

CT 453537

PER UNA STORIA DELLA POPOLAZIONE ITALIANA: PROBLEMI DI METODO

Contributi di: D. Argelli, S. Bertino, M. Breschi, O. Bussini, A. Coppa,
L. Del Panta, G. De Santis, M.E. Forini, D. Grassi, P. Guerresi, A. Lubisco,
C. Martelli, G.E. Montanari, D. Pettener, L. Pozzi, A. Rosina, F. Rossi,
R. Rettaroli, E. Sonnino, F. Martuzzi Veronesi

*Italia - Popolazione - Sec. XVIII - XIX - Ricerche Complessive - Porretta
Terme - 1993*

300.600.007



Questo volume raccoglie le comunicazioni presentate al seminario di Porretta Terme (3-4 luglio 1993) su «Metodi aggregativi in demografia storica» organizzato a cura di Marco Breschi e Gustavo De Santis. Il Seminario si è svolto nell'ambito del progetto di ricerca «Modelli di sviluppo demografico in Italia tra XVIII e XIX secolo» coordinato da Lorenzo Del Panta e finanziato con un contributo 40% del MURST, che ha permesso la stampa del volume. Nel secondo anno della ricerca, il gruppo era costituito da 7 Unità operative, delle seguenti sedi universitarie (tra parentesi i nomi dei responsabili locali): Udine (M. Breschi), Padova (F. Rossi), Bologna (L. Del Panta), Firenze (M. Livi Bacci), Perugia (O. Bussini), Roma (E. Sonnino), Messina (G. De Santis). Partecipano a titolo personale alla ricerca anche C. Grandi (Trento), C. Ge Rondi (Pavia) e G. Da Molin (Bari).

IBL. DIP. DI SCIENZE STATISTICHE

Statistica
Q 3



14992

UNIVERSITÀ DEGLI STUDI DI BOLOGNA

Dipartimento di Scienze Statistiche
"Paolo Fortunati"

UNIVERSITÀ DEGLI STUDI DI BOLOGNA
1993

INDICE

M. Breschi - L. Del Panta - G. De Santis, <i>Linee di ricerca per una storia della popolazione italiana: dai problemi ai metodi</i>	5
M. Breschi - A. Coppa - A. Lubisco - C. Martelli - L. Pozzi - R. Rettaroli, <i>Obiettivo BAGDAD</i>	17
M. Breschi - A. Coppa - A. Lubisco - C. Martelli - L. Pozzi - R. Rettaroli, <i>BIDDES: nascita di una banca informativa</i>	29
G. E. Montanari - O. Bussini, <i>Considerazioni sul campionamento in demografia storica</i>	37
F. Rossi, <i>La qualità dei dati nei documenti storici: una correzione dell'attrazione nella distribuzione delle età</i>	57
A. Rosina, <i>Una generalizzazione dell'Inverse projection</i>	73
G. De Santis, <i>Una modellizzazione degli effetti della mortalità sulle stime di fecondità con il metodo dei figli propri</i>	81
E. Sonnino, <i>Studio sulle caratteristiche di una popolazione, con riferimento ad un modello stabile in regime di nuzialità costante</i>	97
L. Del Panta - D. Argelli, <i>Modelli di mortalità in demografia storica: possibilità offerte dalle statistiche toscane pre-unitarie</i>	117

Finito di stampare nel mese di Dicembre 1993
presso le Officine Grafiche TECNOPRINT S.N.C.
Via del Legatore 3, Bologna.

L. Del Panta - M. E. Forini, <i>Disponibilità alimentari e mortalità in Italia tra l'Unità e la Grande Guerra: problemi di analisi di serie temporali</i>	134
P. Gueresi - F. Martuzzi Veronesi - D. Pettener, <i>Problemi di metodo nelle indagini biodemografiche su popolazioni di ambiente montano</i>	147
S. Bertino - D. Grassi - E. Sonnino, <i>Un modello di macrosimulazione del ciclo di vita di una coorte matrimoniale</i>	161

MARCO BRESCHI - LORENZO DEL PANTA - GUSTAVO DE SANTIS

**LINEE DI RICERCA
PER UNA STORIA DELLA POPOLAZIONE ITALIANA:
DAI PROBLEMI AI METODI**

A partire dai primi decenni del XVIII secolo, le popolazioni dell'Europa occidentale entrano in una nuova fase evolutiva, caratterizzata da una crescita ormai continua, non più interrotta dalle frequenti catastrofi demografiche che avevano caratterizzato le epoche precedenti. La storia demografica dei diversi paesi è però segnata, nel lungo periodo che precede la caduta irreversibile della fecondità e della mortalità (a partire, nella maggioranza dei casi, dalla seconda metà del secolo XIX), da una grande varietà di esperienze. I meccanismi demografici alla base dell'impressionante aumento della popolazione inglese a partire dalla metà del Settecento sono ben conosciuti, grazie alla ricostruzione effettuata da Wrigley e Schofield (1981). Al modello inglese, che incarna una strategia demografica dove il binomio nuzialità-fecondità ha un ruolo dominante, si contrappone il caso svedese. Qui, la ricca documentazione disponibile ha evidenziato un modello di crescita dove la diminuzione della mortalità ha giocato un ruolo preminente. La Francia rappresenta infine una sorta di "compromesso" tra i casi estremi dell'Inghilterra e della Svezia, avendo fatto registrare fin dalla fine del XVIII secolo un lento e graduale declino sia della mortalità che della fecondità.

Per l'Italia, i meccanismi demografici alla base della crescita, moderata ma continua, del XVIII e del XIX secolo sono ancora largamente ignoti, in quanto sono a tutt'oggi assai poche le indagini

che consentono di quantificare il "peso" della fecondità e della mortalità nel generale processo di sviluppo. La letteratura su questo tema è assai limitata, e - con poche eccezioni - neppure molto recente. I lavori più importanti sono ancora quelli di C. M. Cipolla (1965) e di A. Bellettini (1973) che, con accenti diversi, hanno collegato la crescita della popolazione italiana alla diminuzione della mortalità, dando però una diversa spiegazione delle cause che hanno determinato tale diminuzione: alla spiegazione "epidemiologica" del primo autore citato si contrappone infatti l'interpretazione in chiave "malthusiana" del secondo.

Una recente indagine sulla Toscana (Breschi 1990) ha messo in discussione queste interpretazioni. Il caso della Toscana - che certamente non può essere esteso, allo stato attuale delle conoscenze, all'intera popolazione italiana - sembra infatti costituire, per alcuni aspetti, un modello demografico non dissimile da quello inglese, nel senso che la fecondità risulta aver giocato un ruolo determinante. La varietà dei contesti economici ed ambientali presenti in Italia, così come le numerose, se pur sommarie, indicazioni sui differenti regimi demografici che sembrano aver caratterizzato, nel periodo in questione, vaste aree regionali e sub-regionali, consigliano in ogni caso di orientare le indagini - sulla scia di quanto è stato fatto per la Toscana - verso lo studio degli specifici meccanismi di riproduzione demografica che hanno caratterizzato i diversi territori, piuttosto che ricercare, fin d'ora, un ipotetico "modello demografico italiano", da contrapporre o da paragonare a quelli di altre nazioni europee.

A partire dalla metà del XVII secolo, le fonti censuarie esistenti permettono di mettere in evidenza con sufficiente approssimazione, per gran parte degli antichi stati italiani, la tendenza demografica complessiva. Sulla base di questa documentazione, è possibile individuare modalità di sviluppo demografico diversificate per le grandi ripartizioni in cui viene usualmente suddivisa la penisola. Sappiamo che, dopo una fase di forti differenziazioni (seconda metà del XVII secolo), determinate essenzialmente dalla diversa incidenza

temporale della peste nelle varie zone, la crescita si stabilizza, ed interessa tutte le aree geografiche. Nel corso del Settecento, l'espansione demografica maggiore risulta però quella dell'Italia meridionale ed insulare, mentre i territori settentrionali e quelli centrali hanno andamenti tra loro opposti: la crescita rallenta nel Nord nel secondo cinquantennio, e subisce invece un'accelerazione nelle regioni dell'Italia centrale. Nella prima metà del secolo successivo, il Mezzogiorno continentale risulta ancora l'area a più rapido sviluppo demografico, anche se le differenze tra le varie circoscrizioni sono abbastanza attenuate. Nella seconda parte del XIX secolo, infine, il divario nei tempi di inizio del declino della natalità e della mortalità, e soprattutto nei tempi e nell'entità della grande emigrazione portano a nuove differenziazioni nei ritmi di crescita, delle quali i dati per grandi ripartizioni non possono dar conto in modo distinto.

Più difficile risulta, a tutt'oggi, per la popolazione di vaste aree regionali o sub-regionali, la determinazione dei *meccanismi demografici* che sottostanno alle differenze di crescita complessiva. Senza contare, poi, che anche popolazioni con analoghi tassi di crescita possono essere caratterizzate da regimi demografici profondamente diversi. Le statistiche italiane consentono, in alcuni casi, di individuare le componenti naturali (natalità e mortalità) dell'evoluzione demografica degli antichi stati, o di aree corrispondenti ad attuali regioni o province, già per la seconda metà del XVIII, e più spesso a partire dalla prima metà del XIX secolo. Si può comunque presumere, sulla base di numerosi indizi, che le specificità nel regime demografico che si riscontrano in tali periodi rappresentino, di norma, delle caratteristiche strutturali, di lunga durata, valide anche per epoche precedenti.

Indubbiamente i tassi generici - i soli che è talvolta possibile conoscere, in epoca preunitaria, su scala regionale - spiegano assai poco dei meccanismi demografici sottostanti, ed ancor meno dei fattori che condizionano gli stessi meccanismi demografici. La strada

per arrivare ad accrescere le nostre conoscenze al riguardo passa, quindi, attraverso analisi più dettagliate su fonti di prima mano, relative ad aree più ristrette.

In linea di massima, si può affermare - con riferimento alle popolazioni rurali, di gran lunga maggioritarie in Italia fino alla fine del secolo XIX - che le aree nelle quali si ritrova un regime demografico *a rapido ricambio* (con alta natalità e alta mortalità) sono, il più delle volte, caratterizzate da strutture agrarie basate sul latifondo e sul bracciantato e, spesso, da pressioni ambientali sfavorevoli. Le aree per le quali si riscontra un regime demografico *a più lento ricambio* (con natalità e mortalità più moderate) sono invece solitamente accomunate, pur nella diversità di contesti ambientali, da uno sfruttamento più intensivo del territorio e da un rapporto più stabile dei lavoratori con la terra (mezzadria o piccola proprietà). Esempio, a questo proposito, rimane lo studio di G. Delille (1977), dove è messo in chiara evidenza il diverso regime demografico che - tra Sette e Ottocento - opponeva, all'interno del Regno di Napoli, aree come quelle costiere della Campania, contraddistinte dalla policoltura, dalla piccola proprietà contadina e da un'alta densità demografica, a territori con le caratteristiche del Tavoliere delle Puglie, con una bassa densità demografica, e dove predominavano la monocultura cerealicola e la proprietà latifondista. A tale varietà di situazioni economico-ambientali corrispondeva dunque - per quanto a tutt'oggi è stato possibile evidenziare - un'accentuata variabilità di regimi demografici che, globalmente, hanno comunque determinato, in Italia, una crescita abbastanza modesta tra la metà del XVII e la fine del XVIII secolo, ed una discreta accelerazione (inferiore peraltro alla media europea) a partire dagli anni '20 dell'Ottocento.

Se, dunque, sulla base dei dati a tutt'oggi disponibili, risulta abbastanza chiara l'evoluzione nel lungo periodo della popolazione italiana, molti lati oscuri permangono circa i suoi meccanismi riproduttivi. L'interrogativo che si pongono oggi, in Italia, molti cultori di demografia storica è, in sostanza, quello di quale sia la

strada migliore da seguire per arrivare, in un futuro non troppo lontano, ad una "spiegazione" esauriente di tali meccanismi. L'idea che sta alla base del progetto di ricerca¹ del quale questo volume propone i primi risultati di carattere metodologico è, in effetti, che sia preferibile, al momento attuale, approfondire la conoscenza dei meccanismi evolutivi di specifiche aree regionali o sub-regionali (differentemente caratterizzate sotto il profilo economico ed ambientale, e probabilmente anche in termini demografici), per arrivare, in una fase successiva, ad un tentativo di sintesi a livello nazionale. Le indagini che i vari gruppi di ricerca stanno portando avanti dovrebbero insomma costituire un primo passo in vista di una futura ricostruzione delle vicende demografiche del nostro paese, che potrà scaturire, secondo questa impostazione, da uno sforzo di comparazione e di sintesi di specifiche realtà territoriali.

In ogni caso, lo stato della ricerca storico-demografica in Italia non è tale da suggerire, al momento attuale, un tentativo di sintesi. Un tale tentativo, compiuto sulla base dei risultati a tutt'oggi disponibili, sarebbe probabilmente poco produttivo, nel senso che non porterebbe a decisivi passi avanti nella comprensione del *funzionamento del sistema demografico* italiano. Occorre, piuttosto, impostare una serie di nuove indagini, che abbiano però come obiettivo finale l'individuazione di modelli di sviluppo demografico, tali da poter essere messi a confronto tra di loro e con quelli che già sono stati delineati per altri paesi o regioni dell'Europa (in particolare la

¹ Il progetto, finanziato dal Ministero dell'Università e della Ricerca Scientifica e Tecnologica, è coordinato da L. Del Pantà e ad esso aderiscono attualmente (nel secondo anno di attuazione) ricercatori appartenenti alle seguenti sedi universitarie (tra parentesi è indicato il nome del responsabile dell'Unità operativa locale): Udine (Breschi), Padova (Rossi), Bologna (Del Pantà), Firenze (Livi Bacci), Perugia (Bussini), Roma (Sonnino), Messina (De Santis). A titolo personale collaborano al progetto anche C. Grandi (Trento), C. Ge Rondi (Pavia) e G. Da Molin (Bari).

Francia, l'Inghilterra, la Svezia).

Anche se è, dunque, nostra convinzione che questo progetto debba essere in qualche misura concepito come "somma" di progetti "regionali", questi debbono essere però organizzati secondo una metodologia il più possibile uniforme, pur se con qualche margine di elasticità, per tener conto delle specificità locali, sia in termini di strutture economiche e demografiche che di disponibilità di fonti statistiche.

Questo volume, che raccoglie le comunicazioni presentate ad un seminario, su temi di metodo inerenti alla ricerca in oggetto, tenuto a Porretta (Bologna) nei giorni 3-4 luglio del 1993, vuole fornire un primo contributo rivolto a favorire le possibilità di comparazione, ed eventualmente di sintesi, dei risultati che i vari gruppi di lavoro² vanno via via ottenendo. Non vi è dubbio, infatti, che esistono a tutt'oggi, in Italia, una massa ingente di dati demografici già raccolti ed una mole assai cospicua di lavoro già svolto, che è però difficile utilizzare in termini comparativi (ancor più difficili appaiono i tentativi di sintesi), in ragione del fatto che le procedure di raccolta e di vaglio critico dei dati e le metodologie di elaborazione e di produzione dei risultati non vengono solitamente rese note e sono comunque disomogenee.

Il nostro gruppo di ricerca³ ha dunque avviato una riflessione sugli aspetti di metodo delle indagini storico-demografiche, con particolare riguardo alle problematiche connesse con l'analisi e con i tentativi di spiegazione dei meccanismi demografici di vaste aree

² E' nostro auspicio che anche altri ricercatori, impegnati come noi nello studio dei meccanismi demografici delle popolazioni di specifiche aree del nostro paese nei secoli passati, possano recepire ed eventualmente discutere ed integrare le indicazioni e le proposte che il nostro gruppo di ricerca va elaborando.

³ Per gli aspetti di metodo, il progetto è articolato in sottoprogetti che hanno favorito l'aggregazione di gruppi di lavoro "intersedi".

territoriali. In tale ambito è importante sottolineare, tra gli obiettivi prioritari, quello della predisposizione di una *metodologia standardizzata* di raccolta, archiviazione (e successiva utilizzazione) di dati demografici di tipo aggregativo.

In questa ottica si devono leggere i due contributi del gruppo di lavoro intersede su "Modalità e problemi di rilevazione dei dati per la costituzione di una banca-dati di tipo aggregativo". La prima comunicazione illustra le caratteristiche e il modello informatico di BIDDES (Banca Informatica sulla Documentazione in Demografia Storica), una banca informativa (e, pertanto, una meta banca) sulla documentazione raccolta relativamente a dati di popolazione e di movimento. BIDDES, una volta attivata, potrà rispondere, mediante apposite funzioni di interrogazione, a domande volte a conoscere le località per le quali è stata effettuata una rilevazione, il periodo e l'oggetto della rilevazione (popolazione, battesimi, nascite, sepolture, etc.), le modalità di spoglio adottate (nominativo, anonimo individuale, ecc.), le variabili considerate nello spoglio (sesso, età, ecc.). La seconda comunicazione presenta le linee essenziali del progetto BAGDAD (Banca AGgregativa di DATi Demografici), il cui obiettivo è la predisposizione di una banca automatizzata e standardizzata per la memorizzazione e la gestione di dati di movimento rilevati in forma anonima individuale o aggregativa. BAGDAD, offrendo una serie di metodologie standard, vuole favorire l'adozione di criteri uniformi nella fase di predisposizione e di controllo della base dei dati (codifica, memorizzazione, controllo, ecc.).

Se i due precedenti contributi toccano problemi inerenti, in diverso modo, alla memorizzazione e gestione delle informazioni, la comunicazione di Montanari e di Bussini affronta un tema centrale della ricerca demografica - in particolare, di quella macrodemografica, che precede (o dovrebbe sempre precedere) le fasi della rilevazione e della memorizzazione dei dati. Per il passato, la ricostruzione delle vicende di grandi aggregati demografici (nazioni e regioni) deve basarsi su "porzioni" (spesso ridotte) dell'intera popolazione, sia per

l'indisponibilità totale (mancanza di registri all'origine o per successiva perdita) o parziale (lacune nelle registrazioni, ecc.) della documentazione di base sia per la mole di lavoro necessaria ad una rilevazione estensiva. Si comprende allora che il campionamento, cioè l'osservazione di un numero limitato di parrocchie, è una scelta obbligata per stimare i dati di movimento complessivi. Risulta così fondamentale "sapere" scegliere le parrocchie da inserire nel campione e, soprattutto, misurare la precisione e l'affidabilità delle stime. Proprio questo cruciale aspetto è l'argomento della comunicazione di Montanari e Bussini.

Connesse alle precedenti problematiche sono quelle relative alla qualità dei dati rilevati. Il "vaglio" del dato è uno dei momenti, se non il momento, centrale in demografia storica. La comunicazione di Rossi affronta un classico tema della qualità dei dati nei documenti storici: il fenomeno dell'attrazione nelle distribuzioni per età. In particolare, Rossi propone un metodo di correzione delle distribuzioni meno "meccanico" di quelli propri a tutte le procedure di "lisciamiento".

All'analisi dei dati sono dedicati tre contributi che propongono interessanti spunti in riferimento a metodi diversi, tutti caratterizzati da procedure di stima indirette.

Il lavoro di Rosina illustra un adattamento della tecnica dell'inverse projection per sfruttare convenientemente una caratteristica della documentazione di base esistente in Italia. Le registrazioni parrocchiali delle sepolture riportano infatti con frequenza l'età al decesso del defunto: la versione dell'inverse projection sviluppata da Rosina introduce un semplice accorgimento per tenere conto di questa informazione aggiuntiva. La ricostruzione della popolazione ottenibile con questa nuova versione dell'inverse projection risulta più realistica: l'informazione sulla distribuzione dei decessi per età consente infatti di tenere conto non solo delle variazioni di livello ma, anche, di forma della mortalità.

Anche De Santis, nelle sue riflessioni sul metodo dei figli propri, considera alcune problematiche connesse all'azione della mortalità

soffermandosi, in particolare, sull'effetto prodotto, sulle stime della fecondità, da una vasta gamma di variazioni nella mortalità (modello, livello, evoluzione temporale). Le simulazioni condotte dimostrano che le stime della fecondità ottenute con il metodo dei figli propri sono relativamente robuste rispetto alle ipotesi di mortalità adottate.

All'interno della teoria della popolazione stabile si muove il contributo di Sonnino che avanza un nuovo modo di ingresso basato sulle informazioni relative ai matrimoni e, in senso più ampio alla nuzialità. Questa variante può risultare utile, in particolare, in tutte le situazioni caratterizzate da informazioni limitate su natalità e mortalità.

La mortalità è l'oggetto specifico di due lavori condotti da Del Panta in collaborazione, distintamente, con D. Argelli e E. Forini. Il primo di questi è un tentativo di individuare strutture tipiche di mortalità - differenziate sul piano geografico ed ambientale - all'interno del territorio toscano nella prima metà dell'Ottocento. Le applicazioni condotte a partire dalle sole statistiche dei decessi per età hanno consentito di saggiare la "robustezza" del metodo dei decessi generalizzato al variare delle ipotesi e delle diverse informazioni aggiuntive utilizzate per la valutazione del tasso d'incremento. Nell'altro lavoro, Del Panta e Forini utilizzano invece alcuni modelli a ritardi distribuiti per verificare l'esistenza di una relazione tra le variabili alimentari e quelle di mortalità.

Gli ultimi due contributi richiamano due diversi approcci nella ricerca demografica che, soltanto a prima vista, possono sembrare "distanti" dall'approccio macrodemografico proposto in questo progetto di ricostruzione della storia della popolazione italiana.

Il contributo di Guerresi, Martuzzi Veronesi e Pettener si ricollega, da un lato, all'approccio microdemografico e, dall'altro, alla componente biologica negli studi della popolazione. Privilegiare l'approccio aggregativo e macrodemografico non significa disconoscere la rilevanza di quello microdemografico: un modo di "studiare" la popolazione insostituibile per la comprensione sia dei meccanismi

demografici più minuti sia delle relazioni fra variabili demografiche, economiche, culturali e, non ultime, bio-genetiche. Una ricostruzione della storia della popolazione italiana, contrassegnata, ad un tempo, da continui travasi di popolazioni e dal permanere di impermeabili barriere, non potrà infatti dirsi compiuta senza un'adeguata lettura di questi aspetti peculiari.

Il lavoro di Bertino, Grassi e Sonnino "appartiene" infine al campo della simulazione, poco percorso dai demografi italiani e, in particolare, in ambito storico. Un approccio che offre interessanti spunti per la comprensione dell'interazione esistente tra le variabili considerate nel modello, per la verifica dei risultati originati da ricerche dirette e, infine, per la definizione di ipotesi e stime da assumere nella ricerca. In particolare, il modello di macrosimulazione illustrato nella comunicazione consente di prospettare un ventaglio di possibili "storie" del ciclo riproduttivo di una coorte matrimoniale in funzione del diverso agire di vari fattori demografici.

Bibliografia

A. Bellettini (1973), *La popolazione italiana dall'inizio dell'era volgare ai giorni nostri. Valutazioni e tendenze*, in *Storia d'Italia*, V (1), Einaudi, Torino.

M. Breschi (1990), *La popolazione della Toscana dal 1640 al 1940. Un'ipotesi di ricostruzione*, Dipartimento Statistico, Firenze.

C. M. Cipolla (1965), *Four Centuries of Italian Demographic Development*, in D. V. Glass - D. E. C. Eversley (editors), *Population in History*, Arnold, London.

G. Delille (1977), *Agricoltura e demografia nel Regno di Napoli nei secoli XVIII e XIX*, Guida Editore, Napoli.

A. E. Wrigley - R. S. Schofield (1981), *The Population History of England, 1541-1871*, Arnold, London.

MARCO BRESCHI - ALFREDO COPPA - ALESSANDRO LUBISCO
CRISTINA MARTELLI - LUCIA POZZI - ROSELLA RETTAROLI

OBIETTIVO BAGDAD

1. Premessa

Il progetto BAGDAD - acronimo di Banca AGgregativa di DATi Demografici - ha come obiettivo la messa a punto di un sistema automatizzato per la memorizzazione e gestione di dati storico demografici rilevati con modalità aggregativa (non nominativa). La costruzione di BAGDAD si inserisce coerentemente nel quadro della ricerca MURST 40% "Modelli di sviluppo demografico in Italia tra 18° e 19° secolo"¹. L'obiettivo primario di tale ricerca è, infatti, quello di approfondire i meccanismi di riproduzione demografica di ampi aggregati di popolazione con un approccio prevalentemente macrodemografico (aggregativo). In questo campo, sono oggi disponibili tecniche sofisticate di ricostruzione dell'ammontare e della struttura della popolazione, e di stima dell'andamento nel tempo del livello dei principali parametri demografici (speranza di vita, misure di riproduttività). Queste tecniche sono fondate sull'impiego di modelli di mortalità e di fecondità, e basate principalmente sull'utilizzazione di serie storiche di eventi demografici. La tecnica più idonea, in relazione alle fonti italiane, è quella ideata da R. Lee e conosciuta come "inverse projection".

¹ Il gruppo nazionale di ricerca riunisce numerose sedi universitarie italiane e, più precisamente: Bologna, Firenze, Roma, Perugia, Bari, Padova, Udine e Messina.

All'interno della ricerca MURST 40%, BAGDAD vuole offrire un sistema standardizzato ed automatizzato per la memorizzazione e gestione delle serie storiche di eventi demografici (nascite, decessi e matrimoni). Tali serie costituiscono, assieme ad informazioni sul totale della popolazione, la base primaria di dati per ricostruire la popolazione italiana in epoca storica mediante procedure aggregative.

Per conseguire un simile risultato sarà necessario raccogliere e riorganizzare un vasto materiale di base, fissare delle norme comuni per la rilevazione e la memorizzazione, stabilire una serie di test standard per verificare la completezza e la qualità dei dati, adottare un sistema efficiente per la gestione di centinaia di serie storiche, ecc. Tutte esigenze che difficilmente potranno trovare adeguate soluzioni in risposte occasionali ed estemporanee.

L'obiettivo principale di BAGDAD è offrire uno strumento per superare uno dei principali limiti della ricerca storico demografica italiana: l'assenza di standard comuni nella rilevazione e nel trattamento dei dati, nonché nel calcolo anche dei più semplici indici demografici, con la grave conseguenza di rendere difficile, spesso impossibile, il confronto fra i risultati osservati in diverse realtà².

² Gli esempi a questo proposito sono numerosi fra coloro che, di volta in volta, hanno cercato di offrire dei quadri nazionali su alcuni aspetti demografici in epoca preunitaria. Fra i lavori più recenti, si vedano le osservazioni critiche di M. Livi Bacci e M. Breschi (1990), *Italian Fertility: An Historical Account*, *Journal of Family History*, n. 15, pp. 385-408, sulle misure di fecondità tratte dalla trentina di ricostruzioni nominative delle famiglie effettuate in Italia: esse si riferiscono a famiglie di tipo diverso (famiglie complete; famiglie incomplete, ecc.) e spesso il tipo di famiglia non è indicato; le famiglie con zero figli sono talvolta escluse; le età medie sono calcolate con criteri diversi. Anche L. Del Panta (1992) nel sua recente relazione su *Infant and Child Mortality in Italy from XVIIIth to XXth Century: Long-Term Trend and Territorial Differencies*, presentata al convegno IUSSP "Seminar on Child and Infant Mortality in the Past -

BAGDAD rappresenta di fatto un'occasione per la diffusione di criteri uniformi nella fase di predisposizione della base dei dati (rilevazione, codifica, memorizzazione, controllo, ecc.).

A questa breve premessa seguono due paragrafi. Il primo sintetizza le caratteristiche delle banche aggregative di dati storico-demografici esistenti; l'ultimo illustra l'architettura di BAGDAD.

2. Le esperienze precedenti

Soltanto negli ultimi anni, l'approccio aggregativo e macrodemografico ha catalizzato l'interesse crescente di numerosi cultori della demografia storica. A questo interesse non corrisponde ancora un adeguato sviluppo del software sia nel settore della creazione di banche di dati demografici non nominativi, sia nel campo della messa a punto di pacchetti informatici relativi a tecniche di analisi demografica di tipo aggregativo³. Proprio in questi ultimi mesi, è stata annunciata l'uscita di alcuni nuovi programmi applicativi nel settore delle tecniche di proiezione inversa e della misura indiretta della fecondità (in particolare, metodo dei figli

Montreal, 7-9 October 1992", esprime non poche perplessità sulla comparabilità dei numerosi dati sulla mortalità infantile e giovanile reperibili per diverse località italiane. In questo ultimo caso, si tratta di un indicatore tutto sommato di "facile" calcolo, ma le incertezze sono comunque tante e tali da limitare di fatto il confronto. Spesso non è chiaro come sono stati considerati i nati morti; non di rado si confonde fra tasso e probabilità, ecc.

³ Su questo tema cfr. M. Breschi e L. Pozzi (1992), *Demografia informatica: tra vecchio e nuovo*, relazione presentata al "Settimo Congresso Internazionale dell'Associazione for History & Computing, Bologna, agosto - 2 settembre 1992", i cui atti sono in corso di pubblicazione.



propri)⁴. I pacchetti informatici disponibili si configurano però come prodotti sperimentali: la ricerca nel campo delle metodologie aggregate per lo studio delle popolazioni del passato è del resto in una fase di vivace crescita e soggetta ad interessanti avanzamenti⁵.

Nel settore del software per le banche aggregative di dati storico-demografici il ritardo è ancora più marcato, in particolare rispetto a quanto è ormai disponibile per le banche di tipo nominativo. I motivi di questo ritardo sono facilmente individuabili. In larga parte, sono da ricondursi all'interno della stessa disciplina storico demografica: l'interesse nei confronti di un approccio aggregativo nella ricerca è ancora troppo recente e, da questo, ne consegue un numero ancora limitato di studi concreti.

Il lavoro pionieristico di Wrigley e Schofield sulla popolazione inglese (1981)⁶, vero e proprio volano al nuovo interesse nei confronti degli studi macrodemografici, è la ricostruzione che ha utilizzato il numero più elevato di serie parrocchiali (oltre 500) per stimare i dati nazionali di movimento. Le successive applicazioni delle tecniche di proiezione inversa sono state condotte con serie di nati, decessi e

⁴ Robert Mc Caa sta terminando lo sviluppo di una nuova versione di "Populate" con un algoritmo dell'inverse projection con passo annuale anziché quinquennale; copie sperimentale di GIP (Generalized Inverse Projection) sono fornite da Jim Oeppen; A. Rosina e F. Rossi hanno messo a punto una nuova versione dell'inverse projection che tiene conto di eventuali informazioni sulla distribuzione dei decessi per età. De Santis e Breschi hanno annunciato l'uscita di APPLAUSI (APPLICazione AUTomatizzata di un Metodo di Sopravvivenza Inversa) per l'applicazione del metodo dei figli propri in demografia storica.

⁵ In questo senso si possono consultare gli atti del convegno di Palma de Mallorca (12-15 giugno 1991) promosso dalla locale Università e dallo IUSSP su "Old and New Methods in Historical Demography".

⁶ A. Wrigley and R. Schofield (1981), *The Population History of England, 1541-1871: A Reconstruction*, London.

matrimoni già esistenti a livello nazionale (è questo il caso delle applicazioni per la Svezia e la Norvegia effettuate per testare la tecnica e non per fini investigativi), oppure partendo da poche decine di serie (come nel caso della Toscana, della provincia di Valencia e della Castiglia) o, addirittura, da un'unica serie di movimento nei vari studi fatti su singole città o comunità (Lucerna, Amsterdam, Berna, ecc.).

La necessità di consolidare ed approfondire le innovazioni a livello metodologico, combinandosi ad esigenze applicative ancora limitate nel numero e nella dimensione della base dei dati, hanno finora relegato in secondo piano il problema della memorizzazione e gestione di dati di tipo aggregativo. Un problema invece prioritario per il conseguimento di validi risultati come, del resto, indica la recente costituzione di un gruppo di "Coordinamento delle inchieste nazionali di ricostruzione" all'interno della Commissione di Demografia Storica dell'IUSSP.

Inoltre, la specificità dell'oggetto - la creazione di una banca aggregativa di dati storico-demografici - non consente un immediato e facile adattamento dell'architettura di banche dati nate per rispondere ad altre esigenze. Innanzitutto, la maggior parte delle banche gestiscono dati di tipo nominativo e questo anche nel più ristretto campo demografico: si pensi, ad esempio, ai numerosi pacchetti per la complessa gestione delle anagrafi e degli Uffici di stato civile. Lo sviluppo di banche dati nominative in campo storico-demografico ha quindi potuto avvalersi (a costo zero o quasi) dell'esperienza maturata e sviluppata per rispondere ad esigenze della nostra società. A questo proposito è utile riflettere sul "nominal record linkage": un algoritmo determinante per il successo, ad esempio, di una ricostruzione nominativa delle famiglie (a parità di tutte le altre condizioni è il numero di incroci - a determinare il "successo" di uno studio nominativo) ma, anche, per l'efficiente gestione e funzionamento di molti servizi (istituti di credito, anagrafi, etc.).

In definitiva, la letteratura informatica (e, in particolare, quella informatico-demografica) è ricca e altamente specializzata nel campo delle banche di dati nominativi; ancora contenuta e ad uno stadio sperimentale in quello delle banche aggregative di dati demografici e, in particolare, per dati storici.

Negli ultimi decenni sono state effettuate due sole importanti campagne di rilevazione aggregativa di dati di movimento in epoca storica: una promossa, a partire dagli anni '60, dal Cambridge Group for the History of Population and Social Structure che ha fornito il materiale di base utilizzato da Wrigley e Schofield per la loro ricostruzione della popolazione inglese dal 1541 al 1871⁷; l'altra, coordinata dall'INED in collaborazione con l'EHESS, condotta all'interno del progetto di studio della popolazione francese in epoca moderna ancora in corso ed iniziato da oltre quaranta anni⁸.

Nessuna di queste due grandi rilevazioni ha portato alla creazione di una vera e propria banca-dati per la gestione del materiale raccolto. Ciò si può spiegare, in parte, con l'epoca di inizio delle campagne di rilevazione, gli anni sessanta-settanta, e, in parte, con le stesse caratteristiche dei piani di rilevazione e con gli obiettivi della ricerca.

In quegli anni, caratterizzati da poche, centralizzate (mainframe) e costose risorse informatiche, si stavano muovendo, per il software, i primi passi nelle metodologie di modellizzazione ed organizzazione dei dati⁹.

⁷ Per un'illustrazione dell'intera strategia di raccolta e gestione delle serie di movimento, Cfr. E.A. Wrigley and R. Schofield (1981), op. cit. e la relativa bibliografia per gli specifici riferimenti.

⁸ Fra i vari lavori si rimanda a J.N. Biraben (1985), *Le point de l'enquete sur le mouvement de la population en France avant 1670*, *Population*, n. 1, pp. 47-70.

⁹ Oggi le caratteristiche del materiale rilevato non porrebbero particolari problemi di gestione informatica, essendo sufficienti un personal computer e

La rilevazione inglese si limitò alla sola conta degli eventi (nati, matrimoni e decessi) per anno e mese, una scelta troppo riduttiva per gli stessi Wrigley e Schofield che, a posteriori, osservarono che sarebbe stato opportuno tenere almeno conto del sesso.

Il caso inglese è tuttavia istruttivo ed interessante nel campo dei metodi per il controllo della qualità dei dati e, in particolare, dell'individuazione di lacune nelle serie.

L'analisi di 530 serie parrocchiali ridotte poi a 404 e corrispondenti, comunque, a 3.688.884 di frequenze mensili di nascite, decessi e matrimoni richiese la messa a punto di procedure standardizzate per la verifica della qualità dei dati. A tale fine Wrigley e Schofield con l'aiuto dei loro collaboratori inventarono, talvolta riadattando, una serie di semplici ma efficaci procedure (ampiamente descritte nel loro volume) che possono essere utilmente inserite all'interno di una banca aggregativa di dati demografici.

Il caso francese è invece istruttivo per altri aspetti. La rilevazione di tipo aggregativo si inserisce all'interno di un progetto di ricerca di impianto sostanzialmente nominativo (basato sull'applicazione della tecnica di ricostruzione nominativa della famiglia ad alcune centinaia di villaggi) e, pertanto, i demografi francesi hanno approfondito alcuni problemi connessi all'utilizzazione incrociata fra dati raccolti in forma nominativa, individuale anonima e aggregativa. I manuali di demografia storica della scuola francese dedicano infatti ampio spazio a questa particolare classe di problematiche: scelta del campione; controllo incrociato delle informazioni; ecc.

3. L'architettura di BAGDAD

BAGDAD è stata concepita come una banca per la gestione di dati di movimento (nati, decessi e matrimoni) rilevati in forma non

un foglio elettronico o un database.

nominativa e, precisamente, in forma:

- anonima individuale
- aggregata (serie storiche).

La struttura essenziale di BAGDAD è data dalla combinazione di quattro moduli interconnessi ed abilitati alla soluzione delle seguenti classi di problemi:

- A) catalogazione e descrizione della fonte;
- B) memorizzazione dei dati;
- C) controllo e verifica dei dati;
- D) gestione dei dati.

A) Catalogazione e descrizione della fonte. BAGDAD offre una serie di "schede" per la descrizione della fonte (libri parrocchiali, registri di stato civile, ecc.). La descrizione è articolata in sezioni. La prima, assimilabile alla "carta d'identità" della fonte, contiene le informazioni di base del documento: collocazione archivistica, titolo, aggregato demografico considerato, anni di copertura e poco altro. La seconda sezione serve alla descrizione esteriore del documento: dimensioni, stato di conservazione, tipo di rilegatura, lingua usata, grado di leggibilità, ecc. Infine la terza sezione è organizzata in forma tabellare con la specifica delle informazioni usualmente riportate nei vari atti e con caratteristiche diverse a seconda del tipo di rilevazione (battesimi, defunti, matrimoni).

B) Memorizzazione dei dati. Una delle maggiori difficoltà nella costruzione di BAGDAD è garantire al ricercatore una sufficiente flessibilità nella fase di predisposizione di adeguati supporti per la memorizzazione dei dati in funzione del piano di spoglio adottato, delle specifiche caratteristiche del materiale documentario utilizzato ed, anche, degli obiettivi della ricerca. In un certo senso, lo spoglio nominativo completo pone minori problemi in quanto la selezione ed organizzazione del materiale documentario può essere rinviata ad una fase successiva a quella della rilevazione e memorizzazione e, inoltre, il ricercatore potrà sempre riorganizzare e sezionare la sua base di dati - omologa o quasi a quella contenuta nella fonte

originaria. In uno spoglio non nominativo (aggregativo o anonimo individuale), la selezione delle informazioni precede la rilevazione e, una volta effettuata, è in un certo senso irreversibile (a meno che non si proceda ad un nuovo spoglio della fonte originaria).

Ecco, quindi, l'importanza di offrire al ricercatore una sufficiente flessibilità che tenga conto delle specifiche caratteristiche della fonte in esame e degli obiettivi di ricerca, ma senza, al contempo, rinunciare ad alcuni "criteri" standard necessari per avere una base documentaria caratterizzata da un minimo livello di coerenza ed uniformità.

Il precedente problema ha una rilevanza minore nel caso di spogli aggregati ossia di semplice conta degli eventi per anno; per anno e mese; per anno, mese, sesso; ecc. Una rilevazione aggregativa risulta infatti inefficiente (in termini di tempo impiegato e di qualità della rilevazione) se la conta viene effettuata considerando contemporaneamente troppe variabili: ad esempio, nati per anno, mese, sesso, vitalità e legittimità. Esiste infatti una soglia "critica" oltre la quale conviene procedere ad uno spoglio anonimo individuale: nel caso del precedente esempio, il maggior tempo impiegato per uno spoglio anonimo individuale (circa un terzo di tempo in più rispetto ad uno spoglio aggregato) garantisce un minore numero di errori in fase di spoglio e, inoltre, la base dati costruita permette controlli più efficienti e consente un'ampia flessibilità nella fase di trattamento ed elaborazione.

In pratica, lo spoglio di tipo aggregato deve essere condotto soltanto con un numero limitato di variabili (tre, al massimo quattro) e, pertanto, è relativamente facile predisporre degli schemi "standard" per la memorizzazione, in particolare se l'anno e il mese di accadimento dell'evento sono considerati sempre due delle tre o quattro variabili da rilevare. Una scelta dettata non solo dalla rilevanza informativa delle due variabili temporali ma necessaria per individuare la presenza di lacune come suggerisce l'esperienza del Cambridge Group.

La costruzione di schemi "standard" per lo spoglio anonimo individuale è invece più difficile e, pertanto, è data al ricercatore la facoltà di costruirsi direttamente gli schemi per la memorizzazione dei dati rilevati. Il ricercatore dovrà selezionare, per ciascun tipo di atto (ad esempio, nascita), le informazioni (sesso, legittimità, vitalità, provenienza, ecc.) che saranno oggetto di spoglio e di ordinarle nella sequenza più rispondente alle sue esigenze.

In definitiva BAGDAD, oltre ad offrire una serie di schemi "standard" già predisposti, dà al ricercatore la possibilità di costruirsi gli schemi di memorizzazione nel modo più congeniale alle sue esigenze di spoglio. Questa libertà è in parte temperata dal fatto che il ricercatore dovrà attenersi alle norme di codifica predisposte per ciascuna variabile (sesso, legittimità, vitalità, ecc.). Norme che, comunque, possono essere parzialmente "adattate" mediante la specifica di codici particolari. L'adozione del sistema di codifica previsto permette di utilizzare le procedure di controllo per l'immissione dei dati.

C) Controllo e verifica dei dati. Questo modulo contiene le procedure di controllo dell'immissione dei dati. Esse segnalano eventuali incongruenze imputabili in gran parte ad errori di rilevazione, codifica ed immissione dei dati ma, talvolta, presenti nella stessa fonte originaria (ad. esempio, decesso di una persona di 5 anni indicata come sposata). Queste procedure di controllo operano automaticamente nella fase di immissione dei dati e, come si è detto al punto precedente, funzionano soltanto se viene adottato il sistema di codifica previsto.

Inoltre, BAGDAD dispone di una serie di procedure per la verifica della qualità dei dati. Il grado di complessità di queste procedure è funzione del tipo di spoglio (aggregato o anonimo individuale) e del numero di informazioni rilevate. Nella versione prototipo di BAGDAD si prevede di installare soltanto alcune delle molte e possibili

procedure di controllo¹⁰. Del resto, l'architettura a moduli e l'adozione di linguaggi di programmazione "trasparenti" garantiscono implementazioni successive.

D) Gestione dei dati. Questo modulo contiene alcune delle funzioni tipiche di qualsiasi sistema di DBMS (Data Base Management System). BAGDAD si pone l'obiettivo di adattare alle sue specifiche esigenze la ricca gamma di capacità e di funzioni offerte dai numerosi DBMS esistenti sul mercato.

¹⁰ A questo proposito, la letteratura specializzata, ma anche lo strumentario informatico, è assai vasto. Oltre ai già citati lavori della scuola inglese e francese, basti pensare ai manuali delle Nazioni Unite (in particolare il Manuale IV e il Manuale X) e al sempre utile volume di H. Shryock and J. Siegel (eds.), *The Methods and Materials of Demography*, U.S. Department of Commerce, Bureau of Census, Vol. 1-2, Washington.

MARCO BRESCHI - ALFREDO COPPA - ALESSANDRO LUBISCO

CRISTINA MARTELLI - LUCIA POZZI - ROSELLA RETTAROLI

**BIDDES:
NASCITA DI UNA BANCA INFORMATIVA**

Premessa

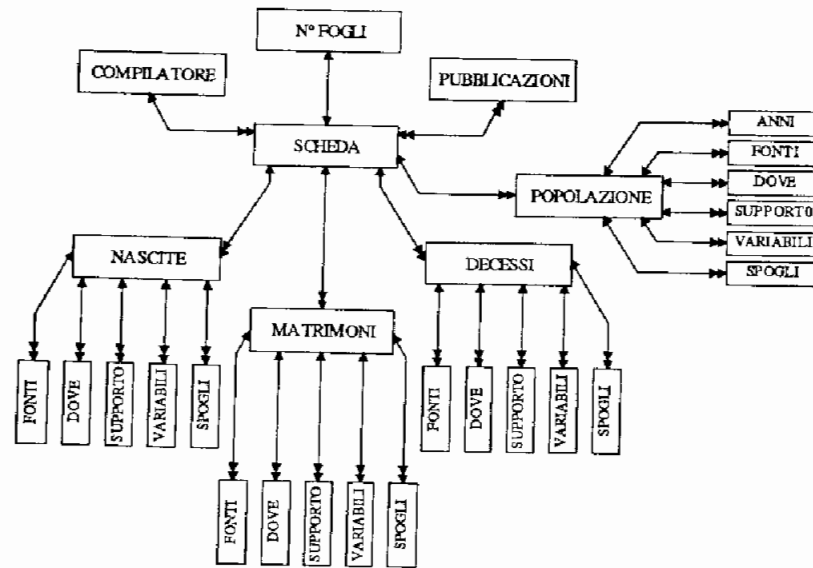
È dalla constatazione dell'esistenza di un ricchissimo patrimonio di dati storico-demografici, troppo spesso disperso presso i singoli ricercatori ed ignorato anche dagli stessi "addetti ai lavori" che nasce l'idea di una banca informativa con l'obiettivo di censire, raccogliere, gestire e valorizzare proprio questo genere di informazioni.

Non si tratta di una raccolta di dati di base, ed è per questo che si è voluto evitare il termine forse più familiare di "banca dati"; più correttamente, si tratta esclusivamente di un luogo di documentazione sui dati demografici già raccolti, e, in tal senso, risponde ad una logica comune, ad esempio, alle banche dati di tipo bibliografico, o alle cosiddette "banche dati di banche dati".

Nelle sue ambizioni, BIDDES (acronimo di Banca Informativa sulla Documentazione in DEMografia Storica) si vuole caratterizzare come uno strumento che consenta agli studiosi di avere un quadro completo su quanto già raccolto, ad esempio, su specifiche aree o collettività e di poter individuare, nel contempo, gli interlocutori più appropriati.

Come è noto, inoltre, le basi di dati possono svolgere, oltre alle funzioni più strettamente istituzionali di aggiornamento e di comunicazione, anche una azione di standardizzazione sulle modalità di gestione delle informazioni; questo aspetto potrebbe risultare

Figura 1



particolarmente prezioso per la valorizzazione delle informazioni storico-demografiche.

Il primo passo è stato così quello di censire il materiale già raccolto; a tal fine è stato usato un questionario che, con la base di dati, ha condiviso, non solo le informazioni, ma lo stesso impianto progettuale e di metodo.

1. Strutture e criteri per la definizione del questionario: progetto e costruzione del modello

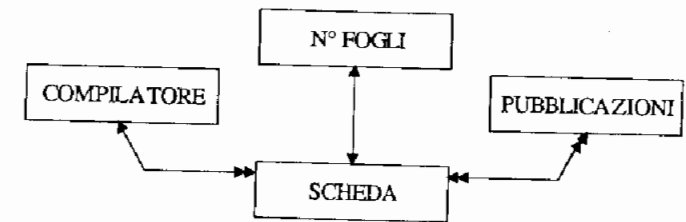
Nella presentazione degli obiettivi di BIDDES si è già anticipato come esista un rapporto strettissimo, tra il questionario proposto e l'architettura della banca informativa; può essere, dunque,

interessante osservare più da vicino l'organizzazione e la strutturazione del questionario¹; proprio per capire meglio la filosofia di base della banca: questionario ed archivio sono infatti le due facce di un unico progetto di modellizzazione delle informazioni che si voleva conoscere e gestire (Figura 1).

Vengono richieste innanzitutto quelle informazioni necessarie all'individuazione dell'area geografica e della collettività studiata; è questa infatti il soggetto delle informazioni ed il più plausibile oggetto dei bisogni informativi dell'utente della banca; queste notizie sono immediatamente arricchite da quelle di tipo bibliografico per poter sapere quali siano stati gli studi che già si sono occupati di questa realtà demografica. Completa la prima scheda la richiesta di tipo anagrafico sul curatore della raccolta di dati per consentire all'utente di risalire al rilavatore (ricercatore o istituto).

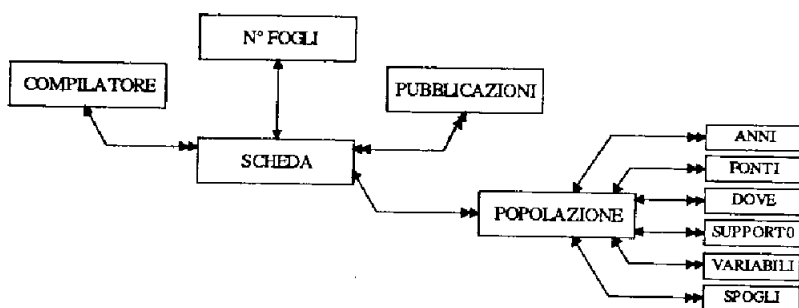
Si noti come in questa prima fase di definizione del modello, le

Figura 2



¹ In Appendice viene riportata la parte del questionario comprendente la Scheda Informativa e la Scheda dei Battesimi. Le restanti schede relative alla Popolazione, ai Matrimoni ed ai Decessi hanno una struttura simile a quella dei Battesimi.

Figura 3



entità² coinvolte siano la Scheda, il Compilatore e la Pubblicazione; con questo tipo di modellizzazione (Figura 2), il database, valorizzato con le risposte date alla prima scheda del questionario, sarebbe già in grado di rispondere a domande del tipo: chi ha compilato la scheda che riguarda l'area di interesse? Quali altre rilevazioni ha eseguito lo studioso di quest'area? Esistono già altre pubblicazioni che la riguardano e, se sì, quali? E di quali altre aree si occupavano queste pubblicazioni?

La seconda pagina del questionario raccoglie informazioni sulla popolazione che fa riferimento all'unità geografica o alla collettività descritta nella prima scheda. Come si può notare, le informazioni sulla popolazione, così come quelle sulle nascite, sui decessi e sui matrimoni, si possono ricondurre essenzialmente a due filoni informativi; si richiedono, infatti, sia notizie sulle modalità di gestione e di immagazzinamento fisico dei dati, sia chiarimenti sulle loro caratteristiche storico-demografiche e sulle loro fonti. Ecco,

² Nell'ambito di questa trattazione si utilizzerà spesso il termine "entità", che oltre all'ovvio significato generale, ha anche un'accezione specifica della teoria dei database; per semplicità si possono considerare le entità come i soggetti che, in reciproco rapporto, costituiscono il database, e a specificazione dei quali vengono valorizzate una serie di variabili.

dunque, coesistere all'interno dello stesso foglio di questionario, da un lato, domande relative al supporto del dato (cartaceo, microfilm, magnetico) e, dall'altro, notizie sui caratteri rilevati, come, ad esempio, il sesso, lo stato civile e la professione. L'inserimento di questi quesiti ha un immediato riscontro nella costruzione del modello e, quindi, della progettazione dell'architettura della banca dati (Figura 3).

Il database può a questo punto rispondere a domande sulla popolazione del tipo: per quali anni è stata condotta la rilevazione? quale è il supporto di rilevazione adottato? questo supporto è lo stesso per tutte le rilevazioni? la professione è stata rilevata per tutti gli anni o, viceversa, ci sono delle lacune?

Ovviamente si possono legare questi quesiti alle informazioni contenute nella scheda descritta precedentemente; ad esempio: quali sono le variabili rilevate per la popolazione di una certa zona e quali sono gli studiosi che se ne sono occupati?

Il modello si estende in maniera analoga alle informazioni relative alle nascite, ai matrimoni ed ai decessi ottenendo così l'architettura finale già anticipata nella Figura 1.

2. Progettazione e definizione del database fisico

In termini di file e tracciati record, il database assumerà una configurazione che sarà la traduzione in termini informatici della modellizzazione descritta.

Grande attenzione deve essere posta nella costruzione e nella definizione delle chiavi d'accesso ai file. In particolare, @Nascite, @Decessi, @Matrimoni e @Popolazione sono chiavi composte dalle variabili numero di scheda, tipo di processo studiato e numero progressivo del foglio di questionario che descrive quel processo (ad esempio, se un ricercatore ha riempito quattro fogli di questionario relativi alla popolazione di una certa zona, ogni foglio sarà

individuabile dal numero della scheda nella quale è inserito, dalla sigla che lo identifica come scheda di popolazione e dal numero d'ordine, da 1 a 4, del foglio stesso). Nella progettazione del modello di dati, si è utilizzato il classico metodo di normalizzazione³, e, in particolare, è stato sufficiente applicare solamente i primi tre criteri, partendo da un'entità indistinta identificata con il questionario nel suo complesso che, nell'organizzazione prescelta, era in rapporto univoco con una certa zona o una certa collettività; con l'applicazione dei primi tre criteri canonici, si sono potuti così individuare e gestire sia i rapporti di tipo gerarchico insiti nel modello (come ad esempio il fatto che un ricercatore avesse potuto inviare più schede ma che, comunque, ogni singola scheda dovesse essere ricondotta ad un solo compilatore), sia di tipo reticolare (che si manifestano, ad esempio, tutte le volte che su scheda vengono citati in bibliografia più libri, ognuno dei quali, a sua volta, può essere citato in più schede).

3. Conclusioni

Il database appena descritto fa parte del progetto di un sistema di gestione e valorizzazione delle informazioni sul patrimonio di dati disponibili in demografia storica. Affinché sia pienamente fruibile, oltre al modello di architettura, servono anche delle procedure che traducano in operazioni facilmente eseguibili tutte le funzionalità necessarie alla gestione, al controllo della qualità dei dati immessi e alla salvaguardia della robustezza dell'impianto informatico; è su questo versante che la progettazione del sistema è, attualmente, impegnata.

³ E. F. Codd (1971), A Relational Model of Data for Large Shared Banks, in *Communications of the A.C.M.*, Volume 13, Number 6, June.
E. F. Codd (1979), Extending the Data Base Relational Model to Capture more Meaning, in *A.C.M. Transactions on Database System*, Volume 4, Number 4, December.

Appendice

Scheda informativa

Sezione generale relativa alla località rilevata e al rilevatore

1. Compilatore della scheda informativa

Nome Cognome

Indirizzo: via n° civico

città prov.

Telefono/..... Fax

Posta elettronica

2. Unità rilevata

2.1 parrocchia Nome
(nome del santo e della località, es. S. Venanzio di Galliera, S. Nicola di Bari)

Diocesi di

2.2 comune/ità/
popolo

Attualmente nel Comune di Prov.

2.3 Altro (specificare)
(ad esempio sottogruppi o aree particolari, province, distretti, circondari, ecc.)

3. Lista delle pubblicazioni che hanno utilizzato i dati rilevati:

(autore, anno di edizione, titolo, nome della rivista, numero, anno, n° pagine / editore, città di edizione).

1.

2.

3.

(allegare lista nel caso in cui lo spazio non sia sufficiente)

NASCITE/BATTESIMI

AD Periodo rilevazione		Ci sono lacune significative?		SI NO	
Anno inizio					
Anno fine					
Dom. A.	Dom. B.	Dom. C.	Dom. D.	Dom. E.	Dom. F.
Fonte	La fonte è presso il rilevatore?	Qual è il supporto di rilevaz.	METODO di rilevazione	Variabili rilevate: eventi per...	Spogli disponibili:
Chiesa	Religiosa	SI, in fotocopia	Aggregativo	anno	Bventi per anno
A.1	B.1	B.1	D.1	anno	per anno e mese
A.2	B.2	B.2	D.2	anno	per anno e trimestre
	B.3	B.3	D.3	anno	per anno e semestre
	B.4	B.4	D.4	anno	per anno e per sesso
	B.5	B.5	D.5	anno	per anno e per età
	B.6	B.6	D.6	anno	per anno e per luogo
	B.7	B.7	D.7	anno	per anno e per altro
	B.8	B.8	D.8	anno	per anno e per altro
	B.9	B.9	D.9	anno	per anno e per altro
	B.10	B.10	D.10	anno	per anno e per altro
	B.11	B.11	D.11	anno	per anno e per altro
	B.12	B.12	D.12	anno	per anno e per altro
	B.13	B.13	D.13	anno	per anno e per altro
	B.14	B.14	D.14	anno	per anno e per altro
	B.15	B.15	D.15	anno	per anno e per altro
	B.16	B.16	D.16	anno	per anno e per altro
	B.17	B.17	D.17	anno	per anno e per altro
	B.18	B.18	D.18	anno	per anno e per altro
	B.19	B.19	D.19	anno	per anno e per altro
	B.20	B.20	D.20	anno	per anno e per altro
	B.21	B.21	D.21	anno	per anno e per altro
	B.22	B.22	D.22	anno	per anno e per altro
	B.23	B.23	D.23	anno	per anno e per altro
	B.24	B.24	D.24	anno	per anno e per altro
	B.25	B.25	D.25	anno	per anno e per altro
	B.26	B.26	D.26	anno	per anno e per altro
	B.27	B.27	D.27	anno	per anno e per altro
	B.28	B.28	D.28	anno	per anno e per altro
	B.29	B.29	D.29	anno	per anno e per altro
	B.30	B.30	D.30	anno	per anno e per altro
	B.31	B.31	D.31	anno	per anno e per altro
	B.32	B.32	D.32	anno	per anno e per altro
	B.33	B.33	D.33	anno	per anno e per altro
	B.34	B.34	D.34	anno	per anno e per altro
	B.35	B.35	D.35	anno	per anno e per altro
	B.36	B.36	D.36	anno	per anno e per altro
	B.37	B.37	D.37	anno	per anno e per altro
	B.38	B.38	D.38	anno	per anno e per altro
	B.39	B.39	D.39	anno	per anno e per altro
	B.40	B.40	D.40	anno	per anno e per altro
	B.41	B.41	D.41	anno	per anno e per altro
	B.42	B.42	D.42	anno	per anno e per altro
	B.43	B.43	D.43	anno	per anno e per altro
	B.44	B.44	D.44	anno	per anno e per altro
	B.45	B.45	D.45	anno	per anno e per altro
	B.46	B.46	D.46	anno	per anno e per altro
	B.47	B.47	D.47	anno	per anno e per altro
	B.48	B.48	D.48	anno	per anno e per altro
	B.49	B.49	D.49	anno	per anno e per altro
	B.50	B.50	D.50	anno	per anno e per altro
	B.51	B.51	D.51	anno	per anno e per altro
	B.52	B.52	D.52	anno	per anno e per altro
	B.53	B.53	D.53	anno	per anno e per altro
	B.54	B.54	D.54	anno	per anno e per altro
	B.55	B.55	D.55	anno	per anno e per altro
	B.56	B.56	D.56	anno	per anno e per altro
	B.57	B.57	D.57	anno	per anno e per altro
	B.58	B.58	D.58	anno	per anno e per altro
	B.59	B.59	D.59	anno	per anno e per altro
	B.60	B.60	D.60	anno	per anno e per altro
	B.61	B.61	D.61	anno	per anno e per altro
	B.62	B.62	D.62	anno	per anno e per altro
	B.63	B.63	D.63	anno	per anno e per altro
	B.64	B.64	D.64	anno	per anno e per altro
	B.65	B.65	D.65	anno	per anno e per altro
	B.66	B.66	D.66	anno	per anno e per altro
	B.67	B.67	D.67	anno	per anno e per altro
	B.68	B.68	D.68	anno	per anno e per altro
	B.69	B.69	D.69	anno	per anno e per altro
	B.70	B.70	D.70	anno	per anno e per altro
	B.71	B.71	D.71	anno	per anno e per altro
	B.72	B.72	D.72	anno	per anno e per altro
	B.73	B.73	D.73	anno	per anno e per altro
	B.74	B.74	D.74	anno	per anno e per altro
	B.75	B.75	D.75	anno	per anno e per altro
	B.76	B.76	D.76	anno	per anno e per altro
	B.77	B.77	D.77	anno	per anno e per altro
	B.78	B.78	D.78	anno	per anno e per altro
	B.79	B.79	D.79	anno	per anno e per altro
	B.80	B.80	D.80	anno	per anno e per altro
	B.81	B.81	D.81	anno	per anno e per altro
	B.82	B.82	D.82	anno	per anno e per altro
	B.83	B.83	D.83	anno	per anno e per altro
	B.84	B.84	D.84	anno	per anno e per altro
	B.85	B.85	D.85	anno	per anno e per altro
	B.86	B.86	D.86	anno	per anno e per altro
	B.87	B.87	D.87	anno	per anno e per altro
	B.88	B.88	D.88	anno	per anno e per altro
	B.89	B.89	D.89	anno	per anno e per altro
	B.90	B.90	D.90	anno	per anno e per altro
	B.91	B.91	D.91	anno	per anno e per altro
	B.92	B.92	D.92	anno	per anno e per altro
	B.93	B.93	D.93	anno	per anno e per altro
	B.94	B.94	D.94	anno	per anno e per altro
	B.95	B.95	D.95	anno	per anno e per altro
	B.96	B.96	D.96	anno	per anno e per altro
	B.97	B.97	D.97	anno	per anno e per altro
	B.98	B.98	D.98	anno	per anno e per altro
	B.99	B.99	D.99	anno	per anno e per altro
	B.100	B.100	D.100	anno	per anno e per altro

Dom. A: _____

Dom. B: _____

Dom. C: _____

Dom. D: _____

Dom. E: _____

Dom. F: _____

Dom. G: _____

Dom. G: Dispone per singolo anno di calendario del numero del matrimonio compresi nella rilevazione dei fatti?

NO SI Non so

Dom. G: Ha escluso i matrimoni?

NO SI Non so

NOTE (in questo spazio il rispondente può specificare aspetti particolari)

Dom. A0: _____

Giorgio E. Montanari - Odoardo Bussini

CONSIDERAZIONI SUL CAMPIONAMENTO IN DEMOGRAFIA STORICA

Nell'ambito del progetto di ricerca 40% sui "Modelli di sviluppo demografico in Italia tra XVIII e XIX secolo", ci si è posti preliminarmente l'obiettivo di risolvere alcuni problemi campionari. In quest'ottica, essenzialmente interdisciplinare, che vede il coinvolgimento di statistici metodologi e di demografi, viene presentato questo primo contributo, frutto comune di lunghe e proficue discussioni svolte anche a livello del sottogruppo campionamento. Ovviamente, la parte metodologica dell'articolo è da attribuire a G. E. Montanari.

Introduzione

In demografia storica la ricostruzione di popolazioni del passato in una determinata regione geografica richiede la stima dei dati di movimento (nascite, morti, matrimoni) di quelle popolazioni, allo scopo di studiarne le principali caratteristiche demografiche. I dati elementari, come è noto, provengono nella gran parte dei casi dalle registrazioni ecclesiastiche dei battesimi, matrimoni e decessi che si celebravano nei territori di competenza delle singole parrocchie. Tali fonti sono conservate negli archivi storici parrocchiali, in quelli diocesani, negli archivi di Stato o in quelli comunali.

Tuttavia non è sempre possibile procedere alla rilevazione dei dati relativi all'intervallo di tempo considerato in tutte le parrocchie che

insistono sulla regione esaminata, specie se di grandi dimensioni. In primo luogo, perché i registri parrocchiali a volte sono andati distrutti o risultano indecifrabili e lacunosi o ne viene negata la consultazione; in secondo luogo perché, anche se tutti i registri fossero integri e disponibili, la mole dei dati da rilevare sarebbe tale da richiedere risorse superiori a quelle disponibili.

Si comprende allora come il campionamento, cioè l'osservazione di un numero limitato di parrocchie, sia una scelta obbligata se si vuole comunque pervenire ad una stima dei dati di movimento complessivi dell'area oggetto di studio. In questo caso diventa cruciale poter misurare la precisione e l'affidabilità delle stime, e in questa nota si intendono illustrare le metodologie che consentono il raggiungimento dell'obiettivo. A tal fine si ridurranno al minimo i dettagli tecnici, per concentrarsi sulla sostanza delle metodologie esaminate.

2 - Posizione del problema

Si supponga di voler stimare le intensità in una data regione di uno o più fenomeni di movimento in intervalli di tempo consecutivi e della stessa lunghezza, sulla base dei dati rilevati da un campione di parrocchie di quella regione. Per semplificare, si assumerà qui che le parrocchie costituiscano una partizione della regione in studio e che siano in vita per tutto il periodo di tempo considerato. Inoltre, in questa nota si farà riferimento solo alle morti, potendosi in modo analogo trattare le nascite e i matrimoni.

L'impostazione del lavoro è la seguente: il totale dei morti in un dato intervallo di tempo può essere visto come il totale di una variabile quantitativa nella popolazione finita delle parrocchie appartenenti alla regione in esame. Sia allora N la dimensione della popolazione e T il numero degli intervalli considerati. Denoteremo con Y_{it} il numero dei morti conteggiati nell'intervallo di tempo t -esimo ($t=1,2,\dots,T$) presso la parrocchia i -esima ($i=1,2,\dots,N$). Per brevità nel

seguito parleremo semplicemente di tempo t -esimo anziché intervallo.

Il totale dei morti nella regione al tempo t , $Y_t = \sum_{i=1}^N Y_{it}$, è un parametro descrittivo della popolazione di cui si vuole stimare i valori ai tempi $t=1,2,\dots,T$.

L'efficienza delle stime aumenta se si utilizzano variabili ausiliarie correlate a quelle oggetto di studio e con valori noti in ogni unità della popolazione. Una di queste è sicuramente il numero delle anime della parrocchia e con X_{it} si denoterà la variabile che quantifica, anche in modo approssimativo, il numero di anime della parrocchia i -esima al tempo t -esimo (in questo lavoro si è attribuito ad X_{it} il valore desunto dagli stati delle anime quando erano disponibili o i valori ottenuti con una interpolazione lineare per punti in caso contrario) ed avente X_t come totale al tempo t . Qualunque altra informazione sulle singole parrocchie (tipologia del territorio, altitudine, grado di urbanizzazione, attività economica, ceto sociale della popolazione residente, ecc.), è altresì utile ai fini della suddivisione delle stesse in sottogruppi più omogenei, con riferimento alle variabili indagate. Queste informazioni consentono di configurare dei caratteri di stratificazione.

3 - Approcci all'inferenza su popolazioni finite

Sono sostanzialmente due gli approcci all'inferenza su popolazioni finite: l'approccio classico e l'approccio predittivo.

L'approccio classico si basa sul campionamento probabilistico che presuppone la selezione delle parrocchie con criteri casuali, in modo tale che ognuna di esse abbia una probabilità positiva e conosciuta di appartenere al campione. Questa condizione si realizza attraverso la formulazione di appropriati piani di campionamento. È un approccio perseguibile quando siano verificate due condizioni: a) è possibile procedere a nuove rilevazioni, eventualmente da affiancare a quelle

già effettuate, allo scopo di campionare la popolazione di parrocchie residua (tolte cioè le parrocchie già rilevate ma non campionate con criteri probabilistici); b) è realistico attendersi un tasso contenuto, cioè inferiore al 10-20%, di parrocchie campione non osservabili per indisponibilità dei registri. Questa seconda condizione è necessaria per poter ricorrere alle metodologie di trattamento delle mancate risposte, senza inficiare sostanzialmente l'inferenza che si va a fare (cfr. Särndal *et al.*, 1992; cap. 15).

L'approccio predittivo parte dalla specificazione di un modello di superpopolazione che si presume in grado di descrivere in termini probabilistici la variabilità propria delle variabili indagate. Comunque il campione sia stato estratto, il modello può essere poi identificato sulla base dei dati campionari, in quanto la validità dell'inferenza poggia sulla corretta specificazione del modello. Questo approccio è particolarmente interessante quando si vogliono utilizzare dati già raccolti, ovvero quando il campione non è casuale e non è possibile integrarlo con un campionamento probabilistico.

4 - Gli schemi di campionamento probabilistici per le indagini ripetute

Quando si vuol stimare il valore che un parametro descrittivo assume in tempi diversi, si è in presenza di una indagine ripetuta nel tempo. In questo caso le metodologie di campionamento più seguite sono quelle appresso descritte.

4.1 - Il campione panel

Si supponga che lo stesso campione di parrocchie venga utilizzato per tutto il periodo oggetto di indagine. Per estrarre il campione risulta più efficiente utilizzare la relazione di approssimata

proporzionalità tra la variabile indagata Y_{it} ed il numero delle anime, mediante l'impiego del campionamento con probabilità variabili e proporzionali alla dimensione delle parrocchie, quantificata dalla variabile X_{it} . Questo significa che se n è la dimensione (che si assume fissa) del campione da estrarre e si fa riferimento al numero delle anime al tempo 1 (in alternativa si possono considerare i valori di X_{it} , $i=1,2,\dots,N$, al tempo centrale o i valori medi nell'arco del periodo), la probabilità che la parrocchia i -esima appartenga al campione è $\pi_i = nX_{i1} / X_{1.}$

Lo stimatore non distorto del totale Y_t (i valori attesi sono calcolati rispetto all'estrazione ripetuta del campione) è quello di Horvitz-Thompson dato dalla espressione

$$\hat{Y}_t = \sum_{i \in s} \frac{Y_{it}}{\pi_i} = X_{1.} \sum_{i \in s} \frac{Y_{it}}{nX_{i1}}, \quad (1)$$

dove s contiene gli indici delle parrocchie campione. La varianza di questo stimatore è ben nota (cfr. Cicchitelli *et al.*, 1992, cap. 4); per semplicità riportiamo qui quella propria del campionamento con ripetizione (in questo caso le π_i sono delle frequenze attese di inclusione). Essa risulta così più trasparente ed assume, nel caso in esame, l'espressione

$$V(\hat{Y}_t) = \frac{X_{1.}}{n} \sum_{i=1}^N X_{i1} \left(\frac{Y_{it}}{X_{i1}} - \frac{Y_{t.}}{X_{1.}} \right)^2. \quad (2)$$

Si osservi come la varianza dello stimatore dipenda dalla variabilità dei tassi di mortalità generici Y_{it}/X_{i1} delle singole parrocchie rispetto a quello regionale $Y_{t.}/X_{1.}$: se ogni parrocchia

avesse lo stesso tasso di mortalità e il numero delle anime rimanesse costante nel tempo, la varianza dello stimatore sarebbe nulla.

Se invece il numero delle anime varia considerevolmente nell'arco di tempo considerato, uno stimatore più efficiente è quello per quoziente,

$$\hat{Y}_{iq} = X_i \frac{\hat{Y}_i}{\hat{X}_i} = X_i \frac{\sum_{i \in S} Y_{it} / \pi_i}{\sum_{i \in S} X_{it} / \pi_i},$$

che, tra l'altro, non richiede la conoscenza dei valori X_{ti} nelle unità non campionate. Questo stimatore è asintoticamente non distorto e richiede perciò campioni di adeguata dimensione, almeno trenta unità (per la sua varianza si veda Cicchitelli *et al.*, 1992, p. 117). Inoltre esso è efficiente anche nel caso di campionamento con probabilità uguali ($\pi_i = n/N$), tuttavia può risultare preferibile campionare con maggiore probabilità le parrocchie più grandi, per le quali è solitamente più facile accedere alle fonti archivistiche.

Nel caso siano disponibili dei caratteri di stratificazione, la stima non distorta del totale e la varianza dello stimatore si ottengono mediante la somma, rispettivamente, degli stimatori (1) e delle varianze (2) relative ai singoli strati. La precisione dello stimatore del totale dei morti dipende ora dalla variabilità dei tassi di mortalità entro gli strati. Gli strati devono essere quindi costruiti mettendo insieme parrocchie omogenee rispetto alle cause di morte. Se poi si vuole utilizzare lo stimatore per quoziente, questo dovrebbe essere del tipo combinato (Cicchitelli *et al.*, 1992; p. 158), essendo verosimilmente poche le parrocchie campione per strato. Si dimostra che, nel caso di variabilità costante dei tassi di mortalità entro gli strati, l'efficienza è massima se il numero delle parrocchie campione negli strati è proporzionale al numero totale delle anime negli strati.

A titolo esemplificativo, utilizzando i dati noti delle 35 parrocchie urbane di Perugia per gli anni 1730-1789, si sono calcolati i coefficienti di variazione degli stimatori sopra proposti, relativamente ad intervalli temporali di lunghezza diversa e per campioni di ampiezza 1. La popolazione complessiva si aggira sulle 12.000 anime, mentre il tasso di mortalità medio dell'intero periodo risulta pari al 32,2 per mille, con valori oscillanti da un minimo di 22,0 ad un massimo di 56,5 per mille. È stata considerata anche una semplice stratificazione dividendo le parrocchie in due strati, quelle più centrali e quelle periferiche. Nella tabella 1 sono riportate le medie (\overline{CV}) e le deviazioni standard (σ_{CV}) dei coefficienti di variazione ottenuti da intervalli della stessa lunghezza.

Tabella 1

Lunghezza intervallo	Numero dei casi	Senza stratificazione			Con stratificazione		
		\overline{CV}	σ_{CV}	n^*	\overline{CV}	σ_{CV}	n^*
1 anno	60	37,7	5,0	57	36,2	5,1	53
2 anni	30	30,4	3,7	37	29,0	3,8	34
3 anni	20	27,4	3,4	31	26,1	3,0	28
5 anni	12	26,0	3,1	28	24,7	2,8	24
10 anni	6	24,1	3,0	24	22,6	2,3	21
60 anni	1	23,6	-	23	21,7	-	19

Si noti come il coefficiente di variazione medio delle stime diminuisce all'aumentare della lunghezza dell'intervallo passando, nel caso non stratificato, dal 37,7% delle stime annuali al 24,1% delle stime dei morti per decenni. Valori di poco inferiori si ottengono con la stratificazione utilizzata. Quando si stima il totale dei morti nell'arco dei 60 anni si ottiene comunque un coefficiente di variazione del 23,6%. Questi risultati segnalano che gran parte della variabilità delle stime dipende da tassi di mortalità parrocchiali sistematicamente diversi dal tasso medio regionale, mentre già con le

stime triennali la variabilità naturale del fenomeno (cioè la componente di varianza entro le parrocchie degli scostamenti dei tassi di mortalità parrocchiali da quello regionale) è in gran parte neutralizzata. Infine, a titolo di esempio, nella tabella sono riportate le dimensioni (n^*) del campione, necessarie per ottenere un coefficiente di variazione del 5%, immaginando di campionare una popolazione infinita di parrocchie dove i tassi di mortalità presentino la stessa variabilità registrata per i dati di Perugia (n^* si ottiene dal rapporto tra il quadrato di CV e il quadrato del coefficiente di variazione desiderato). Si noti come le stime per quinquenni consentano di dimezzare la dimensione del campione a parità di precisione.

È chiaro che l'utilizzo di uno stesso campione per tutti i tempi esaminati genera delle stime tra loro correlate. È plausibile attendersi una correlazione positiva, e quindi un aumento di efficienza quando si vada a stimare la differenza nel numero dei morti relativi a tempi diversi. La stessa correlazione va a discapito però dello stimatore del valore medio di Y_t in tempi diversi quando, ad esempio, si vogliono calcolare delle medie mobili. In questo caso sarebbe più efficiente selezionare un diverso campione per ogni tempo, una possibilità non percorribile operativamente.

Nella tabella 2 sono riportati il valore medio (\bar{p}) e la deviazione standard (σ_p) delle correlazioni tra gli stimatori \hat{Y}_t e \hat{Y}_{t+s} che si ottengono al variare di t , assegnando ad s , per esempio, i valori che vanno da 1 a 20 per le stime annuali e da 1 a 4 nel caso delle stime dei totali quinquennali.

Per quanto riguarda le stime annuali, la correlazione (di entità modesta) sembra decadere molto lentamente all'aumentare di s , forse a causa dell'invecchiamento delle probabilità π_t che sono basate, nei calcoli effettuati, sul numero delle anime al tempo t . Tale diminuzione è più evidente per le stime quinquennali, le cui

Tabella 2

Stime annuali						Stime quinquennali		
s	\bar{p}	σ_p	s	\bar{p}	σ_p	s	\bar{p}	σ_p
1	0,42	0,21	11	0,29	0,13	1	0,49	0,12
2	0,27	0,11	12	0,18	0,12	2	0,47	0,33
3	0,22	0,21	13	0,17	0,14	3	0,38	0,33
4	0,34	0,15	14	0,22	0,11	4	0,28	0,47
5	0,19	0,24	15	0,27	0,09			
6	0,36	0,18	16	0,22	0,20			
7	0,27	0,25	17	0,15	0,13			
8	0,20	0,12	18	0,14	0,22			
9	0,21	0,17	19	0,27	0,21			
10	0,29	0,18	20	0,20	0,22			

correlazioni sono più elevate a causa del minor contributo della variabilità naturale dei tassi di mortalità alla varianza delle stime.

4.2 - I campioni ruotati

In alternativa ai campioni panel, risultano più efficienti i campioni ruotati, che comportano però costi maggiori dal punto di vista operativo. In sintesi il campione iniziale delle n parrocchie viene suddiviso casualmente in k subcampioni di ampiezza $m=n/k$, assimilabili ad altrettanti campioni indipendenti della stessa popolazione. Denotato con l un intero positivo, ogni l tempi il subcampione più anziano (o uno a caso quando ve ne sia più di uno) viene sostituito con un nuovo subcampione riestratto *ex novo* dalla popolazione di parrocchie. In questo modo, a regime, ogni parrocchia viene osservata per un periodo di $k \times l$ tempi. Si è così definito uno schema di rotazione del campione. Ciascun subcampione consente di costruire delle stime corrette (o per quoziente) del totale dei morti Y_t nei diversi tempi. Tali stime, che diremo elementari, sono

indipendenti se calcolate con subcampioni diversi, sono invece correlate positivamente (per quanto visto nel paragrafo precedente) se calcolate con lo stesso subcampione. È proprio questa correlazione che consente di aumentare l'efficienza delle stime come di seguito indicato. Tutte le stime elementari possono essere inserite in un modello di regressione $Y = X\beta + e$ dove Y è il vettore delle $k \times T$ stime elementari; β è il vettore dei T totali Y_t , $t=1,2,\dots,T$; X è una matrice di zeri e di 1 che associa ad ogni stima elementare l'elemento di β di cui è uno stimatore corretto; e è un vettore di residui con valore atteso nullo e con varianza $V = \text{Var}(e)$ data dalla matrice di varianze e covarianze delle stime elementari. Fissata una qualsiasi combinazione lineare $L'\beta$ di β , il suo stimatore lineare non distorto con varianza minima è $L'\hat{\beta} = L'(XV^{-1}X)^{-1}XV^{-1}Y$, con varianza $V(L'\hat{\beta}) = L'(XV^{-1}X)^{-1}L$. La soluzione è ottima se la matrice V è nota, diversamente occorre procedere a una sua valutazione sulla base di alcune semplificazioni o ad una stima con qualche perdita di efficienza.

Il guadagno in efficienza consentito da questa metodologia è tanto maggiore quanto più grande è la correlazione tra stime elementari da uno stesso subcampione. A titolo di esempio, si considerino le stime quinquennali e si ponga $k=2$ e $l=10$. Sotto l'ipotesi semplificatrice della costanza nel tempo della varianza e della covarianza tra le stime elementari, si trova che la varianza dello stimatore $\hat{\beta}_t$ di Y_t è non superiore a $V(\hat{Y}_t)(1-0.5\rho^2)$, dove ρ è la correlazione tra stime elementari quinquennali successive e $V(\hat{Y}_t)$ è la varianza dello stimatore (1). Poiché dalla tabella 2 si desume un valore di ρ pari a 0.49, si ha che rispetto a \hat{Y}_t , lo stimatore $\hat{\beta}_t$ consente un decremento percentuale della varianza superiore al 12%, e quindi nella stessa misura della dimensione del campione per una prefissata precisione.

I valori di correlazione trovati non sono per la verità elevati e l'aumento di efficienza delle stime con campioni ruotati potrebbe non compensare il maggiore onere della rilevazione che, coinvolgendo un maggior numero di parrocchie, pone più problemi di ricerca ed utilizzo delle fonti storiche. La scelta di uno schema di rotazione rientra nel problema del dimensionamento ottimale del campione, tenendo conto della precisione desiderata e del costo della rilevazione. Per un approfondimento di questi aspetti si veda, tra gli altri, Jones (1980).

5 - L'approccio predittivo

Nell'approccio predittivo si ipotizza che il numero dei morti nella parrocchia i al tempo t sia il valore assunto da una variabile casuale Y_{it} , con $i=1,2,\dots,N$ e $t=1,2,\dots,T$. Le caratteristiche della distribuzione di probabilità congiunta delle $N \times T$ variabili casuali Y_{it} sono specificate dal cosiddetto modello di superpopolazione. Quest'ultimo può essere più o meno articolato in base alle informazioni che si hanno sulla popolazione indagata: da semplici condizioni sulla forma dei primi momenti delle Y_{it} , alla esatta forma funzionale della funzione di probabilità congiunta.

In questa nota si considererà un semplice modello di superpopolazione in grado di produrre comunque buoni risultati senza richiedere eccessivi dettagli tecnici. Esso specifica semplicemente il valore atteso e la varianza delle Y_{it} nonché la covarianza tra di esse:

Le quantità β_t , σ_t e ρ sono i parametri incogniti del modello secondo cui il valore atteso di Y_{it} è pari alle anime della parrocchia al tempo t per il tasso di mortalità β_t proprio del tempo t , mentre la varianza è assunta proporzionale allo stesso valore atteso, sulla base della considerazione che Y_{it} dovrebbe seguire la distribuzione di

Poisson; la correlazione tra coppie di variabili casuali relative ad una stessa parrocchia è costante e pari a ρ .

$$\begin{cases} E(Y_{it}) = X_{it}\beta_t & i=1,2,\dots,N; t=1,2,\dots,N; \\ V(Y_{it}) = X_{it}\sigma_t^2 & i=1,2,\dots,N; t=1,2,\dots,N; \\ Cov(Y_{it}, Y_{it'}) = \begin{cases} 0 & \text{se } i \neq i' \\ \rho\sigma_t\sigma_{t'}\sqrt{X_{it}X_{it'}} & \text{se } i = i' \text{ e } t \neq t' \end{cases} \end{cases} \quad (3)$$

In questo contesto la quantità Y_t non è più un parametro da stimare ma una variabile casuale che occorre prevedere o predirre. Da qui il nome di approccio predittivo alla inferenza su popolazioni finite. Una funzione dei dati campionari, \hat{Y}_t , utilizzata per prevedere il valore di Y_t , viene chiamata predittore. Il valore atteso e la varianza di un predittore, vengono calcolati sulla base della distribuzione di probabilità congiunta, ipotizzata dal modello di superpopolazione e condizionatamente al campione estratto. In particolare, un predittore \hat{Y}_t è detto non distorto se $(\hat{Y}_t - Y_t | s) = 0$, quali che siano i parametri del modello di superpopolazione ed il campione estratto. Si chiama errore di previsione medio la quantità $(\hat{Y}_t - Y_t | s)^2$, cioè il valore atteso condizionato ad s dello scostamento al quadrato tra predittore e variabile casuale da prevedere. Un predittore si dirà ottimo in una determinata classe se in essa non esiste un altro predittore con un errore di previsione più piccolo, quali che siano i valori dei parametri del modello ed il campione estratto s .

Venendo al problema di inferenza che ci interessa, sia s , l'insieme delle parrocchie campione al tempo t e si assuma per ora $\rho=0$. In questo caso, non è difficile mostrare che il predittore lineare di Y_t , non distorto e con errore di previsione medio minimo è

$$\hat{Y}_{t0} = \sum_{i \in s_t} Y_{it} + \hat{\beta}_t \sum_{i \in s_t} X_{it} \quad (4)$$

dove $\hat{\beta}_t = \sum_{i \in s_t} Y_{it} / \sum_{i \in s_t} X_{it}$ non è altro che il tasso di mortalità campionario. Nella (4) si riconosce la struttura caratteristica di un predittore del totale: alla somma dei valori osservati nelle unità campionarie viene addizionata la somma dei valori predetti delle variabili casuali non osservate. Poiché però vale l'identità $\hat{Y}_{t0} = \hat{\beta}_t \sum_{i \in s_t} X_{it}$, trova piena giustificazione il metodo di stima intuitivo consistente nell'applicare il tasso campionario di mortalità all'intera popolazione considerata. Per quanto riguarda l'errore di previsione medio si trova

$$E(\hat{Y}_{t0} - Y_t | s)^2 = \frac{\sum_{i \in s_t} X_{it}}{\sum_{i \in s_t} X_{it}} X_{it} \sigma_t^2 \quad (5)$$

La (5) è una quantità tanto più piccola quanto maggiore è il numero delle anime delle parrocchie selezionate: il campione ottimo è allora quello costituito dalle n parrocchie più grandi. Trova così piena legittimità la prassi di scegliere le parrocchie più grandi per la rilevazione dei dati. L'approccio predittivo quindi non solo non richiede la selezione casuale del campione ma, al contrario, suggerisce l'impiego di un campione per scelta ragionata. Tuttavia, il campione che garantisce il minimo errore di previsione è anche quello che genera le distorsioni più grandi quando il modello non è correttamente specificato. Lo stimatore dell'errore di previsione medio si ottiene sostituendo nella (5) lo stimatore non distorto del parametro σ_t^2 dato da

$$\hat{\sigma}_r^2 = \frac{1}{n_r - 1} \sum_{i \in S_r} \frac{(Y_{it} - \hat{\beta}_r X_{it})^2}{X_{it}}$$

dove n_r è il numero delle parrocchie campione al tempo t . Rapportando l'errore di previsione medio (5) al quadrato del valore atteso di Y_t ed estraendo poi la radice quadrata, si ottiene il coefficiente di variazione del predittore che può essere posto nella forma

$$CV = (\hat{Y}_{t0}) = \frac{\sigma_r}{\beta_r \sqrt{X_t}} \sqrt{\frac{1-f_t}{f_t}}$$

dove $f_t = \sum_{i \in S_r} X_{it} / X_t$. Questa espressione permette di definire la quota di anime che deve essere inclusa nel campione di parrocchie per conseguire una determinata precisione del predittore.

In presenza di caratteri di stratificazione, l'efficienza del predittore può essere aumentata ricorrendo alla stratificazione delle parrocchie ed all'utilizzo del modello (3) nei singoli strati. Si dimostra che se il parametro σ_r^2 non varia molto da strato a strato, la ripartizione ottimale delle parrocchie campione agli strati è quella per cui il rapporto tra il totale delle anime delle parrocchie campione ed il totale di strato è lo stesso in ogni strato.

Se il parametro ρ del modello (3) non è nullo, il predittore ottimo \hat{Y}_{t0} non è più tale. La correlazione tra osservazioni relative a tempi diversi in una stessa parrocchia consente infatti la costruzione di un predittore più efficiente che non verrà esaminato in questo lavoro per mancanza di spazio. Il predittore (4) rimane tuttavia ottimo nella classe dei predittori lineari che utilizzano solo i dati del tempo t , e i predittori \hat{Y}_{t0} , $t=1,2,\dots,T$, non sono più incorrelati. Si dimostra infatti

che nel caso di un campione panel con parrocchie aventi numero di anime costante e costanza del parametro σ_r , si ottiene $E[(\hat{Y}_{t0} - Y_t)(\hat{Y}_{r0} - Y_r) | S] = \rho V(\hat{Y}_{t0})$, la covarianza cioè tra gli errori di previsione è una quota ρ dell'errore di previsione medio dei singoli predittori. Un valore positivo di ρ aumenta l'efficienza della previsione della differenza $Y_t - Y_r$ mediante il predittore $\hat{Y}_{t0} - \hat{Y}_{r0}$.

In sintesi, nell'approccio predittivo, l'inferenza è condizionata al campione e un predittore rimane ottimo indipendentemente da come il campione è stato estratto. La validità dell'inferenza è però subordinata alla corretta specificazione del modello di superpopolazione. Un modello non correttamente specificato può dar luogo a delle istorsioni che invalidano il processo inferenziale. Per questo motivo in letteratura si consiglia l'introduzione di elementi di casualità nella estrazione delle parrocchie, in modo da aumentare la rappresentatività del campione rispetto a eventuali grandezze incognite ignorate erroneamente dal modello assunto. Altri metodi tendenti a rendere robusta l'inferenza fanno uso del bilanciamento del campione o fanno ricorso ad un misto tra campionamento probabilistico e approccio predittivo, imponendo la condizione che un predittore non distorto rispetto al modello sia anche asintoticamente non distorto rispetto al piano di campionamento. Per un approfondimento dell'approccio predittivo, si veda Cicchitelli *et al.* (1992, cap. 12) e Bolfarine e Zacks (1992).

5.1 - Applicazione numerica

Le 35 parrocchie urbane di Perugia descritte in precedenza costituiscono una popolazione nota utilizzabile per delle simulazioni. L'approccio predittivo discusso nel paragrafo precedente è stato sperimentato assumendo come campione le 7 parrocchie utilizzate da

Tittarelli (1977) e scelte tra le più grandi in base a criteri di rappresentatività e accessibilità delle fonti. Esse costituiscono un panel osservato per 60 anni, dal 1730 al 1789. La percentuale di anime contenute nelle 7 parrocchie si aggira intorno al 30%. Per ciascun anno sono stati calcolati il valore del predittore \hat{Y}_{t0} , la stima dell'errore di previsione medio e del coefficiente di variazione del predittore e gli estremi dell'intervallo di confidenza al 95% sulla base della approssimazione normale, benché la dimensione del campione non autorizzi tale approssimazione.

Nell'arco di tutti i 60 anni, solo in tre casi l'intervallo di confidenza non contiene il valore noto di Y_t , risultato sempre superiore, anche se si tratta di anni con mortalità media. Mediamente il coefficiente di variazione stimato è del 10,2%, mentre il valore medio delle stime di $\hat{\beta}_1$ è risultato 0,03174 a fronte di un valore effettivo di 0,03224. La media delle stime di σ_t^2 è risultata pari a 0,0570, un valore che segnala la presenza di fenomeni di iperdispersione rispetto alla distribuzione di Poisson, ovvero di un effetto parrocchia. Per una verifica empirica del comportamento del predittore (4), si è calcolato lo scarto relativo $(\hat{Y}_{t0} - Y_t)/Y_t$ e il relativo quadrato per ciascuno dei 60 anni esaminati. La media degli scarti relativi è risultata pari a -1,6%, un valore in accordo con la teoria secondo cui il predittore è non distorto; la media quadratica degli scarti relativi, assimilabile ad un coefficiente di variazione empirico, è invece pari a 9,4%. Il sostanziale accordo tra il coefficiente di variazione empirico e quello medio stimato dal campione suggerisce che il modello (6) è adeguato a descrivere la variabilità del fenomeno. Mediante i residui standardizzati $(Y_n - \hat{\beta}_1 X_n)/\hat{\sigma}_t \sqrt{X_n}$ relativi alle 7 parrocchie campione, si è stimata la correlazione tra coppie di residui appartenenti ad una stessa parrocchia (omogeneità intraparrocchiale), risultata pari a

0,27, un valore prossimo a quello medio registrato sotto l'approccio classico.

Nella tabella 3 sono riportati a titolo di esempio i risultati relativi agli anni 1761-1780. Le colonne contengono nell'ordine il totale delle anime (X_t), i morti effettivi (Y_t), i morti stimati con il predittore (\hat{Y}_{t0}), il coefficiente di variazione stimato del predittore (\hat{CV}), la stima di σ_t^2 , gli estremi inferiore (U) e superiore (V) dell'intervallo di confidenza al 95%.

Tabella 3

Anno	X_t	Y_t	\hat{Y}_{t0}	\hat{CV}	$\hat{\sigma}_t^2$	U	V
1761	12449	360	332	11,1	0,046	258	405
1762	12468	377	406	10,1	0,058	324	488
1763	12497	396	339	11,4	0,050	262	417
1764	12525	709	788	8,6	0,154	653	922
1765	12551	408	408	8,5	0,040	339	477
1766	12453	571	566	10,2	0,112	450	681
1767	12366	699	649	7,4	0,078	553	744
1768	12131	397	397	9,2	0,048	324	470
1769	12073	434	479	9,0	0,066	393	565
1770	12096	301	297	13,2	0,056	219	376
1771	12135	275	242	7,5	0,012	206	278
1772	12164	279	273	6,4	0,010	238	307
1773	12179	473	512	7,9	0,058	430	593
1774	12222	310	325	8,9	0,028	267	383
1775	12252	419	373	11,5	0,062	287	458
1776	12276	270	275	13,5	0,046	201	349
1777	12301	335	370	13,6	0,086	270	471
1778	12318	312	281	10,0	0,028	224	337
1779	12335	491	515	7,7	0,054	435	594
1780	12316	326	336	12,9	0,062	249	422

6. - Conclusioni

Il campionamento probabilistico rappresenta una modalità di fare inferenza su popolazioni finite non dipendente da assunzioni aprioristiche ed è quindi per certi versi preferibile, data la sua robustezza. Tuttavia l'approccio predittivo costituisce una efficiente alternativa specie per piccoli campioni, qualora il campionamento non sia probabilistico, come spesso accade in demografia storica. Se da una parte però non si richiede il campionamento casuale, dall'altra l'approccio è meno robusto perchè necessita della corretta specificazione del modello di superpopolazione. Il modello presentato nel paragrafo 5 sembra sufficientemente adeguato, visti i risultati prodotti, e costituisce un buon punto di partenza per modellazioni più sofisticate. Ulteriori studi empirici possono poi consolidare le metodologie utilizzate.

A questo punto si dispone di metodi per costruire stime di serie storiche aggregate, con una precisione misurabile e predeterminabile. Il quesito fondamentale che resta da sciogliere è stabilire quale sia il livello di precisione opportuno per l'utilizzazione che di questi dati si farà. In altre parole, occorre sapere in che misura i risultati dell'analisi demografica che utilizzano le serie storiche di movimento aggregate, dipendono dalla precisione con cui queste ultime sono stimate. A questa domanda si dovrà rispondere con ulteriori studi e ricerche.

Bibliografia

- H. Bolfarine, S. Zacks (1992), *Prediction theory for finite populations*, Springer-Verlag, New York.
- G. Cicchitelli, A. Herzel, G. E. Montanari (1992), *Il campionamento statistico*, Il Mulino, Bologna.
- R. G. Jones (1980), Best linear unbiased estimators for repeated sampling, *Journal of the Royal Statistical Society*, B42, 221-226.
- C. E. Särndal, B. Swensson, J. Wretman (1992), *Model Assisted Survey Sampling*, Springer-Verlag, New York.
- L. Tittarelli (1977), Sulle crisi di mortalità in Perugia nei secoli XVII-XVIII e XIX, *Quaderni dell'Istituto di Statistica*, Università di Perugia, n. 1.

LA QUALITÀ DEI DATI NEI DOCUMENTI STORICI
Una correzione dell'attrazione nella distribuzione delle età

I. Il problema

Come è noto, tra le caratteristiche dei documenti storici che riportano elenchi di persone di cui fosse richiesta l'indicazione dell'età, due rischiano di inquinare più o meno pesantemente l'analisi dei contenuti più sostanziali: l'incompletezza dell'elenco e la cattiva qualità dell'età riportata. Nel linguaggio odierno, in termini più corretti, si parlerebbe di errori rispettivamente di copertura e di accuratezza; ma è improprio il riferimento ad "errori" in documenti storici quali gli *Stati delle Anime* (S.A.) e i *Registri dei Defunti* (R.D.), compilati con finalità tutt'altro che statistiche, caso mai "amministrative", ma non nel senso odierno, quanto forse in un'accezione più "pratica" che burocratica.

Per quanto concerne la *copertura*, oltre alla consueta mancanza negli S.A., come è noto (Bellettini, 1973), di particolari categorie di persone non appartenenti alla comunità parrocchiale (ebrei, carcerati, militari, monaci, suore, ricoverati in ospedale, ecc.), l'altra mancanza più frequente è quella di una certa quota di bambini. Questa può essere compresa in funzione dello scopo principale di tali enumerazioni, ossia la verifica del soddisfacimento del precetto pasquale (riservato agli adulti), ed eventualmente in funzione degli altri Sacramenti ivi riportati, Cresima e Confessione (somministrati a partire dai 7-8 e fino ai 13-14 anni).

Nei R.D. appare comprensibile la mancanza di alcuni atti di

decesso di bambini, quando la morte avviene immediatamente dopo la nascita, in modo tale che l'atto di Battesimo funge, con qualche modifica nella forma, anche da atto di morte. Più frequente sembra il caso di registri di morte riservati ai bambini, separati da quelli degli adulti e di questi assai meno importanti, perché per i bambini mai sarebbe capitato il caso di dover certificare la morte nel caso di seconde nozze del vedovo, né di verificare l'esistenza di legami con il padrino o la madrina sufficienti ad impedire il matrimonio. Il registro di morte dei bambini non ha ovviamente una consultazione così frequente come quello degli adulti, e, una volta finito un registro, può essere conservato non con la stessa cura, o anzi può anche non essere conservato, proprio perché non servirà più.

Per quanto riguarda invece l'*accuratezza*, in particolare quella sull'età, bisogna precisare che non solo l'età esatta, ma una descrizione precisa del soggetto da indicare non era, credo, neppure richiesta. Ad esempio, negli S.A. era sufficiente che una persona fosse identificata, senza possibilità di confusione con un'altra: pertanto, se due donne adulte vivevano nella stessa famiglia, bastava - al limite - indicare un'età approssimata tale da far capire quale fosse la madre e quale la figlia, o, con riferimento al capo famiglia, quale fosse la moglie e quale la madre.

In modo analogo, nel R.D. bastava che una persona deceduta fosse identificata chiaramente, e pertanto era sempre presente il riferimento alla paternità, o comunque, se femmina, ad un uomo (padre, marito, marito defunto), insieme all'età di morte - anzi prima ancora di questa -, che pertanto non è necessario riportare con esattezza estrema.¹

¹ Nella *Forma Describendi Defunctos juxta Rituale Romanu^m Pauli V Pont. Max. jussu Editu^m*, riportata all'inizio del primo registro dei morti di Adria (1615-1674), si trova scritto: "*N. filius N. ex loco N. Aetatis N. (si haec sciri possunt), in Domo N. in comunione Sancta Matris Ecclesiae animam Deo redidit*", ecc.

2. Gli errori di copertura

Errori di copertura possono essere difficilmente misurati e corretti disponendo della sola fonte in esame. Si può giungere, tutt'al più, a qualche stima di massima, suggerita dall'esperienza, esaminando le distribuzioni per età dei censiti o dei deceduti, presentando queste un profilo abbastanza tipico in popolazioni storiche, caratterizzate da elevata natalità e mortalità. Si può pertanto dare un primo giudizio sulla copertura, in particolare sull'omissione più o meno pronunciata di atti di bambini, quando le relative frequenze sono troppo basse, e sempre che qualche causa eccezionale non abbia modificato la struttura per età della popolazione o dei decessi.²

Il confronto di più fonti, nominativo o aggregato, può invece essere di maggiore utilità per valutare la mancanza di bambini.

(a) *Stati delle Anime*. Un confronto può essere effettuato con i dati del Registro dei Battesimi, unitamente a quello dei Defunti, degli ultimi anni (dagli ultimi due a, poniamo, gli ultimi cinque). Nominativamente, si guarderà se i Battezzati non ancora deceduti al tempo della compilazione dello S.A. siano in questo effettivamente elencati. C'è il pericolo di una distorsione a causa delle migrazioni, dal momento che se un bambino è emigrato, non si trova nello S.A. neppure la sua famiglia, ma se è immigrato non si trova nei registri parrocchiali il suo atto di battesimo. Usando invece dati aggregati, occorrerà vedere se l'ammontare complessivo di una generazione di nati di un certo anno, diminuito dei morti negli anni e alle età corrispondenti, si ritrova, in età corrispondente, nello S.A. Le difficoltà di questa procedura sono numerose: tra queste, quella di

² Occorre porre attenzione al fatto che nonostante alcune ipotesi possano essere avanzate, non sappiamo in realtà quale fosse l'uso prevalente nel modo di indicare l'età: anni compiuti, anni iniziati, compleanno più vicino, o altro ancora. Fosse il compleanno più vicino, certe presunte sottostime dell'età 0 non sussisterebbero, dal momento che occorrerebbe slittare tutte le età di mezzo anno.

identificare tra i morti bambini quelli di una certa generazione (possibile solo se le prime età di morte sono indicate almeno in mesi); quella di poter precisare una data, per quanto indicativa, di compilazione dello S.A.; quella del possibile effetto distorsivo delle migrazioni.

(b) *Registro dei Defunti*. Un confronto nominativo sembra possibile solo in presenza di due S.A. ravvicinati, uno prima e uno dopo l'anno del R.D. (e purché, beninteso, gli S.A. siano completi e accurati). Con dati aggregati, un controllo può essere effettuato a partire dal Registro dei battesimi: applicando ai nati di un anno (o alla media dei nati di due anni adiacenti) opportune probabilità di morte, si dovrebbe ottenere un ordine di grandezza di morti in età 0, 1, ecc. comparabile con quello del R.D.³

3. Gli errori di accuratezza

E' all'età che si fa di solito riferimento, per le due fonti in questione, dal momento che per le altre variabili riportate, può, con frequenza diversa (raramente per il sesso, più spesso per lo stato matrimoniale, o per quello professionale), mancare l'indicazione, ma è improbabile che essa sia errata o imprecisa. Il fenomeno più frequente è la tendenza all'arrotondamento nella dichiarazione delle età, sebbene non sia da escludere anche una tendenza a preferire certe cifre più di altre.

Un primo controllo della precisione delle età riportate è la distribuzione secondo la cifra terminale: conviene, per entrambe le fonti, escludere da questa verifica le età iniziali (sotto i 6-8 anni, ad esempio, o anche meno), e forse anche quelle terminali (oltre i 70-80). Se si effettua il conto su un numero di età multiplo di dieci (ad

³ La difficoltà maggiore, in questo caso, è ovviamente la scelta della tavola di mortalità da cui trarre le probabilità di morte.

esempio, quelle comprese tra gli 8 e i 77 anni), le frequenze di ciascuna cifra terminale non dovrebbero allontanarsi molto dal 10%.⁴

Se la cifra "0", ed eventualmente anche "5" o altre cifre, hanno frequenze sistematicamente superiori, ciò non significa che i dati siano del tutto inaffidabili. Il più delle volte c'è una attrazione pressoché uguale di cifre superiori ed inferiori, sì che l'errore *in media* è piccolo, se non nullo. In questi casi, mentre sembra sconsigliabile usare una classificazione delle età in cui le cifre "preferite" siano piazzate sulla soglia della classe (come le consuete 10-14, 15-19, ecc.), sarà preferibile costruire classi di età in modo che la cifra tonda sia al centro della classe (ad esempio, 8-12, 13-17, 18-22, 23-27, ecc.): le frequenze ricadrebbero così comunque nella classe giusta. L'analisi successiva (ad esempio, costruzione di piramidi delle età, o di tavole di mortalità abbreviate), fatta con queste classi, non dovrebbe soffrirne molto.

L'ipotesi che l'attrazione sia "bilanciata" è però verificabile solo con dati nominativi, incrociando informazioni sugli atti di battesimo - la cui data è certa - con quelle degli S.A. o dei R.D.: ma servirebbero atti di battesimo di almeno un'ottantina d'anni, e comunque con i soliti effetti dei movimenti migratori (mancherebbero i battesimi degli immigrati, penalizzando tra l'altro il controllo dell'età delle mogli).

Ci sono casi tuttavia in cui ai fini dell'analisi successiva sarebbe preferibile avere le età classificate in singoli anni, e non in classi, per giunta non convenzionali. Casi simili si verificano, ad esempio, nell'applicazione del metodo dei *figli propri* per ottenere misure di fecondità per età della madre da dati di S.A., in cui l'uso di classi (ad

⁴ La particolare forma della distribuzione (popolazione per età, o decessi per età) potrebbe portare a distorsioni sistematiche rispetto a tale percentuale attesa. Tuttavia, alcune verifiche hanno mostrato che le differenze dovute a questo motivo sono minime; in ogni caso, esiste un indice ottenuto "ruotando" la cifra iniziale, e facendo la media dei dieci risultati (indice di Myers; per la descrizione del metodo si veda Shryock, Siegel, 1976).

esempio quinquennali) nell'età dei figli non consente di cogliere le variazioni congiunturali della fecondità, mentre nell'età delle donne comporta distorsioni sia nella stima dell'età media al parto che in quella della fecondità totale (Rossi, 1993). Un altro caso è la necessità di costruire tavole di mortalità, anche abbreviate, quando l'uso di classi, anche non convenzionali, porta a risultati non convincenti, quali probabilità di morte sistematicamente più alte nella classe che contiene lo 0 rispetto a quella successiva contenente il 5.⁵ Ancora, il metodo usato da Beltrami (1954), che consente di ricostruire la popolazione totale e per età con soli dati di movimento (in particolare i decessi per età), come pure quello di Lee (1974), che usando dati di stato e di movimento si propone lo stesso obiettivo, potrebbero essere indubbiamente più efficaci con dati corretti. In tutti questi casi, se il controllo nominativo non è possibile, nemmeno a campione⁶, occorre cercare una via per correggere dati evidentemente arrotondati.

4. Una correzione dell'attrazione nella distribuzione delle età

Nel caso in cui non ci siano problemi di copertura e la distribuzione presenti solo evidenti ed improbabili punte in corrispondenza di certe età, una prima soluzione potrebbe essere l'uso di medie mobili per "lisciare" la distribuzione. Ma, a parte altri

⁵ E' vero che i primi risultati della tavola ottenuta con il metodo di Halley, anche nell'estensione di Santini (Santini, Del Panta, 1982), possono essere usati per "entrare" in tavole tipo, ma è anche vero che da più parti sussistono dubbi sulla reale possibilità delle moderne tavole tipo a rappresentare la mortalità in epoca storica.

⁶ A questo scopo, sembra proponibile un eventuale campione dove gli strati entro cui estrarre unità (o grappoli) da controllare siano non sottoperiodi genericamente ottenuti suddividendo l'intero periodo in k parti, ma i periodi corrispondenti alla "gestione" dei successivi parroci, dipendendo spesso da questi, come è noto, la qualità delle registrazioni.

problemi, solo in parte superabili,⁷ la media mobile è un procedimento troppo meccanico, che fa perdere, tra l'altro, la specificità delle situazioni, dovute agli andamenti congiunturali, e questo tanto più quando la serie in esame è riferita a piccoli aggregati, dove la variabilità spiccata è una caratteristica costante.

In sostanza, se si ritiene errata una successione di frequenze in cui i casi per un'età "tonda" (ad esempio per l'età 50), sono quattro o cinque volte più numerosi di quelli delle età vicine (nell'esempio, 49 e 51 e magari anche 48 e 52), è perché appare improbabile un numero di nati così disuguale nelle generazioni corrispondenti. La media mobile liscia tutto, facendo perdere anche la variabilità (vera) dovuta alle oscillazioni nel numero delle nascite.

Conviene allora cercare un meccanismo che non faccia perdere il riferimento, se è noto, all'ammontare delle generazioni corrispondenti. Benché la logica sia molto simile, conviene distinguere il caso di dati di stato da quello relativo a defunti.

4.1. Dati di stato

L'ipotesi su cui si basa il modello è che il numero di persone di età x , PA_x^t , attese in vita al tempo t (che per comodità supponiamo essere la fine dell'anno), sia la risultante dell'ammontare di nascite $t-x$ anni prima, N^{t-x} , moltiplicato per le componenti che l'hanno fatto variare, cioè mortalità, migratorietà, oltre a una componente casuale:

$$PA_x^t = N^{t-x} \cdot s_x \cdot r_x + E \quad [1]$$

dove s_x è la probabilità che un nato x anni prima sia in vita all'età x , ed è esprimibile come $s_x = L_x/l_0$; r_x è un coefficiente di "permanenza"

⁷ Tra questi, la scelta della ponderazione, il trattamento delle età iniziali e terminali della distribuzione, il fatto che la somma delle frequenze dopo il calcolo della media mobile può essere diverso da quella della serie originale.

determinato dalla migratorietà (o meglio il suo complemento) tra la nascita e l'età x (e sarà un valore maggiore o minore di 1, a seconda che la migratorietà complessiva tra le età 0 e t abbia saldo attivo o negativo); E rappresenta l'errore casuale. Con quest'ultimo si intende esprimere il fatto che, anche se s_x ed r_x rappresentano i valori medi della sopravvivenza e della permanenza dalla nascita all'età x , il numero reale delle persone in vita all'età x può differire da quello calcolato in base a questi parametri (se non altro per l'arrotondamento).⁸

Normalmente PA_x^t risulterà diverso dal corrispondente valore rilevato P_x^t , per effetto degli errori nella dichiarazione delle età, e di altri fattori casuali, che per ora trascuriamo. Si tratta dunque di correggere i valori P_x^t in base a quelli ottenuti. Il vantaggio rispetto ad un normale lisciamento con medie mobili consiste nel fatto che si tiene conto dell'ammontare iniziale delle varie generazioni, e della sua variabilità, che verrebbe altrimenti annullata. Per la correzione si possono seguire diverse vie:

(a) trovare mortalità e migratorietà tali che

$$\sum_x PA_x^t = P^t = \sum_x P_x^t$$

per x estesa a tutte le età;

(b) trovare mortalità e migratorietà tali che

$$\sum_x PA_x^t = \sum_x P_x^t$$

per alcune grandi classi di età, con $x = a, \dots, b$, essendo a, b i limiti di tali classi (ad esempio, 0-7, 8-47, 48-w);

⁸ La componente dovuta all'errore casuale potrà essere una Normale con media = 0 e varianza = σ^2 ; alternativamente, nella [1] l'errore E può essere espresso con un termine moltiplicativo, con E tale che $\log E \sim N(0, \sigma^2)$. E potrebbe essere definita eventualmente anche in funzione del tempo t e dell'età x .

(c) trovare mortalità e migratorietà tali che

$$\sum_x PA_x^t = \sum_x P_x^t$$

per numerose classi di età (decennali o quinquennali), con $x = \alpha, \dots, \beta$, essendo α, β i limiti di tali classi (ad esempio, decennali, tranne la prima: 0-7, 8-17, 18-27, ecc.).

Le differenze fra le tre vie consistono nella fiducia riposta nella veridicità delle frequenze ottenute sui dati rilevati in classi più o meno ampie: nel primo caso si ritiene affidabile solo il totale e non la distribuzione per età, nel secondo anche una divisione per ampie classi, nel terzo si ritiene affidabile una divisione in classi più ristrette, anche se vengono evitate le cifre tonde alla soglia delle classi.

Supponendo nulla la migratorietà, si tratterà di trovare, per ogni classe di età, una tavola di mortalità tale che, mediante la [1], sia soddisfatta una delle tre relazioni precedenti. Se si adotta la prima, si troverà un'unica tavola di mortalità, con la seconda due o tre, con la terza, invece, ce ne saranno parecchie. In ogni caso, si suppone che la mortalità in ogni gruppo di generazioni che forma una classe di età sia la stessa (nel senso di "estratta dalla stessa tavola": le età sono diverse!) nel corso della loro vita fino al tempo t . Anche nel caso che tale arco di tempo comprenda una (o più) crisi di mortalità, l'effetto si suppone uguale nelle generazioni appartenenti alla stessa classe di età. Tali ipotesi sembrano più plausibili se la classe di età è ristretta; ma non potendo le classi essere inferiori a 5 o 10 anni, occorrerà riporre particolare cura nella determinazione della loro ampiezza e delle loro soglie.

4.2. Alcune precisazioni operative

Per l'individuazione della tavola di mortalità in ognuna delle classi

di età, si può partire da tavole tipo, con una certa preferenza per le più note, quelle di Coale e Demeny (1966, 1983); queste si presentano suddivise in quattro "famiglie", distinte secondo diversi profili di mortalità, e di ogni famiglia sono tabulate tavole (maschili e femminili) per 24 "livelli" (con vita media femminile alla nascita da 20 a 77,5 anni, con passo 2,5).

Scelta una famiglia, l'individuazione della tavola si può ottenere interpolando le probabilità di morte q_x dei livelli più bassi. E' dimostrato infatti che le variazioni delle q_x in tavole a mortalità alta - indicativamente, livelli da 1 a 10, cioè con vita media compresa tra 20 e 42.5 anni - sono pressoché lineari (Coale, 1972).

E' da rilevare però il fatto che le q_x sono lineari, non invece L_x (né d_x , che servirà nel seguito), per cui occorrerà esprimere tali funzioni in termini di q_x . Considerando solo l'effetto della sopravvivenza, per la popolazione per età si ha, per ciascuna classe di età

$$\sum_x P A_x^t = \sum_x N^{t-x} \cdot s_x = \sum_x P_x^t$$

e, considerando l'uguaglianza tra i due ultimi termini

$$\begin{aligned} \sum_x N^{t-x} \cdot s_x &= \sum_x N^{t-x} \cdot L_x / l_0 = \\ \sum_x N^{t-x} \cdot \Pi_i (1-q'_i) &= \sum_x P_x^t \end{aligned} \quad [2]$$

per $i = -1, \dots, x-1$, avendo indicato con q'_i le probabilità "prospettive" di morte e con q'_{-1} una probabilità simile, ma riferita ai nati.⁹

⁹ Per esteso, si ha

$$L_x / l_0 = L_0 / l_0 \cdot L_1 / L_0 \cdot \dots \cdot L_x / L_{x-1} \Pi_1 = (1-q'_n)_{-1} (1-q'_i)$$

per i estesa da 0 a $x-1$, e avendo indicato con q'_i le probabilità prospettive di morte, i cui complementi all'unità sono

L'interpolazione delle probabilità di morte in una stessa famiglia si esegue in modo analogo a quello usato in altre occasioni (Lee, 1974): prese le q'_x di due tavole (I e II), e la loro differenza A_x , si determina la frazione k di A_x tale che le q'_x ottenute

$$q'_x = q'_x(\text{II}) + k \cdot A_x \quad [3]$$

soddisfino, per ognuna delle classi di età in cui la popolazione è stata suddivisa, l'equazione [2].

4.3. Dati sui decessi

Il problema, come pure la soluzione proposta, è molto simile al caso precedente. L'ipotesi sottostante è che il numero di decessi attesi in età x nell'anno t , DA_x^t , sia la risultante dell'ammontare di nascite $t-x$ anni prima, N^{t-x} , moltiplicato per le componenti che concorrono a determinarlo, cioè mortalità e migratorietà, oltre a una componente casuale:

$$DA_x^t = N^{t-x} \cdot d_x \cdot e_x + E \quad [4]$$

dove d_x è la probabilità che un nato x anni prima muoia proprio all'età x , ed è esprimibile come $d_x / L_x \cdot L_x / l_0 = d_x / l_0$,¹⁰ e_x è un coefficiente di migratorietà relativo all'età x per un nato x anni prima; E rappresenta l'errore casuale.

Anche in questo caso, DA_x^t risulterà normalmente diverso dal corrispondente valore rilevato D_x^t , per effetto degli errori nella

$$1-q'_x = (L_{x+1}) / L_x$$

$$1-q'_n = 1-q'_{-1} = L_0 / l_0$$

¹⁰ In realtà, il rapporto d_x / l_0 potrebbe essere ottenuto anche con un L'_x diverso da L_x . Per la soluzione del problema si veda più avanti.

dichiarazione delle età, e di altri fattori casuali. Si procederà quindi alla correzione dei valori D_x^t in base a quelli ottenuti. Il vantaggio rispetto al consueto lisciamento con medie mobili consiste ancora nel fatto che qui si tiene conto dell'ammontare iniziale delle varie generazioni, e della sua variabilità. Per la correzione si possono seguire anche in questo caso diverse strade, che consistono comunque nel determinare mortalità e migratorietà tali che

$$\sum_x DA_x^t = \sum_x D_x^t$$

per x esteso a tutte le età, o ad alcune grandi classi, o a classi decennali o quinquennali, a seconda della diversa fiducia riposta nell'esattezza del solo dato totale, o dei dati raggruppati in poche, o in molte classi.

Nel caso di migratorietà nulla, il problema si riduce alla scelta, per ogni classe di età, di una tavola di mortalità tale che mediante la [4] sia soddisfatta una delle tre relazioni. Si troverà un'unica tavola di mortalità, o due-tre, o parecchie, rispettivamente, se si adotta la prima, la seconda o la terza soluzione. Anche in questo caso si suppone comunque che la mortalità in ogni gruppo di generazioni che forma una classe di età sia la stessa (estratta dalla stessa tavola) nel corso della loro vita fino al tempo t . Se tale arco di tempo comprendesse una (o più) crisi di mortalità, bisognerebbe supporre uguale l'effetto nelle generazioni appartenenti alla stessa classe di età. Tali ipotesi sembrano più plausibili con classi di età ristrette; ma non potendo le classi essere inferiori a 5 o 10 anni, occorrerà una particolare cura nella determinazione della loro ampiezza e delle loro soglie.¹¹

¹¹ Di minore importanza pratica, anche se rilevante dal punto di vista formale, appare la questione dei decessi per età D_x^t provenienti da due generazioni contigue, e non da una sola. Se la differenza tra i due ammontari di nascite è rilevante, occorrerebbe attribuire i deceduti D_x^t a una media delle due generazioni di nati.

A causa della non linearità della funzione d_x occorrerà esprimere l'equazione in termini di q_x . Considerando gli effetti della sola mortalità, si ha, per ciascuna classe di età

$$\sum_x DA_x^t = \sum_x N^{t-x} \cdot d_x = \sum_x D_x^t$$

Considerando l'uguaglianza tra i due ultimi termini, si ha

$$\sum_x N^{t-x} \cdot d_x = \sum_x N^{t-x} \cdot d'_x / L_x \cdot L_x / l_0 =$$

$$\sum_x N^{t-x} \cdot q'_x \cdot \Pi_i (1-q'_i) = \sum_x D_x^t \quad [5]$$

per $i = -1, \dots, x-1$, dove i simboli hanno lo stesso significato già visto.¹²

5. Commenti finali

Questa nota non è molto più di una riflessione ad alta voce (anzi, davanti alla tastiera di un computer), su cui bisognerà, se la strada si rivelasse buona, lavorare ancora. Vorrei però evidenziare da un lato alcuni vantaggi di questa procedura e dall'altro alcuni inconvenienti già individuati.

I presunti pregi: (a) non ci si vincola, a meno di seguire la prima variante che considera una unica classe di età, ad una sola tavola di mortalità, ma si prendono segmenti di più tavole, e questo potrebbe essere da un lato una precauzione se si ritiene che la mortalità di tavole tipo mal si adatti al profilo di mortalità di epoche storiche,

¹² Per esteso, si ha

$$d'_x / l_0 = d'_x / L_x \cdot L_x / l_0 = q'_x \cdot \Pi_i (1-q'_i)$$

per i estesa da -1 a $x-1$, e avendo dato ai simboli il significato già usato in precedenza.

dall'altro una verifica di tale adattamento (quante tavole: una - o due, ma di livello contiguo - o invece diverse, sono necessarie per descrivere la mortalità della popolazione in esame); (b) si elimina la variabilità dovuta all'attrazione delle età; (c) si recupera invece la variabilità dovuta all'oscillazione del numero di nati, tipica della particolare popolazione in esame.

Gli inconvenienti fin d'ora prevedibili: (a) si perde la variabilità dovuta al caso, che è tanto più importante quanto è ridotta la popolazione in esame; (b) questo a maggior ragione se si tratta della distribuzione di decessi, che sono il 3-4-5 % della popolazione; (c) non ho ancora provato questi modelli su dati concreti: qualche problema di ordine matematico o informatico dovrebbe essere superabile¹³; (d) ho il timore però che, per il caso di dati sui decessi, seri inconvenienti si presentino a causa della numerosità osservata, che in ogni caso è molto bassa. Penso al numero annuo di defunti in una parrocchia di dimensione anche non piccolissima: in una località di 1000 abitanti, ad esempio, le sepolture sono valutabili in numero da 30 a 50, delle quali da un terzo a metà in età inferiore al primo o ai primissimi anni di vita. I rimanenti 15-25 sarebbero da suddividere tra le altre età, almeno 70 anni di vita. Con dati di questo ordine di grandezza, le vie da seguire nel caso di gravi e provate scorrettezze nella distribuzione delle età alla morte potrebbero essere altre, in cui, ad esempio, diventi importante non solo il modello di mortalità di riferimento, ma anche l'effetto casuale, che produce, come è noto, conseguenze di rilievo sui piccoli numeri.

Quest'ultimo punto, delle variazioni per gli effetti del caso, qui

¹³ Il dott. Alessandro Rosina mi fa notare che quando le classi di età siano ampie, ci potrebbero essere dei problemi nella soluzione delle equazioni [2] e [5] mediante la [3], che risulterebbero di grado n rispetto a k , con n pari al numero delle età comprese nella classe. Con $n=2$, una sola della soluzioni dovrebbe essere plausibile (cioè k compreso tra 0 e 1); con $n=3$ o più, le soluzioni plausibili potrebbero essere diverse (ma normalmente, in problemi di questo genere, lo è una sola).

appena accennato, necessita di essere approfondito. Ad esempio, effetti casuali nella determinazione del numero dei decessi (o della popolazione) possono essere introdotti, sia nel semplice aggiustamento per ottenere cifre intere, sia come ulteriore passaggio, sottoponendo l'ammontare di riferimento (popolazione o decessi) a un "bombardamento" con stimoli casuali (ma occorrerebbe almeno conoscere il modello da utilizzare: lo s.q.m., se si pensa ad una Normale con media nulla).

Bibliografia

Bellettini A. (s.d. ma 1973), Gli 'Status Animarum': caratteristiche e problemi di utilizzazione nelle ricerche di demografia storica, in C.I.S.D.S., *Le fonti della demografia storica in Italia*, CISP, Roma.

Beltrami D. (1954), *Storia della popolazione di Venezia dalla fine del secolo XVI alla caduta della Repubblica*, CEDAM, Padova.

Coale A. J. (1972), *The Growth and Structure of Human Populations*, Princeton University Press, Princeton.

Coale A. J., Demeny P. (1966), *Regional Model Life Tables and Stable Populations*, Princeton University Press, Princeton.

Coale A. J., Demeny P., Vaughan S. (1983), *Regional Model Life Tables and Stable Populations, 2nd edition*, Academic Press, New York.

Lee R. D. (1974), Estimating Series of Vital Rates and Age Structure from Baptism and Burials: A New Technique, with Application to Pre-industrial England, *Population Studies*, 28, 3.

Rossi F. (1992), "L'uso del metodo dei figli propri in demografia storica", *SIDeS, Bollettino di Demografia Storica*, 17.

Santini A., Del Panta L. (1982), *Problemi di analisi delle popolazioni del passato in assenza di dati completi*, CLUEB, Bologna.

Shryock H. S., Siegel J. S. (1976), *The Methods and Materials of Demography*, Academic Press, London.

ALESSANDRO ROSINA

Una generalizzazione dell'Inverse Projection

1. L'Inverse Projection

Nelle ricostruzioni delle popolazioni in epoca storica, la tecnica che in questi ultimi anni ha riscosso il maggior consenso ed ha dato i contributi più fecondi è l'inverse projection.

Ideata da R. Lee all'inizio degli anni settanta, e da allora continuamente perfezionata, l'inverse projection ha il pregio di essere una tecnica semplice da applicare e molto versatile, in grado di adattarsi alle diverse esigenze dei dati disponibili in epoca storica, permettendo di stimare misure demografiche molto raffinate, quali la speranza di vita ed i tassi di riproduttività.

L'importanza e la proficuità di tale procedura è testimoniata dalle sue numerose e recenti applicazioni, tra cui per l'Italia è da ricordare quella che ha permesso a M. Breschi (1990) di ricostruire le vicende demografiche della Toscana, "primo tentativo, in Italia, di delineare un modello di crescita regionale".

La logica di funzionamento della tecnica di Lee (1974 e 1985) è abbastanza semplice, e consiste nel cercare di individuare la popolazione sottostante che ha generato le serie storiche osservate.

Nel caso di popolazione chiusa (saldo migratorio nullo), a partire dalla distribuzione iniziale si ottiene la popolazione all'istante temporale successivo aggiungendo i nati sopravvissuti e detraendo da ogni classe d'età i rispettivi decessi. I decessi distinti per età vengono stimati ricorrendo alle tavole tipo di mortalità. All'interno di una "famiglia" di tavole, verrà scelta quella le cui probabilità prospettive

di morte, applicate alla popolazione iniziale, daranno proprio i decessi totali osservati.

Una volta stimata la distribuzione per età della popolazione è possibile poi stimare tutte le più importanti misure di mortalità e fecondità.

2. Una estensione dell'inverse projection con mortalità differenziata per età

Il limite fondamentale della procedura di Lee è quello di assumere un modello costante di mortalità. Basandosi infatti sul solo ammontare totale dei decessi e su una famiglia di tavole tipo, la curva di mortalità stimata dall'inverse projection risulta avere sempre stesso andamento, spostandosi semplicemente verso l'alto o verso il basso a seconda dell'entità dei decessi osservati. Ciò rende la procedura particolarmente inadeguata negli anni di crisi, generando risultati tanto più distorti quanto più è marcata la selettività della crisi nelle diverse classi d'età¹.

Una risposta a tale limite viene suggerita dalla ricchezza delle informazioni disponibili nelle fonti storico-demografiche italiane. L'idea è quella di riuscire a generalizzare la procedura di Lee in modo da farle tener conto dell'ulteriore informazione sulla struttura per età dei decessi. Ciò consentirebbe di far variare non solo il livello della curva di mortalità, ma anche l'inclinazione. In tal modo le tavole tipo, che in genere mal si prestano ad interpretare la mortalità in epoca

¹ Vari aggiustamenti alla procedura di Lee sono stati suggeriti in questi ultimi anni per superare tale limite. "E' possibile utilizzare apposite tavole di mortalità da usare negli anni in cui si è individuata una crisi" (Lee, 1991). Purtroppo le crisi di mortalità sono un fenomeno molto complesso e vario, non di sempre agevole individuazione, e difficilmente imbrigliabile in schemi predefiniti.

storica, verrebbero adattate alla situazione reale e contingente della popolazione in esame, permettendo altresì di catturare l'eventuale incidenza differenziata delle crisi nelle diverse classi d'età.

Nelle fonti storico-demografiche si trovano spesso segnate le età alla morte, ma in modo molto approssimato, con forti concentrazioni nelle età terminanti in zero, ed una non trascurabile percentuale di decessi ad età ignota.

Un modo per utilizzare tale informazione, senza operare stime arbitrarie, è quello di raggruppare i decessi in grandi classi di età. L'operazione più semplice consiste nel considerare due classi di decessi, una giovanile ed una adulta. Disponendo dei decessi totali distinti in decessi giovanili e decessi adulti la versione estesa dell'inverse projection, chiamata inverse projection differenziata, andrà a cercare due tavole di mortalità: la prima le cui probabilità di morte applicate alle età giovanili diano proprio i decessi giovanili osservati, la seconda le cui probabilità di morte applicate alle età adulte diano proprio i decessi adulti osservati. Alla popolazione iniziale verranno così applicati due livelli di mortalità, permettendo di catturare l'eventuale comportamento differenziato delle prime classi d'età rispetto alle altre.

La generalizzazione della procedura di Lee consiste quindi nel considerare come input, invece che solamente i decessi totali, un numero G di classi di decessi a cui corrisponderanno G diverse tavole di mortalità da applicare alla popolazione.

E' chiaro che, se la qualità dei dati di base lo consente, si può arrivare al massimo dettaglio di disaggregazione dei decessi totali fornendo come input dell'inverse projection differenziata i decessi in ogni classe annuale d'età. In tal caso la distribuzione per età dei decessi viene fornita completamente come input, annullando il ruolo delle tavole-tipo. Per $G=1$ invece l'inverse projection differenziata si ridurrà all'inverse projection di Lee, e la distribuzione per età dei decessi verrà effettuata ricorrendo completamente alle tavole-tipo.

Per G compreso tra questi due estremi, la distribuzione per età dei decessi, e quindi la serie delle probabilità di morte da applicare alla popolazione, risulta dalla combinazione delle informazioni sulla struttura dei decessi fornite come input, e le tavole-tipo.

Per saggiare le potenzialità della versione differenziata ed il miglioramento delle stime rispetto alla versione originaria di Lee, sono state effettuate le seguenti applicazioni, di cui riporto succintamente i risultati più importanti.

3. Applicazione dimostrativa sui dati della Toscana

Viene presentata qui di seguito una applicazione dimostrativa dell'inverse projection differenziata sui dati utilizzati da M. Breschi (1990) per la ricostruzione della popolazione della Toscana. Più precisamente si è fatto riferimento alla crisi di mortalità verificatasi nel quinquennio 1815-19.

Per tale crisi si sono ipotizzati due diversi scenari. Nel primo viene simulata una distribuzione dei decessi per età caratteristica di una crisi che colpisce soprattutto le classi giovanili (ad esempio una epidemia di vaiolo). Nel secondo una crisi che colpisce prevalentemente le classi adulte (ad esempio una epidemia di tifo)².

Disponendo quindi delle distribuzioni per età dei decessi possiamo calcolare, per ognuno dei due scenari, la "vera" curva di mortalità, le cui probabilità prospettive di morte vengono ottenute direttamente

² La distribuzione per età dei decessi è stata effettuata ipotizzando che i 231990 decessi totali, verificatisi nel quinquennio 1815-19, possano essere scomposti in 180000 decessi base, ovvero morti non causati dalla crisi, ed un surplus di 51990 decessi imputabili all'epidemia. I decessi base sono stati poi distribuiti per età ricorrendo alle tavole tipo di Coale e Demeny (1966), mentre per la ripartizione per età del surplus, in base al tipo di epidemia, ci si è ispirati agli esempi riportati da Del Panta (1980).

dal rapporto tra i decessi di una data classe d'età e la popolazione corrispondente.

Per ognuno dei due scenari è quindi possibile mettere a confronto la "vera" curva di mortalità, e le corrispondenti stime della speranza di vita alla nascita e dei tassi di riproduttività, con la curva e le stime prodotte dall'inverse projection e della versione differenziata.

Per l'applicazione dell'inverse projection differenziata è stato adottato un valore di $G=3$ (la prima classe contiene i decessi fino ai 4 anni, la seconda dai 5 ai 49 anni, la terza oltre i 50 anni).

I risultati sono riportati nei prospetti 1 e 2. La curva di mortalità stimata dall'inverse projection differenziata risulta molto più vicina alla curva "vera" rispetto a quella stimata dall'inverse projection. Ciò si ripercuote sulle stime delle misure di mortalità e fecondità: fatta 100 la differenza tra i valori stimati dall'inverse projection e quelli veri, il miglioramento delle stime prodotte dalla versione differenziata risulta superiore all'80%.

Prospetto 1 : Primo scenario (crisi alle età giovanili)

	"VERI"	IP	IPD
e_0	25.39	27.53	25.51
GRR	2.122	2.151	2.127
NRR	0.953	0.998	0.962

Prospetto 2 : Secondo scenario (crisi età adulte)

	"VERI"	IP	IPD
e_0	31.09	27.53	30.55
GRR	2.181	2.151	2.180
NRR	1.098	0.998	1.053

4. Applicazione alla Podesteria di Murano

I risultati che si ottengono dall'applicazione dell'inverse projection alla Podesteria di Murano (dal 1665 al 1784) si inseriscono coerentemente nel contesto dell'evoluzione demografica di Venezia e della terraferma veneta, che vede nell'ultimo secolo di vita della Serenissima un invecchiamento della popolazione ed un innalzamento della mortalità, specie infantile (Beltrami 1954, Rossi 1970).

L'inverse projection differenziata (con $G=3$: la prima con i decessi fino ai 4 anni, la seconda dai 5 ai 59, la terza oltre i 60) conferma l'invecchiamento della popolazione, accentuandolo, rendendolo più coerente con i risultati delle "Anagrafi" veneziane che indicano alla fine del settecento una popolazione di età superiore ai 60 anni del 10%, rispetto al 7% stimato dall'inverse projection. Ne consegue che l'età media della popolazione indicata dalla versione differenziata risulta alla fine del periodo di applicazione superiore ai 31 anni, inferiore ai 29 anni per l'inverse projection.

Ancora più interessanti sono le differenze per quanto riguarda l'innalzamento della mortalità. La speranza di vita alla nascita diminuisce di oltre cinque anni nel corso di tutto il secolo XVIII, sia nella curva stimata dall'inverse projection che in quella stimata dalla versione differenziata. La curva della mortalità infantile risulta invece avere una crescita molto più accentuata nella versione differenziata, mentre la speranza di vita oltre i cinque anni d'età risulta tendenzialmente costante. La versione differenziata quindi si "accorge" che l'aumento di mortalità avvenuto a Murano nel secolo XVIII è da attribuire quasi esclusivamente all'aumento della mortalità infantile, ciò non accade per l'inverse projection che fa scontare l'aumento della mortalità proporzionalmente a tutte le età.

5. Ulteriori informazioni sull'inverse projection differenziata

Una descrizione più dettagliata dell'inverse projection differenziata e dell'applicazione dimostrativa sui dati della Toscana si può trovare in un articolo di prossima pubblicazione su "Statistica".

Segnalo infine che è in via di stesura un manuale di applicazione dell'inverse projection e della sua versione differenziata, con allegato un programma informatico per la ricostruzione aggregativa delle popolazioni con l'inverse projection differenziata (da cui l'inverse projection di Lee può essere ottenuta come caso particolare in assenza di dati sulle età alla morte).

Bibliografia

D. Beltrami (1954), *Storia della popolazione di Venezia dalla fine del secolo XVI alla caduta della Repubblica*, CEDAM, Padova.

M. Breschi (1989), Due tecniche di ricostruzione aggregativa: l'inverse projection e la back projection, *Bollettino di Demografia Storica*, 8.

M. Breschi (1990), *La popolazione della Toscana dal 1640 al 1940. Un'ipotesi di ricostruzione*, Università di Firenze, Dipartimento statistico, Firenze.

A. J. Coale e P. Demeny (1966), *Regional Model Life Tables and Stable Population*, Princeton University Press, Princeton.

L. Del Panta (1980), *Le epidemie nella storia demografica italiana (secoli XVI-XIX)*, Loescher editore, Torino.

R. D. Lee (1974), Estimating Series of Vital Rates and Age Structure from Baptism and Burials: A New Technique, *Population Studies*, v. 28, n. 3.

R. D. Lee (1985), Inverse Projection and Back Projection: A Critical Appraisal and Comparative Result for England, 1539-1871, *Population Studies*, v. 39, n. 2.

R. D. Lee (1991), *Methods and Models in Macro-Demographic History: An Update and Assessment*, Seminar on "Old and New Methods in Historical Demography", Palma de Mallorca, Balears, Spain, 12-15 June 1991.

A. Rosina (1991-92), *Una generalizzazione dell'inverse projection: teoria ed applicazioni*, Tesi di Laurea in Scienze Statistiche e Demografiche, Facoltà di Scienze Statistiche Demografiche ed Attuariali, Università di Padova, relatore F. Rossi.

F. Rossi (1970), Storia della popolazione di Adria dal XVI al XIX secolo, *Genus*, vol. XXVI, n.1-2.

GUSTAVO DE SANTIS

UNA MODELLIZZAZIONE DEGLI EFFETTI DELLA MORTALITÀ SULLE STIME DI FECONDITÀ CON IL METODO DEI FIGLI PROPRI

Introduzione

In questa comunicazione si presentano i risultati di alcune simulazioni volte a quantificare l'effetto prodotto dall'azione della mortalità sulle stime di fecondità ricavabili da dati di stato, con il metodo dei figli propri. Nell'ambito di questo metodo (Cho, Retherford e Choe, 1986), si fa riferimento in particolare alla versione proposta da Breschi e De Santis (1992; in preparazione), cui si rimanda anche per tutti gli approfondimenti dei temi qui trattati. Del metodo e della versione adottata, le linee essenziali ai fini della comprensione di quanto segue sono esposte nella sezione 1, mentre il resto della comunicazione è dedicato alla presentazione dei risultati.

1. Il metodo dei figli propri e il problema della mortalità

Ogni fonte di stato che specifichi età e sesso degli individui rilevati consente di calcolare il cd. indice di carico di figli per donna, tramite il rapporto tra bambini (B) e donne (D) in particolari (e convenzionali) classi di età

$$ic = B_{0-4}/D_{15-49}$$

Se poi la fonte specifica anche le relazioni di parentela, e consente quindi di determinare l'età delle madri dei bambini, l'indice può essere reso più specifico, sia al numeratore (si possono considerare solo i bambini di età t , e tra questi solo quelli nati da donne di età x , $B_{t,(x)}$), sia al denominatore (dove si considereranno solo le donne di età x , D_x), per ottenere

$$c_{t,x} = B_{t,(x)} / D_x$$

L'indice $c_{t,x}$ è in qualche rapporto con il tasso di fecondità f_{x-t} , cioè con la fecondità manifestata t anni prima (l'anno di nascita dei bambini¹) da donne che avevano allora $x-t$ anni, perché in entrambe le misure entrano gli effettivi delle stesse due generazioni, sia pure considerati in momenti diversi della loro vita: al numeratore si trovano sempre e soltanto alcuni dei membri della coorte dei nati nell'anno t (quelli generati da donne nate nell'anno x), e al denominatore le donne della generazione x . In generale, le due misure c e f non coincidono, perché nel corso del tempo, gli effettivi delle due generazioni a confronto, *così come risultano dalle fonti*, si modificano con velocità diverse,² per effetto di un complesso di "altre" cause (tra cui movimenti migratori, errori nelle fonti, ecc.), di cui non

¹ Per semplicità si ometterà qui di distinguere tra misure calcolate per età e per anno di nascita, cosa che non porta a gravi inconvenienti, soprattutto se si utilizza un sistema di numerazione convenzionale degli anni di tempo trascorsi, in cui 0 indica i 365 giorni prima della rilevazione, 1 i 366-730 giorni precedenti, ecc.

² E, talvolta, anche in direzioni diverse, come può avvenire, per esempio, se vi sono correnti di immigrazione sufficientemente forti e selettive per età e per sesso.

ci si interesserà in questa sede,³ e della mortalità.

La relazione che si considera è di tipo moltiplicativo

$$c_{t,x} = f_{x-t} h_{t,x}$$

dove il fattore $h_{t,x}$, in generale minore di 1, descrive come, per il solo effetto della mortalità (delle donne della generazione x e dei bambini della generazione t), il rapporto originario f si è trasformato nel tempo, per assumere, al momento dell'indagine, il valore c . Poiché però, come precisato, il problema in questo caso è inverso (si conosce c e si vuol determinare f), risulta più conveniente considerare invece il fattore $k (=h^{-1})$, per cui vale la relazione

$$f_{x-t} = c_{t,x} k_{t,x}$$

Più che i singoli tassi di fecondità f , i demografi ambiscono a determinare il tasso di fecondità totale (TFT)

$${}_tTFT = \sum_x f = \sum c k$$

Su di esso, l'influenza dei singoli fattori di correzione k diventa difficile da misurare, perché varia con i pesi c . Dal punto di vista analitico, quindi, si rivela più conveniente definire prima il "tasso di carico totale" (di figli nati nell'anno t sulle varie coorti di donne che li hanno generati),

$${}_tTCT = \sum_x c$$

e poi un opportuno fattore di correzione K ,⁴ tale per cui

³ Per semplicità si assumerà dunque che le eventuali distorsioni avvengano nello stesso senso e, approssimativamente, con la stessa intensità per i membri delle due generazioni, t e x .

$${}_i TFT = {}_i TCT \cdot K$$

Il metodo dei figli propri, in un certo senso, consiste proprio in questo: partire da una fonte di stato, calcolare i ${}_i TCT$, correggerli con i fattori K e determinare i ${}_i TFT$.

Qui, come accennato, l'interesse è concentrato sul solo ruolo della mortalità, e cioè sui fattori K . In particolare, quattro sono le domande cui si cercherà di fornire una risposta, e precisamente:

1) è proprio necessario tener conto della mortalità? Non potrebbe essere che gli effetti della mortalità dei bambini (numeratore) tendano a annullarsi con quelli della mortalità delle donne (denominatore), lasciando pressoché invariato il risultato del rapporto (cioè $TFT \approx TCT$)?

2) ammettendo che la risposta alla prima domanda sia negativa, quali conseguenze avrebbe un errore nelle ipotesi sul livello della mortalità?

3) quali conseguenze avrebbe invece un errore nelle ipotesi sul tipo di mortalità?

4) è necessario "partire" da e_0 per descrivere la mortalità? Cosa succede se si sceglie invece un altro punto di partenza, in particolare q_0 , che le fonti parrocchiali consentono spesso di stimare con discreta precisione?

⁴ Si verifica che $K = TFT/TCT = \sum f/\sum c = \sum f/\sum (f/k)$, e cioè che K è la media armonica ponderata dei singoli k , con pesi pari ai tassi di fecondità f , che sono ignoti. Il risultato finale, tuttavia, è quasi indipendente dalla serie degli f (per valori coerenti con le esperienze concrete delle popolazioni umane), per cui si può scegliere (quasi) arbitrariamente una curva modello di fecondità e procedere con quella nell'analisi. Qui si è scelta la curva standard di fecondità di Brass (UN, 1983), con età media al parto pari a 29,1 anni.

Per rispondere a queste domande si è utilizzato il programma di "APPLicazione AUTomatizzata di un metodo di Sopravvivenza Inversa" (APPLAUSI), allegato al volume di Breschi e De Santis (in preparazione), che sfrutta curve modello di mortalità di Brass (1971; UN, 1983) e di Coale e Demeny (1983).

2. Le "dimensioni" della mortalità

Prima di procedere, occorre considerare che la mortalità è un fenomeno multidimensionale, di cui, schematicamente, si possono indicare almeno quattro aspetti rilevanti in questo contesto:

1) il *tipo* (che in queste applicazioni può essere lo Standard generale di Brass, oppure uno dei quattro modelli di Coale e Demeny: Ovest, Est, Sud e Nord);

2) il *livello* (sintetizzato nella speranza di vita alla nascita e_0);

3) l'*evoluzione temporale* del fenomeno negli anni precedenti la rilevazione; e

4) il *sexo* della popolazione considerata (perché i coefficienti K risentono della mortalità sia delle donne che dei bambini, tra i quali circa il 50% è di sesso maschile).

Per non complicare eccessivamente la presentazione conviene introdurre alcune semplificazioni: qui si è scelto di ignorare gli ultimi due aspetti, cioè l'evoluzione temporale della mortalità (il che equivale a supporla costante nei 12 anni prima della rilevazione) e la differenza di mortalità tra i sessi (supposte quindi irrilevanti).

3. Effetti del livello della mortalità sul valore dei coefficienti K

Nella Tabella 1 è presentata una serie di coefficienti moltiplicativi K che, come si vede, sono forniti per 13 anni (da 0 a 12 prima dell'istante in cui sono stati rilevati i dati), per cinque modelli di

mortalità (i quattro di Coale e Demeny e lo Standard generale di Brass) e per vari livelli di mortalità (con e_0 variabile da circa 22,5 a circa 62,5 anni). Per il solo modello Ovest, il corrispondente grafico è presentato nella Figura 1.

La tabella fornisce una risposta a tutti e quattro i quesiti elencati poc'anzi. La risposta alla prima domanda è immediata: solo in regimi di bassissima mortalità si può prescindere senza gravi danni dagli effetti della mortalità. Ma anche con una speranza di vita alla nascita di 62,5 anni (impensabile in epoca storica), ignorare la mortalità comporta una sottostima della fecondità non trascurabile, compresa tra il 4 e il 7%, a seconda degli anni. Al crescere della mortalità, gli effetti della sua (eventuale) mancata considerazione si fanno sempre più pesanti: con una speranza di vita alla nascita di 22,5 anni, per esempio, la sottostima arriva a superare, in certi anni, il 70%.

Consideriamo ora il secondo interrogativo (effetti di un'ipotesi errata sul livello della mortalità), cominciando con un esempio. Ammettiamo che, correttamente, si reputi che la mortalità della comunità XY sia del modello Sud, ma che, erroneamente, si stimi che e_0 sia pari a 42,5, invece del vero 37,5. Si commette dunque una sopravvalutazione della sopravvivenza di 5 anni in senso assoluto e del $(5/37,5=)$ 13% in senso relativo. Per l'anno 12 si dovrebbe applicare il coefficiente moltiplicativo 1,3919, mentre invece si applica, erroneamente, il coefficiente 1,3127, che è più basso del 6% circa. Il 6% è quindi l'errore per difetto che si commette nella stima del *TFT*. Per esempio, se il vero *TFT* fosse pari a 5, la stima ottenuta con il metodo dei figli propri porterebbe in questo caso a un valore pari al 94% di 5, cioè 4,7.

La relazione più generale che si osserva tra i coefficienti K e il valore di e_0 è del tipo illustrato nella fig. 2 (in cui si è scelto come esempio l'anno 6) e può essere resa quasi perfettamente lineare con una doppia trasformazione logaritmica (fig. 3). I coefficienti di regressione calcolati su queste trasformate (tab. 2, colonna 2)

forniscono l'elasticità dei K rispetto a e_0 ,⁵ che, come si vede, si mantiene al di sotto di -0,5. Questo significa che l'esempio precedente ha una valenza più generale: un errore di sovrastima di e_0 pari a +2% causa, approssimativamente, un errore di sottostima del *TFT* pari a -1/2%.⁶

4. Effetti del tipo di mortalità sul valore dei coefficienti K

Con l'ausilio della tab. 1 e della fig. 3 si può poi rispondere anche al terzo interrogativo: in generale, è importante scegliere un corretto modello di mortalità, benché non sempre la scelta di un modello errato comporti distorsioni gravi nei risultati. Questo avviene perché vi sono alcuni casi fortunati in cui gli errori per eccesso e per difetto (sulla mortalità dei bambini e delle donne) tendono a compensarsi. In particolare, si possono osservare, per l'anno 6, considerato a mo' di esempio, due gruppi relativamente omogenei: i modelli Sud e Est, che hanno bisogno di fattori di correzione più forti si contrappongono in questo caso ai modelli Nord e Ovest.⁷

Ammettiamo, per esempio, di sapere che la speranza di vita alla nascita di un particolare villaggio è di 35 anni, e di voler calcolare il

⁵ Con questi coefficienti è anche possibile stimare il valore di K senza ricorrere al programma APPLAUSI. Per esempio, ammettiamo di essere interessati al valore di ${}_2K$ per una comunità in cui si suppone $e_0=31$. Si calcola il logaritmo naturale di e_0 (quindi 3,434), lo si trasforma con i parametri a e b della tabella ($1,8845 - 0,439 \times 3,434 = 0,376$) e se ne calcola l'antilogaritmo (1,458). Il risultato è il coefficiente moltiplicativo ${}_2K$ cercato, anche se in questo modo non si è tenuto conto del tipo di mortalità, che non è invece irrilevante (cfr. più avanti).

⁶ Come indicato nella tab. 2, per gli anni 0, 1 e 2 l'elasticità è ancora minore.

⁷ Lo standard generale di Brass risulta invece intermedio tra i due.

valore del coefficiente ${}_4K$. Se non si dispone di informazioni sufficientemente dettagliate per questo villaggio, si supporrà probabilmente applicabile il modello Ovest, ottenendo ${}_4K = 1,3846$ (tab. 3). Ora, se il vero modello di mortalità è non Ovest, ma Nord o Brass, pazienza: l'errore commesso nella stima del *TFT* risulta quasi irrilevante ($\pm 1\%$ circa). Se invece il vero modello di mortalità è Est o Sud, l'errore commesso nella stima del *TFT* può salire fino quasi al 7%, e questo nonostante che il livello generale della mortalità (sintetizzato da e_0) sia stato correttamente individuato.

5. Determinazione dei coefficienti K a partire da q_0

La tab. 1 consente, infine, di determinare i coefficienti K non solo sulla base di e_0 , come si è fatto sinora, ma anche sulla base di q_0 . Il modo di procedere è analogo a quello sin qui considerato. In particolare, si può verificare che la relazione tra i fattori K e il valore di q_0 è (con qualche approssimazione) lineare, come illustrato, a titolo di esempio, nella fig. 4, e se ne possono calcolare i parametri (tab. 2, ultime 4 colonne).⁸

Il coefficiente angolare delle rette interpolate, a eccezione dell'anno 0, è approssimativamente costante e vicino a $2/1000$. Questo significa che un errore assoluto pari a z (in punti millesimali) nella valutazione di q_0 porta a un errore assoluto di K pari a $2/1000$ di z .

⁸ Anche questi parametri, come i precedenti, possono essere usati per stimare il valore di K senza ricorrere al programma APPLAUSI. Per esempio, ammettiamo di essere interessati al valore di ${}_5K$ per un villaggio in cui si suppone $q_0=200$ (per mille). Si opera semplicemente la trasformazione lineare $(0,9176 + 0,0024 \times 200 = 1,3976)$ e il risultato è il coefficiente moltiplicativo ${}_5K$ cercato. Anche in questo caso, occorre ricordare che così facendo non si è tenuto conto del modello di mortalità, che non è irrilevante ai fini dei risultati (cfr. più avanti nel testo).

Per esempio, se il vero valore di q_0 è 200 per mille, ma (erroneamente) si stima che sia invece 250 per mille ($z=50$), K risulta di circa $(50 \times 2/1000 =) 0,1$ troppo alto. In pratica, al posto del valore corretto di 1,3976 calcolato nella nota 8, si otterrà un valore di 1,4976, con sovrastima del $(0,1/1,3976 =) 7\%$ circa.

Si noti comunque che, in questo caso, le stime di fecondità risultano decisamente più affidabili che non nel caso precedente. Reputare che la mortalità infantile sia pari a 250 per mille quando invece è del 200 per mille significa compiere un errore macroscopico (pari al 25%), che si traduce però in una distorsione dei risultati comparativamente contenuta, pari, nell'esempio appena visto, al 7%. Anche se, in questo caso, non si osserva un'elasticità costante tra errore relativo delle ipotesi e errore relativo nei risultati, si verifica comunque che le stime di fecondità tendono a "assorbire" eventuali errori nelle ipotesi, e sono quindi piuttosto robuste.

Per contro, anche in questo caso, si conferma e anzi cresce l'importanza di scegliere correttamente il modello di mortalità, dato che si configurano due gruppi di modelli relativamente omogenei tra di loro rispetto a questa particolare applicazione: da una parte i modelli Sud e Nord, che richiedono coefficienti più elevati, e dall'altra gli altri tre modelli, per i quali sono invece sufficienti correzioni minori.⁹

Si può considerare anche in questo caso l'esempio, analogo al precedente, di una comunità in cui si sappia che la mortalità infantile è del 200 per mille e di cui si voglia calcolare il fattore K relativo all'anno 8. Se si adotta il modello Ovest, si otterrà un ${}_8K = 1,3281$ (tab. 4). Se poi il vero modello di mortalità è non Ovest, ma Est o Brass, pazienza: l'errore commesso nella stima del *TFT* risulta quasi irrilevante (circa -1%). Se però il vero modello di mortalità è Sud o

⁹ Si noti che questi gruppi non sono gli stessi individuati partendo invece da e_0 .

Nord, le cose peggiorano sensibilmente, perché l'errore commesso nella stima del *TFT* può salire fino al 10%.

6. Conclusioni

Lo scopo principale di questo lavoro era la presentazione delle tabb. 1 e 2, che consentono di determinare rapidamente, e in maniera relativamente precisa, l'effetto di una vasta gamma di possibili scenari di mortalità sulle stime di fecondità ottenute con il metodo dei figli propri.

Le simulazioni condotte dimostrano che occorre stare attenti a definire con il massimo di precisione possibile il contesto di mortalità all'interno del quale ci si muove, perché ipotesi errate sul livello o sul tipo della mortalità¹⁰ possono produrre risultati distorti. I dati delle tabb. 1 e 2 consentono di quantificare, e quindi di rendere meno banale, questa osservazione.

In generale, sembra di poter dire che le stime di fecondità ottenute con il metodo dei figli propri sono relativamente robuste rispetto alle ipotesi di mortalità adottate, anche se nelle applicazioni concrete converrà spesso definire, oltre a quello principale, anche una piccola serie di scenari di mortalità alternativi per meglio apprezzare il possibile campo di variazione delle stime ottenute. I coefficienti delle tabb. 1 e 2 consentono di farlo rapidamente e con poca fatica.

¹⁰ O, peggio, sia sul livello che sul tipo, un caso non considerato in questo lavoro. Nei casi più fortunati, gli effetti distorsivi possono anche tendere a annullarsi, ma non si può escludere il rischio che essi, invece, si cumulino.

Bibliografia

W. Brass (1971), On the scale of mortality, in W. BRASS (Ed.) *Biological aspects of demography*, Taylor and Francis, London.

M. Breschi e G. De Santis (1992), Hacia una nueva utilización de las matrículas de feligreses. El método de los hijos propios y su aplicación en demografía histórica, *Boletín de la Asociación de Demografía Histórica*, n. 2 (trad. it.: Verso una nuova utilizzazione degli Stati delle Anime. Il metodo dei figli propri e la sua applicazione in demografia storica, *Bollettino SIDES*, n. 17, 1992).

M. Breschi e G. De Santis (in preparazione), *Il metodo dei figli propri in demografia storica*.

L.J. Cho, R.D. Retherford e M.K. Choe (1986), *The own-children method of fertility estimation*, Un. of Haway Press, East-West Center.

A. Coale e P. Demeny (1983), *Regional model life tables and stable populations*, 2nd Ed., Academic Press, New York.

UN (1983), *Manual X*, New York.

Tab. 1 - Coefficienti di correzione K per alcune curve di mortalità selezionate

Modello	Funzioni		Valore dei coefficienti K per ciascun anno di riferimento												
	biometriche														
	a(0)	q(0)	12	11	10	9	8	7	6	5	4	3	2	1	0
Ovest	22.41	311.0	1.5567	1.5839	1.6107	1.6352	1.6577	1.6787	1.6978	1.7039	1.7013	1.6776	1.6270	1.5008	1.2398
Ovest	27.44	269.1	1.4339	1.4549	1.4756	1.4943	1.5109	1.5256	1.5381	1.5478	1.5468	1.5307	1.4981	1.4042	1.2017
Ovest	32.47	250.4	1.3407	1.3572	1.3734	1.3882	1.4012	1.4128	1.4224	1.4306	1.4304	1.4191	1.3934	1.3282	1.1687
Ovest	37.50	194.4	1.2673	1.2806	1.2935	1.3062	1.3156	1.3248	1.3328	1.3391	1.3394	1.3314	1.3125	1.2621	1.1401
Ovest	42.47	162.0	1.2078	1.2185	1.2284	1.2383	1.2465	1.2540	1.2606	1.2656	1.2662	1.2606	1.2484	1.2084	1.1186
Ovest	47.46	131.2	1.1577	1.1663	1.1746	1.1822	1.1889	1.1949	1.2000	1.2042	1.2080	1.2098	1.1906	1.1622	1.0938
Ovest	52.41	101.4	1.1133	1.1202	1.1269	1.1330	1.1384	1.1433	1.1478	1.1509	1.1518	1.1491	1.1416	1.1298	1.0741
Ovest	57.34	74.4	1.0764	1.0819	1.0872	1.0921	1.0964	1.1003	1.1037	1.1065	1.1074	1.1087	1.1005	1.0887	1.0667
Ovest	62.28	50.6	1.0465	1.0508	1.0549	1.0587	1.0620	1.0650	1.0677	1.0699	1.0708	1.0698	1.0665	1.0586	1.0408
Sud	22.48	271.8	1.7652	1.7886	1.8116	1.8344	1.8563	1.8767	1.8950	1.8903	1.8573	1.7922	1.6870	1.4797	1.2040
Sud	27.47	244.2	1.6090	1.6264	1.6435	1.6601	1.6757	1.6894	1.6730	1.6737	1.6660	1.6332	1.5579	1.3983	1.1809
Sud	32.46	215.9	1.4881	1.5014	1.5144	1.5240	1.5300	1.5345	1.5378	1.5384	1.5322	1.5092	1.4637	1.3338	1.1578
Sud	37.49	188.0	1.3919	1.4022	1.4123	1.4198	1.4245	1.4282	1.4306	1.4314	1.4271	1.4100	1.3685	1.2772	1.1360
Sud	42.50	160.7	1.3127	1.3208	1.3286	1.3345	1.3383	1.3412	1.3432	1.3440	1.3409	1.3281	1.2968	1.2270	1.1186
Sud	47.42	134.2	1.2464	1.2528	1.2589	1.2636	1.2666	1.2690	1.2707	1.2718	1.2692	1.2596	1.2389	1.1835	1.0959
Sud	52.37	109.6	1.1913	1.1963	1.2011	1.2048	1.2072	1.2091	1.2105	1.2111	1.2096	1.2024	1.1846	1.1441	1.0804
Sud	57.36	86.6	1.1448	1.1488	1.1523	1.1552	1.1571	1.1584	1.1597	1.1603	1.1591	1.1539	1.1407	1.1106	1.0685
Sud	62.40	65.0	1.1048	1.1074	1.1102	1.1123	1.1138	1.1150	1.1158	1.1163	1.1156	1.1119	1.1024	1.0808	1.0514
Est	22.44	370.9	1.7877	1.7882	1.8123	1.8346	1.8512	1.8645	1.8742	1.8766	1.8734	1.8448	1.7688	1.6013	1.3062
Est	27.45	322.6	1.6904	1.6111	1.6213	1.6480	1.6606	1.6706	1.6743	1.6821	1.6775	1.6590	1.6198	1.6013	1.2865
Est	32.47	277.8	1.6440	1.4799	1.4964	1.5083	1.5181	1.5260	1.5320	1.5362	1.5393	1.5191	1.4906	1.4182	1.2137
Est	37.47	236.6	1.3657	1.3781	1.3902	1.4003	1.4080	1.4143	1.4190	1.4217	1.4199	1.4103	1.3802	1.3360	1.1770
Est	42.46	197.9	1.2859	1.2957	1.3052	1.3131	1.3193	1.3243	1.3282	1.3306	1.3295	1.3222	1.3068	1.2658	1.1460
Est	47.46	169.5	1.2139	1.2207	1.2275	1.2336	1.2385	1.2425	1.2466	1.2498	1.2518	1.2460	1.2344	1.2039	1.1155
Est	52.42	124.5	1.1586	1.1645	1.1708	1.1769	1.1799	1.1821	1.1847	1.1874	1.1871	1.1825	1.1761	1.1636	1.0907
Est	57.34	93.3	1.1122	1.1171	1.1218	1.1267	1.1288	1.1314	1.1335	1.1360	1.1340	1.1325	1.1264	1.1162	1.0699
Est	62.30	64.9	1.0724	1.0771	1.0808	1.0835	1.0850	1.0860	1.0867	1.0894	1.0909	1.0884	1.0833	1.0742	1.0613
Nord	22.36	264.5	1.6351	1.6595	1.6801	1.6984	1.7144	1.7282	1.7402	1.7463	1.7490	1.7437	1.7284	1.6802	1.5825
Nord	27.39	225.8	1.5033	1.5197	1.5366	1.5424	1.5406	1.5382	1.5365	1.5324	1.5343	1.4734	1.4145	1.3232	1.1838
Nord	32.44	196.9	1.3993	1.4122	1.4246	1.4303	1.4296	1.4284	1.4270	1.4262	1.4123	1.3823	1.3386	1.2694	1.1410
Nord	37.48	168.4	1.3175	1.3279	1.3377	1.3425	1.3424	1.3419	1.3412	1.3402	1.3306	1.3046	1.2765	1.2226	1.1205
Nord	42.50	142.3	1.2506	1.2587	1.2666	1.2706	1.2704	1.2707	1.2705	1.2700	1.2632	1.2464	1.2214	1.1809	1.1017
Nord	47.50	117.8	1.1949	1.2025	1.2074	1.2111	1.2116	1.2118	1.2114	1.2117	1.2067	1.1942	1.1763	1.1444	1.0849
Nord	52.42	92.2	1.1442	1.1496	1.1547	1.1576	1.1582	1.1587	1.1590	1.1591	1.1556	1.1464	1.1324	1.1094	1.0682
Nord	57.36	69.1	1.1024	1.1068	1.1109	1.1133	1.1140	1.1146	1.1150	1.1154	1.1130	1.1064	1.0963	1.0797	1.0634
Nord	62.29	49.0	1.0688	1.0722	1.0754	1.0773	1.0780	1.0787	1.0792	1.0797	1.0782	1.0737	1.0667	1.0581	1.0398
Bras	22.50	390.5	1.7509	1.7742	1.8049	1.8294	1.8513	1.8706	1.8863	1.8990	1.9020	1.8861	1.8367	1.6999	1.3282
Bras	27.50	314.5	1.6187	1.6377	1.6592	1.6790	1.6967	1.6124	1.6259	1.6364	1.6402	1.6303	1.6067	1.5003	1.2476
Bras	32.50	252.4	1.3622	1.3794	1.3970	1.4133	1.4279	1.4409	1.4522	1.4612	1.4652	1.4591	1.4348	1.3602	1.1888
Bras	37.50	200.5	1.2548	1.2698	1.2843	1.2978	1.3098	1.3207	1.3301	1.3374	1.3417	1.3380	1.3208	1.2708	1.1449
Bras	42.50	157.4	1.1796	1.1919	1.2038	1.2148	1.2248	1.2337	1.2416	1.2480	1.2516	1.2495	1.2373	1.2005	1.1119
Bras	47.50	121.5	1.1254	1.1365	1.1452	1.1541	1.1622	1.1694	1.1758	1.1812	1.1843	1.1832	1.1746	1.1481	1.0870
Bras	52.50	91.8	1.0802	1.0943	1.1020	1.1091	1.1165	1.1213	1.1264	1.1307	1.1333	1.1328	1.1268	1.1078	1.0677
Bras	57.50	67.3	1.0675	1.0642	1.0701	1.0756	1.0805	1.0849	1.0888	1.0922	1.0943	1.0942	1.0901	1.0767	1.0520
Bras	62.50	47.5	1.0376	1.0422	1.0468	1.0507	1.0543	1.0575	1.0604	1.0629	1.0645	1.0646	1.0619	1.0528	1.0386

*) Valori generati con il programma APPLAUSI (Breschi e De Santis, in preparazione)

Tab. 2 - Alcuni risultati delle regressioni per il calcolo dei coefficienti K

Anni	log(K) log(e0)				K q0			
	Cost.	Coef. regr.	Errore stand.	R quadro	Cost.	Coef. regr.	Errore stand.	R quadro
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0	0.7804	-0.1784	0.0099	0.8836	0.9925	0.0008	0.0000	0.9901
1	1.5540	-0.3619	0.0172	0.9117	0.9414	0.0018	0.0000	0.9818
2	1.8845	-0.4398	0.0171	0.9387	0.9257	0.0022	0.0001	0.9588
3	2.0298	-0.4741	0.0165	0.9503	0.9203	0.0024	0.0001	0.9406
4	2.1007	-0.4911	0.0157	0.9580	0.9182	0.0025	0.0001	0.9299
5	2.1217	-0.4963	0.0154	0.9604	0.9177	0.0025	0.0001	0.9232
6	2.1146	-0.4949	0.0156	0.9592	0.9180	0.0025	0.0001	0.9178
7	2.1007	-0.4920	0.0158	0.9577	0.9189	0.0024	0.0001	0.9111
8	2.0816	-0.4879	0.0160	0.9556	0.9202	0.0024	0.0001	0.9031
9	2.0579	-0.4827	0.0164	0.9527	0.9219	0.0024	0.0001	0.8935
10	2.0217	-0.4747	0.0168	0.9487	0.9236	0.0023	0.0001	0.8857
11	1.9739	-0.4640	0.0173	0.9438	0.9250	0.0022	0.0001	0.8803
12	1.9245	-0.4529	0.0177	0.9381	0.9263	0.0021	0.0001	0.8741

Tab. 3 - Esempio di calcolo del fattore di correzione K per l'anno 4 e per $e(0)=35$

Modello	Limiti	Informazioni iniziali (1)			Risultati	
		$e(0)$	K	semi impliciti	K	Errore % rispetto al mod. Ovest
Ovest	inferiore	32.47	1.4304	0.4970	1.3846	0
	superiore	37.50	1.3394	0.5030		
Sud	inferiore	32.49	1.5322	0.4980	1.4794	6.8
	superiore	37.49	1.4271	0.5020		
Est	inferiore	32.47	1.5323	0.4940	1.4754	6.6
	superiore	37.47	1.4199	0.5060		
Nord	inferiore	32.44	1.4123	0.4931	1.3710	-1.0
	superiore	37.49	1.3308	0.5069		
Brass	inferiore	32.50	1.4652	0.5000	1.4035	1.4
	superiore	37.50	1.3417	0.5000		

1) Cfr. tab. 1.

Tab. 4 - Esempio di calcolo del fattore di correzione K per l'anno 8 e per $q(0)=200$ per mille

Modello	Limiti	Informazioni iniziali (1)			Risultati	
		$q(0)$ $\times 1.000$	K	semi impliciti	K	Errore % rispetto al mod. Ovest
Ovest	superiore	230.4	1.4012	0.1461	1.3281	0
	inferiore	194.8	1.3156	0.8539		
Sud	superiore	215.9	1.5300	0.4301	1.4699	10.7
	inferiore	188.0	1.4245	0.5699		
Est	superiore	236.5	1.4080	0.0544	1.3241	-0.3
	inferiore	197.9	1.3193	0.9456		
Nord	superiore	225.8	1.5405	0.1073	1.4415	8.5
	inferiore	196.9	1.4296	0.8927		
Brass	superiore	200.5	1.3098	0.9884	1.3088	-1.5
	inferiore	157.4	1.2248	0.0116		

1) Cfr. tab. 1.

Fig. 1 - Valore dei coefficienti di correzione K per anno e per alcuni valori di $e(0)$ del modello Ovest

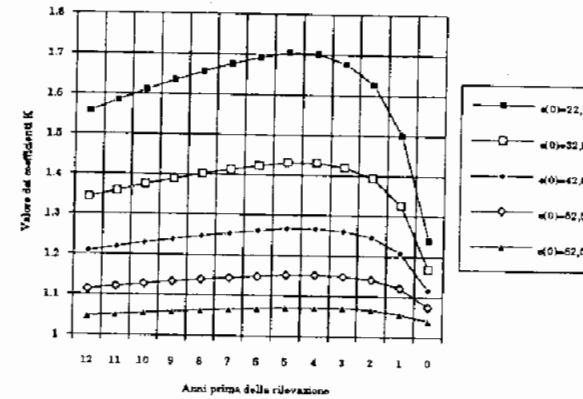
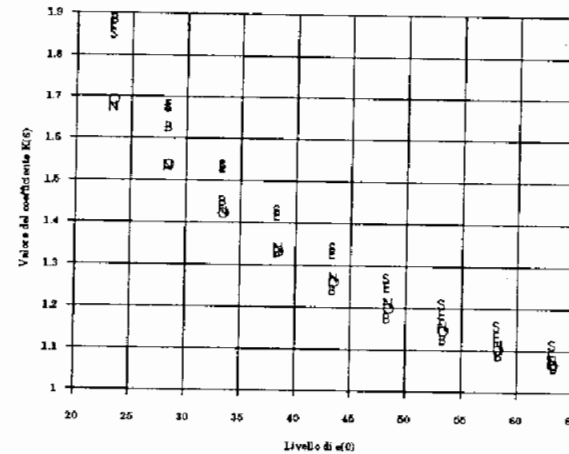


Fig. 2 - Relazione tra $e(0)$ e $K(t)$ per vari modelli e livelli di mortalità



**STUDIO DELLE CARATTERISTICHE DI UNA POPOLAZIONE,
CON RIFERIMENTO AD UN MODELLO STABILE
IN REGIME DI NUZIALITÀ COSTANTE**

1. Nel presente contributo viene illustrato un tentativo di valorizzazione delle ricche fonti archivistiche italiane relative ai matrimoni, nell'ambito degli studi che utilizzano il ricorso a tavole di popolazioni stabili "modello". Tale tentativo presuppone una forzatura del modello stabile, destinata a consentire l'inserimento del regime nuziale e conseguentemente a permettere l'ingresso in una tavola mediante l'utilizzazione di un indicatore appositamente determinato. Lo scopo di questo tentativo è quello di aumentare le possibilità di ricorrere a procedure di stima delle ignote caratteristiche di una popolazione a partire dalla conoscenza di un ridotto numero di informazioni sui suoi processi demografici. Finora, infatti, nell'ipotesi che una popolazione oggetto di studio si approssimi nella sua evoluzione ad un regime stabile e che non sia interessata a fenomeni migratori di rilievo, ci si è basati sulla conoscenza di un paio di parametri significativi dei suoi processi e della sua struttura (scelti tra il tasso di natalità, il tasso di mortalità, il livello di fecondità, il tasso di incremento, la composizione per età) al fine di entrare nelle tavole di popolazioni stabili modello e di desumere da queste sia un controllo di coerenza delle informazioni disponibili, sia la stima di altri parametri che non è possibile, o sarebbe troppo oneroso determinare nella popolazione studiata. In quest'ambito di valutazioni l'utilizzazione delle statistiche sui matrimoni e sulla nuzialità - che sono in genere disponibili - viene, quindi, del tutto

Fig. 3 - Relazione tra i log. naturali di $e(t)$ e di $K(t)$ per vari modelli e livelli di mortalità

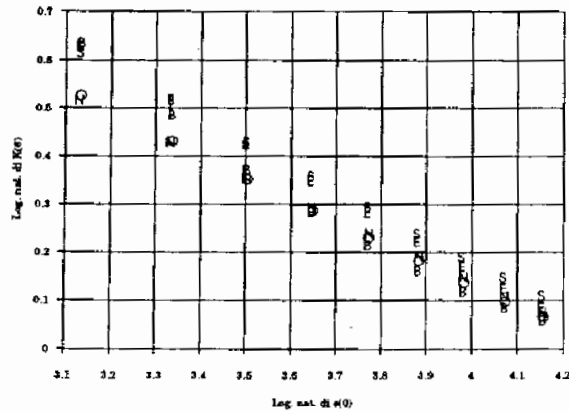
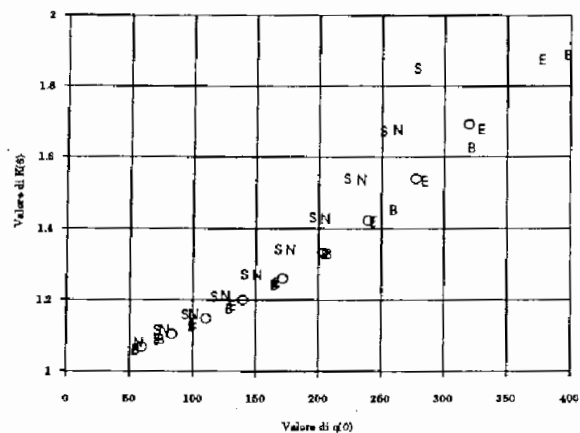


Fig. 4 - Relazione tra $q(t)$ e $K(t)$ per vari modelli e livelli di mortalità



trascurata. Eppure si tratta di fonti altamente affidabili, almeno nel senso della loro rispondenza al reale numero di eventi matrimoniali intervenuti in una popolazione. Mentre in vari casi è noto che le registrazioni dei battesimi e delle sepolture spesso comportano sottoregistrazioni dei nati morti e in parte anche delle nascite seguite a breve distanza da un decesso (mortalità neonatale precoce). Sembra quindi utile tentare di utilizzare le serie storiche dei matrimoni e dei tassi di nuzialità negli studi rivolti a stimare le caratteristiche di una popolazione,¹ avendo l'accortezza di riferirsi a collettività sufficientemente ampie, in modo da eliminare o minimizzare i problemi connessi con il manifestarsi di importanti fenomeni di mobilità causati dal matrimonio. Al tempo stesso sarà opportuno avere un'idea dell'entità delle nascite non matrimoniali nella collettività studiata.

2. Supponiamo che in una popolazione stabile il numero annuo di nascite (N) venga prodotto esclusivamente in regime matrimoniale e che il numero di matrimoni (Ma) segua nel corso del tempo un incremento pari a $1+r$, così come, secondo il modello stabile, avviene per le nascite e per il totale della popolazione; in tale situazione il tasso di nuzialità $nu = Ma/P$ sarà costante e pari, nell'anno 0, a:

$${}_0nu = \frac{{}_0Ma}{{}_0N \cdot \sum e^{-rx} L_x} \quad (1)$$

indicando, come è noto, l'espressione posta al denominatore l'ammontare della popolazione al momento 0, ed essendo

¹ L'assunzione della nuzialità in un modello stabile è stata già sperimentata in uno studio di Demonet-Dupâquier-Les Bras. Gli autori non forniscono purtroppo indicazioni dettagliate sulla metodologia adottata.

$$L_x = \frac{L_x}{l_0}$$

Ricordando che, nell'anno 0, il tasso di natalità della popolazione stabile è pari a:

$$\frac{{}_0N}{{}_0P} = \frac{{}_0N}{{}_0N \cdot \sum e^{-rx} L_x} = \frac{1}{\sum e^{-rx} L_x}$$

il rapporto tra nascite e matrimoni, nell'anno 0, sarà:

$$\frac{{}_0N}{{}_0Ma} = \frac{{}_0P}{{}_0Ma} \cdot \frac{1}{{}_0P} = \frac{1}{{}_0nu \cdot \sum e^{-rx} L_x} = \frac{{}_0^n}{{}_0^{nu}} \quad (2)$$

come del resto è già evidente nella (1)

In virtù delle condizioni di stabilità tale uguaglianza si otterrà in un qualsiasi anno, per cui la (2) può indicarsi più in generale come

$$\frac{n}{nu} = \frac{1}{nu \sum e^{-rx} L_x} \quad (3)$$

La (2) indica che in una popolazione stabile, il rapporto tra nati nell'anno e matrimoni dell'anno è in ogni anno in funzione del tasso di incremento r , della legge di mortalità L_x e del tasso di nuzialità nu .

Si può prendere in considerazione, inoltre, il caso particolare in cui il flusso delle nascite sia costante ed $r = 0$, e cioè ci si trovi nelle condizioni di una popolazione stazionaria con natalità uguale alla mortalità. In tal caso l'espressione (2) diventa uguale a

$$\frac{1}{nu \cdot \sum e^{-rx} L_x} = \frac{1}{nu \cdot e_0} \quad (4)$$

tale rapporto è particolarmente significativo in quanto indica il numero stazionario di nascite per matrimonio, che è determinato dal regime di mortalità e di nuzialità in funzione.

3. I valori di n/nu , nelle diverse condizioni di stabilità, possono essere agevolmente calcolati assumendo il numeratore dalle tavole di popolazioni stabili ed attribuendo al denominatore il livello desiderato. Sono state utilizzate, a scopo sperimentale, le tavole di Coale-Demeny-Vaughan relative al modello Sud, per tutti i livelli di mortalità e tutti i tassi di incremento indicati, determinando il valore medio del tasso di natalità tra la popolazione stabile femminile e quella maschile. Per quanto riguarda la nuzialità, i tassi assunti variano dal 5 al 10 per mille. Risulta da questi computi, pertanto, una tavola di valori di n/nu per ogni livello di mortalità, dove ciascun dato di n/nu è relativo alla combinazione di ogni singolo livello di r e di nu . Per ovvie ragioni di spazio, in questa sede non può essere presentato l'intero set di tavole. A scopo esemplificativo si presentano qui di seguito la tavola relativa al livello 20 (tab. 1) di mortalità (cioè con speranza di vita alla nascita pari a 67.5 e 63.7 anni, rispettivamente per le femmine e per i maschi) ed il relativo grafico.

I valori del parametro n/nu variano da un minimo di 1.02, in una popolazione caratterizzata da un tasso di nuzialità di 10 per mille e da un incremento negativo del 10 per mille, ad un massimo di 11.5 nascite annue per matrimonio necessarie per assicurare la stabilità della popolazione con un tasso di nuzialità del 5 per mille ed un incremento del 50 per mille.

Naturalmente, a parità del tasso d'incremento, il valore di n/nu diminuisce al crescere della nuzialità, e cresce invece quando sia fissa la nuzialità ma aumenti l'incremento.

Nella successiva tavola 2 e nel grafico 2, sono contenuti i dati relativi a valori costanti della nuzialità (pari a 7 per mille) e al

Tabella 1: Valori di n/nu per tasso di nuzialità (per 1000) e incremento della popolazione (per 1000) mortalità modello SUD, livello 20 ($e_0^f=67.5$ e $e_0^m=63.7$)

r per 1000	nu per 1000					
	5	6	7	8	9	10
-10.00	2.04	1.70	1.45	1.27	1.13	1.02
-5.00	2.51	2.09	1.79	1.57	1.39	1.26
0.00	3.05	2.54	2.18	1.91	1.70	1.53
5.00	3.66	3.05	2.62	2.29	2.04	1.83
10.00	4.34	3.62	3.10	2.71	2.41	2.17
15.00	5.08	4.23	3.63	3.18	2.82	2.54
20.00	5.88	4.90	4.20	3.67	3.26	2.94
25.00	6.73	5.60	4.80	4.20	3.74	3.36
30.00	7.62	6.35	5.44	4.76	4.23	3.81
35.00	8.55	7.12	6.11	5.34	4.75	4.28
40.00	9.51	7.93	6.80	5.95	5.29	4.76
45.00	10.51	8.75	7.50	6.57	5.84	5.25
50.00	11.52	9.60	8.23	7.20	6.40	5.76

Grafico 1

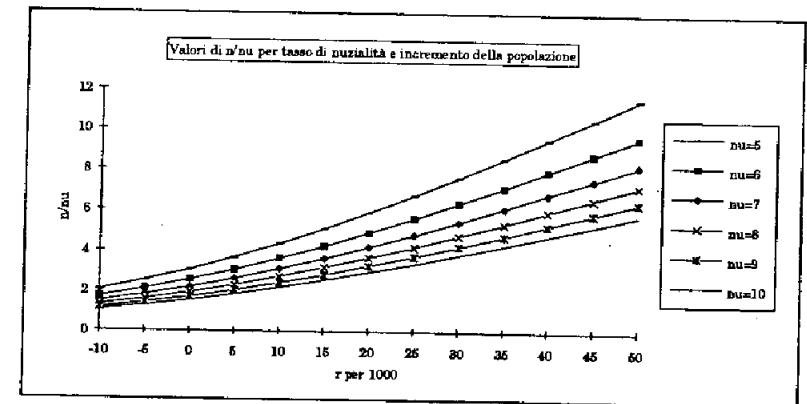


Tabella 2: Valori di n/nu per livello di mortalità e incremento della popolazione (per 1000) ad un tasso costante di nuzialità del 7 per mille

r per 1000	LIVELLO DI MORTALITÀ					
	6	7	8	9	10	11
	valori di n/nu					
-10.00	3.21	2.97	2.76	2.57	2.42	2.28
-5.00	3.80	3.53	3.29	3.08	2.90	2.74
0.00	4.45	4.14	3.88	3.64	3.44	3.25
5.00	5.16	4.82	4.45	4.26	4.03	3.82
10.00	5.92	5.54	5.21	4.92	4.67	4.44
15.00	6.73	6.32	5.95	5.64	5.36	5.11
20.00	7.58	7.13	6.74	6.39	6.09	5.81
25.00	8.47	7.98	7.55	7.18	6.85	6.56
30.00	9.38	8.86	8.41	8.00	7.65	7.33
35.00	10.33	9.77	9.28	8.85	8.48	8.13
40.00	11.29	10.70	10.18	9.72	9.32	8.96
45.00	12.27	11.65	11.10	10.61	10.19	9.80
50.00	13.27	12.61	12.03	11.52	11.07	10.66

Grafico 2

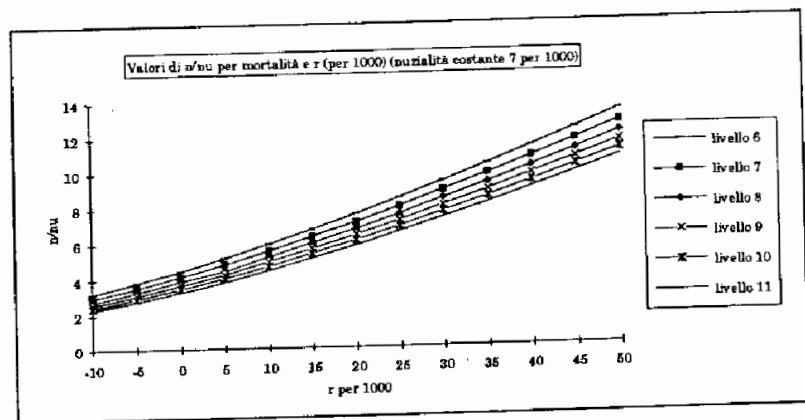
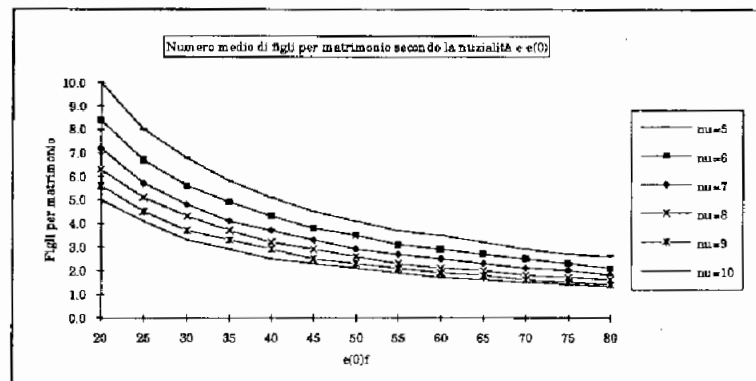


Tabella 3: Valori di n/nu in regime stazionario

Speranza di vita		Numero medio di figli per matrimonio secondo la nuzialità (tasso di nuz. per 1000)					
$e(0)f$	$e(0)m$	5	6	7	8	9	10
20.00	19.92	10.0	8.4	7.2	6.3	5.6	5.0
25.00	24.66	8.0	6.7	5.7	5.1	4.5	4.1
30.00	29.34	6.8	5.6	4.8	4.3	3.7	3.3
35.00	33.95	5.8	4.9	4.1	3.7	3.3	2.9
40.00	38.50	5.1	4.3	3.7	3.2	2.9	2.5
45.00	42.86	4.5	3.8	3.3	2.9	2.5	2.3
50.00	44.37	4.1	3.5	2.9	2.6	2.3	2.1
55.00	51.87	3.7	3.1	2.7	2.3	2.1	1.9
60.00	56.34	3.5	2.9	2.5	2.1	1.9	1.7
65.00	61.25	3.2	2.7	2.3	2.0	1.8	1.6
70.00	66.08	2.9	2.5	2.1	1.8	1.6	1.5
75.00	70.99	2.7	2.3	2.0	1.7	1.5	1.4
80.00	76.00	2.6	2.1	1.8	1.6	1.4	1.3

Grafico 3



variare del tasso di incremento per 6 livelli della mortalità (dal livello 6 al livello 11 delle tavole di Coale-Demeny-Vaughan, cioè da 32.5 a 45 anni di speranza di vita alla nascita femminile). Si può osservare come, una volta fissata la nuzialità, e ad esempio con un incremento del 5 per mille, il rapporto n/nu passi da 5.2 a 3.8 al migliorare delle condizioni di sopravvivenza: meno nascite con migliore sopravvivenza assicurano lo stesso incremento di più nascite sottoposte ad una mortalità più elevata.

Sembra anche interessante osservare che cosa accada in generale in condizioni di stazionarietà (cfr. tab. 3, graf. 3). In questo caso infatti, come sappiamo, le diverse serie di dati indicano il numero di figli prodotti da ogni matrimonio necessario per mantenere stazionaria la popolazione nel dato regime di nuzialità e di mortalità.

4. Come procedere nell'utilizzazione di questo indicatore per entrare nelle tavole di popolazioni stabili modello? E' bene avvertire che la procedura che verrà ora illustrata può essere perfezionata e che i risultati ai quali perverremo sono da considerarsi, per ora, solo in termini di prima approssimazione.

Si supponga quindi di conoscere di una determinata popolazione, per la quale possa essere ipotizzata una condizione di stabilità, soltanto il numero annuo di componenti e il numero annuo di nati e di matrimoni per un lasso di tempo sufficientemente lungo. Avremo quindi la possibilità di determinare il tasso di incremento e i tassi di nuzialità, oltre al rapporto tra nascite e matrimoni. Utilizzando queste informazioni, sarà sufficiente andare a ricercare, nelle tavole dei valori di n/nu precedentemente presentate, il dato corrispondente ai livelli di incremento e di nuzialità noti. Il fatto che il dato individuato sia più o meno prossimo al valore noto del rapporto tra nascite e matrimoni nella popolazione costituirà così una prima verifica dell'ipotesi che le caratteristiche della popolazione si uniformino, nel periodo considerato, ad un regime di stabilità, tenuto

anche conto del fatto che ogni valore di n/nu è collocato in una tavola corrispondente ad un determinato livello di mortalità.

Qualora la coerenza di N/Ma con n/nu sia soddisfacente potremo, a questo punto, entrare nella tavola della popolazione stabile designata da quel livello di mortalità e assumere in primo luogo i diversi parametri di mortalità della tavola-tipo associata ed inoltre - con riferimento al tasso di incremento noto - desumere tutti i parametri della popolazione stabile corrispondente (natalità (CBR), mortalità (CDR), tasso lordo di riproduzione (GRR), struttura per età, ecc.). L'insieme di queste informazioni ci consente di stimare le ignote caratteristiche della popolazione oggetto di studio. E' forse superfluo aggiungere che, qualora i parametri della nostra popolazione non corrispondano esattamente a quelli del tasso di nuzialità, secondo cui sono classificati i valori di n/nu , e del tasso di incremento nelle stesse tavole e nelle tavole di popolazioni stabili, sarà necessario calcolare mediante interpolazione i valori di n/nu e, nelle tavole di popolazioni stabili, i diversi parametri cercati.

5. Allo scopo di valutare l'efficacia della procedura, è stato assunto come riferimento il caso della popolazione della Toscana studiato da Marco Breschi, limitando l'analisi al periodo 1820-1899. In realtà, solo forzando un po' le ipotesi di partenza la popolazione toscana può essere assimilata, in questo lasso di tempo, ad una popolazione stabile. E' pur vero che il saldo migratorio regionale, oscillando tra un valore negativo massimo pari a 4.41 per mille abitanti ed uno positivo massimo di 0.7 per mille abitanti, appare sostanzialmente di scarso rilievo, pur se di segno prevalentemente negativo negli ottanta anni considerati; ma il saldo naturale, invece, nello stesso periodo, scende da valori del 14-15 per mille, all'inizio del periodo, fino a 4-5 per mille nel corso del primo quarantennio, per poi risalire fino al 10-11 per mille nel quarantennio successivo. Ciò è conseguenza di un calo abbastanza regolare della natalità, mentre la

mortalità dapprima aumenta, poi diminuisce. Siamo quindi abbastanza lontani da un modello evolutivo stabile, ed anche da uno cosiddetto "quasi stabile". Tuttavia, è illusorio sperare di trovare nel passato, per lunghi lassi di tempo, popolazioni che si avvicinano in modo pienamente soddisfacente ai requisiti di tali modelli; ed inoltre lo studio condotto da Breschi offre, ai nostri fini, altri vantaggi. Infatti, esso è l'unico riuscito tentativo di ricostruzione di una popolazione regionale italiana per un vasto arco temporale: esso copre in realtà i tre secoli che vanno dal 1640 al 1940. Inoltre tale ricostruzione è stata condotta facendo ricorso ad una tecnica del tutto diversa da quella adottata nel nostro lavoro, e cioè utilizzando lo schema dell'"inverse projection" ideato da R. Lee nel 1974. Ciò consente di istituire confronti dei nostri risultati con alcuni di quelli risultanti dalla ricostruzione di Breschi senza tema di poter registrare fittizie convergenze.

In effetti, nel nostro lavoro ci si è mossi partendo dalla sola utilizzazione dei dati relativi alla popolazione, alle nascite, ai matrimoni della Toscana, nei quinquenni dal 1820-24 al 1895-99. Si è poi passati ad applicare a tali dati la procedura precedentemente illustrata, ricavando i valori di N/Ma da mettere a confronto coi rapporti n/nu calcolati nell'ipotesi di stabilità e poi stimando sulle tavole di Coale-Demeny-Vaughan i parametri CBR, CDR, e_0 , GRR. Considerato lo scarso rilievo esercitato dal fenomeno migratorio, si è deciso di utilizzare il valore del saldo naturale (S_n) come valore del tasso di incremento r . I parametri stimati sono stati, infine, confrontati con quelli ottenuti da Breschi mediante l'"inverse projection".

Nella tabella 4 sono elencati, quinquennio per quinquennio, sia i dati di partenza (nelle prime tre colonne), sia i parametri ricostruiti

da Breschi, sia, a destra, le nostre stime². I grafici successivi consentono di ottenere un'immagine di sintesi dei confronti.

Tabella 4: Indicatori demografici della popolazione della Toscana 1820-1899

Valori ricostruiti da Breschi (*)

Periodo	nu	S_n	N/Ma	$e(0)$	GRR	CBR	CDR
1820-24	9.28	13.97	4.43	37.29	2.70	41.15	27.18
1825-29	7.62	14.59	5.39	38.91	2.81	41.04	26.45
1830-34	6.97	9.99	5.41	36.59	2.68	37.72	27.73
1835-39	6.48	10.12	5.45	38.43	2.60	35.34	25.22
1840-44	7.30	7.45	4.67	36.24	2.51	34.11	26.66
1845-49	7.76	9.16	4.60	36.70	2.53	35.70	26.54
1850-54	7.12	9.56	5.03	35.98	2.42	36.87	27.31
1855-59	8.68	3.59	4.10	29.79	2.28	36.64	33.05
1860-64	8.56	8.81	4.55	33.26	2.48	38.96	30.15
1865-69	6.62	9.83	6.00	34.09	2.62	39.67	29.84
1870-74	6.97	5.54	5.41	31.17	2.55	37.69	32.15
1875-79	7.41	6.63	4.94	32.71	2.52	36.59	29.96
1880-84	7.62	7.89	4.54	35.87	2.40	34.57	26.68
1885-89	8.12	10.68	4.41	38.32	2.49	35.82	25.14
1890-94	7.54	10.01	4.47	40.10	2.34	33.74	23.73
1895-99	6.96	10.28	4.52	43.46	2.18	31.43	21.15

(*) Breschi, tabelle 6.1, 6.5, V.1, V.4

(continua)

² In fondo al testo viene illustrato, con esempi, il criterio di interpolazione seguito per la determinazione delle stime.

Tabella 4: Indicatori demografici della popolazione della Toscana 1820-1899 (continua)

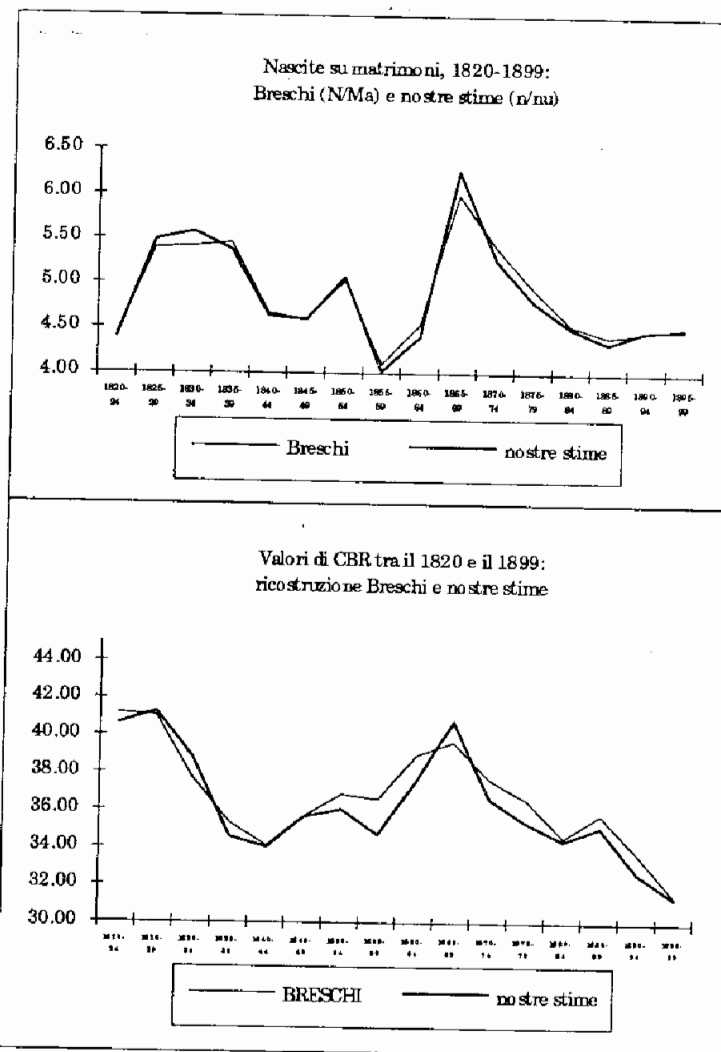
Nostre stime (***)

Periodo	level	n/nu(**)	e(0)	GRR	CBR	CDR
1820-24	8	4.39	36.87	2.71	40.61	26.64
1825-29	8	5.48	36.87	2.75	41.25	26.66
1830-34	7	5.57	34.48	2.57	38.79	28.80
1835-39	9	5.37	39.25	2.30	34.58	24.46
1840-44	8	4.64	36.87	2.26	34.01	26.56
1845-49	8	4.60	36.87	2.36	35.67	26.51
1850-54	8	5.07	36.87	2.39	36.06	26.50
1855-59	6	4.02	32.08	2.29	34.74	31.15
1860-64	7	4.41	34.48	2.49	37.59	28.78
1865-69	6	6.27	32.08	2.73	40.80	31.45
1870-74	6	5.27	32.08	2.42	36.66	31.17
1875-79	7	4.80	34.48	2.34	35.38	28.75
1880-84	8	4.51	36.87	2.28	34.44	26.55
1885-89	9	4.34	39.25	2.34	35.14	24.46
1890-94	10	4.48	41.57	2.18	32.71	22.70
1895-99	11	4.50	43.93	2.09	31.35	21.07

(**) nostro indice

(***) Nostre elaborazioni di dati contenuti in Coale-Demeny, tavole di popolazioni stabili, modello sud, (i livelli considerati sono indicati nella colonna "level" della presente tabella)

Grafici 4 e 5



Grafici 6 e 7

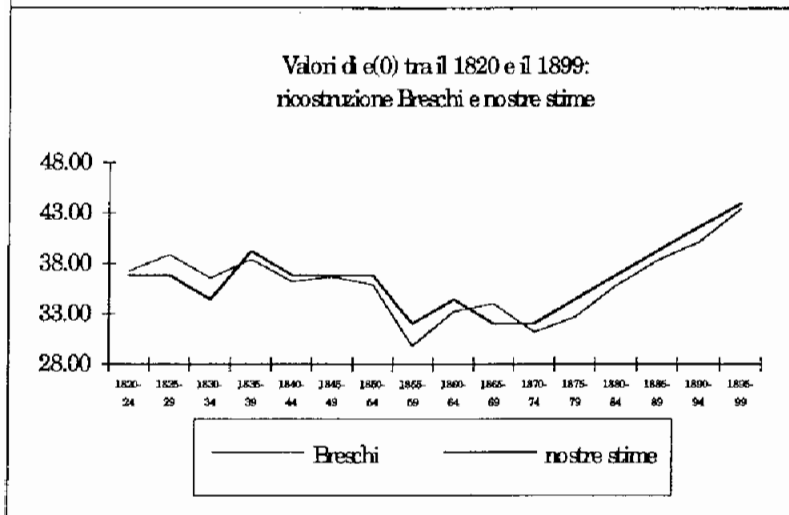
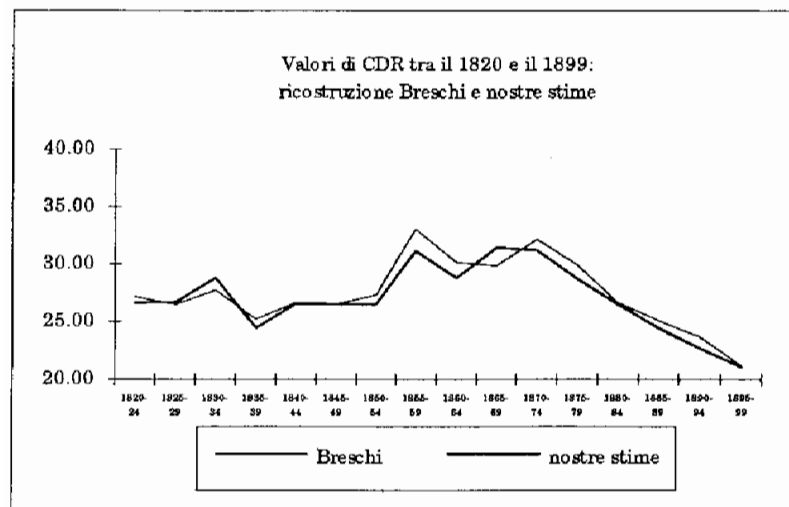
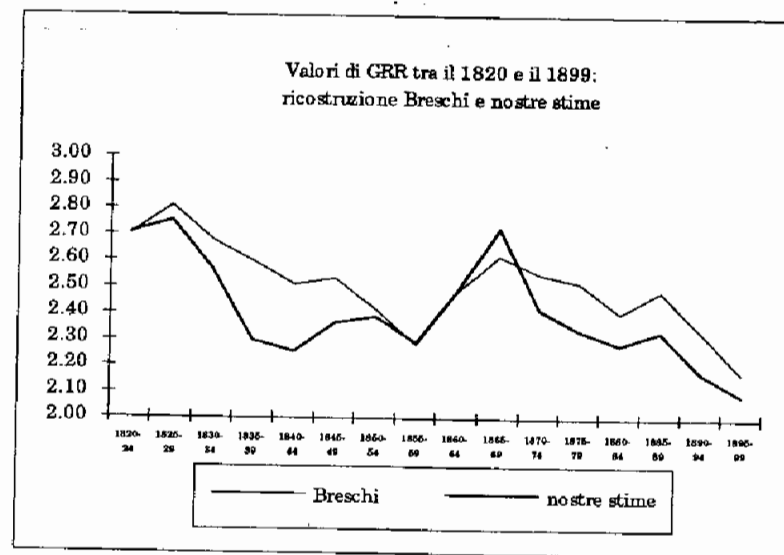


Grafico 8



Il confronto tra il rapporto N/Ma , desunto dai dati di base, e il parametro n/nu corrispondente ai livelli di r e di nu della popolazione toscana nei diversi quinquenni, ha dato ottimi risultati (cfr. graf.4), incoraggiando l'utilizzazione delle tavole di popolazioni stabili associate ai livelli di mortalità propri del modello sud, che è stato scelto in quanto segnalato dallo stesso Breschi come il più idoneo ad illustrare la mortalità toscana.

Ciò che appare particolarmente interessante è non soltanto il buon accordo dei dati (in effetti il maggiore scostamento tra le due serie si osserva in corrispondenza del quinquennio 1865-69, quando il rapporto N/Ma è pari a 6.0 e il rapporto n/nu dà 6.27: una differenza del 4.5%) ma soprattutto la stretta uniformità degli andamenti delle due spezzate, pur in una fase caratterizzata da forti oscillazioni del rapporto N/Ma , che varia tra un minimo di 4.1 e un massimo di 6.0,

mentre il rapporto n/nu presenta un minimo di 4.0 e un massimo di 6.27

Una analoga sintonia di andamenti si può osservare anche per tutti gli altri parametri presi in considerazione, anche se gli adattamenti tra i vari punti delle due spezzate messe a confronto non sono sempre ugualmente soddisfacenti. Nel tasso di natalità (CBR) si osservano valori molto prossimi fino al quinquennio 1845-49 e nell'ultimo ventennio; nel periodo intermedio si osservano scostamenti più accentuati, ma la massima distanza è comunque, nel 1855-59, dell'ordine del 5.2% (CBR calcolato da Breschi 36.64, nostra stima 34.74). Il quoziente di mortalità (CDR) presenta tra il 1855-59 e il 1870-74 qualche divergenza di livelli e anche di andamento, ma la massima differenza tra i due dati, situata in corrispondenza del quinquennio 1855-59 non supera il 5.75%. Passando ai due parametri più significativi dell'evoluzione demografica, la speranza di vita alla nascita (e_0) ed il tasso lordo di riproduzione (GRR), gli adattamenti tra i valori ricostruiti da Breschi e le nostre stime risultano ugualmente significativi, nella corrispondenza generale delle fasi evolutive dei parametri. Si osservano, tuttavia, alcune differenze più sensibili nei valori degli indici, per alcuni quinquenni. Le distanze maggiori riguardano, nel caso di e_0 , il 1855-59 (valore ricostruito da Breschi 29.79 anni, nostra stima 32.08 anni: differenza 7.7%) e il 1865-69 (34.09 contro 32.08: differenza 6.2%). Mentre per il GRR si osserva una differenza massima nel 1840-44 dell'ordine di 11.5% (tasso lordo di riproduzione ricostruito da Breschi 2.6, nostra stima 2.3) e differenze più contenute in altri quinquenni.

Non è certo il caso di entrare, qui, nel merito storico-demografico dei fattori e delle caratteristiche dell'evoluzione della popolazione toscana durante il secolo XIX; a questo riguardo si rinvia senz'altro a quanto illustrato nell'opera di Breschi. Ma rispetto ai confronti istituiti tra i parametri di Breschi e le nostre stime degli stessi, si può trarre qualche conclusione, sia pur provvisoria, tutto sommato soddisfacente. Le buone sintonie riscontrate tra le due serie di

parametri, negli andamenti ed anche nei livelli, sembrano deporre a favore dell'utilità della procedura proposta, soprattutto ove si consideri che a questo si è pervenuti - nell'un caso e nell'altro - seguendo due strade molto distinte tra loro. E' importante poi, a nostro avviso, aver recuperato, nel modo descritto, una utilizzazione delle serie dei matrimoni e della nuzialità in procedure di stima delle caratteristiche di popolazioni del passato. Infatti, la variabile nuzialità si è rivelata in varie ricerche uno strumento importante nello studio dei processi demografici e le sue modificazioni un fattore decisivo nei successivi cambiamenti dei regimi riproduttivi. Perché, quindi, trascurarla?

Prima di concludere, vale forse la pena di ritornare su quanto osservato sopra circa i possibili miglioramenti da introdurre nella procedura proposta. Questi sembra, per ora, debbano essere ricercati in due direzioni. Innanzi tutto mediante un approfondimento della metodologia, cioè entrando meglio nel merito delle diverse conseguenze che possono derivare, sul piano formale e su quello strumentale, dall'introduzione del regime nuziale nel modello stabile. Se questa sembra essere una strada da perseguire, che, tuttavia, si profila indubbiamente ardua, più semplice sembra poter sperimentare qualche altro perfezionamento. Ad esempio, varrà la pena di percorrere anche l'utilizzazione degli altri parametri noti che vengono tradizionalmente utilizzati (natalità e mortalità, tasso d'incremento e natalità, ecc.), per verificare se ne deriverebbero, col ricorso alle tavole di popolazioni stabili, risultati diversi, e in che senso, rispetto a quelli sperimentati con la nuova procedura. Quest'ultima poi può, per suo conto, essere attuata anche costruendo altre serie di valori del parametro n/nu basati su modelli diversi da quello adottato (i modelli nord, ovest e est delle tavole di Coale-

Demeny-Vaughan), in modo da poter ampliare il ventaglio delle possibilità di stima dei parametri ignoti³

Nota sui criteri di interpolazione.

Si supponga di dover calcolare il valore di n/nu corrispondente al livello di mortalità 8, e al tasso di nuzialità $nu=7.30$ per mille e al tasso di incremento $r=7.45$ per mille che caratterizzano la popolazione toscana nel quinquennio 1840-1844. I dati a nostra disposizione, che rappresentano la media tra i valori maschili e femminili di n/nu per il livello 8, sono i seguenti:

	$nu=7$	$nu=8$
$r=5$	4,45	3,89
$r=10$	5,21	4,56

Il valore di n/nu cercato si ottiene semplicemente applicando le tre formule seguenti:

$$x_1 = \frac{7.45 - 5}{10 - 5} * (5.21 - 4.45) + 4.45 = 4.82$$

$$x_2 = \frac{7.45 - 5}{10 - 5} * (4.56 - 3.89) + 3.89 = 4.22$$

$$n / nu = \frac{7.30 - 7}{8 - 7} * (x_2 - x_1) + x_1 = 4.64$$

³ Tale ulteriore verifica è stata già compiuta, ma l'illustrazione dei risultati non può trovare spazio in questa sede. In generale, gli adattamenti risultano buoni anche con riferimento agli altri modelli di mortalità; è da segnalare, anzi, che il modello sud presenta gli adattamenti minori.

Se invece si vuole calcolare il GRR per le femmine, sempre al livello 8 di mortalità e sempre con un tasso di crescita di 7.45 per mille, è sufficiente una sola interpolazione lineare, tra il valore di GRR corrispondente al tasso d'incremento $r=5$ per mille e quello corrispondente al tasso d'incremento $r=10$ per mille. Cioè si può usare la relazione:

$$GRR = \frac{7.45 - 5}{10 - 5} * (2.406 - 2.089) + 2.089 = 2.244$$

In questo esempio si è cercato il GRR relativo al sesso femminile e i dati citati sono desunti dalla corrispondente tavola del modello sud, livello 8 di Coale-Demeny-Vaughan. E' bene precisare che i valori delle stime di CBR, CDR, GRR ed e_0 riportati nella tabella 4 del testo, corrispondono alle medie delle stime calcolate, rispettivamente per i maschi e per le femmine, seguendo la procedura interpolatoria indicata.

Bibliografia

- M. Breschi (1990), *La popolazione della Toscana dal 1640 al 1940. Un'ipotesi di ricostruzione*, Dipartimento Statistico, Università degli Studi di Firenze, Serie ricerche empiriche, n.17, Firenze.
- A. Coale - P. Demeny - B. Vaughan (1983), *Regional Model Life Tables and Stable Populations*, Second Edition, Academic Press, New York.
- L. Del Pantà (1982), *Il modello della popolazione stabile: applicazioni in demografia storica*, in A. Santini - L. Del Pantà, Problemi di analisi delle popolazioni del passato in assenza di dati completi, Editrice CLUEB, Bologna.
- M. Demonet - J. Dupâquier - H. Le Bras (1977), *A Repertory of Stable Populations*, in R. D. Lee (Ed.), *Population Patterns in the Past*, Academic Press, New York.
- R. Lee (1974), *Estimating Series of Vital Rates and Age Structure from Baptisms and Burials: A New Technique, with Application to Pre-industrial England*, *Population Studies*, 28 (3).
- M. Livi Bacci (1967), *Sull'applicazione delle tecniche di analisi basate sulla teoria della popolazione stabile agli studi di demografia storica*, Atti della Riunione scientifica della S.I.S., Bologna.
- M. Livi Bacci (1990), *Introduzione alla demografia*, Nuova edizione, Loescher Editore, Torino.

LORENZO DEL PANTA - DEBORAH ARGELLI

MODELLI DI MORTALITÀ IN DEMOGRAFIA STORICA: POSSIBILITÀ OFFERTE DALLE STATISTICHE TOSCANE PRE-UNITARIE

1. Premessa

L'impiego dei modelli, ed in particolare dei modelli di mortalità, è certamente una delle strade più produttive per arrivare a migliorare le nostre conoscenze riguardo ai meccanismi evolutivi delle popolazioni del passato. Il dibattito sull'opportunità di impiegare in demografia storica modelli - a partire da quello della popolazione stabile (Del Pantà 1982a) - ideati principalmente per applicazioni a popolazioni contemporanee è aperto oramai da alcuni decenni (Van De Walle 1972). Si può affermare che la tecnica di ricostruzione aggregativa proposta da R. Lee (1974) ha dato nuovo impulso a questo filone di ricerca, ed in particolare ha riproposto all'attenzione dei demografi storici il problema dell'adattamento dei modelli di mortalità (in particolare le tavole-tipo) alle specifiche condizioni cui erano soggette le popolazioni del passato (Santini 1982), con particolare riferimento agli effetti che le grandi crisi epidemiche o alimentari potevano avere sulla struttura della mortalità. In questa ottica, particolarmente efficace sembra essere l'impiego del modello proposto da Brass (1971), anche se risultati non disprezzabili possono essere ottenuti in molti casi utilizzando, ad esempio, le ben note tavole-tipo di Coale e Demeny (1983) o quelle (altrettanto note, e per certi aspetti anche più idonee all'impiego nell'ambito della demografia storica) di Ledermann (1969).

Questa nota, nella quale si riportano alcuni risultati di una ricerca tuttora in corso, si inserisce nella problematica cui abbiamo adesso accennato. Ci proponiamo infatti di dar conto di un tentativo di individuare strutture tipiche di mortalità - differenziate sul piano geografico ed ambientale - all'interno del territorio toscano nella prima metà del XIX secolo, utilizzando una fonte di dati (della quale daremo qui alcuni cenni) che consente di ricavare indicazioni sufficientemente attendibili riguardo ai livelli e alle strutture della mortalità, proprio grazie all'ausilio della modellistica demografica.

2. Le "statistiche decennali" e il loro impiego per lo studio della mortalità

Le "statistiche decennali" del Granducato di Toscana¹, che hanno fornito la base-dati per il nostro studio, fanno parte di quella vasta mole di statistiche aggregate che furono ricavate dallo spoglio dei dati individuali dello stato civile². Questa fonte offre ampie possibilità di analisi dettagliate sul piano territoriale, in quanto fornisce, per ciascuna delle circa 250 comunità del Granducato, i dati annuali (dal 1818 al 1847) relativi alla popolazione residente complessiva, alle nascite e ai decessi distinti per sesso (sia i decessi che le nascite sono

¹ Le "statistiche decennali" sono conservate nel fondo *Stato civile toscano* (filze 12182 e 12183) dell'Archivio di Stato di Firenze.

² Una prima analisi della "geografia" della mortalità toscana basata sulle "statistiche decennali" è stata compiuta da M. Breschi e S. Salvini (1993). Nell'ambito del progetto "Modelli di sviluppo demografico in Italia tra XVIII e XIX secolo", del quale questo volume fornisce i primi risultati, si è costituito un sottogruppo di lavoro (del quale è responsabile il prof. A. Santini, e al quale partecipano tra gli altri anche Breschi e Salvini) specificamente orientato allo studio dei modelli di mortalità in demografia storica.

riferiti alla popolazione residente), ai matrimoni, ai nati illegittimi e (ciò che più interessa per i nostri obbiettivi) ai decessi ripartiti per gruppi di età³. Purtroppo, non disponiamo invece della popolazione ripartita per età. Ci troviamo dunque in un caso tipico di disponibilità di dati parziali per l'analisi della mortalità. La classificazione per età della popolazione è in realtà desumibile, per il solo anno 1841, dal censimento nominativo⁴. Ferma restando, dunque, la possibilità di

³ La ripartizione per età dei decessi indicata nelle statistiche decennali è la seguente: da un giorno a un anno; dai 2 ai 5 anni; dai 5 ai 10 anni; dai 10 ai 20 anni; successivamente analoghe classi decennali fino a: dai 90 ai 99 anni; centenari; di età ignota. Non è possibile in questa sede fornire dettagli riguardo ai problemi di interpretazione e valutazione della fonte. Possiamo solo aggiungere che nell'ultimo dei tre decenni per i quali vengono forniti i dati (1838-47) sono anche indicati separatamente i nati-morti (inclusi invece tra i morti nella prima classe di età per i primi due decenni). Sono state effettuate (sulla base di informazioni desunte da controlli sui dati individuali dello stato civile) una serie di operazioni di aggiustamento e di stima per riportare i dati ad una idonea classificazione. Per chiarimenti e dettagli su queste operazioni, e più in generale sulle caratteristiche della fonte utilizzata, rinviamo alle due tesi di laurea che hanno consentito di realizzare la memorizzazione completa del documento (cfr. F. Bonventi, *Analisi delle strutture di mortalità in Toscana sulla base delle "statistiche decennali 1818-47"*, tesi di laurea in Scienze statistiche e demografiche (rel. L. Del Panta), Facoltà di Scienze statistiche, demografiche ed attuariali, Università di Bologna, a.a. 1991-92; D. Argelli, *Modelli di mortalità in demografia storica: un'applicazione alle statistiche toscane preunitarie*, tesi di laurea in Scienze statistiche e demografiche (rel. L. Del Panta), Facoltà di Scienze statistiche, demografiche ed attuariali, Università di Bologna, a.a. 1992-93.

⁴ Il censimento nominativo eseguito nel Granducato di Toscana nel 1841 è conservato presso l'Archivio di Stato di Firenze. Nel periodo successivo all'esecuzione del censimento, non fu purtroppo eseguito uno spoglio completo dei principali caratteri rilevati. Attualmente, nell'ambito di una

utilizzare i dati relativi al 1841, abbiamo in linea di massima puntato a sfruttare le nostre statistiche per la costruzione di tavole di mortalità con il *metodo dei decessi* che, come è noto (Santini 1982), consente di prescindere, sotto certe condizioni, dalla conoscenza della popolazione classificata per età.

Abbiamo dunque deciso di ricavare dalle nostre statistiche una serie di tavole di mortalità (abbreviate), riferite all'intero trentennio 1818-47, che consentissero - pur rinunciando in questa fase ad analisi di tipo diacronico - di evidenziare le differenze di tipo territoriale. In sostanza, abbiamo operato una duplice aggregazione delle circa 250 comunità, arrivando a definire dapprima 148 circoscrizioni (ponendo come vincolo un livello minimo di popolazione di 4000 abitanti⁵), e quindi 17 "vallate", aree assai più ampie ricavate (con alcune modifiche) dall'atlante dello Zuccagni Orlandini (1832).

Per ciascuna circoscrizione abbiamo dunque costruito una tavola di mortalità, applicando il metodo dei decessi. Il problema principale da risolvere è stato, evidentemente, quello di valutare il tasso d'incremento r da utilizzare per correggere la distribuzione per età dei decessi. In sostanza, nell'ipotesi che l'andamento della popolazione di ciascuna circoscrizione sia stato, quanto meno, caratterizzato da una relativa *stabilità*, si trattava di valutare, di volta in volta, di quanto tale andamento si discostasse dalla *stazionarietà*, che avrebbe determinato una perfetta corrispondenza tra le nascite (che

ricerca diretta dal prof. Carlo Corsini è in corso la memorizzazione integrale del documento (il censimento fu effettuato su base parrocchiale, e le parrocchie raggruppate per comunità). Il prof. Corsini ha messo a nostra disposizione i dati sulla struttura per età della popolazione relativi ad una serie di comunità, per le quali è stato quindi possibile costruire delle tavole di mortalità partendo dal calcolo di tassi specifici per classi di età.

⁵ Si tenga conto del fatto che i decessi utilizzati per costruire le tavole si riferiscono ad un intero trentennio: ciò significa che la numerosità "virtuale" di ciascuna circoscrizione è di almeno 120 mila abitanti.

costituiscono la "radice" della tavola) ed il totale dei decessi, rendendo dunque inutile ogni correzione.

In effetti, per la Toscana esistono, soprattutto a partire dall'inizio del periodo lorenese (1737), numerose rilevazioni di tipo censuario - più o meno disaggregate dal punto di vista territoriale - che consentirebbero di ricavare qualche stima sulla velocità di crescita della popolazione delle diverse circoscrizioni anche con riferimento ad un periodo (precedente a quello per il quale si intende valutare la mortalità col metodo dei decessi) sufficientemente ampio⁶. Una stima non troppo imprecisa dei tassi di crescita delle 148 circoscrizioni da noi definite, per un periodo di circa un secolo (approssimativamente tra la metà del XVIII e la metà del XIX secolo), avrebbe richiesto però un impegno molto oneroso. Abbiamo ritenuto accettabile, in questa prima fase della ricerca, limitarsi a considerare la velocità di crescita di ciascuna circoscrizione nel trentennio in esame (1818-47), per il quale disponiamo (anno per anno) dell'ammontare della popolazione oltre che delle nascite.

3. Sull'importanza di una corretta stima del tasso r nel calcolo di tavole di mortalità col metodo dei decessi

I tassi r utilizzati nella costruzione delle tavole col metodo dei decessi si riferiscono dunque allo stesso periodo (il trentennio 1818-47) per il quale sono disponibili i decessi classificati per età. Si tratta, come è facile comprendere, di un periodo troppo breve, in quanto i decessi sui quali si opera derivano in realtà da nascite avvenute lungo un arco di tempo assai più ampio. Le tavole che abbiamo

⁶ Tenendo conto che i dati relativi ai decessi per età si riferiscono al trentennio 1818-47, sarebbe utile riuscire a conoscere il tasso di crescita della popolazione all'incirca tra il 1733 e il 1833 (anno intermedio tra il 1818 e il 1847).

costruito soffrono quindi, probabilmente, di una sorta di "errore sistematico", dovuto essenzialmente al fatto che il periodo per il quale disponiamo dei dati (1818-47) è un periodo di forte crescita della popolazione, mediamente ben più vivace rispetto al periodo precedente (Del Panta 1982b).

Si può quindi dire con relativa certezza che le tavole costruite per le diverse circoscrizioni sopravvalutano in generale la speranza di vita alla nascita, ed in particolare è probabile che le probabilità di morte nelle prime classi di età siano sottostimate. Del resto, l'interesse dei dati calcolati risiede soprattutto nel fatto che è possibile mettere in evidenza una serie di *differenziazioni geografiche nei livelli e nella struttura della mortalità*, che probabilmente si conserverebbero in gran parte anche se fosse possibile calcolare l'effettivo tasso di incremento di ogni circoscrizione. La distorsione interessa infatti, probabilmente, in misura analoga la maggior parte delle circoscrizioni considerate.

Per cercare di apprezzare il margine di errore cui si è andati presumibilmente incontro, per il fatto di non conoscere l'effettivo tasso di crescita della popolazione delle singole circoscrizioni, si è pensato di fare alcune valutazioni meno imprecise, operando a livello di intero Granducato. Considerando la Toscana (granducale) nel suo complesso, infatti, si offrono diverse possibilità di arrivare ad una valutazione dei rischi di morte, e quindi di costruire delle tavole di mortalità. Una prima possibilità deriva dal fatto che si dispone (Bandettini 1956) della distribuzione per età della popolazione stimata (sulla base di un'indagine campionaria) alla data del censimento del 1841⁷.

⁷ I dati presentati da Bandettini non differiscono in maniera rilevante da quelli che è possibile trarre dal censimento del 1881, né da quelli calcolati, mediante la tecnica dell'*inverse projection*, da Breschi (1990) nel suo lavoro di ricostruzione della popolazione dell'intera Toscana.

Utilizzando questi dati, si è quindi costruita una prima tavola per il triennio 1840-42 (si tratta della tavola (a) della tabella 1), partendo dal calcolo di tassi specifici di mortalità, e successivamente stimando le relative probabilità di morte.

La seconda tavola (b) è quella costruita - sempre per l'intero Granducato, ma per il periodo 1818-47 - col metodo dei decessi generalizzato, sulla base del tasso medio annuo d'incremento calcolato sul medesimo trentennio. Si tratta, in sostanza, della tavola generale calcolata con le stesse modalità con cui sono state calcolate tutte le 148 tavole delle singole circoscrizioni e le 17 tavole delle "vallate".

Si è quindi pensato di costruire un'ulteriore tavola (c) sfruttando, nella maniera più analitica possibile, le informazioni disponibili relative al ritmo di crescita della popolazione del Granducato lungo l'arco secolare che precede il periodo per il quale possiamo disporre delle statistiche sui decessi. Abbiamo perciò "riproporzionato" i decessi delle diverse classi di età non con un tasso r costante, ma applicando tassi variabili che tenessero conto della velocità media di crescita della popolazione tra gli anni di nascita degli individui deceduti in ogni classe di età e il periodo⁸ per il quale si considerano i decessi⁹.

⁸ Nel nostro caso si tratta di un trentennio, ragione per la quale le stime sono necessariamente approssimative.

⁹ La scelta dei periodi ha dovuto, ovviamente, tener conto anche dell'esistenza di "enumerazioni" che fornissero i dati occorrenti. Ad esempio, per la classe di età 50-59 è stato calcolato un tasso d'incremento medio annuo r (pari a 0.0058) relativo all'intervallo 1765-1845, in quanto per il 1765 si conosce l'ammontare della popolazione; analogamente, per la classe 10-19, il calcolo è stato fatto sul periodo 1806-45 ($r = 0.0082$) perché è noto il dato di popolazione dell'anno 1806. Per alcune classi, il calcolo del tasso d'incremento è stato fatto partendo da dati ottenuti tramite interpolazione.

Tabella 1: Granducato di Toscana: confronto tra indicatori di mortalità ricavati da tavole costruite con metodi differenti

età	numeri indice delle probabilità di morte			
	(a)	(b)	(c)	(d)
0-4	100.0	74.9	88.8	88.2
5-9	100.0	80.4	104.3	102.3
10-19	100.0	83.7	107.7	103.9
20-29	100.0	87.4	107.3	104.5
30-39	100.0	87.9	106.6	101.5
40-49	100.0	92.6	107.8	103.4
50-59	100.0	93.5	103.3	101.0
60-69	100.0	98.4	104.2	103.1
e_0	29.4	36.7	31.4	31.9
q_0	0.4680	0.3505	0.4155	0.4130

- (a) tavola (1840-42) costruita disponendo d. popolaz. per età
 (b) metodo dei decessi (r cost. relativo al periodo 1818-47)
 (c) metodo dei decessi (r variabile)
 (d) metodo dei decessi (r cost. relativo al periodo 1745-852)

Infine, abbiamo calcolato un'ultima tavola (d) (ancora col metodo dei decessi) con un tasso r fisso pari all'incremento medio annuo verificatosi in Toscana tra il 1745 e il 1852. Questo periodo¹⁰

Si rammenti, tuttavia, che la popolazione è conosciuta per tutti gli anni a partire dal 1818.

¹⁰ Il periodo 1745-1852 è stato scelto perché, per tali date, era già disponibile una valutazione "a confini costanti" della popolazione del Granducato (Del Panta 1982b).

corrisponde abbastanza bene all'intervallo da cui derivano le diverse coorti di nati che danno origine ai decessi utilizzati per la costruzione delle tavole.

Nella Tabella 1 sono riportate (insieme ad altri indicatori) le variazioni percentuali delle probabilità di morte delle tre tavole (b, c, d) costruite col metodo dei decessi rispetto a quella (a) costruita utilizzando il censimento.

In sintesi, si può affermare che - come era prevedibile - la tavola che più si discosta da quella costruita utilizzando il censimento (e che si può ritenere essere sufficientemente affidabile) è la tavola (b) costruita, col metodo dei decessi generalizzato, usando come tasso d'incremento r quello relativo al periodo 1818-47. Questa tavola sottostima sensibilmente la mortalità (la distorsione più macroscopica si ottiene per la classe di età 0-4) in ragione del fatto che il tasso r è chiaramente sovrastimato. La mortalità infantile calcolata in questa tavola, in particolare, deriva da un confronto tra decessi stimati in maniera abbastanza corretta (il tasso r che è stato utilizzato per correggere i decessi delle prime classi di età è abbastanza vicino a quello reale) e "nascite" ottenute come somma di decessi "riproporzionati" (per ricondursi all'ipotesi di stazionarietà) in funzione di un tasso di crescita che - man mano che cresce l'età - è sempre più elevato rispetto a quello effettivo.

E' infine interessante osservare che le tavole (c) e (d), ottenute applicando il metodo dei decessi con tasso r variabile e con tasso r fisso, ma correttamente calcolato su un periodo di circa un secolo (procedendo all'indietro rispetto al periodo per il quale si costruisce la tavola), mostrano andamenti della mortalità sostanzialmente analoghi.

Queste due tavole denotano una differenza abbastanza sensibile (superiore al 10 per cento) nella probabilità di morte 0-4 rispetto alla tavola (a) (costruita utilizzando il censimento del 1841), ma in questo caso si può pensare che anche il valore ricavato dalla tavola (a) non

sia del tutto affidabile¹¹. Per le altre classi, si può dire che le differenze sono veramente molto ridotte¹². In sostanza, si può affermare che - qualora si sia in grado di conoscere l'effettivo tasso di crescita della popolazione per un periodo di circa un secolo (precedente all'anno o al periodo per il quale si dispone della distribuzione per età dei decessi) - il metodo dei decessi appare fornire (è questo almeno il caso da noi esaminato) risultati soddisfacenti, sia che si consideri un tasso r fisso per tutto il periodo, sia che si considerino tassi variabili per le singole classi di età. Ovviamente, tanto meno *stabile* è l'andamento della popolazione nel corso del tempo, tanto più sarà invece importante poter conoscere tale andamento con un certo dettaglio, in modo da effettuare il "riproporzionamento" dei decessi con tasso variabile per le singole classi di età.

4. Un tentativo di individuare strutture tipiche di mortalità all'interno del territorio granducale

Pur ammettendo che le tavole che abbiamo ricavato col metodo dei decessi sono probabilmente affette da un certo grado di sottostima del

¹¹ Infatti la struttura per età del censimento risulta abbastanza anomala nelle prime classi, e sarebbe forse stato più opportuno (come è peraltro possibile fare) calcolare separatamente la probabilità di morte nel primo anno di vita partendo dai nati.

¹² Del resto, è ovvio che non possiamo attenderci una somiglianza perfetta comparando tavole costruite - oltre che con metodi differenti - anche per periodi non coincidenti (il triennio 1840-42 per la tavola (a), il trentennio 1818-47 per le altre tavole).

livello della mortalità¹³, abbiamo voluto effettuare un primo tentativo di individuare, all'interno del territorio granducale, strutture tipiche di mortalità, che potessero eventualmente risultare caratteristiche di aree geografiche in qualche modo definibili da un punto di vista economico e/o ambientale. Partendo dalle 148 tavole costruite col metodo dei decessi, dovevamo quindi, per prima cosa, individuare una tavola di riferimento alla quale poter riferire, per confronto, le probabilità di morte di ciascuna delle nostre tavole. Per individuare questa tavola di riferimento abbiamo in sostanza seguito il procedimento preliminare seguito da Coale e Demeny (1983) per arrivare alla costruzione delle tavole-tipo (delle quattro "regioni"). Si sono cioè considerate separatamente le probabilità di morte di ciascuna delle dieci classi di età che le "statistiche decennali" consentono di considerare. Sono stati poi ordinati i 148 valori di qx , dal più alto al più basso, per ogni classe di età, e sulla base di questi ordinamenti si sono costruite delle tavole di mortalità preliminari (teoriche)¹⁴. Si è infine considerata come tavola standard (di riferimento) quella che, tra tutte le tavole preliminari, presentava il valore di e mediano¹⁵.

Utilizzando come termine di confronto questa tavola standard, per ciascuna classe di età si sono calcolati gli scarti tra le probabilità di morte di questa tavola di riferimento e quelle corrispondenti di ciascuna delle 148 tavole empiriche. Il successivo esame comparato

¹³ Ciò che disturba maggiormente, per l'utilizzazione che abbiamo cercato di fare di questa numerosa serie di tavole, è il fatto che, presumibilmente, il livello di sottostima è più accentuato per la mortalità infantile.

¹⁴ Questo sistema di costruzione delle tavole preliminari ha il vantaggio della semplicità e quello di trattare in maniera simmetrica le probabilità di morte alle diverse età (Coale-Demeny 1983).

¹⁵ E' da notare che questa tavola è risultata molto simile a quella costruita, per l'intero Granducato, utilizzando la popolazione per età fornita da Bandettini (1956)

degli scostamenti nelle varie classi di età, riferiti alle singole circoscrizioni, ha infine consentito di individuare alcuni raggruppamenti di circoscrizioni che risultavano accomunate non tanto da un analogo livello di mortalità quanto da una struttura abbastanza simile.

La Tabella 2 sintetizza il risultato di questa operazione. In sostanza, all'incirca un centinaio di circoscrizioni (sulle 148 considerate) sembrano potersi collocare nell'ambito di specifiche strutture di mortalità, anche se solo alcuni dei cinque "modelli" individuati presentano caratteristiche assai marcate.

Tabella 2: Valor medio e varianza di e_0 per cinque gruppi di circoscrizioni (modelli) caratterizzati da strutture "tipiche" di mortalità

modelli	media	var.	n° circ.	descrizione
	e_0	e_0		
A	33.7	2.9	16	mortal.infant. alta mortal. adulti bassa
B	35.8	1.3	9	mortal. infant. bassa mortal. adulti alta
C	35.3	18.4	10	"modello malaria"
D	40.5	6.2	48	mortalità bassa ad ogni età
E	29.4	2.8	18	mortalità alta ad ogni età

Tra i cinque gruppi di circoscrizioni che abbiamo potuto individuare, quello che è indubbiamente più caratterizzato dal punto di vista geo-ambientale è il gruppo C, nel quale si collocano in pratica gran parte delle comunità della Maremma litoranea, e cioè quelle più

infestate dalla malaria. Occorre peraltro notare che in questo gruppo risulta molto elevata la variabilità del livello della mortalità, contrariamente a quanto si riscontra negli altri quattro gruppi per i quali ad una somiglianza di struttura corrisponde anche una relativa omogeneità in termini di livello della mortalità¹⁶. Come è noto, la malaria incide, oltre che sulla mortalità dei bambini, anche e forse più - pur se in via indiretta - su quella degli adulti. In effetti le 10 circoscrizioni del gruppo C mostrano una struttura abbastanza simile alle 9 circoscrizioni del gruppo B, con mortalità proporzionalmente più elevata nelle età adulte. Alcune di queste ultime, tra l'altro, si collocano nell'area senese, e dunque sono geograficamente assai prossime a quelle maremmane. I territori del senese, peraltro, sono soprattutto contrassegnati dalla presenza di circoscrizioni incluse nel gruppo E, quello contraddistinto da una mortalità elevata praticamente in ogni classe di età. E' ben noto, del resto, che insieme alla Maremma la provincia senese è stata caratterizzata in passato da livelli di mortalità assai più elevati a confronto con la parte settentrionale della regione.

Gli altri due gruppi individuati riguardano da un lato le numerose (48) circoscrizioni (gruppo D) caratterizzate da livelli di mortalità che ad ogni età risultano inferiori alla media toscana, dall'altro un certo numero di circoscrizioni (gruppo A) per le quali si riscontra una mortalità infantile particolarmente bassa che si accompagna però ad

¹⁶ La forte variabilità del livello di e_0 nel gruppo di circoscrizioni maremmane si accompagna ad un valor medio di non molto inferiore a quello del Granducato nel suo complesso: secondo la tavola calcolata col metodo dei decessi (con r relativo al trentennio 1818-47, e quindi senz'altro sovrastimato) il valore di e_0 per il Granducato risulta infatti di 36.7. E' possibile, in realtà, che l'errore nella stima del tasso di crescita sia assai più sensibile per le circoscrizioni maremmane, in quanto per questi territori la differenza nel ritmo di crescita tra la seconda metà del XVIII e la prima metà del XIX è senz'altro molto netta.

una mortalità degli adulti superiore alla norma. Le circoscrizioni del gruppo D si collocano tutte nella parte settentrionale della Toscana. In particolare, ritroviamo in questo "modello" zone ben note di bassa mortalità quali una vasta area che da Pisa si estende soprattutto verso nord-est, il Casentino e gran parte dei territori dell'Appennino orientale (in direzione della Val Tiberina), buona parte della Montagna Pistoiese, infine i territori della Lunigiana e della Garfagnana soggetti al Granducato. Assai meno netta è la delimitazione geografica delle circoscrizioni del gruppo A, che comunque si collocano anch'esse in prevalenza nelle aree centro-settentrionali, ed in buona parte sono situate tra il valdarno aretino e la pianura che da Firenze si estende in direzione di Pistoia.

Come si è già sottolineato, questo quadro potrebbe essere modificato in misura forse non irrilevante quando si potessero ricalcolare le 148 tavole delle singole circoscrizioni sulla base di una informazione più precisa e completa riguardo al tasso medio di crescita per un periodo più ampio di tempo. Non è dunque il caso, in questa breve nota preliminare, di ampliare il commento dei risultati ottenuti, mentre sarebbe utile delineare altre possibili vie da seguire per utilizzare al meglio la base di dati che abbiamo a disposizione. Sotto questo aspetto, però, l'assoluta mancanza di spazio ci consente appena di accennare alle possibilità offerte dall'impiego del modello di Brass (1971)¹⁷, che ha consentito innanzitutto di verificare il buon

¹⁷ Definita la funzione:

$$\text{logit } (y) = 1/2 \ln y/(1-y)$$

Brass ha sperimentato che, per due popolazioni qualsiasi P ed S, è in genere verificata molto strettamente la relazione:

$$\text{logit } ({}^P q_x^0) = a + b \text{logit } ({}^S q_x^0)$$

adattamento delle 17 tavole costruite, per le "vallate" del Granducato, col metodo dei decessi rispetto a quelle che, utilizzando come parametri d'ingresso i valori calcolati di q_0 e di ${}_{10}q_{40}$, sono state appunto ricavate tramite la funzione proposta da Brass. Confortati da questo risultato, abbiamo inoltre predisposto (a titolo essenzialmente esemplificativo) due distinte "reti" di tavole, la prima caratterizzata dal valore costante del parametro a di Brass di + 0.71 (si tratta del valore più basso riscontrato tra quelli delle 17 vallate) e da valori variabili del parametro b , la seconda, al contrario, dal valore più elevato (+ 1.36) del parametro a e ancora da valori variabili del parametro b . Si tratta, in sostanza, di un primo tentativo di fornire uno strumento operativo molto semplice per poter operare degli immediati raffronti con situazioni reali, anche se questo genere di elaborazione andrà senz'altro riproposto nel momento in cui sarà meno incerto il grado di affidabilità sia della tavola standard¹⁸ che delle tavole in base alle quali sono stati scelti i due valori del parametro a .

Disponendo, in pratica, di una tavola di mortalità *standard* (S), si possono stimare i parametri a e b disponendo di due sole informazioni (parziali) sulla mortalità della popolazione P, e dunque ricavare l'intera tavola. In pratica, il parametro a varia attorno al valore 0 (in genere tra -0.8 e +0.8): un valore maggiore di 0 significa una mortalità più elevata rispetto allo *standard* (a è, dunque, un parametro di "livello"). Il parametro b varia invece attorno al valore 1 (in genere tra 0.7 e 1.4): un valore maggiore di 1 significa mortalità proporzionalmente più bassa tra i giovani e più elevata tra gli anziani, viceversa per un valore minore di 1 (b è, insomma, un parametro di "struttura").

¹⁸ E' stata usata come standard, in questo caso, la tavola di mortalità costruita per l'intero Granducato col metodo dei decessi.

Bibliografia

P. Bandettini (1956), *La popolazione della Toscana alla metà dell'Ottocento*, Archivio Economico dell'Unificazione Italiana, Roma

W. Brass (1971), *On the Scale of Mortality*, in W. Brass (ed.), *Biological Aspects of Demography*, Taylor and Francis Ltd., London

M. Breschi (1990), *La popolazione della Toscana dal 1640 al 1940. Un'ipotesi di ricostruzione*, Dipartimento Statistico, Università degli Studi di Firenze

M. Breschi - S. Salvini (1993), *Differenze territoriali nella mortalità del Granducato di Toscana nella prima metà dell'Ottocento*, in S.I.D.E.S., *La popolazione delle campagne italiane in età moderna*, CLUEB, Bologna

A. J. Coale - P. Demeny (1983), *Regional Model Life Tables and Stable Populations*, second edition, Academic Press, New York

L. Del Panta (1982a), *Il modello della popolazione stabile: applicazioni in demografia storica*, in A.Santini-L.Del Panta, *Problemi di analisi delle popolazioni del passato in assenza di dati completi*, CLUEB, Bologna

L. Del Panta (1982b), *Chronologie de la croissance démographique urbaine: le Grand-Duché de Toscane entre la moitié du XVI et la moitié du XIX siècle*, "Genus", vol. XXXVIII, n.1-2

S. Ledermann (1969), *Nouvelles tables-types de mortalité*, INED, Paris

R. Lee (1974), *Estimating Series of Vital Rates and Age Structure from Baptisms and Burials: a New Technique, with Applications to Pre-Industrial England*, "Population Studies", 28 (3)

A. Santini (1982), *Problemi di analisi della mortalità*, in A. Santini - L. Del Panta, *Problemi di analisi delle popolazioni del passato in assenza di dati completi*, CLUEB, Bologna

E. Van De Walle (1972), *De l'emploi des modèles en démographie historique*, "Annales de démographie historique"

A. Zuccagni Orlandini (1832), *Atlante geografico, fisico e storico del Granducato di Toscana*, Firenze

LORENZO DEL PANTA - MARIA ELENA FORINI

**DISPONIBILITÀ ALIMENTARI E MORTALITÀ IN ITALIA
TRA L'UNITÀ E LA GRANDE GUERRA:
PROBLEMI DI ANALISI DI SERIE TEMPORALI**

1. Premessa

Nel periodo compreso tra l'unificazione nazionale (1861) e la prima guerra mondiale, il sistema demografico italiano è caratterizzato da trasformazioni rilevanti (Del Panta 1984). In particolare, la struttura ed il livello della mortalità iniziano a modificarsi molto rapidamente a partire dagli ultimi decenni del XIX secolo. Una serie di studi anche recenti ha ormai delineato in modo abbastanza completo la dinamica del declino della mortalità, così come sono state analizzate con notevole dettaglio le tendenze delle principali cause di morte¹. Il dibattito riguardo ai fattori che hanno contribuito a determinare le tendenze della mortalità, nel periodo che ha visto sul piano nazionale l'inizio del declino irreversibile e insieme l'emergere di nuove differenziazioni sul piano territoriale, rimane però ancora aperto. Del resto, qualsiasi tentativo di discutere la realtà italiana in tale periodo

¹ Non è possibile fornire qui una bibliografia, sia pure parziale, sul tema in questione. Ci limitiamo a citare un lavoro abbastanza recente (1990) di G. Caselli, che ha apportato i contributi più importanti riguardo all'analisi e alle possibili spiegazioni del declino della mortalità in Italia. Occorre inoltre citare la tesi di dottorato di L. Pozzi (1989-90), che ha affrontato in particolare il tema dell'interpretazione delle differenze territoriali nella struttura per cause della mortalità.

deve riferirsi anche ad un contesto temporale e territoriale più ampio, e far tesoro di quanto è già stato scritto riguardo ai fattori esplicativi della mortalità e più in generale ai legami tra sistema demografico, sistema economico ed ecosistema². Anche se è evidente che il declino della mortalità è la conseguenza di un insieme di fattori, che interagiscono tra di loro e che sono, spesso, difficilmente quantificabili e misurabili, qualche contributo di conoscenza può talvolta derivare anche da analisi che, isolando un determinato fattore - nel caso specifico il fattore "alimentare" - tentino di quantificarne l'influenza³.

Occorre in effetti considerare che, mentre per quanto riguarda gli indicatori delle disponibilità alimentari si dispone di dati seriali (per quanto scarsamente affidabili⁴), che possono essere messi in relazione con gli indicatori della mortalità, ciò non avviene per quanto riguarda eventuali indicatori relativi alle condizioni igienico-sanitarie o ad altri fattori, indubbiamente importanti per spiegare il declino della mortalità, quali ad esempio la diffusione delle conoscenze in campo medico o il grado di affollamento delle abitazioni.

Nell'ambito di uno studio rivolto ad indagare - per la popolazione italiana nel periodo sopra indicato - i possibili legami tra regime

² Anche in questo caso, nell'impossibilità di fornire indicazioni bibliografiche dettagliate, ci limitiamo a rinviare ad un recente lavoro di M. Livi Bacci (1989) che sintetizza in maniera molto efficace i problemi in questione.

³ Il tema delle relazioni tra livello nutritivo e mortalità in Italia tra XIX e XX secolo è stato già affrontato da L. Del Panta e A. Samoggia (1991), sia con un primo tentativo di analisi diacronica a livello nazionale che con un'analisi territoriale che ha utilizzato anche indicatori alimentari tratti dall'inchiesta sanitaria effettuata nel 1885.

⁴ I dati seriali (disponibili a livello nazionale a partire dal 1861) cui ci riferiamo sono quelli pubblicati dall'ISTAT (1958). Quanto alla valutazione sul loro grado di affidabilità rinviemo a Del Panta-Samoggia (1991).

alimentare e mortalità, ci proponiamo qui di fissare l'attenzione sull'impiego di alcuni *modelli econometrici*, non frequentemente utilizzati nel contesto della ricerca demografica italiana, che consentono di verificare l'esistenza o meno di tali legami.

Disponendo di serie storiche annuali relative sia ad indicatori della mortalità⁵ che ad indicatori delle disponibilità alimentari⁶, possiamo dunque affrontare un'analisi statistica per cercare di capire se le tendenze della mortalità siano in relazione con quelle del presunto regime alimentare⁷. Occorre però precisare che un'analisi basata sui dati ora richiamati può in realtà avere più obiettivi, sostanzialmente diversi tra loro. Da un lato, ci possiamo infatti domandare se, nel generale processo di declino della mortalità che tendenzialmente si verifica in Italia tra la data dell'unificazione nazionale e la Grande Guerra, un peso rilevante sia attribuibile ad eventuali miglioramenti nel regime alimentare. Dall'altro, possiamo invece tentare di verificare se le variazioni congiunturali dei diversi indicatori di mortalità utilizzati risultano influenzate da corrispondenti variazioni riscontrabili (nei medesimi anni o in anni precedenti) in termini di disponibilità alimentari complessive o di singoli nutrienti⁸. In effetti, non è semplice scindere i due aspetti in

⁵ Sono stati utilizzati, in particolare, il tasso generico di mortalità e le probabilità di morte nel primo anno di vita, tra il primo e il quinto compleanno e tra il settantesimo e il settantacinquesimo compleanno.

⁶ Le statistiche pubblicate dall'ISTAT permettono tra l'altro di conoscere, anno per anno, la disponibilità giornaliera pro-capite di calorie (totali e di origine animale), oltre che quella di cereali, di legumi e di carne.

⁷ Un limite dei dati disponibili riguarda il fatto che non conosciamo l'effettivo consumo medio pro-capite di nutrienti, ma solo la sua disponibilità.

⁸ Sia il possibile legame "congiunturale" tra livello nutritivo e mortalità, che l'ipotesi che il fattore "alimentare" sia da considerare come prevalente per l'interpretazione delle tendenze di lungo periodo della mortalità sono

un'analisi statistica basata esclusivamente su serie di dati annuali⁹. Il periodo in esame, tra l'altro, è caratterizzato da una tendenza generale al declino per quanto riguarda la mortalità, mentre le diverse serie relative alle disponibilità alimentari presentano andamenti difformi, anche se la caratteristica più evidente è costituita da una forte diminuzione delle disponibilità alimentari pro-capite nel periodo della grande crisi agraria (tra l'inizio degli anni '80 e la fine del secolo XIX). Inizialmente, abbiamo dunque deciso di suddividere il periodo complessivamente considerato (1861-1921) in quattro sottoperiodi, differentemente caratterizzati sul piano economico generale e, presumibilmente, dal punto di vista della qualità e della quantità dei fattori nutritivi disponibili per la popolazione italiana¹⁰. Senza entrare qui nei dettagli, riguardo ai risultati dell'analisi di correlazione effettuata, per i singoli sottoperiodi, tra i diversi indicatori di mortalità e quelli delle

stati ampiamente dibattuti. Ci limitiamo qui a citare Mc Keown (1976), favorevole all'ipotesi alimentare, e Livi Bacci (1993), che invece respinge l'ipotesi alimentare soprattutto per quanto riguarda l'interpretazione di lungo termine delle tendenze della mortalità.

⁹ Abbiamo già sottolineato la carenza di indicatori seriali relativi ad altri fattori (oltre a quello alimentare) che possono aver influito sulla mortalità.

¹⁰ Il primo periodo comprende i 20 anni successivi all'unificazione, caratterizzati da una situazione economica scarsamente evolutiva. L'ultimo ventennio del secolo è caratterizzato dalla grande crisi agraria, che sembra ridurre soprattutto la disponibilità di calorie di origine non animale. Nel quindicennio che precede lo scoppio della guerra, anche in relazione al decollo industriale, si verifica un certo aumento del reddito pro-capite che favorisce qualche miglioramento (anche qualitativo) della dieta, con una parziale sostituzione dei cereali minori col frumento. L'ultimo breve periodo (1914-21) comprende gli anni della Grande Guerra ed è evidentemente caratterizzato da un generale peggioramento delle condizioni di vita e dei livelli nutritivi.

disponibilità alimentari, si può sinteticamente affermare che, nella maggioranza dei casi, non sono state riscontrate correlazioni significative (del segno atteso) tra indicatori alimentari e mortalità. Solo nel periodo 1901-13, i dati indicano l'esistenza di una correlazione negativa abbastanza elevata tra la disponibilità di calorie di origine animale e gli indicatori della mortalità.

Nell'intento di verificare, con strumenti statistici più idonei, queste prime sommarie indicazioni, abbiamo quindi deciso di utilizzare dei modelli econometrici a variabili ritardate, nell'ipotesi che le variazioni (ed in particolare eventuali brusche diminuzioni) delle disponibilità pro-capite di nutrienti potessero spiegare i loro effetti negativi sulla sopravvivenza della popolazione non solo nell'anno in cui esse si verificavano ma anche negli anni successivi¹¹.

2. I modelli utilizzati

Prima di illustrare i modelli utilizzati nella nostra analisi, occorre precisare che non è apparso opportuno, per la brevità delle serie, applicare tali modelli separatamente ai quattro sottoperiodi citati in precedenza, e che quindi l'analisi è stata in questo caso eseguita sull'intero arco di tempo 1861-1921. In questa breve nota, per motivi di spazio, non verranno forniti i risultati derivanti dall'applicazione dei modelli. Ci limiteremo a richiamare, come conclusione, il riscontro di un'assai scarsa influenza del regime alimentare sulle tendenze della mortalità.

¹¹ Un modello a ritardi distribuiti è stato impiegato anche da Lee (1981) in uno studio sulla relazione tra tassi demografici, prezzi e tempo atmosferico in Inghilterra nel periodo 1540-1840 e da Ferrari e Livi Bacci (1985) in un saggio sulla relazione tra temperature e mortalità nello stesso periodo da noi esaminato.

L'ipotesi da cui si parte è che la relazione tra le variabili sia lineare, per l'errore quindi si pone che:

$$E(e_t) = 0$$

$$E(e_t, e_t) = \sigma^2 I.$$

Il primo modello utilizza come regressori serie di dati espressi in termini di scarti dalle medie campionarie.

La variabile Y indica le serie di mortalità: di volta in volta sono state utilizzate le probabilità di morte nel primo anno di vita, tra il primo e il quinto compleanno, tra il settantesimo e il settantaquattresimo anno e il tasso generico di mortalità; la X indica le serie alimentari (disponibilità media pro capite di: cereali, legumi, carne, proteine e calorie) e $M(Y)$ ed $M(X)$ indicano rispettivamente le medie delle due serie.

Il primo modello (F1) è dunque:

$$\Delta Y_t = B_0 S_t + B_1 S_{t-1} + B_2 S_{t-2} + B_3 S_{t-3} + e_t$$

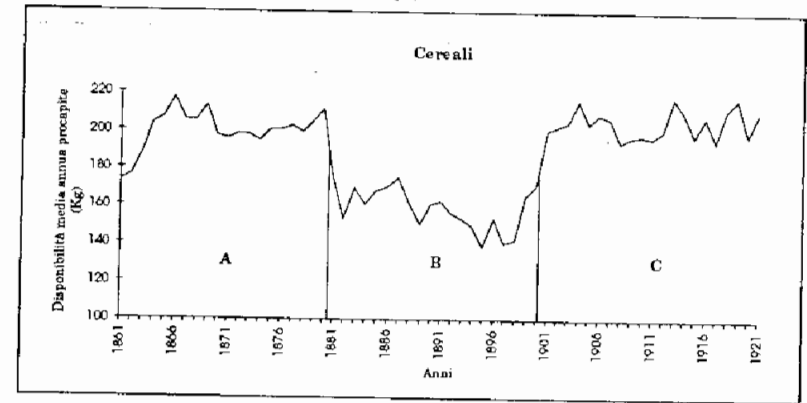
$$\text{dove } S_t = X_t - M(X)$$

Inoltre la non stazionarietà delle serie di mortalità è stata modellata dalle differenze

$$\Delta Y_t = Y_t - Y_{t-1}$$

Considerando poi che l'andamento delle disponibilità alimentari (X) ha subito un crollo nel periodo della crisi agraria (1880-1900), si è tenuto conto della significativa differenza (vedi Grafico 1) tra le medie dei periodi A e C da un lato e B dall'altro riferite agli indicatori alimentari.

Grafico 1



Si è ritenuto opportuno quindi utilizzare, nel modello precedente, per le disponibilità alimentari, gli scarti dalla media calcolati nei tre sottoperiodi indicati dal grafico, quindi:

$$S_t = X_t - M(X_j) \quad j=1, \dots, 3$$

Il modello (F1) a questo punto può essere riscritto come segue (F2):

$$\Delta Y_t = B_0 S_t + B_1 S_{t-1} + B_2 S_{t-2} + B_3 S_{t-3} + D_1 + D_2 + E_t$$

D_1 e D_2 sono due variabili *dummy*, dove il pedice 1 si riferisce ai periodi precedente e seguente la crisi, e il pedice 2 al periodo di crisi; i coefficienti di queste variabili dovrebbero verificare l'esistenza di una relazione tra dieta e mortalità nei periodi di crisi e di «normalità».

I modelli descritti hanno mostrato un buon adattamento in particolare a due serie di mortalità: probabilità di morte nel primo anno di vita e tasso generico di mortalità; mentre nel caso della mortalità tra il primo e il quinto compleanno e tra il settantesimo e

settantaquattresimo anno si verifica una forte autocorrelazione degli errori¹². In quest'ultimo caso probabilmente i termini di disturbo rappresentano l'influenza complessiva delle variabili omesse.

Si è considerato inoltre il caso in cui le serie alimentari siano integrate: a tale scopo è stato applicato il modello a variabili cointegrate Engle-Granger (1991), che tiene conto della eventuale non stazionarietà delle serie. Il modello è rappresentato dalle seguenti equazioni:

$$\Delta Y_t = M + NZ_{t-1} + \text{lags}(\Delta X_t, \Delta Y_t) + E_t$$

dove $\Delta X_t = X_t - X_{t-1}$

$$Z_t = Y_t - (A + BX_t)$$

L'applicazione del modello si divide in due parti:

1) Si stimano i coefficienti A e B dalla relazione tra i livelli di cointegrazione

$$Y_t = A + BX_t + Z_t$$

dove Y indica le serie di mortalità e X quelle alimentari. Successivamente si ricava la serie dei residui Z_t ; se quest'ultima non è stazionaria viene a mancare l'equilibrio.

2) Si formula poi l'equazione di regressione (F3):

$$\Delta Y_t = A_0 \Delta Y_{t-1} + B_0 \Delta X_t + B_1 \Delta X_{t-1} + B_2 \Delta X_{t-2} + B_3 \Delta X_{t-3} + F_1 Z_{t-1} + E_t$$

Z_t in valore assoluto rappresenta la distanza dalla condizione di equilibrio del sistema. Per verificare la stazionarietà della suddetta

¹² Per verificare l'autocorrelazione è stato utilizzato il test Durbin-Watson.

serie è stato applicato il *test di radice unitaria* Dickey-Fuller¹³ il quale ha mostrato la non stazionarietà della serie Z_t . Poiché l'errore Z_t è stato considerato irrilevante ai fini dell'analisi, si è ipotizzato $F_1 = 0$ nella (F3).

Per la mortalità tra il primo e il quinto compleanno si è reso necessario aggiungere, tra le variabili indipendenti un ulteriore lag per la mortalità allo scopo di ridurre l'autocorrelazione.

Come avevamo anticipato, in questo breve testo abbiamo ritenuto preferibile illustrare con qualche dettaglio le caratteristiche dei modelli utilizzati, rinviando ad un successivo lavoro l'analisi ed il commento dei risultati ottenuti. Nei casi in cui non sia presente l'autocorrelazione degli errori, l'applicazione dei modelli presentati conferma comunque quanto già era emerso dalle prime analisi di correlazione, e cioè che il fattore alimentare non sembra aver contribuito in misura rilevante a determinare le tendenze della mortalità della popolazione italiana nel periodo esaminato.

¹³ Il test (cfr. Engle-Granger 1991) è basato sull'equazione

$$\Delta Z_t = A + BZ_{t-1} + \sum_{i=1}^J C_i \Delta Z_{t-i} + D_t$$

dove J è stato assunto uguale a 4. L'ipotesi nulla è che Z_t sia non stazionaria ossia I(1) (integrata di ordine uno); la tavola contenente i valori critici è stata approntata da Dickey e Fuller: tali valori all'1% e al 5% sono, rispettivamente, -3.77 e -3.17. L'ipotesi viene confutata se B è negativo e significativamente diverso da zero.

Bibliografia

- G. Caselli (1990), *Mortalità e sopravvivenza in Italia dall'Unità agli anni '30*, in S.I.D.E.S., *Popolazione, società e ambiente*, CLUEB, Bologna
- L. Del Panta (1984), *Evoluzione demografica e popolamento nell'Italia dell'Ottocento (1796-1914)*, CLUEB, Bologna
- L. Del Panta - A. Samoggia (1991), *Malnutrition, Disease and Mortality in Late XIXth and Early XXth Century Italy*, comunicazione presentata al Summer Meeting della "Society for the Social History of Medicine" su *Famine and Disease*, Christ's College, Cambridge (5-7 July 1991)
- R. F. Engle - C. W. J. Granger (1991), *Long-Run Economic Relationship. Readings in cointegration*, Oxford Univ. Press, Oxford
- G. Ferrari - M. Livi Bacci (1985), *Sulle relazioni tra temperatura e mortalità nell'Italia unita, 1861-1914*, in S.I.D.E.S., *La popolazione italiana nell'Ottocento. Continuità e mutamenti*, CLUEB, Bologna
- ISTAT (1958), *Sommario di statistiche storiche italiane 1861-1955*, Roma
- J. Johnston (1993), *Econometrica*, a cura di M. Costa e P. Paruolo, Franco Angeli Editore, Milano
- R. Lee (1989), *Short-term variation: vital rates, prices, and weather*, in E. A. Wrigley - R. S. Schofield, *The Population History of England 1541-1871. A Reconstruction*, Cambridge University Press, Cambridge (seconda edizione)

M. Livi Bacci (1989), *Storia minima della popolazione del mondo*, Loescher, Torino

M. Livi Bacci (1993), *Popolazione e alimentazione. Saggio sulla storia demografica europea*, Il Mulino, Bologna (seconda edizione)

T. Mc Keown (1976), *The Modern Rise of Population*, Arnold, London (ed. italiana: *L'aumento della popolazione nell'era moderna*, Feltrinelli, Milano, 1979)

L. Pozzi (1989-90), *La mortalità per cause nelle province italiane dal 1890 al 1950*, tesi di dottorato in Demografia, Università di Firenze, Roma e Padova.

PAOLA GUERESI - FOSCA MARTUZZI VERONESI
DAVIDE PETTENER

**PROBLEMI DI METODO NELLE INDAGINI
BIODEMOGRAFICHE SU POPOLAZIONI
DI AMBIENTE MONTANO**

Introduzione

L'utilizzazione delle fonti proprie della demografia storica ha consentito un nuovo approccio allo studio delle popolazioni del passato anche da parte dell'antropologo, con particolare riferimento ai comportamenti matrimoniali; sono infatti tali comportamenti che traducono un fenomeno sociale nel fenomeno biologico della riproduzione. Poiché i comportamenti matrimoniali dipendono da numerosi fattori, culturali, religiosi ed ambientali in genere, il loro studio diviene ancor più interessante in ambiente montano, dove si possono determinare condizioni di isolamento riproduttivo, incremento dell'omozigosi, tendenza alla suddivisione in sottopopolazioni, differenziazioni genetiche tra popolazioni.

Nell'ambito del vasto piano di ricerca da noi iniziato (Martuzzi Veronesi e Pettener, 1990), riguardante aspetti di biodemografia storica in popolazioni di alcune vallate alpine (val di Sole, val di Non, valle del Fersina) dalla fine del XVI secolo agli inizi del XX, abbiamo ritenuto di un certo interesse illustrare la metodologia applicata, perchè possano essere correttamente considerati i risultati che via via si vanno ottenendo (Gueresi e coll., 1993).

Gli indicatori biodemografici che utilizziamo sono il tasso di endogamia, le distanze matrimoniali, il coefficiente di consanguineità, nonché alcuni coefficienti il cui calcolo si basa sulla

distribuzione dei cognomi, i quali possono consentire un'analisi sia della struttura matrimoniale interna alla popolazione che delle interrelazioni genetiche tra popolazioni. I dati di partenza necessari al calcolo di questi indicatori sono forniti dalle registrazioni, religiose o civili, degli atti di matrimonio, e consistono essenzialmente nel cognome degli sposi, nel loro luogo di nascita e residenza, e nel grado di consanguineità del matrimonio.

I problemi di carattere generale connessi alla utilizzazione dei registri parrocchiali sono ben noti: illeggibilità, lacune o duplicazioni degli atti matrimoniali, ecc. Ci soffermeremo quindi solo su alcuni aspetti che assumono particolare rilievo nelle nostre ricerche, in quanto sono in grado di influire sulla valutazione del livello di isolamento riproduttivo e dei rapporti di affinità genetica tra popolazioni.

1. Endogamia e movimenti matrimoniali

I tassi di endogamia e le distanze matrimoniali sono indicatori utilizzati per stimare il livello di isolamento genetico di una popolazione e la dispersione del patrimonio genetico nel territorio, data la relazione esistente tra movimento di coniugi e movimento di geni.

Vengono classificati come endogamici i matrimoni in cui entrambi i coniugi appartengono al collettivo in esame (parrocchia, valle o etnia) e come esogamici i matrimoni in cui uno dei contraenti non vi appartiene, considerando a parte i matrimoni in cui entrambi i contraenti risultano provenienti da altre comunità. Si calcolano quindi i tassi di endogamia ed esogamia, in un intervallo temporale che nelle nostre ricerche approssima l'intervallo tra due generazioni.

Di particolare interesse è il confronto tra tassi di esogamia per nascita o per residenza, al fine di ricavare indicazioni sui movimenti migratori prematrimoniali. Purtroppo accade sovente che gli atti di matrimonio, specialmente i più antichi, non riportino i luoghi di

nascita e di residenza, oppure presentino una sola, generica, indicazione di provenienza. In questi casi, per non rinunciare all'analisi, si conviene che, in mancanza di indicazioni, il luogo di nascita e di residenza sia la parrocchia in cui è registrato il matrimonio, mentre, in presenza di un'unica indicazione, questa valga sia come luogo di nascita che di residenza.

Le distanze matrimoniali tra i luoghi di nascita o di residenza dei coniugi possono essere calcolate in linea d'aria o lungo le principali direttrici di viabilità; nelle nostre ricerche, considerando che le vie impiegate per gli spostamenti variano nel tempo e che le popolazioni alpine spesso utilizzano sentieri di montagna, si è preferito misurare le distanze in linea d'aria.

Abbiamo quindi studiato la distribuzione delle distanze in classi la cui ampiezza tiene conto della dislocazione dei centri abitati nel territorio, calcolato la media e la mediana, quest'ultima perchè meno sensibile ai valori estremi della distribuzione, e quindi osservato le variazioni nel tempo.

Lo studio dei movimenti matrimoniali è completato dalla analisi della distribuzione dei luoghi di provenienza dei coniugi, che permette di individuare asimmetrie nei movimenti matrimoniali, con direzioni e versi di spostamento privilegiati, in dipendenza da fattori di ordine geografico, sociale ed economico.

Va detto che questi indicatori di mobilità hanno, anche dal punto di vista biologico, un valore relativo. Tra l'altro, sfuggono alla rilevazione quei matrimoni che, pur coinvolgendo un membro della parrocchia, vengono celebrati e registrati in un'altra. Questi matrimoni non comportano sempre l'uscita definitiva dell'individuo dal collettivo: esiste infatti la consuetudine, peraltro non generalizzabile ad ogni regione o periodo storico, di celebrare il matrimonio nella parrocchia della sposa, anche nel caso frequente in cui i coniugi vadano a stabilirsi nel paese del marito (migrazione virilocale). Accade così che non vengano registrate le spose che entrano nella comunità, mentre i matrimoni registrati come esogamici, che apparentemente indicano l'ingresso di un maschio

dall'esterno, in realtà sottintendono l'uscita di una donna. Nelle nostre ricerche, consultando i registri matrimoniali di tutte le comunità di una valle, possiamo se non altro ricostruire i movimenti matrimoniali a breve raggio; in ogni caso andrebbero tenute in considerazione altre fonti storiche e demografiche (ad esempio i registri dei nati e dei morti).

2. Consanguineità

Molto frequente nelle popolazioni alpine per evidenti motivi di isolamento e quindi di endogamia, e in particolare nelle minoranze etnico-linguistiche per motivazioni anche culturali, è il matrimonio tra consanguinei. Come è noto, l'effetto dell'unione tra consanguinei, cioè dell'inincrocio (inbreeding), è l'aumento delle frequenze dei genotipi omozigoti rispetto alla proporzione attesa secondo l'equilibrio di Hardy-Weinberg. Ciò comporta l'aumento della probabilità che si manifestino geni deleteri recessivi; di qui il grande interesse dal punto di vista biologico.

Ad ogni grado di consanguineità associato ad un matrimonio, corrisponde un coefficiente di consanguineità F (cfr. Bodmer e Cavalli Sforza, 1977) definito come la probabilità che due geni, estratti a caso dai due coniugi al medesimo locus, siano identici per discesa, cioè derivati dalla replicazione di un singolo allele appartenuto ad un antenato comune ai due coniugi, ed equivalente alla probabilità che dai due consanguinei nasca un omozigote per due alleli identici per discesa.

La registrazione nei libri dei matrimoni delle dispense per l'impedimento di consanguineità costituisce una delle fonti per la valutazione dell'inincrocio in una popolazione. Sulla base di queste informazioni è possibile stimare il livello medio di consanguineità in una popolazione (α), come

$$\alpha = \sum_i F_i n_i / N$$

dove F_i rappresenta il coefficiente di consanguineità dei coniugi di grado i -esimo, n_i è il numero dei matrimoni consanguinei di grado i -esimo, e N è il numero totale di matrimoni.

Il coefficiente α rappresenta una stima per difetto del reale livello di consanguineità di una popolazione, sia che venga calcolato tramite gli alberi genealogici, sia in base alle dispense matrimoniali, per la difficoltà a risalire indietro oltre la terza-quarta generazione. Con il metodo delle dispense, che noi adottiamo, si deve anche tener conto della possibilità di lacune o imprecisioni nelle registrazioni, o perchè le dispense non vengono segnalate, o perchè non è specificato il grado di consanguineità, oppure perchè non si distingue tra dispense per consanguineità o per affinità. Inoltre le formule utilizzate variano da parroco a parroco, in relazione alla scuola di formazione, e nel tempo: ad esempio un matrimonio tra secondi cugini e mezzo, di grado 3 in 4 secondo il computo canonico che considera il numero di generazioni che separa ciascun consanguineo dall'antenato comune (cfr. Moroni, 1960), dalla seconda metà dell'800 viene comunemente indicato come "quarto tangente il terzo", precedentemente come "terzo e quarto". Per questi motivi abbiamo ritenuto opportuno, nella fase di rilevazione, trascrivere integralmente la porzione di testo dell'atto matrimoniale che riguarda la concessione di dispense per consanguineità, per facilitare il computo e la revisione dei casi dubbi.

3. Il metodo dell'isonimia per il calcolo del coefficiente di inincrocio in una popolazione

E' possibile valutare il livello di consanguineità di una popolazione anche attraverso lo studio dei cognomi, in base al presupposto che essi siano assimilabili, in virtù del loro modello di eredità patrilineare, a geni legati al cromosoma Y, e che pertanto cognomi uguali possano essere considerati come alleli identici per discesa.

Crow e Mange (1965), in seguito alla osservazione che la probabilità di isonimia tra coniugi consanguinei è, salvo rarissimi casi, quattro volte il coefficiente di consanguineità di quel tipo di unione, hanno proposto un metodo di stima del coefficiente di consanguineità F che si basa sulla frequenza dei matrimoni isonimici in una popolazione, e che permette di scomporre il valore di F nelle due componenti casuale (F_r) e non casuale (F_n).

La componente casuale è data dalla probabilità che due persone di sesso opposto prese a caso nella popolazione abbiano lo stesso cognome. Varia in funzione delle dimensioni della popolazione e del suo livello di isolamento riproduttivo e comprende anche la consanguineità remota, accumulatasi nel tempo anche per effetti di deriva genica.

La componente non casuale misura invece lo scostamento dal modello di unione casuale e assume valori positivi o negativi secondo che i matrimoni consanguinei siano preferiti o evitati.

Crow e Mange hanno proposto due metodi per il calcolo di F e delle componenti casuale e non casuale: il primo (metodo A) adatto a studiare popolazioni formatesi da poche generazioni, il secondo (metodo B) applicabile a popolazioni formatesi in epoca più remota (Crow 1980) e pertanto più ampiamente impiegato in letteratura.

Secondo il metodo B, che anche noi adottiamo, data la relazione tra F e le componenti casuale e non casuale:

$$F = F_r + F_n - F_r F_n,$$

i valori di F_r , F_n , F riferiti alla generazione $t+1$ si ottengono nel seguente modo:

$$F_{r\ t+1} = P_{r\ t} / 4;$$

$$F_{n\ t+1} = P_{t} - P_{r\ t} / 4 (1 - P_{r\ t});$$

$$F_{t+1} = P_{t} / 4 + 3 P_{r\ t} (P_{t} - P_{r\ t}) / 16 (1 - P_{r\ t})$$

con P_t e $P_{r\ t}$ rispettivamente isonimia osservata e attesa alla generazione t . L'isonimia attesa si ricava come

$$Pr = (\sum_i S_{fi} S_{mi}) / M^2$$

dove S_{fi} e S_{mi} sono rispettivamente il numero di spose e di sposi aventi il cognome i -esimo, e M è il numero di matrimoni considerati.

Vi sono importanti limitazioni all'applicabilità del metodo dell'isonimia per la stima del coefficiente di consanguineità. Tra le condizioni richieste ricordiamo che la condivisione dello stesso cognome deve indicare effettivamente l'esistenza di un antenato comune dal quale esso è derivato (origine monofiletica dei cognomi), inoltre gli alberi genealogici devono essere simmetrici rispetto al sesso degli antenati intermedi e i due sessi devono essere ugualmente rappresentati tra i migranti. In realtà queste assunzioni sono raramente verificate: l'origine polifiletica dei cognomi più comuni e la pratica della migrazione virilocale portano ad una sovrastima dell'effettivo valore della consanguineità.

Critiche al metodo dell'isonimia sono state mosse da vari autori: ad esempio Libet (1983) ha messo in discussione il significato attribuito alle componenti casuale e non casuale, mentre Roger (1991) ha sostenuto che la semplice ipotesi di monofiletismo non è sufficiente a garantire l'applicabilità del metodo dell'isonimia, ma è richiesta un'ulteriore, più restrittiva assunzione. Secondo l'autore occorre ipotizzare che, in una passata generazione, nel gruppo fondatore, a N individui maschi corrispondessero N cognomi: per alcune zone alpine, nell'ipotesi che la colonizzazione sia avvenuta da parte di un ristretto gruppo fondatore, particolarmente nel caso di minoranze etniche, le condizioni iniziali potrebbero approssimare quelle richieste.

La tecnica dell'isonimia presenta un altro inconveniente: la quantità di dati usati nei calcoli è molto bassa, data la scarsa frequenza delle unioni isonimiche, anche in popolazioni endogame.

Nelle parrocchie di piccole dimensioni possono verificarsi notevolissime variazioni temporali dei valori di F in conseguenza di oscillazioni anche molto piccole del numero di matrimoni isonimici.

Nonostante questi limiti, gli studi sulle variazioni relative dell'isonimia forniscono una valutazione attendibile dei cambiamenti del grado di isolamento riproduttivo di una popolazione: l'andamento dei valori di F risulta infatti correlato con l'andamento dell'endogamia (Pettener, 1988; 1990).

4. La valutazione del grado di suddivisione di una popolazione: il metodo delle coppie ripetute

Se i membri di una popolazione, per motivi culturali, sociali ed economici, preferiscono scegliere il coniuge all'interno del gruppo di appartenenza piuttosto che in altri gruppi, nella popolazione potrà instaurarsi un certo grado di suddivisione in sottopopolazioni che avrà come effetto biologico quello di produrre un eccesso di omozigotità (effetto Wahlund).

Lasker e Kaplan nel 1985 hanno proposto un metodo per stimare il livello di suddivisione in una popolazione che, utilizzando la matrice completa delle unioni matrimoniali, misura la tendenza di alcune coppie di cognomi a ripetersi.

La misura proposta è stata chiamata RP (repeated pairs):

$$RP = \sum [S_{ij} (S_{ij} - 1)] / [N(N-1)]$$

dove S_{ij} rappresenta il numero di coppie in cui il marito ha il cognome i e la moglie il cognome j , N è il totale delle coppie registrate.

RP è un indice normalizzato che può variare tra 0 (quando nessuna combinazione di cognomi compare più di una volta) e 1 (quando tutte le coppie presentano la medesima combinazione di cognomi). Dalla formula si evince che al valore finale di RP

concorrono solo le combinazioni di coppie che si ripetono almeno due volte.

Per avere una valutazione del livello di suddivisione della popolazione occorre confrontare il valore di RP osservato con quello atteso in regime di panmissia (RPr). Il calcolo di RPr si esegue generalmente usando la formula proposta da Chakraborty (1985).

Nel 1986 è stata proposta una variante di RP, detta RP2. La formula di RP2 è la stessa di RP, ma cambia il significato da attribuire a S_{ij} , che viene ad indicare il numero di coppie in cui uno dei due coniugi, non importa se marito o moglie, ha il cognome i e l'altro coniuge ha il cognome j .

RP2r si può ottenere come media degli RP2 calcolati su una serie di 12 permutazioni casuali dei cognomi.

Il confronto tra RP e RP2 può evidenziare la presenza di interscambi matrimoniali preferenziali tra gruppi familiari, tenendo conto anche del sesso dei coniugi scambiati.

5. Uso dei cognomi per la valutazione dei rapporti di affinità genetica tra popolazioni

Un'importante estensione del metodo dell'isonimia è stata introdotta da Lasker (1977) per valutare la affinità genetica tra popolazioni, in base alla ipotesi che due distinte popolazioni sono tanto più simili tra loro dal punto di vista genetico, quanto maggiore è il numero di cognomi in comune e quanto più simili sono le distribuzioni dei cognomi.

Il coefficiente di relazione isonimica (R_i) misura la probabilità che due persone scelte a caso da due popolazioni diverse abbiano lo stesso cognome, e quindi lo stesso gene, per comune discendenza:

$$R_i = \sum_i (S_{i1} S_{i2}) / 2 \sum_i S_{i1} \sum_i S_{i2}$$

dove S_{i1} e S_{i2} sono rispettivamente le numerosità del cognome i .

esimo nella prima e nella seconda popolazione.

Un altro coefficiente frequentemente utilizzato è stato proposto da Hedrick (1971) originariamente con il significato di misura della probabilità di identità genotipica tra due popolazioni:

$$I_{12} = \sum_i S_{i1} S_{i2} / [1/2 (\sum_i S^2_{i1} + \sum_i S^2_{i2})]$$

dove S_{i1} e S_{i2} sono le numerosità del cognome i -esimo nelle popolazioni 1 e 2. Qui la probabilità di estrarre cognomi identici dalle due popolazioni (al numeratore) è pesata rispetto alla probabilità media di estrarre cognomi identici dalla stessa popolazione nel corso di due estrazioni indipendenti: la misura così ottenuta non presenta gli svantaggi di altre misure di similarità che sono costruite utilizzando le frequenze dei cognomi.

Di particolare interesse è l'impiego di tali metodi nel confronto tra più popolazioni e/o periodi di tempo, attraverso l'applicazione di tecniche di analisi quali lo scaling multidimensionale o l'analisi delle componenti principali alle matrici dei coefficienti di relazione isonimica (Lucchetti e Soliani, 1989).

Rispetto ai marcatori genetici classici, quali ad esempio i sistemi gruppoematici, i cognomi forniscono risultati sostanzialmente concordanti, tuttavia le loro peculiari caratteristiche (origine più recente, grado di polimorfismo più elevato, frequenze più variabili), ne consentono l'uso nei confronti tra gruppi all'interno di aree geografiche circoscritte e rendono la loro distribuzione particolarmente sensibile ai movimenti di popolazione. I cognomi offrono anche il vantaggio di consentire lo studio delle variazioni temporali nei rapporti di affinità genetica tra popolazioni negli ultimi quattro secoli, aspetto questo che non è possibile analizzare utilizzando i sistemi eritrocitari.

Tra i diversi usi dei cognomi in biologia delle popolazioni umane, questo tipo di applicazione, grazie all'ampia disponibilità di dati e all'elevato grado di informazione associato al sistema cognomi, è attualmente molto diffuso e suscettibile di ulteriori interessanti

sviluppi.

L'uso dei cognomi nello studio della struttura genetica delle popolazioni si basa sul parallelismo tra eredità dei cognomi ed eredità biologica, ma la origine e la evoluzione dei cognomi dipendono da fattori non genetici (ad es. linguistici o culturali) che in molti casi non sono assimilabili alla eredità biologica. Analizzando i libri dei matrimoni di una parrocchia relativamente ad un lungo arco di tempo, è facile imbattersi in forme cognominali diverse che sono in realtà trasformazioni di uno stesso cognome di partenza, la cui presenza è dovuta a disattenzione dei parroci nella trascrizione, o a "mutazioni" dei cognomi nel corso del tempo.

Nella val di Rabbi, ad esempio, per molti cognomi si verifica un raddoppio di consonanti (Ciati-Ciatti, Pangrazi-Pangrazzi, Ruati-Ruatti) o l'aggiunta di una vocale finale (Molignon-Molignoni, Misseron-Misseroni) o una fusione (Dalla Serra-Dallaserra). Considerando queste varianti come cognomi differenti, si rischia di sottostimare il grado di affinità genetica tra popolazioni, o di enfatizzare le differenze tra periodi diversi nell'ambito della stessa popolazione. Anche il livello di consanguineità potrebbe risultare sottostimato, se venissero trascurate unioni da considerarsi in realtà come isonimiche.

Da un punto di vista pratico, il problema che ci si presenta è quello di ricondurre ad un cognome unico le varie forme cognominali che si ritiene possano essere "mutate" dalla forma originale.

Si deve pertanto procedere alla revisione e alla correzione dei cognomi, costruendo un secondo archivio. Un aspetto metodologico particolarmente interessante è dato dal confronto tra risultati ottenuti utilizzando l'archivio originale o l'archivio soggetto ad una revisione ragionata.

Particolare attenzione si deve porre nell'individuare le "famiglie" di cognomi, evitando di accorpate in una unica forma cognomi che, pur molto simili per struttura, sono diversi per significato. Il criterio di uniformare le varie forme cognominali in base a quella più ricorrente può essere valido se l'intervallo temporale considerato è

breve (ad esempio 25 anni); diversamente occorrerà tener presente la possibilità di evoluzione dei cognomi, e quindi scegliere se ricondurre i cognomi alla forma più antica, dalla quale le successive sono derivate per "mutazione", oppure alla più recente, criterio questo che forse operativamente è più utile, perchè può consentire confronti con la distribuzione delle forme cognominali attuali.

I dati biodemografici sin qui raccolti hanno portato alla costituzione di un archivio di grande interesse per lo studio delle variazioni negli ultimi due secoli delle principali caratteristiche biodemografiche di alcune popolazioni delle alpi trentine.

I primi risultati hanno messo in rilievo alti valori di endogamia e di consanguineità, indicativi di una marcata tendenza all'isolamento riproduttivo, nella val di Sole e nella Val di Non, e in particolare nella minoranza etnico-linguistica dei Mocheni della valle del Fersina, tuttavia soltanto con l'apporto delle conoscenze riguardo alla dinamica demografica si potrà avere un quadro più completo delle diverse situazioni che possono avere influito sulla struttura biologica di tali popolazioni.

Bibliografia

- W. F. Bodmer - L. L. Cavalli Sforza (1977), *Genetica, evoluzione uomo*, Vol. II, Edizioni Scientifiche e Tecniche Mondadori, Milano.
- R. Chakraborty (1985), A note on the calculation of random RP and its sampling variance. *Human Biology*, 57, pp. 713-717.
- J. F. Crow (1980), The Estimation of Inbreeding from Isonymy, *Human Biology*, 52, pp.1-14.
- J. F. Crow (1989), Update to "The Estimation of Inbreeding from Isonymy", *Human Biology*, 61, pp.949-954.
- J. F. Crow - A. P. Mange (1965), Measurement of inbreeding from the frequency of marriages between persons of the same surname, *Eugenics Quarterly*, 12, pp.199-203.
- P. Guerresi - F. Martuzzi Veronesi - D. Pettener (1993), *Aspetti biodemografici delle valli trentine: l'alta Valle di Sole dal 1750 al 1923*, Bollettino di Demografia Storica (in corso di pubblicazione).
- P. W. Hedrick (1971), A new approach to measuring genetic similarity, *Evolution*, 25, pp.276-280.
- G. W. Lasker (1977), A coefficient of relationship by isonymy: a method for estimating the genetic relationship between populations, *Human Biology*, 49, pp.489-493.
- G. W. Lasker - B. A. Kaplan (1985), Surnames and genetic structure: repetition of the same pair of names of married couples, a measure of subdivision of the population, *Human Biology*, 57, pp.431-440.

- G. W. Lasker - C. G. N. Mascie-Taylor - D. A. Coleman (1986), Repeating pairs of surnames in marriages in Reading (England), and their significance for population structure, *Human Biology*, 58, pp.421-425.
- M. L. Libet (1983), Expected versus actual random inbreeding: A reinterpretation of the random-non random inbreeding formula, *Human Biology*, 55, pp.323-329.
- E. Lucchetti - L. Soliani (1989), Similarità tra popolazioni esaminate mediante i cognomi, *Rivista di Antropologia*, Vol. LXVII, pp.181-198.
- F. Martuzzi Veronesi - D. Pettener (1990), Aspetti biodemografici in popolazioni chiuse: prospettive di ricerca, *Bollettino di Demografia Storica*, 11, pp.43-48.
- A. Moroni (1960), Analisi metodologica della consanguineità. Parte Prima. Legislazione e materiale religioso e civile italiano per il rilievo della consanguineità, *Folia Hered. et Path.*, 9, 3, pp. 199-247.
- D. Pettener (1988), Aspetti biodemografici nello studio delle popolazioni umane. In: F. Facchini, *Evoluzione, uomo e ambiente*, pp. 458-465.
- D. Pettener (1990) Temporal trends in marital structure and isonymy in S.Paolo Albanese, Italy, *Human Biology*, 62, pp.837-851.
- A. R. Rogers (1991) Doubts about isonymy, *Human Biology*, 63, pp.663-668.

SALVATORE BERTINO - DONATELLA GRASSI - EUGENIO SONNINO

UN MODELLO DI MACROSIMULAZIONE DEL CICLO DI VITA DI UNA COORTE MATRIMONIALE

1. Obiettivi del modello

Gli studi della fecondità basati sulla ricostruzione nominativa delle vicende riproduttive di coorti matrimoniali hanno costituito durante l'ultimo quarantennio un terreno coltivato con particolare cura dai demografi storici. In Italia, tuttavia, la rilevanza dell'investimento di risorse umane e materiali richiesto per l'esame analitico delle storie familiari da ricostruire e il bilancio non sempre soddisfacente dei risultati raggiunti - valutabile sulla scorta delle ricche esperienze maturate in vari paesi -, unitamente a perplessità di ordine metodologico, hanno prodotto un ricorso assai ridotto da parte degli studiosi a questa procedura di analisi. Probabilmente questa limitata diffusione deve essere fatta anche risalire alla difficoltà di riprodurre in Italia il lancio di un vasto progetto nazionale del tipo di quelli condotti, in vario modo, in altri paesi (Francia, Inghilterra, Canada). Difficoltà connessa non solo con la mole delle risorse da investire, ma anche con la specificità del territorio italiano e delle sue diverse realtà storiche. D'altro canto, la stessa straordinaria abbondanza e molteplicità delle fonti storico-demografiche italiane ha richiamato l'attenzione ed il lavoro degli studiosi su un numero molto ampio di versanti di ricerca, lasciando in secondo piano l'esame analitico dei processi riproduttivi delle coorti in epoca pre-unitaria.

Tuttavia, tale studio rimane di importanza primaria e per molti aspetti insostituibile, sia di per sé sia anche per gli apporti che esso può fornire a ricerche di portata più ampia sulle caratteristiche evolutive delle popolazioni, le quali non possono prescindere dalla conoscenza approfondita delle vicende riproduttive.

Il modello di macrosimulazione che viene qui presentato, sia pur - per limiti di spazio - con una illustrazione succinta delle sue caratteristiche e dei risultati della sua sperimentazione, vuole costituire un apporto al tentativo di definire un ventaglio di possibili determinazioni del ciclo riproduttivo matrimoniale, risultante dal diverso giuoco di fattori e condizioni di ordine demografico quali l'età al matrimonio, la lunghezza dell'intervallo tra i parti, il profilo della fecondità, i livelli di mortalità, la frequenza del "remariage". L'utilità di un modello di questo tipo può essere valutata su diversi versanti. In primo luogo c'è l'utilità per così dire "didattica", in quanto la sperimentazione del modello mette chiaramente in luce la diversa importanza ed il vario giuoco delle variabili assunte in "input" e gli effetti "infiniti" dell'interazione delle diverse combinazioni di ipotesi. In secondo luogo, i risultati del modello possono essere confrontati con casi realmente studiati di coorti ricostruite; ciò può servire tanto a testare l'efficacia della procedura di simulazione quanto a consolidare ipotesi interpretative dei risultati della ricostruzione delle famiglie effettuata. Infine, nel caso di aggregati demografici per i quali non si può o non si vuole procedere allo studio della fecondità mediante ricostruzione nominativa e per i quali si dispone di informazioni parziali su alcune variabili da assumere in input, il modello può essere sperimentato a fini di definizione di ipotesi e determinazione di stime da assumere nella ricerca.

Simulare vuol dire necessariamente semplificare. Le manifestazioni del ciclo familiare sono indubbiamente il prodotto dell'operare di fattori certo enormemente più numerosi di quelli che vengono assunti nel modello. Tuttavia le prove finora effettuate ed i risultati ottenuti appaiono incoraggianti. Valutare gli effetti delle

variabili assunte i termini di sopravvivenza dei matrimoni, di fecondità dei medesimi, di numero di orfani e vedovi prodotti dall'operare dei diversi regimi, di effetti del "remariage", consente di proporre - con riferimento alle ipotesi e ai vincoli assunti nelle diverse situazioni - scenari non solo di interesse demografico ma di grande rilievo storico-sociale nella vita delle società oggetto di studio.

Il presente testo è, come si è detto, necessariamente succinto. Qualche approfondimento di carattere tecnico e qualche riflessione sulle diverse problematiche evocate sono contenuti in scritti precedenti degli autori (1991-1992). È in corso di preparazione una più dettagliata illustrazione del lavoro.

2. Le ipotesi di lavoro

Si riportano molto brevemente una serie di ipotesi formulate per la costruzione del modello.

Alla base del processo viene collocata l'evoluzione dei matrimoni, si vuole cioè sapere quanti dei matrimoni iniziali sopravvivono o si sciolgono nei successivi anni di studio. Chiaramente l'evoluzione risponde alle leggi di sopravvivenza cui sono soggetti gli individui che compongono la coorte. A tale riguardo la prima ipotesi che viene formulata consiste nel considerare, come unica possibilità di separazione tra i coniugi, la morte ed, in connessione a ciò, la morte dell'uno si ipotizza indipendente rispetto a quella dell'altro. Inoltre il compleanno di ogni individuo viene fatto coincidere con l'anniversario di matrimonio ed i figli sono soggetti alla stessa legge di mortalità dei genitori. L'origine dei tempi coincide con la data del matrimonio.

Il momento centrale nella costruzione del modello consiste nel calcolo dell'ammontare dei matrimoni fecondi dopo ogni tempo t dal matrimonio. In questo contesto si è convenuto che non sono ammessi concepimenti prematrimoniali.

3. Il modello

3.1 I matrimoni sopravvivenenti

Lo scioglimento di un matrimonio provoca come conseguenza, oltre all'impossibilità di procreare e ad un incremento del numero di orfani nella popolazione, un incremento di una unità nell'ammontare dei vedovi o delle vedove o delle coppie estinte. Tenendo conto di ciò, il procedimento di simulazione dell'evoluzione dei matrimoni viene elaborato come un processo aleatorio in cui la variabile aleatoria studiata può essere definita "status matrimoniale" e gli stati che si possono raggiungere nel processo sono: i) coppia, ii) vedovo, iii) vedova, iv) coppia estinta. Le probabilità di transizione relative a questo processo sono il risultato dell'interazione delle probabilità di sopravvivenza dei coniugi e gli unici passaggi di stato che si possono verificare sono "coppia-coppia", "coppia - vedovo/a", "vedovo/a - coppia estinta", "coppia estinta - coppia estinta" che vengono evidenziati nelle seguenti relazioni:

Matrimoni sopravvivenenti al tempo t : $M(t) = M(t-1) p(y+t-1, 1) p(x+t-1, 1)$

Vedovi al tempo t : $V(t) = V(t-1) p(y+t-1, 1) + M(t-1) p(y+t-1, 1) (1-p(x+t-1, 1))$

Vedove al tempo t : $V(t) = V(t-1) p(x+t-1, 1) + M(t-1) (1-p(y+t-1, 1)) p(x+t-1, 1)$

Coppie estinte al tempo t :

$$E(t) = E(t-1) + V(t-1) (1-p(y+t-1, 1)) + V(t-1) (1-p(x+t-1, 1)) + M(t-1) (1-p(y+t-1, 1)) (1-p(x+t-1, 1))$$

in cui x è l'età della sposa al matrimonio ed y quella dello sposo. Le seconde nozze vengono prese in considerazione solo in fase di rielaborazione dell'output primario.

3.2 I matrimoni fecondi

In modo più complesso, si ottiene il numero di matrimoni fecondi ad ogni istante t . La complicazione risiede nel fatto che non sempre

una coppia può trovarsi nello stato fecondo. In modo specifico un matrimonio si può considerare fecondo nell'istante t se: i) il marito sopravvive almeno fino all'istante $t-3/4$; ii) la moglie sopravvive almeno fino a t ; iii) la coppia non abbia generato una nascita nell'intervallo $[t-t, t]$ dove t rappresenta l'intervallo minimo medio tra due nascite che si è supposto costante per ogni ordine di nascita e per ogni età al parto ed è un parametro da definire a priori. Altro elemento di importanza è rappresentato dalla funzione tasso istantaneo di fecondità $j(x)$, legata concettualmente ai quozienti specifici di fecondità, definita in modo che si possa calcolare la probabilità che una donna di età x , sposa all'età x , dia luogo ad una nascita nell'intervallo $[x, x+Dx)$. Con questo dato sarà possibile calcolare l'ammontare delle coppie che hanno dato luogo ad una nascita in un determinato intervallo di tempo. Viene dunque introdotta, come input al modello, una curva di "fecondità teorica" $f(x)$, che opportunamente manipolata permette di definire la funzione $j(x)$ e di ricavare il numero di matrimoni fecondi al tempo $t+Dt$ mediante la seguente relazione:

$$M(t+Dt) = M(t) \cdot (1-m_m(y+t-3/4)Dt) \cdot (1-m_f(x+t)Dt) \cdot (1-j(x+t)Dt) + M(t-t) \cdot j(x+t-t)Dt \cdot p_m(y+t-3/4+Dt-t, t) \cdot p_f(x+tDt-t, t)$$

dalla quale si ottiene la seguente equazione differenziale:

$$M'(t) = -M(t) \cdot [m_m(y+t-3/4) + m_f(x+t) + j(x+t)] + M(t-t) \cdot j(x+t-t) \cdot p_m(y+t-3/4-t, t) \cdot p_f(x+t-t, t)$$

Nelle variabili h (anni di matrimonio), i (trimestri) e u (porzione di trimestre), in modo che sia $t = h+i/4+u$ questa equazione differenziale ha per soluzione la funzione:

$$M(h+i/4+u) = \exp\{u \cdot S(h+i/4)\} \cdot (h+i/4) + j(x+h-t+i/4) \cdot$$

$$\int_0^u M(h-t+i/4+u) \cdot b(x, y, h, 1, i, t, u) \cdot \exp\{u_1 S(h+i/4)\} du_1$$

dove: $0 \leq u \leq 25$; $i=0, 1, 2, 3$;

$$S(h+i/4) = m_m(y+h+(i-3)/4) + m_f(x+h+i/4) + j(x+h+i/4);$$

$$b(x, y, h, s, i, t, v) = p_m(y+h-st+(i-3)/4+v, t) \cdot p_f(x+h-s-t+i/4+v, t)$$

in cui m_m e m_f rappresentano i tassi istantanei di mortalità rispettivamente per maschi e femmine ed $i/4$ il trimestre dell'anno per il quale si vuole determinare l'ammontare dei matrimoni fecondi.

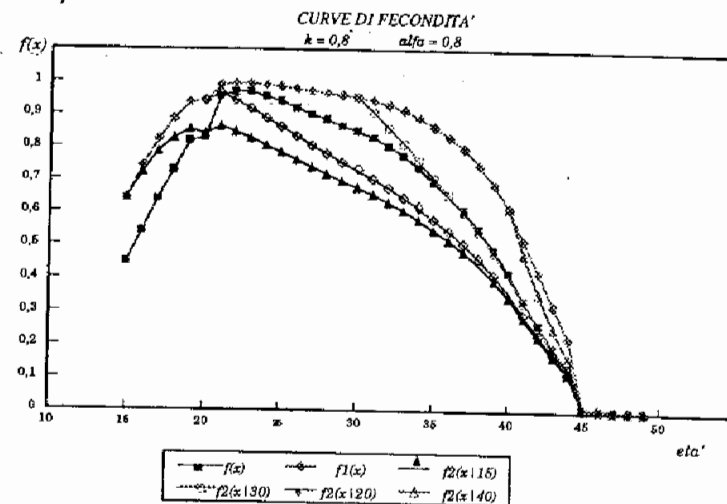
Dunque la sopravvivenza dei coniugi e il fatto di non aver generato tra $t-t$ e t fanno di un "matrimonio sopravvivente" un "matrimonio fecondo". Si noti infine che la funzione tasso di fecondità $j(x)$ viene associata all'intervallo t ed in tal modo la probabilità "potenziale" di dar luogo ad una nascita viene riportata a livelli pari a quelli realmente osservati.

Alcune osservazioni saranno utili per illustrare come il modello sviluppa la dinamica riproduttiva della coppia trasformando opportunamente la curva di fecondità teorica. Per prima cosa il modello tiene conto di quello che si può indicare come "effetto luna di miele" che si concretizza in un aumento dei valori della curva di fecondità teorica per i primi anni di matrimonio e in relazione all'età della sposa al matrimonio. Dopo di ciò si è introdotta l'ipotesi che con il procedere degli anni di matrimonio, oltre a diminuire la frequenza dei rapporti sessuali tra i coniugi, possano intervenire patologie varie che riducono più del naturale la fecondità delle donne. Per tenere conto di questo ulteriore elemento che modifica la dinamica della curva di fecondità naturale è stato assunto un andamento di tipo esponenziale negativo.

Osservando il grafico (Grafico 1) si può avere una illustrazione visiva della curva di fecondità assunta dal modello per una donna che si sposa all'età x_0 (nel grafico si considerano le età 15, 20, 30 e 40). Sia $f(x_0)$ il valore della curva di fecondità teorica in input. Sarà: $f_1(x_0) = f(x_0) + a (f(x_0) (1 - f(x_0)))$ con $a < 1$ parametro assegnato, il valore della funzione di fecondità considerato al momento del matrimonio. Per $x > x_0$ la funzione di fecondità assunta dal modello sarà:

$$f_2(x_0) = f_1(x_0) \exp \left\{ -k (x - x_0) / (50 - x_0) \right\}.$$

Grafico 1



Dunque per la determinazione dell'ammontare dei matrimoni fecondi si devono introdurre come input: i) le probabilità di sopravvivenza; ii) le probabilità di fecondità teorica; iii) un t di ampiezza stabilita a priori ed iv) il parametro a ed il parametro k dell'esponenziale. Le diverse combinazioni nei livelli e negli andamenti di questi elementi permetteranno di volta in volta di descrivere situazioni assimilabili a quelle realmente osservate.

3.3 Le altre relazioni

Una volta determinati i matrimoni, sopravviventi prima e fecondi poi, si può passare alla specificazione di tutte le altre relazioni del modello ottenendo:

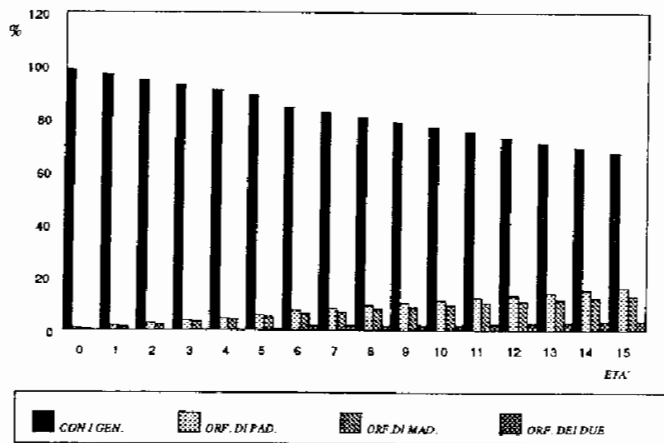
Nati per trimestre $N(t+i/4, t+(i+1)/4) = \int_0^{1/4} M(t+i/4+z)\varphi(x+t)dz$

$$\begin{aligned} \text{Nati per sesso} \quad N_m(t+i/4) &= 0,52N(t+i/4) \\ N_f(t+i/4) &= 0,48N(t+i/4) \end{aligned}$$

Sopravvivenuti alla fine dell'anno con o senza i genitori per i quali si suppone che le nascite siano uniformemente distribuite nel periodo di nascita. Per motivi di spazio, tralasciamo di riportare le equazioni costruite, si riporta invece un grafico (Grafico 2) in cui viene evidenziata la distribuzione dei nati in una coorte matrimoniale in cui valga l'ipotesi di stazionarietà.

Grafico 2

SOPRAVVIVENTI DI 0-15 ANNI DISTRIBUITI SECONDO LA SOPRAVVIVENZA DEI GENITORI
IPOTESI DI STAZIONARIETA'



3.4 L'elaborazione dei risultati

Per ogni possibile combinazione di età degli sposi alle nozze viene effettuata una simulazione, l'insieme dei risultati ottenuti per il complesso dei matrimoni viene poi rielaborato come corpo unico con riferimento cioè ad un'unica coorte matrimoniale. Dunque le informazioni sull'evoluzione dei matrimoni, sulle nascite, sugli orfani e sulla vedovanza, ottenuti applicando il modello, vengono rilette unitariamente.

Come rielaborazione dell'output fornito dalle simulazioni, è possibile poi desumere informazioni sia sull'effettiva fecondità della coorte, cioè il risultato complessivo dell'interazione delle probabilità di fecondità teorica con t , a e k , sia sull'incidenza delle seconde nozze per cui è necessario conoscere i) una serie di probabilità per età di contrarre un secondo matrimonio, ii) una serie di intervalli tra prime e seconde nozze classificati per età o per classi di età con cui si ottiene l'ammontare delle donne che rimaste vedove ad una certa età contrarranno, allo scadere dell'intervallo prefissato, un secondo matrimonio valutando così l'effettiva presenza delle vedove e degli orfani nella popolazione.

4. Le elaborazioni: applicazione e confronti

Di seguito si riportano i risultati di una serie di elaborazioni con specifico riferimento agli input e ai risultati ottenuti.

4.1 Gli input

Per punti i dati introdotti sono:

- Una tavola di mortalità per ciascun sesso con riferimento a quella costruita per l'Italia del 1881/82.

- b) Due curve di fecondità teorica di base (le $f(x)$ per intendersi) che sono state ottenute mediando una serie di curve riferite a popolazioni, reali (storiche ed attuali) e fittizie, per le quali si può ritenere non agisca alcun controllo volontario dei concepimenti. Le curve risultanti sono state traslate verso l'alto, mantenendone invariato l'andamento, per ottenere la fecondità potenziale per cui il massimo è pari ad 1.
- c) Per a e k sono state considerate diverse ipotesi specificate di volta in volta.
- d) Anche per t sono stati testati diversi valori e dopo una serie di elaborazioni preliminari si è ritenuto di effettuare i confronti utilizzando tre diversi intervalli: 15, 18 e 21 mesi. Valori molto più alti o molto più bassi di quelli considerati, combinati con le curve di fecondità teorica prescelte fornivano risultati non soddisfacenti.
- e) Una distribuzione dei matrimoni secondo la doppia classificazione "età dello sposo-età della sposa" osservata per il complesso dell'Italia nel 1865 in cui l'età media al primo matrimonio delle donne è pari a 24,5 anni.
- f) La distribuzione per età degli intervalli tra primo e secondo matrimonio con riferimento a quella proposta da Schofield e Wrigley (1981) e calcolata su 14 parrocchie inglesi per il periodo 1600-1799.
- g) La distribuzione per età delle probabilità di contrarre un secondo matrimonio per le donne ottenuta ridistribuendo quella calcolate per alcune grandi classi di età da Schiaffino (1981) per Bologna.

4.2 I risultati e i confronti

Il programma di simulazione fornisce una serie di risultati che sono poi quelli elencati in precedenza, per ogni singola combinazione d'età degli sposi pesando in modo opportuno, secondo la distribuzione da noi inserita come input, l'ammontare dei matrimoni della coorte. L'insieme di risultati preliminari vengono poi accorpati ipotizzando di

lavorare su una coorte di 10^n matrimoni distribuiti secondo la combinazione di pesi considerati.

In modo schematico i risultati che è possibile ottenere attraverso questa rielaborazione dell'output sono: la distribuzione per età della donna e per età al matrimonio delle vedove, delle vedove sopravvivenenti, degli orfani, del rapporto orfani/vedove; la distribuzione delle vedove i) all'anno di vedovanza, ii) dopo l'intervallo tra primo e secondo matrimonio e iii) al secondo matrimonio; i quozienti specifici di fecondità per età della donna nel complesso e per le singole età al matrimonio; la distribuzione per anno di matrimonio di donne, vedove ed orfani sopravvivenenti e del rapporto orfani/vedove.

Poichè si sono considerati solo i primi matrimoni di donne con età inferiore o uguale ai 50 anni, il numero complessivo di coppie che compongono la coorte è pari a 99000. I pesi, cioè, attribuiti alle singole combinazioni di età sommano ad uno per il complesso dei matrimoni osservati nell'anno 1865, dunque non considerando le spose di età superiore ai 50 anni, si sottostimano i matrimoni dell'1% circa e l'età media al matrimonio della coorte da noi considerata è pari a 24,4 anni contro i 24,5 anni calcolati per la coorte 1865.

Delle 99000 donne in vita all'inizio del primo anno di osservazione 45563 restano vedove entro i 65 anni di età ovvero il 46% dei matrimoni si rompe per morte del marito quando ancora la donna può avere figli di età inferiore o uguale ai 15 anni (età considerata come limite nel rapporto di dipendenza dalla famiglia). Cabourdin (1981) stima che nel XVII/XVIII secolo a Crulay tale percentuale è pari circa a 63, 56 a Tourouvre, 56 ad Argenteuil. Il valore ottenuto con le simulazioni è piuttosto più basso di quelli osservati da Cabourdin, ma questo può dipendere dal fatto che il livello di mortalità della popolazione fittizia considerata è probabilmente più basso, per entrambi i sessi, di quelli che agivano nella popolazione francese del XVII secolo.

Sul complesso delle vedove ben il 90% (42107) arriva in vita alla fine dell'intervallo tra prime e seconde nozze introdotto come input,

ma di queste solo il 18% (7772) si risposa, valore questo molto vicino a quello stimato da Livi Bacci (1981) per l'Italia del 1883/86 e pari al 16%.

Tra le vedove che si risposano il 31% (7263) lo fa entro i 50 anni, Livi Bacci stima che questo valore si aggira attorno al 40% per il complesso dell'Italia variando dal 28 al 49% nei vari compartimenti.

L'anno di matrimonio in cui mediamente una donna resta vedova è il 23°. Cabourdin calcola, sempre per le popolazioni nominate in precedenza, un valore medio di 16-18 anni, con una forbice di 13-22 a seconda delle epoche, confermando dunque l'ipotesi fatta in precedenza sui livelli di mortalità.

Nella tabella 1 si riporta la distribuzione percentuale delle donne rimaste vedove ad età successive ($v(x)$). Confrontando il risultato ottenuto attraverso le simulazioni con quelli riportati da Cabourdin si osservano delle differenze anche consistenti ancora una volta riconducibili alle diversità concernenti le leggi di mortalità, ma che questa volta potrebbero dipendere anche da difformità nel metodo utilizzato per il calcolo dell'indicatore.

Tabella 1: percentuale di vedove ad x anni di età

età	$v(x)$		
	simul	Argenteuil 1770-1815	Conflans 1730-99
25	5,62	7,7	-
30	9,06	17,3	8
35	14,30	30,6	18
40	22,03	45,8	29
45	32,92	58,3	38
50	49,38	67,5	48

Prescindendo dal significato storico, demografico e sociale degli specifici risultati, si può desumere dal confronto con i pochi dati

esistenti riguardo a popolazioni reali del passato, che il modello sia in generale ben formulato.

A conferma della buona adattabilità del modello si riportano anche alcuni dei risultati ottenuti circa la fecondità espletata dalla coorte matrimoniale. Sono state considerate molteplici combinazioni degli elementi che interagiscono nella definizione dei livelli di fecondità. Si è potuto notare che i migliori risultati in termini di adattabilità dei risultati si hanno quando si assumano degli andamenti di fecondità con un massimo tra i 19 ed i 23 anni associati ad un intervallo intergenesico minimo medio variabile 15 ed i 21 mesi e a parametri α e k compresi tra 0,4 e 0,8.

Le curve di fecondità per età e per età al matrimonio vengono confrontate dunque con gli andamenti osservati in alcune popolazioni dell'Italia e della Francia dei secoli scorsi. Osservando i grafici di seguito riportati, si può notare una sostanziale uniformità negli andamenti della fecondità reale e di quella fittizia. Il grafico 3 ed il grafico 4 se sovrapposti infatti presentano numerosi punti di sovrapposizione e lo stesso vale per i grafici 5 e 6. Per motivi di spazio non si possono riportare i grafici relativi alle altre combinazioni i quali comunque sono ugualmente assimilabili a situazioni realistiche di fecondità.

In questa fase il modello non è ancora giunto alla sua formulazione definitiva in quanto permangono ancora alcune difficoltà, non tanto nell'adattamento dei risultati ai dati reali, che come si è visto, risulta generalmente soddisfacente, quanto piuttosto nella gestione dell'output che deve essere resa più agevole.

Grafico 3

QUOZIENTI SPECIFICI DI FECONDITA' PER ETA'
E PER ETA' AL MATRIMONIO
 $f(x)_{max}=23 - \tau=1,75 - \alpha=0,6 - k=0,8$

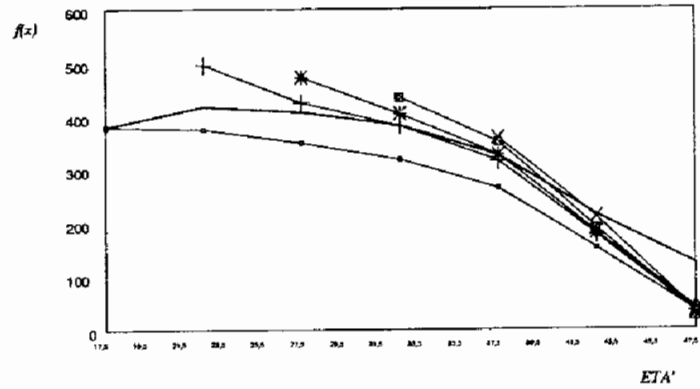


Grafico 4

QUOZIENTI SPECIFICI DI FECONDITA' PER ETA' E PER ETA' AL MATRIMONIO
(TREPPIO 1790-1839, FECONDITA' COMPLETA)

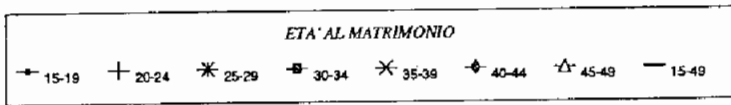
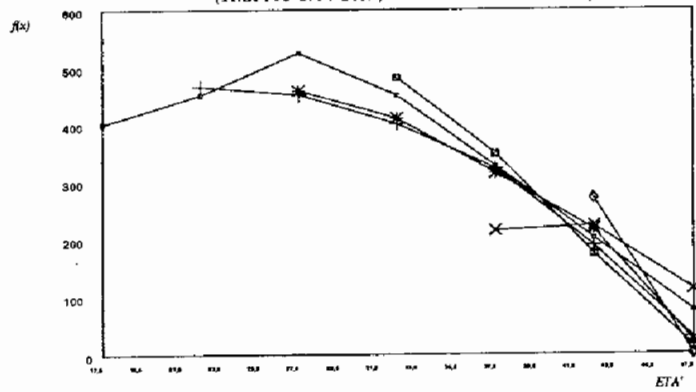


Grafico 5

QUOZIENTI SPECIFICI DI FECONDITA' PER ETA'
E PER ETA' AL MATRIMONIO
 $f(x)_{max}=19 - \tau=1,50 - \alpha=0,8 - k=0,6$

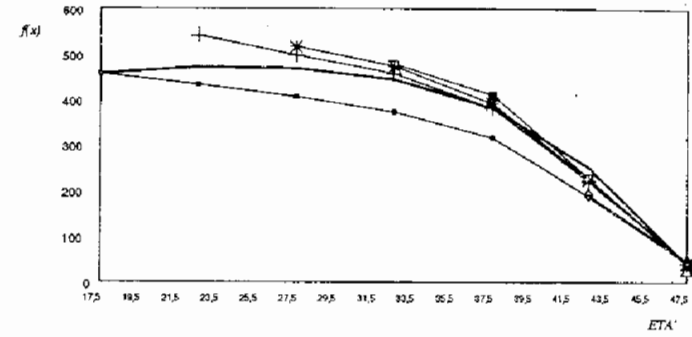
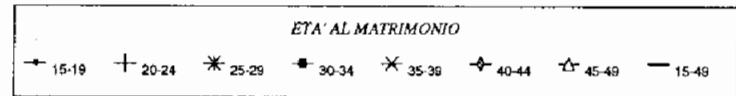
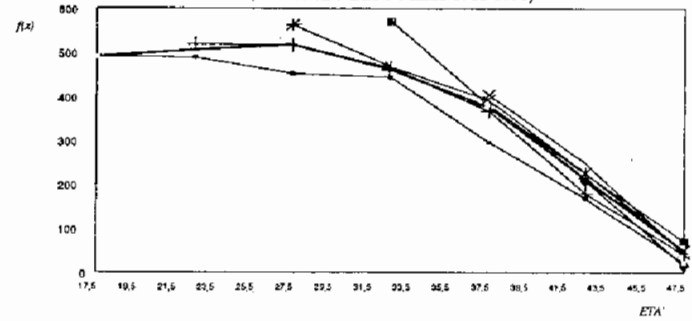


Grafico 6

QUOZIENTI SPECIFICI DI FECONDITA' PER ETA' E PER ETA' AL MATRIMONIO
(VILLEDIEU-LES-POELES 1711-1790)



Bibliografia

S. Bertino - E. Sonnino (1991), *Nota preliminare sullo studio di una coorte di matrimoni*, Dipartimento di statistica, Probabilità e Statistiche Applicate, Università degli studi di Roma "La Sapienza", Pubblicazioni, Serie A, n.32.

M. Breschi (1985), *La fecondità di una parrocchia della Montagna Pistoiese: Treppio, 1790-1889*, S.I.D.E.S., La popolazione italiana nell'ottocento, CLUEB, Bologna.

G. Cabourdin (1981), *Le remariage en France sous l'Ancien Régime (seizième-dixhuitième siècles)*, AA.VV., Marriage and Remarriage in Population of the Past, Accademic Press.

D. Grassi (1992), *La sopravvivenza e la riproduzione di una coorte matrimoniale in funzione di regimi demografici determinati. Variazioni storiche dei processi e simulazioni con un modello macro*, Tesi di Laurea in Demografia, Facoltà di Scienze Statistiche Demografiche ed Attuariali, Università degli studi di Roma "La Sapienza".

M. Livi Bacci (1981), *On The frequency of remarriage in nineteenth century Italy: methods and results*, AA.VV., Marriage and Remarriage in Population of the Past, Accademic Press.

M. H. Jouan (1969), *Les originalités démographiques d'un bourg artisanal normand au XVIII siècle: Villedieu-les-Poêles (1711-1790)*, Annales de Démographie Historique.

R. Schofield - E. Wrigley (1981), *Remarriage intervals and the effect of marriage order on fertility*, AA.VV., Marriage and Remarriage in Population of the Past, Accademic Press.

A. Schiaffino (1981), *Quelques données sur le remariage dans un milieu urbain: Bologne aux dix-neuvième et vingtième siècles*, AA.VV., Marriage and Remarriage in Population of the Past, Accademic Press.

E. Sonnino (1992), *Precarietà sociale, fragilità familiare e ruoli delle istituzioni di assistenza ai poveri e alle orfane (Roma, secoli XVII-XVIII)*, La famiglia ieri e oggi, a cura di G. Da Molin, Cacucci Editore, Bari, Vol.1.