Taxa Bayesiana Empírica

Milton Pifano

Orientador : Marcos Oliveira Prates
Co-orientador : Adrian Pablo Hinojosa Luna

Star4Good - 08 de Abril, 2016

Objetivos

1. Permitir comparações entre diferentes populações no espaço ou no tempo, variáveis devem ser padronizadas.
2. Padronizar as população de risco por tamanho, faixa etária e sexo, etc.
3. Padronização pode ser também por área, por tempo de exposição, etc. empresas, comércio entre países, etc.

Pad. tamanho da população

- i = índice da área. $i = 1,...,N$
- e_i = número de eventos em i
- pop_i = população sob risco na área i
- $r_i = e_i \ / \ pop_i$ = taxa na área i.
  Às vezes, usa-se $t_i = 100.000 * r_i$, taxa por 100 mil habitantes na área i

Taxa como variável aleatória

- Número observado e_i de eventos na área i é variável aleatória.
- $e_i \sim Poisson(E_i)$,onde E_i número esperado na área i
- Se o risco é constante na região (conjunto de áreas), então $E_i = r * pop_i$, onde $r = \frac{\sum e_i}{\sum pop_i}$

Estimação em áreas pequenas

- Valores extremos ocorrem nas áreas com pequenas populações.
- O que mais chama a atenção num mapa (os valores extremos), é o menos confiável !
- As maiores oscilações não estarão, em geral, associadas com variações no risco. Serão apenas flutuação aleatória casual.

Solução : Abordagens Bayesianas

- Empírica: fácil de implementar
- Puramente bayesiana: dificuldades na definição nos parâmetros da distribuição *a priori*.

Abordagem Bayesiana empírica

- Assume que riscos das diferentes áreas não são totalmente "desconectados" e assim pede uma "força para os vizinhos".
- Idéia: contrair taxa em direção à média global. Fator de contração depende da população da área.
- Cada área i possui uma taxa e_i desconhecida. Embora diferentes, estas taxas possuem certa estrutura.

Objetivo : Estimar θ

- Numa área, observa-se um número aleatório e_i de casos.
- NÃO se assume risco constante: e_i possui distribuição de Poisson com número esperado de casos igual a $\theta_i * pop_i$
- Assume-se que as taxas $theta_i$ possuem distribuição (não específica) com média m e variância V.
- Qual é a melhor estimativa $\hat{\theta}_i$ possível dos θ_i? Melhor no sentido de minimizar a soma dos erros de estimação de todas as áreas $\sum_i(\hat{\theta}_i - \theta_i)^2$
- Buscar estimativa ótima APENAS DENTRE as estimativas que podem ser escritas como médias ponderadas de m e de e_i (taxa observada na área i)

Estimação de θ : Solução

- $\hat{\theta}_i = w_i * r_i + (1 - w_i) * m$, onde $w_i = \frac{V}{V + \frac{m}{pop_i}}$
- Problema : V e m são desconhecidos.
- Bayes empírico estima estes valores a partir dos dados. Daí o nome.
- $m = \frac{\sum_i e_i}{\sum_i pop_i}$ = Taxa da região
- $V = \frac{\sum_i pop_i * (r_i - m)^2}{\sum_i pop_i} - \frac{m}{popmedia}$

Finalmente

- Fazer estimativa bayesiana localmente: contrair em direção a uma média local e não, a uma média global
- Basta aplicar o método anterior em cada área considerando como "região" a sua vizinhança.
- Isto é equivalente a supor que as taxas da vizinhança da área i possuem média m_i e variância V_i

Fim

OBRIGADO