

Appendice 1

Approfondimento dei metodi statistici

APPROFONDIMENTO DEI METODI STATISTICI

TASSO STANDARDIZZATO PER ETÀ DI MORTALITÀ (TSDM) E DI OSPEDALIZZAZIONE (TSDH).

Il Tasso Standardizzato (TSD) è calcolato come media pesata dei tassi specifici per classi di età, ognuna delle quali assume un peso dato dalla struttura per età di una popolazione di riferimento considerata come standard. Questo metodo permette di confrontare tassi di popolazioni che hanno strutture per età diverse tra di loro. La popolazione di riferimento è quella europea fittizia, costante nel tempo e definita uguale per ciascun genere.

Il TSD si ottiene come:

$$\text{TSD} = \frac{\sum t_i N_i}{\sum N_i} * 10000$$

dove:

t_i è il tasso di mortalità o di ricovero per la i-esima classe di età nella popolazione in studio;

N_i la numerosità della popolazione standard nella i-esima classe di età.

RAPPORTO DI MORTALITÀ E DI OSPEDALIZZAZIONE STANDARDIZZATI PER ETÀ (RISPETTIVAMENTE SMR E SHR)

L'SMR e l'SHR si calcolano come rapporto tra casi osservati e casi attesi:

$$\text{SMR, SHR} = \frac{\sum d_i}{\sum T_i * n_i} * 100$$

dove

- d_i sono i decessi o i ricoveri osservati per la classe i-esima di età della popolazione in studio;
- T_i rappresentano i tassi di mortalità o di ricovero della popolazione di riferimento per classe di età;
- $\sum T_i * n_i$ sono i decessi o i ricoveri attesi per la classe i-esima di età della popolazione in studio.

I casi attesi sono il numero di casi che ci si attenderebbe se la popolazione in studio sperimentasse il tasso di mortalità della popolazione di riferimento al netto dell'effetto attribuibile all'età.

Un valore di SMR vicino a 100 indica che il numero di osservati è simile all'atteso, mentre un SMR pari a 200 indica un rischio di morte doppio nell'area in studio rispetto all'area di riferimento.

L'intervallo di confidenza dell'SMR è dato dal metodo approssimato di Byar.

ANALISI E TEST DEGLI ANDAMENTI TEMPORALI (TREND)

Se gli eventi si presentano nel tempo, è utile modellare il tasso usando variabili esplicative. Questo implica un processo di Poisson non omogeneo.

Di seguito è riportato il procedimento per gli SMR che vale anche per gli SHR.

L'SMR è per definizione pari a $\frac{d}{e}$ dove d sono i casi osservati ed e sono i casi attesi.

Essendo l'SMR un indicatore non negativo, allora deve necessariamente essere modellato usando modelli lineari nel logaritmo, cioè:

$$\log \frac{d_i}{e_i} = \alpha + \beta x \Rightarrow \log d_i = \log e_i + \alpha + \beta x$$

- β è il coefficiente angolare della retta di regressione;
- x assume valori interi compresi tra 1 (1° periodo) e 3 (3° periodo);
- $\log e_i$ è chiamato *offset* e non è un parametro.

Di seguito è riportato il significato di β :

$$\log \text{SMR}(x) = \alpha + \beta x \Rightarrow \text{SMR}(x) = e^\alpha e^{\beta x}$$

e

$$\log(\text{SMR}(x+1)) = \alpha + \beta(x+1) \Rightarrow \text{SMR}(x+1) = e^\alpha e^{\beta x} e^\beta$$

Quindi:

$$\log(\text{SMR}(x+1)) = e^\beta \log(\text{SMR}(x))$$

$$g = \frac{SMR_{x+1}}{SMR_x} = \frac{\exp(\alpha + \beta + \beta x)}{\exp(\alpha + \beta x)} = \exp(\beta)$$

Un incremento di x di un'unità provoca un incremento del rischio di e^β con un effetto che è chiamato in tal caso *moltiplicativo*. Quindi l'esponentiale di β rappresenta la statistica test sul *trend* degli SMR nei tre periodi considerati. L'ipotesi di non presenza di *trend* ($\beta = 0$) è stata verificata con il test Z.

INDICATORI STATISTICI BAYESIANI

Il BMR è calcolato mediante la somma di due effetti casuali:

$$BMR_i = e^{\alpha + U_i + V_i}$$

- $V[i] \sim N(0, \text{prec.v})$, componente a priori normale del rischio che tiene conto della sovradisersione non spaziale nei dati;
- $U[i] \sim \text{car.N}(\text{adj}[i], \text{prec.u})$, componente normale del rischio che tiene conto della variabilità strutturata spazialmente, la distribuzione a priori *car.normal* (Conditional Auto Regression) permette di stimare il rischio in un comune in funzione delle stime calcolate nei comuni adiacenti: più è alta la correlazione spaziale nella distribuzione della causa di morte, maggiore sarà il peso dei comuni adiacenti nella stima del rischio relativo; più i rischi stimati sono basati su pochi casi, più essi si avvicinano ad una media locale dei rischi fortemente influenzata dalle stime dei comuni adiacenti.

Il calcolo dei BMR e la ricerca delle informazioni a priori utilizzate (Mollié, 2000), prevedono complicati calcoli computazionali. Gli intervalli di "credibilità" al 95% (IC 95%) vengono stimati mediante procedure iterative di simulazione di Monte Carlo (Catene di Markov) e di campionamento di Gibbs (Gelfand et al., 1990) e sono serviti per valutare l'ampiezza dell'incertezza della stima e la significatività statistica dell'eccesso o del difetto di rischio.

Il controllo della convergenza del metodo è stato effettuato attraverso una serie di test statistici (Gelman et al., 1992).

Il modello è stato implementato e compilato tramite il software WinBUGS (Spiegelhalter et al., 1999).

ANALISI DI TENDENZA ALL'ADDENSAMENTO DEGLI EVENTI (CLUSTERING)

L'analisi è stata effettuata mediante la tecnica *Spatial Scan Statistic* di Kulldorff (Kulldorff & Nagarwalla, 1995).

Il metodo utilizzato ipotizza che il numero di casi sia distribuito mediante un modello poissoniano ed è basato sullo spostamento di una finestra ellittica con assi variabili su tutto il territorio in studio; la finestra è centrata su ognuna delle possibili coordinate (x_i, y_i) rappresentanti il comune. Il metodo identifica gli aggregati comunali con SMR più elevato e seleziona quelli statisticamente significativi in base alla massimizzazione del rapporto di verosimiglianza. In dettaglio, per ogni finestra identificata viene testata l'ipotesi nulla (H_0 : nessuna differenza tra il numero di casi osservati nel *cluster* e quelli attesi) mediante il test del rapporto di verosimiglianza (LRT). Il *cluster* è considerato statisticamente significativo quando $p < 0,05$.

ANALISI DI REGRESSIONE DI POISSON PER LO STUDIO DI CORRELAZIONE TRA INQUINANTI ED EVENTI SANITARI

METODOLOGIA PER IL CALCOLO DEL RAPPORTO TRA SMR

I rapporti tra SMR vengono stimati utilizzando l'analisi di regressione lineare di Poisson.

L'ipotesi di partenza è che il numero di eventi siano distribuiti secondo una distribuzione di Poisson.

Partendo dal modello di regressione: $\log(O_i/A_i) = \beta_1 + \beta_2 X_2 + \beta_3 X_3$,

dove:

- i assume valori da 1 a 3 (prima, seconda e terza classe di concentrazione media dell'indicatore ambientale considerato)
- O_i è il valore osservato dei morti/ricoverati del gruppo di comuni nell' i -esima classe di esposizione,
- A_i rappresenta il numero morti/ricoverati attesi rispetto al tasso di riferimento regionale del gruppo di comuni nell' i -esima classe di esposizione,
- X_2 è una variabile dicotomica che prende valore 1 quando $X=2$ e 0 quando $X \neq 2$,

- X_3 è una variabile dicotomica che prende valore 1 quando $X=3$ e 0 quando $X \neq 3$,

si ottiene che il rischio stimato nei diversi profili di esposizione dipende dalla classe di appartenenza ($X=1^a$ classe, 2^a classe, 3^a classe); il rischio SMR_i per il gruppo di comuni con X pari ad i è dato dalla seguente formula:

$$SMR_i = O_i / A_i = e^{\beta_1 + \beta_2 X_i + \beta_3 X_i^2}$$

Per stimare l'SMR nella prima classe ($X=1$): $SMR_1 = O_1 / A_1 = e^{\beta_1}$.

Per stimare l'SMR nella seconda classe ($X=2$): $SMR_2 = O_2 / A_2 = e^{\beta_1 + \beta_2}$.

Per stimare l'SMR nella terza classe ($X=3$): $SMR_3 = O_3 / A_3 = e^{\beta_1 + \beta_2}$.

Il modello fornisce il rapporto tra SMR, $SMR_i / SMR_1 = e^{\beta_i}$, per $i=2, 3$.

METODOLOGIA PER IL CALCOLO DEL TREND

L'andamento del rischio di mortalità/ricoverati al variare dell'indicatore ambientale (*trend*) viene stimato utilizzando l'analisi di regressione lineare di Poisson.

L'ipotesi di partenza è che il numero di eventi sia distribuito secondo una distribuzione di Poisson.

Partendo dal modello di regressione $\log(O_i / A_i) = \alpha + \beta X$,

dove:

- i assume valori da 1 a 3 (prima, seconda e terza classe di concentrazione media dell'indicatore ambientale considerato)
- O_i è il valore osservato dei morti/ricoverati del gruppo di comuni nell' i -esima classe di esposizione
- A_i rappresenta il numero morti/ricoverati attesi rispetto al tasso di riferimento regionale del gruppo di comuni nell' i -esima classe di esposizione
- X è una variabile categorica che prende valore 1, 2, 3 a seconda della classe di appartenenza del comune

si ottiene che l'aumento stimato di rischio all'aumentare unitario della X ($1 \rightarrow 2$, $2 \rightarrow 3$) è dato dalla seguente formula:

$$SMR = O/A = e^{\alpha + \beta X}.$$

Per stimare l'SMR nella prima classe (X=1): $SMR_1 = O_1/A_1 = e^{\alpha + \beta}$.

Per stimare l'SMR nella seconda classe (X=2): $SMR_2 = O_2/A_2 = e^{\alpha + 2\beta}$.

Per stimare l'SMR nella terza classe (X=3): $SMR_3 = O_3/A_3 = e^{\alpha + 3\beta}$.

La stima del *trend* è quindi data da $SMR_{x+1}/SMR_x = e^{\beta}$.