

# Atlante italiano di mortalità

**Cesare Cislaghi, Maura Dal Cason** <sup>(1)</sup>

**Maurizio Camnasio, Claudia Montalbetti** <sup>(2)</sup>

(1) Istituto di Statistica medica e Biometria, Università di Milano

(2) CILEA

## Abstract

L'Atlante di mortalità a livello comunale, giunto alla versione 5.3, è il risultato dell'attività di ricerca finanziata dal Ministero della Sanità alla Regione Emilia Romagna. La ricerca è una collaborazione fra Regione Emilia Romagna, Istituto di Statistica Medica e Biometria dell'Università di Milano e CILEA; la responsabilità scientifica è di Cesare Cislaghi. L'Atlante comprende sofisticati metodi statistici integrati con adeguate rappresentazioni grafiche che consentono di studiare la geografia delle cause di mortalità. È già in fase di distribuzione il **CD-Rom** che contiene questo Atlante. Speriamo che l'Atlante sia un piccolo contributo allo sviluppo degli interventi di prevenzione primaria in Italia.

L'Atlante già nelle sue versioni precedenti presentava numerose innovazioni metodologiche e di contenuto rispetto alle analoghe opere che lo avevano preceduto.

La descrizione dei dati utilizzati, e dei metodi di analisi contenuti nella versione precedente può essere letta sul *Bollettino del CILEA - n. 51 (1996) - "Atlante epidemiologico portatile ad utilizzo flessibile"*.

## Obiettivi

Obiettivo del progetto è in primo luogo la possibilità di ottenere la descrizione delle cause di mortalità per piccole aree e in secondo luogo l'individuazione di quei punti di aggregazioni anomale di decessi che suggeriscono la possibile presenza di aumenti reali di rischio.

Il progetto è stato studiato anche per garantire una ampia flessibilità d'uso e di indagine che si articola nelle seguenti caratteristiche:

- possibilità di analizzare tutti i comuni del territorio italiano che diventano, di volta in volta, centroidi di un'area di indagine con

raggio e forma (aree anche non circolari) variabili;

- possibilità di analizzare qualsiasi causa di decesso, e relative sottocause, con un qualsiasi tipo di aggregazione tra di loro;

- possibilità di applicare in modo semplice ed efficiente le analisi statistiche più consolidate sui dati delle aree selezionate;

- prodotto aperto ad accogliere ed aggiungere altri tipi di analisi statistiche particolari necessarie all'utente finale;

- possibilità di

aggiornare in modo rapido e trasparente gli archivi del prodotto (ad esempio l'estensione degli anni analizzati);

- utilizzo di un prodotto software che possa avere la più ampia trasportabilità su diverse piattaforme hardware;

- controllo dei problemi di riservatezza (legge 675/96) mediante criptazione degli archivi e limitazione delle ampiezze minime delle aree da analizzare.



## Risultati finali conseguiti

Il progetto ha risposto in modo adeguato agli obiettivi sopra menzionati ed in particolare:

- si può scegliere un'area specifica di comuni tra loro anche non adiacenti o si può scegliere di analizzare i comuni contenuti all'interno di una zona circolare;
- la definizione del cerchio di indagine viene determinato solo dai confini comunali risultando indipendente dalle province e regioni di appartenenza. Il raggio che delimita il cerchio di indagine può essere liberamente determinato in vari modi;
- la scelta delle cause di decesso lascia la più ampia libertà di aggregazione utilizzando i codici I.C.D. IX delle cause di mortalità;
- l'analisi statistica sulle cause di decesso all'interno dell'area di indagine predefinita consente molteplici scelte che l'utente può liberamente comporre;
  - è prevista la possibilità di indicare il gruppo di deceduti per sesso, per classi di età, per tipologia (stabili o migranti);
  - è consentita la scelta di numerosi opzioni di elaborazione dei dati (ad esempio: statistiche generali, istogramma indici comunali, grafico dei rischi relativi per età, autocorrelazione spaziale, lista dati e indici comunali, lista comuni outliers, tests di struttura, aggregati spaziali, mappa indice di frequenza, mappa stime di densità;
  - ogni opzione selezionata richiede la definizione dei parametri necessari all'elaborazione;
  - è possibile la scelta di alcune metodologie per le stime di densità: Kernel, Ripartizioni Casuali, Byesiani Empirici, Trend Surface.

## Innovazioni della nuova versione

### La Privacy

Il primo aspetto che ha occupato molta parte della nostra attenzione sono state le procedure necessarie per il rispetto delle norme riguardanti il rispetto della Legge 675/96 sulla privacy.

Si noti innanzitutto che i dati, di fonte ISTAT, sono stati forniti alla Regione Emilia Romagna dal Ministero della Sanità, dopo essere stati sottoposti ad una prima elaborazione per oscurare i campi non necessari e per riaggregare le modalità di alcune variabili in

modo da rendere pressochè impossibile il riferimento individuale di ciascun record.

L'unica informazione che potrebbe indurre qualche problema di riservatezza è quella relativa al comune di residenza, informazione ovviamente indispensabile per realizzare un atlante a livello comunale.

Per tale motivo i dati sono stati organizzati nelle matrici di un database criptato, leggibile mediante SAS e inserito nel CD di questo Atlante; il database è leggibile solo mediante una parola chiave. La parola chiave è conosciuta dal solo sistemista del CILEA che ha effettuato la criptazione nascosta in una parte compilata delle procedure SAS e quindi in nessun modo recuperabile da parte dell'utente finale dell'Atlante.

La versione prevista per "comunicare" i dati di mortalità all'interno del Sistema Sanitario Nazionale non comporta ulteriori limitazioni, se non quelle sopra elencate.

La versione invece destinata alla "diffusione" è ulteriormente restrittiva, ed in particolare :

- è soppressa la tabella dei decessi per età a livello di singoli comuni ( e quindi anche la tabella dei vivi per età che risulterebbe inutile);
- non è possibile elaborare i dati per aree complessive inferiori ai 10.000 abitanti;
- non è possibile elaborare dati per periodi di calendario inferiori ai tre anni, anche non contigui.

Queste regole fanno sì che non sia possibile elaborare dati per un numero di anni-uomo inferiore a circa 30.000, cui corrispondono circa 300 decessi. Pur nella disaggregazione per età e sesso è oltremodo improbabile arrivare a frequenze unitarie per cella relative alla mortalità generale.

### I nuovi dati

I dati relativi ai decessi riguardano ora il periodo 1981-1994, avendo aggiunto alla precedente versione il periodo 1989-1994. Per ogni anno analizzato sono stati esclusi alcuni records individuali relativi ai seguenti casi:

- nel record è codificata erroneamente una variabile essenziale;
- non è codificato il comune di residenza;
- la residenza del deceduto era all'estero.

L'esclusione dei primi due (circa 1000 decessi l'anno) di questi casi può provocare delle lievi differenze tra i totali provinciali dell'atlante e le statistiche ufficiali Istat; infatti la residenza può esser sconosciuta nel comune (e quindi noi abbiamo eliminato il caso) ma conosciuta nella provincia, e quindi considerata nelle tabelle Istat per provincia e/o regione.

Sono stati inoltre aggiornati gli archivi del periodo 1981-1988 per adeguarli alle nuove province.

I comuni che hanno avuto delle variazioni territoriali sono stati accorpati al fine di poter svolgere delle analisi su dati omogenei. L'elenco dei comuni accorpati è presente nel CD-ROM.

I dati di popolazione nella versione precedente erano quelli del censimento 1981. Nella nuova versione è stata introdotta una stima della popolazione comunale al 31.12.91 elaborata da R. Capocaccia dell'Istituto Superiore di Sanità.

### Le nuove statistiche

L'Atlante può avere diverse modalità di utilizzo: una di queste, introdotta nella nuova versione, è la possibilità di verifica di ipotesi a priori, di concentrazione di eventi in un'area o in un punto. L'analisi della struttura interna di un'area è eseguita mediante il **test di autocorrelazione I di Moran** che ha valori tra +1 e -1. Un valore positivo elevato significa che i valori di ciascuna area sono funzione dei valori delle aree più vicine e cioè che vi è dipendenza spaziale. L'elaborazione è eseguibile sia sui valori originali, sia sulle stime di densità; è chiaro che in questo caso l'autocorrelazione aumenta notevolmente proprio perché nelle stime di densità i valori di un'area sono stimati utilizzando i valori di quelle più prossime. Viene elaborato anche l'autocorrelogramma per un numero di lags definito: l'autocorrelazione con lag=1 evidenzia se vi sia autocorrelazione tra le aree adiacenti. L'analisi della autocorrelazione assorbe molte risorse di calcolo e richiede una certa conoscenza dei metodi per poter essere interpretata; ciò nonostante rappresenta un importante aiuto interpretativo.

Il nuovo **test di Stone** permette di testare se attorno ad un punto vi sia o meno una concentrazione particolare di casi. L'uso corretto di questa procedura è quello di avere una ipotesi forte a priori, quale ad esempio una fonte di inquinamento, scegliere il punto

centrale del supposto cluster di decessi come centroide dell'analisi, e quindi selezionare il test di Stone. Per interpretare meglio i dati si è predisposto per il solo centroide un grafico che ha in ordinata i valori dell'indice prescelto ed in ascissa le distanze dal centroide: in questo grafico la linea azzurra evidenzia i limiti di probabilità dell'indice cumulativo che è invece rappresentato in violetto, in rosso infine è riportato l'indice della corona esterna progressiva.

Per quanto riguarda le stime di densità, accanto agli indicatori Kernel sono stati introdotti altri metodi di stima delle densità di frequenza di decessi, ed in particolare le **Ripartizioni Casuali iterative (RCI), la Trend Surface Analysis (TSA) e gli Indicatori Bayesiani Empirici (IBE)**. La presenza di quattro sistemi molto diversi tra loro permette di verificare quanto una immagine cartografica sia dipendente dal metodo di analisi scelto oppure no.


### La nuova procedura

L'interfaccia grafica della nuova procedura è completamente rinnovata:

- la navigazione risulta più lineare;
- sono stati inseriti **help in linea**. Per ogni pagina di menù sono state inserite pagine di help, più un help generale richiamabile dalla prima schermata dell'atlante e help secondari per illustrare le finestre contenenti parametri di non immediata comprensione;
- gli **output** (file di testo, grafici e mappe) sono **visibili dall'interno della procedura**, mediante due nuove finestre chiamate Risultati Grafici e Risultati Testi;
- nella finestra Geografia è ora possibile **selezionare un'area** scegliendo i comuni che la determinano. In altre parole l'area di analisi può essere costituita o da un'aggregato di comuni anche non adiacenti oppure da un cerchio definito da un centroide e da un raggio. Si faccia attenzione che nel caso di presenza di un'area il raggio dovrà essere tale da poter iscrivere completamente l'area nel cerchio.

### L'interfaccia utente

Per mandare in esecuzione il programma è necessario compiere i seguenti passi:

- lanciare il programma SAS;
- dalla finestra di Program Editor scegliere: *File->Open...->* Posizionarsi sull'unità lettore di CD, aprire il direttorio Atlante98 e scegliere il file *lanciami.sas.*;
- eseguire tale file con il comando *Locals->Submit* o utilizzando l'apposito pulsante  presente sulla barra degli strumenti della finestra di SAS;
- comparirà la seguente schermata:



L'interfaccia utente si articola in vari livelli che prendono avvio da un menù principale:



nel quale è assolutamente necessario, per prima cosa, definire il proprio disco e la propria directory di lavoro, nella quale verranno memorizzati, sotto forma di file i risultati della elaborazione stessa. Per default il direttorio dell'analisi viene posto sul disco C e si chiama *Analisi*. L'opzione di default di questa directory, come delle seguenti, può essere ridefinita modificando il file *lanciami.sas*.

L'interfaccia utente si articola in vari livelli che prendono avvio da questo menù principale nel quale, oltre a poter scegliere quale procedura SAS inviare in esecuzione (dopo aver indicato i parametri richiesti), consente di definire una

Area di Lavoro (posta per default a *C:\Analisi*) (un direttorio su disco fisso) nella quale verranno memorizzati i risultati (sotto forma di file) della elaborazione stessa.

I livelli in cui è stata articolata l'applicazione consistono nelle seguenti fasi:

- Geografia (Individuazione di un Area di indagine)
- Nosografia (Indicazione delle Cause di decesso)
- Statistica (Elaborazioni Statistiche)
- Directories (Gestione della eliminazione dei direttori-risultato)
- Risultati Grafici (visualizzazione dei grafici e delle mappe)
- Risultati testi (visualizzazione dei file di testo)

Prima di analizzare più in dettaglio gli argomenti sopra elencati, è necessario descrivere la soluzione adottata per gestire i file che vengono prodotti come risultato finale o come file di dati di appoggio per successive elaborazioni.

I livelli su cui si articola l'applicazione vengono tradotti in una struttura di direttori che vengono creati e gestiti sul disco fisso dell'utente; direttori che vengono successivamente riempiti con i file prodotti dalle singole procedure SAS.

Come primo passo, l'utente individua un Area di lavoro che viene tradotta in un direttorio principale sul disco fisso (esempio "*C:\ANALISI*").

Sotto tale direttorio vengono poi creati altri sottodirettori, uno per ogni Comune epicentro scelto; in questo sottodirettorio vengono memorizzati i file della elaborazione circa le Aree di indagine.

Per ogni sottodirettorio degli epicentri vengono creati altri sottodirettori, uno per ogni causa (o gruppi di cause) di decesso che viene elaborata. Infine, ogni sottodirettorio delle cause contiene altri sottodirettori, uno per ogni statistica effettuata sui dati della causa indicata.

La struttura sopra descritta viene memorizzata in un Data Set SAS (archivio SAS) che viene registrato sotto il direttorio dell'Area di lavoro. Questo Data Set genera delle liste di selezione che vengono di volta in volta proposte all'utente e che gli consentono di scegliere quali dati utilizzare in ogni passo dell'applicazione.

Se si clicca sul pulsante **Geografia**, comparirà la seguente schermata:



che permette la selezione dell'area di analisi. L'area di analisi è costituita o da un'aggregato di comuni anche non adiacenti o da un cerchio definito da un centroide e da un raggio.

Questa scelta deve essere effettuata cliccando sugli appositi pulsanti.

**Se si sceglie** di costruire o di modificare un'AREA, (Crea e modifica un'area specifica) si aprirà il menù "Selezione comuni dell'area" in cui si potrà richiamare, costruire, modificare un'area. Ad ogni area corrisponde comunque sempre un cerchio definito da un centroide e da un raggio; in questa finestra si può scegliere di definire automaticamente il cerchio con centroide il baricentro dei comuni e con cerchio tale che lambisca il comune più lontano dal baricentro. Da questa finestra si rientra nel menù **GEOGRAFIA** dove si potrà scegliere o modificare il cerchio di riferimento che, in presenza di un'area, dovrà esser sempre capace di contenerla.

Si può invece scegliere di analizzare direttamente **un cerchio di comuni**. In tal caso va data innanzitutto l'indicazione del Comune Centroide dell'Area di indagine che si vuole analizzare. La scelta viene effettuata attingendo direttamente dalla lista dei comuni d'Italia. In entrambi i casi, questa prima fase si conclude con la conferma dei dati ora descritti, mediante il tasto Verifica e conferma.

Tale operazione risulta obbligatoria per consentire alla procedura di creare il direttorio in cui verranno memorizzati i file elaborati successivamente. Per le altre opzioni da specificare in questa finestra si rimanda all'help in linea.



Il menù **NOSOGRAFIA** consente di indicare, in un'Area di indagine già precedentemente definita, sia gli anni da considerare sia la causa di decesso da analizzare. Pertanto, la prima informazione che viene richiesta all'utente è quella di scegliere l'Area di indagine tra tutte quelle costruite nel passo precedente. Questa scelta avviene attingendo il dato da una delle Aree di indagine precedentemente calcolate il cui nome viene presentato automaticamente a video. Anche in questo passo viene richiesto il nome da assegnare ad un direttorio (direttorio cause) sotto il quale la procedura inserisce i file relativi alla elaborazione che viene attivata da un pulsante di conferma. La scelta delle cause di decesso lascia la più ampia libertà di aggregazione utilizzando i codici delle cause di mortalità I.C.D. IX. Le modalità per indicare tali valori sono molteplici e per la spiegazione si rimanda all'help in linea. Novità della versione 5.3 è l'aggiunta del pulsante *Cause* che richiama l'elenco delle cause con il relativo codice. Per problemi di privacy non è possibile elaborare uno o due anni separatamente ma solo tre anni, anche non contigui, congiuntamente. In caso contrario uscirà una finestra di avvertimento e si dovrà riprendere l'elaborazione dalla scelta degli anni e della causa.



Il menù **STATISTICA** serve per l'elaborazione statistica dei file prodotti nei due menù precedenti.

La prima parte della finestra è relativa alla scelta delle directories in cui sono stati precedentemente memorizzati i files dell'area di indagine e della causa di decesso già elaborati ed all'indicazione della directory relativa alle statistiche che si elaboreranno.

La seconda propone una consistente quantità di valori e scelte al fine di orientare l'applicazione statistica nella direzione desiderata dall'utente. L'analisi statistica sulle cause di decesso all'interno dell'area di indagine predefinita consente molteplici scelte che l'utente può liberamente comporre.

La definizione di popolazione Stabile e Migrante può creare problemi nel caso in cui si calcolino dei tassi o dei rapporti tra tassi; infatti la definizione del censimento potrebbe non corrispondere a quella delle schede di mortalità. Di conseguenza si consiglia all'utente di utilizzare la specifica "*stabile o migrante*" solo se utilizza come indice l'SPMR, cioè se effettua una elaborazione di mortalità proporzionale.

Per molte delle analisi in questa finestra riportate, si aprirà una apposita finestra per la scelta dei parametri da impostare, con annesso un help per quelle più complesse.

Le **statistiche** che si possono effettuare sono le seguenti:

#### **Statistiche generali**

Elabora le statistiche globali dell'area e delle popolazioni di riferimento possibili.

#### **Istogramma indici comunali**

Elabora l'istogramma degli indici comunali e delle stime di densità opportunamente smussati. Viene anche segnato il valore dell'indice medio e della media degli indici.

#### **Rischi relativi per età' ( segue menù )**

Calcola il rischio relativo per ciascuna classe di età tra il cerchio analizzato ed una popolazione selezionabile.

#### **Autocorrelazione spaziale ( segue menù )**

Calcola l'I di Moran per gli indici originali e/o per le stime di densità. Esegue il calcolo dell'indice globale e dell'autocorrelogramma per un numero di lags definibile. L'ampiezza dei lags viene determinata automaticamente. Questa procedura richiede molto tempo di calcolo, soprattutto se i comuni analizzati sono molti.

#### **Lista dati e indici comunali**

Elabora l'elenco dei comuni per distanza dal centroide con i valori osservati, attesi, l'indice prescelto e la sua stima di densità.

#### **Allarmi sui singoli comuni (segue menù)**

Esegue un elenco dei comuni che singolarmente oltrepassano dei livelli di mortalità; si può definire il livello di probabilità ed il valore minimo degli osservati per attribuire un allarme.

#### **Test di struttura**

Elabora il test di Stone.

#### **Aggregati spaziali (segue menù con help)**

Elabora la ricerca di massimi locali e di aggregati.

#### **Mappa indice di frequenza**

Elabora la mappa con gli indici comunali originali.

#### **Mappa stime di densità'**

Elabora la mappa con le stime comunali di densità prescelte.

Le **stime di densità** che si possono applicare sono le seguenti:

- stime kernel;
- ripartizioni casuali iterative;
- indicatori bayesiani empirici;
- Trend surfaces.

Per ognuno di questi metodi si aprirà una finestra dove specificare i parametri.

Gli indicatori che si possono scegliere sono i seguenti:

- SMR (standardizzazione indiretta) con la popolazione come denominatore
- CMF (standardizzazione diretta)
- SPMR (standardizzazione indiretta) con i decessi per tutte le cause come denominatori, e quindi utili per un'analisi di mortalità proporzionale.

I valori dei rapporti standardizzati proporzionali di mortalità (SPMR) non differiscono molto dai valori dei rapporti tra tassi (SMR) a condizione che la mortalità generale non abbia variazioni eccessive e che la causa analizzata sia relativamente rara: per quest'ultima ragione è bene calcolare gli SMR per la mortalità generale e per i grandi settori nosologici, e gli SPMR per le singole cause di decesso. **Si consiglia l'utente inesperto di elaborare solo SMR e SPMR** (vedi help alla voce CMF).

Nella finestra delle statistiche sono stati inseriti due pulsanti:

**RISULTATI GRAFICI:** invia alla finestra per esaminare i files grafici elaborati.

**RISULTATI TESTI:** invia alla finestra per esaminare i files di testo elaborati.

### Requisiti hardware e software

Il programma gira in ambiente Windows '95 o NT nel quale è stato installato il prodotto SAS base nella version 6.12 con i moduli aggiuntivi Graph e IML. L'hardware consigliato è un pentium a 166Mz con 32Mega di memoria Ram. La configurazione consigliata permette l'elaborazione di aree di qualsiasi dimensione e in tempi ragionevoli si possono trattare anche zone di mille comuni.

### Bibliografia

- [1] Bollettino del CILEA n. 51 - M. Camnasio, C. Cislaghi, R. Ferrari, C. Montalbetti *Atlante epidemiologico portatile ad utilizzo flessibile*
- [2] A.D.Cliff e J.K.Ord, *Spatial Autocorrelation*, London, 1973
- [3] C.Cislaghi et al., *Un metodo per l'identificazione automatica di aggregati in mappe di eventi sanitari*
- [4] Roman E. et al., *A Comparison of Standardized and Proportional Mortality Ratios*, Statistics in Medicine, 1984
- [5] B.W.Silverman, *Density estimation*, London, 1986
- [6] C.Tasco et al., Spatial Components of Variability in Cancer Mortality Distribution, in *"Statistics of Spatial Processes: teory and applications"*, Bari,1993.
- [7] A. De Carli *"Confronto tra due indici di mortalità SPMR ed SMR"*, 1994
- [8] C. Cislaghi, M. Braga, A. Biggeri, *Analisi della Concentrazione Spaziale di Eventi per mezzo delle Superfici di Densità*, Epid Prev 1995
- [9] C.Cislaghi, M. Braga, G. Luppi, C. Tasco *Un metodo per l'identificazione automatica di aggregati in mappe di eventi sanitari*, Epid Prev, 1995
- [10] C.Cislaghi, G. Luppi, M. Camnasio, G.Benedetti, I. Covezzi *L'Atlante italiano di mortalità a livello comunale*, Epid Prev, 1995
- [11] C. Cislaghi et al., *Exploratory tools for disease mapping in geographical epidemiology*, Journal Article Statics in Medicine, 1995.

### Esempio di Elaborazione

Di seguito riportiamo un esempio di elaborazione, riportando anche il contenuto del file parametr.txt.

#### PARAMETRI IMPOSTATI

##### Geografia

- 1) popolazione di riferimento: 1991
- 2) usa cerchio di riferimento
- 3) centroide cerchio: CAPRIOLO
- 4) scelta del raggio di rif.  
Raggio in Km.: 134  
Analizzata area: LOMBARDIA

##### Nosografia

- 2) anni da analizzare: 1992 1993 1994
- 3) causa/cause di decesso: 140-239
- 5) causa intestazione: TUTTI I TUMORI

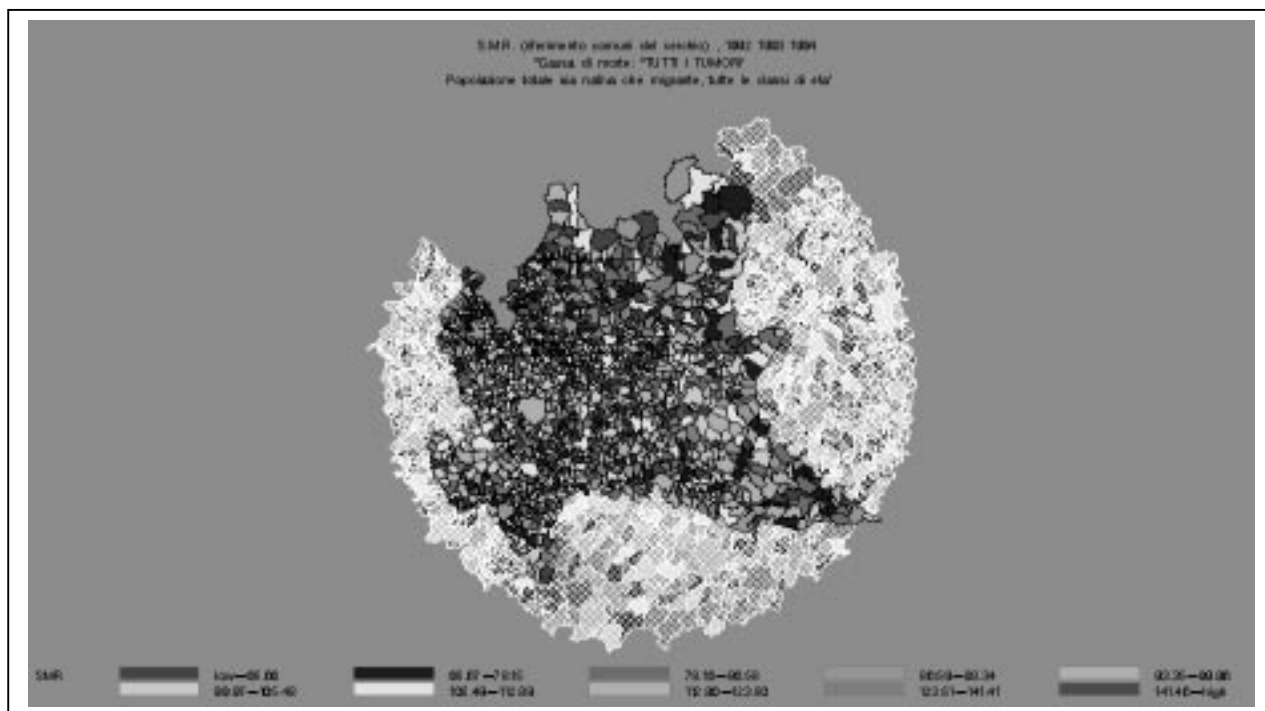
##### Statistiche

- 3) specifiche grafiche:  
colore  
device:BMP 922x516
- 4) selezione decessi da sviluppare  
sesso: totale  
tipo popolazione: totale  
classe di età: 0-95+
- 5) standard di riferimento  
sesso: totale  
tipo popolazione: totale  
popolazione di riferimento:cerchio
- 6) analisi  
statistiche generali
- 7) stime di densità  
kernel  
sigma della gaussiana: 1  
moltiplicatore del sigma: 1  
tipo fisso
- 8) denominatori  
popolazione: Mortalità (SMR)  
mappa indice di frequenza  
mappa stime di densità

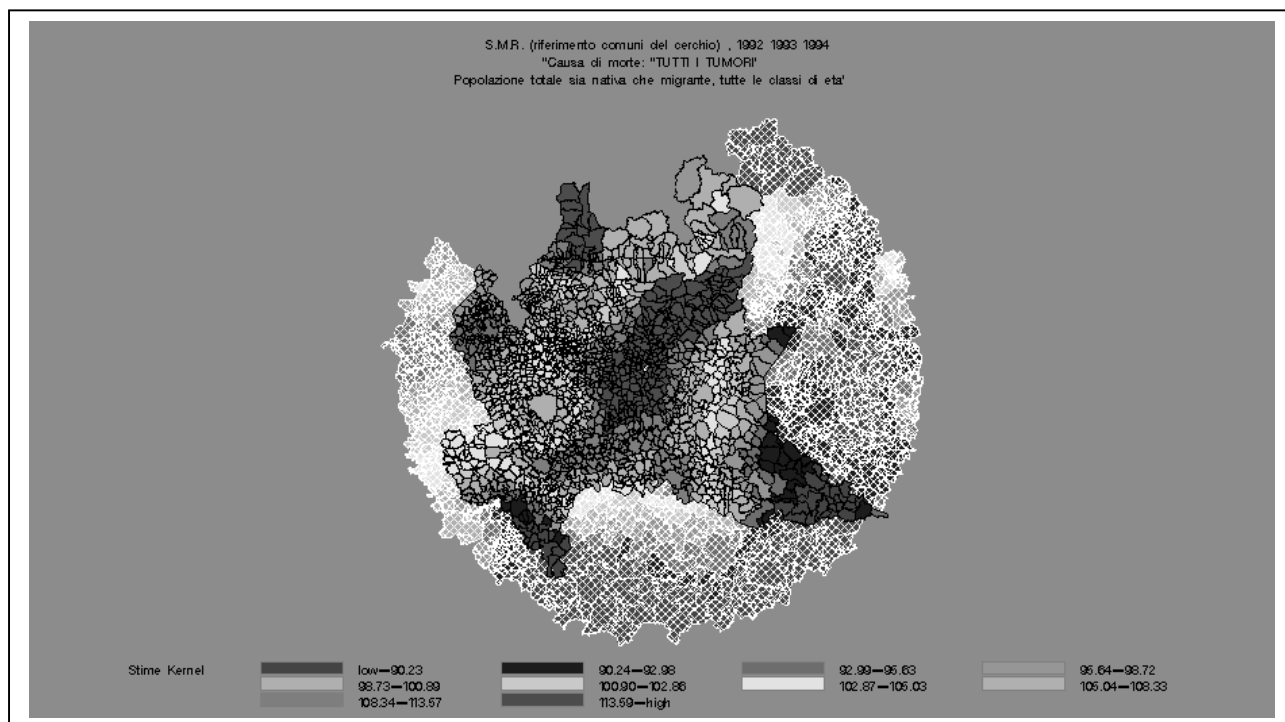
Una copia del CD contenente i programmi ed il data base criptato può essere richiesto alla Regione Emilia Romagna (dott.ssa Arianna Tedeschi - tel 051/283305)

Di seguito riportiamo le mappe prodotte.

**Mappa indice di frequenza**

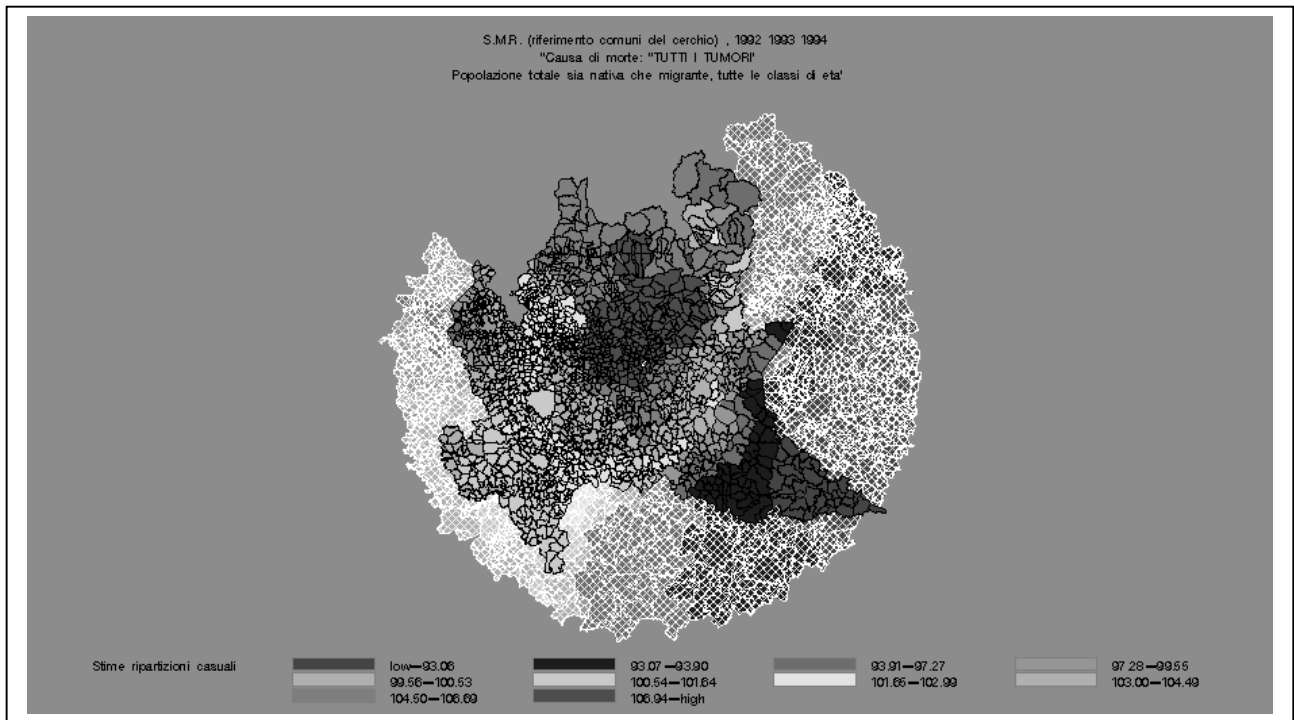


**Mappa stime di densità : Kernel**





**Mappa stime di densità : Ripartizioni casuali**



**Mappa stime di densità : Trend surface ordine del polinomio 4°**

