



UNIVERSITA' DEGLI STUDI DI PADOVA

FACOLTA' DI SCIENZE STATISTICHE

CORSO DI LAUREA IN SCIENZE STATISTICHE DEMOGRAFICHE E SOCIALI

TESI DI LAUREA SPECIALISTICA

**RICOSTRUZIONE NOMINATIVA DELLE FAMIGLIE IN UNA
COMUNITA' VENETA: BATTAGLIA TERME (1607 – 1871)**

RELATORE: CH.MO PROF. FIORENZO ROSSI

LAUREANDO: ANDREA PEGORARO

MATRICOLA: 564965 – SD

ANNO ACCADEMICO 2009 – 2010

A tutti coloro che mi sono stati
e che mi staranno vicino

SOMMARIO

I REGISTRI PARROCCHIALI DI BATTAGLIA TERME	7
1.1 LA STORIA DI BATTAGLIA TERME.....	7
1.2 I REGISTRI PARROCCHIALI.....	8
1.3 QUALITA' E COMPLETEZZA DEI DATI	10
1.4 I DATI DI BATTAGLIA TERME.....	11
1.4.1 LIBRI DEI BATTESIMI	12
1.4.2 LIBRI DELLE SEPOLTURE.....	14
UN ESAME DELLE VARIABILI DEMOGRAFICHE.....	17
2.1 LE NASCITE.....	17
2.2 I DECESSI	21
2.3 I MATRIMONI.....	29
LA RICOSTRUZIONE NOMINATIVA DELLE FAMIGLIE.....	33
3.1 LE TECNICHE NOMINATIVE.....	33
3.2 LA RICOSTRUZIONE NOMINATIVA DELLE FAMIGLIE.....	33
3.3 IL PROBLEMA DEI "LINKAGE"	35
3.4 LA SCHEDA DI FAMIGLIA.....	36
3.5 MISURE DEMOGRAFICHE RICAVABILI DALLA SCHEDA DI FAMIGLIA.....	37
3.5.1 LA MORTALITA' INFANTILE	37
3.5.2 MISURE DI FECONDITA'	39
3.6 LA RICOSTRUZIONE NOMINATIVA NEL CASO DI BATTAGLIA TERME	41
3.6.1 LA RICOSTRUZIONE E I PROBLEMI DI "LINKAGE"	41
3.6.2 LA SCHEDA DI FAMIGLIA.....	45
3.6.3 LA MORTALITA' INFANTILE	46
3.6.4 MISURE DI FECONDITA' DI BATTAGLIA TERME	49
UN ESEMPIO DI ANALISI DEI DATI.....	53
4.1 CHE COSA E' L'EVENT HISTORY ANALYSIS	53

4.2 DURATE CENSURATE	53
4.2.1 CENSURA A DESTRA E CENSURA A SINISTRA.....	54
4.2.2 CENSURE INFORMATIVE E NON INFORMATIVE	55
4.2.3 EFFETTO DELLE CENSURE (NON INFORMATIVE) SULLE ANALISI STATISTICHE.....	55
4.3 TEMPO DISCRETO	56
4.4 TEMPO CONTINUO	58
4.5 LE FUNZIONI STATISTICHE NEL CONTINUO	59
4.5.1 METODI NON PARAMETRICI PER DATI NEL TEMPO CONTINUO	61
4.6. IL MODELLO SEMIPARAMETRICO DI COX.....	65
4.7 L'INTERVALLO TRA PARTI NELLE FAMIGLIE DI BATTAGLIA TERME	67
4.8 OSSERVAZIONI CONCLUSIVE	73

INTRODUZIONE

Questo lavoro vuole apportare un modesto contributo alla conoscenza della storia demografica del paese di Battaglia Terme. Lo scopo fondamentale consiste nell'individuare il maggior numero di informazioni possibili desumibili dai libri dei Battesimi, dei Matrimoni e dei Decessi. A rendere più difficile il compito di ricostruzione si è aggiunto che il Paese era esente dall'obbligo di eseguire la numerazione delle anime, preziosissima fonte d'informazione per gli studiosi di demografia storica.

Si è deciso, quindi, di utilizzare i tre libri presenti nella Parrocchia di San Giacomo in Battaglia Terme, dove sono state consultate le fonti esistenti utilizzando il metodo nominativo come tecnica per la raccolta dei dati.

Nel Capitolo 1, oltre ad una breve descrizione storiografica del paese oggetto di studio, vi è una breve spiegazione dei registri utilizzati per eseguire la ricostruzione. Verrà compiuta anche una prima analisi per verificare la qualità e la copertura dei dati raccolti, sia per i Battesimi sia per le Sepolture.

Il Capitolo 2 presenta una prima analisi descrittiva dei tre libri (in ordine Battesimi, Decessi e Matrimoni), utile per capire la grandezza del fenomeno e per riuscire a inquadrare il comportamento e l'evoluzione, anche dal punto di vista culturale, di Battaglia Terme dal 1607 al 1871.

Gli ultimi due capitoli presentano la vera e propria analisi effettuata sui dati.

Il capitolo 3 spiega il metodo della ricostruzione nominativa, tecnica utilizzata per legare le informazioni presenti tra i vari registri. Dopo una spiegazione sui problemi che si possono incontrare nell'effettuare tale giuntura, si passerà a compiere operativamente tale operazione sui dati in nostro possesso. Solo alla fine di questo capitolo si cercherà di ottenere stime di mortalità e di fecondità.

Anche il capitolo 4 è strutturalmente composto come quello precedente: vi è una prima parte teorica nella quale si spiegano le tecniche utilizzate per l'analisi di fenomeni sociali e demografici (la "Event History Analysis"), in particolare si descrivono dalle durate censurate alle funzioni nel continuo e nel discreto, fino a giungere al modello semiparametrico di Cox. Solo successivamente si utilizzano le informazioni assimilate nei dati riguardanti la Parrocchia di San Giacomo per poter stimare un modello riguardante la distanza inter-parto per le donne che hanno avuto almeno un figlio.

CAPITOLO 1

I REGISTRI PARROCCHIALI DI BATTAGLIA TERME

1.1 LA STORIA DI BATTAGLIA TERME

I primi documenti attendibili sulle origini di “Batalia” risalgono all’inizio del XIII secolo, in concomitanza con lo scavo del canale operato dal Comune di Padova tra il 1189 e il 1201. Si tratta di due statuti che parlano di ponti, case e mulini costruiti dove il Canale di sotto riceve il naviglio, fatto scavare dal Comune per la navigazione tra Padova e Monselice.

In poco tempo questo borgo, da termale ad agricolo, si trasforma in un centro industriale e commerciale.

Alla fine del XVIII secolo anche le vicende riguardanti le acque termali rivestono un ruolo rilevante. Grazie a Pietro Estense Selvatico viene costruito un lussuoso e confortevole stabilimento che ospita fin da subito personaggi illustri e consolida la tradizione termale di Battaglia. Abbattuto il vecchio edificio, nel 1936 venne inaugurato lo stabilimento I.N.P.S. “Pietro d’Abano”:

Contemporaneamente a quanto sopra, la segheria e l’officina del maglio, nei primi anni del secolo XX, si trasformano nelle Officine Rinaldi. La crescita della fabbrica è talmente impetuosa che le officine diventano in breve tempo la maggiore industria locale. La fabbrica si amplia, diventa Officine di Battaglia, si fonde con le Officine Galileo di Firenze, diventa Officine Elettromeccaniche Galileo di Battaglia Terme e poi Magrini-Galileo. Con oltre un migliaio di operai, tecnici e impiegati diventa quasi il simbolo di Battaglia.

Attualmente, riscoperta la propria specificità geografica, storica e culturale, Battaglia si rilancia sul piano turistico come “porta” del Parco dei Colli Euganei, rivalutando il

considerevole patrimonio di bellezze naturali e storico artistiche che possiede, molte delle quali davvero uniche.¹

1.2 I REGISTRI PARROCCHIALI

Le prime disposizioni di carattere generale circa la tenuta regolare di registri parrocchiali furono impartite con il “Concilio di Trento” del 1563. Già da alcuni secoli si riscontrano, comunque, esempi di tenuta di tali registri, soprattutto per quanto riguarda i battesimi.

Occorre ricordare che il “Concilio di Trento” rese obbligatoria la custodia di registri di battesimo e di matrimonio, mentre solo nel 1614 (col Rituale Romanum di Papa Paolo V) si istituì l’obbligo delle sepolture.

Alla base delle registrazioni parrocchiali stanno motivazioni di carattere religioso, in particolare il controllo della regolarità della somministrazione dei sacramenti. Tenendo conto del fatto che l’Italia era per la stragrande maggioranza di religione Cattolica, sono assai poco frequenti le fonti nominative di tipo civile relative sia allo stato che al movimento della parrocchia. Bisogna quindi concludere che i registri parrocchiali di battesimo, matrimonio e sepoltura, insieme agli stati delle anime, costituiscono una base di dati veramente insostituibile per lo studio della struttura e delle caratteristiche evolutive delle popolazioni d’età moderna.

Sia per battesimi che per matrimoni che per sepolture si tratta di registrazioni nominative, contenenti quindi le generalità della persona ed, eventualmente, dei congiunti. Bisogna però considerare che, al momento dell’introduzione della normativa trentina, il processo di formazione dei cognomi era in molte parti ancora in atto. Di conseguenza le registrazioni più antiche spesso non forniscono con regolarità le generalità complete e definite degli individui e mal si presentano ad un impiego di tipo nominativo. In ogni caso, al di là di un’ovvia evoluzione temporale nella qualità delle registrazioni, vanno anche segnalate forte differenziazioni da zona a zona, legate in gran parte al livello culturale ed alla sensibilità dei singoli parroci, o anche dei vescovi, che esercitavano il controllo regolare dei registri.

¹ Informazioni ricavate dal sito internet <http://www.comune.battaglia-terme.pd.it>

Le informazioni usualmente desumibili dal libro dei Battesimi (baptismorum) sono:

- Data di battesimo
- Data di nascita
- Sesso (desumibile solamente dal nome)
- Nome del bambino
- Nome e cognome del padre
- Nome e talvolta cognome della madre
- Nome dei nonni
- Indicazione dell'eventuale illegittimità
- Luogo di domicilio dei genitori
- Professione del Padre (solo a partire dal XIX secolo)
- Padrino e madrina del battezzato

Il registro dei matrimoni (matrimoniorum) presenta invece le seguenti informazioni:

- Data delle pubblicazioni
- Data del matrimonio
- Nome, paternità e maternità degli sposi
- Età degli sposi (quasi mai prima del XIX secolo)
- Professione degli sposi (mai prima del XIX secolo)
- Luogo di residenza degli sposi (il matrimonio viene normalmente celebrato nella parrocchia della sposa)
- Stato civile al matrimonio
- Testimoni degli sposi
- Eventuali dispense per consanguineità o affinità

Il registro delle sepolture (mortuorum), infine, recita le seguenti variabili:

- Data di sepoltura
- Data di morte
- Nome e cognome del defunto
- Sesso

- Età alla morte
- Stato civile
- Paternità del defunto
- Luogo di residenza del defunto
- Professione (non prima del XIX secolo)
- Causa del decesso (sporadicamente)
-

Tra i caratteri rilevabili dai registri di battesimo non figura l'età della madre alla nascita del figlio. E' senza dubbio questa la mancanza più grave per l'impiego dei registri parrocchiali come fonte demografica. Questa lacuna è comunque parzialmente superabile nell'ambito delle tecniche di tipo nominativo che utilizzeremo successivamente. Occorre inoltre precisare che, negli atti di sepoltura, l'età del decesso è di norma abbastanza curata per quanto riguarda i bambini, mentre con il crescere dell'età l'imprecisione aumenta sensibilmente.

Altra informazione importante è che almeno fino al XIX secolo non tutte le parrocchie erano dotate di fonte battesimale. In molti casi, quindi, si dispone esclusivamente dei registri di matrimonio e di sepoltura delle singole parrocchie ed è necessario rintracciare gli atti di battesimo degli appartenenti alla parrocchia studiata nei registri delle pievi o dei battisteri.

Infine, per il fatto stesso di riferirsi a singole parrocchie, i registri parrocchiali mal si presentano a studi demografici in larga scala, mentre possono essere considerati le fonti primarie per compiere analisi approfondite su popolazioni di ridotte dimensioni.

1.3 QUALITA' E COMPLETEZZA DEI DATI

I maggiori problemi di interpretazione dei dati ricavati dalle fonti ecclesiastiche possono essere raggruppati in due categorie sintetiche: errori di copertura ed errori di accuratezza. Per copertura si intende l'universalità della rilevazione. Come già accennato, le registrazioni parrocchiali non riguardano la totalità delle persone, ma solamente coloro

che appartengono alla confessione religiosa cui si riferiscono i registri. Questa limitazione, però, non preclude l'analisi attraverso una loro utilizzazione, in quanto la proporzione di popolazione che solitamente sfugge è quantitativamente poco consistente e spesso ben conosciuta. I casi di omissioni selettive o sistematiche, la cui presenza è necessario svelare se si vuole avere limiti accettabili di sicurezza sulle misure ricavabili dai dati, sono in particolar modo relativi ad atti di battesimo o sepoltura. Per omissione selettiva ci si riferisce, per esempio, quando non si trascrive nell'apposito registro ecclesiastico un battesimo di un bimbo morto immediatamente dopo il parto. L'omissione sistematica avviene quando vi è una abituale mancata registrazione dei decessi dei bambini, da cui risultano seri problemi per la misura della mortalità nei primi anni di vita. Meno soggetti ad omissioni sembrano le rilevazioni dei matrimoni, probabilmente per la maggiore pubblicità dell'evento e per la sua chiara rilevanza sociale. Per quanto riguarda gli stati delle anime, le omissioni possono riguardare particolari classi di età, più precisamente quelle della prima e della primissima infanzia.

Per quanto riguarda gli errori relativi all'accuratezza, sicuramente la variabile di maggiore interesse è l'età. La correzione di tali distorsioni diventa compito importante proprio per poter attribuire con precisione gli eventi della popolazione delle singole classi di età che li ha generati.

1.4 I DATI DI BATTAGLIA TERME

I dati necessari per analizzare la popolazione di Battaglia Terme sono stati rilevati dai registri conservati presso la parrocchia di San Giacomo e trascritti in Excel dal dott. Luciano Donato.

Questi registri dovrebbero di norma essere quattro, ma in questa parrocchia ve ne sono presenti solamente tre: non è presente il registro degli stati delle anime.

1.4.1 LIBRI DEI BATTESIMI

I libri delle nascite contengono 17868 rilevazioni effettuate dal Maggio del 1608 fino al Giugno del 1871.

Le variabili registrate (ove presenti) in questi 12 libri sono:

- Data di nascita e di battesimo
- Sesso
- Nome e cognome
- Nome, cognome e soprannome del padre
- Mestiere del padre
- Data e luogo di nascita del padre
- Nome del nonno da parte del padre
- Nome, cognome e soprannome della madre
- Mestiere della madre
- Data e luogo di nascita della madre
- Nome del nonno da parte della madre
- Data e luogo di matrimonio dei genitori
- Indirizzo dei genitori
- Eventuali nati gemelli
- Nomi e cognomi dei padrini
- Nome e cognome della levatrice
- Note varie
- Pio luogo
- Incogniti

Uno dei possibili mezzi di analisi per valutare la bontà di copertura degli atti di battesimo passa per la determinazione del rapporto dei sessi alla nascita. L'eccedenza dei maschi alla nascita è un carattere tipico della specie umana, ed è misurata tramite il rapporto di

mascolinità, dato dal numero di maschi per ogni 100 femmine. Tale rapporto assume, per numerosità consistenti, valori oscillanti nel rapporto 104-107.

Si è suddiviso l'intervallo di tempo dalla prima all'ultima rilevazione in sei periodi (di diversa ampiezza). Contando il numero di maschi e di femmine per ogni intervallo e dividendo il primo valore per il secondo si ottiene il rapporto di mascolinità.

Periodo	Maschi	Femmine	Sconosciuto	Totale
1607-1649	1364	1249	5	2618
1650-1699	1764	1729	5	3498
1700-1749	1918	1739	3	3660
1750-1799	1759	1673	0	3432
1800-1849	1744	1645	6	3395
1859-1871	662	603	0	1265
Totale	9211	8638	19	17868

Tabella 1.4.1.1 Numero di nascite suddivise per sesso e periodo di nascita

Periodo	Rapporto di mascolinità	I.C. Inferiore	I.C. Superiore
1607-1649	109,2	98,8	112,33
1650-1699	102	99,65	111,36
1700-1749	110,3	99,78	111,22
1750-1799	105,1	99,6	111,41
1800-1849	106	99,57	111,45
1859-1871	109,8	96,06	115,54

Tabella 1.4.1.2 Rapporto di mascolinità suddiviso per periodo di nascita

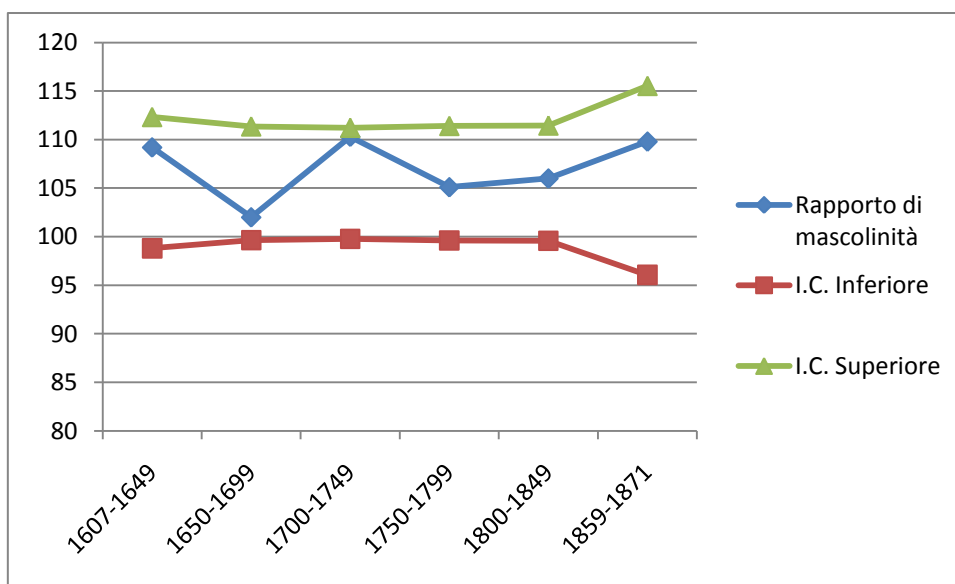


Grafico 1.4.1.1 Rapporto di mascolinità e intervalli di confidenza al 95% per anno di nascita

Nel grafico 1.4.1.1 si evidenzia come tale valore si trovi all'interno delle linee rappresentanti un intervallo di confidenza al 95%, il che porta a concludere che i dati, almeno per quanto riguarda questa componente, sono stati raccolti in maniera sufficientemente adeguata.

1.4.2 LIBRI DELLE SEPOLTURE

Nei libri delle sepolture di Battaglia Terme sono stati registrati 16310 decessi, suddivisi tra il 1664 e il 1871.

Le informazioni riportate nei suddetti registri sono:

- Nome, cognome e soprannome del defunto
- Data di nascita e di morte del defunto
- Età presunta della morte
- Nome e soprannome del padre
- Nome e cognome della madre
- Parrocchia di appartenenza

- Patria
- Luogo della sepoltura
- Gemellarità
- Causa di morte
-

La forma dell'atto di sepoltura assume caratteristiche differenti a seconda che la persona deceduta fosse un bambino, una donna sposata oppure una donna nubile. Per quanto riguarda i bambini, per indicare la illegittimità, poteva comparire nell'atto le frasi "Ca' di Dio", "Padre incognito" oppure "Genitori sconosciuti". Nel caso di donne sposate veniva riportato anche il nome, il cognome e l'eventuale soprannome del marito.

In questi registri la popolazione di riferimento è quella "presente" nella parrocchia. Mancano, quindi, dalle registrazioni, i dati relativi alle sepolture di parrochiani morti al di fuori di essa. Un'altra categoria che può facilmente dare adito ad omissioni selettive è quella dei nati morti: questi casi possono mancare sia dal registro di morte (ed essere presenti in quello dei battesimi), sia contemporaneamente dal registro dei battesimi e delle sepolture.

Per quanto riguarda i problemi relativi all'accuratezza nella rilevazione delle sepolture, la prima considerazione da farsi è sull'età dei deceduti. Il numero di casi che solitamente si addensa attorno ad alcune età al decesso risulta più elevato di quello che si riscontra in altre età limitrofe. I valori usualmente preferiti nella dichiarazione sono quelli terminanti in 0, in 5 o in cifre pari.

Nel nostro caso gli anni di morte sono stati suddivisi in 24 classi di età: la prima classe include i bambini morti entro la prima settimana, la seconda quelli morti dalla prima settimana al primo mese, le terza dal primo mese al primo anno, la quarta dal primo anno ai cinque anni, le successive classi hanno intervalli di 5 anni, fino all'ultima classe che va dai 101 ai 105 anni. Con questa suddivisione possiamo solamente valutare il rapporto tra gli intervalli che includono le età che terminano in 0 e quella in 5.

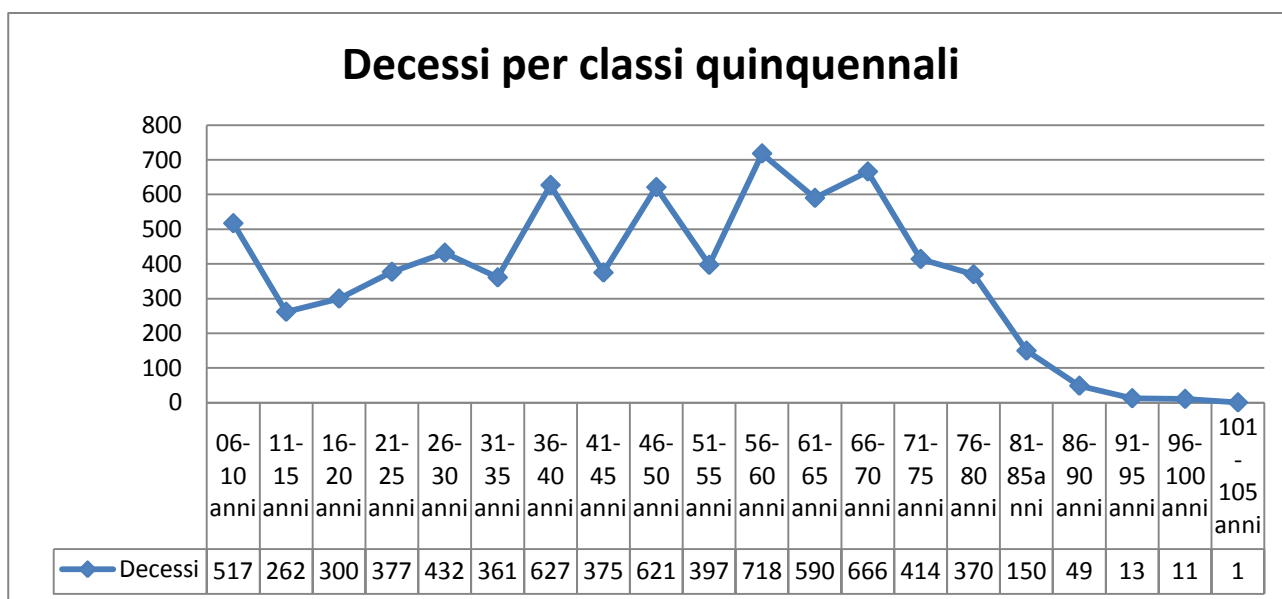


Grafico 1.4.2.1 Decessi per classi quinquennali

Il grafico in figura 1.4.1.2 illustra solamente i dati della quinta classe poiché quelle precedenti hanno ampiezze diverse e quindi non sono confrontabili. Risulta evidente quello detto precedentemente soprattutto analizzando le classi centrali, dai 31-35 ai 71-75 anni: i picchi presenti nella curva indicano come l'arrotondamento alle età che terminano con la cifra "0" sia presente nei dati in esame.

CAPITOLO 2

UN ESAME DELLE VARIABILI DEMOGRAFICHE

Nel seguente capitolo procederemo con una analisi descrittiva delle nascite, dei decessi e dei matrimoni avvenuti a Battaglia Terme tra il 1607 e il 1871.

2.1 LE NASCITE

Le registrazioni effettuate nel libro dei battesimi sono un buon strumento per analizzare le nascite avvenute a Battaglia Terme nel periodo considerato.

Come prima analisi calcoliamo le medie decennali (tranne per la prima classe che sarà una media quadriennale e l'ultima che sarà una media undecennale) dei battesimi nel periodo di cui disponiamo le rilevazioni.

Il grafico 2.1.1 mette in luce come l'andamento delle nascite sia particolarmente altalenante. Sono comunque evidenti quattro bruschi cali causati principalmente dalle epidemie di peste, tifo e colera ed ancora tifo avvenute in Veneto tra il 1600 e il 1850.

Anni	Valore assoluto	Media decennale	Anni	Valore assoluto	Media decennale
1607-1610	228	57	1741-1750	726	72,6
1611-1620	654	65,4	1751-1760	755	75,5
1621-1630	527	52,7	1761-1770	754	75,4
1631-1640	671	67,1	1771-1780	658	65,8
1641-1650	493	49,3	1781-1790	593	59,3
1651-1660	597	59,7	1791-1800	615	61,5
1661-1670	730	73	1801-1810	759	75,9
1671-1680	784	78,4	1811-1820	779	77,9
1681-1690	703	70,3	1821-1830	666	66,6
1691-1700	639	63,9	1831-1840	619	61,9
1701-1710	744	74,4	1841-1850	554	55,4
1711-1720	728	72,8	1851-1860	561	56,1
1721-1730	752	75,2	1861-1871	652	59,27
1731-1740	695	69,5			

Tabella 2.1.1 Valore assoluto e media decennale dei battesimi per periodi

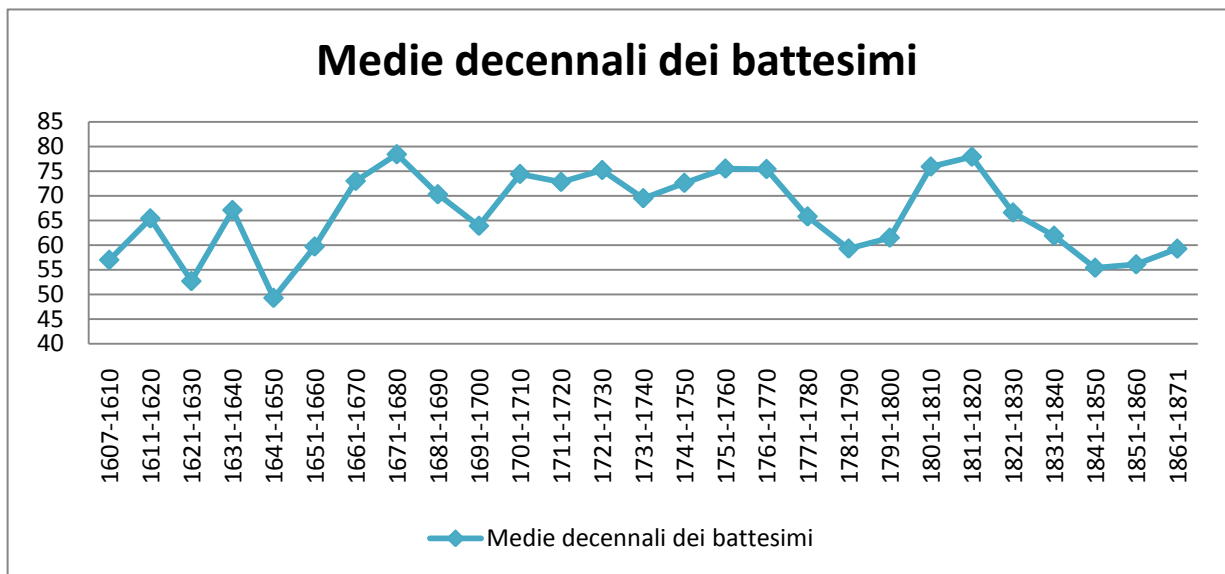


Grafico 2.1.1 Medie decennali dei battesimi di Battaglia Terme dal 1607 al 1871

Un altro elemento importante per analizzare le nascite è lo studio della stagionalità. Un problema che risiede in questo tipo di analisi sta nel fatto che i mesi hanno una durata diversa l'uno dall'altro. Per ovviare a questo problema si procede, dopo aver sommato le nascite dei periodi considerati (il primo periodo di 43 anni, i successivi quattro di 50 e l'ultimo di 21) come se ogni mese fosse composto da 30 giorni, modificando quindi i valori dei mesi con durata differente. Alla fine si moltiplicherà il valore risultante per 100, ottenendo così una percentuale mensile dei nati per ogni periodo. Dopo aver effettuato questi calcoli si tratterà un grafico dove si possano confrontare tra di loro i periodi.

	1607-1649	1650-1699	1700-1749	1750-1799	1800-1849	1850-1871	Media mensile
Gennaio	10,6	9,8	9,6	9,5	8,5	8,7	9,5
Febbraio	11,9	11,7	11,9	9,7	9,6	8,5	10,7
Marzo	12,6	11,4	11,2	11	9	9,2	10,8
Aprile	8,9	9,9	9,9	11,3	10,7	9,4	10,2
Maggio	7	8,8	8,8	9	9,5	9,7	8,8
Giugno	5,5	5,7	5,7	6,4	6,2	7,5	6
Luglio	6,1	5,8	5,9	7,7	8,6	8,2	7
Agosto	6,5	7	7,1	7,1	7,9	8,8	7,3
Settembre	7,9	6,9	6,8	6,2	7,7	8,4	7,2
Ottobre	8,3	7,4	7,5	7,1	8,1	6,8	7,6
Novembre	8,1	8	7,9	7,8	7	6,6	7,6
Dicembre	6,7	7,6	7,7	7,2	7,2	8,3	7,4

Tabella 2.1.2 Percentuale e media mensile per le nascite suddivise per mese e periodo

	1607-1649	1650-1699	1700-1749	1750-1799	1800-1849	1850-1871	Media mensile
Gennaio	3	4	4	4	6	5	4
Febbraio	2	1	1	3	2	6	2
Marzo	1	2	2	2	4	3	1
Aprile	4	3	3	1	1	2	3
Maggio	8	5	5	5	3	1	5
Giugno	12	12	11	11	12	10	12
Luglio	11	11	7	7	5	9	11
Agosto	10	9	9	9	8	4	9
Settembre	7	10	12	12	9	7	10
Ottobre	5	8	9	9	7	11	6
Novembre	6	6	6	6	11	12	6
Dicembre	9	7	8	8	10	8	8

Tabella 2.1.3 Cardinalità delle nascite suddivise per mese e periodo

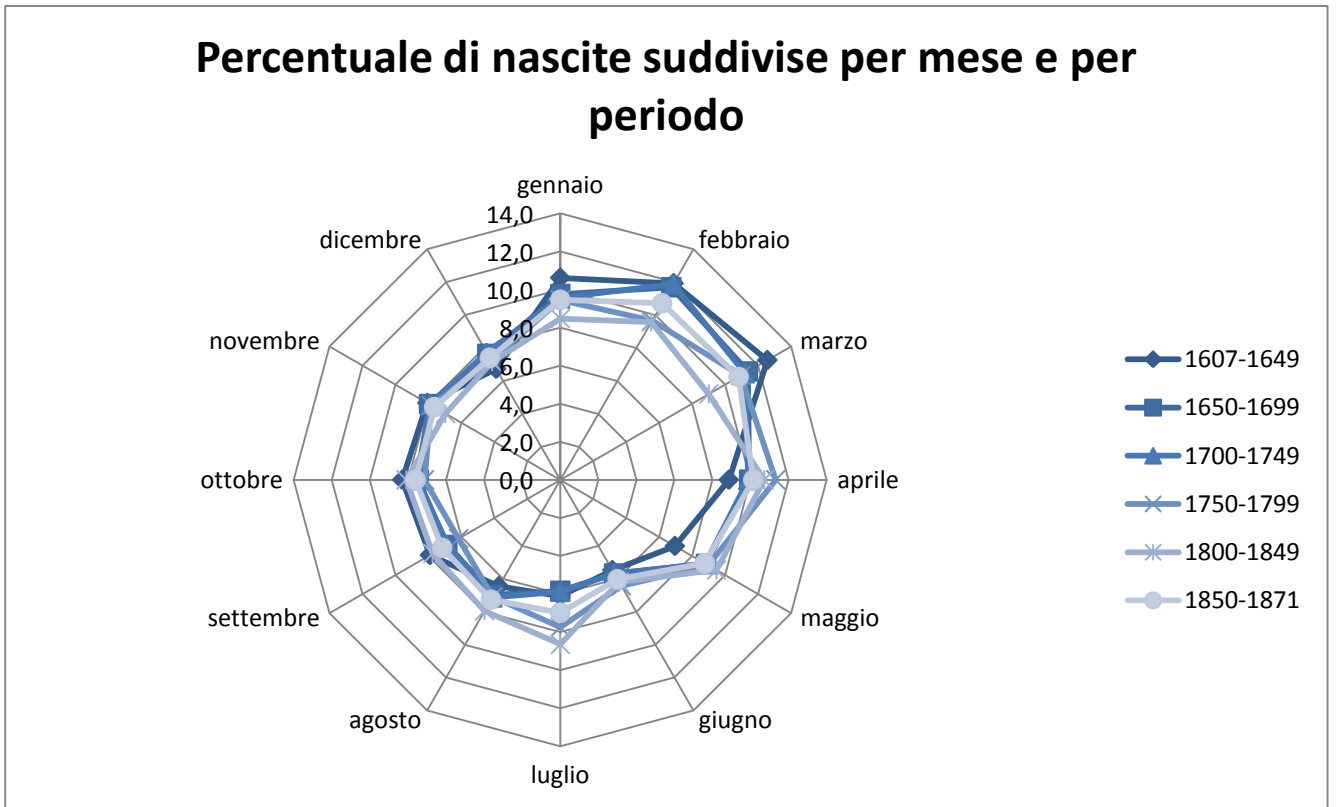


Grafico 2.1.4 Percentuali di nascite suddivise per mese e per classi

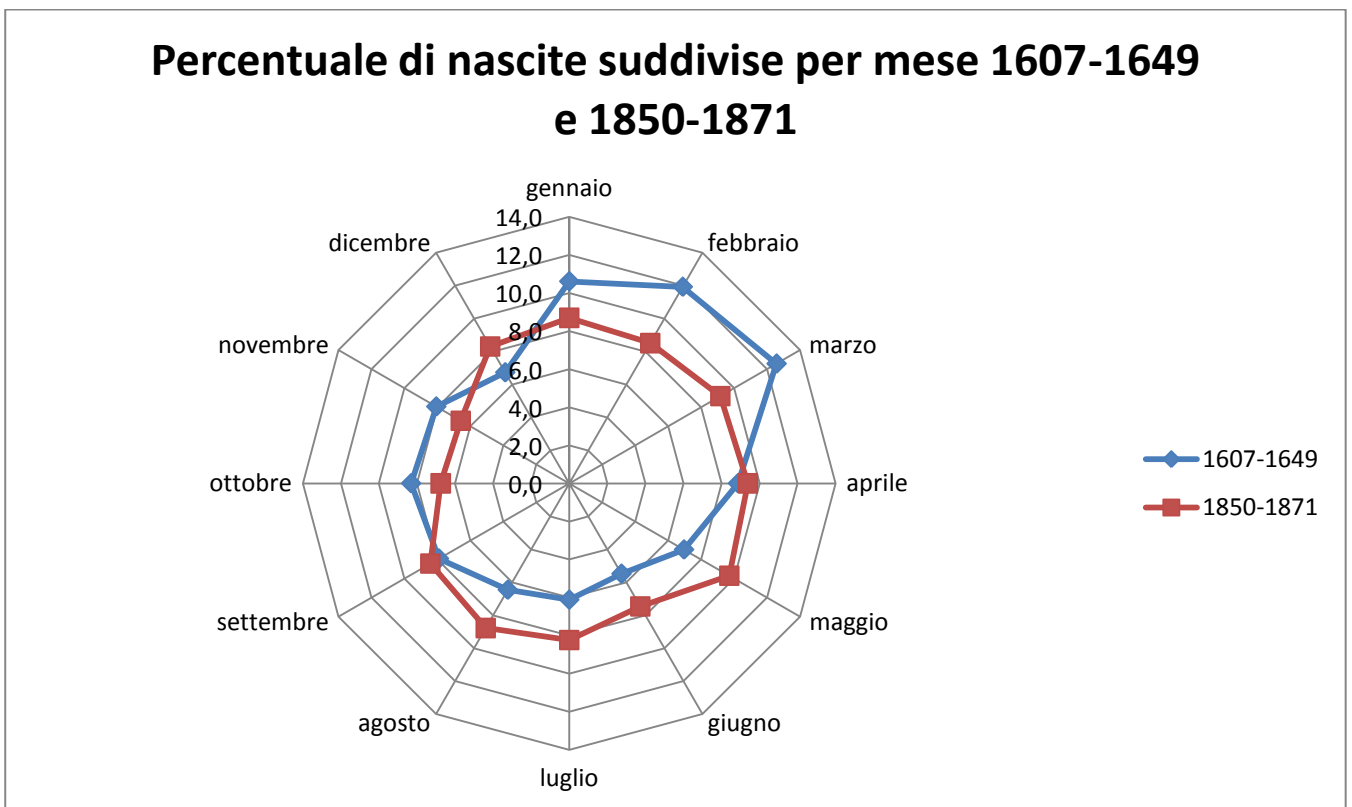


Grafico 2.1.5 Percentuali di nascite suddivise per mese e tra il 1607 e il 1649 e tra il 1850 e il 1871

Dai dati della tabella 2.1.2 si intuisce come con l'avanzare degli anni la percentuale di nascite attorno ai mesi invernali tende a diminuire. Questo cambiamento è maggiormente percepibile confrontando le percentuali del primo e dell'ultimo periodo. All'inizio del Seicento i mesi con maggiore natalità erano marzo e febbraio (12,6% e 11,9% delle nascite), ridotti 250 anni dopo al terzo e al sesto posto, a discapito di mesi più caldi come maggio e aprile. Anche i mesi che tra il 1607 e il 1649 avevano una natalità minore (giugno 5,5% e luglio 6,1%) lasciano successivamente posto a mesi più freddi come novembre ed ottobre.

Era consuetudine nelle popolazioni rurali far coincidere le nascite nei mesi di gennaio, febbraio e marzo, quando il lavoro nei campi era minore. Questo slittamento della nascite può quindi voler dire un cambiamento di stile di vita nel paese di Battaglia Terme, passando da una vita rurale ad una più urbana.

2.2 I DECESSI

Una prima analisi effettuabile è quella di considerare i decessi secondo il sesso, suddividendoli in periodi decennali. I dati disponibili vanno dal 1664 al 1871, un periodo di tempo inferiore rispetto a quello delle nascite.

	Maschi	Femmine	Incerti	Totale		Maschi	Femmine	Incerti	Totale
1664-1670	328	275	3	606	1771-1780	352	401	3	756
1671-1680	471	498	13	982	1781-1790	378	350	0	728
1681-1690	403	345	1	749	1791-1800	421	366	1	788
1691-1700	393	449	7	849	1801-1810	384	348	0	732
1701-1710	394	378	6	778	1811-1820	403	393	101	897
1711-1720	425	406	0	831	1821-1830	358	304	1	663
1721-1730	387	367	12	766	1831-1840	382	358	7	747
1731-1740	418	452	14	884	1841-1850	369	329	6	704
1741-1750	412	441	4	857	1851-1860	304	300	2	606
1751-1760	458	433	3	894	1861-1871	308	247	1	556

Tabella 2.2.1 Numero di defunti suddivisi per sesso e per periodi decennali

Periodo	Maschi	Femmine	Incerti	Totale
1664-1700	1595	1567	24	3186
1701-1750	2036	2044	36	4116
1751-1800	2066	2017	13	4096
1801-1850	1896	1732	115	3743
1851-1871	612	547	3	1162
Totale	8205	7907	191	16303

Tabella 2.2.2 Numero di defunti suddivisi per sesso e per periodi di 50 anni

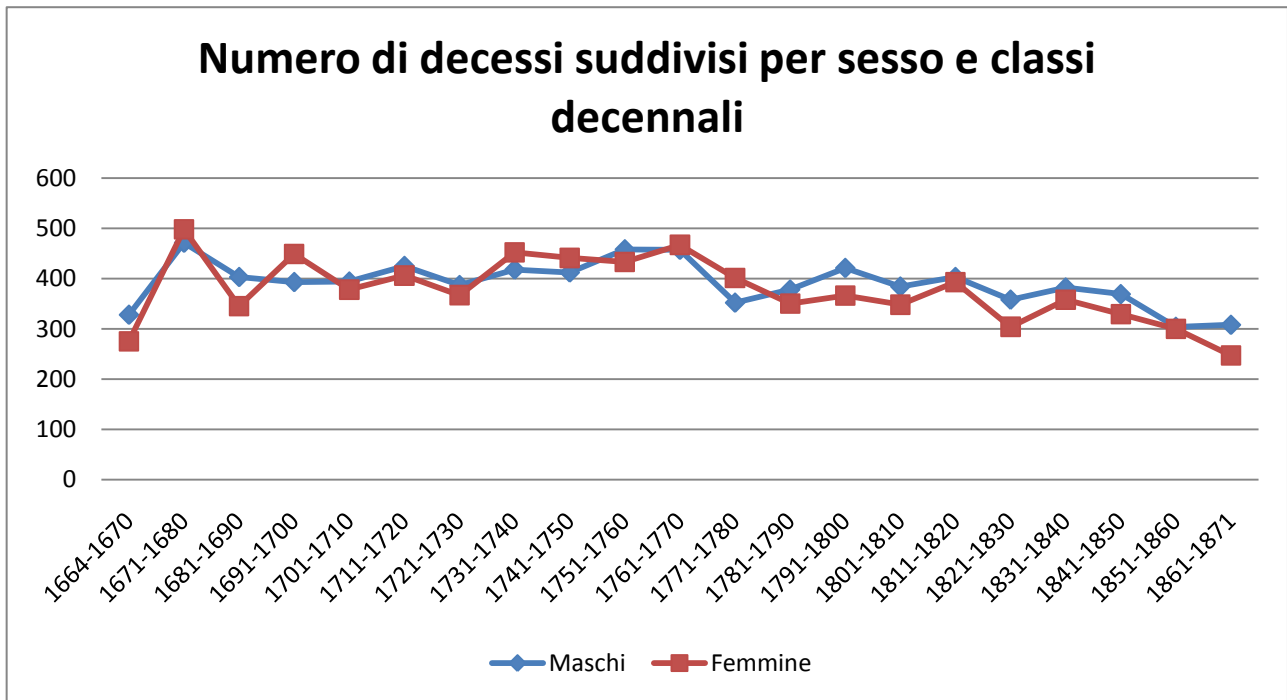


Grafico 2.2.1 Decessi suddivisi per sesso e periodi decennali

Interessante è notare come il numero di decessi di sesso maschile sia quasi sempre maggiore rispetto a quello femminile. Le poche volte in cui si manifesta il contrario, riusciamo a trovare una corrispondenza nella tabella delle nascite: quando nascono più femmine, ne muoiono anche di più. Questo può farci capire come il totale dei decessi sia fortemente influenzato dalle morti infantili.

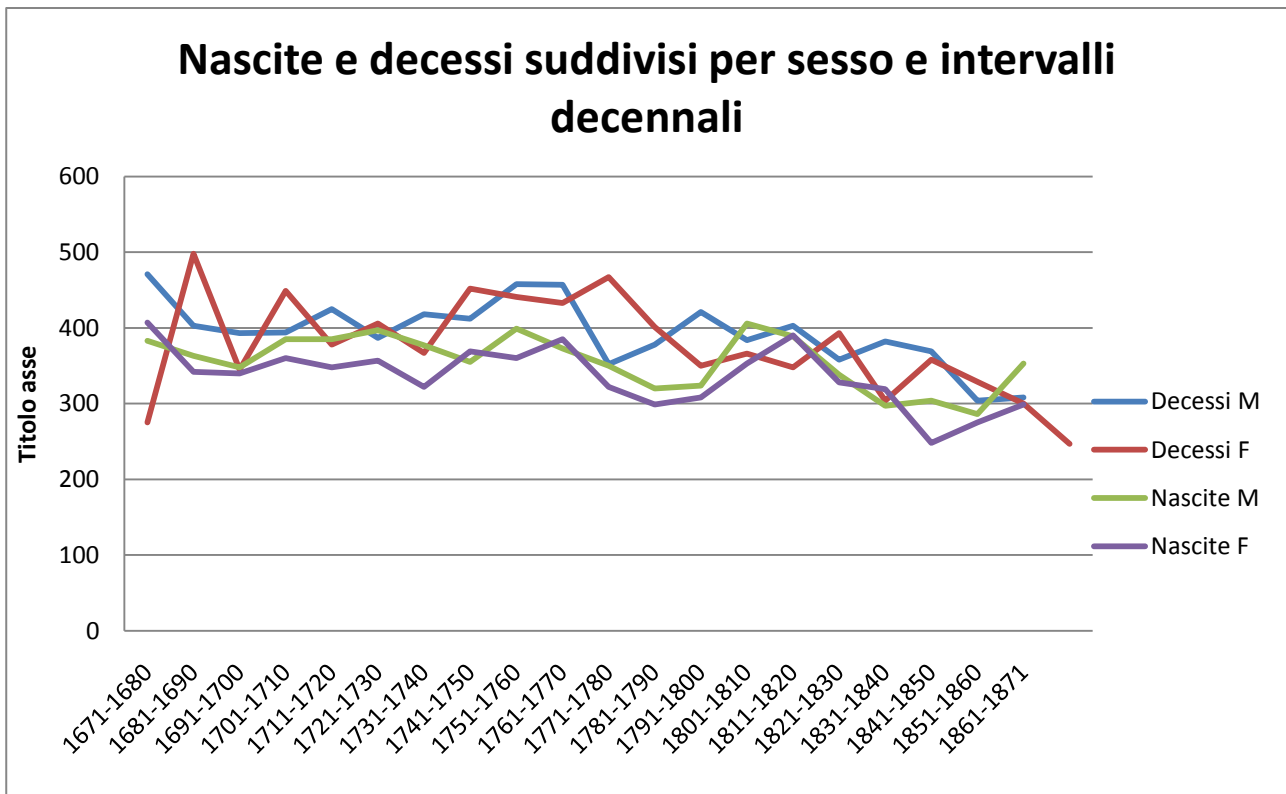


Grafico 2.2.2 nascite e decessi suddivisi per sesso e periodi decennali

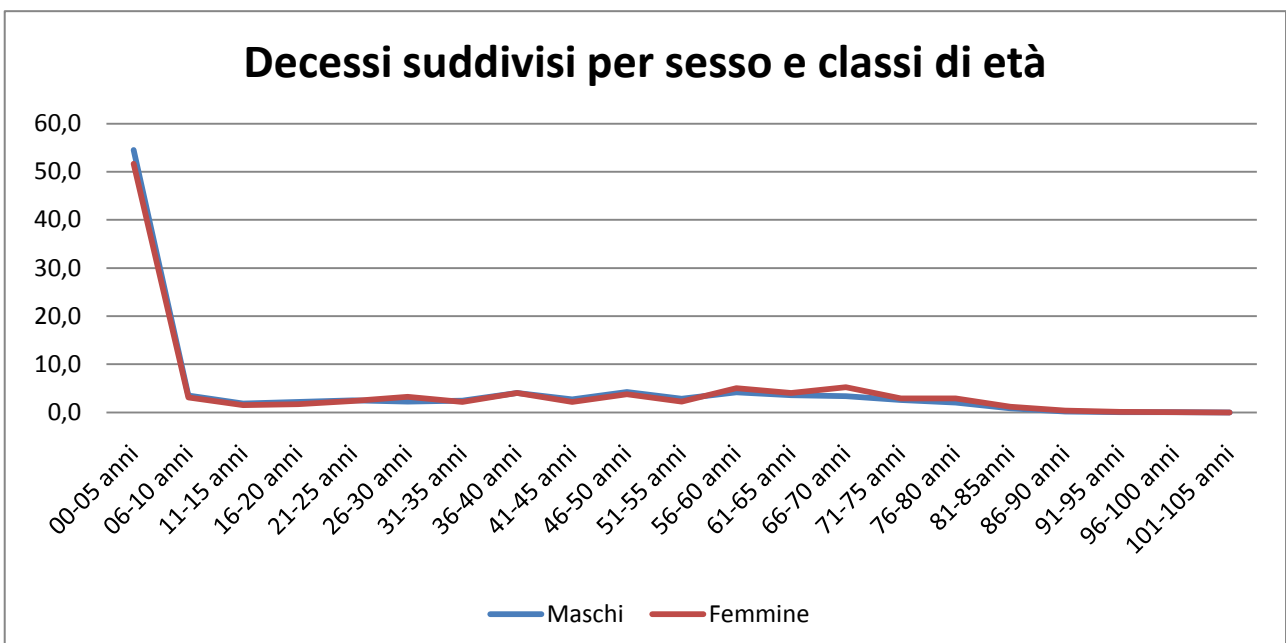


Grafico 2.2.3 Nascite e decessi suddivisi per sesso e periodi decennali. Medie dell'intero periodo

Il grafico 2.2.3 illustra come la maggior parte dei decessi coinvolga i bambini fino ai 5 anni di età, colpendo principalmente i maschi rispetto alle femmine. Eliminando dall'analisi proprio la prima classe risulta evidente, nonostante i picchi dovuti, come visto nel capitolo precedente, all'arrotondamento alle cifre tonde, che fino ai 55 anni la mortalità colpisce maggiormente i maschi, mentre, nelle età più avanzate, le donne. Il valore maggiore per i maschi si trova nell'intervallo 46-50 anni (4,3% delle morti dell'intera serie), mentre per le donne tra i 66 e i 70 anni (5,3% delle morti).

Inoltre, analizzando nello specifico i primi 5 anni di vita, si vede che il 16,9% dei maschi e il 15,5% delle femmine muoiono nella prima settimana di vita. Alla fine dei cinque anni, la percentuale di decessi raggiunge il 54,5% per i maschi e il 51,6% per le femmine.

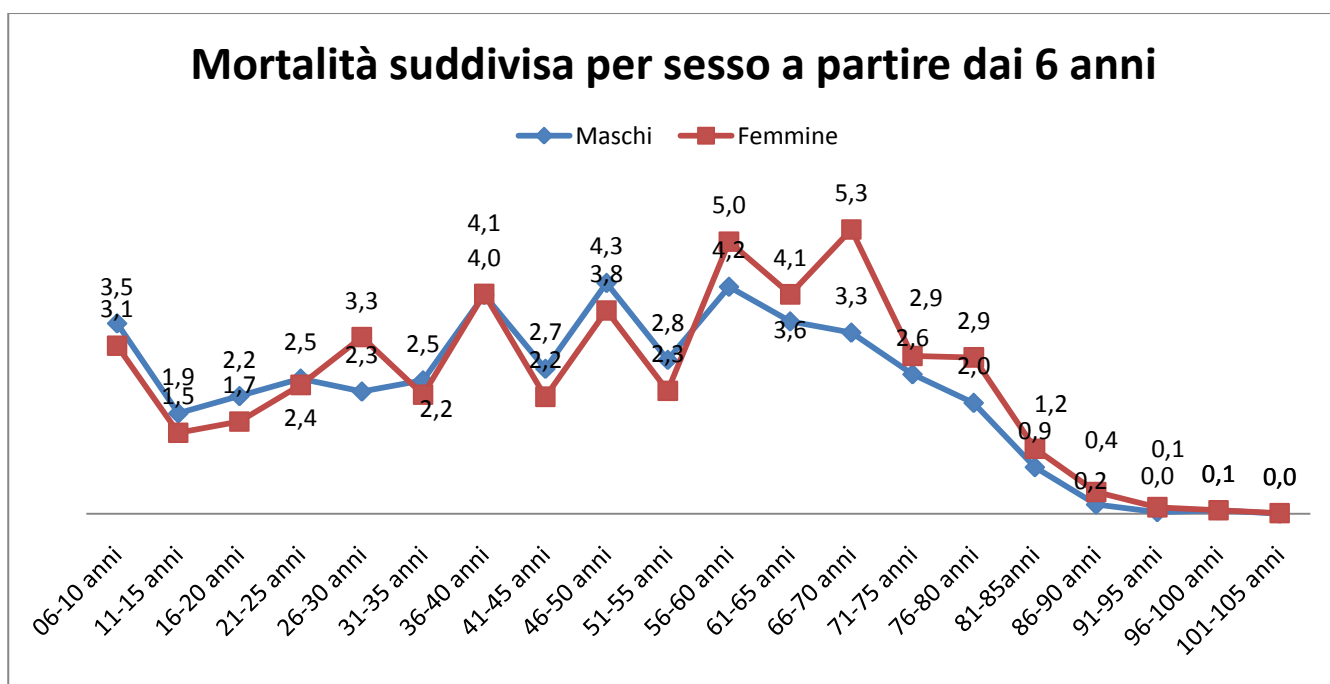


Grafico 2.2.4 Decessi suddivisi per sesso e classi decennali (a partire dal sesto anno di età). Medie percentuali dell'intero periodo.

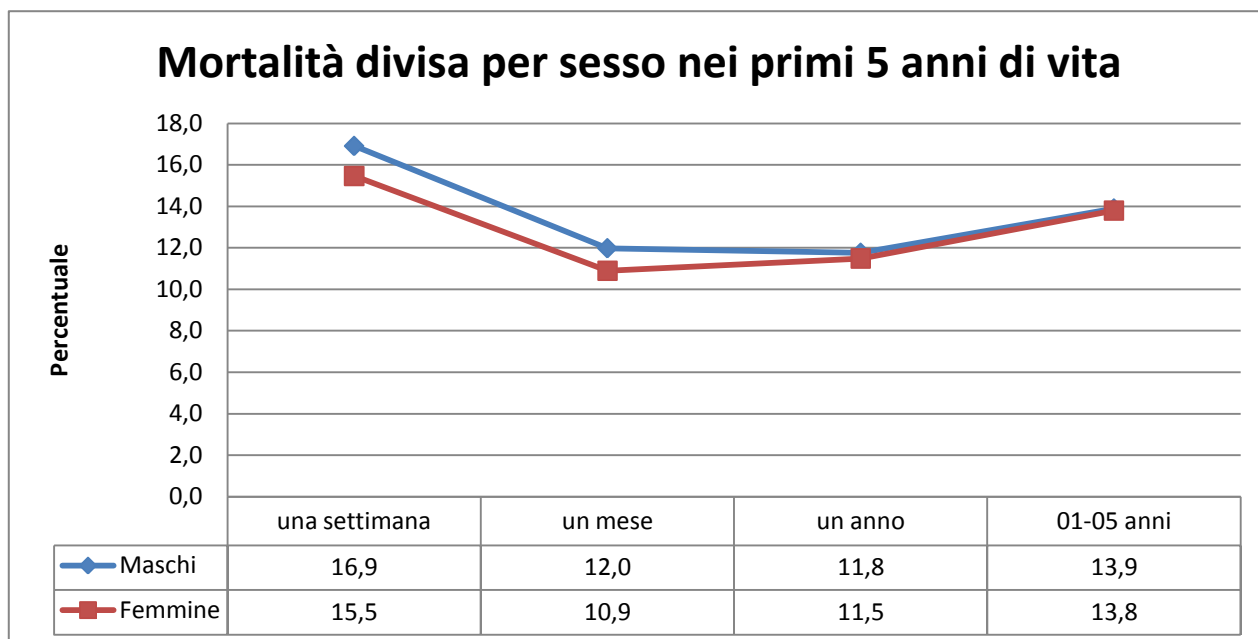


Grafico 2.2.5 Decessi nei primi anni di vita suddivisi per sesso. Medie percentuali dell'intero periodo sul totale dei decessi.

Oltre al confronto tra maschi e femmine, è interessante studiare anche l'andamento dei decessi nel corso del tempo. Proprio per questo motivo studiamo 4 serie (1664-1670, 1701-1710, 1751-1760 e 1861-1871) e le confrontiamo tra di loro.

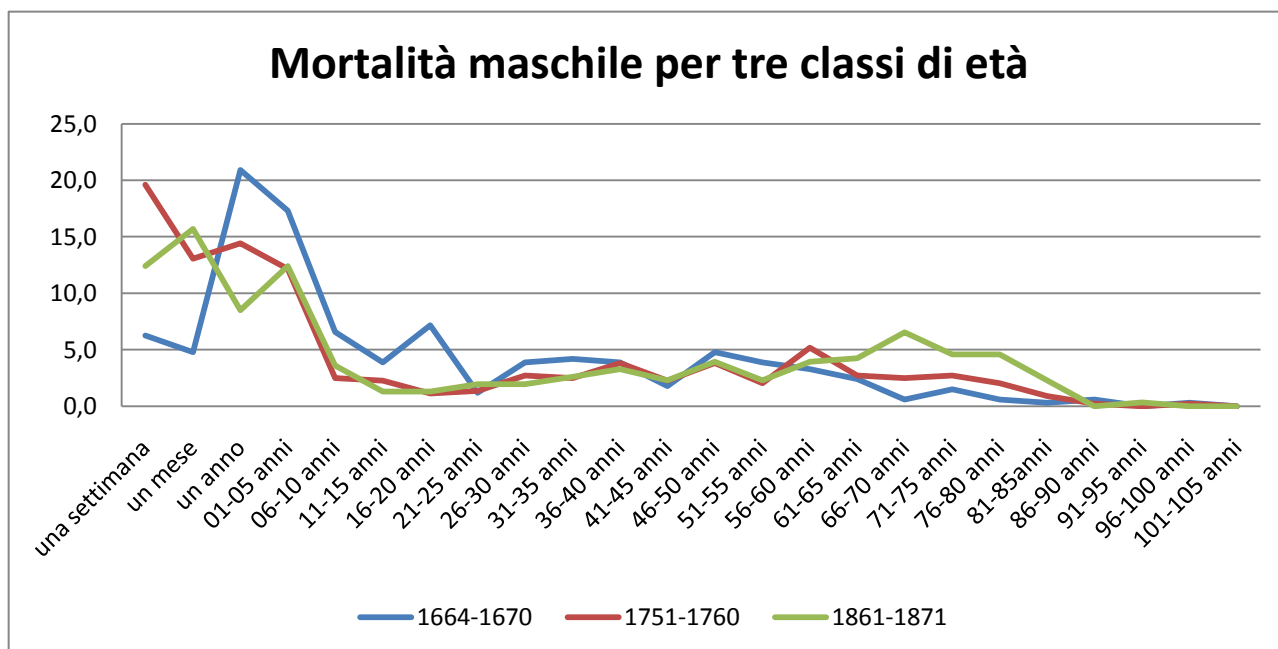


Grafico 2.2.6 Mortalità maschile per i periodi 1664-1670, 1751-1760 e 1861-1871

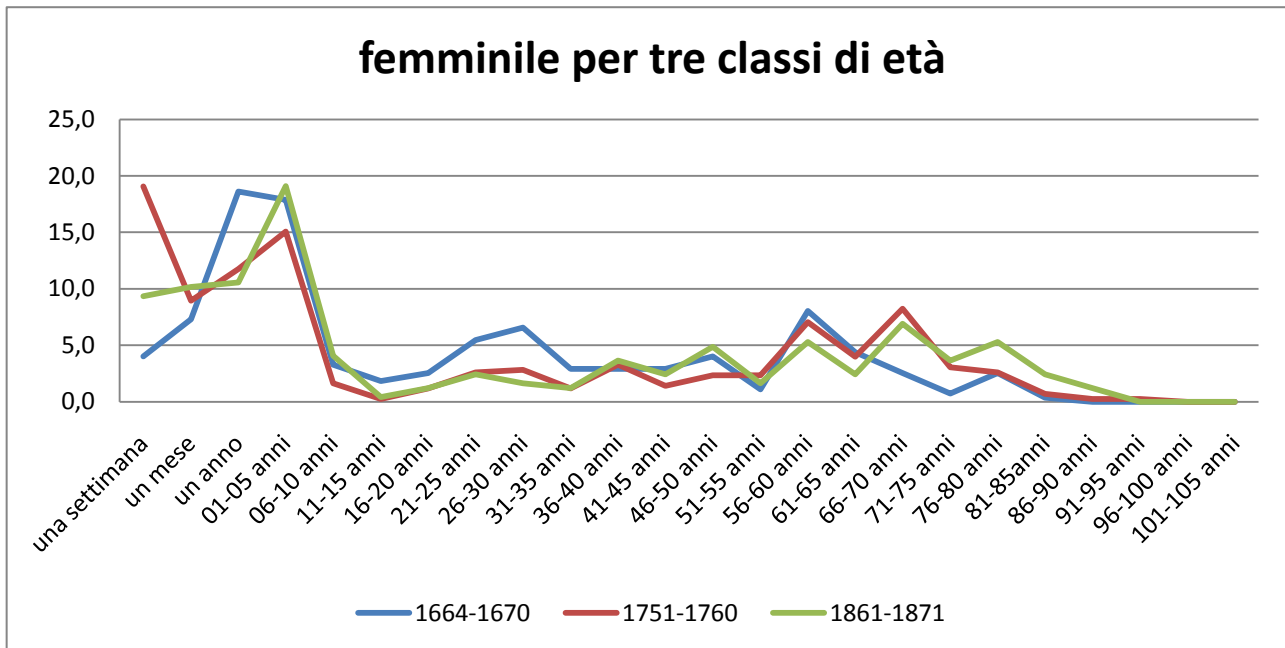


Grafico 2.2.7 Mortalità femminile per i periodi 1664-1670, 1751-1760 e 1861-1871

Tralasciando le prime quattro classi di età che sono irregolari e che sono quelle maggiormente soggette ad errori, si può notare come col passare del tempo la vita media si allunghi, sia per i maschi che per le femmine. Questo serve a sottolineare come nel corso di 250 anni le condizioni di vita a Battaglia Terme siano sensibilmente migliorate.

Oltre alla mortalità, anche la stagionalità dei decessi può essere considerato un indice sul fatto che un popolazione si sia sviluppata o meno.

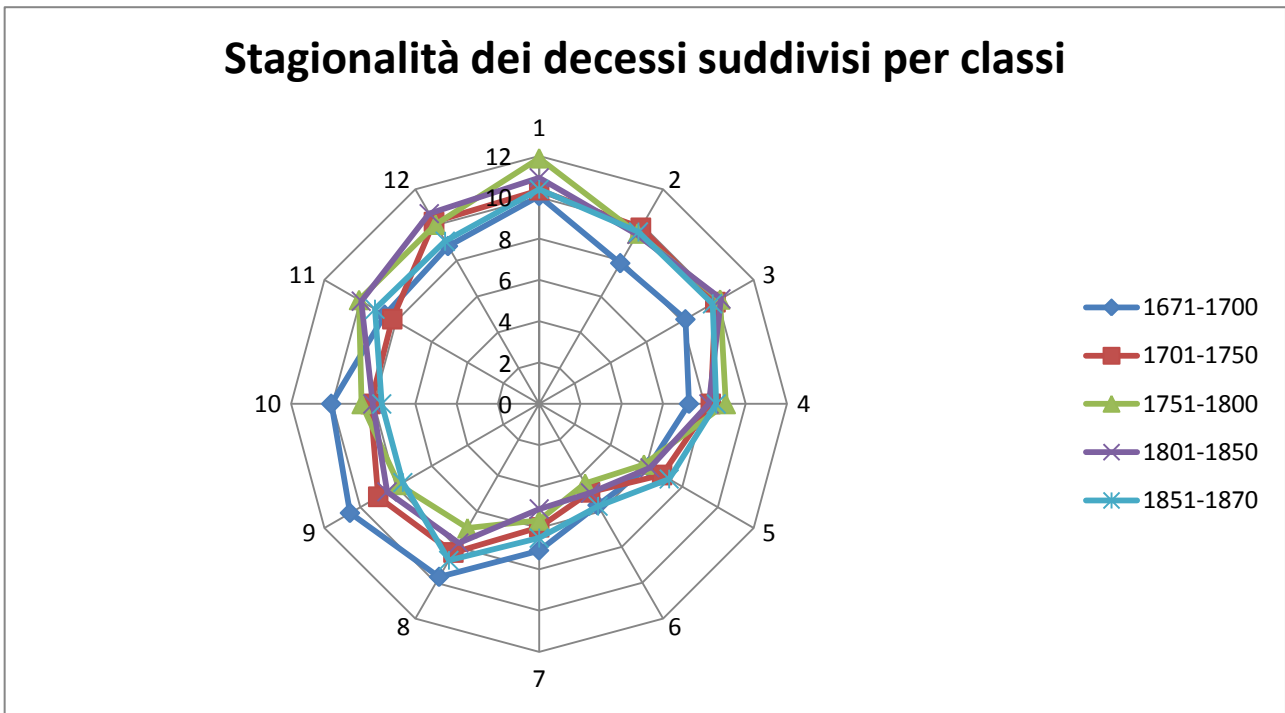


Grafico 2.2.8 Stagionalità dei decessi suddivisi per periodi

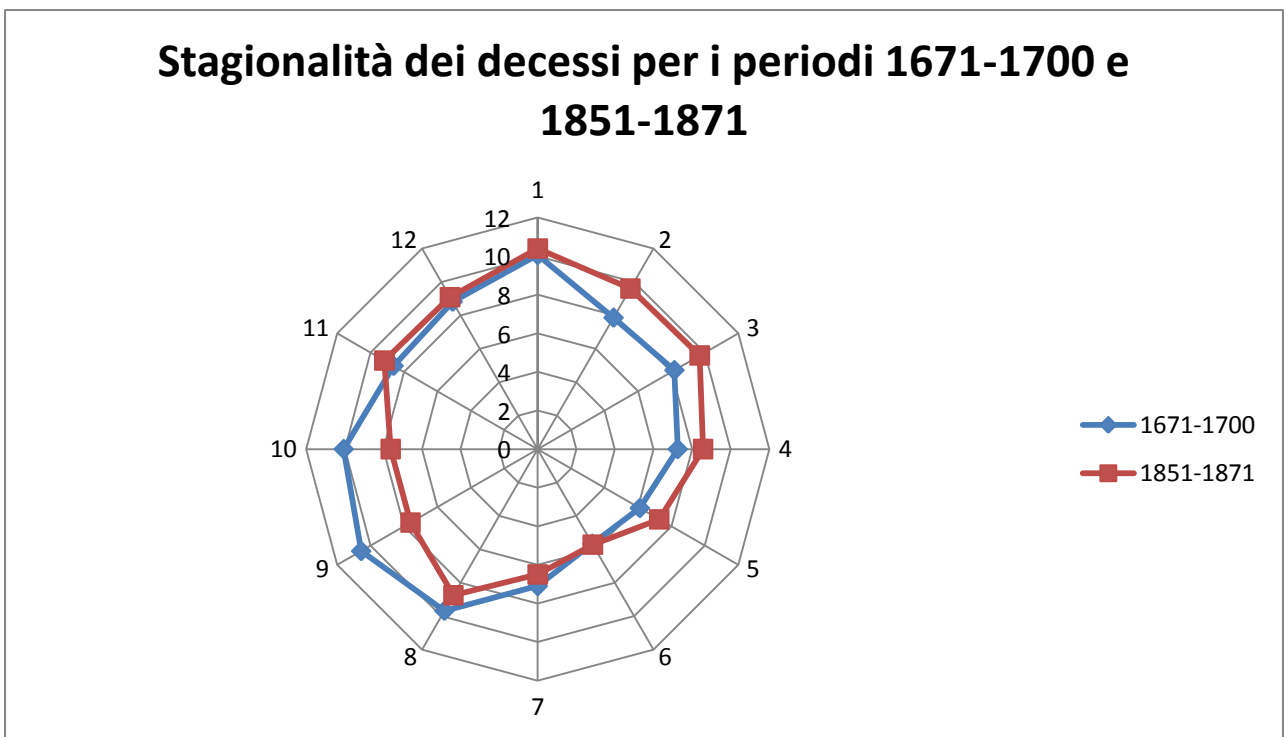


Grafico 2.2.9 Mortalità stagionalità dei decessi dei periodi 1671-1700 e 1851-1871

Nel corso degli anni l'effetto della stagionalità sui decessi si va affievolendo, anche se non si è totalmente esaurito. Questo indica che Battaglia Terme è un paese in movimento che non ha ancora esaurito il proprio ciclo evolutivo.

2.3 I MATRIMONI

Abbiamo suddiviso i 2783 matrimoni di cui disponiamo in periodi decennali, per studiare se esiste una relazione tra gli anni e il valore assoluto di questi.

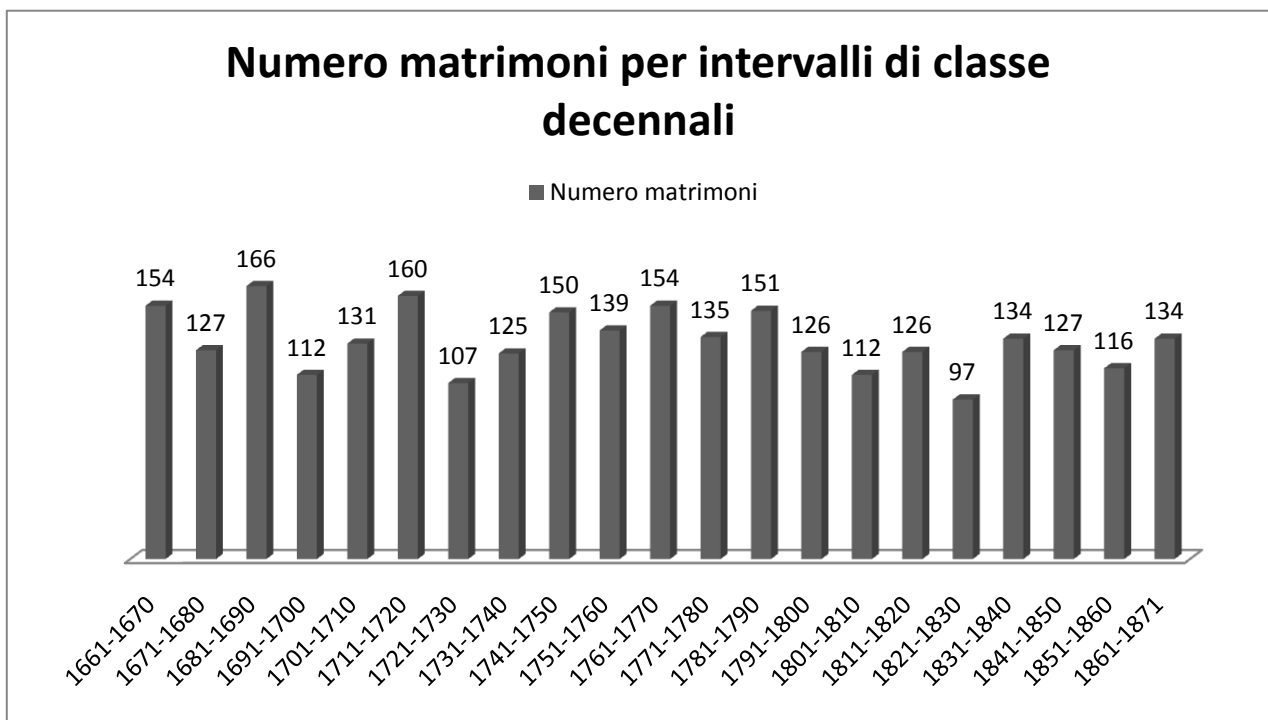


Grafico 2.3.1 Matrimoni suddivisi in periodi decennali

Apparentemente non si nota un trend particolare riguardante il comportamento matrimoniale. Invece si può intuire una certa corrispondenza tra questi ultimi e il numero di decessi: quando i primi aumentano, i secondi diminuiscono, e viceversa. Questo poiché nei periodi di maggior crisi (come nel caso di epidemie) la gente tende a sposarsi di meno.

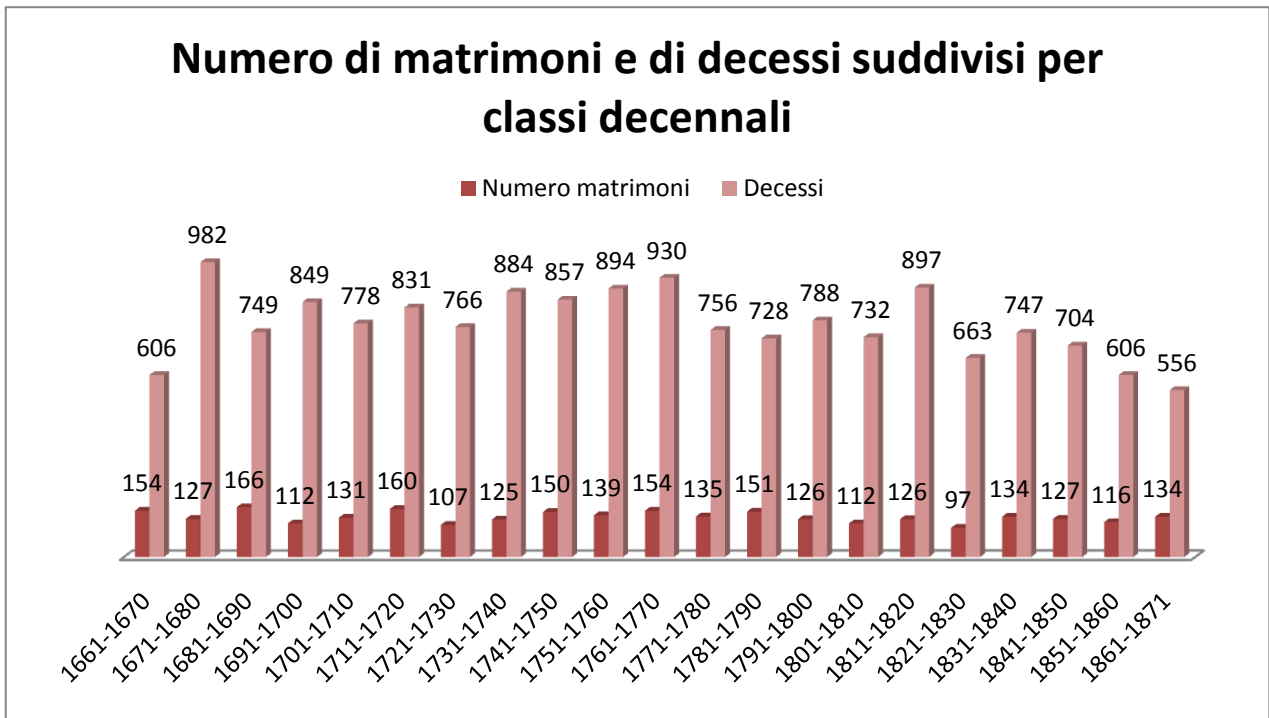


Grafico 2.3.2 Matrimoni e decessi suddivisi in periodi decennali

Per quanto riguarda invece la stagionalità dei matrimoni, nell'intero periodo preso in considerazione, il mese in cui ci si è sposati di più è quello di febbraio, con 715 matrimoni, seguito da Novembre con 464. Dicembre, Marzo e Maggio i mesi in cui ci si sposa di meno.

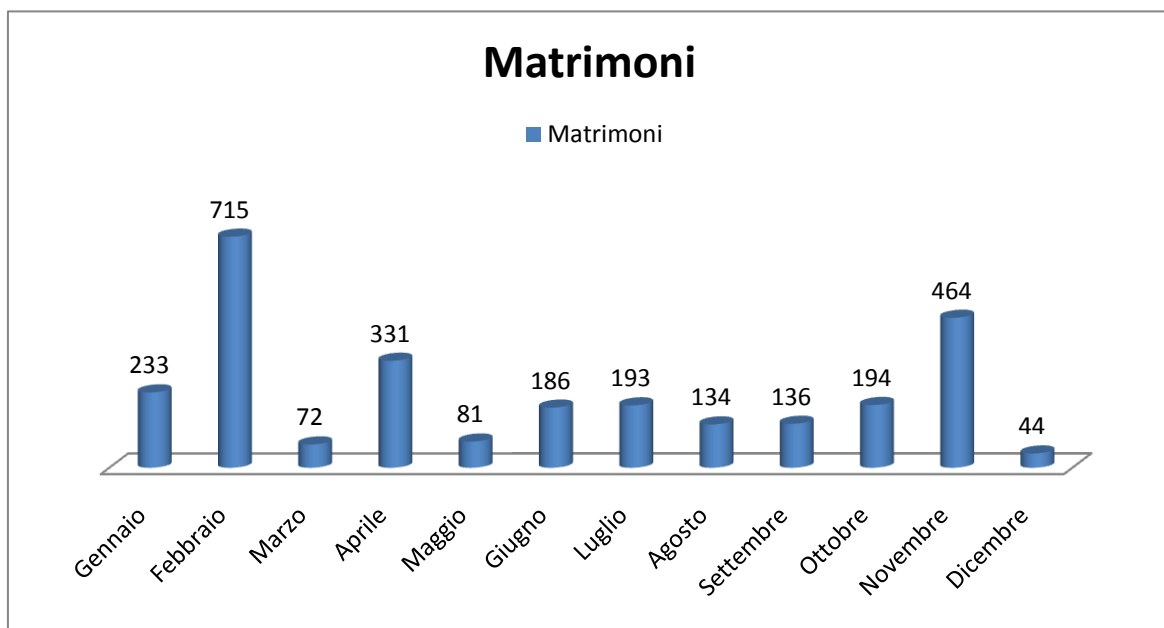


Grafico 2.3.3 Grafico dei matrimoni e dei decessi suddivisi in periodi decennali

Il motivo di queste depressioni è dovuto al rispetto del calendario liturgico degli abitanti di Battaglia Terme. Tra le regole in cui la Chiesa vietava i matrimoni vi erano anche indicati i periodi dell'anno: in periodo di Avvento (cioè in Dicembre), e durante la Quaresima (Marzo e inizio Aprile).

Il motivo dei picchi di Febbraio e di Novembre, ma anche di quelli più limitati di Aprile e Gennaio, oltre ad un recupero rispetto ai mesi vicini, proprio quelli in cui non ci si poteva sposare, sta anche nel fatto che Battaglia Terme, nonostante tutto, rimane comunque un paese agricolo legato in particolar modo al lavoro dei campi, quasi esclusivamente estivo.

	1661-1670	1751-1760	1801-1810	1861-1871
Gennaio	9,1	13,7	7,1	9
Febbraio	23,4	28,8	28,6	18,7
Marzo	0,6	0	4,5	2,2
Aprile	9,1	20,1	16,1	11,2
Maggio	2,6	1,4	3,6	5,2
Giugno	14,9	7,2	4,5	9
Luglio	9,7	2,9	7,1	2,2
Agosto	7,8	2,9	4,5	7,5
Settembre	3,2	4,3	1,8	3
Ottobre	8,4	7,2	9,8	4,5
Novembre	9,7	9,4	11,6	23,9
Dicembre	1,3	2,2	0,9	3,7

Tabella 2.3.1 Percentuali di matrimoni per quattro periodi decennali

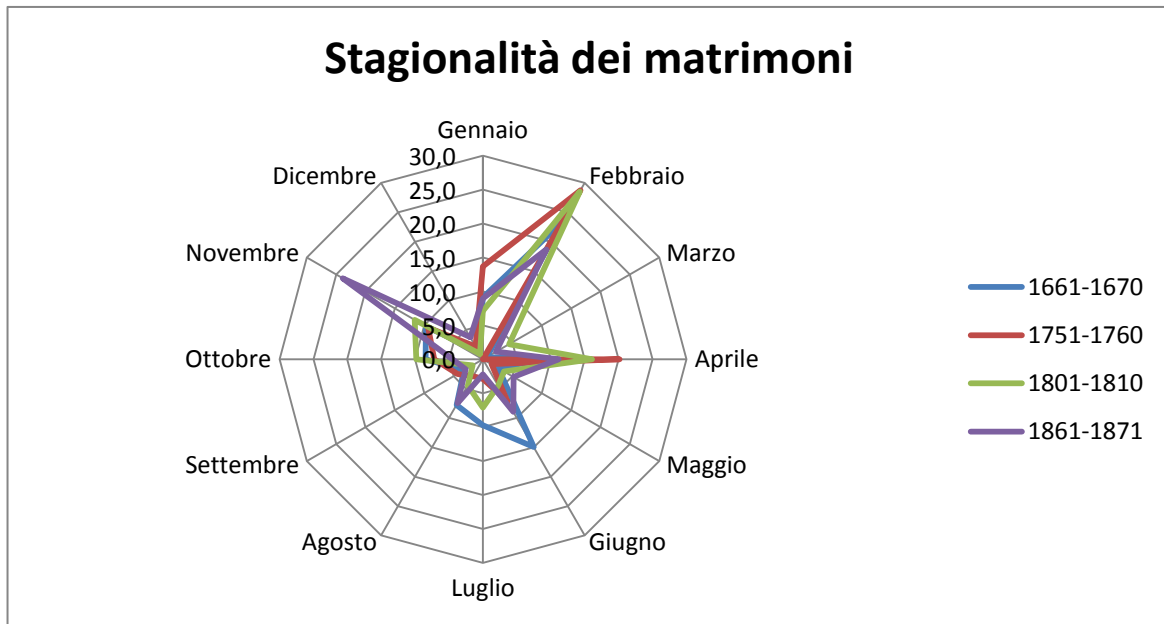


Grafico 2.3.4 Stagionalità dei matrimoni per quattro periodi decennali

Nonostante quello che ci si sarebbe potuti aspettare, la stagionalità dei matrimoni non cambia con il passare degli anni. I mesi di Febbraio, Novembre ed aprile rimangono sempre quelli più favorevoli, a discapito di Dicembre, Marzo, Maggio e Settembre.

Questo serve comunque a confermare quanto detto precedentemente: durante questo periodo Battaglia Terme è un paese che si sta industrializzando, ma questa transizione deve ancora completarsi.

CAPITOLO 3

LA RICOSTRUZIONE NOMINATIVA DELLE FAMIGLIE

3.1 LE TECNICHE NOMINATIVE

"Per tecniche nominative si intende un insieme di modalità di impiego delle fonti storico-demografiche che hanno in comune una rilevazione dei dati di tipo nominativo, e il successivo uso di tali caratteri in funzione della ricostruzione di vicende, situazioni o biografie individuali." (Del Panta, Rettaroli, p152) E' generalmente in una successiva fase dell'analisi che le informazioni relative a singoli individui o nuclei familiari vengono in qualche modo riaggregate, per arrivare quindi a misure demografiche analoghe a quelle che si ottengono con le tradizionali tecniche aggregative. Tutti i documenti che forniscono le generalità degli individui (come nel caso dei registri parrocchiali) si prestano ad essere analizzati in forma individuale.

La tecnica della ricostruzione nominativa delle famiglie ha la particolarità di poter operare esclusivamente su fonti relative al movimento naturale della popolazione consentendo comunque di pervenire a misure corrette della fecondità, della mortalità e della nuzialità.

3.2 LA RICOSTRUZIONE NOMINATIVA DELLE FAMIGLIE

La tecnica nominativa è stata messa a punto da Louis Henry, demografo francese, all'inizio degli anni Cinquanta. L'obiettivo principale di Henry era quello di arrivare a stimare una corretta misura della fecondità delle popolazioni del passato superando un doppio limite delle fonti storico-demografiche: la quasi totale assenza dei "stati delle anime" e la mancata presenza nei registri dei battesimi dell'età della madre alla nascita del figlio.

Lo sfruttamento nominativo congiunto dai registri del movimento della popolazione consente sia di ricavare in maniera precisa determinate informazioni non contenute nei registri (età della donna alla nascita del figlio, età della donna al matrimonio, ecc...), sia di

valutare, in qualsiasi intervallo di età della donna, il numero di anni che questa rimane effettivamente esposta al rischio di procreare e, quindi, di ricavare un comun denominatore per la valutazione dei tassi specifici di fecondità.

La tecnica di ricostruzione nominativa si può quindi ricondurre allo sviluppo di due fasi successive: a) La prima riguarda la ricostruzione della vicenda demografica di ciascun nucleo familiare a partire dalle informazioni contenute nei registri individuali (di nascita, morte e matrimonio); b) La seconda fase si riferisce, utilizzando le informazioni ricavate per ogni nucleo e raccolte in una particolare scheda denominata “scheda di famiglia”, alla riaggregazione dei dati per arrivare alle misure di fecondità, nuzialità e mortalità del tutto analoghe a quelle abitualmente ottenibili, con le usuali tecniche manualistiche, operando su dati aggregati per popolazioni con statistiche complete.

Dunque, la tecnica messa a punto da Henry prevede in primo luogo una rilevazione di tipo nominativo. Successive procedure di ordine alfabetico, e quindi di accoppiamento (“linkage”) tra informazioni relative ad uno stesso individuo consentiranno poi di trasferire alla “scheda di famiglia” sia le date di nascita, matrimonio, morte, sia altre informazioni relative ai membri di una stessa famiglia, e cioè i coniugi e i loro figli.

Per concludere si ricostruiscono biografie individuali all’interno di storie familiari. La storia familiare è delimitata da due date: quella di inizio-osservazione e quella di fine-osservazione. Se il matrimonio è stato celebrato nella zona, la data di inizio-osservazione coincide con quella di inizio dell’unione. La data di fine unione invece coincide con quella di decesso del coniuge che muore per primo. Per concludere, si prefissano quattro possibilità riguardo la presenza o all’assenza delle informazioni sulle date di inizio e di fine unione, per le quali è consuetudine usare la simbologia introdotta da Henry: 1) MF Famiglie per cui sono note le date d’inizio e di fine unione; 2) MO Famiglie in cui è ignota la data di fine unione; 3) EF Famiglie in cui è ignota la data di inizio unione; 4) EO Famiglie in cui non è nota nessuna delle due informazioni. Le famiglie MF si dividono a loro volta in 3 classi: 1) MF1 L’età della donna è nota con certezza 2) MF2 L’età della donna è nota in modo approssimato 3) L’età della donna è ignota.

Il problema cruciale della tecnica di ricostruzione nominativa delle famiglie rimane, comunque, il suo basso rendimento (inteso come rapporto tra il numero di famiglie

ricostruite e quello totale delle famiglie che si sono formate nel medesimo periodo di tempo nella zona considerata, cioè il totale dei matrimoni celebrati). I risultati delle ricostruzioni si riferiscono, di solito, in prevalenza, alle famiglie più stabili e geograficamente endogame. Sfuggono invece le famiglie più mobili che certamente si differenziano, abitualmente, dalle famiglie stabili per quanto attiene alle caratteristiche sociali e che forse, il più delle volte, hanno differenti “comportamenti” demografici (in termini di nuzialità e fecondità) nonché rischi di morte probabilmente superiori.

3.3 IL PROBLEMA DEI “LINKAGE”

Nella pratica concreta le operazioni di “linkage”, e cioè di accoppiamento di informazioni relative e a uno stesso individuo ma desunte da fonti diverse, non risultano sempre agevoli. Come noto, i registri parrocchiali sono spesso redatti con scarsa cura, la lettura delle generalità degli individui non è sempre facile, e spesso il parroco stesso riporta nomi e cognomi in maniera imprecisa, o comunque in forma diversa in differenti registrazioni. Col passare del tempo, inoltre, si notano spesso variazioni di tipo morfologico nei nomi e nei cognomi, mutamenti di lingua e modificazioni nelle identificazioni dei luoghi.

Le numerose informazioni che posso essere utilizzate per stabilire se un link debba essere accettato o rifiutato non sempre sono presenti nei registri. In piccole popolazioni isolate i casi di omonimia sono assai frequenti. E' inoltre comune l'uso di attribuire ai bambini lo stesso nome di loro fratelli premorti.

Prima di iniziare la vera fase di accoppiamento delle informazioni è opportuno procedere a operazioni di controllo dei dati e, in particolar modo, alla standardizzazione dei nomi e dei cognomi e più in generale di tutte quelle variabili che contengono informazioni di tipo alfanumerico relative a persone o luoghi.

Un problema generale è quello di determinare quali forme nominali rappresentano lo stesso nome. Vi sono due principali strategie per procedere alla standardizzazione dei nomi tramite computer. La più semplice, ma anche la meno generalizzabile, comporta l'utilizzo di un archivio di riferimento che contenga una lista alfabetica di tutti i nomi

presenti nelle fonti e nei periodi esaminati, sulla quale il ricercatore opera manualmente, riportando a fianco di ogni nome la forma standardizzata che ritiene più appropriata. Questa strategia ha il vantaggio di essere flessibile, poiché vi è una libertà completa da parte del ricercatore nella scelta delle forme standard, ma, al contrario, non risulta certamente la più veloce e quella maggiormente generalizzabile, questo perché difficilmente gli archivi che contengono gli standard sono utilizzabili per aree geografiche e il lavoro dovrebbe quindi essere in gran parte ripetuto. Il secondo metodo è quello di ricorrere ad algoritmi matematici che rendano possibile la traduzione in forme standard mediante regole che, esaminando lettera per lettera il nome, siano in grado di sostituire ad ognuna, o a loro combinazioni (sequenza di grafemi), altre lettere o altre sequenze prestabilite.

Per concludere, i programmi di linkage si basano in genere sul principio che due atti saranno attribuiti allo stesso individuo se risulta verificata tutta una serie di condizioni, relative soprattutto all'identità delle informazioni nominative oltre che a determinate compatibilità temporali.

3.4 LA SCHEDA DI FAMIGLIA

Una volta ottenute le informazioni dei singoli individui, non rimane altro che “ricostruire” la famiglia. Per questo motivo vengono riportate su di una speciale scheda, detta “scheda di famiglia”, sia le date di nascita, morte, matrimonio sia altre informazioni relative ai membri di una stessa famiglia coniugale, vale a dire i coniugi e i relativi figli. Poiché la ricostruzione delle famiglie si dimostra una tecnica tanto più efficiente quante più famiglie vengono ricostruite, conviene, quando è possibile, ricorrere anche a date di morte presunte, purché non vengano introdotti fattori selettivi nell'analisi.

Per effettuare analisi approfondite si studiano o le famiglie di tipo MF, ovvero quelle, come detto precedentemente, quelle in cui è nota l'età della donna al matrimonio. In questa maniera, oltre alle informazioni reperibili in qualunque scheda di famiglia, quali le

informazioni base sui coniugi, anche informazioni sull'età della donna alla nascita di un figlio, indicatore utile per calcolare indici di fecondità per la coppia.

Il rapporto tra il numero di famiglie ricostruite (solamente quelle del tipo MF) e quello totale delle famiglie che si sono formate nel medesimo periodo di tempo nella zona considerata, fornisce il tasso di rendimento della ricostruzione il quale consente di apprezzare l'eshaustività della ricostruzione stessa in termini di rappresentatività delle famiglie ricostruite rispetto a quelle ricostruibili in quel segmento di realtà storico-demografica che si sta considerando. Non va comunque dimenticato che, per quanto elevato possa risultare, il tasso di rendimento della ricostruzione non garantisce di per sé né la rappresentatività della collettività studiata rispetto a collettività più vaste di appartenenza, né la significatività di risultati raggiunti, essendo quest'ultima legata anche alla sufficiente numerosità dei casi osservati.

3.5 MISURE DEMOGRAFICHE RICAVABILI DALLA SCHEDA DI FAMIGLIA

Come già detto, una volta che si è effettuata la ricostruzione e che quindi si disponga di una scheda di famiglia, occorrerà impostare un processo di riaggregazione dei dati individuali con lo scopo di arrivare a misure di fecondità, nuzialità e mortalità riferite, non a singoli individui o nuclei familiari, ma all'aggregato demografico nel suo complesso. Per evitare distorsioni legate ad eventuali comportamenti demografici differenziali tra famiglie "stabili" e famiglie "mobili", è opportuno operare su famiglie "chiuse", ovvero quelle famiglie identificate con la sigla MF.

3.5.1 LA MORTALITA' INFANTILE

In questo paragrafo si vuole fornire la procedura per ricavare una tavola di mortalità fino al quinto compleanno.

La tecnica proposta da Henry segue la medesima logica delle procedure utilizzate ancora oggi nel calcolo di tavole di eliminazione ricavate da storie di vita ricostruite, nel caso fra di esse vi siano incluse serie tronche, cioè composte da quei soggetti la cui osservazione non si è ancora conclusa.

Innanzitutto si calcolano le durate di osservazione dei bambini dalla nascita fino all'uscita di osservazione della famiglia. Si individuano successivamente le durate dei bambini deceduti prima del compimento del quinto compleanno, o comunque prima dell'uscita di osservazione della famiglia. A questo punto si predispone un prospetto di spoglio come quello in figura 3.5.1.1

Età	Decessi	Usciti di osservazione
Primo mese		
1-11 mesi		
1 anno		
2-4 anni		
5 anni e più		

Figura 3.5.1.1 Prospetto di spoglio

Completato il prospetto di spoglio si passa ad uno riepilogativo, nel quale le prime due colonne riporteranno i totali delle due colonne della scheda iniziale.

Età	Decessi	Usciti di osservazione	Decessi cumulati	Usciti di osservazione cumulati	Decessi e usciti di osservazione cumulati	1000 q(x)
Primo mese						
1-11mesi						
1 anno						
2-4 anni						
5 anni e più						
Totale						

Figura 3.5.1.2 Prospetto riepilogativo

A questo punto, partendo dalle probabilità di morte dell'ultima colonna, ottenute dal rapporto tra i decessi e la somma dei decessi e gli uscite di osservazione cumulati, moltiplicati per 1000, si ricava la tavola di eliminazione:

Età	l_x	d_x	1000 q_x	1000 p_x
0	1000			
1 mese				
1 anno				
2 anni				
5 anni				

Figura 3.5.1.2 Tavola di eliminazione fino al quinto compleanno

Ponendo la radice della tavola pari a 1000, si utilizzano le note formule:

- $D_x = l_x q_x$
- $L_{x+1} = l_x - d_x$

La probabilità di morire tra la nascita e il primo compleanno risulterà quindi:

$$(l_0 - l_1) / l_0 = (1.000 - l_1) / 1000$$

La probabilità, invece, di morire tra la nascita e il quinto compleanno sarà invece:

$$(l_0 - l_5) / l_0 = (1000 - l_5) / 1000$$

3.5.2 MISURE DI FECONDITA'

Come già detto, volendo calcolare tassi specifici di fecondità partendo da registri parrocchiali, bisogna tener conto che nei libri di battesimo non è presente l'età della madre alla nascita del figlio. Occorre inoltre trovare il modo di stimare la popolazione femminile nelle diverse classi di età. I dati contenuti nelle schede di famiglia permettono di conoscere esattamente l'età della madre alla nascita del figlio e di valutare, in ogni classe di età, usualmente partendo dai 15-19 anni, il numero di anni che la donna trascorre esposta al rischio di rimanere incinta. In altre parole si dice che si misura il numero di anni-donna, cioè il numero di anni per i quali ogni donna contribuisce alla formazione del denominatore del tasso specifico di fecondità. Il numeratore sarà quindi fornito dalla

somma di tutti gli eventi nascita conteggiati, nelle varie schede di famiglia, come derivanti da donne coniugate per le diverse classi di età alla nascita dei figli.

Anche in questo caso si partirà dalle schede di famiglia ricostruite, in modo da poter compilare un prospetto di spoglio analogo a quello in figura 3.5.2.1

Totale nati	15-19	20-24	25-29	30-34	35-39	40-44	Nati f.o

Figura 3.5.2.1 Prospetto di spoglio per il calcolo di tassi specifici di fecondità legittima

Ogni riga rappresenta, in sostanza, i dati relativi ad una storia familiare. Si tratta di conteggiare, per ogni donna in ciascuna classe di età, il numero di figli avuti e il numero di donne-anno. Per nati f.o. si intendono i bambini nati fuori osservazione, cioè i bambini nati nell'ultima classe di età non vissuta per intero).

A questo punto sorge l'esigenza di arrivare ad indicatori sintetici, che compendino in un'unica cifra il livello di fecondità coniugale delle coppie o delle donne studiate.

Si possono innanzi tutto calcolare i tassi di fecondità totale legittima per donne sposate ad una certa età: questi tassi vengono solitamente indicati con $TFTL_{20}$, $TFTL_{25}$ ecc.. Per esempio, l'indice $TFTL_{20}$ è ottenuto sommando i tassi specifici di fecondità per le classi da 20-24 in avanti, e moltiplicando il risultato per 5. Il significato è quello di misura "lorda" di fecondità, cioè in assenza di mortalità. In sostanza, $TFTL_{20}$ rappresenta il numero medio di figli messo al modo da una generazione di donne sposate a 20 anni esatti e il cui matrimonio non si interrompe fino ai 45 o 50 anni. Questo indicatore ha però il vantaggio di permettere un confronto corretto della fecondità legittima tra popolazioni diverse o tra periodi diversi per una stessa popolazione.

Un altro indicatore sintetico comunemente calcolato sulla base dei risultati delle ricostruzioni nominative è il numero medio di figli. Tale valore si ricava direttamente dal prospetto di spoglio, rapportando i nati al numero di righe del prospetto. Si può effettuare lo stesso tipo di rapporto, ma operando solamente nell'ambito delle famiglie a fecondità completa, per evitare il riconteggio delle donne risposate che andrebbero a falsare il valore precedente. In questo caso ci si è ricondotti ad una misura lorda, cioè ad una misura non

influenzata dalla rottura delle unioni coniugali prima del termine dell'età feconda. Tale indice sarebbe comunque influenzato dall'età media al matrimonio.

3.6 LA RICOSTRUZIONE NOMINATIVA NEL CASO DI BATTAGLIA TERME

Una volta spiegata la teoria sul come operare per effettuare un ricostruzione nominativa e le relative analisi, si passa ad una esemplificazione del tutto.

3.6.1 LA RICOSTRUZIONE E I PROBLEMI DI “LINKAGE”

I dati di partenza rimangono i file trascritti in Excel dal dott. Luciano Donato e derivanti dai registri dei battesimi, delle sepolture e dei matrimoni custoditi nella Parrocchia di San Giacomo a Battaglia Terme.

Il registro di partenza è stato quello dei matrimoni, dal quale, grazie al nome e al cognome degli sposi, si è potuto ricostruire la vita familiare di numerose famiglie. Le variabili chiave utilizzate per legare la data di matrimonio dei coniugi a quelle di nascita e di morte sono state: nome, cognome, nome del padre, il nome e il cognome della madre (qualora fossero presenti) e una plausibile concordanza temporale. Sia dal file dei battesimi che da quello delle sepolture, conoscendo il nome del padre e della madre, si potevano conoscere le date di nascita e di morte dei relativi figli. Una volta selezionato un componente della famiglia, questo viene evidenziato di azzurro nel caso del padre, di rosa nel caso della madre e di giallo nel caso di un figlio. Questo procedimento è indispensabile per poter ricondurre, qualora nell'analisi fosse comparsa un'altra volta la medesima persona, un figlio alla famiglia di origine o un genitore al precedente matrimonio. I record non colorati indicano, infine, una non corrispondenza tra i dati di matrimonio e quelli di nascita e di morte.

Ricostruzione nominativa delle famiglie in una comunità veneta: Battaglia Terme (1607 – 1871)

	battesimo	Data nasc	sex	Cognome	Nome	Padre	Cognome	Madre
12694	23/11/1791	22/11/1791	M	Lotti	Adamo	Antonio	Trimignon	Giustina
10629	26/04/1761	23/04/1761	M	Lotti	Amadio	Bortolo	Trvisana	Cattarina
10357	12/08/1757	11/08/1757	M	Lotti	Andrea	Antonio	Caonegro	Cattarina
7808	20/09/1722	16/09/1722	F	Lotti	Angela	Domenico		Vittoria
8058	09/12/1725	06/12/1725	F	Lotti	Angela	Zuane		Anzola
8257	29/05/1728	28/05/1728	F	Lotti	Angela	Domenico		Vittoria
9743	26/07/1749	25/07/1749	F	Lotti	Angela	Giovani	Manzeta	Laura
10963	07/09/1765	06/09/1765	F	Lotti	Angela	Bortolo	Trvisan	Catarina
12428	31/03/1787	31/03/1787	F	Lotti	Angela	Domenico	Pierato	Rosa
8000	29/01/1725	26/01/1725	M	Lotti	Angelo	Santo		Domenica
9914	20/01/1752	19/01/1752	M	Lotti	Angelo	Antonio	Caonegro	Cattarina
10043	17/12/1753	16/12/1753	M	Lotti	Angelo	Antonio	Caonegra	Cattarina
10870	18/04/1764	18/04/1764	M	Lotti	Angelo	Giuseppe	Cipriana	Anna
8285	13/10/1728	30/09/1728	F	Lotti	Anna	Zuane		Anzola
10697	21/04/1762	21/04/1762	F	Lotti	Anna Maria	Bortolo	Trevisana	Cattarina

Figura 3.6.1.1 Porzione di fogli di lavoro delle nascite dopo la ricostruzione delle famiglie

COGNOME	NOME	SOPRANNOME	PADRE	MADRE	NOME MADRE	MARITO	DATANASCI	DATAMORI
BODON	DANIELE		ANTONIO					17/09/1812
BODON	DOMENICO		ANTONIO					31/01/1665
BODON	DOMENICO		ANGELO					07/12/1765
BODON	DOMENICO		GIUSEPPE	CIPRIAN	ANNA			08/01/1766
BODON	DOMENICO		GIUSEPPE	TAMIO	ANTONIA			05/02/1778
BODON	DOMENICO		FRANCO					28/01/1805
BODON	DOMENICO		ANTONIO					16/10/1805
BODON	DOMENICO		GIUSEPPE	SANAVIO	ANTONIA	LUNARDI		24/08/1820
Bodon	LUCIANO		GIUSEPPE	CAPPELLARI	ANTONIA			18/01/1854
BODON	ELISABETTA		GIUSEPPE	PAMIO	ANTONIA			28/12/1776
BODON	ELISABETTA		ISEPPO	PAMIO	ANTONIA			08/02/1782
BODON	ELISABETTA		GIUSEPPE	PAMIO'	ANTONIA	MENGOTTO		20/07/1854
BODON	CATTARINA		BORTOLO	ZARAMELLA	ANTONIA			28/06/1828

Figura 3.6.1.2 Porzione di foglio di lavoro dei decessi dopo la ricostruzione delle famiglie

Nella figura 3.6.1.1 viene presentata una porzione del foglio di lavoro delle nascite dopo aver concluso l'operazione di ricostruzione delle famiglie. Per esempio, Lotti Angelo, evidenziato di azzurro, sta ad indicare che si tratta di un marito, Lotti Angela e Lotti Anna Maria, sono, invece, due mogli. I record evidenziati di giallo stanno ad indicare la presenza di figli, mentre quelli in bianco indicano che non vi è stata trovata nessuna concordanza possibile con i matrimoni.

Uguualmente, nella figura 3.6.1.2, dove abbiamo invece una porzione di foglio di lavoro dei decessi, il record blu indica il decesso di un marito, il record rosa il decesso di una moglie, mentre quello arancione il decesso di un figlio.

3262	30/04/1823	MENGOTTO	LUIGI	BATTAGLIA	11/06/1796	VINCENZO	MASI	MARIA	BODON
12978	12/06/1796	11/06/1796	M	Mengotto	Luigi	Vincenzo	Masi	Maria	Carturan Gir Maccoppe
15787	MENGOTTO	LUIGI	TOCCIA	VINCENZO	MASI	MARIA		05/05/1863	MASCHILE
12082	08/11/1781	07/11/1781	F	Lotti	Elisabetta	Iseppo	Pamio	Antonia	Vendrain Fr Laura M Ca
15287	BODON	ELISABETTA		GIUSEPPE	PAMIO'	ANTONIA	MENGOTTO	20/07/1854	FEMMINILE
15377	13/03/1829	13/03/1829	M	Mengotto	Domenica	Luigi	Bodon	Elisabetta	
15055	16/04/1824	16/04/1824	M	Mengotto	Gio Batta	Luigi	Bodon	Elisabetta	
15133	27/04/1825	26/04/1825	M	Mengotto	Giuseppe	Luigi	Bodon	Elisabetta	
15271	19/05/1827	19/05/1827	F	Mengotto	Maria	Luigi	Bodon	Elisabetta	
13498	MENGOTTO	MARIA ANNA		LUIGI	BODON	ELISABETTA		25/03/1829	FEMMINILE
13197	MENGOTTO	FORTUNATO		LUIGI	BODON	ELISABETTA		19/04/1824	MASCHILE
13327	MENGOTTO	DOMENICO		LUIGI	BODON	ELISABETTA		20/07/1826	MASCHILE

Figura 3.6.1.3a Ricostruzione finale di una famiglia

BODON	ELISABETTA	BATTAGLIA	15/07/1798	GIUSEPPE	PAMIO	0	ANTONIA	BATTAGLIA	CELIBE/NUF	BATTAGLIA	CELIBE/NUBILE
Maccoppe Mmaddalena											
	FALSO	#####									
Laura M Cartuan Bortol											
	FALSO	#####									
	FALSO	#####									
	FALSO	#####									
	FALSO	#####									

Figura 3.6.1.3b Ricostruzione finale di una famiglia

Alla fine di questa operazione si ottiene una scheda colorata, come in figura 3.6.1.3a e 3.6.1.3b, dove il rosso indica i dati del matrimonio dal quale si è partiti, l'azzurro sono le informazioni relative alla nascita e al decesso del marito, il rosa sono le medesime informazioni per la moglie, in giallo i dati sulla nascita dei figli e in arancione le informazioni sulla morte di quest'ultimi.

Il problema principale avvenuto in questa parte di analisi è stato quello di collegare cognomi diversi ma equivalenti. Partendo da una lista di cognomi fornitaci dal dott. Donato, si è arrivati ad una stesura completa di un elenco di corrispondenze. I cognomi riportati non sono quelli che cambiano leggermente nel corso del tempo (per esempio da Saramella varia a Zaramella, e viceversa), ma quelli che cambiano radicalmente la propria radice e che, nel passato, erano soliti usare indifferentemente.

Cognome	Riferimento	Cognome	Riferimento	Cognome	Riferimento
BALDO	- BUSON	CANELLA	- NOGARA	MENESELLO	- MENORELLO
BALLARIN	- DE SANTI	CAPARO	- TREMIGNON	MENON	- BOTTARO
BARATIN	- TROGNACARA	CASTRELLI	- BERTOLA	MENORELLO	- MENOSELLO
BASARO	- ZAONELLO	CENO	- FERRARI	MILAN	- RAVANELLO
BEDOCCO	- SCALZARETTO	CESTARO	- TOSATO	MORARI	- MARAN
BERALDO	- FINCHO	CHINCHIO	- BOARO	NIETO	- BORTOLATO
BERTOLA	- CASTRELLI	CORTI	- BOSCHEO	NOGARA	- CANELLA
BIANCHI	- BORDIN	DE BIANCHI	- BORDINO	NORO	- BISACCO
BIANCO	- TURCATO	DE SANTI	- BALLARIN	OSSI	- MENEGOTTI
BIOTTO	- FASOLATO	DEGLI OSTI	- MANZETTO	PASORELLO	- CAMBURAN
BISACCO	- NORO	DONI	- FURLAN	PASTORELLO	- BONAFE'
BISO	- FERRATO	FASOLATO	- BIOTTO	PICCOLO	- ROSSIGNOLO
BOARO	- CHINCHIO	FERRARI	- CENO	PISTORE	- MAISTRO
BOCCARDO	- SCHIOPPETTO	FERRATO	- BISO	RAVANELLO	- MILAN
BODON	- LOTTI	FINCHO	- BERALDO	ROSSIGNOLO	- PICCOLO
BONAFE'	- PASTORELLO	FURLAN	- BONI	SARTORI	- GHIRRO
BORDIN	- BIANCHI	GASTALDELLO	- TOSO	SAVELLON	- BOZZA
BORDINO	- DE BIANCHI	GHIRRO	- SARTORI	SCALZARETTO	- BEDOCCO
BORIN	- GRANCIERO	GRANCIERO	- BORIN	SCHIOPPETTO	- BOCCARDO
BORTOLATO	- NIETO	LOTTI	- BODON	TOSATO	- CESTARO
BOSCHEO	- CONTI	LUNARDI	- ZANON	TOSO	- GASTALDELLO
BOTTARO	- MENON	MAISTRO	- PISTORE	TREMIGNON	- CAPARO
BOZON	- MENEGHETTI	MANZETTO	- DEGLI OSTI	TROGNACARA	- BARATIN
BOZZA	- SAVELLON	MARAN	- MORARI	TURCATO	- BIANCO
BUSON	- BALDO	MENEGHETTI	- BOZON	ZANON	- LUNARDI
CAMBURAN	- PASTORELLO	MENEGOTTI	- OSSI	ZAONELLO	- BASARO

Figura 3.6.1.4 Tabella di corrispondenza dei cognomi

Operativamente questo implica che ogni volta che si cerca una persona con uno dei cognomi presenti nella lista, si dovevano cercare informazioni anche su quello corrispondente. Un esempio pratico è evidente dalla figura 3.6.1.3a, dove, sulle informazioni riguardanti la nascita Lotti Elisabetta, mentre, tra le informazioni del decesso, il nome diventa Bodon Elisabetta. Questo proprio perché, anche guardando la lista, il cognomi Bodon e Lotti erano utilizzati indifferentemente.

3.6.2 LA SCHEDA DI FAMIGLIA

Ricostruite, quindi, manualmente tutte le famiglie, si passa ad una redazione di una carta di famiglia. Per costruire ciò è stato utilizzato il programma Microsoft Access.

Tutte le informazioni interessanti riguardanti la famiglia sono state introdotte in una tabella del programma, altre informazioni non presenti sono state invece desunte dalle date che si avevano in possesso. A questo punto, per poter visualizzare la scheda di famiglia nell'insieme, sono state create tre sottomaschere, una riguardante il padre, una riguardante la madre e una riguardante i figli, le quali sono state introdotte tutte in una maschera principale. Il risultato è quanto si può vedere nelle figure 3.6.2.1a e 3.6.2.1b.

Codice del matrimonio: <input type="text" value="2277"/>	Record di riferimento: <input type="text"/>	
Nome: <input type="text" value="ANTONIO"/> Cognome: <input type="text" value="CALAPPO"/> Sesso: <input type="text" value="M"/>	Nome: <input type="text" value="TERESA"/> Cognome: <input type="text" value="TIOZZI"/> Sesso: <input type="text" value="F"/>	
Data di nascita: <input type="text" value="08/05/1798"/> Data di morte: <input type="text"/>	Data di nascita: <input type="text" value="14/11/1800"/> Data di morte: <input type="text"/>	
Luogo di nascita: <input type="text" value="BATTAGLIA"/> Luogo di provenienza: <input type="text" value="BATTAGLIA"/>	Luogo di nascita: <input type="text" value="CHIOGGIA"/> Luogo di provenienza: <input type="text" value="BATTAGLIA"/>	
Figliola di: <input type="text" value="FRANCESCO E CAM"/> Vedovola di: <input type="text"/>	Figliola di: <input type="text" value="ANTONIO E PESCAI"/> Vedovola di: <input type="text"/>	
Rango del matrimonio: <input type="text" value="1"/>	Data matrimonio: <input type="text" value="04/12/1823"/>	Rango del matrimonio: <input type="text" value="1"/>
Età al matrimonio: <input type="text" value="25,554035567715"/>	Data fine evento: <input type="text"/>	Età al matrimonio: <input type="text" value="23,036935704514"/>
Età fine evento: <input type="text"/>	Durata matrimonio: <input type="text"/>	Età fine evento: <input type="text"/>
Età morte: <input type="text"/>		Età morte: <input type="text"/>
NuoveNozze: <input type="text"/>		NuoveNozze: <input type="text"/>
Con: <input type="text"/>		Con: <input type="text"/>
Durata vedovanza: <input type="text"/>		Durata vedovanza: <input type="text"/>

Figura 3.6.2.1a Porzione di Scheda di famiglia riguardante i genitori

Nome	Cognome	Sesso	Data di nascita	Data di morte	Giorni	Anni
CHIARA	CALAPPO	F		05/04/1824		

Età madre alla nascita:	Anni di matrimonio alla nascita:	Data di matrimonio:	Età al matrimonio:	Nome Coniuge:	Gemello:
					<input type="checkbox"/>

Figura 3.6.2.1 b Porzione di Scheda di famiglia riguardante i figli

Le informazioni ricavate nella porzione di scheda riguardante i genitori sono le età al matrimonio, le età alla fine dell'evento, le età di morte, l'eventuale durata di vedovanza e la durata del matrimonio. Per quanto riguarda i figli, le informazioni ricavate sono i giorni vissuti, gli anni vissuti, l'età della madre alla nascita del figlio e gli anni di matrimonio alla nascita del figlio.

3.6.3 LA MORTALITA' INFANTILE

Tutte queste informazioni sono state esportate in un foglio Excel tramite il quale sono state possibili tutte le analisi per calcolare le misure demografiche descritte precedentemente.

Poiché nella nostra analisi siamo partiti dalle date di matrimonio, tutti le nostre rilevazioni saranno di tipo MF o MO, anzi, per la precisione, abbiamo 719 rilevazioni di tipo MF e 2049 di tipo MO.

Inoltre, dei 5160 figli rilevati (da tutte le famiglie, sia quelle di tipo MF che quelle di tipo MO), di 3137 non sappiamo le età di morte.

Proviamo allora a costruire delle tavole di mortalità, prima di tutte le famiglie e poi solo di quelle del tipo MF, per vedere se esistono sostanziali differenze. Eliminiamo dall'analisi le famiglie che non hanno avuto nemmeno un figlio.

Analizzando il complesso delle famiglie, otteniamo 542 record. I bambini utilizzabili per il calcolo della mortalità infantile sono 2019. Un esempio di prospetto di spoglio utilizzato per Battaglia Terme è quello illustrato in figura 3.6.3.1

Età	Decessi	Usciti di osservazione
Prima settimana	2	
Primo Mese		
Primo Anno		
1-4 Anni	1	
Più di 5 anni		1

Figura 3.6.3.1 Prospetto di spoglio di Battaglia Terme

Il passo successivo riguarda la costruzione del prospetto riepilogativo, come in figura 3.6.3.2

Età	Decessi	Usciti di oss.	Decessi cumulati	Usciti oss. Cumulati	Decessi e usciti oss. Cum.	1000 q(x)
Prima settimana	191		660	292	952	201
Primo mese	131		469	292	761	172
Primo anno	147		338	292	630	233
2-5 anni	191		191	292	483	395
Più di 5 anni		292				

Figura 3.6.3.2 Prospetto riepilogativo di Battaglia Terme

Con questo metodo, partendo dalle probabilità di morte, si ottiene la seguente tavola di eliminazione:

Età	l_x	d_x	1000 q_x	1000 p_x
Prima settimana	1000	201	201	799
Primo mese	799	138	172	828
Primo anno	662	154	233	767
2-5 anni	507	201	395	605
Più di 5 anni	307			

Figura 3.6.3.3 Tavola di eliminazione fino al quinto anno di Battaglia Terme

La probabilità di morire entro la prima settimana di vita si può calcolare facendo:

$$(1000 - 799) / 1000 = 0,201$$

La probabilità di morire invece tra la nascita e il quinto compleanno sarà invece:

$$(1000 - 307) / 1000 = 0,693$$

Analizzando solamente le famiglie di tipo MF con figli, otteniamo 303 record, per un totale di 1414 figli, dei quali però solo 571 sono quelli di cui possediamo l'età di morte. Il prospetto di spoglio è uguale a quello visto precedentemente, mentre la tavola di eliminazione assume un aspetto differente.

Età	Decessi	Usciti di oss.	Decessi cumulati	Usciti oss. Cumulati	Decessi e usciti oss. Cum.	1000 $q(x)$
Prima settimana	117		390	181	571	205
Primo mese	81		273	181	454	178
Primo anno	84		192	181	373	225
2-5 anni	108		108	181	289	374
Più di 5 anni		181				

Figura 3.6.3.4 Prospetto riepilogativo di Battaglia Terme

Ne consegue che la tavola di eliminazione assume il seguente aspetto:

Età	l_x	d_x	$1000 q_x$	$1000 p_x$
Prima settimana	1000	205	205	795
Primo mese	795	142	178	822
Primo anno	653	147	225	775
2-5 anni	506	189	374	626
Più di 5 anni	317			

Figura 3.6.3.5 Tavola di eliminazione fino al quinto anno di Battaglia Terme

Da questa tavola si evince che la probabilità di morire entro la prima settimana di vita è uguale a :

$$(1000 - 795) / 1000 = 0,205;$$

La probabilità di morire entro il quinto anno di vita risulta, invece:

$$(1000 - 317) / 1000 = 0,683.$$

Entrambe i valori non si discostano a seconda che si analizzino le famiglie MF oppure quelle MF ed MO congiunte. Questo è dovuto probabilmente al fatto che il fatto di non aver trovato le date di fine unione non indichi che queste famiglie si siano trasferite altrove, ma che risulti solamente una distorsione dei dati non selettiva.

3.6.4 MISURE DI FECONDITA' DI BATTAGLIA TERME

Nel caso di Battaglia Terme costruirsi una misura come il TFT non risulta semplice. A causa dell'elevato numero di dati mancanti, i risultati non sarebbero affidabili. Per calcolarsi quindi una misura di fecondità più precisa si decide di utilizzare solamente donne che hanno avuto almeno un figlio. Inoltre si selezionano quei figli di cui si conosce la data di nascita e quelle madri che hanno sia la data di nascita che la data di morte. Con tali informazioni si può infatti calcolare gli anni al parto della madre e il numero di anni vissuti dalla madre dal matrimonio alla fine dell'evento (fine matrimonio o morte).

Il periodo di fecondità della donna, 15-49 anni, è stato suddiviso sette classi quinquennali, all'interno delle quali è stato contato il numero di figli avuti e il numero di anni vissuti da

ogni donna (i cosiddetti anni-donna). Dividendo la somma dei figli per ogni classe per il numero di anni-donna vissuti (sempre per ciascuna classe) e moltiplicando il tutto per cinque si ottiene il tasso di fecondità per ogni intervallo. La somma di tutti questi tassi fornisce il tasso di fecondità totale della popolazione di Battaglia Terme dal 1607 al 1871.

Età donna	Tasso di fecondità
15-19	1,06
20-24	1,74
25-29	1,65
30-34	1,67
35-39	1,09
40-44	0,51
45-49	0,1
Totale	7,83

Figura 3.6.4.1 Tasso di fecondità per classi d'età

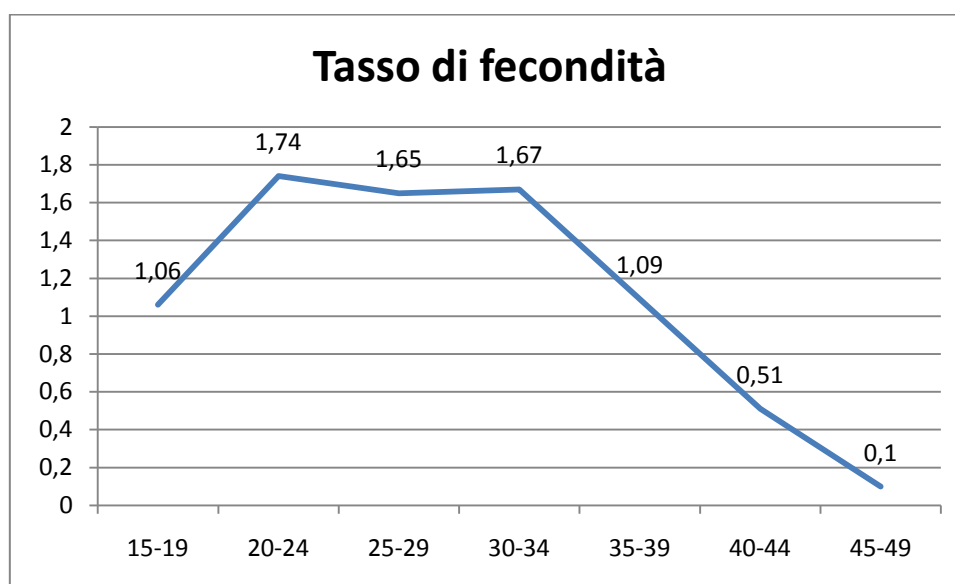


Grafico3.6.4.1 Tasso di fecondità per classi d'età

Il TFT (Tasso di Fecondità Totale) di Battaglia Terme dell'intero periodo è di 7,83, ciò vuol dire che ogni donna ha partorito in media quasi otto figli. L'età in cui le donne hanno avuto in media più figli è quella dai 20 ai 24 anni. Molto elevato è il valore anche per le classi 25-29 e 30-34, rispettivamente con 1,65 e 1,67 figli per donna. Le classi con TFT minore sono

le ultime, in quanto si può pensare che una donna abbia già raggiunto il numero di figli desiderato e che quindi non sia propensa ad averne ancora.

CAPITOLO 4

UN ESEMPIO DI ANALISI DEI DATI

4.1 CHE COSA E' L'EVENT HISTORY ANALYSIS

Per Event History Analysis si intende lo studio della durata di non sperimentazione di uno o più eventi durante il periodo a rischio (Yamaguchi, 1991), intendendo come periodo a rischio il periodo in cui è possibile sperimentare un determinato evento.

Si tratta di metodi adatti a studiare una grande varietà di fenomeni (sociali, naturali, biometrici, ecc...) e che assumono nomi differenti a seconda dell'ambito disciplinare di applicazione: EHA in sociologia, analisi di dati di durata in economia, affidabilità in ingegneria, sopravvivenza in statistica o medicina.

I dati di sopravvivenza hanno caratteristiche particolari e non sono sempre trattabili con le tecniche standard. Queste peculiarità sono:

- La variabile tempo definita positivamente
- Variabili dipendenti fisse nel tempo o tempo-dipendenti
- Presenza di censure nei dati (per questo l'inferenza deve essere basata sulla verosimiglianza e non sul metodo dei minimi quadrati)
- Centralità della funzione di rischio
- Tempo reso discreto o continuo

4.2 DURATE CENSURATE

Prima di introdurre i modelli a tempo discreto o a tempo continuo occorre effettuare un breve digressione sulle durate censurate. Non sempre l'osservazione permette di poter disporre di informazioni complete circa il tempo di permanenza in uno stato da parte di un

individuo e conoscere l'esatta durata del periodo di esposizione al rischio di subire l'evento oggetto di studio. In questo caso si parla di dati censurati.

4.2.1 CENSURA A DESTRA E CENSURA A SINISTRA

Un'osservazione è censurata a destra se tutto quello che si sa della vera durata di T è che questa sia maggiore di una certa durata C . Simmetricamente, una osservazione è invece censurata a sinistra se tutto ciò che si sa di T è che la sua vera durata è inferiore del medesimo valore C . Nelle scienze sociali il concetto di censura a destra e di censura a sinistra assumono però significati leggermente differenti e ricadono tutti nel caso sopracitato di censura a destra poiché entrambe le definizioni prevedono che il vero tempo T sia superiore al valore di C .

Nelle scienze sociali la censura a sinistra si verifica quando gli eventi che definiscono le durate degli episodi avvengono prima dell'inizio di osservazione. Al contrario, la censura a destra si ha quando l'osservazione finisce prima che sia completamente vissuta l'esperienza in un certo stato da parte dell'individuo.

La censura a destra è quella più comune negli studi sociali e naturali. Essa nasce da due ragioni: alcuni individui non sperimentano mai l'oggetto di studio; alcuni individui lo sperimentano alla fine del periodo di rilevazione. Dal punto di vista pratico tale distinzione viene trattata univocamente poiché in entrambi i casi non possiamo sapere se ci sarà, nel caso, quando avverrà l'evento. Quindi, se la censura a sinistra non permette di conoscere quando è avvenuto l'evento origine, quella a destra non permette di conoscere quando avverrà l'evento oggetto di studio. La censura può essere essa stessa considerata come un vero e proprio vento che interrompe l'osservazione.

A seconda del disegno di indagine è possibile distinguere diversi tipi di censura (a destra): censura di I tipo, censura di II tipo, censura random. La censura di I tipo si ha quando il tempo di censura è controllato dal ricercatore ed è fissato a priori; la censura di II tipo ha condizioni sperimentali che sono un po' più controllate in quanto prevede di terminare l'osservazione quando si è raggiunto un certo numero di eventi. La censura random invece

si ha quando l'osservazione termina per motivazioni che sono sotto il controllo del ricercatore. Il numero di censure dipende da due fattori: il tasso di accadimento dell'evento oggetto di osservazione; la lunghezza del periodo di osservazione. Se l'evento è raro il periodo di osservazione è breve il numero di censure sarà inevitabilmente alto.

4.2.2 CENSURE INFORMATIVE E NON INFORMATIVE

Un meccanismo di censura non informativo opera indipendentemente dal tempo di attesa dell'evento oggetto di studio, cioè è indipendente da esso. Se la censura accade a individui che hanno maggiori o minori probabilità di sperimentare l'evento, il meccanismo di censura è informativo.

Le censure di I e II tipo non creano problemi in questo senso perché i metodi di sopravvivenza trattano questi tipi di censura senza gravi distorsioni delle stime di massima verosimiglianza e verosimiglianza parziale dei modelli. La censura random, invece, per non produrre stime distorte impone che le censure casuali siano non informative. E' difficile conoscere la dimensione e la direzione della distorsione nel caso vi siano presenti censure informative: proprio per questo motivo si cerca di inserire tra le variabili da rilevare qualsiasi fattore che possa influire sul rischio di censura. In ogni caso, se ci sono numerose censure bisogna prendere con cautela le stime ottenute.

4.2.3 EFFETTO DELLE CENSURE (NON INFORMATIVE) SULLE ANALISI STATISTICHE

A questo punto il problema maggiore è sul come trattare questa tipologia di dati. Non si può per esempio lavorare solamente sui casi non censurati poiché si finirebbe per sottostimare la durata della permanenza nello stato iniziale. Poco corretta sarebbe anche la soluzione di assegnare ai censurati una durata effettiva pari a quella del tempo di censura. Una soluzione non scorretta, ma non sempre soddisfacente, è quella di

dicotomizzare le storie degli individui fissando un periodo di osservazione utile per tutto il campione e considerando quelli che hanno sperimentato l'evento entro una certa durata. Un'estensione di questa strategia spesso usata in demografia è quella di analizzare storie tronche di generazioni confrontando i comportamenti a durate simili. La soluzione migliore resta tuttavia quella di trattare sia le censure che le non censure sfruttando al massimo l'informazione della censura, cosa che ci permettono di fare i metodi di analisi che si tratterà nei paragrafi successivi.

4.3 TEMPO DISCRETO

Il concetto fondamentale per valutare la propensione a sperimentare un evento in un intervallo temporale discreto è quello di rischio (hazard).

Sia T una variabile casuale discreta i cui valori T_i indichino il periodo di tempo j in cui l'individuo i -esimo sperimenta un determinato evento. Usualmente la distribuzione di una variabile T è descritta dalla funzione di densità di probabilità $\Pr[T_i=j]$, cioè la probabilità che l'individuo i sperimenti l'evento prima dell'intervallo j -esimo; alternativamente si può utilizzare la funzione di ripartizione $F=\Pr[T_i<j]$, cioè la probabilità che l'individuo sperimenti l'evento prima dell'intervallo j -esimo. La migliore rappresentazione di T rimane comunque la funzione di probabilità condizionata, in quanto un evento può accadere solamente se questo non è già avvenuto, cioè dalla distribuzione di probabilità che l'individuo i sperimenti l'evento nell'intervallo j dato che non l'ha ancora sperimentato precedentemente, altrimenti definita come funzione di rischio e tempi discreti:

$$h(t_{ij}) = \Pr\left[\frac{T_i = j}{T_i \geq j}\right]$$

Per convenzione successivamente considereremo individui appartenenti ad uno stesso gruppo omogeneo, eliminando così dalle notazioni il pedice appartenente alle diverse popolazioni.

Definendo con N_j il numero di individui a rischio durante l'intervallo j -esimo e con E_j il numero di individui che hanno sperimentato l'evento nell'intervallo j -esimo, la stima della funzione di rischio per ogni intervallo risulta:

$$h(t_{ij}) = \frac{E_j}{N_j}$$

Altra funzione fondamentale nell' Event History Analysis è la funzione di sopravvivenza, definita come la probabilità che un individuo non sperimenti l'evento nel j -esimo intervallo o in qualsiasi altro periodo di tempo:

$$S(t_{ij}) = Pr[T_i > j]$$

$S(t_{ij})$ è una funzione che con il passare del tempo passa da 1 (nessun individuo ha ancora sperimentato l'evento) a 0 (tutti gli individui hanno sperimentato l'evento).

Per calcolare i valori della funzione si può utilizzare il metodo diretto o quello indiretto. Il primo si effettua rapportando il numero di coloro che non hanno sperimentato l'evento entro la fine del periodo j e il totale di individui ad inizio periodo. In altre parole

$$S(t_j) = Pr\left[\frac{N_j}{N_0}\right]$$

Il metodo indiretto, invece, sfrutta la conoscenza del rischio relative. Se si considera che il complemento a 1 della probabilità condizionata esprime la probabilità di non sperimentare l'evento nell'intervallo j -esimo a chi è arrivato senza evento all'inizio dell'intervallo, quanti

sopravvivono all'evento entro la fine dell'anno j-esimo (e dunque sono sopravvissuti all'inizio del j+1-esimo intervallo) sono dati da:

$$S(t_j) = S(t_j - 1) * [1 - h(t_j)]$$

E' possibile quindi stimare la funzione di sopravvivenza alla fine dell'anno j-esimo indirettamente attraverso la funzione di rischio:

$$S(t_j) = [1 - h(t_j)] * [1 - h(t_j - 1)] * [1 - h(t_j - 2)] * ... * [1 - h(t_0)] * S(t_0)$$

La presenza di censure fa sì che dalla funzione di sopravvivenza non si possa passare alla stima della funzione di rischio.

A partire dalla funzione di sopravvivenza si può invece calcolare la vita mediana in un determinato stato, indicata come quel valore di T per cui il valore della funzione di sopravvivenza è pari a 0,5. Nel caso in cui non sia possibile calcolare la mediana poiché più della metà degli individui non sperimenta l'evento, si possono calcolare altri percentili (per esempio il 75esimo percentile) oppure si può ricorrere a tassi di sopravvivenza cumulati, ovvero valori della stima della funzione di sopravvivenza dopo definiti periodi di tempo. Occorre ricordare che la vita mediana è solamente un indicatore sintetico che può riassumere distribuzioni di funzioni di rischio molto diverse tra di loro.

4.4 TEMPO CONTINUO

Nel caso si abbiano suddivisioni del tempo più dettagliate (mesi, settimane, giorni) si può parlare di tempo continuo. Si rende quindi necessario modificare tutte le definizioni elencate precedentemente. Un problema rilevante è che nel tempo continuo la funzione di sopravvivenza è definita e stimata facilmente, mentre quella di rischio no, implicando una messa a punto di metodi di stima e di analisi differenti.

Essendo il tempo una variabile continua e quindi infinitamente suddivisibile, la distribuzione dei tempi di accadimento ha le seguenti probabilità:

- 1) La probabilità di osservare un evento in un particolare tempo è infinitamente piccola, prossima a zero
- 2) Anche la probabilità di avere due eventi che avvengono simultaneamente (ties) è infinitamente piccola e pari a zero

Sia T una variabile aleatoria continua definita su uno spazio degli eventi Ω a valori positivi $(0, +\infty)$, cioè il tempo intercorrente tra t_0 e il momento in cui si verifica il cambiamento di stato. In questo caso per l'individuo i -esimo la funzione di sopravvivenza, cioè la probabilità di sopravvivere all'evento al tempo t_j è data da:

$$S(t_{ij}) = Pr[T_i > t_j]$$

Per definire invece la funzione di rischio è necessario introdurre il concetto di intervallo infinitesimale. Il rischio è il rapporto tra la probabilità condizionata che un evento accada in un certo intervallo e l'ampiezza di quell'intervallo, quando questo diventa molto piccolo:

$$h(t_{ij}) = \lim_{\Delta t \rightarrow 0} \left[Pr \frac{[t_j \leq T_i \leq t_j + \Delta t]}{T_i \geq t_j} \right]$$

Questa definizione indica che la funzione di rischio nel continuo non è una probabilità, ma un tasso (tasso istantaneo di transizione nel punto t_j) ovvero una probabilità condizionata per unità di tempo. Il tasso si può intendere come una propensione a cambiare di stato al tempo t_j e può assumere valori superiori ad 1.

4.5 LE FUNZIONI STATISTICHE NEL CONTINUO

La funzione di densità è descritta come:

$$f(t) = \lim_{\Delta t \rightarrow 0} \left[Pr \frac{[t \leq T \leq t + \Delta t]}{\Delta t} \right]$$

Ed esprime la probabilità che l'evento accada nell'intervallo Δt quando l'intervallo di tempo diventa infinitesimo.

La funzione di ripartizione, calcolata come:

$$F(t) = Pr[T \leq t] = \int_0^t f(\tau) d\tau$$

Indica la probabilità che la durata dell'episodio sia minore o uguale a t .

Come detto precedentemente, nell'analisi di dati di durata la funzione di ripartizione viene sostituita dalla funzione di sopravvivenza, calcolata come complemento a 1 rispetto alla funzione di ripartizione.

$$S(t) = 1 - F(t) = Pr[T > t]$$

Come si può intuire, essa indica la probabilità che la durata di un episodio sia superiore a t , cioè che l'evento dal quale l'episodio è chiuso accada dopo t , altrimenti detta come la probabilità di non aver sperimentato l'evento entro l'istante t .

Inoltre, sapendo che

$$\lim_{T \rightarrow +\infty} F(T) = F(+\infty) = 1 \quad T \rightarrow +\infty$$

E' facile capire che $S(+\infty) = 0$, tranne nei casi in cui la distribuzione di T risulti difettiva. In questo caso, che si verifica quando non tutti sperimentino l'evento alla fine del periodo dell'esposizione a rischio, $F(+\infty)$ sarà uguale a $1-g$, e, di conseguenza, $S(+\infty)$ sarà uguale a g , dove g sarà la probabilità di non sperimentare l'evento.

Partendo dall'equivalenza tra le varie funzioni esposte fino ad ora si ottengono le seguenti equivalenze:

$$h(t) = \frac{f(t)}{S(t)}$$

$$h(t) = -\frac{\partial \ln S(t)}{\partial t}$$

$$S(t) = \exp \left\{ -\int_0^t h(\tau) d\tau \right\}$$

4.5.1 METODI NON PARAMETRICI PER DATI NEL TEMPO CONTINUO

Verranno descritti tre metodi per lo studio di dati nel continuo tramite metodi non parametrici: i primi due prevedono di raggruppare i tempi in intervalli per costruire tavole di eliminazione, il terzo (Kaplan-Meier) è invece peculiare per dati espressi nel continuo.

Per costruire una tavola di eliminazione occorre che i tempi rilevati siano raggruppati in intervalli di tempo contigui. Il concetto è quello di partizionare l'asse in un certo numero di intervalli anche di ampiezza differente. Tali intervalli includono l'estremo inferiore ed escludono quello superiore. Il vantaggio di questo sistema è che può essere utilizzato per una grande quantità di dati, gli svantaggi sono che i risultati dipendono dalla suddivisione che viene effettuata dell'asse temporale e che per ogni intervallo vi debba essere una quantità sufficiente di eventi.

a) Metodo a tempi discreti

Dati E_j come numero di episodi che terminano nell'intervallo j -esimo, C_j come numero di episodi censurati nell'intervallo j -esimo e N_j come il numero di individui che arrivano all'istante t_j senza avere sperimentato l'evento.

La probabilità condizionata che un individuo a rischio all'inizio dell'intervallo j sperimenti l'evento durante l'intervallo si calcola come:

$$p(t_j) = \frac{E_j}{N_j}$$

Ne consegue che la stima della funzione di sopravvivenza nel continuo alla fine dell'intervallo j -esimo è:

$$S(t_j) = [1 - p(t_1)] * [1 - p(t_2)] * \dots * [1 - p(t_j)]$$

Poiché la stima della funzione di rischio è ora un tasso e non una probabilità, la stima di tale funzione è data da:

$$h(t_j) = \frac{p(t_j)}{\text{AmpiezzaIntervallo}_j - \text{esimo}}$$

Tale misura esprime la probabilità media di accadimento dell'evento per unità di tempo.

b) Il metodo attuariale (Life table method)

Mentre precedentemente si assumeva che gli eventi e le censure accadessero alla fine dell'intervallo, con questo metodo si presume che essi siano uniformemente distribuiti lungo l'intervallo e che si collochino quindi in media a metà intervallo. Ne deriva che cambia, quindi, la definizione della popolazione a rischio.

Mentre la definizione dei censurati e del numero di individui che arrivano all'istante t_j senza sperimentare l'evento rimangono identiche a quelle utilizzate nel metodo precedente, la popolazione a rischio di sopravvivere alla fine dell'intervallo j -esimo utile per stimare la funzione di sopravvivenza è data da:

$$R_j = N_j - 0,5 * C_j$$

A questo punto cambia anche la definizione di probabilità condizionata e diviene:

$$p(t_j) = \frac{E_j}{R_j}$$

La funzione di sopravvivenza diventa:

$$S(t_j) = \prod_s (1 - p(t_s))$$

Per stimare la funzione di rischio è necessario ridefinire il significato di individui a rischio in quanto a noi interessa non la popolazione a rischio di sopravvivere trascorso l'intervallo, ma coloro che sono a rischio di subire l'evento durante l'intervallo.

Il nuovo numero di persone a rischio è dato da:

$$RR_j = N_j - 0,5 * C_j - 0,5 * E_j$$

Poiché la stima della funzione di densità in corrispondenza al punto centrale dell'intervallo è data da:

$$f_j = \frac{S_j - S_{j+1}}{\text{AmpiezzaIntervallo}}$$

si ricava che il tasso di transizione nel punto centrale dell'intervallo è dato da:

$$h(t_j) = E_j * [\text{Amp.Intervallo} * (R_j - 0,5 * E_j)]$$

c) Lo stimatore di Kaplan-Meier

Un vantaggio dello stimatore di Kaplan-Meier è che ha un trattamento ottimale dell'informazione in quanto gli intervalli non sono definiti dal ricercatore ma dai tempi stessi di accadimento degli eventi. Inoltre ha una solida base teorica: Kaplan e Meier nel 1958 dimostrarono che la funzione di sopravvivenza stimata con questo metodo è la stima di massima verosimiglianza della funzione di sopravvivenza.

Al contrario, poiché il metodo non permette una facile rappresentazione tabellare dei dati per un numero elevato di intervalli, tale metodo è particolarmente adatto per il trattamento di piccoli dataset. Ulteriore inconveniente di tale metodo è che non consente il calcolo diretto dei tassi di transizione.

Ciascun intervallo di KM inizia con un tempo di evento osservato e finisce appena prima che cada l'evento successivo. Viene convenzionalmente definito anche l'intervallo iniziale che comincia con t_0 (generalmente pari a 0) e finisce appena prima che accada l'evento.

Inoltre, se un individuo è censurato in corrispondenza di un tempo in cui accade un evento, si assume che l'evento preceda il caso censurato: ciò fa sì che il caso censurato cada nell'intervallo di tempo dell'evento osservato. Analogo trattamento hanno i ties.

La funzione di sopravvivenza (a gradini) è quindi calcolabile come:

$$S(t_j) = \prod_s (1 - p(t_s)) = \prod_s 1 - \frac{E_s}{N_s}$$

Il metodo di KM non permette comunque di stimare bene la funzione di rischio.

La formula:

$$h_{KM}(t_j) = \frac{p_{KM}(t_j)}{\text{AmpiezzaIntervallo}_j - \text{esimo}}$$

non è una buona stima poiché al crescere del tempo il numeratore di questa equazione cresce e dividendo per l'ampiezza dell'intervallo le stime risultanti variano molto da un intervallo all'altro e sono molto erratiche.

Si ricorre quindi ad un'altra funzione, chiamata funzione di rischio cumulato:

$$H(t_{ij}) = \int_t^j h(\tau) d\tau$$

La stima di tale funzione si può calcolare tramite due metodi:

$$H(t_{ij}) = h_{KM}(t_1) * \text{amp.int.1} + h_{KM}(t_2) * \text{amp.int.2} + \dots + h_{KM}(t_j) * \text{amp.int.j}$$

$$H(t_j) = -\ln(S_{KM}(t_j))$$

Il tasso di transizione integrato non è né una probabilità né un tasso. Poiché esso tuttavia non ha un significato di immediata comprensione, si preferisce raffigurarlo piuttosto che tabularlo.

Da tale funzione è possibile ricavare informazioni sull'andamento della funzione di rischio. Il rischio cumulato parte da 0 ed è una funzione non decrescente: se il tasso è costante nel tempo, il rischio cumulato cresce linearmente; se il tasso cresce la sua cumulata cresce più che linearmente; se il tasso diminuisce quello integrato cresce meno che linearmente.

4.6. IL MODELLO SEMIPARAMETRICO DI COX

Non è sempre possibile avere a priori una idea di come possa essere distribuito il tasso di transizione con il tempo né è sempre possibile avere funzioni che si adattano bene alla distribuzione sostantiva del rischio nel tempo. Il modello semiparametrico di Cox tende a specificare la forma funzionale di influenza delle covariate, mentre non specifica la forma del tasso di transizione.

Questo metodo risulta utile quando non si ha una idea della forma di dipendenza del tasso dal tempo, quando abbiamo una debole teoria che supporta lo specifico metodo parametrico, quando il ricercatore ha una idea sulla dipendenza del tasso dal tempo ma non sa come parametrizzarla oppure quando si è interessati solamente alla grandezza e alla direzione degli effetti delle covariate considerate controllando per la dipendenza nel tempo.

Nella sua forma base il modello assume la seguente forma:

$$h_i(t) = h_{0i} * \exp \{ \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} \}$$

Il tasso di transizione per l'individuo i-esimo al tempo t risulta quindi un prodotto di due fattori distinti: la funzione di rischio di base che non è specificata tranne per il fatto che deve essere non negativa e una funzione lineare di un set di k covariate fisse nel tempo esponenziata.

Se applichiamo la trasformata logaritmica al tasso otteniamo:

$$\ln(h_i(t)) = \ln(h_{0i}) + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}$$

Il metodo utilizzato per la stima dei parametri delle variabili esplicative è quello della verosimiglianza parziale invece di quella “completa”, in quanto consente comunque di ottenere stimatori con buone proprietà campionarie.

Occorre precisare che tale modello è a rischi proporzionali, cioè che il rapporto tra i rischi di due individui è costante nel tempo:

$$\frac{h_i(t)}{h_j(t)} = \exp[\beta_1(x_{i1} - x_{j1}) + \dots + \beta_k(x_{ik} - x_{jk})]$$

Proprio per questo motivo prima di applicare il modello nella sua forma base occorre verificare che questa assunzione di proporzionalità degli effetti delle covariate sia confermata. Generalmente si utilizza il metodo grafico, utile specialmente quando vi sono poche covariate preferibilmente categoriali. Si suddivide il campione in tanti sottogruppi quante sono le modalità della covariata che si intende analizzare e si traccia il grafico di $\log(-\log(S(t))) = \log H(t)$ e verificare se tali distribuzioni differiscono tra di loro a meno di una costante di proporzionalità k .

Per quanto riguarda i risultati che si ottengono dal modello, non abbiamo informazioni circa la baseline, ma solo sull’effetto delle varie covariate.

Occorre ricordare che β esprime l’effetto sul logaritmo del tasso di una differenza di una unità nel predittore associato, al netto dell’effetto delle restanti covariate, mentre $\exp(\beta)$ esprime la medesima variazione riferita al tasso (sempre al netto delle rimanenti covariate). Nel caso nel modello si abbiano variabili categoriali, valori inferiori a 1 vanno letti come quote di rischio inferiore a quello del gruppo di base, valori superiori a 1 si interpretano come quote di rischio superiori a quelle del fattore preso come baseline. Se, invece, la variabile predittiva è continua, il valore va interpretato all’aumentare di una

unità. Per calcolare la significatività dei parametri si utilizza il test di Wald con il corrispondente p-value.

Per valutare la bontà totale del modello, poiché il modello di Cox si stima usando il metodo di massima verosimiglianza, la statistica log-likelihood (LL) è quella più utilizzata. Alternativamente si può utilizzare una trasformata del modello LL, cioè $-2LL$, che in questo caso non si può parlare di devianza perché nel modello con verosimiglianza parziale non esiste il modello completamente saturato. Tale valore è però utilizzabile per costruire una statistica test basata sul rapporto di verosimiglianza che verifica la uguaglianza del modello con quello nullo (in cui l'ipotesi nulla è che tutti i parametri risultino nulli) oppure con un altro modello che esclude alcuni parametri (dove l'ipotesi nulla prevede che i parametri oggetto di osservazione siano effettivamente uguali a 0).

Il modello di Cox presenta però un tallone d'Achille riguardante i ties. Poiché i ties sono inevitabili nei dati reali sono stati sviluppati accorgimenti per tenere sotto controllo il problema. Così, se un tempo censurato è uguale al tempo di osservazione di un evento si assume che l'evento precede il caso censurato: in questo modo il caso censurato è collocato tra coloro che sono a rischio di subire l'evento. Nel caso vi siano due tempi di accadimento dell'evento uguali si adottano soluzioni diverse. Il metodo più raffinato è quello di calcolare tutte le possibili combinazioni degli eventi che possano esistere

4.7 L'INTERVALLO TRA PARTI NELLE FAMIGLIE DI BATTAGLIA TERME

Nei paragrafi successivi si intende studiare il fenomeno della distanza inter-parto tra le famiglie di Battaglia Terme dal 1607 al 1871.

Una volta ricostruite le famiglie si è deciso di creare un nuovo dataset utilizzando come elemento di riferimento il figlio. Sono stati scelti solamente i figli di madri di cui si conoscevano le date di nascita e di morte di entrambi.

Alla fine si è costruito un dataset di 3651 record contenenti le seguenti variabili:

Ricostruzione nominativa delle famiglie in una comunità veneta: Battaglia Terme (1607 – 1871)

- Età della madre espressa in giorni
- Durata del matrimonio in giorni
- Distanza tra le nascite in giorni
- Rango del figlio
- Morte del figlio precedente (valore 1 se il figlio precedente è morto prima della nascita del successivo, 0 altrimenti)
- Età delle madre alla nascita del figlio espressa in giorni
- Età del figlio alla morte in giorni
- Luogo di nascita e di provenienza della madre (valore 1 se nata o proveniente da Battaglia Terme, 0 altrimenti)
- Anno di nascita del figlio

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	CodMat	MatN	MatM	DatM1	DatFE1	etamad	duram	etamatu	distanza	Rangofig	mortefiglo	etafigl	etafigl	Provenienz	Luogonascit
42		18/10/1668					5663		445	3	0	8645		1	1
43		09/03/1672	11/01/1679				5663		1238	4		9883	2499	1	1
44		02/01/1677	04/01/1677				5663		1760	5	0	11643	2	1	1
45		02/01/1677	04/01/1677				5663		0	6	1	11643	2	1	1
46		07/01/1679					5663		735	7	1	12378		1	1
47		07/01/1679					5663		0	8	0	12378		1	1
48	10	24/05/1632		31/07/1664											
49	10			31/07/1664											
50	13	25/12/1664		02/10/1664											
51	13			02/10/1664											
52	14			12/02/1665											
53	14	15/02/1647		12/02/1665											
54		26/07/1667						6572		1		7466		1	1
55		02/11/1669						6572	830	2		8296		1	1
56		24/05/1673						6573	1299	3		9595		1	1
57		06/01/1676	04/03/1672					6574	957	4		10552		1	1
58		27/11/1678						6575	1056	5		11608		1	1
59		02/09/1681	22/09/1681					6576	1010	6		12618	20	1	1
60		12/11/1683						6576	801	7	1	13419		1	1
61		02/11/1686						6576	1086	8		14505		1	1
62	15			15/02/1665											
63	15	21/12/1644		15/02/1665											
64	16	10/07/1625		17/06/1665											
65	16			17/06/1665											
66	17	17/04/1635	01/07/1683	02/07/1665	05/03/1681										
67	17		05/03/1681	02/07/1665	05/03/1681										
68	18	01/11/1668	14/11/1668			05/03/81	5725	02/07/65		1		01/11/68	13	0	0
69	18			12/10/1665											
70	18			12/10/1665											
71	19	23/08/1643	14/05/1666	24/11/1665	14/05/1666										
72	19			24/11/1665	14/05/1666										
73	23		14/09/1673	21/01/1666											
74	23			21/01/1666											
75		28/12/1667	05/08/1696												
76		06/05/1674													
77	24			04/02/1666											
78	24			04/02/1666											

Immagine 4.7.1.1: Dataset per lo studio della distanza interparto

Per lo studio della distanza tra i parti si è utilizzato il programma statistico SAS. Il modello che si intende utilizzare è quello di Cox nella sua versione base, verificando precedentemente l'esistenza di una proporzionalità tra le covariate. Vengono considerati dati censurate quelli che hanno una distanza tra parto superiore ai 5 anni (1825 giorni).

Alla fine il dataset era così ripartito:

- Il numero di maschi presenti è di 1838, pari al 50,34% dell'intero campione studiato. Ne consegue che il numero di femmine è di 1813 (49,66%)
- I dati non censurati sono 3232 (88,55%), mentre quelli censurati sono 419 (11,55%)

Per poter stimare un modello di Cox occorre verificare l'assunto di proporzionalità degli effetti delle covariate. Per questo fine si decide di utilizzare il metodo grafico: verrà utilizzata il metodo attuariale (o "Life Table Method"), presentato precedentemente, sia per il sesso che per il rango del figlio.

Come prima cosa viene suddiviso il tempo che intercorre tra un parto all'altro in classi di 250 giorni.

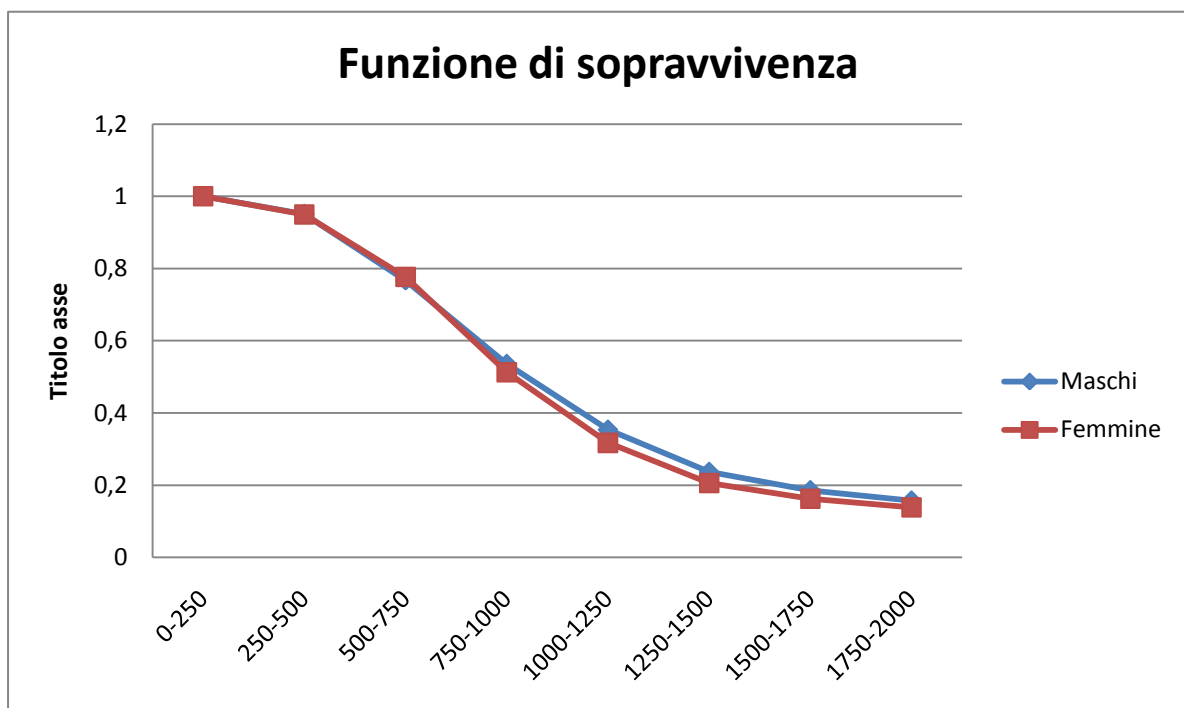


Grafico 4.7.1.1: Funzione di sopravvivenza suddivisa per genere

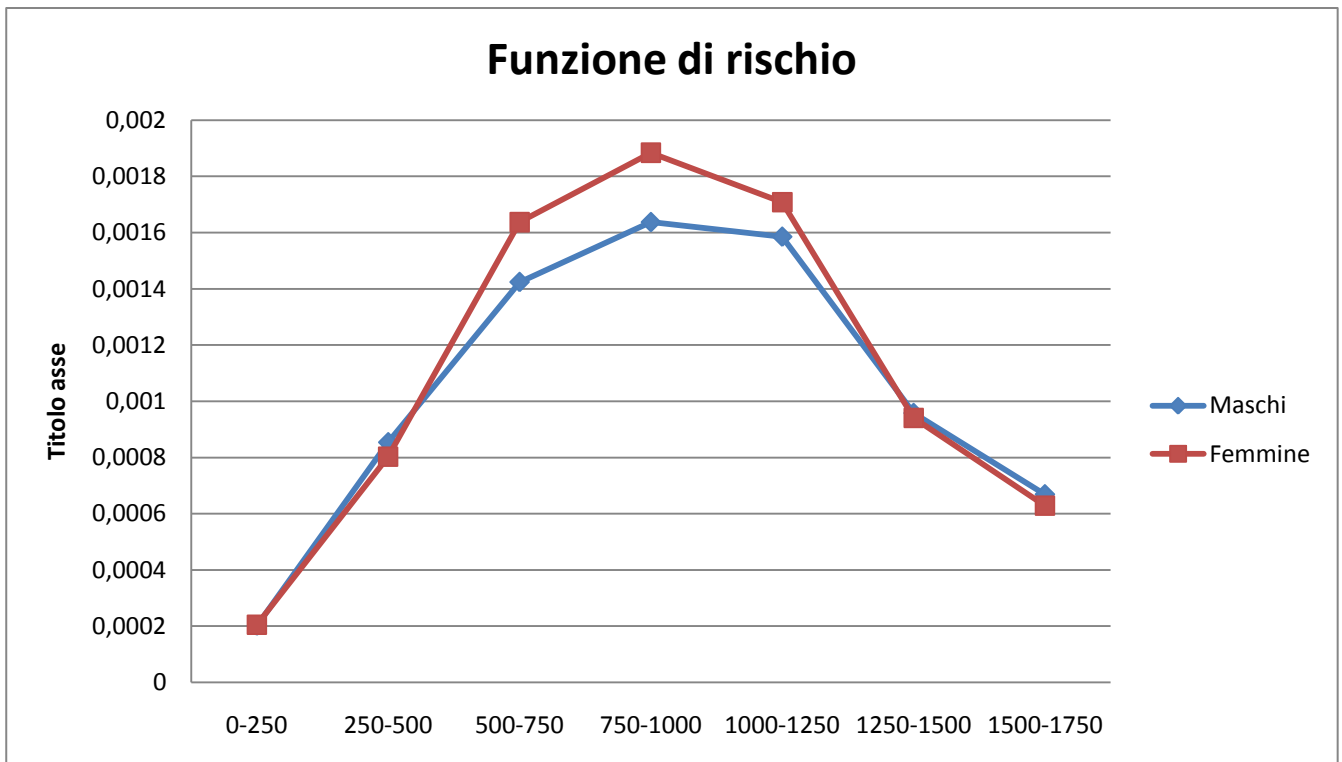


Grafico 4.7.1.2: Funzione di rischio suddivisa per genere

Le funzioni di sopravvivenza sono simili tra di loro sia nella parte iniziale che in quella finale. Una leggera differenza avviene nella parte centrale dove le femmine subiscono l'evento in misura maggiore rispetto ai maschi.

Nello studio per verificare l'effetto di proporzionalità bisogna escludere i figli che hanno rango 1, in quanto non presentano una distanza interparto con il figlio precedente. Inoltre tutti i figli con rango superiore a 5 sono stati racchiusi in una unica classe chiamata rango6.

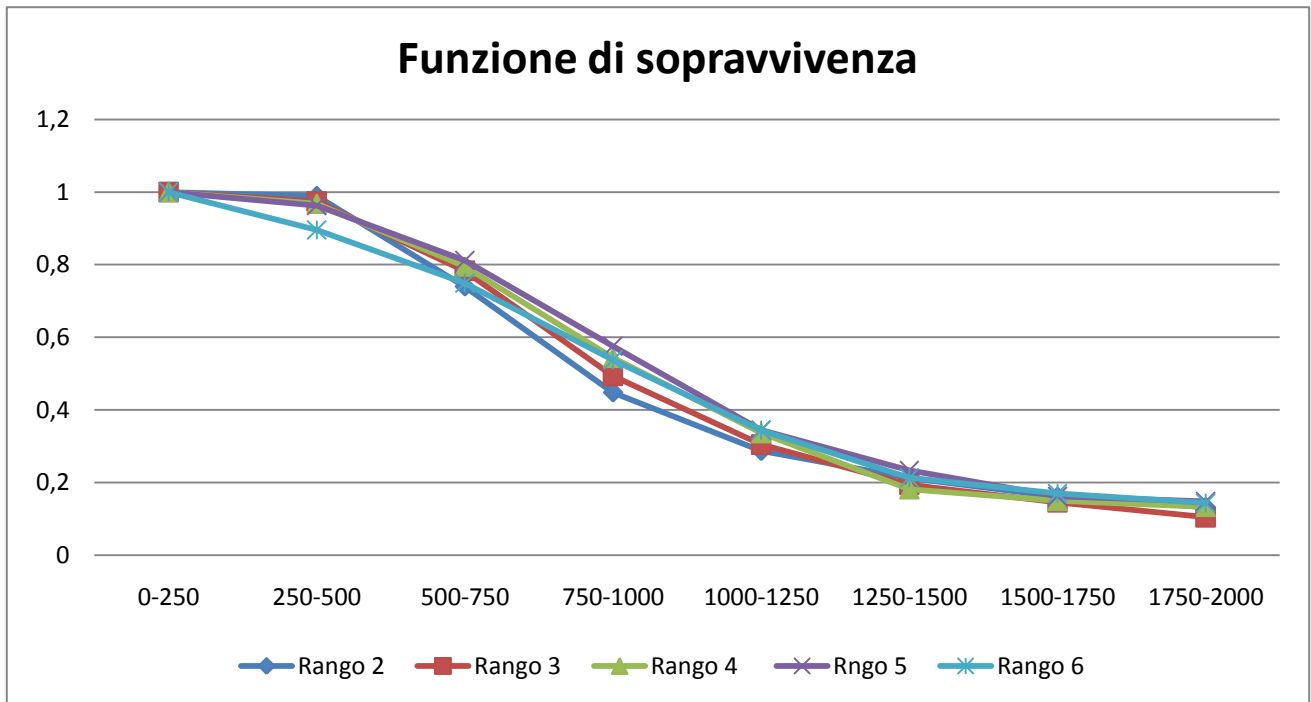


Grafico 4.7.1.3: Funzione di sopravvivenza suddivisa per rango

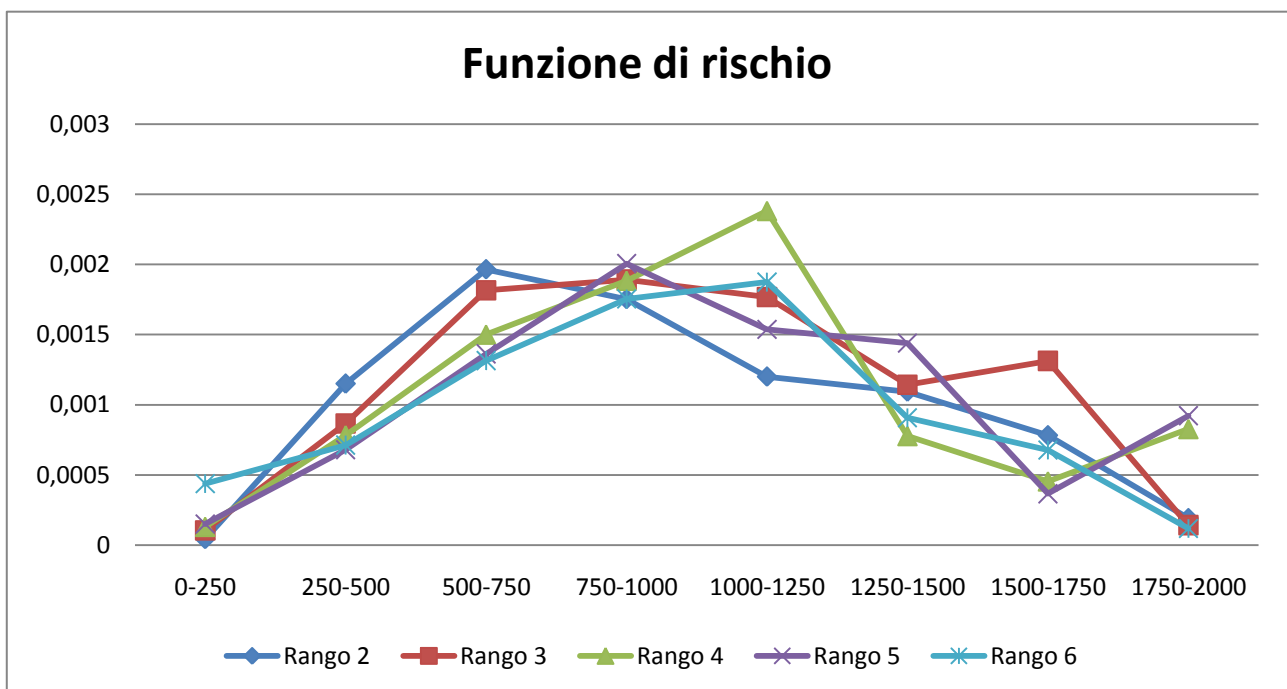


Grafico4.7.1.4: Funzione di rischio suddivisa per rango

La funzione di sopravvivenza suddivisa per rango indica che anche in questo caso gli effetti che abbiamo sono proporzionali in quanto le linee non si differenziano molto l'una dall'altra. La funzione di rischio, grazie alla scala di misura più dettagliata, evidenzia come

le curve in generale siano simili tra di loro, ma che confrontate una ad una possono subire delle differenze. Se si confronta la curva dei figli di rango 2 con quelli di rango 4, per esempio, si nota come i primi tendano a subire l'evento nei periodi iniziali, mentre i secondi in periodisuccessivi

Una volta verificato l'assunto di proporzionalità non rimane altro che stimare il modello per la distanza interparto.

Analisi delle stime di massima verosimiglianza

Stima dei Variabile	Errore		Rapporto		Descrizione		
	DF	parametri	standard	Chi-quadrato	Pr > ChiQuadr	rischio	variabile
sex	1	-0.10887	0.05192	4.3969	0.0360	0.897	
etamatr	1	8.73507E-6	2.77082E-6	9.9384	0.0016	1.000	etamatr
etaalfiglio	1	-0.0000204	2.46504E-6	68.4084	<.0001	1.000	etaalfiglio

Figura 4.7.1.1 Modello di Cox per l'intervallo tra nascite

Le variabili che entrano a far parte del modello sono il sesso del bambino (sex), l'età al matrimonio della madre (etamatr) a l'età della madre alla nascita del figlio (etaalfiglio). La categoria di riferimento per il sesso è quella maschile. Ne deriva che il rischio di subire l'evento nelle femmine è il 10% inferiore rispetto che negli uomini. Etamatr ed etaalfiglio sono invece variabili continue e il valore dei parametri va moltiplicato per il numero di giorni in cui intercorre l'event. Mentre per l'età al matrimonio il rischio di avere un figlio prima aumenta con il crescere dell'età, viceversa tale rischio diminuisce all'aumentare dell'età in cui si ha avuto il figlio. Ciò vuol dire che chi si sposa tardi tenderà ad allungare le nascite tra i figli, mentre chi ha un figlio da giovane tenderà ad avere il successivo in tempi più ristretti rispetto a chi sarà più avanti con l'età.

4.8 OSSERVAZIONI CONCLUSIVE

Il lavoro effettuato nella parrocchia di Battaglia Terme è risultato essere molto interessante in quanto si è potuto lavorare su un periodo di tempo molto lungo, potendo quindi studiare l'evoluzione di diverse dinamiche nel corso del tempo.

Molti risultati sono risultati significativi del periodo considerato, anche se in altri casi sono emersi dei dubbi circa la bontà dei dati su cui si è lavorato.

Per quanto riguarda la costruzione nominativa, essa si è potuta svolgere in maniera sufficientemente adeguata, anche se il legame tra marito, moglie e figli è stato molte volte difficoltoso, soprattutto per quanto riguarda gli anni iniziali dello studio dove il numero di informazioni era minore.

Anche la stima del modello sulla distanza tra i parti è risultato essere di difficile realizzazione in quanto il numero di unità su cui lavorare si è ridotto drasticamente a causa dell'elevato numero di informazioni richieste. Le stime del modello andrebbero quindi verificate con altri studi simili.

Il lavoro non si può considerare quindi del tutto completo in quanto le analisi ancora effettuabili sono molteplici e col passare del tempo si potrà usufruire anche di metodi migliori per quanto riguarda la ricostruzione delle famiglie.

BIBLIOGRAFIA

C. Agostinelli, S. Sartorelli : Quaderini ASID – Introduzione al linguaggio di SAS System – 2002

Agostini F. (1989), Fonti e studi di storia veneta. Anagrafi parrocchiali e popolazione nel Veneto tra XVII e XIX secolo, Istituto per le ricerche di storia sociale e religiosa, Vicenza
SIDES, Società italiana di demografia storica, Popolazione e storia, 2000, pp. 177-198,

M. Breschi, L. Pozzi, R. Rettaroli (1994), Analogie e differenze territoriali nella crescita della popolazione italiana, 1730-1911, Bollettino di Demografia Storica, 20, 1994.

M. Breschi, R. Derosas, C. Lagazio, M. Manfredini (1999), L'influenza del contesto familiare sulla sopravvivenza dei bambini. Risultati di indagini microdemografiche sull'Italia dell'Ottocento, Bollettino di Demografia Storica, 30-31, 1999.

M. Breschi, R. Derosas e M. Manfredini (1999), “New perspectives in demographic micro-studies Italy 19th century”,

M. Breschi, R. Derosas, P.P. Viazzo (a cura di, 2003), Piccolo è bello. Approcci microanalitici nella ricerca storicodemografica, Forum, Udine, 2003.

D.R. Cox (1972), Regression Model and Life Tables (with discussion), Journal of the Royal Statistical Society, B, 34, 1972

G. Dalla Zuanna , A. Rosina, F. Rossi (2004) “Il Veneto”, Storia della popolazione dalla caduta di Venezia ad oggi, 8, 129-141

L. Del Panta, R. Rettaroli (1994), Introduzione alla demografia storica, Editori Laterza, Bari-Roma.

L. Del Panta, L. Pozzi, R. Rettaroli, E. Sonnino (a cura di, 2002), *Dinamiche di popolazione, mobilità e territorio in Italia. Secoli XVII-XX*, Forum, Udine, 2002.

L. Del Panta, V. Rodilossi (2004), *Testing Inverse Projection, Differentiated Inverse Projection and Stochastic Inverse Projection: A Reconstruction of the Population of Sardinia between 1861 and 1921 Using Three Different Techniques*, in Barbi, Bertino, Sonnino (2004).

R.Derosas, "Between identity and assimilation: Jewish fertility in nineteenth-century Venice"

L. Fabbri (1997): *Statistica multivariata – Analisi esplorativa dei dati – McGrawHill*

P. Festy, F. Prioux (2002), *An Evaluation of the Fertility and Family Surveys Project*, United Nations, New York and Geneva, 2002

M. Furegato, "I registri parrocchiali di Battaglia Terme (1607-1871)", *Tesi di laurea, Anno accademico 2006-2007*

P.R. Galloway (1994), *A Reconstruction of the Population of North Italy from 1650 to 1881 using Annual Inverse Projection with Comparison to England, France and Sweden*, *European Journal of Population*, 10, 1994.

C. Ge Rondi (1988), *L'analisi nominativa in demografia storica: metodi e problemi. Il caso di una parrocchia*, Giuffrè, Milano, 1988.

L. Henry (1961), *Some Data on Natural Fertility*, *Eugenics Quarterly*, 8, 1961.

M. Livi Bacci (1975), Una disciplina in rapido sviluppo: la demografia storica, in E. Sori (a cura di), Demografia storica, il Mulino, Bologna, 1975

M. Livi Bacci (1987), Prefazione all'edizione italiana di J. Dupâquier, Per la demografia storica, SEI, Torino, 1987 (ed. originale: J. Dupâquier, Pour la démographie historique, PUF, Paris, 1984).

Livi Bacci, Introduzione alla demografia (Terza edizione, 1999)

M. Livi Bacci (2005), Storia minima della popolazione del mondo, il Mulino, Bologna, 2005.

F. Ongaro : Corso di analisi dei fenomeni socio demografici – Appunti delle lezioni – A.A. 2005-2006

M. Oris (2003), Demografia storica e storia della famiglia. Due genealogie intellettuali, in Breschi, Derosas, Viazzo (2003).

F. Rossi, A. Rosina (1994), Ricostruzioni aggregate dei processi evolutivi delle popolazioni, CLEUP, Padova, 1994.

F. Rossi, A. Rosina (1998), Il Veneto tra Sette e Ottocento, Bollettino di Demografia Storica, 28, 1998.

<http://www.comune.battaglia-terme.pd.it>

http://it.wikipedia.org/wiki/Archivi_parrocchiali

RINGRAZIAMENTI

Desidero ringraziare il Professor Rossi per la disponibilità che ha sempre avuto nei miei confronti.

Ringrazio il dott. Donato per la raccolta dei dati: senza di Lui questo lavoro non sarebbe stato possibile.

La professoressa Meggiolaro per l'aiuto nella stesura del modello finale

Matteo per l'aiuto nella costruzione della "Scheda di Famiglia" con Access

Ringrazio i miei genitori che in questo periodo mi hanno lasciato preparare questa Tesi in tutta tranquillità senza assillarmi.

Ringrazio mia sorella che finalmente si è trovata il "moroso".

Ringrazio Veronica che è stata molto comprensiva nei miei confronti anche se pure per lei è stato un periodo piuttosto difficile e stressante.

Ringrazio i miei nonni paterni che dall'alto mi hanno sempre incoraggiato a non mollare anche quando le situazioni si sono rivelate molto difficoltose.

Ringrazio nonni, zii, cugini vari e altri parenti che si sono interessati sempre a me.

Ringrazio infine, ma non per ordine di importanza, i miei amici. Quelli dell'università, tutti quanti, non servono i nomi, tanto si sa già a chi mi riferisco. E gli amici di Thiene, quelli che vedo settimanalmente e quelli che è da molto tempo che non frequento ma che so che ogni tanto si ricordano di me, nel bene e nel male.

E faccio i complimenti a me stesso... è stata dura ma finalmente ho finito.

