6020600	0.0906191
3	0,4771213	0,7781513	2,0334238	0,3182638
4	0,6020600	1,3802112	4,4416637	0,6807400
5	0,6989700	2,0791812	7,9365137	1.1692991
6	0,7781513	2,8573325	12,6054212	1,7748184
7	0,8450980	3,7024305	18,5211075	2,4890091
8	0,9030900	4,6055205	25,7458274	3,3045806
9	0,9542425	5,5597630	34,3340100	4,2151594
10	1,0000000	6,5597630	44,3340100	5,2151594
11	1,0413927	7,6011557	55,7893295	6,2996581
12	1,0791812	8,6803370	68,7395045	7,4642903
13	1.1139434	9,7942803	83,2207681	8,7051601
14	1.1461280	10,9404084	99,2665606	10,0187697
15	1,1760913	12,1164996	116,9079295	11,4019602
16	1,2041200	13,3206196	136,1738492	12,8518651
17	1,2304489	14,5510685	157,0914808	14,3658697
18	1,2552725	15,8063410	179,6863859	15,9415788
19	1,2787536	17,0850946	203,9827044	17,5767895
20	1,3010300	18,3861246	230,0033043	19,2694686
21	1,3222193	19.7083439	257,7699095	21,0177324
22	1,3424227	21,0507666	287,3032084	22,8198311
23	1,3617278	22,4124944	318,6229487	24.6741338
24	1,3802112	23,7927057	351,7480185	26,5791169
25	1,3979400	25,1906457	386,6965187	28,5333531
26	1,4149733	26,6056190	423,4858257	30,5355027
27	1,4313638	28,0369828	462,1326474	32,5843049
28	1.4471580	29,4841408	502,6530722	34,6785713
29	1.4623980	30,9465388	545,0626142	36,8171792
30	1,4771213	32,4236601	589,3762518	38,9990664
31	1.4913617	33,9150218	635,6084643	41.2232261
32	1,5051500	35,4201717	683,7732636	43,4887026
33	1,5185139	36,9386857	733,8842237	45,7945871
34	1.5314789	38,4701646	785,9545068	48,1400148
35	1,5440680	40,0142326	839,9968884	50,5241609
36	1,5563025	41,5705351	896,0237784	52,9462384
37	1,5682017	43,1387369	954,0472422	55,4054951
38	1,5797836	44,7185205	1 014,0790189	57,9012113
39	1.5910646	46,3095851	1 076,1305385	60,4326979
40	1,6020600	47,9116451	1 140,2129382	62,9992941
41	1,6127839	49,5244289	1 206,3370763	65,6003659
42	1,6232493	51,1476782	1 274,5135465	68,2353041
43	1.6334685	52,7811467	1 344,7526901	70,9035233
44	1,6434527	54,4245993	1 417,0646079	73,6044600
45	1,6532125	56,0778119	1 491,4591710	76,3375716
46	1,6627578	57,7405697	1 567,9460313	79,1023352
47	1,6720979	59,4126676	1 646,5346306	81,8982465
48	1,6812412	61,0939088	1 727,2342100	84,7248186
49	1,6901961	62,7841049	1 810,0538179	87,5815814
50	1.6989700	64,4830749	1 895,0023181	90,4680804

Segue: **Logaritmi, somme di logaritmi ecc., dei numeri interi sino a 100.**

n	$\log n$	$\Sigma \log n$	$\Sigma n \log n$	$\Sigma (\log n)^2$
51	1,7075702	66,1906450	1 982,0883971	93,3838763
52	1,7160033	67,9066484	2 071,3205710	96,3285438
53	1,7242759	69,6309243	2 162,7071920	99,3016711
54	1,7323938	71,3633180	2 256,2564551	102,3028592
55	1,7403627	73,1036807	2 351,9764030	105,3317215
56	1,7481880	74,8518687	2 449,8749325	108,3878829
57	1,7558749	76,6077436	2 549,9597993	111,4709794
58	1,7634280	78,3711716	2 652,2386229	114,5806577
59	1,7708520	80,1420236	2 756,7188916	117,7165745
60	1,7781513	81,9201748	2 863,4079666	120,8783964
61	1,7853298	83,7055047	2 972,3130866	124,0657990
62	1,7923917	85,4978964	3 083,4413713	127,2784670
63	1,7993405	87,2972369	3 196,7998259	130,5160934
64	1,8061800	89,1034169	3 312,3953443	133,7783793
65	1,8129134	90,9163303	3 430,2347124	137,0650341
66	1,8195439	92,7358742	3 550,3246122	140,3757742
67	1,8260748	94,5619490	3 672,6716240	143,7103234
68	1,8325089	96,3944579	3 797,2822300	147,0684123
69	1,8388491	98,2333070	3 924,1628173	150,4497786
70	1,8450980	100,0784050	4 053,3196801	153,8541654
71	1,8512583	101,9296634	4 184,7590228	157,2813229
72	1,8573325	103,7869959	4 318,4869626	160,7310069
73	1,8633229	105,6503187	4 454,5095314	164,2029790
74	1,8692317	107,5195505	4 592,8326786	167,6970062
75	1,8750613	109,3946117	4 733,4622734	171,2128609
76	1,8808136	111,2754253	4 876,4041064	174,7503207
77	1,8864907	113,1619160	5 021,6638922	178,3091670
78	1,8920946	115,0540106	5 169,2472713	181,8891890
79	1,8976271	116,9516377	5 319,1598115	185,4901776
80	1,9030900	118,8547277	5 471,4070104	189,1119291
81	1,9084850	120,7632127	5 625,9942970	192,7542442
82	1,9138139	122,6770266	5 782,9270329	196,4169276
83	1,9190781	124,5961047	5 942,2105145	200,0997884
84	1,9242793	126,5203840	6 103,8499746	203,8026391
85	1,9294189	128,4498029	6 267,8505832	207,5252965
86	1,9344985	130,3843013	6 434,2174510	211,2675808
87	1,9395193	132,3238206	6 602,9556260	215,0293157
88	1,9444827	134,2683033	6 774,0701012	218,8103286
89	1,9493900	136,2176933	6 947,5658118	222,6104500
90	1,9542425	138,1719358	7 123,4476376	226,4295137
91	1,9590414	140,1309772	7 301,7204043	230,2673568
92	1,9637878	142,0947650	7 482,3888844	234,1238194
93	1,9684829	144,0632480	7 665,4577986	237,9987445
94	1,9731279	146,0363758	7 850,9318169	241,8919781
95	1,9777236	148,0140994	8 038,8155594	245,8033687
96	1,9822712	149,9963707	8 229,1135977	249,7327590
97	1,9867717	151,9831424	8 421,8304560	253,6800209
98	1,9912261	153,9743685	8 616,9706114	257,6450022
99	1,9956352	155,9700037	8 814,5384957	261,6275620
100	2,0000000	157,9700037	9 014,5384957	265,6275620

B) V. – **Lunghezze degli archi da 1° a 180° e per frazioni di grado.**

Gradi	Archi	Gradi	Archi	Gradi	Archi	Gradi	Archi
1	0,0174533	46	0,8028515	91	1,5882496	136	2,3736478
2	0349066	47	8203047	92	1,6057029	137	2,3911011
3	0523599	48	8377580	93	1,6231562	138	2,4085544
4	0698132	49	8552113	94	1,6406095	139	2,4260077
5	0872665	50	8726646	95	1,6580628	140	2,4434610
6	1047198	51	8901179	96	1,6755161	141	2,4609142
7	1221730	52	9075712	97	1,6929694	142	2,4783675
8	1396263	53	9250245	98	1,7104227	143	2,4958208
9	1570796	54	9424778	99	1,7278760	144	2,5132741
10	1745329	55	9599311	100	1,7453293	145	2,5307274
11	1919862	56	9773844	101	1,7627825	146	2,5481807
12	2094395	57	9948377	102	1,7802358	147	2,5656340
13	2268928	58	1,0122910	103	1,7976891	148	2,5830873
14	2443461	59	1,0297443	104	1,8151424	149	2,6005406
15	2617994	60	1,0471976	105	1,8325957	150	2,6179939
16	2792527	61	1,0646508	106	1,8500490	151	2,6354472
17	2967060	62	1,0821041	107	1,8675023	152	2,6529005
18	3141593	63	1,0995574	108	1,8849556	153	2,6703538
19	3316126	64	1,1170107	109	1,9024089	154	2,6878070
20	3490659	65	1,1344640	110	1,9198622	155	2,7052603
21	3665191	66	1,1519173	111	1,9373155	156	2,7227136
22	3839724	67	1,1693706	112	1,9547688	157	2,7401669
23	4014257	68	1,1868239	113	1,9722221	158	2,7576202
24	4188790	69	1,2042772	114	1,9896753	159	2,7750735
25	4363323	70	1,2217305	115	2,0071286	160	2,7925268
26	4537856	71	1,2391838	116	2,0245819	161	2,8099801
27	4712389	72	1,2566371	117	2,0420352	162	2,8274334
28	4886922	73	1,2740904	118	2,0594885	163	2,8448867
29	5061455	74	1,2915436	119	2,0769418	164	2,8623400
30	5235988	75	1,3089969	120	2,0943951	165	2,8797933
31	5410521	76	1,3264502	121	2,1118484	166	2,8972466
32	5585054	77	1,3439035	122	2,1293017	167	2,9146999
33	5759587	78	1,3613568	123	2,1467550	168	2,9321531
34	5934119	79	1,3788101	124	2,1642083	169	2,9496064
35	6108652	80	1,3962634	125	2,1816616	170	2,9670597
36	6283185	81	1,4137167	126	2,1991149	171	2,9845130
37	6457718	82	1,4311700	127	2,2165682	172	3,0019663
38	6632251	83	1,4486233	128	2,2340214	173	3,0194196
39	6806784	84	1,4660766	129	2,2514747	174	3,0368729
40	6981317	85	1,4835299	130	2,2689280	175	3,0543262
41	7155850	86	1,5009832	131	2,2863813	176	3,0717795
42	7330383	87	1,5184364	132	2,3038346	177	3,0892328
43	7504916	88	1,5358897	133	2,3212879	178	3,1066861
44	7679449	89	1,5533430	134	2,3387412	179	3,1241394
45	7853982	90	1,5707963	135	2,3561945	180	3,1415927

Segue: **Lunghezze degli archi da 1° a 180° e per frazioni di grado.**

'	Gradi	Archi	'	Gradi	Archi	''	Gradi	Archi	''	Gradi	Archi
1	0,01667	0,0002909	31	0,51667	0,0090175	1	0,00028	0,0000048	31	0,00861	0,0001503
2	03333	0005818	32	53333	0093084	2	00056	0000097	32	00889	0001551
3	05000	0008727	33	55000	0095993	3	00083	0000145	33	00917	0001600
4	06667	0011636	34	56667	0098902	4	00111	0000194	34	00944	0001648
5	08333	0014544	35	58333	0101811	5	00139	0000242	35	00972	0001697
6	10000	0017453	36	60000	0104720	6	00167	0000291	36	01000	0001745
7	11667	0020362	37	61667	0107629	7	00194	0000339	37	01028	0001794
8	13333	0023271	38	63333	0110538	8	00222	0000388	38	01056	0001842
9	15000	0026180	39	65000	0113446	9	00250	0000436	39	01083	0001891
10	16667	0029089	40	66667	0116355	10	00278	0000485	40	01111	0001939
11	18333	0031998	41	68333	0119264	11	00306	0000533	41	01139	0001988
12	20000	0034907	42	70000	0122173	12	00333	0000582	42	01167	0002036
13	21667	0037815	43	71667	0125082	13	00361	0000630	43	01194	0002085
14	23333	0040724	44	73333	0127991	14	00389	0000679	44	01222	0002133
15	25000	0043633	45	75000	0130900	15	00417	0000727	45	01250	0002182
16	26667	0046542	46	76667	0133809	16	00444	0000776	46	01278	0002230
17	28333	0049451	47	78333	0136717	17	00472	0000824	47	01306	0002279
18	30000	0052360	48	80000	0139626	18	00500	0000873	48	01333	0002327
19	31667	0055269	49	81667	0142535	19	00528	0000921	49	01361	0002376
20	33333	0058178	50	83333	0145444	20	00556	0000970	50	01389	0002424
21	35000	0061087	51	85000	0148353	21	00583	0001018	51	01417	0002473
22	36667	0063995	52	86667	0151262	22	00611	0001067	52	01444	0002521
23	38333	0066904	53	88333	0154171	23	00639	0001115	53	01472	0002570
24	40000	0069813	54	90000	0157080	24	00667	0001164	54	01500	0002618
25	41667	0072722	55	91667	0159989	25	00694	0001212	55	01528	0002666
26	43333	0075631	56	93333	0162897	26	00722	0001261	56	01556	0002715
27	45000	0078540	57	95000	0165806	27	00750	0001309	57	01583	0002763
28	46667	0081449	58	96667	0168715	28	00778	0001357	58	01611	0002812
29	48333	0084358	59	98333	0171624	29	00806	0001406	59	01639	0002860
30	50000	0087266	60	1,00000	0174533	30	00833	0001454	60	01667	0002909

INDICE

DEL VOLUME PRIMO

APPENDICI ALLA PARTE PRIMA

Prezzo del presente volume Lire 25 —