



Associazione Italiana di Epidemiologia

Metodi statistici per l'epidemiologia

(Moderatori: Lorenzo Richiardi e Elias Allara)

- Sorveglianza epidemiologica sulla frequenza di tumori (Lorenzo Cecconi)
- Disease mapping in aree geografiche di piccole dimensioni. Confronto tra due modelli bayesiani (Nicola Bartolomeo)
- Strumenti per l'epidemiologia nutrizionale nell'era della globalizzazione (Maria Parpinel)
- Analisi geografica dell'incidenza di tumori infantili nelle province di Firenze e Prato (Giorgia Stoppa)
- Mortalità per incidenti stradali e suicidi tra i maschi in Toscana utilizzando la nuova popolazione standard europea (Tommaso Grassi)
- Applicazione del metodo di imputazione multipla per blocchi monotoni ordinati nell'analisi di serie storiche (Michela Baccini)
- L'uso del propensity score nella valutazione degli effetti della esposizione cronica a inquinamento atmosferico nello studio longitudinale romano (Giulia Cesaroni)

Sorveglianza epidemiologica sulla frequenza di tumori.

Lorenzo Cecconi*, D. Catelan, G. Stoppa, L. Miligi, S. Piro, E.Croce,
G. Manneschi, A. Barchielli, A. Biggeri

*Dipartimento di Statistica, Informatica, Applicazioni «G.Parenti»
Università di Firenze

Stop and Listen to the People: An Enhanced Approach to Cancer Cluster Investigations

(2014 American Journal of Public Health)

| Brian W. Simpson, MPH, MA, Patti Truant, MPH, CPH, and Beth A. Resnick, MPH, CPH

“Every year in the United States, more than 1000 reports of suspected cancer clusters are made to public health authorities”

“A review of 428 suspected cancer cluster investigations in the United States from 1990 to 2011 found that only 1 was linked to a clear cause and that was occupation related”.

1990 editorial in The Lancet noted, “The cluster alarm can become the epidemiologist’s nightmare”.

However, people who have lost loved ones to what they believe is a cancer cluster have a very different definition of “nightmare.” They want to know what caused the suspected cluster, who is at fault, and how they can protect themselves and their loved ones in the future.

Fort Detrick community reaction in a public meeting:

[...] a resident declared to a health official at a public meeting,

“I have no confidence in your confidence intervals.”

Letter from the Editors: Florida Pediatric Cancer Studies

(2014 Statistics and Public Policy)

Dear Readers,

This journal attempts to provide a forum where good statistical methods are applied to data sets that inform public policy. To that end, we have provided records on the locations of pediatric cancer cases in the state of Florida to five different groups of researchers, and invited them to apply their different methodologies to discover whether or not there are regions for which the incidence of pediatric cancer is relatively high.

Over the next few weeks, a series of five papers, along with an introduction and a discussion, will be posted on-line at the *Statistics and Public Policy* website. Each paper is marked by the tag “Analyses of Florida Pediatric Cancer Data.”

These papers study the Florida data from multiple perspectives, under different sets of assumptions. We encourage interested parties to read those papers, weight the evidence, and decide what next steps, if any, are warranted.

Sincerely,
David Banks and Sharon Lohr

Problema decisionale - test multipli

N test di ipotesi $H^1, \dots, H^N: H_0^i: \theta_i = 1$

	H_0 falsa	H_0 vera	
Rifiuto H_0	Vero positivo	Falso positivo	R
Accetto H_0	Falso negativo	Vero negativo	N-R
	N_1	N_0	N

$$\alpha = P(\text{rifiuto } H_0 \mid H_0 \text{ è vera}) \quad \beta = P(\text{accetto } H_0 \mid H_0 \text{ falsa})$$

Cosa ci interessa?

False discovery proportion = $P(H_0 \text{ vera} \mid \text{rifiuto } H_0) = \text{FP}/R$

False negative proportion = $P(H_0 \text{ falsa} \mid \text{accetto } H_0) = \text{FN}/(N-R)$

Regole di decisione – approcci classici ($\alpha = 0.05$)

Correzione di Bonferroni:

- rifiuta H_0^i se $p^{(i)}$ -value $< \alpha/N$

Controlla il FWER, la probabilità di commettere almeno un errore di I tipo su N test

Procedura di Benjamini e Hochberg (BH per test ind.) e Yekutieli (BY con correz.)

- ordinare i p-value $p^{(1)} < \dots < p^{(n)}$

- Confrontare $p^{(i)}$ con $\alpha * i/N$

- Sia $r = \max\{ i \text{ tale che } p^{(i)} \leq \alpha * i/N \}$

- Rifiuta $H_0^{(1)}, \dots, H_0^{(r)}$

Controlla il false discovery rate, E(FDP), minimizzando il false non-discovery rate

Regole di decisione – approcci bayesiani

Bayesian approach to FDR

- Storey – empirical bayes (q-values)
- Efron – two groups probabilistic model

Indichiamo con $\pi_0 = P(H_0 \text{ è vera})$

$$\begin{aligned} H_0 \text{ vera} &\rightarrow \theta = \theta_{H_0} \\ H_0 \text{ falsa} &\rightarrow \theta = \theta_{H_a} \end{aligned}$$

$$\theta = \theta_{H_0} \pi_0 + \theta_{H_a} (1 - \pi_0)$$

Il modello Bayesiano gerarchico (3livelli)

$$Y_i \sim \text{Poisson}(\theta_i E_i)$$

$$\begin{aligned} \mu_i &= \mu_{0,i} r_i + \mu_{1,i} (1 - r_i) \\ \mu_{0,i} &= 0 \quad \mu_{1,i} = \log(\theta_i) \end{aligned}$$

$$\begin{aligned} \theta_i &\sim \text{Gamma}(\kappa, \nu) \\ r_i &\sim \text{Bernoulli}(\pi_i) \end{aligned}$$

$$\pi_i \sim \text{Beta}(a, b)$$

La stima a posteriori di π_i è chiamata probabilità di classificazione, $\pi_i|Y$

$$\begin{aligned} \pi_i|Y &= \text{P}(\text{FP}) \\ 1 - \pi_i|Y &= \text{P}(\text{VP}) \quad \text{se rifiuto } H_0^i \end{aligned}$$

$$\begin{aligned} \pi_i|Y &= \text{P}(\text{VN}) \\ 1 - \pi_i|Y &= \text{P}(\text{FN}) \quad \text{se accetto } H_0^i \end{aligned}$$

$$P(H_0) \simeq 0.9 \quad (\text{Scott and Berger 2006})$$

Due problemi

1. Specificazione di una distribuzione a-priori per H_0
« quanti comuni su N mi aspetto con un eccesso di casi? »

>> Elicitazione di una distribuzione a-priori

2. Problema di classificazione in base a $\pi_i|Y$: rifiuto H_0^i se $\pi_i|Y < \text{cut-off}$
« Come scelgo il cut-off? 0.05? 0.10? 0.20? »

>> Definire una funzione di perdita per FP, FN e determinare il cut-off che minimizza la funzione.

Due esempi

- Il caso «Salto di Quirra»
 - Informazioni a-priori sono disponibili per alcuni comuni. La zona a rischio è individuata nell'area del poligono militare.
 - Una diversa a-priori deve essere specificata per questi comuni!
- Tumori infantili
 - Nessuna informazione a priori su comuni specifici
 - È opportuno definire una funzione di utilità FP/FN a seconda dell'obiettivo dello studio

Due opportunità!

Specificazione di una distribuzione a-priori per H_0 «*area specifica*»

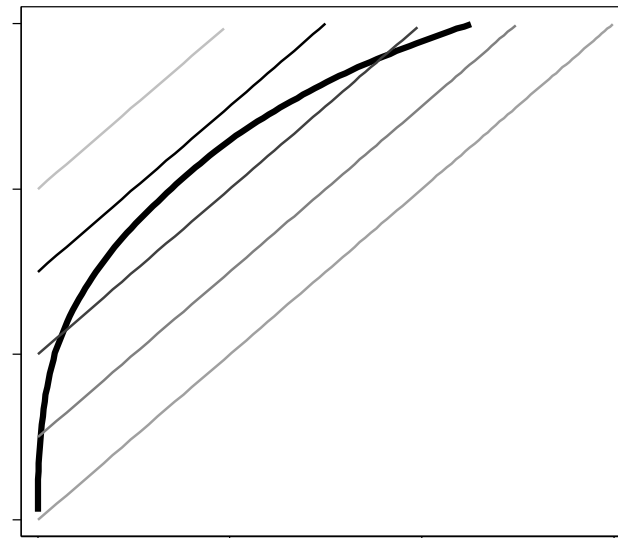
$$\begin{aligned} Y_i &\sim \text{Poisson}(\theta_i E_i) \\ \mu_i &= \mu_{0,i} r_i + \mu_{1,i} (1 - r_i) \\ \mu_{0,i} &= 0 \\ \mu_{1,i} &= \log(\theta_i) \\ \theta_i &\sim \text{Gamma}(\kappa, \nu) \\ r_i &\sim \text{Bernoulli}(\pi_i) \end{aligned}$$

$$\begin{aligned} \pi_i &\sim \text{Beta}(a', b') \text{ se } i \in R \\ \pi_i &\sim \text{Beta}(a'', b'') \text{ se } i \notin R \end{aligned}$$

Specificazione di una funzione di perdita

$$\text{Es. } L_N = c_1 FP + c_2 FN$$

>> determinazione del cut-off ottimale tramite curve ROC (tangente alla curva con pendenza c_1/c_2)



Conclusioni

- È possibile incorporare nella modellazione statistica informazioni a priori per esplorare la presenza di cluster di malattia all'interno di un insieme di comuni correggendo simultaneamente per i test multipli e ottimizzando il rapporto falsi positivi/falsi negativi.
- Le informazioni a priori locali/globali:
 - $P(H_0)$
 - rapporto FP/FN