

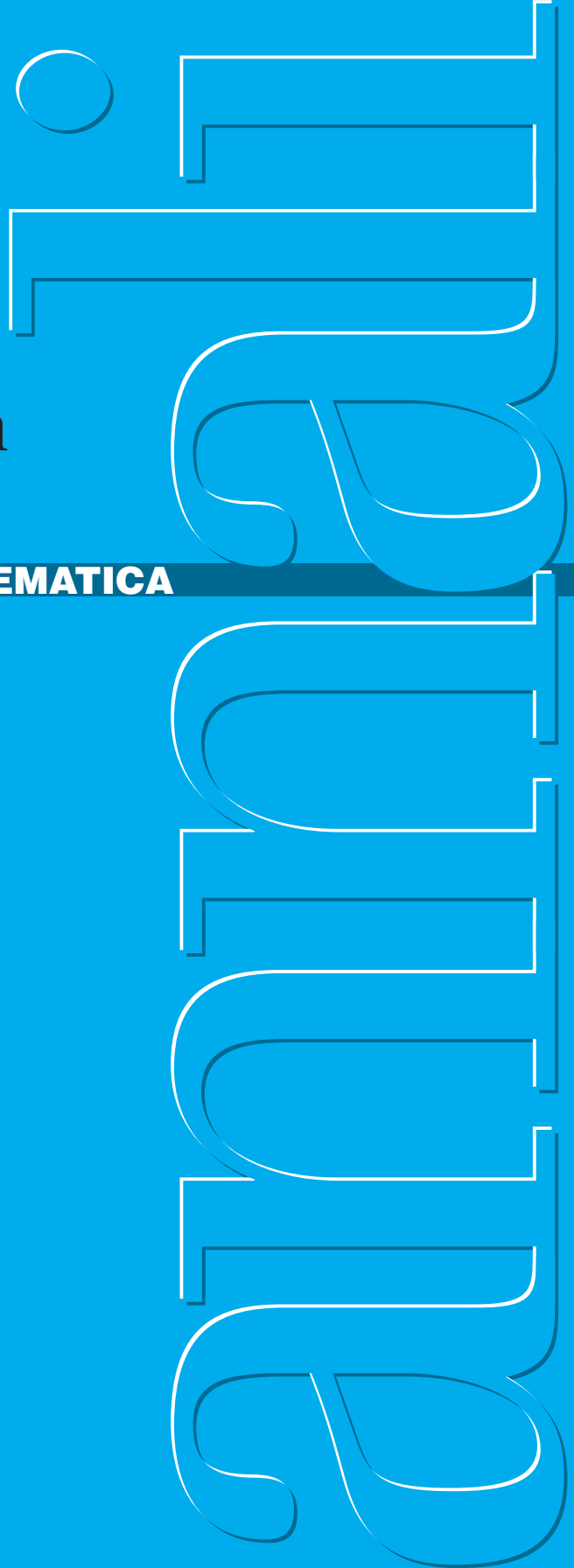


UNIVERSITÀ DEGLI STUDI
DI PALERMO

Annali della Facoltà
di Economia

AREA **STATISTICO-MATEMATICA**

2007
ANNO LXI





Università degli Studi di Palermo

Annali della Facoltà di Economia

AREA **STATISTICO-MATEMATICA**

2007
ANNO LXI

PROPRIETÀ LETTERARIA RISERVATA

Gli articoli riflettono esclusivamente le opinioni dei rispettivi Autori

PRESIDE DELLA FACOLTÀ
Prof. CARLO DOMINICI

DIRETTORE SCIENTIFICO
Prof. VINCENZO LO IACONO

DIRETTORE RESPONSABILE
Prof. GIUSEPPE INGRASSIA

COMITATO SCIENTIFICO
Prof. F. ANDRIA - L. BELLAVISTA VIANELLI
P. BUSETTA - V. CAPURSI - M. CHIODI - G. CUSIMANO
R. GIAIMO - G. INGRASSIA - S. LA ROSA
V. LO IACONO - G. LOVISON - F. VACCINA

ANNALI DELLA FACOLTÀ DI ECONOMIA - UNIVERSITÀ DI PALERMO

RIVISTA SCIENTIFICA

Aderente al Centro Italiano ISSN e all'Agenzia ISBN

ISSN 1827-8388

Iscrizione al Tribunale di Palermo

27 luglio 2005

EDITORE

Facoltà di Economia - Biblioteca Centrale

DIREZIONE E REDAZIONE

Facoltà di Economia
Viale delle Scienze, Ed. 13
90128 Palermo
E-mail: ingra@unipa.it

INDICE

MARIA CALIRI - <i>Giuseppe Bagnera e la matematica finanziaria</i>	pag. 9
MARIA DAVÌ, GIUSEPPE NOTARSTEFANO, ERASMO VASSALLO - <i>Recente evoluzione della struttura produttiva siciliana: una nota con riferimento agli 82 sistemi locali del lavoro</i>	» 29
STEFANO DE CANTIS, ANNA MARIA TAORMINA - <i>Dichotomizing continuous prognostic variables and minimum p-value approach: the prognostic role of age in hodgkin's disease</i>	» 71
FRANCESCO GARGANO - <i>Una stima statistica, effettuata mediante i metodi di ricampionamento Jackknife e Bootstrap, dei confini della normalità o regolarità delle oscillazioni delle frequenze delle misure di un fenomeno quantitativo collegabile ad un fenomeno qualitativo</i>	» 83
MARIA GIUSEPPA MANGANO - <i>Offerta e domanda di lavoro in Sicilia: un'indagine sulle forze di lavoro</i> ...	» 145
ANTONINO MINEO, ANGELO M. MINEO - <i>Analisi delle strutture di interdipendenza correlativa e associativa e della variabilità relativa di variabili statistiche multiple</i>	» 159

ANGELO M. MINEO - *Di un nuovo coefficiente per la corretta misura dell'associazione in tabelle dicotomiche e in tabelle di contingenza* pag. 187

ALIDA PERNICE - *Recente dinamica socio-demografica delle provincie siciliane e probabile evoluzione fino al 2011 della popolazione in età lavorativa* » 203

MARIA CALIRI

GIUSEPPE BAGNERA E LA MATEMATICA FINANZIARIA

1. Introduzione

Prestigioso rappresentante della Scuola matematica palermitana di fine '800 - inizio '900 fu Giuseppe Bagnera (1) di cui Severi, nella Commemorazione che lesse ai Lincei (2), sottolineò l'amore per "la perfezione e l'armonia logica" della matematica. Secondo Brigaglia e Masotto che lo presentano quale importante protagonista del Circolo matematico di Palermo: "Egli quanto a capacità matematiche, lo abbiamo già sottolineato, è di altissima statura. Ma il suo atteggiamento è contemplativo e staccato. I problemi matematici lo affasciano profondamente, ma egli finisce con il porsi al di fuori delle vivaci discussioni di quei tempi".

Di Bagnera sono noti i validi contributi e gli importanti riconoscimenti ufficiali (3) tanto che Bartolozzi lo ha collocato nella "triade divina palermitana" (4) e recentemente, nell'ambito della "rivisitazione storica della grande tradizione matematica siciliana" che il Circolo matematico di Palermo porta avanti da alcuni anni, ne è stata ripresentata alla comunità scientifica l'intera produzione con la stampa del volume delle sue *Opere* (5). Nastasi ha considerato ciò come una "riparazione" dovuta, lasciando immaginare il trasferimento a Roma come la conseguenza di possibili dissensi creatisi nell'ambito matematico paler-

(1) Vincitore di un concorso di Algebra e Geometria analitica, nel 1902, fu chiamato all'Università di Messina dove rimase fino al 1908. Era nato a Bagheria nel 1865. Ritornato a Palermo nel 1909, vi si fermò fino al suo trasferimento a Roma dove morì nel 1927.

(2) Nella seduta della Reale Accademia Nazionale dei Lincei del 6 maggio 1928. Il riferimento è stato reso possibile dal servizio offerto dalla Biblioteca "Roberto Stroffolini" del Dipartimento di Scienze Fisiche dell'Università Federico II di Napoli.

(3) Ad esempio, nel 1908 vinse, insieme a M. De Franchis, il Premio Bordin dell'Accademia delle Scienze di Parigi.

(4) Bagnera, Cipolla e De Franchis.

(5) Il Volume, edito dal Circolo matematico di Palermo, è stato presentato all'Accademia dei Lincei il 9 maggio 2000 da G. Zappa.

mitano, dato che il Circolo non ne pubblicò sui Rendiconti alcuna Commemorazione in occasione della morte.

Meno nota forse è, invece, la sua posizione “senza esclusivismi gretti”, pure ricordata da Severi nella Commemorazione, nei confronti di “qualunque ramo della matematica o della scienza ... (Egli) sapeva passare dalle più alte astrazioni alle cose concrete, con praticità ed equilibrio”.

Ed infatti Bagnera non disdegnò di assumere l’insegnamento di Matematica finanziaria e, successivamente, la direzione dell’Istituto Superiore di Commercio di Palermo. L’Istituto di Palermo nacque nel 1920 (6) ed ottenne il riconoscimento legale proprio ad opera di Bagnera, del quale Severi così descrisse l’impegno a questo proposito ... “lo riordinò e lavorò alacremente perché rispondesse sempre meglio alla sua funzione, e con mirabile tenacia volle che l’Istituto diventasse statale e, dopo lunghissime insistenze, vi riuscì”. Anche se a quell’epoca Bagnera si era già trasferito a Roma e, forse, vi riuscì proprio per questo.

L’esperienza presso l’Istituto lo portò a contatto con le applicazioni economiche e finanziarie della matematica e forse per tale motivo a Severi, che lo ricorda nella Commemorazione, aveva detto: ... *Caro amico, non illudiamoci: la matematica che noi facciamo non serve ai fini utilitari immediati (e lo stesso può ripetersi di quasi tutte le parti in formazione della scienza)*. Ma, e qui si ritrova il Bagnera descritto da Brigaglia e Masotto, nel contempo affermava ... *Se quello che facciamo non è bello, che gusto c’è a lavorare? E per me non è bello se non quello che è finito e perfetto*.

Allora egli aveva già concluso la sua produzione scientifica e si occupava soltanto dell’elaborazione di trattati e corsi di lezioni universitarie (Zappa-Zacher, pp. XXIII e XXVIII). Per il Corso di Matematica finanziaria, scrisse (7): “Lezioni di calcolo bancario e commer-

(6) Gli Istituti Superiori di scienze economiche e commerciali, che avevano preso l’avvio alla fine dell’Ottocento, furono regolati dalla legge istitutiva n. 268 del 20 marzo 1913 e dal Regolamento generale R.D. n. 1227 dell’8 luglio 1925, art. 1. Per l’iniziativa ed il sostegno finanziario di tre Enti fondatori: Camera di Commercio, Comune e Provincia di Palermo, il 20 febbraio 1920, nacque a Palermo l’*Istituto Superiore per gli Studi Economici e Coloniali*. Soltanto con il R.D. n. 829 dell’8 marzo 1925 si ebbe il riconoscimento legale dell’Istituto palermitano con il nome di *Istituto Superiore Libero di Scienze Economiche e Commerciali* con personalità giuridica, autonomia didattica amministrativa e disciplinare, sotto la vigilanza del Ministero dell’Economia nazionale. A questo proposito e in seguito vedansi L. Vianelli Bellavista (1995) e M. Caliri-L. Vianelli Bellavista (1999).

(7) Vedansi a questo proposito Brigaglia-Masotto a pag. 122, Zappa-Zacher a pag. XXVIII e Severi a pag. XX.

ciale” e “Lezioni di Matematica finanziaria”. Purtroppo presso le diverse biblioteche locali non ho trovato traccia di queste ultime e la stessa biblioteca di Bagnera, donata dagli eredi alla Biblioteca comunale di Bagheria e ospitata a Palazzo Cutò, non conserva alcun testo di Matematica finanziaria né del Bagnera né di altri.

Le “Lezioni di calcolo bancario e commerciale” (8) di cui, invece, ho potuto prendere visione presso la Biblioteca comunale di Palermo, mi hanno fatto nascere la curiosità di verificare come un “matematico di razza” di quell’epoca, ma con quell’apertura che gli riconosceva Severi, “guardasse” i problemi di Matematica finanziaria.

È di quel periodo, infatti, il polemico intervento di Insolera (9) (1922) ...”Fra i molti son primi a non comprendere un otto o quasi della nostra disciplina, salvo lodevoli eccezioni, i professori delle facoltà matematiche: costoro adusati alle più ardue speculazioni della mente, o si immaginano una Matematica finanziaria siccome accozzaglia di esercitazioni aritmetiche e computistiche o non nascondono un sufficiente quanto ridicolo dispregio... E non esitano, anche se putacaso insegnano Geometria descrittiva a spacciarsi per competenti di Matematica finanziaria e assumersi, con incosciente disinvoltura, il ruolo di giudici in materia... Eppure nessuno di costoro si sente preparato ad insegnare, per esempio, elettrotecnica o scienza delle costruzioni o altra disciplina di applicazioni matematiche...”.

Prima di entrare in argomento è sicuramente opportuno inquadrare rapidamente la situazione della disciplina a quel tempo. Nel piano di studi di tutti gli Istituti era previsto l’inserimento di una matematica “specificata che insegnasse ad impostare e trattare con adeguati strumenti matematici quei problemi che la continua evoluzione della realtà economica già poneva in campo finanziario ed assicurativo”, la matematica finanziaria, e poiché tale studio presupponeva la conoscenza di una matematica di base che gli studenti provenienti dagli istituti secondari non possedevano, il Ministero delegava il docente di Matematica finan-

(8) Nel seguito denominate semplicemente *Lezioni*.

(9) Filadelfo Insolera, vincitore nel 1914 del concorso a cattedra di Matematica finanziaria nell’Istituto Superiore di Scienze Economiche e Commerciali di Torino, riversò nella didattica e nella ricerca tutta la competenza acquisita, precedentemente al concorso, come capo dell’Ufficio attuariale dell’Istituto Nazionale della Previdenza Sociale. Fondò, insieme a S. Ortu-Carboni, il *Giornale di Matematica finanziaria*, pubblicato a Torino dal 1920 al 1959, del quale fu direttore fino alla morte (1955).

ziaria a fornire, “man mano che se ne presentava la necessità, le nozioni di matematica necessarie alla comprensione degli argomenti trattati”.

La scelta del piano di studi degli Istituti Superiori era dettata dalla legge istitutiva del 1913, già citata, che riconosceva gli Istituti come enti autonomi dotati di personalità giuridica propria e che all’art. 3 stabiliva che uno statuto organico avrebbe determinato “il carattere di ciascuna scuola e gli scopi ad essa prestabiliti per il maggior vantaggio del luogo ove avrà sede e delle classi di cittadini che specialmente dovranno approfittare”.

Gli Istituti erano dunque autorizzati a darsi piani di studio adeguati alle figure professionali richieste dalla società locale come, ad esempio, il commercio marittimo a Genova e l’industria assicurativa a Trieste; a Palermo si riconosceva necessaria una preparazione adatta a figure professionali localmente emergenti quali commercianti, dirigenti presso Amministrazioni dello Stato, ecc. Per questo nel programma di Matematica finanziaria dell’Istituto di Palermo non era fatto riferimento alcuno alla teoria delle assicurazioni che invece era già presente nel programma degli Istituti tecnici di ragioneria (pur non essendo nelle intenzioni del legislatore fare degli attuari, ma solo dare ai futuri ragionieri la possibilità di saper leggere un bilancio tecnico di una Società di assicurazione).

Per un’attenta analisi delle *Lezioni* non si può prescindere dal confronto con alcuni dei testi usati all’epoca nei diversi Istituti.

Il testo di paragone più significativo, sia per l’autorevolezza specifica che per la personalità dell’Autore, è certamente il “Corso di Matematica finanziaria” (1923) di Insolera che nell’Introduzione, con lo spirito polemico che spesso contraddistingueva i suoi scritti, precisava di supporre ...” già acquisiti al lettore ogni argomento e materia di studio che pur essendo indispensabili a una vasta e compiuta comprensione della Matematica finanziaria, tuttavia di esse non sono che premesse. Niente, dunque, geometria analitica, calcolo infinitesimale e equazioni differenziali; niente Calcolo delle probabilità e Statistica matematica presupposti tutti, utili come tali ma, a nostro avviso, sommaramente ingombranti e pregiudizievoli alla organicità e al serio sviluppo della disciplina, se comunque incastonati in un Corso di Matematica finanziaria. Questo che al di là delle Alpi riesce così pacifico da parer banale affermarlo in Italia ha il suo significato”. Lo stesso Autore aveva scritto per gli Istituti tecnici un testo (1916) presentato come una riduzione, “entro limiti elementari”, delle sue lezioni svolte nel R. Istituto Superiore di Torino, e dedicato alla scuola e “non a quanti, già

lontani dalla scuola, sotto l'assillo immanente delle funzioni assunte, anelino più al formulario, che li tragga, alla meno peggio, d'imbarazzo anzi che al trattato che li costringa a nuove fatiche intellettuali".

In entrambi i testi traspare la competenza specifica dell'Autore sugli argomenti trattati; Insolera, infatti, essendo uno studioso della disciplina, vi ha riversato parte delle sue ricerche.

Nelle "Lezioni..." di Amoroso (10), invece, si teneva già conto delle indicazioni del Ministero che delegava al docente i richiami necessari di Matematica generale. Ad ogni capitolo, infatti, egli fa precedere brevi, ma consistenti, cenni di matematica: nel primo capitolo presenta le operazioni di borsa preceduti da cenni di geometria analitica utili per la descrizione del profitto per l'acquirente e il venditore di contratti a premio; passa poi allo studio dei polinomi e delle curve esponenziali e logaritmiche per la capitalizzazione, delle funzioni in generale per risolvere i problemi inversi legati alle rendite, ecc.

Diversa la "logica" che guida il Dell'Agnola (11) nella stesura del suo libro (la copia consultata è del 1930) che nella Prefazione scrive: "L'esposizione della materia è, nell'ordine e nell'essenza, quella stessa che io seguo nelle mie lezioni presso l'Istituto Superiore di Scienze Economiche e Commerciali di Venezia. Assai più che della trattazione di questioni particolari, mi sono preoccupato di stabilire, da un punto di vista generale e scientifico, i principi e i metodi indispensabili per la risoluzione dei problemi che s'incontrano nella pratica".

Vi erano anche altri testi come, ad esempio, quello di Spinedi o quello di Barriol, che già dalla sede di pubblicazione mostravano di essere rivolti ad un diverso genere di utenti e ai quali si farà riferimento nei prossimi paragrafi.

2. Le "Lezioni di calcolo bancario e commerciale"

In queste *Lezioni* l'Autore, segue soltanto in parte le indicazioni del legislatore, soffermandosi sui problemi in cui a quel tempo pote-

(10) Luigi Amoroso era un matematico che aveva vinto il concorso di Matematica finanziaria e...ciò forse lo metteva al riparo dalle critiche di Insolera. Insegnava all'Istituto Superiore di scienze commerciali di Napoli.

(11) Carlo Alberto Dell'Agnola fu vincitore del concorso a cattedra di Matematica finanziaria a Venezia. L'Istituto Superiore di Commercio di Venezia, trasformatosi poi in Istituto Superiore di Scienze Economiche e Commerciali, era il più antico d'Italia essendo stato fondato nel 1868.

vano imbattersi le figure “professionali” individuate come destinatari. Vi risulta alleggerito ogni contenuto strettamente matematico, come ad esempio l’uso dei limiti, ma vi si insiste, e forse esageratamente, su tutti i passaggi algebrici; ciò è difficilmente giustificabile con la sola mancanza di un corso propedeutico di Matematica generale che, al di là dei contenuti, facesse acquistare agli studenti la manualità necessaria a trattare le formule e i passaggi algebrici.

La presentazione degli argomenti è mirata, mostrando l’abilità didattica di Bagnera, così ben descritta da Severi ...“egli non era un rigorista pedante. Ove l’esperienza didattica lo consigliava, sapeva sorvolare. Altrove aveva infatti già scritto: *una trattazione perfettamente logica è spesso un bisogno della nostra mente, che non è affatto condiviso dai giovani; anzi alle volte una particolarità di dimostrazione serve soltanto a sviarli dal concetto fondamentale*”.

L’analisi delle *Lezioni* non si propone di ricercare un contributo originale per contenuti e impostazione, non solo perché si tratta di dispense ad uso degli studenti (anche se Insolera, invece, usava il suo libro come canale per la diffusione delle nuove idee in Matematica finanziaria) e sia perché al Bagnera mancava una specifica competenza. Vuole, invece, dare atto della serietà con la quale egli si è accinto al compito didattico affidatogli, proponendo un suo testo (anche per non presentarsi come un ripetitore di idee altrui, tratte da un qualsiasi altro testo adottato); vuole, inoltre, sottolineare l’approccio che con il rigore consueto e con l’impostazione mentale del matematico, pur con i limiti precisati, egli segue in questo scritto. Le *Lezioni*, infatti, affrontano temi generali che danno un respiro più ampio ad un volume che in fondo si occupa di ...“computisteria”. E ciò ha una certa rilevanza in un’epoca nella quale alla Matematica finanziaria non si riconosceva ancora un ruolo di “pari dignità” all’interno del mondo matematico.

Riportiamo qui di seguito l’indice delle *Lezioni*:

Operazioni a breve scadenza

Calcolo del percento ed applicazione alle questioni di interesse e di sconto (nn. 1 a 6).

Effetti equivalenti (nn. 7 a 14).

Titoli in deposito (nn. 15 a 20).

Operazioni a lunga scadenza

Interesse composto (nn. 21 a 23).

Sconto composto (n. 24).

Tasso equivalente (nn. 25 a 26).

Tasso continuo (nn. 27 a 31).

Rendite certe (nn. 32 a 37).

Costituzione di un capitale (n. 38).

Ammortamenti (nn. 39, 40).

Piano di ammortamento col metodo francese (nn. 41, 42, 43).

2.1 *Le operazioni a breve scadenza*

L'Autore inizia le *Lezioni* così: *Il calcolo bancario e commerciale, che occorre applicare in tutte le operazioni di banca e di commercio, si può dividere in tre parti: problemi del tanto per cento; operazioni di cambio e borsa; questioni sulle annualità. ...Il per cento va ora aggiunto ora tolto a secondo delle questioni che si presentano; talvolta va calcolato come compenso.* Seguono esempi anche di argomento non finanziario (si tenga presente il titolo delle dispense!) come quello del calcolo del peso netto di una merce nota la tara e il peso lordo, ma nel seguito si trattano solo problemi tipicamente finanziari.

Nel caso di un capitale C e ad un tasso r per un tempo t il per cento diventa $I = Crt$ e pertanto, a seconda che esso *vada aggiunto oppure tolto*, si ottengono le formule $M = C + Crt = C(1 + rt)$ e $M = C - Crt = C(1 - rt)$ delle quali la prima si applica nelle operazioni di interesse e la seconda nelle operazioni di sconto.

Bagnera precisa che *nelle operazioni di sconto ordinariamente la durata dell'operazione non è un anno ma una frazione di anno.* Non deve quindi sorprendere che nel caso in cui *il per cento va tolto* egli, come l'Insolera, non si ponga il problema della positività del risultato. Infatti la formula usata $M = C(1 - rt)$ – che si riferisce all'uso del cosiddetto sconto commerciale – dovrebbe essere accompagnata dal vincolo $1 - rt > 0$ (nessun operatore richiederebbe lo sconto di un effetto per trovarsi debitore!) che ne limita l'uso per $t < 1/r$ e, d'altra parte, nel caso del problema inverso della determinazione del valore nominale dell'effetto, $C = \frac{M}{1 - rt}$, il non porre tale limitazione, oltre a far perde-

re di significato finanziario il risultato nel caso di un valore negativo, potrebbe far cadere in difetto – per $1 - rt = 0$ – la stessa formula.

Come già accennato, ho consultato altri testi; in particolare ho potuto consultare i testi di Barriol, Dell’Agnola, Hart, Insolera, Sonnet e Spinedi. Ho constatato che, a questo proposito, diversamente si poneva il Barriol che a pag. 39 osservava: «cette remarque montre l’absurdité de la formule de l’escompte en dehors, qui ne peut s’appliquer pour des durées trop longues; mais, en pratique, son emploi est limité à des durées de 2 à 3 mois au plus, et, dan ce cas, son application, bien qu’avantageuse pour le banquier escompteur, est adopté d’une manière générale». Egli precisava che lo sconto razionale era preferito allo sconto commerciale in quanto il primo «... n’est généralement pas employé dans la pratique à cause des calculus laborieux qu’elle entraîne». Diceva, inoltre, il Sonnet alla voce escompte: «En France, l’escompte se calcule comme l’intérêt simple... Mais si c’est un banquier qui escompte le billet, il prend ordinairement en outre une commission qui est presque toujours de $\frac{1}{2}$ pour 100 de la somme énoncée... L’escompte, tel qu’on le pratique en France, porte le nom d’escompte en dehors. Mais, dans quelques autres pays, on prend l’escompte en dedans; c’est-à-dire que l’on considère la somme énoncée dans le billet comme un capital augmenté de son intérêt pour le temps indiqué par l’échéance; et c’est cet intérêt que l’on retranche de la somme énoncée lorsqu’on escompte le billet. Cette méthode, beaucoup moins commode que celle que l’on suit en France, est cependant plus rationnelle ». Sembrerebbe una contraddizione con la terminologia ancora oggi in uso dato che allo sconto usato come interesse semplice si attribuisce il nome di “sconto razionale”, denominando l’altro (quello che il Sonnet ritiene più razionale) “sconto commerciale”.

L’Hart, invece, introduceva tale sconto, come “simple discount”, senza porre alcuna limitazione, mentre lo Spinedi poneva la limitazione, pur avendo precisato che lo sconto commerciale era usato nella pratica “quando si tratta di capitali scadenti dopo breve tempo” (12), anche se ... “lo sconto razionale è l’unico ammesso dalla scienza” (13).

Ma c’era tanta confusione nella letteratura corrente; ad esempio, lo Spinedi con sconto semplice intendeva lo sconto razionale! Inoltre era ancora l’epoca nella quale si dissertava sul confronto tra i due scon-

(12) Pag. 73 e seguenti.

(13) Nota a pag. 85.

ti ma la questione, di particolare interesse per le banche anche per migliorare l'offerta a fini concorrenziali, poteva essere liquidata in poche battute e non giustificava, quindi, gli appassionati interventi pubblicati fino agli anni '50 sulle pagine del Giornale di Matematica finanziaria.

Bagnera, dunque, dava per acquisito che lo sconto non superasse la somma da scontare, ritenendo forse da criticare non la formula ma l'uso improprio di questa, introdotta solo per operazioni di breve periodo. Da tutto ciò si trae un segnale che deve guidare nella lettura del testo: l'Autore vi porta il rigore nell'impostazione che si traduce, come già detto, in una visione generale dei problemi; ne segue che più che fermarsi sugli esempi numerici, tipici di altri testi, che spesso limitano l'interpretazione dei risultati, egli preferisce spaziare, come si vedrà, su questioni generali.

Bagnera prosegue con il determinare la condizione di equivalenza per due effetti, diversi per valore nominale e scadenza, e osserva che in tal caso *l'effetto che ha valore nominale maggiore ha scadenza più lontana* e, infine, dimostra che:

Se ad una certa epoca due effetti di valori nominali C_1 e C_2 e distanti dalla scadenza rispettivamente t_1 e t_2 non sono equivalenti, perché possano divenire equivalenti ad un'epoca anteriore alla scadenza di entrambi bisogna che:

1° quello avente valore attuale maggiore all'epoca che si considera abbia valore nominale minore e scadenza più vicina;

2° inoltre lo sconto dell'effetto di valore nominale maggiore per un tempo uguale alla differenza dei tempi che mancano ai due effetti per la scadenza deve essere minore della differenza dei valori nominali dei due effetti: $C_2 r (t_2 - t_1) < C_2 - C_1$;

3° il tempo t dopo il quale i due effetti diventano equivalenti è dato

$$dalla formula \quad t = \frac{C_2 t_2 - C_1 t_1}{C_2 - C_1} - \frac{1}{r}.$$

Osserviamo che un'interpretazione in termini finanziari dei risultati avrebbe reso più scorrevole il testo; ad esempio la relazione riportata al punto 2°, così riscritta $C_2 [1 - r (t_2 - t_1)] > C_1$, si riduce al confronto tra i capitali C_1 e C_2 riportati alla stessa scadenza t_1 , dato che due capitali sono confrontabili solo se riferiti ad una stessa scadenza.

Ma non è la mancanza, già sottolineata, di una specifica competenza, che lo porta ad appesantire il commento dei risultati, ma una mentalità allora corrente che portava a privilegiare i legami formali a quel-

li sostanziali (finanziari), perché lo stesso punto 2° è presentato in ugual modo anche da Insolera, che certo era un maestro nel campo finanziario. D'altra parte, anche in tempi moderni, si è indugiato spesso su questioni di tal genere (problemi di unificazione dei capitali o dei tempi) per mere esercitazioni, magari allo scopo di mettere il risultato sotto una forma compatta e nota. Sarebbe, invece, di gran lunga più interessante considerare casi della pratica commerciale; in questo contesto, ad esempio, un problema con un preciso significato finanziario è quello di chi ha sottoscritto più effetti con scadenze diverse e, disponendo ad un certo momento di una somma di denaro, voglia assolvere anticipatamente i suoi obblighi (dunque unificazione dei capitali). La somma

necessaria per l'estinzione in t è: $C = \frac{\sum C_s (1 - rt_s)}{1 - rt}$, dove il generico

addendo $\frac{C_s (1 - rt_s)}{1 - rt}$ è la somma da pagare in t per assolvere l'obbligazione

relativa all'effetto di valore nominale C_s . Si noti che, essendo la legge finanziaria usata non "scindibile", la valutazione di ogni somma C_s non può essere fatta direttamente in t ma occorre prima valutarla in 0 (moltiplicando per $1 - rt_s$ perché questa legge pone l'equivalenza finanziaria in 0) e poi riportarla in t (capitalizzarla fino in t , dividendo per $1 - rt$).

L'argomento delle operazioni a breve scadenza viene completato accennando all'operazione sui *titoli in deposito* che riportiamo qui di seguito integralmente perché vi traspare l'attenzione per un problema, importante ancora oggi, che consente di speculare sul mercato.

“Per avere denaro da una banca talvolta, invece di girare cambiali che rappresentano una garanzia indiretta fondata sulla fiducia, si danno in deposito titoli. Sulla variazione di prezzo di questi titoli si può speculare. Supponiamo che una società abbia emessi dei titoli e che per lo sviluppo che accenna a prendere l'industria di cui si occupa la società questi titoli sia presumibile che aumentino di valore.

Allora una persona che abbia un capitale C e che voglia speculare sulla possibilità di rialzo dei titoli in discorso potrà acquistare di tali titoli per un capitale C .

Volendo acquistare titoli per una somma maggiore si può ricorrere all'operazione del deposito dei titoli; essa consiste nell'ottenere un prestito da una banca depositando a garanzia di tale prestito i titoli che già si erano acquistati col capitale C posseduto prima.

La banca però non dà l'ammontare C del valore dei titoli ma dà alquanto

di meno per essere al coperto da eventuali ribassi che potessero sopravvenire per i titoli che riceve in deposito; ordinariamente dà soltanto i 4/5 del valore C dei titoli; onde depositando dei titoli per un valore C si ottiene la somma $C/5$.

In generale se la banca dà la frazione k ($k < 1$) del valore dei titoli che si offrono in deposito, la somma che si ricava depositando dei titoli per un valore C è Ck . Questa somma Ck si può investire nell'acquisto di nuovi titoli e depositandoli pure alla banca si ricava la somma $(Ck)k = Ck^2$. Acquistando con questa somma altri titoli e depositandoli alla banca si ottiene $(Ck^2)k = Ck^3$, e ripetendo l'operazione di deposito n volte la somma ricevuta complessivamente dalla banca è:

$$S = Ck + Ck^2 + Ck^3 + \dots + Ck^n = Ck(1 + k + k^2 + k^3 + \dots + k^{n-1})$$

Per calcolare poi la somma messa tra parentesi, cioè la somma $1 + k + k^2 + k^3 + \dots + k^{n-1}$ si calcoli (14) il prodotto di questa somma per $1 - k$; si ha:

$$\begin{array}{r} 1 + k + k^2 + k^3 + \dots + k^{n-1} \\ \text{per} \quad 1 - k \\ \hline 1 + k + k^2 + k^3 + \dots + k^{n-1} \\ - k - k^2 - k^3 - \dots - k^{n-1} \\ \hline 1 \qquad \qquad \qquad - k^n \end{array}$$

perché tutti gli altri termini si eliminano. Dalla relazione: $(1 + k + k^2 + \dots + k^{n-1})(1 - k) = 1 - k^n$ si ricava allora: $(1 + k + k^2 + \dots + k^{n-1}) = \frac{1 - k^n}{1 - k}$ e quindi si ha:

$$S = Ck(1 + k + k^2 + \dots + k^{n-1}) = Ck \frac{1 - k^n}{1 - k}.$$

La somma complessiva ricevuta dalla banca ripetendo n volte l'operazione di deposito dei titoli è dunque:

$$Ck \frac{1 - k^n}{1 - k}.$$

Essendo $k < 1$, le potenze k^2, k^3, \dots di k vanno decrescendo col crescere dell'esponente onde prendendo n abbastanza grande la potenza k^n

diventa un numero abbastanza piccolo; se nella formula $Ck \frac{1 - k^n}{1 - k}$

(14) Bagnera, evidentemente, riteneva che gli studenti non conoscessero la formula per il calcolo della somma di termini in progressione geometrica.

quando n è abbastanza grande e quindi k^n è abbastanza piccolo si trascura la potenza k^n , si trova che la somma ricevuta dalla banca ripetendo n volte l'operazione di deposito dei titoli è $S = \frac{Ck}{1-k}$. Se per esempio la Banca dà ogni volta i $4/5$ del valore dei titoli depositati, la somma che si riceve dalla banca ripetendo l'operazione di deposito un numero di volte sufficientemente grande, si ottiene ponendo nella formula precedente $k = 4/5$; essendo allora $1 - k = 1 - 4/5 = 1/5$ si ha che tale somma è $\frac{C \cdot 4/5}{1/5} = 4C$.

La persona quindi che facesse l'operazione di acquistare dei titoli per depositarli in una banca ed investire il ricavato nell'acquisto di nuovi titoli, possedendo un capitale iniziale C e ripetendo l'operazione un numero sufficiente di volte, potrebbe avere complessivamente dalla banca una somma che è circa 4 volte il capitale che possedeva in principio. Avendo poi investito nell'acquisto di titoli il capitale posseduto in principio e le somme man mano ottenute dalla banca, l'ammontare dei titoli acquistati è $C + 4C = 5C$ circa.

Di modo che operando ad esempio con un capitale iniziale di 50.000 lire si possono acquistare dei titoli per circa 250.000 lire.

Effettuandosi quindi per quei titoli il rialzo sperato la persona che ha fatto l'operazione viene a ricavare un guadagno molto maggiore di quello che avrebbe fatto se avesse investito in quei titoli soltanto le 50.000 lire che possedeva.

S'intende che in caso d'insperato ribasso anche la perdita sarebbe maggiore di quella che si sarebbe subito acquistando soltanto 50.000 lire di titoli.

Il problema trattato ha una certa rilevanza perché il ricorso ai successivi prestiti mette a disposizione di uno speculatore che gioca al rialzo una somma considerevole, pur possedendo egli inizialmente un capitale modesto. Bagnera si ferma qui, senza considerare che per accedere al prestito si sostiene un costo che riduce il guadagno in caso di rialzo e appesantisce la perdita in caso di ribasso.

Lo stesso problema è presentato nel testo del Barriol (15) il quale approfondisce la questione tenendo conto delle spese sostenute, per il prestito che la banca ha concesso, anche se poi le trascura limitando-

(15) Op. cit., Cap. III, par. 234, pag. 362 e seguenti.

si a considerare il guadagno in caso di rialzo dei titoli e sottolineando la possibilità di rovina nel caso in cui un eventuale ribasso assorba completamente il capitale impiegato.

Anche Insolera osserva che se il valore dei titoli alla scadenza dell'operazione fosse proprio pari alla frazione k del loro valore, il valore dei titoli in deposito si ridurrebbe a $\frac{Ck}{1-k} < S$ e, in questo caso, "lo speculatore, nella restituzione alla banca delle somme ricevute in anticipo perderebbe l'intero capitale $\frac{C}{1-k} - \frac{Ck}{1-k} = \frac{C(1-k)}{1-k} = C$ oltre, ben inteso, il pagamento degli interessi e delle spese inerenti l'operazione".

2.2 Le operazioni a lunga scadenza

Bagnera introduce l'interesse composto e calcola il montante di un capitale C impiegato per n periodi di durata t , alla fine di ciascuno dei quali gli interessi maturati vengono capitalizzati al tasso r , che è dato da $C(1+rt)^n$. Se il periodo di capitalizzazione è la m -esima parte dell'anno allora in n anni vi saranno nm periodi e *calcolando che il tasso per questo nuovo periodo sia r/m si trova che il montante alla fine degli n anni, cioè dopo gli nm periodi, è $C\left(1+\frac{r}{m}\right)^{mn}$.*

Per esempio il montante alla fine di 5 anni quando il periodo di capitalizzazione è un anno è $C(1+r)^5$, se per periodo di capitalizzazione si prende il semestre il montante viene $C\left(1+\frac{r}{2}\right)^{10}$. Questo montante è maggiore dell'altro che si aveva quando il periodo di capitalizzazione era l'anno. In generale, quando come nuovo periodo di capitalizzazione si sceglie la m -esima parte del periodo primitivo e come tasso relativo a questo nuovo periodo di capitalizzazione si prende la m -esima parte del tasso relativo al primitivo periodo di capitalizzazione, il montante aumenta al diminuire del periodo di capitalizzazione".

L'Autore pone, quindi, il problema in termini generali: *la determinazione del tasso j_m tale che il montante alla fine degli n anni sia lo stesso di quando il periodo di capitalizzazione era un anno. Dalla $C(1+r)^m = C\left(1+\frac{j_m}{m}\right)^{mn}$ si ottiene $j_m = m[(1+r)^{1/m} - 1]$, che è detto tasso equivalente.*

Da qui passa al tasso continuo ipotizzando un continuo rimpicciolimento del periodo di capitalizzazione, ottenuto in corrispondenza di un continuo aumento di m .

Al crescere di m il tasso j_m diminuisce (nel prodotto che definisce j_m la diminuzione del fattore $[(1+r)^{1/m} - 1]$ non è compensata dall'aumento del fattore m). Quando m è grandissimo, anche m/j_m è grandissimo e l'espressione $(1 + \frac{j_m}{m})^{m/j_m}$ si avvicina, o come si dice tende, al numero e ed il tasso j_m si avvicina ad un tasso limite δ , per cui si ha:

$1+r = e^\delta$ dalla quale si ricava $\delta = \frac{\log(1+r)}{\log e} = \log(1+r)$ ed in questo

caso si dice che il capitale C è impiegato a fruttificazione continua. Da qui segue che il montante di un capitale C impiegato per n anni ad interesse composto e a fruttificazione continua è $M = Ce^{\delta n}$.

Come si nota, si tratta di un'impostazione teorica che fa ricorso a conoscenze matematiche che gli studenti certamente non possedevano e impreziosisce le dispense, nonostante il titolo che è certamente riduttivo del contenuto. Ma forse il matematico non ha saputo resistere alla tentazione!

Bagnera passa poi alla valutazione delle rendite costanti ma, purtroppo, non usa i simboli finanziari (16), già introdotti nei testi dell'epoca a indubbio vantaggio sia della sintesi e della snellezza dei procedimenti che dell'eleganza formale del risultato. Forse è proprio la mancanza di tali simboli, i cui valori numerici sono raccolti in apposite tavole, che porta Bagnera ad ignorare il problema inverso della ricerca del tasso al quale è stato calcolato il valore attuale o il montante di una rendita. Eppure si tratta di una questione finanziariamente rilevante perché tale tasso viene spesso usato per la valutazione (quale misura del costo oppure del rendimento effettivo) dell'operazione che utilizza quella rendita. È però anche probabile che egli abbia visto la questione con il "distacco" del matematico, ritenendola "troppo contabile", pur se l'impossibilità di risolvere algebricamente tale problema ha determinato la "fioritura" di un'infinità di metodi matematici di approssimazione della

(16) Il simbolo relativo alla valutazione delle annualità è stato introdotto da Milne (vedasi Cajori) nel 1815. All'epoca del Bagnera, oltre ai simboli del valore attuale e del montante di una rendita unitaria, era stato introdotto il loro inverso ad indicare, rispettivamente, il termine costante dell'ammortamento unitario e quello costitutivo di un capitale, ossia:

$$a_{\bar{n}/i}, s_{\bar{n}/i}, \alpha_{\bar{n}/i}, \sigma_{\bar{n}/i}$$

soluzione, il più semplice dei quali è quello dell'interpolazione lineare mediante secante, che era presente in molti altri testi dell'epoca.

L'Autore presenta poi le prime applicazioni delle rendite, ossia la costituzione di un capitale e l'ammortamento di un debito. Egli si sofferma sull'ammortamento francese ossia l'estinzione di un debito mediante il versamento di una rendita costante le cui rate, per definizione di ammortamento, servono in parte a remunerare il creditore (la cosiddetta quota interessi) ed in parte a rimborsare via via il prestito (la cosiddetta quota capitale). Riportiamo esattamente l'esposizione del Bagnera:

La rata annuale da pagare per ammortizzare con n rate il debito di una lira per quanto abbiamo visto al n. 39, è $\frac{r(1+r)^n}{(1+r)^n-1}$, e di questa rata

la somma r serve per pagare gli interessi di una lira per il primo anno e la parte restante a scomuto del capitale.

La somma che va a scomuto del capitale nella prima rata è quindi $\frac{r}{(1+r)^n-1}$. Alla fine del primo anno la lira era diventata $1+r$ e il

debito alla fine del primo anno è: $\frac{(1+r)^n-(1+r)}{(1+r)^n-1}$.

Nel secondo anno bisogna pagare gli interessi su questo debito restante e perciò tali interessi sono: $\frac{r(1+r)^n-r(1+r)}{(1+r)^n-1}$, onde della seconda rata

di ammortamento la parte che va in conto interessi è: $\frac{r(1+r)^n-r(1+r)}{(1+r)^n-1}$

e la parte che va in conto capitale è la differenza tra l'ammontare della rata e questa somma pagata per interessi, cioè $\frac{r(1+r)^n-r(1+r)+r(1+r)}{(1+r)^n-1}$

e questo risultato come si vede facilmente è $\frac{r(1+r)}{(1+r)^n-1}$. Alla fine del

secondo anno il debito era dato dal debito rimasto dopo il pagamento della prima rata aumentato dagli interessi del secondo anno, onde il debito alla fine del secondo anno è $\frac{r(1+r)^n-r(1+r)+r(1+r)^2}{(1+r)^n-1}$ e

dopo pagata la seconda rata tale debito rimane $\frac{r(1+r)^n-r(1+r)^2}{(1+r)^n-1}$ e

così di seguito.

Abbiamo quindi alla fine del primo anno:

$$\text{rata pagata } \frac{r(1+r)^n}{(1+r)^n - 1}; \text{ debito restante } \frac{(1+r)^n - (1+r)}{(1+r)^n - 1};$$

$$\text{in conto interessi } r; \text{ in conto capitale } \frac{r}{(1+r)^n - 1}.$$

Alla fine del secondo anno, dopo pagata la seconda rata:

$$\text{rata pagata } \frac{r(1+r)^n}{(1+r)^n - 1}; \text{ debito restante } \frac{(1+r)^n - (1+r)^2}{(1+r)^n - 1};$$

$$\text{in conto interessi } \frac{r(1+r)^n - r(1+r)}{(1+r)^n - 1}; \text{ in conto capitale } \frac{r(1+r)}{(1+r)^n - 1}.$$

Alla fine del terzo anno si avrebbe:

$$\text{rata pagata } \frac{r(1+r)^n}{(1+r)^n - 1}; \text{ debito restante } \frac{(1+r)^n - (1+r)^3}{(1+r)^n - 1};$$

$$\text{in conto interessi } \frac{(1+r)^n - (1+r)^2}{(1+r)^n - 1}; \text{ in conto capitale } \frac{r(1+r)^2}{(1+r)^n - 1}.$$

E si costruisce così il piano di ammortamento.

L'Autore chiude in questo modo le "Lezioni", certamente con la consapevolezza che le formule espone descrivono un processo banale nell'esecuzione ma esageratamente pesante nella presentazione; nei fatti, la conclusione non appaga né il matematico né il contabile. All'osservazione di un matematico, infatti, non può sfuggire la ricorrenza di certe espressioni, da cogliere per ottenere una formalizzazione ben più elegante di quella che porta alla stesura di un piano di ammortamento. D'altra parte, alle necessità del contabile, che spesso non è in grado di modificare per gli anni successivi il procedimento indicato, il metodo empirico è più utile di un'elencazione di formule. Inoltre sia il contabile che il matematico, richiedono una generalizzazione delle formule per poter verificare in un qualsiasi anno l'esattezza dei valori del piano, senza doverne effettuare una revisione totale. A volte, infatti, non basta che il piano "bilanci" – ossia che il debito residuo dell'ultimo anno sia nullo e che la somma di tutte le quote in conto capitale sia pari all'ammontare del debito contratto – in quanto errori in quote diverse potrebbero complessivamente compensarsi.

Per tener conto delle aspettative di entrambi sarebbe bastato generalizzare le espressioni che precedono, ottenendo per il k -esimo anno:

$$\text{il debito residuo } D_k = \frac{(1+r)^n - (1+r)^k}{(1+r)^n - 1},$$

$$\text{la quota interessi } I_k = \frac{r(1+r)^n - r(1+r)^{k-1}}{(1+r)^n - 1}$$

$$\text{la quota capitale } C_k = \frac{r(1+r)^{k-1}}{(1+r)^n - 1}.$$

È facile riconoscere che, indicata con R la rata pagata, segue:

$$D_k = R \frac{1 - (1+r)^{-(n-k)}}{r}; \quad I_k = rD_{k-1}; \quad C_k = R(1+r)^{-(n-k+1)},$$

espressioni che consentono la verifica dei valori di un anno qualsiasi di ammortamento.

Dall'ultima relazione si deduce che è anche $C_k = C_{k-1}(1+r)$, ossia che le quote in conto capitale sono crescenti in progressione geometrica di ragione $1+r$; tale osservazione facilita la stesura e la verifica di un piano di ammortamento perché, essendo $I_1 = r$, basta calcolare la prima quota capitale, $C_1 = R - r$, e da questa, moltiplicando per $1+r$, ottenere successivamente tutte le altre quote capitale.

Tale formalizzazione era già presente nei volumi di Insolera (che affrontava anche l'ammortamento ad interessi anticipati), nel testo di Dell'Agnola (che si spingeva fino all'ammortamento americano) e nel volumetto di Spinedi.

Le *Lezioni*, dunque, si concludono con le prime applicazioni delle rendite, senza ulteriori approfondimenti ed estensioni, ad esempio ai prestiti divisi in titoli. Di certo Bagnera non ha fatto in tempo a riscrivere ed ampliare le dispense, perché la sua permanenza nell'Istituto è stata brevissima.

In questo testo, come in tutti gli altri dell'epoca, si nota l'assenza di ogni rappresentazione geometrica, ben presente invece nei testi di oggi, che avrebbe consentito di visualizzare meglio i problemi e di interpretare appropriatamente i risultati. Un tale approccio sarebbe stato certamente congeniale a Bagnera che era ingegnere e – ci fa piacere pensarlo! – forse è stato lo schema che egli ha seguito mentalmente nello studio dei problemi.

BIBLIOGRAFIA

- AMOROSO L. (1921), *Lezioni di Matematica finanziaria raccolte dalla prof.ssa Emma Sciolette*, Vol. I, G. Majo Libraio-Editore.
- BAGNERA G. (1920), *Lezioni di calcolo bancario e commerciale* (Litografie ad uso degli studenti), Palermo.
- BAGNERA G. (1999) *Opere*, a cura di Zappa G. - Zacher G., Supplemento ai Rendiconti del Circolo matematico di Palermo, serie II, n. 60.
- BARRIOL A. (1914), *Théorie et pratique des opérations financières*, Encyclopédie, Bibliotheque de Mathématiques appliquées, II ed., G. Doin, Paris.
- BARTOLOZZI F. (1994), *Giovan Battista Guccia e il Circolo Matematico di Palermo. La grande generazione matematica palermitana*, Acc. Naz. di Scienze, Lettere e Arti di Palermo, 34, pp. 20-43.
- BRIGAGLIA A.-MASOTTO G. (1982), *Il circolo matematico di Palermo*, Ed. Dedalo, pp. 119-120.
- CAJORI F. (1993), *A history of mathematical notations*, Dover Publications, New York, p. 71.
- CALIRI M.-VIANELLI BELLAVISTA L. (1999), *La didattica delle discipline matematiche nella Facoltà di Economia di Palermo*, Annali della Facoltà di Economia dell'Università di Palermo, Area statistico-matematica, 1996-99.
- I° Circolo didattico Scuola elementare G. BAGNERA (a cura di) (1997), *Giuseppe Bagnera tra memoria e aneddoti*, 12 maggio 1997, Città di Bagheria.
- DELL'AGNOLA C.A. (1930), *Matematica finanziaria*, Scarabellin, Venezia.
- HART W.L. (1924), *Mathematics of investment*, Heath and Company.
- INSOLERA F. (1916), *Elementi di Matematica Finanziaria ed Attuariale ad uso degli istituti commerciali*, Lattes.
- INSOLERA F. (1922), *Sull'insegnamento della Matematica Finanziaria negli Istituti Superiori di Commercio*, Giornale di Matematica Finanziaria, vol. IV, n. 4, pp. 154-60.
- INSOLERA F. (1923), *Corso di Matematica finanziaria*, Lattes.
- NASTASI P. (1999), *Giuseppe Bagnera, uno della triade "divina" palermitana* in "Francesco Castronovo sacerdote ed educatore", Atti del Convegno nel 1° Centenario della morte, Lions Club Bagheria, 5 giugno 1999, pp. 75-76.

SEVERI F. (1928), *Commemorazione del Socio corrispondente Giuseppe Bagnera*, Rend. R. Acc. Naz. Lincei, vol. VIII, serie 6°, 1928-VII, App. XII-XX.

SONNET H. (1874), *Dictionnaire de Mathématiques appliquées*, Librairie Hachette et C., Paris.

SPINEDI S. (1917), *Elementi di Matematica Finanziaria*, Biblioteca di Ragioneria e d'Amministrazione, Vallardi.

VIANELLI BELLAVISTA L. (1995), *Una nota per la storia della Facoltà di Economia di Palermo*, preprint dell'Istituto di Matematica per la Ricerca Operativa.

MARIA DAVÌ, GIUSEPPE NOTARSTEFANO, ERASMO VASSALLO (1)

RECENTE EVOLUZIONE DELLA STRUTTURA PRODUTTIVA SICILIANA: UNA NOTA CON RIFERIMENTO AGLI 82 SISTEMI LOCALI DEL LAVORO

Abstract: Il lavoro intende offrire alcune riflessioni sulla evoluzione delle attività industriali negli 82 Sistemi Locali del Lavoro in Sicilia. Si tratta dei primi risultati di una ricerca il cui obiettivo è verificare i recenti percorsi evolutivi della struttura produttiva siciliana, in relazione alla più complessiva evoluzione strutturale delle altre regioni europee. Vi sono evidenze verso uno scenario di declino fortemente determinato dalla riduzione delle attività produttive, dal numero di addetti ma, soprattutto, dalla composizione settoriale degli apparati produttivi locali che penalizza i settori maggiormente innovativi. Tale processo viene rappresentato attraverso lo studio di alcuni indici di localizzazione e specializzazione che consentono di sintetizzare a livello territoriale alcune importanti macro tendenze.

1. Introduzione

In quale direzione si sta evolvendo la Sicilia? È possibile trovare gli indizi di un processo di sviluppo che tenta, pur faticosamente, la via di una evoluzione virtuosa? Oppure, irrimediabilmente, occorrerà rassegnarsi a processi di involuzione che suggeriscono il naufragio di ogni azione non soltanto endogenamente stimolata ma anche esogenamente guidata?

In questo lavoro, nel contesto di un più ampio percorso di ricerca sull'argomento, ci si sofferma su alcune considerazioni descrittive cir-

(1) Dipartimento di Contabilità Nazionale e Analisi dei Processi Sociali, Università degli Studi di Palermo. Il lavoro, pur frutto di comune riflessione, può essere attribuito agli Autori nel modo seguente: i paragrafi 1 e 2 e l'appendice A a Erasmo Vassallo, il paragrafo 3 con l'esclusione del 3.4 a Maria Davì, il sottoparagrafo 3.4 ed i paragrafi 4 e 5 e le appendici B e C a Giuseppe Notarstefano.

ca la recente evoluzione della struttura produttiva siciliana tentando di fornire un quadro generale che incrocia informazioni censuarie e stime per Sistema Locale del Lavoro.

Nel rispetto di quanto appena tratteggiato, il lavoro appare suddiviso in tre parti. Nella prima (paragrafo 2 e appendice A), si contestualizza la realtà siciliana rispetto alle recenti dinamiche italiane ed europee. In particolare, il sottoparagrafo 2.1 evidenzia la necessità di ricorrere ad informazioni territorialmente disaggregate; ciò consente di individuare dinamiche di sviluppo che altrimenti verrebbero nascoste. Il sottoparagrafo 2.2 mostra l'intrinseca debolezza della Regione con riferimento agli analoghi aggregati territoriali europei: emerge una rinnovata condizione di marginalità che aiuta ad interpretare e relativizzare alcuni pur presenti positivi percorsi di crescita, analizzati in parte nel sottoparagrafo 2.3 con riferimento agli 82 SLL siciliani.

La seconda parte (paragrafo 3 e appendice B) si concentra sui dati censuari evidenziando un processo di progressiva despecializzazione delle attività economiche, anche se la transizione è tutt'altro che omogenea nel territorio isolano come reso evidente nella terza parte (paragrafo 4 e appendice C).

Infine, il paragrafo 5 conclude ricapitolando i principali risultati di analisi.

2. Aspetti dello sviluppo economico e ruolo del territorio

2.1 L'importanza del territorio nello studio delle dinamiche di sviluppo economico

Nella prima metà degli anni '90 la regione Sicilia ha registrato variazioni tendenziali annue negative e consecutive del PIL; bisognerà attendere il 1996 per osservare nuovamente valori positivi, il cui massimo sarà raggiunto nel 2001 con il 3,16% (a prezzi costanti '95), ma già nel 2002, anche a seguito delle sfavorevoli condizioni internazionali, si riduce allo 0,70% seguito da un nuovo aumento a 1,79% nel 2003 (Istat, 2004a). Anche i consumi hanno registrato le condizioni economiche sfavorevoli, con variazioni della spesa per consumi finali della famiglia che, dopo un tendenziale annuo del 2,37% nel 2000, raggiungono lo 0,71% nel 2001, lo 0,36% nel 2002 e lo 0,86% nel 2003. In questo periodo va notato, inoltre, il crollo degli investimenti: in Sici-

lia nel 2002, ultimo anno disponibile nel momento in cui si scrive, gli investimenti fissi lordi si riducono di ben $-7,20\%$ rispetto al 2001 (2).

Le condizioni macroeconomiche della Regione non appaiono dunque incoraggianti, anche se il dato generale nasconde i diversi andamenti settoriali e territoriali; un maggiore dettaglio è quindi quanto mai opportuno, visto che il sistema produttivo regionale si presenta fortemente eterogeneo al proprio interno (Asmundo e Mazzola, 2002).

A tal proposito, il settore che contribuisce maggiormente alla produzione è il terziario, con un incremento tra il 1980 ed il 2003 di quasi 10 punti: l'incidenza sul valore aggiunto ai prezzi base (al lordo Sifim) passa dal $68,19\%$ al $78,17\%$.

E' interessante osservare che mentre la quota del commercio aumenta dal $21,40\%$ al $24,05\%$ e così l'intermediazione monetaria dal $16,96\%$ al $27,14\%$, si riducono gli altri servizi da $29,83\%$ a $26,97\%$ ma, soprattutto, le costruzioni da $9,10\%$ a $5,75\%$, l'industria in senso stretto da $15,79\%$ a $11,47\%$ e, naturalmente, l'agricoltura da $6,92\%$ a $4,61\%$.

Il settore primario, in continua e fisiologica diminuzione pressoché in tutta Italia, mantiene in Sicilia un valore superiore a quello nazionale ma anche a quello del Mezzogiorno (rispettivamente $2,77\%$ e $4,43\%$ nel 2003), mentre di molto inferiore è l'incidenza dell'industria in senso stretto ($11,47\%$ in Sicilia a fronte del $15,08\%$ del Mezzogiorno e $22,92\%$ dell'Italia, sempre nel 2003). Il resto dei valori è simile nelle tre aree con l'importante eccezione degli altri servizi ($26,97\%$ contro il $24,67\%$ del Mezzogiorno ed il $18,39\%$ dell'Italia), interpretabile con il maggiore peso che nelle regioni meridionali ha il settore pubblico.

Da queste considerazioni emerge uno stato di intrinseca fragilità sistemica ed un ritardo di sviluppo con le aree più ricche ancora ben lontano dall'essere colmato (Svimez, 2004a e 2004b); emerge, peraltro, una condizione di forte debolezza industriale, vista anche la scarsa presenza nell'Isola di quelli che vengono definiti come fattori localizzativi in grado di attrarre e far crescere le imprese (Banca d'Italia, 2004).

Anche tra le nove province siciliane si scorgono interessanti elementi di differenziazione con riferimento al 2002, ultimo anno disponibile per dettaglio provinciale: per esempio sul valore aggiunto totale ai prezzi base (al lordo Sifim), l'incidenza dell'agricoltura varia tra il minimo di $2,08\%$ di Palermo al massimo di $8,75\%$ di Ragusa; analo-

(2) Tutti gli aggregati monetari, ove necessario, sono deflazionati a prezzi costanti '95.

gamente per l'industria in senso stretto, tra il minimo di 7,72% di Agrigento ed il massimo di 18,83% di Caltanissetta (Istat, 2004b).

Differenze dunque rilevanti, che non possono essere tralasciate da un approccio di analisi che non evidenzia sufficientemente i livelli di eterogeneità presenti all'interno dell'Isola.

D'altro lato, in anni recenti la letteratura ha enfatizzato l'importanza del contesto locale, il cui ruolo è notevolmente cresciuto nella prospettiva di un *pattern* endogeno di sviluppo implementato dall'evoluzione delle politiche comunitarie e dalla moltiplicazione delle forme di progettualità locale legate alle diverse forme di programmazione negoziata, di concertazione e di azioni localizzate in cui il rilievo dell'intervento pubblico è tutt'altro che secondario (Dps, 2005).

Di conseguenza, per un'analisi dello sviluppo che sposa nella sostanza il paradigma dell'approccio locale, è auspicabile per non dire necessario un riferimento areale quanto più possibile dettagliato dal punto di vista spaziale. La provincia non soddisfa; il ricorso ai Sistemi Locali del Lavoro (SLL) pare in questo senso il più opportuno, sia perché offre un dettaglio territorialmente ampio (la Sicilia è divisa in 82 SLL esaustivi e non sovrapposti) con buona disponibilità di informazioni statistiche, sia perché tali aggregati areali sono costruiti non con riferimento a criteri amministrativi e geografici (come invece le province) ma economici e sociali basati sui flussi di pendolarismo casa-lavoro rilevati in occasione delle indagini censuarie (Istat, 1997).

Questa scelta, peraltro, dà ragione di una maggiore esigenza informativa, base di partenza per azioni sempre più decentrate sul territorio come nel caso della Politica di Sviluppo Regionale della Commissione Europea, che pone specifica attenzione alla coesione economica e sociale dei territori regionali (European Commission, 2005).

Inoltre, soprattutto in prospettiva futura di implementazione di *policy* localmente mirate, in grado di leggere correttamente la Sicilia come entità eterogenea che necessita di una serie di interventi differenti e tra loro coerenti, la scelta dei SLL potrebbe consentire una più semplice individuazione di dinamiche "distrettuali" del sistema produttivo regionale, con riferimento per esempio ad alcuni sistemi locali legati alle nuove tecnologie dell'informazione (il cosiddetto paradistretto dell'Etna Valley) oppure alle attività artigianali tradizionali (Santo Stefano di Camastra); ciò potrà diventare elemento concreto se la realtà produttiva saprà superare una fase caratterizzata dalla spontaneità e dalla informalità per imboccare la strada di uno sviluppo indu-

striale moderno ed innovativo, capace di far leva sulla costituzione di sistemi di impresa a rete e di accordi strategici a livello territoriale (Becattini, 2000).

2.2 *La Sicilia nella realtà europea*

La discussione sulle recenti trasformazioni del sistema produttivo siciliano va opportunamente inserita nel contesto dell'ampio dibattito sulla Politica di Sviluppo Regionale, nell'idea che il territorio abbia specificità imprescindibili che richiedono politiche di governo localmente tarate (3).

Questa necessità diviene tanto più stringente quanto più eterogenea è al proprio interno l'area di riferimento; è il caso delle forti e persistenti differenziazioni tra il Nord ed il Sud d'Italia (Svimez, 2005), rese ancor più evidenti nel confronto regionale di tutte le 25 nazioni della Unione Europea (4).

Nello specifico, si considerano il PIL pro capite in "standard dei poteri d'acquisto" (SPA) (anno 2001), l'incidenza degli occupati in agricoltura, industria e servizi, il tasso di occupazione totale, maschile e femminile e, infine, il tasso di disoccupazione totale (nel 2003) (Eurostat, 2004) (5).

(3) Sul punto si sofferma autorevolmente la Commissione Europea (European Commission, 2004), mentre è interessante notare la rilevanza data in Italia alla questione tramite l'individuazione di priorità espressamente segnalate in agenda di Politica Economica (Dps, 2005).

(4) La suddivisione regionale italiana corrisponde sostanzialmente alla classificazione statistica NUTS 2; ad essa, perciò, si farà qui riferimento per individuare aree correttamente confrontabili negli altri Paesi europei. Va notato che la NUTS 2 include, invece del totale del Trentino, le due province autonome di Bolzano e Trento; le unità territoriali NUTS 2 per l'Italia sono quindi 21 e non 20. Va evidenziato, inoltre, che la costituzione dell'Europa a 25 è relativamente recente e ha acuito, come è facile immaginare, una serie di problemi di comparabilità ed omogeneità statistica dei dati diffusi dai singoli Istituti Nazionali di Statistica; le difficoltà sono sostanzialmente irrilevanti nel dettaglio generale e crescono, talvolta in modo non ancora risolto, nel confronto temporale e per un numero elevato di indicatori. Per tali ragioni, qui ci si limiterà agli ultimi dati ufficiali disponibili per gli indicatori di maggiore rilievo sostanziale, di fatto puntando all'obiettivo della massima omogeneità statistica fornendo un quadro quanto più aggiornato possibile, accettando il rischio di incrociare riferimenti temporali differenti; d'altro lato, gli indicatori di seguito considerati esprimono differenze strutturali che non mutano significativamente tra un anno e l'altro.

(5) La trasformazione in Standard dei Poteri di Acquisto (SPA) o Parità dei Poteri di Acquisto (PPA) consente la corretta confrontabilità di aggregati monetari di diversi Paesi parametrizzando rispetto ad uno standard comune le eventuali diversità sia delle valute nazionali (prezzi esterni) che dei divari tra i livelli dei prezzi interni.

Le informazioni statistiche di interesse sono disponibili per 253 o 254 regioni; di queste ben 103, tra cui la Sicilia, sono ammissibili quali aree “obiettivo 1” per il periodo di programmazione 2000-2006 (6).

La Tab.1 riporta alcune statistiche descrittive per gli indicatori scelti.

Tab.1 - Statistiche descrittive di alcuni indicatori per le regioni d'Europa.

Indicatori	Casi validi	Mediana	Media	Minimo	Massimo	Coeff. var.	SICILIA	ITALIA
PIL pro capite (SPA, 2001)	254	19.866	19.509	6.576	58.881	0,366	14.627	22.428
Quota sul totale degli occupati nell'agricoltura (% ,2003)	253	3,77	6,20	0,04	37,52	1,147	8,32	4,88
Quota sul totale degli occupati nell'industria (% ,2003)	253	27,27	27,60	8,33	46,26	0,269	20,93	31,82
Quota sul totale degli occupati nei servizi (% ,2003)	253	66,37	66,20	42,61	91,23	0,148	70,74	63,30
Ammissibilità “Obiettivo 1”	103	—	—	—	—	—	SI	—
Tasso di occupazione totale (% ,2003)	253	64,00	63,35	40,10	78,60	0,129	42,00	56,10
Tasso di occupazione maschile (% ,2003)	253	72,30	71,22	46,60	85,60	0,108	60,60	69,60
Tasso di occupazione femminile (% ,2003)	253	57,50	55,45	24,00	76,10	0,185	24,00	42,70
Tasso di disoccupazione totale (% ,2003)	254	7,35	8,95	2,00	31,80	0,634	20,10	8,70

Vale la pena notare che la maggiore variabilità (coefficiente di variazione) caratterizza l'incidenza degli occupati in agricoltura (1,147); la minima il tasso di occupazione maschile (0,108).

Con riferimento al PIL pro capite espresso in SPA (anno 2001), la Sicilia (con 14.627 euro) occupa posizioni medio-basse, con un 196° posto nella graduatoria decrescente, mentre la Polonia chiude con il minimo di 6.576 euro della Lubelskie.

(6) L'Obiettivo 1 dei Fondi Strutturali Europei attiene al sostegno dello sviluppo delle regioni meno ricche e si può considerare la priorità principale della politica di coesione dell'Unione Europea. Le regioni sono ammissibili all'Obiettivo 1 se in esse, tra l'altro, il PIL è inferiore al 75% della media comunitaria.

Il basso livello di sviluppo economico della nostra Regione è quindi ben evidente; dati ancora più recenti, relativi ad alcune caratteristiche del mercato del lavoro nell'anno 2003, sono esplicativi di altre debolezze strutturali.

Le aree dell'Europa dell'Est, tradizionalmente rurali, nonché le regioni lontane dai centri urbani, presentano nell'anno 2003 le quote più elevate dell'occupazione agricola: è il caso ancora della Polonia (Lubelskie con il 37,52% sul totale degli occupati) ma anche della Grecia (Peloponnisos, 37,28%); per contro, le aree urbane del Regno Unito occupano le ultime posizioni (con un ovvio zero per cento per Inner London). La Sicilia presenta un'incidenza dell'agricoltura relativamente elevata, con una quota di occupazione all'8,32% che corrisponde alla 56° posizione nella graduatoria decrescente dei valori.

Ragionamento analogo può proporsi per l'occupazione in industria e servizi; relativamente bassa in Sicilia è l'incidenza dell'occupazione industriale (20,93% che situa la regione al 210° posto in Europa), mentre l'alta incidenza degli occupati nei servizi (83° posto con un 70,74%) si spiega con il forte peso delle attività del pubblico impiego.

Le pessime *performance* del mercato del lavoro siciliano rispetto al contesto europeo sono rese evidenti dai tassi di occupazione e di disoccupazione.

In particolare, la Sicilia mostra un livello molto basso dell'occupazione totale, con un tasso di 42,00% che classifica la Regione al penultimo posto per l'anno 2003, avanti solo a Reunion (40,10%), a fronte dei più elevati valori di Regno Unito e Finlandia (sino al 78,60%).

Questa è anche conseguenza di una forte caratterizzazione di genere del mercato del lavoro isolano: infatti, in Sicilia, il tasso di occupazione maschile è pari al 60,60% (223° posizione tra le 253 regioni europee considerate) a fronte del 24,00% corrispondente al tasso femminile (ultima posizione in graduatoria); una forbice, questa, comune a gran parte dei Paesi euromediterranei (la Spagna, per esempio, è rispettivamente al 68,40% e 29,10% con la regione Ciudad Autónoma de Ceuta, a fronte di 77,70% e 76,10% rilevati in Svezia o 81,4% e 75,70% in Finlandia).

Il livello del tasso di disoccupazione totale è coerente con quanto finora evidenziato, risultando fortemente influenzato, oltre che dalla difficoltà nella ricerca di un lavoro della componente femminile delle forze di lavoro, anche e soprattutto da quella giovanile; la Sicilia occupa le prime posizioni per livello di disoccupazione (un tasso del

20,10% la pone al 19° posto). Va notato, infine, che il miglior valore europeo è registrato in Italia nella provincia autonoma di Bolzano (appena un tasso del 2,00% nell'anno 2003) ed il peggiore in Francia, a Reunion (31,80%).

Il Graf. A1 in appendice riporta la dispersione regionale tra il PIL pro capite espresso in euro SPA per l'anno 2001 ed il tasso di disoccupazione totale per l'anno 2003.

Si notano i valori estremi in ascissa, con il massimo in Inner London (58.881 euro pro capite), seguito dal valore di Bruxelles (48.721) e Lussemburgo (43.600), ed il massimo in ordinata, con l'elevato tasso di disoccupazione nella regione di Reunion (31,80%).

La posizione della Sicilia, con rispettivamente 14.627 euro e 20,10%, mostra, nel confronto europeo, lo stato di ritardo di sviluppo già sottolineato.

Si apprezza, poi, la generale validità della relazione inversa tra sviluppo e disoccupazione, nonché la forte differenziazione tra regioni italiane del Centro-Nord e del Mezzogiorno. Tra le due serie, per tutte le 254 regioni europee considerate, si ottiene una correlazione negativa di $-0,511$, mentre si registra un più contenuto $-0,361$ per le sole ammissibili all'Obiettivo 1; da notare l'elevato $-0,889$ quando si considerano soltanto le regioni italiane (7).

Appare di generale interesse anche la dispersione tra il tasso di occupazione maschile e femminile come riportato nel Graf. A2 in appendice riferito all'anno 2003.

La bisettrice evidenzia una comune caratterizzazione del mercato del lavoro europeo, pur con importanti distinguo: il tasso di occupazione maschile è regolarmente maggiore di quello femminile, ma nei Paesi del Nord Europa, come già detto, i valori dei due tassi sono molto prossimi (ed entrambi elevati) mentre nell'area euromediterranea il primo supera di parecchio il secondo e, comunque, entrambi si posizionano su livelli relativamente bassi; la Sicilia è nelle peggiori posizioni.

Da questo primo generico quadro emerge, soprattutto se effettuato il confronto tra le regioni d'Europa, una condizione desolante di ritardato sviluppo economico e sociale della Sicilia e ben lungi dall'essere colmato nel medio periodo.

(7) Non va sottaciuto che i confronti suggeriscono cautela stante la diversa numerosità dei gruppi di casi considerati.

In particolare, il basso livello dello sviluppo economico (tipicamente misurato dal PIL pro capite) e gli alti livelli di disoccupazione, evidenziano la necessità di una politica di intervento correttamente calibrata sul territorio, stante una differenziazione evidente già con il solo riferimento alle regioni italiane.

Ma questo quadro generale è utile nel contestualizzare l'analisi che seguirà, non va dimenticato che la Sicilia appare piuttosto eterogenea al proprio interno, evidenziandosi, per esempio, differenze degne di nota tra aree costiere ed interne oppure aree urbane e rurali. La necessità di un'analisi che chiarisca alcuni dei percorsi evolutivi della struttura economica siciliana, ed in particolare della struttura industriale, suggerisce quindi il riferimento ad aree locali quali ben potrebbero rappresentare i Sistemi Locali del Lavoro (SLL), il cui vantaggio è di non risultare costruiti su confini geo-politici, quindi di tipo amministrativo ma, semmai, risultare suddivisi tramite opportuni algoritmi di auto-contenimento degli spostamenti casa-lavoro dei residenti, in tal modo determinandosi 82 zone areali esaustive e non sovrapposte dotate di significatività economica; a tal proposito il sottoparagrafo successivo si sofferma su alcuni aspetti degli 82 SLL siciliani.

2.3 *Alcune differenze all'interno della regione: il riferimento agli 82 SLL*

L'obiettivo di un'analisi della struttura produttiva della Sicilia suggerisce il ricorso ad una classificazione sufficientemente fine. I SLL soddisfano questo criterio; inoltre permettono il riferimento ad aree costruite in base a criteri economici e sociali non dipendenti da vincoli amministrativi predeterminati.

I SLL uniscono due o più comuni contigui sulla base dell'auto-contenimento dei flussi di pendolarismo giornaliero tra luogo di residenza e luogo di lavoro; i flussi sono rilevati in occasione del Censimento, quindi rivisti ogni decennio circa.

Al momento in cui si scrive, l'Istat individua in tutta Italia 784 SLL sulla base del Censimento del 1991 (per approfondimenti, Istat, 1997); è in corso l'aggiornamento con i flussi relativi all'anno 2001 che probabilmente, quando ultimato, introdurrà alcune modificazioni non irrilevanti.

La dimensione media dei SLL appare oggi piuttosto bassa: circa 73.000 residenti considerando il dato del censimento della popolazio-

ne al 2001 (un po' più alto, circa 74.000, con le stime al 2000) (vedi Tab. A1 in appendice).

La Sicilia ha il numero maggiore di SLL con 82 unità (10,46% del totale); di questi ultimi, però, nessuno fa distretto (8).

L'assenza di distretti nell'Isola è in naturale coerenza con una netta predominanza di Sistemi Locali del Lavoro senza una qualche specializzazione produttiva, ben 65 su 82 (79,27%, a fronte del 39,67% per l'intera nazione) (Tab. 2).

Tab. 2 - Distribuzione dei SLL per specializzazione e ripartizione geografica.

	Specializzazione	Nord-Ovest	Nord-Est	Centro	Sud	Isole	Totale	Quota % Spec. ITA	SICILIA	Quota % Spec. SIC
A	Sistemi senza specializzazione	24	22	36	139	90	311	39,67	65	79,27
CA1	Sistemi del "made in Italy" (manifattura leggera)	62	49	41	54	6	212	27,04	4	4,88
CB1	Sistemi dei materiali da costruzione	22	22	15	6	7	72	9,18	3	3,66
BA3	Sistemi turistici	13	30	6	11	11	71	9,06	3	3,66
CA3	Sistemi del cuoio e della pelletteria	1	10	23	6	2	42	5,36	1	1,22
BA1	Sistemi urbani	7	5	11	8	8	39	4,97	5	6,10
CB2	Sistemi dei mezzi di trasporto	4	0	1	7	1	13	1,66	1	1,22
CB3	Sistemi degli apparecchi radiotelevisivi	1	0	2	6	0	9	1,15	0	0,00
CA2	Sistemi del tessile	6	0	1	0	0	7	0,89	0	0,00
CA4	Sistemi dell'occhialeria	0	5	0	0	0	5	0,64	0	0,00
BA2	Sistemi estrattivi	0	0	0	0	3	3	0,38	0	0,00
	<i>Totale</i>	140	143	136	237	128	784	100,00	82	100,00

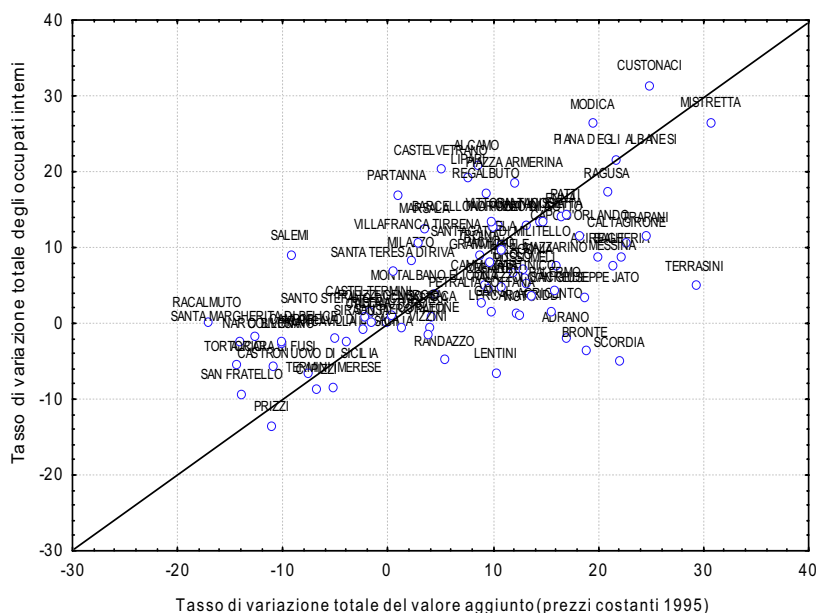
(8) Il numero di SLL oscilla tra il minimo di 4 della Valle d'Aosta ed il massimo di 82 appunto in Sicilia. Invece i distretti industriali, per costruzione, risultano particolarmente concentrati nelle regioni del Nord, mentre la Sicilia ne risulta priva. D'altro lato, la forte variabilità geografica di questi è spiegata anche dall'algorithm adottato; se al posto della media nazionale i dati venissero parametrizzati sulla media del Mezzogiorno, si individuerebbero 45 distretti meridionali di cui 2 in Sicilia: Customaci e Santo Stefano di Camastra. Esistono varie proposte e tecniche di misura per individuare i distretti industriali ma il punto non interessa gli scopi di questo lavoro (si rimanda, tra i tanti, a Cappellin, 1999).

Si nota anche una maggiore incidenza dei SLL classificati come sistemi urbani (BA1): 6,10% contro il 4,97% dell'Italia.

Queste evidenze ed in particolare l'assenza di distretti industriali in Sicilia, non devono suggerire l'idea di SLL statici o comunque con dinamiche necessariamente contenute.

A tal proposito, la Tab. A2 in appendice riporta per tutti gli 82 SLL i valori del tasso di variazione totale nel periodo 1996-2002 degli occupati interni e del valore aggiunto a prezzi costanti '95, suddivisi nel contributo dovuto all'agricoltura, industria e servizi.

Si notano alcuni casi particolari, con oscillazioni anche molto forti, in parte colti anche nel Graf. 1 e nella Tab. 3.



Graf. 1 - Dispersione dei SLL per variazione degli occupati interni e del valore aggiunto (1996-2002).

Pur a rischio di una qualche confusione, nel grafico si è preferito indicare per esteso i nomi dei SLL in modo da poter subito evidenziare i casi estremi.

Per la variazione degli occupati è il caso di Prizzi (-13,48%, a causa di un crollo dell'agricoltura e dei servizi) o di San Fratello (-9,19%,

che registra un forte arretramento del settore primario). Al contrario, Modica, Mistretta e Custonaci presentano i risultati migliori (rispettivamente 26,49%, 26,58% e 31,50%, con un forte incremento nel settore industriale e dei servizi). Per la variazione del valore aggiunto troviamo, da un lato, il pessimo risultato di Racalmuto e Tortorici (-17,18% e -14,51%, dovuti fondamentalmente al crollo dei servizi) e Naro (-14,16%, con una riduzione notevole dell'agricoltura), dall'altro, invece, gli alti e positivi valori di Custonaci (24,85%, dovuto ad incrementi di industria e servizi), Terrasini (29,20%, forte aumento nei servizi) e Mistretta (30,62%, anche qui un forte aumento nei servizi).

Si nota, comunque, la positiva correlazione (0,586) delle due serie dei tassi di variazione totale, ad indicare che la crescita economica si è sostanzialmente accompagnata ad una crescita dell'occupazione. In particolare Piana degli Albanesi, Catania, Caltavuturo e Gela registrano oscillazioni simili tra le due serie (si dispongono lungo la bisettrice del grafico con distanze assolute rispettivamente di 0,001, 0,002, 0,003 e 0,004), al contrario di Adrano, Bronte, Terrasini e Scordia (con distanze rispettivamente di 0,186, 0,223, 0,241 e 0,269). Naturalmente i risultati cambiano se si scompone il dato totale nei settori di attività economica; sul punto, per brevità, non ci si sofferma oltre.

Per quanto riguarda le statistiche riportate in Tab. 3, è interessante notare che i valori della media si posizionano al di sopra di quelli del-

Tab. 3 - Alcune statistiche descrittive per gli 82 SLL siciliani.

	Contributo settoriale OCC			Contributo settoriale VA			Tasso di var. tot. %	
	AGR	IND	SER	AGR	IND	SER	OCC	VA
Media	-2,790	1,761	6,956	-2,339	1,936	7,879	5,927	7,476
Mediana	-2,474	1,343	5,747	-1,794	0,956	8,367	5,137	9,562
Minimo	-12,933	-11,168	-7,152	-11,566	-11,471	-15,687	-13,478	-17,185
Massimo	4,682	20,197	30,742	4,071	16,988	30,393	31,503	30,622
1° Quartile	-4,218	-1,020	3,763	-3,223	-1,340	2,454	-0,541	0,215
3° Quartile	-0,743	3,809	11,434	-0,577	3,939	14,024	11,707	15,440
Range	17,615	31,365	37,894	15,637	28,459	46,079	44,981	47,807
Range interquant.	3,474	4,829	7,671	2,646	5,279	11,569	12,247	15,225
Dev. std.	3,252	4,583	6,431	2,580	5,105	9,261	8,926	11,117
Coeff. variazione	1,166	2,602	0,925	1,103	2,637	1,175	1,506	1,487
Asimmetria	-0,641	1,356	0,464	-1,349	0,879	-0,183	0,447	-0,330
Curtosi	1,107	5,308	1,455	2,672	1,383	-0,019	0,077	-0,510

la mediana per i tassi di crescita totali, mentre ancora una volta è diverso il comportamento in funzione del settore di attività economica considerato. In particolare, si evidenziano i valori negativi del contributo dell'agricoltura che, semmai ve ne fosse bisogno, confermano anche in Sicilia il progressivo indebolimento del settore primario, sia con riferimento alla variazione dell'occupazione che a quella del valore aggiunto; le stime *kernel* di densità riportate in appendice (Graf. A3) consentono di rafforzare questa considerazione (9).

È chiaro, infine, che le diverse dinamiche del valore aggiunto e degli occupati interni spiegano le oscillazioni talvolta molto ampie della produttività del lavoro (qui approssimata dal rapporto tra valore aggiunto ed occupati interni), i cui tassi di variazione nel periodo 1996-2002 sono riportati in appendice (Tab. A3) per il totale dei settori e, al solito, per la scomposizione in agricoltura, industria e servizi. Si nota appena per il totale delle attività economiche il massimo valore negativo di Racalmuto (-17,38%), seguito da Salemi (-16,85%) e Partanna (-13,86%), al contrario di quanto si registra a Terrasini (22,92%), Bron-
te (23,11%) e Scordia (28,25%).

In tutti i SLL citati, ed in genere in gran parte degli 82 casi, quando la produttività si riduce (quindi si hanno tassi di variazione negativi), il contributo dell'industria tende ad essere di segno opposto o comunque molto contenuto; viceversa, quando la produttività cresce (quindi tassi di variazione positivi), il ruolo dell'industria è preminente (10).

Queste considerazioni suggeriscono un'ultima riflessione circa i modelli di declino industriale che investirebbero in modo preoccupante parte della Sicilia.

Come si è verificato anche qui per gli 82 SLL negli anni 1996-2002, la struttura produttiva isolana si è spostata sempre più verso le attività terziarie, con un peso via via inferiore del settore primario e, talvolta, anche di quello industriale; ciò è evidente nelle dinamiche del valore aggiunto e degli occupati interni.

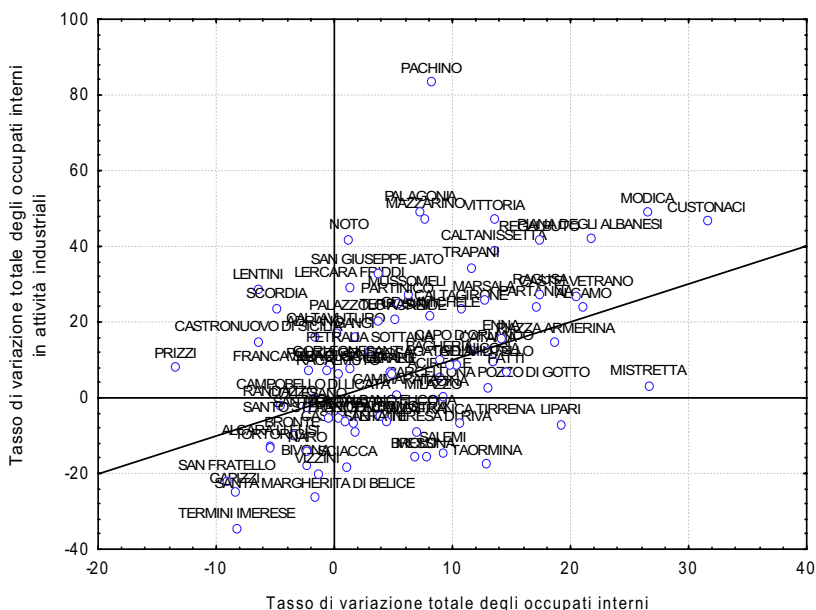
Quest'ultimo aspetto può generare una qualche preoccupazione: una riduzione dell'occupazione industriale (in modo specifico quella

(9) Si è utilizzato uno stimatore non parametrico di densità kernel di tipo gaussiano mentre la costante di *smoothing* è stata scelta applicando la formula di Silverman (per un primo approfondimento si rimanda a Silverman, 1986).

(10) Prendendo ad esempio i due SLL estremi, Racalmuto registra una variazione della produttività del lavoro nelle attività industriali di -0,82% a fronte del -17,38% del totale, mentre Scordia ha un elevato 82,59% dell'industria per un 28,25% del totale (Tab. A3 in appendice).

manifatturiera), se non accompagnata da una variazione positiva del totale o, quantomeno, un mantenimento dei livelli ad opera di un effetto sostituzione del terziario, manifesterebbe un preoccupante indebolimento della struttura economica con il manifestarsi di processi di declino industriale in grado di rendere più difficili i tentativi di promuovere lo sviluppo (Erba e Pellegrini, 2000).

A tal proposito, il Graf. 2 riporta il diagramma di dispersione tra il tasso di variazione totale degli occupati interni nel complesso delle attività economiche e quello degli occupati interni nelle attività industriali; il riferimento è sempre a tutti gli 82 SLL siciliani nel periodo 1996-2002.



Graf. 2 - SLL in declino industriale.

I SLL con processi di deindustrializzazione in atto, in numero di 29 (il 35,37% del totale), si posizionano nel quadrante inferiore del grafico. Ciò non basta a denunciare una condizione di intrinseca debolezza di queste realtà locali né un'ipotesi di assenza di prospettive di sviluppo nel medio periodo. Infatti, tra questi 29 SLL, va effettuata un'ulte-

riori distinzione tra quelli con variazione positiva di tutti gli occupati (deindustrializzazione positiva) e quelli con variazione negativa (deindustrializzazione negativa). Nel primo caso è attivo un processo di compensazione con i servizi, con una struttura produttiva in evoluzione, nel secondo è attivo un processo di indebolimento ed involuzione. Al primo gruppo appartengono, pur con intensità differenti, 15 SLL: Agrigento, Brolo, Casteltermeni, Lipari, Messina, Milazzo, Montalbano Elicona, Palermo, Ribera, Salemi, Santa Teresa di Riva, Santo Stefano di Camastra, Sciacca, Taormina e Villafranca Tirrena; al secondo, 14 SLL: Alcara Li Fusi, Bivona, Bronte, Campobello di Licata, Capizzi, Collesano, Naro, Randazzo, San Fratello, San Piero Patti, Santa Margherita di Belice, Termini Imerese, Tortorici e Vizzini.

Va notato, poi, che dei 53 SLL restanti soltanto 8 si posizionano nel grafico in alto a sinistra, cioè mostrano un incremento dell'occupazione industriale ma una riduzione dell'occupazione totale: Adrano, Castronuovo di Sicilia, Corleone, Francavilla di Sicilia, Lentini, Prizzi, Scordia e Siracusa. Gli ultimi 45, posti nel grafico in alto a destra, mostrano incrementi sia dell'occupazione industriale che di quella totale.

3. La struttura produttiva siciliana tramite l'analisi degli ultimi censimenti

3.1 *Il sistema industriale italiano: aspetti generali*

Il sistema industriale italiano presenta, rispetto alla media degli altri Paesi europei, alcune caratteristiche peculiari: a) un denso tessuto di medie e piccole imprese (MPI) efficienti e dinamiche; b) una struttura produttiva articolata per sistemi locali (SL), nei quali si realizza una complessa divisione del lavoro; c) una specializzazione produttiva in tecnologie mature dalle quali, tuttavia, fluiscono prodotti aventi caratteristiche di qualità (Balloni e Cucculelli, 1998).

I prodotti cosiddetti "tradizionali" rappresentano ancora una quota assai consistente della produzione italiana, e queste produzioni (tessile, abbigliamento, pelli e mobilio) mantengono pressoché invariata (poco meno del 20%) la loro incidenza sull'output manifatturiero.

Il modello di specializzazione produttiva dell'Italia è stato, perciò, caratterizzato dalla prevalenza di settori tradizionali, dalla meccanica specializzata, e in misura marginale da settori ad alta tecnologia o

capaci di economie di scala. Un modello, questo, che ha favorito le piccole imprese più flessibili e che ha creato e mantenuto occupazione, soprattutto nelle aree dei distretti industriali dove nascono efficienti sistemi di relazioni tra le imprese che si scambiano tecnologie, servizi, *know how*.

A fronte di un numero limitato di piccole imprese operanti sulla frontiera della tecnologia, la maggior parte di esse realizza infatti innovazioni di tipo adattivo o incrementale, che svolgono un ruolo molto importante nel sostenere la competitività dei settori in cui queste imprese operano (tipicamente quelli più “maturi”), ma hanno un impatto assai poco significativo in termini di *spillover* all’interno del sistema produttivo.

Inoltre, il consolidarsi del modello di specializzazione italiano si è tradotto in un ulteriore ampliamento della rilevanza delle imprese di medio-piccola dimensione: ancora negli anni Novanta, così come già avvenuto nei due decenni precedenti, si è assistito ad una costante ascesa, non solo in termini produttivi ma anche in termini di occupazione, dello *small business sector*.

3.2 *La distribuzione territoriale delle MPI del settore industriale*

È opportuno però sottolineare, al di là dei noti vantaggi, particolarmente in termini di flessibilità, di creazione di nuova occupazione e di contenimento dei costi di produzione, quali sono, in genere, gli inconvenienti della piccola dimensione delle unità produttive: a) difficoltà di crescita; b) instabilità; c) bassa sopravvivenza; d) limitate competenze manageriali (Carone e Iacobucci, 1999).

Nonostante i problemi e i limiti che le piccole e medie imprese possono incontrare nella loro attività di produzione, è diffuso il convincimento che esse continueranno a giocare un ruolo preminente nell’industria manifatturiera italiana. Tale conclusione viene supportata dalla verifica del processo di convergenza verso la piccola dimensione (10-49 addetti), considerata la più efficiente dati le tendenze prevalenti nella domanda e nella tecnologia (Istituto Tagliacarne, 2005).

La convinzione che il sentiero evolutivo che sta percorrendo l’economia italiana conduca ad un modello di capitalismo nel quale è centrale il ruolo delle medie e piccole imprese e dei sistemi locali si basa, come già detto, sul fatto che le imprese minori sono ordinate in sistemi territoriali diffusi e speciali.

Le analisi compiute sulla struttura e sulle modalità di sviluppo dei distretti industriali hanno fornito spiegazioni interessanti sul fenomeno: i distretti sono il prodotto dell'interazione tra imprese di dimensione medio-piccola e contesto istituzionale, economico e culturale.

Nell'ambito di tali sistemi produttivi l'integrazione territoriale delle imprese tende, pertanto, a mitigare gli effetti negativi della contenuta dimensione aziendale e del modello prevalente di specializzazione produttiva.

Ciò è evidente con specifico riferimento ai Sistemi Locali del Lavoro; si è già visto come dei 784 SLL individuati dall'Istat, 419 si trovano nel Centro-Nord e 365 nel Mezzogiorno. Questi ultimi sono caratterizzati – ancor più di quelli del Centro-Nord – dalla prevalenza delle unità produttive di minori dimensioni: quasi il 97% delle unità locali del Mezzogiorno ha meno di 10 addetti e occupa circa il 60% degli addetti della ripartizione. Nel Centro-Nord, l'incidenza di questo segmento dimensionale è – in termini di unità locali e di addetti – rispettivamente del 94% e 48%.

Le specificità del sistema industriale italiano, attinenti in particolare alla specializzazione settoriale e all'articolazione dimensionale della struttura produttiva, si ripresentano quindi in modo più accentuato nel settore industriale siciliano mentre la presenza delle MPI non assolve al ruolo di struttura portante delle realtà socio-produttive locali, come invece avviene nelle aree distrettuali del resto del Paese.

Le differenze (sostanziali, per certi versi) tra i due livelli territoriali emergono principalmente in relazione alla distribuzione settoriale delle unità economiche, ma anche con riferimento alla struttura organizzativa dei processi produttivi, alla forma giuridica ed al rapporto con il contesto socio-economico in cui le imprese operano.

3.3 *Caratterizzazione delle MPI del sistema produttivo siciliano*

Dalla peculiarità e dall'interazione dei diversi fenomeni che condizionano l'organizzazione e l'attività delle imprese, deriva il quadro tipico che connota il sistema produttivo isolano, riassumibile nei punti seguenti.

1) Modalità di conduzione dell'unità economica: a) accentramento delle responsabilità gestionali nella persona dell'imprenditore; b) orientamento dei processi decisionali verso obiettivi di breve periodo; c) predominanza delle strategie competitive su quelle animate da uno spiri-

to di cooperazione nell'ambito di imprese caratterizzate dalla stessa specializzazione produttiva; d) scarso interesse, o addirittura diffidenza, come effetto dei comportamenti indicati sub b) e c), ad instaurare o ad aderire a moderne forme di aggregazione imprenditoriale basate sulla coordinazione delle attività produttive.

2) Organizzazione della produzione: a) processi produttivi caratterizzati da scarsa propensione all'innovazione; b) riluttanza a superare le soglie dimensionali che garantiscono benefici in termini di esenzioni fiscali e che, inoltre, consentono una maggiore capacità di controllo del fattore lavoro (meno costoso e con un minor grado di sindacalizzazione rispetto a quello occupato nelle imprese più grandi).

3) Modalità di assunzione dei dipendenti: a) limitato ricorso ai canali convenzionali; b) impiego diffuso dei canali informali (sulla base di informazioni reperite tra la cerchia ristretta delle proprie conoscenze e di indicazioni provenienti dalle unità del settore produttivo di appartenenza o dalle associazioni di categoria).

4) Collocazione della produzione: a) i prodotti sono destinati al mercato locale e solo in pochi casi superano l'ambito della regione di appartenenza; b) i rapporti con i clienti sono caratterizzati da tecniche tradizionali di superamento degli opposti interessi ma solo in casi sporadici vengono perseguite moderne strategie di vendita tendenti alla fidelizzazione della clientela (campagne promozionali, assistenza post vendita, certificazione di qualità, *customer satisfaction*).

5) Fattori di freno, esterni all'impresa, che ne limitano il processo evolutivo: a) perifericità rispetto ai mercati di approvvigionamento delle materie prime e a quelli di destinazione dei prodotti; b) mancanza di competitività rispetto alle unità produttive di altre realtà territoriali con particolare riferimento alla dotazione di adeguate risorse in campo manageriale; c) criminalità organizzata; d) carenza di infrastrutture; e) eccessiva onerosità del credito.

Le caratteristiche del sistema produttivo siciliano, sinteticamente descritte in precedenza, condizionano in buona misura la richiesta di forza lavoro, specialmente di quella più qualificata che, nonostante gli incrementi dei tassi di occupazione (cfr. par. 2.3), incontra molte difficoltà nell'inserirsi in modo attivo nel mercato del lavoro locale e, in misura crescente, è costretta a cercare nel resto del Paese la possibilità di un impiego corrispondente alle proprie competenze e la realizzazione delle legittime aspettative riguardo ad una adeguata remunerazione del lavoro prestato.

Assume, pertanto, rilievo la possibilità di condurre un'analisi strutturale, sia dal profilo settoriale che territoriale, a partire dai censimenti del 1981, del 1991 e del 2001.

3.4 *Aspetti metodologici dell'analisi della struttura produttiva*

Nell'ambito delle analisi strutturali e quindi nel processo di costruzione della base informativa di statistica economica volta a misurare e valutare il livello di crescita e trasformazione di un sistema economico assume particolare rilievo l'analisi della struttura produttiva dove per struttura produttiva s'intende il complesso di attività condotte in forma organizzata da parte di soggetti con un fine istituzionale orientato alla generazione di profitto.

In particolare poiché l'obiettivo di questo lavoro è quello di esaminare le caratteristiche localizzative e le forme insediative delle unità economiche e le specializzazioni produttive delle singole zone ci occuperemo delle analisi territoriali.

Oltre al tempo, un importante criterio ordinatore della informazione statistico-economica è costituito dallo spazio.

Le rilevazioni censuarie, come è noto, producono una notevole quantità di informazioni che consentono di definire accuratamente i profili economici di unità territoriali quali le regioni, le province, i comuni. I dati di un censimento sono rappresentabili sotto forma di matrici di dati $n \times p$. La generica riga della matrice definisce il *profilo* della struttura economica dell'unità territoriale i attraverso le determinazioni delle p variabili ad essa corrispondenti, mentre una generica colonna fornisce una rappresentazione spaziale del fenomeno definito dalla variabile h .

Una parte notevole delle analisi statistiche territoriali è riconducibile allo studio di matrici di questo tipo ed al confronto nel tempo. Esse presentano, oltre ai tratti comuni con le analisi che possono essere effettuate su matrici di dati qualsiasi, alcune specificità connesse alle caratteristiche proprie dei dati economici territoriali.

La localizzazione e la diffusione nello spazio dei fenomeni economici, la loro interdipendenza, la diversificazione interregionale costituiscono i temi che ricorrono più di frequente nelle analisi territoriali dei dati censuari.

Tali aspetti possono essere studiati utilizzando in modo appropriato specifiche tecniche statistiche tra cui i coefficienti di localizzazione e di specializzazione.

La distribuzione delle attività economiche nel territorio e tra i settori produttivi costituisce, nell'analisi dei dati del censimento industriale, uno degli aspetti di più immediato interesse. Essa dipende da diversi fattori di tipo economico, sociale, storico, geografico, che rappresentano gli elementi fondamentali delle teorie della localizzazione.

Dal punto di vista statistico, l'analisi dell'articolazione spaziale e settoriale dell'occupazione, ma anche delle unità locali o delle imprese, può avvalersi di diversi indici descrittivi di facile quantificazione ed interpretazione.

Per studiare la specializzazione produttiva dell'insieme delle unità locali operanti in ogni singola area, impieghiamo dapprima l'indice di localizzazione settoriale, poiché fornisce informazioni dettagliate, a livello locale e in ciascuna attività economica. Considerando la matrice $n \times p$ (n unità territoriali $i = 1, \dots, n$ e p settori $h = 1, \dots, p$) illustrata precedentemente, tale indice sarà uguale a:

$$Q_{ih} = (A_{ih}/A_i) / (A_{.h}/A_{..}) \quad [1]$$

dove:

Q_{ih} = coefficiente di localizzazione del settore h nell'area i ;

A_{ih} = numero di addetti alla sottoclasse di attività economica h nell'area i ;

$A_{.h}$ = $\sum_i A_{ih}$ ovvero il numero complessivo di addetti alla sottoclasse h ;

A_i = $\sum_h A_{ih}$ ovvero il numero complessivo di addetti dell'area i ;

$A_{..}$ = $\sum_i A_i = \sum_h A_{.h}$ ovvero totale complessivo degli addetti delle aree.

I rapporti pongono a confronto le quote di occupazione settoriale, calcolate a livello di area, con quelle al livello complessivo delle aree. Quanto all'interpretazione dei valori del quoziente, un valore superiore all'unità ($Q_{ih} > 1$) indica una specializzazione dell'area nella particolare sottoclasse di attività produttiva; il contrario avviene se il quoziente assume valori inferiori all'unità ($Q_{ih} < 1$).

Il quoziente di localizzazione [1] fornisce dunque elementi di giudizio sulla specializzazione settoriale delle economie regionali: quozienti sensibilmente superiori all'unità indicano, infatti, un elevato grado di specializzazione nei settori corrispondenti.

Per avere una visione d'insieme della specializzazione produttiva, utilizziamo i rapporti di composizione dei coefficienti di localizzazio-

ne per definire il coefficiente di specializzazione (secondo una nota formulazione di Beguin, cfr. Del Colle e Esposito, 2000):

$$S_i = 1/2 \sum_h |(A_{ih}/A_i) - (A_{.h}/A_{..})| \quad [2]$$

dove:

S_i = coefficiente di specializzazione nell'area i ;

A_{ih} = numero di addetti alla sottoclasse di attività economica h nell'area i ;

$A_{.h}$ = $\sum_i A_{ih}$ ovvero il numero complessivo di addetti alla sottoclasse h ;

A_i = $\sum_h A_{ih}$ ovvero il numero complessivo di addetti dell'area i ;

$A_{..}$ = $\sum_i A_i = \sum_h A_{.h}$ ovvero totale complessivo degli addetti delle aree.

L'indice di specializzazione [2] è un indice di dissomiglianza che assume un valore minimo, pari a zero, quando l'area considerata presenta una composizione settoriale dell'occupazione identica a quella nazionale (quindi assenza di specializzazione) e valore massimo, pari ad uno, quando tutta l'occupazione è concentrata in una sola attività (massima specializzazione).

Il Graf. B1 in appendice riporta la rappresentazione degli indici di Specializzazione delle nove province siciliane nelle ultime tre rilevazioni censuarie: la sintesi grafica è ottenuta tramite la rappresentazione di un grafico "scatola a baffi" che consente di confrontare la distribuzione degli indici di specializzazione comunale di ogni singola provincia, evidenziando di volta in volta eventuali valori anomali o valore "estremi alla distribuzione.

L'analisi del grafico permette di valutare la sostanziale e diffusa riduzione della specializzazione nelle diverse province siciliane: si osserva in particolare come i valori mediani provinciali nel 1981 si attestano sopra lo 0,10, nelle successive rilevazioni censuarie si abbassano sensibilmente.

3.5 Profili settoriali nei SLL siciliani

La disponibilità dei coefficienti di localizzazione ha consentito di analizzare il complesso delle attività industriali a livello delle sezioni C (estrazione di minerali), E (produzione e distribuzione di energia elettrica, gas e acqua) ed F (costruzioni), mentre per la sezione D (attività

manifatturiere) si è proceduto ad un esame dettagliato delle 12 sotto-sezioni in essa comprese.

Come già precisato in precedenza, sono 65 su 82 i SLL siciliani classificati come sistemi senza specializzazione, mentre i restanti 17 sono definiti come segue: 5 sistemi urbani (BA1), 3 sistemi turistici (BA3), 4 sistemi del *made in Italy*, (CA1), 1 sistema del cuoio e della pelletteria (CA3), 3 sistemi dei materiali da costruzione (CB1) e 1 sistema dei mezzi di trasporto (CB2). È evidente, quindi, la prevalenza di SLL non caratterizzati e se a questi si aggiungono i SLL la cui specializzazione non rientra fra le attività di trasformazione dei prodotti (precisamente, i sistemi urbani e quelli turistici), la percentuale dei SLL “non industriali” ammonta all’89% circa.

Attraverso l’esame del livello e dell’andamento nel periodo 1981-2001 dei quozienti di localizzazione riferiti a ciascuna sezione (e sotto-sezione) del comparto industriale, è possibile verificare i mutamenti strutturali avvenuti nei periodi intercensuari.

Le attività comprese nella sezione C (Estrazione di minerali) sono presenti in quasi tutti i sistemi locali della Sicilia, anche se con una tendenziale diminuzione dei livelli dei quozienti di localizzazione, mentre una forte concentrazione permane nei SLL specializzati in questo tipo di attività (CB1).

Per quanto riguarda le industrie alimentari, delle bevande e del tabacco (sottosezione DA), i già bassi livelli dei quozienti si vanno ulteriormente riducendo a motivo di una progressiva despecializzazione che caratterizza tutti i SLL, particolarmente nell’intervallo tra i due ultimi censimenti.

Relativamente all’industria del tessile e dell’abbigliamento (DB), all’elevata e crescente specializzazione nei sistemi del *made in Italy* (CA1) e del cuoio e della pelletteria (CA3), contrasta la progressiva diminuzione che si riscontra nel resto dei SLL siciliani, presso i quali l’attività considerata perde rapidamente importanza, specialmente per il gruppo che rientra nella tipologia A.

In modo sorprendente, le attività rivolte alla lavorazione del cuoio e della pelletteria (DC) scompaiono dall’unico sistema classificato secondo tale specializzazione (CA3) e, inoltre, perdono rilevanza nei sistemi despecializzati (A) i cui quozienti di localizzazione si abbassano notevolmente.

Per l’industria del legno e dei prodotti in legno (DD), diffusa solamente nei sistemi despecializzati, si nota la permanenza di livelli con-

sistenti dei quozienti di localizzazione nel periodo di tempo considerato (1981-2001).

Riguardo alle attività rivolte alla fabbricazione della carta e alla stampa ed editoria (DE), solo il sistema con la tipologia CA3 mostra valori dei quozienti di localizzazione consistenti e in crescita, mentre la diffusione di questa attività nei sistemi senza specializzazione (A) si attenua notevolmente dal 1981 al 2001, sia come numero di SLL interessati a questo tipo di lavorazione che come intensità della sua presenza, come denunciano i bassi livelli di specializzazione.

Per le attività riunite nella sottosezione DF (fabbricazione di coke, raffinazione del petrolio), soltanto tre sistemi locali (Siracusa, Milazzo e Gela) registrano valori elevati dei quozienti di localizzazione che per Gela, però, mostrano una tendenza alla diminuzione denunciando, pertanto, una progressiva diminuzione di importanza di tali attività industriali nel sistema produttivo gelese.

Anche per le attività produttive rivolte alla fabbricazione di prodotti chimici e di fibre sintetiche (DG) emergono gli apprezzabili valori della specializzazione dei SLL di Siracusa e Gela (anche se in calo negli intervalli censuari); tra i sistemi senza specializzazione, livelli consistenti dei quozienti di localizzazione si notano relativamente ai SLL di Ragusa e Salemi, pur essi in calo negli ultimi censimenti.

Per la sottosezione DH, comprendente le attività rivolte alla fabbricazione di articoli in gomma e materie plastiche, solo alcuni dei SLL in cui è suddiviso il territorio siciliano, indipendentemente dalla tipologia di appartenenza, presentano quozienti di localizzazione elevati ma in calo.

Per le attività della sottosezione DI (fabbricazione di prodotti della lavorazione di minerali non metalliferi) soltanto i tre sistemi classificati nella tipologia CB1 (materiali da costruzione) mostrano quozienti di localizzazione con livelli molto elevati e due di essi (Custonaci e S. Stefano di Camastra) manifestano altresì una crescita robusta nei tre censimenti.

E' opportuno sottolineare che circa la metà dei sistemi privi di specializzazione (A) registra una lieve tendenza alla crescita dei quozienti di localizzazione per le attività comprese in DI.

La sottosezione DJ (produzione di metallo e fabbricazione di prodotti in metallo) è presente soltanto in due sistemi urbani (Gela e Siracusa) con livelli elevati dei corrispondenti quozienti di localizzazione ed una tendenza all'aumento negli anni censuari. Inoltre, nel folto gruppo dei sistemi senza specializzazione, il 25% circa mostra valori non trascurabili dei quozienti di localizzazioni relativi a questa attività.

Per le attività dirette alla fabbricazione di macchine ed apparecchi meccanici (DK), i valori più alti dei quozienti mostrano una distribuzione territoriale simile a quella verificata per la precedente sottosezione DJ; emergono, inoltre, i SLL di Custonaci e Termini Imerese e, con valori consistenti, il 25% dei SLL senza specializzazione.

Per le attività concernenti la fabbricazione di macchine elettriche ed apparecchiature elettriche ed ottiche (DL) si notano valori elevati dei quozienti solo nell'ambito dei SLL di Catania e di Palermo (sistemi urbani) e, nell'ultimo Censimento, in qualcuno dei SLL senza specializzazione.

Solo Palermo, poi, presenta valori dei quozienti superiori a 1 (ma in calo nel periodo considerato) con riferimento ai quozienti della sottosezione DM (mezzi di trasporto) oltre, naturalmente, all'elevatissima specializzazione (relativamente alla Sicilia) di Termini Imerese con livelli, tuttavia, tendenti alla diminuzione.

Riguardo alle altre industrie manifatturiere (DN), i quozienti di localizzazione denunciano un certo addensamento delle eterogenee attività classificate in questa sottosezione all'interno dei SLL di Palermo e di Catania e, in misura molto più limitata, in alcuni dei SLL senza specializzazione.

Le attività di produzione e distribuzione di energia, gas e acqua (sezione E) mostrano una certa concentrazione nei SLL di Siracusa e Palermo, tra i sistemi urbani, ed anche a Termini e a Lipari per evidenti ragioni connesse alla attività economica prevalente, nel primo caso, e alla collocazione geografica nel secondo. Come in altri casi, anche tali valori mostrano una netta tendenza al declino.

Invece, in buona parte dei SLL senza specializzazione le attività della precedente sezione registrano valori consistenti ed in crescita.

Infine, le attività della sezione F (costruzioni) sono presenti nella maggior parte dei SLL indipendentemente dalla tipologia di appartenenza; in aggiunta, i dati mostrano livelli elevati di concentrazione di tali attività in più della metà dei sistemi senza specializzazione.

3.6 *Una sintesi delle dinamiche industriali emerse*

L'analisi precedente consente di effettuare alcune valutazioni: il principale fatto stilizzato è l'evidente processo di despecializzazione produttiva che coinvolge la gran parte degli 82 SLL siciliani nelle branche produttive dell'Industria.

Solo i tre sistemi dei materiali da costruzione presentano una specializzazione elevata non soltanto nelle attività estrattive ma anche nella lavorazione dei minerali non metalliferi, nell'industria del legno e nella fabbricazione di macchine e apparecchi meccanici (comparti strettamente collegati all'attività principale dei SLL della tipologia CB1).

Anche per i cinque sistemi urbani, presso i quali si concentrano le attività di trasporto e gli altri servizi di supporto all'attività economica, si rilevano livelli consistenti dei quozienti di localizzazione in diverse attività del settore industriale.

I 65 SLL che risultano privi di fattori di localizzazione specifici, data la loro numerosità (79% del totale) e la loro eterogeneità, mostrano livelli non trascurabili della specializzazione in diversi comparti delle attività di trasformazione dei prodotti.

Tuttavia, la tendenza che emerge per quasi tutti gli SLL di questo folto gruppo è sostanzialmente quella della perdita di rilievo delle attività industriali presenti all'inizio del periodo in esame.

Infine, per i SLL che rientrano nelle altre categorie (BA3, CA1, CA3 e CB2), complessivamente in numero di nove, sono in atto processi di despecializzazione a ritmo sostenuto in molte attività industriali più o meno connesse alla stessa tipologia di specializzazione; alcuni si stanno despecializzando anche con riferimento all'attività principale in base alla quale erano stati inizialmente classificati.

Le mappe costruite sulla base degli indici di specializzazione a livello di SLL, negli anni 1981, 1991 e 2001, mostrano in modo immediato quanto risulta dall'esame dei quozienti (si veda il Graf. B2 in appendice).

La continua despecializzazione verificatasi nell'ultimo ventennio ha ridotto l'estensione delle aree a più elevata specializzazione che, attualmente, si concentrano prevalentemente nelle province di Palermo e Messina.

Si è determinato, pertanto, un progressivo isolamento dei SLL a più alta specializzazione che formano, ora, nuclei che si distaccano in modo rilevante dagli altri SLL con diversa specializzazione perdendo, in tal modo, i benefici collegati alla prossimità spaziale tra le unità produttive anche di settori differenti.

4. La despecializzazione produttiva: luci ed ombre

L'approfondimento dell'analisi strutturale condotto nei paragrafi precedenti, ha consentito di supportare l'ipotesi che la struttura pro-

duittiva siciliana ha assistito in questi anni ad una progressiva frammentazione del tessuto produttivo, fortemente collegata ad una bassa caratterizzazione delle aree produttive, che conseguentemente ha impedito l'emergere di specializzazioni produttive significative.

Tale macro tendenza, tuttavia, non nasconde alcune interessanti articolazioni: per evidenziarle meglio sono state osservate le variazioni dell'Indice di Specializzazione nei settori dell'Industria nei due intervalli censuari 1981-1991 e 1991-2001.

Un fenomeno che si valuta in maniera più marcata qualora si confrontino le tendenze territoriali dei sistemi produttivi regionali con il totale dei SLL del Mezzogiorno.

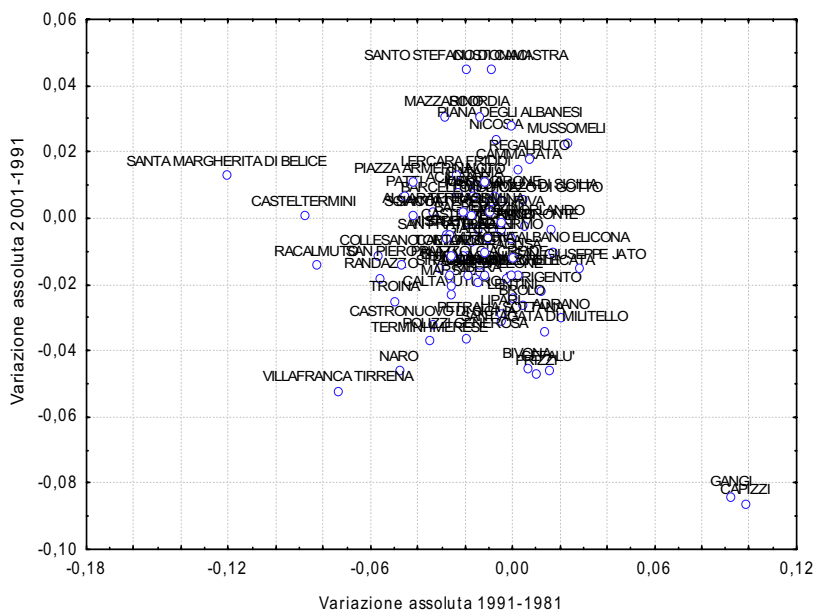
L'indice di Specializzazione, come già detto in precedenza (par. 3.4), offre una misura sintetica della caratterizzazione produttiva di un'area sub-regionale (il SLL) relativamente ad un'area territorialmente sovraordinata (in questo caso il Mezzogiorno).

Le sue variazioni nel tempo, quindi, indicano un incremento (o un decremento) di tale differenza relativa, che può essere interpretato come specializzazione (o riduzione della despecializzazione) o decremento (o riduzione della specializzazione).

Dopo aver calcolato le variazioni per differenza dell'indice nei due intervalli considerati, gli 82 SLL siciliani sono stati classificati secondo il segno della variazione (positivo o negativo): il Graf. 3 evidenzia come nel primo decennio analizzato in questo studio oscillano su un campo di variazione più ampio di quello del secondo decennio.

Ciò suggerisce che, in termini relativi rispetto alla distribuzione regionale delle attività industriali i SLL hanno avuto un forte ridimensionamento tra l'81 e il '91, mentre nel decennio successivo si nota una seppur leggera inversione di tendenza.

Una simile evidenza empirica va compresa alla luce della generale e continuata contrazione degli addetti nei settori industriali: mentre il peggioramento della specializzazione tra il 1981 e il 1991 avviene in una sostanziale stazionarietà del numero degli addetti industriali, tra il 1991 e il 2001 il totale degli addetti nelle attività dell'Industria diminuisce fortemente, pertanto il "miglioramento" degli indici di specializzazione va interpretato come una diversa redistribuzione degli addetti nelle varie branche produttive. Si tratta dunque di una debole ma interessante inversione di tendenza che mette in luce una "condensazione" di aree produttive che si coglie attraverso la partizione territoriale dei sistemi locali.



Graf. 3 - Variazioni dell'Indice di Specializzazione Industriale (1981-1991 e 1991-2001).

Del resto una tale dinamicità, contestuale alla maggiore attenzione alle politiche territoriali e di sviluppo locale, è stata più volte messa in luce dagli osservatori più attenti (11).

Una simile interpretazione viene anche offerta da chi ha osservato come nel processo di despecializzazione della struttura produttiva meridionale, è possibile riconoscere i segni degli effetti dell'integrazione che tende ad attenuare le diversità regionali favorendo la diffusione delle attività manifatturiere e, quindi "conducendo al graduale affievolimento della relazione fra economie di scala e localizzazione nelle regioni periferiche" (Scalera, 2001, pag. 87).

Un'ulteriore considerazione può essere effettuata mettendo a confronto le variazioni nei due intervalli: se si rapportano le variazioni positive e negative rispetto ai rispettivi valori medi, si possono cogliere delle differenti dinamiche che suggeriscono una più complessa articola-

(11) Tra gli altri si vedano le considerazioni di Trigilia in Crouch *et alii* (2004), pp. 114-119.

zione della generale despecializzazione produttiva che investe i comparti dell'Industria siciliana.

Dalle due variazioni negli intervalli censuari sono state costruite due variabili politomiche (*Despec1* e *Despec2*) che vengono rappresentate attraverso una "tabella di mobilità" (Tab. C1 in appendice): l'analisi descrittiva di tale tabella prefigura l'ipotesi che vi sia stato un forte rallentamento del processo di despecializzazione, infatti il numero di SLL che passa da un'accelerazione ad un rallentamento del processo è inferiore a quello che, negli intervalli considerato, compie il percorso inverso (12).

Soffermandosi soprattutto sul significato di tali risultati, è possibile concludere con una valutazione sostanzialmente più positiva: in termini relativi a quella meridionale, la struttura produttiva regionale, nel periodo di riferimento, mostra un processo lento ma interessante di attenuazione in termini di riduzione dell'indice di specializzazione.

Ciò può essere spiegato da una crescente attenzione alla riqualificazione dei territori e di una attenzione mirata alla valorizzazione delle vocazioni produttive locali da parte di politiche economiche locali mirate.

Occorrerà certamente continuare a monitorare tali fenomeni, approfondendone in termini più qualitativi lo spessore e l'entità.

5. Conclusioni

La crescente attenzione verso una territorializzazione delle politiche economiche suggerisce una forte attenzione alle dinamiche evolutive di sistemi locali.

L'apparato produttivo siciliano nell'ultimo trentennio si è caratterizzato per un forte e radicale cambiamento: la frammentazione delle unità produttive, la riduzione della dimensione media, il processo di terziarizzazione che coinvolge soprattutto le piccole imprese (13).

(12) Tale dinamica può ulteriormente essere messa in evidenza dallo studio dei "rapporti fra prodotti incrociati" della tabella di mobilità: in particolare è possibile osservare che per un SLL che subisce un'accelerazione (forte o debole) ve ne sono 9 che si avvantaggiano di un rallentamento (forte o debole) della despecializzazione.

(13) La specializzazione terziaria è tuttavia legata in settori dei servizi non avanzati in particolare alla presenza rilevante del commercio e della piccola e media distribuzione.

La struttura produttiva regionale caratterizzata da un forte processo di despecializzazione, si presenta a prima vista caratterizzata da una riduzione significativa delle caratterizzazioni produttive locali.

Un'analisi più approfondita ha consentito di mettere in evidenza una differenziazione di tale dinamica, cogliendo una interessante inversione di tendenza nella distribuzione territoriale delle attività produttive, in particolare nei comparti dell'industria.

Si tratta di una tendenza, già forte se la si rapporta al sistema regionale delle imprese, ma che appare confermata nel confronto con l'intero apparato produttivo meridionale.

Nell'ultimo decennio considerato nel presente studio, la realtà produttiva meridionale ha messo in luce numerose aggregazioni territoriali di imprese talvolta caratterizzate come aree di "dinamismo imprenditoriale", altre volte interpretate come forme innovative di distretti di tipo marshalliano.

Un dinamismo che accompagna l'evoluzione delle politiche industriali del Mezzogiorno, da un lato maggiormente mirate sui singoli progetti industriali (ad esempio la legge n. 488/92), da un lato più attente alla creazione di aree sistema radicate nel tessuto economico locale (si veda l'esperienza dei patti territoriali o dei progetti integrati territoriali).

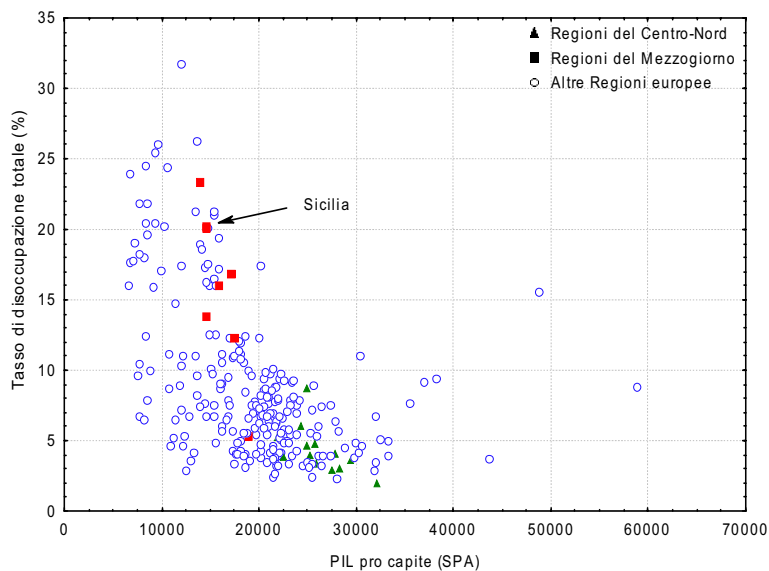
Agli aspetti territoriali si aggiungono quelli settoriali: le imprese operanti nel territorio regionale sono piccole ed hanno, in genere, una bassa propensione all'innovazione ed all'apertura verso i mercati esterni.

Il consolidamento delle specializzazioni produttive nei settori tradizionali si allinea con un modello produttivo sostanzialmente poco competitivo: anche se emergono alcune importanti tendenze nella riorganizzazione di alcune filiere, in particolar modo quelle legate all'agroalimentare, nella maggiore attenzione alla qualità e alle produzioni ad elevato valore aggiunto (si pensi alle emergenti imprese del vitivinicolo, particolarmente vivaci nel proporsi ai mercati esteri).

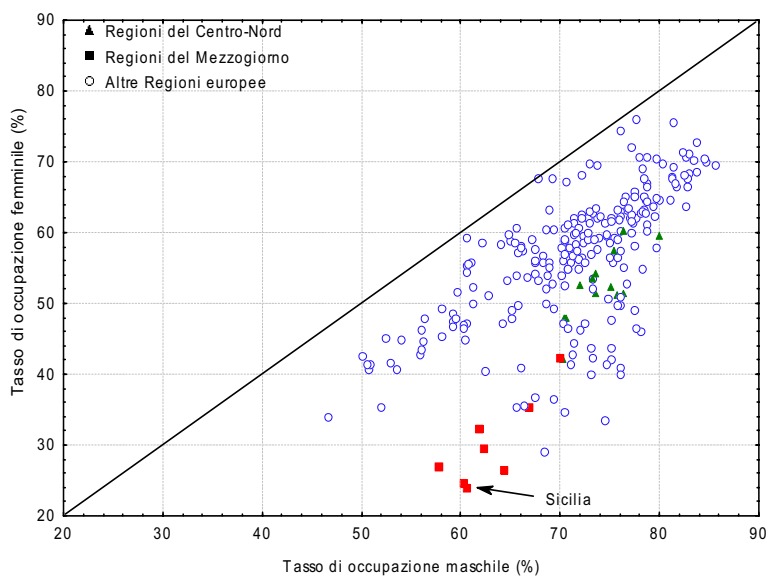
Tuttavia l'analisi condotta nel presente contributo ha inteso soprattutto sottolineare gli aspetti strutturali, le macrotendenze di fondo che coinvolgono le imprese del settore industriale regionale; in tale senso è emersa una generale evidenza verso un sistema produttivo isolano in lento declino, al cui rallentamento ed arresto contribuirebbe in misura rilevante la capacità degli operatori economici a livello locale di cogliere sempre più e sempre meglio la sfida dell'innovazione e della valo-

rizzazione delle vocazioni territoriali e dei saperi locali. D'altronde, se è vero che i processi dell'imprenditorialità locale da un lato devono maggiormente cogliere le sfide di una competizione a livello globale, dall'altro non possono non fare leva su una serie di fattori endogeni che, opportunamente qualificati, costituiscono una riserva importante per l'evoluzione dell'economia regionale verso livelli simili a quelli delle altre regioni europee.

APPENDICE A



Graf. A1 - Dispersione tra PIL pro capite e tasso di disoccupazione totale.



Graf. A2 - Dispersione tra tasso di occupazione maschile e femminile.

Tab. A1 - Distribuzione regionale dei SLL e dimensione media.

	SLL	Di cui distretto	Multi-regionali	Multi-provinc.	Numero comuni	Popolazione Censimento 2001	Nr. medio com. x SLL	Popolazione per SLL	Superf. terr. (Kmq.)
Piemonte . . .	50	16	1	20	1.185	4.193.250	24	83.865	498,8
Valle d'Aosta .	4	0	1	1	78	123.800	20	30.950	830,9
Lombardia . . .	70	42	5	28	1.555	9.066.840	22	129.526	343,3
Trentino A.A..	35	4	2	4	341	945.240	10	27.007	392,7
Veneto	48	34	4	15	581	4.507.860	12	93.914	383,8
Friuli V.G. . .	12	3	1	3	216	1.190.370	18	99.198	643,7
Liguria	16	1	5	6	245	1.573.500	15	98.344	356,9
Emilia Rom. . .	48	24	6	13	348	3.983.830	7	82.996	463,9
Toscana	51	19	4	18	287	3.518.110	6	68.983	448,9
Umbria	16	5	1	2	85	798.850	5	49.928	510,3
Marche	42	34	4	7	249	1.496.580	6	35.633	240,6
Lazio	27	2	4	8	376	5.122.230	14	189.712	638,4
Abruzzo	24	6	2	6	310	1.229.850	13	51.244	445,7
Molise	10	0	1	3	135	320.880	14	32.088	439,5
Campania . . .	65	4	2	9	545	5.692.160	8	87.572	207,2
Puglia	39	3	2	5	263	4.043.200	7	103.672	517,0
Basilicata . . .	25	0	1	3	122	565.870	5	22.635	360,6
Calabria	74	2	2	6	413	2.022.460	6	27.331	206,0
Sicilia	82	0	0	11	390	4.968.990	5	60.597	313,5
Sardegna	46	0	0	2	377	1.631.880	8	35.476	523,7
Centro-Nord . .	419	184	38	125	5.546	36.520.460	13	87.161	426,1
Nord-Ovest . . .	140	59	12	55	3.063	14.957.390	22	106.839	414,3
Nord-Est	143	65	13	35	1.486	10.627.300	10	74.317	434,7
Centro	136	60	13	35	997	10.935.770	7	80.410	429,4
Mezzogiorno . .	365	15	10	45	2.555	20.475.290	7	56.097	336,4
Italia	784	199	48	170	8.101	56.995.750	10	72.699	384,4

Tab. A2 - Tassi di variazione degli occupati interni e del valore aggiunto e corrispondenti contributi settoriali (1996-2002).

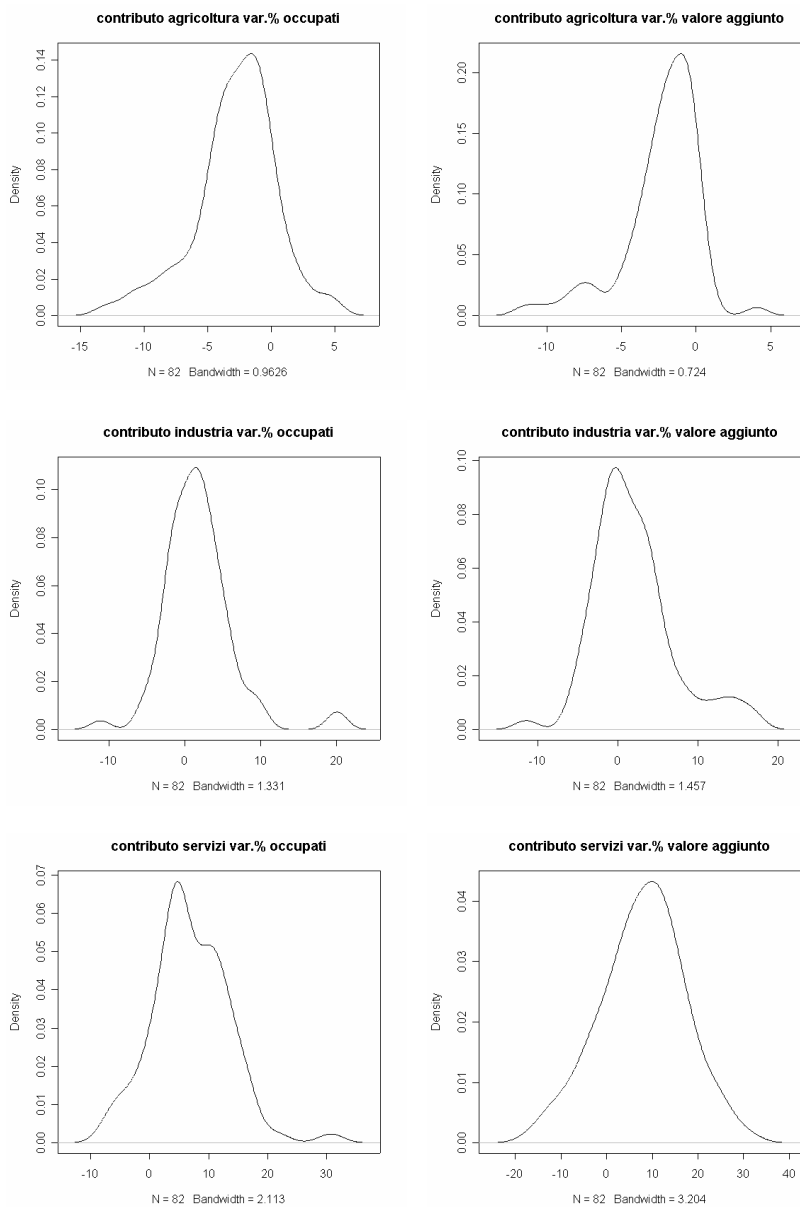
Nr.	Codice Spec	Denominazione SII	Occupati interni 1996-2002				Valore aggiunto a prezzi costanti 1996-2002			
			AGR	IND	SER	TOT	AGR	IND	SER	TOT
1	A	Acireale	-1,86	1,20	9,49	8,82%	-0,58	0,64	19,77	19,83%
2	A	Adrano	-9,87	2,34	5,78	-1,75%	-0,63	-0,68	18,16	16,85%
3	A	Agrigento	-1,56	-1,13	4,25	1,56%	-0,78	-1,34	17,56	15,44%
4	A	Alcamo	1,45	5,74	13,86	21,05%	-4,11	3,35	9,22	8,45%
5	CA1	Alcara Li Fusi	-12,93	-2,31	9,76	-5,48%	-3,12	-0,47	-7,35	-10,94%
6	A	Bagheria	-2,47	2,27	9,03	8,83%	-1,79	11,38	12,53	22,12%

Tab. A2 (segue) - Tassi di variazione degli occupati interni e del valore aggiunto e corrispondenti contributi settoriali (1996-2002).

Nr.	Codice Spec	Denominazione SII	Occupati interni 1996-2002				Valore aggiunto a prezzi costanti 1996-2002			
			AGR	IND	SER	TOT	AGR	IND	SER	TOT
7	A	Barcellona P. Gotto ..	-0,67	0,56	13,08	12,97%	-0,23	2,84	7,77	10,38%
8	A	Bivona	-2,87	-3,44	3,85	-2,46%	-2,09	-3,24	-4,93	-10,26%
9	CA1	Brolo	-4,13	-5,09	15,96	6,74%	-1,78	3,46	10,18	11,86%
10	CA1	Bronte	-5,86	-2,35	4,69	-3,52%	0,42	3,90	14,45	18,77%
11	A	Caltagirone	-1,49	3,04	9,20	10,75%	-2,30	-0,05	24,95	22,60%
12	A	Caltanissetta	-2,20	5,61	10,05	13,46%	-1,36	5,61	10,16	14,42%
13	A	Caltavuturo	-6,44	2,82	3,74	0,12%	-4,37	3,08	1,10	-0,20%
14	A	Cammarata	0,49	0,28	4,38	5,16%	-2,47	2,56	8,97	9,06%
15	A	Campobello Licata ..	-4,07	-0,05	2,23	-1,89%	-2,74	0,21	-2,63	-5,17%
16	A	Capizzi	-10,47	-4,77	6,72	-8,52%	0,51	-5,08	-2,26	-6,84%
17	A	Capo D'Orlando	-2,52	3,11	11,12	11,71%	-1,02	11,39	7,81	18,18%
18	A	Casteltermini	-0,97	-1,62	4,23	1,64%	-3,38	-0,88	2,45	-1,80%
19	A	Castelvetrano	2,33	5,88	12,23	20,44%	-6,99	0,51	11,41	4,94%
20	A	Castronuovo Sicilia ..	-4,91	1,42	-2,94	-6,44%	-4,77	1,52	-4,42	-7,67%
21	BA1	Catania	-0,80	2,34	11,43	12,97%	-0,34	-0,46	13,96	13,16%
22	A	Cefalù	-0,63	0,87	4,45	4,69%	-1,90	3,85	7,55	9,50%
23	A	Collesano	-1,21	-0,31	-0,82	-2,34%	-3,43	4,49	-11,32	-10,26%
24	A	Corleone	-3,86	0,84	2,62	-0,40%	-1,21	2,14	2,93	3,87%
25	CB1	Custonaci	-0,26	19,94	11,83	31,50%	-3,72	16,99	11,59	24,85%
26	A	Enna	-0,91	2,39	12,64	14,12%	-0,07	-0,25	16,69	16,38%
27	CA3	FrancaVillia Sicilia ...	-6,85	1,19	3,44	-2,22%	-4,41	-0,76	1,21	-3,96%
28	A	Gangi	-7,86	2,80	6,75	1,69%	-2,55	9,04	3,26	9,75%
29	BA1	Gela	-0,93	2,50	8,78	10,35%	0,03	-3,32	14,02	10,74%
30	A	Giarre	-2,06	1,46	5,39	4,80%	-1,90	1,39	11,20	10,69%
31	A	Grammichele	0,82	2,31	4,96	8,10%	-1,80	-1,42	12,77	9,55%
32	A	Lentini	-4,02	4,67	-7,15	-6,49%	-2,47	-1,61	14,31	10,23%
33	A	Lercara Friddi	-2,01	3,66	-0,37	1,29%	-1,92	7,93	6,18	12,19%
34	BA3	Lipari	-1,45	-1,02	21,65	19,19%	-0,58	-5,17	13,31	7,57%
35	A	Marsala	-1,93	4,95	9,62	12,63%	-3,22	1,06	5,58	3,42%
36	A	Mazzarino	-8,64	4,18	12,09	7,63%	-6,74	0,38	22,35	15,99%
37	BA1	Messina	-0,28	-1,89	9,89	7,72%	-0,15	-2,92	24,35	21,28%
38	A	Milazzo	-1,12	-0,11	9,53	8,30%	-0,53	-3,13	5,76	2,10%
39	A	Mistretta	-4,52	0,35	30,74	26,58%	-0,16	0,39	30,39	30,62%
40	A	Modica	0,18	9,77	16,55	26,49%	-5,56	7,10	17,91	19,44%
41	A	Montalbano Elicona ..	-4,80	-0,98	9,62	3,84%	-1,69	6,38	-0,33	4,36%
42	A	Mussomeli	-4,08	5,62	4,61	6,15%	-3,55	1,94	14,59	12,98%
43	A	Naro	-4,22	-2,03	3,91	-2,34%	-8,97	-1,71	-3,48	-14,16%
44	A	Nicosia	-4,20	1,66	15,97	13,43%	0,80	2,44	11,38	14,62%

Tab. A2 (*segue*) - Tassi di variazione degli occupati interni e del valore aggiunto e corrispondenti contributi settoriali (1996-2002).

Nr.	Codice Spec	Denominazione SII	Occupati interni 1996-2002				Valore aggiunto a prezzi costanti 1996-2002			
			AGR	IND	SER	TOT	AGR	IND	SER	TOT
45	A	Noto	-3,56	8,46	-3,74	1,16%	-2,11	-0,31	14,93	12,50%
46	A	Pachino	-6,79	20,20	-5,21	8,20%	-3,50	-1,89	14,96	9,57%
47	A	Palagonia	1,85	3,81	1,60	7,25%	-2,99	3,59	12,16	12,75%
48	A	Palazzolo Acreide . . .	-2,05	4,11	1,60	3,67%	-1,40	3,94	10,99	13,53%
49	BA1	Palermo	-0,43	-0,84	5,67	4,40%	-0,17	-1,73	17,70	15,80%
50	A	Partanna	4,68	3,99	8,37	17,04%	-8,01	3,47	5,36	0,82%
51	A	Partinico	-3,07	4,19	4,21	5,33%	-1,96	8,77	6,23	13,04%
52	A	Patti	-0,74	1,27	13,99	14,51%	-0,43	0,69	16,57	16,83%
53	A	Petralia Sottana	-3,18	1,90	4,19	2,90%	-1,44	4,00	6,25	8,81%
54	A	Piana Albanesi	-3,21	8,72	16,14	21,66%	-1,60	13,23	9,92	21,56%
55	A	Piazza Armerina	-0,18	2,84	16,00	18,65%	-0,31	4,68	7,64	12,00%
56	A	Polizzi Generosa	-3,45	0,93	3,76	1,25%	-1,29	1,94	-0,44	0,21%
57	A	Prizzi	-7,80	0,83	-6,50	-13,48%	-3,95	1,26	-8,41	-11,10%
58	A	Racalmuto	-1,26	1,18	0,31	0,24%	-2,36	0,86	-15,69	-17,18%
59	A	Ragusa	0,78	4,61	11,98	17,37%	-2,85	3,95	19,70	20,79%
60	A	Randazzo	-4,64	-0,32	0,29	-4,67%	-3,04	-3,14	11,50	5,31%
61	A	Regalbutto	-4,32	9,89	11,74	17,32%	-10,31	14,23	5,40	9,32%
62	A	Ribera	0,78	-0,75	0,27	0,31%	0,02	2,30	-3,90	-1,58%
63	A	Salemi	4,46	-2,53	7,15	9,09%	-7,30	-3,62	1,63	-9,29%
64	A	San Fratello	-11,19	-2,04	4,04	-9,19%	-0,09	-4,17	-9,69	-13,95%
65	A	San Giuseppe Jato . . .	-3,40	6,48	0,52	3,60%	4,07	16,27	-1,68	18,66%
66	A	San Piero Patti	-5,65	-0,68	5,79	-0,54%	-1,24	-0,17	2,60	1,19%
67	A	Santa Marg. Belice . . .	2,72	-5,50	1,15	-1,63%	-5,00	-3,77	-3,93	-12,70%
68	A	Santa Teresa di Riva . .	-2,80	-1,50	11,22	6,92%	-2,50	-1,07	4,06	0,49%
69	A	Sant'Agata Militello . .	-4,55	1,44	12,84	9,73%	-1,08	2,19	9,58	10,69%
70	CB1	Santo Stefano Cam. . .	-3,17	-1,30	5,39	0,91%	-2,49	-0,66	0,92	-2,24%
71	A	Sciaccia	-0,24	-3,80	5,08	1,04%	-0,75	-1,72	6,53	4,06%
72	A	Scordia	-3,68	3,32	-4,60	-4,95%	-1,38	14,36	8,92	21,90%
73	BA1	Siracusa	-0,71	1,83	-1,84	-0,73%	-0,46	-5,38	3,37	-2,48%
74	BA3	Taormina	-1,80	-2,28	16,96	12,88%	-0,64	-1,20	11,77	9,93%
75	CB2	Termini Imerese	-1,78	-11,17	4,67	-8,28%	-0,80	-11,47	7,01	-5,26%
76	BA3	Terrasini	-2,48	2,44	5,15	5,11%	-1,37	5,10	25,48	29,20%
77	A	Tortorici	-8,64	-1,55	4,77	-5,42%	-0,48	-1,68	-12,35	-14,51%
78	A	Trapani	-0,69	4,72	7,58	11,61%	-2,35	3,91	22,93	24,49%
79	CA1	Troina	-2,67	0,16	11,69	9,19%	-0,16	5,21	3,55	8,59%
80	CB1	Villafranca Tirrena . . .	-1,25	-2,07	13,94	10,62%	-0,84	-0,66	4,30	2,80%
81	A	Vittoria	-2,96	6,77	9,65	13,46%	-8,01	6,32	11,44	9,75%
82	A	Vizzini	-5,05	-2,08	5,72	-1,41%	-11,57	-2,35	17,72	3,80%



Graf. A3 - Stime Kernel di densità per gli 82 SLL siciliani.

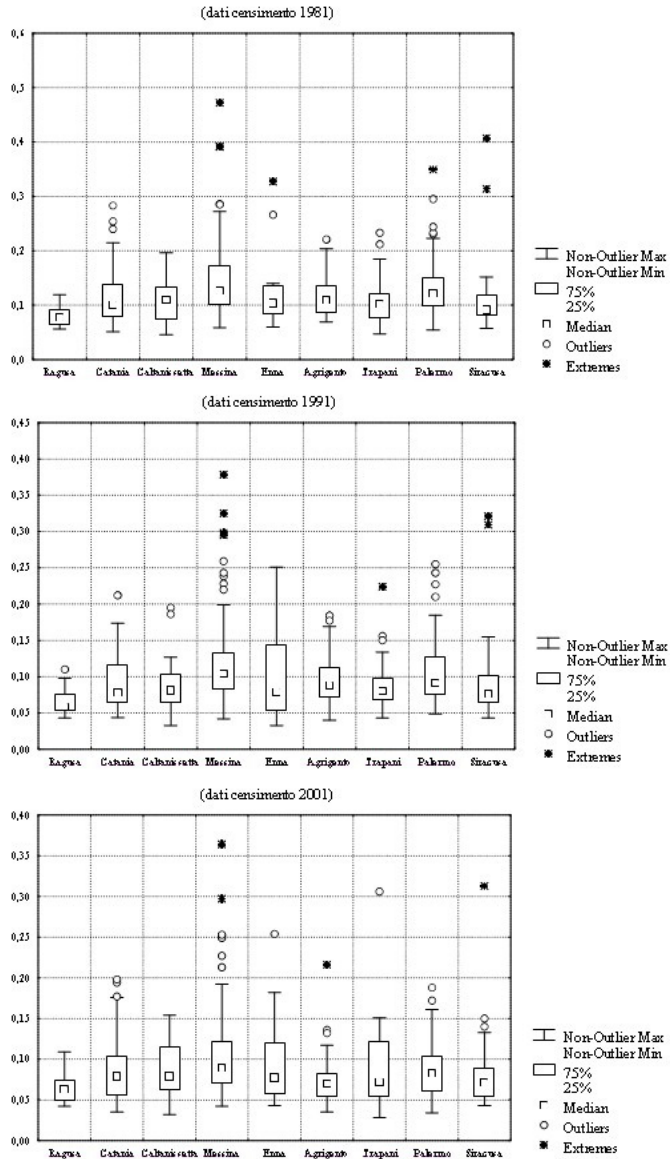
Tab. A3 - Tassi di variazione % della produttività del lavoro per settore economico e totale (1996-2002).

Numero	Spec.	SLL	AGR	IND	SER	TOT
1	A	Acireale	1,57	-2,16	10,73	10,12
2	A	Adrano	57,27	-18,85	11,90	18,93
3	A	Agrigento	-4,42	-0,75	15,90	13,67
4	A	Alcamo	-35,43	-5,77	-9,18	-10,41
5	CA1	Alcara Li Fusi	24,08	10,47	-26,54	-5,78
6	A	Bagheria	-3,69	64,65	2,32	12,21
7	A	Barcellona Pozzo di Gotto ..	1,65	14,82	-7,30	-2,29
8	A	Bivona	-7,27	-8,55	-12,29	-8,00
9	CA1	Brolo	0,92	34,32	-13,12	4,79
10	CA1	Bronte	33,04	35,67	9,34	23,11
11	A	Caltagirone	-22,38	-19,51	15,76	10,70
12	A	Caltanissetta	-23,44	2,46	-0,83	0,84
13	A	Caltavuturo	0,13	7,95	-6,80	-0,32
14	A	Cammarata	-20,88	14,83	4,01	3,71
15	A	Campobello di Licata	-10,33	1,75	-7,01	-3,34
16	A	Capizzi	25,66	4,33	-24,61	1,83
17	A	Capo D'Orlando	2,51	41,44	-6,14	5,79
18	A	Casteltermini	-21,27	2,62	-3,68	-3,39
19	A	Castelvetrano	-47,26	-19,39	-2,92	-12,87
20	A	Castronovo di Sicilia	-2,95	6,09	-0,85	-1,32
21	BA1	Catania	-5,36	-13,05	2,74	0,16
22	A	Cefalù	-17,37	37,92	2,64	4,60
23	A	Collesano	-9,82	45,28	-16,78	-8,11
24	A	Corleone	6,14	22,52	-0,59	4,29
25	CB1	Custonaci	-46,67	-1,48	-3,05	-5,06
26	A	Enna	8,56	-15,09	3,18	1,98
27	CA3	Francavilla di Sicilia	-8,45	-13,42	-4,59	-1,78
28	A	Gangi	12,00	57,07	-7,89	7,93
29	BA1	Gela	5,93	-15,45	9,69	0,35
30	A	Giarre	-15,87	0,99	5,58	5,62
31	A	Grammichele	-13,16	-31,79	6,84	1,34
32	A	Lentini	-0,65	-31,99	39,53	17,88
33	A	Lercara Friddi	-8,05	27,63	8,49	10,76
34	BA3	Lipari	-3,23	-27,01	-8,50	-9,75
35	A	Marsala	-29,08	-16,07	-6,57	-8,18
36	A	Mazarrino	-9,47	-28,37	5,73	7,77
37	BA1	Messina	8,43	-5,32	15,55	12,59
38	A	Milazzo	4,68	-9,25	-4,55	-5,73
39	A	Mistretta	23,09	2,77	-6,52	3,19
40	A	Modica	-31,37	-3,58	-1,98	-5,58
41	A	Montalbano Elicona	4,69	40,77	-16,77	0,49

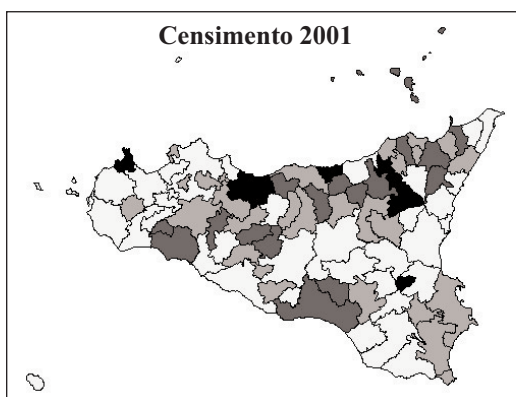
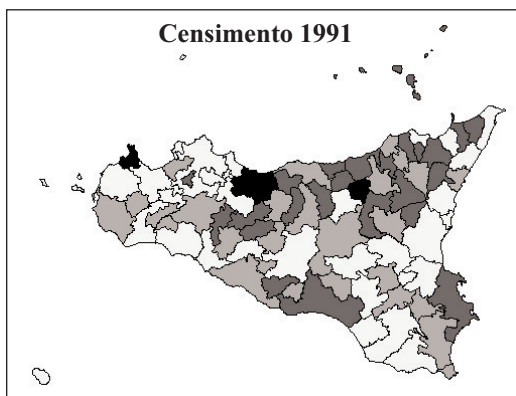
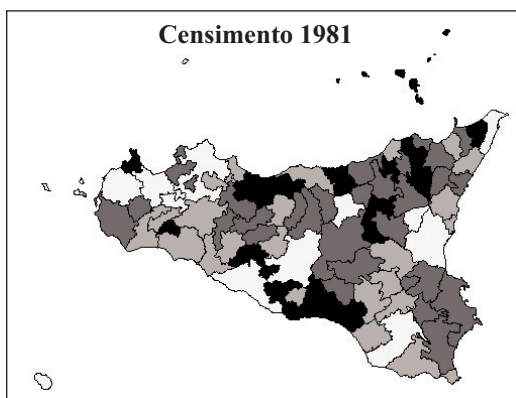
Tab. A3 (segue) - Tassi di variazione % della produttività del lavoro per settore economico e totale (1996-2002).

Numero	Spec.	SLL	AGR	IND	SER	TOT
42	A	Mussomeli	-11,47	-12,50	11,83	6,44
43	A	Naro	-24,55	1,44	-15,05	-12,10
44	A	Nicosia	31,72	4,19	-9,15	1,05
45	A	Noto	3,64	-31,28	32,94	11,21
46	A	Pachino	20,25	-56,17	32,80	1,27
47	A	Palagonia	-13,62	-2,44	14,33	5,12
48	A	Palazzolo Acreide	-2,30	0,61	12,93	9,52
49	BA1	Palermo	6,23	-5,72	13,48	10,92
50	A	Partanna	-38,73	-1,05	-8,95	-13,86
51	A	Partinico	-4,04	42,76	1,50	7,32
52	A	Patti	-4,64	-1,80	0,98	2,02
53	A	Petralia Sottana	3,16	20,95	1,42	5,74
54	A	Piana degli Albanesi	0,37	32,19	-10,36	-0,08
55	A	Piazza Armerina	-1,66	8,32	-13,31	-5,60
56	A	Polizzi Generosa	6,00	22,27	-5,74	-1,02
57	A	Prizzi	4,11	10,79	-0,12	2,75
58	A	Racalmuto	-28,33	-0,82	-20,62	-17,38
59	A	Ragusa	-31,71	-4,68	8,22	2,91
60	A	Randazzo	-9,20	-17,79	15,66	10,47
61	A	Regalbuto	-29,74	21,64	-14,56	-6,81
62	A	Ribera	-2,35	26,36	-5,78	-1,88
63	A	Salemi	-37,74	-9,65	-11,72	-16,85
64	A	San Fratello	24,15	-27,76	-24,76	-5,24
65	A	San Giuseppe Jato	69,03	68,64	-3,08	14,53
66	A	San Piero Patti	10,40	3,39	-6,03	1,74
67	A	Santa Margherita di Belice ...	-26,77	8,58	-9,27	-11,26
68	A	Santa Teresa di Riva	-15,81	-0,51	-9,77	-6,01
69	A	Sant'Agata di Militello	16,70	5,07	-5,30	0,87
70	CB1	Santo Stefano di Camastra ..	-14,95	1,23	-7,16	-3,12
71	A	Sciacca	-6,42	9,33	0,33	2,98
72	A	Scordia	0,93	82,59	20,59	28,25
73	BA1	Siracusa	-7,51	-22,12	8,14	-1,76
74	BA3	Taormina	-1,03	5,38	-6,63	-2,61
75	CB2	Termini Imerese	2,02	2,34	3,25	3,29
76	BA3	Terrasini	-18,15	65,56	20,07	22,92
77	A	Tortorici	25,37	-5,96	-23,12	-9,62
78	A	Trapani	-39,27	-4,53	17,16	11,55
79	CA1	Troina	10,09	16,35	-16,19	-0,54
80	CB1	Villafranca Tirrena	-0,30	4,18	-13,64	-7,06
81	A	Vittoria	-29,05	-0,83	0,18	-3,27
82	A	Vizzini	-31,94	-12,66	13,65	5,29

APPENDICE B



Graf. B1 - Distribuzione provinciale degli indici di specializzazione ai censimenti 1981, 1991 e 2001.



Le 4 classi di valori
(tonalità crescenti di grigio)

1981

- 1 0,145 ≤ 0,243
- 2 0,243 ≤ 0,307
- 3 0,307 ≤ 0,397
- 4 0,397 ≤ 0,580

1991

- 1 0,128 ≤ 0,245
- 2 0,245 ≤ 0,319
- 3 0,319 ≤ 0,442
- 4 0,442 ≤ 0,602

2001

- 1 0,129 ≤ 0,226
- 2 0,226 ≤ 0,296
- 3 0,296 ≤ 0,397
- 4 0,397 ≤ 0,665

Graf. B2 - Georappresentazione degli indici di specializzazione in 4 classi di valori.

APPENDICE C

Tab. C1 - La "mobilità" della despecializzazione dei SLL siciliani.

		Calcolo della variazione dell'Indice di Specializzazione per ciascun intervallo di tempo e osservazione dell'entità (segno) della variazione	Calcolo delle medie per sottogruppo (comuni con var. > 0 e comuni con var. <= 0) e confronto tra ogni singolo valore della var. del singolo SLL con la var. media del gruppo di riferimento			
		Variazione maggiore di 0	Valori superiori alla media			
		Variazione maggiore di 0	Valori inferiori o uguali alla media			
		Variazione minore o uguale a 0	Valori superiori o uguali alla media			
		Variazione minore o uguale a 0	Valori inferiori alla media			
DESPEC 2 1991-2001						
		<i>Forte rallentamento</i>	<i>Debole rallentamento</i>	<i>Debole accelerazione</i>	<i>Forte accelerazione</i>	
	<i>forte rallentamento</i>	1	0	1	2	4
DESPEC 1	<i>debole rallentamento</i>	2	1	4	7	14
1981-1991	<i>debole accelerazione</i>	5	11	20	4	40
	<i>forte accelerazione</i>	1	6	11	6	24
		9	18	36	19	82
DESPEC 2 1991-2001						
		<i>Forte rallentamento</i>	<i>Debole rallentamento</i>	<i>Debole accelerazione</i>	<i>Forte accelerazione</i>	
	<i>forte rallentamento</i>	25,00	0,00	25,00	50,00	100,00
DESPEC 1	<i>debole rallentamento</i>	14,29	7,14	28,57	50,00	100,00
1981-1991	<i>debole accelerazione</i>	12,50	27,50	50,00	10,00	100,00
	<i>forte accelerazione</i>	4,17	25,00	45,83	25,00	100,00
		10,98	21,95	43,90	23,17	100,00
DESPEC 2 1991-2001						
		<i>Forte rallentamento</i>	<i>Debole rallentamento</i>	<i>Debole accelerazione</i>	<i>Forte accelerazione</i>	
	<i>forte rallentamento</i>	11,11	0,00	2,78	10,53	4,88
DESPEC 1	<i>debole rallentamento</i>	22,22	5,56	11,11	36,84	17,07
1981-1991	<i>debole accelerazione</i>	55,56	61,11	55,56	21,05	48,78
	<i>forte accelerazione</i>	11,11	33,33	30,56	31,58	29,27
		100,00	100,00	100,00	100,00	100,00
DESPEC 2 1991-2001						
		<i>Forte rallentamento</i>	<i>Debole rallentamento</i>	<i>Debole accelerazione</i>	<i>Forte accelerazione</i>	
	<i>forte rallentamento</i>	1,22	0,00	1,22	2,44	4,88
DESPEC 1	<i>debole rallentamento</i>	2,44	1,22	4,88	8,54	17,07
1981-1991	<i>debole accelerazione</i>	6,10	13,41	24,39	4,88	48,78
	<i>forte accelerazione</i>	1,22	7,32	13,41	7,32	29,27
		10,98	21,95	43,90	23,17	100,00

BIBLIOGRAFIA

- ASMUNDO A., MAZZOLA F. (2002), *Modelli di industrializzazione ed evoluzione dei sistemi locali manifatturieri in Sicilia*, in Butera S., Ciaccio G. (a cura di) "Aspetti e tendenze dell'economia siciliana", Il Mulino, Bologna.
- BANCA D'ITALIA (2004), *Economie locali, modelli di agglomerazione e apertura internazionale*, Banca d'Italia, Roma.
- BALLONI V., CUCCULELLI M. (1998), *Modelli di impresa e di industria nei contesti di competizione globale: stili organizzativi e rapporti tra imprese nelle aree ad alta intensità industriale*, L'Industria, n. 2.
- BECATTINI G. (2000), *Il distretto industriale. Un nuovo modo di interpretare il cambiamento economico*, Rosenberg & Sellier, Torino.
- CAPPELLIN R. (1999) (a cura di), *Criteri di definizione dei distretti industriali e strategie di politica industriale in Lombardia*, Ricerca Regione Lombardia - Liuc, Regione Lombardia.
- CARONE A., IACOBUCCI D. (1999), *I gruppi di piccole e medie imprese nell'industria italiana*, in "La questione dimensionale nell'industria italiana", di Traù F. (a cura di), Il Mulino, Bologna.
- CROUCH C., LE GALÈS P., TRIGILIA C., VOELZKOW H. (2004), *I sistemi di produzione in Europa*, Il Mulino, Bologna.
- DEL COLLE E., ESPOSITO G.F. (2000), *Economia e statistica per il territorio*, Franco Angeli, Milano.
- DPS (2005), *Rapporto Annuale 2004*, Ministero dell'Economia, Roma.
- ERBA A., PELLEGRINI G. (2000), *Modelli di declino industriale*, Atti della XL Riunione Scientifica SIS, Firenze, 26-28 aprile.
- EUROPEAN COMMISSION (2004), *Al servizio delle Regioni*, (trad. it.), European Commission, Bruxelles.
- EUROPEAN COMMISSION (2005), *Towards a New Partnership for Growth, Jobs and Cohesion - Third Progress Report on Cohesion*, European Commission, Bruxelles.
- EUROSTAT (2004), *Indicatori socio-economici delle regioni dell'Unione Europea (25 Paesi)*, Eurostat, Bruxelles.
- ISTAT (1997), *I Sistemi Locali del Lavoro 1991*, Argomenti n.10, Istat, Roma.

ISTAT (2004a), *Conti economici regionali, anni 1980-2003*, Istat, Roma.

ISTAT (2004b), *Conti economici provinciali, anni 1995-2002*, Istat, Roma.

ISTITUTO TAGLIACARNE (2005), *Le piccole e medie imprese nell'economia italiana, rapporto 2004*, Franco Angeli, Milano.

SCALERA D. (2001), *Integrazione europea e localizzazione industriale: uno studio sulle regioni del Mezzogiorno italiano*, Studi economici, n. 75.

SILVERMAN B.W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, Londra.

SVIMEZ (2004a), *Il Mezzogiorno nell'Europa, ed il mondo mediterraneo e balcanico*, Quaderno Nuova Serie n. 2, Svimez, Roma.

SVIMEZ (2004b), *Rapporto 2004 sull'economia del Mezzogiorno*, Svimez, Roma.

SVIMEZ (2005), *La coesione del Sud, macroregione "debole", con le aree "forti" dell'Italia e dell'Europa*, Quaderno Nuova Serie n. 5, Svimez, Roma.

DICHOTOMIZING CONTINUOUS PROGNOSTIC VARIABLES AND MINIMUM p -VALUE APPROACH: THE PROGNOSTIC ROLE OF AGE IN HODGKIN'S DISEASE (1)

Abstract. In this paper we reviewed methods to dichotomize a quantitative prognostic explanatory variable maximizing the difference in prognosis between low and high risk group respect to a response variable. We briefly reviewed the partial solutions to the methodological difficulties with this approach: high false positive rate in the prognostic assessment of explanatory variable (p -value inflation and over-estimation of difference between groups). Explanatory and graphical methods for categorization and methods separating best risk groups were used and the results of a study about the prognostic role of age (and the determination of its best cut-off) in Hodgkin's disease on GISL (Gruppo Italiano Studio Linfomi) database followed.

1. Minimum p -value approach

Several methods have been elaborated searching for criteria to categorize continuous variables minimizing the "loss of information" [Cox (1957), Connor (1972), Stefansky and Kaiser (1973), Johari and Sclove (1976)]. The maximally selected statistic, also known as the minimum p -value approach, is generally recognized for optimum dichotomization of an explanatory continuous variable when it is con-

(1) The working paper is the result of a joint work; however, S. De Cantis wrote sections 2, 3 and 4, A.M. Taormina wrote section 1. The authors thank GISL (Gruppo Italiano Studio Linfomi), in particular the President Prof. M. Federico for permission to use Hodgkin's disease dataset. Special thanks go to Emilio Iannitto (MD) and his workgroup for some basic ideas inside this paper and for his valuable suggestions and comments.

(2) Dipartimento di Metodi Quantitativi per le Scienze Umane, Viale delle Scienze, Edificio 13, Università degli Studi di Palermo, 90128 Palermo.

(3) Dipartimento di Scienze Statistiche e Matematiche "Silvio Vianelli", Viale delle Scienze, Edificio 13, Università degli Studi di Palermo, 90128 Palermo.

sidered in relation to a response variable. In this paper, we use the minimum p -value approach to dichotomize a quantitative prognostic explanatory variable maximizing the difference in prognosis between the low and the high risk group with respect to a response variable. There are many advantages in categorizing prognostic variables [for an interesting discussion, see Altman (1991)]. Methods used to categorize prognostic variables include: biological determination (most desirable but often unavailable), arbitrary selection of a cutpoint at the median value, graphical examination of the data for a threshold effect, systematic exploration of all observed values for the one which best separates the risk groups. Exploratory and graphical analyses may reveal obvious thresholds that suggest potential cutpoints: the analysis of the relationship between an explicative continuous (prognostic) variable (X) and an outcome (response variable, Y), that we point out as being of two kinds, dichotomous or survival (with the presence of censored observations), allows us to conjecture some typical patterns: a) an ideal situation in which the relationship (step function) between response and prognostic variable highlights one or more cutpoints; b) a situation in which the function linking response and prognostic variable is approximately monotone; as a result, no cutpoint can efficaciously discriminate between high and low risk patients; singling out one or more cutpoints leads to a not negligible and not easily valuable loss of information; c) a final situation in which the relationship appears consistently not monotone: dichotomization is absolutely not adequate and introduces heavy interpretative biases and serious consequences. In practice, for a binary outcome variable a scatterplot with the means of Y (proportions) for classes of X can be useful to reveal the kind of relationship between Y and X . When Y is a survival time (with censored observations) a simple scatterplot of observed times of the value of the prognostic variable X is not adequate. Kaplan-Meier median survival estimates can be plotted versus groups of X ; but this procedure requires that a sufficient proportion of non-censored times should be observed in each group. Alternatively, some authors [Heller and Simonoff (1992) and Mazumdar and Glassman (2000)] proposed to plot the predicted time at which a certain proportion of patients remain alive for each value of X . The predicted time can be derived from the well-known Cox survival time model.

Whether or not graphical examination of the data suggests an obvious cut-point, systematic (certainly useful and not replaceable) meth-

ods are also very useful in the cutpoint selection process. All observed values of the prognostic variable, except a proportion of the extremes, are examined as candidates for the cutpoint. The chosen value is one that best separates outcomes of patients in accordance with the maximally selected statistic and the corresponding minimum p -value.

In the case of a dichotomous response, the optimum cutpoint maximizes the association between outcome and dichotomized prognostic variable. Let Y be a dichotomous response and X be a continuous explanatory (prognostic) variable. Let x_i be a hypothetical cutpoint of X , where x_i assumes each observed value of X , excluding a $1 - (\epsilon_{low} + \epsilon_{high})$ proportion of values. For each x_i , form a 2×2 table of counts and calculate the corresponding chi-square statistic $X_{i, \epsilon_{low}, \epsilon_{high}}^2$. Optimum cutpoint, x_c , is the value of X that corresponds to: $X_{c, \epsilon_{low}, \epsilon_{high}}^2 = \max_i X_{i, \epsilon_{low}, \epsilon_{high}}^2$ or, specularly, $p_{emp} = p_{c, \epsilon_{low}, \epsilon_{high}} = \min_i Pr\{\chi_1^2 > X_{i, \epsilon_{low}, \epsilon_{high}}^2\}$ where χ_1^2 is the well-known chi-square distribution with one degree of freedom.

In the case of a survival response variable (with censored observations), for each potential cutpoint, a logrank statistic is computed in order to measure the strength of the association between the dichotomized prognostic variable and the survival endpoint. Let the standardized logrank test statistic be $SLR_{i, \epsilon_{low}, \epsilon_{high}}$ as a function of x_i . Optimum cutpoint x_c is the value of X that corresponds to $SLR_{c, \epsilon_{low}, \epsilon_{high}} = \max_i SLR_{i, \epsilon_{low}, \epsilon_{high}}$ or, specularly, $p_{emp} = p_{c, \epsilon_{low}, \epsilon_{high}} = \min_i Pr\{\chi_1^2 > X_{i, \epsilon_{low}, \epsilon_{high}}^2\}$ where χ_1^2 is the well-known chi-square distribution with one degree of freedom. For details and a review, see Mazumdar and Glassman (2000); for further methodological developments, see Lausen et al. (2002), Hothorn and Lausen (2003).

Several authors have addressed some of the methodological pitfalls with the (maximally selected statistic or) minimum p -value approach. The minimum p -value approach involves multiple testing which inflates the type I error rate: minimum p -value (in the following sections, p_{emp}) is found by performing several not independent tests on the same dataset. Several methods for adjusting the inflated p -value have been suggested. In this framework, what is the true upper percentile point of $X_{c, \epsilon_{low}, \epsilon_{high}}^2$ and with what probability does $X_{c, \epsilon_{low}, \epsilon_{high}}^2$ exceed the upper percentile point of the chi-square distribution with one degree of freedom? For large samples approximate answers to these questions are given by Miller and Siegmund (1982). Referring to the setting of a binary outcome variable, they derived a p -value adjustment formula

based on the asymptotic distribution for a maximum chi-square statistic: $p_{MS,82}$ [Miller and Siegmund (1982); p. 1013, Eq. (8)]. Since there are no guidelines for the choice of ε_{low} and ε_{high} , we accept the suggest of Altman et al. (1994), who eliminate the top and bottom 5% or 10% of the extreme values in the data, providing the following simplifications for $p_{MS,82}$. For $\varepsilon_{low} = \varepsilon_{high} = 10\%$, $p_{A. et al., 94} = -1.63 p_{emp} [1 + 2.35 \ln(p_{emp})]$. This approximation is nearly identical to $p_{MS,82}$ for small p_{emp} ($.0001 < p_{emp} < .1$).

The p -value correction question was also considered by Lausen and Schumacher (1996). They derived a modified version of the standard Bonferroni correction (p_{BON}), which works well only if the consecutive test statistics calculated are independent (which is not the case for the minimum p -value cutpoint search), and proposed $p_{LS,96}$ [Lausen and Schumacher (1996), p. 309, Eq. (4)]. Lausen and Schumacher (1992) showed that under the assumption of a random censorship model the maximally selected logrank statistic has the same asymptotic distribution as the maximally selected statistics above. Consequently, $p_{MS,82}$ and $p_{LS,96}$ also apply when the outcome variable is censored (Lausen and Schumacher, 1996). The complete logic algorithm and the program (in SPSS) required for calculations are provided by the authors. Appropriate libraries in R and S-plus were also elaborated [Hothorn and Lausen (2002); Mazumdar and Glassman (2000)].

Formulas for p -value correction do not allow a relative risk unbiased estimation. The following two approaches use data-split techniques to avoid inflation both of p -value and of relative risk. In the two-fold cross-validation approach the data are randomly split in two subset of approximately equal size: for each subset the minimum p -value approach is used to find two “optimal” cutpoints (x_{c1} and x_{c2}), according to which each subset is divided into two groups, high and low risk group [for details, see Faraggi and Simon (1996)]. The split-sample approach (Hilsenbeck et al., 1992) divides the entire dataset into two subsets called the training set and the validation set. An optimal cutpoint is found by using a maximally selected statistic search on the training set. This cutpoint is used to categorize the prognostic variable in the validation set; in line with this dichotomization a p -value and a relative risk estimation are obtained. These estimates are expected to be unbiased because the cutpoint used to categorize each observation in the validation set has been optimally selected from the training set. Mazumdar et al. (2003) compared two approaches and established that the two-fold

cross-validation method performs better than the split-sample method in their simulation study.

2. **Prognostic role of age in Hodgkin's disease: preliminary analysis**

The results of the treatment of Hodgkin's disease (HD) have improved considerably over the last decades and in this improvement studies of prognostic factors have played a very important role. Several studies evaluated the prognostic (single or joint) significance of most important clinical and biological indicators and proposed prognostic models to explain individual variability characterizing HD [e.g. see Gobbi et al. (1994), Fermé et al. (1997), Lee et al. (1997), Hasenclever and Diehl (1998) and Gobbi et al. (2001)]. Nevertheless, studies vary with regard to the factors found to be significant, and with regard to the relative importance attributed to these factors [for two interesting reviews, see Specht (1991) and Specht and Hasenclever (1999)]. A small number of studies specifically considered the analysis and the prognostic assessment of "age at diagnosis" (henceforth simply "age") even if it is highly associated with the outcomes of the disease in most investigations.

The empirical analysis to investigate the prognostic role of age in HD and to explore the impact factor of dichotomized age was performed on a series of 494 HD patients, who were collected and studied by GISL (Gruppo Italiano Studio Linfomi). These patients were enrolled between June 1987 and July 1997. Median follow-up was approximately 55 months, while 75% of the patients had follow-up of over 34 months. Enrolment criteria, clinical presentation features, staging procedures, and treatment were described elsewhere [Gobbi et al. (1993) and Gobbi et al. (1998)]. We considered different outcomes (endpoints): a simple dichotomous outcome variable could be "response to therapy" (yes or no) or "survival" (surviving or deceased).

However the vast majority of HD patients respond to therapy (and fortunately the survival rate is high too), so a response to therapy is not an effective indicator of cure or long-term survival; so that in addition to the overall mortality rate (OMR) and the treatment failure rate (TFR), the overall survival (OS, time from the date of diagnosis to the date of last observation or death) and time to treatment failure (TTF, time from

the start of treatment to one of the following events: disease progression during treatment, no complete remission (CR) at the end of treatment, disease recurrence or death from disease) would be important outcome variables and, consequently, taken into account. Deaths unrelated to the disease or treatment effects were censored for OS calculations; deaths due to causes other than HD were censored for TTF calculations. Hodgkin's disease affects every age class but the literature shows that between 15 and 30 years it presents maximal frequency. This holds in our series too. In order to check the monotonicity of the relation between age and outcomes in both cases (dichotomous or continuous survival response) we plotted in Fig.1 the OMRs and the TFRs versus 5-year age and the expected survival time (in months) at which a certain proportion of patients remain alive (95%) and remain treatment failure free (90%) as a function of the 5-year age values. Scatter-

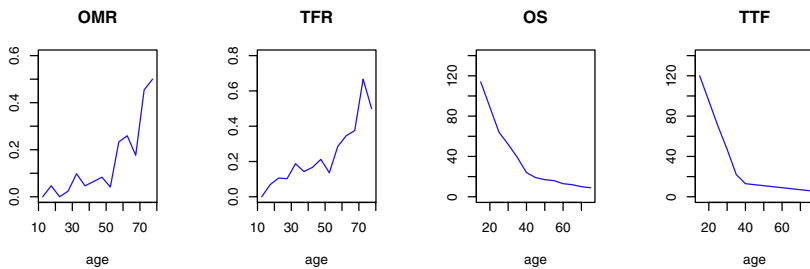


Fig. 1 - OMRs and TFRs for 5-year age, and the expected overall survival time (OS, in months) at which 95% of the patients survived, and the expected time to treatment failure (TTF, in months) at which 90% of patients remained treatment failure free, as a function of 5-year age values.

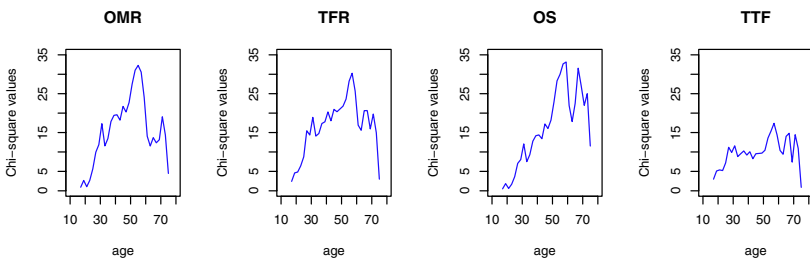


Fig. 2 - Chi-squared values measuring the association between outcomes (OMR, TFR, OS, TTF) and dichotomized age at diagnosis plotted as a function of the cupoint age values.

plots highlight a non-ideal situation because the underlying relationship between outcomes and age seems not to be a step function and therefore no cutpoint clearly divides patients into high and low risk groups; nevertheless, the necessary condition to use a single cutpoint model seems to be satisfied. Finally, results seem to be homogeneous for all outcomes: older patients show very worse expected responses (Fig. 1).

3. Optimum cutpoint search, p -value correction and relative risk estimation using data-split techniques

Fig. 2 reports chi-square values measuring the association between outcomes (OMR, TFR, OS, TTF) and dichotomized age at diagnosis, plotted as a function of the cutpoint age values. It can be seen that the optimum cutpoint interval was restricted to 55-59 years (depending on different outcomes); the optimum value was thus far from the cutpoint (45 years) found in the literature [Hasenclever and Diehl (1998), Gobbi et al. (2001), Specht and Hasenclever (1999)]. Finally, for each outcome, Tab. 1 reports: corresponding optimum cutpoints and relative risk estimates; chi-square values and corresponding p -values; and the p -value corrections proposed by various authors (see section 1). In relation to these results we choose 57 years as the “optimum” cutpoint. Cumulative overall survival rates at 7 years for patients over and under 57 years are 94.1% and 68.9%, respectively; the proportions of treatment failure free patients at 7 years are 87.6% and 66.3%, respectively.

Table 1 - Optimum cutpoints (minimum p -value approach) for age at diagnosis and corresponding relative risk estimates (Odds Ratio and Hazard Rate); chi-square values ($df=1$) and corresponding p -values; p -value corrections proposed by various authors.

	OMR	TFR	OS	TTF
Optimum cutpoints	55	57	59	57
Rel. risk estimates	OR = 6.72	OR = 4.35	HR = 5.26	HR = 3.12
Chi-squared values	32.38	30.27	33.15	17.38
p_{emp}	<.00001	<.00001	<.00001	<.00001
$p_{MS,82}$	//	//	<.0001	<.0020
$p_{A, et al., 94}$	//	//	<.0001	<.0030
p_{BON}	//	//	<.0001	<.0020
$p_{LS,96}$	//	//	<.0001	<.0004

Table 2 - Relative risk estimates (Odds Ratio and Hazard Rate) of age at diagnosis for different endpoints in line with the two-fold cross validation approach and to the split-sample approach (and corresponding p -values).

	Cross-validation approach				Split-sample approach			
	OMR	TFR	OS	TTF	OMR	TFR	OS	TTF
Rel. Risk estimates	OR = 6.37	OR = 4.07	HR = 5.27	HR = 2.79	OR = 7.78	OR = 5.51	HR = 2.58	HR = 3.15
p -value	< .00001	< .00001	< .00001	0.00068	< .00001	< .00001	0,031	0.014

In line with the two-fold cross-validation approach (Faraggi and Simon, 1996) and the split-sample approach (Hilsenbeck et al., 1992) described in section 1, in Tab. 2 we report relative risk estimates and corresponding p -values for four different outcomes. Although the two groups are randomly chosen, “optimum” cutpoint and corresponding relative risk estimates are substantially the same for the two subsets. Similar results are obtained in the validation set in line with the split-sample approach.

4. Conclusions

Older age has frequently been associated with poor survival in studies of prognostic factors in HD. This result was confirmed by our analysis. In the absence of clear criteria of biological determination of a cutpoint, exploratory and graphical analysis revealed a monotonic relationship between the prognostic role of age and the outcomes. Categorization (dichotomization) of age can thus be examined; categorization is often necessary for clinical decision-making (determining criteria for study eligibility, patient stratification and therapeutic assignment; control for varying patient prognoses in the design of randomized and controlled clinical trials). The statistical methods applied in this paper followed an approach based on the intensive analysis of data. The results revealed statistically significant optimum cutpoints (also using p -value corrections for multiple testing) that are sufficiently homogeneous (for different outcomes) and very different from that found in the literature and commonly applied in clinical practice. Nevertheless, we had a not immediately evaluable loss of information when we used univariate cutpoint models where there was a continuous relationship between prognostic covariate values and relative risk. The data-based di-

chotomization of an inherently continuous biological relationship is therefore problematic: apart from losing information, this procedure produces a biologically unrealistic model where the hazard (risk) has a sudden jump at the cut-off level, with all values below the cut-off (Altman, 1991). On the other hand, dichotomization will continue as an important part of the prognostic evaluation, partly for the very practical reason that simple categorical prognostic factors are easier to interpret and integrate into clinical practice. In conclusion, it has still not been established from a clinical point of view whether HD in itself behaves more aggressively in older than in younger patients, even when other well-known prognostic factors are taken into account. The question is crucial when deciding whether or not to allow age *per se* to influence choice of treatment in individual patients. If the poorer prognosis in older patients is not caused by a more aggressive disease but rather by inability of some older patients to tolerate intensive treatment, further intensification of treatment in older patients will only compromise their chance of survival. Further research on this complex issue is clearly needed.

BIBLIOGRAPHY

- ALTMAN D.G. (1991), *Categorizing continuous variables*. British Journal of Cancer, 64, 975.
- ALTMAN D.G., LAUSEN B., SAUERBREI W., SCHUMACHER M. (1994), *Dangers of using "optimal" cutpoints in the evaluation of prognostic factors*. Journal of the National Cancer Institute, 86, 11, 829-835.
- CONNOR R.J. (1972), *Grouping for Testing Trends in Categorical Data*. J.A.S.A., 67, 601-604.
- COX D.R. (1957), *Note on Grouping*. J.A.S.A., 52, 543-547.
- FARAGGI D., SIMON R. (1996), *A simulation study of cross-validation for selecting an optimal cutpoint in univariate survival analysis*. Statistics in Medicine, 15, 2203-2213.
- FERMÉ C., BASTION Y., BRICE P., LEDERLIN P., DIVIN M., GABARRE J., ASSOULINE D., FERRANTAND A., BERGER F., LEPAGE E. (1997), *Prognosis of patients with advanced Hodgkin's disease. Evaluation of four prognostic models using 344 patients included in the Group d'Etudes des Lymphomes de l'Adulte Study*. American Cancer Society, 80.
- GOBBI P.G., PIERESCA C., FEDERICO M., RENZO N.D., NARNI F., IANNITTO E. ET AL. (1993), *MOPP/EBV/CAD hybrid chemotherapy with or without limited radiotherapy in advanced or unfavorably presenting Hodgkin's Disease: report from the Italian Lymphoma Study Group*. Journal of Clinical Oncology, 11, 712-719.
- GOBBI P.G., COMELLI M., GRIGNANI G., PIERESCA C., BERTOLONI D., ASCARI E. (1994), *Estimate of expected survival at diagnosis in Hodgkin's disease: a mean of weighting prognostic factors and a tool for treatment choice and clinical research. A report from the International Database on Hodgkin's Disease (IDHD)*. Haematologica, 79, 241-255.
- GOBBI P.G., PIERESCA C., GHIRARDELLI M.L., FEDERICO M., RENZO N.D., ET AL. (1998), *Long-term results from MOPP/EBV/CAD chemotherapy with optional limited radiotherapy in advanced Hodgkin's Disease*. Blood, 91, 2704-2712.
- GOBBI P.G., ZINZANI P.G., BROGLIA C., COMELLI M., MAGAGNOLI M., FEDERICO M., MERLI F., IANNITTO E., TURA S., ASCARI E. (2001), *Comparison of Prognostic Models in Patients with Advanced Hodgkin Disease*. Cancer, 91/8, 1467-1478.
- HASENCLEVER D., DIEHL V. (1998), *A prognostic score for advanced hodgkin's disease*. The New England Journal of Medicine, 19, 1506-1514.
- HELLER G., SIMONOFF J.S. (1992), *Prediction in censored survival data: a comparison of the proportional hazards and linear regression models*. Biometrics, 48, 111-115.

-
- HILSENBECK S., CLARK G.M., MCGUIRE W.L. (1992), *Why do so many prognostic factors fail to pan out?* Breast Cancer Research Treatment, 22, 197-206.
- HOTHORN T., LAUSEN B. (2002), *On maximally selected rank statistics.* R news, 2, 3-5.
- HOTHORN T., LAUSEN B. (2003), *On exact distribution of maximally selected statistics.* Computational Statistics & Data Analysis, 43, 121-137.
- JOHARI S., SCLOVE S.L. (1976), *Partitioning a Distribution.* Communications in Statistics - Theory and Methods, 5, 133-147.
- LAUSEN B., SCHUMACHER M. (1992), *Maximally selected rank statistics.* Biometrics, 48, 73-85.
- LAUSEN B., SCHUMACHER M. (1996), *Evaluating the effect of optimized cutoff values in the assessment of prognostic factors.* Computational Statistics & Data Analysis, 21, 307-326.
- LAUSEN B., LERCHE R., SCHUMACHER M. (2002). *Maximally selected rank statistics for dose-response problems.* Biometrical Journal, 44, 131-147.
- LEE S., RADFORD J., RYDER W., COLLINS C., DEAKIN D., CROWTHER D. (1997), *Prognostic factors for disease progression in advanced Hodgkin's disease: an analysis of patients aged under 60 years showing no progression in the first 6 months after starting primary chemotherapy.* British Journal of Cancer, 75 (1), 110-115.
- MAZUMDAR M., GLASSMAN J.R. (2000), *Categorizing a prognostic variable: review of methods, code for easy implementation and applications to decision-making about cancer treatments.* Statistics in Medicine, 19, 113-132.
- MAZUMDAR M., SMITH A., BACIK J. (2003), *Methods for categorizing a prognostic variable in a multivariable setting.* Statistics in Medicine, 22/4, 559-571.
- MILLER R., SIEGMUND D. (1982), *Maximally selected chi square statistics.* Biometrics, 38, 1011-1016.
- SPECHT L. (1991), *Prognostic factors in Hodgkin's disease.* Cancer Treatment Reviews, 18, 21-53.
- SPECHT L., HASENCLEVER D. (1999), *Prognostic factors of Hodgkin's disease.* In Mauch P.M., Armitage J.O., Diehl V. et al. (eds.) Hodgkin's disease. Lippincot Williams and Wilkins, Philadelphia.
- STEFANSKY W., KAISER H.F. (1973), *Note on Discrete Approximations.* J.A.S.A., 68, 232-234.

FRANCESCO GARGANO (*)

**UNA STIMA STATISTICA,
EFFETTUATA MEDIANTE I METODI
DI RICAMPIONAMENTO
JACKKNIFE E BOOTSTRAP,
DEI CONFINI DELLA NORMALITÀ
O REGOLARITÀ DELLE OSCILLAZIONI
DELLE FREQUENZE DELLE MISURE
DI UN FENOMENO QUANTITATIVO
COLLEGABILE AD UN FENOMENO
QUALITATIVO**

1. Premesse

A margine dell'argomento già svolto (Gargano F. 1999) in cui è stato proposto il criterio statistico per confermare l'omogeneità di determinate unità statistiche (1) e, quindi, dei soggetti o degli oggetti portatori di un fenomeno qualitativo detto anche descrittivo o collettivo, ed il criterio per determinare i confini convenzionali della regolarità o normalità delle oscillazioni delle frequenze delle misure di un carattere metrico collegabile alla qualità (2) dei soggetti o degli oggetti medesimi, seguono alcune considerazioni necessarie per chiarire il significato dei termini usati e per riferire l'esito di un tentativo di stima – effettuata

(*) Professore associato presso l'Università degli Studi di Palermo. Facoltà di Economia. Dipartimento di metodi quantitativi per le scienze umane. Viale delle Scienze. 90128 Palermo. Telefono: 091-6626286. Fax: 091-426452; e-mail: gargano@economia.unipa.it.

(1) Ossia, determinati individui od unità materiali esistenti in realtà che risultano omogenei in base alla diagnosi di una loro qualità e costituiscono, pertanto, un immaginabile insieme finito, od una classe, di soggetti od oggetti, anziché un immaginario insieme infinito di campioni casuali di soggetti o di oggetti caratterizzati da un'immaginaria serie, tendente ad infinito, di probabili fenomeni accidentali od eventi.

(2) Come, ad es., lo stato di salute del “non diabete” o della malattia diabetica collegabile, secondo la logica consueta, a determinate misure di glicemia dei soggetti di una popolazione umana osservata in un determinato tempo e spazio, oppure la condizione della “ricchezza” o della “povertà” collegabile a determinate misure del reddito procapite delle persone fisiche o giuridiche della predetta popolazione.

mediante i metodi di ricampionamento Jackknife e Bootstrap – dei confini (3) di “normalità statistica” dei casi o, per meglio dire, dei soggetti o degli oggetti scelti in base alla diagnosi di un prestabilito fenomeno qualitativo.

Come è noto, gli anzidetti metodi di ricampionamento sono fondati sull’ipotesi di una relazione di simmetria tra due variabili (una dipendente e l’altra indipendente) e, quindi, resuppongono l’uso di un modello probabilistico, avente un teorico ed immaginario intervallo aperto, per rappresentare schematicamente la variabilità imputabile esclusivamente ad una immaginaria serie, tendente ad infinito, di cause imprevedibili cosiddette accidentali. Pertanto, ciò esclude l’esistenza di un nesso causale tra le predette variabili.

In seguito, sarà chiarito meglio l’esito negativo del tentativo di stima dei confini di “normalità statistica”. Infatti, gli anzidetti metodi di ricampionamento non presuppongono di censire e classificare, secondo prestabilite qualità distintive, la popolazione complessiva dei soggetti o degli oggetti di riferimento.

S’intuisce, inoltre, che la classificazione di un insieme finito di soggetti o di oggetti eterogenei, rispetto a prestabiliti fenomeni qualitativi, può evitare d’ipotizzare il fenomeno della cosiddetta “transvariazione” (Gini C. 1916) o l’accavallamento delle teoriche misure distanti dalle teoriche medie aritmetiche nascoste dall’unica media aritmetica della distribuzione teorica di densità di probabilità, data a priori, della variabile continua casuale gaussiana semplice standardizzata, da cui può scaturire qualsiasi distribuzione empirica di frequenza della variabile medesima.

Gli anzidetti metodi di ricampionamento, dunque, non sono applicabili ad un insieme finito, o ad una classe, di soggetti o di oggetti omogenei distribuiti secondo le misure di un carattere metrico, poiché, in tal caso, si creano le condizioni per ottenere una distribuzione empi-

(3) Onde separare, per motivi di cautela, dalle code della distribuzione empirica unimodale (asimmetrica) di frequenza della cosiddetta variabile potenzialmente continua “non casuale” o statistica semplice, scaturita dalla valutazione di un fenomeno qualitativo in base all’esperienza e con l’ausilio della misura di un fenomeno quantitativo collegabile a quello qualitativo, una prestabilita percentuale (95%) dell’insieme finito, o della classe, dei soggetti o degli oggetti omogenei, rispetto al predetto fenomeno qualitativo, ai quali, per convenzione, si possono associare talune immaginabili misure del fenomeno quantitativo perturbate probabilmente da immaginabili errori casuali non molto distanti dall’errore nullo (misura esatta, o immaginaria media aritmetica degli errori casuali standardizzati delle teoriche misure del fenomeno quantitativo considerato, per lo schema di un’ipotesi, una grandezza astratta o matematica).

rica univariata unimodale (asimmetrica), anziché plurimodale, di frequenza delle misure del fenomeno quantitativo, caratterizzata da variabilità, rispetto alla moda, in realtà stabile a causa della qualità dei soggetti o degli oggetti che limita l'ordine delle misure del fenomeno quantitativo. La forma asimmetrica e con una cuspidè della predetta distribuzione empirica può essere, quindi, nota dopo la valutazione, come detto prima, del fenomeno qualitativo collegabile a quello quantitativo. Perciò, essa è citata in parentesi.

Una qualità provocata dalla natura o prodotta artificialmente dall'uomo può essere in realtà più o meno frequente, o addirittura rara, e, talvolta, può essere valutata in base all'esperienza (4) e con l'ausilio della misura di un carattere metrico (informazione statistica) collegabile, secondo la logica consueta, alla qualità medesima. Pertanto, l'entità dei soggetti o degli oggetti omogenei, rispetto ad un fenomeno qualitativo, non è mai l'infinito matematico ma è sempre un numero più o meno grande che può stabilire un esperto in grado di valutare, come detto prima, il fenomeno qualitativo non misurabile.

La variabilità, rispetto alla moda (5), di un carattere metrico collegabile ad un fenomeno qualitativo è imputabile, quindi, ad una prevedibile causa naturale od artificiale determinante, detta anche causa sistematica, regolare o normale, che caratterizza determinati soggetti od oggetti (omogenei). Essa, però, non può essere separata dalla variabilità imputabile ad un imprecisabile numero di cause imprevedibili, dette cause accidentali od irregolari (6), che, in concreto, si traducono in un altrettanto imprecisabile numero di errori casuali di misura, da

(4) Ossia, la conoscenza del fenomeno qualitativo che scaturisce da un'immaginabile empirica serie (finita) di osservazioni su determinati soggetti od oggetti caratterizzati, appunto, dalla predetta qualità, anziché in base all'ipotesi che il fenomeno qualitativo sia un fenomeno accidentale, per cui bisogna effettuare la prova per verificare se esso sia uno di un'immaginaria (teorica) serie, tendente ad infinito, di fenomeni accidentali od eventi.

(5) Detta anche "standard", misura prevalente, regola o norma, di un insieme limitato chiuso da due misure estreme, o di soglia, di un fenomeno quantitativo collegabile ad un fenomeno qualitativo, prestabilito da una comunità nazionale od internazionale di esperti del fenomeno qualitativo per essere collegate a taluni di un insieme finito, o di una classe, di soggetti o di oggetti omogenei, rispetto alla qualità valutabile in base all'esperienza e con l'ausilio della misura del predetto fenomeno quantitativo. Perciò, in tal caso, è logico ragguagliare all'unità l'insieme finito, o la classe, dei soggetti o degli oggetti in questione per significare che essi sono accumulati da un'unica qualità. Quindi, lo "standard" o la moda prepuppone una sola qualità dei soggetti o degli oggetti ai quali essa è da riferire.

(6) Obiettivamente è difficile o, addirittura, impossibile separare un imprecisabile numero di cause accidentali da una causa sistematica determinante. Tuttavia, l'influenza di una causa sistematica è quella che deve essere ricercata in ogni caso.

supporre a due a due uguali e di segno contrapposto (negativo e positivo) ma non d'illimitata intensità, che possono essere commessi dall'esperto che valuta, come detto prima, il fenomeno qualitativo.

Perciò, si ribadisce che due indici di variabilità assoluta (Gargano F. 1997), inerenti un fenomeno quantitativo collegabile ad un fenomeno qualitativo o la cosiddetta variabile "non casuale" o statistica semplice, detta anche variabile causale, "scarto dalla moda" oppure variabile "di partenza", potenzialmente continua in intervallo limitato chiuso, o compatto, possono consentire il calcolo di un rapporto utile per:

a) Confermare la lunghezza o misura, in realtà stabile, dell'intervallo limitato chiuso dalle misure estreme, o di soglia, di un fenomeno quantitativo collegabile ad un fenomeno qualitativo che limita, in realtà, l'ordine delle misure di quello quantitativo e, quindi, la lunghezza dell'intervallo della variabile in questione. Cioè, il criterio statistico per confermare l'omogeneità di determinati soggetti od oggetti che costituiscono un insieme finito, od una classe, la cui entità è sempre un numero che può essere stabilito in base all'esperienza.

b) Scegliere, nell'ambito della cosiddetta popolazione statistica o dell'insieme limitato chiuso dalle misure estreme, o di soglia, del fenomeno quantitativo collegabile a quello qualitativo, le misure del carattere metrico della maggior parte dei soggetti o degli oggetti (omogenei) perturbate probabilmente dagli errori casuali non molto distanti dall'errore nullo. Cioè, il criterio statistico per scegliere due misure del fenomeno quantitativo, collegabile a quello qualitativo, poste minori od uguali rispetto ad esse stesse, cui corrispondono due percentili empirici, a cavallo ma non necessariamente equidistanti dalla moda, necessarie per calcolare la lunghezza di un intervallo per così dire parziale all'interno dell'anzidetto intervallo limitato chiuso, in cui comprendere, in realtà, una prestabilita quota percentuale (95%) del totale dei soggetti o degli oggetti omogenei concentrati in prossimità del massimo della frequenza cui corrisponde la moda della distribuzione empirica unimodale (asimmetrica) di frequenza delle misure del fenomeno quantitativo. Ciò è utile per effettuare il calcolo della lunghezza di un "range interpercentile" della distribuzione empirica di frequenza cumulata dell'anzidetta variabile potenzialmente continua "non casuale" o statistica semplice.

Quindi, per scegliere due misure di un fenomeno quantitativo, collegabile ad un fenomeno qualitativo, poste minori od uguali rispetto ad esse stesse, cui corrispondono due percentili empirici, per stabilire la

lunghezza di un intervallo per così dire parziale dell'insieme compatto dell'anzidetta variabile (7), occorre fare riferimento alla distribuzione empirica unimodale (asimmetrica) di frequenza della variabile in questione, priva di distribuzione teorica di densità di probabilità, ed alla corrispondente distribuzione empirica unimodale di frequenza cumulata.

2. Obiettivi specifici

S'intende dimostrare che il rapporto tra due indici di variabilità assoluta già proposto [formula (2)] può essere utile per l'impiego di un criterio per:

a) Confermare l'omogeneità, rispetto ad un fenomeno qualitativo di determinati soggetti od oggetti che costituiscono un insieme finito, od una classe, la cui entità è sempre un numero che può stabilire un esperto del fenomeno qualitativo.

b) Calcolare la lunghezza di un "range" interpercentile della distribuzione empirica unimodale (asimmetrica) di frequenza cumulata della cosiddetta variabile potenzialmente continua statistica semplice, per scegliere le estreme frequenze regolari o normali, corrispondenti a due misure, a cavallo ma non necessariamente equidistanti dalla moda, di un carattere metrico collegabile, secondo la logica consueta, ad una qualità di determinati soggetti od oggetti (omogenei) che costituiscono un insieme finito, od una classe. Cioè, i cosiddetti confini di "normalità statistica" dei casi enumerati in occasione di una diagnosi, effettuata come detto prima, di un fenomeno qualitativo.

Il concetto di "normalità statistica" così espresso non allude, quindi, alla gaussianità (8) o casualità delle oscillazioni delle frequenze delle misure di un carattere metrico associabile, senza la ricerca di un'effettiva corrispondenza logica, ad un fenomeno qualitativo supposto un fenomeno casuale. Significa, invece, la regolarità delle oscillazioni del-

(7) Che ha per modalità le misure distanti dalla moda della distribuzione empirica unimodale (asimmetrica) di frequenza della variabile medesima potenzialmente continua in un intervallo limitato chiuso.

(8) Come è noto, la curva della distribuzione teorica di probabilità degli errori casuali di misura è stata chiamata da Gauss, nel 1883, curva normale. Il Laplace (1749-1827) per primo e, poi, il Gauss (1777-1885) introdussero la distribuzione di probabilità degli errori casuali di misura che Gauss chiamò "legge normale degli errori". In seguito, essa è stata impiegata nel contesto dell'analisi degli errori di misura in Astronomia, in Fisica ed in Geodesia (Johnson R. A. Bhat-tacharya G.K. 1992., pp. 236-237. Oppure consulta: Lombardo E. 1993., pag. 343).

le frequenze intorno al loro massimo, cui corrisponde la moda del carattere metrico collegabile, secondo la logica consueta, ad un fenomeno qualitativo di determinati soggetti od oggetti che, essendo omogenei, rispetto al predetto fenomeno qualitativo, sono da ammassare o cumulare e costituiscono un insieme finito od una classe. Perciò, soltanto i fenomeni qualitativi sono da considerare fenomeni di massa.

Per convenzione, dunque, è da considerare regolare o normale la somma integrale di livello percentuale 95% delle frequenze della distribuzione empirica unimodale (asimmetrica) dell'anzidetta variabile "non casuale" o statistica semplice standardizzata potenzialmente continua in un intervallo limitato chiuso.

Bisogna impiegare, perciò, l'indice di variabilità assoluta per così dire totale imputabile ad una causa sistematica determinante difficilmente separabile da un imprecisabile numero di cause accidentali, per standardizzare, per scopi comparativi in un determinato tempo e spazi diversi, oppure in un determinato spazio e tempi diversi, la predetta variabile, espressa con $\bar{\varepsilon} = \bar{x} - m_{\bar{w}}$, e la lunghezza in realtà stabile, espressa con $\bar{w} = \bar{x}_N - \bar{x}_1$, del suo intervallo limitato chiuso, o compatto, dalle misure estreme, o di soglia, da prestabilire in base all'esperienza, oppure per standardizzare una parte dell'intervallo medesimo. Tale indice di variabilità è utile, dunque, per l'esame della reale stabilità della variabilità, rispetto alla moda, di un fenomeno quantitativo collegabile ad un fenomeno qualitativo.

Il criterio proposto per standardizzare l'immaginabile lunghezza dell'intervallo limitato chiuso della variabile in questione, detto anche intervallo totale di rilevazione, e per standardizzare un "range" interpercentile a cavallo della moda del predetto intervallo limitato chiuso è diverso – dunque – dal criterio per standardizzare l'immaginaria lunghezza dell'intervallo aperto, detto campo, della variabile continua casuale gaussiana semplice, oppure una parte di codesto intervallo aperto.

L'uso dell'indice approssimato della deviazione dalla moda della distribuzione empirica unimodale (asimmetrica) di frequenza della cosiddetta variabile potenzialmente continua "non casuale" o statistica semplice consente, quindi:

I) Standardizzare la variabile in questione e calcolare gli estremi (inferiore e superiore) della cosiddetta regione di tolleranza degli errori casuali semplici che probabilmente perturbano le misure dell'empirica distribuzione unimodale (asimmetrica) di frequenza cumulata dell'anzidetta variabile "non casuale".

Infatti, gli estremi errori casuali semplici che possono essere commessi da chi valuta un fenomeno qualitativo in base all'esperienza e con l'ausilio della misura di un fenomeno quantitativo non possono essere in realtà maggiori, ma minori od uguali, alle misure estreme prestabilite per segnalare il cosiddetto passaggio del grado di qualità dei soggetti o degli oggetti esaminati. Nel caso contrario, infatti, si tratta di misure anomale collegabili ad un fenomeno qualitativo diverso da quello valutato, cioè misure esterne a quelle estreme del carattere metrico dei soggetti o degli oggetti omogenei di un insieme finito, o di una classe.

II) Calcolare, all'interno della predetta regione di tolleranza, gli estremi di un intervallo (di tolleranza) di errori casuali semplici per così dire tollerabili che probabilmente perturbano la maggior parte delle misure distanti dalla moda standardizzate, rispetto all'indice approssimato della deviazione dalla moda, del carattere metrico dei soggetti o degli oggetti omogenei, rispetto al fenomeno qualitativo. Per convenzione, tali errori possono essere considerati quelli cui corrisponde un prestabilito livello percentuale (95%) della somma integrale o cumulo delle frequenze dell'empirica distribuzione unimodale (asimmetrica) delle misure distanti dalla moda del fenomeno quantitativo collegabile a quello qualitativo standardizzate come detto prima.

3. Metodi

I) *Il criterio statistico per confermare l'omogeneità di determinati soggetti od oggetti selezionati in base alla diagnosi di un fenomeno qualitativo, detto anche fenomeno descrittivo o collettivo, effettuata in base all'esperienza e con l'ausilio della misura di un fenomeno quantitativo collegabile, secondo la logica consueta, a quello qualitativo. Cioè, il criterio per confermare l'immaginabile lunghezza in realtà stabile, espressa con $\bar{w} = \bar{x}_N - \bar{x}_1$ dell'intervallo limitato chiuso, o compatto (9), della cosiddetta variabile "non casuale" o statistica semplice, detta anche variabile causale, scarto dalla moda oppure variabile "di partenza", espressa con $\bar{\epsilon} = \bar{x} - m_{\bar{w}}$, potenzialmente continua.*

(9) La lunghezza di tale intervallo sarà in realtà stabile quando esso sarà limitato da due misure estreme, o di soglia, del carattere metrico prestabilite da una comunità di esperti per aiutare la valutazione del fenomeno qualitativo o descrittivo.

Per quanto detto al punto I) si ritiene utile, come termine di paragone, l'immaginaria lunghezza approssimata dell'intervallo aperto degli errori casuali standardizzati che probabilmente perturbano le teoriche misure di un fenomeno quantitativo considerato, per lo schema di un'ipotesi, una grandezza astratta X di un immaginario insieme infinito di campioni casuali di soggetti o di oggetti. Tale intervallo risulta di lunghezza approssimata, anziché esatta, perché limitato da due errori casuali standardizzati estremi considerati noti, perché desumibili dalla tavola prontuario ottenuta da E.S. Pearson e da H. O. Hartely con l'integrale improprio proposto da Sheppard ("Biometrika". 1954. Oppure consulta: Vianelli S. 1959. *Prontuario XXIX*, pag. 136). Cioè, l'immaginaria lunghezza approssimata dell'intervallo aperto, di cui non se ne conosce la lunghezza esatta, della variabile continua casuale gaussiana

standardizzata $k = \frac{X - M_X}{\sigma_X}$ che è l'argomento della funzione di ripar-

tizione, o somma integrale, delle densità delle probabilità degli errori casuali standardizzati:

$$W \cong \frac{(X_N - M_X) - (X_1 - M_X)}{\sigma_X} \cong k_N - k_1. \quad (1)$$

La lunghezza approssimata dell'anzidetto intervallo aperto può essere, quindi, $W \cong 6,18$; oppure $W \cong 7,98$; od - infine - $W \cong 12$; a seconda che concorrano al calcolo della somma integrale o cumulo delle densità delle probabilità degli errori casuali standardizzati, immaginati minori od uguali, rispetto ad essi stessi, associabili alle misure del carattere metrico dei soggetti o degli oggetti caratterizzati da una qualità valutabile come detto prima, supposta un fenomeno casuale, rispettivamente, non più di 100, più di 100 ma non più di 1.000, oppure più di 1.000 soggetti od oggetti che se sono in realtà omogenei, rispetto alla qualità valutata, costituiscono un insieme finito, od una classe (10).

(10) Abbiamo considerato, perciò, l'immaginaria lunghezza approssimata [formula (1)] dell'intervallo aperto che corrisponde alla probabilità integrale, rispettivamente, di livello 99,99%, oppure di livello 999,99%^o od, infine, di livello 9999,99%^{oo} della distribuzione teorica delle densità delle probabilità cumulate degli errori casuali standardizzati, desumibile dalla citata tavola prontuario.

Il rapporto tra gli anzidetti due indici di variabilità assoluta (11):

$$\frac{\sqrt{u_{m_w^{-2}}}}{\sigma_w^-} = \frac{\sqrt{\sigma_w^{-2} + (M_w^- - m_w^-)^2}}{\sigma_w^-} = \frac{\sqrt{\left[\frac{x_N^- - x_1^-}{k_N^- - k_1^-} \right]^2 + (M_w^- - m_w^-)^2}}{\frac{x_N^- - x_1^-}{k_N^- - k_1^-}}, \quad (2)$$

(11) Al numeratore di questo rapporto è stato posto l'indice approssimato della deviazione dalla moda dell'insieme limitato chiuso da due misure estreme, o di soglia, del fenomeno qualitativo, prestabilite da una comunità di esperti del fenomeno qualitativo valutabile come detto prima. Esso serve, come già detto, per calcolare la variabilità imputabile ad una prevedibile causa sistemática, regolare o normale, ascrivibile ad un fenomeno qualitativo, che non può essere separata dalla variabilità imputabile ad un imprecisabile numero di cause imprevedibili, dette accidentali od irregolari, che si traducono, in concreto, in un altrettanto imprecisabile numero di errori casuali di misura del predetto fenomeno quantitativo.

Tale indice di variabilità assoluta per così dire totale presuppone, quindi, che le anzidette cause di variabilità siano inseparabili e, perciò, congiunte. La struttura di codesto indicatore tiene conto, infatti, di due origini o sorgenti di variabilità.

L'indice di variabilità assoluta posto al denominatore dell'anzidetto rapporto serve, invece, per calcolare la variabilità imputabile esclusivamente ad un imprecisabile numero di cause accidentali che, come già detto, non sono in realtà separabili dalla causa sistemática. Però, a scanso di equivoci, bisogna dire che tale indice di variabilità assoluta per così dire parziale non può essere calcolato isolatamente, in quanto si riferisce alla distribuzione empirica unimodale (asimmetrica) di frequenza della variabile "non casuale" o statistica semplice che soggiace all'ipotesi distributiva gaussiana ma non mesocurtica. Ciò equivale a dire che la distribuzione empirica unimodale (asimmetrica) di frequenza della variabile in questione si mescola, in realtà, ad un'inseparabile distribuzione empirica campanulare o gaussiana ma non realmente simmetrica e stabilizzata di frequenza di errori casuali di misura, la cui entità è un numero non precisabile.

È da sottolineare che una variabile continua può essere considerata casuale, oppure "non casuale", a seconda del meccanismo di selezione dei soggetti o degli oggetti sui quali osservare le misure del carattere metrico (selezione in base ad una tavola di numeri casuali, oppure in base alla diagnosi o proposizione, come detto prima, di una qualità). Tuttavia, non è un controsenso ipotizzare la casualità o gaussianità della variabile "non casuale" o statistica semplice, perché la selezione di determinati soggetti od oggetti mediante la diagnosi, effettuata come detto prima, di una loro qualità, detta anche causa sistemática, induce a pensare che possano esistere in realtà anche imprevedibili cause, cosiddette accidentali, di numero imprecisabile difficilmente separabili dalla causa sistemática determinante.

Perciò, l'esame della variabilità di un fenomeno quantitativo da effettuare dopo la valutazione di un fenomeno qualitativo in base all'esperienza e con l'ausilio della misura del fenomeno quantitativo, induce a considerare, come già detto, due origini o sorgenti di variabilità.

Invece, come è noto, la teoria statistica della stima presuppone che si debba effettuare l'inferenza induttiva mediante il processo di rilevamento delle informazioni statistiche per così dire incompleto riferito, cioè, ad uno od a più di un sottoinsieme di soggetti o di oggetti da selezionare con un meccanismo di sorteggio casuale, per cui può essere effettuato l'esame dell'omogeneità della varianza dei campioni statistici, mediante la funzione test "F". Pertanto, le rilevazioni campionarie ed il conseguente calcolo dell'errore probabile, dipendente dal numero ristretto delle osservazioni, inducono a volte gli statistici che usano i metodi di stima a ritenere questo calcolo il più importante requisito della rilevazione. Ciò è stato ritenuto esagerato, perché il

può essere impiegato come moltiplicatore della lunghezza dell'intervallo limitato chiuso dell'anzidetta variabile potenzialmente continua statistica semplice standardizzata, rispetto all'indice approssimato della deviazione dalla moda della distribuzione empirica unimodale (asimmetrica) di frequenza della variabile statistica semplice:

$$\bar{W}' = \frac{(\bar{x}'_N - m'_w) - (\bar{x}'_1 - m'_w)}{\sqrt{\left[\frac{\bar{x}'_N - \bar{x}'_1}{k'_N - k'_1}\right]^2 + (M'_w - m'_w)^2}} = \bar{k}'_N - \bar{k}'_1 \quad (3)$$

Infatti, il prodotto tra le formule (2) e (3) determina un risultato uguale alla lunghezza approssimata [Formula (1)] dell'anzidetto immaginario intervallo aperto, desumibile dalla citata tavola proutuario. Pertanto, la formula (2) è utile come fattore di conversione della formula (3) per confermare la lunghezza o misura, in realtà stabile, espressa con $\bar{w}' = \bar{x}'_N - \bar{x}'_1$, dell'intervallo limitato chiuso dalle misure estreme, o di soglia, di un carattere metrico collegabile ad un fenomeno qualitativo e, quindi, per confermare la lunghezza dell'intervallo limitato chiuso della variabile potenzialmente continua statistica semplice. L'anzidetto moltiplicatore è, dunque, il criterio statistico per confermare l'omogeneità di determinati soggetti od oggetti che costituiscono un insieme finito, od una classe, la cui entità è un numero che può stabilire chi valuta un fenomeno qualitativo in base all'esperienza e con l'ausilio della misura di un fenomeno quantitativo collegabile, secondo la logica consueta, a quello qualitativo.

La teoria sulla stabilità della variabilità imputabile ad un'immaginaria serie, tendente ad infinito, di cause accidentali esiste da quando è stato enunciato il teorema di Bernoulli, noto come "legge dei grandi numeri" (Boldrini M. 1962., pp. 1172-1173). Inoltre, una misura della stabilità della variabilità per queste cause fu introdotta dal Lexis

nel 1876 con il rapporto: $Q = \frac{\sigma}{\mu}$ che denominò coefficiente di divergenza. In seguito, R.A. Fisher osservò, opportunamente, che il para-

primo requisito della rilevazione è che essa risponda allo scopo della ricerca ed il secondo requisito è che si possa calcolare l'errore probabile del campione (Gini C. 1962. La logica nella statistica, ecc., pag. 28). Secondo il nostro punto di vista, quindi, occorre un adeguato metodo statistico per l'esame della variabilità, rispetto alla moda, di un fenomeno quantitativo collegabile ad un fenomeno qualitativo o descrittivo che limita, in realtà, l'ordine delle misure di quello quantitativo, per cui è possibile l'esame del variabile stabile in realtà.

metro X^2 di Pizzetti-Pearson è una funzione del quoziente di divergenza e precisamente: $Q^2 v = X^2$, dove v è il numero dei gradi di indipendenza. Quindi, in un'empirica serie bernoulliana di n estrazioni sarebbe soddisfacente l'accordo fra le probabilità e le frequenze effettive dei successi se risultasse, almeno in modo approssimato, l'uguaglianza tra lo scarto quadratico medio empirico σ della serie sperimentale e lo

scarto quadratico medio teorico della serie bernoulliana, $\mu = \sqrt{\frac{pq}{n}}$,

ricavato in base alle probabilità date a priori p e q . In altri termini, la variabilità imputabile ad un'empirica serie di cause imprevedibili, cosiddette accidentali, sarebbe stabile se il rapporto tra gli anzidetti due scarti quadratici medi fosse sensibilmente prossimo all'unità. Il Lexis, però, non ha mancato di mettere in guardia contro i significati erronei che si possono attribuire al quoziente di divergenza, anche se si tratta di fenomeni misurabili con dispersione gaussiana. Infatti, tale quoziente non è mai uguale all'unità e, perciò, subisce una certa influenza del numero delle osservazioni. Per questo motivo, in seguito, gli statistici hanno cercato di determinare i limiti di attendibilità del quoziente di divergenza (Boldrini M. 1962., *op. cit.*, pp. 1182-1183).

Invece, il rapporto tra i due indici di variabilità assoluta proposto per confermare l'omogeneità di determinati soggetti od oggetti, rispetto ad una qualità valutabile come detto prima, e per determinare i confini della normalità o regolarità delle oscillazioni delle frequenze intorno al loro massimo cui corrisponde la moda del carattere metrico collegabile alla qualità dei soggetti o degli oggetti medesimi, può risultare in realtà maggiore dell'unità, senza che ciò possa fare escludere l'effettiva stabilità della variabilità, rispetto alla moda, della distribuzione empirica unimodale (asimmetrica) di frequenza della variabile "non casuale" o statistica semplice potenzialmente continua in un intervallo limitato chiuso.

Come è noto, il Gini ha proposto l'indice di eterogeneità delle unità statistiche incluse in una distribuzione empirica di frequenze relative di una mutabile statistica (12) con due o più di due modalità (13),

(12) Come, ad es., il sesso degli studenti di una sede di un ateneo universitario che abbiano frequentato le lezioni durante un anno accademico, oppure il corso di laurea a cui risultano scritti gli studenti stessi.

(13) Come, ad es., l'insieme costituito dai maschi e dalle femmine degli studenti di una sede di ateneo universitario che hanno frequentato le lezioni durante un anno accademico, oppure l'insieme costituito dai differenti corsi di laurea a cui risultano scritti gli studenti medesimi.

ossia: $S = 1 - \sum_{i=1}^n \left(\frac{y_i}{\sum_{i=1}^n y_i} \right)^2$. Cioè un indice da impiegare nel caso di un

insieme finito di soggetti o di oggetti eterogenei, rispetto a prestabilite qualità distintive, il cui totale è posto uguale ad 1. Ciò è menzionato in due testi destinati alla didattica (Leti G. 1983., pag. 261 e pag. 266. Inoltre consulta: Fraire M. Rizzi A. 1999., pag. 210).

Il Leti, invece, ha curato una rassegna degli indici e delle distanze che appartengono al metodo statistico da impiegare nel caso degli insiemi dei soggetti o degli oggetti aventi in realtà qualità anche diverse, ossia di collezioni di elementi nel senso matematico (Leti G. 1990). A tutto oggi, invece, non ci pare che sia stato suggerito il criterio statistico per confermare l'omogeneità – rispetto ad un fenomeno qualitativo valutabile in base all'esperienza e con l'ausilio della misura di un fenomeno quantitativo collegabile, secondo la logica consueta, a quello qualitativo – di determinati soggetti od oggetti che costituiscono un insieme finito, od una classe, da includere in una seriazione, detta anche distribuzione di quantità (Leti G. 1983., op., cit., pp. 107-111).

Gli studi sulla stabilità dell'immaginario valore che potrebbe assumere lo scarto quadratico medio, espresso con $\bar{\sigma}$, delle teoriche misure di un carattere metrico (grandezza astratta) di un insieme infinito di campioni casuali di soggetti o di oggetti, perturbate da un'immaginaria serie, tendente a infinito, di errori di misura dovuti al caso, riconducibile alla legge empirica del caso nota anche come legge dei grandi numeri dovuta a Bernoulli, indussero – com'è noto – K. Pearson ad evocare l'idea della forma distributiva gaussiana mesocurtica della variabile continua casuale gaussiana standardizzata, rispetto al teorico valore di $\bar{\sigma}$ non determinabile e, quindi, incognito delle teoriche misure dell'anzidetto carattere metrico dell'immaginario insieme infinito dei campioni casuali dei soggetti o degli oggetti. Ciò, però, è utile per effettuare l'inferenza induttiva con il processo di rilevamento delle informazioni statistiche per così dire incompleto che privilegia il principio che tali informazioni debbano essere quelle inerenti un sottoinsieme di soggetti o di oggetti selezionati con il meccanismo del sorteggio casuale previsto dall'anzidetto schema di probabilità dato a priori.

Si ribadisce, quindi, che la forma distributiva campanulare simmetrica e stabilizzata o mesocurtica degli errori casuali standardizzati

delle teoriche misure di un fenomeno quantitativo da considerare, per lo schema di un'ipotesi, una grandezza matematica, e, per conseguenza, la stabilità della lunghezza dell'intervallo aperto della variabile continua casuale gaussiana semplice standardizzata non può essere riscontrata in realtà, essendo un'ipotesi e nulla più. Inoltre, determinati soggetti od oggetti notevolmente numerosi selezionati in base all'anzidetto meccanismo del sorteggio casuale costituiscono in realtà una parte di un insieme d'ordine superiore che è l'immaginario insieme infinito dei campioni casuali dei soggetti o degli oggetti da non osservare.

II) *Il criterio statistico per scegliere le estreme frequenze regolari o normali corrispondenti a due percentili empirici, cui corrispondono due misure di un intervallo per così dire parziale o "range" interpercentile compreso nell'intervallo limitato chiuso dalle misure estreme, o di soglia, di un fenomeno quantitativo collegabile ad un fenomeno qualitativo. Cioè, il criterio per scegliere le estreme frequenze regolari o normali corrispondenti a due misure, poste minori od uguali rispetto ad esse stesse, a cavallo ma non necessariamente equidistanti dalla moda, della distribuzione empirica unimodale (asimmetrica) di frequenza cumulata della cosiddetta variabile potenzialmente continua "non casuale" o statistica semplice standardizzata, rispetto all'indice approssimato della deviazione dalla moda della distribuzione empirica unimodale (asimmetrica) di frequenza della variabile statistica semplice priva di distribuzione teorica di densità di probabilità.*

In sostanza, il criterio statistico per separare una notevole quota percentuale (95%) di un insieme finito, o di una classe, di soggetti o di oggetti omogenei che, per convenzione, hanno frequenza regolare o normale, perché associabili alle misure di un loro carattere metrico perturbate probabilmente dagli errori casuali non molto distanti dall'errore nullo (14).

Per quanto detto al punto II) si ritiene utile come termine di paragone il valore critico o di equilibrio (15) dell'immaginaria lunghezza

(14) Come già detto, la teoria statistica della stima presuppone che si debba effettuare l'inferenza induttiva mediante il processo di rilevamento delle informazioni statistiche per così dire incompleto riferito, cioè, ad uno od a più di un sottoinsieme di soggetti o di oggetti da selezionare con un meccanismo di sorteggio casuale, per cui può essere effettuato l'esame dell'omogeneità della varianza dei campioni statistici mediante la funzione test "F".

(15) In Matematica, come è noto, una delle principali nozioni relative ad un sistema di equazioni autonome è quello di punto di equilibrio (o punto critico) e la definizione più importante è la stabilità di un punto di equilibrio del sistema. Altre nozioni della teoria della stabilità

di una parte dell'intervallo aperto (16) della variabile continua casuale gaussiana standardizzata. Ossia, la lunghezza dell'intervallo aperto limitato ai due lati da due errori casuali standardizzati, rispetto allo s.q.m. non determinabile e, quindi, incognito, k_{\max} e k_{\min} , in teoria noti, che corrispondono ai due percentili teorici 97,5 ($\bar{p}_{97,5}$) e 2,5 ($\bar{p}_{2,5}$) della distribuzione teorica di densità di probabilità cumulata degli errori casuali standardizzati.

Cioè:

$$\begin{aligned} W_{\bar{p}_{97,5} - \bar{p}_{2,5}}^- &= \frac{(X_{\max} - M_X) - (X_{\min} - M_X)}{\sigma_X} = k_{\max} - k_{\min} = \\ &= (1,959964) - (-1,959964) = 3,919928. \end{aligned} \quad [4]$$

In pratica, può essere scelta una coppia di percentili empirici fra cui comprendere il 95% del totale della distribuzione empirica di frequenza cumulata della variabile potenzialmente continua statistica semplice, tale che la lunghezza dell'intervallo delle misure del carattere metrico limitato dalle misure corrispondenti ai due percentili di tale coppia, standardizzata, rispetto all'indice approssimato della deviazione dalla moda della distribuzione empirica unimodale (asimmetrica) di frequenza della variabile in questione, cioè:

$$\bar{W}'_{p_{\max} - p_{\min}} = \frac{\bar{x}'_{\max} - \bar{m}'_w}{\sqrt{\mu_{m_w'^{-2}}}} - \frac{\bar{x}'_{\min} - \bar{m}'_w}{\sqrt{\mu_{m_w'^{-2}}}} = \frac{\bar{x}'_{\max} - \bar{x}'_{\min}}{\sqrt{\mu_{m_w'^{-2}}}} = \bar{k}'_{\max} - \bar{k}'_{\min} \quad [5],$$

risulti, dopo l'effettuazione del prodotto con l'anzidetto fattore di conversione (formula [2]), la più vicina al valore teorico critico. Occorre quindi:

a) Scegliere diverse coppie sperimentali di percentili empi-

per i punti di equilibrio di un sistema di equazioni autonome sono: lo spazio delle fasi, l'orbita, la traiettoria, la caratteristica, il ciclo, il ritratto di fase o quadro generale dell'andamento delle traiettorie e l'integrale primo (Salsa S. Squellati A. 1995., pp. 241-242).

(16) Cui corrisponde la probabilità integrale di livello 95%, concentrata in prossimità della mediana cui corrisponde il percentile teorico 50 (\bar{p}_{50}) della distribuzione teorica di densità di probabilità cumulata degli errori casuali standardizzati delle teoriche misure di una grandezza astratta, desumibile dalla citata tavola prontuario.

rici (17), cui corrispondono coppie di misure a cavallo della moda, tra cui comprendere in modo approssimato una notevole quota percentuale (95%) del totale dell'empirica distribuzione unimodale delle frequenze delle misure del carattere metrico dei soggetti o degli oggetti con la qualità valutabile come detto prima.

b) Standardizzare, rispetto all'indice approssimato della deviazione dalla moda, la variabile potenzialmente continua statistica semplice e, quindi, la lunghezza di una parte dell'intervallo, detto anche intervallo parziale, della variabile in questione che si ottiene per differenza tra le due misure corrispondenti ai due percentili empirici d'ogni coppia, in modo da ottenere diversi valori dell'intervallo espresso con la formula [5].

c) Convertire la lunghezza o misura dei diversi intervalli parziali della variabile potenzialmente continua statistica semplice standardizzata, con distribuzione empirica di frequenza cumulata, moltiplicando i diversi valori che si ottengono con la formula [5] per il quoziente proposto come fattore di conversione (formula [2]) e scegliere il risultato più vicino al valore teorico critico (3,919928) della lunghezza di una parte dell'anzidetto intervallo aperto, cui corrisponde la somma integrale di livello 95% delle densità delle probabilità cumulate calcolate con la funzione di ripartizione della funzione di densità di probabilità della variabile continua casuale gaussiana semplice standardizzata, il cui limite è l'integrale improprio proposto da Sheppard.

In tal modo, è possibile scegliere un intervallo limitato da due misure corrispondenti ad una coppia di percentili empirici della distribuzione empirica di frequenza cumulata della variabile potenzialmente continua statistica semplice standardizzata. Cioè, un intervallo limitato da due misure del fenomeno quantitativo collegabile a quello qualitativo, poste minori od uguali rispetto ad esse stesse, in cui è compreso in realtà il 95% del totale dei soggetti o degli oggetti omogenei concentrati in prossimità del massimo di frequenza cui corrisponde la moda.

d) Considerare le misure corrispondenti ai due percentili empirici di tale coppia quelle a cavallo, ma non necessariamente equidistanti dalla moda del fenomeno quantitativo, cui corrispondono le estreme frequenze regolari o normali dell'empirica distribuzione dei soggetti o degli oggetti omogenei che costituiscono un insieme finito, od una classe.

(17) Il numero delle coppie sperimentali dei percentili empirici da prescegliere è quello che si può al massimo considerare per comprendere nell'intervallo limitato da ogni loro coppia il 95% del totale dei soggetti o degli oggetti omogenei. Cioè le coppie dei percentili: 96 e 1; 97 e 2; 97,5 e 2,5; 98 e 3; 99 e 4 ed, infine, 100 e 5.

Ciò che è detto al punto (c) sarebbe, pertanto, il criterio statistico per scegliere le estreme frequenze regolari o normali, corrispondenti a due misure di un carattere metrico collegabile ad un fenomeno qualitativo. Cioè, il criterio per scegliere una notevole quota percentuale dei soggetti o degli oggetti omogenei, rispetto al fenomeno qualitativo valutabile come detto prima, che costituiscono un insieme finito od una classe.

4. **Discussione**

I due criteri statistici proposti, rispettivamente, per confermare l'omogeneità, rispetto ad una qualità naturale, od artificiale, di determinati soggetti od oggetti che costituiscono un insieme finito, od una classe, e per determinare i confini della regolarità o normalità delle oscillazioni delle frequenze intorno al loro massimo cui corrisponde la moda del carattere metrico dei soggetti o degli oggetti in questione, possono essere utili nelle indagini sul controllo della qualità della produzione industriale, dove è da considerare l'immaginabile lunghezza di un intervallo limitato chiuso da due misure estreme, o di soglia, di un carattere metrico collegabili a taluni di un insieme finito, o di una classe, d'oggetti omogenei, rispetto al fenomeno qualitativo prodotto artificialmente. Inoltre, tali criteri possono essere utili nelle indagini epidemiologiche, in occasione di uno screening, quando la comunità scientifica del settore decide che una misura di soglia debba indicare verosimilmente l'anormalità clinica e la decisione che un individuo sia verosimilmente malato debba essere presa sulla base dello screening medesimo, mentre in realtà, di solito, essa viene basata sulla "anormalità statistica", piuttosto che sull'anormalità clinica, esistendo il problema degli individui border-line.

Quando si effettua l'inferenza induttiva con il processo di rilevamento delle informazioni statistiche per così dire incompleto riferito, cioè, ad un sottoinsieme di soggetti o di oggetti (18) selezionati con un

(18) Distribuiti secondo le misure di un carattere metrico ordinate come una successione crescente con cui, però, non è possibile calcolare, ma stimare, i confini di normalità statistica dei casi selezionati, come già detto, dalla popolazione dei soggetti o degli oggetti di riferimento, aventi in realtà qualità anche diverse. Infatti, il meccanismo del sorteggio casuale previsto dallo schema d'ipotesi o probabilistico gaussiano standardizzato presuppone che i soggetti o gli oggetti appartengano, in teoria, ad un immaginario insieme infinito di campioni casuali di soggetti o di oggetti caratterizzati da una immaginaria serie, tendente ad infinito, di cause accidentali.

meccanismo di sorteggio casuale si parla, spesso, di casualità o normalità di una qualità naturale od artificiale associabile, senza la ricerca di un'effettiva corrispondenza logica, ad un fenomeno quantitativo. Cioè, un fenomeno qualitativo si considera, per lo schema di un'ipotesi, un probabile fenomeno casuale od eventuale, di una serie tendente ad infinito. Per conseguenza, non si allude alla normalità o regolarità delle oscillazioni delle frequenze intorno al loro massimo cui corrisponde la moda di un carattere metrico collegabile ad una qualità che può essere valutata, come detto prima, da un esperto, la cui comunità ha il compito di prestabilire le misure estreme, o di soglia (19), dell'anzidetto fenomeno quantitativo da collegare a taluni di un insieme finito, o di una classe, di soggetti o di oggetti omogenei, rispetto al fenomeno qualitativo (20).

Perciò, il ricercatore che effettua un'indagine statistica campionaria tende, talvolta, ad esaminare la stabilità di variabilità imputabile ad un'immaginaria serie, tendente a infinito, di cause accidentali e decide in senso probabilistico se è vera l'ipotesi che la distribuzione empirica di frequenza per così dire casuale delle misure del fenomeno quantitativo esaminato (21) abbia una forma gaussiana mesocurtica (22) o simile a quella della distribuzione teorica delle densità delle probabilità degli errori casuali standardizzati delle misure del fenomeno quantitativo considerato una grandezza astratta di un immaginario insieme infinito di campioni casuali di soggetti o di oggetti da non osservare. Per accettare la suddetta ipotesi, quindi, egli è orientato ad assumere come livello di probabilità integrale delle empiriche misure del fenomeno quantitativo, considerato una grandezza astratta, la quota del 95%

(19) Come, ad es., le misure di soglia del reddito collegabili a taluni della classe delle persone fisiche omogenee, rispetto al fenomeno qualitativo della ricchezza, o della povertà. Oppure, le misure di soglia della glicemia collegabili a taluni della classe dei soggetti omogenei, rispetto al fenomeno qualitativo del buono stato di salute, oppure della malattia diabetica, ecc.

(20) Spesso, la distribuzione empirica di frequenza delle misure di un carattere metrico di un insieme finito di soggetti o di oggetti omogenei, rispetto ad una qualità valutata da un esperto, non è apparsa di forma simile a quella della distribuzione teorica di densità di probabilità degli errori casuali standardizzati delle teoriche misure di una grandezza astratta. Perciò, l'immaginaria forma distributiva gaussiana mesocurtica non può essere in realtà un indizio di omogeneità, rispetto ad una qualità, di determinati soggetti od oggetti distribuiti secondo le empiriche misure di un fenomeno quantitativo.

(21) Associabile ad un fenomeno qualitativo considerato – per lo schema di un'ipotesi – un probabile fenomeno casuale, di una serie tendente ad infinito.

(22) Da considerare un indizio di stabilità di variabilità, rispetto alla teorica media aritmetica delle misure del carattere metrico considerato una grandezza astratta dell'immaginario insieme infinito dei campioni casuali dei soggetti o degli oggetti, in realtà da non osservare.

dell'area situata a cavallo del valore massimo di probabilità, rappresentato da una delle ordinate avente come ascissa la teorica media aritmetica della distribuzione di densità di probabilità a cui fa riferimento per effettuare l'inferenza. Questo livello di probabilità integrale corrisponde, dunque, alla lunghezza di una parte dell'intervallo aperto, o campo $-W_{\bar{p}_{97,5} - \bar{p}_{2,5}} = 3,919928$ – limitato ai due lati da due errori casuali semplici standardizzati di uguale intensità e di segno contrario, in teoria noti, espressi in valore assoluto con $|\pm k| = 1,959964$ (23) ed inerenti i due percentili teorici 97,5 ($\bar{p}_{97,5}$) e 2,5 ($\bar{p}_{2,5}$) della distribuzione teorica di densità di probabilità integrale della variabile casuale gaussiana standardizzata, tabulata in un'apposita tavola prontuario (Vianelli S. 1959., *op.*, *cit.*), utile per effettuare stime statistiche. Il livello della probabilità integrale (area) esclusa dall'anzidetta distribuzione teorica, riferita ad un immaginario insieme infinito di campioni casuali di soggetti o di oggetti in realtà da non osservare, è noto come “livello di significatività” necessario per effettuare il saggio dell'ipotesi della stabilità di variabilità imputabile ad un'immaginaria serie tendente ad infinito di cause accidentali riconducibile all'anzidetta forma gaussiana mesocurtica della distribuzione teorica di densità di probabilità degli errori casuali standardizzati, da cui scaturisce la distribuzione empirica di frequenza per così dire casuale od eventuale delle misure del fenomeno quantitativo esaminato per effettuare l'inferenza induttiva con il processo di rilevamento delle informazioni statistiche per così dire incompleto.

Tale livello di probabilità integrale rappresenta, dunque, l'area di rifiuto, anziché di accettazione, della predetta ipotesi. Però, se l'indagine campionaria induce alla conclusione del rifiuto dell'ipotesi distributiva gaussiana mesocurtica, allora, il ricercatore è costretto a rimandare, ad un'ulteriore indagine statistica campionaria più numerosa (24), la conclusione sulla casualità del fenomeno qualitativo e sulla stabilità

(23) Cioè, due delle teoriche misure di una grandezza matematica, distanti dalla teorica media aritmetica non determinabile e, quindi, incognita, rapportate al teorico scarto quadratico medio non determinabile e, quindi, incognito delle predette misure.

(24) Come ha deciso, ad esempio, la commissione sanitaria nominata dal governo italiano per stabilire se l'uranio impoverito, sparso sul territorio balcanico dalle armi usate dalle forze militari della Nato, fosse la causa dei linfomi nel sangue dei soldati reduci dalla missione di pace in Bosnia. Oppure, come ha sostenuto un famoso genetista italiano, in occasione di una rassegna canora, allorchè invitò il pubblico televisivo alla raccolta di risorse finanziarie sempre più cospicue per gli studi sulla formazione della mappa cromosomica del genere umano.

della variabilità, rispetto all'empirica media aritmetica delle misure del fenomeno quantitativo, imputabile ad un'immaginaria serie di cause accidentali (25). Ciò, però, secondo il nostro punto di vista, è un discorso circolare inarrestabile, perché la forma distributiva gaussiana mesocurtica non è riscontrabile in realtà. Quindi, ogni volta, è possibile rinviare il saggio dell'anzidetta ipotesi ad un sottoinsieme di soggetti o di oggetti più numerosi.

5. Considerazioni su alcune tecniche di ricampionamento per stimare, anziché calcolare, i cosiddetti confini di “normalità statistica” dei casi enumerati in occasione della diagnosi di un fenomeno qualitativo collegato ad un fenomeno quantitativo

Se un ricercatore valuta, in base all'esperienza e con l'ausilio della misurazione di un carattere metrico, un fenomeno qualitativo non misurabile – come, per es., uno stato di salute – dà luogo alla formazione di un insieme finito, o di una classe, di soggetti (termine usato in Epidemiologia) o di oggetti omogenei, rispetto alla qualità valutata come detto prima. Però, egli non può impiegare un modello probabilistico, dato a priori, per rappresentare, in forma schematica, il nesso causale tra le due variabili (la variabile dipendente costituita da determinati soggetti od oggetti omogenei, rispetto alla qualità valutata come detto prima, e la variabile indipendente costituita dalle misure del carattere metrico che lo hanno aiutato a valutare il fenomeno qualitativo o descrittivo). Quindi, egli è indotto a cercare una soluzione alternativa al problema della stima dei parametri incogniti di una funzione di densità di

(25) Come è noto, la dimensione n di un campione è legata alla variabilità, rispetto alla teorica media aritmetica non determinabile e, quindi, incognita delle misure di un fenomeno quantitativo considerato, per lo schema di un'ipotesi, una grandezza matematica di un'immaginario insieme infinito di campioni casuali di soggetti o di oggetti. Quindi, a parità di altre condizioni, la dimensione del campione dovrebbe essere tanto più grande, quanto più grande sarebbe codesta variabilità (Vianelli S. Ferreri C. 1966., pag. 309). Però, secondo il nostro punto di vista, l'esame della variabilità, rispetto alla moda, di un fenomeno quantitativo collegabile ad un fenomeno qualitativo o descrittivo naturale od artificiale che limita in realtà l'ordine delle misure di quello quantitativo di determinati soggetti od oggetti esistenti in realtà, comporta la valutazione dell'anzidetto fenomeno qualitativo in base all'esperienza e con l'ausilio della misura del carattere metrico. Perciò, come già detto, un fenomeno qualitativo o descrittivo caratterizza determinati soggetti od oggetti, la cui entità è sempre in realtà un numero che può essere stabilito in base all'esperienza. Quindi, se si tratta di un raro fenomeno quantitativo, questo numero può essere in realtà esiguo.

probabilità. A questo proposito, se si fa riferimento al problema della determinazione dei limiti di normalità statistica con tecniche di stima non parametrica come, ad es., i metodi di ricampionamento Jackknife e quello di Bootstrap, che sono alla base della costruzione degli intervalli di confidenza non parametrici (Mignani S. 1990., pag. 17), bisogna ammettere che tali metodi sono indipendenti dalla distribuzione delle empiriche frequenze delle misure del fenomeno quantitativo esaminato ed, in genere, però, il loro contributo risiede sulla possibilità di risolvere problemi di stima dei parametri di un modello stocastico. Infatti, il loro concetto base consiste nel ricampionare adeguatamente le osservazioni originarie, mantenendo la struttura stocastica iniziale e studiando le proprietà dello stimatore (26) prescelto con riferimento ai “pseudo-dati”. L’idea di fondo di tali tecniche è, perciò, la creazione di uno “pseudo-universo” da cui trarre tutte le informazioni che sarebbero da cercare, invece, nell’universo campionario (Mignani S. 1990., *op. cit.*, pag. 4). Quindi, i metodi Jackknife e Bootstrap, anche se sono diversi per taluni aspetti non trascurabili, scaturiscono dalla stessa idea di riorganizzare o, per meglio dire, ricombinare le osservazioni campionarie.

Abbiamo applicato la tecnica di ricampionamento Bootstrap, effettuando 1.000 replicazioni con il pacchetto applicativo informatico S-Plus 4.5 sulla distribuzione empirica univariata plurimodale asimmetrica di frequenza delle misure di glicemia a digiuno e dopo due ore dal carico di 75 g di glucosio dell’insieme finito dei soggetti eterogenei, per stato di salute, di ambo i sessi, di 20 anni di età ed oltre censiti a Pantelleria, in occasione di un’indagine epidemiologica sul diabete mellito. Tali distribuzioni empiriche (27) sono rappresentate in tabelle

(26) Come è noto, lo stimatore è un’opportuna funzione (di probabilità) dei dati campionari dei quali presuppone una sintesi quantitativa come, ad es., la media aritmetica campionaria, oppure la varianza campionaria. La funzione di probabilità che si adatta ai dati campionari o stimatore può assumere una certa forma data a priori. Inoltre, può essere la funzione di stima di un parametro di un insieme finito di soggetti naturali o di unità materiali od oggetti esistenti in realtà che hanno qualità anche diverse, supposte cause accidentali. Tale insieme è detto anche popolazione finita di riferimento (Mignani S. 1992., pag. 52, pag. 97 e pag. 463).

(27) Esse hanno consentito l’esame di un’altra analisi statistica svolta da un cultore della teoria statistica della stima (Rossi C. 1988., pp. 429-442). Lo studio ebbe inizio nella fase di programmazione della predetta inchiesta epidemiologica, effettuata nel 1981, i cui risultati preliminari sono stati pubblicati nel 1985 (Savagnone e coll. 1985). Allora, si doveva decidere se effettuare un’indagine statistica descrittiva per conoscere l’effettiva incidenza e prevalenza della malattia diabetica, oppure ripiegare – per motivi di costo – su un’indagine statistica campionaria. Il dilemma era, cioè, se effettuare il censimento della popolazione classificabile secondo il numero degli

numeriche che si trovano in appendice a questo testo. Come già detto, l'uso di queste tecniche non è utile per stimare i confini di normalità statistica ed, in questo caso, invece, servono per separare una prestabilita quota percentuale (95%) di soggetti aventi qualità diverse e, quindi, non classificati ma censiti in occasione di un'indagine epidemiologica e distribuiti secondo le misure di glicemia. È da precisare, però, che tali soggetti sono stati opportunamente classificati, secondo il loro stato di salute, da taluni medici (Savagnone E., e coll. 1985., *op., cit.*, pp. 71- 80).

L'esito della procedura di ricampionamento Bootstrap è stato negativo, in quanto non è stato possibile stimare con i dati disponibili il parametro di interesse che è la media aritmetica supposta uguale alla mediana ed alla moda del fenomeno quantitativo che si considera, per lo schema di un'ipotesi, variabile casuale gaussiana standardizzata (28). Lo stesso esito è stato ottenuto con la procedura di ricampionamento Jackknife.

Per la verità, prima dell'approccio a queste tecniche, si pensava che fosse scontato tale risultato, per i motivi che si possono meglio intendere approfondendo il significato delle tecniche medesime per le quali si rinvia alla letteratura specifica (29). Si ritiene, dunque, che, ai

stati di salute che di solito è sconosciuto ma è sommariamente stabilito dai medici che, spesso, distinguono la popolazione diabetica da quella predisposta alla malattia diabetica (con ridotta tolleranza agli idrati di carbonio) e da quella non diabetica, detta anche normale, essendo quella prevalente. Oppure, effettuare l'estrazione casuale di un campione di soggetti stratificati in base agli stati di salute previsti dai medici, oltre che per sesso ed età, dalla popolazione di Pantelleria, caratterizzata – in realtà – da diverse presumibili qualità. Cioè, si doveva decidere se considerare tre classi della popolazione umana di quest'isola del Mediterraneo classificata dai medici e, quindi, considerare le corrispondenti mode glicemiche e gli indici della deviazione dalla moda degli insiemi limitati chiusi dalle misure di soglia di glicemia prestabilite dagli stessi medici. Oppure, considerare campioni casuali di soggetti appartenenti a queste classi e, quindi, stimare le medie aritmetiche e gli scarti quadratici medi dei campioni delle misure di glicemia che ne sarebbero scaturiti. A quell'epoca, però, mancava in letteratura la formula dell'indice approssimato della deviazione dalla moda della distribuzione empirica unimodale (asimmetrica) di frequenza della variabile "non casuale" o statistica semplice potenzialmente continua in un intervallo limitato chiuso, o compatto.

(28) È da sottolineare che la media aritmetica, la moda e la mediana delle teoriche misure, perturbate dagli errori casuali standardizzati, di un carattere metrico considerato, per lo schema di un'ipotesi, una grandezza astratta di un'immaginario collettivo infinito di campioni casuali di soggetti o di oggetti sono uguali. Tuttavia, l'anzidetta teorica media aritmetica è da riferire alla distribuzione delle densità delle probabilità degli errori casuali standardizzati, mentre la teorica misura mediana, cui corrisponde il teorico percentile 50, è da riferire alla distribuzione cumulativa delle densità delle probabilità degli errori casuali standardizzati delle predette teoriche misure.

(29) Nella teoria classica della stima statistica assumono, comunque, particolare importanza tre ordini di fattori (Mignani S. 1992., *op., cit.*, pag. 3):

- il modello di campionamento;
- il tipo di parametro;
- la distribuzione del carattere nella popolazione.

fini del calcolo, anziché della stima, dei confini della “normalità statistica” dei casi o, per meglio dire, dei soggetti o degli oggetti da scegliere in base alla diagnosi di un loro prestabilito fenomeno qualitativo valutabile in base all’esperienza e con l’ausilio della misura di un fenomeno quantitativo collegabile, secondo la logica consueta, a quello qualitativo, conti di più la valutazione dell’anzidetto fenomeno qualitativo non misurabile. Quindi, piuttosto che la ricombinazione delle osservazioni campionarie, è necessario il censimento e la classificazione dei soggetti o degli oggetti di riferimento, poiché questi ultimi sono, di solito, eterogenei, rispetto a prestabilite qualità valutabili come detto prima.

6. Considerazioni sul significato dei termini usati

In tutte le discussioni scientifiche, la prima condizione per arrivare a conclusioni univoche è di conoscere bene il significato dei termini che si usano. Ciò si fa sempre perché il significato delle parole deve essere univoco e noto universalmente (Gini C. 1962. *La logica nella statistica*, ecc., op., cit., pag. 12).

Il Gini ha osservato, inoltre, che bisogna interpretare con una certa cautela i risultati della ricerca empirica effettuata mediante l’indagine statistica. Infatti, l’inferenza induttiva svolta con il processo di rilevamento delle informazioni statistiche per così dire incompleto è diversa dalla comune induzione insegnata dalla logica (Gini C. 1962. *La logica nella statistica*, ecc., op., cit., pag. 10). L’intuizione umana trova, infatti, riscontro in risultati a cui si può arrivare consciamente solo mediante procedimenti complessi (Gini C. 1962. *La logica nella statistica*, ecc., op., cit., pag. 85).

Secondo il nostro punto di vista, il metodo statistico descrittivo sarebbe quello che dovrebbe avere per oggetto lo studio di una o più di

In relazione a questi tre fattori, si delineano, perciò, particolari forme distributive e particolari proprietà degli stimatori ed, in tale contesto, spesso, è necessario imporre talune condizioni non verificabili, prima fra tutte l’ipotesi della normalità nel senso della gaussianità della forma distributiva del fenomeno misurabile nella popolazione (al limite è richiesta la sola simmetria della distribuzione). Tuttavia, in molte situazioni ed, in genere, quando si tratta di un fenomeno qualitativo naturale od artificiale collegabile, secondo la logica consueta, ad un carattere metrico di taluni soggetti od oggetti (omogenei), l’ipotesi della gaussianità della forma distributiva del fenomeno quantitativo risulta fuorviante, in quanto è possibile osservare, invece, dopo la valutazione, come già detto, del fenomeno qualitativo, una forma grafica cuspidata asimmetrica, rispetto al massimo di frequenza, cui corrisponde la moda, dell’empirica distribuzione.

una qualità e, quindi, l'esame della variabilità, rispetto alla moda, di un insieme limitato chiuso da due misure estreme, o di soglia, di un carattere metrico collegabile, appunto, ad una o a più di una qualità. Perciò, appositi esperti devono censire e classificare la popolazione dei soggetti o degli oggetti di riferimento che, di solito, sono eterogenei, rispetto a presumibili qualità distintive valutabili in base all'esperienza e con l'ausilio della misura di un fenomeno quantitativo. In tal modo è possibile formare classi di soggetti o di oggetti omogenei, rispetto ad ognuna delle prestabilite qualità distintive.

L'anzidetto metodo dovrebbe, perciò, imporre nei suoi procedimenti particolari accorgimenti. È noto, infatti, che i procedimenti statistici sono in parte tecnici: tabelle, procedimenti matematici e via dicendo, che costituiscono la parte essenziale della metodologia, ed in parte, invece, sono procedimenti trascurati dagli statistici, a causa, forse, del loro carattere più soggettivo. Si tratta delle considerazioni logiche o, secondo le parole del Gini, degli avvedimenti necessari per ragionare sui fenomeni collettivi – che secondo gli statistici della scuola italiana sono i fenomeni qualitativi e quelli quantitativi – che si manifestano in una massa di osservazioni (Gini C. 1962. *La logica nella statistica*, ecc., *op.*, *cit.*, pag. 9), per cui sono detti anche fenomeni di massa. Inoltre, come ha sottolineato il Treloar (1951) in un volume dedicato all'analisi biometrica, la logica delle decisioni statistiche deve essere come una compagna essenziale della tecnica che ad essa conduce. Si ritiene, dunque, che per forgiare un linguaggio proprio del metodo statistico descrittivo è necessaria una terminologia che deve scaturire dalla logica consueta.

Perciò, avendo considerato che la variabilità di un fenomeno misurabile può essere in realtà imputata ad un imprecisabile numero di cosiddette cause accidentali e, soprattutto, ad una causa sistematica, regolare o normale, determinante ascrivibile ad un fenomeno qualitativo non misurabile, detto anche fenomeno descrittivo o collettivo, è stato introdotto (Gargano F. 1999., *op.*, *cit.*, pp. 187-188) l'immaginabile concetto di variabile "non casuale" o statistica semplice, potenzialmente continua in un intervallo limitato chiuso, detta anche variabile "di partenza" come ebbe a scrivere in passato uno statistico matematico (Brambilla F. 1968., pag. 9 e pag. 539). Tale concetto è da aggiungere a quello di fenomeno misurabile che, per lo schema di un'ipotesi, è da considerare una grandezza matematica variabile casuale gaussiana semplice continua in un intervallo aperto.

Quindi, per variabile “non casuale” o statistica semplice s’intende qualsiasi fenomeno misurabile collegabile, secondo la logica consueta, ad un fenomeno qualitativo non misurabile, detto anche fenomeno descrittivo o collettivo, che limita in realtà l’ordine delle misure di quello quantitativo, per cui è possibile l’esame del variabile stabile in realtà.

Le modalità di tale variabile sono le misure, distanti da quella più frequente, di un insieme limitato chiuso da due misure estreme, o di soglia (30), di un carattere metrico di determinati soggetti od oggetti omogenei, rispetto al fenomeno qualitativo valutabile come detto prima.

Capita, quindi, che un esperto – cui viene demandato, per ovvî motivi, il compito della diagnosi di prestabiliti fenomeni qualitativi – abbia da classificare la popolazione umana di riferimento ed avverta la necessità di basarsi sulla cosiddetta “anormalità statistica” per decidere se, in effetti, un soggetto di una classe da egli formata abbia veramente la qualità valutata (31). Ad esempio, nell’ambito delle indagini epidemiologiche, la decisione che un soggetto sia verosimilmente malato si basa, spesso, su tale concetto di anormalità statistica (di Iorio F. 1988., pp. 51-52) che bisogna, perciò, esprimere con termini da coniare secondo la logica consueta e, quindi, in base ad una teoria da cui deve scaturire la metodologia statistica descrittiva vera e propria.

In questa nota, perciò, si raccomanda il calcolo dei confini della “normalità statistica” dei casi che costituiscono una prestabilita quota percentuale (95%) di una classe di soggetti o di oggetti omogenei selezionati, in base alla diagnosi di un loro fenomeno qualitativo collegabile ad un fenomeno quantitativo, per stabilire la regolarità delle oscillazioni delle frequenze intorno al loro massimo cui corrisponde la moda del predetto fenomeno quantitativo.

(30) Come le misure di soglia glicemica stabilite nel 1997 per la diagnosi e la classificazione del diabete mellito, in base ai criteri esposti nella Tabella 3 contenuta in un rapporto della commissione degli esperti del NDDG che hanno modificato le misure di soglia glicemica stabilite dal NDDG nel 1979 e dall’OMS nel 1985.

(31) Come, ad es., in occasione di un’indagine epidemiologica, quando occorre effettuare la valutazione di uno stato di salute, per cui la comunità scientifica nazionale od internazionale di questo settore decide, prima del censimento e della classificazione della popolazione genetica di riferimento, che una misura di soglia di un carattere metrico segnali l’anormalità clinica. Inoltre, tale comunità stabilisce che la diagnosi che un individuo sia verosimilmente malato debba essere emessa in base allo screening medesimo. Però, talvolta, come già detto, tale decisione si basa sull’anormalità statistica, esistendo – come è noto – il problema degli individui border-line, cioè di quelli che risultano avere una misura del carattere metrico molto distante dalla moda.

Se si effettua l'inferenza induttiva con il processo di rilevamento delle informazioni statistiche per così dire completo sorge, perciò, la necessità di sottrarre dall'unità – che può rappresentare un insieme finito, od una classe, di soggetti o di oggetti omogenei – una prestabilita quota di frequenza relativa empirica integrale ($F = 0,95$) che rappresenta, secondo la logica consueta, una quota di maggioranza del totale dei soggetti o degli oggetti in questione caratterizzati da una qualità valutabile come detto prima. Quindi, è possibile ottenere la quantità $1 - F = 0,05$ che si può interpretare come numero relativo – espresso come frazione compresa nell'intervallo $[0,1]$ – dei soggetti o degli oggetti selezionati come detto prima che avrebbero misure di un loro carattere metrico perturbate probabilmente dagli errori casuali molto distanti dall'errore nullo. Per convenzione, allora, il valore percentuale della frequenza empirica integrale $F = 0,95$ può essere considerato un ottimo livello d'affidabilità o di sicurezza sulla qualità dei soggetti o degli oggetti (omogenei) che costituiscono un insieme finito, od una classe.

La distribuzione empirica di frequenza di un carattere metrico collegabile ad un fenomeno qualitativo non è altro, quindi, che la suddivisione di un insieme finito, o di una classe, di soggetti o di oggetti omogenei, in base alle misure del predetto carattere metrico che aiutano in realtà a valutare il fenomeno qualitativo. Occorre, quindi, il collegamento, da effettuare in base all'esperienza, tra i soggetti o gli oggetti caratterizzati da una qualità (funzione empirica) che limita l'ordine delle misure del fenomeno quantitativo e le misure di quest'ultimo fenomeno (argomento della funzione empirica) che aiutano alla diagnosi del fenomeno qualitativo (32).

Inoltre, sono necessarie alcune considerazioni sul termine “normale” che è stato usato, fino ad ora, nell'ambito della letteratura specifica, per definire la funzione di probabilità degli errori casuali di Gauss.

Il concetto di “normalità statistica” dei casi che possono scaturire da una diagnosi di un fenomeno qualitativo collegabile ad un feno-

(32) Invece, la distribuzione empirica di frequenza per così dire casuale od eventuale delle misure di un fenomeno quantitativo è la suddivisione del sottoinsieme dei soggetti o degli oggetti – selezionati dall'insieme dei soggetti o degli oggetti di riferimento in base al meccanismo di sorteggio casuale previsto da un prestabilito schema di probabilità dato a priori – secondo prestabilite classi del sottoinsieme delle misure del carattere metrico associabile, senza la ricerca di un'effettiva corrispondenza logica, ad un fenomeno qualitativo supposto un fenomeno casuale.

meno quantitativo appartiene, infatti, ad una teoria (33) da cui deve scaturire, come già detto, il metodo statistico descrittivo o investigativo-comparativo vero e proprio che presuppone la selezione di determinati soggetti od oggetti omogenei, rispetto ad una qualità naturale o artificiale valutabile in base all'esperienza e con l'ausilio dell'informazione statistica costituita, appunto, dalla misura di un fenomeno quantitativo collegabile, secondo logica consueta, a quello qualitativo.

Nell'ambito della terminologia statistica vigente, il termine "gaussianità", considerato sinonimo di "normalità statistica", allude alla distribuzione teorica di densità di probabilità delle misure perturbate esclusivamente dagli errori casuali che, in teoria, potrebbero essere anche illimitatamente grandi, di un carattere metrico considerato, per codesto schema di un'ipotesi, una grandezza astratta di un immaginario insieme infinito, detto universo ipotetico, di campioni casuali di soggetti o di oggetti. In realtà, però, la distribuzione teorica di densità di probabilità degli errori casuali standardizzati assumerebbe graficamente una forma campanulare simmetrica, oltre che stabilizzata o mesocurtica (34), se si potesse effettuare e non solo immaginare una serie, tendente ad infinito, di misurazioni di un carattere metrico associabile, senza la ricerca di un'effettiva corrispondenza logica, ad un fenomeno qualitativo supposto un fenomeno casuale.

Secondo il linguaggio proprio delle scienze sociali, comprese quelle economiche, e di quelle giuridiche, sanitarie, biologiche, antropolo-

(33) Che non è da confondere con la teoria della matematica e non può essere condizionata dalla teoria statistica della stima che sfrutta la teoria del calcolo delle probabilità ed i dettami della teoria dei campioni casuali dei soggetti o degli oggetti.

(34) Termine introdotto, come è noto, da K. Pearson. Tuttavia, taluni ritengono che non sia reale la capacità diagnostica degli indici di curtosi e di moderata asimmetria che il Pearson ha proposto per verificare, nell'ambito della ricerca empirica, la forma gaussiana mesocurtica (Bressan F. 1979., pp. 437-441). Secondo il Bressan, infatti, non esiste una relazione tra il valore di β_2 e la forma della distribuzione. Ciò è stato già evidenziato da Kaplansky (1945, pag. 259) e ribadito da Lombardo E. (1974, pp. 555-556). Inoltre, il Bressan ha voluto precisare nel suo articolo, in corrispondenza del punto ordinato con la lettera "F", che non esiste una relazione tra mesocurtosi e gaussianità. Cioè, esistono delle distribuzioni simmetriche e con $\beta_2 = 3$ molto dissimili dalla distribuzione gaussiana detta normale. Inoltre, non è stato ancora dimostrato che $\beta_2 = 3$ e $\gamma_1 = 0$ comportino, sempre, rispettivamente la normalità (gaussianità o campanularità) e la simmetria della distribuzione. Infatti, alcuni autori (Mood e Graybill, 1963., pp. 109-110) hanno sottolineato che esistono distribuzioni asimmetriche e con $\gamma_1 = 0$. Però, essi non hanno definito l'espressione analitica delle distribuzioni medesime. Infine, il Bressan ha sottolineato che sarebbe poco adeguato l'utilizzo dei termini "indice di normalità" e "indice di simmetria", in quanto non esiste una relazione stretta tra ciò che si è indicato e ciò che si vorrebbe indicare.

giche, fisiche (35), della medicina costituzionalistica (36), ecc., l'aggettivo "normale" significa conforme alla norma (37). Ciò che tende alla stabilizzazione (38). Inoltre, ciò che è conforme alla legge, ad una regola generale o valida per tutti, od, anche, conforme al canone, al precetto, all'ordine, alla direttiva, all'indicazione. È regolare o normale, quindi, tutto ciò che può essere standardizzato perché è stabile. Sono normali le cose, le opinioni, i comportamenti, le situazioni, le persone, gli individui in generale, che corrispondono ai criteri di valutazione o alle aspettative dei più. Mentre il termine contrario, cioè l'aggettivo "anormale", irregolare, inconsueto, insolito, non comune, eccezionale, espresso secondo il linguaggio statistico attuale con i termini: iponor-

(35) In fisica, l'accelerazione normale della gravità è il valore convenzionale dell'accelerazione di gravità a 45° di latitudine al livello del mare.

(36) Dove il sostantivo "tipo", oppure il sostantivo composto da *normo* e *tipo* (*normotipo*) significa il soggetto umano caratterizzato da una proporzione regolare tra varie misure di pre-stabilite parti del suo corpo.

(37) In generale, la norma od il modello, detto anche tipo, è l'esemplare di confronto per una misurazione. In etica, invece, la norma è la regola per un giusto agire. Quindi, una norma etica o morale è qualsiasi proposizione generale che ingiunge un certo tipo di azione in un certo tipo di situazione. Ad esempio: "si devono mantenere le promesse fatte". In estetica, invece, è il modello per giudicare la bellezza o l'arte. Nella logica, è la regola per un'inferenza valida. Inoltre, la norma di fede è in generale l'asserzione autoritaria di una credenza ed, in particolare, nel cattolicesimo comprende l'intero insegnamento apostolico e le sue ulteriori elaborazioni. La norma, quindi, è un sostantivo per indicare un singolo precetto morale, giuridico, tecnico. In genere si parla di norma dell'agire o, rispettivamente, di quella vigente, oppure del bello scrivere riferibile ad una formulazione imperativa determinata.

In origine, la norma fu un termine che indicava la "squadra geometrica predisposta per tracciare una linea perpendicolare", utile per la misurazione o realizzazione di un angolo retto. Da qui è scaturito il principale significato di regola o di modello da seguire ed osservare. Nell'economia industriale, la norma o tipo è la regola della produzione standardizzata, cioè la produzione organizzata secondo il principio della normalizzazione o della generale convenienza. In matematica, la norma di un numero complesso è, invece, il quadrato del suo modulo. Secondo la teoria da cui scaturisce l'attuale metodo statistico della stima, per norma, detta anche *moda* o, in versione inglese, "standard", s'intende la misura di un carattere metrico da osservare il massimo numero di volte nell'ambito della seriazione o distribuzione empirica di frequenza, per così dire casuale od eventuale, di un fenomeno quantitativo di certi soggetti od oggetti da scegliere con un meccanismo di sorteggio casuale. Invece, secondo la teoria da cui deve scaturire il metodo statistico per effettuare l'esame della variabilità di un fenomeno quantitativo collegabile ad un fenomeno qualitativo o descrittivo e, quindi, secondo l'inferenza induttiva con il processo di rilevamento delle informazioni statistiche per così dire completo riferito, cioè, ad un insieme finito, o ad una classe, di soggetti o di oggetti omogenei, rispetto ad un fenomeno qualitativo valutabile in base all'esperienza e con l'ausilio della misurazione di un fenomeno quantitativo, la norma è la misura del carattere metrico da osservare il massimo numero di volte nell'ambito dei soggetti o degli oggetti da selezionare mediante la diagnosi, come detto prima, del fenomeno qualitativo.

(38) Secondo la scuola economica classica, il valore normale o naturale dei beni, contrapposto al valore corrente, è ciò che tende alla stabilizzazione.

male, ipernormale, è tutto ciò che, in effetti, è al di fuori dei confini della regolarità o normalità delle oscillazioni delle frequenze delle misure, distanti dalla norma e, quindi, da interpretare con incertezza o sospetto. A sua volta, il sostantivo “normalità” è la condizione, o situazione, riconducibile alla consuetudine o all’ordine preesistente, oppure alla generalità interpretabile come “regolarità”. Quindi, la condizione o situazione in cui le cose sono normali, o regolari (Gabrielli A. 1967. Oppure consulta: Devoto G. Oli G. C. 1991). Ad ogni modo, poiché lo “standard” o la norma di un fenomeno quantitativo presuppone una sola qualità naturale od artificiale dei soggetti o degli oggetti ai quali essa è da riferire, anche la normalità o regolarità delle oscillazioni delle frequenze, intorno al loro massimo cui corrisponde la norma del predetto fenomeno quantitativo, evoca l’idea di una qualità provocata dalla natura o prodotta dall’uomo.

Se ci si riferisce all’unità biologica, caso mai ancora esista, come la razza umana, la specie animale o la varietà vegetale, anziché a tutto il genere umano, animale, vegetale o batterico, e si usa il termine “normalità” bisogna pensare, quindi, all’omogeneità, rispetto alla razza, specie o varietà, anziché alla generalità, degli individui. Lo stesso discorso vale se ci si riferisce allo stato epidemico per una specifica malattia degli individui esistenti in realtà, in un determinato tempo e spazio, oppure agli oggetti caratterizzati da una qualità artificiale prodotta dall’industria. Quindi, nelle indagini statistiche sul controllo della qualità della produzione industriale, il termine “normalità” è da interpretare come livello di sicurezza o d’affidabilità sulla qualità che accomuna determinati oggetti che, quindi, sono omogenei, rispetto alla qualità valutabile come detto prima e, perciò, costituiscono un insieme finito, od una classe.

In Biometria, oppure in Epidemiologia od, infine, nell’ambito della Statistica industriale e del controllo della qualità, il termine “normalità statistica” non può fare riferimento – perciò – alla generalità dei soggetti o degli oggetti esistenti in realtà che costituiscono, di solito, un insieme finito di soggetti o di oggetti eterogenei, rispetto alle loro qualità intrinseche, ma ai soggetti o agli oggetti caratterizzati, rispettivamente, da una razza, specie o varietà, oppure da una malattia epidemica, od, infine, da una qualità artificiale prodotta dall’industria (39).

(39) Nell’ambito del controllo della qualità degli oggetti da distruggere per sottoporli a verifica, come il controllo sulla resistenza di taluni materiali ed, in particolare, il controllo sulla durata delle lampadine, ecc., non è conveniente il controllo dell’intera produzione industriale di

Il concetto di “normalità statistica” deve essere desunto, quindi, in base alla logica consueta. Cioè deve significare la prevalenza dei soggetti o degli oggetti caratterizzati da una qualità, valutabile come detto prima, che può essere rara o molto frequente come, ad es., in ambito epidemiologico, il buono stato di salute. Tuttavia, tra le affermazioni finora espresse dagli statistici, è da registrare che la distribuzione della statura è normale, nel senso che essa segue la curva di Gauss (Gini C. Trezza A. 1962. *Anormalità delle distribuzioni statistiche*, ecc.). Però s'intuisce, dal titolo della ricerca pubblicata da Gini e coll., che il termine “anormalità” vuol dire “non gaussianità” nel senso di una distribuzione empirica di frequenza non casuale di forma campanulare asimmetrica. Tutto ciò, però, senza che sia stato precisato in quale insieme finito, o classe, di soggetti o di oggetti omogenei, rispetto ad una qualità non misurabile più o meno frequente come una razza, valutabile come detto prima, sia stato riscontrato questo aspetto non campanulare ed asimmetrico piuttosto uniforme della distribuzione empirica di frequenza delle misure del fenomeno quantitativo esaminato.

Secondo la logica consueta – che deve essere, secondo il Gini, la logica da usare nella statistica – sarebbe opportuno intendere per “normalità statistica” di determinati soggetti od oggetti omogenei, rispetto ad una qualità più o meno frequente, oppure rara, valutabile come detto prima, la regolarità delle oscillazioni delle frequenze concentrate intorno al loro massimo cui corrisponde la moda del carattere metrico collegabile al fenomeno qualitativo. Inoltre, per convenzione, la regolarità delle oscillazioni delle frequenze dell'anzidetta distribuzione empirica di frequenza non casuale di forma unimodale asimmetrica sarebbe quella riscontrabile in una prestabilita quota di maggioranza dell'insieme dei soggetti o degli oggetti omogenei enumerati in occasione della valutazione, come detto prima, del fenomeno qualitativo.

Perciò, è necessario il parametro che indichi la regola, norma, moda o “standard”, di un fenomeno misurabile, collegabile ad un fenomeno qualitativo, potenzialmente continuo in un intervallo limitato chiuso, o compatto. La lunghezza di tale intervallo è utile per calcolare la reale stabilità della variabilità, rispetto alla norma, del predetto fenomeno quantitativo. Un fenomeno misurabile è da considerare, quindi,

tali oggetti (Girone G. Salvemini T. 1983., pag. 34). In tal caso, infatti, è necessaria una rilevazione parziale o campionaria degli oggetti supposti omogenei che costituiscono una classe, in modo da controllare la loro qualità programmata dalla produzione industriale.

variabile per una causa sistematica determinante detta anche variabile “non casuale” o statistica semplice ed occorre, perciò, il parametro che indichi la reale stabilità della variabilità, rispetto alla norma, di un fenomeno quantitativo imputabile ad un fenomeno qualitativo o causa sistematica difficilmente separabile da un imprecisabile numero di cosiddette cause accidentali che si traducono, in concreto, in un altrettanto numero imprecisabile di errori casuali semplici con alternanza del segno più o meno, ma non d’illimitata intensità, che possono essere commessi da chi valuta il fenomeno qualitativo, in base all’esperienza e con l’ausilio della misura del fenomeno quantitativo.

La conoscenza di una regola o norma statistica e delle regolari o normali oscillazioni delle frequenze intorno al loro massimo cui corrisponde la norma di un carattere metrico collegabile ad un fenomeno qualitativo presuppone, quindi, il riferimento ad un immaginabile insieme limitato chiuso, o compatto, da due misure estreme, o di soglia, ed il calcolo dell’indice di variabilità, rispetto alla misura più frequente, di codesto insieme, anziché la stima della media aritmetica e dello scarto quadratico medio non determinabili e, quindi, incogniti, delle teoriche misure di un immaginario insieme aperto che scaturisce dall’ipotesi che un fenomeno quantitativo sia una grandezza astratta di un immaginario insieme infinito di campioni casuali di soggetti o di oggetti in realtà da non osservare. Per scopi statistici descrittivi, quindi, è necessario fare riferimento ad un insieme finito, o ad una classe, di soggetti o di oggetti omogenei, rispetto ad un fenomeno qualitativo naturale od artificiale valutabile come detto prima, anziché ad un sottoinsieme di soggetti o di oggetti selezionati mediante il meccanismo del sorteggio casuale previsto da un prestabilito schema di probabilità dato a priori.

Perciò, l’esame della variabilità, rispetto alla moda, di un fenomeno quantitativo collegabile ad un fenomeno qualitativo o descrittivo, più o meno frequente, presuppone che si abbia un’informazione completa, anziché la sola informazione statistica (40). Cioè è neces-

(40) Taluni distinguono l’inferenza descrittiva dall’inferenza analitica. La prima sarebbe l’inferenza sui parametri – o sulle costanti caratteristiche – della popolazione reale (finita). Ossia, le tecniche di stima necessarie per assegnare valori approssimati ai parametri, come la media, oppure il totale della popolazione stessa, sfruttando le informazioni contenute nel campione osservato. Invece, l’inferenza analitica estenderebbe l’analisi anche alle relazioni statistiche e, quindi, tenderebbe non già a “conoscere” le caratteristiche della popolazione reale sotto indagine, ma a spiegare i fenomeni (qualitativi) esaminati. Secondo Frosini e coll. (1994, pp. 14-15), Cicchitelli

saria la diagnosi del fenomeno qualitativo da effettuare come detto prima. In tal caso, l'analisi statistica è da riferire all'insieme finito, o alla classe, dei soggetti o degli oggetti omogenei, rispetto alla qualità da valutare (41). Perciò, in questo contesto il termine "normalità statistica" non è sinonimo del termine probabilistico "gaussianità" che, nell'ambito della teoria statistica della stima, equivale, invece, come già detto, all'assunzione dell'ipotesi che un fenomeno qualitativo naturale, od artificiale, sia un fenomeno casuale, per cui è necessario che i soggetti o gli oggetti da osservare siano selezionati mediante il meccanismo di sorteggio casuale previsto da uno schema di probabilità dato a priori (42).

È da dire, inoltre, che normalizzare significa ricondurre ad una condizione di normalità o di regolarità, oppure all'ordine consueto. In matematica, significa rendere normali certi elementi. Si normalizzano, ad esempio, due o più numeri reali dividendoli per la radice della somma dei loro quadrati, in modo che la somma dei quadrati dei nuovi numeri sia uguale ad uno. Nell'industria, invece, il verbo normalizzare è sinonimo di standardizzare che, come già detto, significa organizzare la produzione secondo il principio della generale convenienza.

e coll. (1992, op., cit., pp. 44-45 e pag. 52), il confine tra i due tipi di inferenza è assai tenue per l'assenza di modelli riguardanti la popolazione da cui è estratto il campione. Invece, se si ipotizza l'esistenza di appropriati modelli e si vogliono analizzare certi fenomeni quantitativi con predefiniti livelli di significatività, si è pienamente negli ambiti dell'inferenza analitica.

(41) In uno screening epidemiologico, il termine "normale" è usato giustamente per qualificare un soggetto senza segni né sintomi di malattia che è, appunto, ai nostri giorni, lo stato di salute prevalente della gente. Inoltre, secondo la teoria queteletiana dell'uomo medio aritmetico, la ricerca del tipo umano ideale è da riferire all'uomo in cui si può riscontrare la maggior parte delle qualità di tutti gli uomini e, quindi, secondo la logica consueta, l'uomo molto frequente detto anche normale.

(42) Nell'ambito della teoria da cui deve scaturire il metodo statistico descrittivo vero e proprio, la curva "normale" degli errori casuali non è da considerare una curva regolare, dato che la predetta teoria presuppone una curva che rappresenti la reale stabilità della variabilità, rispetto alla moda, di un insieme limitato chiuso da due misure estreme, o di soglia, di un fenomeno quantitativo collegabile ad un fenomeno qualitativo valutabile come detto prima. Inoltre, come già detto, la variabilità per una causa sistematica difficilmente può essere separata dalla variabilità per un imprecisabile numero di cause cosiddette accidentali che, in concreto, si traducono in un altrettanto numero imprecisabile di errori casuali di misura del fenomeno quantitativo da esaminare, commessi da chi valuta, come detto prima, il fenomeno qualitativo.

Come è noto, nell'ambito della teoria statistica della stima, il termine "curva normale" è giustificato dalla persuasione, non del tutto corretta, che i fenomeni fisici e biologici si distribuiscono regolarmente secondo la curva gaussiana. Inoltre, la denominazione di curva degli errori accidentali deriva dal fatto che si può ipotizzare che essa rappresenti la distribuzione degli errori accidentali commessi nel misurare ripetute volte una grandezza (Leti G. 1983., op. cit., pag. 588).

Secondo la teoria da cui deve scaturire il metodo statistico descrittivo vero e proprio, il verbo normalizzare o standardizzare deve significare ridurre o dividere la cosiddetta variabile potenzialmente continua “non casuale” o statistica semplice rispetto all’indice della deviazione dalla norma della distribuzione empirica unimodale (asimmetrica) di frequenza della variabile medesima e, quindi, ridurre la lunghezza dell’intervallo limitato chiuso o parte di codesto intervallo in cui è definita tale variabile.

Quindi, vale la pena sottolineare che:

a) la variabile per una causa sistematica determinante, detta anche variabile “non casuale”, “scarto dalla moda” o variabile statistica semplice, è un fenomeno quantitativo collegabile ad un fenomeno qualitativo, detto anche fenomeno collettivo o descrittivo, che limita in realtà l’ordine delle misure di quello quantitativo. Perciò, talvolta, un fenomeno misurabile è da considerare potenzialmente continuo in un intervallo limitato chiuso da due misure estreme, o di soglia, da prestabilire in base all’esperienza e collegare a taluni di un insieme finito, o di una classe, di soggetti o di oggetti omogenei, rispetto ad una qualità valutabile in base all’esperienza e con l’ausilio della misura del predetto carattere metrico.

b) Il calcolo della standardizzazione, o normalizzazione, dell’anzidetta variabile e del suo intervallo limitato chiuso è utile per confrontare, in un determinato tempo e spazi diversi, oppure in un determinato spazio e tempi diversi, la reale stabilità della variabilità, rispetto alla norma, di un fenomeno quantitativo collegabile ad un fenomeno qualitativo, valutabile come detto prima.

L’immaginabile standardizzazione della variabile in questione (43) può essere effettuata rapportando le misure, distanti dalla moda, del carattere metrico di determinati soggetti od oggetti omogenei, rispetto ad un fenomeno qualitativo valutabile come detto prima, diviso l’indice approssimato della deviazione dalla moda del carattere metrico dei sog-

(43) Invece, l’immaginaria standardizzazione o normalizzazione della variabile casuale gaussiana semplice, continua in un intervallo aperto e, quindi, di lunghezza non limitata, è utile per l’analisi della stabilità della variabilità, rispetto alla teorica media aritmetica, delle misure di un fenomeno quantitativo considerato, per lo schema di un’ipotesi, una grandezza matematica. Cioè, il rapporto tra le teoriche misure, distanti dalla media aritmetica non determinabile e, perciò, incognita, diviso l’indice della deviazione dalla media aritmetica incognita, o scarto quadratico medio non determinabile e, quindi, pur esso incognito della predetta grandezza dell’immaginario collettivo infinito dei campioni casuali dei soggetti o degli oggetti in realtà da non osservare.

getti o degli oggetti medesimi da selezionare, quindi, in base ad una diagnosi o proposizione del fenomeno qualitativo.

Pertanto, sia il criterio statistico proposto per confermare l'omogeneità, rispetto ad una qualità, valutabile come detto prima, di determinati soggetti od oggetti che costituiscono un insieme finito, od una classe, sia il criterio per determinare i confini convenzionali della normalità o regolarità delle oscillazioni delle frequenze intorno al loro massimo cui corrisponde la moda di un carattere metrico collegabile alla qualità dei soggetti o degli oggetti medesimi, non scaturiscono dalla teoria statistica della stima. Sono, invece, dei criteri che presuppongono il calcolo di due parametri caratteristici della distribuzione empirica unimodale (asimmetrica) di frequenza dell'anzidetta variabile potenzialmente continua "non casuale" o statistica semplice, da calcolare dopo la valutazione, come detto prima, di un fenomeno qualitativo o descrittivo.

Il linguaggio usato in questa nota non deriva, dunque, dalla teoria statistica della stima e serve per commentare l'inferenza induttiva con il processo di rilevamento delle informazioni statistiche per così dire completo riferito, cioè, ad un insieme finito, o ad una classe, di soggetti o di oggetti omogenei. Quindi, sarebbe il linguaggio che appartiene alla teoria da cui scaturirebbe il metodo statistico descrittivo o investigativo-comparativo vero e proprio, poiché solo i fenomeni qualitativi non misurabili sono considerati, nell'ambito delle scienze naturali, fenomeni descrittivi o collettivi e possono, in realtà, limitare l'ordine delle misure di un fenomeno quantitativo.

7. Conclusioni

I fenomeni provocati dalla natura o prodotti artificialmente dall'uomo, ai quali si riferisce la ricerca empirica ed in genere la ricerca scientifica, non sono misurabili ma valutabili, essendo fenomeni qualitativi, detti anche fenomeni descrittivi (44) o collettivi.

(44) Come è noto, nell'ambito delle scienze naturali, i fenomeni descrittivi sono quelli che riguardano particolari esteriori o superficiali e, quindi, le forme organizzate in sistema (aspetti morfologici) dei soggetti o degli oggetti esistenti in realtà. Inoltre, il termine "morfologia", nell'ambito delle scienze biologiche che per prime lo hanno adottato, significa lo studio delle strutture degli organismi viventi e delle loro parti.

Inoltre, nell'ambito delle scienze della natura e di quelle umane, la ricerca empirica non può prescindere dal considerare che un fenomeno qualitativo:

1) Condiziona sempre, in realtà, la scelta, in un determinato tempo e spazio, di determinati soggetti od oggetti omogenei, rispetto alla predetta qualità naturale od artificiale che unifica o identifica i soggetti o gli oggetti medesimi, la cui entità è sempre un numero che può stabilire un esperto, anziché una collezione di soggetti o di oggetti che, secondo la logica matematica, sono quelli che in realtà possono avere qualità anche diverse e costituire un sottoinsieme o campione di un'immaginario insieme infinito, detto universo ipotetico, oppure quelli che possono avere qualità anche diverse e costituire un immaginabile insieme finito, detto universo non ipotetico. Quindi, in realtà, è logico ammassare o cumulare determinati soggetti od oggetti omogenei che costituiscono un insieme finito, od una classe.

2) Può essere, in realtà, molto frequente, oppure poco frequente o raro.

3) Talvolta può essere collegabile, secondo la logica consueta, ad un fenomeno quantitativo e, quindi, in tal caso, il fenomeno qualitativo limita in realtà l'ordine delle misure di quello quantitativo.

Per conseguenza, un fenomeno qualitativo può determinare la reale stabilità della lunghezza dell'intervallo limitato chiuso, o compatto, da due misure estreme, o di soglia, di un fenomeno quantitativo collegabile a quello qualitativo. Le anzidette misure estreme, o di soglia, devono essere prestabilite da una comunità nazionale od internazionale di esperti per essere collegate a taluni di un insieme finito, o di una classe, di soggetti o di oggetti omogenei, rispetto al fenomeno qualitativo valutabile come detto prima.

Se un fenomeno qualitativo limita l'ordine delle misure di un fenomeno quantitativo, non bisogna alludere, perciò, alla stabilità di variabilità imputabile ad un'immaginaria serie, tendente ad infinito, di cause accidentali, riconducibile alla legge empirica del caso, nota anche come legge dei grandi numeri. In tal caso, non serve un'indagine statistica campionaria di notevoli dimensioni che presuppone il ricorso ad un modello probabilistico adattativo, qual è la funzione di densità di probabilità della variabile casuale gaussiana standardizzata continua in un intervallo aperto, per sostenere l'idea immaginaria della stabilità di variabilità, rispetto alla teorica media aritmetica del predetto modello dato a priori, la cui funzione di ripartizione rappre-

senta un immaginario insieme infinito di campioni casuali di soggetti o di oggetti distribuiti, in teoria, secondo le misure di un fenomeno quantitativo, considerato una grandezza matematica, perturbate probabilmente dagli errori casuali standardizzati. Quindi, per fini statistici descrittivi, non serve un modello probabilistico perché esso non evoca l'idea del variabile stabile in realtà, ma un'immaginaria stabilità di variabilità.

Nell'ambito delle scienze naturali ed umane, la ricerca empirica, da non confondere con la ricerca scientifica (45), deve fare riferimento ad una popolazione finita di soggetti naturali o di unità materiali od oggetti esistenti in realtà che, spesso, sono eterogenei, rispetto a presumibili qualità distintive. Quindi, per sostenere l'idea della reale stabilità della variabilità, rispetto alla moda, delle empiriche misure di un carattere metrico collegabile ad un fenomeno qualitativo naturale od artificiale che limita, in realtà, l'ordine delle misure di quello quantitativo è necessario censire e classificare le popolazioni genetiche, o le popolazioni delle unità materiali, di riferimento.

Se taluni fenomeni qualitativi sono, perciò, valutabili in base all'esperienza e con l'ausilio della misura di un fenomeno quantitativo collegabile, secondo la logica consueta, alle predette qualità distintive può essere usato il cosiddetto modello statistico descrittivo (46) per

(45) Secondo l'opinione di molti (Fabris L. 1997., pag. 1), fare ricerca scientifica significa investigare il fenomeno (qualitativo) che interessa seguendo il metodo che, nell'accezione ormai consueta, consiste nel:

- raccogliere, ordinare, collegare informazioni sul predetto fenomeno;
- formulare precise ipotesi di ricerca;
- vagliare le ipotesi ponendole a confronto con la realtà e, pertanto, concludere, sulla base di un'indagine campionaria. Cioè, effettuare l'inferenza induttiva con il processo di rilevamento delle informazioni statistiche per così dire incompleto, in un precisato contesto sperimentale, accettando o rifiutando le ipotesi stesse. Secondo il nostro punto di vista, invece, fare ricerca scientifica o valutare esattamente un fenomeno qualitativo significa usare l'esperienza e, quindi, occorre la conoscenza del fenomeno qualitativo che scaturisce da un'immaginabile empirica serie (finita) di osservazioni su determinati soggetti od oggetti caratterizzati dal predetto fenomeno, anziché fare l'ipotesi e, quindi, l'esperimento che il fenomeno qualitativo sia un fenomeno accidentale od evento di un'immaginaria serie, tendente ad infinito, per cui occorre effettuare la prova, mediante l'indagine statistica campionaria, per verificare tale ipotesi.

(46) Invece, qualsiasi indagine statistica che si avvale di uno schema probabilistico o di un'ipotesi ha come scopo quello di esemplificare una popolazione finita di soggetti o di oggetti, aventi qualità anche diverse, oppure una popolazione infinita di soggetti o di oggetti in realtà da non osservare. In tal caso, la popolazione di riferimento, detta anche popolazione d'origine, finisce per essere quella a cui bisogna riferire i risultati dell'indagine effettuata con il metodo statistico condizionato dal predetto schema dato a priori che non presuppone, dunque, alcun procedimento classificatorio (Fabris L. 1997., *op. cit.*, pp. 5-8).

esemplificare il numero delle frequenze enumerate dopo la valutazione, come detto prima, di un fenomeno qualitativo.

In realtà, quindi, per stabilire qual è la moda è necessario considerare la cosiddetta popolazione statistica o l'insieme limitato chiuso, o compatto, da due misure estreme, o di soglia, di un fenomeno quantitativo, il cui ordine delle misure è limitato da un fenomeno qualitativo. Le misure che costituiscono la popolazione statistica sono, perciò, quelle che aiutano un esperto alla valutazione di un fenomeno qualitativo che caratterizza determinati soggetti od oggetti (omogenei), la cui entità è sempre un determinato numero. Come già detto, tali misure possono essere ordinate come una successione crescente (o decrescente) ed anche non decrescente, o poste minori od uguali rispetto ad esse stesse cui corrispondono i percentili empirici (47).

Come è noto, nell'ambito delle scienze umane, il Quetelet ha enunciato la teoria dell'uomo medio aritmetico per valutare due razze diverse, come i lapponi di alta e di bassa statura, ed ha avvertito, quindi, per primo, la necessità di escogitare criterî obiettivi per la ricerca dei tipi in un determinato tempo e spazio, al fine d'identificare gruppi di soggetti o di oggetti omogenei, rispetto a prestabiliti fenomeni qualitativi, descrittivi o collettivi non misurabili, come le razze, le specie o le varietà. Ciò acclara, quindi, la necessità di dover sempre censire e classificare, rispetto a prestabilite qualità distintive, l'insieme dei soggetti o degli oggetti di riferimento, a causa della loro diversità.

L'inferenza induttiva da effettuare con il processo di rilevamento delle informazioni statistiche per così dire completo riferito, cioè, ad un insieme finito, o ad una classe, di soggetti o di oggetti omogenei, rispetto ad una qualità valutabile in base all'esperienza e con l'ausilio della misura di un fenomeno quantitativo collegabile a quello qualitativo, può essere utile per confermare i risultati della ricerca scientifica e, perciò, la metodologia statistica in genere non è da considerare una

(47) L'esame statistico della variabilità, rispetto alla moda, della distribuzione empirica unimodale (asimmetrica) di frequenza della variabile "non casuale" o statistica semplice, potenzialmente continua in un intervallo limitato chiuso, quindi, induce ad ordinare le misure di un carattere metrico collegabile ad un fenomeno qualitativo come una successione crescente (o decrescente) ed anche non decrescente poiché, secondo la logica consueta, determinati soggetti od oggetti omogenei, rispetto ad una qualità valutabile in base all'esperienza e con l'ausilio della misura di un fenomeno quantitativo collegabile a quello qualitativo, possono essere ammassati o cumulati, a partire da quelli con la misura estrema inferiore ed arrivare a quelli con la misura estrema superiore, o misure di soglia, prestabilite in base all'esperienza.

scienza. Però, se si vuole confermare un risultato scientifico o l'esatta valutazione di un fenomeno qualitativo, il metodo statistico idoneo è quello descrittivo, piuttosto che quello che scaturisce dalla teoria statistica della stima.

Si sottolinea, inoltre, che la distribuzione empirica di frequenza delle misure di un fenomeno quantitativo collegabile ad un fenomeno qualitativo è priva di distribuzione teorica di densità di probabilità data a priori ed è libera dall'ipotesi distributiva gaussiana mesocurtica, poiché è caratterizzata dalla reale stabilità di variabilità, rispetto alla moda, del fenomeno quantitativo collegabile a quello qualitativo. Tale distribuzione empirica presuppone, quindi, la selezione dei soggetti o degli oggetti in base ad una proposizione o diagnosi di un esperto di un fenomeno qualitativo. Perciò, la stabilità di variabilità, rispetto alla moda, di un fenomeno quantitativo è da imputare sempre ad una causa sistematica, regolare o normale, determinante più o meno frequente, oppure rara, ascrivibile ad una qualità valutabile come detto prima.

Con riferimento alla distribuzione empirica unimodale (asimmetrica) di frequenza della cosiddetta variabile "non casuale" o statistica semplice potenzialmente continua in un intervallo limitato chiuso ed alla corrispondente distribuzione cumulativa di frequenza si propone, dunque, il criterio per confermare l'omogeneità di determinati soggetti od oggetti che costituiscono in realtà un insieme finito, od una classe, ed il criterio per determinare i confini convenzionali della normalità o regolarità delle oscillazioni delle frequenze delle misure del carattere metrico collegabile alla qualità dei soggetti o degli oggetti medesimi (48).

Tali criteri statistici presuppongono il calcolo di due parametri caratteristici della distribuzione empirica di frequenza dell'anzidetta variabile, cioè:

a) La moda, norma o "standard", diversa dall'empirica media aritmetica, di un insieme limitato chiuso da due misure estreme, o di soglia, di un fenomeno quantitativo collegabile ad un fenomeno qualitativo, prestabilite da una comunità di esperti del fenomeno qualitativo valutabile come detto prima. Ossia, il parametro che, secondo la logica consueta, si presta ad essere considerato la fonte di verità o l'origine

(48) Ad esempio, in occasione di un'indagine epidemiologica e per scopi di medicina preventiva è necessario stabilire le estreme frequenze regolari o normali delle misure di molti costituenti del sangue e delle urine degli esseri viventi da diagnosticare "di buona salute".

della variabilità delle misure del fenomeno quantitativo collegabile a quello qualitativo. Quindi, tale parametro è una logica sintesi delle misure che costituiscono il predetto insieme compatto.

b) L'indice approssimato della deviazione dalla moda della distribuzione empirica unimodale (asimmetrica) di frequenza dell'anzidetta variabile "non casuale" o statistica semplice potenzialmente continua in un intervallo limitato chiuso, necessario per effettuare il calcolo della reale stabilità di variabilità rispetto alla moda.

La struttura di tale indice di variabilità assoluta per così dire totale, tiene conto, tra l'altro, della lunghezza, in realtà stabile, dell'anzidetto intervallo, detto anche intervallo totale di rilevazione. Tale lunghezza può essere ottenuta per differenza tra la misura estrema superiore e quella estrema inferiore, o misure di soglia, poste minori od uguali rispetto ad esse stesse, di un fenomeno quantitativo collegabile ad un fenomeno qualitativo. Cioè, in base alla differenza tra le misure estreme corrispondenti ai due percentili empirici estremi della distribuzione empirica unimodale (asimmetrica) di frequenza cumulata dell'anzidetta variabile "non casuale" o statistica semplice.

L'indice di variabilità, rispetto alla norma, consente di normalizzare o standardizzare per fini comparativi – in un determinato tempo e spazi diversi, oppure in un determinato spazio e tempi diversi – la variabile "non casuale" o statistica semplice ed anche la lunghezza o misura, in realtà stabile, dell'intervallo limitato chiuso dalle misure estreme, o di soglia, della variabile in questione. Il criterio per standardizzare la lunghezza dell'intervallo totale di rilevazione ed anche la lunghezza di un intervallo per così dire parziale o "range" interpercentile limitato da due misure posizionate a cavallo ma non necessariamente equidistanti dalla moda, poste minori od uguali rispetto ad esse stesse, cui corrispondono due percentili empirici fra cui comprendere un prestabilito livello percentuale (95%) della somma integrale delle frequenze delle misure della distribuzione empirica unimodale (asimmetrica) di frequenza della predetta variabile, è diverso dal criterio proposto per standardizzare la variabile continua casuale gaussiana semplice ed anche la lunghezza dell'intervallo aperto, detto campo, o di una parte dell'intervallo di codesta variabile casuale.

La conoscenza di una regola o norma statistica e dei confini delle oscillazioni regolari o normali delle frequenze delle misure distanti dalla norma di un fenomeno quantitativo collegabile ad un fenomeno qualitativo non è, quindi, definitiva se è fondata sull'empirica media

aritmetica e sull'empirico scarto quadratico medio che sono stime dei corrispondenti parametri delle teoriche misure di un fenomeno quantitativo considerato, per lo schema di un'ipotesi, una grandezza astratta e, quindi, una variabile casuale gaussiana semplice standardizzata associabile – senza la ricerca di un'effettiva corrispondenza logica – ad un fenomeno qualitativo supposto un fenomeno casuale.

Una conclusione ottenuta senza la conoscenza della distribuzione empirica dell'anzidetta variabile “non casuale” o statistica semplice è, perciò, riconducibile a cause che si possono considerare, per lo schema di un'ipotesi, solo accidentali.

I criteri statistici raccomandati in questa nota presuppongono, dunque, l'uso dei percentili empirici della distribuzione univariata unimodale (asimmetrica) di frequenza cumulata delle misure di un carattere metrico collegabile ad un fenomeno qualitativo o descrittivo.

In passato, diversi cultori di chimica-clinica (Reed e coll. 1971., pp. 275-284) hanno usato i percentili empirici per effettuare la stima statistica della normalità della variabilità delle misure dei costituenti del sangue e delle urine e, quindi, per effettuare l'inferenza induttiva con il processo di rilevamento delle informazioni statistiche per così dire incompleto o riferito ad un sottoinsieme di soggetti o di oggetti scelti con un meccanismo di sorteggio casuale, secondo un prestabilito schema di probabilità dato a priori.

Pertanto, occorre distinguere:

a) I percentili empirici cui corrispondono le misure, poste minori od uguali rispetto ad esse stesse, di un carattere misurabile di un certo numero di soggetti o di oggetti da selezionare in base al meccanismo del sorteggio casuale previsto dallo schema di probabilità degli errori casuali standardizzati. Tali soggetti od oggetti possono avere in realtà qualità anche diverse e possono costituire un sottoinsieme o campione di un immaginario insieme infinito di campioni casuali. Inoltre, il fenomeno quantitativo da esaminare è associabile, senza la ricerca di un'effettiva corrispondenza logica, ad un fenomeno qualitativo supposto un fenomeno casuale.

b) I percentili empirici di un carattere misurabile di determinati soggetti od oggetti omogenei, rispetto ad un fenomeno qualitativo valutabile come detto prima, che costituiscono un insieme finito, od una classe. In tal caso, i soggetti o gli oggetti o, in genere, le unità statistiche sono da selezionare in base alla diagnosi del predetto fenomeno qualitativo che limita, in realtà, l'ordine delle misure di quello quanti-

tativo. Perciò, essi sono i percentili empirici cui corrispondono le misure poste minori od uguali rispetto ad esse stesse di un fenomeno quantitativo collegabile ad un fenomeno qualitativo.

Di solito, l'analisi statistica cosiddetta parametrica, da svolgere secondo il metodo condizionato dallo schema di un'ipotesi, presuppone di inferire sulle caratteristiche di un insieme, inteso come popolazione matematica o collezione di soggetti o di oggetti aventi in realtà qualità anche diverse, sulla base delle informazioni statistiche di un campione casuale di soggetti o di oggetti appartenenti al predetto insieme. Quindi, l'anzidetta analisi ha lo scopo di ottenere stime di taluni indici e valori segnaletici, ritenuti valori essenziali del fenomeno quantitativo da esaminare che, per lo schema di un'ipotesi, si considera una grandezza astratta di un immaginario insieme infinito di campioni casuali. Ma se si rileva per lo stesso fenomeno misurabile un altro identico numero di osservazioni, ottenute con un procedimento di uguale e sufficiente grado di precisione, si possono ottenere per gli stessi indici e valori segnaletici risultati diversi dai precedenti, se non altro a causa dei fattori sperimentali incontrollabili e di cause sistematiche influenti che determinano i cosiddetti errori di osservazione.

Perciò, bisogna interpretare con la dovuta congruenza il concetto della cosiddetta "significatività statistica" (Gini C. 1943. *I test di significatività*, ecc.) che, nell'ambito delle ricerche sperimentali, è da intendere come un particolare giudizio probabilistico sulla validità di un'ipotesi espressa mediante una relazione matematica, detta anche identità o equazione (49). Cioè, un giudizio di conformità delle osservazioni ad

(49) Secondo la logica matematica, il termine "identità o equazione" è, come è noto, l'uguaglianza tra due espressioni algebriche che può essere verificata per mezzo di particolari valori, chiamati parametri, attribuiti alle variabili incognite dell'equazione. Ossia la relazione matematica che rappresenta una semplificazione simbolica per esprimere, ad esempio, l'idea immaginaria che associa un fenomeno quantitativo di certi soggetti od oggetti, scelti in base ad un meccanismo di sorteggio casuale, ad un fenomeno qualitativo supposto un fenomeno casuale.

Pertanto, la ricerca empirica da effettuare in base all'indagine statistica campionaria pone il problema della verifica della bontà di adattamento dell'equazione ai casi osservati, ossia il problema dell'identità matematica. Ad esempio, la verifica dell'ipotesi citata prima e, quindi, l'uso di un modello probabilistico da adattare ad un'empirica distribuzione di frequenza delle misure di un carattere metrico considerato una grandezza astratta di un immaginario insieme infinito di campioni casuali di soggetti o di oggetti (modello probabilistico adattativo).

Invece, secondo il nostro punto di vista, la relazione per rappresentare in maniera simbolica l'idea immaginabile della corrispondenza tra determinati soggetti od oggetti omogenei, rispetto ad un fenomeno qualitativo valutabile come detto prima, e le misure di un carattere metrico che possono aiutare a valutare il predetto fenomeno qualitativo che limita, in realtà, l'ordine

un'ipotesi, come quella che presuppone lo schema di probabilità degli errori casuali standardizzati (Vianelli S. Ferreri C. 1966., *op. cit.*, pp. 269-270). In tal caso, il test di significatività prescelto è utile per verificare se è corretta, o meno, l'applicazione dell'anzidetto schema. Ossia, per verificare la casualità di un fenomeno qualitativo da associare, senza la ricerca di un'effettiva corrispondenza logica, al fenomeno quantitativo da esaminare mediante un campione casuale di soggetti o di oggetti. Se il risultato di un'analisi statistica di questo genere è significativo o conforme alla suddetta ipotesi vuol dire, in definitiva, che l'analisi è fondata su un vero e proprio campione casuale di soggetti o di oggetti che in realtà, però, possono avere qualità anche diverse, anziché una sola che li accomuni.

Non deve sorprendere, perciò, il fatto che a volte il risultato di una ricerca empirica, effettuata mediante un'indagine statistica campionaria, può essere smentito da un altro, come spesso accade nel settore medico (50).

delle misure di quello quantitativo sarebbe la funzione della potenziale densità di frequenza della cosiddetta variabile "non casuale" o statistica semplice potenzialmente continua in un intervallo limitato chiuso di lunghezza in realtà stabile e, quindi, da standardizzare, rispetto all'indice approssimato della deviazione dalla moda della distribuzione empirica unimodale (asimmetrica) di frequenza della variabile statistica semplice. Cioè, il modello per rappresentare, in uno spazio geometrico bidimensionale, la tendenza delle oscillazioni in parte irregolari e soprattutto regolari delle frequenze intorno al loro massimo cui corrisponde la moda del carattere metrico collegabile al fenomeno qualitativo di determinati soggetti od oggetti (omogenei) che costituiscono in realtà un insieme finito, od una classe. Perciò, è logico ammassare o cumulare tali soggetti od oggetti dato che, peraltro, sono un determinato numero anziché l'infinito matematico.

Per conseguenza, l'unificazione, l'identificazione o la conformità di determinati soggetti od oggetti, rispetto ad una qualità valutabile in base all'esperienza e con l'ausilio della misura di un carattere metrico, può essere rappresentata mediante la funzione di ripartizione, somma integrale o cumulo, della funzione della potenziale densità di frequenza della cosiddetta variabile potenzialmente continua "non casuale" o statistica semplice standardizzata, rispetto all'indice approssimato della deviazione dalla moda della distribuzione empirica unimodale (asimmetrica) di frequenza della variabile "non casuale" o statistica semplice (modello statistico descrittivo).

Il limite della somma integrale della funzione della potenziale densità di frequenza della variabile potenzialmente continua "non casuale" o statistica semplice standardizzata sarebbe, quindi, un integrale definito, idoneo a rappresentare il numero dei soggetti o degli oggetti omogenei, rispetto ad una qualità che limita in realtà l'ordine delle misure di un carattere metrico. In tal modo, non si allude al concetto d'identità o di equazione che scaturisce dalla logica della matematica ma si esprime, secondo la logica consueta, il concetto dell'identificazione o della conformità di determinati soggetti od oggetti (omogenei), rispetto ad una qualità valutabile come detto prima. Pertanto, tale concetto appartiene alla teoria da cui scaturirebbe il metodo statistico per l'esame della variabilità delle misure di un fenomeno quantitativo collegabile ad un fenomeno qualitativo, detto anche fenomeno descrittivo o collettivo.

(50) Secondo il "Journal of the Medical Association", nota rivista dell'Associazione dei medici americani, e la rivista inglese "Lancet", è stato constatato che almeno un terzo delle tera-

Un'enorme quantità di lavori scientifici, compresi quelli svolti in medicina, culminano, infatti, con un'analisi statistica effettuata in base al metodo che scaturisce dalla teoria della stima che sfrutta i dettami della teoria dei campioni casuali.

Di solito, le riviste scientifiche pubblicano lavori che documentano risultati significativi in base all'anzidetto metodo statistico condizionato dallo schema dell'ipotesi della casualità dei fenomeni qualitativi che implica di verificare l'ipotesi medesima e, quindi, l'uso di un prestabilito livello di significatività (Girone G. Salvemini T. 1992., pp. 221-226).

La maggior parte delle ricerche del settore medico dimostrano differenze significative, ad esempio, tra le frequenze percentuali di guarigione da una malattia dopo un trattamento farmacologico e senza il trattamento farmacologico. Si sottolinea, però, che una singola indagine, oppure un'empirica serie di indagini statistiche campionarie, non consente mai conclusioni definitive. Inoltre, l'idea di esaminare casistiche di notevole entità numerica, al fine di evocare il concetto della stabilità della variabilità imputabile ad una teorica serie, tendente a infinito, di cause accidentali o, per meglio dire, la stabilità di variabilità, rispetto alla teorica media aritmetica, delle misure di un fenomeno quantitativo considerato una grandezza astratta di un immaginario insieme infinito di campioni casuali di soggetti o di oggetti in realtà da non osservare, non significa che si possano in realtà ottenere i risultati ipotizzati o, per meglio dire, pretesi (51).

pie farmacologiche ritenute necessarie, oppure no, sono state in seguito smentite o ridimensionate. Infatti, dopo la consultazione di molte ricerche pubblicate tra il 1990 ed il 2003, per una sorta di revisionismo terapeutico, sono stati ribaltati i risultati di diversi studi e ridimensionati i risultati di altri. Ad esempio: l'assunzione di pillole di ormoni, da parte di milioni di donne in menopausa, che proteggerebbe dalle malattie al cuore e accrescerebbe, invece, secondo uno studio più ampio, il rischio cardiaco e aumenterebbe quello di tumore. Inoltre, l'ipotesi che una cura con anticorpi aumenterebbe la sopravvivenza ad un'infezione mortale del sangue, detta anche setticemia, è stata in seguito smentita. Il problema delle ipotesi contrastanti non riguarda soltanto i farmaci. Infatti, secondo uno studio recente, le sostanze antiossidanti contenute nel tè, nel vino, nella frutta e nella verdura che ridurrebbero di molto il rischio di malattie cardiache, avrebbero, invece, benefici soltanto modesti.

Talvolta, queste contraddizioni sono state attribuite al problema del cosiddetto protocollo di studio. Infatti, quando tali studi sono stati condotti su casi controllati ed è stato considerato un campione di soggetti che hanno assunto la terapia ed un altro campione di soggetti di controllo che hanno assunto, invece, placebo è emersa la verità sull'efficacia o meno del trattamento farmacologico.

(51) Se si ammette, infatti, che esiste uno scarso numero di mosche bianche, annidate in una piramide d'Egitto, ed è possibile valutare tale specie d'insetti in base all'esperienza e con l'ausilio della misura della grande apertura delle loro ali, ciò non implica che bisogna osservare codesto fenomeno quantitativo (apertura delle ali) in un notevole numero di mosche bianche, perché in effetti esse sarebbero rare.

Come già detto, talvolta, nell'ambito delle scienze naturali ed umane, è possibile effettuare un collegamento tra un fenomeno quantitativo ed un fenomeno qualitativo che limita, in realtà, l'ordine delle misure di quello quantitativo. Ciò implica l'idea della reale stabilità di variabilità, rispetto alla moda, del fenomeno quantitativo collegabile a quello qualitativo (52) e consente di effettuare un'indagine statistica descrittiva o con il processo di rilevamento delle informazioni statistiche per così dire completo riferito, cioè, ad un insieme finito, o ad una classe, di soggetti o di oggetti omogenei, rispetto alla qualità valutabile in base all'esperienza e con l'ausilio della misura del fenomeno quantitativo. Perciò, il totale delle frequenze empiriche, per così dire non casuali, o numero che scaturisce dalla somma dei soggetti o degli oggetti caratterizzati da un fenomeno qualitativo collegabile ad un fenomeno quantitativo, si deduce dalla valutazione, come detto prima, dell'anzidetto fenomeno qualitativo, anziché da una stima in base ad una distribuzione teorica di densità di probabilità adattata ai casi osservati. Quindi, ai fini del calcolo del variabile stabile in realtà, non si pone il problema se il risultato ottenuto sia conforme (o significativo), oppure no, allo schema di un'ipotesi dato a priori. Cioè, non è necessario che il fenomeno qualitativo che limita l'ordine delle misure di quello quantitativo si consideri, per lo schema di un'ipotesi, un fenomeno casuale da associare, senza la ricerca di un'effettiva corrispondenza logica, al fenomeno quantitativo. Perciò, non occorre stimare il numero dei probabili soggetti od oggetti caratterizzati da codesta qualità.

In conclusione, si raccomandano due criterî del metodo statistico descrittivo o investigativo-comparativo vero e proprio tendenti a:

1) Confermare la valutazione o diagnosi di una qualità che può effettuare, come detto prima, un esperto.

2) Stabilire i confini convenzionali della regolarità o normalità delle oscillazioni delle frequenze delle misure di un carattere metrico collegabile, secondo la logica consueta, ad una qualità valutabile in base all'esperienza e con l'ausilio della misura del predetto carattere metrico (Gargano F. 2002).

Con questi criterî si possono ottenere, in un determinato tempo e luoghi diversi, oppure in un determinato luogo e tempi diversi, diffe-

(52) È ovvio che la reale stabilità della variabilità, rispetto alla moda, di un fenomeno quantitativo collegabile ad un fenomeno qualitativo non dipende dal numero notevole dei soggetti o degli oggetti omogenei, rispetto alla qualità valutabile come detto prima, ma dalla giusta diagnosi della qualità stessa.

renti confini convenzionali delle regolari o normali oscillazioni delle frequenze intorno al loro massimo cui corrisponde la moda del carattere metrico collegabile alla qualità considerata determinante, come – ad esempio – una causa ereditaria, valutabile come detto prima, anche se è da supporre altre cause sistematiche, come – ad esempio – l'ambiente o le differenti condizioni di vita, che possono avere in realtà minore influenza sulla variabilità, rispetto alla moda, del carattere metrico collegabile al fenomeno qualitativo ritenuto maggiormente influente sulla variabilità del fenomeno quantitativo.

Quindi, per fini statistici descrittivi possono essere impiegati gli anzidetti criteri, anziché quelli di analisi della varianza (53) cui spesso si ricorre nelle indagini statistiche campionarie, dato che in realtà non è possibile separare l'effetto dovuto a diverse prevedibili cause sistematiche dall'effetto dovuto ad un'imprecisabile numero di cause imprevedibili, cosiddette accidentali, dette anche fattori o, per meglio dire, addendi sperimentali, considerato che l'analisi della varianza è fondata sulla proprietà addittiva delle cause di variabilità supposte accidentali.

A parità di qualità dei soggetti o degli oggetti (omogenei) che, in un determinato tempo e luoghi diversi, oppure in un determinato luogo e tempi diversi, costituiscono due o più di due insiemi finiti, oppure due o più di due classi, di diversa numerosità si può, cioè, confrontare la normalità o regolarità delle oscillazioni delle frequenze intorno al loro massimo cui corrisponde la moda del carattere metrico collegabile alla qualità dei soggetti o degli oggetti medesimi.

(53) Con cui si prefigura una relazione di tipo lineare – che, in realtà, può essere, invece, non lineare – tra talune cause sistematiche prevedibili (influenze ereditarie ed ambientali), supposte casuali, ed altre cause accidentali non prevedibili che costituiscono un'immaginaria serie, tendente ad infinito, che in teoria si tradurrebbero in un'altrettanta serie di probabili errori casuali standardizzati di misura da immaginare con segni alterni positivi e negativi, di uguale intensità ed anche d'intensità illimitata, a partire da zero o dalla misura esatta del carattere metrico di un immaginario insieme infinito di campioni casuali di oggetti o di soggetti in realtà da non osservare.

Nell'ambito delle ricerche antropologiche, le indagini statistiche campionarie tendono a separare, perciò, in base ai criteri di analisi della varianza proposti dalla teoria statistica della stima, l'effetto delle influenze ereditarie da quelle ambientali, entrambe supposte fenomeni casuali, addizionate all'effetto di altre imprevedibili cause, cosiddette accidentali, che si tradurrebbero in un'immaginaria serie, tendente a infinito, di probabili errori casuali standardizzati di misura. Inoltre, talvolta, nell'ambito delle ricerche antropologiche viene usato il cosiddetto metodo dei gemelli, per cui si confrontano percentuali di coppie gemellari (monovulari) caratterizzate da differenti qualità, supposte fenomeni casuali, e si impiega, perciò, la funzione test X^2 per stabilire se è significativa o meno la differenza tra le percentuali poste a confronto.

Pertanto, l'indagine statistica riferita ad un insieme finito, o ad una classe, di soggetti o di oggetti omogenei, rispetto ad una qualità valutabile in base all'esperienza e con l'ausilio della misura di un fenomeno quantitativo induce all'uso di un indice di variabilità assoluta, per così dire totale, per standardizzare, normalizzare, regolarizzare o ridurre, la lunghezza in realtà stabile dell'intervallo limitato chiuso, o compatto, della cosiddetta variabile potenzialmente continua "non casuale" o statistica semplice. Se tale lunghezza si divide al predetto indice di variabilità assoluta si elimina l'influenza della predetta causa sistematica di variabilità. Quindi, tale operazione aritmetica è necessaria per fini statistici descrittivi o investigativi - comparativi, in quanto elimina l'influenza della stabile variabilità, rispetto alla moda, di un fenomeno quantitativo collegabile ad un fenomeno qualitativo o causa sistematica prevedibile determinante, difficilmente separabile da un imprecisabile numero di cause accidentali non prevedibili.

Ringraziamenti

È doveroso un ringraziamento alla professoressa Carla Rossi dell'Università "Tor Vergata" di Roma che nel 1984 ci consigliò di considerare i percentili empirici per formulare l'indice di variabilità assoluta rispetto alla moda.

BIBLIOGRAFIA

- GARGANO F. 1999. *L'impiego, per mezzo di dati reali, di un criterio statistico per confermare l'omogeneità e determinare, con un prestabilito livello ottimo di affidabilità, la prevalenza o normalità di taluni oggetti riuniti insieme*. In "Rivista Italiana di Economia Demografia e Statistica", n. 3.
- GINI C. 1916. *Il concetto di "transvariazione" e le sue prime applicazioni*. In "Giornale degli Economisti e Rivista di Statistica". Vol. LII. Ristampato nel 1939 con ulteriori sviluppi ed esempi in "Memorie di Metodologia Statistica". Vol. I. Giuffrè. Ed. Milano. Oppure consulta:
- SALVEMINI T. 1959, *Transvariation and discriminant analysis*. In "Memorie di Metodologia Statistica". Vol. II. Libreria Goliardica. Roma.
 - BOLDRINI M. 1962. *Statistica. Teoria e metodi*. Giuffrè. Ed. Milano.
 - VAJANI L. 1990. *Statistica descrittiva*. Gruppo editoriale Fabbri. Etas. S.p.A. Ed. Sonzognò.
- GARGANO F. 1997. *L'indice approssimato della deviazione standard di una variabile statistica o non casuale*. In "Antropologia Contemporanea". Atti del XX Congresso Ass. Antropolog. Ital. Palermo - Alia. 16-20 Settembre.
- JOHNSON R.A. BHATTACHARYA G. K. 1992. *Statistics. Principles and Methods*. John Wiley e Sons. Second Ed. New York.
- LOMBARDO E. 1993. *I dati statistici in pedagogia. Esplorazione e analisi*. La Nuova Italia. Ed. Firenze.
- BIOMETRIKA 1954. *Tables for statisticians*. Vol. 1. Edited by Pearson E.S., and Hartely H.O. Cambridge University Press. Oppure consulta:
- VIANELLI S. 1959. *Prontuari per calcoli statistici. Tavole numeriche e complementi*. Ed. Abbaco. Palermo - Roma.
- GINI C. 1962. *La logica nella statistica. Serie di statistica. Teoria e applicazioni*. 18. Ed. Boringhieri. Torino.
- LETI G., 1983. *Statistica descrittiva*. Ed. Il Mulino. Bologna.
- FRAIRE M., RIZZI A. 1999. *Statistica*. Ed. Carocci. Roma.
- LETI G. 1990. *Distanze e indici*. Ed. Universitaria di Roma. La Goliardica.
- SALSA S., SQUELLATI A. 1995. *Esercizi di analisi matematica 2. Parte prima*. Masson. Ed. Milano.

- VIANELLI S., FERRERI C. 1966. *Calcolo delle probabilità e principi di statistica metodologica*. Calderini. Ed. Bologna.
- MIGNANI S. 1990. *Le tecniche Jackknife e Bootstrap per la stima della media quadratica: una simulazione con valori estremi*. In "Quaderni di Statistica e Econometria". Vol. XII. Università degli studi di Bologna.
- MIGNANI S. 1992. *Le tecniche di ricampionamento Jackknife e Bootstrap*. Serie Strumenti per la Didattica n. 3. Dipartimento di Scienze Statistiche "Paolo Fortunati". Università degli Studi di Bologna. Inoltre consulta:
 - MIGNANI S. MONTANARI A. *Appunti di analisi statistica multivariata*. Società Editrice Esculapio. Bologna.
- ROSSI C. 1988. *Classificazione bayesiana: applicazione alla diagnosi di diabete mellito relativa ad un'indagine epidemiologica nell'isola di Pantelleria*. In "Rivista di Statistica Applicata". Vol. 21, n. 4. Edizioni Scientifiche Tecniche Europee, s.r.l. Milano.
- SAVAGNONE E. CAPPELLO F. CARUSO V. GALVANO L. MANGIONE G. NICOLOSI G. RAIMONDO F. SCHIRÒ V. SPICOLA L. SPINELLI M. SUNSERI V. GARGANO F. 1985. *Indagine epidemiologica sul diabete mellito della popolazione di Pantelleria (Sicilia). Note preliminari*. In "Giornale Italiano di Diabetologia". Vol. V, n. 1.
- TRELOAR A. 1951. *Biometric Analysis. An Introduction*. Burgess. Minnesota.
- BRAMBILLA F. 1968. *Trattato di Statistica. I. La variabilità*. UTET. Torino.
- WORLD HEALTH ORGANIZATION. 1997. *Report of the expert committee. The committee on the diagnosis and classification of diabetes mellitus*.
- DI IORIO F. 1988. *Statistica Medica. Le basi quantitative della ricerca biomedica*. Ed. NIS. La Nuova Italia Scientifica. Roma.
- BRESSAN F. 1979. *Nota sulla simmetria e normalità*. In "Statistica". Anno XXXIX, n. 3.
- KAPLANSKY I. 1945. *A common error concerning kurtosis*. In "Journal of the Am. Stat. Ass." 40. June.
 - LOMBARDO E. 1974. *Nota sulla curtosi*. In "Statistica". Anno XXXIV, n. 3.
 - MOOD A. M. GRAYBILL F. A. 1963. *Introduction to the theory of statistics*. Second Ed. Mc Graw Hill. New York.
- GABRIELLI A. 1967. *Dizionario dei sinonimi e dei contrari*. Ed. Istituto editoriale italiano. Milano. Inoltre consulta:
- Devoto G. Oli G. C. 1991. *Nuovo vocabolario illustrato della lingua italiana. Selezione dal Readers Digest*. Ed. Milano.
- GINI C. TREZZA A. 1962. *Anormalità delle distribuzioni statistiche dei caratteri antropologici*. In "Metron". Vol. 21.
- GIRONE G. SALVEMINI T. 1983. *Lezioni di Statistica*. Volume primo. Cacucci. Ed. Bari.
- FROSINI B. MONTINARO M. NICOLINI G. 1994. *Il campionamento da popolazioni finite. Metodi e applicazioni*. Ed. UTET. Torino.

- CICCHITELLI G. HERZEL A. MONTANARI G. E. 1992. *Il campionamento statistico*. Ed. Il Mulino, Bologna.
- FABRIS L. 1997. *Statistica multivariata. Analisi esplorativa dei dati*. Ed. McGraw Libri Italia, s.r.l. Milano.
- REED A. H. HENRY J. R. MASON W. B. 1971. *Influence of statistical method used on the resulting estimate of normal range*. In "Clinical Chemistry", n° 4.
- GINI C. 1943. *I test di significatività*. In "Atti della VII Riunione Scientifica della Società Italiana di Statistica".
- GIRONE G. SALVEMINI T. 1992. *Lezioni di Statistica*. Volume secondo. Cacucci. Ed. Bari.
- GARGANO F. 2002. *Un criterio statistico per confermare l'omogeneità degli oggetti di una classe e per determinare i confini convenzionali della regolarità, o normalità, delle oscillazioni delle frequenze delle misure del carattere metrico collegabile alla qualità degli oggetti medesimi*. In un cd-rom sul "Convegno dell'aria e salute nelle aree urbane". Organizzato dall'Università degli studi di Verona, dall'European Community, Respiratory Health Survey-Italy e dall'Agenzia Regionale Prevenzione e Protezione Ambiente del Veneto.

TABELLE NUMERICHE

Tabella 1 - Distribuzione empirica plurimodale di frequenza delle misure di glicemia a digiuno di un insieme di 3.288 soggetti eterogenei, per stato di salute, di ambo i sessi di 20 anni di età ed oltre censiti a Pantelleria da alcuni epidemiologi.

Misura della glicemia a digiuno in mg % di sangue capillare (misura del fenomeno quantitativo)		Numero dei soggetti eterogenei, rispetto allo stato di salute (diabete, predisposizione al diabete, non diabete) (frequenza assoluta)	Numero dei soggetti eterogenei, rispetto allo stato di salute, relativo al totale dell'insieme dei soggetti medesimi moltiplicato per 100 (frequenza relativa percentuale) $100 \left[y_i / \sum_{i=1}^n y_i \right] = 100 f'_i$
Classe delle misure	Semisomma delle misure estreme di ogni classe		
x_i		y_i	
Col. 1	Col. 2	Col. 3	Col. 4
34,5 - 35,5	35	1	0,00030
35,5 - 36,5	36	2	0,00061
39,5 - 40,5	40	1	0,00030
40,5 - 41,5	41	1	0,00030
41,5 - 42,5	42	1	0,00030
42,5 - 43,5	43	1	0,00030
43,5 - 44,5	44	1	0,00030
44,5 - 45,5	45	8	0,00243
45,5 - 46,5	46	11	0,00335
46,5 - 47,5	47	7	0,00213
47,5 - 48,5	48	8	0,00243
48,5 - 49,5	49	10	0,00304
49,5 - 50,5	50	13	0,00395
50,5 - 51,5	51	7	0,00213
51,5 - 52,5	52	7	0,00213
52,5 - 53,5	53	13	0,00395
53,5 - 54,5	54	11	0,00335
54,5 - 55,5	55	22	0,00669
55,5 - 56,5	56	18	0,00547
56,5 - 57,5	57	17	0,00517
57,5 - 58,5	58	19	0,00578
58,5 - 59,5	59	19	0,00578
59,5 - 60,5	60	24	0,00730
60,5 - 61,5	61	35	0,01064
61,5 - 62,5	62	24	0,00730
62,5 - 63,5	63	31	0,00943
63,5 - 64,5	64	41	0,01247
64,5 - 65,5	65	36	0,01095
65,5 - 66,5	66	40	0,01217
66,5 - 67,5	67	36	0,01095
67,5 - 68,5	68	42	0,01277

Col. 1	Col. 2	Col. 3	Col. 4
68,5 - 69,5	69	56	0,01703
69,5 - 70,5	70	41	0,01247
70,5 - 71,5	71	44	0,01338
71,5 - 72,5	72	50	0,01521
72,5 - 73,5	73	54	0,01642
73,5 - 74,5	74	40	0,01217
74,5 - 75,5	75	68	0,02068
75,5 - 76,5	76	62	0,01886
76,5 - 77,5	77	75	0,02281
77,5 - 78,5	78	61	0,01855
78,5 - 79,5	79	44	0,01338
79,5 - 80,5	80	90	0,02737
80,5 - 81,5	81	54	0,01642
81,5 - 82,5	82	53	0,01612
82,5 - 83,5	83	62	0,01886
83,5 - 84,5	84	62	0,01886
84,5 - 85,5	85	81	0,02464
85,5 - 86,5	86	59	0,01794
86,5 - 87,5	87	73	0,02220
87,5 - 88,5	88	68	0,02068
88,5 - 89,5	89	70	0,02129
89,5 - 90,5	90	87	0,02646
90,5 - 91,5	91	39	0,01186
91,5 - 92,5	92	64	0,01946
92,5 - 93,5	93	71	0,02159
93,5 - 94,5	94	37	0,01125
94,5 - 95,5	95	82	0,02494
95,5 - 96,5	96	64	0,01946
96,5 - 97,5	97	48	0,01460
97,5 - 98,5	98	48	0,01460
98,5 - 99,5	99	35	0,01064
99,5 - 100,5	100	112	0,03406
100,5 - 101,5	101	49	0,01490
101,5 - 102,5	102	44	0,01338
102,5 - 103,5	103	32	0,00973
103,5 - 104,5	104	41	0,01247
104,5 - 105,5	105	43	0,01308
105,5 - 106,5	106	27	0,00821
106,5 - 107,5	107	33	0,01004
107,5 - 108,5	108	14	0,00426
108,5 - 109,5	109	17	0,00517
109,5 - 110,5	110	40	0,01217
110,5 - 111,5	111	20	0,00608
111,5 - 112,5	112	22	0,00669
112,5 - 113,5	113	8	0,00243
113,5 - 114,5	114	11	0,00335

Col. 1	Col. 2	Col. 3	Col. 4
114,5 - 115,5	115	35	0,01064
115,5 - 116,5	116	12	0,00365
116,5 - 117,5	117	11	0,00335
117,5 - 118,5	118	8	0,00243
118,5 - 119,5	119	13	0,00395
119,5 - 120,5	120	36	0,01095
120,5 - 121,5	121	7	0,00213
121,5 - 122,5	122	12	0,00365
122,5 - 123,5	123	14	0,00426
123,5 - 124,5	124	10	0,00304
124,5 - 125,5	125	13	0,00395
125,5 - 126,5	126	4	0,00122
126,5 - 127,5	127	3	0,00091
127,5 - 128,5	128	7	0,00213
128,5 - 129,5	129	5	0,00152
129,5 - 130,5	130	8	0,00243
130,5 - 131,5	131	3	0,00091
131,5 - 132,5	132	6	0,00183
132,5 - 133,5	133	3	0,00091
133,5 - 134,5	134	2	0,00061
134,5 - 135,5	135	4	0,00122
135,5 - 136,5	136	4	0,00122
136,5 - 137,5	137	2	0,00061
137,5 - 138,5	138	3	0,00091
138,5 - 139,5	139	6	0,00182
139,5 - 140,5	140	10	0,00304
140,5 - 141,5	141	4	0,00122
141,5 - 142,5	142	5	0,00152
142,5 - 143,5	143	5	0,00152
143,5 - 144,5	144	4	0,00122
144,5 - 145,5	145	3	0,00091
145,5 - 146,5	146	4	0,00122
146,5 - 147,5	147	3	0,00091
147,5 - 148,5	148	2	0,00061
148,5 - 149,5	149	3	0,00091
149,5 - 150,5	150	3	0,00091
150,5 - 151,5	151	5	0,00152
151,5 - 152,5	152	2	0,00061
152,5 - 153,5	153	3	0,00091
154,5 - 155,5	155	8	0,00243
155,5 - 156,5	156	2	0,00061
156,5 - 157,5	157	2	0,00061
157,5 - 158,5	158	2	0,00061
158,5 - 159,5	159	1	0,00030
160,5 - 161,5	161	1	0,00030
161,5 - 162,5	162	3	0,00091

Col. 1	Col. 2	Col. 3	Col. 4
162,5 - 163,5	163	2	0,00061
164,5 - 165,5	165	1	0,00030
165,5 - 166,5	166	2	0,00061
166,5 - 167,5	167	1	0,00030
167,5 - 168,5	168	3	0,00091
168,5 - 169,5	169	1	0,00030
170,5 - 171,5	171	1	0,00030
171,5 - 172,5	172	2	0,00061
172,5 - 173,5	173	3	0,00091
173,5 - 174,5	174	2	0,00061
171,5 - 175,5	175	3	0,00091
175,5 - 176,5	176	1	0,00030
176,5 - 177,5	177	2	0,00061
177,5 - 178,5	178	1	0,00030
178,5 - 179,5	179	1	0,00030
179,5 - 180,5	180	3	0,00091
180,5 - 181,5	181	2	0,00061
181,5 - 182,5	182	2	0,00061
183,5 - 184,5	184	1	0,00030
184,5 - 185,5	185	2	0,00061
185,5 - 186,5	186	2	0,00061
187,5 - 188,5	188	4	0,00122
188,5 - 189,5	189	1	0,00030
189,5 - 190,5	190	2	0,00061
190,5 - 191,5	191	1	0,00030
193,5 - 194,5	194	3	0,00091
195,5 - 196,5	196	1	0,00030
199,5 - 200,5	200	2	0,00061
201,5 - 202,5	202	1	0,00030
202,5 - 203,5	203	1	0,00030
203,5 - 204,5	204	2	0,00061
209,5 - 210,5	210	6	0,00182
212,5 - 213,5	213	2	0,00061
213,5 - 214,5	214	1	0,00030
214,5 - 215,5	215	1	0,00030
216,5 - 217,5	217	2	0,00061
217,5 - 218,5	218	1	0,00030
220,5 - 221,5	221	1	0,00030
222,5 - 223,5	223	1	0,00030
224,5 - 225,5	225	2	0,00061
225,5 - 226,5	226	1	0,00030
226,5 - 227,5	227	1	0,00030
227,5 - 228,5	228	1	0,00030
229,5 - 230,5	230	3	0,00091
232,5 - 233,5	233	2	0,00061
234,5 - 235,5	235	2	0,00061

Col. 1	Col. 2	Col. 3	Col. 4
240,5 - 241,5	241	1	0,00030
242,5 - 243,5	243	1	0,00030
245,5 - 246,5	246	1	0,00030
248,5 - 249,5	249	1	0,00030
249,5 - 250,5	250	2	0,00061
256,5 - 257,5	257	1	0,00030
258,5 - 259,5	259	1	0,00030
259,5 - 260,5	260	3	0,00091
264,5 - 265,5	265	1	0,00030
267,5 - 268,5	268	1	0,00030
274,5 - 275,5	275	1	0,00030
277,5 - 278,5	278	1	0,00030
282,5 - 283,5	283	1	0,00030
285,5 - 286,5	286	1	0,00030
294,5 - 295,5	295	2	0,00061
299,5 - 300,5	300	2	0,00061
300,5 - 301,5	301	1	0,00030
301,5 - 302,5	302	1	0,00030
304,5 - 305,5	305	1	0,00030
305,5 - 306,5	306	1	0,00030
317,5 - 318,5	318	1	0,00030
299,5 - 320,5	320	1	0,00030
336,5 - 337,5	337	1	0,00030
341,5 - 342,5	342	1	0,00030
344,5 - 345,5	345	1	0,00030
350,5 - 351,5	351	1	0,00030
354,5 - 355,5	355	1	0,00030
374,5 - 375,5	375	1	0,00030
376,5 - 377,5	377	1	0,00030
398,5 - 399,5	399	1	0,00030
399,5 - 400,5	400	2	0,00061
Totale		3.288	100

Come è noto, i fenomeni misurabili sono di solito continui e, quindi, da esprimere mediante numeri interi e frazioni dei numeri interi medesimi. Tuttavia, le misure di glicemia che costituiscono il materiale documentario con cui è stata effettuata quest'analisi risultano avere salti di discontinuità. Quindi, per segnalare formalmente la continuità della variazione delle predette misure, anziché la loro discontinuità così come risulta dai valori singoli di glicemia rilevati dai medici, sono state adottate classi di misure di lunghezza costante e con semisomma coincidente con il loro valore singolo effettivamente rilevato.

La media aritmetica e la moda, o misura più frequente, della semisomma delle misure estreme delle classi, di lunghezza unitaria, delle misure espresse in mg % di sangue capillare della glicemia a digiuno sono risultate, rispettivamente, 93,026 e 100. Non è stata calcolata, invece, la mediana, cui corrisponde il percentile empirico 50, delle predette misure. Perciò, non è stata considerata l'empirica distribuzione cumulativa di frequenza delle misure di glicemia a digiuno e dopo due ore dal carico di 75 g di glucosio, in quanto non è stato ritenuto opportuno effettuare il cumulo dei soggetti eterogenei, rispetto ai diversi stati di salute diagnosticati dai medici in occasione dell'anzidetta indagine epidemiologica.

Tabella 2 - Distribuzione empirica plurimodale di frequenza delle misure di glicemia dopo due ore dal carico di 75 g di glucosio di un insieme di 3.259 soggetti eterogenei, per stato di salute, di ambo i sessi di 20 anni di età ed oltre censiti a Pantelleria da alcuni epidemiologi.

Misura della glicemia dopo due ore dal carico di 75 g di glucosio in mg % di sangue capillare (misura del fenomeno quantitativo)		Numero dei soggetti eterogenei, rispetto allo stato di salute (diabete, predisposizione al diabete, non diabete) (frequenza assoluta)	Numero dei soggetti eterogenei, rispetto allo stato di salute, relativo al totale dell'insieme dei soggetti medesimi moltiplicato per 100 (frequenza relativa percentuale)
Classe delle misure	Semisomma delle misure estreme di ogni classe	y_i	$100 \left[y_i / \sum_{i=1}^k y_i \right] = 100 f_i$
x_i			
Col. 1	Col. 2	Col. 3	Col. 4
29,5 - 30,5	30	1	0,00030
37,5 - 38,5	38	1	0,00030
38,5 - 39,5	39	1	0,00030
39,5 - 40,5	40	1	0,00030
40,5 - 41,5	41	3	0,00091
41,5 - 42,5	42	4	0,00122
42,5 - 43,5	43	1	0,00030
43,5 - 44,5	44	4	0,00122
44,5 - 45,5	45	3	0,00030
45,5 - 46,5	46	6	0,00182
46,5 - 47,5	47	3	0,00091
47,5 - 48,5	48	4	0,00122
48,5 - 49,5	49	6	0,00182
49,5 - 50,5	50	7	0,00213
50,5 - 51,5	51	8	0,00243
51,5 - 52,5	52	7	0,00213
52,5 - 53,5	53	1	0,00030
53,5 - 54,5	54	6	0,00182
54,5 - 55,5	55	13	0,00395
55,5 - 56,5	56	5	0,00152
56,5 - 57,5	57	13	0,00395
57,5 - 58,5	58	10	0,00304
58,5 - 59,5	59	10	0,00304
59,5 - 60,5	60	14	0,00426
60,5 - 61,5	61	9	0,00274
61,5 - 62,5	62	15	0,00456
62,5 - 63,5	63	25	0,00760
63,5 - 64,5	64	15	0,00456
64,5 - 65,5	65	20	0,00608
65,5 - 66,5	66	16	0,00487

Col. 1	Col. 2	Col. 3	Col. 4
66,5 - 67,5	67	17	0,00517
67,5 - 68,5	68	16	0,00487
68,5 - 69,5	69	14	0,00426
69,5 - 70,5	70	16	0,00487
70,5 - 71,5	71	17	0,00517
71,5 - 72,5	72	23	0,00700
72,5 - 73,5	73	19	0,00578
73,5 - 74,5	74	10	0,00304
74,5 - 75,5	75	36	0,01095
75,5 - 76,5	76	21	0,00639
76,5 - 77,5	77	28	0,00852
77,5 - 78,5	78	23	0,00700
78,5 - 79,5	79	23	0,00700
79,5 - 80,5	80	43	0,01308
80,5 - 81,5	81	33	0,01004
81,5 - 82,5	82	31	0,00943
82,5 - 83,5	83	36	0,01095
83,5 - 84,5	84	25	0,00760
84,5 - 85,5	85	34	0,01034
85,5 - 86,5	86	36	0,01095
86,5 - 87,5	87	31	0,00942
87,5 - 88,5	88	41	0,01247
88,5 - 89,5	89	29	0,00882
89,5 - 90,5	90	39	0,01186
90,5 - 91,5	91	27	0,00821
91,5 - 92,5	92	35	0,01064
92,5 - 93,5	93	33	0,01004
93,5 - 94,5	94	25	0,00760
94,5 - 95,5	95	43	0,01308
95,5 - 96,5	96	34	0,01034
96,5 - 97,5	97	32	0,00973
97,5 - 98,5	98	34	0,01034
98,5 - 99,5	99	25	0,00760
99,5 - 100,5	100	47	0,01429
100,5 - 101,5	101	41	0,01247
101,5 - 102,5	102	41	0,01247
102,5 - 103,5	103	20	0,00608
103,5 - 104,5	104	49	0,01490
104,5 - 105,5	105	45	0,01369
105,5 - 106,5	106	35	0,01064
106,5 - 107,5	107	38	0,01156
107,5 - 108,5	108	26	0,00791
108,5 - 109,5	109	34	0,01034
109,5 - 110,5	110	69	0,02099
110,5 - 111,5	111	52	0,01581
111,5 - 112,5	112	57	0,01734
112,5 - 113,5	113	18	0,00547

Col. 1	Col. 2	Col. 3	Col. 4
113,5 - 114,5	114	35	0,01064
114,5 - 115,5	115	31	0,00943
115,5 - 116,5	116	27	0,00821
116,5 - 117,5	117	32	0,00973
117,5 - 118,5	118	39	0,01186
118,5 - 119,5	119	27	0,00821
119,5 - 120,5	120	33	0,01004
120,5 - 121,5	121	26	0,00791
121,5 - 122,5	122	31	0,00943
122,5 - 123,5	123	26	0,00790
123,5 - 124,5	124	21	0,00639
124,5 - 125,5	125	58	0,01764
125,5 - 126,5	126	26	0,00791
126,5 - 127,5	127	13	0,00395
127,5 - 128,5	128	21	0,00639
128,5 - 129,5	129	13	0,00395
129,5 - 130,5	130	41	0,01247
130,5 - 131,5	131	14	0,00426
131,5 - 132,5	132	24	0,00730
132,5 - 133,5	133	26	0,00791
133,5 - 134,5	134	25	0,00760
134,5 - 135,5	135	25	0,00760
135,5 - 136,5	136	26	0,00791
136,5 - 137,5	137	29	0,00882
137,5 - 138,5	138	16	0,00487
138,5 - 139,5	139	15	0,00456
139,5 - 140,5	140	38	0,01156
140,5 - 141,5	141	4	0,00122
141,5 - 142,5	142	15	0,00456
142,5 - 143,5	143	10	0,00304
143,5 - 144,5	144	5	0,00152
144,5 - 145,5	145	21	0,00639
145,5 - 146,5	146	15	0,00456
146,5 - 147,5	147	14	0,00426
147,5 - 148,5	148	9	0,00274
148,5 - 149,5	149	6	0,00182
149,5 - 150,5	150	15	0,00456
150,5 - 151,5	151	11	0,00335
151,5 - 152,5	152	5	0,00152
152,5 - 153,5	153	5	0,00152
153,5 - 154,5	154	11	0,00335
154,5 - 155,5	155	18	0,00547
155,5 - 156,5	156	8	0,00243
156,5 - 157,5	157	6	0,00182
157,5 - 158,5	158	6	0,00182
158,5 - 159,5	159	3	0,00091
159,5 - 160,5	160	13	0,00395

Col. 1	Col. 2	Col. 3	Col. 4
160,5 - 161,5	161	10	0,00304
161,5 - 162,5	162	8	0,00243
162,5 - 163,5	163	9	0,00274
163,5 - 164,5	164	5	0,00152
164,5 - 165,5	165	10	0,00304
165,5 - 166,5	166	6	0,00182
166,5 - 167,5	167	11	0,00335
167,5 - 168,5	168	9	0,00274
168,5 - 169,5	169	7	0,00213
169,5 - 170,5	170	8	0,00243
170,5 - 171,5	171	5	0,00152
171,5 - 172,5	172	11	0,00335
172,5 - 173,5	173	12	0,00365
173,5 - 174,5	174	1	0,00030
174,5 - 175,5	175	4	0,00122
175,5 - 176,5	176	7	0,00213
176,5 - 177,5	177	9	0,00274
177,5 - 178,5	178	3	0,00091
178,5 - 179,5	179	9	0,00274
179,5 - 180,5	180	15	0,00456
180,5 - 181,5	181	5	0,00154
181,5 - 182,5	182	27	0,00821
182,5 - 183,5	183	5	0,00152
183,5 - 184,5	184	3	0,00091
184,5 - 185,5	185	9	0,00274
185,5 - 186,5	186	8	0,00243
186,5 - 187,5	187	2	0,00061
187,5 - 188,5	188	7	0,00213
188,5 - 189,5	189	1	0,00030
189,5 - 190,5	190	9	0,00273
190,5 - 191,5	191	1	0,00030
191,5 - 192,5	192	6	0,00182
193,5 - 194,5	194	4	0,00122
194,5 - 195,5	195	5	0,00152
195,5 - 196,5	196	5	0,00152
196,5 - 197,5	197	5	0,00152
197,5 - 198,5	198	5	0,00152
198,5 - 199,5	199	3	0,00091
199,5 - 200,5	200	10	0,00304
200,5 - 201,5	201	1	0,00030
201,5 - 202,5	202	8	0,00243
202,5 - 203,5	203	1	0,00030
203,5 - 204,5	204	2	0,00061
204,5 - 205,5	205	4	0,00122
205,5 - 206,5	206	2	0,00061
206,5 - 207,5	207	3	0,00091
207,5 - 208,5	208	4	0,00122

Col. 1	Col. 2	Col. 3	Col. 4
209,5 - 210,5	210	11	0,00335
210,5 - 211,5	211	6	0,00182
211,5 - 212,5	212	4	0,00122
212,5 - 213,5	213	2	0,00061
213,5 - 214,5	214	7	0,00213
214,5 - 215,5	215	8	0,00243
215,5 - 216,5	216	4	0,00122
216,5 - 217,5	217	2	0,00061
217,5 - 218,5	218	3	0,00091
218,5 - 219,5	219	2	0,00061
219,5 - 220,5	220	7	0,00213
220,5 - 221,5	221	5	0,00152
221,5 - 222,5	222	1	0,00030
222,5 - 223,5	223	4	0,00122
224,5 - 225,5	225	3	0,00091
225,5 - 226,5	226	1	0,00030
227,5 - 228,5	228	2	0,00091
228,5 - 229,5	229	4	0,00122
229,5 - 230,5	230	9	0,00274
230,5 - 231,5	231	4	0,00122
231,5 - 232,5	232	2	0,00061
233,5 - 234,5	234	4	0,00122
234,5 - 235,5	235	1	0,00030
235,5 - 236,5	236	2	0,00061
237,5 - 238,5	238	2	0,00061
238,5 - 239,5	239	3	0,00091
239,5 - 240,5	240	6	0,00182
240,5 - 241,5	241	1	0,00030
241,5 - 242,5	242	2	0,00061
243,5 - 244,5	244	2	0,00061
244,5 - 245,5	245	2	0,00061
245,5 - 246,5	246	3	0,00091
246,5 - 247,5	247	4	0,00122
247,5 - 248,5	248	2	0,00061
248,5 - 249,5	249	2	0,00061
249,5 - 250,5	250	6	0,00182
250,5 - 251,5	251	1	0,00030
251,5 - 252,5	252	1	0,00030
252,5 - 253,5	253	1	0,00030
253,5 - 254,5	254	2	0,00061
255,5 - 256,5	256	4	0,00122
256,5 - 257,5	257	5	0,00152
257,5 - 258,5	258	4	0,00122
258,5 - 259,5	259	1	0,00030
259,5 - 260,5	260	2	0,00061
261,5 - 262,5	262	2	0,00061
264,5 - 265,5	265	1	0,00030

Col. 1	Col. 2	Col. 3	Col. 4
266,5 - 267,5	267	2	0,00061
267,5 - 268,5	268	2	0,00061
269,5 - 270,5	270	1	0,00030
270,5 - 271,5	271	6	0,00182
272,5 - 273,5	273	2	0,00061
273,5 - 274,5	274	1	0,00030
274,5 - 275,5	275	2	0,00061
275,5 - 276,5	276	2	0,00061
277,5 - 278,5	278	3	0,00091
278,5 - 279,5	279	1	0,00030
279,5 - 280,5	280	4	0,00122
280,5 - 281,5	281	2	0,00061
281,5 - 282,5	282	2	0,00061
282,5 - 283,5	283	3	0,00091
283,5 - 284,5	284	4	0,00122
284,5 - 285,5	285	1	0,00030
285,5 - 286,5	286	2	0,00061
287,5 - 288,5	288	5	0,00152
288,5 - 289,5	289	2	0,00061
289,5 - 290,5	290	3	0,00091
290,5 - 291,5	291	1	0,00030
291,5 - 292,5	292	4	0,00122
294,5 - 295,5	295	1	0,00030
296,5 - 297,5	297	3	0,00091
297,5 - 298,5	298	1	0,00030
299,5 - 300,5	300	3	0,00091
300,5 - 301,5	301	2	0,00061
301,5 - 302,5	302	3	0,00091
309,5 - 310,5	310	1	0,00030
312,5 - 313,5	313	3	0,00091
316,5 - 317,5	317	2	0,00061
317,5 - 318,5	318	2	0,00061
319,5 - 320,5	320	1	0,00030
323,5 - 324,5	324	2	0,00061
324,5 - 325,5	325	3	0,00091
326,5 - 327,5	327	1	0,00030
327,5 - 328,5	328	1	0,00030
329,5 - 330,5	330	2	0,00061
332,5 - 333,5	333	2	0,00061
333,5 - 334,5	334	1	0,00030
335,5 - 336,5	336	1	0,00030
342,5 - 343,5	343	2	0,00061
349,5 - 350,5	350	7	0,00213
350,5 - 351,5	351	1	0,00030
352,5 - 353,5	353	1	0,00030
357,5 - 358,5	358	2	0,00061
360,5 - 361,5	361	2	0,00061

Col. 1	Col. 2	Col. 3	Col. 4
361,5 - 362,5	362	1	0,00030
365,5 - 366,5	366	1	0,00030
367,5 - 368,5	368	2	0,00061
369,5 - 370,5	370	1	0,00030
371,5 - 372,5	372	1	0,00030
372,5 - 373,5	373	2	0,00061
378,5 - 379,5	379	1	0,00030
379,5 - 380,5	380	1	0,00030
380,5 - 381,5	381	1	0,00030
382,5 - 383,5	383	2	0,00061
383,5 - 384,5	384	1	0,00030
389,5 - 390,5	390	2	0,00061
393,5 - 394,5	394	2	0,00061
398,5 - 399,5	399	2	0,00061
399,5 - 400,5	400	2	0,00061
Totali		3.259	100

La media aritmetica e la moda, o misura più frequente, della semisomma delle misure estreme delle classi delle misure espresse in mg % di sangue capillare della glicemia dopo due ore dal carico di glucosio sono risultate, rispettivamente, 125,06 e 110.

MARIA GIUSEPPA MANGANO

OFFERTA E DOMANDA DI LAVORO IN SICILIA: UN'INDAGINE SULLE FORZE DI LAVORO

1. Premessa

In questo lavoro ci si prefigge di esaminare, attraverso la documentazione statistica disponibile, l'evoluzione del mercato di lavoro nelle province siciliane.

Verranno prese in esame le caratteristiche strutturali della domanda e dell'offerta di lavoro sotto l'aspetto demografico e socio-economico avvalendosi dei tassi di attività, di occupazione, di disoccupazione e di disoccupazione giovanile.

I tassi sono stati calcolati per sesso e per provincia sia ai censimenti del 1991 e 2001, che nel periodo 2000-2004. Un'attenzione particolare è stata posta all'analisi della disoccupazione attraverso la disamina dei tassi di disoccupazione giovanile, dei tassi di disoccupazione di lunga durata ed, infine, ove possibile dei tassi di disoccupazione per classi di età (1).

2. Le forze di lavoro

Le forze di lavoro, cioè il complesso di persone occupate o in cerca di occupazione (disoccupate), passano in Sicilia da 1.776 migliaia di unità del 2000 – pari a 1.201 mila maschi e 575 mila femmine (7,5%) – a 1.739 migliaia di unità nel 2004 – pari a 1.140 mila maschi e 599 mila femmine (7,1%) –.

Pertanto, nel periodo le forze di lavoro totali diminuiscono di 37 migliaia di unità. Il contributo di tale variazione è molto differenziato per i due sessi, infatti tale decremento sarebbe stato anche maggiore se non fosse aumentata la forza di lavoro femminile che registra una ten-

(1) I dati utilizzati sono forniti dalle pubblicazioni ISTAT.

denza inversa rispetto a quella maschile (-61 migliaia di unità per i maschi, +24 migliaia di unità per le femmine).

Generalmente la variazione dell'offerta di lavoro in un certo intervallo temporale è determinata sia da fattori di carattere demografico, cioè dalla variazione della popolazione in età di lavoro, sia da fattori di carattere socio-economico legati dalla propensione della popolazione in età lavorativa ad offrirsi sul mercato.

A partire dagli anni '70 un elemento fondamentale nella dinamica delle forze lavoro è la spinta derivante dal progressivo aumento della partecipazione femminile al mercato del lavoro, contro un decremento della forza lavoro maschile da imputarsi essenzialmente a fattori di tipo socio-demografico.

3. I tassi di attività tra il 1991 e il 2004

I tassi di attività, espressi dal rapporto tra forze lavoro e la popolazione in età lavorativa, rappresentano una misura sintetica del grado di partecipazione della popolazione alle attività economiche, misurano cioè la propensione al lavoro esistente all'interno di un determinato sistema economico.

Nel decennio 1991-2001 la quota delle persone occupate o alla ricerca di un impiego sul totale delle forze di lavoro (tassi di attività) è aumentata complessivamente passando da 36,8% del 1991 al 42,9% del 2001 (Tav. 1).

Tab. 1 - Tassi di attività per sesso nelle province siciliane ai censimenti 1991-2001.

	MASCHI		FEMMINE		TOTALE	
	1991	2001	1991	2001	1991	2001
Trapani . . .	52,0	56,2	22,6	28,7	36,9	41,9
Palermo . . .	51,6	57,6	22,3	30,2	36,6	43,2
Messina . . .	50,2	55,9	28,7	34,1	39,0	44,5
Agrigento . .	50,0	53,8	23,2	26,8	36,3	39,7
Caltanissetta .	50,2	56,7	19,6	26,8	34,5	41,0
Enna	50,5	54,2	21,1	28,0	35,3	40,4
Catania . . .	52,5	58,9	21,4	30,5	36,5	44,1
Ragusa . . .	53,3	59,5	22,6	30,3	37,6	44,4
Siracusa . . .	52,4	56,3	21,8	28,4	36,9	41,9
SICILIA . . .	51,5	57,0	22,9	30,0	36,8	42,9

Questa crescita è da attribuire soprattutto alla componente femminile; infatti, i tassi di attività relativi al sesso femminile passano da 22,9% del 1991 al 30% del 2001, mentre quelli relativi al sesso maschile registrano un incremento meno significativo.

Nonostante la forte crescita della componente femminile, la differenza tra i tassi di attività globali per gli uomini e per le donne risulta ancora elevata, nel 2001 di ben 27 punti percentuali, a fronte dei 28,6 registrati nel 1991. La differenza è sintomatica del perdurare di un forte ritardo nell'ingresso delle donne nel mercato del lavoro, tipica caratteristica, peraltro, delle popolazioni meridionali.

La sintesi tra la variazione dell'offerta di lavoro e l'evoluzione della popolazione, diversificata per i due sessi, si manifesta attraverso la crescita del tasso di attività, passato da 42,9% nel 2000 a 52,3% nel 2004 con quasi 10 punti di variazione per il totale maschile e femminile (Tav. 2).

Uno sguardo ai tassi provinciali rileva la tendenza al rialzo, comune a tutte le province e per entrambi i sessi, con una punta massima per la provincia di Ragusa (ben 18 punti percentuali per i maschi e qua-

Tab. 2 - Tassi di attività per sesso nelle province siciliane.

	TP	PA	ME	AG	CL	EN	CT	RG	SR	SICILIA
MASCHI										
2000	56,8	59,9	60,8	55,7	63,2	58,6	62,4	64,2	62,4	60,5
2001	58,2	59,9	61,6	54,9	61,7	57,3	60,3	62,8	61,8	59,9
2002	61,3	59,1	59,8	55,6	62,0	57,3	58,9	66,9	61,1	59,8
2003	63,3	57,8	56,9	54,9	63,3	56,9	60,8	68,6	61,8	59,8
2004	72,1	70,6	67,9	67,6	67,7	69,7	68,7	82,1	64,8	69,8
FEMMINE										
2000	23,2	25,3	32,5	21,7	23,5	25,3	29,8	26,7	26,5	26,8
2001	22,9	26,4	34,5	25,4	23,3	25,3	28,1	26,8	25,9	27,2
2002	24,7	24,7	33,9	24,1	22,6	23,1	29,1	27,6	23,8	26,8
2003	26,8	23,6	33,0	25,0	25,0	20,5	28,9	28,6	22,5	26,6
2004	34,2	37,1	40,1	33,5	32,5	35,3	32,1	42,2	32,9	35,5
TOTALE										
2000	39,6	41,7	46,1	37,8	42,5	40,0	45,5	45,2	43,7	42,9
2001	39,9	42,4	47,4	39,3	41,7	40,1	43,5	44,9	43,1	42,9
2002	42,5	41,3	46,2	39,0	41,7	39,0	43,2	46,6	41,6	42,6
2003	44,5	40,1	44,2	39,3	43,4	38,0	44,2	47,9	41,7	42,5
2004	52,9	53,5	53,8	50,2	49,6	52,1	50,0	62,1	48,8	52,3

si 16 punti percentuali per le femmine) e l'incremento minore per la provincia di Siracusa (con una variazione di appena 2,5 punti percentuali per il sesso maschile) e per la provincia di Catania per le femmine con un incremento di poco più di 2 punti percentuali.

Un aspetto particolare che caratterizza la popolazione siciliana è la scarsa partecipazione della popolazione femminile, che, nonostante aumenti nell'arco temporale esaminato, si presenta costantemente inferiore a quella degli uomini in tutte le province.

La minore presenza delle donne sul mercato del lavoro rispetto agli uomini è dunque un dato generalizzato, che ha motivazioni di tipo istituzionale e culturale prima ancora che economico. Infatti, è ancora radicata nella popolazione siciliana, la tendenza ad attribuire agli uomini un diritto-dovere al lavoro rappresentato dal loro ruolo nella famiglia e nella società, mentre le donne restano relegate al ruolo riproduttivo piuttosto che a quello produttivo.

Per le donne, la partecipazione alle attività di mercato resta, purtroppo, ancora un fatto marginale legato ad una serie di condizionamenti quali età, stato civile e livello di istruzione, oltre a condizionamenti di tipo familiare, socio-culturali ed economici.

L'interazione di tutte queste variabili rende alquanto ardua l'analisi dell'offerta di lavoro da parte delle donne e sarebbe stato necessario scindere l'aggregato nelle varie componenti che lo caratterizzano, laddove le statistiche ufficiali lo avrebbero consentito.

Le forze di lavoro hanno mostrato, anche attraverso i dati censuari, una persistente ed ampia tendenza all'incremento nonostante il progressivo ampliarsi dello squilibrio tra domanda e offerta di lavoro.

I tassi di attività globali, sin qui utilizzati, forniscono una prima schematica immagine del comportamento dell'offerta di lavoro al netto dell'evoluzione complessiva della popolazione.

Nell'analisi dei tassi di attività del quinquennio 2000-2004, essi sono, però, influenzati e per certi versi distorti, a livello di confronti spazio-tempo, dalla dinamica demografica che porta mutamenti e differenze sostanziali nella composizione per età della popolazione.

Dall'analisi sin qui compiuta, sembra potersi evidenziare il particolare comportamento che la componente femminile ha nella partecipazione al lavoro. Il comportamento femminile ha subito le variazioni più rilevanti, determinando la misura in cui l'offerta potenziale di lavoro, rappresentata dalla popolazione in età lavorativa, si traduce in effettiva offerta.

Si è visto, inoltre, come il comportamento femminile nei confronti del mercato del lavoro, e le trasformazioni che esso ha subito nel tempo, sia complesso e difficile da esaminare, data la molteplicità di fattori, operanti e dal lato dell'offerta e dal lato della domanda.

Anche se la Sicilia è il fanalino di coda dell'Italia, l'aumento della propensione femminile al lavoro, intervenuto nel decennio 1991-2001 e confermato nell'ultimo quinquennio 2000-2004, è da ricercare soprattutto in relazione a fattori sociali e culturali, legati all'emancipazione femminile ed al nuovo ruolo che la donna tende sempre più a rivestire nella famiglia e nella società. Indubbiamente a tali fattori sono da aggiungere altri, non meno importanti, quali la progressiva riduzione della fecondità ed il miglioramento del grado di istruzione.

4. I tassi di occupazione tra il 1991 e il 2004

L'analisi dei tassi di occupazione serve a confrontare l'occupazione con la potenziale offerta di lavoro rappresentata essenzialmente dalla popolazione in età lavorativa. La Tav. 3 riporta i tassi di occupazione, espressi dal rapporto tra occupati e popolazione in età superiore a 15 anni.

Attraverso i dati degli ultimi due censimenti (1991-2001) si vede che i tassi di occupazione presentano analogo andamento a quelli relativi alle forze di lavoro ma con un gap di gran lunga inferiore. A fronte della variazione dei 10 punti percentuali verificatisi tra il 2000 ed il 2004 riscontriamo una variazione di appena 2 punti percentuali.

Tab. 3 - Tassi di occupazione per sesso nelle province siciliane ai censimenti 1991-2001.

	MASCHI		FEMMINE		TOTALE	
	1991	2001	1991	2001	1991	2001
Trapani . . .	45,9	45,5	14,2	18,9	29,4	31,6
Palermo . . .	43,3	43,1	15,6	18,9	28,9	30,4
Messina . . .	44,6	45,4	20,6	24,3	32,0	34,3
Agrigento . .	40,9	40,7	14,3	15,9	27,1	27,8
Caltanissetta .	43,6	45,2	12,2	16,8	27,2	30,3
Enna	42,3	41,8	14,0	17,9	27,5	29,2
Catania . . .	47,1	45,5	15,5	20,0	30,7	32,0
Ragusa . . .	51,7	51,7	16,7	22,8	33,6	36,7
Siracusa . . .	47,4	44,6	14,0	18,0	30,3	30,9
SICILIA . . .	45,1	44,6	15,7	19,5	29,8	31,5

Il valore più elevato, relativo al totale dei sessi, lo troviamo per il 2001 sempre in corrispondenza di Ragusa con 36,7% a fronte del 31,5% della media regionale.

I valori più elevati sono riportati dagli uomini i cui tassi oscillano tra il 41,8% della provincia di Enna nel 2001 ed il 51,7% della provincia di Ragusa.

Uno sguardo ai tassi di occupazione del sesso maschile evidenzia una stazionarietà degli stessi, se non addirittura in alcuni casi una lieve diminuzione, nel decennio 1991-2001.

Andamento alquanto diverso presentano invece i tassi di occupazione relativi al sesso femminile che si attestano tutti in misura crescente mediamente di 4 punti tra il 1991 ed il 2001, ad eccezione della provincia di Agrigento +1,6 circa e della provincia di Ragusa + 6 punti percentuali circa.

La Tab. 4 riporta i tassi di occupazione, per sesso e per provincia, calcolati sulle forze di lavoro per il periodo 2000-2004.

Dalla citata tavola risulta che in Sicilia nel 2000 il 32,6 % della popolazione attiva è occupata, tale valore cresce gradualmente fino al 2003 e registra, poi, un'impennata nel 2004 attestandosi a 43,2%.

Tab. 4 - Tassi di occupazione per sesso nelle province siciliane.

	TP	PA	ME	AG	CL	EN	CT	RG	SR	SICILIA
MASCHI										
2000	50,6	45,4	47,7	45,4	54,0	45,0	50,4	58,4	54,3	49,2
2001	52,4	47,0	50,4	45,8	51,7	47,7	48,8	57,9	56,2	49,9
2002	54,8	46,6	49,9	46,0	52,1	49,0	48,9	60,3	55,5	50,1
2003	57,3	46,2	47,1	46,1	54,4	47,3	50,5	61,8	54,7	50,3
2004	63,4	59,3	58,8	55,5	55,1	57,6	59,9	78,0	56,4	60,0
FEMMINE										
2000	17,2	15,6	21,6	12,8	15,0	13,3	19,1	19,7	18,6	17,4
2001	18,4	17,8	23,9	15,7	14,2	15,8	19,2	20,2	19,1	18,8
2002	20,0	17,6	24,7	16,1	13,9	15,4	19,8	22,1	18,7	19,2
2003	21,4	16,8	23,1	16,9	15,7	13,7	19,7	22,3	18,5	19,0
2004	26,1	26,5	31,7	24,9	24,6	24,7	25,7	36,4	23,9	27,0
TOTALE										
2000	33,4	29,7	34,1	28,3	33,7	28,2	34,2	39,1	35,8	32,6
2001	34,7	31,8	36,5	30,0	32,2	30,6	33,4	38,9	36,9	33,7
2002	36,9	31,6	36,7	30,3	32,3	31,1	33,7	40,8	38,3	34,0
2003	38,8	30,9	34,4	30,8	34,4	29,8	34,4	41,4	35,9	34,0
2004	44,4	42,5	45,1	39,9	39,4	40,7	42,4	57,1	40,1	43,2

La dinamica occupazionale si manifesta molto diversificata tra i due sessi, infatti la componente maschile presenta sia nel tempo che nello spazio tassi di gran lunga superiori a quelli relativi al sesso femminile.

Il tasso di occupazione medio maschile è pari a 49,2% nel 2000 e raggiunge quota 60% nel 2004, ad esso si contrappone un tasso di occupazione medio femminile che oscilla tra il 17,4% del 2000 al 27% del 2004.

Se si passa ad osservare i valori dei tassi di occupazione maschili a livello provinciale si rileva come il valore più elevato è da imputare alla provincia di Ragusa per tutti gli anni osservati, variando dal 58,4% del 2000 al 78% del 2004. Il valore più basso si riscontra in corrispondenza della provincia di Palermo, dove assume valori pari a 45,4% nel 2000, 47% nel 2001, 46,6% nel 2003. Nel 2004 invece il valore più basso si riscontra in corrispondenza della provincia di Caltanissetta con 55,1% di occupati sul totale della popolazione superiore a 14 anni.

Diversa è la situazione per le donne, il cui primato, per i primi 4 anni spetta alla provincia di Messina che viene "sorpasata" dalla provincia di Ragusa con quota 36,4% nel 2004.

Peculiare risulta il comportamento delle donne il cui tasso di occupazione, per tutte le province e per tutti gli anni, si presenta poco più di 1/3 rispetto a quello degli uomini. Ciò conferma quanto esplicitato anche dai tassi di attività: la scarsa propensione delle donne a presentarsi sul mercato del lavoro sebbene questa propensione sia cresciuta nel tempo.

Dal confronto con i tassi di attività emerge come, per entrambi i sessi, l'offerta cresce in misura maggiore rispetto alla domanda. Ciò non significa naturalmente che queste variazioni siano legate soprattutto ad una crescita della domanda, piuttosto, come si è avuto modo di rilevare, per le donne influiscono una molteplicità di fattori relativi a tradizioni e modelli di comportamento variati nel tempo.

Ovviamente, considerati i valori dei tassi di attività e di occupazione, soprattutto femminili, si è ancora molto lontani da una reale possibilità di inserimento nel mercato del lavoro da parte di tutte le donne che oggi sembra riservato a poche.

E' sembrato opportuno calcolare, anche, il peso degli occupati su tutta la popolazione (Tab. 5), e si è riscontrato che, mediamente, poco più di 40 persone di sesso maschile su 100 residenti risultano occupate a fronte di 15 persone per il sesso femminile.

Tab. 5 - Peso occupati su totale popolazione nelle province siciliane.

	TP	PA	ME	AG	CL	EN	CT	RG	SR	SICILIA
MASCHI										
2000	43,2	37,0	40,2	37,9	43,0	37,4	41,4	48,6	43,7	40,5
2001	43,3	38,4	42,1	38,8	41,7	38,8	41,1	48,6	45,4	41,3
2002	44,6	38,1	42,0	38,1	42,7	38,9	40,9	50,4	44,8	41,3
2003	46,7	37,3	39,2	38,0	44,4	40,4	42,1	48,9	44,2	41,2
2004	41,9	40,3	40,2	37,1	37,0	38,2	40,5	51,7	39,0	40,5
FEMMINE										
2000	14,6	13,1	18,3	11,1	11,9	10,8	15,8	15,9	15,8	14,5
2001	15,5	14,7	20,4	13,8	11,3	13,0	16,3	16,5	16,3	15,8
2002	16,4	14,4	21,3	13,8	11,3	13,1	17,0	18,4	15,8	16,1
2003	17,2	13,7	19,3	14,4	12,6	12,1	16,9	18,1	15,3	15,7
2004	16,6	17,4	20,5	16,1	16,1	15,4	16,8	23,1	16,2	17,6
TOTALE										
2000	28,4	24,6	28,9	24,1	26,9	23,6	28,2	31,8	29,4	27,1
2001	28,9	26,1	30,8	25,7	26,3	25,4	28,3	32,2	30,5	28,1
2002	30,1	25,8	31,2	25,5	26,4	25,5	28,4	34,0	30,0	28,3
2003	31,5	25,1	28,8	25,8	27,9	25,7	29,0	33,2	29,4	28,1
2004	28,9	28,4	30,1	26,3	25,8	26,4	28,4	37,2	27,4	28,7

Ciò a ulteriore conferma che la fetta maggiore del mercato di lavoro viene assorbita dagli uomini.

5. I tassi di disoccupazione tra il 1991 e il 2004

Nella Tav. 6 e segg. sono riportati i tassi disoccupazione dati dal rapporto tra le persone in cerca di occupazione e la forza di lavoro.

Generalmente, il tasso di disoccupazione, in quanto indicatore prettamente congiunturale, si presta, essenzialmente, ad un'analisi di breve periodo.

La dinamica dei tassi di disoccupazione non sempre è conseguenza della dinamica della domanda di lavoro (misurata indirettamente dal numero di occupati) ma è ovvio che le variazioni avvenute in essa, e quelle relative alla struttura per età della popolazione, hanno un peso determinante per la valutazione del fenomeno della disoccupazione.

La disoccupazione è un fenomeno alquanto complesso in quanto in essa interagiscono vari fattori sia economici che demografici; la sua analisi non può, quindi, prescindere dall'esame congiunto di variabili

demografiche (ammontare e struttura per età della popolazione) ed economico-sociali (andamento forze-lavoro e occupazione).

Per le motivazioni fin qui esposte, la riflessione sulla disoccupazione si colloca alla fine di questa analisi ed è articolata in modo più ampio rispetto ai fenomeni prima trattati, al fine di meglio cogliere il peso che tale fenomeno assume sulla segmentazione dell'offerta di lavoro.

Uno sguardo ai dati censuari (Tav. 6) rimarca il decremento dei tassi di disoccupazione per tutte le province e per entrambi i sessi, sempre a favore del sesso maschile, con punta minima in ciascun sesso e per il complesso dei sessi per la provincia di Ragusa (rispettivamente 13,2% - 24,9% - 17,3%) e valore massimo relativo alla provincia di Agrigento (rispettivamente 24,3% - 40,5% e 30,0%).

Prima di passare all'esame puntuale dei tassi di disoccupazione, per il periodo 2000-2004, è utile sottolineare l'andamento delle forze di lavoro e delle persone in cerca di occupazione, espressi in valore assoluto, per l'intera Regione.

All'inizio è stato riportato il valore delle forze di lavoro che evidenzia una diminuzione di 37 migliaia di unità tra il 2000 e il 2004 per il complesso della popolazione siciliana. Analogo andamento, ma con una intensità maggiore, avviene per le persone in cerca di occupazione che passano da 426 mila del 2000 (pari a 225 mila uomini e 201 mila donne) a 300 mila del 2004 (pari a 158 mila uomini e 142 mila donne).

È facilmente intuibile che tanto più elevato è il divario tra questi segmenti di popolazione tanto più ne risentirà il tasso di disoccupazione che è il rapporto tra le due quote.

Tab. 6 - Tassi di disoccupazione per sesso nelle province siciliane ai censimenti 1991-2001.

	MASCHI		FEMMINE		TOTALE	
	1991	2001	1991	2001	1991	2001
Trapani . . .	29,4	19,0	48,6	34,2	35,4	24,5
Palermo . . .	34,6	25,1	43,9	37,2	37,5	29,5
Messina . . .	28,5	18,9	40,5	28,8	33,1	22,9
Agrigento . .	35,5	24,3	50,3	40,5	40,3	30,0
Caltanissetta .	32,2	20,4	50,1	37,1	37,4	26,1
Enna	33,5	23,0	46,2	36,0	37,4	27,7
Catania . . .	30,3	22,8	42,0	35,5	33,8	27,4
Ragusa . . .	22,8	13,2	39,8	24,9	28,0	17,3
Siracusa . . .	28,3	20,9	48,1	36,6	34,2	26,4
SICILIA . . .	31,2	21,8	44,4	34,8	35,4	26,5

Detto ciò la Tav. 7 evidenzia un andamento al ribasso dei tassi di disoccupazione, tale decremento è comune ad entrambi i sessi e per tutte le province, ad eccezione della provincia di Trapani per il sesso maschile (11,1% nel 2000, 12% nel 2004) e della provincia di Caltanissetta (14,8% nel 2000 e 18,5% nel 2004).

Scomponendo l'indicatore in base al sesso esso si presenta in maniera composita, infatti, i tassi di disoccupazione femminile si attestano a quasi il doppio rispetto a quelli della componente maschile.

La disoccupazione femminile incide in misura ben più rilevante, e ciò esprime, in modo evidente l'incapacità della struttura produttiva di assorbire questa considerevole quota di lavoro disponibile.

Quanto fin qui osservato, non intende sottovalutare la presenza e la rilevanza di meccanismi sociali tipici delle zone meridionali, la cui incidenza è, già, stata sottolineata.

Gli elevati tassi di disoccupazione femminile, in confronto a quelli maschili, esprimono certamente, in misura rilevante, anche la delusione delle aspettative delle donne che lottano per porsi allo stesso status degli uomini.

Tab. 7 - Tassi di disoccupazione per sesso nelle province siciliane.

	TP	PA	ME	AG	CL	EN	CT	RG	SR	SICILIA
MASCHI										
2000	11,1	24,1	21,6	18,5	14,8	22,1	19,3	9,2	13,2	18,7
2001	10,2	21,5	18,3	16,1	16,1	16,3	18,9	8,4	9,1	16,8
2002	10,3	21,1	16,5	17,1	16,0	14,7	16,8	9,7	8,8	16,0
2003	9,5	20,4	16,9	16,0	13,7	17,2	16,9	9,7	11,8	15,9
2004	12,0	15,9	13,2	18,0	18,5	17,2	12,6	4,9	12,9	13,8
FEMMINE										
2000	26,3	38,3	33,8	40,4	36,0	46,8	35,7	25,2	29,5	35,0
2001	20,4	32,6	30,7	38,2	38,6	36,9	31,4	24,8	26,9	31,2
2002	19,3	28,6	26,9	33,6	39,3	33,2	31,9	20,4	21,7	28,4
2003	20,6	28,7	30,1	32,2	37,1	35,1	32,0	22,6	18,7	28,9
2004	23,7	28,5	21,0	25,3	24,3	29,8	19,9	13,8	26,8	23,7
TOTALE										
2000	15,7	28,6	26,1	25,1	20,8	30,2	24,8	13,9	18,3	24,0
2001	13,3	25,1	23,0	23,6	22,6	23,2	23,1	13,3	14,7	21,5
2002	13,0	23,4	20,5	22,4	22,6	20,5	22,1	13,0	12,7	20,1
2003	12,9	22,9	22,1	21,4	20,6	22,3	22,0	13,7	13,8	20,1
2004	15,8	20,3	16,1	20,4	20,4	21,6	15,0	8,0	17,7	17,2

Relativamente al 2004, il ventaglio di valori cui si attestano i tassi è abbastanza ampio passando da un minimo della provincia di Ragusa 4,9% ad un massimo della provincia di Caltanissetta 18,5% per il sesso maschile; per il sesso femminile i valori oscillano da 13,8% della provincia di Ragusa a 29,8% della provincia di Enna.

I dati dei censimenti hanno permesso di osservare, anche, la disoccupazione giovanile cioè quella quota parte di disoccupazione relativa alle persone di età 15-29 (Tav. 8).

Tab. 8 - Tassi di disoccupazione giovanile per sesso nelle province siciliane ai censimenti 1991-2001.

	MASCHI		FEMMINE		TOTALE	
	1991	2001	1991	2001	1991	2001
Trapani . . .	54,1	49,8	76,5	69,9	62,7	58,3
Palermo . . .	63,3	64,1	72,8	76,2	66,8	69,3
Messina . . .	55,1	53,4	67,0	67,4	60,3	59,5
Agrigento . .	62,0	62,0	77,0	78,6	67,7	68,8
Caltanissetta .	57,7	54,5	78,2	73,7	64,9	62,1
Enna	57,7	57,3	76,5	74,3	64,6	64,3
Catania . . .	54,1	56,0	68,9	71,6	59,2	62,4
Ragusa	41,7	34,7	65,5	53,9	50,3	42,2
Siracusa . . .	53,4	55,4	76,4	73,4	61,8	62,8
SICILIA . . .	56,8	56,3	72,2	72,1	62,6	62,9

Dal confronto tra i tassi globali di disoccupazione e quelli giovanili risulta immediata e chiara la difficoltà dei giovani ad inserirsi nel mondo del lavoro, infatti i tassi di disoccupazione giovanile sono alquanto elevati per tutte le province.

Dall'esame dei dati nell'arco di tempo considerato, si è visto dunque che i tassi mostrano un andamento decrescente: non ci sembra azzardato attribuire questo calo più alla variazione avvenuta nella struttura per età della popolazione, che tende ad invecchiare per effetto della diminuita natalità, che ad un peggioramento della situazione economica.

La minore disoccupazione, in termini di tassi, si presenta come una costante relativamente alla provincia di Ragusa, invece il valore massimo oscilla, posizionandosi nella provincia di Palermo per il sesso maschile sia nel 1991 (63,3%) che nel 2001 (64,1%) e per il complesso dei sessi rispettivamente 66,8% e 69,3%; il dato fa eccezione per le donne dove la punta massima è raggiunta nella provincia di Caltanissetta per il 1991 (78,2%) e di Agrigento per il 2001 (78,6%).

Le differenze dei tassi di disoccupazione giovanile tra maschi e femmine (Tav. 9) evidenziano valori sempre a sfavore, come era prevedibile, delle donne con il massimo nella provincia di Ragusa per il 1991 (23,8%) e nella provincia di Trapani per il 2001 (20,2%).

Tab. 9 - Differenze tra tassi di disoccupazione giovanile.
Differenza assoluta tra tassi di disoccupazione giovanile maschile e femminile.

	1991	2001
Trapani	22.4	20.2
Palermo	9.5	12.1
Messina	11.9	14.0
Agrigento	15.0	16.6
Caltanissetta	20.5	19.2
Enna	18.8	17.0
Catania	14.8	15.5
Ragusa	23.8	19.2
Siracusa	23.0	17.9
SICILIA	15.4	15.8

Vediamo ora analiticamente, l'andamento dei tassi di disoccupazione, secondo la variabile demograficamente più rilevante quale è l'età, per gli anni 2000-2003.

Le classi di età giovanili sono quelle maggiormente interessate al fenomeno disoccupazione, intendendo per classi giovanili quelle in età 25-29. Se ci soffermiamo ad osservare i dati del 2000 la maggiore concentrazione dei tassi di disoccupazione, per i maschi, si colloca nelle classi di età 15-24 (51,8%) e diminuisce nelle classi 25-29 (41%). Tale struttura determina il maggior peso, nell'ambito della disoccupazione giovanile, delle classi di età più giovani ed attesta l'ingresso tardivo sul mercato del lavoro, sintomo di una esasperata penuria di sbocchi occupazionali per i più giovani.

Se guardiamo la componente femminile, la quota di disoccupati aumenta: 69,9% nelle classi 15-24 e 62,7% nelle classi di età 25-29.

Negli altri anni osservati i tassi di disoccupazione per età presentano analogo andamento con maggiore concentrazione nelle classi di età 15-24 (47,2% per i maschi nel 2001): tale incidenza aumenta sempre per le donne.

La quota di disoccupati tra il 2000 ed il 2003 presenta un andamento decrescente, più marcato per il sesso femminile; meno 3 punti

percentuali per le classi di età 15-64 per i maschi (rispettivamente 19% e 16%) e meno 6 punti percentuali per le donne (rispettivamente 35,3% e 29%).

Volendo sintetizzare gli elementi salienti sin qui emersi, la disoccupazione siciliana, a prescindere dalle sue dimensioni, rivela connotati di gravità rispetto alla variabile sesso.

Nonostante vi sia un cospicuo coinvolgimento delle forze-lavoro maschile, ciò sembra meno grave rispetto alla situazione delle forze lavoro femminili.

La disoccupazione femminile in Sicilia tra il 2000 ed il 2003 diminuisce anche se si attesta attorno a valori preoccupanti.

Una caratteristica alquanto significativa della disoccupazione è rappresentata da quella relativa alla sua durata in relazione alla percezione immediata delle difficoltà che incontrano le persone che cercano lavoro ad inserirsi nell'area degli occupati.

È lecito supporre che la durata della disoccupazione è direttamente proporzionale alle difficoltà di inserimento nel mercato del lavoro. Tale difficoltà si ricollega al deprezzamento del capitale umano subito da chi, per lungo tempo, è estromesso dal mercato del lavoro.

Diversamente dalla disoccupazione in senso stretto si presentano i tassi di disoccupazione di lunga durata, cioè i tassi relativi a persone che cercano lavoro da più di 12 mesi. Questi ultimi tra il 1995 ed il 2003 aumentano passando da 57,1% a 66,6% per il complesso dei sessi; variano da 59,8% del 1995 a 71% del 2003 per la quota parte relativa alle donne.

6. Conclusioni

Sebbene sia da ritenersi poco sufficiente l'approfondimento di specifici aspetti relativi al mercato del lavoro siciliano, il presente lavoro ha, tuttavia, delineato caratteri e tendenze relativi al territorio esaminato ed ha messo in luce la complessità crescente di talune problematiche.

Sul piano strutturale è emersa la situazione sfavorevole per la componente femminile con conseguente difficoltà delle donne a collocarsi nel sistema produttivo: ciò ha radici storiche e motivazioni sociali e culturali oltre che economiche.

Il rapporto occupati/popolazione penalizza anche i giovani di sesso maschile, e sembra che soltanto gli uomini delle classi di età cen-

trali, anche se con crescente difficoltà, trovino sufficienti occasioni di lavoro.

Determinante, nel rapporto domanda/offerta di lavoro, è il fattore demografico in quanto la popolazione siciliana presenta una struttura per età che tende ad invecchiare essenzialmente per la denatalità oltre che per le migrazioni.

La dinamica fortemente espansiva dei tassi di sesso femminile è il fenomeno che maggiormente ha caratterizzato il mercato di lavoro in epoca recente. Il comportamento delle donne siciliane segue le trasformazioni culturali in atto nelle società evolute.

ANTONINO MINEO, ANGELO M. MINEO

ANALISI DELLE STRUTTURE DI INTERDIPENDENZA CORRELATIVA E ASSOCIATIVA E DELLA VARIABILITÀ RELATIVA DI VARIABILI STATISTICHE MULTIPLE (*)

1. Premesse metodologiche al problema

I concetti di intensità e di variabilità di una serie di valori di una variabile statistica semplice e le loro misure sintetiche, valori medi e indici di variabilità, espressi nelle stesse unità di misura dei valori, risultano solitamente di immediata comprensione. Qualche problema sorge quando si è in presenza di due serie di valori, relativi a due variabili semplici, e si vogliono confrontare le misure sintetiche dei due aspetti precedenti o misurare la relazione di interdipendenza tra le due variabili, mediante, ad esempio, la misura della covarianza. In questo caso risulta spesso difficile effettuare il confronto tra le due serie di valori o interpretare il significato della covarianza per almeno due ordini di motivi:

- 1) le due serie potrebbero riferirsi a valori espressi mediante unità di misura diverse (si pensi, ad esempio, al confronto della variabilità tra una serie di valori relativi alle altezze e una serie di valori relativi ai pesi di n soggetti);
- 2) le due serie, pur riferendosi a valori espressi nella stessa unità di misura, possono presentare ordini di grandezza diversi (si pensi, ad esempio, ad una serie di valori relativi al peso delle madri subito dopo il parto e una serie di valori relativi al peso dei corrispondenti neonati).

Per risolvere il primo problema si ricorre, come è noto, ai cosiddetti indici di variabilità relativa, ottenuti riscaldando gli indici di variabilità assoluta con la grandezza media dei valori, ottenendo il duplice

(*) Questa ricerca è stata parzialmente finanziata da fondi MURST. Inoltre il lavoro è frutto della stretta collaborazione tra i due autori ed entrambi ne condividono il contenuto; è corretto rilevare comunque che i paragrafi 1, 2 e 5 sono in comune, il paragrafo 3 è stato sviluppato da Antonino Mineo, mentre il paragrafo 4 è stato sviluppato da Angelo M. Mineo.

risultato di eliminare l'effetto della diversa unità di misura, perché espressi in numeri puri (punto 1.), e dell'ordine medio di grandezza (punto 2.).

Per quanto concerne il secondo problema, l'impiego della covarianza quale misura della interdipendenza delle due variabili, espressa dalla media dei prodotti degli scarti di ciascuna variabile dal proprio valore medio:

$$\sigma_{xy} = \frac{\sum_{i=1}^n (x_i - M_x)(y_i - M_y)}{n} \quad (1)$$

risulterà misurata anch'essa dal prodotto delle due unità di misura che, se sono diverse, determinerà un'unità di misura di non facile significato fisico. Si pensi, ad esempio, a due variabili, delle quali una espressa in centimetri e l'altra in chilogrammi; che significato fisico si può attribuire a tale misura espressa in $Kg \cdot cm$ o in $cm \cdot Kg$? Se inoltre si considera la matrice

$$\begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix} \quad (2)$$

a seconda dei significati che si vogliono dare ai simboli, può essere intesa:

- a) come “matrice delle covarianze” delle due variabili X e Y per indicare tutte le possibili interdipendenze (concordanze o discordanze tra le due serie di valori x_i e y_i), per cui σ_{xx} e σ_{yy} indicano le covarianze delle x_i e delle y_i con se stesse, mentre $\sigma_{xy} = \sigma_{yx}$ indica la covarianza tra le x_i e le y_i ; oppure
- b) come “matrice di varianze” delle due variabili X e Y per indicare con $\sigma_{xx} = \sigma_x^2$ e $\sigma_{yy} = \sigma_y^2$ le varianze delle due variabili (in quanto valori massimi di σ_{xx} e σ_{yy}) e $\sigma_{xy} = \sigma_{yx}$ la “componente” di varianza (concordanza tra i valori x_i e y_i) contenuta nelle due serie di valori, che può variare tra un minimo e un massimo compreso tra i valori $-\sigma_x \sigma_y$ e $\sigma_x \sigma_y$.

In questi due possibili significati (o interpretazioni) si pone in evidenza nel primo caso l'aspetto della interdipendenza tra le due variabili, nel secondo caso l'aspetto della variabilità contenuta nelle due variabili; così, nel primo caso appare lecito l'appellativo di “matrice delle covarianze”, nel secondo caso quello di “matrice delle varianze”, ma in nessun caso appare lecito chiamarla “matrice di varianze e covarianze”, perché farebbe riferimento a due significati diversi delle quan-

tà riportate nella matrice. Inoltre, l'aver posto nella (2) le possibili covarianze tra le due variabili, ha prodotto un nuovo aspetto della struttura di interdipendenza tra le due variabili il cui significato sarà chiaro nelle pagine successive.

Quindi, per misurare la concordanza tra le due serie di valori si ricorre alla standardizzazione delle variabili, mediante la nota relazione

$z_{x_i} = \frac{x_i - M_x}{\sigma_x}$, che fornisce valori z_{x_i} che sono adimensionali, perché

espressi da numeri puri, e che non risentono dell'effetto del diverso ordine medio di grandezza; questo comporta un vantaggio anche sulla variabilità. In questo modo, il calcolo della covarianza tra le due variabili standardizzate $z_{x_i} = \frac{x_i - M_x}{\sigma_x}$ e $z_{y_i} = \frac{y_i - M_y}{\sigma_y}$ coincide, come è noto,

con il coefficiente di correlazione lineare di Bravais-Pearson calcolato per le due variabili X e Y , per cui risulta:

$$r_{z_{x_i}, z_{y_i}} = \text{cov}(z_{x_i}, z_{y_i}) = r_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (3)$$

Se, ora, nella (2) sostituiamo le covarianze con le correlazioni, per la (3) possiamo scrivere la matrice di correlazione equivalente

$$\begin{bmatrix} 1 & r_{xy} \\ r_{yx} & 1 \end{bmatrix} \quad (4)$$

che ha la stessa struttura di interdipendenza della (2), con l'ovvia differenza che quando si impiegano le correlazioni nel significato del punto b), cioè di "varianze" tra le variabili standardizzate, si deve parlare non di varianze assolute, bensì di varianze relative espresse in numeri puri. Ne consegue che sulla diagonale principale della (4) troviamo valori unitari, cioè le varianze delle variabili standardizzate, mentre $r_{xy} = r_{yx}$ indicano la "componente" di varianza relativa contenuta nelle due serie di valori (z_{x_i}, z_{y_i}) , che può variare entro limiti compresi tra -1 e $+1$.

Il problema fin qui esposto si complica notevolmente quando si vuole esaminare la variabilità e la struttura di interdipendenza di variabili statistiche multiple omogenee (1) con ordine di molteplicità $k > 2$.

(1) Una variabile multipla si dice "omogenea" se tutte le k variabili semplici che la compongono sono misurabili nella stessa scala di misura, ad esempio una scala di intervallo o una scala di rapporto.

In questo caso la matrice delle covarianze

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \dots & \sigma_{kk} \end{bmatrix} \quad (5)$$

conserva i due significati a) e b), gli inconvenienti lamentati per i motivi 1) e 2) e, in più, presenta un'ulteriore complicazione nella ricerca delle due tipologie di strutture di interdipendenza, perché esse interagiscono tra loro e anche con la misura della variabilità multipla. Anche in questo caso una semplificazione del problema, che consente di analizzarlo più agevolmente, è fornito dalla matrice di correlazione:

$$\begin{bmatrix} 1 & r_{12} & \dots & r_{1k} \\ r_{21} & 1 & \dots & r_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & \dots & 1 \end{bmatrix} \quad (6)$$

per la quale è ancora valida la relazione (3) tra i coefficienti di correlazione tra le coppie delle variabili semplici X_1, X_2, \dots, X_k che compongono la variabile multipla e le covarianze tra le coppie corrispondenti delle variabili standardizzate $z_{X_1}, z_{X_2}, \dots, z_{X_k}$.

Il presente lavoro è così organizzato: nel prossimo paragrafo introdurremo il metodo che seguiremo per analizzare le strutture di interdipendenza e per misurare la variabilità relativa multipla. Nel paragrafo 3 presenteremo le soluzioni al problema particolare, difficilmente riscontrabile in pratica, per l'analisi di una variabile multipla, in cui le coppie di variabili semplici risultano tutte perfettamente correlate. Dopo aver indicato che cosa intendiamo per struttura di interdipendenza "correlativa" e struttura di interdipendenza "associativa", procederemo alla loro valutazione e alla misura della variabilità relativa multipla, proponendo infine un indice di variabilità relativa multipla che varia tra 0 e 1 e consente confronti tra variabili multiple con k diverso. Nel paragrafo 4 esamineremo il problema generale, utile in pratica, di variabili multiple in cui le coppie di variabili semplici sono diversamente correlate. Nel paragrafo 5 presenteremo un'applicazione dei criteri e metodi più importanti del paragrafo 4 ad un problema ormai classico in

letteratura, mostrando la validità dei risultati conseguiti. Infine nel paragrafo 6 presenteremo alcune considerazioni conclusive.

2. Introduzione al metodo

Si abbia una matrice di $(n \times k)$ osservazioni x_{ij} , $i = 1, 2, \dots, n$, $j = 1, 2, \dots, k$ relative a n unità statistiche (oggetti o soggetti) sulle quali sono state rilevate le modalità quantitative (valori) di k caratteri e si voglia studiare la struttura di interdipendenza correlativa e/o associativa (2) tra le k variabili e misurarne la variabilità multipla. La soluzione a questi due problemi può essere data dalla determinazione della matrice delle covarianze σ_{ij} , i e $j = 1, 2, \dots, k$ che, per le considerazioni fatte nel paragrafo precedente sulla influenza delle diverse unità di misura e del diverso ordine medio di grandezza che quasi sempre caratterizzano i valori delle variabili semplici che compongono la variabile multipla, fornisce valori informativi difficilmente interpretabili e per questo difficilmente utilizzabili in pratica.

Come abbiamo visto, il modo migliore per superare tali difficoltà consiste nella determinazione della matrice delle k^2 correlazioni in coppia, r_{ij} , tra le k variabili X_1, X_2, \dots, X_k , cioè la matrice delle covarianze tra le k variabili standardizzate $z_{X_1}, z_{X_2}, \dots, z_{X_k}$.

Se si orla la matrice di correlazione $\{r_{ij}\}$ con i totali marginali

$$R_i = 1 + \sum_{\substack{j=1 \\ j \neq i}}^k r_{ij} \quad (7)$$

$$R_j = 1 + \sum_{\substack{i=1 \\ i \neq j}}^k r_{ij} \quad (8)$$

e con il totale generale

$$R_{..} = \sum_{i=1}^k R_i = \sum_{j=1}^k R_j = k + \sum_{\substack{i=1 \\ i \neq j}}^k \sum_{\substack{j=1 \\ j \neq i}}^k r_{ij} = k + 2 \sum_{i < j}^k r_{ij} \quad (9)$$

(2) Con il termine *interdipendenza correlativa* intendiamo la struttura di interdipendenza tra tutte le possibili correlazioni tra coppie di variabili (X_i, X_j) , mentre con il termine *interdipendenza associativa* intendiamo la struttura di interdipendenza tra ciascuna variabile e tutte le altre.

e si calcola la colonna delle proporzioni $P_i = R_i/R_{..}$ si ottiene la Tab. 1:

Tab. 1 - Matrice di correlazione.

$\{r_{ij}\}$				R_i	P_i
1	r_{12}	\dots	r_{1k}	$R_{1.}$	$P_{1.}$
r_{21}	1	\dots	r_{2k}	$R_{2.}$	$P_{2.}$
\vdots	\vdots	\ddots	\vdots	\vdots	\vdots
r_{k1}	r_{k2}	\dots	1	$R_{k.}$	$P_{k.}$
$R_{.1}$	$R_{.2}$	\dots	$R_{.k}$	$R_{..}$	1

dalla quale si ottengono i seguenti elementi descrittivi:

- 1) i singoli coefficienti di correlazione r_{ij} , $i < j$, che descrivono la struttura e l'intensità dell'interdipendenza correlativa tra le k variabili X_i e l'intensità della variabilità di ciascuna coppia di variabili (3) (z_{X_i}, z_{X_j});
- 2) i k totali marginali R_i che forniscono le interdipendenze associative di ciascuna variabile con tutte le altre;
- 3) il totale generale $R_{..}$ che fornisce la variabilità multipla contenuta nella matrice di correlazione data dalla somma delle due quantità:
 - (a) la somma k degli elementi sulla diagonale principale che rappresentano ciascuno il valore massimo, 1, della covarianza di ciascuna variabile con se stessa che, come abbiamo detto, coincide con la componente di varianza di una variabile standardizzata;
 - (b) la somma, $\sum_{i \neq j} r_{ij}$, delle $k(k-1)$ covarianze relative contenute nei rimanenti elementi r_{ij} della matrice di correlazione che, per la simmetria della matrice, è uguale a $2 \sum_{i < j} r_{ij}$ e che coincide con la componente di varianza (concorrenza) contenuta nelle covarianze tra tutte le coppie di variabili standardizzate z_{X_i}, z_{X_j} , $i \neq j$.

È da notare che i k totali marginali R_i possono risultare uguali o diversi tra loro in relazione alla struttura di interdipendenza correlativa descritta dai $k(k-1)/2$ valori r_{ij} , $i < j$, corrispondenti agli elementi della

(3) Per la nota proprietà della simmetria della matrice di correlazione, la struttura dell'interdipendenza correlativa è uguale sia se si considerano i $k(k-1)/2$ valori di r_{ij} , $i < j$, della semimatrice superiore, sia i $k(k-1)/2$ valori di r_{ij} , $i > j$, della semimatrice inferiore. Per questo nelle pagine successive faremo sempre riferimento agli r_{ij} della semimatrice superiore.

semimatrice superiore. In particolare, se si suppone che $r_{ij} = r, \forall i \neq j$, cioè se gli r_{ij} sono tutti uguali tra loro, anche gli R_i e i P_i risulteranno uguali tra loro e risulterà $P_i = \frac{1}{k}$. Se gli r_{ij} sono diversi tra loro, lo saranno anche i k valori R_i e P_i .

Inoltre, poiché nelle applicazioni pratiche possono ottenersi infinite strutture di interdipendenza tra k variabili, descritte ciascuna da una particolare matrice di correlazione, l'intensità della interdipendenza correlativa può variare da un valore minimo (quando tutti gli $r_{ij}, i < j$, sono nulli) a un valore massimo $\sum_{i < j} r_{ij}$, mentre il totale della variabilità tra le k variabili standardizzate può variare anch'esso da un valore minimo uguale a k a un valore massimo $k + \sum_{i \neq j} r_{ij} = R_{..}$. In più, poiché al variare di k e degli r_{ij} variano anche il totale $R_{..}$, il valore massimo dell'intensità dell'interdipendenza correlativa e i valori minimo e massimo del totale della variabilità, risulta difficile eseguire confronti tra le variabilità relative di due variabili multiple con valori di k o con valori degli r_{ij} diversi.

Questi confronti potrebbero ottenersi se fosse possibile costruire un indice che vari tra zero e uno al variare di $R_{..}$, in corrispondenza a ciascun valore di k e della struttura di interdipendenza correlativa. Per fare questo si rende necessario avere una descrizione più dettagliata della struttura di interdipendenza che consenta anche di valutare il contributo di ciascuna variabile semplice standardizzata alla variabilità multipla.

Per descrivere quest'ultimo aspetto del problema esamineremo separatamente il caso particolare, ma più semplice, di una matrice di correlazione con struttura di dipendenza correlativa uguale, cioè con tutti gli r_{ij} uguali tra loro per i quali è facile ricavare la corrispondente matrice di interdipendenza associativa, da quello più generale e più complesso di matrici di correlazione con struttura di interdipendenza correlativa diversa, cioè con tutti gli r_{ij} diversi tra loro, la cui struttura di interdipendenza associativa richiede la risoluzione di un problema di massimo.

3. Caso di variabili perfettamente correlate: matrici di correlazione con struttura di dipendenza correlativa costante

Quando le k variabili semplici sono tra loro perfettamente correlate, la matrice di correlazione presenta una struttura di interdipendenza

correlativa formata da r_{ij} tutti uguali tra loro e uguali a r . In questo caso la tabella diventa:

Tab. 2 - Matrice di correlazione con $r_{ij} = r, \forall i, j$.

$\{r_{ij}\}$				R_i	P_i
1	r	...	r	R_1	$1/k$
r	1	...	r	R_2	$1/k$
\vdots	\vdots	\ddots	\vdots	\vdots	\vdots
r	r	...	1	R_k	$1/k$
$R_{1.}$	$R_{2.}$...	$R_{k.}$	$R_{..}$	1

nella quale sono:

$$R_{1.} = R_{2.} = \dots = R_{k.} = 1 + (k-1)r \quad (10)$$

$$R_{..} = k + k(k-1)r \quad (11)$$

$$P_{1.} = P_{2.} = \dots = P_{k.} = \frac{1 + (k-1)r}{k + k(k-1)r} = \frac{1 + (k-1)r}{k[1 + (k-1)r]} = \frac{1}{k} \quad (12)$$

Teoricamente si possono pensare infinite matrici di correlazione di ordine $k \geq 2$ che, in corrispondenza allo stesso ordine k , danno luogo a infinite altre matrici di correlazione con la stessa struttura di interdipendenza correlativa, al variare di r con continuità da 0 a 1. Ad esempio, le tabelle corrispondenti a matrici di correlazione di ordine $k = 4$ e ai valori di $r = 1$ (la massima), $r = 0.825$ (intermedia) e $r = 0$ (la minima), con i rispettivi valori di R_i , P_i e $R_{..}$ sono date nell'ordine dalle Tabb. 3, 4 e 5.

Tab. 3 - Matrice di correlazione con $r_{ij} = 1, \forall i, j$.

$\{r_{ij}\}$				R_i	P_i
1	1	1	1	4	0.25
1	1	1	1	4	0.25
1	1	1	1	4	0.25
1	1	1	1	4	0.25
4	4	4	4	16	1.00

Tab. 4 - Matrice di correlazione con $r_{ij} = 0.825, \forall i, j$.

{ r_{ij} }				R_i	P_i
1	0.825	0.825	0.825	3.475	0.25
0.825	1	0.825	0.825	3.475	0.25
0.825	0.825	1	0.825	3.475	0.25
0.825	0.825	0.825	1	3.475	0.25
3.475	3.475	3.475	3.475	13.900	1.00

Tab. 5 - Matrice di correlazione con $r_{ij} = 0, \forall i, j$.

{ r_{ij} }				R_i	P_i
1	0	0	0	1	0.25
0	1	0	0	1	0.25
0	0	1	0	1	0.25
0	0	0	1	1	0.25
1	1	1	1	4	1.00

Per ciascuna di queste matrici di correlazione è, inoltre, facile determinare una matrice corrispondente che presenta una struttura di interdipendenza associativa tra le k variabili uguale a quella della matrice originaria. Gli elementi di questa nuova matrice, tutti uguali tra loro, determinati dalla semplice relazione

$$C = \frac{R_i R_j}{R_{..}} = \frac{1}{k} + \frac{k-1}{k} r \quad (13)$$

forniscono il contributo di ciascuna variabile semplice alla intensità dell'interdipendenza associativa e, ovviamente, in questo caso i valori di R_i , $R_{..}$ e P_i sono tutti uguali.

Ad esempio, le matrici di interdipendenza associativa corrispondenti alle matrici di correlazione delle Tab. 3, 4 e 5 sono rispettivamente:

Tab. 6 - Tabella di interdipendenza associativa relativa alla Tab. 3.

				R_i	P_i
1	1	1	1	4	0.25
1	1	1	1	4	0.25
1	1	1	1	4	0.25
1	1	1	1	4	0.25
4	4	4	4	16	1.00

Tab. 7 - Tabella di interdipendenza associativa relativa alla Tab. 4.

				R_i	P_i
0.86875	0.86875	0.86875	0.86875	3.47500	0.25
0.86875	0.86875	0.86875	0.86875	3.47500	0.25
0.86875	0.86875	0.86875	0.86875	3.47500	0.25
0.86875	0.86875	0.86875	0.86875	3.47500	0.25
3.47500	3.47500	3.47500	3.47500	13.90000	1.00

Tab. 8 - Tabella di interdipendenza associativa relativa alla Tab. 5.

$\{r_{ij}\}$				R_i	P_i
0.25	0.25	0.25	0.25	1	0.25
0.25	0.25	0.25	0.25	1	0.25
0.25	0.25	0.25	0.25	1	0.25
0.25	0.25	0.25	0.25	1	0.25
1.00	1.00	1.00	1.00	4	1.00

ottenute attraverso la relazione: $C = 0.25 + 0.75r$.

Altri elementi interessanti di queste matrici di correlazione sono:

- 1) I valori minimi e massimi del totale della variabilità multipla $R_{..}$ variano con l'ordine k della matrice; però, mentre i valori minimi ($r = 0$) risultano uguali a k , i valori massimi ($r = 1$) risultano uguali a k^2 , per cui risulta che al crescere dell'ordine k della matrice diminuisce l'importanza relativa del contributo degli elementi sulla diagonale principale e cresce sempre più quella del totale dell'interdipendenza correlativa, data da $k(k-1)r$.
- 2) I valori minimi e massimi delle k interdipendenze associative R_i di ciascuna variabile rispetto a tutte le altre variano anch'essi con l'ordine k da $R_{i.,min} = 1$ a $R_{i.,max} = k$, mentre rimangono sempre costanti e uguali a $\frac{1}{k}$ i valori di P_i . Ad esempio, in matri-

ci di correlazione di ordine $k = 5, 10, 50, 100$ i valori minimi e massimi di $R_{..}$, di R_i , del totale dell'interdipendenza correlativa e di P_i assumono i valori della seguente tabella:

Tab. 9 - Valori calcolati per matrici di correlazione di ordine $k = 5, 10, 50, 100$.

k	$R_{..,min}$	$R_{..,max}$	$R_{i.,min}$	$R_{i.,max}$	Tot. interd. corr.		P_i
					min	max	
5	5	25	1	5	0	20	$1/k$
10	10	100	1	10	0	90	$1/k$
50	50	2500	1	50	0	2450	$1/k$
100	100	10000	1	100	0	9900	$1/k$

- 3) Il valore dell'indice $IA = \frac{R_{..}}{(R_{..,max} - R_{..,min})}$ varia, come il valore di r , da 0 a 1, risulta uguale a C e indica il contributo fornito all'interdipendenza associativa anche dagli elementi sulla diagonale principale, indicazione che non era possibile ottenere dalle corrispondenti matrici di correlazione. Ad esempio, risulta che nelle tre matrici di correlazione delle Tab. 3, 4 e 5 il contributo all'interdipendenza associativa degli elementi, tutti unitari, della diagonale principale è funzione dell'intensità dell'interdipendenza associativa e risultano dati, rispettivamente, dai valori 1, 0.86875 e 0.25, come può evincersi anche dall'impiego dell'indice IA .

Alcune relazioni interessanti tra gli elementi descrittivi delle due tipologie di tabelle, quella fondata sulla struttura di interdipendenza correlativa e quella fondata sulla struttura di interdipendenza associativa, sono:

- (a) a ciascuna matrice di correlazione descritta dagli elementi $R_{i.}$, $R_{..}$ e $P_{i.}$ corrisponde una matrice associativa con identici elementi descrittivi $C_{i.}$, $C_{..}$ e $P_{i.}$. Questa è una conseguenza della particolare tipologia della struttura di interdipendenza correlativa costante e uguale a r , a cui corrisponde una struttura di interdipendenza associativa costante e uguale a $\frac{1}{k}$.
- (b) Le coppie di elementi descrittivi delle due tipologie di matrici $(R_{i.}, C_{i.})$, $(R_{..}, C_{..})$ risultano uguali e variano soltanto in funzione dell'intensità della variabilità totale multipla $R_{..}$, per cui la relazione che lega tutte le coppie di valori $(R_{i.}, C_{i.})$ sono disposti lungo una retta del tipo $R_{i.} = C_{i.}$, con coefficiente angolare pari a 1, cioè sono disposti lungo la prima bisettrice di un piano cartesiano con $x = R_{i.}$ e $y = C_{i.}$.

Vedremo nel paragrafo seguente come e perché queste relazioni tra i valori del totale della variabilità multipla $R_{..}$, delle k interdipendenze associative $R_{i.}$, del totale dell'interdipendenza correlativa e dei contributi delle singole variabili all'interdipendenza associativa variano al variare della struttura di interdipendenza correlativa quando gli r_{ij} sono tutti diversi tra loro e al variare di k .

4. Caso di variabili diversamente correlate: matrici di correlazione con struttura di interdipendenza diversa

Quando le k variabili semplici sono tra loro diversamente correlate, ogni matrice di correlazione presenta una struttura di interdipendenza correlativa descritta dalla relazione tra i valori r_{ij} , $i < j$. In questo caso la matrice di correlazione e le quantità informative sono quelle della Tab. 1.

Come abbiamo detto, a questa matrice generale corrispondono infinite matrici che possono variare tra loro sia per l'ordine k , ma anche, supponendo matrici di correlazione con lo stesso k , per la struttura di interdipendenza. Esaminiamo subito le relazioni tra le matrici di correlazione che presentano lo stesso ordine k e la stessa struttura di interdipendenza correlativa o associativa.

4.1 Matrici di correlazione con lo stesso ordine k e la stessa struttura di interdipendenza correlativa

A questa matrice di correlazione corrispondono infinite altre matrici di correlazione dello stesso ordine k che presentano la stessa struttura di interdipendenza correlativa, ma diverso totale della variabilità multipla R . Per chiarire meglio questa affermazione osserviamo che due matrici di correlazione dello stesso ordine k hanno uguale struttura di interdipendenza correlativa quando i $k(k-1)/2$ rapporti interni h_{ij} , ottenuti dividendo ciascuno dei valori r_{ij} , $i < j$, per un particolare valore $r_{ij} \neq 0$, sono uguali nelle due matrici, anche se i valori r_{ij} sono diversi da matrice a matrice.

La semi-matrice triangolare superiore dei $k(k-1)/2$ valori h_{ij} corrispondenti ai valori r_{ij} , $i < j$, quando si assume quale divisore r_{1k} , è la seguente:

$$\begin{aligned} h_{12} &= r_{12}/r_{1k}, & h_{13} &= r_{13}/r_{1k}, & h_{14} &= r_{14}/r_{1k}, & \dots & h_{1k} &= r_{1k}/r_{1k} = 1 \\ h_{23} &= r_{23}/r_{1k}, & h_{24} &= r_{24}/r_{1k}, & \dots & h_{2k} &= r_{2k}/r_{1k} \\ h_{34} &= r_{34}/r_{1k}, & \dots & h_{3k} &= r_{3k}/r_{1k} \\ & & \dots & & & & \\ & & & & & & h_{(k-1)k} &= r_{(k-1)k}/r_{1k} \end{aligned}$$

Questa semi-matrice superiore costituisce la struttura di interdipendenza correlativa costante delle infinite matrici di correlazione che

è possibile costruire dalla matrice di correlazione osservata al variare di $R_{..}$, cioè del totale della variabilità multipla, da $R_{..,max}$ a k . Per determinare la matrice di correlazione corrispondente a $R_{..,max}$ basta moltiplicare tutti i valori r_{ij} della matrice osservata, tranne quelli sulla diagonale principale, per il valore $p = \frac{1}{\max r_{ij}}$, mentre la matrice di corre-

lazione minima è costituita dagli elementi unitari sulla diagonale principale con tutti gli $r_{ij} = 0$, per $i \neq j$.

Così, assumendo che la matrice di correlazione osservata sia quella indicata nella tabella 1, la matrice di correlazione massima corrispondente e le sue quantità informative sono riportate nella tabella seguente:

Tab. 10 - Matrice di correlazione massima relativa alla Tab. 1.

{max r_{ij} }				$\max R_i$	P_i
1	pr_{12}	...	pr_{1k}	$1 + pR'_1$	$P_{1.}$
pr_{21}	1	...	pr_{2k}	$1 + pR'_2$	$P_{2.}$
⋮	⋮	⋮	⋮	⋮	⋮
pr_{k1}	pr_{k2}	...	1	$1 + pR'_k$	$P_{k.}$
$1 + pR'_{.1}$	$1 + pR'_{.2}$...	$1 + pR'_{.k}$	$k + 2p \sum_{i < j} r_{ij}$	1

nella quale è

$$R'_i = \sum_{\substack{i=1 \\ j \neq i}}^k r_{ij}, \quad R'_j = \sum_{\substack{i=1 \\ i \neq j}}^k r_{ij}, \quad R_{..,max} = k + 2p \sum_{i < j} r_{ij} \quad (14)$$

Prima di indicare il criterio per costruire matrici di correlazione intermedie con variabilità relativa multipla $R_{..}$ compresa tra $R_{..,max}$ e k , è importante osservare che non tutte le matrici intermedie sono utili per lo studio della variabilità relativa multipla e della struttura di interdipendenza correlativa e associativa delle k variabili standardizzate. Questo perché alcune matrici intermedie, compresa la massima, presentano l'inconveniente di non essere matrici semi-definite positive e quindi non possono costituire possibili matrici di correlazione nel senso impiegato in questo lavoro, dato che, come è noto, sia le matrici di covarianza delle variabili X_1, X_2, \dots, X_k , sia le matrici di correlazione go-

dono della proprietà di essere semi-definite positive (4). Si rende pertanto necessario trovare la matrice di correlazione massima che sia semi-definita positiva.

4.1.1 Determinazione della matrice di correlazione massima

Per poter determinare la matrice di correlazione massima, rispettando comunque il vincolo che la matrice risulti semi-definita positiva, ricordando che una matrice semi-definita positiva deve avere determinante maggiore o uguale a zero, si può ricercare un valore r_{max} risolvendo l'equazione:

$$\det(\mathbf{R}_{max}) = \begin{vmatrix} 1 & \frac{r_{12}}{r_{max}} & \dots & \frac{r_{1k}}{r_{max}} \\ \frac{r_{21}}{r_{max}} & 1 & \dots & \frac{r_{2k}}{r_{max}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{r_{1k}}{r_{max}} & \frac{r_{2k}}{r_{max}} & \dots & 1 \end{vmatrix} = 0 \quad (15)$$

con \mathbf{R}_{max} data da:

$$\mathbf{R}_{max} = \begin{vmatrix} 1 & \frac{r_{12}}{r_{max}} & \dots & \frac{r_{1k}}{r_{max}} \\ \frac{r_{21}}{r_{max}} & 1 & \dots & \frac{r_{2k}}{r_{max}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{r_{1k}}{r_{max}} & \frac{r_{2k}}{r_{max}} & \dots & 1 \end{vmatrix} \quad (16)$$

Supponendo, quindi, di avere una matrice di correlazione osservata di rango pieno (questa ipotesi non intacca la generalità del ragio-

(4) È importante osservare, tuttavia, che delle volte per lo studio di problemi particolari diversi da quello trattato in questo lavoro vengono costruite matrici di correlazione particolari che possono anche non essere semi-definite positive, cioè possono anche presentare determinanti negativi, pur mantenendo l'importante proprietà che i vettori U della matrice di correlazione sono funzione dei rapporti che descrivono la struttura di interdipendenza correlativa. Esempi di matrici che non sono semi-definite positive sono le matrici iniziali di covarianze e correlazioni genetiche (Rouvier, 1966, 1969; Hashiguchi e Morishima, 1969). Altre matrici di correlazione particolari sono quelle formate da correlazioni tetracoriche o da correlazioni parziali (Jackson, 1991).

namento, dato che una matrice di correlazione non è di rango pieno quando almeno una variabile dipende in modo perfettamente lineare da altre variabili, caso statisticamente poco interessante e comunque facilmente rilevabile) si risolve la precedente equazione ottenendo k soluzioni r_{max} , se k è il rango della matrice. Tra le k soluzioni ottenute $r_{1,max}$, $r_{2,max}$, \dots , $r_{k,max}$ si sceglie la più grande, si sostituisce nella (16) al posto di r_{max} e si calcolano gli elementi r_{ij} della matrice di correlazione massima R_{max} , che ovviamente presenta la stessa struttura di interdipendenza correlativa della matrice di correlazione osservata e gode della proprietà di essere semi-definita positiva. In questo modo conosciamo anche l'intervallo di variazione $R_{max} - k$ entro cui sono comprese le infinite matrici di correlazione intermedie semi-definite positive con la stessa struttura di interdipendenza correlativa della matrice osservata.

4.1.2 Criterio per la determinazione degli elementi di una matrice intermedia

A questo punto si rende necessario un criterio che, dalla conoscenza della struttura di interdipendenza correlativa descritta dai $k(k-1)/2$ rapporti h_{ij} e di un valore dato della variabilità totale R_{\cdot} , permette di ricostruire una corrispondente matrice di correlazione intermedia. Questo criterio consiste nel determinare dalla relazione

$R_{\cdot} = k + 2 \sum_{i<j}^k r_{ij}$ la quantità $\sum_{i<j}^k r_{ij} = R = \frac{R_{\cdot} - k}{2}$ e nel dividere questa quantità tra i nuovi $k(k-1)/2$ valori r_{ij} in proporzione dei $k(k-1)/2$ valori h_{ij} mediante la relazione $r_{ij} = \frac{R}{H} h_{ij}$, nella quale $H = \sum_{i<j}^k h_{ij}$.

Siamo ora in condizione di potere definire un indice di variabilità che varia tra zero e uno, che consente di confrontare la variabilità di due matrici di correlazione osservate che possono presentare o lo stesso ordine k , ma struttura di interdipendenza correlativa diversa, o ordini k diversi tra loro. Questo indice è dato dalla semplice relazione:

$$IV_{DC} = \frac{R_{\dots,oss} - k}{R_{\dots,max} - k} \quad (17)$$

nella quale IV_{DC} indica l'indice di variabilità della matrice osservata, $R_{\dots,oss}$ e $R_{\dots,max}$ i valori del totale della variabilità multipla corrispondente alla matrice osservata e a quella massima, rispettivamente.

variabili. Pertanto, i coefficienti $C_{11}, C_{21}, \dots, C_{k1}$ corrispondenti alla prima componente principale, in corrispondenza dell'autovalore β_1 , costituiscono i termini che esprimono la relazione tra le k variabili che consentono di costruire la matrice di interdipendenza associativa corrispondente alla matrice di correlazione osservata.

Tab. 11 - Matrice di interdipendenza associativa.

C_{ij}				C_i	A_i
C_{11}	C_{12}	\dots	C_{1k}	$C_{1.}$	$A_{1.}$
C_{21}	C_{22}	\dots	C_{2k}	$C_{2.}$	$A_{2.}$
\vdots	\vdots	\ddots	\vdots	\vdots	\vdots
C_{k1}	C_{k2}	\dots	C_{kk}	$C_{k.}$	$A_{k.}$
$C_{.1}$	$C_{.2}$	\dots	$C_{.k}$	$C_{..}$	1

In questa matrice associativa, i valori C_{ij} descrivono la struttura di interdipendenza associativa tra ciascuna variabile z_{X_i} e tutte le altre; i valori C_i , diversi tra loro, sono corrispondenti ai valori R_i della matrice di correlazione e dipendono dalla struttura di interdipendenza correlativa della matrice di correlazione; il vettore $A_i = C_i/C_{.}$, descrive il contributo unitario di ciascuna variabile z_{X_i} alla struttura di interdipendenza associativa; infine, il valore $C_{..}$ corrisponde al contributo massimo del totale della variabilità multipla $R_{..}$. A questo punto è importante osservare che:

- 1) $C_{..}$ non risulta più uguale a $R_{..}$, come accadeva per le matrici del paragrafo 3, per effetto della struttura di interdipendenza correlativa descritta dai rapporti h_{ij} . In più, è facile mostrare come a ciascuna delle infinite matrici di correlazione a partire dalla tabella 10 con valori di $R_{..}$ compresi nell'intervallo $(k, R_{..,max})$ corrisponde una matrice di associazione, rappresentata dalla Tab. 11, con valori $C_{..}$ compresi nell'intervallo $(C_{..,min}, C_{..,max})$ e che esiste una relazione lineare tra tutti i possibili valori di $R_{..}$ e i corrispondenti valori $C_{..}$;
- 2) il vettore A_i risulta sempre lo stesso per tutte le infinite matrici di interdipendenza associativa intermedie che è possibile ottenere in corrispondenza ai valori di $C_{..}$ compresi nell'intervallo $(C_{..,min}, C_{..,max})$. È questo forse il risultato statisticamente più importante di tutta l'analisi, perché mostra

come il vettore A_i dipenda soltanto dalla struttura di interdipendenza degli r_{ij} , $i < j$, descritta dai valori h_{ij} ; cioè, detto in altre parole, esiste una corrispondenza biunivoca tra la struttura di interdipendenza correlativa e la struttura di interdipendenza associativa.

Purtroppo, non risulta agevole determinare analiticamente i parametri di questa relazione lineare che dipende dai rapporti h_{ij} , perché in corrispondenza di $C_{\dots, min}$, cioè del limite inferiore di C_{\dots} che dovrebbe corrispondere al limite inferiore k delle matrici di correlazione con la stessa struttura di interdipendenza correlativa, l'azzeramento dei valori r_{ij} non consente di ottenere la corrispondente matrice associativa appartenente alla Tab. 11, cioè il valore di $C_{\dots, min}$ con valori A_i uguali. Daremo per ora a questo problema una soluzione numerica che mostriamo con dettaglio descrivendo un'applicazione basata su un problema ormai diventato classico, riservandoci di tornare sull'argomento in un prossimo lavoro.

Una volta determinata la relazione lineare $R_{\dots} = a + bC_{\dots}$ e il valore $C_{\dots, min}$, si può definire il seguente indice relativo di variabilità:

$$IA_{DA} = \frac{C_{\dots, oss} - C_{\dots, min}}{C_{\dots, max} - C_{\dots, min}} \quad (21)$$

che varia tra zero e uno e fornisce, come deve essere, indicazioni analoghe all'indice IV_{DC} corrispondente, in quanto le matrici di correlazione e quella di associazione si riferiscono allo stesso aspetto di variabilità (si potrebbe dire che sono due facce della stessa medaglia).

5. Applicazione del metodo

Descriviamo meglio quanto detto con un'applicazione a un problema concreto che fornirà anche tutti i passi necessari per ottenere la descrizione completa di una qualsiasi delle infinite matrici di correlazione osservate, degli indici che consentono tale descrizione e dei confronti tra matrici di correlazione diverse.

La matrice di correlazione presa in considerazione è forse la più nota in letteratura: si tratta, infatti, della matrice considerata nell'articolo originale di Hotelling (1933) sulle componenti principali ed è fon-

data su dati di Kelley (1928). Come è noto, si tratta di quattro test psicologici estratti da una batteria di test somministrati a 140 bambini del settimo grado scolastico. I quattro test scelti sono: x_1 = velocità di lettura, x_2 = potenza di lettura, x_3 = velocità aritmetica e x_4 = potenza aritmetica.

La matrice di correlazione è la seguente:

	x_1	x_2	x_3	x_4
x_1	1.000	0.698	0.264	0.081
x_2	0.698	1.000	-0.061	0.092
x_3	0.264	-0.061	1.000	0.594
x_4	0.081	0.092	0.594	1.000

In base al metodo descritto nelle pagine precedenti, andiamo innanzitutto al calcolo dei totali di riga e di colonna per la matrice, ottenendo quindi la seguente tabella osservata (l'ultima colonna riporta i valori dei marginali di riga, presenti nella penultima colonna, rapportati al loro totale):

1.000	0.698	0.264	0.081	2.043	0.2785
0.698	1.000	-0.061	0.092	1.729	0.2357
0.264	-0.061	1.000	0.594	1.797	0.2450
0.081	0.092	0.594	1.000	1.767	0.2408
2.043	1.729	1.797	1.767	7.336	1.0000

Calcoliamo, quindi, i rapporti h_{ij} tra i singoli coefficienti di correlazione della matrice e l'elemento r_{14} :

$$h_{12} = \frac{r_{12}}{r_{14}} = 8.61728; \quad h_{13} = \frac{r_{13}}{r_{14}} = 3.25926; \quad h_{14} = \frac{r_{14}}{r_{14}} = 1.00000;$$

$$h_{23} = \frac{r_{23}}{r_{14}} = -0.75309; \quad h_{24} = \frac{r_{24}}{r_{14}} = 1.13580;$$

$$h_{34} = \frac{r_{34}}{r_{14}} = 7.33333;$$

questi rapporti descrivono la struttura di interdipendenza correlativa. La somma totale di questi rapporti è pari a 20.59258.

A questo punto si può procedere al calcolo della matrice di correlazione massima, che presenta analoga struttura di interdipendenza cor-

relativa della matrice originaria, rispettando comunque il vincolo che sia semi-definita positiva; per fare questo risolviamo l'equazione:

$$\det(\mathbf{R}_{max}) = \begin{vmatrix} 1.000 & \frac{0.698}{r_{max}} & \frac{0.264}{r_{max}} & \frac{0.081}{r_{max}} \\ \frac{0.698}{r_{max}} & 1.000 & \frac{-0.061}{r_{max}} & \frac{0.092}{r_{max}} \\ \frac{0.264}{r_{max}} & \frac{-0.061}{r_{max}} & 1.000 & \frac{0.594}{r_{max}} \\ \frac{0.081}{r_{max}} & \frac{0.092}{r_{max}} & \frac{0.594}{r_{max}} & 1.000 \end{vmatrix} = 0 \quad (22)$$

La precedente equazione ammette le seguenti radici:

$$r_{1,max} = -0.8474 \quad (23)$$

$$r_{2,max} = -0.4642 \quad (24)$$

$$r_{3,max} = 0.4779 \quad (25)$$

$$r_{4,max} = 0.8337 \quad (26)$$

Tra le quattro radici ottenute, quella "utile" è $r_{4,max} = 0.8337$, dato che $r_{1,max}$ e $r_{2,max}$ sono da scartare in quanto soluzioni negative che cambierebbero la struttura di correlazione della matrice originaria, mentre $r_{3,max}$ è da scartare perché se utilizzata si otterrebbe una matrice con alcuni elementi superiori ad 1 e quindi non avremmo più una matrice di correlazione. La matrice massima è quindi data da:

$$\mathbf{R}_{max} = \begin{bmatrix} 1.0000 & 0.8372 & 0.3167 & 0.0972 \\ 0.8372 & 1.0000 & -0.0732 & 0.1104 \\ 0.3167 & -0.0732 & 1.0000 & 0.7125 \\ 0.0972 & 0.1104 & 0.7125 & 1.0000 \end{bmatrix}$$

Questa è la matrice di correlazione massima che rispetta la struttura di correlazione della matrice originaria e che gode comunque della proprietà di essere semi-definita positiva. "Orlando" anche questa matrice si ha:

1.0000	0.8372	0.3167	0.0972	2.2511	0.2813
0.8372	1.0000	-0.0732	0.1104	1.8744	0.2343
0.3167	-0.0732	1.0000	0.7125	1.9560	0.2444
0.0972	0.1104	0.7125	1.0000	1.9201	0.2400
2.2511	1.8744	1.9560	1.9201	8.0016	1.0000

Come può facilmente verificarsi, questa matrice ha gli stessi rapporti tra i coefficienti di correlazione rispetto a quelli calcolati nella matrice osservata, ma totali marginali unitari diversi da quelli della matrice osservata. Con queste caratteristiche esistono, ovviamente, infinite matrici di correlazione con la stessa struttura di interdipendenza correlativa, ma con totale generale diverso.

Sulla base dei rapporti costanti tra i coefficienti di correlazione, determiniamo le 10 matrici di correlazione che presentano un totale generale che oscilla dal massimo (8.0016) al minimo $k=4$, diminuendo gli elementi della matrice di correlazione di una quantità “proporzionale” a $0.1R_{\dots, max}$, per arrivare quindi alla tabella seguente:

Totale gen.	(Totale gen.-4)/2	r_{12}	r_{13}	r_{14}	r_{23}	r_{24}	r_{34}
8.0016	2.0008	0.8373	0.3167	0.0972	-0.0732	0.1104	0.7125
7.6014	1.8007	0.7535	0.2850	0.0874	-0.0659	0.0993	0.6413
7.2013	1.6007	0.6698	0.2533	0.0777	-0.0585	0.0883	0.5700
6.8011	1.4006	0.5861	0.2217	0.0680	-0.0512	0.0773	0.4988
6.4010	1.2005	0.5024	0.1900	0.0583	-0.0439	0.0662	0.4275
6.0008	1.0004	0.4186	0.1583	0.0486	-0.0366	0.0552	0.3563
5.6006	0.8003	0.3349	0.1267	0.0389	-0.0293	0.0441	0.2850
5.2005	0.6003	0.2512	0.0950	0.0291	-0.0220	0.0331	0.2138
4.8003	0.4002	0.1675	0.0633	0.0194	-0.0146	0.0221	0.1425
4.4002	0.2001	0.0837	0.0317	0.0097	-0.0073	0.0110	0.0713
4.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Questa tabella consente di costruire le 11 matrici a partire da quella massima di totale pari a 8.0016, fino a quella minima di totale 4.0000; il procedimento seguito è il seguente: prendendo il totale della matrice massima si determinano i primi due valori della prima riga della tavola precedente e cioè 8.0016, 2.0008, dove quest'ultima quantità corrisponde alla sommatoria delle r_{ij} della parte triangolare superiore della matrice; si considera quindi un decremento costante nel passaggio da una riga alla successiva pari a $\frac{2.0008}{10}$ e a partire da questa quantità si costruiscono i totali per le rimanenti matrici: così, ad esempio, per

calcolare il totale della seconda matrice basta calcolare la quantità

$$2.0008 - \frac{2.0008}{10} = 1.8007.$$

Da questo valore si ottiene il totale della seconda riga, dato da $1.8007 \cdot 2 + 4 = 7.6014$. Procedendo in questo modo si ottengono tutti gli altri valori delle prime due colonne della tavola precedente, con significati ovvi.

Dagli elementi della seconda colonna si ottengono poi tutti i valori degli r_{ij} corrispondenti alla semi-matrice triangolare superiore di ciascuna matrice successiva. Ad esempio, per passare dalla matrice di correlazione massima alla successiva si divide il valore della seconda colonna 1.8007 per la sommatoria dei rapporti h_{ij} , cioè 20.59258, ottenendo il valore 0.087444118; si moltiplicano, quindi, tutti i valori dei singoli h_{ij} per questo valore ottenendo i corrispondenti valori r_{ij} . In definitiva, per la seconda matrice si ha:

$$r_{12} = 8.61728 * 0.087444118 = 0.7535;$$

$$r_{13} = 3.25926 * 0.087444118 = 0.2850;$$

$$r_{14} = 1.00000 * 0.087444118 = 0.0874;$$

$$r_{23} = -0.75309 * 0.087444118 = -0.0659;$$

$$r_{24} = 1.13580 * 0.087444118 = 0.0993;$$

$$r_{34} = 7.33333 * 0.087444118 = 0.6413.$$

Procedendo allo stesso modo si ottengono le altre matrici di correlazione che hanno analoga struttura di interdipendenza correlativa, ma diversa variabilità totale. A questo punto si passa alla risoluzione del sistema (20) per la matrice massima e per quella che “dista” dalla massima di 1/10 della variabilità totale, cioè si passa al calcolo della prima componente principale per queste due matrici e alle corrispondenti matrici di interdipendenza associativa.

La matrice di interdipendenza associativa che corrisponde alla prima componente principale della matrice massima è la seguente:

0.7224	0.6119	0.5425	0.5158	2.3926	0.3019
0.6119	0.5184	0.4596	0.4369	2.0268	0.2558
0.5425	0.4596	0.4074	0.3874	1.7969	0.2267
0.5158	0.4369	0.3874	0.3683	1.7084	0.2156
2.3926	2.0268	1.7969	1.7084	7.9247	1.0000

La matrice di interdipendenza associativa che corrisponde alla prima componente principale della seconda matrice presa in considerazione è la seguente:

0.6860	0.5811	0.5152	0.4898	2.2721	0.3019
0.5811	0.4923	0.4364	0.4149	1.9247	0.2558
0.5152	0.4364	0.3869	0.3678	1.7063	0.2267
0.4898	0.4149	0.3678	0.3497	1.6222	0.2156
2.4830	2.1034	1.8647	1.7727	7.5253	1.0000

Come può vedersi, queste due matrici (ma le stesse considerazioni si potrebbero fare per le altre matrici di interdipendenza associativa che si potrebbero calcolare) differiscono dalle due matrici di correlazione corrispondenti per i seguenti aspetti:

- 1) le dipendenze associative di ciascuna variabile con se stessa e con tutte le altre sono nella stessa relazione proporzionale tra di loro e uguali a quelle dei totali marginali;
- 2) i contributi unitari di ciascuna delle quattro variabili all'interdipendenza associativa sono uguali nelle due matrici, come si può notare dall'ultima colonna delle due matrici;
- 3) il massimo di variabilità multipla spiegata dalla matrice di interdipendenza associativa, 7.9247, differisce dal massimo della variabilità multipla spiegata dalla matrice di interdipendenza correlativa massima, 8.0016, di -0.0769 ;
- 4) la differenza tra il totale (massimo) della variabilità relativa multipla spiegata dalla seconda matrice di interdipendenza associativa, 7.5253, e il totale della variabilità relativa multipla spiegata dalla seconda matrice di interdipendenza correlativa, 7.6014, è di -0.0762 , cioè uno 0.0007 in meno della differenza precedente (proporzionale a 0.01);
- 5) questi due valori forniscono i due parametri di un'equazione lineare che descrive la relazione di corrispondenza tra le varianze multiple spiegate dalle due forme di interdipendenza, quella correlativa e quella associativa. Il valore -0.0769 è l'ordinata all'origine e -0.0007 è il coefficiente angolare della retta.

Da questi risultati si ottengono i parametri della retta $TVC = -0.0762 - 0.0007 \cdot TVA$ che lega i totali TVC della variabilità multipla $R_{..}$

delle matrici di correlazione a quelli *TVA* delle matrici associative. Da queste relazioni si ottengono i dati della seguente tabella:

<i>T.V.C.</i>		<i>T.V.A.</i>
8.0016	0.0769	7.9247
7.6014	0.0762	7.5253
7.2013	0.0755	7.1258
6.8011	0.0748	6.7263
6.4009	0.0741	6.3268
6.0008	0.0734	5.9274
5.6006	0.0728	5.5279
5.2005	0.0721	5.1284
4.8003	0.0714	4.7289
4.4002	0.0707	4.3295
4.0000	0.0700	3.9300

che indicano come dai totali delle varianze $R_{..}$ delle matrici di correlazione è possibile ottenere i totali $C_{..}$ delle corrispondenti matrici di associazione. Noti i valori $C_{..}$, variabili da $C_{..,min} = 3.9300$ a $C_{..,max} = 7.9247$, è possibile ricostruire le matrici di associazione corrispondenti: basta infatti moltiplicare ciascun totale della relativa matrice di variabilità associativa per i coefficienti unitari costanti riportati a fianco delle prime due tavole, cioè 0.3019, 0.2558, 0.2267 e 0.2156, mediante le quali si ottengono i totali marginali delle matrici successive da cui costruire poi gli elementi interni, attraverso il prodotto dei relativi marginali diviso il totale della matrice.

A questo punto possiamo andare al calcolo del primo indice di variabilità multipla che risulta:

$$IV_{DC} = \frac{R_{..,oss} - k}{R_{..,max} - k} = \frac{7.336 - 4}{8.0016 - 4} = 0.8337 \quad (27)$$

e del secondo indice che risulta:

$$IV_{DA} = \frac{C_{..,oss} - C_{..,min}}{C_{..,max} - C_{..,min}} = \frac{7.2602 - 3.9300}{7.9247 - 3.9300} = 0.8337 \quad (28)$$

dove il valore $C_{\dots, oss}$ è ricavato dal totale degli elementi della matrice di interdipendenza associativa corrispondente alla matrice osservata, data da:

0.6618	0.5606	0.4971	0.4725	2.1920	0.3019
0.5606	0.4749	0.4211	0.4003	1.8569	0.2558
0.4971	0.4211	0.3733	0.3548	1.6463	0.2267
0.4725	0.4003	0.3548	0.3374	1.5650	0.2156
2.1920	1.8569	1.6463	1.5650	7.2602	1.0000

Come era da attendersi e come notato alla fine del paragrafo precedente, i due indici hanno lo stesso valore.

Inoltre, dal vettore dei contributi unitari A_i si ottengono i contributi unitari di ciascuna variabile all'interdipendenza associativa. Si ha così che i contributi percentuali delle quattro variabili osservate sono:

la prima variabile - velocità di lettura - contribuisce con il 30.19%
 la seconda variabile - potenza di lettura - contribuisce con il 25.58%
 la terza variabile - velocità aritmetica - contribuisce con il 22.67%
 la quarta variabile - potenza aritmetica - contribuisce con il 21.56%

100.00%

Come può vedersi, i contributi più importanti risultano, nell'ordine, quelli forniti dalla velocità e dalla potenza di lettura che assommano assieme al 55.77%, mentre la velocità e la potenza aritmetica danno un contributo del 44.23%. D'altro canto, i contributi della velocità di lettura e della velocità aritmetica risultano complessivamente pari al 52.86% contro il 47.14% della potenza di lettura e potenza aritmetica. Ovviamente, si tratta di risultati che potrebbero essere molto informativi per specialisti in psicologia.

6. Conclusioni

In questo lavoro si è affrontato il problema della variabilità multipla, con l'obiettivo di mettere a punto un opportuno indice che misuri questo particolare aspetto in k variabili standardizzate. Per potere fare questo, si è fatto riferimento a una matrice di correlazione tra k variabili semplici X_i che costituiscono una variabile multipla omogenea e si sono introdotti i concetti di interdipendenza correlativa e di interdi-

pendenza associativa. Per interdipendenza correlativa s'intende la struttura d'interdipendenza tra tutte le possibili correlazioni tra coppie di variabili (X_i, X_j) , mentre per interdipendenza associativa s'intende la struttura d'interdipendenza tra ciascuna variabile e tutte le altre. Fatto questo, si è messo a punto un metodo per arrivare alla definizione di due opportuni indici per misurare, ciascuno, questi due aspetti, pervenendo a quelli che nel lavoro sono stati indicati come indici *TVC*, per misurare l'interdipendenza correlativa, e *TVA*, per misurare l'interdipendenza associativa, separando il problema in due parti: il caso in cui abbiamo a che fare con matrici di correlazione con struttura di dipendenza correlativa costante e il caso di matrici di correlazione con struttura d'interdipendenza diversa.

I principali risultati ottenuti in questo lavoro possono essere sintetizzati nei seguenti punti:

- l'aver individuato due tipi d'interdipendenza, quella associativa e quella correlativa;
- l'aver messo a punto due indici coerenti per la misura di questi due aspetti d'interdipendenza;
- l'aver trovato un legame funzionale, attraverso una semplice retta, che lega i due indici *TVC* e *TVA* e che quindi lega i due tipi d'interdipendenza, come deve essere;
- il potere determinare i contributi all'interdipendenza associativa delle k variabili semplici X_i che compongono la variabile multipla sotto studio.

Nel metodo utilizzato per arrivare alla determinazione dei due indici *TVC* e *TVA*, fondamentale è il calcolo della prima componente principale a partire dalla matrice di correlazione osservata: un aspetto in questo senso importante è il fatto che, nell'impostazione seguita in questo lavoro, la prima componente principale è l'unica componente, tra le k che si possono determinare in una matrice di correlazione di rango k , a dare informazioni significative sulla struttura d'interdipendenza, sia associativa che correlativa, tra le k variabili.

Ancora rimangono da investigare con maggiore attenzione alcuni aspetti particolari, quale ad esempio la determinazione analitica della matrice che abbiamo chiamato $C_{\dots, \min}$ e che rappresenta il limite inferiore delle matrici di correlazione con la stessa struttura d'interdipendenza correlativa, cosa che ci ripromettiamo di fare in un prossimo lavoro.

BIBLIOGRAFIA

HASHIGUCHI S., MORISHIMA H. (1969), *Estimation of genetic contribution of principal components to individual variates concerned*, *Biometrics*, 25, 9-15.

HOTELLING H. (1933), *Analysis of a complex of statistical variables into principal components*, *Journal of Educational Psychology*, 24, 417-441, 498-520.

JACKSON J. (1991), *A User's Guide to Principal Components*, J. Wiley, New York.

KELLEY T. (1928), *Crossroads in the Mind of Man*, Stanford University Press, Palo Alto.

ROUVIER R. (1966), *L'analyse en composantes principales: son utilisation en genetique et ses rapports avec l'analyse discriminatoire*, *Biometrics*, 22, 343-357.

ROUVIER R. (1969), *Ponderation des valeurs genotypiques dans la selection par index sur plusieurs caracteres*, *Biometrics*, 25, 295-307.

ANGELO M. MINEO (**)

DI UN NUOVO COEFFICIENTE PER LA CORRETTA MISURA DELL'ASSOCIAZIONE IN TABELLE DICOTOMICHE E IN TABELLE DI CONTINGENZA (*)

1. Introduzione

Un importante capitolo della metodologia statistica riguarda, come è noto, lo studio delle relazioni di dipendenza e di interdipendenza tra "categorie" (classi) di due o più caratteri statistici (1).

L'aspetto di interdipendenza più studiata è storicamente l'associazione tra le categorie di due variabili che, essendo le categorie relative a modalità simboliche e non valori, viene studiata mediante le informazioni fornite dalle frequenze di una cosiddetta tabella di contingenza. Supponiamo di avere la seguente tabella di contingenza quadrata:

Tab. 1 - Tabella di contingenza quadrata.

	A_1	A_2	...	A_r	TOTALE
B_1	n_{11}	n_{12}	...	n_{1r}	$n_{1.}$
B_2	n_{21}	n_{22}	...	n_{2r}	$n_{2.}$
⋮	⋮	⋮	⋮	⋮	⋮
B_r	n_{r1}	n_{r2}	...	n_{rr}	$n_{r.}$
TOTALE	$n_{.1}$	$n_{.2}$...	$n_{.r}$	$n_{..}$

(*) Questa ricerca è stata parzialmente finanziata da fondi MURST.

(**) Dipartimento di Scienze Statistiche e Matematiche "Silvio Vianelli", Università di Palermo, Viale delle Scienze - 90128 Palermo.

(1) Queste "variabili" categoriali possono derivare: a) da valori di variabili quantitative misurate su scale di misura per intervalli o per rapporti; b) da "gradi" di variabili qualitative espresse su scale di misura ordinali; c) da modalità di variabili qualitative espresse su scale di misura non ordinali.

Le frequenze della tabella 1 descrivono l'associazione tra le modalità delle variabili categoriali $A_i, B_i, i = 1, 2, \dots, r$, le quali assumono:

a) la forma di completa associazione quando la maggior parte delle frequenze n_{ij} si concentra sulla diagonale principale, compatibilmente con i totali marginali osservati;

b) la forma di completa dissociazione quando la maggior parte delle frequenze n_{ij} si concentra sulla diagonale secondaria, sempre rispettando i totali marginali osservati;

c) la forma di indipendenza quando tutte le frequenze n_{ij} coincidono con le frequenze "teoriche" corrispondenti alla quantità (2) $\frac{n_{i.}n_{.j}}{n_{..}}$.

Per misurare l'associazione in una tabella di frequenze osservata è necessario, pertanto, disporre di un coefficiente di associazione che in corrispondenza delle forme a), b) e c) assuma i valori +1, -1 e 0, rispettivamente.

Tradizionalmente la ricerca di questo coefficiente è stata affrontata esaminando prima il caso particolare, più semplice, delle tavole dicotomiche (2×2) e poi il caso più generale delle tavole ($r \times r$), con $r > 2$.

Per misurare l'associazione nelle tavole dicotomiche sono stati proposti in letteratura diversi coefficienti di associazione, tra cui i più importanti, non fosse altro che per l'autorevolezza degli statistici che li hanno proposti, i due coefficienti Q e Y di Yule e il coefficiente V di Karl Pearson che, nonostante la grande polemica che allora esplose tra i due autori in difesa della validità dei propri coefficienti, scrivendo per questo centinaia di pagine di cui la rivista "Biometrika" dei primi decenni del XX secolo è principale testimone, ancora oggi presentano seri inconvenienti.

Per misurare l'associazione nelle tabelle di contingenza ($r \times r$), con $r > 2$, Karl Pearson ha proposto l'impiego dell'indice X^2 , che non è una misura di associazione, sia perché il suo limite superiore cresce al crescere di $n_{..}$ tendendo ad infinito, sia perché non rispetta le condizioni a) e b) viste in precedenza.

Nemmeno il coefficiente di associazione P , proposto sempre da Karl Pearson nel tentativo di limitare all'intervallo (0,1) i suoi valori possibili, è un buon coefficiente perché, anche nel caso di tabelle qua-

(2) Le forme a) e b) si possono logicamente ottenere solo da tabelle di frequenze quadrate, in quanto tabelle di frequenze rettangolari ($r \times c$) corrispondenti a due variabili categoriali con numero di categorie r e c diversi presentano seri problemi di significato delle associazioni. La forma c) corrisponde all'ipotesi di indipendenza stocastica tra le due distribuzioni marginali fissate.

drate e nel caso di completa associazione nella tabella, assume un valore che dipende dal numero delle categorie r e quindi anche questo coefficiente non assume il valore limite $+1$.

Migliori, perché consentono di ottenere il valore $+1$ in corrispondenza di completa associazione, sono l'indice T di Tchuprov, che coincide con l'indice C di Cramer, anche se entrambi non rispettano la condizione b).

Nel presente lavoro proporremo un nuovo coefficiente, sia per il caso più semplice delle tavole dicotomiche, sia per il caso più generale delle tabelle di contingenza quadrate che, come si vedrà, rispetta in pieno le condizioni a), b) e c) in presenza di tutti i possibili tipi di tabelle quadrate corrispondenti a distribuzioni marginali qualsiasi. In più, mostreremo che questo nuovo coefficiente è una vera misura che varia in tutti i punti della scala da -1 a 0 (dissociazione) e da 0 a $+1$ (associazione) della stessa quantità per lo spostamento di una unità di frequenza all'interno delle celle della tabella.

Il lavoro è organizzato nel modo seguente: nel paragrafo 2 discuteremo i principali inconvenienti dei coefficienti di associazione Q , Y e C proposti per le tabelle dicotomiche, dell'indice X^2 e dei coefficienti di associazione P , T e V per le tabelle di contingenza.

Nel paragrafo 3 proporremo un nuovo coefficiente di associazione M e ne mostreremo l'impiego sia al caso particolare delle tabelle dicotomiche, sia al caso generale di tabelle di contingenza ($r \times r$). Alcune applicazioni numeriche a problemi particolari mostreranno le notevoli proprietà di correttezza delle misure fornite dal coefficiente M e il rispetto completo delle tre condizioni poste in questo lavoro. Nel paragrafo 4 presenteremo, infine, alcune considerazioni conclusive.

2. Alcune considerazioni sulle principali misure di associazione proposte per tabelle dicotomiche e per tabelle di contingenza

Abbandonando la tradizionale notazione che indica una tabella dicotomica nella forma:

CATEGORIE	A	$NON A$	TOTALE
B	a	b	$a + b$
$NON B$	c	d	$c + d$
TOTALE	$a + c$	$b + d$	n

preferiamo, per uniformità di trattamento, indicarla nella forma della tabella di frequenze 1, ossia

Tab. 2 - Tabella di contingenza (2 x 2).

CATEGORIE	<i>A</i>	<i>NON A</i>	TOTALE
<i>B</i>	n_{11}	n_{12}	$n_{11} + n_{12} = n_{1.}$
<i>NON B</i>	n_{21}	n_{22}	$n_{21} + n_{22} = n_{2.}$
TOTALE	$n_{11} + n_{21} = n_{.1}$	$n_{12} + n_{22} = n_{.2}$	$n_{11} + n_{12} + n_{21} + n_{22} = n_{..}$

La particolarità di questa tabella di frequenze è che, fissati i totali marginali, basta la conoscenza di una sola delle frequenze n_{ij} per determinare quelle di tutte le altre celle, per cui un ragionamento sull'associazione può essere fondato su una sola di tali frequenze. Così, se si confronta la frequenza n_{11} di una tavola di frequenze osservata con la corrispondente frequenza teorica $\frac{(n_{11} + n_{12})(n_{11} + n_{21})}{n_{..}}$ si può affermare che se

$$n_{11} = \frac{(n_{11} + n_{12})(n_{11} + n_{21})}{n_{..}} \quad (1)$$

nella tabella c'è indipendenza, se

$$n_{11} > \frac{(n_{11} + n_{12})(n_{11} + n_{21})}{n_{..}} \quad (2)$$

nella tabella c'è associazione positiva, se

$$n_{11} < \frac{(n_{11} + n_{12})(n_{11} + n_{21})}{n_{..}} \quad (3)$$

nella tabella c'è dissociazione o associazione negativa.

Poiché risulta che nel caso (1) di indipendenza la quantità

$$D_{11} = n_{11} - \frac{(n_{11} + n_{12})(n_{11} + n_{21})}{n_{..}} \quad (4)$$

assume anche la forma

$$D_{11} = \frac{n_{11}n_{22} - n_{12}n_{21}}{n_{..}} \quad (5)$$

per costruire un coefficiente di associazione sono state poste le due condizioni:

- 1) che il coefficiente sia funzione di D_{11} ;
- 2) che sia tale da risultare crescente (positivamente nel caso di associazione, negativamente nel caso di dissociazione) al crescere di D_{11} .

Come può facilmente verificarsi, sia i due coefficienti di Yule:

$$Q = \frac{n_{11}n_{22} - n_{12}n_{21}}{n_{11}n_{22} + n_{12}n_{21}} \quad (6)$$

e

$$Y = \frac{(n_{11}n_{22})^{1/2} - (n_{12}n_{21})^{1/2}}{(n_{11}n_{22})^{1/2} + (n_{12}n_{21})^{1/2}} \quad (7)$$

sia quello di Karl Pearson

$$V = \frac{n_{11}n_{22} - n_{12}n_{21}}{[(n_{11} + n_{12})(n_{11} + n_{21})(n_{12} + n_{22})(n_{21} + n_{22})]^{1/2}} \quad (8)$$

rispettano entrambe le condizioni precedenti. Tuttavia, nessuno dei tre coefficienti (6), (7) e (8) può considerarsi una vera misura perché, seppure per diversa motivazione, nessuno di essi assume sempre il valore -1 in corrispondenza di completa dissociazione e il valore $+1$ in corrispondenza di completa associazione, né presenta la proprietà di crescere o decrescere costantemente su una scala lineare. Quest'ultima proprietà rappresenta una "terza condizione" che un buon coefficiente di associazione a nostro avviso deve presentare. In base a questa condizione deve, cioè, verificarsi che il coefficiente quando varia da -1 a 0 e da 0 a $+1$ per effetto dello spostamento di una unità di frequenza entro le celle della tabella 2, con la condizione che i totali marginali rimangono invariati, aumenti di una quantità costante in corrispondenza a ciascuna delle infinite tabelle dicotomiche che è possibile osservare al variare dei totali marginali di riga e di colonna.

Peggiora è la situazione per i coefficienti proposti per misurare l'associazione tra due variabili con r categorie in tabelle di frequenze ($r \times r$). In questo caso, come è noto, sfruttando la sola condizione che

l'indice sia funzione della quantità $D_{ij} = n_{ij} - \frac{n_i n_j}{n_{..}}$ è stata proposta la quantità

$$X^2 = \sum_{ij} \frac{D_{ij}^2}{\frac{n_i n_j}{n_{..}}} = n_{..} \sum_{ij} \frac{D_{ij}^2}{n_i n_j} = n_{..} \left\{ \sum_{ij} \frac{n_{ij}^2}{n_i n_j} - 1 \right\} \quad (9)$$

che, come abbiamo già osservato, non può essere considerata un coefficiente di associazione perché il suo limite superiore cresce indefinitamente con $n_{..}$. Inoltre, non fornisce alcuna indicazione sul segno dell'associazione per distinguerlo dalla dissociazione.

Le cose non migliorano né con il cosiddetto coefficiente di associazione di Pearson

$$P = \left\{ \frac{X^2}{n+X^2} \right\}^{1/2} \quad (10)$$

che per tavole ($r \times r$) assume la forma

$$P = \left\{ \frac{r-1}{r} \right\}^{1/2} \quad (11)$$

né con l'indice di Tchuprov

$$T = \left\{ \frac{X^2}{n(r-1)} \right\}^{1/2} \quad (12)$$

che per tavole quadrate coincide con l'indice C di Cramer, perché non distinguono l'associazione dalla dissociazione, né assumono valore +1 nel caso di completa associazione.

Nel prossimo paragrafo presentiamo un nuovo coefficiente di associazione che, in corrispondenza delle due tipologie di tabelle (2×2) e ($r \times r$), fornisce misure dell'associazione che rispettano tutte le condizioni fin qui discusse.

3. Un nuovo coefficiente di associazione

Con riferimento al caso più generale di una tabella di frequenze ($r \times r$), un coefficiente di associazione che è funzione della quantità

$$D_{ij} = n_{ij} - \frac{n_i n_j}{n_{..}}, \text{ che cresce costantemente al crescere di } D_{ij} \text{ su una}$$

scala lineare, che coglie correttamente il segno positivo nel caso di associazione e negativo nel caso di dissociazione e che assume correttamente i valori -1 nel caso di completa dissociazione e $+1$ nel caso di completa associazione è dato da:

$$M = \pm \frac{\sum_{i,j} |D_{ij}|}{\sum_{i,j} |\max D_{ij}|} = \pm \frac{\sum_{i,j} |n_{ij} - \frac{n_i n_j}{n_{..}}|}{\sum_{i,j} |\max (n_{ij} - \frac{n_i n_j}{n_{..}})|} = \pm \frac{\sum_{i,j} |n_{ij} n_{..} - n_i n_j|}{\sum_{i,j} |\max (n_{ij} n_{..} - n_i n_j)|} \quad (13)$$

nel quale $\sum_{i,j} |D_{ij}|$ rappresenta la somma dei valori assoluti di tutte le differenze $D_{ij} = n_{ij} - \frac{n_i n_j}{n_{..}}$ tra le frequenze osservate n_{ij} e quelle teoriche $\frac{n_i n_j}{n_{..}}$, corrispondenti all'ipotesi di indipendenza stocastica, mentre

$\sum_{i,j} |\max D_{ij}|$ rappresenta la somma dei valori assoluti di tutte le differenze calcolate su una tabella di frequenze massimante dell'associazione o della dissociazione, cioè una tabella corrispondente alla completa associazione o alla completa dissociazione, che ha gli stessi totali marginali della tabella di frequenze osservata. È facile verificare che il coefficiente M varia da -1 a 0 nel caso di dissociazione e da 0 a $+1$ nel caso di associazione e fornisce la misura dell'associazione o della dissociazione tra le r categorie di due variabili categoriali, variando tale misura costantemente su una scala lineare. Il segno di M sarà positivo nel caso di associazione e negativo nel caso di dissociazione, come risulterà evidente in seguito.

3.1 Impiego del nuovo coefficiente di associazione al caso particolare di tavole dicotomiche

Nel caso di una tabella dicotomica, come abbiamo detto, la conoscenza dei totali marginali e della frequenza n_{11} equivale alla conoscenza di tutte le altre frequenze della tabella, per cui la determinazione della quantità $|D_{11}| = |n_{11} - \frac{n_{1.} n_{.1}}{n_{..}}|$ sulla tabella osservata e della quantità

$|\max D_{11}| = |\max \left(n_{11} - \frac{n_{1.} n_{.1}}{n_{..}} \right)|$ sono sufficienti per determinare il coefficiente di associazione M per questo tipo di tabelle che si semplifica nella forma (3)

(3) La semplificazione deriva dal fatto che tutti i $|D_{ij}|$ sono uguali, per cui si ha $\sum_{i,j} |D_{ij}| = 4 |D_{11}|$; lo stesso accade per i termini $|\max D_{ij}|$ per cui si ha anche $\sum_{i,j} |\max D_{ij}| = 4 |\max D_{11}|$; ne consegue che è $M = \pm \frac{\sum_{i,j} |D_{ij}|}{\sum_{i,j} |\max D_{ij}|} = \pm \frac{4 |D_{11}|}{4 |\max D_{11}|} = \pm \frac{|D_{11}|}{|\max D_{11}|}$.

$$M = \pm \frac{|D_{11}|}{|\max D_{11}|} = \pm \frac{|n_{11}n_{..} - n_{1.}n_{.1}|}{|\max(n_{11}n_{..} - n_{1.}n_{.1})|} \quad (14)$$

Il calcolo di M per una tabella di frequenze osservata con totali marginali fissati si sviluppa nei seguenti semplici passi.

1. Si calcola la differenza $[(n_{11} + n_{22}) - (n_{12} + n_{21})]$ tra la somma degli elementi della diagonale principale e quella degli elementi della diagonale secondaria. Se questa differenza è positiva, nella tabella c'è associazione positiva per cui il valore del coefficiente M sarà compreso tra 0 e 1 e quindi si rende necessario determinare la tabella massimante dell'associazione per potere calcolare $|\max D_{11}|$; se la differenza è negativa, allora nella tabella c'è dissociazione per cui il valore di M sarà compreso tra -1 e 0 e si rende necessario determinare la tabella massimante della dissociazione per poter calcolare $|\max D_{11}|$.

2. Dai totali marginali della tabella osservata si determinano le tabelle massimanti dell'associazione o della dissociazione procedendo, rispettivamente, nel modo seguente:

(a) si confrontano successivamente le coppie di totali $n_{1.}$ con $n_{.1}$ e $n_{2.}$ con $n_{.2}$, corrispondenti agli elementi della diagonale principale, si pongono sulla diagonale principale i valori più piccoli che scaturiscono dai confronti e si ricostruisce la tabella massimante dell'associazione sulla base dei totali marginali fissati;

(b) si confrontano successivamente le coppie di totali $n_{1.}$ con $n_{.2}$ e $n_{2.}$ con $n_{.1}$, corrispondenti agli elementi della diagonale secondaria, si pongono sulla diagonale secondaria i valori più piccoli che scaturiscono dai confronti e si ricostruisce la tabella massimante della dissociazione sulla base dei totali marginali fissati.

Un esempio numerico chiarirà meglio questo processo di calcolo del coefficiente M nelle due situazioni precedenti e mostrerà i suoi pregi e la sua correttezza informativa rispetto agli altri coefficienti proposti da Yule e da Pearson.

Kendall e Stuart (1979), per chiarire meglio come è possibile ottenere un effetto di mascheramento dell'associazione quando in essi si sommano gli "effetti uguali e contrari" di un altro attributo, riportano il seguente esempio (4): si consideri il caso in cui alcuni pazienti siano trattati per una malattia e altri non lo siano. Se A indica il ricovero in

(4) Ci riferiamo all'esempio 33.3 di pagina 574 dell'edizione citata.

ospedale e B indica il trattamento, supponiamo che le frequenze della tabella 2×2 siano:

	B	β	TOTALE
A	100	200	300
α	50	100	150
TOTALE	150	300	450

In questo caso si ha $n_{AB} = 100 = \frac{n_A \cdot n_B}{n}$, cosicché gli attributi sono indipendenti, per cui, come può vedersi, il trattamento non sembra esercitare alcun effetto sul ricovero.

Indicando il sesso maschile con C e quello femminile con γ si supponga che le frequenze per maschi e femmine siano, rispettivamente,

	BC	βC	TOTALE
AC	80	100	180
αC	40	80	120
TOTALE	120	180	300

	$B\gamma$	$\beta\gamma$	TOTALE
$A\gamma$	20	100	120
$\alpha\gamma$	10	20	30
TOTALE	30	120	150

Per queste tavole viene calcolato il coefficiente Q e si ottiene per il gruppo dei maschi $Q = 0.231$ e per il gruppo delle femmine $Q = -0.429$. Kendall e Stuart commentano così i risultati ottenuti: *thus treatment was positively associated with recovery among the males and negatively associated with it among the females. The apparent independence in the combined table is due to the cancelling of these associations*, non considerando che in base alle indicazioni dell'indice Q rimane pur sempre una indicazione di un'associazione negativa data da $-0.429 + 0.231 = -0.198$ che tuttavia non dovrebbe esistere.

Procediamo allora al calcolo del coefficiente M per questo stesso esempio. Per la tabella dei maschi si ha che $(80 + 80) - (100 + 40) = 20$

è positivo, per cui M sarà positivo e compreso tra 0 e 1. La tabella massimante dell'associazione è la seguente

	BC	βC	TOTALE
AC	120	60	180
αC	0	120	120
TOTALE	120	180	300

per cui risulta (5):

$$|D_{11}| = \left| 80 - \frac{180 \times 120}{300} \right| = |80 - 72| = 8 \quad (15)$$

$$|\max D_{11}| = \left| 120 - \frac{180 \times 120}{300} \right| = |120 - 72| = 48 \quad (16)$$

$$M = \frac{8}{48} = 0.1667 \quad (17)$$

(5) Se si impiega la relazione generale $M = \pm \frac{\sum_{i,j} |D_{ij}|}{\sum_{i,j} |\max D_{ij}|}$ si ha

$$\begin{aligned} \sum_{i,j} |D_{ij}| &= \left| 80 - \frac{180 \times 120}{300} \right| + \left| 100 - \frac{180 \times 180}{300} \right| + \left| 40 - \frac{120 \times 120}{300} \right| + \left| 80 - \frac{180 \times 120}{300} \right| = \\ &= |80 - 72| + |100 - 108| + |40 - 48| + |80 - 72| = 8 + 8 + 8 + 8 = 4 \times 8 = 32 \end{aligned}$$

$$\begin{aligned} \sum_{i,j} |\max D_{ij}| &= \left| 120 - \frac{180 \times 120}{300} \right| + \left| 60 - \frac{180 \times 180}{300} \right| + \left| 0 - \frac{120 \times 120}{300} \right| + \left| 120 - \frac{180 \times 120}{300} \right| = \\ &= |120 - 72| + |60 - 108| + |0 - 48| + |120 - 72| = 48 + 48 + 48 + 48 = 4 \times 48 = 192 \end{aligned}$$

$$M = \frac{32}{192} = \frac{4 \times 8}{4 \times 48} = 0.1667$$

Per la tabella delle femmine si ha che $(80 + 20) - (100 + 10) = -10$ è negativo, per cui M sarà negativo e compreso tra -1 e 0 . La tabella massimante della dissociazione è la seguente

	$B\gamma$	$\beta\gamma$	TOTALE
$A\gamma$	0	120	120
$\alpha\gamma$	30	0	30
TOTALE	30	120	150

per cui risulta:

$$|D_{11}| = \left| 20 - \frac{120 \times 30}{150} \right| = |20 - 24| = 4 \quad (18)$$

$$|\max D_{11}| = \left| 0 - \frac{120 \times 30}{150} \right| = |0 - 24| = 24 \quad (19)$$

$$M = -\frac{4}{24} = 0.1667 \quad (20)$$

Come si vede, i due valori si compensano esattamente e danno lo stesso valore $M = 0$ che si era ottenuto nel caso della tabella composta iniziale.

3.2 Impiego del nuovo coefficiente di associazione in tabelle di frequenze ($r \times r$)

Per misurare l'associazione in tabelle di frequenze ($r \times r$) bisogna impiegare la (13) nella sua espressione più generale e calcolare tutti i $|D_{ij}|$ della tabella osservata e tutti i $|\max D_{ij}|$ di una delle due tabelle massimanti l'associazione o la dissociazione, a seconda del segno positivo o negativo da assegnare ad M . Ovviamente, per ottimizzare i tempi di elaborazione conviene impiegare il coefficiente nella forma

$$M = \pm \frac{\sum_{i,j} |n_{ij}n_{..} - n_{i.}n_{.j}|}{\sum_{i,j} |\max(n_{ij}n_{..} - n_{i.}n_{.j})|} \quad (21)$$

Anche in questo caso occorre determinare prima il segno di M dalla tabella delle frequenze osservate stabilendo così se siamo in presenza di associazione o di dissociazione e decidere quale delle due tabelle massimanti utilizzare. Il criterio per la determinazione del segno può essere fondato ancora sulla differenza della somma delle frequenze sulla diagonale principale, $\sum_{i=1}^r n_{ii}$, con quella delle frequenze sulla diagonale secondaria, $\sum_{i=1}^r n_{i(r-i+1)}$, e fissare il segno più se il risultato è positivo, il segno meno se è negativo.

Si procede, quindi, alla determinazione della tabella massimante l'associazione se il segno di M è positivo o della tabella massimante la dissociazione se il segno di M è negativo. I criteri da seguire sono analoghi a quelli impiegati per le tabelle (2 x 2), solo che in questo caso le coppie di totali marginali da considerare sono r e cioè:

a) $(n_1, n_1), (n_2, n_2), \dots, (n_r, n_r)$ per la tabella massimante l'associazione e

b) $(n_1, n_r), (n_2, n_{(r-1)}), \dots, (n_r, n_1)$ per la tabella massimante la dissociazione.

A questo punto è doveroso osservare che seguendo il nostro metodo per stabilire il segno dell'indice M anche i coefficienti

$$X_r^2 = \pm \frac{X^2}{\max X^2} \quad (22)$$

e

$$P_r = \pm \frac{P}{\max P} \quad (23)$$

variano tra -1 e $+1$; tuttavia, come è facile verificare, anche in questo caso X_r^2 e P_r non rispettano la condizione che abbiamo fissato di essere una vera misura che varia linearmente al modificarsi di una unità elementare di frequenza nella tabella.

Per far vedere la validità di questa affermazione abbiamo calcolato l'andamento, per lo spostamento di unità di frequenze all'interno delle celle della tabella, dei tre coefficienti M , X_r^2 e P_r , con il vincolo

che i totali marginali della tabella siano fissati. Si è considerata, in particolare, la seguente tabella di indipendenza:

10	10	10	30
10	10	10	30
10	10	10	30
30	30	30	90

Se si sposta una delle unità estreme verso la diagonale principale si ottiene la tabella

11	10	9	30
10	10	10	30
9	10	11	30
30	30	30	90

che, per il vincolo sui totali marginali, prevede lo spostamento di un'altra unità simmetrica sulla diagonale principale. La seconda tabella è quindi

12	10	8	30
10	10	10	30
8	10	12	30
30	30	30	90

e così via, fino all'ultima tabella che costituisce la tabella massimante l'associazione

30	0	0	30
0	30	0	30
0	0	30	30
30	30	30	90

Se ora si rappresentano graficamente (Fig. 1) i trenta valori di ciascuno dei coefficienti M , X_r^2 e P_r , calcolati su ciascuna delle trenta tabelle che è possibile ottenere, tante quante risultano dalla quantità $\sum_{i < j} n_{ij}$ a partire dalla tabella di indipendenza, si ottengono le evoluzioni di ciascuno di essi da 0 a 1 per il cambiamento costante di due

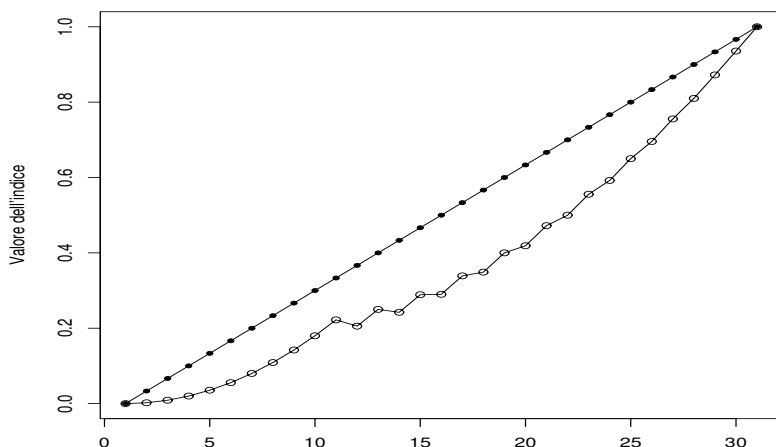


Fig. 1 - Andamento dei tre indici M (punti pieni), X_r^2 e P_r (punti vuoti).

unità nella composizione di ciascuna tabella successiva. Come può vedersi, mentre i valori del coefficiente M seguono l'andamento di una retta, i due coefficienti X_r^2 e P_r seguono, come era da attendersi, lo stesso andamento curvilineo, a dimostrazione del fatto che nei diversi punti della scala gli indici non misurano, come avrebbero dovuto, la stessa quantità di variazione.

4. Conclusioni

In questo lavoro abbiamo proposto un nuovo coefficiente di associazione per la misura della interdipendenza di due variabili categoriali sulla base delle frequenze di tabelle quadrate ($r \times r$). Nel corso della esposizione abbiamo dimostrato le proprietà ottimali del coefficiente che rispetta tutte e tre le condizioni che a nostro avviso un buon indice deve presentare e lo abbiamo confrontato con i coefficienti Q , Y e V da sempre molto utilizzati nel caso di tabelle dicotomiche (2×2) e con i coefficienti X_r^2 e P_r per tabelle di frequenze ($r \times r$), mostrando i seri limiti che questi coefficienti presentano. Rinviamo a lavori successivi lo studio delle sue proprietà elementari quando affronteremo problemi inferenziali connessi con ipotetiche distribuzioni sottostanti, soprattutto per variabili categoriali ottenute a partire da una variabile quantitativa.

BIBLIOGRAFIA

- CRAMER H. (1946), *Mathematical Methods of Statistics*, Princeton University Press, Princeton.
- KENDALL M., STUART A. (1979). *The Advanced Theory of Statistics*, Volume 2, C. Griffin, London.
- PEARSON K. (1904), *On the theory of contingency and its relation to association and normal correlation*, Drapers' Co. Memoirs, Biometric Series, London.
- YULE G. (1900), *On the association of attributes in statistics*, Philosophical Transaction (A), 194, 257.
- YULE G. (1912), *On the methods of measuring association between two attributes*, Journal of the Royal Statistical Society, 75, 579.

ALIDA PERNICE

RECENTE DINAMICA SOCIO-DEMOGRAFICA DELLE PROVINCIE SICILIANE E PROBABILE EVOLUZIONE FINO AL 2011 DELLA POPOLAZIONE IN ETÀ LAVORATIVA

1. Introduzione

Per l'analisi del mercato del lavoro è fondamentale la conoscenza sia dei fattori di natura socio-economica sia delle variabili e della struttura demografica del territorio di riferimento.

Il calo della natalità, iniziato in Italia dopo il boom delle nascite degli anni sessanta e l'aumento della vita media, fenomeni entrambi in corso in tutti i Paesi evoluti, hanno determinato i processi d'invecchiamento della popolazione, che tuttavia in Sicilia sono attenuati da livelli di natalità che si mantengono fino ad oggi superiori alla media nazionale.

I segni di questa grande trasformazione demografica sono ben evidenti nelle dinamiche produttive ed economiche. La denatalità degli anni passati investe oggi sia la popolazione che frequenta la scuola dell'obbligo, sia quella parte di popolazione che si immette nel mercato del lavoro, producendo probabilmente degli effetti di riduzione dell'offerta di lavoro. Inoltre l'invecchiamento della popolazione tende a ridurre il peso delle classi centrali d'età, cioè di quella parte di popolazione che, come è noto, si fa carico degli oneri sociali per il mantenimento della popolazione in età senile.

La situazione della Sicilia, osservata in relazione a questi due fenomeni così appariscenti e pressanti dell'attuale realtà demografica, appare fino ad oggi meno drammatica rispetto a quella di altri Paesi europei e dell'Italia.

In particolare, la Sicilia registra nel 2004 un tasso di fecondità totale pari a 1,43 figli per donna in età feconda, valore superiore –

insieme a quello di Campania e Trentino – alla media nazionale pari a 1,30 (1).

Anche il processo d'invecchiamento è più lento: la durata media di vita (2) ha continuato ad allungarsi negli anni. Le donne hanno superato la soglia degli ottanta anni: al 2003 l'aspettativa di vita femminile alla nascita è di 82,5 anni con un'eccedenza di circa sei anni rispetto agli uomini; l'indice di vecchiaia, ottenuto quale rapporto fra il numero di individui di oltre sessantacinque anni ed il totale, è pari a 17,2% per entrambi i sessi, a fronte del valore medio nazionale di 19% (Tab. 1).

Tab. 1 - Numero medio di figli per donna, speranza di vita alla nascita, percentuale popolazione oltre 65 anni al 1° gennaio 2003.

Ripartizioni territoriali	Numero medio figli per donna	Speranza di vita alla nascita		% popolazione oltre 65 anni		
		Maschi	Femmine	Maschi	Femmine	Totale
Sicilia	1,43	76,70	81,60	15,20	19,10	17,20
ITALIA	1,30	76,80	82,50	16,20	21,70	19,00

Fonte: www.demo.istat.it

2. La dinamica della popolazione siciliana

Al Censimento del 1991 la popolazione residente nell'Isola è risultata di 4.966.186 abitanti, al censimento del 2001 risulta pari a 4.968.991, con un incremento di sole 2.805 unità (0,56%) nel decennio.

La sostanziale stazionarietà della popolazione è stata determinata dal controbilanciarsi di due fenomeni: le immigrazioni di ritorno e di lavoratori stranieri, provenienti per lo più dai Paesi del bacino del mediterraneo, che vanno a coprire posti di lavoro respinti dalla forza lavoro locale e la fuga di giovani siciliani verso l'estero e verso il Nord Italia.

Per comprendere la recente evoluzione del movimento migratorio, nella Tab. 2 sono state disaggregate le migrazioni interne da quelle con l'estero.

(1) Il tasso di fecondità totale TFT misura l'intensità del ricambio generazionale: se è inferiore a 2 non viene garantita nel tempo la sostituzione della generazione dei genitori con quella dei figli.

(2) La speranza di vita (o aspettativa di vita) all'età zero esprime il numero medio di anni che un soggetto alla nascita può attendersi di vivere.

Tab. 2 - Tassi di migratorietà per mille abitanti.

Ripartizioni territoriali	Tasso migratorio interno			Tasso migratorio con l'estero			Tasso migratorio totale		
	2002	2003	2004	2002	2003	2004	2002	2003	2004
Sicilia	-2,8	-2,3	-2,6	0,5	2,1	1,3	0,3	5,6	0,8
ITALIA	—	—	—	3	7,1	7,7	6,1	10,5	9,4

N.B. - Il saldo migratorio totale è comprensivo del saldo migratorio non dovuto ad un effettivo trasferimento di residenza, ma ad operazioni di rettifica anagrafica.

Fonte: www.demo.istat.it

Al negativo saldo migratorio interno nel triennio 2002-2004, si contrappone il positivo saldo migratorio con l'estero, sia per le immigrazioni di ritorno che per i flussi immigratori.

In particolare nel triennio 2002-04 la crescita della Sicilia è determinata da una lieve crescita naturale (divario fra tassi di natalità e di mortalità) e da un tasso migratorio positivo.

Passando ad un'analisi a livello provinciale, per il 2004 risaltano i tassi di migratorietà di Ragusa (8,1‰) e Trapani (7,0‰). Il primo legato alla continua richiesta di manodopera per la raccolta dei prodotti dell'agricoltura, il secondo dovuto alla pressante richiesta nell'attività della pesca. Di contro il più alto saldo migratorio negativo viene registrato ad Enna (-7,6‰) provincia che presenta nel complesso una situazione allarmante di spopolamento con tasso di decremento di -8,7‰; Palermo ha un saldo migratorio negativo di -2,0‰, ma una crescita naturale positiva (1,8‰); infine Messina registra un saldo migratorio negativo (-0,3‰) che sommandosi al decremento naturale (-1,8‰) della popolazione, porta a una diminuzione di popolazione pari al 2,1‰ (Tab. 3).

Tab. 3 - Tasso di crescita naturale, di migratorietà e di crescita totale per 1.000 abitanti.

Province	Tasso di crescita naturale			Tasso migratorio totale			Tasso di crescita totale		
	2002	2003	2004	2002	2003	2004	2002	2003	2004
Palermo	2,0	1,7	1,8	-0,8	-0,2	-2,0	1,2	1,5	-0,2
Ragusa	1,0	0,8	1,0	4,0	24,3	8,1	5,0	25,1	9,1
Trapani	0,5	-0,4	0,0	1,0	7,6	7,0	1,5	7,2	7,0
Agrigento . . .	0,8	0,1	0,5	4,4	14,9	-1,0	5,2	15,0	-0,5
Catania	2,3	2,2	2,3	1,7	6,4	1,9	4,0	8,6	4,2
Enna	-0,3	-1,1	-1,1	-2,3	-5,5	-7,6	-2,6	-6,6	-8,7
Messina	-1,9	-2,4	-1,8	-1,5	1,4	-0,3	-3,4	-1,0	-2,1
Siracusa	0,8	0,9	0,8	0,1	1,3	0,7	0,9	2,2	1,5
Caltanissetta .	1,1	1,0	1,3	-6,3	12,7	2,8	-5,2	13,7	4,1
SICILIA	1,0	0,7	0,9	0,3	5,6	0,8	1,3	6,3	1,7

Fonte: www.demo.istat.it

3. La struttura della popolazione siciliana

La distribuzione per le tre grandi classi d'età (0-14), (15-64), (65 e oltre) mette in evidenza un lento ma graduale processo d'invecchiamento. Le tendenze evolutive mostrano un decremento di giovani in età inferiore ai 14 anni, una certa stazionarietà nella classe di età lavorativa ed un incremento di popolazione nell'età senile.

Più precisamente, la fascia d'età centrale presenta valori pressoché costanti e di ammontare cospicuo, in quanto alimentata dall'ingresso delle leve particolarmente numerose nate agli inizi degli anni '60 (baby boom), che andranno ad alimentare fra circa 20 anni la terza fascia di età, 65 – oltre.

Nella Tab. 4 figura come nel decennio 1991-2001 il segmento di popolazione 0-14 anni si è ridotto di 3 punti (da 20,13 a 17,13 %); mentre quello fra 15 e 64 anni è costante, intorno al 66%. La popolazione senile invece aumenta di circa 3 punti percentuali confermando la tendenza nazionale (Graf. 1).

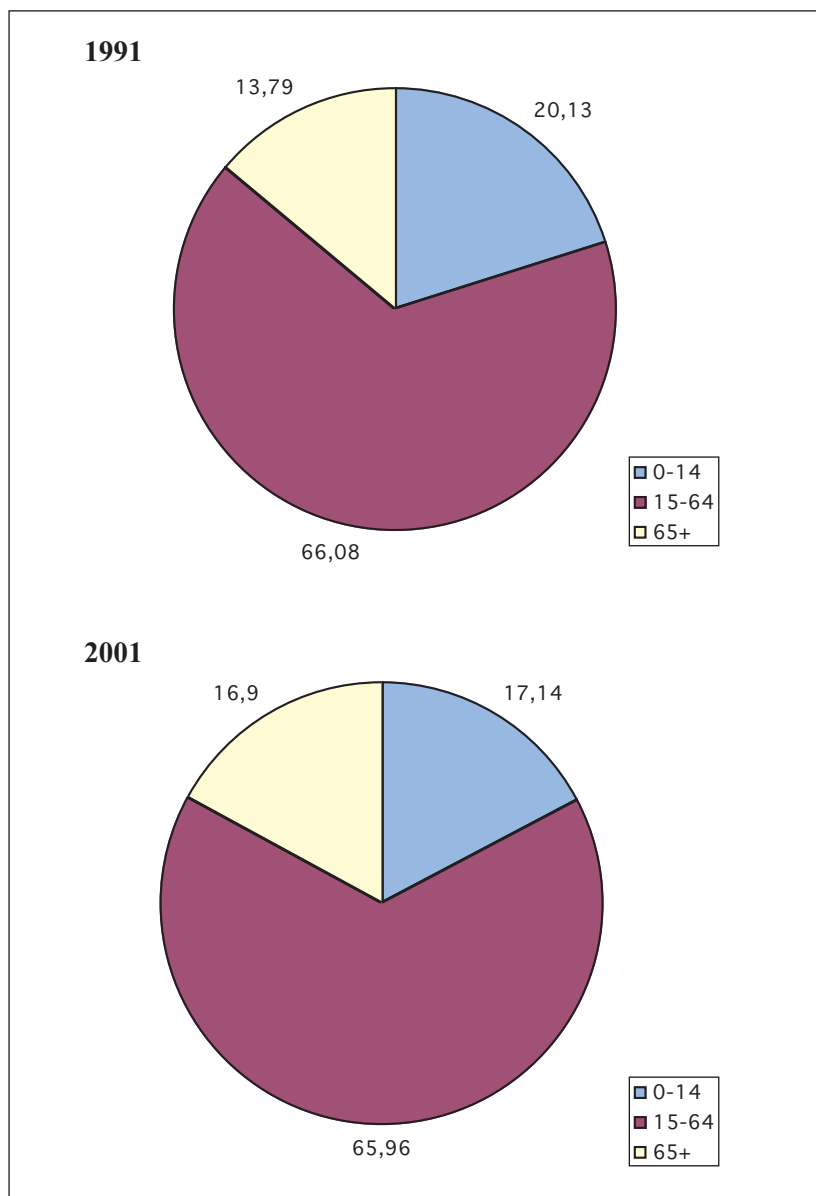
Le province con maggiore incidenza di popolazione anziana sulla popolazione totale sono Enna e Messina. Quest'ultima, inoltre, ha il più basso peso di popolazione fra 0-14 anni (15,37%).

Riguardo alle differenze per sesso, particolare rilievo assume l'ultimo segmento di vita, dove nell'intera isola nel 1991, il divario tra popolazione maschile e femminile è di 2,85 punti; mentre nel 2001 è di 3,7. In particolare Messina con 5,8 e Trapani con 4,0 punti di differenza tra M e F, sono le province in cui il processo di femminilizzazione tipico dell'età avanzata, è più evidente.

Per meglio analizzare le interrelazioni fra situazione demografica e realtà socio-economica della Sicilia, sono stati calcolati, alla luce dei dati censuari del 1991 e del 2001, alcuni indicatori delle principali caratteristiche strutturali della popolazione siciliana sia a livello regionale che provinciale.

4. L'invecchiamento della popolazione siciliana

Gli effetti notevoli del declino della fecondità e dell'allungamento della vita, sono messi in chiara luce dall'indice di vecchiaia (I.V.), calcolato come rapporto tra l'ammontare della popolazione dai 65 anni in poi e la popolazione in età inferiore ai 14 anni.



Graf. 1 - Distribuzione percentuale della popolazione siciliana, per grandi classi di età. Censimenti 1991-2001.

Tale indicatore, nel 1991, era pari al 68,50 anziani per ogni 100 giovanissimi per entrambi i sessi; nell'arco di 10 anni esso è passato a 98,67%.

Da un'analisi per sesso, emerge come la Sicilia si comporta al pari di tutti i Paesi europei, in cui è già accentuato il fenomeno della "femminilizzazione" della popolazione, dovuto alla maggiore longevità del sesso femminile (3).

Nel 1991 la popolazione anziana femminile era di 79,70%; nel 2001 il valore si è attestato a 115,59 ultra sessantacinquenni per ogni 100 femmine in età 0-14.

Dal confronto dei dati provinciali, si evince che la struttura demografica maggiormente compromessa dal processo di senilizzazione, è quella di Messina con un'indice di vecchiaia di 150,14% per le femmine, solo 100,67% per i maschi e 124,74% per il complesso dei sessi.

Tab. 5 - Indici di vecchiaia della popolazione siciliana per sesso. Censimenti 1991-2001.

Province	1991			2001		
	Maschi	Femmine	Totale	Maschi	Femmine	Totale
Palermo	51,56	73,10	61,94	74,47	107,01	90,32
Ragusa	64,77	85,89	75,08	86,78	117,78	101,85
Trapani	67,75	91,82	79,47	93,11	127,90	110,13
Agrigento	62,14	79,43	70,53	89,33	120,59	104,53
Catania	49,81	69,41	59,32	72,83	101,17	86,69
Enna	69,01	86,75	77,81	94,47	125,95	109,93
Messina	72,10	104,59	87,92	100,66	150,14	124,74
Siracusa	55,61	74,81	64,93	84,38	113,64	98,65
Caltanissetta	56,90	75,69	66,03	78,81	106,78	92,45
SICILIA	57,93	79,70	68,50	82,56	115,59	98,67

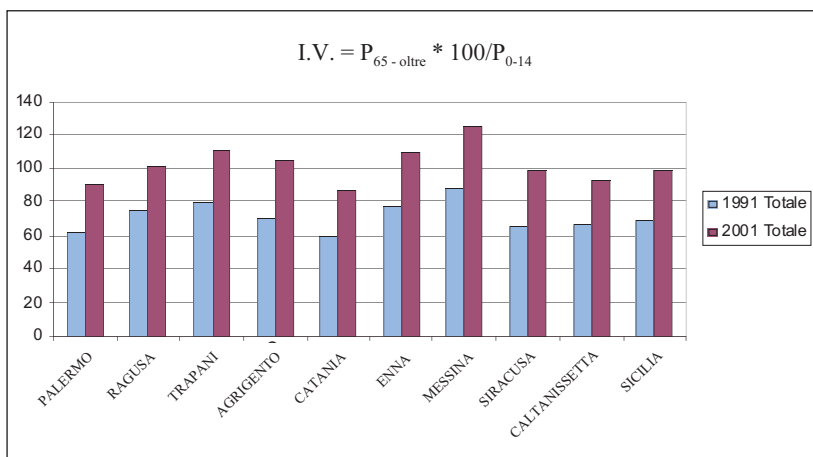
Fonte: Ns. elaborazioni su dati ISTAT dei censimenti della popolazione 1991 e 2001.

Ciò è da attribuirsi probabilmente ai fenomeni di esodo migratorio di popolazioni in età lavorativa e riproduttiva dai numerosi comuni montani del messinese.

La provincia più giovane è Catania, il cui indice nel 1991 è del 59,32% mentre nel 2001 assume il valore di 86,69% (per entrambi i

(3) Cfr. Tavole di mortalità della popolazione italiana - Regioni, province e grandi comuni. Anno 2000. ISTAT. MASSIMO LIVI BACCI, *Introduzione alla Demografia*, 1997, Loescher.

sessi); l'indice per l'intera isola è di 68,50 al censimento del 1991 e di 98,67 % al censimento del 2001.



Graf. 2 - Indice di vecchiaia della popolazione siciliana. Censimenti 1991-2001.

5. Gli indicatori demo-economici

Le variazioni osservate nella struttura demografica per il decennio intercensuale, se dovessero perdurare nei prossimi anni, potrebbero a breve portare a sconvolgimenti nel sistema economico.

In Sicilia nessuno elemento fa ipotizzare un'inversione della tendenza demografica attuale; piuttosto il calo della natalità e l'aumento della speranza di vita fanno presagire processi d'invecchiamento determinati, alla luce di quanto esposto, tanto dall'aumento di popolazione anziana quanto dalla diminuzione di quella giovanile.

I dati censuari forniscono ulteriori elementi di riflessione circa il significato delle trasformazioni strutturali in atto. Se si analizza l'entità delle popolazioni in età lavorativa rispetto alla popolazione che grava su di essa, sia in termini assistenziali che economici, ci si rende conto che il rapido processo d'invecchiamento ha già prodotto e purtroppo continuerà a produrre una serie di problemi di natura sociale e di politica economica.

Con l'indice di dipendenza (I.D.), gli anziani e i giovanissimi sono posti in rapporto alle persone che devono farsene carico in termini sia assistenziali che economici.

Il valore di I.D. a livello regionale si mantiene nei due anni considerati (1991 e 2001) intorno al 50%. Infatti come già rilevato la quota di popolazione in età lavorativa è rimasta nel decennio pressoché invariata, mentre è crescente la quota di popolazione in età non attiva. L'analisi provinciale mette in luce che Siracusa è la provincia con l'indice I.D. più basso; difatti si rilevano nel 2001, 48 persone appartenenti alle due fasce d'età 0-14 e 65 e oltre, per ogni 100 fra 15 e 64 anni; mentre Agrigento, Enna e Caltanissetta sono le province con l'indice più elevato (Tab. 6).

Generalmente però la stabilità dell'indice di dipendenza nasconde delle modificazioni nelle caratteristiche strutturali della popolazione a carico. Difatti la riduzione della popolazione giovane viene controbilanciata dalla crescita della popolazione anziana.

L'indice di struttura (I.S.) è un indicatore del grado di invecchiamento della popolazione in età lavorativa. Si ottiene rapportando la popolazione fra i 40 e i 64 anni a quella giovane fra i 15 e i 39 anni che sarà destinata a sostituirla. Se la popolazione giovane è crescente, il valore I.S. risulterà inferiore a 100; se la popolazione è a lento rinnovo, cioè in regresso demografico, I.S. sarà superiore a 100 esprimendo così che la popolazione fra i 15 e 39 anni non è sufficiente a coprire i posti di lavoro che nel tempo verranno lasciati liberi dall'attuale popolazione in età 40-64 anni.

Tab. 6 - Indici di dipendenza, di struttura e di ricambio della popolazione siciliana.

Province	I.D.		I.S.		I.R.	
	1991	2001	1991	2001	1991	2001
Palermo	50,90	50,50	70,18	80,31	58,28	76,38
Ragusa	51,97	52,59	75,63	82,15	67,27	87,23
Trapani	51,67	52,97	76,98	84,02	65,91	92,32
Agrigento	52,09	53,95	71,98	76,69	64,12	85,16
Catania	50,63	50,67	69,73	78,92	59,47	73,08
Enna	53,46	55,15	75,10	79,50	70,84	86,94
Messina	52,78	52,75	77,37	85,60	71,56	86,77
Siracusa	49,03	48,04	73,08	81,31	57,54	84,94
Caltanissetta	51,94	53,60	71,67	77,65	60,69	79,83
SICILIA	51,33	51,60	72,54	80,67	62,42	80,90

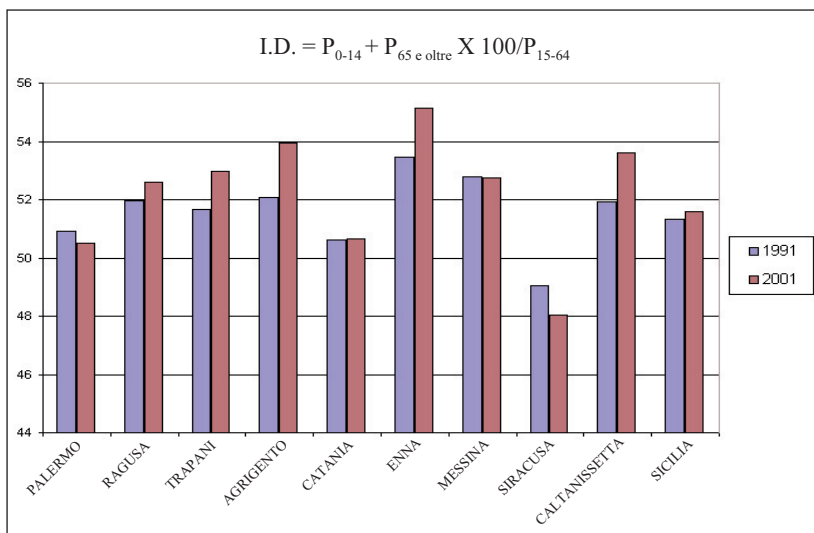
Fonte: Ns. elaborazioni su dati ISTAT dei censimenti della popolazione 1991 e 2001.

Dal confronto tra il 1991 ed il 2001 emerge che l'indice tende a salire, pur mantenendosi al di sotto di cento, indicando un basso ricambio generazionale.

L'indice a livello regionale passa da 72,54 nel 1991 a 80,67% nel 2001. Nei prossimi anni comunque sarà molto evidente il progressivo invecchiamento della popolazione in età lavorativa: la quota di persone in età più matura (40-64), continuerà ad aumentare per l'ingresso del consistente ammontare di popolazione che al Censimento del 2001 si trova ancora nella fascia 35-39, che proviene dalle generazioni dei nati negli anni dell'alta natalità del '60.

Per valutare nel suo pieno significato il processo d'invecchiamento della popolazione in età attiva, viene adoperato l'indice di ricambio della popolazione attiva (I.R.), anch'esso legato come gli indicatori precedenti, al mondo del lavoro. Esso esprime il ricambio generazionale tra le coorti di lavoratori in via di pensionamento e quindi in uscita dal mercato del lavoro e le coorti più giovani che si apprestano a farvi ingresso.

Generalmente questo rapporto supera il valore di 100 nelle popolazioni in decremento, dove il carico sociale della componente anziana non produttiva sarà oneroso per la mancanza di ricambio da parte delle leve giovani.



Graf. 3 - Indice di dipendenza della popolazione siciliana. Censimenti 1991-2001.

I valori sotto il limite di 100 sono invece indicativi di scarsa disponibilità di posti di lavoro per i giovani (pochi escono dal mercato del lavoro in relazione a coloro che vi entrano).

In Sicilia nel 1991, l'indice di ricambio era di 62,42%, mentre nel 2001 esso assume il valore di 80,90%. Dall'analisi provinciale emerge che Trapani e Ragusa presentano situazioni più favorevoli con indici rispettivamente di 92,32% e 87,23%; Siracusa è la provincia che nel decennio ha registrato la variazione più alta: qui l'indice passa da 57,54% nel 1991 a 84,94% nel 2001; Palermo e Catania infine presentano situazioni meno favorevoli al ricambio.

6. Previsioni della popolazione in età lavorativa in Sicilia

Per l'analisi delle problematiche del mercato del lavoro e delle tendenze di breve periodo dell'occupazione e della disoccupazione, è stata stimata la struttura della popolazione in età lavorativa al prossimo Censimento del 2011.

Si è proceduto con il metodo delle componenti, che consiste nel determinare l'ammontare futuro della popolazione per sesso e per classi d'età, applicando iterativamente per il periodo di previsione i tassi di fecondità, di sopravvivenza e di migratorietà, ipotizzati sulla base di convenienti valutazioni di tutti gli elementi demografici conoscitivi disponibili.

In linea con gli obiettivi, si è operato solo con riferimento alla popolazione in età lavorativa al 2001. Ciò ha consentito di evitare valutazioni sia sulle future nascite, sia sulla sopravvivenza delle età senili. Per il segmento di popolazione 15-64 anni, può essere accettabile ipotizzare fino al 2011, la costanza della legge di mortalità, sintetizzata nelle probabilità di sopravvivenza decennali calcolate sulla base delle tavole di mortalità e sopravvivenza a livello provinciale del 2000 pubblicate dalle dall'ISTAT.

Per quanto riguarda le migrazioni, sono caratterizzate è noto, da irregolarità del loro andamento, in quanto connesse con la struttura socio-economica e con le normative di legge sulla regolarizzazione di flussi di immigrazione dai paesi extracomunitari.

Le statistiche dei movimenti migratori si dimostrano imprecise e lacunose ed in particolare a livello provinciale, non viene effettuata da fonti ufficiali alcuna rilevazione per sesso e per classi di età; tuttavia

valutazioni delle tendenze della popolazione e dei suoi effetti sul mercato del lavoro, basate solo sulla sviluppo naturale porterebbe a dei risultati distorti non rispondenti alla realtà socio-economica del territorio.

Nel modello previsivo oltre alle ipotesi sulla sopravvivenza sono state fatte delle valutazioni sulla migratorietà, con stime dei saldi migratori ottenute con l'ausilio di tavole di mortalità dell'intervallo intercensuario 1991-2001.

L'analisi è stata articolata in due fasi successive: nella prima, la popolazione per classi quinquennali di età e per sesso al censimento del 1991 è stata stimata al 2001 con l'ausilio delle probabilità prospettiche di sopravvivenza e in assenza di movimento migratorio, indi confrontata con la popolazione rilevata al censimento del 2001; per differenza fra i due valori è stato ottenuta una stima del movimento migratorio nel decennio.

In ipotesi di tassi di migratorietà costanti e pari a quelli stimati nell'intervallo 1991-2001 e con probabilità di sopravvivenza desunte dalle tavole attuariali a livello provinciale del 2000, è stata valutata la popolazione in età lavorativa al 2011, per gruppi quinquennali di età, per sesso nelle nove province della Sicilia.

Lo schema di calcolo utilizzato è dato dalla relazione:

$${}_{t+10}\bar{P}_{x+10, x+14} = {}_tP_{x, x+4} * p_{x, x+4}$$

— ${}_{t+10}\bar{P}_{x+10, x+14}$ è l'ammontare teorico della popolazione al censimento 2001, in assenza di migrazioni.

— ${}_tP_{x+4}$ è l'ammontare della popolazione nella classe di età quinquennale all'epoca iniziale (censimento 1991);

— $p_{x, x+4}$ è la probabilità prospettica di sopravvivenza all'epoca intermedia 1996 (4).

La differenza tra popolazione censita e ammontare teorico al 2001 fornisce la stima del numero netto di migranti nel decennio, nella coorte che in età $x, x+4$ al censimento del 1991, è passata in età $x+10, x+14$ al censimento del 2001.

$${}_{t,t+10}SM_{(x, x+4), (x+10, x+14)} = {}_{t+10}P_{x+10, x+14} - {}_{t+10}\bar{P}_{x+10, x+14}$$

(4) Cfr. Tavole provinciali di mortalità - informazioni n. 12-2000. ISTAT.

Nel valutare l'andamento dei flussi migratori fra il 2001 ed il 2011, è stato ipotizzato un comportamento analogo al decennio 1991-2001, con tassi migratori negativi nelle classi di età lavorative per le province di Enna, Palermo, Agrigento e Messina, mentre positivi a Ragusa e Trapani, province entrambe di fonte di lavoro per le popolazioni dell'area mediterranea disposte ad accettare rapporti di lavoro caratterizzati da livelli di garanzia, di sicurezza e di redditi rifiutati dai siciliani.

In definitiva la popolazione nelle classi quinquennali di età lavorativa al 2011 per sesso e per le nove province è stata ottenuta secondo il seguente schema di calcolo:

$${}_{2011}P_{x+10, x+14} = {}_{2001}P_{x, x+4} \cdot {}_{2000}P_{x, x+4} \pm {}_{2011, 2001}SM_{(x, x+4) (x+10, x+14)}$$

I risultati delle elaborazioni sono contenuti nella Tab. 7.

Dall'analisi delle tendenze evolutive della popolazione siciliana si rileva al 2011 un ammontare pari a 3.224.606 individui in età 15-64 anni; mostrando un lieve declino di 53.059 unità rispetto al 2001.

Il metodo previsivo applicato ha consentito di stimare l'evoluzione della popolazione a livello provinciale per sesso e per classi quinquennali di età.

L'ammontare della popolazione che al 2011 si trova nelle prime due classi di età è maggiore di quella che ne uscita nel decennio nella classe di età 55-64; tuttavia fatta eccezione solo per Catania e Messina, in tutte le altre province si prevede una lieve riduzione di popolazione, per il processo migratorio stimato di intensità pari a quello del decennio 1991-2001.

I principali aspetti delle conseguenze sul mercato del lavoro, delle variazioni della popolazione siciliana in età lavorativa al 2011, sono stati quantificati negli indicatori demo-economici già calcolati ai censimenti del 1991 e del 2001.

La variazione della struttura per età della popolazione in età lavorativa e precisamente il suo invecchiamento è misurata dall'indice di struttura (I.S.).

È stato stimato, sulla base della probabile evoluzione della popolazione, un indice di 107,36 individui fra 40-64 anni per ogni 100 fra 15-39 anni, per cui emerge un processo d'invecchiamento con conseguente difficoltà di adattamento della popolazione alle esigenze del mercato.

Tab. 7 - Previsione della popolazione siciliana in età lavorativa per sesso al censimento 2011.

Classi di età	PA	RG	TP	AG	CT	EN	ME	SR	CL	Sicilia
Maschi										
15-19	37.126	8.804	12.008	12.142	32.341	5.010	18.208	10.693	8.545	144.877
20-24	38.311	8.874	11.778	12.619	33.557	4.916	19.782	11.697	8.088	149.622
25-29	35.620	8.748	11.388	11.497	31.787	4.752	18.638	11.459	7.376	141.265
30-34	39.145	9.791	12.657	12.743	34.372	4.942	21.332	12.547	7.933	155.462
35-39	42.977	11.007	14.711	14.214	37.987	5.684	24.436	14.400	9.307	174.723
40-44	43.764	11.154	15.105	14.914	38.368	5.886	25.192	14.315	9.530	178.228
45-49	43.961	10.836	14.633	14.667	37.750	5.922	24.800	14.163	9.576	176.308
50-54	40.636	9.707	13.386	13.389	34.885	5.305	23.038	12.841	8.441	161.628
55-59	37.581	8.773	12.658	12.182	32.136	5.012	21.708	11.664	7.538	149.252
60-64	37.013	8.626	12.836	12.206	30.360	5.074	20.879	12.294	7.503	146.791
15-64	396.134	96.320	131.160	130.573	343.543	52.504	218.013	126.073	83.837	1.578.156
Femmine										
15-19	36.521	8.331	11.664	11.881	31.529	4.710	17.425	10.467	8.322	140.850
20-24	37.619	8.669	11.925	11.990	33.239	4.974	19.154	11.038	8.181	146.789
25-29	36.268	8.683	11.547	11.810	31.639	4.775	18.784	11.200	7.593	142.299
30-34	40.913	9.569	13.100	12.816	35.857	5.136	21.338	12.669	8.395	159.793
35-39	45.633	10.758	15.202	14.936	39.505	5.907	24.715	14.502	9.697	180.855
40-44	47.016	11.064	16.050	16.273	41.068	6.234	25.720	14.723	10.123	188.271
45-49	48.009	10.855	15.590	15.654	40.981	6.219	26.498	14.531	10.479	188.816
50-54	44.390	10.165	14.351	14.733	38.378	6.087	24.864	13.662	9.417	176.047
55-59	40.981	9.508	13.578	13.626	35.469	5.616	22.570	12.564	8.690	162.542
60-64	40.254	9.620	13.793	13.538	33.633	5.653	22.407	12.992	8.298	160.188
15-64	417.604	97.222	136.800	37.258	361.298	55.312	223.475	128.348	89.195	164.645

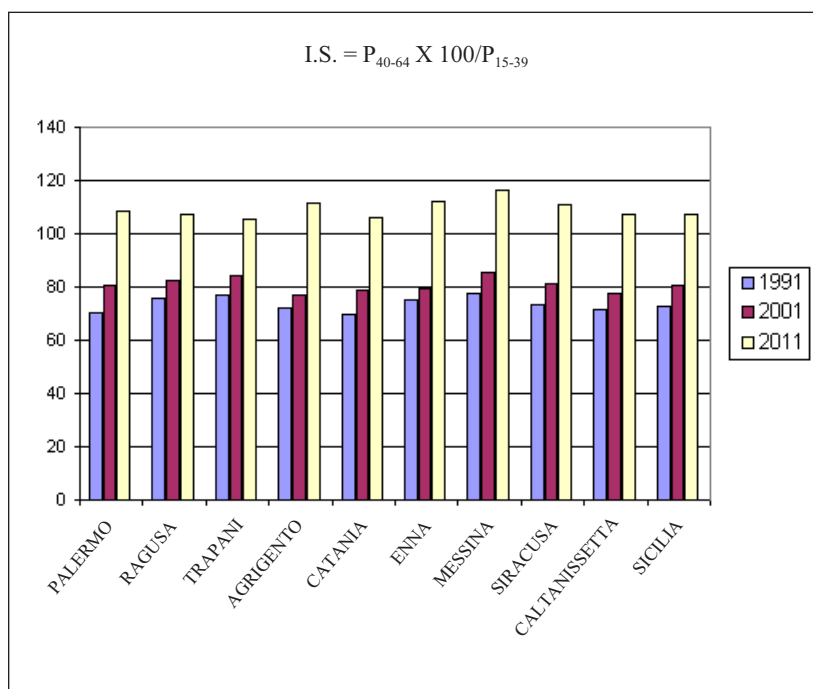
Fonte: Ns. elaborazione.

Crescerà anche l'indice di ricambio (I.R.) con effetti negativi sul sistema economico e pensionistico, in quanto coloro che tendono a uscire dalla fascia attiva di popolazione non vengono sufficientemente sostituiti dai giovani in età 15-19 anni.

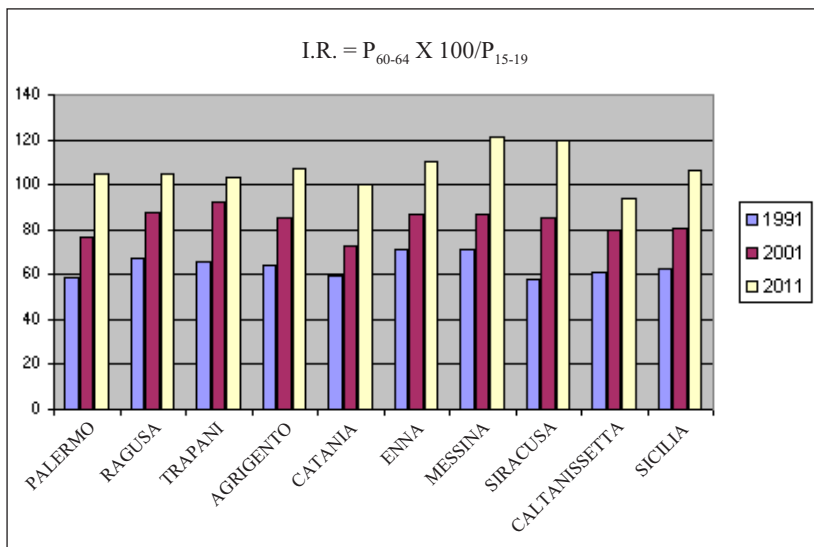
Il valore medio è di 100,54; le province in cui il carico sociale della componente non produttiva si accentuerà, sono Enna con l'indice di ricambio pari a 110,35, Messina con 121,48 e Siracusa con 119,50; Caltanissetta registrerà un indice leggermente al disotto di 100 (Tab. 8 e Graf. 4).

Tab. 8 - Indici di struttura e di ricambio della popolazione siciliana. Previsioni al 2011.

Province	I.S.	I.R.
Palermo	108,58	104,92
Ragusa	107,30	104,89
Trapani	105,71	103,49
Agrigento	111,48	107,16
Catania	106,21	100,19
Enna	112,20	110,35
Messina	116,61	121,48
Siracusa	110,84	119,50
Caltanissetta	107,38	93,68
SICILIA	107,36	106,54



Graf. 4 - Indici di struttura della popolazione siciliana in età lavorativa. Censimenti 1991-2001 e previsioni al 2011.



Graf. 5 - Indici di ricambio della popolazione siciliana in età lavorativa. Censimenti 1991-2001 e previsioni al 2011.

7. Conclusioni

Dall'analisi delle attuali tendenze demografiche è emerso in definitiva, una lieve riduzione della popolazione in età lavorativa ed all'interno di essa un invecchiamento che tenderà ad assumere dimensioni sempre più notevoli, in particolare per la riduzione dei livelli di natalità che produrrà differenze tra il flusso di entrata (popolazione fra 15-19 e 20-24) ed il flusso di uscita (popolazione fra 15-59 e 60-64 anni).

La progressiva riduzione della popolazione giovanile e quindi degli ingressi nel mercato del lavoro all'interno di una situazione economica poco favorevole allo sviluppo, quale è quella della Sicilia, crea una discrasia tra lavoro offerto e domandato, in quanto le nuove leve avranno un livello d'istruzione sempre più elevato ed aspirazione di vita che non trovano corrispondenza nella domanda di lavoro.

In assenza di adeguati interventi di politica economica volti a favorire uno stabile sviluppo economico, la popolazione giovanile andrà ad aumentare i flussi emigratori, mentre si farà sempre più forte la pressione di offerta di lavoro da parte dei Paesi del bacino del Mediterraneo.

Finito di stampare
dalla Officine Grafiche Riunite
Palermo, Giugno 2007

annali

AREA **STATISTICO-MATEMATICA**