GIORGIO MORTARA

SOMMARIO
DI
STATISTICA

OPERA EDITA SOTTO GLI AUSPICI
DELLA UNIVERSITÀ BOCCONI
DI MILANO
1931 - IX

GIORGIO MORTARA

SOMMARIO DI STATISTICA

OPERA EDITA SOTTO GLI AUSPICI
DELLA UNIVERSITÀ BOCCONI
DI MILANO
1931 - IX

A

RODOLFO BENINI

CON DEVOZIONE DI DISCEPOLO

E AFFETTO DI AMICO

INDICE ANALITICO

AGLI STUDENTI

Poichè questo Sommario *è destinato sopratutto a Voi, miei discepoli, ho il dovere di spiegarvi i criteri che mi hanno guidato nel compilarlo.*

Rinunziando all'ambizione di presentarvi un vero e proprio manuale, ho cercato di offrirvi una semplice e piana introduzione allo studio della statistica metodologica.

Se v'è taluno fra Voi che si ribelli alla diffusa concezione del periodo universitario come una lunga e noiosa corsa con ostacoli, e lo riguardi invece, con giovanile entusiasmo, come l'epoca della vita più adatta per estendere ed approfondire la propria coltura, egli non si accontenterà di un'esposizione elementare e sentirà il bisogno di ricorrere a trattazioni ampie ed elevate, quali non mancano nella nostra letteratura: il Sommario *gli darà soltanto la preparazione e l'incitamento a più alte mete.*

Allo « studente medio », che vuol compiere il proprio dovere ma nulla più (è forse ottimista questa definizione dello « studente medio » ?), il presente volumetto offre quanto basta per la preparazione all'esame, a patto di essere adoperato con intelligenza: sfruttato, cioè, non soltanto nel testo ma anche negli annessi quesiti ed esercizi, ed integrato con la frequenza alle lezioni, che non sono mai semplici ripetizioni del testo.

Quanto allo studente che non studia, e si preoccupa unicamente di carpire col minimo sforzo una stiracchiata approvazione all'esame, — sia lieto goliardo spensierato o sia scorato combattente della dura battaglia per l'esistenza —, Vi dirò che non posso ignorarlo e non debbo trascurarlo (troppi sarebbero i trascurati, e non tutti meritevoli dell'abbandono!). Anzi Vi confesso che ho pensato sopratutto a lui, e che ho cercato di rendergli breve ed agevole il cammino, affinchè egli non possa scusare a se stesso ed agli altri la propria ignavia col pretesto della *mole o della difficoltà della materia. Il* Sommario *contiene, a mio avviso, il minimo indispensabile per un corso universitario di statistica; ma appunto perciò il giorno dell'esame bisogna averlo in mente tutto, senza lacune. Vi chiedo poco, ma quel poco dev'essere conosciuto a fondo.*

Badate, però, che il discepolo ideale non è quello che s'imprime fedelmente nel cervello le parole del maestro, bensì quello che assimila e discute le idee da lui esposte. Compito del professore universitario non è soltanto insegnare dogmaticamente una disciplina: è anche, e sopratutto, far pensare *i giovani. Poche menti privilegiate sono in grado di contribuire al progresso della scienza, ma ogni intelletto normale può migliorarsi ed innalzarsi nell'abitudine della serena discussione scientifica. La statistica — ricca di problemi che ammettono molteplici soluzioni, disseminata di trabocchetti pronti ad ingoiare l'incauto manipolatore di cifre, aperta agli artifizi del falso ragionatore truccato da logico rigoroso — si presta mirabilmente a temprare lo spirito critico mercè l'analisi dei metodi e delle loro applicazioni.*

*Gli esercizi, annessi ad ogni capitolo, mirano appunto a stimolare l'attività intellettuale autonoma dello studente. Alcuni richiedono semplicemente l'applicazione delle cognizioni tratte dalla lettura del testo; altri portano l'esecutore a dare forma algebrica, cioè generale, a definizioni ivi esposte per via di esempi particolari, o con parole; altri ancora permettono al volonteroso di aggiungere nuove conoscenze a quelle già apprese; altri, infine, essendo volutamente espressi in forma errata o proponendo compiti non attuabili, conducono chi li tenti a rendersi conto di errori frequenti nella pratica, o a meglio intendere le condizioni di applicabilità di certi metodi. La maggior parte degli esercizi consiste in elaborazioni di dati desunti dall'*Annuario Statistico Italiano (ASI) *dell'*ISTITUTO CENTRALE DI STATISTICA, *che prescrivo ai miei discepoli come testo sussidiario. La ricchezza delle notizie contenute nell'Annuario rende possibile una grande varietà di esempi e di applicazioni; la natura e l'attualità di tali notizie fanno sì che il giovane, se comincia ad esaminarle dall'aspetto formale come oggetto di elaborazioni intese a consolidare la conoscenza dei metodi, finisca poi coll'appassionarsi al loro aspetto sostanziale, abituandosi a giudicare la vita economica e sociale della nazione attraverso la propria interpretazione dei fatti, invece che attraverso le interpretazioni datene da altri, che talora li ignorano.*

Ad evitare equivoci, sia ben chiaro che ogni singolo studente non deve eseguire tutti gli esercizi contenuti nel volume. Questi sono molto *numerosi, affinchè molti studenti possano svolgere contemporaneamente temi diversi, con maggiore soddisfazione di ciascuno di loro e con minore possibilità di dubbi sull'originalità del lavoro, per il professore o l'assistente che dovrà rivederlo. Credo indispensabile che ogni studente nel corso della sua preparazione svolga almeno una trentina di esercizi scelti qua e là nel volume (non sono molti: uno alla settimana, nell'anno scolastico!): solo così egli potrà assicurarsi di non avere invano affaticato la memoria e di essere atto ad applicare ragionando le nozioni apprese.*

I quesiti, intercalati agli esercizi, sono destinati ad agevolare la preparazione all'esame. Molte volte l'interrogato rimane a bocca aperta di fronte ad una domanda; qualche volta egli risponde poi correntemente alla stessa domanda proposta in forma un po' diversa: i quesiti possono dare un'idea del genere delle interrogazioni che costituiranno l'esame (non dovrei parlarvi tanto di esame? dovrei fingere d'ignorare che esso costituisce la massima preoccupazione per molti di Voi, come la costituiva per molti di noi, che ora sediamo in cattedra, quando sedevamo sui banchi?).

Le indicazioni bibliografiche sono frammentarie: non ho voluto schiacciarvi sotto una massa di richiami, mercè i quali foste messi in grado di approfondire la più minuscola questione e di informarvi sul meno significante divario di opinioni. Mi sono limitato ad additarvi poche opere, che possono essere utilmente consultate da chi voglia ampliare e migliorare la sua coltura statistica; nelle opere stesse, e in altre fonti facilmente accessibili, troverete ulteriori indicazioni.

Un'ultima avvertenza, per debito di lealtà. I metodi elementari esposti nel Sommario *sono in gran parte da lungo tempo patrimonio comune degli studiosi, ed è arduo rintracciare il loro primo ideatore: chi ha inventato la media aritmetica, gran pilastro del tempio della statistica? chi ha delineato il primo diagramma? chi ha formato il primo rapporto di frequenza? Dal mio silenzio sui nomi degli inventori non siate indotti a pensare che tutto quanto è esposto nel* Sommario *sia stato ideato dall'autore. Adottate piuttosto l'opinione diametralmente opposta; nè io me ne dorrò, perchè il mio dovere non è quello d'impartire l'insegnamento più originale possibile, bensì il più utile ed il più efficace.*

Al fine appunto dell'efficacia dell'insegnamento confido non siano del tutto vane queste franche dichiarazioni che Vi ho voluto fare, tanto per intenderci fino dal principio del cammino che dobbiamo percorrere insieme.

Milano, 15 giugno 1931-IX

GIORGIO MORTARA.

INTRODUZIONE

CAPITOLO I.

Nozioni generali.

Indagini scientifiche e indagini pratiche — Fenomeni individualmente tipici, collettivamente tipici, atipici — La statistica, disciplina dei metodi per le indagini sui fenomeni collettivamente tipici — Statistica metodologica e statistica applicata — I compiti della statistica metodologica.

1. — L'uomo è tratto, per istinto e per ragionamento, ad osservare quanto avviene nell'ambiente che lo circonda e nella sua stessa persona, ed a ricercare se si presentino regolarità nei fenomeni osservati e se si manifestino relazioni tra singoli fenomeni e le circostanze nelle quali essi si presentano, oppure relazioni tra vari fenomeni. Quest'attività investigatrice può essere diretta allo scopo pratico di determinare le direttive più convenienti della condotta individuale, o allo scopo scientifico di accertare uniformità e relazioni più o meno largamente valide. Le due mete dell'investigazione non sono opposte, anzi una parte delle vie che conducono ad esse è comune: in generale la meta scientifica è più lontana della meta pratica, e perciò la descrizione del processo d'indagine delle scienze di osservazione mostra anche il cammino segnato alle indagini dirette a fini pratici.

2. — Nelle scienze di osservazione, si comincia col determinare, per mezzo dell'*osservazione*, in quali condizioni e con quali caratteri si presentino i fenomeni che formano oggetto d'indagine. Mediante la *descrizione* dei risultati dell'osservazione si determina la corrispondenza esistente tra il presentarsi del fenomeno, o tra i caratteri di esso, e la presenza, o i caratteri, di date condizioni. Con l'*induzione* si estende il risultato delle osservazioni eseguite a casi

non osservati nei quali concorrano condizioni uguali od analoghe: tale generalizzazione, che mira a rendere possibili le previsioni, può assumere la forma di enunciazione di una uniformità (*legge empirica*): talora semplice accertamento di costanza, tal altra accertamento di una relazione tra la manifestazione di un fenomeno e alcune delle condizioni nelle quali esso si presenta. Delle uniformità accertate si chiede poi conferma all'esperienza, sia col cercarne nuove testimonianze, sia col verificare se nella realtà si avverino conseguenze tratte mediante *deduzione* dalle uniformità formulate per induzione. Infine lo scienziato, collegando logicamente fra loro varie leggi empiriche, col sussidio di *ipotesi*, giunge alla costruzione di *teorie*, che riassumono e collegano le relazioni accertate fra vari ordini di fenomeni. Anche le teorie si sottopongono alla prova dell'esperienza, che sola è in grado d'indicare se ed entro quali limiti esse possano riuscire utili come mezzi di sintesi scientifica, come strumenti per nuove indagini, come basi per la previsione dei fenomeni.

La ricerca diretta a fini pratici spesso si ferma all'accertamento di date relazioni che concretamente la interessano, ma talora percorre anch'essa tutte le tappe della più completa ricerca diretta a fini scientifici. D'altro canto, l'indagine scientifica non di rado si compie senza giungere alla determinazione di uniformità, e tanto meno alla formulazione di teorie. Piuttosto che il metodo, è il fine quello che distingue le due categorie di indagini.

3. — Certe volte l'osservazione di un fenomeno insegna che quando concorrono determinate circostanze esso si presenta *sempre*, ovvero assume *sempre* una determinata modalità o misura. Fenomeni di tal sorta si dicono *individualmente tipici*, perchè una sola osservazione dà norma per ogni caso nel quale concorrano le stesse circostanze. Ponendo a contatto in date condizioni determinati elementi chimici in determinate proporzioni, si ottiene un determinato composto: ecco un fenomeno individualmente tipico.

Altre volte l'osservazione di un fenomeno insegna che quando concorrano determinate circostanze esso si presenta *più o meno frequentemente, ma non sempre*, ovvero assume più o meno frequentemente una determinata modalità o misura. Un uomo che compie sessant'anni potrà sopravvivere a sessantuno ma potrà invece non sopravvivere. Una moneta lanciata per aria potrà ricadere sopra una determinata delle sue faccie ma potrà invece ricadere sull'altra. La temperatura dell'aria alle 7 di mattina, in un dato giorno, potrà essere più bassa che alle 8, ma potrà invece essere più alta. La statura di un Italiano ventenne potrà essere di cm. 164 ma potrà essere differente. Fenomeni di tal sorta non sono individualmente tipici, perchè una sola osservazione non dà alcuna norma per altri casi nei quali concorrano le stesse circostanze. Ma per lo più essi presentano uniformità quando vengano considerati non isolatamente bensì per collettività di casi: allora si dicono *collettivamente tipici*, perchè l'osservazione di una collettività di casi dà norma per altre collettività di casi. La frequenza delle morti in più gruppi di sessantenni, la proporzione delle cadute della moneta su una certa faccia in più serie di lanci, la proporzione degli individui alti cm. 164 in più leve successive presentano una approssimativa costanza; la temperatura dell'aria in una serie di giorni è normalmente più bassa alle ore 7 di mattina che alle 8.

Esistono però fenomeni (tali sono, almeno da certi aspetti i terremoti, le guerre) che non presentano uniformità nè considerati individualmente nè considerati collettivamente: essi diconsi *atipici* e si sottraggono per la loro stessa natura così alle indagini pratiche come a quelle scientifiche, oltre lo stadio della descrizione.

La distinzione tra fenomeni individualmente tipici, collettivamente tipici e atipici ha una base puramente soggettiva: col progresso dell'indagine scientifica sono condotti nella prima categoria fenomeni che andavano classificati nella seconda, e sono trasferiti nella seconda categoria fenomeni che erano relegati nella terza. Uno stesso fenomeno, secondo che si prescinda, o non, dai suoi caratteri qualitativi o quantitativi, può essere classificato nell'una o nell'altra categoria.

In molti casi si passa in modo così graduale dal fenomeno individualmente tipico a quello collettivamente tipico, che è difficile stabilire dove finisca l'uno e cominci l'altro. Si consideri, per esempio, l'effetto di una determinata sostanza tossica sull'organismo umano. L'esperienza ci mostra che l'ingestione di dosi superiori a un grammo determina *sempre* la morte, mentre l'ingestione di dosi inferiori a un milligrammo non determina *mai* la morte: qui possiamo stabilire una regola valida per ogni caso singolo, sicchè siamo di fronte ad un fenomeno individualmente tipico. Ma proviamo ad osservare gli effetti dell'ingestione di dosi di uno, di due, di tre decigrammi di veleno: talvolta vedremo sopravvenire la morte, tal altra no: eccoci di fronte ad un fenomeno collettivamente tipico; e diciamo «collettivamente tipico», cioè che presenta regolarità per gruppi di casi, perchè la proporzione dei morti tra gli individui sottoposti all'azione del veleno si mantiene press'a poco costante in più serie di osservazioni se resta costante la dose ingerita, e va crescendo se si accresce tale dose. Qual è il punto in cui la dose di veleno, crescendo, da innocua diviene — sia pure in rarissimi casi — letale? E qual è il punto in cui, crescendo ancora, diviene letale senza eccezioni? Dove sono, cioè, i limiti tra fenomeno individualmente tipico e fenomeno collettivamente tipico? Riuscirà praticamente impossibile accertarlo. Un altro esempio, completamente diverso: il risultato del getto di un dado da giuoco. Se questo è costruito in modo che il centro di gravità sia molto avvicinato al centro di una determinata faccia, il dado cadrà sempre su questa faccia: il fenomeno è individualmente tipico. Se ora a grado a grado, nella costruzione del dado, si sposta il centro di gravità verso il centro di figura, da un certo punto in poi il dado comincerà a non cadere sempre sulla solita faccia: cadrà su di essa con frequenza a grado a grado decrescente. Il fenomeno sarà divenuto collettivamente tipico: ma non potremo stabilire con precisione da quale posizione del centro di gravità abbia cominciato ad essere tale. Gli esempi potrebbero moltiplicarsi all'infinito; ma quello che importa ritenere è che la distinzione tra fenomeni collettivamente tipici e individualmente tipici ha fondamento soggettivo, e non si può sempre segnare con precisione, nella pratica.

4. — Fenomeni individualmente tipici s'incontrano spesso nel campo delle scienze fisiche, meno spesso in quello delle scienze biologiche, quasi mai in quello delle scienze sociali. Fenomeni collettivamente tipici s'incontrano nel campo di ciascuno di codesti tre ordini di scienze, ma assumono la maggior importanza relativa nel campo delle scienze sociali, che essi occupano quasi interamente.

L'investigazione di tali fenomeni richiede speciali metodi, che sono riuniti e coordinati in una particolare disciplina, detta *statistica* appunto perchè il primo impiego sistematico dei metodi stessi ha avuto luogo nelle indagini sui fenomeni più interessanti per la vita degli Stati. Si può dunque definire la statistica come la disciplina dei metodi per lo studio dei fenomeni collettivamente tipici. *Statistico* (e non statista, che vuol dire uomo di Stato) è il cultore della statistica, la quale, come risulta chiaramente da quanto si è detto prima, dà luogo ad applicazioni anche fuori del campo delle scienze sociali.

Dal tronco della *statistica metodologica* si dipartono numerosi rami di *statistica applicata* che vanno dalla statistica economica alla giudiziaria, dalla statistica zootecnica alla botanica, dalla statistica fisica all'astronomica. In ciascuno di questi rami, poi, si hanno applicazioni specialmente importanti che possono dar luogo a trattazione autonoma; esempi: la statistica aziendale, la statistica dell'eredità biologica, la statistica molecolare. I metodi sono sempre i medesimi: variano soltanto, col variare dell'oggetto, l'opportunità e le combinazioni delle applicazioni.

Negli studi superiori di scienze economiche e di scienze giuridiche in Italia è materia d'insegnamento la statistica metodologica, appunto perchè può porre il discente in grado di interpretare quei fenomeni sociali che, riguardati da vari loro aspetti, formano oggetto dei suoi studi; sono impartiti anche insegnamenti di statistica economica, finanziaria, demografica, giudiziaria, intesi a far conoscere ad un tempo i fatti e i metodi idonei per le indagini intorno ad essi.

5. — La statistica metodologica ha come primo compito quello di determinare le norme più adatte per l'osservazione dei fenomeni collettivamente tipici, dei caratteri coi quali essi si manifestano, delle condizioni nelle quali si presentano. Stabilisce poi i procedimenti atti a collegare, nella descrizione, i fenomeni e i loro caratteri — esposti in forme idonee a rendere più agevole l'esame e la visione d'insieme — con le condizioni nelle quali si presentano, in modo da spianare la via all'ulteriore corso delle investigazioni. Indi insegna come, eliminati dalla descrizione gli elementi non essenziali dei fenomeni osservati, si passi alle indagini sulle uniformità e sulle relazioni dei fenomeni stessi, per quali vie tali indagini possano essere condotte, in quali modi possa essere saggiata l'attendibilità dei loro risultati. Mette infine in risalto i particolari caratteri delle uniformità e delle relazioni dei fenomeni collettivamente tipici e ne desume limiti e norme per l'impiego di esse al fine della previsione.

Un problema generale e fondamentale per la metodologia statistica sorge dalla opportunità, o meglio dalla necessità, di stabilire in modo preciso che cosa s'intenda per « collettività di casi ». Quando si definisce « individualmente tipico » il fenomeno che presenta uniformità nell'osservazione di casi singoli, si dà una definizione precisa; ma quando si definisce «collettivamente tipico» quello che presenta uniformità nell'osservazione di collettività di casi, resta a chiarire che cosa sia una «collettività di casi», e se esistano relazioni e quali tra la grandezza del numero dei casi osservati e il presentarsi di uniformità. La discussione di questo problema, che logicamente dovrebbe precedere ogni trattazione di metodi d'osservazione e di descrizione, didatticamente invece può essere con maggior opportunità collegata con lo studio delle uniformità dei fenomeni collettivamente tipici.

Come risulta da quanto finora si è detto, la statistica metodologica non è una scienza nella quale trovi sistemazione e coordinamento tutto un dato ordine di fenomeni; è semplicemente la disciplina coordinatrice di un complesso di metodi impiegabili in molti e svariati campi così a fine scientifico come a fine pratico. Per l'osservazione prende norma dalle scienze e dalle arti nel cui campo l'osservazione si svolge; per la descrizione attinge specialmente alle matematiche; per l'interpretazione, per la ricerca delle uniformità e delle relazioni chiede ausilio alla logica.

Indicazioni bibliografiche. — GABAGLIO A., *Teoria generale della statistica*, 2a ed., Milano, Hoepli, 1888. — BENINI R., *Principii di statistica metodologica*, Torino, Utet, 1906. — GINI C., *Appunti di statistica*, Padova, La Litotipo, 1915. — MORTARA G., *Lezioni di statistica metodologica*, Città di Castello, Edizione del *Giornale degli Economisti*, 1922. — NICEFORO A., *Il metodo statistico*, Messina, Principato, 1923. — LIVI L., *Elementi di statistica*, Padova, Cedam, 1926. — VINCI F., *Introduzione al metodo statistico*, Padova, Cedam, 1930. — JULIN A., *Statistique théorique*, Paris, Rivière, 1921. — BOWLEY A. L., *Elements of statistics*, 5a ed., London, King, 1926 (traduz. francese, ediz. Giard, 1929). — YULE G. U., *An introduction to the theory of statistics*, 8a ediz., London, Griffin, 1927. — SECRIST H., *An introduction to statistical methods*, ed. riv., New York, Macmillan, 1925. — VON MAYR G., *Statistik und Gesellschaftslehre*, Leipzig, Mohr, 1895. — WESTERGAARD H. und NYBOLLE H. C., *Grundzüge der Theorie der Statistik*, Jena, Fischer, 1928.

Quesiti ed esercizi: 1. — Quali possono essere i fini dell'investigazione dei fenomeni?

2. — Quali sono le tappe dell'investigazione diretta a fini scientifici?

3. — Esempi di investigazione diretta a fini scientifici. A fini pratici.

4. — Si delineino i metodi e i possibili fini scientifici e pratici di una investigazione sulle vendite a rate; di una investigazione sulla delinquenza di sangue; di una investigazione comparativa sul rendimento per ettaro della coltivazione di due varietà di grano.

5. — In che differiscono i fenomeni collettivamente tipici dai fenomeni individualmente tipici? da quelli atipici?

6. — Perchè diciamo che il fondamento di questa classificazione dei fenomeni è soggettivo e non oggettivo?

7. — Esempi di fenomeni delle tre categorie. Si desumano esempi di fenomeni collettivamente tipici dall'*Annuario statistico italiano*, pubblicato dall'ISTITUTO CENTRALE DI STATISTICA (che da qui innanzi per brevità sarà citato con la sigla ASI), capitoli: «Climatologia, popolazione, igiene, agricoltura, lavoro, mercato monetario». Si spieghi perchè i fenomeni scelti siano collettivamente tipici, e non atipici nè individualmente tipici.

8. — S'incontrano nell'ASI esempi di fenomeni atipici? individualmente tipici?

9. — Si adducano esempi di passaggio graduale dal fenomeno individualmente tipico al fenomeno collettivamente tipico.

10. — Che cos'è la statistica? Chi è lo statista? Chi lo statistico?

11. — Si esaminino comparativamente le definizioni della statistica date nei manuali di Gabaglio, Benini, Niceforo, Livi, Julin, Bowley, Yule; si pongano in rilievo gli elementi comuni e gli elementi differenziali di esse; si ricerchi quale corrisponda meglio alle pratiche applicazioni dei metodi statistici note al lettore.

12. — Quali rami della statistica applicata vi possono specialmente interessare? Di quale di essi possedete già qualche nozione?

13. — Quali sono le tappe dell'applicazione dei metodi statistici? Si provi a tracciare sommariamente la via per lo studio statistico della natalità in una popolazione; per l'analisi statistica delle esportazioni di un paese; per un'indagine sulla litigiosità giudiziaria; per lo studio della distribuzione e dell'andamento delle vendite di una grande azienda.

14. — Siete in grado di dire di quanti casi almeno debba essere composta una collettività di assicurati sulla vita per poter presentare uniformità nella frequenza delle morti in funzione dell'età?

15. — Conoscete esempi di uniformità di fenomeni collettivamente tipici? Se ne conoscete qualcuno, cercate di rendervi conto fino da ora delle differenze che intercedono fra tali uniformità e quelle dei fenomeni individualmente tipici.

16. — Conoscete esempi di fenomeni che col progresso della scienza siano stati trasferiti dalla categoria degli atipici a quella dei collettivamente tipici? dalla categoria dei collettivamente tipici a quella degli individualmente tipici?

LIBRO PRIMO
L'osservazione dei fenomeni

—

CAPITOLO II.
L'osservazione statistica.

L'osservazione statistica: di quali operazioni consta — Enumerazione, misurazione, graduazione, classificazione — Possibilità di stabilire norme generali per applicazioni particolari — Definizione dell'oggetto dell'osservazione — Delimitazione di essa nel tempo e nello spazio — Definizione delle circostanze da rilevare — Errori di enumerazione; errori di accertamento di circostanze — Mezzi dell'osservazione: simboli isolati, isolabili, non isolabili.

1. — Dicesi *osservazione* (o *rilevazione*) *statistica* il complesso delle operazioni occorrenti per l'accertamento delle manifestazioni dei fenomeni collettivamente tipici e delle condizioni nelle quali essi si manifestano.

Un'operazione peculiare dell'osservazione statistica consiste nell'*enumerazione* di collettività di casi. Si suol parlare di *casi osservati* quando si enumerano gli elementi che costituiscono in un certo momento una collettività in relazione alla quale si vogliono investigare fenomeni collettivamente tipici, per esempio gli individui che costituiscono una popolazione. Si parla invece di casi *accertati* quando si enumerano le manifestazioni individuali, avvenute in un certo intervallo di tempo, del fenomeno che è oggetto di studio, per esempio i casi di morte avvenuti in un anno in una popolazione.

L'enumerazione non è una parte costitutiva assolutamente indispensabile dell'osservazione statistica: talvolta essa è surrogata dalla *misurazione*, cioè dalla determinazione della misura di circostanze quantitative, o dalla *graduazione*, cioè dalla determinazione del grado di circostanze che non ammettono misurazione ma ammettono ordinamento, o dalla *classificazione*, cioè dalla determinazione della modalità di circostanze qualitative. Tali operazioni, d'altronde, in generale si accompagnano a quella dell'enumerazione, quando non la surrogano; e ne costituiscono un complemento necessario affinchè i risultati dell'osservazione possano servire per lo studio dei caratteri coi quali si presenta il fenomeno osservato, e per la ricerca delle relazioni esistenti tra il fenomeno stesso e le condizioni nelle quali esso si manifesta. Per esempio, il numero dei morti è un dato di fatto necessario a conoscere in un'indagine sulla mortalità; ma in generale non è sufficiente: converrà anche rilevare il sesso, l'età, la condizione sociale, la sede dei morti, la malattia od altra causa che ha cagionato il decesso, ecc.

2. — L'infinita varietà delle applicazioni del metodo statistico rende impossibile dettare norme complete idonee a guidare l'osservazione in tanti differenti campi. Gli oggetti dell'enumerazione possono essere passivi di fronte all'osservatore, come nella statistica stellare, od attivi, come nella statistica fiscale; anche quando essi sono passivi, possono presentarsi grandi difficoltà nella enumerazione e nella determinazione dei caratteri degli oggetti osservati, come mostra lo stesso primo esempio ora addotto; mentre quando sono attivi, può accadere che la loro attività sia volta al buon successo dell'indagine, ma può anche accadere proprio il contrario, come in generale avviene appunto nel secondo esempio di dianzi. Soltanto la scienza o l'arte nel cui dominio cade l'oggetto dell'osservazione potrà additare i mezzi e segnare le norme più adatte per superare le difficoltà dell'osservazione derivanti dalla particolare natura di essa: l'astronomo insegnerà come si debba procedere nell'enumerazione delle stelle e nell'accertamento dei loro caratteri; il perito finanziario detterà le norme più adatte per superare le difficoltà che si incontrano nell'enumerazione dei patrimoni e dei redditi e nell'accertamento della loro consistenza. Lo statistico non può pretendere di possedere un'universale competenza in questi e negli innumerevoli altri campi d'indagine che sono aperti al suo metodo.

Ma se nell'organizzazione delle osservazioni aventi diverso oggetto si presentano differenti esigenze, da altra parte si presentano anche analogie e identità di compiti e di mezzi: per contare, e per classificare poi secondo i loro caratteri, le stelle, si potrà ricorrere a mezzi identici a quelli che servono a contare ed a classificare i contribuenti; la natura di certi possibili errori sarà la medesima; le avvertenze utili per evitarli saranno simili. Accanto alla tecnica particolare delle rilevazioni in ciascun campo, e come introduzione ad essa, si può dunque costruire non una teoria generale delle rilevazioni statistiche ma una raccolta di precetti che l'esperienza e il buon senso suggeriscono come generalmente o largamente applicabili nell'osservazione dei fenomeni collettivamente tipici.

3. — Per quanto riguarda l'enumerazione, è ovvio il precetto che essa debba essere ordinata in modo da evitare così errori in eccesso, che possono derivare dal contare casi che non dovrebbero essere contati o dal ripetere il computo di singoli casi, come errori in difetto, che possono derivare dall'omettere il computo di casi che avrebbero dovuto essere contati. Non si vuol alludere ad errori che derivino dal contar male: questi si evitano con l'impiego di persone e di macchine appropriate e si correggono con adeguati controlli; si allude invece ad errori che derivino da equivoca, o male intesa, definizione dell'oggetto dell'enumerazione. Una precisa, chiara ed univoca definizione di tale oggetto è tanto più necessaria quanto maggiore è il numero dei collaboratori nell'osservazione e quanto più ampie sono le possibilità di interpretazioni o di definizioni diverse, per se medesime non assurde. Molte volte un termine o un'espressione, che a prima vista appare di chiaro significato, ammette interpretazioni diverse, solo in parte interferenti, onde conviene precisare il significato che si assume. Si pensi, per esempio, a quante diverse interpretazioni si presti la parola «famiglia» e quanto differenti possano riuscire i risultati di un censimento delle famiglie secondo che si adotti una definizione od un'altra, quanto eterogenei possano riuscire se i diversi esecutori si attengono a diverse definizioni.

4. — Se si considera l'osservazione riguardo al tempo, si può distinguere l'osservazione diretta ad accertare lo stato di date circostanze in un determinato istante (*fenomeno di stato*) da quella diretta ad accertare la variazione di date circostanze in un determinato intervallo di tempo (*fenomeno di movimento*).

L'osservazione di un fenomeno di stato può essere eseguita una volta tanto, o ad intervalli irregolari: dicesi allora *saltuaria* (esempio: un'inchiesta sull'alimentazione di una popolazione); ovvero può essere ripetuta ad intervalli regolari: dicesi allora *periodica* (esempio: censimento della popolazione, ogni dieci anni; misurazione della temperatura, ogni ora). Può essere eseguita in modo *continuo*, così da rappresentare lo stato di fatto in ogni istante; ma allora indica anche le variazioni dello stato di fatto in qualsiasi intervallo di tempo compreso nel periodo di osservazione ed è quindi ad un tempo osservazione di un fenomeno di stato e di un fenomeno di movimento (esempio: l'accertamento della pressione atmosferica mediante un barometro registratore: indica la pressione in qualsiasi istante e la variazione della pressione in qualsiasi intervallo di tempo).

Quando l'osservazione non sia continua, ma riferita a un dato istante o ad un determinato intervallo, conviene definire con la massima precisione l'istante o l'intervallo, per evitare equivoci nell'interpretazione da parte di chi deve collaborare all'indagine o utilizzarne i risultati.

5. — Se si considera l'osservazione riguardo ad una circostanza variabile in modo continuo, diversa dal tempo, in relazione alla quale varii il fenomeno osservato, si possono eseguire distinzioni simili a quelle precedenti. Si compie l'osservazione di un fenomeno di stato se si mira ad accertare lo stato di date circostanze in corrispondenza ad un determinato valore della variabile; di un fenomeno di movimento, se si mira ad accertare la variazione di date circostanze in corrispondenza ad un determinato intervallo di valori della variabile. Nella prima ipotesi l'osservazione potrà essere ripetuta a intervalli regolari, o potrà essere eseguita in modo continuo. Esempio: una schiera di marciatori parte per una gara di resistenza. Possiamo accertare quanti rimangono in gara alla fine della decima ora: ecco l'osservazione di un fenomeno di stato; se accertiamo quanti rimangono in gara alla fine di ogni ora, l'osservazione è periodica; se accertiamo quanti rimangono in gara ad ogni istante, l'osservazione è continua. Se accertiamo quanti si ritirano dalla gara nel corso della decima ora abbiamo l'osservazione di un fenomeno di movimento. Fin qui siamo nelle ipotesi del paragrafo precedente. Ma sostituiamo ora lo spazio al tempo: possiamo accertare quanti rimangono in gara alla fine del decimo chilometro, ovvero alla fine di ciascun chilometro, ovvero in ogni punto dell'itinerario; possiamo accertare quanti si ritirano nel corso del decimo chilometro. Il parallelismo tra le due applicazioni è completo; e del tutto simili sono le avvertenze dell'opportunità di definire in modo univoco il punto o l'intervallo cui va riferita l'osservazione.

6. — Se si considera l'osservazione riguardo ad una circostanza variabile in modo discontinuo, in relazione alla quale varii il fenomeno osservato, si può distinguere l'osservazione diretta ad accertare lo stato di date circostanze in corrispondenza ad un determi-

nato valore della variabile (fenomeno di stato) da quella diretta ad accertare la variazione di date circostanze in corrispondenza ad un determinato valore della variabile (esempio: accertamento del numero delle coppie coniugali che hanno avuto cinque figli: fenomeno di stato; accertamento delle coppie coniugali che vengono allietate da un sesto figlio: fenomeno di movimento).

Per continuare il linguaggio geometrico, diremo che qui l'accertamento non solo del fenomeno di stato ma anche di quello di movimento avviene in determinati punti e non più in determinati intervalli.

7. — In generale l'osservazione è determinata nello spazio, e ad evitare errori di rilevazione bisogna che anche i limiti di spazio, come quelli di tempo, siano precisamente definiti.

In certi casi non appare, almeno esplicitamente, una delimitazione delle osservazioni nello spazio o nel tempo, perchè superflua o perchè già implicita nell'enunciazione dell'oggetto dell'osservazione. Così se un antropologo misura e classifica i crani umani della sua collezione; se un maestro valuta e gradua le attitudini intellettuali dei suoi discepoli; se un criminologo registra le successive condanne di un dato gruppo di delinquenti abituali. Ma nella massima parte delle indagini statistiche di vasta estensione su fenomeni sociali dev'essere esplicitamente e precisamente definito il campo d'osservazione così nello spazio come nel tempo.

8. — Al pari dell'oggetto fondamentale dell'osservazione, devono essere precisamente ed univocamente definite le circostanze inerenti all'oggetto stesso od inerenti al campo d'osservazione, che si vogliono rilevare. Ove occorra, devono essere precisamente stabiliti i mezzi, i metodi, i criteri, le unità di misura da applicare nelle operazioni di misurazione, di graduazione e di classificazione. Infine, quando non sia già implicita nelle precedenti indicazioni, deve darsi quella delle varie modalità o misure che si vogliono tener distinte nella rilevazione delle circostanze qualitative o quantitative. Talvolta le diverse modalità o misure differiscono nettamente tra loro, non essendo possibile il passaggio graduale ma soltanto il passaggio brusco da ciascuna di esse alla successiva; tal altra si passa da una modalità all'altra, per variazioni insensibili; nel primo caso le circostanze si dicono *discontinue*, nel secondo *continue*.

Gli errori nell'accertamento di circostanze possono consistere: in errori in eccesso o in difetto nella misurazione di circostanze quantitative; in errori nella collocazione in un ordinamento per gradi; in errori nell'attribuzione di modalità nella rilevazione di circostanze qualitative.

Predisponendo un'indagine, conviene prospettarsi le possibilità di errori in modo da poterli prevenire. In generale le conseguenze degli errori intervenuti nell'osservazione sono più gravi quando si tratta di circostanze continue, perchè più spesso si concretano in una irreparabile divergenza dalla realtà.

9. — È necessario evitare errori nell'osservazione perchè in generale non possono essere corretti in modo rigoroso, e spesso non possono essere corretti neppure in modo approssimativo, senza ripetere l'osservazione: il che molte volte riesce impossibile. Invece eventuali errori commessi, più tardi, nell'elaborazione dei risultati d'osservazione, possono essere in generale corretti.

Ad ogni elaborazione ed interpretazione dei risultati dell'osservazione conviene far precedere un'accurata analisi dei risultati stessi, diretta ad accertarne il grado di attendibilità. Si deve cioè ricercare se e quali errori possono essere incorsi nella rilevazione dell'oggetto dell'osservazione e delle circostanze intrinseche ed estrinseche all'oggetto stesso che siano state rilevate.

10. — I mezzi dell'osservazione possono avere la più diversa estensione: dal nulla che occorre per una semplice enumerazione compiuta mentalmente, al complicato attrezzamento di strumenti scientifici che si richiede per certe osservazioni fisiche o al poderoso complesso di mezzi materiali e personali che è indispensabile per il censimento di un grande Stato.

Per l'enumerazione di pochi casi, che si possono direttamente contare, non si richiedono mezzi nè norme particolari. Ma per l'enumerazione di grandi masse di casi si ricorre in generale all'enumerazione indiretta: si fa cioè corrispondere un simbolo a ciascun caso da enumerare, e si enumerano poi questi simboli. La rappresentazione simbolica più usata consiste nella scheda di carta o di cartoncino, sulla quale possono essere registrate circostanze intrinseche od estrinseche inerenti a ciascun caso da enumerare. Le schede sono simboli isolati, e perciò suscettibili di aggruppamento secondo una o più delle suddette circostanze: aggruppamento i cui risultati consentono poi lo studio delle relazioni tra il fenomeno osservato e le circostanze nelle quali si manifesta; è questo il loro maggior vantaggio. I simboli non isolati e non isolabili, come le annotazioni in registri, hanno invece il principale vantaggio di mantenere un determinato ordine di successione e di serbare uniti determinati gruppi di casi. Il simbolo non isolato ma isolabile (registro o foglio divisibile in tante schede) unisce in sè i vantaggi dei due tipi di simboli.

Per l'accertamento della modalità di circostanze qualitative o della misura di circostanze quantitative possono occorrere mezzi e strumenti estremamente vari secondo l'oggetto dell'indagine. Uno dei pochi precetti di carattere generale che si possano dare in questa materia consiste nell'avvertenza di evitare la predilezione per le cifre rotonde, per le cifre pari, ecc., che si manifesta sistematicamente nella lettura e nella interpretazione dei risultati di misurazioni, e in generale nelle indicazioni numeriche non indipendenti dall'apprezzamento di chi le fornisce. È opportuno, inoltre, indicare in modo completo, e con precisione in ogni caso in cui possano sorgere ambiguità, le varie modalità possibili delle circostanze qualitative rilevate: indicare, per esempio, le varie categorie predisposte in una classificazione per professioni, che dev'essere compilata in modo tale da non lasciare lacune, affinchè ciascun censito possa denunziare la categoria cui appartiene, senza che sorgano incertezze ed errori nella denunzia o più tardi nella interpretazione di essa. I confronti tra i risultati dei vari censimenti italiani, nella parte relativa alla composizione professionale della popolazione agraria, offrono molti indizi di errori di tal sorta.

Indicazioni bibliografiche. — KAUFMANN A., *Theorie und Methoden der Statistik*, Tübingen, Mohr, 1913.

Quesiti ed esercizi: 1. — Di quali operazioni consta l'osservazione statistica? In che consiste ciascuna di tali operazioni? Perchè in generale l'osservazione non è limitata ad una enumerazione?

2. — Si desuma dall'ASI qualche esempio di enumerazione di elementi (casi *osservati*); di enumerazione di eventi (casi *accertati*); di accertamento delle modalità di circostanze qualitative; di accertamento delle misure di circostanze qualitative.

3. — Si tracci uno schema di possibili operazioni di graduazione di un gruppo di persone secondo un loro carattere intellettuale o morale, che non ammetta misurazione, ma ammetta graduazione (memoria, coraggio).

4. — Si ricerchi nell'ASI qualche esempio di osservazione statistica che non abbia richiesto enumerazione di casi; qualche esempio nel quale l'osservazione invece sia stata limitata all'enumerazione dei casi.

5. — Scelto un dato fenomeno, si indichino le circostanze qualitative e quantitative che sembrano degne di essere rilevate in relazione ad esso. Si indichino le suddivisioni di modalità e di misure da adottare.

6. — Come procedereste se doveste graduare i componenti di una classe di studenti secondo l'intelligenza?

7. — Si provi a far un elenco delle difficoltà che s'incontrano per la precisa definizione o delimitazione dell'oggetto dell'indagine: in un censimento del bestiame bovino, in una rilevazione degli omicidii, in una rilevazione dei patrimoni trasmessi per successione, in una statistica dei salari, in una statistica delle esportazioni, nella registrazione dei risultati di un dato metodo terapeutico.

8. — In ciascuno degli esempi precedenti, si pensi a qualche causa che possa determinare errori in eccesso o errori in difetto nell'enumerazione; errori in eccesso o errori in difetto nella misurazione di circostanze quantitative; errori di accertamento di circostanze qualitative. E si rifletta come debba essere predisposta l'osservazione per evitare simili errori.

9. — Qual è il fondamento della distinzione tra fenomeni di stato e fenomeni di movimento?

10. — Si cerchino nell'ASI esempi di rilevazioni di fenomeni di stato e di fenomeni di movimento; di rilevazioni continue e di rilevazioni discontinue.

11. — Si metta in risalto la differenza tra rilevazione continua e discontinua: *a*) nel caso di fenomeni di stato; *b*) nel caso di fenomeni di movimento.

12. — Si cerchino nell'ASI esempi di rilevazioni nelle quali l'indicazione generica dell'oggetto dell'osservazione non basti ad assicurare un'interpretazione univoca. Si provi a completare la definizione in ciascun caso.

13. — Si desumano dall'ASI esempi di rilevazioni di circostanze qualitative continue e discontinue; di circostanze quantitative continue e discontinue. Si pensi e si accenni agli errori che possono intervenire nella rilevazione di esse, e alle loro probabili conseguenze sui risultati dell'osservazione.

14. — Si esamini come siano delimitate nello spazio e nel tempo le indagini i cui risultati sono riferiti nei capitoli «Climatologia, territorio e popolazione» dell'ASI. Si cerchi di classificare i vari criteri adottati.

15. — Si compili un modello di scheda per un censimento degli studenti universitari; un modello per una inchiesta sui disoccupati; un modello per una indagine sulle vendite d'automobili da parte di una fabbrica; un modello per una indagine sulla delinquenza dei minorenni.

16. — Quali sono i vantaggi e gli inconvenienti: *a*) del simbolo isolato, *b*) del simbolo non isolabile, *c*) del simbolo isolabile?

17. — Si contrappongano i vantaggi e gli svantaggi del simbolo isolato (scheda) e dei simboli non isolabili (registro) nei seguenti esempi: statistica dei risultati di esami universitari; statistica dei matrimoni; statistica delle vendite di immobili. Vantaggi e svantaggi del foglio di famiglia e della scheda individuale nel censimento della popolazione.

18. — Si cerchino nell'ASI esempi di rilevazioni che possano presumersi affette da errori non trascurabili e si mettano in evidenza le cause di errore che si suppone abbiano agito.

CAPITOLO III.

L'osservazione approssimativa; l'osservazione parziale.

Stime — Ragioni che consigliano di ricorrervi — Criteri per la stima — Indagini rappresentative — Criteri per l'esecuzione — Vantaggi e svantaggi.

1. — Riesce talvolta impossibile l'osservazione precisa dei fatti che interesserebbe rilevare. L'impossibilità può essere assoluta, come nella rilevazione del valore dei beni trasmessi per successione, che evidentemente non si possono mettere in vendita al solo intento di determinarne il vero valore (d'altronde con una tal massa di vendite si turberebbe il mercato e si determinerebbe un valore artificialmente depresso per l'eccesso dell'offerta); oppure soltanto relativa, come nella rilevazione della produzione frumentaria di un paese, per l'esecuzione della quale non si possono obbligare centinaia di migliaia di produttori a pesare il loro raccolto e a denunziarne il peso esatto, nè riesce attuabile un controllo completo delle denunzie. In simili casi si supplisce alla mancanza di informazioni precise mediante informazioni approssimative ricavate da procedimenti di *stima*. A ciascun bene trasmesso per successione si attribuisce un valore presunto, dedotto dai valori di beni affini per natura e per caratteri, trasferiti a titolo oneroso. A ciascun ettaro coltivato a frumento in una zona agraria omogenea si attribuisce un rendimento pari alla media dedotta da informazioni raccolte qua e là nella zona stessa. In questi esempi si attribuiscono a determinati oggetti d'osservazione, che non si prestano alla rilevazione diretta, caratteri uguali od analoghi a quelli di altri oggetti affini, direttamente osservabili ed osservati. Quando sussista una grande analogia tra questi oggetti osservati e gli altri non osservati, la stima darà buoni risultati; potrà darli buoni anche se l'analogia è minore, purchè l'esecutore della stima sappia tener conto, nell'eseguirla, dei divari esistenti fra gli oggetti osservati e quelli non osservati.

L'attendibilità dei risultati talvolta è diminuita dalla circostanza che essi implicano non una ma più operazioni di stima reciprocamente collegate. Ad esempio, per stimare il raccolto del frumento, si parte da un dato sulla superficie coltivata a frumento che è esso medesimo il risultato di una stima, non essendo possibile procedere in ogni campagna agraria ad una rigorosa misurazione di tale superficie, che varia da campagna a campagna. Una seconda stima occorre per stabilire qual frazione dell'area coltivata sia stata abbandonata prima della mietitura; collegando i risultati delle due stime, si determina la superficie sulla quale si è mietuto. A questa poi si applica un rendimento medio per ettaro, che risulta da una terza stima. Col moltiplicarsi delle stime occorrenti per giungere al risultato finale, si moltiplicano le possibilità di errore. Il successo del procedimento dipende dalla competenza degli stimatori e dai mezzi d'informazione ond'essi dispongono.

2. — In altri casi la stima è fondata sull'ipotesi dell'esistenza di una certa relazione di grandezza fra una quantità nota e quella ignota che si desidera conoscere. Si calcola la ricchezza privata di un paese ammettendo che essa costituisca un certo multiplo del valore accertato dei beni trasmessi annualmente per successione o donazione; il consumo annuo della carne bovina nazionale ammettendo che stia in una certa proporzione al numero accertato dei capi bovini esistenti; la spesa totale di una popolazione supponendo che essa corrisponda ad un certo multiplo della spesa, nota, per l'alimentazione. I coefficienti che vengono applicati in questi calcoli costituiscono il punto vulnerabile dell'operazione: se sono desunti da esperienza insufficiente, o da errate interpretazioni di fatti, o se vengono applicati senza opportune correzioni, la stima può risultare profondamente divergente dalla realtà. Anche qui spesso l'attendibilità del risultato dipende strettamente dall'abilità dello stimatore.

3. — È ovvia la conclusione che conviene sostituire alla stima l'osservazione rigorosa sempre quando sia possibile; ma non è meno ovvio il rilievo che in molti casi la necessità di pronta informazione, l'opportunità di evitare eccessive spese di rilevazione, il desiderio di evitare vessazioni o fastidi al pubblico, ed altre considerazioni di carattere pratico, impongono o consigliano il ricorso alla stima.

Talvolta la stima è insurrogabile perchè dà qualche indicazione che l'osservazione rigorosa non potrebbe fornire: per esempio una valutazione approssimativa del raccolto del frumento quindici giorni o un mese prima della mietitura, quando una misurazione precisa non è possibile nè praticamente nè teoricamente.

4. — Non di rado la stima viene fondata sui risultati di quella che si chiama *indagine rappresentativa*: di una rilevazione, cioè, rigorosa ma limitata ad una sola frazione del campo di osservazione prefisso. Si possono accertare con precisione i rendimenti per ettaro

in alcune aziende agricole per applicare poi il rendimento medio così accertato all'intera zona agraria cui quelle aziende appartengono; registrare le entrate e le spese di qualche decina di famiglie operaie per applicare poi i risultati all'intera popolazione operaia della città dove quelle famiglie vivono; determinare i caratteri antropologici di uno o due iscritti di leva su ogni cento visitati, per attribuire poi i risultati all'intera massa degli iscritti.

5. — La frazione del campo d'osservazione cui va ristretta l'indagine rappresentativa può essere scelta con due criteri, in apparenza profondamente diversi: quello della scelta ragionata e quello della scelta a caso.

Nella scelta ragionata si procura di far sì che l'insieme degli oggetti sottoposti ad osservazione costituisca, per i suoi caratteri, una immagine ridotta del più vasto insieme che non è possibile osservare totalmente: si cerca per esempio che gli iscritti di leva sottoposti a speciali misurazioni siano distribuiti per luogo di nascita, per condizione sociale, ecc., press'a poco nelle stesse proporzioni nelle quali è distribuita secondo le medesime circostanze l'intera classe di leva.

Nella scelta a caso, invece, si procura che nessun carattere particolare dei singoli oggetti osservati, rilevante agli effetti della indagine in corso, influisca sulla scelta: per esempio, si determina quali iscritti di leva debbano essere sottoposti a speciali misurazioni, non scegliendoli secondo la statura o la professione, o la legittimità dei natali, ma estraendo alcuni nomi da un'urna dove siano contenuti i nomi di tutti, o prendendo quelli che compaiono in una stessa riga nelle varie pagine del registro degli iscritti.

Nonostante la profonda differenza dei procedimenti, anche con la scelta a caso si tende ad ottenere una immagine ridotta dell'insieme che non si può osservare totalmente, l'esperienza avendo dimostrato che tale è appunto il risultato della scelta a caso, quando il numero dei casi scelti sia sufficiente. V'è tutta una teoria della scelta a caso, nata dalla teoria dei giuochi di sorte, alla quale si possono attingere le opportune norme. Invece il risultato della scelta ragionata è strettamente dipendente dalla competenza di chi l'esegue.

Talvolta è indifferente adottare l'uno o l'altro metodo di scelta dei casi da sottoporre ad indagine rappresentativa; tal altra è indispensabile adottare l'uno piuttosto che l'altro di essi: sarebbe assurdo, per esempio, scegliere a caso le famiglie operaie cui si vuole chiedere la compilazione del proprio bilancio, mentre certamente molte famiglie scelte a caso si rifiuterebbero di compilarlo; sarebbe difficile, nell'altro esempio, scegliere razionalmente gli iscritti di leva da misurare. È consigliabile, talora, la combinazione dei due metodi: fissato da prima per ciascuna circoscrizione territoriale il numero degli iscritti di leva da sottoporre a misurazioni speciali (scelta ragionata), determinare poi le persone mediante estrazione a sorte fra gli iscritti di ciascuna circoscrizione (scelta a caso).

6. — Lo svantaggio dell'indagine rappresentativa è palese: la frazione osservata non darà mai l'immagine, ridotta sì ma precisa, dell'insieme che si vorrebbe conoscere; ne differirà più o meno non soltanto quantitativamente ma anche qualitativamente. Il vantaggio maggiore è quello che la limitazione del campo di osservazione permette di approfondire l'indagine più di quanto sarebbe possibile in un campo maggiormente vasto e di controllarne minutamente i risultati in modo da conseguire un alto grado di attendibilità. Sono ovvi gli ulteriori vantaggi del risparmio di tempo e di spesa, che talora possono riuscire decisivi a favore di questo metodo d'investigazione. Il quale talvolta viene impiegato non in sostituzione di rilevazioni complete, ma nella preparazione di esse, per così dire come sondaggio: l'esecuzione dell'indagine in un campo ristretto indica se convenga semplificare, rettificare o modificare i criteri di rilevazione nell'estensione ad un campo più vasto.

Quesiti ed esercizi: 1. — Perchè si applicano procedimenti di stima? In che consistono? Quali ne sono i fondamenti logici?

2. — Si ricerchino esempi di stime nei capitoli dell'ASI relativi all'agricoltura, alle industrie, alle finanze dello Stato. Si mettano in rilievo le principali possibilità di errori nell'esecuzione di tali stime.

3. — Si pensi se e come potrebbe compilarsi, mediante procedimenti di stima, una statistica delle nascite o delle morti avvenute in un anno in Italia; una statistica dei furti; un calcolo della somma dei redditi privati; una rilevazione della distribuzione, nei vari quartieri della città, delle vendite di un grande magazzino; una statistica dei trasferimenti a titolo oneroso della proprietà rurale. Come e in qual misura si incontrerebbero possibilità di errori in tali stime?

4. — Che cos'è una *indagine rappresentativa?*

5. — In quali fra gli esempi dell'esercizio 3 sarebbe utilmente applicabile il procedimento dell'indagine rappresentativa? Con quali criteri si dovrebbe istituire tale indagine in ciascun caso? Quando sarebbe preferibile la scelta a caso; quando la scelta ragionata?

6. — In qual modo le stime preliminari, simultanee e posteriori al raccolto di un prodotto agricolo possono dare norma al mercato di esso? In qual momento può subentrare alla stima l'osservazione precisa?

7. — Come si potrebbe eseguire una indagine rappresentativa sul traffico dei veicoli nelle strade di una grande città? Quali vantaggi e quali svantaggi presenterebbe in confronto ad un'indagine completa? A quali errori potrebbe dar luogo?

8. — Come potrebbe un fabbricante di automobili istituire un'indagine rappresentativa sui risultati dati nell'uso dalle macchine vendute alla clientela? Quali cautele dovrebbe adoperare nella scelta dei casi da rilevare?

9. — Come si potrebbe eseguire un'indagine rappresentativa sulle corrispondenze postali in partenza da uno Stato? Quali sarebbero le principali cause di errore? Come si potrebbe cercar di correggere i risultati della indagine tenendo conto delle vendite di francobolli, ecc.?

10. — Come andrebbe correttamente impiantata un'indagine rappresentativa sul peso netto di carni ricavabile da ciascun capo bovino macellato in una grande città? Quali sarebbero le più ovvie cause di errore?

11. — Come si potrebbe utilmente eseguire una indagine rappresentativa sul rendimento dei bozzoli in seta, per una regione italiana?

12. — Si rifletta se e come si potrebbero eseguire stime senza ricorrere ad indagini rappresentative. E se le indagini rappresentative siano necessariamente dirette a fornire elementi per stime o possano avere altri fini.

13. — Si pensi come potrebbero eseguirsi stime preliminari del traffico ferroviario di un anno sull'esperienza del primo trimestre; delle importazioni di frumento in una campagna granaria sull'esperienza del primo semestre.

CAPITOLO IV.

La sintesi dei risultati dell'osservazione.

I risultati numerici dell'osservazione — Le operazioni di aggruppamento — Aggruppamento delle osservazioni per circostanze isolate; per circostanze combinate — Scopo dell'aggruppamento — Delimitazione dei gruppi — Esecuzione con mezzi meccanici delle operazioni di aggruppamento e di enumerazione.

1. — Quando l'osservazione di fenomeni collettivamente tipici si riduce ad una semplice enumerazione, come in un censimento limitato all'accertamento del numero degli abitanti senza alcuna rilevazione di loro caratteri individuali intrinseci od estrinseci, dalla enumerazione stessa si ottiene immediatamente l'unico risultato cui si mira: un numero. Quando l'osservazione si riduce ad una semplice misurazione, come nell'accertamento delle precipitazioni atmosferiche avvenute in ciascun giorno dell'anno, dalla misurazione stessa si ottiene immediatamente un risultato numerico. I risultati dell'enumerazione delle popolazioni nelle diverse provincie, i risultati della misurazione delle precipitazioni nei diversi giorni dell'anno si possono riguardare come tanti risultati parziali, che po-

tranno venir aggruppati con vari criteri, o potranno venir comparati individualmente fra loro. Quando l'osservazione si riduce ad una graduazione di oggetti osservati, ciascuno di questi, ad osservazione compiuta, è caratterizzato da un posto, o numero d'ordine, che indica il risultato delle osservazioni, la sintesi completa delle quali consiste nell'intero ordinamento degli oggetti osservati: così per esempio se si sono ordinati i discepoli di una classe secondo la loro memoria. E infine, quando l'osservazione sia limitata all'accertamento della modalità di una circostanza qualitativa, si riassumerà in tanti numeri quante sono le modalità distinte; per esempio in due numeri, quello dei giorni sereni e quello dei giorni nuvolosi, se per ciascun giorno dell'anno si è accertato lo stato del cielo: sereno o coperto. Le diverse operazioni delle quali consta l'osservazione statistica, separatamente eseguite, conducono dunque sempre a numeri, che ne esprimono i risultati.

2. — Anche eseguite congiuntamente, le operazioni stesse conducono a risultati numerici. Se nell'enumerare la popolazione si sono rilevati caratteri qualitativi e quantitativi dei singoli abitanti, per esempio stato civile ed età, si potrà avere, accanto al numero che indica l'ammontare totale degli abitanti, un complesso di numeri che indicano quanti abitanti sono contrassegnati da determinate modalità o misure di singoli caratteri: per esempio quanti sono i celibi, quanti sono coloro che hanno da 70 a 71 anni; o da determinate combinazioni di due o più caratteri: per esempio quanti sono i celibi, di 70 a 71 anni, nativi di una determinata provincia. A numeri di questo genere si giunge attraverso operazioni di aggruppamento, di enumerazione, di addizione, compiute sugli oggetti dell'osservazione, o più spesso sui loro simboli: nell'esempio precedente bisognerà aver aggruppato le schede che rappresentano i singoli abitanti censiti, secondo lo stato civile; in ciascuna classe di stato civile aver proceduto ad un aggruppamento per singoli anni d'età, e in ciascuno dei gruppi così formati aver compiuto infine un aggruppamento per provincie di nascita.

È chiaro che si possono eseguire aggruppamenti secondo singole circostanze (*aggruppamenti di prim'ordine*), o secondo più circostanze congiuntamente considerate (*aggruppamenti di secondo, terzo, ecc., ordine*): gli uni permetteranno di studiare il fenomeno osservato in relazione a ciascuna circostanza, gli altri permetteranno di studiarlo in relazione a combinazioni di circostanze e quindi anche di ricercare eventuali relazioni esistenti fra tali circostanze.

Dato questo fine cui mira l'aggruppamento, è palese che lo schema di esso dovrà essere normalmente stabilito prima dell'inizio della osservazione; ed a maggior ragione dovranno essere stabilite le singole circostanze da rilevare. La scelta di queste dev'essere compiuta da persona specialmente competente nella materia che è oggetto d'indagine, perchè è determinata appunto dall'utilità che si presume di poter trarre dalla loro conoscenza nello studio del fenomeno, che si esaminerà in relazione ad esse; se le circostanze non fossero state rilevate insieme col fenomeno, tale indagine riuscirebbe spesso difficile od impossibile.

3. — Quando la circostanza in relazione alla quale si compie l'aggruppamento è discontinua, molte volte la modalità o misura accertata indica senz'altro il gruppo nel quale ciascun caso va inserito. Così l'indicazione dello stato civile del censito, contenuta nella scheda, definisce immediatamente la sua appartenenza al gruppo dei celibi, o a quello dei coniugati, o a quello dei divorziati, o a quello dei vedovi; e l'aggruppamento consiste soltanto nel riunire insieme le schede di ciascuna classe di stato civile. Tuttavia se le modalità o misure della circostanza osservata sono molto numerose, conviene in generale radunarle in un minor numero di gruppi, e qui ha campo l'arbitrio dell'ordinatore dell'indagine: così in un censimento professionale, dove si raduneranno in un sol gruppo occupazioni affini.

Tale arbitrio diviene poi inevitabile quando la circostanza è continua: allora, prima di iniziare l'aggruppamento, e per poterlo iniziare, conviene fissare i limiti, altrimenti indeterminati, dei gruppi che si vogliono distinguere: per esempio stabilire che i censiti devono essere riuniti per gruppi annuali d'età, da 0 a 1 anno, da 1 a 2, da 2 a 3, ecc.; che le misure della statura devono essere riunite in gruppi dell'ampiezza di un centimetro, da m. 1,495 a 1,505, da 1,505 a 1,515, da 1,515 a 1,525, ecc. Col fissare i limiti dei gruppi si determina senz'altro lo schema dell'aggruppamento secondo la singola circostanza; associando le varie modalità o misure possibili di più caratteri si determina poi lo schema dell'aggruppamento per circostanze combinate. La determinazione dei limiti da assegnare ai vari gruppi conviene sia eseguita da persona competente nella materia dell'indagine.

Uno schema d'aggruppamento di prim'ordine contiene tanti gruppi quante sono le modalità o misure della singola circostanza, che si vogliono tener distinte; in uno schema d'aggruppamento di second'ordine il numero dei gruppi è uguale al prodotto dei numeri dei gruppi fissati per l'una e per l'altra circostanza, ecc.

4. — Stabilito lo schema di aggruppamento, eseguito l'aggruppamento dei casi o dei loro simboli secondo lo schema, resta da contare il numero dei casi compresi in ciascun gruppo, e talvolta restano da sommare i valori di circostanze quantitative caratteristiche di essi. Per esempio, aggruppati i contribuenti secondo la professione, resta da contare quanti contribuenti siano compresi in ciascun gruppo professionale e restano da sommare i redditi dei contribuenti di ciascun gruppo. Tutte queste operazioni, compreso l'aggruppamento, si possono compiere con mezzi meccanici.

Per la sintesi meccanica dei risultati dell'osservazione si procede così. A ciascun caso da enumerare si fa corrispondere una scheda speciale di cartoncino, le cui dimensioni — nel tipo più comune poco differenti da quelle di una cartolina postale — sono fissate in modo da corrispondere ad esigenze di carattere pratico: maneggevolezza, resistenza, spazio sufficiente per le indicazioni da apporvi, ecc. La scheda è divisa orizzontalmente in tante fascie di uguale altezza e verticalmente in tante colonne di uguale larghezza, numerate progressivamente; nel tipo più comune, da 0 a 9 le fascie, da 1 a 45 le colonne. Una colonna consente la registrazione di numeri da 0 a 9 (e quindi di dieci modalità o misure di una circostanza); due colonne consecutive, una assegnata alle decine e l'altra alle unità consentono la registrazione dei numeri da 0 a 99 (e quindi di cento modalità o misure), e così via. Si assegnano, secondo il bisogno, una o più colonne a ciascuna circostanza: per esempio, in un censimento, la colonna 1 al sesso, la colonna 2 allo stato civile, le colonne 3 e 4 all'età del censito, ecc. Il numero contrassegnato in una colonna, o la combinazione di numeri contrassegnati in più colonne, indica una data modalità o misura di un carattere: per esempio nella prima colonna si assegnerà la fascia 0 al sesso maschile, la fascia 1 al sesso femminile; nella seconda la fascia 0 ai celibi, la 1 ai coniugati, la 2 ai divorziati, la 3 ai vedovi; nella terza la fascia 0 al primo decennio d'età, la 1 al secondo decennio, ecc.; nella quarta la fascia 0 al primo anno d'età di ciascun decennio, la 1 al secondo, ecc. (così che per esempio volendo indicare l'età di 24 anni si dovrà contrassegnare il 2 nella terza colonna e il 4 nella quarta). Si contrassegnano le misure o modalità rilevate, non graficamente, bensì mediante fori circolari, eseguiti con una apposita macchina che li localizza in modo preciso, così che tutti i fori eseguiti nel rettangolo determinato dall'incrocio di una data fascia con una data colonna si sovrappongono esattamente quando sono esattamente sovrapposte le schede.

Tradotte sulla scheda perforata le indicazioni rilevate, si passa all'aggruppamento, procedendo colonna per colonna. Le schede vengono introdotte in una macchina classificatrice, nella quale, per lo stabilirsi di un diverso contatto elettrico (o meccanico) secondo che la colonna è perforata in corrispondenza ad una o ad altra delle varie fascie, la scheda è avviata ad una o ad altra cassettina raccoglitrice. Le cassettine sono in numero almeno uguale a quello delle fascie; in ciascuna di esse si ritrovano, ad operazione compiuta, tutte le schede perforate in una stessa fascia della stessa colonna, cioè tutte le schede corrispondenti ad una determinata modalità o misura della circostanza cui la colonna è dedicata. In generale v'è una cassettina in più per raccogliere le schede che non presentino nessuna perforazione nella colonna data. Un contatore annesso a ciascuna cassettina enumera le schede che vi si raccolgono: si compiono, così, simultaneamente le due operazioni dell'aggruppamento e della enumerazione dei gruppi. La grandissima rapidità con la quale tali operazioni vengono eseguite compensa largamente, in rilevazioni vaste e particolareggiate, il maggior tempo occorrente per la preparazione delle schede perforate, e rende possibile la pronta sintesi dei risultati di amplissime indagini. Questo vantaggio talora è decisivo, perchè certe informazioni prontamente conosciute possono essere di utilità molto maggiore che se conosciute più tardi.

Una terza macchina permette di addizionare i valori di circostanze quantitative registrate sulla scheda: per esempio di sommare le età dei censiti di ciascun gruppo per poterne poi computare l'età media; anche quest'operazione è compiuta automaticamente e celermente.

Indicazioni bibliografiche: MINISTERO DELLA GUERRA, DIREZIONE GENERALE DEI SERVIZI LOGISTICI, *Corso di calcolo meccanico*, Roma, Libreria dello Stato, 1927.

Quesiti ed esercizi: 1. — In quale forma si presentano i risultati dell'osservazione statistica?

2. — Che cosa significa « aggruppamento di prim'ordine », « aggruppamento di second'ordine »? A che serve l'aggruppamento delle osservazioni?

3. — Si desumano dall'ASI esempi di aggruppamenti statistici di secondo ordine; si ricerchino gli intenti che possono averli suggeriti; si pensi se gli

stessi fenomeni, osservati con maggiori particolari, avrebbero potuto dar luogo utilmente ad aggruppamenti di terzo, di quarto ordine. Quali sono le ragioni che inducono ad eseguire aggruppamenti secondo circostanze combinate, a preferenza di quelli secondo circostanze isolate?

4. — Si riscontri come da un aggruppamento secondo n circostanze combinate possano desumersi tutti i possibili aggruppamenti secondo $(n-1)$, $(n-2)$,..., 2, 1 circostanze. È possibile da due aggruppamenti, ciascuno eseguito secondo una circostanza, desumere l'aggruppamento secondo le due circostanze combinate?

5. — Si ricerchino nell'ASI esempi nei quali si abbiano aggruppamenti secondo due circostanze separate mentre sarebbe stato più utile per lo studio del fenomeno avere l'aggruppamento secondo le due circostanze combinate.

6. — Partendo da una scheda predisposta per il censimento degli studenti universitari, si preparino schemi di aggruppamento dei risultati del censimento secondo circostanze isolate e secondo circostanze combinate, dove ciò sembri utile.

7. — Come si può eseguire l'enumerazione e la classificazione meccanica delle schede contenenti i risultati di una rilevazione statistica? Si adoperano a tal uopo le schede originali?

8. — Data una scheda per l'enumerazione e la classificazione meccanica, di tipo normale, si studi come si potrebbe utilizzarla per una statistica dei risultati dei diversi esami universitari subiti dagli studenti giunti al termine dei loro studi (si farà corrispondere una scheda a ciascuno studente o a ciascun esame?).

9. — Come si potrebbe utilizzare la stessa scheda per una statistica delle vendite di una fabbrica di macchine utensili? per una statistica delle società per azioni? per una statistica delle nascite illegittime? per una statistica dei risultati delle corse al galoppo in una stagione?

10. — Con quali criteri si potrebbero raggruppare le osservazioni in una statistica dei redditi privati di una popolazione? in una rilevazione del colore degli occhi dei coscritti? in una rilevazione dello stato del mare nei vari giorni dell'anno?

LIBRO SECONDO
La descrizione dei fenomeni

—

CAPITOLO V
Il dato statistico.

Dati statistici greggi e dati statistici elaborati — Criteri per l'apprezzamento di un dato statistico — Errore di un dato statistico — Rappresentazione grafica.

1. — Come abbiamo visto, i risultati dell'osservazione statistica assumono, nella loro espressione definitiva, la forma di numeri, i quali esprimono: il risultato di enumerazioni parziali o dell'enumerazione totale dei casi sottoposti all'osservazione, oppure il risultato di singole misurazioni di circostanze quantitative, o una somma di siffatte misurazioni, oppure ancora il risultato di una graduazione, oppure infine il risultato di una stima. Tutti questi numeri si dicono *dati statistici*; si aggiunge la qualificazione di *greggi* o *diretti*, per distinguerli dai *dati statistici elaborati* o *indiretti*, che ne vengono ricavati mediante operazioni di calcolo o di ordinamento. Il salario di un operaio, la somma dei salari degli operai di una fabbrica, sono dati greggi; il salario medio giornaliero degli operai di una fabbrica, il salario medio per ogni unità di prodotto, sono dati elaborati.

2. — Un dato statistico isolatamente considerato rappresenta la misura di una grandezza espressa in una determinata unità, rispetto alla quale soltanto il dato può venir apprezzato finchè non si offrano altri termini di riferimento. Dicendo che l'Italia ha 41 milioni di abitanti, siamo in grado di valutare la popolazione del nostro paese in confronto all'unità uomo; ma se vogliamo poi renderci conto di quello che significhi per l'Italia, come Stato o come nazione, avere quel numero di abitanti dovremo confrontarlo con la

superficie territoriale, con la popolazione vivente sul territorio stesso in altre epoche, con quella attuale di altri paesi, con quella del mondo intero, ecc; e secondo che sceglieremo l'uno o l'altro termine di comparazione, quel numero ci potrà apparire grande, piccolo, mezzano, ecc. In generale per l'apprezzamento del dato statistico è necessario compararlo con altre grandezze: la comparazione con un'altra grandezza soltanto si esegue mediante differenza o mediante rapporto; la comparazione con più grandezze ad un tempo si esegue coll'inserire il dato in un insieme di dati omogenei con esso (serie statistica) ai quali viene poi comparato con metodi vari. Tali metodi di comparazione dei dati, e quindi di apprezzamento dei fenomeni da essi rappresentati, saranno studiati più avanti.

3. — Conviene intanto fissare la nozione dell'errore di un dato statistico, che sarà utile tener presente appunto nello studio dei metodi di comparazione, e in generale di elaborazione, dei dati. Indicato come « dato noto » il risultato dell'osservazione e come « dato esatto » il risultato che l'osservazione stessa avrebbe fornito se scevra da errore, la differenza tra il dato noto e il dato esatto misura *l'errore del dato statistico.* Se tale differenza risulta positiva, diremo che il dato noto è errato *in eccesso*, se negativa che è errato *in difetto*.

Poichè un uguale errore ha rilevanza molto diversa, in generale, secondo la grandezza cui si riferisce, spesso conviene mettere in rapporto l'errore del dato statistico noto col dato esatto; il valore di questo rapporto dicesi *errore relativo* del dato statistico; intendasi: errore del dato noto in rapporto al dato esatto. Se la statistica agraria ci indica un raccolto di 64 milioni di quintali di frumento, mentre il raccolto in realtà è asceso a soli 60 milioni, diremo che il dato noto è affetto da un errore in eccesso di 4 milioni di quintali, che in misura relativa corrisponde al 6,66 °/₀ del dato esatto. In pratica, quando non si ha alcuna cognizione del senso dell'errore, ma si ha qualche cognizione della sua presumibile misura, si riferisce talvolta la misura presunta dell'errore al dato noto, invece che al dato esatto, tanto per avere un'approssimativa idea della importanza relativa dell'errore. Per esempio, nel caso precedente, se si sapesse soltanto che il dato sul raccolto del grano può essere errato al massimo di 4 milioni di quintali in eccesso o in difetto, si potrebbe dire, col criterio ora accennato, che la sua approssimazione è di 4 su 64, cioè del 6,25 °/₀.

Ma parlando di errore relativo è bene riferirsi esclusivamente all'errore riferito al dato esatto, come lo abbiamo definito dianzi.

4. — Un dato statistico si può rappresentare graficamente mediante un segmento od un'area ad esso proporzionale, in una scala preventivamente fissata. Il metodo, assolutamente inutile per la mancanza di ogni termine di comparazione, quando si applichi a un unico dato, riesce invece vantaggioso in molti casi per il confronto fra più dati, come spiegheremo più avanti.

Quesiti ed esercizi: 1. — Che cos'è un dato statistico? La longitudine e la latitudine di un dato punto della superficie terrestre sono dati statistici? La quantità di pioggia caduta in un anno a Milano è un dato statistico? La superficie della Lombardia è un dato statistico? La superficie coltivata a grano in Lombardia è un dato statistico? Il numero dei matrimoni avvenuto in un anno in un comune è un dato statistico? L'età di una sposa all'atto del matrimonio è un dato statistico?

2. — Si desumano dall'ASI esempi di dati statistici greggi i quali rappresentino: *a*) un numero totale di casi osservati; *b*) numeri di casi osservati appartenenti a singoli gruppi caratterizzati da una o più circostanze; *c*) un numero totale di casi accertati; *d*) numeri di casi accertati appartenenti a singoli gruppi caratterizzati da una o più circostanze; *e*) misure di un fenomeno di stato in un determinato istante; *f*) misure di un fenomento di movimento in un determinato intervallo di tempo; *g*) misure di una circostanza quantitativa caratteristica del singolo caso osservato; *h*) una somma di misure come sopra per un gruppo di casi.

3. — Si ricerchi nell'ASI quali dati elaborati vengono ricavati dai dati greggi precedentemente desunti (esercizio 2) dall'annuario stesso.

4. — Si indichi con quali grandezze possano opportunamente essere comparati, da chi vuol giungere a un adeguato apprezzamento, i seguenti dati, i cui valori numerici possono essere desunti dall'ASI: la portata massima del Po; il numero dei nati vivi illegittimi nell'Emilia; il numero dei morti di vaiuolo; il numero degli alunni iscritti alle scuole elementari; il numero delle separazioni personali di coniugi avvenute in un anno; il numero degli omicidii volontari denunziati; la produzione del frumento; la quantità prodotta di energia idroelettrica; il consumo del tabacco; l'esportazione della seta tratta greggia; la quantità delle merci imbarcate e sbarcate nei porti; il numero delle comunicazioni telefoniche interurbane; il numero degli operai occupati in una data industria in una determinata epoca; il numero degli infortuni del lavoro avvenuti in una data industria; l'ammontare dei depositi presso le casse di risparmio ordinarie; l'ammontare degli sconti fatti dalla Banca d'Italia in Piemonte; il getto della imposta fondiaria.

5. — Che cos'è l'errore di un dato statistico? Quando diciamo che il dato è errato in eccesso? in difetto? Che cos'è l'errore relativo di un dato statistico? Se un censimento ci indica un dato numero di abitanti, che cosa chiamiamo « errore » di questo numero? Se un medico militare ci dà un certo numero di centimetri come perimetro toracico di un determinato coscritto, che cosa chiamiamo « errore » di questo numero?

6. — Confrontando le « cifre rilevate » (dati noti) con le « cifre corrette »

del valore del commercio con l'estero, riferite nell'ASI, si determini l'errore di ciascuna delle cifre rilevate, supponendo che le cifre corrette rappresentino i dati esatti (avvertasi che si tratta d'ipotesi arbitraria, fatta al solo scopo dell'esercitazione). Si calcoli poi l'errore relativo di ciascuna delle cifre rilevate.

7. — Si calcoli l'errore e l'errore relativo dei dati sul movimento generale dell'emigrazione, assunti come dati esatti i « numeri più attendibili degli emigrati » indicati nell'ASI.

8. — Si traducano in una formola algebrica le definizioni di errore e di errore relativo di un dato statistico.

CAPITOLO VI

La comparazione fra due dati statistici mediante differenza.

Differenza, eccedenza, deficienza — Eccedenze e deficienze relative — Indici del grado di disuguaglianza, del grado di eccedenza, del grado di deficienza — Errore di una differenza — Rappresentazione grafica.

1. — Per l'apprezzamento di un dato statistico occorre in generale compararlo con un'altra grandezza, che può essere omogenea con esso od eterogenea. Per esempio la quantità di granturco raccolta quest'anno si può comparare con la quantità raccolta l'anno scorso (dato omogeneo) o con la superficie coltivata (dato eterogeneo). Per il confronto tra due dati omogenei può servire la *differenza*, la quale invece non si può istituire tra due dati eterogenei.

Possiamo computare in due modi la differenza tra due dati: o considerando la differenza tra il più grande e il più piccolo, che risulta sempre positiva e che chiameremo *eccedenza*, o la differenza tra il più piccolo e il più grande, che risulta sempre negativa e che chiameremo *deficienza*. La preferenza per l'una o per l'altra forma di comparazione dipende da ragioni pratiche: per esempio se il raccolto di quest'anno è maggiore di quello dell'anno scorso calcoleremo l'eccedenza dell'uno sull'altro; se è minore calcoleremo la deficienza: prenderemo quindi in entrambi i casi come minuendo il raccolto di quest'anno perchè più immediatamente ci interessa.

2. — La differenza misura l'eccedenza del dato maggiore sul minore, o la deficienza del minore rispetto al maggiore, ma fa perdere di vista entrambe le grandezze che si confrontano, se, come di solito avviene, ci si limita a considerare il valore della differenza, invece di considerare insieme con essa i due termini che la costi-

tuiscono. In codesto modo diviene difficile apprezzare la rilevanza della differenza stessa, il che non è privo d'inconvenienti: una diminuzione di 100 mila quintali in un raccolto agrario ha significato ben diverso secondo che si riferisce a un raccolto di 100 milioni, o di 10 milioni, o di 200 mila quintali. Appunto per apprezzare il significato di una differenza, si usa calcolare la differenza relativa quando, come in generale avviene, i due dati che si confrontano hanno lo stesso segno. Il rapporto tra l'eccedenza e il dato minore misura l'*eccedenza relativa* che è sempre positiva, il rapporto tra la deficienza e il dato maggiore misura la *deficienza relativa* che è sempre negativa. Se il raccolto di quest'anno è stato di 70 milioni di quintali, quello dell'anno scorso di 60 milioni, l'eccedenza relativa del primo sul secondo è uguale a 0,1666; la deficienza relativa del secondo rispetto al primo è uguale a — 0,1429. In pratica si dirà: un'eccedenza di 16,66 %; una deficienza di 14,29 %.

Il valore assoluto della deficienza relativa (cioè il valore di essa riguardato come positivo) può variare fra i due limiti di 0 nel caso di uguaglianza tra i due dati e di 1 nel caso di massima disuguaglianza possibile (cioè quando il dato minore è nullo mentre il maggiore è positivo); perciò è adatto a servire come *indice del grado di disuguaglianza* tra i due dati. Se, però, il dato minore per sua natura non può discendere a zero, ed incontra un limite determinato, mentre il dato maggiore è invariabile, od è assunto come tale, conviene invece calcolare un *indice del grado di deficienza* ponendo in rapporto il valore assoluto della deficienza accertata col suo massimo valore assoluto possibile, che è uguale alla differenza tra il dato maggiore e il minimo valore che può assumere il dato minore. In un'industria dove il costo unitario di produzione più basso ottenibile nelle attuali condizioni tecniche è 60, un'azienda riesce a ridurre il proprio costo da 100 a 80. La deficienza relativa è uguale a — 0,20 e significa che il costo è stato ridotto del 20 %; l'indice del grado di deficienza è 0,50 e significa che la riduzione conseguita nel costo è la metà della massima riduzione possibile.

Il valore dell'eccedenza relativa ha come limite inferiore 0 ma in generale non ha alcun limite superiore determinato; se però il dato maggiore per sua natura non può superare un determinato limite, mentre il dato minore è invariabile od è assunto come tale, conviene calcolare un *indice del grado di eccedenza* ponendo in rapporto il valore assoluto dell'eccedenza accertata col suo massimo valore possibile, che è uguale alla differenza fra il massimo valore che può assumere il dato maggiore e il valore effettivo del dato minore. Su 100 vaccinati 20 sono colpiti dalla malattia che la vaccinazione tendeva a prevenire; su 100 non vaccinati sono colpiti 80. L'eccedenza relativa di 80 su 20, cioè 3, indica che il numero dei colpiti è maggiore del 300 °/₀ fra i non vaccinati che fra i vaccinati; l'indice del grado di eccedenza 0,75 denota che l'eccedenza dei colpiti accertata fra i non vaccinati, in confronto ai vaccinati, corrisponde al 75 °/₀ del massimo possibile.

3. — Viene talvolta adottato come indice del grado di disuguaglianza, invece di quello definito nel paragrafo precedente, il rapporto fra l'eccedenza del dato maggiore sul minore e la somma dei due dati: rapporto che varia anch'esso tra 0 nel caso di uguaglianza fra i due dati e 1 nel caso di massima disuguaglianza, ma che, eccettuati i due casi estremi, resta sempre inferiore all'altro indice di dianzi. Questo rapporto misura l'eccedenza relativa del dato maggiore rispetto ad un livello ideale di uguaglianza equidistante dai due dati (semisomma di essi); perciò in certi casi può essere opportunamente impiegato nelle comparazioni. Fra un reddito di 3000 lire ed uno di 1000 quest'indice del grado di disuguaglianza risulta uguale a 0,5, mentre l'altro indice precedentemente definito risulta uguale a 0,666.... Le definizioni date mostrano il diverso significato dei due indici.

4. — L'errore di una differenza è uguale alla differenza degli errori dei termini, tenuto conto del segno. Perciò anche termini fortemente errati potranno dare una differenza lievemente errata purchè gli errori siano dello stesso segno e press'a poco della stessa grandezza; se poi gli errori sono uguali tra loro la differenza risulterà esatta. Invece termini affetti da errori trascurabili possono dare differenze non trascurabili se gli errori sono di segno opposto.

L'errore relativo di una differenza è dato dal quoziente tra la differenza degli errori dei termini e il valore esatto della differenza.

5. — Una differenza si può rappresentare graficamente mediante un'area proporzionale al più grande dei due dati, entro la quale viene inserita un'area proporzionale al più piccolo: la differenza tra le due aree risulta proporzionale alla differenza tra i due dati. Spesso tale rappresentazione grafica pone in efficace rilievo una differenza.

Quesiti ed esercizi : 1. — Quante differenze si possono calcolare fra due dati statistici?

2. — In che differiscono l'eccedenza del dato *a* sul dato *b* e la deficienza del dato *b* sul dato *a*? In che differiscono le corrispondenti eccedenza relativa e deficienza relativa?

3. — Si eseguano confronti, a due a due, mediante differenze, tra dati statistici desunti dall'ASI, indicanti l'ammontare di una determinata produzione industriale in differenti anni, l'ammontare di una determinata produzione agraria in differenti compartimenti, il salario di due differenti categorie di lavoratori, il salario di una stessa categoria in epoche diverse.

4. — Si metta in evidenza mediante esempi come l'apprezzamento di un dato statistico possa fortemente variare col mutare del termine di confronto.

5. — Si esprimano in formole algebriche l'eccedenza relativa e la deficienza relativa.

6. — Sui dati del commercio internazionale dell'Italia forniti nell'appendice retrospettiva dell'ASI, si calcolino per singole merci le eccedenze o le deficienze relative delle importazioni [esportazioni] degli anni 1922 e 1928 su quelle del 1913.

7. — Si esprimano in formole algebriche i due indici del grado di disuguaglianza indicati nel testo. Si cerchi di chiarire la differenza del fondamento logico dell'uno e dell'altro indice.

8. — Una persona ha un reddito di 10.000 lire, un'altra ha un reddito di 5.000 lire. Si calcoli l'indice del grado di disuguaglianza fra i due redditi col due metodi indicati nel testo, cercando di rendersi conto delle ragioni di preferenza per l'uno e per l'altro metodo.

9. — In un paese nel 1880 sopravvivevano a 1 anno d'età 70 su ogni 100 nati; nel 1920 ne sopravvivono 90. Si calcoli l'eccedenza e l'eccedenza relativa dei sopravviventi a un anno nel 1920 in confronto ai sopravviventi a un anno nel 1880. Si calcoli l'indice del grado di eccedenza; e se ne chiarisca il significato.

10. — Si esprima in formola algebrica l'indice del grado di eccedenza. Si esprima in formola algebrica l'indice del grado di deficienza.

11. — Ai prezzi correnti, occorrono almeno lire 1,20 per acquistare alimenti aventi come equivalente energetico tremila calorie. In un bilancio familiare si spendono effettivamente lire 5,40 per ogni tremila calorie, in un altro lire 3,20. Si calcoli la deficienza e la deficienza relativa della seconda spesa rispetto alla prima. Si calcoli l'indice del grado di deficienza e se ne chiarisca il significato.

12. — Si rappresenti algebricamente l'errore di una differenza in funzione degli errori dei due termini.

13. — Ammesso che il valore delle importazioni indicato dalle statistiche di un dato paese sia errato in eccesso di 1 miliardo di lire, e il valore delle esportazioni sia errato in difetto di 2 miliardi di lire, in che senso e di quanto risulterà errata l'eccedenza delle importazioni sulle esportazioni?

14. — Confrontando mediante differenze i valori rilevati e i valori corretti (in lire carta) dall'Istituto centrale di statistica, delle esportazioni italiane degli anni 1920 e 1926 (vedasi ASI, capitolo «Commercio con l'estero»), si metta in

evidenza l'influsso degli errori dei dati originari sui risultati della comparazione. Analogo confronto si esegua per le importazioni dei due anni indicati.

15. — Supposto (avvertasi che si tratta di ipotesi arbitrarie, fatte al solo scopo dell'esercitazione) che il dato ufficiale sul raccolto italiano del frumento nel 1917 sia errato in difetto del 10°/₀ e quello del 1925 sia errato in eccesso del 5°/₀, si calcoli il vero ammontare del raccolto in ciascuno dei due anni; poi si calcoli l'eccedenza apparente e l'eccedenza reale del secondo raccolto sul primo; e infine si calcoli l'errore relativo dell'eccedenza apparente rispetto all'eccedenza reale.

16. — Si rappresentino graficamente mediante aree alcune delle differenze precedentemente calcolate. Accanto alla scala delle grandezze rappresentate, si pongano altre due scale che permettano di leggere immediatamente il valore della eccedenza relativa e quello della deficienza relativa.

17. — Si provi a rappresentare una differenza mediante un unico segmento, in parte a tratto continuo, in parte a punti o a trattini.

18. — Si confrontino diverse forme di rappresentazione grafica di una stessa differenza (segmenti, aree rettangolari, aree circolari), ricercando la forma più efficace.

CAPITOLO VII.

La comparazione fra due dati statistici mediante rapporto.

I rapporti: definizioni — Errore di un rapporto — Rappresentazione grafica — Classificazione dei rapporti — Rapporti fra dati omogenei: rapporti di composizione, rapporti di incremento, rapporti indici — Rapporti fra dati eterogenei: rapporti di intensità, rapporti di estensione, rapporti di coordinamento.

1. — I *rapporti* servono per la comparazione così fra due dati omogenei come fra due dati eterogenei. Al primo intento servono in concorrenza con le differenze, le quali nella pratica quotidiana sono usate più spesso, sia perchè riesce più facile calcolarle mentalmente, sia perchè a molti fini interessa conoscere piuttosto il divario assoluto che la relazione di grandezza fra i dati che si confrontano. Se sappiamo che quest'anno si sono raccolti 60 milioni di quintali di grano in confronto ai 70 milioni dell'anno scorso, riusciamo subito ad enunciare la deficienza di 10 milioni di quintali del presente raccolto in confronto a quello passato, mentre dobbiamo metterci a tavolino per calcolare che l'uno corrisponde a 0,86 dell'altro; inoltre la deficienza ci indica subito la quantità di grano che si dovrà importare se non si vuole restringere il consumo. La differenza permette di confrontare tra loro dati

aventi segno diverso, mentre il rapporto può servire a tale intento solo in certi casi e con l'accompagnamento di spiegazioni verbali. Se il bilancio di un esercizio si chiude con un avanzo di 100 mila lire, mentre quello dell'esercizio precedente si era chiuso con un disavanzo di 50 mila, l'eccedenza del primo dato sul secondo misura immediatamente il miglioramento di 150 mila lire conseguito nell'esercizio attuale in confronto al precedente; ma il rapporto fra i due dati, il cui valore è uguale a -2, risulta a prima vista oscuro, benchè il suo significato aritmetico sia chiaro: esso indica infatti che il saldo dell'ultimo bilancio risulta doppio in valore assoluto, ma di segno opposto, di quello del bilancio scorso.

Quando poi si vogliano comparare tra loro dati eterogenei, per esempio il grano mietuto con la superficie coltivata a grano, bisogna ricorrere necessariamente al rapporto. Il quale serve anche per la comparazione fra un dato statistico e un'altra grandezza che non sia un dato statistico.

Il rapporto è una comparazione fra due grandezze, delle quali l'una (*denominatore*) viene presa come unità per la misura dell'altra (*numeratore*). Si assume come numeratore il dato del quale si vuol compiere l'apprezzamento in relazione all'altro: per esempio, il raccolto di quest'anno, di cui vogliamo giudicare la grandezza in relazione a quello dell'anno scorso. In certi casi si può rendere più efficace la comparazione eseguendola così nell'uno come nell'altro senso: rapporto tra popolazione e superficie di un paese, che ci dà il numero medio di abitanti corrispondente a ciascun chilometro quadrato; e rapporto tra superficie e popolazione, che ci dà la frazione media di chilometro quadrato corrispondente a ciascun abitante.

Avvertasi che quando si assume a denominatore il più grande dei due numeri il quoziente del rapporto non può superare l'unità, mentre quando si assume il più piccolo il quoziente non incontra alcun limite.

Il *quoziente* o valore del rapporto, che si ottiene mercè la divisione del numeratore per il denominatore, indica quante unità del primo corrispondano a ciascuna unità del secondo. Può essere espresso, con approssimazione grande quanto si vuole, nella forma di numero decimale, oppure in quella di frazione decimale, oppure ancora in quella di proporzione a 10, a 100, a 1000, ecc. Per esempio si può dire che il raccolto di quest'anno corrisponde a 0,86 di quello dell'anno scorso, o che corrisponde a 86 centesimi di esso, o che corrisponde all'86 per 100 (°/₀), o all'860 per 1000 di esso. In forme differenti si dice la stessa cosa. Secondo la natura dei numeri che si confrontano, secondo lo scopo del confronto, secondo il pubblico al quale esso è destinato, può convenire in pratica esprimere il quoziente del rapporto nell'una o nell'altra di queste forme. In certi casi il quoziente di un rapporto viene denominato *quoziente*, senz'altro, oppure *saggio*, *tasso*, *coefficiente*, *indice* (quoziente di mortalità, saggio o tasso di invalidità, coefficiente di variabilità, indice del prezzo del pane).

È indispensabile, per la retta interpretazione dei rapporti statistici, tener presente che il quoziente di un rapporto non vuole sempre enunciare il risultato di un'operazione concretamente eseguibile sulle unità rappresentate dal numeratore col distribuirle in tanti gruppi, tutti uguali fra loro, quante sono le unità del denominatore; anzi in generale vuole semplicemente indicare il risultato d'una operazione eseguita su simboli numerici, che può non trovare riscontro in alcuna operazione eseguibile sulle quantità reali.

È ovvia l'osservazione che il quoziente di un rapporto è uguale alla differenza relativa tra il numeratore e il denominatore, aumentata dell'unità: per esempio se il raccolto di quest'anno (60 milioni di quintali) presenta una deficienza relativa di — 0,14 rispetto a quello dell'anno scorso (70 milioni), il rapporto fra il primo ed il secondo raccolto è uguale a (— 0,14 + 1), cioè a 0,86. È dunque ovvio che dall'eccedenza o dalla deficienza relativa si può ricavare il rapporto, e viceversa.

2. — L'errore di un rapporto, cioè la differenza fra il valore noto di esso ed il valore esatto, dipende, com'è ovvio, dagli errori dei termini. Esaminando il modo in cui i due termini concorrono a formare il rapporto, si trova facilmente che l'errore relativo del rapporto noto è uguale al quoziente della differenza tra l'errore relativo del numeratore e quello del denominatore, per l'unità aumentata dell'errore relativo del denominatore. Per esempio l'errore relativo di un rapporto fra un dato errato del 20 °/₀ in eccesso ed uno errato del 20 °/₀ in difetto risulta del 50 °/₀ in eccesso. L'errore del rapporto noto è uguale al prodotto dell'errore relativo dianzi definito per il valore del rapporto esatto.

Poichè l'errore relativo del rapporto si annulla quando l'errore relativo del numeratore è uguale a quello del denominatore, si vede come anche dati gravemente errati possano fornire il valore esatto del rapporto, purchè errati nello stesso senso e nella stessa proporzione. D'altra parte, la formola dell'errore relativo di un rapporto mostra come errori non gravi del numeratore e del denominatore possano dar luogo ad errori gravi nella formazione del rapporto, quando siano di senso opposto, come nell'esempio precedente.

3. — Un rapporto si può rappresentare graficamente in modo semplice e completo mediante un rettangolo del quale una dimensione sia proporzionale al denominatore del rapporto, l'altra al quoziente; l'area del rettangolo risulta, per conseguenza, proporzionale al numeratore del rapporto, del quale vengono così simultaneamente rappresentati numeratore, denominatore e quoziente.

Si può in altro modo rappresentare graficamente il rapporto costruendo un angolo che abbia il seno proporzionale al numeratore e il coseno proporzionale al denominatore; la tangente trigonometrica dell'angolo risulta proporzionale al quoziente del rapporto.

4. — Per meglio vedere come i rapporti vengano impiegati nella descrizione dei fenomeni collettivamente tipici, conviene classificarli.

Una prima distinzione in due grandi categorie si può fare, secondo la natura dei termini del rapporto, separando i rapporti fra dati omogenei dai rapporti fra dati eterogenei (in questa seconda categoria comprendiamo anche rapporti nei quali uno dei termini non è un dato statistico).

I rapporti fra dati omogenei si possono suddividere, secondo il fine per cui vengono istituiti, in tre classi principali: quella dei *rapporti di composizione*, fra dati dei quali l'uno costituisce una parte dell'altro; quella dei *rapporti di incremento*, fra dati dei quali l'uno rappresenta l'incremento dell'altro; quella dei *rapporti indici*, fra dati omogenei che non stanno nella relazione di parte a tutto, nè in quella di incremento a consistenza iniziale.

I rapporti fra dati eterogenei si possono suddividere, secondo il fine per cui vengono istituiti, in tre classi principali: quella dei *rapporti d'intensità*, quella dei *rapporti di estensione* e quella dei *rapporti di coordinamento*.

5. — Cominciamo col prendere in esame le principali classi di rapporti fra dati omogenei. I *rapporti di composizione* indicano quale frazione di un certo complesso, rappresentato dal denominatore, sia costituita da una determinata parte di esso, rappresentata dal numeratore: servono, pertanto, per l'apprezzamento dell'importanza di

una parte in confronto al tutto, e — adoperati in serie — per l'apprezzamento dell'importanza comparativa delle varie parti di uno stesso tutto.

Come risulta dalla sua definizione, il valore del rapporto di composizione è in generale inferiore all'unità, può divenire uguale ad essa, ma non può mai superare l'unità. E quindi se il valore stesso viene messo in forma di proporzione a 100 non può mai superare 100; espresso in questa forma si suole denominare *proporzione percentuale*, o semplicemente *percentuale*. Tale denominazione, che in questo caso è corretta, spesso viene applicata a sproposito ad altri rapporti anche se espressi in forma di numeri o frazioni decimali, o in forma di proporzioni a potenze di 10 diverse da 100: errore che bisogna assolutamente evitare.

Affinchè il rapporto di composizione corrisponda bene al suo fine, bisogna che il numeratore indichi in modo completo la consistenza di quella parte che si vuol confrontare col tutto; altrimenti risulterà inferiore al vero. Suppongasi che in un censimento sia stata accertata, fra gli altri caratteri, l'età di ciascun abitante di un paese, e che per un certo numero di persone non sia stato possibile conoscere l'età. Ne risulterà un insieme di dati, che indicheranno quante persone siano comprese nei singoli gruppi d'età; un ultimo dato indicherà il numero delle persone di età ignota. Evidentemente queste si ripartiscono, nella realtà, fra i vari gruppi d'età: così che i dati indicanti la consistenza accertata di questi sono in generale inferiori al vero; e inferiori al vero risultano i corrispondenti rapporti di composizione. Si potrà mettere in rapporto il numero degli individui di ciascun gruppo con la popolazione di età nota (cioè con la popolazione totale meno il gruppo degli individui di età ignota), invece che con la popolazione totale. Ma anche con questo sistema non sempre si ottengono risultati soddisfacenti.

Quando così il numeratore come il denominatore del rapporto di composizione sono costituiti da casi individualmente distinti, il rapporto viene talora denominato *frequenza relativa*.

Invertendo il rapporto di composizione, cioè assumendo il tutto come numeratore e la parte come denominatore, si ottiene un altro rapporto che esprime la medesima relazione di grandezza e che talvolta può essere preferito al rapporto di composizione per ragioni di opportunità. Ma nella maggior parte dei casi si suole preferire il rapporto di composizione.

6. — I *rapporti d'incremento* indicano qual frazione costituisca un incremento, positivo o negativo, di una certa grandezza, rispetto alla grandezza medesima, considerata nella sua misura iniziale. Recano, dunque, al numeratore il dato che indica l'incremento; al denominatore quello che indica la consistenza iniziale. Se, rispetto a questa, avvengono da una parte aumenti e dall'altra diminuzioni, si potranno istituire tre rapporti, aventi come denominatore comune la consistenza iniziale e come numeratori rispettivamente: l'aumento, la diminuzione, l'incremento netto (aumento meno diminuzione). Esempio: la popolazione all'inizio di un anno (consistenza iniziale), la somma dei nati e degli immigrati (aumento), la somma dei morti e degli emigrati (diminuzione), la differenza tra queste due somme (incremento netto).

7. — I *rapporti indici* confrontano un dato statistico con un altro dato omogeneo col primo: desunto però da osservazioni compiute in altro tempo, in altro luogo, o in altra sezione dello stesso campo d'osservazione (esempio: prezzi del frumento a due date diverse, in due regioni diverse, di due qualità diverse). Sostituiscono, quindi, alla misura assoluta di un fenomeno, la misura relativa alla dimensione assunta dal fenomeno stesso in altre circostanze.

Il valore del rapporto indice espresso nella forma di proporzione a 100 o a 1000 viene denominato *numero indice*. Se si tratta di proporzioni a 100 si può correttamente parlare anche in questo caso di *proporzione percentuale*. Il denominatore del rapporto indice viene talvolta chiamato *base* del numero indice.

Vedremo in seguito come i numeri indici vengano utilmente impiegati per riassumere fenomeni complessi, dei quali possono offrire espressioni sintetiche quantitative non ottenibili per altra via.

8. — Passiamo ora all'esame delle principali classi di rapporti fra dati eterogenei. I *rapporti d'intensità* presentano una grande varietà di tipi, i quali però hanno in comune il carattere che il numeratore rappresenta la misura accertata del fenomeno che è oggetto di esame e il denominatore rappresenta la misura di una circostanza quantitativa caratteristica del campo d'osservazione (si potrebbe dire di una « dimensione » del campo d'osservazione); ovvero rappresenta la misura combinata di due o più di tali circostanze (dimensioni). Si ragguaglia la misura del fenomeno alla dimensione, od alle dimensioni, del campo di osservazione per rendere

possibile l'apprezzamento della intensità di manifestazione del fenomeno stesso.

Il fenomeno può essere di stato (numero degli invalidi esistenti in una data popolazione in un dato momento) o di movimento (numero degli individui che sono divenuti invalidi in una data popolazione in un dato intervallo di tempo). Esso può essere misurato da un certo numero di *casi accertati* (matrimoni avvenuti in un anno); da una *somma di circostanze quantitative* caratteristiche dei singoli casi accertati (giornate di malattia rilevate in un anno in una popolazione, cioè somma delle giornate di malattia corrispondenti ai singoli casi di malattia); da una *grandezza scindibile* in singoli casi, cioè in unità aventi una individualità propria, la quale però si consideri come un sol tutto (quantità di grano raccolta sopra una determinata area); da una *grandezza non scindibile* in singoli casi o in corrispondenza a singoli casi (quantità di pioggia caduta in una giornata in una data zona).

Le circostanze caratteristiche del campo d'osservazione, alle quali — considerate separatamente o congiuntamente — si ragguaglia la misura del fenomeno per apprezzarne l'intensità, possono essere costituite: dal *tempo* per il quale l'osservazione è protratta (riferimento escluso, naturalmente, per i fenomeni di stato), dallo *spazio* (lunghezza, superficie, volume) al quale l'osservazione è estesa, da un numero di *casi osservati* in relazione al quale interessi porre il fenomeno accertato o perchè deriva da essi o perchè influisce su di essi o per altre ragioni, da una *grandezza scindibile* (nel senso dianzi spiegato) o da una *grandezza non scindibile* in singoli casi ovvero in corrispondenza a singoli casi, in relazione alla quale interessi porre il fenomeno.

Ragguagliando la misura del fenomeno a quella di una dimensione del campo d'osservazione, si vuol determinare qual misura del fenomeno corrisponderebbe a ciascuna unità di dimensione del campo d'osservazione se il fenomeno fosse uniformemente distribuito in relazione alla dimensione stessa. Così dividendo il numero delle vetture tramviarie passate davanti a un dato punto d'osservazione per il numero delle ore dell'osservazione (tempo), si determina quante vetture sarebbero passate in ciascuna ora nell'ipotesi di distribuzione uniforme dei passaggi nel tempo; dividendo la popolazione italiana per la superficie del Regno (spazio), si determina quanti abitanti dimorerebbero su ciascun chilometro quadrato, nell'ipotesi di distribuzione uniforme della popolazione nello spazio; dividendo il numero dei limoni raccolti in un limoneto per il numero delle piante fruttifere (casi osservati), si determina quanti limoni darebbe ciascuna pianta, nell'ipotesi di distribuzione uniforme dei frutti sulle piante; dividendo la quantità di acqua defluita da un bacino imbrifero attraverso un dato fiume per la quantità di acqua raccolta dal bacino stesso mercè le precipitazioni atmosferiche (grandezza non scindibile, nel senso sopra spiegato), si determina qual frazione di ciascun metro cubo d'acqua caduta nel bacino sarebbe defluita attraverso quel fiume nell'ipotesi che questo avesse attinto uniformemente all'acqua caduta in ogni parte del bacino. Il riferimento della misura del fenomeno ad una dimensione del campo d'osservazione è logicamente ammissibile quando il campo d'osservazione sia effettivamente contrassegnato da tale dimensione: nel primo dei precedenti esempi non avrebbe significato alcuno il riferimento allo spazio, perchè l'osservazione è riferita ad un punto dello spazio; nel secondo esempio non avrebbe significato il riferimento al tempo, perchè l'osservazione è riferita ad un istante del tempo. Nè tutti i riferimenti logicamente ammissibili sono utili: nulla vieta di mettere i morti per malaria in una provincia in rapporto con la superficie della provincia, o della parte malarica di essa; ma si preferirà metterli in rapporto con la popolazione della provincia o della zona.

Il riferimento della misura del fenomeno a due o più dimensioni del campo d'osservazione congiuntamente considerate è logicamente ammissibile quando sia logicamente ammissibile la ideale successiva e coordinata suddivisione del campo d'osservazione secondo due o più dimensioni. Non sarebbe possibile, ad esempio, riferire il numero dei morti per malaria simultaneamente alla superficie ed alla popolazione della zona malarica, perchè il campo d'osservazione può suddividersi bensì secondo la popolazione in tante sezioni uguali, o secondo la superficie in tante sezioni uguali, ma l'una suddivisione esclude l'altra. Sarebbe invece possibile riferire il numero dei morti per malaria simultaneamente al numero dei giorni di osservazione e alla superficie della zona malarica, oppure al numero dei giorni di osservazione e al numero degli abitanti, perchè l'una suddivisione non esclude l'altra. E così potremo riferire il numero degli infortuni accaduti a viaggiatori ferroviari simultaneamente al numero dei viaggiatori trasportati e al tempo (viaggiatori-ora), ovvero al numero dei viaggiatori trasportati e allo spazio (viaggiatori-chilometro), ovvero allo spazio ed al tempo (infortuni per giorno e per chilometro di linea), ma non potremo riferirlo simultaneamente al tempo, allo spazio e al numero dei viaggiatori non essendo logicamente ammissibile la suddivisione successiva dei singoli viaggi secondo lo spazio e secondo il tempo. Anche per i riferimenti composti va ripetuta l'avvertenza espressa per quelli semplici: che non sempre il riferimento possibile è utile. Tra i riferimenti composti più spesso possibili ed utili, ricordiamo quelli che consentono l'apprezzamento dell'intensità del fenomeno simultaneamente in relazione allo spazio e al numero dei casi osservati (intensità per unità di spazio e per caso osservato); al tempo e al numero dei casi osservati (intensità per unità di tempo e per caso osservato); allo spazio ed al tempo (intensità per unità di spazio e per unità di tempo); allo spazio, al tempo e al numero dei casi osservati (intensità per unità di spazio, per unità di tempo e per caso osservato).

Associando i vari tipi di numeratore con i vari tipi di denominatore possibili, in quanto siano logicamente associabili fra loro, si ottengono tanti tipi di rapporti d'intensità, che non solo esistono in teoria, ma trovano anche applicazione in pratica. Senza enumerare tutte le possibili combinazioni, ci fermeremo su alcuni tipi più spesso usati.

a) *Rapporti di frequenza.* — S'impiegano per la descrizione di quei fenomeni di movimento la cui manifestazione è costituita da casi individualmente distinti l'uno dall'altro e reciprocamente equivalenti nel senso che, pur differendo per circostanze particolari, ciascuno di essi rappresenta un caso accertato del fenomeno in esame (un matrimonio, una condanna penale, un infortunio). L'intensità di fenomeni siffatti è misurata dal rapporto tra il numero dei casi accertati e una dimensione del campo d'osservazione (spazio, tempo, numero dei casi osservati); talvolta giovano i riferimenti composti dianzi indicati. Non sempre sono utili nè sempre logicamente ammissibili *tutti* i possibili riferimenti semplici e composti: per vedere come talora possano avere tutti un significato ed un'utilità si consideri l'esempio di un bombardamento, costituito da C colpi sparati (casi accertati) da P pezzi d'artiglieria (casi osservati), in un tempo T, distribuiti sopra un'estensione di spazio S. In questo caso ciascuno dei possibili riferimenti semplici o composti ha un proprio significato, differente da quello di ciascun altro.

Rapporti di frequenza s'impiegano anche per descrivere fenomeni di stato, la cui manifestazione è costituita da casi nettamente distinti ed equivalenti. Naturalmente vien meno la possibilità di riferimento al tempo: restano le possibilità di riferimento allo spazio e al numero dei casi osservati (esempio: autoveicoli censiti in una provincia in rapporto alla superficie della provincia stessa, o alla lunghezza delle strade, o al numero degli abitanti).

b) *Probabilità statistiche.* — Per definire questi rapporti, partiamo — per comodità didattica — dai rapporti di frequenza per fenomeni di movimento. In questi rapporti il numeratore indica sempre il numero dei casi accertati di un fenomeno; il denominatore, coll'indicare il numero dei casi osservati, indica talvolta anche il massimo numero possibile dei casi del fenomeno, ossia il massimo valore possibile del numeratore. Quindi il rapporto indica quale frazione dei casi possibili sia costituita dai casi effettivamente accertati del fenomeno. Il rapporto di frequenza assume tale carattere quando il fenomeno non può ripetersi in corrispondenza al singolo caso osservato. Se osserviamo, per esempio, un gruppo di assicurati sulla vita, il loro numero rappresenta il massimo numero di casi di morte che potremo accertare nel gruppo stesso; se osserviamo un gruppo di partenti per una corsa, il loro numero rappresenta il massimo numero di casi di arrivo al traguardo che potremo accertare nel gruppo stesso. Nelle ipotesi sopra indicate, il rapporto di frequenza — il cui valore non può mai superare l'unità — viene denominato *probabilità statistica* (probabilità di morte, probabilità di arrivo al traguardo, ecc.).

La probabilità statistica non è che una frequenza, una descrizione di fenomeni avvenuti in passato; non va confusa con la probabilità matematica, la quale è il rapporto fra casi che *a priori* risultano favorevoli a un dato evento e casi che *a priori* risultano ugualmente possibili; nè va confusa con la probabilità del linguaggio comune, la quale costituisce una previsione empirica dell'avvenire; ad evitare le confusioni, converrebbe anzi lasciar da parte l'equivoca denominazione di « probabilità » per i rapporti di frequenza. Ed è pure soltanto nel significato della descrizione, non in quello della previsione, che rapporti di tal sorta si assumono a misura del *rischio* di manifestazione di un certo evento, in quanto indicano in qual frazione dei casi possibili l'evento si è avverato.

Quando il fenomeno può manifestarsi una sola volta, esclusivamente in corrispondenza ad alcuni dei casi osservati, mentre non può manifestarsi in corrispondenza ad altri (esempio: casi accertati: persone che contraggono matrimonio in un mese; casi osservati: popolazione totale, che comprende molte persone non atte al matrimonio), il rapporto di frequenza non ha il carattere di probabilità statistica.

Quando il fenomeno può manifestarsi più volte in corrispondenza a ciascun caso osservato (esempio: casi accertati: casi di malattia in un dato intervallo di tempo; casi osservati: popolazione totale), il rapporto non ha il carattere di probabilità statistica, a meno che ripetuti casi del fenomeno accertati in corrispondenza ad uno stesso caso osservato vengano contati per uno solo (esempio: casi accertati: persone che sono state ammalate una o più volte in un dato intervallo di tempo; casi osservati: popolazione totale).

La denominazione di *probabilità statistica* viene applicata anche a rapporti destinati alla descrizione di fenomeni di stato. Mediante tali rapporti il numero dei casi osservati contrassegnati da una data circostanza viene comparato col numero totale dei casi osservati (esempio: i ciechi con l'intera popolazione). Si tratta di rapporti fra una parte e il tutto: che quindi sono rapporti di composizione; che tuttavia per il fine al quale vengono impiegati hanno il carattere di rapporti di frequenza.

c) *Rapporti di densità.* — Costituiscono un tipo parallelo a quello dei rapporti di frequenza. S'impiegano per la descrizione di quei fenomeni di movimento la cui manifestazione è misurata da circostanze quantitative caratteristiche dei singoli casi accertati. L'intensità del fenomeno è qui misurata dal rapporto fra la somma delle circostanze quantitative caratteristiche dei casi accertati e una dimensione del campo d'osservazione (spazio, tempo, numero dei casi osservati); possono essere utili i riferimenti composti precedentemente indicati. Per vedere la possibilità dei riferimenti semplici e composti, si riprenda l'esempio del bombardamento, addotto a proposito dei rapporti di frequenza; ma, invece di assumere a misura dell'intensità del bombardamento il numero C dei colpi sparati, si assuma la carica complessiva di esplosivo E contenuta nei proietti lanciati (ogni proietto è caratterizzato da una circostanza quantitativa, la carica di esplosivo; E rappresenta la somma delle circostanze quantitative caratteristiche dei singoli casi accertati, che sono i colpi sparati, cioè i proietti lanciati).

Anche i rapporti di densità s'impiegano per descrivere fenomeni di stato, con riferimento allo spazio o al numero dei casi osservati (esempio: peso o volume del legname ottenibile da una data zona forestale in rapporto all'area di essa, o in rapporto al numero delle piante).

d) *Rapporti di diffusione.* — S'impiegano per la descrizione di quei fenomeni di movimento la cui manifestazione è indicata da una grandezza non scindibile se non in modo assolutamente arbitrario. L'intensità del fenomeno è misurata dal rapporto fra il dato che indica il fenomeno e una dimensione del campo d'osservazione (spazio, tempo, numero dei casi osservati). La quantità di pioggia caduta sopra un'area coltivata a frutteto può essere messa in rapporto con la superficie del frutteto, con la durata dell'osservazione, col numero delle piante costituenti il frutteto; potrà essere utile anche qualche riferimento composto (spazio e tempo, tempo e numero delle piante).

Anche i rapporti di diffusione servono per descrivere fenomeni di stato, in relazione allo spazio o al numero dei casi osservati (esempio: volume delle acque contenute in un serbatoio, in rapporto all'area da irrigare mediante le acque stesse, o in rapporto al numero delle piante esistenti sull'area stessa).

9. — I *rapporti di estensione* si definiscono facilmente come i reciproci dei rapporti d'intensità. Questi mostrano quale misura del fenomeno corrisponderebbe all'unità di dimensione del campo d'osservazione se il fenomeno fosse uniformemente distribuito nel campo stesso; invece i rapporti di estensione mostrano quale misura dimensionale del campo d'osservazione corrisponderebbe, nella stessa ipotesi, all'unità di misura del fenomeno.

Ragioni di opportunità inducono a preferire ora il rapporto d'intensità ora il rapporto di estensione, per la descrizione di un fenomeno. In generale si preferisce il primo, perchè varia in ragione diretta della misura del fenomeno, mentre il secondo varia in ragione inversa. Se si sente dire che la provincia A possiede 5 automobili per 100 abitanti mentre la provincia B ne possiede 10, si intende subito che la dotazione relativa di automobili è doppia nella provincia B; se invece si sente dire che nella provincia A si hanno 20 abitanti per un'automobile mentre nella provincia B se ne hanno 10, occorre una breve riflessione per intendere che il numero minore corrisponde ad una frequenza maggiore di automobili.

10. — *Rapporti di coordinamento.* — Servono per apprezzare la misura comparativa di due fenomeni differenti. Numeratore è la

misura del fenomeno che si vuol apprezzare; denominatore è la misura del fenomeno che serve di riferimento. Riferendo, ad esempio, il numero dei nati in un anno al numero dei morti nello stesso anno, si perde di vista la grandezza assoluta del numero dei nati, non si può dir neppure che si misuri l'intensità del rinnovamento della popolazione, ma si vede subito di qual frazione le nascite eccedano le morti o restino ad esse inferiori.

11. — Chi rifletta sulla nostra classificazione dei rapporti statistici noterà come essa sia utile a porre in rilievo le molteplici possibilità di applicazione del metodo; ma per poco che sia dotato di spirito critico noterà anche la possibilità che spesso si presenta di catalogare un rapporto nell'una o nell'altra classe. Il numero dei corridori che non giungono alla meta rappresenta una parte del numero dei corridori che hanno iniziato il percorso, e quindi il rapporto tra i due è un rapporto di composizione; ma il numeratore rappresenta anche il decremento subito dal gruppo dei corridori durante il percorso, e quindi il rapporto è anche un rapporto di incremento; e il numeratore rappresenta anche il numero degli eliminati durante la corsa, onde il rapporto misura la frequenza delle eliminazioni. In questo, come in altri esempi, è l'aspetto dal quale si considera il rapporto, il fine cui esso serve, che detta norma per la classificazione. La mancanza di una netta separazione tra una classe e l'altra, le interferenze tra le varie classi sono inconvenienti inevitabili di ogni classificazione dei rapporti statistici che non voglia mantenersi tanto sommaria da divenire inutile.

12. — Dopo avere studiato i due metodi di comparazione fra coppie di dati: differenza e rapporto, siamo in grado di vedere i vantaggi e gli svantaggi di ciascuno di essi. La differenza mantiene in evidenza la grandezza assoluta della disuguaglianza fra i due dati; perciò va preferita quando importi appunto far risaltare una variazione assoluta: così, ad esempio, quando si confronti l'ammontare di un raccolto o di una produzione industriale in un anno, col corrispondente ammontare dell'anno precedente. Il rapporto mette in evidenza la relazione di grandezza fra i due dati, e quindi è preferibile quando si voglia far risaltare il divario relativo che intercede fra essi. Ma la differenza non dà alcuna idea del divario relativo fra i dati, il rapporto non dà alcuna idea del divario assoluto. La differenza relativa partecipa dei pregi e dei difetti del rapporto, al quale è legata, come sappiamo, da una semplicissima relazione.

Quesiti ed esercizi : 1. — Che cos'è un rapporto? Quale dei due termini del rapporto si assume come riferimento per l'apprezzamento dell'altro?

2. — Si cerchino nell'ASI esempi di comparazioni tra due dati statistici che possano eseguirsi così mediante differenze come mediante rapporti. Si mettano in evidenza, in ogni singolo esempio, i vantaggi e gli svantaggi di ciascuno dei due metodi.

3. — Si cerchino nell'ASI esempi di rapporti che potrebbero utilmente essere invertiti ed altri esempi di rapporti che invertiti diverrebbero meno facilmente comprensibili o disadatti al fine cui debbono servire.

4. — Che cos'è il quoziente d'un rapporto? Con quali nomi si trova talora indicato? In quali modi può essere espresso?

5. — Sui dati del capitolo « Viabilità » dell'ASI si calcolino rapporti il cui valore sia efficacemente espresso nella forma di numero decimale; nella forma di frazione decimale; nella forma di proporzione a 100 o ad altra potenza di 10.

6. — Si traduca in formola algebrica la relazione che esiste tra il quoziente di un rapporto e la differenza relativa fra il numeratore e il denominatore del rapporto stesso.

7. — Si ricerchino nell'ASI esempi nei quali il valore di un rapporto indichi il risultato di un'operazione concretamente eseguibile ed esempi nei quali il valore di un rapporto indichi il risultato di una operazione concretamente ineseguibile, e quindi puramente simbolica ed astratta.

8. — Dai rapporti tra i prezzi dei generi di consumo popolare nel dopoguerra e i prezzi stessi nell'anteguerra, presentati nel capitolo « Prezzi e consumi » dell'ASI si desumano le eccedenze relative dei prezzi postbellici su quelli prebellici.

9. — Come si definisce l'errore di un rapporto? l'errore relativo?

10. — Ammesse come valori esatti del commercio italiano con l'estero le « cifre corrette » *in lire carta* indicate nell'ASI, si esamini l'influenza degli errori che viziano le « cifre rilevate », mettendo in rapporto: l'esportazione con l'importazione del 1920; l'esportazione del 1926 con quella del 1920; l'esportazione del 1920 con quella del 1926; l'esportazione del 1926 con quella del 1919. In ciascuno di questi esempi si calcolino: gli errori e gli errori relativi dei dati rilevati rispetto ai dati esatti; gli errori e gli errori relativi dei rapporti fra i dati rilevati rispetto ai corrispondenti rapporti fra i dati esatti. Si ricerchi quale avrebbe dovuto essere il valore rilevato delle esportazioni del 1920 affinchè, essendo 26.822 milioni di lire il valore rilevato delle importazioni, 18.105 il valore esatto delle importazioni, 8.742 il valore esatto delle esportazioni, risultasse scevro da errore il rapporto fra i dati rilevati.

11. — Ammesse ora come valori esatti del commercio italiano le « cifre corrette » *in lire oro* indicate nell'ASI e come valori errati le « cifre corrette » in lire carta, si mostri come la svalutazione e la rivalutazione della lira influiscano sui risultati del confronto tra le importazioni (o le esportazioni) dei singoli anni bellici e postbellici e quelle del 1913.

12. — Si esprima in una formola l'errore relativo di un rapporto in funzione degli errori relativi del numeratore e del denominatore. Servendosi della formola, si ricerchi in quali casi l'errore del rapporto risulta maggiore

di quelli: *a*) del numeratore, *b*) del denominatore, *c*) di entrambi i termini; in quali casi risulta minore di essi.

13. — Per effetto di erronee dichiarazioni di età il numero dei censiti di 70-71 anni risulta in un paese di 357.451 invece che di 201.709 qual è realmente al 30 giugno, data del censimento; il numero dei morti in età di 70-71 anni nel paese stesso e nello stesso anno risulta di 18.138 invece che di 12.043 qual è realmente. Si determinino gli errori relativi del numeratore e del denominatore del saggio di mortalità calcolato sui dati errati, e l'errore relativo del saggio stesso.

14. — Si rappresentino graficamente, coi due metodi indicati nel testo, i rapporti tra la produzione del frumento e la superficie coltivata a frumento in Italia in ciascuno degli ultimi cinque anni. Si eseguano analoghe rappresentazioni per ciascuna regione italiana (compartimento) in uno stesso anno.

15. — Si rappresenti graficamente, col secondo dei procedimenti indicati nel testo, il rapporto fra il numero dei nati e il numero dei morti in Italia in ciascun anno dal 1881 in poi. Unendo, in ordine cronologico, con una linea punteggiata, i vertici dei vari segmenti che rappresentano il valore del rapporto nei successivi anni, si tracci la storia grafica del rapporto stesso.

16. — Quali sono le principali classi di rapporti fra dati omogenei? fra dati eterogenei? Si enumerino e si definiscano. Che cos'è una « percentuale »? Si può dire che la percentuale degli abitanti per kmq. è di 130 in Italia? che la percentuale dei nati maschi è di 1055 per 1000 femmine? che la percentuale annua dei morti è di 160 per 10.000 abitanti?

17. — Si cerchino nell'ASI esempi di rapporti fra dati statistici omogenei; fra dati statistici eterogenei; fra dati statistici ed altre grandezze.

18. — Si calcolino rapporti di composizione dai dati sulla distribuzione per compartimenti dei depositi a risparmio: *a*) nelle casse ordinarie; *b*) nelle casse postali. Si utilizzino i rapporti calcolati per mettere in evidenza la differente distribuzione regionale delle due categorie di risparmi. Sono adatti questi rapporti a misurare la tendenza al risparmio delle popolazioni delle singole regioni? Quali altri rapporti si potrebbero adoperare all'uopo?

19. — Come influisce la presenza di circa 200 mila censiti di età ignota sull'attendibilità dei rapporti di composizione della popolazione secondo l'età, che l'ASI desume dai risultati del censimento del 1921? Per quali gruppi d'età tale influenza si può ritenere trascurabile, per quali si può presumere non trascurabile? Sarebbe stato preferibile calcolare i rapporti stessi sulla popolazione di età nota invece che sulla popolazione totale?

20. — Si calcolino rapporti di composizione dai dati sulla distribuzione per compartimenti delle istanze di separazione personale di coniugi accolte. Sono questi rapporti adatti a misurare la tendenza alla separazione nei singoli compartimenti? A quali altri rapporti si potrebbe meglio ricorrere per tale intento?

21. — Si calcolino rapporti d'incremento mettendo in relazione i dati sulle nascite, sulle morti, sull'eccedenza delle nascite sopra le morti, sull'incremento effettivo della popolazione, in ciascun anno dal 1881 in poi, col dato sulla popolazione italiana all'inizio dell'anno.

22. — Si calcolino rapporti d'incremento di singole produzioni industriali (acciaio, mercurio, zucchero, seta greggia) da anno ad anno nell'ultimo decennio.

23. — Qual differenza intercede fra i rapporti di composizione e i rapporti indici?

24. — Che cos'è un numero indice? Che cos'è la base di un numero indice?

25. — Si calcolino rapporti indici e numeri indici: dello sviluppo della rete ferroviaria italiana dal 1881 in poi; della produzione per ettaro del frumento nelle varie regioni italiane in un determinato anno; dei salari medi nelle varie industrie ad una medesima data. Si adottino successivamente, in ciascuno di questi esempi, riferimenti diversi, e si mettano in risalto le differenze fra i risultati dei vari calcoli.

26. — Si cerchino nell'ASI esempi nei quali sia stato eseguito il calcolo di rapporti indici e di numeri indici; ed altri esempi nei quali sarebbe stato utile eseguire un simile calcolo.

27. — Si cerchino nell'ASI esempi di rapporti di intensità corrispondenti alle varie ipotesi sulla natura del numeratore e sulla natura del denominatore indicate nella prima parte del paragrafo 8 del testo. Si metta in rilievo con esempi pratici il diverso significato di diverse misure relative allo stesso fenomeno e l'opportunità di adottare diversi riferimenti per una stessa misura del fenomeno.

28. — Si cerchino, nella sezione del capitolo « Territorio e popolazione » dell'ASI relativa al movimento della popolazione, esempi di rapporti di frequenza e si indichino altri rapporti di frequenza che potrebbero essere calcolati mercè dati contenuti nella sezione stessa.

29. — Si calcolino rapporti di frequenza dai dati sulla istruzione elementare contenuti nell'ASI.

30. — Si indichino i rapporti di frequenza già calcolati e quelli che potrebbero calcolarsi dai dati sulle ferrovie contenuti nell'ASI.

31. — In una gara ciclistica di resistenza, 31 corridori soffrono infortuni. La gara si è svolta per dieci giorni, sopra un percorso di 3000 chilometri. La somma dei tempi impiegati dai partecipanti alla gara è di 56.871 ore; la somma dei percorsi da loro compiuti è di 143.247 chilometri. Si calcolino i vari rapporti di frequenza che sembrino utili ad indicare l'intensità con la quale si sono manifestati infortuni durante la gara, e si chiarisca il significato di ciascuno.

32. — In una grande città gli autoveicoli hanno dato luogo nel corso di un anno a 7.251 infortuni stradali, con 1.201 vittime (morti o feriti) tra persone che si trovavano sugli autoveicoli e 9.783 vittime tra altre persone. Quali riferimenti sarebbero utili per valutare l'intensità di manifestazione degli infortuni nella città? Quale rapporto sarebbe adatto a misurare il rischio di esser vittima d'infortunio automobilistico per ciascun abitante della città? Quale rapporto sarebbe adatto a misurare il rischio di cagionare un infortunio per ciascun autoveicolo circolante nella città? Quali difficoltà si incontrerebbero volendo istituire praticamente tali rapporti?

33. — In uno stabilimento lavorano all'inizio dell'anno 7.450 operai; nel corso dell'anno vengono assunti 1.672 operai, vengono dimessi 1.423 operai. Con quale rapporto, o con quali rapporti, si potrebbe misurare l'intensità del ricambio della massa operaia? Se durante l'anno sono avvenuti, nello stabilimento, 73 infortuni nel lavoro, mediante quali rapporti si può misurare la frequenza degli infortuni per operaio impiegato, per operaio - giornata, per

operaio - ora? Si può misurare la probabilità statistica d'infortunio per ogni operaio? e come?

34. — Quali differenze intercedono tra la probabilità statistica, la probabilità matematica e la probabilità come s'intende nel parlar comune?

35. — In una popolazione esistono al principio dell'anno 517.611 donne coniugate. Nel corso dell'anno avvengono 24.367 matrimoni, 9.891 morti di uomini coniugati, 8.703 morti di donne coniugate; immigrano 1.031 donne coniugate, ne emigrano 756. Si calcoli, con opportuni criteri approssimativi: *a*) la frequenza delle morti, delle vedovanze, delle emigrazioni, per 100.000 coniugate-anno; *b*) la probabilità statistica di morte, di vedovanza, di emigrazione, per ciascuna coniugata (se si ritiene di poterla calcolare). Si chiarisca il significato dei rapporti ottenuti. Si calcolino i vari rapporti di incremento positivo e negativo della popolazione femminile coniugata consentiti dai precedenti dati e si calcoli il saggio d'incremento netto.

36. — Si cerchi una espressione algebrica generale del rapporto di frequenza per caso osservato e per unità di tempo, nell'ipotesi che al complesso dei casi osservati affluiscano nel corso delle osservazioni nuovi casi, e defluiscano da esso casi sia per il manifestarsi del fenomeno osservato sia per altre cause (come avviene nell'esempio del precedente n. 35: fenomeno osservato: *morti*; afflusso di nuovi casi nella massa delle coniugate: per *matrimonio* e per *immigrazione*; deflusso dalla massa delle coniugate: per *morte*, per *vedovanza* e per *emigrazione*). Avvertasi che per giungere ad una formola generale bisogna supporre di conoscere l'istante in cui si verifica ogni afflusso o deflusso individuale; ovvero ammettere, in via di approssimazione, che ciascuna categoria di afflussi e di deflussi sia uniformemente distribuita per l'intero periodo di osservazione; o fare altra ipotesi che sembri più attendibile.

37. — Si graduino le varie regioni italiane secondo la frequenza degli omicidii, rilevabile dalle percentuali per 100.000 abitanti riferite nell'ASI. Si ha nulla da obiettare alla formulazione del precedente quesito?

38. — Sui dati del capitolo « Istruzione » dell'ASI relativi alle scuole elementari, si calcoli, per ciascuna regione (compartimento scolastico): la probabilità statistica che un fanciullo obbligato alla scuola sia effettivamente iscritto, quella che un fanciullo iscritto giunga ad essere esaminato, quella che un fanciullo obbligato alla scuola sia promosso negli esami, quella che un fanciullo iscritto giunga ad essere promosso, quella che un fanciullo esaminato sia promosso. Si chiarisca il significato di ciascuna di tali probabilità statistiche. Si calcoli il rischio di non essere promosso, per il fanciullo esaminato, in ciascun compartimento.

39. — Si classifichino secondo la natura del riferimento i rapporti di frequenza contenuti nel capitolo « Viabilità » dell'ASI.

40. — Mediante quali riferimenti si potrebbe meglio apprezzare il significato dei dati sulla disoccupazione contenuti nell'ASI?

41. — Si cerchino nell'ASI esempi di rapporti di densità e si classifichino i rapporti stessi secondo la natura del riferimento. Quali rapporti di densità potrebbero calcolarsi mediante i dati sui trasferimenti di proprietà contenuti nell'ASI? Quali mediante i dati sull'industria elettrica (numero e potenza degli impianti e quantità di energia prodotta)? Si chiarisca il significato dei rapporti enumerati.

42. — Quali rapporti di diffusione si potrebbero calcolare su dati del capitolo «Climatologia, ecc.» dell'ASI? Quale sarebbe il significato di ciascuno di essi?

43. — In che differisce il rapporto di diffusione da quello di densità? Il rapporto di densità da quello di frequenza?

44. — Dai rapporti di intensità calcolati nei precedenti esercizi si desumano rapporti di estensione e si indichi se ed in quali casi l'impiego di questi sia preferibile all'impiego di quelli.

45. — Dai dati sul numero degli autoveicoli esistenti nelle singole regioni italiane si desumano rapporti di intensità, con riferimento: *a*) alla superficie territoriale; *b*) alla popolazione. Indi si calcolino i corrispondenti rapporti di estensione e se ne illustri il significato.

46. — Si calcolino, comparando fra loro dati tolti da differenti capitoli dell'ASI, rapporti di coordinamento che siano logicamente ammissibili e praticamente utili.

CAPITOLO VIII.

La comparazione fra più dati: la serie statistica.

Serie statistiche di primo, secondo, terzo ordine — Funzioni statistiche — Serie geografiche, serie cronologiche, distribuzioni di frequenze — Campo di variazione dei termini della serie — Disposizione dei termini della serie — Opportunità di visione sintetica ed analitica dei termini della serie — Metodi di rappresentazione sintetica, sintetico-analitica, analitica della serie.

1. — Dall'osservazione di uno stesso fenomeno in circostanze differenti si desumono più dati statistici omogenei, i quali di solito vengono ordinati secondo la circostanza, o secondo le circostanze, per cui differiscono, in modo atto a facilitare l'apprezzamento comparativo e ad agevolare la visione di eventuali relazioni tra le variazioni del fenomeno e le variazioni di modalità o di misura della circostanza o delle circostanze considerate. Se abbiamo dati sulle importazioni di frumento in una serie di anni, li disporremo in ordine cronologico, per meglio scorgere la tendenza di questa corrente commerciale nel tempo; se abbiamo dati sui salari degli operai in un grande numero di stabilimenti, li ordineremo per industrie per discernere come varii il salario col variare dell'occupazione; se abbiamo dati sulla delinquenza nelle varie provincie, li disporremo nell'ordine geografico, o nell'ordine alfabetico delle provincie, o nell'ordine di crescente o decrescente intensità del fenomeno, per renderci più prontamente conto della distribuzione territoriale della delinquenza. Se abbiamo accertato la religione degli uomini e quella delle donne che contraggono matrimonio in un dato paese in un

dato anno, potremo ordinare i dati statistici sui matrimoni secondo la religione degli sposi, o secondo quella delle spose, o secondo l'una e l'altra ad un tempo; se abbiamo accertato la statura ed il perimetro toracico di ciascun coscritto, potremo ordinare i dati statistici sui coscritti secondo la statura, o secondo il perimetro toracico, o secondo l'una e l'altro ad un tempo.

Dati statistici omogenei ordinati secondo una, secondo due, secondo *n* circostanze differenziali, costituiscono *serie statistiche* di primo, di secondo, di ennesimo ordine. (Si rammentino gli aggruppamenti di primo, di secondo, di ennesimo ordine, dai quali appunto derivano le serie statistiche di ordine corrispondente).

Quando le circostanze che servono di base all'ordinamento sono quantitative, e a ciascun valore della singola circostanza o a ciascuna combinazione di valori delle diverse circostanze corrisponde un solo dato statistico, si possono riguardare i dati statistici come tanti valori di una funzione di una o di più variabili, che, in considerazione della natura delle variabili, si può chiamare *funzione statistica*. Per esempio, una serie di dati sulla popolazione censita può essere presentata quale funzione dell'età, una serie di dati sugli appartamenti esistenti quale funzione del numero delle camere e del numero degli occupanti.

Tratteremo per ora delle serie di prim'ordine, riservandoci di trattare più avanti delle serie d'ordine superiore.

Tra le serie ordinate secondo circostanze qualitative meritano particolare rilievo quelle che rappresentano distribuzioni di fenomeni secondo circoscrizioni territoriali, o più generalmente distribuzioni *geografiche* o *topografiche*: serie che possono venir tradotte in speciali forme grafiche. Tra le serie ordinate secondo circostanze quantitative sono degne di nota quelle *cronologiche*, che mostrano lo sviluppo o la manifestazione di un fenomeno attraverso il tempo, e le così dette *distribuzioni di frequenze*, o *seriazioni*, che indicano come si distribuiscano secondo la grandezza le misure d'una circostanza quantitativa caratteristica dei singoli casi osservati od accertati (peso del corpo, reddito, durata della vita). Ma è prematuro fermarci ad una classificazione delle serie statistiche, sia perchè oltre le serie di dati greggi s'incontrano serie di dati elaborati, alcuni tipi dei quali non conosciamo ancora, sia perchè i metodi di rappresentazione delle serie statistiche, che tra breve studieremo, in parte sono applicabili a tutte le classi od a più classi di serie. Più utilmente potremo intrattenerci, dunque, sulle varie classi di serie dopo avere studiato i metodi per la loro rappresentazione.

I dati che costituiscono una serie si dicono *termini* della serie. Il minimo ed il massimo fra essi delimitano il *campo di variazione* dei termini della serie (e non il *campo di variabilità*, come talvolta impropriamente si dice: questa seconda espressione è atta ad indicare i limiti fra i quali *possono variare* le misure del fenomeno, non quelli fra i quali esse effettivamente si mantengono in una serie di osservazioni). La delimitazione del campo di variazione assume speciale importanza quando i dati della serie indicano l'intensità di un fenomeno in varie sezioni di uno stesso campo d'osservazione, o in varie epoche nello stesso campo d'osservazione, o costituiscono misure di circostanze quantitative caratteristiche di singoli casi osservati od accertati.

Una serie statistica di prim'ordine essendo costituita da una successione di dati ordinati secondo una sola circostanza viene contenuta in una sola riga o in una sola colonna. Ma, a rendere palese il criterio e il modo dell'ordinamento, è indispensabile un'altra riga, o un'altra colonna, parallela alla prima, nella quale siano indicate le varie modalità o misure della circostanza secondo la quale i dati sono stati ordinati. Solo eccezionalmente si omette questa indicazione, quando si tratti di serie brevissime, il cui criterio di ordinamento risulti in modo non dubbio dal testo nel quale sono incluse. Si preferisce la disposizione per riga oppure quella per colonna seguendo ragioni di opportunità; in generale quando i termini della serie sono numerosi vengono disposti in colonna, perchè in tal modo l'occhio riesce più rapidamente a paragonarli fra loro, essendo incolonnate le unità dello stesso ordine.

Quando la circostanza secondo la quale sono ordinati i dati di una serie è la grandezza dei dati medesimi (come in una serie di stature, di salari, di prezzi, ecc.), l'ordinamento ha importanza puramente formale; e se è stato alterato può sempre venire ricostruito perchè ogni dato reca in se medesimo il criterio fondamentale dell'ordinamento.

Ma quando i dati sono ordinati secondo una circostanza diversa dalla loro grandezza, in generale l'ordine in cui sono disposti ha importanza sostanziale: per la retta interpretazione della serie conviene che ciascun dato non sia disgiunto dalla corrispondente indicazione della modalità d'una circostanza qualitativa o della misura di una circostanza quantitativa.

Possiamo studiare ed elaborare la serie di dati che ci rappresentano i redditi dei singoli capifamiglia italiani pur ignorando quali siano il nome e cognome che corrispondono a ciascun reddito; ma possiamo trarre scarso frutto dalla serie dei dati sulla mortalità nei singoli comuni italiani se ignoriamo a qual comune corrisponda ciascun dato.

2. — La serie statistica costituisce ad un tempo l'ordinamento dei singoli termini che permette di ritrovare prontamente il dato corrispondente ad una determinata modalità o misura della circostanza regolatrice dell'ordinamento, e la raccolta dei vari dati che permette di desumere dall'ispezione generale di essi un'impressione d'insieme sul fenomeno rappresentato. Secondo la natura dei dati che abbiamo innanzi e secondo il fine che perseguiamo, ora prevale in noi il bisogno della sintesi, ossia quello di scorgere ciò che i vari dati della serie hanno in comune; ora prevale il bisogno dell'analisi, cioè quello di discernere le differenze che intercedono tra essi. Seguiamo i prezzi delle varie merci dopo la stabilizzazione della lira, ricercando sopratutto le analogie dei loro movimenti; esaminiamo i dati sulla frequenza dei delitti di sangue nelle varie provincie, col principale intento di renderci conto delle speciali caratteristiche di ciascuna di queste. Talvolta si desidera una doppia sintesi: una visione d'insieme dei dati della serie, un'altra visione d'insieme delle loro disuguaglianze. Conoscendo i redditi di tutti i cittadini cerchiamo tanto di formarci un'idea generale del livello delle entrate quanto di avere un'impressione riassuntiva sulla disuguaglianza di distribuzione delle entrate stesse.

Quando si esamina una sola serie statistica, composta di pochi termini, riesce abbastanza facile anche all'inesperto ottenerne una visione sintetica ed analitica nel tempo stesso. Ma quanto più numerosi sono i termini della serie, tanto più difficili divengono l'analisi e la sintesi. Ponetevi innanzi una serie costituita da parecchie migliaia di termini rappresentanti la mortalità per tubercolosi nei singoli comuni italiani, e cercate di giungere con la semplice ispezione dei dati ad impressioni particolari e ad una impressione generale: vedrete le ardue difficoltà che si oppongono al raggiungimento di tale meta. D'altra parte, se dovrete paragonare tra loro più serie di dati, anche composte di non molti termini ciascuna, la comparazione analitica vi riuscirà penosa; e dopo aver confrontato tra loro i termini corrispondenti delle diverse serie vi toccherà poi ancora ritrarre un unico giudizio dai numerosi confronti: il che non vi riuscirà facile. O per la numerosità dei termini della singola serie o per la numerosità delle serie da comparare, s'impone in molti casi la ricerca di mezzi che consentano il confronto sintetico. È vero che, dopo esser ascesi alla sintesi, spesso si sente il bisogno di ridiscendere all'analisi, per completare, rettificare e precisare il proprio giudizio; ma è anche vero che per lo più soltanto la sintesi può consentire il primo orientamento indispensabile appunto per giungere ad un illuminato giudizio.

Perciò fra i vari metodi di rappresentazione delle serie statistiche cominceremo col considerare il metodo puramente e perfettamente sintetico della *media*, col quale ad una serie numerosa quanto si voglia di termini si sostituisce un unico dato atto a riassumerla. Ma poichè in generale questo metodo ha il difetto di far perdere di vista le differenze non irrilevanti esistenti fra i dati della serie, vedremo poi come si cerchi di recarvi rimedio mercè l'uso di *dati sussidiari alla media*. I quali, col loro moltiplicarsi per divenir più efficaci, ci condurranno a discorrere di metodi analitici di rappresentazione, che vogliono essere nel tempo stesso sintetici: *rappresentazioni grafiche* di vari tipi, e *interpolazione analitica* cioè rappresentazione della serie mediante una formola che esprime simbolicamente i termini in funzione della circostanza quantitativa secondo la quale essi sono ordinati. Esporremo infine il metodo dei *numeri indici,* il quale pur essendo puramente analitico nell'applicazione alla serie singola ci permetterà di passare dal problema della rappresentazione di singole serie a quello della sintesi di più serie di dati ordinati secondo la medesima circostanza.

Quesiti ed esercizi: 1. — Basta mettere in colonna o in riga un certo numero di dati statistici quali si siano per ottenere una serie statistica? Come si può definire la serie statistica? Quando una serie statistica si dice di prim'ordine? quando di secondo?

2. — Si cerchino nell'ASI serie statistiche di primo, di secondo, di terzo, di quarto ordine.

3. — Si distinguano, nell'ASI, serie di prim'ordine ordinate secondo circostanze quantitative e ordinate secondo circostanze qualitative; si tenti una classificazione delle une e delle altre.

4. — Che cos'è una funzione statistica?

5. — Si ricerchi quali delle serie ordinate secondo circostanze quantitative (es. 3) possano riguardarsi come successioni di valori di funzioni statistiche; quali invece come successioni di somme di valori di funzioni statistiche.

6. — Si cerchino nell'ASI esempi di serie geografiche, di serie cronologiche, di distribuzioni di frequenze.

7. — Che cos'è il campo di variazione di una serie statistica? Perchè va preferita l'espressione «campo di variazione» all'altra «campo di variabilità»?

8. — Si determini per il periodo considerato nelle «Notizie statistiche retrospettive» dell'ASI il campo di variazione: della frequenza delle nascite per 1000 abitanti, dell'eccedenza dei nati sui morti per 1000 abitanti, della frequenza degli emigranti per 1000 abitanti, del numero dei fallimenti, della produzione del frumento, della produzione della ghisa, della produzione dello zucchero, dell'importazione del caffè, dell'importazione del carbon fossile, dell'esportazione di seta tratta greggia.

9. — Si determini il campo di variazione della densità della popolazione per kmq. nelle varie provincie italiane, il campo di variazione della frequenza delle nascite per 1000 abitanti nelle provincie stesse.

10. — Si ricerchino nell'ASI, capitoli «Territorio e popolazione, agricoltura, commercio coll'estero», serie di dati che potrebbero utilmente essere presentati anche in un ordine diverso da quello ivi adottato.

11. — Si espongano applicazioni, che già siano note per precedenti studi, dei vari metodi di rappresentazione delle serie statistiche; e si mettano fino da ora in evidenza i differenti fini ai quali corrispondono i vari metodi.

CAPITOLO IX.

La rappresentazione sintetica della serie statistica: la media.

Sintesi di una serie statistica mercè un unico numero: la media — Media oggettiva e media soggettiva — Definizione matematica della media — Definizione logica della media — Condizioni alle quali viene subordinata la scelta di una media: condizioni di equilibrio, di equivalenza, di accostamento — Gli scostamenti — Condizioni cui soddisfano le principali medie in uso: la mediana, la media aritmetica, la media geometrica, il valore equidistante dagli estremi, il valore più frequente — Proprietà di queste medie e criteri generali per l'impiego di esse — Altre medie di secondaria importanza: media armonica, media quadratica, valore divisorio — Medie ponderate e medie semplici — Pesi, pesi relativi — Avvertenze per il computo delle medie — Criteri per la scelta fra le diverse medie.

1. — Sostituire un unico numero a ciascuno dei vari termini di una serie statistica è sempre possibile aritmeticamente; ma non è sempre ammissibile logicamente. La sostituzione in generale è lecita soltanto quando i vari termini della serie misurano la manifestazione di un fenomeno in corrispondenza ad altrettanti singoli casi osservati od accertati, ad altrettanti intervalli di un unico campo di osservazione, ad altrettanti valori o inter-

valli di valore di una variabile, che possano riguardarsi reciprocamente fungibili, perchè esemplari diversi di una medesima specie, ai fini dell'indagine che si vuol compiere. Così normalmente potremo surrogare un sol numero a ciascuno dei numeri che rappresentano i salari dei diversi operai di una industria, o i raccolti di barbabietole ottenuti in diversi campi aventi uguale estensione, o le esportazioni italiane in diversi anni successivi, o le temperature dell'aria al principio di ciascuna delle 24 ore di una giornata. Ma normalmente non surrogheremo un sol numero a ciascuno di quelli che rappresentano i valori dei beni trasmessi per successione nelle diverse regioni italiane, o i raccolti di granturco ottenuti in diverse aziende, o le consistenze dei diversi gruppi professionali oppure dei diversi gruppi d'età di una popolazione, o i prezzi unitari di merci diverse. La media non è dunque un metodo universale di rappresentazione delle serie statistiche, sebbene a prima vista e dal puro aspetto aritmetico possa sembrar tale.

Nel seguito del presente capitolo non ripeteremo questa riserva ad ogni passo, ma chi legge deve averla sempre presente. Vedremo subito come ad essa si aggiungano altre riserve, le quali restringono ancora il campo — che tuttavia rimane vastissimo — delle applicazioni della media.

Se una serie statistica è costituita da termini poco differenti tra loro (compresi cioè in un campo di variazione ristretto), un solo numero può rappresentarla con sufficiente approssimazione a molti fini per i quali non si richiede una precisione rigorosa. Per esempio, se, in un reggimento di truppe scelte, le stature dei soldati variano fra un minimo di cm. 164 e un massimo di 168, qualunque numero compreso fra questi estremi ci darà un'idea sufficientemente approssimata della statura di quelle truppe. Non sarebbe sensato scegliere al nostro fine un numero inferiore alla statura minima o superiore alla massima; sarà ragionevole scegliere un numero intermedio fra l'una e l'altra: se conosciamo soltanto i due estremi dianzi indicati probabilmente assumeremo come espressione sintetica della serie delle stature il numero di cm. 166, ugualmente distante dai due estremi.

Se, invece, una serie statistica è costituita da termini molto differenti tra loro (compresi cioè in un campo di variazione ampio) nessun numero potrà da solo rappresentarla con sufficiente approssimazione, poichè ogni numero vicino ad alcuni dei termini, e quindi adatto a rappresentarli, sarà necessariamente lontano da altri, e quindi disadatto a rappresentarli. Per esempio, se i redditi dei capifamiglia di uno Stato variano fra un minimo di 5.000 lire e un massimo di 5.000.000, ci guarderemo bene dall'assumere ad espressione sintetica dei redditi stessi la semisomma dei due estremi, come abbiamo fatto nell'esempio precedente, perchè una conoscenza anche superficiale della distribuzione dei redditi ci ammonisce che il numero così determinato risulterebbe relativamente prossimo a pochi redditi alti, ma relativamente lontanissimo da molti redditi bassi. Analoghi inconvenienti presenterebbe qualsiasi altro numero, compreso tra il reddito minimo e il reddito massimo, che volessimo assumere ad unico rappresentante dell'intera serie dei redditi.

Abbiamo detto che un numero intermedio fra i termini della serie può rappresentarli tutti, in via approssimativa, quando essi siano poco differenti tra loro; non può rappresentarli quando siano molto differenti. Avvertiamo ora che il nostro modo di esprimerci dà forma solo apparentemente precisa ad un concetto che è necessariamente impreciso, perchè secondo la natura dei dati, secondo il fine per il quale s'impiegano, le stesse differenze possono apparire rilevanti od irrilevanti. Non si può stabilire una rigorosa regola generale che indichi in quali casi un sol numero è adatto a sostituire una serie statistica e in quali non lo è. Considerazioni di carattere prevalentemente pratico ispirano il giudizio in ogni caso concreto. Ci è parso dianzi assurdo prendere un solo numero a rappresentare il reddito individuale dei capifamiglia di un paese; eppure ci pare ragionevole confrontare il reddito medio del capofamiglia italiano, diciamo 10.000 lire, col reddito medio del capofamiglia nordamericano, diciamo 40.000 lire, per concludere che il livello delle entrate è circa quattro volte più alto negli Stati Uniti che in Italia. Ma che cos'è questo « reddito medio » ? Non è altro che un unico reddito assunto a rappresentare sinteticamente tutta la serie dei redditi: ed ecco che abbiamo applicato proprio quel procedimento che or ora avevamo scartato.

Il criterio, che ci si è presentato ovvio da principio, della « piccolezza delle differenze » tra i termini della serie non è dunque decisivo nella determinazione della opportunità dell'uso della media, sebbene in molti casi venga utilmente applicato. Ad escluderne l'applicabilità generale concorre un'altra considerazione: quella che talvolta una serie statistica non solo *può* ma *deve* essere rappresentata sinteticamente da un unico numero, perchè i suoi termini rappresentano tante misure di un'unica grandezza (per esempio tante misure della distanza fra due punti della superficie terrestre), della quale attraverso le varie misurazioni cerchiamo l'espressione più attendibile. I dati di una serie di questo genere possono pur differire fortemente gli uni dagli altri, ma nessuno negherà che sia legittimo, anzi necessario, cercare di riassumerli in un sol numero. Ed anche quando i diversi termini di una serie rappresentano misure di grandezze diverse, talvolta le differenze tra essi esistenti dipendono in parte da errori di rilevazione, così che un solo numero può differire poco dai dati esatti, ignoti, pur differendo fortemente dai dati errati, noti.

Per non racchiuderci in una cerchia più angusta di quella che i bisogni della pratica hanno assegnata alle applicazioni del metodo di rappresentazione sintetica delle serie statistiche, dobbiamo limitarci a constatare che molte volte una serie può essere utilmente rappresentata mediante un solo numero. Questo numero viene chiamato *media* della serie. La media dicesi *oggettiva* se le corrisponde un unico oggetto concreto di cui i termini della serie costituiscono varie misure; *soggettiva* se rappresenta una sintesi astratta delle misure di tanti oggetti differenti quanti sono i termini della serie. La media dei risultati di diverse misurazioni dell'altezza del monte Cervino è una media oggettiva, la media dei pesi dei bambini d'una classe scolastica è una media soggettiva. Si noterà che gli attributi « oggettiva » e « soggettiva » sono impiegati in un significato non coincidente con quello che si dà loro nel parlar comune.

La media può essere determinata con criteri diversissimi: qualsiasi serie di dati ammette non una ma infinite medie. Il nome stesso richiama un criterio generale: quello che il numero scelto a rappresentare in forma sintetica tutti i termini sia « intermedio » tra essi.

2. — Matematicamente, infatti, la media va definita come « un numero non inferiore al termine minimo e non superiore al termine massimo della serie statistica ». Se la definiamo come « un numero superiore al termine minimo e inferiore al termine massimo » veniamo a privare di media una serie composta di termini tutti uguali: cioè proprio quella serie di cui la media è sintesi esatta, e non soltanto approssimata.

Nella definizione matematica generale non possiamo aggiungere altre condizioni. Se aggiungiamo che « la media dev'essere funzione simmetrica di tutti i termini », che cioè i termini devono concorrere tutti ugualmente a determinarla, veniamo bensì a delimitare un complesso di medie che appunto corrispondono a tale condizione, ma ne escludiamo altre, per esempio quel numero che nella serie si presenta più spesso di ogni altro e che appunto perciò talvolta viene assunto a rappresentare la serie (statura più frequente, salario più frequente). Se aggiungiamo che «la media deve coincidere con uno o più termini della serie» veniamo a delimitare, con un diverso criterio, un diverso complesso di medie, rispondenti a tale condizione, ma ne escludiamo altre, per esempio quella che ciascuno di noi calcola ed usa frequentemente quando addiziona i termini di una serie e divide la somma per il numero dei termini stessi (così, dicendo che la media dei voti riportati da un laureando negli esami speciali è 26,83 enunciamo un voto che lo studente non può avere riportato in nessun esame, poichè i voti si danno in numeri interi, ma che tuttavia riassume l'intera serie dei voti riportati). Se aggiungiamo che «la media deve accostarsi il più possibile ai termini della serie» (o, sarebbe lo stesso, che «deve discostarsi il meno possibile dai termini della serie») esprimiamo un criterio che con varie interpretazioni serve di fondamento a molte medie, ma che è estraneo alla determinazione di molte altre, per esempio della semisomma dei due termini estremi che pure abbiamo dianzi adoperata.

3. — Logicamente, la media va definita come «un numero atto a rappresentare da solo l'intera serie statistica». La definizione logica implicitamente contiene quella matematica; aggiunge la condizione che la media abbia qualche proprietà che la renda idonea al fine della rappresentazione sintetica. Ma tale condizione è espressa in modo generico e non suscettibile di univoca traduzione matematica: si può dire che in tal forma comprenda come casi particolari le varie condizioni enumerate nel precedente paragrafo e possa comprenderne altre ancora. Di fatto, a rappresentare una medesima serie vengono utilmente impiegate medie varie, che talora differiscono di molto, numericamente, l'una dall'altra.

4. — Le condizioni che si sogliono porre, in concreto, nella ricerca della media sono assai svariate e definiscono tipi diversi di medie, rispondenti a diverse esigenze della scienza e della pratica, che si presentano chiare alla mente di chi conosce bene la materia nel campo della quale la media va applicata. Ma tali condizioni si possono ricondurre a tre classi: condizioni di equilibrio, tendenti ad ottenere che la media occupi nel campo di

variazione dei termini della serie una posizione atta a conferirle attitudine rappresentativa; condizioni di equivalenza, tendenti ad ottenere che la media abbia proprietà uguali a quelle dei termini; condizioni di accostamento, tendenti ad ottenere che la media si avvicini il più possibile ai termini, o ad una parte di essi. È bene avvertire che spesso, ma non sempre, condizioni dell'una classe implicano, logicamente e matematicamente, condizioni delle altre due classi, o di una di esse.

Le condizioni di equilibrio per lo più consistono in relazioni che devono sussistere tra una funzione dei termini precedenti alla media in ordine di grandezza e una uguale funzione dei termini seguenti alla media (per esempio: il numero dei termini precedenti alla media sia uguale a quello dei termini seguenti, la somma degli uni sia uguale a quella degli altri).

Le condizioni di equivalenza sono definite coll'imporre alla media qualche proprietà matematica comune coi termini della serie, così che, surrogata a ciascun termine la media, rimanga costante una funzione dei termini (per esempio: la somma, o il prodotto, o la somma dei quadrati).

Le condizioni di accostamento sono per lo più determinate col rendere minima una funzione dei valori assoluti delle differenze tra la media e i termini, o tra i termini e la media, differenze che si dicono brevemente *scostamenti* (per esempio: sia la minima possibile la somma dei valori assoluti degli scostamenti, o la somma dei loro quadrati).

Soffermiamoci sulla nozione di scostamento. Notiamo anzitutto che gli scostamenti della media dai termini differiscono soltanto nel segno, e non nel valore assoluto, dagli scostamenti dei termini dalla media; ma se si passa agli scostamenti relativi, questi differiscono anche in valore. Poichè secondo che si tratta di media soggettiva o di media oggettiva conviene considerare gli uni o gli altri scostamenti, la differenza ora notata non è priva d'importanza.

Avvertiamo poi che gli scostamenti sono chiamati anche *scarti*, *deviazioni* od *errori*. Dei due primi nomi è evidente il significato; il terzo si può giustificare così: se si tratta d'una media soggettiva le differenze tra la media e i termini indicano gli errori che si compiono col sostituire la media, rappresentazione sintetica astratta, alle misure delle varie grandezze concrete indicate dai singoli termini; se si tratta di una media oggettiva, che si assume come il vero valore dell'unica grandezza misurata dai vari termini, le differenze tra i termini e la media indicano gli errori intervenuti nelle singole misurazioni.

Vedremo ora, passando in rassegna le medie di uso più comune, quali siano le principali condizioni di equilibrio, di accostamento e di equivalenza cui esse soddisfano. È utile conoscere le proprietà delle varie medie per poter intendere come corrispondano a bisogni diversi, per poter giudicare a volta a volta quale di esse convenga impiegare in ciascuna occorrenza, per poter discernere se altri abbia correttamente od erroneamente impiegato una determinata media.

5. — La condizione di accostamento più semplice che possa immaginarsi è quella che la somma degli scostamenti, riguardati tutti come positivi perchè è la loro grandezza e non il loro segno che importa al fine dell'accostamento, sia la minima possibile. Cominciamo dunque col cercare qual è la media che rende minima, in confronto ad ogni altra, la somma dei valori assoluti degli scostamenti.

Supponiamo che la nostra serie statistica consti di n termini, $u_1, u_2, \ldots, u_n$, disposti in ordine crescente di grandezza. La somma dei valori assoluti degli scostamenti dei termini da uno dei termini stessi, diciamo da u_k, è costituita da k differenze tra u_k e i termini che in ordine di grandezza occupano i posti da 1 a k, e di $(n - k)$ differenze fra i termini che occupano i posti da $(k + 1)$ ad n e u_k. Se assumiamo come media il numero M maggiore di u_k ma non maggiore di u_{k+1}, e calcoliamo la somma dei valori assoluti degli scostamenti rispetto a M, invece che rispetto a u_k, il numero delle differenze sia dell'uno sia dell'altro tipo rimane invariato, ma ciascuna delle prime viene aumentata di $(M - u_k)$, ciascuna delle seconde viene diminuita dello stesso ammontare; onde la somma dei valori assoluti degli scostamenti subisce una variazione uguale a $(2k - n)(M - u_k)$, cioè diminuisce se è $2k < n$, resta invariata se è $2k = n$, aumenta se è $2k > n$. È facile ora vedere che se si fanno attraversare a M tutti i valori intermedi fra u_1 e u_n la somma dei valori assoluti degli scostamenti diminuisce progressivamente fino a quando M viene a coincidere col termine centrale della serie se n è dispari, o col primo dei due termini centrali se n è pari; si mantiene costante, nella seconda ipotesi, finchè M rimane non inferiore al più piccolo e non superiore al più grande dei due termini centrali; indi aumenta progressivamente quando il valore di M si sposta oltre il termine centrale, od oltre il secondo dei due termini centrali. Dunque la media che rende minima la somma dei valori assoluti degli scostamenti corrisponde al termine centrale della serie se il numero dei termini è dispari, e ad un qualsiasi numero intermedio fra i due termini centrali (in pratica si usa prendere la semisomma di essi) se il numero dei termini è pari.

La media così determinata si denomina *valore centrale, valore mediano*, o più semplicemente *mediana*. Tutte queste denominazioni sono corrette: non così quella di valore *probabile* che talvolta viene impiegata (*vita probabile* per una generazione di individui vien detta la durata mediana della vita, cioè quella durata che è raggiunta dalla metà del numero totale dei componenti la generazione mentre non è raggiunta dall'altra metà; sebbene impropria, questa espressione è entrata nell'uso).

La mediana soddisfa la più semplice e più razionale condizione di accostamento ai termini della serie; ma soddisfa nel tempo stesso la più semplice condizione di equilibrio fra i termini stessi, poichè una metà del numero totale di questi la precede, l'altra metà la segue, in ordine di grandezza.

La condizione ora enunciata può riguardarsi come un caso particolare della condizione generale che il numero dei termini precedenti alla media in ordine di grandezza costituisca una data frazione del numero totale dei termini; vedremo più avanti che non solo la mediana ma anche altre medie di questo genere sono utilizzate nella pratica.

La mediana ha inoltre la proprietà, per se stessa evidente, che il reciproco della mediana coincide colla mediana dei reciproci dei termini (proprietà importante per alcune applicazioni, nelle quali possono ad arbitrio considerarsi i termini della serie oppure i loro reciproci); e che l'm^{ma} potenza della mediana coincide con la mediana delle m^{me} potenze dei termini.

La mediana non corrisponde a condizioni di equivalenza (nel senso precedentemente spiegato). Non conviene impiegarla quando si desidera una media sensibile ad ogni variazione, anche lieve, dei termini o magari di un solo termine della serie, perchè essa può rimanere immutata nonostante forti variazioni dei termini, com'è ovvio e come appare dagli esempi che riferiamo più avanti al paragrafo 14. Ma in molti casi pratici la diffidenza contro l'uso della mediana, motivata con la « scarsa sensibilità » di essa, non è giustificata: quando i termini della serie sono numerosi e variano prevalentemente in un dato senso, la mediana per lo più ne riflette i movimenti non meno bene di altre medie.

Un vantaggio veramente grande della mediana consiste nella facilità di determinazione, poichè ad ottenerla basta ordinare per grandezza i dati e scegliere quello di mezzo: se si ha un reggimento in piazza d'armi, basta disporre i soldati in ordine di statura e misurare soltanto quello che occupa il posto centrale per ottenere una semplice ed efficace sintesi delle numerosissime stature non misurate.

Insieme con altre medie caratterizzate dal posto che occupano nella serie o nel campo di variazione dei termini di questa, la mediana appartiene alla classe delle *medie di posizione*.

6. — Rispetto alla mediana *il numero* degli scostamenti positivi uguaglia quello dei negativi. Si può pensare ad una media contrassegnata invece dalla condizione d'equilibrio che *la somma* degli scostamenti positivi uguagli quella dei negativi, di modo che la somma di tutti gli scostamenti (qui considerati col loro segno, perchè si tratta di una condizione di equilibrio, e non di accostamento) si annulli. Dalla condizione $(M-u_1)+(M-u_2)+\ldots+(M-u_n)=0$ si desume immediatamente che la media cercata è uguale alla somma dei termini divisa per il loro numero. È questa la « media » per antonomasia, che ognuno conosce ed applica in mille circostanze: la *media aritmetica*. Si chiama così perchè in una progressione aritmetica (contrassegnata cioè dalla costante differenza tra due termini consecutivi) il suo valore coincide con quello del termine di mezzo, ossia col valore mediano. La proprietà di equilibrio ora indicata ne implica un'altra: quella che la media aritmetica rende nulla la somma degli scostamenti relativi della media dai termini; essa implica inoltre una notevole proprietà di equivalenza: quella che la media aritmetica lascia immutata la somma dei termini, quando sia sostituita a ciascuno di questi. (E perciò la somma dei rapporti fra i termini e la media risulta uguale al numero dei termini). La stessa proprietà di equilibrio implica, come può facilmente scorgere chi abbia qualche nozione di matematiche superiori, una condizione di accostamento tra le più semplici che si possano stabilire: la media aritmetica rende minima, in confronto a qualsiasi altra media, la somma dei quadrati degli scostamenti. Rappresentando coi soliti simboli i termini della serie, con M la loro media

aritmetica, si trova che la somma dei quadrati degli scostamenti rispetto a questa media è uguale a $u_1^2 + u_2^2 \ldots + u_n^2 - n M^2$. Rispetto ad un'altra media che differisca, in eccesso o in difetto, di a dalla media aritmetica, sviluppando la somma dei quadrati degli scostamenti la si trova uguale alla precedente espressione aumentata del termine positivo na^2, e quindi maggiore che rispetto alla media aritmetica. Ecco dunque verificata la proprietà sopra enunciata.

La grande diffusione dell'uso della media aritmetica non è ingiustificata, date le notevoli condizioni di equilibrio, di accostamento e di equivalenza cui essa simultaneamente soddisfa. In molti problemi pratici è opportuno, in altri addirittura necessario valersi di una media tale che surrogata a tutti i termini della serie lasci immutata la somma dei termini stessi. Ciò avviene specialmente quando i termini rappresentano quantità non solo astrattamente sommabili ma anche concretamente cumulabili. Se abbiamo misurato le stature di 100 persone, possiamo sommare i risultati delle misurazioni e dividere la somma per 100: otterremo così la statura media aritmetica. In questo caso l'operazione aritmetica non può avere, neppure in via d'ipotesi, riscontro nella realtà, essendo impossibile cumulare le cento stature e poi dividere il cumulo in cento parti uguali. Se abbiamo invece registrato i redditi di 100 persone, possiamo ancora sommare i dati e dividere la somma per 100: otterremo il reddito medio aritmetico. Qui l'operazione può avere riscontro nella realtà: possiamo immaginare che i cento redditi siano versati in una sola cassa, e indi ridistribuiti in cento parti uguali. La statura media aritmetica è quella che toccherebbe a ciascuno se la somma delle stature venisse ripartita in parti uguali, il reddito medio aritmetico è quello che toccherebbe a ciascuno se la somma dei redditi venisse ripartita in parti uguali: formalmente le due medie hanno lo stesso significato, ma sostanzialmente la prima ha valore rappresentativo inferiore a quello della seconda perchè l'operazione astratta attraverso la quale è stata determinata non ha alcuna possibilità di concreta attuazione. Un altro esempio: un imprenditore acquista a prezzi diversi varie partite della stessa materia prima; vuol calcolare il prezzo medio pagato, che per lui vuol dire costo medio unitario della materia prima. Non può servirsi che della media aritmetica, poichè, data la natura del problema, bisogna che la media, applicata ipoteticamente a ciascuna unità acquistata, riproduca nella somma il prezzo totale effettivamente pagato.

Si deve, tuttavia, riconoscere che la vastità del campo di corretta applicazione della media aritmetica l'ha fatta entrare nell'uso a tal punto da farla talvolta impiegare in casi nei quali sarebbero meglio adatte altre medie, come avremo occasione di illustrare con qualche esempio nel corso del nostro studio.

7. — Se invece di porre la condizione d'equivalenza che la somma dei termini sia lasciata immutata dalla media — condizione la quale conduce alla media aritmetica — si pone quella che sia lasciato immutato il prodotto quando la media venga sostituita ai termini della serie, dalla condizione $M^n = u_1 u_2 \ldots \ldots u_n$ si ricava immediatamente l'espressione della media cercata, che risulta uguale alla radice ennesima del prodotto degli n termini. È questa la *media geometrica*, così denominata perchè in una progressione geometrica (contrassegnata cioè dalla costanza del rapporto fra due termini successivi) il suo valore coincide con quello del termine di mezzo, ossia col valore mediano. La media geometrica è applicabile soltanto a serie statistiche composte di dati tutti positivi: la presenza di dati nulli o negativi (quali possono comparire in una serie di scostamenti, in una serie esprimente aumenti e diminuzioni alternati, in una serie esprimente saldi creditori e saldi debitori, ecc.) esclude la possibilità dell'uso di essa. Per calcolarla conviene servirsi delle tavole di logaritmi: come risulta dalla condizione sopra indicata, il logaritmo della media geometrica è uguale al quoziente della somma dei logaritmi dei termini per il loro numero; perciò il calcolo della media geometrica sui numeri si riconduce al calcolo della media aritmetica sui logaritmi. E il logaritmo della media geometrica ha, rispetto ai logaritmi dei termini, tutte le proprietà della media aritmetica.

Il reciproco della media geometrica è eguale alla media geometrica dei reciproci dei termini; la sua n^{ma} potenza è eguale alla media geometrica delle n^{me} potenze dei termini: proprietà parallele a quelle già notate nella mediana.

8. — Ad una condizione di equilibrio corrisponde il *valore equidistante dagli estremi*, che è dato dalla semisomma del termine minimo e del termine massimo della serie; esso corrisponde anche ad una condizione particolare di accostamento poichè rende minima, in confronto a qualsiasi altra media, la somma dei quadrati degli scostamenti dai due termini. In generale può essere utilmente adoperato soltanto quando le due metà nelle quali esso divide il campo di variazione dei termini contengano un numero di termini non

molto differente l'una dall'altra. Trova scarsa applicazione nell'indagine scientifica ma è spesso impiegato nella pratica, per la semplicità della sua determinazione.

9. — Sul principio della suddivisione del campo di variazione si fonda anche, sebbene con diverso criterio, un'altra media: il *valore più frequente* (*valore normale, norma, moda, valore di massima densità*). Se si divide il campo di variazione dei termini della serie in tanti intervalli uguali, può avvenire che partendo da uno degli estremi il numero dei termini contenuto in ciascun intervallo tenda ad aumentare fino ad un certo intervallo; che indi movendo da questo tenda a diminuire fino all'altro estremo. Così, in una certa distribuzione di stature, se si parte dalla statura minima, poniamo cm. 154, il numero delle stature contenute in ciascun intervallo di un centimetro tende a crescere fin verso il 166mo cm.; poi tende a diminuire se si procede ancora verso la statura massima di cm. 192. Così pure, in una certa distribuzione di redditi, se si parte dal reddito minimo, poniamo 5.000 lire, il numero dei redditi contenuti in ciascun intervallo di 1.000 lire tende a crescere fin verso 12.000 lire, mentre da questo limite in su tende a diminuire fino al reddito massimo di 5.000.000. In simili casi, è palese che le misure individuali del fenomeno considerato si addensano intorno ad un dato valore, diradandosi gradualmente, con maggiore o minore regolarità, di mano in mano che da quel valore si procede verso l'alto o si retrocede verso il basso; e sono questi i casi nei quali il valore più frequente ha una spiccata attitudine a rappresentare la serie. Quando le misure individuali del fenomeno si diradano gradualmente di mano in mano che dal valore minimo della serie si procede verso l'alto, o di mano in mano che dal valore massimo si retrocede verso il basso, v'è ancora nella serie un valore più frequente, ma poiche esso coincide col valore minimo nel primo caso, col massimo nel secondo, non è in generale adatto a rappresentare sinteticamente la serie. Quando le serie hanno andamento meno semplice e meno regolare di quello finora supposto, si può ancora presentare talvolta in esse un valore più frequente nettamente segnato, che però in generale ha scarsa o nulla attitudine rappresentativa. In certe serie, poi, è addirittura esclusa la possibilità di determinare un valore più frequente degli altri: così se i termini disposti in ordine di grandezza costituiscono una progressione aritmetica e ciascuno di essi compare una volta sola. Nelle serie composte di pochi termini il valore più frequente in generale non è adatto alla rappresentazione sintetica.

Avvertasi che, col dividere il campo di variazione in intervalli uguali, come finora si è supposto, non si determina *un* valore più frequente, ma si determinano *due* valori limiti dell'intervallo che comprende un numero di termini maggiore di quello compreso in ogni altro intervallo. Quando diciamo che alla statura di cm. 166 corrisponde il massimo numero di coscritti, in realtà ci riferiamo all'intervallo che comprende le stature fra mm. 1655 e mm. 1665, e in via di approssimazione ne indichiamo il punto centrale. Criterio che può essere seguito senza gravi inconvenienti ogni volta che gli intervalli siano relativamente piccoli. Avvertasi inoltre che, secondo l'ampiezza attribuita agli intervalli nei quali vien suddiviso il campo di variazione, la zona di massimo addensamento dei termini della serie può apparire un po' più in alto o un po' più in basso (solo in rari casi si hanno forti spostamenti); talvolta soltanto attraverso successive prove e col sussidio del procedimento dell'interpolazione, del quale tratteremo più avanti, si può giungere ad una attendibile determinazione del valore più frequente. Notiamo che questo valore soddisfa una condizione particolare di accostamento, poichè coincide (rigorosamente o approssimativamente) con un numero di termini massimo in confronto a quello dei termini coincidenti con ogni altro singolo valore. E soddisfa una condizione di equilibrio poichè costituisce il valore intorno al quale tendono ad addensarsi gli altri.

Il valore più frequente è una media di posizione; per determinarlo non si richiedono operazioni aritmetiche, basta l'ordinamento per grandezza dei termini. Se in piazza d'armi si fanno disporre in tante file adiacenti i soldati di un reggimento, in modo che i capifila siano disposti su una stessa riga in ordine di statura e ciascuna fila comprenda soldati aventi tutti la medesima statura, la fila più lunga ci indicherà qual è la statura più frequente, e basta misurare un soldato della fila stessa per conoscere tale statura, che in questo caso può valere come una delle medie maggiormente rappresentative.

10. — Tra le infinite possibili medie, le cinque fin qui studiate — mediana, media aritmetica, media geometrica, valore equidistante dagli estremi, valore più frequente — bastano alla maggior parte dei fini teorici e pratici. Tuttavia è opportuno conoscere alcune altre medie, che hanno applicazioni non prive d'importanza.

Una condizione di equivalenza che talvolta occorre assumere come base per la determinazione di una media è quella che la media stessa, sostituita ai termini, lasci immutata la somma dei reciproci dei termini: condizione che si può scrivere nel modo seguente:

$$n M^{-1} = u_1^{-1} + u_2^{-1} + \ldots\ldots + u_n^{-1}.$$

Ne risulta la *media armonica*, così chiamata perchè in una progressione armonica (costituita cioè dai reciproci dei termini di una successione aritmetica di numeri positivi; in senso più ristretto, generalmente adottato, costituita dai reciproci della successione dei numeri positivi interi da 1 in poi) il suo valore coincide con quello del termine di mezzo, ossia col valore mediano.

La media armonica possiede una notevole proprietà di equilibrio: essa rende nulla la somma degli scostamenti relativi della media dai termini.

Perchè si possa applicare la media armonica, bisogna che la serie sia costituita interamente di termini positivi. Il calcolo di questa media non riesce laborioso quando si disponga di una tavola dei reciproci, perchè la media stessa risulta uguale al reciproco della media aritmetica dei reciproci: definizione che indica la via più rapida per computarla.

11. — Una condizione di equivalenza spesso assunta a criterio per la determinazione d'una media è quella che la media stessa, sostituita ai termini, lasci immutata la somma dei quadrati. o dei cubi. o in generale delle m^{me} potenze, dei termini. Dalla condizione: $n M^m = u_1^m + u_2^m + \ldots\ldots + u_n^m$ si desume facilmente la formola della media cercata: questa risulta uguale alla radice emmesima della media aritmetica delle emmesime potenze dei termini. E quindi l'm^{ma} potenza di tale media possiede rispetto alle m^{me} potenze dei termini tutte le proprietà della media aritmetica.

Il calcolo di queste medie è laborioso: bisogna anzitutto calcolare le m^{me} potenze dei singoli termini (operazione che può essere risparmiata quando si disponga di tavole delle m^{me} potenze dei numeri interi), poi computare la media aritmetica di esse, infine da questa estrarre la radice m^{ma} (col sussidio dei logaritmi).

Per $m = 2$, la precedente condizione dà la *media quadratica* (corrispondente alla radice quadrata della media aritmetica dei quadrati dei termini), cui spesso si ricorre quando si vuole eliminare l'influenza del segno nel calcolare una media di termini in parte positivi e in parte negativi. Per $m = 3$ si ha la *media cubica*, la quale invece risente ed amplifica l'influenza del segno dei termini, così che talvolta viene impiegata appunto come indicatrice della prevalenza dei termini positivi o di quelli negativi. È ovvia l'osservazione che la media aritmetica corrisponde anch'essa alla condizione generale dianzi indicata, per $m = 1$; la media armonica le corrisponde per $m = -1$. Poichè si può dimostrare, col sussidio delle matematiche superiori, che, quando i termini sono tutti positivi e non sono tutti uguali fra loro, la media corrispondente alla condizione generale suddetta cresce al crescere di m, considerata come una variabile continua, ne deriva che la media cubica è superiore alla media quadratica, la quale a sua volta supera l'aritmetica, che è maggiore dell'armonica: disuguaglianze di validità generale che è opportuno conoscere. Si dimostra anche che, quando i termini sono tutti positivi, col tendere di m a zero i valori medi desumibili dalla condizione stessa tendono verso il valore della media geometrica, la quale pertanto risulta costantemente inferiore all'aritmetica e costantemente superiore all'armonica.

12. — Una semplice condizione di equilibrio conduce alla determinazione del *valore divisorio*: media tale che la somma dei termini ad essa precedenti in ordine di grandezza sia uguale alla somma dei termini ad essa seguenti (media che quindi necessariamente risulta maggiore così della mediana come della media aritmetica). In generale tale condizione non può essere soddisfatta in modo rigoroso ma soltanto in modo approssimativo: s'incontra un termine per il quale la somma dei termini precedenti è ancora inferiore a quella dei termini seguenti, mentre per il termine successivo la somma dei termini precedenti è superiore a quella dei termini seguenti: fra questi due termini si assume a valore divisorio quello per cui è minore la differenza tra le due somme, e quindi maggiore l'approssimazione alla condizione fissata; se però la differenza relativa tra le due somme è grande, conviene rinunciare all'uso del valore divisorio, che perde ogni rispondenza alla realtà. Quando la somma dei primi k termini della serie è esattamente uguale a quella dei successivi $(n - k)$ termini, si può assumere come valore divisorio la semisomma di u_k e di u_{k+1}.

Il valore divisorio corrisponde anche ad una condizione di accostamento: rende minima, in confronto ad ogni altra media, la somma dei valori assoluti degli scostamenti relativi dei termini dalla media.

Il valore divisorio si può riguardare come una specie particolare della media corrispondente alla condizione che la somma dei termini ad essa precedenti in ordine di grandezza costituisca una determinata frazione della somma di tutti i termini; anche altre medie di questo genere trovano pratica applicazione.

13. — Quando i termini di una serie statistica, della quale si ricerca la media, non sono tutti differenti tra loro, anzi coincidono a gruppi, la determinazione di quelle medie che richiedono operazioni aritmetiche può essere semplificata, col sostituire operazioni di moltiplicazione ad operazioni di addizione, operazioni di innalzamento a potenza ad operazioni di moltiplicazione.

Se, per esempio, si cerca il prezzo medio del pane negli n comuni italiani partendo da un singolo dato per ciascun comune, non si avranno certamente tanti prezzi diversi quanti sono i comuni, ma in un certo numero n_1 di comuni si avrà il prezzo u_1, in un altro numero n_2 di comuni si avrà il prezzo maggiore u_2, in un altro numero di comuni n_3 il prezzo ancora maggiore u_3, e così via. Se vogliamo calcolare la media aritmetica, invece di sommare n_1 termini u_1 con n_2 termini u_2, con n_3 termini u_3, ecc., e poi dividere la somma per il numero totale dei termini, potremo più rapidamente eseguire i prodotti $n_1 u_1$, $n_2 u_2$, $n_3 u_3$, ecc., sommarli e dividere la somma per il numero dei termini, cioè per $(n_1 + n_2 + n_3 + \ldots)$. Se vogliamo calcolare la media geometrica, invece di eseguire la moltiplicazione di n_1 fattori u_1 per n_2 fattori u_2 per n_3 fattori $u_3 \ldots$, ecc., potremo innalzare u_1 alla potenza n_1^{ma}, u_2 alla potenza n_2^{ma}, u_3 alla potenza n_3^{ma}, ecc. e moltiplicare poi tra loro i valori ottenuti; indi estrarremo dal prodotto la radice $(n_1 + n_2 + n_3 + \ldots)^{ma}$.

Il numero n_i, che indica quanti termini aventi il valore u_i compaiono nella serie di cui si cerca la media, dicesi *peso* col quale il valore u_i concorre nella formazione della media; *peso relativo* è il rapporto $n_i : n$ fra il singolo peso e la somma di tutti i pesi, rapporto che indica quale frazione del numero totale dei termini sia costituita dai termini aventi valore u_i. Il nome di « peso » richiama taluna delle più comuni applicazioni, come la ricerca del prezzo medio di una merce della quale siano acquistate diverse partite, di differente peso, a prezzi differenti. Una media nel determinare la quale si tiene conto dei differenti pesi dei vari valori che concorrono a formarla si dice *media ponderata*, in contrapposizione a *media semplice*. La media semplice però si può riguardare come un caso particolare della media ponderata, nel quale a ciascun valore è attribuito peso uguale all'unità.

Il precedente esempio mostra come i pesi compaiano nella media aritmetica ponderata e nella media geometrica ponderata. E' facile rendersi conto in modo analogo dell'applicazione di essi alla media armonica, alla media quadratica, ecc. Il valore mediano può essere ora definito come quel valore per cui la somma dei pesi dei valori precedenti in ordine di grandezza è uguale alla somma dei pesi dei valori seguenti; il valore più frequente come quel valore cui corrisponde il massimo peso in una serie in cui, ordinati i valori per grandezza, i corrispondenti pesi costituiscano una successione da prima crescente e di poi decrescente.

Il peso non è rappresentato necessariamente da un numero intero, come nel precedente esempio. Se è noto il rendimento per ettaro ottenuto in una data coltura su diversi campi di estensione differente, per calcolare il rendimento medio aritmetico ponderato dovremo assumere come pesi le estensioni dei campi: ed ecco un esempio di pesi che normalmente non saranno numeri interi.

Talvolta, poi, i pesi rappresentano semplicemente coefficienti applicati ad arbitrio al fine di far concorrere in differente misura i differenti termini alla formazione della media, secondo il grado di attendibilità loro attribuito. In una medesima zona agraria abbiamo incaricato diversi osservatori di indicarci il rendimento medio per ettaro della coltura frumentaria, ed uno di essi ci indica 16 quintali, un altro 16,5, un altro 17,5, un altro 18. Se abbiamo uguale fiducia nei vari osservatori calcoleremo una media semplice delle quattro indicazioni, per esempio la media aritmetica: 17. Se invece ci fidiamo del primo 2 volte più che del secondo, 4 volte più che del terzo, 5 volte più che del quarto, ma tuttavia non reputiamo dover trascurare le informazioni degli osservatori meno valenti, determineremo una media attribuendo peso 1 all'indicazione dei 16 quintali, peso 0,50 a quella dei 16,5, peso 0,25 a quella dei 17,5 quintali, peso 0,20 a quella dei 18 quintali. Calcolando la media aritmetica ponderata con questo criterio, otteniamo al numeratore 32,225 come somma dei prodotti dei valori per i pesi loro attribuiti, al denominatore 1,95 come somma dei pesi, e quindi 16,53 quintali per ettaro come rendimento medio unitario.

Si può ora definire in modo più generale il *peso* come un coefficiente numerico che indica l'importanza da attribuire a ciascun valore compreso in una serie statistica nella determinazione di una media della serie stessa.

14. — I termini della serie che si vuol rappresentare sinteticamente possono essere affetti da errori. Come si riflettono questi er-

rori nella media? Non si può dare una risposta generale al quesito: si possono dare risposte particolari riferendosi alle singole medie, o alle singole famiglie di medie. Vediamo qualche caso.

Il valore mediano è determinato attraverso l'ordinamento per grandezza dei termini della serie e non può esprimersi quale funzione di essi; quindi anche il suo errore non può esprimersi in funzione degli errori dei termini. Per quanto gravi siano questi errori, finchè essi non alterano l'ordinamento dei termini in modo da determinare una modificazione del valore che sta nel posto centrale, non alterano neppure il valore mediano. Ma d'altra parte basta che sia errato l'unico termine centrale, pur essendo esatti tutti gli altri, perchè la mediana sia errata nella stessa proporzione. Valgano i seguenti esempi:

a) Serie di dati esatti	10	30	50	70	90
b) Serie di dati in parte errati	5	15	50	90	99
c) »	15	40	50	60	80
d) »	10	30	33	70	90

Nei casi *b*) e *c*) i forti errori della maggior parte dei termini non influiscono sulla mediana, nel caso *d*) il solo errore di un termine influisce fortemente su di essa. Se tutti i termini sono errati in eccesso (o tutti in difetto) la mediana risulta anch'essa errata in eccesso (o in difetto).

L'errore della media aritmetica ponderata (o semplice) si può, invece, esprimere in funzione degli errori dei termini e risulta uguale alla loro media aritmetica ponderata (o semplice). Se gli errori dei termini hanno tutti lo stesso segno, l'errore della media aritmetica coincide pertanto, in valore assoluto, con la media aritmetica dei valori assoluti degli errori dei termini; ma se gli errori dei termini sono in parte positivi e in parte negativi, l'errore della media aritmetica è inferiore, in valore assoluto, alla media aritmetica dei valori assoluti dei termini, e addirittura si annulla se la somma degli errori positivi dai quali una parte dei termini è affetta uguaglia la somma degli errori negativi dai quali è affetta la rimanente parte. Perciò si suol dire che nella formazione della media aritmetica gli errori dei termini tendono a compensarsi.

In una media ponderata influiscono sulla determinazione non solo gli errori dei termini ma anche gli eventuali errori dei pesi. Si può dimostrare che l'errore della media aritmetica ponderata è funzione simmetrica degli errori relativi dei termini e di quelli dei pesi relativi, che cioè dipende in ugual modo dagli uni e dagli altri errori relativi. Però gli errori dei pesi relativi, essendo necessariamente in parte positivi ed in parte negativi (perchè la somma dei pesi relativi errati, al pari di quella dei pesi relativi esatti, è uguale all'unità), tendono a compensarsi reciprocamente in maggiore o minor misura nella formazione della media, mentre gli errori dei termini possono essere anche tutti positivi, o tutti negativi, in modo da escludere ogni possibilità di compensazione. Appunto perciò qualche statistico afferma che gli errori dei termini influiscono maggiormente di quelli dei pesi a determinare l'errore della media aritmetica ponderata: è questa l'espressione inesatta di un'idea sostanzialmente esatta. Considerazioni analoghe a quelle ora esposte valgono per la media geometrica ponderata.

Il valore più frequente risente l'influenza degli errori dei termini soltanto quando questi siano tali da spostare l'intervallo, o il valore, cui corrisponde il massimo numero di termini. E risente l'influenza degli errori dei pesi soltanto quando questi siano tali da spostare il massimo peso da un intervallo, o da un valore, ad un altro. Così che questa media può talvolta risultare esatta anche se tutti i termini sono errati nello stesso senso.

In via generale, e con qualche imprecisione di linguaggio, si può dire che quando gli errori dei termini sono in parte positivi e in parte negativi tendono a compensarsi nella formazione di una media. Ed è perciò che quando si hanno differenti misure di uno stesso oggetto si cerca di desumere dalla loro media (« media oggettiva ») un più attendibile valore: si presume che gli errori in senso opposto si eliminino a vicenda, in tutto o almeno in parte, nella formazione della media. Se si ritiene che nelle misure eseguite abbiano potuto avvenire con uguale frequenza errori in eccesso ed errori in difetto, si può assumere come valore più attendibile la mediana; se si ritiene che abbiano potuto avvenire con uguale frequenza errori in eccesso ed errori in difetto della stessa grandezza, si può assumere la media aritmetica; nè mancano buoni argomenti a pro del valore più frequente, se esso appare ben marcato.

15. — Il computo delle medie può essere accelerato mercè l'uso di opportuni sussidi: tavole di logaritmi, di reciproci, di potenze, ecc., e naturalmente mercè l'impiego di macchine calcolatrici per l'esecuzione delle operazioni aritmetiche.

Semplici artifizi consentono in qualche caso un notevole risparmio di tempo. Per esempio, invece di computare direttamente la media aritmetica dei termini, può convenire di computare la media aritmetica delle differenze fra i termini ed un valore intermedio tra essi, e aggiungendo l'una all'altro ottenere poi la media aritmetica della serie.

La determinazione di una media può presentare qualche difficoltà quando i termini della serie non sono noti individualmente, ma sono soltanto indicati per gruppi i limiti nei quali essi sono compresi. In generale una statistica dei redditi non indicherà i singoli redditi: indicherà, per esempio, il numero dei redditi da 5.000 a 6.000 lire, da 6.000 a 7.000, da 7.000 a 8.000, ecc. L'espediente più usato in simili casi, nel computo di medie che richiedono operazioni aritmetiche, è quello di attribuire a tutti i termini compresi in un intervallo valore uguale alla semisomma dei limiti dell'intervallo: nel nostro esempio, rispettivamente: 5.500, 6.500, 7.500, ecc.; s'intende che in questo modo si giunge ad un valore approssimato e non al valore esatto della media: ma codesta è conseguenza inevitabile dell'ignoranza dei singoli termini; nel nostro caso: dei singoli redditi. Quando si ha ragione di presumere che i termini compresi in un intervallo non vi siano uniformemente distribuiti ma si addensino maggiormente verso l'inizio o verso la fine di esso, conviene attribuire ai termini compresi nell'intervallo un valore inferiore, o superiore, al valore centrale dell'intervallo stesso. Nel caso in cui si adotta il valore centrale, l'errore che deriva alla media dall'applicazione del procedimento approssimativo in luogo del procedimento esatto non può superare un valore corrispondente alla metà dell'ampiezza dell'intervallo: 500 lire nel nostro esempio. Per il computo di medie di posizione (mediana, valore più frequente, valore divisorio) si parte dall'ipotesi della distribuzione uniforme dei termini in ciascun intervallo, oppure da appropriate ipotesi di distribuzione non uniforme.

Talvolta i valori sono raccolti per intervalli di disuguale ampiezza: per esempio: redditi da 5.000 a 6.000, da 6.000 a 8.000, da 8.000 a 12.000, da 12.000 a 20.000, ecc. Si può procedere come nel caso precedente; ma è chiaro che il margine d'errore è più grande.

Può anche accadere che di alcuni intervalli sia indicato soltanto il limite superiore: per esempio, redditi *fino a* 5.000 lire; oppure soltanto il limite inferiore: per esempio, redditi *oltre* 5.000.000 di lire. È chiaro che in simili casi il limite noto indica semplicemente il massimo, oppure il minimo, valore che può attribuirsi ai termini compresi nell'intervallo; soltanto se si ha ragione di ritenere che i termini si addensino molto vicino al limite noto si assumerà questo valore come approssimato in eccesso, o in difetto; altrimenti bisognerà adottare ipotesi che soltanto una profonda conoscenza dei fatti rappresentati dai dati statistici può suggerire attendibili.

Quando si deve calcolare una media di rapporti, avviene talora che il denominatore di ciascuno di questi indichi una dimensione del campo d'osservazione nel quale si svolge il fenomeno espresso nel numeratore, e che mercè il calcolo della media si voglia determinare il valore che avrebbe assunto un rapporto desunto dalla osservazione di un unico campo formato con la riunione di quelli cui si riferiscono i singoli rapporti. Si deve allora determinare la media aritmetica ponderata dei rapporti, assegnando a ciascuno un peso uguale al suo denominatore. La media risulta uguale alla somma dei numeratori divisa per la somma dei denominatori: questa osservazione indica la via più spedita per il calcolo. Cosi se si conosce la frequenza delle nascite in rapporto alla popolazione delle singole provincie italiane, la frequenza media italiana è data dalla media ponderata delle frequenze provinciali, in cui alla frequenza di ciascuna provincia si attribuisce come peso la popolazione della provincia stessa: il che equivale, com'è facile verificare, a dividere la somma dei nati nelle diverse provincie per la somma delle popolazioni delle diverse provincie, cioè il numero dei nati in Italia per la popolazione italiana.

16. — Sui criteri per la scelta tra le differenti medie è stato molto discusso; ma ogni discussione su tale argomento degenera in una casistica appena si esce dal campo di quei criteri che corrispondono alle proprietà matematiche delle varie medie e che abbiamo a volta a volta accennato nei precedenti paragrafi. Ormai nessuno scienziato si affanna più a dimostrare che l'ottima fra le medie è l'aritmetica, o l'armonica, o un'altra qualsiasi, perchè ognuno è convinto che secondo il fine per il quale si vuol impiegare la media è opportuno servirsi di questa o di quella media.

Nel caso della media oggettiva, la scelta in generale è limitata fra la media aritmetica, la mediana, il valore più frequente, mancando ogni ragione di ricorrere ad altre medie; e quando le singole misurazioni cui corrispondono i singoli termini della serie sono compiute con mezzi adeguati e con le opportune cure e cautele, molto spesso i valori delle tre medie coincidono o quasi, così che la questione della scelta fra esse non ha grande rilevanza.

Nel caso della media soggettiva, invece, possono aversi grandi, e talora grandissime, differenze tra i vari termini della serie, e quindi rilevanti differenze fra le varie medie. Se consideriamo la serie dei redditi dei capifamiglia in uno Stato potremo trovare, per esempio, uguale a 7.000 lire il reddito più frequente, a 8.000 il reddito mediano, a 10.000 il reddito medio aritmetico, a 15.000 il reddito divisorio. Certo non sarà indifferente assumere l'una o l'altra di queste medie a rappresentazione sintetica della serie; ma non possiamo dire che una di esse sia sempre da preferire alle altre. Se vogliamo porre in evidenza la condizione economica dominante, adopreremo il reddito più frequente; se vogliamo segnare quel limite sopra il quale si trova una metà dei redditi e sotto il quale si trova l'altra metà ci varremo del reddito mediano; se vogliamo mostrare il risultato che si avrebbe dividendo in parti uguali tra i capifamiglia la somma di tutti i redditi, ricorreremo al reddito medio aritmetico; se vogliamo dare un'idea del concentramento dei redditi più alti in poche mani, ci serviremo del reddito divisorio.

Come abbiamo già notato, la preferenza che nel campo delle applicazioni pratiche viene data alla media aritmetica in parte è fondata sulle sue proprietà, che la rendono atta ad un larghissimo uso, ma in parte anche alla secolare abitudine dell'uso stesso. Aggiungasi la facilità del calcolo: se anche il ragionamento induca a preferire la media geometrica in una data applicazione, chi non abbia familiarità coi logaritmi è portato ad attenersi alla media aritmetica. Dall'aspetto della facilità di determinazione, la mediana, e il valore più frequente nei casi di sua legittima applicazione, vincono la media aritmetica; tuttavia la prima è poco usata nella vita quotidiana, mentre il secondo, specialmente in forma approssimativa, è usato molto spesso. Così quando diciamo: l'operaio meccanico è pagato 4 lire l'ora, la statura dello studente universitario è di cm. 165-170, l'età dell'uomo che contrae matrimonio supera di 5 anni quella della donna, da casa a scuola ci sono 25 minuti di cammino, ecc., non intendiamo riferirci a medie determinate attraverso operazioni aritmetiche, ma esprimiamo quelle che riteniamo le misure normali — cioè le più frequenti — dei fenomeni considerati.

Chi voglia operare razionalmente ed efficacemente, ogni volta che debba usare una media farà bene a chiedersi quale sia più adatta al fine che egli persegue. Abbiamo già avvertito che la mediana, la media aritmetica, la media geometrica, il valore equidistante dagli estremi, il valore più frequente bastano alla massima parte dei bisogni della pratica e della scienza; pertanto la scelta è quasi sempre limitata in un ristretto campo e non riesce laboriosa per chi conosca le proprietà delle varie medie concorrenti.

È ovvia, ma non superflua perchè spesso dimenticata, l'avvertenza che, volendosi paragonare tra loro più serie di dati col metodo delle medie, si deve prendere di tutte le serie la stessa specie di media. Non sarebbe corretto comparare, ad esempio, la media aritmetica dei prezzi del pane nelle città italiane con la media geometrica dei prezzi del pane nelle città francesi, perchè la differenza tra le due medie in parte deriverebbe dalle differenze del prezzo del pane tra i due paesi ma in parte deriverebbe semplicemente dall'essere la media geometrica necessariamente differente dall'aritmetica e inferiore ad essa.

17. — Quando s'impiega una media, qualunque essa sia, per rappresentare una serie statistica, non si deve mai dimenticare che la media offre una rappresentazione dei singoli termini solamente approssimata: e « approssimata » può significare talvolta molto lontana dal vero. Perciò non si deve mai ragionare come se i singoli termini coincidessero con la media, applicando a ciascuno di essi conclusioni che solo in tale ipotesi sarebbero valide. L'avere suddiviso aritmeticamente la somma dei redditi privati in modo da stabilire che il reddito medio della famiglia italiana è di 10.000 lire non ci autorizza a considerare la popolazione italiana come costituita da tante famiglie, ciascuna delle quali abbia proprio 10.000 lire di reddito: la reale distribuzione dei redditi, e quindi del benessere, è estremamente lontana da questa ipotetica uguaglianza. E l'avere stabilito che la durata mediana della vita di una data generazione è stata di 54 anni non ci autorizza a riguardare la generazione stessa come composta da tante persone che abbiano vissuto proprio 54 anni ciascuna: in realtà alcune avranno vissuto pochi istanti, altre 95 o 100 anni. Il più elementare buon senso basta a segnare, nella maggior parte dei casi, i limiti entro i quali può essere lecito ed utile l'impiego della media: esso basta, però, a chi abbia sufficiente conoscenza del fenomeno rappresentato, cioè sufficiente idea della distribuzione per grandezza dei termini della serie; chè altrimenti anche il dotto in altra materia potrà incorrere in errori d'applicazione della media, atti a suscitare legittima diffidenza e giustificata ironia, come quella dell'affamato, dipinto da un

nostro poeta satirico, che conforta il digiuno con la deliziosa visione del pollo che « in media » gli tocca ogni settimana.

La media è il mezzo ideale per la comparazione fra più serie statistiche perchè è il solo che conduca ad un risultato univoco. Ma, per ovvie ragioni, che sono rese chiare dalle considerazioni ora esposte, cade in errore chi presuma di poter applicare il giudizio comparativo desunto dal confronto delle medie di due serie, a due termini rispettivamente compresi nelle due serie. Il reddito medio dell'Italiano è molto inferiore a quello dello Svizzero, ma non perciò siamo autorizzati ad asserire che un Tizio italiano ha un reddito minore di un Sempronio svizzero. Qui l'errore logico è palese; ma molte volte nella vita quotidiana si sentono esprimere conclusioni ugualmente infondate, che vengono accolte da molti, incapaci di scorgerne il punto debole, perchè ignoranti della materia cui le conclusioni si riferiscono.

Indicazioni bibliografiche. — MESSEDAGLIA A., *Il calcolo dei valori medii e le sue applicazioni statistiche*, Torino, Utet, 1908.

Quesiti ed esercizi: 1. — Si distinguano nel capitolo « Territorio e popolazione » dell'ASI le serie per le quali è logicamente ammissibile la sintesi in un unico numero da quelle per le quali non è ammissibile.

2. — Come si può definire in modo generale la media? Quando una media può bastare alla rappresentazione di una serie statistica?

3. — Fra le serie di dati sulla produzione industriale contenute nelle « Notizie statistiche retrospettive » dell'ASI, quali si prestano ad essere rappresentate con medie? E perchè?

4. — I dati sulla produzione del frumento per ettaro nelle varie provincie italiane si possono utilmente riassumere in una media? Il dato medio per l'Italia corrisponde a tale fine?

5. — Che cosa aggiunge la definizione logica della media alla definizione matematica di essa? A qual fine si pongono, nella scelta di una media, condizioni di accostamento, di equilibrio, di equivalenza? Qual è il significato di tali espressioni convenzionali?

6. — Si esponga qualche esempio di media oggettiva; di media soggettiva.

7. — Si possono calcolare gli scostamenti della media dai termini o dei termini dalla media; quale criterio si preferirà nel caso di una media oggettiva? di una media soggettiva?

8. — La serie statistica costituita dai numeri 15, 17, 20, 21, 24, 25, 28 può avere come media il numero 14? il numero 28,1? il numero 27,5? il numero 21? il numero 15? il numero 21,4? Quale di questi numeri è più adatto a servire come media? Perchè?

9. — Conoscete già qualche media tra quelle di uso più frequente? Provatevi a definirla in modo preciso e verificate a quali condizioni corrisponda.

10. — Può esistere una media la quale non varii se un termine della serie, ad essa superiore, aumenta di a o uno ad essa inferiore diminuisce di a?

11. — La mediana è una funzione simmetrica di tutti i termini della serie?

12. — In qual misura varia la media aritmetica di una serie se tutti i termini variano di una stessa quantità a? E se tutti vengono moltiplicati per a?

13. — In qual misura varia la media geometrica di una serie se tutti i termini vengono moltiplicati per una stessa quantità a?

14. — In qual misura varia la media aritmetica se un solo termine della serie varia di a?

15. — In qual misura varia la media geometrica se un solo termine della serie è moltiplicato per a?

16. — In qual misura varia la mediana se un solo termine della serie varia di a?

17. — A quali condizioni di equilibrio corrispondono le principali medie di uso comune? A quali condizioni di accostamento? A quali condizioni di equivalenza?

18. — Si dimostri che la mediana rende minima la somma dei valori assoluti degli scostamenti.

19. — Si calcolino gli scostamenti delle frequenze delle nascite per 1000 abitanti nelle singole provincie italiane dalla media nazionale. In qual senso tali scostamenti si possono chiamare « errori »?

20. — Quali sono le « medie di posizione » e perchè si chiamano così?

21. — Diverse misurazioni topografiche, che si ritengono ugualmente accurate, danno le seguenti espressioni, in metri, di una medesima distanza: 13.041, 13.042, 13.042, 13.044, 13.045, 13.047, 13.047. Quale media si può assumere come espressione più attendibile della distanza misurata?

22. — Proseguendo il precedente esercizio: si calcolino gli scostamenti dei valori sopra indicati dalla media assunta. In qual senso tali scostamenti si possono chiamare errori?

23. — In undici prove di allenamento compiute in undici giorni successivi un corridore ha percorso la distanza che dovrà percorrere in gara, nei seguenti tempi, in minuti secondi: 123, 120, 122, 121, 119, 121, 119, 118, 117, 119, 117. Qual è la durata probabile del percorso? Questa « durata probabile » esprime il tempo che presumibilmente il corridore impiegherà in gara?

24. — Si dimostri che la media aritmetica rende nulla la somma dei valori assoluti degli scostamenti.

25. — Un aeroplano percorre un tratto di dieci km. in linea retta, in sei successivi passaggi, nei seguenti tempi: 61″, 58″, 57″, 60″, 63″, 61″. Qual è il tempo medio impiegato dall'aeroplano per percorrere 1 km.? È questa una media oggettiva o soggettiva? Se i tempi si riferiscono invece a sei aeroplani differenti è media oggettiva o soggettiva? Tornando al caso dell'unico aeroplano, qual è la velocità di esso in ciascuna prova? E qual è la velocità media nel corso delle sei prove? Quale media conviene impiegare in questo caso? Perchè? V'è una ragione speciale per preferire la media aritmetica sem-

plice? o quella ponderata? o la media geometrica? o il valore mediano? Quale relazione esiste tra la media aritmetica semplice dei tempi impiegati per percorrere 1 km. e la media armonica semplice degli spazi percorsi nell'unità di tempo?

26. — A quale media corrisponde il reciproco della media aritmetica dei reciproci dei termini?

27. — Si metta in formola la somma dei valori assoluti degli scostamenti da una media, compresa in ordine di grandezza fra il k^{mo} e il $(k+1)^{mo}$ termine della serie.

28. — Che cos'è la media quadratica? Quali sono le sue proprietà?

29. — Come si può calcolare rapidamente la somma dei quadrati degli scostamenti rispetto ad una qualsiasi media, quando sia nota la somma dei quadrati degli scostamenti rispetto alla media aritmetica? Che proprietà ha quest'ultima media rispetto alla somma dei quadrati degli scostamenti? Che proprietà ha rispetto alla media quadratica degli scostamenti?

30. — Si determini il valore mediano delle seguenti serie cronologiche contenute nelle «Notizie statistiche retrospettive» dell'ASI: numero assoluto dei nati vivi, produzione di minerale di ferro, produzione di zolfo, importazione di caffè, esportazione di seta tratta greggia. È adatta la mediana a rappresentare sinteticamente taluna di queste serie?

31. — Sui dati dell'ASI, capitolo «Credito e previdenza» si determinino l'età mediana, l'età media aritmetica, l'età più frequente delle operaie puerpere sussidiate dalla Cassa nazionale per le assicurazioni sociali. Si esamini il significato e l'attitudine rappresentativa di ciascuna di queste medie. Si calcoli la somma dei valori assoluti degli scostamenti rispetto a ciascuna di esse. Si calcoli la somma dei quadrati degli scostamenti rispetto alle medie stesse: si esegua anzitutto la somma dei quadrati degli scostamenti rispetto alla media aritmetica; da questa si desumano le altre.

32. — In quali casi la media aritmetica va preferita ad ogni altra?

33. — Qual è la media che si può determinare più rapidamente?

34. — Quale media s'impiegherà se si vuole una media sensibile ad ogni variazione di ogni termine della serie?

35. — Si determini l'età mediana delle navi a propulsione meccanica costituenti la marina mercantile italiana (v. ASI), prima partendo dai dati sul numero delle navi e poi da quelli sul tonnellaggio. Si spieghi il divario tra i due risultati. Quale di essi è preferibile? Si esamini se si può calcolare con precisione anche l'età media aritmetica. Si può calcolarla in via approssimativa? Si può determinare una età normale per le navi stesse?

36. — Si scriva la formola generale del genere di medie studiato nel paragrafo 11. Si indichi come in pratica si dovrà procedere nel calcolo. Si calcoli il valore di tale media successivamente per $m = 1, 2, 3, 4, 5$, per la serie costituita dai primi venti numeri interi positivi (da 1 a 20).

37. — Che cos'è il valore divisorio? Sapreste indicarne qualche opportuna applicazione?

38. — Si calcolino la media aritmetica, la media geometrica, la media quadratica, la media armonica della seguente serie di temperature, espresse in gradi centigradi: 3, 2, 1, 3, 0, — 1, 0, 2, 4, 3.

39. — Calcolando la media aritmetica dei logaritmi dei termini di una serie si ottiene la media geometrica della serie?

40. — Si calcolino le medie geometriche dei prezzi del pane nelle principali città italiane al 1° di ciascun mese del 1929 (ASI, capitolo «Grandi città»). Si calcolino le medie aritmetiche e le mediane degli stessi prezzi. Perchè la media geometrica risulta solo di poco differente dalla media aritmetica e dalla mediana?

41. — Si esprima in una formola generale la media geometrica. Si dimostri che il reciproco della media geometrica è uguale alla media geometrica dei reciproci.

42. — Si determini il valore equidistante dagli estremi, per la nuzialità, la natalità, la mortalità nelle varie provincie italiane. Lo si confronti con la media aritmetica (media per l'Italia, indicata nell'ASI), con la mediana, coi valore più frequente. Qual è il significato dei divari fra queste medie? Quale di esse ha migliore attitudine rappresentativa?

43. — Si può calcolare con buona approssimazione la statura normale degli iscritti nelle liste di leva, dai dati riferiti nell'ASI? la statura media aritmetica? la mediana?

44. — Si calcoli la media cubica di ciascuna delle due seguenti serie di dati:

a) — 7, — 7, — 5, — 3, — 2, — 2, — 1, 2, 4, 5, 5, 6, 8, 12, 17, 19, 21, 21, 22, 27.

b) — 27, — 22, — 21, — 21, — 19, — 17, — 12, — 8, — 6, — 5, — 5, — 4, — 2, 1, 2, 2, 3, 5, 7, 7.

45. — Si calcoli la media quadratica di ciascuna delle due precedenti serie; si calcoli la media aritmetica di ciascuna di esse. Quali considerazioni suggerisce la comparazione fra i risultati dei tre calcoli? (compreso quello dell'esercizio 44).

46. — Che cos'è la media armonica? Quali proprietà ha?

47. — Qual è l'origine delle denominazioni delle medie aritmetica, armonica e geometrica?

48. — Si provi a dimostrare che il valore divisorio rende minima la somma dei valori assoluti degli scostamenti relativi dei termini dalla media.

49. — Partendo dai dati dell'ASI sulla distribuzione dei comuni italiani, o di quelli d'una singola regione, secondo il numero degli abitanti, si determini la popolazione media di ciascun comune: *a*) media aritmetica, *b*) valore mediano, *c*) valore più frequente, *d*) valore divisorio.

50. — In che differisce una media ponderata dalla corrispondente media semplice? Come si definisce il «peso»? il peso relativo?

51. — Partendo dai dati dell'ASI sulla produzione del frumento per ettaro nelle varie regioni (compartimenti), si calcoli quella media ponderata di essi che coincide con la produzione del frumento per ettaro in Italia. Oppure, partendo dai dati delle varie provincie, si esegua il calcolo per una regione. Si calcolino i pesi relativi.

52. — I «pesi» sono sempre rappresentati da numeri interi?

53. — Partendo dai dati dell'ASI sull'età dei colpiti da infortuni industriali, si calcoli l'età media aritmetica ponderata.

54. — Si raggruppino per valori uguali i dati sull'umidità relativa dell'aria nei 12 mesi dell'anno in 12 osservatorii italiani, raccolti nell'ASI, capitolo « Climatologia » : se ne determinino poi la media aritmetica ponderata, la media geometrica ponderata, la media armonica ponderata, la media quadratica ponderata, il valore mediano, il valore normale. Per il calcolo della media armonica si adoperi una tavola dei reciproci, per quello della media quadratica una tavola dei quadrati.

55. — Si esprima in una formola generale : *a*) la media aritmetica ponderata, *b*) la media quadratica ponderata, *c*) la media armonica ponderata, *d*) la media geometrica ponderata.

56. — Si calcoli il capitale medio aritmetico delle società per azioni in ciascuna categoria di attività (v. ASI, capitolo « Mercato monetario e finanziario »); e poi il capitale medio aritmetico ponderato per l'insieme di tutte le categorie.

57. — Si dimostri che quando i pesi dei vari valori sono tutti uguali tra loro, pur differendo dall'unità, la determinazione di una media ponderata (aritmetica, geometrica, quadratica, armonica, ecc.) si può ricondurre alla determinazione di una media semplice.

58. — Si cerchi di esprimere in una formola l'errore assoluto, e poi l'errore relativo, di una media aritmetica ponderata, in funzione dei valori esatti dei termini e dei pesi relativi, e degli errori relativi dei termini e dei pesi relativi. Si interpreti poi questa formola, chiarendo l'influenza degli errori relativi dei termini, e di quelli dei pesi relativi, nella formazione della media.

59. — Quando è lecito affermare che gli errori dei termini tendono ad eliminarsi reciprocamente nella formazione della media?

60. — Per avere un'idea dell'influenza di variazioni nei termini e nei pesi sulla formazione di una media ponderata si calcoli la media aritmetica degli stipendi corrisposti da ciascuna di quattro aziende, delle quali la prima ha 100 impiegati di prima classe con 48.000 lire di stipendio, 100 di seconda con 24.000 lire, 100 di terza con 12.000 lire ; la seconda ha 100 impiegati a 64.000 lire, 100 a 32.000 lire, 100 a 16.000 lire ; la terza ha 100 impiegati a 48.000 lire. 400 a 24.000 lire, 800 a 12.000 lire : la quarta ha 100 impiegati a 64.000 lire. 400 a 32.000 lire, 800 a 16.000 lire. Si noti, fra altro, che mentre nelle singole classi gli stipendi sono maggiori nella quarta azienda che nella prima, lo stipendio medio risulta minore.

61. — Si calcoli la media aritmetica delle eccedenze dei nati sui morti accertate in Italia nei 10 anni dal 1901 al 1910 e di quelle accertate nei 10 anni dal 1920 al 1929, con l'artificio indicato nel secondo comma del paragrafo 15, partendo per esempio dal numero di 300.000.

62. — Si calcolino l'età media aritmetica, l'età mediana e l'età normale degli uomini che hanno contratto matrimonio in Italia (« sposi » nell'ASI). Si rifletta bene sulle ipotesi da adottare, nel calcolo della media aritmetica, per l'età media in ciascun gruppo poliennale di età.

63. — Si esegua un analogo calcolo a quello del precedente esercizio, per le spose.

64. — Dividendo la somma delle m^{me} potenze dei termini di una serie statistica per la somma delle $(m-1)^{me}$ potenze di esse. si ottiene una media. Lo si dimostri. Per $m=2$ si ottiene la così detta media *antiarmonica*. Quali

medie si ottengono per $m = 1$, per $m = 0$? Si calcolino col metodo ora indicato le medie dei primi 20 numeri interi, per $m = 1, 2, 3, 4$; si verifichi che al crescere di m cresce la media così ottenuta.

65. — Quali sono i criteri fondamentali da seguire nella scelta fra le varie medie? Esiste una media che meriti di essere sempre preferita a tutte le altre?

CAPITOLO X.

Rappresentazione sintetico-analitica della serie statistica: i dati sussidiari alla media.

Il difetto di qualsiasi media: dissimulare le differenze fra i termini della serie — Il possibile rimedio: impiego di dati sussidiari alla media — I due termini estremi — Uso simultaneo di più medie della stessa specie — Uso simultaneo di più medie di diversa specie — Medie degli scostamenti da una media — Indici di precisione — Medie delle differenze tra i termini.

1. — La media soggettiva, qualunque essa sia, che si sceglie per rappresentare una serie statistica, ha l'insanabile difetto che a valori tra loro differenti sostituisce un unico valore, escludendo così la possibilità di mettere in rilievo le variazioni, le tendenze, le disuguaglianze, che si manifestano attraverso le differenze tra i vari termini.

Se le differenze fra i termini sono piccole, la media può servir bene, in molti casi — come abbiamo avvertito fin dal principio del precedente capitolo — a rappresentare la serie statistica, e può bastare da sola a tale intento. La regola non è senza eccezioni: in certe serie statistiche, come in quella delle temperature di un ammalato il cui stato morboso è caratterizzato da lievissime variazioni di temperatura che si ripetono ad intervalli regolari, una rappresentazione puramente sintetica, come quella data da una media, dissimula proprio le caratteristiche per conoscere le quali si è eseguita la rilevazione. In generale si può dire che la media offre una inadeguata rappresentazione della serie quando differenze relativamente piccole fra i termini possono essere rilevanti ai fini dell'indagine.

Se le differenze fra i termini sono grandi, la media può riuscire bensì utile a rappresentare in via di prima approssimazione la serie statistica, ma non è mai sufficiente a rappresentarla in modo completo. Se sappiamo che in una data provincia il reddito medio aritmetico dei capi famiglia è di 12.000 lire, abbiamo certamente una nozione preferibile all'ignoranza assoluta, ma ci sfugge ogni cognizione delle disuguaglianze caratteristiche della distribuzione della ricchezza, che sono espresse nella serie dei singoli redditi.

Appunto per mettere in risalto le disuguaglianze esistenti fra i termini d'una serie statistica, si uniscono talvolta ad una media altri dati che mirano ad offrire una visione sintetica di tali disuguaglianze: tali sono i due termini estremi della serie, la cui comparazione indica la massima differenza esistente fra i termini della serie. le medie di differenze fra i termini, le medie di scostamenti. Anche con l'uso simultaneo di più medie si può cercare di dare risalto alle disuguaglianze esistenti fra i termini.

Esamineremo successivamente queste varie categorie di dati sussidiari alle medie.

2. — I due termini estremi della serie — estremi, s'intende, in ordine di grandezza — delimitano, come sappiamo, il campo di variazione, e perciò posti accanto alla media ci mostrano i limiti dell'errore che si commette coll'assumere la media stessa a rappresentare i singoli termini della serie. Quando questi limiti sono bassi, in molti casi la media ha sufficiente attitudine rappresentativa, e viceversa. Se sappiamo che la mediana dei prezzi praticati dai vari fornai di Milano per una data qualità di pane è di lire 2, il minimo prezzo è di 1,90, e il massimo di 2,10, vediamo subito che la media è buono strumento di sintesi, quale non sarebbe se, per esempio, il minimo fosse di 1 lira e il massimo di 5 lire.

Una medesima ampiezza del campo di variazione dei termini può avere rilevanza molto differente secondo la grandezza dei termini stessi: per esempio le 4 lire di scarto fra il prezzo minimo e il prezzo massimo indicano una grande disuguaglianza di prezzi nel precedente esempio in cui il minimo è di lire 1, il massimo di 5 e la mediana di 2; ma indicherebbero una minuscola disuguaglianza se il minimo fosse di 998, il massimo di 1002 e la mediana di 1000. Per avere una misura dell'ampiezza relativa del campo di variazione, si può mettere in rapporto l'ampiezza assoluta di esso, cioè la differenza fra il termine massimo e il termine minimo, con la media dei termini (rapporto 4: 2 nel primo esempio, 4: 1000 nel secondo). Diviso per 2, il numeratore di tal rapporto rappresenta la media aritmetica dei valori assoluti dei due massimi scostamenti in eccesso e in difetto, di modo che il rapporto così modificato viene a costituire una misura relativa dei limiti dell'errore che s'incontra surrogando la media ai termini.

Talvolta i termini estremi di una serie rappresentano anomalie od eccezioni, così che hanno scarso valore al fine della rappresentazione dell'intera serie. Quasi in ogni classe di leva s'incontra qualche nano dell'altezza di 75 o 80 cm. e qualche gigante di 210 o 220 cm., ma queste stature singolari sono ben lontane dalla massima parte di quelle degli altri giovani misurati; servono bensì ad indicare i limiti estremi fra i quali varia la statura, ma si attribuisce loro soverchia importanza assumendole, insieme con una media, a rappresentare sinteticamente l'intera serie delle stature.

Anche prescindendo dai casi nei quali i valori estremi della serie rappresentano misure assolutamente eccezionali nella serie stessa, l'indicazione dei valori stessi riesce in generale insufficiente a dare un' idea, anche largamente approssimativa, della distribuzione dei termini nel campo di variazione. Tale insufficienza, che appare palese al ragionamento ed all'esperienza, può essere resa intuitiva mercè il riferimento alla rappresentazione grafica.

Immaginiamo di avere rappresentato graficamente i termini di una serie statistica, disposti in ordine di grandezza, mediante altrettanti segmenti proporzionali ad essi, che avremo disposti verticalmente, paralleli ed equidistanti fra loro, col piede sopra una stessa retta orizzontale. Uniamo ora con una spezzata i vertici dei segmenti: otterremo un'immagine grafica dell'andamento della serie. Se per rappresentare in numeri la serie noi prendiamo la mediana ed i due termini estremi, ciò equivale a prendere nel grafico il segmento mediano e i due segmenti estremi: proviamo a riunire i vertici di questi tre soli segmenti, e otterremo una spezzata che per lo più sarà insufficiente a rappresentare con buona approssimazione l'altra che descrive in modo completo l'andamento della serie. Se invece di conoscere tre soli punti di quest'ultima spezzata ne conoscessimo un maggior numero, avremmo una rappresentazione maggiormente approssimata dell'andamento della serie.

3. — Sorge così l'idea di assumere a rappresentare la serie, accanto al valore mediano ed ai termini estremi, qualche altro valore intermedio fra il minimo ed il mediano, fra il mediano e il massimo. Poichè il valore mediano divide la serie ordinata per grandezza in due semi-serie composte di ugual numero di termini, la più semplice attuazione di questa idea consisterà nel determinare il valore mediano della prima semi-serie (detto *primo quartile* perchè preceduto in ordine di grandezza da un quarto del numero totale dei termini) e il valore mediano della seconda semi-serie (detto

terzo quartile perchè preceduto in ordine di grandezza da tre quarti del numero totale dei termini). Avremo ora, a rappresentare la serie, i tre quartili (il secondo di essi è la mediana) e i due estremi: cinque punti della spezzata, invece dei tre di poc'anzi, e quindi una rappresentazione maggiormente approssimata. Ma possiamo far di meglio: determinare addirittura i *decili* (primo decile è il valore preceduto in ordine di grandezza da un decimo del numero totale dei termini, secondo decile quello preceduto da due decimi, ecc.); così avremo non solo cinque ma ben undici punti della spezzata. E perchè fermarci qui? Specialmente se la serie è molto numerosa, ci converrà determinare i *centili*, in modo da conoscere centouno punti della spezzata; nè sono ancora queste le colonne d'Ercole, se i dati della serie sono molto numerosi. Di mano in mano la nostra rappresentazione ridotta della serie va guadagnando in approssimazione.

Però l'entusiasmo per la feconda idea che ci ha condotti a tale risultato non deve dissimularci gli inconvenienti di questo metodo sintetico-analitico. La suprema virtù della media è quella di racchiudere in un sol numero la rappresentazione d'una successione di numeri che può comprendere milioni di termini. Di mano in mano che si accompagnano ad essa dati sussidiari in numero crescente, cresce l'approssimazione, ma svanisce la semplicità della descrizione, anzi si passa gradualmente dal metodo di rappresentazione che abbiamo chiamato sintetico-analitico perchè mira ancora principalmente alla sintesi, a metodi sempre più vicini a quelli analitico-sintetici, che mirano principalmente a conservare la visione dei particolari nella visione dell'insieme. I confronti tra più serie divengono a grado a grado più laboriosi e meno concludenti: se ci contentiamo di confrontare le mediane delle varie serie, giungiamo a risultati univoci; se confrontiamo i quartili, non sempre arriveremo a risultati concordanti nei vari confronti; e se confrontiamo i decili o i centili, ciò accadrà più di rado. Perciò in pratica spesso si usano i quartili, raramente i decili e quasi mai i centili.

Per esporre il metodo dell'uso simultaneo di più medie della stessa specie, ci siamo riferiti alla mediana. Ma è ovvio che altre medie si prestano all'applicazione dello stesso metodo (non, però, il valore più frequente, per ovvia ragione). Per esempio, se si parte dalla media aritmetica M, si possono successivamente porle accanto la media aritmetica M_1 dei termini ad essa non superiori e la media aritmetica M_3 dei termini ad essa superiori; e poi la media aritmetica dei termini non superiori a M_1, quella dei termini compresi fra M_1 e M, quella dei termini compresi fra M e M_3, quella dei termini superiori a M_3; e così via con criterio analogo. Oppure si può determinare la media aritmetica della prima metà dei termini in ordine di grandezza, e quella della seconda metà; e poi la media aritmetica del primo quarto, sempre in ordine di grandezza del numero totale dei termini, quella del secondo quarto, del terzo, dell'ultimo; e così via con analogo criterio. E se si parte dal valore divisorio, si può accompagnarlo col valore per cui la somma dei termini precedenti in ordine di grandezza è un quarto della somma di tutti i termini, e con quello per cui la somma dei termini precedenti è tre quarti della somma di tutti i termini, ecc. Non vale la pena di moltiplicare gli esempi, nè occorre ripetere quale sia l'inconveniente del metodo.

4. — L'uso simultaneo di medie di diversa specie, della stessa serie (per esempio: mediana, media aritmetica, valore più frequente, valore divisorio), certamente può mettere in rilievo diversi aspetti del fenomeno rappresentato, e in qualche caso particolare può rivelare allo statistico esperto notevoli caratteri della distribuzione dei termini per grandezza; ma in generale offre soltanto varie rappresentazioni sintetiche della serie, che mal suppliscono all'assenza di una sola rappresentazione analitica.

5. — Nella rappresentazione grafica supposta nel precedente paragrafo 2, a ciascuna media corrisponde una retta orizzontale che interseca, più su o più giù secondo la media che si sceglie, la spezzata che delinea l'andamento della serie. Alle varie medie corrispondono dunque altrettanti livelli costanti, cioè altrettante ipotesi di uguaglianza; gli scostamenti misurano le deviazioni dei singoli termini da codesti livelli, cioè le deviazioni dagli ipotetici stati di uguaglianza. Se i termini della serie differiscono poco tra loro, sono piccole le deviazioni di essi da un qualsiasi livello medio; se differiscono molto tra loro, sono grandi, almeno in parte, le loro differenze da qualsiasi media. Una funzione degli scostamenti (per esempio la somma dei loro valori assoluti, o una media di essi) può servire come indice delle disuguaglianze esistenti fra i vari termini.

Non è indifferente la scelta di un livello o dell'altro come riferimento per la misura delle deviazioni dei termini della serie da un livello costante; non è indifferente, cioè, calcolare gli scostamenti rispetto ad una o ad un'altra media. Per esempio, sappiamo già (capitolo IX, paragrafo 5) che la somma dei valori assoluti degli scostamenti, partendo da un massimo quando la media coincide col termine minimo della serie, scende gradualmente col crescere della media fino a toccare un minimo quand'essa coincide col valore mediano, e poi gradualmente risale fino a raggiungere un altro massimo quando la media coincide col termine massimo della serie. E sappiamo (capitolo IX, paragrafo 6) che la somma dei quadrati degli scostamenti ha un andamento in tutto analogo, salvo che il minimo di essa corrisponde alla media aritmetica. Sarebbe facile ripetere simili considerazioni rispetto ad altre funzioni degli scostamenti.

Uno dei criteri che si possono seguire nella scelta della funzione degli scostamenti da assumere ad indice di disuguaglianza fra i termini della serie consiste nell'accompagnare a ciascuna media quella funzione degli scostamenti che essa rende minima: alla mediana la media aritmetica dei valori assoluti degli scostamenti, alla media aritmetica la media quadratica degli scostamenti, ecc. Il criterio si può applicare soltanto a quelle medie che hanno la proprietà di rendere minima una funzione degli scostamenti, e si applica utilmente soltanto a quelle poche medie per cui tale funzione è semplice e di significato intuitivo.

Un criterio più generale, e più razionale in quanto permette di confrontare il grado di approssimazione col quale varie medie rappresentano la stessa serie, è quello di accompagnare a qualsiasi media una medesima funzione degli scostamenti: la media aritmetica oppure la mediana dei valori assoluti di essi. La prima viene brevemente designata come *scostamento medio assoluto*; la seconda come *scostamento mediano* (è molto usata anche la denominazione di *scostamento*, *scarto* od *errore probabile*). È largamente impiegato, spesso senza alcuna plausibile ragione di preferenza rispetto alle più semplici funzioni or ora indicate, lo *scostamento medio quadratico* (talvolta detto *deviazione tipo*: nomenclatura esotica che conviene conoscere ma non adottare). Le medie degli scostamenti hanno il vantaggio e lo svantaggio caratteristici di ogni media: riassumono nel modo più breve possibile l'intera serie degli scostamenti, ma li riducono tutti ad un unico livello, impedendo così la percezione delle differenze che esistono tra essi; danno una misura sintetica della distribuzione o *dispersione* dei termini intorno alla media, ma non presentano una visione completa di tale dispersione.

Quando si accompagna uno scostamento medio alla media dei termini, si suole farlo precedere dal segno $\pm$ (per esempio, statura media cm. 166 ± 4; reddito medio lire 10.000 ± 3.750), ad indicare che si tratta di una media dei valori assoluti di grandezze che in parte sono positive, in parte sono negative. In certi casi può convenire di calcolare separatamente una media degli scostamenti positivi ed una degli scostamenti negativi: così specialmente quando tali due medie sono molto differenti tra loro (come normalmente in una distribuzione di redditi). Il valore mediano degli scostamenti positivi dei termini rispetto alla mediana è uguale alla differenza fra il terzo quartile e la mediana, quello degli scostamenti negativi è uguale alla differenza fra la mediana e il primo quartile.

Per eseguire rapidamente il calcolo dello scostamento medio assoluto, quando non occorra conoscere i singoli scostamenti, si tenga presente che, per una media compresa fra il k^{mo} e il $(k+1)^{mo}$ termine in ordine di grandezza, la somma dei valori assoluti degli scostamenti è data dalla somma degli $(n-k)$ termini susseguenti alla media, diminuita della somma dei k termini precedenti ad essa, ed aumentata del prodotto della media per $(2k-n)$. Nel caso della mediana, quest'ultimo termine si annulla essendo $2k=n$; nel caso della media aritmetica, essendo il valore assoluto della somma degli scostamenti negativi uguale alla somma degli scostamenti positivi, per ottenere la somma dei valori assoluti di tutti gli scostamenti basta sottrarre la somma dei k termini precedenti alla media dal prodotto di k per la media, e raddoppiare il numero ottenuto.

6. — Una media degli scostamenti dà una visione sintetica delle deviazioni dei termini da quel livello d'uguaglianza che la media dei termini stessi rappresenta; ma una data deviazione media ha rilevanza maggiore o minore secondo che è meno o più grande la media dei termini. In un paese il reddito medio aritmetico è di lire 10.000, con uno scostamento medio assoluto di 2.000; in un altro paese il reddito medio è di 40.000, lo scostamento medio di 2.000. Lo stesso scostamento medio indica una disuguaglianza relativa dei redditi molto maggiore nel primo paese che nel secondo. Per metterlo in risalto, è ovvia l'idea di porre in rapporto lo scostamento medio con la media alla quale esso si riferisce: nel caso nostro diremo che nel primo paese lo scostamento medio è del 20 %, nel secondo del 5 %. Rapporti di questo tipo, tra una media degli scostamenti e una media dei termini, si chiamano *coeffi-*

cienti di variazione (meglio che *coefficienti di variabilità*, denominazione comunemente adottata).

I coefficienti di variazione si possono riguardare come medie di scostamenti relativi *dei termini dalla media*, considerati in valore assoluto; perciò sono teoricamente adatti per indicare il grado di approssimazione di più misure di una stessa grandezza rispetto alla media oggettiva, assunta come vero valore di tale grandezza. Quando si tratta di medie soggettive, come nella massima parte delle indagini statistiche, sarebbe teoricamente più opportuno adoperare medie di scostamenti relativi *della media dai termini* (sempre considerati in valore assoluto) per indicare il grado di approssimazione col quale la media, qui pura astrazione, rappresenta i termini che misurano grandezze concrete. Per una ragione formale, ma decisiva — la laboriosità dei calcoli che occorrerebbero per applicare questo secondo procedimento — si applica, in generale, il primo procedimento anche nel caso di medie soggettive.

Quando i termini della serie rappresentano quantità che potrebbero, almeno teoricamente, essere ridistribuite, la media aritmetica dei termini acquista speciale importanza, come sappiamo, perchè indica quell'unico livello di uguaglianza che sarebbe concretamente possibile (per esempio il patrimonio che toccherebbe a ciascuno se la somma dei patrimoni esistenti venisse ridistribuita fra tutti in parti uguali). Se si definisce disuguaglianza ogni deviazione da codesto livello di uguaglianza, si può dire che lo scostamento medio assoluto, essendo direttamente proporzionale alla somma delle disuguaglianze, costituisca la migliore misura della disuguaglianza media, e il coefficiente di variazione che se ne desume offra la migliore misura della disuguaglianza media relativa.

Avvertasi che il coefficiente di variazione desunto dallo scostamento medio assoluto rispetto alla media aritmetica non può superare, com'è facile verificare, il valore $(2n - 2) : n$, limite che, quando n è un grande numero, praticamente coincide con 2. È utile conoscere tale limite per poter interpretare il significato del valore che il coefficiente di variazione assume in ciascun caso concreto.

7. — Poichè le medie degli scostamenti ed i coefficienti di variazione che se ne possono desumere variano in senso inverso a quello dell'approssimazione con cui la media rappresenta i termini della serie, il reciproco di una media degli scostamenti o di un coefficiente di variazione può servire come indice dell'approssimazione, variando nello stesso senso di essa. Ma un siffatto indice dell'ap-

prossimazione o *precisione* con la quale la media rappresenta i termini (o con la quale i termini rappresentano la media se si tratta d'una media oggettiva) non è facilmente interpretabile, avendo un campo di variabilità illimitato. Perciò in pratica si preferiscono gli indici studiati nei paragrafi 5 e 6, che possono considerarsi come indici inversi dell'approssimazione ed hanno un significato proprio ben chiaro.

8. — Le funzioni degli scostamenti sono indici, non misure dirette, delle disuguaglianze esistenti fra i vari termini di una serie. Così lo scostamento medio assoluto dal reddito medio aritmetico misura la deviazione media dei singoli redditi individuali da un ipotetico stato di uguaglianza, ma non misura direttamente la disuguaglianza media fra i vari redditi: se vogliamo misurare questa, dobbiamo partire dalle disuguaglianze, cioè dalle differenze, tra i redditi comparati tra loro a due a due. Dati n redditi individuali $u_1, u_2, \ldots\ldots, u_n$, che supponiamo, al solito, numerati in ordine crescente di grandezza, potremo comparare ciascuno di essi con ciascuno dei rimanenti $(n-1)$; otterremo quindi in complesso $n\,(n-1)$ differenze, le quali sono a due a due uguali in valore assoluto ma differenti per il segno, poichè ciascuna comparazione è eseguita due volte, nei due possibili sensi: per esempio $(u_k - u_l)$ e $(u_l - u_k)$. Una media di queste differenze, considerate in valore assoluto perchè altrimenti si compenserebbero reciprocamente a due a due, si può assumere come espressione sintetica delle disuguaglianze esistenti fra i redditi nel nostro esempio, e fra i termini della serie in generale. Avvertasi che qui si parte da una definizione di disuguaglianza diversa da quella adottata nel penultimo comma del paragrafo 6: là si definiva disuguaglianza la deviazione di ciascun termine da un livello medio, qui si definisce disuguaglianza la deviazione di ciascun termine da ciascun altro. Sono due concezioni non contraddittorie ma neppure coincidenti, che corrispondono a due dei significati che si attribuiscono all'espressione disuguaglianza nell'uso comune: non si può dire che l'una definizione vada sempre preferita all'altra, anzi l'esperienza insegna che secondo i fini dell'indagine può apparire più opportuno definire la disuguaglianza ora nell'uno ora nell'altro modo.

Se per calcolare una media delle $n\,(n-1)$ differenze fosse indispensabile calcolare prima tutte le differenze stesse, riuscirebbe estremamente laborioso l'impiego di questo metodo appena n superasse poche decine, e in pratica esso non verrebbe quasi mai adoperato. Ma nè per calcolare la somma dei valori assoluti delle differenze, la quale divisa per $n\,(n-1)$ dà la *differenza media assoluta*, nè per calcolare la somma dei quadrati delle differenze, la quale mediante divisione per $n(n-1)$ e successiva estrazione di radice quadrata dal quoziente ottenuto dà la *differenza media quadratica*, occorre il previo computo delle singole differenze. Esso occorre, invece, per il computo della *differenza mediana*, la quale perciò non viene usata.

La somma dei valori assoluti delle differenze si può ottenere abbastanza rapidamente, col sussidio di una macchina calcolatrice, eseguendo i prodotti $i\,u_i$, per tutti i valori di i da 1 ad n (nel nostro esempio: il prodotto di ciascun reddito individuale per il numero del posto che esso occupa nella serie dei redditi individuali disposti in ordine crescente di grandezza); quadruplicando la somma di tali prodotti: e infine detraendo dal risultato il prodotto della somma dei termini per $2\,(n+1)$. Oppure, se si sono computati gli scostamenti dei termini dalla loro media aritmetica M, basta eseguire i prodotti $i\,(u_i - M)$ e quadruplicarne la somma algebrica per ottenere la somma dei valori assoluti delle differenze.

La somma dei quadrati delle differenze risulta uguale al prodotto per $2\,n$ della somma dei quadrati dei termini, diminuito del doppio del quadrato della somma dei termini: se i termini sono numeri interi non molto grandi, col sussidio di una tavola dei quadrati il calcolo si può eseguire abbastanza rapidamente. È facile verificare che la somma dei quadrati delle differenze è uguale al prodotto per $2\,n$ della somma dei quadrati degli scostamenti dalla media aritmetica. Perciò se si conosce lo scostamento medio quadratico dalla media aritmetica, basta moltiplicarlo per la radice quadrata del quoziente $2\,n : (n-1)$, che quando n non è un piccolo numero è espressa con sufficiente approssimazione dal valore 1,414, per ottenere la differenza media quadratica. Data l'esistenza di un rapporto praticamente costante fra la differenza media quadratica e lo scostamento medio quadratico dalla media aritmetica, non v'è ragione di preferire l'una espressione all'altra, più rapidamente calcolabile.

Invece la differenza media assoluta, diretta misura sintetica delle disuguaglianze fra i termini, meriterebbe di essere introdotta nell'uso. La grande semplicità del calcolo dello scostamento medio assoluto dalla media aritmetica è causa che esso le venga spesso preferito.

9. — Una media delle differenze offre una misura riassuntiva delle disuguaglianze esistenti fra i termini della serie. Essa può venir usata indipendentemente da qualsiasi media (al contrario di

una media degli scostamenti che acquista preciso significato soltanto accompagnata alla media rispetto alla quale sono stati computati gli scostamenti). Però se si vuol giudicare la rilevanza delle disuguaglianze esistenti fra i termini conviene mettere in rapporto la differenza media assoluta con la media aritmetica dei termini: il valore di questo rapporto varia entro limiti strettamente definiti, non potendo superare 2, così che è adatto a servire come indice del grado di disuguaglianza fra i termini. Se in un paese il reddito medio aritmetico è di 10.000 lire, la differenza media assoluta tra i redditi di 4.000 lire, il rapporto fra questa e quello ci indica che la differenza media assoluta corrisponde al 40 per cento della media aritmetica dei redditi. Se Tizio ha un patrimonio di 10.000 lire, Sempronio di 0, il rapporto tra la differenza media assoluta 10.000 e il patrimonio medio aritmetico 5.000 indica che la differenza media assoluta corrisponde al 200 per 100 della media aritmetica dei patrimoni (è questo un esempio della massima possibile disuguaglianza).

10. — Mercè l'impiego coordinato delle medie e dei dati sussidiari che abbiamo imparato a conoscere, è possibile in molti casi ottenere una sufficiente rappresentazione sintetico-analitica della serie statistica. In altri casi questi metodi offrono bensì una sufficiente visione sintetica del fenomeno considerato, ma non ne presentano una adeguata visione analitica. Si prenda, per esempio, una serie cronologica che mostri una netta e vivace tendenza attraverso il tempo — quella della produzione di energia idroelettrica in Italia, o quella dell'importazione di cotone greggio — e si provi a rappresentarla per mezzo di medie e di dati sussidiari. Nemmeno col moltiplicare questi ultimi (ricorrendo per esempio ai decili) si ottiene una rappresentazione soddisfacente, perchè in questo caso al fine di un giudizio sull'andamento del fenomeno non importa soltanto l'ordine dei termini secondo la grandezza ma anche il loro ordine cronologico. Occorrono metodi di rappresentazione coi quali quest'ordine non venga sconvolto, anzi venga mantenuto e messo in risalto. Si prenda, poi, una serie geografica, per esempio quella della natalità nelle varie provincie italiane. Neppur qui le medie e i dati sussidiari danno risultati soddisfacenti, perchè distruggono l'ordine geografico dei termini della serie: occorrono metodi di rappresentazione atti a conservarlo ed a porlo in evidenza. Nè solo per questi tipi di serie la media, pur corredata di dati sussidiari,

appare insufficiente per la rappresentazione; anche altre serie di ogni tipo richiedono spesso metodi più adatti a porgere, insieme con la visione d'insieme, una visione dei particolari. Fra tali metodi emergono i procedimenti grafici, che studieremo nel prossimo capitolo.

Quesiti ed esercizi: 1. — Qual è il difetto irreparabile della media cui si cerca di rimediare mediante l'impiego di dati sussidiari? Quali sono i principali dati sussidiari a tal uopo impiegati?

2. — Qual è il modo più adatto per mettere in evidenza, sinteticamente, le disuguaglianze esistenti fra i termini di una serie statistica?

3. — Considerando i prezzi al minuto di singoli generi alimentari in diverse città italiane ad un stessa data (V. ASI, capitolo « Grandi città ») si ricerchi se l'indicazione dei termini estremi accanto alla media basti a dare adeguata idea delle disuguaglianze di prezzi esistenti fra le varie città.

4. — La serie dei dati sul consumo medio per abitante del tabacco nelle varie regioni italiane può essere adeguatamente rappresentata da una media e dai termini estremi?

5. — Serie cronologiche come quelle raccolte nelle « Notizie statistiche retrospettive » dell'ASI possono essere ben rappresentate da una media e dai termini estremi in ordine di grandezza? o vanno preferiti a questi gli estremi in ordine di tempo?

6. — Si esegua per una di tali serie la rappresentazione grafica indicata nel paragrafo 2 del testo, con le applicazioni ivi esposte (mediana, quartili, ecc.).

7. — Serie cronologiche possono essere abbreviate e quindi rese più facilmente comparabili con altre, coll'estendere l'intervallo cui ciascun dato si riferisce (per esempio col considerare quinquennii invece che anni), attribuendo poi all'intervallo più ampio la somma o una media dei dati riferentisi agli intervalli più brevi ond'esso è costituito (per esempio: somma quinquennale, o media annua del quinquennio). Si applichi il procedimento alle serie cronologiche retrospettive dell'ASI, e se ne ricerchino attraverso l'esperienza i vantaggi e gli svantaggi.

8. — Utilizzando i dati sulla portata media, massima e minima dei vari fiumi italiani (V. ASI, capitolo « Idrometria »), si mettano in risalto le differenti caratteristiche di essi, comparando l'ampiezza del campo di variazione della portata con la portata media.

9. — Per la serie delle età delle puerpere sussidiate dalla Cassa nazionale per le assicurazioni sociali (V. ASI, capitolo « Credito e previdenza »), si determinino i quartili e i decili; l'età media aritmetica del primo decimo, del secondo decimo, ecc., delle puerpere.

10. — Si determinino i quartili per la serie delle frequenze delle nascite, delle morti, dei matrimoni, nelle varie provincie italiane. Quale dei tre fenomeni appare maggiormente variabile?

11. — Si determinino i quartili per la serie dei prezzi dell'energia elettrica nelle città italiane (V. ASI, capitolo « Grandi città »).

12. — Considerando una distribuzione di redditi secondo l'ammontare (se ne trovano esempi negli Annuari statistici degli Stati germanici), se ne determino la media aritmetica, la mediana, il valore più frequente, il valore divisorio; e poi la media aritmetica dei redditi inferiori al reddito medio aritmetico e di quelli superiori; il primo e il terzo quartile, i decili; i redditi sopra i quali si trovano nove decimi, otto decimi, ecc., della somma di tutti i redditi.

13. — Data la serie costituita dai numeri interi da 1 a 11, si calcoli la somma dei valori assoluti degli scostamenti dei termini dalle seguenti medie: 2, 4, 6. 8, 10. Si calcoli la stessa somma rispetto alle medie 5,9 e 6.1. Qual è il valore della media che rende minima tal somma? Perchè, per la stessa serie. si trova che la minima somma dei quadrati degli scostamenti corrisponde al valore mediano? Si possono calcolare per la stessa serie gli scostamenti dal valore più frequente?

14. — Che cos'è lo scostamento medio assoluto? mediano? medio quadratico?

15. — Per la serie delle età delle puerpere sussidiate dalla Cassa nazionale per le assicurazioni sociali (V. ASI, capitolo « Credito e previdenza ») si determino: lo scostamento medio assoluto, lo scostamento medio quadratico, lo scostamento mediano.

16. — Si calcoli lo scostamento medio assoluto semplice dei dati sulla frequenza delle nascite nelle varie provincie italiane rispetto alla loro media aritmetica semplice. Volendo calcolare invece lo scostamento medio assoluto ponderato dalla media italiana quale procedimento si dovrebbe seguire?

17. — Si esegua analogo calcolo per la frequenza delle morti. Si calcoli poi, così per la natalità come per la mortalità, il coefficiente di variazione desumibile dallo scostamento medio assoluto.

18. — Quando è data una media M dei termini di una serie e si conosce così la media aritmetica dei termini che seguono in ordine di grandezza la media M come la media aritmetica dei termini che la precedono, si può calcolare rapidamente la media aritmetica degli scostamenti positivi e quella degli scostamenti negativi della media stessa?

19. — Si metta in formola l'espressione della media aritmetica dei valori assoluti degli scostamenti da una media compresa fra il k^{mo} e il $(k+1)^{mo}$ termine (in ordine crescente di grandezza) di una serie statistica.

20. — Che cos'è un « coefficiente di variazione »? Perchè va preferita questa espressione all'altra: « coefficiente di variabilità »?

21. — Si calcolino la mediana, lo scostamento mediano e il coefficiente di variazione per le due serie statistiche che rappresentano la produzione italiana del frumento e l'importazione di frumento in Italia dal 1881 in poi. La media coi dati sussidiari suddetti rappresenta adeguatamente l'andamento dei due fenomeni?

22. — Si calcoli lo scostamento medio assoluto dei prezzi al minuto di un singolo genere di consumo nelle varie città italiane, considerate tutte ad una stessa data, dalla mediana dei prezzi stessi. Ripetuto il calcolo per vari generi si confronti l'ampiezza relativa delle variazioni dei prezzi dei diversi generi, impiegando coefficienti di variazione.

23. — Si ripeta il precedente esercizio, impiegando invece la differenza media assoluta e il rapporto di essa alla media aritmetica.

24. — Data una media M, compresa in ordine crescente di grandezza fra il termine u_k e il termine u_{k+1}, si cerchi di esprimere in una formola generale: *a*) la media aritmetica dei valori assoluti degli scostamenti relativi della media dai termini; *b*) la media aritmetica dei valori assoluti degli scostamenti relativi dei termini dalla media. Considerando poi una breve serie statistica (p. es. prezzi al minuto di un genere di consumo nelle varie città italiane), si calcolino le due medie *a*) e *b*) rispetto alla media aritmetica dei termini.

25. — Si calcoli la differenza media assoluta per ciascuna delle seguenti successioni di numeri interi: da 5 a 7, da 4 a 8, da 3 a 9, da 2 a 10, da 1 a 11. La si confronti con lo scostamento medio assoluto. Si calcoli la differenza media quadratica e la si confronti con lo scostamento medio quadratico.

26. — Si traduca in formola la relazione tra la somma dei quadrati delle differenze tra i termini e la somma dei quadrati degli scostamenti dei termini dalla media aritmetica.

27. — Si ricerchino nell'ASI serie statistiche le quali possano venire rappresentate in modo soddisfacente con medie accompagnate da dati sussidiari ed altre serie per le quali invece tale metodo di rappresentazione riesca insufficiente. Si spieghi perchè si ritiene il metodo adatto, o disadatto, alla rappresentazione di ciascuna delle serie considerate.

CAPITOLO XI.

La rappresentazione analitico-sintetica della serie statistica in forme grafiche.

Scopo delle rappresentazioni grafiche — Procedimenti geometrici e procedimenti cromatici — Vari procedimenti geometrici: estensione comparativa delle loro applicazioni. — Il diagramma: criteri per la scelta della scala; diagramma a canne d'organo; diagramma logaritmico — Rappresentazione grafica delle funzioni statistiche: diagramma in coordinate cartesiane; diagramma a doppia scala logaritmica; criteri per la scelta delle scale; effetti ottici della modificazione delle scale; artifizi ed errori nella costruzione dei diagrammi — Diagramma a gradinata — Impiego di colori o di tratteggi nei diagrammi — Diagramma in coordinate polari — L'istogramma: criteri per la scelta della scala; errori più frequenti nella esecuzione: istogramma ad immagine: impiego di colori o tratteggi negli istogrammi — Lo stereogramma — Il cartogramma: cartogramma a punti; cartogramma ad istogrammi; impiego di colori o tratteggi; artifizi ed errori di esecuzione; cartogramma a stereogrammi in proiezione piana; cartogramma a nastri; cartogramma a diagrammi. Cartogramma a tinte graduali; criteri per la determinazione della corrispondenza fra una successione di tinte e una successione di numeri; errori più comuni — Influenza degli errori di osservazione sulle rap-

presentazioni grafiche — La comparazione fra più serie statistiche mediante rappresentazioni grafiche.

1. — La rappresentazione grafica della serie statistica è una semplice traduzione dei dati numerici e nulla aggiunge ad essi; giova però a rendere possibile la rapida formazione d'impressioni generali sul fenomeno rappresentato, attraverso la visione simultanea degli aspetti particolari di esso che corrispondono ai singoli termini. Le relazioni di grandezza fra i termini sono percepite con celerità e con approssimazione spesso assai maggiori nella rappresentazione grafica, interpretata per intuizione, che nella rappresentazione numerica, interpretata per ragionamento. Si rivelano, anche all'occhio meno esperto, certe regolarità nell'andamento del fenomeno; emergono certe irregolarità, le quali talvolta corrispondono all'azione saltuaria di fattori che vengono così rivelati, talvolta corrispondono ad errori d'osservazione, che sono così scoperti mentre sfuggirebbero forse all'esame dei dati numerici.

Si provi a prendere in esame una lunga serie cronologica di dati statistici, per esempio quella della natalità in Italia negli ultimi cinquant'anni, espressa sia nella forma numerica sia in quella grafica di una linea spezzata: si vedrà quanto più rapide e più precise siano le impressioni desunte da questa seconda rappresentazione. Oppure si consideri da un canto la serie numerica dei dati sulla densità della popolazione nei singoli comuni italiani, dall'altro la rappresentazione dei dati stessi mediante tinte graduali sopra una carta geografica: occorreranno ore e forse giorni di studio per ritrovarsi nel labirinto dei 7.306 dati numerici, mentre un solo sguardo al cartogramma basterà a dare un buon orientamento generale. Certo, se si vuol conoscere un determinato dato (la natalità in un certo anno, o la densità della popolazione in un certo comune), lo si troverà con maggiore prontezza ed esattezza mediante la rappresentazione numerica: nella rappresentazione grafica il singolo dato compare piuttosto come una pennellata in un quadro che come una immagine autonoma. Ma lo scopo della rappresentazione grafica non è tanto quello di mettere in rilievo i particolari per se medesimi quanto quello di farli concorrere alla descrizione dell'insieme cui appartengono. E d'altronde i dati che fortemente divergono da quelli ad essi vicini nell'ordinamento della serie, e che perciò possono più degli altri interessare, risaltano immediatamente nella rappresentazione grafica: così avviene, nell'esempio di poc'anzi, della natalità perturbata nel periodo bellico.

Se per intendere una rappresentazione grafica non basta guardarla, ma occorre meditare lungamente, o rendersi familiari complicate convenzioni, se insomma essa non parla all'occhio senza bisogno d'intermediari, si può dire che non consegua il suo fine principale: quello di porgere rapidamente una visione dei particolari, inquadrati in una visione generale.

2. — Le rappresentazioni grafiche di serie statistiche si valgono di procedimenti geometrici e di procedimenti cromatici.

Far corrispondere a ciascun dato di una serie un segmento di retta di lunghezza proporzionale al dato stesso: tale è il procedimento geometrico col quale si costruisce il *diagramma*. Fargli corrispondere una figura piana di forma determinata (rettangolare, circolare, ecc.), di area proporzionale al dato: ecco un altro procedimento geometrico, col quale si costruisce l'*istogramma*. Fargli corrispondere una figura solida di forma determinata (cubica, sferica, ecc.), di volume proporzionale al dato, costruire cioè uno *stereogramma*, è ancora un procedimento geometrico, il quale diviene procedimento grafico se la figura solida viene rappresentata in proiezione sul piano, come per lo più si usa.

Far corrispondere alla successione dei numeri compresi nel campo di variazione di una serie una successione di gradazioni d'una medesima tinta, in modo che a numeri di mano in mano più grandi corrispondano gradazioni di tinta di mano in mano più intense: è questo un procedimento cromatico. Un analogo effetto di crescente intensità si può ottenere mediante punti di mano in mano più densi (intendasi «punti» nel significato grafico, non in quello geometrico) oppure mediante tratti paralleli od incrociati di mano in mano più fitti. In questi casi le differenze cromatiche indicano differenze quantitative; ma con l'impiego di tinte diverse (scelte cioè in settori non adiacenti della scala cromatica) si fanno servire i colori anche ad indicare differenze qualitative. Avvertiamo una volta per sempre che tra i «colori» al fine della rappresentazione cromatica di serie statistiche, comprendiamo anche il nero, che a rigor di termini non è un colore.

Qualche metodo grafico è dubbio se vada classificato tra i geometrici od i cromatici: tale l'impiego di singoli punti («punti» nel significato grafico, cioè minuscole aree circolari, non in quello geometrico) per rappresentare singoli casi di un fenomeno, o singoli gruppi di *n* casi: procedimento che ha un aspetto geometrico nella proporzionalità tra il numero dei «punti» e quindi tra l'area che essi ricoprono, da un canto, e il dato statistico, dall'altro; ed ha un aspetto cromatico nell'effetto di maggiore o minore intensità di tinta che viene a corrispondere alla maggiore o minore densità dei « punti ».

Altre forme di rappresentazione grafica si ottengono mediante l'associazione dei procedimenti grafici con procedimenti cromatici: associazione cui si aprono le più varie possibilità di combinazioni. Ripetiamo però l'avvertenza già esposta: quanto più si complica la rappresentazione grafica, tanto meno essa giova al suo fine: le forme di rappresentazione che sono impiegate più largamente e più efficacemente nella pratica sono le più semplici.

3. — Il diagramma, l'istogramma, lo stereogramma, cioè i procedimenti geometrici di rappresentazione dei dati statistici, sono di universale applicabilità: qualunque serie statistica può essere rappresentata con uno di tali procedimenti. In pratica il primo procedimento è impiegato molto spesso, il secondo abbastanza frequentemente, il terzo quasi mai. A prescindere dalla difficoltà materiale della costruzione dello stereogramma, devesi notare che l'occhio umano in generale è poco avvezzo ad apprezzare relazioni di grandezza fra volumi diversi; con minore difficoltà giudica relazioni di grandezza fra aree; e con massima facilità valuta relazioni di grandezza fra lunghezze.

4. — Il diagramma, come sappiamo, si ottiene sostituendo a ciascun dato della serie un segmento di retta, di lunghezza proporzionale ad esso, secondo una determinata scala di ragguaglio. Quindi per preparare un diagramma si deve anzitutto determinare l'unità di lunghezza corrispondente all'unità di misura nella quale è espresso il dato statistico. Per questa determinazione non si può dare nessuna norma teorica: purchè, adottata una qualsiasi unità, ci si attenga fedelmente ad essa nella traduzione dei numeri in segmenti, nulla si può obiettare alla correttezza del procedimento. Ma dal punto di vista pratico non è indifferente la scelta dell'una o dell'altra unità di lunghezza. Se, per rappresentare graficamente la serie dei dati sulla produzione della ghisa in Italia negli ultimi cinquant'anni, facciamo corrispondere la lunghezza di un millimetro ad ogni mille tonnellate prodotte, da principio ci troviamo bene: negli anni dal 1881 al 1905 la produzione oscilla fra un minimo di 7 mila e un massimo di 143 mila tonnellate, e quindi i corrispondenti segmenti variano tra un minimo di 7 mm. e un massimo di 143 mm., che rientrano comodamente in un quadro adeguato alle

consuete dimensioni della pagina d'un libro. Ma andiamo avanti: già nel 1909 si sale a 208 mila tonnellate, cioè a 208 mm.; nel 1913 si giunge a 427 mila tonnellate, cioè a 427 mm., che trascendono la dimensione di una pagina di grande formato; e nel 1929 si raggiungono le 671 mila tonnellate, cioè i 671 mm. Sopra un cartone murale potremo eseguire la nostra rappresentazione alla scala fissata, e magari a scala più grande; ma se vogliamo contenere il diagramma in una pagina normale, dobbiamo mutare la scala. Proviamo allora a far corrispondere un millimetro a centomila tonnellate di ghisa: così il nostro diagramma, raggiungendo l'altezza massima di 6,71 mm. entrerà comodamente anche in una pagina di piccolissimo formato. Ma vedremo manifestarsi un altro inconveniente: i segmenti destinati a rappresentare la produzione dei primi anni riusciranno tanto piccoli, variando fra 0,07 e 1,43 mm., che sarà praticamente impossibile disegnarli nelle giuste proporzioni; e quand'anche ciò fosse possibile, l'occhio mal ne distinguerebbe le differenze. Se adottiamo una scala intermedia, facendo per esempio corrispondere un millimetro a cinquemila tonnellate, potremo giungere a dimensioni più soddisfacenti: i segmenti varieranno fra un minimo di 1,4 mm. e un massimo di 134,2 mm. e quindi comporranno un quadro di dimensioni convenienti ma nel tempo stesso atto a mostrare chiaramente l'andamento della serie.

In pratica, le dimensioni del foglio nel quale va inserito il diagramma danno norma alla scelta della scala, o almeno le segnano un limite verso l'alto; mentre un limite verso il basso è segnato dalla necessità che i vari segmenti risultino abbastanza grandi perchè l'occhio possa confrontarli reciprocamente. Entro questi limiti, il senso artistico e il senso pratico del compilatore del diagramma concorrono a determinare la scelta della scala.

Per poter bene confrontare tra loro i vari segmenti, si suole disporli ordinatamente, col piede su una stessa linea orizzontale, ad uguali distanze (quando non vi siano ragioni per proporzionare le distanze ad intervalli disuguali di tempo, di spazio, ecc. che separino tra loro le varie rilevazioni: esempio, censimenti eseguiti prima di 10 in 10 anni e poi di 5 in 5). Si possono disporre, invece, col piede su una stessa linea verticale; questo procedimento è usato meno spesso del primo, probabilmente perchè l'occhio è avvezzo a comparare tra loro lunghezze specialmente disposte in direzione normale alla superficie terrestre, cioè verticalmente.

Nel diagramma, come nelle altre forme di rappresentazione geometrica, è sempre bene indicare la scala, per agevolare l'interpretazione e il controllo; conviene indicarla anche se accanto alla rappresentazione grafica sono posti i dati numerici, come si fa talvolta per agevolare la lettura dei singoli dati.

5. — Se invece di segmenti sostituiamo ai vari termini altrettanti rettangoli aventi basi uguali fra loro e altezze proporzionali ai numeri rappresentati, regolandone le dimensioni in modo che la base risulti normalmente molto stretta in confronto all'altezza, otteniamo il così detto *diagramma a canne d'organo* o *a nastri.* A rigore esso è un istogramma, perchè ai numeri sono state sostituite aree ad essi proporzionali; ma tuttavia si suol considerare alla stregua di un diagramma perchè, essendo uguali le basi delle diverse « canne » o « sezioni di nastro » che costituiscono il grafico, la relazione di grandezza fra i vari termini è tradotta nella relazione tra le altezze di esse. Questa forma di rappresentazione in confronto alla precedente ha il pregio della migliore visibilità, specialmente se le « canne » sono interamente colorate in nero o in altra tinta uniforme.

6. — Quando il campo di variazione dei termini di una serie statistica supera certi limiti, diviene praticamente impossibile trovare un'unità di lunghezza che renda chiaramente visibile la relazione di grandezza fra i termini più piccoli senza dare eccessive dimensioni ai termini più grandi. Si provi, per esempio, a rappresentare graficamente la produzione dello zucchero in Italia negli ultimi cinquant'anni, che varia tra un minimo di 635 quintali nel 1881 ed un massimo di 3.822.072 nell'anno finanziario 1924-25: si incontreranno già notevoli difficoltà per trovare una scala soddisfacente. Ma se poi si vogliono rappresentare le vicende della moneta germanica nel decennio bellico e postbellico, durante il quale l'equivalenza del marco oro in marchi carta ha variato fra un minimo di 1 e un massimo di 1 trilione, nessun'arte di disegnatore può dare una rappresentazione adeguata.

In simili casi si ricorre all'espediente di rappresentare graficamente, invece dei dati statistici, i logaritmi di essi (*diagramma logaritmico*): così l'ampiezza del campo di variazione viene enormemente ridotta e la rappresentazione è resa praticamente possibile. Nell'ultimo esempio, invece di dover rappresentare graficamente segmenti di dimensioni comprese fra 1 e 1 trilione, avremo da rappresentare segmenti di dimensioni comprese fra 0 (logaritmo decimale di 1) e 12 (logaritmo di 1 trilione).

Mentre nel diagramma ordinario la differenza tra due segmenti è proporzionale *alla differenza* tra i due numeri che essi rappresentano, nel diagramma logaritmico essa è proporzionale *al rapporto* tra i due numeri: perciò questo tipo di diagramma rende utili servigi quando interessa mettere in rilievo le variazioni relative piuttosto che le variazioni assolute di un fenomeno. Oggi si tende ad abusarne, adoperandolo in molti casi nei quali sarebbe preferibile il diagramma ordinario: così si suscitano talvolta impressioni non corrispondenti alla realtà, anche perchè è molto ristretta la schiera di coloro che sono in grado di interpretare correttamente un diagramma logaritmico.

7. — La rappresentazione grafica di una funzione statistica (vedasi capitolo VIII, paragrafo 1) si vale dei metodi escogitati dai matematici per la rappresentazione grafica delle funzioni in generale.

Tracciamo in un piano due rette perpendicolari fra loro: l'una orizzontale, l'altra verticale; fissiamo su ciascuna di esse un senso positivo (rispettivamente da sinistra a destra e dal basso in alto) e un senso negativo; adottiamo per ciascuna di esse un'unità di misura delle lunghezze e su ciascuna riportiamo una scala di misura, assumendo come origine di entrambe le scale il punto d'incontro delle due rette. Ogni punto del piano è ora indicato univocamente quando siano date, in grandezza e in segno, le distanze del punto stesso dalle due rette di riferimento (*assi coordinati* o *assi cartesiani*). Si dice *ordinata* la distanza del punto dalla retta orizzontale, che si misura sulla retta verticale, detta perciò *asse delle ordinate.* Si dice *ascissa* la distanza del punto dalla retta verticale, che si misura sulla retta orizzontale, detta perciò *asse delle ascisse.* A ciascun punto del piano corrispondono due *coordinate ortogonali* determinate, cioè una coppia di valori dell'ascissa e dell'ordinata; e reciprocamente a ciascuna coppia di valori che si assumono l'uno ad ascissa e l'altro ad ordinata corrisponde univocamente un punto del piano.

Se ora abbiamo una quantità variabile y che sia funzione di un'altra quantità variabile x (sia tale cioè che a ciascun valore della *variabile indipendente* x compreso in un determinato intervallo corrisponda un solo e determinato valore della *variabile dipendente* y), possiamo valerci del sistema delle coordinate ortogonali per rappresentare la relazione fra le due variabili, purchè queste — com'è regola generale per le funzioni statistiche — possano assumere soltanto valori reali.

Ad ogni coppia di valori associati di y e di x corrisponde infatti, se assumiamo come ordinata il valore dato di y e come ascissa il valore dato di x, un punto del piano; e l'insieme dei punti così determinati rappresenta la funzione. Per esempio, alla coppia di valori associati $y = 9$, $x = 3$ corrisponde il punto avente ordinata 9 e ascissa 3, e la funzione $y = x^2$ è rappresentata dall'insieme dei punti aventi l'ordinata uguale al quadrato dell'ascissa, cioè dalla linea che è il luogo di tali punti; alla coppia di valori associati $y = 42$, $x = 4$ corrisponde il punto avente ordinata 42 e ascissa 4, e la funzione $y = 10x + 2$ è rappresentata dall'insieme dei punti aventi l'ordinata uguale al decuplo dell'ascissa aumentato di 2, cioè dalla linea che è il luogo di tali punti. E passando alle funzioni statistiche, il numero y dei superstiti di una data generazione all'età x è rappresentato dal punto avente y per ordinata e x per ascissa, e la funzione di sopravvivenza è rappresentata dall'insieme dei punti così determinati; il numero y delle persone aventi un reddito x è rappresentato dal punto avente y per ordinata ed x per ascissa, e la funzione di distribuzione dei redditi è rappresentata dall'insieme dei punti così determinati. Questa forma di rappresentazione corrisponde perfettamente a quella già spiegata nel paragrafo 4, poichè ogni dato statistico viene rappresentato da un segmento di retta (che qui è l'ordinata), di lunghezza proporzionale al dato stesso. Ma nel caso speciale che ora consideriamo i vari segmenti sono simboli di altrettanti valori di una medesima funzione, mentre nel caso generale del paragrafo 4 non hanno tale carattere; così che non avrebbe ivi alcun significato la linea condotta per i vertici dei segmenti, la quale qui rappresenta la funzione statistica. (Si pensi, per esempio, alla rappresentazione grafica di queste due serie: mortalità nei singoli anni d'età, in Italia: avente il carattere di funzione statistica; e mortalità nelle singole provincie italiane: non avente tale carattere). Avvertasi a questo proposito che il *diagramma lineare* viene chiamato così (in contrapposto al *diagramma areale* o istogramma) perchè in esso ciascun dato è *sempre* rappresentato da una linea (segmento) e non perchè *talvolta* può venir tracciata una spezzata o curva a rappresentare l'andamento della serie.

Una quantità variabile dicesi continua nell'intervallo x, $(x + a)$ se nel passare da un valore x ad un altro valore $(x + a)$ attraversa tutti i valori reali intermedi: tale il tempo. Se codesta condizione non è soddisfatta, la variabile si dice discontinua: tale la successione dei numeri interi. Una funzione d'una variabile continua si dice continua in un certo intervallo di valori della variabile indipendente, se col tendere a zero della differenza fra due successivi valori della variabile indipendente compresi in tale intervallo, tende a zero anche la differenza fra i corrispondenti valori della funzione. In statistica s'incontrano funzioni continue, come una dimensione somatica di un animale in funzione dell'età, e funzioni discontinue, come il numero dei componenti una popolazione in funzione del tempo.

Nella rappresentazione grafica, ad una funzione discontinua corrisponde un insieme di punti nettamente distinti l'uno dall'altro: così se si rappresenta il numero delle famiglie italiane in funzione del numero dei loro componenti; ad una funzione continua corrisponde una linea ininterrotta: così se si rappresenta la statura y di una persona in funzione dell'età x. Quando, pur essendo la funzione discontinua, a piccole variazioni della x corrispondono piccole variazioni della y, i punti che rappresentano la funzione danno l'immagine approssimativa di una linea continua: linea che si suole tracciare per rendere più evidente l'andamento del fenomeno: così se si vuol rappresentare il numero y dei superstiti di una data generazione in funzione dell'età x (*curva di sopravvivenza*); così anche nell'altro esempio di poc'anzi, della distribuzione dei redditi (*curva dei redditi*). Quando i punti noti sono meno regolarmente disposti, si traccia invece una linea spezzata allo stesso fine ora detto. Ma comunemente anche queste spezzate si comprendono nel nome generico di *curve statistiche*.

Si tratti di funzioni veramente continue, o di funzioni in via approssimativa riguardate tali, l'osservazione statistica soltanto in casi eccezionali permette di tracciare completamente la linea rappresentatrice della funzione in un dato intervallo: in generale ce ne indica soltanto alcuni punti. Il barometro registratore ci delinea direttamente la curva della pressione atmosferica in funzione dell'ora, cioè del tempo; un barometro usuale ci consente solo di leggere e registrare la pressione ad intervalli, che potranno essere di sei ore, di un'ora, di mezz'ora, magari di un minuto, ma che per quanto vicini sono sempre separati l'uno dall'altro da infiniti istanti. Quanto più brevi sono gli intervalli, tanto maggiore è l'approssimazione con la quale i punti noti ci indicano l'andamento della curva della pressione atmosferica. Analogamente un'ipotetica macchina che istantaneamente ed automaticamente registrasse ogni nascita, ogni morte, ogni caso individuale di immigrazione, ogni caso individuale di emigrazione, ci potrebbe tracciare la curva della popolazione in funzione del tempo ; ma in mancanza di una simile macchina dobbiamo contentarci dei dati forniti dai censimenti, eseguiti ad intervalli di cinque o di dieci anni l'uno dall'altro, i quali certo non bastano a darci una visione molto approssimata al vero dell'andamento della curva della popolazione.

8. — A chi voglia rappresentare graficamente una funzione statistica, si presenta anzitutto il problema della scelta della scala delle ascisse e di quella delle ordinate. Dall'aspetto teorico non v'è alcun limite all'arbitrio nella scelta ; dall'aspetto pratico, invece, bisogna tener presenti le considerazioni già esposte nel paragrafo 4 intorno alla determinazione di quei segmenti che ora, trattando delle funzioni statistiche, chiamiamo ordinate. Per le ascisse valgono considerazioni analoghe : bisogna contenersi nei limiti dello spazio disponibile e utilizzarlo nel miglior modo possibile.

Se il campo di variazione delle ordinate è molto esteso, conviene talvolta ricorrere all'espediente già esposto nel paragrafo 6: tracciare cioè un *diagramma a scala logaritmica*, rappresentando il logaritmo della variabile y in funzione della variabile x ; se è molto esteso anche il campo di variazione delle ascisse si può tracciare un *diagramma a doppia scala logaritmica*, rappresentando il logaritmo di y in funzione del logaritmo di x. Ma bisogna poi prestare molta attenzione per non incorrere in equivoci nella interpretazione di questi diagrammi logaritmici. E bisogna tener presente che a piccole variazioni dei logaritmi possono corrispondere grandi variazioni dei dati rappresentati.

Se ascisse ed ordinate sono espresse nella medesima unità di misura (esempio : prezzo medio del pane in funzione del prezzo della farina, indicati entrambi in lire per chilogrammo) per lo più conviene far corrispondere a tale unità di misura la medesima unità di lunghezza così nella scala delle ascisse come in quella delle ordinate. Ma se sono espresse in unità di misura differenti (esempio : numero degli infortuni nel lavoro in funzione del tempo trascorso dall'inizio del lavoro giornaliero), non si può dare alcuna regola intorno alla relazione di grandezza fra il segmento da assumere ad unità di ascissa e quello da assumere ad unità di ordinata : a determinare tale relazione prevalgono di solito ragioni di carattere pratico. Il rapporto tra la larghezza e l'altezza dello spazio disponibile per la rappresentazione grafica, il campo di variazione dei valori della x e di quelli della y, l'opportunità di mettere in chiaro rilievo l'andamento della serie statistica, il desiderio di mantenere uniformi le dimensioni di più diagrammi, ed altre simili circostanze, influiscono sulla scelta delle due scale. Gli statistici che si sono affaticati a ricercare norme generali applicabili in questa materia non hanno potuto conseguire alcun sicuro successo: hanno potuto bensì dare suggerimenti opportuni per singole categorie di applicazioni, ma non dettare un regolamento adatto a contenere l'infinita varietà dei casi che la pratica presenta. Con prescrizioni troppo rigide — d'altronde assolutamente arbitrarie — si corre il rischio di togliere al procedimento grafico quell'elasticità, quella adattabilità ai più vari bisogni, che costituisce uno dei suoi maggiori pregi.

Nel determinare le scale delle ascisse e delle ordinate (scale che in ogni caso vanno chiaramente indicate nel diagramma) non si deve però procedere alla cieca o a tentoni: è anzi necessario sapere quali siano gli effetti di una modificazione dell'una o dell'altra scala sull'aspetto della rappresentazione grafica. Per chiarirli ci gioveremo di un esempio, supponendo di voler rappresentare in diagramma la produzione della ghisa in Italia dal 1881 al 1930, in funzione del tempo. Cominciamo coll'assumere come unità di ascissa, corrispondente ad un anno, la lunghezza di 5 mm.; e come unità di ordinata, corrispondente a 10.000 tonnellate prodotte, la lunghezza di 2 mm. Dal 1903 al 1913 la produzione aumenta da 75 mila a 427 mila tonnellate; ossia in un intervallo orizzontale di 50 mm. la curva della produzione sale di 70,4 mm. Facciamo ora corrispondere non più 5 ma 10 mm. all'intervallo di un anno, mantenendo immutata la scala della produzione: vedremo ancora ascendere di 70,4 mm. la curva della produzione, ma non più in un intervallo orizzontale di 50 mm. bensì in uno di 100 mm.: abbiamo ancora l'impressione dell'ascesa ma di un'ascesa più lenta. Se invece facciamo corrispondere mm. 2,5 ad ogni anno, lasciando sempre immutata la scala della produzione, otteniamo un'impressione di ascesa più rapida. In generale: impiccolendo l'unità di ascissa diamo rilievo, ingrandendola togliamo rilievo alle differenze esistenti fra i dati della serie. Proviamo ora a modificare l'unità d'ordinata, senza mutare l'unità di ascissa: se facciamo corrispondere 4 mm. ad ogni 10.000 tonnellate di produzione, nell'intervallo orizzontale di 50 mm. la curva della produzione sale di 140,8 mm. invece che di 70,4 come dianzi; se facciamo corrispondere soltanto 1 mm. a 10.000 t., nell'intervallo orizzontale di 50 mm. la curva sale di 35,2 mm. appena; l'impressione dell'ascesa è accentuata nel primo caso, attenuata nel secondo. In generale: ingrandendo l'unità di ordinata diamo rilievo, impiccolendola togliamo rilievo alle differenze esistenti fra i dati della serie.

Comunque si modifichino le unità di ascissa e di ordinata, una curva ascendente non si può mai far apparire discendente, e viceversa. Ma è certo che regolando abilmente le due scale si possono ottenere effetti notevoli di attenuazione o di accentuazione delle differenze esistenti fra i dati della serie, senza uscire dai limiti della più rigorosa correttezza tecnica. Chi voglia essere obiettivo nella rappresentazione grafica cercherà di regolare le scale in modo che le differenze tra i dati abbiano rilievo adeguato alla loro importanza relativa: se codeste differenze sono relativamente grandi, tali devono apparire all'occhio di chi guarda il diagramma; se sono relativamente piccole, devono essere appena percepite.

Secondo il fine cui deve servire una rappresentazione grafica, può essere tuttavia opportuno attenuare differenze grandi (così se si vuol mostrare la tendenziale costanza del raccolto granario di un paese attraverso le forti oscillazioni da anno ad anno), ovvero accentuare differenze piccole (così se si vogliono seguire le variazioni, relativamente deboli, della temperatura di un ammalato). Soltanto chi conosca bene il fenomeno rappresentato e lo scopo della rappresentazione può dare sagge direttive per simili operazioni.

9. — Un metodo spesso usato per mettere in rilievo le differenze fra i dati della serie consiste nel presentare il diagramma amputato della fascia orizzontale che sta al di sotto dell'ordinata minima, ovvero di una parte di tale fascia. È come se per confrontare le stature di più uomini li disponiamo dietro un muricciuolo, dal quale si veda spuntare appena il sommo del capo del più basso, la fronte di un altro, il viso intero di un terzo, mentre un quarto emerge fino al collo e si scorge anche il busto di un quinto. Le differenze di statura appaiono così molto più rilevanti di quanto apparirebbero nel confrontare le intere stature col porre le persone davanti, invece che dietro, al muricciuolo. L'adozione di questo metodo chirurgico di presentazione dei diagrammi, che spesso è ispirata, specialmente nel campo della pubblicità, dal deliberato proposito di dare l'impressione d'una variazione più ampia del vero, in certi casi è consigliata dall'opportunità di risparmiare spazio e di meglio utilizzare quello disponibile, in altri è imposta dalla necessità di rendere chiaramente visibili differenze relativamente piccole fra i dati, che altrimenti sarebbero a stento

percepite. Ma nel guardare il grafico amputato bisogna ricostruirlo mentalmente intero, ad evitare fallaci impressioni: a ciò aiuta la scala delle ordinate, che non deve mancare se la mutilazione del diagramma è stata eseguita senza fini subdoli.

Non di rado i compilatori di diagrammi fanno corrispondere uguali intervalli di ascissa o di ordinata a disuguali intervalli delle variabili rappresentate (così se rappresentano mediante ordinate equidistanti i risultati di successivi censimenti eseguiti a disuguali intervalli di tempo; o se rappresentano mediante ascisse equidistanti gruppi di età di disuguale ampiezza dei quali le ordinate indicano la consistenza in una popolazione). Questo è un vero e proprio errore, che non può in alcun caso essere tollerato perchè deforma irreparabilmente la rappresentazione grafica.

10. — Quando non si conoscono i valori della variabile dipendente y corrispondenti a singoli valori della variabile indipendente x, ma si conoscono soltanto le somme dei valori della y corrispondenti ai valori di x compresi in singoli dati intervalli (non si conosce per esempio l'ammontare dei singoli redditi, ma si sa quanti sono compresi fra 0 e 1000 lire, quanti fra 1000 e 2000, quanti fra 2000 e 3000, ecc.), si possono costruire graficamente i rettangoli aventi per base l'intervallo x, $(x+h)$ e per altezza il quoziente per h della somma dei valori di y corrispondenti ai valori di x compresi nell'intervallo. Invece di una curva si ottiene in questo modo una specie di gradinata (i cui gradini sono ugualmente o disugualmente larghi secondo che l'ampiezza h degli intervalli di valore della x sia costante o variabile). Attraverso questa gradinata si può in molti casi tracciare l'approssimativo andamento della curva che il modo di presentazione dei dati statistici non ci permette di delineare esattamente: l'area della gradinata nota è uguale alla corrispondente area della curva ignota, così che se gli intervalli non sono molto ampi la curva può essere delineata con buona approssimazione. Quando gli intervalli sono uguali tra loro, il *diagramma a gradinata* si può far rientrare nel tipo, che già conosciamo, del diagramma, o meglio istogramma, a canne d'organo (v. paragrafo 5).

11. — Nei diagrammi le possibilità d'impiego dei colori o dei tratteggi sono ristrette.

Se nel diagramma compaiono esplicitamente i singoli segmenti che rappresentano i singoli dati della serie, ciascun segmento può essere scomposto in più segmenti di diversi colori, o delineati diversamente in un unico colore (a tratto grosso, a tratto fino, a lineette, a punti), corrispondenti a diverse parti del dato. Per esempio, in una rappresentazione della popolazione delle varie provincie, il segmento che corrisponde alla popolazione di ciascuna provincia può essere suddiviso in due parti diversamente colorate corrispondenti ai due sessi, o in più parti corrispondenti alle varie classi di professioni. Ma è ovvio che l'applicazione cromatica riesce più efficace se invece che su segmenti è eseguita su nastri (« canne d'organo »): le varie sezioni del nastro possono essere variamente colorate o tratteggiate, in modo assai meglio visibile.

Se nel diagramma compare la curva, o la spezzata, o la gradinata, che rappresenta l'andamento di una funzione statistica, l'area di questa può essere suddivisa, nel senso orizzontale, in più parti diversamente colorate o tratteggiate, corrispondenti ad altrettante parti nelle quali si scompone il fenomeno considerato.

Se, infine, su una stessa tavola grafica si raccolgono più diagrammi per poterli meglio confrontare tra loro, si possono delineare le diverse curve con tinte diverse o con tratti diversi, o si possono colorare o tratteggiare diversamente le aree, per distinguerle a colpo d'occhio in modo da evitare ogni confusione.

12. — Il diagramma si può tracciare anche col sistema delle *coordinate polari*, che però viene usato raramente perchè la rappresentazione grafica che ne risulta non è interpretata con rapidità e approssimazione paragonabili a quelle che si ottengono col sistema delle coordinate ortogonali (cartesiane).

Se si congiunge il punto di ascissa x e di ordinata y con l'origine degli assi cartesiani, mediante un segmento rettilineo, la lunghezza di questo segmento (*raggio vettore*) e la misura dell'angolo che esso determina con l'asse delle ascisse (*anomalia*) bastano ad individuare la posizione, nel piano, del punto dato, purchè si sia stabilito il senso dell'angolo: sono le sue coordinate polari. Una funzione è rappresentata, in questo sistema, dall'insieme dei punti aventi per anomalie i valori della variabile indipendente e per raggi vettori i corrispondenti valori della funzione.

Il metodo può essere convenientemente impiegato per la rappresentazione di fenomeni che si svolgono in successivi periodi di ampiezza costante, con tendenza crescente o con tendenza decrescente: in tal caso si farà corrispondere un angolo di 360° a ciascun periodo; la spezzata o la curva condotta per gli estremi dei raggi vettori mostrerà un andamento simile a quello delle spire d'una chiocciola: perciò talvolta queste rappresentazioni grafiche son dette « *diagrammi a chiocciola* ».

13. — Il principio cui s'informa la rappresentazione grafica mediante aree (*istogramma* o *diagramma areale*) è quello, semplicissimo, della proporzionalità tra l'area di ciascuna figura geometrica e la grandezza del corrispondente termine della serie.

Bisogna però che le aree adottate per rappresentare i diversi termini di una serie siano tutte di uguale forma, perchè l'occhio possa confrontarle con speranza di approssimarsi al vero nel loro apprezzamento comparativo. Se si rappresenta un dato con un rombo, un altro con un quadrato, un terzo con un trapezio, un quarto con un triangolo, un quinto con un cerchio, e così via, l'occhio si confonde nella diversità delle forme e cade facilmente in errore nell'apprezzamento comparativo delle grandezze (in ogni libro di giuochi scientifici si trovano esempi di simili illusioni ottiche). Le forme maggiormente usate sono quelle rettangolari e circolari, che essendo familiari all'occhio consentono più facilmente un corretto giudizio.

La determinazione della scala, nell'esecuzione di un istogramma, è regolata esclusivamente da considerazioni di opportunità pratica, analoghe a quelle che abbiamo esposte trattando del diagramma. La scala dev'essere esplicitamente indicata nel grafico.

Se si impiegano rettangoli: o questi hanno tutti uguale base e allora ciascuno di essi deve avere altezza proporzionale al dato da rappresentare, come nell'istogramma a canne d'organo; o hanno basi differenti, e allora dev'essere proporzionale al dato il prodotto della base per l'altezza: in particolare, se si impiegano quadrati, ciascuno di questi deve avere il lato proporzionale alla radice quadrata del dato da rappresentare. Se si impiegano cerchi, ciascuno di questi deve avere il raggio proporzionale alla radice quadrata del dato da rappresentare, affinchè l'area riesca proporzionale al dato.

Talvolta i compilatori di istogrammi proporzionano al dato statistico, e non alla radice quadrata di esso, il lato del quadrato o il raggio del cerchio; e così ottengono quadrati o cerchi che hanno aree proporzionali non al dato statistico bensì al quadrato di esso, e perciò danno luogo ad apprezzamenti molto divergenti dalla realtà. È anche questo un errore inescusabile ed irreparabile, che pur troppo spesso s'incontra.

Analogo errore è implicito, per lo più, nell'impiego, per la rappresentazione di serie statistiche, di aree irregolari il cui contorno richiama la natura del fenomeno rappresentato (immagini di patate se si rappresentano raccolti di questo tubero in anni successivi; di navi se si rappresenta il tonnellaggio delle flotte dei vari paesi; di bambini in fasce se si rappresentano le nascite nelle diverse provincie). Certo l'immagine ravviva la rappresentazione grafica e attrae l'attenzione del profano; ma se viene eseguita con criterio corretto, in modo cioè che l'area da essa circoscritta sia pure proporzionale al dato, difficilmente raggiunge l'effetto voluto dal compilatore, che in generale è quello di mettere in forte risalto le differenze fra i dati della serie: effetto che più semplicemente si consegue con un diagramma a canne d'organo. E allora, per ignoranza o per malizia, il compilatore per lo più proporziona al dato da rappresentare non l'area ma l'altezza della immagine; ad evitare deformazioni, modifica press'a poco in uguale proporzione anche la larghezza, così che l'area riesce proporzionale all'incirca al quadrato del dato rappresentato e non al dato stesso. Un paese ha un esercito doppio di quello d'un altro: l'immagine di un soldato due volte più alto, ma anche due volte più largo, di un altro ci dà l'impressione visuale di uno squilibrio di potenza militare molto maggiore di quello realmente esistente, facendo supporre un rapporto di 4 ad 1 e non di 2 ad 1 qual è di fatto. Se non si vuol rinunziare a questa forma di rappresentazione grafica adatta a far colpo sugli incolti, ma non si vuole nel tempo stesso impiegarla scorrettamente (cioè, in molti casi, disonestamente), si può adottare un semplice espediente, quello di iscrivere ciascuna immagine in un rettangolo o in un cerchio di area proporzionale al dato da rappresentare; in tal modo essa dà un'impressione adeguata alla realtà e non falsa ed esagerata come quella che viene suscitata dall'errato sistema consueto. Conviene inoltre delineare le immagini a semplice contorno, perchè se si cerca di dare all'occhio l'impressione delle tre dimensioni mediante artifici di ombreggiatura o di prospettiva esse possono venire confuse, specialmente se hanno forma regolare (casse di merce, ammassi di minerali, ecc.), con proiezioni di stereogrammi: dal che sorgono nuovi dubbi nell'interpretazione.

Data la maggior difficoltà che l'occhio incontra nel confronto tra aree, rispetto al confronto tra lunghezze, conviene sempre unire alla rappresentazione grafica eseguita con questo metodo l'indicazione numerica dei dati, ciascuno dei quali può essere inserito nell'area corrispondente, o sopra o sotto di essa.

14. — L'area che rappresenta un dato statistico può essere suddivisa in tante parti proporzionali a parti costituenti del dato. Le forme che meglio si prestano alla suddivisione sono il rettangolo, che può essere diviso sia verticalmente sia orizzontalmente, ed il cerchio, che può essere diviso sia in settori (è il modo preferibile) sia in anelli concentrici.

Colori o tratteggi differenti possono essere impiegati per distinguere le varie parti di ogni area: dev'essere in tal caso chiaramente indicato il significato di ciascun colore o tratteggio.

Se in una stessa tavola si raccolgono istogrammi che rappresentano diverse serie, ciascuno di essi può essere contrassegnato da un colore o da un tratteggio speciale.

15. — Nella preparazione di uno *stereogramma* si devono osservare norme per così dire parallele a quelle esposte per l'istogramma: impiegare figure solide della stessa forma per rappresentare i vari dati di una medesima serie; se s'impiegano cubi o sfere, proporzionare il lato o il raggio alla radice cubica del dato da rappresentare; se s'impiegano immagini scolpite, inserire ciascuna di esse in un cubo o in una sfera di materia trasparente, o in un mezzo cubo od emisfero cavo di volume proporzionale al dato da rappresentare, ecc.

Colori o tratteggi differenti possono essere impiegati per indicare differenze qualitative.

Se invece di presentare gli stereogrammi, si presentano le loro proiezioni sul piano, bisogna avvertirlo esplicitamente, affinchè non vengano confusi con istogrammi. In ogni caso è bene accompagnare allo stereogramma l'indicazione numerica del dato corrispondente e quella della scala.

Non ci tratteniamo più a lungo su questo metodo di rappresentazione, perchè, come abbiamo già avvertito, è poco usato per la difficoltà di apprezzamento che presenta all'occhio non specialmente esercitato.

16. — Accanto ai procedimenti grafici e geometrici di applicabilità universale fin qui esposti, troviamo tutto un altro complesso di procedimenti grafici che sono applicabili soltanto a serie indicanti la distribuzione di fenomeni nello spazio: per Stati, per circoscrizioni territoriali, per regioni o zone geografiche, per linee di comunicazione, per località, ecc. Queste rappresentazioni grafiche, le

quali hanno sempre per base una carta geografica o topografica, per lo più semplificata o schematizzata, si dicono *cartogrammi*.

Consideriamo anzitutto il caso in cui i dati della serie rappresentano la misura assoluta di un fenomeno in tante circoscrizioni o zone territoriali diverse. Conosciamo, per esempio, il numero degli abitanti di ciascuna provincia italiana: ne vogliamo una rappresentazione grafica. Immaginiamo di poter distribuire gli abitanti di ciascuna provincia uniformemente su tutto il territorio di essa e di fotografare poi da una grande altezza l'Italia intera. Ogni persona sarà rappresentata, sulla nostra colossale carta geografica dell'Italia, da un punto; in alcune provincie i punti appariranno radi, in altre meno radi, in altre densi, in altre densissimi: avremo così una visione diretta della varia densità della popolazione nelle varie provincie. Senza disturbare il prossimo, possiamo assai più semplicemente ottenere lo stesso effetto ottico: basta che sopra una carta geografica a semplice contorno, dove siano segnati i confini delle provincie, inseriamo nello spazio corrispondente a ciascuna provincia tanti punti uniformemente distribuiti (intendasi punti nel senso grafico non in quello geometrico) quanti sono gli abitanti, o meglio quante sono le migliaia di abitanti, chè altrimenti la carta si trasformerebbe in una sola macchia nera. Secondo la grandezza dei numeri da rappresentare e secondo la grandezza della carta geografica, faremo corrispondere un punto ad 1 abitante, a 100, a 1.000, a 10.000, o ad un altro qualsiasi numero n, in modo da ottenere che in nessuna parte del cartogramma i punti riescano tanto fitti da confondersi tra loro, e possibilmente in nessuna parte riescano eccessivamente radi. Si mira ad ottenere, mediante la varia densità dei punti, l'effetto di una colorazione più o meno intensa.

Questo metodo (*cartogramma a punti*), se abilmente applicato, riesce assai efficace: bisogna notare, però, che essendo le aree delle varie circoscrizioni nella carta geografica proporzionali alle aree reali, distribuire su di esse un numero di punti proporzionale alla misura assoluta del fenomeno equivale a mettere in rapporto tale misura con l'estensione territoriale. Nel nostro esempio della popolazione, ciò risulta opportuno, perchè si ottiene una visione della densità di essa nelle varie provincie; ma in molti altri casi può risultare inopportuno trattandosi di fenomeni che si vogliono far apprezzare nella loro misura assoluta e non in rapporto al territorio: così per esempio se si vuol rappresentare la serie statistica degli omicidii avvenuti in un anno, suddivisi per provincie.

17. — Qui soccorre un altro espediente. Torniamo al precedente esempio. Invece di distribuire la popolazione uniformemente sul territorio di ciascuna provincia, prima di fotografare dall'alto la nostra Italia, potremmo radunare la popolazione di ciascuna provincia in un luogo aperto situato nella parte centrale della circoscrizione, indi disporla su tante fitte righe di uguale lunghezza, in modo da formare un rettangolo o un quadrato, e infine eseguire la fotografia dall'alto. In questa immagine, sull'area corrispondente a ciascuna provincia vedremmo disegnato un rettangolo od un quadrato nero di superficie proporzionale al numero degli abitanti. Più semplicemente giungiamo a tale risultato tracciando sulla carta geografica, nello spazio corrispondente a ciascuna provincia, un rettangolo, un quadrato, oppure anche un cerchio od altra figura, di area proporzionale alla popolazione della provincia: otteniamo così un *cartogramma ad istogrammi.* Bisogna scegliere la forma delle figure e regolare la scala in modo da evitare possibilmente che esistano interferenze tra le figure riguardanti circoscrizioni o zone territoriali prossime le une alle altre, ma da ottenere nel tempo stesso che anche le figure più piccole riescano nettamente discernibili e comparabili tra loro. Talvolta riesce impossibile conciliare questi *desiderata*; sugli espedienti opportuni a risolvere il contrasto che allora sorge ci intratterremo fra poco.

18. — Nel cartogramma a punti si possono simultaneamente rappresentare più serie mediante punti diversamente colorati (per esempio: distribuzione per provincie del bestiame bovino con punti rossi, del bestiame equino con punti azzurri, del bestiame ovino con punti gialli, del bestiame suino con punti verdi); ma in generale queste rappresentazioni simultanee risultano confuse ed esigono un paziente lavoro d'interpretazione. Nel cartogramma ad istogrammi, mediante l'impiego di colori o di tratteggi coi quali si contrassegnano varie sezioni di ciascuna figura, si può rappresentare la suddivisione di ciascun dato totale in più dati parziali: nel nostro esempio, il cerchio che rappresenta la popolazione di ciascuna provincia può essere suddiviso in tanti settori, diversamente colorati o tratteggiati, aventi aree proporzionali ad altrettante classi nelle quali la popolazione si suddivida (secondo il sesso, o la professione, od altra circostanza). In que-

sto modo il cartogramma viene a rappresentare simultaneamente una serie di dati totali e tante serie di dati parziali quante sono le circoscrizioni o zone territoriali. Quando non si ecceda nelle suddivisioni e quando il numero delle circoscrizioni territoriali non sia molto grande, questo metodo di rappresentazione può riuscire molto efficace: valga come esempio il cartogramma che indica la suddivisione dei depositi fiduciari in Italia secondo i compartimenti nei quali sono raccolti e secondo il genere degli enti raccoglitori, inserito nell'ASI del 1913.

19. — Un procedimento analogo a quello dianzi descritto si può seguire quando i dati della serie misurano, in cifre assolute, la manifestazione di un fenomeno in singole località, che sulla carta geografica o topografica appaiono come punti. Facendo centro in ciascuno di questi punti, si possono tracciare cerchi o quadrati, od altre figure, di aree proporzionali ai dati da rappresentare; tratteggi o colori diversi permettono di rappresentare suddivisioni dei dati stessi. Se vogliamo rappresentare la serie dei dati sul peso delle merci spedite per ferrovia dai vari centri ferroviari italiani, possiamo servirci di questo metodo del cartogramma ad istogrammi; colori o tratteggi diversi ci consentono di distinguere diverse categorie di merci. In questo esempio riesce molto difficile evitare interferenze tra le figure relative a diversi centri: affinchè il cerchio o il quadrato corrispondente al grande traffico delle stazioni di Milano non giunga a coprire quelli corrispondenti al traffico dei prossimi centri ferroviari di Monza, di Saronno, di Rho, ecc., bisogna adottare una scala talmente piccola che le figure corrispondenti a centri di medio traffico risultano minuscole e mal comparabili tra loro, e quelle corrispondenti a centri di scarso traffico risultano addirittura invisibili; oppure bisogna tracciare il cartogramma sopra una carta grandissima. In simili casi, si possono usare due espedienti. Il primo espediente, corretto, consiste nell'ammettere le interferenze, rappresentando però entro le figure maggiori, e quasi viste per trasparenza, le figure minori che esse ricoprirebbero (avvertasi che così diviene impossibile la suddivisione di ciascuna figura in settori o sezioni diversamente colorati o tratteggiati: bisogna contentarsi di rappresentare i dati complessivi). Il secondo espediente, scorretto, consiste nell'evitare le interferenze coll'adottare scale differenti per la rappresentazione dei dati di differente grandezza: nel far corrispondere, per esempio, un centimetro quadrato ad un milione di tonnellate nelle stazioni di grande traffico, a centomila

tonnellate in quelle di medio traffico, a diecimila tonnellate in quelle di piccolo traffico. Questo secondo sistema è ammissibile se si scinde la rappresentazione in più cartogrammi distinti (nel nostro esempio, uno per le stazioni di grande traffico, uno per quelle di medio traffico, uno per quelle di piccolo traffico); ma non può essere in alcun modo giustificato quando si esegua un unico cartogramma, perchè gli istogrammi in questo contenuti devono essere disegnati tutti alla stessa scala per poter essere interpretati dall'occhio senza necessità di rettificare col ragionamento le impressioni visive. Un terzo possibile espediente, corretto e non privo di eleganza, consiste nel surrogare agli istogrammi proiezioni di stereogrammi, col che si riduce di molto la dimensione delle figure più grandi; ma talvolta questo rimedio può essere insufficiente, e in ogni caso presenta il solito inconveniente della difficoltà grande che incontra il nostro occhio nella comparazione di stereogrammi, ancor più se rappresentati in proiezione. Sopratutto per questa seconda ragione il *cartogramma a stereogrammi in proiezione piana* trova scarso impiego.

20. — Ancora il cartogramma ad istogrammi serve per la rappresentazione grafica di fenomeni che si svolgano lungo linee di comunicazione, rappresentabili mediante linee sulla carta geografica o topografica. La linea grafica viene assunta come asse di un nastro, la cui lunghezza risulta pertanto proporzionale, nella scala della carta, alla lunghezza della linea di comunicazione, e la cui larghezza viene proporzionata alla misura della manifestazione del fenomeno. Con cartogrammi di questo genere, detti *a nastri*, si può rappresentare il traffico di linee di trasporto terrestri, sotterranee, aeree, acquatiche; quello di linee telegrafiche o telefoniche; la consistenza di correnti commerciali, di correnti migratorie, ecc. Colori o tratteggi diversi applicati a suddivisioni longitudinali dei nastri permettono di suddividere le correnti rappresentate in più parti costituenti (viaggiatori di I, II e III classe, emigranti per gruppi professionali, ecc.). Anche in questa forma di rappresentazione sono molte volte inevitabili interferenze tra le figure che rappresentano i vari dati, cioè tra i vari nastri: intorno ai rimedi adottabili si possono ripetere considerazioni analoghe a quelle del precedente paragrafo.

21. — La sovrapposizione di diagrammi, invece che di cartogrammi, alle aree corrispondenti a varie circoscrizioni o zone in una carta geografica o topografica, può servire a mettere in rilievo differenze esistenti nello spazio tra l'andamento di un medesimo feno-

meno nel tempo, o la distribuzione di un medesimo carattere. Per esempio, in corrispondenza alle varie regioni italiane, si potrebbe così rappresentare il vario andamento della natalità negli ultimi cinquant'anni, o la varia distribuzione per stature degli iscritti di leva. Il metodo può servire vantaggiosamente quando le circoscrizioni o zone siano poco numerose e quando l'andamento del fenomeno in esame presenti forti differenze da zona a zona, prontamente rilevabili con uno sguardo alla rappresentazione grafica. Negli altri casi, data anche la necessità, che in generale s'impone, di adottare per i diagrammi scale molto piccole, il *cartogramma a diagrammi* è più atto a confondere le idee che a chiarirle. Pare superfluo avvertire che la scala dev'essere la medesima per tutti i diagrammi.

22. — I tipi di cartogramma fin qui studiati sono tutti fondati su criteri di proporzionalità tra il simbolo grafico e il dato numerico da rappresentare. Il *cartogramma a tinte graduali* è invece fondato sopra un criterio di analogia, col quale si fanno corrispondere, attraverso una iniziale associazione di idee ed una successiva convenzione da questa ispirata, variazioni d'intensità d'una tinta a variazioni di grandezza d'una quantità. Esso è generalmente applicato nella rappresentazione di misure relative di singoli fenomeni in varie circoscrizioni o zone territoriali, oppure in varie regioni o zone geografiche; è usato, cioè, specialmente quando non si hanno cifre assolute bensì rapporti o medie. Se vogliamo rappresentare la densità della popolazione nelle varie provincie italiane, applicheremo sulla carta gradazioni più o meno intense di una tinta, per esempio di rosso, all'area corrispondente a ciascuna provincia, secondo che è maggiore o minore la densità della popolazione. Se vogliamo rappresentare la piovosità di varie regioni geografiche nelle quali si possa dividere il territorio nazionale, applicheremo alle regioni gradazioni di tinta più o meno intense secondo la maggiore o minore quantità delle precipitazioni. Il principio è semplice, ma l'applicazione dà luogo a qualche difficoltà.

Anzitutto ci chiediamo: come proporzionare l'intensità della tinta alla grandezza del numero? Nelle forme di rappresentazione geometrica dianzi studiate, ad un numero n volte maggiore di un altro corrisponde un numero di punti, o una lunghezza, o un'area, o un volume, n volte maggiore; e l'occhio può percepire in modo più o meno rapido e preciso il rapporto di grandezza fra i due numeri, aiutato ove occorra dalle indicazioni di strumenti di misura. Ma in questa rappresentazione cromatica, anche ammesso che sia possibile assumere ad unità un leggerissimo strato di colore e che col sussidio di mezzi meccanici si possano sovrapporre successivamente numerosissimi strati unitari, come si può ottenere l'effetto ottico della proporzionalità? Si sovrappongano in un foglio 50 strati unitari di colore; in un altro se ne sovrappongano 100; indubbiamente si distinguerà che la seconda tinta è più intensa della prima; ma sarà estremamente difficile giudicare con approssimazione la proporzione di intensità esistente fra l'una e l'altra. Sarà poi assolutamente impossibile, all'occhio normale, pur aiutato da una scala delle gradazioni di tinta, discernere la gradazione corrispondente a 98 o 99 strati sovrapposti da quella corrispondente a 100 o 101; anche gradazioni più distanti fra loro saranno facilmente confuse. Quindi, se pure non manca la possibilità tecnica (almeno quando il campo di variazione dei dati della serie non sia molto grande) di proporzionare esattamente l'intensità della rappresentazione cromatica alla grandezza del numero da rappresentare, manca però la convenienza di farlo, perchè l'occhio non giunge a percepire differenze di tinta ottenute laboriosamente e con alto costo. In pratica, gli esecutori di cartogrammi a tinte graduali sogliono accontentarsi di poche gradazioni: talvolta cinque o sei soltanto; al massimo una ventina. A ciascuna gradazione si fanno corrispondere i numeri compresi in un dato intervallo: per esempio, volendo rappresentare con otto gradazioni la densità della popolazione nelle varie provincie italiane, poichè essa varia fra un minimo di 29 e un massimo di 705 abitanti per chilometro quadrato, potremo far corrispondere la prima gradazione di tinta alle densità da 1 a 100 abitanti, la seconda a quelle da 101 a 200, la terza a quelle da 201 a 300, e così via. L'inconveniente di questo metodo è quello di qualsiasi classificazione di grandezze: si assegnano ad una stessa classe, cioè si rappresentano con la stessa gradazione di tinta, grandezze che possono essere assolutamente o almeno relativamente molto differenti tra loro (nel nostro esempio, la provincia di Nuoro con 29 abitanti per chilometro quadrato e quella di Arezzo con 95, ossia più del triplo, verrebbero contrassegnate con la stessa gradazione di tinta); e da altra parte si assegnano a classi diverse, cioè si rappresentano con gradazioni di tinta diverse, grandezze che possono essere assolutamente e relativamente poco differenti tra loro (nel nostro esempio la provincia di Milano con 685 abitanti per kmq. e quella di Napoli con 705 verrebbero contrassegnate con gradazioni di tinta diverse). L'inconveniente è inevitabile se si adotta il criterio di far corrispondere ogni gradazione ad un intervallo di grandezza e non ad un singolo numero: espediente che in pratica bisogna quasi sempre necessariamente seguire. Si può cercare di attenuarlo coll'impiegare il maggior numero possibile di gradazioni e con lo stabilire nel modo più razionale la corrispondenza fra gradazioni di tinte e intervalli di grandezza.

23. — Qui conviene avvertire che in pratica, per difficoltà tecniche, o per scarsa capacità degli esecutori di cartogrammi, o per convinzione della poca utilità del sistema, non viene quasi mai seguito il metodo della sovrapposizione di successivi strati di tinta di costante intensità. Vengono invece, ad occhio, scelte gradazioni successive d'intensità crescente. Talvolta, per ottenere maggior numero di gradazioni nettamente discernibili, si passa da una tinta ad un'altra tinta contigua nella scala cromatica, per esempio dal giallo chiaro all'azzurro scuro, attraverso il verde: metodo al quale non si possono opporre serie obiezioni perchè l'occhio discerne senza essere guidato da speciali convenzioni la naturale successione delle gradazioni. Altre volte si preferisce passare gradualmente dalla massima intensità di una tinta ad una intensità minima, da questa alla minima di un'altra tinta, indi gradualmente alla massima intensità di quest'ultima: per esempio dall'azzurro scuro gradualmente al celeste chiaro, da questo al rosso chiaro e poi gradualmente al rosso scuro. Questo metodo è irrazionale, perchè fa corrispondere alle minime intensità del fenomeno, come alle massime, le più forti intensità di tinta, e alle intensità intermedie del fenomeno le più deboli intensità di tinta; esso può essere opportunamente impiegato solo quando la misura del fenomeno assuma valori in parte positivi (contrassegnati dalle gradazioni di una tinta), in parte negativi (contrassegnati dalle gradazioni dell'altra tinta): così se si tratta di rappresentare la variazione relativa della popolazione dei vari comuni fra due censimenti, e in una parte dei comuni la variazione è stata positiva, in un'altra parte negativa. (Del pari in una carta geografica si adoperano opportunamente gradazioni di due tinte, per rappresentare con l'una le altezze sul livello del mare, con l'altra le profondità sotto tale livello). Ancor più irrazionale di quello dei due colori è il criterio col quale si fanno corrispondere non gradazioni successive di tinta, ma tinte nettamente diverse, alle diverse misure dell'intensità di un fenomeno: per esempio l'azzurro a densità di popolazione fra 1 e 100 abitanti per kmq., il giallo a densità da 101 a 200, il rosso a densità fra 201 e 300, ecc. L'occhio scorge un mosaico di vi-

vaci colori, ma non può giudicare del loro significato, poichè questo è puramente convenzionale e non fondato su proprietà fisiche dei colori stessi. Consultando le indicazioni sulla corrispondenza delle varie tinte, che devono essere annesse al cartogramma, si può anche in questo caso intendere faticosamente il significato delle colorazioni: ma il cartogramma vien meno al suo fine principale, che è quello di dare a colpo d'occhio una visione d'insieme delle variazioni dell'intensità del fenomeno attraverso lo spazio.

Un'altra avvertenza: quando si impiegano le gradazioni tra il bianco e il nero, si cerca di ottenere l'effetto di una crescente intensità di tinta con punti o con tratti che ricoprano in nero una frazione gradualmente crescente della superficie bianca. Anche qui è possibile l'adozione di un criterio di proporzionalità: si possono successivamente inserire n punti, $2n$ punti, $3n$ punti ecc. in ciascun centimetro quadrato di superficie, o tracciare n linee parallele, $2n$ linee parallele, $3n$ linee parallele, ecc., attraverso ciascun centimetro lineare. Ma, in pratica, per lo più si procede ad occhio; non di rado invece di punti o di linee parallele e incrociate si adottano disegni relativamente complicati; e la maggior parte dei cartogrammi così compilati riesce difficilmente interpretabile a prima vista, per la difficoltà di giudicare l'ordine di successione delle varie gradazioni.

24. — Torniamo ora al problema che ci si era presentato: data una successione di gradazioni di tinta, stabilire la più opportuna corrispondenza tra essa e la successione numerica delle grandezze da rappresentare. La soluzione più semplice è quella di suddividere il campo di variazione dei termini della serie in tanti intervalli uguali quante sono le gradazioni di tinta disponibili, facendo poi ordinatamente corrispondere una gradazione a ciascun intervallo. Nel nostro esempio della densità della popolazione nelle varie provincie, abbiamo visto che si va da un minimo di 29 a un massimo di 705: il campo di variazione ha un'ampiezza di 676 unità. Se abbiamo a disposizione dieci gradazioni di tinta, possiamo suddividerlo in dieci intervalli, dell'ampiezza di 67,6 unità ciascuno, o più comodamente, spostando di due punti così il massimo come il minimo, dell'ampiezza di 68 unità: al primo intervallo — da oltre 27 fino a 95 abitanti per kmq. — faremo corrispondere la tinta meno intensa di tutte, al secondo — da oltre 95 fino a 163 abitanti per kmq. — la tinta subito successiva in ordine di intensità, e così via fino all'ultima tinta che sarà applicata alle provincie aventi da oltre 639 fino a 707 abitanti per kmq. Dall'aspetto teo-

rico questa soluzione è senza dubbio la migliore, specialmente se le successive gradazioni sono ottenute mediante la successiva sovrapposizione di strati di tinta di costante intensità: si può dire infatti che così si applica ancora, almeno in forma approssimativa, un criterio di proporzionalità. Ma dall'aspetto pratico la soluzione stessa presenta lo svantaggio di non dare un'adeguata visione della distribuzione dell'intensità del fenomeno, quando le intensità ac certate nelle varie zone o circoscrizioni siano molto disugualmente ripartite nel campo di variazione. Nel caso nostro, alla prima gradazione di tinta (densità fra 28 e 95) corrispondono 24 provincie, alla seconda (densità fra 96 e 163) ne corrispondono 35, alla terza (densità fra 164 e 231) ne corrispondono 23, alla quarta (densità fra 232 e 299) ne corrispondono 6 sole, alla quinta (densità fra 300 e 367) ne corrisponde appena una, alla sesta (densità fra 368 e 435) non ne corrisponde nessuna, alla settima (densità fra 436 e 503) ne corrisponde una, all'ottava ed alla nona (densità fra 504 e 571 e fra 572 e 639) non ne corrisponde nessuna, alla decima (densità fra 640 e 707) ne corrispondono 2. Tre gradazioni di tinta non sono affatto utilizzate, ad altre quattro corrispondono in complesso soltanto 10 provincie, mentre le tre rimanenti gradazioni comprendono da sole le altre 82 provincie. In altri esempi l'inconveniente può essere ancor più accentuato: basta che vi siano poche intensità altissime, magari una sola, di fronte a molte intensità basse, perchè la massima parte delle gradazioni di tinta resti inutilizzata. Se, nel nostro esempio, 91 provincie avessero avuto densità fra 28 e 95, e una sola avesse avuto densità di 707, il cartogramma ci sarebbe apparso colorato con una sola gradazione di tinta, la meno intensa di tutte, con l'unica eccezione d'una provincia che sarebbe stata colorata con la gradazione più intensa di tutte. Benchè correttamente eseguita, la rappresentazione grafica avrebbe fallito completamente al suo fine di mostrarci le caratteristiche della distribuzione geografica della densità degli abitanti.

L'inconveniente della disuguale ripartizione dei dati da rappresentare fra le diverse tinte, che si può presentare nell'applicazione del metodo ora descritto, ha suggerito l'adozione di un altro metodo, che elimina del tutto l'inconveniente stesso: quello di ordinare per grandezza i dati da rappresentare, di suddividerli in tanti gruppi di ugual numero di termini quante sono le gradazioni di tinta disponibili e di far corrispondere a ciascuno di questi gruppi, ordinatamente, una gradazione di tinta. Nel caso nostro, abbiamo 92 provincie e 10 gradazioni; non potendo dividere le provincie in 10 gruppi proprio uguali, le divideremo in 8 gruppi di 9 e 2 gruppi di 10: prenderemo, per esempio, le prime 10 provincie in ordine di densità della popolazione, le successive 9, le 9 ancora successive, e così via fino alle ultime 10; e ai gruppi così formati faremo corrispondere, ordinatamente, le gradazioni di tinta dalla meno intensa alla più intensa. Ma con questo metodo si manifesta un inconveniente non meno grave di quello del precedente: l'estensione dell'intervallo di valore corrispondente ad una gradazione di tinta varia grandemente, secondo che i termini ordinati per grandezza compresi nell'intervallo stesso differiscono di poco, oppure di molto, tra loro. Nel nostro esempio, la quinta gradazione si applica a provincie con densità che vanno da 132 a 141 abitanti per kmq., la decima si applica a provincie con densità che vanno da 233 a 705: l'una gradazione abbraccia un intervallo di 10 unità, l'altra uno di 473 unità, cioè 47 volte più grande. Così che talvolta all'uguaglianza di gradazione di tinta fra due provincie corrisponde un'approssimativa uguaglianza della densità della popolazione, tal altra le corrisponde una differenza grandissima; mentre nell'applicazione del primo metodo sappiamo che la differenza tra due provincie ugualmente colorate non può superare la differenza, uguale per tutti gli intervalli, tra i limiti dell'intervallo.

Da un punto di vista, diremo così, matematico il secondo metodo sarebbe da ripudiare. Ma talora le esigenze della pratica consigliano di adottarlo, quando l'inconveniente cui esso dà luogo non si manifesti intollerabile. Più spesso si adottano corrispondenze tra gradazioni di colore ed intervalli di valori numerici, che costituiscono una via di mezzo fra quelle suggerite dai due criteri estremi; e solo chi non si è trovato di fronte alle difficoltà pratiche di compilazione d'un cartogramma a tinte graduali può condannare a priori questi tentativi di conciliazione, che abilmente eseguiti possono dare buoni risultati, cioè consentire un apprezzamento visivo adeguato alla realtà. Si dà apparenza sistematica a tali procedimenti empirici, col porre qualche norma a base della determinazione dell'ampiezza degli intervalli: per esempio, invece di assumere intervalli uguali come nel primo metodo, si possono assumere intervalli di ampiezza crescente dal basso in alto (criterio conveniente quando i termini della serie si diradano sempre più di mano in mano che si ascende nella scala della grandezza), o di ampiezza decrescente dal basso in alto (criterio conveniente quando i termini si addensano sempre più di mano in mano che si sale), o infine di ampiezza decrescente nel salire dal valore minimo al valore più frequente e poi crescente nel salire dal valore più frequente al massimo (criterio conveniente quando i termini si addensano intorno ad un valore normale, nel senso spiegato nel capitolo sulle medie). Un'applicazione non rara del primo criterio consiste nel fissare gli intervalli in modo che i loro limiti costituiscano una progressione geometrica, invece che una progressione aritmetica come nel metodo esposto da principio. Forme particolari di distribuzione per grandezza dei termini possono suggerire altri criteri o speciali modalità di applicazione dei criteri ora indicati.

25. — In tutte le forme di cartogrammi devono essere chiaramente indicate le corrispondenze tra valori numerici e punti, lunghezze, aree, volumi, colori impiegati, in modo che l'interpretazione data alla rappresentazione grafica dopo un primo sguardo possa essere poi controllata e resa più precisa e particolareggiata con l'esame minuto della rappresentazione stessa. Nell'interpretazione dei particolari, si tenga sempre presente che i limiti indicati accanto a ciascuna gradazione di tinte o a ciascun tratteggio nelle scale cromatiche annesse ai cartogrammi non segnano limiti entro i quali *varia* la manifestazione del fenomeno in ciascuna delle zone così colorate o tratteggiate, bensì limiti entro i quali *è compresa* la manifestazione del fenomeno in ciascuna di esse. Così, nel nostro esempio, se troviamo la provincia di Belluno contrassegnata con la gradazione di tinta che indica densità di popolazione da 28 a 95 abitanti per kmq., non dobbiamo intendere che la densità della popolazione nelle varie parti della provincia varia fra 28 e 95, ma che la densità media nell'intera provincia è compresa fra 28 e 95: di fatto essa è di 66 abitanti per kmq., ma il cartogramma non ci può dare quest'indicazione precisa, ci dà solo l'indicazione largamente approssimativa ora esposta.

26. — Poichè le rappresentazioni grafiche non sono che traduzioni dei dati numerici, gli errori dai quali questi siano affetti si riflettono fedelmente in esse. Però nel diagramma una imprevista punta verso l'alto o verso il basso, nel cartogramma una brusca variazione di tinta in una zona uniformemente colorata, può dare pronto indizio all'occhio dell'esistenza di errori di osservazione, che sarebbero sfuggiti, o soltanto dopo laboriosi ragionamenti sarebbero stati sospettati, nell'esame dei dati numerici. Sebbene non giovino, come talvolta giova la media, ad eliminare o ad attenuare errori di os-

servazione, le rappresentazioni grafiche servono pertanto di ausilio per la scoperta e per la correzione di tali errori.

27. — I metodi di rappresentazione grafica riescono molto utili per la comparazione tra più serie statistiche.

Il cartogramma si presta bene alla comparazione fra un piccolo numero di serie: due o tre, quattro al massimo. Se si moltiplicano maggiormente i cartogrammi posti l'uno accanto all'altro, diviene molto difficile la comparazione tra essi.

Il diagramma può servire alla comparazione fra un numero anche maggiore di serie: in molti casi una decina di curve o di spezzate, che rappresentano l'andamento di altrettante funzioni statistiche d'una medesima variabile, possono essere comodamente raccolte e comparate in un sol quadro. Artifizi di esecuzione, come l'impiego simultaneo di più coppie di assi coordinati, rendono possibile il confronto tra un numero anche maggiore di serie.

L'istogramma serve anch'esso assai bene, in certi casi, alla comparazione fra più serie, come risulta anche da quanto abbiamo detto intorno alla sua applicazione al cartogramma.

Ovvia norma generale è quella che nell'esecuzione delle diverse rappresentazioni grafiche da confrontare reciprocamente siano adottati gli stessi criteri: altrimenti la comparazione può riuscire difficile e fallace, se non impossibile.

Indicazioni bibliografiche. — BRINTON W. C., *Graphic methods for presenting facts*, New York, Engineering Magazine Company, 1923. — RIGGLEMAN J. R. *Graphic methods for presenting business statistics*, London, Mc Graw-Hill, 1926. — RASERI E., *Atlante di demografia e geografia medica d'Italia*, Roma, Istituto geografico De Agostini, 1906. — BARATTA M. e VISINTIN L., *Atlante della produzione e dei commerci*, 2ª edizione, Novara, Istituto geografico De Agostini, 1929.

Quesiti ed esercizi: 1. — Che cos'è un diagramma? Come si costruisce? Quale campo è lasciato all'arbitrio nella costruzione del diagramma?

2. — Che cos'è un istogramma? Quali regole si seguono nella costruzione? Dove ha possibilità di manifestarsi l'abilità del costruttore?

3. — Che cos'è uno stereogramma? Perchè questo metodo è poco usato?

4. — Si esaminino le rappresentazioni grafiche contenute nell'ASI cercando d'intendere qual è il fine principale cui mira ciascuna di esse e quali sono gli aspetti del fenomeno rappresentato che appaiono in speciale rilievo nel diagramma o nel cartogramma.

5. — Quali delle suddette rappresentazioni possono essere interpretate a prima vista dal lettore poco esperto? Quali richiedono invece conoscenze tec-

niche e paziente studio? Quali sembrano meglio corrispondenti al fine di dare una immediata visione dell'insieme del fenomeno rappresentato e dei suoi particolari più importanti?

6. — In un sistema di coordinate cartesiane si segnino con una crocetta i punti che corrispondono alle seguenti combinazioni di valori dell'ascissa x e dell'ordinata y:

$$x = +1 \quad +5 \quad -2 \quad -4 \quad +3 \quad +7 \quad -6 \quad -10$$
$$y = +3 \quad -7 \quad -8 \quad +5 \quad +2 \quad -1 \quad -1 \quad +2$$

7. — Si classifichino le rappresentazioni grafiche dell'ASI secondo il tipo: diagrammi, istogrammi, stereogrammi, cartogrammi.

8. — Che cos'è un istogramma circolare? rettangolare? a canne d'organo?

9. — Si rappresentino in diagramma: *a*) la produzione del frumento in Italia negli ultimi venti anni; *b*) la produzione del frumento nei vari compartimenti italiani (regioni) in un dato anno. Si traccino i segmenti proporzionali ai dati, se ne riuniscano poi i vertici mediante una linea spezzata. Che cosa significa questa spezzata nel diagramma *a*? nel diagramma *b*?

10. — Che cos'è un diagramma logaritmico? Quando conviene adottarlo?

11. — Si rappresenti in diagramma la frequenza delle nascite per 1000 abitanti, in Italia, negli ultimi cinquant'anni. Si descrivano le tendenze della natalità, quali sono messe in evidenza dal diagramma. Si esegua lo stesso diagramma mutando le scale delle ascisse e delle ordinate: si raddoppi o si dimezzi il segmento unitario nell'una e nell'altra scala; si raddoppi nell'una e si dimezzi nell'altra; si dimezzi nell'una e si raddoppi nell'altra. Come si modifica l'impressione di chi guarda il diagramma, secondo le diverse modificazioni arrecate alle scale?

12. — Come influisce, in generale, sull'impressione di chi guarda un diagramma, una modificazione della scala delle ordinate? una modificazione della scala delle ascisse?

13. — Si rappresenti in diagramma il valore delle merci importate in un dato anno in Italia, suddivise nelle «categorie» della tariffa doganale. Conviene in questo caso adottare la scala logaritmica?

14. — Si rappresenti in diagramma logaritmico la circolazione monetaria complessiva, secondo i dati dell'ASI, negli ultimi cinquant'anni. Quali vantaggi e quali inconvenienti presenta questo diagramma in confronto al diagramma ordinario? Si confrontino i due diagrammi e si indichi quale di essi è preferibile.

15. — Si rappresenti in diagramma ordinario e in diagramma logaritmico lo sviluppo della rete ferroviaria italiana negli ultimi cinquant'anni. Perchè e come le impressioni desunte dal secondo diagramma possono integrare quelle desunte dal primo?

16. — Che cos'è un diagramma a doppia scala logaritmica? Quando conviene adottarlo?

17. — Si rappresenti in diagramma: *a*) la distribuzione dei comuni italiani secondo il numero degli abitanti presenti in ciascun comune; *b*) la distribuzione della popolazione presente in Italia per categorie di popolazione dei co-

muni. Quali difficoltà si presentano nell'eseguire tali diagrammi sulla base dei dati dell'ASI? Sono superabili, e come?

18. — Come si può rappresentare una funzione statistica in coordinate cartesiane?

19. — Si cerchino nei capitoli « Territorio e popolazione, igiene e sanità » dell'ASI esempi di funzioni statistiche utilmente rappresentabili in diagramma.

20. — A quali scopi possono essere impiegati i colori nella costruzione di diagrammi? di istogrammi?

21. — Come si può rappresentare in diagramma la distribuzione delle stature degli iscritti di leva, qual è indicata nell'ASI? Quali difficoltà si presentano e come si possono superare?

22. — Si rappresentino in un solo diagramma i dati sulle temperature mensili medie, minime e massime per uno degli osservatorii considerati nel capitolo « Climatologia » dell'ASI.

23. — In qual modo si potrebbe rappresentare in un solo diagramma l'andamento nel tempo del valore delle esportazioni e la sua suddivisione secondo i paesi di destinazione delle merci?

24. — Come si può rappresentare una funzione statistica in coordinate polari?

25. — Si rappresentino in coordinate polari i dati sulle frequenze annuali delle varie direzioni del vento per uno degli osservatorii considerati nel capitolo « Climatologia » dell'ASI.

26. — Si rappresenti in coordinate cartesiane e in coordinate polari il traffico delle ferrovie italiane nei singoli mesi di un dato anno. Si confrontino le due rappresentazioni: quale di esse è più adatta a mettere in risalto le caratteristiche stagionali del traffico?

27. — Si rappresenti in istogramma a canne d'organo, in istogramma a quadrati e in istogramma a cerchi la distribuzione della produzione vinicola italiana di un dato anno fra i vari compartimenti. Quale di queste forme di rappresentazione riesce più efficace? Quale si presterebbe meglio per l'esecuzione di un cartogramma ad istogrammi?

28. — Si rappresentino graficamente i dati sulla portata dei principali fiumi italiani nelle diverse stagioni, contenuti nel capitolo « Climatologia » dell'ASI. Come si potrebbero rappresentare gli stessi dati in un unico cartogramma a nastro?

29. — Volendo rappresentare l'andamento dei prezzi di un prodotto agricolo nei singoli mesi della campagna agraria, conviene tracciare un diagramma in coordinate cartesiane? o in coordinate polari? o un istogramma circolare diviso in settori?

30. — Nella compilazione di istogrammi circolari è il raggio che dev'essere proporzionale al dato statistico?

31. — Si esaminino i diagrammi contenuti nell'atlante del Raseri. Si può muovere qualche critica alla loro compilazione? Quale errore è intervenuto nell'esecuzione della tavola 12? Quali conseguenze ne derivano nell'aspetto delle curve raffigurate? Quali indicazioni si desumono dalla suddetta tavola? Come si può spiegare l'andamento a zig-zag di alcune curve ivi contenute? (si pensi a possibili errori di rilevazione dell'età dei censiti). Sapendo che il Ve-

neto è la regione italiana dove la statura media è più alta, la Sardegna quella dov'è più bassa, si correggano le indicazioni errate apposte alla tavola 20. È completa la tavola 43, o vi manca una parte del grafico? Perchè nella tavola 7 la curva nera ha andamento irregolare e quella rossa lo ha regolare?

32. — L'istogramma contenuto nella tavola 5 dell'atlante Raseri corrisponde bene al fine delle rappresentazioni grafiche? Tenete nota del tempo che vi occorre per interpretarne il significato; e poi provatevi a fare un confronto fra due regioni, cercando di giungere a un giudizio comparativo, senza il sussidio di strumenti di misura.

33. — Si traducano in istogramma i dati sulle popolazioni delle varie grandi città italiane in uno stesso censimento, o di una stessa città nei successivi censimenti (ASI 1930, capitolo « Grandi città »).

34. — Si interpreti l'istogramma contenuto nella tavola 10 dell'atlante del Raseri.

35. — Si esaminino e si riassumano i principali esempi di errori nella compilazione di diagrammi e di istogrammi, raccolti nel libro del Brinton.

36. — Si rappresenti in istogramma la consistenza delle varie biblioteche italiane, per categorie di opere, impiegando colori o tratteggi per distinguere le varie categorie.

37. — Si traduca la serie dei dati sul corso del cambio su Londra — medie annuali per gli ultimi cinquant'anni — in diagramma lineare a segmenti, in diagramma lineare a spezzata, in cartogramma a nastri, in istogramma a canne d'organo, in istogramma a quadrati. Quale di queste forme è più adatta a mostrare l'andamento del cambio?

38. — Si tracci la curva della produzione italiana dei vari minerali in un dato anno.

39. — Si rappresenti con istogrammi circolari suddivisi in settori la produzione totale dei cereali e la produzione di ogni singolo cereale (frumento, segala, orzo, avena, granturco, riso) in ciascun compartimento italiano.

40. — Si rappresenti con istogrammi circolari suddivisi in settori il numero delle centrali elettriche (suddivise in idriche e termiche) e la potenza installata nei vari compartimenti italiani.

41. — Assumendo raggi proporzionali ai dati da rappresentare, si traducano in istogrammi circolari i dati per compartimenti sull'ammontare dei depositi presso le casse di risparmio postali.

42. — Che cos'è un cartogramma in generale? Quali sono i vari tipi di cartogrammi?

43. — Si rappresenti in cartogramma ad istogrammi circolari il traffico delle varie linee aeree italiane.

44. — Si rappresenti in cartogramma a nastri il traffico dei vari aeroporti italiani.

45. — Si ricerchino nel capitolo « Territorio e popolazione » dell'ASI i casi nei quali è applicabile il cartogramma ad istogrammi circolari, e nel capitolo « Viabilità » i casi nei quali è applicabile il cartogramma a nastri.

46. — Con quali diversi criteri si può determinare la corrispondenza fra tinte e numeri in un cartogramma a tinte graduali? Perchè in pratica riesce difficile seguire rigorosamente l'uno o l'altro dei criteri indicati nel testo? Come si rimedia?

47. — Si rappresentino in cartogramma a punti, in cartogramma ad istogrammi e in cartogramma a tinte graduali: *a*) i dati sulle nascite nelle diverse provincie italiane; *b*) i dati sulla natalità (nati per 1000 abitanti) nelle provincie stesse. Qual è il procedimento di rappresentazione più adatto nel caso *a*? nel caso *b*?

48. — Si esaminino criticamente i cartogrammi contenuti nell'atlante del Raseri. Per rappresentare i vari gradi d'intensità di un fenomeno appare più efficace il metodo delle tinte graduali seguito nelle tavole 2 e 3 o quello delle tinte diverse seguito nelle tavole 54 e 56? Perchè nel determinare la corrispondenza fra tinte e numeri sono stati seguiti criteri diversi nei vari grafici; e qual è il criterio dominante? Quali sono gli inconvenienti che derivano dai criteri adottati; quali i vantaggi?

49. — Come si potrebbe rappresentare sopra la carta topografica di una grande città la distribuzione degli infortuni stradali? quella dei casi accertati d'una malattia infettiva? l'intensità del traffico nelle diverse vie? l'afflusso dei viaggiatori alle diverse fermate tramviarie?

50. — Si esaminino criticamente i diagrammi, gli istogrammi e i cartogrammi contenuti nell'ASI. Vi si riscontrano errori? imperfezioni? Si suggeriscano desiderabili miglioramenti.

51. — Si comparino mediante cartogrammi le distribuzioni regionali delle seguenti produzioni agrarie o d'industrie agricole: frumento, olio d'oliva, vino, foraggi, agrumi, bozzoli; e si espongano le caratteristiche di ciascuna distribuzione.

52. — Si comparino mediante diagrammi gli andamenti negli ultimi cinquant'anni delle seguenti produzioni: frumento, vino, zolfo, acciaio.

53. — Si compari mediante cartogrammi la distribuzione regionale della natalità con quella della mortalità.

54. — Si compari mediante diagrammi l'andamento della natalità e della mortalità negli ultimi cinquant'anni.

55. — Si rappresenti in cartogramma il traffico dei vari porti italiani in un determinato anno.

56. — Si rappresenti in un unico cartogramma a tinte graduali il rendimento medio per ettaro della coltura del frumento nelle varie regioni nel 1928 e nel 1929.

57. — Si esaminino criticamente i cartogrammi dell'atlante del Baratta, indicando eventuali errori od imperfezioni.

CAPITOLO XII.

La rappresentazione analitico-sintetica della serie statistica mediante l'interpolazione.

Sostituzione di una curva continua alla rappresentazione discontinua ottenuta mediante la traduzione in diagramma di una serie statistica — Sostituzione di una curva regolare a quella così tracciata: cenni sui fini e sui criteri del procedimento; esempio di applicazione — Contrapposizione fra questo procedimento e quello della media — Rappresentazione di una curva geometrica regolare mediante un'equazione: alcuni esempi — Criteri per la scelta della forma della curva regolare da sostituire alla spezzata o curva desunta dall'osservazione — Procedimenti sistematici per la sostituzione — Definizione dell'interpolazione: varie condizioni cui corrispondono i diversi metodi d'interpolazione: condizioni di equilibrio, di equivalenza, di accostamento; cenni sui metodi dei minimi quadrati, delle somme, delle aree, dei momenti, di Cauchy — Considerazioni riassuntive sull'interpolazione analitica — L'estrapolazione — L'interpolazione grafica — Influenza degli errori di osservazione sui risultati dell'interpolazione — Applicazioni del procedimento interpolatorio a vari fini — Il procedimento della media mobile e quello della perequazione come surrogati dell'interpolazione.

1. — Una *curva statistica* è, a stretto rigore, la rappresentazione grafica di una funzione statistica continua. Ma, come abbiamo avvertito nel precedente capitolo, solo eccezionalmente l'osservazione ci dà il tracciato della curva; normalmente essa ci dà soltanto alcuni punti della curva stessa, così che sorge il problema della ricostruzione dell'andamento della curva nell'intervallo considerato, da eseguire sulla scorta dei punti noti.

Abbiamo anche avvertito come, in via approssimativa, spesso si sostituisca una funzione continua ad una funzione discontinua di una variabile continua: si parla, per esempio, di una « curva di sopravvivenza », benchè il numero dei sopravviventi, in funzione dell'età, varii per unità intere e non per quantità infinitesime. La funzione di sopravvivenza è in realtà una funzione discontinua (numero dei superstiti all'età x) di una variabile continua (l'età), ma poichè a brevissimi intervalli d'età corrispondono variazioni relativamente piccolissime del numero dei superstiti, non si commette un errore sensibile nel sostituirle una funzione continua. Graficamente si surroga una curva a quella che in realtà sarebbe una gradinata, dai gradini molto bassi e molto stretti.

Si fa di più talvolta: si sostituisce, cioè, sempre in via di approssimazione, una funzione continua ad una funzione discontinua di una variabile discontinua: così si parla correntemente di « curva di distribuzione dei redditi » sebbene il numero dei possessori di redditi corrispondente ai successivi valori del reddito varii per unità intere, e sebbene due redditi disuguali fra loro non possano mai differire di una quantità infinitesima: si riduca pure la differenza ad una lira o ad un centesimo, è piccola, ma finita. E si parla di una « curva di distribuzione delle famiglie secondo il numero dei componenti » benchè tanto il numero delle famiglie quanto quello dei componenti non possano variare che di intere unità.

La funzione continua è, dall'aspetto matematico, più maneggevole, e suscettibile di analisi e di elaborazioni alle quali non si presta o mal si presta la funzione discontinua: perciò è comoda, quando riesce possibile, la surrogazione dell'una all'altra; la buona approssimazione che in molti casi si raggiunge ne giustifica l'applicazione nel campo statistico. Ma per eseguire tale applicazione bisogna risolvere un problema analogo, non tuttavia identico, a quello posto al principio di questo paragrafo: il problema, cioè, di costruire, sulla scorta dei punti corrispondenti a valori dati di una funzione discontinua, una curva corrispondente ad una funzione continua atta a sostituire con buona approssimazione la prima.

Nonostante il differente fondamento teorico, i due problemi nella pratica si confondono in uno solo: quello di tracciare una curva continua della quale siano dati soltanto alcuni punti. Il problema qui posto in forma grafica si può tradurre in forma numerica: ma allora non si può tendere alla determinazione degli infiniti valori corrispondenti agli infiniti punti di un ramo, per quanto breve, di curva; si tratta invece di determinare un numero maggiore o minore di tali valori, per lo più scelti in modo che corrispondano a ordinate equidistanti. L'analisi della soluzione grafica ci può agevolare la ricerca della soluzione aritmetica.

Cominciamo col congiungere fra loro i punti noti, ordinatamente a due a due, mediante segmenti di retta, in modo da costruire una linea spezzata. Ciò equivale a supporre che la variazione della funzione sia ripartita in modo uniforme in ciascuno degli intervalli delimitati dai valori noti: numericamente, cioè, equivale a far corrispondere una progressione aritmetica di valori ad una successione di ordinate equidistanti comprese fra due ordinate note. Talvolta l'andamento della spezzata suggerisce immediatamente l'idea di sostituire una vera e propria curva alla spezzata stessa, perchè la curva appare più adatta a rappresentare la graduale variazione della differenza fra successive ordinate equidistanti (e qui notiamo che la spezzata è designata in statistica come « curva » appunto perchè spesso costituisce la rappresentazione approssimativa di una vera e propria curva). Altre volte l'esame del diagramma induce a preferire la spezzata alla curva, perchè l'andamento del fenomeno è così capriccioso od accidentato che la curva tracciata per i punti noti non avrebbe il vantaggio di offrire una rappresentazione maggiormente approssimata di quella che la spezzata offre, mentre avrebbe lo svantaggio di confondere in sè i punti noti, che la spezzata mette in risalto.

Il procedimento della spezzata ha un preciso riscontro numerico, come risulta da quanto è stato detto or ora. Se, conoscendo il numero degli abitanti di un paese secondo più censimenti successivi, abbiamo tracciato la linea spezzata che indica lo sviluppo della popolazione nel tempo, possiamo leggere sul grafico, col sussidio delle scale, il numero degli abitanti in qualsiasi momento intermedio fra due consecutivi censimenti. Questo numero, che in generale solo approssimativamente riusciremo a leggere sul grafico, ci è dato esattamente (esattamente, intendasi, non rispetto alla realtà, ma rispetto alla nostra ipotesi di uniforme ripartizione nel tempo della variazione di popolazione accertata fra due censimenti consecutivi) dalla risoluzione di una semplice proporzione aritmetica.

Al tracciamento, per i punti noti, di una curva, che ad occhio si cerca di rendere il più regolare possibile nel suo andamento ma che non corrisponde ad un preciso criterio matematico, non fa riscontro invece un procedimento numerico sistematico. Le ordinate intermedie fra quelle note si possono leggere sul grafico; si possono anche determinare numericamente per via di tentativi, cercando di ottenere successioni di numeri procedenti con una certa regolarità: ma regolarità nel senso empirico ed intuitivo della parola, e non in senso precisamente ed univocamente definito.

A parità di ogni altra condizione, l'andamento della funzione statistica si ricostruisce con approssimazione tanto maggiore quanto maggiore è il numero dei punti noti. Lo abbiamo già osservato nel trattare della rappresentazione di una serie per mezzo di più medie dello stesso genere, quando abbiamo visto che col moltiplicare il numero delle medie simultaneamente impiegate — col passare, per esempio, dai quartili ai decili, dai decili ai centili, ecc. — si può accrescere gradualmente l'approssimazione della rappresentazione. Quanto più i punti sono vicini tra loro, tanto meglio ci danno l'immagine della funzione; ma qui calza una osservazione: se l'andamento della funzione è molto regolare, anche punti relativamente lontani possono darcene una buona visione; se l'andamento è irregolare, anche punti relativamente vicini possono darcene un quadro inadeguato. Si consideri lo sviluppo di una popolazione civile, in epoche normali, attraverso il tempo: censimenti eseguiti ogni dieci anni potranno bastare a tracciarne con soddisfacente approssimazione la curva; si considerino invece, sempre attraverso il tempo, le variazioni nella forza di un esercito operante in guerra: stati della forza redatti ogni mese potranno essere assolutamente insufficienti.

2. — Chi guardi la curva o la spezzata rappresentatrice di una funzione statistica in certi casi riceve l'impressione che le variazioni del fenomeno rappresentato siano del tutto irregolari. In altri casi, attraverso ondulazioni od oscillazioni, più o meno ampie e profonde, scorge un andamento ben definito, una forma regolare: la curva statistica gli appare quasi l'immagine, tracciata dalla mano incerta di un fanciullo o dalla mano tremante di un vecchio, di una curva geometrica regolare, o almeno di una curva regolare all'occhio, se pure non precisamente definibile mediante una proprietà comune dei suoi punti. Si presenta ovvia l'idea di surrogare la rappresentazione approssimativa data dalla curva regolare alla rappresentazione precisa data dalla curva statistica, all'intento di far meglio risaltare i tratti principali dell'andamento del fenomeno mercè l'eliminazione dei tratti accessori, e di ottenere una rappresentazione semplificata sì, ma in compenso più facilmente traducibile in parole o in formola.

Supponiamo che in un paese la proporzione dei coscritti analfabeti fosse in una certa epoca di 72 per 100; dieci anni dopo di 61, vent'anni dopo di 54, trent'anni dopo di 40, quarant'anni dopo di 33, cinquant'anni dopo di 22. Tracciando la spezzata nella quale si può tradurre questa serie di dati, vediamo che il suo andamento non si discosta molto da quello di una linea retta (che è la forma più semplice di linea regolare per l'occhio e quella che più prontamente viene riconosciuta). Col sussidio della riga, possiamo provare a tracciare un segmento di retta idoneo a surrogare la nostra spezzata. Di queste rette ne esistono infinite; ma se dopo una serie di tentativi ci fermiamo ad una soluzione soddisfacente e la confrontiamo con la soluzione ritenuta soddisfacente da un altro operatore autonomo, probabilmente troveremo che le due rette hanno andamento molto simile tra loro: sono entrambe discendenti e l'inclinazione dell'una non differisce molto da quella dell'altra. Consideriamo una fra le tante possibili rette: per esempio quella che passa per i due punti noti di ordinata 72 e 22. E chiediamoci: perchè, dopo vari tentativi, questa retta ci è apparsa meglio idonea di altre sperimentate a rappresentare in modo semplificato il variare dell'analfabetismo? Certo ci allontana alquanto dalla realtà: dieci anni dopo l'inizio dell'osservazione l'analfabetismo non è di 62 per 100 come indica la nostra retta ma di 61, vent'anni dopo non è di 52 ma di 54, trent'anni dopo non è di 42 ma di 40, quarant'anni dopo non è di 32 ma di 33. Però le differenze tra i dati rilevati e quelli che ad essi sostituiamo sono relativamente non grandi, ossia i dati surrogati si accostano abbastanza a quelli da surrogare; e d'altro canto tali differenze sono in parte positive e in parte negative, anzi le une compensano le altre, così che la somma dei dati surrogati uguaglia quella dei dati da surrogare. Valendoci di espressioni analoghe a quelle che abbiamo adoperato a proposito delle medie, possiamo dunque dire che nella surrogazione della retta alla spezzata abbiamo soddisfatto una condizione di accostamento, una di equilibrio ed una di equivalenza.

3. — Sui criteri di equilibrio e di equivalenza non occorre fermarsi: le condizioni soddisfatte dalla nostra retta del precedente paragrafo sono in tutto simili a quelle che, nel campo delle medie, soddisfa la media aritmetica; e sarebbero state soddisfatte anche se alla spezzata desunta dalle osservazioni avessimo sostituito una retta orizzontale avente ordinata costante uguale appunto a 47, media aritmetica dei dati noti. La differenza tra il procedimento della media e quello ora seguito consiste in ciò: che nel primo siamo vincolati a surrogare alla spezzata desunta dalle osservazioni una linea retta orizzontale, così che la rappresentazione è necessariamente statica; nel secondo invece siamo liberi di scegliere quella retta che per la sua inclinazione, o quella curva che per la sua forma, meglio si adatta all'andamento della spezzata, così che la rappresentazione non è più statica ma dinamica, e può riprodurre con approssimazione assai grande i dati dall'osservazione. Nel nostro esempio, se ci contentiamo di sostituire la retta orizzontale alla spezzata, incontriamo uno scostamento medio assoluto di 15,3; se le sostituiamo la retta discendente, lo scostamento medio assoluto si riduce ad 1: il vantaggio che si consegue nell'accostamento della rappresentazione semplificata alla realtà è dunque molto grande.

Ma la media, si osserverà, ha il pregio incomparabile di riassumere in un sol numero l'intera serie; la retta inclinata invece è una rappresentazione grafica, che non possiede simile virtù sintetica. Si può replicare che alla rappresentazione grafica della nostra retta è facile far corrispondere una rappresentazione analitica (adoperiamo qui l'aggettivo « analitica » per indicare una rappresentazione compiuta coi metodi dell'analisi matematica) assai semplice. Questa retta, infatti, ci indica la variazione dell'analfabetismo y in funzione del tempo x; all'inizio delle osservazioni (cioè per $x = 0$) è y, analfabetismo, uguale a 72; per $x = 10$ è $y = 62$; per $x = 20$ è $y = 52$, ecc. Esaminando la successione delle coppie di valori corrispondenti delle due variabili, si vede subito la relazione che li lega: a valori di x che si seguono in progressione aritmetica corrispondono valori di y anch'essi costituenti una progressione aritmetica; e precisamente ad ogni valore di x corrisponde un valore di y uguale a $(72 - x)$. Possiamo esprimere in forma generale questa relazione scrivendo: $y = 72 - 1\ x$. Ed ecco che la nostra retta, se non è contenuta in un sol numero come la media, è tuttavia rappresentata in modo completo ed univoco da due numeri soli: 72 che indica il valore di y corrispondente al valore 0 di x, e -1 che indica il rapporto costante tra la variazione di y e la corrispondente variazione di x, cioè misura la variazione di y corrispondente alla variazione unitaria di x. Questi due numeri sono caratteristici della nostra retta e di essa sola: ogni retta che nel punto di ascissa 0 abbia ordinata 72, nel punto di ascissa 1 abbia ordinata 71, nel punto di ascissa 2 abbia ordinata 70, ecc. coincide con essa, anzi non è altro che la nostra retta.

Sostituendo la retta inclinata, invece che la retta orizzontale (media), alla spezzata originaria, abbiamo dunque lo svantaggio di adottare una rappresentazione caratterizzata da due numeri invece che da uno solo, ma abbiamo il vantaggio di ottenere una rappresentazione molto più approssimata, perchè dinamica invece che statica. Nel nostro esempio pare indubitabile che il vantaggio superi di gran lunga lo svantaggio.

4. — La possibilità di essere rappresentata mediante un'equazione che esprime una relazione fra ordinata ed ascissa non è un privilegio della nostra retta: qualsiasi retta può essere rappresentata da un'equazione della forma $y = a\,x + b$, che ne esprime la proprietà caratteristica, cioè la costanza del rapporto tra la variazione dell'ordinata e la corrispondente variazione dell'ascissa. In questa equazione, y ed x rimangono indeterminate come indicazioni generiche del valore dell'ordinata e del corrispondente valore dell'ascissa, mentre a e b sono due numeri determinati caratteristici della particolare retta che l'equazione rappresenta (nel nostro esempio è $a = -1$,

$b = 72$). Poichè per $x = 0$ risulta $y = b$, è chiaro che la costante b rappresenta il valore dell'ordinata corrispondente al valore 0 dell'ascissa: nel nostro caso possiamo dire il valore dell'ordinata iniziale. E poichè se x varia di 1, y varia di a, è chiaro che la costante a rappresenta la variazione costante dell'ordinata che corrisponde alla variazione unitaria dell'ascissa, ossia la ragione della progressione aritmetica dei valori delle ordinate che corrisponde alla progressione aritmetica dei valori interi delle ascisse. Se a è positiva, la retta è ascendente; se è nulla, la retta è orizzontale; se è negativa, la retta è discendente: tale è il significato del segno di a; alla maggiore o minore grandezza di questa costante corrisponde poi, com'è ovvio, una maggiore o minore inclinazione della retta sull'asse delle ascisse. Nel nostro esempio, se invece di considerare le percentuali degli analfabeti avessimo considerato quelle, complementari, degli alfabeti, la retta surrogata alla spezzata avrebbe avuto per equazione $y = 28 + 1\,x$, cioè sarebbe stata ascendente invece che discendente; se partendo da dati diversi avessimo trovato $y = 28 + 0{,}5\,x$ avremmo espresso in questa relazione un incremento dell'alfabetismo della metà più lento; se avessimo trovato $y = 28 + 1{,}2\,x$ avremmo espresso un incremento di due decimi più rapido. I numeri a e b, costanti per una data retta ma variabili da retta a retta, si dicono *parametri* della retta.

Non la sola retta, ma ogni curva geometrica regolare può essere rappresentata mediante una equazione, che esprime una proprietà comune ai suoi punti: proprietà nella quale appunto consiste la regolarità della curva. Se consideriamo, per esempio, un arco di cerchio, vediamo facilmente che un raggio costituisce l'ipotenusa di un triangolo rettangolo i cui cateti sono l'ascissa e l'ordinata del punto della circonferenza cui mette capo il raggio; e quindi se assumiamo il centro del cerchio a origine degli assi coordinati, per il teorema di Pitagora, indicando con r il raggio, possiamo scrivere: $r^2 = x^2 + y^2$, ossia: $y^2 = r^2 - x^2$. Il quadrato dell'ordinata è uguale al quadrato del raggio, che è il parametro caratteristico del cerchio, diminuito del quadrato dell'ascissa.

5. — Incontrando ora l'equazione di una curva non saremo imbarazzati ad interpretarla. Dalla relazione $y = b\,a^x$, dove a e b rappresentano numeri positivi, ed a è supposto diverso da 1, conchiuderemo che la curva così rappresentata è caratterizzata dal variare dell'ordinata secondo una progressione geometrica col variare dell'ascissa secondo una progressione aritmetica. Il parametro b ci indica

il valore dell'ordinata corrispondente all'ascissa 0, il parametro a non è altro che la ragione della progressione geometrica delle ordinate corrispondente alla progressione aritmetica dei valori interi delle ascisse. Traducendo in logaritmi la precedente equazione abbiamo: $\log y = \log b + x \log a$; da questa formola è facile vedere che, tradotta in diagramma logaritmico, la curva in esame (detta « curva logaritmica ») dà luogo ad una retta caratterizzata dai parametri $\log a$ e $\log b$.

E se incontriamo l'equazione $y = a x^3 + b x^2 + c x + d$, intenderemo che rappresenta una curva (« curva parabolica di terzo grado ») la cui ordinata è uguale alla somma dei prodotti di un parametro a per il cubo dell'ascissa, di un parametro b per il quadrato dell'ascissa, di un parametro c per l'ascissa, e di un parametro d per l'unità. Col variare dell'ascissa secondo una progressione aritmetica le differenze delle differenze delle corrispondenti ordinate variano secondo una progressione aritmetica.

Riuscirà meno facile intendere di primo acchito il significato dell'equazione $y = a : e^{\frac{x^2}{b}}$, dove la costante $e = 2{,}718 \ldots.$ rappresenta la base dei logaritmi naturali ed i parametri a e b rappresentano numeri positivi. Ma una breve riflessione basta a mostrarci che il valore dell'espressione $e^{\frac{x^2}{b}}$ è sempre positivo; è il medesimo in corrispondenza ad un determinato valore di x, sia questo positivo o negativo; parte da un minimo uguale all'unità per $x = 0$ e aumenta con crescente rapidità col crescere in valore assoluto di x, tendendo ad infinito col tendere ad infinito di x. Perciò l'ordinata della nostra curva assume un valore massimo uguale al parametro a per $x = 0$; così a destra come a sinistra di questa ordinata massima l'ordinata decresce con rapidità relativa crescente al crescere di x (ma con rapidità assoluta crescente fino ad un flesso e indi decrescente), e tende a 0 col tendere di x ad infinito. La curva rappresentata dalla nostra equazione è dunque simmetrica rispetto all'ordinata massima a; di forma campanulare; aperta (« asintotica rispetto all'asse delle ascisse », in linguaggio matematico), perchè per quanto cresca x non si annulla mai il corrispondente valore di y. È la *curva degli errori*, detta così perchè descrive con buona approssimazione la distribuzione per grandezza di errori casuali relativamente piccoli che intervengono in ripetute misurazioni di una data grandezza fisica. La stessa curva descrive anche, con approssimazione in molti casi sufficiente, la distribuzione per grandezza di caratteri quantitativi di esseri viventi (per esempio: distribuzione delle stature di coetanei). Il parametro a si può determinare dividendo il numero dei dati noti (numero delle stature misurate) per lo scostamento medio quadratico dei dati stessi (stature) dalla loro media aritmetica (statura media) moltiplicato per $\sqrt{2\pi} = 2{,}507$; il parametro b si può determinare moltiplicando lo stesso scostamento medio quadratico per $\sqrt{2} — 1{,}414$. In pratica, dunque, basta calcolare lo scostamento medio quadratico per poter determinare poi facilmente i valori dei due parametri.

6. — Nel nostro esempio dei paragrafi 2-3 abbiamo surrogato alla spezzata desunta dall'osservazione una retta discendente, e abbiamo ottenuto una buona approssimazione. Ma in moltissimi casi una spezzata mostra andamento regolare, non però tale da potersi rappresentare con sufficiente approssimazione mediante una retta discendente od ascendente: così per esempio se per un certo tratto la spezzata discende e per un altro tratto ascende, o viceversa; oppure se la spezzata ascende con inclinazione gradualmente crescente o discende con inclinazione gradualmente decrescente (in modo da accostarsi all'immagine di una curva concava verso l'alto); oppure ancora se essa ascende con inclinazione decrescente o discende con inclinazione crescente (in modo da accostarsi all'immagine di una curva convessa verso l'alto). Conviene allora sostituire alla spezzata non una retta, che darebbe una rappresentazione contrastante con la realtà, ma un'altra curva geometrica regolare il cui andamento si accosti il più possibile a quello della spezzata. Quale curva? All'occhio è facile distinguere se l'andamento di una spezzata si accosti a quello d'una retta, ma è sempre meno facile, spesso difficile, e talora impossibile, discernere se si accosti maggiormente a quello di un arco di circonferenza, ovvero a quello di un ramo di parabola, o piuttosto a quello d'un ramo d'iperbole o di altro tipo di curva. Certo il matematico esclude a colpo d'occhio molte ipotesi; ma anch'egli può restare incerto tra parecchie sostituzioni alla spezzata di curve geometriche regolari. Diciamo *parecchie* e non *infinite*, perchè la curva da sostituire alla spezzata non solo deve avere andamento simile a quello della spezzata stessa ma dev'essere anche rappresentata da un'equazione semplice: caratterizzata cioè da un piccolo numero di parametri; altrimenti invece di una rappresentazione semplificata ci darebbe una rappresentazione complicata della serie statistica.

Bisogna dunque anzitutto esaminare con cura il diagramma e cercar di tradurre in forma ben determinata quell'impressione di regolarità che esso ci ha dato (se non ce l'avesse data, non avremmo pensato a sostituire alla spezzata una curva geometrica regolare); ove a ciò non si riesca, esaminare pazientemente i dati numerici. Se, in corrispondenza a successivi valori della variabile indipendente costituenti una progressione aritmetica, i valori della variabile dipendente si seguono press'a poco in progressione aritmetica, proveremo la retta; se si seguono press'a poco in progressione geometrica, proveremo la curva logaritmica; se le loro differenze formano press'a poco una progressione aritmetica, proveremo una curva parabolica di secondo grado ($y = a x^2 + b x + c$). Insomma, esaminando i dati, i loro rapporti, le loro differenze, le differenze delle differenze, ecc., ricercheremo se la successione dei valori noti ha « press'a poco » una certa proprietà, una certa regolarità, e proveremo poi a sostituirle una funzione analitica tale che i valori di essa corrispondenti a quelli noti abbiano « rigorosamente » quella stessa proprietà, espressa nella sua equazione. Sull'interpretazione di questo « press'a poco » è arduo dare direttive generali; soltanto la lunga pratica può insegnarne i limiti, che variano da caso a caso. Nel nostro esempio di dianzi, i dati sull'analfabetismo sono rappresentati con buona approssimazione da una retta: in altri termini, l'ipotesi di una velocità costante nella diminuzione della percentuale degli analfabeti attraverso il tempo corrisponde da vicino alla realtà. Ma avrebbe il lettore a prima vista giudicato che i numeri 72, 61, 54, 40, 33, 22 costituiscono « press'a poco » una progressione aritmetica? avrebbe cioè affermato che le loro differenze 11, 7, 14, 7, 11 sono « press'a poco » costanti, che cioè è press'a poco costante la velocità media decennale di diminuzione dell'analfabetismo?

Si badi che mentre basta, in generale, un'approssimativa uguaglianza di proprietà tra la funzione surrogata e quella da surrogare, bisogna però che tale uguaglianza sussista non solo nei punti noti, ma anche negli intervalli fra essi; altrimenti la rappresentazione riesce infedele e fallace.

Accanto al criterio sostanziale dell'analogia tra le proprietà, nell'intervallo che si considera, della funzione surrogata e quelle della funzione statistica da surrogare, ha spesso una parte decisiva nella scelta della funzione surrogata la semplicità dell'espressione analitica di essa: semplicità che rende meno laboriosi i calcoli occorrenti per la determinazione dei valori dei parametri e che rende meglio utilizzabile l'equazione della curva interpolata per i confronti con altre curve.

Avvertasi che la scelta della funzione surrogata va eseguita con riguardo alla forma che assumerebbe la funzione statistica se fosse immune da errori di osservazione, perchè si mira alla rappresentazione dei dati esatti e non degli errori coi quali essi sono noti. Perciò talvolta si giudica ottima una rappresentazione che pure si allontana molto dalla curva statistica, ritenendo che in parte preponderante gli scostamenti indichino errori di osservazione dai quali si presumono affetti i dati noti.

7. — Ora che sappiamo come una curva geometrica regolare sia rappresentata da un'equazione, possiamo intendere come il procedimento della sostituzione della curva stessa alla spezzata possa essere reso sistematico: corrispondente cioè a condizioni precisamente definite. Nell'esempio svolto abbiamo trovato, per così dire a tentoni, una retta che *si accosta molto* alla spezzata desunta dall'osservazione; ma non potremmo proporci di trovare la retta la quale *si accosta più di ogni altra* alla spezzata stessa? Graficamente il problema non si può risolvere in modo preciso: possiamo solo, fra numerosi tentativi, scegliere il meglio riuscito, ma nulla ci garantisce che esso sia l'ottimo. Coi procedimenti dell'analisi matematica, invece, possiamo porre condizioni precise: per esempio cercare quella retta per la quale la somma dei quadrati delle differenze (*scostamenti*) dei dati surrogati dai dati osservati sia la minima possibile: sia inferiore, cioè, a quella che si avrebbe adottando qualsiasi altra retta. (Perchè, si chiederà, porre la condizione di minimo per la somma dei quadrati e non per la somma dei valori assoluti degli scostamenti? Per l'unica ragione che questa seconda condizione, logicamente preferibile per la sua maggiore semplicità e per il suo più intuitivo significato, matematicamente non si può tradurre in un sistema di equazioni in generale risolubile senza difficoltà, come la prima).

Indicando con u i valori noti della funzione statistica (nel nostro esempio, le percentuali dell'analfabetismo ricavate dall'osservazione) e con y i corrispondenti valori della funzione che ad essa surroghiamo ($y = a\,x + b$), possiamo tradurre la condizione che la somma dei quadrati delle differenze $(u - y)$ sia la minima possibile, nell'altra che sia la minima possibile la somma dei quadrati delle espressioni $(u - a\,x - b)$. Chi abbia qualche nozione di calcolo infinitesimale intende che, affinchè sia soddisfatta tale condizione, dev'essere soddisfatta l'altra condizione che siano nulle le derivate parziali del primo ordine della funzione « somma dei quadrati delle differenze » rispetto alle due variabili a e b (cioè ai due parametri da determinare). Devono essere, cioè, soddisfatte simultaneamente dai valori di a e b le condizioni: che sia nulla la somma delle espressioni $x\ (u - a\,x - b)$, e che sia nulla la somma delle espressioni $(u - a\,x - b)$, che possono formarsi mediante le coppie di valori dati di u e di x.

Nel nostro esempio abbiamo: per $x = 0$, $u = 72$; per $x = 10$, $u = 61$; per $x = 20$, $u = 54$; per $x = 30$, $u = 40$; per $x = 40$, $u = 33$; per $x = 50$, $u = 22$. Quindi le due condizioni risultano espresse nelle due equazioni:

$$5310 - 5500\,a - 150\,b = 0; \qquad 282 - 150\,a - 6\,b = 0.$$

Risolvendo questo sistema di due equazioni di primo grado a due incognite, troviamo i valori dei parametri

$$a = -0{,}9943 \qquad \text{e} \qquad b = 71{,}8577,$$

che soddisfano la condizione posta. L'equazione, che ora possiamo scrivere, della retta che rende minima in confronto ad ogni altra la somma dei quadrati degli scostamenti, cioè $y = 71{,}8577 - 0{,}9943\,x$, differisce ben poco da quella cui eravamo giunti precedentemente per tentativi. Anche i valori che la precedente equazione ci permette di surrogare a quelli noti (ordinatamente: 71,86; 61,91; 51,97; 42,03; 32,09; 22,14) differiscono poco da quelli dianzi adottati; lo scostamento medio quadratico si riduce a 1,28 in confronto a 1,29, ed è tutto qui il vantaggio di approssimazione conseguito (anzi se ci serviamo dello scostamento medio assoluto troviamo uno svantaggio, poichè esso è passato da 1,00 a 1,03; ma il metodo qui applicato rende il minimo possibile lo scostamento medio quadratico, non quello assoluto). In questo nostro esempio il procedimento sistematico ci ha dato un piccolissimo vantaggio di approssimazione, ma in altri casi può dare vantaggi assai maggiori. Ha inoltre l'incontestabile pregio di escludere l'arbitrio: per tentativi, cento diversi esecutori giungono a cento diverse, se pur vicine, soluzioni; mercè il metodo sistematico, sono condotti tutti alla stessa soluzione.

Possiamo determinare la retta da surrogare alla spezzata con altri procedimenti sistematici: porre, per esempio, la condizione che la somma dei valori surrogati sia uguale a quella dei corrispondenti valori noti, sia per i primi tre valori dati, sia per gli ultimi tre. Abbiamo posto la condizione separatamente per *due* intervalli all'intento di ottenere le *due* equazioni che ci occorrono per determinare i valori dei *due* parametri. Otteniamo così le due condizioni: $187 - 30\,a - 3\,b = 0$ e $95 - 120\,a - 3\,b = 0$. Risolvendo, si ha $a = -1{,}0222$ e $b = 72{,}5555$; si hanno, cioè, ancora valori poco differenti da quelli che avevamo adottati nel primo tentativo empirico. Dall'equazione $y = 72{,}5555 - 1{,}0222\,x$ desumiamo i valori da surrogare a quelli noti: ordinatamente 72,55; 62,33; 52,11; 41,89; 31,67; 21,45: valori anche questi poco differenti da quelli adottati nel nostro primo tentativo empirico. Lo scostamento medio assoluto risulta di 1,26, lo scostamento medio quadratico di 1,37: entrambi leggermente più alti di quelli ottenuti sopra.

8. — Senza accennare per ora ad altri possibili procedimenti, riflettiamo un momento alle precedenti applicazioni.

Graficamente, che cosa abbiamo fatto? Abbiamo sostituito una retta alla spezzata; cioè invece di tracciare una spezzata *per* i punti noti abbiamo tracciato una retta *fra* i punti noti. Se i punti fossero stati due soli, avremmo dovuto tracciare la retta *per* i punti dati; se due dei punti noti ci fossero apparsi per qualche ragione meglio adatti degli altri a segnare l'andamento del fenomeno, avremmo potuto tracciare la retta per quei due punti (e quindi verosimilmente ancora, riferendoci all'insieme dei punti, *fra* i punti dati). Se la retta si fosse allontanata troppo dalla realtà, avremmo potuto tracciare una curva geometrica regolare meglio conforme all'andamento dei dati: l'avremmo tracciata ancora *fra* i dati noti, eventualmente facendola passare per alcuni dei punti noti meglio adatti a segnare l'andamento del fenomeno; se il numero dei parametri della curva fosse stato uguale a quello dei punti noti, avremmo tracciato la curva *per* i punti noti.

L'operazione eseguita, cioè la sostituzione di una curva geometrica regolare ad una spezzata o ad una curva ricavata direttamente dall'osservazione, o, se si vuole, la sostituzione di una funzione analitica ad una funzione statistica, operata in modo che ai valori noti di questa si accostino i corrispondenti valori di quella, o addirittura coincidano con essi, è detta *interpolazione*, e al nome si aggiunge l'aggettivo *analitica* per indicare che si adoperano gli strumenti dell'analisi matematica. I saggi che abbiamo eseguito or ora ci mostrano l'applicazione del *metodo d'interpolazione dei minimi quadrati* e del *metodo d'interpolazione delle somme.* Col primo, adottata una determinata forma di funzione interpolatrice, si cercano quei valori dei parametri che rendono minima la somma dei quadrati degli scostamenti dei dati interpolati dai corrispondenti dati osservati: minima, s'intende, rispetto a tutte le funzioni di quella determinata forma (nel nostro esempio rispetto a tutte le funzioni della forma $y = ax + b$); si tende, cioè, a soddisfare una condizione di accostamento ai dati osservati. Col secondo, adottata una determinata forma di funzione interpolatrice, si cercano i valori dei parametri che rendono la somma dei dati interpolati uguale a quella dei dati osservati, simultaneamente in tanti intervalli, nei quali viene suddiviso il campo di variazione della x, quanti sono i parametri stessi.

Si tende cioè a soddisfare una condizione di equivalenza rispetto ai dati osservati, che implica anche una condizione di equilibrio: l'uguaglianza fra la somma degli scostamenti positivi e quella dei negativi.

A proposito delle condizioni di accostamento e di equilibrio riferite agli scostamenti, notiamo che in certi casi (specialmente se i dati noti non sono cifre assolute ma rapporti, o se essendo cifre assolute hanno un campo di variazione molto vasto) può convenire di porre tali condizioni rispetto agli scostamenti relativi (rapporti tra scostamento e corrispondente dato noto): per esempio porre la condizione che sia minima possibile la somma dei quadrati degli scostamenti relativi (metodo dei minimi quadrati) o quella che sia nulla in più intervalli la somma degli scostamenti relativi (metodo delle somme).

Ritroviamo nel campo dell'interpolazione condizioni di accostamento, di equilibrio e di equivalenza del tutto simili a quelle che abbiamo incontrate nel campo delle medie. La differenza è questa: che con la media si sostituisce un unico valore a tutti i termini della serie, talvolta molto disuguali fra loro; con l'interpolazione invece si tende a sostituire a ciascun termine un valore atto a rappresentarlo con buona approssimazione, e perciò le condizioni di accostamento possono operare con efficacia assai maggiore. Appunto in questa differenza, come abbiamo già notato, sta la superiorità del procedimento interpolatorio, che riproduce l'andamento della serie, mentre la media lo nasconde.

Se, nel nostro esempio di applicazione, avessimo voluto partire da condizioni simili a quelle che definiscono le medie di posizione, avremmo potuto, con la riga alla mano, tracciare la retta in modo da lasciare al di sopra o al di sotto di essa un ugual numero di punti noti (condizione analoga a quella che definisce la mediana), ovvero tracciarla in modo da farla passare per il maggior numero possibile di punti noti (condizione analoga a quella che definisce il valore più frequente). Ma nel campo dell'interpolazione queste condizioni non conducono ad *una* soluzione; ciascuna di esse può essere conciliabile con infinite soluzioni; e quindi non v'è adito ad un procedimento sistematico.

Si giunge a procedimenti sistematici d'interpolazione anche per vie diverse da quelle che abbiamo già mostrato. Nel *metodo d'interpolazione delle aree* si pone la condizione che l'area della curva interpolata sia uguale a quella della curva osservata, simultaneamente in tanti intervalli, nei quali viene diviso il campo di variazione della x, quanti sono i parametri da determinare. A tale condizione di equivalenza fa riscontro, com'è ovvio, una condizione di equilibrio: quella che la somma delle aree rappresentanti deviazioni positive dell'una curva dall'altra sia uguale alla somma delle aree rappresentanti deviazioni negative. Un'altra condizione di equivalenza che si può porre è quella che, — date tante serie di coefficienti o moltiplicatori quanti sono i parametri da determinare, composta ciascuna di tanti termini quanti sono i dati noti, in modo che a ciascun dato noto, e quindi al dato interpolato che deve surrogarlo, corrisponda ordinatamente un determinato coefficiente in ciascuna serie —, la somma dei prodotti dei coefficienti di ciascuna serie per i corrispondenti dati interpolati sia uguale a quella dei prodotti dei coefficienti stessi per i corrispondenti dati noti. Secondo il criterio di scelta dei coefficienti, questa condizione conduce a diversi metodi d'interpolazione. Può condurre al *metodo dei momenti*, così denominato perchè, i coefficienti essendo le successive potenze dei valori dati di x che corrispondono ai valori noti u, la somma dei prodotti del tipo $x^m u$ costituisce l'emmesimo « momento » della funzione statistica, e la somma dei prodotti del tipo $x^m y$ costituisce l'emmesimo « momento » della funzione interpolatrice (a questo metodo corrispondono i criteri per la determinazione dei parametri della curva degli errori indicati al paragrafo 5). Può condurre al *metodo d'interpolazione di Cauchy*, così detto dal nome del suo ideatore, nel quale i coefficienti sono scelti in modo da poter assumere soltanto i valori $+1, -1, 0$. Può anche condurre al metodo, che già conosciamo, dei minimi quadrati, nel quale i coefficienti sono i valori delle derivate parziali del primo ordine della funzione interpolatrice rispetto ai parametri, nei punti corrispondenti ai valori noti di x. Con l'applicazione di altri coefficienti, lo stesso criterio conduce ad altri metodi meno usati.

9. — Non è sistematico, ma è, in compenso, rapido, il procedimento d'interpolazione che consiste nel determinare i valori dei parametri della funzione interpolatrice col sussidio non di tutti i punti noti ma di una parte di essi, e precisamente di tanti punti quanti sono i parametri da determinare (due per la retta, tre per la curva parabolica di secondo grado, quattro per la curva parabolica di terzo grado, uno per il cerchio avente il suo centro nell'origine delle coordinate, due per la curva logaritmica, ecc.). L'inconveniente di questo procedimento consiste nell'arbitrarietà della scelta dei punti per i quali vien condotta la curva interpolatrice; ma è appunto tale arbitrarietà che consente all'operatore abile di raggiungere una soddisfacente approssimazione facendo passare la curva per i punti che meglio sono atti a segnare le caratteristiche dominanti dell'andamento del fenomeno. L'applicazione è semplice: ponendo la condizione che sia nulla la differenza tra ciascun valore noto u corrispondente ad uno dei punti scelti e il corrispondente valore interpolato, che si esprimerà quale funzione di x, otterremo le equazioni occorrenti per la determinazione dei parametri. Per la retta, ad esempio, avremo due equazioni del tipo $u - a x - b = 0$; per la curva parabolica di secondo grado avremo tre equazioni del tipo $u - a x^2 - b x - c = 0$.

10. — Quando la funzione interpolatrice prescelta si adatta bene a rappresentare i dati dell'osservazione è, in generale, press'a poco indifferente servirsi dell'uno o dell'altro fra i vari metodi di interpolazione. Ma quando l'adattamento è imperfetto, si possono ottenere risultati molto differenti secondo il metodo che si applica, e in certi casi anche secondo il procedimento che si segue nell'applicazione di taluno di essi che ammetta più vie.

Non si possono ridurre a sistema i criteri per la scelta fra i vari metodi e procedimenti: criteri che soltanto una lunga esperienza può suggerire all'operatore, secondo il campo in cui si svolge l'indagine e secondo il fine perseguito. Le attitudini personali dell'operatore hanno maggiori possibilità di esplicarsi nei metodi delle somme e delle aree, nell'applicazione dei quali egli deve dividere il campo di variazione della variabile indipendente in tanti intervalli quanti sono i parametri da determinare. Poichè tale divisione si può eseguire in più modi differenti, e spesso in un grande numero di modi, senza che *a priori* appaia nettamente preferibile l'uno o l'altro, è assai vasto il campo aperto all'arbitrio. È questa una ragione d'inferiorità dei metodi delle somme e delle aree in confronto a quelli dei minimi quadrati e dei momenti, dall'aspetto sistematico; ma può tramutarsi in ragione di superiorità, nell'applicazione, se l'operatore sappia abilmente valersi dell'arbitrio che gli è concesso, eseguendo la divisione degli intervalli in modo atto ad agevolare la fedele riproduzione, nella curva interpolata, di caratteri particolari della curva statistica.

In un corso elementare come il nostro non importa tanto imparare l'applicazione pratica dei vari metodi d'interpolazione, che normalmente è riservata ai tecnici della statistica, quanto intendere i principii generali cui il procedimento interpolatorio è informato, per poterne apprezzare criticamente l'impiego da altri compiuto. D'altronde l'applicazione si risolve in una serie di operazioni aritmetiche ed algebriche, che chiunque può eseguire sulla scorta di un manuale o di un formulario, mentre soltanto chi conosca bene la materia alla quale viene eseguita l'applicazione, e sia dotato di mente equilibrata, può giudicare se l'interpolazione sia o non sia opportuna e, una volta eseguita, se sia o non sia riuscita all'intento.

11. — Ricapitolando le nozioni già esposte nei precedenti paragrafi, notiamo:

che l'interpolazione è procedimento applicabile soltanto alle serie statistiche i cui termini possono considerarsi come una successione di valori di una funzione;

che, entro tali limiti, l'interpolazione analitica può essere correttamente applicata soltanto alle serie il cui diagramma richiama l'andamento di una curva geometrica regolare; altrimenti conduce ad una rappresentazione troppo lontana dalla realtà;

che, anche entro questi più ristretti confini, normalmente conviene applicare l'interpolazione analitica soltanto quando la curva geometrica regolare, idonea a rappresentare con buona approssimazione la curva statistica, sia caratterizzata da un numero di parametri molto piccolo; altrimenti, invece di una rappresentazione semplice e chiara, si consegue una rappresentazione complicata ed oscura.

Aggiungiamo che, per quanto il racchiudere in una breve formola una serie statistica — specialmente se composta di un numero molto grande di termini — sia operazione per se stessa vantaggiosa, la massima utilità dell'interpolazione non consiste tanto nella possibilità di sintesi d'una serie isolata quanto nella possibilità di comparazione tra gli andamenti di più serie. Il confronto tra le equazioni delle varie curve permette talora di giungere rapidamente e sicuramente a risultati che solo adagio e con incertezza sarebbero stati raggiunti per altra via; e dà il modo di accertare e di misurare concordanze, discordanze e in generale relazioni tra gli andamenti di più serie, che sarebbe molto difficile conoscere e precisare senza il sussidio dell'interpolazione.

12. — Per poter giudicare se una formola interpolatrice sia veramente adatta a rappresentare una serie statistica, bisogna verificare il grado di approssimazione col quale i dati y interpolati mercè la formola stessa riproducono i dati u osservati. Poichè un grado di approssimazione maggiore corrisponde a scostamenti minori, una media dei valori assoluti degli scostamenti $(y - u)$ può essere sufficiente in molti casi come misura dell'approssimazione; ove occorra, si può metterla in rapporto con una media dei dati u per avere un'idea della rilevanza relativa degli scostamenti; e se si vuol essere ancora più precisi, si può calcolare a tale intento una media dei valori assoluti degli scostamenti relativi $(y - u) : u$, che misurano gli errori relativi incontrati nella sostituzione dei singoli dati interpolati ai corrispondenti dati osservati. Se l'approssimazione raggiunta è scarsa, conclusioni desunte dall'esame della equazione della curva interpolatrice possono risultare non conformi, o addirittura contrastanti, con una corretta interpretazione della realtà.

È difficile, per non dire impossibile, stabilire un criterio generale che permetta di giudicare se l'approssimazione raggiunta mediante l'interpolazione di una data curva sia soddisfacente. Secondo la natura dei dati statistici e secondo lo scopo dell'interpolazione, i criteri atti a giudicare la bontà dell'approssimazione possono variare grandemente. In certi casi, per esempio quando si adopera la interpolazione per correggere approssimativamente errori d'osservazione, *devono* rimanere forti scostamenti tra la curva interpolata e la curva statistica; in altri casi, per esempio quando si adopera l'interpolazione per scindere le variazioni d'un fenomeno dovute a circostanze operanti in modo continuo da quelle operanti in modo ciclico o in modo saltuario, è normale la presenza di notevoli scostamenti. Errerebbe chi sempre applicasse la formola interpolatrice che dà minori scostamenti dall'osservazione: anche a prescindere dai casi ora accennati, non conviene mai spingere l'approssimazione ad un punto tale da eliminare gli scostamenti che corrispondono a variazioni non significative dei dati noti, cioè a variazioni le quali si attenuerebbero o addirittura scomparirebbero coll'estendersi del campo di osservazione.

13. — La rigorosa regolarità, caratteristica della curva geometrica, non si può estendere alla curva statistica cui la surroghiamo in via di approssimazione. Anche se, nei limiti dell'osservazione, la curva geometrica sostituisce con ottima approssimazione la curva statistica, non se ne può trarre alcun fondamento logico per affermare che possa sostituirla con approssimazione ugualmente buona fuori di quei limiti. Nel nostro esempio l'assurdità di una ipotesi simile è evidente: assunto un decremento annuo dell'1 °/₀ circa nell'analfabetismo, basterebbe risalire pochi anni verso il passato per trovare una percentuale di analfabeti del 100 °/₀; risalendo ancora si troverebbero percentuali superiori al 100 °/₀ (risultato privo di significato); e basterebbe procedere pochi anni verso l'avvenire per incontrare una percentuale dello 0 °/₀ e successivamente percentuali negative (prive anch'esse di significato). Non sempre chi si trova in possesso di una formola interpolatrice atta a descrivere la realtà entro dati limiti, e si lascia sedurre ad estenderne la validità oltre codesti limiti, — a compiere cioè una *estrapolazione* —, incontra una così palese assurdità di conclusioni. atta a frenare l'entusiasmo per l'agevole profezia; perfino scienziati di gran valore sono caduti talora in gravi errori percorrendo questa via con soverchia foga. Soltanto in certi casi, e per intervalli prossimi ai limiti dell'osservazione, l'estrapolazione, cautamente eseguita da chi conosca profondamente la materia che è oggetto d'indagine, può essere lecita ed utile, come potremo meglio intendere dopo che avremo analizzato i caratteri delle regolarità statistiche.

Qui torna opportuna l'osservazione che curve geometriche regolari di forma differente possono avere *in un dato intervallo* andamento poco diverso e quindi essere tutte atte a rappresentare con buona approssimazione una medesima curva statistica *in quell'intervallo*. Ma fuori dell'intervallo stesso le varie curve divergono, così che, secondo si sia interpolata l'una o l'altra di esse, si giunge nell'estrapolazione a risultati diversi, e talvolta contrastanti.

In generale, se anche si riesce a rappresentare con ottima approssimazione mediante una curva geometrica una curva statistica, non bisogna immaginare senz'altro che l'equazione della curva rappresenti « la legge » del fenomeno. L'interpolazione è un procedimento descrittivo, e se la rappresentazione semplificata dell'andamento di un fenomeno collettivamente tipico può metterne in risalto la regolarità o le regolarità, non è però quasi mai l'espressione di una « legge » nel vero significato scientifico della parola. Anche questo punto riuscirà più chiaro dopo che avremo analizzato i caratteri delle regolarità statistiche; ma il fatto solo che equazioni di curve di forma differente, e quindi corrispondenti a « leggi » differenti, possano rappresentare con buona approssimazione una medesima curva statistica basta a mostrare con quanta precauzione si debba procedere su terreno così infido.

14. — All'espressione *interpolazione grafica* si può dare un'interpretazione più ampia di quella che la fa consistere nella semplice traduzione grafica dell'interpolazione analitica. Molte volte, come sappiamo, la spezzata che rappresenta una serie statistica si presta ad essere sostituita graficamente da una curva avente andamento regolare all'occhio. Ma se, dopo avere tracciato questa curva, ricerchiamo, per tentativi, se essa coincida almeno approssimativamente con una curva geometrica regolare, la cui equazione possieda i requisiti di semplicità richiesti per rendere conveniente l'interpolazione, non di rado la ricerca dà risultati negativi, nonostante l'apparente regolarità della curva tracciata. Ne indurremo che tale regolarità non si presta ad essere sintetizzata in forma semplice; e rinunziando alla ricerca di una formola che riassuma in sè la serie statistica, ci accontenteremo del risultato grafico, il quale può essere sufficiente ed utile per separare le variazioni principali del fenomeno da quelle che, almeno a certi fini, riguardiamo come secondarie e trascurabili. Può ancora convenire di fermarsi all'interpolazione grafica quando nella dominante regolarità di andamento di una curva si vogliano mantenere evidenti alcune anomalie che per la loro origine presentano speciale importanza: così se nello sviluppo regolare di una popolazione si vogliano far risaltare i turbamenti determinati da una lunga guerra, se nella regolare distribuzione per età degli sposi si voglia serbare quella irregolarità che deriva dal servizio militare obbligatorio per causa del quale i matrimoni sono diradati in certe età, ecc. Nell'interpolazione analitica sarebbe estremamente difficile, e talora impossibile, riprodurre tali irregolarità; in ogni modo, per giungervi bisognerebbe ricorrere a formole interpolatrici complicatissime e quindi disadatte allo scopo della rappresentazione semplificata.

Nell'esecuzione dell'interpolazione grafica si seguono criteri analoghi a quelli dell'interpolazione analitica, ma più semplici, perchè non sarebbe praticamente possibile applicare graficamente condizioni come quelle che stanno a base dei metodi dei minimi quadrati o dei momenti. La condizione più semplice che si può porre è quella che la curva interpolatrice passi fra i punti noti lasciandone un ugual numero al di sopra e al di sotto; perchè questa condizione di equilibrio divenga anche efficace condizione di accostamento conviene porla simultaneamente per due o più intervalli nei quali si sia diviso il campo di variazione della variabile indipendente. In qualche caso, per ottenere una curva di andamento regolare, l'operatore dovrà rassegnarsi a porre la condizione in forma approssimativa: press'a poco tanti punti sopra la curva quanti sotto di essa. Se davvero la spezzata desunta dall'osservazione ha andamento fondamentalmente regolare, non riesce difficile in generale ottenere che sia soddisfatta la condizione ora indicata, o in forma precisa o almeno in forma approssimativa.

Meno rapide nell'attuazione grafica, perchè richiedono pazienti letture del diagramma, sono le condizioni che la somma degli scostamenti positivi sia uguale a quella dei negativi, o che la somma delle aree comprese fra la curva interpolata e la curva osservata sia uguale al di sopra e al di sotto di una di queste curve: condizioni rispettivamente corrispondenti a quelle dei metodi delle somme e delle aree, che conviene porre simultaneamente per più intervalli allo scopo di ottenere un miglior accostamento. A condizioni più complicate si ricorre di rado nella pratica, perchè riuscirebbero di troppo laboriosa applicazione. Il buon successo dell'interpolazione grafica dipende dall'abilità dell'operatore e dalla conoscenza ch'egli ha del fenomeno cui i dati si riferiscono: l'elemento personale ha qui importanza anche maggiore che nell'interpolazione analitica.

15. — È difficile esporre in modo generale, e nel tempo stesso semplice, una teoria sull'influenza degli errori nell'interpolazione. Ma possono, ai fini del nostro corso, essere sufficienti alcune brevi e piane considerazioni.

Se tutti i termini della serie sono errati nello stesso senso, la deformazione della curva statistica nota, rispetto a quella esatta (che si sarebbe avuta, cioè, in assenza di errori), si riflette in generale irreparabilmente nella curva interpolata ; la misura media dell'errore non viene nella maggior parte dei casi notevolmente accresciuta nè diminuita per effetto dell'interpolazione.

Altrettanto accade, di solito, se procedendo dal primo all'ultimo termine della serie, nell'ordine normale di essa, si passa gradualmente da errori in un senso ad errori in senso opposto: se per esempio s'incontrano nei primi termini forti errori in eccesso, poi ancora errori in eccesso ma di mano in mano decrescenti, poi lievi errori in difetto, e poi a grado a grado errori in difetto crescenti in valore assoluto.

Nei due casi ora considerati, la regolarità della distribuzione degli errori impedisce all'operatore di averne indizio. La curva statistica può apparirgli, anzi, assai regolare: e nulla gli indica che quella ch'egli ha innanzi non sia la vera regolarità della serie. Per esempio, i redditi denunziati dai contribuenti al fisco sono sempre inferiori al vero, e in misura spesso rilevante; ciò non toglie che la loro curva di distribuzione per grandezza presenti una notevole regolarità.

Quando, invece, percorrendo la serie nel suo ordine normale, s'incontrano alternativamente errori in eccesso ed errori in difetto, il diagramma spesso presenta oscillazioni verso l'alto e verso il basso, che alterano una dominante regolarità di andamento; tali oscillazioni, corrispondenti agli errori, possono in molti casi venire eliminate nell'interpolazione, così nella sua forma analitica come in quella grafica. È raro che l'eliminazione avvenga automaticamente, anzi ha gran parte nell'ottenerla l'abilità dell'operatore. Su questo argomento torneremo fra breve.

16. — Oltre che per la rappresentazione delle serie statistiche, l'interpolazione può servire per la ricostruzione approssimativa dell'andamento di serie note soltanto in parte. Così, se l'applichiamo ad una serie di dati sulla popolazione d'un paese secondo successivi censimenti, la curva interpolata ci indicherà il numero approssimativo degli abitanti in qualsiasi momento intermedio fra due rilevazioni. È superfluo aggiungere che il procedimento si può ammettere soltanto se è plausibile l'ipotesi dell'andamento regolare della curva fra i punti noti: se conosciamo la temperatura in una data località alle 7 della mattina di trenta giorni successivi, la curva che possiamo tracciare non ci indica, in generale, con sufficiente approssimazione l'andamento della temperatura in quel periodo, perchè tra le 7 di ciascun giorno e le 7 del giorno successivo il termometro di solito prima sale e poi discende.

L'interpolazione serve anche, nello stesso modo, alla scissione di somme di valori di una funzione statistica, corrispondenti ad intervalli di valori della variabile indipendente, in somme corrispondenti ad intervalli più ristretti, o in singoli valori corrispondenti a valori singoli della variabile. Se conosciamo la distribuzione di una popolazione per gruppi quinquennali di età, possiamo porre la serie dei dati nella forma seguente: primo dato, popolazione di età superiore a 0 anni (cioè l'intera popolazione); secondo dato, popolazione di età superiore a 5 anni (cioè il dato precedente diminuito del numero degli abitanti in età da 0 a 5 anni); terzo dato, popolazione di età superiore a 10 anni, ecc. Abbiamo così una serie (*serie cumulativa*, per il modo in cui è stata ottenuta) di dati che indicano il numero degli abitanti di età superiore ad x, in funzione di x. Se riusciamo a rappresentarla in modo soddisfacente con una funzione interpolatrice, potremo poi leggere sulla rappresentazione grafica di questa, o calcolare mediante la sua rappresentazione analitica, il numero degli individui di età superiore ad x, non solo per $x = 0, 5, 10$, ecc. ma anche per qualsiasi valore intermedio. E calcolato il numero degli individui di età superiore a 0, a 1, a 2, a 3, ecc., anni, ci sarà facile determinare per sottrazione il numero degli individui in età da 0 a 1, da 1 a 2, da 2 a 3, ecc., anni. Avremo così scisso in gruppi annuali d'età la popolazione che ci era data per gruppi quinquennali. Naturalmente anche questo procedimento si può ammettere soltanto se è plausibile l'ipotesi del regolare andamento della curva negli intervalli fra i punti noti.

Quando i dati noti siano affetti da errori di osservazione i quali alterino la regolarità di andamento della serie, l'interpolazione, ricostituendo tale regolarità, può servire talvolta — come abbiamo già accennato — per l'approssimativa correzione degli errori. Accade, per esempio, quasi sempre nei censimenti di popolazioni ignoranti che le dichiarazioni di età si addensino intorno agli anni espressi da numeri multipli di 10, secondariamente intorno agli anni espressi da multipli di 5, e poi ancora intorno a quelli espressi da numeri pari: la distribuzione per età degli abitanti indicata dal censimento risulta quindi profondamente diversa da quella reale. Un accorto impiego dei procedimenti interpolatorii spesso permette di ricostruire in modo approssimativo la vera distribuzione della popolazione secondo l'età.

Un ufficio analogo a quello dianzi spiegato esercita l'interpolazione quando è adoperata per eliminare dalla descrizione di un fenomeno quelle variazioni che si dicono non significative perchè tendono ad attenuarsi od a scomparire coll'estendersi del campo di osservazione. Potremo meglio intendere l'importanza di questo ufficio quando avremo studiato le relazioni che intercedono tra il modo di presentarsi delle regolarità statistiche e l'ampiezza del campo di osservazione.

L'interpolazione è promossa da procedimento descrittivo a procedimento investigativo quando viene impiegata per scindere le variazioni connesse con certe circostanze da quelle connesse con altre: così quando, nell'esame delle variazioni cronologiche di un fenomeno, si cerca di separare quelle derivanti da fattori di azione evolutoria da quelle derivanti da fattori di azione ciclica o saltuaria; così quando, nello studio della distribuzione di un carattere quantitativo (p. es. il peso del corpo), si cerca di scindere la parte della variazione di esso che si connette con la variazione di un altro carattere (p. es. la statura) da quella che non vi si connette. In queste applicazioni, la forma che si attribuisce alla funzione interpolatrice rispecchia un'ipotesi sul modo di agire di un dato fattore, o gruppo di fattori; ipotesi che a sua volta in molti casi è suggerita dall'esame dell'andamento dominante della curva statistica, o delle deviazioni, che si manifestano, da questo andamento.

17. — Quando la variazione di un fenomeno attraverso il tempo dà l'impressione, attraverso la rappresentazione grafica, di una dominante regolarità turbata da oscillazioni ora verso l'alto ora verso il basso, l'interpolazione grafica può facilmente consentire l'eliminazione delle oscillazioni e la netta visione della regolarità. L'ordinata della curva interpolata graficamente non coincide, in generale, con la corrispondente ordinata nota, ma può considerarsi quasi sempre come una media di più ordinate note consecutive, poichè appunto raggruppando in una media più ordinate consecutive si eliminano, o almeno si attenuano, le oscillazioni ora positive ora negative che alterano la regolarità della curva. Il procedimento si può applicare in forma numerica più precisamente che in forma grafica, col surrogare a ciascun dato della serie una media di più dati fra i quali esso è compreso nell'ordine cronologico (in generale si adopera la media aritmetica, ma se la serie ha un andamento molto diverso da quello di una progressione aritmetica possono convenire altre medie): di solito una media dei 3, 5, 7, 9, ... dati, fra i quali quello che si vuol sostituire occupa il posto di mezzo in ordine di tempo.

Questo procedimento della *media mobile* non è che un surrogato dell'interpolazione: al pari di altri procedimenti empirici di *perequazione* delle serie statistiche tende a mettere in rilievo una regolarità fondamentale che è dissimulata in parte da irregolarità secondarie. Non è un procedimento di interpolazione analitica, nè un procedimento di interpolazione grafica (sebbene, una volta applicato, si possa tradurlo graficamente); si potrebbe dire piuttosto un procedimento di *interpolazione numerica*, al pari degli altri procedimenti di perequazione, ora ricordati, che sostituiscono a ciascun dato una media ponderata di più dati fra i quali esso è compreso. L'interpolazione numerica si può applicare anche a serie non cronologiche, aventi però il carattere di funzioni statistiche, sempre allo stesso fine ora detto. Appunto perchè procedimento empirico e non sistematico, essa dà risultati molto diversi secondo l'abilità tecnica di chi l'impiega e la conoscenza ch'egli ha dei fenomeni rappresentati.

Il metodo della media mobile è opportunamente applicato per l'eliminazione di oscillazioni aventi carattere ciclico o periodico. Per esempio, se attraverso i dati mensili sulle esportazioni, che risentono l'influenza di fattori stagionali, vogliamo indagare la tendenza dominante (espansione, contrazione, stazionarietà) del commercio d'esportazione, per eliminare le variazioni periodiche sostituiremo a ciascun dato la media dei dodici dati mensili dei quali esso è l'ultimo in ordine cronologico. È chiaro però che in tal modo non metteremo più in rilievo la tendenza delle esportazioni nel singolo mese, ma nel periodo annuale che con quel mese si compie. Come risulta da questo esempio, il numero dei dati che si riuniscono a formare la media deve corrispondere all'ampiezza del periodo o ciclo entro il quale si ritiene che le oscillazioni positive e negative si compensino reciprocamente: nel caso nostro dodici mesi. Qui non poteva sorgere dubbio sull'ampiezza del periodo; altre volte l'abilità dell'operatore ha campo a manifestarsi nella scelta, non sempre facile, della più plausibile tra varie ammissibili ipotesi sull'ampiezza del ciclo.

Lo stesso metodo della media mobile può essere usato per l'approssimativa eliminazione di errori di osservazione. Quando si abbia ragione di ritenere che, formando tante successioni di k termini consecutivi, a partire dal primo, dal secondo, dal terzo, ecc., termine di una serie statistica, entro ciascuna di queste successioni gli errori in eccesso siano compensati, all'incirca, da errori in difetto, si può in certi casi assumere la media di k termini a rappresentare approssimativamente quello che occupa il posto centrale fra essi. Il metodo è rigorosamente corretto se i veri valori dei termini costituiscono una progressione aritmetica e se entro la successione di k termini la compensazione fra errori positivi e negativi è perfetta; ma in altre ipotesi può condurre a risultati poco attendibili. Perciò dev'essere impiegato con cautela e da gente esperta.

Indicazioni bibliografiche. — Fra i manuali di statistica già citati, vedansi specialmente quelli di WESTERGAARD, BOWLEY, MORTARA per un più ampio sviluppo degli argomenti trattati nel presente capitolo.

Vedansi inoltre: CASSINIS G., *Calcoli numerici, grafici e meccanici*, Pisa, Mariotti-Pacini, 1928. — RUNNING T. R., *Empirical formulas*, London, Chapman and Hall, 1917. — WHITTAKER AND ROBINSON, *The calculus of observations*, London, Blackie and Son, 1924. — ELDERTON W. P., *Frequency curves and correlation*, 2ª ed., London, Layton, 1927. — PEARSON K., *Tables for statisticians and biometricians*, Cambridge University Press, 1914.

In qualche esercizio è richiamato l'*Atlante* del RASERI già citato.

Quesiti ed esercizi: 1. — Che cos'è una curva statistica?

2. — Una curva statistica rappresenta sempre una funzione statistica continua? Come può una curva continua rappresentare una funzione discontinua?

3. — Si tracci la curva della popolazione italiana dal 1862 in poi, servendosi soltando dei dati dei censimenti, riportati nell'ASI, nelle due forme di spezzata e di curva vera e propria. Si calcoli la popolazione nel 1921, dall'una e dall'altra rappresentazione, e la si confronti col risultato dei calcoli dell'Istituto centrale di statistica, indicato nelle « Notizie statistiche retrospettive » dello stesso ASI.

4. — Dovendo rappresentare graficamente l'andamento dell'emigrazione italiana negli ultimi cinquant'anni, si preferirà tracciare una linea spezzata od una curva (nel senso più proprio di questa parola)? Si giustifichi la preferenza espressa.

5. — Dovendo rappresentare graficamente l'andamento della mortalità in Italia dal 1884 al 1913, si preferirà tracciare una linea spezzata od una curva?

6. — Dovendo rappresentare graficamente l'andamento della produzione del grano in Italia negli ultimi cinquant'anni, si preferirà tracciare una spezzata od una curva?

7. — Si provi a rappresentare graficamente la frequenza dei reati denunciati in Italia, per ogni 100.000 abitanti, negli ultimi cinquant'anni, servendosi solo dei dati per gli anni 1881, 1886, 1891, 1896, ecc. Tracciata la curva della delinquenza sulla scorta di questi dati, la si confronti poi con quella tracciata mediante i dati per tutti gli anni. Può la prima curva rappresentare con buona approssimazione la seconda?

8. — Si eseguano le operazioni indicate nell'esercizio precedente, partendo invece dai dati sull'eccedenza dei nati sui morti per 1000 abitanti.

9. — Nell'ASI e nell'Atlante del Raseri si cerchino esempi di curve che presentino regolarità di andamento, e di altre che si mostrino irregolari.

10. — Si provi a rappresentare in via approssimativa, mediante una linea retta, ciascuna delle seguenti curve statistiche descriventi l'andamento di fenomeni demografici ed economici in Italia negli ultimi cinquant'anni: *a*) densità della popolazione per kmq.; *b*) nati vivi per 1000 abitanti; *c*) produzione

dell'olio d'oliva; *d*) produzione della ghisa; *e*) importazione di caffè; *f*) esportazione di vini; *g*) sviluppo della rete ferroviaria; *h*) risparmio medio per abitante. Si spieghi in ciascun caso il criterio seguito nel tracciare la retta. Si giudichi in ciascun caso se questa rappresenti con sufficiente approssimazione l'andamento del fenomeno. Si tracci in ciascun caso anche la linea orizzontale corrispondente alla mediana e si confronti l'approssimazione delle due rappresentazioni.

11. — Quali condizioni si cercherà di soddisfare nel tentare la sostituzione di una linea regolare (retta o curva) ad una spezzata che rappresenta una serie di dati statistici?

12. — Riprendendo le serie indicate nell'esercizio 10, si esegua l'interpolazione grafica di una curva in quei casi nei quali non sia apparsa soddisfacente l'approssimazione raggiunta con l'interpolazione di una retta. Si raggiunge ora in tutti i casi un risultato soddisfacente?

13. — Esaminando le serie di dati retrospettivi sulle principali produzioni agrarie e industriali negli ultimi cinquant'anni, contenute nell'ASI, si ricerchi quali di esse possano essere ragionevolmente oggetto di interpolazione e quali no. Si spieghino le ragioni del giudizio per ciascun caso.

14. — Che cos'è l'interpolazione? In che differisce l'interpolazione grafica da quella analitica? In che differisce l'interpolazione per punti dati da quella fra punti dati?

15. — Procedendo analogamente al modo tenuto nel testo al paragrafo 7, si determini col metodo dei minimi quadrati l'equazione di una retta atta a rappresentare l'andamento della natalità in Italia (nati vivi per 1000 abitanti): *a*) dal 1881 al 1892, *b*) dal 1893 al 1912, *c*) dal 1920 al 1929. Si eseguano gli stessi calcoli col metodo delle somme e si confrontino le equazioni trovate coi due metodi. Che cosa ci dicono queste equazioni sulle tendenze della natalità nei tre periodi considerati?

16. — Si eseguano le operazioni indicate nell'esercizio precedente sui dati sulla mortalità in Italia (morti per 1000 abitanti), per gli stessi periodi ivi indicati. Si confrontino le equazioni della mortalità con quelle della natalità, periodo per periodo, e se ne traggano conclusioni sull'andamento comparato dei due fenomeni demografici.

17. — Si misuri l'approssimazione con la quale le rette interpolate negli esercizi 15 e 16 rappresentano i dati dell'osservazione, confrontando i valori calcolati coi valori osservati.

18. — Si eseguano graficamente le interpolazioni degli esercizi 15 e 16; si confrontino i risultati dell'interpolazione grafica, e l'approssimazione in questa ottenuta, coi risultati e con l'approssimazione ottenuti nella interpolazione analitica.

19. — Si eseguano interpolazioni grafiche sui diagrammi delle seguenti serie, desumibili dall'ASI, capitolo « Climatologia »: *a*) temperatura media registrata da un osservatorio nei vari mesi di un anno; *b*) quantità dell'acqua caduta nei vari mesi di un anno, secondo le registrazioni di un dato osservatorio; *c*) numero delle scosse di terremoto registrate in Italia nei vari mesi di un dato anno. Si prestano ugualmente bene le diverse serie all'interpolazione grafica? Le curve di distribuzione presentano qualche regolarità nel singolo anno e qualche costanza nel tempo e nello spazio? (Si confrontino,

per rispondere all'ultimo quesito, più osservatorii nei casi *a*) e *b*), più anni nel caso *c*).

20. — Si verifichi se si prestino all'interpolazione grafica le seguenti serie, desumibili dall'ASI, capitolo « Territorio e popolazione »: *a*) popolazione italiana secondo l'età (poichè è data per gruppi quinquennali, si parta da un istogramma a canne d'organo); *b*) comuni italiani (o comuni di un compartimento) ordinati secondo il numero degli abitanti; *c*) sposi secondo l'età; *d*) nati-morti in un anno nelle varie regioni italiane. Nei casi *a*) e *b*) si può provare ad eseguire l'interpolazione grafica, invece che sulla serie originaria, sulla serie cumulativa che da essa si ricava nel modo spiegato al paragrafo 16 del testo. Il primo dato della serie cumulativa (p. es. popolazione da 0 anni in su) risulta uguale alla somma di tutti i termini della serie originaria; l'ultimo termine (p. es. popolazione da ω anni in su, se ω è l'età immediatamente superiore a quella del più longevo tra i censiti) risulta uguale a zero. Con questo artificio si può tracciare un diagramma lineare anche nei casi *a*) e *b*), nonostante che la serie originaria non indichi i valori di una funzione y corrispondenti a singoli valori di una variabile x, ma indichi invece le somme di valori della funzione y corrispondenti ai valori di x compresi in un certo intervallo. Coll'artificio indicato si passa infatti all'altra funzione: « somma dei valori di y corrispondenti ai valori di x non inferiori a un dato limite » (p. es. somma dei numeri di censiti corrispondenti alle età da x in su); e di questa funzione si ottengono i valori corrispondenti a singoli valori di x, così che si può tracciare direttamente il diagramma lineare e su questo eseguire l'interpolazione grafica, ove sia il caso. Dai dati interpolati corrispondenti ai valori della serie cumulativa si possono poi ricavare per sottrazione i valori calcolati da contrapporre ai valori osservati della serie originaria (p. es. dai valori calcolati della popolazione di età da 5 anni in su e da 10 anni in su si può desumere per sottrazione il valore calcolato della popolazione in età da 5 a 10 anni, da contrapporre a quello osservato).

21. — Si provi a rappresentare mediante interpolazione grafica la distribuzione per statura degli iscritti di leva (ASI, capitolo « Igiene e sanità »; o più dettagliatamente MINISTERO DELLA GUERRA, *Della leva di terra*, relazione annuale). Quale forma di curva, descritta nel testo, si adatta a rappresentare tale distribuzione?

22. — Si dispongano in ordine di grandezza i dati sulla natalità (o sulla mortalità, o sulla nuzialità) nelle varie provincie italiane in un dato anno; e se la distribuzione per grandezza presenta qualche regolarità si provi ad eseguire un'interpolazione grafica atta a rappresentarla. Si ottengono qui distribuzioni del tipo della curva degli errori?

23. — Esercizio analogo al precedente, da compiere sui dati del rendimento medio per ettaro della coltura del frumento nelle varie provincie italiane.

24. — Si interpoli graficamente la distribuzione per età delle operaie sussidiate dalla Cassa nazionale assicurazioni sociali in occasione di un parto (ASI, capitolo « Credito e previdenza »). Si verifichi anche in questo caso se la curva degli errori è idonea alla rappresentazione.

25. — Si traduca in diagramma a scala logaritmica la serie delle probabilità di morte per la popolazione maschile per le età fra 30 e 90 anni (ASI

1929, pag. 42). Quale curva geometrica regolare è atta a rappresentare l'andamento della mortalità in funzione dell'età nella rappresentazione logaritmica? È possibile indurre dalla risposta al precedente quesito quale curva si presti a tale intento se invece dei logaritmi si considerino i dati originari?

26. — Quali condizioni sono caratteristiche del metodo d'interpolazione delle somme? di quello delle aree? di quello dei minimi quadrati? di quello dei momenti?

27. — Si potrebbe rappresentare mediante una retta l'andamento della percentuale dei nati illegittimi in Italia dal 1863 al 1904 (curva B della tavola 41 dell'atlante Raseri)? Sarebbe opportuna l'interpolazione di una retta inclinata o potrebbe bastare una media (retta orizzontale) per rappresentare l'andamento della percentuale dei nati-morti nello stesso periodo (curve A e C della citata tavola)?

28. — A quali condizioni formali e sostanziali è subordinata l'applicabilità del procedimento interpolatorio ad una serie statistica?

29. — Sarebbe adatta una linea retta a rappresentare l'andamento del numero dei sopravviventi in funzione dell'età (V. ASI 1929, pagg. 42-44, e atlante Raseri, tav. 48)?

30. — Come si misura il grado di approssimazione raggiunto in una interpolazione?

31. — Partendo dalle equazioni determinate negli esercizi 15 e 16, che descrivono l'andamento della natalità e della mortalità dal 1893 al 1912, si calcolino per estrapolazione i valori della natalità e della mortalità negli anni 1881 a 1892 e 1913 a 1930. Si confrontino questi dati calcolati coi dati osservati (v. ASI) e si espongano le considerazioni che tale confronto suggerisce.

32. — Le due curve rappresentate nella tavola 36 dell'atlante del Raseri si prestano all'interpolazione grafica? Per la prima di esse (matrimoni) potrebbe bastare una media? Per la seconda (frequenza delle nascite legittime) le oscillazioni sono veramente molto grandi, come appare a prima vista dal diagramma, e tali da sconsigliare l'interpolazione?

33. — Determinata l'equazione di una curva geometrica atta a sostituire con buona approssimazione una curva statistica, si può senz'altro affermare di avere scoperto «la legge» del fenomeno?

34. — A quali condizioni l'interpolazione può servire per la ricostruzione approssimativa di serie note soltanto in parte? Se s'ignora la natalità italiana degli anni 1916-1919, conoscendo quella degli anni 1906-1915 e 1920-1929, si può colmare la lacuna? Se si conosce il raccolto del grano soltanto degli anni dispari dal 1909 al 1929, ci si può valere di tali dati per calcolare il raccolto degli anni pari dal 1910 al 1930? Se si conosce la lunghezza delle ferrovie italiane al principio di ciascun decennio dal 1881 ai 1921, si può calcolarne la lunghezza al principio di ciascun anno del quarantennio?

35. — In qual modo si può adoperare l'interpolazione per la scissione di dati di una serie statistica? A quali serie è applicabile il procedimento?

36. — Si provi, col sussidio dell'interpolazione grafica, a scindere in gruppi annuali di età la distribuzione della popolazione italiana per gruppi quinquennali d'età indicata nell'ASI, adottando l'artificio spiegato nel paragrafo 16 del testo e richiamato nell'esercizio 20: calcolando cioè la serie cumulativa e su questa eseguendo l'interpolazione.

37. — Come e perchè l'interpolazione può servire per l'approssimativa correzione di errori di rilevazione? A quali condizioni può essa adempiere tale ufficio?

38. — Partendo dai dati sulla distribuzione per età della popolazione femminile italiana secondo il censimento del 1872, dati che sono affetti da gravi errori, si provi a ricostruire mediante interpolazione grafica il presumibile vero andamento della distribuzione per età.

39. — Nella tavola 12 dell'atlante del Raseri si provi a correggere mediante interpolazione grafica i presumibili errori della distribuzione per età delle proporzioni di abitanti che sapevano leggere, secondo i censimenti 1872 e 1882.

40. — Nella curva della mortalità per morbillo rappresentata nell'atlante del Raseri (tav. 50) si cerchi di scindere la variazione evolutoria corrispondente al progresso igienico e sanitario, dalla variazione ondulatoria corrispondente al succedersi di cicli epidemici.

41. — Nella curva del traffico ferroviario, che si può costruire sui dati mensili sulle merci trasportate negli ultimi cinque anni, si cerchi di scindere le variazioni stagionali da quella evolutoria e da quelle saltuarie.

42. — Si cerchi di scindere la variazione evolutoria dalle rimanenti variazioni, nella curva dei suicidi (atlante Raseri, tav. 73).

43. — in che consiste il procedimento della media mobile? A quali scopi può servire?

44. — Si applichi il procedimento della media mobile alle seguenti serie di dati retrospettivi contenuti nell'ASI, per gli ultimi cinquant'anni: *a*) mortalità (provando successivamente: medie triennali, medie quinquennali, medie settennali); *b*) produzione del frumento (medie quadriennali, medie quinquennali: v'è una ragione per preferire queste o quelle?); *c*) produzione dell'olio d'oliva (medie triennali, medie novennali, medie decennali: quali sembrano preferibili, e perchè?); *d*) produzione dello zucchero; *e*) importazione del cotone greggio; *f*) esportazione della seta greggia; *g*) corso del cambio su Londra. In quali di questi casi il procedimento appare di opportuna applicazione, in quali no?

45. — Si applichi il procedimento della media mobile ai dati sull'esportazione italiana negli ultimi cinque anni, nel modo spiegato al paragrafo 17 del testo, cercando di scindere la variazione evolutoria dalle variazioni stagionali e da altre minori variazioni.

46. — Applicando il procedimento della media mobile (medie settennali), si provi a correggere gli errori della distribuzione per età della popolazione femminile italiana indicata dal censimento italiano del 1872, fra i 30 e gli 80 anni.

CAPITOLO XIII.

La rappresentazione analitica della serie statistica: i numeri indici semplici.

Rapporto indice e numero indice — Rappresentazione di serie di vari tipi mediante numeri indici: scopo della rappresentazione; vantaggi e inconvenienti — Numeri indici con riferimento costante e numeri indici con riferimento variabile; numeri indici a catena con riferimento immediato o mediato — Criteri per la scelta del riferimento — Spostamento del riferimento — Serie di variazioni percentuali corrispondenti alle serie di numeri indici; errori che intervengono nella loro interpretazione — Influenza degli errori dei dati originari nella formazione dei numeri indici — Comparazione tra più serie di numeri indici — Determinazione grafica dei numeri indici — Rappresentazione di serie statistiche mediante rapporti di composizione.

1. — Sappiamo già che cos'è un *numero indice*. Il rapporto fra due dati statistici omogenei, desunti da osservazioni compiute in diversi tempi, in diversi luoghi, o in diverse sezioni dello stesso campo d'osservazione, è un rapporto indice; e il suo valore moltiplicato per 100 o per altra potenza di 10, per comodità di espressione, è un numero indice. È facile intendere che ad una serie di dati greggi od elaborati (medie, rapporti, ecc.), che da ora innanzi chiameremo per brevità «serie originaria», si può surrogare una serie di numeri indici, ma è meno facile intendere il vantaggio dell'operazione. Il numero dei termini è il medesimo nella serie derivata di numeri indici come nella serie originaria di dati, quindi non si consegue alcun vantaggio di sintesi; la grandezza assoluta di ciascun termine non si ritrova, attraverso la serie dei numeri indici che sono misure di grandezza relativa. Ma appunto queste misure di grandezza relativa possono essere idonee a mettere in risalto certi tratti caratteristici dell'andamento della serie statistica, che specialmente importa conoscere od apprezzare a determinati fini.

2. — Consideriamo, per esempio, i dati sull'importazione di frumento in Italia negli ultimi cinquant'anni, e desumiamo da questa serie di dati assoluti una serie di numeri indici. Come? I criteri possono essere svariati. Se vogliamo porre in rilievo le tendenze che si manifestano nell'importazione granaria, attraverso il tempo, calcoleremo i numeri indici mettendo in rapporto tutti i dati annuali col dato per l'anno iniziale del cinquantennio, o con una media dei primi due, quattro, sei dati se vogliamo riferirci non alla situazione di un anno particolare ma alla situazione che si poteva ritenere normale verso il principio del periodo considerato (conviene prendere un numero pari di dati perchè spesso si alternano annate favorevoli con annate avverse per il raccolto nazionale, e quindi oscilla fortemente da un anno all'altro il fabbisogno d'importazioni). Potremmo anche assumere come riferimento il dato finale del cinquantennio, o una media dei due, quattro, sei dati più recenti; ma se volessimo poi proseguire il calcolo l'anno venturo con lo stesso criterio dovremmo rifarlo tutto da capo. Per la stessa

ragione, e perchè il riferimento non avrebbe un riscontro concreto nella realtà, non assumeremmo in questo caso come denominatore dei rapporti indici una media dei cinquanta dati annuali, a meno che volessimo mettere in rilievo la forte variabilità dell'importazione granaria paragonando la sua distribuzione effettiva nel tempo con la distribuzione uniforme che corrisponderebbe, per esempio, alla media aritmetica. Immaginiamo, per fissare le idee, di avere scelto come riferimento il dato per il 1881, anno in cui vennero importati in Italia 1.540.440 quintali di frumento. Prendiamo ora un altro anno qualsiasi, per esempio il 1923: il numero indice risulta uguale a 1.811, posto uguale a 100 il dato per il 1881. Qual è il vantaggio della trasformazione eseguita del dato assoluto in numero indice? Non scorgiamo più, è vero, che nel 1923 sono stati importati 27.887.420 quintali di frumento, ma vediamo subito che l'importazione è stata circa 18 volte maggiore che nel 1881. Tutti i dati della serie sono ragguagliati al livello del 1881: la serie così trasformata ci mostra a colpo d'occhio in che relazione stia l'importazione di ciascun anno con quella dell'anno scelto a rappresentare la situazione iniziale.

Se abbiamo invece una serie che ci indichi l'importazione del frumento nei diversi Stati d'Europa in uno stesso anno, potremo tradurre anche questa in numeri indici, mettendo in rapporto tutti i dati con quello che si riferisce ad un determinato paese. Incontrando tale serie di dati nel preparare uno studio sull'approvvigionamento granario dell'Italia, probabilmente assumeremo come riferimento il dato che riguarda il nostro paese; se nella serie così tradotta in numeri indici troviamo per un certo Stato il numero 39, abbiamo bensì perduto di vista l'ammontare assoluto delle importazioni frumentarie di questo Stato, ma vediamo subito che corrisponde a circa due quinti dell'ammontare delle importazioni italiane, che abbiamo posto uguale a 100. La serie è stata trasformata in modo da servire più prontamente al nostro fine dell'esame comparativo dell'approvvigionamento granario dei vari Stati, dal punto di vista italiano. L'esempio ci mostra che il metodo dei numeri indici non si applica soltanto alla rappresentazione delle serie cronologiche, come suol credere taluno poco versato nelle applicazioni dei metodi statistici, ma anche ad altre serie. Si può dire anzi che sia difficile trovare una serie statistica alla quale esso non possa, per determinati fini, venire utilmente applicato.

Vediamo un altro esempio: abbiamo la serie dei salari medi giornalieri nelle varie industrie. Se li esaminiamo da un punto di vista generale, desiderosi d'indagare le condizioni comparative della rimunerazione del lavoratore nelle diverse occupazioni, potremo tradurli in numeri indici riferendoli tutti al salario medio aritmetico ponderato per il complesso degli operai; se li esaminiamo dal punto di vista particolare di una singola industria, alle condizioni della quale specialmente c'interessiamo, per esempio di quella della trattura della seta, potremo assumere come riferimento il salario medio in questa industria. Supposto il salario medio generale di 16 lire, il salario nella trattura della seta di 8 lire, il salario nella siderurgia di 24 lire, traducendo questi dati in numeri indici, col primo criterio, otteniamo 50 per gli operai dell'industria serica, 150 per quelli della siderurgia: vediamo subito che il salario giornaliero è nell'una industria inferiore del 50 %, nell'altra superiore del 50 % alla media generale. Col secondo criterio, otteniamo il numero indice 300 per la siderurgia, che ci dice come l'operaio vi percepisca un salario triplo di quello dell'operaio occupato nella trattura della seta. Anche in questo caso lo svantaggio della perduta visione del dato assoluto è compensato dal vantaggio della rapida e precisa visione della grandezza di ciascun dato in relazione a quella del dato scelto come riferimento.

3. — Senza che occorra moltiplicare gli esempi, è chiaro che il solo problema da risolvere nella traduzione di una serie originaria di dati statistici in una serie derivata di numeri indici è quello della scelta del riferimento, o della *base*, come si suol dire. Problema per la cui soluzione non si può dare alcuna norma generale, dovendo la scelta essere suggerita dal fine cui si mira nel calcolo dei numeri indici. A volta a volta potrà convenire di assumere a riferimento un dato della serie originaria, o una media di più dati, o una media dell'intera serie, oppure anche un dato estraneo alla serie ma omogeneo coi suoi termini (così, se mettiamo in rapporto i rendimenti per ettaro ottenuti dalla coltura della barbabietola nelle varie provincie italiane in cui è praticata, col rendimento medio per ettaro ottenuto nella Cecoslovacchia, per confrontare le condizioni della bieticoltura nazionale con quelle d'un mercato specialmente favorito da fattori fisici).

Abbiamo fin qui considerato casi nei quali una serie statistica viene tradotta in numeri indici aventi tutti la medesima base, nei quali cioè per tutti i termini della serie si adotta il medesimo riferimento. Ma accanto a queste *serie di numeri indici con riferi-*

mento costante esistono anche *serie di numeri indici con riferimento variabile*, ottenute col mettere in rapporto ciascun termine della serie con un diverso dato ad esso omogeneo e coordinato. Un fine del riferimento variabile può essere quello di agevolare l'apprezzamento comparativo dell'andamento di due fenomeni in relazione al tempo, o in relazione allo spazio, o in relazione ad altra circostanza qualitativa o quantitativa.

4. — Per formarci un'idea dell'applicazione pratica del procedimento dei numeri indici con riferimento variabile, consideriamo la serie dei dati sull'importazione della ghisa in Italia negli ultimi cinquant'anni, ponendole accanto la serie dei dati sulla produzione nazionale della ghisa negli stessi anni: avremo così per ciascun anno due dati, che metteremo in rapporto fra loro assumendo a numeratore quello che esprime l'importazione, a denominatore quello che esprime la produzione. Nel 1921 l'importazione è di tonnellate 65.692, la produzione di 61.381; nel 1928 l'importazione è di tonnellate 138.936, la produzione di 507.482. I numeri indici 107 per il 1921 e 27 per il 1928 ci nascondono le grandi variazioni avvenute nell'ammontare così dell'importazione come della produzione, ma ci fanno subito scorgere che nel primo anno la quantità importata supera del 7 °/₀ la quantità prodotta, nel secondo le rimane inferiore del 73 °/₀.

Dal precedente esempio potrebbe apparire che una serie di numeri indici con riferimento variabile esiga sempre per il calcolo due serie di dati ordinati secondo la medesima circostanza: quella che è oggetto diretto di indagine e un'altra, costituita da ugual numero di termini che ordinatamente servano di riferimento per i termini della prima. Quando le cose stanno così, si può dire che la serie dei numeri indici sia ad un tempo qualcosa di più e qualcosa di meno che la rappresentazione analitica dell'andamento di *una* serie di dati: essa costituisce, infatti, la rappresentazione dell'andamento comparativo di *due* serie di dati; ma appunto perciò ciascun termine di essa è funzione dei due termini corrispondenti delle due serie che concorrono inscindibilmente a formarlo. Se, nel nostro esempio, troviamo che ad un anno corrisponde il numero indice 50, all'anno successivo il numero indice 60, non possiamo conchiudere niente sulla variazione avvenuta nell'importazione della ghisa nè su quella avvenuta nella produzione di essa, separatamente considerate: possiamo soltanto affermare che il rapporto fra importazione e produzione si è modificato in senso favorevole all'importazione, nella misura che indica il confronto tra i due numeri indici; ed è chiaro che la modificazione può essere avvenuta perchè sia aumentata l'importazione in proporzione maggiore della produzione, o diminuita in proporzione minore; oppure perchè sia aumentata l'importazione e diminuita o rimasta stazionaria la produzione; oppure ancora perchè sia rimasta stazionaria l'importazione e sia diminuita la produzione. Quale delle cinque possibili ipotesi corrisponda al vero, il numero indice non dirà mai; soli i dati assoluti potranno chiarirlo.

5. — Ma una serie di numeri indici con riferimento variabile non deriva sempre dalla comparazione di grandezza fra i termini di due serie ordinatamente corrispondentisi provenienti da rilevazioni diverse; talvolta la seconda serie, assunta come riferimento, non è altro che la prima con tutti i numeri d'ordine dei termini spostati di un'unità, o di più unità (ma in questo caso dello stesso numero d'unità per tutti i termini).

Consideriamo ancora la serie dei dati sulla produzione della ghisa in Italia negli ultimi cinquant'anni; riferiamo ora ciascuno di essi a quello che lo precede in ordine cronologico, per esempio il dato del 1920 a quello del 1919, il dato del 1922 a quello del 1921 (la serie di numeri indici si ridurrà a 49 termini, a meno che ci sia possibile conoscere la produzione della ghisa anche nell'anno antecedente a quello iniziale della nostra serie). I numeri indici così ottenuti segnano immediatamente la variazione della produzione da ciascun anno al successivo: il numero indice 37 per il 1920 ci mostra che in quell'anno la produzione è diminuita del 63 °/₀ in confronto al 1919, il numero indice 255 per il 1922 ci mostra che la produzione è aumentata del 155 °/₀ in confronto al 1921. Il non avere sott'occhio i numeri assoluti (239.710 t. nel 1919, 88.072 nel 1920, 61.381 nel 1921, 157.599 nel 1922) ci rende arduo un apprezzamento generale dell'andamento della serie, ma l'esame dei numeri indici ci mostra immediatamente il senso e la grandezza relativa della variazione avvenuta da ciascun anno al successivo.

Serie di numeri indici come quella ora descritta diconsi *a catena, con riferimento immediato*: normalmente quando si parla di numeri indici a catena, senz'altro, si accenna a questo tipo. La serie di rapporti indici dai quali è tratta la serie dei numeri indici è costituita da rapporti concatenati, perchè ciascuno di essi ha come denominatore lo stesso dato che compare come numeratore nel rapporto che lo precede. Il prodotto dei primi k rapporti indici risulta pertanto uguale al valore del rapporto fra il $(k+1)^{mo}$ termine e il primo termine della serie originaria di dati; così che dalla serie dei numeri indici a catena si può agevolmente desumere per successive moltiplicazioni una serie di numeri indici aventi come riferimento costante il termine iniziale della serie originaria.

Serie di numeri indici a catena con riferimento immediato si possono ricavare non soltanto da serie cronologiche, ma anche da altre serie ordinate secondo i valori d'una circostanza quantitativa. Se u_0 è il numero dei velivoli partiti in una gara a più tappe, come il giro d'Italia, e u_1 il numero di quelli arrivati alla prima tappa, u_2 il numero degli arrivati alla seconda tappa, e così via, dalla serie dei dati u_0, u_1, u_2, ... possiamo desumere una serie di rapporti indici a catena $(u_1 : u_0)$, $(u_2 : u_1)$, $(u_3 : u_2)$, ..., i cui dati non ci dicono più il numero assoluto dei velivoli arrivati a ciascuna tappa, ma ce lo esprimono in rapporto a quello degli arrivati alla tappa precedente.

In certi casi conviene riferire ciascun termine della serie originaria, per la formazione del numero indice, non al termine immediatamente precedente ma al termine precedente di due, di tre o di più posti: o perchè si voglia mettere in evidenza la variazione del fenomeno in un intervallo più ampio di quello unitario, o perchè il fenomeno stesso presenti periodicità di movimento che consiglino di cercare l'indicazione della tendenza generale nel paragone tra i dati riferentisi a fasi reciprocamente corrispondenti, piuttosto che tra quelli riferentisi a fasi immediatamente successive l'una all'altra. Numeri indici così calcolati si possono chiamare *a catena con riferimento mediato*.

Se un raccolto agrario presenta spiccate alternative di abbondanza e di scarsezza da un anno all'altro, per scorgere meglio l'andamento dominante attraverso il tempo possiamo calcolare numeri indici mettendo in rapporto il raccolto di ciascun anno con quello di due anni prima: l'anno pari con l'anno pari precedente, l'anno dispari con l'anno dispari precedente. Qui il denominatore precede di due posti il numeratore, invece che di un posto solo come nella catena immediata. Se conosciamo il numero dei superstiti di una certa generazione, alle età di 0, di 1, di 2, di 3 anni, e così via di anno in anno, possiamo riferire ciascun dato a quello che lo precede di tre posti (i superstiti a 3 anni ai superstiti a 0 anni, i superstiti a 4 anni ai superstiti a 1 anno, ecc.), per vedere prontamente come varii il numero dei superstiti di tre in tre anni. Se abbiamo i dati sul valore delle importazioni, mese per mese, per un certo numero di anni, possiamo tradurli in numeri indici a catena mediata mettendo in rapporto il dato per il mese di gennaio di un anno col dato per il mese di gennaio dell'anno precedente, quello pel febbraio col dato per il febbraio precedente, e così via. Qui l'intervallo fra i dati messi in rapporto è di dodici posti: il procedimento mira a mettere in rilievo le tendenze che si manifestano nelle importazioni, prescindendo dalle variazioni stagionali.

Nel presentare una serie di numeri indici a catena con riferimento mediato è bene spiegare sempre chiaramente la costruzione di essa, affinchè non venga confusa con una serie a catena immediata e quindi interpretata erroneamente.

6. — Nelle serie di numeri indici a catena il riferimento di ciascun dato è determinato, in modo indipendente dall'arbitrio dell'osservatore, dall'ordinamento stesso della serie, e, nel caso della catena con riferimento mediato, da ben definiti caratteri di periodicità del fenomeno, o dallo scopo dell'indagine. Ma in ogni altro caso la traduzione di una serie di dati in numeri indici presuppone la scelta del riferimento, costante o variabile, da parte dell'operatore. Questa scelta ha importanza decisiva, perchè secondo il riferimento adottato possono molto differire le impressioni di chi esamina la serie dei numeri indici. Se riferiamo i dati sulla produzione italiana del frumento negli ultimi cinquant'anni al dato iniziale, la serie dei numeri indici ci dà l'impressione di una dominante tendenza all'aumento; se li riferiamo invece, anno per anno, ai corrispondenti dati sull'importazione del frumento, vediamo dominare la tendenza alla diminuzione rispetto a quelli; se li riferiamo, ancora anno per anno, ai corrispondenti dati sulla produzione del granturco, vediamo prevalere la tendenza all'aumento; se calcoliamo i numeri indici a catena, vediamo alternarsi aumenti e diminuzioni, con prevalenza degli uni sulle altre. Le varie serie di numeri indici così ottenute ci danno impressioni diverse e contrastanti, ma non contraddittorie poichè corrispondono a diversi punti di vista dai quali abbiamo considerato lo stesso fenomeno. La prima serie ci indica il progresso della produzione frumentaria nazionale; la seconda ci mostra che tale progresso è inadeguato all'aumento del consumo; la terza ci avverte della crescente preferenza per la coltura del frumento in confronto a quella del granturco; la quarta ci rivela, meglio della prima, le forti variazioni della produzione da un anno all'altro. È chiaro che scegliendo accortamente

l'uno o l'altro di questi riferimenti, o un altro ancora degli innumerevoli riferimenti possibili, si potrà, secondo si desidera, suscitare nel lettore scarsamente dotato di senso critico un'impressione favorevole, oppure un'impressione sfavorevole, sull'andamento della produzione frumentaria italiana, con facilità tanto maggiore in quanto chi si trova innanzi soltanto la serie dei numeri indici non ha la possibilità di completare e correggere la propria impressione mediante l'esame delle cifre assolute. Contro quest'insidia, talvolta volontariamente tal altra involontariamente, tesa dai compilatori di numeri indici ai loro lettori, bisogna esser sempre vigili, perchè specialmente chi conosca poco la materia cui i dati si riferiscono può facilmente incapparvi.

7. — Occorre talora, a chi disponga d'una serie di numeri indici con riferimento costante, di mutare il riferimento. Se, per esempio, vogliamo mettere in rilievo le ripercussioni dirette e indirette della guerra sull'economia italiana, potremo giudicare opportuno di trasportare il riferimento dei dati sulla produzione annuale della ghisa dal 1881 al 1913. È ovvio che a tale intento possiamo ricalcolare l'intera serie di numeri indici; ed è ciò che faremo se disponiamo dei dati originari. Ma se non ne disponiamo potremo ugualmente calcolare i numeri indici riferiti alla nuova base (1913) dividendo i numeri indici riferiti alla vecchia base (1881) per il valore del rapporto tra la produzione del 1913 e quella del 1881, cioè del rapporto indice che compare nella vecchia serie in corrispondenza all'anno 1913 (per eseguire più rapidamente il calcolo, sopratutto se disponiamo d'una macchina calcolatrice, invece che dividere i numeri indici per il valore del suddetto rapporto ci converrà moltiplicarli per il reciproco di esso). In via generale, potremo sempre trasformare una serie di numeri indici calcolati con riferimento costante ad una base u_k in una serie di numeri indici riferiti ad un'altra base u_h, moltiplicando ciascun termine della prima serie per il valore del rapporto ($u_k : u_h$), cioè per il reciproco del rapporto ($u_h : u_k$) che compare nella nostra serie. Questo procedimento ha il vantaggio che può applicarsi anche se non si conoscono i dati originari.

8. — A ciascun rapporto indice corrisponde una differenza relativa fra il numeratore e il denominatore del rapporto stesso, uguale al valore del rapporto diminuito dell'unità. Moltiplicando il rapporto per 100, 1000, ecc. si moltiplica per lo stesso fattore anche la differenza relativa, che (così moltiplicata) si può ottenere

sottraendo dal numero indice 100, 1000, ecc. Se abbiamo posto uguale a 100 la base d'una serie di numeri indici, sottraendo 100 dai termini di questa serie otteniamo pertanto un'altra serie di numeri (che possono essere positivi, nulli o negativi), i quali ci rappresentano le deviazioni o differenze percentuali dei dati della serie originaria rispetto alla base o alle basi cui questi sono stati riferiti. Se il numero indice del prezzo del pane nell'anno 1929, con base il prezzo del 1913 posto uguale a 100, risulta uguale a 500, la differenza (500 — 100) misura la variazione percentuale avvenuta nel prezzo del pane dal 1913 al 1929, segnando un aumento del 400 %.

Spesso viene inesattamente espressa una simile relazione: incontrando il numero indice 500 si dice che il prezzo del pane è aumentato *del* 500 % o *del* quintuplo, il che evidentemente è errato: si può dire che è aumentato *al* 500 % o *al* quintuplo, ma è preferibile, ad evitare equivoci, dire che si è avuto un aumento del 400 %.

Un altro errore che talvolta si commette nell'enunciare le variazioni relative misurate dai numeri indici consiste nel riferire la variazione al numeratore invece che al denominatore del rapporto indice. Posta uguale a 100 l'esportazione italiana della seta greggia nel 1906 (anno che ha segnato il massimo sviluppo di questa corrente commerciale), l'esportazione del 1922 risulta espressa dal numero indice 42, e la diminuzione dal 1906 al 1922 da 58. Diminuzione del 58 %, dunque; e non diminuzione del 138 %, come direbbe chi misurasse la diminuzione relativa dal rapporto 58 : 42 invece che dal rapporto 58 : 100: procedimento assurdo in quanto riferisce la diminuzione allo stato finale, invece che allo stato iniziale come la logica richiede. (In certi casi può servire anche il riferimento allo stato finale, non però allo scopo di misurare la grandezza relativa della diminuzione. Se avevo 100 lire e me ne ritrovo sole 20, potrò bensì dire che le 80 lire perdute corrispondono al 400 % delle 20 che mi restano; ma non potrò dire che ho sofferto una perdita relativa del 400 % del mio avere; dovrò dire che ne ho perduto l'80 %. Non sempre il riferimento di una diminuzione allo stato finale è possibile: se perdo tutte 100 le lire che possedevo, posso dire di aver perduto il 100 % del mio avere, riferendomi allo stato iniziale; ma poichè mi restano 0 lire non posso riferire la perdita allo stato finale: otterrei un valore del rapporto privo di ogni corrispondenza numerica concreta, e superiore ad ogni numero finito).

Qui torna opportuno avvertire che nel calcolo dei numeri indici con riferimento costante non si potrà mai scegliere come riferimento un dato uguale a zero; e che nel calcolo dei numeri indici con riferimento variabile nessuno dei denominatori da assumersi per i rapporti indici deve essere nullo.

9. — Poichè il numero indice non è altro che il valore di un rapporto moltiplicato per 100 o per un'altra potenza di 10, le nozioni che abbiamo dato intorno alle relazioni tra l'errore di un rapporto e gli errori dei termini trovano qui applicazione, nè occorre esporle nuovamente. Soltanto vale la pena di notare che, se tutti i termini di una serie sono errati nello stesso senso, traducendoli in numeri indici con riferimento costante ad un termine o ad una media di più termini della serie stessa si attenueranno gli errori relativi; se i termini sono errati in parte in eccesso e in parte in difetto, il riferimento ad una base costante errata anch'essa aggraverà alcuni degli errori relativi, attenuerà altri; e ancora in quest'ultimo caso, la determinazione di numeri indici a catena in generale aggraverà gli errori, perchè spesso accadrà di mettere in rapporto un dato errato in eccesso con uno errato in difetto, e viceversa. Questi non sono che esempi dell'influenza degli errori dei termini nella formazione dei numeri indici: una teoria generale non si può costruire, data l'infinita molteplicità dei casi che si possono presentare. In ogni applicazione pratica conviene tener presente che nella traduzione di una serie in numeri indici si possono talvolta attenuare ma tal altra aggravare gli errori relativi dai quali sono affetti i dati della serie: bisogna pertanto usare il metodo con l'opportuna circospezione.

10. — La rappresentazione di una singola serie statistica mediante numeri indici non ha, in generale, grande utilità. Il vantaggio del metodo si palesa specialmente quando si hanno da comparare numerose serie di dati, ordinati secondo la stessa circostanza in modo tale che a ciascun dato di una serie corrisponda univocamente un dato in ciascuna delle altre serie. Tradotte in numeri indici le serie, con un unico criterio nella scelta del riferimento, i dati corrispondenti delle varie serie divengono agevolmente paragonabili tra loro, non nella grandezza assoluta, s'intende, ma nella grandezza relativa alla base adottata.

Supponiamo di conoscere il rendimento medio per ettaro di diverse colture agrarie nelle varie regioni italiane. Avremo tante serie, di 18 dati ciascuna, quanti sono i prodotti considerati. Ognuna di queste serie si può tradurre in numeri indici, assumendo come riferimento il dato medio generale per l'Italia o il dato per una regione che particolarmente ci interessi, per esempio la Lombardia. Assunto quest'ultimo riferimento, calcoliamo i numeri indici per un'altra regione: la Campania. Ricaviamo, dai dati del 1929, i seguenti numeri indici: frumento 48, segale 63, orzo 94, avena 54, granturco 38, patate 35, ecc. (potremmo aggiungere altri prodotti, ma al nostro fine è superfluo). Non vediamo più, da questi numeri indici, quale sia il rendimento delle varie colture nella Campania, ma vediamo che, per tutte le colture considerate, esso è più basso che nella Lombardia: il rendimento campano costituisce infatti, secondo i prodotti, dal 35 al 94 °/₀ del rendimento lombardo. Saremmo quasi tentati di assumere una media dei nostri numeri indici (opportunamente completati, s'intende) come espressione sintetica del rendimento conseguito dall'agricoltore campano in relazione a quello dell'agricoltore lombardo; e poichè altrettanto potremmo fare per ogni altra regione, saremmo alla fine in grado di ordinare le regioni italiane secondo il rendimento unitario conseguito sui loro territori. Ma di simili tentativi di sintesi parleremo in altro capitolo: qui ci basti notare che la comparazione tra i vari dati delle serie, diremo così, parallele è grandemente agevolata mercè la traduzione in numeri indici.

11. — Il diagramma lineare che rappresenta la serie originaria può servire anche per la rappresentazione della serie tradotta in numeri indici con riferimento costante: basta porre accanto alla scala dei dati originari un'altra scala, lo 0 della quale corrisponda allo 0 della prima e il 100 della quale corrisponda a quel valore della prima scala che si assume come riferimento nel calcolo dei numeri indici. Mediante le due scale si possono leggere così i dati originari come i corrispondenti numeri indici.

Volendo raccogliere in una sola tavola più diagrammi che rappresentino altrettante serie ordinate secondo la stessa circostanza, in modo tale che consenta di leggere sia i valori originari sia i numeri indici, dobbiamo anzitutto fissare la scala comune per i numeri indici di tutte le serie, poi porle accanto tante diverse scale quante sono le serie di dati originari, in modo tale che allo 0 della scala dei numeri indici corrisponda lo 0 di tutte le scale dei dati originari e al 100 della scala dei numeri indici corrisponda nella scala di ciascuna serie il valore che assumiamo come riferimento per la traduzione della serie stessa in numeri indici: valore che deve essere stabilito con un unico criterio per tutte le serie originarie, altrimenti verrebbe meno un'essenziale condizione per la comparabilità fra le diverse serie di numeri indici. Se, per esempio, traduciamo in numeri indici i dati sul valore delle esportazioni italiane verso vari Stati nel corso degli ultimi cinquant'anni, assumendo come riferimento il dato iniziale di ciascuna serie, nella rappresentazione grafica faremo corrispondere al valore 100 della scala dei numeri indici l'ordinata iniziale di tutte le curve che rappresentano le esportazioni verso i vari Stati; vedremo indi le curve divergere, e poi svolgersi separatamente od incrociarsi secondo le vicende dei traffici. Se assumiamo invece come riferimento il dato per un anno intermedio del cinquantennio, vedremo tutte le curve convergere nel punto di ordinata 100 e di ascissa corrispondente a quell'anno, indi divergere di nuovo.

12. — Un altro metodo di rappresentazione delle serie statistiche, affine a quello dei numeri indici, consiste nel riferimento dei termini della serie alla loro somma: nella traduzione, cioè, dei dati originari in rapporti di composizione, i valori dei quali si sogliono poi esprimere nella forma di proporzioni a 100, a 1000 o ad altra potenza di 10. Con questo procedimento si riducono, idealmente, alla stessa dimensione, e si rendono così comparabili fra loro, complessi di osservazioni che realmente sono di grandezza diversa e quindi non direttamente comparabili. Volendo, per esempio, confrontare la costituzione economica di due regioni, Lombardia e Basilicata, ci converrà paragonare le proporzioni nelle quali i singoli gruppi di occupazioni concorrono a formare la popolazione delle due regioni, piuttosto che le cifre assolute. Com'è ovvio, il procedimento è adatto specialmente, se non esclusivamente, alla rappresentazione di quelle serie i cui termini si possono riguardare come misure delle parti di un unico complesso, la misura del quale corrisponde alla somma dei termini stessi (suddivisione per età di una popolazione, suddivisione per destinazioni del valore delle merci esportate, suddivisione secondo le cause di morte dei decessi accertati in un paese). Poichè serie di tal sorta s'incontrano molto spesso, il procedimento di cui ora discorriamo è largamente impiegato, anche per merito della sua grande semplicità. Gioverà tener presenti nell'applicazione le avvertenze che abbiamo esposto trattando dei rapporti di composizione.

Quesiti ed esercizi: 1. — Che cosa significa «tradurre in numeri indici una serie statistica»? Tutte le serie si prestano a tal modo di rappresentazione? Qual è il fine di esso? Quali sono i vantaggi e quali gli svantaggi della traduzione di una serie in numeri indici?

2. — Oltre che alle serie cronologiche, a quali altre serie è utilmente applicabile la traduzione in numeri indici? Si cerchino nell'ASI serie che sono state effettivamente tradotte in numeri indici ed altre che potrebbero essere utilmente presentate in tal forma.

3. — Tradurreste in numeri indici la serie dei dati sulla popolazione delle varie provincie italiane? quella sulla densità della popolazione nelle varie provincie? A quali fini potrebbe servire l'una operazione; a quali l'altra? Quale riferimento adottereste nel primo caso? nel secondo?

4. — Con quale criterio di riferimento tradurreste in numeri indici i dati sui prezzi medi mensili dei principali prodotti agricoli contenuti nell'ASI, capitolo «Prezzi»? Provate l'applicazione di vari criteri che vi sembrino plausibili ed esponete le considerazioni che quest'esperienza vi suggerisce.

5. — Si traducano in numeri indici le seguenti serie di «Notizie statistiche retrospettive» contenute nell'ASI: *a*) popolazione; *b*) produzione del frumento; *c*) produzione del vino; *d*) produzione della birra. I dati si riferiscono all'ultimo cinquantennio; con quali criteri si può fissare il riferimento per i numeri indici? A quali fini corrispondono i diversi criteri? Quale criterio vi pare preferibile, volendo confrontare lo sviluppo delle varie produzioni sopra enumerate? Paragonate fra loro le serie *a*) e *b*); le serie *c*) e *d*); le serie *b*) e *c*).

6. — Si rappresentino in diagramma due delle serie cronologiche indicate nel precedente esercizio, in modo che il diagramma permetta di leggere così i dati assoluti come i numeri indici.

7. — Che cos'è la «base» di una serie di numeri indici?

8. — Volendo tradurre in numeri indici la serie cronologica dei dati sulla produzione italiana dell'acciaio, con riferimento al 1913, sarà opportuno porre il dato del 1913 uguale a 100 o a 1000?

9. — Si traduca in numeri indici la serie dei dati sulla produzione italiana della ghisa negli anni 1881-1910, adottando successivamente i seguenti riferimenti: *a*) la produzione del 1881, *b*) la produzione del 1910, *c*) la produzione del 1896, *d*) la produzione del 1929, *e*) la produzione media aritmetica del periodo 1881-1910. Quali considerazioni suggerisce la comparazione tra i risultati dei vari calcoli? A quali fini corrisponde ciascuno di questi?

10. — Esaminando i numeri indici raccolti nel capitolo «Prezzi, salari e consumi» dell'ASI, si cerchi d'intendere: a quali fini siano stati calcolati, con quali criteri siano stati scelti i riferimenti, quali vantaggi presentino le serie di numeri indici in confronto alle serie originarie dalle quali sono state ricavate.

11. — Si traducano in numeri indici i dati sui salari delle diverse industrie ad una stessa epoca contenuti nell'ASI.

12. — Si traducano in numeri indici i dati sulla frequenza dei furti denunciati nei vari compartimenti giudiziari in un determinato anno. Con quale criterio si sceglierà il riferimento?

13. — Quali tipi conoscete di numeri indici con riferimento variabile? Che cosa sono i numeri indici a catena?

14. — Si traduca in numeri indici a catena la serie dei dati sulla frequenza delle nascite nelle varie provincie italiane in un determinato anno.

15. — Si traducano in numeri indici i dati sulla produzione italiana della birra negli ultimi cinquant'anni, assumendo come riferimento i corrispondenti dati sulla produzione del vino. Quali considerazioni suggeriscono i risultati del calcolo?

16. — Si traducano in numeri indici i dati sulle importazioni italiane dalla Gran Bretagna negli ultimi cinquant'anni, assumendo come riferimento i dati sulle importazioni italiane dagli Stati Uniti negli anni corrispondenti. Si interpretino i risultati.

17. — Si traducano in numeri indici a catena i dati sul corso medio del cambio su Parigi negli anni dal 1884 al 1928 contenuti nell'appendice retrospettiva dell'ASI.

18. — Si traducano in numeri indici le serie cronologiche: *a*) dei massimi annuali, *b*) dei minimi annuali, del corso del cambio su Londra, assumendo come riferimento per ciascun anno il corso medio dell'anno stesso. Si interpretino i risultati del calcolo. Perchè si hanno risultati tanto differenti per il 1910 e per il 1920?

19. — Si traducano in numeri indici a catena i dati sull'eccedenza dei nati sui morti in Italia negli anni dal 1910 al 1929. Quali considerazioni suggeriscono i risultati del calcolo? Si confrontino tali risultati con quelli ottenuti assumendo per riferimento costante il dato del 1910. Come avrebbe potuto ricavarsi dalla prima questa seconda serie di numeri indici?

20. — Si traducano in numeri indici i dati mensili sulle nascite avvenute in Italia negli ultimi quattro anni, adottando come riferimento per ciascun dato mensile il dato per il mese omonimo dell'anno immediatamente precedente. Sono, questi, numeri indici a catena con riferimento immediato?

21. — Si traducano in numeri indici i dati mensili sulle nascite avvenute in Italia negli ultimi quattro anni, riferendo il dato di ciascun mese a quello del mese immediatamente precedente. A qual fine corrispondono i numeri indici così calcolati? a quale i numeri indici calcolati nell'esercizio precedente?

22. — Come si possono trasferire ad una nuova base i numeri indici riferiti ad una data base? Si applichi il procedimento ai numeri indici calcolati nell'esercizio 9 con riferimento al 1910, per trasferirli alla base 1896; si confrontino i risultati con quelli ottenuti dal calcolo diretto.

23. — Se il numero indice della produzione della seta artificiale passa da 100 a 323, possiamo affermare che tale produzione è aumentata del 323 per 100?

24. — Come influiscono nella traduzione di una serie in numeri indici gli errori dai quali siano affetti i termini della serie stessa?

25. — Si può tradurre in numeri indici la serie dei dati sulle esportazioni italiane verso la Germania, assumendo come riferimento il 1917?

26. — Si traducano in diagramma, in un unico quadro, i numeri indici dei prezzi al minuto del pane, del riso, delle uova, dello zucchero, del burro, nei vari mesi di uno stesso anno, riportati nell'ASI. Quali considerazioni suggerisce l'esame del diagramma?

27. — Si calcolino numeri indici della natalità e della mortalità per i vari compartimenti italiani in un dato anno, e si pongano a confronto in un istogramma a canne d'organo e in un diagramma a curva continua.

28. — Dalle serie di numeri indici calcolate negli esercizi 5, 9, 11, 12 si desumano le corrispondenti serie di deviazioni percentuali.

29. — Se si è rappresentata in diagramma una serie di numeri indici, come si può sul diagramma stesso rappresentare la serie delle deviazioni percentuali?

30. — Come si possono rappresentare in un unico diagramma più serie di numeri indici in modo tale che si possano leggere così i numeri indici come i corrispondenti dati originari?

31. — Se dal 1913 al 1930 il numero indice del prezzo del pane è salito da 100 a 425 in una città, da 100 a 410 in un'altra, in quale delle due città è più basso il prezzo del pane?

32. — Quali serie sono specialmente adatte ad essere rappresentate mediante la traduzione dei loro termini in rapporti di composizione? Ricercate nell'ASI esempi di applicazione di questo procedimento e in ciascun caso mettete in risalto il vantaggio ottenuto mediante tale applicazione. Provate ad eseguire tra due serie così tradotte comparazioni che sarebbe stato malagevole eseguire sui dati originari.

33. — Traducete in rapporti di composizione la serie delle temperature medie mensili registrate in un osservatorio italiano in un dato anno; traducete in numeri indici la serie delle quantità d'acqua cadute nei vari mesi di un anno secondo gli accertamenti di un osservatorio italiano (ASI, capitolo «Climatologia»). Vi sembra opportuna l'applicazione dei due procedimenti? Il vostro buon senso suggerisce qualche rettifica al precedente tema?

34. — Come potreste utilmente ridurre a numeri indici i valori stagionali delle portate dei principali fiumi italiani riportati nell'ASI?

CAPITOLO XIV.

L'applicazione dei vari modi di rappresentazione ai vari tipi di serie.

Limitazioni all'applicazione dei vari modi di rappresentazione — Applicazione di essi ai vari tipi di serie: serie cronologiche; altre serie indicanti l'andamento di un fenomeno in funzione di una circostanza quantitativa; distribuzioni di caratteri quantitativi; serie di termini disposti in ordine di grandezza; distribuzioni di caratteri qualitativi; distribuzioni di fenomeni nelle varie sezioni di un campo d'osservazione; serie geografiche e topografiche.

1. — I vari modi di rappresentazione delle serie statistiche convengono in diverso grado alla rappresentazione dei vari tipi di serie. Alcuni di questi modi servono esclusivamente per certi tipi di serie: così il cartogramma per le serie geografiche o topografiche; altri servono per più tipi, contraddistinti da caratteri comuni: così l'interpolazione può applicarsi alle serie cronologiche, alle distribuzioni di caratteri quantitativi e ad altri tipi di serie che

rientrano nella categoria delle funzioni statistiche; altri ancora, come il diagramma, sono di applicazione assolutamente generale. Passando in rassegna i vari modi di rappresentazione abbiamo già indicato i limiti dell'uso di ciascuno di essi; ora passando invece in rassegna, molto più brevemente, i principali tipi di serie statistiche, potremo meglio delimitare il campo delle applicazioni di ciascun modo di rappresentazione.

2. — Le serie cronologiche in generale offrono una nitida visione delle loro caratteristiche più spiccate quando siano rappresentate in diagramma (non ugualmente utile riesce l'istogramma, perchè meno facilmente e meno prontamente interpretabile). Se il fenomeno si svolge con qualche regolarità di tendenza — oscillazioni intorno ad un livello costante, variazione prolungata in un determinato senso, ciclicità o periodicità delle variazioni, regolare alternativa di variazioni verso l'alto e verso il basso, ecc. — l'esame della spezzata o della curva del diagramma può rivelare tale regolarità anche ad un occhio inesperto. L'occhio esperto, poi, riesce a distinguere fra l'intreccio delle diverse specie di variazioni il dominare d'una tendenza, l'agire di fattori ciclici o periodici, il comparire di fattori saltuari. La scissione delle varie specie di variazioni può essere tentata, in certi casi, mercè l'impiego del procedimento interpolatorio, applicato ad eliminare da prima la variazione evolutoria, poi — nel residuo — le variazioni cicliche e periodiche, poi — nel successivo residuo — le variazioni saltuarie, ecc.; ma di codeste eleganti applicazioni conviene usare con grande cautela, e solo conoscendo profondamente la materia cui si riferiscono, altrimenti è facile abusarne e andar lontano dalla realtà. Il metodo della media mobile può non di rado surrogare quello dell'interpolazione; anch'esso, però, ha bisogno di essere applicato con competenza.

La traduzione in numeri indici è utile specialmente se si hanno da paragonare, o da sintetizzare insieme, più serie cronologiche: nell'eseguirla si deve porre la massima cura nella scelta del riferimento, per non determinare impressioni inesatte in chi si troverà poi innanzi soltanto la serie dei numeri indici. Per l'analisi delle variazioni d'un fenomeno nel tempo può essere vantaggioso il procedimento dei numeri indici a catena, che appunto in questo campo trova il suo principale impiego. Anche il riferimento dei termini di una serie cronologica a quelli corrispondenti di un'altra serie dello stesso tipo è frequentemente usato.

La media può riassumere in modo soddisfacente una serie cronologica, sia quando l'andamento del fenomeno è oscillante senza netta tendenza all'aumento o alla diminuzione, sia quando le deviazioni da un livello costante si possono riguardare per la massima parte dovute all'azione di fattori ciclici o periodici; in entrambi i casi si potranno utilmente calcolare medie degli scostamenti, o numeri indici col riferire alla media i singoli termini della serie. Ma normalmente la media non è adatta a rappresentare le serie cronologiche perchè nasconde completamente quel movimento che è il carattere precipuo di molte di esse; nè offrono sufficienti rimedi i dati sussidiari perchè, pure indicando l'ampiezza delle variazioni, non mostrano la distribuzione di queste attraverso il tempo che è caratteristica di ciascun fenomeno, eccettuati i casi nei quali l'ordine cronologico dei dati coincide con l'ordine di essi per grandezza. Quanto finora si è detto vale tanto per le serie cronologiche di dati greggi quanto per quelle di dati elaborati (rapporti, medie, ecc.).

3. — Considerazioni analoghe a quelle dianzi esposte si possono ripetere riguardo a serie non rigorosamente cronologiche, ma affini, come quelle che indicano variazioni di un fenomeno o di un carattere in funzione dell'età; e più generalmente riguardo alla maggior parte delle serie che indicano l'andamento di un fenomeno in funzione di una circostanza quantitativa continua (che può essere lo spazio lineare, od altra), e ad una parte delle funzioni di variabili discontinue.

4. — Valgono considerazioni diverse per le serie che si indicano col nome di «distribuzioni di frequenze» o «distribuzioni di caratteri quantitativi» e che presentano il risultato dell'ordinamento di un insieme di casi individuali, osservati od accertati, secondo la misura di una circostanza quantitativa che è caratteristica di ciascuno di essi. Qui la media ha maggiore attitudine rappresentativa, specialmente se il carattere si presenta con una certa dimensione normale che appare più frequente di ogni altra e intorno alla quale le altre si addensano; in tal caso, il valore normale è tanto più adatto a rappresentare la serie quanto più strettamente e quanto più simmetricamente intorno ad esso si addensano i termini della serie. Della dispersione dei termini intorno alla media offrono misure dirette alcuni dati sussidiari (medie di scostamenti, coefficienti di variazione); altri ne porgono indizi (termini estremi della serie; quartili, decili ed altri dati analoghi); altri infine ne danno misure indirette (medie di differenze). Dell'asimmetria dà un indice la media aritmetica semplice degli scostamenti dal valore normale, o la

loro media cubica. Per le distribuzioni di caratteri quantitativi una media dei termini ed una media degli scostamenti talora sono sufficienti a rappresentare in modo adeguato la serie: abbiamo visto, per esempio, nel capitolo sull'interpolazione, che questi due dati bastano per determinare i parametri della curva degli errori, e quindi per fornire una descrizione completa della serie quando tale forma di curva le si adatta bene.

La forma della distribuzione del carattere quantitativo si rivela prontamente, anche all'indagatore novellino, attraverso la rappresentazione in diagramma. Una forma che spesso si presenta è quella in cui, nel graduale passaggio dalle misure più basse alle più alte del carattere, la frequenza con la quale si presentano le varie misure prima cresce gradualmente e poi gradualmente diminuisce, di modo che la curva assume aspetto piramidale o più sovente campanulare se la distribuzione è simmetrica attorno al valore normale, e si deforma tanto più quanto più la distribuzione è asimmetrica (esempi di distribuzione campanulare: approssimativamente simmetrica, stature di coetanei; moderatamente asimmetrica, salari degli operai di un'industria; notevolmente asimmetrica, età degli sposi o delle spose; fortemente asimmetrica, redditi individuali in un paese, in regime capitalista). Meno frequente è la forma in cui si presentano due o più massimi, in cui cioè nel passare gradualmente dalla misura più bassa del carattere alla più alta, la frequenza prima aumenta, poi diminuisce, poi aumenta ancora, poi diminuisce ancora, ecc.: forma che talvolta si può considerare determinata dall'interferenza di due o più distribuzioni di tipo campanulare. Rara è la forma di distribuzione detta ad U, che è in certo modo opposta alla prima considerata sopra perchè la frequenza con la quale si presentano le varie misure del carattere segna un massimo in corrispondenza alla misura minima del carattere, poi discende gradualmente fino a toccare un minimo in corrispondenza ad una misura intermedia del carattere, per indi risalire fino ad un nuovo massimo corrispondente alla misura massima del carattere. Meno raramente s'incontrano le distribuzioni dette a J, nelle quali la frequenza sale gradualmente a partire dalla misura minima fino alla misura massima (esempio: distribuzione per età dei morti di una generazione dai 20 anni in su), o viceversa (esempio: distribuzione per età di una popolazione). Per tutte queste forme di curve sono state trovate funzioni atte a rappresentarle, e quindi riesce in generale possibile l'interpolazione analitica: non però sempre utile, perchè parecchie di tali funzioni mancano dei requisiti di semplicità necessari affinchè l'interpolazione sia praticamente di aiuto nell'apprezzamento della serie e nella sua comparazione con altre dello stesso genere.

È raramente impiegato per le distribuzioni di caratteri quantitativi il metodo dei numeri indici: il riferimento costante è aritmeticamente possibile ma in generale logicamente inutile; il riferimento a catena per lo più non serve ad alcun fine d'indagine sul fenomeno rappresentato; soltanto il riferimento variabile, nella forma di comparazione tra due serie, termine a termine, può talvolta giovare. È invece frequentemente usata la traduzione dei termini della serie in rapporti di composizione, sia perchè agevola la comparazione fra i vari termini della stessa serie, sia perchè rende più facili le comparazioni tra più serie.

5. — Una qualsiasi serie statistica può essere disposta in modo che i suoi termini si succedano in ordine di grandezza, senza riguardo ad alcun altro criterio di classificazione. Per esempio, se abbiamo i dati sulla natalità in ciascun comune italiano, potremo proporci di studiare la distribuzione per grandezza di questi dati, prescindendo dalla situazione geografica dei vari comuni, dalla loro importanza demografica, ecc. A siffatte serie di dati disposti in ordine di grandezza si possono applicare, in generale, le considerazioni esposte or ora a proposito delle distribuzioni di caratteri quantitativi.

6. — Le serie che indicano la distribuzione di caratteri qualitativi, cioè la suddivisione di un insieme di casi osservati od accertati secondo la modalità di una circostanza qualitativa, possono essere utilmente rappresentate mediante istogrammi (diagrammi a canne d'organo, settori di cerchio, ecc.). Non trattandosi di funzioni statistiche, è inapplicabile sia il diagramma a spezzata o a curva continua (che pur talvolta si trova applicato; ma è un errore non giustificabile) sia il procedimento interpolatorio; e sebbene aritmeticamente si possa calcolare una media dei termini, logicamente essa ha poca rilevanza. Trova scarse possibilità d'impiego il metodo dei numeri indici, salvo che nella forma dell'ordinato riferimento dei termini di una serie a quelli corrispondenti di un'altra serie; è invece largamente usata la traduzione dei dati in rapporti di composizione, che permette di vedere immediatamente quale frazione costituiscano i casi corrispondenti ad una certa modalità del carattere, rispetto al complesso delle osservazioni.

7. — Le serie che indicano la misura della manifestazione di un fenomeno nelle varie sezioni di un medesimo campo d'osserva-

zione, delimitate con riguardo a circostanze qualitative, possono essere rappresentate per mezzo di medie e di dati sussidiari quando siano composte di dati elaborati (rapporti, medie ecc.), mentre in generale tal forma di rappresentazione non è logicamente applicabile quando le serie siano composte di dati greggi. Se conosciamo il numero degli infortuni nel lavoro avvenuti nelle singole industrie, poco ci servirà calcolare la media di tali numeri, cioè il numero medio degli infortuni avvenuti in una industria; ma se conosciamo la frequenza degli infortuni in relazione al numero degli operai di ciascuna industria, una media aritmetica ponderata ci darà la frequenza media degli infortuni per il complesso delle industrie; e una media degli scostamenti, dalla quale potremo trarre un coefficiente di variazione, ci darà una misura della deviazione delle frequenze accertate nelle varie industrie dalla media generale. Così alle serie di dati greggi come a quelle di dati elaborati, del tipo che ora consideriamo, è applicabile il procedimento dei numeri indici, sia con riferimento costante (specialmente utile quanto si tratta di medie o di rapporti) sia con riferimento dei singoli termini ai termini corrispondenti di un'altra serie; hanno rarissime possibilità di applicazione i numeri indici a catena. Alle serie di dati greggi è vantaggiosamente applicabile anche la traduzione in rapporti di composizione, che indica prontamente qual frazione dei casi osservati od accertati cada in ciascuna sezione del campo d'osservazione. Tra i procedimenti grafici, specialmente l'istogramma trova conveniente impiego; è da escludere il diagramma a spezzata o a curva continua, talora erroneamente usato.

8. — Le serie geografiche e topografiche rientrano nel tipo cui è dedicato il precedente paragrafo, così che vale per esse quanto ivi è detto. Ma dispongono di un mezzo di rappresentazione grafica loro esclusivo: il cartogramma, che al fine della rappresentazione della singola serie è spesso più efficace di ogni altro. Sul cartogramma non ci intratteniamo, avendone già trattato a lungo nel capitolo sulle rappresentazioni grafiche.

9. — Le considerazioni precedentemente esposte si riferiscono soltanto ad alcuni principali tipi di serie statistiche. Non ci sembra necessario proseguire il ragionamento col prendere in esame altri tipi secondari, perchè, messo ormai sulla via, il lettore potrà agevolmente intendere da sè quali siano i procedimenti rappresentativi di più opportuna applicazione.

Quesiti ed esercizi: 1. — Quali sono i procedimenti più adatti per la rappresentazione delle serie cronologiche?

2. — Si prendano in esame le seguenti serie cronologiche riportate nell'ASI, fra le « Notizie statistiche retrospettive »; per ciascuna di esse si studino e si indichino le possibilità e le opportunità di applicazione dei vari procedimenti di rappresentazione e si spieghi perchè appaiono preferibili soluzioni in parte diverse nei diversi casi.

Serie da esaminare: *a*) popolazione italiana; *b*) ammontare dei depositi a risparmio; *c*) numero degli emigranti per la Francia; *d*) numero degli emigranti per il Brasile; *e*) produzione della pirite di ferro; *f*) produzione dello zolfo; *g*) corso medio del cambio su Parigi; *h*) corso medio del cambio su Londra.

3. — Si studino e si indichino le possibilità e le opportunità di applicazione dei vari procedimenti di rappresentazione alle seguenti serie cronologiche di dati mensili riportate nell'ASI: *a*) nati vivi negli ultimi quattro anni (si ricerchi anche la possibilità di rappresentazione comparativa dell'andamento delle nascite maschili e di quelle femminili); *b*) matrimoni nello stesso periodo; *c*) temperature medie registrate da diversi osservatorii nei vari mesi di un anno (si ricerchino i modi più adatti per la comparazione tra i diversi osservatorii); *d*) numero dei disoccupati nei diversi compartimenti; *e*) valore delle importazioni e delle esportazioni negli ultimi cinque anni; *f*) merci importate da reti estere ed esportate su reti estere, dai vari transiti ferroviari; *g*) contributi per assicurazioni obbligatorie riscossi dalla Cassa nazionale per le assicurazioni sociali negli ultimi cinque anni. Per tutte queste serie si cerchi di mettere in evidenza le variazioni periodiche, se ve ne sono.

4. — Quali sono i procedimenti più adatti per la rappresentazione delle distribuzioni di caratteri quantitativi?

5. — Si ricerchino e si indichino le forme meglio adatte per la rappresentazione delle seguenti distribuzioni di circostanze quantitative, riportate nell'ASI o desumibili dai dati ivi riportati: *a*) distribuzione delle scosse sismiche secondo l'intensità; *b*) distribuzione delle provincie italiane secondo la densità della popolazione per kmq.; *c*) distribuzione delle provincie italiane secondo l'incremento relativo della popolazione negli anni successivi all'ultimo censimento; *d*) distribuzione della popolazione italiana per età (si cerchino forme idonee per la comparazione tra i due sessi); *e*) distribuzione dei comuni, nei singoli compartimenti, secondo il numero degli abitanti; *f*) distribuzione per età dei morti per suicidio; *g*) distribuzione per statura degli iscritti di leva; *h*) distribuzione per valore delle cambiali protestate; *i*) distribuzione dei condannati alla reclusione secondo la durata della pena loro inflitta; *j*) distribuzione per età dei senatori; *k*) distribuzione delle provincie italiane secondo il rendimento per ettaro della coltivazione del frumento; *l*) distribuzione per età dei colpiti da infortuni nel lavoro, nelle industrie; *m*) distribuzione delle linee ferroviarie secondo il prodotto medio chilometrico; *n*) distribuzione per età delle puerpere sussidiate dalla Cassa nazionale per le assicurazioni sociali.

6. — Quali sono i procedimenti più adatti per la rappresentazione di serie di termini disposti in ordine di grandezza?

7. — Quali sono i procedimenti più adatti per la rappresentazione di serie di termini che misurano la manifestazione di un fenomeno nelle varie sezioni di un medesimo campo d'osservazione?

8. — Si ricerchino e si indichino i procedimenti più convenienti per la rappresentazione delle seguenti serie, contenute nell'ASI: *a*) classificazione dei morti secondo le cause di morte; *b*) produzione delle cave; *c*) distribuzione del valore delle importazioni per categorie di merci; *d*) distribuzione delle attività delle casse di risparmio ordinarie; *e*) distribuzione delle loro passività; *f*) suddivisione per professioni degli iscritti di leva arruolati nell'esercito.

9. — Quali sono i procedimenti più adatti per la rappresentazione di serie geografiche o topografiche?

10. — A quali serie si può applicare il procedimento di rappresentazione mediante proporzioni desunte da rapporti di composizione?

11. — Si ricerchino i procedimenti più adatti per la rappresentazione delle seguenti serie contenute nell'ASI: *a*) distribuzione per provincie delle scosse sismiche, graduate secondo l'intensità; *b*) distribuzione per compartimenti dei nati vivi, classificati secondo la legittimità o illegittimità; *c*) natalità, mortalità, eccedenza dei nati sui morti per 1000 abitanti (rappresentazione simultanea dei tre fenomeni) nelle varie provincie; *d*) distribuzione degli emigranti secondo il paese di destinazione; *e*) fanciulli soggetti all'obbligo scolastico, iscritti alle scuole elementari, esaminati, promossi (rappresentazione simultanea dei quattro ordini di dati) nei vari compartimenti; *f*) iscritti ai diversi istituti di istruzione superiore, distribuiti secondo le sedi degli istituti e i generi degli insegnamenti; *g*) distribuzione regionale dei cinematografi; *h*) distribuzione regionale delle varie forme di delinquenza (indicata dalle cifre assolute dei reati denunciati e dalle cifre proporzionali alla popolazione); *i*) distribuzione regionale della superficie coltivata a grano, del raccolto del grano, dei rendimenti per ettaro; *j*) distribuzione regionale delle centrali elettriche, della loro potenza, della loro produzione di energia; *k*) distribuzione regionale del consumo del tabacco: dati complessivi e medie per abitante; *l*) distribuzione del valore delle importazioni per paesi di provenienza; *m*) movimento (numero degli apparecchi partiti) e traffico (numero dei passeggeri partiti) dei vari aeroporti; *n*) movimento (chilometri volati) e traffico (passeggeri-chilometro) delle varie linee aeree; *o*) suddivisione regionale del getto dell'imposta sui terreni.

CAPITOLO XV.

La sintesi di più serie statistiche in una: i numeri indici composti.

Il problema — Sintesi di dati che rappresentano parti sommabili di un unico insieme — Sintesi di dati che rappresentano aspetti diversi di un medesimo fenomeno, non sommabili — Sintesi per addizione e sintesi mediante numeri indici composti: generale applicabilità del secondo procedimento — I tre problemi particolari da risolvere nell'applicazione dei numeri indici composti: la scelta del riferimento, delle serie da riassumere, del metodo di sintesi — Applicazioni: numeri indici del costo della vita, numeri indici dei prezzi in grosso — Spostamenti del riferimento.

1. — Si presentano spesso nelle indagini statistiche problemi di questo genere:

Si hanno tante serie di dati, ordinate secondo la stessa circostanza e con lo stesso criterio, che indicano lo stato o l'andamento di singoli fenomeni costituenti altrettante parti od altrettanti aspetti di un fenomeno più generale. Ciascuna di queste serie permette di formare un apprezzamento sul fenomeno particolare che essa rappresenta. Si vuole giungere, sulla base di siffatti apprezzamenti particolari, ad un apprezzamento sintetico del fenomeno generale.

2. — Quando i fenomeni particolari sono tante parti che riunite costituiscono un unico insieme, omogeneo nei riguardi dell'indagine cui si attende, i dati greggi che li rappresentano — espressi tutti nella stessa unità — sono non solo aritmeticamente ma anche logicamente, ai fini dell'indagine, sommabili; e le somme dei termini corrispondenti delle varie serie particolari formano un'unica serie generale dalla quale si può desumere un'impressione d'insieme. Conosciamo, per esempio, le quantità di ogni singola merce esportate in tanti anni successivi: per ciascuna merce siamo in grado di giudicare quali siano le tendenze dell'esportazione attraverso il tempo. Sommiamo ora le quantità delle diverse merci esportate in ciascun dato anno: otteniamo così, anno per anno, la quantità totale delle esportazioni; e mercè la serie dei dati ottenuti possiamo giudicare quali siano le tendenze dell'esportazione in generale attraverso il tempo. Ma in tal modo l'esportazione è considerata dall'aspetto della quantità fisica, chè se volessimo considerarla dall'aspetto della quantità economica non potremmo più sommare come quantità omogenee il chilogrammo di seta ed il chilogrammo di zolfo, o il chilogrammo di limoni e il chilogrammo di bottoni esportati: dovremmo invece sommare i prezzi totali delle quantità esportate di queste merci e delle altre, per ottenere il prezzo complessivo di tutte le merci esportate. In una tale sintesi economica delle esportazioni potrà rientrare anche qualche merce, come l'energia elettrica, che per la sua natura sarebbe sfuggita alla sintesi fisica in unità di peso o in unità di volume; ma naturalmente il significato delle varie sintesi è diverso ed esse servono a fini diversi.

3. — Talvolta le statistiche adottano unità di misura differenti per la rappresentazione di fenomeni particolari che possono tuttavia riguardarsi come parti di un unico insieme: questo è un ostacolo

formale, non sostanziale, alla sintesi; e spesso con adatti artifizi può essere superato. Per esempio, le nostre statistiche del commercio con l'estero non indicano il peso ma il numero dei capi di bestiame, il volume dei vini spediti in fusti o in damigiane, la stazza lorda delle navi esportate; tuttavia è chiaro che, se si ha modo di conoscere in via approssimativa il peso medio, rispettivamente, del capo di bestiame, dell'ettolitro di vino, della tonnellata di stazza lorda di naviglio, si possono facilmente ridurre i dati sulle esportazioni delle varie merci a quella omogeneità che permette di riassumerli mediante addizione.

4. — In altri casi, invece, ci troviamo di fronte a fenomeni particolari che possono bensì riguardarsi come tanti aspetti d'un fenomeno generale, ma non come tante parti di un unico insieme omogeneo, anche se sono rappresentati da dati statistici espressi nella stessa unità di misura. La produzione del frumento, la quantità di merci trasportata dalle ferrovie, il consumo della carne costituiscono le misure di tre importanti aspetti dello stato economico di un paese, che possono essere espressi tutti tre in unità di peso; ma non possiamo sommare il grano prodotto, le merci trasportate, la carne consumata, per ottenere un dato complessivo atto a servire come misuratore dello stato economico nazionale. L'aritmetica lo consente, ma la logica lo vieta, perchè nè lo stato economico d'un paese è misurabile in unità di peso, nè l'apprezzamento particolare di ciascuno degli aspetti dello stato economico or ora considerati si può far concorrere a formare l'apprezzamento sintetico proprio in proporzione alle tonnellate o ai quintali rappresentati dal dato statistico che esprime il sintomo economico.

5. — In altri casi, infine, la sintesi mediante addizione è impossibile tanto logicamente quanto aritmeticamente perchè i fenomeni particolari, pur costituendo tanti aspetti di un fenomeno generale, nè possono riguardarsi parti di un unico insieme omogeneo, nè possono esprimersi, ai fini dell'indagine, in dati reciprocamente comparabili. Il numero dei soldati, il numero degli autocarri, il numero dei cannoni sono tanti elementi della forza materiale di un esercito, ma non possiamo sommare i numeri dei soldati, degli autocarri, dei cannoni di ciascun esercito per ottenere misure comparative della potenza militare di vari Stati; come non potremmo sommare il numero degli immigranti, l'ammontare del capitale affluente dall'estero, l'area delle terre dissodate in ciascun anno in singoli paesi coloniali, per avere un'idea riassuntiva della rapidità del loro svi-

luppo economico, benchè indubbiamente di tale sviluppo i tre dati illustrino tre aspetti.

6. — Nei casi considerati ai paragrafi 2 e 3 si può eseguire la sintesi di più serie mediante addizione; nei casi dei paragrafi 4 e 5 tale procedimento non è applicabile, e se si vuol giungere ad una sintesi è necessario anzitutto trasformare le varie serie, rappresentatrici dei fenomeni particolari, in modo tale che i loro dati divengano aritmeticamente e logicamente comparabili ai fini dell'indagine. A tale intento si ricorre al procedimento dei numeri indici.

Se vogliamo formare un completo giudizio sullo stato economico di un paese, non ci basterà conoscere la produzione del frumento, l'eccedenza delle nascite sulle morti, l'incremento dei depositi a risparmio: ricorreremo anche a molti altri sintomi delle condizioni economiche. Ma poichè il problema metodologico della sintesi di più serie in una sola è il medesimo, o si tratti di tre serie o di trenta o di trecento, supponiamo per semplicità di voler desumere un'impressione d'insieme sull'andamento economico di un paese, in una successione di anni, dall'andamento di quei tre sintomi particolari. Potremo considerare quanti anni vogliamo, ma per esporre il metodo ci basterà considerarne due: uno di questi assumeremo per riferimento, e lo chiameremo anno 1; l'altro, che compareremo col primo, chiameremo anno 2. Calcolati i numeri indici, troviamo che, posto uguale a 100 il dato per l'anno 1, nell'anno 2 la produzione del frumento è espressa dal numero indice 110, l'eccedenza di nascite da 103, l'incremento del risparmio da 120. Ammesso che la variazione positiva di ciascuno dei fenomeni considerati costituisca un indizio di miglioramento delle condizioni economiche del paese, se noi vogliamo fondarci sopra un solo dei sintomi scelti, possiamo dire che le condizioni economiche hanno segnato, dall'anno 1 all'anno 2, un miglioramento del 3, o del 10, o del 20 $^{0}/_{0}$, secondo il sintomo che scegliamo. Ma se preferiamo tener conto delle indicazioni di tutti tre i sintomi (e non sarebbero tre soli, ma un numero maggiore se invece di studiare teoricamente il metodo lo applicassimo realmente), dobbiamo conchiudere che la misura più plausibile del miglioramento economico non è nè 3, nè 10, nè 20 $^{0}/_{0}$, bensì una proporzione non inferiore al 3, nè superiore al 20 $^{0}/_{0}$: insomma una media dei tre incrementi percentuali. In altri termini, dobbiamo determinare una media dei tre numeri indici particolari 103, 110 e 120, riferentisi a speciali aspetti dell'economia del paese; ce ne potremo poi servire come di un numero indice delle condizioni economiche. Se adottiamo la media aritmetica, otteniamo come numero indice sintetico 111.

Non discutiamo, per ora, se convenga meglio questa od altra media: fermiamoci piuttosto un momento ad esaminare il significato del numero indice sintetico. Che cosa vogliamo denotare quando affermiamo che le condizioni economiche del nostro ipotetico paese nell'anno 2 stanno a quelle dell'anno 1 come 111 sta a 100?

Se trovassimo, in un'altra applicazione dello stesso procedimento, che le esportazioni dell'anno 2 stanno a quelle dell'anno 1 come 111 sta a 100, l'interpretazione sarebbe facile: il peso totale delle merci esportate se siamo partiti da dati di peso, o il valore totale di esse se siamo partiti da dati di valore, è aumentato dell'11 °/₀. Ma lo stato economico di un paese non è una quantità concreta, misurabile, come il peso o il valore delle esportazioni; è un intricato complesso di fatti, sui quali l'osservatore forma tanti giudizi separati che poi fonde mentalmente in uno solo, quasi determinando la risultante di tante componenti aventi diversa intensità e diversa direzione. E mentre l'apprezzamento quantitativo delle variazioni dei singoli fenomeni economici è compiuto sulla solida base della misurazione, di modo che si sottrae all'arbitrio dell'operatore, l'apprezzamento sintetico dello stato economico del paese, sebbene compiuto sul fondamento degli apprezzamenti particolari, non conduce ad una *misura* della variazione dello stato economico, poichè non abbiamo un metro atto a misurare tale stato e le sue variazioni; conduce soltanto ad una *indicazione* numerica della direzione e dell'importanza della variazione dello stato economico: indicazione che può risultare differente secondo i criteri adottati nella scelta dei sintomi da esaminare e da riassumere, del riferimento, del procedimento di sintesi.

7. — Svincolandoci ora dall'esempio, possiamo affermare in modo generale che nelle ipotesi dei paragrafi 2 e 3, quando cioè i termini corrispondenti delle varie serie sommati ricostituiscono un insieme omogeneo che è proprio l'oggetto cui mira la sintesi, l'addizione è il modo più semplice possibile per riassumere in una sola serie generale le varie serie particolari; volendo, potremo poi tradurre questa serie generale in numeri indici.

Ma nelle ipotesi dei paragrafi 4 e 5, quando cioè le varie serie rappresentano aspetti diversi di un unico fenomeno, la cui unità però non ha una manifestazione concreta misurabile, ma è soltanto l'espressione del risultato di un processo di sintesi astratta, il modo di tradurre questa sintesi in un'unica serie di dati numerici — che non *misurano* ma *indicano* lo stato o l'andamento del fenomeno — è offerto dal procedimento dei numeri indici composti, ossia dalla costruzione di una serie di numeri indici del fenomeno generale, ciascun termine della quale è funzione dei corrispondenti termini delle serie di numeri indici particolari.

8. — Allo stesso procedimento dei numeri indici composti si può sempre ricorrere anche nelle ipotesi dei paragrafi 2 e 3: debitamente applicato esso permette di giungere direttamente, ma per una via più lunga, ai numeri indici del fenomeno generale, che noi supponevamo or ora di calcolare dai dati della serie ottenuta mediante addizione dei termini corrispondenti delle varie serie. La maggiore lunghezza del procedimento è compensata dal vantaggio, che esso ci offre, di metterci in possesso anche dei numeri indici dei fenomeni particolari oltre che di quelli del fenomeno generale: quando non importi conoscere tali indici particolari, converrà procedere per addizione. Spieghiamoci meglio con un esempio: se vogliamo soltanto conoscere l'andamento complessivo delle esportazioni, sommeremo anno per anno le quantità delle varie merci esportate; otterremo così la quantità totale delle esportazioni; dalla serie ottenuta desumeremo poi i numeri indici dell'esportazione totale. Se invece c'interessa seguire comparativamente l'andamento delle varie esportazioni, cominceremo col ridurre a numeri indici ciascuna serie che rappresenta le quantità d'una certa merce esportate nei vari anni, assumendo naturalmente come riferimento lo stesso anno per tutte le merci. Dopo di che, per ottenere il numero indice dell'esportazione totale in un dato anno, ci basterà determinare una media dei numeri indici delle esportazioni delle varie merci. Poichè queste concorrono in misura molto differente tra loro a costituire l'esportazione totale, adotteremo una media ponderata: al numero indice di ciascuna merce potremo assegnare un « peso » uguale alla quantità della merce stessa esportata nell'anno che serve di riferimento. (Nel precedente esempio abbiamo parlato di « quantità » esportata: alla parola « quantità » si possono a volta a volta sostituire le altre più precise: « peso », « volume », « valore », secondo l'applicazione che si preferisce). Traducendo in formola il procedimento ora descritto, si vede subito che i suoi risultati finali coincidono con quelli del procedimento di addizione con successivo calcolo dei numeri indici generali.

9. — Anche quando i termini corrispondenti delle varie serie par-

ticolari ricostituiscono, addizionati, un insieme omogeneo, può darsi che non sia questo l'oggetto cui mira la sintesi, o che essa non vi si arresti. In certi casi il fenomeno generale che si vuol ricostruire mediante le osservazioni dei fenomeni particolari non è quello reale nella sua complessità, ma una semplificazione di esso: così accade quando il fenomeno generale essendo nella realtà modificato, sia da circostanze che fanno variare i fenomeni particolari, sia da circostanze che fanno variare la partecipazione di ciascuno di questi alla costituzione del fenomeno generale, lo si vuol ricostruire quale sarebbe se fosse modificato soltanto dalle une o soltanto dalle altre circostanze. Gioverà anche qui ricorrere ad un esempio. È chiaro che il valore complessivo delle esportazioni non è altro che la somma dei valori delle esportazioni delle varie merci; e non è meno chiaro che il valore dell'esportazione di una merce dipende: primo, dalla quantità esportata; secondo, dal prezzo medio unitario di essa. Dall'anno 1 all'anno 2 potrebbero restare immutate le quantità esportate di tutte le singole merci, e pur variare i valori perchè sono variati i prezzi unitari; o potrebbero restare immutati i prezzi e pur variare ugualmente i valori esportati perchè sono variate le quantità. Queste sono ipotesi estreme cui ricorriamo soltanto per distinguere l'influenza dei due ordini di circostanze; ma nella realtà così i prezzi come le quantità continuamente variano. Ora se troviamo che il valore totale delle esportazioni dell'anno 2 sta a quello dell'anno 1 come 105 sta a 100, non possiamo affermare che siano aumentate del 5 % le quantità esportate, o che siano saliti del 5 % i prezzi: abbiamo di fronte una risultante e non possiamo scinderla idealmente nelle sue componenti. Per rintracciare queste componenti dobbiamo percorrere un lungo cammino: dobbiamo cioè ricercare: primo, come sarebbe variato il valore totale delle esportazioni se, variando dall'anno 1 all'anno 2 le quantità esportate, nella misura che l'osservazione ci ha mostrato, fossero rimasti immutati i prezzi unitari; secondo, come sarebbe variato il valore totale delle esportazioni se, variando dall'anno 1 all'anno 2 i prezzi unitari, nella misura che l'osservazione ci ha mostrato, fossero rimaste immutate le quantità esportate. La prima elaborazione ci indicherà la variazione che sarebbe stata determinata nel valore totale delle esportazioni dalle modificazioni avvenute nelle quantità esportate; la seconda ci indicherà la variazione che sarebbe stata determinata dalle modificazioni avvenute nei prezzi unitari. Nè l'una nè l'altra è la variazione reale, poichè questa è determinata da simultanee modificazioni di quantità e di prezzi; ma l'una ci rivela una prima componente, l'altra una seconda componente di quella risultante che già conosciamo e che è espressa nei dati sull'esportazione totale ottenuti per semplice addizione dei dati originari sulle esportazioni delle varie merci. Naturalmente potremo tradurre le nostre serie in numeri indici. Ponendo in rapporto le esportazioni italiane del 1929 con quelle del 1925, troviamo che il valore delle une corrisponde all'81,5 $^0/_0$ di quello delle altre. La diminuzione del 18,5 $^0/_0$ nel valore totale è risultante di variazioni di quantità e di variazioni di prezzi: l'indice delle quantità, 113,8, ci dice che se i prezzi di tutte le merci esportate fossero rimasti immutati dal 1925 al 1929 il valore complessivo delle esportazioni sarebbe aumentato del 13,8 $^0/_0$, per effetto delle variazioni di quantità; l'indice dei prezzi, 72,2 $^0/_0$, ci dice che se le quantità di tutte le merci esportate fossero rimaste immutate dal 1925 al 1929 il valore complessivo delle esportazioni sarebbe diminuito del 27,8 $^0/_0$, per effetto delle variazioni di prezzi. La diminuzione effettiva del 18,5 $^0/_0$ è dunque la risultante di una diminuzione relativamente più forte (del 27,8 $^0/_0$) dei prezzi e di un aumento (del 13,8 $^0/_0$) delle quantità.

10. — Nel precedente esempio abbiamo supposto di surrogare anzitutto ai dati delle serie originarie (valori effettivi dell'esportazione delle singole merci) altri dati calcolati in corrispondenza a determinate ipotesi (invariabilità delle quantità esportate, oppure dei prezzi unitari); di addizionare poi anno per anno i dati surrogati; di tradurre infine in numeri indici la serie delle somme ottenute. Ma anche in questo caso è possibile, e talvolta opportuno, applicare, invece del procedimento di addizione preceduto dalle elaborazioni sopra descritte, il procedimento dei numeri indici composti. Per giungere al numero indice degli effetti delle variazioni di prezzi avremmo potuto calcolare una media aritmetica ponderata dei numeri indici dei prezzi delle diverse merci, per ciascun dato anno: poichè le varie merci partecipano in varia proporzione alla esportazione totale, avremmo attribuito al numero indice del prezzo di ciascuna merce un peso uguale al valore complessivo dell'esportazione della merce stessa nell'anno assunto a riferimento. E per giungere al numero indice degli effetti delle variazioni di quantità avremmo potuto calcolare una media aritmetica ponderata dei numeri indici delle quantità delle diverse merci, per ciascun dato anno, attribuendo i pesi con lo stesso criterio detto dianzi. Traducendo in formole i due procedimenti, si vede che essi conducono allo

stesso risultato del metodo di addizione, preceduto — s'intende — dalle occorrenti elaborazioni.

11. — Quanto abbiamo esposto nei paragrafi 9 e 10, riferendoci al caso particolare delle esportazioni considerate in valore, si può facilmente estendere a qualsiasi caso in cui ogni termine delle serie di dati originari sia il prodotto di due fattori indipendenti che corrispondono a due circostanze di natura differente, delle quali può essere utile considerare separatamente gli effetti.

Suppongasi, per esempio, di esaminare m corridori che sono in allenamento per n giorni. Ogni corridore percorre in ciascun giorno uno spazio s che egli fissa a suo arbitrio; da giorno a giorno tale spazio varia, col variare del tempo t che il corridore dedica all'allenamento e della velocità media v che egli mantiene durante codesto tempo. I dati statistici sugli spazi percorsi dai singoli corridori nei singoli giorni costituiscono m serie di n dati ciascuna. Sommando gli m dati riferentisi al k^{mo} giorno, sapremo quanto spazio hanno percorso in complesso i corridori in quel giorno: troveremo, per esempio, che hanno percorso km. 454,6 in confronto a 354 percorsi il primo giorno dell'allenamento; e se assumiamo il primo giorno come riferimento avremo 128 come numero indice dello spazio percorso nel k^{mo} giorno. Ma il progresso del 28 °/₀ rispecchia un progresso delle velocità tenute dai nostri corridori, o deriva da maggior tempo che essi abbiano dedicato all'esercizio? Per rispondere a tale quesito possiamo anzitutto moltiplicare i tempi effettivi di allenamento dei singoli corridori nel k^{mo} giorno per le corrispondenti velocità del 1° giorno; sommando i prodotti così ottenuti, abbiamo lo spazio totale (per esempio, km. 420,5) che i corridori avrebbero percorso nel k^{mo} giorno se dal 1° al k^{mo} giorno fosse variato il tempo dedicato all'allenamento ma non fosse variata la velocità. E poi, moltiplicando le velocità effettive dei singoli corridori nel k^{mo} giorno per i corrispondenti tempi di allenamento del 1° giorno, e sommando i prodotti così ottenuti, abbiamo lo spazio totale (per esempio, km. 383) che i corridori avrebbero percorso nel k^{mo} giorno se dal 1° al k^{mo} giorno fosse variata la velocità ma non fosse variato il tempo dedicato all'allenamento. Dai tre dati totali di cui ora disponiamo: quello corrispondente alla realtà, cioè alla variazione dei tempi e delle velocità, quello corrispondente all'ipotesi dei tempi variati e delle velocità invariate, quello corrispondente all'ipotesi delle velocità variate e dei tempi invariati, possiamo ora ricavare tre numeri indici, cioè rispettivamente: il numero indice 128 dello spazio percorso; il numero indice 119 degli effetti, sullo spazio percorso, delle variazioni avvenute nei tempi di allenamento; il numero indice 108 degli effetti, sullo spazio percorso, delle variazioni avvenute nelle velocità dei corridori.

12. — Dopo aver posto, al principio del presente capitolo, il problema generale che viene risoluto mercè il metodo dei numeri indici composti; dopo avere distinto, nei precedenti paragrafi, i vari aspetti nei quali quel problema può presentarsi, veniamo ora ad esaminare i tre problemi particolari, ad un tempo logici e tecnici, che devono risolversi in ogni applicazione del suddetto metodo.

Primo problema: la scelta del riferimento («base») nel calcolo dei numeri indici. Naturalmente per tutte le serie che si vogliono poi riassumere in una sola è indispensabile adottare il medesimo riferimento: e questo dev'essere suggerito dal fine stesso della indagine. Se abbiamo i prezzi medi annuali di tante diverse merci dal 1913 al 1930 e vogliamo riassumerli in un'unica serie atta a mostrarci le tendenze dominanti dell'andamento dei prezzi delle merci, dovremo tradurre i prezzi delle varie merci in numeri indici adottando per tutte le merci il riferimento ad uno stesso anno, al 1913 per esempio; se operassimo diversamente, non potremmo poi correttamente riassumere le varie serie in una. Il riferimento può essere in certi casi precisamente determinato dallo stesso scopo della indagine: se si vuol vedere come abbiano variato i prezzi delle merci nei periodi bellico e postbellico in confronto all'anteguerra, è ovvia l'opportunità del riferimento al 1913, ultimo anno anteriore allo scoppio del conflitto europeo. Ma se il 1913, invece che un anno nè di eccezionale espansione nè di eccezionale contrazione dell'attività economica, fosse stato un anno di carestia o un anno di sovraproduzione, adotteremmo un diverso riferimento: il 1912, per esempio, o la media di più anni anteriori alla guerra: 1909-13, per esempio. Il problema della scelta del riferimento è più difficile a risolvere nel caso dei numeri indici composti che in quello dei numeri indici semplici, perchè il riferimento che appare opportuno per alcune delle serie che si vogliono riassumere può non essere adatto per altre: così il riferimento al 1913 può essere conveniente per il prezzo di una derrata agricola che in quell'anno sia stata prodotta in quantità normale, non conveniente per il prezzo di un'altra il cui raccolto sia stato eccezionalmente scarso o eccezionalmente abbondante. Molte volte nella pratica riesce impossibile trovare un riferimento ugualmente soddisfacente per tutte le serie da riassumere, e quindi si finisce coll'adottare quel riferimento che appare opportuno per il maggior numero di serie, tenuto sempre presente il problema generale che si vuol risolvere. Nell'esempio di poc'anzi, se avessimo voluto studiare gli effetti della politica economica fascista sull'andamento dei prezzi, avremmo dovuto adottare come riferimento il 1922, ultimo anno anteriore all'attuazione di tale politica; se avessimo voluto studiare gli effetti della guerra sottomarina inasprita, o della cessazione dei prestiti esteri, o della stabilizzazione della lira, avremmo dovuto a volta a volta adottare riferimenti diversi.

13. — Secondo problema: la scelta delle serie da riassumere. Quando si tratta di ricostituire mediante le serie di dati su fenomeni particolari una serie di dati sopra un fenomeno generale che si può riguardare come la somma dei primi, la soluzione teorica è semplice: si baderà a prendere tutte le parti per poter ricostruire interamente il tutto. Prenderemo i dati sull'esportazione di tutte le singole merci, nessuna eccettuata, per ricostruire l'esportazione totale. La soluzione pratica può essere differente: se le parti sono molto numerose e alcune di esse costituiscono grosse frazioni del tutto, altre ne costituiscono frazioni molto piccole, potremo spesso trascurare queste ultime ed ottenere tuttavia una soddisfacente visione approssimativa delle variazioni del tutto. Se constatiamo che, di 2000 merci comprese nella statistica delle esportazioni, 400 costituiscono da sole il 90 % del valore complessivo, mentre le rimanenti 1.600 ne costituiscono appena il 10 %, per seguire l'andamento dell'esportazione potremo accontentarci, in via di prima approssimazione, di seguire le esportazioni di quelle 400 merci principali, desumendo poi dai dati riferentisi ad esse numeri indici del valore totale effettivo dell'esportazione (cioè numeri indici degli effetti delle variazioni di quantità e delle variazioni di prezzi); del valore totale nell'ipotesi di prezzi immutati (cioè numeri indici degli effetti delle variazioni di quantità); del valore totale nell'ipotesi di quantità immutate (cioè numeri indici degli effetti delle variazioni di prezzi). In questo caso tutti i dati parziali erano noti: ne trascuriamo una parte perchè la migliore approssimazione raggiungibile col tenerne conto non ci sembra compensi la maggiore fatica a tal uopo occorrente. Ma in altri casi ci si deve accontentare, per necessità, di considerare soltanto alcune delle serie parziali che addizionate termine a termine danno una serie generale, perchè le altre non sono note: l'espediente è lecito se le parti note costituiscono una frazione pre-

ponderante del tutto. Così le statistiche italiane permettono di seguire con buona approssimazione l'andamento della produzione agraria, sebbene non comprendano che un numero relativamente piccolo di prodotti, perchè questi da soli costituiscono una frazione preponderante, sia per valore sia per quantità, della produzione complessiva. Sarebbe invece erroneo il criterio di chi volesse seguire l'andamento del risparmio nazionale italiano attraverso i dati sui depositi nelle casse di risparmio, perchè la variazione della consistenza di tali depositi in un certo intervallo di tempo rappresenta soltanto una frazione non grande e non costante del risparmio nazionale che si è formato nell'intervallo stesso. Potremo assumere come indici dell'andamento di un fenomeno totale quelli desunti da un fenomeno parziale che non costituisca una frazione preponderante del primo soltanto quando abbiamo ragione di ritenere che il fenomeno parziale varii press'a poco proporzionalmente a quello totale.

Il problema della scelta delle serie da riassumere si pone diversamente quando le varie serie non corrispondono a parti sommabili di un sol tutto, ma a diversi aspetti non sommabili di un unico fenomeno. Quest'unico fenomeno non si può più ricostruire, diremo così, meccanicamente mediante ricorso alle sue varie parti: dev'essere ricostruito mediante tutto un processo mentale diretto a stabilire quali sono gli aspetti particolari del fenomeno generale, quali relazioni intercedono fra essi, qual è la loro importanza comparativa, quali di essi ed in qual modo convenga considerare a chi vuol formarsi un giudizio sintetico sul fenomeno generale. Quando vogliamo giudicare l'andamento complessivo delle esportazioni, il più elementare buon senso c'insegna che dobbiamo sommare le esportazioni delle varie merci per ottenere dati sull'esportazione totale, atti a servire come basi di giudizio: una macchina calcolatrice potrà eseguire l'operazione. Ma quando vogliamo giudicare lo stato economico di un paese dobbiamo chiedere all'economista e all'uomo d'affari: quali siano i fatti da rilevare, quali tra questi siano essenziali e quali accessori, se alcuni non siano che il riflesso di altri, come dai dati particolari sia possibile risalire ad una sintesi generale. Prima di mettere in moto la macchina calcolatrice, dobbiamo scegliere i sintomi da considerare, attribuire a ciascuno di essi un dato peso, stabilire fra le diverse possibili vie di sintesi quale sia la migliore, al nostro fine.

È evidente che non si può fissare alcun criterio generale di carattere sostanziale per la scelta delle serie, corrispondenti a singoli aspetti d'un fenomeno, che vanno riassunte da chi voglia ottenere un'impressione d'insieme dello stato o dell'andamento del fenomeno stesso. Il perito di ciascuna materia può stabilire, caso per caso, le norme più adatte, sulla scorta della teoria e dell'esperienza; e nell'applicazione queste norme possono essere provate e riprovate, corrette e perfezionate, in modo da assurgere talvolta a sistema. Così il medico delle assicurazioni ha davanti a sè un questionario che gli indica quali siano i sintomi essenziali da esaminare per accertare le condizioni di salute dell'assicurando. Innumerevoli altri sintomi potrebbero essere esaminati, ma l'esperienza ha mostrato che normalmente la laboriosità del lungo esame non sarebbe compensata in misura adeguata da un più illuminato giudizio sullo stato fisico del visitato. Così pure il compilatore di termometri o barometri economici ha imparato, dall'esperienza altrui e dalla propria, quali siano i sintomi più adatti dello stato di salute economica del paese: anch'egli distingue sintomi principali, come il consumo del pane, e sintomi secondari, come il consumo del prezzemolo; o in altro campo rispettivamente la produzione delle automobili e quella dei mandolini; o in altro campo ancora rispettivamente il traffico d'una grande linea ferroviaria e quello d'una minuscola tramvia; e incontra sintomi facili ad accertare, come il peso delle merci scaricate nei porti, ed altri difficili od impossibili ad accertare, come la somma dei redditi privati; e osserva sintomi rilevanti piuttosto come espressioni indirette che come espressioni dirette dello stato di salute: tale il consumo dei tabacchi; e preferisce sintomi che presentano uniformità di comportamento nei vari paesi e nelle varie epoche, ad altri che hanno andamento così capriccioso che sembrano sottrarsi ad ogni regola. Soltanto la sicura conoscenza del fenomeno che forma oggetto d'indagine può illuminare la scelta degli aspetti da considerare per la sintesi: nessuna ricetta tecnica può mettere l'incompetente in grado di compiere tale scelta. Si può stabilire soltanto qualche criterio generale di carattere formale: se alcuni aspetti hanno rilevanza preponderante in confronto a tutti gli altri, potrà bastare la conoscenza dei primi per ottenere una visione sintetica sufficientemente approssimata, con grande risparmio di lavoro; se alcuni aspetti si riflettono in certi altri, potrà supplire la conoscenza di questi all'ignoranza di quelli; se sopra alcuni aspetti si hanno dati attendibili e sopra altri poco attendibili, converrà dare la preferenza ai primi, escludendo i secondi o tenendone conto solo accessoriamente; se alcuni aspetti si mantengono paragonabili attraverso il tempo ed attraverso lo spazio, si guarderanno piuttosto questi che altri i quali si sottraggano alla possibilità di corretta comparazione; e se gli aspetti del fenomeno sono molti non si porrà la maggior cura a considerarne il massimo numero possibile, bensì a considerare quelli che, coordinati, ci presentano una più soddisfacente visione delle grandi linee del fenomeno. Così il pittore nell'effigiare un volto cerca la rassomiglianza nel coglierne i tratti fondamentali e non nel riprodurne con precisione i più minuscoli particolari.

14. — Terzo problema: la scelta del metodo di sintesi. Fissato il riferimento, stabilito quali siano le serie da riassumere e ridotte queste in numeri indici, rimane da scegliere la via per la sintesi. Come risulta dagli esempi già esposti, il problema si riduce a quello della scelta di una media: poichè i numeri indici che reciprocamente si corrispondono nelle varie serie segnano lo stato o l'andamento di altrettante parti o di altrettanti aspetti del fenomeno, un'indicazione numerica dello stato o dell'andamento di questo non può essere inferiore al minimo nè superiore al massimo di codesti numeri indici. Ma il problema della scelta di una media è tutt'altro che facile a risolvere, specialmente quando i numeri indici che reciprocamente si corrispondono nelle varie serie sono molto differenti tra loro, perchè in tal caso secondo che si adotta l'una media o l'altra, una media semplice od una media ponderata, si possono avere profondi divari nei risultati.

Bisogna distinguere il caso nel quale i fenomeni particolari sono parti sommabili che addizionate ricostituiscono il fenomeno generale, da quello in cui sono aspetti particolari non sommabili dello stesso fenomeno. Nel primo caso non sorgono dubbi: attraverso il procedimento dei numeri indici composti si vuol giungere alla stessa serie di numeri indici cui si sarebbe giunti addizionando termine a termine le serie originarie e calcolando poi i numeri indici dalla serie delle somme. E perciò si deve computare una media aritmetica ponderata dei numeri indici reciprocamente corrispondenti, attribuendo a ciascuno di essi come peso il denominatore del rapporto indice dal quale il numero indice è tratto. Così abbiamo operato nell'esempio del paragrafo 8.

Nel secondo caso, quando cioè i fenomeni particolari rappresentano altrettanti aspetti di un unico fenomeno generale, ma non parti sommabili di esso, il problema della scelta di una media diventa complicato e ammette non una sola soluzione corretta, come nel primo caso, ma più soluzioni: in generale, anzi, infinite soluzioni. Quale peso attribuire nella formazione di un indice sintetico del progresso economico dell'Argentina rispettivamente al numero indice delle immigrazioni di uomini, delle importazioni di capitali, dei dissodamenti di nuove terre? Quale peso attribuire nella formazione di un indice sintetico della forza materiale di ciascuno dei vari eserciti europei, rispettivamente, al numero indice della forza in truppa, al numero indice della quantità di autocarri, al numero indice della potenza dei cannoni? Problemi di questo genere sono difficili non perchè sia impossibile risolverli, ma perchè, come avvertivamo dianzi, ammettono moltissime soluzioni plausibili, tra le quali un perito della materia sceglierà come ottima l'una, un altro perito una seconda, un altro ancora una terza; e se per avventura i periti si trovassero d'accordo nella scelta d'una stessa soluzione potrebbero bensì dichiarare che essa è l'ottima fra tutte, ma non potrebbero dimostrarlo, essendo fondato il loro giudizio sopra l'apprezzamento comparativo di fenomeni che in realtà non ammettono alcuna misura comune.

Anche qui, come nel problema della scelta delle serie da riassumere, è d'importanza decisiva che colui il quale stabilisce i pesi da attribuire ai vari indici particolari per la formazione dell'indice generale conosca a fondo il fenomeno che in quest'indice si vuole esprimere. Ma è necessario inoltre che egli conosca le proprietà delle diverse medie, per poter valutare le conseguenze che derivano dalla scelta dell'una o dell'altra di esse.

In pratica, la grande difficoltà di determinare con criterio obiettivo i pesi da attribuire ai numeri indici particolari nella formazione del numero indice generale fa spesso rinunziare alla ponderazione e calcolare medie semplici. In molti casi si applica la media aritmetica a preferenza delle altre: trovano però applicazione anche la media geometrica, la mediana, il valore più frequente; di rado sono usate altre medie. La semplicità del calcolo, ma più ancora l'abitudine che si tramuta in inerzia, ha fatto preferire talvolta la media aritmetica in casi nei quali sarebbero state più adatte altre medie. Quando si applicano pesi si ricorre in generale alla media aritmetica ponderata; qualche volta alla geometrica. Non bisogna certo trascurare le esigenze di carattere pratico; non si deve però subordinare ad esse esclusivamente la scelta di una media, con assoluto dispregio delle esigenze di carattere teorico.

Piuttosto che altre considerazioni di carattere generale, crediamo che possa chiarire l'applicazione del procedimento dei numeri indici composti l'esposizione di due esempi particolarmente importanti.

15. — Son oggi spesso citati i « numeri indici del costo della vita ». Questi rientrano nel primo tipo: il « costo della vita » per una famiglia in un dato intervallo di tempo è, infatti, semplicemente la somma di tutte le spese occorse in quell'intervallo per il mantenimento — nel significato più ampio della parola — della famiglia stessa.

Se noi consideriamo una determinata famiglia, che registri scrupolosamente le sue spese, potremo avere settimana per settimana, o mese per mese, dati sulle singole spese compiute e sui beni o servigi acquistati. Sapremo, per esempio, che in un dato mese la famiglia ha speso tanto per acquistare un dato numero di chilogrammi di pane, tanto per acquistare un dato numero di dozzine di uova, tanto per acquistare un dato numero di metri di tela, tanto per acquistare un dato numero di kilowatt-ora di energia elettrica, tanto per pagare un dato numero di visite mediche, tanto per pagare un dato numero di corse tramviarie, tanto per pagare l'affitto mensile dell'abitazione, ecc. Se vogliamo seguire l'andamento di ciascuna spesa attraverso il tempo, calcoleremo numeri indici delle singole spese; se vogliamo seguire l'andamento della spesa totale, sommeremo mese per mese le singole spese e poi calcoleremo i numeri indici. Ci riferiremo, nel calcolo, a quell'epoca che per una fondata ragione ci sembri più adatta come punto di partenza: o al primo mese per il quale abbiamo dati, o ad un mese in cui riteniamo che la spesa sia stata normale — non alterata cioè da circostanze eccezionali —, o ad un mese immediatamente anteriore all'inizio di un'epoca di turbamenti economici derivanti da circostanze esteriori alla famiglia (guerra, ecc.). I numeri indici della spesa totale così determinati rispecchieranno ad un tempo gli effetti delle variazioni avvenute nei prezzi e di quelle avvenute nei consumi.

La spesa mensile occorsa per ciascun bene o servigio, divisa per la quantità acquistata, ci dà il prezzo medio unitario. Se vogliamo indagare l'influenza delle variazioni dei prezzi sulla spesa familiare potremo ora calcolare come varierebbe da mese a mese la spesa totale se le quantità acquistate dei vari beni o servigi rimanessero sempre le stesse del mese assunto come riferimento, variando invece i prezzi unitari conformemente alla realtà. Dai dati così ottenuti sulla spesa familiare « a consumi immutati e a prezzi mutati » potremo poi facilmente desumere numeri indici. Se vogliamo invece indagare l'influenza delle variazioni dei consumi sulla spesa familiare potremo calcolare come varierebbe da mese a mese la spesa totale se i prezzi unitari dei vari beni o servigi rimanessero sempre gli stessi del mese assunto come riferimento, variando invece le quantità acquistate, conformemente alla realtà. Dai dati così ottenuti sulla spesa familiare « a prezzi immutati e a consumi mutati » potremo calcolare numeri indici. Le due serie di numeri indici, poste accanto a quella ottenuta prima « a prezzi mutati e a consumi mutati », ci mostreranno l'influenza rispettivamente delle variazioni dei prezzi e delle variazioni dei consumi sulla spesa familiare.

Se abbiamo calcolato numeri indici dei prezzi unitari e numeri indici delle quantità acquistate dei singoli beni o servigi, potremo servirci di questi per determinare i numeri indici composti. E precisamente: l'indice generale « a consumi immutati e a prezzi mutati » sarà uguale alla media aritmetica ponderata dei numeri indici dei prezzi unitari dei diversi beni o servigi, computata coll'assegnare a ciascuno di essi un peso uguale alla spesa occorsa nel mese assunto come riferimento per l'acquisto di quel bene o servigio; l'indice generale « a prezzi immutati e a consumi mutati » sarà uguale alla media aritmetica ponderata dei numeri indici delle quantità acquistate dei diversi beni o servigi, computata con lo stesso criterio di ponderazione; e l'indice generale « a prezzi mutati e a consumi mutati » sarà uguale alla media aritmetica ponderata dei prodotti dei numeri indici dei prezzi unitari dei singoli beni o servigi per i numeri indici delle quantità degli stessi beni, computata col solito criterio di ponderazione. Traducendo il procedimento in formole, appare immediatamente l'identità del risultato con quello ottenuto senza il passaggio per i numeri indici particolari.

Aggiungiamo, ora, che i numeri indici del costo della vita di solito non servono per seguire l'andamento delle spese di una determinata famiglia (argomento d'interesse privato) ma per seguire l'andamento delle spese familiari che può considerarsi normale di una data classe sociale. Perciò si considera un insieme più o meno numeroso di famiglie di tale classe e si costruisce la immagine astratta della famiglia media dell'operaio urbano milanese, o dell'impiegato statale romano, o del contadino cremonese. Poichè anche la composizione media di un insieme di famiglie si modifica nel tempo, quando le indagini si protraggono per un lungo periodo si costruisce addirittura una famiglia ipotetica, la cui costituzione si suppone costante: si attribuiscono per esempio al padre 40 anni, alla madre 35, a tre figli rispettivamente 15, 10, 5 anni. Fin che dura l'indagine, questa gente immaginaria sopravvive immortale perchè per essa il tempo non passa: nè bare, nè imenei, nè culle vengono a modificare la composizione della famiglia. I consumi di questa ipotetica famiglia son quelli che l'osservazione insegna essere i consumi d'una famiglia così composta in ciascuna epoca d'osservazione; i prezzi son quelli che segna il mercato. Si può supporre, se si vuole, che i consumi restino immutati; ovvero che restino immutati i prezzi; e così si possono calcolare le spese totali anche in queste due ipotesi. Tutti i calcoli che dianzi supponevamo di eseguire per una famiglia reale si possono eseguire per questa famiglia ipotetica, col vantaggio che essa non si estingue nè si modifica mai.

Se indichiamo con p_1 e p_2 i prezzi unitari di un bene o servigio rispettivamente nel periodo 1 assunto a riferimento e in un altro dato periodo 2, e con q_1 e q_2 le quantità acquistate nei due periodi, otterremo i tre numeri indici del costo della vita assumendo come denominatore per tutti tre la somma dei prodotti del tipo $p_1\ q_1$ riferentisi a tutti i beni o servigi che entrano nel consumo familiare, e per numeratore: la somma dei prodotti del tipo $p_2\ q_2$ per il numero indice a prezzi mutati e a consumi mutati; la somma dei prodotti del tipo $p_2\ q_1$ per il numero indice a prezzi mutati e a consumi immutati; la somma dei prodotti $p_1\ q_2$ per il numero indice a prezzi immutati e a consumi mutati.

Queste stesse formole sono adatte ad innumerevoli applicazioni per la soluzione di problemi formalmente analoghi benchè sostanzialmente diversi.

16. — È oggetto di costante attenzione nel mondo economico anche l'andamento, nel tempo, dei « numeri indici dei prezzi delle merci in grosso », i quali, come dice la denominazione, riguardano i prezzi che si formano nello scambio di grosse partite di merci tra il produttore e l'intermediario, o tra un intermediario ed un altro, o tra il produttore e l'intermediario da un canto e l'imprenditore di un'industria trasformatrice o consumatrice della merce dall'altro; mentre i numeri indici del costo della vita or ora esaminati riguardano invece i prezzi pagati dal consumatore individuale per piccole partite di merce destinate all'uso personale o familiare (prezzi al minuto).

Se si considerano tutti gli scambi di merci, in grosso, che avvengono in un mercato, si può riguardare il valore totale di essi come la somma dei valori degli scambi delle singole merci; e per ciascuna merce il valore degli scambi si può riguardare come il prodotto della quantità scambiata per il prezzo medio unitario. Con elaborazioni parallele a quelle del paragrafo precedente si può indi calcolare una serie di numeri indici del valore totale effettivo degli scambi di merci in grosso sul mercato in esame, una serie di numeri indici del valore degli scambi computato nell'ipotesi di costanza dei prezzi, una serie di numeri indici del valore degli scambi computato nell'ipotesi di costanza delle quantità scambiate. La prima serie mostra come il valore totale degli scambi varii per l'effetto combinato di variazioni dei prezzi e di variazioni delle quantità scambiate, la seconda come esso varii per effetto di variazioni delle quantità scambiate, la terza come esso varii per effetto di variazioni dei prezzi.

Ma ora ci porremo da un punto di vista alquanto differente. L'esame dell'andamento dei prezzi delle diverse merci, nel tempo, mostra che essi presentano tendenze generali comuni, le quali si rivelano attraverso i movimenti particolari dei prezzi di singole merci. Se traduciamo in diagramma i numeri indici dei prezzi di molte merci, considerati nella loro variazione in un dato intervallo di tempo, di solito vediamo che le varie curve, a partire dal punto comune a tutte (ordinata = 100, ascissa = data di riferimento comune), pur non seguendo precisamente lo stesso cammino, hanno andamenti molto simili nell'aspetto generale: così che alternativamente divergendo e convergendo, e ora intersecandosi ora sovrapponendosi, finiscono col delimitare nel piano una fascia entro la quale sono contenute le variazioni dei prezzi, se non proprio di tutte le merci, della maggior parte di esse. Se tracciamo una linea curva o spezzata che si mantenga costantemente nel mezzo di questa fascia, essa può da sola rappresentarci la tendenza dominante nell'andamento dei prezzi: è chiaro che a questo procedimento grafico corrisponde il procedimento numerico del calcolo della mediana dei numeri indici dei prezzi delle diverse medie in ciascun intervallo unitario di tempo. Ecco già una via di sintesi. Ma ci conviene ripigliare il cammino dal principio per vedere come si pongano e come si risolvano in questo caso i soliti problemi.

La scelta del riferimento è determinata dal fine stesso dell'indagine. Nel 1913 molti indici dei prezzi in grosso venivano calcolati con riferimento ai primi anni del secolo ventesimo: evidentemente i compilatori si proponevano di mettere in rilievo le variazioni dei prezzi che si andavano manifestando nel corso del secolo, come deviazioni dalla situazione iniziale. Oggi molti indici sono invece riferiti al 1913: si vogliono porre in evidenza le deviazioni dell'attuale livello dei prezzi, soggetto a brusche ed intense variazioni, da quel livello solo lentamente variabile che era caratteristico degli ultimi anni antecedenti alla guerra mondiale. E nei paesi che più presto sono giunti, od hanno creduto di giungere, ad un nuovo equilibrio, si comincia già ad assumere come riferimento l'anno postbellico nel quale codesto nuovo equilibrio è stato raggiunto: l'Ufficio di statistica del lavoro degli Stati Uniti ha adottato il 1926, la rivista inglese *The Economist* il 1927. In Italia si potrebbe ragionevolmente adottare il riferimento al 1928, primo anno successivo alla stabilizzazione legale della lira; ma il riferimento al 1913 offre ancora un « ubi consistam » più stabile, perchè i prezzi del 1913 differivano relativamente poco da quelli del 1912, del 1911, del 1910, ecc.: e di un livello di prezzi mantenutosi a lungo con lievi variazioni dura molto più a lungo il ricordo, con la conseguente possibilità di apprezzamenti comparativi, di quanto duri il ricordo d'un livello di prezzi molto differente dal livello degli anni più prossimi, precedenti e successivi, come quello del 1928. Invece del riferimento fisso, o accanto ad esso, è sempre possibile il riferimento variabile a catena, che permette di determinare direttamente la deviazione del livello dei prezzi di ciascun intervallo unitario di tempo da quello dell'intervallo immediatamente precedente.

La scelta delle serie di dati da riassumere è determinata da considerazioni di carattere pratico e di carattere teorico. Di carattere pratico: per certe merci gli scambi si compiono pubblicamente, nelle borse o in altre apposite sedi, e si hanno quindi prezzi sicuramente accertati; per altre gli scambi si compiono privatamente, e quindi o mancano informazioni sui prezzi oppure si hanno informazioni dai compratori o dai venditori, che possono non corrispondere al vero, anzi essere alterate per la tutela del reale o supposto tornaconto di questi o di quelli. Talvolta si possono controllare le indicazioni di un contraente con quelle dell'altro, ma non sempre si può escludere che compratore e venditore siano d'accordo nel dare un'informazione errata; e d'altronde in molti casi mancano assolutamente informazioni o se ne hanno, sospette, e manca la possibilità di controllarle. È ovvio che si preferiranno, nella scelta, le merci per le quali si possono avere dati attendibili sui prezzi, lasciando da parte le altre anche se teoricamente sarebbe desiderabile tenerne conto. Ancora: certe merci si mantengono identiche a sè medesime, fisicamente e chimicamente, nel tempo; altre variano continuamente, così che un nome costante corrisponde ad un oggetto mutevole. Una data qualità di ghisa, un certo tipo di carbone, l'acido borico, possono essere precisamente definiti: il prezzo che per essi segna il mercato è sempre il prezzo della stessa merce. Una data qualità di grano, di legname, di lana non resta rigorosamente identica a sè medesima nel tempo, tuttavia può riguardarsi approssimativamente tale per lunghi periodi. I prezzi di merci siffatte sono adatti ad entrare nella composizione di un numero indice sintetico. Ma merci come un'automobile, una locomotiva, uno scialle, mutano aspetto e composizione col progredire della tecnica produttiva e col cambiare dei gusti, così che il prezzo d'uno di questi prodotti in varie epoche non è il prezzo d'uno stesso oggetto ma di oggetti talora profondamente diversi. Merci di tal genere non convengono per la formazione di numeri indici sintetici dei prezzi: ecco perchè nella composizione di questi vediamo in generale predominare le materie prime e le derrate agrarie. Un'altra considerazione di carattere pratico conduce ad escludere quei prezzi che si formano in scambi di scarsa rilevanza, come aventi scarso valore rappresentativo. Altre considerazioni, che per brevità omettiamo, restringono sempre più il campo di scelta.

Dall'aspetto teorico, poichè vogliamo mettere in evidenza ciò che vi è di comune nei movimenti dei prezzi delle varie merci, ci converrebbe escludere quelle merci che per cause particolari si allontanino di molto dall'andamento generale: ma è chiaro che il principio, facile ad enunciare, è difficile ad applicare, a meno di voler modificare da settimana a settimana o da mese a mese l'elenco delle merci considerate per la costruzione del numero indice, il che turberebbe la comparabilità dei dati. Perciò in pratica si suole mantenere invariato nel tempo l'elenco. Ancora dall'aspetto teorico, non dovrebbe influire sulla scelta l'importanza economica della merce, cioè il valore complessivo degli scambi cui essa dà luogo, perchè il prezzo d'una merce importantissima, come il carbone, può segnare variazioni fortissime per cause particolari (per esempio, uno sciopero dei minatori), mentre quello d'una merce pochissimo importante, come il talco, può invece seguire fedelmente le tendenze generali dei prezzi. Ma in pratica, sia perchè i numeri indici dei prezzi in grosso hanno anche il fine di segnare le variazioni del potere d'acquisto della moneta rispetto alle merci, sia perchè non v'è ragione di ritenere che a lungo andare le merci economicamente più importanti siano meno atte delle altre a rivelare l'andamento dominante dei prezzi anzi vi sono buone ragioni per la tesi contraria, sia perchè sui prezzi delle merci principali si possiedono in generale dati più attendibili, i numeri indici sintetici finiscono coll'essere formati sui prezzi delle merci che danno luogo a maggior valore di scambi.

La scelta del procedimento di sintesi può essere diversa secondo il fine cui principalmente si mira. Se, come noi ci siamo proposti, si tende sopratutto a descrivere la parte comune delle variazioni dei prezzi, non v'è ragione di attribuire peso diverso ai prezzi delle diverse merci e si può calcolare una media semplice. Il valore più frequente (se appare ben definito, per il che di solito occorre avere un buon numero di merci) è forse la media più adatta al nostro intento; sarebbe adatta anche la mediana; ma in pratica talvolta vien preferita la media geometrica, e spesso la media aritmetica. Quest'ultima non sembra ammissibile. Infatti è ugualmente legittimo l'esame delle variazioni dei prezzi così attraverso i dati sui prezzi delle merci espressi in moneta, come attraverso i dati — che sono i reciproci dei primi — sui prezzi della moneta espressi nelle varie merci. Perciò bisogna adottare un procedimento tale che la media dei numeri indici dei prezzi delle merci in moneta risulti uguale al reciproco della media dei numeri indici dei prezzi della moneta in merci; altrimenti secondo che si considera l'aspetto diretto o l'aspetto inverso dei prezzi si arriva a risultati discordanti e reciprocamente incompatibili. Ora la media geometrica, la mediana, il valore più frequente hanno appunto, come sappiamo, la proprietà che il reciproco di ciascuno di essi è uguale alla corrispondente media dei reciproci; e quindi, applicati ai prezzi ed ai loro reciproci, conducono a risultati concordanti; la media aritmetica non ha tale proprietà, e perciò va scartata.

Se, invece, si mira specialmente ad accertare in quale proporzione sia variato il potere d'acquisto della moneta rispetto alle merci comprate in grosso, e quindi conviene tener conto dell'importanza economica delle varie merci, si può ricorrere ad una media ponderata. Il numero indice adatto a quest'intento sarebbe quello enumerato come terzo nel secondo comma del presente paragrafo; ma in pratica è impossibile computarlo: per molte merci, come abbiamo detto, mancano dati sui prezzi praticati e ancor più spesso sulle quantità scambiate; così che in pratica ci si limita a calcolare una media aritmetica ponderata dei numeri indici dei prezzi delle merci principali per le quali si hanno informazioni attendibili, attribuendo a ciascun numero indice un peso uguale al valore degli scambi della corrispondente merce avvenuti nel periodo che si è adottato per riferimento; e talora ci si accontenta di una media semplice, supplendo alla mancanza della ponderazione col considerare più volte (in diverse qualità, o in diversi stadi della loro trasformazione) le merci più importanti, e una volta sola le merci meno importanti.

Su questo argomento della scelta della media più adatta per la sintesi dei numeri indici dei prezzi in grosso sono stati scritti libri di discreta mole e si sono svolte vivaci discussioni. In gran parte le divergenze di vedute tra i vari studiosi rispecchiano diverse concezioni del fine per il quale si calcolano tali numeri indici sintetici; vi accenniamo soltanto per mostrare quante questioni possano sorgere in un sola applicazione particolare di quel procedimento dei numeri indici composti, che nel nostro corso dobbiamo esporre in modo generale senza poterci fermare a lungo sulle sue applicazioni.

17. — Data una serie di numeri indici composti, calcolati con un certo riferimento costante, occorre talvolta di dover spostare il riferimento (data per esempio una serie di numeri indici dei prezzi in grosso riferita al 1913, potrà occorrere di trasformarla in una serie riferita al 1923). A tale intento non si può, in generale, procedere come si procederebbe per una serie di numeri indici semplici (V. Cap. XIII, paragrafo 6); anzi è necessario ricalcolare tutte le serie di numeri indici semplici col nuovo riferimento e poi ricalcolare la serie dei numeri indici composti. Ma se i numeri indici composti sono medie geometriche dei numeri indici semplici, non è necessario questo laborioso procedimento: basta dividere i numeri indici composti calcolati col vecchio riferimento per il rapporto indice composto corrispondente al nuovo riferimento nella serie stessa (nell'esempio di dianzi, dividere i numeri indici dei prezzi riferiti al 1913 per il rapporto indice dei prezzi del 1923 riferiti al 1913); traducendo in formole il procedimento, si vede come esso conduca allo stesso risultato cui si giungerebbe per la via più lunga. Altrettanto non accade per la media aritmetica, nè per la mediana, nè per il valore più frequente; anzi di regola col procedimento più rapido si giunge a risultati differenti da quelli cui si giungerebbe col più lento procedimento del nuovo calcolo dei numeri indici semplici prima, e di quelli composti poi. In pratica il procedimento

sbrigativo viene spesso applicato anche a numeri indici composti calcolati per medie aritmetiche, e non sono sempre trascurabili gli errori che ne derivano.

Indicazioni bibliografiche. — La maggior parte delle trattazioni sui numeri indici composti hanno di mira specialmente i numeri indici dei prezzi. Vedasi FISHER I., *The making of index numbers*, Boston, Houghton Mifflin, 1922 (opera suggestiva, ma pericolosa perchè con rigorosa logica l'autore giunge a conclusioni molto discutibili, essendo partito da premesse in parte arbitrarie ed avendo trascurato premesse necessarie). — VON BORTKIEWICZ L., *Zweck und Struktur einer Preisindexzahl*, in *Nordisk Statistisk Tidskrift*, 1923-24. — GINI C., *Quelques considérations sur la construction des nombres-indices des prix*, ecc., in *Metron*, Vol. IV. — JULIN A., *Statistique des prix et méthode des index-numbers*, Paris, Rivière, 1928 (con un'abbondante bibliografia sull'argomento). Sui numeri indici delle condizioni economiche, vedasi MITCHELL C. W., *Business cycles*, New York, National Bureau of Economic Research, 1927, capitolo III, per quanto riguarda le variazioni cicliche; e MORTARA G., *Lezioni di statistica economica e demografica*, Roma, Athenaeum, 1920, parte IV, per quanto riguarda le variazioni secolari.

Quesiti ed esercizi: 1. — Qual è il problema generale alla cui soluzione mira il metodo dei numeri indici composti? Con quali diverse modalità può presentarsi tal problema? In quali casi la sintesi si può eseguire mediante semplice addizione, in quali no? Come si giunge alla sintesi nella seconda ipotesi? Come si applica il procedimento dei numeri indici composti? È applicabile soltanto a serie cronologiche?

2. — Si prendano in esame i numeri indici composti riportati nell'ASI, capitoli « Prezzi, salari e consumi; commercio con l'estero; mercato monetario e finanziario », cercando d'intendere come siano stati risoluti in ciascuna applicazione i problemi della scelta del riferimento, della scelta delle serie da riassumere, della scelta della media, della ponderazione della media. Si espongano eventuali osservazioni critiche.

3. — Si scelgano tra le serie riportate nell'ASI, « Notizie statistiche retrospettive », quelle che sembrano più adatte per la costruzione di numeri indici sintetici delle variazioni delle condizioni economiche dell'Italia negli ultimi cinquant'anni; si tracci la via per il calcolo di tale indice sintetico; si provi a calcolarlo per qualche anno. Si mettano in rilievo le differenze di apprezzamento che possono derivare dall'adozione di differenti criteri: *a*) nella scelta della base, *b*) nella scelta delle serie da riassumere, *c*) nella scelta della media da adottare, *d*) nella scelta dei pesi.

4. — Si scelgano, tra le serie riportate nei vari capitoli dell'ASI, quelle che sembrano meglio adatte per la costruzione di numeri indici sintetici comparativi delle condizioni economiche delle varie regioni a una certa data; si prepari lo schema delle elaborazioni necessarie e si provi a calcolare il numero indice per qualche regione.

5. — Si predisponga, e si esegua per qualche regione, il calcolo di numeri

indici sintetici regionali della delinquenza, elaborando i dati offerti nel capitolo « Giustizia » dell'ASI.

6. — Se è noto a varie date successive il numero degli operai occupati in ciascuna industria e il salario medio dell'operaio in ciascuna industria, quale significato avrebbero numeri indici sintetici dei salari, ottenuti col supporre immutati i numeri di operai delle varie industrie attraverso il tempo e modificati i salari? numeri indici ottenuti col supporre modificati i numeri di operai e immutati i salari? numeri indici ottenuti col supporre modificati i numeri di operai ed i salari? a quali diversi fini potrebbero servire i diversi indici?

7. — Si predisponga, e si esegua per qualche regione, il calcolo di numeri indici sintetici regionali della produzione agraria, utilizzando i dati sulla produzione e sui prezzi riportati nell'ASI.

8. — Si calcolino numeri indici sintetici, a catena, dei prezzi mensili delle derrate agrarie negli ultimi anni.

9. — Si calcolino numeri indici regionali dei depositi a risparmio (casse ordinarie e casse postali), sia partendo dai dati assoluti sia dalle medie per abitante.

10. — Si calcolino numeri indici del progresso della produzione industriale in Italia negli ultimi trent'anni utilizzando i seguenti sintomi diretti e indiretti intorno ai quali forniscono dati le « Notizie statistiche retrospettive » dell'ASI: produzione del ferro e dell'acciaio, produzione dei perfosfati, produzione dello zucchero, importazione di cotone in bioccoli o in massa, importazione di rame in pani, ecc., importazione di carbon fossile, esportazione di seta tratta greggia. Si metta poi in rapporto il numero indice sintetico della produzione industriale così ottenuto col rapporto indice della popolazione, al fine di ottenere un numero indice per abitante.

11. — Esercizio analogo al n. 10, per la produzione agraria. Sintomi: Produzioni del frumento, del granturco, del vino, dell'olio di oliva; esportazioni di canapa greggia, di agrumi.

12. — Partendo dai numeri indici dei prezzi in grosso di singole merci riportati mensilmente nel Listino del Consiglio provinciale dell'economia di Milano, si calcoli per un dato mese il numero indice sintetico dei prezzi in grosso, adottando successivamente le seguenti medie: media aritmetica, media armonica, mediana, valore più frequente, valore equidistante dagli estremi. Si confrontino tali medie tra loro e con la media geometrica già calcolata dal Consiglio: si espongano le considerazioni che tale confronto suggerisce. Si calcoli, rispetto alle diverse medie, lo scostamento medio assoluto dei numeri indici singoli dal numero indice sintetico: si interpretino i risultati del calcolo. Si dichiari, da ultimo, quale delle medie provate appaia più adatta al fine del numero indice sintetico dei prezzi in grosso, avendo ben cura di precisare quale sia codesto fine.

13. — Si traducano in formole i tre numeri indici del costo della vita ai quali si riferiscono i due ultimi commi del paragrafo 15 del testo; si chiarisca il significato di ciascuna formola e lo scopo dell'applicazione di essa.

14. — Si dimostri che si giunge allo stesso risultato nel calcolo di un numero indice sintetico del costo della vita a consumi immutati e a prezzi mutati seguendo queste due diverse vie: *a*) moltiplicando le quantità dei vari beni (merci o servigi) consumati nel periodo 1 per i prezzi del periodo 2, som-

mando i prodotti così ottenuti e dividendo la somma per la somma dei prodotti delle quantità del periodo 1 per i prezzi del periodo 1 (metodo di addizione e consecutivo calcolo di numeri indici semplici); *b*) calcolando i numeri indici dei prezzi dei singoli beni nel periodo 2 con riferimento al periodo 1 e computandone poi la media aritmetica ponderata, nella quale al numero indice del prezzo di ciascun bene sia assegnato come peso la spesa complessiva fatta per quel bene nel periodo 1, cioè il prodotto della quantità del periodo 1 per il prezzo del periodo 1 (metodo dei numeri indici composti).

15. — Si dia una dimostrazione analoga a quella richiesta nell'esercizio precedente, previe le opportune modificazioni nell'enunciazione della tesi, riferendosi al numero indice a consumi mutati e a prezzi immutati.

16. — Come il precedente esercizio: ma per il numero indice a consumi e prezzi mutati.

17. — Si trasformino in numeri indici a catena, partendo dal mese di dicembre di un dato anno, i numeri indici dei prezzi in grosso calcolati dal Consiglio provinciale dell'economia di Milano per i singoli mesi dell'anno successivo.

18. — Nel calcolare un numero indice del costo della vita a consumi immutati si possono adottare come consumi sia quelli del periodo 1 sia quelli del periodo 2: i due criteri appaiono ugualmente plausibili. Perciò qualche statistico considera come «formola ideale» una che concilii in sè i due criteri: a tal uopo si ritiene specialmente adatta la media geometrica dei risultati ottenuti con l'applicazione separata dei due criteri. Sembra accettabile questa tesi? Sebbene i due criteri separatamente applicati diano un'indicazione degli effetti della variazione dei prezzi sul costo della vita, applicati congiuntamente hanno la stessa efficacia? o non danno un'indicazione degli effetti della variazione dei prezzi e degli effetti della variazione dei consumi, tale da non poter distinguere separatamente gli uni e gli altri effetti? Per meglio scorgere l'inconveniente del metodo, può giovare di non fermarsi al periodo 2: si supponga anzi di continuare il calcolo per i periodi 3, 4, 5, ecc., per ciascuno dei quali si potrà ripetere il ragionamento fatto in principio relativamente al periodo 2.

19. — Si ricerchi e si esponga come potrebbero applicarsi le tre formole di numeri indici composti considerate nell'ultima parte del paragrafo 15 del testo allo studio dei seguenti fenomeni: *a*) variazione dei salari industriali nel tempo, denotando p il salario in ciascuna industria e q il numero degli operai occupati in ciascuna industria; *b*) variazione della produzione agraria nel tempo, denotando p il prezzo unitario di ciascun prodotto e q la quantità raccolta di ciascun prodotto. Si chiarisca bene quale sarebbe il significato di ciascuna delle tre formole nell'una e nell'altra applicazione. Che cosa significherebbe un numero indice 120? un numero indice 80?

20. — Una gara atletica consta di tre prove: corsa, salto in lungo, sollevamento di pesi. Dieci concorrenti ottengono i seguenti risultati (la lettera designa un concorrente; il primo numero successivo i minuti secondi impiegati nella prova di corsa, il secondo i centimetri saltati, il terzo i chilogrammi sollevati): *A* 57, 630, 71 ; *B* 55, 520, 64 ; *C* 58, 570, 66 ; *D* 56, 600, 80 ; *E* 56, 600, 75 ; *F* 57, 580, 69 ; *G* 59, 610, 72 ; *H* 56, 620, 59 ; *I* 58, 640, 79 ; *J* 58, 590, 77.

Si traducano in numeri indici i risultati delle singole prove (attenzione all'artifizio da adottare per la prima prova, per far sì che ai migliori risultati corrispondano numeri indici più alti, e reciprocamente!); si calcolino poi numeri indici sintetici atti a servire per la graduazione dei concorrenti secondo il merito. Si confronti la graduazione così ottenuta con quella che si avrebbe graduando invece i concorrenti secondo il posto conseguito nelle varie prove (cioè assegnando a ciascuno una somma di punti uguale alla somma dei numeri d'ordine dei posti occupati nelle varie prove e graduando poi i concorrenti in ordine inverso a quello del numero dei punti conseguiti). Perchè le due graduazioni possono differire? A quale criterio corrisponde l'una, a quale l'altra?

21. — Quali problemi si presentano nell'elaborazione di un numero indice dei prezzi in grosso? Si contrappongano tali problemi a quelli che sorgono nella preparazione di un numero indice del costo della vita e si ricerchino e si espongano le ragioni delle diverse posizioni e soluzioni di problemi.

22. — Si prendano in esame i numeri indici composti dei noli marittimi pubblicati nella rivista *The Statist*, cercando di intenderne il fine e il modo di costruzione.

23. — Come si potrebbe calcolare un numero indice dei salari industriali in Italia mercè i dati forniti dall'ASI? Come si è proceduto a tal fine nell'ASI 1930?

24. — Con quali diversi criteri si possono determinare numeri indici composti delle quotazioni di borsa dei titoli azionari? In che differiscono gli indici del Bachi da quelli del Guarneri (riferiti entrambi nell'ASI)?

25. — Si sposti la base dei numeri indici dei prezzi in grosso del Consiglio provinciale dell'economia di Milano dal 1913 al 1928.

26. — Si dimostri che, date tre merci i cui prezzi in moneta siano rispettivamente p_1', p_1'', p_1''' in un primo periodo e p_2', p_2'', p_2''' in un secondo periodo, si arriva a risultati concordanti nei calcoli dei numeri indici sintetici, per il secondo periodo, dei prezzi delle merci in moneta e dei prezzi della moneta in merci, quando si adoperi la media geometrica; a risultati discordanti quando si adoperi la media aritmetica.

27. — Se da più serie di numeri indici a catena (p. es. numeri indici dei prezzi di varie merci in tanti mesi successivi) si è desunta una serie di numeri indici composti (p. es. numeri indici dei prezzi in generale), si può poi convertire quest'ultima serie in una serie di numeri indici composti con riferimento costante (p. es. con riferimento al primo mese), senza ricalcolare i numeri indici semplici? Si risponda al quesito considerando separatamente le ipotesi che i numeri indici composti siano stati calcolati: *a*) per medie aritmetiche, *b*) per medie geometriche, *c*) per mediane.

CAPITOLO XVI.

Il coordinamento di più serie per la descrizione di fenomeni di movimento: le tavole di eliminazione.

Fenomeni di eliminazione; tavola di eliminazione — Descrizione di una particolare tavola di eliminazione: la tavola di sopravvivenza; serie che la costituiscono — Generalizzazione: distinzioni tra vari tipi di tavole di eliminazione, secondo la natura del complesso osservato e il modo dell'eliminazione; funzioni caratteristiche della tavola di eliminazione: la funzione di permanenza, la funzione di eliminazione, la velocità di eliminazione, il saggio di eliminazione, il saggio di permanenza, la permanenza mediana, la permanenza media aritmetica — Relazioni tra queste varie funzioni — Ricostruzione dello svolgimento del fenomeno di eliminazione nelle sue fasi successive mediante l'osservazione simultanea delle fasi stesse: tavole di sopravvivenza per individui viventi simultaneamente e tavole di sopravvivenza per individui nati simultaneamente — Casi più complicati di fenomeni di eliminazione: esistenza simultanea di più fattori di eliminazione e di fattori di ingresso — Calcolo del saggio di eliminazione per un singolo fattore, in queste ipotesi — Costruzione di tavole di eliminazione, supposto agente un solo fattore di eliminazione; costruzione di tavole di ingresso e di eliminazione — Riferimento, alla tavola di eliminazione, di fenomeni che si svolgono nel complesso osservato.

1. — Numerosi fenomeni, osservati nel loro svolgimento attraverso il tempo, ci si presentano nell'aspetto del graduale esaurimento di un complesso inizialmente osservato, la consistenza del quale si vede progressivamente diminuire col trascorrere del tempo. Un serbatoio di nafta che si vuota per il graduale deflusso del suo contenuto, una generazione d'individui che si va gradualmente estinguendo per la morte di suoi componenti, una schiera di marciatori in gara che si va di mano in mano assottigliando col fermarsi di quelli divenuti incapaci di proceder oltre: sono esempi di fenomeni del genere accennato.

Altri fenomeni, — o gli stessi, osservati da un diverso punto di vista —, ci si presentano ancora nell'aspetto del graduale esaurimento di un complesso inizialmente osservato: ma non più esaurimento in funzione del tempo, bensì esaurimento in funzione di un'altra circostanza quantitativa, di un'altra « variabile ». Il progressivo assottigliarsi della schiera di marciatori supposta dianzi può essere studiato in relazione allo spazio percorso, invece che in relazione al tempo trascorso, dall'inizio della marcia. Di un gruppo di donne le quali contraggono matrimonio si può studiare il graduale esaurimento in relazione al numero dei figli procreati: alcune si fermano al punto stesso di partenza, poichè non mettono al mondo figli; altre si fermano alla tappa *uno* cioè al primo figlio, altre alla tappa *due* cioè al secondo; e così via procedendo di tappa in tappa vediamo assottigliarsi la schiera.

Fenomeni come questi, che si presentano nell'aspetto del graduale esaurimento di un complesso osservato, in funzione di una variabile, possono dirsi in generale *fenomeni di eliminazione.* Una sola serie di dati statistici (o una sola curva, o la formola rappresentatrice della curva stessa) è sufficiente a descriverne l'andamento: quella che indica la consistenza del complesso osservato, in corrispondenza ai successivi valori della variabile in funzione della quale esso viene osservato. Ma normalmente a questa serie se ne pongono accanto altre, per rendere più efficace la descrizione; e raccogliendole tutte in una sola tabella — com'è facile, poichè sono tutte ordinate secondo la medesima circostanza — si ottiene una *tavola di eliminazione.*

2. — L'esame di una particolare tavola di eliminazione ci consentirà di chiarire i fini di questa forma di descrizione in un caso speciale: dopo di che ci sarà più agevole esporre tali fini in modo generale senza essere oscuri o prolissi.

Ci varremo della tavola di sopravvivenza, che è una delle più comuni e più utili tavole di eliminazione. Nell'ASI, 1929, pagine 41-44, sono riportate tavole di sopravvivenza per l'Italia; nel nostro discorso ci riferiremo per qualche esempio numerico alla prima di esse: quella relativa alla popolazione maschile (pag. 42).

In una prima colonna troviamo indicate le età, espresse in anni interi: (0 l'istante della nascita), 1 (il primo compleanno), 2 (il secondo compleanno), ecc. In funzione appunto dell'età sono ordinate le varie serie di dati, che nel loro insieme costituiscono la tavola di sopravvivenza.

a) La serie fondamentale, dalla quale possono facilmente desumersi, come vedremo, tutte le altre, è quella contenuta nella colonna 3 della tavola: la serie dei sopravviventi. I numeri che la costituiscono indicano rispettivamente i numeri dei sopravviventi all'età 0 (cioè dei nati vivi), al primo compleanno, al secondo, ecc. Questi numeri non sono che alcuni tra gli infiniti valori che assume la funzione « sopravviventi all'età x » col variare di x fra 0 e ω, limite superiore della vita umana: cioè alcuni tra gli infiniti punti che costituiscono la curva di sopravvivenza in tale intervallo. Se conosciamo il numero dei sopravviventi all'inizio di ciascun mese di età (invece che di ciascun anno di età) potremo tracciare con maggior approssimazione la curva di sopravvivenza; se conosciamo il numero dei sopravviventi all'inizio di ciascun giorno di età potremo tracciarla con approssimazione anche più grande.

b) Una seconda serie, utile per la descrizione del progressivo estinguersi d'una generazione, si può costruire immediatamente mercè la prima: è quella formata dai numeri dei « morti fino all'età x ». La differenza tra il numero iniziale dei sopravviventi (il numero, cioè, dei nati vivi costituenti la generazione che si considera) e il numero dei sopravviventi all'età x indica il numero dei morti fino all'età x. I valori di questa funzione non sono esplicitamente indicati nella tavola dell'ASI, perchè partendo essa da un numero tondo di nati (100.000) riesce facile calcolare i valori stessi da quelli dei sopravviventi. Per esempio, sapendo che a 65 anni sopravvivono 45.333 dei 100.000 nati, si vede subito che non sopravvivono, cioè sono morti fino all'età di 65 anni, 54.667 di essi.

c) La tavola cui ci riferiamo presenta invece, nella colonna 4, un'altra serie di dati sui morti: non i morti *fino* all'età x, bensì i « morti *fra l'x^{mo} e l'$(x+1)^{mo}$ compleanno* ». Per $x = 65$, vi si legge il numero (1.579) dei morti fra il 65° e il 66° compleanno, che è uguale alla differenza tra i sopravviventi a 65 anni e i sopravviventi a 66 anni (45.333 — 43.754). In generale i dati di questa serie rappresentano le eliminazioni fra due successivi compleanni. Essi ci indicano la velocità di eliminazione: fra 2 e 3 anni d'età si hanno 1.876 eliminati, fra 68 e 69 anni se ne hanno 1.903; in questi due anni d'età la velocità dell'eliminazione è press'a poco uguale.

d) Ma un ugual numero di morti ha diverso significato, riguardo all'intensità con la quale la morte miete le schiere dei viventi, secondo che proviene da un gruppo di viventi più o meno numeroso. Per misurare tale intensità conviene dunque mettere in rapporto i morti in ciascun anno di età coi sopravviventi all'inizio dell'anno stesso. Il rapporto tra 1.876 morti fra 2 e 3 anni d'età e 81.982 superstiti a 2 anni risulta uguale a 0,023; quello tra 1.903 morti fra 68 e 69 anni e 40.315 superstiti a 68 anni risulta uguale a 0,047. Questi rapporti hanno il carattere di probabilità statistiche, avendo al numeratore il numero dei morti in ciascun anno di età e al denominatore il numero di coloro che avrebbero potuto morire in quell'anno di età. Appunto con la denominazione di « probabilità di morte » li troviamo raccolti nella colonna 2 della tavola. Invece che per intervalli di un anno d'età, la probabilità di morte si può calcolare per intervalli di più anni o per intervalli di una frazione di anno.

e) La probabilità di morte misura l'intensità dell'eliminazione in ciascun anno d'età: abbiamo visto or ora che su 1000 sopravviventi a 2 anni 23 muoiono nel corso del terzo anno di età. È chiaro che gli altri 977 su 1000 sopravvivono alla fine dell'anno stesso; poichè le eventualità possibili ed escludentisi a vicenda sono soltanto quelle di morire o di sopravvivere, se la probabilità di morte è 0,023, la probabilità di sopravvivenza è (1 — 0,023), cioè 0,977. In generale la differenza tra l'unità e la probabilità di morte in un dato intervallo d'età c'indicherà la « probabilità di sopravvivenza » alla fine dello stesso intervallo d'età. Appunto perchè si possono immediatamente ricavare dalle probabilità di morte, nella tavola dell'ASI non troviamo esplicitamente riportate le probabilità di sopravvivenza. Anche la probabilità di sopravvivenza si può calcolare per intervalli maggiori o minori di quello di un anno d'età.

f) La « durata mediana della vita dei sopravviventi all'età x » corrisponde a quell'età nella quale i sopravviventi stessi sono ridotti alla metà; infatti metà del numero totale delle durate di vita considerate è inferiore a tale limite e l'altra metà è superiore. Per esempio, a 24 anni sopravvivono 72.608 individui; la durata mediana della vita del sopravvivente a 24 anni corrisponde dunque all'età in cui i sopravviventi sono ridotti a 36.304. Scorrendo i dati della tavola, troviamo che a 70 anni i sopravviventi sono 36.384, a 71 anni 34.253; la durata mediana cercata è dunque lievemente superiore a 70 anni: risulta di 70 anni e 14 giorni se si ammette che i morti fra 70 e 71 anni siano uniformemente distribuiti nel corso di codesto anno d'età. La vita mediana residua del superstite a 24 anni si ottiene sottraendo 24 anni dal numero così trovato: risulta pertanto di 46 anni e 14 giorni. Tale è il significato dei dati esposti nella tavola dell'ASI alla colonna 6, sotto il titolo di « vita probabile ». Accanto alla mediana si possono determinare i quartili, i decili, ecc., per caratterizzare la distribuzione delle durate di vita individuali.

g) La « durata media aritmetica della vita residua dei sopravviventi all'età x » è uguale alla somma degli anni vissuti al di là dell'età x dalla generazione considerata nella tavola di sopravvivenza, divisa per il numero dei sopravviventi all'età x. Tale somma si può suddividere in tante somme parziali: anni vissuti fra l'età x e l'età $(x+1)$, anni vissuti fra l'età $(x+1)$ e l'età $(x+2)$, e così via. La tavola di sopravvivenza non ci offre indicazioni precise sul valore di queste somme parziali, ma solo indicazioni approssimative: essa ci dice, per esempio, che a 75 anni sopravvivevano 25.020 individui, a 76 ne sopravvivevano 22.603, e che perciò la generazione considerata ha vissuto fra 75 e 76 anni al massimo 25.020 anni (se tutti i morti fra 75 e 76 anni sono deceduti proprio nell'istante in cui stavano per compiere 76 anni) e al minimo 22.603 anni (se tutti i morti fra 75 e 76 anni sono deceduti proprio nell'istante in cui avevano compiuto 75 anni). Perchè la tavola di sopravvivenza ci desse indicazioni precise, bisognerebbe che segnasse esattamente l'età in cui è morto ciascuno dei 2.417 deceduti fra 75 e 76 anni. La prima ipotesi che abbiamo fatto or ora ci conduce a un dato errato in eccesso, la seconda ad uno errato in difetto; per accostarci maggiormente al vero sceglieremo l'ipotesi intermedia e supporremo che gli anni vissuti dalla generazione fra 75 e 76 anni d'età non siano stati nè 25.020, limite superiore, nè 22.603, limite inferiore, ma 23.811,5, semisomma dei due limiti (il che equivale a supporre che i morti fra 75 e 76 anni siano deceduti all'età di anni 75 e mezzo: ipotesi esatta nel caso di uniforme distribuzione delle morti nell'intero anno d'età, approssimata in ogni altro caso). Procedendo in modo analogo, cioè assumendo che il numero degli anni vissuti dalla generazione in ciascun anno d'età sia uguale alla semisomma dei sopravviventi all'inizio dell'anno stesso e dei sopravviventi all'inizio del successivo, la somma degli anni vissuti al di là dell'età x risulta espressa da metà del numero dei sopravviventi all'età x aumentata della somma dei numeri dei sopravviventi a tutte le età successive. Dividendo per il numero dei sopravviventi all'età x si ottiene la durata media aritmetica della vita. Tale è il significato dei dati raccolti nella colonna 5 della tabella dell'ASI sotto il titolo «vita media». Accanto alla media si può porre lo scostamento medio assoluto, od altra misura riassuntiva delle deviazioni delle durate individuali di vita dalla loro media; si può anche calcolare abbastanza facilmente la differenza media fra le durate individuali di vita.

Oltre i dati fin qui descritti, molti altri potrebbero aggiungersi per rendere ancor più completa la descrizione dell'esaurimento d'una generazione in funzione dell'età. Ma in una trattazione elementare come la nostra non conviene eccedere in particolari: ci basta avere mostrato come i dati sui sopravviventi all'inizio dei successivi anni di età, quelli sui morti nel corso dei singoli anni di età, quelli sulla probabilità di morte e sulla probabilità di sopravvivenza in ciascun intervallo annuale d'età, quelli sulla vita mediana o media aritmetica residua a ciascuna età, pur essendo tutti desumibili dalla prima delle serie ora ricordate, ossia da quella dei sopravviventi, contribuiscano a rendere più efficace la descrizione dell'ordine di estinzione di una generazione, presentando quest'unico fenomeno da tanti punti di vista differenti.

3. — L'esempio di tavola di eliminazione illustrato nel precedente paragrafo ci permette ora di esporre senza difficoltà in forma generale le nozioni relative a tale metodo di descrizione dei fenomeni collettivamente tipici.

Come appare dagli esempi stessi riferiti nel paragrafo 1, il complesso osservato può essere indivisibile in unità individualmente distinte (tale la nafta che defluisce dal serbatoio), ovvero divisibile in siffatte unità (tale la generazione che si va estinguendo).

E il graduale esaurimento del complesso può avvenire:

in modo continuo, rispetto alla variabile in funzione della quale esso è osservato: così per la nafta, se defluisce ininterrottamente;

o in modo discontinuo, tale però che l'eliminazione possa avverarsi in corrispondenza a qualsiasi valore di una variabile continua: così per le eliminazioni a causa di morte, le quali possono corrispondere a qualsiasi valore della variabile continua « età », ma avvengono saltuariamente;

o in modo discontinuo, tale che l'eliminazione possa avvenire soltanto in corrispondenza a dati valori della variabile: così le donne coniugate possono essere eliminate dopo aver messo al mondo 0 figli, 1 figlio, 2 figli, ecc.: possono cioè essere eliminate soltanto in corrispondenza ad un valore intero della variabile « numero dei figli ».

La precedente enumerazione dei modi di esaurimento non vuol essere completa; essa indica soltanto i modi che hanno maggior importanza dal punto di vista dello studio formale delle tavole di eliminazione.

Come abbiamo mostrato con l'esempio della tavola di sopravvivenza, lo svolgimento di un fenomeno di eliminazione può essere descritto mediante più funzioni della variabile in relazione alla quale l'eliminazione viene osservata: funzioni dai cui valori si possono desumere altrettante serie statistiche. Nel passare in rassegna queste funzioni, considerate nel loro significato più generale, seguiremo lo stesso ordine che abbiamo adottato nel caso particolare della tavola di sopravvivenza.

a) La funzione fondamentale, dalla quale possono desumersi tutte le altre, è quella che ci indica la consistenza del complesso osservato in corrispondenza ai successivi valori della variabile in relazione alla quale esso viene osservato: per esempio la quantità di nafta che rimane nel serbatoio in ogni dato momento, il numero delle donne che giungono ad avere un dato numero di figli. Per brevità la chiameremo *funzione di permanenza.*

La funzione di permanenza può essere continua o discontinua; nel primo caso la sua descrizione completa non può essere costituita da una serie statistica ma soltanto da un diagramma o da una formola.

b) La funzione che ci indica la differenza tra la consistenza iniziale del complesso osservato e la sua consistenza corrispondente a ciascun successivo valore della variabile si ricava immediatamente dalla precedente, ogni valore di questa seconda funzione essendo uguale alla differenza tra due valori della prima. Per esempio, la quantità di nafta defluita dal serbatoio fino ad un dato momento è uguale alla quantità iniziale diminuita della quantità esistente in quel dato momento; il numero delle donne che non giungono ad avere un dato numero di figli è uguale al numero iniziale diminuito del numero delle donne che giungono ad avere quel dato numero di figli. Questa si può chiamare *funzione di eliminazione.*

Ripetiamo l'avvertenza data or ora al capoverso sotto *a*). Aggiungiamo che capovolgendo il diagramma della funzione di permanenza si ha senz'altro il diagramma della funzione di eliminazione.

c) La differenza tra due valori successivi della funzione di permanenza ci indica l'ammontare dell'eliminazione avvenuta in corrispondenza a un dato intervallo di valori della variabile, o a un dato valore di essa (secondo il modo di eliminazione). Per esempio, la differenza tra la quantità di nafta che resta nel serbatoio alla fine del 12° minuto e quella che vi resta alla fine del 13° misura l'eliminazione (deflusso) nel corso del 13° minuto; la differenza tra il numero delle donne che son giunte ad avere 5 figli e quello delle donne che sono poi giunte ad averne 6 misura l'eliminazione in corrispondenza al 5° figlio. Eliminazione *nell'intervallo* delimitato dai punti 12 e 13 nel primo esempio, eliminazione *nel punto* 5 nel secondo esempio.

Se l'intervallo che si considera ha un'ampiezza diversa dall'unità, mettendo in rapporto l'ammontare dell'eliminazione corrispondente all'intervallo stesso con l'ampiezza di esso si ottiene la corrispondente misura dell'eliminazione media aritmetica per l'intervallo unitario. Così, dividendo per 2 la quantità di nafta defluita tra la fine del 12° minuto e la fine del 14°, o dividendo per 0,5 la quantità defluita tra la fine del 12° minuto e la metà del 13°, si ottiene l'eliminazione media aritmetica per minuto.

L'ammontare dell'eliminazione avvenuta in un intervallo unitario, o in un punto, misura la *velocità d'eliminazione* in quell'intervallo, o in quel punto.

Se la funzione di permanenza è continua, si può restringere indefinitamente l'intervallo considerato, in modo che il rapporto tra l'eliminazione avvenuta nell'intervallo e l'ampiezza dell'intervallo misuri la velocità di eliminazione in un intervallo sempre più ristretto a partire dal punto iniziale; e, ricorrendo al calcolo infinitesimale, con un passaggio al limite si può cercare la misura della velocità *nel punto* iniziale dell'intervallo (velocità indicata dalla derivata, in quel punto, della funzione di permanenza). Per esempio si può misurare la velocità di deflusso della nafta *nell'istante* iniziale del 12° minuto.

d) Per misurare l'intensità dell'eliminazione in relazione al complesso osservato bisogna tener conto della consistenza di questo all'inizio dell'intervallo, o nel punto, in cui avviene l'eliminazione. Dal punto di vista della velocità di eliminazione, 10 litri di nafta defluiti dal serbatoio in un minuto si equivalgono, sia che all'inizio del minuto il serbatoio contenesse 10.000 litri, sia che ne contenesse 100. Ma dal punto di vista dell'intensità dell'eliminazione, nel primo caso 10 litri corrispondono al deflusso di un millesimo della quantità iniziale, nel secondo caso al deflusso di un decimo.

In generale, il rapporto tra l'ammontare delle eliminazioni avvenute in un certo intervallo, o in un certo punto, e la consistenza del complesso osservato all'inizio di quell'intervallo, o in quel punto, ci dà il *saggio di eliminazione*. Quando il complesso osservato è costituito da elementi individualmente distinti, il saggio di eliminazione assume il carattere di probabilità statistica (*probabilità di eliminazione*, *probabilità di morte*, ecc.).

Se il numeratore di tal rapporto è costituito dall'eliminazione media aritmetica per l'intervallo unitario, calcolata come è detto al capoverso sotto *c*), avremo un saggio unitario medio aritmetico di eliminazione; se il numeratore è costituito (nelle ipotesi considerate al terzo capoverso sotto *c*) dalla velocità di eliminazione in un dato punto, avremo un saggio di eliminazione nel punto dato (*saggio istantaneo di eliminazione*, se le eliminazioni sono accertate in funzione del tempo).

e) Per misurare l'intensità della non eliminazione, ossia della permanenza, in relazione al complesso osservato, si mette in rapporto la consistenza del complesso stesso alla fine dell'intervallo, o al di là del punto, in cui avviene l'eliminazione, con la consistenza all'inizio dell'intervallo, o nel punto suddetto. Il rapporto così ottenuto, che possiamo chiamare *saggio di permanenza*, è uguale alla differenza tra l'unità e il saggio d'eliminazione e quindi si può calcolare immediatamente da questo. Quando il complesso osservato è costituito da elementi individualmente distinti, il saggio di permanenza assume il carattere di probabilità statistica (*probabilità di permanenza, probabilità di sopravvivenza*, ecc.).

f) La *permanenza mediana* dell'oggetto dell'osservazione nel complesso osservato, a partire da un dato punto (cioè da un dato valore della variabile), si ottiene facilmente ricercando il punto in corrispondenza al quale la consistenza del complesso è ridotta alla metà di quella che era nel punto dato. All'inizio delle osservazioni abbiamo nel serbatoio 10.000 litri di nafta; dopo 25 minuti e 33 secondi ve ne rimangono 5.000 litri: appunto il valore di 25′33″ indica la permanenza mediana della nafta nel serbatoio, poichè metà della quantità iniziale v'è rimasta per un tempo più breve e metà vi rimane per un tempo più lungo. Analogamente si determinano i quartili, i decili, ecc.

Quando il complesso osservato è divisibile in elementi individualmente distinti, la permanenza di ciascun elemento è data dalla differenza tra il valore della variabile corrispondente all'eliminazione dell'elemento stesso e il valore iniziale della variabile. Il dato che misura la permanenza mediana viene talvolta indicato in questo caso come *valore probabile* della permanenza stessa (vita probabile, numero probabile dei figli, ecc.): denominazione a nostro modo di vedere poco opportuna, perchè atta a generare equivoci nell'interpretazione, ma che è bene conoscere perchè frequentemente usata.

g) La *permanenza media aritmetica* dell'oggetto dell'osservazione nel complesso osservato, a partire da un dato punto, si determina nel modo seguente:

Se la funzione di permanenza è continua, non si può determinare rigorosamente la permanenza media aritmetica senza ricorrere al calcolo infinitesimale: essa è data dal quoziente dell'integrale definito della funzione stessa nell'intervallo fra il punto x, dal quale si parte, e il punto ω in corrispondenza al quale nel suo decrescere si annulla la funzione di permanenza — nel nostro esempio, fra l'istante x in cui s'inizia l'osservazione e l'istante ω in cui il serbatoio della nafta rimane completamente vuoto per l'avvenuto deflusso di tutto il liquido che conteneva —, per il valore della funzione nel punto x — nel nostro esempio la quantità di nafta esistente nel serbatoio nell'istante x. Approssimativamente, però, si può calcolare la permanenza totale, dalla quale si desume poi la permanenza media aritmetica, col supporre che in ciascun intervallo l'eliminazione avvenga in modo uniforme: di norma l'approssimazione è tanto maggiore quanto più ristretti sono gli intervalli che si considerano. Sapendo, per esempio, che alla fine del 12° minuto restavano nel serbatoio 9.000 litri di nafta e alla fine del 15° minuto ve ne restavano 6.000, calcoleremo in via approssimativa la permanenza totale della nafta nel serbatoio in quell'intervallo di tre minuti, moltiplicando per 3 (minuti) 7.500 (litri), semisomma della consistenza iniziale e della consistenza finale. Se in realtà il deflusso della nafta avviene in modo uniforme, il risultato del calcolo è esatto e non soltanto approssimato.

Se la funzione di permanenza è discontinua e non si conosce con precisione il punto in cui avviene ciascuna eliminazione (come nel caso della tavola di sopravvivenza), si può adottare il procedimento di approssimazione or ora indicato, come appunto abbiamo fatto per calcolare la durata media aritmetica della vita.

Se la funzione di permanenza è discontinua e le eliminazioni avvengono in punti determinati (come nel caso delle donne coniugate seguite attraverso i successivi parti), il calcolo preciso della permanenza media aritmetica è facile. Se si parte, nel calcolo, dal punto x, la somma dei numeri che indicano la consistenza del complesso nei punti $(x+1)$, $(x+2)$, $(x+3)$, ... misura la permanenza totale; divisa per il numero che indica la consistenza nel punto x, essa ci dà la permanenza media aritmetica. La somma dei numeri delle donne che sono giunte ad avere rispettivamente 5, 6, 7,.... figli, divisa per il numero delle donne che sono giunte ad avere 4 figli, ci dà il numero medio dei figli avuti, successivamente al quarto, da queste ultime.

Accanto alla permanenza media aritmetica si può porre spesso utilmente una misura dello scostamento medio, atta a riassumere le deviazioni delle permanenze individuali dalla permanenza media.

4. — Abbiamo mostrato come dalla funzione di eliminazione si desumano le altre funzioni più spesso impiegate per descrivere l'andamento del fenomeno di eliminazione. È chiaro, tuttavia, che — date le relazioni esistenti tra le funzioni stesse — la conoscenza di una qualsiasi di quelle considerate sotto *a*), *b*), *c*), *d*), *e*) basta per ritrovare le altre.

Se conosciamo il numero dei morti, di una stessa generazione, anno per anno d'età, potremo ricostruire la consistenza iniziale della generazione sommando tutti i morti dall'età 0 all'età ω. Otterremo così il numero dei sopravviventi all'età 0; diminuendolo del numero dei morti fra 0 e 1 anno ricaveremo i sopravviventi all'età 1; diminuendo questo del numero dei morti fra 1 e 2 anni ricaveremo i sopravviventi all'età 2, e così via. Potremo indi ricostruire tutte le altre serie utili a descrivere l'ordine di eliminazione della generazione considerata.

Se conosciamo soltanto la probabilità di sopravvivenza (o soltanto la probabilità di morte, dalla quale la prima può essere immediatamente desunta) in ciascun anno d'età, applicheremo la probabilità di sopravvivenza per l'intervallo da 0 a 1 anni di età ad un numero ipotetico N di sopravviventi a 0 anni; otterremo così il numero dei sopravviventi a 1 anno. A questo numero applicheremo la probabilità di sopravvivenza per l'intervallo da 1 a 2 anni; otterremo così il numero dei sopravviventi a 2 anni. Procedendo analogamente in tutti i successivi anni d'età, fino all'esaurimento della generazione, potremo ricostruire l'intera serie dei sopravviventi. O meglio, avremo ricostruito una serie di dati rigorosamente proporzionali ai dati originari sui sopravviventi: per ricostruire questi ci sarebbe stato necessario e sufficiente conoscere il numero vero dei sopravviventi a 0 anni, cui abbiamo sostituito il numero ipotetico N.

La tavola di sopravvivenza italiana alla quale ci siamo riferiti nel paragrafo 2 è costruita nel modo ora descritto; ed è perciò che la serie delle probabilità di morte (colonna 2) viene premessa a quella dei sopravviventi (colonna 3). Ad un ipotetico numero di 100.000 sopravviventi all'età 0 si è applicata la probabilità di sopravvivenza per il primo anno d'età, 0,86437 (uguale alla differenza fra 1 e la probabilità di morte 0,13563 indicata nell'ASI); si è così ottenuto il numero dei sopravviventi a 1 anno d'età: 86.437. Applicando a questo la probabilità di sopravvivenza per il secondo anno d'età, 0,94846, si è ottenuto il numero dei sopravviventi a 2 anni d'età: 81.982, e così via di seguito. Se le probabilità di morte dell'ASI fossero state desunte dall'osservazione di una generazione di individui nati simultaneamente, seguita attraverso più di un secolo dalla nascita fino alla completa estinzione, i dati sui sopravviventi ottenuti nel modo ora descritto risulterebbero proporzionali ai veri numeri dei sopravviventi della generazione stessa. Se, per esempio, questa fosse stata composta all'inizio di 456.234 individui, moltiplicando per il coefficiente fisso 4,56234 i numeri di sopravviventi indicati nell'ASI otterremmo gli effettivi numeri dei sopravviventi di quella generazione alle varie età.

Leggendo l'intestazione della tavola dell'ASI vediamo invece che le probabilità di morte non sono state desunte dall'osservazione di una sola generazione. Mercè il censimento del 1921 è stato possibile suddividere la popolazione italiana in tanti gruppi annuali d'età; d'altra parte le statistiche del movimento della popolazione secondo gli atti dello stato civile indicavano, per ciascun anno solare, il numero dei morti in ciascun anno di età. Mettendo in relazione, con opportuni accorgimenti, il numero dei morti in ciascun anno d'età (media annua 1921-1922) col numero dei viventi nello stesso anno di età alla fine del 1921, si è ottenuta la probabilità di morte in ciascun anno di età.

Sarebbero bastati anche i dati per un solo anno, ad esempio per il 1922, per giungere allo stesso risultato. In un anno solo di osservazione si possono calcolare le probabilità di morte per cento e più anni d'età. Le quali però, com'è ovvio, non sono desunte dall'osservazione di una stessa generazione seguita attraverso cento e più anni d'età, bensì dall'osservazione di cento e più generazioni seguite attraverso un solo anno d'età. Formalmente le serie dei sopravviventi, degli eliminati, delle probabilità di morte, delle probabilità di sopravvivenza, ecc. si presentano nello stesso modo in entrambi i casi, ma sostanzialmente differiscono profondamente.

La tavola di sopravvivenza di una generazione realmente esistita, per esempio dei nati nell'anno 1801 in Italia, ci descrive come si sia andata esaurendo attraverso il tempo quella generazione: è la rappresentazione di un fenomeno di eliminazione *realmente avvenuto*.

La tavola di sopravvivenza di una generazione ipotetica, che s'immagina sottoposta in ciascun anno d'età ad una mortalità pari a quella accertata in Italia nel 1922 (o nella media del biennio 1921-22), ci descrive come si esaurirebbe attraverso il tempo la nostra generazione ipotetica se la mortalità in ciascun anno d'età si mantenesse costantemente uguale a quella accertata in Italia nel periodo di osservazione: è dunque la rappresentazione di un fenomeno di eliminazione *ipotetico*.

Supponiamo che la prima tavola dianzi accennata ci avesse indicato una durata media aritmetica della vita di 49 anni e 3 mesi: questa sarebbe stata la vita media dei nati nel 1801.

La seconda tavola, quella dell'ASI, ci indica appunto una durata media della vita di 49 anni e 3 mesi. È la vita media dei nati nel

1921-22? o dei morti nel 1921-22? Nè l'una nè l'altra: è la vita media di una generazione ipotetica, la cui probabilità di morte in ciascun anno d'età è supposta uguale a quella accertata in Italia nella media annua del biennio 1921-22.

Il procedimento presenta qualche analogia con quello che si può impiegare in cinematografia per ottenere mediante una successione di fotografie simultanee la visione delle successive fasi di un fenomeno. Tante fotografie, prese successivamente nel tempo, di una medesima pianta ci permettono di ottenere la visione cinematografica dello sviluppo della pianta; ma tante fotografie, prese simultaneamente nel tempo, di altrettante piante diverse giunte a gradi successivi di sviluppo, ci permettono di ottenere la stessa visione.

Un'altra analogia: se vogliamo avere un'idea dello sviluppo della statura dell'uomo in funzione dell'età possiamo misurare uno stesso gruppo d'individui ad ogni successivo compleanno: ci occorrerà un buon secolo per completare il nostro lavoro. Osservando invece simultaneamente lo sviluppo della statura fra la nascita e il primo compleanno in un primo gruppo d'individui, fra il primo e il secondo compleanno in un secondo gruppo, fra il secondo ed il terzo compleanno in un terzo gruppo, e così via, potremo in un solo anno ricostruire lo sviluppo della statura quale avviene nel corso dell'esistenza umana. Ma non sarà lo sviluppo della statura di una generazione realmente esistita, bensì quello di una generazione ipotetica che si suppone abbia in ciascun anno d'età l'incremento di statura accertato, per quell'anno d'età, nell'anno di osservazione.

La tavola di sopravvivenza dei nati nel 1801 riflette nel suo andamento le circostanze che hanno influito sulla mortalità nel corso di un secolo e più. La tavola di sopravvivenza secondo la mortalità del 1921-22 riflette le circostanze che hanno influito sulla mortalità in quel biennio. L'una descrive uno stato di cose in gran parte lontano e sorpassato: ha grande valore storico e scarso valore attuale. L'altra descrive uno stato di cose vicino e non molto dissimile dal presente: ha scarso valore storico, perchè non descrive fenomeni realmente avvenuti, ma ha grande valore attuale. Analogo ragionamento si potrebbe ripetere per l'esempio delle stature.

Passando dal caso particolare della tavola di sopravvivenza ad una formulazione generale, possiamo dire che l'osservazione simultanea del fenomeno di eliminazione in tanti intervalli, corrispondenti ad intervalli successivi di valori della variabile in relazione alla quale esso viene osservato, può consentire la ricostruzione dello svolgimento del fenomeno stesso in funzione della variabile considerata: ricostruzione subordinata all'ipotesi che i saggi di eliminazione nei singoli intervalli si mantengano costanti.

5. — I casi di fenomeni di eliminazione che abbiamo fin qui considerato sono più semplici di molti casi che la realtà ci presenta. Abbiamo supposto un serbatoio di nafta che si vuota gradualmente, una generazione d'individui che si estingue gradualmente. Le eliminazioni dai complessi osservati, ai quali finora ci siamo riferiti, avvenivano per una sola circostanza: l'uscita della nafta dal serbatoio, la morte dell'individuo; ed all'uscita non si contrapponeva nessuna entrata. Ma nella realtà talvolta possono avvenire eliminazioni per circostanze diverse, che interessa studiare, e quindi descrivere, separatamente: la nafta può eliminarsi per deflusso ma anche per evaporazione, l'individuo può eliminarsi per morte ma anche per emigrazione. E mentre defluisce nafta dal serbatoio, altra nafta può affluire in esso; mentre si eliminano alcuni individui, altri possono subentrare per immigrazione. Potranno esservi parecchie bocche di afflusso e parecchie di deflusso; parecchie vie di entrata e parecchie vie di uscita. Si voglia, per esempio, costruire una tavola di sopravvivenza non per la popolazione maschile complessiva di un paese privo di immigrazioni e di emigrazioni, ma per i soli coniugati in un paese con movimenti immigratori. Fra un dato compleanno e il successivo, per esempio fra 40 e 41 anni, si avranno ancora eliminazioni per morte, come nel caso più semplice considerato al paragrafo 2; ma si avranno altresì: eliminazioni per vedovanza ed eliminazioni per emigrazione di coniugati; ingressi per matrimonio e ingressi per immigrazione di coniugati. Ed allora, si presenteranno due principali possibilità a chi osserva il fenomeno; descrivere l'andamento di esso in tutta la sua complicazione, o sceverare nella descrizione uno dei fattori di variazione del complesso osservato — in ispecie uno dei fattori di eliminazione — per mettere in rilievo l'influenza di esso. Se, nel seguire una generazione d'individui attraverso il corso della sua esistenza, accerteremo non soltanto quanti sopravvivano ad ogni compleanno ma anche quanti dei sopravviventi siano rispettivamente celibi, coniugati, divorziati, vedovi, giungeremo ad una esatta descrizione della realtà. Non potremo, però, confrontare i dati sui sopravviventi celibi a due età consecutive coi dati sui sopravviventi coniugati alle stesse età, per trarne conclusioni sulla diversa mortalità nelle due classi di stato civile, perchè le differenze fra le due coppie di dati non dipendono

solo dalle eliminazioni per morte ma anche dagli altri fattori di aumento e di diminuzione or ora ricordati. Se vogliamo ottenere due descrizioni comparabili dell'esaurimento per morte di un gruppo di coniugati e di un gruppo di celibi dobbiamo prescindere da tutti gli altri fattori di diminuzione e di aumento e ricercare come si esaurirebbero in funzione dell'età due gruppi uguali, uno di coniugati ed uno di celibi, per conseguenza della diversa mortalità delle due classi di stato civile. Dobbiamo, cioè: anzitutto calcolare le probabilità di morte, anno per anno di età; poi mediante esse costruire l'ordine di eliminazione di un ipotetico gruppo di coniugati e l'ordine di eliminazione di un ipotetico gruppo di celibi, a partire dalla stessa età, per esempio 18 anni.

La seconda operazione non presenta alcuna difficoltà: abbiamo chiarito nel paragrafo 4 come essa si svolga. Meno semplice è la prima: per spiegarla più semplicemente ci varremo di un esempio numerico. Supponiamo che il numero dei coniugati sopravviventi a 40 anni sia di 12.132, quello dei sopravviventi a 41 anni sia di 12.102. Poichè la differenza di 30 unità fra i due numeri dipende non solo dalle morti ma anche dalle altre circostanze sopra enumerate, non possiamo ricavare per sottrazione il numero dei morti. Bisogna che il numero dei morti venga accertato direttamente e che vengano pure accertate direttamente le entrate e le uscite dovute ad altre circostanze. Supponiamo di averle accertate, e sia risultato che fra il 40° e il 41° compleanno la schiera dei coniugati perde 84 individui per morte, 72 per vedovanza, 54 per emigrazione; mentre acquista 139 individui per matrimonio e 41 per immigrazione. La differenza fra la somma delle uscite, 210, e la somma delle entrate, 180, corrisponde appunto alla differenza tra il numero dei sopravviventi a 40 anni e il numero dei sopravviventi a 41. (Le entrate avrebbero potuto anche superare le uscite; in tal caso il numero dei coniugati sopravviventi a 41 anni avrebbe superato quello dei sopravviventi a 40: è quanto avviene nelle età più giovanili). Muniti dei precedenti dati cerchiamo di risolvere il nostro problema: qual è la probabilità di morte dei coniugati fra 40 e 41 anni di età?

Sono stati esposti a morire coniugati fra 40 e 41 anni non soltanto i 12.132 individui che sopravvivevano coniugati a 40 anni, ma anche i 180 che si sono loro aggiunti, in seguito a matrimonio o ad immigrazione, nel corso del 41° anno d'età. I casi possibili di morte sono dunque 12.312, e a prima vista sembra che debba essere questo il denominatore della probabilità di morte, il cui numeratore è 84, numero dei morti. Ma notiamo che i 180 nuovi coniugati od immigrati non sono stati esposti a morire in uguale misura degli altri; se supponiamo che in media si siano aggiunti al complesso da noi osservato ad un'età di circa 40 anni e mezzo, possiamo dire che per circa mezzo anno sono stati esposti a morire in altre classi di stato civile o in altre popolazioni, e per altro mezzo anno lo sono stati nel complesso da noi osservato. Perciò, in via di approssimazione, non conteremo i 180 come 180, bensì come 90 esposti a morire tra i coniugati da noi osservati (ai 139 nuovi coniugati faremo corrispondere 69,5 esposti a morire coniugati e 69,5 esposti a morire nella classe di stato civile cui prima appartenevano; ai 41 coniugati immigrati faremo corrispondere 20,5 esposti a morire nella popolazione da noi osservata e 20,5 esposti a morire nelle popolazioni cui prima appartenevano). D'altra parte notiamo che i 72 vedovati e i 54 emigrati per una frazione del 41mo anno d'età (in media per mezzo anno: è questa l'ipotesi più semplice, ma se fosse più conforme alla realtà si potrebbe così qua come dianzi adottarne un'altra) sono esposti a morire non come coniugati nella popolazione da noi osservata, ma gli uni come vedovi nella popolazione stessa e gli altri come coniugati o come vedovi nelle popolazioni di cui vanno a far parte: anche questi 126 individui, dunque, vanno contati soltanto per metà nel numero degli esposti a morire. Così che il numero degli esposti a morire coniugati fra 40 e 41 anni si ricostruisce nel modo seguente: 12.132 sopravviventi a 40 anni, più 90 che rappresenta la metà dei nuovi entrati, meno 63 che rappresenta la metà degli eliminati per circostanze diverse dalla morte: cioè 12.159. E la probabilità di morte risulta uguale a 84 : 12.159 = 0,00691.

Quella così calcolata è la probabilità di morte che si avrebbe fra i coniugati in assenza di entrate e in assenza di uscite per vedovanza e per emigrazione. È una misura di ciò che avverrebbe, in date ipotesi solo in parte corrispondenti al vero, se la realtà fosse diversa da quella che è; perciò è una misura necessariamente arbitraria ed approssimativa. Tuttavia può servire utilmente. Determinate le probabilità di morte nel modo anzidetto per ogni anno d'età dal 18° compleanno in poi, si potrà desumerne una tavola di sopravvivenza dei coniugati, la quale ci indicherà come questa classe di stato civile tenda ad esaurirsi per il solo effetto della mortalità. Si sarà così sceverata l'azione di questo fattore di variazione da quella degli altri.

Generalizzando, per i casi nei quali le eliminazioni avvengono nel corso di intervalli dati, possiamo dire che, ove sussistano più fattori di eliminazione e fattori di ingresso, un'espressione approssimativa del saggio (probabilità) di eliminazione che si avrebbe in un dato intervallo per l'azione isolata di un determinato fattore di eliminazione si ottiene mettendo in rapporto il numero delle eliminazioni dovute a quel fattore con la consistenza del complesso osservato all'inizio dell'intervallo, aumentata di metà del numero degli ingressi avvenuti nel corso dell'intervallo stesso e diminuita di metà del numero delle eliminazioni avvenute nel corso dell'intervallo stesso per fattori diversi da quello considerato. Calcolati in tal modo i saggi di eliminazione, è facile costruire poi una tavola di eliminazione nella quale l'azione del fattore considerato sia sceverata da quella degli altri.

Se le eliminazioni avvengono soltanto in dati punti, il procedimento per il calcolo dei saggi di eliminazione è più semplice ed è rigoroso. Supponiamo di seguire un complesso di donne coniugate attraverso i successivi parti: 2.000 son giunte ad avere tre figli, 1.400 giungono ad averne quattro. Se il complesso osservato è chiuso, non riceve cioè immigrazioni e non dà emigrazioni, la differenza fra 2.000 e 1.400 ci indica il numero delle donne che si sono fermate al terzo figlio; e il rapporto tra 600, numero di queste eliminate, e 2.000, numero iniziale, ci dà un saggio di eliminazione di 0,30. Ma se vi sono movimenti migratori, e precisamente la schiera delle 2.000 donne con tre figli si è accresciuta per immigrazione di 200 donne che hanno avuto tre figli ed è diminuita per emigrazione di 100, il numero delle donne che avrebbero potuto giungere al quarto figlio diviene 2.000 più 200 meno 100, cioè 2.100; il numero delle eliminate diviene 700; il saggio di eliminazione 0,33.

Generalizzando, se le eliminazioni dovute al fattore studiato avvengono in punti dati, mentre eliminazioni dovute ad altri fattori ed ingressi avvengono nel corso degli intervalli delimitati da tali punti, si procederà nel modo seguente per il calcolo del saggio di eliminazione.

Alla consistenza del complesso osservato in un punto dato si aggiungerà il numero degli ingressi avvenuti nell'intervallo tra esso ed il successivo punto di eliminazione; e se ne detrarrà il numero delle eliminazioni derivate da fattori diversi da quello studiato, nell'intervallo stesso. La differenza tra il numero così ottenuto, che esprime la consistenza rettificata nel punto dato, e la consistenza (non rettificata) nel successivo punto di eliminazione indica il numero delle eliminazioni avvenute nel punto dato per effetto del fattore studiato: essa costituisce il numeratore del saggio di eliminazione, di cui la consistenza rettificata è il denominatore.

6. — Ove sia opportuno, si possono costruire tavole di eliminazione nelle quali si tenga simultaneamente conto delle eliminazioni dovute a più fattori. Si possono costruire tavole nelle quali si tenga conto anche degli ingressi (ma in questo caso non si diranno più tavole d'eliminazione, bensì « tavole d'ingresso e d'eliminazione »); per esempio si può costruire la tavola di sopravvivenza di una generazione suddivisa per classi di stato civile, tenendo conto della frequenza delle morti e della frequenza degli ingressi e degli egressi in ciascun anno d'età per ciascuna classe di stato civile.

7. - La ricostruzione dello svolgimento del fenomeno di eliminazione attraverso le sue fasi successive, eseguita mediante l'osservazione simultanea delle fasi stesse, permette, come abbiamo visto, di misurare l'influenza delle circostanze attuali sullo svolgimento del fenomeno. La ricostruzione dello svolgimento del fenomeno di eliminazione permette di misurare l'influenza di un singolo fattore di eliminazione, esclusa l'influenza di altri fattori di eliminazione e di fattori d'ingresso. La combinazione dei due metodi riesce a rendere comparabili i risultati di rilevazioni che non si potrebbero direttamente paragonare in modo corretto.

Supponiamo, per esempio, di voler confrontare la mortalità dell'Italia con quella della Nuova Zelanda nel biennio 1921-22. Mettendo in rapporto il numero medio annuo dei morti col numero medio annuo degli abitanti, si trova una frequenza di 17,54 morti per 1000 abitanti in Italia, di 8,75 nella Nuova Zelanda. La mortalità appare due volte maggiore in Italia. Basta una breve riflessione per far intendere che la comparazione fatta in questo modo ci mostra bensì qual frazione dell'una e dell'altra popolazione viene ogni anno falciata dalla morte, ma non può indicarci la differenza esistente tra le condizioni igieniche e sanitarie dei due paesi. È noto, infatti, che la frequenza delle morti varia molto fortemente in funzione dell'età: molto alta nelle età infantili, scende rapidamente fino a toccare un minimo in corrispondenza agli anni dell'adolescenza; indi sale lentamente nelle età giovanili, e in seguito con rapidità crescente al crescere dell'età, divenendo nuovamente molto alta nelle età senili. Ora l'Italia, paese di incremento demografico relativo lento — in confronto alla Nuova Zelanda — e di forte emigrazione, ha relativamente molti vecchi e molti bambini, relativamente pochi adulti: predominano — sempre in confronto alla Nuova Zelanda — le età nelle quali la mortalità è alta. La Nuova Zelanda, paese d'incremento demografico rapido ma di natalità inferiore alla nostra, e di forte immigrazione, ha invece molti adulti e relativamente pochi vecchi e bambini: predominano le età nelle quali la mortalità è bassa.

Il mezzo più semplice per giungere alla corretta comparazione della mortalità dei due paesi parrebbe quello di supporre in entrambi composta ugualmente, per gruppi annuali d'età, la popolazione: di calcolare cioè la mortalità generale come media ponderata delle mortalità accertate nei singoli anni d'età, assumendo per ciascun anno d'età lo stesso « peso » nelle due popolazioni. Ma in questo metodo è implicita una contraddizione, poichè si adotta una composizione per età che, almeno per uno dei due paesi, è incompatibile con l'andamento effettivo della mortalità in funzione dell'età.

Meno semplice, ma più corretto, è il metodo della tavola di sopravvivenza. Con esso costruiamo l'ordine di estinzione di due generazioni ipotetiche: una di queste incontra nei vari anni di età la mortalità accertata in quegli anni nella Nuova Zelanda secondo l'esperienza del 1921-22, l'altra incontra la mortalità accertata in Italia nello stesso periodo. Dalle due tavole di sopravvivenza così costruite desumiamo la durata media aritmetica della vita: 64 anni nella Nuova Zelanda, 50 anni in Italia. Ciò significa che in una popolazione dove il numero dei nati fosse costantemente uguale a quello dei morti, e la mortalità in ciascun anno d'età fosse quella accertata nella Nuova Zelanda nel 1921-22, si vivrebbe in media 64 anni, cioè si rinnoverebbe in ogni anno solare un sessantaquattresimo della popolazione: la mortalità sarebbe di 1 : 64, cioè di 15,63 per 1000 abitanti. Invece in un'analoga popolazione, ma con mortalità in ciascun anno d'età pari a quella italiana, si vivrebbe in media 50 anni e si rinnoverebbe in ogni anno solare un cinquantesimo della popolazione: la mortalità sarebbe di 1 : 50, cioè di 20 per 1000 abitanti. Con questo metodo di confronto corretto vediamo che la mortalità italiana non supera quella neozelandese della metà, come ci era parso da principio, ma solo di un quarto o poco più.

Se ora volessimo comparare la natalità dei due paesi nello stesso periodo 1921-22, potremmo facilmente renderci conto che il semplice rapporto fra nati e abitanti indica bensì correttamente qual frazione si aggiunga annualmente all'una e all'altra popolazione per conseguenza delle nascite, ma non può servire per una corretta comparazione della tendenza alla procreazione nei due paesi. Predominano relativamente, infatti, nella Nuova Zelanda i gruppi di età idonei alla procreazione, in Italia quelli non idonei. Per giungere ad un confronto corretto potremmo anzitutto calcolare la frequenza delle nascite in singoli gruppi annuali di età delle madri (o dei padri) e poi applicare le frequenze così calcolate ai numeri di viventi in ciascun anno d'età secondo la tavola di sopravvivenza femminile (o secondo quella maschile). Troveremmo in tal modo in quale misura si riprodurrebbe una generazione ipotetica nella quale la mortalità e la fecondità nei singoli anni d'età fossero quelle osservate rispettivamente in Italia o nella Nuova Zelanda.

Sarebbe facile esporre in modo generale le applicazioni delle tavole di eliminazione, esemplificate nel presente paragrafo; ma, mentre l'esposizione riuscirebbe pesante, aggiungerebbe poco o nulla alla suggestione che il lettore può trarre dagli esempi recati. Sarebbe ancor più facile moltiplicare gli esempi. Invece che la natalità, potremmo confrontare col sussidio della tavola di sopravvivenza lo svolgimento in due o più paesi di un altro qualsiasi fenomeno sociale osservabile e variabile in funzione dell'età, per esempio quello della delinquenza. Invece che riferirci alla tavola di sopravvivenza, potremmo riferirci ad altra tavola di eliminazione. Ma neppure accumulando esempi aggiungeremmo gran che di nuovo a quanto abbiamo esposto; e dato il sistema del nostro corso preferiamo lasciare al lettore l'elaborazione di ulteriori applicazioni.

8. — Ogni serie di dati statistici che rappresentino altrettante misure di una circostanza quantitativa può essere facilmente trasformata, dopo essere stata disposta in ordine di grandezza, in una serie di valori d'una funzione di permanenza. Per esempio, i dati sui redditi dei singoli censiti possono essere disposti in ordine crescente di grandezza; dalla loro serie si potrà poi ottenere, sommandoli tutti, il numero u_0 dei redditi superiori a 0 lire, e successivamente, sottraendo da u_0 il numero dei redditi non superiori ad ogni dato valore x, il numero u_x dei redditi superiori ad x. La serie dei numeri u_x, progressivamente decrescenti al crescere di x, può essere considerata ed elaborata come una funzione di permanenza; a certi fini di analisi riesce utile questo trattamento.

Indicazioni bibliografiche. — LEXIS W., *Abhandlungen zur Theorie der Bevölkerungs-und Moralstatistik*, Jena, Fischer, 1903 (capitoli II, IV). — BROGGI U., *Matematica attuariale*, Milano, Hoepli, 1906. — WESTERGAARD, già citato, capitolo VI B. — MOIR H., *Sources and characteristics of the principal mortality tables*, New York, Actuarial Society of America, 1919. — ISTITUTO CENTRALE DI STATISTICA, *Tavole di mortalità della popolazione italiana*, Roma, 1931. — STATISTIQUE GÉNÉRALE DE LA FRANCE, *Statistique internationale du mouvement de la population*, Paris, I voi. 1907, II vol. 1913. — MORTARA G., *Tavole di criminalità e di recidività*, in *Giornale degli Economisti*, gennaio 1910; *Tavole di sopravvivenza e delle variazioni di stato civile e tavola di natalità per Milano*, Napoli, Atti del R. Istituto d'Incoraggiamento di Napoli, Serie VI, Vol. V.

Quesiti ed esercizi: 1. — Si definisca il fenomeno di eliminazione e la tavola di eliminazione e si dia qualche esempio dell'uno e dell'altra.

2. — Partendo dai dati di una tavola di sopravvivenza (per esempio di quella riferita nell'ASI 1929, pag. 42), si determini: *a*) il numero degli eliminati fino all'età di 20, di 40, di 60, di 80, di 100 anni, *b*) il numero degli eliminati fino al quindicesimo compleanno, dal quindicesimo al sessantesimo, oltre il sessantesimo (cioè degli eliminati prima del normale inizio della vita economicamente produttiva, nel corso di essa, dopo la normale sua fine); *c*) la durata mediana della vita residua, per il sopravvivente a 0 anni, a 1 anno, a 5 anni, a 15 anni, a 60 anni, a 80 anni (Perchè questa durata, che è la così detta *vita probabile*, risulta maggiore per il sopravvivente a 1 anno che per il sopravvivente a 0 anni?); *d*) i quartili e i decili nella serie delle durate di vita; *e*) lo scostamento mediano delle durate di vita inferiori alla durata mediana, rispetto alla mediana stessa; lo scostamento mediano di quelle superiori; lo scostamento mediano generale.

3. — Ancora partendo dai dati di una tavola di sopravvivenza, si determini: *a*) la durata più frequente dell'esistenza, per il sopravvivente all'età di 3 anni, di 20 anni, di 80 anni; *b*) il numero totale degli anni vissuti dalla generazione cui la tavola si riferisce, rispettivamente: fra 0 e 15 anni, fra 15 e 60 anni, fra 60 e ω anni (Quale frazione del numero totale degli anni vissuti è compresa fra 15 e 60 anni d'età, cioè nel periodo economicamente produttivo dell'esistenza?); *c*) il numero medio degli anni vissuti da ciascun sopravvivente all'età 0 nei tre periodi d'età ora indicati; *d*) la durata media aritmetica della vita residua, per il superstite a 0 anni, a 1 anno, a 5 anni, a 15 anni, a 60 anni (Perchè la vita media aritmetica residua del sopravvivente a 0 anni è molto inferiore alla vita mediana residua, mentre per il superstite a 80 anni a vita media aritmetica è fortemente superiore alla vita mediana?).

4. — Ancora partendo dai dati di una tavola di sopravvivenza, si determini: *a*) la probabilità, per il sopravvivente a 0 anni, di sopravvivere a 15 anni, a 37 anni e 6 mesi, a 60 anni; *b*) la probabilità, per il sopravvivente a 0 anni, di morire fra 0 e 15, fra 15 e 60, fra 60 e ω anni; *c*) la probabilità, per il sopravvivente a 15 anni, di sopravvivere a 60 anni; e da questa si desuma la sua probabilità di morire fra 15 e 60 anni.

5. — Nella tavola di sopravvivenza riferita nell'ASI 1929, pag. 42, la probabilità di morte (col. 2) cresce ininterrottamente da anno ad anno di età, dai 50 anni in poi. Perchè il numero dei morti (col. 4), che cresce anch'esso

coll'età fino al settantasettesimo anno, va poi invece gradualmente diminuendo?

6. — Dalle probabilità di morte riferite nella suddetta tavola si ricavino, anno per anno di età, le corrispondenti probabilità di sopravvivenza.

7. — Perchè le tavole dell'ASI, 1929, pag. 42-44, sono intitolate *tavole di mortalità* e non *tavole di sopravvivenza*? Come si è potuto dalle osservazioni compiute in un solo biennio (1921-22) trarre elementi sufficienti per descrivere l'ordine di estinzione di una generazione attraverso cento e più anni di età?

8. — Si descriva in modo generale una tavola di sopravvivenza, indicando le principali serie che la compongono.

9. — Si traducano in formole i vari dati che compaiono nella tavola di sopravvivenza, indicando con u_x il numero dei sopravviventi all'età x ed esprimendo tutti gli altri dati in funzione esclusivamente dei dati u_x. In particolare si rappresenti in formole il numero degli eliminati fra 0 e 60 anni, fra 60 e 61 anni; la probabilità di morte fra 0 e 60 anni, fra 60 e 61 anni; la probabilità di soppravvivenza fra 0 e 60 anni, fra 60 e 61 anni; la durata media aritmetica della vita residua a 60 anni.

10. — Quale differenza intercede tra la tavola di sopravvivenza per la generazione dei nati in un dato anno e la tavola di sopravvivenza desunta dalle osservazioni di una popolazione in un dato anno? Quale dato ci potrebbe fornire, durante la compilazione di una tavola del primo tipo, un'espressione della durata media della vita della generazione, molto prima che la generazione stessa fosse totalmente estinta?

11. — I dati della tavola di sopravvivenza possono essere utilizzati per lo studio di un'altra tavola di eliminazione formalmente analoga. Si supponga che i dati sui sopravviventi indichino invece i numeri dei marciatori che, in una prova di resistenza, rimangono in gara rispettivamente al chilometro 0 (linea di partenza), al chilometro 1, al chilometro 2, ecc. Qual significato assumono, in tale ipotesi, i dati indicati nell'ASI 1929, pag. 42, come: «probabilità di morte», «morti», «vita media», «vita probabile»? E qual significato corrisponde alla probabilità di sopravvivenza? Quale media sarebbe la più adatta, nella stessa ipotesi, a misurare il percorso normale del marciatore? Quale indicherebbe il percorso mediano?

12. — Si spieghi la differenza fra tavole di eliminazione corrispondenti a complessi divisibili ed a complessi indivisibili in unità individualmente distinte; la differenza fra le tavole con eliminazioni in intervalli dati e quelle con eliminazioni in punti dati. Si diano esempi dei vari tipi.

13. — Si definiscano in modo generale la funzione di permanenza, la funzione di eliminazione, la velocità di eliminazione, il saggio di permanenza, il saggio di eliminazione, la permanenza media.

14. — Si spieghi chiaramente quali ipotesi siano implicite nel consueto metodo di calcolo della durata media aritmetica della vita secondo una tavola di sopravvivenza.

15. — Come si calcola il saggio di eliminazione quando si ha un solo fattore di eliminazione e non si hanno fattori d'ingresso? Come si modifica il calcolo quando agiscono simultaneamente più fattori di eliminazione e agiscono anche fattori di ingresso?

16. — La tavola di sopravvivenza per la Germania, costruita secondo l'esperienza del triennio 1924-26 indica una vita media (aritmetica) di anni 55,97

per i maschi e di 58,82 per le femmine. Qual è il significalo di questi dati? Sono i nati nel 1924-26, oppure i morti nel 1924-26, che vivono in media il numero di anni indicato?

Ancora per la Germania, secondo l'esperienza del 1924-26, la vita media residua a 25 anni è di anni 38,76 per i celibi, 43,92 per i coniugati, 38,32 per i vedovi. Si chiarisca il significato di questi dati.

In Italia la durata media aritmetica della vita risulla di anni 35 e mesi 5 secondo l'esperienza del 1881-82, di anni 42 e mesi 9 secondo l'esperienza del 1899-1902, di anni 50 secondo l'esperienza del 1921-22. Si spieghi il significato di queste cifre.

17. — In quale anno è nata la generazione femminile cui si riferisce la tavola di sopravvivenza dell'ASI 1929, pag. 43?

18. — Si spieghi in qual modo mediante l'osservazione simultanea delle successive fasi di un fenomeno di eliminazione si possa ricostruire l'andamento del fenomeno stesso dall'inizio alla fine dell'esaurimento del complesso osservato.

19. — Qual vantaggio si ottiene comparando la mortalità di due paesi, o dello stesso paese in due epoche diverse, mediante tavole di sopravvivenza, invece che mediante il rapporto tra il numero dei morti e quello degli abitanti?

20. — Per calcolare la probabilità di morte in un anno d'età, per esempio fra 40 e 41 anni, si suol mettere in rapporto il numero dei morti in quell'anno d'età durante un anno solare col numero medio dei viventi nello stesso anno d'età durante lo stesso anno solare, accresciuto di metà del numero dei morti. Come si può giustificare questo procedimento approssimativo?

21. — Si traducano in diagramma le serie dei sopravviventi maschi e femmine contenuti nelle tavole dell'ASI 1929, pagg. 42, 43. Dal diagramma si determini la durata mediana della vita per i maschi e per le femmine. Si traducano in diagramma le serie delle probabilità di morte per i due sessi; indi si descriva l'andamento della mortalità in funzione dell'età e si mettano in rilievo le differenze esistenti al riguardo tra i due sessi. Si traducano in diagramma le serie dei morti secondo l'età e si desuma dal diagramma la durata normale della vita.

Diagrammi di questo genere, per altre epoche, si trovano nell'atlante Raseri (tavole 47 a 49). Come si può ricavare dal diagramma della probabilità di morte quello della probabilità di sopravvivenza?

22. — Da un silo granarlo contenente 10.000 quintali di frumento viene estratto il contenuto alla velocità costante di 1.000 quintali all'ora. Si descriva il fenomeno in una tavola di eliminazione, assunto come variabile il tempo espresso in ore intere.

Si traducano in diagramma la funzione di permanenza, la funzione di eliminazione, il saggio di permanenza, il saggio di eliminazione.

23. — I seguenti dati indicano quante donne di un dato complesso abbiano avuto nel corso della loro esistenza rispettivamente 0, 1, 2, 3, ... figli. Si costruisca la tavola di eliminazione per il suddetto complesso di donne, includendovi le seguenti serie, disposte tutte in funzione del numero dei figli: *a*) Numero delle donne che hanno avuto *almeno* x figli; *b*) numero delle donne che hanno avuto *meno di* x figli; *c*) numero delle donne che hanno avuto soltanto x figli (i nostri dati); *d*) probabilità, per la donna che giunge ad avere

x figli, di averne almeno un altro, l'$(x+1)^o$; *e*) probabilità, per la donna che giunge ad avere x figli di fermarsi a questo numero. Si determini poi il numero mediano dei figli avuti da ciascuna donna del complesso osservato, il numero medio aritmetico, il numero normale. Si traduca in diagramma la funzione di permanenza.

I seguenti dati rappresentano: colonne x numero d'ordine dei figli, colonne u_x numero delle donne che hanno avuto soltanto x figli.

x	u_x	x	u_x	x	u_x	x	u_x
0	59	4	118	8	133	12	44
1	46	5	136	9	120	13	28
2	63	6	151	10	101	14	13
3	90	7	145	11	67	15	5

24. — Al primo anno di corso d'una scuola s'iscrivono inizialmente 1.000 studenti. Nei 4 anni di studio questa schiera d'iscritti riceve le seguenti variazioni:

Variazioni	I Corso	II Corso	III Corso	IV Corso
Aumenti per trasferimenti da altre scuole	25	20	15	10
Diminuzioni per trasferimenti ad altre scuole	10	15	20	25
Diminuzioni per riprovazione agli esami finali del corso . . .	160	120	80	40
Diminuzioni per morte . . .	5	4	4	3

Si costruisca la tavola di eliminazione, coi seguenti dati: *a*) studenti presenti all'inizio di ciascuno dei quattro corsi e all'inizio del V corso (cioè del primo corso della scuola immediatamente superiore); *b*) studenti eliminati in ciascun corso per trasferimento; *c*) eliminati per riprovazione; *d*) eliminati per morte; *e*) entrati per trasferimento; *f*) saggio di eliminazione, in ciascun corso, per trasferimento; *g*) per riprovazione; *h*) per morte.

Si costruisca poi, mediante i saggi *g*), una tavola di eliminazione subordinata all'ipotesi che durante i quattro corsi non vi siano nuovi ingressi di studenti e non vi siano altre eliminazioni che quelle per riprovazione (si parta da un numero ipotetico di 1.000 iscritti all'inizio).

Si confronti la tavola stessa con quella costruita per un'altra scuola dello stesso tipo, nella quale il saggio di eliminazione per riprovazione è 0,037 nel primo corso, 0,054 nel secondo, 0,076 nel terzo e 0,112 nel quarto.

Qual è, nelle due scuole, la probabilità per l'iscritto all'inizio del I corso, di raggiungere regolarmente la fine del IV corso?

25. — Una banca accetta depositi a risparmio vincolati per un mese. All'inizio del mese successivo il depositante può richiedere il rimborso o rinnovare il deposito col vincolo di un altro mese, e cosi via. Indichiamo, per l'inizio del primo, del secondo, del terzo mese, ecc. (tempo 0, 1, 2, 3, ecc.), il numero u e l'ammontare v dei depositi che rimangono presso la banca.

Tempo x	Numero dei depositi u_x	Ammontare dei depositi (lire) v_x
0	2.731	2.845.687
1	2.584	2.731.014
2	2.103	2.579.999
3	1.948	2.307.485
4	1.716	2.004.863
5	1.491	1.845.751
6	1.037	1.512.692
7	872	1.198.211
8	539	819.488
9	266	603.189
10	147	451.513
11	51	215.502
12	0	0

In questo caso si possono formare due tavole di eliminazione partendo dalle due serie fondamentali sopra riferite: nella prima l'unità è il deposito, nella seconda è la lira. Si formino le due tavole, comprendendo in ciascuna di esse, oltre la serie fondamentale, le seguenti serie: *a*) eliminazioni all'inizio di ciascun mese; *b*) saggio di eliminazione all'inizio di ciascun mese; *c*) saggio di permanenza all'inizio di ciascun mese; *d*) giacenza media aritmetica residua dei depositi, a partire dall'inizio di ciascun mese.

Si rappresentino in diagramma sopra uno stesso foglio le funzioni di permanenza delle due tavole, regolando le scale in modo da far coincidere l'ordinata iniziale dei due diagrammi. Si esamini e si commenti la rappresentazione grafica così ottenuta.

26. — Una generazione realmente esistita viene seguita dal momento della nascita a quello della completa estinzione. La generazione comprendeva all'inizio 100.000 individui, dei quali 48.751 femmine. Indichiamo il numero u_x delle femmine sopravviventi a ciascun compleanno fra il 15° e il 50°, limiti d'età fra i quali è stato contenuto l'esercizio dell'attività riproduttiva da parte delle donne della generazione considerata. Dai numeri delle sopravviventi alle età x e $(x+1)$ si calcoli il numero medio delle donne viventi fra l'età x e l'età $(x+1)$. A questo numero si applichi il saggio di natalità f_x, che indica il numero dei nati vivi per ogni 1000 donne di età fra x e $(x+1)$; si tro-

verà così il numero assoluto dei nati da donne di età fra x e $(x+1)$. Le tre serie delle donne viventi in ciascun anno di età, dei saggi di natalità, dei nati, ordinate l'una accanto all'altra in funzione dell'età delle donne, costituiranno una tavola di natalità. Ecco i dati da elaborare.

x	u_x	f_x	x	u_x	f_x	x	u_x	f_x
15	39.444	0	27	36.970	197	39	34.141	154
16	39.273	2	28	36.749	203	40	33.881	142
17	39.095	8	29	36.525	207	41	33.622	120
18	38.917	19	30	36.295	211	42	33.359	100
19	38.722	39	31	36.063	208	43	33.099	78
20	38.516	64	32	35.830	208	44	32.837	54
21	38.311	90	33	35.594	203	45	32.570	34
22	38.101	117	34	35.362	198	46	32.302	18
23	37.882	139	35	35.127	193	47	32.038	9
24	37.655	160	36	34.880	186	48	31.776	4
25	37.427	176	37	34.635	179	49	31.494	2
26	37.198	188	38	34.394	166	50	31.198	1

Dalla tavola di natalità, calcolata sulla base dei precedenti dati, si desumano le seguenti indicazioni: *a*) numero totale dei nati da donne della generazione data (ammesso che tali donne si siano unite esclusivamente con uomini della stessa generazione, e reciprocamente, tale numero indicherà in qual misura la generazione stessa si è riprodotta: tenendo presente che in origine la generazione era composta di 100.000 individui, si misuri l'eccedenza relativa della generazione riprodotta su quella riproduttrice); *b*) numero medio complessivo dei nati per ogni 100 donne sopravviventi a 15 anni; *c*) età media della madre all'atto della nascita di un figlio (valore mediano, media aritmetica, valore più frequente).

Si tracci il diagramma del numero dei nati in funzione dell'età dalla madre; lo si esamini e lo si commenti.

27. — In una corsa con ostacoli partono 53 cavalli: di questi vengono eliminati al primo ostacolo 2, al secondo 0, al terzo 1, al quarto 3, al quinto 1, al sesto 5, al settimo 2, all'ottavo 7, al nono 1, al decimo 4. In funzione del numero d'ordine del punto d'osservazione (0 partenza, 1 primo ostacolo, 2 secondo ostacolo, ecc., 11 arrivo) si descriva la corsa mediante una tavola di eliminazione. Si calcoli la probabilità, per un cavallo in partenza, di venir eliminato al primo, al secondo, ecc., ostacolo; e la probabilità di arrivare al traguardo. Si tracci il diagramma della funzione di permanenza.

28. — Date le stature di un certo numero di individui coetanei, come si potrebbe formare una tavola di eliminazione degli individui stessi in funzione della statura? Quali dati offrirebbe questa tavola?

CAPITOLO XVII.

La comparazione tra più serie statistiche.

Avvertenze sull'impiego dei vari metodi di rappresentazione delle serie statistiche nella comparazione fra più serie; associazione fra diversi metodi — Comparazione termine a termine fra due serie statistiche ordinate secondo la stessa circostanza, mediante differenze o rapporti; differenze, loro valori assoluti, medie di differenze; misure della disuguaglianza assoluta o relativa; misure del grado di coincidenza e del grado di non coincidenza tra due serie — Avvertenze per l'uso dei metodi grafici nella comparazione.

1. — Per comparare tra loro le manifestazioni di un medesimo fenomeno collettivamente tipico in diversi campi d'osservazione, o nello stesso campo osservato in epoche diverse, ovvero le manifestazioni di diversi fenomeni in uno stesso campo d'osservazione, occorre spesso confrontare tra loro più serie statistiche.

Confronti siffatti si eseguono coll'aiuto dei vari modi di rappresentazione delle serie statistiche, che abbiamo studiato nei capitoli IX-XIV. La media permette di graduare le diverse serie omogenee tra loro, che si vogliono comparare, secondo la misura della manifestazione del fenomeno (statura media per classi di leva successive, reddito medio individuale per popolazioni diverse); i dati sussidiari alla media aggiungono a codesta sommaria comparazione indicazioni sulla distribuzione per grandezza delle misure del fenomeno nelle singole serie. Le rappresentazioni grafiche rendono agevole il confronto anche tra le manifestazioni di fenomeni eterogenei (andamento comparativo della nuzialità e dell'accumulazione di risparmio, attraverso il tempo; distribuzione del pauperismo e dei reati contro la proprietà, per circoscrizioni territoriali); conformemente al carattere di questi metodi, la comparazione riesce ad un tempo sintetica come quella offerta dalla media — ma senza il vantaggio della sintesi numerica — ed analitica assai meglio di quella offerta dai dati sussidiari. L'interpolazione, quand'è applicabile, riunisce in sè una parte dei vantaggi della media e di quelli delle rappresentazioni grafiche; può servire anch'essa per la comparazione di fenomeni eterogenei (andamento comparativo della circolazione monetaria e del livello dei prezzi, attraverso il tempo; distribuzione delle stature e distribuzione dei pesi del corpo in un complesso di individui). I numeri indici semplificano il paragone tra serie, omogenee od eterogenee, ordinate secondo la stessa circostanza; giovano specialmente quando occorra paragonare, piuttosto che le misure reciprocamente corrispondenti di due fenomeni, le deviazioni di esse da un certo livello stabilito per entrambi con criterio uniforme (salario del minatore di carbone e salario del tessitore di cotone in diversi paesi, assunti come riferimento i salari britannici). I rapporti di composizione rendono più facili i confronti tra serie, col ragguagliare all'unità la somma dei termini di ciascuna serie (classificazione per professioni della popolazione d'un paese secondo più censimenti successivi); sono utili specialmente quando, piuttosto che le grandezze assolute, interessi confrontare le proporzioni nelle quali le varie parti entrano a costituire il tutto.

È opportuno in molti casi associare insieme, per la comparazione, varie forme di rappresentazione della serie statistica. Si potranno talvolta utilmente ridurre le medie o i loro dati sussidiari a numeri indici, per agevolarne il confronto; si potranno rappresentare e confrontare graficamente medie e dati sussidiari, risultati d'interpolazione, numeri indici, rapporti di composizione; si potrà eseguire l'interpolazione su numeri indici o su rapporti di composizione per renderne più semplicemente comparabili i risultati; si potranno calcolare medie e dati sussidiari da serie di numeri indici invece che da serie originarie. Tutte queste associazioni di metodi vengono suggerite dalle esigenze delle singole comparazioni che si vogliono compiere e mal si prestano ad essere ridotte a sistema: d'altronde chi conosca bene i fini e le applicazioni dei metodi stessi non sarà imbarazzato a predisporne le più adatte associazioni.

Per la correttezza della comparazione tra due serie è necessario presupposto l'impiego di identici criteri di rappresentazione. L'avvertenza è tanto ovvia che parrebbe superflua; che non lo sia, è dimostrato dall'esperienza quotidiana, poichè si vedono spesso eseguire confronti tra rappresentazioni informate a criteri differenti: per esempio si confrontano numeri indici composti dei prezzi, ottenuti per medie geometriche, con altri ottenuti per medie aritmetiche; si confrontano, senza usare le necessarie cautele, diagrammi disegnati ad una certa scala con diagrammi disegnati a scala differente e non opportunamente proporzionata alla prima, ecc.

2. — Due serie statistiche ordinate secondo la stessa circostanza, in modo che a ciascun termine dell'una corrisponda un termine dell'altra, talora richiedono di essere comparate fra loro numericamente

termine a termine. Se conosciamo per un certo complesso di studenti i risultati di due esami, per esempio di quello di statistica e di quello di economia, potremo confrontare tra loro i risultati medi (numero medio dei voti riportati) delle due prove, potremo confrontare la distribuzione dei risultati individuali dell'una prova con quella dei risultati dell'altra, ma potremo anche chiederci in che relazione stiano fra loro i risultati conseguiti nelle due prove da ogni singolo studente. A quest'ultimo quesito possiamo rispondere soltanto confrontando i voti riportati da ogni determinato studente nelle due prove.

Se i dati delle due serie che si vogliono confrontare sono omogenei, come nel nostro esempio, la comparazione numerica si può eseguire mediante differenze o mediante rapporti; se sono eterogenei, solo mediante rapporti. Per le possibilità di sviluppo delle comparazioni ha importanza sopratutto il primo caso, anche perchè spesso serie eterogenee possono essere rese omogenee tra loro, all'effetto della comparazione, mediante riduzione a numeri indici od a rapporti di composizione. Ci fermeremo, pertanto, sul confronto mediante differenze, accennando ai possibili sviluppi di esso.

La prima tappa di questo confronto consiste nel calcolo delle differenze fra i termini corrispondenti delle due serie. Nel nostro esempio potremo calcolare la differenza tra il voto riportato da ciascuno studente nell'esame di statistica e il voto riportato dallo stesso studente nell'esame di economia.

Di queste differenze potremo poi considerare:

a) una funzione atta a riassumerle, per esempio la loro somma, o una loro media;

b) i valori assoluti, e una funzione atta a riassumere tali valori assoluti, per esempio la loro somma, o una loro media;

c) i valori riferiti ai dati assunti come sottraendi, che potranno essere quelli dell'una serie o quelli dell'altra serie: cioè le differenze relative;

d) i valori assoluti di tali differenze relative.

Riguardo al criterio *a*), osserviamo che la media aritmetica delle differenze coincide con la differenza tra le medie aritmetiche delle due serie, così che per determinarla non occorre calcolare le differenze termine a termine.

Riguardo al criterio *c*), notiamo che esso è utile per un esame particolareggiato delle differenze tra le due serie, ma si presta poco alla sintesi; quest'osservazione vale anche per il criterio *d*).

Richiede meno breve discorso il criterio *b*). Il valore assoluto di ciascuna differenza misura la disuguaglianza esistente fra due termini corrispondenti delle due serie; la somma dei valori assoluti delle differenze misura la disuguaglianza totale esistente fra i termini delle due serie: divisa per il numero dei termini dell'una o dell'altra serie, ci indica la disuguaglianza media aritmetica.

La grandezza della somma dei valori assoluti delle differenze dipende, a parità di ogni altra condizione, dalla grandezza dei numeri fra i quali sono state calcolate le differenze. Nel nostro esempio, se i voti sono espressi in cifre proporzionali a 30 la somma dei valori assoluti delle differenze fra i risultati dei due esami è tripla di quella che si avrebbe se i voti fossero espressi in cifre proporzionali a 10. Per rendersi conto della grandezza relativa della disuguaglianza conviene riferire la somma dei valori assoluti delle differenze alla somma delle somme dei termini delle due serie.

Il rapporto così ottenuto varia tra 0 e 1; assume il valore 0 quando a ciascun dato dell'una serie corrisponde nell'altra serie un dato uguale, assume il valore 1 quando a ciascun dato differente da zero nell'una serie corrisponde nell'altra serie un dato nullo o di segno contrario. Tale rapporto è perciò adatto a servire come *indice del grado di disuguaglianza*, specialmente quando la distribuzione per grandezza dei dati in ciascuna delle due serie non sia limitata da alcun vincolo. Nel nostro esempio, se consideriamo i risultati di esami *sostenuti*, la distribuzione dei risultati è libera: in teoria potrebbe avvenire che tutti gli studenti approvati in statistica fossero riprovati col voto zero in economia, e viceversa, nel qual caso la somma dei valori assoluti delle differenze raggiungerebbe appunto un valore uguale alla somma dei voti riportati in entrambe le materie dagli studenti considerati.

Ma spesso non corrisponde al vero l'ipotesi della libera distribuzione per grandezza dei dati nelle due serie; ed allora si può cercar di ottenere un indice del grado di disuguaglianza col riferire la somma dei valori assoluti delle differenze non al massimo generico bensì al massimo specifico che essa avrebbe potuto raggiungere. Nel nostro esempio, se consideriamo i risultati di esami *superati*, nessun dato nelle due serie può essere inferiore a 18, e poichè d'altra parte nessuno può essere superiore a 30, la somma dei valori assoluti delle differenze non può superare il prodotto di 12 per il numero degli studenti considerati. Se poi prendiamo come dati di fatto i voti assegnati in uno dei due esami, la massima disuguaglianza possibile non è più neppur quella ora indicata: è quella che si avrebbe se a ciascun voto dato nell'una materia corrispondesse nell'altra materia il dato più lontano possibile da esso, compreso fra 18 e 30. Se, infine, prendiamo come dati di fatto i voti assegnati in entrambe le materie, ammettendo soltanto che possa variare la loro distribuzione fra gli studenti, troviamo che la massima disuguaglianza si avrebbe se ai voti disposti in ordine crescente nell'una materia corrispondessero i voti disposti in ordine decrescente nell'altra (o in ogni altra distribuzione che pur modificando la grandezza non modificasse il segno di ciascuna delle differenze così formate). Potremmo proseguire l'esemplificazione, ma ci fermiamo qui perchè volevamo soltanto mostrare come, mentre la somma dei valori assoluti delle differenze è quella che l'osservazione indica, il massimo possibile valore di essa può essere determinato con criteri diversi. Nè il minimo valore possibile di essa si può assumere sempre uguale a zero: se, per esempio, prendiamo come dati di fatto i voti assegnati in entrambe le materie, la minima disuguaglianza si avrebbe se ai voti disposti in ordine crescente nell'una materia corrispondessero i voti disposti in ordine crescente nell'altra materia (o in ogni altra distribuzione che pur modificando la grandezza non modificasse il segno di ciascuna delle differenze così formate). Insomma, tanto il massimo quanto il minimo valore possibile della somma dei valori assoluti delle differenze possono essere determinati con criteri diversi, secondo la natura e il modo di variare dei dati delle due serie; e ai diversi criteri corrispondono modi diversi di calcolare indici del grado di disuguaglianza. In generale non conviene abusare di questi indici: per molti fini è sufficiente la misura della disuguaglianza relativa, ottenuta col riferire la somma dei valori assoluti delle differenze alla somma delle somme dei termini delle due serie.

La differenza fra due numeri misura la parte che essi non hanno comune; la parte che essi hanno comune è misurata dal più piccolo in valore assoluto dei due numeri se questi hanno entrambi lo stesso segno, ed è nulla se essi hanno segno opposto. La somma dei valori assoluti delle differenze fra i termini corrispondenti di due serie misura pertanto la parte che essi non hanno comune; mentre la parte che essi hanno comune è misurata dalla somma dei termini dell'una serie minori od uguali in valore assoluto dei corrispondenti termini dell'altra aventi ugual segno, e dei termini dell'altra serie minori in valore assoluto dei corrispondenti termini dell'altra aventi ugual segno. Mettendo in rapporto la seconda somma col totale della prima e della seconda, si ottiene un *indice del grado di coincidenza* tra le due serie; mettendo in rapporto la prima somma col totale della prima e della seconda si ottiene un *indice del grado di non coincidenza* fra le due serie. La somma dei due indici è sempre uguale all'unità. Se i dati, per loro natura positivi, non possono scendere sotto un certo valore minimo (e quindi hanno necessariamente comune una parte uguale a questo minimo), si potrà partire dal minimo stesso nella determinazione degli indici di coincidenza e di non coincidenza.

Si potrebbe calcolare un indice del grado di coincidenza con criterio diverso dal precedente, computando due volte, invece che una sola, la parte comune alle due serie, e computando una volta sola la parte non comune: l'indice così determinato sarebbe uguale alla differenza fra l'unità e l'indice del grado di disuguaglianza ottenuto dal rapporto tra la somma dei valori assoluti delle differenze e la somma delle somme dei termini delle due serie.

Un esempio di applicazione chiarirà quanto abbiamo finora esposto. Supponiamo di conoscere per dieci studenti i risultati degli esami di statistica e di economia, espressi in voti proporzionali a 30.

Numero d'ordine dello studente	I	II	III	IV	V	VI	VII	VIII	IX	X
Voti riportati nella statistica	18	19	21	22	24	25	26	27	28	30
Voti riportati nell'economia	21	20	24	26	23	30	24	28	25	29
Differenze tra i due risultati	— 3	— 1	— 3	— 4	+ 1	— 5	+ 2	— 1	+ 3	+ 1

La somma delle differenze è uguale a — 10; la loro media aritmetica è uguale a — 1: coincide con la differenza tra la media aritmetica dei voti riportati nella statistica (24) e quella dei voti riportati nell'economia (25). La somma dei valori assoluti delle differenze è 24, la loro media aritmetica 2,4. Il rapporto fra la somma dei valori assoluti delle differenze (24) e la somma delle somme dei termini delle due serie (240 + 250) è 0,049. L'indice del grado di coincidenza è uguale a 233 : 257 = 0,907; se però nel determinarlo non si parte da zero ma dal minimo di diciotto voti necessari per l'approvazione, diventa uguale a 53 : 77 = 0,688.

Qualche volta, invece di confrontare i dati di due serie che si corrispondono a vicenda, si confrontano le loro deviazioni dalle rispettive medie; ovvero, se si tratta di serie cronologiche, si confrontano le variazioni avvenute da ciascun intervallo di tempo, o da ciascun istante, ad un intervallo o ad un istante successivo: analogamente si può procedere per serie ordinate secondo altre circostanze quantitative. Nel nostro esempio, confrontando gli scostamenti dei risultati degli esami di statistica dal voto medio aritmetico (24) con gli scostamenti corrispondenti degli esami di economia dal voto medio aritmetico (25), si trova un indice del grado di coincidenza uguale a 18 : 40 = 0,450. Naturalmente col variare del criterio di comparazione variano i risultati della comparazione: è anche questa una buona ragione per attenersi ai metodi più semplici e di più facile interpretazione.

Quando le due serie da confrontare costituiscano due graduatorie (cioè due serie di numeri d'ordine attribuiti agli stessi elementi con criteri diversi: per esempio numero d'ordine dello studente secondo il risultato del suo esame di statistica e numero d'ordine secondo il risultato del suo esame di economia) si possono calcolare indici del grado di coincidenza o di non coincidenza tra le due graduatorie, applicando il procedimento sopra descritto agli scostamenti dei numeri d'ordine dalla media aritmetica (che è la stessa per le due serie). Se le graduatorie coincidono, l'indice del grado di coincidenza risulta uguale all'unità, se si corrispondono in ordine inverso risulta uguale a zero. Nel nostro esempio l'indice del grado di coincidenza fra i risultati dei due esami risulta uguale a 16 : 34 = 0,471.

Tutti questi indici, altri numerosi che sono stati proposti ed applicati, altri innumerevoli che possono immaginarsi, sono semplicemente mezzi atti a semplificare la comparazione preliminare tra le serie statistiche; in generale essi danno soltanto un indizio delle relazioni esistenti fra le due serie: relazioni che vanno accertate con un paziente esame analitico.

3. — Aggiungiamo qualche avvertenza intorno alla comparazione tra serie statistiche mediante diagrammi.

In uno stesso sistema di coordinate cartesiane si possono rappresentare insieme due o più serie statistiche mediante diagrammi a spezzata; quando le linee s'intrecciano conviene distinguerle mediante colori o tratti differenti. Riesce invece difficile sovrapporre efficacemente più di due diagrammi a canne d'organo: due si possono sovrapporre raffigurandoli come trasparenti se le differenze non sono dello stesso segno fra tutti i dati corrispondenti delle due serie.

Nel diagramma a spezzata o a curva continua, se importa confrontare le grandezze assolute dei dati corrispondenti delle due serie si adotterà per entrambe la stessa scala. Se importa confrontare le grandezze assolute degli scostamenti dei dati delle due serie dalle rispettive medie, si rappresenteranno graficamente questi scostamenti invece che i dati, oppure si disporranno le due scale, spostando l'origine di una di esse, in modo che coincidano tra loro i valori corrispondenti alle medie delle due serie.

Se importa confrontare le grandezze dei dati relative alle rispettive medie [o a due dati valori, scelti come riferimento] si assumerà come unità d'ordinata in entrambe le scale un'uguale frazione della media [o del valore scelto come riferimento]. Se importa confrontare le grandezze degli scostamenti relativi alle rispettive medie delle serie, si assumerà come unità d'ordinata per ciascuna serie un'uguale frazione della media; se importa confrontare le grandezze degli scostamenti relative allo scostamento medio, si assumerà come unità d'ordinata per ciascuna serie un'ugual frazione dello scostamento medio.

Se i dati delle due serie sono ordinati secondo una circostanza quantitativa, e interessa comparare la misura nella quale gli uni e gli altri variano col variare di questa circostanza: o importano le variazioni assolute, e ci si gioverà di un diagramma normale; o importano le variazioni relative ai valori iniziali di ciascun intervallo, e ci si gioverà del diagramma logaritmico, nel quale la variazione dell'ordinata è proporzionale alla variazione relativa intesa nel senso ora indicato.

Quesiti ed esercizi: 1. — Come s'impiegano i vari metodi di rappresentazione delle serie statistiche, nella comparazione tra più serie? Quali metodi possono servire per il confronto tra serie eterogenee; quali soltanto per il confronto tra serie omogenee? Come si possono associare tra loro i diversi metodi per rendere più efficace la comparazione?

2. — Si esegua, nei modi che si giudicano più opportuni, e ricorrendo ad associazioni di vari metodi dove sembri utile, la comparazione tra le seguenti serie cronologiche contenute nelle « Notizie statistiche retrospettive » dell'ASI: *a*) nati e morti in Italia negli ultimi trent'anni; *b*) natalità (nati per 1000 abitanti) e mortalità (morti per 1000 abitanti) negli ultimi trent'anni; *c*) produzione del frumento e produzione del vino; *d*) importazioni ed esportazioni; *e*) importazioni di materie prime gregge per le industrie ed esportazione di prodotti fabbricati; *f*) importazioni dalla Gran Bretagna, importazioni dalla Germania e importazioni dagli Stati Uniti negli ultimi trent'anni; *g*) depositi nelle casse di risparmio ordinarie e depositi nelle casse di risparmio postali.

3. — Si esegua la comparazione tra le seguenti serie contenute nell'ASI: *a*) matrimoni e nati vivi, per regioni; *b*) nuzialità (matrimoni per 1000 abitanti) e natalità (nati per 1000 abitanti), per regioni; *c*) distribuzione, secondo le cause di morte, dei morti maschi e femmine; *d*) fallimenti dichiarati, per regioni, in due anni successivi; *e*) varie categorie di reati denunciati per regioni; per esempio si confrontino: lesioni personali volontarie e furti; *f*) rendimento medio per ettaro nella coltura dei seguenti cereali: frumento, avena, riso, granturco, negli ultimi vent'anni; *g*) prezzi mensili del frumento, in più anni successivi (confrontare tra loro le diverse serie di prezzi mensili che si riferiscono ai diversi anni); *h*) prezzi mensili di più generi di consumo popolare (p. es. pane, uova, zucchero); *i*) distribuzione per regioni dei salari medi orari in due o più mesi diversi.

4. — Dieci persone hanno percepito in due anni successivi i redditi qui sotto indicati. Si confrontino le due serie di redditi coi metodi esposti nel paragrafo 2 del testo. Per ciascuno degli individui, designati con numeri romani, il primo dato indica il reddito nell'anno 1929, il secondo dato il reddito nell'anno 1930: I. 4.500, 3,800; II. 4.800, 4.500; III. 4.800, 4.900; IV. 5.100, 3.900; V. 5.400, 4.800; VI. 5.500, 5.700; VII. 6.000, 6.000; VIII. 9.500, 7.300; IX. 12.000, 10.000; X. 12.100, 11.300.

5. — Si confrontino tra loro termine a termine, applicando con discernimento i procedimenti esposti nel paragrafo 2 del testo, le seguenti serie cronologiche contenute nelle «Notizie statistiche retrospettive» dell'ASI: *a*) corso massimo e corso minimo del consolidato; *b*) produzione e importazione di frumento; *c*) importazione di cotone in bioccoli o in massa e esportazione di tessuti di cotone; *d*) tonnellaggio delle navi a vela e tonnellaggio delle navi a propulsione meccanica; *e*) entrate e spese effettive dello Stato; *f*) corso massimo e corso minimo del cambio su Parigi; *g*) importazione di carbon fossile e importazione di petrolio ed altri oli minerali.

6. — Si confrontino tra loro termine a termine le seguenti serie contenute nell'ASI: *a*) rendimento per ettaro della coltura del frumento, per regioni, in due anni successivi; *b*) numeri indici dei prezzi di borsa delle azioni di varie categorie di società, in due mesi diversi; *c*) merci esportate per ferrovia, secondo i transiti di uscita, in due anni finanziari diversi; *d*) valore delle importazioni, per paesi di provenienza, in due anni diversi.

7. — Con quali criteri si possono confrontare tra loro due serie statistiche termine a termine? Due serie statistiche quali si siano si possono confrontare tra loro termine a termine; oppure per la comparabilità occorre che le due serie abbiano qualche requisito comune? Si possono confrontare termine a termine soltanto serie omogenee tra loro per la natura dei dati che le costituiscono?

8. — Che cosa ci indica la media aritmetica delle differenze fra i dati corrispondenti di due serie che si confrontano termine a termine? Che cosa ci indica invece la media aritmetica dei valori assoluti delle differenze stesse? Si illustri il diverso significato delle due medie, utilizzando i dati del precedente esercizio 4.

9. — Come si può ottenere un indice del grado di disuguaglianza fra due serie che vengono confrontate termine a termine? In quali degli esempi con-

siderati nell'esercizio 6 riuscirebbe opportuna la determinazione di un siffatto indice?

10. — Si confrontino le variazioni da anno ad anno della natalità nel periodo 1881-1913 con le variazioni da anno ad anno della mortalità nello stesso periodo (V. ASI, «Notizie statistiche retrospettive»).

11. — Si esegua la comparazione mediante diagrammi delle serie considerate negli esercizi 2, 3, 5, 6, facendo precedere, ove sembri opportuno, alla comparazione grafica l'elaborazione dei dati che può giovare a renderli più utilmente comparabili.

Si provi ad applicare il diagramma logaritmico ai seguenti casi (indichiamo l'esercizio e la lettera): 2*a*, 2*d*, 2*g*, 6*a*, 6*d*.

12. — Quali sono le principali avvertenze che vanno tenute presenti nell'eseguire la comparazione fra due o più serie statistiche mediante diagrammi? Quando è opportuno l'uso del diagramma logaritmico?

13. — Si può applicare la comparazione mediante diagramma a spezzata ai casi considerati nell'esercizio 6?

CAPITOLO XVIII.

La comparazione fra più dati ordinati secondo due circostanze: la serie statistica di second'ordine. I metodi numerici per la sua rappresentazione sintetica od analitica.

Serie di ordine superiore al primo; tabella di second'ordine — Rappresentazione sintetica od analitica della serie di second'ordine — Impiego di medie: media generale e medie parziali dei dati della serie; medie dei valori delle circostanze quantitative secondo le quali è ordinata la serie; dati sussidiari alle medie — Impiego di rapporti indici — Impiego di rapporti di composizione: con riferimento al totale generale; con riferimento ai totali delle singole linee o delle singole colonne — Comparazione tra rapporti di composizione come mezzo per l'accertamento dell'esistenza di relazioni tra le due circostanze secondo le quali la serie è ordinata; indici di relazione — Comparazione tra la serie osservata ed una serie calcolata per l'ipotesi di assenza di relazione tra le due circostanze; contingenze, contingenza media assoluta; indici del grado di eccedenza o del grado di deficienza dell'osservazione rispetto al calcolo. Analisi critica di questo metodo. Se e quando sia possibile determinare la distribuzione del fenomeno che si avrebbe in assenza di relazione fra le due circostanze con riguardo alle quali è stato osservato — Varia natura delle relazioni che possono intercedere tra le due circostanze; vari modi nei quali tali relazioni si presentano — Sui metodi per la comparazione tra due serie di second'ordine.

1. — Nel definire la serie statistica, abbiamo distinto le serie di prim'ordine, cioè ordinate secondo una circostanza, da quelle di secondo, di terzo, di ennesimo ordine, cioè ordinate ad un tempo

secondo due, tre, *n* circostanze. Le une — che abbiamo finora studiato — ci indicano la misura di un fenomeno che corrisponde a ciascuna modalità o misura d'una circostanza, con riguardo alla quale esso è stato osservato; le altre — che studieremo in questo capitolo — ci indicano la misura di un fenomeno che corrisponde a ciascuna combinazione delle modalità o misure di due o più circostanze con riguardo alle quali esso è stato osservato. L'aggruppamento di un certo numero d'individui secondo la statura dà luogo ad una serie di prim'ordine, che permette di studiare come gli individui osservati si distribuiscano secondo tale carattere; l'aggruppamento secondo la statura ed il peso dà luogo ad una serie di second'ordine, che, presentando la distribuzione secondo entrambi i caratteri, consente di verificare se e quale relazione esista fra la statura ed il peso individuale; l'aggruppamento secondo la statura, il peso ed il perimetro toracico dà luogo ad una serie di terz'ordine, lo studio della quale dà modo di accertare le relazioni eventualmente esistenti fra questi tre caratteri.

Una serie di prim'ordine è contenuta in una colonna, o in una riga, coordinata ad un'altra ove sono indicate le diverse modalità o misure della circostanza secondo la quale i dati sono ordinati. Una serie di second'ordine invece viene esposta nel modo più adatto in una tabella, avente tante righe quante sono le modalità o misure considerate dell'una circostanza — che vengono specificamente indicate all'inizio delle singole righe —, e tante colonne quante sono le modalità o misure considerate dell'altra circostanza — che vengono specificamente indicate in testa alle singole colonne. Per esempio, nella serie di second'ordine che rappresenta la distribuzione dei matrimoni avvenuta in Italia secondo le combinazioni di stato civile degli sposi e delle spose, a ciascuna delle tre classi di stato civile degli sposi corrisponde una riga, a ciascuna delle tre classi di stato civile delle spose corrisponde una colonna.

Il numero dei posti, o *caselle*, esistenti nella tabella che racchiude la serie di second'ordine è uguale al prodotto del numero delle modalità o misure dell'una circostanza per il numero delle modalità o misure dell'altra circostanza: nel nostro esempio $3 \times 3 = 9$. Il numero dei dati contenuti nella tabella può essere minore di quello dei posti: ciò avviene quando a determinate combinazioni delle due circostanze non corrisponde alcun risultato dell'osservazione (vedasi, per esempio, nell'ASI, la tabella che indica la distribuzione dei matrimoni secondo le combinazioni d'età degli sposi e delle spose: in quella per l'anno 1928, su 196 caselle soltanto 176 contengono dati differenti da zero; alle altre 20 combinazioni delle età dello sposo e della sposa non corrisponde alcun matrimonio).

Chiameremo *tabella di second'ordine* quella che contiene una serie di second'ordine; ed analogamente parleremo di tabelle di terzo, di quarto ordine, ecc. (La tabella di second'ordine è detta anche *tabella a doppia entrata*, perchè per intendere il significato di un numero in essa contenuto è indispensabile leggere sia le indicazioni scritte all'inizio della riga sia quelle scritte in testa della colonna ove il numero è contenuto). Di rado s'impiegano tabelle di ordine superiore al quarto: già una tabella di quart'ordine riesce in molti casi ponderosa: se per esempio si volessero suddividere i matrimoni secondo le combinazioni dell'età e dello stato civile dello sposo con l'età e lo stato civile della sposa, ciascuna delle 196 caselle della tabella di second'ordine che indica la distribuzione dei matrimoni per combinazioni d'età dovrebbe essere suddivisa in nove caselle, corrispondenti alle nove possibili combinazioni di stato civile: la tabella di quart'ordine comprenderebbe dunque 9 volte 196, cioè 1.764, posti o caselle: avendo 42 righe e 42 colonne richiederebbe almeno una doppia pagina di formato normale.

In un corso elementare come il nostro, è sufficiente studiare la tabella di second'ordine: le nozioni ad essa relative possono, d'altronde, estendersi facilmente alle tabelle d'ordine superiore.

Normalmente la tabella di second'ordine costituita mediante dati greggi comprende, oltre i dati della serie di second'ordine, quelli delle due serie di prim'ordine che si possono formare ordinando i risultati dell'osservazione secondo l'una e secondo l'altra circostanza, separatamente considerate. Nel nostro esempio, in ciascuna riga della tabella tre dati indicano come si suddividano nei tre gruppi di stato civile della sposa i matrimoni nei quali lo sposo appartiene ad un determinato gruppo di stato civile; un quarto ed ultimo dato è la somma dei tre precedenti; indica cioè il numero totale dei matrimoni nei quali lo sposo appartiene a quel determinato gruppo di stato civile. La quarta colonna, che comprende tali somme, ci presenta dunque la serie dei matrimoni suddivisi secondo lo stato civile dello sposo. Analogamente, in ciascuna colonna della tabella sono tre dati che indicano come si suddividano nei tre gruppi di stato civile dello sposo i matrimoni nei quali la sposa appartiene a un determinato gruppo di stato civile; un quarto dato è la somma dei tre precedenti: indica cioè il numero totale dei matrimoni nei quali la sposa appartiene a quel gruppo di stato civile. La quarta riga, che comprende tali somme, ci presenta la serie dei matrimoni suddivisi secondo lo stato civile della sposa. La somma dei dati dell'ultima colonna della tabella è necessariamente uguale alla somma dei dati dell'ultima riga e rappresenta la somma della serie di second'ordine: nel nostro esempio il totale dei matrimoni.

Matrimoni avvenuti in Italia nel 1928, per combinazioni di stato civile

Stato civile dello sposo	Stato civile della sposa			Totale
	Nubile	Divorziata	Vedova	
Celibe	257.933	47	4.882	262.862
Divorziato	35	1	5	41
Vedovo	15.847	6	6.492	22.345
TOTALE . . .	273.815	54	11.379	285.248

2. — È spesso necessario, o almeno utile, elaborare i dati di una serie di second'ordine in modo da potere meglio dominare e più chiaramente intendere i fatti che essi rappresentano, o in modo atto a semplificare le comparazioni con altre serie analoghe. Talvolta si mira ad ottenere una rappresentazione sintetica, tal altra una rappresentazione ad un tempo sintetica ed analitica, tal altra ancora una rappresentazione soltanto analitica. A questi fini s'impiegano, come vedremo, metodi simili a quelli che si applicano per gli stessi fini alle serie di prim'ordine.

Non a tutte le serie di second'ordine convengono le medesime elaborazioni; avviene altrettanto, del resto, per le serie di prim'ordine: abbiamo visto, per esempio, che le rappresentazioni mediante diagramma continuo e mediante interpolazione sono applicabili soltanto alle serie che si possono riguardare come costituite da tanti valori di una funzione d'una variabile. Le serie di second'ordine possono essere ordinate secondo due circostanze quantitative (esempio: numero delle navi esistenti, secondo il tonnellaggio e l'età), oppure secondo due circostanze qualitative (esempio: numero delle navi, secondo la nazionalità e l'impiego), oppure secondo una circostanza quantitativa ed una qualitativa (esempio: numero dei censiti, secondo l'età e la professione). Esamineremo principalmente i metodi atti a rappresentare serie del primo e del secondo tipo: il lettore intenderà facilmente quali dei metodi stessi possano venir applicati alle serie del terzo tipo. Ricordiamo fino da ora che quando le circostanze sono entrambe quantitative, e a ciascuna combinazione di valori di esse corrisponde un solo dato statistico, i dati della serie di second'ordine possono essere riguardati come valori d'una funzione di due variabili (esempio: il numero dei matrimoni in funzione dell'età dello sposo e dell'età della sposa).

3. — La determinazione di una media dei termini della serie di second'ordine è sempre possibile, aritmeticamente, ma non sempre riesce utile, logicamente: valgono al riguardo le riserve espresse al principio del capitolo IX. In ogni caso la media ci fa perdere di vista irreparabilmente ogni relazione eventualmente esistente tra il fenomeno osservato e le circostanze secondo le quali le osservazioni sono state ordinate. Ma poichè essa ci indica il numero che dovrebbe comparire in ciascuna casella in certe ipotesi di distribuzione uniforme del fenomeno rispetto alle varie combinazioni di modalità o misure delle due circostanze, se non è adatta al fine della sintesi, è invece adatta a servire come livello di riferimento, quando si voglia apprezzare la non uniformità della distribuzione reale del fenomeno.

Gioverà un esempio. Sono stati coltivati a frumento, in una grande azienda rurale, dodici appezzamenti di dieci ettari ciascuno: su quattro di essi è stato seminato grano di una certa razza che chiameremo α, su altri quattro grano di razza β, sugli ultimi quattro grano di razza γ. Nei quattro appezzamenti seminati con grano di ciascuna razza sono stati applicati quattro diversi sistemi di coltura (differenti sopratutto per la qualità e la quantità della concimazione chimica), che chiameremo rispettivamente sistema *A*, *B*, *C*, *D*. La tabella che segue indica i raccolti ottenuti sui vari appezzamenti.

Raccolti ottenuti su appezzamenti di dieci ettari, con sementi diverse e sistemi diversi.

(quintali)

Razza di frumento	Sistema di coltura			
	A	*B*	*C*	*D*
α	187	175	212	241
β	243	251	257	296
γ	275	220	301	342

La media aritmetica dei dodici dati è 250 (quintali, raccolti su 10 ettari; cioè 25 quintali per ettaro). Essa ci riassume i dodici dati in uno solo. Ma avremmo potuto ricavarla senza bisogno di conoscere le razze dei grani seminati e i metodi di coltivazione adottati; e se ci fermiamo ad essa perdiamo il beneficio dell'ordinamento di second'ordine, che deve consentirci di riscontrare se la razza coltivata e il metodo di coltivazione influiscano sul prodotto. La media, indicandoci quale sarebbe stato il prodotto di ciascun appezzamento se — assunta come dato di fatto la quantità totale prodotta — tale influenza fosse stata nulla, ci può servire però come termine di comparazione per i singoli dati della serie, i quali presentano rispetto alla media stessa le differenze indicate nella seguente tabella.

Differenze tra i raccolti ottenuti sui singoli appezzamenti e il raccolto medio.
(quintali)

Razza di frumento	Sistema di coltura			
	A	*B*	*C*	*D*
α	— 63	— 75	— 38	— 9
β	— 7	+ 1	+ 7	+ 46
γ	+ 25	— 30	+ 51	+ 92

La somma dei valori assoluti delle precedenti differenze (scostamenti) è 444, la loro media aritmetica è 37 (scostamento medio assoluto). Il rapporto $37 : 250 = 0{,}148$ (coefficiente di variazione) costituisce un indice sintetico dell'influenza combinata della razza del seme e del sistema colturale sulla produzione. Ma soltanto l'esame dei singoli dati della tabella ci permette di discernere come varii il raccolto in relazione alle due circostanze: vediamo, per esempio, che il minimo raccolto si ottiene coltivando la razza α col metodo *B*; che il massimo raccolto è dato dalla razza γ, col metodo *D*; vediamo inoltre che i risultati dei vari sistemi di cultura sono diversi secondo le razze cui si applicano; e che i risultati delle varie razze sono diversi secondo i sistemi di coltura adottati.

Chi volesse rendersi conto del risultato dell'applicazione dei diversi sistemi a ciascuna razza di frumento potrebbe utilmente calcolare medie di produzione e scostamenti, riga per riga. Troverebbe:

Variazione, secondo il sistema di coltura, del raccolto fornito dalle diverse razze di frumento.

Razza di frumento	Media aritmetica	Scostamento, secondo il sistema di coltura			
		A	*B*	*C*	*D*
α	204	-17	-29	$+8$	$+37$
β	262	-19	-11	-5	$+34$
γ	284	-9	-64	$+17$	$+58$

e conchiuderebbe che il metodo *D* dà in tutti i casi i migliori risultati, mentre i risultati peggiori si ottengono col metodo *B* per i grani α e γ, col metodo *A* per il grano β; conchiuderebbe anche che, ove mancasse la possibilità di applicare il metodo *D*, converrebbe in ogni caso applicare il metodo *C*, ecc,

Chi invece volesse rendersi conto del risultato dell'impiego delle diverse sementi nell'applicazione di ciascun metodo di coltura calcolerebbe medie di produzione e scostamenti, colonna per colonna. Troverebbe:

Variazioni, secondo la razza del frumento, del raccolto ottenuto con diversi sistemi di coltura.

Sistema di coltura		*A*	*B*	*C*	*D*
Media aritmetica		235	215	257	293
Scostamento, secondo la razza del frumento	α	-48	-40	-45	-52
	κ	$+8$	$+36$	0	$+3$
	γ	$+40$	$+5$	$+44$	$+49$

e conchiuderebbe che i migliori risultati sono dati dalla razza β ove s'impieghi il metodo *B*, ma dalla razza γ ove s'impieghi uno degli altri metodi; i peggiori risultati sono dati sempre dalla razza α, ecc.

Il precedente esempio ci mostra come possano essere impiegate la media generale dei dati della serie di second'ordine e le medie particolari delle singole righe o delle singole colonne, nello studio del fenomeno osservato. La media generale ci dà una misura sintetica del fenomeno, senza alcun riguardo alle circostanze secondo le quali è stato eseguito l'aggruppamento di second'ordine: sarà utile determinarla quando le singole caselle corrispondano — come nel nostro esempio — ad altrettanti intervalli omogenei e correttamente comparabili tra loro del campo d'osservazione; in altri casi servirà poco o non servirà affatto. Le medie per righe ci danno misure sintetiche del fenomeno in corrispondenza alle varie misure o modalità dell'una circostanza (la razza del frumento, nel nostro esempio); le medie per colonne ci danno misure sintetiche del fenomeno in corrispondenza alle varie misure o modalità dell'altra circostanza (metodo di coltura, nel nostro esempio).

4. — La media si utilizza anche in altro modo per la rappresentazione della serie di second'ordine, quando i dati della serie indicano il numero dei casi individualmente distinti che corrispondono a ciascuna combinazione di misure di due circostanze quantitative. Si può, allora, per ciascun valore dell'una circostanza calcolare il corrispondente valore medio dell'altra, e viceversa.

Si abbiano, per esempio, i seguenti dati sulla distribuzione di un certo complesso di individui adulti, secondo la statura e il perimetro toracico.

Numero degli individui aventi un dato perimetro toracico e una data statura.

Statura (centimetri)	Perimetro toracico (centimetri)					Totale
	83	84	85	86	87	
163	139	183	210	213	202	947
164	128	173	210	211	204	926
165	132	161	204	207	213	917
166	102	142	172	183	182	781
167	84	118	147	162	168	679
TOTALE	585	777	943	976	969	4.250

Nè la media delle stature (cm. 164,84, media aritmetica) nè quella dei perimetri toracici (cm. 85,23), separatamente considerate, possono riguardarsi espressioni sintetiche della serie di second'ordine. L'una riassume la serie di prim'ordine costituita dalle stature, l'altra riassume la serie di prim'ordine costituita dai perimetri toracici: se le poniamo accanto vediamo bensì qual rapporto di grandezza esista tra la misura media dell'uno e dell'altro carattere individuale, ma non vediamo se e come tenda a variare l'uno al variare dell'altro: perdiamo cioè il vantaggio dell'ordinamento di second'ordine. Conserviamo tale vantaggio, e nel tempo stesso semplifichiamo la serie di second'ordine riducendola a serie di primo ordine, se da essa ricaviamo l'una o l'altra delle due serie seguenti, che ci indicano rispettivamente la misura media aritmetica del perimetro toracico corrispondente a ciascuna statura e la misura media aritmetica della statura corrispondente a ciascun perimetro toracico.

Torace in relazione alla statura		Statura in relazione al torace	
Statura	Perimetro toracico medio	Perimetro toracico	Statura media
163	85,16	83	164,77
164	85,21	84	164,79
165	85,23	85	164,83
166	85,26	86	164,87
167	85,31	87	164,91

La prima serie ci mostra che al crescere della statura corrisponde un incremento, più moderato, del perimetro toracico; la seconda ci mostra che al crescere del perimetro toracico corrisponde un incremento, più moderato, della statura (conclusioni che non sono contraddittorie, come pare a primo aspetto, anzi si confermano a vicenda). La prima serie ci permetterebbe (se fosse completa, invece che limitata ad un breve intervallo di stature) di rappresentare in funzione della statura il corrispondente valore medio del perimetro toracico; la seconda ci permetterebbe di rappresentare in funzione del perimetro toracico il corrispondente valore medio della statura: entrambe le rappresentazioni concorrerebbero a descriverci la relazione esistente fra i due caratteri.

Accanto alle medie ora calcolate potremmo porre gli scostamenti medi corrispondenti: per esempio alla statura di cm. 163 corrisponde la misura media aritmetica del perimetro toracico di cm. 85,16, con uno scostamento medio assoluto di cm. 1,16, cioè di circa 1,36 °/₀. Lo scostamento medio misura l'approssimazione con cui il perimetro toracico medio rappresenta i vari perimetri corrispondenti a quell'unica statura: approssimazione nel nostro caso abbastanza buona.

5. — Il procedimento ora esemplificato si può applicare anche quando i dati della serie indicano il numero dei casi individualmente distinti che corrispondono a ciascuna combinazione della misura di una circostanza quantitativa con la modalità di una circostanza qualitativa. In questo caso si può calcolare il valore medio

della circostanza quantitativa che corrisponde a ciascuna modalità della circostanza qualitativa.

Diamo un esempio. Ciascuno dei dodici appezzamenti considerati nel paragrafo 3 era diviso in dieci campi della superficie di un ettaro: per ognuno di questi campi è stato rilevato separatamente il raccolto del grano. La seguente tabella indica come si distribuiscano i 120 campi secondo la semente usata e la quantità di grano raccolta. (Per risparmio di spazio, abbiamo diviso a mezzo la tabella ed abbiamo affiancato le due metà).

Numero dei campi di un ettaro che hanno dato un determinato raccolto con l'uso di una determinata semente.

Raccolto (quintali)	Razza di frumento			Raccolto (quintali)	Razza di frumento		
	α	β	γ		α	β	γ
16	2	—	—	26	—	3	5
17	4	—	—	27	—	6	3
18	3	1	—	28	—	7	5
19	4	—	—	29	—	4	6
20	9	1	1	30	—	2	4
21	4	2	—	31	—	2	4
22	7	1	1	32	—	—	3
23	1	3	2	33	—	1	3
24	5	5	—	34	—	—	1
25	1	2	2	35	—	—	—

Calcolando per colonne le medie aritmetiche dei raccolti ottenuti sui 40 campi coltivati con ciascuna razza, troviamo quintali 20,4 per la razza α, 26,2 per la β, 28,4 per la γ (dati che già risultavano dal calcolo del paragrafo 4). Ma qui possiamo inoltre vedere come variano da campo a campo i risultati ottenuti con ciascuna razza di frumento; lo scostamento medio assoluto è di quintali 1,94 per la α, 1,32 per la β, 2,53 per la γ: la disuguaglianza fra i risultati è massima per la terza razza e minima per la seconda.

6. — Al procedimento sintetico delle medie si contrappone il procedimento analitico dei rapporti indici (di solito espressi poi nella forma di numeri indici), mediante il quale si surrogano ai dati della serie di second'ordine dati ad essi proporzionali, meglio adatti a mostrare a colpo d'occhio le deviazioni del fenomeno da

un determinato livello di riferimento, in corrispondenza alle varie combinazioni di modalità o misure delle due circostanze con riguardo alle quali il fenomeno è osservato. In generale i numeri indici si possono utilmente calcolare soltanto in quei casi nei quali è logicamente ammissibile la determinazione di una media dei dati della serie.

Per esempio, riducendo a numeri indici i dati della prima tabella del paragrafo 2, assunta come riferimento la media aritmetica dei raccolti ottenuti sui dodici appezzamenti, troviamo:

Numeri indici dei raccolti ottenuti su diversi appezzamenti di dieci ettari.
(Raccolto medio aritmetico = 100)

Razza di frumento	Sistema di coltura			
	A	*B*	*C*	*D*
α	75	70	85	96
β	97	100	103	118
γ	110	88	120	137

I numeri indici giovano a mostrare subito lo scostamento relativo dei raccolti ottenuti coi vari sistemi, dalla media generale.

Si potrebbero calcolare anche numeri indici per righe, per comparare più comodamente i risultati dei vari sistemi di coltura, e numeri indici per colonne per comparare i risultati delle varie sementi.

7. — Un altro procedimento analitico, specialmente adatto a semplificare le comparazioni tra più tabelle di second'ordine, consiste nella surrogazione di rapporti di composizione ai dati della serie di second'ordine. Questo procedimento in generale può essere utilmente applicato quando i dati della serie di second'ordine rappresentano altrettante parti di un sol tutto. Ogni rapporto ci dice qual frazione di questo tutto corrisponda ad una data combinazione di modalità o misure delle due circostanze secondo le quali il tutto è stato suddiviso in parti. Per esempio, traducendo in rapporti di composizione i dati della tabella riportata alla fine del paragrafo 1 (dopo averli riassunti in modo da distinguere soltanto le prime nozze dalle successive), si ottengono le seguenti proporzioni.

Matrimoni avvenuti in Italia nel 1928, per combinazioni di stato civile.
(Proporzioni a 10.000 matrimoni)

Stato civile dello sposo	Stato civile della sposa		Totale
	Nubile	Divorziata o vedova	
Celibe	9.042	173	9.215
Divòrziato o vedovo . .	557	228	785
TOTALE . .	9.599	401	10.000

8. — Quando i dati della serie di second'ordine si possono riguardare come parti di un unico insieme, ciascuna riga della tabella mostra come si distribuisca il fenomeno osservato, secondo le modalità e misure di una circostanza, in corrispondenza ad una determinata modalità o misura dell'altra circostanza. Possiamo desumere dai dati di ciascuna riga rapporti di composizione aventi per denominatore il totale della riga stessa : la somma di questi rapporti risulterà, in ciascuna riga, uguale all'unità. Eseguendo un simile calcolo per tutte le righe della tabella potremo verificare se e come varii la distribuzione delle modalità o misure dell'una circostanza col variare della modalità o misura dell'altra. Potremo per esempio, vedere come varii la distribuzione per età delle spose col variare dell'età dello sposo.

Analoga elaborazione possiamo eseguire per colonne, invece che per righe: vedremo come varii la distribuzione delle modalità o misure dell'altra circostanza col variare della modalità o misura dell'una (per esempio, come varii la distribuzione per età degli sposi col variare dell'età della sposa).

Questa semplice elaborazione della tabella di second'ordine basta in certi casi a presentarci nella forma più adatta tutti gli elementi di giudizio che la tabella stessa fornisce intorno al fenomeno rappresentato. Lo mostreremo con un esempio.

Alla vigilia dello scoppio di un'epidemia in un reggimento era stato accertato quanti soldati non fossero stati mai vaccinati, quanti fossero stati vaccinati da più di due anni, quanti fossero stati vaccinati da non più di due anni. È stato più tardi accertato quanti dègli stessi soldati siano stati colpiti dalla malattia epidemica e ne siano morti, quanti ammalatisi siano guariti, quanti siano rimasti immuni. Ecco i dati.

Distribuzione osservata.

Condizione rispetto alla vaccinazione	Comportamento di fronte all'epidemia			Totale
	Colpiti e morti	Colpiti e guariti	Non colpiti	
Non vaccinati	95	307	84	486
Vaccinati da oltre 2 anni. . .	61	239	1.197	1.497
Vaccinati da non oltre 2 anni .	8	90	943	1.041
TOTALE . . .	164	636	2.224	3.024

Calcoliamo dai precedenti dati rapporti di composizione, riga per riga, e traduciamoli, per comodità di scrittura, in proporzioni a 1000.

Distribuzione proporzionale a 1000 individui in ciascun gruppo.

Condizione rispetto alla vaccinazione	Comportamento di fronte all'epidemia			Totale
	Colpiti e morti	Colpiti e guariti	Non colpiti	
Non vaccinati	196	631	173	1.000
Vaccinati da oltre 2 anni . .	41	160	799	1.000
Vaccinati da non oltre 2 anni .	8	86	906	1.000

La distribuzione dei soldati secondo il loro comportamento durante l'epidemia è molto diversa nei vari gruppi formati secondo la condizione rispetto alla vaccinazione. La proporzione dei soldati che sono rimasti immuni dall'epidemia è massima nel gruppo dei vaccinati da minor tempo, ancora molto alta in quello dei vaccinati da maggior tempo, minima in quello dei non vaccinati. Andamento opposto presenta la proporzione dei morti.

Se le proporzioni fossero risultate ordinatamente uguali tra loro nelle tre righe della nostra tabellina, se cioè tra i non vaccinati, tra i vaccinati da più di due anni, tra i vaccinati da non più di due anni, avessimo riscontrato uguali proporzioni di colpiti e morti, uguali proporzioni di colpiti e guariti, uguali proporzioni di non colpiti, avremmo conchiuso che la condizione rispetto alla vaccinazione non sembra influire in alcun modo sul comportamento di fronte all'epidemia. Ma in realtà le proporzioni sono differenti nelle tre righe, e perciò siamo condotti alla conclusione opposta:

a quella, cioè, che la condizione rispetto alla vaccinazione influisca sul comportamento di fronte all'epidemia. Quanto più differiscono le proporzioni, tanto più grande ci appare quest'influenza: se, per esempio, la proporzione dei colpiti e morti fosse di 500 per 1000 tra i non vaccinati e di 5 per 1000 tra i vaccinati da non più di due anni, l'influenza ci apparirebbe maggiore di quanto risulti dal confronto delle proporzioni di 196 e di 8 che ci dà la nostra tabellina.

Si presenta ovvia l'idea di ricavare dalle differenze tra le proporzioni delle varie righe un indice dell'influenza dell'una circostanza (condizione rispetto alla vaccinazione) sull'altra (comportamento di fronte all'epidemia), poichè quanto maggiore è tale influenza tanto maggiori saranno quelle differenze. Si può giungere ad un siffatto indice per diverse vie: una delle più semplici è la seguente. 1° Calcolare la media aritmetica delle proporzioni, colonna per colonna: nel nostro esempio, rispettivamente: 82 (morti), 292 (guariti), 626 (immuni). 2° Calcolare lo scostamento medio assoluto, colonna per colonna, delle varie proporzioni dalla loro media: nel nostro esempio, rispettivamente 76, 226, 302. 3° Mettere in rapporto la somma degli scostamenti assoluti (604) con la somma delle medie (1.000). Nel nostro esempio il valore del rapporto così calcolato risulta 0,604.

In una serie costituita da n termini il rapporto tra lo scostamento medio assoluto dalla media aritmetica e la media stessa non può superare, come abbiamo altrove avvertito, il valore $2 - (2 : n)$. Perciò il rapporto calcolato nel modo di cui abbiamo dato ora un esempio, il quale può riguardarsi come una media ponderata di rapporti di tal sorta, non può neppur esso superare il medesimo limite, essendo n il numero delle righe della tabella di second'ordine (esclusa la riga dei totali). Essendo proporzionale alla somma dei valori assoluti delle differenze, che sono tanto più grandi quanto più stretta è la relazione tra le due circostanze, il valore del rapporto stesso è adatto a servire come *indice della relazione* esistente fra le due circostanze secondo le quali è stato eseguito l'ordinamento di second'ordine. Diciamo «indice di relazione», e non «indice dell'influenza dell'una circostanza sull'altra». Il nostro esempio porterebbe a preferire, come più precisa, quest'ultima denominazione; ma diamo la preferenza alla prima denominazione, più generale, perchè meglio risponde alla varietà delle relazioni che possono esistere tra le due circostanze: varietà che più avanti metteremo in risalto.

Vogliamo subito avvertire che indici di relazione come quello dianzi calcolato, o come quelli che possono ottenersi con altri procedimenti, non sono che mezzi grossolani, idonei ad avvertire che esiste una relazione tra le due circostanze e a dar un indizio, difficilmente interpretabile, della strettezza o intensità di tale relazione. Quanto c'insegna la nostra tabellina non si può racchiudere in un solo indice numerico: essa ci mostra che la proporzione dei colpiti dalla malattia epidemica discende da 827 per 1000 fra i non vaccinati a 201 per 1000 tra i vaccinati da oltre due anni ed a 94 per 1000 tra i vaccinati da non oltre due anni: ci indica così le differenze di morbosità fra i tre gruppi; ci mostra poi che la proporzione dei morti discende da 196 per 1000 nel primo gruppo a 41 per 1000 nel secondo e ad 8 per 1000 nel terzo: proporzioni che per se medesime misurano la mortalità derivata dall'epidemia nei tre gruppi, e che messe in rapporto con quelle dei colpiti ci mostrano la diversa letalità del morbo nei tre gruppi. Un indice unico non può riassumere in sè tante diverse indicazioni: esso ci fa supporre l'esistenza di relazioni: quali siano queste converrà poi indagare attraverso l'analisi dei dati ed esprimere con una serie di giudizi e di misure.

9. — Esaminiamo un altro esempio. Anche in questo i dati della tabella di second'ordine rappresentano tante parti di un unico tutto.

In una università è stato registrato, per ciascuno studente giunto alla laurea, il risultato delle prove scientifiche e quello delle prove sportive superate. Riassumendo, per semplicità, in tre soli gruppi i risultati conseguiti in ciascuno dei due campi, otteniamo la seguente tabella.

Distribuzione osservata.

Risultati scientifici	Risultati sportivi			Totale
	Buoni	Mediocri	Cattivi	
Buoni	625	315	233	1.173
Mediocri	867	751	419	2.037
Cattivi	131	612	507	1.250
TOTALE . .	1.623	1.678	1.159	4.460

Calcoliamo anche in questo caso i rapporti di composizione riga per riga, e traduciamoli in proporzioni a 1000:

Risultati scientifici	Risultati sportivi			Totale
	Buoni	Mediocri	Cattivi	
Buoni	533	268	199	1.000
Mediocri	425	369	206	1.000
Cattivi	105	490	405	1.000

Queste proporzioni ci permettono di confrontare efficacemente tra loro i risultati sportivi degli studenti che nelle prove scientifiche hanno avuto risultati rispettivamente buoni, mediocri, cattivi. Se tra i risultati dei due ordini di prove non esistesse alcuna relazione, dovremmo trovare proporzioni uguali nelle diverse righe della seconda tabellina. Troviamo proporzioni differenti e ne induciamo che esiste relazione fra i risultati delle due prove.

Ma il concetto di « relazione » indica qui una forma di collegamento fra le due circostanze differente da quella rilevata nell'esempio precedente. La condizione rispetto alla vaccinazione costituiva un antecedente al comportamento di fronte alla malattia: antecedente nel senso cronologico e antecedente anche nel senso logico, perchè l'una circostanza poteva influire a modificare l'altra, e non viceversa. Qui, invece, i risultati scientifici non possono riguardarsi come antecedenti dei risultati sportivi: trovando una certa analogia nella distribuzione degli uni e degli altri fra gli individui osservati siamo portati a supporre che determinati requisiti fisici, intellettuali e morali concorrano a far eccellere l'individuo nell'uno e nell'altro campo, che altri requisiti concorrano a mantenerlo oscuro nell'uno e nell'altro campo. Se invece di trovare analogia avessimo trovato contrasto tra le due distribuzioni, avremmo conchiuso che i requisiti favorevoli alla riuscita nell'un campo sono spesso sfavorevoli nell'altro campo.

Non colleghiamo le due circostanze tra loro come antecedente e conseguente: le colleghiamo piuttosto come conseguenti entrambe di antecedenti in parte comuni. Perciò sarà legittimo calcolare, accanto alla seconda tabella sopra riferita, l'altra che qui sotto riportiamo, nella quale sono raccolte proporzioni a 1000 desunte dai rapporti di composizione calcolati colonna per colonna (invece che riga per riga).

Risultati scientifici	Risultati sportivi		
	Buoni	Mediocri	Cattivi
Buoni	385	188	201
Mediocri	534	447	362
Cattivi.	81	365	437
TOTALE . . .	1.000	1.000	1.000

Mediante queste proporzioni confrontiamo tra loro i risultati scientifici degli studenti che nelle prove sportive hanno avuto risultati rispettivamente buoni, mediocri, cattivi. Non è una semplice ripetizione del confronto fatto dianzi, poichè nella seconda tabella abbiamo comparato i risultati sportivi di gruppi ugualmente numerosi di studenti aventi conseguito diversi risultati scientifici ; in questa terza tabella compariamo i risultati scientifici di gruppi ugualmente numerosi di studenti aventi conseguito diversi risultati sportivi. I due confronti ci permettono di rispondere a due quesiti differenti, sebbene collegati fra loro : primo : « come riescono nello sport i giovani rispettivamente buoni, mediocri, scadenti nella scienza ? » ; secondo : « come riescono nella scienza i giovani rispettivamente buoni, mediocri, scadenti nello sport ? ». Le risposte sono in parte differenti : per esempio, su 1000 buoni nella scienza 533 sono buoni anche nello sport, ma su 1000 buoni nello sport soltanto 385 sono buoni anche nella scienza. Appunto perchè la composizione del complesso osservato cui si riferisce la tabella seconda (1000 buoni, 1000 mediocri, 1000 cattivi *scientificamente*) è diversa da quella del complesso cui si riferisce la tabella terza (1000 buoni, 1000 mediocri, 1000 cattivi *sportivamente*), se cerchiamo un indice della relazione tra le due circostanze — risultati scientifici e risultati sportivi — questo ci risulterà normalmente differente nel calcolo dalle due tabelline. Col procedimento usato nel paragrafo precedente otteniamo dalla seconda tabellina un indice di 0,333, dalla terza (eseguendo, naturalmente, i calcoli per righe invece che per colonne) un indice di 0, 284. Questi indici ci denunziano l'esistenza di una relazione tra le due circostanze, ma soltanto l'esame analitico delle tabelline seconda e terza ci permette di renderci conto della natura e dell'intensità di tale relazione.

10. — Partendo dalla tabella di dati originari elaborati nel pre-

cedente paragrafo avremmo potuto ragionare anche nel modo seguente.

Assunto come dato di fatto quello che fra i 4.460 studenti si hanno 1.173 buoni, 2.037 mediocri, 1.250 cattivi dall'aspetto scientifico, e 1.623 buoni, 1.678 mediocri, 1.159 cattivi dall'aspetto sportivo, quanti studenti corrisponderebbero a ciascuna delle possibili combinazioni dei risultati dei due ordini, se fra tali risultati non vi fosse alcuna relazione?

Si risponde: nell'ipotesi di assenza d'ogni relazione, in ciascuno dei gruppi distinti secondo i risultati sportivi si dovrebbe avere la medesima ripartizione proporzionale degli studenti secondo i risultati scientifici, e in ciascuno dei gruppi distinti secondo i risultati scientifici si dovrebbe avere la medesima ripartizione proporzionale degli studenti secondo i risultati sportivi. E perciò, se si assumono come dati di fatto così la distribuzione dei 4.460 studenti secondo i risultati scientifici, come quella secondo i risultati sportivi, separatamente considerate, si dovrebbero avere nei vari gruppi formati secondo il risultato scientifico proporzioni di studenti sportivamente buoni tutte eguali a 1.623 : 4.460, di sportivamente mediocri tutte uguali a 1.678 : 4.460, di sportivamente cattivi tutte uguali a 1.159 : 4.460; e si dovrebbero avere nei vari gruppi formati secondo il risultato sportivo proporzioni di studenti scientificamente buoni tutte uguali a 1.173 : 4.460, di scientificamente mediocri tutte uguali a 2.037 : 4.460, di scientificamente cattivi tutte uguali a 1.250 : 4.460. Partendo dall'una o dall'altra successione di proporzioni si può calcolare una serie di second'ordine corrispondente alle ipotesi espresse nel secondo comma di questo paragrafo. È indifferente partire dall'una o dall'altra successione, perchè le due condizioni che si pongono per la determinazione del numero x da inserire in una data casella:

x : totale della riga = totale della colonna : totale generale,

x : totale della colonna = totale della riga : totale generale,

si equivalgono.

Ecco la serie di second'ordine così calcolata.

Distribuzione calcolata.

Risultati scientifici	Risultati sportivi			Totale
	Buoni	Mediocri	Cattivi	
Buoni	427	441	305	1.173
Mediocri	741	767	529	2.037
Cattivi	455	470	325	1.250
TOTALE . .	1.623	1.678	1.159	4.460

Per semplicità d'espressione indicheremo come « distribuzione osservata » quella rappresentata nella prima tabella del paragrafo 9, e come « distribuzione calcolata » quella della tabella precedente: si sottintenda: « calcolata per l'ipotesi di assenza di relazione tra le due circostanze, ammessa come dato di fatto la distribuzione osservata di ciascuna di esse ».

Contrapponendo ordinatamente i dati della tabella osservata a quelli della tabella calcolata, possiamo computare le seguenti differenze tra i primi ed i secondi.

Differenze tra dati osservati e dati calcolati.

Risultati scientifici	Risultati sportivi		
	Buoni	Mediocri	Cattivi
Buoni	+ 198	— 126	— 72
Mediocri	+ 126	— 16	— 110
Cattivi.	— 324	+ 142	+ 182

Poichè le somme dei dati delle due tabelle, osservata e calcolata, sono uguali tra loro sia in ciascuna riga sia in ciascuna colonna, le somme delle differenze tra gli uni e gli altri devono risultare nulle sia in ciascuna riga sia in ciascuna colonna.

L'esistenza di differenze è sintomo dell'esistenza di una relazione fra la distribuzione delle due circostanze. La grandezza delle differenze fra i dati osservati e quelli calcolati indica l'intensità della relazione. Naturalmente la grandezza delle differenze dipende, a pari intensità di relazione, dalla grandezza dei numeri che si confrontano: per eliminare quest'ultima influenza si può riferire ciascuna differenza al dato calcolato corrispondente. Ecco i rapporti, espressi in forma di proporzioni a 100.

Differenze relative tra dati osservati e dati calcolati.

Risultati scientifici	Risultati sportivi		
	Buoni	Mediocri	Cattivi
Buoni	+46	—29	—24
Mediocri	+17	— 2	—21
Cattivi	—71	+30	+39

Queste differenze relative mettono in rilievo le relazioni esistenti fra le distribuzioni dei due caratteri. Si vede che la combinazione così di buone attitudini nel campo scientifico e in quello sportivo, come di cattive attitudini in entrambi, è più frequente di quel che sarebbe in assenza di relazione tra i due ordini di attitudini; si vede che i mediocri nella scienza danno una proporzione relativamente alta di buoni risultati nello sport e che invece i mediocri nello sport danno una proporzione relativamente alta di cattivi risultati nella scienza, ecc. Se, rinunziando all'analisi, ci contentiamo soltanto di un indice dell'intensità della relazione fra i due ordini di risultati, possiamo ricavare quest'indice dalle precedenti differenze relative. Le considereremo in valore assoluto, affinchè non avvenga una parziale compensazione fra valori positivi e valori negativi; e ne prenderemo una media ponderata, non una media semplice, per tener conto della varia grandezza dei numeri dai quali quelle differenze sono desunte. Se calcoliamo la media aritmetica ponderata, assegnando come peso a ciascuna differenza relativa il corrispondente dato della distribuzione calcolata, tale media risulta — com'è facile verificare — uguale alla somma dei valori assoluti delle differenze tra dati osservati e dati calcolati (*differenze*, si badi, e non differenze relative: nel nostro esempio, cioè, quelle contenute nella penultima tabella) divisa per la somma dei dati osservati (o di quelli calcolati, che è la stessa). Nel nostro esempio la somma delle differenze, considerate in valore assoluto, è 1.296, la somma dei dati osservati 4.460: il rapporto risulta uguale a 0,291.

Differenze del tipo di quelle della penultima tabella vengono talora denominate *contingenze*; in armonia con questo nome le differenze relative come quella dell'ultima tabella vengono chiamate *contingenze relative*, e la loro media ponderata, calcolata nel modo ora descritto, viene denominata *contingenza media assoluta*. La contingenza media assoluta non può mai raggiungere il valore 2; ed appunto perchè, mentre è proporzionale alla somma dei valori assoluti delle contingenze, varia entro limiti definiti, può servire come indice dell'intensità della relazione fra le due circostanze. Dalla stessa contingenza media assoluta oppure dalla contingenza media quadratica, oppure dalle contingenze relative, si possono facilmente ricavare altri indici di relazione variabili fra 0 e 1.

Noi riteniamo che, almeno nelle applicazioni ai fenomeni sociali, simili indici sintetici abbiano scarsa importanza.

È importante accertare se fra le due circostanze secondo le quali è ordinata la serie di second'ordine esista una relazione: questa relazione, se v'è, richiede in generale di essere studiata particolarmente, e di rado può esprimersi in modo soddisfacente in un sol numero, come ci mostra anche il precedente esempio. Studio particolare vuol dire esame delle singole contingenze, se si adotta il metodo descritto nel presente paragrafo; e poichè le contingenze non son altro che differenze tra due dati statistici, si potranno applicare ad esse le elaborazioni indicate nel capitolo VI.

Cerchiamo, per esempio, di renderci conto del significato dell'eccedenza del dato osservato 625 sul dato calcolato 427, nella combinazione di buoni risultati scientifici e sportivi. L'eccedenza è di 198, e nella tabella delle differenze relative l'abbiamo messa in rapporto col dato calcolato, traducendola in un'eccedenza relativa di 0,46. Ma ci possiamo anche, domandare: qual valore massimo avrebbe potuto raggiungere quest'eccedenza? Conoscendo tale massimo potremo apprezzare in relazione ad esso il significato dell'eccedenza effettiva di 198. Orbene, poichè gli studenti sportivamente buoni sono 1.623 ma quelli scientificamente buoni sono soltanto 1.173, il massimo numero possibile di studenti buoni nei due campi non può superare 1.173. Rispetto al numero calcolato di 427 questo massimo darebbe un'eccedenza di 746; il rapporto 198 : 746 ci dice che l'eccedenza accertata corrisponde al 27 $^{0}/_{0}$ della massima possibile. Abbiamo così calcolato un indice del grado di eccedenza, nel senso attribuito a questa espressione alla pagina 30.

Nella combinazione di cattivi risultati scientifici con buoni risultati sportivi troviamo una deficienza del dato osservato 131 rispetto al dato calcolato 455. Nella tabella delle differenze relative questa deficienza è stata tradotta in una deficienza relativa di 0,71. Ma qual valore massimo avrebbe potuto raggiungere questa deficienza? Poichè gli studenti scientificamente non cattivi sono 3.210 e quelli sportivamente non cattivi 3.301, ciascuno dei 1.250 scientificamente cattivi avrebbe potuto appartenere ai non cattivi sporvamente, e ciascuno dei 1.159 sportivamente cattivi avrebbe potuto appartenere ai non cattivi scientificamente. Alla combinazione di risultati cattivi sia nella scienza sia nello sport avrebbero allora corrisposto 0 studenti; la deficienza avrebbe raggiunto 455. In questo esempio la deficienza relativa (71 %) ci dà nel tempo stesso l'indice del grado di deficienza definito alla pagina 30.

Abbiamo calcolato gli indici del grado di eccedenza, o di deficienza, del dato osservato rispetto al dato calcolato, per tutte le caselle della nostra tabella: riferiamo i risultati, avvertendo che per distinguere il senso delle deviazioni abbiamo premesso il segno + agli indici del grado di eccedenza e il segno — agli indici del grado di deficienza.

Indici del grado di eccedenza o di deficienza.

Risultati scientifici	Risultati sportivi		
	Buoni	Mediocri	Cattivi
Buoni	+ 0,27	— 0,29	— 0,24
Mediocri	+ 0,14	— 0,02	— 0,21
Cattivi	— 0,71	+ 0,18	— 0,22

Nello studio delle combinazioni matrimoniali secondo vari caratteri individuali degli sposi e delle spose, indici del grado di eccedenza e del grado di deficienza calcolati coi criteri ora esposti sono stati ideati ed impiegati dal BENINI, col nome rispettivamente di *indici d'attrazione* e di *indici di repulsione*: nomi specialmente convenienti per quella particolare applicazione.

11. — Col metodo esposto nel paragrafo precedente si confronta la distribuzione effettivamente osservata dei casi di un fenomeno, secondo le combinazioni di modalità o di misure di due circostanze, con la distribuzione che si avrebbe — a quanto si presume — se tra le due circostanze non esistesse alcuna relazione. La caratteristica fondamentale del metodo — per cui esso si distingue dai metodi seguiti nei paragrafi 8 e 9 — sta appunto in questa compa-

razione tra risultati osservati e risultati calcolati per l'ipotesi di assenza di ogni relazione. Che poi tale comparazione si esegua mediante differenze, mediante differenze relative, mediante rapporti; che si riassuma, o non si riassuma, in un unico dato numerico da prendersi ad indice della relazione fra le due circostanze; che tale dato sintetico venga determinato con un criterio o con un altro: tutto ciò ha importanza del tutto accessoria. Sicchè, senza fermarci su questi particolari, crediamo utile esaminare il fondamento logico del confronto tra distribuzione osservata e distribuzione calcolata.

La distribuzione osservata è quella che è: su di essa non v'è ragione di discutere.

Ma la distribuzione calcolata per il caso di assenza d'ogni relazione fra le due circostanze non è un dato di fatto. Nel paragrafo 10 l'abbiamo determinata partendo dalle seguenti ipotesi:

1) che, in assenza di relazione tra le due circostanze, in corrispondenza alle diverse modalità o misure di ciascuna circostanza si avrebbe la stessa distribuzione proporzionale delle modalità o misure dell'altra;

2) che codesta distribuzione proporzionale sarebbe uguale a quella riscontrata nell'insieme dei casi osservati.

La prima di queste ipotesi equivale alla seguente definizione dell'assenza di relazione tra due circostanze: «Non v'è relazione tra due circostanze, con riguardo alle quali è osservato un fenomeno, se al variare della modalità o misura di ciascuna di esse non varia la distribuzione proporzionale delle modalità o misure dell'altra» Per essere rigorosa la definizione dovrebbe enunciare esplicitamente l'ipotesi della parità di ogni altra circostanza estranea alle due considerate: ipotesi che vi è implicita; per essere applicabile ai fenomeni collettivamente tipici dovrebbe contenere, dopo «non varia», la riserva: «salvo che per variazioni non significative» (variazioni che più avanti impareremo a conoscere). Comunque, l'esperienza ha dimostrato che la prima ipotesi, anche se non sempre logicamente inattaccabile, costituisce un utile strumento di lavoro.

La seconda ipotesi è fondata a sua volta sulla prima, poichè per giungere ad essa si ragiona così: «In assenza di relazione fra le due circostanze, la distribuzione proporzionale delle modalità o misure di ciascuna di esse sarebbe la medesima in corrispondenza alle diverse modalità o misure dell'altra, *e coinciderebbe con quella che si ha nel complesso delle osservazioni*». Accettata la prima ipotesi, non si può non accettare la conseguenza di essa, espressa nell'ultima proposizione in corsivo. Ma il ragionamento continua : « Dunque, la distribuzione proporzionale di ciascuna circostanza, che di fatto sussiste nel complesso delle osservazioni, indica la distribuzione che si avrebbe in corrispondenza ad ogni modalità o misura dell'altra, se tra le due circostanze non esistesse alcuna relazione ». Questa seconda parte del ragionamento è priva di base logica, perchè estende alla generalità dei casi quella conseguenza della prima ipotesi che sussiste soltanto in quanto corrisponda al vero l'ipotesi stessa : in quanto, cioè fra le due circostanze non vi sia relazione. Se, invece v'è relazione, la distribuzione proporzionale delle due circostanze nel complesso delle osservazioni, o almeno quella di una di esse, è modificata dalla relazione stessa, e quindi non può rappresentare la distribuzione che si avrebbe in assenza di relazione.

Tale, a nostro parere, è il difetto insanabile del metodo esposto nel paragrafo precedente. Esso non consiste, come vorrebbe, nella comparazione fra la distribuzione effettiva delle combinazioni delle due circostanze e la distribuzione che si avrebbe in assenza di una relazione tra esse; consiste semplicemente nella comparazione fra la distribuzione effettiva e la distribuzione che si avrebbe se la distribuzione proporzionale delle modalità o misure dell'una circostanza in corrispondenza alle diverse modalità o misure dell'altra fosse sempre quella — dipendente anche dalle relazioni eventualmente esistenti fra le due circostanze — che si ha nel complesso delle osservazioni. È un confronto lecito, e talvolta istruttivo, ma non ha il significato che gli si vuol attribuire.

Se la distribuzione che si calcola col procedimento del paragrafo 10 *non è* — come abbiamo dimostrato — quella che si avrebbe in assenza di ogni relazione tra le due circostanze, qual è dunque la distribuzione che si avrebbe in tale ipotesi ? Non v'è un criterio sistematico per determinarla, fuorchè nel caso in cui la distribuzione proporzionale dell'una circostanza che si osserva in corrispondenza ad una determinata modalità o misura dell'altra circostanza appaia essere quella che si dovrebbe avere anche in corrispondenza alle altre modalità o misure di essa se non sussistesse alcuna relazione tra le due circostanze. Fuori di questo caso, in generale non è possibile stabilire quale distribuzione si avrebbe in assenza di relazione fra le due circostanze; e fa d'uopo accontentarsi del confronto tra le distribuzioni proporzionali di ciascuna circostanza che corrispondono alle diverse modalità o misure dell'altra.

12. — Un esempio servirà a chiarire meglio quanto abbiamo esposto dianzi. Proviamo ad applicare il procedimento del paragrafo 10 ad un caso analogo a quello esaminato nel paragrafo 8, ma più semplice. Un complesso d'individui è stato classificato secondo la sua condizione rispetto alla vaccinazione e secondo il suo comportamento di fronte all'epidemia: n'è risultata la seguente distribuzione.

Distribuzione osservata.

Condizione rispetto alla vaccinazione	Comportamento di fronte all'epidemia		Totale
	Colpiti	Non colpiti	
Non vaccinati	270	30	300
Vaccinati	70	630	700
TOTALE . .	340	660	1.000

Si ragiona così: « In assenza di relazione tra la condizione rispetto alla vaccinazione e il comportamento di fronte all'epidemia, la proporzione dei colpiti sarebbe uguale tra i vaccinati e tra i non vaccinati, e altrettanto avverrebbe della proporzione dei non colpiti ». Questa prima parte del ragionamento è accettabile, specialmente se dopo la prima virgola si aggiunga l'inciso: « a parità di ogni altra circostanza », e dopo la seconda virgola si aggiunga l'inciso: « a prescindere da differenze non significative ».

Ma il ragionamento prosegue « E così tra i vaccinati come tra i non vaccinati — nella supposta assenza di relazione — si avrebbe la stessa proporzione di colpiti e la stessa proporzione di non colpiti che si ha di fatto nel complesso dei casi osservati ». Qui il ragionamento non corre più: l'affermazione sarebbe vera soltanto se nella realtà non vi fosse relazione tra le due circostanze; se v'è relazione, l'affermazione è falsa perchè la proporzione dei colpiti nel complesso dei casi osservati — media ponderata delle proporzioni, *differenti tra loro*, che si riscontrano tra i vaccinati ed i non vaccinati — dipende: primo: da tali proporzioni; secondo: dai « pesi » coi quali i due gruppi concorrono a costituire il complesso osservato.

Il ragionamento, dunque, è sbagliato. Nel nostro esempio, esso ci porterebbe ad assumere che in assenza di relazione fra le due circostanze si avrebbe così fra i vaccinati come tra i non vaccinati una proporzione di colpiti del 34°/₀ ed una proporzione di non colpiti del 66°/₀. Ne risulterebbe la seguente distribuzione.

Distribuzione calcolata (I).

Condizione rispetto alla vaccinazione	Comportamento di fronte all'epidemia		Totale
	Colpiti	Non colpiti	
Non vaccinati	102	198	300
Vaccinati	238	462	700
TOTALE . .	340	660	1.000

Contrapponendo la distribuzione calcolata a quella osservata potremo calcolare le contingenze; la somma dei valori assoluti di esse è 672, e poichè i casi osservati sono 1000 la contingenza media assoluta risulta uguale a 0,672.

Nel nostro esempio la proporzione dei colpiti è uguale a 90°/₀ fra i non vaccinati, a 10°/₀ fra i vaccinati. Se, mantenendo fisse queste proporzioni (mantenendo cioè immutata la relazione tra le due circostanze: condizione rispetto alla vaccinazione e comportamento di fronte all'epidemia), modifichiamo le proporzioni nelle quali i due gruppi concorrono a formare il complesso osservato, vediamo anche variare la proporzione media generale dei colpiti nel complesso stesso, che si assume come quella che si avrebbe in assenza di relazioni tra le due circostanze. In un complesso di 700 non vaccinati e 300 vaccinati troveremo una proporzione media di colpiti del 66°/₀; in uno di 900 non vaccinati e 100 vaccinati, una proporzione media dell'82°/₀; in uno di 100 non vaccinati e 900 vaccinati una proporzione media del 18°/₀; in uno di 500 non vaccinati e 500 vaccinati una proporzione media del 50°/₀. La contingenza media assoluta risulta uguale a 0,288 nella ripartizione di 900 a 100, o di 100 a 900; a 0,672 in quella di 700 a 300, o di 300 a 700; di 0,80 in quella di 500 a 500: varia, cioè, fortemente l'indice di una relazione che in realtà si mantiene costante. E può variare ancor più: se consideriamo 999.900 non vaccinati e 100 vaccinati, la proporzione media dei colpiti sale a 89,992°/₀ e la contingenza media assoluta si riduce a 0,00036. Il metodo appare dunque assolutamente fallace, poichè della medesima relazione fornisce indici tanto discordanti. L'applicazione degli indici del grado di eccedenza e di deficienza (metodo di BENINI) può attenuare, ma non eliminare, l'inconveniente.

In questo esempio è invece possibile e facile paragonare correttamente la distribuzione osservata con una distribuzione calcolata rappresentante la *vera* distribuzione che si avrebbe in assenza di relazione. Infatti, se tra la condizione rispetto alla vaccinazione e il comportamento rispetto all'epidemia non esistesse alcuna relazione, dovrebbe aversi tra i vaccinati (a parità di ogni altra circostanza, e a prescindere dalle variazioni non significative) la stessa proporzione di colpiti che si ha tra i vaccinati. (Ragioniamo sicuramente così perchè sappiamo che la relazione eventualmente esistente consisterebbe in una influenza della vaccinazione sulla proporzione dei colpiti). Con questo più semplice e più corretto criterio possiamo ora stabilire la distribuzione che si avrebbe in assenza di relazioni tra le due circostanze.

Distribuzione calcolata (II).

Condizione rispetto alla vaccinazione	Comportamento di fronte all'epidemia		Totale
	Colpiti	Non colpiti	
Non vaccinati	270	30	300
Vaccinati	630	70	700
TOTALE . .	900	100	1.000

Il confronto tra questa distribuzione calcolata e quella osservata indica che il numero osservato dei colpiti tra i vaccinati (70) è inferiore di 560 a quello calcolato (630): che cioè la vaccinazione evita l'89°/₀ dei casi di malattia. Comunque si facciano variare le proporzioni dei due gruppi (vaccinati e non vaccinati), mantenendo ferme nell'uno e nell'altro le proporzioni dei colpiti, si giungerà sempre alla stessa espressione numerica della relazione tra le due circostanze considerate: espressione che non è un *indice* ma una vera e propria *misura*.

13. — Nel precedente paragrafo abbiamo potuto, per la speciale natura della relazione intercedente tra le due circostanze, determinare correttamente la distribuzione che si avrebbe nel caso di assenza d'ogni relazione. Ciò è possibile soltanto quando una delle due circostanze sia antecedente logico e cronologico dell'altra e quando ad una delle modalità o misure della prima circostanza

corrisponda una distribuzione proporzionale della seconda che può correttamente assumersi come quella che si avrebbe in assenza di relazione.

Quando non si verificano tali condizioni, è in generale impossibile determinare la distribuzione che si avrebbe in assenza di relazione tra le due circostanze. Per esempio: si conosce la distribuzione degli scioperi avvenuti in un paese secondo la loro causa e il loro esito, congiuntamente considerati. Si vorrebbe confrontarla con la distribuzione che si avrebbe se tra causa ed esito dello sciopero non vi fosse alcuna relazione. Ma quale sia questa distribuzione ipotetica nessuno può dire. Col metodo del paragrafo 10 si ammette che la distribuzione proporzionale degli esiti, nell'ipotesi di indipendenza dell'esito dalla causa dello sciopero, sarebbe in ciascun gruppo, costituito secondo la causa dello sciopero, uguale a quella che si osserva di fatto per il complesso degli scioperi. Ma quest'ultima dipende: primo: dalle distribuzioni proporzionali degli esiti, che si riscontrano nei vari gruppi costituiti secondo la causa: distribuzioni che di fatto differiscono tra loro proprio per l'esistenza di quella relazione che si suppone assente; secondo: dai «pesi» coi quali codesti gruppi entrano a costituire il complesso osservato. Quindi il riferimento è del tutto arbitrario. Conviene, pertanto, limitarsi a confrontare tra loro le varie distribuzioni proporzionali degli esiti che corrispondono ai vari gruppi costituiti secondo le cause dello sciopero: procedere, cioè, come abbiamo fatto nel paragrafo 8.

14. — Abbiamo più volte accennato nelle precedenti pagine, alla varia natura della relazione che può sussistere tra le due circostanze secondo le quali è ordinata una serie di second'ordine. Riteniamo ora utile esporre in proposito alcune considerazioni riassuntive.

Le principali relazioni che possono sussistere tra le due circostanze sono le seguenti:

a) L'una circostanza è antecedente dell'altra, nel senso che opera a modificare la distribuzione delle modalità o misure di essa, mentre non può accadere il reciproco (la condizione rispetto alla vaccinazione può influire sul comportamento dell'individuo di fronte alla malattia, ma il comportamento non può influire sulla vaccinazione; la causa dello sciopero può influire sull'esito, ma l'esito non può influire sulla causa). In questo caso la distribuzione delle modalità o misure della circostanza «conseguente» nel complesso dei casi osservati dipende dalla proporzione in cui compaiono in tale complesso le diverse modalità o misure della circostanza « antecedente » e quindi non si può assumere, senza contraddizione in termini, a rappresentare la distribuzione che si avrebbe se non vi fosse alcuna relazione tra le due circostanze. Conviene limitarsi al confronto tra le distribuzioni delle modalità o misure della circostanza « conseguente » che corrispondono alle diverse modalità o misure della circostanza « antecedente ».

b) Ciascuna delle due circostanze può essere a volta a volta antecedente o conseguente rispetto all'altra, nel senso che il modificarsi dell'una tende a modificare la distribuzione dell'altra e il modificarsi dell'altra tende a modificare la distribuzione dell'una (l'offerta di una merce influisce sul prezzo e il prezzo influisce sull'offerta; la mortalità influisce sulla natalità e la natalità influisce sulla mortalità; l'età dell'uomo che vuol contrarre matrimonio concorre a fargli preferire nella scelta una donna di età compresa entro determinati limiti, ma l'età della donna che viene richiesta in matrimonio concorre a sua volta a farle preferire nell'accettazione dell'offerta un uomo di età compresa entro determinati limiti). Per questo caso si può ripetere l'osservazione esposta sotto *a*); e conviene limitarsi al confronto ivi indicato, assumendo però successivamente come antecedente sia l'una sia l'altra circostanza.

c) Entrambe le circostanze in parte dipendono da antecedenti comuni, i quali influiscono sulla loro distribuzione. Possono distinguersi due casi:

c_1) una delle due circostanze è antecedente dell'altra, o soltanto cronologicamente, od anche logicamente nel senso che concorre a modificarne la distribuzione (statura del padre e statura del figlio dipendono da circostanze in parte comuni, ma l'una è antecedente cronologico dell'altra; professione del padre e professione del figlio dipendono da circostanze in parte comuni, e l'una è antecedente cronologico e logico dell'altra, concorrendo a determinarla). Anche in questo caso si può ripetere l'osservazione esposta sotto *a*);

c_2) non si può considerare alcuna delle due circostanze come antecedente dell'altra (statura e perimetro toracico individuali, attitudini intellettuali e attitudini corporali). In questo caso conviene considerare a volta a volta come antecedente ciascuna delle due circostanze (come abbiamo fatto nel paragrafo 9); inoltre si può tollerare l'applicazione del metodo del paragrafo 10, sebbene neppure qui siano logicamente fondati i suoi presupposti: si può applicarlo come confronto tra una distribuzione osservata e una distribuzione ipotetica, che *non è* quella (non determinabile) che si avrebbe nell'ipotesi di assenza di relazione tra le due circostanze, ma è quella che si avrebbe nell'ipotesi di distribuzione delle due circostanze, separatamente considerate, uguale a quella osservata, e di distribuzione proporzionale delle modalità o misure di ciascuna delle due circostanze uguale in corrispondenza alle diverse modalità dell'altra.

15. — Quanto al modo in cui si presenta la relazione tra le due circostanze secondo le quali è ordinata la serie di second'ordine, notiamo che in generale tale relazione è denunziata dalla diversa distribuzione delle modalità o misure dell'una circostanza in corrispondenza alle diverse modalità o misure dell'altra circostanza, e che non sempre questa diversità di distribuzione si può esprimere in modo semplice e sintetico.

Quando, però, le due circostanze sono omogenee tra loro (professione del padre e professione del figlio) o almeno sono omogenee le modalità nelle quali esse si manifestano (risultati scientifici e risultati sportivi, se rilevati come nel nostro esempio del paragrafo 9), la relazione si può manifestare come un'analogia o come un contrasto, come una concordanza o come una discordanza, tra le distribuzioni delle due circostanze. Nel nostro esempio del paragrafo 9, la massima analogia si avrebbe se tutti i buoni scientificamente fossero anche buoni sportivamente, i cattivi scientificamente fossero cattivi anche sportivamente, e reciprocamente; il massimo contrasto si avrebbe se tutti i buoni scientificamente fossero cattivi sportivamente, tutti i cattivi scientificamente fossero buoni sportivamente e reciprocamente. L'esame della prima tabella del paragrafo 9 mostra che non si verifica nella realtà nè l'una nè l'altra delle due ipotesi. Non è difficile escogitare metodi che in simili casi permettano una misura del grado di analogia o del grado di contrasto tra le due circostanze; spesso però l'applicazione di tali metodi ha l'inconveniente di suscitare l'illusione che le relazioni osservate siano più semplici di quanto sono nella realtà.

Quando entrambe le circostanze sono quantitative può accadere che al crescere dell'una si veda crescere la misura media corrispondente dell'altra (al crescere della statura dei padri si vede crescere la statura media dei figli), o viceversa che al crescere dell'una si veda diminuire la misura media corrispondente dell'altra (al crescere della quantità di stoffa prodotta in una tessitura si vede diminuire il costo medio di produzione per metro di stoffa). Simili casi di *correlazione diretta* e *correlazione inversa* non sono che forme particolari — le più semplici — delle relazioni che possono intercedere tra circostanze quantitative: relazioni sulle quali ci intratterremo nel capitolo XX.

16. — I metodi per la comparazione tra una serie di second'ordine quale risulta dall'osservazione e quale si presume che sarebbe in assenza di relazione fra le due circostanze non costituiscono che un'applicazione particolare dei metodi adatti per la comparazione fra due serie di second'ordine omogenee tra loro per la natura dei dati, ordinate ugualmente in modo che a ciascun termine dell'una corrisponda un termine dell'altra, ed aventi ugual somma di termini.

Il procedimento comparativo si può riassumere schematicamente così:

1) I dati corrispondenti delle due serie vengono comparati fra loro; cioè casella per casella si calcola la loro differenza o il loro rapporto. L'esame di queste differenze o di questi rapporti riesce già per se medesimo istruttivo.

2) Se si sono calcolate differenze, si possono desumere da queste le differenze relative, utili per l'ulteriore sviluppo della comparazione. (Se si sono calcolati rapporti si possono ugualmente desumere da essi le differenze relative, sottraendo l'unità dal valore del rapporto).

3) Sommando i valori assoluti delle differenze e dividendo poi la loro somma per la somma dei dati dell'una o dell'altra serie (si ricordi che queste due ultime somme sono uguali fra loro), si ottiene un'unica misura della disuguaglianza fra i termini corrispondenti delle due serie, che in certi casi costituisce una sufficiente espressione sintetica dei risultati della comparazione. (La stessa misura può essere riguardata come una media ponderata dei valori assoluti delle differenze relative).

4) Il rapporto fra la misura della disuguaglianza così ottenuta e il massimo valore che essa avrebbe potuto assumere (massimo che risulta diverso secondo che si pongono o non si pongono vincoli alla distribuzione ipotetica dei dati nelle due serie) dà un indice del grado di disuguaglianza: altra espressione sintetica, talora utile, dei risultati della comparazione.

5) Le differenze pertinenti alle singole caselle possono venire comparate coi massimi valori assoluti che avrebbero potuto assumere in date ipotesi; si ottengono così indici del grado di eccedenza o

indici del grado di deficienza, secondo che si tratta di differenze positive o negative.

6) Dai valori assoluti di questi indici, per media semplice o ponderata, si può ricavare un unico indice del grado di disuguaglianza tra le due serie, che secondo il metodo di calcolo potrà coincidere o non coincidere con quello calcolato nel modo esposto sotto 4.

Alcune tappe del procedimento consentono varianti: per esempio invece che sui valori assoluti delle differenze si può operare sui loro quadrati; l'indice del grado di disuguaglianza si può ottenere con diversi criteri: nel calcolo di medie si può adottare l'una o l'altra media, ecc. Ma questi particolari, se possono influire sensibilmente sui risultati numerici, non influiscono a modificare la sostanza del procedimento e quindi il significato dei risultati.

Varia, invece, grandemente il significato della comparazione col variare della serie che viene assunta come termine di comparazione per quella osservata. Nei precedenti paragrafi abbiamo avuto occasione di eseguire simili comparazioni con riferimenti diversi; riprendendo uno degli esempi svolti possiamo ancor meglio vedere la molteplicità dei possibili riferimenti, senza certamente esaurire le comparazioni immaginabili.

Disponendo della distribuzione osservata di un complesso di studenti secondo le combinazioni dei risultati delle loro prove scientifiche e dei risultati delle loro prove sportive possiamo compararla:

con la distribuzione calcolata per l'ipotesi che gli studenti si suddividano in parti uguali fra le varie combinazioni; questa comparazione ha lo scopo di mettere in rilievo la disuguaglianza esistente nella distribuzione delle due circostanze combinate;

con la distribuzione calcolata per l'ipotesi che tutti gli studenti rispettivamente buoni, mediocri, cattivi nella scienza siano anche buoni, mediocri, cattivi nello sport; questa comparazione mira ad accertare la divergenza della distribuzione delle attitudini sportive da quella delle attitudini scientifiche;

con la distribuzione calcolata per l'ipotesi che tutti gli studenti rispettivamente buoni, mediocri, cattivi nello sport, siano anche buoni, mediocri, cattivi nella scienza; questa comparazione mira ad accertare la divergenza della distribuzione delle attitudini scientifiche da quella delle attitudini sportive;

con la distribuzione calcolata per l'ipotesi che nei diversi gruppi formati secondo i risultati scientifici [sportivi] la distribuzione proporzionale degli studenti secondo i risultati sportivi [scientifici] sia quella che si riscontra in uno determinato dei gruppi stessi; con questo confronto si vuol porre in luce la diversa distribuzione delle attitudini sportive [scientifiche] nei diversi gruppi formati secondo le attitudini scientifiche [sportive]. Secondo il fine dell'indagine, si assumerà come termine di comparazione la distribuzione riscontrata in uno od in un altro determinato gruppo;

con la distribuzione calcolata nell'ipotesi che la distribuzione proporzionale per colonne nelle singole righe e la distribuzione proporzionale per righe nelle singole colonne, sia uguale a quella che si riscontra nell'intero complesso osservato; è questa la comparazione eseguita nel paragrafo 10; come abbiamo dimostrato, la distribuzione così calcolata *non è* quella che si avrebbe in assenza di relazione tra le due circostanze, tuttavia in certi casi può essere assunta per termine di comparazione, come corrispondente a certi requisiti di uniformità;

con la distribuzione osservata delle combinazioni dei risultati delle prove scientifiche e delle prove sportive di un altro complesso di individui (p. es. degli studenti di un'altra università);

con la distribuzione osservata delle combinazioni dei risultati delle prove scientifiche degli stessi studenti e dei risultati di altre prove diverse da quelle sportive: per esempio prove delle doti morali, che potrebbero dar luogo anch'esse ad una graduazione in buoni, mediocri, cattivi.

Non continuiamo l'enumerazione, perchè non vogliamo approfondire un'applicazione particolare, ma soltanto mostrare da quanto numerosi punti di vista si possa procedere in comparazioni come quelle che qui studiamo. Gli indici numerici dell'affinità o del contrasto fra le due distribuzioni che si confrontano possono moltiplicarsi quasi all'infinito col moltiplicarsi dei possibili criteri di comparazione e dei procedimenti di computo degli indici; e si sono effettivamente assai moltiplicati nella statistica moderna. Nella maggior parte dei casi questi indici, come abbiamo già per taluno di essi avvertito, sono insufficienti ad esprimere la complessità delle conclusioni che si possono ricavare dalla comparazione onde furono ricavati; e se lo statistico esperto può aver qualche vantaggio nell'impiegarli con le debite cautele in un primo esame dei fenomeni, il principiante, cui è destinato questo libro, farà bene in generale ad astenersi dall'usarli, preferendo la comparazione analitica ad una comparazione sintetica, spesso d'interpretazione dubbia e difficile.

Indicazioni bibliografiche. — Intorno agli argomenti trattati nei capitoli XVII-XX vedansi specialmente i già citati manuali di BENINI, GINI, BOWLEY, YULE e il recente manuale di MARCH L., *Les principes de la méthode statistique*, Paris, Alcan, 1930. — Vedansi anche gli ampi studi monografici del GINI: specialmente *Variabilità e mutabilità*, Bologna, Cuppini, 1912; *Nuovi contributi alla teoria delle relazioni statistiche; Indici di omofilia, ecc.*; *Di una misura della dissomiglianza tra due gruppi di quantità, ecc.* (questi tre studi sono inseriti negli *Atti del Reale Istituto Veneto*, Anno accademico 1914-15, Tomo LXXIV, parte seconda); *Di una misura delle relazioni tra le graduatorie di due caratteri*, Roma, Cecchini, 1914.

Quesiti ed esercizi: 1. — Si definisca la serie statistica di secondo, di terzo, di quarto ordine e si corredi la definizione con qualche esempio.

2. — Che cos'è una tabella di second'ordine? Quante caselle contiene? Quali e quante serie di prim'ordine si possono desumere da una serie di second'ordine? Come compaiono, in generale, queste serie di prim'ordine nella tabella di second'ordine? Quali e quante serie di second'ordine si possono desumere da una serie di quart'ordine (per esempio da quella dei matrimoni aggruppati secondo l'età e lo stato civile dello sposo e della sposa).

3. — Quando e come si possono adoperare le medie per la rappresentazione sintetica della serie di second'ordine? Si può sostituire un'unica media alla serie di second'ordine? Qual e il significato delle medie delle singole righe? di quelle delle singole colonne?

4. — In qual modo si può rappresentare per mezzo di medie la tabella dei matrimoni secondo l'età dello sposo e l'età della sposa, contenuta nell'ASI? Si provi ad eseguire tale rappresentazione, adottando opportune ipotesi sull'età media degli individui compresi in ciascun intervallo di più anni d'età. Si traducano in diagramma i risultati ottenuti e si cerchi di interpretare il significato del diagramma.

5. — Si rappresentino mediante diagrammi i dati della tabella riferita nel paragrafo 5, in modo da poter confrontare graficamente tra loro i dati delle tre colonne.

6. — Si traggano dai dati dell'ASI per le provincie italiane tabelle di second'ordine, che indichino la distribuzione delle provincie stesse: *a*) secondo la densità della popolazione e la mortalità; *b*) secondo la nuzialità (matrimoni per 1000 abitanti) e la natalità (nati vivi per 1000 abitanti); *c*) secondo la natalità e la mortalità; *d*) secondo la natalità e l'eccedenza dei nati sui morti per 1000 abitanti; *e*) secondo il rendimento per ettaro nella coltura del frumento in due anni differenti.

7. — Si traggano dai dati dell'ASI sui prezzi medi mensili dei prodotti agricoli per un quinquennio tabelle di second'ordine, che indichino la distribuzione dei 60 mesi d'osservazione secondo i prezzi: *a*) del frumento tenero e del frumento duro; *b*) del frumento tenero e del granturco. Converrà raggruppare i prezzi per intervalli di 2 o di 5 lire.

8. — Si prendano in esame, eseguendo le opportune elaborazioni, le tabelle compilate negli esercizi 6 e 7 e si espongano le conclusioni che tale esame suggerisce.

9. — In un esperimento di coltivazione del grano compiuto sugli stessi appezzamenti di 10 ettari ciascuno considerati nel paragrafo 3 del testo, sono state somministrate diverse dosi di concimi chimici ai diversi appezzamenti. La seguente tabella indica i risultati ottenuti.

Raccolti ottenuti su appezzamenti di dieci ettari con diverse dosi di concimi fosfatici ed azotati.

(quintali)

Dose di concimi fosfatici (quintali)	Dose di concimi azotati (quintali)			
	3	4	5	6
6	281	304	317	322
7	309	312	331	344
8	325	338	357	360

Si confrontino i singoli risultati col risultato medio generale. Si confrontino tra loro i risultati di ciascuna riga e quelli di ciascuna colonna e si espongano le conclusioni desunte dal confronto.

Sapendo che un undicesimo appezzamento coltivato senza concimazione chimica ha dato un raccolto di 157 quintali su 10 ettari, come si possono integrare le precedenti conclusioni?

10. — Si traduca in numeri indici la tabella del precedente esercizio assumendo successivamente come riferimento: *a*) la media generale; *b*) le medie per righe; *c*) le medie per colonne; *d*) il raccolto ottenuto nell'appezzamento non concimato. Si esaminino e si commentino i numeri indici.

11. — Si traducano in rapporti di composizione (che potranno comodamente essere espressi nella forma di proporzioni a 100.000) i dati della tabella dell'ASI sui matrimoni secondo l'età dello sposo e l'età della sposa.

12. — Si traducano in rapporti di composizione: *a*) per righe, *b*) per colonne, i dati citati nell'esercizio precedente.

13. — Si traducano in rapporti di composizione, sia per righe sia per colonne, i dati dell'ASI intorno: *a*) agli ammalati di mente secondo il genere della malattia e l'età; *b*) ai suicidi secondo il sesso e il mezzo adoperato; *c*) agli infortuni nel lavoro secondo l'età e il sesso; *d*) ai trasferimenti di proprietà secondo la regione e secondo la causa del trasferimento; *e*) ai redditi di ricchezza mobile secondo la regione e secondo la categoria di reddito. — NB. Negli ultimi due esempi si può considerare il numero ovvero l'ammontare (dei trasferimenti, dei redditi). Si espongano le conclusioni che possono desumersi dall'esame dei dati così elaborati.

14. — Nell'ASI si trova già tradotta in rapporti di composizione, così per righe come per colonne, la serie di second'ordine che indica la distribuzione dei morti in Italia secondo la causa di morte e l'età. Si cerchi di rendersi conto del modo in cui tale forma di presentazione dei dati agevola l'esame di essi.

15. — Quando e come si possono adoperare rapporti indici per la rappresentazione della serie di second'ordine? Qual è il significato dei rapporti indici calcolati per righe? per colonne?

16. — Quando e come si possono adoperare rapporti di composizione per la rappresentazione della serie di second'ordine? Qual è il significato dei rapporti di composizione aventi come denominatore il totale della serie? i totali delle singole righe? i totali delle singole colonne? Si chiariscano le risposte con esempi, riferendosi a quelli considerati negli esercizi 11 a 13.

17. — Si applichi alla tabella del paragrafo 7 (matrimoni secondo le combinazioni di stato civile degli sposi e delle spose) il procedimento esemplificato nel paragrafo 8 per il calcolo di un indice di relazione.

18. — Si descriva in modo generale il procedimento seguito nel paragrafo 8 per il calcolo di un indice di relazione. Qual è il massimo valore che tale indice può raggiungere?

19. — In che differisce il metodo seguito nel paragrafo 9 per l'analisi delle relazioni esistenti fra le due circostanze secondo le quali è ordinata la tabella di second'ordine, da quello seguito nel paragrafo 10? Qual elemento di arbitrio è inerente al secondo metodo, mentre manca nel primo?

20. — Qual è il significato della «distribuzione calcolata» del paragrafo 10? Da quale ipotesi si parte per eseguire il calcolo della distribuzione stessa? Si può ravvisare qualche contraddizione fra tali ipotesi?

21. — Come si esegue il confronto fra la «distribuzione osservata» di un fenomeno, secondo due circostanze, e la «distribuzione calcolata» per l'ipotesi di assenza di relazione fra le due circostanze? Si definiscano le «contingenze», le «contingenze relative», la «contingenza media assoluta».

22. — Si calcoli la contingenza media assoluta per la tabella riferita nel paragrafo 7.

23. — In qual modo, calcolate le contingenze, si possono da esse desumere indici del grado di eccedenza e indici del grado di deficienza? Che cosa sono gli «indici di attrazione» del Benini? Si calcolino indici di attrazione dalla tabella del paragrafo 7. Dai dati dell'ASI si calcolino indici di attrazione secondo lo stato civile degli sposi, per anni successivi.

24. — Si calcolino indici di attrazione dalla tabella dell'ASI sui matrimoni secondo le combinazioni d'età degli sposi e delle spose. Si confronti, casella per casella, mediante le contingenze e le contingenze relative, la distribuzione osservata dei matrimoni con la distribuzione calcolata per l'ipotesi di assenza d'ogni relazione fra le età dello sposo e della sposa. Si paragonino le conclusioni deducibili dai precedenti confronti con quelle ottenute negli esercizi 11 e 12.

25. — Il massimo numero che potrebb'essere contenuto in una casella di una tabella di second'ordine è uguale al minore fra i due numeri che rappresentano rispettivamente la somma della riga e la somma della colonna cui appartiene la casella.

Il minimo numero che potrebb'essere contenuto in una casella è zero se la somma della riga più la somma della colonna cui la casella appartiene non supera la somma di tutti i dati della serie di second'ordine; se la supera, la differenza rappresenta il minimo numero che potrebb'essere contenuto nella casella.

Si dimostri quanto sopra è affermato; e si verifichi se nello svolgimento degli esercizi 23 e 24 si è seguito il criterio corretto per la determinazione del massimo e del minimo numero contenibile in ciascuna casella.

26. — Si provi ad applicare il metodo esemplificato nel paragrafo 10 alla seguente tabella di second'ordine.

Distribuzione di un complesso di dattilografi secondo l'ordine progressivo dell'ora di lavoro e il numero degli errori commessi su ogni 1000 parole battute.

Numero progressivo dell'ora di lavoro	Errori commessi su 1000 parole											Totale
	0	1	2	3	4	5	6	7	8	9	10	
1	—	2	3	5	4	8	2	1	—	—	—	25
2	2	4	7	5	3	—	3	1	—	—	—	25
3	1	5	3	6	1	2	4	2	1	—	—	25
4	—	3	2	4	5	7	1	—	2	1	—	25
5	—	—	4	1	2	6	6	3	1	2	—	25
6	—	—	2	—	3	1	5	4	7	2	1	25
Totale	3	14	21	21	18	24	21	11	11	5	1	150

Si applichi alla tabella stessa il metodo esemplificato nel paragrafo 9.

Si confrontino le conclusioni desunte dall'applicazione dei due metodi. Si ricerchi se col sussidio di medie possa esprimersi semplicemente in pochi dati la relazione che intercede fra la durata del lavoro e la frequenza degli errori di scrittura.

(NB. I dati della tabella si riferiscono agli stessi 25 dattilografi osservati in sei consecutive ore di lavoro).

27. — All'applicazione, nel precedente esercizio, del metodo esemplificato nel paragrafo 10, possono muoversi le obiezioni espresse nel paragrafo 11? Si provi ad applicare a questo caso particolare il ragionamento del paragrafo 11. Si può in questo esempio, come in quello del paragrafo 12, determinare con procedimento logicamente corretto la distribuzione che si avrebbe in assenza di relazione tra la durata del lavoro e la frequenza degli errori?

28. — Si esponga e si chiarisca, con esempi diversi da quelli recati nel testo, la varia natura delle relazioni che possono sussistere fra le due circostanze secondo le quali è ordinata una serie di second'ordine.

29. — Si esponga e si chiarisca, con esempi diversi da quelli del testo, in quali modi si possa presentare la relazione tra le due circostanze secondo le quali è ordinata una serie di second'ordine.

30. — Si riepiloghino i vari metodi studiati per la comparazione fra due serie di second'ordine e si paragonino i loro fini e i loro risultati.

CAPITOLO XIX

La rappresentazione analitico-sintetica della serie di second'ordine in forme grafiche

Rappresentazione nel piano mediante diagrammi — Rappresentazione nello spazio mediante stereogramma — Interpolazione — Grafico a curve di livello — Difficoltà che si oppongono alla diffusione di tali metodi.

1. — Qualunque serie di second'ordine si può rappresentare graficamente; tale forma di rappresentazione presenta, generalmente però in minor grado, gli stessi vantaggi che offre nella sua applicazione alle serie di prim'ordine.

Ciascuna riga, o ciascuna colonna, di una tabella di second'ordine contiene una serie di prim'ordine che può essere tradotta in diagramma: l'insieme dei diagrammi corrispondenti alle varie righe ordinati uno sotto l'altro, o di quelli corrispondenti alle varie colonne ordinati uno di fianco all'altro, offre la visione analitico-sintetica dell'intera serie di second'ordine. Possiamo, per esempio, tracciare la curva di distribuzione delle stature dei figli in corrispondenza a ciascuna statura dei padri; ponendo poi uno sotto l'altro i diagrammi vediamo come al crescere della statura dei padri corrisponda una modificazione della distribuzione dei figli secondo la statura.

2. — Un procedimento più sistematico di quello ora accennato consiste nella rappresentazione geometrica nello spazio. Consideriamo anzitutto il caso in cui le due circostanze sono qualitative: supponiamo di avere la classificazione di un certo complesso d'individui secondo la professione da loro esercitata e quella che era od è esercitata dal padre di ciascuno di essi. Col metodo sopra descritto, potremmo per ciascuna professione del padre tracciare un diagramma a canne d'organo rappresentante la suddivisione dei figli tra i diversi gruppi professionali; mettere poi uno sotto l'altro i diagrammi così ottenuti. Ci riuscirebbe facile il confronto tra l'altezza delle varie «canne» in ciascun diagramma; meno facile il confronto tra l'altezza delle «canne» che a vicenda si corrispondono nei vari diagrammi. Se potessimo, per così dire, far alzare in piedi le canne che giacciono sul foglio, entrambi i confronti ci riuscirebbero ugualmente facili. Basterà disporre di un paio di forbici per poter mettere in piedi le «canne»: ne ritaglieremo i contorni

lasciandole unite al foglio soltanto per la base; indi piegandole lungo la base, le disporremo perpendicolarmente al foglio.

3. — Dal precedente sistema allo stereogramma è breve il passo: per mantenere verticali le nostre canne potremo disporre dietro ed a contatto di ciascuna di esse un pilastrino con base rettangolare, avente altezza e larghezza uguali a quelle della corrispondente « canna »: ed ecco che l'insieme dèi pilastrini ci dà uno stereogramma. Possiamo ottenerlo direttamente col rappresentare ciascun dato della serie mediante un prisma rettangolare, a base quadrata e uguale per tutti i prismi (del tipo di quei pilastrini di legno che adoperano i bambini nel giuoco delle costruzioni), e di altezza proporzionale al dato da rappresentare. I prismi vanno poi disposti verticalmente l'uno accanto all'altro, ordinati per righe secondo l'una circostanza (professione del padre, nel nostro esempio) e per file secondo l'altra circostanza (professione del figlio, nel nostro esempio). Avvertasi che, se qualche casella della tabella di second'ordine è vuota, risulterà in corrispondenza ad essa un vuoto, anche nello stereogramma.

Lo stereogramma ottenuto nel modo ora descritto ci permetterà, nel nostro esempio, di scorgere con uguale facilità la distribuzione dei figli per professioni in ciascun dato gruppo di professione dei padri, e la distribuzione dei padri per professioni in ciascun dato gruppo di professione dei figli, e di comparare la rappresentanza di un dato gruppo di professione dei figli nei vari gruppi di professione dei padri, o la rappresentanza di un dato gruppo di professione dei padri nei vari gruppi di professione dei figli. La costruzione di un simile stereogramma non riuscirebbe molto più laboriosa di quella dei diagrammi a canne d'organo, per chi disponesse dei mezzi necessari; se è vero che lo stereogramma non si può inserire in una pagina d'un libro, è pur vero che vi si può inserire la sua rappresentazione prospettica sul piano (per esempio lo stereogramma che or ora abbiamo supposto di costruire potrebb'essere fotografato dal punto di vista più adatto, e la fotografia potrebb'essere poi riprodotta).

4. — Se le due circostanze sono quantitative:

o le loro misure sono date per singole unità, o per intervalli uguali, e ciascun prisma dovrà avere base quadrata — uguale per tutti i prismi — come nel caso precedente;

o sono dati per intervalli diversi, e le due dimensioni della base di ciascun prisma dovranno essere rispettivamente proporzionali alle misure degli intervalli cui corrispondono, in modo che la base riesca proporzionale al prodotto delle misure dei due intervalli; l'altezza, invece, dovrà essere proporzionale al quoziente del dato da rappresentare per il prodotto delle misure dei due intervalli. Per esempio, nella rappresentazione dei matrimoni secondo le combinazioni d'età degli sposi e delle spose, dovendo rappresentare 8.624 matrimoni avvenuti fra uomini di 18-21 anni (intervallo di 3 anni d'età) e donne di 15-21 anni (intervallo di 6 anni d'età), formeremo la base del prisma con lati rispettivamente proporzionali a 3 e a 6, in modo che la sua area riesca proporzionale a 18, e daremo al prisma un'altezza proporzionale a

$$8.624 : 18 = 479,$$

numero medio aritmetico dei matrimoni corrispondenti a ciascuna combinazione fra un anno di età dello sposo compreso tra 18 e 21 e un anno di età della sposa compreso tra 15 e 21.

L'insieme dei prismi ordinati secondo le due circostanze costituisce una figura solida, che in parecchi casi (specialmente se si tratta di fenomeni fisici, biologici, o di fenomeni sociali determinati in parte considerevole da fattori biologici) presenta una conformazione abbastanza regolare. In tali casi, come con l'interpolazione nel piano si arrotondano gli spigoli di un diagramma a spezzata, così con un'interpolazione nello spazio a tre dimensioni si possono arrotondare gli spigoli di uno stereogramma; e come alla linea tracciata nel piano si può talora sostituire utilmente, con buona approssimazione, una curva geometrica regolare avente una espressione analitica abbastanza semplice, così alla superficie tracciata nello spazio si può talora sostituire utilmente, con buona approssimazione, una superficie geometrica regolare cui corrisponda un'espressione analitica abbastanza semplice. Criteri analoghi a quelli che disciplinano l'interpolazione di una funzione d'una variabile nel piano si possono adottare per l'interpolazione di una funzione di due variabili nello spazio.

5. — Col sistema delle curve di livello, mediante il quale si rappresenta sulle carte geografiche la configurazione orografica del terreno, si può rappresentare sul piano, e quindi sopra un foglio, uno stereogramma di conformazione regolare. In pratica per tracciare le curve di livello non occorre aver eseguito lo stereogramma: basta delineare ordinatamente sul foglio le basi dei prismi cui ci siamo dianzi riferiti, segnare su ciascuna base la quota corrispon-

dente, cioè il numero indicante l'altezza del prisma che vi si dovrebbe costruire, e congiungere poi ordinatamente fra loro i punti centrali delle basi aventi quota uguale (o quota compresa fra dati limiti).

6. — Tutti questi procedimenti geometrici non possono divenire di uso generale: vi fa ostacolo sia la laboriosità dell'applicazione, sia la difficoltà dell'interpretazione. Qualsiasi persona di intelligenza normale riesce facilmente ad interpretare presto e bene un diagramma correttamente eseguito; pochi privilegiati riescono, dopo lunga pratica, ad interpretare prontamente e sicuramente uno stereogramma; e chiunque abbia avuto occasione di servirsi spesso di carte topografiche, come militare o come escursionista, sa per esperienza quanto sia difficile formarsi un'idea esatta del rilievo del terreno mediante la rappresentazione a curve di livello.

Anche l'interpretazione di formole rappresentatrici di funzioni di due variabili riesce in generale malagevole, eccettuati alcuni casi, non frequenti nella pratica, di funzioni eccezionalmente semplici.

Specialmente per la descrizione di fenomeni sociali, l'instabilità delle cui forme sconsiglia l'uso di metodi richiedenti lungo lavoro, si preferiscono elaborazioni della tabella di second'ordine che richiedano minor fatica e che si prestino alla determinazione di indici numerici atti a caratterizzare la serie e ad agevolare le comparazioni. Esporremo brevemente nel prossimo capitolo un metodo che corrisponde a tali requisiti.

Quesiti ed esercizi: 1. In quali modi si può rappresentare graficamente una serie di second'ordine?

2. — Si esaminino gli esempi di stereogrammi destinati a rappresentare serie di second'ordine, riportati nel manuale di statistica di YULE, e si spieghi il criterio di esecuzione di tali rappresentazioni geometriche. Altrettanto per gli esempi di GABAGLIO e di VINCI.

3. — Si rappresentino mediante diagrammi a canne d'organo i dati delle seguenti tabelle riportate o indicate nel capitolo XVIII: *a*) prima tabella del paragrafo 3 (raccolti secondo i sistemi di coltura e le razze di frumento); *b*) prima tabella del paragrafo 8 (condizione rispetto alla vaccinazione e comportamento di fronte all'epidemia); *c*) distribuzione degli studenti secondo i risultati scientifici e i risultati sportivi (si confronti graficamente la « distribuzione osservata » indicata nel paragrafo 9 con la « distribuzione calcolata » indicata nel paragrafo 10); *d*) tabelle indicate nell'esercizio 13.

4. — Si rappresentino mediante diagrammi a spezzata i dati delle seguenti tabelle riportate o indicate nel capitolo XVIII: *a*) matrimoni secondo l'età dello sposo e l'età della sposa (tabella dell'ASI); *b*) tabelle indicate nell'eser-

cizio 6; *c*) tabelle indicate nell'esercizio 13; *d*) tabella riportata nell'esercizio 26.

5. — Si provi a rappresentare mediante stereogramma (a prismi vuoti, di cartoncino leggero, che si possono facilmente e rapidamente costruire) qualcuna delle serie di second'ordine riportate o indicate nel capitolo XVIII.

6. — I rettangoli, determinati nel modo indicato al principio del paragrafo 3, per servire di base ai prismi delio stereogramma, possono servire invece per un'altra forma di rappresentazione grafica. Su ciascuno di questi rettangoli si segnino tanti punti quante sono le unità contenute nella corrispondente casella delia tabella di second'ordine. La diversa densità dei punti nelle varie parti del grafico avverte l'occhio esperto della relazione eventualmente esistente fra le due circostanze. Si applichi questo modo di rappresentazione alle seguenti tabelle del capitolo XVIII: *a*) rendimenti per ettaro secondo la razza del frumento: paragrafo 5; *b*) risultati scientifici e risultati sportivi: seconda tabella del paragrafo 9; *c*) durata del lavoro ed errori commessi dai dattilografi: esercizio 26. Si applichi lo stesso metodo alla tabella dei risultati combinati di due esami universitari, contenuta nel capitolo XX, paragrafo 7.

CAPITOLO XX.

Rappresentazione della relazione fra due circostanze quantitative secondo le quali è ordinata una serie di second'ordine.

Posizione del problema. Come si ricerca: se esista una dipendenza fra le due circostanze; quale sia la forma di tale dipendenza; in quale proporzione le variazioni dell'una circostanza dipendano da quelle dell'altra. Indici del grado di dipendenza, indici del grado di indipendenza: critica del procedimento usuale. Correlazione: correlazione diretta e inversa.

1. — Abbiamo ripetutamente avvertito che i termini di una serie di second'ordine, ordinati secondo due circostanze quantitative, i cui valori variabili indichiamo rispettivamente coi simboli x e y, si possono riguardare come altrettanti valori di una funzione delle due variabili x e y; da questi valori noti si può talvolta ricavare, con procedimento interpolatorio, un'espressione generale della relazione funzionale, ma spesso tale espressione risulta troppo complicata e troppo difficilmente interpretabile per essere usata da un pubblico non composto di esperti della statistica matematica.

Conviene, pertanto, cercare metodi di rappresentazione sintetica della serie di second'ordine ordinata secondo circostanze quantitative, che possano applicarsi anche quando non sussiste quella regolarità nella distribuzione dei termini della serie in funzione delle due variabili che è presupposto necessario per l'applicazione dei procedimenti d'interpolazione, e che presentino nel tempo stesso requisiti di semplicità.

Eseguiremo tale ricerca considerando il tipo di serie di second'ordine in cui le circostanze quantitative esprimono le misure di due caratteri corrispondenti ai casi osservati, e i dati della serie esprimono i numeri di casi corrispondenti alle varie combinazioni delle misure dei due caratteri (numero degli studenti che hanno riportato un dato voto nell'esame di economia e un dato voto nell'esame di statistica; numero degli individui che hanno una data statura e son figli di individui che avevano una data statura; numero degli individui che hanno un dato perimetro toracico e un dato peso).

Una serie di tal sorta viene utilizzata, in generale, per lo studio della relazione eventualmente esistente tra le due circostanze secondo le quali è ordinata. È evidente che questa relazione non corrisponde al concetto usuale di funzione, secondo il quale a ciascun dato valore di x (voto in economia, statura del padre, perimetro toracico) dovrebbe corrispondere un unico determinato valore di y (voto in statistica, statura del figlio, peso), poichè vediamo invece corrispondere ad ogni valore di x più valori di y, almeno in parte differenti tra loro.

Possiamo tuttavia conciliare questo stato di fatto con l'ipotesi che y vari in funzione di x, ammettendo che y dipenda non solo da x ma anche da altre circostanze. Poichè nella realtà non varia soltanto x ma variano anche queste altre circostanze, l'osservazione non ci può indicare l'unico determinato valore di y che corrisponderebbe a ciascun dato valore di x se y fosse *soltanto* funzione di x. Per esempio, a ciascuna statura del padre corrispondono tante differenti stature del figlio, perchè la statura del figlio non dipende soltanto dalla statura del padre ma anche da altre circostanze.

2. — Ponendoci da questo punto di vista, ci possiamo chiedere in ogni caso concreto:

primo: se esista una dipendenza fra x e y;

secondo: in caso affermativo, quale sia la forma di tale dipendenza, che nella realtà ci è dissimulata dalla molteplicità delle circostanze operanti a modificare y;

terzo: ammessa una data forma di dipendenza, per qual frazione le variazioni di y dipendano da quelle di x e per qual frazione non ne dipendano.

3. — Se non esistesse dipendenza fra x e y, a ciascun dato valore di x corrisponderebbe la stessa distribuzione proporzionale dei valori di y, e quindi anche lo stesso valore medio di y (p. es. a ciascuna statura del padre corrisponderebbe la stessa distribuzione proporzionale delle stature dei figli, e quindi anche la stessa statura media dei figli). In pratica si suole assumere la distribuzione proporzionale dei valori di y che si ha nel complesso dei casi osservati come quella che si avrebbe anche in corrispondenza a ciascun valore di x, se tra le due variabili non sussistesse alcuna dipendenza: l'ipotesi non è logicamente rigorosa, come abbiamo dimostrato nel capitolo XVIII; perciò bisogna impiegarla con cautela, anche se l'esperienza insegna ch'essa in parecchi casi può costituire un utile sussidio alle indagini.

4. — L'esame dei valori medi di y che, secondo l'osservazione, corrispondono ai diversi valori dati di x ci può dunque, anzitutto, dare indizio della dipendenza, o non dipendenza. Nel caso di non dipendenza tali valori medi saranno tutti uguali fra loro; e se li troviamo tutti uguali avremo ragione di presumere che non esista dipendenza. Se non sono uguali, appunto il modo del loro variare, al variare di x, ci darà indizio della forma della relazione tra le due variabili (ci indicherebbe con precisione assoluta tale forma soltanto se le variazioni dipendenti da altre circostanze si compensassero a vicenda nella formazione di ciascuno dei suddetti valori medi). Se, per esempio, al variare della statura del padre secondo una progressione aritmetica vediamo variare secondo una progressione aritmetica anche la statura media dei figli, intendiamo subito che potremo esprimere molto semplicemente la relazione fra l'una e l'altra statura mercè il rapporto fra le ragioni delle due progressioni.

In via generale:

potremo assumere, in prima approssimazione, i valori medi di y in funzione di x a rappresentare la forma della dipendenza tra le due variabili;

se tale forma si accosta a quella di una funzione analitica di semplice espressione, potremo, in via di seconda approssimazione, assumere questa funzione a rappresentare la dipendenza tra le due variabili.

In casi speciali la natura stessa delle due variabili potrà suggerire certe ipotesi sulla forma della loro relazione (si potrà supporre, per esempio, che il voto in statistica tenda ad uguagliare il voto in economia; che la statura del figlio tenda ad uguagliare quella del padre; che il prodotto del campo tenda a proporzionarsi alla quantità di seme impiegata, ecc.).

Normalmente non v'è alcun procedimento *sicuro* che ci conduca a determinare la forma della relazione tra y ed x; è bene insistere su questo punto, perchè secondo la forma supposta varia poi l'apprezzamento del grado di dipendenza di y da x, da un canto, e da altre circostanze, dall'altro canto.

Di solito la forma della relazione tra y e x viene desunta dall'andamento delle medie di y in funzione di x, come sopra abbiamo detto: metodo arbitrario ed empirico, cui tuttavia non è facile sostituirne uno migliore.

5. — Supposto che si sia adottata una rappresentazione, che si ritiene soddisfacente, della ipotetica relazione funzionale tra la x e la y, nel nostro esempio tra la statura del padre e quella del figlio, rimane da risolvere il problema della misura comparativa del grado in cui la y (statura del figlio) dipende dalla x (statura del padre), e del grado in cui dipende da altre circostanze.

Indichiamo con u i valori accertati, con y i corrispondenti valori teorici della funzione; con U una media generale dei valori accertati (in generale converrà adottare la media aritmetica).

Se consideriamo un determinato valore accertato u (la statura di un figlio), il corrispondente valore teorico y (la statura che il figlio avrebbe se l'altezza del figlio dipendesse esclusivamente da quella del padre, e la forma della dipendenza fosse quella supposta), e la media U (media generale delle stature dei figli), possiamo istituire le seguenti differenze:

$(u - y) =$ variazione di u non dipendente da variazione di x (deviazione della statura accertata dalla statura teorica, ossia: variazione della statura del figlio non dipendente da variazione della statura del padre);

$(y - U) =$ variazione di y dipendente da variazione di x (deviazione della statura teorica dalla statura media generale, corrispondente alla relazione supposta tra statura del figlio e statura del padre, ossia: variazione della statura del figlio dipendente da variazione della statura del padre);

$(u - U) =$ variazione risultante di u (deviazione della statura accertata dalla media generale; ossia variazione della statura del figlio risultante dall'azione combinata delle circostanze espresse nella statura del padre e delle circostanze estranee alla statura del padre). Algebricamente, quest'ultima differenza è la somma, e fisicamente potremmo dire la risultante, delle due precedenti. È infatti $(u - U) = (u - y) + (y - U)$.

Per brevità chiameremo da ora innanzi *variazioni risultanti* quelle del tipo $(u - U)$, *variazioni indipendenti* quelle del tipo $(u - y)$, *variazioni dipendenti* quelle del tipo $(y - U)$. Tutte queste differenze o variazioni possono essere positive, nulle o negative.

Se la statura del figlio dipendesse *esclusivamente* dalla statura del padre, le variazioni indipendenti sarebbero tutte nulle, poichè i valori accertati u coinciderebbero coi corrispondenti valori teorici y.

Se la statura del figlio dipendesse *esclusivamente* da circostanze differenti e indipendenti dalla statura del padre, le variazioni dipendenti sarebbero tutte nulle, poichè i valori teorici y coinciderebbero con la media generale U dei valori accertati.

Nei casi intermedi tra questi due estremi della dipendenza completa e della indipendenza completa, quanto maggiori sono le variazioni dipendenti e quanto minori sono le indipendenti, tanto maggiore è il grado di dipendenza della statura del figlio da quella del padre.

Un *indice del grado di dipendenza* conviene che varii entro limiti ben definiti, per poter adempiere la sua funzione: per esempio fra 0 e 1, come è consuetudine per simili indici. Il rapporto tra la somma dei valori assoluti delle variazioni dipendenti e la somma stessa aumentata della somma dei valori assoluti delle variazioni indipendenti corrisponde al requisito ora indicato, perchè si annulla nel caso di indipendenza completa, diviene uguale all'unità nel caso di dipendenza completa e, fra questi limiti, cresce al crescere del rapporto fra la somma dei valori assoluti delle variazioni dipendenti e la somma dei valori assoluti delle variazioni indipendenti.

L'indice del grado di dipendenza così calcolato soddisfa un'esigenza logica assai importante, come ora vedremo. Ci siamo chiesti un indice del grado di dipendenza: avremmo, invece, potuto chiederci un indice del grado di indipendenza: logicamente le due richieste sono ugualmente giustificate, anzi non rappresentano che due modi di porre un unico problema. Se diciamo che la statura del figlio dipende per 60 % dalla statura del padre (se, cioè, enunciamo un indice del grado di dipendenza uguale a 0,6) implicitamente affermiamo che per 40 % non ne dipende, ossia dipende da altre circostanze (il che equivale ad enunciare un indice del grado di indipendenza uguale a 0,4). Bisogna, pertanto, che se si procede alla determinazione diretta dell'indice del grado di indipendenza, con un metodo parallelo a quello seguito per l'indice del grado di dipendenza, si ottenga come risultato 0,4, nel nostro esempio particolare, e non un altro valore: bisogna, cioè, in generale, che la somma dei due indici sia sempre uguale all'unità.

Il procedimento parallelo a quello seguito per l'indice del grado di dipendenza, che ci può condurre a un indice del grado di indipendenza, è ovvio: basta mettere in rapporto la somma dei valori assoluti delle variazioni indipendenti con la somma stessa, aumentata della somma dei valori assoluti delle variazioni dipendenti. Questo rapporto si annulla nel caso di dipendenza completa, è uguale all'unità nel caso di indipendenza completa.

Per conseguenza del modo di formazione dei due indici del grado di dipendenza e del grado d'indipendenza, la loro somma risulta sempre uguale all'unità: cioè le loro indicazioni concordano.

Invece di adoperare, nella formazione degli indici, somme di valori assoluti delle variazioni considerate, si possono adoperare medie aritmetiche di valori assoluti: è perfettamente la stessa cosa. Invece della media aritmetica si può adoperare la media quadratica od altra media dei valori assoluti; non è più la stessa cosa, ma, se non si è stravaganti nella scelta della media, normalmente si otterranno risultati non molto differenti. L'impiego della media quadratica è talvolta suggerito da ragioni pratiche, perchè i calcoli eseguiti per l'interpolazione ne agevolano il computo. Ma in generale non v'è ragione teorica per preferirla, almeno nel campo delle applicazioni statistiche ai fenomeni sociali.

Questi indici del grado di dipendenza acquistano vera importanza nell'indagine scientifica, quando ripetute osservazioni sopra la relazione tra due fenomeni mostrano una certa costanza degli indici stessi. A mo' d'esempio, l'accertare un indice del grado di dipendenza uguale a 0,6 fra la statura del figlio e quella del padre per un dato gruppo di individui non ha grande rilevanza. Ma se in ogni analoga indagine si accertano indici press'a poco uguali a 0,6, si è scoperta una uniformità, sia pure approssimativa, che può essere di notevole sussidio nello studio dei fattori influenti sulla statura.

6. — Gli indici del grado di dipendenza più comunemente usati non sono calcolati nel modo da noi suggerito. La precedente spiegazione ci permette però di chiarirne semplicemente il significato e di porne in rilievo il difetto fondamentale.

Ripigliamo la somma dei valori assoluti delle variazioni dipendenti, e proviamo a metterla in rapporto con la somma dei valori assoluti delle variazioni risultanti. Nel caso d'indipendenza completa anche questo rapporto risulta nullo; nel caso di dipendenza completa risulta uguale all'unità; nei casi intermedi varia proporzionalmente alla somma delle variazioni dipendenti. Anch'esso sembra, a prima vista, un idoneo indice del grado di dipendenza.

Proponiamoci ora di determinare, con criterio analogo, un indice del grado di indipendenza: dovremo a tal uopo mettere in rapporto la somma dei valori assoluti delle variazioni indipendenti con la somma dei valori assoluti delle variazioni risultanti. L'indice ottenuto si annulla nel caso di dipendenza completa; diviene uguale all'unità nel caso di indipendenza completa; varia proporzionalmente alla somma dei valori assoluti delle variazioni indipendenti, nei casi intermedi. Pare, dunque, un buon indice del grado d'indipendenza.

Poniamo ora accanto i due indici — di dipendenza e d'indipendenza — per vedere se le loro indicazioni concordino (nel qual caso la loro somma dovrebbe sempre risultare uguale all'unità come nelle ipotesi estreme), oppure discordino. Sommandoli, otteniamo:

$$\frac{\text{Somma dei val. ass. delle var. dip.} + \text{Somma dei val. ass. delle var. indip.}}{\text{Somma dei valori ass. delle var. risultanti.}}$$

Ciascuna di quelle che abbiamo chiamato variazioni risultanti non è che la somma algebrica di una variazione dipendente e di una variazione indipendente, e poichè questi due termini in molti casi hanno segno opposto, la somma dei valori assoluti delle variazioni risultanti è normalmente minore della somma dei valori assoluti delle variazioni dipendenti aumentata della somma dei valori assoluti delle variazioni risultanti. Quindi la somma dei due indici del grado di dipendenza e del grado d'indipendenza ottenuti con questo secondo criterio è sempre maggiore dell'unità, salvo, come abbiamo già avvertito, nei due casi estremi di dipendenza completa e di indipendenza completa. Sicchè può avvenire che l'indice del grado di dipendenza ci assicuri che il 71 % delle variazioni della statura del figlio dipende da quelle della statura del padre, mentre nel tempo stesso l'indice del grado di indipendenza ci insegna che il 71 % delle variazioni della statura del figlio non dipende da quelle della statura del padre. Contraddizione intollerabile, che, non potendosi in alcun modo evitare, induce a rigettare l'uso degli indici di questo secondo tipo. Quali rapporti fra le deviazioni dei dati osservati dai dati interpolati e le deviazioni dei dati osservati dalla media, gli indici così calcolati, che non sono atti a misurare il grado di indipendenza, misurano invece il vantaggio che si ottiene col sostituire il procedimento dell'interpolazione a quello della media nella rappresentazione della relazione fra due variabili.

Per il calcolo di questi indici, invece che somme di valori assoluti si possono adoperare medie aritmetiche dei valori stessi, o altre medie, per esempio medie quadratiche.

Se con U si designa la media aritmetica delle u, se la funzione interpolatrice è una funzione lineare (cui graficamente corrisponde una linea retta), i parametri della quale vengono determinati col metodo dei minimi quadrati, e se nel calcolo dell'indice del grado di dipendenza si adoperano medie quadratiche, si ottiene, col metodo qui criticato, il valore assoluto del « coefficiente di correlazione » di BRAVAIS-PEARSON. (Il segno di questo coefficiente, calcolato con altro procedimento, indica se la retta è ascendente o discendente, se cioè la correlazione è diretta o inversa: vedansi più avanti il paragrafo 10 e l'esercizio 11).

Se con U si designa la media aritmetica delle u, se come valori teorici y si assumono le medie aritmetiche dei valori u corrispondenti a ciascun dato valore di x, e se nel calcolo dell'indice del grado di dipendenza si adoperano medie quadratiche, si giunge al « rapporto di correlazione » di PEARSON.

A prescindere dalla frequente insufficienza dell'interpolazione lineare, che rende poco consigliabile l'impiego del coefficiente di correlazione (difetto sanabile, poichè il metodo si può facilmente generalizzare per curve interpolatrici di qualsiasi altra forma), i due indici del grado di dipendenza sono affetti dal grave vizio poc'anzi notato e perciò non dovrebbero essere impiegati, poichè tendono ad esagerare l'apprezzamento del grado di dipendenza.

È bensì vero che all'uno e all'altro indice si può giungere per via diversa da quella qui tracciata; anzi al coefficiente di correlazione si può giungere per parecchie vie, con procedimenti che sembrano logicamente inattaccabili. Ma per nessuna via si sfugge alla critica dianzi esposta, perchè le definizioni delle variazioni dipendenti e delle variazioni indipendenti — ossia le sole premesse del nostro ragionamento — vengono in ogni caso, implicitamente od esplicitamente, accolte.

7. — Per dare un'esempio d'applicazione del metodo esposto

nei paragrafi 2-5 ci varremo dei seguenti dati sui risultati combinati di due esami superati dagli stessi studenti.

Studenti, secondo i voti riportati in due esami.

Voti nell'esame di economia	Voti nell'esame di statistica													Totale
	18	19	20	21	22	23	24	25	26	27	28	29	30	
18	87	13	19	26	7	2	6	—	2	4	3	—	4	173
19	19	7	3	4	5	—	1	—	—	2	—	1	—	42
20	28	5	14	22	13	1	1	1	—	2	—	—	1	88
21	32	9	23	35	21	2	10	2	1	7	2	—	4	148
22	8	4	11	18	9	—	4	3	2	5	2	1	—	67
23	7	—	5	11	8	1	7	3	1	4	1	1	3	52
24	9	5	2	26	1	8	13	15	12	10	2	2	1	106
25	—	2	1	17	—	12	19	17	3	8	3	—	—	82
26	3	—	—	6	—	2	6	7	8	1	2	—	1	36
27	5	1	—	8	7	6	21	3	7	27	2	7	3	97
28	1	—	2	4	2	—	2	2	12	1	3	—	7	36
29	—	—	1	2	—	—	2	3	4	1	2	2	1	18
30	2	—	2	3	1	3	3	2	9	6	6	7	11	55
Totale	201	46	83	182	74	37	95	58	61	78	28	21	36	1.000

Se non esistesse alcuna relazione tra il risultato x dell'esame di economia e il risultato u dell'esame di statistica, in corrispondenza a ciascun voto in economia dovremmo trovare un'uguale distribuzione proporzionale dei voti in statistica. Se si ammette che questa distribuzione sarebbe la stessa che si osserva per il complesso degli studenti (cioè quella indicata nell'ultima riga della precedente tabella), si deve anche ammettere che — sempre nell'ipotesi di indipendenza tra x e u — a ciascun valore di x corrisponderebbe lo stesso valore medio di u e che tal valore medio coinciderebbe col valore medio generale.

La media aritmetica dei voti nell'esame di statistica è 22,35; nell'ipotesi di indipendenza tra i risultati dei due esami, a qualsiasi voto riportato in economia dovrebbe corrispondere una media aritmetica dei voti riportati in statistica uguale a 22,35. In realtà invece vediamo che al crescere del voto in economia tende a crescere il voto in statistica, come ci mostrano le due prime colonne seguenti.

Voto nell'esame di economia x	Corrispondente voto medio aritmetico nell'esame di statistica	Voto nell'esame di statistica: valori interpolati y
18	19,92	19,63
19	19,90	20,20
20	20,26	20,77
21	21,16	21,33
22	21,78	21,90
23	22,69	22,47
24	23,23	23,04
25	23,78	23,61
26	24,06	24,18
27	24,82	24,75
28	25,47	25,32
29	25,56	25,89
30	26,35	26,46

Rappresentando in diagramma lineare i dati della seconda colonna in funzione di quelli della prima, si vede che la spezzata ottenuta può essere con grande approssimazione surrogata da una linea retta. Decidiamo pertanto di assumere una linea retta a rappresentare la dipendenza fra le due variabili, e determiniamo i parametri di questa retta col metodo dei minimi quadrati. L'equazione della retta risulta $y = 0{,}5693\ x + 9{,}3796$; da tale equazione abbiamo desunto i valori raccolti nella terza colonna qui sopra.

Confrontando i voti in statistica effettivamente riportati dagli studenti compresi in ciascuna riga della tabella originaria col voto medio teorico così determinato otteniamo le differenze $(u-y)$. Computando, naturalmente, ciascuna di tali differenze tante volte quanti sono gli studenti che hanno riportato il voto u, la somma dei valori assoluti delle differenze stesse risulta uguale a 2.205. Questa è la somma delle *variazioni indipendenti*, nella nostra nomenclatura.

Confrontando i voti medi teorici in statistica con la media generale (22,35) otteniamo le differenze $(y-U)$. Computando ciascuna differenza tante volte quanti sono gli studenti compresi nella riga cui si riferisce la media teorica y, la somma dei valori assoluti delle differenze stesse risulta uguale a 1.759. Questa è la somma delle *variazioni dipendenti*.

Confrontando, infine, i voti in statistica effettivamente riportati con la media generale, otteniamo le differenze $(u-U)$. Computando ciascuna differenza tante volte quanti sono gli studenti che hanno ottenuto il voto u, la somma dei valori assoluti delle differenze stesse risulta uguale a 2.988. Questa è la somma delle *variazioni risultanti.*

Il nostro indice del grado di dipendenza risulta uguale a 1.759: (1.759 + 2.205) = 0,444; l'indice del grado d'indipendenza a 2.205: (1.759 + 2.205) = 0,556. Si può concludere che per circa 44 % le variazioni del risultato dell'esame di statistica dipendono (nel senso della dipendenza funzionale) da quelle del risultato dell'esame di economia, per circa 56 % non ne dipendono.

Se avessimo riferito la variazione dipendente alla variazione risultante, avremmo ottenuto un indice del grado di dipendenza uguale a 0,589; se le avessimo riferito la variazione indipendente, avremmo ottenuto un indice del grado d'indipendenza uguale a 0,738. I due indici sono inconciliabili tra loro, e quindi vanno respinti entrambi, poichè sono calcolati con procedimenti paralleli.

8. — Quando le circostanze quantitative secondo le quali è ordinata la serie di second'ordine sono, come nel nostro esempio, tali che ciascuna di esse può a volta a volta riguardarsi antecedente o conseguente dell'altra, si potranno ricavare dalla tabella di second'ordine due indici del grado di dipendenza [d'indipendenza], secondo che si riguarda y come funzione di x, o x come funzione di y. Nel nostro esempio, se consideriamo il voto di economia come funzione del voto di statistica, e non viceversa come abbiamo fatto dianzi, possiamo riprendere da capo i calcoli: arriveremo ad indici non coincidenti con quelli or ora calcolati. Ed è naturale che sia così, poichè noi abbiamo cercato di misurare quale frazione delle variazioni dei voti di statistica dipenda dalle variazioni dei voti di economia, mentre col nuovo calcolo ricercheremo invece quale frazione delle variazioni dei voti di economia dipenda dalle variazioni dei voti di statistica.

9. — Il metodo esposto nei paragrafi 2-5 ed esemplificato nel paragrafo 7 si riduce, sostanzialmente, alla determinazione empirica della forma di una relazione di dipendenza funzionale fra le due circostanze quantitative secondo le quali è ordinata la serie di second'ordine e al computo di un indice del grado di approssimazione col quale tale relazione sussiste secondo i risultati dell'osservazione. In tale significato, il metodo può essere accolto anche da chi trovasse troppo complicate o troppo dubbie le ipotesi sulle quali è fondato nella precedente esposizione.

10. — Le forme di dipendenza tra due circostanze quantitative studiate in questo capitolo si sogliono comprendere, come abbiamo già avvertito, nella designazione di «correlazione». Quando la dipendenza è espressa da una funzione crescente, si suol parlare di *correlazione diretta* o *positiva*; quando è espressa da una funzione decrescente, di *correlazione inversa* o *negativa*. In tali casi si usano talora artifizi che rendono positivi nel primo caso e negativi nel secondo gli indici della correlazione, in modo da far sì che essi segnino ad un tempo l'intensità della relazione (cioè il grado della dipendenza) e il senso della relazione.

Indicazioni bibliografiche. — Intorno alla teoria della correlazione vedasi specialmente il citato manuale di YULE; si mantenga però sempre desto, nel corso dello studio di esso, il senso critico.

Quesiti ed esercizi: 1. — Quali sono i problemi che si mira a risolvere, mediante l'elaborazione di una serie di second'ordine ordinata secondo due circostanze quantitative? Si enuncino tali problemi con riferimento concreto alla tabella dell'ASI sui matrimoni secondo le combinazioni d'età degli sposi e delle spose.

2. — In qual modo si accerta se esista una dipendenza fra le due circostanze? Si applichi il criterio alla tabella dei matrimoni secondo le combinazioni di età.

3. — In quali modi si determina la forma della dipendenza fra le due circostanze? Si applichino i vari criteri che sembrano plausibili all'esempio indicato nei precedenti esercizi.

4. — Ammesso che le età medie delle spose, in funzione dell'età dello sposo, indichino la forma della relazione fra le due età, si determini in quale misura l'età della sposa dipenda da quella dello sposo e in quale misura non ne dipenda.

5. — Si calcoli l'indice del grado di dipendenza nell'esempio del paragrafo 7, assumendo come variabile indipendente il voto in statistica e come variabile dipendente il voto in economia (al contrario di quanto è stato fatto nel testo). Se si vuol determinare col metodo dei minimi quadrati l'equazione di una retta che dia la rappresentazione teorica del voto in economia quale funzione del voto in statistica (analogamente al modo tenuto nel testo per ottenere la rappresentazione del voto in statistica quale funzione lineare del voto in economia), si devono ricavare i due parametri a' e b' come è stato indicato a pag. 140, risolvendo il seguente sistema di due equazioni di primo grado a due incognite:

$$22.348\,a' + 1.000\,b' - 22.779 = 0$$
$$511.662\,a' + 22.348\,b' - 516.435 = 0\,.$$

I numeri sopra indicati rappresentano: 22.348 la somma dei voti in statistica, 22.779 la somma dei voti in economia, 511.662 la somma dei quadrati dei voti in statistica, 516.435 la somma dei prodotti del voto in statistica per il voto in economia riportato da ciascuno studente, 1.000 il numero dei casi osservati.

(Per determinare l'equazione riferita nel testo è stato invece risoluto il sistema:

$$22.779\,a + 1.000\,b - 22.348 = 0$$
$$531.835\,a + 22.779\,b - 516.435 = 0\,.$$

Qui 531.835 è la somma dei quadrati dei voti in economia).

Mediante opportuni artifici si potrà semplificare grandemente il calcolo della somma delle variazioni dipendenti.

6. — Si calcoli l'indice del grado di dipendenza nell'esempio del paragrafo 7, assumendo come valori teorici le medie ivi riportate nella seconda colonna della seconda tabella, invece dei valori interpolati ivi riportati nella terza colonna.

7. — Si calcoli l'indice del grado di dipendenza della frequenza degli errori dei dattilografi dalla durata del lavoro assumendo come valore teorico la frequenza media corrispondente ad ogni ora di lavoro (dati dell'esercizio 26 annesso al capitolo XVIII).

8. — A quali criteri è ispirato il calcolo di indici del grado di dipendenza col procedimento esposto nel paragrafo 5? In che differiscono tali criteri da quelli che ispirano il calcolo del « coefficiente di correlazione » e del « rapporto di correlazione », riguardati come misure del grado di dipendenza? Si definiscano le « variazioni dipendenti », le « variazioni non dipendenti », le « variazioni risultanti », e si spieghi come esse vengano impiegate secondo i due diversi criteri nel calcolo degli indici del grado di dipendenza.

9. — Potrebbero applicarsi ai dati della prima tabella del paragrafo 7 i metodi impiegati nel capitolo XVII per confrontare mediante differenze i risultati di due esami? Se si ritiene di sì, si provi ad applicarli. In che differiscono i metodi del capitolo XVII da quelli del capitolo XX? Quali ipotesi si introducono in questo secondo capitolo, che non occorrono nel primo? Qual diverso significato hanno i risultati ottenuti? In che differisce un indice del grado di dipendenza da un indice del grado di coincidenza?

10. — Che cosa s'intende per « correlazione »? Quando la correlazione si dice « diretta », quando « inversa »?

11. — Data la grande diffusione dell'uso del « coefficiente di correlazione », crediamo utile esporre alcuni chiarimenti sulla natura di esso e sul procedimento più rapido per calcolarlo: chiarimenti che gioveranno per l'interpretazione di questo indice e per l'esecuzione dei successivi esercizi.

Siano date n coppie di valori reciprocamente corrispondenti di due variabili χ e υ. Siano rispettivamente X e Y le medie aritmetiche dei valori dati delle due variabili. Se υ è funzione lineare di χ, se cioè sussiste la relazione $\upsilon = a\chi + b$, dove a e b sono due parametri, sommando le n relazioni simili alla precedente che sussistono fra le n coppie di valori delle due variabili e dividendo poi la somma per n, si ottiene l'altra relazione $Y = a\,X + b$. Sottraendo quest'ultima relazione dalla precedente, si ha: $(\upsilon - Y) = a\,(\chi - X)$. Cioè: gli scostamenti delle υ dalla loro media aritmetica sono proporzionali agli scostamenti delle χ dalla loro media aritmetica.

Se, ora, disponendo di n coppie di valori corrispondenti di due variabili empiriche χ ed υ, vogliamo rappresentare approssimativamente l'una variabile quale funzione lineare dell'altra, potremo semplificare l'interpolazione operando sugli scostamenti dei valori dati delle due variabili dalle rispettive medie aritmetiche, invece di operare sui valori stessi.

Posto $y = \upsilon - Y$ e $x = \chi - X$, supporremo che la relazione approssimativa esistente fra y e x sia della forma $y = a\,x$, essendo a un parametro da determinare. Se per determinarlo adottiamo il metodo dei minimi quadrati, dobbiamo porre la condizione $\Sigma\,(y - a\,x)^2 =$ minimo, dove la sommatoria si estende alle n coppie di valori corrispondenti delle due variabili. Perchè sia soddisfatta tale condizione deve annullarsi la derivata rispetto ad a della funzione $\Sigma\,(y - a\,x)^2$; dev'essere cioè: $\Sigma\,a\,x^2 - \Sigma\,x\,y = 0$. Di qui si ricava il valore di $a = \Sigma\,x\,y : \Sigma\,x^2$.

Procedendo in modo del tutto analogo per rappresentare x quale funzione di y mediante una relazione della forma $x = a'\,y$, troviamo: $a' = \Sigma\,x\,y : \Sigma\,y^2$.

Si voglia, ora, misurare il grado di dipendenza fra le due variabili col metodo criticato nel paragrafo 6 del testo: col riferire, cioè, una media delle variazioni dipendenti ad una media delle variazioni risultanti. Ci converrà servirci di medie quadratiche per evitare nuovi calcoli. Essendo nulla la media aritmetica così delle y come delle x (variabili che rappresentano entrambe scostamenti dalla media aritmetica), l'espressione dell'indice risulta molto semplice.

Se partiamo dalla rappresentazione di y in funzione di x, otteniamo come espressione dell'indice il quoziente $\sqrt{\Sigma\,(a\,x)^2} : \sqrt{\Sigma\,y^2}$, che avuto riguardo alla definizione di a si può scrivere: $\Sigma\,x\,y : \sqrt{\Sigma\,x^2\,\Sigma\,y^2}$.

Se partiamo dalla rappresentazione di x in funzione di y, otteniamo come espressione dell'indice il quoziente $\sqrt{\Sigma\,(a'\,y)^2} : \sqrt{\Sigma\,x^2}$, che avuto riguardo alla definizione di a' si può scrivere: $\Sigma\,x\,y : \sqrt{\Sigma\,x^2\,\Sigma\,y^2}$. L'indice risulta identico dai due calcoli.

L'indice così ottenuto, uguale alla media geometrica dei valori assoluti dei parametri a e a', coincide col valore assoluto del « coefficiente di correlazione », generalmente designato con la lettera r, la cui espressione è data di solito nella forma $r = \Sigma\,x\,y : \sqrt{\Sigma\,x^2\,\Sigma\,y^2}$.

Il coefficiente di correlazione risulta positivo, nullo o negativo secondo che è positiva o negativa la somma $\Sigma\,x\,y$ dei prodotti degli scostamenti che si corrispondono delle due variabili originarie dalle rispettive medie. Questa somma a sua volta risulta positiva o negativa secondo che prevalgono combinazioni fra scostamenti di ugual segno o fra scostamenti di segno opposto; risulta nulla quando le combinazioni dei due tipi si compensano. Sicchè, quando le due variabili tendono a variare nello stesso senso il coefficiente di correlazione risulta positivo, quando tendono a variare in senso opposto risulta negativo.

La somma dei quadrati delle differenze dei valori interpolati dai valori osservati, nelle due interpolazioni eseguite, si può rispettivamente ridurre alla forma:

$$\Sigma\,(a\;x - y)^2 = \Sigma\,y^2\,(1 - r^2)$$
$$\Sigma\,(a'\,y - x)^2 = \Sigma\,x^2\,(1 - r^2)\,.$$

Poichè il primo membro è necessariamente positivo, la differenza $(1 - r^2)$ non può mai divenire negativa; ossia r non può mai superare in valore assoluto l'unità. Esso diviene uguale all'unità quando tutti i valori interpolati coincidono coi corrispondenti valori osservati, cioè nel caso di dipendenza completa; si annulla, come sappiamo, quando si annulla l'espressione $\Sigma x y$, come avviene nel caso d'indipendenza completa fra le due variabili.

Per queste sue proprietà r viene assunto come indice della relazione fra le due variabili: fornisce ad un tempo col suo valore assoluto un indice del grado di dipendenza (e come tale lo abbiamo criticato nel paragrafo 6) e col suo segno un indice del senso della dipendenza (e come tale merita fede soltanto quando la funzione lineare rappresenta con buona approssimazione la relazione esistente fra le due variabili).

Nel nostro esempio del paragrafo 7, indicando con x lo scostamento del voto in economia e con y lo scostamento del voto in statistica dalle rispettive medie aritmetiche, si ha: $\Sigma x y = 7.370$, $\sqrt{\Sigma x^2} = 113{,}8$, $\sqrt{\Sigma y^2} = 110{,}6$; e quindi risulta $r = 7.370 : 12.586 = + 0{,}586$. Il valore assoluto di questo coefficiente di correlazione poteva calcolarsi anche col mettere in rapporto la media quadratica delle variazioni dipendenti, 2,05, con la media quadratica delle variazioni risultanti 3,50; si sarebbe ottenuto lo stesso valore, data l'equivalenza delle due formole per la determinazione del valore assoluto di r.

Volendo calcolare con procedimento parallelo un indice del grado d'indipendenza, dovremo formare il rapporto fra la media quadratica delle variazioni dipendenti, 2,83, e la media quadratica delle variazioni risultanti, 3,50. Questo rapporto è uguale al quoziente $\sqrt{\Sigma(a x - y)^2} : \sqrt{\Sigma y^2}$, il quale a sua volta, come risulta dalla relazione sopra riportata, è uguale a $\sqrt{1 - r^2}$. Un criterio parallelo a quello che fa assumere il valore assoluto di r come indice del grado di dipendenza conduce dunque ad assumere $\sqrt{1 - r^2}$ come indice del grado d'indipendenza. La somma di questi due indici è sempre maggiore dell'unità, eccettuati soltanto i due casi estremi nei quali si ha $r = 0$ o $r = 1$; perciò il criterio secondo il quale i due indici sono determinati non può, a nostro modo di vedere, essere accettato. Nel nostro esempio abbiamo trovato $r = 0{,}586$ come indice del grado di dipendenza; troviamo ora $\sqrt{1 - r^2} = 0{,}810$ come indice del grado di indipendenza: il primo indice sembra mostrare che 59 % delle variazioni del voto di statistica dipendano da quelle del voto di economia, il secondo indice sembra mostrare che 81 % non ne dipendano: i due indici sono reciprocamente inconciliabili.

La contraddizione si può eliminare (anche servendosi di medie quadratiche) col ricorrere al nostro criterio. Il rapporto fra la media quadratica delle variazioni dipendenti, 2,05, e la somma della media stessa e della media quadratica delle variazioni indipendenti ci dà un indice del grado di dipendenza uguale a 0,420. Il rapporto fra la media quadratica delle variazioni indipendenti, 2,83, e lo stesso denominatore ci dà un indice del grado d'indipendenza uguale a 0,580: la somma dei due indici è uguale all'unità: le loro indicazioni sono perfettamente concordanti e non differiscono molto da quelle che abbiamo trovate alla fine del paragrafo 7, impiegando invece di medie quadratiche medie aritmetiche dei valori assoluti.

12. — Si confrontino gli indici del grado di dipendenza calcolati nel pre-

cedente esercizio con quelli calcolati dagli stessi dati nel paragrafo 7 e nell'esercizio 5.

13. — Si calcoli il coefficiente di correlazione tra la frequenza degli errori dei dattilografi e la durata del lavoro. In questo esempio corrisponde al vero l'ipotesi di una relazione lineare tra le due variabili? In caso negativo, come si può scegliere un'ipotesi più plausibile, fondandola sui dati dell'osservazione? Scelta tale ipotesi, determinata cioè una forma di relazione diversa da quella lineare, si calcoli l'indice del grado di dipendenza col nostro criterio. (Vedasi il precedente esercizio 7).

14. — La seguente tabella indica i risultati di due prove di salto — in lungo e in alto — eseguite dagli stessi ginnasti.

Ginnasti, secondo i risultati di due prove di salto.

Salto in lungo (centimetri)	Salto in alto (centimetri)										Totale
	163	166	169	172	175	178	181	184	187	190	
540	4	1	—	1	—	1	—	—	—	—	7
550	2	1	—	—	2	3	—	1	—	—	9
560	1	3	3	2	—	—	—	—	1	—	10
570	2	—	2	4	2	1	3	—	—	—	14
580	1	4	—	3	1	2	—	—	—	1	12
590	3	5	2	—	—	4	—	—	1	—	15
600	1	1	4	2	7	1	—	1	—	—	17
610	—	1	1	2	3	4	—	—	—	1	12
620	—	2	1	2	1	5	2	1	—	—	14
630	3	—	—	—	3	1	4	2	—	—	13
640	—	1	3	1	—	4	2	—	1	—	12
650	—	1	2	3	1	4	5	1	1	—	18
660	—	—	1	1	4	1	3	2	—	1	13
670	2	1	—	—	1	1	2	1	2	—	10
680	—	—	—	1	3	2	3	—	1	—	10
690	—	—	—	—	1	—	2	2	3	—	8
700	—	—	1	—	—	—	—	2	1	2	6
Totale .	19	21	20	22	29	34	26	13	11	5	200

Si calcolino indici del grado di dipendenza fra le due variabili (lunghezza raggiunta dal ginnasta nel salto in lungo e altezza raggiunta nel salto in alto). Si calcoli il coefficiente di correlazione.

Si rappresenti in diagramma l'andamento dei valori medi di ciascuna variabile in funzione dei valori dell'altra. In questo caso appare plausibile l'ipotesi della relazione lineare fra le due variabili?

15. — La seguente tabella indica i risultati di un esperimento di marcia, anteriormente al quale i marciatori avevano ingerito quantità diverse di alcool.

Marciatori, secondo la quantità di alcool ingerita e la velocità media oraria tenuta.

Alcool ingerito (centimetri cubi)	Velocità oraria (ettometri)													Totale
	60	61	62	63	64	65	66	67	68	69	70	71	72	
0	—	—	—	1	—	1	1	—	3	6	4	2	2	20
15	—	1	1	2	—	2	—	1	2	5	5	1	—	20
30	1	—	1	2	1	2	3	4	1	3	1	1	—	20
45	—	2	—	3	2	4	6	1	2	—	—	—	—	20
60	—	1	1	8	5	2	1	1	—	1	—	—	—	20
Totale.	1	4	3	16	8	11	11	7	8	15	10	4	2	100

Si calcoli l'indice del grado di dipendenza e il coefficiente di correlazione. Si riscontra in questo esempio correlazione diretta?

Si cerchi di esprimere in modo semplice ed efficace le conclusioni che possono desumersi dall'esperimento riassunto nella precedente tabella. In questo esempio possono ritenersi corrispondenti al vero le ipotesi sulle quali è fondato il procedimento degli indici del grado di dipendenza?

Si rappresenti graficamente la relazione fra la quantità di alcool ingerita e la velocità raggiunta.

16. — In un sistema di coordinate cartesiane, si misuri sull'asse delle ascisse il voto in economia, sull'asse delle ordinate il voto in statistica; e si tracci la retta ascendente, determinata col metodo dei minimi quadrati, che rappresenta la relazione tra le due variabili (v. pag. 292); si tracci inoltre la retta orizzontale che rappresenta la media aritmetica dei voti in statistica. Considerando ora sei studenti che hanno riportato rispettivamente i seguenti voti nelle due materie: I. E. 20, S. 27; II. E. 24, S. 19; III. E. 21, S. 21; IV. E. 21, S. 18; V. E. 30, S. 24; VI. E. 30, S. 30; si traccino graficamente per ciascuno di questi casi individuali: la variazione dipendente, la variazione non dipendente, la variazione risultante.

LIBRO TERZO
L'interpretazione dei fenomeni

—

CAPITOLO XXI
La stabilità statistica. Le variazioni non significative.

Un problema da risolvere: se e quali relazioni esistano fra il presentarsi di uniformità e il numero dei casi osservati — Ricerca empirica della soluzione: esame del variare, col numero delle osservazioni, delle oscillazioni che si manifestano nella misura di alcuni fenomeni — Come tendano a restringersi tali oscillazioni col crescere del numero dei casi osservati; le variazioni non significative; la legge dei grandi numeri; la stabilità statistica — Alcuni caratteri delle variazioni non significative: loro distribuzione per segno e per grandezza — Criteri che permettono di distinguere le variazioni non significative da quelle significative — Considerazioni sul significato di alcuni metodi di descrizione dei fenomeni collettivamente tipici.

1. — Gli elementi raccolti mediante l'osservazione e coordinati mediante la descrizione ci forniscono rappresentazioni dei fenomeni collettivamente tipici e delle variazioni che questi presentano al modificarsi delle circostanze d'osservazione. Tali rappresentazioni rendono possibile la ricerca di uniformità e di relazioni, e noi dovremo in quest'ultima parte del nostro corso studiare appunto i metodi che si applicano in codesta ricerca. Ma s'impone un'indagine preliminare, la cui necessità abbiamo accennato fino dal principio della nostra trattazione. Abbiamo definito fenomeno « collettivamente tipico » quello che presenta uniformità nell'osservazione di collettività di casi, ma non abbiamo chiarito che cosa debba intendersi per « collettività di casi ». Quanti casi occorrono per costituire una collettività che consenta l'accertamento di uniformità? E quali relazioni esistono tra il numero dei casi osservati e il presentarsi di uniformità? Sappiamo che un dato uomo di vent'anni può morire

prima di un dato uomo di novant'anni; ma sappiamo anche che in un numeroso gruppo di ventenni la frequenza delle morti è molto minore che in un numeroso gruppo di novantenni. *Un* caso non basta a mostrarci l'uniformità, *molti* casi bastano; ma quand'è che più casi si possono chiamare, al nostro fine, « molti »? E come influisce sul manifestarsi dell'uniformità il numero dei casi osservati?

Chiediamo una risposta all'esperienza: proviamo cioè ad osservare lo stesso fenomeno in campi di differente ampiezza, e guardiamo se e come si modifichino le manifestazioni di uniformità. Per semplificare la ricerca, consideriamo un fenomeno, che, attraverso un'esperienza ormai secolare, si è dimostrato abbastanza stabile nella sua misura: la mascolinità delle nascite, cioè la proporzione dei maschi sul totale dei nati vivi. Se considerassimo invece un fenomeno tendente a variare in un determinato senso, come la mortalità o la natalità, che negli ultimi decenni si sono andate abbassando, ci riuscirebbe difficile discernere le variazioni nelle quali si esprime la tendenza dominante da quelle eventualmente connesse col maggiore o minor numero delle osservazioni.

Ecco, anzitutto, le proporzioni percentuali dei maschi sui nati vivi in Italia nei dieci anni dal 1903 al 1912; sotto ciascuna di esse abbiamo segnato il suo scostamento dalla media decennale, che è 51,34.

ITALIA	1903	1904	1905	1906	1907	1908	1909	1910	1911	1912
Maschi per 100 nati	51,40	51,38	51,35	51,38	51,40	51,35	51,23	51,27	51,35	51,31
Scostamenti. . . .	+ 0,06	+ 0,04	+ 0,01	+ 0,04	+ 0,06	+ 0,01	— 0,11	— 0,07	+ 0,01	— 0,03

Il numero dei casi osservati è molto grande: in media 1.097.193 nascite all'anno. Le differenze fra i dati della serie sono così minuscole che essi possono dirsi quasi uguali fra loro: il campo di variazione è contenuto fra il minimo di 51,23 e il massimo di 51,40; lo scostamento medio assoluto 0,04 è inferiore ad un millesimo della media.

Passiamo ad una regione italiana: la Campania. La percentuale media dei maschi nel decennio è poco differente da quella italiana: 51,31. Riferiamo i dati per i singoli anni.

CAMPANIA	1903	1904	1905	1906	1907	1908	1909	1910	1911	1912
Maschi per 100 nati	51,66	51,40	51,42	51,37	51,17	51,29	51,13	51,15	51,47	51,08
Scostamenti	+ 0,35	+ 0,09	+ 0,11	+ 0,06	— 0,14	— 0,02	— 0,18	— 0,16	+ 0,16	— 0,23

Il numero dei casi osservati è circa dieci volte minore che in Italia, poichè i nati nella Campania sono, in media annua, 105.221. Le differenze fra i dati della serie sono ancora piccole; tuttavia il campo di variazione è più esteso che nel precedente esempio: va da un minimo di 51,08 ad un massimo di 51,66: lo scostamento medio assoluto sale a 0,15.

Restringiamo ancora il numero delle osservazioni, prendendo in esame i dati per una provincia: quella di Mantova, dove la percentuale media dei nati maschi nel solito decennio è 51,75.

MANTOVA	1903	1904	1905	1906	1907	1908	1909	1910	1911	1912
Maschi per 100 nati	52,45	51,87	51,92	51,66	51,66	52,11	51,89	51,50	50,69	51,73
Scostamenti . . .	+ 0,70	+ 0,12	+ 0,17	— 0,09	— 0,09	+ 0,36	+ 0,14	— 0,25	— 1,06	— 0,02

Il numero dei casi osservati, 11.051, è circa cento volte minore che in Italia. Le differenze fra i dati della serie sono maggiori che negli esempi precedenti; il campo di variazione si allarga: da un minimo di 50,69 si va fino ad un massimo di 52,45; lo scostamento medio assoluto ascende a 0,30.

Passando ad un comune, quello di Massa, otteniamo un'ulteriore forte riduzione del numero dei casi osservati. La percentuale media dei maschi nelle nascite è 51,15.

MASSA	1903	1904	1905	1906	1907	1908	1909	1910	1911	1912
Maschi per 100 nati	50,20	52,14	51,30	51,31	50,74	49,24	52,55	51,80	52,44	49,79
Scostamenti . . .	— 0,95	+ 0,99	+ 0,15	+ 0,16	— 0,41	— 1,91	+ 1,40	+ 0,65	+ 1,29	— 1,36

Il numero medio annuo delle nascite, 1.180, corrisponde a circa un millesimo di quello osservato in Italia. Le differenze fra i dati cominciano a divenire sensibili: il campo di variazione si estende da un minimo di 49,24 ad un massimo di 52,55; lo scostamento medio assoluto balza a 0,93.

Un altro comune, quello di Gaeta, ci offre un campo d'osservazione ancor più ristretto. La proporzione media dei maschi è di 50,11 per 100 nati.

GAETA	1903	1904	1905	1906	1907	1908	1909	1910	1911	1912
Maschi per 100 nati	43,30	56,19	47,73	50,50	46,74	52,04	51,16	48,49	54,95	50,00
Scostamenti . . .	— 6,81	+ 6,08	— 2,38	+ 0,39	— 3,37	+ 1,93	+ 1,05	— 1,62	+ 4,84	— 0,11

Il numero medio annuo delle nascite è soltanto di 95. Le differenze fra i dati annuali sono considerevoli: il campo di variazione è compreso fra il minimo di 43,30 e il massimo di 56,19; lo scostamento medio assoluto raggiunge 2,86, cioè 57 millesimi della media.

Possiamo ora fermarci e riassumere i risultati del nostro esame. Il fenomeno la cui stabilità ci appare solo lievemente turbata da minuscole oscillazioni quando consideriamo un grandissimo numero di osservazioni, ci si presenta alterato da oscillazioni d'ampiezza crescente col progressivo restringersi del numero dei casi osservati. Siamo tratti ad indurre che in un paese con un numero di nascite molto maggiore di quello italiano vedremmo quasi scomparire le oscillazioni delle percentuali di mascolinità, mentre in una popolazione con numero di nascite molto minore di quello di Gaeta vedremmo ampliarsi tali oscillazioni così da rendere difficile l'accertamento della tendenziale uniformità.

2. — Una seconda serie di esempi ci sarà fornita da un fenomeno di tutt'altra natura: esamineremo alcuni risultati del giuoco del lotto. Non sembri strana la scelta: il risultato dell'estrazione del lotto, come quello d'un altro qualsiasi giuoco di sorte, rientra nella nostra definizione di fenomeno collettivamente tipico. Aggiungasi che le condizioni di osservazione, per la parte in cui vengono preordinate dall'opera umana, sono note e sono mantenute costanti, così che si può assimilare il giuoco ad un vero e proprio esperimento scientifico.

I seguenti dati indicano le proporzioni percentuali nelle quali si sono presentati nelle estrazioni i numeri compresi rispettivamente fra 1 e 9, fra 10 e 18, fra 19 e 27, ecc., in 41.810 estrazioni (cioè in 209.050 numeri sorteggiati) di tutte le Ruote italiane. Indichiamo anche gli scostamenti dalla percentuale media, che, per l'ordinamento stesso del giuoco, è necessariamente uguale a 10.

(209.050 sorteggi)	1-9	10-18	19-27	28-36	37-45	46-54	55 63	64-72	73 81	82-90
Percent. accertate	10,03	10,07	10,04	10,04	10,08	10,04	9,94	9,96	9,80	10,00
Scostamenti . . .	+ 0,03	+ 0,07	+ 0,04	+ 0,04	+ 0,08	+ 0,04	— 0,06	— 0,04	— 0,20	0

Le percentuali variano fra un minimo di 9,80 e un massimo di 10,07, cioè entro limiti molto ristretti. Lo scostamento medio assoluto è soltanto di 0,06.

Restringiamo il numero dei casi osservati: 1.440 estrazioni (cioè 7.200 numeri sorteggiati) della Ruota di Palermo ci danno i seguenti risultati.

(7.200 sorteggi)	1-9	10-18	19-27	28-36	37-45	46-54	55-63	64-72	73-81	82-90
Percent. accertate.	10,77	9,49	10,60	10,08	10,11	9,68	9,83	9,78	9,65	10,01
Scostamenti . .	+ 0,77	— 0,51	+ 0,60	+ 0,08	+ 0,11	— 0,32	— 0,17	— 0,22	— 0,35	+ 0,01

Il minimo scende a 9,49, il massimo sale a 10,77; il campo di variazione è molto più ampio che nell'esempio precedente; lo scostamento medio assoluto sale a 0,31.

Consideriamo infine una serie di estrazioni ancor meno numerosa: 144 (cioè 720 sorteggi) della stessa Ruota di Palermo. Otteniamo le seguenti percentuali.

(720 sorteggi)	1-9	10-18	19-27	28-36	37-45	46-54	55-63	64-72	73-81	82-90
Percent. accertate.	10,69	10,28	11,81	9,44	11,25	8,06	8,89	9,17	10,83	9,58
Scostamenti . . .	+ 0,69	+ 0,28	+ 1,81	— 0,56	+ 1.25	— 1,94	— 1,11	— 0,83	+ 0,83	— 0,42

I limiti del campo di variazione, 8,06 minimo e 11,81 massimo, sono molto più lontani che nei precedenti esempi; lo scostamento medio assoluto aumenta a 0,97.

Anche in questi esempi, come in quelli del paragrafo 1, col restringersi del numero dei casi osservati vediamo di mano in mano oscillare più ampiamente la misura del fenomeno.

3. — Scegliamo in un altro campo una terza serie di esempi, ricorrendo alle statistiche della leva. Queste ci indicano, fra altro, il numero degli iscritti di leva sottoposti a visita medica e i numeri dei riformati suddivisi secondo la causa della riforma. Abbiamo calcolato per dieci leve successive la proporzione dei dichiarati inabili al servizio militare per la mancanza totale d'una mano o di un piede, su ogni 10.000 visitati.

In Italia il numero medio degli iscritti di leva visitati per ciascuna delle classi di leva considerate è 410.240; la proporzione media dei riformati per mancanza di un arto è di 6,9 per 10.000. Ecco i dati per le singole classi, e i loro scostamenti dalla media.

ITALIA	1883	1884	1885	1886	1887	1888	1889	1890	1891	1892
Riformati per 10.000 visitati	6,9	6,2	6,8	7,1	7,9	6,9	7,0	6,7	7,0	6,8
Scostamenti . . .	0	− 0,7	− 0,1	+ 0,2	+ 1,0	0	− 0,1	− 0,2	+ 0,1	− 0,1

Lo scostamento medio assoluto 0,25 corrisponde a meno di 4 centesimi della media: il campo di variazione è contenuto fra il minimo di 6,2 e il massimo di 7,9.

Calcoliamo ora le corrispondenti proporzioni per una singola provincia: quella di Arezzo, che annovera in media annua 3.884 visitati, cioè meno della centesima parte del totale italiano. La proporzione media dei riformati per mancanza di un arto, 6,0 per 10.000, non si discosta molto dalla media nazionale. I dati per singole classi di leva presentano fortissime oscillazioni.

AREZZO	1883	1884	1885	1886	1887	1888	1889	1890	1891	1892
Riformati per 10.000 visitati	5,5	11,5	8,1	2,6	2,5	2,7	4,5	5,1	7,4	9,8
Scostamenti . . .	− 0,5	+ 5,5	+ 2,1	− 3,4	− 3,5	− 3,3	− 1,5	− 0,9	+ 1,4	+ 3,8

Il campo di variazione si è grandemente allargato, poichè va da un minimo di 2,5 ad un massimo di 11,5; lo scostamento medio assoluto 2,6 corrisponde a 43 centesimi della media.

4. — Sarebbe facile moltiplicare gli esempi; ma preferiamo che il lettore, studiando ed elaborando personalmente le statistiche demografiche, biologiche, fisiche, o i risultati di giuochi di sorte, si procuri da sè nuove conferme della regolarità che abbiamo messo in risalto: così ne resterà maggiormente persuaso. D'altronde i nostri esempi non fanno che precisare una nozione di comune dominio: quella che le uniformità dei fenomeni collettivamente tipici appaiono tanto meglio quanto maggiore è il numero dei casi osservati. Diciamo «precisare», perchè esaminando i dati da noi riferiti si può vedere non solo che le oscillazioni tendono ad attenuarsi col crescere del numero delle osservazioni, ma anche in qual misura esse tendano ad attenuarsi.

Ripercorrendo in ordine inverso gli esempi di oscillazioni della mascolinità delle nascite esposti nel paragrafo 1, vediamo subito che l'ampiezza del campo di variazione delle percentuali e la grandezza del loro scostamento medio assoluto non crescono in ragione inversa del numero dei casi osservati, ma molto più adagio.

Popolazione osservata	Numero dei casi osservati	Campo di variazione della mascolinità	Scostamento medio assoluto
Comune di Gaeta . . .	95	12,89	2,86
Comune di Massa . . .	1.180	3,31	0,93
Provincia di Mantova . .	11.051	1,76	0,30
Campania	105.221	0,58	0,15
Italia	1.097.193	0,17	0,04

Certo non si riscontra una relazione rigorosa fra le variazioni dei dati delle ultime due colonne e quelle dei dati della prima; ma *grosso modo* si può dire che gli uni variino press'a poco in ragione inversa della radice quadrata degli altri. Passando dall'Italia alla provincia di Mantova il numero delle osservazioni si riduce a circa un centesimo, l'ampiezza delle oscillazioni — sia nella misura data dall'estensione del campo di variazione, sia in quella data dallo scostamento medio — aumenta circa dieci volte; passando dalla provincia di Mantova al comune di Gaeta il numero delle osservazioni si riduce a circa un centesimo, l'ampiezza delle oscillazioni aumenta circa dieci volte. Lo scostamento medio assoluto y deve dunque potersi rappresentare con sufficiente approssimazione in funzione del numero x dei casi osservati, mediante una relazione del tipo $y = a : \sqrt{x}$. Determinando col metodo delle somme il parametro a lo troviamo uguale a 29,5 e dall'equazione $y = 29{,}5 : \sqrt{x}$ determiniamo i valori interpolati dello scostamento medio assoluto: 3,02; 0,86; 0,28; 0,09; 0,03, i quali non differiscono molto dai valori osservati, raccolti nell'ultima colonna della tabellina precedente.

Negli esempi del paragrafo 2, tratti dal giuoco del lotto, vediamo che restringendo circa trenta volte il numero dei casi osservati lo scostamento medio assoluto aumenta poco più di cinque volte; che restringendo ancora dieci volte il numero dei casi osservati lo scostamento medio aumenta poco più di tre volte. Si ritrova anche qui l'approssimativa proporzionalità inversa tra l'ampiezza delle oscillazioni e la radice quadrata del numero dei casi osservati.

Esempio	Numero dei sorteggi	Campo di variazione delle percentuali	Scostamento medio assoluto
I	209.050	0,27	0,06
II	7.200	1,28	0,31
III	720	3,75	0,97

E, infine, negli esempi del paragrafo 3, tratti dalle statistiche della leva, il numero dei casi osservati nella provincia d'Arezzo è circa cento volte minore che nell'insieme del Regno; lo scostamento medio assoluto è circa dieci volte maggiore.

Siamo dunque in grado di esprimere in forma precisa i risultati del nostro esame. Considerando rapporti tra i numeri dei casi di alcuni fenomeni contrassegnati da certi caratteri e i corrispondenti numeri totali di casi osservati, abbiamo trovato che le oscillazioni di questi rapporti intorno alla loro media tendono a variare in ragione inversa della radice quadrata del numero dei casi osservati: tendono cioè a restringersi indefinitamente coll'indefinito crescere del numero dei casi osservati.

Possiamo ora intendere il significato della così detta « legge dei grandi numeri » assunta come formulazione di una regolarità empirica: soltanto da grandi numeri di osservazioni intorno a fenomeni collettivamente tipici si possono ottenere misure medie dei fenomeni stessi che siano approssimativamente costanti.

Chiameremo da ora innanzi *variazioni non significative* quelle oscillazioni delle misure medie di fenomeni collettivamente tipici che tendono ad attenuarsi progressivamente col crescere del numero dei casi osservati, fino a divenire piccolissime quando tale numero diviene grandissimo: oscillazioni che appaiono non inerenti alla natura del fenomeno osservato ma solo dipendenti dal numero dei casi osservati. Esse vengono spesso denominate *variazioni accidentali.*

Quando un fenomeno presenta esclusivamente variazioni non significative si dice che esso è « statisticamente stabile ». La *stabilità statistica* non è dunque una vera e propria stabilità nel senso di costanza rigorosa della misura di manifestazione di un fenomeno; essa consiste invece in una tendenza del fenomeno alla costanza: tendenza che si manifesta sempre più palese col crescere del numero delle osservazioni, ma che non raggiunge mai la sua meta per quanto grande sia tale numero.

5. — Qui apriamo una parentesi. Se invece di considerare, co-

me abbiamo fatto, i rapporti tra i numeri dei casi contrassegnati da certi caratteri ed i numeri dei casi osservati, avessimo considerato i numeri assoluti dei primi, come sarebbe stata espressa la regolarità che abbiamo trovata?

Per rendercene conto consideriamo ancora due serie di estrazioni del lotto, quella di 7.200 sorteggi e quella di 720; ma invece di prendere in esame i rapporti fra il numero di uscite di numeri compresi fra determinati limiti e il numero totale dei sorteggi, prendiamo in esame i dati assoluti, cioè i numeratori di tali rapporti. I seguenti dati indicano appunto quanti dei numeri sorteggiati siano compresi rispettivamente fra 1 e 9, fra 10 e 18, fra 19 e 27, ecc.

Num. dei sorteggi	1-9	10-18	19-27	28 36	37-45	46-54	55-63	64-72	73-81	82-90
7.200	775	683	763	726	728	697	708	704	695	721
720	77	74	85	68	81	58	64	66	78	69

Per la prima serie il campo di variazione si estende dal minimo di 683 al massimo di 775: ha dunque un'ampiezza di 92 unità; lo scostamento medio assoluto è uguale a 22,6. Per la seconda serie il campo d'osservazione si estende da 58 a 85, comprendendo un intervallo di 27 unità; lo scostamento medio è uguale a 7,0. Mentre il numero dei casi osservati diminuisce nella proporzione di 10 a 1, l'ampiezza delle oscillazioni si riduce nella misura di 3,3 a 1, cioè varia all'incirca in ragione diretta della radice quadrata del numero dei casi osservati.

Una breve riflessione basta a mostrare che abbiamo semplicemente espresso in altra forma la regolarità già accertata. Dire che lo scostamento medio assoluto s dei risultati assoluti (numeri di casi contrassegnati da certi caratteri) tende a variare in ragione diretta della radice quadrata del numero n dei casi osservati equivale a dire che lo scostamento medio assoluto (s: n) dei risultati relativi (rapporti fra casi contrassegnati da certi caratteri e casi osservati) tende a variare in ragione inversa della radice quadrata di n. Infatti dalla relazione $s = a\sqrt{n}$, dove a è un coefficiente di proporzionalità, si desume l'altra $s : n = a : \sqrt{n}$.

6. — Un'altra, più breve, parentesi. Abbiamo parlato nei paragrafi precedenti di « numero dei casi osservati ». Ma sappiamo che molte volte l'oggetto dell'osservazione non consiste di unità individual-

mente distinte. Vale anche in questa ipotesi la legge dei grandi numeri?

Certamente vale, e l'esperienza dimostra che invece di parlare di «grande numero delle osservazioni», quale condizione per il miglior accertamento delle uniformità statistiche, si può parlare in generale di «grande estensione delle osservazioni». Supponiamo di disporre di duemila ettari di terreno perfettamente omogeneo; seminiamo su mille ettari una data varietà di grano, sugli altri mille un'altra varietà: otterremo nelle due zone due diversi rendimenti medi per ettaro, che, se la coltivazione è stata condotta con metodi uguali nelle due zone, ci indicheranno con buona approssimazione la differente produttività delle due sementi. Se dividiamo ciascuna delle due zone in tante sottozone di cento ettari ciascuna, troveremo rendimenti diversi in ogni singola sottozona, più o meno dispersi intorno alla media della corrispondente zona. Se dividiamo ancora ogni sottozona in dieci appezzamenti di dieci ettari ciascuno, troveremo oscillazioni anche maggiori; se dividiamo gli appezzamenti da dieci ettari in campi di un ettaro, l'ampiezza delle oscillazioni del rendimento aumenterà ancora, e così via. Viceversa, se potessimo disporre di tante zone di terreno omogeneo, non di soli mille ettari ciascuna, ma di diecimila, di centomila, e così via crescendo, vedremmo sempre più attenuarsi le oscillazioni dei rendimenti medi di ciascuna varietà di grano: potremmo cioè determinare con crescente precisione il rendimento medio tipico di ciascuna varietà.

7. — Nello studio dei fenomeni reali, importa discernere le variazioni non significative, che non presentano interesse per l'indagine sui fenomeni stessi, da quelle significative, che sono le sole rilevanti agli effetti di tale indagine. Questa separazione, non agevole nella pratica, può essere facilitata dalla conoscenza di alcuni caratteri delle variazioni non significative.

Un primo carattere è quello che le variazioni non significative tendono a presentarsi in numero press'a poco uguale nei due sensi: positivo e negativo.

Un secondo carattere consiste in una certa regolarità nella loro distribuzione per grandezza.

I due caratteri insieme si riassumono nella constatazione più precisa che la distribuzione per grandezza e per segno delle variazioni non significative si presta ad essere rappresentata mediante la curva degli errori (v. pagg. 136-137), se, misurando sull'asse delle ascisse la grandezza delle variazioni, si fa corrispondere la variazione nulla all'origine delle ascisse, mentre sull'asse delle ordinate si misura il numero delle variazioni aventi una data grandezza.

Per dare un esempio della distribuzione delle variazioni non significative, conviene prendere serie composte di maggior numero di termini di quelle riportate nei paragrafi precedenti. Ci varremo di dieci gruppi di estrazioni del lotto comprendenti ciascuno 144 estrazioni (720 sorteggi); al nostro fine possiamo riguardare come una sola serie di cento dati l'insieme delle dieci serie di dieci dati ciascuna.

INTERVALLO	Uscite di numeri compresi nell'intervallo indicato, nel gruppo di estrazioni									
	I	II	III	IV	V	VI	VII	VIII	IX	X
da 1 a 9.	80	77	67	76	80	84	64	78	87	82
da 10 a 18.	69	74	77	75	59	77	65	55	63	69
da 19 a 27.	78	85	71	76	85	72	71	63	82	80
da 28 a 36.	77	68	77	69	74	69	78	73	72	69
da 37 a 45.	69	81	76	60	74	71	80	74	71	72
da 46 a 54.	75	58	52	81	82	64	75	75	62	73
da 55 a 63.	64	64	78	69	81	62	76	81	69	64
da 64 a 72.	70	66	73	72	57	78	68	63	68	89
da 73 a 81.	63	78	80	82	70	70	74	68	59	51
da 82 a 90.	75	69	69	60	58	73	69	90	87	71

La media aritmetica dei cento dati è 72; il numero degli scostamenti positivi (49) differisce poco da quello dei negativi (47); vi sono 4 scostamenti nulli. Lo scostamento medio assoluto è 6,5.

Gli scostamenti piccoli sono più frequenti dei grandi: con maggior esattezza di espressione si può dire che il numero degli scostamenti uguali in valore assoluto ad x tende a decrescere rapidamente col crescere di x. Dei 100 scostamenti, 58 non superano in valore assoluto lo scostamento medio assoluto, altri 32 non superano il doppio di esso, altri 8 non ne superano il triplo e solo 2 lo superano (le corrispondenti percentuali caratteristiche di una distribuzione perfettamente conforme alla curva degli errori sarebbero 57,5; 31,4; 9,4; 1,7; come si vede, le percentuali accertate concordano bene con quelle teoriche).

8. — Ricapitoliamo: Le misure dei fenomeni collettivamente tipici presentano, fra le altre, certe variazioni le quali nei dati asso-

luti tendono ad assumere una grandezza media direttamente proporzionale alla radice quadrata del numero dei casi osservati, e nei dati relativi (rapporti tra dati assoluti e numero dei casi osservati) tendono ad assumere una grandezza media inversamente proporzionale alla radice quadrata del numero dei casi osservati. Queste variazioni, che chiamiamo «non significative» perchè tendono ad eliminarsi, nei dati relativi, coll'indefinito crescere del numero dei casi osservati, così che appaiono nettamente indipendenti dalla particolare natura del fenomeno osservato, tendono a distribuirsi secondo il segno e la grandezza in modo adeguatamente rappresentato dalla curva degli errori.

Qualunque uniformità dei fenomeni collettivamente tipici, qualunque relazione tra la manifestazione di essi e le circostanze di osservazione, qualunque relazione tra fenomeni diversi, ci appare sempre dissimulata dalle variazioni non significative, le quali conferiscono andamento oscillante a tutte le serie statistiche. Poichè queste variazioni tendono a diminuire d'ampiezza relativa col crescere del numero dei casi osservati, divengono nel tempo stesso sempre meglio visibili le uniformità, le relazioni, le variazioni significative.

9. — Da codesti insegnamenti dell'esperienza si può trarre qualche norma pratica per l'interpretazione delle variazioni di dati statistici:

a) Cercar di estendere il più possibile il numero delle osservazioni per giungere a risultati più attendibili; non trarre, di regola, conclusioni da piccoli numeri d'osservazioni, ma solo trarne ipotesi da verificare con ulteriori esperienze.

b) Per quanto sia grande il numero delle osservazioni, non dimenticare mai che sussistono ancora variazioni non significative; quindi non dare per sicure conclusioni che potrebbero essere smentite da ulteriori esperienze.

c) Per discernere la natura di variazioni accertate, ricercare come esse si comportino col variare del numero dei casi osservati; se la loro ampiezza media assoluta varia press'a poco proporzionalmente al numero dei casi osservati (e quindi la loro ampiezza relativa si mantiene press'a poco costante), vi è ragione per ritenerle significative; se la loro ampiezza media assoluta varia press'a poco in ragione diretta (e quindi la loro ampiezza relativa in ragione inversa) della radice quadrata del numero dei casi osservati, vi è ragione per ritenerle non significative. Per compiere quest'esame non sarà sempre possibile accrescere il numero dei casi osservati; ma sarà spesso possibile diminuirlo scindendo in più parti il complesso delle osservazioni, e quindi si potrà ugualmente verificare come si modifichi l'ampiezza delle variazioni del fenomeno col variare del numero dei casi osservati.

d) Ancora per discernere la natura di date variazioni accertate, verificare se esse si ripetano in ugual senso e in poco differente misura in diversi tempi, in diversi luoghi, in diversi campi d'osservazione o in diverse sezioni dello stesso campo. Se esperienze ripetute, anche desunta ciascuna da un numero non grande d'osservazioni, mostrano costantemente una data variazione del fenomeno in corrispondenza alla variazione di una data circostanza, vi è ragione per ritenerla significativa.

Nella maggior parte delle applicazioni dei metodi statistici allo studio dei fenomeni sociali, basta tener presenti le precedenti avvertenze per evitare i gravi errori di giudizio cui troppo spesso conduce la confusione tra variazioni non significative e variazioni significative.

La statistica matematica offre metodi che permettono di verificare sistematicamente se date variazioni possano verosimilmente assumersi come significative ovvero si presentino invece coi caratteri delle non significative. L'esposizione di questi metodi esce dai limiti di un corso elementare di statistica, come il nostro; d'altra parte, mentre essi presentano grande importanza teorica, sono raramente applicati nella pratica, ai bisogni della quale per lo più basta l'oculato uso delle regole or ora esposte.

10. — La conoscenza delle variazioni non significative e dei loro caratteri ci permette di aggiungere alle nozioni date nel secondo libro alcune considerazioni.

Nell'interpretare le descrizioni dei fenomeni collettivamente tipici ottenute mediante l'elaborazione dei risultati dell'osservazione non si dovrà mai dimenticare l'esistenza delle variazioni non significative. Ad esempio, non si potrà ritenere senz'altro e per intero significativa la differenza fra due dati statistici, poichè ciascuno di questi è normalmente affetto da variazioni non significative. Analoga avvertenza per il rapporto.

Qualche metodo, che ci era apparso esclusivamente diretto alla descrizione dei fenomeni, ora ci appare anche rivolto alla eliminazione, o meglio all'attenuazione, delle variazioni non significative. Quando calcoliamo una media dei rapporti che indicano la mascolinità delle nascite in più anni successivi, come abbiamo fatto nel paragrafo 1, non solo cerchiamo l'espressione sintetica di una serie di dati, ma ci proponiamo anche di ottenere un'espressione maggiormente approssimata al vero della misura del fenomeno, poichè, considerando i dati di più anni invece di quelli di un anno solo, veniamo ad aumentare il numero dei casi osservati e quindi a diminuire l'ampiezza media delle variazioni non significative dalle quali può essere affetto il rapporto di mascolinità. Quando eseguiamo un'interpolazione, mettendo così in rilievo certe variazioni dei fenomeni e lasciando da parte certe altre, talvolta trascuriamo queste ultime perchè le riteniamo secondarie rispetto alle prime, ma tal altra le trascuriamo perchè le riteniamo addirittura non significative: cioè presumiamo che se potessimo accrescere indefinitamente il numero delle osservazioni ne ricaveremmo una curva regolare come quella interpolata, e non una curva oscillante come quella desunta dalla nostra ristretta esperienza.

NOTA AL CAPITOLO XXI.

Intorno ad una rappresentazione schematica dei fenomeni collettivamente tipici.

1. — Eseguendo una serie di osservazioni intorno ad un fenomeno o ad un carattere collettivamente tipico, lo vediamo presentarsi in modalità o misura variabile da osservazione ad osservazione. Chiamiamo *frequenza relativa* di una determinata modalità o misura del fenomeno, o carattere, il rapporto fra il numero delle osservazioni nelle quali essa si è presentata, e il numero complessivo delle osservazioni eseguite.

In certi casi, come abbiamo visto, al crescere del numero delle osservazioni, gli scostamenti tra le frequenze relative delle varie modalità o misure del fenomeno osservato si vanno attenuando. Così avviene per le frequenze relative con cui si presentano al sorteggio i singoli numeri contenuti nell'urna del lotto; esse variano, per esempio, da un minimo di 0,00139 ad un massimo di 0,02222 in serie di 720 sorteggi (Palermo), da un minimo di 0,01050 a un massimo 0,01186 in 209.050 sorteggi (Italia). Mentre aumenta il numero delle osservazioni, si restringe il campo di variazione delle frequenze relative, da 0,02083 nel primo caso a 0,00136 nel secondo.

Se, fondandoci sopra un imponente corredo di dati sperimentali concordanti, estendiamo, anche oltre i limiti del numero di osservazioni praticamente eseguibile, la validità della relazione accertata, possiamo esprimerla così: all'indefinito crescere del numero delle osservazioni le frequenze relative delle varie modalità o misure del fenomeno tendono ad uguagliarsi fra loro. Per esempio, possiamo dire che, all'indefinito crescere del numero dei sorteggi, le frequenze relative del sorteggio dei vari numeri dall'urna del lotto tendono ad uguagliarsi.

Analizzando le condizioni, o circostanze, di esperimento, in quei casi nei quali l'esperienza mostra che le frequenze relative tendono ad uguagliarsi, siamo condotti a classificarle in due categorie.

Una prima categoria è costituita da *condizioni generali*, rispetto alle quali i vari risultati possibili sono da ritenere *equivalenti*: tali, nel lotto, le dimensioni, la costituzione, il peso, delle sferette riposte nell'urna.

Una seconda categoria è costituita da condizioni *particolari*, rispetto alle quali i vari risultati possibili si possono ritenere *non equivalenti*: tali la posizione iniziale delle sferette, il loro percorso durante la rotazione dell'urna.

Poichè rispetto alle condizioni generali i vari risultati possibili sono equivalenti, il manifestarsi dell'uno di essi in una osservazione, dell'altro in un'altra, non può essere attribuito se non alla differenza delle condizioni particolari sussistenti nei due casi. E dobbiamo imputare soltanto alla ignoranza di queste ultime condizioni la nostra incapacità di prevedere quale risultato si presenterà in ciascuna osservazione.

L'esperienza, quando mostra che col crescere del numero delle osservazioni le frequenze relative dei vari risultati tendono ad uguagliarsi, non illumina l'oscurità in cui ci troviamo riguardo al genere, alla modalità od alla misura delle condizioni particolari presenti in ciascuna singola osservazione. Essa però ci attesta che i vari risultati possibili, — già equivalenti *nell'osservazione singola* rispetto alle condizioni generali —, tendono, *nell'insieme di una numerosa serie di osservazioni*, a divenire equivalenti anche rispetto alle condizioni particolari.

Siamo pertanto indotti a conchiudere che fra le tante possibili combinazioni di condizioni particolari, che possono presentarsi nelle singole osservazioni, siano in proporzioni presso che uguali quelle che mettono capo a ciascuno dei vari risultati: che, per esempio, tra le possibili combinazioni di situazioni iniziali, di successivi percorsi, ecc., delle sferette del lotto, siano in proporzioni all'incirca uguali quelle che conducono all'estrazione di ciascuna delle novanta sferette, ossia rispettivamente al sorteggio del numero uno, del due, del tre, ... del novanta.

La tendenza che mostrano le frequenze relative dei vari risultati possibili, equivalenti fra loro per certe condizioni generali, ad uguagliarsi fra loro col crescere del numero delle osservazioni, suol esprimersi col dire che i vari risultati sono *ugualmente possibili*. Adottato questo termine, non bisogna dimenticarne il significato convenzionale; nè, sopratutto, credere che indichi equivalenza dei vari risultati rispetto a *tutte* le condizioni in ogni singola osservazione. Se esistesse una siffatta equivalenza, o non dovrebbe presentarsi nessuno dei risultati, o dovrebbero presentarsi tutti ad un tempo.

2. — Dall'accertamento concreto della frequenza press'a poco uguale della manifestazione di vari risultati, non riesce difficile passare alla concezione astratta di una pluralità di risultati ugualmente possibili, rigorosamente equivalenti in ciascuna osservazione rispetto ad una certa categoria di condizioni (condizioni generali), e tendenti con l'indefinito crescere del numero delle osservazioni a divenire equivalenti anche rispetto alle altre condizioni (condizioni particolari). Diciamo « concezione astratta », perchè in pratica non è ottenibile *l'assoluta* equivalenza dei vari risultati rispetto alle condizioni generali: non si potranno costruire, p. es., novanta sferette *esattamente* uguali tra loro, per racchiudervi

i novanta numeri del lotto. Ma in teoria si può supporre perfetta l'equivalenza.

Partendo da tale concezione, si possono dedurre le forme di regolarità che presenterebbe la manifestazione di un fenomeno collettivamente tipico, per il quale sussistesse in ciascuna osservazione una pluralità di risultati ugualmente possibili.

3. — Ad ognuno dei risultati ugualmente possibili si può sempre far corrispondere un simbolo numerico. Distinguiamo due casi:

a) ciascun risultato è contrassegnato da una caratteristica quantitativa, che ne costituisce il simbolo numerico ed è oggetto immediato dell'osservazione. Tale, per esempio, il numero contenuto in ciascuna sferetta del lotto;

b) alcuni dei risultati ugualmente possibili possiedono, altri non possiedono, una data caratteristica quantitativa o qualitativa (misura o modalità); col presentarsi di uno dei primi risultati la caratteristica appare una volta, col presentarsi d'uno dei secondi appare zero volte; pertanto si farà corrispondere a ciascuno dei primi il simbolo 1, a ciascuno dei secondi il simbolo 0. Per esempio, classificati i numeri del lotto secondo che superano, o non, 45, si indicherà con 1 ciascun risultato superiore, con 0 ciascun risultato inferiore a codesto limite.

Diremo «*valore medio*» (è usata anche la denominazione di «*valore probabile*», ma la evitiamo perchè equivoca) la media aritmetica dei numeri che rappresentano un insieme di risultati ugualmente possibili. In particolare diremo «valore medio» del risultato di una data osservazione la media aritmetica dei risultati ugualmente possibili in quella osservazione. Nell'ipotesi *b*) fatta or ora, questo «valore medio» indica qual frazione del numero totale dei risultati ugualmente possibili sia contraddistinta da una determinata caratteristica: dicesi allora *probabilità* della caratteristica («probabilità matematica», da non confondersi con la «probabilità statistica»; vedansi pagg. 42, 43). Se p è la probabilità del presentarsi d'una caratteristica, $(1-p)$ è la probabilità del non presentarsi di essa, come risulta immediatamente dalla definizione. Poichè una probabilità può variare fra 0 ed 1, diremo che è piccolissima se molto prossima a zero, che è grandissima se molto prossima ad uno.

Esempi: il «valore medio» del risultato del sorteggio d'un numero dall'urna del lotto è 45,5, media dei novanta risultati ugualmente possibili. La probabilità di un risultato superiore a 45 è 0,5, media di 45 possibili casi in cui si ha 1 volta un risultato superiore a 45 e di 45 casi in cui si ha 0 volte un risultato superiore a 45; è quindi uguale a $1 - 0{,}5 = 0{,}5$ la probabilità di un risultato non superiore a 45. E' piccola la probabilità che esca un dato numero (0,011); è grande quella che esca uno qualsiasi degli altri 89 numeri (0,989).

In una serie di più osservazioni, i risultati ugualmente possibili o rimangono i medesimi o variano da osservazione ad osservazione; il «valore medio» resta costante nel primo caso, può variare nel secondo. Diciamo «può variare», perchè differenti assortimenti di risultati ugualmente possibili possono avere uguali medie. Distingueremo dunque i «valori medi» in *costanti* e *variabili*.

Così il «valore medio» del risultato d'un sorteggio varierebbe da Ruota a Ruota se l'urna di Napoli contenesse i numeri da 1 a 90, quella di Roma i numeri da 7 a 96, quella di Bari i numeri da 71 a 160, ecc.

Immaginiamo eseguita una serie di n osservazioni: la somma dei risultati numerici ottenuti nelle singole osservazioni dicesi *risultato totale*; la media aritmetica *risultato medio* delle n osservazioni.

Per esempio, una delle possibili serie di risultati di cinque sorteggi d'un numero dall'urna del lotto essendo costituita dai numeri 7, 72, 15, 43, 33 ,il corrispondente risultato totale è 170, il risultato medio 34. La stessa serie darebbe 1 come risultato totale e 0,2 come risultato medio a chi ricercasse soltanto se il numero estratto supera 45.

Se ciascuno di m_1 risultati ugualmente possibili in una osservazione può associarsi con ciascuno di m_2 risultati ugualmente possibili in un'altra, $m_1\, m_2$ rappresenta il numero dei risultati (totali o medi) ugualmente possibili delle due osservazioni. E, in generale, il prodotto $m_1\, m_2 \ldots m_n$ rappresenta il numero dei risultati (totali o medi) ugualmente possibili di n osservazioni.

Così se si estraggono successivamente più numeri dall'urna del lotto, senza riporre di volta in volta nell'urna il numero sorteggiato, a ciascuno dei 90 risultati ugualmente possibili del primo sorteggio si associano 89 risultati ugualmente possibili del secondo; si ha quindi $90 \times 89 = 8.010$ come numero dei risultati ugualmente possibili dei due sorteggi. A ciascuno di questi risultati si associano 88 risultati ugualmente possibili nel terzo sorteggio; si ha quindi 704.880 come numero complessivo dei risultati ugualmente possibili dei tre sorteggi, e così via.

La media aritmetica dei risultati (totali o medi) ugualmente possibili di n osservazioni dicesi « valore medio » del risultato (totale o medio) ~~di n~~ osservazioni. Nell'ipotesi *b*) fatta dianzi, il « valore medio » del risultato medio dicesi probabilità media.

Segue immediatamente dalle nostre definizioni che il « valore medio » del risultato medio di n osservazioni è uguale alla media aritmetica (e quindi il « valore medio » del risultato totale è uguale alla somma) dei « valori medi » dei risultati delle singole osservazioni.

Quando il « valore medio » non varia da osservazione ad osservazione, il « valore medio » del risultato medio è uguale al « valore medio » del risultato di ciascuna osservazione.

Il « valore medio » della somma dei numeri estratti in dieci sorteggi dall'urna del lotto è 455; mentre 5 è il « valore medio » del risultato totale di ugual numero di sorteggi, quando si guardi soltanto se il numero estratto supera 45. Il « valore medio » del risultato medio è uguale in entrambi i casi al « valore medio » del risultato di ciascuna osservazione (rispettivamente 45,5 e 0,5).

4. — Date tutte le serie ugualmente possibili di risultati di più osservazioni, diremo che due osservazioni sono reciprocamente *indipendenti*, se a ciascun determinato risultato dell'una troviamo associati i medesimi risultati ugualmente possibili dell'altra; che sono reciprocamente *dipendenti*, se ai vari risultati dell'una si associano diverse pluralità di risultati ugualmente possibili dell'altra.

Sono reciprocamente indipendenti il primo estratto della Ruota di Napoli e il primo estratto della Ruota di Bari: qualunque sia il primo estratto in una delle due Ruote, il primo estratto nell'altra può essere uno qualsiasi dei novanta numeri dell'urna. Sono, invece, reciprocamente dipendenti il terzo estratto ed il primo estratto di Napoli: se il terzo estratto è 27, il primo può essere 1, 2, 3, ..., 26, 28, ..., 90, ma non 27; se il primo estratto è 12, il terzo estratto può essere uno qualsiasi degli 89 numeri che restano nell'urna tolto il 12.

Nel seguito della presente esposizione ci limiteremo a considerare l'ipotesi di osservazioni reciprocamente indipendenti, che è la più semplice.

5. — In generale i singoli risultati ugualmente possibili differiscono dal « valore medio » del risultato di ciascuna osservazione. Come misura sintetica della loro variazione si può assumere lo scostamento medio quadratico; poichè questo è la media quadratica degli errori che s'incontrano sostituendo i singoli risultati possibili al « valore medio » del risultato di una data osservazione, lo diremo per brevità *errore medio di una singola osservazione*. (Parrebbe più ovvio servirsi dello scostamento medio assoluto; ma con ciò si complicherebbe molto il trattamento analitico. Accanto a ragioni teoriche, questa ragione pratica induce a preferire lo scostamento medio quadratico). Quando i risultati ugualmente possibili possono assumere soltanto i valori 0 od 1, quando cioè il « valore medio » è una probabilità p, il loro scostamento medio quadratico rispetto a p, cioè l'errore medio di una singola osservazione, risulta uguale, com'è facile dimostrare, a $\sqrt{p(1-p)}$.

Esempio: la media quadratica degli scostamenti dei novanta numeri del lotto dal « valore medio » 45,5 è 25,98, errore medio di una singola osservazione. Quando si bada soltanto se il numero sorteggiato supera 45, dei novanta possibili risultati 45 si discostano di 0,5 in più, e 45 di 0,5 in meno, dalla probabilità 0,5; l'errore medio di una singola osservazione è dunque 0,5.

In modo analogo a quello dianzi tenuto, partendo dagli scostamenti dei vari risultati totali [medi] ugualmente possibili, dal corrispondente « valore medio » si definisce l'errore medio quadratico del risultato totale [medio] di n osservazioni.

Per una serie di n osservazioni, diremo per brevità *errore medio di un'osservazione* la media quadratica degli errori medi delle singole osservazioni, e *errore medio di n osservazioni* l'errore medio quadratico del risultato medio di n osservazioni. L'errore medio quadratico del risultato totale di n osservazioni è uguale al prodotto, per n, dell'errore medio quadratico del risultato medio, come deriva immediatamente dalle definizioni di risultato totale e di risultato medio.

Quando le osservazioni sono reciprocamente indipendenti, l'errore medio di n osservazioni risulta uguale al quoziente per $\sqrt{n}$ dell'errore medio di un'osservazione, e l'errore medio del risultato totale di n osservazioni risulta uguale al prodotto per $\sqrt{n}$ dell'errore medio di un'osservazione. Si ritrovano quelle relazioni che abbiamo determinato empiricamente tra l'ampiezza media delle variazioni non significative e il numero dei casi osservati.

Così, nel giuoco del lotto, l'errore medio del primo estratto è, per ciascuna Ruota, 25,98; l'errore medio della media dei primi estratti nelle otto Ruote è dunque $9{,}19 = 25{,}98 : \sqrt{8}$, e l'errore medio della media dei primi estratti in 100 estrazioni è 2,598.

Quando i risultati ugualmente possibili in ciascuna osservazione possono assumere soltanto i valori 0 od 1, quando cioè il « valore medio » di ciascuna osservazione è una probabilità p, uguale per tutte le osservazioni, l'errore medio di n osservazioni risulta uguale a $\sqrt{p(1-p)} : \sqrt{n}$. L'errore medio si può desumere dunque immediatamente dalla probabilità e dal numero delle osservazioni. Per esempio, essendo 0,5 la probabilità che il primo estratto superi 45, l'errore medio di una osservazione si trova uguale a $0{,}5 = \sqrt{0{,}5 \times 0{,}5}$, e quello di 100 osservazioni a $0{,}5 : \sqrt{100} = 0{,}05$.

Nelle stesse ipotesi, l'errore medio del risultato totale risulta uguale a $\sqrt{np(1-p)}$.

6. — Quando siano ignoti i risultati ugualmente possibili nelle singole osservazioni e quindi siano ignoti i « valori medi », eseguendo n osservazioni intorno ad un fenomeno che in ciascuna osservazione presenti una pluralità di casi ugualmente possibili, verremo a conoscere uno dei tanti possibili risultati medi di n osservazioni. Assumendolo come espressione approssimata del « valore medio » del risultato medio di n osservazioni, commetteremo un errore, che *in media* va diminuendo col crescere del numero n. Si può dimostrare che col crescere di n cresce indefinitamente la frazione dei risultati medi di n osservazioni che non differiscono in valore assoluto di più di h (numero positivo, scelto ad arbitrio) dal « valore medio ». Per esempio, la media dei primi estratti di n estrazioni è compresa fra 45 e 46 (cioè non differisce di più di $h = \pm 0{,}5$ dal « valore medio » 45,5) in una frazione del numero dei casi ugualmente possibili non inferiore a 0,73 per $n = 10.000$, non inferiore a 0,9973 per $n = 1.000.000$, non inferiore a 0,999973 per $n = 100.000.000$. Ritroviamo così, per via deduttiva, quella *legge dei grandi numeri* cui eravamo pervenuti per via induttiva: col crescere del numero dei casi osservati i risultati medi tendono ad uguagliarsi tra loro e ad uguagliare il « valore medio ».

Sempre nelle ipotesi di costanza del « valore medio » nelle diverse osservazioni e di indipendenza tra le diverse osservazioni, si dimostra che la distribuzione degli scostamenti dei risultati medi (o dei risultati totali) ugualmente possibili di n osservazioni dal loro « valore medio » si può rappresentare in via approssimativa con la curva degli errori. Si ritrova così deduttivamente un'altra caratteristica delle variazioni non significative che avevamo accertata induttivamente.

7. — Nello schema del fenomeno collettivamente tipico costruito sulla base dell'ipotesi di una pluralità di risultati ugualmente possibili in ciascuna osservazione, si possono introdurre varianti col supporre che il « valore medio » differisca da osservazione ad osservazione, oppure che le diverse osservazioni siano reciprocamente dipendenti. A tali varianti corrispondono variazioni nella misura degli errori medi e nella distribuzione per grandezza degli scostamenti: così che si può determinare *a priori* in qual modo ed in qual misura date forme di variazione del « valore medio » o di reciproca dipendenza fra le osservazioni tendano ad influire sulle variazioni dei risultati dell'osservazione.

La comparazione dei risultati di osservazioni di fenomeni reali con quelli dedotti dalla rappresentazione schematica giova ad accertare se i primi si comportino conformemente alle ipotesi d'indipendenza fra le varie osservazioni e di costanza del « valore medio » nelle diverse osservazioni, oppure invece conformemente a date ipotesi di dipendenza tra le osservazioni e di variazione del « valore medio ». Assunta come *normale* l'ampiezza media delle variazioni che si ha nelle prime ipotesi, dicesi *subnormale* un'ampiezza inferiore, *supernormale* un'ampiezza maggiore. La variazione (o *dispersione*) normale fa supporre assenza di variazioni significative; la variazione subnormale o supernormale denota la presenza di esse, e mediante il confronto con quella normale ne consente la misurazione.

8. — Concludendo: col percorrere la via in senso inverso a quello seguito nel capitolo XXI, col muovere cioè dalla concezione di una pluralità di risultati ugualmente possibili, si giunge alla previsione di forme di regolarità simili a quelle che ci hanno condotto alla concezione dell'uguale possibilità di più risultati. Ci si ritrova dunque, apparentemente, al punto di partenza. Ma il procedimento, appena accennato nei precedenti paragrafi, che il lettore potrà trovare più ampiamente sviluppato nelle nostre *Lezioni di statistica metodologica*, non è vano, come non è vana l'opera del geometra, che dall'accertamento delle proprietà di certi oggetti concreti assurge alla definizione generale ed astratta del triangolo, e dalla definizione deduce poi, mediante il ragionamento, proprietà della figura, già implicite, almeno in un certo senso, nella definizione.

Le aree, dalla cui visione è suggerita la concezione astratta del triangolo, soltanto in via approssimativa sono equivalenti fra loro riguardo a certe proprietà; del pari i risultati, le cui frequenze relative tendono ad uguagliarsi in serie numerose di osservazioni, presentano in realtà un'equivalenza solo approssimativa riguardo a certe condizioni. Nella rappresentazione schematica, invece, l'equivalenza si suppone completa e rigorosa; e l'ipotesi permette di dedurre in forma generale nell'un caso le proprietà del triangolo, nell'altro le regolarità del fenomeno collettivamente tipico. Inoltre il passaggio dal ragionamento induttivo al deduttivo rende più facile semplificare, distinguere e ridurre in classi la varietà delle condizioni che possono presentarsi; consente quindi di spingere l'analisi dove sarebbe stato arduo, od impossibile, condurla mediante il solo esame di risultati d'osservazione.

9. — Ricordando che la regolarità del fenomeno individualmente tipico consiste nella costanza del suo presentarsi, o nella costanza della modalità o misura in cui esso si presenta, in determinate condizioni, e nella conseguente costanza di variazione del fenomeno nel passaggio da date condizioni a date altre, avremmo potuto immaginare, per analogia, che la regolarità del fenomeno collettivamente tipico dovesse consistere nella costante frequenza del suo presentarsi, o nella costante frequenza relativa del suo presentarsi in data modalità o misura, in determinate condizioni. Ora siamo in grado di rettificare quest'ipotesi.

Infatti abbiamo visto che in più serie di osservazioni, aventi comuni certe condizioni, la manifestazione del fenomeno collettivamente tipico non resta costante, anzi varia da serie a serie. Non si può dire in modo generale che varii poco nè che varii molto; secondo la natura dei fatti osservati, secondo il numero delle osservazioni, la variazione media può esser piccola o grande, in relazione alla misura del fenomeno.

Ma in certi casi, con l'estendersi del campo di osservazione la variazione media dei risultati medi tende a ridursi, e con l'indefinito aumentare del numero delle osservazioni tende a ridursi indefinitamente; così che serie numerosissime di osservazioni forniscono risultati approssimativamente costanti. Non un'assoluta costanza dunque, ma una *tendenza* alla costanza, costituisce la forma di regolarità caratteristica dei fenomeni collettivamente tipici: quella che abbiamo chiamato *stabilità statistica*.

In questa forma di regolarità, le condizioni per cui differiscono tra loro le singole osservazioni e le singole serie di osservazioni influiscono sulle manifestazioni individuali, ma tendono ad influire sempre meno sulla manifestazione collettiva del fenomeno col crescere del numero delle osservazioni. Le successive estrazioni del lotto di una medesima Ruota, le estrazioni di Ruote diverse, differiscono l'una dall'altra per innumerevoli circostanze, ma ne hanno comuni alcune. Quella differenza ci vieta la previsione del risultato d'ogni singolo sorteggio, questa costanza ci consente di prevedere, con approssimazione relativa crescente al crescere del numero delle osservazioni, il risultato medio o totale di più sorteggi. In un certo senso, pertanto, la manifestazione di un fenomeno collettivamente tipico di questo genere si può dire indipendente dalle condizioni per cui le varie serie di osservazioni differiscono tra loro; e reciprocamentequeste condizioni possono dirsi trascurabili riguardo alla manifestazione del fenomeno.

10. — La forma di regolarità dianzi accennata non si presenta però tutte le volte in cui più serie di osservazioni hanno comuni date circostanze; in realtà, appare soltanto in qualche fenomeno fisico e non si presenta quasi mai nei fenomeni biologici e mai in quelli sociali, ove, nonostante la costanza di talune condizioni di osservazione, non si vedono i risultati tendere alla costanza, nè quindi si vede l'errore relativo nella previsione del risultato medio tendere ad annullarsi al crescere del numero delle osservazioni. In altri termini: le circostanze per cui differiscono le varie osservazioni e serie d'osservazioni non appaiono più trascurabili, riguardo alla manifestazione del fenomeno. Se i numeri dell'urna variano da Ruota a Ruota, da estrazione ad estrazione, non solo resta vietata la previsione del risultato del singolo sorteggio, ma vien meno anche la possibilità di previsione del risultato medio o totale di più sorteggi.

Nel caso di costanza del « valore medio » e di reciproca indipendenza delle osservazioni, si può passare dai risultati empirici ad una rappresentazione schematica generale, valida anche per casi non osservati. La descrizione si trasforma senz'altro in induzione: approssimativa sì, ma induzione.

Invece, nel caso di variazione del « valore medio » e di reciproca dipendenza delle osservazioni, i risultati empirici possono fornire bensì indizio della presenza di codeste circostanze perturbatrici, ed anche consentire una valutazione approssimativa del loro effetto cumulativo; ma non porgono un'esatta misura dell'azione di ciascuna di esse. Finchè la variazione si presenta con lieve intensità, possono ancora adoperarsi, in via di larga approssimazione, le previsioni dedotte dall'ipotesi di costanza del « valore medio » ma di fronte a variazioni più intense vien meno ogni attendibilità di siffatte previsioni.

Si presenta allora un nuovo problema: risalire dalle variazioni nella manifestazione del fenomeno collettivamente tipico alle variazioni nel « valore medio ». A tale intento, occorre: anzitutto discernere, tra le variazioni della misura del fenomeno, quelle significative da quelle non significative; poi accertare se sussista qualche relazione fra le variazioni significative ed una o più delle condizioni per cui differiscono le varie serie di osservazioni. Come si operi al primo di tali fini, abbiamo già accennato; come si operi al secondo di essi, vedremo nel capitolo XXII.

Indicazioni bibliografiche. — Intorno alla stabilità statistica vedansi specialmente i manuali di WESTERGAARD, YULE, MORTARA, citati a pag. 6; quello di MARCH, citato a pag. 275; la *Matematica attuariale* di BROGGI; e F. VINCI, *Statistica metodologica*, Padova, La Litotipo, 1924. Rimangono fondamentali gli studi di W. LEXIS, riassunti in forma elementare nelle *Abhandlungen* citate a pag. 227 (vedansi specialmente i capitoli VII e VIII).

Quesiti ed esercizi: 1. — Si può rispondere in modo preciso al seguente quesito: Quanti casi si devono osservare per poter riscontrare uniformità in un fenomeno collettivamente tipico? E all'altro quesito: V'è qualche relazione tra il numero dei casi osservati e il presentarsi delle uniformità dei fenomeni collettivamente tipici?

2. — Ricavando i dati dal *Movimento della popolazione secondo gli atti dello stato civile*, pubblicazione annuale dell'ISTITUTO CENTRALE DI STATISTICA, si studi come varia, col variare del numero dei casi osservati, l'ampiezza delle oscillazioni dei seguenti fenomeni:

a) Mascolinità delle nascite. Si possono studiare le variazioni attraverso il tempo, come è stato fatto nel paragrafo I del testo, scegliendo periodi diversi da quello ivi considerato, o popolazioni diverse. Si possono, invece, studiare le variazioni nello spazio, esaminando le oscillazioni della mascolinità nelle dieci regioni più popolose e in dieci provincie con 300-500 mila abitanti ciascuna (NB. Non v'è alcuna ragione sostanziale per prendere proprio il numero di dieci anni o di dieci circoscrizioni; v'è soltanto la ragione formale che i calcoli riescono semplificati).

b) Mascolinità delle morti. Si proceda con analoghi criteri. Sarà interessante contrapporre i risultati di questa indagine *b* a quelli dell'indagine *a*, per vedere il differente comportamento dei due fenomeni col variare del numero delle osservazioni.

c) Natalità (nati per 1000 abitanti; per le regioni — compartimenti — e per le provincie il calcolo è già eseguito nel *Mov. d. pop.*; per i principali comuni lo si può eseguire riferendo i dati sulle nascite a quelli sulla popolazione, riportati nell'introduzione ai singoli volumi della citata pubblicazione). In questo esempio si vedranno in generale prevalere fortemente le variazioni significative su quelle non significative.

d) Percentuale dei vedovi fra gli sposi di 35 a 39 anni. Nel *Mov. d. pop.* sono forniti dati per l'Italia in complesso, per regioni e per provincie, sui quali si possono calcolare le percentuali. Analoga indagine si può eseguire per le spose. Così per gli sposi come per le spose si può prendere un gruppo d'età diverso da quello sopra indicato.

e) Percentuale dei parti multipli sul totale dei parti. Si hanno dati per l'Italia, per regioni e per provincie. In questo esempio pare lecito ritenere che le variazioni riscontrate siano, almeno in buona parte, non significative?

3. — Ricavando i dati dalla *Statistica delle cause di morte*, pubblicazione annuale dell'ISTITUTO CENTRALE DI STATISTICA, si studi come varia, col variare del numero dei casi osservati, l'ampiezza delle oscillazioni dei seguenti fenomeni:

a) Percentuale dei suicidi consumati con un determinato mezzo sul totale dei suicidi (p. es. suicidi con arma da fuoco; si hanno dati per l'Italia, per regioni, per provincie);

b) Mascolinità dei suicidi, secondo il mezzo impiegato per il suicidio (si hanno dati per l'Italia e per regioni; converrà costituire serie di proporzioni accertate nelle singole regioni in più anni successivi).

4. — Ricavando i dati dalle statistiche della leva militare contenute nella pubblicazione annuale del MINISTERO DELLA GUERRA, *Della leva di terra*, si studi come varia, col variare del numero dei casi osservati, l'ampiezza delle oscillazioni, nel tempo, dei seguenti fenomeni:

a) Percentuale dei riformati per singole cause sul totale degli iscritti di leva visitati (come nel paragrafo 3 del testo; si hanno dati per l'Italia, per regioni e per provincie).

b) Percentuale degli individui aventi una data statura (p. es. m. 1,67) sul totale degli iscritti di leva misurati (si hanno dati per l'Italia, per regioni, per provincie).

5. — Ricavando i dati dalle statistiche agrarie, riassunte nell'ASI, si studi come varia l'ampiezza delle oscillazioni, attraverso il tempo, del rendimento medio per ettaro nella coltivazione del frumento, passando da singole provincie a singole regioni e poi all'Italia. In questo esempio prevalgono le variazioni non significative o quelle significative?

6. — Quali delle variazioni di fenomeni collettivamente tipici si dicono « non significative »? In quale relazione sta l'ampiezza media di queste variazioni col numero dei casi osservati? Come tendono a distribuirsi secondo il segno le variazioni non significative? come secondo la grandezza? come secondo la grandezza ed il segno?

7. — I seguenti dati indicano la mascolinità delle nascite in Francia in ciascuno dei cent'anni dal 1806 al 1905 (in ciascuna colonna i dati sono in ordine cronologico). Il numero medio annuo delle nascite in questi cent'anni è stato di 936.616, con variazioni non grandi da un anno all'altro (minimo 807.291, massimo 1.017.896).

Nati maschi per 10.000 nati in Francia.

1806-1825	1826-1845	1846-1865	1866-1885	1886-1905
5.149	5.154	5.135	5.129	5.099
5.142	5.156	5.112	5.104	5.106
5.145	5.136	5.132	5.116	5.114
5.163	5.143	5.133	5.123	5.114
5.161	5.133	5.133	5.117	5.109
5.176	5.160	5.113	5.119	5.116
5.170	5.155	5.128	5.119	5.115
5.172	5.165	5.148	5.122	5.110
5.163	5.157	5.131	5.129	5.105
5.146	5.155	5.136	5.124	5.081
5.164	5.148	5.129	5.114	5.102
5.168	5.147	5.127	5.110	5.102
5.157	5.148	5.122	5.116	5.111
5.153	5.147	5.129	5.110	5.101
5.154	5.139	5.117	5.099	5.102
5.164	5.147	5.121	5.128	5.095
5.150	5.152	5.128	5.109	5.101
5.149	5.140	5.120	5.121	5.095
5.158	5.135	5.131	5.112	5.094
5.169	5.139	5.124	5.131	5.102

La media aritmetica dei cento dati è 5.131. Si calcolino gli scostamenti dei singoli dati dalla media; si esamini la loro distribuzione secondo il segno e secondo la grandezza; si calcoli la loro media aritmetica e la loro media quadratica. Si provi a calcolare medie decennali; poi si calcolino gli scostamenti di esse dalla media generale e si verifichi se la media di questi scostamenti stia a quella desunta dai dati annuali press'a poco nel rapporto di 1 a $\sqrt{10}$.

Si rappresenti in diagramma la percentuale degli scostamenti che non superano, in grandezza, una frazione progressivamente crescente dello scostamento medio assoluto o dello scostamento medio quadratico (per esempio, un quinto, due quinti, tre quinti, ecc., di esso). Si confronti il diagramma con quello che può ricavarsi dalla seguente tabellina, che indica come si distribuiscano, secondo la grandezza, gli scostamenti, nell'ipotesi che si adatti perfettamente a rappresentare la loro distribuzione la curva degli errori (la quale è contrassegnata dalla costanza — comunque variino i suoi parametri — delle percentuali degli scostamenti non superiori ad una data frazione dello scostamento medio).

Percentuale teorica degli scostamenti non superiori a *m* volte lo scostamento medio.

m	Di 100 scostamenti, non superano *m* volte lo scostamento medio	
	assoluto	quadratico
0,1	6,36	7,97
0,2	12,68	15,85
0,3	18,92	23,58
0,4	25,04	31,08
0,5	31,00	38,29
0,6	36,78	45,14
0,7	42,35	51,61
0,8	47,67	57,63
0,9	52,73	63,19
1,0	57,51	68,27
1,2	66,17	76,99
1,4	73,60	83,85
1,6	79,83	89,04
1,8	84,90	92,81
2,0	88,95	95,45
2,5	95,39	98,76
3,0	98,33	99,73
3,5	99,48	99,95
4,0	99,86	99,99

L'esame degli scostamenti dei dati annuali dalla media secolare, compiuto nei modi sopra indicati, induce a ritenere che tali scostamenti corrispondano per intero a variazioni non significative?

Si esamini l'andamento delle medie decennali della mascolinità. Basta questo esame a far escludere che i suddetti scostamenti corrispondano per la maggior parte a variazioni non significative? Consente esso invece di affermare il contrario?

Eseguendo sulle medie decennali un'interpolazione lineare col metodo delle somme, si può surrogare alla serie dei dati osservati, riferita all'inizio di questo esercizio, una progressione aritmetica decrescente, il cui primo dato è 5.165,8088 e l'ultimo 5.096,2712. I termini intermedi si possono facilmente computare conoscendo la ragione della progressione, che è — 0,7024. Eseguito il computo, si calcolino le differenze fra dati osservati e dati calcolati e si verifichi poi, coi soliti procedimenti, se esse presentino il carattere di variazioni non significative. In questa ipotesi le deviazioni dei dati interpolati dalla media generale misurerebbero le variazioni significative, le deviazioni dei dati osservati dai dati interpolati misurerebbero le variazioni non significative della mascolinità delle nascite in Francia.

8. — Con procedimenti simili a quelli adottati nella prima parte del precedente esercizio, si verifichi se le deviazioni dei cento risultati di estrazioni del lotto, riferiti nel paragrafo 7 del testo, dalla loro media generale 72 possano riguardarsi come variazioni non significative.

9. — Come tende a variare, col numero dei casi osservati, la grandezza media delle variazioni non significative nel risultato totale dell'osservazione (numero dei casi contrassegnati da un dato carattere) e nel risultato medio dell'osservazione (rapporto tra il numero dei casi contrassegnati da un certo carattere e quello dei casi osservati)?

10. — Come si può esprimere la « legge dei grandi numeri »? Quando un fenomeno si dice « statisticamente stabile »? Quale differenza intercede fra « stabilità statistica » e « stabilità » nella consueta accezione di questa parola?

11. — Con quali criteri si può cercar di discernere se date variazioni di fenomeni reali siano significative oppure non significative? Quali cautele bisogna usare nell'interpretazione dei dati statistici, ad evitare di prendere per significative variazioni che non siano tali?

12. — Com'è detto nel paragrafo 5 della nota al capitolo XXI, quando il « valore medio » di ciascuna osservazione è una probabilità costante p, l'errore medio del risultato medio di n osservazioni è uguale a $\sqrt{p(1-p)} : \sqrt{n}$. Se la mascolinità delle nascite, calcolata in ciascun anno per l'Italia, potesse riguardarsi come l'espressione empirica di una probabilità costante, lo scostamento medio quadratico delle percentuali annuali, riportate nel paragrafo 1 del testo, dalla loro media 51,34, dovrebbe essere press'a poco uguale a $\sqrt{0{,}5134 \cdot 0{,}4866} : \sqrt{1.097.193} = 0{,}4998 : 1.047 = 0{,}000477$. Esso risulta in realtà uguale a 0,000535, cioè poco differente dal suo « valore medio »: così che si ha ragione di ritenere verosimile che le oscillazioni della mascolinità siano totalmente o per la massima parte non significative.

Si esegua analoga verifica per gli altri esempi dei paragrafi 1, 2 e 3 del testo e per l'esempio dell'esercizio 7.

Tra la media quadratica s_2 e la media aritmetica dei valori assoluti s_1 di scostamenti che si distribuiscano conformemente alla curva degli errori sussiste la relazione $s_2 = 1{,}25\ s_1$, ovvero: $s_1 = 0{,}80\ s_2$. È quindi facile calcolare il valore teorico dello scostamento medio assoluto da quello dello scostamento medio quadratico determinato nel modo sopra ricordato, per poi paragonare il valore osservato con questo valore teorico. Nell'esempio di dianzi, poichè il valore teorico dello scostamento medio quadratico è 0,000477, quello dello scostamento medio assoluto è 0,000382; il valore realmente accertato è 0,000440.

13. — Indagini rappresentative. — Siano u individui o casi individuali contraddistinti da un certo carattere quantitativo e sia M la media aritmetica delle misure in cui codesto carattere si presenta nei singoli individui, s_2 lo scostamento medio quadratico delle misure individuali dalla media.

Scelta a caso (ossia in modo che le u misure siano risultati ugualmente possibili della scelta) una fra le u misure; sceltane una seconda fra le $u - 1$ rimanenti, una terza fra le $u - 2$ che ancora restano, e così via fino ad averne scelte n, la media aritmetica m delle n misure scelte rappresenterà un valore approssimato di M; e la media quadratica degli scostamenti delle misure stesse da m fornirà un valore approssimato di s_2.

I valori ugualmente possibili del risultato medio m, i quali sono in numero di $u\,(u-1)\,(u-2)\ldots(u-n+1)$, si discostano da M, in media quadratica, di:

$$\frac{s_2}{\sqrt{n}}\sqrt{1-\frac{n-1}{u-1}}.$$

Mercè la precedente formola si può determinare il numero n di osservazioni sufficiente a conferire un certo grado di approssimazione al risultato medio, quale valore approssimato di M.

Per esempio, dati 10.000 ($=u$) individui, la cui statura media aritmetica è di 163 cm. ($=M$) e lo scostamento medio quadratico 7 cm. ($=s_2$), prendendo a caso 100 ($=n$) di essi, la loro statura media differirà, in media quadratica, di cm. 0,7 dalla media generale; prendendone 2.500 lo scostamento medio quadratico si ridurrà a cm. 0,12.

Dati ancora 10.000 individui, se tra essi 1.000 sono celibi e 9.000 coniugati o vedovi, il carattere «celibe» si presenta nella misura 1 nei primi 1.000, nella misura 0 negli altri 9.000; in media, nella misura 0,1. Lo scostamento medio quadratico è 0,3. Scegliendo a caso 100 individui, la frequenza relativa dei celibi tra essi differirà, in media quadratica, di 0,03 da 0,1; scegliendone 2.500, ne differirà in media di 0,0052.

Appare dai precedenti esempi come l'applicazione di criteri assai semplici permetta di ottenere da osservazioni parziali risultati sufficientemente prossimi a quelli che darebbe l'osservazione completa di un complesso di elementi.

Allo stesso intento si può anche, talora, procedere in modo un po' diverso da quello esemplificato.

Suppongansi 90.000 individui divisi in 900 centurie: la statura media e la variabilità delle stature variino da una centuria all'altra; sia $s_2 = 7$ cm. la media quadratica degli scostamenti medi quadratici accertati nelle diverse centurie. Se prendiamo a caso un individuo da ciascuna centuria, la statura media dei 900 individui scelti differirà, in media quadratica di $s_2 : \sqrt{900} =$ cm. 0,23, dalla media generale. Se gli individui fossero divisi in decurie e $s_2 = 10$ cm. fosse la media quadratica degli scostamenti medi quadratici accertati nelle diverse decurie, preso a caso un individuo per decuria, la statura media dei 9.000 scelti differirebbe, in media quadratica, di cm. 0,105 dalla media generale.

Avvertasi che la scelta si suppone sempre fatta a caso, ossia in modo che per ciascun elemento, compreso nel complesso alla cui osservazione completa si rinunzia, sia ugualmente possibile venir scelto. E nell'ottener ciò sta, molte volte, la maggiore difficoltà pratica.

Talvolta nelle indagini rappresentative, studiando fenomeni o caratteri contraddistinti dal presentarsi di una misura più frequente, o *normale,* nella loro manifestazione, si limita l'indagine ad alcuni elementi scelti nell'insieme, non a caso, ma fra quelli che presentano il fenomeno o carattere studiato nella misura normale. È evidente la differenza tra le due forme d'indagine rappresentativa: la prima tende a dare un'immagine ridotta dell'intero complesso studiato, la seconda d'una parte sola di esso, che si reputa la più importante ai fini dell'indagine.

CAPITOLO XXII.

La ricerca delle relazioni tra le variazioni significative e le circostanze d'osservazione.

Posizione del problema — Come il problema venga risoluto per i fenomeni individualmente tipici e quale sia il significato delle conclusioni tratte dall'applicazione dei metodi comparativi: esposizione schematica di tali metodi — L'esperimento come strumento logico; in qual modo lo si supplisca quando esso non riesce possibile — Induzione; leggi empiriche — Limiti dell'applicabilità dei metodi comparativi ai fenomeni collettivamente tipici — Esempi di applicazione.

1. — Abbiamo mostrato nel capitolo precedente come i fenomeni collettivamente tipici presentino variazioni non inerenti alla natura del singolo fenomeno ma esclusivamente dipendenti dall'estensione delle osservazioni e tendenti a scomparire, nei risultati medi, coll'indefinito aumentare di tale estensione. Discernere ed eliminare tali variazioni « non significative » è compito che precede ogni ricerca di relazioni tra le variazioni inerenti alla natura del fenomeno (« variazioni significative ») e le circostanze d'osservazione.

Spesso, però, nelle indagini su fenomeni sociali si dispone di numeri d'osservazioni molto grandi e si ha perciò ragione di ritenere che le variazioni non significative siano relativamente piccole; d'altra parte le variazioni palesemente significative sono tanto grandi da far apparire sicuro che quelle non significative siano, al confronto, trascurabili. In simili casi si sogliono riguardare come significative tutte le variazioni del fenomeno; e normalmente si commette con ciò un errore assai lieve. Per esempio, le variazioni della natalità in Italia da anno ad anno negli ultimi cinquant'anni, o le variazioni del consumo medio individuale del vino da provincia a provincia, o le variazioni del salario medio da industria a industria, si assumono per intero come significative, e si sbaglia di poco.

Quando il numero delle osservazioni è molto grande, e quindi sono da presumere piccole le variazioni non significative, ma appaiono assai piccole anche le variazioni significative, è invece indispensabile cercare anzitutto di distinguere queste da quelle. Ci limitiamo a richiamare, a questo proposito, l'esempio della mascolinità delle nascite in Italia esposto nel capitolo precedente.

Tale distinzione preliminare è, poi, necessaria ogni volta che l'estensione delle osservazioni sia molto ristretta.

Se, eliminate le variazioni non significative, il fenomeno appare « statisticamente stabile » (nel senso chiarito nel capitolo precedente) comunque variino le circostanze d'osservazione, il compito dell'indagatore è finito: egli ha accertato, infatti, che il modificarsi delle circostanze d'osservazione non influisce sulla misura collettiva della manifestazione del fenomeno. Ma, come abbiamo già avvertito, questo caso non si presenta nei fenomeni sociali: in generale la misura di questi tende anzi a modificarsi col variare delle circostanze d'osservazione, come d'altronde avviene anche per molti fenomeni collettivamente tipici che rientrano nel campo della fisica o della biologia. Così che, di regola, nelle indagini su fenomeni collettivamente tipici si presenta il problema della determinazione delle relazioni esistenti tra la manifestazione dei fenomeni stessi e le circostanze nelle quali essi si presentano.

In pratica il problema si pone in due forme, apparentemente diverse: o si vuole stabilire *se e come una determinata circostanza influisca* sulla manifestazione di un fenomeno, oppure si vuole stabilire *quali circostanze influiscano* sulla manifestazione di un determinato fenomeno, *e come influiscano*. È soltanto apparente la distinzione tra le due forme del problema, perchè quand'esso venga proposto nella seconda non si può far altro che ridurlo alla prima, cioè provare a volta a volta se e come singole date circostanze influiscano sulla manifestazione del fenomeno. Ed è perciò necessario che l'indagatore sia non soltanto perito nei metodi statistici ma anche profondo conoscitore dei fenomeni ai quali i metodi stessi sono stati applicati: così egli eviterà lunghe ricerche sulla influenza di circostanze che effettivamente non agiscono sul fenomeno e metterà alla prova soltanto le ipotesi più plausibili.

2. — Per semplificare l'esposizione dei metodi diretti a stabilire se e come una determinata circostanza influisca sulla manifestazione di un fenomeno, richiameremo anzitutto i procedimenti che si applicano nello studio dei fenomeni individualmente tipici. Vedremo poi se e con quali limitazioni essi siano applicabili ai fenomeni collettivamente tipici. Questi procedimenti si riducono tutti alla comparazione fra i risultati di più osservazioni, ed hanno per riscontro altrettanti modi di comparazione, mediante i quali si ricerca se al manifestarsi ed al variare del fenomeno studiato corrisponda costantemente la presenza o la variazione d'una determinata circostanza (o combinazione di circostanze: si sottintenda questa alternativa in tutto il seguito del presente paragrafo), oppure se alla presenza od

alla variazione di una data circostanza (o combinazione di circostanze) corrisponda costantemente il manifestarsi od il variare del fenomeno.

La comparazione fra i risultati di diverse osservazioni non conduce, in generale, a conclusioni sicure e definitive. Esaminando schematicamente le varie forme di conclusioni cui si può giungere direttamente mercè la comparazione, vedremo come la maggior parte di esse abbia il carattere di semplici ipotesi.

I. Se si trova che, tutte le volte in cui è presente [si modifica] una data circostanza, si manifesta [si modifica] il fenomeno, si può supporre che la circostanza sia sufficiente per determinare la manifestazione del fenomeno [che la modificazione della circostanza sia sufficiente per determinare la modificazione del fenomeno]; non si può escludere che sia sufficiente nè che sia necessaria.

II. Se si trova che, tutte le volte in cui si manifesta [si modifica] il fenomeno, è presente [si modifica] una determinata circostanza, si può supporre che la circostanza sia necessaria per determinare la manifestazione del fenomeno [che la modificazione della circostanza sia necessaria per determinare la modificazione del fenomeno]; non si può escludere che sia necessaria nè che sia sufficiente.

III. Se si trova che, alcune volte in cui è presente [si modifica] una data circostanza, si manifesta [si modifica] il fenomeno, altre volte no, si può escludere che la circostanza sia sufficiente per determinare la manifestazione del fenomeno [che la modificazione della circostanza sia sufficiente per determinare la modificazione del fenomeno]; non si può escludere che sia necessaria.

IV. Se si trova che, alcune volte in cui si manifesta [si modifica] il fenomeno, è presente [si modifica] una data circostanza, altre volte no, si può escludere che la circostanza sia necessaria per determinare la manifestazione del fenomeno [che la modificazione della circostanza sia necessaria per determinare la modificazione del fenomeno]; non si può escludere che sia sufficiente.

Il carattere ipotetico delle conclusioni che si traggono mediante l'applicazione di un singolo modo di comparazione risulta chiaro al considerare — per esempio — la variazione di un fenomeno (z), suscettibile di espressione quantitativa, in relazione alle variazioni di due circostanze (x, y), anch'esse esprimibili quantitativamente, dalle quali il fenomeno dipende. Posto, p. es.:

$$z = a\,x + by, \qquad (a,\ b \text{ costanti})$$

è facile vedere che z varia se varia la sola x o la sola y; può variare se variano entrambe; resta costante se entrambe restano tali; può restare costante anche se entrambe variano.

Confrontando, in uno dei modi indicati nel testo, la variazione di z con quella di x soltanto, o di y soltanto, non si può, in via generale, desumere con sicurezza la relazione fra z e la circostanza variabile scelta. La variazione di una determinata delle due circostanze appare condizione sufficiente ma non necessaria per la variazione del fenomeno; la variazione di almeno una di esse è condizione necessaria; la variazione di entrambe insieme non è necessaria nè sufficiente.

Combinando tra loro i risultati di varie comparazioni eseguite nei modi detti dianzi, si può proceder oltre nello studio delle relazioni tra il fenomeno e le circostanze di osservazione: il simultaneo verificarsi dei casi I e II suggerisce l'ipotesi che la circostanza in esame sia necessaria e sufficiente per la manifestazione del fenomeno; il simultaneo verificarsi di I e IV consente l'ipotesi che sia sufficiente, ma esclude quella che sia necessaria; il simultaneo verificarsi di II e III consente l'ipotesi che sia necessaria, ma esclude quella che sia sufficiente; il simultaneo verificarsi di III e IV esclude così l'ipotesi che sia necessaria come quella che sia sufficiente.

Coi metodi di comparazione applicati nei modi sopra esposti si accerta, dunque, semplicemente l'esistenza di alcune concomitanze, o non concomitanze, tra la manifestazione [la modificazione] d'un fenomeno e la presenza [la modificazione] d'una circostanza: tale accertamento suggerisce ipotesi intorno alla relazione tra il fenomeno e la circostanza in esame.

3. — Ma per stabilire con certezza (s'intende, con quella certezza empirica che è propria del procedimento induttivo, e non con assoluta certezza logica) quale sia la relazione esistente tra la manifestazione [la modificazione] di un fenomeno e la presenza [la modificazione] di una circostanza, è indispensabile muovere un ulteriore passo.

Nel caso I, si cercherà di verificare se, *comunque variino tutte le altre circostanze*, il fenomeno si manifesti [si modifichi] purchè sia presente [si modifichi] la circostanza sulla quale si è fermata l'attenzione. Ove ciò avvenga, applicando il metodo d'induzione detto « di concordanza » [« delle variazioni concomitanti »] si dichiarerà che la circostanza data è sufficiente a determinare la manifestazione [la modificazione] del fenomeno.

Nel caso II, si cercherà di verificare se, *restando immutate tutte le altre circostanze*, il fenomeno cessi di manifestarsi [di modificarsi] quando non sia presente [non si modifichi] la circostanza sulla quale si è fermata l'attenzione. Ove ciò avvenga, applicando il metodo d'induzione detto «di differenza», si dichiarerà che la circostanza data è necessaria a determinare la manifestazione [la modificazione] del fenomeno.

Nel caso III, essendo esclusa la sufficienza della circostanza data a determinare [a modificare] il fenomeno, si cercherà di verificare se essa sia a ciò necessaria, cioè si indagherà se il fenomeno possa presentarsi [modificarsi] anche in assenza della circostanza stessa.

Nel caso IV, essendo esclusa la necessità della circostanza data a determinare [a modificare] il fenomeno, si cercherà di verificare se essa sia a ciò sufficiente, cioè si indagherà se il fenomeno si presenti [si modifichi] ogni volta che è presente [si modifica] la circostanza stessa.

È facile scorgere che anche i procedimenti ora accennati sono comparativi; differiscono però da quelli del paragrafo precedente in quanto per applicarli l'indagatore deve poter modificare nel modo più opportuno le condizioni d'osservazione. Per conseguire tale intento egli può battere una via diretta e sicura, quella dell'*esperimento*.

Nell'esperimento l'indagatore mantiene costanti date circostanze d'osservazione e ne fa variare (cioè introduce, esclude, modifica) una o più altre, per rendersi conto dell'eventuale influenza di tale variazione sulla manifestazione del fenomeno. Ma, com'è ovvio, l'esperimento è possibile soltanto quando l'indagatore ha la facoltà di modificare a proprio talento le circostanze di osservazione: almeno quelle che presumibilmente esercitano sensibile influenza sul fenomeno (nessun indagatore ha facoltà, per esempio, di arrestare il decorso del tempo, e solo in quanto il tempo non influisca sensibilmente sulla manifestazione del fenomeno studiato sarà lecito generalizzare conclusioni tratte dall'esperimento). Ora molte volte è negata, totalmente o parzialmente, all'indagatore la facoltà di modificare le condizioni di osservazione: intere scienze, come l'astronomia, e intere categorie di scienze, come quella delle scienze sociali, si vedono normalmente preclusa la possibilità dell'esperimento.

Si può battere allora un'altra via, meno diretta, però, e meno sicura. All'esperimento, col quale si tengono ferme alcune circostanze d'osservazione e se ne modifica una particolare, si sostituisce la comparazione fra due (o più) osservazioni, aventi comuni tutte le circostanze di osservazione, eccettuata una, per la presenza, modalità o misura della quale differiscono. La differenza che si riscontra nella manifestazione del fenomeno tra due osservazioni, differenti solo per una circostanza, viene attribuita appunto a questa circostanza. Tale procedimento costituisce un perfetto surrogato dell'esperimento, purchè sia possibile: 1° comparare osservazioni che abbiano veramente comuni tutte le circostanze tranne una; 2° ottenere con sicurezza la presenza, l'assenza, o una data modificazione di quest'*una* circostanza. In pratica riesce spesso impossibile soddisfare la prima condizione, o la seconda, o entrambe; così che di regola il procedimento si deve considerare un surrogato più o meno imperfetto dell'inattuabile esperimento. È codesta una delle cause d'inferiorità delle scienze alle quali è negato l'esperimento in confronto a quelle cui è concesso.

4. — Ricapitolando: l'esame comparativo delle osservazioni eseguite sopra un fenomeno suggerisce ipotesi sopra le relazioni esistenti tra il fenomeno e le circostanze nelle quali esso si manifesta. Queste ipotesi vengono poi messe alla prova mediante nuove comparazioni, col sussidio, dov'è possibile, di nuove osservazioni eseguite in condizioni opportunamente modificate. Uno dei più efficaci mezzi di prova consiste nel ricercare se si trovino verificate, nell'osservazione della realtà, conseguenze dedotte, mediante il ragionamento, dalle ipotesi adottate.

Se le ipotesi vengono confermate in queste prove, potranno talora essere estese anche a casi non osservati, mediante la generalizzazione dei risultati delle osservazioni eseguite. Si compirà così una *induzione*.

Le relazioni tra fenomeni e circostanze d'osservazione, stabilite mediante induzione, son dette, quando abbiano un campo di validità assai ampio, *leggi empiriche*. Con questa denominazione si vuole indicare che si tratta di regolarità desunte direttamente dall'osservazione, e non indirettamente da rappresentazioni astratte e semplificate della realtà.

Quale sia il fondamento logico dell'induzione, è stato molto discusso. Generalmente si è finito col riporlo in qualche principio generale, per esempio in quello dell'uniformità della natura, che deve riguardarsi esso medesimo come una generalizzazione dell'esperienza umana. Si è giunti così, per diverse vie, a conchiudere press'a poco questo: che il fondamento dell'induzione sta in una induzione;

così che la discussione è ancora aperta e verosimilmente resterà sempre aperta.

Lo studioso dei fenomeni concreti si accontenta di sapere che una millenaria esperienza giustifica appieno l'uso di questo procedimento logico, il quale si è dimostrato fecondo di preziosi risultati nella scienza e nella pratica.

5. — Tutto quanto abbiamo detto finora si riferisce allo studio dei fenomeni individualmente tipici; chè per quelli collettivamente tipici, mancando la costanza della manifestazione individuale nella costanza di date condizioni, vien meno la possibilità di applicare i metodi comparativi ai risultati delle singole osservazioni. È possibile invece applicarli ai risultati medi di gruppi d'osservazioni?

Quando, in determinate condizioni, il fenomeno tende ad un determinato livello col crescere del numero dei casi osservati — quando cioè si presenta una stabilità statistica, i metodi comparativi possono essere applicati ai risultati medi. Se in corrispondenza alle circostanze A, B, C_1 la misura di un fenomeno collettivamente tipico tende al livello costante z_1, mentre in corrispondenza alle circostanze A, B, C_2 tende al livello costante z_2, potremo supporre che la variazione da C_1 a C_2 della circostanza C sia sufficiente a determinare la variazione del fenomeno dalla misura media z_1 alla misura media z_2. Saremo poi tratti a supporre che sia necessaria se ogni volta che si avvera la variazione del fenomeno da z_1 a z_2 la circostanza C ha variato da C_1 a C_2. Avvertasi, però, che, per quanto grande sia il numero delle osservazioni eseguite nelle une e nelle altre circostanze, non potremo mai assumere che la variazione del fenomeno da z_1 a z_2 rappresenti precisamente l'effetto della variazione di C da C_1 a C_2: una parte della differenza tra i due risultati medi è « non significativa », così che la differenza significativa determinata dalla variazione di C da C_1 a C_2 potrà essere alquanto maggiore o alquanto minore della differenza $z_1 - z_2$. Quindi la conclusione è necessariamente imprecisa.

Se nell'urna del lotto sostituiamo ai numeri da 1 a 10 i numeri da 91 a 100, il « valore medio » del risultato medio di più sorteggi passa da 45,5 a 55,5: la surrogazione eseguita determina una variazione di 10 unità nel « valore medio » del risultato medio. Ma solo nello schema teorico del giuoco troviamo questa relazione rigorosa; nella pratica, eseguendo un numero molto grande di sorteggi dall'urna contenente i numeri da 1 a 90 otterremo un risultato medio vicino a 45,5; eseguendo un numero molto grande di sorteggi dall'urna contenente i numeri da 11 a 100 otterremo un risultato medio vicino a 55,5; ma in generale la differenza tra i due risultati medi non sarà proprio uguale a 10: differirà cioè, per variazioni non significative, da quella che la surrogazione attuata tende a determinare. Col crescere del numero dei sorteggi nelle due serie, la differenza osservata tenderà ad avvicinarsi indefinitamente al suo « valore medio » 10, ma per quanto si aumenti il numero dei sorteggi non si giungerà mai ad un punto oltre il quale essa sia e si mantenga costantemente uguale a 10.

Analogamente potremo applicare i metodi comparativi a risultati medi per ricercare se la variazione di una circostanza *non* sia sufficiente o *non* sia necessaria a determinare variazione nella misura media di un fenomeno collettivamente tipico. Ma anche in questo caso la conclusione sarà imprecisa: dovremo tener conto delle variazioni non significative.

I procedimenti comparativi che permettono di stabilire relazioni tra la manifestazione del fenomeno collettivamente tipico e le circostanze in cui esso viene osservato sono dunque simili a quelli che vengono impiegati per lo studio dei fenomeni individualmente tipici; ma si applicano ai risultati medi di singoli gruppi di osservazioni invece che a singole osservazioni, e conducono a conclusioni approssimative invece che a conclusioni precise; l'approssimazione è normalmente tanto maggiore quanto maggiore è il numero delle osservazioni nei vari gruppi.

Come sappiamo, soltanto per un numero relativamente ristretto di fenomeni collettivamente tipici, specialmente nel campo dei fenomeni sociali, si riesce ad accertare una stabilità statistica, si riesce cioè ad individuare un complesso di circostanze nelle quali la misura media del fenomeno tenda alla stabilità col crescere del numero delle osservazioni. Per moltissimi fenomeni non si riesce ad individuare un tal complesso di circostanze, anzi si nota che, pur rimanendo costante tutto un complesso di circostanze, la misura media di manifestazione del fenomeno tende a variare. Se ne induce che, mentre restano costanti le circostanze anzidette, altre circostanze, influenti in modo non trascurabile sulla manifestazione del fenomeno, si modificano; e si tenterà di ricercare — coi soliti procedimenti comparativi — quali siano queste circostanze e come influiscano sul fenomeno. Ma è chiaro che, fino a quando il fenomeno non si presenta come statisticamente stabile, le conclusioni desunte mercè le applicazioni dei metodi comparativi non sono soltanto approssimative ma anche incerte.

6. — La possibilità dell'esperimento non è affatto esclusa per il fenomeno collettivamente tipico, come tale; anzi innumerevoli fenomeni collettivamente tipici, sopratutto nel campo delle scienze fisiche e di quelle biologiche, vi si prestano. Ma, per riuscire logicamente efficace, l'esperimento dev'essere esteso a collettività di casi; eseguito sopra il caso singolo non ha altro valore che quello di un singolo elemento dell'esperimento sulla collettività. D'altra parte, le conclusioni che si possono trarre dall'esperimento collettivo non sono rigorose ma approssimative, perchè il risultato è affetto da variazioni non significative, per quanto grande sia l'estensione data all'osservazione sperimentale.

Volendo dar esempi di fenomeni collettivamente tipici che consentano l'esperimento, prescinderemo da quello, troppo ovvio, dei giuochi di sorte, e ci riferiremo ad applicazioni statistiche nel campo agrario o in quello zootecnico. Seminando su due appezzamenti, uguali per dimensioni, per qualità di terreno e per ogni condizione esteriore, due varietà diverse di frumento, e coltivandole con gli identici metodi, si esegue un esperimento diretto a determinare l'influenza della varietà coltivata sul rendimento della coltura. La differenza tra i rendimenti ottenuti sui due appezzamenti è la somma algebrica d'una differenza significativa determinata dall'unica circostanza per cui differiscono le due colture, cioè dalla diversa varietà delle sementi, e d'una differenza non significativa; soltanto se quest'ultima è molto piccola in confronto alla prima si può con lieve errore assumere come significativa l'intera differenza osservata. Analogamente, scegliendo da un vasto gregge due numerosi gruppi di agnelli equivalenti per razza, per età, per robustezza, ed allevandoli in condizioni uguali, però con differente regime alimentare, si esegue un esperimento diretto a determinare l'influenza dell'alimentazione sulla produzione della lana. Il differente peso medio del vello nei due gruppi di ovini in parte deriva da differenze determinate dal diverso regime alimentare, in parte risulta da variazioni non significative; se queste sono relativamente piccole in confronto a quelle significative, si sbaglierà di poco misurando dalla differenza accertata la conseguenza del differente regime alimentare. Gli errori, tutt'altro che rari, nell'interpretazione dei risultati di simili esperimenti su fenomeni collettivamente tipici talora derivano dalla confusione di variazioni non significative con variazioni significative, tal altra da imperfetta parità delle condizioni d'esperimento diverse da quella che l'indagatore ha fatto variare: così nei nostri esempi, se i due terreni sono di qualità diverse o vengono diversamente coltivati, se gli agnelli dei due gruppi sono di razze diverse o vengono allevati in climi diversi, manca quella «parità di ogni altra condizione» che è necessaria per la correttezza e l'utilità dell'esperimento.

Le scienze sociali, che più direttamente c'interessano in questo corso per le possibilità d'applicazione della statistica nel loro campo, di regola si vedono precluso l'esperimento e sono costrette a ricorrere a quel procedimento suppletivo di suddivisione delle osservazioni, del quale abbiamo accennato le debolezze. Pare superfluo avvertire che a tale procedimento si deve ricorrere non solo nell'interpretazione delle statistiche sociali, ma in ogni altra indagine su fenomeni collettivamente tipici che non possa valersi dell'esperimento.

7. — In tali casi si cerca di giungere, mediante successive suddivisioni, a gruppi d'osservazioni aventi comuni tutto un complesso di circostanze e differenti a due a due per una circostanza. La differenza accertata nella manifestazione media del fenomeno in due gruppi viene attribuita, in quanto sia da ritenere significativa, appunto alla circostanza per cui i due gruppi differiscono. Il procedimento è simile a quello che si adotta per i fenomeni individualmente tipici, salvo che dev'essere applicato a gruppi di osservazioni (mentre per i fenomeni individualmente tipici si può scendere fino all'osservazione singola) e che conduce a conclusioni soltanto approssimative, per la presenza di variazioni non significative, le quali perturbano tanto più il giudizio quanto più innanzi si spinge l'applicazione del metodo. Per esigenza logica il processo di suddivisione delle osservazioni dovrebb'essere spinto fino alla formazione di gruppi di osservazioni differenti tra loro, a due a due, soltanto per una circostanza, in modo da poter poi attribuire *soltanto* a questa circostanza la differenza che eventualmente presenti la manifestazione media del fenomeno nei due gruppi; praticamente, invece, il processo di suddivisione è presto arrestato dalla necessità di evitare il frazionamento delle osservazioni in gruppi poco numerosi di casi, nei quali le variazioni non significative verrebbero a sopraffare ed a nascondere quelle significative. Così che nella realtà non si giunge quasi mai a formare gruppi che differiscano, a due a due, esclusivamente per una circostanza: si riesce soltanto, nei casi più favorevoli, a formare gruppi che differiscono, a due

a due, *principalmente* per una circostanza. Per conseguenza, anche ammesso che si possano discernere ed eliminare le variazioni non significative del fenomeno, è sempre arbitrario attribuire *tutta* la differenza significativa che si riscontra tra i due gruppi nella manifestazione di esso a quella circostanza per cui essi differiscono si *principalmente* — a nostro fallibile giudizio —, ma non *esclusivamente*. Indi un nuovo fattore d'incertezza delle conclusioni: fattore che abbiamo già avuto occasione di mettere in rilievo quando abbiamo trattato degli indici di relazione.

In pratica questo procedimento destinato a supplire l'esperimento ammette numerose varianti nell'applicazione, secondo la natura dei fenomeni e delle circostanze d'osservazione e secondo i fini dell'indagine. O si suddividono le osservazioni in tanti gruppi, differenti l'uno dall'altro almeno per una circostanza, scelta fra quelle che si sa o si suppone possano influire sulla manifestazione del fenomeno; o si ordinano le osservazioni secondo la misura in cui si presenta in esse una data circostanza; oppure si suddividono le osservazioni in due gruppi, secondo che in esse si è manifestato o non si è manifestato il fenomeno; o si ordinano le osservazioni secondo la misura assunta dal fenomeno in ciascuna di esse; o si suddividono i casi del fenomeno in gruppi contrassegnati da determinati caratteri, che indicano la loro relazione con circostanze d'osservazione.

Con tutte queste operazioni di aggruppamento e di ordinamento si mira soltanto a presentare i risultati dell'osservazione nella forma più adatta per la determinazione delle relazioni tra il fenomeno e le circostanze dalle quali esso dipende: determinazione fondata essenzialmente sull'applicazione di procedimenti comparativi.

O si confronta la manifestazione del fenomeno nei vari gruppi formati secondo i due primi criteri; o si confrontano le circostanze caratteristiche dei vari gruppi formati secondo i due successivi criteri; o si esamina come le variazioni nelle manifestazioni parziali del fenomeno delimitate coll'ultimo criterio si connettano alle diverse circostanze con le quali sono in relazione, e come quindi tali circostanze vengano ad influire sulla manifestazione del fenomeno.

Si riesce così a stabilire se alla presenza od alla variazione di date circostanze corrispondano modificazioni nel fenomeno studiato. La ripetizione di simili comparazioni nel tempo, nello spazio, in diverse sezioni del campo d'osservazione, consente a sua volta di verificare l'attendibilità delle ipotesi che ciascuna comparazione ha potuto suggerire, e di giungere così, a grado a grado, ad espressioni sempre più generali delle relazioni esistenti tra il fenomeno e le circostanze d'osservazione.

Quale sia il valore di tali espressioni vedremo in seguito. Qui ci limiteremo ad esemplificare schematicamente il procedimento di suddivisione delle osservazioni.

Poniamoci un problema apparentemente semplice: come influisce l'occupazione sulla morbosità tubercolare? Per risolverlo non possiamo ricorrere all'esperimento. Esperimentare vorrebbe dire prendere vari numerosi gruppi di uomini — gruppi reciprocamente equivalenti per il sesso, l'età, la costituzione fisica, lo stato di salute, ecc., dei loro componenti — e imporre a ciascun gruppo una data occupazione, mantenendo rigorosamente identiche per tutti i gruppi quelle condizioni di vita che non differiscono per diretta e necessaria conseguenza della diversa occupazione. È chiaro che ciò riesce assolutamente impossibile. Ed allora non resta altro che surrogare all'esperimento la suddivisione di una popolazione osservata in tanti gruppi, che differiscano soltanto per l'occupazione, e in questi gruppi misurare poi la frequenza della tubercolosi. A stretto rigore anche questa meta è irraggiungibile. Scegliamo due occupazioni quali si siano, per esempio lavoranti sarti e operai meccanici: per poter correttamente comparare la morbosità tubercolare delle due categorie professionali dovremo scindere ciascuna di queste in tanti gruppi secondo il sesso, l'età, il luogo di dimora, ecc., ecc., all'intento di giungere alla formazione di gruppi identici a due a due per tutte le circostanze meno che per l'occupazione. Se non provvediamo a tale suddivisione, la differenza che riscontriamo nella morbosità tubercolare di due gruppi non deriverà solo dalla differente occupazione ma anche dalle differenze di sesso, d'età, di luogo di dimora, ecc., che sussisteranno tra la composizione del gruppo dei sarti e di quello dei meccanici. D'altra parte, per quanto spingiamo innanzi la suddivisione, non giungeremo mai alla meta, perchè non esistono un solo sarto e un solo meccanico che siano in tutto identici fra loro, salvo che nell'occupazione. Dove ci fermeremo? Dovremo fermarci ad un punto in cui i gruppi ottenuti con le successive suddivisioni siano abbastanza numerosi da assicurare una moderata ampiezza delle variazioni non significative; ma d'altra parte dovremo spingerci fino ad un punto in cui possiamo ragionevolmente ritenere che le differenze significative accertate fra due gruppi derivino *principalmente*, se non esclusivamente, dalla differenza dell'occupazione.

Come si vede, non è possibile dare norme generali per la corretta attuazione di questo processo d'analisi: in ogni applicazione concreta, la conoscenza che l'indagatore ha della materia d'indagine, la pratica ch'egli possiede dell'interpretazione dei dati statistici, il buon senso di cui egli è dotato, e fors'anche una certa particolare attitudine naturale, che può essere stata negata ad un valente teorico della statistica ed essere stata concessa a taluno che conosce appena i metodi più elementari di questa disciplina, concorrono a determinare il successo.

Un altro semplice problema: determinare come influisca una modificazione del prezzo sopra lo smercio di un prodotto. L'esperimento è escluso: il monopolista stesso può bensì modificare quando vuole e come vuole il prezzo di vendita, ma non può mantenere immutati i prezzi di tutti gli altri beni economici, nè può, come occorrerebbe per la perfetta parità di ogni altra condizione, mantenere immutate tutte le condizioni dell'equilibrio economico. Non resta che la suddivisione delle osservazioni, la quale sarà eseguita in modo da eliminare di mano in mano l'influenza di altre circostanze che possano avere concorso a modificare lo smercio del prodotto dato, di modo che la variazione significativa residua possa essere attribuita alla variazione del prezzo: compito molto più facile a prescrivere che ad eseguire.

Nei due precedenti esempi si tratta di cercare se e come una determinata circostanza influisca sopra un determinato fenomeno. Ma spesso, come sappiamo, il problema da risolvere è più vasto: si tratta di cercare quali circostanze abbiano influito sopra un determinato fenomeno. La mortalità in Italia negli ultimi sessant'anni si è ridotta da poco più di 30 a poco più di 15 per 1000 abitanti: quali sono le circostanze che hanno concorso a determinare questa grande diminuzione del tributo pagato alla morte? Se vogliamo ricercarle, e misurare approssimativamente l'influenza di ciascuna di esse, potremo cominciare col suddividere la popolazione italiana secondo il sesso; indi la suddivideremo secondo l'età: esamineremo poi come abbia variato, attraverso il tempo, la frequenza delle morti nei singoli gruppi formati secondo il sesso e l'età. Troveremo che la diminuzione della mortalità non è ugualmente forte nei diversi gruppi: questa constatazione ci aiuterà a restringere le ipotesi intorno ai fattori della diminuzione stessa, richiamando la nostra attenzione su quelli che possono aver agito nei gruppi favoriti e ponendo invece in seconda linea quelli che non possono avere ivi agito. Continueremo la suddivisione della popolazione, per esempio per regioni: ed ora che potremo determinare la frequenza delle morti in ciascuna regione in ciascun gruppo di sesso e di età saremo in grado di localizzare ancor meglio le circostanze attenuatrici della mortalità, perchè noteremo che in certe regioni esse hanno più intensamente operato, in altre meno. E se, proseguendo, divideremo ancora le popolazioni urbane da quelle rurali, potremo restringere ulteriormente il campo delle ipotesi plausibili. Qui giunti, per variare il procedimento, passeremo a suddividere i morti, in ciascun gruppo che avremo formato, secondo le cause di morte; vedremo che alcune malattie hanno grandemente diminuito la loro frequenza, o almeno la loro letalità, altre hanno perduto minor terreno, altre conservano le loro posizioni, altre infine sono divenute più spesso causa di morte. Le nostre ipotesi sulle cause della diminuzione della mortalità dovranno essere messe d'accordo con quanto avremo accertato: taluna di quelle che ci erano parse attendibili verrà scartata, ci sembrerà invece degna di considerazione tal altra che avevamo messa da parte, o ce ne apparirà qualcuna nuova, o verrà rafforzata qualcuna che già ci si era presentata. Potremo procedere ancora oltre, ma i nostri gruppi d'osservazioni diverrebbero sempre meno numerosi e le variazioni non significative dissimulerebbero sempre più quelle significative. Ad evitare siffatto inconveniente potremo eseguire separatamente le varie suddivisioni: per esempio scindere dapprima la popolazione in gruppi per sesso ed età; poi, separatamente, suddividerla per regioni; poi, separatamente, suddividere i morti per cause di morte, ecc., ecc. Così si moltiplicano le possibilità di suddivisione senza eccessiva riduzione della consistenza dei gruppi, ma in compenso diviene meno feconda la suddivisione perchè non riesce possibile determinare l'influenza della combinazione di varie circostanze sulla manifestazione del fenomeno.

Un ultimo esempio. Ci chiediamo quali circostanze abbiano influito sullo sviluppo delle esportazioni italiane in un determinato periodo di anni. Suddivideremo le esportazioni per categorie di merci, o addirittura per singole merci; di ciascuna merce esportata distingueremo le varie destinazioni; eseguiremo quell'analisi anno per anno. Così verremo a localizzare nel tempo, nelle varie categorie di merci, nelle varie destinazioni, i progressi ed i regressi; e la localizzazione eseguita ci consentirà di restringere di mano in mano il campo delle ipotesi ammissibili per spiegare le variazioni riscontrate.

È superfluo dire che in ciascuno dei precedenti esempi le ipotesi successivamente formulate potranno essere messe alla prova mediante la verifica di conseguenze che ne dovrebbero derivare nella realtà.

8. — Dopo gli esempi di applicazione supposta del procedimento di suddivisione, daremo ora due esempi di applicazione effettiva, mantenuta — per non trascendere il nostro fine — entro i limiti di un semplice saggio, e quindi non recata fino al massimo sviluppo che avrebbe potuto raggiungere.

Il primo esempio è fondato su dati demografici intorno alla natalità. I dati qui sotto riportati, per il periodo 1886-95, indicano una grande variabilità nella frequenza dei nati, da paese a paese.

PAESI	Numero medio annuo dei nati vivi per 10.000 ab.
Romania	462
Ungheria.	428
Sassonia	407
Prussia	371
Italia	365
Olanda	339
Svezia.	281
Svizzera	281
Irlanda.	231
Francia	229

Quali differenze di condizioni concorrono a determinare tanto forti differenze di natalità fra alcune popolazioni; quali analogie concorrono a determinare approssimativa uguaglianza fra altre? Non presumiamo qui di rispondere in modo esauriente a questi e ad altri quesiti, che si presentano spontanei alla mente di chi consideri i nostri dati; e neppur possiamo dilungarci in un esame generale dei dati stessi. Proveremo tuttavia ad accennare la via per la soluzione, limitandoci a sommari confronti.

Due popolazioni, anche considerate nello stesso intervallo di tempo, differiscono per tali e tante circostanze intrinseche ed estrinseche, che l'opera del più sagace e paziente studioso dei fenomeni sociali non giungerebbe a distinguerle ed enumerarle completamente. Nessuna di coteste circostanze si può dichiarare *a* *priori* non influente sulla natalità; conviene però iniziare l'analisi da quelle il cui modo di azione è noto e si presume rilevante, prima di cercare se e come altre agiscano.

Osserviamo, pertanto, che il numero delle nascite è limitato da quello delle donne atte alla procreazione. Se, in rapporto alla popolazione, queste son molto più numerose in un paese che in un altro, si avrà nel primo molto maggior frequenza di nascite, a pari intensità di esercizio dell'attività riproduttiva.

In fatto non troviamo forti divergenze tra i vari paesi. Su 10.000 abitanti, sono donne in età da 15 a 50 anni, e quindi presumibilmente idonee alla procreazione, 2.560 in Francia, 2.500 in Irlanda, 2.480 in Ungheria, 2.590 in Sassonia. Paesi d'alta e di bassa natalità presentano proporzioni poco dissimili.

Se in Francia ed in Irlanda la frequenza delle nascite tra le donne in età feconda (15-50) salisse al livello della Sassonia, il numero annuo dei nati per ogni 10.000 abitanti ascenderebbe rispettivamente a 402 ed a 393; resterebbe cioè ancora inferiore di 5 e rispettivamente di 14 alla cifra sassone, per effetto della minore proporzione di donne in età feconda. Questa minore proporzione giustifica dunque soltanto una piccola parte (5 su 178 per la Francia, 14 su 176 per l'Irlanda) del disavanzo di nascite riscontrato nelle popolazioni francese ed irlandese in confronto a quella sassone.

Riferendo il numero dei nati al numero delle donne in età feconda, vediamo persistere un profondo divario tra la natalità della Francia (89 nati per ogni 1000 donne da 15 a 50 anni) e dell'Irlanda (92), da un canto, e quella della Sassonia (157) e dell'Ungheria (172), dall'altro.

Siamo ora certi che la bassa natalità dei due paesi deriva realmente da scarsa attività riproduttiva, e non da debole rappresentanza dei gruppi idonei alla procreazione. Se indi ci proponiamo di rintracciare le circostanze che determinano una siffatta infecondità, una breve riflessione sul modo in cui le istituzioni e le consuetudini sociali limitano la riproduzione della specie umana c'induce a rivolgere la nostra attenzione alla varia frequenza dei matrimoni, come circostanza da tenere in gran conto. Esaminando i dati sulla nuzialità del periodo 1886-95, troviamo 89 matrimoni per 10.000 abitanti (media annua) in Ungheria, 90 in Sassonia, 46 in Irlanda, 79 in Francia. Comincia a delinearsi in questi numeri una antitesi, non soltanto tra l'Ungheria e la Sassonia, paesi d'alta nuialità e d'alta natalità, e l'Irlanda, paese di bassa nuzialità e di bassa natalità, ma anche tra la Francia e l'Irlanda. In Francia la bassa frequenza di nascite contrasta con una frequenza di nozze non olto inferiore a quella dei paesi più fecondi.

Eccoci dunque tratti a indagare in modo più preciso l'influenza el matrimonio sulla natalità. A tale intento ci varremo anzitutto dei censimenti eseguiti intorno al 1890, i quali ci dicono che su 1000 donne da 15 a 50 anni sono coniugate 704 in Ungheria, 549 n Sassonia, 364 in Irlanda, 545 in Francia; che quindi, su 10.000 abitanti, sono donne da 15 a 50 anni:

	Ungheria	Sassonia	Irlanda	Francia
coniugate	1.746	1.422	910	1.395
non coniugate.	734	1.168	1.590	1.165

E poi provvederemo, com'è necessario, a distinguere i nati legittimi dagli illegittimi, nelle cifre proporzionali a 10.000 abitanti riferite all'inizio del presente paragrafo:

	Ungheria	Sassonia	Irlanda	Francia
legittimi	391	356	223	209
illegittimi.	37	51	8	20

Siamo così in grado di misurare separatamente l'intensità d'esercizio dell'attività riproduttiva nel matrimonio e fuori di esso.

	Ungheria	Sassonia	Irlanda	Francia
Nati leg. per 1000 con. di 15-50 anni	224	250	245	150
Nati ill. per 1000 non con. di 15-50 anni	50	44	5	17

Ora possediamo elementi bastanti per stabilire in modo ben chiaro le differenze tra i vari paesi.

Nella proporzione delle donne in età feconda non si trovano differenze notevoli, nè concordanti con le differenze di natalità. La Francia, anzi, paese di minima natalità, ha una proporzione più alta che l'Ungheria, paese di natalità massima.

Quanto alle altre circostanze:

l'Ungheria ha un'altissima proporzione di donne coniugate; e queste sono abbastanza feconde; anche le non coniugate son più feconde che negli altri paesi;

la Sassonia ha minor proporzione di coniugate, ma più feconde che le ungheresi; mentre le donne non coniugate presentano una frequenza di parti poco minore di quella accertata in Ungheria;

l'Irlanda ha una bassissima proporzione di coniugate, però molto feconde. Fuori del matrimonio, la fecondità è debole;

la Francia, superiore all'Irlanda per proporzione di coniugate e per fecondità illegittima, segna la natalità più bassa fra i quattro paesi, a cagione della scarsa fecondità dei matrimoni.

La debole natalità deriva principalmente da scarsa frequenza di nascite legittime in Francia, da scarsa frequenza di nozze in Irlanda.

I dati statistici raccolti nella seguente tabella permettono di estendere i confronti agli altri paesi per i quali abbiamo recato notizia della natalità. Ne lasciamo la cura al lettore, per non prolungare soverchiamente l'esposizione.

PAESI	Nati vivi per 10.000 abitanti	Donne in età da 15 a 50 anni su 10.000 ab.	Coniugate su 1000 donne in età feconda	Nati vivi per 1000 donne in età feconda		
				in complesso	legitt. per 1000 coniug.	illegitt. per 1000 non coniug.
Romania	462	2.510	700	184	240	54
Ungheria	428	2.480	704	172	224	50
Sassonia	407	2.590	549	157	250	44
Prussia	371	2.490	519	149	265	24
Italia	365	2.450	556	149	249	24
Olanda	339	2.390	479	142	286	9
Svezia	281	2.400	454	117	231	22
Svizzera	281	2.550	457	110	230	9
Irlanda	231	2.500	364	92	245	5
Francia	229	2.560	545	89	150	17

I nostri dati permettono anche di valutare numericamente l'influenza delle varie circostanze che deprimono la natalità nei due paesi sui quali abbiamo fermato la nostra attenzione. Proseguendo col metodo già sopra applicato, assumeremo per termine di confronto la Sassonia.

Abbiamo già visto che soltanto una piccola parte del disavanzo di nascite, sia in Francia sia in Irlanda, deriva da minore proporzione di donne da 15 a 50 anni nelle loro popolazioni.

Qual parte — ci chiediamo ora — deriva da minore proporzione di coniugate tra le donne da 15 a 50 anni d'età? Osserviamo che, annoverandosi in Irlanda, su 10.000 abitanti, 910 donne coniugate e 1.590 non coniugate da 15 a 50 anni, mentre in Sassonia si contano 1.422 coniugate e 1.168 non coniugate da 15 a 50 anni, se ciascuna delle due classi di stato civile fosse dotata della fecondità della corrispondente classe sassone, si avrebbero in Irlanda 227 nascite legittime e 70 illegittime — in tutto 297 — per 10.000 abitanti.

Dianzi si è detto che, se in Irlanda la fecondità delle donne da 15 a 50 anni (considerate senza distinzione di stato civile) fosse al livello sassone, si avrebbero 393 nascite per ogni 10.000 abitanti; vediamo ora che la diversa composizione per stato civile riduce tale cifra a 297. Un disavanzo di 96 nascite per 10.000 abitanti è dunque determinato da quest'ultima circostanza.

Per la Francia, con analogo calcolo, si trova che soltanto un disavanzo di 2 nascite per ogni 10.000 abitanti è imputabile alla medesima circostanza.

In fatto, però, non troviamo in Irlanda 227 nascite legittime, come dal precedente calcolo, bensì 223; e 8 illegittime invece di 70. Dunque, un ulteriore disavanzo di 4 nascite proviene da minor fecondità delle donne coniugate ed uno di 62 da minor fecondità delle non coniugate. In Francia, poi, non trovando 349 nascite legittime e 51 illegittime — quante avremmo trovato se ciascuna classe di stato civile fosse stata dotata della fecondità della corrispondente classe sassone — ma rispettivamente 209 e 20, dobbiamo attribuire alla minor fecondità legittima un disavanzo di 140 nascite, alla minor fecondità illegittima un disavanzo di 31.

Ricapitolando: in confronto alla Sassonia, si hanno:

In Irlanda	In Francia	Disavanzi di nascite
14	5	nascite in meno, per minor proporzione di donne in età feconda, nella popolazione;
96	2	nascite in meno, per minor proporzione di coniugate, fra le donne in età feconda;
4	140	nascite in meno, per minore fecondità delle donne coniugate;
62	31	nascite in meno, per minor fecondità delle donne non coniugate.

Proviamoci a riassumere il nostro ragionamento in forma schematica.

Abbiamo riscontrato differenze tra i vari paesi nel rapporto che misura la frequenza delle nascite in relazione al numero degli abitanti. Tale rapporto:

$$\frac{\text{Nati}}{\text{Abitanti}} = \frac{\text{Donne atte a procreare}}{\text{Abitanti}} \times \frac{\text{Nati}}{\text{Donne atte a procreare}}$$

è uguale al prodotto di due rapporti, uno dei quali misura la proporzione in cui le donne atte a procreare entrano a comporre la popolazione totale, mentre l'altro misura la fecondità delle donne stesse. Abbiamo scorto la necessità di ricercare se le differenze tra i vari paesi corrispondessero a differenze nel primo rapporto o nel secondo, ed abbiamo accertato che le differenze stanno principalmente nel secondo.

Ora, essendo:

$$\begin{aligned} &\text{Nati} &&= \text{Legittimi} + \text{Illegittimi}, \\ &\text{Donne atte a procreare} &&= \text{Coniugate} + \text{Non coniugate}, \end{aligned}$$

il rapporto che misura la fecondità femminile si può scindere in due prodotti di rapporti:

$$\frac{\text{Nati}}{\text{Donne}} = \frac{\text{Coniugate}}{\text{Donne}} \times \frac{\text{Legittimi}}{\text{Coniugate}} + \frac{\text{Non coniugate}}{\text{Donne}} \times \frac{\text{Illegittimi}}{\text{Non coniugate}}.$$

Abbiamo determinato i valori dei rapporti che compaiono nella precedente uguaglianza; ed abbiamo trovato che le differenze tra i vari paesi dipendono in generale da differenze nei vari rapporti e non in uno solo di essi: rivelano cioè divergenza di varie circostanze, e non di una sola. Delle diverse circostanze abbiamo anche tentato di misurar l'influenza; tuttavia l'indagine non può dirsi compiuta, anzi è appena iniziata, o, meglio, accennata. Il primo passo, già mosso, nella ricerca delle relazioni tra il fenomeno e le circostanze di osservazione, renderà meno difficili i successivi.

Abbiamo accertato, per esempio, che il basso livello della natalità francese dipende principalmente da scarsa fecondità dei matrimoni. Siamo, dunque tratti a fissare la nostra attenzione sopratutto sulla fecondità delle unioni legali, ed a farne oggetto di speciale analisi.

Un'alta frequenza di coppie sterili, un'esigua proporzione di connubi molto fecondi, un generale basso livello della prolificità, ovvero il concorso di varie di queste circostanze, potrebbe determinare la povertà di figli caratteristica dei matrimoni francesi. Cercheremo di stabilire quali delle menzionate circostanze agiscano.

Assumeremo qui a termine di confronto la natalità ungherese, riferendoci tanto per la Francia quanto per l'Ungheria agli anni prossimi al 1900. Per utilizzare i soli dati disponibili, siamo costretti a valerci di cifre che comprendono anche i nati morti. Perciò i dati che riferiamo non sono correttamente paragonabili con quelli precedentemente riportati, relativi al periodo 1886-95, che comprendono i soli nati vivi.

Ecco i dati.

Dati		Ungheria	Francia
Nati per 10.000 abitanti		396	232
Donne da 15 a 50 anni su 10.000 abitanti		2.460	2.570
Coniugate su 1000 donne da 15 a 50 anni		654	577
Nati per 1000 donne da 15 a 50 anni	in complesso	161	90
	legittimi per 1000 coniugate	224	142
	illegittimi per 1000 non coniugate	42	19

Il disavanzo di 164 nascite, che presenta la Francia in confronto all'Ungheria, si scompone così:

(in più) + 18 nascite per maggiore proporzione di donne in età feconda, nella popolazione;
(in meno) — 36 nascite per minore proporzione di coniugate tra le donne in età feconda;
» — 122 nascite per minore fecondità delle donne coniugate;
» — 24 nascite per minore fecondità delle donne non coniugate.
» — 164

La differenza in meno di 164 nascite per 10.000 ab., che mostra la popolazione francese, è dunque per la massima parte dovuta a minore fecondità dei matrimoni; a questa circostanza si può attribuire, infatti, un disavanzo di circa 122 nascite.

La sterilità, patologica o fisiologica, è senza dubbio più frequente tra le donne più mature; se, per estrema ipotesi, tutte le coniugate francesi avessero più di 40 anni e tutte le ungheresi meno di 30, questa circostanza varrebbe a spiegare una notevole differenza in meno della natalità francese. Sarà dunque certamente utile ricercare come si distribuisca, per più brevi intervalli d'età, la schiera delle donne coniugate da 15 a 50 anni nei due paesi. Nel periodo di osservazione, si annoveravano in Francia 1.483 donne coniugate in età feconda, per ogni 10.000 abitanti. Fra queste 1.483 donne erano in età di anni:

15-20	20-25	25-30	30-35	35-40	40-50	
21	171	260	284	275	472.	[A]

Se invece la distribuzione proporzionale per gruppi d'età fosse stata uguale a quella accertata in Ungheria, si sarebbero avute queste altre proporzioni:

15-20	20-25	25-30	30-35	35-40	40-50	
59	235	266	270	246	407.	[B]

Ecco messa in luce una delle circostanze che determinano la minor natalità francese: essa consiste nella più bassa proporzione di coniugate giovani, maggiormente feconde. Con metodo semplicissimo possiamo ricercare quanto influisca questa circostanza.

Da rilevazioni eseguite in Ungheria risulta che, per ogni 1000 donne coniugate di ciascun gruppo d'età, si ha il seguente numero annuo di nati.

15-20	20-25	25-30	30-35	35-40	40-50	
327	346	304	249	191	89.	[C]

Se le 1.483 donne coniugate, che si contano in Francia su ogni 10.000 abitanti, fossero distribuite per gruppi d'età come le coniugate ungheresi (ossia come in [B]), e dotate in ciascun gruppo d'età della fecondità ungherese (dati [C]), si avrebbero da loro 332 nascite annue, così distribuite secondo l'età delle madri:

15-20	20-25	25-30	30-35	35-40	40-50	
19	82	81	67	47	36.	[D]

Invece, data l'effettiva distribuzione per età delle coniugate francesi (indicata in [A]), pur supposta in ciascun gruppo d'età la fecondità ungherese, si avrebbero sole 311 nascite (e non 332), così distribuite per età delle madri:

15-20	20-25	25-30	30-35	35-40	40-50	
7	59	79	71	53	42.	[E]

Un disavanzo di 21 nascite dipende dunque dalla meno favorevole distribuzione per età delle coniugate francesi.

In realtà, le donne francesi non sono dotate della fecondità ungherese, ma di una fecondità molto più bassa; per ogni 1.000

donne coniugate, in ciascun gruppo d'età, si ha infatti il seguente numero annuo di nascite.

15-20	20-25	25-30	30-35	35-40	40-50	
305	310	240	162	108	28.	[F]

Perciò le 1.483 donne francesi forniscono soltanto 210 nascite annue; ne concludiamo che una differenza in meno di 101 nascite è da attribuire alla minor fecondità delle coniugate francesi, in confronto alle ungheresi, a pari età.

Le 210 nascite si distribuiscono, per età delle madri, così:

15-20	20-25	25-30	30-35	35-40	40-50	
6	53	62	46	30	13.	[G]

Paragonando i precedenti dati con quelli in [E], si può vedere come si distribuisca nei vari gruppi d'età il disavanzo di nascite derivante da minor fecondità delle coniugate francesi.

15-20	20-25	25-30	30-35	35-40	40-50	
— 1	— 6	— 17	— 25	— 23	— 29.	[H]

Riassumendo i precedenti dati, vediamo che si hanno 121 nascite, invece di 145, da coniugate di età inferiore a 30 anni; e 89, invece di 166, da coniugate d'età superiore a 30 anni. Neppure una quarta parte del complessivo disavanzo dipende da minor fecondità delle coniugate giovani; mentre più di tre quarti di esso dipendono da minor fecondità delle coniugate mature. Nei gruppi d'età sotto trent'anni il disavanzo ammonta a 17 %, mentre nei gruppi sopra trent'anni sale a 46 %.

Tali i fatti. Ricerchiamo ora meglio il significato di essi. Suddividendo le coniugate, come abbiamo fatto dianzi, secondo l'età superiore ovvero non superiore ai trent'anni, formiamo da una parte un gruppo di donne sposate (in media) da minor tempo, dall'altra un gruppo di donne sposate (in media) da maggior tempo. I nati dalle prime saranno in buona parte primogeniti o secondogeniti; mentre fra i nati dalle seconde prevarranno i terzogeniti, quartogeniti, ecc.

Ora, poichè il disavanzo si manifesta principalmente tra i nati dalle coniugate più anziane, è ovvia l'ipotesi che esso non sia tanto notevole rispetto ai primogeniti e secondogeniti, quanto rispetto ai nati con numero d'ordine più alto. Ipotesi tutt'altro che insignificante; giacchè, se verrà confermata, potremo asserire che la scarsa prolificità media dei matrimoni francesi dipenda dalla presenza di moltissime coppie con un figlio solo o con due, piuttosto che da un'eccezionale frequenza di coppie prive di figli.

Per verificare l'attendibilità dell'ipotesi, conviene eseguire un'indagine diretta intorno al numero dei figli nati dalle singole coppie coniugali. Una siffatta indagine, compiuta in Francia nel 1906, ha accertato che su ogni 1000 capi di famiglia in età da 50 a 55 anni, 87 non avevano avuto figli, 526 ne avevano avuto da uno a tre, 259 da quattro a sei e 128 più di sei. Per meglio valutare il significato di questi dati, li metteremo a confronto con altri simili, relativi alla popolazione d'Australia nell'anno 1911. Ivi, su 1.000 capi di famiglia in età da 50 a 55 anni, 87 non avevano avuto figli, 244 ne avevano avuto da uno a tre, 316 da quattro a sei e 353 più di sei. I dati francesi non sono rigorosamente paragonabili con quelli australiani, perchè comprendono anche i divorziati e i vedovi, oltre i coniugati. Nonostante questa circostanza, che concorre ad innalzare la percentuale delle famiglie meno numerose, la quota delle unioni assolutamente sterili non appare più alta che in Australia. (Del resto la circostanza sopra accennata non può esercitare grande influenza, perchè su 100 capi di famiglia in età da 50 a 55 anni, censiti in Francia, soli 9 sono vedovi o divorziati).

Risulta pienamente confermato che la scarsa fecondità media dei matrimoni francesi è determinata sopratutto dalla presenza di numerose coppie poco prolifiche. Ed è pertanto escluso che la debole riproduttività della popolazione francese consegua da un'eccezionale frequenza dei casi di completa sterilità (patologica) di uno dei coniugi, perchè le coppie senza figli son poche — ed è raro il sopravvenire della sterilità subito dopo il primo od il secondo parto. Si presenta allora quasi necessaria l'ipotesi che il fenomeno dipenda principalmente da volontaria limitazione del numero dei figli da parte dei genitori (ipotesi che vien suffragata da molte prove).

Le ricerche dianzi disegnate ci hanno condotto ad una particolare analisi delle relazioni tra il fenomeno della natalità e le circostanze nelle quali esso si svolge; e nulla vieterebbe di proseguire con lo stesso metodo, studiando — per esempio — la frequenza delle nascite in relazione all'occupazione dei genitori, per determinare l'influsso delle condizioni economiche sulla limitazione volontaria della prole. Ma qui ci è bastato delineare a larghi tratti il metodo di ricerca; sarebbe inopportuno sostare troppo su questa particolare applicazione.

Il metodo, in sostanza, è questo. Non potendo far variare ad arbitrio le circostanze di osservazione, ci siamo contentati di passare dall'esame della manifestazione complessiva del fenomeno a quello della sua manifestazione in gruppi, delimitati secondo la presenza o assenza di certe circostanze, ovvero secondo la modalità o misura di certe altre. Abbiamo poi cercato di riconnettere il variare del fenomeno da gruppo a gruppo con la variazione delle circostanze di osservazione, ricorrendo ad ipotesi. Delle quali, poi, con nuove ricerche, abbiamo provato l'attendibilità, giungendo ad ottenerne conferma.

L'analisi della manifestazione del fenomeno per gruppi di osservazioni mostra la concomitanza di certe caratteristiche della manifestazione stessa coi caratteri distintivi di singoli gruppi. Tale concomitanza suggerisce, o almeno limita, la scelta delle ipotesi meglio idonee a spiegare i fatti accertati.

9. — Per dare un altro saggio del procedimento di suddivisione delle osservazioni, cercheremo di renderci conto della reazione del consumatore di una merce ad un rincaro del prezzo di questa.

La lavorazione e la vendita dei tabacchi in Italia formano oggetto di monopolio fiscale; si hanno quindi precise notizie intorno ai prezzi di vendita ed alle quantità consumate.

A partire dal 4 gennaio 1914 furono rialzati i prezzi di vendita di alcune qualità di tabacco da fumo. Ci chiedamo come tale circostanza abbia influito sul consumo.

Se ci riferiamo al consumo complessivo (compresi i tabacchi da fiuto), troviamo, per gli ultimi quattro anni finanziari (1° luglio-30 giugno) anteriori alla guerra europea, le seguenti cifre di consumo medio per abitante, espresso in grammi:

1910-11	1911-12	1912-13	1913-14
538	550	554	525

La tendenza all'aumento, che si osserva nei primi tre anni, è bruscamente interrotta nell'ultimo e lascia posto ad una non lieve diminuzione. Si ha dunque concomitanza tra il fenomeno dell'aumento di prezzo e quello della diminuzione di consumo. La semplice concomitanza non permette tuttavia di asserire che il secondo fatto dipenda dal primo; chè anzi notevoli variazioni di consumo possono avvenire senza cambiamento di tariffe: dal 1906-07 al 1907-08 si riscontra, per esempio, un aumento di 24 grammi nel consumo medio per abitante; mentre dal 1886-87 al 1887-88 si era manifestata una diminuzione di 21. Questa volta la variazione è più forte (29 grammi in meno); ma non è lecito per ciò escludere a *priori* che in parte abbiano concorso a determinarla circostanze diverse da quella che più attrae l'attenzione, ovvero che essa sia la risultante d'una maggiore variazione in meno dovuta all'accennata circostanza, e d'una variazione in più dovuta ad altra causa.

Se, per esempio, ammettiamo che la tendenza progressiva del consumo nel primo triennio sia derivata dal miglioramento delle condizioni economiche del paese, e se riteniamo che questa circostanza abbia continuato ad agire in ugual misura nel quarto anno, siamo condotti ad assumere la diminuzione di 29 grammi come risultante di un aumento di 8 proveniente dalle migliorate condizioni, e di una diminuzione di 37, cagionata dal rialzo dei prezzi. Scomposizione, codesta, che sarebbe bensì arbitraria, ma non più di tante analoghe valutazioni che le quotidiane necessità di ogni grande azienda sogliono richiedere. Potremmo anzi proceder oltre, ragionando così: i prezzi furono rialzati soltanto alla metà dell'anno finanziario 1913-14; nel primo semestre dunque il consumo medio dev'essere stato poco differente da quello medio semestrale del precedente anno. In ragione di 554 grammi all'anno, avrebbe dovuto ascendere a 277 grammi; poniamo 281 per tener conto del cresciuto benessere. Restano allora 244 grammi per il secondo semestre; e la differenza tra le due cifre semestrali si può scomporre così: variazione di grammi 4 in più per miglioramento di condizioni economiche; variazione di grammi 37 in meno per aumento di prezzi; risultante: variazione di 33 grammi in meno. Il rialzo dei prezzi per se solo avrebbe dunque determinata una riduzione di 13 $^{0}/_{0}$ nel consumo.

La precedente analisi è fondata sull'ipotesi che la diminuzione del consumo medio dal penultimo all'ultimo esercizio sia per intero un effetto delle modificazioni dei prezzi. Ipotesi, questa, che taluno forse accetterà; mentre altri potrà desiderare di vederla confermata da validi indizi. Cerchiamone dunque la conferma, o la correzione.

Le statistiche disponibili indicano il consumo medio per abitante anche per regioni. Ne ricaviamo i seguenti dati, espressi in grammi.

REGIONI	Consumo medio per abitante 1884-85 1885-86	1912-13	1913-14	REGIONI	Consumo medio per abitante 1884-85 1885-86	1912-13	1913-14
Piemonte . . .	643	559	528	Lazio	686	592	566
Liguria. . . .	935	840	812	Abruzzi . . .	336	400	372
Lombardia . .	710	653	608	Campania. . .	553	551	530
Veneto. . . .	814	736	692	Puglie	424	453	431
Emilia	895	730	686	Basilicata. . .	231	271	257
Toscana . · .	701	597	568	Calabria . . .	287	310	298
Marche. . . .	402	325	310	Sicilia	316	370	356
Umbria. . . .	321	312	298	Sardegna . . .	465	410	379

Si scorge subito che in tutte le regioni è scemato il consumo dal 1912-13 al successivo anno finanziario: indizio che almeno in parte la variazione deve dipendere dall'aumento dei prezzi, risentito dovunque. D'altra parte, la misura della diminuzione varia molto da compartimento a compartimento; da 45 grammi per abitante (Lombardia) a 12 grammi (Calabria). Nel Veneto e nell'Emilia si riscontrano diminuzioni poco inferiori a quella della Lombardia; nell'Umbria, in Basilicata, in Sicilia, si trovano invece diminuzioni appena lievemente superiori a quella della Calabria. Differenze tanto considerevoli fanno pensare che in parte la diminuzione di consumo non dipenda dal rincaro: che anzi nelle diverse regioni circostanze diverse concorrano ad attenuarne, o ad aggravarne, le conseguenze. I dati della precedente tabella consentono di intravedere taluna di codeste circostanze.

Anzitutto osserviamo che le più forti diminuzioni si hanno nelle regioni di più alto consumo, e viceversa. La media semplice delle cifre di consumo individuale per le otto regioni di massimo consumo, nel 1912-13, era uguale a grammi 657; la media delle corrispondenti cifre per l'anno finanziario successivo scende a 624; si ha dunque una diminuzione di grammi 33. Per le otto regioni di minimo consumo, la media discende da grammi 356 a 338; diminuzione di grammi 18. Il consumo appare più elastico dov'è più elevato; in cifra relativa però le diminuzioni si equivalgono, essendo entrambe poco superiori al 5 °/₀.

Notiamo, in secondo luogo, che le più forti diminuzioni si accertano in quelle regioni dove già il consumo accennava a sensibile decremento; mentre minori diminuzioni si riscontrano là dove il consumo tendeva ad aumentare, o meno rapidamente scemava. Le otto regioni che hanno segnato maggior diminuzione del consumo medio per abitante dal 1912-13 al 1913-14 (media semplice delle diminuzioni accertate grammi 35) presentavano già una diminuzione molto forte (grammi 72) dal 1884-86 al 1912-13; le otto regioni che hanno segnato minor diminuzione (media grammi 17) presentavano invece una lievissima diminuzione media (grammi 5) in quell'intervallo di tempo. Si è dunque tratti a pensare che, insieme con la brusca azione del rialzo dei prezzi, continui a manifestarsi quella, più lenta, della modificazione dei gusti. Che così sia, può verificarsi facilmente, esaminando se e come abbia variato il consumo dei tabacchi da fiuto, le cui tariffe di vendita non furono modificate. Si trovano sensibili contrazioni del consumo nel Veneto (8 grammi), in Liguria, Emilia, Sardegna (4 grammi), in Piemonte, Lombardia, Toscana (3 grammi), nelle Marche (2 grammi), nell'Umbria e nel Lazio (1 grammo); la sola Calabria segna un lievissimo aumento (1 grammo); nelle altre regioni non si riscontrano variazioni. Lo scemar del consumo riflette indubbiamente la graduale scomparsa dell'abitudine di fiutare tabacco.

A chi rifletta sui possibili fattori di variazione del consumo si presenta l'opportunità di tener conto anche di un'altra circostanza, oltre quelle già accennate. L'anno agrario 1913 è stato assai più favorevole del precedente per la massima parte del paese, ed è ragionevole pensare che i copiosi raccolti, recando un certo agio alle popolazioni rurali, abbiano dato impulso al consumo dei tabacchi, a dispetto dell'aumento dei prezzi. In proposito è bene ricordare che, se il consumo medio individuale è sceso da 554 a 525 grammi, è invece salita da L. 9,04 a 9,39 la spesa media per abitante.

I prezzi di molte merci, in special modo delle derrate alimentari, sono andati diminuendo dall'anno finanziario 1912-13 al 1913-14; si può ritenere che il costo dell'alimentazione, per le classi meno abbienti, si sia ridotto di circa 4 %. Possiamo credere che questa circostanza non abbia influito per nulla sul consumo del tabacco? Ci sembra di no; dobbiamo anzi ritenere che anch'essa abbia concorso ad attenuare gli effetti del rincaro dei prezzi, poichè diminuendo la spesa per i generi di prima necessità ha lasciato più ampio margine per le spese voluttuarie.

Sarebbe inutile proseguire l'analisi; a noi bastava mostrare che, pur in un caso apparentemente semplice come quello ora considerato, può riuscir impossibile l'isolamento dell'effetto di una determinata circostanza, anche se questa si mostra predominante su ogni altra.

A ragione il Direttore generale delle Privative, nella Relazione sull'Azienda dei Tabacchi per l'esercizio 1913-14, scriveva: « Sarebbe interessante conoscere in quale misura abbia veramente influito sugli introiti dell'esercizio in esame la nuova tariffa. Se non che ogni indagine diretta a tal uopo riuscirebbe ardua e potrebbe condurre a risultati, se non fallaci, almeno di scarsa attendibilità, per l'intrecciarsi di elementi opposti che si confondono od elidono a vicenda; imperocchè di fronte all'ascesa naturale del contributo in via assoluta dovuta all'accrescimento della popolazione ed a quella proporzionale dipendente dal progressivo intensificarsi del contributo individuale, sta il fatto che i consumatori, ad ogni inasprimento di tariffa, generalmente reagiscono — specie nei primi tempi — sia limitando l'uso, sia volgendosi a prodotti di minor costo, e ben pochi son quelli che si adattano in sul subito al maggior dispendio ».

Crediamo tuttavia che gli elementi forniti nei quadri statistici allegati alla citata relazione permettano un'analisi degli effetti del rincaro sul consumo, meno rudimentale di quella dianzi tentata. Si hanno, infatti, per ciascuna qualità di prodotti, notizie intorno alle vendite compiute nel periodo dell'anno finanziario 1913-14 in cui restò in vigore la vecchia tariffa, ed in quello in cui fu in vigore la nuova tariffa: notizie veramente preziose al nostro intento. Il primo dei suddetti periodi va dal 1° luglio 1913 al 3 gennaio 1914; il secondo dal 4 gennaio al 30 giugno 1914; l'uno comprende 187 giorni, l'altro 178; e pertanto, allo scopo di ottenere dati paragonabili assumeremo come unità di tempo il giorno. Avvertasi che da qui innanzi considereremo il consumo totale in Italia, e non più il consumo per abitante, per semplicità di esposizione; non derivano da ciò sensibili inconvenienti.

Il rialzo dei prezzi fu limitato ad alcune qualità di tabacchi trinciati, sigari e spagnolette, nazionali; per le altre qualità nazionali e per i tabacchi esteri non furono modificate le tariffe. Taluno potrebbe pensare che le variazioni del consumo di queste qualità non rincarate fossero idoneo indice dell'effetto risultante di tutte le circostanze diverse dal rincaro. Sia o non sia giusta questa ipotesi, per ora poco importa stabilire; in ogni modo converrà esaminare anche le variazioni di consumo delle qualità non rincarate. Ma vediamo, prima di tutto, quali modificazioni siano avvenute nel consumo delle qualità rincarate.

Consumo medio giornaliero (« levate per contanti », in kg.) delle qualità di tabacchi da fumo rincarate dal 4 gennaio 1914.

Generi	nel periodo anteriore al rincaro	nel periodo posteriore al rincaro
Trinciati	14.520	13.620
Sigari	23.104	17.303
Spagnolette	7.209	5.934

Nell'insieme, da 44.833 chilogrammi il consumo giornaliero si abbassa a 36.857, diminuendo di quasi un quinto (18 °/₀); per i trinciati diminuisce di 6 °/₀, per i sigari di 25 °/₀, per le spagnolette di 18 °/₀.

Di fronte a così forti decrementi può sorgere il dubbio che i rivenditori avessero prelevato nel secondo semestre del 1913 eccezionali quantità di tabacchi e che in parte a tal circostanza sia da attribuire la scarsezza dei prelevamenti del successivo semestre. È bene chiarire il dubbio; ci limiteremo ad esaminare le variazioni del consumo dei sigari per i quali si ebbe rialzo di prezzi. Confronteremo il consumo medio giornaliero per ciascuno dei due periodi nei quali abbiamo diviso l'ultimo esercizio, col corrispondente consumo per ciascuno dei due precedenti esercizi. Ecco i dati.

Consumo medio giornaliero (in kg.)

Qualità dei sigari	1911-12	1912-13	1913-14	
			periodo anteriore al rincaro	periodo posteriore al rincaro
5a, foggia estera	52	56	58	30
1a, comuni	21.474	21.692	21.997	16.561
2a, comuni	609	576	538	360
3a, comuni	527	506	520	360

Nel periodo anteriore al rincaro, il consumo medio giornaliero non presenta nulla di anormale; si ritrovano in esso le tendenze all'aumento od alla diminuzione riscontrate negli ultimi tempi precedenti. Invece nel secondo semestre la diminuzione è forte per tutte quattro le qualità considerate.

Il dubbio che si era presentato non trova dunque conferma nell'esame dei fatti. Resterebbe a vedere se i dati del primo semestre 1914 sul consumo corrispondano al vero con la medesima approssimazione di quelli del precedente semestre. Perchè in realtà questi dati non indicano la quantità comprata dai consumatori, bensì quella acquistata dai rivenditori; e potrebbe il rialzo delle tariffe avere, nei primi mesi, indotto i rivenditori a tenere minori scorte. Essendo nota, però, la generale esiguità delle provviste tenute dai rivenditori, è lecito asserire che questa circostanza non può aver esercitato grande influenza sui dati da noi assunti a misura del consumo.

Se guardiamo soltanto le variazioni dello spaccio delle qualità rincarate, gli effetti del mutamento di tariffa possono apparire a prima vista disastrosi per l'Azienda dei Tabacchi; giacchè la spesa media giornaliera dei consumatori, in tabacchi delle qualità rincarate, è scesa da 835.426 lire nel primo periodo a 797.892 nel secondo, diminuendo di oltre 4 %. Poichè la quantità di merce venduta si è ridotta in misura assai maggiore (18 %), è probabile che in fatto l'azienda abbia tratto un più lauto utile netto dalle 798 mila lire del secondo periodo che dalle 835 mila del primo. In ogni caso, è certo che i consumatori hanno notevolmente ridotta la loro spesa complessiva in tabacchi delle qualità rincarate.

Estendiamo ora l'indagine alle qualità che non subirono variazioni di tariffa.

Consumo medio giornaliero («levate per contanti», in kg.) delle qualità di tabacco da fumo non rincarate dal 4 gennaio 1914.

Generi	1° periodo	2° periodo
Trinciati	437	632
Sigari	908	1.502
Spagnolette	3.083	4.040

Il consumo complessivo di queste qualità balza da 4.428 a 6.174 chilogrammi per giorno, aumentando di 45 % per i trinciati, di 65 % per i sigari, di 31 % per le spagnolette.

Aumenti così cospicui e repentini devono attribuirsi per la massima parte a quella stessa circostanza che ha prevalso nel determinare le grandi e brusche diminuzioni poc'anzi accertate, ossia al mutamento dei prezzi. Il consumatore, per eludere il rialzo, si volge a qualità non rincarate, e, parzialmente o totalmente, le sostituisce nell'uso a quelle rincarate. È ovvio supporre che si rivolga a qualità di minor prezzo.

Vediamo se quest'ultima ipotesi regga alla prova dei fatti.

Nessuna qualità di trinciati, fra quelle non rincarate, aveva un prezzo inferiore al più alto dei nuovi prezzi delle qualità rincarate; il consumatore si è dunque rivolto a qualità di maggior prezzo.

Quanto ai sigari, l'aumento di consumo giornaliero delle qualità non rincarate si manifesta per 593 chilogrammi nelle qualità di prezzo non inferiore al più alto dei nuovi prezzi delle qualità rincarate, e per un solo chilogrammo in una qualità di prezzo uguale a quello nuovo della meno costosa qualità rincarata.

Per le spagnolette, infine, l'aumento di 957 chilogrammi è dovuto per 46 chilogrammi a maggior consumo di qualità di prezzo superiore a quello nuovo di tutte le qualità rincarate e per 911 chilogrammi a maggior consumo di qualità di prezzo inferiore.

Si hanno dunque aumenti di consumo così nelle qualità più costose come in quelle meno costose; par quasi che il consumatore, pur di evitare l'onere imposto dall'aumentato prezzo della qualità consueta, si sottoponga volontariamente ad un onere più grave, volgendosi a qualità migliori.

I consumatori hanno accresciuto la loro spesa media giornaliera nelle qualità non rincarate, di 46.594 lire (da 111.720 a 158.314), cioè più di quanto abbiano ridotto la spesa nelle qualità rincarate. Sicchè, nell'insieme, dal primo al secondo periodo dell'esercizio 1913-14, la spesa media dei consumatori in tabacchi da fumo è aumentata di 9.060 lire al giorno (circa di 1 %), mentre la quantità consumata scemava di 13 %.

Anche qui fermiamo l'indagine prima del limite che i dati disponibili consentirebbero di raggiungere, a noi importando delineare il metodo piuttosto che approfondire una ricerca, la quale c'interessa dal solo aspetto formale.

Il procedimento che abbiamo seguito non si discosta da quello tenuto nel precedente esempio. Accertata la successione di certe variazioni del consumo a certe modificazioni dei prezzi, abbiamo scisso l'insieme delle quantità consumate, secondo diversi caratteri (tempo e luogo del consumo, qualità, prezzo, ecc.), in tanti gruppi. Abbiamo così potuto confrontare tra loro gruppi di osservazioni differenti per varie circostanze, e quindi accertare come codeste circostanze abbiano influito. L'analisi ci ha rivelato spostamenti, riduzioni ed estensioni dei consumi, e ci ha permesso di connettere tali mutamenti con le modificazioni dei prezzi e di altre condizioni.

10. — Ulteriori esempi riuscirebbero monotoni, perchè, pur variando l'oggetto dello studio e venendo correlativamente modificati gli artifizi dell'analisi, il procedimento si può sempre ridurre alle stesse linee fondamentali, cioè:

suddivisione delle osservazioni in gruppi, differenti tra loro per una o più circostanze;

esame comparativo della manifestazione del fenomeno nei vari gruppi;

collegamento, mediante ipotesi, tra le speciali caratteristiche della manifestazione del fenomeno studiato nei singoli gruppi e le circostanze distintive dei gruppi stessi;

riprova delle ipotesi fatte, mediante l'accertamento empirico del verificarsi di conseguenze da esse dedotte.

Le difficoltà del procedimento derivano, in generale, dall'impossibilità di scindere le osservazioni in gruppi differenti o concordanti tra loro per una sola circostanza, come sarebbe necessario per la rigorosa applicazione dello schema logico. Un'altra difficoltà, di indole pratica, sta nel rapido accrescimento del lavoro di preparazione dei dati statistici, che si ha nel passaggio dalla semplice enumerazione, o dall'aggruppamento secondo caratteri isolati, all'aggruppamento secondo caratteri combinati. Di quest'ultima forma di aggruppamento appaiono chiari i vantaggi, dopo i precedenti esempi di analisi. Riferendoci al primo esempio, è ovvio che se avessimo disposto per ciascun paese di una classificazione della popolazione secondo il sesso, di un'altra secondo l'età, di una terza secondo lo stato civile; e di una classificazione dei nati secondo l'età delle madri, di un'altra secondo lo stato civile, invece di avere a disposizione una classificazione della popolazione per sesso, età e stato civile combinati, ed una dei nati secondo l'età e lo stato civile delle madri, ci sarebbe riuscito impossibile rintracciare, come abbiamo fatto, alcune delle circostanze che influiscono sulla misura della natalità.

11. — Non ci siamo preoccupati, nello svolgimento dei nostri esempi, delle variazioni non significative, perchè così nell'uno come nell'altro di essi anzitutto disponevamo di gruppi molto ampi d'osservazioni, e in secondo luogo le variazioni significative erano

grandi: di modo che le variazioni non significative si potevano ritenere trascurabili di fronte a quelle significative.

Saremmo stati inescusabili se avessimo proceduto con lo stesso criterio nell'interpretazione dei seguenti dati sulla mascolinità delle nascite, considerata in relazione a varie circostanze.

(1906-10)		(MILANO)		(SASSONIA)		(BERLINO)	
PAESI	Maschi su 1000 nati	Durata del matrimonio Anni	Maschi su 1000 nati	Madri che avevano avuto figli in numero di	Maschi su 1000 ultimi nati	Differenza fra l'età del padre e quella della madre	Maschi su 1000 nati
Romania. . .	515	< 1	516	1	514		
Ungheria . .	514			2	517	oltre 3 anni in meno	514
Sassonia. . .	512	1 — 3	507	3	512		
Prussia . . .	514			4	514		
Italia	513	3 — 5	514	5	513		
Olanda . . .	512			6	513	da 3 anni in meno a 3 anni in più	514
Svezia . . .	516	5 — 10	514	7	515		
Svizzera. . .	511	10 — 20	511	8	510		
Irlanda . . .	513			9	515		
Francia . . .	511	> 20	517	10	512	oltre 3 anni in più	514

Se anche le variazioni della mascolinità da paese a paese, secondo la durata del matrimonio, secondo il numero d'ordine della nascita, secondo la differenza d'età fra i genitori non sono per intero non significative, certamente lo sono in parte considerevole e probabilmente in parte preponderante.

In casi come questo l'applicazione del metodo comparativo serve piuttosto ad escludere l'influenza di date circostanze sul fenomeno studiato, che a determinarla o a misurarla. Comunque variino le condizioni d'osservazione, il fenomeno appare statisticamente stabile, o almeno si accosta a tale forma di regolarità.

12. — Dopo gli esempi dati di indagini sulle relazioni tra i fenomeni collettivamente tipici e le circostanze nelle quali essi si manifestano, ci appaiono in nuova luce alcuni metodi che nel nostro precedente studio avevamo presentato come semplicemente intesi alla descrizione dei fenomeni.

I procedimenti interpolatorii, le tavole di eliminazione e di frequenza, i vari indici della dipendenza e della correlazione, non sono altro, infatti, che mezzi per la descrizione dei fenomeni. Ma siamo ora in grado d'intendere come la descrizione, eseguita con particolari artifici, giovi ad illuminare l'indagine sulle circostanze dalle quali dipendono i fenomeni studiati. Col formare una tavola di mortalità, si ordinano le frequenze delle morti in funzione dell'età, e quindi ci si mette in condizione di poter facilmente stabilire relazioni tra la mortalità e la circostanza « età », o meglio fra la mortalità e le molteplici circostanze (resistenza organica, alimentazione, esposizione ad agenti esterni patogeni, occupazione, ecc.), che variano col variare dell'età. Col rappresentare in una formola il consumo d'una merce in funzione del prezzo, si fa un passo verso la determinazione della relazione d'interdipendenza esistente tra le due variabili economiche. Col riassumere, in un coefficiente o indice, la correlazione esistente tra la statura dei genitori e quella dei figli, si tende a determinare l'influenza della circostanza « eredità » sul fenomeno « statura ». La stessa serie statistica, ordinamento delle misure di un fenomeno secondo la modalità o misura di una circostanza, ci presenta già l'andamento del fenomeno in corrispondenza al modificarsi di tale circostanza, di modo che ci agevola la ricerca della relazione eventualmente esistente fra l'uno e l'altra.

L'impiego di quelli che ci erano apparsi semplici strumenti per la descrizione dei fenomeni non conduce sempre direttamente all'accertamento delle relazioni che può interessar di stabilire tra i fenomeni stessi e le circostanze nelle quali essi si manifestano. Ma permette, in generale, di limitare grandemente il campo delle ipotesi che conviene mettere alla prova, e così facilita la scelta delle ipotesi più attendibili.

Quesiti: 1. — La ricerca di relazioni tra le variazioni di fenomeni collettivamente tipici e le condizioni d'osservazione dovrebb'essere, in massima, preceduta dall'eliminazione delle variazioni non significative. In pratica, quando si può omettere tale operazione? quando, invece, essa è indispensabile?

2. — Se col variare delle condizioni di osservazione un fenomeno collettivamente tipico presenta soltanto variazioni non significative, come si può classificare tale fenomeno?

3. — Perchè il problema: « quali circostanze influiscano, e come, sopra un dato fenomeno » si riduce normalmente all'altra forma « se e come una data circostanza influisca sopra un dato fenomeno »? Con quali procedimenti comparativi si cerca di risolvere questo secondo problema riferito a fenomeni individualmente tipici? Quale carattere hanno le conclusioni che si possono

raggiungere coll'applicazione di tali procedimenti? In qual modo, mediante l'ulteriore applicazione dei procedimenti comparativi, si cerca di dare maggiore sicurezza alle conclusioni o di rettificarle? Qual è, in questa fase, la funzione dell'esperimento? Perchè, come strumento logico, l'esperimento di regola è più efficace che l'osservazione del fenomeno nel suo libero svolgimento? Che cosa s'intende per « legge empirica » ?

4. — Con quali modificazioni i procedimenti comparativi che s'impiegano nel ragionamento sui fenomeni individualmente tipici possono essere applicati a fenomeni collettivamente tipici? Si applica ancora il procedimento comparativo a casi singoli? Si ottengono conclusioni rigorose? Perchè, anche quando si tratti di fenomeni statisticamente stabili, le conclusioni possono essere soltanto approssimative? Perchè le conclusioni divengono incerte quando si tratti di fenomeni che non presentano una stabilità statistica?

5. — È esclusa la possibilità dell'esperimento sui fenomeni collettivamente tipici, come tali? Come va applicato ad essi? Qual è il carattere delle conclusioni che se ne possono trarre?

6. — Come si applica ai fenomeni collettivamente tipici il procedimento di progressiva suddivisione delle osservazioni, inteso a supplire l'esperimento dove questo non è attuabile? Qual ragione consiglia a non spingere la suddivisione agli estremi limiti possibili? Come si riflette l'imperfetta applicabilità del procedimento di suddivisione sulla validità delle conclusioni?

7. — Si spieghi come il procedimento di suddivisione sia stato applicato negli esempi esposti nel testo e si esaminino criticamente le conclusioni raggiunte. Si mostri come il procedimento stesso consenta una progressiva limitazione delle ipotesi ammissibili intorno alle relazioni tra il fenomeno e le circostanze d'osservazione.

Esercizi: — NOTA. Esercizi di ragionamento sui risultati dell'osservazione escono dai limiti di un corso elementare di statistica metodologica; essi sono invece sommamente opportuni nei corsi di statistica applicata (statistica economica, demografica, antropologica, giudiziaria), che vengono tenuti in Università ed Istituti Superiori. Per poter ragionare sui fatti, bisogna che lo studente abbia già una conoscenza non superficiale della materia cui questi si riferiscono: il che di rado è possibile agli studenti di primo anno che seguono il corso di statistica metodologica nelle nostre Facoltà giuridiche od economiche.

Per addestrarsi al ragionamento sui fatti, espressi nei dati statistici, il giovine potrà utilmente esaminare, con spirito critico, ragionamenti altrui. I giornali quotidiani, le riviste economiche, i libri su questioni sociali, offrono al volonteroso una miniera inesauribile di esempi di ricerca delle relazioni tra fenomeni sociali e circostanze nelle quali essi si manifestano; e molto spesso il lettore non stenta a trovare errori logici, specialmente negli scritti a tesi che sono di gran lunga più numerosi di quelli ispirati a rigorosa obiettività. Sarà utile al giovine non solo rintracciare gli errori, ma anche cercare di correggerli, sostituendo al ragionamento errato un ragionamento corretto, in quanto glielo consentano gli elementi offerti dagli scritti criticati e quelli che egli potrà desumere da altre fonti (per ciò che riguarda dati su fenomeni sociali, specialmente l'ASI e le pubblicazioni ivi citate). L'esperienza didattica attesta che il desiderio di rettificare le altrui conclusioni errate costituisce uno dei più efficaci stimoli allo studio scolastico, come del resto è un poderoso propulsore del progresso scientifico.

È istruttiva anche la lettura delle relazioni, fondate su dati statistici, che vengono pubblicate annualmente da amministrazioni pubbliche o semi-pubbliche, come quelle dell'ISTITUTO CENTRALE DI STATISTICA sul movimento della popolazione e sulle cause di morte, quelle del MINISTERO DELLE FINANZE sulle aziende statali dei tabacchi e dei sali, quelle del MINISTERO DELLE COMUNICAZIONI sulle aziende statali delle ferrovie e delle poste, quella del MINISTERO DELLE CORPORAZIONI sull'industria mineraria, quella della BANCA D'ITALIA, quella dell'ISTITUTO NAZIONALE DELLE ASSICURAZIONI, ed altre, che non solo presentano al lettore ragionamenti da esaminare ma gli offrono anche elementi per ulteriori sviluppi delle indagini sulle relazioni tra i fenomeni descritti e le circostanze d'osservazione.

Alcune pubblicazioni periodiche, le quali si propongono di presentare una cronaca dei fenomeni economici, redatta con criteri scientifici, come *L'Italia economica* di R. BACHI, le *Prospettive economiche* di G. MORTARA, *La vita economica italiana* dell'ISTITUTO DI STATISTICA DELLA R. UNIVERSITÀ DI ROMA, possono essere studiate con grande vantaggio, in quanto l'interpretazione ivi data ai fatti e le previsioni esposte per l'avvenire possono venire rettificate mediante lo studio degli avvenimenti posteriori, in modo da far risaltare gli errori che erano insiti nei ragionamenti degli autori.

Lo studente potrà consultare proficuamente opere dedicate all'interpretazione scientifica di fenomeni sociali: per l'analisi critica del ragionamento condotto sui risultati dell'osservazione statistica si prestano meglio, in generale, le monografie che i trattati o i manuali. Indichiamo, a titolo di esempio: BENINI R., *Di alcuni punti oscuri della demografia* (in *Giornale degli economisti*, 1896); BENINI R., *La demografia italiana nell'ultimo cinquantennio* (in *Cinquanta anni di storia italiana*, Milano, Hoepli, 1911: pubblicazione dell'ACCADEMIA DEI LINCEI); BOLDRINI M., *Sviluppo corporeo e predisposizioni morbose*, Milano, Soc. ed. Vita e Pensiero, 1925; BOSCO A., *Divorzi e separazioni personali di coniugi* (in *Annali di statistica*, serie IV, vol. 94 bis: pubblicazione della DIREZIONE GENERALE DELLA STATISTICA); BRESCIANI-TURRONI C., *Le vicende del marco germanico* (in *Annali di economia*, vol. VII, 1931: pubblicazione dell'UNIVERSITÀ BOCCONI); COLETTI F., *La popolazione rurale in Italia*, Piacenza, Federazione Italiana dei Consorzi Agrari, 1925; EINAUDI L., *La guerra e il sistema tributario italiano*, Bari, Laterza, 1927; GINI C., *Il sesso dal punto di vista statistico*, Palermo, Sandron, 1908; GINI C., *L'ammontare e la composizione della ricchezza delle nazioni*, Torino, Bocca, 1914; LIVI L., *Gli Ebrei alla luce della statistica*, Firenze, Vallecchi, 1920; MORTARA G., *Effetti economici della diminuzione della mortalità* (parte III delle *Lezioni di statistica economica e demografica*, Roma, Athenaeum, 1920); MORTARA G., *La salute pubblica in Italia durante e dopo la guerra*, Bari, Laterza, 1925; NICEFORO A., *Considérations sur les rapports présumés entre le cancer et la race* (in collaborazione con E. PITTARD: pubblicazione della SOCIÉTÉ DES NATIONS, Genève, 1926); PANTALEONI M., *La crisi del 1905-1907* (in *Annali di economia*, vol. I, 1924); PARETO V., *Cours d'économie politique*, Lausanne, Rouge, 1896-1897.

Tra le riviste italiane aventi fini prevalentemente scientifici ricordiamo il *Giornale degli economisti*, la *Riforma sociale*, gli *Annali di economia*, il *Metron*, la *Rivista italiana di statistica*. Anche la *Rivista di politica economica*, la *Rivista bancaria* e l'*Economia* contengono talvolta articoli che possono essere utilmente studiati al fine che qui c'interessa.

È superfluo avvertire che ragionamenti di valorosi scienziati possono essere talora mal condotti o addirittura errati; lo studente, pertanto, cercherà d'apprendere dall'esempio dei maestri il corretto ragionare, ma non li seguirà mai ad occhi chiusi nelle loro induzioni, anzi dovrà essere più fiero se scoprirà un errore nel ragionamento disinteressato dell'uomo di scienza che se coglierà in fallo — com'è assai più facile — il polemista interessato o il gazzettiere frettoloso.

CAPITOLO XXIII.

Le uniformità statistiche.

Stabilità statistica e leggi statistiche — Riserve sull'uso del termine « legge » — Un esempio di labilità d'una presunta legge statistica — Caratteri di alcuni principali tipi di uniformità statistiche: variazione del fenomeno entro limiti determinati; esistenza di una misura o modalità normale di manifestazione del fenomeno; lenta variazione nel tempo; approssimativa costanza nella distribuzione di modalità o misure; costanza o ripetizione di tendenze: variazioni evolutorie, cicliche, periodiche.

1. — Dobbiamo ora chiederci qual significato abbiano le relazioni che, coi procedimenti descritti nel capitolo precedente, possono scoprirsi tra fenomeni collettivamente tipici e le circostanze nelle quali si manifestano; e quale sia il campo di validità di tali relazioni.

A questo proposito bisogna premettere una distinzione tra i casi nei quali si riscontra una vera e propria stabilità statistica nella manifestazione dei fenomeni — stabilità, nel senso spiegato nel capitolo XXI — e tutti gli altri casi.

Talvolta l'esame della manifestazione di un fenomeno collettivamente tipico mostra che, in determinate circostanze, esso tende a presentarsi in misura costante col crescere del numero delle osservazioni. In questo caso, se ci riferiamo ad un grandissimo numero di osservazioni, possiamo quasi assimilare ad un coefficiente fisico l'uniformità accertata. Il numero 45,5, cui tende la media aritmetica dei risultati di una serie di sorteggi dall'urna del lotto; il numero 0,25 cui tende la frequenza relativa dei conigli bianchi nella seconda generazione proveniente dall'incrocio tra razza di pelame bianco e razza di pelame nero, hanno significato e validità non molto dissimile dal numero 8,83 che esprime il peso specifico del manganese o dal numero 0,000022 che rappresenta il coefficiente di dilatazione termica (lineare) dell'alluminio. A prima vista, tra i numeri i quali esprimono leggi fisiche e quelli che esprimono leggi statistiche pare intercedere una fondamentale differenza, essendo i secondi solo *approssimativamente validi per un grande numero di osservazioni*, e non validi per l'osservazione singola. Nè si può negare che appunto tale differenza distingua le due categorie di leggi; ma se si riflette che alla « singola osservazione » può fare riscontro la singola molecola o il singolo atomo, si vede attenuarsi l'apparente contrasto. La costanza del risultato si manifesta quando l'osservazione è estesa ad un grandissimo numero di molecole; in molti casi è probabile che la costanza apparirebbe assai meno approssimativamente nell'osservazione di un piccolo numero di molecole, e scomparirebbe addirittura nell'osservazione delle singole molecole.

Le regolarità dei fenomeni collettivamente tipici, del genere di quelle ora citate, sono vere e proprie *leggi statistiche*: hanno cioè indefinita validità nel tempo e nello spazio.

Ma, come più volte abbiamo avvertito, nella maggior parte dei casi lo studioso del fenomeno collettivamente tipico non riesce a stabilire un insieme di condizioni tali che la misura del fenomeno tenda alla costanza coll'estendersi delle osservazioni. Qual significato e qual validità resta allora alle uniformità accertate, alle relazioni riscontrate tra i fenomeni e le circostanze di osservazione? Hanno esse il puro valore storico di descrizioni d'un passato che non ritornerà mai, o possono giovare, ed entro quali limiti, per la previsione del futuro?

2. — Sarà bene prendere in esame un caso concreto. Risaliremo poi, meno difficilmente, dal particolare al generale.

Seguendo, nella popolazione olandese, i nati negli anni 1819-23 per il corso della loro esistenza, si è potuto accertare quanti di essi abbiano raggiunto varie età; ed in ispecie quanti abbiano toccato il cinquantesimo compleanno, il cinquantunesimo, il cinquantaduesimo, ecc. I risultati della statistica delle morti e quelli dei censimenti, controllandosi a vicenda, hanno permesso di calcolare misure, che devono ritenersi molto prossime al vero, della frequenza delle morti nei singoli anni d'età. La seconda colonna della tabella più avanti riferita raccoglie i valori di codesti rapporti di frequenza per gli intervalli annuali d'età compresi fra il cinquantesimo ed il settantesimo compleanno. L'insieme dei dati ci indica come varii il fenomeno in esame (frequenza delle morti) col variare dell'età. Applicando procedimenti interpolatorii potremmo agevolmente ottenere una rappresentazione della frequenza delle morti in funzione dell'età; ma preferiamo qui riassumere in altro modo le indicazioni desunte dall'indagine in esame. Questa fornisce tutti gli elementi necessari al calcolo di una tavola di sopravvivenza per la generazione studiata, per le età da 50 a 70 anni. Dalla tavola — che per brevità omettiamo — si può desumere poi il rapporto fra il numero dei sopravviventi a 70 anni e quello dei sopravviventi a 50; esso risulta uguale a 0,5135; ciò significa che poco più di metà dei superstiti a 50 anni (5.135 su ogni 10.000) hanno raggiunto il settantesimo compleanno, mentre gli altri son morti prima di toccarlo.

Immaginiamo ora che gli Olandesi nati nel 1840, cioè all'incirca vent'anni dopo la generazione dianzi considerata, avessero voluto utilizzare i risultati dell'esperienza, per assicurarsi un capitale nella vecchiaia, riunendosi in associazione mutua. Facendo assegnamento sulla costanza dei saggi di mortalità, avrebbero potuto ragionare così: La somma da impiegare oggi, al saggio del 4 %, per ottenere 1000 fiorini in capo a vent'anni, è uguale a fiorini 456,38. Così che per essere sicuro di ottenere 5.000 fiorini all'età di 70 anni, ciascuno di noi potrebbe, nel giorno del suo cinquantesimo compleanno, investire al 4 % la somma di fiorini 2.281,93. Ma se ci riuniamo in associazione mutua è sufficiente al medesimo intento una somma più modesta: poichè su ogni 10.000 superstiti a cinquant'anni, soli 5.135 raggiungono i settanta, basta che ciascuno dei cinquantenni associati investa 5.135 decimillesimi della somma di fiorini 2.281,93, ossia fiorini 1.171,77. Il calcolo è rigoroso: supponendo, per semplicità, nulle le spese di amministrazione, e nullo il rischio dell'investimento, ogni gruppo di 10.000 viventi a 50 anni investirà 11.717.735 fiorini, che dopo quattro lustri saranno cresciuti a 25.675.000, quanti occorrono appunto per versarne 5.000 a ciascun superstite settantenne.

Così avrebbero forse ragionato i nati nel 1840, quand'erano prossimi a raggiungere i cinquant'anni. Da allora in poi è trascorso un tempo sufficiente perchè noi ora possiamo verificare se la realtà abbia corrisposto alle previsioni. Anche per i nati nel 1840 è stato osservato l'ordine di estinzione secondo l'età; anche per loro è stata poi computata la frequenza delle morti nei singoli anni d'età fra il cinquantesimo e il settantesimo compleanno (vedasi la terza colonna della tabella). Il confronto di questi dati con quelli che si riferiscono alla generazione del 1819-23 palesa una forte riduzione della mortalità in tutti gli intervalli d'età considerati: riduzione che è indicata nella sesta ed ultima colonna della tabella.

Dividendo il ventennio di età in quattro intervalli quinquennali, vediamo che nel primo di questi la mortalità è discesa di 16 %₀, nel secondo di 15, nel terzo di 16, nel quarto di 14 %₀.

ETÀ x	Morti in età $x, x+1$ per 100.000 sopravviventi all'età x				Differenza tra la mortalità dei nati nel 1840 e quella dei nati nel 1819-23
	nati negli anni		viventi negli anni		
	1819-23	1840	1870-74	1889-93	
(1)	(2)	(3)	(4)	(5)	(6) = (3) — (2)
50	1.806	1.637	1.806	1.622	— 169
51	1.905	1.590	1.965	1.668	— 315
52	1.989	1.560	1.999	1.746	— 429
53	2.029	1.719	2.172	1.900	— 310
54	2.142	1.799	2.255	1.922	— 343
55	2.221	1.916	2.409	2.124	— 305
56	2.362	1.929	2.501	2.222	— 433
57	2.535	1.967	2.601	2.433	— 568
58	2.699	2.295	2.980	2.642	— 404
59	2.864	2.692	3.271	2.792	— 172
60	3.013	2.587	3.295	3.004	— 426
61	3.065	2.720	3.607	3.332	— 345
62	3.397	2.971	3.879	3.496	— 426
63	3.767	2.987	4.273	3.830	— 780
64	4.044	3.332	4.379	4.196	— 712
65	4.267	3.645	4.855	4.438	— 622
66	4.588	4.160	4.961	4.787	— 428
67	5.131	4.398	5.510	5.177	— 733
68	5.473	4.725	5.806	5.555	— 748
69	6.087	5.118	6.558	6.087	— 969

Per avere una rappresentazione sintetica della diminuzione di mortalità, ricerchiamo anche per questa generazione quanti dei superstiti a 50 anni raggiungano i 70. Troviamo 5.674 su 10.000. Se i superstiti a 50 anni avessero realmente istituito l'associazione mutua nel modo dianzi spiegato, i 25.675.000 fiorini spettanti in media a ciascun gruppo composto in origine di 10.000 persone andrebbero dunque divisi fra 5.674 — e non fra soli 5.135 — superstiti a 70 anni. Occorrerebbero 2.695.000 fiorini in più dei disponibili, per dare a tutti lo sperato capitale, oppure bisognerebbe che ciascuno si acconciasse a vedere ridotto da 5.000 a 4.525 fiorini il suo patrimonietto. Nell'associazione mutua, ciascun socio potrebbe trovar conforto al sacrifizio pensando che egli è forse uno dei 539 superstiti in più dei previsti; ma se invece l'assicurazione fosse esercitata da un'impresa avente fine di lucro, qual delusione si sarebbe questa preparata! Se pur avesse richiesto a ciascun cinquantenne della generazione 1840 il versamento di 1.300 fiorini invece che di 1.171,77, dovrebbe ancora rimettere del proprio per pagare le somme assicurate, senza tener conto della perdita per il mancato rimborso delle spese di esercizio.

3. — Abbiamo illustrato i dati sulla mortalità col calcolo dei risultati di un'assicurazione soltanto allo scopo di rendere più tangibile la rilevanza delle variazioni di frequenza delle morti nel tempo; variazioni sulla cui realtà e grandezza non è lecito il dubbio, perchè innumerevoli esperienze le confermano.

Il numero che misura la frequenza delle morti in un breve intervallo d'età non è dunque assimilabile in alcun modo ad un coefficiente fisico, per quanto riguarda la variabilità nel tempo e nello spazio. Neppure per generazioni abbastanza prossime fra loro nel tempo, come quelle considerate, e appartenenti alla medesima popolazione, esso tende a rimaner costante; maggiormente varia, poi, tra epoche più distanti e da paese a paese.

Ma — possiamo chiederci — v'è almeno qualche cosa di costante nella variazione della mortalità in funzione dell'età? L'andamento generale della curva di frequenza delle morti è analogo, a prescindere dalle variazioni non significative, per le due generazioni considerate; solo le ordinate della curva più recente appaiono ridotte nella proporzione di 14 a 16 $^0/_0$, in confronto a quelle della più antica. Calcolando la ragione media geometrica di aumento della mortalità per ogni anno d'età fra il cinquantunesimo e il settantesimo, la troviamo uguale a 1,063 per la generazione vecchia, a 1,063 per la nuova. Ecco un'uniformità; saremmo quasi tentati di enunciare una *legge statistica* dell'aumento della mortalità con l'età, assumendola poi come espressione del decadere della resistenza organica con l'invecchiamento dell'organismo.

Ed invero, se le condizioni d'ambiente si mantenessero immutate, il variare della frequenza delle morti in relazione all'età indicherebbe appunto variazione della resistenza organica. Ma se le condizioni d'ambiente si modificano mentre una generazione attraversa un certo periodo della sua esistenza, la diversa frequenza delle morti nei diversi anni d'età compresi in quel periodo dipende non soltanto da variazione della resistenza organica bensì anche da mutamento delle condizioni di vita. Se il miglioramento delle condizioni esterne fosse molto rapido, ne potrebbe derivare una stazionarietà, o perfino un decremento, della mortalità col crescere dell'età. Dunque, perchè sia lecito assumere la variazione della mortalità con l'età ad indice della variazione di resistenza organica, si deve procurar di ottenere uguaglianza di condizioni esterne nell'osservazione delle successive età.

A ciò si giunge facilmente considerando la frequenza delle morti, per gruppi d'età, tra le persone *viventi* in un dato periodo di tempo, invece che tra le persone *nate* in un dato periodo. I nati nel quinquennio 1819-23 incontrano diverse condizioni d'ambiente nei successivi anni solari che attraversano tra il loro cinquantesimo ed il settantesimo compleanno. Ma, invece, le persone in età da 50 a 51, da 51 a 52..., da 69 a 70 anni, viventi *contemporaneamente* in un dato periodo, per esempio nel 1870-74, si trovano in uguali condizioni esterne; e quindi la diversa mortalità delle diverse età misura adeguatamente la diminuzione della resistenza organica col crescere dell'età.

Nelle colonne quarta e quinta della nostra tabella abbiamo appunto indicato la frequenza delle morti per intervalli annuali d'età, dal cinquantesimo al settantesimo compleanno, fra i viventi in due diversi periodi: dal 1870 al 1874 e dal 1889 al 1893. Anche qui si nota una riduzione della mortalità, nel tempo; ma la grandezza relativa della riduzione va diminuendo col crescere dell'età: essa è di 13 % per le età da 50 a 55 anni, di 11 % per il successivo lustro d'età, e rispettivamente di 8 e di 6 % per i due seguenti. Non è dunque indifferente considerare l'uno o l'altro periodo quando si ricerca una misura della ragione d'incremento della mortalità: per il periodo più antico la troviamo uguale a 1,070, per il più recente a 1,072. Ragioni entrambe notevolmente più alte di quella (1,063) che avevamo trovato prima senza tener conto del miglioramento delle condizioni di ambiente.

Ma neppur ora possiamo sottrarci ad una obiezione: giacchè, se prima non tenevamo conto delle differenze d'ambiente fra i diversi anni d'osservazione, attraverso i quali seguivamo una generazione, ora non teniamo conto delle differenze di resistenza organica tra le varie generazioni che attraversano diversi intervalli d'età nello stesso periodo di osservazione; di modo che sarebbe vana lusinga quella di aver ottenuto una misura attendibile dell'influsso della circostanza *età* sulla frequenza delle morti.

Ad illustrare meglio la difficoltà dell'indagine, mostreremo come non vada trascurata un'altra circostanza. Oltre la tendenza continua al miglioramento, connessa coi progressi sanitari ed igienici, si nota anche una tendenza della mortalità a variare concordemente nei diversi intervalli annuali d'età, da anno ad anno solare; tendenza che deve trovare la sua spiegazione principalmente nelle modificazioni di fattori meteorologici e nella variabile diffusione delle malattie epidemiche. Ecco alcuni dati in proposito, sempre per l'Ol anda

ETÀ x	Morti in età x, $x+1$ per 100.000 sopravviventi all'età x			
	1900	1901	1902	1903
50	1.322	1.212	1.356	1.195
51	1.441	1.319	1.300	1.414
52	1.447	1.392	1.473	1.338
53	1.576	1.431	1.555	1.594
54	1.888	1.630	1.814	1.612
55	2.079	1.905	1.804	1.541
56	2.045	1.991	2.064	1.741
57	2.014	2.009	1.957	2.014
58	2.297	2.223	2.280	2.011
59	2.692	2.293	2.352	2.112

Dal 1900 al 1901 in ciascun anno d'età fra il cinquantesimo e il sessantesimo compleanno la frequenza delle morti è diminuita (in media di 8 ‰); essa è invece cresciuta, in quasi tutti gli anni d'età, dal 1901 al 1902 (in media di 3 ‰); dal 1902 al 1903 è di nuovo diminuita (in media di 8 ‰), scendendo ancora più giù che nel 1901. Queste oscillazioni si sovrappongono e s'intrecciano alle variazioni connesse coi mutamenti della resistenza organica, e concorrono a renderne ardua la misurazione.

Quanto molteplici siano le circostanze che influiscono sulla mortalità non fa d'uopo ulteriormente chiarire. Eppure si tratta proprio d'uno di quei fenomeni la cui osservazione suggerì la prima idea d'una costanza dei fenomeni collettivamente tipici; e appunto la mortalità fu assunta spesso a modello di tale costanza. Ebbene: pur fondandosi la previsione della mortalità sopra dati vecchi di due o tre decenni soltanto, essa rimane lontana, e non poco, dal vero. Ciò deriva principalmente dalla difficoltà di isolare gli effetti, sulla mortalità, di singole circostanze. È l'impossibilità dell'esperimento causa principale di tale difficoltà, e in questo esempio essa costituisce un ostacolo forse insormontabile all'accertamento di vere e proprie *leggi*.

4. — Dal precedente esempio, scelto tra quei fenomeni che pure suscitano l'impressione di una mirabile regolarità degli eventi della vita sociale, dobbiamo dunque concludere che, fuori dei casi di stabilità statistica accennati all'inizio di questo capitolo — casi frequenti nelle scienze fisiche, rari nelle biologiche, rarissimi od inesistenti nelle sociali — non si possa parlare di *leggi statistiche*? Crediamo che appunto questa sia la conclusione corretta, se per legge s'intende un'uniformità che si manifesti in modo assolutamente costante nel tempo e nello spazio.

In pratica si parla di leggi in moltissimi casi nei quali si riscontrano uniformità approssimative e tutt'altro che permanenti. Una brevissima analisi dei caratteri di tali uniformità, che meglio potrebbero dirsi *quasi-uniformità*, servirà a completare il nostro studio, mostrando come i dati dell'esperienza possano fornire elementi utili per la previsione, anche fuori dei casi di vera e propria costanza dei fenomeni collettivamente tipici. Diciamo fino da ora che ci sembra preferibile evitare la denominazione di *leggi* per queste uniformità, valide soltanto entro ristretti limiti di tempo e di spazio; ma insistiamo nell'avvertire che la scienza, ed ancor più la pratica, dove non sono in grado di riscontrare vere e proprie leggi, non possono nè devono disprezzare l'accertamento di simili uniformità, o quasi-uniformità. Molti progressi delle società umane sono stati resi possibili dalla conoscenza di esse; e tutta la vita sociale si fonda sulla previsione di certe uniformità di fenomeni collettivamente tipici, non rigorose ed immutabili, bensì instabili ed approssimative. Di tali uniformità passeremo in rassegna i principali tipi.

5. — Notiamo, anzitutto, che in molti casi, mentre la misura d'un fenomeno collettivamente tipico (o della relazione di grandezza tra due fenomeni: si sottintenda quest'alternativa nel seguito del presente paragrafo) appare variabile nel tempo e nello spazio e mal si presta alla previsione, rimane però sempre, o quasi sem-

pre, contenuta entro determinati limiti. Talvolta si possono fissare limiti dai quali il fenomeno non esce *mai*, tal altra si possono fissare limiti dai quali esce soltanto in rarissimi casi. Non per la scienza, ma per la pratica, i due casi si equivalgono. Se pure l'antropologo non è in grado di stabilire con sicurezza che nessun uomo può superare la statura di due metri e mezzo, l'architetto fissa tranquillo a due metri e mezzo l'altezza delle porte nell'appartamento che costruisce, sicuro praticamente che i futuri inquilini non avranno bisogno di abbassare il capo per passarvi. In questo esempio sono le misure individuali del fenomeno che rimangono contenute entro certi limiti, così che si può dire che se la misura individuale del fenomeno è collettivamente tipica, nel senso che presenta regolarità solo al considerare la media o la distribuzione in un grande numero d'osservazioni, i limiti entro i quali essa varia sono invece individualmente tipici, nel senso che il fenomeno è sempre contenuto fra essi.

In altri casi sono i limiti non della misura individuale del fenomeno ma della misura collettiva di esso — totale o media — che possono riguardarsi come individualmente tipici. Così l'agricoltore, mentre non è in grado di prevedere qual rendimento otterrà dal suo campo coltivato a barbabietole, è però in grado di stabilire un minimo numero di quintali, che quasi certamente sarà superato, ed un massimo, che certamente non sarà raggiunto. La previsione dei due limiti, se non vale quanto una previsione *precisa*, è tuttavia utile sussidio alla previsione approssimativa che l'agricoltore fa sul rendimento. Così pure il tessitore, benchè non possa conoscere anticipatamente a qual prezzo venderà fra qualche mese la stoffa fabbricata col cotone che compra oggi, sa che codesto prezzo sarà compreso fra un certo minimo ed un certo massimo, praticamente determinabili; e questa conoscenza lo assiste nei suoi calcoli di tornaconto. Ripetiamo, a scanso di equivoci, che in certi casi i limiti valgono soltanto in condizioni normali, mentre non valgono più in condizioni eccezionali. Ma la nostra esistenza è organizzata in gran parte sulla previsione del normale; nè può essere fondata sulla previsione dell'eccezionale senza danni spesso superiori ai vantaggi. Un istituto raccoglitore del risparmio tiene in cassa una somma sufficiente per rimborsare la massima quantità di depositi che *normalmente* si possano prevedere richiesti dai depositanti; se volesse far fronte alla massima richiesta possibile dovrebbe tener in cassa una somma uguale all'intero ammontare dei depositi raccolti, così che diverrebbe un semplice organo per la custodia del risparmio invece che un organo per l'impiego del capitale, e dovrebbe richiedere un compenso di custodia ai depositanti, invece di corrispondere loro un interesse. S'intende che, ove si manifesti un diffuso pànico, l'istituto può trovarsi nell'impossibilità di far fronte alle richieste di rimborso, ma questo danno eventuale del depositante è più che compensato dal vantaggio sicuro dell'interesse percepito.

Gli esempi si possono moltiplicare indefinitamente: il droghiere, fornito di scarso capitale, che deve rinnovare le sue provviste di settimana in settimana, le acquisterà in base ad una previsione minima sul bisogno dei clienti; il pubblico amministratore, che predispone la costruzione di edifici scolastici, dovrebbe attenersi ad una previsione massima sul numero dei futuri frequentatori. Se pur fosse possibile al droghiere conoscere esattamente il consumo dei suoi clienti nella prossima settimana, ed all'amministratore comunale sapere con assoluta precisione quanti ragazzi andranno a scuola in ciascuno dei prossimi cent'anni, l'uno e l'altro trarrebbero scarso giovamento pratico dalla più precisa conoscenza, come direttiva di azione.

6. — In altri casi, la previsione s'informa alla conoscenza della manifestazione più frequente del fenomeno (o della relazione di grandezza tra due fenomeni), ossia di quella che si presenta nella maggior parte dei casi. Ciascun uomo sano di trent'anni prevede e provvede, nella vita quotidiana, come se avesse avanti a sè almeno venti o trent'anni di vita. Eppure nessuno ignora che un certo numero di coloro che vivono a trent'anni non giungeranno a compiere i cinquanta od i sessanta; ma poichè la maggior parte vi giunge, tutti operano come se fossero sicuri di appartenere a questa schiera più fortunata. Il contadino che semina il frumento non si cura di sapere se tutti i granelli ch'egli sparge germoglieranno: gli basta poter prevedere che la maggior parte di essi recheranno frutto. L'impiegato il quale ha comprato coi suoi modesti risparmi un buono del Tesoro che potrà recargli un premio d'un milione di lire, continua a compiere con la solita assiduità il lavoro quotidiano, senza fare assegnamento sopra un patrimonio che gli potrebbe toccare ma che verosimilmente non gli toccherà. Ognuno di noi non fa testamento prima d'uscire di casa per i propri affari, sebbene sappiamo che ogni giorno qualche persona uscita di casa con simile intento non vi ha fatto ritorno, perchè travolta da un'au-

tomobile o da un tram; la grande maggioranza degli usciti rientrano, e noi contiamo di essere fra questa maggioranza.

Anche qui, come dianzi, non occorre molta fantasia per aggiungere nuovi esempi.

7. — In terzo luogo, la manifestazione di molti fenomeni collettivamente tipici (o la relazione di grandezza tra due fenomeni) varia solo lentamente nel tempo, come lentamente variano le condizioni dalle quali essa dipende. Così che i risultati dell'esperienza di un dato periodo valgono, in via approssimativa, per un periodo molto vicino.

Secondo la più o meno rapida variazione del fenomeno, l'approssimazione è più o meno grande. Potremmo prevedere con l'approssimazione del centesimo la proporzione dei maschi fra i nati dell'anno venturo, partendo da quella dei nati di quest'anno; riusciremmo forse ancora a calcolare con l'approssimazione di circa un decimo il traffico postale di un anno, partendo da quello del precedente. Commetteremmo invece errori gravissimi istituendo una consimile previsione sull'incremento del risparmio nazionale; il trovare un incremento di 49 milioni nel risparmio italiano durante l'anno 1893 non ci autorizza a prevedere un aumento di 44-54, o di 39-59 milioni, per l'anno successivo. In realtà si accerta, nel 1894, una diminuzione di 2 milioni; la quale a sua volta non dà norma neppure per l'immediato avvenire, perchè nel 1895 si riscontra un nuovo aumento di 97 milioni. Qui il fenomeno considerato non presenta la caratteristica della lenta variabilità nel tempo; ed è quindi impossibile la previsione approssimativa, fondata sull'ipotesi della costanza.

In tempi normali i vari cespiti dell'entrata e i vari capitoli della spesa di uno Stato si modificano lentamente attraverso il tempo: è quindi possibile sulla base del bilancio consuntivo di un esercizio finanziario formare il bilancio preventivo per l'esercizio seguente. Lo stesso principio, applicato in tempi di rapida modificazione delle entrate e delle spese può condurre a gravissimi errori nelle previsioni.

Molto spesso nel campo dei fenomeni fisici, meno spesso in quello dei fenomeni biologici, di rado in quello dei fenomeni sociali, la variazione è così lenta da permettere previsioni assai prossime al vero. Tali sono, per esempio — fuori del caso di cataclismi tellurici, guerre, epidemie, e in genere avvenimenti eccezionali — le previsioni delle imprese assicuratrici intorno ai rischi ai quali esse si propongono di far fronte.

La maggior parte di quelle uniformità che impropriamente vengono innalzate alla dignità di leggi statistiche rientrano appunto nel quadro della lenta variazione, cui abbiamo ora accennato. Variazione lenta a breve distanza di tempo, ma che può apparire rapida o rapidissima se si considerano più ampi intervalli. La quantità delle corrispondenze trasportate dalle poste italiane, pur variando poco da anno ad anno, varia molto nel corso di qualche lustro; il numero dei nati per 1000 abitanti in Italia, pur variando in generale soltanto di una frazione d'unità da anno ad anno, è sceso da 39 nel 1884 a 25 nel 1929. L'uno e l'altro esempio mostra come le piccole variazioni, sommandosi, finiscano col dar luogo a variazioni grandi.

Ma, perchè ciò avvenga, bisogna che le variazioni susseguentisi nel tempo siano tutte, o siano in prevalenza, dirette nello stesso senso. Altrimenti tenderebbero a compensarsi reciprocamente.

Siamo così condotti alla considerazione di un'altra uniformità: l'uniformità nella direzione delle successive variazioni di fenomeni collettivamente tipici.

8. — Prima di passare all'esame delle uniformità or ora accennate, dobbiamo aggiungere che la lenta variazione dei *fenomeni* collettivamente tipici si traduce molte volte in una approssimativa costanza della forma di distribuzione di *caratteri* collettivamente tipici. Se, per esempio, in un paese varia soltanto lentamente, attraverso il tempo, il numero delle persone di statura compresa fra cm. 163 e 164, tra i coscritti ventenni, e se una simile approssimativa costanza si ha per ogni altra statura, ne segue che la distribuzione dei coscritti in funzione delle stature si mantiene press'a poco costante nel tempo, o meglio varia solo lentamente da anno ad anno.

Notiamo anche qui che il sommarsi di tante piccole variazioni dirette nello stesso senso finisce col determinare differenze grandi fra lo stato iniziale e lo stato finale.

Così la distribuzione degli sposi secondo l'età, in un paese, suol variare pochissimo di anno in anno, e tuttavia si trova spesso radicalmente trasformata in capo ad un quarto di secolo od a mezzo secolo, perchè gradualmente diminuiscono le proporzioni di tutti i gruppi d'età giovanili e crescono quelle dei più maturi, o viceversa.

9. — Come abbiamo detto poc'anzi, un quarto genere di uniformità consiste in questo: che, anche dove non si trova costanza

nella manifestazione di fenomeni collettivamente tipici, spesso si osserva una costante tendenza ascendente o discendente nella misura di tale manifestazione (o nella misura della relazione di grandezza tra due fenomeni): in simili casi si suol parlare di « tendenza evolutoria » (o semplicemente « tendenza ») nella manifestazione del fenomeno e di « variazioni evolutorie » (talvolta dette anche « variazioni secolari ») nella misura di esso. Nella maggior parte dei casi la costanza è limitata a più o meno lunghi intervalli di tempo, e ascese e discese si alternano, cedendo talora il campo ad una approssimativa stazionarietà: così che la previsione del futuro proseguimento di una tendenza accertata in un dato periodo, se talvolta è confermata dai fatti, tal altra è invece contraddetta.

Chi, intorno al 1890, presi in esame i dati sul consumo medio annuo del grano per abitante, in Francia, negli ultimi settant'anni, avesse trovato che da decennio a decennio il consumo era aumentato di 20-30 litri per abitante, passando da 184 litri nel 1821-30 a 323 nel 1881-90, avrebbe potuto prevedere — sulla base della ipotesi di una continua tendenza all'aumento — un consumo medio di almeno 343 litri per abitante nel decennio 1891-1900 e di almeno 363 nel decennio 1901-10. Col protrarre la profezia per qualche altro decennio, balza agli occhi l'assurdità della ipotesi, in essa implicita, che il consumo individuale del grano possa aumentare indefinitamente; si giunge infatti ben presto a cifre superiori alla capacità digestiva del più poderoso divoratore. Se non altro per la ragione della limitata capacità di consumo individuale, l'ipotesi di un continuo aumento del consumo medio non regge. In fatto, troviamo nel decennio 1891-1900 un consumo medio individuale di soli 319 litri, e nel successivo di 304. La tendenza ascendente è dunque sostituita da una tendenza discendente. La quale neppure può servir di base ad attendibili previsioni, perchè già nel triennio 1911-13 ritroviamo indizi di una nuova ascesa (consumo medio litri 338), arrestata poi dalla guerra.

Il precedente esempio vale a mettere in avvertenza contro i rischi di un'affrettata estensione al futuro di tendenze accertate nel passato. Troppe volte nella vita pratica si prevede ascendente o discendente l'andamento di un fenomeno collettivamente tipico, senz'altra base logica che l'esperienza di un siffatto andamento nel più recente periodo d'osservazione. Base che non merita questo nome, perchè il solo fatto di essersi avverate ieri certe circostanze non assicura ch'esse siano per rinnovarsi domani. Eppure, specialmente nel campo dei rapporti economici, la cieca previsione di una continuità nelle tendenze degli eventi domina le azioni umane, ed è causa non ultima di quegli svariati fenomeni patologici che si raccolgono nell'unica denominazione di crisi economiche. È facile scorgere, dopo gli eventi, l'errore delle previsioni fatte col criterio ora accennato; ma nessuno può scagliare la prima pietra contro i rei di simili errori, perchè tutti, più o meno spesso, vi siamo incorsi e v'incorreremo.

10. — Un quinto genere di uniformità consiste nella ripetizione, ad intervalli di tempo più o meno ampi e più o meno regolari, di certe variazioni nella misura di manifestazione dei fenomeni (o nella misura della relazione di grandezza tra due fenomeni). Queste uniformità vengono spesso designate col nome di « movimenti ciclici » o di « variazioni cicliche » dei fenomeni: quando il ciclo si svolge entro un intervallo di tempo costante, si parla di « variazioni periodiche ». Sono movimenti ciclici dei fenomeni economici quelli che si traducono nell'alternarsi di periodi d'espansione e di depressione: movimenti che si manifestano in cicli di varia durata; sono movimenti periodici quelli che tendono a ripetersi col rinnovarsi delle stagioni o dei mesi, in ciascun anno.

Un esempio di variazioni periodiche si può desumere dai dati sul traffico delle Ferrovie dello Stato esposti qui sotto.

Quantità media giornaliera delle merci trasportate per conto di privati.
(tonnellate)

MESI	Esercizio 1906-07	Esercizio 1907-08	Esercizio 1908-09	Esercizio 1909-10	Esercizio 1910-11	Esercizio 1911-12
Luglio	75.954	81.157	84.230	90.010	91.125	98.859
Agosto	79.201	80.677	84.406	88.058	95.427	105.630
Settembre. . .	78.339	82.680	92.126	93.726	101.018	108.466
Ottobre . . .	77.617	82.972	96.674	91.566	99.848	103.544
Novembre. . .	72.251	82.132	87.084	93.948	91.625	97.699
Dicembre . . .	67.962	75.208	81.356	86.003	84.975	89.550
Gennaio . . .	65.725	79.984	75.964	82.481	83.074	90.101
Febbraio . . .	65.876	83.286	75.405	86.897	93.729	97.729
Marzo	73.435	82.210	80.992	90.576	98.498	99.501
Aprile	79.214	85.088	88.463	89.569	92.740	91.636
Maggio. . . .	78.826	80.861	88.714	85.244	96.015	100.596
Giugno	79.856	80.997	87.306	90.615	96.581	96.387
Luglio-Giugno .	74.561	81.414	85.269	89.044	93.702	98.309

Uno sguardo alle medie annuali, contenute nell'ultima riga della tabella, mostra che il traffico tende ad aumentare, nel tempo, in modo che in via di prima approssimazione si può ritenere uniforme. Si presenta, cioè, una variazione evolutoria, che dobbiamo eliminare se vogliamo mettere in risalto le variazioni periodiche. Una interpolazione lineare indica che l'incremento medio annuo è di 4.554 tonnellate: i valori medi annuali desunti dall'interpolazione differiscono poco da quelli accertati:

	1906-07	1907-08	1908-09	1909-10	1910-11	1911-12
Interpolazione . .	75.665	80.219	84.773	89.327	93.881	98.435
Osservazione . . .	74.561	81.414	85.269	89.044	93.702	98.389

Estendendo alle cifre mensili l'ipotesi dell'incremento uniforme, è ovvio che da mese a mese si dovrebbe accertare un aumento di 379 o 380 tonnellate, uguale ad un dodicesimo dell'aumento medio annuo: dovrebbe, cioè, la quantità media giornaliera delle merci trasportate salire gradualmente, secondo una progressione aritmetica, da 73.577 tonnellate nel luglio 1906, a 100.522 nel giugno 1912.

Confrontando le cifre così calcolate con quelle desunte dall'osservazione, si trova che in alcuni mesi dell'anno il traffico accertato è sempre (agosto, settembre, ottobre), o quasi sempre (luglio), superiore a quello calcolato; mentre in altri mesi è sempre (dicembre, gennaio), o quasi sempre (febbraio), inferiore. Meno concordanti risultati si hanno per gli altri mesi (anche perchè, eliminata la variazione evolutoria, le variazioni residue non sono per intero di natura periodica, come per semplicità abbiamo supposto, ma sono anche in parte « variazioni saltuarie », dovute cioè a circostanze agenti saltuariamente nel tempo). In media il carico giornaliero effettivo supera quello calcolato: dell'8 % in settembre, del 7 in ottobre, del 4 in agosto; mentre rimane ad esso inferiore: del 9 % in gennaio, del 7 in dicembre, del 4 in febbraio.

Ai mesi autunnali corrisponde una larga espansione del traffico, ai mesi invernali una forte contrazione. Queste caratteristiche si ripetono in ciascun anno, come si ripetono le circostanze ond'esse hanno origine (stagionalità dei raccolti, difficoltà delle comunicazioni invernali, ecc.); e ne tien conto l'amministrazione ferroviaria nel predisporre i servizi, per ottenere la massima sollecitudine dei trasporti e la più completa utilizzazione del materiale.

Variazioni cicliche e variazioni periodiche si riscontrano in moltissimi di quei fenomeni collettivamente tipici che danno luogo a larghe applicazioni dei metodi statistici: così nei fenomeni meteorologici, in quelli economici, in quelli demografici e in altri fenomeni sociali: per esempio nella criminalità. Si riscontrano periodicità stagionali, mensili, settimanali, giornaliere, orarie; s'incontrano anche fenomeni con periodi o con cicli più ampi del singolo anno.

La conoscenza dell'andamento ciclico o periodico dei fenomeni costituisce sempre un utile sussidio alle previsioni.

11. — Abbiamo fin qui enumerato ed esemplificato alcuni generi di uniformità, che più spesso s'incontrano nello studio dei fenomeni collettivamente tipici. Non pretendiamo che l'enumerazione sia completa; ma ci sembra di avere passato in rassegna i principali generi di uniformità, ai quali si possono facilmente ricongiungere o ricondurre altri tipi cui si sarebbe tentati di attribuire autonomia.

D'altronde, non ci preme tanto compilare un elenco scrupolosamente completo di tutti i generi d'uniformità che si riscontrano nei fenomeni collettivamente tipici, quanto mettere in evidenza il carattere contingente ed approssimativo di tali uniformità. Fuori delle rade occasioni nelle quali s'imbatte in veri e propri casi di stabilità statistica, l'indagatore non desume dalle indagini su fenomeni collettivamente tipici leggi rigorosamente e indefinitamente valide, ma soltanto uniformità approssimativamente valide entro ristretti limiti di tempo e di spazio. Non per ciò si deve stimar minore la dignità degli studi che mirano a quest'ultimo fine, nè si può contestare ad essi il carattere scientifico. L'opera del sociologo, che si sforza di risalire alle cause della criminalità, col che renderà poi più agevole all'uomo di Stato la ricerca di adeguati rimedi a questa piaga sociale, non è meno alta nè meno scientifica dell'opera dell'astronomo, che osserva e traduce in formole il corso degli astri. Le conclusioni dell'uno varranno per una generazione, le conclusioni dell'altro attraverseranno quasi immutate i secoli; ma queste e quelle del pari costituiscono un utile contributo all'inesauribile lavoro della scienza. Dall'aspetto pratico, poi, la conoscenza delle uniformità statistiche riesce veramente preziosa: si può dire, come abbiamo già ricordato, che sopra di essa sia fondata l'intera organizzazione della vita sociale.

Quesiti ed esercizi: 1. — Quali sono i caratteri delle « leggi statistiche »? In che differiscono dalle « leggi » dei fenomeni individualmente tipici?

2. — L'esempio esposto nei paragrafi 2 e 3 del testo mira a mostrare i veri caratteri delle «leggi» della mortalità; si espongano in forma riassuntiva tali caratteri.

3. — Quali sono i principali tipi di uniformità statistiche, che, pur non potendosi ritenere «leggi» nel significato più corretto della parola, possono servire come fondamento per la previsione?

4. — Si ricerchino nell'ASI esempi di fenomeni collettivamente tipici i quali, pur variando nel tempo o nello spazio, normalmente non escano da determinati limiti. Analogamente si ricerchino esempi di relazioni di grandezza tra due fenomeni, che normalmente rimangano entro determinati limiti: si ricerchino specialmente casi nei quali, l'un fenomeno precedendo cronologicamente l'altro, l'uniformità accertata possa essere utile nella previsione.

5. — Si traggano dall'esperienza quotidiana della vita sociale esempi di organizzazioni fondate sulla previsione che dati fenomeni collettivamente tipici abbiano a mantenersi, nelle loro variazioni, entro dati limiti. Si adducano esempi di simili previsioni, assunte come norme della condotta individuale.

6. — Si cerchino nell'ASI esempi di fenomeni collettivamente tipici i quali presentino una misura normale di manifestazione. Si esegua analoga ricerca di relazioni di grandezza tra due fenomeni che presentino una misura normale, tenendo conto dell'avvertenza esposta nell'esercizio 4.

7. — Si traggano dall'esperienza quotidiana esempi di organizzazioni sociali fondate sulla previsione che dati fenomeni collettivamente tipici abbiano ad oscillare intorno ad una data misura normale. Si adducano esempi di simili previsioni, assunte come norma della condotta individuale.

8. — Si cerchino nell'ASI esempi di fenomeni collettivamente tipici i quali presentino il carattere di variare molto lentamente nel tempo, almeno in condizioni normali. Si metta in evidenza come circostanze eccezionali possano fortemente alterare, e addirittura annullare, questa regolarità. Si compia analoga ricerca con riferimento alla relazione di grandezza tra due fenomeni.

9. — Si traggano dall'esperienza della vita sociale esempi di fenomeni che normalmente variano con grande lentezza e si mostri come la condotta sociale e quella individuale si fondino sulla previsione della permanenza di tale carattere. Si rilevi come impreviste brusche modificazioni dei fenomeni vengano a sconvolgere tanto più profondamente la vita sociale quanto più contrastano con la normale lenta variabilità dei fenomeni stessi.

10. — Si chiarisca come la lenta variazione dei fenomeni collettivamente tipici attraverso il tempo si traduca in una approssimativa costanza della forma di distribuzione di caratteri collettivamente tipici. Si cerchino nell'ASI esempi di tale costanza.

11. — Si distinguano, tra le variazioni dei fenomeni collettivamente tipici, le saltuarie dalle evolutorie e dalle cicliche. Quando le variazioni cicliche si dicono periodiche?

12. — Si cerchino nell'ASI esempi di variazioni saltuarie, evolutorie, cicliche, periodiche di fenomeni collettivamente tipici (e di relazioni di grandezza tra due fenomeni). Si traggano dalla conoscenza della vita sociale esempi di previsioni individuali o collettive fondate sulla previsione della continuità di tendenze evolutorie o della regolarità di andamenti ciclici o periodici.

13. — Si studino le variazioni negli ultimi cinquant'anni dei seguenti fenomeni per i quali sono date notizie nell'appendice retrospettiva dell'ASI : *a*) popolazione italiana (numero degli abitanti); *b*) natalità; *c*) mortalità ; *d*) nuzialità ; *e*) frequenza delle emigrazioni ; *f*) frequenza dei reati denunziati; *g*) produzione del frumento; *h*) produzione dell'olio d'oliva; *i*) produzione della pirite; *j*) produzione della ghisa; *k*) produzione dello zolfo; *l*) produzione della birra; *m*) importazione di caffè; *n*) importazione di cotone greggio; *o*) importazione di carbon fossile; *p*) importazione di frumento; *q*) esportazione di canapa greggia; *r*) esportazione di seta greggia; *s*) esportazione di paste di frumento; *t*) esportazione di uova di pollame; *u*) tonnellate di merce imbarcate e sbarcate nei porti; *v*) lunghezza delle ferrovie; *w*) lunghezza delle linee telegrafiche. Si ricerchi se i fenomeni considerati presentino qualche regolarità nelle loro variazioni, ed in ispecie si cerchi di determinarne la tendenza evolutoria, ricorrendo, ove sembri utile, all'interpolazione grafica.

14. — Esaminando serie di dati mensili riportate nell'ASI, si cerchi di separare le variazioni periodiche di certi fenomeni da quelle evolutorie, ed eventualmente da quelle cicliche e da quelle saltuarie.

INDICE ALFABETICO

(NB. — Nei richiami al testo e alla bibliografia è indicata soltanto la pagina; nei richiami agli esercizi è indicato anche, tra parentesi, il numero d'ordine dell'esercizio. Avvertasi che sono richiamati soltanto quegli esercizi che danno occasione a sviluppi delle nozioni esposte nel testo).

ERRATA-CORRIGE

Pag. 216, riga 8 dal basso: si legga « funzione di permanenza » in luogo di « funzione di eliminazione ».

Prezzo: LIRE QUARANTA