ISTITUTO CENTRALE DI STATISTICA
DEL REGNO D'ITALIA

SERIE VI - VOL. IV. 1929 - VII.

Annali di Statistica

C. GINI E L. GALVANI — Di una applicazione del metodo rappresentativo all'ultimo censimento italiano della popolazione (1° dicembre 1921)

ROMA
PROVVEDITORATO GENERALE DELLO STATO
LIBRERIA
1929 - ANNO VII

(6615) — Roma, 1929 - Anno VII — Stabilimento Poligrafico per l'Amministrazione dello Stato - G. C.

INDICE

I. — Il problema pratico.

1. Occasione all'indagine e suo scopo .... *Pag.* 1
2. Circoscrizioni costitutive del campione .... » 3
3. Criterio di designazione del campione .... » 4
4–5. Caratteri strumentali della scelta e procedura per la designazione del campione .... » 6
6. Confronto delle intensità medie dei caratteri e confronto delle quantità fondamentali nella totalità del Regno e nel campione » 8
7. Dipendenza lineare di un carattere da altri, o di una quantità fondamentale da altre .... » 10
8. Quantità fondamentali di verificazione .... » 12
9. Caratteri di verificazione .... » 13
10. Esame della concentrazione dei caratteri .... » 14
11. Esame della distribuzione dei caratteri .... » 16
12. Esame delle reciproche relazioni fra i caratteri .... » 17
13. Confronto mediante rette interpolatrici .... » 19

II. — Considerazioni generali sulla rappresentatività.

14. Critica della consueta definizione di rappresentatività .... *Pag.* 21
15. La rappresentatività è un concetto relativo e non assoluto ...... » 22
16. Come dovrebbe delinearsi una teoria generale della rappresentatività » 24
17. Problemi diretti ed inversi della rappresentatività .... » 25
18. Rappresentatività subordinate ad altre rappresentatività .... » 28
19. Subordinazione di alcuni aspetti di un carattere ad altri aspetti dello stesso carattere .... » 28
20. Rappresentatività subordinate dalla numerosità del campione ... » 29
21. Rappresentatività subordinate da relazioni lineari fra quantità fondamentali .... » 29

22–26. Rappresentatività subordinate da relazioni lineari fra caratteri:

a) conservazione della media aritmetica .... *Pag.* 32

b) scarti fra le medie dei caratteri nella totalità e nel campione » 36

c) uso dei caratteri di verificazione quando essi dipendano linearmente dai caratteri strumentali........................ » 37

d) condizioni per la conservazione dello scostamento quadratico medio .... » 38

e) condizioni per la conservazione della differenza media e per la conservazione del rapporto di concentrazione.......... » 39

APPENDICE A.

Probabilità che lo scarto percentuale fra il valore medio di un carattere quantitativo nella totalità, costituita da K circondarî, e in un campione, costituito da k circondarî ($k \leq K$) estratti a sorte, non superi ε (§§ 1–5).. *Pag.* 44

APPENDICE B.

Sul valore medio di caratteri diversi da quelli strumentali della scelta (§§ 1–5) .... *Pag.* 56

Tavole (1–36) .... *Pag.* 71

Di una applicazione del metodo rappresentativo all'ultimo censimento italiano della popolazione (1° dicembre 1921) (1).

I. — Il problema pratico.

Occasione all'indagine e suo scopo.

1. Nel novembre 1926, dovendosi rapidamente procedere, presso l'Istituto Centrale di Statistica del Regno d'Italia, allo sgombero dei locali in cui erano accumulati i fogli di famiglia del censimento della popolazione al 1° dicembre 1921, onde far posto ad altro materiale, il Presidente dell' Istituto ritenne opportuno derogare dalla consuetudine invalsa in passato di distruggere tutta la documentazione dei censimenti, una volta compiuta l'elaborazione e pubblicati i risultati; e giudicò miglior partito *conservare fra tali fogli quelli relativi ad alcune circoscrizioni*, al duplice scopo di potere:

a) ottenere un campione rappresentativo di tutto il Paese, rispetto alle sue principali caratteristiche demografiche, sociali, economiche e geografiche, nella considerazione dell'interesse che eventualmente si potrebbe avere a paragonare i censimenti passati con quelli futuri, anche relativamente a certi caratteri o a certi aspetti che precedentemente non erano stati giudicati meritevoli d'esame, ma di cui l'importanza potrebbe affermarsi in avvenire;

b) eseguire comparazioni locali fra le zone più particolarmente interessanti sotto taluni dei menzionati o di altri aspetti.

(1) L'indagine, di cui tratta la presente memoria, ha già dato occasione ad una comunicazione all'ultimo Congresso dell'Istituto Internazionale di Statistica, tenutosi al Cairo nel dicembre 1927 (C. Gini, *Une application de la méthode représentative aux matériaux du dernier recensement de la population italienne (1er décembre 1921)*, « Bulletin de l'Institut International de Statistique », Tome XXIII, 2ème livraison). Il lettore non dovrà quindi meravigliarsi di trovare nella presente memoria alcuni passi quasi testualmente riprodotti da quella comunicazione. Alcuni errori materiali, in parte tipografici, sfuggiti nella correzione delle bozze della detta comunicazione, sono stati corretti in questa memoria.

Notiamo subito che, trovandosi la totalità dei fogli già distribuita per provincie, circondarî e comuni, si poteva con tutta facilità reperire il gruppo parziale inerente a ciascuna di queste unità amministrative; cosicchè il criterio di costituire il campione mediante i gruppi di fogli di talune tra queste unità, oltre ad essere il solo che potesse consentire delle comparazioni locali, assicurava nel tempo stesso la necessaria rapidità delle operazioni.

Le indicazioni principali contenute in quei fogli, uno per ciascuna famiglia o convivenza, erano le seguenti: provincia; circondario; mandamento; comune; parrocchia; frazione; nome della via o della piazza; numero della casa; località dell'abitazione (se in centro abitato o in isolata campagna); situazione dell'abitazione (se al piano sotterraneo, terreno, 1., 2.,... piano, in soffitta); uso dell'abitazione (se esclusivo ad una famiglia oppure in comune con altre); numero delle stanze; cognome, nome, paternità, relazione di parentela o di convivenza col capo famiglia; sesso; data e luogo di nascita; stato civile; saper leggere o meno; occupazione o professione principale od accessoria; condizione (per i censiti da 10 anni in su); imposte fondiarie eventualmente pagate; eventuale cittadinanza straniera di qualche membro della famiglia, con distinzione fra i presenti nell'abitazione alla data del censimento e quelli temporaneamente assenti o dimoranti all'estero; luogo degli assenti eventuali; eventuale prestazione di servizio militare, per obblighi di leva, degli assenti.

Il campione cercato doveva, fra i requisiti esteriori, aver quello di comprendere intorno al 15 % di tutto il materiale, e ciò sia per commisurarsi alla capacità dell'ambiente in cui avrebbe dovuto essere conservato, sia per non essere nè così ristretto da fornire risultati poco attendibili, nè troppio ampio da riuscire difficilmente indagabile. D'altronde, concordemente allo scopo dell'indagine, il suo primo requisito intrinseco doveva essere che *i valori di alcuni fra i più importanti caratteri statistici risultassero praticamente uguali nel campione e nella totalità.*

Per l'apprezzamento sommario di una tale concordanza si stabilì di riguardare il campione come *ottimo*, o *soddisfacente*, o *sufficiente* o *insufficiente*, dal punto di vista della sua rappresentatività riguardo a un certo carattere, secondo che la differenza fra il valore del carattere stesso nel campione e nella popolazione totale risultasse non superiore all'1,5 %, o compresa fra l'1,5 % e il 5 %, o compresa fra il 5 % e il 10 %, o infine superiore al 10 %.

Fu poi necessario scartare dall'indagine le nuove Provincie, non conoscendosene alcuni dati statistici e in particolare i valori dei due caratteri *altitudine sul livello del mare* e *reddito medio*, ai quali si era stabilito di ricorrere per sperimentare la rappresentatività del campione formato; cosicchè il materiale da cui tale campione doveva essere estratto rimase costituito dagli 8.347.995 fogli relativi alle 69 Provincie, o 214 Circondari od 8.354 Comuni in cui, alla data del censimento, era amministrativamente diviso il territorio del Regno, escluse le Provincie della Venezia Tridentina, della Venezia Giulia e di Zara.

Circoscrizioni costitutive del campione.

2. La condizione fissata che il campione dovesse prestarsi anche a comparazioni locali, imponeva dunque di costituirlo sulla base di convenienti circoscrizioni fra quelle che erano già servite nella classificazione e nell'elaborazione del materiale di censimento. Tali circoscrizioni, o puramente territoriali, o amministrative, o giudiziarie, o ecclesiastiche, erano: la provincia, il circondario, il mandamento, il comune, la parrocchia, la frazione, la sezione di censimento.

Per varie ragioni si esclusero senz'altro le divisioni per mandamento, parrocchia, frazione, sezione di censimento. Apparve anche immediatamente l'impossibilità di assumere come base della scelta le provincie, sia per la loro vastità territoriale, sia per l'esiguità del loro numero. D'altra parte gli 8.354 comuni costituivano un insieme troppo numeroso per il caso che la designazione del campione avesse dovuto essere affidata non ad un sorteggio, ma ad una scelta giudiziosa (così come effettivamente si è trovato conveniente di fare), specialmente quando, al fine di migliorare il campione, si fossero dovute eseguire successive sostituzioni di comuni, per tener conto non di uno solo, ma di diversi caratteri statistici; senza dire che le frequenti variazioni territoriali, a cui i comuni stessi vanno soggetti da qualche anno a questa parte avrebbero reso assai difficili i confronti con l'avvenire, frustrando lo scopo che doveva essere raggiunto.

Non rimaneva dunque se non assumere, come elementi costitutivi del campione, i circondarî. È vero che posteriormente all'inizio dell'indagine, il R. decreto legge 2 gennaio 1927 n. 1 sopprimeva i circondarî stessi come unità amministrative; ma l'Istituto Centrale di Statistica, nella previsione di eventuali comparazioni locali, riconosciuta l'opportunità di basarsi su circoscrizioni meno vaste delle provincie, e l'impossibilità di riferirsi ai comuni, perchè troppo instabili, aveva già da tempo determinato di assumere come unità statistiche i circondarî, quali erano costituiti al 31 dicembre 1921. Questi infatti, in occasione di indagini future, si sarebbero idealmente ricostruiti mediante i comuni che ne facevano parte a quella data, tenendo conto delle variazioni territoriali eventualmente intervenute nel frattempo nei comuni perimetrali, potendosi evidentemente prescindere dalle variazioni fra loro compensative dei comuni interni. D'altronde, all'infuori delle circostanze esclusive che obbligavano a ricorrere ai circondarî, appariva a priori che il loro ordine di grandezza avrebbe consentito non soltanto confronti fra unità territoriali di diversa struttura fisica e di diversa posizione geografica, ma anche indagini in seno alle unità stesse; e ciò avrebbe accresciuto il valore di consultazione del materiale accantonato.

Criterio di designazione del campione.

3. Bisognava ora decidere se il campione dovesse designarsi a sorte o per giudiziosa scelta; e poichè si era già stabilito di riguardarlo come ottimo, ossia praticamente concordante con la totalità rispetto ad un certo carattere, se la differenza fra i valori di questo nel campione e nella popolazione del Regno risultasse non superiore all'1,5 %, così si trattava di calcolare la probabilità affinchè un campione di k unità amministrative, prese a sorte fra le K della totalità, avesse ottemperato a quella condizione: e si sarebbe così avuta una base plausibile per determinare se da una scelta giudiziosa si sarebbe potuto attendere un risultato migliore.

Non sembra che il problema enunciato sia stato prima d'ora considerato; infatti gli Autori che si sono occupati della questione generale della rappresentatività, ed in particolare il Bowley (1), che, più estesamente degli altri, l'ha sottoposta ad analisi matematica, hanno considerato il caso che l'estrazione a sorte sia fatta per unità statistiche semplici

(1) *Measurement of the Precision attained in Sampling*, in appendice al « Bulletin de l'Institut International de Statistique », Tome XXII, Ière livraison. V. anche, nello stesso fascicolo, un'ampia relazione dello Jensen sul metodo rappresentativo: *Report on the Representative Method in Statistics; The Representative Method in practice.*

(nel nostro caso *individui*) e non per unità complesse (nel nostro caso *unità amministrative o territoriali*, costituite dai circondarî) (1).

La differenza fra i problemi che si presentano nell'uno e nell'altro caso risulta evidente dal paragone dei due seguenti, che si enunciano come esempio:

a) Supposto che nel Regno esistano N redditieri, e che il loro reddito medio sia A, qual'è la probabilità che estratti a sorte n redditieri, il loro reddito medio a sia affetto da un errore non maggiore di ε rispetto ad A?

b) Supposto che gli N redditieri siano distribuiti nei K circondarî del Regno secondo i numeri $p_1, p_2, \ldots p_K$, qual'è la probabilità che, estratti a sorte k di questi circondarî, il reddito medio a dei redditieri che vi appartengono sia affetto da un errore non maggiore di ε rispetto al reddito medio A constatato nel Regno?

Basta riflettere al fatto che, pur restando fisso il numero k, potrà, a seconda del sorteggio, variare il numero dei redditieri estratti dalla totalità, per vedere subito che il problema *b*) è ben distinto da *a*), e che non si potrebbe neanche far rientrare negli schemi teorici recentemente trattati dal MARCH (2) e che in certo modo si connettono a quella che il BOWLEY dice « stratified selection ».

Nell'Appendice A è trattato il problema generale del tipo *b*), nella supposizione, plausibile in prima approssimazione, che fra il numero medio degli individui che in ciascuno dei possibili gruppi di k circoscrizioni possiedono un certo carattere e il valore medio dell'ammontare globale del carattere nelle singole circoscrizioni degli stessi gruppi interceda una correlazione normale; e del risultato ottenuto è, nella stessa Appendice, fatta applicazione al calcolo della probabilità Π che in un campione formato estraendo a sorte $k = 29$ circondarî dai $K = 214$ del Regno, il red-

(1) Poichè i caratteri statistici solitamente considerati in una popolazione sono, dal punto di vista aritmetico, riferibili ai singoli individui (quali ad esempio, i coefficienti di natalità, di mortalità e di nuzialità, esprimibili per testa invece che per 1000 secondo la consuetudine, il reddito medio dei redditieri, le proporzioni o percentuali della popolazione agricola o della popolazione agglomerata rispetto alla totalità ecc.) così, assumendo come unità del sorteggio i singoli fogli di famiglia, che riguardano, ciascuno, un esiguo numero di persone (in media 4,45 persone), ci si sarebbe praticamente avvicinati alla condizione necessaria per potere applicare i criteri dati dal BOWLEY nel Rapporto testè citato, cioè alla condizione di eseguire l'estrazione a sorte per singole unità statistiche. Ma ciò avrebbe importato la rinuncia alla possibilità di comparazioni territoriali.

(2) *L'analyse de la variabilité* in « Metron », vol. VI, n. 2.

dito medio a abbia rispetto al reddito medio A constatato nel Regno uno scarto ε non maggiore dell'1,5 %; si trova $\Pi = 0{,}08$ circa. Altre consimili applicazioni ad altri caratteri vennero risparmiate, poichè già bastò quel valore trovato per Π a indicare con la sua piccolezza che una scelta sistematica avrebbe potuto fornire un campione assai meglio rappresentativo della totalità di quanto sarebbe stato da attendersi mediante un sorteggio. A ciò si aggiunga che la considerazione simultanea, oltre che del reddito medio, di altri caratteri, sia pure correlati con questo, avrebbe ulteriormente diminuito il valore di quella probabilità (1).

Conclusione di tutta questa indagine preliminare fu dunque che il campione dovesse costituirsi mediante opportuna scelta di un adeguato numero di circondarî (da 25 a 30).

Caratteri strumentali della scelta e procedura per la designazione del campione.

4. Affinchè il campione risultasse rappresentativo non solo per i più salienti caratteri demografici, ma anche per alcuni caratteri sociali, economici e geografici, si stabilì di operare la scelta tenendo in vista i seguenti, che si potranno perciò dire « caratteri strumentali » della scelta stessa:

1) la *natalità*,

2) la *mortalità*,

3) la *nuzialità*, come numero dei nati e rispettivamente dei morti e dei matrimonî per 1000 abitanti;

4) la *percentuale della popolazione agricola maschile* nell'insieme dei maschi di età superiore ai 10 anni;

5) la *percentuale della popolazione agglomerata ;*

(1) È qui opportuno notare, considerando in sè il criterio dell'estrazione a sorte e prescindendo per un momento dalle ragioni che hanno impedito di assumere come unità i comuni, che il sorteggio di un conveniente quantitativo fra gli 8354 comuni avrebbe indubbiamente dato risultati migliori che il sorteggio per circondarî, e che, presumibilmente, il miglioramento sarebbe stato anche più notevole estraendo il campione col metodo di « stratificazione », il quale consisterebbe nel distribuire la totalità dei comuni in tanti gruppi di comuni « simili », e nell'estrarne a sorte, in eguale proporzione, un certo numero di ogni gruppo; ma non è chi non veda quali ostacoli si incontrerebbero nell'eseguire tale « stratificazione » simultaneamente rispetto a diversi caratteri. Per quanto concerne il metodo di « stratificazione » si vedano i già citati Rapporti del Bowley e dello Jensen.

6) il *reddito medio* accertato ai contribuenti dell'imposta di Ricchezza Mobile delle categorie *B* e *C* (1);

7) l'*altitudine media*, considerando come altitudine media sul livello del mare di ciascun circondario quella del suo capoluogo.

Nella Tav. 1 è dato il valore di tali medie e percentuali per ciascuno dei 214 circondarî, disposti per ordine di natalità decrescente. Essa contiene inoltre la densità della popolazione per chilometro quadrato, l'accrescimento naturale, la popolazione e l'area di ogni circondario. Nelle Tavole da 2 a 8 gli stessi circondarî sono distribuiti per gradi equidistanti di ciascuno dei sette caratteri strumentali della scelta. Infine la Tav. 9 fornisce la distribuzione dei circondarî rispetto alle popolazioni assolute.

Si rileva a prima vista che quasi tutte queste distribuzioni danno luogo a notevole dispersione, e ciò, tenuto anche conto del piccolo numero di circondarî, conferma la previsione che una estrazione a sorte per circondarî non avrebbe potuto fornire un campione sufficientemente rappresentativo della totalità.

5. Per la scelta effettiva dei circondarî si usò il criterio di considerare anzitutto il carattere statistico di importanza prevalente, e successivamente gli altri meno essenziali; e poichè si ritenne che il fenomeno più saliente fosse quello della natalità, sia perchè su tutti il più essenziale per lo sviluppo demografico della Nazione, sia per l'interesse sempre maggiore che esso suscita nel campo scientifico e in quello politico, così si tentò, in prima approssimazione, di costituire il campione desiderato, tenendo presente il solo valore medio della natalità, salvo a modificare il risultato della scelta provvisoria con la considerazione dei rimanenti caratteri.

Si assunsero come denominatori dei quozienti di natalità le popolazioni presenti censite al 1 dicembre 1921 nei singoli circondarî, e come numeratori (allo scopo di conseguire una conveniente perequazione) la semisomma dei nati vivi nel 1921 e nel 1922; si disposero i 214 circondarî per ordine di natalità decrescente dal massimo di 44,91 ‰ (Circondario di Adria) al minimo di 15,58 ‰ (Circondario di Mortara); si iscrissero successivamente alla natalità il quoziente di nuzialità (dedotto dalla semisomma dei matrimoni nel 1921 e nel 1922), il quoziente di mortalità (dedotto ana-

(1) La tassa di Ricchezza Mobile categoria *B* colpisce i redditi industriali e commerciali; quella di categoria *C* colpisce i redditi vitalizi e quelli temporanei dipendenti dall'opera dell'uomo, senza intervento di capitali (stipendi degli impiegati di aziende private, guadagni professionali, ecc.).

logamente) ed i coefficienti o medie concernenti gli altri caratteri (Tav. 1), in modo che fu possibile apprezzare succintamente se e come gli ultimi sei caratteri considerati si cograduassero a quello fondamentale della natalità.

Ciò fatto, s'incominciò con lo scegliere in quell'elenco, ordinato dunque per valori decrescenti della natalità, i circondarî di cui il numero del posto terminava con la cifra 1, cioè quelli occupanti il 1°, l'11°, il 21°... posto. Ma poichè era desiderabile che nel campione tutte le regioni d'Italia fossero rappresentate, e che esso comprendesse delle grandi e delle medie città le quali, per la loro ricchezza, per il loro carattere industriale, per la loro altitudine sul livello del mare, ecc, potessero considerarsi come sufficientemente rappresentative dell'insieme delle grandi e rispettivamente delle medie città italiane, così a quel primo gruppo di circondarî convenne aggiungerne qualche altro, pervenendo a un campione provvisorio di 27 circondarî che, rispetto alla natalità, era ottimamente rappresentativo.

Senonchè, come era da attendersi, tale campione, la cui ampiezza era conforme al criterio di quantità già fissato in principio, risultò ben lontano dall'adempiere a quel ragionevole minimum di condizioni che potevano, da un primo esame, esigersi in una generica rappresentazione della totalità, per quei caratteri che non si erano tenuti presenti nella cernita fatta.

Fu dunque necessario modificare quel primo gruppo di circondarî, sia nel numero, sia nella composizione; e lo si fece tenendo simultaneamente sottocchio le scale numeriche di tutti i caratteri, facendo seguire ad ogni aggiunta o sostituzione una prova sommaria per vedere fino a qual punto la media del gruppo parziale, per ciascun carattere, si avvicinasse alla media generale dello stesso carattere, e mettendo in opera tutti quegli accorgimenti che l'esperienza suggeriva e che potevano in qualche modo circoscrivere il campo di scelta di ciascun nuovo circondario.

La scelta definitiva cadde sul gruppo di 29 circondarî che nella Tav. 1 vengono indicati in carattere grassetto.

Confronto delle intensità medie dei caratteri e confronto delle quantità fondamentali nella totalità del Regno e nel campione.

6. Vediamo ora se e come le intensità medie dei sette caratteri tenuti in vista nell'operare la scelta si adeguino alle medie corrispondenti per la totalità del Regno. La Tav. 10 contiene appunto i valori di tali medie

e gli scarti assoluti e relativi fra le medie corrispondenti (1), e contiene altresì gli analoghi valori per altri due caratteri, la *densità della popolazione* e l'*accrescimento naturale* di cui sarà fatta parola più avanti. Essa dimostra che, essendo lo scarto relativo: inferiore all'1,5 % per la natalità, la mortalità, la nuzialità, e l'altitudine media; compreso fra l'1,5 % ed il 5 % per il reddito medio; e compreso fra il 5 % e il 10 % per le percentuali della popolazione agricola e della popolazione agglomerata, la scelta eseguita è ottima per i primi quattro caratteri ora nominati, è soddisfacente per il quinto, ed è sufficiente per gli ultimi due. Fra i caratteri strumentali della scelta non ve n'è dunque alcuno rispetto al quale si debba dire che la rappresentatività è insufficiente. Inoltre, poichè in tre casi sopra sette gli scarti sono negativi e nei rimanenti positivi, non si rivela nessuna tendenza dello scarto ad avere un segno piuttosto che l'altro.

Per concludere, la scelta eseguita ottempera in modo soddisfacente alla condizione della pratica concordanza fra i valori medî dei sette caratteri considerati, nel gruppo parziale e nella totalità.

Al confronto ora eseguito si può dare un'altra forma, mettendo a riscontro le quantità fondamentali da cui si deducono i valori medî della Tav. 10.

La Tav. 11 presenta appunto, per la totalità dei circondarî e pel complesso scelto, la popolazione presente censita al 1° dicembre 1921, la semisomma dei nati vivi, quella dei morti e quella dei matrimoni nel 1921 e nel 1922, la popolazione agglomerata, la popolazione maschile in età superiore ai 10 anni e quella agricola maschile oltre la stessa età, il reddito complessivo (di Ricchezza Mobile, categorie B e C), il numero dei redditieri, la somma dei prodotti delle popolazioni dei circondarî per le altitudini dei rispettivi capoluoghi. E da essa si constata che i rapporti di

---

(1) Si ottiene lo scarto relativo fra la media generale A e la media a del campione dividendo lo scarto assoluto $a - A$ per la media generale A, cioè eseguendo $\frac{a - A}{A}$. Nondimeno, quando si tratti di un carattere la cui intensità non può superare un certo limite L, è necessario introdurre un coefficiente di correzione; in questo caso è consigliabile la formula $\frac{a - A}{A \frac{L - A}{L}}$, adottata nella presente ricerca. Vedi a tale proposito, C. Gini: *Sul massimo degli indici di variabilità assoluta e sulle sue applicazioni agli indici di variabilità relativa ed al rapporto di concentrazione*, in « Metron » Vol. VIII, di prossima pubblicazione.

tutte queste quantità, fra il gruppo parziale ed il totale, si aggirano intorno al valore costante del 15 %.

È dunque lecito affermare che la scelta è stata eseguita in modo da verificare, praticamente, il criterio di una costante proporzionalità fra le quantità fondamentali, nel passare dalla totalità al campione, cioè, anche per questo riguardo, la scelta stessa è riuscita in modo del tutto soddisfacente.

Infine, per quanto concerne la distribuzione geografica, allo scopo di giudicare fino a qual punto sia stato possibile ottenere che fra i circondarî scelti risultassero equamente rappresentate tutte le regioni del Regno, si potrà osservare il cartogramma rappresentativo di tali circondarî (Tavola 12). Di qui si rileva che la zona litoranea tirrena è più abbondantemente rappresentata di quella adriatica; che la zona centro-settentrionale del Piemonte e quella centro-occidentale della Sicilia non forniscono alcun contigente al campione. In compenso però risultano convenientemente rappresentate la regione alpina e prealpina e quella appenninica, e, d'altra parte, i circondarî segnati coi numeri 12, 13 e 14 come anche quelli dei numeri 21, 22, 23 e 24 costituiscono, per così dire, due anelli, nell'Italia Centrale e nella Meridionale, lungo i quali si può, con una certa continuità, seguire l'andamento dei diversi fenomeni, nel passaggio dalle zone costiere a quelle interne. Pertanto, all'infuori di qualche lieve e inevitabile manchevolezza, si può dire che il gruppo costituito risponda sufficientemente al requisito di rappresentare tutte le regioni e, nella misura del possibile, le varie condizioni di ambiente.

Dipendenza lineare di un carattere da altri, o di una quantità fondamentale da altre.

7. Conviene ora esaminare come si comporti il nostro campione rispetto ad alcuni altri caratteri non tenuti presenti nell'operare la scelta, o rispetto a quantità fondamentali diverse da quelle che intervengono a definire i caratteri stessi. Questi nuovi caratteri e queste nuove quantità fondamentali diremo « di verificazione » per distinguerli dai caratteri tenuti presenti per operare la scelta, ossia « strumentali della scelta ». Ma, a tale esame, premettiamo alcune osservazioni, per la cui facile dimostrazione rimandiamo alla seconda parte di questo lavoro.

α) Se nella totalità e in un campione formato di alcuni circondarî si considerano diverse « quantità fondamentali » M, N, P, Q.... (per esempio area, popolazione, numero dei redditieri, ecc.) delle quali una, M, si possa riguardare come funzione lineare omogenea delle altre N, P, Q...., e se i valori globali di queste ultime nella totalità e cioè ΣN_i, ΣP_i, ΣQ_i,... sono proporzionali ai valori globali delle stesse nel campione, vale a dire, alle $\Sigma' N_i$, $\Sigma' P_i$, $\Sigma' Q_i$,... allora anche i valori globali di M nella totalità e nel campione staranno nello stesso rapporto; cioè se

$$\Sigma N_i / \Sigma' N_i = \Sigma P_i / \Sigma' P_i = \Sigma Q_i / \Sigma' Q_i = k, \text{ anche } \Sigma M_i / \Sigma' M_i = k.$$

Se la relazione fra M ed N, P, Q.... è soltanto lineare in senso statistico, cioè sia turbata da errori accidentali, cosicchè si debba parlare di relazione a « tendenza lineare » piuttosto che di relazione lineare nel preciso senso matematico della parola, quella conclusione sarà applicabile soltanto se il numero dei circondarî del campione sia abbastanza ampio perchè l'azione perturbatrice del caso si possa ritenere praticamente eliminata. Alla stessa conclusione si perviene, se il campione è abbastanza numeroso, qualora M sia indipendente da N, P, Q....; mentre essa non vale se si può ancora parlare di una « relazione » fra M ed N, P, Q...., ma questa non sia lineare.

β) Se nella totalità e in un campione formato da alcuni circondarî si considerano diversi « caratteri » T, X, Y,... i quali siano definiti come « rapporti di quantità fondamentali » (per esempio: natalità, mortalità, reddito medio, ecc.), ed uno di essi T sia funzione lineare, omogenea o no, degli altri X, Y...., allora se il campione è conservativo rispetto alla totalità del valore medio dei caratteri X, Y...., esso sarà pure conservativo del valore medio di T.

Deve anche qui osservarsi che nel caso di linearità in senso statistico e non in senso strettamente matematico, quella conclusione è legittima soltanto se il campione sia sufficientemente numeroso.

La conclusione stessa vale se T è indipendente da X, Y..., sempre nella supposizione di una sufficiente ampiezza del campione; non vale, invece, se tra T ed X, Y... intercede una relazione non lineare.

γ) Se, come accade quasi sempre in pratica, le medie dei caratteri X, Y..... non coincidono perfettamente nel campione e nella totalità, ma danno luogo a certi scarti relativi, pure ammettendo che tali scarti siano trascurabili nei limiti della precisione che si vuole raggiungere, potrà avvenire che lo scarto relativo di T non sia trascurabile, anche nel caso in cui T sia funzione lineare di X, Y....

Quantità fondamentali di verificazione.

8. Ciò posto, incominciamo coll'esaminare il nostro campione rispetto all'*area* e al *numero dei fogli di famiglia*, che sono due quantità fondamentali le quali non intervengono a definire i 7 caratteri tenuti presenti nella scelta (vedi Tav. 11), mentre nella definizione di questi concorrono: la popolazione, le semisomme dei nati, dei morti e dei matrimonî nel 1921 e nel 1922, la popolazione agglomerata, la popolazione maschile in età superiore ai 10 anni, la popolazione agricola maschile, il reddito complessivo, il numero dei redditieri.

Poichè si può ammettere che fra i numeri di individui di gruppi sufficientemente ampî di popolazione (quali sono le popolazioni dei circondarî) e i numeri dei rispettivi fogli di famiglia, impiegati pel censimento, vi sia una relazione di proporzionalità (relazione lineare omogenea) (cfr. nota (1) a pag. 5 ed anche vedi n. 21), per quanto si è detto ad α del n. 7, sarà da attendersi che il numero di tali fogli nel gruppo scelto stia al numero totale nello stesso rapporto delle popolazioni: e ciò è appunto confermato dai dati al n. 12 della Tavola 11.

Invece, per la variabilissima conformazione ambientale dei circondarî non si può pensare che vi sia proporzionalità fra la loro area e la loro popolazione. Non sembra neanche ragionevole presumere che l'area sia una funzione lineare omogenea di diverse quantità fondamentali, fra quelle per le quali abbiamo giá constatato la pratica uguaglianza dei rapporti, nel passare della totalità al campione; tuttavia non si può negare l'esistenza di una relazione fra l'area e la popolazione (tanto è vero che il coefficiente di correlazione fra le aree e le popolazioni della totalità dei circondari raggiunge il rilevante valore di 0.340), e similmente fra l'area e il numero dei nati ecc. Ma, trattandosi di relazioni che non sono lineari omogenee, nè in senso matematico, nè in senso statistico, vi potrà essere un notevole divario fra il rapporto dell'area complessiva dei circondarî scelti all'area totale, ed il rapporto analogo delle popolazioni (e delle altre quantità fondamentali enumerate); e ciò effettivamente risulta dal n. 11 della Tavola 11.

Caratteri di verificazione.

9. Venendo all'altra forma di esame del campione, vediamo come si presentino i valori medî dell'intensità di alcuni caratteri diversi da quei sette che s'impiegarono come strumento di scelta, e a tal fine assumiamo l'*accrescimento naturale* e la *densità della popolazione* per chilometro quadrato. Ritorniamo alla Tav. 10 in cui è anche effettuato, come si era già detto, il confronto delle medie di tali caratteri.

Per quanto concerne l'accrescimento naturale, notiamo che esso è, per definizione, la differenza fra la natalità e la mortalità, ossia è una funzione lineare di questi due caratteri. Pertanto, se il campione conservasse esattamente i valori medî della natalità e della mortalità, per l'osservazione del n. 7, β), esso conserverebbe anche esattamente il valore medio dell'accrescimento naturale. Ma, poichè (vedi numeri 1 e 2 della stessa Tavola) il campione è conservativo dei valori medî di quei due primi caratteri soltanto in via approssimata (per quanto l'approssimazione sia tale da far riguardare il campione stesso come ottimo nel rappresentare la loro intensità media, a norma della convenzione posta), così per l'osservazione γ) del n. 7, non si potrà, senz'altro, concludere che il campione sia anche ottimo rispetto all'intensità media dell'accrescimento naturale. Ed effettivamente si vede che l'accordo fra le due medie, essendo il loro scarto relativo dell'1,99 %, è non già ottimo, ma soltanto soddisfacente.

Ci si rende ragione di tale apparente anomalia notando che, mentre lo scarto assoluto per l'accrescimento naturale T dipende dagli scarti assoluti per la natalità X e per la mortalità Y mediante la relazione

$$a - A = (b - B) - (c - C),$$

in cui A, B, C ed a, b, c sono le medie di T, X, Y nella totalità e nel campione, lo scarto relativo di T è collegato a quelli di X ed Y per mezzo della uguaglianza (cfr. n. 23):

$$\frac{a - A}{A} = \frac{b - B}{B} \cdot \frac{B}{B - C} - \frac{c - C}{C} \cdot \frac{C}{B - C},$$

in cui il secondo membro dipende dalla differenza $B - C$; cosicchè, se questa è abbastanza piccola, lo scarto $(a - A)/A$ può superare qualunque numero prefissato, pure essendo trascurabili gli scarti $(b - B)/B$ e $(c - C)/C$.

Se, in via d'ipotesi, essendo stata la natalità del 30,23 ‰ nel Regno e del 30.54 ‰ nel campione, la mortalità fosse stata del 30,23 ‰ nel Regno e nel campione, questo sarebbe stato ottimo nel rappresentare la natalità, assolutamente perfetto per la mortalità e del tutto insufficiente per l'accrescimento naturale, di cui il valore sarebbe stato uguale a 0 per il Regno ed uguale a 0,31 per il campione, con uno scarto relativo infinito.

Quanto all'altro carattere di verificazione, cioè densità della popolazione, esso è evidentemente in relazione con parecchi caratteri strumentali della scelta, per esempio con l'altitudine sul livello del mare, con la percentuale della popolazione rurale, con quella della popolazione agglomerata, con la natalità e con la mortalità; ma tali relazioni non sono lineari, e neanche si può affermare che siano a tendenza lineare. Non v'è, dunque, alcuna ragione per ritenere che un campione conservativo delle medie di questi ultimi caratteri, lo sia anche per la media della densità della popolazione. Ed effettivamente lo scarto relativo per la densità supera l'11 %, cioè il campione non è affatto rappresentativo per quanto riguarda il valore medio di questo carattere (1).

Esame della concentrazione dei caratteri.

10. Poichè il solo criterio in base al quale usualmente si giudica se un campione sia rappresentativo della totalità è quello di osservare se si conservino i valori medî dell'intensità dei caratteri ed i rapporti fra il numero delle unità aventi un certo attributo ed il numero totale delle unità, così, da questo punto di vista, potrebbe ritenersi esaurito il confronto fra il nostro gruppo di 29 circondarî e l'insieme complessivo dei 214 circondarî, da cui era costituito il Regno d'Italia, escluse le nuove Provincie, al 1º dicembre 1921.

Tuttavia, volendo eseguire un'indagine comparativa più stringente della consueta, ci proponiamo di prendere in esame altri aspetti quantita-

---

(1) La circostanza dell'avere paragonato il campione alla totalità sia sui dati della Tav. 10 (valori medî di alcuni caratteri quantitativi) sia su quelli della Tav. 11 (quantità fondamentali che intervengono a definire quei caratteri) ci induce ad osservare che i due confronti sono logicamente equivalenti, ma che, peraltro, quello eseguito mediate la Tav. 10 rende meglio evidenti le diversità fra l'insieme parziale e il totale; esso è anche più opportuno in quanto mette a riscontro dei rapporti o percentuali costituenti dei « caratteri » delle diverse unità statistiche, mentre come tali non possono considerarsi le « quantità fondamentali » della Tav. 11, prese ciascuna a sè. Il forte errore (11,46 %) da cui risulta affetta la densità della popolazione nel campione, rispetto a quella del Regno, pure essendo implicito nei dati della Tav. 11, non poteva avere in questa immediato risalto. Potremo dunque ritenere che i risultati di tutti i confronti fin qui eseguiti fra totalità e campione siano quelli che appaiono dalla Tav. 10 e dal cartogramma della Tav. 12.

tivi dei caratteri, non meno importanti dei valori medî, quali sono la variabilità o concentrazione e la distribuzione delle quantità. Non solo: ma l'indagine si potrà ulteriormente estendere, considerando, oltre gli accennati aspetti quantitativi che riguardano i diversi caratteri presi singolarmente, anche le mutue relazioni fra questi; cosicchè si farà luogo alla domanda, se il campione sia rappresentativo della totalità, anche nel senso di conservare gli indici di connessione fra i caratteri.

Incominciamo con lo sperimentare quale sia la concentrazione dei sette caratteri strumentali della scelta e dei due caratteri di verificazione nella totalità e nel gruppo scelto. Nella Tav. 13 sono inscritti i valori R_a ed R_e, per la totalità e per il campione, del *rapporto di concentrazione* R (1) calcolato mediante la formula

$$R = \frac{\sum_{l=1}^{s} (i_l + i_{l-1} - 1) f_l x_l}{(n-1) A_n}$$

dove x_l denota uno degli s valori distinti assunti dal carattere considerato, f_l è la rispettiva frequenza, i_l è il numero dei valori del carattere inferiori od uguali ad x_l, n è il numero totale dei casi, A_n è la somma di quei valori in tutti questi casi (2). Si constata che lo scarto relativo fra R_a ed R_e non è mai inferiore al 5 %; è compreso fra il 5 e il 10 % per la natalità e la mortalità; supera il 10 % in tutti gli altri casi. Secondo il solito criterio di apprezzamento sommario dell'accordo fra totalità e campione, potremo dire che questo accordo è sufficiente in due casi, ed è insufficiente nei rimanenti sette. Si vede dunque che, nell'insieme, il campione risulta meno rappresentativo per ciò che riguarda la concentrazione di quanto non lo fosse per l'intensità dei caratteri. Non solo: ma quei caratteri pei quali la scelta eseguita è meglio rappresentativa nel conservare i loro valori medî, non sono necessariamente quegli stessi pei quali la medesima scelta è meglio rappresentativa per quanto riguarda la loro concentrazione. Un tale parallelismo talora si verifica e talora no. Così per la natalità

---

(1) C. Gini: *Sulla misura della concentrazione e della variabilità dei caratteri* « Atti del R. Ist. Ven. di S. L. A. », Anno 1913-14, Tomo LXXV, Parte II.

(2) A scopo di verifica dei valori ottenuti, abbiamo anche eseguito l'esame della concentrazione dei caratteri descrivendo, per ciascuno di questi, la curva di concentrazione nella totalità e nel gruppo parziale, e valutando l'area di concentrazione sia a mezzo di un planimetro Amsler, sia per via grafica, mediante il tracciamento delle curve integrali.

e la mortalità si osserva un certo accordo fra gli scarti relativi dell'intensità *media* (1,06 % e rispettivamente 0,29 %) e della concentrazione (7,5 % e rispettivamente 9,1 %), ossia il campione è ottimamente rappresentativo rispetto all'intensità e sufficientemente rappresentativo rispetto alla concentrazione. Invece esso è ottimamente rappresentativo rispetto alla nuzialità media (scarto relativo dell'1,06 %) e lo è insufficientemente per la concentrazione dello stesso carattere (scarto relativo del 22,4 %). L'*indice di cograduazione* (1) fra gli scarti relativi per la media e per la concentrazione è — 0,2.

In conclusione, nel nostro campione non si osserva nessuna analogia di comportamento per quanto riguarda la rappresentatività rispetto all'intensità dei caratteri e rispetto alla loro concentrazione; ciò che, d'altronde, non costituisce nulla di sorprendente perchè, in generale, il fatto della conservazione del valore medio di un carattere passando dalla totalità al campione non subordina quello del conservare la concentrazione del carattere stesso.

Vedremo più avanti se e quando la rappresentatività per quanto concerne un aspetto di un carattere possa determinare quella di altri aspetti dello stesso carattere (n. 18 e segg.).

Esame della distribuzione dei caratteri.

11. In secondo luogo paragoniamo le distribuzioni dei soliti 9 caratteri per gradi equidistanti di intensità, quali appaiono dalle Tavole 14-22. Il paragone viene facilitato sostituendo, per ciascun carattere, alle distribuzioni originarie nella totalità e nel campione (coll. *a* e *b*) altre due distribuzioni rispettivamente simili (coll. *c* e *d*) contenenti lo stesso numero di unità, per es. 10 milioni. Le Tavole grafiche 23-31 traducono in diagrammi le coll. *c* e *d*.

Per esprimere sinteticamente il risultato del confronto di due distribuzioni è opportuno impiegare l'*indice relativo di dissomiglianza*, che si ottiene ragguagliando alla intensità media della distribuzione totale l'indice assoluto di dissomiglianza, il quale a sua volta si deduce dalla media della differenze fra le quantità di ugual posto (cograduate)

(1) Per la teoria di questo indice vedi C. Gini: *Di una misura delle relazioni fra le graduatorie di due caratteri*. Roma, Cecchini, 1914, ed *Indici di concordanza* « Atti del R. Istituto Ven. di S. L. A. », Anno 1915-16, Tomo LXXV, Parte II.

nelle due distribuzioni (1). L'indice relativo di dissomiglianza può essere riguardato come lo scarto relativo fra le due distribuzioni, e quindi esso si annulla se e soltanto se queste sono simili.

La Tav. 32 raccoglie i valori degli indici assoluti e relativi di dissomiglianza fra le corrispondenti distribuzioni nel complesso totale e nel campione. Seguendo il solito criterio adottato per gli scarti relativi delle intensità medie e dei rapporti di concentrazione, si constaterà che l'accordo fra totalità e campione non è ottimo in nessun caso, è soddisfacente in tre, è sufficiente in due, è insufficiente negli altri quattro casi.

Va notato che nel caso della densità l'indice relativo di dissomiglianza raggiunge il 93 %, e ciò dipende dalla circostanza che nel campione non sono rappresentati i circondarî di densità più elevata, quali sono Napoli, Milano, Livorno ecc.

L'indice di dissomiglianza fra due distribuzioni dipende anzitutto dalla differenza fra le medie delle due distribuzioni e poi dalle differenze nell'intensità e nella forma della dispersione delle quantità delle due distribuzioni intorno alle medie rispettive. Anche la differenza fra le concentrazioni di due distribuzioni dipende dalla differenza fra le intensità di queste, ma non dipende affatto nè dalla differenza tra le forme della dispersione, nè dalla differenza fra le medie dei due caratteri.

Perciò non v'è luogo a pensare che possa verificarsi una stretta concordanza fra lo scarto relativo dei rapporti di concentrazione e l'indice relativo di dissomiglianza delle distribuzioni; e difatti la cograduazione fra queste quantità risulta bassa (indice = 0,10). Al contrario la cograduazione fra lo scarto relativo delle medie e l'indice relativo di dissomiglianza risulta abbastanza elevato (indice = 0,55).

Esame delle reciproche relazioni fra i caratteri.

12. Passiamo ora a considerare gli *indici di correlazione* e quelli *di omofilia* (2), nella totalità e nel gruppo dei circondarî scelti, per le 36 coppie a cui danno luogo i 9 caratteri combinati a 2 a 2 in tutti i modi possi-

---

(1) Per la teoria dell'indice di dissomiglianza vedi C. Gini: *Di una misura della dissomiglianza tra due gruppi di quantità* ecc. « Atti del R. Istituto Ven. di S. L. A. ». Anno accad. 1914-15, Tomo LXXIV, Parte II.

(2) Per gli indici di omofilia e per le loro relazioni con quelli di correlazione vedi C. Gini: *Indici di omofilia e di rassomiglianza e loro relazioni col coefficiente di correlazione e con gli indici di attrazione* « Atti del R. Istituto Ven. di S. L. A. », Anno accademico 1914-15, Tomo LXXIV, Parte II. Vedasi anche, dello stesso autore: *Indici di concordanza*, ibid., 1915-16, Tomo LXXV, Parte II.

bili. La Tav. 33, che contiene tali indici ed i corrispondenti scarti assoluti e relativi, dimostra che in 32 casi per gli indici di correlazione e pure in 32 per quelli di omofilia, essendo lo scarto relativo superiore al 10 %, il campione risulta insufficientemente rappresentativo.

Riferendosi all'indice di correlazione, il campione è ottimo per la coppia natalità-popolazione agricola, soddisfacente per la coppia nuzialità-accrescimento naturale, sufficiente per le due coppie natalità-altitudine sul livello del mare e natalità-accrescimento naturale. Riferendosi, invece, all'indice di omofilia, il campione è soddisfacente in un caso (nuzialità-accrescimento naturale) e sufficiente in tre altri (natalità-popolazione agricola; natalità-altitudine sul mare; popolazione agricola-densità).

Lo scarto relativo raggiunge talvolta valori elevatissimi: in 16 casi, tanto per gli indici di correlazione che per quelli di omofilia, supera il 100 % e in uno raggiunge il 4000 %. Infine va rilevato che un buon accordo fra totalità e campione nei valori medî di due caratteri, o nelle loro concentrazioni o nelle loro distribuzioni non garantisce affatto l'accordo fra gli indici di omofilia e fra quelli di correlazione. Lo scarto relativo più elevato fra gli indici di omofilia (4100 %) e fra quelli di correlazione (3900 %) si verifica per la coppia di caratteri nuzialità-reddito, nonostante che per quanto concerne la nuzialità il campione fosse ottimo pel valore medio (Tav. 10) e soddisfacente per la distribuzione (Tav. 32) e per quanto concerne il reddito esso fosse soddisfacente per la media (Tav. 10) e sufficiente per la distribuzione (Tav. 32).

Inoltre fra nuzialità e reddito, che hanno dunque presso a poco gli stessi valori medî nella totalità e nel campione (scarti relativi: -1,06 %, 2,80 %, cfr. Tav. 10), si avverte negli indici di correlazione e di omofilia uno scarto assai maggiore (3900 % e rispettivamente 4100 %) che fra popolazione agricola e densità (scarti relativi degli indici di correlazione e di omofilia 90 % e rispettivamente — 7 %), pure avendo questi due caratteri dei valori medî molto diversi fra di loro (scarti relativi — 7,98 % e rispettivamente — 11,46 %, cfr. Tav. 10).

La natalità e la mortalità, che presentano fra totalità e campione un accordo sufficiente per il rapporto di concentrazione (Tav. 13), danno luogo per gli indici di correlazione e per quelli di omofilia a degli scarti relativi del 31 % e 50 % (accordo insufficiente); e, al contrario, la nuzialità e l'accrescimento naturale per cui l'accordo del rapporto di concentrazione nella totalità e nel campione è insufficiente, presentano un soddisfacente accordo sia negli indici di correlazione che in quelli di omofilia (scarti relativi — 4 %).

Fra le 6 coppie a cui danno luogo i caratteri: natalità, mortalità, nuzialità, altitudine, pei quali il campione era ottimamente rappresentativo dei valori medî (vedi Tav. 10), la sola coppia natalità-altitudine presenta accordo sufficiente, nel totale e nel gruppo parziale, per gli indici di correlazione e di omofilia (scarti relativi — 8 % e — 7 %).

La coppia che si può formare coi caratteri: natalità e mortalità, pei quali il campione è sufficientemente rappresentativo del rapporto di concentrazione (vedi Tav. 13), è tale che per essa il campione non è rappresentativo nei riguardi dei rapporti di correlazione e di omofilia.

Fra le 10 coppie costituite coi caratteri: natalità, mortalità, nuzialità, reddito, accrescimento naturale, pei quali il campione è almeno sufficientemente rappresentativo delle distribuzioni, come appare dagli indici di dissomiglianza della Tav. 32, la sola coppia nuzialità-accrescimento naturale dà luogo a soddisfacente accordo fra totalità e campione, sia per gli indici di correlazione che di omofilia; e la sola coppia natalità-accrescimento naturale dà luogo ad un accordo sufficiente per il solo indice di correlazione.

Conclusione di tutti questi confronti è dunque che *il campione formato non è generalmente rappresentativo delle relazioni intercedenti fra le varie coppie di caratteri considerati, nè valutate mediante gli indici quadratici di correlazione, nè mediante gli indici quadratici di omofilia.*

Ed invero non v'è alcuna ragione per far ritenere che un campione rappresentativo per quanto concerne la media o la distribuzione o la concentrazione dei diversi caratteri, lo sia anche per ciò che riguarda le loro mutue relazioni.

Confronto mediante rette interpolatrici.

13. Infine un ulteriore esame delle reciproche relazioni fra i diversi fenomeni, può essere eseguito confrontando le distribuzioni di ciascun carattere rispetto ai rimanenti, nella totalità e nel campione. A tal fine, per ognuna delle 72 disposizioni a cui danno luogo i soliti 9 caratteri abbiamo rappresentato i due sistemi di circondarî, complessivo e parziale, mediante punti di un piano, assumendo come ascissa il valore del primo e come ordinata il valore del secondo carattere. Per ognuna delle disposizioni stesse abbiamo poi condotto le rette che interpolano il sistema totale e il sistema

parziale di punti, minimizzando la somma dei quadrati delle distanze nella direzione delle ordinate. Benchè una retta interpolatrice abbia uno scarso valore indicativo nel sintetizzare l'andamento di una distribuzione, a meno che questa non sia approssimativamente lineare, pur nondimeno il confronto dei coefficienti angolari e delle ordinate all'origine, effettuato per ciascuna coppia di rette interpolatrici, conferma anche esso (come si rileva dalla Tav. 34 in cui abbiamo raccolto i valori di tali elementi) l'esistenza di notevoli divarî nelle reciproche relazioni fra i varî caratteri, secondo che questi si considerino nella totalità e nel campione.

Notiamo altresì che, se invece di assumere come coordinate dei punti immagini dei circondarî i valori dei loro caratteri, quali risultano dalla Tav. 1, avessimo misurato ciascun carattere a partire dal suo valore medio ed assunto come unità di misura il corrispondente scarto quadratico medio, allora il coefficiente angolare della retta interpolatrice per le singole coppie di caratteri sarebbe stato uguale al relativo coefficiente di correlazione (1). Ma, trattandosi di eseguire una comparazione fra la totalità e il campione, non si è ritenuto necessario eseguire quel cambiamento di unità. Comunque, più che sulle rette interpolatrici, è sui coefficienti di correlazione e su quelli di omofilia, considerati nel numero precedente, che un confronto decisivo, circa le mutue relazioni dei caratteri nella massa e nel campione, può essere basato.

Per questa ragione, omettendo il paragone delle rette interpolatrici, si sono sintetizzati i risultati di tutti i precedenti confronti nella Tav. 35: la quale esprime dunque, nel suo complesso, che *l'avere scelto il nostro campione in modo che risultasse ottimo o soddisfacente o sufficiente nel conservare l'intensità media di sette caratteri diversi, non è valso ad assicurare che esso fosse sempre per lo meno sufficientemente rappresentativo della intensità media di altri caratteri non tenuti presenti nella scelta, e tanto meno della variabilità, della distribuzione e delle mutue relazioni dei varî caratteri considerati.*

---

(1) Se $y = a\,x + b$ è la retta che interpola i punti $(x_i\,y_i)$ minimizzando la somma $\Sigma\,(y_i - a\,x_i - b)^2$ ed x_i, y_i denotano gli scostamenti che i valori di due caratteri quantitativi X, Y nel caso i^o hanno dalle medie aritmetiche rispettive, allora si trova facilmente $b = o$ ed $a = \frac{\Sigma\,x_i\,y_i}{\Sigma\,x_i^2} = \frac{\Sigma\,x_i\,y_i}{n\,\sigma_x\,\sigma_y}\,\frac{\sigma_y}{\sigma_x} = r_{xy}\,\frac{\sigma_y}{\sigma_x}$, che è il coefficiente di regressione di Y rispetto ad X. (Cfr. p. es. G. U. YULE, *An Introduction to the Theory of Statistics*, London, 1911). Se inoltre si assume $\sigma_x = \sigma_y = 1$, risulta semplicemente $a = r_{xy}$.

II. — Considerazioni generali sulla rappresentatività.

Critica della consueta definizione di rappresentatività.

14. I risultati ai quali siamo pervenuti nella ricerca prevalentemente pratica alla quale ci eravamo accinti, hanno evidentemente una portata che trascende quella particolare ricerca e che investe il problema generale della rappresentatività, di cui sembrerebbero dover modificare in alcuni punti le idee correnti fra gli autori che di esso si sono finora occupati.

In primo luogo, è la stessa definizione di rappresentazione e di rappresentatività che va discussa. Nella concezione comune la rappresentazione dovrebbe fornire una specie di *riduzione di tutte le parti dell'insieme* (1), e il metodo rappresentativo mirerebbe a rendere possibile *la generalizzazione dei risultati che si traggono da una ricerca parziale* (2); cosicchè, secondo tale modo assoluto e generico di intendere la rappresentatività, *questa dovrebbe comprendere qualunque carattere del campione e qualunque aspetto del carattere*. Ma è veramente possibile che nella pratica un campione ottemperi a un tale complesso di condizioni?

Secondariamente, vien fatto di osservare che *il mezzo usualmente impiegato per verificare*, nella pratica, *se un campione sia rappresentativo della totalità, consiste nel prendere soltanto in considerazione l'intensità media dei caratteri e la frequenza delle diverse modalità sul totale, omettendo qualsiasi altra verifica la quale concerna l'ulteriore contenuto logico della definizione consueta di rappresentatività.*

Si manifesta, dunque, una vera incoerenza fra la definizione usuale di rappresentatività e il modo di verificarla. Nè vale obbiettare che anche nella matematica si pongono alcune definizioni astratte, la cui verifica pratica non può essere compiuta rigorosamente per gli errori inevitabili in qualunque forma di confronto e di misura.

Così, per esempio, realizzati con modelli due triangoli uguali nel senso comune della parola, non si potrà mai dire se essi siano geometricamente

---

(1) Cfr. L. March: *Observations sur la méthode représentative et sur le projet de rapport relatif à cette méthode* « Bulletin de l'Institut International de Statistique », Tome XXII, 1ère livraison, Rome 1926, p. 446.

(2) Cfr. A. Jensen: *Report on the Representative Method in Statistics*, già citato.

uguali. Gli è che, in casi come questo, la definizione riguarda un oggetto ideale e la sua applicazione ad oggetti concreti è necessariamente viziata da errori di valutazione, per quanto questi possano talora essere contenuti in certi limiti di approssimazione.

Nel nostro caso si tratta di ben altro, e cioè di insufficienza logica del criterio impiegato per verificare la definizione, rispetto al contenuto totale di questa.

Del resto, tale discordanza si connette al procedimento che comunemente si segue quando si voglia designare un gruppo parziale, rappresentativo della totalità nel senso usuale della parola, non per via di una estrazione a sorte, ma per mezzo di una scelta giudiziosa. Difatti, sia perchè il simultaneo paragone dei diversi aspetti dei caratteri richiederebbe un esame così lungo e minuzioso da rendere l'operazione della scelta assolutamente impraticabile (mentre il confronto fra le medie si effettua rapidamente nel corso di quel procedimento e consente di eseguire via via le necessarie sostituzioni di un aggruppamento ad un altro), sia anche per la presunzione che, se il campione è abbastanza ampio, le sue caratteristiche non risultino molto dissimili da quelle della totalità, il fatto è che *soltanto la coincidenza dei valori medî dei caratteri quantitativi è quella di cui ci si appaga quando si operi una «scelta giudiziosa» al fine di avere un campione genericamente rappresentativo della totalità.*

E altrettanto si fa anche quando si voglia saggiare se un gruppo parziale estratto a sorte sia conforme alla totalità.

Orbene, che questo sia un mezzo spedito di scelta, nessuno potrà negare: ma che esso sia adeguato al soddisfacimento di tutte le condizioni implicite nel comune concetto di rappresentatività, non si potrebbe certamente affermare.

La rappresentatività è un concetto relativo e non assoluto.

15. Comunque, indipendentemente da qualsiasi considerazione di ordine teorico, l'esperienza da noi fatta dimostra che un campione può essere dichiarato rappresentativo per quanto riguarda alcuni aspetti di determinati caratteri, ma, almeno generalmente, non può venir dichiarato tale per tutti i caratteri e per tutti i loro aspetti.

La segnalata incoerenza può, pertanto, essere eliminata soltanto a patto di riconoscere che *la rappresentatività è un concetto relativo e non assoluto*, e quindi:

a) *sostituendo alla consueta definizione di rappresentatività una definizione particolaristica che possa riferirsi ad uno o ad alcuni determinati aspetti di uno o alcuni determinati caratteri, e subordinatamente:*

b) *sostituendo all'accennato consueto mezzo di verificazione, quello che, caso per caso, risulterà idoneo per il particolare aspetto che nei caratteri considerati dovrà essere colto.*

Con ciò non si vuole escludere l'astratta possibilità che un campione possa per tutti i caratteri esaminabili e per tutti i loro aspetti, essere perfettamente rappresentativo della totalità (1), ma poichè i caratteri e gli aspetti dei caratteri pei quali si può porre la questione della rappresentatività di un campione sono teoricamente infiniti e praticamente troppo numerosi per poterli tutti verificare, così *in generale manca praticamente la possibilità di affermare la rappresentatività del campione se non per quei caratteri e quei loro aspetti, in numero necessariamente ristretto, che sono stati sottoposti a verificazione.*

Non si può dunque mai ritenere che un campione possa completamente sostituire la totalità delle osservazioni; e perciò una rilevazione incompleta non potrà mai essere considerata come equivalente alla rilevazione completa dell'insieme.

Siamo tuttavia ben lontani dallo sconsigliare in modo assoluto l'impiego del metodo rappresentativo, e riconosciamo che vi sono dei casi in cui vi si deve forzatamente ricorrere, come è avvenuto per l'applicazione che ha dato occasione al presente studio. Si può anche consentire che in talune evenienze, essendo il fine della ricerca limitato ad alcuni aspetti di alcuni caratteri, il metodo rappresentativo può raggiungere lo scopo con maggiore economia; ma non ci dobbiamo nascondere che la validità del campione è sempre limitata e che, se non si vogliono preventivamente imporre dei limiti all'impiego dei dati, è necessario eseguire la rilevazione totale.

---

(1) Basta pensare al caso banale di un insieme $A+A'$ costituito da due insiemi A ed A' perfettamente uguali. L'insieme parziale A costituirà un campione rappresentativo in senso assoluto dell'insieme totale $A+A'$.

Come dovrebbe delinearsi una teoria generale della rappresentatività.

16. Se si volesse tracciare una teoria generale della rappresentatività sarebbe dunque necessario prendere le mosse da una definizione particolaristica e dire che *un campione è rappresentativo per quanto concerne uno o più determinati aspetti di uno o più determinati caratteri se la parziale rilevazione a cui esso dà luogo fornisce, per tali aspetti di tali caratteri, gli stessi risultati che si avrebbero dalla rilevazione della totalità dei casi*. Pertanto, se il carattere che si ha in vista è unico, gli aspetti di volta in volta esaminabili potranno essere: l'*intensità* (mediante i valori medî), la *forma della distribuzione* (mediante i rapporti di frequenza o mediante le funzioni descrittive delle distribuzioni, a seconda dei casi), la *variabilità* (mediante gli indici di concentrazione o di variabilità), ecc. Quando, invece, si tratti di due o più caratteri potranno, oltre quegli aspetti, considerarsi anche le *mutue relazioni* fra i caratteri stessi (mediante gli indici di dissomiglianza, di connessione, di concordanza, di cograduazione, mediante le funzioni descrittive delle superficie di correlazione, ecc.).

Così, per es., una rilevazione parziale sarà rappresentativa rispetto ad un fenomeno o carattere per quanto concerne il suo valore medio (dato che tale carattere sia misurabile) se il valore medio da essa dedotto coinciderà con quello deducibile dalla rilevazione totale.

Similmente la rappresentatività rispetto ad un carattere, per quanto concerne la sua distribuzione, esigerà che le classi di frequenza ottenute dalla rilevazione incompleta, in corrispondenza alle diverse modalità di quel carattere, siano ordinatamente proporzionali alle classi analoghe ottenute dalla rilevazione completa. E, pertanto, se il carattere considerato, essenzialmente qualitativo, ammette modalità che si escludono a vicenda o disgiuntive, come per es. il sesso (M, F), lo stato civile (celibe, coniugato o vedovo), ecc., le classi di frequenza risulteranno senz'altro formate nelle due rilevazioni completa e incompleta; cosicchè indicando, per es., con $M_{\mathfrak{N}}$ ed $F_{\mathfrak{N}}$ i numeri dei maschi e delle femmine di una popolazione totale di $\mathfrak{N}$ individui, e con M_n ed F_n i maschi e le femmine del gruppo parziale di n individui, si dovrà avere: $M_n : M_{\mathfrak{N}} = F_n : F_{\mathfrak{N}} = n : \mathfrak{N}$. Se invece il carattere considerato è qualitativo, od anche quantitativo ma tale che le sue modalità costituiscano, per loro natura o convenzionalmente, un aggregato discontinuo (per es. distribuzione dei redditieri a seconda del reddito, distinto in classi di determinata ampiezza), le classi di frequenza saranno subordinate agli elementi di quell'aggregato, e si potrà ripetere per questo caso il già detto nel caso precedente. Se infine il carattere è quantitativo (od anche qualitativo misurabile) e tale che le sue modalità si debbano o convenga considerare come un aggregato continuo, si potrà parlare di un valore della frequenza Y in corrispondenza ad ogni valore X del carattere dentro i limiti di definizione di questo; ed allora, se le curve di frequenza ottenute dalla rilevazione parziale e da quella completa sono rispettivamente:

$$Y = n \,.\, \varphi(X) \text{ e } Y = \mathfrak{N} \,.\, \psi(X), \; \left(\int_{-\infty}^{\infty} \varphi(X)\, dX = 1, \; \int_{-\infty}^{\infty} \psi(X)\, dX = 1 \right)$$

si dovrà avere identicamente, cioè per ogni valore di X, $\varphi(X) = \psi(X)$.

Come ultimo esempio, atto a lumeggiare le diverse possibili specie di rappresentatività implicite nella nostra definizione, consideriamo quello di una rilevazione parziale la quale debba essere rappresentativa rispetto a due caratteri A e B per quanto concerne un loro indice di concordanza, per esempio l'indice quadratico di omofilia. Perchè ciò accada non sarà affatto necessaria la conservazione della proporzionalità fra le ordinate corrispondenti delle curve di distribuzione dei singoli caratteri nel gruppo parziale e nella totalità, e nemmeno la conservazione del valore medio di ciascun carattere, ma sarà sufficiente che l'indice di omofilia dedotto per la coppia dei due caratteri nella rilevazione parziale sia quello stesso che si dedurrebbe dalla totalità.

In conclusione, la premessa definizione di rappresentatività ha un diverso contenuto non soltanto a seconda della qualità e del numero dei caratteri che si intendono considerare, ma anche a seconda degli aspetti, intrinseci o mutui, che nei caratteri stessi si vogliono cogliere.

Problemi diretti ed inversi della rappresentatività.

17. Circa l'effettiva applicazione della posta definizione di rappresentatività è appena necessario avvertire che se di un insieme di unità statistiche si sia eseguita la rilevazione completa ed una rilevazione incompleta, la verifica se quest'ultima sia rappresentativa, rispetto ai caratteri che si vogliono considerare e per quanto concerne uno o più dei loro aspetti, potrà eseguirsi applicando senz'altro la definizione stessa.

Che se, invece, sia stata eseguita soltanto una delle due rilevazioni, allora quella definizione, pur conservando tutto il suo valore per indicare un ideale stato di fatto, dovrà praticamente sostituirsi con il calcolo della probabilità che le misure dei varî aspetti che si vogliono esaminare, dedotte dalla rilevazione effettuata, siano affette da errori non oltrepassanti certi limiti rispetto alle analoghe che si potrebbero dedurre dall'altra rilevazione, qualora eseguita.

E, precisamente, si dovranno risolvere problemi diretti quando, eseguita la rilevazione totale, se ne voglia inferire quale potrà essere il comportamento delle caratteristiche considerate in un gruppo parziale; oppure quando si vogliano stabilire dei criterî di scelta di tale gruppo, affinchè gli errori di quelle caratteristiche siano probabilmente contenuti entro certi limiti; ed anche quando si vorrà calcolare la probabilità affinchè un campione, di estensione prefissata, fornisca per una certa caratteristica un valore differente per meno di ε dall'analogo tratto dalla totalità. Di tale natura è, p. es., il problema considerato nell'Appendice *A*.

Si dovranno, invece, risolvere problemi inversi quando si vorranno dedurre dai valori delle caratteristiche calcolate nelle rilevazioni parziali, gli analoghi valori probabili per la totalità del gruppo statistico.

L'altezza media degli individui di una popolazione totale è H; qual'è la probabilità che in un gruppo parziale, estratto a caso, l'altezza media sia compresa tra $H - \varepsilon$ ed $H + \varepsilon$? Oppure: una popolazione di $\mathfrak{N}$ individui si divide, rispetto ai gradi di un carattere, in classi parziali di $q_1 \mathfrak{N}$, $q_2 \mathfrak{N}$, $q_3 \mathfrak{N}, \ldots$ individui; qual'è la probabilità che estratti a caso n individui e dette $q'_1 n$, $q'_2 n$, $q'_3 n, \ldots$ le classi parziali, si trovi

$$(q_1 - q'_1)^2 + (q_2 - q_2')^2 + (q_3 - q'_3)^2 + \ldots < \varepsilon ?$$

o che si trovi

$$q_1 - \varepsilon_1 < q_1' < q_1 + \varepsilon_1 , \quad q_2 - \varepsilon_2 < q_2' < q_2 + \varepsilon_2 , \ldots ?$$

Così pure: in un certo insieme totale la connessione fra due caratteri *A* e *B* è misurata da un certo valore r dell'indice di correlazione; qual'è la probabilità che in un campione, estratto dall'insieme, l'indice di correlazione fra gli stessi caratteri *A* e *B* sia compreso fra $r \pm \varepsilon$?

Ed anche: un campione è rappresentativo del valore medio (o di un'altra caratteristica) di certi caratteri: con quale approssimazione il valore medio di un nuovo carattere tratto dal campione rappresenterà il valore medio del medesimo carattere nella totalità? (Si veda, a questo proposito, l'Appendice *B*).

Di tale natura possono essere i più semplici problemi diretti ed inversi che si presentano nella questione della rappresentatività; ma è facile concepirne altri, ad essi analoghi o più complicati, riflettenti, come si è detto, altri aspetti dei singoli caratteri o delle loro mutue relazioni.

Di alcuni fra essi ebbero recentemente ad occuparsi, come già si disse, lo JENSEN (1) e segnatamente il BOWLEY (2) nei loro Rapporti sul Metodo rappresentativo, presentati alla XVI Sessione dell'Istituto Internazionale di Statistica. In particolare, l'indagine del BOWLEY approfondisce, dal punto di vista matematico, sotto particolari ipotesi, i problemi diretti ed inversi che si presentano nel caso della scelta giudiziosa e in quello della estrazione a sorte (supposto che tale estrazione a sorte sia fatta per singole unità statistiche e non per gruppi di unità) per quanto concerne le seguenti caratteristiche:

a) distribuzione di attributi che si escludono a vicenda e particolarmente proporzione, rispetto al totale, delle unità aventi un certo attributo;

b) valore medio di un carattere quantitativo.

Ma, nè questo, nè altri Autori accennano agli altri problemi che possono inquadrarsi nella questione generale della rappresentatività (3) e che a maggior ragione debbono essere considerati quando si voglia assumere la consueta definizione di rappresentatività, che è più comprensiva di quella che abbiamo proposto.

---

(1) *Report on the Representative Method*, etc., già citato.

(2) *Measurement of the Precision attained in Sampling*, già citato.

(3) A meno che non si voglia accogliere come una vaga indicazione in tale senso una riserva avanzata dal MARCH là ove dichiara che « dès que l'on s'écarte du fait simple et contrôlé pour examiner les circonstances qui accompagnent ce fait, l'exactitude peut devenir illusoire », *Observations sur la Méthode représentative*, etc., già citato. Lo stesso Autore (*ibid.*) dichiara che in una rilevazione parziale rappresentativa possono essere considerati, oltre i valori medî, anche gli scostamenti quadratici medî; ma nella breve nota matematica che chiude il suo scritto, per dare un'idea dell'influenza che il numero n degli elementi del campione (preso a caso) ha sull'approssimazione con la quale la media di un carattere quantitativo rilevata nel campione stesso fornisce un valore approssimato della media dello stesso carattere nella totalità, si limita a trovare una relazione fra lo scostamento quadratico medio delle medie rilevabili in tutti i possibili campioni di n elementi con lo scostamento quadratico medio del carattere nella totalità.

Rappresentatività subordinate ad altre rappresentatività.

18. Un'ultima questione alla quale vogliamo accennare è la seguente: Abbiamo ripetutamente dichiarato che quello di rappresentatività è un concetto relativo e non assoluto, e abbiamo formulato una definizione che, a differenza di quella usualmente data, si accorda con questa esatta concezione e consente che un campione possa essere rappresentativo di alcuni determinati aspetti di certi caratteri fissati, senza nulla affermare relativamente ad altri caratteri o ad altri aspetti di quei primi caratteri. Ma non potrebbe darsi che, sotto opportune ipotesi:

a) la rappresentatività, per quanto concerne uno o alcuni aspetti di un carattere, subordinasse quella relativa ad altri aspetti dello stesso carattere? Ed anche che:

b) la rappresentatività relativa ad uno o a più aspetti di un carattere subordinasse quella degli stessi aspetti di altri caratteri?

Le risposte sono affermative, come del resto abbiamo avuto occasione di osservare nel corso della nostra applicazione pratica. Ma vale la pena di soffermarsi a considerare succintamente le due questioni, senza peraltro avere la pretesa di esaurirle, dato che esse possono riferirsi a parecchi casi particolari: ci limiteremo dunque a pochi rilievi, alcuni dei quali richiameremo appunto da quella applicazione, o suffragheremo coi risultati che da essa sono emersi.

Subordinazione di alcuni aspetti di un carattere ad altri aspetti dello stesso carattere.

19. In relazione alla questione *a*) è ovvio che, se un determinato aspetto di un carattere è implicito in uno o più altri aspetti del medesimo carattere, pei quali si sia già constatata la rappresentatività del campione, questo risulterà anche rappresentativo per quanto riguarda quel primo aspetto.

Così è ovvio che la conservazione della distribuzione di un carattere, passando dalla rilevazione totale alla parziale, assicurerà quella del valore medio dello stesso carattere, *ma non reciprocamente*.

Così pure, la conservazione del valore medio e della differenza media di un carattere subordinerà quella del rapporto di concentrazione dello stesso carattere, ma non inversamente, etc.

Rappresentatività subordinate dalla numerosità del campione.

20. Per quanto concerne la questione *b*) si può anzitutto notare che se due caratteri X ed Y sono assolutamente indipendenti, non vi sarà alcuna ragione per ritenere che la rappresentatività di un certo aspetto di X subordini senz'altro la rappresentatività dello stesso aspetto di Y. Ma se il campione sarà abbastanza numeroso, in se stesso e rispetto alla totalità, perchè l'azione perturbatrice del caso sia eliminata, allora le due rappresentatività si verificheranno insieme e l'accordo fra di esse sarà tanto migliore quanto più numeroso sia il campione (cfr. n. 7 α).

Qui, dunque, è la mole del campione che permette di trarre qualche conclusione circa il connettersi della rappresentatività relativamente ad un aspetto di un carattere con l'analoga rappresentatività per un altro carattere. Quasi tutti i problemi trattati dagli Autori, nel campo della rappresentatività, riguardano un tale lato della questione (1).

Ma vi sono altri casi in cui quella concomitanza può, almeno prevalentemente, verificarsi e ciò accade quando fra i caratteri considerati intercedano determinate relazioni. Fra queste le più semplici sono quelle lineari, e su tale supposizione si basano appunto le considerazioni che seguono, delle quali parte concernono le quantità fondamentali dell'aggregato esaminato, parte i suoi caratteri statistici propriamente detti.

Rappresentatività subordinate da relazioni lineari fra quantità fondamentali.

21. In primo luogo consideriamo, in una totalità e in un campione estrattone, diverse quantità fondamentali (cfr. n. 7 e n. 8).

a) Se di più quantità fondamentali M, N, P, Q... che possono essere considerate nelle varie unità statistiche (per es. circondarî) una è funzione lineare omogenea delle altre, cioè:

$$M = \alpha N + \beta P + \gamma Q + \dots$$

ed i valori globali che tutte le quantità, tranne una, assumono nella totalità e in un gruppo estrattone sono proporzionali, cioè:

$$\frac{\Sigma N_i}{\Sigma' N_i} = \frac{\Sigma P_i}{\Sigma' P_i} = \frac{\Sigma Q_i}{\Sigma' Q_i} = \dots = k$$

(1) Cfr. per es. Bowley, Rapporto citato.

essendo ΣN_i, $\Sigma P_i, \ldots$ i valori globali di N, $P, \ldots$ nella totalità, e $\Sigma' N_i$, $\Sigma' P_i, \ldots$ i valori delle stesse quantità nel gruppo scelto, anche i valori globali della rimanente quantità, nella totalità e nel medesimo campione, saranno fra loro nello stesso rapporto; e difatti:

$$\frac{\Sigma M_i}{\Sigma' M_i} = \frac{\alpha \Sigma N_i + \beta \Sigma P_i + \gamma \Sigma Q_i + \ldots}{\alpha \Sigma' N_i + \beta \Sigma' P_i + \gamma \Sigma' Q_i + \ldots} = k .$$

Come esempio, se avvenisse che, considerando i singoli circondarî, il numero dei morti fosse funzione lineare omogenea del numero dei nati e della popolazione, estratto un campione tale che, in questo e nella totalità, i numeri globali dei nati e degli abitanti stessero in un medesimo rapporto, anche i numeri globali dei morti starebbero nello stesso rapporto.

In particolare, se fosse semplicemente $M = \alpha N$, dall'essere $\Sigma N_i / \Sigma' N_i = k$ risulterebbe senz'altro: $\Sigma M_i / \Sigma' M_i = k$ (cfr. n. 8).

Nelle applicazioni pratiche di questa osservazione bisognerà tenere conto della circostanza che i fattori di perturbazione nelle relazioni fra le quantità M, N, $P, \ldots$ avranno come effetto che le supposte relazioni, anche quando abbiano la tendenza ad essere lineari, non siano lineari nel senso matematico della parola. Si potrà allora parlare, piuttosto che di relazioni rigorosamente lineari, di *relazioni a tendenza lineare* ossia di *relazioni lineari in senso statistico*. In tal caso l'osservazione sarà dunque applicabile a solo patto che il numero delle unità considerate nel campione sia sufficientemente grande affinchè risulti eliminata l'azione del caso. Nel senso ora dichiarato si può, per es., parlare di proporzionalità fra il numero dei fogli di famiglia impiegati in una rilevazione censuaria e il corrispondente numero dei censiti; cosicchè il rapporto fra il numero totale dei fogli di famiglia del censimento e il numero dei fogli relativi al campione, avrà la tendenza ad essere uguale al rapporto fra la popolazione totale e quella del campione (cfr. n. 8).

b) Conservando le stesse notazioni, se la quantità fondamentale M fosse indipendente dalla N, i rapporti $\Sigma M_i / \Sigma' M_i$ e $\Sigma N_i / \Sigma' N_i$ sarebbero diversi, ma generalmente tanto più prossimi fra loro quanto più numeroso fosse il gruppo parziale rispetto al totale (1).

---

(1) Si supponga che la totalità sia costituita da $\mathfrak{M}$ unità e che se ne estragga un campione, senza ripetizioni, di n unità. Posto $\Sigma M_i = \sum_1^{\mathfrak{M}} M_i = S$, cioè $M(M_i) = \frac{S}{\mathfrak{M}}$, dove M è, come d'uso, il

Si ritorna dunque ad un caso di rappresentatività che si connette ad un'altra rappresentatività per solo effetto della numerosità del campione (cfr. n. 20).

c) Se, infine, la M fosse funzione di $N, P, \ldots$, ma non lineare omogenea, per es. fosse

$$M = \alpha + \beta N + \gamma N^2$$

ne seguirebbe:

$$\frac{\Sigma M_i}{\Sigma' M_i} = \frac{\Sigma \alpha + \beta \Sigma N + \gamma \Sigma N^2}{\Sigma' \alpha + \beta \Sigma' N + \gamma \Sigma' N^2},$$

e non vi sarebbe quindi alcuna ragione perchè i due rapporti $\Sigma M / \Sigma' M$ e $\Sigma N / \Sigma' N$ fossero o tendessero a divenire uguali (cfr. n. 8).

---

simbolo di valore medio, si dedurrà subito che il valore medio della somma dei termini dei campioni di n termini è:

$$M\left(\sum_1^n M_i\right) = M(\Sigma' M_i) = \sum_1^n M(M_i) = \frac{n}{\mathfrak{N}} S.$$

Si consideri ora lo scostamento quadratico medio σ_n dei valori delle singole somme di n termini dalla rispettiva media, cioè:

$$\sigma_n = \sqrt{M\left[\left(Y - \frac{n}{\mathfrak{N}} S\right)^2\right]},$$

dove si intende essere:

$$Y = \sum_1^n M_i = X_1 + X_2 + \ldots + X_n$$

ed $X_1, X_2, \ldots X_n$ sono n variabili casuali capaci di assumere i valori degli M_i di ciascun campione. Poichè:

$$Y - \frac{n}{\mathfrak{N}} S = \left(X_1 - \frac{S}{\mathfrak{N}}\right) + \left(X_2 - \frac{S}{\mathfrak{N}}\right) + \ldots + \left(X_n - \frac{S}{\mathfrak{N}}\right)$$

ed il valore medio di ciascun termine in parentesi è zero, si avrà (cfr., per es., CASTELNUOVO: *Calcolo delle Probabilità*, Vol. I, § 25):

$$M\left[\left(Y - \frac{n}{\mathfrak{N}} S\right)^2\right] = \sum_1^n M\left[\left(X_i - \frac{S}{\mathfrak{N}}\right)^2\right] + 2M\left[\left(X_1 - \frac{S}{\mathfrak{N}}\right)\left(X_2 - \frac{S}{\mathfrak{N}}\right)\right] + \ldots +$$
$$+ 2M\left[\left(X_{n-1} - \frac{S}{\mathfrak{N}}\right)\left(X_n - \frac{S}{\mathfrak{N}}\right)\right] = n\sigma^2 + 2M\left[\left(X_1 - \frac{S}{\mathfrak{N}}\right)\left(X_2 - \frac{S}{\mathfrak{N}}\right)\right] + \ldots +$$
$$+ 2M\left[\left(X_{n-1} - \frac{S}{\mathfrak{N}}\right)\left(X_n - \frac{S}{\mathfrak{N}}\right)\right]$$

essendo σ lo scostamento quadratico medio dei valori M_i nella totalità. Risulta dunque

[1] $$\sigma_n = \sqrt{n\sigma^2 + \ldots + 2M\left(X_{n-1} - \frac{S}{\mathfrak{N}}\right)\left(X_n - \frac{S}{\mathfrak{N}}\right)}$$

Questa espressione esatta di σ_n è troppo complicata perchè possa prestarsi a mostrare il suo comportamento al crescere di n. Osserviamo peraltro che σ_n deve essere una funzione simmetrica di n e di $\mathfrak{N} - n$, perchè ad ogni singola somma (senza ripetizioni) di n termini scelti fra gli M_i ne corrisponde una residua di $\mathfrak{N} - n$ termini, complementare alla prima rispetto ad S: queste due

Anche qui la relazione $M = \alpha + \beta N + \gamma N^2$ si dovrebbe, nelle applicazioni pratiche, considerare non già come verificantesi in senso strettamente matematico, ma piuttosto in senso statistico (relazione a tendenza parabolica).

RAPPRESENTATIVITÀ SUBORDINATE DA RELAZIONI LINEARI FRA CARATTERI.

a) *Conservazione della media aritmetica.*

22. Secondariamente consideriamo, in una totalità e in un campione estrattone, diversi « caratteri » che si possano riguardare come rapporti di quantità fondamentali. Se uno di tali caratteri ha in diversi aggregati

---

somme avranno dunque dai rispettivi valori medî $\frac{n}{\mathfrak{N}} S$ ed $\frac{\mathfrak{N} - n}{\mathfrak{N}} S$ differenze di segno contrario ma di uguale valore assoluto. Inoltre $\sigma_n = 0$ per $n = 0$ e per $n = \mathfrak{N}$.

Ora, per n abbastanza piccolo rispetto ad $\mathfrak{N}$, σ_n avrà un comportamento prossimo a quello che avrebbe se le variabili casuali $X_1\, X_2 \ldots X_n$ fossero indipendenti (cioè se in ciascun campione le M_i si potessero presentare con ripetizione), e quindi sarà approssimativamente:

$$\sigma_n = \sigma \sqrt{n} \tag{2}$$

perchè nella espressione [1] si potranno trascurare i valori medî dei doppi prodotti. Se invece n sarà prossimo ad N, per la già avvertita simmetria si avrà approssimativamente;

$$\sigma_n = \sigma_{\mathfrak{N} - n} = \sigma \sqrt{\mathfrak{N} - n} = \sigma \sqrt{n} \sqrt{\frac{\mathfrak{N}}{n} - 1} . \tag{3}$$

Dalle [2] e [3] si vede dunque che al crescere di n, σ_n andrà prima crescendo e poi decrescendo.

Se poi, invece di considerare lo scostamento quadratico medio assoluto σ_n si considererà lo scostamento quadratico medio σ_n' relativo alla media $\frac{n}{\mathfrak{N}} S$, si avrà approssimativamente e rispettivamente per n piccolo e per n grande a paragone di $\mathfrak{N}$:

$$\sigma_n' = \frac{\sigma}{\sqrt{n}} \tag{2'}$$

$$\sigma_n' = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{\mathfrak{N}}{n} - 1} . \tag{3'}$$

Al crescere di n l'espressione [3′] di σ'_n decresce più rapidamente di quella [2′].

In questo senso si può quindi dire che al crescere di n il quoziente $\frac{\Sigma' M_i}{\Sigma M_i}$ si avvicina sempre più ad $\frac{n}{\mathfrak{N}}$.

Se N è un'altra quantità fondamentale, sia pure indipendente dalla M, si potrà nello stesso modo concludere che al crescere di n il quoziente analogo $\frac{\Sigma' N_i}{\Sigma N_i}$ si avvicinerà sempre più ad $\frac{n}{\mathfrak{N}}$.

E pertanto, crescendo n, i rapporti $\frac{\Sigma' M_i}{\Sigma M_i}$ e $\frac{\Sigma' N_i}{\Sigma N_i}$, ed anche i loro inversi, considerati nel testo, saranno generalmente sempre più prossimi fra loro.

parziali di $P_1\ P_2 \ldots P_i \ldots$ unità rispettivamente i valori $X_1\ X_2 \ldots X_i \ldots$, il valore dello stesso carattere nell'insieme di tutti quegli aggregati sarà evidentemente costituito dalla media ponderata $\frac{\Sigma\, P_i X_i}{\Sigma\, P_i}$. Ciò posto:

a) Se di due caratteri X, Y che possono essere considerati in varî aggregati (per es. nei circondarî) di pesi P_1, P_2, ... P_i, ... uno è funzione lineare, omogenea o no, dell'altro, cioè:

$$Y = h\,X + k,$$

e si sceglie un insieme di aggregati o campione tale che in esso la media ponderata di X abbia lo stesso valore che nella totalità, ossia

$$\frac{\Sigma\, P_i\, X_i}{\Sigma\, P_i} = \frac{\Sigma'\, P_i X_i}{\Sigma'\, P_i},$$

allora anche la media ponderata di Y avrà nel campione e nella totalità valori uguali, cioè:

$$\frac{\Sigma\, P_i\, Y_i}{\Sigma\, P_i} = \frac{\Sigma'\, P_i\, Y_i}{\Sigma'\, P_i}.$$

La verifica è immediata, perchè dalla relazione lineare supposta viene subito:

$$\frac{\Sigma\, P_i\, Y_i}{\Sigma\, P_i} = h\,\frac{\Sigma\, P_i X_i}{\Sigma\, P_i} + k \quad , \quad \frac{\Sigma'\, P_i\, Y_i}{\Sigma'\, P_i} = h\,\frac{\Sigma'\, P_i X_i}{\Sigma'\, P_i} + k.$$

Così se X_i, Y_i, P_i, rappresentassero rispettivamente la natalità, la mortalità e la popolazione del circondario i^{mo}, e la mortalità fosse funzione lineare della natalità, poichè

$$\frac{\Sigma\, P_i X_i}{\Sigma\, P_i},\ \frac{\Sigma'\, P_i X_i}{\Sigma'\, P_i},\ \frac{\Sigma\, P_i\, Y_i}{\Sigma\, P_i},\ \frac{\Sigma'\, P_i\, Y_i}{\Sigma'\, P_i}$$

costituiscono, evidentemente, i valori medî della natalità e della mortalità nella totalità e nel campione, si concluderebbe che un campione atto a conservare il valore medio della natalità, conserverebbe pure il valore medio della mortalità, rispetto all'insieme totale (1). In particolare la proprietà

(1) Del caso *a*, ora considerato, è opportuno dare una illustrazione meccanica. Poichè ciascuna unità è dotata di un peso P_i, rappresentate le coppie di valori $(X_i\ Y_i)$ mediante punti di un piano, riferiti, per es. a un sistema cartesiano ortogonale, conduciamo la retta che interpola tale sistema di punti, minimizzando la somma dei quadrati delle distanze nella direzione di Y, dopo avere conferito ad ogni punto (X_i, Y_i) il relativo peso P_i. Tale retta conterrà, come si sa, il baricentro dei punti ponderati

vale se tutti i pesi hanno il valore unitario. Come nel caso *a*) del n. 21, se la linearità della relazione fra X ed Y si verifica soltanto in senso statistico, come accade in tutti i problemi concreti ai quali sarebbero applicabili le presenti considerazioni, la proprietà esposta troverà riscontro nel solo caso che il campione sia abbastanza numeroso da far ritenere eliminate le perturbazioni accidentali. Notiamo anzi, a questo proposito, che se la relazione fra X ed Y fosse lineare in senso matematico, non vi sarebbe ragione di considerare X ed Y come due caratteri distinti, ottenendosi ogni valore di Y dal corrispondente valore di X mediante un cambiamento dell'unità di misura e dell'origine.

b) Più in generale, se di diversi caratteri X, Y, Z,... T uno è funzione lineare degli altri, se cioè:

$$T = h\,X + k\,Y + \ldots + l$$

$(X_i\,,\ Y_i)$. Similmente si conduca la retta che interpola il sistema parziale di punti pesati $(X_i\,,\ Y_i)$ corrispondente alle unità del campione, ed anche questa retta conterrà il baricentro di tale gruppo parziale di punti. D'altronde

$$\frac{\Sigma\, P_i\, X_i}{\Sigma\, P_i},\ \frac{\Sigma'\, P_i\, X_i}{\Sigma'\, P_i},\ \left(\frac{\Sigma\, P_i\, Y_i}{\Sigma\, P_i},\ \frac{\Sigma\,'P_i\, Y_i}{\Sigma\,'P_i}\right)$$

rappresentano i baricentri sull'asse X (e sull'asse Y) del sistema totale e del sistema parziale dei valori X_i (e rispettivamente Y_i), ponderati con gli stessi pesi; si sa pure che tali baricentri sono le proiezioni ortogonali su X (e su Y) dei baricentri giacenti sulla retta interpolatrice del sistema totale e sulla retta interpolatrice del sistema parziale di punti $(X_i\,,\ Y_i)$. Perciò se fra X ed Y intercede esattamente la relazione lineare $Y = h\,X + k$, le due rette interpolatrici coincideranno fra loro (e coincideranno anche con la retta che interpola i soliti punti $(X_i\ Y_i)$ non ponderati) e quindi dalla uguaglianza:

$$\frac{\Sigma\, P_i\, X_i}{\Sigma\, P_i} = \frac{\Sigma\,'P_i\, X_i}{\Sigma\,'P_i}$$

risulterà senz'altro l'uguaglianza:

$$\frac{\Sigma\, P_i\, Y_i}{\Sigma\, P_i} = \frac{\Sigma\,'P_i\, Y_i}{\Sigma\,'P_i}\,.$$

Se, invece, la relazione fra X ed Y è lineare soltanto in senso statistico, le due solite rette interpolatrici generalmente non coincideranno, ma (nell'ambito dei valori X_i) saranno tanto più « prossime » fra loro quanto più numeroso sia il campione rispetto alla totalità. Perciò dalla solita uguaglianza $\dfrac{\Sigma\, P_i\, X_i}{\Sigma\, P_i} = \dfrac{\Sigma'\, P_i\, X_i}{\Sigma'\, P_i}$ si dedurrà che i baricentri dei sistemi totale e parziale di punti $(X_i\ Y_i)$, giacendo all'intersezione di quelle due rette con una stessa parallela all'asse Y, di equazione:

$$X = \frac{\Sigma\, P_i\, X_i}{\Sigma\, P_i}$$

saranno tanto più prossimi quanto più lo siano quelle due rette, il che è come dire che i valori:

$$\frac{\Sigma\, P_i\, Y_i}{\Sigma\, P_i} \quad \text{e} \quad \frac{\Sigma'\, P_i\, Y_i}{\Sigma'\, P_i}$$

saranno tanto più vicini quanto meglio la linearità in senso statistico, che si suppone verificarsi fra X ed Y, si avvicini alla linearità in senso matematico.

ed un campione è scelto in modo che le medie ponderate di X, Y,... abbiano in esso i medesimi rispettivi valori che nella totalità, anche la media ponderata di T avrà nel campione e nell'insieme totale valori uguali (intendendo che tutte le ponderazioni siano fatte con uno stesso sistema di pesi attribuiti alle diverse unità statistiche).

Come caso particolare, poichè l'accrescimento naturale è funzione lineare della natalità e della mortalità, supposto che un campione conservi, rispetto all'insieme totale, i valori medî della natalità e della mortalità, esso conserverà anche il valore medio dell'accrescimento naturale (cfr. n. 9).

c) Come caso limite di una relazione a tendenza lineare, se i caratteri X, Y sono indipendenti fra di loro e si suppone che sia $\Sigma P_i X_i / \Sigma P_i = \Sigma' P_i X_i / \Sigma' P_i$, i valori di $\Sigma P_i Y_i / \Sigma P_i$ e $\Sigma' P_i Y_i / \Sigma' P_i$ saranno diversi, ma generalmente tanto più prossimi fra loro quanto più ampio sia il sistema parziale rispetto al totale (1).

(1) Si può pensare che X_i sia il valore medio del carattere X nelle unità del circondario i^o. — Allora $\frac{\Sigma P_i X_i}{\Sigma P_i}$ è la media del carattere X nella totalità di $\Sigma P_i = \mathfrak{N}$ unità, cioè:

$$M(X) = \frac{\Sigma P_i X_i}{\Sigma P_i}.$$

Invece $\frac{\Sigma' P_i X_i}{\Sigma' P_i}$ è il valore medio dello stesso carattere in un campione di $\Sigma' P_i = n$ unità ($n < \mathfrak{N}$). Come si sa (cfr. CASTELNUOVO, Op. cit., Vol. I, § 29), i valori medî di X dedotti da tutti i possibili campioni di n unità hanno da $M(X)$ uno scostamento quadratico $\frac{\sigma}{\sqrt{n}}$ (essendo σ lo scost. quad. medio del carattere X nella totalità delle $\Sigma P_i = \mathfrak{N}$ unità). Perciò si può dire che al crescere di n, ed anche generalmente al crescere del numero dei circondarî costituenti il campione, l'espressione $\frac{\Sigma' P_i X_i}{\Sigma' P_i}$ si avvicina sempre più alla media generale di X, cioè a $\frac{\Sigma P_i X_i}{\Sigma P_i}$.

Ora, se Y è un altro carattere, sia pure indipendente da X, si potrà, nello stesso modo, concludere che al crescere di n il quoziente $\frac{\Sigma' P_i Y_i}{\Sigma' P_i}$ si approssima sempre più a $\frac{\Sigma P_i Y_i}{\Sigma P_i}$.

Se dunque $\frac{\Sigma P_i X_i}{\Sigma P_i} = \frac{\Sigma' P_i X_i}{\Sigma' P_i}$ ed il campione di $\Sigma' P_i = n$ unità è poco numeroso, non si potrà concludere nulla circa i valori dei due rapporti $\frac{\Sigma P_i Y_i}{\Sigma P_i}$ e $\frac{\Sigma' P_i Y_i}{\Sigma' P_i}$; ma col crescere della numerosità del campione i due quozienti $\frac{\Sigma P_i Y_i}{\Sigma P_i}$ e $\frac{\Sigma' P_i Y_i}{\Sigma' P_i}$ si approssimeranno generalmente sempre più fra di loro.

Si noti, anche qui, che una rappresentatività (circa il valore medio di Y) si connette ad un'altra (circa il valore medio di X) non per effetto di un legame intercedente fra i due caratteri, ma in dipendenza della numerosità del campione (cfr. n. 20).

d) Se, infine, fra due caratteri X, Y, che possono essere considerati nelle varie unità statistiche, intercede una relazione non lineare, non si potrà dire, generalmente, che una scelta conservativa della media ponderata dei valori di X rispetto alla totalità sia anche conservativa della media ponderata dei valori di Y. Così, se fosse:

$$Y = l + k X + h X^2$$

e si ammettesse l'uguaglianza:

$$\frac{\Sigma P_i X_i}{\Sigma P_i} = \frac{\Sigma' P_i X_i}{\Sigma' P_i} ,$$

poichè:

$$\frac{\Sigma P_i Y_i}{\Sigma P_i} = h \frac{\Sigma P_i X_i^2}{\Sigma P_i} + k \frac{\Sigma P_i X_i}{\Sigma P_i} + l$$

$$\frac{\Sigma' P_i Y_i}{\Sigma' P_i} = h \frac{\Sigma' P_i X_i^2}{\Sigma' P_i} + k \frac{\Sigma' P_i X_i}{\Sigma' P_i} + l ,$$

affinchè risultasse $\dfrac{\Sigma P_i Y_i}{\Sigma P_i} = \dfrac{\Sigma' P_i Y_i}{\Sigma' P_i}$ sarebbe anche necessario avere $\dfrac{\Sigma P_i X_i^2}{\Sigma P_i} = \dfrac{\Sigma' P_i X_i^2}{\Sigma' P_i}$ cioè, oltre alla supposta coincidenza delle medie ponderate di X nella totalità e nel campione, si dovrebbe verificare quella delle medie quadratiche ponderate di X negli stessi insiemi.

b) *Scarti fra le medie dei caratteri nella totalità e nel campione.*

23. Nella ipotesi che un carattere T sia funzione lineare di più altri X, Y,... p. es. $T = h X + k Y + l$, diciamo A, B, C ed a, b, c rispettivamente le medie di T, X, Y nella totalità e nel campione, cosicchè:

$$a - A \ , \ b - B \ , \ c - C$$

costituiranno gli scarti assoluti fra i valori medî dei caratteri stessi T, X, Y nella totalità e nel campione, mentre:

$$\frac{a - A}{A} , \frac{b - B}{B} , \frac{c - C}{C}$$

ne saranno gli scarti relativi.

Per l'ipotesi fatta risulterà senz'altro:

$$a - A = \frac{\Sigma' P_i T_i}{\Sigma' P_i} - \frac{\Sigma P_i T_i}{\Sigma P_i} = h (b - B) + k (c - C) \ ,$$

cioè lo scarto assoluto di T sarà funzione lineare omogenea degli scarti assoluti di X e di Y e quindi, essendo h e k finiti, al decrescere indefinito delle differenze $b - B$, $c - C$, decrescerà indefinitamente anche la $a - A$.

Se invece si considera lo scarto relativo $(a - A)/A$ verrà:

$$\frac{a - A}{A} = \frac{h(b - B) + k(c - C)}{hB + kC + l} = \frac{\frac{b - B}{B} \cdot hB + \frac{c - C}{C} \cdot kC}{hB + kC + l} =$$
$$= \frac{b - B}{B} \beta + \frac{c - C}{C} \gamma,$$

la quale esprime che lo scarto relativo fra le medie di T (nella totalità e nel campione) è ancora funzione lineare degli scarti analoghi fra le medie di X e fra quelle di Y; ma i coefficienti β e γ, che sono funzioni di B e C, possono questa volta superare qualunque numero prefissato quando B e C siano tali da tendere:

$$hB + kC + l$$

sufficientemente piccolo, cosicchè lo scarto relativo $\frac{a - A}{A}$ può avere un ordine di grandezza assai diverso dagli ordini di grandezza degli scarti $\frac{b - B}{B}$, $\frac{c - C}{C}$; o, in altri termini, anche nella supposizione che questi due ultimi scarti siano contenuti in limiti secondo i quali la concordanza delle medie di X ed Y, nel campione e nella totalità, debba riguardarsi come soddisfacente, può avvenire che lo scarto $\frac{a - A}{A}$ ecceda quei limiti, cioè che non vi sia soddisfacente concordanza fra le medie di T nel campione e nella totalità (cfr. n. 9).

c) *Uso dei caratteri di verificazione quando essi dipendano linearmente dai caratteri strumentali.*

24. Estratto un campione il quale ottemperi praticamente alla condizione di essere rispetto alla totalità conservativo dei valori medî di alcuni caratteri (caratteri strumentali della scelta), si indaga solitamente se esso conservi anche il valore medio di altri caratteri, e ciò con l'intento di verificare se il campione sia rappresentativo in un senso più generale, vale a dire non soltanto rispetto ai primi, ma anche rispetto a questi nuovi ca-

ratteri (che appunto perciò si possono dire « di verificazione »). Ora, le considerazioni fatte ai nn. 21 e 22 sembrerebbero dimostrare la superfluità del fare uso di caratteri di verificazione quando si abbia ragione di ritenere, in base ad altre ricerche, che essi siano in relazioni lineari coi caratteri strumentali della scelta, cosicchè dalla conservazione dei valori medî di questi ultimi segua la conservazione dei valori medî dei primi. Ma effettivamente accade, nella maggior parte dei casi pratici, che le relazioni stesse si verifichino non in senso rigoroso, ma soltanto in senso statistico; può accadere altresì, per l'osservazione del numero precedente, che gli scarti relativi abbiano, pei caratteri strumentali della scelta e per quelli di verificazione, ordini di grandezza assai diversi fra di loro; di modo che l'uso di tali caratteri di verificazione conserva, anche nei casi in cui intervengano le dette relazioni, tutta la sua pratica opportunità ed efficacia.

d) *Condizioni per la conservazione dello scostamento quadratico medio.*

25. *a*) Se un carattere Y è funzione lineare di un altro carattere X, ed un campione è rappresentativo dello scostamento quadratico medio σ_x di X, esso lo sarà anche per lo scostamento quadratico medio σ_y di Y. La dimostrazione è ovvia.

b) Se Y dipende linearmente da due soli caratteri X, Z ed un campione è rappresentativo dello scostamento quadratico medio di ciascuno di questi, affinchè esso sia anche rappresentativo dello scostamento quadratico medio di Y sarà *necessario* e sufficiente che esso sia anche rappresentativo del coefficiente di correlazione fra X e Z.

Si supponga $Y = h\,X + k\,Z + l$, cosicchè, essendo A, B, C i valori medî di Y, X, Z, risulterà $A = h\,B + k\,C + l$, e quindi $Y - A = h\,(X - B) + k\,(Z - C)$, o più brevemente indicando con y, x, z gli scostamenti rispettivi di Y, X, Z delle medie:

$$y = h\,x + k\,z.$$

Quadrando per ciascuna delle $\mathfrak{N}$ unità statistiche dell'insieme totale, e sommando membro a membro verrà:

$$\Sigma\, y^2 = h^2\, \Sigma\, x^2 + k^2\, \Sigma\, z^2 + 2\,h\,k\, \Sigma\, x\,z = h^2\, \Sigma\, x^2 + k^2\, \Sigma\, z^2 + 2\,h\,k\, \mathfrak{N}\, r_{xz}\, \sigma_x\, \sigma_z\,,$$

cioè:

$$\sigma^2_y = h^2\,\sigma_x^{\,2} + k^2\,\sigma_z^{\,2} + 2\,h\,k\,r_{xz}\,\sigma_x\,\sigma_z$$

dove r_{xz} è il coefficiente di correlazione fra X e Z nella totalità.

Analogamente per le n unità del campione verrà:

$$\sigma'^2_y = h^2 \sigma'^2_x + k^2 \sigma'^2_z + 2\,h\,k\,r'_{xz}\,\sigma'_x\,\sigma'_z$$

dove r'_{xz} è il coefficiente di correlazione fra X e Z nel campione.

E poichè, per ipotesi, $\sigma'_x = \sigma_x$, $\sigma'_z = \sigma_z$, affinchè sia $\sigma'_y = \sigma_y$ sarà *necessario* e sufficiente che gli indici di correlazione nella totalità r_{xz} e nel campione r'_{xz} siano uguali, *c. d. d.*.

c) Se infine Y dipende linearmente da più di due caratteri, X, Z, T, etc., cioè: $Y = h\,X + k\,Z + l\,T + \ldots + m$, si avrà, usando le stesse notazioni di prima, nella totalità, e nel campione:

$$\sigma^2_y = h^2 \sigma^2_x + k^2 \sigma^2_z + l^2 \sigma^2_t + \ldots + 2\,h\,k\,r_{xz}\,\sigma_x\,\sigma_z + 2\,h\,l\,r_{xt}\,\sigma_x \sigma_t + \ldots$$

$$\sigma'^2_y = h^2 \sigma'^2_x + k^2 \sigma'^2_z + l^2 \sigma'^2_t + \ldots + 2\,h\,k\,r'_{xz}\sigma'_x\sigma'_z + 2\,h\,l\,r'_{xt}\,\sigma'_x\sigma'_y + \ldots$$

Ora, nell'ipotesi che sia $\sigma'_x = \sigma_x$, $\sigma'_z = \sigma_z$, $\sigma'_t = \sigma_t$, ..., affinchè sia $\sigma'_y = \sigma_y$ condizione sufficiente (ma non necessaria) sarà $r'_{xz} = r_{xz}$ etc. E pertanto: se Y dipende linearmente da più di due caratteri ed un campione è rappresentativo dello scostamento quadratico medio di questi caratteri, affinchè lo sia anche per lo scostamento quadratico medio di Y sarà *sufficiente* che lo sia per i coefficienti di correlazione degli stessi caratteri, presi a due a due nei diversi modi possibili.

e) *Condizioni per la conservazione della differenza media e per la conservazione del rapporto di concentrazione.*

26. *a*) Se un carattere Y è funzione lineare di un altro carattere X, ed un campione è rappresentativo della differenza media Δ_x di X lo sarà anche della differenza media Δ_y di Y (1).

La dimostrazione è immediata, perchè, supposto $Y = h\,X + k$, e detti $\mathfrak{N}$ ed n i numeri delle unità dell'insieme totale e del campione, si avrà:

$$\mathfrak{N}^2\,\Delta_y = \Sigma\,\Sigma\,|\,Y_i - Y_j\,| = |h|\;\Sigma\,\Sigma\,|X_i - X_j|\;,$$

dove le sommatorie doppie si intendono estese a tutte le possibili disposizioni con ripetizione degli indici i a due a due, ossia

$$\mathfrak{N}^2\,\Delta_y = |\,h\,|\;\mathfrak{N}^2\,\Delta_x$$

e similmente

$$n^2\,\Delta'_y = |\,h\,|\;n^2\,\Delta'_x,$$

(1) In tutto questo numero, le dimostrazioni sono date soltanto per la differenza media con ripetizione; ma esse potrebbero ripetersi, con ovvie modificazioni, anche per le differenze medie semplici.

essendo Δ'_x e Δ'_y le differenze medie di X e di Y nel campione, cosicchè, qualora sia $\Delta'_x = \Delta_x$ sarà anche $\Delta'_y = \Delta_y$.

b) Se Y è funzione lineare di due caratteri X, Z, ed un campione è rappresentativo delle differenze medie Δ_x, Δ_z di questi ultimi, affinchè esso lo sia anche di Δ_y sarà *necessario* e sufficiente che lo sia per la espressione:

$$\frac{\Sigma\,\Sigma p\,|G_i - G_j|}{\upsilon^2},$$

dove υ è il numero delle unità statistiche dell'insieme considerato (1), e dove $\Sigma\,\Sigma p\,|G_i - G_j|$ è la somma dei valori assoluti delle differenze minori in valore assoluto fra quelle corrispondenti di segno contrario a cui danno luogo i caratteri stessi X, Z moltiplicate, tali differenze, per certi parametri che intervengono e definire la data relazione lineare.

Si supponga dunque $Y = h\,X + k\,Z + l$ e si ammetta $h > o$, $k > o$, il che è lecito, poichè se fosse, ad esempio, $h < o$ si potrebbero sostituire i valori di h e di X con quelli di $-h$ e di $-X$.

Ciò posto, la differenza media di Y nell'insieme totale sarà:

$$\mathfrak{N}^2\,\Delta_y = \Sigma\,\Sigma\,|Y_i - Y_j| = \Sigma\,\Sigma\,|h(X_i - X_j) + k\,(Z_i - Z_j)|\,.$$

Ora, poichè $|a + b| = |a| + |b|$ se *sign* a = *sign* b, mentre quando *sign* $a \neq$ *sign* b:

$$|a + b| = |a| + |b| - 2\,|c|\,,$$

essendo $|c|$ il valore assoluto del minore in valore assoluto fra i due numeri a e b, così, ogni volta che le differenze $X_i - X_j$ e $Z_i - Z_j$ siano concordi (od una di esse sia nulla), sarà:

$$|h(X_i - X_j) + k\,(Z_i - Z_j)| = h\,|X_i - X_j| + k\,|Z_i - Z_j|\,,$$

mentre quando le differenze stesse siano discordi sarà:

$$|h\,(X_i - X_j) + k\,(Z_i - Z_j)| = h\,|X_i - X_j| + k\,|Z_i - Z_j| - 2\,p\,|G_i - G_j|$$

dove $p\,(G_i - G_j)$ è fra le due quantità $h\,(X_i - X_j)$ e $k\,(Z_i - Z_j)$ quella minore in valore assoluto.

---

(1) Cioè $v = \mathfrak{N}$ se tale insieme è la totalità, e $v = n$ se l'insieme è il campione.

Si conclude quindi che, in generale:

$$\mathfrak{N}^2 \Delta_y = h \Sigma \Sigma | X_i - X_j | + k \Sigma \Sigma | Z_i - Z_j | - 2 \Sigma \Sigma p | G_i - G_j |$$

ossia che

$$\mathfrak{N}^2 \Delta_y = h \mathfrak{N}^2 \Delta_x + k \mathfrak{N}^2 \Delta_z - 2 \mathfrak{N}^2 \frac{\Sigma \Sigma p | G_i - G_j |}{\mathfrak{N}^2}$$

Similmente nel campione si avrà:

$$n^2 \Delta'_y = h n^2 \Delta'_x + k n^2 \Delta'_z - 2 n^2 \frac{\Sigma' \Sigma' p | G_i - G_j |}{n^2}$$

e poichè, per ipotesi, $\Delta'_x = \Delta_x$, $\Delta'_z = \Delta_z$, affinchè sia $\Delta'_y = \Delta_y$ sarà *necessario* e sufficiente che:

$$\frac{\Sigma \Sigma p | G_i - G_j |}{\mathfrak{N}^2} = \frac{\Sigma' \Sigma' p | G_i - G_j |}{n^2} \quad c. d. d.$$

In particolare, soltanto se i caratteri X e Z sono perfettamente cograduati queste due quantità saranno nulle, cosicchè nella totalità e nel campione avremo rispettivamente:

$$\mathfrak{N}^2 \Delta_y = h \mathfrak{N}^2 \Delta_x + k \mathfrak{N}^2 \Delta_z$$
$$n^2 \Delta'_y = h n^2 \Delta'_x + k n^2 \Delta'_z ,$$

e, avendo fatto l'ipotesi $\Delta'_x = \Delta_x$, $\Delta'_z = \Delta_z$, si concluderà senz'altro che $\Delta'_y = \Delta_y$.

c) Infine, se Y dipende linearmente da più di due caratteri

$$Y = h X + k Z + l T + \ldots + m, \ (h > o , \ k > o , \ l > o , \ldots)$$

si avrà nella totalità:

$$\mathfrak{N}^2 \Delta_y = \Sigma \Sigma | Y_i - Y_j | = \Sigma \Sigma | h (X_i - X_j) + k (Z_i - Z_j) + l (T_i - T_j) + \ldots | .$$

D'altronde:

$$| a + b + c + \ldots + s | = | a | + | b | + \ldots + | s |$$

se $\qquad$ $sign\ a = sign\ b = \ldots = sign\ s\ ;$

ed invece, se i segni dei termini non sono tutti uguali,

$$| a + b + c + \ldots + s | = | \Sigma' + \Sigma'' | = | \Sigma' | + | \Sigma'' | - 2 | \Sigma''' | =$$
$$= | a | + | b | + \ldots + | s | - 2 | \alpha | - 2 | \beta | \ldots$$

dove Σ' denota la somma dei termini positivi, Σ'' la somma dei termini negativi, Σ''' quella fra queste due somme che ha minore valore assoluto, ed infine, α, β, ... sono i termini di Σ'''.

Perciò, quando le differenze $X_i - X_j$, $Z_i - Z_j$, $T_i - T_j$ siano concordi:

$$|h(X_i - X_j) + k(Z_i - Z_j) + l(T_i - T_j) + \ldots| = h|X_i - X_j| + \\ + k|Z_i - Z_j| + l|T_i - T_j| + \ldots$$

mentre, se sono discordi:

$$|h(X_i - X_j) + k(Z_i - Z_j) + l(T_i - T_j) + \ldots| = h|X_i - X_j| + \\ + k|Z_i - Z_j| + l|T_i - T_j| + \ldots - 2p|G_i - G_j| - 2q|H_i - H_j| \ldots$$

essendo $p(G_i - G_j)$, $q(H_i - H_j)$ . . . i termini della minore in valore assoluto fra la somma dei termini positivi e la somma dei termini negativi costistituenti complessivamente il polinomio $h(X_i - X_j) + k(Z_i - Z_j) + l(T_i - T_j) + \ldots$

In generale si avrà dunque, nella totalità di $\mathfrak{N}$ unità statistiche,

$$\mathfrak{N}^2 \Delta_y = h \Sigma \Sigma |X_i - X_j| + k \Sigma \Sigma |Z_i - Z_j| + l \Sigma \Sigma |T_i - T_j| + \ldots - \\ - 2 \Sigma \Sigma p |G_i - G_j| - 2 \Sigma \Sigma q |H_i - H_j| - \ldots$$

che è quanto dire:

$$\mathfrak{N}^2 \Delta_y = h \mathfrak{N}^2 \Delta_x + k \mathfrak{N}^2 \Delta_y + l \mathfrak{N}^2 \Delta_z + \ldots - \\ - 2\mathfrak{N}^2 \frac{\Sigma \Sigma p |G_i - G_j|}{\mathfrak{N}^2} - 2\mathfrak{N}^2 \frac{\Sigma \Sigma q |H_i - H_j|}{\mathfrak{N}^2} - \ldots$$

mentre nel campione di n unità si avrà similmente:

$$n^2 \Delta'_y = h n^2 \Delta'_x + k n^2 \Delta'_y + l n^2 \Delta'_z + \ldots - \\ - 2 n^2 \frac{\Sigma' \Sigma' p |G_i - G_j|}{n^2} - 2 n^2 \frac{\Sigma' \Sigma' q |H_i - H_j|}{n^2} - \ldots$$

Poichè vale l'ipotesi $\Delta'_x = \Delta_x$, $\Delta'_z = \Delta_z$, $\Delta'_t = \Delta_t$, . . . affinchè sia $\Delta'_y = \Delta_y$ sarà necessario che si abbia:

$$\frac{\Sigma \Sigma p |G_i - G_j|}{\mathfrak{N}^2} + \frac{\Sigma \Sigma q |H_i - H_j|}{\mathfrak{N}^2} + \ldots = \\ = \frac{\Sigma' \Sigma' p |G_i - G_j|}{n^2} + \frac{\Sigma' \Sigma' q |H_i - H_j|}{n^2} + \ldots$$

per il che è sufficiente che sia:

$$\frac{\Sigma \Sigma p |G_i - G_j|}{\mathfrak{N}^2} = \frac{\Sigma' \Sigma' p |G_i - G_j|}{n^2}, \frac{\Sigma \Sigma q |H_i - H_j|}{\mathfrak{N}^2} = \frac{\Sigma' \Sigma' q |H_i - H_j|}{n^2}, \ldots$$

Anche qui si osservi che se i caratteri X, Z, T,... sono tutti fra loro cograduati (intendendo di aver sostituito coi loro contrarî quelli ai quali competono coefficienti negativi nella relazione lineare data), e soltanto allora, i due membri di ciascuna delle ultime uguaglianze scritte sono nulli, e quindi la sola ipotesi $\Delta'_x = \Delta_x$, $\Delta'_z = \Delta_z$, $\Delta'_t = \Delta_t$, . . . porta a concludere che $\Delta_y = \Delta'_y$.

d) Come corollario delle osservazioni precedenti si ha:

Se un carattere Y è funzione lineare di più caratteri X, Z,... ed un campione è rappresentativo per quanto concerne sia il valore medio, sia la concentrazione di X, Z,..., affinchè esso lo sia anche per la concentrazione di Y sarà condizione sufficiente che:

$$\frac{\Sigma\Sigma p\,|G_i - G_j|}{\mathfrak{N}^2} = \frac{\Sigma'\Sigma' p\,|G_i - G_j|}{n^2},\quad \frac{\Sigma\Sigma q\,|H_i - H_j|}{\mathfrak{N}^2} = \frac{\Sigma'\Sigma' q\,|H_i - H_j|}{n^2},\ldots$$

Difatti, per l'ipotesi della linearità, il campione sarà senz'altro rappresentativo del valore medio di Y. Inoltre, poichè il rapporto di concentrazione è costituito dal quoziente fra la differenza media del carattere e il doppio della sua media aritmetica, il campione, essendo rappresentativo delle concentrazioni e dei valori medî di X, Z,..., lo sarà anche delle differenze medie di X, Z,... Ora se si verificano le condizioni espresse dalle uguaglianze scritte, il campione sarà anche rappresentativo della differenza media di Y, e quindi della sua concentrazione, *c. d. d.*

È ovvio che, nel caso di perfetta cograduazione di tutti i caratteri, per concludere il verificarsi della rappresentatività della concentrazione basta, oltre la supposta linearità, sapere che il campione è conservativo dei valori medî e delle concentrazioni di X, Z, T,...

Risulta pure evidente che se un carattere Y è funzione lineare di un solo carattere X, ed un campione è rappresentativo sia del valore medio, sia della concentrazione di X, lo sarà anche della concentrazione di Y.

APPENDICE A

Probabilità che lo scarto percentuale fra il valore medio di un carattere quantitativo nella totalità, costituita da K circondarî, e in un campione, costituito da k circondarî ($k \leqq K$) estratti a sorte, non superi ε.

1. Una totalità sia formata da K gruppi parziali (per es. « circondarî del Regno ») $C_1 \ldots C_i \ldots C_K$, i cui elementi siano rispettivamente in numero di $p_1 \ldots p_i \ldots p_K$ (per es. p_i « numero dei redditieri » del circondario C_i). In ciascuno di tali elementi si consideri l'intensità di un carattere quantitativo (per es. « reddito ») e siano

$$q_1 \ldots q_i \ldots q_K$$

i valori globali dello stesso carattere nei singoli gruppi (per es. q_i « ammontare totale dei redditi accertati » nel circondario C_i). Siano infine:

p la media aritmetica non ponderata dei p_i (per es. « numero medio dei redditieri » dei circondarî);

q la media aritmetica non ponderata dei q_i (per es. « reddito totale medio accertato » nei circondarî).

Allora il valore medio individuale A del carattere nella totalità (per es. « reddito medio individuale » nel Regno) sarà:

$$A = \sum_{i=1}^{K} q_i \Big/ \sum_{i=1}^{K} p_i = \frac{1}{K} \Sigma\, q_i \Big/ \frac{1}{K} \Sigma\, p_i = \frac{q}{p}\,.$$

Supposto ora di avere costituito un campione estraendo a sorte k fra i K gruppi parziali della totalità, denotando rispettivamente con a, $\sum_{i=1}^{k}{}' p_i$, $\sum_{i=1}^{k}{}' q_i$, p', q' i valori che si constatano nel campione, analoghi ai valori A, $\sum_{i=1}^{K} p_i$, $\sum_{i=1}^{K} q_i$, p, q constatati nella totalità, si avrà

$$a = \sum_{i=1}^{k}{}' q_i \Big/ \sum_{i=1}^{k}{}' p_i = \frac{1}{k} \Sigma'\, q_i \Big/ \frac{1}{k} \Sigma'\, q_i = \frac{q'}{p'}\,.$$

Infine si ponga

$$\frac{q'}{p'} = \frac{q + x}{p + y}\,.$$

Si tratta di determinare la probabilità Π affinchè a risulti affetto da un errore, rispetto ad A, che non superi ε, cioè la probabilità Π che si abbia

[1] $$\left| \frac{q+x}{p+y} - \frac{q}{p} \right| \leq \varepsilon \frac{q}{p}.$$

2. Supponiamo, in primo luogo, $p + y > 0$, cioè $y > -p$ e quindi $p^2 + py > 0$; dovrà allora essere

$$-\varepsilon q (p + y) \leq p x - q y \leq \varepsilon q (p + y).$$

La disuguaglianza costituita dalla seconda parte di questa limitazione fornisce:

[2] $$y \geq \frac{p x}{q(1+\varepsilon)} - \frac{p q \varepsilon}{q(1+\varepsilon)} = \alpha(x),$$

mentre dall'altra disuguaglianza si trae, notando che $|\varepsilon| < 1$:

[3] $$y \leq \frac{p x}{q(1-\varepsilon)} + \frac{p q \varepsilon}{q(1-\varepsilon)} = \beta(x).$$

Interpretando geometricamente il risultato, si potrà dire che, nell'ipotesi di $y > -p$, il problema enunciato è ridotto a determinare la probabilità affinchè gli errori x ed y da cui possono essere affette le medie q' e p' dedotte dai circondarî estratti a sorte, rispetto alle medie q e p dedotte dalla totalità dei circondarî, costituiscano le coordinate di punti che rispetto alle rette

[2′] $$y = \frac{p x}{q(1+\varepsilon)} - \frac{p q \varepsilon}{q(1+\varepsilon)}$$

[3′] $$y = \frac{p x}{q(1-\varepsilon)} + \frac{p q \varepsilon}{q(1-\varepsilon)}$$

cadano dalla stessa parte in cui si trova l'origine, o cadano anche sulle rette stesse. La regione in cui possono giacere tali punti è dunque l'angolo determinato da quelle due rette, e contenente l'origine: vertice V di tale angolo, cioè punto comune alle rette [2′] e [3′] è $(x = -q,\ y = -p)$.

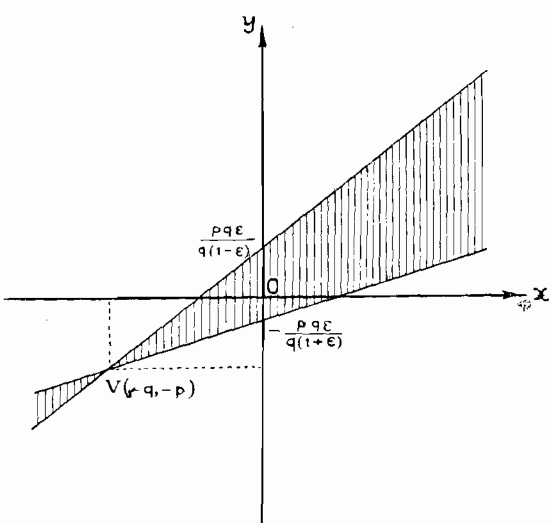


Fig. 1

Se $p + y < o$, cioè $y < -p$, si trova analogamente che i punti di coordinate x, y tali da soddisfare la [1], debbono cadere internamente o sui lati dell'angolo opposto al vertice a quello ora detto. La regione complessiva dei punti (x, y) pei quali si verifica la [1] è dunque quella rappresentata dalla figura della pagina precedente.

Pertanto, se $z(xy)\, dx\, dy$ è la probabilità di commettere, per estrazione a sorte di k gruppi parziali o circondarî, un errore compreso fra $x \pm \frac{1}{2} dx$ nella media q' rispetto alla q, e un errore compreso fra $y \pm \frac{1}{2} dy$ nella media p' rispetto alla p, la probabilità Π di verificare la [1] assumerà la forma:

$$\Pi = \int_{-\infty}^{-r} dx \int_{\beta(x)}^{\alpha(x)} z(xy)\, dy + \int_{-r}^{\infty} dx \int_{\alpha(x}^{\beta(x)} z(xy)\, dy. \tag{4}$$

Alla espressione di Π si può anche dare un'altra forma, che meglio si presta alla sua valutazione effettiva. A tal fine, notiamo che la equazione

$$\frac{q+x}{p+y} - \frac{q}{p} = m\frac{q}{p} \tag{5}$$

determina una retta appartenente al fascio delle [2'] e [3'], e, d'altronde, essa rappresenta l'insieme dei punti (x, y) ai quali corrisponde, per il valore medio individuale del carattere nel campione, un errore m rispetto al valore medio analogo nella totalità. Se dunque si denota con $\psi(m)$ la probabilità di commettere tale errore m (ossia il limite a cui tende il rapporto tra la probabilità che un punto (x, y) cada in un angolo avente come bisettrice la retta [5] e l'ampiezza di tale angolo, quando l'angolo stesso tende a zero) la probabilità cercata sarà data da:

$$\Pi = \left| \int_{-\varepsilon}^{\varepsilon} \psi(m)\, dm \right|. \tag{6}$$

3. Ora, poichè dalla totalità dei K gruppi si possono estrarre $\binom{K}{k}$ campioni costituiti da k gruppi, così per ottenere una espressione almeno approssimata di $\psi(m)$ sarà necessario conoscere l'equazione

$$z = z(x, y)$$

della superficie di correlazione inerente ai valori che le medie $\frac{1}{k}\Sigma' q_i$ e

$\frac{1}{k}\Sigma' p_i$ assumono in tutti quei campioni. Per es., nel caso che il carattere preso in esame sia il reddito, $\frac{1}{k}\Sigma' q_i$ denoterà la media del reddito globale accertato in ciascuno dei k circondarî di un campione, e $\frac{1}{k}\Sigma' p_i$ sarà la media dei numeri di redditieri nei circondarî dello stesso campione.

Assumiamo, in prima approssimazione, che quella correlazione sia normale, cosicchè detto r il coefficiente di correlazione fra i $\binom{K}{k}$ valori di $\frac{1}{k}\Sigma' p_i$ e di $\frac{1}{k}\Sigma' q_i$, detti σ_y e σ_x gli scostamenti quadratici medî del sistema di numeri $\frac{1}{k}\Sigma' p_i$ e del sistema $\frac{1}{k}\Sigma' q_i$, si avrà:

$$[7] \qquad z(x\,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}}\, e^{-\frac{1}{2(1-r^2)}\left\{\frac{x^2}{\sigma_x^2} - \frac{2rxy}{\sigma_x\sigma_y} + \frac{y^2}{\sigma_y^2}\right\}}$$

Si tratta, allora, di determinare i valori dei parametri r, σ_x, σ_y, e a tal fine si osservi che:

a) le medie aritmetiche di tutti i possibili $\binom{K}{k}$ valori di $\frac{1}{k}\Sigma' p_i$ e di $\frac{1}{k}\Sigma' q_i$ coincidono evidentemente con le medie p e q dei p_i e rispettivamente dei q_i;

b) i valori σ_x e σ_y sono legati agli scostamenti quadratici medî s_x ed s_y dei p_i e dei q_i dalle relazioni:

$$(8) \quad \sigma_x = \frac{1}{k}\sqrt{\frac{k(K-k)}{K-1}}\,.\,s_x\,, \quad \sigma_y = \frac{1}{k}\sqrt{\frac{k(K-k)}{K-1}}\,.\,s_y \text{ (1)};$$

c) il coefficiente di correlazione r fra i due sistemi di $\binom{K}{k}$ numeri

---

(1) Difatti, per quanto si riferisce per es. a σ_x, basta osservare che, avendo detto q la media aritmetica dei q_i, si potrà porre: $q_i = q + m_i$ (con $\Sigma\, m_i = 0$) e perciò in un campione generico formato da k gruppi sarà $\frac{1}{k}\Sigma' q_i = q + \frac{1}{k}\Sigma' m_i$, dove $\frac{1}{k}\Sigma' m_i$ costituisce lo scostamento di una

$\frac{1}{k}\Sigma' p_i$ e $\frac{1}{k}\Sigma' q_i$ ha lo stesso valore del coefficiente di correlazione fra i due sistemi di K numeri p_i e q_i (1).

Possediamo dunque tutti gli elementi necessarî per scrivere l'equazione [7], che peraltro giova trasformare in modo che le ellissi concentriche omotetiche di uguale probabilità, rappresentate dall'uguagliare ad una costante arbitraria l'esponente di e nella [7] e cioè le

$$\frac{x^2}{\sigma_x^2} - \frac{2\,r\,xy}{\sigma_x\,\sigma_y} + \frac{y^2}{\sigma_y^2} = cost.$$

---

media generica $\frac{1}{k}\Sigma' q_i$ dalla loro media complessiva q. Il momento secondo μ_2 dell'insieme dei numeri $\frac{1}{k}\Sigma' q_i$ rispetto alla loro media q è quindi definito dalla

$$(\alpha) \qquad \binom{K}{k}\mu_2 = \Sigma\left(\frac{1}{k}\Sigma' m_i\right)^2,$$

cioè dalla

$$k^2\binom{K}{k}\mu_2 = P\,\Sigma\, m_i^2 + Q\,\Sigma\, 2\, m_i\, m_j,$$

dove P e Q denotano rispettivamente i numeri di volte che ciascun termine della forma m_i^2 e ciascun termine della forma $2\,m_i\,m_i$ entra nel secondo membro della (α). È dunque:

$$P = \binom{K}{k}\frac{k}{K},\ Q = \binom{K}{k}\left\{\binom{k-1}{2} : \binom{K-1}{2}\right\},$$

e perciò

$$k^2\,\mu_2 = \frac{k}{K}\,\Sigma\, m_i^2 + \frac{2\,k\,(k-1)}{K\,(k-1)}\,m_i\,m_j,$$

da cui infine, ricordando che

$$\Sigma\, m_i^2 + 2\,\Sigma\, m_i\, m_j = (\Sigma\, m_i)^2 = 0,$$

risulta

$$k^2\,\mu_2 = \frac{k\,(K-k)}{K-1}\,s_x^2,$$

che fornisce la prima delle [8]. Analogamente si dimostra la seconda.

(1) Ponendo, come nella nota precedente,

$$(\beta) \qquad p_i = p + m_i,\ (\Sigma\, m_i = 0),$$

e similmente

$$(\gamma) \qquad q_i = q + n_i\ (\Sigma\, n_i = 0),$$

il coefficiente di correlazione fra i p_i ed i q_i $(i = 1, 2, \ldots K)$ è $\frac{\Sigma\, m_i\, n_i}{K\, s_x\, s_y}$ dove s_x ed s_y sono i soliti scostamenti quadratici medî.

Si tratta di calcolare il coefficiente di correlazione r fra il sistema di numeri $\frac{1}{k}\Sigma' p_i = p + \frac{1}{k}\Sigma' m_i$ e il sistema dei numeri $\frac{1}{k}\Sigma' q_i = q + \frac{1}{k}\Sigma' n_i$, ossia $r = \frac{\sum_1^N \left(\frac{1}{k}\Sigma' m_i \cdot \frac{1}{k}\Sigma' n_i\right)}{N\,\sigma_x\,\sigma_y}$, dove $N = \binom{K}{k}$

si mutino in cerchi concentrici, il che si ottiene ponendo:

$$[9] \quad \xi = \frac{1}{\sqrt{2\,(1+r)}}\left(\frac{x}{\sigma_x}+\frac{y}{\sigma_y}\right), \quad \eta = \frac{1}{\sqrt{2\,(1-r)}}\left(-\frac{x}{\sigma_x}+\frac{y}{\sigma_y}\right)$$

poichè risulta allora

$$\frac{1}{1-r^2}\left\{\frac{x^2}{\sigma_x^{\,2}}-\frac{2\,r\,xy}{\sigma_x\,\sigma_y}+\frac{y^2}{\sigma_y^{\,2}}\right\}=\xi^2+\eta^2$$

cosicchè la [7] si trasforma nella

$$[10] \qquad Z\,(\xi\ \eta) = \frac{1}{2\,\pi}\,e^{-\frac{1}{2}(\xi^2+\eta^2)},$$

---

e le Σ' denotano tutte le possibili somme di k termini tratti dai K della successione (β) e dalla successione (γ).

Allora, ricordando le (8), sarà $r = \dfrac{\sum\limits_1^N (\Sigma' m_i\,.\,\Sigma' n_i)}{N\,\dfrac{k\,(K-k)}{K-1}\,s_x\,s_y}$; d'altronde si trova che:

$$(\delta) \qquad \sum_1^N (\Sigma' m_i\,.\,\Sigma' n_i) = \binom{K-2}{k-2}\,\Sigma\, m_i\, n_j + \binom{K-1}{k-1}\,\Sigma\, m_i\, n_i$$

perchè ogni termine della forma $m_i\ n_j$ (di cui tanti sono i distinti quante sono le disposizioni dei K indici 1, 2, ... K a 2 a 2) si presenterà tante volte quante sono le combinazioni dei $K-2$ indici diversi da i e da j a $k-2$ a $k-2$; ed ogni termine della forma $m_i\ n_i$ (di cui tanti sono i distinti quante sono le disposizioni dei k soliti indici ad 1 ad 1) si presenterà tante volte quante sono le combinazioni dei $K-1$ indici diversi da i, a $k-1$ a $k-1$.

Ma

$$\binom{K-2}{k-2} = \frac{(K-2)\,!}{(k-2)\,!\,(K-k)\,!} = \frac{K\,!\,k\,(k-1)}{k\,!\,(K-k)\,!\,K\,(K-1)} = \binom{K}{k}\frac{k\,(k-1)}{K\,(K-1)};$$

$$\binom{K-1}{k-1} = \frac{(K-1)!}{(k-1)!\,(K-k)\,!} = \frac{K!\,k}{k\,!\,(K-k)\,!\,K} = \binom{K}{k}\frac{k}{K};$$

e perciò la (δ) si può scrivere:

$$\sum_1^N (\Sigma' m_i\ \Sigma' n_i) = N\left[\frac{k\,(k-1)}{K\,(K-1)}\,\Sigma\, m_i\, n_j + \frac{k}{K}\,\Sigma\, m_i\, n_i\right].$$

Ora si noti che

$$0 = \Sigma\, m_i\,\Sigma\, n_i = \Sigma\, m_i\, n_j + \Sigma\, m_i\, n_i$$

dove gli indici $i\,j$ costituiscono una delle generiche disposizioni dei K numeri 1, 2 . . . K, a 2 a 2; e che, per conseguenza,

$$\Sigma\, m_i\, n_j = -\,\Sigma\, m_i\, n_i\,,$$

onde risulterà

$$\sum_1^N (\Sigma' m_i\ \Sigma' n_i) = N\left[-\frac{k\,(k-1)}{K\,(K-1)}\,\Sigma\, m_i\, n_i + \frac{k}{K}\,\Sigma\, m_i\, n_i\right] = N\,\frac{k\,(K-k)}{K\,(K-1)}\,\Sigma\, m_i\, n_i\,.$$

E finalmente, sostituendo questa espressione al numeratore di r, verrà senz'altro:

$$r = \frac{\Sigma\, m_i\, n_i}{K\, s_x\, s_y}, \quad c\,.\,d\,.\,d\,.$$

equazione di una superficie di rotazione avente l'asse Z come asse di rotazione. Alle ellissi di uguale probabilità definite dalla [7] corrispondono per la [10] i cerchi di uguale probabilità:

$$\xi^2 + \eta^2 = R^2$$

e la fig. 1, data poc'anzi, si trasforma nel modo indicato dalla fig. 2, intendendo che $\bar{\xi}$, $\bar{\eta}$ siano le coordinate di V rispetto ai nuovi assi ξ, η (cioè le trasformate di $-p$ e di $-q$ mediante le [9]); che sia:

$$\bar{R} = \overline{OV} = \sqrt{\bar{\xi}^2 + \bar{\eta}^2};$$

e che S (ξ, η) sia un punto qualunque giacente sulla retta $V\,S$ che rispetto agli assi primitivi era rappresentata dalla equazione [5] e che è costituita, come sappiamo, dall'insieme dei punti (ξ, η) ai quali corrisponde, per il valore medio individuale del carattere nel campione, un errore m rispetto al valore medio analogo nella totalità.

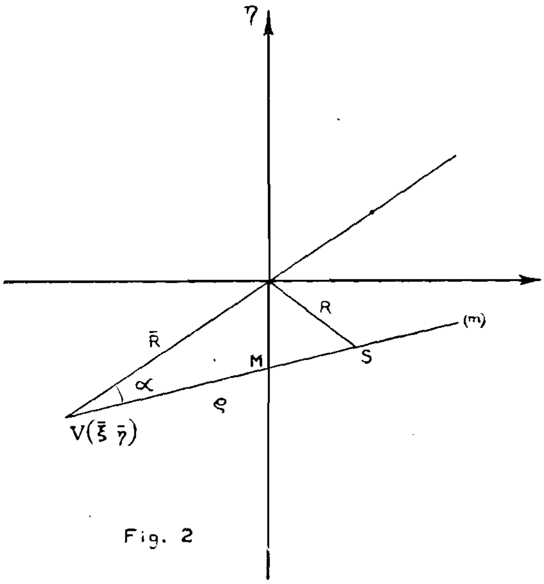


Fig. 2

4. Ciò posto, per $\alpha = \widehat{OVS}$ si ha subito, immaginando di imprimere al sistema di riferimento una traslazione parallela che porti l'origine in V:

$$[11] \quad \alpha = \operatorname{arctg} \frac{-\bar{\eta}}{-\bar{\xi}} - \operatorname{arctg} \frac{\eta - \bar{\eta}}{\xi - \bar{\xi}} = \operatorname{arctg} \frac{\bar{\eta}(\xi - \bar{\xi}) - \bar{\xi}(\eta - \bar{\eta})}{\bar{\xi}(\xi - \bar{\xi}) - \bar{\eta}(\eta - \bar{\eta})}.$$

Si tratta ora di esprimere α (o una sua funzione) mediante l'errore m corrispondente alla retta $V\,S$, e ciò si può ottenere nel modo più semplice considerando anzitutto la $V\,S$ come individuata da V e dal suo punto di intersezione M con l'asse η.

Perciò, facendo $\xi = O$, dalla [11] si dedurrà:

$$(11') \quad [\operatorname{tg}\alpha]_{\xi = o} = \frac{-\bar{\eta}\,\bar{\xi} - \bar{\xi}\,(\eta - \bar{\eta})}{-\bar{\xi}^2 + \bar{\eta}\,(\eta - \bar{\eta})} = \frac{-\bar{\xi}\,\eta}{\bar{\eta}\,\eta - (\bar{\xi}^2 + \bar{\eta}^2)} = \frac{-\bar{\xi}\,\eta}{\bar{\eta}\,\eta - \bar{R}^2}$$

Dopo ciò si osservi che per $\xi = o$ la prima delle [9] fornisce:

$$\frac{x}{\sigma_x} = \frac{-y}{\sigma_y},$$

cosicchè sostituendo questa espressione di x nella [5] si ricava:

$$y\left[\frac{p}{q}\frac{\sigma_x}{\sigma_y} + 1 + m\right] = -mp,$$

da cui

$$y = \frac{-mp}{\left(\frac{p}{q}\frac{\sigma_x}{\sigma_y} + 1\right) + m} = \frac{-mp}{a+m},$$

quando si intenda che $a = \frac{p}{q}\frac{\sigma_x}{\sigma_y} + 1$.

Ne segue, per la seconda delle [9]:

[12] $$\eta = \frac{1}{\sqrt{2(1-r)}}\frac{2y}{\sigma_y} = \frac{-2}{\sqrt{2(1-r)}\,\sigma_y}\cdot\frac{mp}{a+m} = \frac{cm}{a+m},$$

avendo posto

$$c = -\frac{2p}{\sigma_y\sqrt{2(1-r)}}.$$

Le coordinate di M sono dunque:

$$\xi = O \text{ ed } \eta = \frac{cm}{a+m},$$

e perciò la cercata espressione dell'angolo α in funzione di m è:

[13] $$\operatorname{tg}\alpha = \frac{-\bar{\xi}\dfrac{cm}{a+m}}{\bar{\eta}\dfrac{cm}{a+m} - \bar{R}^2} = \frac{-\bar{\xi}\,cm}{\bar{\eta}\,cm - \bar{R}^2 m - \bar{R}^2 a} = \frac{\lambda m}{\beta + \gamma m},$$

con

$$\lambda = -\bar{\xi}c, \quad \beta = -\bar{R}^2 a, \quad \gamma = \bar{\eta}c - \bar{R}^2.$$

Inoltre, indicando con R la lunghezza OS e con ρ la distanza VS, dalla fig. 2 risulterà:

$$R^2 = \overline{R}^2 + \rho^2 - 2\overline{R}\,\rho \cos\alpha = (\rho - \overline{R}\cos\alpha)^2 + \overline{R}^2 \operatorname{sen}^2\alpha$$

e poichè la [10] si può scrivere

$$Z(\xi, \eta) = \frac{1}{2\pi} e^{-\frac{1}{2}R^2}$$

ne verrà:

$$[14] \qquad Z(\xi, \eta) = \frac{1}{2\pi} e^{-\frac{1}{2}(\rho - \overline{R}\cos\alpha)^2} \cdot e^{-\frac{1}{2}\overline{R}^2 \operatorname{sen}^2\alpha}$$

che costituisce l'equazione della medesima superficie di correlazione [10], quando ogni punto S del piano (ξ, η) si pensi individuato dalle coordinate polari (α, ρ). L'elemento dell'integrale da calcolarsi, in coordinate polari, è dunque:

$$Z(\xi, \eta)\, \rho\, d\alpha\, d\rho\,;$$

e perciò eseguendo per prima l'integrazione in ρ avremo:

$$\int_0^\infty \frac{1}{2\pi} e^{-\frac{1}{2}\overline{R}^2 \operatorname{sen}^2\alpha} \cdot e^{-\frac{1}{2}(\rho - \overline{R}\cos\alpha)^2} \rho\, d\rho = \varphi(\alpha),$$

da cui ponendo $\rho - \overline{R}\cos\alpha = t$, cioè $\rho = t + \overline{R}\operatorname{sen}\alpha$,

$$\varphi(\alpha) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\overline{R}^2 \operatorname{sen}^2\alpha} \left[\int_{-\overline{R}\cos\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}\, t.\, dt + \overline{R}\cos\alpha \int_{-\overline{R}\cos\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}\, dt \right].$$

Siamo così ridotti a dovere eseguire integrazioni di funzioni note: e precisamente, poichè:

$$\int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt = \theta(x)\ (1) \qquad \int_x^\infty e^{-\frac{1}{2}t^2}\, t\, dt = e^{-\frac{1}{2}x^2}$$

viene senz'altro:

$$[15] \quad \varphi(\alpha) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\overline{R}^2 \operatorname{sen}^2\alpha} \left[\frac{e^{-\frac{1}{2}\overline{R}^2\cos^2\alpha}}{\sqrt{2\pi}} + \overline{R}\cos\alpha \left\{ \frac{1}{2} + \theta(\overline{R}\cos\alpha) \right\} \right]$$

(1) Il cui valore si può ricavare dalle note Tavole del PEARSON.

Si dovrebbe ora eseguire l'integrazione in α, cioè calcolare:

$$\left[\int_a^b \varphi(\alpha)\, d\alpha\right]$$

dove a e b denotano i valori di α corrispondenti, per la [13], agli estremi $-\varepsilon$ e $+\varepsilon$ dell'integrale [6], ma poichè

$$\varphi\Big(\alpha\,(m)\Big)\frac{d\,\alpha}{dm}\,dm = \psi\,(m)\,dm \quad (1)$$

così la funzione sotto il segno dell'integrale [6] è:

$$\psi\,(m) = \varphi\,\Big(\alpha\,(m)\Big)\,\frac{d\,\alpha}{dm},$$

dove è da intendersi che α sia funzione di m definita mediante la [13]; e siccome di qui risulta che

$$\frac{d\,\alpha}{dm} = \frac{d\mathrm{tg}\alpha}{dm} : \frac{d\mathrm{tg}\alpha}{d\alpha} = \left(\frac{d}{dm}\,\frac{\lambda\, m}{\beta+\gamma\, m}\right) : \left(1 + \left(\frac{\lambda\, m}{\beta+\gamma\, m}\right)^2\right) =$$

$$= \frac{\lambda\,\beta}{(\beta+\gamma\, m)^2 + \lambda^2 + m^2},$$

mentre, per la [13] stessa, è

$$\cos\alpha = \frac{1}{\sqrt{1+\mathrm{tg}^2\alpha}} = \frac{1}{\sqrt{1+\left(\dfrac{\lambda\, m}{\beta+\gamma\, m}\right)^2}},$$

$$\operatorname{sen}\alpha = \frac{\mathrm{tg}\,\alpha}{\sqrt{1+\mathrm{tg}^2\,\alpha}} = \frac{\dfrac{\lambda\, m}{\beta+\gamma\, m}}{\sqrt{1+\left(\dfrac{\lambda\, m}{\beta+\gamma\, m}\right)^2}},$$

$$\operatorname{sen}^2\alpha = \frac{\lambda^2\, m^2}{(\beta+\gamma\, m)^2 + \lambda^2\, m^2},$$

---

(1) A m corrisponde, oltre il valore α, anche $\alpha+\pi$. Invece di $\varphi\ (\alpha)$ dovremmo prendere $\varphi\ (\alpha) + \varphi\ (\alpha+\pi)$; trascurando il secondo termine si commette però un errore minore in valore assoluto a $\frac{1}{\pi} e^{-\frac{1}{2}\bar{R}^2}$. Nell'applicazione che segue tale errore è effettivamente senza influenza sulle cifre decimali fino a cui è spinto il calcolo.

così verrà

$$\psi(m) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\bar{R}^2 \frac{\lambda^2 m^2}{(\beta+\gamma m)^2+\lambda^2 m^2}} \left[\frac{e^{-\frac{1}{2}\bar{R}^2 \cos^2 \alpha}}{\sqrt{2\pi}} + \right.$$

$$\left. + \bar{R} \frac{\theta\left(\bar{R}\cos\alpha + \frac{1}{2}\right)}{\sqrt{1+\frac{\lambda^2 m^2}{(\beta+\gamma m)^2}}} \right] \frac{\lambda\beta}{(\beta+\gamma m)^2+\lambda^2 m^2}$$

e finalmente, tenendo conto della circostanza che al crescere di m l'angolo α decresce, si concluderà che:

$$[16] \qquad \Pi = \left[\int_{-\varepsilon}^{\varepsilon} \psi(m)\, dm\right] = \int_{-\varepsilon}^{\varepsilon} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\bar{R}^2 \frac{\lambda^2 m^2}{(\beta+\gamma m)^2+\lambda^2 m^2}}$$

$$\left[\frac{e^{-\frac{1}{2}\bar{R}^2 \cos^2}}{\sqrt{2\pi}} + \bar{R} \frac{\theta\left(\bar{R}\cos\alpha + \frac{1}{2}\right)}{\sqrt{1+\frac{\lambda^2 m^2}{(\beta+\gamma m)^2}}} \right] \frac{[\lambda\beta]}{(\beta+\gamma m)^2+\lambda^2 m^2}\, d\,m$$

5. Come applicazione di questa formula cerchiamo la probabilità Π affinchè il reddito medio dedotto da un gruppo di 29 circondarî, presi a caso fra i 214, sia affetto da un errore non maggiore dell '1,5 % rispetto al reddito medio dedotto dalla totalità, non dimenticando, peraltro, che il valore che andremo a trovare per Π avrà significato soltanto in prima approssimazione, e cioè assumendo come normale la correlazione fra i valori $\frac{1}{k}\Sigma' p_i$ ed i valori $\frac{1}{k}\Sigma' q_i$.

I valori dei parametri che, via via, intervengono nel calcolo sono:

$$K = 214;\ k = 29;\ \frac{1}{k}\sqrt{\frac{k(K-k)}{K-1}} = 0{,}1731;$$

media aritmetica dei q_i è : 16.404,74 $= q$;
id. id. dei p_i è : 3.950,46 $= p$;
scostamento quadratico medio dei q_i è : $s_x =$ 32.793,46 ;
id. id. id. dei p_i è : $s_y =$ 14.108,00 ;

coefficiente di correlazione r fra i p_i ed i q_i è 0,3002 ;

id. id. fra i $\frac{1}{29} \Sigma' p_i$ ed i $\frac{1}{29} \Sigma' q_i$ è $0{,}3002 = r$;

scostamento quadratico medio dei valori $\frac{1}{29} \Sigma' q_i$ è : $\sigma_x = 5675{,}24$;

id. id. id. $\frac{1}{29} \Sigma' p_i$ è : $\sigma_y = 2441{,}53$;

$$a = 1{,}560 ; \frac{1}{\sqrt{2(1-r)}} = 0{,}845 ; \frac{1}{\sqrt{2(1+r)}} = 0{,}620 ;$$

$$c = -3{,}828 ; \bar{\xi} = 2{,}796 ; \bar{\eta} = -1{,}076 ;$$

distanza del vertice V dall'origine $\overline{R} = 2{,}996$;

$$\lambda = 10{,}704 ; \beta = -4{,}673 ; \gamma = -4{,}856.$$

Due limiti fra i quali sia compreso Π si avranno moltiplicando l'intervallo d'integrazione $0{,}015 - (-0{,}015) = 0{,}030$ per i valori massimo e minimo della funzione integranda $\psi(m)$ nello stesso intervallo, compresi gli estremi. Evidentemente il valore massimo di $\psi(m)$ si ha per $m = 0$, ed è quindi :

$$\psi(o) = \frac{1}{\sqrt{2\pi}} \left(\frac{e^{-\frac{1}{2}\overline{R}^2}}{\sqrt{2\pi}} + \overline{R}. \left\{ \frac{1}{2} + \theta(\overline{R}) \right\} \right) \left| \frac{\lambda}{\beta} \right| = 2{,}738,$$

mentre il valore minimo di $\psi(m)$ ha luogo per $m = -0{,}015$, ed è:

$$\psi(-0{,}015) = 2{,}635.$$

Dunque :

$$\psi(-0{,}015).\ 0{,}03 < \Pi < \psi(0).\ 0{,}03,$$

ossia :

$$\Pi_1 = 0{,}079 < \Pi < 0{,}082 = \Pi',$$

ed anche, tenendo conto della convessità della $\psi(m)$:

$$\Pi > \frac{\Pi_1 + \Pi'}{2} = 0{,}080.$$

Si conclude che :

$$\Pi = 0{,}081 \text{ circa.}$$

APPENDICE B

Sul valore medio di caratteri diversi da quelli strumentali della scelta.

1. Dedotto da un complesso di circoscrizioni un gruppo parziale *il quale sia conservativo dei valori medî di alcuni caratteri u, v, w...*, uno dei problemi (inversi) più ovvî che si presentano, quando si voglia utilizzare quel campione è il seguente: dato che per un certo carattere x, distinto dai precedenti, risulti dal campione un certo valore medio, che cosa si potrà inferire circa il valore medio dello stesso carattere nella totalità ?

Ora il Bowley (1) cerca appunto di determinare la precisione con la quale la media ponderata di x, tratta dal campione, rappresenta il valore medio dello stesso carattere nella totalità, e lo fa assumendo le seguenti ipotesi: 1° che la ponderazione per la determinazione di tutte le medie, sia dei caratteri strumentali u, v, w, sia del nuovo carattere x, sia eseguita con gli stessi pesi (per es. popolazioni dei circondarî); 2° che l'equazione di regressione parziale collegante x ad u, v, w... sia lineare con sufficiente approssimazione; 3° che il campione scelto sia conservativo della media di u, di v, di w..; 4° che i coefficienti di correlazione $r_{xu}, r_{xv}, r_{xw}, \ldots$, fra il carattere x e gli u, v, w . . ., dedotti dal campione differiscano di poco dai corrispondenti dedotti dalla totalità; 5° che le differenze fra le medie ponderate e non ponderate di x, u, v, w siano piccole.

La prima di tali ipotesi esprime che il metodo in questione è applicabile ad un carattere x soltanto nel caso che esso sia ottenuto come media di quantità fondamentali mediante lo stesso sistema di ponderazioni che serve a definire i caratteri strumentali u, v, w... Così, per es., se si assume: u = natalità, v = percentuale di popolazione agglomerata, w = altitudine media, caratteri che si ottengono prendendo come pesi le popolazioni totali dei singoli circondarî, il metodo sarà applicabile (almeno per questo riguardo) al carattere x = accrescimento naturale, che dipende dagli

(1) Rapporto citato, pag. 47.

stessi pesi, ma non, per es., al carattere y = reddito medio, che si ottiene prendendo altri pesi, e cioè i numeri dei redditieri dei singoli circondarî.

La seconda ipotesi non sembra altrettanto essenziale per lo sviluppo analitico, in quanto gli errori risultanti dal calcolare i valori di x mediante l'equazione di regressione, non vengono trascurati, ma entrano in calcolo.

La terza ipotesi significa semplicemente che il campione deve essere rappresentativo della totalità per quanto concerne i valori medî dei caratteri u, v, w...

Finalmente la quarta e la quinta ipotesi appaiono essere quelle più difficili a realizzarsi, per la buona riuscita del metodo: l'ammettere (quinta ipotesi) che siano trascurabili le differenze fra le medie ponderate e non ponderate dei caratteri nella totalità, e soprattutto l'ammettere (quarta ipotesi) che siano trascurabili le differenze fra i coefficienti di correlazione r_{xu}, r_{xv}, r_{xw} . . . dedotti dal campione e gli omologhi ρ_{xu}, ρ_{xv}, ρ_{xw} dedotti dalla totalità, equivale infatti a presupporre nel campione un complesso di requisiti estremamente improbabile a verificarsi; e tanto più quando si tenga presente che la conservazione dei valori medî di u, v, w, non implica assolutamente quella dei mutui coefficienti di correlazione, come abbiamo anche verificato nella nostra applicazione pratica.

2. Comunque, abbiamo voluto applicare l'analisi del Bowley al nostro caso; e poichè è necessario prendere le mosse dalla supposizione che le medie dei caratteri strumentali della scelta siano le stesse nella totalità e nel campione, così per non allontanarci troppo da questa supposizione, abbiamo ammesso di avere eseguito la scelta impiegando come caratteri strumentali soltanto tre di quelli pei quali le medie rilevate nella totalità e nel campione presentano i minimi scarti assoluti (1) e cioè la *natalità*, la *percentuale di popolazione agglomerata* e l'*altitudine*. Del resto, poichè la precisione del risultato si accresce lievemente coll'aumentare il numero dei caratteri strumentali (2), così l'omissione degli altri caratteri non influirà notevolmente sul calcolo che vogliamo eseguire.

Supporremo poi che il nuovo carattere, per il cui valore medio nella totalità vogliamo cercare dei limiti di approssimazione mediante il valore medio constatato nel campione, sia l'*accrescimento naturale:* poichè sap-

---

(1) Nella presente applicazione conviene considerare gli scarti assoluti fra i valori medî e non gli scarti relativi.

(2) Cfr. Bowley, Rapporto citato, pag. 50.

piamo che l'accrescimento naturale medio nel campione è 12,95, che cosa si potrà dire circa il valore medio di tale carattere nella totalità?

Gli elementi del problema risultano dalla seguente tabella:

Simboli	QUANTITÀ FONDAMENTALI E CARATTERI	Per il Regno	Per il gruppo dei circondarî scelti
a	Popolazione	37.142.886	5.724.103
u	Natalità	30,23 ‰	30,54 ‰
v	Popolazione agglomerata su 100 abitanti	73,75	74,69
w	Altitudine	171,35	170,49
x	Accrescimento naturale	X	12,95 ‰ $= X_{29}$

Osserviamo subito che la media non ponderata dell'accrescimento naturale nel campione è $\bar{x}_{29} = 14{,}06$ e quindi $X_{29} - \bar{x}_{29} = 1,\ 11$; si prevede dunque che se questo notevole scarto (rilevato nel campione) non sarà neutralizzato dalle differenze analoghe che si verificano nella totalità per i caratteri strumentali u, v, w, il risultato del calcolo non potrà essere soddisfacente.

Occorre poi conoscere: i coefficienti di correlazione fra x ed u, v, w, tratti, naturalmente, dal campione, i quali, come si rileva dalla Tav. 33, sono:

$$r_{xu} = 0{,}916 \qquad r_{xv} = -0{,}383 \qquad r_{xw} = 0{,}099;$$

i coefficienti di correlazione dei caratteri strumentali fra di loro, dedotti dalla totalità dei circondarî, e cioè (vedi stessa tavola):

$$\rho_{uv} = -0{,}065 \qquad \rho_{uw} = 0{,}165 \qquad \rho_{vw} = 0{,}148;$$

lo scostamento quadratico medio σ_x di x nel gruppo dei circondarî scelti, che è

$$\sigma_x = 4{,}15;$$

gli scostamenti quadratici medi σ_a σ_u σ_v σ_w della popolazione e dei tre caratteri strumentali nella totalità dei circondarî, scostamenti che risultano essere:

$$\sigma_a = 157.848{,}55$$

$$\sigma_u = 9{,}63 \qquad \sigma_v = 16{,}92 \qquad \sigma_w = 278{,}63.$$

Allora, avendo ammesso che l'equazione di regressione collegante il carattere x coi caratteri u, v, w, sia lineare con sufficiente approssimazione, cioè che sia

$$x - \bar{x} = G_u(u - \bar{u}) + G_v(v - \bar{v}) + G_w(w - \bar{w}),$$

dove $\bar{x}, \bar{u}, \bar{v}, \bar{w}$ denotano le medie non ponderate dei caratteri x, u, v, w nella totalità dei circondarî, il BOWLEY dimostra che

$$X = X_{29} - K \pm \sigma_{29},$$

essendo X l'accrescimento naturale medio che si cerca, X_{29} il valore medio dedotto dal gruppo dei 29 circondarî scelti,

$$K = -(X - \bar{x}) + G_u(U - \bar{u}) + G_v(V - \bar{v}) + G_w(W - \bar{w}) \text{ (circa)}$$

$$\sigma_{29} = \frac{\sigma_x}{29} \sqrt{1 + \frac{\sigma_a^2}{\bar{a}^2}} \sqrt{\frac{R}{R_{11}}}.$$

In queste ultime uguaglianze il significato dei simboli è:

$$R = \begin{vmatrix} 1 & r_{xu} & r_{xv} & r_{xw} \\ r_{xu} & 1 & \rho_{uv} & \rho_{uw} \\ r_{xv} & \rho_{uv} & 1 & \rho_{vw} \\ r_{xw} & \rho_{uw} & \rho_{vw} & 1 \end{vmatrix}$$

R_{11}, R_{12}, R_{13}, R_{14} sono ordinatamente i reciproci degli elementi della prima linea di questo determinante, U, V, W sono le medie dei caratteri strumentali, dedotte dalla totalità dei circondarî (e coincidenti per ipotesi con quelle dedotte dal gruppo scelto) e cioè:

$$U = 30{,}23 \qquad V = 73{,}75 \qquad W = 171{,}35;$$

$\bar{a}$ è la media delle popolazioni di tutti i circondarî, cioè:

$$\bar{a} = 173.564{,}89.$$

X ed $\bar{x}$, che indicano la media ponderata e quella non ponderata di x nella totalità, vengono sostituiti con gli analoghi valori rilevati dal campione, e cioè con

$$X_{29} = 12{,}95 \qquad \bar{x}_{29} = 14{,}06.$$

Inoltre

$$\bar{u} = 31{,}12 \qquad \bar{v} = 72{,}65 \qquad \bar{w} = 231{,}95.$$

Infine la determinazione di G_u, G_v, G_w si effettua mediante le

$$\frac{-\sigma_x}{R_{11}} = \frac{G_u \sigma_u}{R_{12}} = \frac{G_v \sigma_v}{R_{13}} = \frac{G_w \sigma_w}{R_{14}} .$$

Eseguiti i calcoli si trova:

$$R_{11} = 0{,}943 \quad R_{12} = 0{,}844 \quad R_{13} = 0{,}306 \quad R_{14} = -0{,}001 \quad R = 0{,}053.$$

$$G_u = 0{,}382 \qquad G_v = -0{,}080 \qquad G_w = 0{,}000$$

$$K = 0{,}68$$

$$\sigma_{29} = 0{,}046$$

$$X = 12{,}95 - 0{,}68 \pm 0{,}05 \begin{cases} 12{,}32 \\ 12{,}22 \end{cases}$$

I limiti trovati *non comprendono* il vero valore di X che è 12,70 (1).

3. Riprendendo lo stesso problema trattato dal BOWLEY, procuriamo di risolverlo in base ad un sistema di ipotesi più semplici e più plausibili. *In particolare prescinderemo dall'ipotesi che i coefficienti di correlazione fra i caratteri considerati abbiano valori praticamente uguali nella totalità e nel campione*, condizione che abbiamo veduto non essere — in generale — verificata, neanche quando le medie dei caratteri strumentali siano bene conservate passando dalla totalità al campione. Aggiungiamo, anzi, che il presupporre senz'altro una stretta concordanza fra i valori dei coefficienti di correlazione nella totalità e nel gruppo scelto, costituisce forse una condizione assai più restrittiva di quanto non sia l'ammettere senz'altro la conservazione del valore medio del nuovo carattere.

Siano $u^{(1)}$, $u^{(2)} \ldots u^{(t)}$ i caratteri strumentali della scelta; $\overline{u^{(1)}}$ $\overline{u^{(2)}} \ldots$ $\overline{u^{(t)}}$ i rispettivi valori medî nel gruppo dei k circondarî scelti fra la totalità dei K circondarî; $u_i^{(1)}$, $u_i^{(2)}$, $\ldots$ $u_i^{(t)}$ i valori degli stessi caratteri nel circondario i^{mo}.

Sia x un nuovo carattere (diciamo « di verificazione ») di cui sono noti i valori x_i *nei soli circondarî scelti*. Ammettiamo anzitutto (prima ipotesi) che tutti questi caratteri siano definiti come medie ponderate di quantità fondamentali, assumendo uno stesso sistema di pesi (nel nostro caso: popolazione dei circondarî).

---

(1) Si noti che se l'intervallo di previsione di X fosse valutato omettendo $K = 0{,}68$ risulterebbe $X = 12{,}95 \pm 0{,}05 = \begin{cases} 13, \\ 12{,}90 \end{cases}$ cioè esso sarebbe più prossimo al vero valore di X di quanto non lo sia l'intervallo calcolato col tenere conto di K.

Consideriamo quella funzione lineare di $u^{(1)}\ u^{(2)} \dots u^{(t)}$

$$f\,(u^{(1)},\ u^{(2)},\ \dots\ u^{(t)})$$

per cui risulti minima la somma ponderata rispetto alle popolazioni dei quadrati delle differenze fra i valori di x e di f nei circondarî scelti, ossia quella funzione f per cui

$$\sum_{i=1}^{k} a_i\,(x_i - f_i)^2 = Min\,.$$

essendo a_i la popolazione del circondario i^{mo} ed

[1] $$f_i = f\,(u_i^{(1)},\ u_i^{(2)},\ \dots\ u_i^{(t)}).$$

Una generica funzione lineare delle $u^{(1)}$, $u^{(2)}$, ... $u^{(t)}$ è:

[1′] $$\bar{x} + b_1\left(u^{(1)} - \overline{u^{(1)}}\right) + b_2\left(u^{(2)} - \overline{u^{(2)}}\right) + \dots + b_t\left(u^{(t)} - \overline{u^{(t)}}\right)$$

e si tratta, pertanto, di determinare i parametri $\bar{x}$, b_1, b_2, ... b_t in tal modo da rendere minimo

$$\varphi\,(\bar{x}, b_1, b_2, \dots b_t) = \sum_{i=1}^{k} a_i\,(x_i - f_i)^2 = \sum_{i=1}^{k} a_i \left\{ x_i - \bar{x} - \left[b_1\,(u_i^{(1)} - \overline{u^{(1)}}) + \right.\right.$$
$$\left.\left. + \dots + b_t\,(u_i^{(t)} - \overline{u^{(t)}}) \right] \right\}^2$$

per il che si richiede

$$\frac{\partial \varphi}{\partial \bar{x}} = -2 \sum_{i=1}^{k} a_i \left\{ x_i - \bar{x} - [\dots] \right\} = 0,\ \frac{\partial \varphi}{\partial b_1} = 0,\ \dots\ \frac{\partial \varphi}{\partial b_t} = 0.$$

La $\dfrac{\partial \varphi}{\partial \bar{x}} = 0$, essendo per definizione $\sum_{i=1}^{k} a_i\,u_i^{(h)} = \overline{u^{(h)}} \sum_{i=1}^{k} a_i$,

dà :

$$\sum_{1}^{k} a_i\,x_i = \bar{x} \sum_{1}^{k} a_i\,,\quad \text{ossia } \bar{x} = \frac{\sum a_i\,x_i}{\sum a_i}\,,$$

cioè $\bar{x}$ è la media (ponderata) del carattere x nel campione.

Dalle prime delle altre condizioni, ponendo:

$$z = x - \bar{x}\,,\quad v^{(h)} = u^{(h)} - \overline{u^{(h)}}$$

si trae :

$$\frac{1}{2}\frac{\partial \varphi}{\partial b_1} = \frac{1}{2}\frac{\partial}{\partial b_1} \sum_{1}^{k} a_i \left\{ z_i - b_1\,u_i^{(1)} - \dots - b_t\,u_i^{(t)} \right\}^2$$

$$= - \sum_{1}^{k} a_i \left\{ z_i - b_1\,u_i^{(1)} - \dots - b_t\,u_i^{(t)} \right\} u_i^{(1)}$$

$$= - \sum_{1}^{k} a_i\,z_i\,u_i^{(1)} + b_1 \sum_{1}^{k} a_i\,u_i^{(1)^2} + b_2 \sum_{1}^{k} a_i\,u_i^{(2)}\,u_i^{(1)} + \dots + b_t \sum a_i\,u_i^{(t)}\,u_i^{(1)} = 0$$

da cui, indicando con r_{1x} l'indice di correlazione nel campione fra i caratteri x ed $u^{(1)}$, e similmente con r_{12} l'indice di correlazione fra $u^{(1)}$ ed $u^{(2)}$, etc. *qualora si prenda come unità l'individuo e non il circondario*, e indicando altresì con σ_x, σ_1, σ_2, ... σ_t gli scostamenti quadratici medî di $x, u^{(1)}, u^{(2)}, \ldots u^{(t)}$ nel campione stesso, *come costituito di individui*, si ha:

$$\frac{1}{2}\frac{\partial \varphi}{\partial b_1} = -r_{1x}\sigma_1\sigma_x\sum_1^k a_i + b_1\sigma_1^2\sum_1^k a_i + b_2 r_{12}\sigma_1\sigma_2\sum_1^k a_i + \ldots + b_t r_{1t}\sigma_1\sigma_t\sum_1^k a_i$$

$$= -k a\,(r_{1x}\sigma_1\sigma_x - b_1\sigma_1^2 - b_2 r_{12}\sigma_1\sigma_2 - \ldots - b_t r_{1t}\sigma_1\sigma_t) = 0$$

dove si è posto $\sum_1^k a_i = ka$, essendo a la popolazione media dei circondarî scelti.

Tenendo dunque conto di tutte le condizioni di minimo si ha il sistema di $t+1$ equazioni:

$$[2] \qquad \begin{cases} \varphi = k a \sigma_x (\sigma_x - b_1\sigma_1 r_{1x} - b_2\sigma_2 r_{2x} - \ldots - b_t\sigma_t r_{tx}) \\ 0 = k a \sigma_1 (r_{1x}\sigma_x - b_1\sigma_1 - b_2\sigma_2 r_{12} - \ldots - b_t\sigma_t r_{1t}) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ 0 = k a \sigma_t (r_{tx}\sigma_x - b_1\sigma_1 r_{1t} - b_2\sigma_2 r_{1,t-1} - \ldots - b_t\sigma_t \quad) \end{cases}$$

dalle quali ultime t, che costituiscono un sistema omogeneo rispetto alle incognite

$$\sigma_x, b_1\sigma_1, b_2\sigma_2, \ldots b_t\sigma_t$$

si deduce, ponendo:

$$R = \begin{vmatrix} 1 & -r_{1x} & -r_{2x} & \ldots & -r_{tx} \\ r_{1x} & -1 & -r_{12} & \ldots & -r_{1t} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ r_{tx} & -r_{1t} & -r_{1,t-1} & \ldots & -1 \end{vmatrix}$$

e indicando, come sempre, con R_{11}, R_{12}, ... $R_{1,t+1}$ i minori degli elementi della prima linea:

$$[3] \qquad \frac{-\sigma_x}{R_{11}} = \frac{b_1\sigma_1}{R_{12}} = \ldots = \frac{b_t\sigma_t}{R_{1,t+1}},$$

mentre la prima delle equazioni fornisce:

$$\sigma_x = \frac{R_{11}}{R}\frac{\varphi}{k a \sigma_x},$$

ossia:

$$[4] \qquad \varphi = \frac{R}{R_{11}}\sigma_x^2 k a.$$

L'errore quadratico medio di x constatato nel campione, *considerato come insieme di individui*, è dunque:

$$\sqrt{\frac{\varphi}{k\,a}} = \sigma_x \sqrt{\frac{R}{R_{11}}}\,.$$

Applicando la formula di regressione [1] a tutti i singoli circondarî, appartenenti o no al campione considerato, e dicendo e_i l'errore corrispondente al circondario i^{mo} si avrà

$$x_i = f_i + e_i = \bar{x} + b_1\left(u_i^{(1)} - \overline{u^{(1)}}\right) + b_2\left(u_i^{(2)} - \overline{u^{(2)}}\right) + \ldots + b_t\left(u_i^{(t)} - \overline{u^{(t)}}\right) + e^i$$

cosicchè il valore medio X del carattere x nella totalità sarà:

$$\frac{\sum_1^K a_i x_i}{\sum_1^K a_i} = \bar{x} + b_1\left(\frac{\sum_1^K a_i u_i^{(1)}}{\sum_1^K a_i} - \overline{u^{(1)}}\right) + \ldots + b_t\left(\frac{\sum_1^K a_i u_i^{(t)}}{\sum_1^K a_i} - \overline{u^{(t)}}\right) + \frac{\sum_1^K a_i e_i}{\sum_1^K a_i}$$

cioè:

$$[5] \qquad X = \bar{x} + b_1\left(U^{(1)} - \overline{u^{(1)}}\right) + \ldots + b_t\left(U^{(t)} - \overline{u^{(t)}}\right) + \frac{\sum_1^K a_i e_i}{\sum_1^K a_i}$$

essendo evidentemente $U^{(1)}, \ldots U^{(t)}$ i valori medî di $u^{(1)}, \ldots u^{(t)}$ nella totalità.

Il problema è così ridotto allo studio dell'errore:

$$E = \frac{\sum_1^K a_i e_i}{\sum_1^K a_i} = \frac{\sum_1^k a_i e_i + \sum_{k+1}^K a_i e_i}{\sum_1^K a_i} = \frac{\sum_{k+1}^K a_i e_i}{\sum_1^K a_i}\,,$$

dove la $\sum_1^k$ è estesa all'insieme dei circondarî scelti; cercheremo il valore probabile di E^2.

a) Supponiamo (seconda ipotesi) che il valore probabile del quadrato dell'errore di x in ciascuno dei circondarî non scelti, quando si applichi l'equazione di regressione, sia lo stesso, e indichiamolo con:

$$\sigma_e^2 = M\left(e_i^2\right) \qquad (i = k+1, k+2, \ldots K);$$

ammettiamo inoltre (terza ipotesi) che tutti questi errori siano indipendenti; avremo allora (1):

$$[6] \qquad \sigma_E^{\,2} = M\,(E^2) = M\left(\frac{\sum\limits_{k+1}^{K} a_i\,e_i}{\sum\limits_{1}^{K} a_i}\right)^2 = \frac{1}{\left(\sum\limits_{1}^{K} a_i\right)^2}\sum_{k+1}^{K} M\,(a_i\,e_i)^2 =$$

$$= \frac{1}{\left(\sum\limits_{1}^{K} a_i\right)^2}\sum_{k+1}^{K} a_i^{\,2}\,\sigma_e^{\,2} = \sigma_e^{\,2}\,\frac{\sum\limits_{k+1}^{K} a_i^{\,2}}{\left(\sum\limits_{1}^{K} a_i\right)^2} \qquad (2)$$

b) Se in luogo di σ_e poniamo l'errore quadratico medio constatato per x nel campione (quarta ipotesi) e cioè:

$$\sqrt{\frac{\varphi}{k\,a}} = \sigma_x \sqrt{\frac{R}{R_{11}}}$$

dovremo attenderci una valutazione di σ_E approssimata per difetto, essendo troppo ottimistica la presunzione che l'errore quadratico medio dei diversi gruppi di k circondarî rispetto all'equazione di regressione calcolata in base al campione, sia lo stesso di quello del campione, che si trova indubbiamente in condizioni più favorevoli degli altri gruppi, poichè i suoi dati sono precisamente quelli utilizzati per determinare i parametri di quella equazione.

---

(1) Cfr. G. Castelnuovo, *Calcolo delle Probabilità*, 2ª ed. Vol. I, § 25.

(2) Questa formula si presta, intanto, ad una semplice interpretazione, e cioè se le popolazioni di tutti i circondarî fossero uguali risulterebbe:

$$\sigma_E^2 = M\,(E^2) = \sigma_e^2\,\frac{K-k}{K^2}$$

cioè:

$$\sigma_E = \sigma_e\,\frac{\sqrt{K-k}}{K}$$

ossia la precisione della valutazione della media di x nel gruppo dei circondarî non scelti starebbe nel rapporto di $\frac{\sqrt{K-k}}{K}$ alla precisione attendibile in un singolo circondario non scelto.

Comunque, il quadrato di tale errore, probabilmente per difetto, avrebbe l'espressione:

$$\sigma_E^2 = \frac{\sum\limits_{k+1}^{K} a_i^2}{\left(\sum\limits_{1}^{K} a_i\right)^2} \frac{R}{R_{11}} \sigma_x^2 ,$$

cosicchè si avrebbe:

$$X = F \pm \sigma_E$$

essendo:

$$F = f\,(U^{(1)}, U^{(2)}, \ldots U^{(t)})$$

ossia:

[7] $$X = \bar{x} + b_1 \left(U^{(1)} - \overline{u^{(1)}}\right) + \ldots + b_t \left(U^{(t)} - \overline{u^{(t)}}\right) \pm \sigma_E$$

Naturalmente, se aggiungiamo l'ipotesi (quinta ipotesi) che i valori medî dei caratteri strumentali nella totalità e nel campione coincidano, si avrà più semplicemente:

[8] $$X = \bar{x} \pm \sigma_E .$$

Il procedimento indicato è adunque applicabile, mediante la formula [7], indipendentemente dal verificarsi di questa ultima condizione; se poi la coincidenza nelle medie si verifica, esattamente o nei limiti di approssimazione prefissati, si potrà invece usare la [8]. Nella applicazione che segue, faremo, per maggiore precisione, uso della [7], in base alle sole prime quattro ipotesi enunciate.

Insistiamo poi ancora nel rilevare che l'ottimismo che informa la quarta ipotesi, potrebbe, mediante la [7] e la [8], fornire un intervallo di approssimazione di X troppo ristretto: nell'esempio che si è già trattato col metodo di Bowley e che stiamo per esaminare nel modo ora indicato, la previsione fornita dalla [7] si avvera egregiamente; il che significa che, nonostante l'ammissione della quarta ipotesi, l'analisi da noi condotta, col criterio di tener conto dei pesi o popolazioni dei diversi circondarî, ha molto aderenza con la realtà.

4. Riprendiamo i dati fondamentali dell'esempio svolto al n. 2 di questa Appendice, ai quali converrà aggiungere, naturalmente, le popolazioni dei circondarî scelti, rilevabili dalla Tav. 1, che interverranno nella

determinazione degli scostamenti quadratici medî e dei mutui indici di correlazione fra i varî caratteri, nel campione, quando si prenda per unità l'individuo (e non il circondario).

Gli scostamenti quadratici medî dei caratteri strumentali u = natalità, v = percentuale di popolazione agglomerata, w = altitudine, e del nuovo carattere x = accrescimento naturale, risultano essere nel campione:

$$\sigma_1 = \sigma_u\,(1) = 6{,}107 \qquad \sigma_2 = \sigma_v = 21{,}232 \qquad \sigma_3 = \sigma_w = 227{,}867$$
$$\sigma_x = 4{,}617\,.$$

I coefficienti di correlazione, dedotti dal campione sono:

$$r_{12} = r_{u\,v} = -0{,}348.800 \quad r_{13} = r_{u\,w} = 0{,}256.827 \quad r_{23} = r_{v\,w} = -0{,}047.823$$
$$r_{1\,x} = r_{x\,u} = 0{,}934.527 \quad r_{2\,x} = r_{x\,v} = -0{,}527.554 \quad r_{1\,w} = r_{x\,w} = 0{,}165{,}394$$

Si trova poi:

$$R_{11} = 0{,}818.660 \quad R_{12} = 0{,}715.140 \quad R_{13} = 0{,}185.178 \quad R_{14} = -0{,}057.121$$
$$R = 0{,}062.098$$

$$\frac{\sigma_x}{R_{11}} = 5{,}639.67$$

$$b_1 = 0{,}660.449 \qquad b_2 = -0{,}049.188 \qquad b_3 = -0{,}001.414$$

$$\frac{\sum\limits_{k+1}^{K} a_i^2}{\left(\sum\limits_{1}^{K} a_i\right)^2} = 0{,}006.838.61$$

$$\sigma_E = 0{,}105$$

Le differenze fra le medie $U^{(i)}$ dei caratteri strumentali nella totalità e quelle $\overline{u^{(i)}}$ nel campione sono:

$$U^{(1)} - \overline{u^{(1)}} = U - \bar{u} = -0{,}31\,; \quad U^{(2)} - \overline{u^{(2)}} = V - \bar{v} = -0{,}94\,;$$
$$U^{(3)} - \overline{u^{(3)}} = W - \bar{w} = 0{,}86.$$

Perciò:

$$X = 12{,}953 + b_1\,(-0{,}31) + b_2\,(-0{,}94) + b_3\,.\,0{,}83 \pm 0{,}105$$

e infine

$$X = 12{,}793 \pm 0{,}105 = \begin{cases} 12{,}898 \\ 12{,}688 \end{cases}$$

(1) Gli scostamenti quadratici medî e gli indici di correlazione sono qui indicati, per semplicità, come nella precedente trattazione dello stesso esempio, al n. 2 della presente Appendice, nonostante il diverso significato.

Il vero valore di X è 12,70, ed è appunto compreso nei limiti di approssimazione previsti.

5. Affinchè il paragone fra i due metodi esposti risulti più conclusivo, trattiamo con ciascuno di essi un altro esempio numerico. La Tav. 36 contiene per 47 unità statistiche complesse (città) il valore della popolazione a (in migliaia) ed i valori di quattro caratteri u, v, w, x. Il valore di un certo carattere per una certa città si deve intendere come la media dei valori che lo stesso carattere ha in *tutti* gli individui della medesima città (1). Nella stessa tavola sono indicate in carattere grassetto 10 città costituenti un campione scelto in modo che i valori medî di u, v, w siano in esso praticamente uguali ai corrispondenti nella totalità. Il valore medio di X nel campione è $X_{10} = 39{,}63$. Che cosa si potrà dire del valore medio X di questo medesimo carattere nella totalità?

Per maggiore chiarezza riuniamo i dati principali del problema nella seguente tabella:

Simboli	Per la totalità	Per il gruppo delle città scelte
a (popolazione)	9993	2850
u	34,74	34,81
v	9,35	9,37
w	20,23	19,83
x	X	$39{,}63 = X_{10}$

(1) I dati numerici della Tav. 36 sono stati cortesemente forniti dal Dr. BOWLEY, che li ha impiegati per applicare il metodo da lui dato, come è detto a pag. 52 del suo più volte citato Rapporto. In tale applicazione u, v, w, x rappresentano rispettivamente i salari medî, nelle varie città, dei compositori, dei muratori, dei meccanici e dei fonditori. Volendo tener conto dei pesi delle diverse unità statistiche (città), è chiaro che tali pesi cambierebbero da carattere a carattere, perchè sarebbero costituiti dai numeri di individui occupati nei varî mestieri in ciascuna città. Verrebbe dunque a mancare una condizione essenziale per la legittima applicabilità sia del metodo del BOWLEY, sia di quello qui dato. Nella sua applicazione questo A. prescinde da tali pesi. Noi, al contrario, per potere saggiare i due metodi in tutta la loro portata, prescinderemo dall'effettivo significato di u, v, w, x, e assumeremo le popolazioni come un unico sistema di pesi applicabile a ciascuno dei caratteri stessi. Quindi u, v, w, x potrebbero rappresentare percentuali di individui aventi certe caratteristiche, o valori medî di caratteri quantitativi posseduti dalle singole persone delle varie città, etc.

a) Applichiamo il metodo del Bowley, utilizzando i medesimi simboli già impiegati al n. 2 di questa Appendice. Troveremo i seguenti valori:

$$r_{xu} = 0{,}3366 \quad r_{xv} = 0{,}3769 \quad r_{xw} = 0{,}4591$$
(nel campione);

$$\rho_{uv} = 0{,}5814 \quad \rho_{uw} = -0{,}0109 \quad \rho_{vw} = 0{,}3749$$
(nella totalità);

$$\sigma_x = 1{,}68$$
(nel campione);

$$\sigma_a = 223{,}3812 \quad \sigma_u = 2{,}1930 \quad \sigma_v = 0{,}4889 \quad \sigma_w = 1{,}3746$$
(nella totalità);

$$U = 34{,}7387 \quad V = 9{,}3509 \quad W = 20{,}2339$$
(media dei caratteri strumentali nella totalità);

$$\bar{a} = 212{,}617$$
(popolazione media delle città);

$$X_{10} = 39{,}63 \quad \bar{x}_{10} = 38{,}89$$
(medie, ponderata e non ponderata, di x nel campione);

$$\bar{u} = 33{,}3564 \quad \bar{v} = 9{,}1489 \quad \bar{w} = 20{,}2660$$
(medie non ponderate di u, v, w, nella totalità);

$$R_{11} = 0{,}5166 \quad R_{12} - = 0{,}1737 \quad R_{13} = 0{,}0048 \quad R_{14} = -0{,}2373$$
$$R = 0{,}3510.$$

Allora i coefficienti G_u G_v G_w dell'equazione di regressione collegante x, ad u, v, w, avranno i valori:

$$G_u = 0{,}1858 \quad G_v = -0{,}0286 \quad G_w = 0{,}8669$$

cosicchè risulterà:

$$K = -(X_{10} - \bar{x}_{10}) + G_u(U - \bar{u}) + G_v(V - \bar{v}) + G_w(W - \bar{w}) = -0{,}5168$$

$$\sigma_{10} = \frac{\sigma_x}{10}\sqrt{1 + \frac{\sigma_a^2}{\bar{a}^2}}\sqrt{\frac{R}{R_{11}}} = 0{,}2009$$

$$X = X_{10} - K \pm \sigma_{10} = 39{,}63 + 0{,}5168 \pm 0{,}2009 = \begin{cases} 40{,}35 \\ 39{,}95 \end{cases}$$

I valori trovati non comprendono il vero valore di X che è 39,80. Tuttavia questo vero valore si allontana dal più prossimo dei limiti trovati per X soltanto di 0,15; mentre nella esemplificazione del Bowley (in cui pure la previsione non si accorda con la realtà) la differenza analoga è 0,426. La maggiore precisione da noi raggiunta, non ostante che il nostro campione comprenda soltanto 10 città, mentre quello scelto dall'A. ne comprendeva 12, può forse dipendere dalla circostanza che pei caratteri u, v, w, egli ha assunto tutti i pesi uguali ad 1, mentre il valore medio del carattere x è stato probabilmente determinato tenendo effettivamente conto delle popolazioni delle diverse città.

Anche qui, come già si notò in fine al n. 2 di questa Appendice, l'intervallo di previsione risulterebbe più favorevole omettendo il valore di K; anzi risulterebbe $X = 39{,}63 \pm 0{,}20 = \begin{cases} 39{,}83 \\ 39{,}43 \end{cases}$ cioè comprenderebbe effettivamente il vero valore di X.

b) Applichiamo ora l'altro metodo proposto e utilizziamo anche qui le stesse notazioni del n. 3, le quali hanno, ricordiamo, un significato diverso da quelle del n. 2.

Si trova che nel campione, *come costituito di individui:*

$$\sigma_1 = \sigma_u = 2{,}029.340 \qquad \sigma_2 = \sigma_v = 0{,}492.250 \qquad \sigma_3 = \sigma_w = 0{,}968.751$$
$$\sigma_x = 1{,}819.222$$
$$r_{12} = r_{uv} = 0{,}528.994 \qquad r_{13} = r_{uw} = 0{,}097.271 \qquad r_{23} = r_{vw} = 0{,}709.202$$
$$r_{1x} = r_{ux} = 0{,}600.627 \qquad r_{2x} = r_{vx} = 0{,}196.701 \qquad r_{3x} = r_{wx} = 0{,}245.434$$

Di qui si trae che:

$$R_{11} = -0{,}286{,}656 \qquad R_{12} = 0{,}276.251 \qquad R_{13} = 0{,}242.902 \qquad R_{14} = 0{,}214.271$$
$$R = -0{,}115.922$$
$$\frac{\sigma_x}{R_{11}} = 6{,}346.359\,;$$

cosicchè i coefficienti dell'equazione di regressione di x rispetto ad u, v, w saranno:

$$b_1 = 0{,}863.924 \qquad b_2 = 3{,}131.630 \qquad b_3 = 1{,}403.707\,;$$

inoltre

$$\frac{\sum\limits_{11}^{47} a_i^2}{\left(\sum\limits_{1}^{47} a_i\right)^2} = \frac{2.888.929}{9.993^2} = 0{,}028.929.8$$

$$\sigma_E = \sqrt{0{,}028.929.8\,\frac{R}{R_{11}}}\,.\,\sigma_x = 0{,}196.771.$$

Le differenze fra le medie $U^{(i)}$ dei caratteri strumentali nella totalità e quelle $\overline{U^{(i)}}$ nel campione sono:

$$U^{(1)} - \overline{u^{(1)}} = U - \overline{u} = -0{,}07\,; \quad U^{(2)} - \overline{u^{(2)}} = V - \overline{v} = -0{,}02\,;$$

$$U^{(3)} - \overline{u^{(3)}} = W - \overline{w} = 0{,}40\,;$$

perciò:

$$X = X_{10} + b_1(-0{,}07) + b_2 \,.\, 0{,}02 + b_3 \,.\, 0{,}40 \pm \sigma_E$$

$$X = 39{,}63 + 0{,}44 \pm 0{,}20 = \begin{cases} 40{,}27 \\ 39{,}87 \end{cases}$$

I limiti trovati per X non comprendono, neanche qui, il vero valore 39,80; tuttavia la differenza fra questo valore e il più prossimo di quei due limiti è 0,07, mentre la differenza analoga riscontrata coll'altro metodo era 0,15. La nostra previsione è dunque migliore, per quanto la quarta ipotesi che si è utilizzata dia sistematicamente luogo, come si è rilevato, ad una valutazione troppo scarsa dell'errore.

TAVOLE

Tav. 1.

ELENCO DEI CIRCONDARI DEL REGNO D'ITALIA (ESCLUSE LE NUOVE PROVINCIE) AL 1° DICEMBRE 1921, DISPOSTI PER ORDINE DI NATALITÀ DECRESCENTE.

(I Circondari costituenti il gruppo scelto sono indicati in carattere grassetto).

N. d'ordine	CIRCONDARIO	Natalità	Mortalità	Nuzialità	Popolazione agricola	Popolazione agglomerata	Reddito medio	Altitudine	Densità	Accrescimento naturale	Popolazione	Area
1	**Adria**	**44,91**	**21,64**	**9,15**	**55, 1**	**49, 9**	**2822**	**4**	**124**	**23, 28**	**99.968**	**805, 38**
2	Comacchio	42,86	20,44	7,57	52, 0	55, 6	3674	1	81	22, 42	57.841	713, 40
3	Asiago	42,55	16,73	14,89	37, 1	61, 8	3821	999	67	25, 82	30.361	452, 11
4	Treviglio	42,45	24,63	11,69	60, 2	69, 9	2920	126	278	17, 83	140.248	504, 90
5	Avezzano	41,87	20,52	11,46	70, 8	92, 4	2252	697	62	21, 35	120.448	1938, 64
6	Chioggia	41,66	22,35	9,02	51, 2	64, 1	3390	2	169	19, 31	67.885	402, 70
7	Bergamo	41,13	24,69	11,57	41, 9	75, 2	3988	366	233	16, 44	331.075	1422, 56
8	Chiari	41,06	22,55	12,16	56, 8	72, 3	3427	148	241	18, 52	102.716	426, 05
9	Melfi	40,85	25,42	11,33	66, 3	92, 7	1753	531	65	15, 43	102.813	1572, 32
10	Casoria	40,66	21,12	10,49	35, 3	95, 5	3494	70	800	19, 54	192.794	241, 10
11	Frosinone	39,72	18,74	11,35	70, 4	51, 7	2000	291	116	20, 98	211.281	1822, 67
12	Santangelo dei Lombardi	39,43	21,29	11,69	69, 5	71, 8	1623	870	89	18, 14	118.523	1338, 13
13	Foggia	39,14	23,44	10,12	48, 4	93, 6	3064	74	75	15, 70	234.254	3110, 92
14	Sansevero	39,02	21,68	10,53	66, 8	96, 9	2741	87	61	17, 34	171.660	2815, 18
15	Cotrone	39,00	21,99	10,93	63, 8	90, 0	2659	43	48	17, 01	82.544	1727, 67
16	Gallipoli	38,80	20,90	10,81	66, 3	96, 0	2131	14	161	17, 90	207.562	1292, 07
17	Paola	38,60	18,08	11,12	69, 0	59, 1	2225	94	104	20, 52	105.979	1019, 77
18	Feltre	38,55	19,04	12,75	50, 4	61, 3	6121	325	121	19, 50	77.626	639, 72
19	**Potenza**	**38,44**	**21,69**	**11,47**	**69, 3**	**89, 0**	**2409**	**823**	**51**	**16, 75**	**151.436**	**2989, 37**
20	Terranova di Sicilia	38,41	23,97	10,21	62, 7	94, 3	3035	45	84	14, 44	87.156	1036, 37
21	Nicastro	38,26	19,57	10,88	68, 9	81, 8	2125	200	136	18, 67	120.099	884, 61
22	Clusone	38,24	23,24	10,38	39, 9	81, 7	2396	648	98	15, 00	84.363	861, 37
23	Brindisi	38,18	21,31	9,91	60, 2	89, 1	2423	11	111	16, 86	188.039	1701, 11
24	Matera	37,79	23,28	10,65	67, 6	96, 7	2424	401	36	14, 50	108.999	3019, 91
25	Belluno	37,55	17,63	11,68	30, 7	73, 4	6264	389	74	19, 92	111.200	1492, 96
26	Ariano di Puglia	37,38	19,54	10,99	71, 0	66, 2	1716	817	114	17, 84	100.604	879, 87
27	**Pavullo nel Frignano**	**37,21**	**16,83**	**11,57**	**70, 7**	**29, 1**	**2028**	**682**	**73**	**20, 38**	**76.808**	**1051, 20**
28	Verolanuova	37,16	18,54	11,94	58, 7	77, 0	4244	64	185	18, 61	70.109	378, 92
29	Velletri	37,13	18,05	11,38	61, 9	79, 2	2157	352	70	19, 08	104.385	1490, 98
30	Lagonegro	37,09	20,60	12,14	66, 0	75, 1	1709	666	44	16, 49	105.309	2405, 83
31	Castellammare di Stabia	37,00	19,77	10,61	32, 1	84, 8	3259	5	760	17, 24	216.622	285, 06
32	Piedimonte d'Alife	36,99	21,45	12,17	69, 0	67, 9	1548	200	67	15, 54	46.855	700, 13

Segue *Tav.* 1.

N d'ordine	CIRCONDARIO	Natalità	Mortalità	Nuzialità	Popolazione agricola	Popolazione agglomerata	Reddito medio	Altitudine	Densità	Accrescimento naturale	Popolazione	Area
33	**Monteleone di Calabria**	**36,98**	**21,21**	**10,62**	**64,3**	**96,1**	**1837**	**556**	**140**	**15,76**	**150.858**	**1080,90**
34	Cosenza	36,95	19,38	10,68	61,7	70,2	2435	240	90	17,56	205.760	2279,56
35	Massa-Carrara	36,72	18,01	10,75	35,5	76,9	3538	65	188	18,71	142.459	755,89
36	Castrovillari	36,70	21,69	11,29	65,0	88,5	1999	350	54	15,01	112.730	2102,61
37	**Bari delle Puglie**	**36,69**	**21,39**	**10,07**	**45,0**	**90,2**	**2788**	**5**	**235**	**15,30**	**430.243**	**1830,10**
38	Breno	36,65	26,26	11,09	47,1	87,9	1612	340	55	10,40	71.756	1303,75
39	Cesena	36,62	18,58	10,97	62,3	39,1	3804	40	158	18,03	114.403	725,38
40	Treviso	36,57	14,03	12,23	56,3	47,9	3887	15	221	22,53	548.487	2476,58
41	**Sala Consilina**	**36,55**	**18,71**	**10,78**	**66,5**	**85,1**	**1435**	**614**	**66**	**17,84**	**73.104**	**1107,85**
42	Urbino	36,52	19,24	11,18	69,0	37,8	1938	451	70	17,29	144.148	2047,40
43	Sora	36,52	17,75	10,75	62,9	46,6	1566	300	124	18,77	170.967	1375,64
44	Vasto	36,49	20,30	10,72	73,2	69,3	1592	144	100	16,18	111.798	1116,99
45	Campagna	36,47	19,23	11,75	69,4	78,8	1633	280	63	17,24	99.111	1561,76
46	Rovigo	36,42	16,31	9,69	57,5	49,1	3482	6	194	20,11	187.270	965,87
47	Mirandola	36,26	16,29	9,32	61,6	34,7	4178	18	173	19,98	79.947	462,34
48	Palmi	36,21	17,84	9,70	65,4	89,2	2224	250	198	18,38	173.357	874,10
49	Lecce	36,20	19,76	10,52	58,2	89,7	2141	51	150	16,44	215.540	1432,62
50	Salerno	36,18	21,63	10,90	39,8	83,8	2382	4	357	14,55	314.008	880,36
51	Benevento	36,11	18,67	11,12	61,5	72,3	1931	135	156	17,44	123.516	791,74
52	Taranto	36,01	18,64	9,39	45,6	90,3	2312	15	113	17,37	274.907	2426,25
53	Lanciano	35,97	18,81	11,03	69,6	63,1	3604	283	130	17,15	123.311	951,65
54	Rocca San Casciano	35,94	18,82	11,21	69,2	39,3	1715	210	59	17,12	59.927	1015,49
55	Nola	35,91	20,43	10,38	55,1	80,7	1431	40	406	15,48	107.108	263,83
56	Caserta	35,86	19,65	10,17	47,1	90,9	1751	68	230	16,20	331.025	1438,63
57	**Isernia**	**35,85**	**21,24**	**11,32**	**72,2**	**77,2**	**1152**	**457**	**70**	**14,61**	**119.111**	**1693,92**
58	Bovino	35,79	23,58	9,43	65,2	92,7	1924	647	51	12,21	52.588	1025,09
59	Vergato	35,73	15,49	11,54	62,7	30,4	1179	195	91	20,25	68.362	750,54
60	Catanzaro	35,52	21,87	10,38	55,4	89,1	2605	343	102	13,65	160.622	1577,08
61	**Rimini**	**35,43**	**18,73**	**9,79**	**54,3**	**39,1**	**3155**	**7**	**236**	**16,70**	**122.564**	**519,65**
62	**Gaeta**	**35,27**	**18,12**	**10,52**	**61,3**	**76,8**	**1617**	**10**	**112**	**17,14**	**160.177**	**1490,49**
63	Penne	35,26	18,02	10,44	69,8	44,0	2335	438	135	17,25	130.876	972,10
64	Larino	35,25	21,78	11,16	70,1	95,2	1935	310	70	13,47	105.353	1468,87
65	San Bartolomeo in Galdo	35,22	20,57	12,03	80,2	83,6	1310	590	91	14,65	60.278	660,47
66	Barletta	35,12	22,00	9,56	57,6	96,5	2546	15	212	13,12	385.697	1821,61
67	**Padova**	**35,03**	**16,11**	**10,69**	**55,9**	**39,4**	**3420**	**12**	**275**	**18,92**	**588.043**	**2140,73**
68	Pozzuoli	34,96	16,94	10,14	42,8	77,1	1834	39	634	18,03	106.736	168,36
69	Venezia	34,92	16,22	10,09	35,5	59,7	6592	2	221	18,70	451.323	2041,66

Segue *Tav.* 1.

N d'ordine	CIRCONDARIO	Natalità	Mortalità	Nuzialità	Popolazione agricola	Popolazione agglomerata	Reddito medio	Altitudine	Densità	Accrescimento naturale	Popolazione	Area
70	Campobasso	34,76	20,67	11,52	68, 7	84, 6	1657	700	95	14, 09	116.445	1220, 54
71	Spoleto	34,70	18,34	11,01	60, 1	62, 8	2438	317	51	16, 36	84.401	1663, 48
72	Avellino	34,64	18,74	10,67	59, 6	76, 9	1673	351	230	15, 90	184.258	800, 35
73	Sulmona	34,53	20,38	11,54	59, 8	92, 1	2068	403	73	14, 15	87.350	1193, 01
74	**Lanusei**	34,52	20,03	9,85	67, 5	97, 6	1278	595	24	14, 49	78.798	3346, 40
75	Gerace Marina	34,48	19,71	10,14	68, 4	72, 7	2112	5	113	14, 76	150.092	1330, 82
76	Ferrara	34,32	16,00	7,59	53, 4	53, 8	6471	10	143	18, 32	244.891	1717, 06
77	Cividale del Friuli	34,26	16,58	10,65	66, 1	81, 4	6056	138	117	17, 68	67.485	574, 40
78	Reggio Calabria	34,21	16,79	8,91	47, 8	78, 8	2870	29	188	17, 42	178.862	952, 60
79	Teramo	34,13	16,78	10,33	67, 7	47, 9	2924	265	107	17, 34	189.114	1773, 20
80	Chieti	34,12	17,61	10,58	62, 9	48, 5	2830	330	159	16, 51	141.133	890, 32
81	Altamura	34,08	20,41	9,37	65, 5	91, 6	2470	473	83	13, 66	136.571	1653, 33
82	Pordenone	34,01	13,75	10,83	51, 9	75, 3	8099	24	107	20, 26	156.945	1465, 27
83	**Vallo della Lucania**	33,93	17,28	10,85	69, 3	85, 0	1477	380	70	16, 65	98.090	1393, 80
84	**Orvieto**	33,92	16,64	12,86	71, 6	35, 3	1826	315	57	17, 29	60.109	1054, 85
85	Pontremoli	33,87	17,30	12,15	65, 7	72, 8	1406	237	86	16, 56	40.568	471, 03
86	Rossano	33,79	18,44	9,72	62, 8	81, 8	3021	297	57	15, 35	71.415	1244, 83
87	Foligno	33,64	17,21	11,30	55, 2	62, 5	2353	234	94	16, 43	88.547	938, 24
88	Perugia	33,63	17,42	11,55	66, 5	34, 7	2994	493	82	16, 21	291.982	3550, 58
89	Crema	33,57	17,63	11,09	56, 1	78, 6	4551	79	220	15, 94	109.300	497, 29
90	Cittaducale	33,44	17,26	10,24	69, 9	89, 7	2153	450	44	16, 18	60.086	1361, 56
91	Rieti	33,39	17,10	11,21	65, 8	64, 8	2297	402	74	16, 28	101.690	1376, 80
92	Udine	33,38	16,88	10,86	45, 0	88, 4	6695	137	148	16, 51	420.938	2840, 57
93	Vicenza	33,37	16,83	11,69	52, 4	50, 3	3867	40	227	16, 55	517.119	2282, 84
94	Patti	33,24	15,37	10,07	63, 8	51, 9	1981	153	156	17, 86	122.095	781, 09
95	Fermo	33,14	17,43	10,90	64, 1	34, 6	3160	320	151	15, 71	133.518	882, 05
96	Castelnuovo di Garfag.na	33,11	15,84	12,61	58, 5	71, 6	2084	277	77	17, 27	42.917	554, 60
97	**Brescia**	33,02	20,77	10,05	44, 4	77, 0	4603	149	204	12, 25	334.948	1639, 80
98	Ascoli Piceno	32,89	17,23	10,23	64, 6	56, 4	2633	153	109	15, 66	131.646	1203, 11
99	Montepulciano	32,88	17,48	12,69	66, 7	44, 5	2700	605	64	15, 40	79.674	1237, 54
100	Reggio Emilia	32,87	16,76	10,56	55, 5	45, 2	4457	60	141	16, 11	266.611	1892, 03
101	**Alghero**	32,81	21,01	10,36	64, 0	98, 1	1753	7	43	11, 80	49.319	1159, 71
102	Viterbo	32,80	18,62	11,21	63, 5	82, 7	2085	327	65	14, 18	196.842	3020, 88
103	Nuoro	32,65	19,32	9,46	72, 3	97, 5	1212	553	24	13, 33	77.033	3154, 56
104	Aquila degli Abruzzi	32,58	20,84	10,80	61, 9	94, 4	3353	721	65	12, 34	127.915	1965, 00
105	**Tolmezzo**	32,57	19,23	11,70	24, 0	90, 2	3278	323	44	13, 35	74.628	1689, 90
106	**Camerino**	32,47	17,94	12,04	69, 3	71, 2	1778	661	48	14, 53	50.324	1059, 20

Segue *Tav.* 1.

N. d'ordine	CIRCONDARIO	Natalità	Mortalità	Nuzialità	Popolazione agricola	Popolazione agglomerata	Reddito medio	Altitudine	Densità	Accrescimento naturale	Popolazione	Area
107	Iglesias	32,38	20,28	9,95	43, 6	83, 0	1795	199	36	12, 11	103.179	2839, 65
108	Modena	32,37	16,83	9,93	48, 7	35, 1	3737	35	220	15, 54	238.758	1083, 86
109	**Arezzo**	**32,27**	**17,65**	**11,80**	**62, 1**	**39, 0**	**2473**	**296**	**90**	**14, 62**	**298.519**	**3302, 39**
110	Guastalla	32,12	14,52	9,74	58, 0	38, 3	5142	25	202	17, 59	80.484	399, 31
111	Pesaro	32,11	18,13	11,39	54, 6	49, 8	2632	11	161	13, 99	136.570	847, 49
112	Sassari	32,09	18,74	10,06	50, 9	88, 7	2289	225	57	13, 34	106.488	1878, 67
113	Cagliari	32,02	19,47	8,66	49, 6	97, 1	2062	10	53	12, 56	211.245	3975, 65
114	**Cerreto Sannita**	**31,98**	**17,37**	**10,72**	**73, 2**	**55, 8**	**1286**	**290**	**124**	**14, 61**	**83.377**	**670, 66**
115	Noto	31,85	16,64	9,94	63, 6	83, 5	1593	159	101	15, 21	111.644	1104, 15
116	Macerata	31,82	17,08	11,41	62, 9	37, 7	2023	311	127	14, 74	217.436	1713, 93
117	Girgenti	31,77	22,63	9,00	54, 5	96, 1	2023	326	175	9, 15	272.880	1558, 94
118	**Castroreale**	**31,73**	**16,57**	**10,68**	**61, 9**	**82, 3**	**2274**	**394**	**143**	**15, 16**	**115.890**	**810, 20**
119	Ancona	31,69	16,59	11,29	48, 9	51, 9	4874	106	173	15, 09	334.654	1937, 70
120	Bivona	31,67	21,42	9,11	61, 2	97, 7	1305	503	92	10, 25	75.624	821, 90
121	Cento	31,61	14,12	10,07	48, 9	34, 4	3185	15	206	17, 49	43.283	209, 72
122	Salò	31,56	20,17	12,02	57, 2	93, 4	2320	75	75	11, 39	72.696	973, 45
123	Monza	31,55	20,23	11,06	25, 9	77, 1	4230	162	717	11, 31	322.032	449, 36
124	Ozieri	31,54	18,03	8,20	65, 3	92, 7	1433	390	22	13, 50	51.688	2338, 84
125	Lodi	30,65	16,89	11,57	46, 9	75, 1	5747	80	218	13, 76	179.419	824, 44
126	**Grosseto**	**30,60**	**15,95**	**11,38**	**61, 8**	**60, 1**	**2146**	**10**	**37**	**14, 65**	**164.990**	**4496, 00**
127	Nicosia	30,57	17,52	8,62	67, 4	94, 9	1851	714	89	13, 05	121.856	1366, 40
128	**Sondrio**	**30,32**	**20,86**	**10,41**	**64, 3**	**83, 5**	**2047**	**298**	**41**	**9, 46**	**131.184**	**3193, 92**
129	Lecco	30,18	20,44	11,42	43, 6	80, 9	4269	212	224	9, 74	157.662	702, 36
130	Mistretta	30,16	19,42	9,51	64, 0	97, 3	2037	950	66	10, 73	58.131	886, 91
131	Acireale	30,11	15,05	11,40	46, 9	81, 0	3762	161	227	15, 06	151.200	667, 06
132	Pieve di Cadore	30,02	18,49	9,50	26, 9	96, 7	3373	878	34	11, 53	39.801	1172, 71
133	Forlì	29,79	16,85	9,76	57, 6	40, 4	2783	34	146	12, 94	94.132	646, 78
134	Mantova	29,70	15,52	10,15	55, 8	53, 1	5763	20	161	14, 18	376.901	2339, 38
135	Verona	29,66	15,44	11,00	50, 6	62, 7	4289	59	169	14, 22	518.256	3071, 20
136	Oristano	29,54	19,70	9,41	65, 2	98, 9	1961	9	42	9, 84	137.010	3253, 52
137	Fiorenzuola d'Arda	29,32	13,99	10,19	67, 6	30, 6	4407	82	91	15, 33	78.920	869, 94
138	Civitavecchia	29,19	14,13	9,41	39, 1	88, 3	2640	11	43	15, 06	46.348	1073, 37
139	Termini Imerese	28,84	20,92	10,22	62, 8	96, 3	2003	113	83	7, 92	90.529	1088, 09
140	**Borgotaro**	**28,78**	**15,78**	**10,52**	**71, 3**	**45, 9**	**1908**	**411**	**51**	**13, 00**	**37.460**	**741, 28**
141	Terni	28,74	16,87	11 07	46, 6	54, 1	2442	130	94	11, 87	111.341	1183, 19
142	Caltanissetta	28,38	19,37	8,16	51, 5	84, 9	2411	588	150	9, 01	160.232	1074, 74
143	Napoli	28,32	18,53	8,60	6, 6	96, 6	5917	19	4452	9, 79	952.488	213, 93

Segue *Tav.* 1.

N. d'ordine	CIRCONDARIO	Natalità	Mortalità	Nuzialità	Popolazione agricola	Popolazione agglomerata	Reddito medio	Altitudine	Densità	Accrescimento naturale	Popolazione	Area
144	Modica	28,24	16,00	8,97	61, 5	91, 6	1682	449	168	8, 97	252.546	1504, 51
145	Parma	28,13	15,79	10,12	48, 3	56, 8	4811	52	131	12, 34	207.680	1585, 06
146	Palermo	27,99	17,45	8,49	29, 4	94, 5	4088	9	353	10, 54	617.662	1750, 85
147	Volterra	27,98	15,49	11,39	49, 4	58, 2	3007	531	67	12, 49	100.814	1504, 07
148	Caltagirone...........	27,92	17,97	8,35	66, 7	91, 9	1827	608	100	9, 95	154.329	1544, 84
149	Borgo San Donnino	27,87	14,00	10,74	61, 0	36, 7	4194	75	116	13, 88	108.238	933, 05
150	Corleone	27,77	21,06	9,84	69, 5	99, 1	1689	594	67	6, 70	51.606	775, 49
151	Messina	27,76	15,28	8,10	33, 1	91, 9	2898	5	371	12, 48	285.948	770, 75
152	Mazzara del Vallo	27,16	20,25	9,25	61, 8	90, 2	2331	8	107	6, 90	97.952	912, 03
153	Imola................	27,14	16,15	10,59	59, 1	37, 7	3167	47	114	11, 00	90.111	790, 30
154	Lucca	27,05	15,01	10,86	45, 8	45, 2	3759	19	241	12, 03	346.602	1436.82
155	**Roma**	**27,03**	**16,50**	**9,78**	**19, 4**	**93, 8**	**6004**	**20**	**205**	**10, 53**	**958.436**	**4674, 82**
156	Pistoia...............	27,02	14,80	12,17	48, 3	46, 1	2995	65	191	12, 22	140.375	735, 86
157	Abbiategrasso	26,80	15,78	11,64	51, 5	77, 5	4206	120	239	11, 03	134.578	562, 58
158	Bobbio	26,76	14,64	11,78	77, 6	69, 5	1739	272	55	12, 12	38.457	695, 59
159	Catania	26,73	15,52	8,02	34, 2	93, 9	4568	38	390	11, 21	448.880	1359, 61
160	**Siracusa**.............	**26,66**	**13,65**	**8,65**	**54, 3**	**92, 5**	**2005**	**5**	**156**	**13, 01**	**172.424**	**1103, 87**
161	San Miniato...........	26,62	14,47	12,39	56, 4	44, 0	3432	117	170	12, 15	143.284	840, 67
162	Alba	26,54	14,98	12,14	70, 5	56, 2	3439	172	149	11, 56	151.039	1016, 79
163	Tempio Pausania	26,51	16,84	8,72	45, 8	65, 1	2141	566	23	9, 67	49.414	2143, 17
164	Casalmaggiore	26,42	15,57	11,65	57, 3	89, 2	4588	26	150	10, 85	45.150	300, 55
165	Cefalù	26,25	17,09	9,41	67, 3	88, 7	2514	30	73	9, 13	100.509	1377, 88
166	Cremona	26,23	18,00	11,22	43, 4	76, 4	5267	45	208	8, 23	203.155	977, 12
167	Piazza Armerina.......	25,97	18,33	7,77	56, 3	86, 2	1905	721	117	7, 64	138.287	1186, 75
168	Siena	25,84	17,24	12,44	58, 0	49, 2	2864	322	65	8, 59	168.168	2574, 02
169	Piacenza	25,79	15,34	9,53	50, 6	58, 7	5368	61	124	10, 45	202.389	1635, 52
170	Gallarate.............	25,69	14,61	11,31	24, 1	88, 8	5900	238	492	11, 09	267.414	543, 52
171	Cuneo	25,61	19,27	10,78	59, 7	58, 4	3494	534	54	6, 34	171.944	3172, 20
172	Pisa	25,46	14,53	11,34	45, 7	67, 5	2957	4	166	10, 93	259.973	1565, 19
173	Domodossola	25,22	19,02	10,25	35, 9	84, 9	2484	277	30	6, 20	43.700	1450, 66
174	Bologna	25,11	14,75	9,81	35, 9	52, 0	5733	55	215	10, 36	484.191	2254, 66
175	Spezia	25,09	13,64	9,34	22, 0	72, 8	3649	3	312	11, 45	195.925	628, 49
176	Como	24,95	18,61	11,12	31, 2	84, 7	3686	202	219	6, 34	299.286	1364, 52
177	Faenza	24,94	15,88	10,50	57, 9	45, 4	3136	35	131	9, 06	79.995	611, 00
178	Acqui	24,84	14,43	12,59	68, 9	46, 1	2513	164	133	10, 41	122.683	764, 19
179	Lugo	24,77	15,76	10,15	56, 4	45, 3	2833	15	206	9, 02	76.419	371, 00
180	Saluzzo	24,76	17,28	10,68	61, 7	55, 5	4301	395	96	7, 49	147.873	1543, 15

Segue *Tav.* 1.

N. d'ordine	CIRCONDARIO	Natalità	Mortalità	Nuzialità	Popolazione agricola	Popolazione agglomerata	Reddito medio	Altitudine	Densità	Accrescimento naturale	Popolazione	Area
181	Ravenna	24,70	15,85	9,74	54, 7	41, 2	3925	3	116	8, 84	101.190	870, 00
182	Aosta	24,64	21,23	9,81	60, 9	81, 2	3187	583	24	3, 41	78.811	3264, 92
183	**Mondovì**	**24,54**	**15,88**	**11,84**	**67, 3**	**59, 4**	**2872**	**559**	**90**	**8, 66**	**152.742**	**1702, 73**
184	Trapani	23,98	13,15	9,60	47, 8	60, 5	3857	3	194	10, 83	198.473	1021, 92
185	Livorno	23,97	17,43	8,52	4, 8	92, 1	6426	3	1159	6, 54	114.809	99, 06
186	Sciacca	23,93	17,35	7,93	63, 8	90, 9	2090	60	95	6, 58	62.777	662, 81
187	Firenze	23,86	15,24	11,06	30, 9	65, 1	6565	49	214	8, 62	698.191	3266, 75
188	Portoferraio	23,31	15,91	9,37	25, 0	78, 4	2193	10	118	7, 40	28.914	244, 35
189	Novi Ligure	22,76	14,79	11,97	51, 5	68, 1	2757	197	111	7, 97	92.560	830, 88
190	Alcamo	22,42	16,17	7,62	68, 9	99, 3	2101	256	196	6, 26	112.822	574, 86
191	Savona	22,10	13,71	10,10	29, 7	72, 1	4545	19	150	8, 39	145.378	971, 42
192	**Susa**	**22,00**	**16,21**	**11,04**	**50, 7**	**78, 4**	**2308**	**503**	**62**	**5, 79**	**86.905**	**1396, 98**
193	Asti	21,44	14,23	13,04	66, 6	57, 9	3613	123	187	7, 20	182.121	974, 47
194	Pinerolo	21,34	16,62	10,86	57, 9	64, 5	3273	377	85	4, 72	120.866	1420, 08
195	Pavia	21,30	14,43	11,89	44, 9	79, 7	4308	77	199	6, 87	159.663	801, 50
196	**Voghera**	**21,19**	**14,27**	**10,99**	**56, 9**	**69, 2**	**2677**	**93**	**180**	**6, 91**	**138.991**	**774, 01**
197	Ivrea	21,13	18,59	11,70	57, 6	77, 7	2388	267	99	2, 55	148.074	1494, 52
198	Chiavari	21,09	15,26	9,65	47, 6	61, 6	3092	3	128	5, 83	117.027	911, 29
199	Albenga	21,05	14,91	10,20	53, 7	85, 3	2157	5	97	6, 14	60.082	618, 40
200	Pallanza	20,91	16,88	11,33	16, 2	90, 2	3029	201	107	4, 03	84.543	788, 62
201	Tortona	20,25	13,60	10,99	59, 7	72, 0	2672	114	114	6, 65	75.603	665, 77
202	Varese	19,99	15,50	11,65	21, 3	81, 5	3446	382	219	4, 48	174.029	794, 58
203	Milano	19,80	14,73	9,88	10, 0	93, 1	9510	122	1280	5, 07	1002.788	783, 33
204	Porto Maurizio	19,62	16,48	10,00	47, 7	95, 7	4822	47	118	3, 14	60.201	511, 57
205	Varallo	19,49	18,49	10,71	15, 8	90, 0	2983	451	42	1, 00	32.121	773, 27
206	**Genova**	**19,02**	**14,70**	**9,01**	**10, 3**	**88, 6**	**8314**	**25**	**679**	**4, 32**	**657.657**	**968, 41**
207	Casale Monferrato	18,93	15,03	12,25	60, 0	77, 2	3489	116	177	3, 90	148.789	840, 59
208	Novara	18,29	14,54	12,52	44, 2	85, 1	4018	164	189	3, 74	261.206	1385, 00
209	San Remo	18,14	16,53	9,79	45, 6	89, 7	3857	15	135	1, 61	90.634	670, 60
210	Alessandria	17,82	14,76	11,53	40, 7	72, 7	5124	95	190	3, 06	159.994	843, 21
211	Vercelli	17,04	13,97	13,20	54, 3	86, 3	3749	131	117	3, 06	146.553	1249, 05
212	Torino	16,66	15,32	10,39	19, 7	88, 3	7694	239	307	1, 33	818.787	2663, 80
213	Biella	16,55	14,87	11,30	24, 9	83, 2	7570	424	164	1, 68	156.732	955, 02
214	Mortara	15,58	13,47	12,32	53, 2	78, 6	6080	108	146	2, 10	155.409	1064, 53

CARATTERISTICHE GENERICHE DELLA TOTALITÀ DEI CIRCONDARI.

Tav. 2.

Quoziente di natalità per 1000 abitanti	Numero dei Circondari
Da 15 a 16 escluso	1
» 16 a 17 »	2
» 17 a 18 »	2
» 18 a 19 »	3
» 19 a 20 »	5
» 20 a 21 »	2
» 21 a 22 »	7
» 22 a 23 »	4
» 23 a 24 »	5
» 24 a 25 »	8
» 25 a 26 »	9
» 26 a 27 »	10
» 27 a 28 »	11
» 28 a 29 »	7
» 29 a 30 »	6
» 30 a 31 »	8
» 31 a 32 »	11
» 32 a 33 »	16
» 33 a 34 »	15
» 34 a 35 »	15
» 35 a 36 »	15
» 36 a 37 »	21
» 37 a 38 »	8
» 38 a 39 »	8
» 39 a 40 »	5
» 40 a 41 »	2
» 41 a 42 »	4
» 42 a 43 »	3
» 43 a 44 »	—
» 44 a 45 »	1
TOTALE ...	214

Tav. 3.

Quoziente di mortalità per 1000 abitanti.	Numero dei Circondari
Da 13 a 14 escluso	9
» 14 a 15 »	23
» 15 a 16 »	26
» 16 a 17 »	32
» 17 a 18 »	26
» 18 a 19 »	28
» 19 a 20 »	18
» 20 a 21 »	20
» 21 a 22 »	19
» 22 a 23 »	4
» 23 a 24 »	5
» 24 a 25 »	2
» 25 a 26 »	1
» 26 a 27 »	1
TOTALE ...	214

Tav. 4.

Quoziente di nuzialità per 1000 abitanti	Numero dei Circondari
Da 7 a 8 escluso	5
» 8 a 9 »	14
» 9 a 10 »	40
» 10 a 11 »	71
» 11 a 12 »	61
» 12 a 13 »	20
» 13 a 14 »	2
» 14 a 15 »	1
TOTALE ...	214

Tav. 5.

Popolazione agricola maschile su 100 maschi di età superiore ai 10 anni	Numero dei Circondari
Fino a 10 escluso	2
da 10 a 20 »	6
» 20 a 30 »	10
» 30 a 40 »	15
» 40 a 50 »	33
» 50 a 60 »	53
» 60 a 70 »	81
» 70 a 80 »	13
» 80 a 90 »	1
TOTALE ...	214

Tav. 6.

Popolazione agglomerata su 100 abitanti	Numero dei Circondari
Da 20 a 30 escluso	1
» 30 a 40 »	18
» 40 a 50 »	20
» 50 a 60 »	21
» 60 a 70 »	22
» 70 a 80 »	37
» 80 a 90 »	45
» 90 a 100 »	50
TOTALE ...	214

Segue CARATTERISTICHE GENERICHE DELLA TOTALITÀ DEI CIRCONDARI.

Tav. 7.

Reddito medio	Numero dei Circondari
Fino a 1500 escluso	12
da 1500 a 2000 »	35
» 2000 a 2500 »	49
» 2500 a 3000 »	28
» 3000 a 3500 »	27
» 3500 a 4000 »	18
» 4000 a 4500 »	13
» 4500 a 5000 »	8
» 5000 a 5500 »	4
» 5500 a 6000 »	5
» 6000 a 6500 »	7
» 6500 a 7000 »	3
» 7000 a 7500 »	—
» 7500 a 8000 »	2
» 8000 a 8500 »	2
» 8500 a 9000 »	—
» 9000 a 9500 »	—
» 9500 a 10000 »	1
TOTALE ...	214

Tav. 8.

Altitudine del Capoluogo in metri s/m	Numero dei Circondari
Fino a 50 escluso	65
da 50 a 100 »	23
» 100 a 150 »	18
» 150 a 200 »	11
» 200 a 250 »	12
» 250 a 300 »	14
» 300 a 350 »	14
» 350 a 400 »	11
» 400 a 450 »	7
» 450 a 500 »	6
» 500 a 550 »	5
» 550 a 600 »	9
» 600 a 650 »	5
» 650 a 700 »	4
« 700 a 750 »	4
» 750 a 800 »	—
» 800 a 850 »	2
» 850 a 900 »	2
» 900 a 950 »	—
» 950 a 1000 »	2
TOTALE ...	214

Tav. 9.

Popolazione assoluta	Numero dei Circondari
Fino a 40.000 escl.	6
da 40.000 a 80.000 »	43
» 80.000 a 120.000 »	50
» 120.000 a 160.000 »	44
» 160.000 a 200.000 »	22
» 200.000 a 240.000 »	12
» 240.000 a 280.000 »	8
» 280.000 a 320.000 »	5
» 320.000 a 360.000 »	6
» 360.000 a 400.000 »	2
» 400.000 a 440.000 »	2
» 440.000 a 480.000 »	2
» 480.000 a 520.000 »	3
» 520.000 a 560.000 »	1
» 560.000 a 600.000 »	1
» 600.000 a 640.000 »	1
» 640.000 a 680.000 »	1
» 680.000 a 720.000 »	1
» 720.000 a 760.000 »	—
» 760.000 a 800.000 »	—
» 800.000 a 840.000 »	1
» 840.000 a 880.000 »	—
» 880.000 a 920.000 »	2
» 920.000 a 960.000 »	—
» 960.000 a 1.000.000 »	1
» 1.000.000 a 1.040.000 »	—
TOTALE ...	214

Tav. 10.

CONFRONTO FRA LE INTENSITÀ MEDIE DEI CARATTERI NELL'INSIEME E NEL CAMPIONE.

CARATTERI	Insieme dei 214 Circondari dell'antico Regno A	Campione (29 Circondari scelti) E	Scarto assoluto $E-A$	Scarto relativo	
				$\frac{E-A}{A}$	$\frac{E-A}{A\,\frac{L-A}{L}}$ (1)
1. Natalità	30,23 ‰	30,54 ‰	0,31 ‰	—	1,06 %
2. Mortalità	17,54 ‰	17,59 ‰	0,05 ‰	—	0,29 %
3. Nuzialità	10,45 ‰	10,34 ‰	— 0,11 ‰	—	— 1,06 %
4. Rapporto fra la popalazione agricola maschile e la popolazione maschile di più di 10 anni	47,36 %	45,37 %	— 1,99 %	—	— 7,98 %
5. Rapporto fra la popolazione agglomerata e la popolazione totale	73,25 %	74,69 %	1,44 %	—	7,35 %
6. Reddito medio per contribuente (redditi del lavoro o misti)	4.152,62	4.268,83	116,21	2,80 %	—
7. Altitudine media sul livello del mare (in metri)	171,35	170,49	— 0,86	— 0,50 %	—
8. Densità della popolazione (abitanti per Km²)	129,59	114,74	— 14,85	— 11,46 %	—
9. Accrescimento naturale (1) - (2)	12,70 ‰	12,95 ‰	0,25 ‰	—	1,99 %

(1) Il limite superiore è $L = 1000$ per la natalità, la mortalità, la nuzialità e l'accrescimento naturale, ed esso è $L = 100$ per i rapporti relativi alla popolazione agricola e alla popolazione agglomerata.

Tav. **11.**

CONFRONTO FRA LE QUANTITÀ FONDAMENTALI NELLA TOTALITÀ E NEL GRUPPO SCELTO.

QUANTITÀ FONDAMENTALI	Totalità dei Circondarî (escluse le nuove Provincie) *a*	Gruppo dei Circondarî scelti *b*	Rapporto percentuale $\frac{b}{a} \times 100$ *c*
1. Popolazione presente	(1) 37.142.886	(1) 5.724.103	15,41
2. Semisomma dei nati 1921-22.....	(2) 1.122.894	(2) 174.826	15,57
3. Semisomma dei morti 1921-22 ...	(2) 651.322,5	(2) 100.684,5	15,46
4. Semisomma dei matrimoni 1921-22	(2) 388.325,5	(2) 59.189	15,24
5. Popolazione agglomerata.........	(3) 27.394.030	(3) 4.275.596	15,61
6. Popolazione maschile in età superiore ai 10 anni	(3) 14.495.343	(3) 2.261.669	15,60
7. Popolazione agricola maschile	6.864.842	1.026.057	14,95
8. Reddito complessivo	(4) 3.510.614.398	(4) 585.158.632	16,67
9. Numero dei redditieri...........	(4) 845.398	(4) 137.077	16,21
10. Somma dei prodotti delle popolazioni dei Circondarî per le altitudini dei rispettivi capoluoghi	6.364.567.082	975.898.501	15,33
11. Area	(5) 286.610,37	(5) 49.888,52	17,41
12. Numero dei fogli di famiglia	(6) 8.347.995	(6) 1.280.999	15,34
13. Numero dei Circondarî	214	29	—

(1) *Risultati sommarî del Censimento della popolazione eseguito il 1° dicembre 1921*, Roma, 1927.

(2) *Movimento della popolazione secondo gli atti dello Stato Civile, in ciascuno degli anni 1919–1923.* Tav. I: Movimento generale della popolazione negli anni 1921–1922; *A*) Nei 214 Circondarî.

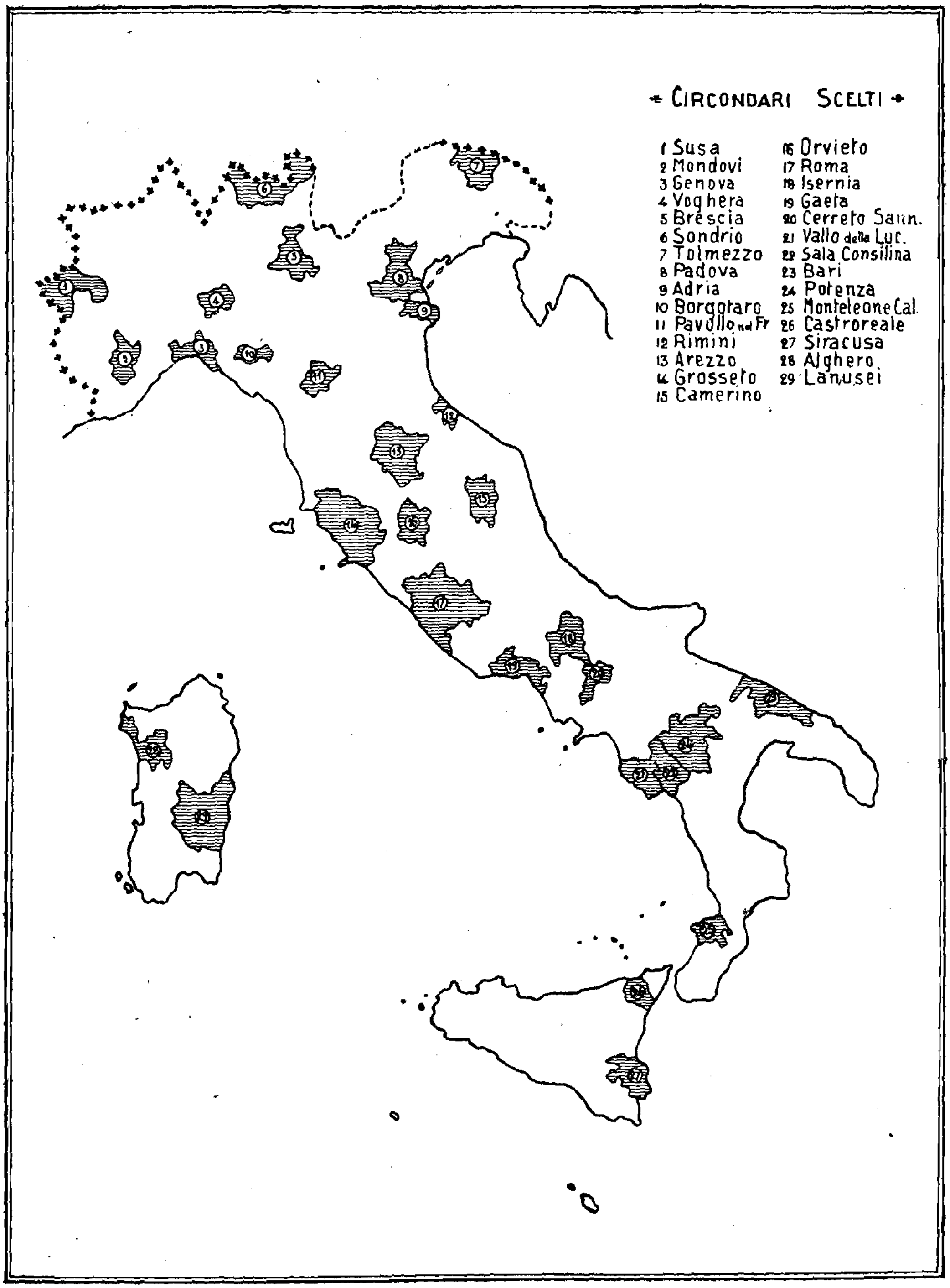
(3) Questi dati subirono successivamente leggere modificazioni per effetto di accertamenti ulteriori. Alla data 30 settembre 1927 in cui vennero sottoposti a calcolo, non si conoscevano esattamente le popolazioni agglomerate e le popolazioni agricole dei Circondarî del Veneto. Essi sono reperibili nei volumi regionali del *Censimento della Popolazione del Regno d'Italia al 1° dicembre 1921*, la cui pubblicazione è ora compiuta.

(4) Cfr. *Elenco dei contribuenti privati possessori di redditi incerti e variabili*, pubblicato dal Ministero delle Finanze, Direzione Generale delle Imposte Dirette, Roma, Libreria dello Stato.

(5) *Annuario Statistico Italiano* 1919–1921, II Serie, Volume VIII.

(6) *Censimento della Popolazione del Regno d'Italia al 1° dicembre 1921.* Volumi regionali, Tavola III: Convivenze, famiglie presenti e interamente assenti nei Centri di almeno 15.000 abitanti, nei Comuni capoluoghi di circondario, nei Circondarî e nelle Provincie.

Tav. 12.


CIRCONDARI SCELTI
1 Susa
2 Mondovi
3 Genova
4 Voghera
5 Brescia
6 Sondrio
7 Tolmezzo
8 Padova
9 Adria
10 Borgotaro
11 Pavullo nel Fr
12 Rimini
13 Arezzo
14 Grosseto
15 Camerino
16 Orvieto
17 Roma
18 Isernia
19 Gaeta
20 Cerreto Sann.
21 Vallo della Luc.
22 Sala Consilina
23 Bari
24 Potenza
25 Monteleone Cal.
26 Castroreale
27 Siracusa
28 Alghero
29 Lanusei

Tav. 13.

COMPARAZIONE TRA I RAPPORTI DI CONCENTRAZIONE DEI CARATTERI NELL'INSIEME (R_a) E NEL CAMPIONE (R_e).

CARATTERI	Insieme (214 Circondari dell'antico Regno) R_a	Campione (29 Circondari scelti) R_e	Scarto assoluto $R_e - R_a$	Scarto relativo $\dfrac{R_e - R_a}{R_a \dfrac{100 - R_a}{100}}$
1. Natalità	12,2 %	11,4 %	— 0,8 %	— 7,5 %
2. Mortalità	8,55 %	7,84 %	— 0,71 %	— 9,1 %
3. Nuzialità	6,27 %	4,95 %	— 1,32 %	— 22,4 %
4. Rapporto fra la popolazione agricola maschile e la popolazione maschile di più di 10 anni	40,1 %	47,2 %	7,1 %	29,7 %
5. Rapporto fra la popolazione agglomerata e la popolazione totale....................	54,6 %	60,1 %	5,5 %	22,1 %
6. Reddito medio per contribuente (redditi di lavoro o misti)	27,2 %	30,3 %	3,1 %	15,7 %
7. Altitudine media sul livello del mare.................	59,0 %	65,0 %	6,0 %	24,8 %
8. Densità della popolazione (abitanti per Km^2)............	39,2 %	43,0 %	3,8 %	16,0 %
9. Accrescimento naturale	22,5 %	20,3 %	— 2,2 %	— 12,6 %

Tav. **14.**

DISTRIBUZIONE DELLA NATALITÀ RISPETTO ALLA POPOLAZIONE.

Quozienti di natalità per 1000 abitanti	Popolazione totale		Popolazione totale ridotta a 10.000.000	
	Regno *a*	Circondari scelti *b*	Regno *c*	Circondari scelti *d*
Da 15 a 16 escluso	155.409	—	41.841	—
» 16 a 17 »	975.519	—	262.640	—
» 17 a 18 »	306.547	—	82.532	—
» 18 a 19 »	500.629	—	134.785	—
» 19 a 20 »	1.926.796	657.657	518.753	1.148.926
» 20 a 21 »	160.146	—	43.116	—
» 21 a 22 »	926.824	138.991	249.529	242.817
» 22 a 23 »	437.665	86.905	117.833	151.823
» 23 a 24 »	1.103.164	—	297.006	—
» 24 a 25 »	1.058.999	152.742	285.115	266.840
» 25 a 26 »	1.931.991	—	520.151	—
» 26 a 27 »	1.486.890	172.424	400.316	301.224
» 27 a 28 »	2.952.073	958.436	794.788	1.674.386
» 28 a 29 »	1.812.276	37.460	487.920	65.443
» 29 a 30 »	1.251.567	—	336.960	—
» 30 a 31 »	1.004.243	296.174	270.373	517.416
» 31 a 32 »	1.701.204	199.267	458.016	348.119
» 32 a 33 »	2.229.235	472.790	600.178	825.964
» 33 a 34 »	2.493.322	493.147	671.279	861 527
» 34 a 35 »	2.374.404	78.798	639.262	137.660
» 35 a 36 »	2.582.042	996.895	695.165	1.741.574
» 36 a 37 »	3.791.224	654.205	1.020.714	1.142.895
» 37 a 38 »	894.036	76.808	240.702	134.183
» 38 a 39 »	1.022.260	151.436	275.224	264.559
» 39 a 40 »	818.262	—	220.301	—
» 40 a 41 »	295.607	—	79.586	—
» 41 a 42 »	622.124	—	167.495	—
» 42 a 43 »	228.450	—	61.506	—
» 43 a 44 »	—	—	—	—
» 44 a 45 »	99.968	99.968	26.914	174.644
	37.142.876	5.724.103	10.000.000	10.000.000

Tav. 15.

DISTRIBUZIONE DELLA MORTALITÀ RISPETTO ALLA POPOLAZIONE.

QUOZIENTI DI MORTALITÀ PER 1000 ABITANTI	POPOLAZIONE TOTALE		POPOLAZIONE TOTALE RIDOTTA A 10.000.000	
	Regno *a*	Circondarî scelti *b*	Regno *c*	Circondarî scelti *d*
Da 13 a 14 escluso	1.325.630	172.424	356.900	301.224
» 14 a 15 »	5.306.050	796.648	1.428.551	1.391.743
» 15 a 16 »	5.650.305	355.192	1.521.235	620.520
» 16 a 17 »	6.557.933	1.886.191	1.765.596	3.295.173
» 17 a 18 »	3.869.397	530.310	1.041.760	926.451
» 18 a 19 »	4.468.872	362.845	1.203.157	633.890
» 19 a 20 »	2.594.550	74.628	698.532	130.375
» 20 a 21 »	2.627.605	544.930	707.432	951.992
» 21 a 22 »	2.700.104	1.000.935	726.951	1.748.632
» 22 a 23 »	829.178	—	223.241	—
» 23 a 24 »	567.360	—	152.751	—
» 24 a 25 »	471.323	—	126.895	—
» 25 a 26 »	102.813	—	27.680	—
» 26 a 27 »	71.756	—	19.319	—
	37.142.876	5.724.103	10.000.000	10.000.000

Tav. 16.

DISTRIBUZIONE DELLA NUZIALITÀ RISPETTO ALLA POPOLAZIONE.

QUOZIENTI DI NUZIALITÀ PER 1000 ABITANTI	POPOLAZIONE TOTALE		POPOLAZIONE TOTALE RIDOTTA A 10.000.000	
	Regno *a*	Circondarî scelti *b*	Regno *c*	Circondarî scelti *d*
Da 7 a 8 escluso	616.618	—	166.012	—
» 8 a 9 »	3.772.383	172.424	1.015.641	301.224
» 9 a 10 »	7.620.532	1.917.423	2.051.681	3.349.736
» 10 a 11 »	12.001.523	2.398.684	3.231.178	4.190.498
» 11 a 12 »	10.194.273	1·125.139	2.744.611	1.965.616
» 12 a 13 »	2.578.512	110·433	694.214	192.926
» 13 a 14 »	328.674	—	88.489	—
» 14 a 15 »	30.361	—	8.174	—
	37.142.876	5.724.103	10.000.000	10.000.000

Tav. 17.

DISTRIBUZIONE DELLA POPOLAZIONE AGRICOLA RISPETTO ALLA POPOLAZ. MASCHILE IN ETÀ SUPERIORE AI 10 ANNI.

POPOLAZIONE AGRICOLA MASCHILE su 100 maschi di età superiore ai 10 anni	Popolazione maschile in età superiore ai 10 anni		Popolazione maschile in età superiore ai 10 anni ridotta a 10.000.000	
	Regno *a*	Circondarî scelti *b*	Regno *c*	Circondarî scelti *d*
Da 0 a 10 escluso	422.450	—	291.438	—
» 10 a 20 »	1.486.383	678.429	1.025.421	2.999.683
» 20 a 30 »	786.989	27.458	542.925	121.406
» 30 a 40 »	1.499.096	—	1.034.191	—
» 40 a 50 »	2.571.147	292.818	1.773.775	1.294.699
» 50 a 60 »	3.462.148	470.612	2.388.456	2.080.817
» 60 a 70 »	3.759.125	649.724	2.593.333	2.872.763
» 70 a 80 »	485.937	142.628	335.237	630.632
» 80 a 90 »	22.068	—	15.224	—
» 90 a 100 »	—	—	—	—
	14.495.343	2.261.669	10.000.000	10.000.000

Tav. 18.

DISTRIBUZIONE DELLA POPOLAZIONE AGGLOMERATA RISPETTO ALLA POPOLAZIONE TOTALE.

POPOLAZIONE AGGLOMERATA per 100 abitanti	Popolazione totale		Popolazione totale ridotta a 10.000.000	
	Regno *a*	Circondarî scelti *b*	Regno *c*	Circondarî scelti *d*
Da 20 a 30 escluso	76.808	76.808	20.679	134.183
» 30 a 40 »	2.818.752	1.069.235	758.895	1.867.952
» 40 a 50 »	3.260.978	137.428	877.955	240.087
» 50 a 60 »	4.549.241	236.119	1.224.795	412.500
» 60 a 70 »	3.370.524	303.981	907.448	531.054
» 70 a 80 »	5.299.239	758.465	1.426.717	1.325.037
» 80 a 90 »	7.503.027	1.227.361	2.020.045	2.144.198
» 90 a 100 »	10.264.307	1.914.706	2.763.466	3.344.989
	37.142.876	5.724.103	10.000.000	10.000.000

Tav. **19.**

DISTRIBUZIONE DEL REDDITO RISPETTO AI REDDITIERI.

REDDITO MEDIO	Redditieri		Redditieri ridotti a 10.000.000	
	Totalità *a*	Gruppo scelto *b*	Totalità *c*	Gruppo scelto *d*
Da 1.000 a 1.500 escluso	20.580	9.678	243.436	819.968
» 1.500 a 2.000 »	73.917	9.110	874.345	771.844
» 2.000 a 2.500 »	119.819	21.319	1.417.309	1.806.251
» 2.500 a 3.000 »	88.192	20.591	1.043.201	1.744.571
» 3.000 a 3.500 »	97.259	19.087	1.150.452	1.617.145
» 3.500 a 4.000 »	87.521	—	1.035.264	—
» 4.000 a 4.500 »	73.738	—	872.228	—
» 4.500 a 5.000 »	41.283	9.099	488.326	770.912
» 5.000 a 5.500 »	18.119	—	214.325	—
» 5.500 a 6.000 »	57.399	—	678.958	—
» 6.000 a 6.500 »	42.716	6.004	505.277	508.689
» 6.500 a 7.000 »	35.021	—	414.254	—
» 7.000 a 7.500 »	—	—	—	—
» 7.500 a 8.000 »	35.961	—	425.374	—
» 8.000 a 8.500 »	24.872	23.141	294.205	1.960.620
» 8.500 a 9.000 »	—	—	—	—
» 9.000 a 9.500 »	—	—	—	—
» 9.500 a 10.000 »	29.001	—	343.046	—
	845.398	118.029	10.000.000	10.000.000

Tav. **20.**

DISTRIBUZIONE DELL'ALTITUDINE RISPETTO ALLA POPOLAZIONE.

ALTITUDINE IN METRI	Popolazione totale		Popolazione totale ridotta a 10.000.000	
	Regno *a*	Circondarî scelti *b*	Regno *c*	Circondarî scelti *d*
Fino a 50 escluso	14.916.664	3.410.821	4.016.023	5.958.700
Da 50 a 100	4.445.880	138.991	1.196.967	242.817
» 100 a 150	3.827.298	334.948	1.030.426	585.154
» 150 a 200	1.545.086	—	415.984	—
» 200 a 250	2.295.936	—	618.136	—
» 250 a 300	1.766.639	513.080	475.633	896.350
» 300 a 350	1.935.439	134.737	521.079	235.385
» 350 a 400	1.552.084	213.980	417.868	373.823
» 400 a 450	875.653	37.460	235.753	65.443
» 450 a 500	784.019	119.111	211.082	208.087
» 500 a 550	538.100	86.905	144.873	151.823
» 550 a 600	859.772	382.398	231.477	668.049
» 600 a 650	444.058	73.104	119.554	127.712
» 650 a 700	352.889	127.132	95.009	222.099
» 700 a 750	504.503	—	135.828	—
» 750 a 800	—	—	—	—
» 800 a 850	252.040	151.436	67.857	264.558
» 850 a 900	158.324	—	42.626	—
» 900 a 950	—	—	—	—
» 950 a 1000	88.492	—	23.825	—
	37.142.876	5.724.103	10.000.000	10.000.000

Tav. **21.**

DISTRIBUZIONE DELLA DENSITÀ RISPETTO ALLA POPOLAZIONE.

ABITANTI PER CHILOMETRO QUADRATO	POPOLAZIONE TOTALE		POPOLAZIONE TOTALE RIDOTTA A 10.000.000	
	Regno *a*	Circondarî scelti *b*	Regno *c*	Circondarî scelti *d*
Da 0 a 25 escluso	335.744	78.798	90.393	137.660
» 25 a 50 »	1.229.542	470.445	331.030	821.867
» 50 a 75 »	3.793.007	703.023	1.021.194	1.228.180
» 75 a 100 »	3.175.466	451.261	854.933	788.352
» 100 a 125 »	3.940.141	350.522	1.060.807	612.361
» 125 a 150 »	2.693.619	266.748	725.205	466.008
» 150 a 175 »	4.023.589	172.424	1.083.273	301.225
» 175 a 200 »	2.527.371	138.991	680.446	242.817
» 200 a 225 »	5.423.068	1.293.384	1.460.056	2.259.540
» 225 a 250 »	2.651.380	552.807	713.833	965.753
» 250 a 275 »	—	—	—	—
» 275 a 300 »	728.291	588.043	196.078	1.027.311
» 300 a 325 »	1.014.712	—	273.191	—
» 325 a 350 »	448.880	—	120.852	—
» 350 in più	5.158.066	657.657	1.388.709	1.148.926
	37.142.876	5.724.103	10.000.000	10.000.000

NB. — Per ragione di spazio si sono qui raccolti in un'unica classe, sia per il Regno, che per il gruppo scelto, i Circondarî aventi densità di popolazione superiore a 350; ma per il calcolo dell'indice di dissomiglianza si è conferito a ciascuno di tali circondarî il posto che gli compete nella graduatoria secondo la densità, con la sola licenza di conglobare alcuni circondarî costituenti un gruppo di densità media 650-675, e cioè i circondarî di Salerno, Messina, Nola, Gallarate, Pozzuoli, Genova, Monza, Castellamare di Stabia, Casoria, Livorno, Milano.

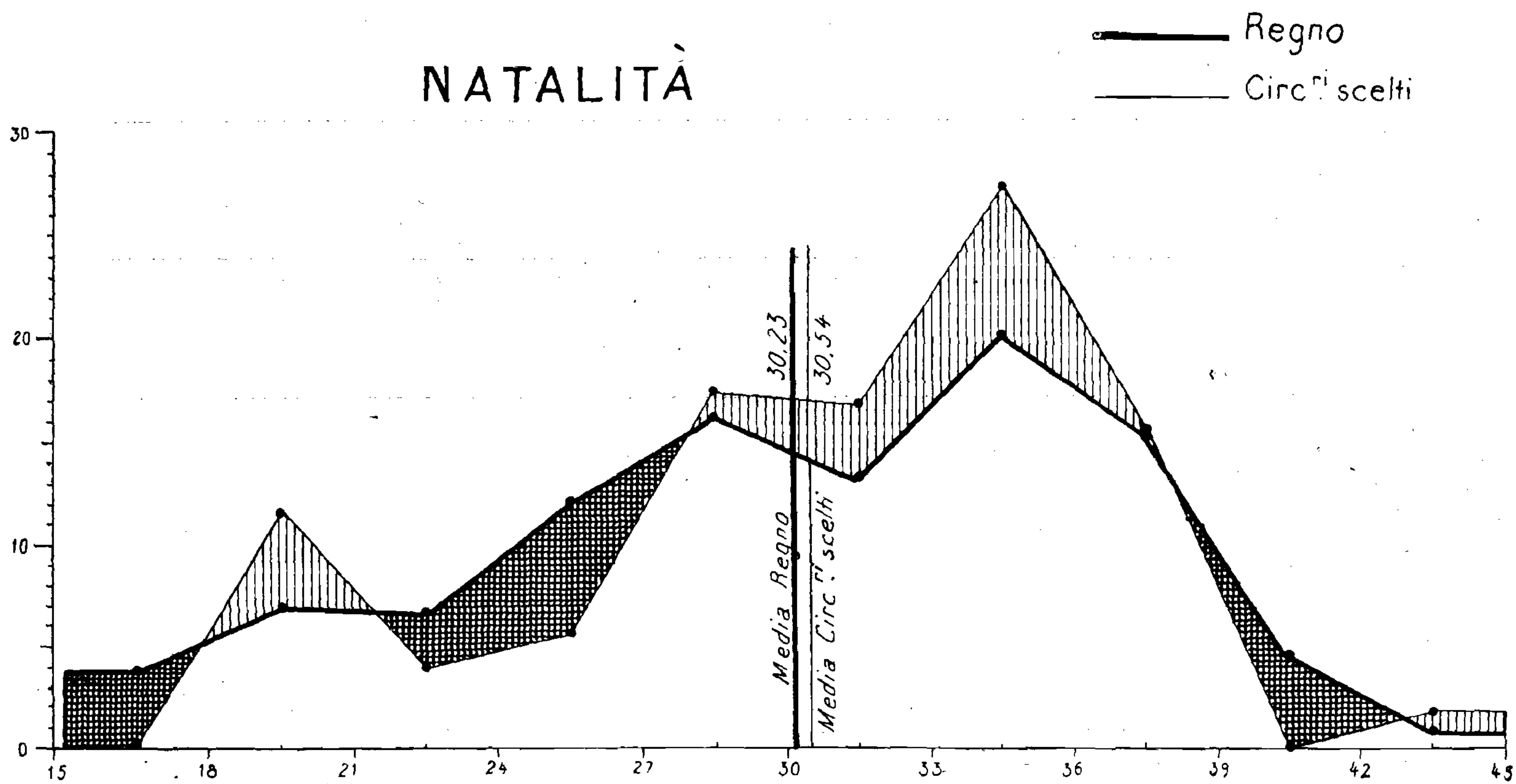
Tav. 22.

DISTRIBUZIONE DELL'ACCRESCIMENTO NATURALE RISPETTO ALLA POPOLAZIONE.

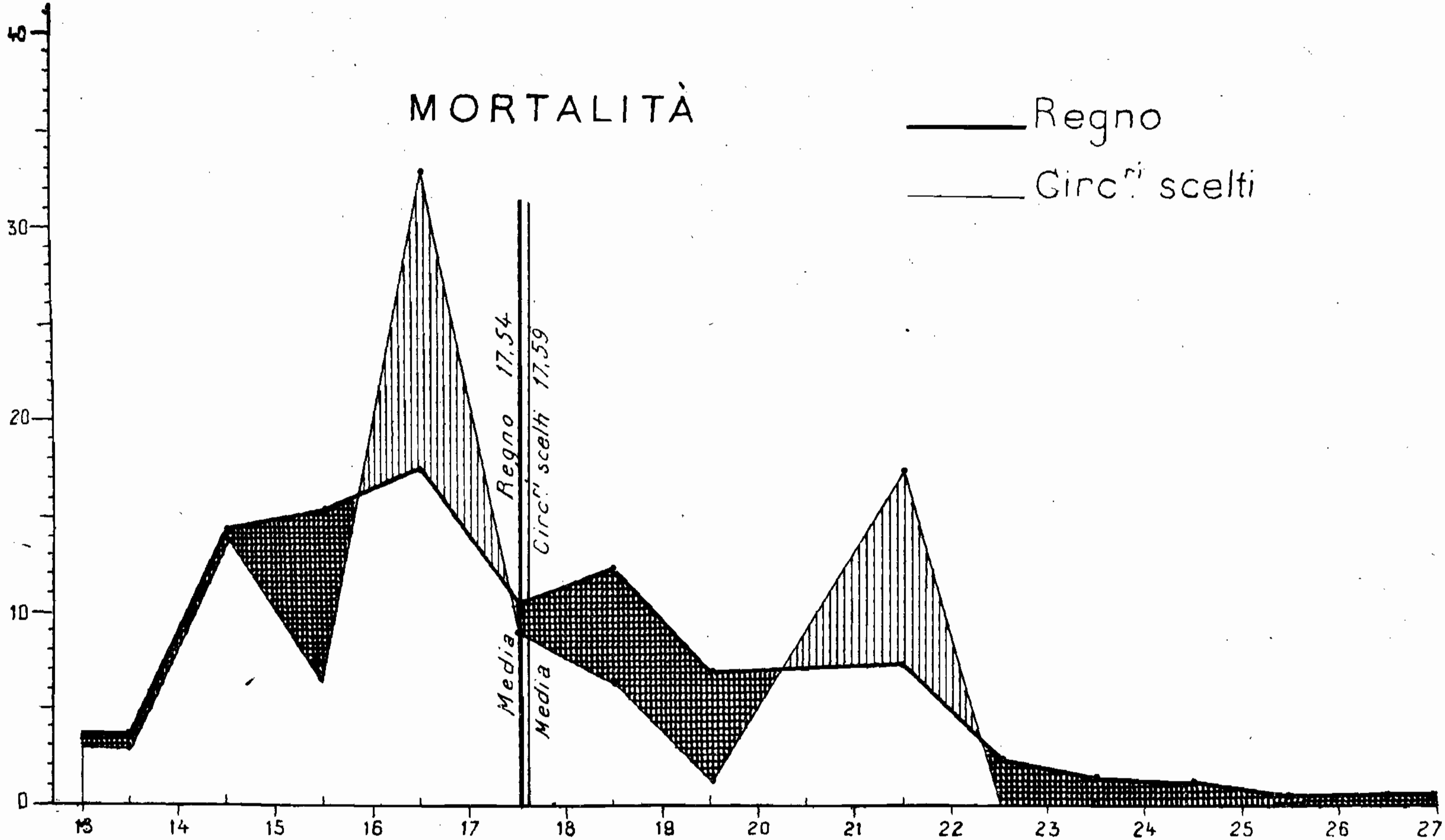
QUOZIENTI DI ACCRESCIMENTO PER 1000 ABITANTI	POPOLAZIONE TOTALE		POPOLAZIONE TOTALE RIDOTTA A 10.000.000	
	Regno *a*	Circondari scelti *b*	Regno *c*	Circondari scelti *d*
Da 1 a 2	1.098.274	—	295.689	—
» 2 a 3	303.483	—	81.707	—
» 3 a 4	855.554	—	230.341	—
» 4 a 5	1.037.095	657.657	279.218	1.148.926
» 5 a 6	1.206.720	86.905	324.886	151.823
» 6 a 7	1.389.235	138.991	374.025	242.817
» 7 a 8	680.284	—	183.153	—
» 8 a 9	1.468.824	152.742	395.452	266.840
» 9 a 10	2.272.122	131.184	611.725	229.178
» 10 a 11	3.094.468	958.436	833.126	1.674.386
» 11 a 12	1.883.136	49.319	506.998	86.160
» 12 a 13	2.439.713	334.948	656.845	585.154
» 13 a 14	1.854.047	284.512	499.166	497.042
» 14 a 15	3.028.882	795.119	815.468	1.389.072
» 15 a 16	3.056.449	696.991	822.890	1.217.642
» 16 a 17	3.667.951	372.090	987.525	650.041
» 17 a 18	3.105.426	300.390	836.076	524.781
» 18 a 19	2.403.626	588.043	647.130	1.027.310
» 19 a 20	633.837	—	170.648	—
» 20 a 21	806.645	76.808	217.174	134.184
» 21 a 22	120.448	—	32.428	—
» 22 a 23	606.328	—	163.242	—
» 23 a 24	99.968	99.968	26.914	174.644
» 24 a 25	—	—	—	—
» 25 a 26	30.361	—	8.174	—
	37.142.876	5.724.103	10.000.000	10.000.000

Tav. 23.

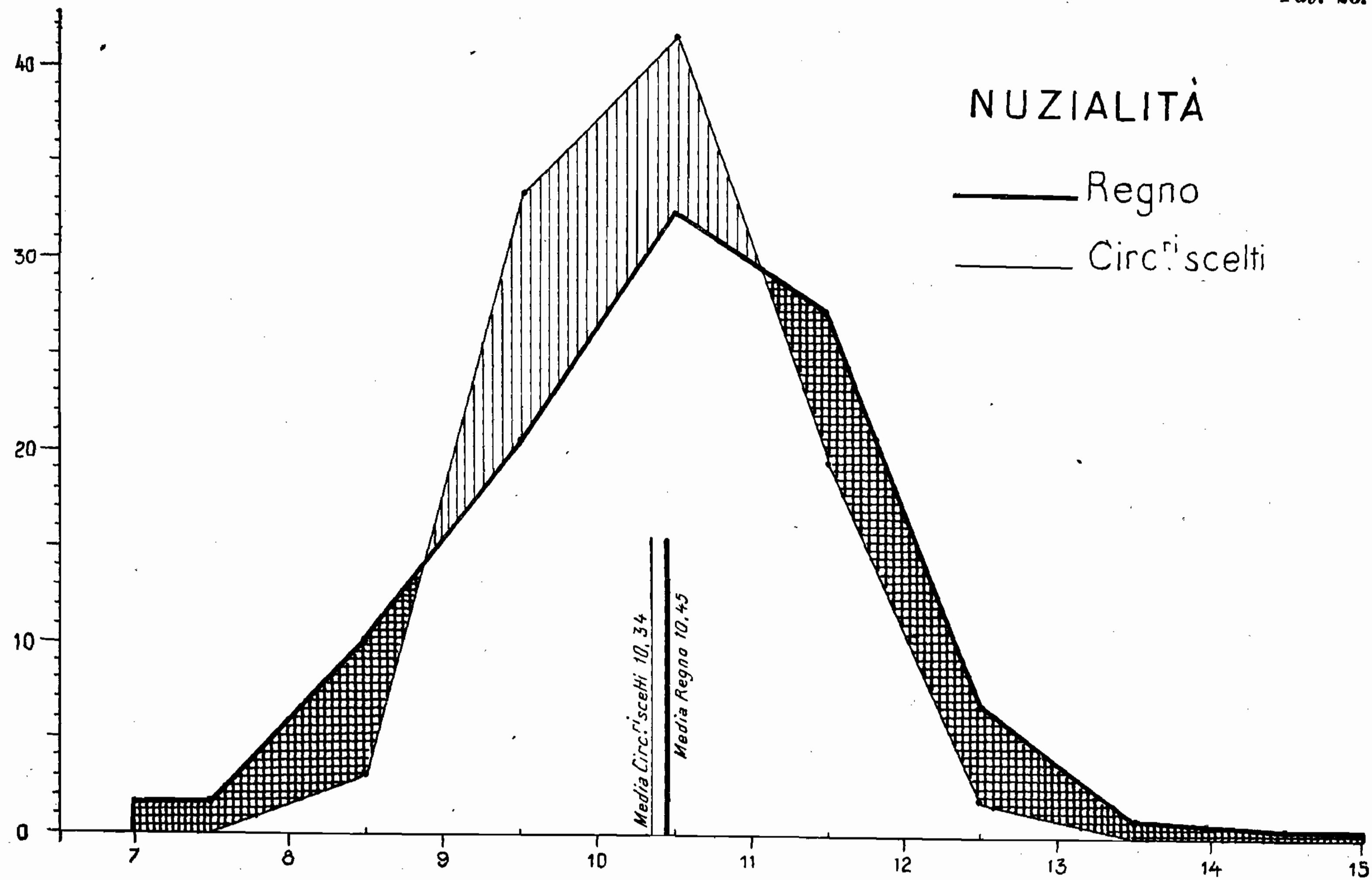
NATALITÀ



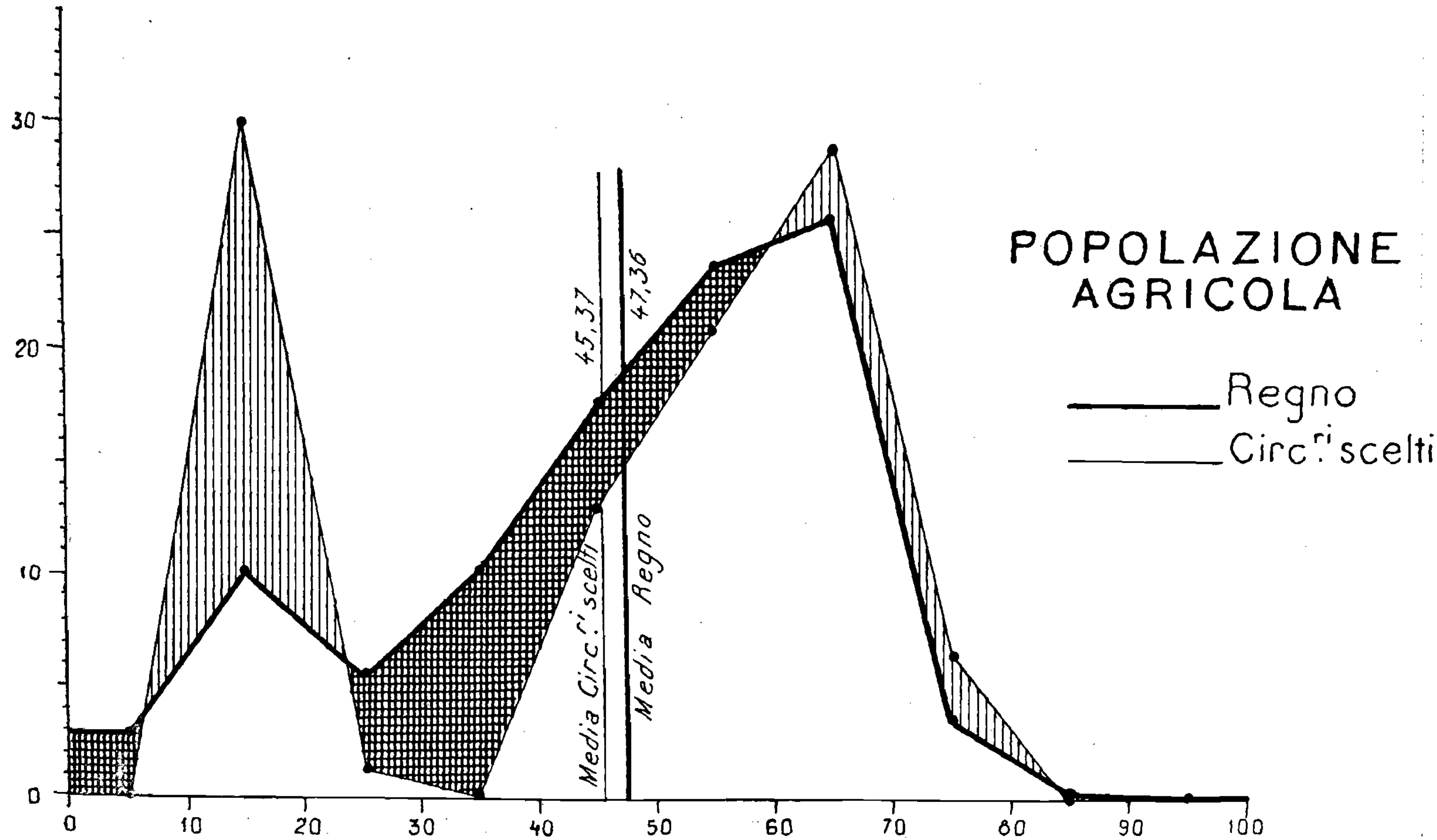
Tav. **24.**


MORTALITÀ
Regno
Circ.ri scelti
Media Regno 17,54
Media Circ.ri scelti 17,59
0
10
20
30
40
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27

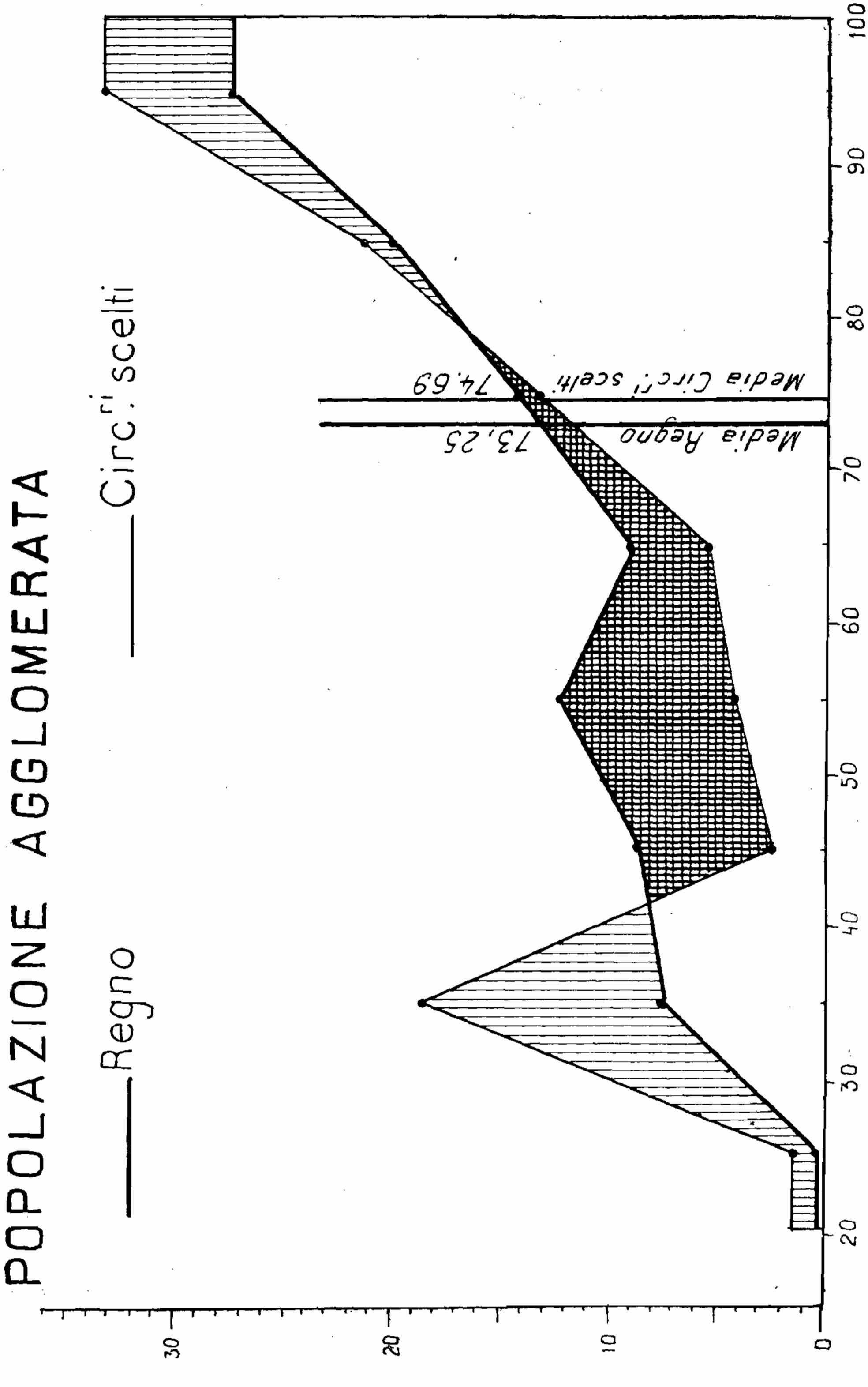
Tav. 25.


NUZIALITÀ
Regno
Circ.ri scelti
Media Circ.ri scelti 10,34
Media Regno 10,45
40
30
20
10
0
7
8
9
10
11
12
13
14
15

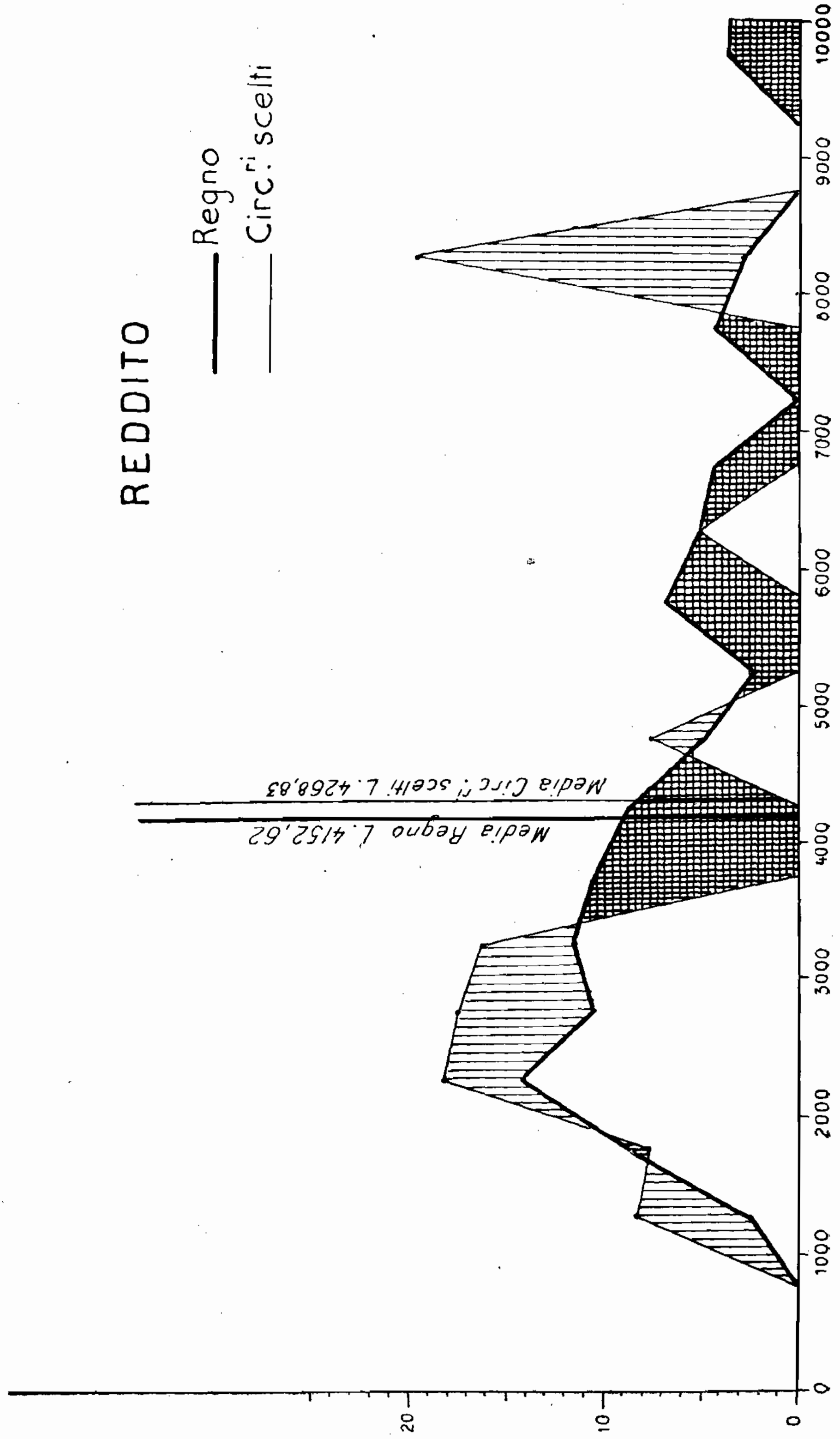
Tav. 26.


POPOLAZIONE
AGRICOLA
Regno
Circ.ri scelti
Media Circ.ri scelti
45,37
Media Regno
47,36
0
10
20
30
0
10
20
30
40
50
60
70
80
90
100

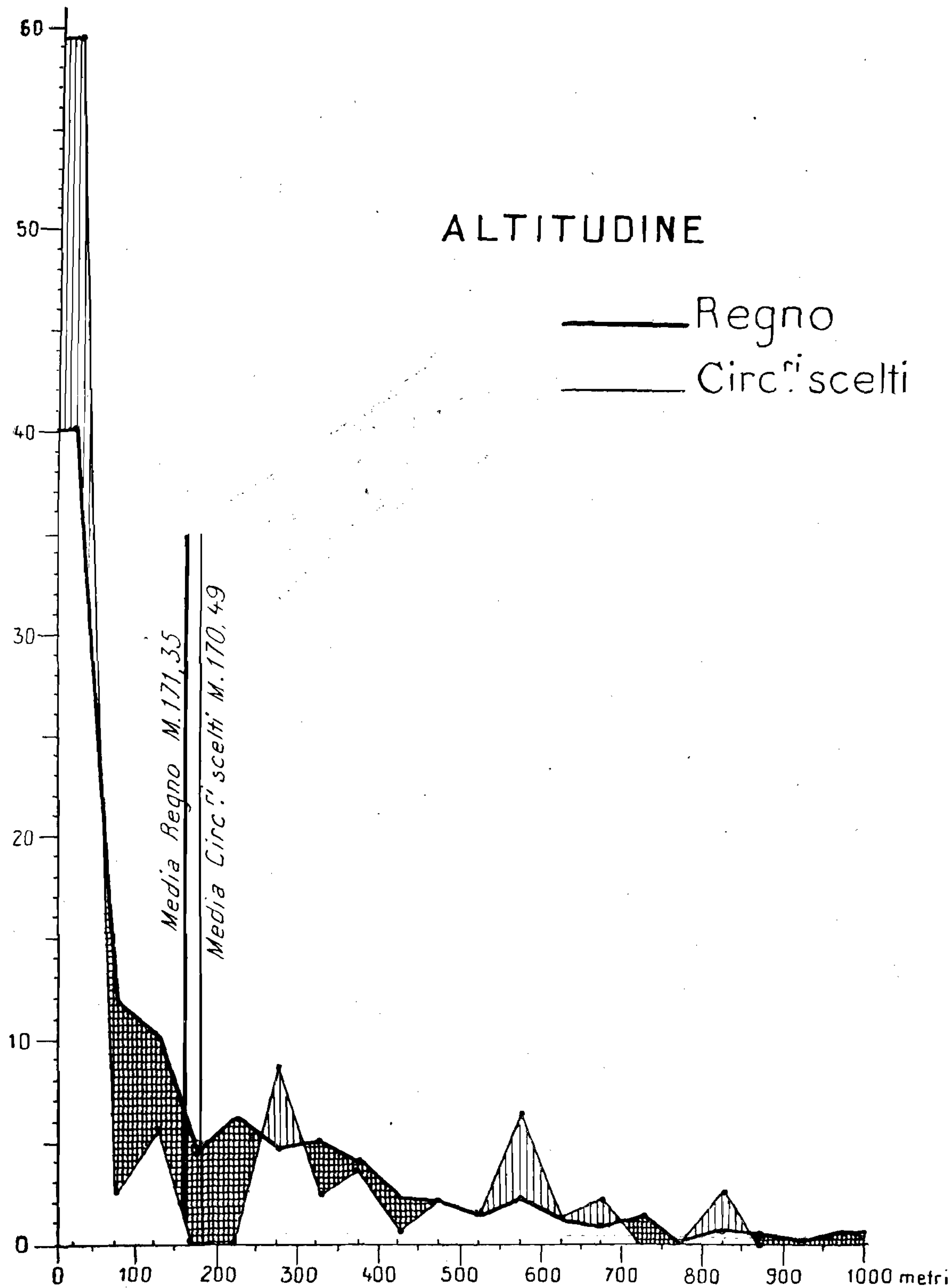
Tav. 27.


POPOLAZIONE AGGLOMERATA
Regno
Circ.ri scelti
Media Circ.ri scelti 74,69
Media Regno 73,25
20
30
40
50
60
70
80
90
100
0
10
20
30

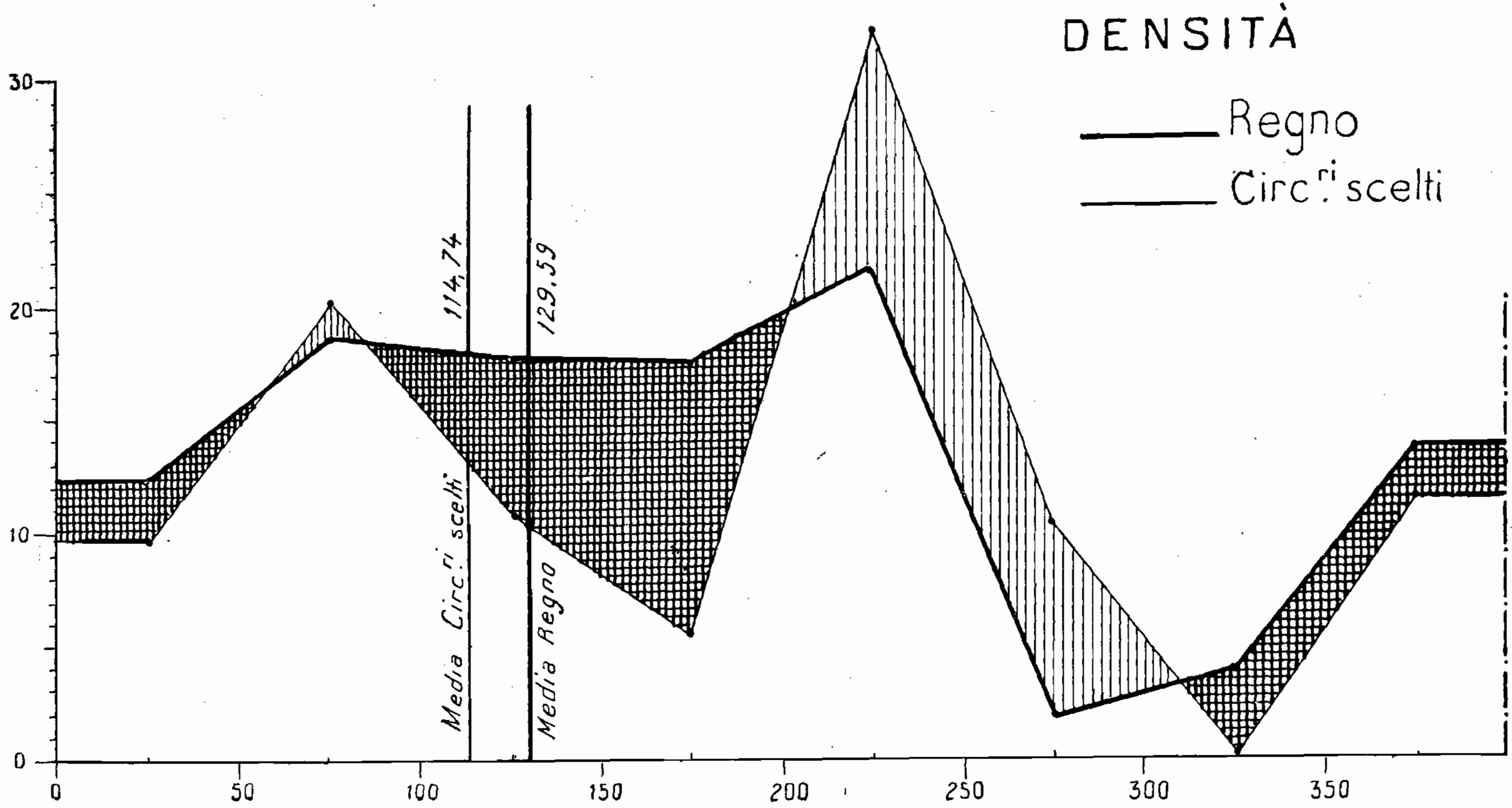
Tav. 28.


REDDITO
Regno
Circ.ri scelti
Media Circ.ri scelti L. 4268,83
Media Regno L. 4152,62
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
0
10
20

Tav. 29.

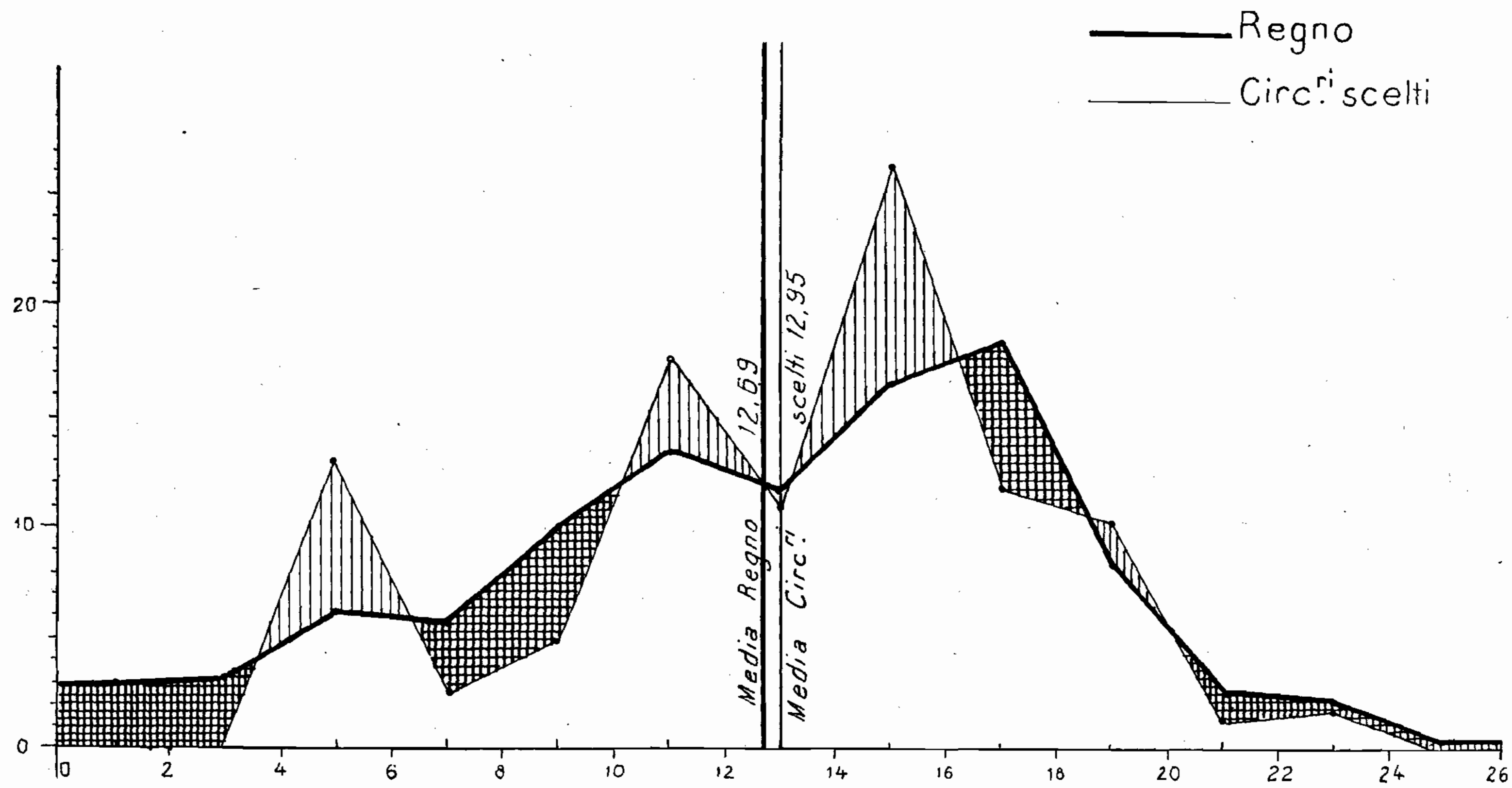

ALTITUDINE
Regno
Circ.ri scelti
Media Regno M.171,35
Media Circ.ri scelti M.170,49
0
10
20
30
40
50
60
0
100
200
300
400
500
600
700
800
900
1000 metri

Tav. 30.


DENSITÀ
Regno
Circ.ri scelti
Media Circ.ri scelti 114,74
Media Regno 129,59
0
10
20
30
0
50
100
150
200
250
300
350

Tav. 31.

ACCRESCIMENTO NATURALE



Tav. 32.

INDICE DI DISSOMIGLIANZA TRA LA DISTRIBUZIONE NELLA TOTALITÀ E LA DISTRIBUZIONE NEL CAMPIONE DEI DIVERSI CARATTERI.

CARATTERI	Indice assoluto di dissomiglianza D	Indice relativo di dissomiglianza	
		$\frac{D}{A}$ (1)	$\frac{D}{A \frac{L-A}{L}}$ (1)
1. Natalità	0,926 ‰	—	3,2 %
2. Mortalità	0,455 ‰	—	2,6 %
3. Nuzialità	0,354 ‰	—	3,4 %
4. Rapporto fra la popolazione agricola maschile e la popolazione maschile di più di 10 anni	4,570 %	—	18,3 %
5. Rapporto fra la popolazione agglomerata e la popolazione totale	4,039 %	—	20,6 %
6. Reddito medio per contribuente (redditi di lavoro o misti)	321,57	7,7 %	—
7. Altitudine media sul livello del mare	36,30	21,2 %	—
8. Densità della popolazione (abitanti per Km²)	120,66	93,1 %	—
9. Accrescimento naturale (1-2)	0,655 ‰	—	5,2 %

(1) Per i valori di A e di L vedere nota alla Tav. 10.

Tav. **33.**

CONFRONTO FRA GLI INDICI DI CORRELAZIONE E GLI INDICI DI OMOFILIA NELL'INSIEME E NEL CAMPIONE, CALCOLATI PER LE DIVERSE COPPIE DI CARATTERI.

COPPIE DI CARATTERI		INDICI DI CORRELAZIONE				INDICI DI OMOFILIA			
		Nell'insieme r_a	Nel campione r_e	Scarto assoluto $r_e - r_a$	Scarto relativo $\frac{r_e - r_a}{r_a(1-\|r_a\|)}$	Nell'insieme ω_a	Nel campione ω_e	Scarto assoluto $\omega_e - \omega_a$	Scarto relativo $\frac{\omega_e - \omega_a}{\omega_a(1-\|\omega_a\|)}$
Natalità –	Mortalità.......	0,642	0,713	0,071	0,309	0,668	0,779	0,111	0,500
	Nuzialità.......	0,063	0,071	0,008	0,136	0,064	0,072	0,008	0,133
	Pop. agricola...	0,396	0,398	0,002	0,008	0,405	0,424	0,019	0,075
	Pop. agglomerata	— 0,065	— 0,166	— 0,101	1,683	— 0,069	— 0,190	— 0,121	1,890
	Reddito........	— 0,350	— 0,439	— 0,089	0,392	— 0,364	— 0,491	— 0,127	0,548
	Altitudine......	0,165	0,154	— 0,011	— 0,080	0,187	0,176	— 0,011	— 0,072
	Densità	— 0,093	— 0,373	— 0,280	3,333	— 0,182	— 0,443	— 0,261	1,752
	Accr. naturale..	0,909	0,916	0,007	0,084	0,911	0,923	0,012	0,148
Mortalità –	Nuzialità.......	— 0,061	— 0,026	0,035	— 0,614	— 0,062	— 0,028	0,034	— 0,586
	Pop. agricola...	0,228	0,157	— 0,071	— 0,402	0,253	0,184	— 0,069	— 0,365
	Pop. agglomerata	0,396	0,286	— 0,110	— 0,460	0,422	0,305	— 0,117	— 0,479
	Reddito........	— 0,403	— 0,238	0,165	— 0,684	— 0,462	— 0,295	0,167	— 0,671
	Altitudine......	0,301	0,182	— 0,119	— 0,567	0,309	0,190	— 0,119	— 0,556
	Densità	— 0,054	— 0,284	— 0,230	4,510	— 0,130	— 0,372	— 0,242	2,142
	Accr. naturale..	0,265	0,371	0,106	0,544	0,276	0,399	0,123	0,615
Nuzialità –	Pop. agricola...	0,174	0,446	0,272	1,889	0,180	0,492	0,312	2,108
	Pop. agglomerata	— 0,274	— 0,368	— 0,094	0,472	— 0,292	— 0,392	— 0,100	0,483
	Reddito........	— 0,010	— 0,400	— 0,390	39,000	— 0,011	— 0,463	— 0,452	41,091
	Altitudine......	0,178	0,540	0,362	2,480	0,197	0,581	0,384	2,430
	Densità	— 0,167	— 0,524	— 0,357	2,568	— 0,323	— 0,646	— 0,323	1,477
	Accr. naturale..	0,113	0,109	— 0,004	— 0,040	0,114	0,110	— 0,004	— 0,040
Pop. agr. –	Pop. agglomerata	— 0,111	— 0,330	— 0,219	2,235	— 0,128	— 0,436	— 0,308	2,753
	Reddito........	— 0,614	— 0,875	— 0,261	1,101	— 0,619	— 0,902	— 0,283	1,199
	Altitudine......	0,297	0,049	— 0,248	— 1,187	0,377	0,065	— 0,312	— 1,328
	Densità	— 0,456	— 0,681	— 0,225	0,907	— 0,766	— 0,754	0,012	— 0,067
	Accr. naturale..	0,362	0,436	0,074	0,320	0,368	0,467	0,099	0,425
Pop. aggl. –	Reddito........	— 0,078	0,111	0,189	— 2,625	— 0,082	0,159	0,241	— 3,213
	Altitudine......	0,148	0,037	— 0,111	— 0,881	0,174	0,041	— 0,133	— 0,924
	Densità	0,115	0,061	— 0,054	— 0,530	0,310	0,093	— 0,217	— 1,014
	Accr. naturale..	— 0,296	— 0,383	— 0,087	0,416	— 0,319	— 0,430	— 0,111	0,512
Reddito –	Altitudine......	— 0,346	— 0,412	— 0,066	0,292	— 0,459	— 0,602	— 0,143	0,575
	Densità	0,330	0,834	0,504	2,280	0,510	0,867	0,357	1,428
	Accr. naturale..	— 0,221	— 0,445	— 0,224	1,302	— 0,229	— 0,500	— 0,271	1,531
Altitud. –	Densità	— 0,211	— 0,476	— 0,265	1,587	— 0,676	— 0,724	— 0,048	0,219
	Accr. naturale..	0,043	0,099	0,056	1,366	0,049	0,112	0,063	1,340
Densità –	Accr. naturale..	— 0,087	— 0,313	— 0,226	2,825	— 0,171	— 0,374	— 0,203	1,435

Tav. **34.**

COEFFICIENTI ANGOLARI E ORDINATE ALL'ORIGINE DELLE RETTE INTERPOLATRICI.

CARATTERI PARAGONATI		Nella totalità		Nel gruppo scelto	
		Coefficiente angolare	Ordinata all' origine	Coefficiente angolare	Ordinata all' origine
Natalità........	– Mortalità	0,288	9,282	0,308	8,172
	- Nuzialità	0,012	10,235	0,012	10,321
	Popolazione agricola...	0,939	25,142	1,136	21,358
	Popolaz. agglomerata..	— 0,202	78,939	— 0,636	91,984
	Reddito..............	— 85,278	5.773,748	—118,399	6.389,135
	Altitudine	6,166	40,039	7,110	64,801
	Densità	— 4,982	330,239	— 8,463	400,536
	Accrescimento naturale.	0,722	— 9,286	0,692	— 8,178
Mortalità	– Nuzialità	— 0,027	11,089	— 0,010	10,895
	Popolazione agricola...	1,251	31,931	1,040	39,070
	Popolaz. agglomerata..	2,863	21,326	2,544	25,585
	Reddito..............	—227,321	7.195,475	—148,331	5.265,139
	Altitudine	26,008	— 234,364	19,479	— 58,695
	Densità	— 6,713	295,551	— 14,890	397,646
	Accrescimento naturale.	0,487	4,472	0,649	2,331
	Natalità	1,485	4,493	1,649	2,335
Nuzialità.......	– Popolazione agricola...	2,172	31,311	7,488	— 22,311
	Popolaz. agglomerata..	— 4,498	120,370	— 8,301	160,424
	Reddito..............	— 13,148	3.259,205	—633,020	9.363,069
	Altitudine	34,940	— 138,710	146,436	— 1.274,683
	Densità	— 47,325	677,243	— 69,748	875,426
	Accrescimento naturale.	0,470	8,206	0,482	8,900
	Natalità	0,332	27,605	0,414	27,692
	Mortalità	— 0,139	19,399	— 0,067	18,785
Popolaz. agricola	– Popolaz. agglomerata..	— 0,146	80,581	— 0,444	97,220
	Reddito..............	— 63,117	6.550,540	— 82,559	7.362,778
	Altitudine............	4,673	— 22,052	7,981	— 168,635
	Densità	— 10,335	736,960	— 5,406	441,469
	Accrescimento naturale.	0,121	6,598	0,115	7,379
	Natalità	0,167	22,048	0,139	24,065
	Mortalità	0,042	15,668	0,026	16,615
	Nuzialità	0,014	9,849	0,027	9,167

Segue *Tav* **34**..

CARATTERI PARAGONATI		Nella totalità		Nel gruppo scelto	
		Coefficiente angolare	Ordinata all'origine	Coefficiente angolare	Ordinata all'origine
Popol. agglomer. –	Reddito	— 6,080	3.561,485	7,770	2.029,264
	Altitudine	1,768	103,476	0,449	261,144
	Densità	1,974	31,751	0,358	103,041
	Accrescimento naturale	— 0,075	18,664	— 0,075	19,457
	Natalità	— 0,021	32,622	— 0,043	35,215
	Mortalità	0,055	13,950	0,032	15,762
	Nuzialità	— 0,017	11,817	— 0,016	11,875
	Popolazione agricola	— 0,084	60,447	— 0,245	75,426
Reddito –	Altitudine	— 0,053	397,488	— 0,071	475,903
	Densità	0,073	— 52,160	0,070	— 52,736
	Accrescimento naturale	— 0,001	15,439	— 0,001	17,288
	Natalità	— 0,001	35,600	— 0,002	36,343
	Mortalità	— 0,001	20,162	— 0,0004	19,054
	Nuzialità	— 0,000008	10,633	— 0,0003	11,360
	Popolazione agricola	— 0,006	72,992	— 0,009	81,836
	Popolaz. agglomerata	— 0,001	75,779	0,002	67,464
Altitudine –	Densità	— 0,304	245,710	— 0,234	197,150
	Accrescimento naturale	0,001	12,981	0,002	13,584
	Natalità	0,004	30,097	0,003	31,151
	Mortalità	0,003	17,121	0,002	17,567
	Nuzialità	0,001	10,398	0,002	10,124
	Popolazione agricola	0,019	49,988	0,012	48,922
	Popolaz. agglomerata	0,012	69,782	0,003	70,469
	Reddito	— 2,260	3.643,894	— 2,405	3.289,138
Densità –	Accrescimento naturale	— 0,001	13,421	— 0,010	15,401
	Natalità	— 0,002	31,424	— 0,016	34,247
	Mortalità	— 0,0004	18,006	— 0,005	18,764
	Nuzialità	— 0,001	10,712	— 0,004	11,213
	Popolazione agricola	— 0,020	57,877	— 0,086	68,905
	Popolaz. agglomerata	0,007	71,487	0,010	70,225
	Reddito	1,497	2.857,370	9,915	1.309,917
	Altitudine	— 0,147	257,646	— 0,970	418,035
Accrescim. natur. –	Natalità	1,142	16,025	1,212	15,086
	Mortalità	0,144	16,027	0,212	15,082
	Nuzialità	0,027	10,253	0,024	10,363
	Popolazione agricola	1,081	40,093	1,649	34,683
	Popolaz. agglomerata	— 1,166	88,042	— 1,947	98,926
	Reddito	— 67,696	4.012,950	— 158,773	4.817,632
	Altitudine	2,033	205,122	6,079	207,767
	Densità	— 5,911	253,196	— 9,380	260,503

Tav. 35.

ACCORDO INTERCEDENTE FRA LA TOTALITÀ DEL REGNO (ESCLUSE LE NUOVE PROVINCIE) E IL GRUPPO DEI 29 CIRCONDARI SCELTI PER QUANTO CONCERNE ALCUNI ASPETTI DEI DIVERSI CARATTERI CONSIDERATI E LE LORO MUTUE RELAZIONI

(L'accordo ottimo viene indicato con 3; soddisfacente con 2; sufficiente con 1; insufficiente con 0.

A) Aspetti dei singoli caratteri.

CARATTERI	Intensità (valori medî)	Variabilità (ind. di concentr.)	Distribuzioni (ind. di dissom.)
1. Natalità	3	1	2
2. Mortalità	3	1	2
3. Nuzialità	3	0	2
4. Rapporto della popolazione agricola maschile alla popol. maschile di più di 10 anni	1	0	0
5. Rapporto della popolazione agglomerata alla popolazione totale	1	0	0
6. Reddito medio per contribuente	2	0	1
7. Altitudine	3	0	0
8. Densità per Km^2	0	0	0
9. Accrescimento naturale	2	0	1

B) Relazioni mutue dei caratteri.

CARATTERI	Natalità	Mortalità	Nuzialità	Popol. agricola	Popol. agglomerata	Reddito	Altitudine	Densità	Accrescimento naturale
1. Natalità	=		0	3	0	0	1	0	1
2. Mortalità	0	=	0	0	0	0	0	0	0
3. Nuzialità	0	0	=	0	0	0	0	0	2
4. Popolazione agricola	1	0	0	=	0	0	0	0	0
5. Popolazione agglomerata	0	0	0	0	=	0	0	0	0
6. Reddito	0	0	0	0	0	=	0	0	0
7. Altitudine	1	0	0	0	0	0	=	0	0
8. Densità	0	0	0	1	0	0	0	=	0
9. Accrescimento naturale	0	0	2	0	0	0	0	0	=

NB. — L'accoppiamento dei caratteri viene eseguito come in una tavola pitagorica; il grado di accordo fra gli indici di correlazione è indicato superiormente alla diagonale, e quello fra gli indici di omofilia inferiormente.

Tav. 36.

ELENCO DI 47 CITTÀ E DEI CORRISPONDENTI VALORI DI ALCUNI CARATTERI

(Le città costituenti il gruppo scelto sono indicate in carattere grassetto).

CITTÀ	*a* (Popolazione in migliaia)	*u*	*v*	*w*	*x*
1. Barrow	64	33	9	20	41
2. Bedford	**39**	**27**	**8**	**20,25**	**38,25**
3. Birmingham	**840**	**36**	**10**	**21**	**40**
4. Blackburn	133	33,5	9,5	25	40
5. Bolton	181	35	9,5	20	40
6. Bradford	288	34	9	19,5	40
7. Bristol	357	33	9,5	19	36
8. Burneley	107	32,5	9	20	40
9. Carlisle	46	30	9	20,5	36,5
10. Chester	39	31,5	9	18	39,5
11. Coventry	106	34	9	22,5	36
12. Derby	123	34	9	19	40
13. Gateshead	117	35	9,5	22,5	39
14. Glocester	50	29	7,5	19	36,5
15. Grinsby	**75**	**29**	**9**	**21**	**40**
16. Halifax	102	32,5	9	19	38
17. Huddersfield	**108**	**30**	**9**	**20,5**	**38**
18. Hull	278	32	9,75	23	40
19. Jarrow	34	32,5	9,5	22,5	38
20. Keighley	43	30	8,5	20	35
21. Leicester	**227**	**34,5**	**9**	**20**	**38**
22. Lincoln	57	30	8,5	20	32

Segue *Tav.* **36.**

CITTÀ	a (Popolazione in migliaia)	u	v	w	x
23. Liverpool	816	38	10	20	40
24. Leeds	446	35,5	9,5	22,5	40,5
25. Manchester	946	36,5	9,5	20	42
26. Middlesbrough	105	33	9	22	40,5
27. Newcastle	267	35	9,5	21,5	39
28. Newport	84	33	9	19	38
29. Nottingham	**260**	**35,5**	**9,5**	**19**	**39**
30. Oldham	147	34	9,5	20	41
31. Preston	**117**	**33,5**	**9,5**	**20**	**40**
32. Rockdale	91	34	9,5	20	40
33. Belfast	387	35	8,5	16,5	40
34. Sheffield	460	35	9,5	22	42
35. South Shields	109	32,5	9,75	22,5	38
36. Stockport	109	33,75	9,5	20	42
37. Stockton	52	32	9	21	38,5
38. Stoke	**235**	**32**	**8,75**	**18**	**35**
39. Sunderland	151	33	9,5	22	38
40. Wigan	89	33,5	9,5	19	40
41. Aberdeen	164	34,5	8,5	20,5	39,375
42. Dundee	**165**	**35**	**9,75**	**19,75**	**39**
43. Edinburgh	401	34	9	20,25	39,375
44. Glasgow	**784**	**36**	**9**	**19,125**	**41,625**
45. Greenock	75	36	9	19,125	41,625
46. Kilmanock	35	33,5	8,5	20,5	38
47. Paisley	84	36	9	19,5	41,625